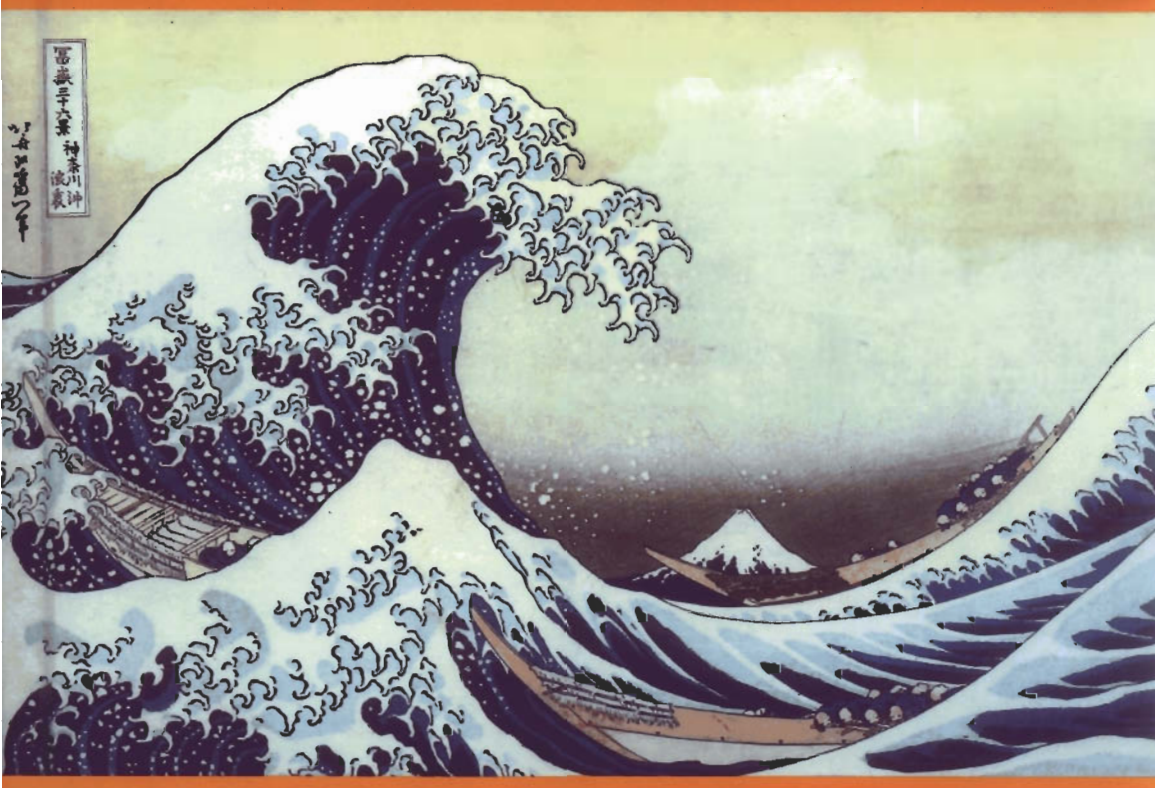
a wavelet tour of signal processing



Second Edition

Stéphane Mallat

A WAVELET TOUR OF SIGNAL PROCESSING

A WAVELET TOUR OF SIGNAL PROCESSING

Second Edition

Stéphane Mallat

École Polytechnique, Paris
Courant Institute, New York University

ACADEMIC PRESS

A Harcourt Science and Technology Company
San Diego San Francisco New York
Boston London Sydney Tokyo

This book is printed on acid-free paper. ♾

Academic Press
An imprint of Elsevier
525 B Street, Suite 1900, San Diego, California 92101-4495, USA
http://www.academicpress.com

Academic Press
84 Theobald's Road, London WC1X 8RR, UK
http://www.academicpress.com

ISBN: 0-12-466606-X

A catalogue record for this book is available from the British Library

Produced by HWA Text and Data Management, Tunbridge Wells
Printed in the United Kingdom at the University Press, Cambridge
PRINTED IN THE UNITED STATES OF AMERICA
03 04 05 06 9 8 7 6 5 4

A mes parents,
Alexandre et Francine

Contents

II

FOURIER KINGDOM

III

DISCRETE REVOLUTION

IV

TIME MEETS FREQUENCY

V

FRAMES

VI
WAVELET ZOOM

VII
WAVELET BASES

VIII

WAVELET PACKET AND LOCAL COSINE BASES

IX
AN APPROXIMATION TOUR

X
ESTIMATIONS ARE APPROXIMATIONS

XI
TRANSFORM CODING

Appendix A
MATHEMATICAL COMPLEMENTS

Appendix B
SOFTWARE TOOLBOXES

Preface

Facing the unusual popularity of wavelets in sciences, I began to wonder whether this was just another fashion that would fade away with time. After several years of research and teaching on this topic, and surviving the painful experience of writing a book, you may rightly expect that I have calmed my anguish. This might be the natural self-delusion affecting any researcher studying his corner of the world, but there might be more.

Wavelets are not based on a "bright new idea", but on concepts that already existed under various forms in many different fields. The formalization and emergence of this "wavelet theory" is the result of a multidisciplinary effort that brought together mathematicians, physicists and engineers, who recognized that they were independently developing similar ideas. For signal processing, this connection has created a flow of ideas that goes well beyond the construction of new bases or transforms.

A Personal Experience At one point, you cannot avoid mentioning who did what. For wavelets, this is a particularly sensitive task, risking aggressive replies from forgotten scientific tribes arguing that such and such results originally belong to them. As I said, this wavelet theory is truly the result of a dialogue between scientists who often met by chance, and were ready to listen. From my totally subjective point of view, among the many researchers who made important contributions, I would like to single out one, Yves Meyer, whose deep scientific vision was a major ingredient sparking this catalysis. It is ironic to see a French pure mathematician, raised in a Bourbakist culture where applied meant trivial, playing a central role

along this wavelet bridge between engineers and scientists coming from different disciplines.

When beginning my Ph.D. in the U.S., the only project I had in mind was to travel, never become a researcher, and certainly never teach. I had clearly destined myself to come back to France, and quickly begin climbing the ladder of some big corporation. Ten years later, I was still in the U.S., the mind buried in the hole of some obscure scientific problem, while teaching in a university. So what went wrong? Probably the fact that I met scientists like Yves Meyer, whose ethic and creativity have given me a totally different view of research and teaching. Trying to communicate this flame was a central motivation for writing this book. I hope that you will excuse me if my prose ends up too often in the no man's land of scientific neutrality.

A Few Ideas Beyond mathematics and algorithms, the book carries a few important ideas that I would like to emphasize.

- *Time-frequency wedding* Important information often appears through a simultaneous analysis of the signal's time and frequency properties. This motivates decompositions over elementary "atoms" that are well concentrated in time and frequency. It is therefore necessary to understand how the uncertainty principle limits the flexibility of time and frequency transforms.

- *Scale for zooming* Wavelets are scaled waveforms that measure signal variations. By traveling through scales, zooming procedures provide powerful characterizations of signal structures such as singularities.

- *More and more bases* Many orthonormal bases can be designed with fast computational algorithms. The discovery of filter banks and wavelet bases has created a popular new sport of basis hunting. Families of orthogonal bases are created every day. This game may however become tedious if not motivated by applications.

- *Sparse representations* An orthonormal basis is useful if it defines a representation where signals are well approximated with a few non-zero coefficients. Applications to signal estimation in noise and image compression are closely related to approximation theory.

- *Try it non-linear and diagonal* Linearity has long predominated because of its apparent simplicity. We are used to slogans that often hide the limitations of "optimal" linear procedures such as Wiener filtering or Karhunen-Loève bases expansions. In sparse representations, simple non-linear diagonal operators can considerably outperform "optimal" linear procedures, and fast algorithms are available.

WAVELAB and LASTWAVE Toolboxes Numerical experimentations are necessary to fully understand the algorithms and theorems in this book. To avoid the painful programming of standard procedures, nearly all wavelet and time-frequency algorithms are available in the WAVELAB package, programmed in MATLAB. WAVELAB is a freeware software that can be retrieved through the Internet. The correspondence between algorithms and WAVELAB subroutines is explained in Appendix B. All computational figures can be reproduced as demos in WAVELAB. LASTWAVE is a wavelet signal and image processing environment, written in C for X11/Unix and Macintosh computers. This stand-alone freeware does not require any additional commercial package. It is also described in Appendix B.

Teaching This book is intended as a graduate textbook. It took form after teaching "wavelet signal processing" courses in electrical engineering departments at MIT and Tel Aviv University, and in applied mathematics departments at the Courant Institute and École Polytechnique (Paris).

In electrical engineering, students are often initially frightened by the use of vector space formalism as opposed to simple linear algebra. The predominance of linear time invariant systems has led many to think that convolutions and the Fourier transform are mathematically sufficient to handle all applications. Sadly enough, this is not the case. The mathematics used in the book are not motivated by theoretical beauty; they are truly necessary to face the complexity of transient signal processing. Discovering the use of higher level mathematics happens to be an important pedagogical side-effect of this course. Numerical algorithms and figures escort most theorems. The use of WAVELAB makes it particularly easy to include numerical simulations in homework. Exercises and deeper problems for class projects are listed at the end of each chapter.

In applied mathematics, this course is an introduction to wavelets but also to signal processing. Signal processing is a newcomer on the stage of legitimate applied mathematics topics. Yet, it is spectacularly well adapted to illustrate the applied mathematics chain, from problem modeling to efficient calculations of solutions and theorem proving. Images and sounds give a sensual contact with theorems, that can wake up most students. For teaching, formatted overhead transparencies with enlarged figures are available on the Internet:

`http://www.cmap.polytechnique.fr/~mallat/Wavetour_fig/`.

Francois Chaplais also offers an introductory Web tour of basic concepts in the book at

`http://cas.ensmp.fr/~chaplais/Wavetour_presentation/`.

Not all theorems of the book are proved in detail, but the important techniques are included. I hope that the reader will excuse the lack of mathematical rigor in the many instances where I have privileged ideas over details. Few proofs are long; they are concentrated to avoid diluting the mathematics into many intermediate results, which would obscure the text.

Course Design Level numbers explained in Section 1.5.2 can help in designing an introductory or a more advanced course. Beginning with a review of the Fourier transform is often necessary. Although most applied mathematics students have already seen the Fourier transform, they have rarely had the time to understand it well. A non-technical review can stress applications, including the sampling theorem. Refreshing basic mathematical results is also needed for electrical engineering students. A mathematically oriented review of time-invariant signal processing in Chapters 2 and 3 is the occasion to remind the student of elementary properties of linear operators, projectors and vector spaces, which can be found in Appendix A. For a course of a single semester, one can follow several paths, oriented by different themes. Here are a few possibilities.

One can teach a course that surveys the key ideas previously outlined. Chapter 4 is particularly important in introducing the concept of local time-frequency decompositions. Section 4.4 on instantaneous frequencies illustrates the limitations of time-frequency resolution. Chapter 6 gives a different perspective on the wavelet transform, by relating the local regularity of a signal to the decay of its wavelet coefficients across scales. It is useful to stress the importance of the wavelet vanishing moments. The course can continue with the presentation of wavelet bases in Chapter 7, and concentrate on Sections 7.1-7.3 on orthogonal bases, multiresolution approximations and filter bank algorithms in one dimension. Linear and non-linear approximations in wavelet bases are covered in Chapter 9. Depending upon students' backgrounds and interests, the course can finish in Chapter 10 with an application to signal estimation with wavelet thresholding, or in Chapter 11 by presenting image transform codes in wavelet bases.

A different course may study the construction of new orthogonal bases and their applications. Beginning with the wavelet basis, Chapter 7 also gives an introduction to filter banks. Continuing with Chapter 8 on wavelet packet and local cosine bases introduces different orthogonal tilings of the time-frequency plane. It explains the main ideas of time-frequency decompositions. Chapter 9 on linear and non-linear approximation is then particularly important for understanding how to measure the efficiency of these bases, and for studying best bases search procedures. To illustrate the differences between linear and non-linear approximation procedures, one can compare the linear and non-linear thresholding estimations studied in Chapter 10.

The course can also concentrate on the construction of sparse representations with orthonormal bases, and study applications of non-linear diagonal operators in these bases. It may start in Chapter 10 with a comparison of linear and non-linear operators used to estimate piecewise regular signals contaminated by a white noise. A quick excursion in Chapter 9 introduces linear and non-linear approximations to explain what is a sparse representation. Wavelet orthonormal bases are then presented in Chapter 7, with special emphasis on their non-linear approximation properties for piecewise regular signals. An application of non-linear diagonal operators to image compression or to thresholding estimation should then be studied in some detail, to motivate the use of modern mathematics for understanding these problems.

A more advanced course can emphasize non-linear and adaptive signal processing. Chapter 5 on frames introduces flexible tools that are useful in analyzing the properties of non-linear representations such as irregularly sampled transforms. The dyadic wavelet maxima representation illustrates the frame theory, with applications to multiscale edge detection. To study applications of adaptive representations with orthonormal bases, one might start with non-linear and adaptive approximations, introduced in Chapter 9. Best bases, basis pursuit or matching pursuit algorithms are examples of adaptive transforms that construct sparse representations for complex signals. A central issue is to understand to what extent adaptivity improves applications such as noise removal or signal compression, depending on the signal properties.

Responsibilities This book was a one-year project that ended up in a never will finish nightmare. Ruzena Bajcsy bears a major responsibility for not encouraging me to choose another profession, while guiding my first research steps. Her profound scientific intuition opened my eyes to and well beyond computer vision. Then of course, are all the collaborators who could have done a much better job of showing me that science is a selfish world where only competition counts. The wavelet story was initiated by remarkable scientists like Alex Grossmann, whose modesty created a warm atmosphere of collaboration, where strange new ideas and ingenuity were welcome as elements of creativity.

I am also grateful to the few people who have been willing to work with me. Some have less merit because they had to finish their degree but others did it on a voluntary basis. I am thinking of Amir Averbuch, Emmanuel Bacry, François Bergeaud, Geoff Davis, Davi Geiger, Frédéric Falzon, Wen Liang Hwang, Hamid Krim, George Papanicolaou, Jean-Jacques Slotine, Alan Willsky, Zifeng Zhang and Sifen Zhong. Their patience will certainly be rewarded in a future life.

Although the reproduction of these 600 pages will probably not kill many trees, I do not want to bear the responsibility alone. After four years writing and rewriting each chapter, I first saw the end of the tunnel during a working retreat at the Fondation des Treilles, which offers an exceptional environment to think, write and eat in Provence. With WaveLab, David Donoho saved me from spending the second half of my life programming wavelet algorithms. This opportunity was beautifully implemented by Maureen Clerc and Jérôme Kalifa, who made all the figures and found many more mistakes than I dare say. Dear reader, you should thank Barbara Burke Hubbard, who corrected my Franglais (remaining errors are mine), and forced me to modify many notations and explanations. I thank her for doing it with tact and humor. My editor, Chuck Glaser, had the patience to wait but I appreciate even more his wisdom to let me think that I would finish in a year.

She will not read this book, yet my deepest gratitude goes to Branka with whom life has nothing to do with wavelets.

Stéphane Mallat

Preface to the second edition

Before leaving this *Wavelet Tour*, I naively thought that I should take advantage of remarks and suggestions made by readers. This almost got out of hand, and 200 pages ended up being rewritten. Let me outline the main components that were not in the first edition.

- *Bayes versus Minimax* Classical signal processing is almost entirely built in a Bayes framework, where signals are viewed as realizations of a random vector. For the last two decades, researchers have tried to model images with random vectors, but in vain. It is thus time to wonder whether this is really the best approach. Minimax theory opens an easier avenue for evaluating the performance of estimation and compression algorithms. It uses deterministic models that can be constructed even for complex signals such as images. Chapter 10 is rewritten and expanded to explain and compare the Bayes and minimax points of view.
- *Bounded Variation Signals* Wavelet transforms provide sparse representations of piecewise regular signals. The total variation norm gives an intuitive and precise mathematical framework in which to characterize the piecewise regularity of signals and images. In this second edition, the total variation is used to compute approximation errors, to evaluate the risk when removing noise from images, and to analyze the distortion rate of image transform codes.
- *Normalized Scale* Continuous mathematics give asymptotic results when the signal resolution N increases. In this framework, the signal support is

fixed, say $[0,1]$, and the sampling interval N^{-1} is progressively reduced. In contrast, digital signal processing algorithms are often presented by normalizing the sampling interval to 1, which means that the support $[0,N]$ increases with N. This new edition explains both points of views, but the figures now display signals with a support normalized to $[0,1]$, in accordance with the theorems.

- *Video Compression* Compressing video sequences is of prime importance for real time transmission with low-bandwidth channels such as the Internet or telephone lines. Motion compensation algorithms are presented at the end of Chapter 11.

Notation

$\langle f, g \rangle$	Inner product (A.6).
$\|f\|$	Norm (A.3).
$f[n] = O(g[n])$	Order of: there exists K such that $f[n] \leq Kg[n]$.
$f[n] = o(g[n])$	Small order of: $\lim_{n \to +\infty} \frac{f[n]}{g[n]} = 0$.
$f[n] \sim g[n]$	Equivalent to: $f[n] = O(g[n])$ and $g[n] = O(f[n])$.
$A < +\infty$	A is finite.
$A \gg B$	A is much bigger than B.
z^*	Complex conjugate of $z \in \mathbb{C}$.
$\lfloor x \rfloor$	Largest integer $n \leq x$.
$\lceil x \rceil$	Smallest integer $n \geq x$.
$n \bmod N$	Remainder of the integer division of n modulo N.

Sets

$\mathbb{N}$	Positive integers including 0.
$\mathbb{Z}$	Integers.
$\mathbb{R}$	Real numbers.
$\mathbb{R}^+$	Positive real numbers.
$\mathbb{C}$	Complex numbers.

Signals

$f(t)$	Continuous time signal.
$f[n]$	Discrete signal.

$\delta(t)$	Dirac distribution (A.30).
$\delta[n]$	Discrete Dirac (3.16).
$\mathbf{1}_{[a,b]}$	Indicator function which is 1 in $[a,b]$ and 0 outside.

Spaces

$\mathbf{C}_0$	Uniformly continuous functions (7.240).
$\mathbf{C}^p$	p times continuously differentiable functions.
$\mathbf{C}^\infty$	Infinitely differentiable functions.
$\mathbf{W}^s(\mathbb{R})$	Sobolev s times differentiable functions (9.5).
$\mathbf{L}^2(\mathbb{R})$	Finite energy functions $\int \lvert f(t)\rvert^2\, dt < +\infty$.
$\mathbf{L}^p(\mathbb{R})$	Functions such that $\int \lvert f(t)\rvert^p\, dt < +\infty$.
$\mathbf{l}^2(\mathbb{Z})$	Finite energy discrete signals $\sum_{n=-\infty}^{+\infty} \lvert f[n]\rvert^2 < +\infty$.
$\mathbf{l}^p(\mathbb{Z})$	Discrete signals such that $\sum_{n=-\infty}^{+\infty} \lvert f[n]\rvert^p < +\infty$.
$\mathbb{C}^N$	Complex signals of size N.
$\mathbf{U} \oplus \mathbf{V}$	Direct sum of two vector spaces.
$\mathbf{U} \otimes \mathbf{V}$	Tensor product of two vector spaces (A.19).

Operators

Id	Identity.
$f'(t)$	Derivative $\frac{df(t)}{dt}$.
$f^{(p)}(t)$	Derivative $\frac{d^p f(t)}{dt^p}$ of order p .
$\vec{\nabla} f(x,y)$	Gradient vector (6.55).
$f \star g(t)$	Continuous time convolution (2.2).
$f \star g[n]$	Discrete convolution (3.17).
$f \circledast g[n]$	Circular convolution (3.57)

Transforms

$\hat{f}(\omega)$	Fourier transform (2.6), (3.23).
$\hat{f}[k]$	Discrete Fourier transform (3.33).
$Sf(u,s)$	Short-time windowed Fourier transform (4.11).
$P_S f(u,\xi)$	Spectrogram (4.12).
$Wf(u,s)$	Wavelet transform (4.31).
$P_W f(u,\xi)$	Scalogram (4.55).
$P_V f(u,\xi)$	Wigner-Ville distribution (4.108).
$Af(u,\xi)$	Ambiguity function (4.24).

Probability

X	Random variable.
$\mathsf{E}\{X\}$	Expected value.
$\mathcal{H}(X)$	Entropy (11.4).

$\mathcal{H}_d(X)$	Differential entropy (11.20).
$\mathrm{Cov}(X_1, X_2)$	Covariance (A.22).
$F[n]$	Random vector.
$R_F[k]$	Autocovariance of a stationary process (A.26).

1

INTRODUCTION TO A TRANSIENT WORLD

After a few minutes in a restaurant we cease to notice the annoying hubbub of surrounding conversations, but a sudden silence reminds us of the presence of neighbors. Our attention is clearly attracted by transients and movements as opposed to stationary stimuli, which we soon ignore. Concentrating on transients is probably a strategy for selecting important information from the overwhelming amount of data recorded by our senses. Yet, classical signal processing has devoted most of its efforts to the design of time-invariant and space-invariant operators, that modify stationary signal properties. This has led to the indisputable hegemony of the Fourier transform, but leaves aside many information-processing applications.

The world of transients is considerably larger and more complex than the garden of stationary signals. The search for an ideal Fourier-like basis that would simplify most signal processing is therefore a hopeless quest. Instead, a multitude of different transforms and bases have proliferated, among which wavelets are just one example. This book gives a guided tour in this jungle of new mathematical and algorithmic results, while trying to provide an intuitive sense of orientation. Major ideas are outlined in this first chapter. Section 1.5.2 serves as a travel guide and introduces the *reproducible experiment* approach based on the WAVELAB and LASTWAVE softwares. It also discusses the use of *level numbers*—landmarks that can help the reader keep to the main roads.

1.1 FOURIER KINGDOM

The Fourier transform rules over linear time-invariant signal processing because sinusoidal waves $e^{i\omega t}$ are eigenvectors of linear time-invariant operators. A linear time-invariant operator L is entirely specified by the eigenvalues $\hat{h}(\omega)$:

$$\forall \omega \in \mathbb{R} \ , \ L e^{i\omega t} = \hat{h}(\omega) e^{i\omega t}. \tag{1.1}$$

To compute Lf, a signal f is decomposed as a sum of sinusoidal eigenvectors $\{e^{i\omega t}\}_{\omega \in \mathbb{R}}$:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega) e^{i\omega t} \, d\omega. \tag{1.2}$$

If f has finite energy, the theory of Fourier integrals presented in Chapter 2 proves that the amplitude $\hat{f}(\omega)$ of each sinusoidal wave $e^{i\omega t}$ is the Fourier transform of f:

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} \, dt. \tag{1.3}$$

Applying the operator L to f in (1.2) and inserting the eigenvector expression (1.1) gives

$$Lf(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega) \hat{h}(\omega) e^{i\omega t} \, d\omega. \tag{1.4}$$

The operator L amplifies or attenuates each sinusoidal component $e^{i\omega t}$ of f by $\hat{h}(\omega)$. It is a frequency *filtering* of f.

As long as we are satisfied with linear time-invariant operators, the Fourier transform provides simple answers to most questions. Its richness makes it suitable for a wide range of applications such as signal transmissions or stationary signal processing. However, if we are interested in transient phenomena—a word pronounced at a particular time, an apple located in the left corner of an image—the Fourier transform becomes a cumbersome tool.

The Fourier coefficient is obtained in (1.3) by correlating f with a sinusoidal wave $e^{i\omega t}$. Since the support of $e^{i\omega t}$ covers the whole real line, $\hat{f}(\omega)$ depends on the values $f(t)$ for all times $t \in \mathbb{R}$. This global "mix" of information makes it difficult to analyze any local property of f from $\hat{f}$. Chapter 4 introduces local time-frequency transforms, which decompose the signal over waveforms that are well localized in time and frequency.

1.2 TIME-FREQUENCY WEDDING

The uncertainty principle states that the energy spread of a function and its Fourier transform cannot be simultaneously arbitrarily small. Motivated by quantum mechanics, in 1946 the physicist Gabor [187] defined elementary time-frequency atoms as waveforms that have a minimal spread in a time-frequency plane. To measure time-frequency "information" content, he proposed decomposing signals over these elementary atomic waveforms. By showing that such decompositions

are closely related to our sensitivity to sounds, and that they exhibit important structures in speech and music recordings, Gabor demonstrated the importance of localized time-frequency signal processing.

Chapter 4 studies the properties of windowed Fourier and wavelet transforms, computed by decomposing the signal over different families of time-frequency atoms. Other transforms can also be defined by modifying the family of time-frequency atoms. A unified interpretation of local time-frequency decompositions follows the time-frequency energy density approach of Ville. In parallel to Gabor's contribution, in 1948 Ville [342], who was an electrical engineer, proposed analyzing the time-frequency properties of signals f with an energy density defined by

$$P_V f(t,\omega) = \int_{-\infty}^{+\infty} f\left(t+\frac{\tau}{2}\right) f^*\left(t-\frac{\tau}{2}\right) e^{-i\tau\omega}\, d\tau.$$

Once again, theoretical physics was ahead, since this distribution had already been introduced in 1932 by Wigner [351] in the context of quantum mechanics. Chapter 4 explains the path that relates Wigner-Ville distributions to windowed Fourier and wavelet transforms, or any linear time-frequency transform.

1.2.1 Windowed Fourier Transform

Gabor atoms are constructed by translating in time and frequency a time window g:

$$g_{u,\xi}(t) = g(t-u)\, e^{i\xi t}.$$

The energy of $g_{u,\xi}$ is concentrated in the neighborhood of u over an interval of size σ_t, measured by the standard deviation of $|g|^2$. Its Fourier transform is a translation by ξ of the Fourier transform $\hat{g}$ of g:

$$\hat{g}_{u,\xi}(\omega) = \hat{g}(\omega-\xi)\, e^{-iu(\omega-\xi)}. \tag{1.5}$$

The energy of $\hat{g}_{u,\xi}$ is therefore localized near the frequency ξ, over an interval of size σ_ω, which measures the domain where $\hat{g}(\omega)$ is non-negligible. In a time-frequency plane (t,ω), the energy spread of the atom $g_{u,\xi}$ is symbolically represented by the Heisenberg rectangle illustrated by Figure 1.1. This rectangle is centered at (u,ξ) and has a time width σ_t and a frequency width σ_ω. The uncertainty principle proves that its area satisfies

$$\sigma_t\, \sigma_\omega \geq \frac{1}{2}.$$

This area is minimum when g is a Gaussian, in which case the atoms $g_{u,\xi}$ are called *Gabor functions*.

The windowed Fourier transform defined by Gabor correlates a signal f with each atom $g_{u,\xi}$:

$$Sf(u,\xi) = \int_{-\infty}^{+\infty} f(t)\, g^*_{u,\xi}(t)\, dt = \int_{-\infty}^{+\infty} f(t)\, g(t-u)\, e^{-i\xi t}\, dt. \tag{1.6}$$

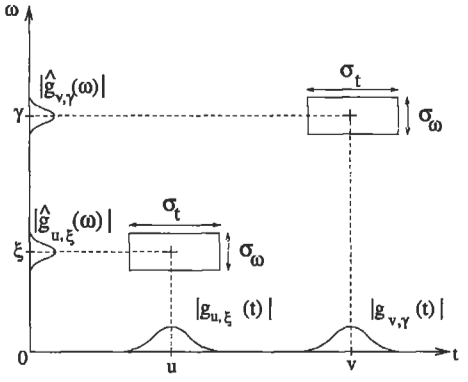


FIGURE 1.1 Time-frequency boxes ("Heisenberg rectangles") representing the energy spread of two Gabor atoms.

It is a Fourier integral that is localized in the neighborhood of u by the window $g(t-u)$. This time integral can also be written as a frequency integral by applying the Fourier Parseval formula (2.25):

$$Sf(u,\xi) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{f}(\omega)\,\hat{g}^*_{u,\xi}(\omega)\,d\omega. \tag{1.7}$$

The transform $Sf(u,\xi)$ thus depends only on the values $f(t)$ and $\hat{f}(\omega)$ in the time and frequency neighborhoods where the energies of $g_{u,\xi}$ and $\hat{g}_{u,\xi}$ are concentrated. Gabor interprets this as a "quantum of information" over the time-frequency rectangle illustrated in Figure 1.1.

When listening to music, we perceive sounds that have a frequency that varies in time. Measuring time-varying harmonics is an important application of windowed Fourier transforms in both music and speech recognition. A spectral line of f creates high amplitude windowed Fourier coefficients $Sf(u,\xi)$ at frequencies $\xi(u)$ that depend on the time u. The time evolution of such spectral components is therefore analyzed by following the location of large amplitude coefficients.

1.2.2 Wavelet Transform

In reflection seismology, Morlet knew that the modulated pulses sent underground have a duration that is too long at high frequencies to separate the returns of fine, closely-spaced layers. Instead of emitting pulses of equal duration, he thus thought of sending shorter waveforms at high frequencies. Such waveforms are simply obtained by scaling a single function called a *wavelet*. Although Grossmann was working in theoretical physics, he recognized in Morlet's approach some ideas that were close to his own work on coherent quantum states. Nearly forty years after Gabor, Morlet and Grossmann reactivated a fundamental collaboration between theoretical physics and signal processing, which led to the formalization of the

continuous wavelet transform [200]. Yet, these ideas were not totally new to mathematicians working in harmonic analysis, or to computer vision researchers studying multiscale image processing. It was thus only the beginning of a rapid catalysis that brought together scientists with very different backgrounds, first around coffee tables, then in more luxurious conferences.

A wavelet ψ is a function of zero average:

$$\int_{-\infty}^{+\infty} \psi(t)\,dt = 0,$$

which is dilated with a scale parameter s, and translated by u:

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\,\psi\left(\frac{t-u}{s}\right). \tag{1.8}$$

The wavelet transform of f at the scale s and position u is computed by correlating f with a wavelet atom

$$Wf(u,s) = \int_{-\infty}^{+\infty} f(t)\,\frac{1}{\sqrt{s}}\,\psi^*\left(\frac{t-u}{s}\right)dt. \tag{1.9}$$

Time-Frequency Measurements Like a windowed Fourier transform, a wavelet transform can measure the time-frequency variations of spectral components, but it has a different time-frequency resolution. A wavelet transform correlates f with $\psi_{u,s}$. By applying the Fourier Parseval formula (2.25), it can also be written as a frequency integration:

$$Wf(u,s) = \int_{-\infty}^{+\infty} f(t)\,\psi^*_{u,s}(t)\,dt = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{f}(\omega)\,\hat{\psi}^*_{u,s}(\omega)\,d\omega. \tag{1.10}$$

The wavelet coefficient $Wf(u,s)$ thus depends on the values $f(t)$ and $\hat{f}(\omega)$ in the time-frequency region where the energy of $\psi_{u,s}$ and $\hat{\psi}_{u,s}$ is concentrated. Time varying harmonics are detected from the position and scale of high amplitude wavelet coefficients.

In time, $\psi_{u,s}$ is centered at u with a spread proportional to s. Its Fourier transform is calculated from (1.8):

$$\hat{\psi}_{u,s}(\omega) = \mathrm{e}^{-iu\omega}\,\sqrt{s}\,\hat{\psi}(s\omega),$$

where $\hat{\psi}$ is the Fourier transform of ψ. To analyze the phase information of signals, a complex analytic wavelet is used. This means that $\hat{\psi}(\omega) = 0$ for $\omega < 0$. Its energy is concentrated in a positive frequency interval centered at η. The energy of $\hat{\psi}_{u,s}(\omega)$ is therefore concentrated over a positive frequency interval centered at η/s, whose size is scaled by $1/s$. In the time-frequency plane, a wavelet atom $\psi_{u,s}$ is symbolically represented by a rectangle centered at $(u, \eta/s)$. The time and frequency spread are respectively proportional to s and $1/s$. When s varies, the height and width of the rectangle change but its area remains constant, as illustrated by Figure 1.2.

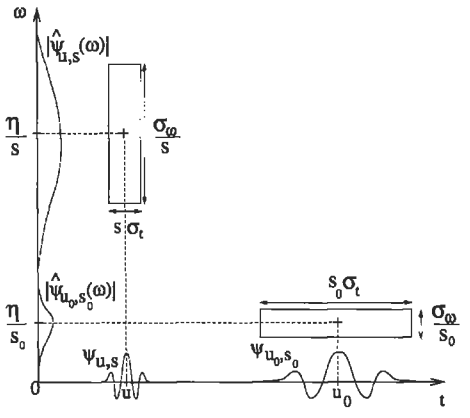


FIGURE I.2 Time-frequency boxes of two wavelets $\psi_{u,s}$ and ψ_{u_0,s_0}. When the scale s decreases, the time support is reduced but the frequency spread increases and covers an interval that is shifted towards high frequencies.

Multiscale Zooming The wavelet transform can also detect and characterize transients with a zooming procedure across scales. Suppose that ψ is real. Since it has a zero average, a wavelet coefficient $Wf(u,s)$ measures the variation of f in a neighborhood of u whose size is proportional to s. Sharp signal transitions create large amplitude wavelet coefficients. Chapter 6 relates the pointwise regularity of f to the asymptotic decay of the wavelet transform $Wf(u,s)$, when s goes to zero. Singularities are detected by following across scales the local maxima of the wavelet transform. In images, high amplitude wavelet coefficients indicate the position of edges, which are sharp variations of the image intensity. Different scales provide the contours of image structures of varying sizes. Such multiscale edge detection is particularly effective for pattern recognition in computer vision [113].

The zooming capability of the wavelet transform not only locates isolated singular events, but can also characterize more complex multifractal signals having non-isolated singularities. Mandelbrot [43] was the first to recognize the existence of multifractals in most corners of nature. Scaling one part of a multifractal produces a signal that is statistically similar to the whole. This self-similarity appears in the wavelet transform, which modifies the analyzing scale. From the global wavelet transform decay, one can measure the singularity distribution of multifractals. This is particularly important in analyzing their properties and testing models that explain the formation of multifractals in physics.

I.3 BASES OF TIME-FREQUENCY ATOMS

The continuous windowed Fourier transform $Sf(u,\xi)$ and the wavelet transform $Wf(u,s)$ are two-dimensional representations of a one-dimensional signal f. This

indicates the existence of some redundancy that can be reduced and even removed by subsampling the parameters of these transforms.

Frames Windowed Fourier transforms and wavelet transforms can be written as inner products in $\mathbf{L}^2(\mathbb{R})$, with their respective time-frequency atoms

$$Sf(u,\xi) = \int_{-\infty}^{+\infty} f(t)\, g^*_{u,\xi}(t)\, dt = \langle f, g_{u,\xi} \rangle$$

and

$$Wf(u,s) = \int_{-\infty}^{+\infty} f(t)\, \psi^*_{u,s}(t)\, dt = \langle f, \psi_{u,s} \rangle.$$

Subsampling both transforms defines a complete signal representation if any signal can be reconstructed from linear combinations of discrete families of windowed Fourier atoms $\{g_{u_n,\xi_k}\}_{(n,k)\in\mathbb{Z}^2}$ and wavelet atoms $\{\psi_{u_n,s_j}\}_{(j,n)\in\mathbb{Z}^2}$. The frame theory of Chapter 5 discusses what conditions these families of waveforms must meet if they are to provide stable and complete representations.

Completely eliminating the redundancy is equivalent to building a basis of the signal space. Although wavelet bases were the first to arrive on the research market, they have quickly been followed by other families of orthogonal bases, such as wavelet packet and local cosine bases.

1.3.1 Wavelet Bases and Filter Banks

In 1910, Haar [202] realized that one can construct a simple piecewise constant function

$$\psi(t) = \begin{cases} 1 & \text{if } 0 \le t < 1/2 \\ -1 & \text{if } 1/2 \le t < 1 \\ 0 & \text{otherwise} \end{cases}$$

whose dilations and translations generate an orthonormal basis of $\mathbf{L}^2(\mathbb{R})$:

$$\left\{ \psi_{j,n}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t - 2^j n}{2^j} \right) \right\}_{(j,n)\in\mathbb{Z}^2}.$$

Any finite energy signal f can be decomposed over this wavelet orthogonal basis $\{\psi_{j,n}\}_{(j,n)\in\mathbb{Z}^2}$

$$f = \sum_{j=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \langle f, \psi_{j,n} \rangle\, \psi_{j,n}. \tag{1.11}$$

Since $\psi(t)$ has a zero average, each partial sum

$$d_j(t) = \sum_{n=-\infty}^{+\infty} \langle f, \psi_{j,n} \rangle\, \psi_{j,n}(t)$$

can be interpreted as detail variations at the scale 2^j. These layers of details are added at all scales to progressively improve the approximation of f, and ultimately recover f.

If f has smooth variations, we should obtain a precise approximation when removing fine scale details, which is done by truncating the sum (1.11). The resulting approximation at a scale 2^J is

$$f_J(t) = \sum_{j=J}^{+\infty} d_j(t).$$

For a Haar basis, f_J is piecewise constant. Piecewise constant approximations of smooth functions are far from optimal. For example, a piecewise linear approximation produces a smaller approximation error. The story continues in 1980, when Strömberg [322] found a piecewise linear function ψ that also generates an orthonormal basis and gives better approximations of smooth functions. Meyer was not aware of this result, and motivated by the work of Morlet and Grossmann he tried to prove that there exists no regular wavelet ψ that generates an orthonormal basis. This attempt was a failure since he ended up constructing a whole family of orthonormal wavelet bases, with functions ψ that are infinitely continuously differentiable [270]. This was the fundamental impulse that lead to a widespread search for new orthonormal wavelet bases, which culminated in the celebrated Daubechies wavelets of compact support [144].

The systematic theory for constructing orthonormal wavelet bases was established by Meyer and Mallat through the elaboration of multiresolution signal approximations [254], presented in Chapter 7. It was inspired by original ideas developed in computer vision by Burt and Adelson [108] to analyze images at several resolutions. Digging more into the properties of orthogonal wavelets and multiresolution approximations brought to light a surprising relation with filter banks constructed with conjugate mirror filters.

Filter Banks Motivated by speech compression, in 1976 Croisier, Esteban and Galand [141] introduced an invertible filter bank, which decomposes a discrete signal $f[n]$ in two signals of half its size, using a filtering and subsampling procedure. They showed that $f[n]$ can be recovered from these subsampled signals by canceling the aliasing terms with a particular class of filters called *conjugate mirror filters*. This breakthrough led to a 10-year research effort to build a complete filter bank theory. Necessary and sufficient conditions for decomposing a signal in subsampled components with a filtering scheme, and recovering the same signal with an inverse transform, were established by Smith and Barnwell [316], Vaidyanathan [336] and Vetterli [339].

The multiresolution theory of orthogonal wavelets proves that any conjugate mirror filter characterizes a wavelet ψ that generates an orthonormal basis of $\mathbf{L}^2(\mathbb{R})$. Moreover, a fast discrete wavelet transform is implemented by cascading these conjugate mirror filters. The equivalence between this continuous time wavelet

theory and discrete filter banks led to a new fruitful interface between digital signal processing and harmonic analysis, but also created a culture shock that is not totally resolved.

Continuous Versus Discrete and Finite Many signal processors have been and still are wondering what is the point of these continuous time wavelets, since all computations are performed over discrete signals, with conjugate mirror filters. Why bother with the convergence of infinite convolution cascades if in practice we only compute a finite number of convolutions? Answering these important questions is necessary in order to understand why throughout this book we alternate between theorems on continuous time functions and discrete algorithms applied to finite sequences.

A short answer would be "simplicity". In $\mathbf{L}^2(\mathbb{R})$, a wavelet basis is constructed by dilating and translating a single function ψ. Several important theorems relate the amplitude of wavelet coefficients to the local regularity of the signal f. Dilations are not defined over discrete sequences, and discrete wavelet bases have therefore a more complicated structure. The regularity of a discrete sequence is not well defined either, which makes it more difficult to interpret the amplitude of wavelet coefficients. A theory of continuous time functions gives asymptotic results for discrete sequences with sampling intervals decreasing to zero. This theory is useful because these asymptotic results are precise enough to understand the behavior of discrete algorithms.

Continuous time models are not sufficient for elaborating discrete signal processing algorithms. Uniformly sampling the continuous time wavelets $\{\psi_{j,n}(t)\}_{(j,n)\in\mathbb{Z}^2}$ does not produce a discrete orthonormal basis. The transition between continuous and discrete signals must be done with great care. Restricting the constructions to finite discrete signals adds another layer of complexity because of border problems. How these border issues affect numerical implementations is carefully addressed once the properties of the bases are well understood. To simplify the mathematical analysis, throughout the book continuous time transforms are introduced first. Their discretization is explained afterwards, with fast numerical algorithms over finite signals.

1.3.2 Tilings of Wavelet Packet and Local Cosine Bases

Orthonormal wavelet bases are just an appetizer. Their construction showed that it is not only possible but relatively simple to build orthonormal bases of $\mathbf{L}^2(\mathbb{R})$ composed of local time-frequency atoms. The completeness and orthogonality of a wavelet basis is represented by a tiling that covers the time-frequency plane with the wavelets' time-frequency boxes. Figure 1.3 shows the time-frequency box of each $\psi_{j,n}$, which is translated by $2^j n$, with a time and a frequency width scaled respectively by 2^j and 2^{-j}.

One can draw many other tilings of the time-frequency plane, with boxes of minimal surface as imposed by the uncertainty principle. Chapter 8 presents

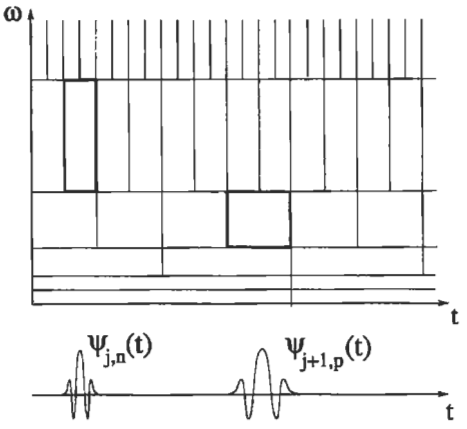


FIGURE I.3 The time-frequency boxes of a wavelet basis define a tiling of the time-frequency plane.

several constructions that associate large families of orthonormal bases of $\mathbf{L}^2(\mathbb{R})$ to such new tilings.

Wavelet Packet Bases A wavelet orthonormal basis decomposes the frequency axis in dyadic intervals whose sizes have an exponential growth, as shown by Figure 1.3. Coifman, Meyer and Wickerhauser [139] have generalized this fixed dyadic construction by decomposing the frequency in intervals whose bandwidths may vary. Each frequency interval is covered by the time-frequency boxes of wavelet packet functions that are uniformly translated in time in order to cover the whole plane, as shown by Figure 1.4.

Wavelet packet functions are designed by generalizing the filter bank tree that relates wavelets and conjugate mirror filters. The frequency axis division of wavelet packets is implemented with an appropriate sequence of iterated convolutions with conjugate mirror filters. Fast numerical wavelet packet decompositions are thus implemented with discrete filter banks.

Local Cosine Bases Orthonormal bases of $\mathbf{L}^2(\mathbb{R})$ can also be constructed by dividing the time axis instead of the frequency axis. The time axis is segmented in successive finite intervals $[a_p, a_{p+1}]$. The local cosine bases of Malvar [262] are obtained by designing smooth windows $g_p(t)$ that cover each interval $[a_p, a_{p+1}]$, and multiplying them by cosine functions $\cos(\xi t + \phi)$ of different frequencies. This is yet another idea that was independently studied in physics, signal processing and mathematics. Malvar's original construction was done for discrete signals. At the same time, the physicist Wilson [353] was designing a local cosine basis with smooth windows of infinite support, to analyze the properties of quantum coherent states. Malvar bases were also rediscovered and generalized by the

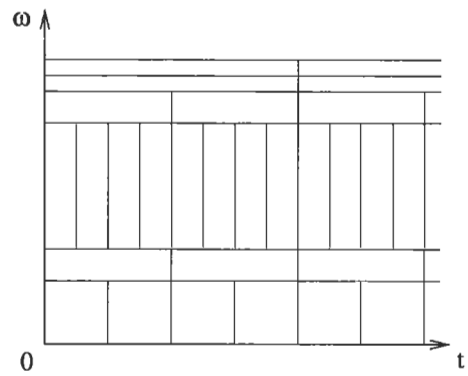


FIGURE 1.4 A wavelet packet basis divides the frequency axis in separate intervals of varying sizes. A tiling is obtained by translating in time the wavelet packets covering each frequency interval.

harmonic analysts Coifman and Meyer [138]. These different views of the same bases brought to light mathematical and algorithmic properties that opened new applications.

A multiplication by $\cos(\xi t+\phi)$ translates the Fourier transform $\hat{g}_p(\omega)$ of $g_p(t)$ by $\pm\xi$. Over positive frequencies, the time-frequency box of the modulated window $g_p(t)\cos(\xi t+\phi)$ is therefore equal to the time-frequency box of g_p translated by ξ along frequencies. The time-frequency boxes of local cosine basis vectors define a tiling of the time-frequency plane illustrated by Figure 1.5.

1.4 BASES FOR WHAT?

The tiling game is clearly unlimited. Local cosine and wavelet packet bases are important examples, but many other kinds of bases can be constructed. It is thus time to wonder how to select an appropriate basis for processing a particular class of signals. The decomposition coefficients of a signal in a basis define a representation that highlights some particular signal properties. For example, wavelet coefficients provide explicit information on the location and type of signal singularities. The problem is to find a criterion for selecting a basis that is intrinsically well adapted to represent a class of signals.

Mathematical approximation theory suggests choosing a basis that can construct precise signal approximations with a linear combination of a small number of vectors selected inside the basis. These selected vectors can be interpreted as intrinsic signal structures. Compact coding and signal estimation in noise are applications where this criterion is a good measure of the efficiency of a basis. Linear and non-linear procedures are studied and compared. This will be the occasion to show that non-linear does not always mean complicated.

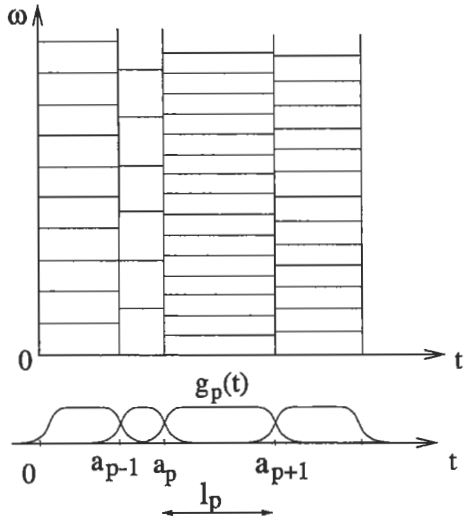


FIGURE 1.5 A local cosine basis divides the time axis with smooth windows $g_p(t)$. Multiplications with cosine functions translate these windows in frequency and yield a complete cover of the time-frequency plane.

1.4.1 Approximation

The development of orthonormal wavelet bases has opened a new bridge between approximation theory and signal processing. This exchange is not quite new since the fundamental sampling theorem comes from an interpolation theory result proved in 1935 by Whittaker [349]. However, the state of the art of approximation theory has changed since 1935. In particular, the properties of non-linear approximation schemes are much better understood, and give a firm foundation for analyzing the performance of many non-linear signal processing algorithms. Chapter 9 introduces important approximation theory results that are used in signal estimation and data compression.

Linear Approximation A linear approximation projects the signal f over M vectors that are chosen *a priori* in an orthonormal basis $\mathcal{B} = \{g_m\}_{m \in \mathbb{N}}$, say the first M:

$$f_M = \sum_{m=0}^{M-1} \langle f, g_m \rangle \, g_m. \tag{1.12}$$

Since the basis is orthonormal, the approximation error is the sum of the remaining squared inner products

$$\epsilon[M] = \|f - f_M\|^2 = \sum_{m=M}^{+\infty} |\langle f, g_m \rangle|^2.$$

The accuracy of this approximation clearly depends on the properties of f relative to the basis $\mathcal{B}$.

A Fourier basis yields efficient linear approximations of uniformly smooth signals, which are projected over the M lower frequency sinusoidal waves. When M increases, the decay of the error $\epsilon[M]$ can be related to the global regularity of f. Chapter 9 characterizes spaces of smooth functions from the asymptotic decay of $\epsilon[M]$ in a Fourier basis.

In a wavelet basis, the signal is projected over the M larger scale wavelets, which is equivalent to approximating the signal at a fixed resolution. Linear approximations of uniformly smooth signals in wavelet and Fourier bases have similar properties and characterize nearly the same function spaces.

Suppose that we want to approximate a class of discrete signals of size N, modeled by a random vector $F[n]$. The average approximation error when projecting F over the first M basis vectors of an orthonormal basis $\mathcal{B} = \{g_m\}_{0 \leq m < N}$ is

$$\epsilon[M] = \mathrm{E}\{\|F - F_M\|^2\} = \sum_{m=M}^{N-1} \mathrm{E}\{|\langle F, g_m \rangle|^2\}.$$

Chapter 9 proves that the basis that minimizes this error is the Karhunen-Loève basis, which diagonalizes the covariance matrix of F. This remarkable property explains the fundamental importance of the Karhunen-Loève basis in optimal linear signal processing schemes. This is however only a beginning.

Non-linear Approximation The linear approximation (1.12) is improved if we choose *a posteriori* the M vectors g_m, depending on f. The approximation of f with M vectors whose indexes are in I_M is

$$f_M = \sum_{m \in I_M} \langle f, g_m \rangle \, g_m. \tag{1.13}$$

The approximation error is the sum of the squared inner products with vectors not in I_M:

$$\epsilon[M] = \|f - f_M\|^2 = \sum_{n \notin I_M} |\langle f, g_m \rangle|^2.$$

To minimize this error, we choose I_M to be the set of M vectors that have the largest inner product amplitude $|\langle f, g_m \rangle|$. This approximation scheme is non-linear because the approximation vectors change with f.

The amplitude of inner products in a wavelet basis is related to the local regularity of the signal. A non-linear approximation that keeps the largest wavelet inner products is equivalent to constructing an adaptive approximation grid, whose resolution is locally increased where the signal is irregular. If the signal has isolated singularities, this non-linear approximation is much more precise than a linear scheme that maintains the same resolution over the whole signal support. The spaces of functions that are well approximated by non-linear wavelet schemes are

thus much larger than for linear schemes, and include functions with isolated singularities. Bounded variation signals are important examples that provide useful models for images.

In this non-linear setting, Karhunen-Loève bases are not optimal for approximating the realizations of a process F. It is often easy to find a basis that produces a smaller non-linear error than a Karhunen-Loève basis, but there is yet no procedure for computing the optimal basis that minimizes the average non-linear error.

Adaptive Basis Choice Approximations of non-linear signals can be improved by choosing the approximation vectors in families that are much larger than a basis. Music recordings, which include harmonic and transient structures of very different types, are examples of complex signals that are not well approximated by a few vectors chosen from a single basis.

A new degree of freedom is introduced if instead of choosing *a priori* the basis $\mathcal{B}$, we adaptively select a "best" basis, depending on the signal f. This best basis minimizes a cost function related to the non-linear approximation error of f. A fast dynamical programming algorithm can find the best basis in families of wavelet packet basis or local cosine bases [140]. The selected basis corresponds to a time-frequency tiling that "best" concentrates the signal energy over a few time-frequency atoms.

Orthogonality is often not crucial in the post-processing of signal coefficients. One may thus further enlarge the freedom of choice by approximating the signal f with M non-orthogonal vectors $\{g_{\gamma_m}\}_{0 \le m < M}$, chosen from a large and redundant dictionary $\mathcal{D} = \{g_\gamma\}_{\gamma \in \Gamma}$:

$$f_M = \sum_{m=0}^{M-1} a_m g_{\gamma_m}.$$

Globally optimizing the choice of these M vectors in $\mathcal{D}$ can lead to a combinatorial explosion. Chapter 9 introduces sub-optimal pursuit algorithms that reduce the numerical complexity, while constructing efficient approximations [119, 259].

1.4.2 Estimation

The estimation of a signal embedded in noise requires taking advantage of any prior information about the signal and the noise. Chapter 10 studies and contrasts several approaches: Bayes versus minimax, linear versus non-linear. Until recently, signal processing estimation was mostly Bayesian and linear. Non-linear smoothing algorithms existed in statistics, but these procedures were often ad-hoc and complex. Two statisticians, Donoho and Johnstone [167], changed the game by proving that a simple thresholding algorithm in an appropriate basis can be a nearly optimal non-linear estimator.

Linear versus Non-Linear A signal $f[n]$ of size N is contaminated by the addition of a noise. This noise is modeled as the realization of a random process $W[n]$, whose

probability distribution is known. The measured data are

$$X[n] = f[n] + W[n] .$$

The signal f is estimated by transforming the noisy data X with an operator D:

$$\tilde{F} = DX .$$

The risk of the estimator $\tilde{F}$ of f is the average error, calculated with respect to the probability distribution of the noise W:

$$r(D,f) = \mathsf{E}\{\|f - DX\|^2\} .$$

It is tempting to restrict oneself to linear operators D, because of their simplicity. Yet, non-linear operators may yield a much lower risk. To keep the simplicity, we concentrate on diagonal operators in a basis $\mathcal{B}$. If the basis $\mathcal{B}$ gives a sparse signal representation, Donoho and Johnstone [167] prove that a nearly optimal non-linear estimator is obtained with a simple thresholding:

$$\tilde{F} = DX = \sum_{m=0}^{N-1} \rho_T(\langle X, g_m \rangle)\, g_m .$$

The thresholding function $\rho_T(x)$ sets to zero all coefficients below T:

$$\rho_T(x) = \begin{cases} 0 & \text{if } |x| < T \\ x & \text{if } |x| \geq T \end{cases} .$$

In a wavelet basis, such a thresholding implements an adaptive smoothing, which averages the data X with a kernel that depends on the regularity of the underlying signal f.

Bayes Versus Minimax To optimize the estimation operator D, one must take advantage of any prior information available about the signal f. In a Bayes framework, f is considered as a realization of a random vector F, whose probability distribution π is known a priori. Thomas Bayes was a XVII century philosopher, who first suggested and investigated methods sometimes referred as "inverse probability methods," which are basic to the study of Bayes estimators. The Bayes risk is the expected risk calculated with respect to the prior probability distribution π of the signal:

$$r(D,\pi) = \mathsf{E}_\pi\{r(D,F)\} .$$

Optimizing D among all possible operators yields the *minimum Bayes risk*:

$$r_n(\pi) = \inf_{all\, D} r(D,\pi) .$$

Complex signals such as images are clearly non-Gaussian, and there is yet no reliable probabilistic model that incorporates the diversity of structures such as edges and textures.

In the 1940's, Wald brought a new perspective on statistics, through a decision theory partly imported from the theory of games. This point of view offers a simpler way to incorporate prior information on complex signals. Signals are modeled as elements of a particular set Θ, without specifying their probability distribution in this set. For example, large classes of images belong to the set of signals whose total variation is bounded by a constant. To control the risk for any $f \in \Theta$, we compute the maximum risk

$$r(D,\Theta) = \sup_{f\in\Theta} r(D,f) .$$

The *minimax risk* is the lower bound computed over all operators D:

$$r_n(\Theta) = \inf_{all\, D} r(D,\Theta).$$

In practice, the goal is to find an operator D that is simple to implement and which yields a risk close the minimax lower bound.

Unless Θ has particular convexity properties, non-linear estimators have a much lower risk than linear estimators. If W is a white noise and signals in Θ have a sparse representation in $\mathcal{B}$, then Chapter 10 shows that thresholding estimators are nearly minimax optimal. In particular, the risk of wavelet thresholding estimators is close to the minimax risk for wide classes of piecewise smooth signals, including bounded variation images. Thresholding estimators are extended to more complex problems such as signal restorations and deconvolutions. The performance of a thresholding may also be improved with a best basis search or a pursuit algorithm that adapts the basis $\mathcal{B}$ to the noisy data. However, more adaptivity does not necessarily means less risk.

1.4.3 Compression

Limited storage space and transmission through narrow band-width channels create a need for compressing signals while minimizing their degradation. Transform codes compress signals by decomposing them in an orthonormal basis. Chapter 11 introduces the basic information theory needed to understand these codes and optimize their performance. Bayes and minimax approaches are studied.

A transform code decomposes a signal f in an orthonormal basis $\mathcal{B} = \{g_m\}_{0\le m<N}$:

$$f = \sum_{m=0}^{N-1} \langle f, g_m\rangle\, g_m .$$

The coefficients $\langle f, g_m\rangle$ are approximated by quantized values $Q(\langle f, g_m\rangle)$. A signal $\tilde{f}$ is restored from these quantized coefficients:

$$\tilde{f} = \sum_{m=0}^{N-1} Q(\langle f, g_m\rangle)\, g_m .$$

A binary code is used to record the quantized coefficients $Q(\langle f, g_m \rangle)$ with R bits. The resulting distortion is

$$d(R, f) = \|f - \tilde{f}\|^2 .$$

At the compression rates currently used for images, $d(R, f)$ has a highly non-linear behavior, which depends on the precision of non-linear approximations of f from a few vectors in the basis $\mathcal{B}$.

To compute the distortion rate over a whole signal class, the Bayes framework models signals as realizations of a random vector F whose probability distribution π is known. The goal is then to optimize the quantization and the basis $\mathcal{B}$ in order to minimize the average distortion rate $d(R, \pi) = \mathsf{E}_\pi\{d(R, F)\}$. This approach applies particularly well to audio signals, which are relatively well modeled by Gaussian processes.

In the absence of stochastic models for complex signals such as images, the minimax approach computes the maximum distortion by assuming only that the signal belongs to a prior set Θ. Chapter 11 describes the implementation of image transform codes in wavelet bases and block cosine bases. The minimax distortion rate is calculated for bounded variation images, and wavelet transform codes are proved to be nearly minimax optimal.

For video compression, one must also take advantage of the similarity of images across time. The most effective algorithms predict each image from a previous one by compensating for the motion, and the error is recorded with a transform code. MPEG video compression standards are described.

1.5 TRAVEL GUIDE

1.5.1 Reproducible Computational Science

The book covers the whole spectrum from theorems on functions of continuous variables to fast discrete algorithms and their applications. Section 1.3.1 argues that models based on continuous time functions give useful asymptotic results for understanding the behavior of discrete algorithms. Yet, a mathematical analysis alone is often unable to predict fully the behavior and suitability of algorithms for specific signals. Experiments are necessary and such experiments ought in principle be reproducible, just like experiments in other fields of sciences.

In recent years, the idea of reproducible algorithmic results has been championed by Claerbout [127] in exploration geophysics. The goal of exploration seismology is to produce the highest possible quality image of the subsurface. Part of the scientific know-how involved includes appropriate parameter settings that lead to good results on real datasets. The reproducibility of experiments thus requires having the complete software and full source code for inspection, modification and application under varied parameter settings.

Donoho has advocated the reproducibility of algorithms in wavelet signal processing, through the development of a WaveLab toolbox, which is a large library of Matlab routines. He summarizes Claerbout's insight in a slogan: [105]

An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software environment and the complete set of instructions which generated the figures.

Following this perspective, all wavelet and time-frequency tools presented in this book are available in WAVELAB. The figures can be reproduced as demos and the source code is available. The LASTWAVE package offers a similar library of wavelet related algorithms that are programmed in C, with a user-friendly shell interface and graphics. Appendix B explains how to retrieve these toolboxes, and relates their subroutines to the algorithms described in the book.

1.5.2 Road Map

Sections are kept as independent as possible, and some redundancy is introduced to avoid imposing a linear progression through the book. The preface describes several possible paths for a graduate signal processing or an applied mathematics course. A partial hierarchy between sections is provided by a level number. If a section has a level number then all sub-sections without number inherit this level, but a higher level number indicates that a subsection is more advanced.

Sections of level ^{1} introduce central ideas and techniques for wavelet and time-frequency signal processing. These would typically be taught in an introductory course. The first sections of Chapter 7 on wavelet orthonormal bases are examples. Sections of level ^{2} concern results that are important but which are either more advanced or dedicated to an application. Wavelet packets and local cosine bases in Chapter 8 are of that sort. Applications to estimation and data compression belong to this level, including fundamental results such as Wiener filtering. Sections of level ^{3} describe advanced results that are at the frontier of research or mathematically more difficult. These sections open the book to research problems.

All theorems are explained in the text and reading the proofs is not necessary to understand the results. Proofs also have a level index specifying their difficulty, as well as their conceptual or technical importance. These levels have been set by trying to answer the question: "Should this proof be taught in an introductory course?" Level ^{1} means probably, level ^{2} probably not, level ^{3} certainly not. Problems at the end of each chapter follow this hierarchy of levels. Direct applications of the course are at the level ^{1}. Problems at level ^{2} require more thinking. Problems of level ^{3} are often at the interface of research and can provide topics for deeper projects.

The book begins with Chapters 2 and 3, which review the Fourier transform properties and elementary discrete signal processing. They provide the necessary background for readers with no signal processing experience. Fundamental properties of local time-frequency transforms are presented in Chapter 4. The wavelet and windowed Fourier transforms are introduced and compared. The measurement of instantaneous frequencies is used to illustrate the limitations of their time-frequency resolution. Wigner-Ville time-frequency distributions give a

global perspective which relates all quadratic time-frequency distributions. Frame theory is explained in Chapter 5. It offers a flexible framework for analyzing the properties of redundant or non-linear adaptive decompositions. Chapter 6 explains the relations between the decay of the wavelet transform amplitude across scales and local signal properties. It studies applications involving the detection of singularities and analysis of multifractals.

The construction of wavelet bases and their relations with filter banks are fundamental results presented in Chapter 7. An overdose of orthonormal bases can strike the reader while studying the construction and properties of wavelet packets and local cosine bases in Chapter 8. It is thus important to read in parallel Chapter 9, which studies the approximation performance of orthogonal bases. The estimation and data compression applications of Chapters 10 and 11 give life to most theoretical and algorithmic results of the book. These chapters offer a practical perspective on the relevance of these linear and non-linear signal processing algorithms.

II

FOURIER KINGDOM

The story begins in 1807 when Fourier presents a memoir to the Institut de France, where he claims that any periodic function can be represented as a series of harmonically related sinusoids. This idea had a profound impact in mathematical analysis, physics and engineering, but it took one and a half centuries to understand the convergence of Fourier series and complete the theory of Fourier integrals.

Fourier was motivated by the study of heat diffusion, which is governed by a linear differential equation. However, the Fourier transform diagonalizes all linear time-invariant operators, which are the building blocks of signal processing. It is therefore not only the starting point of our exploration but the basis of all further developments.

2.1 LINEAR TIME-INVARIANT FILTERING [1]

Classical signal processing operations such as signal transmission, stationary noise removal or predictive coding are implemented with linear time-invariant operators. The time invariance of an operator L means that if the input $f(t)$ is delayed by τ, $f_\tau(t) = f(t-\tau)$, then the output is also delayed by τ:

$$g(t) = Lf(t) \;\Rightarrow\; g(t-\tau) = Lf_\tau(t). \tag{2.1}$$

For numerical stability, the operator L must have a weak form of continuity, which means that Lf is modified by a small amount if f is slightly modified. This weak

continuity is formalized by the theory of distributions [66, 69], which guarantees that we are on a safe ground without further worrying about it.

2.1.1 Impulse Response

Linear time-invariant systems are characterized by their response to a Dirac impulse, defined in Appendix A.7. If f is continuous, its value at t is obtained by an "integration" against a Dirac located at t. Let $\delta_u(t) = \delta(t-u)$:

$$f(t) = \int_{-\infty}^{+\infty} f(u)\,\delta_u(t)\,du.$$

The continuity and linearity of L imply that

$$Lf(t) = \int_{-\infty}^{+\infty} f(u)\,L\delta_u(t)\,du.$$

Let h be the impulse response of L:

$$h(t) = L\delta(t).$$

The time-invariance proves that $L\delta_u(t) = h(t-u)$ and hence

$$Lf(t) = \int_{-\infty}^{+\infty} f(u)\,h(t-u)\,du = \int_{-\infty}^{+\infty} h(u)\,f(t-u)\,du = h \star f(t). \tag{2.2}$$

A time-invariant linear filter is thus equivalent to a convolution with the impulse response h. The continuity of f is not necessary. This formula remains valid for any signal f for which the convolution integral converges.

Let us recall a few useful properties of convolution products:

- Commutativity
$$f \star h(t) = h \star f(t). \tag{2.3}$$
- Differentiation
$$\frac{d}{dt}(f \star h)(t) = \frac{df}{dt} \star h(t) = f \star \frac{dh}{dt}(t). \tag{2.4}$$
- Dirac convolution
$$f \star \delta_\tau(t) = f(t-\tau). \tag{2.5}$$

Stability and Causality A filter is said to be *causal* if $Lf(t)$ does not depend on the values $f(u)$ for $u > t$. Since

$$Lf(t) = \int_{-\infty}^{+\infty} h(u)\,f(t-u)\,du,$$

this means that $h(u) = 0$ for $u < 0$. Such impulse responses are said to be *causal*.

The *stability* property guarantees that $Lf(t)$ is bounded if $f(t)$ is bounded. Since

$$|Lf(t)| \leq \int_{-\infty}^{+\infty} |h(u)|\,|f(t-u)|\,du \leq \sup_{u\in\mathbb{R}} |f(u)| \int_{-\infty}^{+\infty} |h(u)|\,du,$$

it is sufficient that $\int_{-\infty}^{+\infty} |h(u)|\,du < +\infty$. One can verify that this condition is also necessary if h is a function. We thus say that h is *stable* if it is integrable.

Example 2.1 An *amplification* and *delay* system is defined by

$$Lf(t) = \lambda f(t-\tau).$$

The impulse response of this filter is $h(t) = \lambda\delta(t-\tau)$.

Example 2.2 A *uniform averaging* of f over intervals of size T is calculated by

$$Lf(t) = \frac{1}{T}\int_{t-T/2}^{t+T/2} f(u)\,du.$$

This integral can be rewritten as a convolution of f with the impulse response $h = 1/T\,\mathbf{1}_{[-T/2,T/2]}$.

2.1.2 Transfer Functions

Complex exponentials $e^{i\omega t}$ are eigenvectors of convolution operators. Indeed

$$Le^{i\omega t} = \int_{-\infty}^{+\infty} h(u)\,e^{i\omega(t-u)}\,du,$$

which yields

$$Le^{i\omega t} = e^{it\omega}\int_{-\infty}^{+\infty} h(u)\,e^{-i\omega u}\,du = \hat{h}(\omega)\,e^{i\omega t}.$$

The eigenvalue

$$\hat{h}(\omega) = \int_{-\infty}^{+\infty} h(u)\,e^{-i\omega u}\,du$$

is the Fourier transform of h at the frequency ω. Since complex sinusoidal waves $e^{i\omega t}$ are the eigenvectors of time-invariant linear systems, it is tempting to try to decompose any function f as a sum of these eigenvectors. We are then able to express Lf directly from the eigenvalues $\hat{h}(\omega)$. The Fourier analysis proves that under weak conditions on f, it is indeed possible to write it as a Fourier integral.

2.2 FOURIER INTEGRALS [1]

To avoid convergence issues, the Fourier integral is first defined over the space $\mathbf{L^1}(\mathbb{R})$ of integrable functions [57]. It is then extended to the space $\mathbf{L^2}(\mathbb{R})$ of finite energy functions [24].

2.2.1 Fourier Transform in $\mathbf{L}^1(\mathbb{R})$

The Fourier integral

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(t)\, e^{-i\omega t}\, dt \tag{2.6}$$

measures "how much" oscillations at the frequency ω there is in f. If $f \in \mathbf{L}^1(\mathbb{R})$ this integral does converge and

$$|\hat{f}(\omega)| \leq \int_{-\infty}^{+\infty} |f(t)|\, dt < +\infty. \tag{2.7}$$

The Fourier transform is thus bounded, and one can verify that it is a continuous function of ω (Problem 2.1). If $\hat{f}$ is also integrable, the following theorem gives the inverse Fourier transform.

Theorem 2.1 (INVERSE FOURIER TRANSFORM) *If $f \in \mathbf{L}^1(\mathbb{R})$ and $\hat{f} \in \mathbf{L}^1(\mathbb{R})$ then*

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega)\, e^{i\omega t}\, d\omega. \tag{2.8}$$

Proof[2]. Replacing $\hat{f}(\omega)$ by its integral expression yields

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega) \exp(i\omega t)\, d\omega = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(u) \exp[i\omega(t-u)]\, du \right) d\omega.$$

We cannot apply the Fubini Theorem A.2 directly because $f(u)\exp[i\omega(t-u)]$ is not integrable in $\mathbb{R}^2$. To avoid this technical problem, we multiply by $\exp(-\epsilon^2\omega^2/4)$ which converges to 1 when ϵ goes to 0. Let us define

$$I_\epsilon(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(u) \exp\left(\frac{-\epsilon^2\omega^2}{4}\right) \exp[i\omega(t-u)]\, du \right) d\omega. \tag{2.9}$$

We compute I_ϵ in two different ways using the Fubini theorem. The integration with respect to u gives

$$I_\epsilon(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega) \exp\left(\frac{-\epsilon^2\omega^2}{4}\right) \exp(i\omega t)\, d\omega.$$

Since

$$\left| \hat{f}(\omega) \exp\left(\frac{-\epsilon^2\omega^2}{4}\right) \exp[i\omega(t-u)] \right| \leq |\hat{f}(\omega)|$$

and since $\hat{f}$ is integrable, we can apply the dominated convergence Theorem A.1, which proves that

$$\lim_{\epsilon \to 0} I_\epsilon(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega) \exp(i\omega t)\, d\omega. \tag{2.10}$$

Let us now compute the integral (2.9) differently by applying the Fubini theorem and integrating with respect to ω:

$$I_\epsilon(t) = \int_{-\infty}^{+\infty} g_\epsilon(t-u)\, f(u)\, du, \tag{2.11}$$

with

$$g_\epsilon(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp(ix\omega)\exp\left(\frac{-\epsilon^2\omega^2}{4}\right) d\omega.$$

A change of variable $\omega' = \epsilon\omega$ shows that $g_\epsilon(x) = \epsilon^{-1} g_1(\epsilon^{-1}x)$, and it is proved in (2.32) that $g_1(x) = \pi^{-1/2} e^{-x^2}$. The Gaussian g_1 has an integral equal to 1 and a fast decay. The squeezed Gaussians g_ϵ have an integral that remains equal to 1, and thus they converge to a Dirac δ when ϵ goes to 0. By inserting (2.11) one can thus verify that

$$\lim_{\epsilon\to 0}\int_{-\infty}^{+\infty} |I_\epsilon(t) - f(t)|\,dt = \lim_{\epsilon\to 0}\int\int g_\epsilon(t-u)\,|f(u)-f(t)|\,du\,dt = 0.$$

Inserting (2.10) proves (2.8). ■

The inversion formula (2.8) decomposes f as a sum of sinusoidal waves $e^{i\omega t}$ of amplitude $\hat{f}(\omega)$. By using this formula, we can show (Problem 2.1) that the hypothesis $\hat{f} \in \mathbf{L}^1(\mathbb{R})$ implies that f must be continuous. The reconstruction (2.8) is therefore not proved for discontinuous functions. The extension of the Fourier transform to the space $\mathbf{L}^2(\mathbb{R})$ will address this issue.

The most important property of the Fourier transform for signal processing applications is the convolution theorem. It is another way to express the fact that sinusoidal waves $e^{it\omega}$ are eigenvalues of convolution operators.

Theorem 2.2 (CONVOLUTION) *Let $f \in \mathbf{L}^1(\mathbb{R})$ and $h \in \mathbf{L}^1(\mathbb{R})$. The function $g = h \star f$ is in $\mathbf{L}^1(\mathbb{R})$ and*

$$\hat{g}(\omega) = \hat{h}(\omega)\,\hat{f}(\omega). \tag{2.12}$$

Proof[1].

$$\hat{g}(\omega) = \int_{-\infty}^{+\infty} \exp(-it\omega)\left(\int_{-\infty}^{+\infty} f(t-u)\,h(u)\,du\right) dt.$$

Since $|f(t-u)||h(u)|$ is integrable in $\mathbb{R}^2$, we can apply the Fubini Theorem A.2, and the change of variable $(t,u) \to (v = t-u, u)$ yields

$$\begin{aligned}\hat{g}(\omega) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \exp[-i(u+v)\omega]\,f(v)\,h(u)\,du\,dv \\ &= \left(\int_{-\infty}^{+\infty} \exp(-iv\omega)\,f(v)\,dv\right)\left(\int_{-\infty}^{+\infty} \exp(-iu\omega)\,h(u)\,du\right),\end{aligned}$$

which verifies (2.12). ■

The response $Lf = g = f \star h$ of a linear time-invariant system can be calculated from its Fourier transform $\hat{g}(\omega) = \hat{f}(\omega)\hat{h}(\omega)$ with the inverse Fourier formula

$$g(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{g}(\omega)\,e^{i\omega t}\,d\omega, \tag{2.13}$$

which yields

$$Lf(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{h}(\omega)\,\hat{f}(\omega)\,e^{i\omega t}\,d\omega. \tag{2.14}$$

Each frequency component $e^{it\omega}$ of amplitude $\hat{f}(\omega)$ is amplified or attenuated by $\hat{h}(\omega)$. Such a convolution is thus called a *frequency filtering*, and $\hat{h}$ is the *transfer function* of the filter.

The following table summarizes important properties of the Fourier transform, often used in calculations. Most of these formulas are proved with a change of variable in the Fourier integral.

Property	Function	Fourier Transform	
	$f(t)$	$\hat{f}(\omega)$	
Inverse	$\hat{f}(t)$	$2\pi f(-\omega)$	(2.15)
Convolution	$f_1 \star f_2(t)$	$\hat{f}_1(\omega)\,\hat{f}_2(\omega)$	(2.16)
Multiplication	$f_1(t)\,f_2(t)$	$\frac{1}{2\pi}\hat{f}_1 \star \hat{f}_2(\omega)$	(2.17)
Translation	$f(t-u)$	$e^{-iu\omega}\,\hat{f}(\omega)$	(2.18)
Modulation	$e^{i\xi t} f(t)$	$\hat{f}(\omega-\xi)$	(2.19)
Scaling	$f(t/s)$	$\lvert s\rvert\,\hat{f}(s\omega)$	(2.20)
Time derivatives	$f^{(p)}(t)$	$(i\omega)^p\,\hat{f}(\omega)$	(2.21)
Frequency derivatives	$(-it)^p f(t)$	$\hat{f}^{(p)}(\omega)$	(2.22)
Complex conjugate	$f^*(t)$	$\hat{f}^*(-\omega)$	(2.23)
Hermitian symmetry	$f(t) \in \mathbb{R}$	$\hat{f}(-\omega) = \hat{f}^*(\omega)$	(2.24)

2.2.2 Fourier Transform in $\mathbf{L}^2(\mathbb{R})$

The Fourier transform of the indicator function $f = \mathbf{1}_{[-1,1]}$ is

$$\hat{f}(\omega) = \int_{-1}^{1} e^{-i\omega t}\,dt = \frac{2\sin\omega}{\omega}.$$

This function is not integrable because f is not continuous, but its square is integrable. The inverse Fourier transform Theorem 2.1 thus does not apply. This motivates the extension of the Fourier transform to the space $\mathbf{L}^2(\mathbb{R})$ of functions f with a finite energy $\int_{-\infty}^{+\infty} |f(t)|^2\,dt < +\infty$. By working in the Hilbert space $\mathbf{L}^2(\mathbb{R})$, we also have access to all the facilities provided by the existence of an inner product. The inner product of $f \in \mathbf{L}^2(\mathbb{R})$ and $g \in \mathbf{L}^2(\mathbb{R})$ is

$$\langle f, g\rangle = \int_{-\infty}^{+\infty} f(t)\,g^*(t)\,dt,$$

and the resulting norm in $\mathbf{L}^2(\mathbb{R})$ is

$$\|f\|^2 = \langle f, f\rangle = \int_{-\infty}^{+\infty} |f(t)|^2\,dt.$$

The following theorem proves that inner products and norms in $\mathbf{L}^2(\mathbb{R})$ are conserved by the Fourier transform up to a factor of 2π. Equations (2.25) and (2.26) are called respectively the *Parseval* and *Plancherel* formulas.

Theorem 2.3 *If f and h are in $\mathbf{L}^1(\mathbb{R}) \cap \mathbf{L}^2(\mathbb{R})$ then*

$$\int_{-\infty}^{+\infty} f(t)\, h^*(t)\, dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega)\, \hat{h}^*(\omega)\, d\omega. \tag{2.25}$$

For $h = f$ it follows that

$$\int_{-\infty}^{+\infty} |f(t)|^2\, dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\hat{f}(\omega)|^2\, d\omega. \tag{2.26}$$

Proof[1]. Let $g = f \star \bar{h}$ with $\bar{h}(t) = h^*(-t)$. The convolution Theorem 2.2 and property (2.23) show that $\hat{g}(\omega) = \hat{f}(\omega)\hat{h}^*(\omega)$. The reconstruction formula (2.8) applied to $g(0)$ yields

$$\int_{-\infty}^{+\infty} f(t)\, h^*(t)\, dt = g(0) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{g}(\omega)\, d\omega = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega)\, \hat{h}^*(\omega)\, d\omega.$$

■

Density Extension in $\mathbf{L}^2(\mathbb{R})$ If $f \in \mathbf{L}^2(\mathbb{R})$ but $f \notin \mathbf{L}^1(\mathbb{R})$, its Fourier transform cannot be calculated with the Fourier integral (2.6) because $f(t)\, e^{i\omega t}$ is not integrable. It is defined as a limit using the Fourier transforms of functions in $\mathbf{L}^1(\mathbb{R}) \cap \mathbf{L}^2(\mathbb{R})$.

Since $\mathbf{L}^1(\mathbb{R}) \cap \mathbf{L}^2(\mathbb{R})$ is dense in $\mathbf{L}^2(\mathbb{R})$, one can find a family $\{f_n\}_{n \in \mathbb{Z}}$ of functions in $\mathbf{L}^1(\mathbb{R}) \cap \mathbf{L}^2(\mathbb{R})$ that converges to f:

$$\lim_{n \to +\infty} \|f - f_n\| = 0.$$

Since $\{f_n\}_{n \in \mathbb{Z}}$ converges, it is a Cauchy sequence, which means that $\|f_n - f_p\|$ is arbitrarily small if n and p are large enough. Moreover, $f_n \in \mathbf{L}^1(\mathbb{R})$, so its Fourier transform $\hat{f}_n$ is well defined. The Plancherel formula (2.26) proves that $\{\hat{f}_n\}_{n \in \mathbb{Z}}$ is also a Cauchy sequence because

$$\|\hat{f}_n - \hat{f}_p\| = \sqrt{2\pi}\, \|f_n - f_p\|$$

is arbitrarily small for n and p large enough. A Hilbert space (Appendix A.2) is complete, which means that all Cauchy sequences converge to an element of the space. Hence, there exists $\hat{f} \in \mathbf{L}^2(\mathbb{R})$ such that

$$\lim_{n \to +\infty} \|\hat{f} - \hat{f}_n\| = 0.$$

By definition, $\hat{f}$ is the Fourier transform of f. This extension of the Fourier transform to $\mathbf{L}^2(\mathbb{R})$ satisfies the convolution theorem, the Parseval and Plancherel formulas, as well as all properties (2.15-2.24).

Diracs Diracs are often used in calculations; their properties are summarized in Appendix A.7. A Dirac δ associates to a function its value at $t = 0$. Since $e^{i\omega t} = 1$ at $t = 0$ it seems reasonable to define its Fourier transform by

$$\hat{\delta}(\omega) = \int_{-\infty}^{+\infty} \delta(t) e^{-i\omega t} dt = 1. \tag{2.27}$$

This formula is justified mathematically by the extension of the Fourier transform to tempered distributions [66, 69].

2.2.3 Examples

The following examples often appear in Fourier calculations. They also illustrate important Fourier transform properties.

- The *indicator function* $f = \mathbf{1}_{[-T,T]}$ is discontinuous at $t = \pm T$. Its Fourier transform is therefore not integrable:

$$\hat{f}(\omega) = \int_{-T}^{T} e^{-i\omega t} dt = \frac{2\sin(T\omega)}{\omega}. \tag{2.28}$$

- An *ideal low-pass filter* has a transfer function $\hat{h} = \mathbf{1}_{[-\xi,\xi]}$ that selects low frequencies over $[-\xi, \xi]$. The impulse response is calculated with the inverse Fourier integral (2.8):

$$h(t) = \frac{1}{2\pi} \int_{-\xi}^{\xi} e^{i\omega t} d\omega = \frac{\sin(\xi t)}{\pi t}. \tag{2.29}$$

- A *passive electronic circuit* implements analog filters with resistances, capacities and inductors. The input voltage $f(t)$ is related to the output voltage $g(t)$ by a differential equation with constant coefficients:

$$\sum_{k=0}^{K} a_k f^{(k)}(t) = \sum_{k=0}^{M} b_k g^{(k)}(t). \tag{2.30}$$

Suppose that the circuit is not charged for $t < 0$, which means that $f(t) = g(t) = 0$. The output g is a linear time-invariant function of f and can thus be written $g = f \star h$. Computing the Fourier transform of (2.30) and applying (2.22) proves that

$$\hat{h}(\omega) = \frac{\hat{g}(\omega)}{\hat{f}(\omega)} = \frac{\sum_{k=0}^{K} a_k (i\omega)^k}{\sum_{k=0}^{M} b_k (i\omega)^k}. \tag{2.31}$$

It is therefore a rational function of $i\omega$. An ideal low-pass transfer function $\mathbf{1}_{[-\xi,\xi]}$ thus cannot be implemented by an analog circuit. It must be approximated by a rational function. Chebyshev or Butterworth filters are often used for this purpose [14].

- A *Gaussian* $f(t) = \exp(-t^2)$ is a $\mathbf{C}^\infty$ function with a fast asymptotic decay. Its Fourier transform is also a Gaussian:

$$\hat{f}(\omega) = \sqrt{\pi}\exp(-\omega^2/4). \tag{2.32}$$

This Fourier transform is computed by showing with an integration by parts that $\hat{f}(\omega) = \int_{-\infty}^{+\infty} \exp(-t^2)\,e^{-i\omega t}dt$ is differentiable and satisfies the differential equation

$$2\hat{f}'(\omega) + \omega\hat{f}(\omega) = 0. \tag{2.33}$$

The solution of this equation is a Gaussian $\hat{f}(\omega) = K\exp(-\omega^2/4)$, and since $\hat{f}(0) = \int_{-\infty}^{+\infty} \exp(-t^2)\,dt = \sqrt{\pi}$, we obtain (2.32).

- A Gaussian *chirp* $f(t) = \exp[-(a-ib)t^2]$ has a Fourier transform calculated with a similar differential equation:

$$\hat{f}(\omega) = \sqrt{\frac{\pi}{a-ib}}\exp\left(\frac{-(a+ib)\omega^2}{4(a^2+b^2)}\right). \tag{2.34}$$

- A translated *Dirac* $\delta_\tau(t) = \delta(t-\tau)$ has a Fourier transform calculated by evaluating $e^{-i\omega t}$ at $t = \tau$:

$$\hat{\delta}_\tau(\omega) = \int_{-\infty}^{+\infty} \delta(t-\tau)\,e^{-i\omega t}\,dt = e^{-i\omega\tau}. \tag{2.35}$$

- The *Dirac comb* is a sum of translated Diracs

$$c(t) = \sum_{n=-\infty}^{+\infty} \delta(t-nT)$$

that is used to uniformly sample analog signals. Its Fourier transform is derived from (2.35):

$$\hat{c}(\omega) = \sum_{n=-\infty}^{+\infty} e^{-inT\omega}. \tag{2.36}$$

The Poisson formula proves that it is also equal to a Dirac comb with a spacing equal to $2\pi/T$.

Theorem 2.4 (POISSON FORMULA) *In the sense of distribution equalities (A.32),*

$$\sum_{n=-\infty}^{+\infty} e^{-inT\omega} = \frac{2\pi}{T}\sum_{k=-\infty}^{+\infty} \delta\left(\omega - \frac{2\pi k}{T}\right). \tag{2.37}$$

Proof[2]. The Fourier transform $\hat{c}$ in (2.36) is periodic with period $2\pi/T$. To verify the Poisson formula, it is therefore sufficient to prove that the restriction of $\hat{c}$ to $[-\pi/T, \pi/T]$ is equal to $2\pi/T\,\delta$. The formula (2.37) is proved in the sense of a distribution equality (A.32) by showing that for any test function $\hat{\phi}(\omega)$ with a support included in $[-\pi/T, \pi/T]$,

$$\langle \hat{c}, \hat{\phi} \rangle = \lim_{N \to +\infty} \int_{-\infty}^{+\infty} \sum_{n=-N}^{N} \exp(-inT\omega)\, \hat{\phi}(\omega)\, d\omega = \frac{2\pi}{T}\, \hat{\phi}(0).$$

The sum of the geometric series is

$$\sum_{n=-N}^{N} \exp(-inT\omega) = \frac{\sin[(N+1/2)T\omega]}{\sin[T\omega/2]}. \tag{2.38}$$

Hence

$$\langle \hat{c}, \hat{\phi} \rangle = \lim_{N \to +\infty} \frac{2\pi}{T} \int_{-\pi/T}^{\pi/T} \frac{\sin[(N+1/2)T\omega]}{\pi\omega} \frac{T\omega/2}{\sin[T\omega/2]}\, \hat{\phi}(\omega)\, d\omega. \tag{2.39}$$

Let

$$\hat{\psi}(\omega) = \begin{cases} \hat{\phi}(\omega) \frac{T\omega/2}{\sin[T\omega/2]} & \text{if } |\omega| \leq \pi/T \\ 0 & \text{if } |\omega| > \pi/T \end{cases}$$

and $\psi(t)$ be the inverse Fourier transform of $\hat{\psi}(\omega)$. Since $2\omega^{-1}\sin(a\omega)$ is the Fourier transform of $\mathbf{1}_{[-a,a]}(t)$, the Parseval formula (2.25) implies

$$\begin{aligned} \langle \hat{c}, \hat{\phi} \rangle &= \lim_{N \to +\infty} \frac{2\pi}{T} \int_{-\infty}^{+\infty} \frac{\sin[(N+1/2)T\omega]}{\pi\omega}\, \hat{\psi}(\omega)\, d\omega \\ &= \lim_{N \to +\infty} \frac{2\pi}{T} \int_{-(N+1/2)T}^{(N+1/2)T} \psi(t)\, dt. \end{aligned} \tag{2.40}$$

When N goes to $+\infty$ the integral converges to $\hat{\psi}(0) = \hat{\phi}(0)$. ■

2.3 PROPERTIES [1]

2.3.1 Regularity and Decay

The global regularity of a signal f depends on the decay of $|\hat{f}(\omega)|$ when the frequency ω increases. The differentiability of f is studied. If $\hat{f} \in \mathbf{L}^1(\mathbb{R})$, then the Fourier inversion formula (2.8) implies that f is continuous and bounded:

$$|f(t)| \leq \frac{1}{2\pi} \int_{-\infty}^{+\infty} |e^{i\omega t}\, \hat{f}(\omega)|\, d\omega = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\hat{f}(\omega)|\, d\omega < +\infty. \tag{2.41}$$

The next proposition applies this property to obtain a sufficient condition that guarantees the differentiability of f at any order p.

Proposition 2.1 *A function f is bounded and p times continuously differentiable with bounded derivatives if*

$$\int_{-\infty}^{+\infty} |\hat{f}(\omega)|\,(1+|\omega|^p)\,d\omega < +\infty. \tag{2.42}$$

Proof[1]. The Fourier transform of the k^{th} order derivative $f^{(k)}(t)$ is $(i\omega)^k \hat{f}(\omega)$. Applying (2.41) to this derivative proves that

$$|f^{(k)}(t)| \leq \int_{-\infty}^{+\infty} |\hat{f}(\omega)|\,|\omega|^k\,d\omega.$$

Condition (2.42) implies that $\int_{-\infty}^{+\infty} |\hat{f}(\omega)||\omega|^k\,d\omega < +\infty$ for any $k \leq p$, so $f^{(k)}(t)$ is continuous and bounded. ■

This result proves that if there exist a constant K and $\epsilon > 0$ such that

$$|\hat{f}(\omega)| \leq \frac{K}{1+|\omega|^{p+1+\epsilon}}, \quad \text{then} \quad f \in \mathbf{C}^p.$$

If $\hat{f}$ has a compact support then (2.42) implies that $f \in \mathbf{C}^\infty$.

The decay of $|\hat{f}(\omega)|$ depends on the worst singular behavior of f. For example, $f = \mathbf{1}_{[-T,T]}$ is discontinuous at $t = \pm T$, so $|\hat{f}(\omega)|$ decays like $|\omega|^{-1}$. In this case, it could also be important to know that $f(t)$ is regular for $t \neq \pm T$. This information cannot be derived from the decay of $|\hat{f}(\omega)|$. To characterize local regularity of a signal f it is necessary to decompose it over waveforms that are well localized in time, as opposed to sinusoidal waves $e^{i\omega t}$. Section 6.1.3 explains that wavelets are particularly well adapted to this purpose.

2.3.2 Uncertainty Principle

Can we construct a function f whose energy is well localized in time and whose Fourier transform $\hat{f}$ has an energy concentrated in a small frequency neighborhood? The Dirac $\delta(t-u)$ has a support restricted to $t = u$ but its Fourier transform $e^{-iu\omega}$ has an energy uniformly spread over all frequencies. We know that $|\hat{f}(\omega)|$ decays quickly at high frequencies only if f has regular variations in time. The energy of f must therefore be spread over a relatively large domain.

To reduce the time spread of f, we can scale it by $s < 1$ while maintaining constant its total energy. If

$$f_s(t) = \frac{1}{\sqrt{s}} f\left(\frac{t}{s}\right) \quad \text{then} \quad \|f_s\|^2 = \|f\|^2.$$

The Fourier transform $\hat{f}_s(\omega) = \sqrt{s}\,\hat{f}(s\omega)$ is dilated by $1/s$ so we lose in frequency localization what we gained in time. Underlying is a trade-off between time and frequency localization.

Time and frequency energy concentrations are restricted by the Heisenberg uncertainty principle. This principle has a particularly important interpretation in

quantum mechanics as an uncertainty as to the position and momentum of a free particle. The state of a one-dimensional particle is described by a wave function $f \in \mathbf{L}^2(\mathbb{R})$. The probability density that this particle is located at t is $\frac{1}{\|f\|^2}|f(t)|^2$. The probability density that its momentum is equal to ω is $\frac{1}{2\pi\|f\|^2}|\hat{f}(\omega)|^2$. The average location of this particle is

$$u = \frac{1}{\|f\|^2}\int_{-\infty}^{+\infty} t\,|f(t)|^2\,dt, \tag{2.43}$$

and the average momentum is

$$\xi = \frac{1}{2\pi\|f\|^2}\int_{-\infty}^{+\infty} \omega\,|\hat{f}(\omega)|^2\,d\omega. \tag{2.44}$$

The variances around these average values are respectively

$$\sigma_t^2 = \frac{1}{\|f\|^2}\int_{-\infty}^{+\infty} (t-u)^2\,|f(t)|^2\,dt \tag{2.45}$$

and

$$\sigma_\omega^2 = \frac{1}{2\pi\|f\|^2}\int_{-\infty}^{+\infty} (\omega-\xi)^2\,|\hat{f}(\omega)|^2\,d\omega. \tag{2.46}$$

The larger σ_t, the more uncertainty there is concerning the position of the free particle; the larger σ_ω, the more uncertainty there is concerning its momentum.

Theorem 2.5 (HEISENBERG UNCERTAINTY) *The temporal variance and the frequency variance of $f \in \mathbf{L}^2(\mathbb{R})$ satisfy*

$$\sigma_t^2\,\sigma_\omega^2 \geq \frac{1}{4}. \tag{2.47}$$

This inequality is an equality if and only if there exist $(u,\xi,a,b) \in \mathbb{R}^2 \times \mathbb{C}^2$ such that

$$f(t) = a\exp[i\xi t - b(t-u)^2]. \tag{2.48}$$

Proof[2]. The following proof due to Weyl [75] supposes that $\lim_{|t|\to+\infty}\sqrt{t}\,f(t) = 0$, but the theorem is valid for any $f \in \mathbf{L}^2(\mathbb{R})$. If the average time and frequency localization of f is u and ξ, then the average time and frequency location of $\exp(-i\xi t)\,f(t+u)$ is zero. It is thus sufficient to prove the theorem for $u = \xi = 0$. Observe that

$$\sigma_t^2\,\sigma_\omega^2 = \frac{1}{2\pi\|f\|^4}\int_{-\infty}^{+\infty} |t\,f(t)|^2\,dt\int_{-\infty}^{+\infty} |\omega\,\hat{f}(\omega)|^2\,d\omega. \tag{2.49}$$

Since $i\omega\hat{f}(\omega)$ is the Fourier transform of $f'(t)$, the Plancherel identity (2.26) applied to $i\omega\hat{f}(\omega)$ yields

$$\sigma_t^2\sigma_\omega^2 = \frac{1}{\|f\|^4}\int_{-\infty}^{+\infty} |t\,f(t)|^2\,dt\int_{-\infty}^{+\infty} |f'(t)|^2\,dt. \tag{2.50}$$

Schwarz's inequality implies

$$\begin{aligned}\sigma_t^2\sigma_\omega^2 &\geq \frac{1}{\|f\|^4}\left[\int_{-\infty}^{+\infty}|t\,f'(t)\,f^*(t)|\,dt\right]^2\\ &\geq \frac{1}{\|f\|^4}\left[\int_{-\infty}^{+\infty}\frac{t}{2}\left[f'(t)\,f^*(t)+f'^*(t)\,f(t)\right]dt\right]^2\\ &\geq \frac{1}{4\|f\|^4}\left[\int_{-\infty}^{+\infty}t\,(|f(t)|^2)'\,dt\right]^2.\end{aligned}$$

Since $\lim_{|t|\to+\infty}\sqrt{t}\,f(t)=0$, an integration by parts gives

$$\sigma_t^2\sigma_\omega^2 \geq \frac{1}{4\|f\|^4}\left[\int_{-\infty}^{+\infty}|f(t)|^2\,dt\right]^2=\frac{1}{4}. \tag{2.51}$$

To obtain an equality, Schwarz's inequality applied to (2.50) must be an equality. This implies that there exists $b\in\mathbb{C}$ such that

$$f'(t)=-2bt\,f(t). \tag{2.52}$$

Hence, there exists $a\in\mathbb{C}$ such that $f(t)=a\exp(-bt^2)$. The other steps of the proof are then equalities so that the lower bound is indeed reached. When $u\neq 0$ and $\xi\neq 0$ the corresponding time and frequency translations yield (2.48). ■

In quantum mechanics, this theorem shows that we cannot reduce arbitrarily the uncertainty as to the position and the momentum of a free particle. In signal processing, the modulated Gaussians (2.48) that have a minimum joint time-frequency localization are called Gabor chirps. As expected, they are smooth functions with a fast time asymptotic decay.

Compact Support Despite the Heisenberg uncertainty bound, we might still be able to construct a function of compact support whose Fourier transform has a compact support. Such a function would be very useful in constructing a finite impulse response filter with a band-limited transfer function. Unfortunately, the following theorem proves that it does not exist.

Theorem 2.6 *If $f\neq 0$ has a compact support then $\hat{f}(\omega)$ cannot be zero on a whole interval. Similarly, if $\hat{f}\neq 0$ has a compact support then $f(t)$ cannot be zero on a whole interval.*

Proof[2]. We prove only the first statement, since the second is derived from the first by applying the Fourier transform. If $\hat{f}$ has a compact support included in $[-b,b]$ then

$$f(t)=\frac{1}{2\pi}\int_{-b}^{b}\hat{f}(\omega)\exp(i\omega t)\,d\omega. \tag{2.53}$$

If $f(t)=0$ for $t\in[c,d]$, by differentiating n times under the integral at $t_0=(c+d)/2$, we obtain

$$f^{(n)}(t_0)=\frac{1}{2\pi}\int_{-b}^{b}\hat{f}(\omega)\,(i\omega)^n\exp(i\omega t_0)\,d\omega=0. \tag{2.54}$$

Since

$$f(t) = \frac{1}{2\pi} \int_{-b}^{b} \hat{f}(\omega) \exp[i\omega(t-t_0)] \exp(i\omega t_0)\, d\omega, \tag{2.55}$$

developing $\exp[i\omega(t-t_0)]$ as an infinite series yields for all $t \in \mathbb{R}$

$$f(t) = \frac{1}{2\pi} \sum_{n=0}^{+\infty} \frac{[i(t-t_0)]^n}{n!} \int_{-b}^{b} \hat{f}(\omega)\, \omega^n \exp(i\omega t_0)\, d\omega = 0. \tag{2.56}$$

This contradicts our assumption that $f \neq 0$. ■

2.3.3 Total Variation

The total variation measures the total amplitude of signal oscillations. It plays an important role in image processing, where its value depends on the length of the image level sets. We show that a low-pass filter can considerably amplify the total variation by creating Gibbs oscillations.

Variations and Oscillations If f is differentiable, its total variation is defined by

$$\|f\|_V = \int_{-\infty}^{+\infty} |f'(t)|\, dt. \tag{2.57}$$

If $\{x_p\}_p$ are the abscissa of the local extrema of f where $f'(x_p) = 0$, then

$$\|f\|_V = \sum_p |f(x_{p+1}) - f(x_p)|.$$

It thus measures the total amplitude of the oscillations of f. For example, if $f(t) = \exp(-t^2)$, then $\|f\|_V = 2$. If $f(t) = \sin(\pi t)/(\pi t)$, then f has a local extrema at $x_p \in [p, p+1]$ for any $p \in \mathbb{Z}$. Since $|f(x_{p+1}) - f(x_p)| \sim |p|^{-1}$, we derive that $\|f\|_V = +\infty$.

The total variation of non-differentiable functions can be calculated by considering the derivative in the general sense of distributions [66, 79]. This is equivalent to approximating the derivative by a finite difference on an interval h that goes to zero:

$$\|f\|_V = \lim_{h \to 0} \int_{-\infty}^{+\infty} \frac{|f(t) - f(t-h)|}{|h|}\, dt. \tag{2.58}$$

The total variation of discontinuous functions is thus well defined. For example, if $f = \mathbf{1}_{[a,b]}$ then (2.58) gives $\|f\|_V = 2$. We say that f has a *bounded variation* if $\|f\|_V < +\infty$.

Whether f' is the standard derivative of f or its generalized derivative in the sense of distributions, its Fourier transform is $\widehat{f'}(\omega) = i\omega \hat{f}(\omega)$. Hence

$$|\omega|\, |\hat{f}(\omega)| \leq \int_{-\infty}^{+\infty} |f'(t)| dt = \|f\|_V\,,$$

which implies that

$$|\hat{f}(\omega)| \leq \frac{\|f\|_V}{|\omega|} . \tag{2.59}$$

However, $|\hat{f}(\omega)| = O(|\omega|^{-1})$ is not a sufficient condition to guarantee that f has bounded variation. For example, if $f(t) = \sin(\pi t)/(\pi t)$, then $\hat{f} = \mathbf{1}_{[-\pi,\pi]}$ satisfies $|\hat{f}(\omega)| \leq \pi |\omega|^{-1}$ although $\|f\|_V = +\infty$. In general, the total variation of f cannot be evaluated from $|\hat{f}(\omega)|$.

Discrete Signals Let $f_N[n] = f(n/N)$ be a discrete signal obtained with a uniform sampling at intervals N^{-1}. The discrete total variation is calculated by approximating the signal derivative by a finite difference over the sampling distance $h = N^{-1}$, and replacing the integral (2.58) by a Riemann sum, which gives:

$$\|f_N\|_V = \sum_n |f_N[n] - f_N[n-1]| . \tag{2.60}$$

If n_p are the abscissa of the local extrema of f_N, then

$$\|f_N\|_V = \sum_p |f_N[n_{p+1}] - f_N[n_p]| .$$

The total variation thus measures the total amplitude of the oscillations of f. In accordance with (2.58), we say that the discrete signal has a *bounded variation* if $\|f_N\|_V$ is bounded by a constant independent of the resolution N.

Gibbs Oscillations Filtering a signal with a low-pass filter can create oscillations that have an infinite total variation. Let $f_\xi = f \star h_\xi$ be the filtered signal obtained with an ideal low-pass filter whose transfer function is $\hat{h}_\xi = \mathbf{1}_{[-\xi,\xi]}$. If $f \in \mathbf{L}^2(\mathbb{R})$, then f_ξ converges to f in $\mathbf{L}^2(\mathbb{R})$ norm: $\lim_{\xi \to +\infty} \|f - f_\xi\| = 0$. Indeed, $\hat{f}_\xi = \hat{f}\,\mathbf{1}_{[-\xi,\xi]}$ and the Plancherel formula (2.26) implies that

$$\|f - f_\xi\|^2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\hat{f}(\omega) - \hat{f}_\xi(\omega)|^2 \, d\omega = \frac{1}{2\pi} \int_{|\omega|>\xi} |\hat{f}(\omega)|^2 \, d\omega,$$

which goes to zero as ξ increases. However, if f is discontinuous in t_0, then we show that f_ξ has Gibbs oscillations in the neighborhood of t_0, which prevents $\sup_{t \in \mathbb{R}} |f(t) - f_\xi(t)|$ from converging to zero as ξ increases.

Let f be a bounded variation function $\|f\|_V < +\infty$ that has an isolated discontinuity at t_0, with a left limit $f(t_0^-)$ and right limit $f(t_0^+)$. It is decomposed as a sum of f_c, which is continuous in the neighborhood of t_0, plus a Heaviside step of amplitude $f(t_0^+) - f(t_0^-)$:

$$f(t) = f_c(t) + [f(t_0^+) - f(t_0^-)]\, u(t - t_0),$$

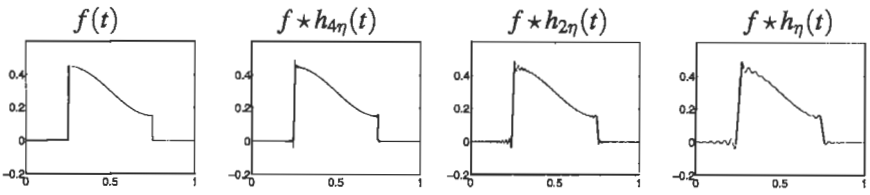


FIGURE 2.1 Gibbs oscillations created by low-pass filters with cut-off frequencies that decrease from left to right.

with

$$u(t) = \begin{cases} 1 & \text{if } t \geq 0 \\ 0 & \text{otherwise} \end{cases} . \tag{2.61}$$

Hence

$$f_\xi(t) = f_c \star h_\xi(t) + [f(t_0^+) - f(t_0^-)]\, u \star h_\xi(t - t_0). \tag{2.62}$$

Since f_c has bounded variation and is uniformly continuous in the neighborhood of t_0, one can prove (Problem 2.13) that $f_c \star h_\xi(t)$ converges uniformly to $f_c(t)$ in a neighborhood of t_0. The following proposition shows that this is not true for $u \star h_\xi$, which creates Gibbs oscillations.

Proposition 2.2 (GIBBS) *For any* $\xi > 0$,

$$u \star h_\xi(t) = \int_{-\infty}^{\xi t} \frac{\sin x}{\pi x}\, dx. \tag{2.63}$$

Proof[2]. The impulse response of an ideal low-pass filter, calculated in (2.29), is $h_\xi(t) = \sin(\xi t)/(\pi t)$. Hence

$$u \star h_\xi(t) = \int_{-\infty}^{+\infty} u(\tau)\, \frac{\sin \xi(t-\tau)}{\pi(t-\tau)}\, d\tau = \int_{0}^{+\infty} \frac{\sin \xi(t-\tau)}{\pi(t-\tau)}\, d\tau.$$

The change of variable $x = \xi(t - \tau)$ gives (2.63). ■

The function

$$s(\xi t) = \int_{-\infty}^{\xi t} \frac{\sin x}{\pi x}\, dx$$

is a sigmoid that increases from 0 at $t = -\infty$ to 1 at $t = +\infty$, with $s(0) = 1/2$. It has oscillations of period π/ξ, which are attenuated when the distance to 0 increases, but their total variation is infinite: $\|s\|_V = +\infty$. The maximum amplitude of the Gibbs oscillations occurs at $t = \pm\pi/\xi$, with an amplitude independent of ξ:

$$A = s(\pi) - 1 = \int_{-\infty}^{\pi} \frac{\sin x}{\pi x} dx - 1 \approx 0.045\ .$$

Inserting (2.63) in (2.62) shows that

$$f(t) - f_\xi(t) = [f(t_0^+) - f(t_0^-)]\, s(\xi(t-t_0)) + \epsilon(\xi,t)\,, \tag{2.64}$$

where $\lim_{\xi\to+\infty} \sup_{|t-t_0|<\alpha} |\epsilon(\xi,t)| = 0$ in some neighborhood of size $\alpha > 0$ around t_0. The sigmoid $s(\xi(t-t_0))$ centered at t_0 creates a maximum error of fixed amplitude for all ξ. This is seen in Figure 2.1, where the Gibbs oscillations have an amplitude proportional to the jump $f(t_0^+) - f(t_0^-)$ at all frequencies ξ.

Image Total Variation The total variation of an image $f(x_1,x_2)$ depends on the amplitude of its variations as well as the length of the contours along which they occur. Suppose that $f(x_1,x_2)$ is differentiable. The total variation is defined by

$$\|f\|_V = \int\int |\vec{\nabla} f(x_1,x_2)|\, dx_1\, dx_2\,, \tag{2.65}$$

where the modulus of the gradient vector is

$$|\vec{\nabla} f(x_1,x_2)| = \left(\left| \frac{\partial f(x_1,x_2)}{\partial x_1} \right|^2 + \left| \frac{\partial f(x_1,x_2)}{\partial x_2} \right|^2 \right)^{1/2}.$$

As in one dimension, the total variation is extended to discontinuous functions by taking the derivatives in the general sense of distributions. An equivalent norm is obtained by approximating the partial derivatives by finite differences:

$$|\Delta_h f(x_1,x_2)| = \left(\left| \frac{f(x_1,x_2) - f(x_1-h,x_2)}{h} \right|^2 + \left| \frac{f(x_1,x_2) - f(x_1,x_2-h)}{h} \right|^2 \right)^{1/2}.$$

One can verify that

$$\|f\|_V \leq \lim_{h\to 0} \int\int |\Delta_h f(x_1,x_2)|\, dx_1\, dx_2 \leq \sqrt{2}\, \|f\|_V. \tag{2.66}$$

The finite difference integral gives a larger value when $f(x_1,x_2)$ is discontinuous along a diagonal line in the (x_1,x_2) plane.

The total variation of f is related to the length of it level sets. Let us define

$$\Omega_y = \{(x_1,x_2) \in \mathbb{R}^2 \;:\; f(x_1,x_2) > y\}\,.$$

If f is continuous then the boundary $\partial\Omega_y$ of Ω_y is the level set of all (x_1,x_2) such that $f(x_1,x_2) = y$. Let $H^1(\partial\Omega_y)$ be the length of $\partial\Omega_y$. Formally, this length is calculated in the sense of the monodimensional Hausdorff measure. The following theorem relates the total variation of f to the length of its level sets.

Theorem 2.7 (CO-AREA FORMULA) *If* $\|f\|_V < +\infty$ *then*

$$\|f\|_V = \int_{-\infty}^{+\infty} H^1(\partial\Omega_y)\, dy. \tag{2.67}$$

Proof[2]. The proof is a highly technical result that is given in [79]. We give an intuitive explanation when f is continuously differentiable. In this case $\partial\Omega_y$ is a differentiable curve $x(y,s) \in \mathbb{R}^2$, which is parameterized by the arc-length s. Let $\vec{\tau}(x)$ be the vector tangent to this curve in the plane. The gradient $\vec{\nabla} f(x)$ is orthogonal to $\vec{\tau}(x)$. The Frenet coordinate system along $\partial\Omega_y$ is composed of $\vec{\tau}(x)$ and of the unit vector $\vec{n}(x)$ parallel to $\vec{\nabla} f(x)$. Let ds and dn be the Lebesgue measures in the direction of $\vec{\tau}$ and $\vec{n}$. We have

$$|\vec{\nabla} f(x)| = \vec{\nabla} f(x) \cdot \vec{n} = \frac{dy}{dn}, \tag{2.68}$$

where dy is the differential of amplitudes across level sets. The idea of the proof is to decompose the total variation integral over the plane as an integral along the level sets and across level sets, which we write:

$$\|f\|_V = \int\int |\vec{\nabla} f(x_1,x_2)|\, dx_1\, dx_2 = \int \int_{\partial\Omega_y} |\vec{\nabla} f(x(y,s))|\, ds\, dn. \tag{2.69}$$

By using (2.68) we can get

$$\|f\|_V = \int \int_{\partial\Omega_y} ds\, dy\,.$$

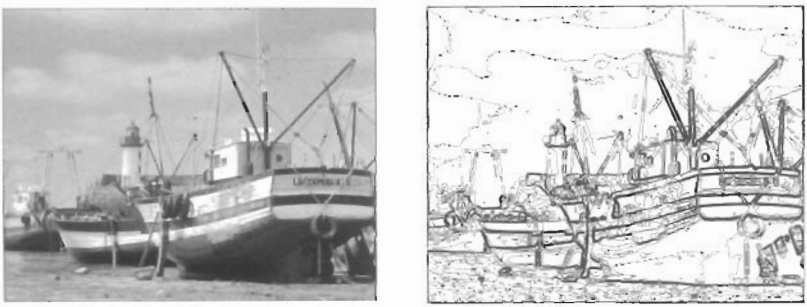
But $\int_{\partial\Omega_y} ds = H^1(\partial\Omega_y)$ is the length of the level set, which justifies (2.67). ■

The co-area formula gives an important geometrical interpretation of the total image variation. Images are uniformly bounded so the integral (2.67) is calculated over a finite interval and is proportional to the average length of level sets. It is finite as long as the level sets are not fractal curves. Let $f = \alpha \mathbf{1}_\Omega$ be proportional to the indicator function of a set $\Omega \subset \mathbb{R}^2$ which has a boundary $\partial\Omega$ of length L. The co-area formula (2.7) implies that $\|f\|_V = \alpha L$. In general, bounded variation images must have step edges of finite length.

Discrete Images A camera measures light intensity with photoreceptors that perform a uniform sampling over a grid that is supposed to be uniform. For a resolution N, the sampling interval is N^{-1} and the resulting image can be written $f_N[n_1,n_2] = f(n_1/N, n_2/N)$. Its total variation is defined by approximating derivatives by finite differences and the integral (2.66) by a Riemann sum:

$$\|f_N\|_V = \frac{1}{N} \sum_{n_1} \sum_{n_2} \Big(\Big| f[n_1,n_2] - f[n_1-1,n_2] \Big|^2 + \Big| f[n_1,n_2] - f[n_1,n_2-1] \Big|^2 \Big)^{1/2}. \tag{2.70}$$

In accordance with (2.66) we say that the image has bounded variation if $\|f_N\|_V$ is bounded by a constant independent of the resolution N. The co-area formula proves



(a) (b)

FIGURE 2.2 (a): The total variation of this image remains nearly constant when the resolution N increases. (b): Level sets $\partial\Omega_y$ obtained by sampling uniformly the amplitude variable y.

that it depends on the length of the level sets as the image resolution increases. The $\sqrt{2}$ upper bound factor in (2.66) comes from the fact that the length of a diagonal line can be increased by $\sqrt{2}$ if it is approximated by a zig-zag line that remains on the horizontal and vertical segments of the image sampling grid. Figure 2.2(a) shows a bounded variation image and Figure 2.2(b) displays the level sets obtained by discretizing uniformly the amplitude variable y. The total variation of this image remains nearly constant as the resolution varies.

2.4 TWO-DIMENSIONAL FOURIER TRANSFORM [1]

The Fourier transform in $\mathbb{R}^n$ is a straightforward extension of the one-dimensional Fourier transform. The two-dimensional case is briefly reviewed for image processing applications. The Fourier transform of a two-dimensional integrable function $f \in \mathbf{L}^1(\mathbb{R}^2)$ is

$$\hat{f}(\omega_1,\omega_2) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x_1,x_2)\exp[-i(\omega_1 x_1+\omega_2 x_2)]\,dx_1\,dx_2. \qquad (2.71)$$

In polar coordinates $\exp[i(\omega_1 x+\omega_2 y)]$ can be rewritten

$$\exp[i(\omega_1 x_1+\omega_2 x_2)] = \exp[i\rho(x_1\cos\theta+x_2\sin\theta)]$$

with $\rho=\sqrt{\omega_1^2+\omega_2^2}$. It is a plane wave that propagates in the direction of θ and oscillates at the frequency ρ. The properties of a two-dimensional Fourier transform are essentially the same as in one dimension. We summarize a few important results.

- If $f \in \mathbf{L}^1(\mathbb{R}^2)$ and $\hat{f} \in \mathbf{L}^1(\mathbb{R}^2)$ then

$$f(x_1,x_2) = \frac{1}{4\pi^2}\int\int \hat{f}(\omega_1,\omega_2)\exp[i(\omega_1 x_1 + \omega_2 x_2)]\,d\omega_1\,d\omega_2. \tag{2.72}$$

- If $f \in \mathbf{L}^1(\mathbb{R}^2)$ and $h \in \mathbf{L}^1(\mathbb{R}^2)$ then the convolution

$$g(x_1,x_2) = f \star h(x_1,x_2) = \int\int f(u_1,u_2)\,h(x_1-u_1,x_2-u_2)\,du_1\,du_2$$

has a Fourier transform

$$\hat{g}(\omega_1,\omega_2) = \hat{f}(\omega_1,\omega_2)\,\hat{h}(\omega_1,\omega_2). \tag{2.73}$$

- The Parseval formula proves that

$$\int\int f(x_1,x_2)\,g^*(x_1,x_2)\,dx_1\,dx_2 = \tag{2.74}$$
$$\frac{1}{4\pi^2}\int\int \hat{f}(\omega_1,\omega_2)\,\hat{g}^*(\omega_1,\omega_2)\,d\omega_1\,d\omega_2 \quad .$$

If $f = g$, we obtain the Plancherel equality

$$\int\int |f(x_1,x_2)|^2\,dx_1\,dx_2 = \frac{1}{4\pi^2}\int\int |\hat{f}(\omega_1,\omega_2)|^2\,d\omega_1\,d\omega_2. \tag{2.75}$$

The Fourier transform of a finite energy function thus has finite energy. With the same density based argument as in one dimension, energy equivalence makes it possible to extend the Fourier transform to any function $f \in \mathbf{L}^2(\mathbb{R}^2)$.

- If $f \in \mathbf{L}^2(\mathbb{R}^2)$ is separable, which means that

$$f(x_1,x_2) = g(x_1)\,h(x_2),$$

then its Fourier transform is

$$\hat{f}(\omega_1,\omega_2) = \hat{g}(\omega_1)\,\hat{h}(\omega_2),$$

where $\hat{h}$ and $\hat{g}$ are the one-dimensional Fourier transforms of g and h. For example, the indicator function

$$f(x_1,x_2) = \begin{cases} 1 & \text{if } |x_1| \le T\ ,\ |x_2| \le T \\ 0 & \text{otherwise} \end{cases} = \mathbf{1}_{[-T,T]}(x_1) \times \mathbf{1}_{[-T,T]}(x_2)$$

is a separable function whose Fourier transform is derived from (2.28):

$$\hat{f}(\omega_1,\omega_2) = \frac{4\sin(T\omega_1)\,\sin(T\omega_2)}{\omega_1\,\omega_2}.$$

- If $f(x_1,x_2)$ is rotated by θ:

$$f_\theta(x_1,x_2) = f(x_1\cos\theta - x_2\sin\theta, x_1\sin\theta + x_2\cos\theta),$$

then its Fourier transform is rotated by $-\theta$:

$$\hat{f}_\theta(\omega_1,\omega_2) = \hat{f}(\omega_1\cos\theta + \omega_2\sin\theta, -\omega_1\sin\theta + \omega_2\cos\theta). \tag{2.76}$$

2.5 PROBLEMS

2.1. [1] Prove that if $f \in \mathbf{L}^1(\mathbb{R})$ then $\hat{f}(\omega)$ is a continuous function of ω, and if $\hat{f} \in \mathbf{L}^1(\mathbb{R})$ then $f(t)$ is continuous.

2.2. [1] Prove the translation (2.18), scaling (2.20) and time derivative (2.21) properties of the Fourier transform.

2.3. [1] Let $f_r(t) = \text{Real}[f(t)]$ and $f_i(t) = \text{Ima}[f(t)]$ be the real and imaginary parts of $f(t)$. Prove that $\hat{f}_r(\omega) = [\hat{f}(\omega) + \hat{f}^*(-\omega)]/2$ and $\hat{f}_i(\omega) = [\hat{f}(\omega) - \hat{f}^*(-\omega)]/(2i)$.

2.4. [1] By using the Fourier transform, verify that

$$\int_{-\infty}^{+\infty} \frac{\sin^3 t}{t^3}\,dt = \frac{3\pi}{4} \quad \text{and} \quad \int_{-\infty}^{+\infty} \frac{\sin^4 t}{t^4}\,dt = \frac{2\pi}{3}.$$

2.5. [1] Show that the Fourier transform of $f(t) = \exp(-(a-ib)t^2)$ is

$$\hat{f}(\omega) = \sqrt{\frac{\pi}{a-ib}}\,\exp\left(-\frac{a+ib}{4(a^2+b^2)}\omega^2\right).$$

Hint: write a differential equation similar to (2.33).

2.6. [2] *Riemann-Lebesgue* Prove that if $f \in \mathbf{L}^1(\mathbb{R})$ then $\lim_{\omega\to\infty} \hat{f}(\omega) = 0$.

Hint: Prove it first for $\mathbf{C}^1$ functions with a compact support and use a density argument.

2.7. [1] *Stability of passive circuits*

(a) Let p be a complex number with $\text{Real}[p] < 0$. Compute the Fourier transforms of $f(t) = \exp(pt)\,\mathbf{1}_{[0,+\infty)}(t)$ and of $f(t) = t^n \exp(pt)\,\mathbf{1}_{[0+\infty)}(t)$.

(b) A passive circuit relates the input voltage f to the output voltage g by a differential equation with constant coefficients:

$$\sum_{k=0}^{K} a_k f^{(k)}(t) = \sum_{k=0}^{M} b_k g^{(k)}(t).$$

Prove that this system is stable and causal if and only if the roots of the equation $\sum_{k=0}^{M} b_k z^k = 0$ have a strictly negative real part.

(c) A Butterworth filter satisfies

$$|\hat{h}(\omega)|^2 = \frac{1}{1+(\omega/\omega_0)^{2N}}.$$

For $N = 3$, compute $\hat{h}(\omega)$ and $h(t)$ so that this filter can be implemented by a stable electronic circuit.

2.8. [1] For any $A > 0$, construct f such that the time and frequency spread measured respectively by σ_t and σ_ω in (2.46, 2.45) satisfy $\sigma_t > A$ and $\sigma_\omega > A$.

2.9. [2] Suppose that $f(t) \geq 0$ and that its support is in $[-T, T]$. Verify that $|\hat{f}(\omega)| \leq \hat{f}(0)$. Let ω_c be the half-power point defined by $|\hat{f}(\omega_c)|^2 = |f(0)|^2/2$ and $|f(\omega)|^2 < |f(0)|^2/2$ for $\omega < \omega_c$. Prove that $\omega_c T \geq \pi/2$.

2.10. [1] *Hilbert transform*

(a) Prove that if $\hat{f}(\omega) = 2/(i\omega)$ then $f(t) = \text{sign}(t) = t/|t|$.

(b) Suppose that $f \in \mathbf{L}^1(\mathbb{R})$ is a causal function, i.e., $f(t) = 0$ for $t < 0$. Let $\hat{f}_r(\omega) = \text{Real}[\hat{f}(\omega)]$ and $\hat{f}_i(\omega) = \text{Ima}[\hat{f}(\omega)]$. Prove that $\hat{f}_r = Hf_i$ and $\hat{f}_i = -Hf_r$ where H is the Hilbert transform operator

$$Hg(x) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{g(u)}{x-u}\, du.$$

2.11. [1] *Rectification* A rectifier computes $g(t) = |f(t)|$, for recovering the envelope of modulated signals [57].

(a) Show that if $f(t) = a(t)\sin\omega_0 t$ with $a(t) \geq 0$ then

$$\hat{g}(\omega) = -\frac{2}{\pi} \sum_{n=-\infty}^{+\infty} \frac{\hat{a}(\omega - 2n\omega_0)}{4n^2 - 1}.$$

(b) Suppose that $\hat{a}(\omega) = 0$ for $|\omega| > \omega_0$. Find h such that $a(t) = h \star g(t)$.

2.12. [2] *Amplitude modulation* For $0 \leq n < N$, we suppose that $f_n(t)$ is real and that $\hat{f}_n(\omega) = 0$ for $|\omega| > \omega_0$.

(a) *Double side-bands* An amplitude modulated multiplexed signal is defined by

$$g(t) = \sum_{n=0}^{N} f_n(t)\cos(2n\omega_0 t).$$

Compute $\hat{g}(\omega)$ and verify that the width of its support is $4N\omega_0$. Find a demodulation algorithm that recovers each f_n from g.

(b) *Single side-band* We want to reduce the bandwidth of the multiplexed signal by 2. Find a modulation procedure that transforms each f_n into a real signal g_n such that $\hat{g}_n$ has a support included in $[-(n+1)\omega_0, -n\omega_0] \cup [n\omega_0, (n+1)\omega_0]$, with the possibility of recovering f_n from g_n. Compute the bandwidth of $g = \sum_{n=0}^{N-1} g_n$, and find a demodulation algorithm that recovers each f_n from g.

2.13. [2] Let $f_\xi = f \star h_\xi$ with $\hat{h}_\xi = \mathbf{1}_{[-\xi,\xi]}$. Suppose that f has a bounded variation $\|f\|_V < +\infty$ and that it is continuous in a neighborhood of t_0. Prove that in a neighborhood of t_0, $f_\xi(t)$ converges uniformly to $f(t)$ when ξ goes to $+\infty$.

2.14. [1] *Tomography* Let $g_\theta(t)$ be the integral of $f(x_1, x_2)$ along the line $-x_1\sin\theta + x_2\cos\theta = t$, which has an angle θ and lies at a distance $|t|$ from the origin:

$$g_\theta(t) = \int_{-\infty}^{+\infty} f(-t\sin\theta + \rho\cos\theta, t\cos\theta + \rho\sin\theta)\, d\rho.$$

Prove that $\hat{g}_\theta(\omega) = \hat{f}(-\omega\sin\theta, \omega\cos\theta)$. How can we recover $f(x_1, x_2)$ from the tomographic projections $g_\theta(t)$ for $0 \leq \theta < 2\pi$?

2.15. [1] Let $f(x_1, x_2)$ be an image which has a discontinuity of amplitude A along a straight line having an angle θ in the plane (x_1, x_2). Compute the amplitude of the Gibbs oscillations of $f \star h_\xi(x_1, x_2)$ as a function of ξ, θ and A, for $\hat{h}_\xi(\omega_1, \omega_2) = \mathbf{1}_{[-\xi,\xi]}(\omega_1)\,\mathbf{1}_{[-\xi,\xi]}(\omega_2)$.

III

DISCRETE REVOLUTION

Digital signal processing has taken over. First used in the 1950's at the service of analog signal processing to simulate analog transforms, digital algorithms have invaded most traditional fortresses, including television standards, speech processing, tape recording and all types of information manipulation. Analog computations performed with electronic circuits are faster than digital algorithms implemented with microprocessors, but are less precise and less flexible. Thus analog circuits are often replaced by digital chips once the computational performance of microprocessors is sufficient to operate in real time for a given application.

Whether sound recordings or images, most discrete signals are obtained by sampling an analog signal. Conditions for reconstructing an analog signal from a uniform sampling are studied. Once more, the Fourier transform is unavoidable because the eigenvectors of discrete time-invariant operators are sinusoidal waves. The Fourier transform is discretized for signals of finite size and implemented with a fast computational algorithm.

3.1 SAMPLING ANALOG SIGNALS [1]

The simplest way to discretize an analog signal f is to record its sample values $\{f(nT)\}_{n\in\mathbb{Z}}$ at intervals T. An approximation of $f(t)$ at any $t \in \mathbb{R}$ may be recovered by interpolating these samples. The Whittaker sampling theorem gives a sufficient condition on the support of the Fourier transform $\hat{f}$ to compute $f(t)$ exactly. Aliasing and approximation errors are studied when this condition is not satisfied. More general sampling theorems are studied in Section 3.1.3 from a vector space point of view.

3.1.1 Whittaker Sampling Theorem

A discrete signal may be represented as a sum of Diracs. We associate to any sample $f(nT)$ a Dirac $f(nT)\delta(t-nT)$ located at $t=nT$. A uniform sampling of f thus corresponds to the weighted Dirac sum

$$f_d(t) = \sum_{n=-\infty}^{+\infty} f(nT)\,\delta(t-nT). \tag{3.1}$$

The Fourier transform of $\delta(t-nT)$ is $e^{-inT\omega}$ so the Fourier transform of f_d is a Fourier series:

$$\hat{f}_d(\omega) = \sum_{n=-\infty}^{+\infty} f(nT)\,e^{-inT\omega}. \tag{3.2}$$

To understand how to compute $f(t)$ from the sample values $f(nT)$ and hence f from f_d, we relate their Fourier transforms $\hat{f}$ and $\hat{f}_d$.

Proposition 3.1 *The Fourier transform of the discrete signal obtained by sampling f at intervals T is*

$$\hat{f}_d(\omega) = \frac{1}{T}\sum_{k=-\infty}^{+\infty} \hat{f}\left(\omega - \frac{2k\pi}{T}\right). \tag{3.3}$$

Proof[1]. Since $\delta(t-nT)$ is zero outside $t=nT$,

$$f(nT)\,\delta(t-nT) = f(t)\,\delta(t-nT),$$

so we can rewrite (3.1) as multiplication with a Dirac comb:

$$f_d(t) = f(t)\sum_{n=-\infty}^{+\infty}\delta(t-nT) = f(t)\,c(t). \tag{3.4}$$

Computing the Fourier transform yields

$$\hat{f}_d(\omega) = \frac{1}{2\pi}\,\hat{f}\star\hat{c}(\omega). \tag{3.5}$$

The Poisson formula (2.4) proves that

$$\hat{c}(\omega) = \frac{2\pi}{T}\sum_{k=-\infty}^{+\infty}\delta\left(\omega - \frac{2\pi k}{T}\right). \tag{3.6}$$

Since $\hat{f}\star\delta(\omega-\xi) = \hat{f}(\omega-\xi)$, inserting (3.6) in (3.5) proves (3.3). ■

Proposition 3.1 proves that sampling f at intervals T is equivalent to making its Fourier transform $2\pi/T$ periodic by summing all its translations $\hat{f}(\omega - 2k\pi/T)$. The resulting sampling theorem was first proved by Whittaker [349] in 1935 in a book on interpolation theory. Shannon rediscovered it in 1949 for applications to communication theory [306].

Theorem 3.1 (SHANNON, WHITTAKER) *If the support of $\hat{f}$ is included in $[-\pi/T, \pi/T]$ then*

$$f(t) = \sum_{n=-\infty}^{+\infty} f(nT)\, h_T(t-nT), \tag{3.7}$$

with

$$h_T(t) = \frac{\sin(\pi t/T)}{\pi t/T}. \tag{3.8}$$

Proof[1]. If $n \neq 0$, the support of $\hat{f}(\omega - n\pi/T)$ does not intersect the support of $\hat{f}(\omega)$ because $\hat{f}(\omega) = 0$ for $|\omega| > \pi/T$. So (3.3) implies

$$\hat{f}_d(\omega) = \frac{\hat{f}(\omega)}{T} \text{ if } |\omega| \leq \frac{\pi}{T}. \tag{3.9}$$

The Fourier transform of h_T is $\hat{h}_T = T\,\mathbf{1}_{[-\pi/T,\pi/T]}$. Since the support of $\hat{f}$ is in $[-\pi/T, \pi/T]$ it results from (3.9) that $\hat{f}(\omega) = \hat{h}_T(\omega)\,\hat{f}_d(\omega)$. The inverse Fourier transform of this equality gives

$$\begin{aligned} f(t) = h_T \star f_d(t) &= h_T \star \sum_{n=-\infty}^{+\infty} f(nT)\,\delta(t-nT) \\ &= \sum_{n=-\infty}^{+\infty} f(nT)\, h_T(t-nT). \end{aligned}$$

■

The sampling theorem imposes that the support of $\hat{f}$ is included in $[-\pi/T, \pi/T]$, which guarantees that f has no brutal variations between consecutive samples, and can thus be recovered with a smooth interpolation. Section 3.1.3 shows that one can impose other smoothness conditions to recover f from its samples. Figure 3.1 illustrates the different steps of a sampling and reconstruction from samples, in both the time and Fourier domains.

3.1.2 Aliasing

The sampling interval T is often imposed by computation or storage constraints and the support of $\hat{f}$ is generally not included in $[-\pi/T, \pi/T]$. In this case the interpolation formula (3.7) does not recover f. We analyze the resulting error and a filtering procedure to reduce it.

Proposition 3.1 proves that

$$\hat{f}_d(\omega) = \frac{1}{T} \sum_{k=-\infty}^{+\infty} \hat{f}\left(\omega - \frac{2k\pi}{T}\right). \tag{3.10}$$

Suppose that the support of $\hat{f}$ goes beyond $[-\pi/T, \pi/T]$. In general the support of $\hat{f}(\omega - 2k\pi/T)$ intersects $[-\pi/T, \pi/T]$ for several $k \neq 0$, as shown in Figure

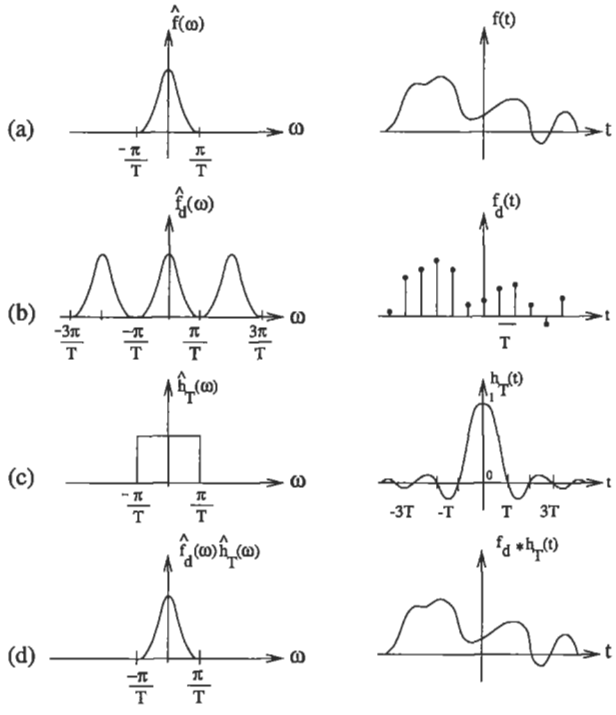


FIGURE 3.1 (a): Signal f and its Fourier transform $\hat{f}$. (b): A uniform sampling of f makes its Fourier transform periodic. (c): Ideal low-pass filter. (d): The filtering of (b) with (c) recovers f.

3.2. This folding of high frequency components over a low frequency interval is called *aliasing*. In the presence of aliasing, the interpolated signal

$$h_T \star f_d(t) = \sum_{n=-\infty}^{+\infty} f(nT)\, h_T(t-nT)$$

has a Fourier transform

$$\hat{f}_d(\omega)\,\hat{h}_T(\omega) = T\,\hat{f}_d(\omega)\,\mathbf{1}_{[-\pi/T,\pi/T]}(\omega) = \mathbf{1}_{[-\pi/T,\pi/T]}(\omega) \sum_{k=-\infty}^{+\infty} \hat{f}\left(\omega - \frac{2k\pi}{T}\right) \tag{3.11}$$

which may be completely different from $\hat{f}(\omega)$ over $[-\pi/T, \pi/T]$. The signal $h_T \star f_d$ may not even be a good approximation of f, as shown by Figure 3.2.

Example 3.1 Let us consider a high frequency oscillation

$$f(t) = \cos(\omega_0 t) = \frac{e^{i\omega_0 t} + e^{-i\omega_0 t}}{2}.$$

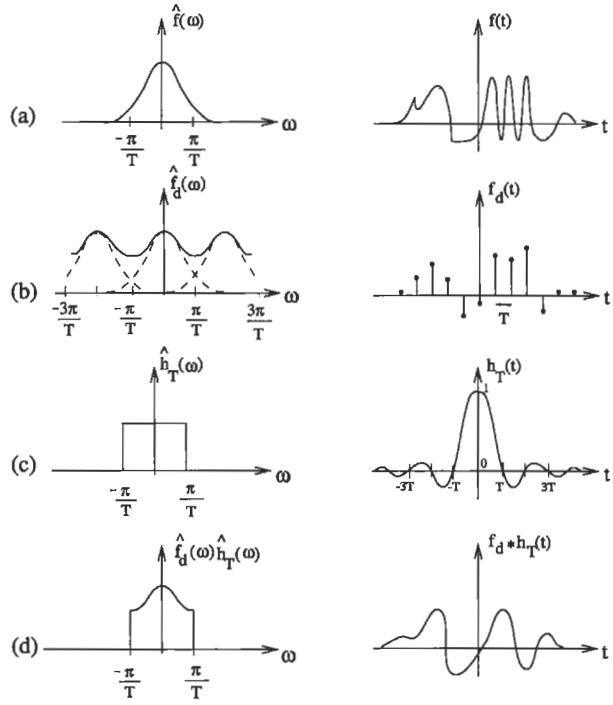


FIGURE 3.2 (a): Signal f and its Fourier transform $\hat{f}$. (b): Aliasing produced by an overlapping of $\hat{f}(\omega - 2k\pi/T)$ for different k, shown in dashed lines. (c): Ideal low-pass filter. (d): The filtering of (b) with (c) creates a low-frequency signal that is different from f.

Its Fourier transform is

$$\hat{f}(\omega) = \pi\Big(\delta(\omega-\omega_0)+\delta(\omega+\omega_0)\Big).$$

If $2\pi/T > \omega_0 > \pi/T$ then (3.11) yields

$$\begin{aligned}
&\hat{f}_d(\omega)\,\hat{h}_T(\omega)\\
&\quad = \pi\,\mathbf{1}_{[-\pi/T,\pi/T]}(\omega)\sum_{k=-\infty}^{+\infty}\left(\delta\Big(\omega-\omega_0-\frac{2k\pi}{T}\Big)+\delta\Big(\omega+\omega_0-\frac{2k\pi}{T}\Big)\right)\\
&\quad = \pi\Big(\delta(\omega-\frac{2\pi}{T}+\omega_0)+\delta(\omega+\frac{2\pi}{T}-\omega_0)\Big),
\end{aligned}$$

so

$$f_d \star h_T(t) = \cos\left[\Big(\frac{2\pi}{T}-\omega_0\Big)t\right].$$

The aliasing reduces the high frequency ω_0 to a lower frequency $2\pi/T - \omega_0 \in [-\pi/T, \pi/T]$. The same frequency folding is observed in a film that samples a fast moving object without enough images per second. A wheel turning rapidly appears as turning much more slowly in the film.

Removal of Aliasing To apply the sampling theorem, f is approximated by the closest signal $\tilde{f}$ whose Fourier transform has a support in $[-\pi/T, \pi/T]$. The Plancherel formula (2.26) proves that

$$\begin{aligned} \|f - \tilde{f}\|^2 &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\hat{f}(\omega) - \widehat{\tilde{f}}(\omega)|^2 \, d\omega \\ &= \frac{1}{2\pi} \int_{|\omega| > \pi/T} |\hat{f}(\omega)|^2 \, d\omega + \frac{1}{2\pi} \int_{|\omega| \leq \pi/T} |\hat{f}(\omega) - \widehat{\tilde{f}}(\omega)|^2 \, d\omega. \end{aligned}$$

This distance is minimum when the second integral is zero and hence

$$\widehat{\tilde{f}}(\omega) = \hat{f}(\omega) \, \mathbf{1}_{[-\pi/T, \pi/T]}(\omega) = \frac{1}{T} \hat{h}_T(\omega) \, \hat{f}(\omega). \tag{3.12}$$

It corresponds to $\tilde{f} = \frac{1}{T} f \star h_T$. The filtering of f by h_T avoids the aliasing by removing any frequency larger than π/T. Since $\widehat{\tilde{f}}$ has a support in $[-\pi/T, \pi/T]$, the sampling theorem proves that $\tilde{f}(t)$ can be recovered from the samples $\tilde{f}(nT)$. An analog to digital converter is therefore composed of a filter that limits the frequency band to $[-\pi/T, \pi/T]$, followed by a uniform sampling at intervals T.

3.1.3 General Sampling Theorems

The sampling theorem gives a sufficient condition for reconstructing a signal from its samples, but other sufficient conditions can be established for different interpolation schemes [335]. To explain this new point of view, the Whittaker sampling theorem is interpreted in more abstract terms, as a signal decomposition in an orthogonal basis.

Proposition 3.2 *If $h_T(t) = \sin(\pi t/T)/(\pi t/T)$ then $\{h_T(t - nT)\}_{n \in \mathbb{Z}}$ is an orthogonal basis of the space $\mathbf{U}_T$ of functions whose Fourier transforms have a support included in $[-\pi/T, \pi/T]$. If $f \in \mathbf{U}_T$ then*

$$f(nT) = \frac{1}{T} \langle f(t), h_T(t - nT) \rangle. \tag{3.13}$$

Proof[2]. Since $\hat{h}_T = T \, \mathbf{1}_{[-\pi/T, \pi/T]}$ the Parseval formula (2.25) proves that

$$\begin{aligned} \langle h_T(t - nT), h_T(t - pT) \rangle &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} T^2 \, \mathbf{1}_{[-\pi/T, \pi/T]}(\omega) \exp[-i(n-p)T\omega] \, d\omega \\ &= \frac{T^2}{2\pi} \int_{-\pi/T}^{\pi/T} \exp[-i(n-p)T\omega] \, d\omega = T \, \delta[n-p]. \end{aligned}$$

The family $\{h_T(t-nT)\}_{n\in\mathbf{Z}}$ is therefore orthogonal. Clearly $h_T(t-nT) \in \mathbf{U}_T$ and (3.7) proves that any $f \in \mathbf{U}_T$ can be decomposed as a linear combination of $\{h_T(t-nT)\}_{n\in\mathbf{Z}}$. It is therefore an orthogonal basis of $\mathbf{U}_T$.

Equation (3.13) is also proved with the Parseval formula

$$\langle f(t), h_T(t-nT)\rangle = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{f}(\omega)\,\hat{h}_T(\omega)\exp(inT\omega)\,d\omega.$$

Since the support of $\hat{f}$ is in $[-\pi/T, \pi/T]$ and $\hat{h}_T = T\,\mathbf{1}_{[-\pi/T,\pi/T]}$,

$$\langle f(t), h_T(t-nT)\rangle = \frac{T}{2\pi}\int_{-\pi/T}^{\pi/T} \hat{f}(\omega)\exp(inT\omega)\,d\omega = T\,f(nT).$$

■

Proposition 3.2 shows that the interpolation formula (3.7) can be interpreted as a decomposition of $f \in \mathbf{U}_T$ in an orthogonal basis of $\mathbf{U}_T$:

$$f(t) = \frac{1}{T}\sum_{n=-\infty}^{+\infty} \langle f(u), h_T(u-nT)\rangle\, h_T(t-nT). \tag{3.14}$$

If $f \notin \mathbf{U}_T$, which means that $\hat{f}$ has a support not included in $[-\pi/T, \pi/T]$, the removal of aliasing is computed by finding the function $\tilde{f} \in \mathbf{U}_T$ that minimizes $\|\tilde{f} - f\|$. Proposition A.2 proves that $\tilde{f}$ is the orthogonal projection $P_{\mathbf{U}_T} f$ of f in $\mathbf{U}_T$.

The Whittaker sampling theorem is generalized by defining other spaces $\mathbf{U}_T$ such that any $f \in \mathbf{U}_T$ can be recovered by interpolating its samples $\{f(nT)\}_{n\in\mathbf{Z}}$. A signal $f \notin \mathbf{U}_T$ is approximated by its orthogonal projection $\tilde{f} = P_{\mathbf{U}_T} f$ in $\mathbf{U}_T$, which is characterized by a uniform sampling $\{\tilde{f}(nT)\}_{n\in\mathbf{Z}}$.

Block Sampler A block sampler approximates signals with piecewise constant functions. The approximation space $\mathbf{U}_T$ is the set of all functions that are constant on intervals $[nT, (n+1)T)$, for any $n \in \mathbb{Z}$. Let $h_T = \mathbf{1}_{[0,T)}$. The family $\{h_T(t-nT)\}_{n\in\mathbf{Z}}$ is clearly an orthogonal basis of $\mathbf{U}_T$. Any $f \in \mathbf{U}_T$ can be written

$$f(t) = \sum_{n=-\infty}^{+\infty} f(nT)\,h_T(t-nT).$$

If $f \notin \mathbf{U}_T$ then (A.17) shows that its orthogonal projection on $\mathbf{U}_T$ is calculated with a partial decomposition in an orthogonal basis of $\mathbf{U}_T$. Since $\|h_T(t-nT)\|^2 = T$,

$$P_{\mathbf{U}_T} f(t) = \frac{1}{T}\sum_{n=-\infty}^{+\infty} \langle f(u), h_T(u-nT)\rangle\, h_T(t-nT). \tag{3.15}$$

Let $\bar{h}_T(t) = h_T(-t)$. Then

$$\langle f(u), h_T(u-nT)\rangle = \int_{nT}^{(n+1)T} f(t)\,dt = f \star \bar{h}_T(nT).$$

This averaging of f over intervals of size T is equivalent to the aliasing removal used for the Whittaker sampling theorem.

Approximation Space The space $\mathbf{U}_T$ should be chosen so that $P_{\mathbf{U}_T} f$ gives an accurate approximation of f, for a given class of signals. The Whittaker interpolation approximates signals by restricting their Fourier transform to a low frequency interval. It is particularly effective for smooth signals whose Fourier transform have an energy concentrated at low frequencies. It is also well adapted to sound recordings, which are well approximated by lower frequency harmonics.

For discontinuous signals such as images, a low-frequency restriction produces the Gibbs oscillations studied in Section 2.3.3. The visual quality of the image is degraded by these oscillations, which have a total variation (2.65) that is infinite. A piecewise constant approximation has the advantage of creating no spurious oscillations, and one can prove that the projection in $\mathbf{U}_T$ decreases the total variation: $\|P_{\mathbf{U}_T} f\|_V \leq \|f\|_V$. In domains where f is a regular function, the piecewise constant approximation $P_{\mathbf{U}_T} f$ may however be significantly improved. More precise approximations are obtained with spaces $\mathbf{U}_T$ of higher order polynomial splines. These approximations can introduce small Gibbs oscillations, but these oscillations have a finite total variation. Section 7.6.1 studies the construction of interpolation bases used to recover signals from their samples, when the signals belong to spaces of polynomial splines and other spaces $\mathbf{U}_T$.

3.2 DISCRETE TIME-INVARIANT FILTERS [1]

3.2.1 Impulse Response and Transfer Function

Classical discrete signal processing algorithms are mostly based on time-invariant linear operators [55, 58]. The time-invariance is limited to translations on the sampling grid. To simplify notation, the sampling interval is normalized $T = 1$, and we denote $f[n]$ the sample values. A linear discrete operator L is time-invariant if an input $f[n]$ delayed by $p \in \mathbb{Z}$, $f_p[n] = f[n-p]$, produces an output also delayed by p:

$$Lf_p[n] = Lf[n-p].$$

Impulse Response We denote by $\delta[n]$ the discrete Dirac

$$\delta[n] = \begin{cases} 1 & \text{if } n = 0 \\ 0 & \text{if } n \neq 0 \end{cases}. \tag{3.16}$$

Any signal $f[n]$ can be decomposed as a sum of shifted Diracs

$$f[n] = \sum_{p=-\infty}^{+\infty} f[p]\, \delta[n-p].$$

Let $L\delta[n] = h[n]$ be the discrete *impulse response*. The linearity and time-invariance implies that

$$Lf[n] = \sum_{p=-\infty}^{+\infty} f[p]\, h[n-p] = f \star h[n]. \tag{3.17}$$

A discrete linear time-invariant operator is thus computed with a discrete convolution. If $h[n]$ has a finite support the sum (3.17) is calculated with a finite number of operations. These are called *Finite Impulse Response* (FIR) filters. Convolutions with infinite impulse response filters may also be calculated with a finite number of operations if they can be rewritten with a recursive equation (3.29).

Causality and Stability A discrete filter L is *causal* if $Lf[p]$ depends only on the values of $f[n]$ for $n \leq p$. The convolution formula (3.17) implies that $h[n] = 0$ if $n < 0$.

The filter is *stable* if any bounded input signal $f[n]$ produces a bounded output signal $Lf[n]$. Since

$$|Lf[n]| \leq \sup_{n \in \mathbb{Z}} |f[n]| \sum_{k=-\infty}^{+\infty} |h[k]|,$$

it is sufficient that $\sum_{n=-\infty}^{+\infty} |h[n]| < +\infty$, which means that $h \in \mathbf{l}^1(\mathbb{Z})$. One can verify that this sufficient condition is also necessary. The impulse response h is thus stable if $h \in \mathbf{l}^1(\mathbb{Z})$.

Transfer Function The Fourier transform plays a fundamental role in analyzing discrete time-invariant operators, because the discrete sinusoidal waves $e_\omega[n] = \mathrm{e}^{i\omega n}$ are eigenvectors:

$$Le_\omega[n] = \sum_{p=-\infty}^{+\infty} \mathrm{e}^{i\omega(n-p)}\, h[p] = \mathrm{e}^{i\omega n} \sum_{p=-\infty}^{+\infty} h[p]\, \mathrm{e}^{-i\omega p}. \tag{3.18}$$

The eigenvalue is a Fourier series

$$\hat{h}(\omega) = \sum_{p=-\infty}^{+\infty} h[p]\, \mathrm{e}^{-i\omega p}. \tag{3.19}$$

It is the filter *transfer function*.

Example 3.2 The uniform discrete average

$$Lf[n] = \frac{1}{2N+1} \sum_{p=n-N}^{n+N} f[p]$$

is a time-invariant discrete filter whose impulse response is $h = (2N+1)^{-1}\mathbf{1}_{[-N,N]}$. Its transfer function is

$$\hat{h}(\omega) = \frac{1}{2N+1}\sum_{n=-N}^{+N} e^{-in\omega} = \frac{1}{2N+1}\frac{\sin(N+1/2)\omega}{\sin\omega/2}. \tag{3.20}$$

3.2.2 Fourier Series

The properties of Fourier series are essentially the same as the properties of the Fourier transform since Fourier series are particular instances of Fourier transforms for Dirac sums. If $f(t) = \sum_{n=-\infty}^{+\infty} f[n]\,\delta(t-n)$ then $\hat{f}(\omega) = \sum_{n=-\infty}^{+\infty} f[n]\,e^{-i\omega n}$.

For any $n \in \mathbb{Z}$, $e^{-i\omega n}$ has period 2π, so Fourier series have period 2π. An important issue is to understand whether all functions with period 2π can be written as Fourier series. Such functions are characterized by their restriction to $[-\pi,\pi]$. We therefore consider functions $\hat{a} \in \mathbf{L}^2[-\pi,\pi]$ that are square integrable over $[-\pi,\pi]$. The space $\mathbf{L}^2[-\pi,\pi]$ is a Hilbert space with the inner product

$$\langle \hat{a}, \hat{b} \rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{a}(\omega)\,\hat{b}^*(\omega)\,d\omega \tag{3.21}$$

and the resulting norm

$$\|\hat{a}\|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |\hat{a}(\omega)|^2\,d\omega.$$

The following theorem proves that any function in $\mathbf{L}^2[-\pi,\pi]$ can be written as a Fourier series.

Theorem 3.2 *The family of functions* $\{e^{-ik\omega}\}_{k\in\mathbb{Z}}$ *is an orthonormal basis of* $\mathbf{L}^2[-\pi,\pi]$.

Proof[2]. The orthogonality with respect to the inner product (3.21) is established with a direct integration. To prove that $\{\exp(-ik\omega)\}_{k\in\mathbb{Z}}$ is a basis, we must show that linear expansions of these vectors are dense in $\mathbf{L}^2[-\pi,\pi]$.

We first prove that any continuously differentiable function $\hat{\phi}$ with a support included in $[-\pi,\pi]$ satisfies

$$\hat{\phi}(\omega) = \sum_{k=-\infty}^{+\infty} \langle \hat{\phi}(\xi), \exp(-ik\xi)\rangle \exp(-ik\omega)\,, \tag{3.22}$$

with a pointwise convergence for any $\omega \in [-\pi,\pi]$. Let us compute the partial sum

$$\begin{aligned} S_N(\omega) &= \sum_{k=-N}^{N} \langle \hat{\phi}(\xi), \exp(-ik\xi)\rangle \exp(-ik\omega) \\ &= \sum_{k=-N}^{N} \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{\phi}(\xi)\exp(ik\xi)\,d\xi\,\exp(-ik\omega) \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{\phi}(\xi)\sum_{k=-N}^{N}\exp[ik(\xi-\omega)]\,d\xi. \end{aligned}$$

The Poisson formula (2.37) proves the distribution equality

$$\lim_{N\to+\infty} \sum_{k=-N}^{N} \exp[ik(\xi-\omega)] = 2\pi \sum_{k=-\infty}^{+\infty} \delta(\xi-\omega-2\pi k),$$

and since the support of $\hat{\phi}$ is in $[-\pi,\pi]$ we get

$$\lim_{N\to+\infty} S_N(\omega) = \hat{\phi}(\omega).$$

Since $\hat{\phi}$ is continuously differentiable, following the steps (2.38-2.40) in the proof of the Poisson formula shows that $S_N(\omega)$ converges uniformly to $\hat{\phi}(\omega)$ on $[-\pi,\pi]$.

To prove that linear expansions of sinusoidal waves $\{\exp(-ik\omega)\}_{k\in\mathbf{Z}}$ are dense in $\mathbf{L}^2[-\pi,\pi]$, let us verify that the distance between $\hat{a}\in\mathbf{L}^2[-\pi,\pi]$ and such a linear expansion is less than ϵ, for any $\epsilon>0$. Continuously differentiable functions with a support included in $[-\pi,\pi]$ are dense in $\mathbf{L}^2[-\pi,\pi]$, hence there exists $\hat{\phi}$ such that $\|\hat{a}-\hat{\phi}\|\leq\epsilon/2$. The uniform pointwise convergence proves that there exists N for which

$$\sup_{\omega\in[-\pi,\pi]} |S_N(\omega)-\hat{\phi}(\omega)| \leq \frac{\epsilon}{2},$$

which implies that

$$\|S_N-\hat{\phi}\|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |S_N(\omega)-\hat{\phi}(\omega)|^2\,d\omega \leq \frac{\epsilon^2}{4}.$$

It follows that $\hat{a}$ is approximated by the Fourier series S_N with an error

$$\|\hat{a}-S_N\| \leq \|\hat{a}-\hat{\phi}\| + \|\hat{\phi}-S_N\| \leq \epsilon.$$

■

Theorem 3.2 proves that if $f\in\mathbf{l}^2(\mathbb{Z})$, the Fourier series

$$\hat{f}(\omega) = \sum_{n=-\infty}^{+\infty} f[n]\,e^{-i\omega n} \tag{3.23}$$

can be interpreted as the decomposition of $\hat{f}\in\mathbf{L}^2[-\pi,\pi]$ in an orthonormal basis. The Fourier series coefficients can thus be written as inner products

$$f[n] = \langle \hat{f}(\omega), e^{-i\omega n}\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{f}(\omega)\,e^{i\omega n}\,d\omega. \tag{3.24}$$

The energy conservation of orthonormal bases (A.10) yields a Plancherel identity:

$$\sum_{n=-\infty}^{+\infty} |f[n]|^2 = \|\hat{f}\|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |\hat{f}(\omega)|^2\,d\omega. \tag{3.25}$$

Pointwise Convergence The equality (3.23) is meant in the sense of mean-square convergence

$$\lim_{N\to+\infty} \left\| \hat{f}(\omega) - \sum_{k=-N}^{N} f[k]\, e^{-i\omega k} \right\| = 0.$$

It does not imply a pointwise convergence at all $\omega \in \mathbb{R}$. In 1873, Dubois-Reymond constructed a periodic function $\hat{f}(\omega)$ that is continuous and whose Fourier series diverges at some points. On the other hand, if $\hat{f}(\omega)$ is continuously differentiable, then the proof of Theorem 3.2 shows that its Fourier series converges uniformly to $\hat{f}(\omega)$ on $[-\pi, \pi]$. It was only in 1966 that Carleson [114] was able to prove that if $\hat{f} \in \mathbf{L}^2[-\pi, \pi]$ then its Fourier series converges almost everywhere. The proof is however extremely technical.

Convolutions Since $\{e^{-i\omega k}\}_{k\in\mathbb{Z}}$ are eigenvectors of discrete convolution operators, we also have a discrete convolution theorem.

Theorem 3.3 *If $f \in \mathbf{l}^1(\mathbb{Z})$ and $h \in \mathbf{l}^1(\mathbb{Z})$ then $g = f \star h \in \mathbf{l}^1(\mathbb{Z})$ and*

$$\hat{g}(\omega) = \hat{f}(\omega)\hat{h}(\omega). \tag{3.26}$$

The proof is identical to the proof of the convolution Theorem 2.2, if we replace integrals by discrete sums. The reconstruction formula (3.24) shows that a filtered signal can be written

$$f \star h[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{h}(\omega)\hat{f}(\omega)\, e^{i\omega n}\, d\omega. \tag{3.27}$$

The transfer function $\hat{h}(\omega)$ amplifies or attenuates the frequency components $\hat{f}(\omega)$ of $f[n]$.

Example 3.3 An ideal *discrete low-pass filter* has a 2π periodic transfer function defined by $\hat{h}(\omega) = \mathbf{1}_{[-\xi,\xi]}(\omega)$, for $\omega \in [-\pi, \pi]$ and $0 < \xi < \pi$. Its impulse response is computed with (3.24):

$$h[n] = \frac{1}{2\pi} \int_{-\xi}^{\xi} e^{i\omega n}\, d\omega = \frac{\sin \xi n}{\pi n}. \tag{3.28}$$

It is a uniform sampling of the ideal analog low-pass filter (2.29).

Example 3.4 A *recursive filter* computes $g = Lf$ which is solution of a recursive equation

$$\sum_{k=0}^{K} a_k\, f[n-k] = \sum_{k=0}^{M} b_k\, g[n-k], \tag{3.29}$$

with $b_0 \neq 0$. If $g[n] = 0$ and $f[n] = 0$ for $n < 0$ then g has a linear and time-invariant dependency upon f, and can thus be written $g = f \star h$. The transfer function is

obtained by computing the Fourier transform of (3.29). The Fourier transform of $f_k[n] = f[n-k]$ is $\hat{f}_k(\omega) = \hat{f}(\omega)\mathrm{e}^{-ik\omega}$ so

$$\hat{h}(\omega) = \frac{\hat{g}(\omega)}{\hat{f}(\omega)} = \frac{\sum_{k=0}^{K} a_k \mathrm{e}^{-ik\omega}}{\sum_{k=0}^{M} b_k \mathrm{e}^{-ik\omega}}.$$

It is a rational function of $\mathrm{e}^{-i\omega}$. If $b_k \neq 0$ for some $k > 0$ then one can verify that the impulse response h has an infinite support. The stability of such filters is studied in Problem 3.8. A direct calculation of the convolution sum $g[n] = f \star h[n]$ would require an infinite number of operation whereas (3.29) computes $g[n]$ with $K+M$ additions and multiplications from its past values.

Window Multiplication An infinite impulse response filter h such as the ideal low-pass filter (3.28) may be approximated by a finite response filter $\tilde{h}$ by multiplying h with a window g of finite support:

$$\tilde{h}[n] = g[n]\, h[n].$$

One can verify that a multiplication in time is equivalent to a convolution in the frequency domain:

$$\widehat{\tilde{h}}(\omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{h}(\xi)\, \hat{g}(\omega - \xi)\, d\xi = \frac{1}{2\pi} \hat{h} \star \hat{g}(\omega). \tag{3.30}$$

Clearly $\widehat{\tilde{h}} = \hat{h}$ only if $\hat{g} = 2\pi\delta$, which would imply that g has an infinite support and $g[n] = 1$. The approximation $\widehat{\tilde{h}}$ is close to $\hat{h}$ only if $\hat{g}$ approximates a Dirac, which means that all its energy is concentrated at low frequencies. In time, g should therefore have smooth variations.

The rectangular window $g = \mathbf{1}_{[-N,N]}$ has a Fourier transform $\hat{g}$ computed in (3.20). It has important side lobes far away from $\omega = 0$. The resulting $\widehat{\tilde{h}}$ is a poor approximation of $\hat{h}$. The Hanning window

$$g[n] = \cos^2\left(\frac{\pi n}{2N}\right) \mathbf{1}_{[-N,N]}[n]$$

is smoother and thus has a Fourier transform better concentrated at low frequencies. The spectral properties of other windows are studied in Section 4.2.2.

3.3 FINITE SIGNALS [1]

Up to now, we have considered discrete signals $f[n]$ defined for all $n \in \mathbb{Z}$. In practice, $f[n]$ is known over a finite domain, say $0 \leq n < N$. Convolutions must therefore be modified to take into account the border effects at $n = 0$ and $n = N-1$. The Fourier transform must also be redefined over finite sequences for numerical computations. The fast Fourier transform algorithm is explained as well as its application to fast convolutions.

3.3.1 Circular Convolutions

Let $\tilde{f}$ and $\tilde{h}$ be signals of N samples. To compute the convolution product

$$\tilde{f} \star \tilde{h}[n] = \sum_{p=-\infty}^{+\infty} \tilde{f}[p]\,\tilde{h}[n-p] \quad \text{for } 0 \leq n < N,$$

we must know $\tilde{f}[n]$ and $\tilde{h}[n]$ beyond $0 \leq n < N$. One approach is to extend $\tilde{f}$ and $\tilde{h}$ with a periodization over N samples, and define

$$f[n] = \tilde{f}[n \bmod N] \ , \ h[n] = \tilde{h}[n \bmod N].$$

The *circular convolution* of two such signals, both with period N, is defined as a sum over their period:

$$f \circledast h[n] = \sum_{p=0}^{N-1} f[p]\,h[n-p] = \sum_{p=0}^{N-1} f[n-p]\,h[p].$$

It is also a signal of period N.

The eigenvectors of a circular convolution operator

$$Lf[n] = f \circledast h[n]$$

are the discrete complex exponentials $e_k[n] = \exp(i2\pi kn/N)$. Indeed

$$Le_k[n] = \exp\left(\frac{i2\pi kn}{N}\right) \sum_{p=0}^{N-1} h[p] \exp\left(\frac{-i2\pi kp}{N}\right),$$

and the eigenvalue is the discrete Fourier transform of h:

$$\hat{h}[k] = \sum_{p=0}^{N-1} h[p] \exp\left(\frac{-i2\pi kp}{N}\right).$$

3.3.2 Discrete Fourier Transform

The space of signals of period N is an Euclidean space of dimension N and the inner product of two such signals f and g is

$$\langle f, g \rangle = \sum_{n=0}^{N-1} f[n]\,g^*[n]. \tag{3.31}$$

The following theorem proves that any signal with period N can be decomposed as a sum of discrete sinusoidal waves.

Theorem 3.4 *The family*

$$\left\{ e_k[n] = \exp\left(\frac{i2\pi kn}{N}\right) \right\}_{0 \le k < N}$$

is an orthogonal basis of the space of signals of period N.

Since the space is of dimension N, any orthogonal family of N vectors is an orthogonal basis. To prove this theorem it is therefore sufficient to verify that $\{e_k\}_{0 \le k < N}$ is orthogonal with respect to the inner product (3.31). Any signal f of period N can be decomposed in this basis:

$$f = \sum_{k=0}^{N-1} \frac{\langle f, e_k \rangle}{\|e_k\|^2} e_k. \tag{3.32}$$

By definition, the *discrete Fourier transform* (DFT) of f is

$$\hat{f}[k] = \langle f, e_k \rangle = \sum_{n=0}^{N-1} f[n] \exp\left(\frac{-i2\pi kn}{N}\right). \tag{3.33}$$

Since $\|e_k\|^2 = N$, (3.32) gives an inverse discrete Fourier formula:

$$f[n] = \frac{1}{N} \sum_{k=0}^{N-1} \hat{f}[k] \exp\left(\frac{i2\pi kn}{N}\right). \tag{3.34}$$

The orthogonality of the basis also implies a Plancherel formula

$$\|f\|^2 = \sum_{n=0}^{N-1} |f[n]|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |\hat{f}[k]|^2. \tag{3.35}$$

The discrete Fourier transform of a signal f of period N is computed from its values for $0 \le n < N$. Then why is it important to consider it a periodic signal with period N rather than a finite signal of N samples? The answer lies in the interpretation of the Fourier coefficients. The discrete Fourier sum (3.34) defines a signal of period N for which the samples $f[0]$ and $f[N-1]$ are side by side. If $f[0]$ and $f[N-1]$ are very different, this produces a brutal transition in the periodic signal, creating relatively high amplitude Fourier coefficients at high frequencies. For example, Figure 3.3 shows that the "smooth" ramp $f[n] = n$ for $0 \le n < N$ has sharp transitions at $n = 0$ and $n = N$ once made periodic.

Circular Convolutions Since $\{\exp(i2\pi kn/N)\}_{0 \le k < N}$ are eigenvectors of circular convolutions, we derive a convolution theorem.

Theorem 3.5 *If f and h have period N then the discrete Fourier transform of $g = f \circledast h$ is*

$$\hat{g}[k] = \hat{f}[k]\,\hat{h}[k]. \tag{3.36}$$

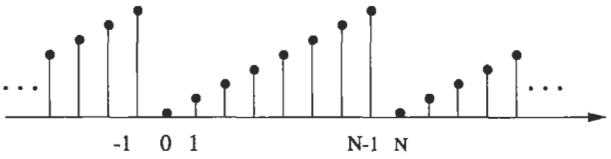


FIGURE 3.3 Signal $f[n] = n$ for $0 \leq n < N$ made periodic over N samples.

The proof is similar to the proof of the two previous convolution Theorems 2.2 and 3.3. This theorem shows that a circular convolution can be interpreted as a discrete frequency filtering. It also opens the door to fast computations of convolutions using the fast Fourier transform.

3.3.3 Fast Fourier Transform

For a signal f of N points, a direct calculation of the N discrete Fourier sums

$$\hat{f}[k] = \sum_{n=0}^{N-1} f[n] \exp\left(\frac{-i2\pi kn}{N}\right) \quad , \text{ for } 0 \leq k < N, \tag{3.37}$$

requires N^2 complex multiplications and additions. The fast Fourier transform (FFT) algorithm reduces the numerical complexity to $O(N \log_2 N)$ by reorganizing the calculations.

When the frequency index is even, we group the terms n and $n+N/2$:

$$\hat{f}[2k] = \sum_{n=0}^{N/2-1} \Big(f[n] + f[n+N/2]\Big) \exp\left(\frac{-i2\pi kn}{N/2}\right). \tag{3.38}$$

When the frequency index is odd, the same grouping becomes

$$\hat{f}[2k+1] = \sum_{n=0}^{N/2-1} \exp\left(\frac{-i2\pi n}{N}\right) \Big(f[n] - f[n+N/2]\Big) \exp\left(\frac{-i2\pi kn}{N/2}\right). \tag{3.39}$$

Equation (3.38) proves that even frequencies are obtained by calculating the discrete Fourier transform of the $N/2$ periodic signal

$$f_e[n] = f[n] + f[n+N/2].$$

Odd frequencies are derived from (3.39) by computing the Fourier transform of the $N/2$ periodic signal

$$f_o[n] = \exp\left(\frac{-i2\pi n}{N}\right) \Big(f[n] - f[n+N/2]\Big).$$

A discrete Fourier transform of size N may thus be calculated with two discrete Fourier transforms of size $N/2$ plus $O(N)$ operations.

The inverse fast Fourier transform of $\hat{f}$ is derived from the forward fast Fourier transform of its complex conjugate $\hat{f}^*$ by observing that

$$f^*[n] = \frac{1}{N} \sum_{k=0}^{N-1} \hat{f}^*[k] \exp\left(\frac{-i2\pi kn}{N}\right). \tag{3.40}$$

Complexity Let $C(N)$ be the number of elementary operations needed to compute a discrete Fourier transform with the FFT. Since f is complex, the calculation of f_e and f_o requires N complex additions and $N/2$ complex multiplications. Let KN be the corresponding number of elementary operations. We have

$$C(N) = 2C(N/2) + KN. \tag{3.41}$$

Since the Fourier transform of a single point is equal to itself, $C(1) = 0$. With the change of variable $l = \log_2 N$ and the change of function $T(l) = \frac{C(N)}{N}$, we derive from (3.41) that

$$T(l) = T(l-1) + K.$$

Since $T(0) = 0$ we get $T(l) = Kl$ and hence

$$C(N) = KN \log_2(N).$$

There exist several variations of this fast algorithm [177, 51]. The goal is to minimize the constant K. The most efficient fast discrete Fourier transform to this date is the split-radix FFT algorithm, which is slightly more complicated than the procedure just described, but which requires only $N \log_2 N$ real multiplications and $3N \log_2 N$ additions. When the input signal f is real, there are half as many parameters to compute, since $\hat{f}[-k] = \hat{f}^*[k]$. The number of multiplications and additions is thus reduced by 2.

3.3.4 Fast Convolutions

The low computational complexity of a fast Fourier transform makes it efficient to compute finite discrete convolutions by using the circular convolution Theorem 3.5. Let f and h be two signals whose samples are non-zero only for $0 \leq n < M$. The causal signal

$$g[n] = f \star h[n] = \sum_{k=-\infty}^{+\infty} f[k]\, h[n-k] \tag{3.42}$$

is non-zero only for $0 \leq n < 2M$. If h and f have M non-zero samples, calculating this convolution product with the sum (3.42) requires $M(M+1)$ multiplications and additions. When $M \geq 32$, the number of computations is reduced by using the fast Fourier transform [11, 51].

To use the fast Fourier transform with the circular convolution Theorem 3.5, the non-circular convolution (3.42) is written as a circular convolution. We define

two signals of period $2M$:

$$a[n] = \begin{cases} f[n] & \text{if } 0 \leq n < M \\ 0 & \text{if } M \leq n < 2M \end{cases} \tag{3.43}$$

$$b[n] = \begin{cases} h[n] & \text{if } 0 \leq n < M \\ 0 & \text{if } M \leq n < 2M \end{cases} . \tag{3.44}$$

Let $c = a \circledast b$, one can verify that $c[n] = g[n]$ for $0 \leq n < 2M$. The $2M$ non-zero coefficients of g are thus obtained by computing $\hat{a}$ and $\hat{b}$ from a and b and then calculating the inverse discrete Fourier transform of $\hat{c} = \hat{a}\hat{b}$. With the fast Fourier transform algorithm, this requires a total of $O(M \log_2 M)$ additions and multiplications instead of $M(M+1)$. A single FFT or inverse FFT of a real signal of size N is calculated with $2^{-1}N \log_2 N$ multiplications, using a split-radix algorithm. The FFT convolution is thus performed with a total of $3M \log_2 M + 11M$ real multiplications. For $M \geq 32$ the FFT algorithm is faster than the direct convolution approach. For $M \leq 16$, it is faster to use a direct convolution sum.

Fast Overlap-Add Convolutions The convolution of a signal f of L non-zero samples with a smaller causal signal h of M samples is calculated with an overlap-add procedure that is faster than the previous algorithm. The signal f is decomposed with a sum of L/M blocks f_r having M non-zero samples:

$$f[n] = \sum_{r=0}^{L/M-1} f_r[n - rM] \text{ with } f_r[n] = f[n + rM]\, 1_{[0,M-1]}[n]. \tag{3.45}$$

For each $0 \leq r < L/M$, the $2M$ non-zero samples of $g_r = f_r \star h$ are computed with the FFT based convolution algorithm, which requires $O(M \log_2 M)$ operations. These L/M convolutions are thus obtained with $O(L \log_2 M)$ operations. The block decomposition (3.45) implies that

$$f \star h[n] = \sum_{r=0}^{L/M-1} g_r[n - rM]. \tag{3.46}$$

The addition of these L/M translated signals of size $2M$ is done with $2L$ additions. The overall convolution is thus performed with $O(L \log_2 M)$ operations.

3.4 DISCRETE IMAGE PROCESSING [1]

Two-dimensional signal processing poses many specific geometrical and topological problems that do not exist in one dimension [23, 34]. For example, a simple concept such as causality is not well defined in two dimensions. We avoid the complexity introduced by the second dimension by extending one-dimensional algorithms with a separable approach. This not only simplifies the mathematics but also leads to faster numerical algorithms along the rows and columns of images. Appendix A.5 reviews the properties of tensor products for separable calculations.

3.4.1 Two-Dimensional Sampling Theorem

The light intensity measured by a camera is generally sampled over a rectangular array of picture elements, called *pixels*. The one-dimensional sampling theorem is extended to this two-dimensional sampling array. Other two-dimensional sampling grids such as hexagonal grids are also possible, but non-rectangular sampling arrays are hardly ever used. We avoid studying them following our separable extension principle.

Let T_1 and T_2 be the sampling intervals along the x_1 and x_2 axes of an infinite rectangular sampling grid. A discrete image obtained by sampling $f(x_1,x_2)$ can be represented as a sum of Diracs located at the grid points:

$$f_d(x_1,x_2) = \sum_{n_1,n_2=-\infty}^{+\infty} f(n_1T_1,n_2T_2)\,\delta(x_1-n_1T_1)\,\delta(x_2-n_2T_2).$$

The two-dimensional Fourier transform of

$$\delta(x_1-n_1T_1)\,\delta(x_2-n_2T_2) \quad \text{is} \quad \exp[-i(n_1T_1\omega_1+n_2T_2\omega_2)].$$

The Fourier transform of f_d is thus a two-dimensional Fourier series

$$\hat{f}_d(\omega_1,\omega_2) = \sum_{n_1,n_2=-\infty}^{+\infty} f(n_1T_1,n_2T_2)\exp[-i(n_1T_1\omega_1+n_2T_2\omega_2)]. \tag{3.47}$$

It has period $2\pi/T_1$ along ω_1 and period $2\pi/T_2$ along ω_2. An extension of Proposition 3.1 relates $\hat{f}_d$ to the two-dimensional Fourier transform $\hat{f}$ of f.

Proposition 3.3 *The Fourier transform of the discrete image obtained by sampling f at intervals T_1 and T_2 along x_1 and x_2 is*

$$\hat{f}_d(\omega_1,\omega_2) = \frac{1}{T_1T_2}\sum_{k_1,k_2=-\infty}^{+\infty} \hat{f}\left(\omega_1-\frac{2k_1\pi}{T_1},\omega_2-\frac{2k_2\pi}{T_2}\right). \tag{3.48}$$

We derive the following two-dimensional sampling theorem, which is analogous to Theorem 3.1.

Theorem 3.6 *If $\hat{f}$ has a support included in $[-\pi/T_1,\pi/T_1]\times[-\pi/T_2,\pi/T_2]$ then*

$$f(x_1,x_2) = \sum_{n_1,n_2=-\infty}^{+\infty} f(n_1T_1,n_2T_2)\,h_{T_1}(x_1-n_1T_1)\,h_{T_2}(x_2-n_2T_2), \tag{3.49}$$

where

$$h_T(t) = \frac{\sin(\pi t/T)}{\pi t/T}. \tag{3.50}$$

Aliasing If the support of $\hat{f}$ is not included in the low-frequency rectangle $[-\pi/T_1, \pi/T_1] \times [-\pi/T_2, \pi/T_2]$, the interpolation formula (3.49) introduces aliasing errors. This aliasing is eliminated by prefiltering f with the ideal low-pass separable filter $h_{T_1}(x_1)h_{T_2}(x_2)/(T_1T_2)$ whose Fourier transform is the indicator function of $[-\pi/T_1, \pi/T_1] \times [-\pi/T_2, \pi/T_2]$.

3.4.2 Discrete Image Filtering

The properties of two-dimensional space-invariant operators are essentially the same as in one dimension. The sampling intervals T_1 and T_2 are normalized to 1. A pixel value located at (n_1, n_2) is written $f[n_1, n_2]$. A linear operator L is space-invariant if for any $f_{p_1,p_2}[n_1, n_2] = f[n_1 - p_1, n_2 - p_2]$, with $(p_1, p_2) \in \mathbb{Z}^2$,

$$Lf_{p_1,p_2}[n_1, n_2] = Lf[n_1 - p_1, n_2 - p_2].$$

Impulse Response Since an image can be decomposed as a sum of discrete Diracs:

$$f[n_1, n_2] = \sum_{p_1,p_2=-\infty}^{+\infty} f[p_1, p_2]\, \delta[n_1 - p_1]\, \delta[n_2 - p_2],$$

the linearity and time invariance implies

$$Lf[n_1, n_2] = \sum_{p_1,p_2=-\infty}^{+\infty} f[p_1, p_2]\, h[n_1 - p_1, n_2 - p_2] = f \star h[n_1, n_2], \tag{3.51}$$

where $h[n_1, n_2]$ is the response of the impulse $\delta_{0,0}[p_1, p_2] = \delta[p_1]\delta[p_2]$:

$$h[n_1, n_2] = L\delta_{0,0}[n_1, n_2].$$

If the impulse response is separable:

$$h[n_1, n_2] = h_1[n_1]\, h_2[n_2], \tag{3.52}$$

the two-dimensional convolution (3.51) is computed as one-dimensional convolutions along the columns of the image followed by one-dimensional convolutions along the rows (or vice-versa):

$$f \star h[n_1, n_2] = \sum_{p_1=-\infty}^{+\infty} h_1[n_1 - p_1] \sum_{p_2=-\infty}^{+\infty} h_2[n_2 - p_2]\, f[p_1, p_2]. \tag{3.53}$$

This factorization reduces the number of operations. For example, a moving average over squares of $(2M+1)^2$ pixels:

$$Lf[n_1, n_2] = \frac{1}{(2M+1)^2} \sum_{p_1=-M}^{M} \sum_{p_2=-M}^{M} f[n_1 - p_1, n_2 - p_2] \tag{3.54}$$

is a separable convolution with $h_1 = h_2 = (2M+1)^{-1}\mathbf{1}_{[-M,M]}$. A direct calculation with (3.54) requires $(2M+1)^2$ additions per pixel whereas the factorization (3.53) performs this calculation with $2(2M+1)$ additions per point.

Transfer Function The Fourier transform of a discrete image f is defined by the Fourier series

$$\hat{f}(\omega_1,\omega_2) = \sum_{n_1=-\infty}^{+\infty} \sum_{n_2=-\infty}^{+\infty} f[n_1,n_2] \exp[-i(\omega_1 n_1 + \omega_2 n_2)]. \tag{3.55}$$

The two-dimensional extension of the convolution Theorem 3.3 proves that if $g = Lf = f \star h$ then its Fourier transform is

$$\hat{g}(\omega_1,\omega_2) = \hat{f}(\omega_1,\omega_2)\,\hat{h}(\omega_1,\omega_2),$$

and $\hat{h}$ is the transfer function of the filter. When a filter is separable $h[n_1,n_2] = h_1[n_1]\,h_2[n_2]$, its transfer function is also separable:

$$\hat{h}(\omega_1,\omega_2) = \hat{h}_1(\omega_1)\,\hat{h}_2(\omega_2). \tag{3.56}$$

3.4.3 Circular Convolutions and Fourier Basis

The discrete convolution of a finite image $\tilde{f}$ raises border problems. As in one dimension, these border issues are solved by extending the image, making it periodic along its rows and columns:

$$f[n_1,n_2] = \tilde{f}[n_1 \bmod N, n_2 \bmod N].$$

The resulting image $f[n_1,n_2]$ is defined for all $(n_1,n_2) \in \mathbb{Z}^2$, and each of its rows and columns is a one-dimensional signal of period N.

A discrete convolution is replaced by a circular convolution over the image period. If f and h have period N along their rows and columns, then

$$f \circledast h[n_1,n_2] = \sum_{p_1,p_2=0}^{N-1} f[p_1,p_2]\,h[n_1-p_1,n_2-p_2]. \tag{3.57}$$

Discrete Fourier Transform The eigenvectors of circular convolutions are two-dimensional discrete sinusoidal waves:

$$e_{k_1,k_2}[n_1,n_2] = \exp\left(\frac{i2\pi}{N}(k_1 n_1 + k_2 n_2)\right).$$

This family of N^2 discrete vectors is the separable product of two one-dimensional discrete Fourier bases $\{\exp(i2\pi kn/N)\}_{0\le k<N}$. Theorem A.3 thus proves that the family

$$\left\{ e_{k_1,k_2}[n_1,n_2] = \exp\left(\frac{i2\pi}{N}(k_1 n_1 + k_2 n_2)\right) \right\}_{0\le k_1,k_2<N}$$

is an orthogonal basis of the space of images that are periodic with period N along their rows and columns. Any discrete periodic image f can be decomposed in this

orthogonal basis:

$$f[n_1,n_2] = \frac{1}{N^2} \sum_{k_1,k_2=0}^{N-1} \hat{f}[k_1,k_2] \exp\left(\frac{i2\pi}{N}(k_1n_1 + k_2n_2)\right), \tag{3.58}$$

where $\hat{f}$ is the two-dimensional discrete Fourier transform of f

$$\hat{f}[k_1,k_2] = \langle f, e_{k_1,k_2}\rangle = \sum_{n_1,n_2=0}^{N-1} f[n_1,n_2] \exp\left(\frac{-i2\pi}{N}(k_1n_1 + k_2n_2)\right). \tag{3.59}$$

Fast Convolutions Since $\exp(\frac{-i2\pi}{N}(k_1n_1 + k_2n_2))$ are eigenvectors of two-dimensional circular convolutions, the discrete Fourier transform of $g = f \circledast h$ is

$$\hat{g}[k_1,k_2] = \hat{f}[k_1,k_2]\,\hat{h}[k_1,k_2]. \tag{3.60}$$

A direct computation of $f \circledast h$ with the summation (3.57) requires $O(N^4)$ multiplications. With the two-dimensional FFT described next, $\hat{f}[k_1,k_2]$ and $\hat{h}[k_1,k_2]$ as well as the inverse DFT of their product (3.60) are calculated with $O(N^2 \log N)$ operations. Non-circular convolutions are computed with a fast algorithm by reducing them to circular convolutions, with the same approach as in Section 3.3.4.

Separable Basis Decomposition Let $\{e_k\}_{0\le k<N}$ be an orthogonal basis of signals of size N. The family $\{e_{k_1}[n_1]\,e_{k_2}[n_2]\}_{0\le k_1,k_2<N}$ is then an orthogonal basis of the space of images of N^2 pixels. The decomposition coefficients of an image f in such a basis is calculated with a separable algorithm. The application to the two-dimensional FFT is explained.

Two-dimensional inner products are calculated with

$$\begin{aligned}\langle f, e_{k_1}e_{k_2}\rangle &= \sum_{n_1=0}^{N-1}\sum_{n_2=0}^{N-1} f[n_1,n_2]\,e^*_{k_1}[n_1]\,e^*_{k_2}[n_2] \\ &= \sum_{n_1=0}^{N-1} e^*_{k_1}[n] \sum_{n_2=0}^{N-1} f[n_1,n_2]\,e^*_{k_2}[n_2].\end{aligned} \tag{3.61}$$

For $0 \le n_1 < N$, we must compute

$$Tf[n_1,k_2] = \sum_{n_2=0}^{N-1} f[n_1,n_2]\,e^*_{k_2}[n_2],$$

which are the decomposition coefficients of the N image rows in the basis $\{e_{k_2}\}_{0\le k_2<N}$. The coefficients $\{\langle f, e_{k_1}e_{k_2}\rangle\}_{0\le k_1,k_2<N}$ are calculated in (3.61) as the inner products of the columns of the transformed image $Tf[n_1,k_2]$ in the same basis $\{e_k\}_{0\le k<N}$. This requires expanding $2N$ one-dimensional signals (N rows and N columns) in $\{e_k\}_{0\le k<N}$.

The fast Fourier transform algorithm of Section 3.3.3 decomposes a signal of size N in the discrete Fourier basis $\{e_k[n] = \exp(-i2\pi kn/N)\}_{0\leq k<N}$ with $KN\log_2 N$ operations. A separable implementation of a two-dimensional FFT thus requires $2KN^2\log_2 N$ operations. A split-radix FFT corresponds to $K = 3$.

3.5 PROBLEMS

3.1. [1] Suppose that $\hat{f}$ has a support in $[-(n+1)\pi/T, -n\pi/T]\cup[n\pi/T, (n+1)\pi/T]$ and that $f(t)$ is real. Find an interpolation formula that recovers $f(t)$ from $\{f(nT)\}_{n\in\mathbb{Z}}$.

3.2. [2] Suppose that $\hat{f}$ has a support in $[-\pi/T, \pi/T]$. Find a formula that recovers $f(t)$ from the average samples

$$\forall n \in \mathbb{Z} \ , \ \tilde{f}(nT) = \int_{(n-1/2)T}^{(n+1/2)T} f(t)\,dt \ .$$

3.3. [1] An interpolation function $f(t)$ satisfies $f(n) = \delta[n]$.

(a) Prove that $\sum_{k=-\infty}^{+\infty} \hat{f}(\omega + 2k\pi) = 1$ if and only if f is an interpolation function.

(b) Suppose that $f(t) = \sum_{n=-\infty}^{+\infty} h[n]\,\theta(t-n)$ with $\theta \in \mathbf{L}^2(\mathbb{R})$. Find $\hat{h}(\omega)$ so that $f(n) = \delta[n]$, and relate $\hat{f}(\omega)$ to $\hat{\theta}(\omega)$. Give a sufficient condition on $\hat{\theta}$ to guarantee that $f \in \mathbf{L}^2(\mathbb{R})$.

3.4. [1] Prove that if $f \in \mathbf{L}^2(\mathbb{R})$ and $\sum_{n=-\infty}^{+\infty} f(t-n) \in \mathbf{L}^2[0,1]$ then

$$\sum_{n=-\infty}^{+\infty} f(t-n) = \sum_{k=-\infty}^{+\infty} \hat{f}(2k\pi)\,e^{i2\pi kt} \ .$$

3.5. [1] Verify that

$$\hat{h}(\omega) = \prod_{k=1}^{K} \frac{a_k^* - e^{-i\omega}}{1 + a_k\,e^{i\omega}}$$

is an all-pass filter, i.e. $|\hat{h}(\omega)| = 1$. Prove that $\{h[n-m]\}_{m\in\mathbb{Z}}$ is an orthonormal basis of $\mathbf{l}^2(\mathbb{Z})$.

3.6. [1] Let $g[n] = (-1)^n h[n]$. Relate $\hat{g}(\omega)$ to $\hat{h}(\omega)$. If h is a low-pass filter can $\hat{g}(\omega)$ be a low-pass filter?

3.7. [1] Prove the convolution Theorem 3.3.

3.8. [2] *Recursive filters*

(a) Compute the Fourier transform of $h[n] = a^n\,\mathbf{1}_{[0,+\infty)}[n]$ for $|a| < 1$. Compute the inverse Fourier transform of $\hat{h}(\omega) = (1 - a\,e^{-i\omega})^{-p}$.

(b) Suppose that $g = f \star h$ is calculated by a recursive equation with real coefficients

$$\sum_{k=0}^{K} a_k\,f[n-k] = \sum_{k=0}^{M} b_k\,g[n-k]$$

Show that h is a stable filter if and only if the equation $\sum_{k=0}^{M} b_k\,z^{-k} = 0$ has roots with a modulus strictly smaller than 1.

(c) Suppose that $|\hat{h}(\omega)|^2 = |P(e^{-i\omega})|^2/|D(e^{-i\omega})|^2$ where $P(z)$ and $D(z)$ are polynomials. If $D(z)$ has no root of modulus 1, prove that one can find two polynomials $P_1(z)$ and $D_1(z)$ such that $\hat{h}(\omega) = P_1(e^{-i\omega})/D_1(e^{-i\omega})$ is the Fourier transform of a stable and causal recursive filter. Hint: find the complex roots of $D(z)$ and compute $D_1(z)$ by choosing the appropriate roots.

(d) A discrete Butterworth filter with cut-off frequency $\omega_c < \pi$ satisfies

$$|\hat{h}(\omega)|^2 = \frac{1}{1+\left(\frac{\tan(\omega/2)}{\tan(\omega_c/2)}\right)^{2N}}$$

Compute $\hat{h}(\omega)$ for $N = 3$ in order to obtain a filter h which is real, stable and causal.

3.9. [1] Let a and b be two integers with many digits. Relate the product ab to a convolution. Explain how to use the FFT to compute this product.

3.10. [1] Let h^{-1} be the inverse of h defined by $h \star h^{-1}[n] = \delta[n]$.

(a) Prove that if h has a finite support then h^{-1} has a finite support if and only if $h[n] = \delta[n-p]$ for some $p \in \mathbb{Z}$.

(b) Find a sufficient condition on $\hat{h}(\omega)$ for h^{-1} to be a stable filter.

3.11. [1] *Discrete interpolation* Let $\hat{f}[k]$ be the DFT of a signal $f[n]$ of size N. We define $\widehat{\hat{f}}[N/2] = \widehat{\hat{f}}[3N/2] = \widehat{f}[N/2]$ and

$$\widehat{\hat{f}}[k] = \begin{cases} 2\hat{f}[k] & \text{if } 0 \le k < N/2 \\ 0 & \text{if } N/2 < k < 3N/2 \\ 2\hat{f}[k-N] & \text{if } 3N/2 < k < 2N \end{cases}.$$

Prove that $\tilde{f}[2n] = f[n]$.

3.12. [1] *Decimation* Let $x[n] = y[Mn]$ with $M > 1$.

(a) Show that $\hat{x}(\omega) = M^{-1}\sum_{k=0}^{M-1} \hat{y}(M^{-1}(\omega - 2k\pi))$.

(b) Give a sufficient condition on $\hat{y}(\omega)$ to recover y from x. Describe the interpolation algorithm.

3.13. [1] *Complexity of FFT*

(a) Find an algorithm that multiplies two complex numbers with 3 additions and 3 multiplications.

(b) Compute the total number of additions and multiplications of the FFT algorithm described in Section 3.3.3, for a signal of size N.

3.14. [2] We want to compute numerically the Fourier transform of $f(t)$. Let $f_d[n] = f(nT)$, and $f_p[n] = \sum_{p=-\infty}^{+\infty} f_d[n-pN]$.

(a) Prove that the DFT of $f_p[n]$ is related to the Fourier series of $f_d[n]$ and to the Fourier transform of $f(t)$ by

$$\hat{f}_p[k] = \hat{f}_d\left(\frac{2\pi k}{N}\right) = \frac{1}{T}\sum_{l=-\infty}^{+\infty} \hat{f}\left(\frac{2k\pi}{NT} - \frac{2l\pi}{T}\right).$$

(b) Suppose that $|f(t)|$ and $|\hat{f}(\omega)|$ are negligible when $t \notin [-t_0, t_0]$ and $\omega \notin [-\omega_0, \omega_0]$. Relate N and T to t_0 and ω_0 so that one can compute an approximation value of $\hat{f}(\omega)$ at all $\omega \in \mathbb{R}$ by interpolating the samples $\hat{f}_p[k]$. Is it possible to compute exactly $\hat{f}(\omega)$ with such an interpolation formula?

(c) Let $f(t) = \left(\sin(\pi t)/(\pi t)\right)^4$. What is the support of $\hat{f}$? Sample f appropriately and compute $\hat{f}$ with the FFT algorithm of MATLAB.

3.15. [1] Suppose that $f[n_1, n_2]$ is an image with N^2 non-zero pixels for $0 \leq n_1, n_2 < N$. Let $h[n_1, n_2]$ be a non-separable filter with M^2 non-zero coefficients for $0 \leq n_1, n_2 < M$. Describe an overlap-add algorithm to compute $g[n_1, n_2] = f \star h[n_1, n_2]$. How many operations does it require? For what range of M is it better to compute the convolution with a direct summation?

IV

TIME MEETS FREQUENCY

When we listen to music, we clearly "hear" the time variation of the sound "frequencies." These localized frequency events are not pure tones but packets of close frequencies. The properties of sounds are revealed by transforms that decompose signals over elementary functions that are well concentrated in time and frequency. Windowed Fourier transforms and wavelet transforms are two important classes of local time-frequency decompositions. Measuring the time variations of "instantaneous" frequencies is an important application that illustrates the limitations imposed by the Heisenberg uncertainty.

There is no unique definition of time-frequency energy density, which makes this topic difficult. Yet, some order can be established by proving that quadratic time-frequency distributions are obtained by averaging a single quadratic form called the Wigner-Ville distribution. This unified framework gives a more general perspective on windowed Fourier transforms and wavelet transforms.

4.1 TIME-FREQUENCY ATOMS [1]

A linear time-frequency transform correlates the signal with a family of waveforms that are well concentrated in time and in frequency. These waveforms are called *time-frequency atoms*. Let us consider a general family of time-frequency atoms $\{\phi_\gamma\}_{\gamma\in\Gamma}$, where γ might be a multi-index parameter. We suppose that $\phi_\gamma \in \mathbf{L}^2(\mathbb{R})$ and that $\|\phi_\gamma\| = 1$. The corresponding linear time-frequency transform of $f \in$

$\mathbf{L}^2(\mathbb{R})$ is defined by

$$Tf(\gamma) = \int_{-\infty}^{+\infty} f(t)\, \phi_\gamma^*(t)\, dt = \langle f, \phi_\gamma \rangle.$$

The Parseval formula (2.25) proves that

$$Tf(\gamma) = \int_{-\infty}^{+\infty} f(t)\, \phi_\gamma^*(t)\, dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega)\, \hat{\phi}_\gamma^*(\omega)\, d\omega. \tag{4.1}$$

If $\phi_\gamma(t)$ is nearly zero when t is outside a neighborhood of an abscissa u, then $\langle f, \phi_\gamma \rangle$ depends only on the values of f in this neighborhood. Similarly, if $\hat{\phi}_\gamma(\omega)$ is negligible for ω far from ξ, then the right integral of (4.1) proves that $\langle f, \phi_\gamma \rangle$ reveals the properties of $\hat{f}$ in the neighborhood of ξ.

Example 4.1 A windowed Fourier atom is constructed with a window g translated by u and modulated by the frequency ξ:

$$\phi_\gamma(t) = g_{\xi,u}(t) = e^{i\xi t}\, g(t-u). \tag{4.2}$$

A wavelet atom is a dilation by s and a translation by u of a *mother wavelet* ψ:

$$\phi_\gamma(t) = \psi_{s,u}(t) = \frac{1}{\sqrt{s}}\, \psi\left(\frac{t-u}{s}\right). \tag{4.3}$$

Wavelets and windowed Fourier functions have their energy well localized in time, while their Fourier transform is mostly concentrated in a limited frequency band. The properties of the resulting transforms are studied in Sections 4.2 and 4.3.

Heisenberg Boxes The slice of information provided by $\langle f, \phi_\gamma \rangle$ is represented in a time-frequency plane (t, ω) by a region whose location and width depends on the time-frequency spread of ϕ_γ. Since

$$\|\phi_\gamma\|^2 = \int_{-\infty}^{+\infty} |\phi_\gamma(t)|^2\, dt = 1,$$

we interpret $|\phi_\gamma(t)|^2$ as a probability distribution centered at

$$u_\gamma = \int_{-\infty}^{+\infty} t\, |\phi_\gamma(t)|^2\, dt. \tag{4.4}$$

The spread around u_γ is measured by the variance

$$\sigma_t^2(\gamma) = \int_{-\infty}^{+\infty} (t-u_\gamma)^2\, |\phi_\gamma(t)|^2\, dt. \tag{4.5}$$

The Plancherel formula (2.26) proves that $\int_{-\infty}^{+\infty} |\hat{\phi}_\gamma(\omega)|^2 d\omega = 2\pi \|\phi_\gamma\|^2$. The center frequency of $\hat{\phi}_\gamma$ is therefore defined by

$$\xi_\gamma = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \omega |\hat{\phi}_\gamma(\omega)|^2 d\omega, \tag{4.6}$$

and its spread around ξ_γ is

$$\sigma_\omega^2(\gamma) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} (\omega - \xi_\gamma)^2 |\hat{\phi}_\gamma(\omega)|^2 d\omega. \tag{4.7}$$

The time-frequency resolution of ϕ_γ is represented in the time-frequency plane (t,ω) by a Heisenberg box centered at (u_γ,ξ_γ), whose width along time is $\sigma_t(\gamma)$ and whose width along frequency is $\sigma_\omega(\gamma)$. This is illustrated by Figure 4.1. The Heisenberg uncertainty Theorem 2.5 proves that the area of the rectangle is at least $1/2$:

$$\sigma_t \sigma_\omega \geq \frac{1}{2}. \tag{4.8}$$

This limits the joint resolution of ϕ_γ in time and frequency. The time-frequency plane must be manipulated carefully because a point (t_0,ω_0) is ill-defined. There is no function that is perfectly well concentrated at a point t_0 and a frequency ω_0. Only rectangles with area at least $1/2$ may correspond to time-frequency atoms.

Energy Density Suppose that for any (u,ξ) there exists a unique atom $\phi_{\gamma(u,\xi)}$ centered at (u,ξ) in the time-frequency plane. The time-frequency box of $\phi_{\gamma(u,\xi)}$ specifies a neighborhood of (u,ξ) where the energy of f is measured by

$$P_T f(u,\xi) = |\langle f, \phi_{\gamma(u,\xi)} \rangle|^2 = \left| \int_{-\infty}^{+\infty} f(t)\, \phi^*_{\gamma(u,\xi)}(t)\, dt \right|^2. \tag{4.9}$$

Section 4.5.1 proves that any such energy density is an averaging of the Wigner-Ville distribution, with a kernel that depends on the atoms ϕ_γ.

4.2 WINDOWED FOURIER TRANSFORM [1]

In 1946, Gabor [187] introduced windowed Fourier atoms to measure the "frequency variations" of sounds. A real and symmetric window $g(t) = g(-t)$ is translated by u and modulated by the frequency ξ:

$$g_{u,\xi}(t) = e^{i\xi t} g(t-u). \tag{4.10}$$

It is normalized $\|g\| = 1$ so that $\|g_{u,\xi}\| = 1$ for any $(u,\xi) \in \mathbb{R}^2$. The resulting windowed Fourier transform of $f \in \mathbf{L}^2(\mathbb{R})$ is

$$Sf(u,\xi) = \langle f, g_{u,\xi} \rangle = \int_{-\infty}^{+\infty} f(t)\, g(t-u)\, e^{-i\xi t}\, dt. \tag{4.11}$$

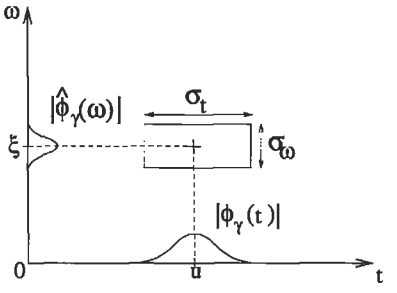


FIGURE 4.1 Heisenberg box representing an atom ϕ_γ.

This transform is also called the *short time Fourier transform* because the multiplication by $g(t-u)$ localizes the Fourier integral in the neighborhood of $t=u$.

As in (4.9), one can define an energy density called a *spectrogram*, denoted P_S:

$$P_S f(u,\xi) = |Sf(u,\xi)|^2 = \left| \int_{-\infty}^{+\infty} f(t)\, g(t-u)\, e^{-i\xi t}\, dt \right|^2 . \tag{4.12}$$

The spectrogram measures the energy of f in the time-frequency neighborhood of (u,ξ) specified by the Heisenberg box of $g_{u,\xi}$.

Heisenberg Boxes Since g is even, $g_{u,\xi}(t) = e^{i\xi t} g(t-u)$ is centered at u. The time spread around u is independent of u and ξ:

$$\sigma_t^2 = \int_{-\infty}^{+\infty} (t-u)^2 \, |g_{u,\xi}(t)|^2 \, dt = \int_{-\infty}^{+\infty} t^2 \, |g(t)|^2 \, dt. \tag{4.13}$$

The Fourier transform $\hat{g}$ of g is real and symmetric because g is real and symmetric. The Fourier transform of $g_{u,\xi}$ is

$$\hat{g}_{u,\xi}(\omega) = \hat{g}(\omega-\xi) \exp[-iu(\omega-\xi)]. \tag{4.14}$$

It is a translation by ξ of the frequency window $\hat{g}$, so its center frequency is ξ. The frequency spread around ξ is

$$\sigma_\omega^2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} (\omega-\xi)^2 \, |\hat{g}_{u,\xi}(\omega)| \, d\omega = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \omega^2 \, |\hat{g}(\omega)| \, d\omega. \tag{4.15}$$

It is independent of u and ξ. Hence $g_{u,\xi}$ corresponds to a Heisenberg box of area $\sigma_t \sigma_\omega$ centered at (u,ξ), as illustrated by Figure 4.2. The size of this box is independent of (u,ξ), which means that a windowed Fourier transform has the same resolution across the time-frequency plane.

Example 4.2 A sinusoidal wave $f(t) = \exp(i\xi_0 t)$ whose Fourier transform is a Dirac $\hat{f}(\omega) = 2\pi\delta(\omega-\xi_0)$ has a windowed Fourier transform

$$Sf(u,\xi) = \hat{g}(\xi-\xi_0) \exp[-iu(\xi-\xi_0)].$$

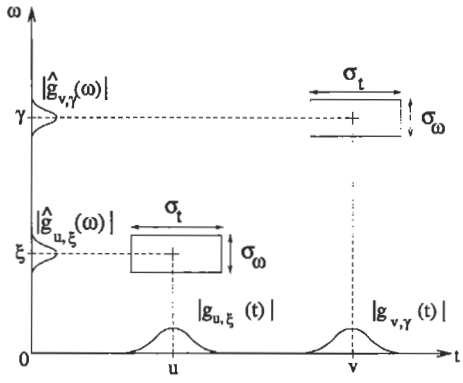


FIGURE 4.2 Heisenberg boxes of two windowed Fourier atoms $g_{u,\xi}$ and $g_{v,\gamma}$.

Its energy is spread over the frequency interval $[\xi_0 - \sigma_\omega/2, \xi_0 + \sigma_\omega/2]$.

Example 4.3 The windowed Fourier transform of a Dirac $f(t) = \delta(t - u_0)$ is

$$Sf(u,\xi) = g(u_0 - u)\exp(-i\xi u_0).$$

Its energy is spread in the time interval $[u_0 - \sigma_t/2, u_0 + \sigma_t/2]$.

Example 4.4 A linear chirp $f(t) = \exp(iat^2)$ has an "instantaneous frequency" that increases linearly in time. For a Gaussian window $g(t) = (\pi\sigma^2)^{-1/4}\exp[-t^2/(2\sigma^2)]$, the windowed Fourier transform of f is calculated using the Fourier transform (2.34) of Gaussian chirps. One can verify that its spectrogram is

$$P_S f(u,\xi) = |Sf(u,\xi)|^2 = \left(\frac{4\pi\sigma^2}{1+4a^2\sigma^4}\right)^{1/2} \exp\left(-\frac{\sigma^2(\xi - 2au)^2}{1+4a^2\sigma^4}\right). \tag{4.16}$$

For a fixed time u, $P_S f(u,\xi)$ is a Gaussian that reaches its maximum at the frequency $\xi(u) = 2au$. Observe that if we write $f(t) = \exp[i\phi(t)]$, then $\xi(u)$ is equal to the "instantaneous frequency," defined as the derivative of the phase: $\omega(u) = \phi'(u) = 2au$. Section 4.4.1 explains this result.

Example 4.5 Figure 4.3 gives the spectrogram of a signal that includes a linear chirp, a quadratic chirp and two modulated Gaussians. The spectrogram is computed with a Gaussian window dilated by $\sigma = 0.05$. As expected from (4.16), the linear chirp yields large amplitude coefficients along the trajectory of its instantaneous frequency, which is a straight line. The quadratic chirp yields large

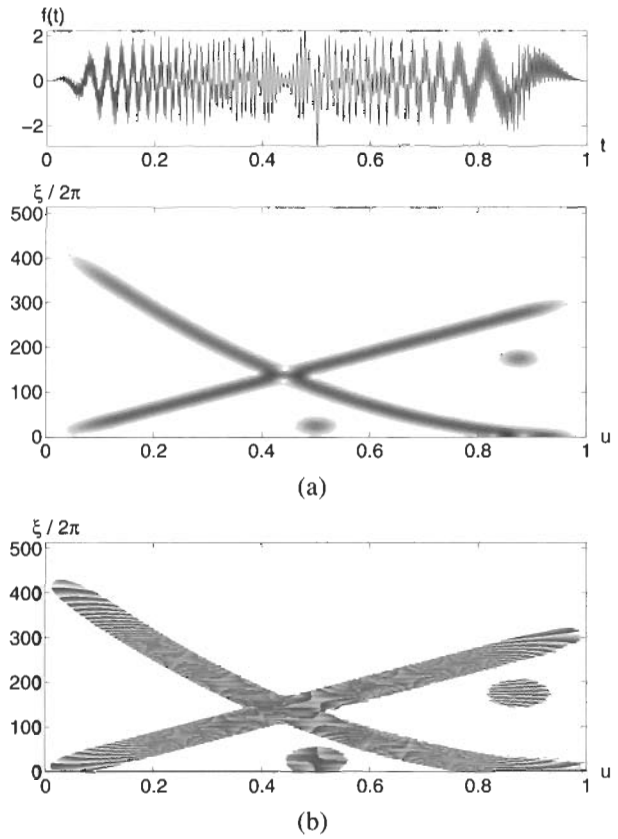


FIGURE 4.3 The signal includes a linear chirp whose frequency increases, a quadratic chirp whose frequency decreases, and two modulated Gaussian functions located at $t = 0.5$ and $t = 0.87$. (a) Spectrogram $P_S f(u,\xi)$. Dark points indicate large amplitude coefficients. (b) Complex phase of $Sf(u,\xi)$ in regions where the modulus $P_S f(u,\xi)$ is non-zero.

coefficients along a parabola. The two modulated Gaussians produce low and high frequency blobs at $u = 0.5$ and $u = 0.87$.

4.2.1 Completeness and Stability

When the time-frequency indices (u,ξ) vary across $\mathbb{R}^2$, the Heisenberg boxes of the atoms $g_{u,\xi}$ cover the whole time-frequency plane. One can thus expect that f can be recovered from its windowed Fourier transform $Sf(u,\xi)$. The following theorem gives a reconstruction formula and proves that the energy is conserved.

Theorem 4.1 *If* $f \in \mathbf{L}^2(\mathbb{R})$ *then*

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} Sf(u,\xi)\, g(t-u)\, e^{i\xi t}\, d\xi\, du \tag{4.17}$$

and

$$\int_{-\infty}^{+\infty} |f(t)|^2\, dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |Sf(u,\xi)|^2\, d\xi\, du. \tag{4.18}$$

Proof[1]. The reconstruction formula (4.17) is proved first. Let us apply the Fourier Parseval formula (2.25) to the integral (4.17) with respect to the integration in u. The Fourier transform of $f_\xi(u) = Sf(u,\xi)$ with respect to u is computed by observing that

$$Sf(u,\xi) = \exp(-iu\xi) \int_{-\infty}^{+\infty} f(t)\, g(t-u)\, \exp[i\xi(u-t)]\, dt = \exp(-iu\xi)\, f \star g_\xi(u),$$

where $g_\xi(t) = g(t)\exp(i\xi t)$, because $g(t) = g(-t)$. Its Fourier transform is therefore

$$\hat{f}_\xi(\omega) = \hat{f}(\omega+\xi)\, \hat{g}_\xi(\omega+\xi) = \hat{f}(\omega+\xi)\, \hat{g}(\omega).$$

The Fourier transform of $g(t-u)$ with respect to u is $\hat{g}(\omega)\exp(-it\omega)$. Hence

$$\frac{1}{2\pi} \left(\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} Sf(u,\xi)\, g(t-u)\, \exp(i\xi t)\, du \right) d\xi \quad =$$
$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} \left(\frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega+\xi)\, |\hat{g}(\omega)|^2\, \exp[it(\omega+\xi)]\, d\omega \right) d\xi\ .$$

If $\hat{f} \in \mathbf{L}^1(\mathbb{R})$, we can apply the Fubini Theorem A.2 to reverse the integration order. The inverse Fourier transform proves that

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega+\xi)\, \exp[it(\omega+\xi)]\, d\xi = f(t).$$

Since $\frac{1}{2\pi}\int_{-\infty}^{+\infty} |\hat{g}(\omega)|^2\, d\omega = 1$, we derive (4.17). If $\hat{f} \notin \mathbf{L}^1(\mathbb{R})$, a density argument is used to verify this formula.

Let us now prove the energy conservation (4.18). Since the Fourier transform in u of $Sf(u,\xi)$ is $\hat{f}(\omega+\xi)\,\hat{g}(\omega)$, the Plancherel formula (2.26) applied to the right-hand side of (4.18) gives

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |Sf(u,\xi)|^2\, du\, d\xi = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\hat{f}(\omega+\xi)\, \hat{g}(\omega)|^2\, d\omega\, d\xi.$$

The Fubini theorem applies and the Plancherel formula proves that

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} |\hat{f}(\omega+\xi)|^2\, d\xi = \|f\|^2,$$

which implies (4.18). ■

The reconstruction formula (4.17) can be rewritten

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \langle f, g_{u,\xi} \rangle \, g_{u,\xi}(t) \, d\xi \, du. \tag{4.19}$$

It resembles the decomposition of a signal in an orthonormal basis but it is not, since the functions $\{g_{u,\xi}\}_{u,\xi \in \mathbb{R}^2}$ are very redundant in $\mathbf{L}^2(\mathbb{R})$. The second equality (4.18) justifies the interpretation of the spectrogram $P_S f(u,\xi) = |Sf(u,\xi)|^2$ as an energy density, since its time-frequency sum equals the signal energy.

Reproducing Kernel A windowed Fourier transform represents a one-dimensional signal $f(t)$ by a two-dimensional function $Sf(u,\xi)$. The energy conservation proves that $Sf \in \mathbf{L}^2(\mathbb{R}^2)$. Because $Sf(u,\xi)$ is redundant, it is not true that any $\Phi \in \mathbf{L}^2(\mathbb{R}^2)$ is the windowed Fourier transform of some $f \in \mathbf{L}^2(\mathbb{R})$. The next proposition gives a necessary and sufficient condition for such a function to be a windowed Fourier transform.

Proposition 4.1 *Let $\Phi \in \mathbf{L}^2(\mathbb{R}^2)$. There exists $f \in \mathbf{L}^2(\mathbb{R})$ such that $\Phi(u,\xi) = Sf(u,\xi)$ if and only if*

$$\Phi(u_0,\xi_0) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Phi(u,\xi) \, K(u_0,u,\xi_0,\xi) \, du \, d\xi, \tag{4.20}$$

with

$$K(u_0,u,\xi_0,\xi) = \langle g_{u,\xi}, g_{u_0,\xi_0} \rangle \, . \tag{4.21}$$

Proof[2]. Suppose that there exists f such that $\Phi(u,\xi) = Sf(u,\xi)$. Let us replace f with its reconstruction integral (4.17) in the windowed Fourier transform definition:

$$Sf(u_0,\xi_0) = \int_{-\infty}^{+\infty} \left(\frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} Sf(u,\xi) \, g_{u,\xi}(t) \, du \, d\xi \right) g^*_{u_0,\xi_0}(t) \, dt. \tag{4.22}$$

Inverting the integral on t with the integrals on u and ξ yields (4.20). To prove that the condition (4.20) is sufficient, we define f as in the reconstruction formula (4.17):

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Phi(u,\xi) \, g(t-u) \exp(i\xi t) \, d\xi \, du$$

and show that (4.20) implies that $\Phi(u,\xi) = Sf(u,\xi)$. ■

Ambiguity Function The reproducing kernel $K(u_0,u,\xi_0,\xi)$ measures the time-frequency overlap of the two atoms $g_{u,\xi}$ and g_{u_0,ξ_0}. The amplitude of $K(u_0,u,\xi_0,\xi)$ decays with $u_0 - u$ and $\xi_0 - \xi$ at a rate that depends on the energy concentration of g and $\hat{g}$. Replacing $g_{u,\xi}$ and g_{u_0,ξ_0} by their expression and making the change of variable $v = t - (u+u_0)/2$ in the inner product integral (4.21) yields

$$K(u_0,u,\xi_0,\xi) = \exp\left(-\frac{i}{2}(\xi_0 - \xi)(u+u_0) \right) A g(u_0 - u, \xi_0 - \xi) \tag{4.23}$$

where

$$Ag(\tau,\gamma) = \int_{-\infty}^{+\infty} g\left(v+\frac{\tau}{2}\right) g\left(v-\frac{\tau}{2}\right) e^{-i\gamma v}\, dv \tag{4.24}$$

is called the *ambiguity function* of g. Using the Parseval formula to replace this time integral with a Fourier integral gives

$$Ag(\tau,\gamma) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{g}\left(\omega+\frac{\gamma}{2}\right) \hat{g}\left(\omega-\frac{\gamma}{2}\right) e^{i\tau\omega}\, d\omega. \tag{4.25}$$

The decay of the ambiguity function measures the spread of g in time and of $\hat{g}$ in frequency. For example, if g has a support included in an interval of size T, then $Ag(\tau,\omega) = 0$ for $|\tau| \geq T/2$. The integral (4.25) shows that the same result applies to the support of $\hat{g}$.

4.2.2 Choice of Window [2]

The resolution in time and frequency of the windowed Fourier transform depends on the spread of the window in time and frequency. This can be measured from the decay of the ambiguity function (4.24) or more simply from the area $\sigma_t \sigma_\omega$ of the Heisenberg box. The uncertainty Theorem 2.5 proves that this area reaches the minimum value $1/2$ if and only if g is a Gaussian. The ambiguity function $Ag(\tau,\gamma)$ is then a two-dimensional Gaussian.

Window Scale The time-frequency localization of g can be modified with a scaling. Suppose that g has a Heisenberg time and frequency width respectively equal to σ_t and σ_ω. Let $g_s(t) = s^{-1/2} g(t/s)$ be its dilation by s. A change of variables in the integrals (4.13) and (4.15) shows that the Heisenberg time and frequency width of g_s are respectively $s\sigma_t$ and σ_ω/s. The area of the Heisenberg box is not modified but it is dilated by s in time and compressed by s in frequency. Similarly, a change of variable in the ambiguity integral (4.24) shows that the ambiguity function is dilated in time and frequency respectively by s and $1/s$

$$Ag_s(\tau,\gamma) = Ag\left(\frac{\tau}{s}, s\gamma\right).$$

The choice of a particular scale s depends on the desired resolution trade-off between time and frequency.

Finite Support In numerical applications, g must have a compact support. Theorem 2.6 proves that its Fourier transform $\hat{g}$ necessarily has an infinite support. It is a symmetric function with a main lobe centered at $\omega = 0$, which decays to zero with oscillations. Figure 4.4 illustrates its behavior. To maximize the frequency resolution of the transform, we must concentrate the energy of $\hat{g}$ near $\omega = 0$. Three important parameters evaluate the spread of $\hat{g}$:

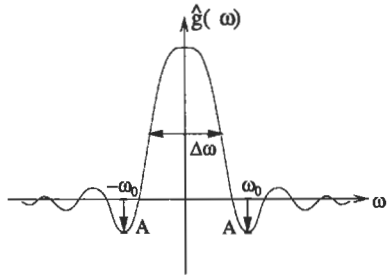


FIGURE 4.4 The energy spread of $\hat{g}$ is measured by its bandwidth $\Delta\omega$ and the maximum amplitude A of the first side-lobes, located at $\omega = \pm\omega_0$.

- The root mean-square bandwidth $\Delta\omega$, which is defined by

$$\frac{|\hat{g}(\Delta\omega/2)|^2}{|\hat{g}(0)|^2} = \frac{1}{2}.$$

- The maximum amplitude A of the first side-lobes located at $\omega = \pm\omega_0$ in Figure 4.4. It is measured in decibels:

$$A = 10\log_{10}\frac{|\hat{g}(\omega_0)|^2}{|\hat{g}(0)|^2}.$$

- The polynomial exponent p, which gives the asymptotic decay of $|\hat{g}(\omega)|$ for large frequencies:

$$|\hat{g}(\omega)| = O(\omega^{-p-1}). \tag{4.26}$$

  Table 4.1 gives the values of these three parameters for several windows g whose supports are restricted to $[-1/2, 1/2]$ [204]. Figure 4.5 shows the graph of these windows.

To interpret these three frequency parameters, let us consider the spectrogram of a frequency tone $f(t) = \exp(i\xi_0 t)$. If $\Delta\omega$ is small, then $|Sf(u,\xi)|^2 = |\hat{g}(\xi - \xi_0)|^2$ has an energy concentrated near $\xi = \xi_0$. The side-lobes of $\hat{g}$ create "shadows" at $\xi = \xi_0 \pm \omega_0$, which can be neglected if A is also small.

If the frequency tone is embedded in a signal that has other components of much higher energy at different frequencies, the tone can still be detected if $\hat{g}(\omega - \xi)$ attenuates these components rapidly when $|\omega - \xi|$ increases. This means that $|\hat{g}(\omega)|$ has a rapid decay, and Proposition 2.1 proves that this decay depends on the regularity of g. Property (4.26) is typically satisfied by windows that are p times differentiable.

Name	$g(t)$	$\Delta\omega$	A	p
Rectangle	1	0.89	−13db	0
Hamming	$0.54+0.46\cos(2\pi t)$	1.36	−43db	0
Gaussian	$\exp(-18t^2)$	1.55	−55db	0
Hanning	$\cos^2(\pi t)$	1.44	−32db	2
Blackman	$0.42+0.5\cos(2\pi t)$ $+0.08\cos(4\pi t)$	1.68	−58db	2

Table 4.1 Frequency parameters of five windows g whose supports are restricted to $[-1/2, 1/2]$. These windows are normalized so that $g(0) = 1$ but $\|g\| \neq 1$.

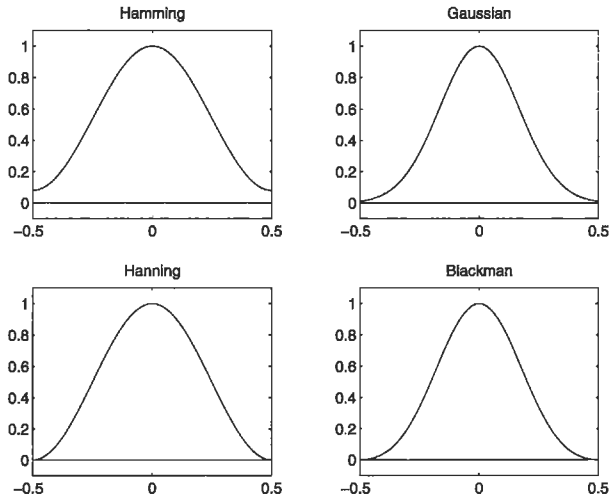


FIGURE 4.5 Graphs of four windows g whose support are $[-1/2, 1/2]$.

4.2.3 Discrete Windowed Fourier Transform [2]

The discretization and fast computation of the windowed Fourier transform follow the same ideas as the discretization of the Fourier transform described in Section 3.3. We consider discrete signals of period N. The window $g[n]$ is chosen to be a symmetric discrete signal of period N with unit norm $\|g\| = 1$. Discrete windowed Fourier atoms are defined by

$$g_{m,l}[n] = g[n-m] \exp\left(\frac{i2\pi ln}{N}\right).$$

The discrete Fourier transform of $g_{m,l}$ is

$$\hat{g}_{m,l}[k] = \hat{g}[k-l] \exp\left(\frac{-i2\pi m(k-l)}{N}\right).$$

The discrete windowed Fourier transform of a signal f of period N is

$$Sf[m,l] = \langle f, g_{m,l}\rangle = \sum_{n=0}^{N-1} f[n]\, g[n-m] \exp\left(\frac{-i2\pi ln}{N}\right), \tag{4.27}$$

For each $0 \leq m < N$, $Sf[m,l]$ is calculated for $0 \leq l < N$ with a discrete Fourier transform of $f[n]g[n-m]$. This is performed with N FFT procedures of size N, and thus requires a total of $O(N^2 \log_2 N)$ operations. Figure 4.3 is computed with this algorithm.

Inverse Transform The following theorem discretizes the reconstruction formula and the energy conservation of Theorem 4.1.

Theorem 4.2 *If f is a signal of period N then*

$$f[n] = \frac{1}{N} \sum_{m=0}^{N-1} \sum_{l=0}^{N-1} Sf[m,l]\, g[n-m] \exp\left(\frac{i2\pi ln}{N}\right) \tag{4.28}$$

and

$$\sum_{n=0}^{N-1} |f[n]|^2 = \frac{1}{N} \sum_{l=0}^{N-1} \sum_{m=0}^{N-1} |Sf[m,l]|^2. \tag{4.29}$$

This theorem is proved by applying the Parseval and Plancherel formulas of the discrete Fourier transform, exactly as in the proof of Theorem 4.1. The reconstruction formula (4.28) is rewritten

$$f[n] = \frac{1}{N} \sum_{m=0}^{N-1} g[n-m] \sum_{l=0}^{N-1} Sf[m,l] \exp\left(\frac{i2\pi ln}{N}\right).$$

The second sum computes for each $0 \leq m < N$ the inverse discrete Fourier transform of $Sf[m,l]$ with respect to l. This is calculated with N FFT procedures, requiring a total of $O(N^2 \log_2 N)$ operations.

A discrete windowed Fourier transform is an N^2 image $Sf[l,m]$ that is very redundant, since it is entirely specified by a signal f of size N. The redundancy is characterized by a discrete reproducing kernel equation, which is the discrete equivalent of (4.20).

4.3 WAVELET TRANSFORMS [1]

To analyze signal structures of very different sizes, it is necessary to use time-frequency atoms with different time supports. The wavelet transform decomposes signals over dilated and translated wavelets. A wavelet is a function $\psi \in \mathbf{L}^2(\mathbb{R})$ with a zero average:

$$\int_{-\infty}^{+\infty} \psi(t)\, dt = 0. \tag{4.30}$$

It is normalized $\|\psi\| = 1$, and centered in the neighborhood of $t = 0$. A family of time-frequency atoms is obtained by scaling ψ by s and translating it by u:

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right).$$

These atoms remain normalized: $\|\psi_{u,s}\| = 1$. The wavelet transform of $f \in \mathbf{L}^2(\mathbb{R})$ at time u and scale s is

$$Wf(u,s) = \langle f, \psi_{u,s} \rangle = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{s}} \psi^*\left(\frac{t-u}{s}\right) dt. \tag{4.31}$$

Linear Filtering The wavelet transform can be rewritten as a convolution product:

$$Wf(u,s) = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{s}} \psi^*\left(\frac{t-u}{s}\right) dt = f \star \bar{\psi}_s(u) \tag{4.32}$$

with

$$\bar{\psi}_s(t) = \frac{1}{\sqrt{s}} \psi^*\left(\frac{-t}{s}\right).$$

The Fourier transform of $\bar{\psi}_s(t)$ is

$$\hat{\bar{\psi}}_s(\omega) = \sqrt{s}\, \hat{\psi}^*(s\omega). \tag{4.33}$$

Since $\hat{\psi}(0) = \int_{-\infty}^{+\infty} \psi(t)\, dt = 0$, it appears that $\hat{\psi}$ is the transfer function of a band-pass filter. The convolution (4.32) computes the wavelet transform with dilated band-pass filters.

Analytic Versus Real Wavelets Like a windowed Fourier transform, a wavelet transform can measure the time evolution of frequency transients. This requires using a complex analytic wavelet, which can separate amplitude and phase components. The properties of this analytic wavelet transform are described in Section 4.3.2, and its application to the measurement of instantaneous frequencies is explained in Section 4.4.2. In contrast, real wavelets are often used to detect sharp signal transitions. Section 4.3.1 introduces elementary properties of real wavelets, which are developed in Chapter 6.

4.3.1 Real Wavelets

Suppose that ψ is a real wavelet. Since it has a zero average, the wavelet integral

$$Wf(u,s) = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{s}} \psi^* \left(\frac{t-u}{s} \right) dt$$

measures the variation of f in a neighborhood of u, whose size is proportional to s. Section 6.1.3 proves that when the scale s goes to zero, the decay of the wavelet coefficients characterizes the regularity of f in the neighborhood of u. This has important applications for detecting transients and analyzing fractals. This section concentrates on the completeness and redundancy properties of real wavelet transforms.

Example 4.6 Wavelets equal to the second derivative of a Gaussian are called *Mexican hats*. They were first used in computer vision to detect multiscale edges [354]. The normalized Mexican hat wavelet is

$$\psi(t) = \frac{2}{\pi^{1/4}\sqrt{3\sigma}} \left(\frac{t^2}{\sigma^2} - 1 \right) \exp\left(\frac{-t^2}{2\sigma^2} \right). \tag{4.34}$$

For $\sigma = 1$, Figure 4.6 plots $-\psi$ and its Fourier transform

$$\hat{\psi}(\omega) = \frac{-\sqrt{8}\,\sigma^{5/2}\pi^{1/4}}{\sqrt{3}} \omega^2 \exp\left(\frac{-\sigma^2\omega^2}{2} \right). \tag{4.35}$$

Figure 4.7 shows the wavelet transform of a signal that is piecewise regular on the left and almost everywhere singular on the right. The maximum scale is smaller than 1 because the support of f is normalized to $[0,1]$. The minimum scale is limited by the sampling interval of the discretized signal used in numerical calculations. When the scale decreases, the wavelet transform has a rapid decay to zero in the regions where the signal is regular. The isolated singularities on the left create cones of large amplitude wavelet coefficients that converge to the locations of the singularities. This is further explained in Chapter 6.

A real wavelet transform is complete and maintains an energy conservation, as long as the wavelet satisfies a weak admissibility condition, specified by the following theorem. This theorem was first proved in 1964 by the mathematician Calderón [111], from a different point of view. Wavelets did not appear as such, but Calderón defines a wavelet transform as a convolution operator that decomposes the identity. Grossmann and Morlet [200] were not aware of Calderón's work when they proved the same formula for signal processing.

FIGURE 4.6 Mexican hat wavelet (4.34) for $\sigma = 1$ and its Fourier transform.

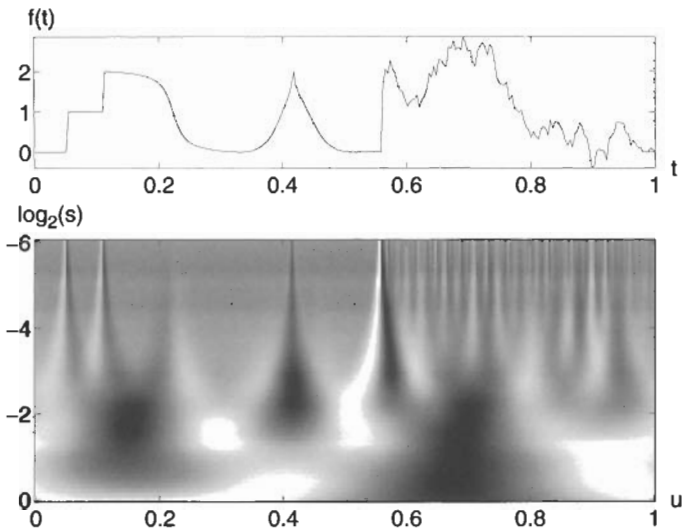


FIGURE 4.7 Real wavelet transform $Wf(u,s)$ computed with a Mexican hat wavelet (4.34). The vertical axis represents $\log_2 s$. Black, grey and white points correspond respectively to positive, zero and negative wavelet coefficients.

Theorem 4.3 (CALDERÓN, GROSSMANN, MORLET) *Let $\psi \in \mathbf{L}^2(\mathbb{R})$ be a real function such that*

$$C_\psi = \int_0^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega}\, d\omega < +\infty. \tag{4.36}$$

Any $f \in \mathbf{L}^2(\mathbb{R})$ satisfies

$$f(t) = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} Wf(u,s)\, \frac{1}{\sqrt{s}}\, \psi\left(\frac{t-u}{s}\right) du\, \frac{ds}{s^2}, \tag{4.37}$$

and

$$\int_{-\infty}^{+\infty} |f(t)|^2 dt = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} |Wf(u,s)|^2 \, du \frac{ds}{s^2}. \tag{4.38}$$

Proof[1]. The proof of (4.38) is almost identical to the proof of (4.18). Let us concentrate on the proof of (4.37). The right integral $b(t)$ of (4.37) can be rewritten as a sum of convolutions. Inserting $Wf(u,s) = f \star \bar{\psi}_s(u)$ with $\psi_s(t) = s^{-1/2}\psi(t/s)$ yields

$$\begin{aligned} b(t) &= \frac{1}{C_\psi} \int_0^{+\infty} Wf(.,s) \star \psi_s(t) \frac{ds}{s^2} \\ &= \frac{1}{C_\psi} \int_0^{+\infty} f \star \bar{\psi}_s \star \psi_s(t) \frac{ds}{s^2}. \end{aligned} \tag{4.39}$$

The "." indicates the variable over which the convolution is calculated. We prove that $b = f$ by showing that their Fourier transforms are equal. The Fourier transform of b is

$$\hat{b}(\omega) = \frac{1}{C_\psi} \int_0^{+\infty} \hat{f}(\omega) \sqrt{s} \hat{\psi}^*(s\omega) \sqrt{s} \hat{\psi}(s\omega) \frac{ds}{s^2} = \frac{\hat{f}(\omega)}{C_\psi} \int_0^{+\infty} |\hat{\psi}(s\omega)|^2 \frac{ds}{s}.$$

Since ψ is real we know that $|\hat{\psi}(-\omega)|^2 = |\hat{\psi}(\omega)|^2$. The change of variable $\xi = s\omega$ thus proves that

$$\hat{b}(\omega) = \frac{1}{C_\psi} \hat{f}(\omega) \int_0^{+\infty} \frac{|\hat{\psi}(\xi)|^2}{\xi} d\xi = \hat{f}(\omega).$$

■

The theorem hypothesis

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

is called the wavelet *admissibility condition*. To guarantee that this integral is finite we must ensure that $\hat{\psi}(0) = 0$, which explains why we imposed that wavelets must have a zero average. This condition is nearly sufficient. If $\hat{\psi}(0) = 0$ and $\hat{\psi}(\omega)$ is continuously differentiable then the admissibility condition is satisfied. One can verify that $\hat{\psi}(\omega)$ is continuously differentiable if ψ has a sufficient time decay

$$\int_{-\infty}^{+\infty} (1 + |t|) \, |\psi(t)| \, dt < +\infty.$$

Reproducing Kernel Like a windowed Fourier transform, a wavelet transform is a redundant representation, whose redundancy is characterized by a reproducing kernel equation. Inserting the reconstruction formula (4.37) into the definition of the wavelet transform yields

$$Wf(u_0,s_0) = \int_{-\infty}^{+\infty} \left(\frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} Wf(u,s) \psi_{u,s}(t) \, du \frac{ds}{s^2} \right) \psi^*_{u_0,s_0}(t) \, dt.$$

Interchanging these integrals gives

$$Wf(u_0,s_0) = \frac{1}{C_\psi}\int_{-\infty}^{+\infty} K(u,u_0,s,s_0)\, Wf(u,s)\, du\, \frac{ds}{s^2}, \tag{4.40}$$

with

$$K(u_0,u,s_0,s) = \langle \psi_{u,s}, \psi_{u_0,s_0} \rangle . \tag{4.41}$$

The reproducing kernel $K(u_0,u,s_0,s)$ measures the correlation of two wavelets $\psi_{u,s}$ and ψ_{u_0,s_0}. The reader can verify that any function $\Phi(u,s)$ is the wavelet transform of some $f \in \mathbf{L}^2(\mathbb{R})$ if and only if it satisfies the reproducing kernel equation (4.40).

Scaling Function When $Wf(u,s)$ is known only for $s < s_0$, to recover f we need a complement of information corresponding to $Wf(u,s)$ for $s > s_0$. This is obtained by introducing a *scaling function* ϕ that is an aggregation of wavelets at scales larger than 1. The modulus of its Fourier transform is defined by

$$|\hat{\phi}(\omega)|^2 = \int_1^{+\infty} |\hat{\psi}(s\omega)|^2 \frac{ds}{s} = \int_\omega^{+\infty} \frac{|\hat{\psi}(\xi)|^2}{\xi}\, d\xi, \tag{4.42}$$

and the complex phase of $\hat{\phi}(\omega)$ can be arbitrarily chosen. One can verify that $\|\phi\| = 1$ and we derive from the admissibility condition (4.36) that

$$\lim_{\omega \to 0} |\hat{\phi}(\omega)|^2 = C_\psi. \tag{4.43}$$

The scaling function can thus be interpreted as the impulse response of a low-pass filter. Let us denote

$$\phi_s(t) = \frac{1}{\sqrt{s}}\phi\left(\frac{t}{s}\right) \quad \text{and} \quad \bar{\phi}_s(t) = \phi_s^*(-t).$$

The low-frequency approximation of f at the scale s is

$$Lf(u,s) = \left\langle f(t), \frac{1}{\sqrt{s}}\phi\left(\frac{t-u}{s}\right)\right\rangle = f \star \bar{\phi}_s(u). \tag{4.44}$$

With a minor modification of the proof of Theorem 4.3, it can be shown that

$$f(t) = \frac{1}{C_\psi}\int_0^{s_0} Wf(.,s) \star \psi_s(t)\, \frac{ds}{s^2} + \frac{1}{C_\psi s_0} Lf(.,s_0) \star \phi_{s_0}(t). \tag{4.45}$$

Example 4.7 If ψ is the second order derivative of a Gaussian whose Fourier transform is given by (4.35), then the integration (4.42) yields

$$\hat{\phi}(\omega) = \frac{2\sigma^{3/2}\pi^{1/4}}{\sqrt{3}}\sqrt{\omega^2 + \frac{1}{\sigma^2}}\, \exp\left(-\frac{\sigma^2\omega^2}{2}\right). \tag{4.46}$$

Figure 4.8 displays ϕ and $\hat{\phi}$ for $\sigma = 1$.

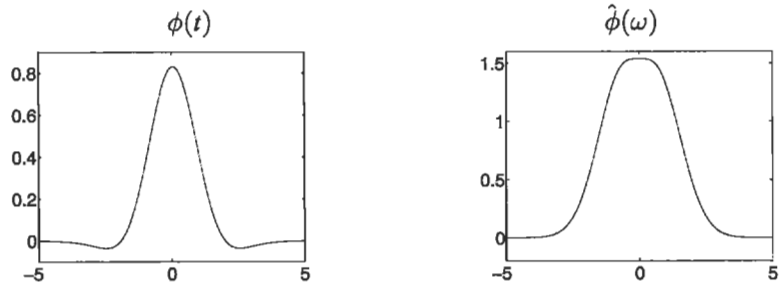


FIGURE 4.8 Scaling function associated to a Mexican hat wavelet and its Fourier transform calculated with (4.46).

4.3.2 Analytic Wavelets

To analyze the time evolution of frequency tones, it is necessary to use an analytic wavelet to separate the phase and amplitude information of signals. The properties of the resulting analytic wavelet transform are studied.

Analytic Signal A function $f_a \in \mathbf{L}^2(\mathbb{R})$ is said to be *analytic* if its Fourier transform is zero for negative frequencies:

$$\hat{f}_a(\omega) = 0 \text{ if } \omega < 0.$$

An analytic function is necessarily complex but is entirely characterized by its real part. Indeed, the Fourier transform of its real part $f = \text{Real}[f_a]$ is

$$\hat{f}(\omega) = \frac{\hat{f}_a(\omega) + \hat{f}_a^*(-\omega)}{2},$$

and this relation can be inverted:

$$\hat{f}_a(\omega) = \begin{cases} 2\hat{f}(\omega) & \text{if } \omega \geq 0 \\ 0 & \text{if } \omega < 0 \end{cases}. \tag{4.47}$$

The analytic part $f_a(t)$ of a signal $f(t)$ is the inverse Fourier transform of $\hat{f}_a(\omega)$ defined by (4.47).

Discrete Analytic Part The analytic part $f_a[n]$ of a discrete signal $f[n]$ of size N is also computed by setting to zero the negative frequency components of its discrete Fourier transform. The Fourier transform values at $k = 0$ and $k = N/2$ must be carefully adjusted so that $\text{Real}[f_a] = f$:

$$\hat{f}_a[k] = \begin{cases} \hat{f}[k] & \text{if } k = 0, N/2 \\ 2\hat{f}[k] & \text{if } 0 < k < N/2 \\ 0 & \text{if } N/2 < k < N \end{cases}. \tag{4.48}$$

We obtain $f_a[n]$ by computing the inverse discrete Fourier transform.

Example 4.8 The Fourier transform of

$$f(t) = a\cos(\omega_0 t + \phi) = \frac{a}{2}\Big(\exp[i(\omega_0 t + \phi)] + \exp[-i(\omega_0 t + \phi)]\Big)$$

is

$$\hat{f}(\omega) = \pi a\Big(\exp(i\phi)\,\delta(\omega - \omega_0) + \exp(-i\phi)\,\delta(\omega + \omega_0)\Big).$$

The Fourier transform of the analytic part computed with (4.47) is $\hat{f}_a(\omega) = 2\pi a \exp(i\phi)\,\delta(\omega - \omega_0)$ and hence

$$f_a(t) = a\exp[i(\omega_0 t + \phi)]. \tag{4.49}$$

Time-Frequency Resolution An analytic wavelet transform is calculated with an analytic wavelet ψ:

$$Wf(u,s) = \langle f, \psi_{u,s}\rangle = \int_{-\infty}^{+\infty} f(t)\,\frac{1}{\sqrt{s}}\,\psi^*\left(\frac{t-u}{s}\right) dt. \tag{4.50}$$

Its time-frequency resolution depends on the time-frequency spread of the wavelet atoms $\psi_{u,s}$. We suppose that ψ is centered at 0, which implies that $\psi_{u,s}$ is centered at $t = u$. With the change of variable $v = \frac{t-u}{s}$, we verify that

$$\int_{-\infty}^{+\infty} (t-u)^2\,|\psi_{u,s}(t)|^2\,dt = s^2\,\sigma_t^2, \tag{4.51}$$

with $\sigma_t^2 = \int_{-\infty}^{+\infty} t^2\,|\psi(t)|^2\,dt$. Since $\hat{\psi}(\omega)$ is zero at negative frequencies, the center frequency η of $\hat{\psi}$ is

$$\eta = \frac{1}{2\pi}\int_0^{+\infty} \omega\,|\hat{\psi}(\omega)|^2\,d\omega. \tag{4.52}$$

The Fourier transform of $\psi_{u,s}$ is a dilation of $\hat{\psi}$ by $1/s$:

$$\hat{\psi}_{u,s}(\omega) = \sqrt{s}\,\hat{\psi}(s\omega)\exp(-i\omega u). \tag{4.53}$$

Its center frequency is therefore η/s. The energy spread of $\hat{\psi}_{u,s}$ around η/s is

$$\frac{1}{2\pi}\int_0^{+\infty} \left(\omega - \frac{\eta}{s}\right)^2 \left|\hat{\psi}_{u,s}(\omega)\right|^2 d\omega = \frac{\sigma_\omega^2}{s^2}, \tag{4.54}$$

with

$$\sigma_\omega^2 = \frac{1}{2\pi}\int_0^{+\infty} (\omega - \eta)^2\,|\hat{\psi}(\omega)|^2\,d\omega.$$

The energy spread of a wavelet time-frequency atom $\psi_{u,s}$ thus corresponds to a Heisenberg box centered at $(u, \eta/s)$, of size $s\sigma_t$ along time and σ_ω/s along frequency. The area of the rectangle remains equal to $\sigma_t\,\sigma_\omega$ at all scales but the resolution in time and frequency depends on s, as illustrated in Figure 4.9.

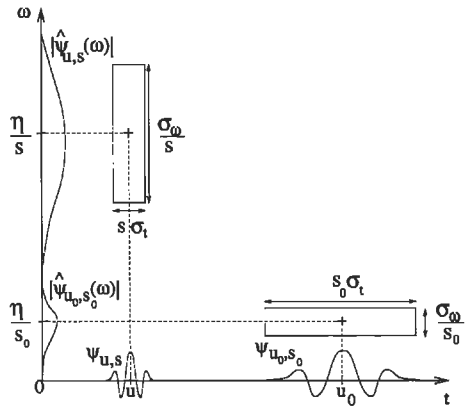


FIGURE 4.9 Heisenberg boxes of two wavelets. Smaller scales decrease the time spread but increase the frequency support, which is shifted towards higher frequencies.

An analytic wavelet transform defines a local time-frequency energy density $P_W f$, which measures the energy of f in the Heisenberg box of each wavelet $\psi_{u,s}$ centered at $(u, \xi = \eta/s)$:

$$P_W f(u,\xi) = |Wf(u,s)|^2 = \left|Wf\left(u, \frac{\eta}{\xi}\right)\right|^2. \tag{4.55}$$

This energy density is called a *scalogram.*

Completeness An analytic wavelet transform of f depends only on its analytic part f_a. The following theorem derives a reconstruction formula and proves that energy is conserved for real signals.

Theorem 4.4 *For any* $f \in \mathbf{L}^2(\mathbb{R})$

$$Wf(u,s) = \frac{1}{2} Wf_a(u,s). \tag{4.56}$$

If $C_\psi = \int_0^{+\infty} \omega^{-1} |\hat{\psi}(\omega)|^2 \, d\omega < +\infty$ *and* f *is real then*

$$f(t) = \frac{2}{C_\psi} \text{Real}\left[\int_0^{+\infty} \int_{-\infty}^{+\infty} Wf(u,s)\, \psi_s(t-u)\, du \frac{ds}{s^2}\right], \tag{4.57}$$

and

$$\|f\|^2 = \frac{2}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} |Wf(u,s)|^2 \, du \frac{ds}{s^2}. \tag{4.58}$$

Proof[1]. Let us first prove (4.56). The Fourier transform with respect to u of

$$f_s(u) = Wf(u,s) = f \star \bar{\psi}_s(u)$$

is

$$\hat{f}_s(\omega) = \hat{f}(\omega)\sqrt{s}\hat{\psi}^*(s\omega).$$

Since $\hat{\psi}(\omega) = 0$ at negative frequencies, and $\hat{f}_a(\omega) = 2\hat{f}(\omega)$ for $\omega \geq 0$, we derive that

$$\hat{f}_s(\omega) = \frac{1}{2}\hat{f}_a(\omega)\sqrt{s}\hat{\psi}^*(s\omega),$$

which is the Fourier transform of (4.56).

With the same derivations as in the proof of (4.37) one can verify that the inverse wavelet formula reconstructs the analytic part of f:

$$f_a(t) = \frac{1}{C_\psi}\int_0^{+\infty}\int_{-\infty}^{+\infty} Wf_a(u,s)\,\psi_s(t-u)\,\frac{ds}{s^2}\,du. \tag{4.59}$$

Since $f = \text{Real}[f_a]$, inserting (4.56) proves (4.57).

An energy conservation for the analytic part f_a is proved as in (4.38) by applying the Plancherel formula:

$$\int_{-\infty}^{+\infty} |f_a(t)|^2\,dt = \frac{1}{C_\psi}\int_0^{+\infty}\int_{-\infty}^{+\infty} |W_a f(u,s)|^2\,du\,\frac{ds}{s^2}.$$

Since $Wf_a(u,s) = 2Wf(u,s)$ and $\|f_a\|^2 = 2\|f\|^2$, equation (4.58) follows. ■

If f is real the change of variable $\xi = 1/s$ in the energy conservation (4.58) proves that

$$\|f\|^2 = \frac{2}{C_\psi}\int_0^{+\infty}\int_{-\infty}^{+\infty} P_W f(u,\xi)\,du\,d\xi.$$

It justifies the interpretation of a scalogram as a time-frequency energy density.

Wavelet Modulated Windows An analytic wavelet can be constructed with a frequency modulation of a real and symmetric window g. The Fourier transform of

$$\psi(t) = g(t)\exp(i\eta t) \tag{4.60}$$

is $\hat{\psi}(\omega) = \hat{g}(\omega - \eta)$. If $\hat{g}(\omega) = 0$ for $|\omega| > \eta$ then $\hat{\psi}(\omega) = 0$ for $\omega < 0$. Hence ψ is analytic, as shown in Figure 4.10. Since g is real and even, $\hat{g}$ is also real and symmetric. The center frequency of $\hat{\psi}$ is therefore η and

$$|\hat{\psi}(\eta)| = \sup_{\omega\in\mathbf{R}} |\hat{\psi}(\omega)| = \hat{g}(0). \tag{4.61}$$

A *Gabor wavelet* $\psi(t) = g(t)\,e^{i\eta t}$ is obtained with a Gaussian window

$$g(t) = \frac{1}{(\sigma^2\pi)^{1/4}}\exp\left(\frac{-t^2}{2\sigma^2}\right). \tag{4.62}$$

The Fourier transform of this window is $\hat{g}(\omega) = (4\pi\sigma^2)^{1/4}\exp(-\sigma^2\omega^2/2)$. If $\sigma^2\eta^2 \gg 1$ then $\hat{g}(\omega) \approx 0$ for $|\omega| > \eta$. Such Gabor wavelets are thus considered to be approximately analytic.

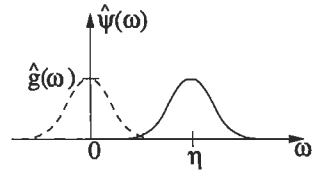


FIGURE 4.10 Fourier transform $\hat{\psi}(\omega)$ of a wavelet $\psi(t) = g(t)\exp(i\eta t)$.

Example 4.9 The wavelet transform of $f(t) = a\exp(i\omega_0 t)$ is

$$Wf(u,s) = a\sqrt{s}\,\hat{\psi}^*(s\omega_0)\exp(i\omega_0 t) = a\sqrt{s}\,\hat{g}(s\omega_0 - \eta)\exp(i\omega_0 t).$$

Observe that the normalized scalogram is maximum at $\xi = \omega_0$:

$$\frac{\xi}{\eta}P_W f(u,\xi) = \frac{1}{s}|Wf(u,s)|^2 = a^2\left|\hat{g}\left(\eta\left(\frac{\omega_0}{\xi} - 1\right)\right)\right|^2.$$

Example 4.10 The wavelet transform of a linear chirp $f(t) = \exp(iat^2) = \exp[i\phi(t)]$ is computed for a Gabor wavelet whose Gaussian window is (4.62). By using the Fourier transform of Gaussian chirps (2.34) one can verify that

$$\frac{|Wf(u,s)|^2}{s} = \left(\frac{4\pi\sigma^2}{1+4s^2a^2\sigma^4}\right)^{1/2}\exp\left(\frac{-\sigma^2}{1+4a^2s^4\sigma^4}(\eta - 2asu)^2\right).$$

As long as $4a^2s^4\sigma^4 \ll 1$, at a fixed time u the renormalized scalogram $\eta^{-1}\xi P_W f(u,\xi)$ is a Gaussian function of s that reaches its maximum at

$$\xi(u) = \frac{\eta}{s(u)} = \phi'(u) = 2au. \tag{4.63}$$

Section 4.4.2 explains why the amplitude is maximum at the instantaneous frequency $\phi'(u)$.

Example 4.11 Figure 4.11 displays the normalized scalogram $\eta^{-1}\xi P_W f(u,\xi)$, and the complex phase $\Phi_W(u,\xi)$ of $Wf(u,s)$, for the signal f of Figure 4.3. The frequency bandwidth of wavelet atoms is proportional to $1/s = \xi/\eta$. The frequency resolution of the scalogram is therefore finer than the spectrogram at low frequencies but coarser than the spectrogram at higher frequencies. This explains why the wavelet transform produces interference patterns between the high frequency Gabor function at the abscissa $t = 0.87$ and the quadratic chirp at the same location, whereas the spectrogram in Figure 4.3 separates them well.

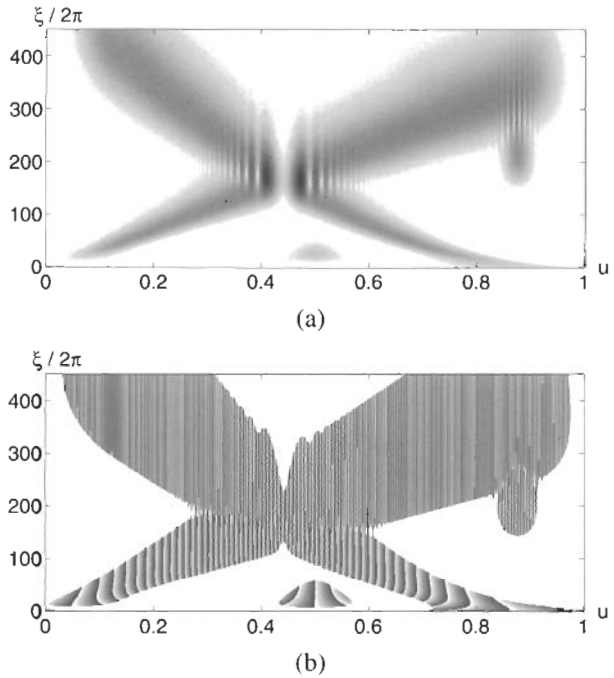


FIGURE 4.11 (a) Normalized scalogram $\eta^{-1}\xi P_W f(u,\xi)$ computed from the signal in Figure 4.3. Dark points indicate large amplitude coefficients. (b) Complex phase $\Phi_W(u,\xi)$ of $Wf(u,\eta/\xi)$, where the modulus is non-zero.

4.3.3 Discrete Wavelets [2]

Let $\dot{f}(t)$ be a continuous time signal that is uniformly sampled at intervals N^{-1} over $[0,1]$. Its wavelet transform can only be calculated at scales $N^{-1} < s < 1$, as shown in Figure 4.7. In discrete computations, it is easier to normalize the sampling distance to 1 and thus consider the dilated signal $f(t) = \dot{f}(N^{-1}t)$. A change of variable in the wavelet transform integral (4.31) proves that

$$W\dot{f}(u,s) = N^{-1/2} Wf(Nu,Ns) .$$

To simplify notation, we concentrate on f and denote $f[n] = f(n)$ the discrete signal of size N. Its discrete wavelet transform is computed at scales $s = a^j$, with $a = 2^{1/v}$, which provides v intermediate scales in each octave $[2^j, 2^{j+1})$.

Let $\psi(t)$ be a wavelet whose support is included in $[-K/2, K/2]$. For $2 \leq a^j \leq NK^{-1}$, a discrete wavelet scaled by a^j is defined by

$$\psi_j[n] = \frac{1}{\sqrt{a^j}} \psi\left(\frac{n}{a^j}\right) .$$

This discrete wavelet has Ka^j non-zero values on $[-N/2, N/2]$. The scale a^j is larger than 2 otherwise the sampling interval may be larger than the wavelet support.

Fast Transform To avoid border problems, we treat $f[n]$ and the wavelets $\psi_j[n]$ as periodic signals of period N. The discrete wavelet transform can then be written as a circular convolution $\bar{\psi}_j[n] = \psi_j^*[-n]$:

$$Wf[n, a^j] = \sum_{m=0}^{N-1} f[m]\, \psi_j^*[m-n] = f \circledast \bar{\psi}_j[n]. \tag{4.64}$$

This circular convolution is calculated with the fast Fourier transform algorithm, which requires $O(N \log_2 N)$ operations. If $a = 2^{1/v}$, there are $v \log_2(N/(2K))$ scales $a^j \in [2N^{-1}, K^{-1}]$. The total number of operations to compute the wavelet transform over all scales is therefore $O(vN(\log_2 N)^2)$ [291].

To compute the scalogram $P_W[n, \xi] = |Wf[n, \frac{\eta}{\xi}]|^2$ we calculate $Wf[n, s]$ at any scale s with a parabola interpolation. Let j be the closest integer to $\log_2 s / \log_2 a$, and $p(x)$ be the parabola such that

$$p(j-1) = Wf[n, a^{j-1}] \ , \ p(j) = Wf[n, a^j] \ , \ p(j+1) = Wf[n, a^{j+1}].$$

A second order interpolation computes

$$Wf[n, s] = p\left(\frac{\log_2 s}{\log_2 a}\right) .$$

Parabolic interpolations are used instead of linear interpolations in order to locate more precisely the ridges defined in Section 4.4.2.

Discrete Scaling Filter A wavelet transform computed up to a scale a^J is not a complete signal representation. It is necessary to add the low frequencies $Lf[n, a^J]$ corresponding to scales larger than a^J. A discrete and periodic scaling filter is computed by sampling the scaling function $\phi(t)$ defined in (4.42):

$$\phi_J[n] = \frac{1}{\sqrt{a^J}} \phi\left(\frac{n}{a^J}\right) \quad \text{for } n \in [-N/2, N/2].$$

Let $\bar{\phi}_J[n] = \phi_J^*[-n]$. The low frequencies are carried by

$$Lf[n, a^J] = \sum_{m=0}^{N-1} f[m]\, \phi_J^*[m-n] = f \circledast \bar{\phi}_J[n]. \tag{4.65}$$

Reconstruction An inverse wavelet transform is implemented by discretizing the integral (4.45). Suppose that $a^J = 2$ is the finest scale. Since $ds/s^2 = d \log_e s / s$ and

the discrete wavelet transform is computed along an exponential scale sequence $\{a^j\}_j$ with a logarithmic increment $d\log_e s = \log_e a$, we obtain

$$f[n] \approx \frac{\log_e a}{C_\psi} \sum_{j=I}^{J} \frac{1}{a^j} Wf[.,a^j] \circledast \psi_j[n] + \frac{1}{C_\psi a^J} Lf[.,a^J] \circledast \phi_J[n]. \tag{4.66}$$

The "." indicates the variable over which the convolution is calculated. These circular convolutions are calculated using the FFT, with $O(vN(\log_2 N)^2)$ operations.

Analytic wavelet transforms are often computed over real signals $f[n]$ that have no energy at low frequencies. In this case do not use a scaling filter $\phi_J[n]$. Theorem 4.4 shows that

$$f[n] \approx \frac{2\log_e a}{C_\psi} \text{Real} \left(\sum_{j=I}^{J} \frac{1}{a^j} Wf[.,a^j] \circledast \psi_j[n] \right). \tag{4.67}$$

The error introduced by the discretization of scales decreases when the number v of voices per octave increases. However, the approximation of continuous time convolutions with discrete convolutions also creates high frequency errors. Perfect reconstructions can be obtained with a more careful design of the reconstruction filters. Section 5.5.2 describes an exact inverse wavelet transform computed at dyadic scales $a^j = 2^j$.

4.4 INSTANTANEOUS FREQUENCY [2]

When listening to music, we perceive several frequencies that change with time. This notion of instantaneous frequency remains to be defined. The time variation of several instantaneous frequencies can be measured with time-frequency decompositions, and in particular with windowed Fourier transforms and wavelet transforms.

Analytic Instantaneous Frequency A cosine modulation

$$f(t) = a\cos(\omega_0 t + \phi_0) = a\cos\phi(t)$$

has a frequency ω_0 that is the derivative of the phase $\phi(t) = \omega_0 t + \phi_0$. To generalize this notion, real signals f are written as an amplitude a modulated with a time varying phase ϕ:

$$f(t) = a(t)\cos\phi(t) \quad \text{with} \quad a(t) \geq 0\,. \tag{4.68}$$

The *instantaneous frequency* is defined as a positive derivative of the phase:

$$\omega(t) = \phi'(t) \geq 0\,.$$

The derivative can be chosen to be positive by adapting the sign of $\phi(t)$. One must be careful because there are many possible choices of $a(t)$ and $\phi(t)$, which implies that $\omega(t)$ is not uniquely defined relative to f.

A particular decomposition (4.68) is obtained from the analytic part f_a of f, whose Fourier transform is defined in (4.47) by

$$\hat{f}_a(\omega) = \begin{cases} 2\hat{f}(\omega) & \text{if } \omega \geq 0 \\ 0 & \text{if } \omega < 0 \end{cases} . \tag{4.69}$$

This complex signal is represented by separating the modulus and the complex phase:

$$f_a(t) = a(t) \exp[i\phi(t)] . \tag{4.70}$$

Since $f = \text{Real}[f_a]$, it follows that

$$f(t) = a(t) \cos\phi(t).$$

We call $a(t)$ the *analytic* amplitude of $f(t)$ and $\phi'(t)$ its *instantaneous frequency*; they are uniquely defined.

Example 4.12 If $f(t) = a(t)\cos(\omega_0 t + \phi_0)$, then

$$\hat{f}(\omega) = \frac{1}{2}\left(\exp(i\phi_0)\,\hat{a}(\omega-\omega_0) + \exp(-i\phi_0)\,\hat{a}(\omega+\omega_0)\right) .$$

If the variations of $a(t)$ are slow compared to the period $2\pi/\omega_0$, which is achieved by requiring that the support of $\hat{a}$ be included in $[-\omega_0, \omega_0]$, then

$$\hat{f}_a(\omega) = \hat{a}(\omega-\omega_0)\exp(i\phi_0)$$

so $f_a(t) = a(t)\exp[i(\omega_0 t + \phi_0)]$.

If a signal f is the sum of two sinusoidal waves:

$$f(t) = a\cos(\omega_1 t) + a\cos(\omega_2 t),$$

then

$$f_a(t) = a\exp(i\omega_1 t) + a\exp(i\omega_2 t) = a\cos\left(\frac{1}{2}(\omega_1-\omega_2)t\right)\exp\left(\frac{i}{2}(\omega_1+\omega_2)t\right).$$

The instantaneous frequency is $\phi'(t) = (\omega_1+\omega_2)/2$ and the amplitude is

$$a(t) = a\left|\cos\left(\frac{1}{2}(\omega_1-\omega_2)t\right)\right| .$$

This result is not satisfying because it does not reveal that the signal includes two sinusoidal waves of the same amplitude. It measures an average frequency value. The next sections explain how to measure the instantaneous frequencies of several spectral components by separating them with a windowed Fourier transform or a wavelet transform. We first describe two important applications of instantaneous frequencies.

Frequency Modulation In signal communications, information can be transmitted through the amplitude $a(t)$ (amplitude modulation) or the instantaneous frequency $\phi'(t)$ (frequency modulation) [65]. Frequency modulation is more robust in the presence of additive Gaussian white noise. In addition, it better resists multi-path interferences, which destroy the amplitude information. A frequency modulation sends a message $m(t)$ through a signal

$$f(t) = a\cos\phi(t) \quad \text{with} \quad \phi'(t) = \omega_0 + k m(t).$$

The frequency bandwidth of f is proportional to k. This constant is adjusted depending on the transmission noise and the available bandwidth. At the reception, the message $m(t)$ is restored with a frequency demodulation that computes the instantaneous frequency $\phi'(t)$ [101].

Additive Sound Models Musical sounds and voiced speech segments can be modeled with sums of sinusoidal *partials*:

$$f(t) = \sum_{k=1}^{K} f_k(t) = \sum_{k=1}^{K} a_k(t)\cos\phi_k(t)\ , \tag{4.71}$$

where a_k and ϕ_k' vary slowly [296, 297]. Such decompositions are useful for pattern recognition and for modifying sound properties [245]. Sections 4.4.1 and 4.4.2 explain how to compute a_k and the instantaneous frequency ϕ_k' of each partial, from which the phase ϕ_k is derived by integration.

To compress the sound f by a factor α in time, without modifying the values of ϕ_k' and a_k, we synthesize

$$g(t) = \sum_{k=1}^{K} a_k(\alpha t)\cos\Big(\frac{1}{\alpha}\phi_k(\alpha t)\Big). \tag{4.72}$$

The partials of g at $t = \alpha t_0$ and the partials of f at $t = t_0$ have the same amplitudes and instantaneous frequencies. If $\alpha > 1$, the sound g is shorter but it is perceived as having the same "frequency content" as f.

A frequency transposition is calculated by multiplying each phase by a constant α:

$$g(t) = \sum_{k=1}^{K} b_k(t)\cos\Big(\alpha\phi_k(t)\Big). \tag{4.73}$$

The instantaneous frequency of each partial is now $\alpha\phi_k'(t)$. To compute new amplitudes $b_k(t)$, we use a resonance model, which supposes that these amplitudes are samples of a smooth frequency envelope $F(t,\omega)$:

$$a_k(t) = F\Big(t, \phi_k'(t)\Big) \ \text{ and } \ b_k(t) = F\Big(t, \alpha\phi_k'(t)\Big)\ .$$

This envelope is called a *formant* in speech processing. It depends on the type of phoneme that is pronounced. Since $F(t,\omega)$ is a regular function of ω, its amplitude at $\omega = \alpha\phi_k'(t)$ is calculated by interpolating the values $a_k(t)$ corresponding to $\omega = \phi_k'(t)$.

4.4.1 Windowed Fourier Ridges

The spectrogram $P_S f(u,\xi) = |Sf(u,\xi)|^2$ measures the energy of f in a time-frequency neighborhood of (u,ξ). The ridge algorithm computes instantaneous frequencies from the local maxima of $P_S f(u,\xi)$. This approach was introduced by Delprat, Escudié, Guillemain, Kronland-Martinet, Tchamitchian and Torrésani [154, 71] to analyze musical sounds. Since then it has found applications for a wide range of signals [201, 71] that have time varying frequency tones.

The windowed Fourier transform is computed with a symmetric window $g(t) = g(-t)$ whose support is equal to $[-1/2, 1/2]$. The Fourier transform $\hat{g}$ is a real symmetric function and $|\hat{g}(\omega)| \leq \hat{g}(0)$ for all $\omega \in \mathbb{R}$. The maximum $\hat{g}(0) = \int_{-1/2}^{1/2} g(t)\,dt$ is on the order of 1. Table 4.1 gives several examples of such windows. The window g is normalized so that $\|g\| = 1$. For a fixed scale s, $g_s(t) = s^{-1/2} g(t/s)$ has a support of size s and a unit norm. The corresponding windowed Fourier atoms are

$$g_{s,u,\xi}(t) = g_s(t-u)\,e^{i\xi t},$$

and the windowed Fourier transform is defined by

$$Sf(u,\xi) = \langle f, g_{s,u,\xi}\rangle = \int_{-\infty}^{+\infty} f(t)\, g_s(t-u)\, e^{-i\xi t}\,dt. \tag{4.74}$$

The following theorem relates $Sf(u,\xi)$ to the instantaneous frequency of f.

Theorem 4.5 *Let $f(t) = a(t)\cos\phi(t)$. If $\xi \geq 0$ then*

$$\langle f, g_{s,u,\xi}\rangle = \frac{\sqrt{s}}{2} a(u) \exp(i[\phi(u) - \xi u]) \left(\hat{g}(s[\xi - \phi'(u)]) + \epsilon(u,\xi)\right). \tag{4.75}$$

The corrective term satisfies

$$|\epsilon(u,\xi)| \leq \epsilon_{a,1} + \epsilon_{a,2} + \epsilon_{\phi,2} + \sup_{|\omega| \geq s\phi'(u)} |\hat{g}(\omega)| \tag{4.76}$$

with

$$\epsilon_{a,1} \leq \frac{s\,|a'(u)|}{|a(u)|}\ ,\ \epsilon_{a,2} \leq \sup_{|t-u| \leq s/2} \frac{s^2\,|a''(t)|}{|a(u)|}\ , \tag{4.77}$$

and if $s\,|a'(u)|\,|a(u)|^{-1} \leq 1$, then

$$\epsilon_{\phi,2} \leq \sup_{|t-u| \leq s/2} s^2 |\phi''(t)|\,. \tag{4.78}$$

If $\xi = \phi'(u)$ *then*

$$\epsilon_{a,1} = \frac{s|a'(u)|}{|a(u)|} \left| \hat{g}'\Big(2s\phi'(u)\Big) \right| . \tag{4.79}$$

Proof[2]. Observe that

$$\begin{aligned} \langle f, g_{s,u,\xi} \rangle &= \int_{-\infty}^{+\infty} a(t) \cos\phi(t)\, g_s(t-u) \exp(-i\xi t)\, dt \\ &= \frac{1}{2}\int_{-\infty}^{+\infty} a(t)\, (\exp[i\phi(t)] + \exp[-i\phi(t)])\, g_s(t-u) \exp[-i\xi t]\, dt \\ &= I(\phi) + I(-\phi). \end{aligned}$$

We first concentrate on

$$\begin{aligned} I(\phi) &= \frac{1}{2}\int_{-\infty}^{+\infty} a(t) \exp[i\phi(t)]\, g_s(t-u) \exp(-i\xi t)\, dt \\ &= \frac{1}{2}\int_{-\infty}^{+\infty} a(t+u) e^{i\phi(t+u)}\, g_s(t) \exp[-i\xi(t+u)]\, dt. \end{aligned}$$

This integral is computed by using second order Taylor expansions:

$$\begin{aligned} a(t+u) &= a(u) + t a'(u) + \frac{t^2}{2}\alpha(t) \text{ with } |\alpha(t)| \le \sup_{h\in[u,t+u]} |a''(h)| \\ \phi(t+u) &= \phi(u) + t\phi'(u) + \frac{t^2}{2}\beta(t) \text{ with } |\beta(t)| \le \sup_{h\in[u,t+u]} |\phi''(h)| . \end{aligned}$$

We get

$$\begin{aligned} 2 \quad & \exp\Big(-i(\phi(u) - \xi u)\Big) I(\phi) = \\ & \int_{-\infty}^{+\infty} a(u)\, g_s(t) \exp\Big(-it(\xi - \phi'(u))\Big) \exp\Big(i\frac{t^2}{2}\beta(t)\Big)\, dt \\ & + \int_{-\infty}^{+\infty} a'(u)\, t\, g_s(t) \exp\Big(-it(\xi - \phi'(u))\Big) \exp\Big(i\frac{t^2}{2}\beta(t)\Big)\, dt \\ & + \frac{1}{2}\int_{-\infty}^{+\infty} \alpha(t) t^2\, g_s(t) \exp\Big(-i(t\xi + \phi(u) - \phi(t+u))\Big)\, dt . \end{aligned}$$

A first order Taylor expansion of $\exp(ix)$ gives

$$\exp\Big(i\frac{t^2}{2}\beta(t)\Big) = 1 + \frac{t^2}{2}\beta(t)\gamma(t) \quad \text{with} \quad |\gamma(t)| \le 1 . \tag{4.80}$$

Since

$$\int_{-\infty}^{+\infty} g_s(t) \exp[-it(\xi - \phi'(u))]\, dt = \sqrt{s}\, \hat{g}(s[\xi - \phi'(u)]),$$

inserting (4.80) in the expression of $I(\phi)$ yields

$$\left| I(\phi) - \frac{\sqrt{s}}{2} a(u) \exp[i(\phi(u) - \xi u)]\, \hat{g}(\xi - \phi'(u)) \right| \le \frac{\sqrt{s}|a(u)|}{4} (\epsilon_{a,1}^{+} + \epsilon_{a,2} + \epsilon_{\phi,2}) \tag{4.81}$$

with

$$\epsilon^+_{a,1} = \frac{2|a'(u)|}{|a(u)|}\left|\int_{-\infty}^{+\infty} t\,\frac{1}{\sqrt{s}}\,g_s(t)\exp[-it(\xi-\phi'(u))]\,dt\right|, \tag{4.82}$$

$$\epsilon_{a,2} = \int_{-\infty}^{+\infty} t^2|\alpha(t)|\,\frac{1}{\sqrt{s}}\,|g_s(t)|\,dt\,, \tag{4.83}$$

$$\epsilon_{\phi,2} = \int_{-\infty}^{+\infty} t^2|\beta(t)|\,\frac{1}{\sqrt{s}}\,|g_s(t)|\,dt + \frac{|a'(u)|}{|a(u)|}\int_{-\infty}^{+\infty}|t^3||\beta(t)|\,\frac{1}{\sqrt{s}}\,|g_s(t)|dt. \tag{4.84}$$

Applying (4.81) to $I(-\phi)$ gives

$$|I(-\phi)| \le \frac{\sqrt{s}\,|a(u)|}{2}\,|\hat{g}(\xi+\phi'(u))| + \frac{\sqrt{s}\,|a(u)|}{4}(\epsilon^-_{a,1}+\epsilon_{a,2}+\epsilon_{\phi,2})\,,$$

with

$$\epsilon^-_{a,1} = \frac{2|a'(u)|}{|a(u)|}\left|\int_{-\infty}^{+\infty} t\,\frac{1}{\sqrt{s}}\,g_s(t)\exp[-it(\xi+\phi'(u))]\,dt\right|. \tag{4.85}$$

Since $\xi \ge 0$ and $\phi'(u) \ge 0$, we derive that

$$|\hat{g}(s[\xi+\phi'(u)])| \le \sup_{|\omega|\ge s\phi'(u)} |\hat{g}(\omega)|\,,$$

and hence

$$I(\phi)+I(-\phi) = \frac{\sqrt{s}}{2}\,a(u)\exp[i(\phi(u)-\xi u)]\left(\hat{g}\left(s[\xi-\phi'(u)]\right)+\epsilon(u,\xi)\right)$$

with

$$\epsilon(u,\xi) = \frac{\epsilon^+_{a,1}+\epsilon^-_{a,1}}{2}+\epsilon_{a,2}+\epsilon_{\phi,2}+\sup_{|\omega|\ge s|\phi'(u)|}|\hat{g}(\omega)|\,.$$

Let us now verify the upper bound (4.77) for $\epsilon_{a,1} = (\epsilon^+_{a,1}+\epsilon^-_{a,1})/2$. Since $g_s(t) = s^{-1/2}g(t/s)$, a simple calculation shows that for $n \ge 0$

$$\int_{-\infty}^{+\infty}|t|^n\,\frac{1}{\sqrt{s}}\,|g_s(t)|\,dt = s^n\int_{-1/2}^{1/2}|t|^n\,|g(t)|\,dt \le \frac{s^n}{2^n}\,\|g\|^2 = \frac{s^n}{2^n}. \tag{4.86}$$

Inserting this for $n=1$ in (4.82) and (4.85) gives

$$\epsilon_{a,1} = \frac{\epsilon^+_{a,1}+\epsilon^-_{a,1}}{2} \le \frac{s\,|a'(u)|}{|a(u)|}\,.$$

The upper bounds (4.77) and (4.78) of the second order terms $\epsilon_{a,2}$ and $\epsilon_{\phi,2}$ are obtained by observing that the remainder $\alpha(t)$ and $\beta(t)$ of the Taylor expansion of $a(t+u)$ and $\phi(t+u)$ satisfy

$$\sup_{|t|\le s/2}|\alpha(t)| \le \sup_{|t-u|\le s/2}|a''(t)|\,,\quad \sup_{|t|\le s/2}|\beta(t)| \le \sup_{|t-u|\le s/2}|\phi''(t)|. \tag{4.87}$$

Inserting this in (4.83) yields

$$\epsilon_{a,2} \leq \sup_{|t-u|\leq s/2} \frac{s^2\,|a''(t)|}{|a(u)|}\,.$$

When $s\,|a'(u)||a(u)|^{-1} \leq 1$, replacing $|\beta(t)|$ by its upper bound in (4.84) gives

$$\epsilon_{\phi,2} \leq \frac{1}{2}\left(1+\frac{s\,|a'(u)|}{|a(u)|}\right) \sup_{|t-u|\leq s/2} s^2|\phi''(t)| \leq \sup_{|t-u|\leq s/2} s^2|\phi''(t)|.$$

Let us finally compute ϵ_a when $\xi = \phi'(u)$. Since $g(t) = g(-t)$, we derive from (4.82) that

$$\epsilon^+_{a,1} = \frac{2|a'(u)|}{|a(u)|}\left|\int_{-\infty}^{+\infty} t\,\frac{1}{\sqrt{s}}\,g_s(t)\,dt\right| = 0\,.$$

We also derive from (2.22) that the Fourier transform of $t\frac{1}{\sqrt{s}}g_s(t)$ is $is\,\hat{g}'(s\omega)$, so (4.85) gives

$$\epsilon_a = \frac{1}{2}\epsilon^-_{a,1} = \frac{s|a'(u)|}{|a(u)|}\,|\hat{g}'(2s\phi'(u))|\,.$$

■

Delprat *et al.* [154] give a different proof of a similar result when $g(t)$ is a Gaussian, using a stationary phase approximation. If we can neglect the corrective term $\epsilon(u,\xi)$ we shall see that (4.75) enables us to measure $a(u)$ and $\phi'(u)$ from $Sf(u,\xi)$. This implies that the decomposition $f(t) = a(t)\cos\phi(t)$ is uniquely defined. By reviewing the proof of Theorem 4.5, one can verify that a and ϕ' are the analytic amplitude and instantaneous frequencies of f.

The expressions (4.77, 4.78) show that the three corrective terms $\epsilon_{a,1}$, $\epsilon_{a,2}$ and $\epsilon_{\phi,2}$ are small if $a(t)$ and $\phi'(t)$ have small relative variations over the support of the window g_s. Let $\Delta\omega$ be the bandwidth of $\hat{g}$ defined by

$$|\hat{g}(\omega)| \ll 1 \quad \text{for } |\omega| \geq \Delta\omega. \tag{4.88}$$

The term $\sup_{|\omega|\geq s|\phi'(u)|} |\hat{g}(\omega)|$ of $\epsilon(u,\xi)$ is negligible if

$$\phi'(u) \geq \frac{\Delta\omega}{s}\,.$$

Ridge Points Let us suppose that $a(t)$ and $\phi'(t)$ have small variations over intervals of size s and that $\phi'(t) \geq \Delta\omega/s$ so that the corrective term $\epsilon(u,\xi)$ in (4.75) can be neglected. Since $|\hat{g}(\omega)|$ is maximum at $\omega = 0$, (4.75) shows that for each u the spectrogram $|Sf(u,\xi)|^2 = |\langle f, g_{s,u,\xi}\rangle|^2$ is maximum at $\xi(u) = \phi'(u)$. The corresponding time-frequency points $(u,\xi(u))$ are called *ridges*. At ridge points, (4.75) becomes

$$Sf(u,\xi) = \frac{\sqrt{s}}{2}\,a(u)\,\exp(i[\phi(u)-\xi u])\left(\hat{g}(0)+\epsilon(u,\xi)\right). \tag{4.89}$$

Theorem 4.5 proves that the $\epsilon(u,\xi)$ is smaller at a ridge point because the first order term $\epsilon_{a,1}$ becomes negligible in (4.79). This is shown by verifying that $|\hat{g}'(2s\phi'(u))|$ is negligible when $s\phi'(u) \geq \Delta\omega$. At ridge points, the second order terms $\epsilon_{a,2}$ and $\epsilon_{\phi,2}$ are predominant in $\epsilon(u,\xi)$.

The ridge frequency gives the instantaneous frequency $\xi(u) = \phi'(u)$ and the amplitude is calculated by

$$a(u) = \frac{2\,|Sf(u,\xi(u))|}{\sqrt{s}\,|\hat{g}(0)|}. \tag{4.90}$$

Let $\Phi_S(u,\xi)$ be the complex phase of $Sf(u,\xi)$. If we neglect the corrective term, then (4.89) proves that ridges are also points of stationary phase:

$$\frac{\partial \Phi_S(u,\xi)}{\partial u} = \phi'(u) - \xi = 0.$$

Testing the stationarity of the phase locates the ridges more precisely.

Multiple Frequencies When the signal contains several spectral lines whose frequencies are sufficiently apart, the windowed Fourier transform separates each of these components and the ridges detect the evolution in time of each spectral component. Let us consider

$$f(t) = a_1(t)\cos\phi_1(t) + a_2(t)\cos\phi_2(t),$$

where $a_k(t)$ and $\phi_k'(t)$ have small variations over intervals of size s and $s\phi_k'(t) \geq \Delta\omega$. Since the windowed Fourier transform is linear, we apply (4.75) to each spectral component and neglect the corrective terms:

$$\begin{aligned} Sf(u,\xi) &= \frac{\sqrt{s}}{2} a_1(u)\,\hat{g}(s[\xi - \phi_1'(u)])\exp(i[\phi_1(u) - \xi u]) \\ &\quad + \frac{\sqrt{s}}{2} a_2(u)\,\hat{g}(s[\xi - \phi_2'(u)])\exp(i[\phi_2(u) - \xi u])\,. \end{aligned} \tag{4.91}$$

The two spectral components are discriminated if for all u

$$\hat{g}(s|\phi_1'(u) - \phi_2'(u)|) \ll 1, \tag{4.92}$$

which means that the frequency difference is larger than the bandwidth of $\hat{g}(s\omega)$:

$$|\phi_1'(u) - \phi_2'(u)| \geq \frac{\Delta\omega}{s}. \tag{4.93}$$

In this case, when $\xi = \phi_1'(u)$, the second term of (4.91) can be neglected and the first term generates a ridge point from which we may recover $\phi_1'(u)$ and $a_1(u)$, using (4.90). Similarly, if $\xi = \phi_2'(u)$ the first term can be neglected and we have a second ridge point that characterizes $\phi_2'(u)$ and $a_2(u)$. The ridge points are

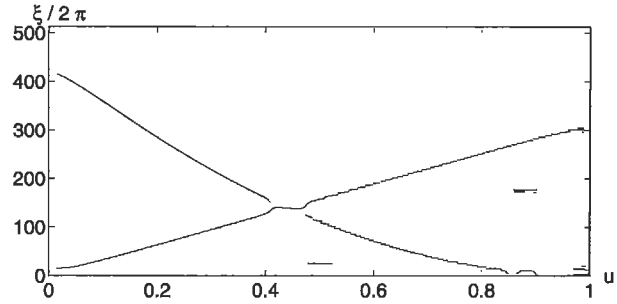


FIGURE 4.12 Larger amplitude ridges calculated from the spectrogram in Figure 4.3. These ridges give the instantaneous frequencies of the linear and quadratic chirps, and of the low and high frequency transients at $t = 0.5$ and $t = 0.87$.

distributed along two time-frequency lines $\xi(u) = \phi_1'(u)$ and $\xi(u) = \phi_2'(u)$. This result is valid for any number of time varying spectral components, as long as the distance between any two instantaneous frequencies satisfies (4.93). If two spectral lines are too close, they interfere, which destroys the ridge pattern.

Generally, the number of instantaneous frequencies is unknown. We thus detect all local maxima of $|Sf(u,\xi)|^2$ which are also points of stationary phase $\frac{\partial \Phi_S(u,\xi)}{\partial u} = \phi'(u) - \xi = 0$. These points define curves in the (u,ξ) planes that are the ridges of the windowed Fourier transform. Ridges corresponding to a small amplitude $a(u)$ are often removed because they can be artifacts of noise variations, or "shadows" of other instantaneous frequencies created by the side-lobes of $\hat{g}(\omega)$.

Figure 4.12 displays the ridges computed from the modulus and phase of the windowed Fourier transform shown in Figure 4.3. For $t \in [0.4, 0.5]$, the instantaneous frequencies of the linear chirp and the quadratic chirps are close and the frequency resolution of the window is not sufficient to discriminate them. As a result, the ridges detect a single average instantaneous frequency.

Choice of Window The measurement of instantaneous frequencies at ridge points is valid only if the size s of the window g_s is sufficiently small so that the second order terms $\epsilon_{a,2}$ and $\epsilon_{\phi,2}$ in (4.77,4.78) are small:

$$\sup_{|t-u| \leq s/2} \frac{s^2 |a_k''(t)|}{|a_k(u)|} \ll 1 \quad \text{and} \quad \sup_{|t-u| \leq s/2} s^2 |\phi_k''(t)| \ll 1 \,. \tag{4.94}$$

On the other hand, the frequency bandwidth $\Delta\omega/s$ must also be sufficiently small to discriminate consecutive spectral components in (4.93). The window scale s must therefore be adjusted as a trade-off between both constraints.

Table 4.1 gives the spectral parameters of several windows of compact support. For instantaneous frequency detection, it is particularly important to ensure that $\hat{g}$ has negligible side-lobes at $\pm\omega_0$, as illustrated by Figure 4.4. The reader can verify

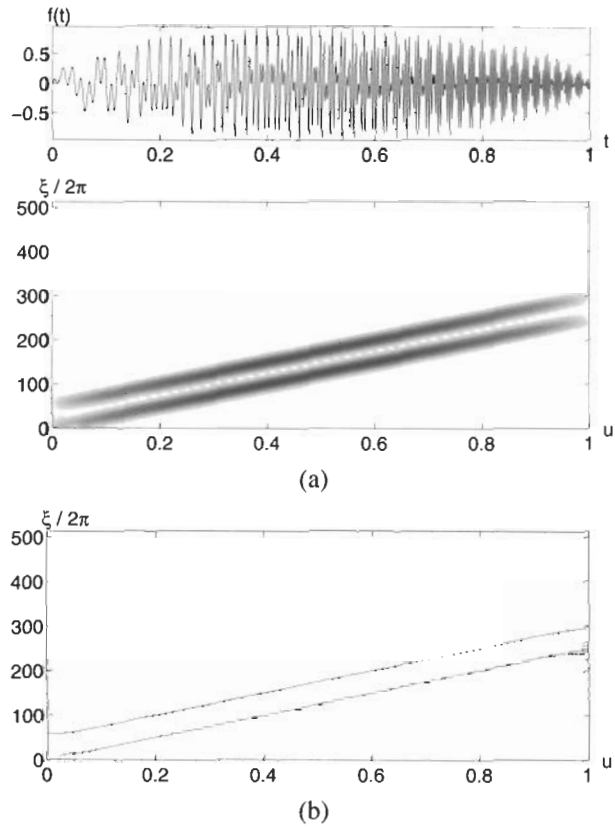


FIGURE 4.13 Sum of two parallel linear chirps. (a): Spectrogram $P_S f(u,\xi) = |Sf(u,\xi)|^2$. (b): Ridges calculated from the spectrogram.

with (4.75) that these side-lobes "react" to an instantaneous frequency $\phi'(u)$ by creating shadow maxima of $|Sf(u,\xi)|^2$ at frequencies $\xi = \phi'(u) \pm \omega_0$. The ratio of the amplitude of these shadow maxima to the amplitude of the main local maxima at $\xi = \phi'(u)$ is $|\hat{g}(\omega_0)|^2 |\hat{g}(0)|^{-2}$. They can be removed by thresholding or by testing the stationarity of the phase.

Example 4.13 The sum of two parallel linear chirps

$$f(t) = a_1 \cos(bt^2 + ct) + a_2 \cos(bt^2) \tag{4.95}$$

has two instantaneous frequencies $\phi_1'(t) = 2bt + c$ and $\phi_2'(t) = 2bt$. Figure 4.13 gives a numerical example.

The window g_s has enough frequency resolution to discriminate both chirps if

$$|\phi_1'(t) - \phi_2'(t)| = |c| \geq \frac{\Delta\omega}{s}. \tag{4.96}$$

Its time support is small enough compared to their time variation if

$$s^2 |\phi_1''(u)| = s^2 |\phi_2''(u)| = 2bs^2 \ll 1. \tag{4.97}$$

Conditions (4.96) and (4.97) prove that we can find an appropriate window g if and only if

$$\frac{c}{\sqrt{b}} \gg \Delta\omega. \tag{4.98}$$

Since g is a smooth window with a support $[-1/2, 1/2]$, its frequency bandwidth $\Delta\omega$ is on the order of 1. The linear chirps in Figure 4.13 satisfy (4.98). Their ridges are computed with the truncated Gaussian window of Table 4.1, with $s = 0.5$.

Example 4.14 The hyperbolic chirp

$$f(t) = \cos\left(\frac{\alpha}{\beta - t}\right)$$

for $0 \leq t < \beta$ has an instantaneous frequency

$$\phi'(t) = \frac{\alpha}{(\beta - t)^2},$$

which varies quickly when t is close to β. The instantaneous frequency of hyperbolic chirps goes from 0 to $+\infty$ in a finite time interval. This is particularly useful for radars. These chirps are also emitted by the cruise sonars of bats [154].

The instantaneous frequency of hyperbolic chirps cannot be estimated with a windowed Fourier transform because for any fixed window size the instantaneous frequency varies too quickly at high frequencies. When u is close enough to β then (4.94) is not satisfied because

$$s^2 |\phi''(u)| = \frac{s^2 \alpha}{(\beta - u)^3} > 1.$$

Figure 4.14 shows a signal that is a sum of two hyperbolic chirps:

$$f(t) = a_1 \cos\left(\frac{\alpha_1}{\beta_1 - t}\right) + a_2 \cos\left(\frac{\alpha_2}{\beta_2 - t}\right), \tag{4.99}$$

with $\beta_1 = 0.68$ and $\beta_2 = 0.72$. At the beginning of the signal, the two chirps have close instantaneous frequencies that are discriminated by the windowed Fourier ridge computed with a large size window. When getting close to β_1 and β_2, the instantaneous frequency varies too quickly relative to the window size. The resulting ridges cannot follow these instantaneous frequencies.

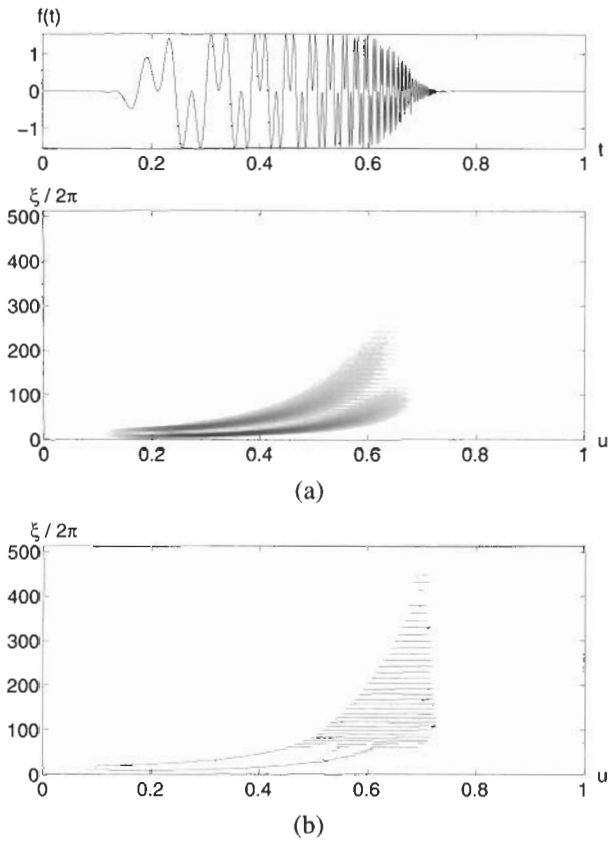


FIGURE 4.14 Sum of two hyperbolic chirps. (a): Spectrogram $P_S f(u,\xi)$. (b): Ridges calculated from the spectrogram

4.4.2 Wavelet Ridges

Windowed Fourier atoms have a fixed scale and thus cannot follow the instantaneous frequency of rapidly varying events such as hyperbolic chirps. In contrast, an analytic wavelet transform modifies the scale of its time-frequency atoms. The ridge algorithm of Delprat *et al.* [154] is extended to analytic wavelet transforms to accurately measure frequency tones that are rapidly changing at high frequencies.

An approximately analytic wavelet is constructed in (4.60) by multiplying a window g with a sinusoidal wave:

$$\psi(t) = g(t)\exp(i\eta t)\,.$$

As in the previous section, g is a symmetric window with a support equal to $[-1/2, 1/2]$, and a unit norm $\|g\| = 1$. Let $\Delta\omega$ be the bandwidth of $\hat{g}$ defined in

(4.88). If $\eta > \Delta\omega$ then

$$\forall \omega < 0 \ , \ \hat{\psi}(\omega) = \hat{g}(\omega - \eta) \ll 1 \ .$$

The wavelet ψ is not strictly analytic because its Fourier transform is not exactly equal to zero at negative frequencies.

Dilated and translated wavelets can be rewritten

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) = g_{s,u,\xi}(t) \exp(-i\xi u) ,$$

with $\xi = \eta/s$ and

$$g_{s,u,\xi}(t) = \sqrt{s}\, g\left(\frac{t-u}{s}\right) \exp(i\xi t).$$

The resulting wavelet transform uses time-frequency atoms similar to those of a windowed Fourier transform (4.74) but in this case the scale s varies over $\mathbb{R}^+$ while $\xi = \eta/s$:

$$Wf(u,s) = \langle f, \psi_{u,s} \rangle = \langle f, g_{s,u,\xi} \rangle \exp(i\xi u) .$$

Theorem 4.5 computes $\langle f, g_{s,u,\xi} \rangle$ when $f(t) = a(t)\cos\phi(t)$, which gives

$$Wf(u,s) = \frac{\sqrt{s}}{2} a(u) \exp[i\phi(u)] \left(\hat{g}(s[\xi - \phi'(u)]) + \epsilon(u,\xi) \right). \tag{4.100}$$

The corrective term $\epsilon(u,\xi)$ is negligible if $a(t)$ and $\phi'(t)$ have small variations over the support of $\psi_{u,s}$ and if $\phi'(u) \geq \Delta\omega/s$.

Wavelet Ridges The instantaneous frequency is measured from ridges defined over the wavelet transform. The normalized scalogram defined by

$$\frac{\xi}{\eta} P_W f(u,\xi) = \frac{|Wf(u,s)|^2}{s} \quad \text{for } \xi = \eta/s$$

is calculated with (4.100):

$$\frac{\xi}{\eta} P_W f(u,\xi) = \frac{1}{4} a^2(u) \left| \hat{g}\left(\eta\left[1 - \frac{\phi'(u)}{\xi}\right]\right) + \epsilon(u,\xi) \right|^2 .$$

Since $|\hat{g}(\omega)|$ is maximum at $\omega = 0$, if we neglect $\epsilon(u,\xi)$, this expression shows that the scalogram is maximum at

$$\frac{\eta}{s(u)} = \xi(u) = \phi'(u) \ . \tag{4.101}$$

The corresponding points $(u, \xi(u))$ are called *wavelet ridges*. The analytic amplitude is given by

$$a(u) = \frac{2\sqrt{\eta^{-1}\xi P_W f(u,\xi)}}{|\hat{g}(0)|}. \tag{4.102}$$

The complex phase of $Wf(u,s)$ in (4.100) is $\Phi_W(u,\xi) = \phi(u)$. At ridge points,

$$\frac{\partial \Phi_W(u,\xi)}{\partial u} = \phi'(u) = \xi. \tag{4.103}$$

When $\xi = \phi'(u)$, the first order term $\epsilon_{a,1}$ calculated in (4.79) becomes negligible. The corrective term is then dominated by $\epsilon_{a,2}$ and $\epsilon_{\phi,2}$. To simplify their expression we approximate the sup of a'' and ϕ'' in the neighborhood of u by their value at u. Since $s = \eta/\xi = \eta/\phi'(u)$, (4.77,4.78) imply that these second order terms become negligible if

$$\frac{\eta^2}{|\phi'(u)|^2}\frac{|a''(u)|}{|a(u)|} \ll 1 \text{ and } \eta^2\frac{|\phi''(u)|}{|\phi'(u)|^2} \ll 1. \tag{4.104}$$

The presence of ϕ' in the denominator proves that a' and ϕ' must have slow variations if ϕ' is small but may vary much more quickly for large instantaneous frequencies.

Multispectral Estimation Suppose that f is a sum of two spectral components:

$$f(t) = a_1(t)\cos\phi_1(t) + a_2(t)\cos\phi_2(t).$$

As in (4.92), we verify that the second instantaneous frequency ϕ_2' does not interfere with the ridge of ϕ_1' if the dilated window has a sufficient spectral resolution at the ridge scale $s = \eta/\xi = \eta/\phi_1'(u)$:

$$\hat{g}(s|\phi_1'(u) - \phi_2'(u)|) \ll 1. \tag{4.105}$$

Since the bandwidth of $\hat{g}(\omega)$ is $\Delta\omega$, this means that

$$\frac{|\phi_1'(u) - \phi_2'(u)|}{\phi_1'(u)} \geq \frac{\Delta\omega}{\eta}. \tag{4.106}$$

Similarly, the first spectral component does not interfere with the second ridge located at $s = \eta/\xi = \eta/\phi_2'(u)$ if

$$\frac{|\phi_1'(u) - \phi_2'(u)|}{\phi_2'(u)} \geq \frac{\Delta\omega}{\eta}. \tag{4.107}$$

To separate spectral lines whose instantaneous frequencies are close, these conditions prove that the wavelet must have a small octave bandwidth $\Delta\omega/\eta$. The bandwidth $\Delta\omega$ is a fixed constant, which is on the order of 1. The frequency η is a free parameter whose value is chosen as a trade-off between the time-resolution condition (4.104) and the frequency bandwidth conditions (4.106) and (4.107).

Figure 4.15 displays the ridges computed from the normalized scalogram and the wavelet phase shown in Figure 4.11. The ridges of the high frequency transient located at $t = 0.87$ have oscillations because of the interferences with the linear chirp above. The frequency separation condition (4.106) is not satisfied. This is also the case in the time interval $[0.35, 0.55]$, where the instantaneous frequencies of the linear and quadratic chirps are too close.

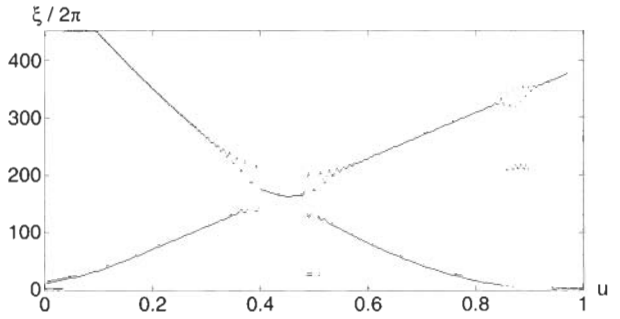


FIGURE 4.15 Ridges calculated from the scalogram shown in Figure 4.11. Compare with the windowed Fourier ridges in Figure 4.12.

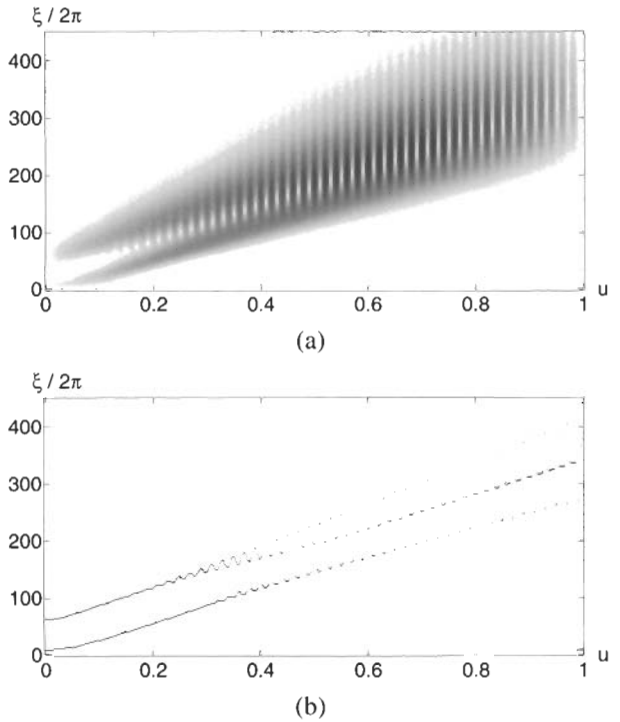


FIGURE 4.16 (a): Normalized scalogram $\eta^{-1}\xi P_W f(u,\xi)$ of two parallel linear chirps shown in Figure 4.13. (b): Wavelet ridges.

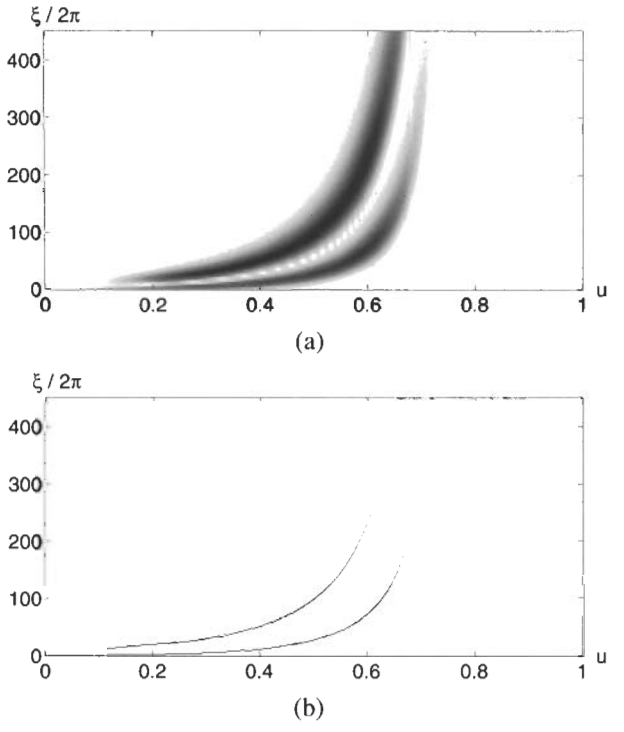


FIGURE 4.17 (a): Normalized scalogram $\eta^{-1}\xi P_W f(u,\xi)$ of two hyperbolic chirps shown in Figure 4.14. (b): Wavelet ridges.

Example 4.15 The instantaneous frequencies of two linear chirps

$$f(t) = a_1 \cos(bt^2 + ct) + a_2 \cos(bt^2)$$

are not well measured by wavelet ridges. Indeed

$$\frac{|\phi_2'(u) - \phi_1'(u)|}{\phi_1'(u)} = \frac{c}{bt}$$

converges to zero when t increases. When it is smaller than $\Delta\omega/\eta$ the two chirps interact and create interference patterns like those in Figure 4.16. The ridges follow these interferences and do not estimate properly the two instantaneous frequencies, as opposed to the windowed Fourier ridges shown in Figure 4.13.

Example 4.16 The instantaneous frequency of a hyperbolic chirp

$$f(t) = \cos\left(\frac{\alpha}{\beta - t}\right)$$

is $\phi'(t) = \alpha(1-t)^{-2}$. Wavelet ridges can measure this instantaneous frequency if the time resolution condition (4.104) is satisfied:

$$\eta^2 \ll \frac{\phi'(t)^2}{|\phi''(t)|} = \frac{\alpha}{|t-\beta|}.$$

This is the case if $|t-\beta|$ is not too large.

Figure 4.17 displays the scalogram and the ridges of two hyperbolic chirps

$$f(t) = a_1 \cos\left(\frac{\alpha_1}{\beta_1 - t}\right) + a_2 \cos\left(\frac{\alpha_2}{\beta_2 - t}\right),$$

with $\beta_1 = 0.68$ and $\beta_2 = 0.72$. As opposed to the windowed Fourier ridges shown in Figure 4.14, the wavelet ridges follow the rapid time modification of both instantaneous frequencies. This is particularly useful in analyzing the returns of hyperbolic chirps emitted by radars or sonars. Several techniques have been developed to detect chirps with wavelet ridges in presence of noise [117, 328].

4.5 QUADRATIC TIME-FREQUENCY ENERGY [1]

The wavelet and windowed Fourier transforms are computed by correlating the signal with families of time-frequency atoms. The time and frequency resolution of these transforms is thus limited by the time-frequency resolution of the corresponding atoms. Ideally, one would like to define a density of energy in a time-frequency plane, with no loss of resolution.

The Wigner-Ville distribution is a time-frequency energy density computed by correlating f with a time and frequency translation of itself. Despite its remarkable properties, the application of Wigner-Ville distributions is limited by the existence of interference terms. These interferences can be attenuated by a time-frequency averaging, but this results in a loss of resolution. It is proved that the spectrogram, the scalogram and all squared time-frequency decompositions can be written as a time-frequency averaging of the Wigner-Ville distribution.

4.5.1 Wigner-Ville Distribution

To analyze time-frequency structures, in 1948 Ville [342] introduced in signal processing a quadratic form that had been studied by Wigner [351] in a 1932 article on quantum thermodynamics:

$$P_V f(u,\xi) = \int_{-\infty}^{+\infty} f\left(u+\frac{\tau}{2}\right) f^*\left(u-\frac{\tau}{2}\right) e^{-i\tau\xi}\, d\tau. \tag{4.108}$$

The Wigner-Ville distribution remains real because it is the Fourier transform of $f(u+\tau/2)f^*(u-\tau/2)$, which has a Hermitian symmetry in τ. Time and

frequency have a symmetrical role. This distribution can also be rewritten as a frequency integration by applying the Parseval formula:

$$P_V f(u,\xi) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}\left(\xi + \frac{\gamma}{2}\right) \hat{f}^*\left(\xi - \frac{\gamma}{2}\right) e^{i\gamma u}\, d\gamma. \tag{4.109}$$

Time-Frequency Support The Wigner-Ville transform localizes the time-frequency structures of f. If the energy of f is well concentrated in time around u_0 and in frequency around ξ_0 then $P_V f$ has its energy centered at (u_0, ξ_0), with a spread equal to the time and frequency spread of f. This property is illustrated by the following proposition, which relates the time and frequency support of $P_V f$ to the support of f and $\hat{f}$.

Proposition 4.2 • *If the support of f is $[u_0 - T/2, u_0 + T/2]$, then for all ξ the support in u of $P_V f(u,\xi)$ is included in this interval.*

• *If the support of $\hat{f}$ is $[\xi_0 - \Delta/2, \xi_0 + \Delta/2]$, then for all u the support in ξ of $P_V f(u,\xi)$ is included in this interval.*

Proof[2]. Let $\bar{f}(t) = f(-t)$. The Wigner-Ville distribution is rewritten

$$P_V f(u,\xi) = \int_{-\infty}^{+\infty} f\left(\frac{\tau + 2u}{2}\right) \bar{f}^*\left(\frac{\tau - 2u}{2}\right) e^{-i\xi\tau}\, d\tau. \tag{4.110}$$

Suppose that f has a support equal to $[u_0 - T/2, u_0 + T/2]$. The supports of $f(\tau/2 + u)$ and $\bar{f}(\tau/2 - u)$ are then respectively

$$[2(u_0 - u) - T, 2(u_0 - u) + T] \text{ and } [-2(u_0 + u) - T, -2(u_0 + u) + T].$$

The Wigner-Ville integral (4.110) shows that $P_V f(u,\xi)$ is non-zero if these two intervals overlap, which is the case only if $|u_0 - u| < T$. The support of $P_V f(u,\xi)$ along u is therefore included in the support of f. If the support of $\hat{f}$ is an interval, then the same derivation based on (4.109) shows that the support of $P_V f(u,\xi)$ along ξ is included in the support of $\hat{f}$. ■

Example 4.17 Proposition 4.2 proves that the Wigner-Ville distribution does not spread the time or frequency support of Diracs or sinusoids, unlike windowed Fourier and wavelet transforms. Direct calculations yield

$$f(t) = \delta(u - u_0) \quad \Longrightarrow \quad P_V f(u,\xi) = \delta(t - u_0)\,, \tag{4.111}$$

$$f(t) = \exp(i\xi_0 t) \quad \Longrightarrow \quad P_V f(u,\xi) = \frac{1}{2\pi}\,\delta(\xi - \xi_0)\,. \tag{4.112}$$

Example 4.18 If f is a smooth and symmetric window then its Wigner-Ville distribution $P_V f(u,\xi)$ is concentrated in a neighborhood of $u = \xi = 0$. A Gaussian $f(t) = (\sigma^2\pi)^{-1/4} \exp(-t^2/(2\sigma^2))$ is transformed into a two-dimensional Gaussian because its Fourier transform is also a Gaussian (2.32) and one can verify that

$$P_V f(u,\xi) = \frac{1}{\pi} \exp\left(\frac{-u^2}{\sigma^2} - \sigma^2\xi^2\right). \tag{4.113}$$

In this particular case $P_V f(u,\xi) = |f(u)|^2 |\hat{f}(\xi)|^2$.

The Wigner-Ville distribution has important invariance properties. A phase shift does not modify its value:

$$g(t) = e^{i\phi} g(t) \implies P_V f(u,\xi) = P_V g(u,\xi) . \tag{4.114}$$

When f is translated in time or frequency, its Wigner-Ville transform is also translated:

$$f(t) = g(t-u_0) \implies P_V f(u,\xi) = P_V g(u-u_0,\xi) , \tag{4.115}$$

$$f(t) = \exp(i\xi_0 t) g(t) \implies P_V f(u,\xi) = P_V g(u,\xi-\xi_0) . \tag{4.116}$$

If f is scaled by s and thus $\hat{f}$ is scaled by $1/s$ then the time and frequency parameters of $P_V f$ are also scaled respectively by s and $1/s$

$$f(t) = \frac{1}{\sqrt{s}} g\Big(\frac{t}{s}\Big) \implies P_V f(u,\xi) = P_V f\Big(\frac{u}{s}, s\xi\Big) . \tag{4.117}$$

Example 4.19 If g is a smooth and symmetric window then $P_V g(u,\xi)$ has its energy concentrated in the neighborhood of $(0,0)$. The time-frequency atom

$$f_0(t) = \frac{a}{\sqrt{s}} \exp(i\phi_0) f\Big(\frac{t-u_0}{s}\Big) \exp(i\xi_0 t) .$$

has a Wigner-Ville distribution that is calculated with (4.114), (4.115) and (4.116):

$$P_V f_0(u,\xi) = |a|^2 P_V g\Big(\frac{u-u_0}{s}, s(\xi-\xi_0)\Big) . \tag{4.118}$$

Its energy is thus concentrated in the neighborhood of (u_0,ξ_0), on an ellipse whose axes are proportional to s in time and $1/s$ in frequency.

Instantaneous Frequency Ville's original motivation for studying time-frequency decompositions was to compute the instantaneous frequency of a signal [342]. Let f_a be the analytic part of f obtained in (4.69) by setting to zero $\hat{f}(\omega)$ for $\omega < 0$. We write $f_a(t) = a(t) \exp[i\phi(t)]$ to define the instantaneous frequency $\omega(t) = \phi'(t)$. The following proposition proves that $\phi'(t)$ is the "average" frequency computed relative to the Wigner-Ville distribution $P_V f_a$.

Proposition 4.3 *If* $f_a(t) = a(t) \exp[i\phi(t)]$ *then*

$$\phi'(u) = \frac{\int_{-\infty}^{+\infty} \xi P_V f_a(u,\xi)\, d\xi}{\int_{-\infty}^{+\infty} P_V f_a(u,\xi)\, d\xi} . \tag{4.119}$$

Proof[2]. To prove this result, we verify that any function g satisfies

$$\int\int \xi\, g\left(u+\frac{\tau}{2}\right) g^*\left(u-\frac{\tau}{2}\right) \exp(-i\tau\xi)\, d\xi\, d\tau = -\pi i \left[g'(u)\, g^*(u) - g(u)\, g^{*\prime}(u) \right]. \tag{4.120}$$

This identity is obtained by observing that the Fourier transform of $i\xi$ is the derivative of a Dirac, which gives an equality in the sense of distributions:

$$\int_{-\infty}^{+\infty} \xi \exp(-i\tau\xi)\, d\xi = -i2\pi\, \delta'(\tau).$$

Since $\int_{-\infty}^{+\infty} \delta'(\tau)\, h(\tau) = -h'(0)$, inserting $h(\tau) = g(u+\tau/2)\, g^*(u-\tau/2)$ proves (4.120). If $g(u) = f_a(u) = a(u)\exp[i\phi(u)]$ then (4.120) gives

$$\int_{-\infty}^{+\infty} \xi\, P_V f_a(u,\xi)\, d\xi = 2\pi a^2(u)\, \phi'(u).$$

We will see in (4.124) that $|f_a(u)|^2 = \int_{-\infty}^{+\infty} P_V f_a(u,\xi)\, d\xi$, and since $|f_a(u)|^2 = a(u)^2$ we derive (4.119). ■

This proposition shows that for a fixed u the mass of $P_V f_a(u,\xi)$ is typically concentrated in the neighborhood of the instantaneous frequency $\xi = \phi'(u)$. For example, a linear chirp $f(t) = \exp(iat^2)$ is transformed into a Dirac located along the instantaneous frequency $\xi = \phi'(u) = 2au$:

$$P_V f(u,\xi) = \delta(\xi - 2au).$$

Similarly, the multiplication of f by a linear chirp $\exp(iat^2)$ makes a frequency translation of $P_V f$ by the instantaneous frequency $2au$:

$$f(t) = \exp(iat^2)\, g(t) \implies P_V f(u,\xi) = P_V g(u,\xi - 2au)\,. \tag{4.121}$$

Energy Density The Moyal [275] formula proves that the Wigner-Ville transform is unitary, which implies energy conservation properties.

Theorem 4.6 (MOYAL) *For any f and g in $\mathbf{L}^2(\mathbb{R})$*

$$\left| \int_{-\infty}^{+\infty} f(t)\, g^*(t)\, dt \right|^2 = \frac{1}{2\pi} \int\int P_V f(u,\xi)\, P_V g(u,\xi)\, du\, d\xi. \tag{4.122}$$

Proof[1]. Let us compute the integral

$$\begin{aligned} I &= \int\int P_V f(u,\xi)\, P_V g(u,\xi)\, du\, d\xi \\ &= \int\int\int\int f\left(u+\frac{\tau}{2}\right) f^*\left(u-\frac{\tau}{2}\right) g\left(u+\frac{\tau'}{2}\right) g^*\left(u-\frac{\tau'}{2}\right) \\ &\qquad \exp[-i\xi(\tau+\tau')]\, d\tau\, d\tau'\, du\, d\xi. \end{aligned}$$

The Fourier transform of $h(t) = 1$ is $\hat{h}(\omega) = 2\pi\delta(\omega)$, which means that we have a distribution equality $\int \exp[-i\xi(\tau+\tau')]d\xi = 2\pi\delta(\tau+\tau')$. As a result,

$$\begin{aligned} I &= 2\pi \iiint f\left(u+\frac{\tau}{2}\right) f^*\left(u-\frac{\tau}{2}\right) g\left(u+\frac{\tau'}{2}\right) g^*\left(u-\frac{\tau'}{2}\right) \delta(\tau+\tau')\, d\tau\, d\tau'\, du \\ &= 2\pi \iint f\left(u+\frac{\tau}{2}\right) f^*\left(u-\frac{\tau}{2}\right) g\left(u-\frac{\tau}{2}\right) g^*\left(u+\frac{\tau}{2}\right) d\tau\, du. \end{aligned}$$

The change of variable $t = u+\tau/2$ and $t' = u-\tau/2$ yields (4.122). ■

One can consider $|f(t)|^2$ and $|\hat{f}(\omega)|^2/(2\pi)$ as energy densities in time and frequency that satisfy a conservation equation:

$$\|f\|^2 = \int_{-\infty}^{+\infty} |f(t)|^2\, dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\hat{f}(\omega)|^2\, d\omega.$$

The following proposition shows that these time and frequency densities are recovered with marginal integrals over the Wigner-Ville distribution.

Proposition 4.4 *For any* $f \in \mathbf{L}^2(\mathbb{R})$

$$\int_{-\infty}^{+\infty} P_V f(u,\xi)\, du = |\hat{f}(\xi)|^2, \tag{4.123}$$

and

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} P_V f(u,\xi)\, d\xi = |f(u)|^2. \tag{4.124}$$

Proof[1]. The frequency integral (4.109) proves that the one-dimensional Fourier transform of $g_\xi(u) = P_V f(u,\xi)$ with respect to u is

$$\hat{g}_\xi(\gamma) = \hat{f}\left(\xi+\frac{\gamma}{2}\right) \hat{f}^*\left(\xi-\frac{\gamma}{2}\right).$$

We derive (4.123) from the fact that is

$$\hat{g}_\xi(0) = \int_{-\infty}^{+\infty} g_\xi(u)\, du.$$

Similarly, (4.108) shows that $P_V f(u,\xi)$ is the one-dimensional Fourier transform of $f(u+\tau/2)f^*(u-\tau/2)$ with respect to τ, where ξ is the Fourier variable. Its integral in ξ thus gives the value for $\tau = 0$, which is the identity (4.124). ■

This proposition suggests interpreting the Wigner-Ville distribution as a joint time-frequency energy density. However, the Wigner-Ville distribution misses one fundamental property of an energy density: positivity. Let us compute for example the Wigner-Ville distribution of $f = \mathbf{1}_{[-T,T]}$ with the integral (4.108):

$$P_V f(u,\xi) = \frac{2\sin\Big(2(T-|u|)\xi\Big)}{\xi} \mathbf{1}_{[-T,T]}(u).$$

It is an oscillating function that takes negative values. In fact, one can prove that translated and frequency modulated Gaussians are the only functions whose Wigner-Ville distributions remain positive. As we will see in the next section, to obtain positive energy distributions for all signals, it is necessary to average the Wigner-Ville transform and thus lose some time-frequency resolution.

4.5.2 Interferences and Positivity

At this point, the Wigner-Ville distribution may seem to be an ideal tool for analyzing the time-frequency structures of a signal. This is however not the case because of interferences created by the quadratic properties of this transform. These interferences can be removed by averaging the Wigner-Ville distribution with appropriate kernels which yield positive time-frequency densities. However, this reduces the time-frequency resolution. Spectrograms and scalograms are examples of positive quadratic distributions obtained by smoothing the Wigner-Ville distribution.

Cross Terms Let $f = f_1 + f_2$ be a composite signal. Since the Wigner-Ville distribution is a quadratic form,

$$P_V f = P_V f_1 + P_V f_2 + P_V[f_1, f_2] + P_V[f_2, f_1], \tag{4.125}$$

where $P_V[h, g]$ is the cross Wigner-Ville distribution of two signals

$$P_V[h,g](u,\xi) = \int_{-\infty}^{+\infty} h\left(u+\frac{\tau}{2}\right) g^*\left(u-\frac{\tau}{2}\right) e^{-i\tau\xi}\, d\tau. \tag{4.126}$$

The interference term

$$I[f_1, f_2] = P_V[f_1, f_2] + P_V[f_2, f_1]$$

is a real function that creates non-zero values at unexpected locations of the (u, ξ) plane.

Let us consider two time-frequency atoms defined by

$$f_1(t) = a_1 e^{i\phi_1} g(t-u_1) e^{i\xi_1 t} \quad \text{and} \quad f_2(t) = a_2 e^{i\phi_2} g(t-u_2) e^{i\xi_2 t},$$

where g is a time window centered at $t = 0$. Their Wigner-Ville distributions computed in (4.118) are

$$P_V f_1(u,\xi) = a_1^2 P_V g(u-u_1, \xi-\xi_1) \text{ and } P_V f_2(u,\xi) = a_2^2 P_V g(u-u_2, \xi-\xi_2).$$

Since the energy of $P_V g$ is centered at $(0,0)$, the energy of $P_V f_1$ and $P_V f_2$ is concentrated in the neighborhoods of (u_1, ξ_1) and (u_2, ξ_2) respectively. A direct calculation verifies that the interference term is

$$I[f_1, f_2](u,\xi) = 2a_1 a_2 P_V g(u-u_0, \xi-\xi_0) \cos\Big((u-u_0)\Delta\xi - (\xi-\xi_0)\Delta u + \Delta\phi\Big)$$

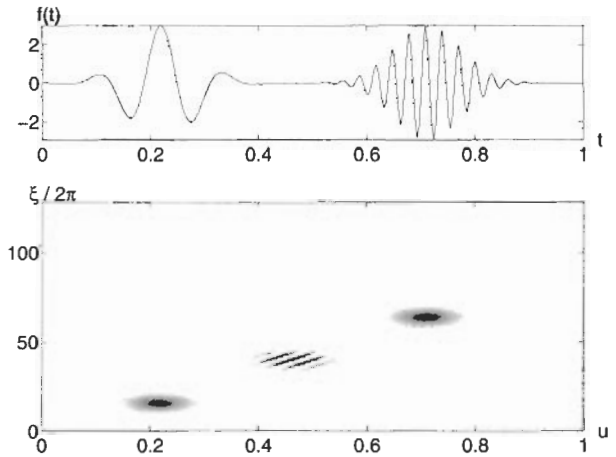


FIGURE 4.18 Wigner-Ville distribution $P_V f(u,\xi)$ of two Gabor atoms shown at the top. The oscillating interferences are centered at the middle time-frequency location.

with

$$u_0 = \frac{u_1 + u_2}{2} \quad , \quad \xi_0 = \frac{\xi_1 + \xi_2}{2}$$
$$\Delta u = u_1 - u_2 \quad , \quad \Delta\xi = \xi_1 - \xi_2$$
$$\Delta\phi = \phi_1 - \phi_2 + u_0\,\Delta\xi.$$

It is an oscillatory waveform centered at the middle point (u_0,ξ_0). This is quite counter-intuitive since f and $\hat{f}$ have very little energy in the neighborhood of u_0 and ξ_0. The frequency of the oscillations is proportional to the Euclidean distance $\sqrt{\Delta\xi^2 + \Delta u^2}$ of (u_1,ξ_1) and (u_2,ξ_2). The direction of these oscillations is perpendicular to the line that joins (u_1,ξ_1) and (u_2,ξ_2). Figure 4.18 displays the Wigner-Ville distribution of two atoms obtained with a Gaussian window g. The oscillating interference appears at the middle time-frequency point.

This example shows that the interference $I[f_1,f_2](u,\xi)$ has some energy in regions where $|f(u)|^2 \approx 0$ and $|\hat{f}(\xi)|^2 \approx 0$. These interferences can have a complicated structure [26, 211] but they are necessarily oscillatory because the marginal integrals (4.123) and (4.124) vanish:

$$\int_{-\infty}^{+\infty} P_V f(u,\xi)\,d\xi = 2\pi |f(u)|^2 \quad , \quad \int_{-\infty}^{+\infty} P_V f(u,\xi)\,du = |\hat{f}(\xi)|^2.$$

Analytic Part Interference terms also exist in a real signal f with a single instantaneous frequency component. Let $f_a(t) = a(t)\exp[i\phi(t)]$ be its analytic part:

$$f = \text{Real}[f_a] = \frac{1}{2}\,(f_a + f_a^*).$$

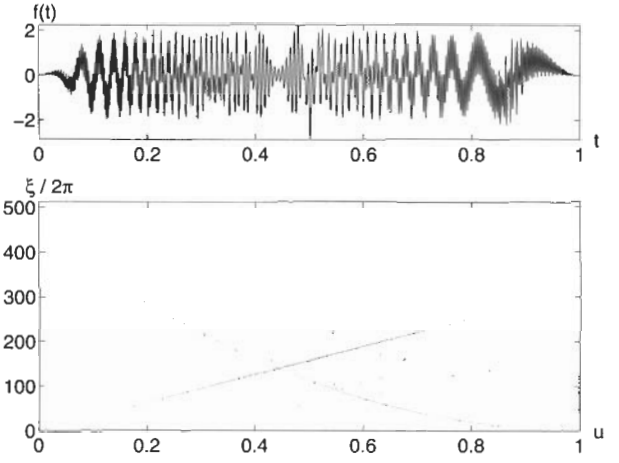


FIGURE 4.19 The bottom displays the Wigner-Ville distribution $P_V f_a(u,\xi)$ of the analytic part of the top signal.

Proposition 4.3 proves that for fixed u, $P_V f_a(u,\xi)$ and $P_V f_a^*(u,\xi)$ have an energy concentrated respectively in the neighborhood of $\xi_1 = \phi'(u)$ and $\xi_2 = -\phi'(u)$. Both components create an interference term at the intermediate zero frequency $\xi_0 = (\xi_1 + \xi_2)/2 = 0$. To avoid this low frequency interference, we often compute $P_V f_a$ as opposed to $P_V f$.

Figure 4.19 displays $P_V f_a$ for a real signal f that includes a linear chirp, a quadratic chirp and two isolated time-frequency atoms. The linear and quadratic chirps are localized along narrow time frequency lines, which are spread on wider bands by the scalogram and the scalogram shown in Figure 4.3 and 4.11. However, the interference terms create complex oscillatory patterns that make it difficult to detect the existence of the two time-frequency transients at $t = 0.5$ and $t = 0.87$, which clearly appear in the spectrogram and the scalogram.

Positivity Since the interference terms include positive and negative oscillations, they can be partly removed by smoothing $P_V f$ with a kernel θ:

$$P_\theta f(u,\xi) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} P_V f(u',\xi')\, \theta(u,u',\xi,\xi')\, du'\, d\xi'. \tag{4.127}$$

The time-frequency resolution of this distribution depends on the spread of the kernel θ in the neighborhood of (u,ξ). Since the interferences take negative values, one can guarantee that all interferences are removed by imposing that this time-frequency distribution remain positive $P_\theta f(u,\xi) \geq 0$ for all $(u,\xi) \in \mathbb{R}^2$.

The spectrogram (4.12) and scalogram (4.55) are examples of positive time-frequency energy distributions. In general, let us consider a family of time-frequency atoms $\{\phi_\gamma\}_{\gamma\in\Gamma}$. Suppose that for any (u,ξ) there exists a unique atom

$\phi_{\gamma(u,\xi)}$ centered in time-frequency at (u,ξ). The resulting time-frequency energy density is

$$Pf(u,\xi) = |\langle f, \phi_{\gamma(u,\xi)}\rangle|^2.$$

The Moyal formula (4.122) proves that this energy density can be written as a time-frequency averaging of the Wigner-Ville distribution

$$Pf(u,\xi) = \frac{1}{2\pi}\int\int P_V f(u',\xi')\, P_V \phi_{\gamma(u,\xi)}(u',\xi')\, du'\, d\xi'. \tag{4.128}$$

The smoothing kernel is the Wigner-Ville distribution of the atoms

$$\theta(u,u',\xi,\xi') = \frac{1}{2\pi} P_V \phi_{\gamma(u,\xi)}(u',\xi').$$

The loss of time-frequency resolution depends on the spread of the distribution $P_V\phi_{\gamma(u,\xi)}(u',\xi')$ in the neighborhood of (u,v).

Example 4.20 A spectrogram is computed with windowed Fourier atoms

$$\phi_{\gamma(u,\xi)}(t) = g(t-u)\, e^{i\xi t}.$$

The Wigner-Ville distribution calculated in (4.118) yields

$$\theta(u,u',\xi,\xi') = \frac{1}{2\pi} P_V \phi_{\gamma(u,\xi)}(u',\xi') = \frac{1}{2\pi} P_V g(u'-u,\xi'-\xi). \tag{4.129}$$

For a spectrogram, the Wigner-Ville averaging (4.128) is therefore a two-dimensional convolution with $P_V g$. If g is a Gaussian window, then $P_V g$ is a two-dimensional Gaussian. This proves that averaging $P_V f$ with a sufficiently wide Gaussian defines a positive energy density. The general class of time-frequency distributions obtained by convolving $P_V f$ with an arbitrary kernel θ is studied in Section 4.5.3.

Example 4.21 Let ψ be an analytic wavelet whose center frequency is η. The wavelet atom $\psi_{u,s}(t) = s^{-1/2}\psi((t-u)/s)$ is centered at $(u, \xi = \eta/s)$ and the scalogram is defined by

$$P_W f(u,\xi) = |\langle f, \psi_{u,s}\rangle|^2 \quad \text{for } \xi = \eta/s.$$

Properties (4.115, 4.117) prove that the averaging kernel is

$$\theta(u,u',\xi,\xi') = \frac{1}{2\pi} P_V \psi\left(\frac{u'-u}{s}, s\xi'\right) = \frac{1}{2\pi} P_V \psi\left(\frac{\xi}{\eta}(u'-u), \frac{\eta}{\xi}\xi'\right).$$

Positive time-frequency distributions totally remove the interference terms but produce a loss of resolution. This is emphasized by the following theorem, due to Wigner [352].

Theorem 4.7 (WIGNER) *There is no positive quadratic energy distribution Pf that satisfies*

$$\int_{-\infty}^{+\infty} Pf(u,\xi)\,d\xi = 2\pi |f(u)|^2 \quad \text{and} \quad \int_{-\infty}^{+\infty} Pf(u,\xi)\,du = |\hat{f}(\xi)|^2. \tag{4.130}$$

Proof[2]. Suppose that Pf is a positive quadratic distribution that satisfies these marginals. Since $Pf(u,\xi) \geq 0$, the integrals (4.130) imply that if the support of f is included in an interval I then $Pf(u,\xi) = 0$ for $u \notin I$. We can associate to the quadratic form Pf a bilinear distribution defined for any f and g by

$$P[f,g] = \frac{1}{4}\Big(P(f+g) - P(f-g)\Big).$$

Let f_1 and f_2 be two non-zero signals whose supports are two intervals I_1 and I_2 that do not intersect, so that $f_1 f_2 = 0$. Let $f = a f_1 + b f_2$:

$$Pf = |a|^2 Pf_1 + ab^* P[f_1,f_2] + a^* b P[f_2,f_1] + |b|^2 Pf_2.$$

Since I_1 does not intersect I_2, $Pf_1(u,\xi) = 0$ for $u \in I_2$. Remember that $Pf(u,\xi) \geq 0$ for all a and b so necessarily $P[f_1,f_2](u,\xi) = P[f_2,f_1](u,\xi) = 0$ for $u \in I_2$. Similarly we prove that these cross terms are zero for $u \in I_1$ and hence

$$Pf(u,\xi) = |a|^2 Pf_1(u,\xi) + |b|^2 Pf_2(u,\xi).$$

Integrating this equation and inserting (4.130) yields

$$|\hat{f}(\xi)|^2 = |a|^2 |\hat{f}_1(\xi)|^2 + |b|^2 |\hat{f}_2(\xi)|^2.$$

Since $\hat{f}(\xi) = a\hat{f}_1(\xi) + b\hat{f}_2(\xi)$ it follows that $\hat{f}_1(\xi)\,\hat{f}_2(\xi) = 0$. But this is not possible because f_1 and f_2 have a compact support in time and Theorem 2.6 proves that $\hat{f}_1$ and $\hat{f}_2$ are $\mathbf{C}^\infty$ functions that cannot vanish on a whole interval. We thus conclude that one cannot construct a positive quadratic distribution Pf that satisfies the marginals (4.130). ■

4.5.3 Cohen's Class [2]

While attenuating the interference terms with a smoothing kernel θ, we may want to retain certain important properties. Cohen [135] introduced a general class of quadratic time-frequency distributions that satisfy the time translation and frequency modulation invariance properties (4.115) and (4.116). If a signal is translated in time or frequency, its energy distribution is just translated by the corresponding amount. This was the beginning of a systematic study of quadratic time-frequency distributions obtained as a weighted average of a Wigner-Ville distribution [10, 26, 136, 210].

Section 2.1 proves that linear translation invariant operators are convolution products. The translation invariance properties (4.115, 4.116) are thus equivalent to imposing that the smoothing kernel in (4.127) be a convolution kernel

$$\theta(u,u',\xi,\xi') = \theta(u-u',\xi-\xi'), \tag{4.131}$$

and hence

$$P_\theta f(u,\xi) = P_V f \star \theta(u,\xi) = \int\int \theta(u-u',\xi-\xi')\, P_V f(u',\xi')\, du'\, d\xi'. \quad (4.132)$$

The spectrogram is an example of Cohen's class distribution, whose kernel in (4.129) is the Wigner-Ville distribution of the window

$$\theta(u,\xi) = \frac{1}{2\pi} P_V g(u,\xi) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} g\left(u+\frac{\tau}{2}\right) g\left(u-\frac{\tau}{2}\right) e^{-i\tau\xi}\, d\tau. \quad (4.133)$$

Ambiguity Function The properties of the convolution (4.132) are more easily studied by calculating the two-dimensional Fourier transform of $P_V f(u,\xi)$ with respect to u and ξ. We denote by $Af(\tau,\gamma)$ this Fourier transform

$$Af(\tau,\gamma) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} P_V f(u,\xi) \exp[-i(u\gamma+\xi\tau)]\, du\, d\xi.$$

Note that the Fourier variables τ and γ are inverted with respect to the usual Fourier notation. Since the one-dimensional Fourier transform of $P_V f(u,\xi)$ with respect to u is $\hat{f}(\xi+\gamma/2)\,\hat{f}^*(\xi-\gamma/2)$, applying the one-dimensional Fourier transform with respect to ξ gives

$$Af(\tau,\gamma) = \int_{-\infty}^{+\infty} \hat{f}\left(\xi+\frac{\gamma}{2}\right) \hat{f}^*\left(\xi-\frac{\gamma}{2}\right) e^{-i\tau\xi}\, d\xi. \quad (4.134)$$

The Parseval formula yields

$$Af(\tau,\gamma) = \int_{-\infty}^{+\infty} f\left(u+\frac{\tau}{2}\right) f^*\left(u-\frac{\tau}{2}\right) e^{-i\gamma u}\, du. \quad (4.135)$$

We recognize the *ambiguity function* encountered in (4.24) when studying the time-frequency resolution of a windowed Fourier transform. It measures the energy concentration of f in time and in frequency.

Kernel Properties The Fourier transform of $\theta(u,\xi)$ is

$$\hat{\theta}(\tau,\gamma) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \theta(u,\xi) \exp[-i(u\gamma+\xi\tau)]\, du\, d\xi.$$

As in the definition of the ambiguity function (4.134), the Fourier parameters τ and γ of $\hat{\theta}$ are inverted. The following proposition gives necessary and sufficient conditions to ensure that P_θ satisfies marginal energy properties like those of the Wigner-Ville distribution. The Wigner Theorem 4.7 proves that in this case $P_\theta f(u,\xi)$ takes negative values.

Proposition 4.5 *For all* $f \in \mathbf{L}^2(\mathbb{R})$

$$\int_{-\infty}^{+\infty} P_\theta f(u,\xi)\,d\xi = 2\pi |f(u)|^2 \ , \ \int_{-\infty}^{+\infty} P_\theta f(u,\xi)\,du = |\hat{f}(\xi)|^2, \qquad (4.136)$$

if and only if

$$\forall (\tau,\gamma) \in \mathbb{R}^2 \ , \ \hat{\theta}(\tau,0) = \hat{\theta}(0,\gamma) = 1. \qquad (4.137)$$

Proof[2]. Let $A_\theta f(\tau,\gamma)$ be the two-dimensional Fourier transform of $P_\theta f(u,\xi)$. The Fourier integral at $(0,\gamma)$ gives

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} P_\theta f(u,\xi)\, e^{-iu\gamma}\,d\xi\,du = A_\theta f(0,\gamma). \qquad (4.138)$$

Since the ambiguity function $Af(\tau,\gamma)$ is the Fourier transform of $P_V f(u,\xi)$, the two-dimensional convolution (4.132) gives

$$A_\theta(\tau,\gamma) = Af(\tau,\gamma)\,\hat{\theta}(\tau,\gamma). \qquad (4.139)$$

The Fourier transform of $2\pi|f(u)|^2$ is $\hat{f} \star \overline{\hat{f}}(\gamma)$, with $\overline{\hat{f}}(\gamma) = \hat{f}^*(-\gamma)$. The relation (4.138) shows that (4.136) is satisfied if and only if

$$A_\theta f(0,\gamma) = Af(0,\gamma)\,\hat{\theta}(0,\gamma) = \hat{f} \star \overline{\hat{f}}(\gamma). \qquad (4.140)$$

Since $P_V f$ satisfies the marginal property (4.123), we similarly prove that

$$Af(0,\gamma) = \hat{f} \star \overline{\hat{f}}(\gamma).$$

Requiring that (4.140) be valid for any $\hat{f}(\gamma)$, is equivalent to requiring that $\hat{\theta}(0,\gamma) = 1$ for all $\gamma \in \mathbb{R}$. The same derivation applied to the other marginal integration yields $\hat{\theta}(\xi,0) = 1$. ■

In addition to requiring time-frequency translation invariance, it may be useful to guarantee that P_θ satisfies the same scaling property as a Wigner-Ville distribution:

$$g(t) = \frac{1}{\sqrt{s}} f\left(\frac{t}{s}\right) \implies P_\theta g(u,\xi) = P_\theta f\left(\frac{u}{s}, s\xi\right).$$

Such a distribution P_θ is *affine* invariant. One can verify (Problem 4.15) that affine invariance is equivalent to imposing that

$$\forall s \in \mathbb{R}^+ \ , \ \theta\left(su, \frac{\xi}{s}\right) = \theta(u,\xi), \qquad (4.141)$$

and hence

$$\theta(u,\xi) = \theta(u\xi,1) = \beta(u\xi).$$

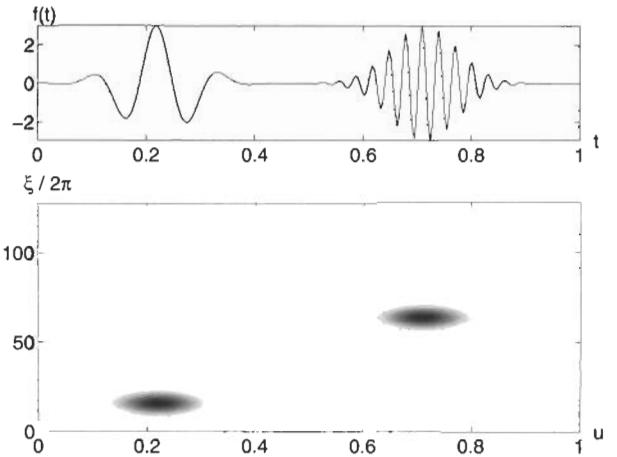


FIGURE 4.20 Choi-William distribution $P_\theta f(u,\xi)$ of the two Gabor atoms shown at the top. The interference term that appears in the Wigner-Ville distribution of Figure 4.18 has nearly disappeared.

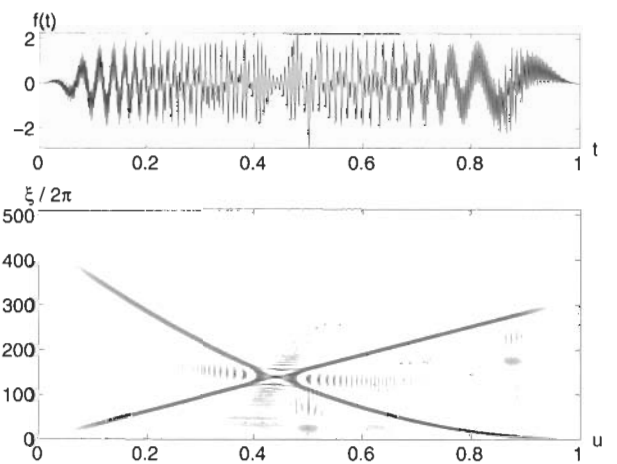


FIGURE 4.21 Choi-William distribution $P_\theta f_a(u,\xi)$ of the analytic part of the signal shown at the top. The interferences remain visible.

Example 4.22 The *Rihaczek* distribution is an affine invariant distribution whose convolution kernel is

$$\hat{\theta}(\tau,\gamma) = \exp\left(\frac{i\tau\gamma}{2}\right). \tag{4.142}$$

A direct calculation shows that

$$P_\theta f(u,\xi) = f(u)\,\hat{f}^*(\xi)\exp(-iu\xi). \tag{4.143}$$

Example 4.23 The kernel of the *Choi-William* distribution is [122]

$$\hat{\theta}(\tau,\gamma) = \exp(-\sigma^2\tau^2\gamma^2)\,. \tag{4.144}$$

It is symmetric and thus corresponds to a real function $\theta(u,\xi)$. This distribution satisfies the marginal conditions (4.137). Since $\lim_{\sigma\to 0}\hat{\theta}(\tau,\gamma) = 1$, when σ is small the Choi-William distribution is close to a Wigner-Ville distribution. Increasing σ attenuates the interference terms, but spreads $\theta(u,\xi)$, which reduces the time-frequency resolution of the distribution.

Figure 4.20 shows that the interference terms of two modulated Gaussians nearly disappear when the Wigner-Ville distribution of Figure 4.18 is averaged by a Choi-William kernel having a sufficiently large σ. Figure 4.21 gives the Choi-William distribution of the analytic signal whose Wigner-Ville distribution is in Figure 4.19. The energy of the linear and quadratic chirps are spread over wider time-frequency bands but the interference terms are attenuated, although not totally removed. It remains difficult to isolate the two modulated Gaussians at $t = 0.5$ and $t = 0.87$, which clearly appear in the spectrogram of Figure 4.3.

4.5.4 Discrete Wigner-Ville Computations [2]

The Wigner integral (4.108) is the Fourier transform of $f(u+\tau/2)f^*(u-\tau/2)$:

$$P_V f(u,\xi) = \int_{-\infty}^{+\infty} f\left(u+\frac{\tau}{2}\right) f^*\left(u-\frac{\tau}{2}\right) \mathrm{e}^{-i\tau\xi}\,d\tau. \tag{4.145}$$

For a discrete signal $f[n]$ defined over $0 \le n < N$, the integral is replaced by a discrete sum:

$$P_V f[n,k] = \sum_{p=-N}^{N-1} f\left[n+\frac{p}{2}\right] f^*\left[n-\frac{p}{2}\right] \exp\left(\frac{-i2\pi kp}{N}\right). \tag{4.146}$$

When p is odd, this calculation requires knowing the value of f at half integers. These values are computed by interpolating f, with an addition of zeroes to its Fourier transform. This is necessary to avoid the aliasing produced by the discretization of the Wigner-Ville integral [126].

The interpolation $\tilde{f}$ of f is a signal of size $2N$ whose discrete Fourier transform $\hat{\tilde{f}}$ is defined from the discrete Fourier transform $\hat{f}$ of f by

$$\hat{\tilde{f}}[k] = \begin{cases} 2\hat{f}[k] & \text{if } 0 \le k < N/2 \\ 0 & \text{if } N/2 < k < 3N/2 \\ 2\hat{f}[k-N] & \text{if } 3N/2 < k < 2N \\ \hat{f}[N/2] & \text{if } k = N/2, 3N/2 \end{cases}.$$

Computing the inverse discrete Fourier transform shows that $\tilde{f}[2n] = f[n]$ for $n \in [0, N-1]$. When $n \notin [0, 2N-1]$, we set $\tilde{f}[n] = 0$. The Wigner summation (4.146) is calculated from $\tilde{f}$:

$$\begin{aligned} P_V f[n,k] &= \sum_{p=-N}^{N-1} \tilde{f}[2n+p]\,\tilde{f}^*[2n-p] \exp\left(\frac{-i2\pi kp}{N}\right) \\ &= \sum_{p=0}^{2N-1} \tilde{f}[2n+p-N]\,\tilde{f}^*[2n-p+N] \exp\left(\frac{-i2\pi(2k)p}{2N}\right). \end{aligned}$$

For $0 \le n < N$ fixed, $P_V f[n,k]$ is the discrete Fourier transform of size $2N$ of $g[p] = \tilde{f}[2n+p-N]\,\tilde{f}^*[2n-p+N]$ at the frequency $2k$. The discrete Wigner-Ville distribution is thus calculated with N FFT procedures of size $2N$, which requires $O(N^2 \log N)$ operations. To compute the Wigner-Ville distribution of the analytic part f_a of f, we use (4.48).

Cohen's Class A Cohen's class distribution is calculated with a circular convolution of the discrete Wigner-Ville distribution with a kernel $\theta[p,q]$:

$$P_\theta[n,k] = P_V \circledast \theta[n,k]. \tag{4.147}$$

Its two-dimensional discrete Fourier transform is therefore

$$A_\theta[p,q] = Af[p,q]\,\hat{\theta}[p,q]. \tag{4.148}$$

The signal $Af[p,q]$ is the discrete ambiguity function, calculated with a two-dimensional FFT of the discrete Wigner-Ville distribution $P_V f[n,k]$. As in the case of continuous time, we have inverted the index p and q of the usual two-dimensional Fourier transform. The Cohen's class distribution (4.147) is obtained by calculating the inverse Fourier transform of (4.148). This also requires a total of $O(N^2 \log N)$ operations.

4.6 PROBLEMS

4.1. [1] *Instantaneous frequency* Let $f(t) = \exp[i\phi(t)]$.

(a) Prove that $\int_{-\infty}^{+\infty} |Sf(u,\xi)|^2\, d\xi = 2\pi$. Hint: $Sf(u,\xi)$ is a Fourier transform; use the Parseval formula.

(b) Similarly, show that

$$\int_{-\infty}^{+\infty} \xi\,|Sf(u,\xi)|^2\,d\xi = 2\pi \int_{-\infty}^{+\infty} \phi'(t)\,|g(t-u)|^2\,dt,$$

and interpret this result.

4.2. [1] Write a reproducing kernel equation for the discrete windowed Fourier transform $Sf[m,l]$ defined in (4.27).

4.3. [1] When $g(t) = (\pi\sigma^2)^{-1/4}\exp(-t^2/(2\sigma^2))$, compute the ambiguity function $Ag(\tau,\gamma)$.

4.4. [1] Let $g[n]$ be a window with L non-zero coefficients. For signals of size N, describe a fast algorithm that computes the discrete windowed Fourier transform (4.27) with $O(N\log_2 L)$ operations. Implement this algorithm in WAVELAB. Hint: Use a fast overlap-add convolution algorithm.

4.5. [1] Let K be the reproducing kernel (4.21) of a windowed Fourier transform.

(a) For any $\Phi \in \mathbf{L}^2(\mathbb{R}^2)$ we define:

$$T\Phi(u_0,\xi_0) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \Phi(u,\xi)\,K(u_0,u,\xi_0,\xi)\,du\,d\xi.$$

Prove that T is an orthogonal projector on the space of functions $\Phi(u,\xi)$ that are windowed Fourier transforms of functions in $\mathbf{L}^2(\mathbb{R})$.

(b) Suppose that for all $(u,\xi) \in \mathbb{R}^2$ we are given $\tilde{S}f(u,\xi) = Q\Big(Sf(u,\xi)\Big)$, which is a quantization of the windowed Fourier coefficients. How can we reduce the norm $\mathbf{L}^2(\mathbb{R}^2)$ of the quantification error $\epsilon(u,\xi) = Sf(u,\xi) - Q\Big(Sf(u,\xi)\Big)$?

4.6. [1] Prove that a scaling function ϕ defined by (4.42) satisfies $\|\phi\| = 1$.

4.7. [2] Let ψ be a real and even wavelet such that $C = \int_0^{+\infty} \omega^{-1}\,\hat{\psi}(\omega)\,d\omega < +\infty$. Prove that

$$\forall f \in \mathbf{L}^2(\mathbb{R}) \;,\; f(t) = \frac{1}{C}\int_0^{+\infty} Wf(t,s)\,\frac{ds}{s^{3/2}}\,. \tag{4.149}$$

4.8. [1] *Analytic Continuation* Let $f \in \mathbf{L}^2(\mathbb{R})$ be a function such that $\hat{f}(\omega) = 0$ for $\omega < 0$. For any complex $z \in \mathbb{C}$ such that $\mathrm{Im}(z) \geq 0$, we define

$$f^{(p)}(z) = \frac{1}{\pi}\int_0^{+\infty} (i\omega)^p\,\hat{f}(\omega)\,e^{iz\omega}\,d\omega\,.$$

(a) Verify that if f is $\mathbf{C}^p$ then $f^{(p)}(t)$ is the derivative of order p of $f(t)$.

(b) Prove that if $\mathrm{Im}(z) > 0$, then $f^{(p)}(z)$ is differentiable relative to the complex variable z. Such a function is said to be *analytic* on the upper half complex plane.

(c) Prove that this analytic extension can be written as a wavelet transform

$$f^{(p)}(x+iy) = y^{-p-1/2}\,Wf(x,y)\,,$$

calculated with an analytic wavelet ψ that you will specify.

4.9. [1] Let $f(t) = \cos(a\cos bt)$. We want to compute precisely the instantaneous frequency of f from the ridges of its windowed Fourier transform. Find a necessary condition on the window support as a function of a and b. If $f(t) = \cos(a\cos bt) + \cos(a\cos bt + ct)$, find a condition on a, b and c in order to measure both instantaneous frequencies with the ridges of a windowed Fourier transform. Verify your calculations with a numerical implementation in WAVELAB.

4.10. [1] *Sound manipulation*

(a) Make a program that synthesizes sounds with the model (4.71) where the amplitudes a_k and phase ϕ_k are calculated from the ridges of a windowed Fourier transform or of a wavelet transform. Test your results on the Tweet and Greasy signals in WAVELAB.

(b) Make a program that modifies the sound duration with the formula (4.72) or which transposes the sound frequency with (4.73).

4.11. [1] Prove that $Pf(u,\xi) = \|f\|^{-2}\,|f(u)|^2\,|\hat{f}(\xi)|^2$ satisfies the marginal properties (4.123,4.124). Why can't we apply the Wigner Theorem 4.7?

4.12. [1] Let g_σ be a Gaussian of variance σ^2. Prove that $P_\theta f(u,\xi) = P_V f \star \theta(u,\xi)$ is a positive distribution if $\theta(u,\xi) = g_\sigma(u)\,g_\beta(\xi)$ with $\sigma\beta \geq 1/2$. Hint: consider a spectrogram calculated with a Gaussian window.

4.13. [2] Let $\{g_n(t)\}_{n\in\mathbb{N}}$ be an orthonormal basis of $\mathbf{L}^2(\mathbb{R})$. Prove that

$$\forall (t,\omega) \in \mathbb{R}^2 \;,\; \sum_{n=0}^{+\infty} P_V g_n(t,\omega) = 1 \;.$$

4.14. [2] Let $f_a(t) = a(t)\exp[i\phi(t)]$ be the analytic part of $f(t)$. Prove that

$$\int_{-\infty}^{+\infty} \Big(\xi - \phi'(t)\Big)^2 P_V f_a(t,\xi)\,d\xi = -\pi a^2(t)\,\frac{d^2 \log a(t)}{dt^2}.$$

4.15. [2] Quadratic affine time-frequency distributions satisfy time shift (4.115), scaling invariance (4.117), and phase invariance (4.114). Prove that any such distribution can be written as an affine smoothing of the Wigner-Ville distribution

$$P_\theta(u,\xi) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \theta\Big(\xi(u-\tau), \frac{\gamma}{\xi}\Big)\,P_V(\tau,\gamma)\,d\tau\,d\gamma, \qquad (4.150)$$

where $\theta(a,b)$ depends upon dimensionless variables.

4.16. [3] To avoid the time-frequency resolution limitations of a windowed Fourier transform, we want to adapt the window size to the signal content. Let $g(t)$ be a window of variance 1. We denote by $S_j f(u,\xi)$ the windowed Fourier transform calculated with the dilated window $g_j(t) = 2^{-j/2} g(2^{-j}t)$. Find a procedure that computes a single map of ridges by choosing a "best" window size at each (u,ξ). One approach is to choose the scale 2^l for each (u,ξ) such that $|S_l f(u,\xi)|^2 = \sup_j |S_j f(u,\xi)|^2$. Test your algorithm on the linear and hyperbolic chirp signals (4.95,4.99). Test it on the Tweet and Greasy signals in WAVELAB.

4.17. [3] The sinusoidal model (4.71) is improved for speech signals by adding a "noise component" $B(t)$ to the partials [245]:

$$F(t) = \sum_{k=1}^{K} a_k(t) \cos \phi_k(t) + B(t). \tag{4.151}$$

Given a signal $f(t)$ that is considered to be a realization of $F(t)$, compute the ridges of a windowed Fourier transform, find the "main" partials and compute their amplitude a_k and phase ϕ_k. These partials are subtracted from the signal. Over intervals of fixed size, the residue is modeled as the realization of an autoregressive process $B(t)$, of order 10 to 15. Use a standard algorithm to compute the parameters of this autoregressive process [60]. Evaluate the audio quality of the sound restored from the calculated model (4.151). Study an application to audio compression by quantizing and coding the parameters of the model.

V

FRAMES

Frame theory analyzes the completeness, stability and redundancy of linear discrete signal representations. A frame is a family of vectors $\{\phi_n\}_{n\in\Gamma}$ that characterizes any signal f from its inner products $\{\langle f,\phi_n\rangle\}_{n\in\Gamma}$. Signal reconstructions from regular and irregular samplings are examples of applications.

Discrete windowed Fourier transforms and discrete wavelet transforms are studied through the frame formalism. These transforms generate signal representations that are not translation invariant, which raises difficulties for pattern recognition applications. Dyadic wavelet transforms maintain translation invariance by sampling only the scale parameter of a continuous wavelet transform. A fast dyadic wavelet transform is calculated with a filter bank algorithm. In computer vision, dyadic wavelet transforms are used for texture discrimination and edge detection.

5.1 FRAME THEORY [2]

5.1.1 Frame Definition and Sampling

The frame theory was originally developed by Duffin and Schaeffer [175] to reconstruct band-limited signals f from irregularly spaced samples $\{f(t_n)\}_{n\in\mathbb{Z}}$. If f has a Fourier transform included in $[-\pi/T,\pi/T]$, we prove as in (3.13) that

$$f(t_n)=\frac{1}{T}\langle f(t),h_T(t-t_n)\rangle \quad \text{with} \quad h_T(t)=\frac{\sin(\pi t/T)}{\pi t/T}. \tag{5.1}$$

This motivated Duffin and Schaeffer to establish general conditions under which one can recover a vector f in a Hilbert space $\mathbf{H}$ from its inner products with a family of vectors $\{\phi_n\}_{n\in\Gamma}$. The index set Γ might be finite or infinite. The following frame definition gives an energy equivalence to invert the operator U defined by

$$\forall n \in \Gamma \quad , \quad Uf[n] = \langle f, \phi_n \rangle. \tag{5.2}$$

Definition 5.1 *The sequence $\{\phi_n\}_{n\in\Gamma}$ is a frame of $\mathbf{H}$ if there exist two constants $A > 0$ and $B > 0$ such that for any $f \in \mathbf{H}$*

$$A\,\|f\|^2 \leq \sum_{n\in\Gamma} |\langle f, \phi_n \rangle|^2 \leq B\,\|f\|^2. \tag{5.3}$$

When $A = B$ the frame is said to be tight.

If the frame condition is satisfied then U is called a frame operator. Section 5.1.2 proves that (5.3) is a necessary and sufficient condition guaranteeing that U is invertible on its image, with a bounded inverse. A frame thus defines a complete and stable signal representation, which may also be redundant. When the frame vectors are normalized $\|\phi_n\| = 1$, this redundancy is measured by the frame bounds A and B. If the $\{\phi_n\}_{n\in\Gamma}$ are linearly independent then it is proved in (5.23) that

$$A \leq 1 \leq B.$$

The frame is an orthonormal basis if and only if $A = B = 1$. This is verified by inserting $f = \phi_n$ in (5.3). If $A > 1$ then the frame is redundant and A can be interpreted as a minimum redundancy factor.

Example 5.1 Let (e_1, e_2) be an orthonormal basis of a two-dimensional plane $\mathbf{H}$. The three vectors

$$\phi_1 = e_1 \ , \ \phi_2 = -\frac{e_1}{2} + \frac{\sqrt{3}}{2} e_2 \ , \ \phi_3 = -\frac{e_1}{2} - \frac{\sqrt{3}}{2} e_2$$

have equal angles of $2\pi/3$ between themselves. For any $f \in \mathbf{H}$

$$\sum_{n=1}^{3} |\langle f, \phi_n \rangle|^2 = \frac{3}{2} \|f\|^2.$$

These three vectors thus define a tight frame with $A = B = 3/2$. The frame bound $3/2$ measures their redundancy in a space of dimension 2.

Example 5.2 For any $0 \leq k < K$, suppose that $\{e_{k,n}\}_{n\in\mathbb{Z}}$ is an orthonormal basis of $\mathbf{H}$. The union of these K orthonormal bases $\{e_{k,n}\}_{n\in\mathbb{Z}, 0\leq k<K}$ is a tight frame with $A = B = K$. Indeed, the energy conservation in an orthonormal basis implies that for any $f \in \mathbf{H}$,

$$\sum_{n\in\mathbb{Z}} |\langle f, e_{k,n} \rangle|^2 = \|f\|^2,$$

hence

$$\sum_{k=0}^{K-1} \sum_{n \in \mathbb{Z}} |\langle f, e_{k,n} \rangle|^2 = K \|f\|^2.$$

Example 5.3 One can verify (Problem 5.8) that a finite set of N vectors $\{\phi_n\}_{1 \leq n \leq N}$ is always a frame of the space $\mathbf{V}$ generated by linear combinations of these vectors. When N increases, the frame bounds A and B may go respectively to 0 and $+\infty$. This illustrates the fact that in infinite dimensional spaces, a family of vectors may be complete and not yield a stable signal representation.

Irregular Sampling Let $\mathbf{U}_T$ be the space of $\mathbf{L}^2(\mathbb{R})$ functions whose Fourier transforms have a support included in $[-\pi/T, \pi/T]$. For a uniform sampling, $t_n = nT$, Proposition 3.2 proves that $\{T^{-1/2} h_T(t - nT)\}_{n \in \mathbb{Z}}$ is an orthonormal basis of $\mathbf{U}_T$. The reconstruction of f from its samples is then given by the sampling Theorem 3.1.

The irregular sampling conditions of Duffin and Schaeffer [175] for constructing a frame were later refined by several researchers [91, 360, 74]. Grochenig proved [197] that if $\lim_{n \to +\infty} t_n = +\infty$ and $\lim_{n \to -\infty} t_n = -\infty$, and if the maximum sampling distance δ satisfies

$$\delta = \sup_{n \in \mathbb{Z}} |t_{n+1} - t_n| < T\,, \tag{5.4}$$

then

$$\left\{ \sqrt{\frac{t_{n+1} - t_{n-1}}{2T}}\, h_T(t - t_n) \right\}_{n \in \mathbb{Z}}$$

is a frame with frame bounds $A \geq (1 - \delta/T)^2$ and $B \leq (1 + \delta/T)^2$. The amplitude factor $2^{-1/2}(t_{n+1} - t_{n-1})^{1/2}$ compensates for the non-uniformity of the density of samples. It attenuates the amplitude of frame vectors where there is a high density of samples. The reconstruction of f requires inverting the frame operator $Uf[n] = \langle f(t), h_T(t - t_n) \rangle$.

5.1.2 Pseudo Inverse

The reconstruction of f from its frame coefficients $Uf[n]$ is calculated with a pseudo inverse. This pseudo inverse is a bounded operator that is expressed with a dual frame. We denote

$$\mathbf{l}^2(\Gamma) = \{x \,:\, \|x\|^2 = \sum_{n \in \Gamma} |x[n]|^2 < +\infty\}\,,$$

and by $\mathbf{Im}\mathrm{U}$ the image space of all Uf with $f \in \mathbf{H}$.

Proposition 5.1 *If $\{\phi_n\}_{n\in\Gamma}$ is a frame whose vectors are linearly dependent, then* **Im**U *is strictly included in* $\mathbf{l}^2(\Gamma)$*, and U admits an infinite number of left inverses* $\tilde{U}^{-1}$*:*

$$\forall f \in \mathbf{H} \ , \ \tilde{U}^{-1}Uf = f. \tag{5.5}$$

Proof[2]. The frame inequality (5.3) guarantees that $\mathbf{Im}\mathrm{U} \subset \mathbf{l}^2(\Gamma)$ since

$$\|Uf\|^2 = \sum_{n\in\Gamma} |\langle f, \phi_n\rangle|^2 \leq B\|f\|^2. \tag{5.6}$$

Since $\{\phi_n\}_{n\in\Gamma}$ is linearly dependent, there exists a non-zero vector $x \in \mathbf{l}^2(\Gamma)$ such that

$$\sum_{n\in\Gamma} x^*[n]\,\phi_n = 0.$$

For any $f \in \mathbf{H}$

$$\sum_{n\in\Gamma} x[n]\,\langle f, \phi_n\rangle = \sum_{n\in\Gamma} x[n]\,Uf[n] = 0.$$

This proves that **Im**U is orthogonal to x and hence that $\mathbf{Im}\mathrm{U} \neq \mathbf{l}^2(\Gamma)$.

A frame operator U is injective (one to one). Indeed, the frame inequality (5.3) guarantees that $Uf = 0$ implies $f = 0$. Its restriction to **Im**U is thus invertible. Let $\mathbf{Im}\mathrm{U}^\perp$ be the orthogonal complement of **Im**U in $\mathbf{l}^2(\Gamma)$. If $\{\phi_n\}_{n\in\Gamma}$ are linearly dependent then $\mathbf{Im}\mathrm{U}^\perp \neq \{0\}$ and the restriction of $\tilde{U}^{-1}$ to $\mathbf{Im}\mathrm{U}^\perp$ may be any arbitrary linear operator. ■

The more redundant the frame $\{\phi_n\}_{n\in\Gamma}$, the larger the orthogonal complement $\mathbf{Im}\mathrm{U}^\perp$ of the image **Im**U. The pseudo inverse $\tilde{U}^{-1}$ is the left inverse that is zero on $\mathbf{Im}\mathrm{U}^\perp$:

$$\forall x \in \mathbf{Im}\mathrm{U}^\perp \ , \ \tilde{U}^{-1}x = 0.$$

In infinite dimensional spaces, the pseudo inverse $\tilde{U}^{-1}$ of an injective operator is not necessarily bounded. This induces numerical instabilities when trying to reconstruct f from Uf. The following theorem proves that a frame operator has a pseudo inverse that is always bounded. We denote by U^* the adjoint of U: $\langle Uf, x\rangle = \langle f, U^*x\rangle$.

Theorem 5.1 (PSEUDO INVERSE) *The pseudo inverse satisfies*

$$\tilde{U}^{-1} = (U^*U)^{-1}U^*. \tag{5.7}$$

It is the left inverse of minimum sup norm. If U is a frame operator with frame bounds A and B then

$$\|\tilde{U}^{-1}\|_S \leq \frac{1}{\sqrt{A}}. \tag{5.8}$$

Proof[2]. To prove that $\tilde{U}^{-1}$ has a minimum sup norm, let us decompose any $x \in \mathbf{l}^2(\Gamma)$ as a sum $x = x_1 + x_2$ with $x_2 \in \mathbf{Im}\mathrm{U}^\perp$ and $x_1 \in \mathbf{Im}\mathrm{U}$. Let $\bar{U}^{-1}$ be an arbitrary left inverse of U. Then

$$\frac{\|\tilde{U}^{-1}x\|}{\|x\|} = \frac{\|\tilde{U}^{-1}x_1\|}{\|x\|} = \frac{\|\bar{U}^{-1}x_1\|}{\|x\|} \leq \frac{\|\bar{U}^{-1}x_1\|}{\|x_1\|}.$$

We thus derive that

$$\|\tilde{U}^{-1}\|_S = \sup_{x\in l^2(\Gamma)-\{0\}} \frac{\|\tilde{U}^{-1}x\|}{\|x\|} \leq \sup_{x\in l^2(\Gamma)-\{0\}} \frac{\|\bar{U}^{-1}x\|}{\|x\|} = \|\bar{U}^{-1}\|_S.$$

Since $x_1 \in \mathbf{Im}\mathrm{U}$, there exists $f \in \mathbf{H}$ such that $x_1 = Uf$. The inequality (5.8) is derived from the frame inequality (5.3) which shows that

$$\|\tilde{U}^{-1}x\| = \|f\| \leq \frac{1}{\sqrt{A}}\|Uf\| \leq \frac{1}{\sqrt{A}}\|x\|.$$

To verify (5.7), we first prove that the self-adjoint operator U^*U is invertible by showing that it is injective and surjective (onto). If $U^*Uf = 0$ then $\langle U^*Uf, f\rangle = 0$ and hence $\langle Uf, Uf\rangle = 0$. Since U is injective then $f = 0$, which proves that U^*U is injective. To prove that the image of U^*U is equal to $\mathbf{H}$ we prove that no non-zero vector can be orthogonal to this image. Suppose that $g \in \mathbf{H}$ is orthogonal to the image of U^*U. In particular $\langle g, U^*Ug\rangle = 0$, so $\langle Ug, Ug\rangle = 0$, which implies that $g = 0$. This proves that U^*U is surjective.

Since U^*U is invertible, proving (5.7) is equivalent to showing that for any x the pseudo inverse satisfies

$$(U^*U)\tilde{U}^{-1}x = U^*x. \tag{5.9}$$

If $x \in \mathbf{Im}\mathrm{U}^\perp$ then $(U^*U)\tilde{U}^{-1}x = 0$ because $\tilde{U}^{-1}x = 0$, and $U^*x = 0$ because

$$\forall f \in \mathbf{H}\ ,\ \langle f, U^*x\rangle = \langle Uf, x\rangle = 0.$$

It thus verifies (5.9) for $x \in \mathbf{Im}\mathrm{U}^\perp$. If $x \in \mathbf{Im}\mathrm{U}$, then $U\tilde{U}^{-1}x = x$ so (5.9) remains valid. We thus derive that (5.9) is satisfied for all $x \in \mathbf{H}$. ■

Dual Frame The pseudo inverse of a frame operator is related to a dual frame family, which is specified by the following theorem.

Theorem 5.2 *Let $\{\phi_n\}_{n\in\mathbb{Z}}$ be a frame with bounds A, B. The dual frame defined by*

$$\tilde{\phi}_n = (U^*U)^{-1}\phi_n$$

satisfies

$$\forall f \in \mathbf{H}\ ,\ \frac{1}{B}\|f\|^2 \leq \sum_{n\in\Gamma} |\langle f, \tilde{\phi}_n\rangle|^2 \leq \frac{1}{A}\|f\|^2, \tag{5.10}$$

and

$$f = \tilde{U}^{-1}Uf = \sum_{n\in\Gamma} \langle f, \phi_n\rangle \tilde{\phi}_n = \sum_{n\in\Gamma} \langle f, \tilde{\phi}_n\rangle \phi_n. \tag{5.11}$$

If the frame is tight (i.e., $A = B$), then $\tilde{\phi}_n = A^{-1}\phi_n$.

Proof[2]. To prove (5.11), we relate U^* to $\{\phi_n\}_{n\in\Gamma}$ and use the expression (5.7) of $\tilde{U}^{-1}$. For any $x \in \mathbf{l}^2(\Gamma)$ and $f \in \mathbf{H}$

$$\langle U^*x, f\rangle = \langle x, Uf\rangle = \sum_{n\in\Gamma} x[n]\langle f, \phi_n\rangle^*.$$

Consequently

$$\langle U^*x, f\rangle = \sum_{n\in\Gamma} \langle x[n]\phi_n, f\rangle,$$

which implies that

$$U^*x = \sum_{n\in\Gamma} x[n]\,\phi_n. \tag{5.12}$$

The pseudo inverse formula (5.7) proves that

$$\tilde{U}^{-1}x = (U^*U)^{-1}U^*x = (U^*U)^{-1}\sum_{n\in\Gamma} x[n]\,\phi_n,$$

so

$$\tilde{U}^{-1}x = \sum_{n\in\Gamma} x[n]\,\tilde{\phi}_n. \tag{5.13}$$

If $x[n] = Uf[n] = \langle f, \phi_n\rangle$ then

$$f = \tilde{U}^{-1}Uf = \sum_{n\in\Gamma} \langle f, \phi_n\rangle\,\tilde{\phi}_n. \tag{5.14}$$

The dual family of vectors $\{\phi_n\}_{n\in\Gamma}$ and $\{\tilde{\phi}_n\}_{n\in\Gamma}$ play symmetrical roles. Indeed (5.14) implies that for any f and g in $\mathbf{H}$,

$$\langle f, g\rangle = \sum_{n\in\Gamma} \langle f, \phi_n\rangle\,\langle \tilde{\phi}_n, g\rangle, \tag{5.15}$$

hence

$$g = \sum_{n\in\Gamma} \langle g, \tilde{\phi}_n\rangle\,\phi_n, \tag{5.16}$$

which proves (5.11).

The expression (5.12) of U^* proves that for $x[n] = Uf[n] = \langle f, \phi_n\rangle$

$$U^*Uf = \sum_{n\in\Gamma} \langle f, \phi_n\rangle\,\phi_n. \tag{5.17}$$

The frame condition (5.3) can thus be rewritten

$$A\,\|f\|^2 \le \langle U^*Uf, f\rangle \le B\,\|f\|^2. \tag{5.18}$$

If $A = B$ then $\langle U^*Uf, f\rangle = A\,\|f\|^2$. Since U^*U is symmetrical, one can show that necessarily $U^*U = A\,Id$ where Id is the identity operator. It thus follows that $\tilde{\phi}_n = (U^*U)^{-1}\phi_n = A^{-1}\phi_n$.

Similarly (5.10) can be rewritten

$$\frac{1}{B}\,\|f\|^2 \le \langle (U^*U)^{-1}f, f\rangle \le \frac{1}{A}\,\|f\|^2. \tag{5.19}$$

because

$$(U^*U)^{-1}f = \sum_{n\in\Gamma} \langle f, \tilde{\phi}_n\rangle\,(U^*U)^{-1}\phi_n = \sum_{n\in\Gamma} \langle f, \tilde{\phi}_n\rangle\,\tilde{\phi}_n.$$

The double inequality (5.19) is derived from (5.18) by applying the following lemma to $L = U^*U$.

Lemma 5.1 *If L is a self-adjoint operator such that there exist $A > 0$ and B satisfying*

$$\forall f \in \mathbf{H} \ , \ A\|f\|^2 \leq \langle Lf, f\rangle \leq B\|f\|^2 \tag{5.20}$$

then L is invertible and

$$\forall f \in \mathbf{H} \ , \ \frac{1}{B}\|f\|^2 \leq \langle L^{-1}f, f\rangle \leq \frac{1}{A}\|f\|^2. \tag{5.21}$$

In finite dimensions, since L is self-adjoint we know that it is diagonalized in an orthonormal basis. The inequality (5.20) proves that its eigenvalues are between A and B. It is therefore invertible with eigenvalues between B^{-1} and A^{-1}, which proves (5.21). In infinite dimensions, the proof is left to the reader. ■

This theorem proves that $\{\tilde{\phi}_n\}_{n\in\Gamma}$ is a dual frame that recovers any $f \in \mathbf{H}$ from its frame coefficients $\{\langle f, \phi_n\rangle\}_{n\in\Gamma}$. If the frame is tight then $\tilde{\phi}_n = A^{-1}\phi_n$, so the reconstruction formula becomes

$$f = \frac{1}{A}\sum_{n\in\Gamma}\langle f, \phi_n\rangle\,\phi_n. \tag{5.22}$$

Biorthogonal Bases A Riesz basis is a frame of vectors that are linearly independent, which implies that $\mathbf{Im}\mathrm{U} = \mathbf{l}^2(\Gamma)$. One can derive from (5.11) that the dual frame $\{\tilde{\phi}_n\}_{n\in\Gamma}$ is also linearly independent. It is called the dual Riesz basis. Inserting $f = \phi_p$ in (5.11) yields

$$\phi_p = \sum_{n\in\Gamma}\langle \phi_p, \tilde{\phi}_n\rangle\,\phi_n,$$

and the linear independence implies that

$$\langle \phi_p, \tilde{\phi}_n\rangle = \delta[p-n].$$

Dual Riesz bases are thus biorthogonal families of vectors. If the basis is normalized (i.e., $\|\phi_n\| = 1$), then

$$A \leq 1 \leq B. \tag{5.23}$$

This is proved by inserting $f = \phi_p$ in the frame inequality (5.10):

$$\frac{1}{B}\|\phi_p\|^2 \leq \sum_{n\in\Gamma}|\langle \phi_p, \tilde{\phi}_n\rangle|^2 = 1 \leq \frac{1}{A}\|\phi_p\|^2.$$

Partial Reconstruction Suppose that $\{\phi_n\}_{n\in\Gamma}$ is a frame of a subspace $\mathbf{V}$ of the whole signal space. The inner products $Uf[n] = \langle f, \phi_n\rangle$ give partial information on f that does not allow us to fully recover f. The best linear mean-square approximation of f computed from these inner products is the orthogonal projection

of f on the space $\mathbf{V}$. This orthogonal projection is computed with the dual frame $\{\tilde{\phi}_n\}_{n\in\Gamma}$ of $\{\phi_n\}_{n\in\Gamma}$ in $\mathbf{V}$:

$$P_{\mathbf{V}} f = \tilde{U}^{-1} U f = \sum_{n\in\Gamma} \langle f, \phi_n \rangle \tilde{\phi}_n. \tag{5.24}$$

To prove that $P_{\mathbf{V}} f$ is the orthogonal projection in $\mathbf{V}$, we verify that $P_{\mathbf{V}} f \in \mathbf{V}$ and that $\langle f - P_{\mathbf{V}} f, \phi_p \rangle = 0$ for all $p \in \Gamma$. Indeed,

$$\langle f - P_{\mathbf{V}} f, \phi_p \rangle = \langle f, \phi_p \rangle - \sum_{n\in\Gamma} \langle f, \phi_n \rangle \langle \tilde{\phi}_n, \phi_p \rangle,$$

and the dual frame property in $\mathbf{V}$ implies that

$$\sum_{n\in\Gamma} \langle \tilde{\phi}_n, \phi_p \rangle \phi_n = \phi_p.$$

Suppose we have a finite number of data measures $\{\langle f, \phi_n \rangle\}_{0\leq n<N}$. Since a finite family $\{\phi_n\}_{0\leq n<N}$ is necessarily a frame of the space $\mathbf{V}$ it generates, the approximation formula (5.24) reconstructs the best linear approximation of f.

5.1.3 Inverse Frame Computations

We describe efficient numerical algorithms to recover a signal f from its frame coefficients $Uf[n] = \langle f, \phi_n \rangle$. If possible, the dual frame vectors are precomputed:

$$\tilde{\phi}_n = (U^* U)^{-1} \phi_n,$$

and we recover each f with the sum

$$f = \sum_{n\in\Gamma} \langle f, \phi_n \rangle \tilde{\phi}_n.$$

In some applications, the frame vectors $\{\phi_n\}_{n\in\Gamma}$ may depend on the signal f, in which case the dual frame vectors $\tilde{\phi}_n$ cannot be computed in advance. For example, the frame (5.1) associated to an irregular sampling depends on the position t_n of each sample. If the sampling grid varies from signal to signal it modifies the frame vectors. It is then highly inefficient to compute the dual frame for each new signal. A more direct approach applies the pseudo inverse to Uf:

$$f = \tilde{U}^{-1} U f = (U^* U)^{-1} (U^* U) f = L^{-1} L f, \tag{5.25}$$

where

$$L f = U^* U f = \sum_{n\in\Gamma} \langle f, \phi_n \rangle \phi_n. \tag{5.26}$$

Whether we precompute the dual frame vectors or apply the pseudo inverse on the frame data, both approaches require an efficient way to compute $f = L^{-1} g$

for some $g \in \mathbf{H}$. Theorems 5.3 and 5.4 describe two iterative algorithms with exponential convergence. The *extrapolated Richardson procedure* is simpler but requires knowing the frame bounds A and B. *Conjugate gradient* iterations converge more quickly when B/A is large, and do not require knowing the values of A and B.

Theorem 5.3 (EXTRAPOLATED RICHARDSON) *Let $g \in \mathbf{H}$. To compute $f = L^{-1}g$ we initialize $f_0 = 0$. Let $\gamma > 0$ be a relaxation parameter. For any $n > 0$, define*

$$f_n = f_{n-1} + \gamma(g - Lf_{n-1}). \tag{5.27}$$

If

$$\delta = \max\{|1-\gamma A|, |1-\gamma B|\} < 1, \tag{5.28}$$

then

$$\|f - f_n\| \leq \delta^n \|f\|, \tag{5.29}$$

and hence $\lim_{n\to+\infty} f_n = f$.

Proof[2]. The induction equation (5.27) can be rewritten

$$f - f_n = f - f_{n-1} - \gamma L(f - f_{n-1}).$$

Let

$$R = Id - \gamma L,$$

$$f - f_n = R(f - f_{n-1}) = R^n(f - f_0) = R^n(f). \tag{5.30}$$

We saw in (5.18) that the frame inequality can be rewritten

$$A\|f\|^2 \leq \langle Lf, f\rangle \leq B\|f\|^2.$$

This implies that $R = I - \gamma L$ satisfies

$$|\langle Rf, f\rangle| \leq \delta \|f\|^2,$$

where δ is given by (5.28). Since R is symmetric, this inequality proves that $\|R\| \leq \delta$. We thus derive (5.29) from (5.30). The error $\|f - f_n\|$ clearly converges to zero if $\delta < 1$. ■

For frame inversion, the extrapolated Richardson algorithm is sometimes called the *frame algorithm* [21]. The convergence rate is maximized when δ is minimum:

$$\delta = \frac{B-A}{B+A} = \frac{1-A/B}{1+A/B},$$

which corresponds to the relaxation parameter

$$\gamma = \frac{2}{A+B}.$$

The algorithm converges quickly if A/B is close to 1. If A/B is small then

$$\delta \approx 1 - 2\frac{A}{B}. \tag{5.31}$$

The inequality (5.29) proves that we obtain an error smaller than ϵ for a number n of iterations, which satisfies:

$$\frac{\|f - f_n\|}{\|f\|} \leq \delta^n = \epsilon.$$

Inserting (5.31) gives

$$n \approx \frac{\log_e \epsilon}{\log_e(1 - 2A/B)} \approx \frac{-B}{2A} \log_e \epsilon. \tag{5.32}$$

The number of iterations thus increases proportionally to the frame bound ratio B/A.

The exact values of A and B are often not known, in which case the relaxation parameter γ must be estimated numerically by trial and error. If an upper bound B_0 of B is known then we can choose $\gamma = 1/B_0$. The algorithm is guaranteed to converge, but the convergence rate depends on A.

The conjugate gradient algorithm computes $f = L^{-1}g$ with a gradient descent along orthogonal directions with respect to the norm induced by the symmetric operator L:

$$\|f\|_L^2 = \|Lf\|^2. \tag{5.33}$$

This L norm is used to estimate the error. Grochenig's [198] implementation of the conjugate gradient algorithm is given by the following theorem.

Theorem 5.4 (CONJUGATE GRADIENT) *Let $g \in \mathbf{H}$. To compute $f = L^{-1}g$ we initialize*

$$f_0 = 0 \ , \ r_0 = p_0 = g \ , \ p_{-1} = 0. \tag{5.34}$$

For any $n \geq 0$, we define by induction

$$\lambda_n = \frac{\langle r_n, p_n \rangle}{\langle p_n, Lp_n \rangle} \tag{5.35}$$

$$f_{n+1} = f_n + \lambda_n p_n \tag{5.36}$$

$$r_{n+1} = r_n - \lambda_n L p_n \tag{5.37}$$

$$p_{n+1} = Lp_n - \frac{\langle Lp_n, Lp_n \rangle}{\langle p_n, Lp_n \rangle} p_n - \frac{\langle Lp_n, Lp_{n-1} \rangle}{\langle p_{n-1}, Lp_{n-1} \rangle} p_{n-1}. \tag{5.38}$$

If $\sigma = \frac{\sqrt{B} - \sqrt{A}}{\sqrt{B} + \sqrt{A}}$ then

$$\|f - f_n\|_L \leq \frac{2\sigma^n}{1 + \sigma^{2n}} \|f\|_L, \tag{5.39}$$

and hence $\lim_{n \to +\infty} f_n = f$.

Proof[2]. We give the main steps of the proof as outlined by Grochenig [198].

Step 1: Let $\mathbf{U}_n$ be the subspace generated by $\{L^j f\}_{1 \le j \le n}$. By induction on n, we derive from (5.38) that $p_j \in \mathbf{U}_n$, for $j < n$.

Step 2: We prove by induction that $\{p_j\}_{0<j<n}$ is an orthogonal basis of $\mathbf{U}_n$ with respect to the inner product $\langle f, h \rangle_L = \langle f, Lh \rangle$. Assuming that $\langle p_n, Lp_j \rangle = 0$, for $j \le n-1$, it can be shown that $\langle p_{n+1}, Lp_j \rangle = 0$, for $j \le n$.

Step 3: We verify that f_n is the orthogonal projection of f onto $\mathbf{U}_n$ with respect to $\langle ., . \rangle_L$ which means that

$$\forall g \in \mathbf{U}_n \ , \ \|f - g\|_L \le \|f - f_n\|_L .$$

Since $f_n \in \mathbf{U}_n$, this requires proving that $\langle f - f_n, p_j \rangle_L = 0$, for $j < n$.

Step 4: We compute the orthogonal projection of f in embedded spaces $\mathbf{U}_n$ of dimension n, and one can verify that $\lim_{n \to +\infty} \|f - f_n\|_L = 0$. The exponential convergence (5.39) is proved in [198].

■

As in the extrapolated Richardson algorithm, the convergence is slower when A/B is small. In this case

$$\sigma = \frac{1 - \sqrt{A/B}}{1 + \sqrt{A/B}} \approx 1 - 2\sqrt{\frac{A}{B}}.$$

The upper bound (5.39) proves that we obtain a relative error

$$\frac{\|f - f_n\|_L}{\|f\|_L} \le \epsilon$$

for a number of iterations

$$n \approx \frac{\log_e \frac{\epsilon}{2}}{\log_e \sigma} \approx \frac{-\sqrt{B}}{2\sqrt{A}} \log_e \frac{\epsilon}{2}.$$

Comparing this result with (5.32) shows that when A/B is small, the conjugate gradient algorithm needs many fewer iterations than the extrapolated Richardson algorithm to compute $f = L^{-1} g$ at a fixed precision.

5.1.4 Frame Projector and Noise Reduction

Frame redundancy is useful in reducing noise added to the frame coefficients. The vector computed with noisy frame coefficients is projected on the image of U to reduce the amplitude of the noise. This technique is used for high precision analog to digital conversion based on oversampling. The following proposition specifies the orthogonal projector on $\mathbf{Im}U$.

Proposition 5.2 *The orthogonal projection from* $\mathbf{l}^2(\Gamma)$ *onto* $\mathbf{Im}U$ *is*

$$Px[n] = U\tilde{U}^{-1}x[n] = \sum_{p \in \Gamma} x[p] \langle \tilde{\phi}_p, \phi_n \rangle . \tag{5.40}$$

Proof[2]. If $x \in \mathbf{Im}U$ then $x = Uf$ and

$$Px = U\tilde{U}^{-1}Uf = Uf = x.$$

If $x \in \mathbf{Im}U^{\perp}$ then $Px = 0$ because $\tilde{U}^{-1}x = 0$. This proves that P is an orthogonal projector on $\mathbf{Im}U$. Since $Uf[n] = \langle f, \phi_n \rangle$ and $\tilde{U}^{-1}x = \sum_{p \in \Gamma} x[p]\,\tilde{\phi}_p$, we derive (5.40). ■

A vector $x[n]$ is a sequence of frame coefficients if and only if $x = Px$, which means that x satisfies the reproducing kernel equation

$$x[n] = \sum_{p \in \Gamma} x[p]\,\langle \tilde{\phi}_p, \phi_n \rangle. \tag{5.41}$$

This equation generalizes the reproducing kernel properties (4.20) and (4.40) of windowed Fourier transforms and wavelet transforms.

Noise Reduction Suppose that each frame coefficient $Uf[n]$ is contaminated by an additive noise $W[n]$, which is a random variable. Applying the projector P gives

$$P(Uf + W) = Uf + PW,$$

with

$$PW[n] = \sum_{p \in \Gamma} W[p]\,\langle \tilde{\phi}_p, \phi_n \rangle.$$

Since P is an orthogonal projector, $\|PW\| \leq \|W\|$. This projector removes the component of W that is in $\mathbf{Im}U^{\perp}$. Increasing the redundancy of the frame reduces the size of $\mathbf{Im}U$ and thus increases $\mathbf{Im}U^{\perp}$, so a larger portion of the noise is removed. If W is a white noise, its energy is uniformly distributed in the space $\mathbf{l}^2(\Gamma)$. The following proposition proves that its energy is reduced by at least A if the frame vectors are normalized.

Proposition 5.3 *Suppose that* $\|\phi_n\| = C$, *for all* $n \in \Gamma$. *If* W *is a zero-mean white noise of variance* $\mathsf{E}\{|W[n]|^2\} = \sigma^2$, *then*

$$\mathsf{E}\{|PW[n]|^2\} \leq \frac{\sigma^2 C^2}{A}. \tag{5.42}$$

If the frame is tight then this inequality is an equality.

Proof[2]. Let us compute

$$\mathsf{E}\{|PW[n]|^2\} = \mathsf{E}\left\{ \left(\sum_{p \in \Gamma} W[p]\,\langle \tilde{\phi}_p, \phi_n \rangle \right) \left(\sum_{l \in \Gamma} W^*[l]\,\langle \tilde{\phi}_l, \phi_n \rangle^* \right) \right\}.$$

Since W is white,

$$\mathsf{E}\{W[p]\,W^*[l]\} = \sigma^2\,\delta[p - l],$$

and therefore

$$\mathsf{E}\{|PW[n]|^2\} = \sigma^2 \sum_{p \in \Gamma} |\langle \tilde{\phi}_p, \phi_n \rangle|^2 \leq \frac{\sigma^2 \|\phi_n\|^2}{A} = \frac{\sigma^2 C^2}{A}.$$

The last inequality is an equality if the frame is tight. ■

Oversampling This noise reduction strategy is used by high precision analog to digital converters. After a low-pass filter, a band-limited analog signal $f(t)$ is uniformly sampled and quantized. In hardware, it is often easier to increase the sampling rate rather than the quantization precision. Increasing the sampling rate introduces a redundancy between the sample values of the band-limited signal. For a wide range of signals, it has been shown that the quantization error is nearly a white noise [194]. It can thus be significantly reduced by a frame projector.

After the low-pass filtering, f belongs to the space $\mathbf{U}_T$ of functions whose Fourier transforms have their support included in $[-\pi/T, \pi/T]$. The Whittaker sampling Theorem 3.1 guarantees perfect reconstruction with a sampling interval T, but f is oversampled with an interval $T_0 = T/K$ that provides K times more coefficients. We verify that the frame projector is then a low-pass filter that reduces by K the energy of the quantization noise.

Proposition 3.2 proves that

$$f(nT_0) = \frac{1}{T} \langle f(t), h_T(t - nT_0) \rangle \text{ with } h_T(t) = \frac{\sin(\pi t/T)}{\pi t/T},$$

and for each $1 \leq k \leq K$ the family $\{h_T(t - kT/K - nT)\}_{n \in \mathbb{Z}}$ is an orthogonal basis of $\mathbf{U}_T$. As a consequence

$$\left\{ \phi_n(t) = h_T(t - nT_0) \right\}_{n \in \mathbb{Z}} = \left\{ h_T\left(t - k\frac{T}{K} - nT\right) \right\}_{1 \leq k \leq K, n \in \mathbb{Z}}$$

is a union of K orthogonal bases, with vectors having a square norm $C^2 = T$. It is therefore a tight frame of $\mathbf{U}_T$ with $A = B = KT = T_0$. Proposition 5.3 proves that the frame projector P reduces the energy of the quantization white noise W of variance σ^2 by a factor K:

$$\mathsf{E}\{|PW[n]|^2\} = \frac{\sigma^2 C^2}{A} = \frac{\sigma^2}{K}. \tag{5.43}$$

The frame $\{\phi_n(t)\}_{n \in \mathbb{Z}}$ is tight so $\tilde{\phi}_n = \frac{1}{T_0}\phi_n$ and (5.40) implies that

$$Px[n] = \frac{1}{T_0} \sum_{p=-\infty}^{+\infty} x[p] \langle h_T(t - pT_0), h_T(t - nT_0) \rangle.$$

This orthogonal projector can thus be rewritten as the convolution

$$Px[n] = x \star h_0[n] \quad \text{with} \quad h_0[n] = \frac{1}{T_0} \langle h_T(t), h_T(t - nT_0) \rangle.$$

One can verify that h_0 is an ideal low-pass filter whose transfer function has a restriction to $[-\pi, \pi]$ defined by $\hat{h}_0 = \mathbf{1}_{[-\pi/K, \pi/K]}$. In this case $\mathbf{Im}\mathrm{U}$ is simply the space of discrete signals whose Fourier transforms have a restriction to $[-\pi, \pi]$ which is non-zero only in $[-\pi/K, \pi/K]$.

The noise can be further reduced if it is not white but if its energy is better concentrated in $\mathbf{Im}U^{\perp}$. This can be done by transforming the quantization noise into a noise whose energy is mostly concentrated at high frequencies. Sigma-Delta modulators produce such quantization noises by integrating the signal before its quantization [82]. To compensate for the integration, the quantized signal is differentiated. This differentiation increases the energy of the quantized noise at high frequencies and reduces its energy at low frequencies. The low-pass filter h_0 thus further reduces the energy of the quantized noise. Several levels of integration and differentiation can be used to better concentrate the quantization noise in the high frequencies, which further reduces its energy after the filtering by h_0 [330].

This oversampling example is analyzed just as well without the frame formalism because the projector is a simple convolution. However, the frame approach is more general and applies to noise removal in more complicated representations such as irregularly oversampled signals or redundant windowed Fourier and wavelet frames [329].

5.2 WINDOWED FOURIER FRAMES [2]

Frame theory gives conditions for discretizing the windowed Fourier transform while retaining a complete and stable representation. The windowed Fourier transform of $f \in \mathbf{L}^2(\mathbb{R})$ is defined in Section 4.2 by

$$Sf(u,\xi) = \langle f, g_{u,\xi} \rangle,$$

with

$$g_{u,\xi}(t) = g(t-u)\, e^{i\xi t}.$$

Setting $\|g\| = 1$ implies that $\|g_{u,\xi}\| = 1$. A discrete windowed Fourier transform representation

$$\{Sf(u_n,\xi_k) = \langle f, g_{u_n,\xi_k} \rangle\}_{(n,k)\in\mathbb{Z}^2}$$

is complete and stable if $\{g_{u_n,\xi_k}\}_{(n,k)\in\mathbb{Z}^2}$ is a frame of $\mathbf{L}^2(\mathbb{R})$.

Intuitively, one can expect that the discrete windowed Fourier transform is complete if the Heisenberg boxes of all atoms $\{g_{u_n,\xi_k}\}_{(n,k)\in\mathbb{Z}^2}$ fully cover the time-frequency plane. Section 4.2 shows that the Heisenberg box of g_{u_n,ξ_k} is centered in the time-frequency plane at (u_n,ξ_k). Its size is independent of u_n and ξ_k. It depends on the time-frequency spread of the window g. A complete cover of the plane is thus obtained by translating these boxes over a uniform rectangular grid, as illustrated in Figure 5.1. The time and frequency parameters (u,ξ) are discretized over a rectangular grid with time and frequency intervals of size u_0 and ξ_0. Let us denote

$$g_{n,k}(t) = g(t-nu_0)\exp(ik\xi_0 t).$$

The sampling intervals (u_0,ξ_0) must be adjusted to the time-frequency spread of g.

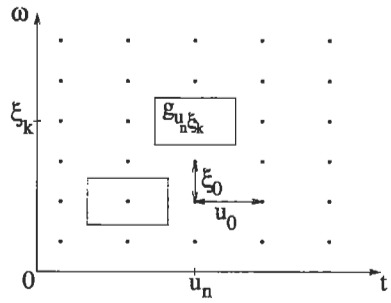


FIGURE 5.1 A windowed Fourier frame is obtained by covering the time-frequency plane with a regular grid of windowed Fourier atoms, translated by $u_n = n u_0$ in time and by $\xi_k = k\xi_0$ in frequency.

Window Scaling Suppose that $\{g_{n,k}\}_{(n,k)\in\mathbb{Z}^2}$ is a frame of $\mathbf{L}^2(\mathbb{R})$ with frame bounds A and B. Let us dilate the window $g_s(t) = s^{-1/2}g(t/s)$. It increases by s the time width of the Heisenberg box of g and reduces by s its frequency width. We thus obtain the same cover of the time-frequency plane by increasing u_0 by s and reducing ξ_0 by s. Let

$$g_{s,n,k}(t) = g_s(t - nsu_0)\exp\left(ik\frac{\xi_0}{s}t\right). \tag{5.44}$$

We prove that $\{g_{s,n,k}\}_{(n,k)\in\mathbb{Z}^2}$ satisfies the same frame inequalities as $\{g_{n,k}\}_{(n,k)\in\mathbb{Z}^2}$, with the same frame bounds A and B, by a change of variable $t' = ts$ in the inner product integrals.

Necessary Conditions Daubechies [21] proved several necessary conditions on g, u_0 and ξ_0 to guarantee that $\{g_{n,k}\}_{(n,k)\in\mathbb{Z}^2}$ is a frame of $\mathbf{L}^2(\mathbb{R})$. We do not reproduce the proofs, but summarize the main results.

Theorem 5.5 (DAUBECHIES) *The windowed Fourier family $\{g_{n,k}\}_{(n,k)\in\mathbb{Z}^2}$ is a frame only if*

$$\frac{2\pi}{u_0\xi_0} \geq 1. \tag{5.45}$$

The frame bounds A and B necessarily satisfy

$$A \leq \frac{2\pi}{u_0\xi_0} \leq B, \tag{5.46}$$

$$\forall t \in \mathbb{R}\ ,\ A \leq \frac{2\pi}{\xi_0}\sum_{n=-\infty}^{+\infty}|g(t-nu_0)|^2 \leq B, \tag{5.47}$$

$$\forall \omega \in \mathbb{R}\ ,\ A \leq \frac{1}{u_0}\sum_{k=-\infty}^{+\infty}|\hat{g}(\omega-k\xi_0)|^2 \leq B. \tag{5.48}$$

The ratio $2\pi/(u_0\xi_0)$ measures the density of windowed Fourier atoms in the time-frequency plane. The first condition (5.45) ensures that this density is greater than 1 because the covering ability of each atom is limited. The inequalities (5.47) and (5.48) are proved in full generality by Chui and Shi [124]. They show that the uniform time translations of g must completely cover the time axis, and the frequency translations of its Fourier transform $\hat{g}$ must similarly cover the frequency axis.

Since all windowed Fourier vectors are normalized, the frame is an orthogonal basis only if $A = B = 1$. The frame bound condition (5.46) shows that this is possible only at the critical sampling density $u_0\xi_0 = 2\pi$. The Balian-Low Theorem [86] proves that g is then either non-smooth or has a slow time decay.

Theorem 5.6 (BALIAN-LOW) *If $\{g_{n,k}\}_{(n,k)\in\mathbb{Z}^2}$ is a windowed Fourier frame with $u_0\xi_0 = 2\pi$, then*

$$\int_{-\infty}^{+\infty} t^2 |g(t)|^2 dt = +\infty \quad or \quad \int_{-\infty}^{+\infty} \omega^2 |\hat{g}(\omega)|^2 d\omega = +\infty. \tag{5.49}$$

This theorem proves that we cannot construct an orthogonal windowed Fourier basis with a differentiable window g of compact support. On the other hand, one can verify that the discontinuous rectangular window

$$g = \frac{1}{\sqrt{u_0}} \mathbf{1}_{[-u_0/2, u_0/2]}$$

yields an orthogonal windowed Fourier basis for $u_0\xi_0 = 2\pi$. This basis is rarely used because of the bad frequency localization of $\hat{g}$.

Sufficient Conditions The following theorem proved by Daubechies [145] gives sufficient conditions on u_0, ξ_0 and g for constructing a windowed Fourier frame.

Theorem 5.7 (DAUBECHIES) *Let us define*

$$\beta(u) = \sup_{0 \le t \le u_0} \sum_{n=-\infty}^{+\infty} |g(t - nu_0)|\,|g(t - nu_0 + u)| \tag{5.50}$$

and

$$\Delta = \sum_{\substack{k=-\infty \\ k\neq 0}}^{+\infty} \left[\beta\left(\frac{2\pi k}{\xi_0}\right)\beta\left(\frac{-2\pi k}{\xi_0}\right)\right]^{1/2}. \tag{5.51}$$

If u_0 and ξ_0 satisfy

$$A_0 = \frac{2\pi}{\xi_0}\left(\inf_{0 \le t \le u_0} \sum_{n=-\infty}^{+\infty} |g(t - nu_0)|^2 - \Delta\right) > 0 \tag{5.52}$$

and

$$B_0 = \frac{2\pi}{\xi_0}\left(\sup_{0\le t\le u_0}\sum_{n=-\infty}^{+\infty}|g(t-nu_0)|^2+\Delta\right) < +\infty, \tag{5.53}$$

then $\{g_{n,k}\}_{(n,k)\in\mathbb{Z}^2}$ *is a frame. The constants* A_0 *and* B_0 *are respectively lower bounds and upper bounds of the frame bounds A and B.*

Observe that the only difference between the sufficient conditions (5.52, 5.53) and the necessary condition (5.47) is the addition and subtraction of Δ. If Δ is small compared to $\inf_{0\le t\le u_0}\sum_{n=-\infty}^{+\infty}|g(t-nu_0)|^2$ then A_0 and B_0 are close to the optimal frame bounds A and B.

Dual Frame Theorem 5.2 proves that the dual windowed frame vectors are

$$\tilde{g}_{n,k} = (U^*U)^{-1}g_{n,k}. \tag{5.54}$$

The following proposition shows that this dual frame is also a windowed Fourier frame, which means that its vectors are time and frequency translations of a new window $\tilde{g}$.

Proposition 5.4 *Dual windowed Fourier vectors can be rewritten*

$$\tilde{g}_{n,k}(t) = \tilde{g}(t-nu_0)\exp(ik\xi_0 t)$$

where $\tilde{g}$ *is the dual window*

$$\tilde{g} = (U^*U)^{-1}g. \tag{5.55}$$

Proof[2]. This result is proved by showing first that $L = U^*U$ commutes with time and frequency translations proportional to u_0 and ξ_0. If $h \in \mathbf{L}^2(\mathbb{R})$ and $h_{m,l}(t) = h(t-mu_0)\exp(il\xi_0 t)$ we verify that

$$Lh_{m,l}(t) = \exp(il\xi_0 t)\,Lh(t-mu_0).$$

Indeed (5.26) shows that

$$Lh_{m,l} = \sum_{(n,k)\in\mathbb{Z}^2}\langle h_{m,l}, g_{n,k}\rangle\, g_{n,k}$$

and a change of variable yields

$$\langle h_{m,l}, g_{n,k}\rangle = \langle h, g_{n-m,k-l}\rangle.$$

Consequently

$$\begin{aligned} Lh_{m,l}(t) &= \sum_{(n,k)\in\mathbb{Z}^2}\langle h, g_{n-m,k-l}\rangle\exp(il\xi_0 t)\,g_{n-m,k-l}(t-mu_0) \\ &= \exp(il\xi_0 t)\,Lh(t-mu_0). \end{aligned}$$

Since L commutes with these translations and frequency modulations we verify that L^{-1} necessarily commutes with the same group operations. Hence

$$\tilde{g}_{n,k}(t) = L^{-1}g_{n,k} = \exp(ik\xi_0)\,L^{-1}g_{0,0}(t-nu_0) = \exp(ik\xi_0)\,\tilde{g}(t-nu_0).$$

■

$u_0\xi_0$	A_0	B_0	B_0/A_0
$\pi/2$	3.9	4.1	1.05
$3\pi/4$	2.5	2.8	1.1
π	1.6	2.4	1.5
$4\pi/3$	0.58	2.1	3.6
1.9π	0.09	2.0	22

Table 5.1 Frame bounds estimated with Theorem 5.7 for the Gaussian window (5.56) and $u_0 = \xi_0$.

Gaussian Window The Gaussian window

$$g(t) = \pi^{-1/4} \exp\left(\frac{-t^2}{2}\right) \tag{5.56}$$

has a Fourier transform $\hat{g}$ that is a Gaussian with the same variance. The time and frequency spreads of this window are identical. We therefore choose equal sampling intervals in time and frequency: $u_0 = \xi_0$. For the same product $u_0\xi_0$ other choices would degrade the frame bounds. If g is dilated by s then the time and frequency sampling intervals must become su_0 and ξ_0/s.

If the time-frequency sampling density is above the critical value: $2\pi/(u_0\xi_0) > 1$, then Daubechies [145] proves that $\{g_{n,k}\}_{(n,k)\in\mathbb{Z}^2}$ is a frame. When $u_0\xi_0$ tends to 2π, the frame bound A tends to 0. For $u_0\xi_0 = 2\pi$, the family $\{g_{n,k}\}_{(n,k)\in\mathbb{Z}^2}$ is complete in $\mathbf{L}^2(\mathbb{R})$, which means that any $f \in \mathbf{L}^2(\mathbb{R})$ is entirely characterized by the inner products $\{\langle f, g_{n,k}\rangle\}_{(n,k)\in\mathbb{Z}^2}$. However, the Balian-Low Theorem 5.6 proves that it cannot be a frame and one can indeed verify that $A = 0$ [145]. This means that the reconstruction of f from these inner products is unstable.

Table 5.1 gives the estimated frame bounds A_0 and B_0 calculated with Theorem 5.7, for different values of $u_0 = \xi_0$. For $u_0\xi_0 = \pi/2$, which corresponds to time and frequency sampling intervals that are half the critical sampling rate, the frame is nearly tight. As expected, $A \approx B \approx 4$, which verifies that the redundancy factor is 4 (2 in time and 2 in frequency). Since the frame is almost tight, the dual frame is approximately equal to the original frame, which means that $\tilde{g} \approx g$. When $u_0\xi_0$ increases we see that A decreases to zero and $\tilde{g}$ deviates more and more from a Gaussian. In the limit $u_0\xi_0 = 2\pi$, the dual window $\tilde{g}$ is a discontinuous function that does not belong to $\mathbf{L}^2(\mathbb{R})$. These results can be extended to discrete window Fourier transforms computed with a discretized Gaussian window [361].

Tight Frames Tight frames are easier to manipulate numerically since the dual frame is equal to the original frame. Daubechies, Grossmann and Meyer [146] give two sufficient conditions for building a window of compact support that generates a tight frame.

Theorem 5.8 (DAUBECHIES, GROSSMANN, MEYER) *Let g be a window whose support is included in $[-\pi/\xi_0, \pi/\xi_0]$. If*

$$\forall t \in \mathbb{R} \ , \ \frac{2\pi}{\xi_0} \sum_{n=-\infty}^{+\infty} |g(t - nu_0)|^2 = A \tag{5.57}$$

then $\{g_{n,k}\}_{(n,k)\in\mathbb{Z}^2}$ is a tight frame with a frame bound equal to A.

The proof is studied in Problem 5.4. If we impose that

$$1 \leq \frac{2\pi}{u_0\xi_0} \leq 2,$$

then only consecutive windows $g(t - nu_0)$ and $g(t-(n+1)u_0)$ have supports that overlap. The design of such windows is studied in Section 8.4.2 for local cosine bases.

5.3 WAVELET FRAMES [2]

Wavelet frames are constructed by sampling the time and scale parameters of a continuous wavelet transform. A real continuous wavelet transform of $f \in \mathbf{L}^2(\mathbb{R})$ is defined in Section 4.3 by

$$Wf(u,s) = \langle f, \psi_{u,s} \rangle,$$

where ψ is a real wavelet and

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right).$$

Imposing $\|\psi\| = 1$ implies that $\|\psi_{u,s}\| = 1$.

Intuitively, to construct a frame we need to cover the time-frequency plane with the Heisenberg boxes of the corresponding discrete wavelet family. A wavelet $\psi_{u,s}$ has an energy in time that is centered at u over a domain proportional to s. Over positive frequencies, its Fourier transform $\hat{\psi}_{u,s}$ has a support centered at a frequency η/s, with a spread proportional to $1/s$. To obtain a full cover, we sample s along an exponential sequence $\{a^j\}_{j\in\mathbb{Z}}$, with a sufficiently small dilation step $a > 1$. The time translation u is sampled uniformly at intervals proportional to the scale a^j, as illustrated in Figure 5.2. Let us denote

$$\psi_{j,n}(t) = \frac{1}{\sqrt{a^j}} \psi\left(\frac{t - nu_0 a^j}{a^j}\right).$$

We give necessary and sufficient conditions on ψ, a and u_0 so that $\{\psi_{j,n}\}_{(j,n)\in\mathbb{Z}^2}$ is a frame of $\mathbf{L}^2(\mathbb{R})$.

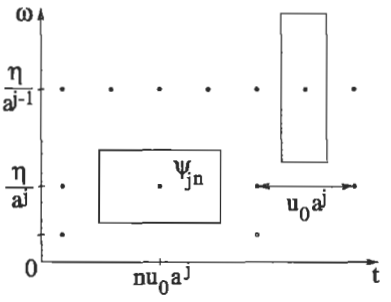


FIGURE 5.2 The Heisenberg box of a wavelet $\psi_{j,n}$ scaled by $s = a^j$ has a time and frequency width proportional respectively to a^j and a^{-j}. The time-frequency plane is covered by these boxes if u_0 and a are sufficiently small.

Necessary Conditions We suppose that ψ is real, normalized, and satisfies the admissibility condition of Theorem 4.3:

$$C_\psi = \int_0^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega} d\omega < +\infty. \tag{5.58}$$

Theorem 5.9 (DAUBECHIES) *If $\{\psi_{j,n}\}_{(j,n)\in\mathbb{Z}^2}$ is a frame of $\mathbf{L}^2(\mathbb{R})$ then the frame bounds satisfy*

$$A \leq \frac{C_\psi}{u_0 \log_e a} \leq B, \tag{5.59}$$

$$\forall \omega \in \mathbb{R} - \{0\} \ , \ A \leq \frac{1}{u_0} \sum_{j=-\infty}^{+\infty} |\hat{\psi}(a^j \omega)|^2 \leq B. \tag{5.60}$$

The condition (5.60) imposes that the Fourier axis is covered by wavelets dilated by $\{a^j\}_{j\in\mathbb{Z}}$. It is proved in [124, 21]. Section 5.5 explains that this condition is sufficient for constructing a complete and stable signal representation if the time parameter u is not sampled. The inequality (5.59), which relates the sampling density $u_0 \log_e a$ to the frame bounds, is proved in [21]. It shows that the frame is an orthonormal basis if and only if

$$A = B = \frac{C_\psi}{u_0 \log_e a} = 1.$$

Chapter 7 constructs wavelet orthonormal bases of $\mathbf{L}^2(\mathbb{R})$ with regular wavelets of compact support.

Sufficient Conditions The following theorem proved by Daubechies [21] provides a lower and upper bound for the frame bounds A and B, depending on ψ, u_0 and a.

Theorem 5.10 (DAUBECHIES) *Let us define*

$$\beta(\xi) = \sup_{1\leq|\omega|\leq a} \sum_{j=-\infty}^{+\infty} |\hat{\psi}(a^j\omega)|\,|\hat{\psi}(a^j\omega+\xi)| \tag{5.61}$$

and

$$\Delta = \sum_{\substack{k=-\infty \\ k\neq 0}}^{+\infty} \left[\beta\left(\frac{2\pi k}{u_0}\right)\beta\left(\frac{-2\pi k}{u_0}\right)\right]^{1/2}.$$

If u_0 and a are such that

$$A_0 = \frac{1}{u_0}\left(\inf_{1\leq|\omega|\leq a} \sum_{j=-\infty}^{+\infty} |\hat{\psi}(a^j\omega)|^2 - \Delta\right) > 0, \tag{5.62}$$

and

$$B_0 = \frac{1}{u_0}\left(\sup_{1\leq|\omega|\leq a} \sum_{j=-\infty}^{+\infty} |\hat{\psi}(a^j\omega)|^2 + \Delta\right) < +\infty, \tag{5.63}$$

then $\{\psi_{j,n}\}_{(j,n)\in\mathbb{Z}^2}$ is a frame of $\mathbf{L}^2(\mathbb{R})$. The constants A_0 and B_0 are respectively lower and upper bounds of the frame bounds A and B.

The sufficient conditions (5.62) and (5.63) are similar to the necessary condition (5.60). If Δ is small relative to $\inf_{1\leq|\omega|\leq a}\sum_{j=-\infty}^{+\infty}|\hat{\psi}(a^j\omega)|^2$ then A_0 and B_0 are close to the optimal frame bounds A and B. For a fixed dilation step a, the value of Δ decreases when the time sampling interval u_0 decreases.

Dual Frame Theorem 5.2 gives a general formula for computing the dual wavelet frame vectors

$$\tilde{\psi}_{j,n} = (U^*U)^{-1}\psi_{j,n}. \tag{5.64}$$

One could reasonably hope that the dual functions $\tilde{\psi}_{j,n}$ would be obtained by scaling and translating a dual wavelet $\tilde{\psi}$. The sad reality is that this is generally not true. In general the operator U^*U does not commute with dilations by a^j, so $(U^*U)^{-1}$ does not commute with these dilations either. On the other hand, one can prove that $(U^*U)^{-1}$ commutes with translations by na^ju_0, which means that

$$\tilde{\psi}_{j,n}(t) = \tilde{\psi}_{j,0}(t - na^ju_0). \tag{5.65}$$

The dual frame $\{\tilde{\psi}_{j,n}\}_{(j,n)\in\mathbb{Z}^2}$ is thus obtained by calculating each elementary function $\tilde{\psi}_{j,0}$ with (5.64), and translating them with (5.65). The situation is much simpler for tight frames, where the dual frame is equal to the original wavelet frame.

a	u_0	A_0	B_0	B_0/A_0
2	0.25	13.091	14.183	1.083
2	0.5	6.546	7.092	1.083
2	1.0	3.223	3.596	1.116
2	1.5	0.325	4.221	12.986
$2^{\frac{1}{2}}$	0.25	27.273	27.278	1.0002
$2^{\frac{1}{2}}$	0.5	13.673	13.639	1.0002
$2^{\frac{1}{2}}$	1.0	6.768	6.870	1.015
$2^{\frac{1}{2}}$	1.75	0.517	7.276	14.061
$2^{\frac{1}{4}}$	0.25	54.552	54.552	1.0000
$2^{\frac{1}{4}}$	0.5	27.276	27.276	1.0000
$2^{\frac{1}{4}}$	1.0	13.586	13.690	1.007
$2^{\frac{1}{4}}$	1.75	2.928	12.659	4.324

Table 5.2 Estimated frame bounds for the Mexican hat wavelet computed with Theorem 5.10 [21].

Mexican Hat Wavelet The normalized second derivative of a Gaussian is

$$\psi(t) = \frac{2}{\sqrt{3}} \pi^{-1/4} (t^2 - 1) \exp\left(\frac{-t^2}{2}\right). \tag{5.66}$$

Its Fourier transform is

$$\hat{\psi}(\omega) = -\frac{\sqrt{8}\pi^{1/4}\omega^2}{\sqrt{3}} \exp\left(\frac{-\omega^2}{2}\right).$$

The graph of these functions is shown in Figure 4.6.

The dilation step a is generally set to be $a = 2^{1/v}$ where v is the number of intermediate scales (voices) for each octave. Table 5.2 gives the estimated frame bounds A_0 and B_0 computed by Daubechies [21] with the formula of Theorem 5.10. For $v \geq 2$ voices per octave, the frame is nearly tight when $u_0 \leq 0.5$, in which case the dual frame can be approximated by the original wavelet frame. As expected from (5.59), when $A \approx B$

$$A \approx B \approx \frac{C_\psi}{u_0 \log_e a} = \frac{v}{u_0} C_\psi \log_2 e.$$

The frame bounds increase proportionally to v/u_0. For $a = 2$, we see that A_0 decreases brutally from $u_0 = 1$ to $u_0 = 1.5$. For $u_0 = 1.75$ the wavelet family is not a frame anymore. For $a = 2^{1/2}$, the same transition appears for a larger u_0.

5.4 TRANSLATION INVARIANCE [1]

In pattern recognition, it is important to construct signal representations that are translation invariant. When a pattern is translated, its numerical descriptors should

be translated but not modified. Indeed, a pattern search is particularly difficult if its representation depends on its location. Continuous wavelet transforms and windowed Fourier transforms provide translation-invariant representations, but uniformly sampling the translation parameter destroys this translation invariance.

Continuous Transforms Let $f_\tau(t) = f(t-\tau)$ be a translation of $f(t)$ by τ. The wavelet transform can be written as a convolution product:

$$Wf(u,s) = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) dt = f \star \bar{\psi}_s(u)$$

with $\bar{\psi}_s(t) = s^{-1/2}\psi(-t/s)$. It is therefore translation invariant:

$$Wf_\tau(u,s) = f_\tau \star \bar{\psi}_s(u) = Wf(u-\tau,s).$$

A windowed Fourier transform can also be written as a linear filtering

$$Sf(u,\xi) = \int_{-\infty}^{+\infty} f(t)\, g(t-u)\, e^{-it\xi}\, dt = e^{-iu\xi}\, f \star \bar{g}_\xi(u),$$

with $\bar{g}_\xi(t) = g(-t)\, e^{it\xi}$. Up to a phase shift, it is also translation invariant:

$$Sf_\tau(u,\xi) = e^{-iu\xi}\, f \star g_\xi(u-\tau) = e^{-i\tau\xi}\, Sf(u-\tau,\xi).$$

Frame Sampling A wavelet frame

$$\psi_{j,n}(t) = \frac{1}{\sqrt{a^j}} \psi\left(\frac{t - n a^j u_0}{a^j}\right)$$

yields inner products that sample the continuous wavelet transform at time intervals $a^j u_0$:

$$\langle f, \psi_{j,n} \rangle = f \star \bar{\psi}_{a^j}(n a^j u_0) = Wf(n a^j u_0, a^j).$$

Translating f by τ gives

$$\langle f_\tau, \psi_{j,n} \rangle = f \star \bar{\psi}_{a^j}(n a^j u_0 - \tau) = Wf(n a^j u_0 - \tau, a^j).$$

If the sampling interval $a^j u_0$ is large relative to the rate of variation of $f \star \bar{\psi}_{a^j}(t)$, then the coefficients $\langle f, \psi_{j,n} \rangle$ and $\langle f_\tau, \psi_{j,n} \rangle$ may take very different values that are not translated with respect to one another. This is illustrated in Figure 5.3. This problem is particularly acute for wavelet orthogonal bases where u_0 is maximum. The orthogonal wavelet coefficients of f_τ may be very different from the coefficients of f. The same translation distortion phenomena appear in windowed Fourier frames.

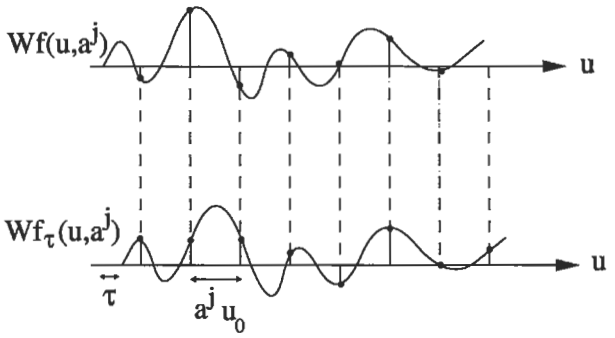


FIGURE 5.3 If $f_\tau(t) = f(t-\tau)$ then $Wf_\tau(u,a^j) = Wf(u-\tau,a^j)$. Uniformly sampling $Wf_\tau(u,a^j)$ and $Wf(u,a^j)$ at $u = na^j u_0$ may yield very different values if $\tau \neq ku_0a^j$.

Translation-Invariant Representations There are several strategies for maintaining the translation invariance of a wavelet transform. If the sampling interval $a^j u_0$ is small enough then the samples of $f \star \bar{\psi}_{a^j}(t)$ are approximately translated when f is shifted. The dyadic wavelet transform presented in Section 5.5 is a translation-invariant representation that does not sample the translation factor u. This creates a highly redundant signal representation.

To reduce the representation size while maintaining translation invariance, one can use an adaptive sampling scheme, where the sampling grid is automatically translated when the signal is translated. For each scale a^j, $Wf(u,a^j) = f \star \bar{\psi}_{a^j}(u)$ can be sampled at locations u where $|Wf(a^j,u)|$ is locally maximum. The resulting representation is translation invariant since the local maxima positions are translated when f and hence $f \star \bar{\psi}_{a^j}$ are translated. This adaptive sampling is studied in Section 6.2.2.

5.5 DYADIC WAVELET TRANSFORM [2]

To construct a translation-invariant wavelet representation, the scale s is discretized but not the translation parameter u. The scale is sampled along a dyadic sequence $\{2^j\}_{j\in\mathbb{Z}}$, to simplify the numerical calculations. Fast computations with filter banks are presented in the next two sections. An application to computer vision and texture discrimination is described in Section 5.5.3.

The dyadic wavelet transform of $f \in \mathbf{L}^2(\mathbb{R})$ is defined by

$$Wf(u,2^j) = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{2^j}} \psi\left(\frac{t-u}{2^j}\right) dt = f \star \bar{\psi}_{2^j}(u), \tag{5.67}$$

with

$$\bar{\psi}_{2^j}(t) = \psi_{2^j}(-t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{-t}{2^j}\right).$$

The following proposition proves that if the frequency axis is completely covered by dilated dyadic wavelets, as illustrated by Figure 5.4, then it defines a complete and stable representation.

Theorem 5.11 *If there exist two constants $A > 0$ and $B > 0$ such that*

$$\forall \omega \in \mathbb{R} - \{0\} \ , \ A \leq \sum_{j=-\infty}^{+\infty} |\hat{\psi}(2^j \omega)|^2 \leq B, \tag{5.68}$$

then

$$A \|f\|^2 \leq \sum_{j=-\infty}^{+\infty} \frac{1}{2^j} \|Wf(u, 2^j)\|^2 \leq B \|f\|^2. \tag{5.69}$$

If $\tilde{\psi}$ satisfies

$$\forall \omega \in \mathbb{R} - \{0\} \ , \ \sum_{j=-\infty}^{+\infty} \hat{\psi}^*(2^j \omega) \hat{\tilde{\psi}}(2^j \omega) = 1, \tag{5.70}$$

then

$$f(t) = \sum_{j=-\infty}^{+\infty} \frac{1}{2^j} Wf(., 2^j) \star \tilde{\psi}_{2^j}(t). \tag{5.71}$$

Proof[2]. The Fourier transform of $f_j(u) = Wf(u, 2^j)$ with respect to u is derived from the convolution formula (5.67):

$$\hat{f}_j(\omega) = \sqrt{2^j}\, \hat{\psi}^*(2^j \omega)\, \hat{f}(\omega). \tag{5.72}$$

The condition (5.68) implies that

$$A |\hat{f}(\omega)|^2 \leq \sum_{j=-\infty}^{+\infty} \frac{1}{2^j} |\hat{f}_j(\omega)|^2 \leq B |\hat{f}(\omega)|^2.$$

Integrating each side of this inequality with respect to ω and applying the Parseval equality (2.25) yields (5.69).

Equation (5.71) is proved by taking the Fourier transform on both sides and inserting (5.70) and (5.72).

■

The energy equivalence (5.69) proves that the normalized dyadic wavelet transform operator

$$Uf[j, u] = \frac{1}{\sqrt{2^j}} Wf(u, 2^j) = \left\langle f, \frac{1}{\sqrt{2^j}} \psi_{2^j}(t - u) \right\rangle$$

satisfies frame inequalities. There exist an infinite number of reconstructing wavelets $\tilde{\psi}$ that verify (5.70). They correspond to different left inverses of U, calculated with (5.71). If we choose

$$\hat{\tilde{\psi}}(\omega) = \frac{\hat{\psi}(\omega)}{\sum_{j=-\infty}^{+\infty} |\hat{\psi}(2^j \omega)|^2}, \tag{5.73}$$

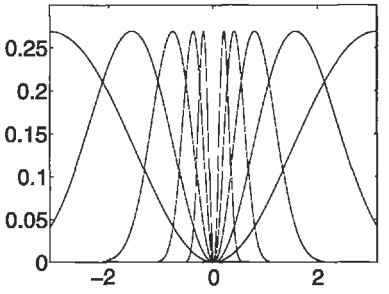


FIGURE 5.4 Scaled Fourier transforms $|\hat{\psi}(2^j\omega)|^2$ computed with (5.84), for $1 \leq j \leq 5$ and $\omega \in [-\pi, \pi]$.

then one can verify that the left inverse is the pseudo inverse $\tilde{U}^{-1}$. Figure 5.5 gives a dyadic wavelet transform computed over 5 scales with the quadratic spline wavelet shown in Figure 5.6.

5.5.1 Wavelet Design

A discrete dyadic wavelet transform can be computed with a fast filter bank algorithm if the wavelet is appropriately designed. The synthesis of these dyadic wavelets is similar to the construction of biorthogonal wavelet bases, explained in Section 7.4. All technical issues related to the convergence of infinite cascades of filters are avoided in this section. Reading Chapter 7 first is necessary for understanding the main results.

Let h and g be a pair of finite impulse response filters. Suppose that h is a low-pass filter whose transfer function satisfies $\hat{h}(0) = \sqrt{2}$. As in the case of orthogonal and biorthogonal wavelet bases, we construct a scaling function whose Fourier transform is

$$\hat{\phi}(\omega) = \prod_{p=1}^{+\infty} \frac{\hat{h}(2^{-p}\omega)}{\sqrt{2}} = \frac{1}{\sqrt{2}} \hat{h}\left(\frac{\omega}{2}\right) \hat{\phi}\left(\frac{\omega}{2}\right). \tag{5.74}$$

We suppose here that this Fourier transform is a finite energy function so that $\phi \in \mathbf{L}^2(\mathbb{R})$. The corresponding wavelet ψ has a Fourier transform defined by

$$\hat{\psi}(\omega) = \frac{1}{\sqrt{2}} \hat{g}\left(\frac{\omega}{2}\right) \hat{\phi}\left(\frac{\omega}{2}\right). \tag{5.75}$$

Proposition 7.2 proves that both ϕ and ψ have a compact support because h and g have a finite number of non-zero coefficients. The number of vanishing moments of ψ is equal to the number of zeroes of $\hat{\psi}(\omega)$ at $\omega = 0$. Since $\hat{\phi}(0) = 1$, (5.75) implies that it is also equal to the number of zeros of $\hat{g}(\omega)$ at $\omega = 0$.

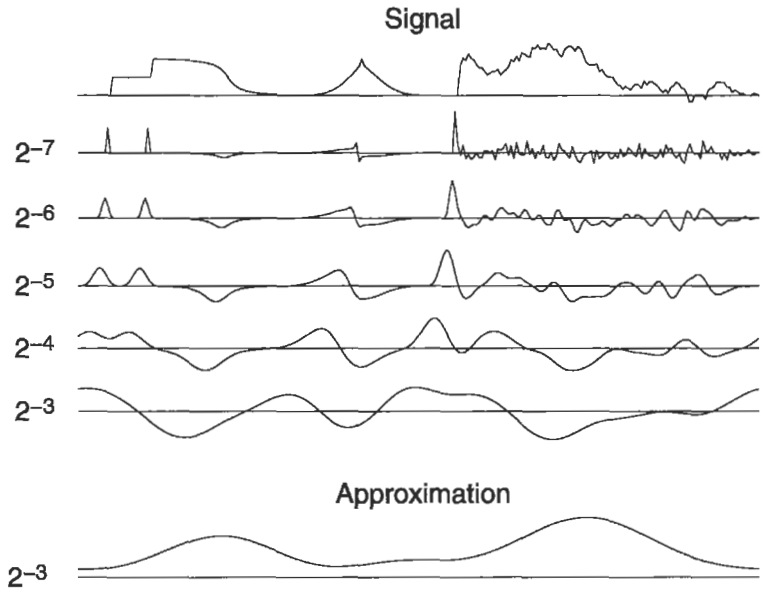


FIGURE 5.5 Dyadic wavelet transform $Wf(u,2^j)$ computed at scales $2^{-7} \leq 2^j \leq 2^{-3}$ with the filter bank algorithm of Section 5.5.2, for signal defined over $[0,1]$. The bottom curve carries the lower frequencies corresponding to scales larger than 2^{-3}.

Reconstructing Wavelets Reconstructing wavelets that satisfy (5.70) are calculated with a pair of finite impulse response dual filters $\tilde{h}$ and $\tilde{g}$. We suppose that the following Fourier transform has a finite energy:

$$\widehat{\tilde{\phi}}(\omega) = \prod_{p=1}^{+\infty} \frac{\widehat{\tilde{h}}(2^{-p}\omega)}{\sqrt{2}} = \frac{1}{\sqrt{2}} \widehat{\tilde{h}}\left(\frac{\omega}{2}\right) \widehat{\tilde{\phi}}\left(\frac{\omega}{2}\right). \tag{5.76}$$

Let us define

$$\widehat{\tilde{\psi}}(\omega) = \frac{1}{\sqrt{2}} \widehat{\tilde{g}}\left(\frac{\omega}{2}\right) \widehat{\tilde{\phi}}\left(\frac{\omega}{2}\right). \tag{5.77}$$

The following proposition gives a sufficient condition to guarantee that $\widehat{\tilde{\psi}}$ is the Fourier transform of a reconstruction wavelet.

Proposition 5.5 *If the filters satisfy*

$$\forall \omega \in [-\pi, \pi] \ , \ \widehat{\tilde{h}}(\omega)\hat{h}^*(\omega) + \widehat{\tilde{g}}(\omega)\hat{g}^*(\omega) = 2 \tag{5.78}$$

then

$$\forall \omega \in \mathbb{R} - \{0\} \ , \ \sum_{j=-\infty}^{+\infty} \hat{\psi}^*(2^j\omega)\widehat{\tilde{\psi}}(2^j\omega) = 1. \tag{5.79}$$

Proof[2]. The Fourier transform expressions (5.75) and (5.77) prove that

$$\widehat{\tilde{\psi}}(\omega)\,\hat{\psi}^*(\omega) = \frac{1}{2}\,\widehat{\tilde{g}}\left(\frac{\omega}{2}\right)\hat{g}^*\left(\frac{\omega}{2}\right)\widehat{\tilde{\phi}}\left(\frac{\omega}{2}\right)\hat{\phi}^*\left(\frac{\omega}{2}\right).$$

Equation (5.78) implies

$$\begin{aligned}\widehat{\tilde{\psi}}(\omega)\,\hat{\psi}^*(\omega) &= \frac{1}{2}\left[2-\widehat{\tilde{h}}\left(\frac{\omega}{2}\right)\hat{h}^*\left(\frac{\omega}{2}\right)\right]\widehat{\tilde{\phi}}\left(\frac{\omega}{2}\right)\hat{\phi}^*\left(\frac{\omega}{2}\right)\\ &= \widehat{\tilde{\phi}}\left(\frac{\omega}{2}\right)\hat{\phi}^*\left(\frac{\omega}{2}\right) - \widehat{\tilde{\phi}}(\omega)\,\hat{\phi}^*(\omega).\end{aligned}$$

Hence

$$\sum_{j=-l}^{k}\widehat{\tilde{\psi}}(2^j\omega)\,\hat{\psi}^*(2^j\omega) = \hat{\phi}^*(2^{-l}\omega)\,\widehat{\tilde{\phi}}(2^{-l}\omega) - \hat{\phi}^*(2^k\omega)\,\widehat{\tilde{\phi}}(2^k\omega).$$

Since $\hat{g}(0)=0$, (5.78) implies $\widehat{\tilde{h}}(0)\,\hat{h}^*(0)=2$. We also impose that $\hat{h}(0)=\sqrt{2}$ so one can derive from (5.74,5.76) that $\widehat{\tilde{\phi}}(0)=\hat{\phi}^*(0)=1$. Since ϕ and $\tilde{\phi}$ belong to $\mathbf{L}^1(\mathbb{R})$, $\hat{\phi}$ and $\widehat{\tilde{\phi}}$ are continuous, and the Riemann-Lebesgue lemma (Problem 2.6) proves that $|\hat{\phi}(\omega)|$ and $|\widehat{\tilde{\phi}}(\omega)|$ decrease to zero when ω goes to ∞. For $\omega \neq 0$, letting k and l go to $+\infty$ yields (5.79). ■

Observe that (5.78) is the same as the unit gain condition (7.122) for biorthogonal wavelets. The aliasing cancellation condition (7.121) of biorthogonal wavelets is not required because the wavelet transform is not sampled in time.

Finite Impulse Response Solution Let us shift h and g to obtain causal filters. The resulting transfer functions $\hat{h}(\omega)$ and $\hat{g}(\omega)$ are polynomials in $e^{-i\omega}$. We suppose that these polynomials have no common zeros. The Bezout Theorem 7.6 on polynomials proves that if $P(z)$ and $Q(z)$ are two polynomials of degree n and l, with no common zeros, then there exists a unique pair of polynomials $\tilde{P}(z)$ and $\tilde{Q}(z)$ of degree $l-1$ and $n-1$ such that

$$P(z)\,\tilde{P}(z) + Q(z)\,\tilde{Q}(z) = 1. \tag{5.80}$$

This guarantees the existence of $\widehat{\tilde{h}}(\omega)$ and $\widehat{\tilde{g}}(\omega)$ that are polynomials in $e^{-i\omega}$ and satisfy (5.78). These are the Fourier transforms of the finite impulse response filters $\tilde{h}$ and $\tilde{g}$. One must however be careful because the resulting scaling function $\widehat{\tilde{\phi}}$ in (5.76) does not necessarily have a finite energy.

Spline Dyadic Wavelets A *box spline* of degree m is a translation of $m+1$ convolutions of $\mathbf{1}_{[0,1]}$ with itself. It is centered at $t=1/2$ if m is even and at $t=0$ if m is odd. Its Fourier transform is

$$\hat{\phi}(\omega) = \left(\frac{\sin(\omega/2)}{\omega/2}\right)^{m+1}\exp\left(\frac{-i\epsilon\omega}{2}\right) \quad \text{with} \quad \epsilon = \begin{cases} 1 & \text{if } m \text{ is even} \\ 0 & \text{if } m \text{ is odd} \end{cases}, \tag{5.81}$$

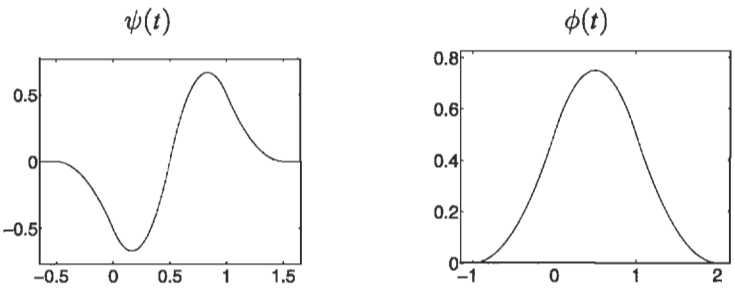


FIGURE 5.6 Quadratic spline wavelet and scaling function.

so

$$\hat{h}(\omega) = \sqrt{2}\frac{\hat{\phi}(2\omega)}{\hat{\phi}(\omega)} = \sqrt{2}\left(\cos\frac{\omega}{2}\right)^{m+1}\exp\left(\frac{-i\epsilon\omega}{2}\right). \tag{5.82}$$

We construct a wavelet that has one vanishing moment by choosing $\hat{g}(\omega) = O(\omega)$ in the neighborhood of $\omega = 0$. For example

$$\hat{g}(\omega) = -i\sqrt{2}\sin\frac{\omega}{2}\exp\left(\frac{-i\epsilon\omega}{2}\right). \tag{5.83}$$

The Fourier transform of the resulting wavelet is

$$\hat{\psi}(\omega) = \frac{1}{\sqrt{2}}\hat{g}\left(\frac{\omega}{2}\right)\hat{\phi}\left(\frac{\omega}{2}\right) = \frac{-i\omega}{4}\left(\frac{\sin(\omega/4)}{\omega/4}\right)^{m+2}\exp\left(\frac{-i\omega(1+\epsilon)}{4}\right). \tag{5.84}$$

It is the first derivative of a box spline of degree $m+1$ centered at $t = (1+\epsilon)/4$. For $m = 2$, Figure 5.6 shows the resulting quadratic splines ϕ and ψ. The dyadic admissibility condition (5.68) is verified numerically for $A = 0.505$ and $B = 0.522$.

To design dual scaling functions $\tilde{\phi}$ and wavelets $\tilde{\psi}$ which are splines, we choose $\widehat{\tilde{h}} = \hat{h}$. As a consequence, $\phi = \tilde{\phi}$ and the reconstruction condition (5.78) implies that

$$\widehat{\tilde{g}}(\omega) = \frac{2 - |\hat{h}(\omega)|^2}{\hat{g}^*(\omega)} = -i\sqrt{2}\exp\left(\frac{-i\omega}{2}\right)\sin\frac{\omega}{2}\sum_{n=0}^{m}\left(\cos\frac{\omega}{2}\right)^{2n}. \tag{5.85}$$

Table 5.3 gives the corresponding filters for $m = 2$.

5.5.2 "Algorithme à Trous"

Suppose that the scaling functions and wavelets ϕ, ψ, $\tilde{\phi}$ and $\tilde{\psi}$ are designed with the filters h, g, $\tilde{h}$ and $\tilde{g}$. A fast dyadic wavelet transform is calculated with a filter bank algorithm called in French the *algorithme à trous*, introduced by Holschneider, Kronland-Martinet, Morlet and Tchamitchian [212]. It is similar to a fast biorthogonal wavelet transform, without subsampling [308, 261].

n	$h[n]/\sqrt{2}$	$\tilde{h}[n]/\sqrt{2}$	$g[n]/\sqrt{2}$	$\tilde{g}[n]/\sqrt{2}$
−2				−0.03125
−1	0.125	0.125		−0.21875
0	0.375	0.375	−0.5	−0.6875
1	0.375	0.375	0.5	0.6875
2	0.125	0.125		0.21875
3				0.03125

Table 5.3 Coefficients of the filters computed from their transfer functions (5.82, 5.83, 5.85) for $m = 2$. These filters generate the quadratic spline scaling functions and wavelets shown in Figure 5.6.

Let $\dot{f}(t)$ be a continuous time signal characterized by N samples at a distance N^{-1} over $[0,1]$. Its dyadic wavelet transform can only be calculated at scales $1 > 2^j \geq N^{-1}$. To simplify the description of the filter bank algorithm, it is easier to consider the signal $f(t) = \dot{f}(N^{-1}t)$, whose samples have distance equal to 1. A change of variable in the dyadic wavelet transform integral shows that $W\dot{f}(u,2^j) = N^{-1/2} Wf(Nu, N2^j)$. We thus concentrate on the dyadic wavelet transform of f, from which the dyadic wavelet transform of $\dot{f}$ is easily derived.

Fast Dyadic Transform We suppose that the samples $a_0[n]$ of the input discrete signal are not equal to $f(n)$ but to a local average of f in the neighborhood of $t = n$. Indeed, the detectors of signal acquisition devices perform such an averaging. The samples $a_0[n]$ are written as averages of $f(t)$ weighted by the scaling kernels $\phi(t-n)$:

$$a_0[n] = \langle f(t), \phi(t-n) \rangle = \int_{-\infty}^{+\infty} f(t)\, \phi(t-n)\, dt.$$

This is further justified in Section 7.3.1. For any $j \geq 0$, we denote

$$a_j[n] = \langle f(t), \phi_{2^j}(t-n) \rangle \quad \text{with} \quad \phi_{2^j}(t) = \frac{1}{\sqrt{2^j}} \phi\left(\frac{t}{2^j}\right).$$

The dyadic wavelet coefficients are computed for $j > 0$ over the integer grid

$$d_j[n] = Wf(n, 2^j) = \langle f(t), \psi_{2^j}(t-n) \rangle.$$

For any filter $x[n]$, we denote by $x_j[n]$ the filters obtained by inserting $2^j - 1$ zeros between each sample of $x[n]$. Its Fourier transform is $\hat{x}(2^j\omega)$. Inserting zeros in the filters creates holes (*trous* in French). Let $\bar{x}_j[n] = x_j[-n]$. The next proposition gives convolution formulas that are cascaded to compute a dyadic wavelet transform and its inverse.

Proposition 5.6 *For any $j \geq 0$,*

$$a_{j+1}[n] = a_j \star \bar{h}_j[n] \quad , \quad d_{j+1}[n] = a_j \star \bar{g}_j[n] \ , \tag{5.86}$$

and

$$a_j[n] = \frac{1}{2}\left(a_{j+1} \star \tilde{h}_j[n] + d_{j+1} \star \tilde{g}_j[n]\right). \tag{5.87}$$

Proof[2]. *Proof of (5.86).* Since

$$a_{j+1}[n] = f \star \bar{\phi}_{2^{j+1}}(n) \quad \text{and} \quad d_{j+1}[n] = f \star \bar{\psi}_{2^{j+1}}(n),$$

we verify with (3.3) that their Fourier transforms are respectively

$$\hat{a}_{j+1}(\omega) = \sum_{k=-\infty}^{+\infty} \hat{f}(\omega + 2k\pi)\, \hat{\phi}^*_{2^{j+1}}(\omega + 2k\pi)$$

and

$$\hat{d}_{j+1}(\omega) = \sum_{k=-\infty}^{+\infty} \hat{f}(\omega + 2k\pi)\, \hat{\psi}^*_{2^{j+1}}(\omega + 2k\pi).$$

The properties (5.76) and (5.77) imply that

$$\hat{\phi}_{2^{j+1}}(\omega) = \sqrt{2^{j+1}}\, \hat{\phi}(2^{j+1}\omega) = \hat{h}(2^j\omega)\sqrt{2^j}\, \hat{\phi}(2^j\omega),$$

$$\hat{\psi}_{2^{j+1}}(\omega) = \sqrt{2^{j+1}}\, \hat{\psi}(2^{j+1}\omega) = \hat{g}(2^j\omega)\sqrt{2^j}\, \hat{\phi}(2^j\omega).$$

Since $j \geq 0$, both $\hat{h}(2^j\omega)$ and $\hat{g}(2^j\omega)$ are 2π periodic, so

$$\hat{a}_{j+1}(\omega) = \hat{h}^*(2^j\omega)\, \hat{a}_j(\omega) \text{ and } \hat{d}_{j+1}(\omega) = \hat{g}^*(2^j\omega)\, \hat{a}_j(\omega). \tag{5.88}$$

These two equations are the Fourier transforms of (5.86).

Proof of (5.87). Equations (5.88) imply

$$\begin{aligned} \hat{a}_{j+1}(\omega)\,\overline{\hat{h}}(2^j\omega) + \hat{d}_{j+1}(\omega)\,\overline{\hat{g}}(2^j\omega) &= \\ \hat{a}_j(\omega)\,\hat{h}^*(2^j\omega)\,\overline{\hat{h}}(2^j\omega) &+ \hat{a}_j(\omega)\,\hat{g}^*(2^j\omega)\,\overline{\hat{g}}(2^j\omega). \end{aligned}$$

Inserting the reconstruction condition (5.78) proves that

$$\hat{a}_{j+1}(\omega)\,\overline{\hat{h}}(2^j\omega) + \hat{d}_{j+1}(\omega)\,\overline{\hat{g}}(2^j\omega) = 2\,\hat{a}_j(\omega),$$

which is the Fourier transform of (5.87). ■

The dyadic wavelet representation of a_0 is defined as the set of wavelet coefficients up to a scale 2^J plus the remaining low-frequency information a_J:

$$\left[\{d_j\}_{1 \leq j \leq J}\,,\; a_J\right]. \tag{5.89}$$

It is computed from a_0 by cascading the convolutions (5.86) for $0 \leq j < J$, as illustrated in Figure 5.7(a). The dyadic wavelet transform of Figure 5.5 is calculated with this filter bank algorithm. The original signal a_0 is recovered from its wavelet representation (5.89) by iterating (5.87) for $J > j \geq 0$, as illustrated in Figure 5.7(b).

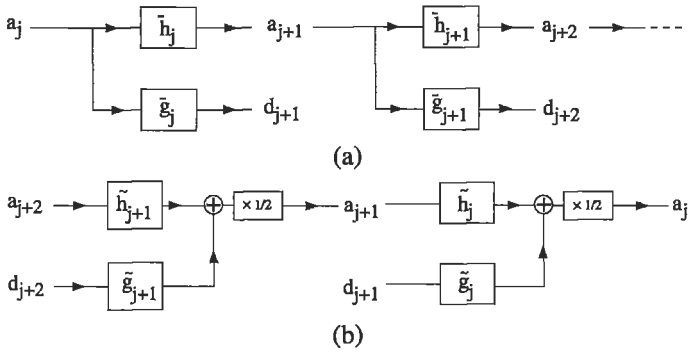


FIGURE 5.7 (a): The dyadic wavelet coefficients are computed by cascading convolutions with dilated filters $\bar{h}_j$ and $\bar{g}_j$. (b): The original signal is reconstructed through convolutions with $\tilde{h}_j$ and $\tilde{g}_j$. A multiplication by $1/2$ is necessary to recover the next finer scale signal a_j.

If the input signal $a_0[n]$ has a finite size of N samples, the convolutions (5.86) are replaced by circular convolutions. The maximum scale 2^J is then limited to N, and for $J = \log_2 N$ one can verify that $a_J[n]$ is constant and equal to $N^{-1/2}\sum_{n=0}^{N-1} a_0[n]$. Suppose that h and g have respectively K_h and K_g non-zero samples. The "dilated" filters h_j and g_j have the same number of non-zero coefficients. The number of multiplications needed to compute a_{j+1} and d_{j+1} from a_j or the reverse is thus equal to $(K_h + K_g)N$. For $J = \log_2 N$, the dyadic wavelet representation (5.89) and its inverse are thus calculated with $(K_h + K_g)N\log_2 N$ multiplications and additions.

5.5.3 Oriented Wavelets for a Vision [3]

Image processing applications of dyadic wavelet transforms are motivated by many physiological and computer vision studies. Textures can be synthesized and discriminated with oriented two-dimensional wavelet transforms. Section 6.3 relates multiscale edges to the local maxima of a wavelet transform.

Oriented Wavelets In two dimensions, a dyadic wavelet transform is computed with several mother wavelets $\{\psi^k\}_{1\le k\le K}$ which often have different spatial orientations. For $x = (x_1, x_2)$, we denote

$$\psi_{2^j}^k(x_1,x_2) = \frac{1}{2^j}\psi^k\left(\frac{x_1}{2^j},\frac{x_2}{2^j}\right) \quad \text{and} \quad \bar{\psi}_{2^j}^k(x) = \psi_{2^j}^k(-x).$$

The wavelet transform of $f \in \mathbf{L}^2(\mathbb{R}^2)$ in the direction k is defined at the position $u = (u_1, u_2)$ and at the scale 2^j by

$$W^k f(u,2^j) = \langle f(x), \psi_{2^j}^k(x-u)\rangle = f \star \bar{\psi}_{2^j}^k(u). \tag{5.90}$$

As in Theorem 5.11, one can prove that the two-dimensional wavelet transform is a complete and stable signal representation if there exist $A > 0$ and B such that

$$\forall \omega = (\omega_1, \omega_2) \in \mathbb{R}^2 - \{(0,0)\} \ , \ A \leq \sum_{k=1}^{K} \sum_{j=-\infty}^{+\infty} |\hat{\psi}^k(2^j \omega)|^2 \leq B. \tag{5.91}$$

Then there exist reconstruction wavelets $\{\tilde{\psi}^k\}_{1 \leq k \leq K}$ whose Fourier transforms satisfy

$$\sum_{j=-\infty}^{+\infty} \frac{1}{2^{2j}} \sum_{k=1}^{K} \widehat{\tilde{\psi}^k}(2^j \omega) \hat{\psi}^{k*}(2^j \omega) = 1, \tag{5.92}$$

which yields

$$f(x) = \sum_{j=-\infty}^{+\infty} \frac{1}{2^{2j}} \sum_{k=1}^{K} W^k f(.,2^j) \star \tilde{\psi}^k_{2^j}(x). \tag{5.93}$$

Wavelets that satisfy (5.91) are called *dyadic wavelets*.

Families of oriented wavelets along any angle α can be designed as a linear expansion of K mother wavelets [312]. For example, a wavelet in the direction α may be defined as the partial derivative of order p of a window $\theta(x)$ in the direction of the vector $\vec{n} = (\cos\alpha, \sin\alpha)$:

$$\psi^\alpha(x) = \frac{\partial^p \theta(x)}{\partial \vec{n}^p} = \left(\cos\alpha \frac{\partial}{\partial x_1} + \sin\alpha \frac{\partial}{\partial x_2} \right)^p \theta(x).$$

This partial derivative is a linear expansion of $K = p+1$ mother wavelets

$$\psi^\alpha(x) = \sum_{k=0}^{p} \begin{pmatrix} p \\ k \end{pmatrix} (\cos\alpha)^k (\sin\alpha)^{p-k} \psi^k(x) \ , \tag{5.94}$$

with

$$\psi^k(x) = \frac{\partial^p \theta(x)}{\partial x_1^k \partial x_2^{p-k}} \quad \text{for } 0 \leq k \leq p.$$

For appropriate windows θ, these $p+1$ partial derivatives define a family of dyadic wavelets. In the direction α, the wavelet transform $W^\alpha f(u,2^j) = f \star \bar{\psi}^\alpha_{2^j}(u)$ is computed from the $p+1$ components $W^k f(u,2^j) = f \star \bar{\psi}^k_{2^j}(u)$ with the expansion (5.94). Section 6.3 uses such oriented wavelets, with $p = 1$, to detect the multiscale edges of an image.

Gabor Wavelets In the cat's visual cortex, Hubel and Wiesel [215] discovered a class of cells, called simple cells, whose responses depend on the frequency and orientation of the visual stimuli. Numerous physiological experiments [283] have shown that these cells can be modeled as linear filters, whose impulse responses have been measured at different locations of the visual cortex. Daugmann [149]

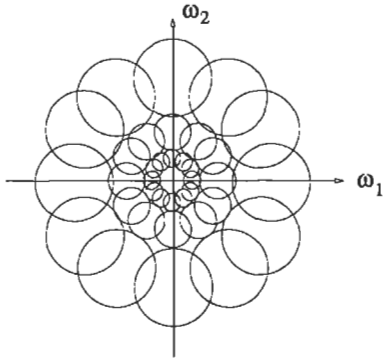


FIGURE 5.8 Each circle represents the frequency support of a dyadic wavelet $\hat{\psi}_{2^j}^k$. This support size is proportional to 2^{-j} and its position rotates when k is modified.

showed that these impulse responses can be approximated by *Gabor wavelets*, obtained with a Gaussian window $g(x_1,x_2)$ multiplied by a sinusoidal wave:

$$\psi^k(x_1,x_2) = g(x_1,x_2)\exp[-i\eta(x_1\cos\alpha_k + x_2\sin\alpha_k)].$$

The position, the scale and the orientation α_k of this wavelet depend on the cortical cell. These findings suggest the existence of some sort of wavelet transform in the visual cortex, combined with subsequent non-linearities [284]. The "physiological" wavelets have a frequency resolution on the order of 1–1.5 octaves, and are thus similar to dyadic wavelets.

Let $\hat{g}(\omega_1,\omega_2)$ be the Fourier transform of $g(x_1,x_2)$. Then

$$\hat{\psi}_{2^j}^k(\omega_1,\omega_2) = \sqrt{2^j}\,\hat{g}(2^j\omega_1 - \eta\cos\alpha_k, 2^j\omega_2 - \eta\sin\alpha_k).$$

In the Fourier plane, the energy of this Gabor wavelet is mostly concentrated around $(2^{-j}\eta\cos\alpha_k, 2^{-j}\eta\sin\alpha_k)$, in a neighborhood proportional to 2^{-j}. Figure 5.8 shows a cover of the frequency plane by such dyadic wavelets. The bandwidth of $\hat{g}(\omega_1,\omega_2)$ and η must be adjusted to satisfy (5.91).

Texture Discrimination Despite many attempts, there are no appropriate mathematical models for "homogeneous image textures." The notion of texture homogeneity is still defined with respect to our visual perception. A texture is said to be homogeneous if it is preattentively perceived as being homogeneous by a human observer.

The texton theory of Julesz [231] was a first important step in understanding the different parameters that influence the perception of textures. The orientation of texture elements and their frequency content seem to be important clues for discrimination. This motivated early researchers to study the repartition of texture

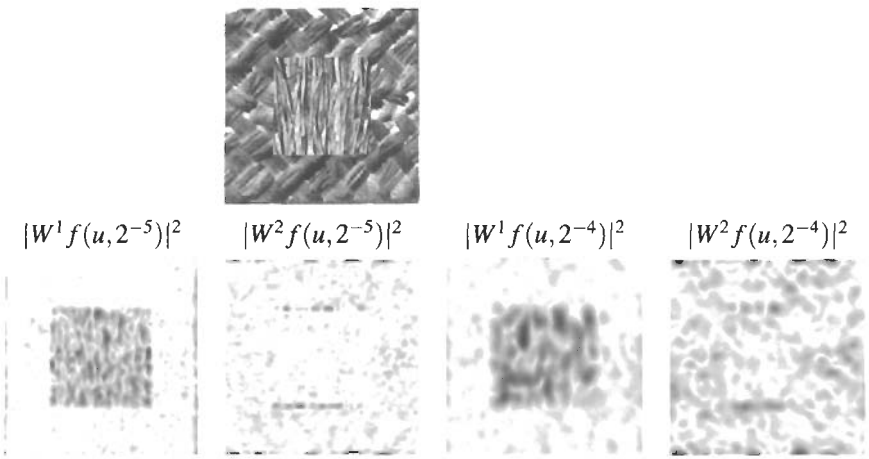


FIGURE 5.9 Gabor wavelet transform $|W^k f(u,2^j)|^2$ of a texture patch, at the scales 2^{-4} and 2^{-5}, along two orientations α_k respectively equal to 0 and $\pi/2$ for $k=1$ and $k=2$. The darker a pixel, the larger the wavelet coefficient amplitude.

energy in the Fourier domain [85]. For segmentation purposes, it is however necessary to localize texture measurements over neighborhoods of varying sizes. The Fourier transform was thus replaced by localized energy measurements at the output of filter banks that compute a wavelet transform [224, 244, 285, 334]. Besides the algorithmic efficiency of this approach, this model is partly supported by physiological studies of the visual cortex.

Since $W^k f(u,2^j) = \langle f(x), \psi_{2^j}^k(x-u)\rangle$, we derive that $|W^k f(u,2^j)|^2$ measures the energy of f in a spatial neighborhood of u of size 2^j and in a frequency neighborhood of $(2^{-j}\eta\cos\alpha_k, 2^{-j}\eta\sin\alpha_k)$ of size 2^{-j}. Varying the scale 2^j and the angle α_k modifies the frequency channel [100]. The wavelet transform energy $|W^k f(u,2^j)|^2$ is large when the angle α_k and scale 2^j match the orientation and scale of high energy texture components in the neighborhood of u. The amplitude of $|W^k f(u,2^j)|^2$ can thus be used to discriminate textures. Figure 5.9 shows the dyadic wavelet transform of two textures, computed along horizontal and vertical orientations, at the scales 2^{-4} and 2^{-5} (the image support is normalized to $[0,1]^2$). The central texture has more energy along horizontal high frequencies than the peripheric texture. These two textures are therefore discriminated by the wavelet oriented with $\alpha_k = 0$ whereas the other wavelet corresponding $\alpha_k = \pi/2$ produces similar responses for both textures.

For segmentation, one must design an algorithm that aggregates the wavelet responses at all scales and orientations in order to find the boundaries of homogeneous textured regions. Both clustering procedures and detection of sharp transitions over wavelet energy measurements have been used to segment the image [224, 285, 334]. These algorithms work well experimentally but rely on ad

hoc parameter settings.

A homogeneous texture can be modeled as a realization of a stationary process, but the main difficulty is to find the characteristics of this process that play a role in texture discrimination. Texture synthesis experiments [277, 313] show that Markov random field processes constructed over grids of wavelet coefficients offer a promising mathematical framework for understanding texture discrimination.

5.6 PROBLEMS

5.1. [1] Prove that if $K \in \mathbb{Z} - \{0\}$ then $\{e_k[n] = \exp(i2\pi kn/(KN))\}_{0 \le k < KN}$ is a tight frame of $\mathbb{C}^N$. Compute the frame bound.

5.2. [1] Prove that if $K \in \mathbb{R} - \{0\}$ then $\{e_k(t) = \exp(i2\pi knt/K)\}_{k \in \mathbb{Z}}$ is a tight frame of $\mathbf{L}^2[0,1]$. Compute the frame bound.

5.3. [1] Let $\hat{g} = \mathbf{1}_{[-\omega_0, \omega_0]}$. Prove that $\{g(t - nu_0) \exp(i2k\pi t/u_0)\}_{(k,n) \in \mathbb{Z}^2}$ is an orthonormal basis of $\mathbf{L}^2(\mathbb{R})$.

5.4. [1] Let $g_{n,k}(t) = g(t - nu_0) \exp(ik\xi_0 t)$, where g is a window whose support is included in $[-\pi/\xi_0, \pi/\xi_0]$.

(a) Prove that $|g(t - nu_0)|^2 f(t) = \sum_{k=-\infty}^{+\infty} \langle f, g_{n,k} \rangle g_{n,k}(t)$.
(b) Prove Theorem 5.8.

5.5. [1] Compute the trigonometric polynomials $\widehat{\tilde{h}}(\omega)$ and $\widehat{\tilde{g}}(\omega)$ of minimum degree that satisfy (5.78) for the spline filters (5.82, 5.83) with $m = 2$. Compute $\tilde{\phi}$ with WAVELAB. Is it a finite energy function?

5.6. [1] Compute a cubic spline dyadic wavelet with 2 vanishing moments using the filter h defined by (5.82) for $m = 3$, with a filter g having 3 non-zero coefficients. Compute in WAVELAB the dyadic wavelet transform of the Lady signal with this new wavelet. Calculate $\tilde{g}[n]$ if $\tilde{h}[n] = h[n]$.

5.7. [1] Let $\{g(t - nu_0) \exp(ik\xi_0 t)\}_{(n,k) \in \mathbb{Z}^2}$ be a windowed Fourier frame defined by $g(t) = \pi^{-1/4} \exp(-t^2/2)$ with $u_0 = \xi_0$ and $u_0 \xi_0 < 2\pi$. With the conjugate gradient algorithm of Theorem 5.4, compute in MATLAB the window $\tilde{g}(t)$ that generates the dual frame, for the values of $u_0 \xi_0$ in Table 5.1. Compare $\tilde{g}$ with g and explain your result. Verify numerically that when $\xi_0 u_0 = 2\pi$ then $\tilde{g}$ is a discontinuous function that does not belong to $\mathbf{L}^2(\mathbb{R})$.

5.8. [1] Prove that a finite set of N vectors $\{\phi_n\}_{1 \le n \le N}$ is always a frame of the space $\mathbf{V}$ generated by linear combinations of these vectors. With an example, show that the frame bounds A and B may go respectively to 0 and $+\infty$ when N goes to $+\infty$.

5.9. [2] *Sigma-Delta converter* A signal $f(t)$ is sampled and quantized. We suppose that $\hat{f}$ has a support in $[-\pi/T, \pi/T]$.

(a) Let $x[n] = f(nT/K)$. Show that if $\omega \in [-\pi, \pi]$ then $\hat{x}(\omega) \neq 0$ only if $\omega \in [-\pi/K, \pi/K]$.
(b) Let $\tilde{x}[n] = Q(x[n])$ be the quantized samples. We now consider $x[n]$ as a random vector, and we model the error $x[n] - \tilde{x}[n] = W[n]$ as a white noise process of variance σ^2. Find the filter $h[n]$ that minimizes

$$\epsilon = \mathsf{E}\{\|\tilde{x} \star h - x\|^2\} ,$$

and compute this minimum as a function of σ^2 and K. Compare your result with (5.43).

(c) Let $\hat{h}_p(\omega) = (1-e^{-i\omega})^{-p}$ be the transfer function of a discrete integration of order p. We quantize $\tilde{x}[n] = Q(x \star h_p[n])$. Find the filter $h[n]$ that minimizes $\epsilon = \mathrm{E}\{\|\tilde{x} \star h - x\|^2\}$, and compute this minimum as a function of σ^2, K and p. For a fixed oversampling factor K, how can we reduce this error?

5.10. [2] Let ψ be a dyadic wavelet that satisfies (5.68). Let $\mathbf{l}^2(\mathbf{L}^2(\mathbb{R}))$ be the space of sequences $\{g_j(u)\}_{j\in\mathbf{Z}}$ such that $\sum_{j=-\infty}^{+\infty} \|g_j\|^2 < +\infty$.

(a) Verify that if $f \in \mathbf{L}^2(\mathbb{R})$ then $\{Wf(u,2^j)\}_{j\in\mathbf{Z}} \in \mathbf{l}^2(\mathbf{L}^2(\mathbb{R}))$. Let $\tilde{\psi}$ be defined by

$$\hat{\tilde{\psi}}(\omega) = \frac{\hat{\psi}(\omega)}{\sum_{j=-\infty}^{+\infty} |\hat{\psi}(2^j\omega)|^2},$$

and W^{-1} be the operator defined by

$$W^{-1}\{g_j(u)\}_{j\in\mathbf{Z}} = \sum_{j=-\infty}^{+\infty} \frac{1}{2^j}\, g_j \star \tilde{\psi}_{2^j}(t).$$

Prove that W^{-1} is the pseudo inverse of W in $\mathbf{l}^2(\mathbf{L}^2(\mathbb{R}))$.

(b) Verify that $\tilde{\psi}$ has the same number of vanishing moments as ψ.

(c) Let $\mathbf{V}$ be the subspace of $\mathbf{l}^2(\mathbf{L}^2(\mathbb{R}))$ that regroups all the dyadic wavelet transforms of functions in $\mathbf{L}^2(\mathbb{R})$. Compute the orthogonal projection of $\{g_j(u)\}_{j\in\mathbf{Z}}$ in $\mathbf{V}$.

5.11. [1] Prove that if there exist $A > 0$ and $B \geq 0$ such that

$$A\,(2 - |\hat{h}(\omega)|^2) \leq |\hat{g}(\omega)|^2 \leq B\,(2 - |\hat{h}(\omega)|^2), \tag{5.95}$$

and if ϕ defined in (5.74) belongs to $\mathbf{L}^2(\mathbb{R})$, then the wavelet ψ given by (5.75) is a dyadic wavelet.

5.12. [2] *Zak transform* The Zak transform associates to any $f \in \mathbf{L}^2(\mathbb{R})$

$$Zf(u,\xi) = \sum_{l=-\infty}^{+\infty} e^{i2\pi l\xi} f(u-l)\,.$$

(a) Prove that it is a unitary operator from $\mathbf{L}^2(\mathbb{R})$ to $\mathbf{L}^2[0,1]^2$:

$$\int_{-\infty}^{+\infty} f(t)\, g^*(t)\, dt = \int_0^1 \int_0^1 Zf(u,\xi)\, Zg^*(u,\xi)\, du\, d\xi\,,$$

by verifying that for $g = \mathbf{1}_{[0,1]}$ it transforms the orthogonal basis $\{g_{n,k}(t) = g(t-n)\exp(i2\pi kt)\}_{(n,k)\in\mathbf{Z}^2}$ of $\mathbf{L}^2(\mathbb{R})$ into an orthonormal basis of $\mathbf{L}^2[0,1]^2$.

(b) Prove that the inverse Zak transform is defined by

$$\forall h \in \mathbf{L}^2[0,1]^2\ ,\ Z^{-1}h(u) = \int_0^1 h(u,\xi)\, d\xi.$$

(c) Prove that if $g \in \mathbf{L}^2(\mathbb{R})$ then $\{g(t-n)\exp(i2\pi kt)\}_{(n,k)\in\mathbb{Z}^2}$ is a frame of $\mathbf{L}^2(\mathbb{R})$ if and only if there exist $A > 0$ and B such that

$$\forall (u,\xi) \in [0,1]^2 \ , \ A \leq |Zg(u,\xi)|^2 \leq B \ , \tag{5.96}$$

where A and B are the frame bounds.

(d) Prove that if (5.96) holds then the dual window $\tilde{g}$ of the dual frame is defined by $Z\tilde{g}(u,\xi) = 1/Zg^*(u,\xi)$.

5.13. [3] Suppose that $\hat{f}$ has a support in $[-\pi/T, \pi/T]$. Let $\{f(t_n)\}_{n\in\mathbb{Z}}$ be irregular samples that satisfy (5.4). With an inverse frame algorithm based on the conjugate gradient Theorem 5.4, implement in MATLAB a procedure that computes $\{f(nT)\}_{n\in\mathbb{Z}}$ (from which f can be recovered with the sampling Theorem 3.1). Analyze the convergence rate of the conjugate gradient algorithm as a function of δ. What happens if the condition (5.4) is not satisfied?

5.14. [3] Develop a texture classification algorithm with a two-dimensional Gabor wavelet transform using four oriented wavelets. The classification procedure can be based on "feature vectors" that provide local averages of the wavelet transform amplitude at several scales, along these four orientations [224, 244, 285, 334].

VI

WAVELET ZOOM

A wavelet transform can focus on localized signal structures with a zooming procedure that progressively reduces the scale parameter. Singularities and irregular structures often carry essential information in a signal. For example, discontinuities in the intensity of an image indicate the presence of edges in the scene. In electrocardiograms or radar signals, interesting information also lies in sharp transitions. We show that the local signal regularity is characterized by the decay of the wavelet transform amplitude across scales. Singularities and edges are detected by following the wavelet transform local maxima at fine scales.

Non-isolated singularities appear in complex signals such as multifractals. In recent years, Mandelbrot led a broad search for multifractals, showing that they are hidden in almost every corner of nature and science. The wavelet transform takes advantage of multifractal self-similarities, in order to compute the distribution of their singularities. This singularity spectrum is used to analyze multifractal properties. Throughout the chapter, the wavelets are real functions.

6.1 LIPSCHITZ REGULARITY [1]

To characterize singular structures, it is necessary to precisely quantify the local regularity of a signal $f(t)$. Lipschitz exponents provide uniform regularity measurements over time intervals, but also at any point v. If f has a singularity at v, which means that it is not differentiable at v, then the Lipschitz exponent at v characterizes this singular behavior.

The next section relates the uniform Lipschitz regularity of f over $\mathbb{R}$ to the

asymptotic decay of the amplitude of its Fourier transform. This global regularity measurement is useless in analyzing the signal properties at particular locations. Section 6.1.3 studies zooming procedures that measure local Lipschitz exponents from the decay of the wavelet transform amplitude at fine scales.

6.1.1 Lipschitz Definition and Fourier Analysis

The Taylor formula relates the differentiability of a signal to local polynomial approximations. Suppose that f is m times differentiable in $[v-h, v+h]$. Let p_v be the Taylor polynomial in the neighborhood of v:

$$p_v(t) = \sum_{k=0}^{m-1} \frac{f^{(k)}(v)}{k!} (t-v)^k. \tag{6.1}$$

The Taylor formula proves that the approximation error

$$\epsilon_v(t) = f(t) - p_v(t)$$

satisfies

$$\forall t \in [v-h, v+h] \ , \ |\epsilon_v(t)| \leq \frac{|t-v|^m}{m!} \sup_{u \in [v-h, v+h]} |f^m(u)|. \tag{6.2}$$

The m^{th} order differentiability of f in the neighborhood of v yields an upper bound on the error $\epsilon_v(t)$ when t tends to v. The Lipschitz regularity refines this upper bound with non-integer exponents. Lipschitz exponents are also called *Hölder* exponents in the mathematical literature.

Definition 6.1 (LIPSCHITZ) • *A function f is pointwise Lipschitz $\alpha \geq 0$ at v, if there exist $K > 0$, and a polynomial p_v of degree $m = \lfloor \alpha \rfloor$ such that*

$$\forall t \in \mathbb{R} \ , \ |f(t) - p_v(t)| \leq K |t-v|^\alpha. \tag{6.3}$$

- *A function f is uniformly Lipschitz α over $[a,b]$ if it satisfies (6.3) for all $v \in [a,b]$, with a constant K that is independent of v.*
- *The Lipschitz regularity of f at v or over $[a,b]$ is the sup of the α such that f is Lipschitz α.*

At each v the polynomial $p_v(t)$ is uniquely defined. If f is $m = \lfloor \alpha \rfloor$ times continuously differentiable in a neighborhood of v, then p_v is the Taylor expansion of f at v. Pointwise Lipschitz exponents may vary arbitrarily from abscissa to abscissa. One can construct multifractal functions with non-isolated singularities, where f has a different Lipschitz regularity at each point. In contrast, uniform Lipschitz exponents provide a more global measurement of regularity, which applies to a whole interval. If f is uniformly Lipschitz $\alpha > m$ in the neighborhood

of v then one can verify that f is necessarily m times continuously differentiable in this neighborhood.

If $0 \leq \alpha < 1$ then $p_v(t) = f(v)$ and the Lipschitz condition (6.3) becomes

$$\forall t \in \mathbb{R} \ , \ |f(t) - f(v)| \leq K |t-v|^{\alpha}.$$

A function that is bounded but discontinuous at v is Lipschitz 0 at v. If the Lipschitz regularity is $\alpha < 1$ at v, then f is not differentiable at v and α characterizes the singularity type.

Fourier Condition The uniform Lipschitz regularity of f over $\mathbb{R}$ is related to the asymptotic decay of its Fourier transform. The following theorem can be interpreted as a generalization of Proposition 2.1.

Theorem 6.1 *A function f is bounded and uniformly Lipschitz α over $\mathbb{R}$ if*

$$\int_{-\infty}^{+\infty} |\hat{f}(\omega)| \, (1 + |\omega|^{\alpha}) \, d\omega < +\infty. \tag{6.4}$$

Proof[1]. To prove that f is bounded, we use the inverse Fourier integral (2.8) and (6.4) which shows that

$$|f(t)| \leq \int_{-\infty}^{+\infty} |\hat{f}(\omega)| \, d\omega < +\infty.$$

Let us now verify the Lipschitz condition (6.3) when $0 \leq \alpha \leq 1$. In this case $p_v(t) = f(v)$ and the uniform Lipschitz regularity means that there exists $K > 0$ such that for all $(t, v) \in \mathbb{R}^2$

$$\frac{|f(t) - f(v)|}{|t-v|^{\alpha}} \leq K.$$

Since

$$\begin{aligned} f(t) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega) \exp(i\omega t) \, d\omega, \\ \frac{|f(t) - f(v)|}{|t-v|^{\alpha}} &\leq \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\hat{f}(\omega)| \, \frac{|\exp(i\omega t) - \exp(i\omega v)|}{|t-v|^{\alpha}} \, d\omega. \end{aligned} \tag{6.5}$$

For $|t-v|^{-1} \leq |\omega|$,

$$\frac{|\exp(i\omega t) - \exp(i\omega v)|}{|t-v|^{\alpha}} \leq \frac{2}{|t-v|^{\alpha}} \leq 2 |\omega|^{\alpha}.$$

For $|t-v|^{-1} \geq |\omega|$,

$$\frac{|\exp(i\omega t) - \exp(i\omega v)|}{|t-v|^{\alpha}} \leq \frac{|\omega| \, |t-v|}{|t-v|^{\alpha}} \leq |\omega|^{\alpha}.$$

Cutting the integral (6.5) in two for $|\omega| < |t-v|^{-1}$ and $|\omega| \geq |t-v|^{-1}$ yields

$$\frac{|f(t) - f(v)|}{|t-v|^{\alpha}} \leq \frac{1}{2\pi} \int_{-\infty}^{+\infty} 2 \, |\hat{f}(\omega)| \, |\omega|^{\alpha} \, d\omega = K.$$

If (6.4) is satisfied, then $K < +\infty$ so f is uniformly Lipschitz α.

Let us extend this result for $m = \lfloor \alpha \rfloor > 0$. We proved in (2.42) that (6.4) implies that f is m times continuously differentiable. One can verify that f is uniformly Lipschitz α over $\mathbb{R}$ if and only if $f^{(m)}$ is uniformly Lipschitz $\alpha - m$ over $\mathbb{R}$. The Fourier transform of $f^{(m)}$ is $(i\omega)^m \hat{f}(\omega)$. Since $0 \leq \alpha - m < 1$, we can use our previous result which proves that $f^{(m)}$ is uniformly Lipschitz $\alpha - m$, and hence that f is uniformly Lipschitz α. ■

The Fourier transform is a powerful tool for measuring the minimum global regularity of functions. However, it is not possible to analyze the regularity of f at a particular point v from the decay of $|\hat{f}(\omega)|$ at high frequencies ω. In contrast, since wavelets are well localized in time, the wavelet transform gives Lipschitz regularity over intervals *and* at points.

6.1.2 Wavelet Vanishing Moments

To measure the local regularity of a signal, it is not so important to use a wavelet with a narrow frequency support, but vanishing moments are crucial. If the wavelet has n vanishing moments then we show that the wavelet transform can be interpreted as a multiscale differential operator of order n. This yields a first relation between the differentiability of f and its wavelet transform decay at fine scales.

Polynomial Suppression The Lipschitz property (6.3) approximates f with a polynomial p_v in the neighborhood of v:

$$f(t) = p_v(t) + \epsilon_v(t) \quad \text{with} \quad |\epsilon_v(t)| \leq K\,|t-v|^{\alpha}\,. \tag{6.6}$$

A wavelet transform estimates the exponent α by ignoring the polynomial p_v. For this purpose, we use a wavelet that has $n > \alpha$ *vanishing moments*:

$$\int_{-\infty}^{+\infty} t^k \psi(t)\,dt = 0 \quad \text{for} \quad 0 \leq k < n\,.$$

A wavelet with n vanishing moments is orthogonal to polynomials of degree $n-1$. Since $\alpha < n$, the polynomial p_v has degree at most $n-1$. With the change of variable $t' = (t-u)/s$ we verify that

$$W p_v(u,s) = \int_{-\infty}^{+\infty} p_v(t)\,\frac{1}{\sqrt{s}}\,\psi\left(\frac{t-u}{s}\right)dt = 0. \tag{6.7}$$

Since $f = p_v + \epsilon_v$,

$$Wf(u,s) = W\epsilon_v(u,s). \tag{6.8}$$

Section 6.1.3 explains how to measure α from $|Wf(u,s)|$ when u is in the neighborhood of v.

Multiscale Differential Operator The following proposition proves that a wavelet with n vanishing moments can be written as the n^{th} order derivative of a function θ; the resulting wavelet transform is a multiscale differential operator. We suppose that ψ has a fast decay which means that for any decay exponent $m \in \mathbb{N}$ there exists C_m such that

$$\forall t \in \mathbb{R}\ ,\ |\psi(t)| \leq \frac{C_m}{1+|t|^m}\ . \tag{6.9}$$

Theorem 6.2 *A wavelet ψ with a fast decay has n vanishing moments if and only if there exists θ with a fast decay such that*

$$\psi(t) = (-1)^n \frac{d^n \theta(t)}{dt^n}. \tag{6.10}$$

As a consequence

$$Wf(u,s) = s^n \frac{d^n}{du^n}(f \star \bar{\theta}_s)(u)\ , \tag{6.11}$$

with $\bar{\theta}_s(t) = s^{-1/2}\theta(-t/s)$. Moreover, ψ has no more than n vanishing moments if and only if $\int_{-\infty}^{+\infty} \theta(t)\,dt \neq 0$.

Proof[1]. The fast decay of ψ implies that $\hat{\psi}$ is $\mathbf{C}^\infty$. This is proved by setting $f = \hat{\psi}$ in Proposition 2.1. The integral of a function is equal to its Fourier transform evaluated at $\omega = 0$. The derivative property (2.22) implies that for any $k < n$

$$\int_{-\infty}^{+\infty} t^k \psi(t)\,dt = (i)^k \hat{\psi}^{(k)}(0) = 0. \tag{6.12}$$

We can therefore make the factorization

$$\hat{\psi}(\omega) = (-i\omega)^n \hat{\theta}(\omega), \tag{6.13}$$

and $\hat{\theta}(\omega)$ is bounded. The fast decay of θ is proved with an induction on n. For $n = 1$,

$$\theta(t) = \int_{-\infty}^{t} \psi(u)\,du = \int_{t}^{+\infty} \psi(u)\,du,$$

and the fast decay of θ is derived from (6.9). We then similarly verify that increasing by 1 the order of integration up to n maintains the fast decay of θ.

Conversely, $|\hat{\theta}(\omega)| \leq \int_{-\infty}^{+\infty} |\theta(t)|\,dt < +\infty$, because θ has a fast decay. The Fourier transform of (6.10) yields (6.13) which implies that $\hat{\psi}^{(k)}(0) = 0$ for $k < n$. It follows from (6.12) that ψ has n vanishing moments.

To test whether ψ has more than n vanishing moments, we compute with (6.13)

$$\int_{-\infty}^{+\infty} t^n \psi(t)\,dt = (i)^n \hat{\psi}^{(n)}(0) = (-i)^n n!\,\hat{\theta}(0).$$

Clearly, ψ has no more than n vanishing moments if and only if $\hat{\theta}(0) = \int_{-\infty}^{+\infty} \theta(t)\,dt \neq 0$.

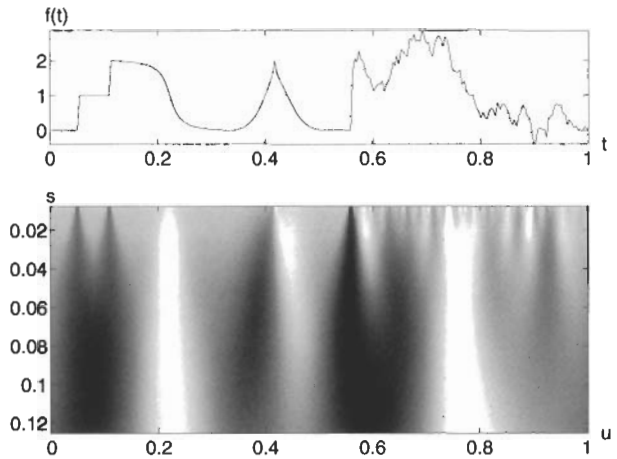


FIGURE 6.1 Wavelet transform $Wf(u,s)$ calculated with $\psi = -\theta'$ where θ is a Gaussian, for the signal f shown above. The position parameter u and the scale s vary respectively along the horizontal and vertical axes. Black, grey and white points correspond respectively to positive, zero and negative wavelet coefficients. Singularities create large amplitude coefficients in their cone of influence.

The wavelet transform (4.32) can be written

$$Wf(u,s) = f \star \bar{\psi}_s(u) \text{ with } \bar{\psi}_s(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{-t}{s}\right). \tag{6.14}$$

We derive from (6.10) that $\bar{\psi}_s(t) = s^n \frac{d^n \bar{\theta}_s(t)}{dt^n}$. Commuting the convolution and differentiation operators yields

$$Wf(u,s) = s^n f \star \frac{d^n \bar{\theta}_s}{dt^n}(u) = s^n \frac{d^n}{du^n}(f \star \bar{\theta}_s)(u).$$

■

If $K = \int_{-\infty}^{+\infty} \theta(t)\,dt \neq 0$ then the convolution $f \star \bar{\theta}_s(t)$ can be interpreted as a weighted average of f with a kernel dilated by s. So (6.11) proves that $Wf(u,s)$ is an n^{th} order derivative of an averaging of f over a domain proportional to s. Figure 6.1 shows a wavelet transform calculated with $\psi = -\theta'$, where θ is a Gaussian. The resulting $Wf(u,s)$ is the derivative of f averaged in the neighborhood of u with a Gaussian kernel dilated by s.

Since θ has a fast decay, one can verify that

$$\lim_{s \to 0} \frac{1}{\sqrt{s}} \bar{\theta}_s = K\,\delta\,,$$

in the sense of the weak convergence (A.30). This means that for any ϕ that is continuous at u,

$$\lim_{s\to 0} \phi \star \frac{1}{\sqrt{s}} \bar{\theta}_s(u) = K\,\phi(u).$$

If f is n times continuously differentiable in the neighborhood of u then (6.11) implies that

$$\lim_{s\to 0} \frac{Wf(u,s)}{s^{n+1/2}} = \lim_{s\to 0} f^{(n)} \star \frac{1}{\sqrt{s}} \bar{\theta}_s(u) = K\,f^{(n)}(u)\ . \tag{6.15}$$

In particular, if f is $\mathbf{C}^n$ with a bounded n^{th} order derivative then $|Wf(u,s)| = O(s^{n+1/2})$. This is a first relation between the decay of $|Wf(u,s)|$ when s decreases and the uniform regularity of f. Finer relations are studied in the next section.

6.1.3 Regularity Measurements with Wavelets

The decay of the wavelet transform amplitude across scales is related to the uniform and pointwise Lipschitz regularity of the signal. Measuring this asymptotic decay is equivalent to zooming into signal structures with a scale that goes to zero. We suppose that the wavelet ψ has n vanishing moments and is $\mathbf{C}^n$ with derivatives that have a fast decay. This means that for any $0 \leq k \leq n$ and $m \in \mathbb{N}$ there exists C_m such that

$$\forall t \in \mathbb{R}\ ,\ |\psi^{(k)}(t)| \leq \frac{C_m}{1+|t|^m}\ . \tag{6.16}$$

The following theorem relates the uniform Lipschitz regularity of f on an interval to the amplitude of its wavelet transform at fine scales.

Theorem 6.3 *If $f \in \mathbf{L}^2(\mathbb{R})$ is uniformly Lipschitz $\alpha \leq n$ over $[a,b]$, then there exists $A > 0$ such that*

$$\forall (u,s) \in [a,b] \times \mathbb{R}^+\ ,\ |Wf(u,s)| \leq A s^{\alpha+1/2}. \tag{6.17}$$

Conversely, suppose that f is bounded and that $Wf(u,s)$ satisfies (6.17) for an $\alpha < n$ that is not an integer. Then f is uniformly Lipschitz α on $[a+\epsilon, b-\epsilon]$, for any $\epsilon > 0$.

Proof[3]. This theorem is proved with minor modifications in the proof of Theorem 6.4. Since f is Lipschitz α at any $v \in [a,b]$, Theorem 6.4 shows in (6.20) that

$$\forall (u,s) \in \mathbb{R} \times \mathbb{R}^+\ ,\ |Wf(u,s)| \leq A s^{\alpha+1/2} \left(1 + \left|\frac{u-v}{s}\right|^{\alpha}\right).$$

For $u \in [a,b]$, we can choose $v = u$, which implies that $|Wf(u,s)| \leq A s^{\alpha+1/2}$. We verify from the proof of (6.20) that the constant A does not depend on v because the Lipschitz regularity is uniform over $[a,b]$.

To prove that f is uniformly Lipschitz α over $[a+\epsilon, b-\epsilon]$ we must verify that there exists K such that for all $v \in [a+\epsilon, b-\epsilon]$ we can find a polynomial p_v of degree $\lfloor \alpha \rfloor$ such that

$$\forall t \in \mathbb{R} \ , \ |f(t) - p_v(t)| \leq K |t-v|^\alpha \ . \tag{6.18}$$

When $t \notin [a+\epsilon/2, b-\epsilon/2]$ then $|t-v| \geq \epsilon/2$ and since f is bounded, (6.18) is verified with a constant K that depends on ϵ. For $t \in [a+\epsilon/2, b-\epsilon/2]$, the proof follows the same derivations as the proof of pointwise Lipschitz regularity from (6.21) in Theorem 6.4. The upper bounds (6.26) and (6.27) are replaced by

$$\forall t \in [a+\epsilon/2, b-\epsilon/2] \ , \ |\Delta_j^{(k)}(t)| \leq K 2^{(\alpha-k)j} \ \text{ for } 0 \leq k \leq \lfloor \alpha \rfloor + 1 \ . \tag{6.19}$$

This inequality is verified by computing an upper bound integral similar to (6.25) but which is divided in two, for $u \in [a,b]$ and $u \notin [a,b]$. When $u \in [a,b]$, the condition (6.21) is replaced by $|Wf(u,s)| \leq A s^{\alpha+1/2}$ in (6.25). When $u \notin [a,b]$, we just use the fact that $|Wf(u,s)| \leq \|f\| \, \|\psi\|$ and derive (6.19) from the fast decay of $|\psi^{(k)}(t)|$, by observing that $|t-u| \geq \epsilon/2$ for $t \in [a+\epsilon/2, b-\epsilon/2]$. The constant K depends on A and ϵ but not on v. The proof then proceeds like the proof of Theorem 6.4, and since the resulting constant K in (6.29) does not depend on v, the Lipschitz regularity is uniform over $[a-\epsilon, b+\epsilon]$. ■

The inequality (6.17) is really a condition on the asymptotic decay of $|Wf(u,s)|$ when s goes to zero. At large scales it does not introduce any constraint since the Cauchy-Schwarz inequality guarantees that the wavelet transform is bounded:

$$|Wf(u,s)| = |\langle f, \psi_{u,s} \rangle| \leq \|f\| \, \|\psi\| .$$

When the scale s decreases, $Wf(u,s)$ measures fine scale variations in the neighborhood of u. Theorem 6.3 proves that $|Wf(u,s)|$ decays like $s^{\alpha+1/2}$ over intervals where f is uniformly Lipschitz α.

Observe that the upper bound (6.17) is similar to the sufficient Fourier condition of Theorem 6.1, which supposes that $|\hat{f}(\omega)|$ decays faster than $\omega^{-\alpha}$. The wavelet scale s plays the role of a "localized" inverse frequency ω^{-1}. As opposed to the Fourier transform Theorem 6.1, the wavelet transform gives a Lipschitz regularity condition that is localized over any finite interval and it provides a necessary condition which is nearly sufficient. When $[a,b] = \mathbb{R}$ then (6.17) is a necessary and sufficient condition for f to be uniformly Lipschitz α on $\mathbb{R}$.

If ψ has exactly n vanishing moments then the wavelet transform decay gives no information concerning the Lipschitz regularity of f for $\alpha > n$. If f is uniformly Lipschitz $\alpha > n$ then it is $\mathbf{C}^n$ and (6.15) proves that $\lim_{s \to 0} s^{-n-1/2} Wf(u,s) = K f^{(n)}(u)$ with $K \neq 0$. This proves that $|Wf(u,s)| \sim s^{n+1/2}$ at fine scales despite the higher regularity of f.

If the Lipschitz exponent α is an integer then (6.17) is not sufficient in order to prove that f is uniformly Lipschitz α. When $[a,b] = \mathbb{R}$, if $\alpha = 1$ and ψ has two vanishing moments, then the class of functions that satisfy (6.17) is called the *Zygmund class* [47]. It is slightly larger than the set of functions that are uniformly Lipschitz 1. For example, $f(t) = t \log_e t$ belongs to the Zygmund class although it is not Lipschitz 1 at $t = 0$.

Pointwise Lipschitz Regularity The study of pointwise Lipschitz exponents with the wavelet transform is a delicate and beautiful topic which finds its mathematical roots in the characterization of Sobolev spaces by Littlewood and Paley in the 1930's. Characterizing the regularity of f at a point v can be difficult because f may have very different types of singularities that are aggregated in the neighborhood of v. In 1984, Bony [99] introduced the "two-microlocalization" theory which refines the Littlewood-Paley approach to provide pointwise characterization of singularities, which he used to study the solution of hyperbolic partial differential equations. These technical results became much simpler through the work of Jaffard [220] who proved that the two-microlocalization properties are equivalent to specific decay conditions on the wavelet transform amplitude. The following theorem gives a necessary condition and a sufficient condition on the wavelet transform for estimating the Lipschitz regularity of f at a point v. Remember that the wavelet ψ has n vanishing moments and n derivatives having a fast decay.

Theorem 6.4 (JAFFARD) *If $f \in \mathbf{L}^2(\mathbb{R})$ is Lipschitz $\alpha \leq n$ at v, then there exists A such that*

$$\forall (u,s) \in \mathbb{R} \times \mathbb{R}^+ \ , \ |Wf(u,s)| \leq A s^{\alpha+1/2} \left(1 + \left|\frac{u-v}{s}\right|^{\alpha}\right) . \tag{6.20}$$

Conversely, if $\alpha < n$ is not an integer and there exist A and $\alpha' < \alpha$ such that

$$\forall (u,s) \in \mathbb{R} \times \mathbb{R}^+ \ , \ |Wf(u,s)| \leq A s^{\alpha+1/2} \left(1 + \left|\frac{u-v}{s}\right|^{\alpha'}\right) \tag{6.21}$$

then f is Lipschitz α at v.

Proof. The necessary condition is relatively simple to prove but the sufficient condition is much more difficult.

• *Proof*[1] *of (6.20)* Since f is Lipschitz α at v, there exists a polynomial p_v of degree $\lfloor \alpha \rfloor < n$ and K such that $|f(t) - p_v(t)| \leq K|t-v|^{\alpha}$. Since ψ has n vanishing moments, we saw in (6.7) that $Wp_v(u,s) = 0$ and hence

$$\begin{aligned} |Wf(u,s)| &= \left| \int_{-\infty}^{+\infty} \left(f(t) - p_v(t)\right) \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) dt \right| \\ &\leq \int_{-\infty}^{+\infty} K|t-v|^{\alpha} \frac{1}{\sqrt{s}} \left|\psi\left(\frac{t-u}{s}\right)\right| dt. \end{aligned}$$

The change of variable $x = (t-u)/s$ gives

$$|Wf(u,s)| \leq \sqrt{s} \int_{-\infty}^{+\infty} K|sx+u-v|^{\alpha} |\psi(x)| dx.$$

Since $|a+b|^{\alpha} \leq 2^{\alpha}(|a|^{\alpha} + |b|^{\alpha})$,

$$|Wf(u,s)| \leq K 2^{\alpha} \sqrt{s} \left(s^{\alpha} \int_{-\infty}^{+\infty} |x|^{\alpha} |\psi(x)| dx + |u-v|^{\alpha} \int_{-\infty}^{+\infty} |\psi(x)| dx \right)$$

which proves (6.20).

• *Proof*[2] *of (6.21)* The wavelet reconstruction formula (4.37) proves that f can be decomposed in a Littlewood-Paley type sum

$$f(t) = \sum_{j=-\infty}^{+\infty} \Delta_j(t) \tag{6.22}$$

with

$$\Delta_j(t) = \frac{1}{C_\psi} \int_{-\infty}^{+\infty} \int_{2^j}^{2^{j+1}} Wf(u,s) \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) \frac{ds}{s^2} du \,. \tag{6.23}$$

Let $\Delta_j^{(k)}$ be its k^{th} order derivative. To prove that f is Lipschitz α at v we shall approximate f with a polynomial that generalizes the Taylor polynomial

$$p_v(t) = \sum_{k=0}^{\lfloor \alpha \rfloor} \left(\sum_{j=-\infty}^{+\infty} \Delta_j^{(k)}(v) \right) \frac{(t-v)^k}{k!} \,. \tag{6.24}$$

If f is n times differentiable at v then p_v corresponds to the Taylor polynomial but this is not necessarily true. We shall first prove that $\sum_{j=-\infty}^{+\infty} \Delta_j^{(k)}(v)$ is finite by getting upper bounds on $|\Delta_j^{(k)}(t)|$. These sums may be thought of as a generalization of pointwise derivatives.

To simplify the notation, we denote by K a generic constant which may change value from one line to the next but that does not depend on j and t. The hypothesis (6.21) and the asymptotic decay condition (6.16) imply that

$$\begin{aligned} |\Delta_j(t)| &= \frac{1}{C_\psi} \int_{-\infty}^{+\infty} \int_{2^j}^{2^{j+1}} A s^\alpha \left(1 + \left|\frac{u-v}{s}\right|^{\alpha'}\right) \frac{C_m}{1+|(t-u)/s|^m} \frac{ds}{s^2} du \\ &\leq K \int_{-\infty}^{+\infty} 2^{\alpha j} \left(1 + \left|\frac{u-v}{2^j}\right|^{\alpha'}\right) \frac{1}{1+|(t-u)/2^j|^m} \frac{du}{2^j} \end{aligned} \tag{6.25}$$

Since $|u-v|^{\alpha'} \leq 2^{\alpha'}(|u-t|^{\alpha'} + |t-v|^{\alpha'})$, the change of variable $u' = 2^{-j}(u-t)$ yields

$$|\Delta_j(t)| \leq K 2^{\alpha j} \int_{-\infty}^{+\infty} \frac{1 + |u'|^{\alpha'} + |(v-t)/2^j|^{\alpha'}}{1+|u'|^m} du'.$$

Choosing $m = \alpha' + 2$ yields

$$|\Delta_j(t)| \leq K 2^{\alpha j} \left(1 + \left|\frac{v-t}{2^j}\right|^{\alpha'}\right). \tag{6.26}$$

The same derivations applied to the derivatives of $\Delta_j(t)$ yield

$$\forall k \leq \lfloor \alpha \rfloor + 1 \;,\; |\Delta_j^{(k)}(t)| \leq K 2^{(\alpha-k)j} \left(1 + \left|\frac{v-t}{2^j}\right|^{\alpha'}\right). \tag{6.27}$$

At $t = v$ it follows that

$$\forall k \leq \lfloor \alpha \rfloor \;,\; |\Delta_j^{(k)}(v)| \leq K 2^{(\alpha-k)j}. \tag{6.28}$$

This guarantees a fast decay of $|\Delta_j^{(k)}(v)|$ when 2^j goes to zero, because α is not an integer so $\alpha > \lfloor\alpha\rfloor$. At large scales 2^j, since $|Wf(u,s)| \leq \|f\|\,\|\psi\|$ with the change of variable $u' = (t-u)/s$ in (6.23) we have

$$|\Delta_j^{(k)}(v)| \leq \frac{\|f\|\,\|\psi\|}{C_\psi} \int_{-\infty}^{+\infty} |\psi^{(k)}(u')|\,du' \int_{2^j}^{2^{j+1}} \frac{ds}{s^{3/2+k}}$$

and hence $|\Delta_j^{(k)}(v)| \leq K 2^{-(k+1/2)j}$. Together with (6.28) this proves that the polynomial p_v defined in (6.24) has coefficients that are finite.

With the Littlewood-Paley decomposition (6.22) we compute

$$|f(t) - p_v(t)| = \left| \sum_{j=-\infty}^{+\infty} \left(\Delta_j(t) - \sum_{k=0}^{\lfloor\alpha\rfloor} \Delta_j^{(k)}(v) \frac{(t-v)^k}{k!} \right) \right| .$$

The sum over scales is divided in two at 2^J such that $2^J \geq |t-v| \geq 2^{J-1}$. For $j \geq J$, we can use the classical Taylor theorem to bound the Taylor expansion of Δ_j:

$$\begin{aligned} I &= \sum_{j=J}^{+\infty} \left| \Delta_j(t) - \sum_{k=0}^{\lfloor\alpha\rfloor} \Delta_j^{(k)}(v) \frac{(t-v)^k}{k!} \right| \\ &\leq \sum_{j=J}^{+\infty} \frac{(t-v)^{\lfloor\alpha\rfloor+1}}{(\lfloor\alpha\rfloor+1)!} \sup_{h\in[t,v]} |\Delta_j^{\lfloor\alpha\rfloor+1}(h)| . \end{aligned}$$

Inserting (6.27) yields

$$I \leq K\,|t-v|^{\lfloor\alpha\rfloor+1} \sum_{j=J}^{+\infty} 2^{-j(\lfloor\alpha\rfloor+1-\alpha)} \left| \frac{v-t}{2^j} \right|^{\alpha'}$$

and since $2^J \geq |t-v| \geq 2^{J-1}$ we get $I \leq K\,|v-t|^\alpha$.

Let us now consider the case $j < J$

$$\begin{aligned} II &= \sum_{j=-\infty}^{J-1} \left| \Delta_j(t) - \sum_{k=0}^{\lfloor\alpha\rfloor} \Delta_j^{(k)}(v) \frac{(t-v)^k}{k!} \right| \\ &\leq K \sum_{j=-\infty}^{J-1} \left(2^{\alpha j} \left(1 + \left| \frac{v-t}{2^j} \right|^{\alpha'} \right) + \sum_{k=0}^{\lfloor\alpha\rfloor} \frac{(t-v)^k}{k!} 2^{j(\alpha-k)} \right) \\ &\leq K \left(2^{\alpha J} + 2^{(\alpha-\alpha')J} |t-v|^{\alpha'} + \sum_{k=0}^{\lfloor\alpha\rfloor} \frac{(t-v)^k}{k!} 2^{J(\alpha-k)} \right) \end{aligned}$$

and since $2^J \geq |t-v| \geq 2^{J-1}$ we get $II \leq K\,|v-t|^\alpha$. As a result

$$|f(t) - p_v(t)| \leq I + II \leq K\,|v-t|^\alpha \tag{6.29}$$

which proves that f is Lipschitz α at v. ■

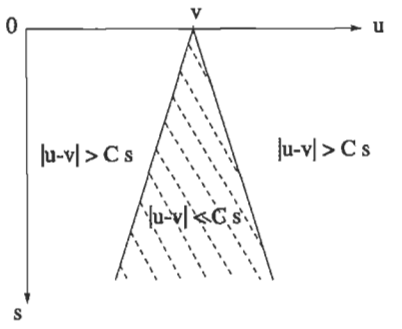


FIGURE 6.2 The cone of influence of an abscissa v consists of the scale-space points (u,s) for which the support of $\psi_{u,s}$ intersects $t = v$.

Cone of Influence To interpret more easily the necessary condition (6.20) and the sufficient condition (6.21), we shall suppose that ψ has a compact support equal to $[-C, C]$. The *cone of influence* of v in the scale-space plane is the set of points (u,s) such that v is included in the support of $\psi_{u,s}(t) = s^{-1/2}\psi((t-u)/s)$. Since the support of $\psi((t-u)/s)$ is equal to $[u - Cs, u + Cs]$, the cone of influence of v is defined by

$$|u - v| \leq Cs. \tag{6.30}$$

It is illustrated in Figure 6.2. If u is in the cone of influence of v then $Wf(u,s) = \langle f, \psi_{u,s} \rangle$ depends on the value of f in the neighborhood of v. Since $|u-v|/s \leq C$, the conditions (6.20,6.21) can be written

$$|Wf(u,s)| \leq A' s^{\alpha+1/2}$$

which is identical to the uniform Lipschitz condition (6.17) given by Theorem 6.3. In Figure 6.1, the high amplitude wavelet coefficients are in the cone of influence of each singularity.

Oscillating Singularities It may seem surprising that (6.20,6.21) also impose a condition on the wavelet transform outside the cone of influence of v. Indeed, this corresponds to wavelets whose support does not intersect v. For $|u - v| > Cs$ we get

$$|Wf(u,s)| \leq A' s^{\alpha-\alpha'+1/2} |u - v|^{\alpha}. \tag{6.31}$$

We shall see that it is indeed necessary to impose this decay when u tends to v in order to control the oscillations of f that might generate singularities.

Let us consider the generic example of a highly oscillatory function

$$f(t) = \sin \frac{1}{t},$$

which is discontinuous at $v = 0$ because of the acceleration of its oscillations. Since ψ is a smooth $\mathbf{C}^n$ function, if it is centered close to zero then the rapid oscillations

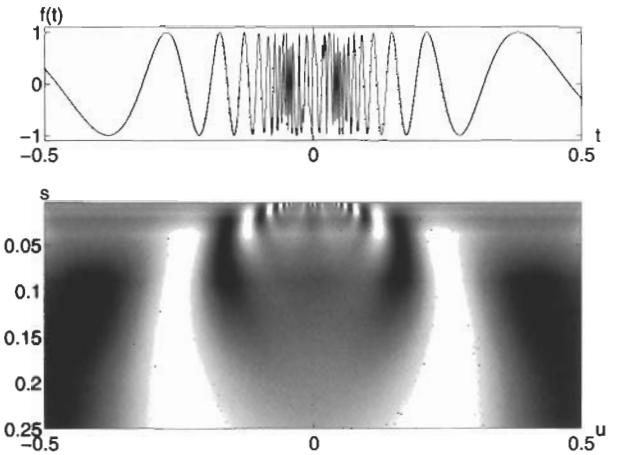


FIGURE 6.3 Wavelet transform of $f(t) = \sin(at^{-1})$ calculated with $\psi = -\theta'$ where θ is a Gaussian. High amplitude coefficients are along a parabola below the cone of influence of $t = 0$.

of $\sin t^{-1}$ produce a correlation integral $\langle \sin t^{-1}, \psi_{u,s} \rangle$ that is very small. With an integration by parts, one can verify that if (u,s) is in the cone of influence of $v = 0$, then $|Wf(u,s)| \leq A s^{2+1/2}$. This looks as if f is Lipschitz 2 at 0. However, Figure 6.3 shows high energy wavelet coefficients below the cone of influence of $v = 0$, which are responsible for the discontinuity. To guarantee that f is Lipschitz α, the amplitude of such coefficients is controlled by the upper bound (6.31).

To explain why the high frequency oscillations appear below the cone of influence of v, we use the results of Section 4.4.2 on the estimation of instantaneous frequencies with wavelet ridges. The instantaneous frequency of $\sin t^{-1} = \sin \phi(t)$ is $|\phi'(t)| = t^{-2}$. Let ψ^a be the analytic part of ψ, defined in (4.47). The corresponding complex analytic wavelet transform is $W^a f(u,s) = \langle f, \psi^a_{u,s} \rangle$. It was proved in (4.101) that for a fixed time u, the maximum of $s^{-1/2}|W^a f(u,s)|$ is located at the scale

$$s(u) = \frac{\eta}{\phi'(u)} = \eta u^2,$$

where η is the center frequency of $\hat{\psi}^a(\omega)$. When u varies, the set of points $(u, s(u))$ define a *ridge* that is a parabola located below the cone of influence of $v = 0$ in the plane (u,s). Since $\psi = \text{Real}[\psi^a]$, the real wavelet transform is

$$Wf(u,s) = \text{Real}[W^a f(u,s)].$$

The high amplitude values of $Wf(u,s)$ are thus located along the same parabola ridge curve in the scale-space plane, which clearly appears in Figure 6.3. Real wavelet coefficients $Wf(u,s)$ change sign along the ridge because of the variations of the complex phase of $W^a f(u,s)$.

The example of $f(t) = \sin t^{-1}$ can be extended to general oscillatory singularities [33]. A function f has an oscillatory singularity at v if there exist $\alpha \geq 0$ and $\beta > 0$ such that for t in a neighborhood of v

$$f(t) \sim |t-v|^{\alpha} g\left(\frac{1}{|t-v|^{\beta}}\right),$$

where $g(t)$ is a C^{∞} oscillating function whose primitives at any order are bounded. The function $g(t) = \sin t$ is a typical example. The oscillations have an instantaneous frequency $\phi'(t)$ that increases to infinity faster than $|t|^{-1}$ when t goes to v. High energy wavelet coefficients are located along the ridge $s(u) = \eta/\phi'(u)$, and this curve is necessarily below the cone of influence $|u-v| \leq Cs$.

6.2 WAVELET TRANSFORM MODULUS MAXIMA [2]

Theorems 6.3 and 6.4 prove that the local Lipschitz regularity of f at v depends on the decay at fine scales of $|Wf(u,s)|$ in the neighborhood of v. Measuring this decay directly in the time-scale plane (u,s) is not necessary. The decay of $|Wf(u,s)|$ can indeed be controlled from its local maxima values.

We use the term *modulus maximum* to describe any point (u_0, s_0) such that $|Wf(u,s_0)|$ is locally maximum at $u = u_0$. This implies that

$$\frac{\partial Wf(u_0,s_0)}{\partial u} = 0.$$

This local maximum should be a strict local maximum in either the right or the left neighborhood of u_0, to avoid having any local maxima when $|Wf(u,s_0)|$ is constant. We call *maxima line* any connected curve $s(u)$ in the scale-space plane (u,s) along which all points are modulus maxima. Figure 6.5(b) shows the wavelet modulus maxima of a signal.

6.2.1 Detection of Singularities

Singularities are detected by finding the abscissa where the wavelet modulus maxima converge at fine scales. To better understand the properties of these maxima, the wavelet transform is written as a multiscale differential operator. Theorem 6.2 proves that if ψ has exactly n vanishing moments and a compact support, then there exists θ of compact support such that $\psi = (-1)^n \theta^{(n)}$ with $\int_{-\infty}^{+\infty} \theta(t)\,dt \neq 0$. The wavelet transform is rewritten in (6.11) as a multiscale differential operator

$$Wf(u,s) = s^n \frac{d^n}{du^n}(f \star \bar{\theta}_s)(u). \tag{6.32}$$

If the wavelet has only one vanishing moment, wavelet modulus maxima are the maxima of the first order derivative of f smoothed by $\bar{\theta}_s$, as illustrated by Figure 6.4. These multiscale modulus maxima are used to locate discontinuities, and edges in images. If the wavelet has two vanishing moments, the modulus maxima correspond to high curvatures. The following theorem proves that if $Wf(u,s)$ has no modulus maxima at fine scales, then f is locally regular.

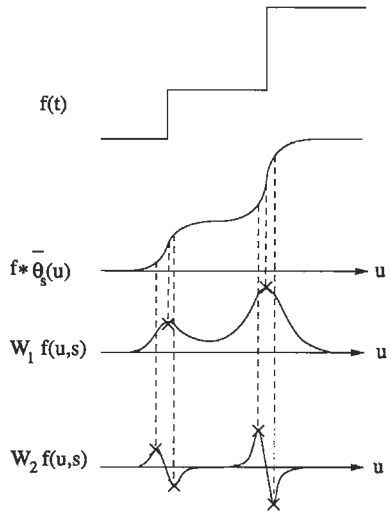


FIGURE 6.4 The convolution $f \star \bar{\theta}_s(u)$ averages f over a domain proportional to s. If $\psi = -\theta'$ then $W_1 f(u,s) = s\frac{d}{du}(f \star \bar{\theta}_s)(u)$ has modulus maxima at sharp variation points of $f \star \bar{\theta}_s(u)$. If $\psi = \theta''$ then the modulus maxima of $W_2 f(u,s) = s^2 \frac{d^2}{du^2}(f \star \bar{\theta}_s)(u)$ correspond to locally maximum curvatures.

Theorem 6.5 (HWANG, MALLAT) *Suppose that ψ is $\mathbf{C}^n$ with a compact support, and $\psi = (-1)^n \theta^{(n)}$ with $\int_{-\infty}^{+\infty} \theta(t)dt \neq 0$. Let $f \in \mathbf{L}^1[a,b]$. If there exists $s_0 > 0$ such that $|Wf(u,s)|$ has no local maximum for $u \in [a,b]$ and $s < s_0$, then f is uniformly Lipschitz n on $[a+\epsilon, b-\epsilon]$, for any $\epsilon > 0$.*

This theorem is proved in [258]. It implies that f can be singular (not Lipschitz 1) at a point v only if there is a sequence of wavelet maxima points $(u_p, s_p)_{p \in \mathbb{N}}$ that converges towards v at fine scales:

$$\lim_{p \to +\infty} u_p = v \quad \text{and} \quad \lim_{p \to +\infty} s_p = 0 .$$

These modulus maxima points may or may not be along the same maxima line. This result guarantees that all singularities are detected by following the wavelet transform modulus maxima at fine scales. Figure 6.5 gives an example where all singularities are located by following the maxima lines.

Maxima Propagation For all $\psi = (-1)^n \theta^{(n)}$, we are not guaranteed that a modulus maxima located at (u_0, s_0) belongs to a maxima line that propagates towards finer scales. When s decreases, $Wf(u,s)$ may have no more maxima in the neighborhood of $u = u_0$. The following proposition proves that this is never the case if θ is a Gaussian. The wavelet transform $Wf(u,s)$ can then be written as the solution of the heat diffusion equation, where s is proportional to the diffusion time. The

maximum principle applied to the heat diffusion equation proves that maxima may not disappear when s decreases. Applications of the heat diffusion equation to the analysis of multiscale averaging have been studied by several computer vision researchers [217, 236, 359].

Proposition 6.1 (HUMMEL, POGGIO, YUILLE) *Let $\psi = (-1)^n \theta^{(n)}$ where θ is a Gaussian. For any $f \in \mathbf{L}^2(\mathbb{R})$, the modulus maxima of $Wf(u,s)$ belong to connected curves that are never interrupted when the scale decreases.*

Proof[3]. To simplify the proof, we suppose that θ is a normalized Gaussian $\theta(t) = 2^{-1}\pi^{-1/2}\exp(-t^2/4)$ whose Fourier transform is $\hat{\theta}(\omega) = \exp(-\omega^2)$. Theorem 6.2 proves that

$$Wf(u,s) = s^n f^{(n)} \star \theta_s(u), \tag{6.33}$$

where the n^{th} derivative $f^{(n)}$ is defined in the sense of distributions. Let τ be the diffusion time. The solution of

$$\frac{\partial g(\tau,u)}{\partial \tau} = -\frac{\partial^2 g(\tau,u)}{\partial u^2} \tag{6.34}$$

with initial condition $g(0,u) = g_0(u)$ is obtained by computing the Fourier transform with respect to u of (6.34):

$$\frac{\partial g(\tau,u)}{\partial \tau} = -\omega^2 \hat{g}(\tau,\omega).$$

It follows that $\hat{g}(\tau,\omega) = \hat{g}_0(\omega)\exp(-\tau\omega^2)$ and hence

$$g(u,\tau) = \frac{1}{\sqrt{\tau}} g_0 \star \theta_\tau(u).$$

For $\tau = s$, setting $g_0 = f^{(n)}$ and inserting (6.33) yields $Wf(u,s) = s^{n+1/2} g(u,s)$. The wavelet transform is thus proportional to a heat diffusion with initial condition $f^{(n)}$.

The maximum principle for the parabolic heat equation [36] proves that a global maximum of $|g(u,s)|$ for $(u,s) \in [a,b] \times [s_0,s_1]$ is necessarily either on the boundary $u = a,b$ or at $s = s_0$. A modulus maxima of $Wf(u,s)$ at (u_1,s_1) is a local maxima of $|g(u,s)|$ for a fixed s and u varying. Suppose that a line of modulus maxima is interrupted at (u_1,s_1), with $s_1 > 0$. One can then verify that there exists $\epsilon > 0$ such that a global maximum of $|g(u,s)|$ over $[u_1-\epsilon, u_1+\epsilon] \times [s_1-\epsilon, s_1]$ is at (u_1,s_1). This contradicts the maximum principle, and thus proves that all modulus maxima propagate towards finer scales. ■

Derivatives of Gaussians are most often used to guarantee that all maxima lines propagate up to the finest scales. Chaining together maxima into maxima lines is also a procedure for removing spurious modulus maxima created by numerical errors in regions where the wavelet transform is close to zero.

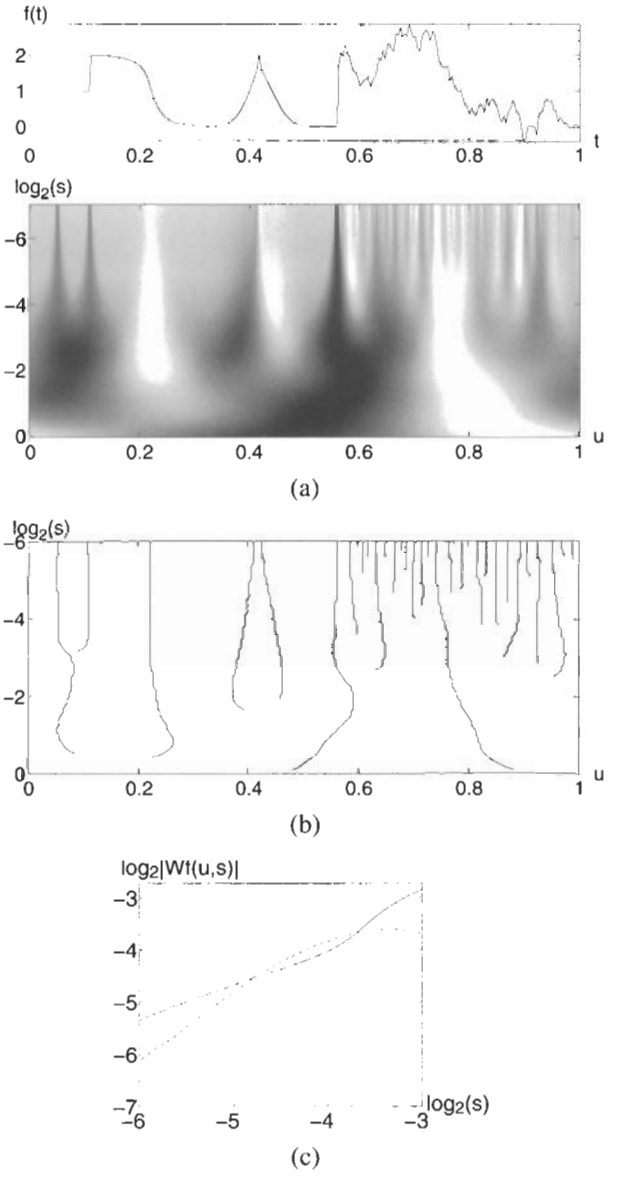


FIGURE 6.5 (a): Wavelet transform $Wf(u,s)$. The horizontal and vertical axes give respectively u and $\log_2 s$. (b): Modulus maxima of $Wf(u,s)$. (c): The full line gives the decay of $\log_2 |Wf(u,s)|$ as a function of $\log_2 s$ along the maxima line that converges to the abscissa $t = 0.05$. The dashed line gives $\log_2 |Wf(u,s)|$ along the left maxima line that converges to $t = 0.42$.

Isolated Singularities A wavelet transform may have a sequence of local maxima that converge to an abscissa v even though f is perfectly regular at v. This is the case of the maxima line of Figure 6.5 that converges to the abscissa $v = 0.23$. To detect singularities it is therefore not sufficient to follow the wavelet modulus maxima across scales. The Lipschitz regularity is calculated from the decay of the modulus maxima amplitude.

Let us suppose that for $s < s_0$ all modulus maxima that converge to v are included in a cone

$$|u - v| \leq Cs. \tag{6.35}$$

This means that f does not have oscillations that accelerate in the neighborhood of v. The potential singularity at v is necessarily isolated. Indeed, we can derive from Theorem 6.5 that the absence of maxima below the cone of influence implies that f is uniformly Lipschitz n in the neighborhood of any $t \neq v$ with $t \in (v - Cs_0, v + Cs_0)$. The decay of $|Wf(u,s)|$ in the neighborhood of v is controlled by the decay of the modulus maxima included in the cone $|u - v| \leq Cs$. Theorem 6.3 implies that f is uniformly Lipschitz α in the neighborhood of v if and only if there exists $A > 0$ such that each modulus maximum (u,s) in the cone (6.35) satisfies

$$|Wf(u,s)| \leq A s^{\alpha+1/2}, \tag{6.36}$$

which is equivalent to

$$\log_2 |Wf(u,s)| \leq \log_2 A + \left(\alpha + \frac{1}{2}\right) \log_2 s. \tag{6.37}$$

The Lipschitz regularity at v is thus the maximum slope of $\log_2 |Wf(u,s)|$ as a function of $\log_2 s$ along the maxima lines converging to v.

In numerical calculations, the finest scale of the wavelet transform is limited by the resolution of the discrete data. From a sampling at intervals N^{-1}, Section 4.3.3 computes the discrete wavelet transform at scales $s \geq \lambda N^{-1}$, where λ is large enough to avoid sampling coarsely the wavelets at the finest scale. The Lipschitz regularity α of a singularity is then estimated by measuring the decay slope of $\log_2 |Wf(u,s)|$ as a function of $\log_2 s$ for $2^J \geq s \geq \lambda N^{-1}$. The largest scale 2^J should be smaller than the distance between two consecutive singularities to avoid having other singularities influence the value of $Wf(u,s)$. The sampling interval N^{-1} must therefore be small enough to measure α accurately. The signal in Figure 6.5(a) is defined by $N = 256$ samples. Figure 6.5(c) shows the decay of $\log_2 |Wf(u,s)|$ along the maxima line converging to $t = 0.05$. It has slope $\alpha + 1/2 \approx 1/2$ for $2^{-4} \geq s \geq 2^{-6}$. As expected, $\alpha = 0$ because the signal is discontinuous at $t = 0.05$. Along the second maxima line converging to $t = 0.42$ the slope is $\alpha + 1/2 \approx 1$, which indicates that the singularity is Lipschitz $1/2$.

When f is a function whose singularities are not isolated, finite resolution measurements are not sufficient to distinguish individual singularities. Section 6.4 describes a global approach that computes the singularity spectrum of multifractals by taking advantage of their self-similarity.

Smoothed Singularities The signal may have important variations that are infinitely continuously differentiable. For example, at the border of a shadow the grey level of an image varies quickly but is not discontinuous because of the diffraction effect. The smoothness of these transitions is modeled as a diffusion with a Gaussian kernel whose variance is measured from the decay of wavelet modulus maxima.

In the neighborhood of a sharp transition at v, we suppose that

$$f(t) = f_0 \star g_\sigma(t), \tag{6.38}$$

where g_σ is a Gaussian of variance σ^2:

$$g_\sigma(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-t^2}{2\sigma^2}\right). \tag{6.39}$$

If f_0 has a Lipschitz α singularity at v that is isolated and non-oscillating, it is uniformly Lipschitz α in the neighborhood of v. For wavelets that are derivatives of Gaussians, the following theorem [261] relates the decay of the wavelet transform to σ and α.

Theorem 6.6 *Let $\psi = (-1)^n \theta^{(n)}$ with $\theta(t) = \lambda \exp(-t^2/(2\beta^2))$. If $f = f_0 \star g_\sigma$ and f_0 is uniformly Lipschitz α on $[v-h, v+h]$ then there exists A such that*

$$\forall (u,s) \in [v-h, v+h] \times \mathbb{R}^+ \ , \ |Wf(u,s)| \leq A s^{\alpha+1/2} \left(1 + \frac{\sigma^2}{\beta^2 s^2}\right)^{-(n-\alpha)/2}. \tag{6.40}$$

Proof[2]. The wavelet transform can be written

$$Wf(u,s) = s^n \frac{d^n}{du^n}(f \star \bar{\theta}_s)(u) = s^n \frac{d^n}{du^n}(f_0 \star g_\sigma \star \bar{\theta}_s)(u). \tag{6.41}$$

Since θ is a Gaussian, one can verify with a Fourier transform calculation that

$$\bar{\theta}_s \star g_\sigma(t) = \sqrt{\frac{s}{s_0}}\, \bar{\theta}_{s_0}(t) \ \text{ with } \ s_0 = \sqrt{s^2 + \frac{\sigma^2}{\beta^2}}. \tag{6.42}$$

Inserting this result in (6.41) yields

$$Wf(u,s) = s^n \sqrt{\frac{s}{s_0}} \frac{d^n}{du^n}(f_0 \star \bar{\theta}_{s_0})(u) = \left(\frac{s}{s_0}\right)^{n+1/2} Wf_0(u, s_0). \tag{6.43}$$

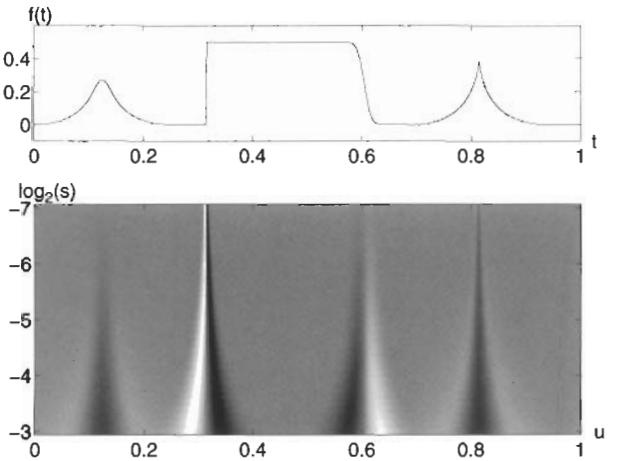
Since f_0 is uniformly Lipschitz α on $[v-h, v+h]$, Theorem 6.3 proves that there exists $A > 0$ such that

$$\forall (u,s) \in [v-h, v+h] \times \mathbb{R}^+ \ , \ |Wf_0(u,s)| \leq A s^{\alpha+1/2}. \tag{6.44}$$

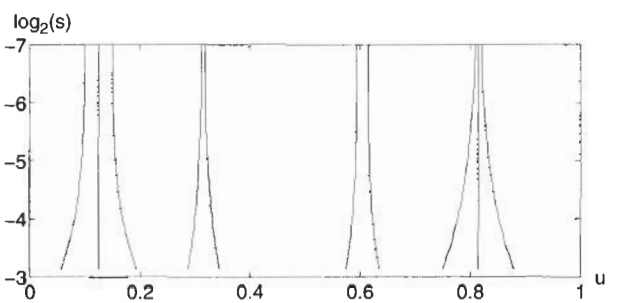
Inserting this in (6.43) gives

$$|Wf(u,s)| \leq A \left(\frac{s}{s_0}\right)^{n+1/2} s_0^{\alpha+1/2}, \tag{6.45}$$

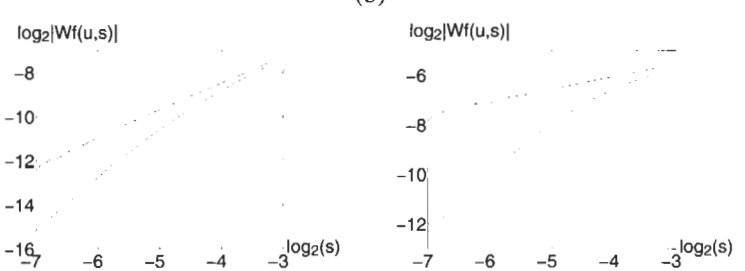
from which we derive (6.40) by inserting the expression (6.42) of s_0. ■



(a)



(b)



(c)

FIGURE 6.6 (a): Wavelet transform $Wf(u,s)$. (b): Modulus maxima of a wavelet transform computed $\psi = \theta''$, where θ is a Gaussian with variance $\beta = 1$. (c): Decay of $\log_2 |Wf(u,s)|$ along maxima curves. In the left figure, the solid and dotted lines correspond respectively to the maxima curves converging to $t = 0.81$ and $t = 0.12$. In the right figure, they correspond respectively to the curves converging to $t = 0.38$ and $t = 0.55$. The diffusion at $t = 0.12$ and $t = 0.55$ modifies the decay for $s \leq \sigma = 2^{-5}$.

Theorem 6.6 explains how the wavelet transform decay relates to the amount of diffusion of a singularity. At large scales $s \gg \sigma/\beta$, the Gaussian averaging is not "felt" by the wavelet transform which decays like $s^{\alpha+1/2}$. For $s \leq \sigma/\beta$, the variation of f at v is not sharp relative to s because of the Gaussian averaging. At these fine scales, the wavelet transform decays like $s^{n+1/2}$ because f is $\mathbf{C}^\infty$.

The parameters K, α, and σ are numerically estimated from the decay of the modulus maxima along the maxima curves that converge towards v. The variance β^2 depends on the choice of wavelet and is known in advance. A regression is performed to approximate

$$\log_2 |Wf(u,s)| \approx \log_2(K) + \left(\alpha + \frac{1}{2}\right) \log_2 s - \frac{n-\alpha}{2} \log_2\left(1 + \frac{\sigma^2}{\beta^2 s^2}\right).$$

Figure 6.6 gives the wavelet modulus maxima computed with a wavelet that is a second derivative of a Gaussian. The decay of $\log_2 |Wf(u,s)|$ as a function of $\log_2 s$ is given along several maxima lines corresponding to smoothed and non-smoothed singularities. The wavelet is normalized so that $\beta = 1$ and the diffusion scale is $\sigma = 2^{-5}$.

6.2.2 Reconstruction From Dyadic Maxima [3]

Wavelet transform maxima carry the properties of sharp signal transitions and singularities. If one can reconstruct a signal from these maxima, it is then possible to modify the singularities of a signal by processing the wavelet transform modulus maxima. The strength of singularities can be modified by changing the amplitude of the maxima and we can remove some singularities by suppressing the corresponding maxima.

For fast numerical computations, the detection of wavelet transform maxima is limited to dyadic scales $\{2^j\}_{j\in\mathbb{Z}}$. Suppose that ψ is a dyadic wavelet, which means that there exist $A > 0$ and B such that

$$\forall \omega \in \mathbb{R} - \{0\} \ , \ A \leq \sum_{j=-\infty}^{+\infty} |\hat{\psi}(2^j\omega)|^2 \leq B. \tag{6.46}$$

Theorem 5.11 proves that the dyadic wavelet transform $\{Wf(u,2^j)\}_{j\in\mathbb{Z}}$ is a complete and stable representation. This means that it admits a bounded left inverse. This dyadic wavelet transform has the same properties as a continuous wavelet transform $Wf(u,s)$. All theorems of Sections 6.1.3 and 6.2 remain valid if we restrict s to the dyadic scales $\{2^j\}_{j\in\mathbb{Z}}$. Singularities create sequences of maxima that converge towards the corresponding location at fine scales, and the Lipschitz regularity is calculated from the decay of the maxima amplitude.

Translation-Invariant Representation At each scale 2^j, the maxima representation provides the values of $Wf(u,2^j)$ where $|Wf(u,2^j)|$ is locally maximum. Figure 6.7(c) gives an example. This adaptive sampling of u produces a translation-invariant representation. When f is translated by τ each $Wf(2^j,u)$ is translated by

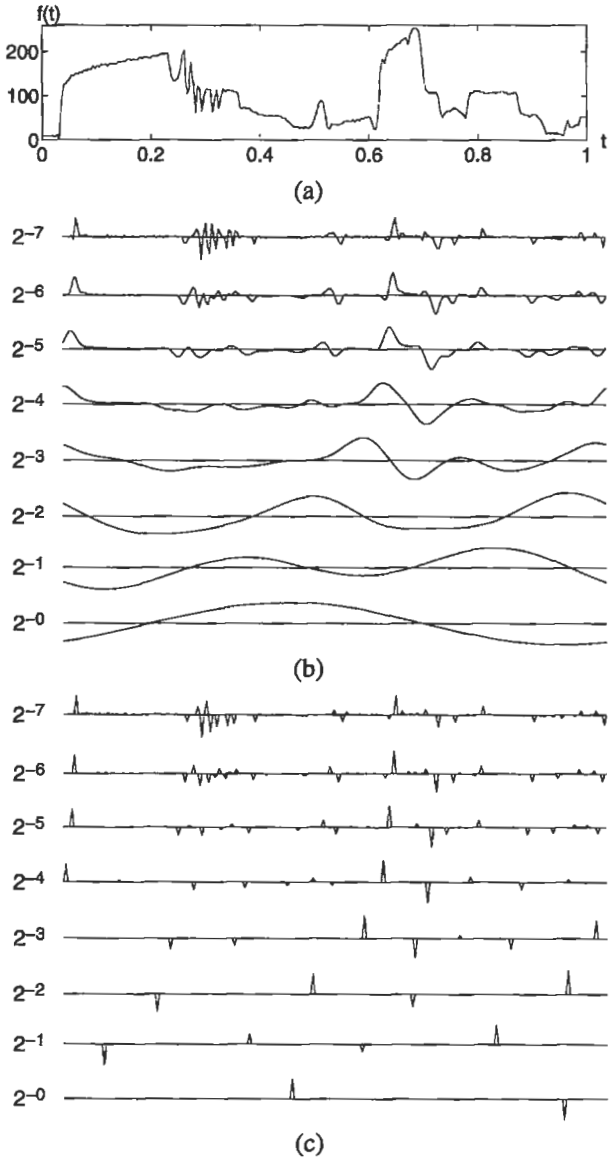


FIGURE 6.7 (a): Intensity variation along one row of the Lena image. (b): Dyadic wavelet transform computed at all scales $2N^{-1} \leq 2^j \leq 1$, with the quadratic spline wavelet $\psi = -\theta'$ shown in Figure 5.6. (c): Modulus maxima of the dyadic wavelet transform.

τ and their maxima are translated as well. This is not the case when u is uniformly sampled as in the wavelet frames of Section 5.3. Section 5.4 explains that this translation invariance is of prime importance for pattern recognition applications.

Reconstruction To study the completeness and stability of wavelet maxima representations, Mallat and Zhong introduced an alternate projection algorithm [261] that recovers signal approximations from their wavelet maxima; several other algorithms have been proposed more recently [116, 142, 199]. Numerical experiments show that one can only recover signal approximations with a relative mean-square error of the order of 10^{-2}. For general dyadic wavelets, Meyer [48] and Berman [94] proved that exact reconstruction is not possible. They found families of continuous or discrete signals whose dyadic wavelet transforms have the same modulus maxima. However, signals with the same wavelet maxima differ from each other only slightly, which explains the success of numerical reconstructions [261]. If the signal has a band-limited Fourier transform and if $\hat{\psi}$ has a compact support, then Kicey and Lennard [235] proved that wavelet modulus maxima define a complete and stable signal representation.

A simple and fast reconstruction algorithm is presented from a frame perspective. Section 5.1 is thus a prerequisite. At each scale 2^j, we know the positions $\{u_{j,p}\}_p$ of the local maxima of $|Wf(u,2^j)|$ and the values

$$Wf(u_{j,p},2^j) = \langle f, \psi_{j,p} \rangle$$

with

$$\psi_{j,p}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t-u_{j,p}}{2^j}\right).$$

Let ψ' be the derivative of ψ and $\psi'_{j,p}(t) = 2^{-j/2}\psi'(2^{-j}(t-u_{j,p}))$. Since $Wf(u,2^j)$ has a local extremum at $u = u_{j,p}$

$$\frac{\partial Wf(u_{j,p},2^j)}{\partial u} = -2^{-j} \langle f, \psi'_{j,p} \rangle = 0 .$$

The reconstruction algorithm should thus recover a function $\tilde{f}$ such that

$$W\tilde{f}(u_{j,p},2^j) = \langle \tilde{f}, \psi_{j,p} \rangle = \langle f, \psi_{j,p} \rangle , \tag{6.47}$$

and

$$\langle \tilde{f}, \psi'_{j,p} \rangle = \langle f, \psi'_{j,p} \rangle = 0 . \tag{6.48}$$

This last condition imposes that the derivative of $W\tilde{f}(u,2^j)$ vanishes at $u = u_{j,p}$, which does not necessarily mean that it has a modulus maxima at $u_{j,p}$. This is further enforced by minimizing $\|\tilde{f}\|$.

Frame Pseudo-Inverse The reconstruction algorithm recovers the function $\tilde{f}$ of minimum norm that satisfies (6.47) and (6.48). The minimization of $\|\tilde{f}\|$ has a tendency to decrease the wavelet transform energy at each scale 2^j:

$$\|W\tilde{f}(u,2^j)\|^2 = \int_{-\infty}^{+\infty} |W\tilde{f}(u,2^j)|^2\,du$$

because of the norm equivalence proved in Theorem 5.11:

$$A\|\tilde{f}\|^2 \leq \sum_{j=-\infty}^{+\infty} 2^{-j}\|W\tilde{f}(u,2^j)\|^2 \leq B\|\tilde{f}\|^2.$$

The norm $\|W\tilde{f}(u,2^j)\|$ is reduced by decreasing $|W\tilde{f}(u,2^j)|$. Since we also impose that $W\tilde{f}(u_{j,p},2^j) = \langle f, \psi_{j,p}\rangle$, minimizing $\|\tilde{f}\|$ generally creates local maxima at $u = u_{j,p}$.

The signal $\tilde{f}$ of minimum norm that satisfies (6.47) and (6.48) is the orthogonal projection $P_{\mathbf{V}}f$ of f on the space $\mathbf{V}$ generated by the wavelets $\{\psi_{j,p}, \psi'_{j,p}\}_{j,p}$. In discrete calculations, there is a finite number of maxima, so $\{\psi_{j,p}, \psi'_{j,p}\}_{j,p}$ is a finite family and hence a basis or a redundant frame of $\mathbf{V}$. Theorem 5.4 describes a conjugate gradient algorithm that recovers $\tilde{f}$ from the frame coefficients with a pseudo-inverse. It performs this calculation by inverting a frame symmetrical operator L introduced in (5.26). This operator is defined by

$$\forall r \in \mathbf{V}\ ,\ Lr = \sum_{j,p}\Big(\langle r, \psi_{j,p}\rangle \psi_{j,p} + \langle r, \psi'_{j,p}\rangle \psi'_{j,p}\Big)\ . \tag{6.49}$$

Clearly $\tilde{f} = L^{-1}Lf = L^{-1}g$ with

$$g = L\tilde{f} = \sum_{j,p}\Big(\langle \tilde{f}, \psi_{j,p}\rangle \psi_{j,p} + \langle \tilde{f}, \psi'_{j,p}\rangle \psi'_{j,p}\Big) = \sum_{j,p}\langle f, \psi_{j,p}\rangle \psi_{j,p}\ . \tag{6.50}$$

The conjugate gradient computes $L^{-1}g$ with an iterative procedure that has exponential convergence. The convergence rate depends on the frame bounds A and B of $\{\psi_{j,p}, \psi'_{j,p}\}_{j,p}$ in $\mathbf{V}$.

A faster reconstruction algorithm does not explicitly impose the fact $\langle \tilde{f}, \psi'_{j,p}\rangle = 0$, by considering a smaller space $\mathbf{V}$ generated by the restricted wavelet family $\{\psi_{j,p}\}_{j,p}$, and performing the calculation with the reduced frame operator:

$$Lr = \sum_{j,p}\langle r, \psi_{j,p}\rangle \psi_{j,p}\ . \tag{6.51}$$

The minimization of $\|\tilde{f}\|$ recovers a function such that $W\tilde{f}(u,2^j)$ is nearly locally maximum at $u = u_{j,p}$, and fewer operations are needed with this reduced frame operator. About 10 iterations are usually sufficient to recover an approximation of f with a relative mean-square error on the order of 10^{-2}. More iterations do not decrease the error much because $\tilde{f} \neq f$. Each iteration requires $O(N\log_2 N)$ calculations if implemented with a fast "à trous" algorithm.

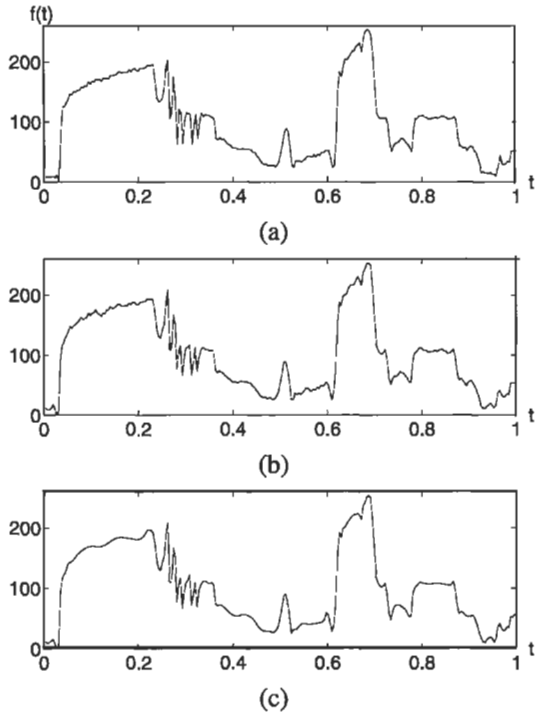


FIGURE 6.8 (a): Original signal. (b): Frame reconstruction from the dyadic wavelet maxima shown in Figure 6.7(c). (c): Frame reconstruction from the maxima whose amplitude is above the threshold $T = 10$.

Example 6.1 Figure 6.8(b) shows the signal $\tilde{f} = P_V f$ recovered with 10 iterations of the conjugate gradient algorithm, from the wavelet tranform maxima in Figure 6.7(c). This reconstruction is calculated with the simplified frame operator (6.51). After 20 iterations, the reconstruction error is $\|f - \tilde{f}\|/\|f\| = 2.5\,10^{-2}$. Figure 6.8(c) shows the signal reconstructed from the 50% of wavelet maxima that have the largest amplitude. Sharp signal transitions corresponding to large wavelet maxima have not been affected, but small texture variations disappear because the corresponding maxima are removed. The resulting signal is piecewise regular.

Fast Discrete Calculations To simplify notation, the sampling interval of the input signal is normalized to 1. The dyadic wavelet transform of this normalized discrete signal $a_0[n]$ of size N is calculated at scales $2 \leq 2^j \leq N$ with the "algorithme à trous" of Section 5.5.2. The cascade of convolutions with the two filters $h[n]$ and $g[n]$ is computed with $O(N \log_2 N)$ operations.

Each wavelet coefficient can be written as an inner product of a_0 with a discrete

wavelet translated by m:

$$d_j[m] = \langle a_0[n], \psi_j[n-m] \rangle = \sum_{n=0}^{N-1} a_0[n]\, \psi_j[n-m] \,.$$

The modulus maxima are located at abscissa $u_{j,p}$ where $|d_j[u_{j,p}]|$ is locally maximum, which means that

$$|d_j[u_{j,p}]| \geq |d_j[u_{j,p}-1]| \text{ and } |d_j[u_{j,p}]| \geq |d_j[u_{j,p}+1]| \,,$$

so long as one of these two inequalities is strict. We denote $\psi_{j,p}[n] = \psi_j[n-u_{j,p}]$.

To reconstruct a signal from its dyadic wavelet transform calculated up to the coarsest scale 2^J, it is necessary to provide the remaining coarse approximation $a_J[m]$, which is reduced to a constant when $2^J = N$:

$$a_J[m] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} a_0[n] = \sqrt{N}\, C \,.$$

Providing the average C is also necessary in order to reconstruct a signal from its wavelet maxima.

The simplified maxima reconstruction algorithm inverts the symmetrical operator L associated to the frame coefficients that are kept, without imposing explicitly the local extrema property:

$$Lr = \sum_{j=1}^{\log_2 N} \sum_{p} \langle r, \psi_{j,p} \rangle\, \psi_{j,p} + C \,. \tag{6.52}$$

The computational complexity of the conjugate gradient algorithm of Theorem 5.4 is driven by the calculation of Lp_n in (5.38). This is optimized with an efficient filter bank implementation of L.

To compute Lr we first calculate the dyadic wavelet transform of $r[n]$ with the "algorithme à trous". At each scale 2^j, all coefficients that are not located at an abscissa $u_{j,p}$ are set to zero:

$$\tilde{d}_j[m] = \begin{cases} \langle r[n], \psi_j[n-u_{j,p}] \rangle & \text{if } m = u_{j,p} \\ 0 & \text{otherwise} \end{cases} \,. \tag{6.53}$$

Then $Lr[n]$ is obtained by modifying the filter bank reconstruction given by Proposition 5.6. The decomposition and reconstruction wavelets are the same in (6.52) so we set $\tilde{h}[n] = h[n]$ and $\tilde{g}[n] = g[n]$. The factor $1/2$ in (5.87) is also removed because the reconstruction wavelets in (6.52) are not attenuated by 2^{-j} as are the wavelets in the non-sampled reconstruction formula (5.71). For $J = \log_2 N$, we initialize $\tilde{a}_J[n] = C/\sqrt{N}$ and for $\log_2 N > j \geq 0$ we compute

$$\tilde{a}_j[n] = \tilde{a}_{j+1} \star h_j[n] + \tilde{d}_{j+1} \star g_j[n]. \tag{6.54}$$

One can verify that $Lr[n] = \tilde{a}_0[n]$ with the same derivations as in the proof of Proposition 5.6. Let K_h and K_g be the number of non-zero coefficients of $h[n]$ and $g[n]$. The calculation of $Lr[n]$ from $r[n]$ requires a total of $2(K_h + K_g)N\log_2 N$ operations. The reconstructions shown in Figure 6.8 are computed with the filters of Table 5.3.

6.3 MULTISCALE EDGE DETECTION [2]

The edges of structures in images are often the most important features for pattern recognition. This is well illustrated by our visual ability to recognize an object from a drawing that gives a rough outline of contours. But, what is an edge? It could be defined as points where the image intensity has sharp transitions. A closer look shows that this definition is often not satisfactory. Image textures do have sharp intensity variations that are often not considered as edges. When looking at a brick wall, we may decide that the edges are the contours of the wall whereas the bricks define a texture. Alternatively, we may include the contours of each brick in the set of edges and consider the irregular surface of each brick as a texture. The discrimination of edges versus textures depends on the scale of analysis. This has motivated computer vision researchers to detect sharp image variations at different scales [44, 298]. The next section describes the multiscale Canny edge detector [113]. It is equivalent to detecting modulus maxima in a two-dimensional dyadic wavelet transform [261]. The Lipschitz regularity of edge points is derived from the decay of wavelet modulus maxima across scales. It is also shown that image approximations may be reconstructed from these wavelet modulus maxima, with no visual degradation. Image processing algorithms can thus be implemented on multiscale edges.

6.3.1 Wavelet Maxima for Images [2]

Canny Edge Detection The Canny algorithm detects points of sharp variation in an image $f(x_1, x_2)$ by calculating the modulus of its gradient vector

$$\vec{\nabla} f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \right). \tag{6.55}$$

The partial derivative of f in the direction of a unit vector $\vec{n} = (\cos\alpha, \sin\alpha)$ in the $x = (x_1, x_2)$ plane is calculated as an inner product with the gradient vector

$$\frac{\partial f}{\partial \vec{n}} = \vec{\nabla} f \,.\, \vec{n} = \frac{\partial f}{\partial x_1}\cos\alpha + \frac{\partial f}{\partial x_2}\sin\alpha \,.$$

The absolute value of this partial derivative is maximum if $\vec{n}$ is colinear to $\vec{\nabla} f$. This shows that $\vec{\nabla} f(x)$ is parallel to the direction of maximum change of the surface $f(x)$. A point $y \in \mathbb{R}^2$ is defined as an edge if $|\vec{\nabla} f(x)|$ is locally maximum at $x = y$ when $x = y + \lambda \vec{\nabla} f(y)$ for $|\lambda|$ small enough. This means that the partial derivatives of f reach a local maximum at $x = y$, when x varies in a one-dimensional

neighborhood of y along the direction of maximum change of f at y. These edge points are inflection points of f.

Multiscale Edge Detection A multiscale version of this edge detector is implemented by smoothing the surface with a convolution kernel $\theta(x)$ that is dilated. This is computed with two wavelets that are the partial derivatives of θ:

$$\psi^1 = -\frac{\partial \theta}{\partial x_1} \text{ and } \psi^2 = -\frac{\partial \theta}{\partial x_2}. \tag{6.56}$$

The scale varies along the dyadic sequence $\{2^j\}_{j\in\mathbb{Z}}$ to limit computations and storage. For $1 \leq k \leq 2$, we denote for $x = (x_1, x_2)$

$$\psi^k_{2^j}(x_1,x_2) = \frac{1}{2^j}\psi^k\left(\frac{x_1}{2^j}, \frac{x_2}{2^j}\right) \text{ and } \bar{\psi}^k_{2^j}(x) = \psi^k_{2^j}(-x).$$

In the two directions indexed by $1 \leq k \leq 2$, the dyadic wavelet transform of $f \in \mathbf{L}^2(\mathbb{R}^2)$ at $u = (u_1, u_2)$ is

$$W^k f(u, 2^j) = \langle f(x), \psi^k_{2^j}(x-u)\rangle = f \star \bar{\psi}^k_{2^j}(u)\,. \tag{6.57}$$

Section 5.5.3 gives necessary and sufficient conditions for obtaining a complete and stable representation.

Let us denote $\theta_{2^j}(x) = 2^{-j}\theta(2^{-j}x)$ and $\bar{\theta}_{2^j}(x) = \theta_{2^j}(-x)$. The two scaled wavelets can be rewritten

$$\bar{\psi}^1_{2^j} = 2^j \frac{\partial \bar{\theta}_{2^j}}{\partial x_1} \text{ and } \bar{\psi}^2_{2^j} = 2^j \frac{\partial \bar{\theta}_{2^j}}{\partial x_2}\,.$$

We thus derive from (6.57) that the wavelet transform components are proportional to the coordinates of the gradient vector of f smoothed by $\bar{\theta}_{2^j}$:

$$\begin{pmatrix} W^1 f(u,2^j) \\ W^2 f(u,2^j) \end{pmatrix} = 2^j \begin{pmatrix} \frac{\partial}{\partial u_1}(f \star \bar{\theta}_{2^j})(u) \\ \frac{\partial}{\partial u_2}(f \star \bar{\theta}_{2^j})(u) \end{pmatrix} = 2^j \vec{\nabla}(f \star \bar{\theta}_{2^j})(u)\,. \tag{6.58}$$

The modulus of this gradient vector is proportional to the wavelet transform modulus

$$Mf(u,2^j) = \sqrt{|W^1 f(u,2^j)|^2 + |W^2 f(u,2^j)|^2}. \tag{6.59}$$

Let $Af(u,2^j)$ be the angle of the wavelet transform vector (6.58) in the plane (x_1, x_2)

$$Af(u,2^j) = \begin{cases} \alpha(u) & \text{if } W^1 f(u,2^j) \geq 0 \\ \pi - \alpha(u) & \text{if } W^1 f(u,2^j) < 0 \end{cases} \tag{6.60}$$

with

$$\alpha(u) = \tan^{-1}\left(\frac{W^2 f(u,2^j)}{W^1 f(u,2^j)}\right)\,.$$

The unit vector $\vec{n}_j(u) = (\cos Af(u,2^j), \sin Af(u,2^j))$ is colinear to $\vec{\nabla}(f \star \bar{\theta}_{2^j})(u)$. An edge point at the scale 2^j is a point v such that $Mf(u,2^j)$ is locally maximum at $u = v$ when $u = v + \lambda \vec{n}_j(v)$ for $|\lambda|$ small enough. These points are also called wavelet transform *modulus maxima*. The smoothed image $f \star \bar{\theta}_{2^j}$ has an inflection point at a modulus maximum location. Figure 6.9 gives an example where the wavelet modulus maxima are located along the contour of a circle.

Maxima curves Edge points are distributed along curves that often correspond to the boundary of important structures. Individual wavelet modulus maxima are chained together to form a maxima curve that follows an edge. At any location, the tangent of the edge curve is approximated by computing the tangent of a level set. This tangent direction is used to chain wavelet maxima that are along the same edge curve.

The level sets of $g(x)$ are the curves $x(s)$ in the (x_1, x_2) plane where $g(x(s))$ is constant. The parameter s is the arc-length of the level set. Let $\vec{\tau} = (\tau_1, \tau_2)$ be the direction of the tangent of $x(s)$. Since $g(x(s))$ is constant when s varies,

$$\frac{\partial g(x(s))}{\partial s} = \frac{\partial g}{\partial x_1}\tau_1 + \frac{\partial g}{\partial x_2}\tau_2 = \vec{\nabla} g \,.\, \vec{\tau} = 0 \,.$$

So $\vec{\nabla} g(x)$ is perpendicular to the direction $\vec{\tau}$ of the tangent of the level set that goes through x.

This level set property applied to $g = f \star \bar{\theta}_{2^j}$ proves that at a maximum point v the vector $\vec{n}_j(v)$ of angle $Af(v,2^j)$ is perpendicular to the level set of $f \star \bar{\theta}_{2^j}$ going through v. If the intensity profile remains constant along an edge, then the inflection points (maxima points) are along a level set. The tangent of the maxima curve is therefore perpendicular to $\vec{n}_j(v)$. The intensity profile of an edge may not be constant but its variations are often negligible over a neighborhood of size 2^j for a sufficiently small scale 2^j, unless we are near a corner. The tangent of the maxima curve is then nearly perpendicular to $\vec{n}_j(v)$. In discrete calculations, maxima curves are thus recovered by chaining together any two wavelet maxima at v and $v + \vec{n}$, which are neighbors over the image sampling grid and such that $\vec{n}$ is nearly perpendicular to $\vec{n}_j(v)$.

Example 6.2 The dyadic wavelet transform of the image in Figure 6.9 yields modulus images $Mf(2^j, v)$ whose maxima are along the boundary of a disk. This circular edge is also a level set of the image. The vector $\vec{n}_j(v)$ of angle $Af(2^j, v)$ is thus perpendicular to the edge at the maxima locations.

Example 6.3 In the Lena image shown in Figure 6.10, some edges disappear when the scale increases. These correspond to fine scale intensity variations that are removed by the averaging with $\bar{\theta}_{2^j}$ when 2^j is large. This averaging also modifies the position of the remaining edges. Figure 6.10(f) displays the wavelet

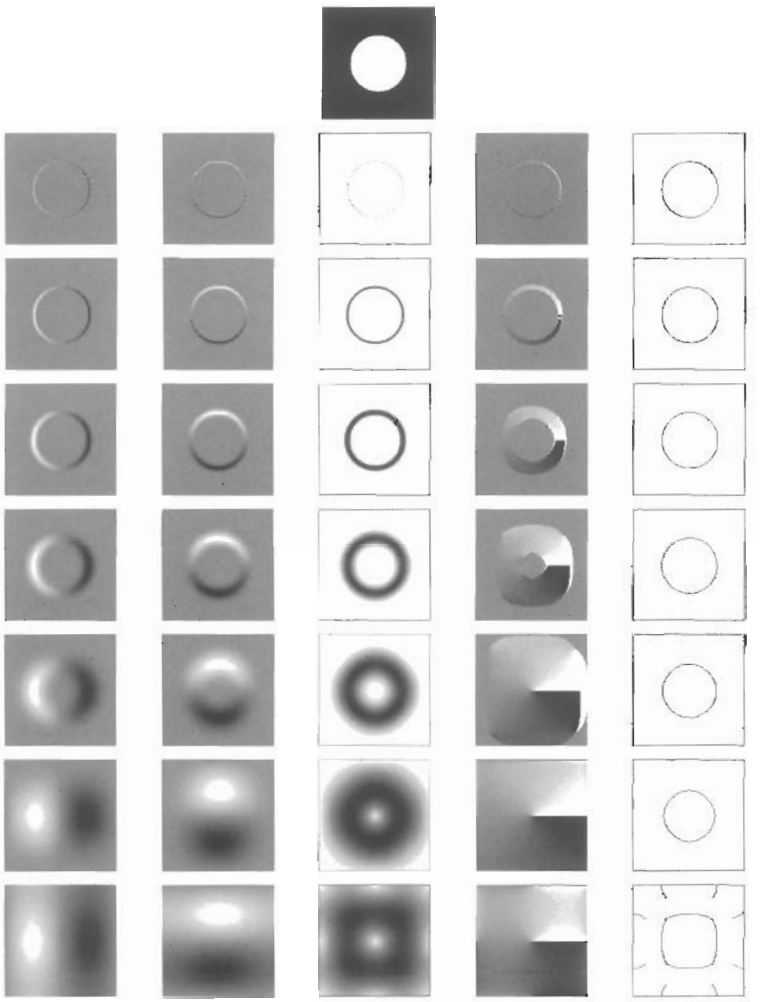


FIGURE 6.9 The top image has $N^2 = 128^2$ pixels. (a): Wavelet transform in the horizontal direction, with a scale 2^j that increases from top to bottom: $\{W^1 f(u,2^j)\}_{-6\leq j\leq 0}$. Black, grey and white pixels correspond respectively to negative, zero and positive values. (b): Vertical direction: $\{W^2 f(u,2^j)\}_{-6\leq j\leq 0}$. (c): Wavelet transform modulus $\{Mf(u,2^j)\}_{-6\leq j\leq 0}$. White and black pixels correspond respectively to zero and large amplitude coefficients. (d): Angles $\{Af(u,2^j)\}_{-6\leq j\leq 0}$ at points where the modulus is non-zero. (e): Wavelet modulus maxima are in black.

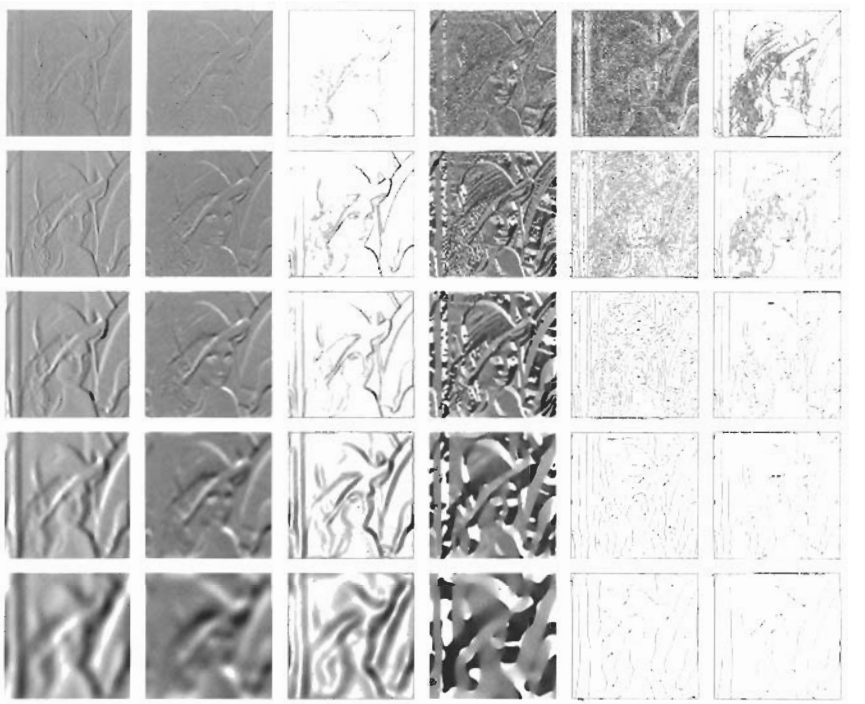


FIGURE 6.10 Multiscale edges of the Lena image shown in Figure 6.11. (a): $\{W^1 f(u,2^j)\}_{-7\le j\le -3}$. (b): $\{W^2 f(u,2^j)\}_{-7\le j\le -3}$. (c): $\{Mf(u,2^j)\}_{-7\le j\le -3}$. (d): $\{Af(u,2^j)\}_{-7\le j\le -3}$. (e): Modulus maxima. (f): Maxima whose modulus values are above a threshold.

maxima such that $Mf(v,2^j) \ge T$, for a given threshold T. They indicate the location of edges where the image has large amplitude variations.

Lipschitz Regularity The decay of the two-dimensional wavelet transform depends on the regularity of f. We restrict the analysis to Lipschitz exponents $0 \le \alpha \le 1$. A function f is said to be Lipschitz α at $v = (v_1, v_2)$ if there exists $K > 0$ such that for all $(x_1, x_2) \in \mathbb{R}^2$

$$|f(x_1,x_2) - f(v_1,v_2)| \le K\,(|x_1 - v_1|^2 + |x_2 - v_2|^2)^{\alpha/2}. \tag{6.61}$$

If there exists $K > 0$ such that (6.61) is satisfied for any $v \in \Omega$ then f is uniformly Lipschitz α over Ω. As in one dimension, the Lipschitz regularity of a function f is related to the asymptotic decay $|W^1 f(u,2^j)|$ and $|W^2 f(u,2^j)|$ in the corresponding neighborhood. This decay is controlled by $Mf(u,2^j)$. Like in Theorem 6.3, one can prove that f is uniformly Lipschitz α inside a bounded domain of $\mathbb{R}^2$ if and

only if there exists $A > 0$ such that for all u inside this domain and all scales 2^j

$$|Mf(u,2^j)| \leq A 2^{j(\alpha+1)}. \tag{6.62}$$

Suppose that the image has an isolated edge curve along which f has Lipschitz regularity α. The value of $|Mf(u,2^j)|$ in a two-dimensional neighborhood of the edge curve can be bounded by the wavelet modulus values along the edge curve. The Lipschitz regularity α of the edge is estimated with (6.62) by measuring the slope of $\log_2 |Mf(u,2^j)|$ as a function of j. If f is not singular but has a smooth transition along the edge, the smoothness can be quantified by the variance σ^2 of a two-dimensional Gaussian blur. The value of σ^2 is estimated by generalizing Theorem 6.6.

Reconstruction from Edges In his book about vision, Marr [44] conjectured that images can be reconstructed from multiscale edges. For a Canny edge detector, this is equivalent to recovering images from wavelet modulus maxima. In two dimensions, whether dyadic wavelet maxima define a complete and stable representation is still an open mathematical problem. However, the algorithm of Mallat and Zhong [261] recovers an image approximation that is visually identical to the original one.

As in Section 6.2.2, a simpler inverse frame reconstruction is described. At each scale 2^j, a multiscale edge representation provides the positions $u_{j,p}$ of the wavelet transform modulus maxima as well as the values of the modulus $Mf(u_{j,p},2^j)$ and the angle $Af(u_{j,p},2^j)$. The modulus and angle specify the two wavelet transform components

$$W^k f(u_{j,p},2^j) = \langle f, \psi^k_{j,p} \rangle \text{ for } 1 \leq k \leq 2, \tag{6.63}$$

with $\psi^k_{j,p}(x) = 2^{-j}\psi^k(2^{-j}(x-u_{j,p}))$. Let $\vec{n}_{j,p}$ be the unit vector in the direction of $Af(u_{j,p},2^j)$ and

$$\psi^3_{j,p}(x) = 2^{2j}\frac{\partial^2 \theta_{2^j}(x-u_{j,p})}{\partial \vec{n}^2_{j,p}}.$$

Since the gradient modulus $Mf(u_{j,p},2^j)$ has a local extremum at $u_{j,p}$ in the direction of $\vec{n}_{j,p}$, one can verify that

$$\langle f, \psi^3_{j,p} \rangle = 0. \tag{6.64}$$

As in one dimension, the reconstruction algorithm recovers a function of minimum norm $\tilde{f}$ such that

$$\langle \tilde{f}, \psi^k_{j,p} \rangle = \langle f, \psi^k_{j,p} \rangle \text{ for } 1 \leq k \leq 3. \tag{6.65}$$

It is the orthogonal projection of f in the closed space $\mathbf{V}$ generated by the family of wavelets

$$\left\{\psi^1_{j,p}, \psi^2_{j,p}, \psi^3_{j,p}\right\}_{j,p}.$$

(a)

(b)

(c)

FIGURE 6.11 (a): Original Lena. (b): Reconstructed from the wavelet maxima displayed in Figure 6.10(e) and larger scale maxima. (c): Reconstructed from the thresholded wavelet maxima displayed in Figure 6.10(f) and larger scale maxima.

If this family is a frame of $\mathbf{V}$, which is true in finite dimension, the associated frame operator is

$$\forall r \in \mathbf{V} \ , \ Lr = \sum_{k=1}^{3} \sum_{j,p} \langle r, \psi_{j,p}^k \rangle \ \psi_{j,p}^k \ . \tag{6.66}$$

We compute $\tilde{f} = L^{-1} g$ from $g = L\tilde{f} = \sum_{k=1}^{2} \sum_{j,p} \langle f, \psi_{j,p}^k \rangle \ \psi_{j,p}^k$, with the conjugate gradient algorithm of Theorem 5.4.

To simplify the numerical implementation, one can restrict the inner product conditions (6.65) to the wavelets $\psi_{j,p}^k$ for $k = 1, 2$. The frame operator (6.66) is

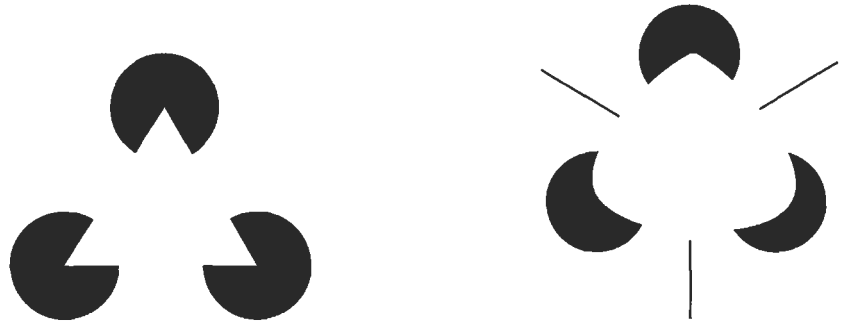


FIGURE 6.12 The illusory edges of a straight and a curved triangle are perceived in domains where the images are uniformly white.

then limited to these two types of wavelets:

$$Lr = \sum_{k=1}^{2} \sum_{j,p} \langle r, \psi_{j,p}^{k} \rangle \, \psi_{j,p}^{k} \, . \tag{6.67}$$

The resulting reconstructed image $\tilde{f}$ is not equal to the original image f but their relative mean-square difference is less than 10^{-2}. Singularities and edges are nearly perfectly recovered and no spurious oscillations are introduced. The images differ slightly in smooth regions, but visually this is not noticeable.

Example 6.4 The image reconstructed in Figure 6.11(b) is visually identical to the original image. It is recovered with 10 conjugate gradient iterations. After 20 iterations, the relative mean-square reconstruction error is $\|\tilde{f} - f\|/\|f\| = 4\,10^{-3}$. The thresholding of edges accounts for the disappearance of image structures from the reconstruction shown in Figure 6.11(c). Sharp image variations are perfectly recovered.

Illusory Contours A multiscale wavelet edge detector defines edges as points where the image intensity varies sharply. This definition is however too restrictive when edges are used to find the contours of objects. For image segmentation, edges must define closed curves that outline the boundaries of each region. Because of noise or light variations, local edge detectors produce contours with holes. Filling these holes requires some prior knowledge about the behavior of edges in the image. The illusion of the Kanizsa triangle [39] shows that such an edge filling is performed by the human visual system. In Figure 6.12, one can "see" the edges of a straight and a curved triangle although the image grey level remains uniformly white between the black disks. Closing edge curves and understanding illusory contours requires computational models that are not as local as multiscale

differential operators. Such contours can be obtained as the solution of a global optimization that incorporates constraints on the regularity of contours and which takes into account the existence of occlusions [189].

6.3.2 Fast Multiscale Edge Computations [3]

The dyadic wavelet transform of an image of N^2 pixels is computed with a separable extension of the filter bank algorithm described in Section 5.5.2. A fast multiscale edge detection is derived [261].

Wavelet Design Edge detection wavelets (6.56) are designed as separable products of one-dimensional dyadic wavelets, constructed in Section 5.5.1. Their Fourier transform is

$$\hat{\psi}^1(\omega_1,\omega_2) = \hat{g}\left(\frac{\omega_1}{2}\right)\hat{\phi}\left(\frac{\omega_1}{2}\right)\hat{\phi}\left(\frac{\omega_2}{2}\right), \tag{6.68}$$

and

$$\hat{\psi}^2(\omega_1,\omega_2) = \hat{g}\left(\frac{\omega_2}{2}\right)\hat{\phi}\left(\frac{\omega_1}{2}\right)\hat{\phi}\left(\frac{\omega_2}{2}\right), \tag{6.69}$$

where $\hat{\phi}(\omega)$ is a scaling function whose energy is concentrated at low frequencies and

$$\hat{g}(\omega) = -i\sqrt{2}\sin\left(\frac{\omega}{2}\right)\exp\left(\frac{-i\omega}{2}\right). \tag{6.70}$$

This transfer function is the Fourier transform of a finite difference filter which is a discrete approximation of a derivative

$$\frac{g[p]}{\sqrt{2}} = \begin{cases} -0.5 & \text{if } p = 0 \\ 0.5 & \text{if } p = 1 \\ 0 & \text{otherwise} \end{cases}. \tag{6.71}$$

The resulting wavelets ψ^1 and ψ^2 are finite difference approximations of partial derivatives along x and y of $\theta(x_1,x_2) = 4\phi(2x)\phi(2y)$.

To implement the dyadic wavelet transform with a filter bank algorithm, the scaling function $\hat{\phi}$ is calculated, as in (5.76), with an infinite product:

$$\hat{\phi}(\omega) = \prod_{p=1}^{+\infty}\frac{\hat{h}(2^{-p}\omega)}{\sqrt{2}} = \frac{1}{\sqrt{2}}\hat{h}\left(\frac{\omega}{2}\right)\hat{\phi}\left(\frac{\omega}{2}\right). \tag{6.72}$$

The 2π periodic function $\hat{h}$ is the transfer function of a finite impulse response low-pass filter $h[p]$. We showed in (5.81) that the Fourier transform of a box spline of degree m

$$\hat{\phi}(\omega) = \left(\frac{\sin(\omega/2)}{\omega/2}\right)^{m+1}\exp\left(\frac{-i\epsilon\omega}{2}\right) \quad \text{with} \quad \epsilon = \begin{cases} 1 & \text{if } m \text{ is even} \\ 0 & \text{if } m \text{ is odd} \end{cases}$$

is obtained with

$$\hat{h}(\omega) = \sqrt{2}\,\frac{\hat{\phi}(2\omega)}{\hat{\phi}(\omega)} = \sqrt{2}\left(\cos\frac{\omega}{2}\right)^{m+1} \exp\left(\frac{-i\epsilon\omega}{2}\right) .$$

Table 5.3 gives $h[p]$ for $m = 2$.

"Algorithme à trous" The one-dimensional "algorithme à trous" of Section 5.5.2 is extended in two dimensions with convolutions along the rows and columns of the image. The support of an image f is normalized to $[0,1]^2$ and the N^2 pixels are obtained with a sampling on a uniform grid with intervals N^{-1}. To simplify the description of the algorithm, the sampling interval is normalized to 1 by considering the dilated image $f(x_1,x_2) = f(N^{-1}x_1, N^{-1}x_2)$. A change of variable shows that the wavelet transform of f is derived from the wavelet transform of f with a simple renormalization:

$$W^k f(u,2^j) = N^{-1}\, W^k f(Nu, N2^j) .$$

Each sample $a_0[n]$ of the normalized discrete image is considered to be an average of f calculated with the kernel $\phi(x_1)\,\phi(x_2)$ translated at $n = (n_1,n_2)$:

$$a_0[n_1,n_2] = \langle f(x_1,x_2)\,,\, \phi(x_1-n_1)\,\phi(x_2-n_2)\rangle .$$

This is further justified in Section 7.7.3. For any $j \geq 0$, we denote

$$a_j[n_1,n_2] = \langle f(x_1,x_2)\,,\, \phi_{2^j}(x_1-n_1)\,\phi_{2^j}(x_2-n_2)\rangle .$$

The discrete wavelet coefficients at $n = (n_1,n_2)$ are

$$d_j^1[n] = W^1 f(n,2^j) \;\text{ and }\; d_j^2[n] = W^2 f(n,2^j) .$$

They are calculated with separable convolutions.

For any $j \geq 0$, the filter $h[p]$ "dilated" by 2^j is defined by

$$\bar{h}_j[p] = \begin{cases} h[-p/2^j] & \text{if } p/2^j \in \mathbb{Z} \\ 0 & \text{otherwise} \end{cases} \tag{6.73}$$

and for $j > 0$, a centered finite difference filter is defined by

$$\frac{\bar{g}_j[p]}{\sqrt{2}} = \begin{cases} 0.5 & \text{if } p = -2^{j-1} \\ -0.5 & \text{if } p = 2^{j-1} \\ 0 & \text{otherwise} \end{cases} . \tag{6.74}$$

For $j = 0$, we define $\bar{g}_0[0]/\sqrt{2} = -0.5$, $\bar{g}_0[-1]/\sqrt{2} = -0.5$ and $\bar{g}_0[p] = 0$ for $p \neq 0, -1$. A separable two-dimensional filter is written

$$\alpha\beta[n_1,n_2] = \alpha[n_1]\,\beta[n_2] ,$$

and $\delta[n]$ is a discrete Dirac. Similarly to Proposition 5.6, one can prove that for any $j \geq 0$ and any $n = (n_1, n_2)$

$$a_{j+1}[n] = a_j \star \bar{h}_j \bar{h}_j[n], \tag{6.75}$$

$$d^1_{j+1}[n] = a_j \star \bar{g}_j \delta[n], \tag{6.76}$$

$$d^2_{j+1}[n] = a_j \star \delta \bar{g}_j[n]. \tag{6.77}$$

Dyadic wavelet coefficients up to the scale 2^J are therefore calculated by cascading the convolutions (6.75-6.77) for $0 < j \leq J$. To take into account border problems, all convolutions are replaced by circular convolutions, which means that the input image $a_0[n]$ is considered to be N periodic along its rows and columns. Since $J \leq \log_2 N$ and all filters have a finite impulse response, this algorithm requires $O(N^2 \log_2 N)$ operations. If $J = \log_2 N$ then one can verify that the larger scale approximation is a constant proportional to the grey level average C:

$$a_J[n_1, n_2] = \frac{1}{N} \sum_{n_1, n_2 = 0}^{N-1} a_0[n_1, n_2] = NC\,.$$

The wavelet transform modulus is $Mf(n, 2^j) = |d^1_j[n]|^2 + |d^2_j[n]|^2$ whereas $Af(n, 2^j)$ is the angle of the vector $(d^1_j[n], d^2_j[n])$. The wavelet modulus maxima are located at points $u_{j,p}$ where $Mf(u_{j,p}, 2^j)$ is larger than its two neighbors $Mf(u_{j,p} \pm \vec{\epsilon}, 2^j)$, where $\vec{\epsilon} = (\epsilon_1, \epsilon_2)$ is the vector whose coordinates ϵ_1 and ϵ_2 are either 0 or 1, and whose angle is the closest to $Af(u_{j,p}, 2^j)$. This verifies that $Mf(n, 2^j)$ is locally maximum at $n = u_{j,p}$ in a one-dimensional neighborhood whose direction is along the angle $Af(u_{j,p}, 2^j)$.

Reconstruction from Maxima The frame algorithm recovers an image approximation from multiscale edges by inverting the operator L defined in (6.67), with the conjugate gradient algorithm of Theorem 5.4. This requires computing Lr efficiently for any image $r[n]$. For this purpose, the wavelet coefficients of r are first calculated with the "algorithme à trous," and at each scale $2 \leq 2^j \leq N$ all wavelets coefficients not located at a maximum position $u_{j,p}$ are set to zero as in the one-dimensional implementation (6.53):

$$\tilde{d}^k_j[n] = \begin{cases} W^k r(n, 2^j) & \text{if } n = u_{j,p} \\ 0 & \text{otherwise} \end{cases}.$$

The signal $Lr[n]$ is recovered from these non-zero wavelet coefficients with a reconstruction formula similar to (6.54). Let $h_j[n] = \bar{h}_j[-n]$ and $g_j[n] = \bar{g}_j[-n]$ be the two filters defined with (6.73) and (6.74). The calculation is initialized for $J = \log_2 N$ by setting $\tilde{a}_J[n] = CN^{-1}$, where C is the average image intensity. For $\log_2 N > j \geq 0$ we compute

$$\tilde{a}_j[n] = \tilde{a}_{j+1} \star h_j h_j[n] + d^1_{j+1} \star g_j \delta[n] + d^2_{j+1}[n] \star \delta g_j[n]\,,$$

and one can verify that $Lr[n] = \tilde{a}_0[n]$. It is calculated from $r[n]$ with $O(N^2 \log_2 N)$ operations. The reconstructed images in Figure 6.11 are obtained with 10 conjugate gradient iterations implemented with this filter bank algorithm.

6.4 MULTIFRACTALS [2]

Signals that are singular at almost every point were originally studied as pathological objects of pure mathematical interest. Mandelbrot [43] was the first to recognize that such phenomena are encountered everywhere. Among the many examples [25] let us mention economic records like the Dow Jones industrial average, physiological data including heart records, electromagnetic fluctuations in galactic radiation noise, textures in images of natural terrain, variations of traffic flow...

The singularities of multifractals often vary from point to point, and knowing the distribution of these singularities is important in analyzing their properties. Pointwise measurements of Lipschitz exponents are not possible because of the finite numerical resolution. After discretization, each sample corresponds to a time interval where the signal has an infinite number of singularities that may all be different. The singularity distribution must therefore be estimated from global measurements, which take advantage of multifractal self-similarities. Section 6.4.2 computes the fractal dimension of sets of points having the same Lipschitz regularity, with a global partition function calculated from wavelet transform modulus maxima. Applications to fractal noises such as fractional Brownian motions and to hydrodynamic turbulence are studied in Section 6.4.3.

6.4.1 Fractal Sets and Self-Similar Functions

A set $\mathcal{S} \subset \mathbb{R}^n$ is said to be self-similar if it is the union of disjoint subsets $\mathcal{S}_1, \ldots, \mathcal{S}_k$ that can be obtained from $\mathcal{S}$ with a scaling, translation and rotation. This self-similarity often implies an infinite multiplication of details, which creates irregular structures. The triadic Cantor set and the Van Koch curve are simple examples.

Example 6.5 The Von Koch curve is a fractal set obtained by recursively dividing each segment of length l in four segments of length $l/3$, as illustrated in Figure 6.13. Each subdivision increases the length by $4/3$. The limit of these subdivisions is therefore a curve of infinite length.

Example 6.6 The triadic Cantor set is constructed by recursively dividing intervals of size l in two sub-intervals of size $l/3$ and a central hole, illustrated by Figure 6.14. The iteration begins from $[0, 1]$. The Cantor set obtained as a limit of these subdivisions is a dust of points in $[0, 1]$.

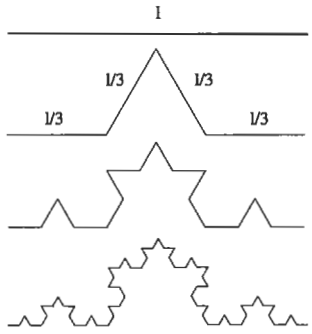


FIGURE 6.13 Three iterations of the Von Koch subdivision. The Von Koch curve is the fractal obtained as a limit of an infinite number of subdivisions.

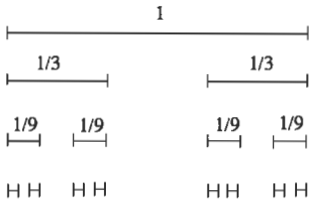


FIGURE 6.14 Three iterations of the Cantor subdivision of $[0,1]$. The limit of an infinite number of subdivisions is a closed set in $[0,1]$.

Fractal Dimension The Von Koch curve has infinite length in a finite square of $\mathbb{R}^2$. The usual length measurement is therefore not well adapted to characterize the topological properties of such fractal curves. This motivated Hausdorff in 1919 to introduce a new definition of dimension, based on the size variations of sets when measured at different scales.

The *capacity dimension* is a simplification of the Hausdorff dimension that is easier to compute numerically. Let $\mathcal{S}$ be a bounded set in $\mathbb{R}^n$. We count the minimum number $N(s)$ of balls of radius s needed to cover $\mathcal{S}$. If $\mathcal{S}$ is a set of dimension D with a finite length ($D=1$), surface ($D=2$) or volume ($D=3$) then

$$N(s) \sim s^{-D},$$

so

$$D = -\lim_{s \to 0} \frac{\log N(s)}{\log s}. \tag{6.78}$$

The capacity dimension D of $\mathcal{S}$ generalizes this result and is defined by

$$D = -\liminf_{s \to 0} \frac{\log N(s)}{\log s}. \tag{6.79}$$

The measure of $\mathcal{S}$ is then

$$M = \limsup_{s \to 0} N(s)\, s^D .$$

It may be finite or infinite.

The Hausdorff dimension is a refined fractal measure that considers all covers of $\mathcal{S}$ with balls of radius smaller than s. It is most often equal to the capacity dimension, but not always. In the following, the capacity dimension is called *fractal dimension.*

Example 6.7 The Von Koch curve has infinite length because its fractal dimension is $D > 1$. We need $N(s) = 4^n$ balls of size $s = 3^{-n}$ to cover the whole curve, hence

$$N(3^{-n}) = (3^{-n})^{-\log 4/\log 3}.$$

One can verify that at any other scale s, the minimum number of balls $N(s)$ to cover this curve satisfies

$$D = -\liminf_{s \to 0} \frac{\log N(s)}{\log s} = \frac{\log 4}{\log 3}.$$

As expected, it has a fractal dimension between 1 and 2.

Example 6.8 The triadic Cantor set is covered by $N(s) = 2^n$ intervals of size $s = 3^{-n}$, so

$$N(3^{-n}) = (3^{-n})^{-\log 2/\log 3}.$$

One can also verify that

$$D = -\liminf_{s \to 0} \frac{\log N(s)}{\log s} = \frac{\log 2}{\log 3}.$$

Self-Similar Functions Let f be a continuous function with a compact support $\mathcal{S}$. We say that f is *self-similar* if there exist disjoint subsets $\mathcal{S}_1, \ldots, \mathcal{S}_k$ such that the graph of f restricted to each $\mathcal{S}_i$ is an affine transformation of f. This means that there exist a scale $l_i > 1$, a translation r_i, a weight p_i and a constant c_i such that

$$\forall t \in \mathcal{S}_i \ , \ f(t) = c_i + p_i f\Big(l_i(t - r_i)\Big). \tag{6.80}$$

Outside these subsets, we suppose that f is constant. Generalizations of this definition can also be used [110].

If a function is self similar, its wavelet transform is also. Let g be an affine transformation of f:

$$g(t) = p f\Big(l(t - r)\Big) + c. \tag{6.81}$$

Its wavelet transform is

$$Wg(u,s) = \int_{-\infty}^{+\infty} g(t) \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) dt.$$

With the change of variable $t' = l(t-r)$, since ψ has a zero average, the affine relation (6.81) implies

$$Wg(u,s) = \frac{p}{\sqrt{l}} Wf\Big(l(u-r), sl\Big).$$

Suppose that ψ has a compact support included in $[-K,K]$. The affine invariance (6.80) of f over $\mathcal{S}_i = [a_i,b_i]$ produces an affine invariance for all wavelets whose support is included in $\mathcal{S}_i$. For any $s < (b_i - a_i)/K$ and any $u \in [a_i + Ks, b_i - Ks]$,

$$Wf(u,s) = \frac{p_i}{\sqrt{l_i}} Wf\Big(l_i(u-r_i), sl_i\Big).$$

The self-similarity of the wavelet transform implies that the positions and values of its modulus maxima are also self-similar. This can be used to recover unknown affine invariance properties with a voting procedure based on wavelet modulus maxima [218].

Example 6.9 A Cantor measure is constructed over a Cantor set. Let $d\mu_0(x) = dx$ be the uniform Lebesgue measure on $[0,1]$. As in the Cantor set construction, this measure is subdivided into three uniform measures, whose integrals over $[0,1/3]$, $[1/3,2/3]$ and $[2/3,1]$ are respectively p_1, 0 and p_2. We impose $p_1 + p_2 = 1$ to obtain a total measure $d\mu_1$ on $[0,1]$ whose integral is equal to 1. This operation is iteratively repeated by dividing each uniform measure of integral p over $[a,a+l]$ into three equal parts whose integrals are respectively $p_1 p$, 0 and $p_2 p$ over $[a,a+l/3]$, $[a+l/3,a+2l/3]$ and $[a+2l/3,a+l]$. This is illustrated by Figure 6.15. After each subdivision, the resulting measure $d\mu_n$ has a unit integral. In the limit, we obtain a Cantor measure $d\mu_\infty$ of unit integral, whose support is the triadic Cantor set.

Example 6.10 A devil's staircase is the integral of a Cantor measure:

$$f(t) = \int_0^t d\mu_\infty(x). \tag{6.82}$$

It is a continuous function that increases from 0 to 1 on $[0,1]$. The recursive construction of the Cantor measure implies that f is self-similar:

$$f(t) = \begin{cases} p_1 f(3t) & \text{if } t \in [0,1/3] \\ p_1 & \text{if } t \in [1/3,2/3] \\ p_1 + p_2 f(3t-2) & \text{if } t \in [2/3,0] \end{cases}.$$

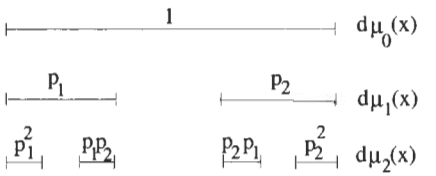


FIGURE 6.15 Two subdivisions of the uniform measure on $[0, 1]$ with left and right weights p_1 and p_2. The Cantor measure $d\mu_\infty$ is the limit of an infinite number of these subdivisions.

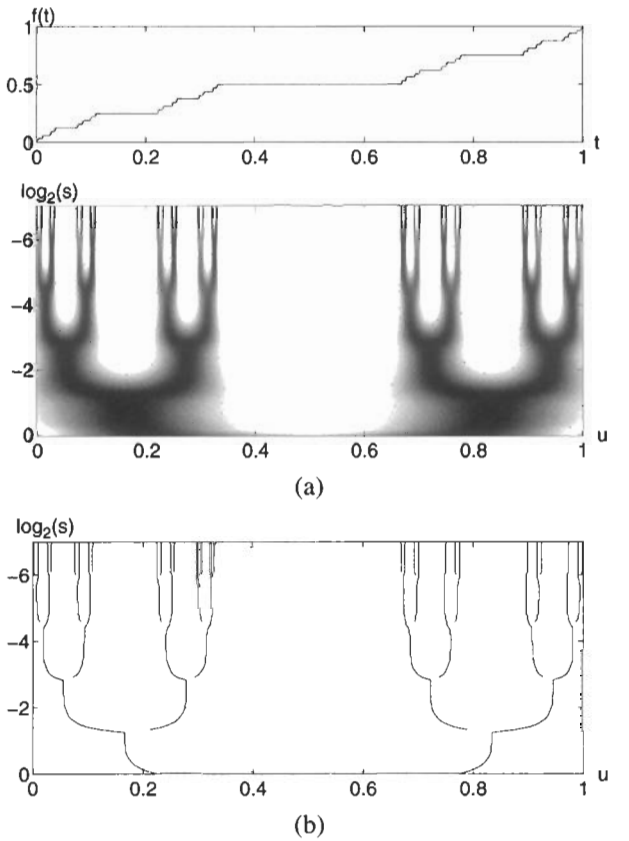


FIGURE 6.16 Devil's staircase calculated from a Cantor measure with equal weights $p_1 = p_2 = 0.5$. (a): Wavelet transform $Wf(u,s)$ computed with $\psi = -\theta'$, where θ is Gaussian. (b): Wavelet transform modulus maxima.

Figure 6.16 displays the devil's staircase obtained with $p_1 = p_2 = 0.5$. The wavelet transform below is calculated with a wavelet that is the first derivative of a Gaussian. The self-similarity of f yields a wavelet transform and modulus maxima that are self-similar. The subdivision of each interval in three parts appears through the multiplication by 2 of the maxima lines, when the scale is multiplied by 3. This Cantor construction is generalized with different interval subdivisions and weight allocations, beginning from the same Lebesgue measure $d\mu_0$ on $[0,1]$ [5].

6.4.2 Singularity Spectrum [3]

Finding the distribution of singularities in a multifractal signal f is particularly important for analyzing its properties. The spectrum of singularity measures the global repartition of singularities having different Lipschitz regularity. The pointwise Lipschitz regularity of f is given by Definition 6.1.

Definition 6.2 (SPECTRUM) *Let $\mathcal{S}_\alpha$ be the set of all points $t \in \mathbb{R}$ where the pointwise Lipschitz regularity of f is equal to α. The spectrum of singularity $D(\alpha)$ of f is the fractal dimension of $\mathcal{S}_\alpha$. The support of $D(\alpha)$ is the set of α such that $\mathcal{S}_\alpha$ is not empty.*

This spectrum was originally introduced by Frisch and Parisi [185] to analyze the homogeneity of multifractal measures that model the energy dissipation of turbulent fluids. It was then extended by Arneodo, Bacry and Muzy [278] to multifractal signals. The fractal dimension definition (6.79) shows that if we make a disjoint cover of the support of f with intervals of size s then the number of intervals that intersect $\mathcal{S}_\alpha$ is

$$N_\alpha(s) \sim s^{-D(\alpha)}. \tag{6.83}$$

The singularity spectrum gives the proportion of Lipschitz α singularities that appear at any scale s. A multifractal f is said to be homogeneous if all singularities have the same Lipschitz exponent α_0, which means the support of $D(\alpha)$ is restricted to $\{\alpha_0\}$. Fractional Brownian motions are examples of homogeneous multifractals.

Partition Function One cannot compute the pointwise Lipschitz regularity of a multifractal because its singularities are not isolated, and the finite numerical resolution is not sufficient to discriminate them. It is however possible to measure the singularity spectrum of multifractals from the wavelet transform local maxima, using a global partition function introduced by Arneodo, Bacry and Muzy [278].

Let ψ be a wavelet with n vanishing moments. Theorem 6.5 proves that if f has pointwise Lipschitz regularity $\alpha_0 < n$ at v then the wavelet transform $Wf(u,s)$ has a sequence of modulus maxima that converges towards v at fine scales. The set of maxima at the scale s can thus be interpreted as a covering of the singular support of f with wavelets of scale s. At these maxima locations

$$|Wf(u,s)| \sim s^{\alpha_0 + 1/2}.$$

Let $\{u_p(s)\}_{p\in\mathbb{Z}}$ be the position of all local maxima of $|Wg(u,s)|$ at a fixed scale s. The partition function $\mathcal{Z}$ measures the sum at a power q of all these wavelet modulus maxima:

$$\mathcal{Z}(q,s) = \sum_p |Wf(u_p,s)|^q. \tag{6.84}$$

At each scale s, any two consecutive maxima u_p and u_{p+1} are supposed to have a distance $|u_{p+1} - u_p| > \epsilon s$, for some $\epsilon > 0$. If not, over intervals of size ϵs, the sum (6.84) includes only the maxima of largest amplitude. This protects the partition function from the multiplication of very close maxima created by fast oscillations.

For each $q \in \mathbb{R}$, the scaling exponent $\tau(q)$ measures the asymptotic decay of $\mathcal{Z}(q,s)$ at fine scales s:

$$\tau(q) = \liminf_{s\to 0} \frac{\log \mathcal{Z}(q,s)}{\log s}.$$

This typically means that

$$\mathcal{Z}(q,s) \sim s^{\tau(q)}.$$

Legendre Transform The following theorem relates $\tau(q)$ to the Legendre transform of $D(\alpha)$ for self-similar signals. This result was established in [83] for a particular class of fractal signals and generalized by Jaffard [222].

Theorem 6.7 (ARNEODO, BACRY, JAFFARD, MUZY) *Let $\Lambda = [\alpha_{\min}, \alpha_{\max}]$ be the support of $D(\alpha)$. Let ψ be a wavelet with $n > \alpha_{\max}$ vanishing moments. If f is a self-similar signal then*

$$\tau(q) = \min_{\alpha\in\Lambda} \Big(q(\alpha + 1/2) - D(\alpha) \Big). \tag{6.85}$$

Proof[3]. The detailed proof is long; we only give an intuitive justification. The sum (6.84) over all maxima positions is replaced by an integral over the Lipschitz parameter. At the scale s, (6.83) indicates that the density of modulus maxima that cover a singularity with Lipschitz exponent α is proportional to $s^{-D(\alpha)}$. At locations where f has Lipschitz regularity α, the wavelet transform decay is approximated by

$$|Wf(u,s)| \sim s^{\alpha+1/2}.$$

It follows that

$$\mathcal{Z}(q,s) \sim \int_\Lambda s^{q(\alpha+1/2)}\, s^{-D(\alpha)}\, d\alpha.$$

When s goes to 0 we derive that $\mathcal{Z}(q,s) \sim s^{\tau(q)}$ for $\tau(q) = \min_{\alpha\in\Lambda}(q(\alpha+1/2) - D(\alpha))$. ■

This theorem proves that the scaling exponent $\tau(q)$ is the Legendre transform of $D(\alpha)$. It is necessary to use a wavelet with enough vanishing moments to measure all Lipschitz exponents up to $\alpha_{\max}$. In numerical calculations $\tau(q)$ is computed by evaluating the sum $\mathcal{Z}(q,s)$. We thus need to invert the Legendre transform (6.85) to recover the spectrum of singularity $D(\alpha)$.

Proposition 6.2 • *The scaling exponent $\tau(q)$ is a convex and increasing function of q.*

• *The Legendre transform (6.85) is invertible if and only if $D(\alpha)$ is convex, in which case*

$$D(\alpha) = \min_{q \in \mathbb{R}} \Big(q(\alpha + 1/2) - \tau(q) \Big). \tag{6.86}$$

The spectrum $D(\alpha)$ of self-similar signals is convex.

Proof[3]. The proof that $D(\alpha)$ is convex for self-similar signals can be found in [222]. We concentrate on the properties of the Legendre transform that are important in numerical calculations. To simplify the proof, let us suppose that $D(q)$ is twice differentiable. The minimum of the Legendre transform (6.85) is reached at a critical point $q(\alpha)$. Computing the derivative of $q(\alpha+1/2) - D(\alpha)$ with respect to α gives

$$q(\alpha) = \frac{dD}{d\alpha}, \tag{6.87}$$

with

$$\tau(q) = q\left(\alpha + \frac{1}{2}\right) - D(\alpha). \tag{6.88}$$

Since it is a minimum, the second derivative of $\tau(q(\alpha))$ with respect to α is negative, from which we derive that

$$\frac{d^2 D(\alpha(q))}{d\alpha^2} \leq 0.$$

This proves that $\tau(q)$ depends only on the values where $D(\alpha)$ has a negative second derivative. We can thus recover $D(\alpha)$ from $\tau(q)$ only if it is convex.

The derivative of $\tau(q)$ is

$$\frac{d\tau(q)}{dq} = \alpha + \frac{1}{2} + q\frac{d\alpha}{dq} - \frac{d\alpha}{dq}\frac{dD(\alpha)}{d\alpha} = \alpha + \frac{1}{2} \geq 0. \tag{6.89}$$

It is therefore increasing. Its second derivative is

$$\frac{d^2\tau(q)}{dq^2} = \frac{d\alpha}{dq}.$$

Taking the derivative of (6.87) with respect to q proves that

$$\frac{d\alpha}{dq}\frac{d^2 D(\alpha)}{d\alpha^2} = 1.$$

Since $\frac{d^2 D(\alpha)}{d\alpha^2} \leq 0$ we derive that $\frac{d^2\tau(q)}{dq^2} \leq 0$. Hence $\tau(q)$ is convex. By using (6.88), (6.89) and the fact that $\tau(q)$ is convex, we verify that

$$D(\alpha) = \min_{q \in \mathbb{R}} \Big(q(\alpha + 1/2) - \tau(q) \Big).$$

■

The spectrum $D(\alpha)$ of self-similar signals is convex and can therefore be calculated from $\tau(q)$ with the inverse Legendre formula (6.86). This formula is also valid for a much larger class of multifractals. For example, it is verified for statistical self-similar signals such as realizations of fractional Brownian motions. Multifractals having some stochastic self-similarity have a spectrum that can often be calculated as an inverse Legendre transform (6.86). However, let us emphasize that this formula is not exact for any function f because its spectrum of singularity $D(\alpha)$ is not necessarily convex. In general, Jaffard proved [222] that the Legendre transform (6.86) gives only an upper bound of $D(\alpha)$. These singularity spectrum properties are studied in detail in [49].

Figure 6.17 illustrates the properties of a convex spectrum $D(\alpha)$. The Legendre transform (6.85) proves that its maximum is reached at

$$D(\alpha_0) = \max_{\alpha \in \Lambda} D(\alpha) = -\tau(0).$$

It is the fractal dimension of the Lipschitz exponent α_0 most frequently encountered in f. Since all other Lipschitz α singularities appear over sets of lower dimension, if $\alpha_0 < 1$ then $D(\alpha_0)$ is also the fractal dimension of the singular support of f. The spectrum $D(\alpha)$ for $\alpha < \alpha_0$ depends on $\tau(q)$ for $q > 0$, and for $\alpha > \alpha_0$ it depends on $\tau(q)$ for $q < 0$.

Numerical Calculations To compute $D(\alpha)$, we assume that the Legendre transform formula (6.86) is valid. We first calculate $\mathcal{Z}(q,s) = \sum_p |Wf(u_p,s)|^q$, then derive the decay scaling exponent $\tau(q)$, and finally compute $D(\alpha)$ with a Legendre transform. If $q < 0$ then the value of $\mathcal{Z}(q,s)$ depends mostly on the small amplitude maxima $|Wf(u_p,s)|$. Numerical calculations may then become unstable. To avoid introducing spurious modulus maxima created by numerical errors in regions where f is nearly constant, wavelet maxima are chained to produce maxima curve across scales. If $\psi = (-1)^p \theta^{(p)}$ where θ is a Gaussian, Proposition 6.1 proves that all maxima lines $u_p(s)$ define curves that propagate up to the limit $s = 0$. All maxima lines that do not propagate up to the finest scale are thus removed in the calculation of $\mathcal{Z}(q,s)$. The calculation of the spectrum $D(\alpha)$ proceeds as follows.

1. *Maxima* Compute $Wf(u,s)$ and the modulus maxima at each scale s. Chain the wavelet maxima across scales.
2. *Partition function* Compute
$$\mathcal{Z}(q,s) = \sum_p |Wf(u_p,s)|^q .$$
3. *Scaling* Compute $\tau(q)$ with a linear regression of $\log_2 \mathcal{Z}(s,q)$ as a function of $\log_2 s$:
$$\log_2 \mathcal{Z}(q,s) \approx \tau(q) \log_2 s + C(q) .$$
4. *Spectrum* Compute
$$D(\alpha) = \min_{q \in \mathbb{R}} \Big(q(\alpha + 1/2) - \tau(q) \Big) .$$

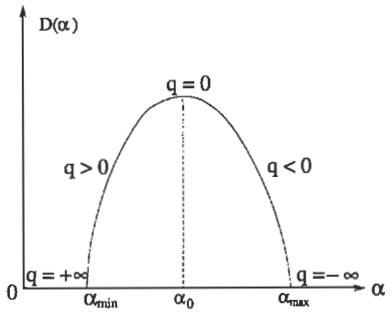


FIGURE 6.17 Convex spectrum $D(\alpha)$.

Example 6.11 The spectrum of singularity $D(\alpha)$ of the devil's staircase (6.82) is a convex function that can be calculated analytically [203]. Suppose that $p_1 < p_2$. The support of $D(\alpha)$ is $[\alpha_{\min}, \alpha_{\max}]$ with

$$\alpha_{\min} = \frac{-\log p_2}{\log 3} \text{ and } \alpha_{\max} = \frac{-\log p_1}{\log 3}.$$

If $p_1 = p_2 = 1/2$ then the support of $D(\alpha)$ is reduced to a point, which means that all the singularities of f have the same Lipschitz $\log 2/\log 3$ regularity. The value $D(\log 2/\log 3)$ is then the fractal dimension of the triadic Cantor set and is thus equal to $\log 2/\log 3$.

Figure 6.18(a) shows a devil's staircase calculated with $p_1 = 0.4$ and $p_2 = 0.6$. Its wavelet transform is computed with $\psi = -\theta'$, where θ is a Gaussian. The decay of $\log_2 \mathcal{Z}(q,s)$ as a function of $\log_2 s$ is shown in Figure 6.18(b) for several values of q. The resulting $\tau(q)$ and $D(\alpha)$ are given by Figures 6.18(c,d). There is no numerical instability for $q < 0$ because there is no modulus maximum whose amplitude is close to zero. This is not the case if the wavelet transform is calculated with a wavelet that has more vanishing moments.

Smooth Perturbations Let f be a multifractal whose spectrum of singularity $D(\alpha)$ is calculated from $\tau(q)$. If a $\mathbf{C}^\infty$ signal g is added to f then the singularities are not modified and the singularity spectrum of $\tilde{f} = f + g$ remains $D(\alpha)$. We study the effect of this smooth perturbation on the spectrum calculation.

The wavelet transform of $\tilde{f}$ is

$$W\tilde{f}(u,s) = Wf(u,s) + Wg(u,s).$$

Let $\tau(q)$ and $\tilde{\tau}(q)$ be the scaling exponent of the partition functions $\mathcal{Z}(q,s)$ and $\tilde{\mathcal{Z}}(q,s)$ calculated from the modulus maxima respectively of $Wf(u,s)$ and $W\tilde{f}(u,s)$. We denote by $D(\alpha)$ and $\tilde{D}(\alpha)$ the Legendre transforms respectively of $\tau(q)$ and $\tilde{\tau}(q)$. The following proposition relates $\tau(q)$ and $\tilde{\tau}(q)$.

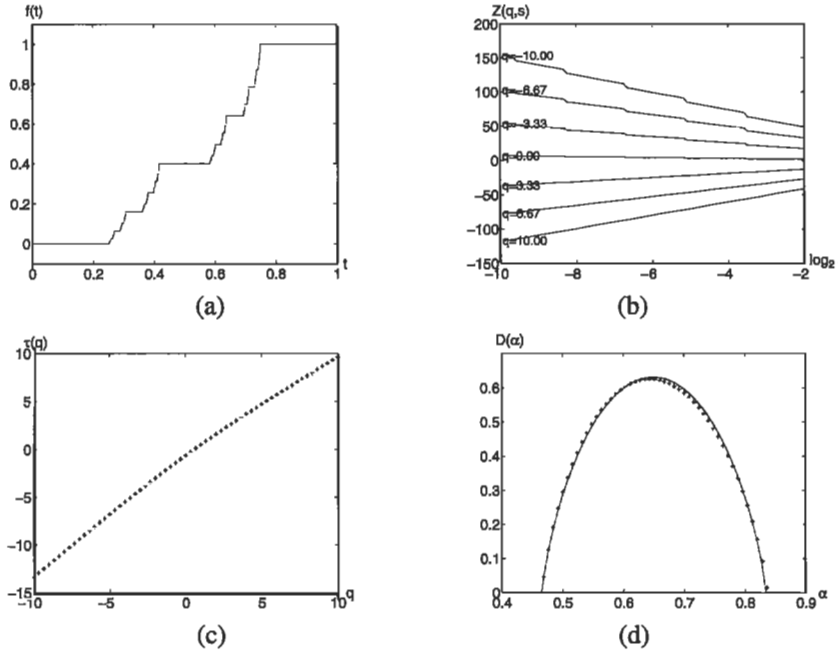


FIGURE 6.18 (a): Devil's staircase with $p_1 = 0.4$ and $p_2 = 0.6$. (b): Partition function $\mathcal{Z}(q,s)$ for several values of q. (c): Scaling exponent $\tau(q)$. (d): The theoretical spectrum $D(\alpha)$ is shown with a solid line. The + are the spectrum values calculated numerically with a Legendre transform of $\tau(q)$.

Proposition 6.3 (ARNEODO, BACRY, MUZY) *Let ψ be a wavelet with exactly n vanishing moments. Suppose that f is a self-similar function.*

- *If g is a polynomial of degree $p < n$ then $\tau(q) = \tilde{\tau}(q)$ for all $q \in \mathbb{R}$.*
- *If $g^{(n)}$ is almost everywhere non-zero then*

$$\tilde{\tau}(q) = \begin{cases} \tau(q) & \text{if } q > q_c \\ (n+1/2)\,q & \text{if } q \leq q_c \end{cases} \tag{6.90}$$

where q_c is defined by $\tau(q_c) = (n+1/2)q_c$.

Proof[3]. If g is a polynomial of degree $p < n$ then $Wg(u,s) = 0$. The addition of g does not modify the calculation of the singularity spectrum based on wavelet maxima, so $\tau(q) = \tilde{\tau}(q)$ for all $q \in \mathbb{R}$.

If g is a $\mathbf{C}^\infty$ function that is not a polynomial then its wavelet transform is generally non-zero. We justify (6.91) with an intuitive argument that is not a proof. A rigorous proof can be found in [83]. Since ψ has exactly n vanishing moments, (6.15) proves that

$$|Wg(u,s)| \sim K\, s^{n+1/2}\, g^{(n)}(u).$$

We suppose that $g^{(n)}(u) \neq 0$. For $\tau(q) \leq (n+1/2)q$, since $|Wg(u,s)|^q \sim s^{q(n+1/2)}$ has a faster asymptotic decay than $s^{\tau(q)}$ when s goes to zero, one can verify that $\tilde{\mathcal{Z}}(q,s)$ and $\mathcal{Z}(q,s)$ have the same scaling exponent, $\tilde{\tau}(q) = \tau(q)$. If $\tau(q) > (n+1/2)q$, which means that $q \leq q_c$, then the decay of $|W\tilde{f}(u,s)|^q$ is controlled by the decay of $|Wg(u,s)|^q$, so $\tilde{\tau}(q) = (n+1/2)q$. ■

This proposition proves that the addition of a non-polynomial smooth function introduces a bias in the calculation of the singularity spectrum. Let α_c be the critical Lipschitz exponent corresponding to q_c:

$$D(\alpha_c) = q_c\,(\alpha_c + 1/2) - \tau(q_c).$$

The Legendre transform of $\tilde{\tau}(q)$ in (6.90) yields

$$\tilde{D}(\alpha) = \begin{cases} D(\alpha) & \text{if } \alpha \leq \alpha_c \\ 0 & \text{if } \alpha = n \\ -\infty & \text{if } \alpha > \alpha_c \text{ and } \alpha \neq n \end{cases}. \tag{6.91}$$

This modification is illustrated by Figure 6.19.

The bias introduced by the addition of smooth components can be detected experimentally by modifying the number n of vanishing moments of ψ. Indeed the value of q_c depends on n. If the singularity spectrum varies when changing the number of vanishing moments of the wavelet then it indicates the presence of a bias.

6.4.3 Fractal Noises [3]

Fractional Brownian motions are statistically self-similar Gaussian processes that give interesting models for a wide class of natural phenomena [265]. Despite their non-stationarity, one can define a power spectrum that has a power decay. Realizations of fractional Brownian motions are almost everywhere singular, with the same Lipschitz regularity at all points.

We often encounter fractal noise processes that are not Gaussian although their power spectrum has a power decay. Realizations of these processes may include singularities of various types. The spectrum of singularity is then important in analyzing their properties. This is illustrated by an application to hydrodynamic turbulence.

Definition 6.3 (FRACTIONAL BROWNIAN MOTION) *A fractional Brownian motion of Hurst exponent $0 < H < 1$ is a zero-mean Gaussian process B_H such that*

$$B_H(0) = 0,$$

and

$$\mathrm{E}\{|B_H(t) - B_H(t-\Delta)|^2\} = \sigma^2 |\Delta|^{2H}. \tag{6.92}$$

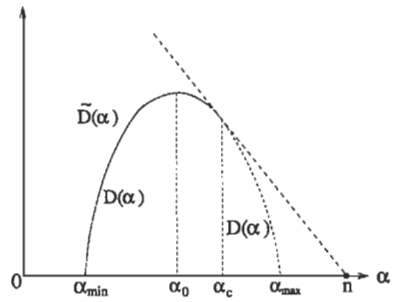


FIGURE 6.19 If ψ has n vanishing moments, in presence of a $\mathbf{C}^\infty$ perturbation the computed spectrum $\tilde{D}(\alpha)$ is identical to the true spectrum $D(\alpha)$ for $\alpha \leq \alpha_c$. Its support is reduced to $\{n\}$ for $\alpha > \alpha_c$.

Property (6.92) imposes that the deviation of $|B_H(t) - B_H(t-\Delta)|$ be proportional to $|\Delta|^H$. As a consequence, one can prove that any realization f of B_H is almost everywhere singular with a pointwise Lipschitz regularity $\alpha = H$. The smaller H, the more singular f. Figure 6.20(a) shows the graph of one realization for $H = 0.7$.

Setting $\Delta = t$ in (6.92) yields

$$\mathsf{E}\{|B_H(t)|^2\} = \sigma^2 |t|^{2H}.$$

Developing (6.92) for $\Delta = t - u$ also gives

$$\mathsf{E}\{B_H(t)\, B_H(u)\} = \frac{\sigma^2}{2}\left(|t|^{2H} + |u|^{2H} - |t-u|^{2H}\right). \tag{6.93}$$

The covariance does not depend only on $t - u$, which proves that a fractional Brownian motion is non-stationary.

The statistical self-similarity appears when scaling this process. One can derive from (6.93) that for any $s > 0$

$$\mathsf{E}\{B_H(st)\, B_H(su)\} = \mathsf{E}\{s^H B_H(t)\, s^H B_H(u)\}.$$

Since $B_H(st)$ and $s^H B_H(t)$ are two Gaussian processes with same mean and same covariance, they have the same probability distribution

$$B_H(st) \equiv s^H B_H(t),$$

where $\equiv$ denotes an equality of finite-dimensional distributions.

Power Spectrum Although B_H is not stationary, one can define a generalized power spectrum. This power spectrum is introduced by proving that the increments of a fractional Brownian motion are stationary, and by computing their power spectrum [78].

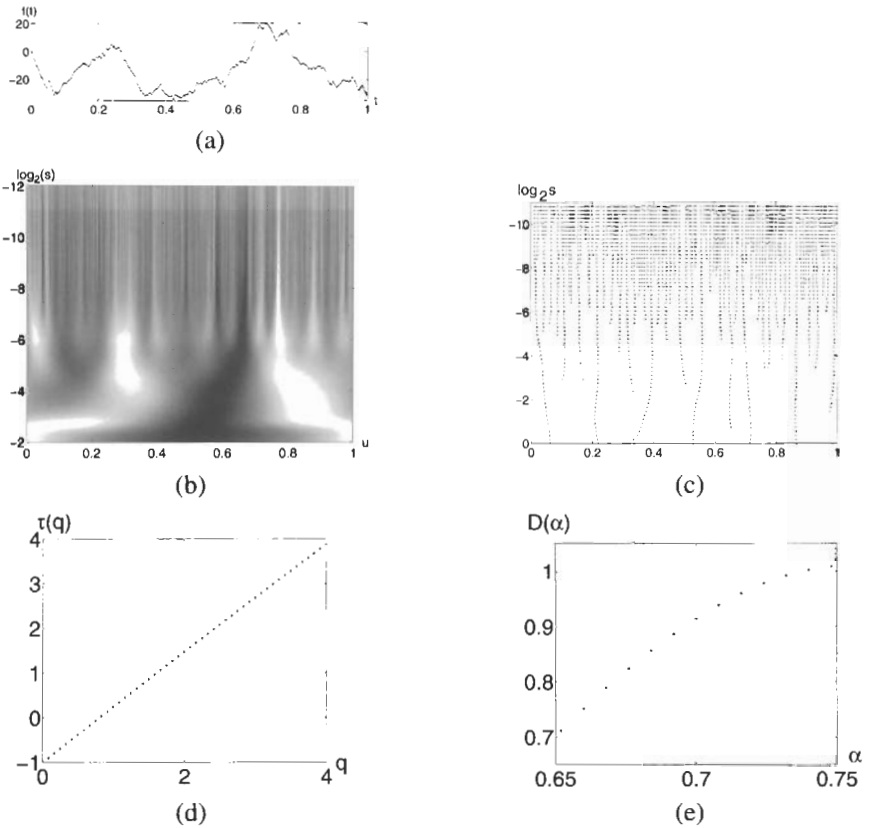


FIGURE 6.20 (a): One realization of a fractional Brownian motion for a Hurst exponent $H = 0.7$. (b): Wavelet transform. (c): Modulus maxima of its wavelet transform. (d): Scaling exponent $\tau(q)$. (e): Resulting $D(\alpha)$ over its support.

Proposition 6.4 *Let* $g_\Delta(t) = \delta(t) - \delta(t-\Delta)$. *The increment*

$$I_{H,\Delta}(t) = B_H \star g_\Delta(t) = B_H(t) - B_H(t-\Delta) \tag{6.94}$$

is a stationary process whose power spectrum is

$$\hat{R}_{I_{H,\Delta}}(\omega) = \frac{\sigma_H^2}{|\omega|^{2H+1}} |\hat{g}_\Delta(\omega)|^2. \tag{6.95}$$

Proof[2]. The covariance of $I_{H,\Delta}$ is computed with (6.93):

$$\mathrm{E}\{I_{H,\Delta}(t)\, I_{H,\Delta}(t-\tau)\} = \frac{\sigma^2}{2}(|\tau-\Delta|^{2H} + |\tau+\Delta|^{2H} - 2|\tau|^{2H}) = R_{I_{H,\Delta}}(\tau). \tag{6.96}$$

The power spectrum $\hat{R}_{I_{H,\Delta}}(\omega)$ is the Fourier transform of $R_{I_{H,\Delta}}(\tau)$. One can verify that the Fourier transform of the distribution $f(\tau) = |\tau|^{2H}$ is $\hat{f}(\omega) = -\lambda_H |\omega|^{-(2H+1)}$,

with $\lambda_H > 0$. We thus derive that the Fourier transform of (6.96) can be written

$$\hat{R}_{I_{H,\Delta}}(\omega) = 2\sigma^2 \lambda_H |\omega|^{-(2H+1)} \sin^2 \frac{\Delta\omega}{2},$$

which proves (6.95) for $\sigma_H^2 = \sigma^2 \lambda_H / 2$. ■

If $X(t)$ is a stationary process then we know that $Y(t) = X \star g(t)$ is also stationary and the power spectrum of both processes is related by

$$\hat{R}_X(\omega) = \frac{\hat{R}_Y(\omega)}{|\hat{g}(\omega)|^2}. \tag{6.97}$$

Although $B_H(t)$ is not stationary, Proposition 6.4 proves that $I_{H,\Delta}(t) = B_H \star g_\Delta(t)$ is stationary. As in (6.97), it is tempting to define a "generalized" power spectrum calculated with (6.95):

$$\hat{R}_{B_H}(\omega) = \frac{\hat{R}_{I_{H,\Delta}}(\omega)}{|\hat{g}_\Delta(\omega)|^2} = \frac{\sigma_H^2}{|\omega|^{2H+1}}. \tag{6.98}$$

The non-stationarity of $B_H(t)$ appears in the energy blow-up at low frequencies. The increments $I_{H,\Delta}(t)$ are stationary because the multiplication by $|\hat{g}_\Delta(\omega)|^2 = O(\omega^2)$ removes the explosion of the low frequency energy. One can generalize this result and verify that if g is an arbitrary stable filter whose transfer function satisfies $|\hat{g}(\omega)| = O(\omega)$, then $Y(t) = B_H \star g(t)$ is a stationary Gaussian process whose power spectrum is

$$\hat{R}_Y(\omega) = \frac{\sigma_H^2}{|\omega|^{2H+1}} |\hat{g}(\omega)|^2. \tag{6.99}$$

Wavelet Transform The wavelet transform of a fractional Brownian motion is

$$WB_H(u,s) = B_H \star \bar{\psi}_s(u). \tag{6.100}$$

Since ψ has a least one vanishing moment, necessarily $|\hat{\psi}(\omega)| = O(\omega)$ in the neighborhood of $\omega = 0$. The wavelet filter $g = \bar{\psi}_s$ has a Fourier transform $\hat{g}(\omega) = \sqrt{s}\hat{\psi}^*(s\omega) = O(\omega)$ near $\omega = 0$. This proves that for a fixed s the process $Y_s(u) = WB_H(u,s)$ is a Gaussian stationary process [181], whose power spectrum is calculated with (6.99):

$$\hat{R}_{Y_s}(\omega) = s|\hat{\psi}(s\omega)|^2 \frac{\sigma_H^2}{|\omega|^{2H+1}} = s^{2H+2} \hat{R}_{Y_1}(s\omega). \tag{6.101}$$

The self-similarity of the power spectrum and the fact that B_H is Gaussian are sufficient to prove that $WB_H(u,s)$ is self-similar across scales:

$$WB_H(u,s) \equiv s^{H+1/2} WB_H\left(\frac{u}{s}, 1\right),$$

where the equivalence means that they have same finite distributions. Interesting characterizations of fractional Brownian motion properties are also obtained by decomposing these processes in wavelet bases [49, 78, 357].

Example 6.12 Figure 6.20(a) displays one realization of a fractional Brownian with $H = 0.7$. The wavelet transform and its modulus maxima are shown in Figures 6.20(b) and 6.20(c). The partition function (6.84) is computed from the wavelet modulus maxima. Figure 6.20(d) gives the scaling exponent $\tau(q)$, which is nearly a straight line. Fractional Brownian motions are homogeneous fractals with Lipschitz exponents equal to H. In this example, the theoretical spectrum $D(\alpha)$ has therefore a support reduced to $\{0.7\}$ with $D(0.7) = 1$. The estimated spectrum in Figure 6.20(e) is calculated with a Legendre transform of $\tau(q)$. Its support is $[0.65, 0.75]$. There is an estimation error because the calculations are performed on a signal of finite size.

Fractal Noises Some physical phenomena produce more general fractal noises $X(t)$, which are not Gaussian processes, but which have stationary increments. As for fractional Brownian motions, one can define a "generalized" power spectrum that has a power decay

$$\hat{R}_X(\omega) = \frac{\sigma_H^2}{|\omega|^{2H+1}}.$$

These processes are transformed into a wide-sense stationary process by a convolution with a stable filter g which removes the lowest frequencies $|\hat{g}(\omega)| = O(\omega)$. One can thus derive that the wavelet transform $Y_s(u) = WX(u,s)$ is a stationary process at any fixed scale s. Its spectrum is the same as the spectrum (6.101) of fractional Brownian motions. If $H < 1$, the asymptotic decay of $\hat{R}_X(\omega)$ indicates that realizations of $X(t)$ are singular functions but it gives no information on the distribution of these singularities. As opposed to fractional Brownian motions, general fractal noises have realizations that may include singularities of various types. Such multifractals are differentiated from realizations of fractional Brownian motions by computing their singularity spectrum $D(\alpha)$. For example, the velocity fields of fully developed turbulent flows have been modeled by fractal noises, but the calculation of the singularity spectrum clearly shows that these flows differ in important ways from fractional Brownian motions.

Hydrodynamic Turbulence Fully developed turbulence appears in incompressible flows at high Reynolds numbers. Understanding the properties of hydrodynamic turbulence is a major problem of modern physics, which remains mostly open despite an intense research effort since the first theory of Kolmogorov in 1941 [237]. The number of degrees of liberty of three-dimensional turbulence is considerable, which produces extremely complex spatio-temporal behavior. No formalism is yet able to build a statistical-physics framework based on the Navier-Stokes equations, that would enable us to understand the global behavior of turbulent flows, at it is done in thermodynamics.

In 1941, Kolmogorov [237] formulated a statistical theory of turbulence. The velocity field is modeled as a process $V(x)$ whose increments have a variance

$$\mathrm{E}\{|V(x+\Delta) - V(x)|^2\} \sim \epsilon^{2/3} \Delta^{2/3}.$$

The constant ϵ is a rate of dissipation of energy per unit of mass and time, which is supposed to be independent of the location. This indicates that the velocity field is statistically homogeneous with Lipschitz regularity $\alpha = H = 1/3$. The theory predicts that a one-dimensional trace of a three-dimensional velocity field is a fractal noise process with stationary increments, and whose spectrum decays with a power exponent $2H+1 = 5/3$:

$$\hat{R}_V(\omega) = \frac{\sigma_H^2}{|\omega|^{5/3}}.$$

The success of this theory comes from numerous experimental verifications of this power spectrum decay. However, the theory does not take into account the existence of coherent structures such as vortices. These phenomena contradict the hypothesis of homogeneity, which is at the root of Kolmogorov's 1941 theory.

Kolmogorov [238] modified the homogeneity assumption in 1962, by introducing an energy dissipation rate $\epsilon(x)$ that varies with the spatial location x. This opens the door to "local stochastic self-similar" multifractal models, first developed by Mandelbrot [264] to explain energy exchanges between fine-scale structures and large-scale structures. The spectrum of singularity $D(\alpha)$ is playing an important role in testing these models [185]. Calculations with wavelet maxima on turbulent velocity fields [5] show that $D(\alpha)$ is maximum at $1/3$, as predicted by the Kolmogorov theory. However, $D(\alpha)$ does not have a support reduced to $\{1/3\}$, which verifies that a turbulent velocity field is not a homogeneous process. Models based on the wavelet transform were recently introduced to explain the distribution of vortices in turbulent fluids [12, 179, 180].

6.5 PROBLEMS

6.1. [1] *Lipschitz regularity*

(a) Prove that if f is uniformly Lipschitz α on $[a,b]$ then it is pointwise Lipschitz α at all $t_0 \in [a,b]$.

(b) Show that $f(t) = t \sin t^{-1}$ is Lipschitz 1 at all $t_0 \in [-1,1]$ and verify that it is uniformly Lipschitz α over $[-1,1]$ only for $\alpha \leq 1/2$. Hint: consider the points $t_n = (n+1/2)^{-1}\pi^{-1}$.

6.2. [1] *Regularity of derivatives*

(a) Prove that f is uniformly Lipschitz $\alpha > 1$ over $[a,b]$ if and only if f' is uniformly Lipschitz $\alpha - 1$ over $[a,b]$.

(b) Show that f may be pointwise Lipschitz $\alpha > 1$ at t_0 while f' is not pointwise Lipschitz $\alpha - 1$ at t_0. Consider $f(t) = t^2 \cos t^{-1}$ at $t = 0$.

6.3. [1] Find $f(t)$ which is uniformly Lipschitz 1 but does not satisfy the sufficient Fourier condition (6.1).

6.4. [1] Let $f(t) = \cos \omega_0 t$ and $\psi(t)$ be a wavelet that is symmetric about 0.

(a) Verify that

$$Wf(u,s) = \sqrt{s}\,\hat{\psi}(s\omega_0)\cos\omega_0 t\,.$$

(b) Find the equations of the curves of wavelet modulus maxima in the time-scale plane (u,s). Relate the decay of $|Wf(u,s)|$ along these curves to the number n of vanishing moments of ψ.

6.5. [1] Let $f(t) = |t|^\alpha$. Show that $Wf(u,s) = s^{\alpha+1/2} Wf(u/s,1)$. Prove that it is not sufficient to measure the decay of $|Wf(u,s)|$ when s goes to zero at $u = 0$ in order to compute the Lipschitz regularity of f at $t = 0$.

6.6. [2] Let $f(t) = |t|^\alpha \sin |t|^{-\beta}$ with $\alpha > 0$ and $\beta > 0$. What is the pointwise Lipschitz regularity of f and f' at $t = 0$? Find the equation of the ridge curve in the (u,s) plane along which the high amplitude wavelet coefficients $|Wf(u,s)|$ converge to $t = 0$ when s goes to zero. Compute the maximum values of α and α' such that $Wf(u,s)$ satisfy (6.21).

6.7. [1] For a complex wavelet, we call *lines of constant phase* the curves in the (u,s) plane along which the complex phase of $Wf(u,s)$ remains constant when s varies.

(a) If $f(t) = |t|^\alpha$, prove that the lines of constant phase converge towards the singularity at $t = 0$ when s goes to zero. Verify this numerically in WAVELAB.

(b) Let ψ be a real wavelet and $Wf(u,s)$ be the real wavelet transform of f. Show that the modulus maxima of $Wf(u,s)$ correspond to lines of constant phase of an analytic wavelet transform, which is calculated with a particular analytic wavelet ψ^a that you will specify.

6.8. [2] Prove that if $f = 1_{[0,+\infty)}$ then the number of modulus maxima of $Wf(u,s)$ at each scale s is larger than or equal to the number of vanishing moments of ψ.

6.9. [1] The spectrum of singularity of the Riemann function

$$f(t) = \sum_{n=-\infty}^{+\infty} \frac{1}{n^2} \sin n^2 t$$

is defined on its support by $D(\alpha) = 4\alpha - 2$ if $\alpha \in [1/2, 3/4]$ and $D(3/2) = 0$ [213, 222]. Verify this result numerically with WAVELAB, by computing this spectrum from the partition function of a wavelet transform modulus maxima.

6.10. [2] Let $\psi = -\theta'$ where θ is a positive window of compact support. If f is a Cantor devil's staircase, prove that there exist lines of modulus maxima that converge towards each singularity.

6.11. [2] Implement in WAVELAB an algorithm that detects oscillating singularities by following the ridges of an analytic wavelet transform when the scale s decreases. Test your algorithm on $f(t) = \sin t^{-1}$.

6.12. [2] Implement in WAVELAB an algorithm that reconstructs a signal from the local maxima of its dyadic wavelet transform, with the frame algorithm of Section 6.2.2. Compare numerically the speed of convergence and the precision of the reconstruction when using the frame operator (6.49) that incorporates the extrema condition and the reduced frame operator (6.51).

6.13. [2] Let $X[n] = f[n] + W[n]$ be a signal of size N, where W is a Gaussian white noise of variance σ^2. Implement in WAVELAB an estimator of f which thresholds at $T = \sigma\sqrt{2 \log_e N}$ the maxima of a dyadic wavelet transform of X. The

estimation of f is reconstructed from the thresholded maxima representation with the frame algorithm of Section 6.2.2. Compare numerically this estimator with a thresholding estimator in a wavelet orthonormal basis.

6.14. [2] Let $\theta(t)$ be a Gaussian of variance 1.

(a) Prove that the Laplacian of a two-dimensional Gaussian

$$\psi(x_1,x_2) = \frac{\partial^2\theta(x_1)}{\partial x^2}\,\theta(x_2) + \theta(x_1)\,\frac{\partial^2\theta(x_2)}{\partial x_2^2}$$

satisfies the dyadic wavelet condition (5.91) (there is only 1 wavelet).

(b) Explain why the zero-crossings of this dyadic wavelet transform provide the locations of multiscale edges in images. Compare the position of these zero-crossings with the wavelet modulus maxima obtained with $\psi^1(x_1,x_2) = -\theta'(x_1)\,\theta(x_2)$ and $\psi^2(x_1,x_2) = -\theta(x_1)\,\theta'(x_2)$.

6.15. [1] The covariance of a fractional Brownian motion $B_H(t)$ is given by (6.93). Show that the wavelet transform at a scale s is stationary by verifying that

$$\mathrm{E}\left\{ WB_H(u_1,s)\,WB_H(u_2,s) \right\} = -\frac{\sigma^2}{2}\,s^{2H+1}\int_{-\infty}^{+\infty} |t|^{2H}\,\Psi\left(\frac{u_1-u_2}{s} - t\right) dt,$$

with $\Psi(t) = \psi \star \bar{\psi}(t)$ and $\bar{\psi}(t) = \psi(-t)$.

6.16. [2] Let $X(t)$ be a stationary Gaussian process whose covariance $R_X(\tau) = \mathrm{E}\{X(t)X(t-\tau)\}$ is twice differentiable. One can prove that the average number of zero-crossings over an interval of size 1 is $-\pi R_X''(0)\,(\pi^2 R_X(0))^{-1}$ [56]. Let $B_H(t)$ be a fractional Brownian motion and ψ a wavelet that is C^2. Prove that the average numbers respectively of zero-crossings and of modulus maxima of $WB_H(u,s)$ for $u \in [0,1]$ are proportional to s. Verify this result numerically in WAVELAB.

6.17. [3] We want to interpolate the samples of a discrete signal $f(n/N)$ without blurring its singularities, by extending its dyadic wavelet transform at finer scales with an interpolation procedure on its modulus maxima. The modulus maxima are calculated at scales $2^j > N^{-1}$. Implement in WAVELAB an algorithm that creates a new set of modulus maxima at the finer scale N^{-1}, by interpolating across scales the amplitudes and positions of the modulus maxima calculated at $2^j > N^{-1}$. Reconstruct a signal of size $2N$ by adding these fine scale modulus maxima to the maxima representation of the signal.

6.18. [3] Implement an algorithm that estimates the Lipschitz regularity α and the smoothing scale σ of sharp variation points in one-dimensional signals by applying the result of Theorem 6.6 on the dyadic wavelet transform maxima. Extend Theorem 6.6 for two-dimensional signals and find an algorithm that computes the same parameters for edges in images.

6.19. [3] Construct a compact image code from multiscale wavelet maxima [261]. An efficient coding algorithm must be introduced to store the positions of the "important" multiscale edges as well as the modulus and the angle values of the wavelet transform along these edges. Do not forget that the wavelet transform angle is nearly orthogonal to the tangent of the edge curve. Use the image reconstruction algorithm of Section 6.3.2 to recover an image from this coded representation.

6.20. [3] A generalized Cantor measure is defined with a renormalization that transforms the uniform measure on $[0,1]$ into a measure equal to p_1, 0 and p_2 respectively on $[0,l_1]$, $[l_1,l_2]$ and $[l_2,1]$, with $p_1+p_2=1$. Iterating infinitely many times this renormalization operation over each component of the resulting measures yields a Cantor measure. The integral (6.82) of this measure is a devil's staircase. Suppose that l_1, l_2, p_1 and p_2 are unknown. Find an algorithm that computes these renormalization parameters by analyzing the self-similarity properties of the wavelet transform modulus maxima across scales. This problem is important in order to identify renormalization maps in experimental data obtained from physical experiments.

VII

WAVELET BASES

One can construct wavelets ψ such that the dilated and translated family

$$\left\{\psi_{j,n}(t) = \frac{1}{\sqrt{2^j}}\psi\left(\frac{t-2^j n}{2^j}\right)\right\}_{(j,n)\in\mathbb{Z}^2}$$

is an orthonormal basis of $\mathbf{L}^2(\mathbb{R})$. Behind this simple statement lie very different points of view which open a fruitful exchange between harmonic analysis and discrete signal processing.

Orthogonal wavelets dilated by 2^j carry signal variations at the resolution 2^{-j}. The construction of these bases can thus be related to multiresolution signal approximations. Following this link leads us to an unexpected equivalence between wavelet bases and conjugate mirror filters used in discrete multirate filter banks. These filter banks implement a fast orthogonal wavelet transform that requires only $O(N)$ operations for signals of size N. The design of conjugate mirror filters also gives new classes of wavelet orthogonal bases including regular wavelets of compact support. In several dimensions, wavelet bases of $\mathbf{L}^2(\mathbb{R}^d)$ are constructed with separable products of functions of one variable.

7.1 ORTHOGONAL WAVELET BASES [1]

Our search for orthogonal wavelets begins with multiresolution approximations. For $f \in \mathbf{L}^2(\mathbb{R})$, the partial sum of wavelet coefficients $\sum_{n=-\infty}^{+\infty} \langle f, \psi_{j,n}\rangle \psi_{j,n}$ can indeed be interpreted as the difference between two approximations of f at the

resolutions 2^{-j+1} and 2^{-j}. Multiresolution approximations compute the approximation of signals at various resolutions with orthogonal projections on different spaces $\{\mathbf{V}_j\}_{j\in\mathbb{Z}}$. Section 7.1.3 proves that multiresolution approximations are entirely characterized by a particular discrete filter that governs the loss of information across resolutions. These discrete filters provide a simple procedure for designing and synthesizing orthogonal wavelet bases.

7.1.1 Multiresolution Approximations

Adapting the signal resolution allows one to process only the relevant details for a particular task. In computer vision, Burt and Adelson [108] introduced a multiresolution pyramid that can be used to process a low-resolution image first and then selectively increase the resolution when necessary. This section formalizes multiresolution approximations, which set the ground for the construction of orthogonal wavelets.

The approximation of a function f at a resolution 2^{-j} is specified by a discrete grid of samples that provides local averages of f over neighborhoods of size proportional to 2^j. A multiresolution approximation is thus composed of embedded grids of approximation. More formally, the approximation of a function at a resolution 2^{-j} is defined as an orthogonal projection on a space $\mathbf{V}_j \subset \mathbf{L}^2(\mathbb{R})$. The space $\mathbf{V}_j$ regroups all possible approximations at the resolution 2^{-j}. The orthogonal projection of f is the function $f_j \in \mathbf{V}_j$ that minimizes $\|f - f_j\|$. The following definition introduced by Mallat [254] and Meyer [47] specifies the mathematical properties of multiresolution spaces. To avoid confusion, let us emphasize that a scale parameter 2^j is the inverse of the resolution 2^{-j}.

Definition 7.1 (MULTIRESOLUTIONS) *A sequence $\{\mathbf{V}_j\}_{j\in\mathbb{Z}}$ of closed subspaces of $\mathbf{L}^2(\mathbb{R})$ is a multiresolution approximation if the following 6 properties are satisfied:*

$$\forall (j,k) \in \mathbb{Z}^2 \ , \ f(t) \in \mathbf{V}_j \Leftrightarrow f(t - 2^j k) \in \mathbf{V}_j \ , \tag{7.1}$$

$$\forall j \in \mathbb{Z} \ , \ \mathbf{V}_{j+1} \subset \mathbf{V}_j \ , \tag{7.2}$$

$$\forall j \in \mathbb{Z} \ , \ f(t) \in \mathbf{V}_j \Leftrightarrow f\left(\frac{t}{2}\right) \in \mathbf{V}_{j+1} \ , \tag{7.3}$$

$$\lim_{j\to+\infty} \mathbf{V}_j = \bigcap_{j=-\infty}^{+\infty} \mathbf{V}_j = \{0\} \ , \tag{7.4}$$

$$\lim_{j\to-\infty} \mathbf{V}_j = \text{Closure}\left(\bigcup_{j=-\infty}^{+\infty} \mathbf{V}_j\right) = \mathbf{L}^2(\mathbb{R}) \ . \tag{7.5}$$

There exists θ such that $\{\theta(t-n)\}_{n\in\mathbb{Z}}$ is a Riesz basis of $\mathbf{V}_0$.

Let us give an intuitive explanation of these mathematical properties. Property (7.1) means that $\mathbf{V}_j$ is invariant by any translation proportional to the scale 2^j. As we shall see later, this space can be assimilated to a uniform grid with intervals 2^j, which characterizes the signal approximation at the resolution 2^{-j}. The inclusion

(7.2) is a causality property which proves that an approximation at a resolution 2^{-j} contains all the necessary information to compute an approximation at a coarser resolution 2^{-j-1}. Dilating functions in $\mathbf{V}_j$ by 2 enlarges the details by 2 and (7.3) guarantees that it defines an approximation at a coarser resolution 2^{-j-1}. When the resolution 2^{-j} goes to 0 (7.4) implies that we lose all the details of f and

$$\lim_{j \to +\infty} \|P_{\mathbf{V}_j} f\| = 0. \tag{7.6}$$

On the other hand, when the resolution 2^{-j} goes $+\infty$, property (7.5) imposes that the signal approximation converges to the original signal:

$$\lim_{j \to -\infty} \|f - P_{\mathbf{V}_j} f\| = 0. \tag{7.7}$$

When the resolution 2^{-j} increases, the decay rate of the approximation error $\|f - P_{\mathbf{V}_j} f\|$ depends on the regularity of f. Section 9.1.3 relates this error to the uniform Lipschitz regularity of f.

The existence of a Riesz basis $\{\theta(t-n)\}_{n \in \mathbb{Z}}$ of $\mathbf{V}_0$ provides a discretization theorem. The function θ can be interpreted as a unit resolution cell; Appendix A.3 gives the definition of a Riesz basis. There exist $A > 0$ and B such that any $f \in \mathbf{V}_0$ can be uniquely decomposed into

$$f(t) = \sum_{n=-\infty}^{+\infty} a[n]\, \theta(t-n) \tag{7.8}$$

with

$$A \|f\|^2 \leq \sum_{n=-\infty}^{+\infty} |a[n]|^2 \leq B \|f\|^2. \tag{7.9}$$

This energy equivalence guarantees that signal expansions over $\{\theta(t-n)\}_{n \in \mathbb{Z}}$ are numerically stable. With the dilation property (7.3) and the expansion (7.8), one can verify that the family $\{2^{-j/2}\theta(2^{-j}t-n)\}_{n \in \mathbb{Z}}$ is a Riesz basis of $\mathbf{V}_j$ with the same Riesz bounds A and B at all scales 2^j. The following proposition gives a necessary and sufficient condition for $\{\theta(t-n)\}_{n \in \mathbb{Z}}$ to be a Riesz basis.

Proposition 7.1 *A family $\{\theta(t-n)\}_{n \in \mathbb{Z}}$ is a Riesz basis of the space $\mathbf{V}_0$ it generates if and only if there exist $A > 0$ and $B > 0$ such that*

$$\forall \omega \in [-\pi, \pi] \ , \ \frac{1}{B} \leq \sum_{k=-\infty}^{+\infty} |\hat{\theta}(\omega - 2k\pi)|^2 \leq \frac{1}{A}. \tag{7.10}$$

Proof[1]. Any $f \in \mathbf{V}_0$ can be decomposed as

$$f(t) = \sum_{n=-\infty}^{+\infty} a[n]\, \theta(t-n). \tag{7.11}$$

The Fourier transform of this equation yields

$$\hat{f}(\omega) = \hat{a}(\omega)\,\hat{\theta}(\omega) \tag{7.12}$$

where $\hat{a}(\omega)$ is the Fourier series $\hat{a}(\omega) = \sum_{n=-\infty}^{+\infty} a[n] \exp(-in\omega)$. The norm of f can thus be written

$$\begin{aligned} \|f\|^2 &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} |\hat{f}(\omega)|^2\,d\omega = \frac{1}{2\pi}\int_0^{2\pi} \sum_{k=-\infty}^{+\infty} |\hat{a}(\omega+2k\pi)|^2\,|\hat{\theta}(\omega+2k\pi)|^2\,d\omega \\ &= \frac{1}{2\pi}\int_0^{2\pi} |\hat{a}(\omega)|^2 \sum_{k=-\infty}^{+\infty} |\hat{\theta}(\omega+2k\pi)|^2\,d\omega\,, \end{aligned} \tag{7.13}$$

because $a(\omega)$ is 2π periodic. The family $\{\theta(t-n)\}_{n\in\mathbf{Z}}$ is a Riesz basis if and only if

$$A\,\|f\|^2 \le \frac{1}{2\pi}\int_0^{2\pi} |\hat{a}(\omega)|^2\,d\omega = \sum_{n=-\infty}^{+\infty} |a[n]|^2 \le B\,\|f\|^2. \tag{7.14}$$

If $\hat{\theta}$ satisfies (7.10) then (7.14) is derived from (7.13). The linear independence of $\{\theta(t-n)\}_{n\in\mathbf{Z}}$ is a consequence of the fact that (7.14) is valid for any $a[n]$ satisfying (7.11). If $f = 0$ then necessarily $a[n] = 0$ for all $n \in \mathbb{Z}$. The family $\{\theta(t-n)\}_{n\in\mathbf{Z}}$ is therefore a Riesz basis of $\mathbf{V}_0$.

Conversely, if $\{\theta(t-n)\}_{n\in\mathbf{Z}}$ is a Riesz basis then (7.14) is valid for any $a[n] \in \mathbf{l}^2(\mathbb{Z})$. If either the lower bound or the upper bound of (7.10) is not satisfied for almost all $\omega \in [-\pi,\pi]$ then one can construct a non-zero 2π periodic function $\hat{a}(\omega)$ whose support corresponds to frequencies where (7.10) is not verified. We then derive from (7.13) that (7.14) is not valid for $a[n]$, which contradicts the Riesz basis hypothesis. ■

Example 7.1 Piecewise constant approximations A simple multiresolution approximation is composed of piecewise constant functions. The space $\mathbf{V}_j$ is the set of all $g \in \mathbf{L}^2(\mathbb{R})$ such that $g(t)$ is constant for $t \in [n2^j, (n+1)2^j)$ and $n \in \mathbb{Z}$. The approximation at a resolution 2^{-j} of f is the closest piecewise constant function on intervals of size 2^j. The resolution cell can be chosen to be the box window $\theta = \mathbf{1}_{[0,1)}$. Clearly $\mathbf{V}_j \subset \mathbf{V}_{j-1}$ since functions constant on intervals of size 2^j are also constant on intervals of size 2^{j-1}. The verification of the other multiresolution properties is left to the reader. It is often desirable to construct approximations that are smooth functions, in which case piecewise constant functions are not appropriate.

Example 7.2 Shannon approximations Frequency band-limited functions also yield multiresolution approximations. The space $\mathbf{V}_j$ is defined as the set of functions whose Fourier transform has a support included in $[-2^{-j}\pi, 2^{-j}\pi]$. Proposition 3.2 provides an orthonormal basis $\{\theta(t-n)\}_{n\in\mathbf{Z}}$ of $\mathbf{V}_0$ defined by

$$\theta(t) = \frac{\sin \pi t}{\pi t}. \tag{7.15}$$

All other properties of multiresolution approximation are easily verified.

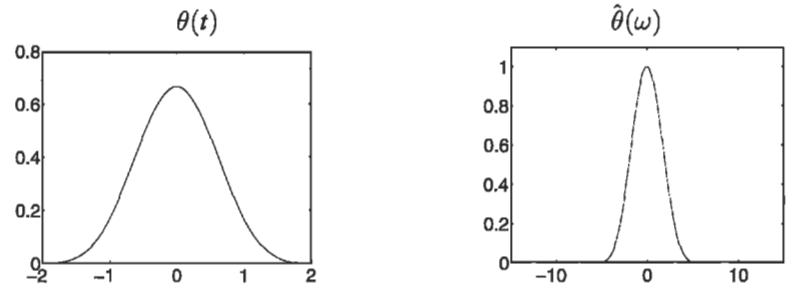


FIGURE 7.1 Cubic box spline θ and its Fourier transform $\hat{\theta}$.

The approximation at the resolution 2^{-j} of $f \in \mathbf{L}^2(\mathbb{R})$ is the function $P_{\mathbf{V}_j} f \in \mathbf{V}_j$ that minimizes $\|P_{\mathbf{V}_j} f - f\|$. It is proved in (3.12) that its Fourier transform is obtained with a frequency filtering:

$$\widehat{P_{\mathbf{V}_j} f}(\omega) = \hat{f}(\omega)\, \mathbf{1}_{[-2^{-j}\pi, 2^{-j}\pi]}(\omega).$$

This Fourier transform is generally discontinuous at $\pm 2^{-j}\pi$, in which case $|P_{\mathbf{V}_j} f(t)|$ decays like $|t|^{-1}$, for large $|t|$, even though f might have a compact support.

Example 7.3 Spline approximations Polynomial spline approximations construct smooth approximations with fast asymptotic decay. The space $\mathbf{V}_j$ of splines of degree $m \geq 0$ is the set of functions that are $m-1$ times continuously differentiable and equal to a polynomial of degree m on any interval $[n2^j, (n+1)2^j]$, for $n \in \mathbb{Z}$. When $m = 0$, it is a piecewise constant multiresolution approximation. When $m = 1$, functions in $\mathbf{V}_j$ are piecewise linear and continuous.

A Riesz basis of polynomial splines is constructed with *box splines*. A box spline θ of degree m is computed by convolving the box window $\mathbf{1}_{[0,1]}$ with itself $m+1$ times and centering at 0 or $1/2$. Its Fourier transform is

$$\hat{\theta}(\omega) = \left(\frac{\sin(\omega/2)}{\omega/2}\right)^{m+1} \exp\left(\frac{-i\epsilon\omega}{2}\right). \tag{7.16}$$

If m is even then $\epsilon = 1$ and θ has a support centered at $t = 1/2$. If m is odd then $\epsilon = 0$ and $\theta(t)$ is symmetric about $t = 0$. Figure 7.1 displays a cubic box spline $m = 3$ and its Fourier transform. For all $m \geq 0$, one can prove that $\{\theta(t-n)\}_{n\in\mathbb{Z}}$ is a Riesz basis of $\mathbf{V}_0$ by verifying the condition (7.10). This is done with a closed form expression for the series (7.24).

7.1.2 Scaling Function

The approximation of f at the resolution 2^{-j} is defined as the orthogonal projection $P_{\mathbf{V}_j} f$ on $\mathbf{V}_j$. To compute this projection, we must find an orthonormal basis of $\mathbf{V}_j$.

The following theorem orthogonalizes the Riesz basis $\{\theta(t-n)\}_{n\in\mathbb{Z}}$ and constructs an orthogonal basis of each space $\mathbf{V}_j$ by dilating and translating a single function ϕ called a *scaling function*. To avoid confusing the resolution 2^{-j} and the scale 2^j, in the rest of the chapter the notion of resolution is dropped and $P_{\mathbf{V}_j} f$ is called an approximation at the scale 2^j.

Theorem 7.1 *Let $\{\mathbf{V}_j\}_{j\in\mathbb{Z}}$ be a multiresolution approximation and ϕ be the scaling function whose Fourier transform is*

$$\hat{\phi}(\omega) = \frac{\hat{\theta}(\omega)}{\left(\sum_{k=-\infty}^{+\infty} |\hat{\theta}(\omega+2k\pi)|^2\right)^{1/2}}. \tag{7.17}$$

Let us denote

$$\phi_{j,n}(t) = \frac{1}{\sqrt{2^j}}\,\phi\left(\frac{t-n}{2^j}\right).$$

The family $\{\phi_{j,n}\}_{n\in\mathbb{Z}}$ is an orthonormal basis of $\mathbf{V}_j$ for all $j \in \mathbb{Z}$.

Proof[1]. To construct an orthonormal basis, we look for a function $\phi \in \mathbf{V}_0$. It can thus be expanded in the basis $\{\theta(t-n)\}_{n\in\mathbb{Z}}$:

$$\phi(t) = \sum_{n=-\infty}^{+\infty} a[n]\,\theta(t-n),$$

which implies that

$$\hat{\phi}(\omega) = \hat{a}(\omega)\,\hat{\theta}(\omega),$$

where $\hat{a}$ is a 2π periodic Fourier series of finite energy. To compute $\hat{a}$ we express the orthogonality of $\{\phi(t-n)\}_{n\in\mathbb{Z}}$ in the Fourier domain. Let $\bar{\phi}(t) = \phi^*(-t)$. For any $(n,p) \in \mathbb{Z}^2$,

$$\begin{aligned}\langle \phi(t-n), \phi(t-p)\rangle &= \int_{-\infty}^{+\infty} \phi(t-n)\,\phi^*(t-p)\,dt \\ &= \phi \star \bar{\phi}(p-n)\,. \end{aligned} \tag{7.18}$$

Hence $\{\phi(t-n)\}_{n\in\mathbb{Z}}$ is orthonormal if and only if $\phi \star \bar{\phi}(n) = \delta[n]$. Computing the Fourier transform of this equality yields

$$\sum_{k=-\infty}^{+\infty} |\hat{\phi}(\omega+2k\pi)|^2 = 1. \tag{7.19}$$

Indeed, the Fourier transform of $\phi \star \bar{\phi}(t)$ is $|\hat{\phi}(\omega)|^2$, and we proved in (3.3) that sampling a function periodizes its Fourier transform. The property (7.19) is verified if we choose

$$\hat{a}(\omega) = \left(\sum_{k=-\infty}^{+\infty} |\hat{\theta}(\omega+2k\pi)|^2\right)^{-1/2}.$$

Proposition 7.1 proves that the denominator has a strictly positive lower bound, so $\hat{a}$ is a 2π periodic function of finite energy. ■

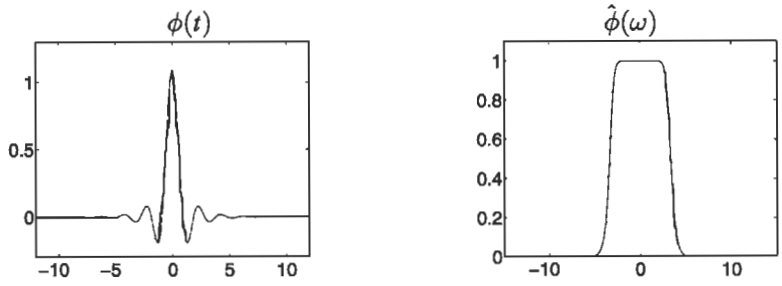


FIGURE 7.2 Cubic spline scaling function ϕ and its Fourier transform $\hat{\phi}$ computed with (7.23).

Approximation The orthogonal projection of f over $\mathbf{V}_j$ is obtained with an expansion in the scaling orthogonal basis

$$P_{\mathbf{V}_j} f = \sum_{n=-\infty}^{+\infty} \langle f, \phi_{j,n} \rangle \, \phi_{j,n}. \tag{7.20}$$

The inner products

$$a_j[n] = \langle f, \phi_{j,n} \rangle \tag{7.21}$$

provide a discrete approximation at the scale 2^j. We can rewrite them as a convolution product:

$$a_j[n] = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{2^j}} \phi\left(\frac{t - 2^j n}{2^j}\right) dt = f \star \bar{\phi}_j(2^j n), \tag{7.22}$$

with $\bar{\phi}_j(t) = \sqrt{2^{-j}} \phi(2^{-j} t)$. The energy of the Fourier transform $\hat{\phi}$ is typically concentrated in $[-\pi, \pi]$, as illustrated by Figure 7.2. As a consequence, the Fourier transform $\sqrt{2^j} \hat{\phi}^*(2^j \omega)$ of $\bar{\phi}_j(t)$ is mostly non-negligible in $[-2^{-j}\pi, 2^{-j}\pi]$. The discrete approximation $a_j[n]$ is therefore a low-pass filtering of f sampled at intervals 2^j. Figure 7.3 gives a discrete multiresolution approximation at scales $2^{-9} \leq 2^j \leq 2^{-4}$.

Example 7.4 For piecewise constant approximations and Shannon multiresolution approximations we have constructed Riesz bases $\{\theta(t - n)\}_{n \in \mathbb{Z}}$ which are orthonormal bases, hence $\phi = \theta$.

Example 7.5 Spline multiresolution approximations admit a Riesz basis constructed with a box spline θ of degree m, whose Fourier transform is given by (7.16). Inserting this expression in (7.17) yields

$$\hat{\phi}(\omega) = \frac{\exp(-i\epsilon\omega/2)}{\omega^{m+1} \sqrt{S_{2m+2}(\omega)}}, \tag{7.23}$$

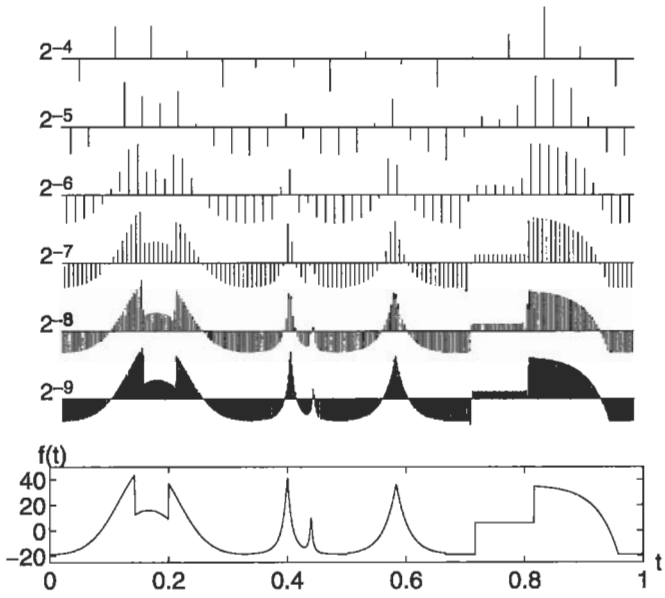


FIGURE 7.3 Discrete multiresolution approximations $a_j[n]$ at scales 2^j, computed with cubic splines.

with

$$S_n(\omega) = \sum_{k=-\infty}^{+\infty} \frac{1}{(\omega + 2k\pi)^n}, \tag{7.24}$$

and $\epsilon = 1$ if m is even or $\epsilon = 0$ if m is odd. A closed form expression of $S_{2m+2}(\omega)$ is obtained by computing the derivative of order $2m$ of the identity

$$S_2(2\omega) = \sum_{k=-\infty}^{+\infty} \frac{1}{(2\omega + 2k\pi)^2} = \frac{1}{4\sin^2\omega}.$$

For linear splines $m = 1$ and

$$S_4(2\omega) = \frac{1 + 2\cos^2\omega}{48\sin^4\omega}, \tag{7.25}$$

which yields

$$\hat{\phi}(\omega) = \frac{4\sqrt{3}\sin^2(\omega/2)}{\omega^2\sqrt{1 + 2\cos^2(\omega/2)}}. \tag{7.26}$$

The cubic spline scaling function corresponds to $m = 3$ and $\hat{\phi}(\omega)$ is calculated with (7.23) by inserting

$$S_8(2\omega) = \frac{5 + 30\cos^2\omega + 30\sin^2\omega\cos^2\omega}{1052^8\sin^8\omega} \tag{7.27}$$

$$+\frac{70\cos^4\omega+2\sin^4\omega\cos^2\omega+2/3\sin^6\omega}{105\,2^8\sin^8\omega}.$$

This cubic spline scaling function ϕ and its Fourier transform are displayed in Figure 7.2. It has an infinite support but decays exponentially.

7.1.3 Conjugate Mirror Filters

A multiresolution approximation is entirely characterized by the scaling function ϕ that generates an orthogonal basis of each space $\mathbf{V}_j$. We study the properties of ϕ which guarantee that the spaces $\mathbf{V}_j$ satisfy all conditions of a multiresolution approximation. It is proved that any scaling function is specified by a discrete filter called a *conjugate mirror filter*.

Scaling Equation The multiresolution causality property (7.2) imposes that $\mathbf{V}_j \subset \mathbf{V}_{j-1}$. In particular $2^{-1/2}\phi(t/2) \in \mathbf{V}_1 \subset \mathbf{V}_0$. Since $\{\phi(t-n)\}_{n\in\mathbb{Z}}$ is an orthonormal basis of $\mathbf{V}_0$, we can decompose

$$\frac{1}{\sqrt{2}}\phi\left(\frac{t}{2}\right) = \sum_{n=-\infty}^{+\infty} h[n]\,\phi(t-n), \tag{7.28}$$

with

$$h[n] = \left\langle \frac{1}{\sqrt{2}}\phi\left(\frac{t}{2}\right), \phi(t-n)\right\rangle. \tag{7.29}$$

This scaling equation relates a dilation of ϕ by 2 to its integer translations. The sequence $h[n]$ will be interpreted as a discrete filter.

The Fourier transform of both sides of (7.28) yields

$$\hat{\phi}(2\omega) = \frac{1}{\sqrt{2}}\hat{h}(\omega)\,\hat{\phi}(\omega) \tag{7.30}$$

for $\hat{h}(\omega) = \sum_{n=-\infty}^{+\infty} h[n]\,e^{-in\omega}$. It is thus tempting to express $\hat{\phi}(\omega)$ directly as a product of dilations of $\hat{h}(\omega)$. For any $p \geq 0$, (7.30) implies

$$\hat{\phi}(2^{-p+1}\omega) = \frac{1}{\sqrt{2}}\hat{h}(2^{-p}\omega)\,\hat{\phi}(2^{-p}\omega). \tag{7.31}$$

By substitution, we obtain

$$\hat{\phi}(\omega) = \left(\prod_{p=1}^{P}\frac{\hat{h}(2^{-p}\omega)}{\sqrt{2}}\right)\hat{\phi}(2^{-P}\omega). \tag{7.32}$$

If $\hat{\phi}(\omega)$ is continuous at $\omega = 0$ then $\lim_{P\to+\infty}\hat{\phi}(2^{-P}\omega) = \hat{\phi}(0)$ so

$$\hat{\phi}(\omega) = \prod_{p=1}^{+\infty}\frac{\hat{h}(2^{-p}\omega)}{\sqrt{2}}\,\hat{\phi}(0). \tag{7.33}$$

The following theorem [254, 47] gives necessary and then sufficient conditions on $\hat{h}(\omega)$ to guarantee that this infinite product is the Fourier transform of a scaling function.

Theorem 7.2 (MALLAT, MEYER) *Let $\phi \in \mathbf{L}^2(\mathbb{R})$ be an integrable scaling function. The Fourier series of $h[n] = \langle 2^{-1/2}\phi(t/2), \phi(t-n)\rangle$ satisfies*

$$\forall \omega \in \mathbb{R} \ , \ |\hat{h}(\omega)|^2 + |\hat{h}(\omega+\pi)|^2 = 2, \tag{7.34}$$

and

$$\hat{h}(0) = \sqrt{2}. \tag{7.35}$$

Conversely, if $\hat{h}(\omega)$ is 2π periodic and continuously differentiable in a neighborhood of $\omega = 0$, if it satisfies (7.34) and (7.35) and if

$$\inf_{\omega \in [-\pi/2, \pi/2]} |\hat{h}(\omega)| > 0 \tag{7.36}$$

then

$$\hat{\phi}(\omega) = \prod_{p=1}^{+\infty} \frac{\hat{h}(2^{-p}\omega)}{\sqrt{2}} \tag{7.37}$$

is the Fourier transform of a scaling function $\phi \in \mathbf{L}^2(\mathbb{R})$.

Proof. This theorem is a central result whose proof is long and technical. It is divided in several parts.

• *Proof[1] of the necessary condition (7.34)* The necessary condition is proved to be a consequence of the fact that $\{\phi(t-n)\}_{n\in\mathbb{Z}}$ is orthonormal. In the Fourier domain, (7.19) gives an equivalent condition:

$$\forall \omega \in \mathbb{R} \ , \ \sum_{k=-\infty}^{+\infty} |\hat{\phi}(\omega + 2k\pi)|^2 = 1. \tag{7.38}$$

Inserting $\hat{\phi}(\omega) = 2^{-1/2}\hat{h}(\omega/2)\,\hat{\phi}(\omega/2)$ yields

$$\sum_{k=-\infty}^{+\infty} |\hat{h}(\frac{\omega}{2} + k\pi)|^2 |\hat{\phi}(\frac{\omega}{2} + k\pi)|^2 = 2.$$

Since $\hat{h}(\omega)$ is 2π periodic, separating the even and odd integer terms gives

$$|\hat{h}(\frac{\omega}{2})|^2 \sum_{p=-\infty}^{+\infty} \left|\hat{\phi}\left(\frac{\omega}{2} + 2p\pi\right)\right|^2 + \left|\hat{h}\left(\frac{\omega}{2} + \pi\right)\right|^2 \sum_{p=-\infty}^{+\infty} \left|\hat{\phi}\left(\frac{\omega}{2} + \pi + 2p\pi\right)\right|^2 = 2.$$

Inserting (7.38) for $\omega' = \omega/2$ and $\omega' = \omega/2 + \pi$ proves that

$$|\hat{h}(\omega')|^2 + |\hat{h}(\omega' + \pi)|^2 = 2.$$

• *Proof[2] of the necessary condition (7.35)* We prove that $\hat{h}(0) = \sqrt{2}$ by showing that $\hat{\phi}(0) \neq 0$. Indeed we know that $\hat{\phi}(0) = 2^{-1/2}\hat{h}(0)\,\hat{\phi}(0)$. More precisely, we verify

that $|\hat{\phi}(0)| = 1$ is a consequence of the completeness property (7.5) of multiresolution approximations.

The orthogonal projection of $f \in \mathbf{L}^2(\mathbb{R})$ on $\mathbf{V}_j$ is

$$P_{\mathbf{V}_j} f = \sum_{n=-\infty}^{+\infty} \langle f, \phi_{j,n} \rangle \, \phi_{j,n}. \tag{7.39}$$

Property (7.5) expressed in the time and Fourier domains with the Plancherel formula implies that

$$\lim_{j \to -\infty} \| f - P_{\mathbf{V}_j} f \|^2 = \lim_{j \to -\infty} 2\pi \| \hat{f} - \widehat{P_{\mathbf{V}_j} f} \|^2 = 0. \tag{7.40}$$

To compute the Fourier transform $\widehat{P_{\mathbf{V}_j} f}(\omega)$, we denote $\phi_j(t) = \sqrt{2^{-j}} \phi(2^{-j} t)$. Inserting the convolution expression (7.22) in (7.39) yields

$$P_{\mathbf{V}_j} f(t) = \sum_{n=-\infty}^{+\infty} f \star \bar{\phi}_j(2^j n) \, \phi_j(t - 2^j n) = \phi_j \star \sum_{n=-\infty}^{+\infty} f \star \bar{\phi}_j(2^j n) \, \delta(t - 2^j n).$$

The Fourier transform of $f \star \bar{\phi}_j(t)$ is $\sqrt{2^j} \hat{f}(\omega) \hat{\phi}^*(2^j \omega)$. A uniform sampling has a periodized Fourier transform calculated in (3.3), and hence

$$\widehat{P_{\mathbf{V}_j} f}(\omega) = \hat{\phi}(2^j \omega) \sum_{k=-\infty}^{+\infty} \hat{f}\left(\omega - \frac{2k\pi}{2^j}\right) \hat{\phi}^*\left(2^j \left[\omega - \frac{2k\pi}{2^j}\right]\right). \tag{7.41}$$

Let us choose $\hat{f} = \mathbf{1}_{[-\pi,\pi]}$. For $j < 0$ and $\omega \in [-\pi, \pi]$, (7.41) gives $\widehat{P_{\mathbf{V}_j} f}(\omega) = |\hat{\phi}(2^j \omega)|^2$. The mean-square convergence (7.40) implies that

$$\lim_{j \to -\infty} \int_{-\pi}^{\pi} \left| 1 - |\hat{\phi}(2^j \omega)|^2 \right|^2 d\omega = 0 \, .$$

Since ϕ is integrable, $\hat{\phi}(\omega)$ is continuous and hence $\lim_{j \to -\infty} |\hat{\phi}(2^j \omega)| = |\hat{\phi}(0)| = 1$. We now prove that the function ϕ whose Fourier transform is given by (7.37) is a scaling function. This is divided in two intermediate results.

• *Proof*[3] *that* $\{\phi(t-n)\}_{n \in \mathbb{Z}}$ *is orthonormal.* Observe first that the infinite product (7.37) converges and that $|\hat{\phi}(\omega)| \leq 1$ because (7.34) implies that $|\hat{h}(\omega)| \leq \sqrt{2}$. The Parseval formula gives

$$\langle \phi(t), \phi(t-n) \rangle = \int_{-\infty}^{+\infty} \phi(t) \, \phi^*(t-n) \, dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\hat{\phi}(\omega)|^2 \, \mathrm{e}^{in\omega} \, d\omega.$$

Verifying that $\{\phi(t-n)\}_{n \in \mathbb{Z}}$ is orthonormal is thus equivalent to showing that

$$\int_{-\infty}^{+\infty} |\hat{\phi}(\omega)|^2 \, \mathrm{e}^{in\omega} \, d\omega = 2\pi \, \delta[n].$$

This result is obtained by considering the functions

$$\hat{\phi}_k(\omega) = \prod_{p=1}^{k} \frac{\hat{h}(2^{-p}\omega)}{\sqrt{2}} \, \mathbf{1}_{[-2^k \pi, 2^k \pi]}(\omega).$$

and computing the limit, as k increases to $+\infty$, of the integrals

$$I_k[n] = \int_{-\infty}^{+\infty} |\hat{\phi}_k(\omega)|^2 \, e^{in\omega} \, d\omega = \int_{-2^k\pi}^{2^k\pi} \prod_{p=1}^{k} \frac{|\hat{h}(2^{-p}\omega)|^2}{2} \, e^{in\omega} \, d\omega.$$

First, let us show that $I_k[n] = 2\pi\delta[n]$ for all $k \geq 1$. To do this, we divide $I_k[n]$ into two integrals:

$$I_k[n] = \int_{-2^k\pi}^{0} \prod_{p=1}^{k} \frac{|\hat{h}(2^{-p}\omega)|^2}{2} \, e^{in\omega} \, d\omega + \int_{0}^{2^k\pi} \prod_{p=1}^{k} \frac{|\hat{h}(2^{-p}\omega)|^2}{2} \, e^{in\omega} \, d\omega.$$

Let us make the change of variable $\omega' = \omega + 2^k\pi$ in the first integral. Since $\hat{h}(\omega)$ is 2π periodic, when $p < k$ then $|\hat{h}(2^{-p}[\omega' - 2^k\pi])|^2 = |\hat{h}(2^{-p}\omega')|^2$. When $k = p$ the hypothesis (7.34) implies that

$$|\hat{h}(2^{-k}[\omega' - 2^k\pi])|^2 + |\hat{h}(2^{-k}\omega')|^2 = 2.$$

For $k > 1$, the two integrals of $I_k[n]$ become

$$I_k[n] = \int_{0}^{2^k\pi} \prod_{p=1}^{k-1} \frac{|\hat{h}(2^{-p}\omega)|^2}{2} \, e^{in\omega} \, d\omega. \tag{7.42}$$

Since $\prod_{p=1}^{k-1} |\hat{h}(2^{-p}\omega)|^2 \, e^{in\omega}$ is $2^k\pi$ periodic we obtain $I_k[n] = I_{k-1}[n]$, and by induction $I_k[n] = I_1[n]$. Writing (7.42) for $k = 1$ gives

$$I_1[n] = \int_{0}^{2\pi} e^{in\omega} \, d\omega = 2\pi \, \delta[n],$$

which verifies that $I_k[n] = 2\pi\delta[n]$, for all $k \geq 1$.

We shall now prove that $\hat{\phi} \in \mathbf{L}^2(\mathbb{R})$. For all $\omega \in \mathbb{R}$

$$\lim_{k\to\infty} |\hat{\phi}_k(\omega)|^2 = \prod_{p=1}^{\infty} \frac{|\hat{h}(2^{-p}\omega)|^2}{2} = |\hat{\phi}(\omega)|^2.$$

The Fatou Lemma A.1 on positive functions proves that

$$\int_{-\infty}^{+\infty} |\hat{\phi}(\omega)|^2 \, d\omega \leq \lim_{k\to\infty} \int_{-\infty}^{+\infty} |\hat{\phi}_k(\omega)|^2 \, d\omega = 2\pi, \tag{7.43}$$

because $I_k[0] = 2\pi$ for all $k \geq 1$. Since

$$|\hat{\phi}(\omega)|^2 \, e^{in\omega} = \lim_{k\to\infty} |\hat{\phi}_k(\omega)|^2 \, e^{in\omega},$$

we finally verify that

$$\int_{-\infty}^{+\infty} |\hat{\phi}(\omega)|^2 \, e^{in\omega} \, d\omega = \lim_{k\to\infty} \int_{-\infty}^{+\infty} |\hat{\phi}_k(\omega)|^2 \, e^{in\omega} \, d\omega = 2\pi \, \delta[n] \tag{7.44}$$

by applying the dominated convergence Theorem A.1. This requires verifying the upper-bound condition (A.1). This is done in our case by proving the existence of a constant C such that

$$\left| |\hat{\phi}_k(\omega)|^2 \, e^{in\omega} \right| = |\hat{\phi}_k(\omega)|^2 \leq C \, |\hat{\phi}(\omega)|^2. \tag{7.45}$$

Indeed, we showed in (7.43) that $|\hat{\phi}(\omega)|^2$ is an integrable function.

The existence of $C > 0$ satisfying (7.45) is trivial for $|\omega| > 2^k\pi$ since $\hat{\phi}_k(\omega) = 0$. For $|\omega| \leq 2^k\pi$ since $\hat{\phi}(\omega) = 2^{-1/2}\,\hat{h}(\omega/2)\,\hat{\phi}(\omega/2)$, it follows that

$$|\hat{\phi}(\omega)|^2 = |\hat{\phi}_k(\omega)|^2\,|\hat{\phi}(2^{-k}\omega)|^2.$$

To prove (7.45) for $|\omega| \leq 2^k\pi$, it is therefore sufficient to show that $|\hat{\phi}(\omega)|^2 \geq 1/C$ for $\omega \in [-\pi, \pi]$.

Let us first study the neighborhood of $\omega = 0$. Since $\hat{h}(\omega)$ is continuously differentiable in this neighborhood and since $|\hat{h}(\omega)|^2 \leq 2 = |\hat{h}(0)|^2$, the functions $|\hat{h}(\omega)|^2$ and $\log_e |\hat{h}(\omega)|^2$ have derivatives that vanish at $\omega = 0$. It follows that there exists $\epsilon > 0$ such that

$$\forall |\omega| \leq \epsilon \ , \ 0 \geq \log_e \left(\frac{|\hat{h}(\omega)|^2}{2} \right) \geq -|\omega|.$$

Hence, for $|\omega| \leq \epsilon$

$$|\hat{\phi}(\omega)|^2 = \exp\left[\sum_{p=1}^{+\infty} \log_e \left(\frac{|\hat{h}(2^{-p}\omega)|^2}{2} \right) \right] \geq e^{-|\omega|} \geq e^{-\epsilon}. \tag{7.46}$$

Now let us analyze the domain $|\omega| > \epsilon$. To do this we take an integer l such that $2^{-l}\pi < \epsilon$. Condition (7.36) proves that $K = \inf_{\omega \in [-\pi/2, \pi/2]} |\hat{h}(\omega)| > 0$ so if $|\omega| \leq \pi$

$$|\hat{\phi}(\omega)|^2 = \prod_{p=1}^{l} \frac{|\hat{h}(2^{-p}\omega)|^2}{2} \left| \hat{\phi}\left(2^{-l}\omega\right) \right|^2 \geq \frac{K^{2l}}{2^l}\, e^{-\epsilon} = \frac{1}{C}.$$

This last result finishes the proof of inequality (7.45). Applying the dominated convergence Theorem A.1 proves (7.44) and hence that $\{\phi(t-n)\}_{n \in \mathbb{Z}}$ is orthonormal. A simple change of variable shows that $\{\phi_{j,n}\}_{j \in \mathbb{Z}}$ is orthonormal for all $j \in \mathbb{Z}$.

• *Proof*[3] *that* $\{\mathbf{V}_j\}_{j \in \mathbb{Z}}$ *is a multiresolution.* To verify that ϕ is a scaling function, we must show that the spaces $\mathbf{V}_j$ generated by $\{\phi_{j,n}\}_{j \in \mathbb{Z}}$ define a multiresolution approximation. The multiresolution properties (7.1) and (7.3) are clearly true. The causality $\mathbf{V}_{j+1} \subset \mathbf{V}_j$ is verified by showing that for any $p \in \mathbb{Z}$,

$$\phi_{j+1,p} = \sum_{n=-\infty}^{+\infty} h[n-2p]\,\phi_{j,n}.$$

This equality is proved later in (7.112). Since all vectors of a basis of $\mathbf{V}_{j+1}$ can decomposed in a basis of $\mathbf{V}_j$ it follows that $\mathbf{V}_{j+1} \subset \mathbf{V}_j$.

To prove the multiresolution property (7.4) we must show that any $f \in \mathbf{L}^2(\mathbb{R})$ satisfies

$$\lim_{j \to +\infty} \|P_{\mathbf{V}_j} f\| = 0. \tag{7.47}$$

Since $\{\phi_{j,n}\}_{n \in \mathbb{Z}}$ is an orthonormal basis of $\mathbf{V}_j$

$$\|P_{\mathbf{V}_j} f\|^2 = \sum_{n=-\infty}^{+\infty} |\langle f, \phi_{j,n} \rangle|^2.$$

Suppose first that f is bounded by A and has a compact support included in $[2^J, 2^J]$. The constants A and J may be arbitrarily large. It follows that

$$\begin{aligned}\sum_{n=-\infty}^{+\infty} |\langle f, \phi_{j,n}\rangle|^2 &\leq 2^{-j}\left[\sum_{n=-\infty}^{+\infty}\int_{-2^J}^{2^J} |f(t)|\,|\phi(2^{-j}t-n)|\,dt\right]^2 \\ &\leq 2^{-j}A^2\left[\sum_{n=-\infty}^{+\infty}\int_{-2^J}^{2^J} |\phi(2^{-j}t-n)|\,dt\right]^2\end{aligned}$$

Applying the Cauchy-Schwarz inequality to $1 \times |\phi(2^{-j}t-n)|$ yields

$$\begin{aligned}\sum_{n=-\infty}^{+\infty} |\langle f, \phi_{j,n}\rangle|^2 &\leq A^2 2^{J+1}\sum_{n=-\infty}^{+\infty}\int_{-2^J}^{2^J} |\phi(2^{-j}t-n)|^2 2^{-j}\,dt \\ &\leq A^2 2^{J+1}\int_{S_j} |\phi(t)|^2\,dt = A^2 2^{J+1}\int_{-\infty}^{+\infty} |\phi(t)|^2\,\mathbf{1}_{S_j}(t)\,dt,\end{aligned}$$

with $S_j = \cup_{n\in\mathbf{Z}}[n-2^{J-j}, n+2^{J-j}]$ for $j > J$. For $t \notin \mathbb{Z}$ we obviously have $\mathbf{1}_{S_j}(t) \to 0$ for $j \to +\infty$. The dominated convergence Theorem A.1 applied to $|\phi(t)|^2\,\mathbf{1}_{S_j}(t)$ proves that the integral converges to 0 and hence

$$\lim_{j\to+\infty}\sum_{n=-\infty}^{+\infty} |\langle f, \phi_{j,n}\rangle|^2 = 0.$$

Property (7.47) is extended to any $f \in \mathbf{L}^2(\mathbb{R})$ by using the density in $\mathbf{L}^2(\mathbb{R})$ of bounded function with a compact support, and Proposition A.3.

To prove the last multiresolution property (7.5) we must show that for any $f \in \mathbf{L}^2(\mathbb{R})$,

$$\lim_{j\to-\infty} \|f - P_{\mathbf{V}_j} f\|^2 = \lim_{j\to-\infty}\left(\|f\|^2 - \|P_{\mathbf{V}_j} f\|^2\right) = 0. \tag{7.48}$$

We consider functions f whose Fourier transform $\hat{f}$ has a compact support included in $[-2^J\pi, 2^J\pi]$ for J large enough. We proved in (7.41) that the Fourier transform of $P_{\mathbf{V}_j} f$ is

$$\widehat{P_{\mathbf{V}_j} f}(\omega) = \hat{\phi}(2^j\omega)\sum_{k=-\infty}^{+\infty} \hat{f}\left(\omega - 2^{-j}2k\pi\right)\hat{\phi}^*\left(2^j\left[\omega - 2^{-j}2k\pi\right]\right).$$

If $j < -J$, then the supports of $\hat{f}(\omega - 2^{-j}2k\pi)$ are disjoint for different k so

$$\begin{aligned}\|P_{\mathbf{V}_j} f\|^2 &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} |\hat{f}(\omega)|^2\,|\hat{\phi}(2^j\omega)|^4\,d\omega \\ &\quad + \frac{1}{2\pi}\int_{-\infty}^{+\infty}\sum_{\substack{k=-\infty \\ k\neq 0}}^{+\infty} |\hat{f}\left(\omega - 2^{-j}2k\pi\right)|^2\,|\hat{\phi}(2^j\omega)|^2\,|\hat{\phi}\left(2^j\left[\omega - 2^{-j}2k\pi\right]\right)|^2\,d\omega.\end{aligned} \tag{7.49}$$

We have already observed that $|\phi(\omega)| \leq 1$ and (7.46) proves that for ω sufficiently small $|\phi(\omega)| \geq e^{-|\omega|}$ so

$$\lim_{\omega\to 0} |\hat{\phi}(\omega)| = 1.$$

Since $|\hat{f}(\omega)|^2|\hat{\phi}(2^j\omega)|^4 \leq |\hat{f}(\omega)|^2$ and $\lim_{j\to-\infty}|\hat{\phi}(2^j\omega)|^4|\hat{f}(\omega)|^2 = |\hat{f}(\omega)|^2$ one can apply the dominated convergence Theorem A.1, to prove that

$$\lim_{j\to-\infty}\int_{-\infty}^{+\infty}|\hat{f}(\omega)|^2\,|\hat{\phi}(2^j\omega)|^4\,d\omega = \int_{-\infty}^{+\infty}|\hat{f}(\omega)|^2\,d\omega = \|f\|^2. \tag{7.50}$$

The operator P_{V_j} is an orthogonal projector, so $\|P_{V_j}f\| \leq \|f\|$. With (7.49) and (7.50), this implies that $\lim_{j\to-\infty}(\|f\|^2 - \|P_{V_j}f\|^2) = 0$, and hence verifies (7.48). This property is extended to any $f \in \mathbf{L}^2(\mathbb{R})$ by using the density in $\mathbf{L}^2(\mathbb{R})$ of functions whose Fourier transforms have a compact support and the result of Proposition A.3. ■

Discrete filters whose transfer functions satisfy (7.34) are called *conjugate mirror filters*. As we shall see in Section 7.3, they play an important role in discrete signal processing; they make it possible to decompose discrete signals in separate frequency bands with filter banks. One difficulty of the proof is showing that the infinite cascade of convolutions that is represented in the Fourier domain by the product (7.37) does converge to a decent function in $\mathbf{L}^2(\mathbb{R})$. The sufficient condition (7.36) is not necessary to construct a scaling function, but it is always satisfied in practical designs of conjugate mirror filters. It cannot just be removed as shown by the example $\hat{h}(\omega) = \cos(3\omega/2)$, which satisfies all other conditions. In this case, a simple calculation shows that $\phi = 1/3\,\mathbf{1}_{[-3/2,3/2]}$. Clearly $\{\phi(t-n)\}_{n\in\mathbb{Z}}$ is not orthogonal so ϕ is not a scaling function. The condition (7.36) may however be replaced by a weaker but more technical necessary and sufficient condition proved by Cohen [17, 128].

Example 7.6 For a Shannon multiresolution approximation, $\hat{\phi} = \mathbf{1}_{[-\pi,\pi]}$. We thus derive from (7.37) that

$$\forall\omega \in [-\pi,\pi]\ ,\ \hat{h}(\omega) = \sqrt{2}\,\mathbf{1}_{[-\pi/2,\pi/2]}(\omega).$$

Example 7.7 For piecewise constant approximations, $\phi = \mathbf{1}_{[0,1]}$. Since $h[n] = \langle 2^{-1/2}\phi(t/2), \phi(t-n)\rangle$ it follows that

$$h[n] = \begin{cases} 2^{-1/2} & \text{if } n = 0,1 \\ 0 & \text{otherwise} \end{cases} \tag{7.51}$$

Example 7.8 Polynomial splines of degree m correspond to a conjugate mirror filter $\hat{h}(\omega)$ that is calculated from $\hat{\phi}(\omega)$ with (7.30):

$$\hat{h}(\omega) = \sqrt{2}\,\frac{\hat{\phi}(2\omega)}{\hat{\phi}(\omega)}. \tag{7.52}$$

Inserting (7.23) yields

$$\hat{h}(\omega) = \exp\left(\frac{-i\epsilon\omega}{2}\right)\sqrt{\frac{S_{2m+2}(\omega)}{2^{2m+1}\,S_{2m+2}(2\omega)}}, \tag{7.53}$$

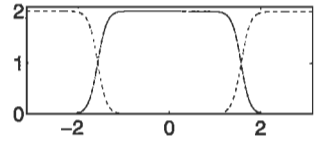


FIGURE 7.4 The solid line gives $|\hat{h}(\omega)|^2$ on $[-\pi, \pi]$, for a cubic spline multiresolution. The dotted line corresponds to $|\hat{g}(\omega)|^2$.

	n	$h[n]$		n	$h[n]$
$m=1$	0	0.817645956	$m=3$	5, −5	0.042068328
	1, −1	0.397296430		6, −6	−0.017176331
	2, −2	−0.069101020		7, −7	−0.017982291
	3, −3	−0.051945337		8, −8	0.008685294
	4, −4	0.016974805		9, −9	0.008201477
	5, −5	0.009990599		10, −10	−0.004353840
	6, −6	−0.003883261		11, −11	−0.003882426
	7, −7	−0.002201945		12, −12	0.002186714
	8, −8	0.000923371		13, −13	0.001882120
	9, −9	0.000511636		14, −14	−0.001103748
	10, −10	−0.000224296		15, −15	−0.000927187
	11, −11	−0.000122686		16, −16	0.000559952
$m=3$	0	0.766130398		17, −17	0.000462093
	1, −1	0.433923147		18, −18	−0.000285414
	2, −2	−0.050201753		19, −19	−0.000232304
	3, −3	−0.110036987		20, −20	0.000146098
	4, −4	0.032080869			

Table 7.1 Conjugate mirror filters $h[n]$ corresponding to linear splines $m = 1$ and cubic splines $m = 3$. The coefficients below 10^{-4} are not given.

where $\epsilon = 0$ if m is odd and $\epsilon = 1$ if m is even. For linear splines $m = 1$ so (7.25) implies that

$$\hat{h}(\omega) = \sqrt{2}\left[\frac{1+2\cos^2(\omega/2)}{1+2\cos^2\omega}\right]^{1/2}\cos^2\left(\frac{\omega}{2}\right). \tag{7.54}$$

For cubic splines, the conjugate mirror filter is calculated by inserting (7.27) in (7.53). Figure 7.4 gives the graph of $|\hat{h}(\omega)|^2$. The impulse responses $h[n]$ of these filters have an infinite support but an exponential decay. For m odd, $h[n]$ is symmetric about $n = 0$. Table 7.1 gives the coefficients $h[n]$ above 10^{-4} for $m = 1, 3$.

7.1.4 In Which Orthogonal Wavelets Finally Arrive

Orthonormal wavelets carry the details necessary to increase the resolution of a signal approximation. The approximations of f at the scales 2^j and 2^{j-1} are respectively equal to their orthogonal projections on $\mathbf{V}_j$ and $\mathbf{V}_{j-1}$. We know that $\mathbf{V}_j$ is included in $\mathbf{V}_{j-1}$. Let $\mathbf{W}_j$ be the orthogonal complement of $\mathbf{V}_j$ in $\mathbf{V}_{j-1}$:

$$\mathbf{V}_{j-1} = \mathbf{V}_j \oplus \mathbf{W}_j. \tag{7.55}$$

The orthogonal projection of f on $\mathbf{V}_{j-1}$ can be decomposed as the sum of orthogonal projections on $\mathbf{V}_j$ and $\mathbf{W}_j$:

$$P_{\mathbf{V}_{j-1}} f = P_{\mathbf{V}_j} f + P_{\mathbf{W}_j} f. \tag{7.56}$$

The complement $P_{\mathbf{W}_j} f$ provides the "details" of f that appear at the scale 2^{j-1} but which disappear at the coarser scale 2^j. The following theorem [47, 254] proves that one can construct an orthonormal basis of $\mathbf{W}_j$ by scaling and translating a wavelet ψ.

Theorem 7.3 (MALLAT, MEYER) *Let ϕ be a scaling function and h the corresponding conjugate mirror filter. Let ψ be the function whose Fourier transform is*

$$\hat{\psi}(\omega) = \frac{1}{\sqrt{2}} \hat{g}\left(\frac{\omega}{2}\right) \hat{\phi}\left(\frac{\omega}{2}\right), \tag{7.57}$$

with

$$\hat{g}(\omega) = \mathrm{e}^{-i\omega} \hat{h}^*(\omega + \pi). \tag{7.58}$$

Let us denote

$$\psi_{j,n}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t - 2^j n}{2^j}\right).$$

For any scale 2^j, $\{\psi_{j,n}\}_{n \in \mathbb{Z}}$ is an orthonormal basis of $\mathbf{W}_j$. For all scales, $\{\psi_{j,n}\}_{(j,n) \in \mathbb{Z}^2}$ is an orthonormal basis of $\mathbf{L}^2(\mathbb{R})$.

Proof[1]. Let us prove first that $\hat{\psi}$ can be written as the product (7.57). Necessarily $\psi(t/2) \in \mathbf{W}_1 \subset \mathbf{V}_0$. It can thus be decomposed in $\{\phi(t-n)\}_{n \in \mathbb{Z}}$ which is an orthogonal basis of $\mathbf{V}_0$:

$$\frac{1}{\sqrt{2}} \psi\left(\frac{t}{2}\right) = \sum_{n=-\infty}^{+\infty} g[n]\, \phi(t-n), \tag{7.59}$$

with

$$g[n] = \frac{1}{\sqrt{2}} \left\langle \psi\left(\frac{t}{2}\right), \phi(t-n) \right\rangle. \tag{7.60}$$

The Fourier transform of (7.59) yields

$$\hat{\psi}(2\omega) = \frac{1}{\sqrt{2}} \hat{g}(\omega)\, \hat{\phi}(\omega). \tag{7.61}$$

The following lemma gives necessary and sufficient conditions on $\hat{g}$ for designing an orthogonal wavelet.

Lemma 7.1 *The family* $\{\psi_{j,n}\}_{n\in\mathbb{Z}}$ *is an orthonormal basis of* $\mathbf{W}_j$ *if and only if*

$$|\hat{g}(\omega)|^2 + |\hat{g}(\omega+\pi)|^2 = 2 \tag{7.62}$$

and

$$\hat{g}(\omega)\,\hat{h}^*(\omega) + \hat{g}(\omega+\pi)\,\hat{h}^*(\omega+\pi) = 0. \tag{7.63}$$

The lemma is proved for $j = 0$ from which it is easily extended to $j \neq 0$ with an appropriate scaling. As in (7.19) one can verify that $\{\psi(t-n)\}_{n\in\mathbb{Z}}$ is orthonormal if and only if

$$\forall \omega \in \mathbb{R}\ ,\ I(\omega) = \sum_{k=-\infty}^{+\infty} |\hat{\psi}(\omega+2k\pi)|^2 = 1. \tag{7.64}$$

Since $\hat{\psi}(\omega) = 2^{-1/2}\,\hat{g}(\omega/2)\,\hat{\phi}(\omega/2)$ and $\hat{g}(\omega)$ is 2π periodic,

$$\begin{aligned} I(\omega) &= \sum_{k=-\infty}^{+\infty} |\hat{g}\left(\frac{\omega}{2}+k\pi\right)|^2\,|\hat{\phi}\left(\frac{\omega}{2}+k\pi\right)|^2 \\ &= |\hat{g}\Big(\frac{\omega}{2}\Big)|^2 \sum_{p=-\infty}^{+\infty} |\hat{\phi}\Big(\frac{\omega}{2}+2p\pi\Big)|^2 + |\hat{g}\Big(\frac{\omega}{2}+\pi\Big)|^2 \sum_{p=-\infty}^{+\infty} |\hat{\phi}\Big(\frac{\omega}{2}+\pi+2p\pi\Big)|^2. \end{aligned}$$

We know that $\sum_{p=-\infty}^{+\infty} |\hat{\phi}(\omega+2p\pi)|^2 = 1$ so (7.64) is equivalent to (7.62).

The space $\mathbf{W}_0$ is orthogonal to $\mathbf{V}_0$ if and only if $\{\phi(t-n)\}_{n\in\mathbb{Z}}$ and $\{\psi(t-n)\}_{n\in\mathbb{Z}}$ are orthogonal families of vectors. This means that for any $n \in \mathbb{Z}$

$$\langle \psi(t), \phi(t-n)\rangle = \psi \star \bar{\phi}(n) = 0.$$

The Fourier transform of $\psi \star \bar{\phi}(t)$ is $\hat{\psi}(\omega)\hat{\phi}^*(\omega)$. The sampled sequence $\psi \star \bar{\phi}(n)$ is zero if its Fourier series computed with (3.3) satisfies

$$\forall \omega \in \mathbb{R}\ ,\ \sum_{k=-\infty}^{+\infty} \hat{\psi}(\omega+2k\pi)\,\hat{\phi}^*(\omega+2k\pi) = 0. \tag{7.65}$$

By inserting $\hat{\psi}(\omega) = 2^{-1/2}\,\hat{g}(\omega/2)\,\hat{\phi}(\omega/2)$ and $\hat{\phi}(\omega) = 2^{-1/2}\,\hat{h}(\omega/2)\,\hat{\phi}(\omega/2)$ in this equation, since $\sum_{k=-\infty}^{+\infty} |\hat{\phi}(\omega+2k\pi)|^2 = 1$ we prove as before that (7.65) is equivalent to (7.63).

We must finally verify that $\mathbf{V}_{-1} = \mathbf{V}_0 \oplus \mathbf{W}_0$. Knowing that $\{\sqrt{2}\phi(2t-n)\}_{n\in\mathbb{Z}}$ is an orthogonal basis of $\mathbf{V}_{-1}$, it is equivalent to show that for any $a[n] \in \mathbf{l}^2(\mathbb{Z})$ there exist $b[n] \in \mathbf{l}^2(\mathbb{Z})$ and $c[n] \in \mathbf{l}^2(\mathbb{Z})$ such that

$$\sum_{n=-\infty}^{+\infty} a[n]\sqrt{2}\,\phi(2[t-2^{-1}n]) = \sum_{n=-\infty}^{+\infty} b[n]\,\phi(t-n) + \sum_{n=-\infty}^{+\infty} c[n]\,\psi(t-n). \tag{7.66}$$

This is done by relating $\hat{b}(\omega)$ and $\hat{c}(\omega)$ to $\hat{a}(\omega)$. The Fourier transform of (7.66) yields

$$\frac{1}{\sqrt{2}}\,\hat{a}\left(\frac{\omega}{2}\right)\hat{\phi}\left(\frac{\omega}{2}\right) = \hat{b}(\omega)\,\hat{\phi}(\omega) + \hat{c}(\omega)\,\hat{\psi}(\omega).$$

Inserting $\hat{\psi}(\omega) = 2^{-1/2}\,\hat{g}(\omega/2)\,\hat{\phi}(\omega/2)$ and $\hat{\phi}(\omega) = 2^{-1/2}\,\hat{h}(\omega/2)\,\hat{\phi}(\omega/2)$ in this equation shows that it is necessarily satisfied if

$$\hat{a}\left(\frac{\omega}{2}\right) = \hat{b}(\omega)\,\hat{h}\left(\frac{\omega}{2}\right) + \hat{c}(\omega)\,\hat{g}\left(\frac{\omega}{2}\right). \tag{7.67}$$

Let us define

$$\hat{b}(2\omega) = \frac{1}{2}\left[\hat{a}(\omega)\,\hat{h}^*(\omega) + \hat{a}(\omega+\pi)\,\hat{h}^*(\omega+\pi)\right]$$

and

$$\hat{c}(2\omega) = \frac{1}{2}\left[\hat{a}(\omega)\,\hat{g}^*(\omega) + \hat{a}(\omega+\pi)\,\hat{g}^*(\omega+\pi)\right].$$

When calculating the right-hand side of (7.67) we verify that it is equal to the left-hand side by inserting (7.62), (7.63) and using

$$|\hat{h}(\omega)|^2 + |\hat{h}(\omega+\pi)|^2 = 2. \tag{7.68}$$

Since $\hat{b}(\omega)$ and $\hat{c}(\omega)$ are 2π periodic they are the Fourier series of two sequences $b[n]$ and $c[n]$ that satisfy (7.66). This finishes the proof of the lemma.

The formula (7.58)

$$\hat{g}(\omega) = \mathrm{e}^{-i\omega}\,\hat{h}^*(\omega+\pi)$$

satisfies (7.62) and (7.63) because of (7.68). We thus derive from Lemma 7.1 that $\{\psi_{j,n}\}_{(j,n)\in\mathbb{Z}^2}$ is an orthogonal basis of $\mathbf{W}_j$.

We complete the proof of the theorem by verifying that $\{\psi_{j,n}\}_{(j,n)\in\mathbb{Z}^2}$ is an orthogonal basis of $\mathbf{L}^2(\mathbb{R})$. Observe first that the detail spaces $\{\mathbf{W}_j\}_{j\in\mathbb{Z}}$ are orthogonal. Indeed $\mathbf{W}_j$ is orthogonal to $\mathbf{V}_j$ and $\mathbf{W}_l \subset \mathbf{V}_{l-1} \subset \mathbf{V}_j$ for $j < l$. Hence $\mathbf{W}_j$ and $\mathbf{W}_l$ are orthogonal. We can also decompose

$$\mathbf{L}^2(\mathbb{R}) = \oplus_{j=-\infty}^{+\infty} \mathbf{W}_j. \tag{7.69}$$

Indeed $\mathbf{V}_{j-1} = \mathbf{W}_j \oplus \mathbf{V}_j$ and we verify by substitution that for any $L > J$

$$\mathbf{V}_L = \oplus_{j=L-1}^{J} \mathbf{W}_j \oplus \mathbf{V}_J. \tag{7.70}$$

Since $\{\mathbf{V}_j\}_{j\in\mathbb{Z}}$ is a multiresolution approximation, $\mathbf{V}_L$ and $\mathbf{V}_J$ tend respectively to $\mathbf{L}^2(\mathbb{R})$ and $\{0\}$ when L and J go respectively to $-\infty$ and $+\infty$, which implies (7.69). A union of orthonormal bases of all $\mathbf{W}_j$ is therefore an orthonormal basis of $\mathbf{L}^2(\mathbb{R})$. ■

The proof of the theorem shows that $\hat{g}$ is the Fourier series of

$$g[n] = \left\langle \frac{1}{\sqrt{2}}\psi\left(\frac{t}{2}\right), \phi(t-n) \right\rangle, \tag{7.71}$$

which are the decomposition coefficients of

$$\frac{1}{\sqrt{2}}\psi\left(\frac{t}{2}\right) = \sum_{n=-\infty}^{+\infty} g[n]\,\phi(t-n). \tag{7.72}$$

Calculating the inverse Fourier transform of (7.58) yields

$$g[n] = (-1)^{1-n}\,h[1-n]. \tag{7.73}$$

This mirror filter plays an important role in the fast wavelet transform algorithm.

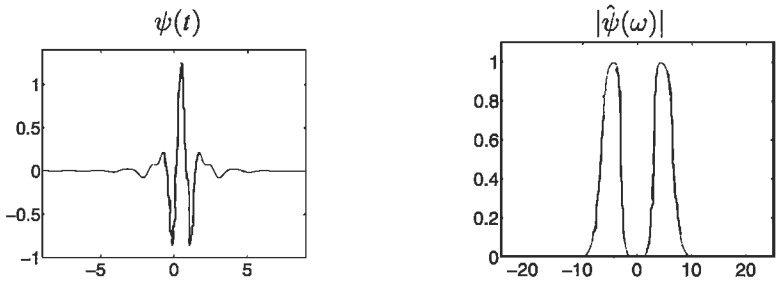


FIGURE 7.5 Battle-Lemarié cubic spline wavelet ψ and its Fourier transform modulus.

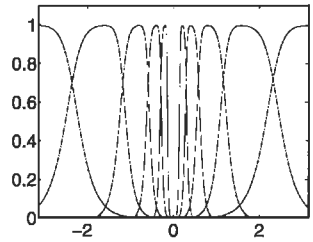


FIGURE 7.6 Graphs of $|\hat{\psi}(2^j\omega)|^2$ for the cubic spline Battle-Lemarié wavelet, with $1 \leq j \leq 5$ and $\omega \in [-\pi, \pi]$.

Example 7.9 Figure 7.5 displays the cubic spline wavelet ψ and its Fourier transform $\hat{\psi}$ calculated by inserting in (7.57) the expressions (7.23) and (7.53) of $\hat{\phi}(\omega)$ and $\hat{h}(\omega)$. The properties of this Battle-Lemarié spline wavelet are further studied in Section 7.2.2. Like most orthogonal wavelets, the energy of $\hat{\psi}$ is essentially concentrated in $[-2\pi, -\pi] \cup [\pi, 2\pi]$. For any ψ that generates an orthogonal basis of $\mathbf{L}^2(\mathbb{R})$, one can verify that

$$\forall \omega \in \mathbb{R} - \{0\} \ , \ \sum_{j=-\infty}^{+\infty} |\hat{\psi}(2^j\omega)|^2 = 1.$$

This is illustrated in Figure 7.6.

The orthogonal projection of a signal f in a "detail" space $\mathbf{W}_j$ is obtained with a partial expansion in its wavelet basis

$$P_{\mathbf{W}_j} f = \sum_{n=-\infty}^{+\infty} \langle f, \psi_{j,n} \rangle \, \psi_{j,n}.$$

A signal expansion in a wavelet orthogonal basis can thus be viewed as an aggregation of details at all scales 2^j that go from 0 to $+\infty$

$$f = \sum_{j=-\infty}^{+\infty} P_{W_j} f = \sum_{j=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \langle f, \psi_{j,n} \rangle \psi_{j,n}.$$

Figure 7.7 gives the coefficients of a signal decomposed in the cubic spline wavelet orthogonal basis. The calculations are performed with the fast wavelet transform algorithm of Section 7.3.

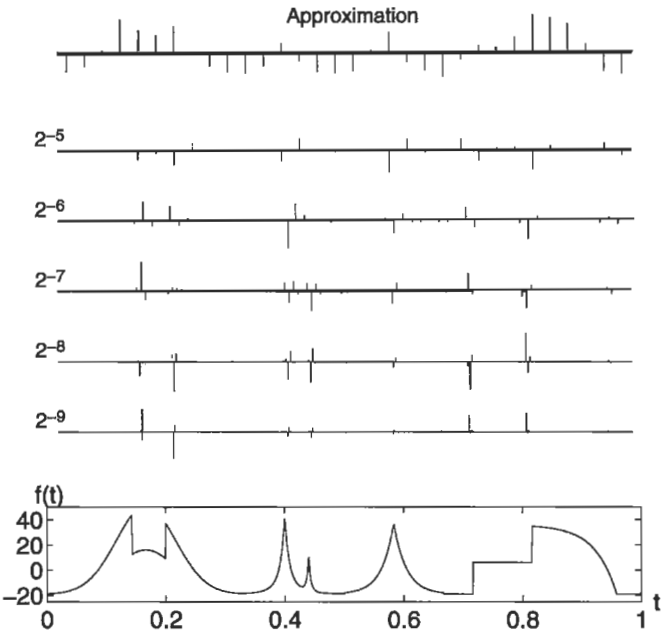


FIGURE 7.7 Wavelet coefficients $d_j[n] = \langle f, \psi_{j,n} \rangle$ calculated at scales 2^j with the cubic spline wavelet. At the top is the remaining coarse signal approximation $a_J[n] = \langle f, \phi_{J,n} \rangle$ for $J = -5$.

Wavelet Design Theorem 7.3 constructs a wavelet orthonormal basis from any conjugate mirror filter $\hat{h}(\omega)$. This gives a simple procedure for designing and building wavelet orthogonal bases. Conversely, we may wonder whether all wavelet orthonormal bases are associated to a multiresolution approximation and a conjugate mirror filter. If we impose that ψ has a compact support then Lemarié [41] proved that ψ necessarily corresponds to a multiresolution approximation. It is however possible to construct pathological wavelets that decay like $|t|^{-1}$ at infinity, and which cannot be derived from any multiresolution approximation. Section

7.2 describes important classes of wavelet bases and explains how to design $\hat{h}$ to specify the support, the number of vanishing moments and the regularity of ψ.

7.2 CLASSES OF WAVELET BASES [1]

7.2.1 Choosing a Wavelet

Most applications of wavelet bases exploit their ability to efficiently approximate particular classes of functions with few non-zero wavelet coefficients. This is true not only for data compression but also for noise removal and fast calculations. The design of ψ must therefore be optimized to produce a maximum number of wavelet coefficients $\langle f, \psi_{j,n} \rangle$ that are close to zero. A function f has few non-negligible wavelet coefficients if most of the fine-scale (high-resolution) wavelet coefficients are small. This depends mostly on the regularity of f, the number of vanishing moments of ψ and the size of its support. To construct an appropriate wavelet from a conjugate mirror filter $h[n]$, we relate these properties to conditions on $\hat{h}(\omega)$.

Vanishing Moments Let us recall that ψ has p vanishing moments if

$$\int_{-\infty}^{+\infty} t^k \psi(t)\, dt = 0 \quad \text{for } 0 \leq k < p. \tag{7.74}$$

This means that ψ is orthogonal to any polynomial of degree $p-1$. Section 6.1.3 proves that if f is regular and ψ has enough vanishing moments then the wavelet coefficients $|\langle f, \psi_{j,n} \rangle|$ are small at fine scales 2^j. Indeed, if f is locally $\mathbf{C}^k$, then over a small interval it is well approximated by a Taylor polynomial of degree k. If $k < p$, then wavelets are orthogonal to this Taylor polynomial and thus produce small amplitude coefficients at fine scales. The following theorem relates the number of vanishing moments of ψ to the vanishing derivatives of $\hat{\psi}(\omega)$ at $\omega = 0$ and to the number of zeros of $\hat{h}(\omega)$ at $\omega = \pi$. It also proves that polynomials of degree $p-1$ are then reproduced by the scaling functions.

Theorem 7.4 (VANISHING MOMENTS) *Let ψ and ϕ be a wavelet and a scaling function that generate an orthogonal basis. Suppose that $|\psi(t)| = O((1+t^2)^{-p/2-1})$ and $|\phi(t)| = O((1+t^2)^{-p/2-1})$. The four following statements are equivalent:*

(i) The wavelet ψ has p vanishing moments.
(ii) $\hat{\psi}(\omega)$ and its first $p-1$ derivatives are zero at $\omega = 0$.
(iii) $\hat{h}(\omega)$ and its first $p-1$ derivatives are zero at $\omega = \pi$.
(iv) For any $0 \leq k < p$,

$$q_k(t) = \sum_{n=-\infty}^{+\infty} n^k \phi(t-n) \quad \textit{is a polynomial of degree } k. \tag{7.75}$$

Proof [2]. The decay of $|\phi(t)|$ and $|\psi(t)|$ implies that $\hat{\psi}(\omega)$ and $\hat{\phi}(\omega)$ are p times continuously differentiable. The k^{th} order derivative $\hat{\psi}^{(k)}(\omega)$ is the Fourier transform of

$(-it)^k \psi(t)$. Hence

$$\hat{\psi}^{(k)}(0) = \int_{-\infty}^{+\infty} (-it)^k \psi(t)\, dt.$$

We derive that (i) is equivalent to (ii).

Theorem 7.3 proves that

$$\sqrt{2}\,\hat{\psi}(2\omega) = \mathrm{e}^{-i\omega}\, \hat{h}^*(\omega+\pi)\, \hat{\phi}(\omega).$$

Since $\hat{\phi}(0) \neq 0$, by differentiating this expression we prove that (ii) is equivalent to (iii).

Let us now prove that (iv) implies (i). Since ψ is orthogonal to $\{\phi(t-n)\}_{n\in\mathbf{Z}}$, it is thus also orthogonal to the polynomials q_k for $0 \leq k < p$. This family of polynomials is a basis of the space of polynomials of degree at most $p-1$. Hence ψ is orthogonal to any polynomial of degree $p-1$ and in particular to t^k for $0 \leq k < p$. This means that ψ has p vanishing moments.

To verify that (i) implies (iv) we suppose that ψ has p vanishing moments, and for $k < p$ we evaluate $q_k(t)$ defined in (7.75). This is done by computing its Fourier transform:

$$\hat{q}_k(\omega) = \hat{\phi}(\omega) \sum_{n=-\infty}^{+\infty} n^k \exp(-in\omega) = (i)^k\, \hat{\phi}(\omega)\, \frac{d^k}{d\omega^k} \sum_{n=-\infty}^{+\infty} \exp(-in\omega)\,.$$

Let $\delta^{(k)}$ be the distribution that is the k^{th} order derivative of a Dirac, defined in Appendix A.7. The Poisson formula (2.4) proves that

$$\hat{q}_k(\omega) = (i)^k\, \frac{1}{2\pi}\, \hat{\phi}(\omega) \sum_{l=-\infty}^{+\infty} \delta^{(k)}(\omega - 2l\pi). \tag{7.76}$$

With several integrations by parts, we verify the distribution equality

$$\hat{\phi}(\omega)\, \delta^{(k)}(\omega - 2l\pi) = \hat{\phi}(2l\pi)\, \delta^{(k)}(\omega - 2l\pi) + \sum_{m=0}^{k-1} a_{m,l}^k\, \delta^{(m)}(\omega - 2l\pi), \tag{7.77}$$

where $a_{m,l}^k$ is a linear combination of the derivatives $\{\hat{\phi}^{(m)}(2l\pi)\}_{0\leq m\leq k}$.

For $l \neq 0$, let us prove that $a_{m,l}^k = 0$ by showing that $\hat{\phi}^{(m)}(2l\pi) = 0$ if $0 \leq m < p$. For any $P > 0$, (7.32) implies

$$\hat{\phi}(\omega) = \hat{\phi}(2^{-P}\omega) \prod_{p=1}^{P} \frac{\hat{h}(2^{-p}\omega)}{\sqrt{2}}. \tag{7.78}$$

Since ψ has p vanishing moments, we showed in (iii) that $\hat{h}(\omega)$ has a zero of order p at $\omega = \pm\pi$. But $\hat{h}(\omega)$ is also 2π periodic, so (7.78) implies that $\hat{\phi}(\omega) = O(|\omega - 2l\pi|^p)$ in the neighborhood of $\omega = 2l\pi$, for any $l \neq 0$. Hence $\hat{\phi}^{(m)}(2l\pi) = 0$ if $m < p$.

Since $a_{m,l}^k = 0$ and $\phi(2l\pi) = 0$ when $l \neq 0$, it follows from (7.77) that

$$\hat{\phi}(\omega)\, \delta^{(k)}(\omega - 2l\pi) = 0 \quad \text{for } l \neq 0.$$

The only term that remains in the summation (7.76) is $l = 0$ and inserting (7.77) yields

$$\hat{q}_k(\omega) = (i)^k \frac{1}{2\pi} \left(\hat{\phi}(0)\, \delta^{(k)}(\omega) + \sum_{m=0}^{k-1} a_{m,0}^k\, \delta^{(m)}(\omega) \right) .$$

The inverse Fourier transform of $\delta^{(m)}(\omega)$ is $(2\pi)^{-1}(-it)^m$ and Theorem 7.2 proves that $\hat{\phi}(0) \neq 0$. Hence the inverse Fourier transform q_k of $\hat{q}_k$ is a polynomial of degree k. ■

The hypothesis (iv) is called the Fix-Strang condition [320]. The polynomials $\{q_k\}_{0 \leq k < p}$ define a basis of the space of polynomials of degree $p-1$. The Fix-Strang condition thus proves that ψ has p vanishing moments if and only if any polynomial of degree $p-1$ can be written as a linear expansion of $\{\phi(t-n)\}_{n \in \mathbb{Z}}$. The decomposition coefficients of the polynomials q_k do not have a finite energy because polynomials do not have a finite energy.

Size of Support If f has an isolated singularity at t_0 and if t_0 is inside the support of $\psi_{j,n}(t) = 2^{-j/2}\psi(2^{-j}t - n)$, then $\langle f, \psi_{j,n} \rangle$ may have a large amplitude. If ψ has a compact support of size K, at each scale 2^j there are K wavelets $\psi_{j,n}$ whose support includes t_0. To minimize the number of high amplitude coefficients we must reduce the support size of ψ. The following proposition relates the support size of h to the support of ϕ and ψ.

Proposition 7.2 (COMPACT SUPPORT) *The scaling function ϕ has a compact support if and only if h has a compact support and their support are equal. If the support of h and ϕ is $[N_1, N_2]$ then the support of ψ is $[(N_1 - N_2 + 1)/2, (N_2 - N_1 + 1)/2]$.*

Proof[1]. If ϕ has a compact support, since

$$h[n] = \frac{1}{\sqrt{2}} \left\langle \phi\left(\frac{t}{2}\right), \phi(t-n) \right\rangle,$$

we derive that h also has a compact support. Conversely, the scaling function satisfies

$$\frac{1}{\sqrt{2}} \phi\left(\frac{t}{2}\right) = \sum_{n=-\infty}^{+\infty} h[n]\, \phi(t-n). \tag{7.79}$$

If h has a compact support then one can prove [144] that ϕ has a compact support. The proof is not reproduced here.

To relate the support of ϕ and h, we suppose that $h[n]$ is non-zero for $N_1 \leq n \leq N_2$ and that ϕ has a compact support $[K_1, K_2]$. The support of $\phi(t/2)$ is $[2K_1, 2K_2]$. The sum at the right of (7.79) is a function whose support is $[N_1 + K_1, N_2 + K_2]$. The equality proves that the support of ϕ is $[K_1, K_2] = [N_1, N_2]$.

Let us recall from (7.73) and (7.72) that

$$\frac{1}{\sqrt{2}} \psi\left(\frac{t}{2}\right) = \sum_{n=-\infty}^{+\infty} g[n]\, \phi(t-n) = \sum_{n=-\infty}^{+\infty} (-1)^{1-n} h[1-n]\, \phi(t-n).$$

If the supports of ϕ and h are equal to $[N_1, N_2]$, the sum in the right-hand side has a support equal to $[N_1 - N_2 + 1, N_2 - N_1 + 1]$. Hence ψ has a support equal to $[(N_1 - N_2 + 1)/2, (N_2 - N_1 + 1)/2]$. ■

If h has a finite impulse response in $[N_1, N_2]$, Proposition 7.2 proves that ψ has a support of size $N_2 - N_1$ centered at $1/2$. To minimize the size of the support, we must synthesize conjugate mirror filters with as few non-zero coefficients as possible.

Support Versus Moments The support size of a function and the number of vanishing moments are a priori independent. However, we shall see in Theorem 7.5 that the constraints imposed on orthogonal wavelets imply that if ψ has p vanishing moments then its support is at least of size $2p - 1$. Daubechies wavelets are optimal in the sense that they have a minimum size support for a given number of vanishing moments. When choosing a particular wavelet, we thus face a trade-off between the number of vanishing moments and the support size. If f has few isolated singularities and is very regular between singularities, we must choose a wavelet with many vanishing moments to produce a large number of small wavelet coefficients $\langle f, \psi_{j,n} \rangle$. If the density of singularities increases, it might be better to decrease the size of its support at the cost of reducing the number of vanishing moments. Indeed, wavelets that overlap the singularities create high amplitude coefficients.

The multiwavelet construction of Geronimo, Hardin and Massupust [190] offers more design flexibility by introducing several scaling functions and wavelets. Problem 7.16 gives an example. Better trade-off can be obtained between the multiwavelets supports and their vanishing moments [321]. However, multiwavelet decompositions are implemented with a slightly more complicated filter bank algorithm than a standard orthogonal wavelet transform.

Regularity The regularity of ψ has mostly a cosmetic influence on the error introduced by thresholding or quantizing the wavelet coefficients. When reconstructing a signal from its wavelet coefficients

$$f = \sum_{j=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \langle f, \psi_{j,n} \rangle \, \psi_{j,n},$$

an error ϵ added to a coefficient $\langle f, \psi_{j,n} \rangle$ will add the wavelet component $\epsilon \psi_{j,n}$ to the reconstructed signal. If ψ is smooth, then $\epsilon \psi_{j,n}$ is a smooth error. For image coding applications, a smooth error is often less visible than an irregular error, even though they have the same energy. Better quality images are obtained with wavelets that are continuously differentiable than with the discontinuous Haar wavelet. The following proposition due to Tchamitchian [327] relates the uniform Lipschitz regularity of ϕ and ψ to the number of zeros of $\hat{h}(\omega)$ at $\omega = \pi$.

Proposition 7.3 (TCHAMITCHIAN) *Let $\hat{h}(\omega)$ be a conjugate mirror filter with p zeros at π and which satisfies the sufficient conditions of Theorem 7.2. Let us perform the factorization*

$$\hat{h}(\omega) = \sqrt{2}\left(\frac{1+e^{i\omega}}{2}\right)^p \hat{l}(\omega).$$

If $\sup_{\omega\in\mathbb{R}} |\hat{l}(\omega)| = B$ then ψ and ϕ are uniformly Lipschitz α for

$$\alpha < \alpha_0 = p - \log_2 B - 1. \tag{7.80}$$

Proof[3]. This result is proved by showing that there exist $C_1 > 0$ and $C_2 > 0$ such that for all $\omega \in \mathbb{R}$

$$|\hat{\phi}(\omega)| \leq C_1 (1+|\omega|)^{-p+\log_2 B} \tag{7.81}$$
$$|\hat{\psi}(\omega)| \leq C_2 (1+|\omega|)^{-p+\log_2 B}. \tag{7.82}$$

The Lipschitz regularity of ϕ and ψ is then derived from Theorem 6.1, which shows that if $\int_{-\infty}^{+\infty} (1+|\omega|^\alpha)\,|\hat{f}(\omega)|\,d\omega < +\infty$, then f is uniformly Lipschitz α.

We proved in (7.37) that $\hat{\phi}(\omega) = \prod_{j=1}^{+\infty} 2^{-1/2}\hat{h}(2^{-j}\omega)$. One can verify that

$$\prod_{j=1}^{+\infty} \frac{1+\exp(i2^{-j}\omega)}{2} = \frac{1-\exp(i\omega)}{i\omega},$$

hence

$$|\hat{\phi}(\omega)| = \frac{|1-\exp(i\omega)|^p}{|\omega|^p} \prod_{j=1}^{+\infty} |\hat{l}(2^{-j}\omega)|. \tag{7.83}$$

Let us now compute an upper bound for $\prod_{j=1}^{+\infty} |\hat{l}(2^{-j}\omega)|$. At $\omega = 0$ we have $\hat{h}(0) = \sqrt{2}$ so $\hat{l}(0) = 1$. Since $\hat{h}(\omega)$ is continuously differentiable at $\omega = 0$, $\hat{l}(\omega)$ is also continuously differentiable at $\omega = 0$. We thus derive that there exists $\epsilon > 0$ such that if $|\omega| < \epsilon$ then $|\hat{l}(\omega)| \leq 1 + K|\omega|$. Consequently

$$\sup_{|\omega|\leq\epsilon} \prod_{j=1}^{+\infty} |\hat{l}(2^{-j}\omega)| \leq \sup_{|\omega|\leq\epsilon} \prod_{j=1}^{+\infty} (1+K|2^{-j}\omega|) \leq e^{K\epsilon}. \tag{7.84}$$

If $|\omega| > \epsilon$, there exists $J \geq 1$ such that $2^{J-1}\epsilon \leq |\omega| \leq 2^J\epsilon$ and we decompose

$$\prod_{j=1}^{+\infty} \hat{l}(2^{-j}\omega) = \prod_{j=1}^{J} |\hat{l}(2^{-j}\omega)| \prod_{j=1}^{+\infty} |\hat{l}(2^{-j-J}\omega)|. \tag{7.85}$$

Since $\sup_{\omega\in\mathbb{R}} |\hat{l}(\omega)| = B$, inserting (7.84) yields for $|\omega| > \epsilon$

$$\prod_{j=1}^{+\infty} \hat{l}(2^{-j}\omega) \leq B^J e^{K\epsilon} = e^{K\epsilon} 2^{J\log_2 B}. \tag{7.86}$$

Since $2^J \leq \epsilon^{-1} 2|\omega|$, this proves that

$$\forall \omega \in \mathbb{R} \ , \ \prod_{j=1}^{+\infty} \hat{l}(2^{-j}\omega) \leq e^{K\epsilon} \left(1 + \frac{|2\omega|^{\log_2 B}}{\epsilon^{\log_2 B}}\right).$$

Equation (7.81) is derived from (7.83) and this last inequality. Since $|\hat{\psi}(2\omega)| = 2^{-1/2} |\hat{h}(\omega+\pi)| |\hat{\phi}(\omega)|$, (7.82) is obtained from (7.81). ■

This proposition proves that if $B < 2^{p-1}$ then $\alpha_0 > 0$. It means that ϕ and ψ are uniformly continuous. For any $m > 0$, if $B < 2^{p-1-m}$ then $\alpha_0 > m$ so ψ and ϕ are m times continuously differentiable. Theorem 7.4 shows that the number p of zeros of $\hat{h}(\omega)$ at π is equal to the number of vanishing moments of ψ. A priori, we are not guaranteed that increasing p will improve the wavelet regularity, since B might increase as well. However, for important families of conjugate mirror filters such as splines or Daubechies filters, B increases more slowly than p, which implies that wavelet regularity increases with the number of vanishing moments. Let us emphasize that the number of vanishing moments and the regularity of orthogonal wavelets are related but it is the number of vanishing moments and not the regularity that affects the amplitude of the wavelet coefficients at fine scales.

7.2.2 Shannon, Meyer and Battle-Lemarié Wavelets

We study important classes of wavelets whose Fourier transforms are derived from the general formula proved in Theorem 7.3,

$$\hat{\psi}(\omega) = \frac{1}{\sqrt{2}} \hat{g}\left(\frac{\omega}{2}\right) \hat{\phi}\left(\frac{\omega}{2}\right) = \frac{1}{\sqrt{2}} \exp\left(\frac{-i\omega}{2}\right) \hat{h}^*\left(\frac{\omega}{2}+\pi\right) \hat{\phi}\left(\frac{\omega}{2}\right). \quad (7.87)$$

Shannon Wavelet The Shannon wavelet is constructed from the Shannon multiresolution approximation, which approximates functions by their restriction to low frequency intervals. It corresponds to $\hat{\phi} = \mathbf{1}_{[-\pi,\pi]}$ and $\hat{h}(\omega) = \sqrt{2}\mathbf{1}_{[-\pi/2,\pi/2]}(\omega)$ for $\omega \in [-\pi,\pi]$. We derive from (7.87) that

$$\hat{\psi}(\omega) = \begin{cases} \exp(-i\omega/2) & \text{if } \omega \in [-2\pi,-\pi] \cup [\pi,2\pi] \\ 0 & \text{otherwise} \end{cases} \quad (7.88)$$

and hence

$$\psi(t) = \frac{\sin 2\pi(t-1/2)}{2\pi(t-1/2)} - \frac{\sin \pi(t-1/2)}{\pi(t-1/2)}.$$

This wavelet is $\mathbf{C}^\infty$ but has a slow asymptotic time decay. Since $\hat{\psi}(\omega)$ is zero in the neighborhood of $\omega = 0$, all its derivatives are zero at $\omega = 0$. Theorem 7.4 thus implies that ψ has an infinite number of vanishing moments.

Since $\hat{\psi}(\omega)$ has a compact support we know that $\psi(t)$ is $\mathbf{C}^\infty$. However $|\psi(t)|$ decays only like $|t|^{-1}$ at infinity because $\hat{\psi}(\omega)$ is discontinuous at $\pm\pi$ and $\pm 2\pi$.

Meyer Wavelets A Meyer wavelet [270] is a frequency band-limited function whose Fourier transform is smooth, unlike the Fourier transform of the Shannon wavelet. This smoothness provides a much faster asymptotic decay in time. These wavelets are constructed with conjugate mirror filters $\hat{h}(\omega)$ that are $\mathbf{C}^n$ and satisfy

$$\hat{h}(\omega) = \begin{cases} \sqrt{2} & \text{if } \omega \in [-\pi/3, \pi/3] \\ 0 & \text{if } \omega \in [-\pi, -2\pi/3] \cup [2\pi/3, \pi] \end{cases} . \tag{7.89}$$

The only degree of freedom is the behavior of $\hat{h}(\omega)$ in the transition bands $[-2\pi/3, -\pi/3] \cup [\pi/3, 2\pi/3]$. It must satisfy the quadrature condition

$$|\hat{h}(\omega)|^2 + |\hat{h}(\omega+\pi)|^2 = 2, \tag{7.90}$$

and to obtain $\mathbf{C}^n$ junctions at $|\omega| = \pi/3$ and $|\omega| = 2\pi/3$, the n first derivatives must vanish at these abscissa. One can construct such functions that are $\mathbf{C}^\infty$.

The scaling function $\hat{\phi}(\omega) = \prod_{p=1}^{+\infty} 2^{-1/2} \hat{h}(2^{-p}\omega)$ has a compact support and one can verify that

$$\hat{\phi}(\omega) = \begin{cases} 2^{-1/2} \hat{h}(\omega/2) & \text{if } |\omega| \leq 4\pi/3 \\ 0 & \text{if } |\omega| > 4\pi/3 \end{cases} . \tag{7.91}$$

The resulting wavelet (7.87) is

$$\hat{\psi}(\omega) = \begin{cases} 0 & \text{if } |\omega| \leq 2\pi/3 \\ 2^{-1/2} \hat{g}(\omega/2) & \text{if } 2\pi/3 \leq |\omega| \leq 4\pi/3 \\ 2^{-1/2} \exp(-i\omega/2) \hat{h}(\omega/4) & \text{if } 4\pi/3 \leq |\omega| \leq 8\pi/3 \\ 0 & \text{if } |\omega| > 8\pi/3 \end{cases} . \tag{7.92}$$

The functions ϕ and ψ are $\mathbf{C}^\infty$ because their Fourier transforms have a compact support. Since $\hat{\psi}(\omega) = 0$ in the neighborhood of $\omega = 0$, all its derivatives are zero at $\omega = 0$, which proves that ψ has an infinite number of vanishing moments.

If $\hat{h}$ is $\mathbf{C}^n$ then $\hat{\psi}$ and $\hat{\phi}$ are also $\mathbf{C}^n$. The discontinuities of the $(n+1)^{th}$ derivative of $\hat{h}$ are generally at the junction of the transition band $|\omega| = \pi/3, 2\pi/3$, in which case one can show that there exists A such that

$$|\phi(t)| \leq A(1+|t|)^{-n-1} \quad \text{and} \quad |\psi(t)| \leq A(1+|t|)^{-n-1} .$$

Although the asymptotic decay of ψ is fast when n is large, its effective numerical decay may be relatively slow, which is reflected by the fact that A is quite large. As a consequence, a Meyer wavelet transform is generally implemented in the Fourier domain. Section 8.4.2 relates these wavelet bases to lapped orthogonal transforms applied in the Fourier domain. One can prove [21] that there exists no orthogonal wavelet that is $\mathbf{C}^\infty$ and has an exponential decay.

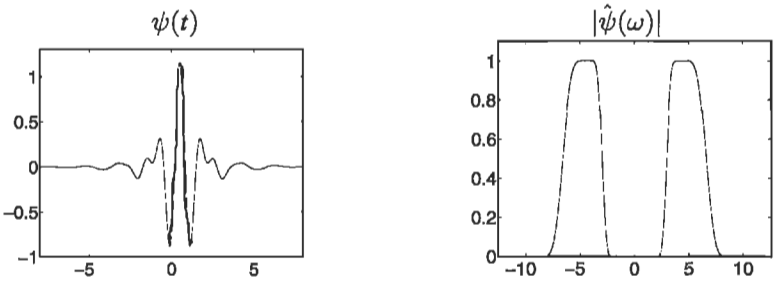


FIGURE 7.8 Meyer wavelet ψ and its Fourier transform modulus computed with (7.94).

Example 7.10 To satisfy the quadrature condition (7.90), one can verify that $\hat{h}$ in (7.89) may be defined on the transition bands by

$$\hat{h}(\omega) = \sqrt{2}\cos\left[\frac{\pi}{2}\beta\left(\frac{3|\omega|}{\pi} - 1\right)\right] \quad \text{for } |\omega| \in [\pi/3, 2\pi/3]\,,$$

where $\beta(x)$ is a function that goes from 0 to 1 on the interval $[0,1]$ and satisfies

$$\forall x \in [0,1]\ ,\ \ \beta(x) + \beta(1-x) = 1. \tag{7.93}$$

An example due to Daubechies [21] is

$$\beta(x) = x^4(35 - 84x + 70x^2 - 20x^3). \tag{7.94}$$

The resulting $\hat{h}(\omega)$ has $n = 3$ vanishing derivatives at $|\omega| = \pi/3\,, 2\pi/3$. Figure 7.8 displays the corresponding wavelet ψ.

Haar Wavelet The Haar basis is obtained with a multiresolution of piecewise constant functions. The scaling function is $\phi = \mathbf{1}_{[0,1]}$. The filter $h[n]$ given in (7.51) has two non-zero coefficients equal to $2^{-1/2}$ at $n = 0$ and $n = 1$. Hence

$$\frac{1}{\sqrt{2}}\psi\left(\frac{t}{2}\right) = \sum_{n=-\infty}^{+\infty} (-1)^{1-n} h[1-n]\,\phi(t-n) = \frac{1}{\sqrt{2}}\Big(\phi(t-1) - \phi(t)\Big),$$

so

$$\psi(t) = \begin{cases} -1 & \text{if } 0 \le t < 1/2 \\ 1 & \text{if } 1/2 \le t < 1 \\ 0 & \text{otherwise} \end{cases} \tag{7.95}$$

The Haar wavelet has the shortest support among all orthogonal wavelets. It is not well adapted to approximating smooth functions because it has only one vanishing moment.

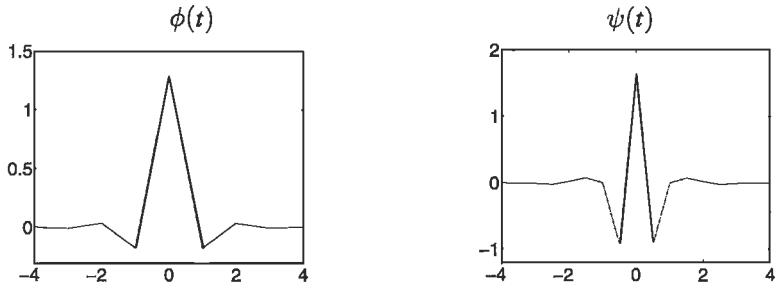


FIGURE 7.9 Linear spline Battle-Lemarié scaling function ϕ and wavelet ψ.

Battle-Lemarié Wavelets Polynomial spline wavelets introduced by Battle [89] and Lemarié [249] are computed from spline multiresolution approximations. The expressions of $\hat{\phi}(\omega)$ and $\hat{h}(\omega)$ are given respectively by (7.23) and (7.53). For splines of degree m, $\hat{h}(\omega)$ and its first m derivatives are zero at $\omega = \pi$. Theorem 7.4 derives that ψ has $m+1$ vanishing moments. It follows from (7.87) that

$$\hat{\psi}(\omega) = \frac{\exp(-i\omega/2)}{\omega^{m+1}} \sqrt{\frac{S_{2m+2}(\omega/2+\pi)}{S_{2m+2}(\omega)\, S_{2m+2}(\omega/2)}}.$$

This wavelet ψ has an exponential decay. Since it is a polynomial spline of degree m, it is $m-1$ times continuously differentiable. Polynomial spline wavelets are less regular than Meyer wavelets but have faster time asymptotic decay. For m odd, ψ is symmetric about $1/2$. For m even it is antisymmetric about $1/2$. Figure 7.5 gives the graph of the cubic spline wavelet ψ corresponding to $m = 3$. For $m = 1$, Figure 7.9 displays linear splines ϕ and ψ. The properties of these wavelets are further studied in [93, 15, 125].

7.2.3 Daubechies Compactly Supported Wavelets

Daubechies wavelets have a support of minimum size for any given number p of vanishing moments. Proposition 7.2 proves that wavelets of compact support are computed with finite impulse response conjugate mirror filters h. We consider real causal filters $h[n]$, which implies that $\hat{h}$ is a trigonometric polynomial:

$$\hat{h}(\omega) = \sum_{n=0}^{N-1} h[n]\, e^{-in\omega}.$$

To ensure that ψ has p vanishing moments, Theorem 7.4 shows that $\hat{h}$ must have a zero of order p at $\omega = \pi$. To construct a trigonometric polynomial of minimal size, we factor $(1+e^{-i\omega})^p$, which is a minimum size polynomial having p zeros at $\omega = \pi$:

$$\hat{h}(\omega) = \sqrt{2} \left(\frac{1+e^{-i\omega}}{2}\right)^p R(e^{-i\omega}). \tag{7.96}$$

The difficulty is to design a polynomial $R(e^{-i\omega})$ of minimum degree m such that $\hat{h}$ satisfies

$$|\hat{h}(\omega)|^2 + |\hat{h}(\omega+\pi)|^2 = 2. \tag{7.97}$$

As a result, h has $N = m+p+1$ non-zero coefficients. The following theorem by Daubechies [144] proves that the minimum degree of R is $m = p-1$.

Theorem 7.5 (DAUBECHIES) *A real conjugate mirror filter h, such that $\hat{h}(\omega)$ has p zeros at $\omega = \pi$, has at least $2p$ non-zero coefficients. Daubechies filters have $2p$ non-zero coefficients.*

Proof[2]. The proof is constructive and computes the Daubechies filters. Since $h[n]$ is real, $|\hat{h}(\omega)|^2$ is an even function and can thus be written as a polynomial in $\cos\omega$. Hence $|R(e^{-i\omega})|^2$ defined in (7.96) is a polynomial in $\cos\omega$ that we can also write as a polynomial $P(\sin^2\frac{\omega}{2})$

$$|\hat{h}(\omega)|^2 = 2\left(\cos\frac{\omega}{2}\right)^{2p} P\left(\sin^2\frac{\omega}{2}\right). \tag{7.98}$$

The quadrature condition (7.97) is equivalent to

$$(1-y)^p P(y) + y^p P(1-y) = 1, \tag{7.99}$$

for any $y = \sin^2(\omega/2) \in [0,1]$. To minimize the number of non-zero terms of the finite Fourier series $\hat{h}(\omega)$, we must find the solution $P(y) \geq 0$ of minimum degree, which is obtained with the Bezout theorem on polynomials.

Theorem 7.6 (BEZOUT) *Let $Q_1(y)$ and $Q_2(y)$ be two polynomials of degrees n_1 and n_2 with no common zeros. There exist two unique polynomials $P_1(y)$ and $P_2(y)$ of degrees n_2-1 and n_1-1 such that*

$$P_1(y)\,Q_1(y) + P_2(y)\,Q_2(y) = 1. \tag{7.100}$$

The proof of this classical result is in [21]. Since $Q_1(y) = (1-y)^p$ and $Q_2(y) = y^p$ are two polynomials of degree p with no common zeros, the Bezout theorem proves that there exist two unique polynomials $P_1(y)$ and $P_2(y)$ such that

$$(1-y)^p P_1(y) + y^p P_2(y) = 1.$$

The reader can verify that $P_2(y) = P_1(1-y) = P(1-y)$ with

$$P(y) = \sum_{k=0}^{p-1} \binom{p-1+k}{k} y^k. \tag{7.101}$$

Clearly $P(y) \geq 0$ for $y \in [0,1]$. Hence $P(y)$ is the polynomial of minimum degree satisfying (7.99) with $P(y) \geq 0$.

Minimum Phase Factorization Now we need to construct a minimum degree polynomial

$$R(e^{-i\omega}) = \sum_{k=0}^{m} r_k e^{-ik\omega} = r_0 \prod_{k=0}^{m} (1 - a_k e^{-i\omega})$$

such that $|R(e^{-i\omega})|^2 = P(\sin^2(\omega/2))$. Since its coefficients are real, $R^*(e^{-i\omega}) = R(e^{i\omega})$ and hence

$$|R(e^{-i\omega})|^2 = R(e^{-i\omega}) R(e^{i\omega}) = P\left(\frac{2 - e^{i\omega} - e^{-i\omega}}{4}\right) = Q(e^{-i\omega}). \tag{7.102}$$

This factorization is solved by extending it to the whole complex plane with the variable $z = e^{-i\omega}$:

$$R(z) R(z^{-1}) = r_0^2 \prod_{k=0}^{m} (1 - a_k z)(1 - a_k z^{-1}) = Q(z) = P\left(\frac{2 - z - z^{-1}}{4}\right). \tag{7.103}$$

Let us compute the roots of $Q(z)$. Since $Q(z)$ has real coefficients if c_k is a root, then c_k^* is also a root and since it is a function of $z + z^{-1}$ if c_k is a root then $1/c_k$ and hence $1/c_k^*$ are also roots. To design $R(z)$ that satisfies (7.103), we choose each root a_k of $R(z)$ among a pair $(c_k, 1/c_k)$ and include a_k^* as a root to obtain real coefficients. This procedure yields a polynomial of minimum degree $m = p - 1$, with $r_0^2 = Q(0) = P(1/2) = 2^{p-1}$. The resulting filter h of minimum size has $N = p + m + 1 = 2p$ non-zero coefficients.

Among all possible factorizations, the minimum phase solution $R(e^{i\omega})$ is obtained by choosing a_k among $(c_k, 1/c_k)$ to be inside the unit circle $|a_k| \leq 1$ [55]. The resulting causal filter h has an energy maximally concentrated at small abscissa $n \geq 0$. It is a Daubechies filter of order p. ■

The constructive proof of this theorem synthesizes causal conjugate mirror filters of size $2p$. Table 7.2 gives the coefficients of these Daubechies filters for $2 \leq p \leq 10$. The following proposition derives that Daubechies wavelets calculated with these conjugate mirror filters have a support of minimum size.

Proposition 7.4 (DAUBECHIES) *If ψ is a wavelet with p vanishing moments that generates an orthonormal basis of $\mathbf{L}^2(\mathbb{R})$, then it has a support of size larger than or equal to $2p - 1$. A Daubechies wavelet has a minimum size support equal to $[-p + 1, p]$. The support of the corresponding scaling function ϕ is $[0, 2p - 1]$.*

This proposition is a direct consequence of Theorem 7.5. The support of the wavelet, and that of the scaling function, are calculated with Proposition 7.2. When $p = 1$ we get the Haar wavelet. Figure 7.10 displays the graphs of ϕ and ψ for $p = 2, 3, 4$.

The regularity of ϕ and ψ is the same since $\psi(t)$ is a finite linear combination of the $\phi(2t - n)$. This regularity is however difficult to estimate precisely. Let $B = \sup_{\omega \in \mathbb{R}} |R(e^{-i\omega})|$ where $R(e^{-i\omega})$ is the trigonometric polynomial defined in (7.96). Proposition 7.3 proves that ψ is at least uniformly Lipschitz α for $\alpha < p - \log_2 B - 1$. For Daubechies wavelets, B increases more slowly than p and Figure 7.10 shows indeed that the regularity of these wavelets increases with p.

	n	$h_p[n]$
p = 2	0	.482962913145
	1	.836516303738
	2	.224143868042
	3	−.129409522551
p = 3	0	.332670552950
	1	.806891509311
	2	.459877502118
	3	−.135011020010
	4	−.085441273882
	5	.035226291882
p = 4	0	.230377813309
	1	.714846570553
	2	.630880767930
	3	−.027983769417
	4	−.187034811719
	5	.030841381836
	6	.032883011667
	7	−.010597401785
p = 5	0	.160102397974
	1	.603829269797
	2	.724308528438
	3	.138428145901
	4	−.242294887066
	5	−.032244869585
	6	.077571493840
	7	−.006241490213
	8	−.012580751999
	9	.003335725285
p = 6	0	.111540743350
	1	.494623890398
	2	.751133908021
	3	.315250351709
	4	−.226264693965
	5	−.129766867567
	6	.097501605587
	7	.027522865530
	8	−.031582039317
	9	.000553842201
	10	.004777257511
	11	−.001077301085
p = 7	0	.077852054085
	1	.396539319482
	2	.729132090846
	3	.469782287405
	4	−.143906003929
	5	−.224036184994
	6	.071309219267
	7	.080612609151
	8	−.038029936935
	9	−.016574541631
	10	.012550998556
	11	.000429577973
	12	−.001801640704
	13	.000353713800

	n	$h_p[n]$
p = 8	0	.054415842243
	1	.312871590914
	2	.675630736297
	3	.585354683654
	4	−.015829105256
	5	−.284015542962
	6	.000472484574
	7	.128747426620
	8	−.017369301002
	9	−.04408825393
	10	.013981027917
	11	.008746094047
	12	−.004870352993
	13	−.000391740373
	14	.000675449406
	15	−.000117476784
p = 9	0	.038077947364
	1	.243834674613
	2	.604823123690
	3	.657288078051
	4	.133197385825
	5	−.293273783279
	6	−.096840783223
	7	.148540749338
	8	.030725681479
	9	−.067632829061
	10	.000250947115
	11	.022361662124
	12	−.004723204758
	13	−.004281503682
	14	.001847646883
	15	.000230385764
	16	−.000251963189
	17	.000039347320
p = 10	0	.026670057901
	1	.188176800078
	2	.527201188932
	3	.688459039454
	4	.281172343661
	5	−.249846424327
	6	−.195946274377
	7	.127369340336
	8	.093057364604
	9	−.071394147166
	10	−.029457536822
	11	.033212674059
	12	.003606553567
	13	−.010733175483
	14	.001395351747
	15	.001992405295
	16	−.000685856695
	17	−.000116466855
	18	.000093588670
	19	−.000013264203

Table 7.2 Daubechies filters for wavelets with p vanishing moments.

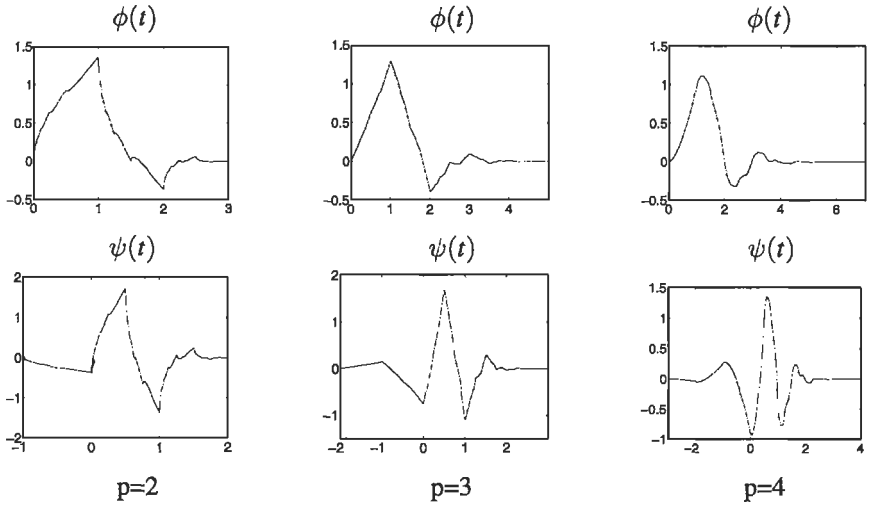


FIGURE 7.10 Daubechies scaling function ϕ and wavelet ψ with p vanishing moments.

Daubechies and Lagarias [147] have established a more precise technique that computes the exact Lipschitz regularity of ψ. For $p = 2$ the wavelet ψ is only Lipschitz 0.55 but for $p = 3$ it is Lipschitz 1.08 which means that it is already continuously differentiable. For p large, ϕ and ψ are uniformly Lipschitz α for α of the order of $0.2\,p$ [129].

Symmlets Daubechies wavelets are very asymmetric because they are constructed by selecting the minimum phase square root of $Q(e^{-i\omega})$ in (7.102). One can show [55] that filters corresponding to a minimum phase square root have their energy optimally concentrated near the starting point of their support. They are thus highly non-symmetric, which yields very asymmetric wavelets.

To obtain a symmetric or antisymmetric wavelet, the filter h must be symmetric or antisymmetric with respect to the center of its support, which means that $\hat{h}(\omega)$ has a linear complex phase. Daubechies proved [144] that the Haar filter is the only real compactly supported conjugate mirror filter that has a linear phase. The *Symmlet* filters of Daubechies are obtained by optimizing the choice of the square root $R(e^{-i\omega})$ of $Q(e^{-i\omega})$ to obtain an almost linear phase. The resulting wavelets still have a minimum support $[-p+1, p]$ with p vanishing moments but they are more symmetric, as illustrated by Figure 7.11 for $p = 8$. The coefficients of the Symmlet filters are in WaveLab. Complex conjugate mirror filters with a compact support and a linear phase can be constructed [251], but they produce complex wavelet coefficients whose real and imaginary parts are redundant when the signal is real.

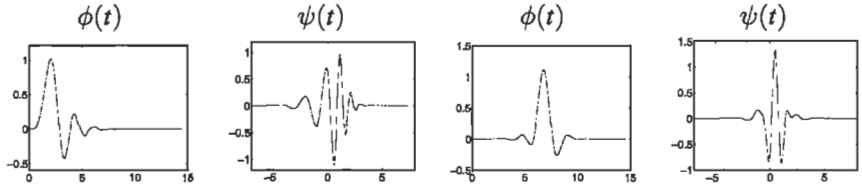


FIGURE 7.11 Daubechies (first two) and Symmlets (last two) scaling functions and wavelets with $p = 8$ vanishing moments.

Coiflets For an application in numerical analysis, Coifman asked Daubechies [144] to construct a family of wavelets ψ that have p vanishing moments and a minimum size support, but whose scaling functions also satisfy

$$\int_{-\infty}^{+\infty} \phi(t)\,dt = 1 \quad \text{and} \quad \int_{-\infty}^{+\infty} t^k \phi(t)\,dt = 0 \quad \text{for } 1 \leq k < p. \tag{7.104}$$

Such scaling functions are useful in establishing precise quadrature formulas. If f is $\mathbf{C}^k$ in the neighborhood of $2^J n$ with $k < p$, then a Taylor expansion of f up to order k shows that

$$2^{-J/2} \langle f, \phi_{J,n} \rangle \approx f(2^J n) + O(2^{(k+1)J}) \,. \tag{7.105}$$

At a fine scale 2^J, the scaling coefficients are thus closely approximated by the signal samples. The order of approximation increases with p. The supplementary condition (7.104) requires increasing the support of ψ; the resulting Coiflet has a support of size $3p - 1$ instead of $2p - 1$ for a Daubechies wavelet. The corresponding conjugate mirror filters are tabulated in WaveLab.

Audio Filters The first conjugate mirror filters with finite impulse response were constructed in 1986 by Smith and Barnwell [317] in the context of perfect filter bank reconstruction, explained in Section 7.3.2. These filters satisfy the quadrature condition $|\hat{h}(\omega)|^2 + |\hat{h}(\omega+\pi)|^2 = 2$, which is necessary and sufficient for filter bank reconstruction. However, $\hat{h}(0) \neq \sqrt{2}$ so the infinite product of such filters does not yield a wavelet basis of $\mathbf{L}^2(\mathbb{R})$. Instead of imposing any vanishing moments, Smith and Barnwell [317], and later Vaidyanathan and Hoang [337], designed their filters to reduce the size of the transition band, where $|\hat{h}(\omega)|$ decays from nearly $\sqrt{2}$ to nearly 0 in the neighborhood of $\pm\pi/2$. This constraint is important in optimizing the transform code of audio signals, explained in Section 11.3.3. However, many cascades of these filters exhibit wild behavior. The Vaidyanathan-Hoang filters are tabulated in WaveLab. Many other classes of conjugate mirror filters with finite impulse response have been constructed [74, 73]. Recursive conjugate mirror filters may also be designed [209] to minimize the size of the transition band for a given number of zeros at $\omega = \pi$. These filters have a fast but non-causal recursive implementation for signals of finite size.

7.3 WAVELETS AND FILTER BANKS [1]

Decomposition coefficients in a wavelet orthogonal basis are computed with a fast algorithm that cascades discrete convolutions with h and g, and subsamples the output. Section 7.3.1 derives this result from the embedded structure of multiresolution approximations. A direct filter bank analysis is performed in Section 7.3.2, which gives more general perfect reconstruction conditions on the filters. Section 7.3.3 shows that perfect reconstruction filter banks decompose signals in a basis of $\mathbf{l}^2(\mathbb{Z})$. This basis is orthogonal for conjugate mirror filters.

7.3.1 Fast Orthogonal Wavelet Transform

We describe a fast filter bank algorithm that computes the orthogonal wavelet coefficients of a signal measured at a finite resolution. A fast wavelet transform decomposes successively each approximation $P_{\mathbf{V}_j} f$ into a coarser approximation $P_{\mathbf{V}_{j+1}} f$ plus the wavelet coefficients carried by $P_{\mathbf{W}_{j+1}} f$. In the other direction, the reconstruction from wavelet coefficients recovers each $P_{\mathbf{V}_j} f$ from $P_{\mathbf{V}_{j+1}} f$ and $P_{\mathbf{W}_{j+1}} f$.

Since $\{\phi_{j,n}\}_{n\in\mathbb{Z}}$ and $\{\psi_{j,n}\}_{n\in\mathbb{Z}}$ are orthonormal bases of $\mathbf{V}_j$ and $\mathbf{W}_j$ the projection in these spaces is characterized by

$$a_j[n] = \langle f, \phi_{j,n} \rangle \text{ and } d_j[n] = \langle f, \psi_{j,n} \rangle .$$

The following theorem [253, 255] shows that these coefficients are calculated with a cascade of discrete convolutions and subsamplings. We denote $\bar{x}[n] = x[-n]$ and

$$\check{x}[n] = \begin{cases} x[p] & \text{if } n = 2p \\ 0 & \text{if } n = 2p+1 \end{cases} . \tag{7.106}$$

Theorem 7.7 (MALLAT) *At the decomposition*

$$a_{j+1}[p] = \sum_{n=-\infty}^{+\infty} h[n-2p]\, a_j[n] = a_j \star \bar{h}[2p], \tag{7.107}$$

$$d_{j+1}[p] = \sum_{n=-\infty}^{+\infty} g[n-2p]\, a_j[n] = a_j \star \bar{g}[2p]. \tag{7.108}$$

At the reconstruction,

$$\begin{aligned} a_j[p] &= \sum_{n=-\infty}^{+\infty} h[p-2n]\, a_{j+1}[n] + \sum_{n=-\infty}^{+\infty} g[p-2n]\, d_{j+1}[n] \\ &= \check{a}_{j+1} \star h[p] + \check{d}_{j+1} \star g[p]. \end{aligned} \tag{7.109}$$

Proof [1]. *Proof of (7.107)* Any $\phi_{j+1,p} \in \mathbf{V}_{j+1} \subset \mathbf{V}_j$ can be decomposed in the orthonormal basis $\{\phi_{j,n}\}_{n\in\mathbb{Z}}$ of $\mathbf{V}_j$:

$$\phi_{j+1,p} = \sum_{n=-\infty}^{+\infty} \langle \phi_{j+1,p}, \phi_{j,n} \rangle\, \phi_{j,n}. \tag{7.110}$$

With the change of variable $t' = 2^{-j}t - 2p$ we obtain

$$\begin{aligned}\langle \phi_{j+1,p}, \phi_{j,n} \rangle &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2^{j+1}}} \phi\left(\frac{t-2^{j+1}p}{2^{j+1}}\right) \frac{1}{\sqrt{2^j}} \phi^*\left(\frac{t-2^j n}{2^j}\right) dt \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2}} \phi\left(\frac{t}{2}\right) \phi^*(t-n+2p)\, dt \\ &= \left\langle \frac{1}{\sqrt{2}} \phi\left(\frac{t}{2}\right), \phi(t-n+2p) \right\rangle = h[n-2p]. \end{aligned} \tag{7.111}$$

Hence (7.110) implies that

$$\phi_{j+1,p} = \sum_{n=-\infty}^{+\infty} h[n-2p]\, \phi_{j,n}. \tag{7.112}$$

Computing the inner product of f with the vectors on each side of this equality yields (7.107).

Proof of (7.108) Since $\psi_{j+1,p} \in \mathbf{W}_{j+1} \subset \mathbf{V}_j$, it can be decomposed as

$$\psi_{j+1,p} = \sum_{n=-\infty}^{+\infty} \langle \psi_{j+1,p}, \phi_{j,n} \rangle\, \phi_{j,n}.$$

As in (7.111), the change of variable $t' = 2^{-j}t - 2p$ proves that

$$\langle \psi_{j+1,p}, \phi_{j,n} \rangle = \left\langle \frac{1}{\sqrt{2}} \psi\left(\frac{t}{2}\right), \phi(t-n+2p) \right\rangle = g[n-2p] \tag{7.113}$$

and hence

$$\psi_{j+1,p} = \sum_{n=-\infty}^{+\infty} g[n-2p]\, \phi_{j,n}. \tag{7.114}$$

Taking the inner product with f on each side gives (7.108).

Proof of (7.109) Since $\mathbf{W}_{j+1}$ is the orthogonal complement of $\mathbf{V}_{j+1}$ in $\mathbf{V}_j$ the union of the two bases $\{\psi_{j+1,n}\}_{n\in\mathbf{Z}}$ and $\{\phi_{j+1,n}\}_{n\in\mathbf{Z}}$ is an orthonormal basis of $\mathbf{V}_j$. Hence any $\phi_{j,p}$ can be decomposed in this basis:

$$\begin{aligned}\phi_{j,p} &= \sum_{n=-\infty}^{+\infty} \langle \phi_{j,p}, \phi_{j+1,n} \rangle\, \phi_{j+1,n} \\ &\quad + \sum_{n=-\infty}^{+\infty} \langle \phi_{j,p}, \psi_{j+1,n} \rangle\, \psi_{j+1,n}. \end{aligned}$$

Inserting (7.111) and (7.113) yields

$$\phi_{j,p} = \sum_{n=-\infty}^{+\infty} h[p-2n]\, \phi_{j+1,n} + \sum_{n=-\infty}^{+\infty} g[p-2n]\, \psi_{j+1,n}.$$

Taking the inner product with f on both sides of this equality gives (7.109). ■

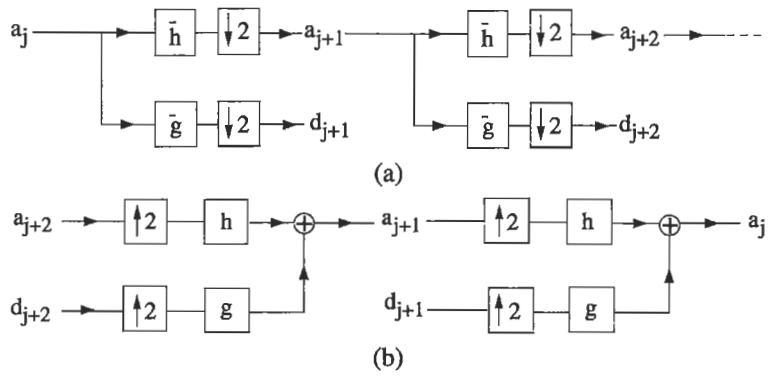


FIGURE 7.12 (a): A fast wavelet transform is computed with a cascade of filterings with $\bar{h}$ and $\bar{g}$ followed by a factor 2 subsampling. (b): A fast inverse wavelet transform reconstructs progressively each a_j by inserting zeros between samples of a_{j+1} and d_{j+1}, filtering and adding the output.

Theorem 7.7 proves that a_{j+1} and d_{j+1} are computed by taking every other sample of the convolution of a_j with $\bar{h}$ and $\bar{g}$ respectively, as illustrated by Figure 7.12. The filter $\bar{h}$ removes the higher frequencies of the inner product sequence a_j whereas $\bar{g}$ is a high-pass filter which collects the remaining highest frequencies. The reconstruction (7.109) is an interpolation that inserts zeros to expand a_{j+1} and d_{j+1} and filters these signals, as shown in Figure 7.12.

An *orthogonal wavelet representation* of $a_L = \langle f, \phi_{L,n} \rangle$ is composed of wavelet coefficients of f at scales $2^L < 2^j \leq 2^J$ plus the remaining approximation at the largest scale 2^J:

$$[\{d_j\}_{L<j\leq J}\,,\, a_J]\,. \tag{7.115}$$

It is computed from a_L by iterating (7.107) and (7.108) for $L \leq j < J$. Figure 7.7 gives a numerical example computed with the cubic spline filter of Table 7.1. The original signal a_L is recovered from this wavelet representation by iterating the reconstruction (7.109) for $J > j \geq L$.

Initialization Most often the discrete input signal $b[n]$ is obtained by a finite resolution device that averages and samples an analog input signal. For example, a CCD camera filters the light intensity by the optics and each photo-receptor averages the input light over its support. A pixel value thus measures average light intensity. If the sampling distance is N^{-1}, to define and compute the wavelet coefficients, we need to associate to $b[n]$ a function $f(t) \in \mathbf{V}_L$ approximated at the scale $2^L = N^{-1}$, and compute $a_L[n] = \langle f, \phi_{L,n} \rangle$. Problem 7.6 explains how to compute $a_L[n] = \langle f, \phi_{L,n} \rangle$ so that $b[n] = f(N^{-1}n)$.

A simpler and faster approach considers

$$f(t) = \sum_{n=-\infty}^{+\infty} b[n]\, \phi\left(\frac{t - 2^L n}{2^L}\right) \in \mathbf{V}_L.$$

Since $\{\phi_{L,n}(t) = 2^{-L/2}\phi(2^{-L}t - n)\}_{n\in\mathbb{Z}}$ is orthonormal and $2^L = N^{-1}$,

$$b[n] = N^{1/2}\langle f, \phi_{L,n}\rangle = N^{1/2} a_L[n]\ .$$

But $\hat{\phi}(0) = \int_{-\infty}^{\infty} \phi(t)\, dt = 1$, so

$$N^{1/2} a_L[n] = \int_{-\infty}^{+\infty} f(t)\, \frac{1}{N^{-1}}\, \phi\left(\frac{t - N^{-1}n}{N^{-1}}\right) dt$$

is a weighted average of f in the neighborhood of $N^{-1}n$ over a domain proportional to N^{-1}. Hence if f is regular,

$$b[n] = N^{1/2} a_L[n] \approx f(N^{-1}n)\ . \tag{7.116}$$

If ψ is a Coiflet and $f(t)$ is regular in the neighborhood of $N^{-1}n$, then (7.105) shows that $N^{-1/2} a_L[n]$ is a high order approximation of $f(N^{-1}n)$.

Finite Signals Let us consider a signal f whose support is in $[0, 1]$ and which is approximated with a uniform sampling at intervals N^{-1}. The resulting approximation a_L has $N = 2^{-L}$ samples. This is the case in Figure 7.7 with $N = 1024$. Computing the convolutions with $\bar{h}$ and $\bar{g}$ at abscissa close to 0 or close to N requires knowing the values of $a_L[n]$ beyond the boundaries $n = 0$ and $n = N - 1$. These boundary problems may be solved with one of the three approaches described in Section 7.5.

Section 7.5.1 explains the simplest algorithm, which periodizes a_L. The convolutions in Theorem 7.7 are replaced by circular convolutions. This is equivalent to decomposing f in a periodic wavelet basis of $\mathbf{L}^2[0, 1]$. This algorithm has the disadvantage of creating large wavelet coefficients at the borders.

If ψ is symmetric or antisymmetric, we can use a folding procedure described in Section 7.5.2, which creates smaller wavelet coefficients at the border. It decomposes f in a folded wavelet basis of $\mathbf{L}^2[0, 1]$. However, we mentioned in Section 7.2.3 that Haar is the only symmetric wavelet with a compact support. Higher order spline wavelets have a symmetry but h must be truncated in numerical calculations.

The most efficient boundary treatment is described in Section 7.5.3, but the implementation is more complicated. Boundary wavelets which keep their vanishing moments are designed to avoid creating large amplitude coefficients when f is regular. The fast algorithm is implemented with special boundary filters, and requires the same number of calculations as the two other methods.

Complexity Suppose that h and g have K non-zero coefficients. Let a_L be a signal of size $N = 2^{-L}$. With appropriate boundary calculations, each a_j and d_j has 2^{-j} samples. Equations (7.107) and (7.108) compute a_{j+1} and d_{j+1} from a_j with $2^{-j}K$ additions and multiplications. The wavelet representation (7.115) is therefore calculated with at most $2KN$ additions and multiplications. The reconstruction (7.109) of a_j from a_{j+1} and d_{j+1} is also obtained with $2^{-j}K$ additions and multiplications. The original signal a_L is thus also recovered from the wavelet representation with at most $2KN$ additions and multiplications.

Wavelet Graphs The graphs of ϕ and ψ are computed numerically with the inverse wavelet transform. If $f = \phi$ then $a_0[n] = \delta[n]$ and $d_j[n] = 0$ for all $L < j \leq 0$. The inverse wavelet transform computes a_L and (7.116) shows that

$$N^{1/2} a_L[n] \approx \phi(N^{-1}n) \ .$$

If ϕ is regular and N is large enough, we recover a precise approximation of the graph of ϕ from a_L.

Similarly, if $f = \psi$ then $a_0[n] = 0$, $d_0[n] = \delta[n]$ and $d_j[n] = 0$ for $L < j < 0$. Then $a_L[n]$ is calculated with the inverse wavelet transform and $N^{1/2} a_L[n] \approx \psi(N^{-1}n)$. The Daubechies wavelets and scaling functions in Figure 7.10 are calculated with this procedure.

7.3.2 Perfect Reconstruction Filter Banks

The fast discrete wavelet transform decomposes signals into low-pass and high-pass components subsampled by 2; the inverse transform performs the reconstruction. The study of such classical multirate filter banks became a major signal processing topic in 1976, when Croisier, Esteban and Galand [141] discovered that it is possible to perform such decompositions and reconstructions with *quadrature mirror filters* (Problem 7.7). However, besides the simple Haar filter, a quadrature mirror filter can not have a finite impulse response. In 1984, Smith and Barnwell [316] and Mintzer [272] found necessary and sufficient conditions for obtaining perfect reconstruction orthogonal filters with a finite impulse response, that they called *conjugate mirror filters*. The theory was completed by the biorthogonal equations of Vetterli [338, 339] and the general paraunitary matrix theory of Vaidyanathan [336]. We follow this digital signal processing approach which gives a simple understanding of conjugate mirror filter conditions. More complete presentations of filter banks properties can be found in [1, 2, 68, 73, 74].

Filter Bank A two-channel multirate filter bank convolves a signal a_0 with a low-pass filter $\bar{h}[n] = h[-n]$ and a high-pass filter $\bar{g}[n] = g[-n]$ and subsamples by 2 the output:

$$a_1[n] = a_0 \star \bar{h}[2n] \ \text{ and } \ d_1[n] = a_0 \star \bar{g}[2n]. \tag{7.117}$$

A reconstructed signal $\tilde{a}_0$ is obtained by filtering the zero expanded signals with a dual low-pass filter $\tilde{h}$ and a dual high-pass filter $\tilde{g}$, as shown in Figure 7.13. With

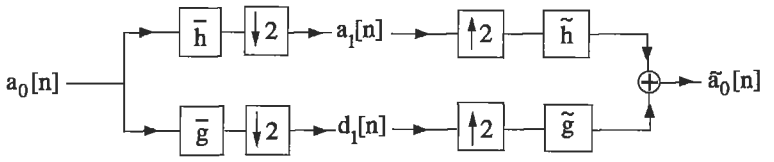


FIGURE 7.13 The input signal is filtered by a low-pass and a high-pass filter and subsampled. The reconstruction is performed by inserting zeros and filtering with dual filters $\tilde{h}$ and $\tilde{g}$.

the zero insertion notation (7.106) it yields

$$\tilde{a}_0[n] = \check{a}_1 \star \tilde{h}[n] + \check{d}_1 \star \tilde{g}[n]. \tag{7.118}$$

We study necessary and sufficient conditions on h, g, $\tilde{h}$ and $\tilde{g}$ to guarantee a perfect reconstruction $\tilde{a}_0 = a_0$.

Subsampling and Zero Interpolation Subsamplings and expansions with zero insertions have simple expressions in the Fourier domain. Since $\hat{x}(\omega) = \sum_{n=-\infty}^{+\infty} x[n]\, e^{-in\omega}$ the Fourier series of the subsampled signal $y[n] = x[2n]$ can be written

$$\hat{y}(2\omega) = \sum_{n=-\infty}^{+\infty} x[2n]\, e^{-i2n\omega} = \frac{1}{2}\left(\hat{x}(\omega) + \hat{x}(\omega+\pi)\right). \tag{7.119}$$

The component $\hat{x}(\omega+\pi)$ creates a frequency folding. This *aliasing* must be canceled at the reconstruction.

The insertion of zeros defines

$$y[n] = \check{x}[n] = \begin{cases} x[p] & \text{if } n = 2p \\ 0 & \text{if } n = 2p+1 \end{cases},$$

whose Fourier transform is

$$\hat{y}(\omega) = \sum_{n=-\infty}^{+\infty} x[n]\, e^{-i2n\omega} = \hat{x}(2\omega). \tag{7.120}$$

The following theorem gives Vetterli's [339] biorthogonal conditions, which guarantee that $\tilde{a}_0 = a_0$.

Theorem 7.8 (VETTERLI) *The filter bank performs an exact reconstruction for any input signal if and only if*

$$\hat{h}^*(\omega+\pi)\,\widehat{\tilde{h}}(\omega) + \hat{g}^*(\omega+\pi)\,\widehat{\tilde{g}}(\omega) = 0, \tag{7.121}$$

and

$$\hat{h}^*(\omega)\,\widehat{\tilde{h}}(\omega) + \hat{g}^*(\omega)\,\widehat{\tilde{g}}(\omega) = 2. \tag{7.122}$$

Proof[1]. We first relate the Fourier transform of a_1 and d_1 to the Fourier transform of a_0. Since h and g are real, the transfer functions of $\bar{h}$ and $\bar{g}$ are respectively $\hat{h}(-\omega) = \hat{h}^*(\omega)$ and $\hat{g}(-\omega) = \hat{g}^*(\omega)$. By using (7.119), we derive from the definition (7.117) of a_1 and d_1 that

$$\hat{a}_1(2\omega) = \frac{1}{2}\left(\hat{a}_0(\omega)\hat{h}^*(\omega) + \hat{a}_0(\omega+\pi)\hat{h}^*(\omega+\pi)\right), \tag{7.123}$$

$$\hat{d}_1(2\omega) = \frac{1}{2}\left(\hat{a}_0(\omega)\hat{g}^*(\omega) + \hat{a}_0(\omega+\pi)\hat{g}^*(\omega+\pi)\right). \tag{7.124}$$

The expression (7.118) of $\tilde{a}_0$ and the zero insertion property (7.120) also imply

$$\widehat{\tilde{a}}_0(\omega) = \hat{a}_1(2\omega)\widehat{\tilde{h}}(\omega) + \hat{d}_1(2\omega)\widehat{\tilde{g}}(\omega). \tag{7.125}$$

Hence

$$\widehat{\tilde{a}}_0(\omega) = \frac{1}{2}\left(\hat{h}^*(\omega)\widehat{\tilde{h}}(\omega) + \hat{g}^*(\omega)\widehat{\tilde{g}}(\omega)\right)\hat{a}_0(\omega) + \frac{1}{2}\left(\hat{h}^*(\omega+\pi)\widehat{\tilde{h}}(\omega) + \hat{g}^*(\omega+\pi)\widehat{\tilde{g}}(\omega)\right)\hat{a}_0(\omega+\pi).$$

To obtain $a_0 = \tilde{a}_0$ for all a_0, the filters must cancel the aliasing term $\hat{a}_0(\omega+\pi)$ and guarantee a unit gain for $\hat{a}_0(\omega)$, which proves equations (7.121) and (7.122). ■

Theorem 7.8 proves that the reconstruction filters $\tilde{h}$ and $\tilde{g}$ are entirely specified by the decomposition filters h and g. In matrix form, it can be rewritten

$$\begin{pmatrix} \hat{h}(\omega) & \hat{g}(\omega) \\ \hat{h}(\omega+\pi) & \hat{g}(\omega+\pi) \end{pmatrix} \times \begin{pmatrix} \widehat{\tilde{h}}^*(\omega) \\ \widehat{\tilde{g}}^*(\omega) \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}. \tag{7.126}$$

The inversion of this 2×2 matrix yields

$$\begin{pmatrix} \widehat{\tilde{h}}^*(\omega) \\ \widehat{\tilde{g}}^*(\omega) \end{pmatrix} = \frac{2}{\Delta(\omega)} \begin{pmatrix} \hat{g}(\omega+\pi) \\ -\hat{h}(\omega+\pi) \end{pmatrix} \tag{7.127}$$

where $\Delta(\omega)$ is the determinant

$$\Delta(\omega) = \hat{h}(\omega)\hat{g}(\omega+\pi) - \hat{h}(\omega+\pi)\hat{g}(\omega). \tag{7.128}$$

The reconstruction filters are stable only if the determinant does not vanish for all $\omega \in [-\pi, \pi]$. Vaidyanathan [336] has extended this result to multirate filter banks with an arbitrary number M of channels by showing that the resulting matrices of filters satisfy paraunitary properties [73].

Finite Impulse Response When all filters have a finite impulse response, the determinant $\Delta(\omega)$ can be evaluated. This yields simpler relations between the decomposition and reconstruction filters.

Theorem 7.9 *Perfect reconstruction filters satisfy*

$$\hat{h}^*(\omega)\widehat{\tilde{h}}(\omega)+\hat{h}^*(\omega+\pi)\widehat{\tilde{h}}(\omega+\pi)=2. \tag{7.129}$$

For finite impulse response filters, there exist $a \in \mathbb{R}$ *and* $l \in \mathbb{Z}$ *such that*

$$\hat{g}(\omega)=a\,\mathrm{e}^{-i(2l+1)\omega}\widehat{\tilde{h}}^*(\omega+\pi) \quad \text{and} \quad \widehat{\tilde{g}}(\omega)=a^{-1}\,\mathrm{e}^{-i(2l+1)\omega}\hat{h}^*(\omega+\pi). \tag{7.130}$$

Proof[1]. Equation (7.127) proves that

$$\widehat{\tilde{h}}^*(\omega)=\frac{2}{\Delta(\omega)}\hat{g}(\omega+\pi) \quad \text{and} \quad \widehat{\tilde{g}}^*(\omega)=\frac{-2}{\Delta(\omega)}\hat{h}(\omega+\pi). \tag{7.131}$$

Hence

$$\hat{g}(\omega)\widehat{\tilde{g}}^*(\omega)=-\frac{\Delta(\omega+\pi)}{\Delta(\omega)}\widehat{\tilde{h}}^*(\omega+\pi)\hat{h}(\omega+\pi). \tag{7.132}$$

The definition (7.128) implies that $\Delta(\omega+\pi)=-\Delta(\omega)$. Inserting (7.132) in (7.122) yields (7.129).

The Fourier transform of finite impulse response filters is a finite series in $\exp(\pm in\omega)$. The determinant $\Delta(\omega)$ defined by (7.128) is therefore a finite series. Moreover (7.131) proves that $\Delta^{-1}(\omega)$ must also be a finite series. A finite series in $\exp(\pm in\omega)$ whose inverse is also a finite series must have a single term. Since $\Delta(\omega)=-\Delta(\omega+\pi)$ the exponent n must be odd. This proves that there exist $l \in \mathbb{Z}$ and $a \in \mathbb{R}$ such that

$$\Delta(\omega)=-2a\exp[i(2l+1)\omega]. \tag{7.133}$$

Inserting this expression in (7.131) yields (7.130). ■

The factor a is a gain which is inverse for the decomposition and reconstruction filters and l is a reverse shift. We generally set $a=1$ and $l=0$. In the time domain (7.130) can then be rewritten

$$g[n]=(-1)^{1-n}\tilde{h}[1-n] \quad \text{and} \quad \tilde{g}[n]=(-1)^{1-n}h[1-n]. \tag{7.134}$$

The two pairs of filters (h,g) and $(\tilde{h},\tilde{g})$ play a symmetric role and can be inverted.

Conjugate Mirror Filters If we impose that the decomposition filter h is equal to the reconstruction filter $\tilde{h}$, then (7.129) is the condition of Smith and Barnwell [316] and Mintzer [272] that defines conjugate mirror filters:

$$|\hat{h}(\omega)|^2+|\hat{h}(\omega+\pi)|^2=2. \tag{7.135}$$

It is identical to the filter condition (7.34) that is required in order to synthesize orthogonal wavelets. The next section proves that it is also equivalent to discrete orthogonality properties.

7.3.3 Biorthogonal Bases of $l^2(\mathbb{Z})$ [2]

The decomposition of a discrete signal in a multirate filter bank is interpreted as an expansion in a basis of $l^2(\mathbb{Z})$. Observe first that the low-pass and high-pass signals of a filter bank computed with (7.117) can be rewritten as inner products in $l^2(\mathbb{Z})$:

$$a_1[l] = \sum_{n=-\infty}^{+\infty} a_0[n]\, h[n-2l] = \langle a_0[k], h[k-2n]\rangle, \tag{7.136}$$

$$d_1[l] = \sum_{n=-\infty}^{+\infty} a_0[n]\, g[n-2l] = \langle a_0[n], g[n-2l]\rangle. \tag{7.137}$$

The signal recovered by the reconstructing filters is

$$a_0[n] = \sum_{l=-\infty}^{+\infty} a_1[l]\, \tilde{h}[n-2l] + \sum_{l=-\infty}^{+\infty} d_1[l]\, \tilde{g}[n-2l]. \tag{7.138}$$

Inserting (7.136) and (7.137) yields

$$a_0[n] = \sum_{l=-\infty}^{+\infty} \langle f[k], h[k-2l]\rangle\, \tilde{h}[n-2l] + \sum_{l=-\infty}^{+\infty} \langle f[k], g[k-2l]\rangle\, \tilde{g}[n-2l]. \tag{7.139}$$

We recognize the decomposition of a_0 over dual families of vectors $\{\tilde{h}[n-2l], \tilde{g}[n-2l]\}_{l\in\mathbb{Z}}$ and $\{h[n-2l], g[n-2l]\}_{l\in\mathbb{Z}}$. The following theorem proves that these two families are biorthogonal.

Theorem 7.10 *If h, g, $\tilde{h}$ and $\tilde{g}$ are perfect reconstruction filters whose Fourier transform is bounded then $\{\tilde{h}[n-2l], \tilde{g}[n-2l]\}_{l\in\mathbb{Z}}$ and $\{h[n-2l], g[n-2l]\}_{l\in\mathbb{Z}}$ are biorthogonal Riesz bases of $l^2(\mathbb{Z})$.*

Proof [2]. To prove that these families are biorthogonal we must show that for all $n \in \mathbb{Z}$

$$\langle \tilde{h}[n], h[n-2l]\rangle = \delta[l] \tag{7.140}$$

$$\langle \tilde{g}[n], g[n-2l]\rangle = \delta[l] \tag{7.141}$$

and

$$\langle \tilde{h}[n], g[n-2l]\rangle = \langle \tilde{g}[n], h[n-2l]\rangle = 0. \tag{7.142}$$

For perfect reconstruction filters, (7.129) proves that

$$\frac{1}{2}\left(\hat{h}^*(\omega)\,\widehat{\tilde{h}}(\omega) + \hat{h}^*(\omega+\pi)\,\widehat{\tilde{h}}(\omega+\pi)\right) = 1.$$

In the time domain, this equation becomes

$$\bar{h} \star \tilde{h}[2l] = \sum_{k=-\infty}^{+\infty} \tilde{h}[n]\, \bar{h}[n-2l] = \delta[l], \tag{7.143}$$

which verifies (7.140). The same proof as for (7.129) shows that

$$\frac{1}{2}\left(\hat{g}^*(\omega)\widehat{\tilde{g}}(\omega)+\hat{g}^*(\omega+\pi)\widehat{\tilde{g}}(\omega+\pi)\right)=1.$$

In the time domain, this equation yields (7.141). It also follows from (7.127) that

$$\frac{1}{2}\left(\hat{g}^*(\omega)\widehat{\tilde{h}}(\omega)+\hat{g}^*(\omega+\pi)\widehat{\tilde{h}}(\omega+\pi)\right)=0,$$

and

$$\frac{1}{2}\left(\hat{h}^*(\omega)\widehat{\tilde{g}}(\omega)+\hat{h}^*(\omega+\pi)\widehat{\tilde{g}}(\omega+\pi)\right)=0.$$

The inverse Fourier transforms of these two equations yield (7.142).

To finish the proof, one must show the existence of Riesz bounds defined in (A.12). The reader can verify that this is a consequence of the fact that the Fourier transform of each filter is bounded. ■

Orthogonal Bases A Riesz basis is orthonormal if the dual basis is the same as the original basis. For filter banks, this means that $h=\tilde{h}$ and $g=\tilde{g}$. The filter h is then a conjugate mirror filter

$$|\hat{h}(\omega)|^2+|\hat{h}(\omega+\pi)|^2=2. \tag{7.144}$$

The resulting family $\{h[n-2l], g[n-2l]\}_{l\in\mathbb{Z}}$ is an orthogonal basis of $\mathbf{l}^2(\mathbb{Z})$.

Discrete Wavelet Bases The construction of conjugate mirror filters is simpler than the construction of orthogonal wavelet bases of $\mathbf{L}^2(\mathbb{R})$. Why then should we bother with continuous time models of wavelets, since in any case all computations are discrete and rely on conjugate mirror filters? The reason is that conjugate mirror filters are most often used in filter banks that cascade several levels of filterings and subsamplings. It is thus necessary to understand the behavior of such a cascade [290]. In a wavelet filter bank tree, the output of the low-pass filter $\bar{h}$ is subdecomposed whereas the output of the high-pass filter $\bar{g}$ is not; this is illustrated in Figure 7.12. Suppose that the sampling distance of the original discrete signal is N^{-1}. We denote $a_L[n]$ this discrete signal, with $2^L=N^{-1}$. At the depth $j-L\geq 0$ of this filter bank tree, the low-pass signal a_j and high-pass signal d_j can be written

$$a_j[l]=a_L\star\bar{\phi}_j[2^{j-L}l]=\langle a_L[n],\phi_j[n-2^{j-L}l]\rangle$$

and

$$d_j[l]=a_L\star\bar{\psi}_j[2^{j-L}l]=\langle a_L[n],\psi_j[n-2^{j-L}l]\rangle.$$

The Fourier transforms of these equivalent filters are

$$\hat{\phi}_j(\omega)=\prod_{p=0}^{j-L-1}\hat{h}(2^p\omega)\ \text{ and }\ \hat{\psi}_j(\omega)=\hat{g}(2^{j-L-1}\omega)\prod_{p=0}^{j-L-2}\hat{h}(2^p\omega). \tag{7.145}$$

A filter bank tree of depth $J-L \geq 0$, decomposes a_L over the family of vectors

$$\left[\left\{\phi_J[n-2^{J-L}l]\right\}_{l\in\mathbb{Z}}\ ,\ \left\{\psi_j[n-2^{j-L}l]\right\}_{L<j\leq J,\,l\in\mathbb{Z}}\right]. \tag{7.146}$$

For conjugate mirror filters, one can verify that this family is an orthonormal basis of $\mathbf{l}^2(\mathbb{Z})$. These discrete vectors are close to a uniform sampling of the continuous time scaling functions $\phi_j(t) = 2^{-j/2}\phi(2^{-j}t)$ and wavelets $\psi_j(t) = 2^{-j/2}\phi(2^{-j}t)$. When the number $L-j$ of successive convolutions increases, one can verify that $\phi_j[n]$ and $\psi_j[n]$ converge respectively to $N^{-1/2}\phi_j(N^{-1}n)$ and $N^{-1/2}\psi_j(N^{-1}n)$. The factor $N^{-1/2}$ normalizes the $\mathbf{l}^2(\mathbb{Z})$ norm of these sampled functions. If $L-j=4$ then $\phi_j[n]$ and $\psi_j[n]$ are already very close to these limit values. The impulse responses $\phi_j[n]$ and $\psi_j[n]$ of the filter bank are thus much closer to continuous time scaling functions and wavelets than they are to the original conjugate mirror filters h and g. This explains why wavelets provide appropriate models for understanding the applications of these filter banks. Chapter 8 relates more general filter banks to wavelet packet bases.

If the decomposition and reconstruction filters of the filter bank are different, the resulting basis (7.146) is non-orthogonal. The stability of this discrete wavelet basis does not degrade when the depth $J-L$ of the filter bank increases. The next section shows that the corresponding continuous time wavelet $\psi(t)$ generates a Riesz basis of $\mathbf{L}^2(\mathbb{R})$.

7.4 BIORTHOGONAL WAVELET BASES [2]

The stability and completeness properties of biorthogonal wavelet bases are described for perfect reconstruction filters h and $\tilde{h}$ having a finite impulse response. The design of linear phase wavelets with compact support is explained in Section 7.4.2.

7.4.1 Construction of Biorthogonal Wavelet Bases

An infinite cascade of perfect reconstruction filters (h, g) and $(\tilde{h}, \tilde{g})$ yields two scaling functions and wavelets whose Fourier transforms satisfy

$$\hat{\phi}(2\omega) = \frac{1}{\sqrt{2}}\hat{h}(\omega)\hat{\phi}(\omega)\ ,\ \widehat{\tilde{\phi}}(2\omega) = \frac{1}{\sqrt{2}}\widehat{\tilde{h}}(\omega)\widehat{\tilde{\phi}}(\omega)\ , \tag{7.147}$$

$$\hat{\psi}(2\omega) = \frac{1}{\sqrt{2}}\hat{g}(\omega)\hat{\phi}(\omega)\ ,\ \widehat{\tilde{\psi}}(2\omega) = \frac{1}{\sqrt{2}}\widehat{\tilde{g}}(\omega)\widehat{\tilde{\phi}}(\omega)\ . \tag{7.148}$$

In the time domain, these relations become

$$\phi(t) = \sqrt{2}\sum_{n=-\infty}^{+\infty} h[n]\,\phi(2t-n)\ ,\ \tilde{\phi}(t) = \sqrt{2}\sum_{n=-\infty}^{+\infty} \tilde{h}[n]\,\tilde{\phi}(2t-n) \tag{7.149}$$

$$\psi(t) = \sqrt{2}\sum_{n=-\infty}^{+\infty} g[n]\,\phi(2t-n)\ ,\ \tilde{\psi}(t) = \sqrt{2}\sum_{n=-\infty}^{+\infty} \tilde{g}[n]\,\tilde{\phi}(2t-n)\ . \tag{7.150}$$

The perfect reconstruction conditions are given by Theorem 7.9. If we normalize the gain and shift to $a = 1$ and $l = 0$, the filters must satisfy

$$\hat{h}^*(\omega)\widehat{\tilde{h}}(\omega) + \hat{h}^*(\omega+\pi)\widehat{\tilde{h}}(\omega+\pi) = 2, \tag{7.151}$$

and

$$\hat{g}(\omega) = e^{-i\omega}\widehat{\tilde{h}}^*(\omega+\pi) \ , \ \widehat{\tilde{g}}(\omega) = e^{-i\omega}\hat{h}^*(\omega+\pi). \tag{7.152}$$

Wavelets should have a zero average, which means that $\hat{\psi}(0) = \widehat{\tilde{\psi}}(0) = 0$. This is obtained by setting $\hat{g}(0) = \widehat{\tilde{g}}(0) = 0$ and hence $\hat{h}(\pi) = \widehat{\tilde{h}}(\pi) = 0$. The perfect reconstruction condition (7.151) implies that $\hat{h}^*(0)\widehat{\tilde{h}}(0) = 2$. Since both filters are defined up to multiplicative constants respectively equal to λ and λ^{-1}, we adjust λ so that $\hat{h}(0) = \widehat{\tilde{h}}(0) = \sqrt{2}$.

In the following, we also suppose that h and $\tilde{h}$ are finite impulse response filters. One can then prove [21] that

$$\hat{\phi}(\omega) = \prod_{p=1}^{+\infty} \frac{\hat{h}(2^{-p}\omega)}{\sqrt{2}} \quad \text{and} \quad \widehat{\tilde{\phi}}(\omega) = \prod_{p=1}^{+\infty} \frac{\widehat{\tilde{h}}(2^{-p}\omega)}{\sqrt{2}} \tag{7.153}$$

are the Fourier transforms of distributions of compact support. However, these distributions may exhibit wild behavior and have infinite energy. Some further conditions must be imposed to guarantee that $\hat{\phi}$ and $\widehat{\tilde{\phi}}$ are the Fourier transforms of finite energy functions. The following theorem gives sufficient conditions on the perfect reconstruction filters for synthesizing biorthogonal wavelet bases of $\mathbf{L}^2(\mathbb{R})$.

Theorem 7.11 (COHEN, DAUBECHIES, FEAUVEAU) *Suppose that there exist strictly positive trigonometric polynomials $P(e^{i\omega})$ and $\tilde{P}(e^{i\omega})$ such that*

$$\left|\hat{h}\left(\frac{\omega}{2}\right)\right|^2 P(e^{i\omega/2}) + \left|\hat{h}\left(\frac{\omega}{2}+\pi\right)\right|^2 P(e^{i(\omega/2+\pi)}) = 2P(e^{i\omega}), \tag{7.154}$$

$$\left|\widehat{\tilde{h}}\left(\frac{\omega}{2}\right)\right|^2 \tilde{P}(e^{i\omega/2}) + \left|\widehat{\tilde{h}}\left(\frac{\omega}{2}+\pi\right)\right|^2 \tilde{P}(e^{i(\omega/2+\pi)}) = 2\tilde{P}(e^{i\omega}) \tag{7.155}$$

and that P and $\tilde{P}$ are unique (up to normalization). Suppose that

$$\inf_{\omega\in[-\pi/2,\pi/2]} |\hat{h}(\omega)| > 0 \ , \quad \inf_{\omega\in[-\pi/2,\pi/2]} |\widehat{\tilde{h}}(\omega)| > 0. \tag{7.156}$$

- *Then the functions $\hat{\phi}$ and $\widehat{\tilde{\phi}}$ defined in (7.153) belong to $\mathbf{L}^2(\mathbb{R})$, and ϕ, $\tilde{\phi}$ satisfy biorthogonal relations*

$$\langle \phi(t), \tilde{\phi}(t-n)\rangle = \delta[n]. \tag{7.157}$$

- *The two wavelet families* $\{\psi_{j,n}\}_{(j,n)\in\mathbb{Z}^2}$ *and* $\{\tilde{\psi}_{j,n}\}_{(j,n)\in\mathbb{Z}^2}$ *are biorthogonal Riesz bases of* $\mathbf{L}^2(\mathbb{R})$.

The proof of this theorem is in [131] and [21]. The hypothesis (7.156) is also imposed by Theorem 7.2, which constructs orthogonal bases of scaling functions. The conditions (7.154) and (7.155) do not appear in the construction of wavelet orthogonal bases because they are always satisfied with $P(e^{i\omega}) = \tilde{P}(e^{i\omega}) = 1$ and one can prove that constants are the only invariant trigonometric polynomials [247].

Biorthogonality means that for any $(j, j', n, n') \in \mathbb{Z}^4$,

$$\langle \psi_{j,n}, \tilde{\psi}_{j',n'} \rangle = \delta[n-n']\,\delta[j-j']. \tag{7.158}$$

Any $f \in \mathbf{L}^2(\mathbb{R})$ has two possible decompositions in these bases:

$$f = \sum_{n,j=-\infty}^{+\infty} \langle f, \psi_{j,n} \rangle \tilde{\psi}_{j,n} = \sum_{n,j=-\infty}^{+\infty} \langle f, \tilde{\psi}_{j,n} \rangle \psi_{j,n}\,. \tag{7.159}$$

The Riesz stability implies that there exist $A > 0$ and $B > 0$ such that

$$A\,\|f\|^2 \;\leq\; \sum_{n,j=-\infty}^{+\infty} |\langle f, \psi_{j,n} \rangle|^2 \leq B\,\|f\|^2, \tag{7.160}$$

$$\frac{1}{B}\,\|f\|^2 \;\leq\; \sum_{n,j=-\infty}^{+\infty} |\langle f, \tilde{\psi}_{j,n} \rangle|^2 \leq \frac{1}{A}\,\|f\|^2. \tag{7.161}$$

Multiresolutions Biorthogonal wavelet bases are related to multiresolution approximations. The family $\{\phi(t-n)\}_{n\in\mathbb{Z}}$ is a Riesz basis of the space $\mathbf{V}_0$ it generates, whereas $\{\tilde{\phi}(t-n)\}_{n\in\mathbb{Z}}$ is a Riesz basis of another space $\tilde{\mathbf{V}}_0$. Let $\mathbf{V}_j$ and $\tilde{\mathbf{V}}_j$ be the spaces defined by

$$\begin{aligned} f(t) \in \mathbf{V}_j \;&\Leftrightarrow\; f(2^j t) \in \mathbf{V}_0, \\ f(t) \in \tilde{\mathbf{V}}_j \;&\Leftrightarrow\; f(2^j t) \in \tilde{\mathbf{V}}_0. \end{aligned}$$

One can verify that $\{\mathbf{V}_j\}_{j\in\mathbb{Z}}$ and $\{\tilde{\mathbf{V}}_j\}_{j\in\mathbb{Z}}$ are two multiresolution approximations of $\mathbf{L}^2(\mathbb{R})$. For any $j \in \mathbb{Z}$, $\{\phi_{j,n}\}_{n\in\mathbb{Z}}$ and $\{\tilde{\phi}_{j,n}\}_{n\in\mathbb{Z}}$ are Riesz bases of $\mathbf{V}_j$ and $\tilde{\mathbf{V}}_j$. The dilated wavelets $\{\psi_{j,n}\}_{n\in\mathbb{Z}}$ and $\{\tilde{\psi}_{j,n}\}_{n\in\mathbb{Z}}$ are bases of two detail spaces $\mathbf{W}_j$ and $\tilde{\mathbf{W}}_j$ such that

$$\mathbf{V}_j \oplus \mathbf{W}_j = \mathbf{V}_{j-1} \;\text{ and }\; \tilde{\mathbf{V}}_j \oplus \tilde{\mathbf{W}}_j = \tilde{\mathbf{V}}_{j-1}\,.$$

The biorthogonality of the decomposition and reconstruction wavelets implies that $\mathbf{W}_j$ is not orthogonal to $\mathbf{V}_j$ but is to $\tilde{\mathbf{V}}_j$ whereas $\tilde{\mathbf{W}}_j$ is not orthogonal to $\tilde{\mathbf{V}}_j$ but is to $\mathbf{V}_j$.

Fast Biorthogonal Wavelet Transform The perfect reconstruction filter bank studied in Section 7.3.2 implements a fast biorthogonal wavelet transform. For any discrete signal input $b[n]$ sampled at intervals $N^{-1} = 2^L$, there exists $f \in \mathbf{V}_L$ such that $a_L[n] = \langle f, \phi_{L,n} \rangle = N^{-1/2} b[n]$. The wavelet coefficients are computed by successive convolutions with $\bar{h}$ and $\bar{g}$. Let $a_j[n] = \langle f, \phi_{j,n} \rangle$ and $d_j[n] = \langle f, \psi_{j,n} \rangle$. As in Theorem 7.7, one can prove that

$$a_{j+1}[n] = a_j \star \bar{h}[2n] \ , \ d_{j+1}[n] = a_j \star \bar{g}[2n] \ . \tag{7.162}$$

The reconstruction is performed with the dual filters $\tilde{h}$ and $\tilde{g}$:

$$a_j[n] = \check{a}_{j+1} \star \tilde{h}[n] + \check{d}_{j+1} \star \tilde{g}[n]. \tag{7.163}$$

If a_L includes N non-zero samples, the biorthogonal wavelet representation $[\{d_j\}_{L<j\leq J}, a_J]$ is calculated with $O(N)$ operations, by iterating (7.162) for $L \leq j < J$. The reconstruction of a_L by applying (7.163) for $J > j \geq L$ requires the same number of operations.

7.4.2 Biorthogonal Wavelet Design [2]

The support size, the number of vanishing moments, the regularity and the symmetry of biorthogonal wavelets is controlled with an appropriate design of h and $\tilde{h}$.

Support If the perfect reconstruction filters h and $\tilde{h}$ have a finite impulse response then the corresponding scaling functions and wavelets also have a compact support. As in Section 7.2.1, one can show that if $h[n]$ and $\tilde{h}[n]$ are non-zero respectively for $N_1 \leq n \leq N_2$ and $\tilde{N}_1 \leq n \leq \tilde{N}_2$, then ϕ and $\tilde{\phi}$ have a support respectively equal to $[N_1, N_2]$ and $[\tilde{N}_1, \tilde{N}_2]$. Since

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

the supports of ψ and $\tilde{\psi}$ defined in (7.150) are respectively

$$\left[\frac{N_1 - \tilde{N}_2 + 1}{2}, \frac{N_2 - \tilde{N}_1 + 1}{2}\right] \ \text{ and } \ \left[\frac{\tilde{N}_1 - N_2 + 1}{2}, \frac{\tilde{N}_2 - N_1 + 1}{2}\right]. \tag{7.164}$$

Both wavelets thus have a support of the same size and equal to

$$l = \frac{N_2 - N_1 + \tilde{N}_2 - \tilde{N}_1}{2}. \tag{7.165}$$

Vanishing Moments The number of vanishing moments of ψ and $\tilde{\psi}$ depends on the number of zeros at $\omega = \pi$ of $\hat{h}(\omega)$ and $\widehat{\tilde{h}}(\omega)$. Theorem 7.4 proves that ψ has $\tilde{p}$ vanishing moments if the derivatives of its Fourier transform satisfy $\hat{\psi}^{(k)}(0) = 0$ for $k \leq \tilde{p}$. Since $\hat{\phi}(0) = 1$, (7.4.1) implies that it is equivalent to impose that $\hat{g}(\omega)$ has a zero of order $\tilde{p}$ at $\omega = 0$. Since $\hat{g}(\omega) = \mathrm{e}^{-i\omega} \widehat{\tilde{h}}^*(\omega + \pi)$, this means that $\widehat{\tilde{h}}(\omega)$ has a zero of order $\tilde{p}$ at $\omega = \pi$. Similarly the number of vanishing moments of $\tilde{\psi}$ is equal to the number p of zeros of $\hat{h}(\omega)$ at π.

Regularity Although the regularity of a function is a priori independent of the number of vanishing moments, the smoothness of biorthogonal wavelets is related to their vanishing moments. The regularity of ϕ and ψ is the same because (7.150) shows that ψ is a finite linear expansion of ϕ translated. Tchamitchian's Proposition 7.3 gives a sufficient condition for estimating this regularity. If $\hat{h}(\omega)$ has a zero of order p at π, we can perform the factorization

$$\hat{h}(\omega) = \left(\frac{1+e^{-i\omega}}{2}\right)^p \hat{l}(\omega). \tag{7.166}$$

Let $B = \sup_{\omega\in[-\pi,\pi]} |\hat{l}(\omega)|$. Proposition 7.3 proves that ϕ is uniformly Lipschitz α for

$$\alpha < \alpha_0 = p - \log_2 B - 1.$$

Generally, $\log_2 B$ increases more slowly than p. This implies that the regularity of ϕ and ψ increases with p, which is equal to the number of vanishing moments of $\tilde{\psi}$. Similarly, one can show that the regularity of $\tilde{\psi}$ and $\tilde{\phi}$ increases with $\tilde{p}$, which is the number of vanishing moments of ψ. If $\hat{h}$ and $\tilde{h}$ have different numbers of zeros at π, the properties of ψ and $\tilde{\psi}$ can therefore be very different.

Ordering of Wavelets Since ψ and $\tilde{\psi}$ might not have the same regularity and number of vanishing moments, the two reconstruction formulas

$$f = \sum_{n,j=-\infty}^{+\infty} \langle f, \psi_{j,n}\rangle \tilde{\psi}_{j,n}, \tag{7.167}$$

$$f = \sum_{n,j=-\infty}^{+\infty} \langle f, \tilde{\psi}_{j,n}\rangle \psi_{j,n} \tag{7.168}$$

are not equivalent. The decomposition (7.167) is obtained with the filters (h, g) at the decomposition and $(\tilde{h}, \tilde{g})$ at the reconstruction. The inverse formula (7.168) corresponds to $(\tilde{h}, \tilde{g})$ at the decomposition and (h, g) at the reconstruction.

To produce small wavelet coefficients in regular regions we must compute the inner products using the wavelet with the maximum number of vanishing moments. The reconstruction is then performed with the other wavelet, which is generally the smoothest one. If errors are added to the wavelet coefficients, for example with a quantization, a smooth wavelet at the reconstruction introduces a smooth error. The number of vanishing moments of ψ is equal to the number $\tilde{p}$ of zeros at π of $\widehat{\tilde{h}}$. Increasing $\tilde{p}$ also increases the regularity of $\tilde{\psi}$. It is thus better to use h at the decomposition and $\tilde{h}$ at the reconstruction if $\hat{h}$ has fewer zeros at π than $\widehat{\tilde{h}}$.

Symmetry It is possible to construct smooth biorthogonal wavelets of compact support which are either symmetric or antisymmetric. This is impossible for orthogonal wavelets, besides the particular case of the Haar basis. Symmetric or

antisymmetric wavelets are synthesized with perfect reconstruction filters having a linear phase. If h and $\tilde{h}$ have an odd number of non-zero samples and are symmetric about $n = 0$, the reader can verify that ϕ and $\tilde{\phi}$ are symmetric about $t = 0$ while ψ and $\tilde{\psi}$ are symmetric with respect to a shifted center. If h and $\tilde{h}$ have an even number of non-zero samples and are symmetric about $n = 1/2$, then $\phi(t)$ and $\tilde{\phi}(t)$ are symmetric about $t = 1/2$, while ψ and $\tilde{\psi}$ are antisymmetric with respect to a shifted center. When the wavelets are symmetric or antisymmetric, wavelet bases over finite intervals are constructed with the folding procedure of Section 7.5.2.

7.4.3 Compactly Supported Biorthogonal Wavelets [2]

We study the design of biorthogonal wavelets with a minimum size support for a specified number of vanishing moments. Symmetric or antisymmetric compactly supported spline biorthogonal wavelet bases are constructed with a technique introduced in [131].

Theorem 7.12 (COHEN, DAUBECHIES, FEAUVEAU) *Biorthogonal wavelets ψ and $\tilde{\psi}$ with respectively $\tilde{p}$ and p vanishing moments have a support of size at least $p+\tilde{p}-1$. CDF biorthogonal wavelets have a minimum support of size $p+\tilde{p}-1$.*

Proof[3]. The proof follows the same approach as the proof of Daubechies's Theorem 7.5. One can verify that p and $\tilde{p}$ must necessarily have the same parity. We concentrate on filters $h[n]$ and $\tilde{h}[n]$ that have a symmetry with respect to $n = 0$ or $n = 1/2$. The general case proceeds similarly. We can then factor

$$\hat{h}(\omega) = \sqrt{2}\exp\left(\frac{-i\epsilon\omega}{2}\right)\left(\cos\frac{\omega}{2}\right)^{p} L(\cos\omega)\ , \tag{7.169}$$

$$\hat{\tilde{h}}(\omega) = \sqrt{2}\exp\left(\frac{-i\epsilon\omega}{2}\right)\left(\cos\frac{\omega}{2}\right)^{\tilde{p}} \tilde{L}(\cos\omega)\ , \tag{7.170}$$

with $\epsilon = 0$ for p and $\tilde{p}$ even and $\epsilon = 1$ for odd values. Let $q = (p+\tilde{p})/2$. The perfect reconstruction condition

$$\hat{h}^*(\omega)\hat{\tilde{h}}(\omega) + \hat{h}^*(\omega+\pi)\hat{\tilde{h}}(\omega+\pi) = 2$$

is imposed by writing

$$L(\cos\omega)\tilde{L}(\cos\omega) = P\left(\sin^2\frac{\omega}{2}\right), \tag{7.171}$$

where the polynomial $P(y)$ must satisfy for all $y \in [0,1]$

$$(1-y)^q P(y) + y^q P(1-y) = 1. \tag{7.172}$$

We saw in (7.101) that the polynomial of minimum degree satisfying this equation is

$$P(y) = \sum_{k=0}^{q-1} \binom{q-1+k}{k} y^k. \tag{7.173}$$

The spectral factorization (7.171) is solved with a root attribution similar to (7.103). The resulting minimum support of ψ and $\tilde{\psi}$ specified by (7.165) is then $p+\tilde{p}-1$. ■

n	$p,\tilde{p}$	$h[n]$	$\tilde{h}[n]$
0		0.70710678118655	0.99436891104358
1, −1	$p=2$	0.35355339059327	0.41984465132951
2, −2	$\tilde{p}=4$		−0.17677669529664
3, −3			−0.06629126073624
4, −4			0.03314563036812
0,1		0.53033008588991	0.95164212189718
−1,2	$p=3$	0.17677669529664	−0.02649924094535
−2,3	$\tilde{p}=7$		−0.30115912592284
−3,4			0.03133297870736
−4,5			0.07466398507402
−5,6			−0.01683176542131
−6,7			−0.00906325830378
−7,8			0.00302108610126

Table 7.3 Perfect reconstruction filters h and $\tilde{h}$ for compactly supported spline wavelets, with $\hat{h}$ and $\hat{\tilde{h}}$ having respectively $\tilde{p}$ and p zeros at $\omega = \pi$.

Spline Biorthogonal Wavelets Let us choose

$$\hat{h}(\omega) = \sqrt{2} \exp\left(\frac{-i\epsilon\omega}{2}\right) \left(\cos\frac{\omega}{2}\right)^p \tag{7.174}$$

with $\epsilon = 0$ for p even and $\epsilon = 1$ for p odd. The scaling function computed with (7.153) is then a box spline of degree $p-1$

$$\hat{\phi}(\omega) = \exp\left(\frac{-i\epsilon\omega}{2}\right) \left(\frac{\sin(\omega/2)}{\omega/2}\right)^p.$$

Since ψ is a linear combination of box splines $\phi(2t-n)$, it is a compactly supported polynomial spline of same degree.

The number of vanishing moments $\tilde{p}$ of ψ is a free parameter, which must have the same parity as p. Let $q = (p+\tilde{p})/2$. The biorthogonal filter $\tilde{h}$ of minimum length is obtained by observing that $L(\cos\omega) = 1$ in (7.169). The factorization (7.171) and (7.173) thus imply that

$$\hat{\tilde{h}}(\omega) = \sqrt{2} \exp\left(\frac{-i\epsilon\omega}{2}\right) \left(\cos\frac{\omega}{2}\right)^{\tilde{p}} \sum_{k=0}^{q-1} \binom{q-1+k}{k} \left(\sin\frac{\omega}{2}\right)^{2k}. \tag{7.175}$$

These filters satisfy the conditions of Theorem 7.11 and thus generate biorthogonal wavelet bases. Table 7.3 gives the filter coefficients for $(p=2, \tilde{p}=4)$ and $(p=3, \tilde{p}=7)$. The resulting dual wavelet and scaling functions are shown in Figure 7.13.

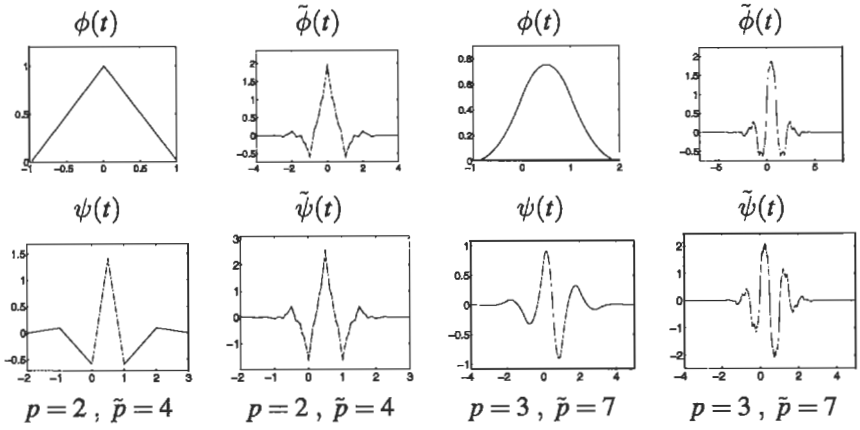


FIGURE 7.14 Spline biorthogonal wavelets and scaling functions of compact support corresponding to the filters of Table 7.3.

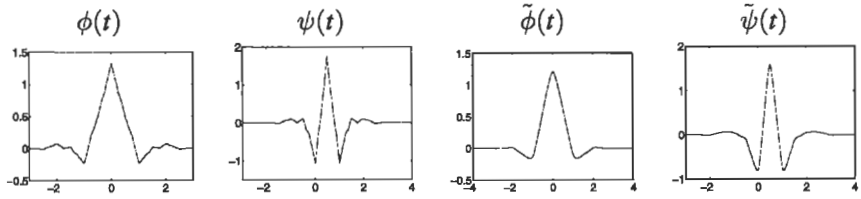


FIGURE 7.15 Biorthogonal wavelets and scaling functions calculated with the filters of Table 7.4, with $p=4$ and $\tilde{p}=4$.

Closer Filter Length Biorthogonal filters h and $\tilde{h}$ of more similar length are obtained by factoring the polynomial $P(\sin^2\frac{\omega}{2})$ in (7.171) with two polynomial $L(\cos\omega)$ and $\tilde{L}(\cos\omega)$ of similar degree. There is a limited number of possible factorizations. For $q=(p+\tilde{p})/2<4$, the only solution is $L(\cos\omega)=1$. For $q=4$ there is one non-trivial factorization and for $q=5$ there are two. Table 7.4 gives the resulting coefficients of the filters h and $\tilde{h}$ of most similar length, computed by Cohen, Daubechies and Feauveau [131]. These filters also satisfy the conditions of Theorem 7.11 and therefore define biorthogonal wavelet bases. Figure 7.15 gives the scaling functions and wavelets corresponding to $p=\tilde{p}=4$. These dual functions are similar, which indicates that this basis is nearly orthogonal. This particular set of filters is often used in image compression. The quasi-orthogonality guarantees a good numerical stability and the symmetry allows one to use the folding procedure of Section 7.5.2 at the boundaries. There are also enough vanishing moments to create small wavelet coefficients in regular image domains. How to design other compactly supported biorthogonal filters is discussed extensively in [131, 340].

$p,\tilde{p}$	n	$h[n]$	$\tilde{h}[n]$
	0	0.78848561640637	0.85269867900889
$p=4$	−1,1	0.41809227322204	0.37740285561283
$\tilde{p}=4$	−2,2	−0.04068941760920	−0.11062440441844
	−3,3	−0.06453888262876	−0.02384946501956
	−4,4	0	0.03782845554969
	0	0.89950610974865	0.73666018142821
$p=5$	−1,1	0.47680326579848	0.34560528195603
$\tilde{p}=5$	−2,2	−0.09350469740094	−0.05446378846824
	−3,3	−0.13670658466433	0.00794810863724
	−4,4	−0.00269496688011	0.03968708834741
	−5,5	0.01345670945912	0
	0	0.54113273169141	1.32702528570780
$p=5$	−1,1	0.34335173921766	0.47198693379091
$\tilde{p}=5$	−2,2	0.06115645341349	−0.36378609009851
	−3,3	0.00027989343090	−0.11843354319764
	−4,4	0.02183057133337	0.05382683783789
	−5,5	0.00992177208685	0

Table 7.4 Perfect reconstruction filters of most similar length.

7.4.4 Lifting Wavelets [3]

A lifting is an elementary modification of perfect reconstruction filters, which is used to improve the wavelet properties. It also leads to fast polyphase implementations of filter bank decompositions. The lifting scheme of Sweldens [325, 324] does not rely on the Fourier transform and can therefore construct wavelet bases over non-translation invariant domains such as bounded regions of $\mathbb{R}^p$ or surfaces. This section concentrates on the main ideas, avoiding technical details. The proofs are left to the reader.

Theorem 7.11 constructs compactly supported biorthogonal wavelet bases from finite impulse response biorthogonal filters $(h, g, \tilde{h}, \tilde{g})$ which satisfy

$$\hat{h}^*(\omega)\,\widehat{\tilde{h}}(\omega) + \hat{h}^*(\omega+\pi)\,\widehat{\tilde{h}}(\omega+\pi) = 2 \tag{7.176}$$

and

$$\hat{g}(\omega) = \mathrm{e}^{-i\omega}\,\widehat{\tilde{h}}^*(\omega+\pi) \;,\quad \widehat{\tilde{g}}(\omega) = \mathrm{e}^{-i\omega}\,\hat{h}^*(\omega+\pi). \tag{7.177}$$

The filters $\tilde{h}$ and h are said to be *dual*. The following proposition [209] characterizes all filters of compact support that are dual to $\tilde{h}$.

Proposition 7.5 (HERLEY, VETTERLI) *Let h and $\tilde{h}$ be dual filters with a finite support. A filter h^l with finite support is dual to $\tilde{h}$ if and only if there exists a finite filter l such that*

$$\hat{h}^l(\omega) = \hat{h}(\omega) + \mathrm{e}^{-i\omega}\,\widehat{\tilde{h}}^*(\omega+\pi)\,\hat{l}^*(2\omega). \tag{7.178}$$

This proposition proves that if $(h, g, \tilde{h}, \tilde{g})$ are biorthogonal then we can construct a new set of biorthogonal filters $(h^l, g, \tilde{h}, \tilde{g}^l)$ with

$$\hat{h}^l(\omega) = \hat{h}(\omega) + \hat{g}(\omega)\hat{l}^*(2\omega) \tag{7.179}$$

$$\widehat{\tilde{g}}^l(\omega) = e^{-i\omega}\hat{h}^{l*}(\omega+\pi) = \widehat{\tilde{g}}(\omega) - \widehat{\tilde{h}}(\omega)\hat{l}(2\omega). \tag{7.180}$$

This is verified by inserting (7.177) in (7.178). The new filters are said to be *lifted* because the use of l can improve their properties.

The inverse Fourier transform of (7.179) and (7.180) gives

$$h^l[n] = h[n] + \sum_{k=-\infty}^{+\infty} g[n-2k]\, l[-k], \tag{7.181}$$

$$\tilde{g}^l[n] = \tilde{g}[n] - \sum_{k=-\infty}^{+\infty} \tilde{h}[n-2k]\, l[k]. \tag{7.182}$$

Theorem 7.10 proves that the conditions (7.176) and (7.177) are equivalent to the fact that $\{h[n-2k], g[n-2k]\}_{k\in\mathbb{Z}}$ and $\{\tilde{h}[n-2k], \tilde{g}[n-2k]\}_{k\in\mathbb{Z}}$ are biorthogonal Riesz bases of $\mathbf{l}^2(\mathbb{Z})$. The lifting scheme thus creates new families $\{h^l[n-2k], g[n-2k]\}_{k\in\mathbb{Z}}$ and $\{\tilde{h}[n-2k], \tilde{g}^l[n-2k]\}_{k\in\mathbb{Z}}$ that are also biorthogonal Riesz bases of $\mathbf{l}^2(\mathbb{Z})$. The following theorem derives new biorthogonal wavelet bases by inserting (7.181) and (7.182) in the scaling equations (7.149) and (7.150).

Theorem 7.13 (SWELDENS) *Let $(\phi, \psi, \tilde{\phi}, \tilde{\psi})$ be a family of compactly supported biorthogonal scaling functions and wavelets associated to the filters $(h, g, \tilde{h}, \tilde{g})$. Let $l[k]$ be a finite sequence. A new family of formally biorthogonal scaling functions and wavelets $(\phi^l, \psi^l, \tilde{\phi}, \tilde{\psi}^l)$ is defined by*

$$\phi^l(t) = \sqrt{2}\sum_{k=-\infty}^{+\infty} h[k]\,\phi^l(2t-k) + \sum_{k=-\infty}^{+\infty} l[-k]\,\psi^l(t-k) \tag{7.183}$$

$$\psi^l(t) = \sqrt{2}\sum_{k=-\infty}^{+\infty} g[k]\,\phi^l(2t-k) \tag{7.184}$$

$$\tilde{\psi}^l(t) = \tilde{\psi}(t) - \sum_{k=-\infty}^{+\infty} l[k]\,\tilde{\phi}(t-k). \tag{7.185}$$

Theorem 7.11 imposes that the new filter h^l should satisfy (7.154) and (7.156) to generate functions ϕ^l and ψ^l of finite energy. This is not necessarily the case for all l, which is why the biorthogonality should be understood in a formal sense. If these functions have a finite energy then $\{\psi^l_{j,n}\}_{(j,n)\in\mathbb{Z}^2}$ and $\{\tilde{\psi}^l_{j,n}\}_{(j,n)\in\mathbb{Z}^2}$ are biorthogonal wavelet bases of $\mathbf{L}^2(\mathbb{R})$.

The lifting increases the support size of ψ and $\tilde{\psi}$ typically by the length of the support of l. Design procedures compute minimum size filters l to achieve specific

properties. Section 7.4.2 explains that the regularity of ϕ and ψ and the number of vanishing moments of $\tilde{\psi}$ depend on the number of zeros of $\hat{h}(\omega)$ at $\omega = \pi$, which is also equal to the number of zeros of $\widehat{\tilde{g}}(\omega)$ at $\omega = 0$. The coefficients $l[n]$ are often calculated to produce a lifted transfer function $\widehat{\tilde{g}}^{\,l}(\omega)$ with more zeros at $\omega = 0$.

To increase the number of vanishing moment of ψ and the regularity of $\tilde{\phi}$ and $\tilde{\psi}$ we use a dual lifting which modifies $\tilde{h}$ and hence g instead of h and $\tilde{g}$. The corresponding lifting formula with a filter $L[k]$ are obtained by inverting h with g and g with $\tilde{g}$ in (7.181) and (7.182):

$$g^L[n] = g[n] + \sum_{k=-\infty}^{+\infty} h[n-2k]\, L[-k], \tag{7.186}$$

$$\tilde{h}^L[n] = \tilde{h}[n] - \sum_{k=-\infty}^{+\infty} \tilde{g}[n-2k]\, L[k]. \tag{7.187}$$

The resulting family of biorthogonal scaling functions and wavelets $(\phi, \psi^L, \tilde{\phi}^L, \tilde{\psi}^L)$ are obtained by inserting these equations in the scaling equations (7.149) and (7.150):

$$\tilde{\phi}^L(t) = \sqrt{2} \sum_{k=-\infty}^{+\infty} \tilde{h}[k]\, \tilde{\phi}^L(2t-k) - \sum_{k=-\infty}^{+\infty} L[k]\, \tilde{\psi}^L(t-k) \tag{7.188}$$

$$\tilde{\psi}^L(t) = \sqrt{2} \sum_{k=-\infty}^{+\infty} \tilde{g}[k]\, \tilde{\phi}^L(2t-k) \tag{7.189}$$

$$\psi^L(t) = \psi(t) + \sum_{k=-\infty}^{+\infty} L[-k]\, \phi(t-k). \tag{7.190}$$

Successive iterations of liftings and dual liftings can improve the regularity and vanishing moments of both ψ and $\tilde{\psi}$ by increasing the number of zeros of $\hat{g}(\omega)$ and $\widehat{\tilde{g}}(\omega)$ at $\omega = 0$.

Lazy Wavelets *Lazy filters* $\tilde{h}[n] = h[n] = \delta[n]$ and $\tilde{g}[n] = g[n] = \delta[n-1]$ satisfy the biorthogonality conditions (7.176) and (7.177). Their Fourier transform is

$$\widehat{\tilde{h}}(\omega) = \hat{h}(\omega) = 1 \text{ and } \widehat{\tilde{g}}(\omega) = \hat{g}(\omega) = \mathrm{e}^{-i\omega}. \tag{7.191}$$

The resulting filter bank just separates the even and odd samples of a signal without filtering. This is also called a *polyphase* decomposition [73]. The *lazy* scaling functions and wavelets associated to these filters are Diracs $\tilde{\phi}(t) = \phi(t) = \delta(t)$ and $\tilde{\psi}(t) = \psi(t) = \delta(t-1/2)$. They do not belong to $\mathbf{L}^2(\mathbb{R})$ because $\widehat{\tilde{g}}(\omega)$ and $\hat{g}(\omega)$ do not vanish at $\omega = 0$. These wavelet can be transformed into finite energy functions by appropriate liftings.

Example 7.11 A lifting of a lazy filter $\widehat{\tilde{g}}(\omega) = e^{-i\omega}$ yields

$$\widehat{\tilde{g}}^{\,l}(\omega) = e^{-i\omega} - \hat{l}(2\omega).$$

To produce a symmetric wavelet $e^{i\omega}\hat{l}(2\omega)$ must be even. For example, to create 4 vanishing moments a simple calculation shows that the shortest filter l has a Fourier transform

$$\hat{l}(2\omega) = e^{-i\omega}\left(\frac{9}{8}\cos\omega - \frac{1}{8}\cos 3\omega\right).$$

Inserting this in (7.178) gives

$$\hat{h}^l(\omega) = -\frac{1}{16}e^{-3i\omega} + \frac{9}{16}e^{-i\omega} + 1 + \frac{9}{16}e^{i\omega} - \frac{1}{16}e^{3i\omega}. \tag{7.192}$$

The resulting ϕ^l is the Deslauriers-Dubuc interpolating scaling function of order 4 shown in Figure 7.21(b), and $\psi^l(t) = \sqrt{2}\,\phi^l(2t-1)$. These interpolating scaling functions and wavelets are further studied in Section 7.6.2. Both ϕ^l and ψ^l are continuously differentiable but $\tilde{\phi}$ and $\tilde{\psi}^l$ are sums of Diracs. A dual lifting can transform these into finite energy functions by creating a lifted filter $\hat{g}^l(\omega)$ with one or more zero at $\omega = 0$.

The following theorem proves that lifting lazy wavelets is a general filter design procedure. A constructive proof is based on the Euclidean algorithm [148].

Theorem 7.14 (DAUBECHIES, SWELDENS) *Any biorthogonal filters $(h, g, \tilde{h}, \tilde{g})$ can be synthesized with a succession of liftings and dual liftings applied to the lazy filters (7.191), up to shifting and multiplicative constants.*

Fast Polyphase Transform After lifting, the biorthogonal wavelet transform is calculated with a simple modification of the original wavelet transform. This implementation requires less calculation than a direct filter bank implementation of the lifted wavelet transform. We denote $a^l_j[k] = \langle f, \phi^l_{j,k}\rangle$ and $d^l_j[k] = \langle f, \psi^l_{j,k}\rangle$.

The standard filter bank decomposition with $(h^l, \tilde{h}, g, \tilde{g}^l)$ computes

$$a^l_{j+1}[k] = \sum_{n=-\infty}^{+\infty} h^l[n-2k]\,a^l_j[n] = a^l_j \star \bar{h}^l[2k], \tag{7.193}$$

$$d^l_{j+1}[k] = \sum_{n=-\infty}^{+\infty} g[n-2k]\,a^l_j[n] = a^l_j \star \bar{g}[2k]. \tag{7.194}$$

The reconstruction is obtained with

$$a^l_j[n] = \sum_{k=-\infty}^{+\infty} \tilde{h}[n-2k]\,a^l_{j+1}[k] + \sum_{k=-\infty}^{+\infty} \tilde{g}^l[n-2k]\,d^l_{j+1}[k]. \tag{7.195}$$

Inserting the lifting formulas (7.181) and (7.182) in (7.193) gives an expression that depends only on the original filter h:

$$a^0_{j+1}[k] = \sum_{n=-\infty}^{+\infty} h[n-2k]\, a^l_j[n] = a^l_j \star \bar{h}[2k]$$

plus a lifting component that is a convolution with l

$$a^l_{j+1}[k] = a^0_{j+1}[k] + \sum_{n=-\infty}^{+\infty} l[k-n]\, d^l_{j+1}[n] = a^0_{j+1}[k] + d^l_{j+1} \star l[k].$$

This operation is simply inverted by calculating

$$a^0_{j+1}[k] = a^l_{j+1}[k] - d^l_{j+1} \star l[k]$$

and performing a reconstruction with the original filters $(\tilde{h}, \tilde{g})$

$$a^l_j[n] = \sum_{k=-\infty}^{+\infty} \tilde{h}[n-2k]\, a^0_j[k] + \sum_{k=-\infty}^{+\infty} \tilde{g}[n-2k]\, d^l_j[k].$$

Figure 7.16 illustrates this decomposition and reconstruction. It also includes the implementation of a dual lifting with L, which is calculated with (7.186):

$$d^L_{j+1}[k] = d^l_{j+1}[k] + a^l_{j+1} \star L[k]\ .$$

Theorem 7.14 proves that any biorthogonal family of filters can be calculated with a succession of liftings and dual liftings applied to lazy filters. In this case, the filters $\bar{h}[n] = \tilde{h}[n] = \delta[n]$ can be removed whereas $\bar{g}[n] = \delta[n+1]$ and $\tilde{g}[n] = \delta[n-1]$ shift signals by 1 sample in opposite directions. The filter bank convolution and subsampling is thus directly calculated with a succession of liftings and dual liftings on the polyphase components of the signal (odd and even samples) [73]. One can verify that this implementation divides the number of operations by up to a factor 2 [148], compared to direct convolutions and subsamplings calculated in (7.193) and (7.194).

Lifted Wavelets on Arbitrary Domains The lifting procedure is extended to signal spaces which are not translation invariant. Wavelet bases and filter banks are designed for signals defined on arbitrary domains D of $\mathbb{R}^p$ or on surfaces such as a spheres.

Wavelet bases of $\mathbf{L}^2(D)$ are derived from a family of embedded vector spaces $\{\mathbf{V}_j\}_{j\in\mathbb{Z}}$ that satisfy similar multiresolution properties as in Definition 7.1. These spaces are constructed from embedded sampling grids $\{\mathcal{G}_j\}_{j\in\mathbb{Z}}$ included in D. For each index j, $\mathcal{G}_j$ has nodes whose distance to all its neighbors is of the order of 2^j. Since $\mathcal{G}_{j+1}$ is included in $\mathcal{G}_j$ we can define a complementary grid C_{j+1} that

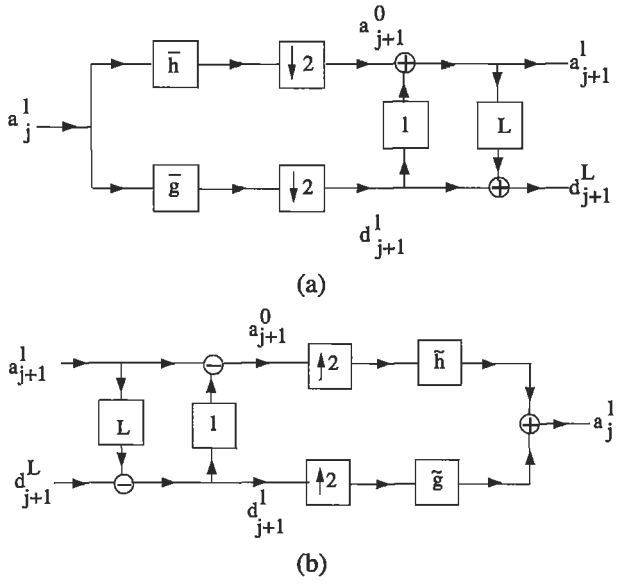


FIGURE 7.16 (a): A lifting and a dual lifting are implemented by modifying the original filter bank with two lifting convolutions, where l and L are respectively the lifting and dual lifting sequences. (b): The inverse lifted transform removes the lifting components before calculating the filter bank reconstruction.

regroups all nodes of $\mathcal{G}_j$ that are not in $\mathcal{G}_{j+1}$. For example, if $D = [0,N]$ then $\mathcal{G}_j$ is the uniform grid $\{2^j n\}_{0 \leq n \leq 2^{-j}N}$. The complementary grid $\mathcal{C}_{j+1}$ corresponds to $\{2^j(2n+1)\}_{0 \leq n < 2^{-j-1}N}$. In two dimensions, the sampling grid $\mathcal{G}_j$ can be defined as the nodes of a regular triangulation of D. This triangulation is progressively refined with a midpoint subdivision illustrated in Figure 7.17. Such embedded grids can also be constructed on surfaces [325].

Suppose that $\{h_{j,k}\}_{k \in \mathcal{G}_{j+1}} \cup \{g_{j,m}\}_{m \in \mathcal{C}_{j+1}}$ is a basis of the space $\mathbf{l}^2(\mathcal{G}_j)$ of finite energy signals defined over $\mathcal{G}_j$. Any $a_j \in \mathbf{l}^2(\mathcal{G}_j)$ is decomposed into two signals defined respectively over $\mathcal{G}_{j+1}$ and $\mathcal{C}_{j+1}$ by

$$\forall k \in \mathcal{G}_{j+1} \ , \ a_{j+1}[k] = \langle a_j, h_{j,k} \rangle = \sum_{n \in \mathcal{G}_j} a_j[n]\, h_{j,k}[n], \tag{7.196}$$

$$\forall m \in \mathcal{C}_{j+1} \ , \ d_{j+1}[m] = \langle a_j, g_{j,m} \rangle = \sum_{n \in \mathcal{G}_j} d_j[n]\, g_{j,m}[n]. \tag{7.197}$$

This decomposition is implemented by linear operators on subsampled grids as in the filter banks previously studied. However, these operators are not convolutions because the basis $\{h_{j,k}\}_{k \in \mathcal{G}_{j+1}} \cup \{g_{j,m}\}_{m \in \mathcal{C}_{j+1}}$ is not translation invariant. The

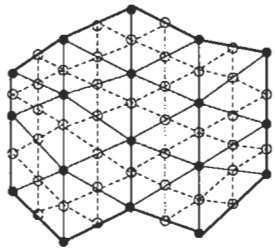


FIGURE 7.17 Black dots are the nodes of a triangulation grid $\mathcal{G}_{j+1}$ of a polygon domain D. This grid is refined with a subdivision, which adds a complementary grid $\mathcal{C}_{j+1}$ composed of all midpoints indicated with white circles. The finer grid is $\mathcal{G}_j = \mathcal{G}_{j+1} \cup \mathcal{C}_{j+1}$.

reconstruction is performed with a biorthogonal basis $\{\tilde{h}_{j,k}\}_{k\in\mathcal{G}_{j+1}} \cup \{\tilde{g}_{j,m}\}_{m\in\mathcal{C}_{j+1}}$:

$$a_j[n] = \sum_{k\in\mathcal{G}_{j+1}} a_{j+1}[k]\,\tilde{h}_{j,k}[n] + \sum_{m\in\mathcal{C}_{j+1}} d_{j+1}[m]\,\tilde{g}_{j,m}[n].$$

Scaling functions and wavelets are obtained by cascading filter bank reconstructions over progressively finer scales. As a result, they satisfy scaling equations similar to (7.112) and (7.114)

$$\phi_{j+1,k} = \sum_{n\in\mathcal{G}_j} h_{j,k}[n]\,\phi_{j,n}\ ,\ \psi_{j+1,m} = \sum_{n\in\mathcal{G}_j} g_{j,m}[n]\,\phi_{j,n}\ , \tag{7.198}$$

$$\tilde{\phi}_{j+1,k} = \sum_{n\in\mathcal{G}_j} \tilde{h}_{j,k}[n]\,\phi_{j,n}\ ,\ \tilde{\psi}_{j+1,m} = \sum_{n\in\mathcal{G}_j} \tilde{g}_{j,m}[n]\,\phi_{j,n}\ . \tag{7.199}$$

These wavelets and scaling functions have a support included in D. If they have a finite energy with respect to an appropriate measure $d\mu$ defined over D then one can verify that for any $J \le \log_2 N$

$$\left[\{\phi_{J,k}\}_{k\in\mathcal{G}_J}, \{\psi_{j,m}\}_{m\in\mathcal{C}_j, j\ge J}\right] \text{ and } \left[\{\tilde{\phi}_{J,k}\}_{k\in\mathcal{G}_J}, \{\tilde{\psi}_{j,m}\}_{m\in\mathcal{C}_j, j\ge J}\right]$$

are biorthogonal bases of $\mathbf{L}^2(D, d\mu)$.

The discrete lazy basis of $\mathbf{l}^2(\mathcal{G}_j)$ is composed of Diracs $h_{j,k}[n] = \delta[n-k]$ for $(k,n) \in \mathcal{G}_{j+1} \times \mathcal{G}_j$ and $g_{j,m}[n] = \delta[n-k]$ for $(k,n) \in \mathcal{C}_{j+1} \times \mathcal{G}_j$. This basis is clearly orthonormal so the dual basis is also the lazy basis. The resulting filter bank just separates samples of $\mathcal{G}_j$ into two sets of samples that belong respectively to $\mathcal{G}_{j+1}$ and $\mathcal{C}_{j+1}$. The corresponding scaling functions and wavelets are Diracs located over these sampling grids. Finite energy wavelets and scaling functions are constructed by lifting the discrete lazy basis.

Theorem 7.15 (SWELDENS) *Suppose that* $\{h_{j,k}\}_{k\in\mathcal{G}_{j+1}} \cup \{g_{j,m}\}_{m\in\mathcal{C}_{j+1}}$ *and* $\{\tilde{h}_{j,k}\}_{k\in\mathcal{G}_{j+1}} \cup \{\tilde{g}_{j,m}\}_{m\in\mathcal{C}_{j+1}}$ *are biorthogonal Riesz bases of* $\mathbf{l}^2(\mathcal{G}_j)$. *Let* $l_j[k,m]$ *be a matrix with a finite number of non-zero values. If*

$$\forall k \in \mathcal{G}_{j+1} \ , \ h^l_{j,k} = h_{j,k} + \sum_{m\in\mathcal{C}_{j+1}} l_j[k,m]\, g_{m,j} \tag{7.200}$$

$$\forall m \in \mathcal{C}_{j+1} \ , \ \tilde{g}^l_{j,m} = \tilde{g}_{j,m} - \sum_{k\in\mathcal{G}_{j+1}} l_j[k,m]\, \tilde{h}_{k,j} \tag{7.201}$$

then $\{h^l_{j,k}\}_{k\in\mathcal{G}_{j+1}} \cup \{g_{j,m}\}_{m\in\mathcal{C}_{j+1}}$ *and* $\{\tilde{h}_{j,k}\}_{k\in\mathcal{G}_{j+1}} \cup \{\tilde{g}^l_{j,m}\}_{m\in\mathcal{C}_{j+1}}$ *are biorthogonal Riesz bases of* $\mathbf{l}^2(\mathcal{G}_j)$.

These formulas generalize the translation invariant lifting (7.181) and (7.182), which corresponds to $l_j[k,m] = l[k-m]$. In the general case, at each scale 2^j, the lifting matrix $l_j[k,m]$ can be chosen arbitrarily. The lifted bases generate new scaling functions and wavelets that are related to the original scaling functions and wavelets by inserting (7.200) and (7.201) in the scaling equations (7.198) and (7.199) calculated with lifted filters:

$$\begin{aligned}
\phi^l_{j+1,k} &= \sum_{n\in\mathcal{G}_j} h_{j,k}[n]\, \phi_{j,n} + \sum_{m\in\mathcal{C}_{j+1}} l_j[k,m]\, \psi_{j+1,m} \\
\psi^l_{j+1,m} &= \sum_{n\in\mathcal{G}_j} g_{j,m}[n]\, \phi^l_{j,n} \\
\tilde{\psi}^l_{j+1,m} &= \tilde{\psi}_{j+1,m} - \sum_{k\in\mathcal{G}_{j+1}} l_j[k,m]\, \tilde{\phi}_{j+1,k}.
\end{aligned}$$

The dual scaling functions $\tilde{\phi}_{j,k}$ are not modified since $\tilde{h}_{j,k}$ is not changed by the lifting.

The fast decomposition algorithm in this lifted wavelet basis is calculated with the same approach as in the translation invariant case previously studied. However, the lifting blocks illustrated in Figure 7.16 are not convolutions anymore. They are linear operators computed with the matrices $l_j[k,m]$, which depend upon the scale 2^j.

To create wavelets $\tilde{\psi}_{j,m}$ with vanishing moments, we ensure that they are orthogonal to a basis of polynomials $\{p_i\}_i$ of degree smaller than q. The coefficients $l[k,m]$ are calculated by solving the linear system for all i and $m \in C_{j+1}$

$$\langle \tilde{\psi}^l_{j+1,m}, p_i \rangle = \langle \psi^l_{j+1,m}, p_i \rangle - \sum_{k\in\mathcal{G}_{j+1}} l_j[k,m]\, \langle \tilde{\phi}^l_{j+1,k}, p_i \rangle = 0.$$

A dual lifting is calculated by modifying $\tilde{h}_{j,k}$ and $g_{j,m}$ instead of $h_{j,k}$ and $\tilde{g}_{j,m}$. It allows one to change $\tilde{\phi}_{j,k}$.

Applications Lifting lazy wavelets is a simple way to construct biorthogonal wavelet bases of $\mathbf{L}^2[0,1]$. One may use a translation invariant lifting, which is modified near the left and right borders to construct filters whose supports remains inside $D=[0,1]$. The lifting coefficients are calculated to design regular wavelets with vanishing moments [325]. Section 7.5 studies other ways to construct orthogonal wavelet bases of $\mathbf{L}^2[0,1]$.

Biorthogonal wavelet bases on manifolds or bounded domains of $\mathbb{R}^p$ are calculated by lifting lazy wavelets constructed on embedded sampling grids. Lifted wavelets on the sphere have applications in computer graphics [326]. In finite two-dimensional domains, lifted wavelet bases are used for numerical calculations of partial differential equations [118].

To optimize the approximation of signals with few wavelet coefficients, one can also construct adaptive wavelet bases with liftings that depend on the signal. Short wavelets are needed in the neighborhood of singularities, but longer wavelets with more vanishing moments can improve the approximation in regions where the signal is more regular. Such a basis can be calculated with a time varying lifting whose coefficients $l_j[k,m]$ are adapted to the local signal properties [325].

7.5 WAVELET BASES ON AN INTERVAL [2]

To decompose signals f defined over an interval $[0,1]$, it is necessary to construct wavelet bases of $\mathbf{L}^2[0,1]$. Such bases are synthesized by modifying the wavelets $\psi_{j,n}(t)=2^{-j/2}\psi(2^{-j}t-n)$ of a basis $\{\psi_{j,n}\}_{(j,n)\in\mathbb{Z}^2}$ of $\mathbf{L}^2(\mathbb{R})$. The *inside* wavelets $\psi_{j,n}$ whose support are included in $[0,1]$ are not modified. The *boundary* wavelets $\psi_{j,n}$ whose supports overlap $t=0$ or $t=1$ are transformed into functions having a support in $[0,1]$, which are designed in order to provide the necessary complement to generate a basis of $\mathbf{L}^2[0,1]$. If ψ has a compact support then there is a constant number of boundary wavelets at each scale.

The main difficulty is to construct boundary wavelets that keep their vanishing moments. The next three sections describe different approaches to constructing boundary wavelets. Periodic wavelets have no vanishing moments at the boundary, whereas folded wavelets have one vanishing moment. The custom-designed boundary wavelets of Section 7.5.3 have as many vanishing moments as the inside wavelets but are more complicated to construct. Scaling functions $\phi_{j,n}$ are also restricted to $[0,1]$ by modifying the scaling functions $\phi_{j,n}(t)=2^{-j/2}\phi(2^{-j}t-n)$ associated to the wavelets $\psi_{j,n}$. The resulting wavelet basis of $\mathbf{L}^2[0,1]$ is composed of 2^{-J} scaling functions at a coarse scale $2^J<1$, plus 2^{-j} wavelets at each scale $2^j\leq 2^J$:

$$\left[\{\phi^{\text{int}}_{J,n}\}_{0\leq n<2^{-J}}\ ,\ \{\psi^{\text{int}}_{j,n}\}_{-\infty<j\leq J,0\leq n<2^{-j}}\right]\ . \tag{7.202}$$

On any interval $[a,b]$, a wavelet orthonormal basis of $\mathbf{L}^2[a,b]$ is constructed with a dilation by $b-a$ and a translation by a of the wavelets in (7.202).

Discrete Basis of $\mathbb{C}^N$ The decomposition of a signal in a wavelet basis over an interval is computed by modifying the fast wavelet transform algorithm of Section

7.3.1. A discrete signal $b[n]$ of N samples is associated to the approximation of a signal $f \in \mathbf{L}^2[0,1]$ at a scale $N^{-1} = 2^L$ with (7.116):

$$N^{-1/2} b[n] = a_L[n] = \langle f, \phi^{\text{int}}_{L,n} \rangle \text{ for } 0 \leq n < 2^{-L} .$$

Its wavelet coefficients can be calculated at scales $1 \geq 2^j > 2^L$. We set

$$a_j[n] = \langle f, \phi^{\text{int}}_{j,n} \rangle \text{ and } d_j[n] = \langle f, \psi^{\text{int}}_{j,n} \rangle \text{ for } 0 \leq n < 2^{-j} . \tag{7.203}$$

The wavelets and scaling functions with support inside $[0,1]$ are identical to the wavelets and scaling functions of a basis of $\mathbf{L}^2(\mathbb{R})$. The corresponding coefficients $a_j[n]$ and $d_j[n]$ can thus be calculated with the decomposition and reconstruction equations given by Theorem 7.7. These convolution formulas must however be modified near the boundary where the wavelets and scaling functions are modified. Boundary calculations depend on the specific design of the boundary wavelets, as explained in the next three sections. The resulting filter bank algorithm still computes the N coefficients of the wavelet representation $[a_J , \{d_j\}_{L<j\leq J}]$ of a_L with $O(N)$ operations.

Wavelet coefficients can also be written as discrete inner products of a_L with discrete wavelets:

$$a_j[n] = \langle a_L[m], \phi^{\text{int}}_{j,n}[m] \rangle \text{ and } d_j[n] = \langle a_L[m], \psi^{\text{int}}_{j,n}[m] \rangle . \tag{7.204}$$

As in Section 7.3.3, we verify that

$$\left[\{\phi^{\text{int}}_{J,n}[m]\}_{0\leq n<2^{-J}} , \{\psi^{\text{int}}_{j,n}[m]\}_{L<j\leq J, 0\leq n<2^{-j}} \right]$$

is an orthonormal basis of $\mathbb{C}^N$.

7.5.1 Periodic Wavelets

A wavelet basis $\{\psi_{j,n}\}_{(j,n)\in\mathbb{Z}^2}$ of $\mathbf{L}^2(\mathbb{R})$ is transformed into a wavelet basis of $\mathbf{L}^2[0,1]$ by periodizing each $\psi_{j,n}$. The periodization of $f \in \mathbf{L}^2(\mathbb{R})$ over $[0,1]$ is defined by

$$f^{\text{per}}(t) = \sum_{k=-\infty}^{+\infty} f(t+k). \tag{7.205}$$

The resulting periodic wavelets are

$$\psi^{\text{per}}_{j,n}(t) = \frac{1}{\sqrt{2^j}} \sum_{k=-\infty}^{+\infty} \psi\left(\frac{t - 2^j n + k}{2^j} \right).$$

For $j \leq 0$, there are 2^{-j} different $\psi^{\text{per}}_{j,n}$ indexed by $0 \leq n < 2^{-j}$. If the support of $\psi_{j,n}$ is included in $[0,1]$ then $\psi^{\text{per}}_{j,n}(t) = \psi_{j,n}(t)$ for $t \in [0,1]$. The restriction to $[0,1]$ of this periodization thus modifies only the boundary wavelets whose supports overlap $t = 0$ or $t = 1$. As indicated in Figure 7.18, such wavelets are

FIGURE 7.18 The restriction to $[0,1]$ of a periodic wavelet $\psi_{j,n}^{\text{per}}$ has two disjoint components near $t=0$ and $t=1$.

transformed into boundary wavelets which have two disjoint components near $t=0$ and $t=1$. Taken separately, the components near $t=0$ and $t=1$ of these boundary wavelets have no vanishing moments, and thus create large signal coefficients, as we shall see later. The following theorem proves that periodic wavelets together with periodized scaling functions $\phi_{j,n}^{\text{per}}$ generate an orthogonal basis of $\mathbf{L}^2[0,1]$.

Theorem 7.16 *For any* $J \leq 0$

$$\left[\{\psi_{j,n}^{\text{per}}\}_{-\infty<j\leq J, 0\leq n<2^{-j}}\ ,\ \{\phi_{J,n}^{\text{per}}\}_{0\leq n<2^{-J}}\right] \tag{7.206}$$

is an orthogonal basis of $\mathbf{L}^2[0,1]$.

Proof[2]. The orthogonality of this family is proved with the following lemma.

Lemma 7.2 *Let* $\alpha(t), \beta(t) \in \mathbf{L}^2(\mathbb{R})$. *If* $\langle \alpha(t), \beta(t+k)\rangle = 0$ *for all* $k \in \mathbb{Z}$ *then*

$$\int_0^1 \alpha^{\text{per}}(t)\,\beta^{\text{per}}(t)\,dt = 0. \tag{7.207}$$

To verify (7.207) we insert the definition (7.205) of periodized functions:

$$\begin{aligned}\int_0^1 \alpha^{\text{per}}(t)\,\beta^{\text{per}}(t)\,dt &= \int_{-\infty}^{+\infty} \alpha(t)\,\beta^{\text{per}}(t)\,dt \\ &= \sum_{k=-\infty}^{+\infty} \int_{-\infty}^{+\infty} \alpha(t)\,\beta(t+k)\,dt = 0.\end{aligned}$$

Since $[\{\psi_{j,n}\}_{-\infty<j\leq J, n\in\mathbf{Z}}\ ,\ \{\phi_{J,n}\}_{n\in\mathbf{Z}}]$ is orthogonal in $\mathbf{L}^2(\mathbb{R})$, we can verify that any two different wavelets or scaling functions α^{per} and β^{per} in (7.206) have necessarily a non-periodized version that satisfies $\langle \alpha(t), \beta(t+k)\rangle = 0$ for all $k \in \mathbb{Z}$. Lemma 7.2 thus proves that (7.206) is orthogonal in $\mathbf{L}^2[0,1]$.

To prove that this family generates $\mathbf{L}^2[0,1]$, we extend $f \in \mathbf{L}^2[0,1]$ with zeros outside $[0,1]$ and decompose it in the wavelet basis of $\mathbf{L}^2(\mathbb{R})$:

$$f = \sum_{j=-\infty}^{J} \sum_{n=-\infty}^{+\infty} \langle f, \psi_{j,n}\rangle\, \psi_{j,n} + \sum_{n=-\infty}^{+\infty} \langle f, \phi_{J,n}\rangle\, \phi_{J,n}\ . \tag{7.208}$$

This zero extension is periodized with the sum (7.205), which defines $f^{\text{per}}(t) = f(t)$ for $t \in [0,1]$. Periodizing (7.208) proves that f can be decomposed over the periodized wavelet family (7.206) in $\mathbf{L}^2[0,1]$. ■

Theorem 7.16 shows that periodizing a wavelet orthogonal basis of $\mathbf{L}^2(\mathbb{R})$ defines a wavelet orthogonal basis of $\mathbf{L}^2[0,1]$. If $J=0$ then there is a single scaling function, and one can verify that $\phi_{0,0}(t)=1$. The resulting scaling coefficient $\langle f,\phi_{0,0}\rangle$ is the average of f over $[0,1]$.

Periodic wavelet bases have the disadvantage of creating high amplitude wavelet coefficients in the neighborhood of $t=0$ and $t=1$, because the boundary wavelets have separate components with no vanishing moments. If $f(0)\neq f(1)$, the wavelet coefficients behave as if the signal were discontinuous at the boundaries. This can also be verified by extending $f\in\mathbf{L}^2[0,1]$ into an infinite 1 periodic signal f^{per} and by showing that

$$\int_0^1 f(t)\,\psi_{j,n}^{\mathrm{per}}(t)\,dt=\int_{-\infty}^{+\infty} f^{\mathrm{per}}(t)\,\psi_{j,n}(t)\,dt. \tag{7.209}$$

If $f(0)\neq f(1)$ then $f^{\mathrm{per}}(t)$ is discontinuous at $t=0$ and $t=1$, which creates high amplitude wavelet coefficients when $\psi_{j,n}$ overlaps the interval boundaries.

Periodic Discrete Transform For $f\in\mathbf{L}^2[0,1]$ let us consider

$$a_j[n]=\langle f,\phi_{j,n}^{\mathrm{per}}\rangle \text{ and } d_j[n]=\langle f,\psi_{j,n}^{\mathrm{per}}\rangle.$$

We verify as in (7.209) that these inner products are equal to the coefficients of a periodic signal decomposed in a non-periodic wavelet basis:

$$a_j[n]=\langle f^{\mathrm{per}},\phi_{j,n}\rangle \text{ and } d_j[n]=\langle f^{\mathrm{per}},\psi_{j,n}\rangle.$$

The convolution formulas of Theorem 7.7 thus apply if we take into account the periodicity of f^{per}. This means that $a_j[n]$ and $d_j[n]$ are considered as discrete signals of period 2^{-j}, and all convolutions in (7.107-7.109) must therefore be replaced by circular convolutions. Despite the poor behavior of periodic wavelets near the boundaries, they are often used because the numerical implementation is particularly simple.

7.5.2 Folded Wavelets

Decomposing $f\in\mathbf{L}^2[0,1]$ in a periodic wavelet basis was shown in (7.209) to be equivalent to a decomposition of f^{per} in a regular basis of $\mathbf{L}^2(\mathbb{R})$. Let us extend f with zeros outside $[0,1]$. To avoid creating discontinuities with such a periodization, the signal is folded with respect to $t=0$: $f_0(t)=f(t)+f(-t)$. The support of f_0 is $[-1,1]$ and it is transformed into a 2 periodic signal, as illustrated in Figure 7.19

$$f^{\mathrm{fold}}(t)=\sum_{k=-\infty}^{+\infty} f_0(t-2k)=\sum_{k=-\infty}^{+\infty} f(t-2k)+\sum_{k=-\infty}^{+\infty} f(2k-t). \tag{7.210}$$

Clearly $f^{\mathrm{fold}}(t)=f(t)$ if $t\in[0,1]$, and it is symmetric with respect to $t=0$ and $t=1$. If f is continuously differentiable then f^{fold} is continuous at $t=0$ and $t=1$, but its derivative is discontinuous at $t=0$ and $t=1$ if $f'(0)\neq 0$ and $f'(1)\neq 0$.

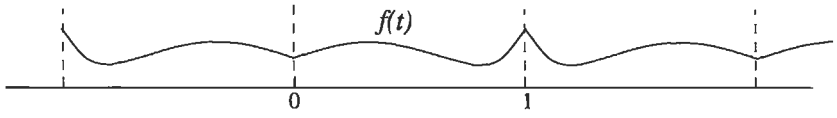


FIGURE 7.19 The folded signal $f^{\text{fold}}(t)$ is 2 periodic, symmetric about $t = 0$ and $t = 1$, and equal to $f(t)$ on $[0, 1]$.

Decomposing f^{fold} in a wavelet basis $\{\psi_{j,n}\}_{(j,n)\in\mathbb{Z}^2}$ is equivalent to decomposing f on a folded wavelet basis. Let $\psi_{j,n}^{\text{fold}}$ be the folding of $\psi_{j,n}$ with the summation (7.210). One can verify that

$$\int_0^1 f(t)\,\psi_{j,n}^{\text{fold}}(t)\,dt = \int_{-\infty}^{+\infty} f^{\text{fold}}(t)\,\psi_{j,n}(t)\,dt. \tag{7.211}$$

Suppose that f is regular over $[0, 1]$. Then f^{fold} is continuous at $t = 0, 1$ and hence produces smaller boundary wavelet coefficients than f^{per}. However, it is not continuously differentiable at $t = 0, 1$, which creates bigger wavelet coefficients at the boundary than inside.

To construct a basis of $\mathbf{L}^2[0, 1]$ with the folded wavelets $\psi_{j,n}^{\text{fold}}$, it is sufficient for $\psi(t)$ to be either symmetric or antisymmetric with respect to $t = 1/2$. The Haar wavelet is the only real compactly supported wavelet that is symmetric or antisymmetric and which generates an orthogonal basis of $\mathbf{L}^2(\mathbb{R})$. On the other hand, if we loosen up the orthogonality constraint, Section 7.4 proves that there exist biorthogonal bases constructed with compactly supported wavelets that are either symmetric or antisymmetric. Let $\{\psi_{j,n}\}_{(j,n)\in\mathbb{Z}^2}$ and $\{\tilde{\psi}_{j,n}\}_{(j,n)\in\mathbb{Z}^2}$ be such biorthogonal wavelet bases. If we fold the wavelets as well as the scaling functions then for $J \leq 0$

$$\left[\{\psi_{j,n}^{\text{fold}}\}_{-\infty<j\leq J,\,0\leq n<2^{-j}},\ \{\phi_{J,n}^{\text{fold}}\}_{0\leq n<2^{-J}}\right] \tag{7.212}$$

is a Riesz basis of $\mathbf{L}^2[0, 1]$ [134]. The biorthogonal basis is obtained by folding the dual wavelets $\tilde{\psi}_{j,n}$ and is given by

$$\left[\{\tilde{\psi}_{j,n}^{\text{fold}}\}_{-\infty<j\leq J,\,0\leq n<2^{-j}},\ \{\tilde{\phi}_{J,n}^{\text{fold}}\}_{0\leq n<2^{-J}}\right]. \tag{7.213}$$

If $J = 0$ then $\phi_{0,0}^{\text{fold}} = \tilde{\phi}_{0,0}^{\text{fold}} = 1$.

Biorthogonal wavelets of compact support are characterized by a pair of finite perfect reconstruction filters $(h, \tilde{h})$. The symmetry of these wavelets depends on the symmetry and size of the filters, as explained in Section 7.4.2. A fast folded wavelet transform is implemented with a modified filter bank algorithm, where the treatment of boundaries is slightly more complicated than for periodic wavelets. The symmetric and antisymmetric cases are considered separately.

Folded Discrete Transform For $f \in \mathbf{L}^2[0, 1]$, we consider

$$a_j[n] = \langle f, \phi_{j,n}^{\text{fold}} \rangle \text{ and } d_j[n] = \langle f, \psi_{j,n}^{\text{fold}} \rangle.$$

We verify as in (7.211) that these inner products are equal to the coefficients of a folded signal decomposed in a non-folded wavelet basis:

$$a_j[n] = \langle f^{\text{fold}}, \phi_{j,n} \rangle \text{ and } d_j[n] = \langle f^{\text{fold}}, \psi_{j,n} \rangle.$$

The convolution formulas of Theorem 7.7 thus apply if we take into account the symmetry and periodicity of f^{fold}. The symmetry properties of ϕ and ψ imply that $a_j[n]$ and $d_j[n]$ also have symmetry and periodicity properties, which must be taken into account in the calculations of (7.107-7.109).

Symmetric biorthogonal wavelets are constructed with perfect reconstruction filters h and $\hat{h}$ of odd size that are symmetric about $n = 0$. Then ϕ is symmetric about 0, whereas ψ is symmetric about $1/2$. As a result, one can verify that $a_j[n]$ is 2^{-j+1} periodic and symmetric about $n = 0$ and $n = 2^{-j}$. It is thus characterized by $2^{-j} + 1$ samples, for $0 \leq n \leq 2^{-j}$. The situation is different for $d_j[n]$ which is 2^{-j+1} periodic but symmetric with respect to $-1/2$ and $2^{-j} - 1/2$. It is characterized by 2^{-j} samples, for $0 \leq n < 2^{-j}$.

To initialize this algorithm, the original signal $a_L[n]$ defined over $0 \leq n < N - 1$ must be extended by one sample at $n = N$, and considered to be symmetric with respect to $n = 0$ and $n = N$. The extension is done by setting $a_L[N] = a_L[N-1]$. For any $J < L$, the resulting discrete wavelet representation $\left[\{d_j\}_{L<j\leq J}, a_J\right]$ is characterized by $N + 1$ coefficients. To avoid adding one more coefficient, one can modify symmetry at the right boundary of a_L by considering that it is symmetric with respect to $N - 1/2$ instead of N. The symmetry of the resulting a_j and d_j at the right boundary is modified accordingly by studying the properties of the convolution formula (7.162). As a result, these signals are characterized by 2^{-j} samples and the wavelet representation has N coefficients. This approach is used in most applications because it leads to simpler data structures which keep constant the number of coefficients. However, the discrete coefficients near the right boundary can not be written as inner products of some function $f(t)$ with dilated boundary wavelets.

Antisymmetric biorthogonal wavelets are obtained with perfect reconstruction filters h and $\hat{h}$ of even size that are symmetric about $n = 1/2$. In this case ϕ is symmetric about $1/2$ and ψ is antisymmetric about $1/2$. As a result a_j and d_j are 2^{-j+1} periodic and respectively symmetric and antisymmetric about $-1/2$ and $2^{-j} - 1/2$. They are both characterized by 2^{-j} samples, for $0 \leq n < 2^{-j}$. The algorithm is initialized by considering that $a_L[n]$ is symmetric with respect to $-1/2$ and $N - 1/2$. There is no need to add another sample. The resulting discrete wavelet representation $\left[\{d_j\}_{L<j\leq J}, a_J\right]$ is characterized by N coefficients.

7.5.3 Boundary Wavelets [3]

Wavelet coefficients are small in regions where the signal is regular only if the wavelets have enough vanishing moments. The restriction of periodic and folded "boundary" wavelets to the neighborhood of $t = 0$ and $t = 1$ have respectively 0 and 1 vanishing moment. These boundary wavelets thus cannot fully take advantage

of the signal regularity. They produce large inner products, as if the signal were discontinuous or had a discontinuous derivative. To avoid creating large amplitude wavelet coefficients at the boundaries, one must synthesize boundary wavelets that have as many vanishing moments as the original wavelet ψ. Initially introduced by Meyer, this approach has been refined by Cohen, Daubechies and Vial [134]. The main results are given without proofs.

Multiresolution of $\mathbf{L^2}[0,1]$ A wavelet basis of $\mathbf{L^2}[0,1]$ is constructed with a multiresolution approximation $\{\mathbf{V}_j^{\text{int}}\}_{-\infty<j\leq 0}$. A wavelet has p vanishing moments if it is orthogonal to all polynomials of degree $p-1$ or smaller. Since wavelets at a scale 2^j are orthogonal to functions in $\mathbf{V}_j^{\text{int}}$, to guarantee that they have p vanishing moments we make sure that polynomials of degree $p-1$ are inside $\mathbf{V}_j^{\text{int}}$.

We define an approximation space $\mathbf{V}_j^{\text{int}} \subset \mathbf{L^2}[0,1]$ with a compactly supported Daubechies scaling function ϕ, associated to a wavelet with p vanishing moments. Theorem 7.5 proves that the support of ϕ has size $2p-1$. We translate ϕ so that its support is $[-p+1, p]$. At a scale $2^j \leq (2p)^{-1}$, there are $2^{-j}-2p$ scaling functions with a support inside $[0,1]$:

$$\phi_{j,n}^{\text{int}}(t) = \phi_{j,n}(t) = \frac{1}{\sqrt{2^j}}\phi\Big(\frac{t-2^j n}{2^j}\Big) \quad \text{for } p \leq n < 2^{-j}-p\,.$$

To construct an approximation space $\mathbf{V}_j^{\text{int}}$ of dimension 2^{-j} we add p scaling functions with a support on the left boundary near $t=0$:

$$\phi_{j,n}^{\text{int}}(t) = \frac{1}{\sqrt{2^j}}\phi_n^{\text{left}}\Big(\frac{t}{2^j}\Big) \quad \text{for } 0 \leq n < p\,,$$

and p scaling functions on the right boundary near $t=1$:

$$\phi_{j,n}^{\text{int}}(t) = \frac{1}{\sqrt{2^j}}\phi_{2^{-j}-1-n}^{\text{right}}\Big(\frac{t-1}{2^j}\Big) \quad \text{for } 2^{-j}-p \leq n < 2^{-j}\,.$$

The following proposition constructs appropriate boundary scaling functions $\{\phi_n^{\text{left}}\}_{0\leq n<p}$ and $\{\phi_n^{\text{right}}\}_{0\leq n<p}$.

Proposition 7.6 (COHEN, DAUBECHIES, VIAL) *One can construct boundary scaling functions ϕ_n^{left} and ϕ_n^{right} so that if $2^{-j} \geq 2p$ then $\{\phi_{j,n}^{\text{int}}\}_{0\leq n<2^{-j}}$ is an orthonormal basis of a space $\mathbf{V}_j^{\text{int}}$ satisfying*

$$\mathbf{V}_j^{\text{int}} \subset \mathbf{V}_{j-1}^{\text{int}}$$

$$\lim_{j\to-\infty} \mathbf{V}_j^{\text{int}} = \text{Closure}\left(\bigcup_{j=-\infty}^{-\log_2(2p)} \mathbf{V}_j^{\text{int}}\right) = \mathbf{L^2}[0,1]\,,$$

and the restrictions to $[0,1]$ of polynomials of degree $p-1$ are in $\mathbf{V}_j^{\text{int}}$.

Proof[2]. A sketch of the proof is given. All details can be found in [134]. Since the wavelet ψ corresponding to ϕ has p vanishing moments, the Fix-Strang condition (7.75) implies that

$$q_k(t) = \sum_{n=-\infty}^{+\infty} n^k \phi(t-n) \tag{7.214}$$

is a polynomial of degree k. At any scale 2^j, $q_k(2^{-j}t)$ is still a polynomial of degree k, and for $0 \leq k < p$ this family defines a basis of polynomials of degree $p-1$. To guarantee that polynomials of degree $p-1$ are in $\mathbf{V}_j^{\text{int}}$ we impose that the restriction of $q_k(2^{-j}t)$ to $[0,1]$ can be decomposed in the basis of $\mathbf{V}_j^{\text{int}}$:

$$\begin{aligned} q_k(2^{-j}t)\,\mathbf{1}_{[0,1]}(t) \quad = \quad & \sum_{n=0}^{p-1} a[n]\,\phi_n^{\text{left}}(2^{-j}t) + \sum_{n=p}^{2^{-j}-p-1} n^k \phi(2^{-j}t-n) + \\ & \sum_{n=0}^{p-1} b[n]\,\phi_n^{\text{right}}(2^{-j}t-2^{-j})\,. \end{aligned} \tag{7.215}$$

Since the support of ϕ is $[-p+1,p]$, the condition (7.215) together with (7.214) can be separated into two non-overlapping left and right conditions. With a change of variable, we verify that (7.215) is equivalent to

$$\sum_{n=-p+1}^{p} n^k \phi(t-n)\,\mathbf{1}_{[0,+\infty)}(t) = \sum_{n=0}^{p-1} a[n]\,\phi_n^{\text{left}}(t), \tag{7.216}$$

and

$$\sum_{n=-p}^{p-1} n^k \phi(t-n)\,\mathbf{1}_{(-\infty,0]}(t) = \sum_{n=0}^{p-1} b[n]\,\phi_n^{\text{right}}(t). \tag{7.217}$$

The embedding property $\mathbf{V}_j^{\text{int}} \subset \mathbf{V}_{j-1}^{\text{int}}$ is obtained by imposing that the boundary scaling functions satisfy scaling equations. We suppose that ϕ_n^{left} has a support $[0, p+n]$ and satisfies a scaling equation of the form

$$2^{-1/2}\,\phi_n^{\text{left}}(2^{-1}t) = \sum_{l=0}^{p-1} H_{n,l}^{\text{left}}\,\phi_l^{\text{left}}(t) + \sum_{m=p}^{p+2n} h_{n,m}^{\text{left}}\,\phi(t-m)\,, \tag{7.218}$$

whereas ϕ_n^{right} has a support $[-p-n, 0]$ and satisfies a similar scaling equation on the right. The constants $H_{n,l}^{\text{left}}$, $h_{n,m}^{\text{left}}$, $H_{n,l}^{\text{right}}$ and $h_{n,m}^{\text{right}}$ are adjusted to verify the polynomial reproduction equations (7.216) and (7.217), while producing orthogonal scaling functions. The resulting family $\{\phi_{j,n}^{\text{int}}\}_{0 \leq n < 2^{-j}}$ is an orthonormal basis of a space $\mathbf{V}_j^{\text{int}}$.

The convergence of the spaces $\mathbf{V}_j^{\text{int}}$ to $\mathbf{L}^2[0,1]$ when 2^j goes to 0 is a consequence of the fact that the multiresolution spaces $\mathbf{V}_j$ generated by the Daubechies scaling function $\{\phi_{j,n}\}_{n \in \mathbf{Z}}$ converge to $\mathbf{L}^2(\mathbb{R})$. ■

The proof constructs the scaling functions through scaling equations specified by discrete filters. At the boundaries, the filter coefficients are adjusted to construct orthogonal scaling functions with a support in $[0,1]$, and to guarantee that polynomials of degree $p-1$ are reproduced by these scaling functions. Table 7.5 gives the filter coefficients for $p = 2$.

Wavelet Basis of $\mathbf{L}^2[0,1]$ Let $\mathbf{W}_j^{\text{int}}$ be the orthogonal complement of $\mathbf{V}_j^{\text{int}}$ in $\mathbf{V}_{j-1}^{\text{int}}$. The support of the Daubechies wavelet ψ with p vanishing moments is $[-p+1, p]$. Since $\psi_{j,n}$ is orthogonal to any $\phi_{j,l}$, we verify that an orthogonal basis of $\mathbf{W}_j^{\text{int}}$ can be constructed with the $2^{-j} - 2p$ inside wavelets with support in $[0,1]$:

$$\psi_{j,n}^{\text{int}}(t) = \psi_{j,n}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t - 2^j n}{2^j}\right) \quad \text{for } p \le n < 2^{-j} - p,$$

to which are added $2p$ left and right boundary wavelets

$$\psi_{j,n}^{\text{int}}(t) = \frac{1}{\sqrt{2^j}} \psi_n^{\text{left}}\left(\frac{t}{2^j}\right) \quad \text{for } 0 \le n < p\,,$$

$$\psi_{j,n}^{\text{int}}(t) = \frac{1}{\sqrt{2^j}} \psi_{2^{-j}-1-n}^{\text{right}}\left(\frac{t-1}{2^j}\right) \quad \text{for } 2^{-j} - p \le n < 2^{-j}\,.$$

Since $\mathbf{W}_j^{\text{int}} \subset \mathbf{V}_{j-1}^{\text{int}}$, the left and right boundary wavelets at any scale 2^j can be expanded into scaling functions at the scale 2^{j-1}. For $j = 1$ we impose that the left boundary wavelets satisfy equations of the form

$$\frac{1}{\sqrt{2}} \psi_n^{\text{left}}\left(\frac{t}{2}\right) = \sum_{l=0}^{p-1} G_{n,l}^{\text{left}} \phi_l^{\text{left}}(t) + \sum_{m=p}^{p+2n} g_{n,m}^{\text{left}} \phi(t-m)\,. \tag{7.219}$$

The right boundary wavelets satisfy similar equations. The coefficients $G_{n,l}^{\text{left}}$, $g_{n,m}^{\text{left}}$, $G_{n,l}^{\text{right}}$, $g_{n,m}^{\text{right}}$ are computed so that $\{\psi_{j,n}^{\text{int}}\}_{0 \le n < 2^{-j}}$ is an orthonormal basis of $\mathbf{W}_j^{\text{int}}$. Table 7.5 gives the values of these coefficients for $p = 2$.

For any $2^J \le (2p)^{-1}$ the multiresolution properties prove that

$$\mathbf{L}^2[0,1] = \mathbf{V}_J^{\text{int}} \oplus_{j=-\infty}^{J} \mathbf{W}_j^{\text{int}},$$

which implies that

$$\left[\{\phi_{J,n}^{\text{int}}\}_{0 \le n < 2^{-J}}\,,\ \{\psi_{j,n}^{\text{int}}\}_{-\infty < j \le J,\, 0 \le n < 2^{-j}}\right] \tag{7.220}$$

is an orthonormal wavelet basis of $\mathbf{L}^2[0,1]$. The boundary wavelets, like the inside wavelets, have p vanishing moments because polynomials of degree $p-1$ are included in the space $\mathbf{V}_j^{\text{int}}$. Figure 7.20 displays the $2p = 4$ boundary scaling functions and wavelets.

Fast Discrete Algorithm For any $f \in \mathbf{L}^2[0,1]$ we denote

$$a_j[n] = \langle f, \phi_{j,n}^{\text{int}} \rangle \ \text{ and } \ d_j[n] = \langle f, \psi_{j,n}^{\text{int}} \rangle \ \text{ for } 0 \le n \le 2^{-j}\,.$$

Wavelet coefficients are computed with a cascade of convolutions identical to Theorem 7.7 as long as filters do not overlap the signal boundaries. A Daubechies filter h is considered here to have a support located at $[-p+1, p]$. At the boundary, the usual Daubechies filters are replaced by the boundary filters that relate the boundary wavelets and scaling functions to the finer-scale scaling functions in (7.218) and (7.219).

k	l	$H_{k,l}^{\text{left}}$	$G_{k,l}^{\text{left}}$	k	m	$h_{k,m}^{\text{left}}$	$g_{k,m}^{\text{left}}$
0	0	0.6033325119	−0.7965436169	0	2	−0.398312997	−0.2587922483
0	1	0.690895531	0.5463927140	1	2	0.8500881025	0.227428117
1	0	0.03751746045	0.01003722456	1	3	0.2238203570	−0.8366028212
1	1	0.4573276599	0.1223510431	1	4	−0.1292227434	0.4830129218
k	l	$H_{k,l}^{\text{right}}$	$G_{k,l}^{\text{right}}$	k	m	$h_{k,m}^{\text{right}}$	$g_{k,m}^{\text{right}}$
−2	−2	0.1901514184	−0.3639069596	−2	−5	0.4431490496	0.235575950
−2	−1	−0.1942334074	0.3717189665	−2	−4	0.7675566693	0.4010695194
−1	−2	0.434896998	0.8014229620	−2	−3	0.3749553316	−0.7175799994
−2	−1	0.8705087534	−0.2575129195	−1	−3	0.2303890438	−0.5398225007
		$h[-1]$	$h[0]$			$h[1]$	$h[2]$
		0.482962913145	0.836516303738			0.224143868042	−0.129409522551

Table 7.5 Left and right border coefficients for a Daubechies wavelet with $p = 2$ vanishing moments. The inside filter coefficients are at the bottom of the table. A table of coefficients for $p \geq 2$ vanishing moments can be retrieved over the Internet at the FTP site ftp://math.princeton.edu/pub/user/ingrid/interval-tables.

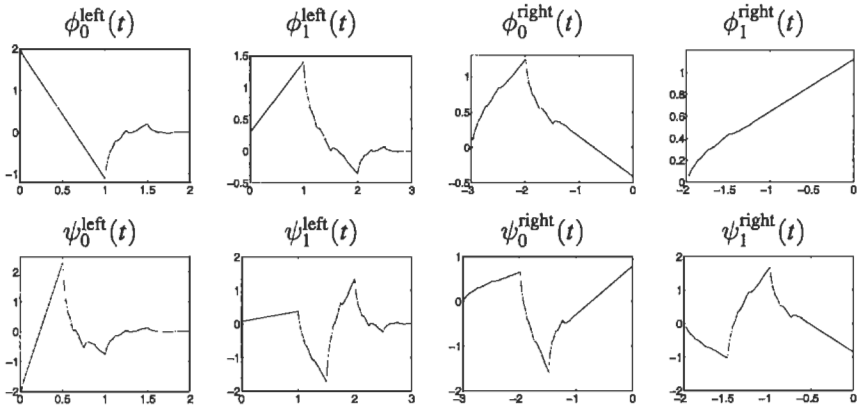


FIGURE 7.20 Boundary scaling functions and wavelets with $p = 2$ vanishing moments.

Theorem 7.17 (COHEN, DAUBECHIES, VIAL)
If $0 \leq k < p$

$$\begin{aligned} a_j[k] &= \sum_{l=0}^{p-1} H_{k,l}^{\text{left}}\, a_{j-1}[l] + \sum_{m=p}^{p+2k} h_{k,m}^{\text{left}}\, a_{j-1}[m], \\ d_j[k] &= \sum_{l=0}^{p-1} G_{k,l}^{\text{left}}\, a_{j-1}[l] + \sum_{m=p}^{p+2k} g_{k,m}^{\text{left}}\, a_{j-1}[m]. \end{aligned}$$

If $p \leq k < 2^{-j} - p$

$$\begin{aligned} a_j[k] &= \sum_{l=-\infty}^{+\infty} h[l-2k]\, a_{j-1}[l], \\ d_j[k] &= \sum_{l=-\infty}^{+\infty} g[l-2k]\, a_{j-1}[l]. \end{aligned}$$

If $-p \leq k < 0$

$$\begin{aligned} a_j[2^{-j}+k] &= \sum_{l=-p}^{-1} H_{k,l}^{\text{right}}\, a_{j-1}[2^{-j+1}+l] + \sum_{m=-p+2k+1}^{-p-1} h_{k,m}^{\text{right}}\, a_{j-1}[2^{-j+1}+m], \\ d_j[2^{-j}+k] &= \sum_{l=-p}^{-1} G_{k,l}^{\text{right}}\, a_{j-1}[2^{-j+1}+l] + \sum_{m=-p+2k+1}^{-p-1} g_{k,m}^{\text{right}}\, a_{j-1}[2^{-j+1}+m]. \end{aligned}$$

This cascade algorithm decomposes a_L into a discrete wavelet transform $[a_J, \{d_j\}_{L<j\leq J}]$ with $O(N)$ operations. The maximum scale must satisfy $2^J \leq$

$(2p)^{-1}$, because the number of boundary coefficients remains equal to $2p$ at all scales. The implementation is more complicated than the folding and periodic algorithms described in Sections 7.5.1 and 7.5.2, but does not require more computations. The signal a_L is reconstructed from its wavelet coefficients, by inverting the decomposition formula in Theorem 7.17.

Theorem 7.18 (COHEN, DAUBECHIES, VIAL)
If $0 \leq l \leq p-1$

$$a_{j-1}[l] = \sum_{k=0}^{p-1} H_{k,l}^{\text{left}} a_j[k] + \sum_{k=0}^{p-1} G_{k,l}^{\text{left}} d_j[k].$$

If $p \leq l \leq 3p-2$

$$a_{j-1}[l] = \sum_{k=(l-p)/2}^{p-1} h_{k,l}^{\text{left}} a_j[k] + \sum_{k=-\infty}^{+\infty} h[l-2k]\, a_j[k] + \sum_{k=(l-p)/2}^{p-1} g_{k,l}^{\text{left}} d_j[k] + \sum_{k=-\infty}^{+\infty} g[l-2k]\, d_j[k].$$

If $3p-1 \leq l \leq 2^{-j+1}-3p$

$$a_{j-1}[l] = \sum_{k=-\infty}^{+\infty} h[l-2k]\, a_j[k] + \sum_{k=-\infty}^{+\infty} g[l-2k]\, d_j[k].$$

If $-p-1 \geq l \geq -3p+1$

$$a_{j-1}[2^{-j+1}+l] = \sum_{k=-p}^{(l+p-1)/2} h_{k,l}^{\text{right}} a_j[2^{-j}+k] + \sum_{k=-\infty}^{+\infty} h[l-2k]\, a_j[2^{-j}+k] + \sum_{k=-p}^{(l+p-1)/2} g_{k,l}^{\text{right}} d_j[2^{-j}+k] + \sum_{k=-\infty}^{+\infty} g[l-2k]\, d_j[2^{-j}+k].$$

If $-1 \geq l \geq -p$

$$a_{j-1}[2^{-j+1}+l] = \sum_{k=-p}^{-1} H_{k,l}^{\text{right}} a_j[2^{-j}+k] + \sum_{k=-p}^{-1} G_{k,l}^{\text{right}} d_j[2^{-j}+k].$$

The original signal a_L is reconstructed from the orthogonal wavelet representation $[a_J, \{d_j\}_{L<j\leq J}]$ by iterating these equations for $L < j \leq J$. This reconstruction is performed with $O(N)$ operations.

7.6 MULTISCALE INTERPOLATIONS [2]

Multiresolution approximations are closely connected to the generalized interpolations and sampling theorems studied in Section 3.1.3. The next section constructs general classes of interpolation functions from orthogonal scaling functions and derives new sampling theorems. Interpolation bases have the advantage of easily computing the decomposition coefficients from the sample values of the signal. Section 7.6.2 constructs interpolation wavelet bases.

7.6.1 Interpolation and Sampling Theorems

Section 3.1.3 explains that a sampling scheme approximates a signal by its orthogonal projection onto a space $\mathbf{U}_T$ and samples this projection at intervals T. The space $\mathbf{U}_T$ is constructed so that any function in $\mathbf{U}_T$ can be recovered by interpolating a uniform sampling at intervals T. We relate the construction of interpolation functions to orthogonal scaling functions and compute the orthogonal projector on $\mathbf{U}_T$.

We call *interpolation function* any ϕ such that $\{\phi(t-n)\}_{n\in\mathbb{Z}}$ is a Riesz basis of the space $\mathbf{U}_1$ it generates, and which satisfies

$$\phi(n) = \begin{cases} 1 & \text{if } n = 0 \\ 0 & \text{if } n \neq 0 \end{cases} . \tag{7.221}$$

Any $f \in \mathbf{U}_1$ is recovered by interpolating its samples $f(n)$:

$$f(t) = \sum_{n=-\infty}^{+\infty} f(n)\,\phi(t-n). \tag{7.222}$$

Indeed, we know that f is a linear combination of the basis vector $\{\phi(t-n)\}_{n\in\mathbb{Z}}$ and the interpolation property (7.221) yields (7.222). The Whittaker sampling Theorem 3.1 is based on the interpolation function

$$\phi(t) = \frac{\sin \pi t}{\pi t}.$$

In this case, the space $\mathbf{U}_1$ is the set of functions whose Fourier transforms are included in $[-\pi, \pi]$.

Scaling an interpolation function yields a new interpolation for a different sampling interval. Let us define $\phi_T(t) = \phi(t/T)$ and

$$\mathbf{U}_T = \left\{ f \in \mathbf{L}^2(\mathbb{R}) \text{ with } f(Tt) \in \mathbf{U}_1 \right\}.$$

One can verify that any $f \in \mathbf{U}_T$ can be written

$$f(t) = \sum_{n=-\infty}^{+\infty} f(nT)\,\phi_T(t-nT)\,. \tag{7.223}$$

Scaling Autocorrelation We denote by ϕ_o an orthogonal scaling function, defined by the fact that $\{\phi_o(t-n)\}_{n\in\mathbf{Z}}$ is an orthonormal basis of a space $\mathbf{V}_0$ of a multiresolution approximation. Theorem 7.2 proves that this scaling function is characterized by a conjugate mirror filter h_o. The following theorem defines an interpolation function from the autocorrelation of ϕ_o [302].

Theorem 7.19 *Let* $\bar{\phi}_o(t) = \phi_o(-t)$ *and* $\bar{h}_o[n] = h_o[-n]$. *If* $|\hat{\phi}_o(\omega)| = O((1+|\omega|)^{-1})$ *then*

$$\phi(t) = \int_{-\infty}^{+\infty} \phi_o(u)\,\phi_o(u-t)\,du = \phi_o \star \bar{\phi}_o(t) \tag{7.224}$$

is an interpolation function. Moreover

$$\phi\left(\frac{t}{2}\right) = \sum_{n=-\infty}^{+\infty} h[n]\,\phi(t-n) \tag{7.225}$$

with

$$h[n] = \sum_{m=-\infty}^{+\infty} h_o[m]\,h_o[m-n] = h_o \star \bar{h}_o[n]. \tag{7.226}$$

Proof[3]. Observe first that

$$\phi(n) = \langle \phi_o(t), \phi_o(t-n)\rangle = \delta[n],$$

which prove the interpolation property (7.221). To prove that $\{\phi(t-n)\}_{n\in\mathbf{Z}}$ is a Riesz basis of the space $\mathbf{U}_1$ it generates, we verify the condition (7.10). The autocorrelation $\phi(t) = \phi_o \star \bar{\phi}_o(t)$ has a Fourier transform $\hat{\phi}(\omega) = |\hat{\phi}_o(\omega)|^2$. Condition (7.10) thus means that there exist $A > 0$ and $B > 0$ such that

$$\forall \omega \in [-\pi,\pi]\ ,\ \frac{1}{B} \leq \sum_{k=-\infty}^{+\infty} |\hat{\phi}_o(\omega - 2k\pi)|^4 \leq \frac{1}{A}. \tag{7.227}$$

We proved in (7.19) that the orthogonality of a family $\{\phi_o(t-n)\}_{n\in\mathbf{Z}}$ is equivalent to

$$\forall \omega \in [-\pi,\pi]\ ,\ \sum_{k=-\infty}^{+\infty} |\hat{\phi}_o(\omega + 2k\pi)|^2 = 1. \tag{7.228}$$

The right inequality of (7.227) is therefore valid for $A = 1$. Let us prove the left inequality. Since $|\hat{\phi}_o(\omega)| = O((1+|\omega|)^{-1})$, one can verify that there exists $K > 0$ such that for all $\omega \in [-\pi,\pi]$, $\sum_{|k|>K} |\hat{\phi}_o(\omega + 2k\pi)|^2 < 1/2$, so (7.228) implies that $\sum_{k=-K}^{K} |\hat{\phi}_o(\omega + 2k\pi)|^2 \geq 1/2$. It follows that

$$\sum_{k=-K}^{K} |\hat{\phi}_o(\omega + 2k\pi)|^4 \geq \frac{1}{4(2K+1)},$$

which proves (7.227) for $B = 4(2K+1)$.

Since ϕ_o is a scaling function, (7.28) proves that there exists a conjugate mirror filter h_o such that

$$\frac{1}{\sqrt{2}}\phi_o\left(\frac{t}{2}\right) = \sum_{n=-\infty}^{+\infty} h_o[n]\,\phi_o(t-n).$$

Computing $\phi(t) = \phi_o \star \bar{\phi}_o(t)$ yields (7.225) with $h[n] = h_o \star \bar{h}_o[n]$. ■

Theorem 7.19 proves that the autocorrelation of an orthogonal scaling function ϕ_o is an interpolation function ϕ that also satisfies a scaling equation. One can design ϕ to approximate regular signals efficiently by their orthogonal projection in $\mathbf{U}_T$. Definition 6.1 measures the regularity of f with a Lipschitz exponent, which depends on the difference between f and its Taylor polynomial expansion. The following proposition gives a condition for recovering polynomials by interpolating their samples with ϕ. It derives an upper bound for the error when approximating f by its orthogonal projection in $\mathbf{U}_T$.

Proposition 7.7 (FIX, STRANG) *Any polynomial $q(t)$ of degree smaller or equal to $p-1$ is decomposed into*

$$q(t) = \sum_{n=-\infty}^{+\infty} q(n)\,\phi(t-n) \tag{7.229}$$

if and only if $\hat{h}(\omega)$ has a zero of order p at $\omega = \pi$.
Suppose that this property is satisfied. If f has a compact support and is uniformly Lipschitz $\alpha \leq p$ then there exists $C > 0$ such that

$$\forall T > 0 \ , \ \|f - P_{\mathbf{U}_T} f\| \leq C\,T^{\alpha}. \tag{7.230}$$

Proof[3]. The main steps of the proof are given, without technical detail. Let us set $T = 2^j$. One can verify that the spaces $\{\mathbf{V}_j = \mathbf{U}_{2^j}\}_{j\in\mathbf{Z}}$ define a multiresolution approximation of $\mathbf{L}^2(\mathbb{R})$. The Riesz basis of $\mathbf{V}_0$ required by Definition 7.1 is obtained with $\theta = \phi$. This basis is orthogonalized by Theorem 7.1 to obtain an orthogonal basis of scaling functions. Theorem 7.3 derives a wavelet orthonormal basis $\{\psi_{j,n}\}_{(j,n)\in\mathbf{Z}^2}$ of $\mathbf{L}^2(\mathbb{R})$.

Using Theorem 7.4, one can verify that ψ has p vanishing moments if and only if $\hat{h}(\omega)$ has p zeros at π. Although ϕ is not the orthogonal scaling function, the Fix-Strang condition (7.75) remains valid. It is thus also equivalent that for $k < p$

$$q_k(t) = \sum_{n=-\infty}^{+\infty} n^k\,\phi(t-n)$$

is a polynomial of degree k. The interpolation property (7.222) implies that $q_k(n) = n^k$ for all $n \in \mathbb{Z}$ so $q_k(t) = t^k$. Since $\{t^k\}_{0\leq k<p}$ is a basis for polynomials of degree $p-1$, any polynomial $q(t)$ of degree $p-1$ can be decomposed over $\{\phi(t-n)\}_{n\in\mathbf{Z}}$ if and only if $\hat{h}(\omega)$ has p zeros at π.

We indicate how to prove (7.230) for $T = 2^j$. The truncated family of wavelets $\{\psi_{l,n}\}_{l \leq j, n \in \mathbf{Z}}$ is an orthogonal basis of the orthogonal complement of $\mathbf{U}_{2^j} = \mathbf{V}_j$ in $\mathbf{L}^2(\mathbb{R})$. Hence

$$\|f - P_{\mathbf{U}_{2^j}} f\|^2 = \sum_{l=-\infty}^{j} \sum_{n=-\infty}^{+\infty} |\langle f, \psi_{l,n} \rangle|^2.$$

If f is uniformly Lipschitz α, since ψ has p vanishing moments, Theorem 6.3 proves that there exists $A > 0$ such that

$$|Wf(2^l n, 2^l)| = |\langle f, \psi_{l,n} \rangle| \leq A 2^{(\alpha+1/2)l}.$$

To simplify the argument we suppose that ψ has a compact support, although this is not required. Since f also has a compact support, one can verify that the number of non-zero $\langle f, \psi_{l,n} \rangle$ is bounded by $K 2^{-l}$ for some $K > 0$. Hence

$$\|f - P_{\mathbf{U}_{2^j}} f\|^2 \leq \sum_{l=-\infty}^{j} K 2^{-l} A^2 2^{(2\alpha+1)l} \leq \frac{K A^2}{1 - 2^{-\alpha}} 2^{2\alpha j},$$

which proves (7.230) for $T = 2^j$. ■

As long as $\alpha \leq p$, the larger the Lipschitz exponent α the faster the error $\|f - P_{\mathbf{U}_T} f\|$ decays to zero when the sampling interval T decreases. If a signal f is $\mathbf{C}^k$ with a compact support then it is uniformly Lipschitz k, so Proposition 7.7 proves that $\|f - P_{\mathbf{U}_T} f\| \leq C T^k$.

Example 7.12 A cubic spline interpolation function is obtained from the linear spline scaling function ϕ_o. The Fourier transform expression (7.5) yields

$$\hat{\phi}(\omega) = |\hat{\phi}_o(\omega)|^2 = \frac{48 \sin^4(\omega/2)}{\omega^4 (1 + 2\cos^2(\omega/2))}. \tag{7.231}$$

Figure 7.21(a) gives the graph of ϕ, which has an infinite support but exponential decay. With Proposition 7.7 one can verify that this interpolation function recovers polynomials of degree 3 from a uniform sampling. The performance of spline interpolation functions for generalized sampling theorems is studied in [123, 335].

Example 7.13 Deslaurier-Dubuc [155] interpolation functions of degree $2p-1$ are compactly supported interpolation functions of minimal size that decompose polynomials of degree $2p-1$. One can verify that such an interpolation function is the autocorrelation of a scaling function ϕ_o. To reproduce polynomials of degree $2p-1$, Proposition 7.7 proves that $\hat{h}(\omega)$ must have a zero of order $2p$ at π. Since $h[n] = h_o \star \bar{h}_o[n]$ it follows that $\hat{h}(\omega) = |\hat{h}_o(\omega)|^2$, and hence $\hat{h}_o(\omega)$ has a zero of order p at π. Daubechies's Theorem 7.5 designs minimum size conjugate mirror filters h_o which satisfy this condition. Daubechies filters h_o have $2p$ non-zero coefficients and the resulting scaling function ϕ_o has a support of size $2p-1$. The

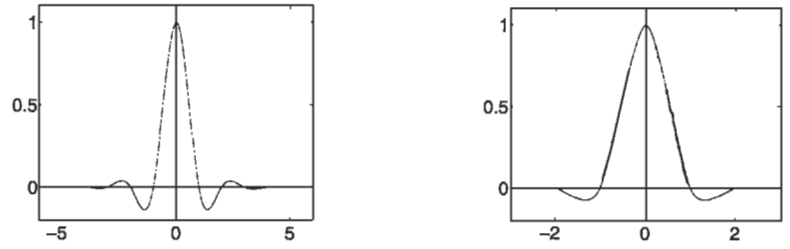


FIGURE 7.21 (a): Cubic spline interpolation function. (b): Deslaurier-Dubuc interpolation function of degree 3.

autocorrelation ϕ is the Deslaurier-Dubuc interpolation function, whose support is $[-2p+1, 2p-1]$.

For $p = 1$, $\phi_o = \mathbf{1}_{[0,1]}$ and ϕ is the piecewise linear tent function whose support is $[-1, 1]$. For $p = 2$, the Deslaurier-Dubuc interpolation function ϕ is the autocorrelation of the Daubechies 2 scaling function, shown in Figure 7.10. The graph of this interpolation function is in Figure 7.21(b). Polynomials of degree $2p-1 = 3$ are interpolated by this function.

The scaling equation (7.225) implies that any autocorrelation filter verifies $h[2n] = 0$ for $n \neq 0$. For any $p \geq 0$, the non-zero values of the resulting filter are calculated from the coefficients of the polynomial (7.173) that is factored to synthesize Daubechies filters. The support of h is $[-2p+1, 2p-1]$ and

$$h[2n+1] = (-1)^{p-n} \frac{\prod_{k=0}^{2p-1}(k-p+1/2)}{(n+1/2)\,(p-n-1)!\,(p+n)!} \quad \text{for } -p \leq n < p. \tag{7.232}$$

Dual Basis If $f \notin \mathbf{U}_T$ then it is approximated by its orthogonal projection $P_{\mathbf{U}_T} f$ on $\mathbf{U}_T$ before the samples at intervals T are recorded. This orthogonal projection is computed with a biorthogonal basis $\{\tilde{\phi}_T(t - nT)\}_{n\in\mathbb{Z}}$, which is calculated by the following theorem [75].

Theorem 7.20 *Let ϕ be an interpolation function. We define $\tilde{\phi}$ to be the function whose Fourier transform is*

$$\hat{\tilde{\phi}}(\omega) = \frac{\hat{\phi}(\omega)}{\sum_{k=-\infty}^{+\infty} |\hat{\phi}(\omega + 2k\pi)|^2}. \tag{7.233}$$

Let $\tilde{\phi}_T(t) = T^{-1}\tilde{\phi}(T^{-1}t)$. Then the family $\{\tilde{\phi}_T(t - nT)\}_{n\in\mathbb{Z}}$ is the biorthogonal basis of $\{\phi_T(t - nT)\}_{n\in\mathbb{Z}}$ in $\mathbf{U}_T$.

Proof[3]. Let us set $T = 1$. Since

$$\hat{\tilde{\phi}}(\omega) = \hat{a}(\omega)\,\hat{\phi}(\omega), \tag{7.234}$$

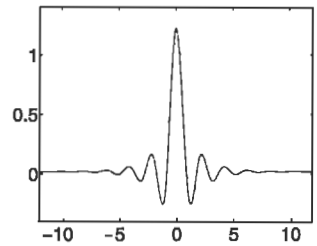


FIGURE 7.22 The dual cubic spline $\tilde{\phi}(t)$ associated to the spline interpolation function $\phi(t)$ shown in Figure 7.21(a).

where $\hat{a}(\omega) \in \mathbf{L}^2[-\pi,\pi]$ is 2π periodic, we derive as in (7.12) that $\tilde{\phi} \in \mathbf{U}_1$ and hence that $\tilde{\phi}(t-n) \in \mathbf{U}_1$ for any $n \in \mathbb{Z}$. A dual Riesz basis is unique and characterized by biorthogonality relations. Let $\bar{\phi}(t) = \phi(-t)$. For all $(n,m) \in \mathbb{Z}^2$, we must prove that

$$\langle \phi(t-n), \tilde{\phi}(t-m) \rangle = \tilde{\phi} \star \bar{\phi}(n-m) = \delta[n-m]. \tag{7.235}$$

Since the Fourier transform of $\tilde{\phi} \star \bar{\phi}(t)$ is $\hat{\tilde{\phi}}(\omega)\hat{\phi}^*(\omega)$, the Fourier transform of the biorthogonality conditions (7.235) yields

$$\sum_{k=-\infty}^{+\infty} \hat{\tilde{\phi}}(\omega + 2k\pi)\, \hat{\phi}^*(\omega + 2k\pi) = 1.$$

This equation is clearly satisfied for $\hat{\tilde{\phi}}$ defined by (7.233). The family $\{\tilde{\phi}(t-n)\}_{n\in\mathbb{Z}}$ is therefore the dual Riesz basis of $\{\phi(t-n)\}_{n\in\mathbb{Z}}$. The extension for any $T > 0$ is easily derived. ■

Figure 7.22 gives the graph of the cubic spline $\tilde{\phi}$ associated to the cubic spline interpolation function. The orthogonal projection of f over $\mathbf{U}_T$ is computed by decomposing f in the biorthogonal bases

$$P_{\mathbf{U}_T} f(t) = \sum_{n=-\infty}^{+\infty} \langle f(u), \tilde{\phi}_T(u - nT) \rangle\, \phi_T(t - nT). \tag{7.236}$$

Let $\bar{\tilde{\phi}}_T(t) = \tilde{\phi}_T(-t)$. The interpolation property (7.221) implies that

$$P_{\mathbf{U}_T} f(nT) = \langle f(u), \tilde{\phi}_T(u - nT) \rangle = f \star \bar{\tilde{\phi}}_T(nT). \tag{7.237}$$

This discretization of f through a projection onto $\mathbf{U}_T$ is therefore obtained by a filtering with $\bar{\tilde{\phi}}_T$ followed by a uniform sampling at intervals T. The best linear approximation of f is recovered with the interpolation formula (7.236).

7.6.2 Interpolation Wavelet Basis [3]

An interpolation function ϕ can recover a signal f from a uniform sampling $\{f(nT)\}_{n\in\mathbb{Z}}$ if f belongs to an appropriate subspace $\mathbf{U}_T$ of $\mathbf{L}^2(\mathbb{R})$. Donoho [162] has extended this approach by constructing interpolation wavelet bases of the whole space of uniformly continuous signals, with the sup norm. The decomposition coefficients are calculated from sample values instead of inner product integrals.

Subdivision Scheme Let ϕ be an interpolation function, which is the autocorrelation of an orthogonal scaling function ϕ_o. Let $\phi_{j,n}(t) = \phi(2^{-j}t - n)$. The constant $2^{-j/2}$ that normalizes the energy of $\phi_{j,n}$ is not added because we shall use a sup norm $\|f\|_\infty = \sup_{t\in\mathbb{R}} |f(t)|$ instead of the $\mathbf{L}^2(\mathbb{R})$ norm, and

$$\|\phi_{j,n}\|_\infty = \|\phi\|_\infty = |\phi(0)| = 1.$$

We define the interpolation space $\mathbf{V}_j$ of functions

$$g = \sum_{n=-\infty}^{+\infty} a[n]\,\phi_{j,n},$$

where $a[n]$ has at most a polynomial growth in n. Since ϕ is an interpolation function, $a[n] = g(2^j n)$. This space $\mathbf{V}_j$ is not included in $\mathbf{L}^2(\mathbb{R})$ since $a[n]$ may not have a finite energy. The scaling equation (7.225) implies that $\mathbf{V}_{j+1} \subset \mathbf{V}_j$ for any $j \in \mathbb{Z}$. If the autocorrelation filter h has a Fourier transform $\hat{h}(\omega)$ which has a zero of order p at $\omega = \pi$, then Proposition 7.7 proves that polynomials of degree smaller than $p-1$ are included in $\mathbf{V}_j$.

For $f \notin \mathbf{V}_j$, we define a simple projector on $\mathbf{V}_j$ that interpolates the dyadic samples $f(2^j n)$:

$$P_{\mathbf{V}_j} f(t) = \sum_{n=-\infty}^{+\infty} f(2^j n)\,\phi_j(t - 2^j n). \tag{7.238}$$

This projector has no orthogonality property but satisfies $P_{\mathbf{V}_j} f(2^j n) = f(2^j n)$. Let $\mathbf{C}_0$ be the space of functions that are uniformly continuous over $\mathbb{R}$. The following theorem proves that any $f \in \mathbf{C}_0$ can be approximated with an arbitrary precision by $P_{\mathbf{V}_j} f$ when 2^j goes to zero.

Theorem 7.21 (DONOHO) *Suppose that ϕ has an exponential decay. If $f \in \mathbf{C}_0$ then*

$$\lim_{j\to-\infty} \|f - P_{\mathbf{V}_j} f\|_\infty = \lim_{j\to-\infty} \sup_{t\in\mathbb{R}} |f(t) - P_{\mathbf{V}_j} f(t)| = 0. \tag{7.239}$$

Proof[3]. Let $\omega(\delta, f)$ denote the modulus of continuity

$$\omega(\delta, f) = \sup_{|h|\le\delta} \sup_{t\in\mathbb{R}} |f(t+h) - f(t)|. \tag{7.240}$$

By definition, $f \in \mathbf{C}_0$ if $\lim_{\delta \to 0} \omega(\delta, f) = 0$.

Any $t \in \mathbb{R}$ can be written $t = 2^j(n+h)$ with $n \in \mathbb{Z}$ and $|h| \leq 1$. Since $P_{\mathbf{V}_j} f(2^j n) = f(2^j n)$,

$$\begin{aligned} |f(2^j(n+h)) - P_{\mathbf{V}_j} f(2^j(n+h))| &\leq |f(2^j(n+h)) - f(2^j n)| \\ &\quad + |P_{\mathbf{V}_j} f(2^j(n+h)) - P_{\mathbf{V}_j} f(2^j n)| \\ &\leq \omega(2^j, f) + \omega(2^j, P_{\mathbf{V}_j} f). \end{aligned}$$

The next lemma proves that $\omega(2^j, P_{\mathbf{V}_j} f) \leq C_\phi \omega(2^j, f)$ where C_ϕ is a constant independent of j and f. Taking a sup over $t = 2^j(n+h)$ implies the final result:

$$\sup_{t \in \mathbb{R}} |f(t) - P_{\mathbf{V}_j} f(t)| \leq (1 + C_\phi) \omega(2^j, f) \to 0 \text{ when } j \to -\infty.$$

Lemma 7.3 *There exists $C_\phi > 0$ such that for all $j \in \mathbb{Z}$ and $f \in \mathbf{C}_0$*

$$\omega(2^j, P_{\mathbf{V}_j} f) \leq C_\phi \omega(2^j, f). \tag{7.241}$$

Let us set $j = 0$. For $|h| \leq 1$, a summation by parts gives

$$P_{\mathbf{V}_0} f(t+h) - P_{\mathbf{V}_0} f(t) = \sum_{n=-\infty}^{+\infty} (f(n+1) - f(n))\, \theta_h(t-n)$$

where

$$\theta_h(t) = \sum_{k=1}^{+\infty} (\phi(t+h-k) - \phi(t-k)).$$

Hence

$$|P_{\mathbf{V}_0} f(t+h) - P_{\mathbf{V}_0} f(t)| \leq \sup_{n \in \mathbb{Z}} |f(n+1) - f(n)| \sum_{n=-\infty}^{+\infty} |\theta_h(t-n)|. \tag{7.242}$$

Since ϕ has an exponential decay, there exists a constant C_ϕ such that if $|h| \leq 1$ and $t \in \mathbb{R}$ then $\sum_{n=-\infty}^{+\infty} |\theta_h(t-n)| \leq C_\phi$. Taking a sup over t in (7.242) proves that

$$\omega(1, P_{\mathbf{V}_0} f) \leq C_\phi \sup_{n \in \mathbb{Z}} |f(n+1) - f(n)| \leq C_\phi \omega(1, f).$$

Scaling this result by 2^j yields (7.241). ■

Interpolation Wavelets The projection $P_{\mathbf{V}_j} f(t)$ interpolates the values $f(2^j n)$. When reducing the scale by 2, we obtain a finer interpolation $P_{\mathbf{V}_{j-1}} f(t)$ which also goes through the intermediate samples $f(2^j(n+1/2))$. This refinement can be obtained by adding "details" that compensate for the difference between $P_{\mathbf{V}_j} f(2^j(n+1/2))$ and $f(2^j(n+1/2))$. To do this, we create a "detail" space $\mathbf{W}_j$ that provides the values $f(t)$ at intermediate dyadic points $t = 2^j(n+1/2)$. This space is constructed from interpolation functions centered at these locations, namely $\phi_{j-1,2n+1}$. We call *interpolation wavelets*

$$\psi_{j,n} = \phi_{j-1,2n+1}\,.$$

Observe that $\psi_{j,n}(t) = \psi(2^{-j}t - n)$ with

$$\psi(t) = \phi(2t - 1) .$$

The function ψ is not truly a wavelet since it has no vanishing moment. However, we shall see that it plays the same role as a wavelet in this decomposition. We define $\mathbf{W}_j$ to be the space of all sums $\sum_{n=-\infty}^{+\infty} a[n]\,\psi_{j,n}$. The following theorem proves that it is a (non-orthogonal) complement of $\mathbf{V}_j$ in $\mathbf{V}_{j-1}$.

Theorem 7.22 *For any $j \in \mathbb{Z}$*

$$\mathbf{V}_{j-1} = \mathbf{V}_j \oplus \mathbf{W}_j .$$

If $f \in \mathbf{V}_{j-1}$ then

$$f = \sum_{n=-\infty}^{+\infty} f(2^j n)\,\phi_{j,n} + \sum_{n=-\infty}^{+\infty} d_j[n]\,\psi_{j,n}$$

with

$$d_j[n] = f\left(2^j (n+1/2)\right) - P_{\mathbf{V}_j} f\left(2^j (n+1/2)\right) . \tag{7.243}$$

Proof[3]. Any $f \in \mathbf{V}_{j-1}$ can be written

$$f = \sum_{n=-\infty}^{+\infty} f(2^{j-1}n)\,\phi_{j-1,n} .$$

The function $f - P_{\mathbf{V}_j} f$ belongs to $\mathbf{V}_{j-1}$ and vanishes at $\{2^j n\}_{n\in\mathbb{Z}}$. It can thus be decomposed over the intermediate interpolation functions $\phi_{j-1,2n+1} = \psi_{j,n}$:

$$f(t) - P_{\mathbf{V}_j} f(t) = \sum_{n=-\infty}^{+\infty} d_j[n]\,\psi_{j,n}(t) \in \mathbf{W}_j .$$

This proves that $\mathbf{V}_{j-1} \subset \mathbf{V}_j \oplus \mathbf{W}_j$. By construction we know that $\mathbf{W}_j \subset \mathbf{V}_{j-1}$ so $\mathbf{V}_{j-1} = \mathbf{V}_j \oplus \mathbf{W}_j$. Setting $t = 2^{j-1}(2n+1)$ in this formula also verifies (7.243). ■

Theorem 7.22 refines an interpolation from a coarse grid $2^j n$ to a finer grid $2^{j-1}n$ by adding "details" whose coefficients $d_j[n]$ are the interpolation errors $f(2^j(n+1/2)) - P_{\mathbf{V}_j} f(2^j(n+1/2))$. The following theorem defines a interpolation wavelet basis of $\mathbf{C}_0$ in the sense of uniform convergence.

Theorem 7.23 *If $f \in \mathbf{C}_0$ then*

$$\lim_{\substack{m\to+\infty \\ l\to-\infty}} \|f - \sum_{n=-m}^{m} f(2^J n)\,\phi_{J,n} - \sum_{j=l}^{J} \sum_{n=-m}^{m} d_j[n]\,\psi_{j,n}\|_\infty = 0. \tag{7.244}$$

The formula (7.244) decomposes f into a coarse interpolation at intervals 2^J plus layers of details that give the interpolation errors on successively finer dyadic grids. The proof is done by choosing f to be a continuous function with a compact support, in which case (7.244) is derived from Theorem 7.22 and (7.239). The density of such functions in $\mathbf{C}_0$ (for the sup norm) allows us to extend this result to any f in $\mathbf{C}_0$. We shall write

$$f = \sum_{n=-\infty}^{+\infty} f(2^J n)\, \phi_{J,n} + \sum_{j=-\infty}^{J} \sum_{n=-\infty}^{+\infty} d_j[n]\, \psi_{j,n},$$

which means that $[\{\phi_{J,n}\}_{n\in\mathbb{Z}}, \{\psi_{j,n}\}_{n\in\mathbb{Z}, j\leq J}]$ is a basis of $\mathbf{C}_0$. In $\mathbf{L}^2(\mathbb{R})$, "biorthogonal" scaling functions and wavelets are formally defined by

$$f(2^J n) = \langle f, \tilde{\phi}_{J,n}\rangle = \int_{-\infty}^{+\infty} f(t)\, \tilde{\phi}_{J,n}(t)\, dt\,,$$

$$d_j[n] = \langle f, \tilde{\psi}_{j,n}\rangle = \int_{-\infty}^{+\infty} f(t)\, \tilde{\psi}_{j,n}(t)\, dt\,. \tag{7.245}$$

Clearly $\tilde{\phi}_{J,n}(t) = \delta(t - 2^J n)$. Similarly, (7.243) and (7.238) implies that $\tilde{\psi}_{j,n}$ is a finite sum of Diracs. These dual scaling functions and wavelets do not have a finite energy, which illustrates the fact that $[\{\phi_{J,n}\}_{n\in\mathbb{Z}}, \{\psi_{j,n}\}_{n\in\mathbb{Z}, j\leq J}]$ is not a Riesz basis of $\mathbf{L}^2(\mathbb{R})$.

If $\hat{h}(\omega)$ has p zeros at π then one can verify that $\tilde{\psi}_{j,n}$ has p vanishing moments. With similar derivations as in the proof of (6.20) in Theorem 6.4, one can show that if f is uniformly Lipschitz $\alpha \leq p$ then there exists $A > 0$ such that

$$|\langle f, \tilde{\psi}_{j,n}\rangle| = |d_j[n]| \leq A\, 2^{\alpha j}.$$

A regular signal yields small amplitude wavelet coefficients at fine scales. We can thus neglect these coefficients and still reconstruct a precise approximation of f.

Fast Calculations The interpolating wavelet transform of f is calculated at scale $1 \geq 2^j > N^{-1} = 2^L$ from its sample values $\{f(N^{-1}n)\}_{n\in\mathbb{Z}}$. At each scale 2^j, the values of f in between samples $\{2^j n\}_{n\in\mathbb{Z}}$ are calculated with the interpolation (7.238):

$$\begin{aligned} P_{\mathbf{V}_j} f\left(2^j (n+1/2)\right) &= \sum_{k=-\infty}^{+\infty} f(2^j k)\, \phi(n-k+1/2) \\ &= \sum_{k=-\infty}^{+\infty} f(2^j k)\, h_i[n-k], \end{aligned} \tag{7.246}$$

where the interpolation filter h_i is a subsampling of the autocorrelation filter h in (7.226):

$$h_i[n] = \phi(n+1/2) = h[2n+1]. \tag{7.247}$$

The wavelet coefficients are computed with (7.243):

$$d_j[n] = f\left(2^j(n+1/2)\right) - P_{\mathbf{V}_j} f\left(2^j(n+1/2)\right).$$

The reconstruction of $f(N^{-1}n)$ from the wavelet coefficients is performed recursively by recovering the samples $f(2^{j-1}n)$ from the coarser sampling $f(2^j n)$ with the interpolation (7.246) to which is added $d_j[n]$. If $h_i[n]$ is a finite filter of size K and if f has a support in $[0,1]$ then the decomposition and reconstruction algorithms require KN multiplications and additions.

A Deslauriers-Dubuc interpolation function ϕ has the shortest support while including polynomials of degree $2p-1$ in the spaces $\mathbf{V}_j$. The corresponding interpolation filter $h_i[n]$ defined by (7.247) has $2p$ non-zero coefficients for $-p \leq n < p$, which are calculated in (7.232). If $p=2$ then $h_i[1] = h_i[-2] = -1/16$ and $h_i[0] = h_i[-1] = 9/16$. Suppose that $q(t)$ is a polynomial of degree smaller or equal to $2p-1$. Since $q = P_{\mathbf{V}_j} q$, (7.246) implies a Lagrange interpolation formula

$$q\left(2^j(n+1/2)\right) = \sum_{k=-\infty}^{+\infty} q(2^j k)\, h_i[n-k]\,.$$

The Lagrange filter h_i of size $2p$ is the shortest filter that recovers intermediate values of polynomials of degree $2p-1$ from a uniform sampling.

To restrict the wavelet interpolation bases to a finite interval $[0,1]$ while reproducing polynomials of degree $2p-1$, the filter h_i is modified at the boundaries. Suppose that $f(N^{-1}n)$ is defined for $0 \leq n < N$. When computing the interpolation

$$P_{\mathbf{V}_j} f\left(2^j(n+1/2)\right) = \sum_{k=-\infty}^{+\infty} f(2^j k)\, h_i[n-k],$$

if n is too close to 0 or to $2^{-j}-1$ then h_i must be modified to ensure that the support of $h_i[n-k]$ remains inside $[0, 2^{-j}-1]$. The interpolation $P_{\mathbf{V}_j} f(2^j(n+1/2))$ is then calculated from the closest $2p$ samples $f(2^j k)$ for $2^j k \in [0,1]$. The new interpolation coefficients are computed in order to recover exactly all polynomials of degree $2p-1$ [324]. For $p=2$, the problem occurs only at $n=0$ and the appropriate boundary coefficients are

$$h_i[0] = \frac{5}{16}\,,\quad h_i[-1] = \frac{15}{16}\,,\quad h_i[-2] = \frac{-5}{16}\,,\quad h_i[-3] = \frac{1}{16}\,.$$

The symmetric boundary filter $h_i[-n]$ is used on the other side at $n = 2^{-j}-1$.

7.7 SEPARABLE WAVELET BASES [1]

To any wavelet orthonormal basis $\{\psi_{j,n}\}_{(j,n)\in\mathbb{Z}^2}$ of $\mathbf{L}^2(\mathbb{R})$, one can associate a separable wavelet orthonormal basis of $\mathbf{L}^2(\mathbb{R}^2)$:

$$\left\{\psi_{j_1,n_1}(x_1)\,\psi_{j_2,n_2}(x_2)\right\}_{(j_1,j_2,n_1,n_2)\in\mathbb{Z}^4}. \tag{7.248}$$

The functions $\psi_{j_1,n_1}(x_1)\,\psi_{j_2,n_2}(x_2)$ mix information at two different scales 2^{j_1} and 2^{j_2} along x_1 and x_2, which we often want to avoid. Separable multiresolutions lead to another construction of separable wavelet bases whose elements are products of functions dilated at the same scale. These multiresolution approximations also have important applications in computer vision, where they are used to process images at different levels of details. Lower resolution images are indeed represented by fewer pixels and might still carry enough information to perform a recognition task.

Signal decompositions in separable wavelet bases are computed with a separable extension of the filter bank algorithm described in Section 7.7.3. Non-separable wavelets bases can also be constructed [78, 239] but they are used less often in image processing. Section 7.7.4 constructs separable wavelet bases in any dimension, and explains the corresponding fast wavelet transform algorithm.

7.7.1 Separable Multiresolutions

As in one dimension, the notion of resolution is formalized with orthogonal projections in spaces of various sizes. The approximation of an image $f(x_1,x_2)$ at the resolution 2^{-j} is defined as the orthogonal projection of f on a space $\mathbf{V}_j^2$ that is included in $\mathbf{L}^2(\mathbb{R}^2)$. The space $\mathbf{V}_j^2$ is the set of all approximations at the resolution 2^{-j}. When the resolution decreases, the size of $\mathbf{V}_j^2$ decreases as well. The formal definition of a multiresolution approximation $\{\mathbf{V}_j^2\}_{j\in\mathbb{Z}}$ of $\mathbf{L}^2(\mathbb{R}^2)$ is a straightforward extension of Definition 7.1 that specifies multiresolutions of $\mathbf{L}^2(\mathbb{R})$. The same causality, completeness and scaling properties must be satisfied.

We consider the particular case of separable multiresolutions. Let $\{\mathbf{V}_j\}_{j\in\mathbb{Z}}$ be a multiresolution of $\mathbf{L}^2(\mathbb{R})$. A separable two-dimensional multiresolution is composed of the tensor product spaces

$$\mathbf{V}_j^2 = \mathbf{V}_j \otimes \mathbf{V}_j \,. \tag{7.249}$$

The space $\mathbf{V}_j^2$ is the set of finite energy functions $f(x_1,x_2)$ that are linear expansions of separable functions:

$$f(x_1,x_2) = \sum_{m=-\infty}^{+\infty} a[m]\, f_m(x_1)\, g_m(x_2) \quad \text{with} \quad f_m \in \mathbf{V}_j \;,\; g_m \in \mathbf{V}_j \,.$$

Section A.5 reviews the properties of tensor products. If $\{\mathbf{V}_j\}_{j\in\mathbb{Z}}$ is a multiresolution approximation of $\mathbf{L}^2(\mathbb{R})$ then $\{\mathbf{V}_j^2\}_{j\in\mathbb{Z}}$ is a multiresolution approximation of $\mathbf{L}^2(\mathbb{R}^2)$.

Theorem 7.1 demonstrates the existence of a scaling function ϕ such that $\{\phi_{j,m}\}_{m\in\mathbb{Z}}$ is an orthonormal basis of $\mathbf{V}_j$. Since $\mathbf{V}_j^2 = \mathbf{V}_j \otimes \mathbf{V}_j$, Theorem A.3 proves that for $x = (x_1,x_2)$ and $n = (n_1,n_2)$

$$\left\{ \phi_{j,n}^2(x) = \phi_{j,n_1}(x_1)\,\phi_{j,n_2}(x_2) = \frac{1}{2^j}\,\phi\Big(\frac{x_1 - 2^j n_1}{2^j}\Big)\,\phi\Big(\frac{x_2 - 2^j n_2}{2^j}\Big) \right\}_{n\in\mathbb{Z}^2}$$

is an orthonormal basis of $\mathbf{V}_j^2$. It is obtained by scaling by 2^j the two-dimensional separable scaling function $\phi^2(x) = \phi(x_1)\phi(x_2)$ and translating it on a two-dimensional square grid with intervals 2^j.

Example 7.14 Piecewise constant approximation Let $\mathbf{V}_j$ be the approximation space of functions that are constant on $[2^j m, 2^j(m+1)]$ for any $m \in \mathbb{Z}$. The tensor product defines a two-dimensional piecewise constant approximation. The space $\mathbf{V}_j^2$ is the set of functions that are constant on any square $[2^j n_1, 2^j(n_1+1)] \times [2^j n_2, 2^j(n_2+1)]$, for $(n_1, n_2) \in \mathbb{Z}^2$. The two dimensional scaling function is

$$\phi^2(x) = \phi(x_1)\phi(x_2) = \begin{cases} 1 & \text{if } 0 \le x_1 \le 1 \text{ and } 0 \le x_2 \le 1 \\ 0 & \text{otherwise} \end{cases}.$$

Example 7.15 Shannon approximation Let $\mathbf{V}_j$ be the space of functions whose Fourier transforms have a support included in $[-2^{-j}\pi, 2^{-j}\pi]$. The space $\mathbf{V}_j^2$ is the set of functions whose two-dimensional Fourier transforms have a support included in the low-frequency square $[-2^{-j}\pi, 2^{-j}\pi] \times [-2^{-j}\pi, 2^{-j}\pi]$. The two-dimensional scaling function is a perfect two-dimensional low-pass filter whose Fourier transform is

$$\hat{\phi}(\omega_1)\hat{\phi}(\omega_2) = \begin{cases} 1 & \text{if } |\omega_1| \le 2^{-j}\pi \text{ and } |\omega_2| \le 2^{-j}\pi \\ 0 & \text{otherwise} \end{cases}.$$

Example 7.16 Spline approximation Let $\mathbf{V}_j$ be the space of polynomial spline functions of degree p that are $\mathbf{C}^{p-1}$, with nodes located at $2^{-j}m$ for $m \in \mathbb{Z}$. The space $\mathbf{V}_j^2$ is composed of two-dimensional polynomial spline functions that are $p-1$ times continuously differentiable. The restriction of $f(x_1, x_2) \in \mathbf{V}_j^2$ to any square $[2^j n_1, 2^j(n_1+1)) \times [2^j n_2, 2^j(n_2+1))$ is a separable product $q_1(x_1)q_2(x_2)$ of two polynomials of degree at most p.

Multiresolution Vision An image of 512 by 512 pixels often includes too much information for real time vision processing. Multiresolution algorithms process less image data by selecting the relevant details that are necessary to perform a particular recognition task [62]. The human visual system uses a similar strategy. The distribution of photoreceptors on the retina is not uniform. The visual acuity is greatest at the center of the retina where the density of receptors is maximum. When moving apart from the center, the resolution decreases proportionally to the distance from the retina center [305].

The high resolution visual center is called the *fovea*. It is responsible for high acuity tasks such as reading or recognition. A retina with a uniform resolution equal to the highest fovea resolution would require about 10,000 times more photoreceptors. Such a uniform resolution retina would increase considerably the size of the optic nerve that transmits the retina information to the visual cortex and the size of the visual cortex that processes this data.

FIGURE 7.23 Multiresolution approximations $a_j[n_1, n_2]$ of an image at scales 2^j, for $-5 \geq j \geq -8$.

Active vision strategies [76] compensate the non-uniformity of visual resolution with eye saccades, which move successively the fovea over regions of a scene with a high information content. These saccades are partly guided by the lower resolution information gathered at the periphery of the retina. This multiresolution sensor has the advantage of providing high resolution information at selected locations, and a large field of view, with relatively little data.

Multiresolution algorithms implement in software [107] the search for important high resolution data. A uniform high resolution image is measured by a camera but only a small part of this information is processed. Figure 7.23 displays a pyramid of progressively lower resolution images calculated with a filter bank presented in Section 7.7.3. Coarse to fine algorithms analyze first the lower resolution image and selectively increase the resolution in regions where more details are needed. Such algorithms have been developed for object recognition, and stereo calculations [196]. Section 11.5.1 explains how to compute velocity vectors in video sequences with a coarse to fine matching algorithm.

7.7.2 Two-Dimensional Wavelet Bases

A separable wavelet orthonormal basis of $\mathbf{L}^2(\mathbb{R}^2)$ is constructed with separable products of a scaling function ϕ and a wavelet ψ. The scaling function ϕ is associated to a one-dimensional multiresolution approximation $\{\mathbf{V}_j\}_{j\in\mathbb{Z}}$. Let $\{\mathbf{V}_j^2\}_{j\in\mathbb{Z}}$ be the separable two-dimensional multiresolution defined by $\mathbf{V}_j^2 = \mathbf{V}_j \otimes \mathbf{V}_j$. Let $\mathbf{W}_j^2$ be the detail space equal to the orthogonal complement of the lower resolution approximation space $\mathbf{V}_j^2$ in $\mathbf{V}_{j-1}^2$:

$$\mathbf{V}_{j-1}^2 = \mathbf{V}_j^2 \oplus \mathbf{W}_j^2 \,. \tag{7.250}$$

To construct a wavelet orthonormal basis of $\mathbf{L}^2(\mathbb{R}^2)$, the following theorem builds a wavelet basis of each detail space $\mathbf{W}_j^2$.

Theorem 7.24 *Let ϕ be a scaling function and ψ be the corresponding wavelet generating a wavelet orthonormal basis of $\mathbf{L}^2(\mathbb{R})$. We define three wavelets:*

$$\psi^1(x) = \phi(x_1)\psi(x_2) \ , \ \psi^2(x) = \psi(x_1)\phi(x_2) \ , \ \psi^3(x) = \psi(x_1)\psi(x_2), \quad (7.251)$$

and denote for $1 \leq k \leq 3$

$$\psi^k_{j,n}(x) = \frac{1}{2^j}\psi^k\left(\frac{x_1 - 2^j n_1}{2^j}, \frac{x_2 - 2^j n_2}{2^j}\right).$$

The wavelet family

$$\left\{\psi^1_{j,n} \ , \ \psi^2_{j,n} \ , \ \psi^3_{j,n}\right\}_{n \in \mathbb{Z}^2} \quad (7.252)$$

is an orthonormal basis of $\mathbf{W}^2_j$ and

$$\left\{\psi^1_{j,n} \ , \ \psi^2_{j,n} \ , \ \psi^3_{j,n}\right\}_{(j,n) \in \mathbb{Z}^3} \quad (7.253)$$

is an orthonormal basis of $\mathbf{L}^2(\mathbb{R}^2)$.

Proof[1]. Equation (7.250) is rewritten

$$\mathbf{V}_{j-1} \otimes \mathbf{V}_{j-1} = (\mathbf{V}_j \otimes \mathbf{V}_j) \ominus \mathbf{W}^2_j \ . \quad (7.254)$$

The one-dimensional multiresolution space $\mathbf{V}_{j-1}$ can also be decomposed into $\mathbf{V}_{j-1} = \mathbf{V}_j \oplus \mathbf{W}_j$. By inserting this in (7.254), the distributivity of $\oplus$ with respect to $\otimes$ proves that

$$\mathbf{W}^2_j = (\mathbf{V}_j \otimes \mathbf{W}_j) \oplus (\mathbf{W}_j \otimes \mathbf{V}_j) \oplus (\mathbf{W}_j \otimes \mathbf{W}_j) \ . \quad (7.255)$$

Since $\{\phi_{j,m}\}_{m \in \mathbb{Z}}$ and $\{\psi_{j,m}\}_{m \in \mathbb{Z}}$ are orthonormal bases of $\mathbf{V}_j$ and $\mathbf{W}_j$, we derive that

$$\{\phi_{j,n_1}(x_1)\psi_{j,n_2}(x_2) \ , \ \psi_{j,n_1}(x_1)\phi_{j,n_2}(x_2) \ , \ \psi_{j,n_1}(x_1)\psi_{j,n_2}(x_2)\}_{(n_1,n_2) \in \mathbb{Z}^2}$$

is an orthonormal basis of $\mathbf{W}^2_j$. As in the one-dimensional case, the overall space $\mathbf{L}^2(\mathbb{R}^2)$ can be decomposed as an orthogonal sum of the detail spaces at all resolutions:

$$\mathbf{L}^2(\mathbb{R}^2) = \oplus_{j=-\infty}^{+\infty} \mathbf{W}^2_j. \quad (7.256)$$

Hence

$$\{\phi_{j,n_1}(x_1)\psi_{j,n_2}(x_2) \ , \ \psi_{j,n_1}(x_1)\phi_{j,n_2}(x_2) \ , \ \psi_{j,n_1}(x_1)\psi_{j,n_2}(x_2)\}_{(j,n_1,n_2) \in \mathbb{Z}^3}$$

is an orthonormal basis of $\mathbf{L}^2(\mathbb{R}^2)$. ■

The three wavelets extract image details at different scales and orientations. Over positive frequencies, $\hat{\phi}$ and $\hat{\psi}$ have an energy mainly concentrated respectively on $[0, \pi]$ and $[\pi, 2\pi]$. The separable wavelet expressions (7.251) imply that

$$\hat{\psi}^1(\omega_1, \omega_2) = \hat{\phi}(\omega_1)\hat{\psi}(\omega_2) \ , \ \hat{\psi}^2(\omega_1, \omega_2) = \hat{\psi}(\omega_1)\hat{\phi}(\omega_2)$$

and $\hat{\psi}^3(\omega_1, \omega_2) = \hat{\psi}(\omega_1)\hat{\psi}(\omega_2)$. Hence $|\hat{\psi}^1(\omega_1, \omega_2)|$ is large at low horizontal frequencies ω_1 and high vertical frequencies ω_2, whereas $|\hat{\psi}^2(\omega_1, \omega_2)|$ is large at high

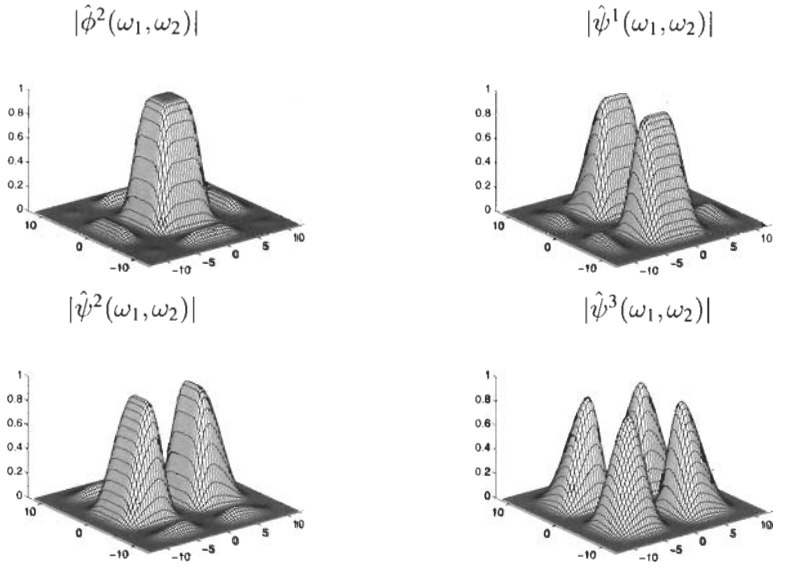


FIGURE 7.24 Fourier transforms of a separable scaling function and of 3 separable wavelets calculated from a one-dimensional Daubechies 4 wavelet.

horizontal frequencies and low vertical frequencies, and $|\hat{\psi}_3(\omega_1,\omega_2)|$ is large at high horizontal and vertical frequencies. Figure 7.24 displays the Fourier transform of separable wavelets and scaling functions calculated from a one-dimensional Daubechies 4 wavelet. Wavelet coefficients calculated with ψ^1 and ψ^2 are large along edges which are respectively horizontal and vertical. This is illustrated by the decomposition of a square in Figure 7.26. The wavelet ψ^3 produces large coefficients at the corners.

Example 7.17 For a Shannon multiresolution approximation, the resulting two-dimensional wavelet basis paves the two-dimensional Fourier plane (ω_1,ω_2) with dilated rectangles. The Fourier transforms $\hat{\phi}$ and $\hat{\psi}$ are the indicator functions respectively of $[-\pi,\pi]$ and $[-2\pi,-\pi]\cup[\pi,2\pi]$. The separable space $\mathbf{V}_j^2$ contains functions whose two-dimensional Fourier transforms have a support included in the low-frequency square $[-2^{-j}\pi,2^{-j}\pi]\times[-2^{-j}\pi,2^{-j}\pi]$. This corresponds to the support of $\hat{\phi}_{j,n}^2$ indicated in Figure 7.25. The detail space $\mathbf{W}_j^2$ is the orthogonal complement of $\mathbf{V}_j^2$ in $\mathbf{V}_{j-1}^2$ and thus includes functions whose Fourier transforms have a support in the frequency annulus between the two squares $[-2^{-j}\pi,2^{-j}\pi]\times[-2^{-j}\pi,2^{-j}\pi]$ and $[-2^{-j+1}\pi,2^{-j+1}\pi]\times[-2^{-j+1}\pi,2^{-j+1}\pi]$. As shown in Figure 7.25, this annulus is decomposed in three separable frequency regions, which are the Fourier supports of $\hat{\psi}_{j,n}^k$ for $1\leq k\leq 3$. Dilating these supports at all scales 2^j yields an exact cover of the frequency plane (ω_1,ω_2).

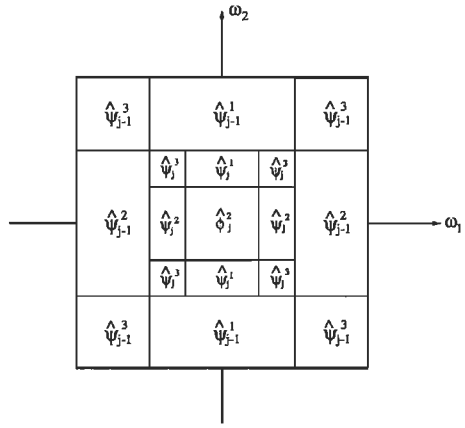


FIGURE 7.25 These dyadic rectangles indicate the regions where the energy of $\hat{\psi}^k_{j,n}$ is mostly concentrated, for $1 \leq k \leq 3$. Image approximations at the scale 2^j are restricted to the lower frequency square.

For general separable wavelet bases, Figure 7.25 gives only an indication of the domains where the energy of the different wavelets is concentrated. When the wavelets are constructed with a one-dimensional wavelet of compact support, the resulting Fourier transforms have side lobes that appear in Figure 7.24.

Example 7.18 Figure 7.26 gives two examples of wavelet transforms computed using separable Daubechies wavelets with $p = 4$ vanishing moments. They are calculated with the filter bank algorithm of Section 7.7.3. Coefficients of large amplitude in d^1_j, d^2_j and d^3_j correspond respectively to vertical high frequencies (horizontal edges), horizontal high frequencies (vertical edges), and high frequencies in both directions (corners). Regions where the image intensity varies smoothly yield nearly zero coefficients, shown in grey. The large number of nearly zero coefficients makes it particularly attractive for compact image coding.

Separable Biorthogonal Bases One-dimensional biorthogonal wavelet bases are extended to separable biorthogonal bases of $\mathbf{L}^2(\mathbb{R}^2)$ with the same approach as in Theorem 7.24. Let ϕ, ψ and $\tilde{\phi}$, $\tilde{\psi}$ be two dual pairs of scaling functions and wavelets that generate biorthogonal wavelet bases of $\mathbf{L}^2(\mathbb{R})$. The dual wavelets of ψ^1, ψ^2 and ψ^3 defined by (7.251) are

$$\tilde{\psi}^1(x) = \tilde{\phi}(x_1)\,\tilde{\psi}(x_2)\ ,\ \tilde{\psi}^2(x) = \tilde{\psi}(x_1)\,\tilde{\phi}(x_2)\ ,\ \tilde{\psi}^3(x) = \tilde{\psi}(x_1)\,\tilde{\psi}(x_2)\ . \quad (7.257)$$

One can verify that

$$\left\{\psi^1_{j,n}\ ,\ \psi^2_{j,n}\ ,\ \psi^3_{j,n}\right\}_{(j,n)\in\mathbb{Z}^3} \quad (7.258)$$

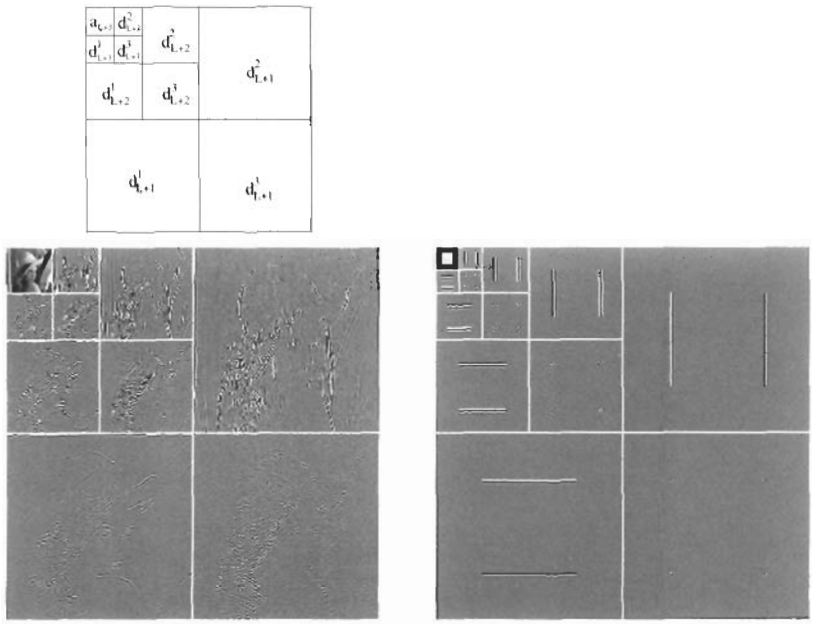


FIGURE 7.26 Separable wavelet transforms of Lena and of a white square in a black background, decomposed respectively on 3 and 4 octaves. Black, grey and white pixels correspond respectively to positive, zero and negative wavelet coefficients. The disposition of wavelet image coefficients $d_j^k[n,m] = \langle f, \psi_{j,n}^k \rangle$ is illustrated at the top.

and

$$\left\{ \tilde{\psi}_{j,n}^1 \, , \, \tilde{\psi}_{j,n}^2 \, , \, \tilde{\psi}_{j,n}^3 \right\}_{(j,n) \in \mathbb{Z}^3} \tag{7.259}$$

are biorthogonal Riesz bases of $\mathbf{L}^2(\mathbb{R}^2)$.

7.7.3 Fast Two-Dimensional Wavelet Transform

The fast wavelet transform algorithm presented in Section 7.3.1 is extended in two dimensions. At all scales 2^j and for any $n = (n_1, n_2)$, we denote

$$a_j[n] = \langle f, \phi_{j,n}^2 \rangle \text{ and } d_j^k[n] = \langle f, \psi_{j,n}^k \rangle \text{ for } 1 \leq k \leq 3 \, .$$

For any pair of one-dimensional filters $y[m]$ and $z[m]$ we write the product filter $yz[n] = y[n_1]\, z[n_2]$, and $\bar{y}[m] = y[-m]$. Let $h[m]$ and $g[m]$ be the conjugate mirror filters associated to the wavelet ψ.

The wavelet coefficients at the scale 2^{j+1} are calculated from a_j with two-dimensional separable convolutions and subsamplings. The decomposition formula are obtained by applying the one-dimensional convolution formula (7.108)

and (7.107) of Theorem 7.7 to the separable two-dimensional wavelets and scaling functions for $n = (n_1, n_2)$:

$$a_{j+1}[n] = a_j \star \bar{h}\bar{h}[2n] , \tag{7.260}$$

$$d^1_{j+1}[n] = a_j \star \bar{h}\bar{g}[2n] , \tag{7.261}$$

$$d^2_{j+1}[n] = a_j \star \bar{g}\bar{h}[2n] , \tag{7.262}$$

$$d^3_{j+1}[n] = a_j \star \bar{g}\bar{g}[2n] . \tag{7.263}$$

We showed in (3.53) that a separable two-dimensional convolution can be factored into one-dimensional convolutions along the rows and columns of the image. With the factorization illustrated in Figure 7.27(a), these four convolutions equations are computed with only six groups of one-dimensional convolutions. The rows of a_j are first convolved with $\bar{h}$ and $\bar{g}$ and subsampled by 2. The columns of these two output images are then convolved respectively with $\bar{h}$ and $\bar{g}$ and subsampled, which gives the four subsampled images a_{j+1}, d^1_{j+1}, d^2_{j+1} and d^3_{j+1}.

We denote by $\check{y}[n] = \check{y}[n_1, n_2]$ the image twice the size of $y[n]$, obtained by inserting a row of zeros and a column of zeros between pairs of consecutive rows and columns. The approximation a_j is recovered from the coarser scale approximation a_{j+1} and the wavelet coefficients d^k_{j+1} with two-dimensional separable convolutions derived from the one-dimensional reconstruction formula (7.109)

$$a_j[n] = \check{a}_{j+1} \star hh[n] + \check{d}^1_{j+1} \star hg[n] + \check{d}^2_{j+1} \star gh[n] + \check{d}^3_{j+1} \star gg[n] . \tag{7.264}$$

These four separable convolutions can also be factored into six groups of one-dimensional convolutions along rows and columns, illustrated in Figure 7.27(b).

Let $b[n]$ be an input image whose pixels have a distance $2^L = N^{-1}$. We associate to $b[n]$ a function $f(x) \in \mathbf{V}^2_L$ approximated at the scale 2^L. Its coefficients $a_L[n] = \langle f, \phi^2_{L,n} \rangle$ are defined like in (7.116) by

$$b[n] = N\, a_L[n] \approx f(N^{-1}n) . \tag{7.265}$$

The wavelet image representation of a_L is computed by iterating (7.260-7.263) for $L \leq j < J$:

$$\left[a_J \, , \, \{d^1_j, d^2_j, d^3_j\}_{L<j\leq J} \right] . \tag{7.266}$$

The image a_L is recovered from this wavelet representation by computing (7.264) for $J > j \geq L$.

Finite Image and Complexity When a_L is a finite image of N^2 pixels, we face boundary problems when computing the convolutions (7.260-7.264). Since the decomposition algorithm is separable along rows and columns, we use one of the three one-dimensional boundary techniques described in Section 7.5. The resulting values are decomposition coefficients in a wavelet basis of $\mathbf{L}^2[0,1]^2$. Depending on the boundary treatment, this wavelet basis is a periodic basis, a folded basis or a boundary adapted basis.

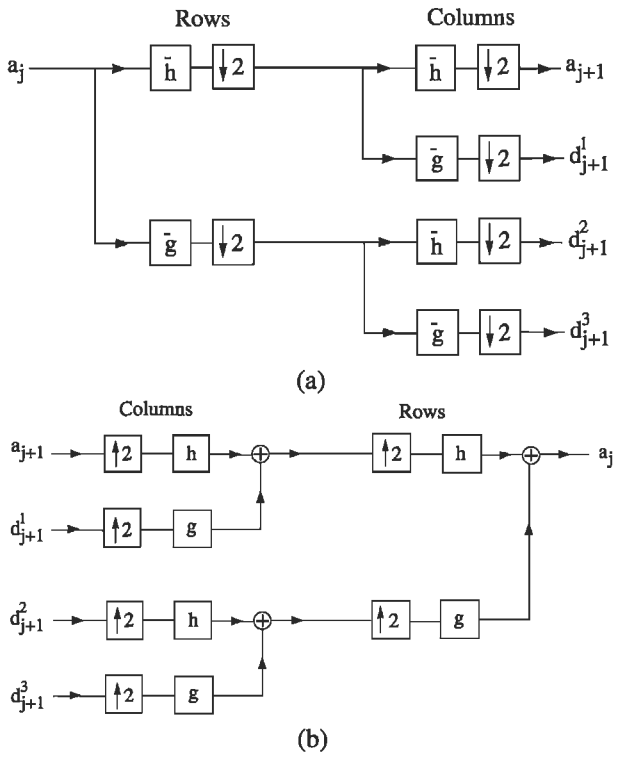


FIGURE 7.27 (a): Decomposition of a_j with 6 groups of one-dimensional convolutions and subsamplings along the image rows and columns. (b): Reconstruction of a_j by inserting zeros between the rows and columns of a_{j+1} and d^k_{j+1}, and filtering the output.

The resulting images a_j and d^k_j have 2^{-2j} samples. The images of the wavelet representation (7.266) thus include a total of N^2 samples. If h and g have size K, the reader can verify that $2K2^{-2(j-1)}$ multiplications and additions are needed to compute the four convolutions (7.260-7.263) with the factorization of Figure 7.27(a). The wavelet representation (7.266) is thus calculated with fewer than $\frac{8}{3}KN^2$ operations. The reconstruction of a_L by factoring the reconstruction equation (7.264) requires the same number of operations.

Fast Biorthogonal Wavelet Transform The decomposition of an image in a biorthogonal wavelet basis is performed with the same fast wavelet transform algorithm. Let $(\tilde{h}, \tilde{g})$ be the perfect reconstruction filters associated to (h, g). The inverse wavelet transform is computed by replacing the filters (h, g) that appear in (7.264) by $(\tilde{h}, \tilde{g})$.

7.7.4 Wavelet Bases in Higher Dimensions [2]

Separable wavelet orthonormal bases of $\mathbf{L}^2(\mathbb{R}^p)$ are constructed for any $p \geq 2$, with a procedure similar to the two-dimensional extension. Let ϕ be a scaling function and ψ a wavelet that yields an orthogonal basis of $\mathbf{L}^2(\mathbb{R})$. We denote $\theta^0 = \phi$ and $\theta^1 = \psi$. To any integer $0 \leq \epsilon < 2^p$ written in binary form $\epsilon = \epsilon_1 \ldots, \epsilon_p$ we associate the p-dimensional functions defined in $x = (x_1, \ldots, x_p)$ by

$$\psi^\epsilon(x) = \theta^{\epsilon_1}(x_1) \ldots \theta^{\epsilon_n}(x_p),$$

For $\epsilon = 0$, we obtain a p-dimensional scaling function

$$\psi^0(x) = \phi(x_1) \ldots \phi(x_p).$$

Non-zero indexes ϵ correspond to $2^p - 1$ wavelets. At any scale 2^j and for $n = (n_1, \ldots, n_p)$ we denote

$$\psi^\epsilon_{j,n}(x) = 2^{-pj/2} \psi^\epsilon\Big(\frac{x_1 - 2^j n_1}{2^j}, \ldots, \frac{x_p - 2^j n_p}{2^j}\Big).$$

Theorem 7.25 *The family obtained by dilating and translating the $2^p - 1$ wavelets for $\epsilon \neq 0$*

$$\left\{\psi^\epsilon_{j,n}\right\}_{1 \leq \epsilon < 2^p, (j,n) \in \mathbb{Z}^{p+1}} \tag{7.267}$$

is an orthonormal basis of $\mathbf{L}^2(\mathbb{R}^p)$.

The proof is done by induction on p. It follows the same steps as the proof of Theorem 7.24 which associates to a wavelet basis of $\mathbf{L}^2(\mathbb{R})$ a separable wavelet basis of $\mathbf{L}^2(\mathbb{R}^2)$. For $p = 2$, we verify that the basis (7.267) includes 3 elementary wavelets. For $p = 3$, there are 7 different wavelets.

Fast Wavelet Transform Let $b[n]$ be an input p-dimensional discrete signal sampled at intervals $N^{-1} = 2^L$. We associate to $b[n]$ an approximation f at the scale 2^L whose scaling coefficients $a_L[n] = \langle f, \psi^0_{L,n} \rangle$ satisfy

$$b[n] = N^{p/2} a_L[n] \approx f(N^{-1}n).$$

The wavelet coefficients of f at scales $2^j > 2^L$ are computed with separable convolutions and subsamplings along the p signal dimensions. We denote

$$a_j[n] = \langle f, \psi^0_{j,n} \rangle \text{ and } d^\epsilon_j[n] = \langle f, \psi^\epsilon_{j,n} \rangle \text{ for } 0 < \epsilon < 2^p.$$

The fast wavelet transform is computed with filters that are separable products of the one-dimensional filters h and g. The separable p-dimensional low-pass filter is

$$h^0[n] = h[n_1] \ldots h[n_p].$$

Let us denote $u^0[m] = h[m]$ and $u^1[m] = g[m]$. To any integer $\epsilon = \epsilon_1 \ldots \epsilon_p$ written in a binary form, we associate a separable p-dimensional band-pass filter

$$g^\epsilon[n] = u^{\epsilon_1}[n_1] \ldots u^{\epsilon_p}[n_p].$$

Let $\bar{g}^\epsilon[n] = g^\epsilon[-n]$. One can verify that

$$a_{j+1}[n] = a_j \star \bar{h}^0[2n] , \tag{7.268}$$

$$d^\epsilon_{j+1}[n] = a_j \star \bar{g}^\epsilon[2n] . \tag{7.269}$$

We denote by $\check{y}[n]$ the signal obtained by adding a zero between any two samples of $y[n]$ that are adjacent in the p-dimensional lattice $n = (n_1, \ldots, n_p)$. It doubles the size of $y[n]$ along each direction. If $y[n]$ has M^p samples, then $\check{y}[n]$ has $(2M)^p$ samples. The reconstruction is performed with

$$a_j[n] = \check{a}_{j+1} \star h^0[n] + \sum_{\epsilon=1}^{2^p-1} \check{d}^\epsilon_{j+1} \star g^\epsilon[n] . \tag{7.270}$$

The 2^p separable convolutions needed to compute a_j and $\{d^\epsilon_j\}_{1 \le \epsilon \le 2^p}$ as well as the reconstruction (7.270) can be factored in $2^{p+1} - 2$ groups of one-dimensional convolutions along the rows of p-dimensional signals. This is a generalization of the two-dimensional case, illustrated in Figures 7.27. The wavelet representation of a_L is

$$\left[\{d^\epsilon_j\}_{1 \le \epsilon < 2^p, L < j \le J} , a_J\right] . \tag{7.271}$$

It is computed by iterating (7.268) and (7.269) for $L \le j < J$. The reconstruction of a_L is performed with the partial reconstruction (7.270) for $J > j \ge L$.

If a_L is a finite signal of size N^p, the one-dimensional convolutions are modified with one of the three boundary techniques described in Section 7.5. The resulting algorithm computes decomposition coefficients in a separable wavelet basis of $\mathbf{L}^2[0,1]^p$. The signals a_j and d^ϵ_j have 2^{-pj} samples. Like a_L, the wavelet representation (7.271) is composed of N^p samples. If the filter h has K non-zero samples then the separable factorization of (7.268) and (7.269) requires $pK2^{-p(j-1)}$ multiplications and additions. The wavelet representation (7.271) is thus computed with fewer than $p(1-2^{-p})^{-1}KN^p$ multiplications and additions. The reconstruction is performed with the same number of operations.

7.8 PROBLEMS

7.1. [1] Let h be a conjugate mirror filter associated to a scaling function ϕ.

(a) Prove that if $\hat{h}(\omega)$ has a zero of order p at π then $\hat{\phi}^{(l)}(2k\pi) = 0$ for any $k \in \mathbb{Z} - \{0\}$ and $l < p$.

(b) Derive that if $q < p$ then $\sum_{n=-\infty}^{+\infty} n^q \phi(n) = \int_{-\infty}^{+\infty} t^q \phi(t)\, dt$.

7.2. [1] Prove that $\sum_{n=-\infty}^{+\infty} \phi(t-n) = 1$ if ϕ is an orthogonal scaling function.

7.3. [1] Let ϕ_m be the Battle-Lemarié scaling function of degree m defined in (7.23). Let ϕ be the Shannon scaling function defined by $\hat{\phi} = \mathbf{1}_{[-\pi,\pi]}$. Prove that $\lim_{m\to+\infty} \|\phi_m - \phi\| = 0$.

7.4. [1] Suppose that $h[n]$ is non-zero only for $0 \le n < K$. We denote $m[n] = \sqrt{2}h[n]$. The scaling equation is $\phi(t) = \sum_{n=0}^{K-1} m[n]\,\phi(2t-n)$.

(a) Suppose that $K = 2$. Prove that if t is a dyadic number that can be written in binary form with i digits: $t = 0.\epsilon_1\epsilon_2\cdots\epsilon_i$, with $\epsilon_k \in \{0,1\}$, then $\phi(t)$ is the product

$$\phi(t) = m[\epsilon_0] \times m[\epsilon_1] \times \cdots \times m[\epsilon_i] \times \phi(0) \ .$$

(b) For $K = 2$, show that if $m[0] = 4/3$ and $m[1] = 2/3$ then $\phi(t)$ is singular at all dyadic points. Verify numerically with WAVELAB that the resulting scaling equation does not define a finite energy function ϕ.

(c) Show that one can find two matrices $M[0]$ and $M[1]$ such that the K-dimensional vector $\Phi(t) = [\phi(t), \phi(t+1), \ldots, \phi(t+K-1)]^T$ satisfies

$$\Phi(t) = M[0]\,\Phi(2t) + M[1]\,\Phi(2t-1) \ .$$

(d) Show that one can compute $\Phi(t)$ at any dyadic number $t = 0.\epsilon_1\epsilon_2\cdots\epsilon_i$ with a product of matrices:

$$\Phi(t) = M[\epsilon_0] \times M[\epsilon_1] \times \cdots \times M[\epsilon_i] \times \Phi(0) \ .$$

7.5. [1] Let us define

$$\phi_{k+1}(t) = \sqrt{2} \sum_{n=-\infty}^{+\infty} h[n]\,\phi_k(2t-n) \ , \tag{7.272}$$

with $\phi_0 = \mathbf{1}_{[0,1]}$, and $a_k[n] = \langle \phi_k(t), \phi_k(t-n) \rangle$.

(a) Let

$$P\hat{f}(\omega) = \frac{1}{2}\left(|\hat{h}(\frac{\omega}{2})|^2\,\hat{f}(\frac{\omega}{2}) + |\hat{h}(\frac{\omega}{2}+\pi)|^2\,\hat{f}(\frac{\omega}{2}+\pi) \right) \ .$$

Prove that $\hat{a}_{k+1}(\omega) = P\hat{a}_k(\omega)$.

(b) Prove that if there exists ϕ such that $\lim_{k\to+\infty} \|\phi_k - \phi\| = 0$ then 1 is an eigenvalue of P and $\hat{\phi}(\omega) = \prod_{p=1}^{+\infty} 2^{-1/2}\hat{h}(2^{-p}\omega)$. What is the degree of freedom on ϕ_0 in order to still converge to the same limit ϕ?

(c) Implement in MATLAB the computations of $\phi_k(t)$ for the Daubechies conjugate mirror filter with $p = 6$ zeros at π. How many iterations are needed to obtain $\|\phi_k - \phi\| < 10^{-4}$? Try to improve the rate of convergence by modifying ϕ_0.

7.6. [1] Let $b[n] = f(N^{-1}n)$ with $2^L = N^{-1}$ and $f \in \mathbf{V}_L$. We want to recover $a_L[n] = \langle f, \phi_{L,n} \rangle$ from $b[n]$ to compute the wavelet coefficients of f with Theorem 7.7.

(a) Let $\phi_L[n] = 2^{-L/2}\phi(2^{-L}n)$. Prove that $b[n] = a_L \star \phi_L[n]$.

(b) Prove that if there exists $C > 0$ such that for all $\omega \in [-\pi,\pi]$

$$\hat{\phi}_d(\omega) = \sum_{k=-\infty}^{+\infty} \hat{\phi}(\omega + 2k\pi) \ge C,$$

then a_L can be calculated from b with a stable filter $\phi_L^{-1}[n]$.

(c) If ϕ is a cubic spline scaling function, compute numerically $\phi_L^{-1}[n]$. For a given numerical precision, compare the number of operations needed to compute a_L from b with the number of operations needed to compute the fast wavelet transform of a_L.

(d) Show that calculating a_L from b is equivalent to performing a change of basis in $\mathbf{V}_L$, from a Riesz interpolation basis to an orthonormal basis.

7.7. [1] *Quadrature mirror filters* We define a multirate filter bank with four filters h, g, $\tilde{h}$, and $\tilde{g}$, which decomposes a signal $a_0[n]$

$$a_1[n] = a_0 \star h[2n] \ , \ d_1[n] = a_0 \star g[2n].$$

Using the notation (7.106), we reconstruct

$$\tilde{a}_0[n] = \check{a}_1 \star \tilde{h}[n] + \check{d}_1 \star \tilde{g}[n].$$

(a) Prove that $\tilde{a}_0[n] = a_0[n-l]$ if

$$\hat{g}(\omega) = \hat{h}(\omega+\pi) \ , \ \hat{\tilde{h}}(\omega) = \hat{h}(\omega) \ , \ \hat{\tilde{g}}(\omega) = -\hat{\tilde{h}}(\omega+\pi) \ ,$$

and h satisfies the quadrature mirror condition

$$\hat{h}^2(\omega) - \hat{h}^2(\omega+\pi) = 2\,\mathrm{e}^{-il\omega} \ .$$

(b) Show that l is necessarily odd.

(c) Verify that the Haar filter (7.51) is a quadrature mirror filter (it is the only finite impulse response solution).

7.8. [1] Let f be a function of support $[0,1]$, that is equal to different polynomials of degree q on the intervals $\{[\tau_k, \tau_{k+1}]\}_{0 \leq k < K}$, with $\tau_0 = 0$ and $\tau_K = 1$. Let ψ be a Daubechies wavelet with p vanishing moments. Depending on p, compute the number of non-zero wavelet coefficients $\langle f, \psi_{j,n} \rangle$. How should we choose p to minimize this number?

7.9. [1] Let θ be a box spline of degree m obtained by $m+1$ convolutions of $\mathbf{1}_{[0,1]}$ with itself.

(a) Prove that

$$\theta(t) = \frac{1}{m!} \sum_{k=0}^{m+1} (-1)^k \binom{m+1}{k} ([t-k]_+)^m,$$

where $[x]_+ = \max(x, 0)$. Hint: write $\mathbf{1}_{[0,1]} = \mathbf{1}_{[0,+\infty)} - \mathbf{1}_{(1,+\infty)}$.

(b) Let A_m and B_m be the Riesz bounds of $\{\theta(t-n)\}_{n \in \mathbb{Z}}$. With Proposition 7.1, prove that $\lim_{m \to +\infty} B_m = +\infty$. Compute numerically A_m and B_m for $m \in \{0, \ldots, 5\}$, with MATLAB.

7.10. [1] Prove that if $\{\psi_{j,n}\}_{(j,n) \in \mathbb{Z}^2}$ is an orthonormal basis of $\mathbf{L}^2(\mathbb{R})$ then for all $\omega \in \mathbb{R} - \{0\}$ $\sum_{j=-\infty}^{+\infty} |\hat{\psi}(2^j \omega)|^2 = 1$. Find an example showing that the converse is not true.

7.11. [2] Let us define

$$\hat{\psi}(\omega) = \begin{cases} 1 & \text{if } 4\pi/7 \leq |\omega| \leq \pi \text{ or } 4\pi \leq |\omega| \leq 4\pi + 4\pi/7 \\ 0 & \text{otherwise} \end{cases}$$

Prove that $\{\psi_{j,n}\}_{(j,n)\in\mathbf{Z}^2}$ is an orthonormal basis of $\mathbf{L}^2(\mathbb{R})$. Prove that ψ is not associated to a scaling function ϕ that generates a multiresolution approximation.

7.12. [1] Express the Coiflet property (7.104) as an equivalent condition on the conjugate mirror filter $\hat{h}(e^{i\omega})$.

7.13. [1] Prove that $\psi(t)$ has p vanishing moments if and only if for all $j > 0$ the discrete wavelets $\psi_j[n]$ defined in (7.145) have p discrete vanishing moments

$$\sum_{n=-\infty}^{+\infty} n^k \psi_j[n] = 0 \quad \text{for } 0 \leq k < p.$$

7.14. [1] Let $\psi(t)$ be a compactly supported wavelet calculated with Daubechies conjugate mirror filters (h, g). Let $\psi'_{j,n}(t) = 2^{-j/2}\psi'(2^{-j}t - n)$ be the derivative wavelets.

(a) Verify that h_1 and g_1 defined by

$$\hat{h}_1(\omega) = 2\hat{h}(\omega)(e^{i\omega} - 1)^{-1} \ , \ \hat{g}_1(\omega) = 2(e^{i\omega} - 1)\hat{g}(\omega)$$

are finite impulse response filters.

(b) Prove that the Fourier transform of $\psi'(t)$ can be written

$$\widehat{\psi'}(\omega) = \frac{\hat{g}_1(2^{-1}\omega)}{\sqrt{2}} \prod_{p=2}^{+\infty} \frac{\hat{h}_1(2^{-p}\omega)}{\sqrt{2}} \ .$$

(c) Describe a fast filter bank algorithm to compute the derivative wavelet coefficients $\langle f, \psi'_{j,n}\rangle$ [95].

7.15. [2] Let $\psi(t)$ be a compactly supported wavelet calculated with Daubechies conjugate mirror filters (h, g). Let $\hat{h}^a(\omega) = |\hat{h}(\omega)|^2$. We verify that $\hat{\psi}^a(\omega) = \hat{\psi}(\omega)\hat{h}^a(\omega/4 - \pi/2)$ is an almost analytic wavelet.

(a) Prove that ψ^a is a complex wavelet such that $\text{Real}[\psi^a] = \psi$.

(b) Compute $\psi^a(\omega)$ in MATLAB for a Daubechies wavelet with four vanishing moments. Explain why $\psi^a(\omega) \approx 0$ for $\omega < 0$.

(c) Let $\psi^a_{j,n}(t) = 2^{-j/2}\psi^a(2^{-j}t - n)$. Using the fact that

$$\hat{\psi}^a(\omega) = \frac{\hat{g}(2^{-1}\omega)}{\sqrt{2}} \frac{\hat{h}(2^{-2}\omega)}{\sqrt{2}} \frac{|\hat{h}(2^{-2}\omega - 2^{-1}\pi)|^2}{\sqrt{2}} \prod_{k=3}^{+\infty} \frac{\hat{h}(2^{-k}\omega)}{\sqrt{2}}$$

show that we can modify the fast wavelet transform algorithm to compute the "analytic" wavelet coefficients $\langle f, \psi^a_{j,n}\rangle$ by inserting a new filter.

(d) Let ϕ be the scaling function associated to ψ. We define separable two-dimensional "analytic" wavelets by:

$$\psi^1(x) = \psi^a(x_1)\phi(x_2) \ , \ \psi^2(x) = \phi(x_1)\psi^a(x_2) \ ,$$

$$\psi^3(x) = \psi^a(x_1)\psi^a(x_2) \ , \ \psi^4(x) = \psi^a(x_1)\psi^a(-x_2) \ .$$

Let $\psi^k_{j,n}(x) = 2^{-j}\psi^k(2^{-j}x - n)$ for $n \in \mathbb{Z}^2$. Modify the separable wavelet filter bank algorithm of Section 7.7.3 to compute the "analytic" wavelet coefficients $\langle f, \psi^k_{j,n}\rangle$.

(e) Prove that $\{\psi^k_{j,n}\}_{1\le k\le 4, j\in\mathbb{Z}, n\in\mathbb{Z}^2}$ is a frame of the space of *real* functions $f \in \mathbf{L}^2(\mathbb{R}^2)$ [95].

7.16. [2] *Multiwavelets* We define the following two scaling functions:

$$\begin{aligned}\phi_1(t) &= \phi_1(2t) + \phi_1(2t-1)\\ \phi_2(t) &= \frac{1}{2}\Big(\phi_2(2t) + \phi_2(2t-1) - \phi_1(2t) + \phi_1(2t-1)\Big)\end{aligned}$$

(a) Compute the functions ϕ_1 and ϕ_2. Prove that $\{\phi_1(t-n), \phi_2(t-n)\}_{n\in\mathbb{Z}}$ is an orthonormal basis of a space $\mathbf{V}_0$ that will be specified.

(b) Find ψ_1 and ψ_2 with a support on $[0,1]$ that are orthogonal to each other and to ϕ_1 and ϕ_2. Plot these wavelets. Verify that they have 2 vanishing moments and that they generate an orthonormal basis of $\mathbf{L}^2(\mathbb{R})$.

7.17. [2] Let f^{fold} be the folded function defined in (7.210).

(a) Let $\alpha(t), \beta(t) \in \mathbf{L}^2(\mathbb{R})$ be two functions that are either symmetric or antisymmetric about $t = 0$. If $\langle \alpha(t), \beta(t+2k)\rangle = 0$ and $\langle \alpha(t), \beta(2k-t)\rangle = 0$ for all $k \in \mathbb{Z}$, then prove that

$$\int_0^1 \alpha^{\text{fold}}(t)\,\beta^{\text{fold}}(t)\,dt = 0 \ .$$

(b) Prove that if ψ, $\tilde{\psi}$, ϕ, $\tilde{\phi}$ are either symmetric or antisymmetric with respect to $t = 1/2$ or $t = 0$, and generate biorthogonal bases of $\mathbf{L}^2(\mathbb{R})$, then the folded bases (7.212) and (7.213) are biorthogonal bases of $\mathbf{L}^2[0,1]$. Hint: use the same approach as in Theorem 7.16.

7.18. [1] A recursive filter has a Fourier transform that is a ratio of trigonometric polynomials as in (2.31).

(a) Let $p[n] = h \star \bar{h}[n]$ with $\bar{h}[n] = h[-n]$. Verify that if h is a recursive conjugate mirror filter then $\hat{p}(\omega) + \hat{p}(\omega+\pi) = 2$ and there exists $\hat{r}(\omega) = \sum_{k=0}^{K-1} r[k]\,e^{-ik\omega}$ such that

$$\hat{p}(\omega) = \frac{2|\hat{r}(\omega)|^2}{|\hat{r}(\omega)|^2 + |\hat{r}(\omega+\pi)|^2}. \tag{7.273}$$

(b) Suppose that K is even and that $r[K/2-1-k] = r[K/2+k]$. Verify that

$$\hat{p}(\omega) = \frac{|\hat{r}(\omega)|^2}{2\,|\hat{r}(\omega) + \hat{r}(\omega+\pi)|^2}. \tag{7.274}$$

(c) If $\hat{r}(\omega) = (1+e^{-i\omega})^{K-1}$ with $K = 6$, compute $\hat{h}(\omega)$ with the factorization (7.274), and verify that it is a stable filter (Problem 3.8). Compute numerically and plot with WAVELAB the graph of the corresponding wavelet $\psi(t)$.

7.19. [1] *Balancing* Suppose that h, $\tilde{h}$ define a pair of perfect reconstruction filters satisfying (7.129).

(a) Prove that

$$h_{new}[n] = \frac{1}{2}\Big(h[n] + h[n-1]\Big) \ , \ \tilde{h}_{new}[n] = \frac{1}{2}\Big(\tilde{h}[n] + \tilde{h}[n-1]\Big)$$

defines a new pair of perfect reconstruction filters. Verify that $\hat{h}_{new}(\omega)$ and $\widehat{\tilde{h}}_{new}(\omega)$ have respectively 1 more and 1 less zero at π than $\hat{h}(\omega)$ and $\widehat{\tilde{h}}(\omega)$ [68].

(b) The Deslauriers-Dubuc filters are $\hat{h}(\omega) = 1$ and

$$\widehat{\tilde{h}}(\omega) = \frac{1}{16}\left(-e^{-3i\omega} + 9e^{-i\omega} + 16 + 9e^{i\omega} - e^{3i\omega}\right) .$$

Compute h_{new} and $\tilde{h}_{new}$ as well as the corresponding biorthogonal wavelets ψ_{new}, $\tilde{\psi}_{new}$, after one balancing and after a second balancing.

7.20. [1] *Lifting* The filter (7.192) is calculated by lifting lazy filters. Find a dual lifting that produces a lifted filter with a support of size 9 so that $\widehat{\tilde{h}}^{l}(\omega)$ has 2 zeros at π. Compute the resulting lifted wavelets and scaling functions. Implement in WAVELAB the corresponding fast wavelet transform and its inverse with the polyphase decomposition illustrated in Figure 7.16.

7.21. [1] For a Deslaurier-Dubuc interpolation wavelet of degree 3, compute the dual wavelet $\tilde{\psi}$ in (7.245), which is a sum of Diracs. Verify that it has 4 vanishing moments.

7.22. [1] Prove that a Deslaurier-Dubuc interpolation function of degree $2p-1$ converges to a sinc function when p goes to $+\infty$.

7.23. [2] Let ϕ be an autocorrelation scaling function that reproduces polynomials of degree $p-1$ as in (7.229). Prove that if f is uniformly Lipschitz α then under the same hypotheses as in Theorem 7.21, there exists $K > 0$ such that

$$\|f - P_{V_j} f\|_\infty \leq K 2^{\alpha j}.$$

7.24. [1] Let $\phi(t)$ be an interpolation function that generates an interpolation wavelet basis of $\mathbf{C}_0(\mathbb{R})$. Construct a separable interpolation wavelet basis of the space $\mathbf{C}_0(\mathbb{R}^p)$ of uniformly continuous p-dimensional signals $f(x_1, \ldots, x_p)$. Hint: construct $2^p - 1$ interpolation wavelets by appropriately translating $\phi(x_1)\cdots\phi(x_p)$.

7.25. [2] *Fractional Brownian* Let $\psi(t)$ be a compactly supported wavelet with p vanishing moments that generates an orthonormal basis of $\mathbf{L}^2(\mathbb{R})$. The covariance of a fractional Brownian motion $B_H(t)$ is given by (6.93).

(a) Prove that $\mathrm{E}\{|\langle B_H, \psi_{j,n}\rangle|^2\}$ is proportional to $2^{j(2H+1)}$. Hint: use Problem 6.15.

(b) Prove that the decorrelation between same scale wavelet coefficients increases when the number p of vanishing moments of ψ increases:

$$\mathrm{E}\{\langle B_H, \psi_{j,n}\rangle \langle B_H, \psi_{l,m}\rangle\} = O\left(2^{j(2H+1)} |n-m|^{2(H-p)}\right) .$$

(c) In two dimensions, synthesize "approximate" fractional Brownian motion images $\tilde{B}_H$ with wavelet coefficients $\langle B_H, \psi^k_{j,n} \rangle$ that are independent Gaussian random variables, whose variances are proportional to $2^{j(2H+2)}$. Adjust H in order to produce textures that look like clouds in the sky.

7.26. [1] *Image mosaic* Let $f_0[n_1, n_2]$ and $f_1[n_1, n_2]$ be two images of N^2 pixels. We want to merge the center of $f_0[n_1, n_2]$ for $N/4 \leq n_1, n_2 < 3N/4$ in the center of f_1. Compute in WAVELAB the wavelet coefficients of f_0 and f_1. At each scale 2^j and orientation $1 \leq k \leq 3$, replace the $2^{-2j}/4$ wavelet coefficients corresponding to the center of f_1 by the wavelet coefficients of f_0. Reconstruct an image from this manipulated wavelet representation. Explain why the image f_0 seems to be merged in f_1, without the strong boundary effects that are obtained when replacing directly the pixels of f_1 by the pixels of f_0.

7.27. [2] *Foveal vision* A foveal image has a maximum resolution at the center, with a resolution that decreases linearly as a function of the distance to the center. Show that one can construct an approximate foveal image by keeping a constant number of non-zero wavelet coefficients at each scale 2^j. Implement this algorithm in WAVELAB. You may build a highly compact image code from such an image representation.

7.28. [1] *High contrast* We consider a color image specified by three color channels: red $r[n]$, green $g[n]$, and blue $b[n]$. The intensity image $(r+g+b)/3$ averages the variations of the three color channels. To create a high contrast image f, for each wavelet $\psi^k_{j,n}$ we set $\langle f, \psi^k_{j,n} \rangle$ to be the coefficient among $\langle r, \psi^k_{j,n} \rangle$, $\langle g, \psi^k_{j,n} \rangle$ and $\langle b, \psi^k_{j,n} \rangle$, which has the maximum amplitude. Implement this algorithm in WAVELAB and evaluate numerically its performance for different types of multispectral images. How does the choice of ψ affect the results?

7.29. [2] *Restoration* Develop an algorithm that restores the sharpness of a smoothed image by increasing the amplitude of wavelet coefficients. Find appropriate amplification functionals depending on the scale and orientation of the wavelet coefficients, in order to increase the image sharpness without introducing important artifacts. To improve the visual quality of the result, study the impact of the wavelet properties: symmetry, vanishing moments and regularity.

7.30. [3] *Smooth extension* Let $f[n]$ be an image whose samples are known only over a domain D, which may be irregular and may include holes. Design and implement an algorithm that computes the wavelet coefficients of a smooth extension $\tilde{f}$ of f over a square domain that includes D, and compute $\tilde{f}$ from these. Choose wavelets with p vanishing moments. Set to zero all coefficients corresponding wavelets whose support do not intersect D, which is equivalent to impose that $\tilde{f}$ is locally a polynomial of degree p. The coefficients of wavelets whose support are in D are calculated from f. The issue is therefore to compute the coefficients of wavelets whose support intersect the boundary of D. You must guarantee that $\tilde{f} = f$ on D as well as the numerical stability of your extension.

VIII

WAVELET PACKET AND LOCAL COSINE BASES

Different types of time-frequency structures are encountered in complex signals such as speech recordings. This motivates the design of bases whose time-frequency properties may be adapted. Wavelet bases are one particular family of bases that represent piecewise smooth signals effectively. Other bases are constructed to approximate different types of signals such as highly oscillatory waveforms.

Orthonormal wavelet packet bases use conjugate mirror filters to divide the frequency axis in separate intervals of various sizes. A discrete signal of size N is decomposed in more than $2^{N/2}$ wavelet packet bases with a fast filter bank algorithm that requires $O(N \log_2 N)$ operations.

If the signal properties change over time, it is preferable to isolate different time intervals with translated windows. Local cosine bases are constructed by multiplying these windows with cosine functions. Wavelet packet and local cosine bases are dual families of bases. Wavelet packets segment the frequency axis and are uniformly translated in time whereas local cosine bases divide the time axis and are uniformly translated in frequency.

8.1 WAVELET PACKETS [2]

8.1.1 Wavelet Packet Tree

Wavelet packets were introduced by Coifman, Meyer and Wickerhauser [139] by generalizing the link between multiresolution approximations and wavelets. A space $\mathbf{V}_j$ of a multiresolution approximation is decomposed in a lower resolution space $\mathbf{V}_{j+1}$ plus a detail space $\mathbf{W}_{j+1}$. This is done by dividing the orthogonal basis $\{\phi_j(t-2^j n)\}_{n\in\mathbb{Z}}$ of $\mathbf{V}_j$ into two new orthogonal bases

$$\{\phi_{j+1}(t-2^{j+1}n)\}_{n\in\mathbb{Z}} \text{ of } \mathbf{V}_{j+1} \quad \text{and} \quad \{\psi_{j+1}(t-2^{j+1}n)\}_{n\in\mathbb{Z}} \text{ of } \mathbf{W}_{j+1}.$$

The decompositions (7.112) and (7.114) of ϕ_{j+1} and ψ_{j+1} in the basis $\{\phi_j(t-2^j n)\}_{n\in\mathbb{Z}}$ are specified by a pair of conjugate mirror filters $h[n]$ and

$$g[n] = (-1)^{1-n} h[1-n].$$

The following theorem generalizes this result to any space $\mathbf{U}_j$ that admits an orthogonal basis of functions translated by $n2^j$, for $n \in \mathbb{Z}$.

Theorem 8.1 (COIFMAN, MEYER, WICKERHAUSER) *Let $\{\theta_j(t-2^j n)\}_{n\in\mathbb{Z}}$ be an orthonormal basis of a space $\mathbf{U}_j$. Let h and g be a pair of conjugate mirror filters. Define*

$$\theta^0_{j+1}(t) = \sum_{n=-\infty}^{+\infty} h[n]\,\theta_j(t-2^j n) \quad \textit{and} \quad \theta^1_{j+1}(t) = \sum_{n=-\infty}^{+\infty} g[n]\,\theta_j(t-2^j n). \tag{8.1}$$

The family

$$\left\{\theta^0_{j+1}(t-2^{j+1}n),\, \theta^1_{j+1}(t-2^{j+1}n)\right\}_{n\in\mathbb{Z}}$$

is an orthonormal basis of $\mathbf{U}_j$.

Proof [2]. This proof is very similar to the proof of Theorem 7.3. The main steps are outlined. The fact that $\{\theta_j(t-2^j n)\}_{n\in\mathbb{Z}}$ is orthogonal means that

$$\frac{1}{2^j}\sum_{k=-\infty}^{+\infty}\left|\hat{\theta}_j\left(\omega+\frac{2k\pi}{2^j}\right)\right|^2 = 1. \tag{8.2}$$

We derive from (8.1) that the Fourier transform of θ^0_{j+1} is

$$\hat{\theta}^0_{j+1}(\omega) = \hat{\theta}_j(\omega)\sum_{n=-\infty}^{+\infty} h[n]\exp(-i2^j n\omega) = \hat{h}(2^j\omega)\,\hat{\theta}_j(\omega). \tag{8.3}$$

Similarly, the Fourier transform of θ^1_{j+1} is

$$\hat{\theta}^1_{j+1}(\omega) = \hat{g}(2^j\omega)\,\hat{\theta}_j(\omega). \tag{8.4}$$

Proving that $\{\theta^0_{j+1}(t-2^{j+1}n)\}$ and $\{\theta^1_{j+1}(t-2^{j+1}n)\}_{n\in\mathbf{Z}}$ are two families of orthogonal vectors is equivalent to showing that for $l=0$ or $l=1$

$$\frac{1}{2^{j+1}}\sum_{k=-\infty}^{+\infty}\left|\hat{\theta}^l_{j+1}\left(\omega+\frac{2k\pi}{2^{j+1}}\right)\right|^2=1. \tag{8.5}$$

These two families of vectors yield orthogonal spaces if and only if

$$\frac{1}{2^{j+1}}\sum_{k=-\infty}^{+\infty}\hat{\theta}^0_{j+1}\left(\omega+\frac{2k\pi}{2^{j+1}}\right)\hat{\theta}^{1*}_{j+1}\left(\omega+\frac{2k\pi}{2^{j+1}}\right)=0. \tag{8.6}$$

The relations (8.5) and (8.6) are verified by replacing $\hat{\theta}^0_{j+1}$ and $\hat{\theta}^1_{j+1}$ by (8.3) and (8.4) respectively, and by using the orthogonality of the basis (8.2) and the conjugate mirror filter properties

$$\begin{aligned}|\hat{h}(\omega)|^2+|\hat{h}(\omega+\pi)|^2 &= 2,\\ |\hat{g}(\omega)|^2+|\hat{g}(\omega+\pi)|^2 &= 2,\\ \hat{g}(\omega)\hat{h}^*(\omega)+\hat{g}(\omega+\pi)\hat{h}^*(\omega+\pi) &= 0.\end{aligned}$$

To prove that the family $\{\theta^0_{j+1}(t-2^{j+1}n)\,,\,\theta^1_{j+1}(t-2^{j+1}n)\}_{n\in\mathbf{Z}}$ generates the same space as $\{\theta_j(t-2^jn)\}_{n\in\mathbf{Z}}$, we must prove that for any $a[n]\in\mathbf{l}^2(\mathbf{Z})$ there exist $b[n]\in\mathbf{l}^2(\mathbf{Z})$ and $c[n]\in\mathbf{l}^2(\mathbf{Z})$ such that

$$\sum_{n=-\infty}^{+\infty}a[n]\,\theta_j(t-2^jn)=\sum_{n=-\infty}^{+\infty}b[n]\,\theta^0_{j+1}(t-2^{j+1}n)+\sum_{n=-\infty}^{+\infty}c[n]\,\theta^1_{j+1}(t-2^{j+1}n). \tag{8.7}$$

To do this, we relate $\hat{b}(\omega)$ and $\hat{c}(\omega)$ to $\hat{a}(\omega)$. The Fourier transform of (8.7) yields

$$\hat{a}(2^j\omega)\,\hat{\theta}_j(\omega)=\hat{b}(2^{j+1}\omega)\,\hat{\theta}^0_{j+1}(\omega)+\hat{c}(2^{j+1}\omega)\,\hat{\theta}^1_{j+1}(\omega). \tag{8.8}$$

One can verify that

$$\hat{b}(2^{j+1}\omega)=\frac{1}{2}\left(\hat{a}(2^j\omega)\,\hat{h}^*(2^j\omega)+\hat{a}(2^j\omega+\pi)\,\hat{h}^*(2^j\omega+\pi)\right)$$

and

$$\hat{c}(2^{j+1}\omega)=\frac{1}{2}\left(\hat{a}(2^j\omega)\,\hat{g}^*(2^j\omega)+\hat{a}(2^j\omega+\pi)\,\hat{g}^*(2^j\omega+\pi)\right)$$

satisfy (8.8). ■

Theorem 8.1 proves that conjugate mirror filters transform an orthogonal basis $\{\theta_j(t-2^jn)\}_{n\in\mathbf{Z}}$ in two orthogonal families $\{\theta^0_{j+1}(t-2^{j+1}n)\}_{n\in\mathbf{Z}}$ and $\{\theta^1_{j+1}(t-2^{j+1}n)\}_{n\in\mathbf{Z}}$. Let $\mathbf{U}^0_{j+1}$ and $\mathbf{U}^1_{j+1}$ be the spaces generated by each of these families. Clearly $\mathbf{U}^0_{j+1}$ and $\mathbf{U}^1_{j+1}$ are orthogonal and

$$\mathbf{U}^0_{j+1}\oplus\mathbf{U}^1_{j+1}=\mathbf{U}_j.$$

Computing the Fourier transform of (8.1) relates the Fourier transforms of θ^0_{j+1} and θ^1_{j+1} to the Fourier transform of θ_j:

$$\hat{\theta}^0_{j+1}(\omega) = \hat{h}(2^j\omega)\,\hat{\theta}_j(\omega) \;\;,\;\; \hat{\theta}^1_{j+1}(\omega) = \hat{g}(2^j\omega)\,\hat{\theta}_j(\omega). \tag{8.9}$$

Since the transfer functions $\hat{h}(2^j\omega)$ and $\hat{g}(2^j\omega)$ have their energy concentrated in different frequency intervals, this transformation can be interpreted as a division of the frequency support of $\hat{\theta}_j$.

Binary Wavelet Packet Tree Instead of dividing only the approximation spaces $\mathbf{V}_j$ to construct detail spaces $\mathbf{W}_j$ and wavelet bases, Theorem 8.1 proves that we can set $\mathbf{U}_j = \mathbf{W}_j$ and divide these detail spaces to derive new bases. The recursive splitting of vector spaces is represented in a binary tree. If the signals are approximated at the scale 2^L, to the root of the tree we associate the approximation space $\mathbf{V}_L$. This space admits an orthogonal basis of scaling functions $\{\phi_L(t - 2^L n)\}_{n\in\mathbb{Z}}$ with $\phi_L(t) = 2^{-L/2}\,\phi(2^{-L}t)$.

Any node of the binary tree is labeled by (j, p), where $j - L \geq 0$ is the depth of the node in the tree, and p is the number of nodes that are on its left at the same depth $j - L$. Such a tree is illustrated in Figure 8.1. To each node (j, p) we associate a space $\mathbf{W}^p_j$, which admits an orthonormal basis $\{\psi^p_j(t - 2^j n)\}_{n\in\mathbb{Z}}$, by going down the tree. At the root, we have $\mathbf{W}^0_L = \mathbf{V}_L$ and $\psi^0_L = \phi_L$. Suppose now that we have already constructed $\mathbf{W}^p_j$ and its orthonormal basis $\mathcal{B}^p_j = \{\psi^p_j(t - 2^j n)\}_{n\in\mathbb{Z}}$ at the node (j, p). The two wavelet packet orthogonal bases at the children nodes are defined by the splitting relations (8.1):

$$\psi^{2p}_{j+1}(t) = \sum_{n=-\infty}^{+\infty} h[n]\,\psi^p_j(t - 2^j n) \tag{8.10}$$

and

$$\psi^{2p+1}_{j+1}(t) = \sum_{n=-\infty}^{+\infty} g[n]\,\psi^p_j(t - 2^j n). \tag{8.11}$$

Since $\{\psi^p_j(t - 2^j n)\}_{n\in\mathbb{Z}}$ is orthonormal,

$$h[n] = \langle \psi^{2p}_{j+1}(u), \psi^p_j(u - 2^j n)\rangle \;\;,\;\; g[n] = \langle \psi^{2p+1}_{j+1}(u), \psi^p_j(u - 2^j n)\rangle. \tag{8.12}$$

Theorem 8.1 proves that $\mathcal{B}^{2p}_{j+1} = \{\psi^{2p}_{j+1}(t - 2^{j+1}n)\}_{n\in\mathbb{Z}}$ and $\mathcal{B}^{2p+1}_{j+1} = \{\psi^{2p+1}_{j+1}(t - 2^{j+1}n)\}_{n\in\mathbb{Z}}$ are orthonormal bases of two orthogonal spaces $\mathbf{W}^{2p}_{j+1}$ and $\mathbf{W}^{2p+1}_{j+1}$ such that

$$\mathbf{W}^{2p}_{j+1} \oplus \mathbf{W}^{2p+1}_{j+1} = \mathbf{W}^p_j. \tag{8.13}$$

This recursive splitting defines a binary tree of wavelet packet spaces where each parent node is divided in two orthogonal subspaces. Figure 8.2 displays the 8

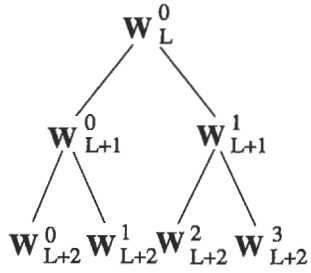


FIGURE 8.1 Binary tree of wavelet packet spaces.

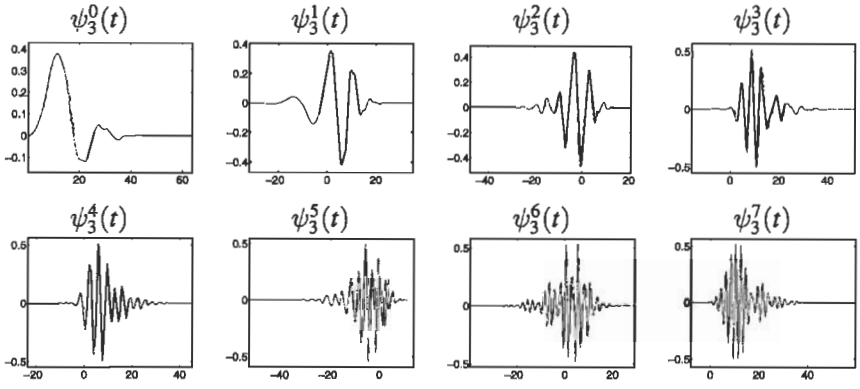


FIGURE 8.2 Wavelet packets computed with the Daubechies 5 filter, at the depth $j-L=3$ of the wavelet packet tree, with $L=0$. They are ordered from low to high frequencies.

wavelet packets ψ_j^p at the depth $j-L=3$, calculated with the Daubechies filter of order 5. These wavelet packets are frequency ordered from left to right, as explained in Section 8.1.2.

Admissible Tree We call *admissible tree* any binary tree where each node has either 0 or 2 children, as shown in Figure 8.3. Let $\{j_i, p_i\}_{1\le i\le I}$ be the leaves of an admissible binary tree. By applying the recursive splitting (8.13) along the branches of an admissible tree, we verify that the spaces $\{\mathbf{W}_{j_i}^{p_i}\}_{1\le i\le I}$ are mutually orthogonal and add up to $\mathbf{W}_L^0$:

$$\mathbf{W}_L^0 = \oplus_{i=1}^I \mathbf{W}_{j_i}^{p_i} . \tag{8.14}$$

The union of the corresponding wavelet packet bases

$$\{\psi_{j_i}^{p_i}(t-2^{j_i}n)\}_{n\in\mathbb{Z},1\le i\le I}$$

thus defines an orthogonal basis of $\mathbf{W}_L^0 = \mathbf{V}_L$.

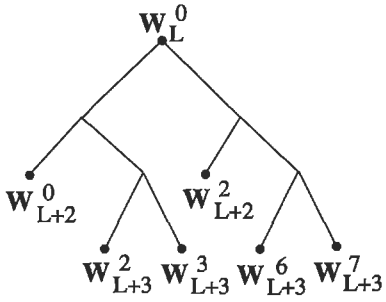


FIGURE 8.3 Example of admissible wavelet packet binary tree.

Number of Wavelet Packet Bases The number of different wavelet packet orthogonal bases of $\mathbf{V}_L$ is equal to the number of different admissible binary trees. The following proposition proves that there are more than $2^{2^{J-1}}$ different wavelet packet orthonormal bases included in a full wavelet packet binary tree of depth J.

Proposition 8.1 *The number B_J of wavelet packet bases in a full wavelet packet binary tree of depth J satisfies*

$$2^{2^{J-1}} \leq B_J \leq 2^{\frac{5}{4}2^{J-1}}. \tag{8.15}$$

Proof[2]. This result is proved by induction on the depth J of the wavelet packet tree. The number B_J of different orthonormal bases is equal to the number of different admissible binary trees of depth at most J, whose nodes have either 0 or 2 children. For $J = 0$, the tree is reduced to its root so $B_0 = 1$.

Observe that the set of trees of depth at most $J+1$ is composed of trees of depth at least 1 and at most $J+1$ plus one tree of depth 0 that is reduced to the root. A tree of depth at least 1 has a left and a right subtree that are admissible trees of depth at most J. The configuration of these trees is a priori independent and there are B_J admissible trees of depth J so

$$B_{J+1} = B_J^2 + 1. \tag{8.16}$$

Since $B_1 = 2$ and $B_{J+1} \geq B_J^2$, we prove by induction that $B_J \geq 2^{2^{J-1}}$. Moreover

$$\log_2 B_{J+1} = 2\log_2 B_J + \log_2(1 + B_J^{-2}).$$

If $J \geq 1$ then $B_J \geq 2$ so

$$\log_2 B_{J+1} \leq 2\log_2 B_J + \frac{1}{4}. \tag{8.17}$$

Since $B_1 = 2$,

$$\log_2 B_{J+1} \leq 2^J + \frac{1}{4}\sum_{j=0}^{J-1} 2^j \leq 2^J + \frac{2^J}{4},$$

so $B_J \leq 2^{\frac{5}{4}2^{J-1}}$. ■

For discrete signals of size N, we shall see that the wavelet packet tree is at most of depth $J = \log_2 N$. This proposition proves that the number of wavelet packet bases satisfies $2^{N/2} \leq B_{\log_2 N} \leq 2^{5N/8}$.

Wavelet Packets on Intervals To construct wavelet packets bases of $\mathbf{L}^2[0,1]$, we use the border techniques developed in Section 7.5 to design wavelet bases of $\mathbf{L}^2[0,1]$. The simplest approach constructs periodic bases. As in the wavelet case, the coefficients of $f \in \mathbf{L}^2[0,1]$ in a periodic wavelet packet basis are the same as the decomposition coefficients of $f^{\text{per}}(t) = \sum_{k=-\infty}^{+\infty} f(t+k)$ in the original wavelet packet basis of $\mathbf{L}^2(\mathbb{R})$. The periodization of f often creates discontinuities at the borders $t=0$ and $t=1$, which generate large amplitude wavelet packet coefficients.

Section 7.5.3 describes a more sophisticated technique which modifies the filters h and g in order to construct boundary wavelets which keep their vanishing moments. A generalization to wavelet packets is obtained by using these modified filters in Theorem 8.1. This avoids creating the large amplitude coefficients at the boundary, typical of the periodic case.

Biorthogonal Wavelet Packets Non-orthogonal wavelet bases are constructed in Section 7.4 with two pairs of perfect reconstruction filters (h,g) and $(\tilde{h},\tilde{g})$ instead of a single pair of conjugate mirror filters. The orthogonal splitting Theorem 8.1 is extended into a biorthogonal splitting by replacing the conjugate mirror filters with these perfect reconstruction filters. A Riesz basis $\{\theta_j(t-2^j n)\}_{n\in\mathbb{Z}}$ of $\mathbf{U}_j$ is transformed into two Riesz bases $\{\theta^0_{j+1}(t-2^{j+1}n)\}_{n\in\mathbb{Z}}$ and $\{\theta^1_{j+1}(t-2^{j+1}n)\}_{n\in\mathbb{Z}}$ of two non-orthogonal spaces $\mathbf{U}^0_{j+1}$ and $\mathbf{U}^1_{j+1}$ such that

$$\mathbf{U}^0_{j+1} \oplus \mathbf{U}^1_{j+1} = \mathbf{U}_j.$$

A binary tree of non-orthogonal wavelet packet Riesz bases can be derived by induction using this vector space division. As in the orthogonal case, the wavelet packets at the leaves of an admissible binary tree define a basis of $\mathbf{W}^0_L$, but this basis is not orthogonal.

The lack of orthogonality is not a problem by itself as long as the basis remains stable. Cohen and Daubechies proved [130] that when the depth $j-L$ increases, the angle between the spaces $\mathbf{W}^p_j$ located at the same depth can become progressively smaller. This indicates that some of the wavelet packet bases constructed from an admissible binary tree become unstable. We thus concentrate on orthogonal wavelet packets constructed with conjugate mirror filters.

8.1.2 Time-Frequency Localization

Time Support If the conjugate mirror filters h and g have a finite impulse response of size K, Proposition 7.2 proves that ϕ has a support of size $K-1$ so $\psi^0_L = \phi_L$ has a support of size $(K-1)2^L$. Since

$$\psi^{2p}_{j+1}(t) = \sum_{n=-\infty}^{+\infty} h[n]\,\psi^p_j(t-2^j n) \;,\; \psi^{2p+1}_{j+1}(t) = \sum_{n=-\infty}^{+\infty} g[n]\,\psi^p_j(t-2^j n), \quad (8.18)$$

an induction on j shows that the support size of ψ^p_j is $(K-1)2^j$. The parameter j thus specifies the scale 2^j of the support. The wavelet packets in Figure 8.2 are

constructed with a Daubechies filter of $K = 10$ coefficients with $j = 3$ and thus have a support of size $2^3(10-1) = 72$.

Frequency Localization The frequency localization of wavelet packets is more complicated to analyze. The Fourier transform of (8.18) proves that the Fourier transforms of wavelet packet children are related to their parent by

$$\hat{\psi}_{j+1}^{2p}(\omega) = \hat{h}(2^j\omega)\,\hat{\psi}_j^p(\omega) \quad , \quad \hat{\psi}_{j+1}^{2p+1}(\omega) = \hat{g}(2^j\omega)\,\hat{\psi}_j^p(\omega). \tag{8.19}$$

The energy of $\hat{\psi}_j^p$ is mostly concentrated over a frequency band and the two filters $\hat{h}(2^j\omega)$ and $\hat{g}(2^j\omega)$ select the lower or higher frequency components within this band. To relate the size and position of this frequency band to the indexes (p, j), we consider a simple example.

Shannon Wavelet Packets Shannon wavelet packets are computed with perfect discrete low-pass and high-pass filters

$$|\hat{h}(\omega)| = \begin{cases} \sqrt{2} & \text{if } \omega \in [-\pi/2 + 2k\pi, \pi/2 + 2k\pi] \text{ with } k \in \mathbb{Z} \\ 0 & \text{otherwise} \end{cases} \tag{8.20}$$

and

$$|\hat{g}(\omega)| = \begin{cases} \sqrt{2} & \text{if } \omega \in [\pi/2 + 2k\pi, 3\pi/2 + 2k\pi] \text{ with } k \in \mathbb{Z} \\ 0 & \text{otherwise} \end{cases}. \tag{8.21}$$

In this case it is relatively simple to calculate the frequency support of the wavelet packets. The Fourier transform of the scaling function is

$$\hat{\psi}_L^0 = \hat{\phi}_L = \mathbf{1}_{[-2^{-L}\pi, 2^{-L}\pi]}. \tag{8.22}$$

Each multiplication with $\hat{h}(2^j\omega)$ or $\hat{g}(2^j\omega)$ divides the frequency support of the wavelet packets in two. The delicate point is to realize that $\hat{h}(2^j\omega)$ does not always play the role of a low-pass filter because of the side lobes that are brought into the interval $[-2^{-L}\pi, 2^{-L}\pi]$ by the dilation. At the depth $j - L$, the following proposition proves that $\hat{\psi}_j^p$ is proportional to the indicator function of a pair of frequency intervals, that are labeled I_j^k. The permutation that relates p and k is characterized recursively [76].

Proposition 8.2 (COIFMAN, WICKERHAUSER) *For any $j - L > 0$ and $0 \leq p < 2^{j-L}$, there exists $0 \leq k < 2^{j-L}$ such that*

$$|\hat{\psi}_j^p(\omega)| = 2^{j/2}\,\mathbf{1}_{I_j^k}(\omega), \tag{8.23}$$

where I_j^k is a symmetric pair of intervals

$$I_j^k = [-(k+1)\pi 2^{-j}, -k\pi 2^{-j}] \cup [k\pi 2^{-j}, (k+1)\pi 2^{-j}]. \tag{8.24}$$

The permutation $k = G[p]$ satisfies for any $0 \leq p < 2^{j-L}$

$$G[2p] = \begin{cases} 2G[p] & \text{if } G[p] \text{ is even} \\ 2G[p]+1 & \text{if } G[p] \text{ is odd} \end{cases} \tag{8.25}$$

$$G[2p+1] = \begin{cases} 2G[p]+1 & \text{if } G[p] \text{ is even} \\ 2G[p] & \text{if } G[p] \text{ is odd} \end{cases} \tag{8.26}$$

Proof[3]. The three equations (8.23), (8.25) and (8.26) are proved by induction on the depth $j-L$. For $j-L=0$, (8.22) shows that (8.23) is valid. Suppose that (8.23) is valid for $j=l \geq L$ and any $0 \leq p < 2^{l-L}$. We first prove that (8.25) and (8.26) are verified for $j=l$. From these two equations we then easily carry the induction hypothesis to prove that (8.23) is true for $j=l+1$ and for any $0 \leq p < 2^{l+1-L}$.

Equations (8.20) and (8.21) imply that

$$|\hat{h}(2^l\omega)| = \begin{cases} \sqrt{2} & \text{if } \omega \in [-2^{-l-1}(4m-1)\pi, 2^{-l-1}(4m+1)\pi] \text{ with } m \in \mathbb{Z} \\ 0 & \text{otherwise} \end{cases} \tag{8.27}$$

$$|\hat{g}(2^l\omega)| = \begin{cases} \sqrt{2} & \text{if } \omega \in [-2^{-l-1}(4m+1)\pi, 2^{-l-1}(4m+3)\pi] \text{ with } m \in \mathbb{Z} \\ 0 & \text{otherwise} \end{cases} \tag{8.28}$$

Since (8.23) is valid for l, the support of $\hat{\psi}_l^p$ is

$$I_l^k = [-(2k+2)\pi 2^{-l-1}, -2k\pi 2^{-l-1}] \cup [2k\pi 2^{-l-1}, (2k+2)\pi 2^{-l-1}].$$

The two children are defined by

$$\hat{\psi}_{l+1}^{2p}(\omega) = \hat{h}(2^l\omega)\hat{\psi}_l^p(\omega) \ , \ \hat{\psi}_{l+1}^{2p+1}(\omega) = \hat{g}(2^l\omega)\hat{\psi}_l^p(\omega) \ .$$

We thus derive (8.25) and (8.26) by checking the intersection of I_l^k with the supports of $\hat{h}(2^j\omega)$ and $\hat{g}(2^j\omega)$ specified by (8.27) and (8.28). ■

For Shannon wavelet packets, Proposition 8.2 proves that $\hat{\psi}_j^p$ has a frequency support located over two intervals of size $2^{-j}\pi$, centered at $\pm(k+1/2)\pi 2^{-j}$. The Fourier transform expression (8.23) implies that these Shannon wavelet packets can be written as cosine modulated windows

$$\psi_j^p(t) = 2^{-j/2+1}\,\theta(2^{-j}t)\cos\left[2^{-j}\pi(k+1/2)(t-\tau_{j,p})\right] , \tag{8.29}$$

with

$$\theta(t) = \frac{\sin(\pi t/2)}{\pi t} \quad \text{and hence} \quad \hat{\theta}(\omega) = \mathbf{1}_{[-\pi/2,\pi/2]}(\omega).$$

The translation parameter $\tau_{j,p}$ can be calculated from the complex phase of $\hat{\psi}_j^p$.

Frequency Ordering It is often easier to label ψ_j^k a wavelet packet ψ_j^p whose Fourier transform is centered at $\pm(k+1/2)\pi 2^{-j}$, with $k = G[p]$. This means changing its position in the wavelet packet tree from the node p to the node k. The resulting wavelet packet tree is frequency ordered. The left child always

corresponds to a lower frequency wavelet packet and the right child to a higher frequency one.

The permutation $k = G[p]$ is characterized by the recursive equations (8.25) and (8.26). The inverse permutation $p = G^{-1}[k]$ is called a *Gray code* in coding theory. This permutation is implemented on binary strings by deriving the following relations from (8.25) and (8.26). If p_i is the i^{th} binary digit of the integer p and k_i the i^{th} digit of $k = G[p]$ then

$$k_i = \left(\sum_{l=i}^{+\infty} p_l \right) \bmod 2, \tag{8.30}$$

and

$$p_i = (k_i + k_{i+1}) \bmod 2. \tag{8.31}$$

Compactly Supported Wavelet Packets Wavelet packets of compact support have a more complicated frequency behavior than Shannon wavelet packets, but the previous analysis provides important insights. If h is a finite impulse response filter, $\hat{h}$ does not have a support restricted to $[-\pi/2, \pi/2]$ over the interval $[-\pi, \pi]$. It is however true that the energy of $\hat{h}$ is mostly concentrated in $[-\pi/2, \pi/2]$. Similarly, the energy of $\hat{g}$ is mostly concentrated in $[-\pi, -\pi/2] \cup [\pi/2, \pi]$, for $\omega \in [-\pi, \pi]$. As a consequence, the localization properties of Shannon wavelet packets remain qualitatively valid. The energy of $\hat{\psi}_j^p$ is mostly concentrated over

$$I_j^k = [-(k+1)\pi 2^{-j}, -k\pi 2^{-j}] \cup [k\pi 2^{-j}, (k+1)\pi 2^{-j}],$$

with $k = G[p]$. The larger the proportion of energy of $\hat{h}$ in $[-\pi/2, \pi/2]$, the more concentrated the energy of $\hat{\psi}_j^p$ in I_j^k. The energy concentration of $\hat{h}$ in $[-\pi/2, \pi/2]$ is increased by having more zeroes at π, so that $\hat{h}(\omega)$ remains close to zero in $[-\pi, -\pi/2] \cup [\pi/2, \pi]$. Theorem 7.4 proves that this is equivalent to imposing that the wavelets constructed in the wavelet packet tree have many vanishing moments.

These qualitative statements must be interpreted carefully. The side lobes of $\hat{\psi}_j^p$ beyond the intervals I_j^k are not completely negligible. For example, wavelet packets created with a Haar filter are discontinuous functions. Hence $|\hat{\psi}_j^p(\omega)|$ decays like $|\omega|^{-1}$ at high frequencies, which indicates the existence of large side lobes outside I_k^p. It is also important to note that contrary to Shannon wavelet packets, compactly supported wavelet packets cannot be written as dilated windows modulated by cosine functions of varying frequency. When the scale increases, wavelet packets generally do not converge to cosine functions. They may have a wild behavior with localized oscillations of considerable amplitude.

Walsh Wavelet Packets Walsh wavelet packets are generated by the Haar conjugate mirror filter

$$h[n] = \begin{cases} \frac{1}{\sqrt{2}} & \text{if } n = 0, 1 \\ 0 & \text{otherwise} \end{cases}.$$

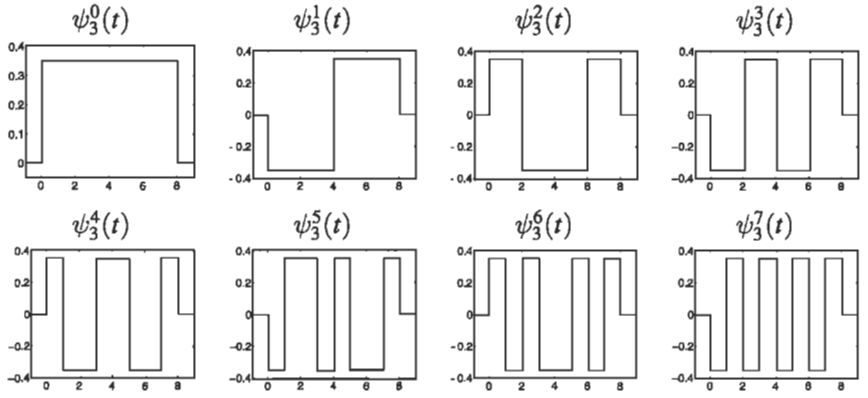


FIGURE 8.4 Frequency ordered Walsh wavelet packets computed with a Haar filter, at the depth $j-L=3$ of the wavelet packet tree, with $L=0$.

They have very different properties from Shannon wavelet packets since the filter h is well localized in time but not in frequency. The corresponding scaling function is $\phi = \mathbf{1}_{[0,1]}$ and the approximation space $\mathbf{V}_L = \mathbf{W}_L^0$ is composed of functions that are constant over the intervals $[2^L n, 2^L(n+1))$, for $n \in \mathbb{Z}$. Since all wavelet packets created with this filter belong to $\mathbf{V}_L$, they are piecewise constant functions. The support size of h is $K=2$, so Walsh functions ψ_j^p have a support of size 2^j. The wavelet packet recursive relations (8.18) become

$$\psi_{j+1}^{2p}(t) = \frac{1}{\sqrt{2}}\psi_j^p(t) + \frac{1}{\sqrt{2}}\psi_j^p(t-2^j), \tag{8.32}$$

and

$$\psi_{j+1}^{2p+1}(t) = \frac{1}{\sqrt{2}}\psi_j^p(t) - \frac{1}{\sqrt{2}}\psi_j^p(t-2^j). \tag{8.33}$$

Since ψ_j^p has a support of size 2^j, it does not intersect the support of $\psi_j^p(t-2^j)$. These wavelet packets are thus constructed by juxtaposing ψ_j^p with its translated version whose sign might be changed. Figure 8.4 shows the Walsh functions at the depth $j-L=3$ of the wavelet packet tree. The following proposition computes the number of oscillations of ψ_j^p.

Proposition 8.3 *The support of a Walsh wavelet packet ψ_j^p is $[0, 2^j]$. Over its support, $\psi_j^p(t) = \pm 2^{-j/2}$. It changes sign $k = G[p]$ times, where $G[p]$ is the permutation defined by (8.25) and (8.26).*

Proof[2]. By induction on j, we derive from (8.32) and (8.33) that the support is $[0, 2^j]$ and that $\psi_j^p(t) = \pm 2^{-j/2}$ over its support. Let k be the number of times that ψ_j^p changes sign. The number of times that ψ_{j+1}^{2p} and ψ_{j+1}^{2p+1} change sign is either $2k$ or $2k+1$ depending on the sign of the first and last non-zero values of ψ_j^p. If k is even, then

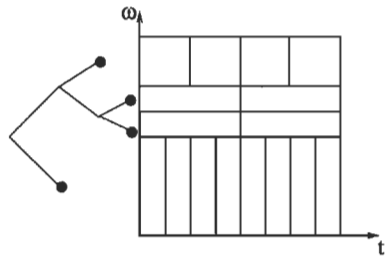


FIGURE 8.5 The wavelet packet tree on the left divides the frequency axis in several intervals. The Heisenberg boxes of the corresponding wavelet packet basis are on the right.

the sign of the first and last non-zero values of ψ_j^p are the same. Hence the number of times ψ_{j+1}^{2p} and ψ_{j+1}^{2p+1} change sign is respectively $2k$ and $2k+1$. If k is odd, then the sign of the first and last non-zero values of ψ_j^p are different. The number of times ψ_{j+1}^{2p} and ψ_{j+1}^{2p+1} change sign is then $2k+1$ and $2k$. These recursive properties are identical to (8.25) and (8.26). ■

A Walsh wavelet packet ψ_j^p is therefore a square wave with $k = G[p]$ oscillations over a support of size 2^j. This result is similar to (8.29), which proves that a Shannon wavelet packet ψ_j^p is a window modulated by a cosine of frequency $2^{-j}k\pi$. In both cases, the oscillation frequency of wavelet packets is proportional to $2^{-j}k$.

Heisenberg Boxes For display purposes, we associate to any wavelet packet $\psi_j^p(t-2^jn)$ a Heisenberg rectangle which indicates the time and frequency domains where the energy of this wavelet packet is mostly concentrated. The time support of the rectangle is set to be the same as the time support of a Walsh wavelet packet $\psi_j^p(t-2^jn)$, which is equal to $[2^jn, 2^j(n+1)]$. The frequency support of the rectangle is defined as the positive frequency support $[k\pi 2^{-j}, (k+1)\pi 2^{-j}]$ of Shannon wavelet packets, with $k = G[p]$. The scale 2^j modifies the time and frequency elongation of this time-frequency rectangle, but its surface remains constant. The indices n and k give its localization in time and frequency. General wavelet packets, for example computed with Daubechies filters, have a time and a frequency spread that is much wider than this Heisenberg rectangle. However, this convention has the advantage of associating a wavelet packet basis to an exact paving of the time-frequency plane. Figure 8.5 shows an example of such a paving and the corresponding wavelet packet tree.

Figure 8.6 displays the decomposition of a multi-chirp signal whose spectrogram was shown in Figure 4.3. The wavelet packet basis is computed with the Daubechies 10 filter. As expected, the coefficients of large amplitude are along the trajectory of the linear and the quadratic chirps that appear in Figure 4.3. We

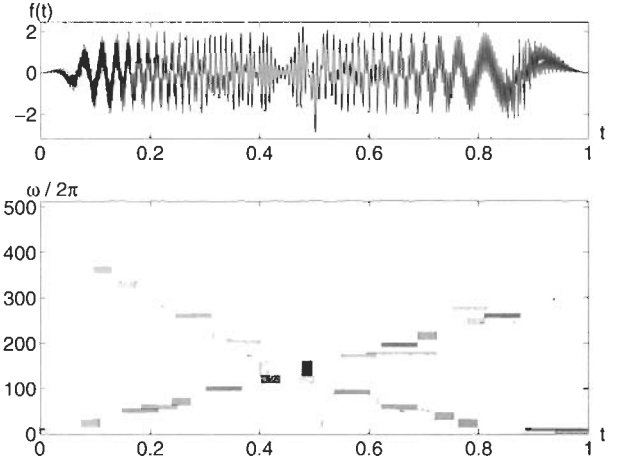


FIGURE 8.6 Wavelet packet decomposition of the multi-chirp signal whose spectrogram is shown in Figure 4.3. The darker the gray level of each Heisenberg box the larger the amplitude $|\langle f, \psi_j^p \rangle|$ of the corresponding wavelet packet coefficient.

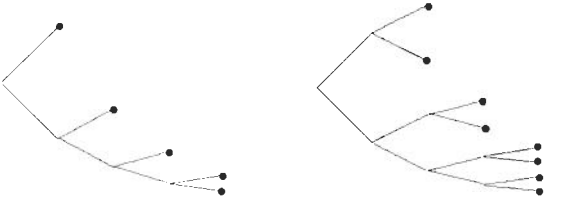


FIGURE 8.7 (a): Wavelet packet tree of a dyadic wavelet basis. (b): Wavelet packet tree of an M-band wavelet basis with $M = 2$.

also see the trace of the two modulated Gaussian functions located at $t = 512$ and $t = 896$.

8.1.3 Particular Wavelet Packet Bases

Among the many wavelet packet bases, we describe the properties of M-band wavelet bases, "local cosine type" bases and "best" bases. The wavelet packet tree is frequency ordered, which means that ψ_j^k has a Fourier transform whose energy is essentially concentrated in the interval $[k\pi 2^{-j}, (k+1)\pi 2^{-j}]$, for positive frequencies.

M-band Wavelets The standard dyadic wavelet basis is an example of a wavelet packet basis of $\mathbf{V}_L$, obtained by choosing the admissible binary tree shown in Figure 8.7(a). Its leaves are the nodes $k = 1$ at all depth $j - L$ and thus correspond to the wavelet packet basis

$$\{\psi_j^1(t - 2^j n)\}_{n \in \mathbb{Z}, j > L}$$

constructed by dilating a single wavelet ψ^1 :

$$\psi_j^1(t) = \frac{1}{\sqrt{2^j}} \psi^1\left(\frac{t}{2^j}\right).$$

The energy of $\hat{\psi}^1$ is mostly concentrated in the interval $[-2\pi, -\pi] \cup [\pi, 2\pi]$. The octave bandwidth for positive frequencies is the ratio between the bandwidth of the pass band and its distance to the zero frequency. It is equal to 1 octave. This quantity remains constant by dilation and specifies the frequency resolution of the wavelet transform.

Wavelet packets include other wavelet bases constructed with several wavelets having a better frequency resolution. Let us consider the admissible binary tree of Figure 8.7(b), whose leaves are indexed by $k = 2$ and $k = 3$ at all depth $j - L$. The resulting wavelet packet basis of $\mathbf{V}_L$ is

$$\{\psi_j^2(t - 2^j n), \psi_j^3(t - 2^j n)\}_{n \in \mathbb{Z}, j > L+1}.$$

These wavelet packets can be rewritten as dilations of two elementary wavelets ψ^2 and ψ^3:

$$\psi_j^2(t) = \frac{1}{\sqrt{2^{j-1}}} \psi^2\left(\frac{t}{2^{j-1}}\right) \quad , \quad \psi_j^3(t) = \frac{1}{\sqrt{2^{j-1}}} \psi^3\left(\frac{t}{2^{j-1}}\right).$$

Over positive frequencies, the energy of $\hat{\psi}^2$ and $\hat{\psi}^3$ is mostly concentrated respectively in $[\pi, 3\pi/2]$ and $[3\pi/2, 2\pi]$. The octave bandwidths of $\hat{\psi}^2$ and $\hat{\psi}^3$ are thus respectively equal to $1/2$ and $1/3$. These wavelets ψ^2 and ψ^3 have a higher frequency resolution than ψ^1, but their time support is twice as large. Figure 8.8(a) gives a 2-band wavelet decomposition of the multi-chirp signal shown in Figure 8.6, calculated with the Daubechies 10 filter.

Higher resolution wavelet bases can be constructed with an arbitrary number of $M = 2^l$ wavelets. In a frequency ordered wavelet packet tree, we define an admissible binary tree whose leaves are the indexes $2^l \leq k < 2^{l+1}$ at the depth $j - L > l$. The resulting wavelet packet basis

$$\{\psi_j^k(t - 2^j n)\}_{M \leq k < 2M, j > L+l}$$

can be written as dilations and translations of M elementary wavelets

$$\psi_j^k(t) = \frac{1}{\sqrt{2^{j-l}}} \psi^k\left(\frac{t}{2^{j-l}}\right).$$

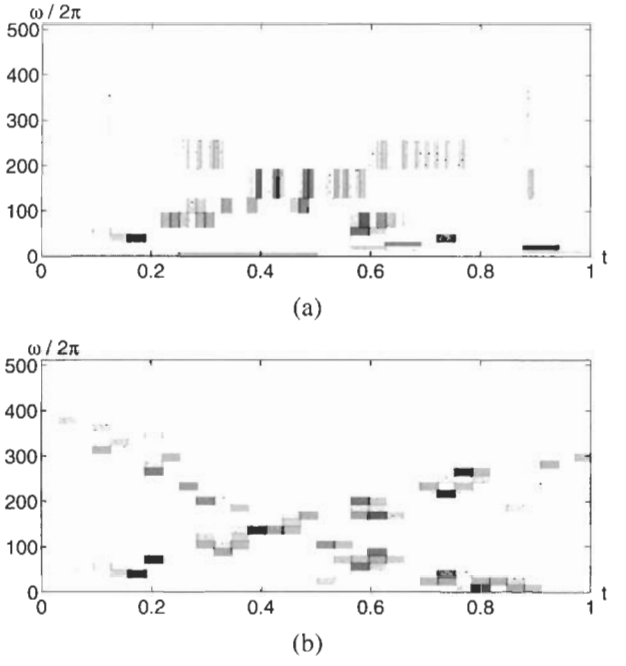


FIGURE 8.8 (a): Heisenberg boxes of a 2-band wavelet decomposition of the multi-chirp signal shown in Figure 8.6. (b): Decomposition of the same signal in a pseudo-local cosine wavelet packet basis.

The support size of ψ^k is proportional to $M = 2^l$. Over positive frequencies, the energy of $\hat{\psi}^k$ is mostly concentrated in $[k\pi 2^{-l}, (k+1)\pi 2^{-l}]$. The octave bandwidth is therefore $\pi 2^{-l}/(k\pi 2^{-l}) = k^{-1}$, for $M \leq k < 2M$. The M wavelets $\{\psi^k\}_{M \leq k < 2M}$ have an octave bandwidth smaller than M^{-1} but a time support M times larger than the support of ψ^1. Such wavelet bases are called M-band wavelets. More general families of M-band wavelets can also be constructed with other M-band filter banks studied in [73].

Pseudo Local Cosine Bases Pseudo local cosine bases are constructed with an admissible binary tree which is a full tree of depth $J - L \geq 0$. The leaves are the nodes indexed by $0 \leq k < 2^{J-L}$ and the resulting wavelet packet basis is

$$\{\psi_J^k(t - 2^J n)\}_{n \in \mathbb{Z}, 0 \leq k < 2^{J-L}}. \tag{8.34}$$

If these wavelet packets are constructed with a conjugate mirror filter of size K, they have a support of size $(K-1)2^J$. Over positive frequencies, the energy of $\hat{\psi}_j^k$ is concentrated in $[k\pi 2^{-J}, (k+1)\pi 2^{-J}]$. The bandwidth of all these wavelet packets is therefore approximately constant and equal to $\pi 2^{-J}$. The Heisenberg

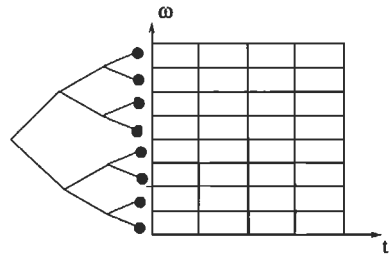


FIGURE 8.9 Admissible tree and Heisenberg boxes of a wavelet packet pseudo local cosine basis.

boxes of these wavelet packets have the same size and divide the time-frequency plane in the rectangular grid illustrated in Figure 8.9.

Shannon wavelet packets ψ_j^k are written in (8.29) as a dilated window θ modulated by cosine functions of frequency $2^{-J}(k+1/2)\pi$. In this case, the uniform wavelet packet basis (8.34) is therefore a local cosine basis, with windows of constant size. This result is not valid for wavelet packets constructed with different conjugate mirror filters. Nevertheless, the time and frequency resolution of uniform wavelet packet bases (8.34) remains constant, like that of local cosine bases constructed with windows of constant size. Figure 8.8(b) gives the decomposition coefficients of a signal in such a uniform wavelet packet basis.

Best Basis Applications of orthogonal bases often rely on their ability to efficiently approximate signals with only a few non-zero vectors. Choosing a wavelet packet basis that concentrates the signal energy over a few coefficients also reveals its time-frequency structures. Section 9.4.2 describes a fast algorithm that searches for a "best" basis that minimizes a Schur concave cost function, among all wavelet packet bases. The wavelet packet basis of Figure 8.6 is calculated with this best basis search.

8.1.4 Wavelet Packet Filter Banks

Wavelet packet coefficients are computed with a filter bank algorithm that generalizes the fast discrete wavelet transform. This algorithm is a straightforward iteration of the two-channel filter bank decomposition presented in Section 7.3.2. It was therefore used in signal processing by Croisier, Esteban and Galand [141] when they introduced the first family of perfect reconstruction filters. The algorithm is presented here from a wavelet packet point of view.

To any discrete signal input $b[n]$ sampled at intervals $N^{-1} = 2^L$, like in (7.116) we associate $f \in \mathbf{V}_L$ whose decomposition coefficients $a_L[n] = \langle f, \phi_{L,n} \rangle$ satisfy

$$b[n] = N^{1/2} a_L[n] \approx f(N^{-1}n) \,. \tag{8.35}$$

For any node (j,p) of the wavelet packet tree, we denote the wavelet packet

coefficients

$$d_j^p[n] = \langle f(t), \psi_j^p(t - 2^j n) \rangle.$$

At the root of the tree $d_L^0[n] = a_L[n]$ is computed from $b[n]$ with (8.35).

Wavelet Packet Decomposition We denote $\bar{x}[n] = x[-n]$ and by $\check{x}$ the signal obtained by inserting a zero between each sample of x. The following proposition generalizes the fast wavelet transform Theorem 7.7.

Proposition 8.4 *At the decomposition*

$$d_{j+1}^{2p}[k] = d_j^p \star \bar{h}[2k] \text{ and } d_{j+1}^{2p+1}[k] = d_j^p \star \bar{g}[2k]. \tag{8.36}$$

At the reconstruction

$$d_j^p[k] = \check{d}_{j+1}^{2p} \star h[k] + \check{d}_{j+1}^{2p+1} \star g[k]. \tag{8.37}$$

The proof of these equations is identical to the proof of Theorem 7.7. The coefficients of wavelet packet children d_{j+1}^{2p} and d_{j+1}^{2p+1} are obtained by subsampling the convolutions of d_j^p with $\bar{h}$ and $\bar{g}$. Iterating these equations along the branches of a wavelet packet tree computes all wavelet packet coefficients, as illustrated by Figure 8.10(a). From the wavelet packet coefficients at the leaves $\{j_i, p_i\}_{1 \le i \le I}$ of an admissible subtree, we recover d_L^0 at the top of the tree by computing (8.37) for each node inside the tree, as illustrated by Figure 8.10(b).

Finite Signals If a_L is a finite signal of size $2^{-L} = N$, we are facing the same border convolution problems as in a fast discrete wavelet transform. One approach explained in Section 7.5.1 is to periodize the wavelet packet basis. The convolutions (8.36) are then replaced by circular convolutions. To avoid introducing sharp transitions with the periodization, one can also use the border filters described in Section 7.5.3. In either case, d_j^p has 2^{-j} samples. At any depth $j - L$ of the tree, the wavelet packet signals $\{d_j^p\}_{0 \le p < 2^{j-L}}$ include a total of N coefficients. Since the maximum depth is $\log_2 N$, there are at most $N \log_2 N$ coefficients in a full wavelet packet tree.

In a full wavelet packet tree of depth $\log_2 N$, all coefficients are computed by iterating (8.36) for $L \le j < 0$. If h and g have K non-zero coefficients, this requires $KN \log_2 N$ additions and multiplications. This is quite spectacular since there are more than $2^{N/2}$ different wavelet packet bases included in this wavelet packet tree.

The computational complexity to recover $a_L = d_L^0$ from the wavelet packet coefficients of an admissible tree increases with the number of inside nodes of the admissible tree. When the admissible tree is the full binary tree of depth $\log_2 N$, the number of operations is maximum and equal to $KN \log_2 N$ multiplications and additions. If the admissible subtree is a wavelet tree, we need fewer than $2KN$ multiplications and additions.

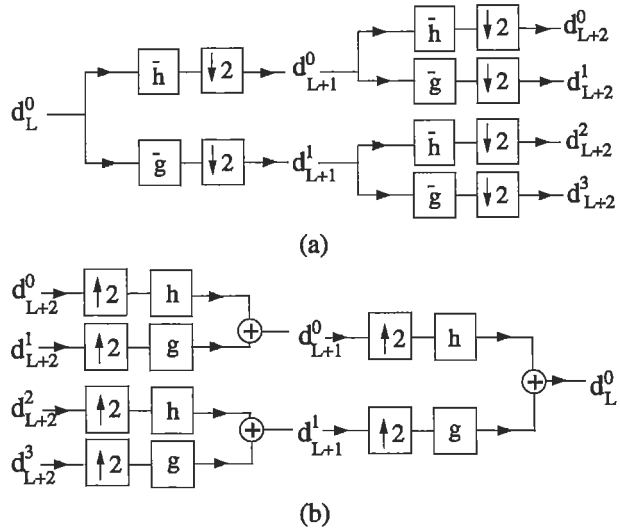


FIGURE 8.10 (a): Wavelet packet filter-bank decomposition with successive filterings and subsamplings. (b): Reconstruction by inserting zeros and filtering the outputs.

Discrete Wavelet Packet Bases of $\mathbf{l}^2(\mathbb{Z})$ The signal decomposition in a conjugate mirror filter bank can also be interpreted as an expansion in discrete wavelet packet bases of $\mathbf{l}^2(\mathbb{Z})$. This is proved with a result similar to Theorem 8.1.

Theorem 8.2 *Let $\{\theta_j[m-2^{j-L}n]\}_{n\in\mathbb{Z}}$ be an orthonormal basis of a space $\mathbf{U}_j$, with $j-L\in\mathbb{N}$. Define*

$$\theta^0_{j+1}[m]=\sum_{n=-\infty}^{+\infty} h[n]\,\theta_j[m-2^{j-L}n]\ ,\ \theta^1_{j+1}[m]=\sum_{n=-\infty}^{+\infty} g[n]\,\theta_j[m-2^{j-L}n]. \quad (8.38)$$

The family

$$\left\{\theta^0_{j+1}[m-2^{j+1-L}n]\,,\,\theta^1_{j+1}[m-2^{j+1-L}n]\right\}_{n\in\mathbb{Z}}$$

is an orthonormal basis of $\mathbf{U}_j$.

The proof is similar to the proof of Theorem 8.1. As in the continuous time case, we derive from this theorem a binary tree of discrete wavelet packets. At the root of the discrete wavelet packet tree is the space $\mathbf{W}^0_L=\mathbf{l}^2(\mathbb{Z})$ of discrete signals obtained with a sampling interval $N^{-1}=2^L$. It admits a canonical basis of Diracs $\{\psi^0_L[m-n]=\delta[m-n]\}_{n\in\mathbb{Z}}$. The signal $a_L[m]$ is specified by its sample values in this basis. One can verify that the convolutions and subsamplings (8.36) compute

$$d^p_j[n]=\langle a_L[m],\psi^p_j[m-2^{j-L}n]\rangle,$$

where $\{\psi_j^p[m-2^{j-L}n]\}_{n\in\mathbb{Z}}$ is an orthogonal basis of a space $\mathbf{W}_j^p$. These discrete wavelet packets are recursively defined for any $j \geq L$ and $0 \leq p < 2^{j-L}$ by

$$\psi_{j+1}^{2p}[m] = \sum_{n=-\infty}^{+\infty} h[n]\,\psi_j^p[m-2^{j-L}n] \;,\; \psi_{j+1}^{2p+1}[m] = \sum_{n=-\infty}^{+\infty} g[n]\,\psi_j^p[m-2^{j-L}n]. \tag{8.39}$$

8.2 IMAGE WAVELET PACKETS [2]

8.2.1 Wavelet Packet Quad-Tree

We construct wavelet packet bases of $\mathbf{L}^2(\mathbb{R}^2)$ whose elements are separable products of wavelet packets $\psi_j^p(x_1-2^jn_1)\,\psi_j^q(x_2-2^jn_2)$ having the same scale along x_1 and x_2. These separable wavelet packet bases are associated to quad-trees, and divide the two-dimensional Fourier plane (ω_1,ω_2) into square regions of varying sizes. Separable wavelet packet bases are extensions of separable wavelet bases.

If images are approximated at the scale 2^L, to the root of the quad-tree we associate the approximation space $\mathbf{V}_L^2 = \mathbf{V}_L \otimes \mathbf{V}_L \subset \mathbf{L}^2(\mathbb{R}^2)$ defined in Section 7.7.1. Section 8.1.1 explains how to decompose $\mathbf{V}_L$ with a binary tree of wavelet packet spaces $\mathbf{W}_j^p \subset \mathbf{V}_L$, which admit an orthogonal basis $\{\psi_j^p(t-2^jn)\}_{n\in\mathbb{Z}}$. The two-dimensional wavelet packet quad-tree is composed of separable wavelet packet spaces. Each node of this quad-tree is labeled by a scale 2^j and two integers $0 \leq p < 2^{j-L}$ and $0 \leq q < 2^{j-L}$, and corresponds to a separable space

$$\mathbf{W}_j^{p,q} = \mathbf{W}_j^p \otimes \mathbf{W}_j^q. \tag{8.40}$$

The resulting separable wavelet packet for $x = (x_1,x_2)$ is

$$\psi_j^{p,q}(x) = \psi_j^p(x_1)\,\psi_j^q(x_2)\;.$$

Theorem A.3 proves that an orthogonal basis of $\mathbf{W}_j^{p,q}$ is obtained with a separable product of the wavelet packet bases of $\mathbf{W}_j^p$ and $\mathbf{W}_j^q$, which can be written

$$\left\{\psi_j^{p,q}(x-2^jn)\right\}_{n\in\mathbb{Z}^2}\;.$$

At the root $\mathbf{W}_L^{0,0} = \mathbf{V}_L^2$ and the wavelet packet is a two-dimensional scaling function

$$\psi_L^{0,0}(x) = \phi_L^2(x) = \phi_L(x_1)\,\phi_L(x_2)\;.$$

One-dimensional wavelet packet spaces satisfy

$$\mathbf{W}_j^p = \mathbf{W}_{j+1}^{2p} \oplus \mathbf{W}_{j+1}^{2p+1} \quad \text{and} \quad \mathbf{W}_j^q = \mathbf{W}_{j+1}^{2q} \oplus \mathbf{W}_{j+1}^{2q+1}.$$

Inserting these equations in (8.40) proves that $\mathbf{W}_j^{p,q}$ is the direct sum of the four orthogonal subspaces

$$\mathbf{W}_j^{p,q} = \mathbf{W}_{j+1}^{2p,2q} \oplus \mathbf{W}_j^{2p+1,2q} \oplus \mathbf{W}_{j+1}^{2p,2q+1} \oplus \mathbf{W}_{j+1}^{2p+1,2q+1}. \tag{8.41}$$

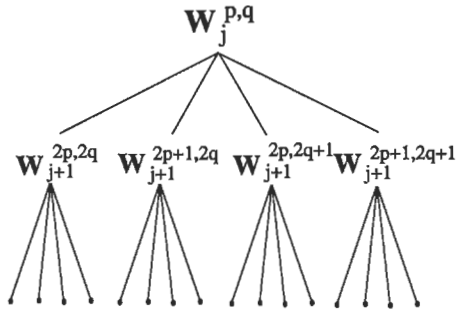


FIGURE 8.11 A wavelet packet quad-tree for images is constructed recursively by decomposing each separable space $\mathbf{W}_j^{p,q}$ in four subspaces.

These subspaces are located at the four children nodes in the quad-tree, as shown by Figure 8.11. We call *admissible quad-tree* any quad-tree whose nodes have either 0 or 4 children. Let $\{j_i, p_i, q_i\}_{0 \le i \le I}$ be the indices of the nodes at the leaves of an admissible quad-tree. Applying recursively the reconstruction sum (8.41) along the branches of this quad-tree gives an orthogonal decomposition of $\mathbf{W}_L^{0,0}$:

$$\mathbf{W}_L^{0,0} = \oplus_{i=1}^{I} \mathbf{W}_{j_i}^{p_i,q_i}.$$

The union of the corresponding wavelet packet bases

$$\left\{ \psi_{j_i}^{p_i,q_i}(x - 2^{j_i} n) \right\}_{(n_1,n_2) \in \mathbb{Z}^2, 1 \le i \le I}$$

is therefore an orthonormal basis of $\mathbf{V}_L^2 = \mathbf{W}_L^{0,0}$.

Number of Wavelet Packet Bases The number of different bases in a full wavelet packet quad-tree of depth J is equal to the number of admissible subtrees. The following proposition proves that there are more than $2^{4^{J-1}}$ such bases.

Proposition 8.5 *The number B_J of wavelet packet bases in a full wavelet packet quad-tree of depth J satisfies*

$$2^{4^{J-1}} \le B_J \le 2^{\frac{49}{48} 4^{J-1}}.$$

Proof[3]. This result is proved with induction, as in the proof of Proposition 8.5. The reader can verify that B_J satisfies an induction relation similar to (8.16):

$$B_{J+1} = B_J^4 + 1. \tag{8.42}$$

Since $B_0 = 1$, $B_1 = 2$, and $B_{J+1} \ge B_J^4$, we derive that $B_J \ge 2^{4^{J-1}}$. Moreover, for $J \ge 1$

$$\log_2 B_{J+1} = 4 \log_2 B_J + \log_2 (1 + B_J^{-4}) \le 4 \log_2 B_J + \frac{1}{16} \le 4^J + \frac{1}{16} \sum_{j=0}^{J-1} 4^j,$$

which implies that $B_J \ge 2^{\frac{49}{48} 4^{J-1}}$. ■

For an image of N^2 pixels, we shall see that the wavelet packet quad-tree has a depth at most $\log_2 N$. The number of wavelet packet bases thus satisfies

$$2^{\frac{N^2}{4}} \leq B_{\log_2 N} \leq 2^{\frac{49}{48}\frac{N^2}{4}}. \tag{8.43}$$

Spatial and Frequency Localization The spatial and frequency localization of two-dimensional wavelet packets is derived from the time-frequency analysis performed in Section 8.1.2. If the conjugate mirror filter h has K non-zero coefficients, we proved that ψ_j^p has a support of size $2^j(K-1)$ hence $\psi_j^p(x_1)\,\psi_j^q(x_2)$ has a square support of width $2^j(K-1)$.

We showed that the Fourier transform of ψ_j^p has its energy mostly concentrated in

$$[-(k+1)2^{-j}\pi, -k2^{-j}\pi] \cup [k2^{-j}\pi, (k+1)2^{-j}\pi],$$

where $k = G[p]$ is specified by Proposition 8.2. The Fourier transform of a two-dimensional wavelet packet $\psi_j^{p,q}$ therefore has its energy mostly concentrated in

$$[k_1 2^{-j}\pi, (k_1+1)2^{-j}\pi] \times [k_2 2^{-j}\pi, (k_2+1)2^{-j}\pi], \tag{8.44}$$

with $k_1 = G[p]$ and $k_2 = G[q]$, and in the three squares that are symmetric with respect to the two axes $\omega_1 = 0$ and $\omega_2 = 0$. An admissible wavelet packet quad-tree decomposes the positive frequency quadrant into squares of dyadic sizes, as illustrated in Figure 8.12. For example, the leaves of a full wavelet packet quad-tree of depth $j-L$ define a wavelet packet basis that decomposes the positive frequency quadrant into squares of constant width equal to $2^{-j}\pi$. This wavelet packet basis is similar to a two-dimensional local cosine basis with windows of constant size.

8.2.2 Separable Filter Banks

The decomposition coefficients of an image in a separable wavelet packet basis are computed with a separable extension of the filter bank algorithm described in Section 8.1.4. Let $b[n]$ be an input image whose pixels have a distance $2^L = N^{-1}$. We associate to $b[n]$ a function $f \in \mathbf{V}_L^2$ approximated at the scale 2^L, whose decomposition coefficients $a_L[n] = \langle f(x), \phi_L^2(x - 2^L n)\rangle$ are defined like in (7.265):

$$b[n] = N\, a_L[n] \approx f(N^{-1}n)\ .$$

The wavelet packet coefficients

$$d_j^{p,q}[n] = \langle f, \psi_j^{p,q}(x - 2^j n)\rangle$$

characterize the orthogonal projection of f in $\mathbf{W}_j^{p,q}$. At the root, $d_L^{0,0} = a_L$.

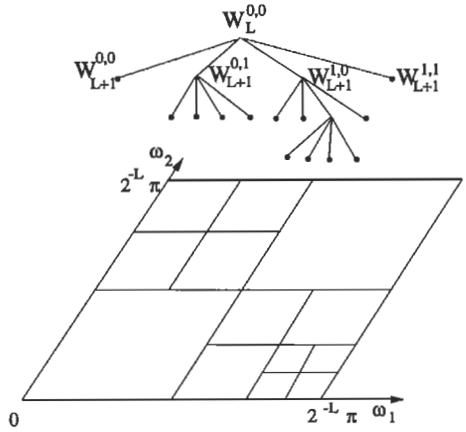


FIGURE 8.12 A wavelet packet quad-tree decomposes the positive frequency quadrant into squares of progressively smaller sizes as we go down the tree.

Separable Filter Bank From the separability of wavelet packet bases and the one-dimensional convolution formula of Proposition (8.4), we derive that for any $n = (n_1, n_2)$

$$d_{j+1}^{2p,2q}[n] = d_j^{p,q} \star \bar{h}\bar{h}[2n] \ , \ d_{j+1}^{2p+1,2q}[n] = d_j^{p,q} \star \bar{g}\bar{h}[2n], \tag{8.45}$$

$$d_{j+1}^{2p,2q+1}[n] = d_j^{p,q} \star \bar{h}\bar{g}[2n] \ , \ d_{j+1}^{2p+1,2q+1}[n] = d_j^{p,q} \star \bar{g}\bar{g}[2n]. \tag{8.46}$$

The coefficients of a wavelet packet quad-tree are thus computed by iterating these equations along the branches of the quad-tree. The calculations are performed with separable convolutions along the rows and columns of the image, illustrated in Figure 8.13.

At the reconstruction

$$\begin{aligned} d_j^{p,q}[n] = {} & \check{d}_{j+1}^{2p,2q} \star hh[n] + \check{d}_{j+1}^{2p+1,2q} \star gh[n] \\ & + \check{d}_{j+1}^{2p,2q+1} \star hg[n] + \check{d}_{j+1}^{2p+1,2q+1} \star gg[n]. \end{aligned} \tag{8.47}$$

The image $a_L = d_L^{0,0}$ is reconstructed from wavelet packet coefficients stored at the leaves of any admissible quad-tree by repeating the partial reconstruction (8.47) in the inside nodes of this quad-tree.

Finite Images If the image a_L has $N^2 = 2^{-2L}$ pixels, the one-dimensional convolution border problems are solved with one of the two approaches described in Sections 7.5.1 and 7.5.3. Each wavelet packet image $d_j^{p,q}$ includes 2^{-2j} pixels. At the depth $j - L$, there are N^2 wavelet packet coefficients in $\{d_j^{p,q}\}_{0 \le p,q < 2^{j-L}}$. A quad-tree of maximum depth $\log_2 N$ thus includes $N^2 \log_2 N$ coefficients. If h

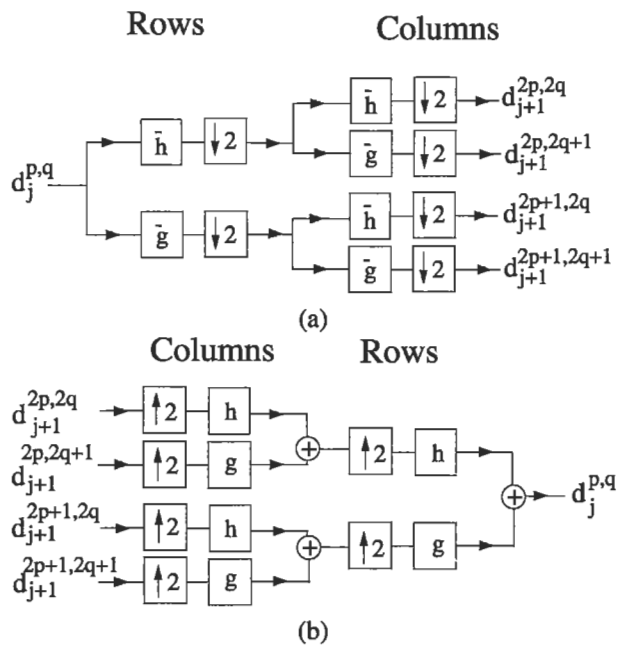


FIGURE 8.13 (a): Wavelet packet decomposition implementing (8.45) and (8.46) with one-dimensional convolutions along the rows and columns of $d_1^{p,q}$. (b): Wavelet packet reconstruction implementing (8.47).

and g have K non-zero coefficients, the one-dimensional convolutions that implement (8.45) and (8.46) require $2K2^{-2j}$ multiplications and additions. All wavelet packet coefficients at the depth $j+1-L$ are thus computed from wavelet packet coefficients located at the depth $j-L$ with $2KN^2$ calculations. The $N^2 \log_2 N$ wavelet packet coefficients of a full tree of depth $\log_2 N$ are therefore obtained with $2KN^2 \log_2 N$ multiplications and additions. The numerical complexity of reconstructing a_L from a wavelet packet basis depends on the number of inside nodes of the corresponding quad-tree. The worst case is a reconstruction from the leaves of a full quad-tree of depth $\log_2 N$, which requires $2KN^2 \log_2 N$ multiplications and additions.

8.3 BLOCK TRANSFORMS [1]

Wavelet packet bases are designed by dividing the frequency axis in intervals of varying sizes. These bases are thus particularly well adapted to decomposing signals that have different behavior in different frequency intervals. If f has properties that vary in time, it is then more appropriate to decompose f in a *block basis* that segments the time axis in intervals whose sizes are adapted to the signal structures. The next section explains how to generate a block basis of $\mathbf{L}^2(\mathbb{R})$ from

any basis of $\mathbf{L}^2[0,1]$. The cosine bases described in Sections 8.3.2 and 8.3.3 define particularly interesting block bases.

8.3.1 Block Bases

Block orthonormal bases are obtained by dividing the time axis in consecutive intervals $[a_p, a_{p+1}]$ with

$$\lim_{p\to-\infty} a_p = -\infty \quad \text{and} \quad \lim_{p\to+\infty} a_p = +\infty.$$

The size $l_p = a_{p+1} - a_p$ of each interval is arbitrary. Let $g = \mathbf{1}_{[0,1]}$. An interval is covered by the dilated rectangular window

$$g_p(t) = \mathbf{1}_{[a_p,a_{p+1}]}(t) = g\left(\frac{t-a_p}{l_p}\right). \tag{8.48}$$

The following theorem constructs a block orthogonal basis of $\mathbf{L}^2(\mathbb{R})$ from a single orthonormal basis of $\mathbf{L}^2[0,1]$.

Theorem 8.3 *If $\{e_k\}_{k\in\mathbb{Z}}$ is an orthonormal basis of $\mathbf{L}^2[0,1]$ then*

$$\left\{ g_{p,k}(t) = g_p(t)\frac{1}{\sqrt{l_p}} e_k\left(\frac{t-a_p}{l_p}\right) \right\}_{(p,k)\in\mathbb{Z}} \tag{8.49}$$

is a block orthonormal basis of $\mathbf{L}^2(\mathbb{R})$.

Proof[1]. One can verify that the dilated and translated family

$$\left\{ \frac{1}{\sqrt{l_p}} e_k\left(\frac{t-a_p}{l_p}\right) \right\}_{k\in\mathbb{Z}} \tag{8.50}$$

is an orthonormal basis of $\mathbf{L}^2[a_p, a_{p+1}]$. If $p \neq q$ then $\langle g_{p,k}, g_{q,k}\rangle = 0$ since their supports do not overlap. The family (8.49) is thus orthonormal. To expand a signal f in this family, it is decomposed as a sum of separate blocks

$$f(t) = \sum_{p=-\infty}^{+\infty} f(t)\, g_p(t),$$

and each block $f(t)g_p(t)$ is decomposed in the basis (8.50). ■

Block Fourier Basis A block basis is constructed with the Fourier basis of $\mathbf{L}^2[0,1]$:

$$\left\{ e_k(t) = \exp(i2k\pi t) \right\}_{k\in\mathbb{Z}}.$$

The time support of each block Fourier vector $g_{p,k}$ is $[a_p, a_{p+1}]$, of size l_p. The Fourier transform of $g = \mathbf{1}_{[0,1]}$ is

$$\hat{g}(\omega) = \frac{\sin(\omega/2)}{\omega/2} \exp\left(\frac{i\omega}{2}\right)$$

and

$$\hat{g}_{p,k}(\omega) = \sqrt{l_p}\,\hat{g}(l_p\omega - 2k\pi)\exp\left(\frac{-i2\pi k a_p}{l_p}\right).$$

It is centered at $2k\pi l_p^{-1}$ and has a slow asymptotic decay proportional to $l_p^{-1}|\omega|^{-1}$. Because of this bad frequency localization, even though a signal f is smooth, its decomposition in a block Fourier basis may include large high frequency coefficients. This can also be interpreted as an effect of periodization.

Discrete Block Bases For all $p \in \mathbb{Z}$, we suppose that $a_p \in \mathbb{Z}$. Discrete block bases are built with discrete rectangular windows whose supports are $[a_p, a_{p+1} - 1]$

$$g_p[n] = \mathbf{1}_{[a_p, a_{p+1}-1]}(n).$$

Since dilations are not defined in a discrete framework, we generally cannot derive bases of intervals of varying sizes from a single basis. The following theorem thus supposes that we can construct an orthonormal basis of $\mathbb{C}^l$ for any $l > 0$. The proof is straightforward.

Theorem 8.4 *Suppose that $\{e_{k,l}\}_{0\le k<l}$ is an orthogonal basis of $\mathbb{C}^l$, for any $l > 0$. The family*

$$\left\{g_{p,k}[n] = g_p[n]\,e_{k,l_p}[n - a_p]\right\}_{0\le k<l_p, p\in\mathbb{Z}} \tag{8.51}$$

is a block orthonormal basis of $\mathbf{l}^2(\mathbb{Z})$.

A discrete block basis is constructed with discrete Fourier bases

$$\left\{e_{k,l}[n] = \frac{1}{\sqrt{l}}\exp\left(\frac{i2\pi kn}{l}\right)\right\}_{0\le k<l}.$$

The resulting block Fourier vectors $g_{p,k}$ have sharp transitions at the window border, and are thus not well localized in frequency. As in the continuous case, the decomposition of smooth signals f may produce large amplitude high frequency coefficients because of border effects.

Block Bases of Images General block bases of images are constructed by partitioning the plane $\mathbb{R}^2$ into rectangles $\{[a_p, b_p] \times [c_p, d_p]\}_{p\in\mathbb{Z}}$ of arbitrary length $l_p = b_p - a_p$ and width $w_p = d_p - c_p$. Let $\{e_k\}_{k\in\mathbb{Z}}$ be an orthonormal basis of $\mathbf{L}^2[0,1]$ and $g = \mathbf{1}_{[0,1]}$. We denote

$$g_{p,k,j}(x,y) = g\left(\frac{x-a_p}{l_p}\right) g_q\left(\frac{y-c_p}{w_p}\right)\frac{1}{\sqrt{l_p w_p}}\,e_k\left(\frac{x-a_p}{l_p}\right) e_j\left(\frac{y-c_p}{w_p}\right).$$

The family $\{g_{p,k,j}\}_{(k,j)\in\mathbb{Z}^2}$ is an orthonormal basis of $\mathbf{L}^2([a_p, b_p] \times [c_p, d_p])$, and hence $\left\{g_{p,k,j}\right\}_{(p,k,j)\in\mathbb{Z}^3}$ is an orthonormal basis of $\mathbf{L}^2(\mathbb{R}^2)$.

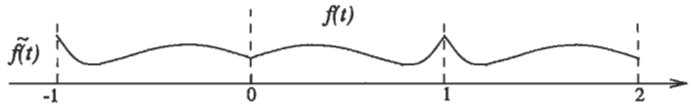


FIGURE 8.14 The function $\tilde{f}(t)$ is an extension of $f(t)$; it is symmetric about 0 and of period 2.

For discrete images, we define discrete windows that cover each rectangle

$$g_p = \mathbf{1}_{[a_p, b_p-1]\times[c_p, d_p-1]}.$$

If $\{e_{k,l}\}_{0\leq k<l}$ is an orthogonal basis of $\mathbb{C}^l$ for any $l > 0$, then

$$\left\{ g_{p,k,j}[n_1,n_2] = g_p[n_1,n_2]\, e_{k,l_p}[n_1-a_p]\, e_{j,w_p}[n_2-c_p] \right\}_{(k,j,p)\in\mathbb{Z}^3}.$$

is a block basis of $\mathbf{l}^2(\mathbb{Z}^2)$.

8.3.2 Cosine Bases

If $f \in \mathbf{L}^2[0,1]$ and $f(0) \neq f(1)$, even though f might be a smooth function, the Fourier coefficients

$$\langle f(u), \mathrm{e}^{i2k\pi u}\rangle = \int_0^1 f(u)\, \mathrm{e}^{-i2k\pi u}\, du$$

have a relatively large amplitude at high frequencies $2k\pi$. Indeed, the Fourier series expansion

$$f(t) = \sum_{k=-\infty}^{+\infty} \langle f(u), \mathrm{e}^{i2k\pi u}\rangle\, \mathrm{e}^{i2k\pi t}$$

is a function of period 1, equal to f over $[0,1]$, and which is therefore discontinuous if $f(0) \neq f(1)$. This shows that the restriction of a smooth function to an interval generates large Fourier coefficients. As a consequence, block Fourier bases are rarely used. A cosine I basis reduces this border effect by restoring a periodic extension $\tilde{f}$ of f which is continuous if f is continuous. High frequency cosine I coefficients thus have a smaller amplitude than Fourier coefficients.

Cosine I Basis We define $\tilde{f}$ to be the function of period 2 that is symmetric about 0 and equal to f over $[0,1]$:

$$\tilde{f}(t) = \begin{cases} f(t) & \text{for } t \in [0,1] \\ f(-t) & \text{for } t \in (-1,0) \end{cases} \tag{8.52}$$

If f is continuous over $[0,1]$ then $\tilde{f}$ is continuous over $\mathbb{R}$, as shown by Figure 8.14. However, if f has a non-zero right derivative at 0 or left derivative at 1, then $\tilde{f}$ is non-differentiable at integer points.

The Fourier expansion of $\tilde{f}$ over $[0,2]$ can be written as a sum of sine and cosine terms:

$$\tilde{f}(t) = \sum_{k=0}^{+\infty} a[k] \cos\left(\frac{2\pi kt}{2}\right) + \sum_{k=1}^{+\infty} b[k] \sin\left(\frac{2\pi kt}{2}\right).$$

The sine coefficients $b[k]$ are zero because $\tilde{f}$ is even. Since $f(t) = \tilde{f}(t)$ over $[0,1]$, this proves that any $f \in \mathbf{L}^2[0,1]$ can be written as a linear combination of the cosines $\{\cos(k\pi t)\}_{k\in\mathbb{N}}$. One can verify that this family is orthogonal over $[0,1]$. It is therefore an orthogonal basis of $\mathbf{L}^2[0,1]$, as stated by the following theorem.

Theorem 8.5 (COSINE I) *The family*

$$\left\{\lambda_k \sqrt{2}\cos(\pi kt)\right\}_{k\in\mathbb{N}} \text{ with } \lambda_k = \begin{cases} 2^{-1/2} & \text{if } k = 0 \\ 1 & \text{if } k \neq 0 \end{cases}$$

is an orthonormal basis of $\mathbf{L}^2[0,1]$.

Block Cosine Basis Let us divide the real line with square windows $g_p = \mathbf{1}_{[a_p, a_{p+1}]}$. Theorem 8.3 proves that

$$\left\{ g_{p,k}(t) = g_p(t) \sqrt{\frac{2}{l_p}} \lambda_k \cos\left(\pi k \frac{t - a_p}{l_p}\right) \right\}_{k\in\mathbb{N}, p\in\mathbb{Z}}$$

is a block basis of $\mathbf{L}^2(\mathbb{R})$. The decomposition coefficients of a smooth function have a faster decay at high frequencies in a block cosine basis than in a block Fourier basis, because cosine bases correspond to a smoother signal extension beyond the intervals $[a_p, a_{p+1}]$.

Cosine IV Basis Other cosine bases are constructed from Fourier series, with different extensions of f beyond $[0,1]$. The cosine IV basis appears in fast numerical computations of cosine I coefficients. It is also used to construct local cosine bases with smooth windows in Section 8.4.2.

Any $f \in \mathbf{L}^2[0,1]$ is extended into a function $\tilde{f}$ of period 4, which is symmetric about 0 and antisymmetric about 1 and -1:

$$\tilde{f}(t) = \begin{cases} f(t) & \text{if } t \in [0,1] \\ f(-t) & \text{if } t \in (-1,0) \\ -f(2-t) & \text{if } t \in [1,2) \\ -f(2+t) & \text{if } t \in [-1,-2) \end{cases}$$

If $f(1) \neq 0$, the antisymmetry at 1 creates a function $\tilde{f}$ that is discontinuous at $f(2n+1)$ for any $n \in \mathbb{Z}$, as shown by Figure 8.15. This extension is therefore less regular than the cosine I extension (8.52).

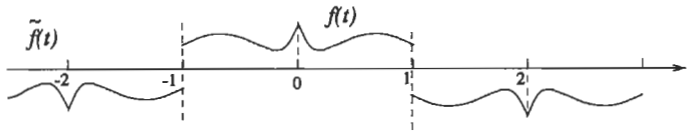


FIGURE 8.15 A cosine IV extends $f(t)$ into a signal $\tilde{f}(t)$ of period 4 which is symmetric with respect to 0 and antisymmetric with respect to 1.

Since $\tilde{f}$ is 4 periodic, it can be decomposed as a sum of sines and cosines of period 4:

$$\tilde{f}(t) = \sum_{k=0}^{+\infty} a[k] \cos\left(\frac{2\pi kt}{4}\right) + \sum_{k=1}^{+\infty} b[k] \sin\left(\frac{2\pi kt}{4}\right).$$

The symmetry about 0 implies that

$$b[k] = \frac{1}{2}\int_{-2}^{2} \tilde{f}(t) \sin\left(\frac{2\pi kt}{4}\right) dt = 0.$$

For even frequencies, the antisymmetry about 1 and -1 yields

$$a[2k] = \frac{1}{2}\int_{-2}^{2} \tilde{f}(t) \cos\left(\frac{2\pi(2k)t}{4}\right) dt = 0.$$

The only non-zero components are thus cosines of odd frequencies:

$$\tilde{f}(t) = \sum_{k=0}^{+\infty} a[2k+1] \cos\left(\frac{(2k+1)2\pi t}{4}\right). \tag{8.53}$$

Since $f(t) = \tilde{f}(t)$ over $[0,1]$, this proves that any $f \in \mathbf{L}^2[0,1]$ is decomposed as a sum of such cosine functions. One can verify that the restriction of these cosine functions to $[0,1]$ is orthogonal in $\mathbf{L}^2[0,1]$, which implies the following theorem.

Theorem 8.6 (COSINE IV) *The family*

$$\left\{\sqrt{2}\cos\left[\left(k+\frac{1}{2}\right)\pi t\right]\right\}_{k\in\mathbf{N}}$$

is an orthonormal basis of $\mathbf{L}^2[0,1]$.

The cosine transform IV is not used in block transforms because it has the same drawbacks as a block Fourier basis. Block Cosine IV coefficients of a smooth f have a slow decay at high frequencies because such a decomposition corresponds to a discontinuous extension of f beyond each block. Section 8.4.2 explains how to avoid this issue with smooth windows.

8.3.3 Discrete Cosine Bases

Discrete cosine bases are derived from the discrete Fourier basis with the same approach as in the continuous time case. To simplify notations, the sampling distance is normalized to 1. If the sampling distance was originally N^{-1} then the frequency indexes that appear in this section must be multiplied by N.

Discrete Cosine I A signal $f[n]$ defined for $0 \leq n < N$ is extended by symmetry with respect to $-1/2$ into a signal $\tilde{f}[n]$ of size $2N$:

$$\tilde{f}[n] = \begin{cases} f[n] & \text{for } 0 \leq n < N \\ f[-n-1] & \text{for } -N \leq n \leq -1 \end{cases}. \tag{8.54}$$

The $2N$ discrete Fourier transform of $\tilde{f}$ can be written as a sum of sine and cosine terms:

$$\tilde{f}[n] = \sum_{k=0}^{N-1} a[k] \cos\left[\frac{k\pi}{N}\left(n+\frac{1}{2}\right)\right] + \sum_{k=0}^{N-1} b[k] \sin\left[\frac{k\pi}{N}\left(n+\frac{1}{2}\right)\right].$$

Since $\tilde{f}$ is symmetric about $-1/2$, necessarily $b[k] = 0$ for $0 \leq k < N$. Moreover $f[n] = \tilde{f}[n]$ for $0 \leq n < N$, so any signal $f \in \mathbb{C}^N$ can be written as a sum of these cosine functions. The reader can also verify that these discrete cosine signals are orthogonal in $\mathbb{C}^N$. We thus obtain the following theorem.

Theorem 8.7 (COSINE I) *The family*

$$\left\{ \lambda_k \sqrt{\frac{2}{N}} \cos\left[\frac{k\pi}{N}\left(n+\frac{1}{2}\right)\right] \right\}_{0 \leq k < N} \quad \textit{with} \quad \lambda_k = \begin{cases} 2^{-1/2} & \textit{if } k = 0 \\ 1 & \textit{otherwise} \end{cases}$$

is an orthonormal basis of $\mathbb{C}^N$.

This theorem proves that any $f \in \mathbb{C}^N$ can be decomposed into

$$f[n] = \frac{2}{N} \sum_{k=0}^{N-1} \hat{f}_I[k]\, \lambda_k \cos\left[\frac{k\pi}{N}\left(n+\frac{1}{2}\right)\right], \tag{8.55}$$

where

$$\hat{f}_I[k] = \left\langle f[n], \lambda_k \cos\left[\frac{k\pi}{N}\left(n+\frac{1}{2}\right)\right] \right\rangle = \lambda_k \sum_{n=0}^{N-1} f[n] \cos\left[\frac{k\pi}{N}\left(n+\frac{1}{2}\right)\right]. \tag{8.56}$$

is the discrete cosine transform I (DCT-I) of f. The next section describes a fast discrete cosine transform which computes $\hat{f}_I$ with $O(N \log_2 N)$ operations.

Discrete Block Cosine Transform Let us divide the integer set $\mathbb{Z}$ with discrete windows $g_p[n] = 1_{[a_p, a_{p-1}]}(n)$, with $a_p \in \mathbb{Z}$. Theorem 8.4 proves that the corresponding block basis

$$\left\{ g_{p,k}[n] = g_p[n]\, \lambda_k \sqrt{\frac{2}{l_p}} \cos\left[\frac{k\pi}{l_p}\left(n + \frac{1}{2} - a_p\right)\right] \right\}_{0 \le k < N, p \in \mathbb{Z}}$$

is an orthonormal basis of $l^2(\mathbb{Z})$. Over each block of size $l_p = a_{p+1} - a_p$, the fast DCT-I algorithm computes all coefficients with $O(l_p \log_2 l_p)$ operations. Section 11.4.3 describes the JPEG image compression standard, which decomposes images in a separable block cosine basis. A block cosine basis is used as opposed to a block Fourier basis because it yields smaller amplitude high frequency coefficients, which improves the coding performance.

Discrete Cosine IV To construct a discrete cosine IV basis, a signal f of N samples is extended into a signal $\tilde{f}$ of period $4N$, which is symmetric with respect to $-1/2$ and antisymmetric with respect to $N - 1/2$ and $-N + 1/2$. As in (8.53), the decomposition of $\tilde{f}$ over a family of sines and cosines of period $4N$ has no sine terms and no cosine terms of even frequency. Since $\tilde{f}[n] = f[n]$, for $0 \le n < N$, we derive that f can also be written as a linear expansion of these odd frequency cosines, which are orthogonal in $\mathbb{C}^N$. We thus obtain the following theorem.

Theorem 8.8 (COSINE IV) *The family*

$$\left\{ \sqrt{\frac{2}{N}} \cos\left[\frac{\pi}{N}\left(k + \frac{1}{2}\right)\left(n + \frac{1}{2}\right)\right] \right\}_{0 \le k < N}$$

is an orthonormal basis of $\mathbb{C}^N$.

This theorem proves that any $f \in \mathbb{C}^N$ can be decomposed into

$$f[n] = \frac{2}{N} \sum_{k=0}^{N-1} \hat{f}_{IV}[k] \cos\left[\frac{\pi}{N}\left(k + \frac{1}{2}\right)\left(n + \frac{1}{2}\right)\right], \tag{8.57}$$

where

$$\hat{f}_{IV}[k] = \sum_{n=0}^{N-1} f[n] \cos\left[\frac{\pi}{N}\left(k + \frac{1}{2}\right)\left(n + \frac{1}{2}\right)\right] \tag{8.58}$$

is the discrete cosine transform IV (DCT-IV) of f.

8.3.4 Fast Discrete Cosine Transforms [2]

The discrete cosine transform IV (DCT-IV) of a signal of size N is related to the discrete Fourier transform (DFT) of a complex signal of size $N/2$ with a formula introduced by Duhamel, Mahieux, and Petit [176, 42]. By computing this DFT with the fast Fourier transform (FFT) described in Section 3.3.3, we need $O(N \log_2 N)$ operations to compute the DCT-IV. The DCT-I coefficients are then calculated through an induction relation with the DCT-IV, due to Wang [346].

Fast DCT-IV To clarify the relation between a DCT-IV and a DFT, we split $f[n]$ in two half-size signals of odd and even indices:

$$\begin{aligned} b[n] &= f[2n], \\ c[n] &= f[N-1-2n]. \end{aligned}$$

The DCT-IV (8.58) is rewritten

$$\begin{aligned} \hat{f}_{IV}[k] &= \sum_{n=0}^{N/2-1} b[n] \cos\left[\left(2n+\frac{1}{2}\right)\left(k+\frac{1}{2}\right)\frac{\pi}{N}\right] + \\ &\quad \sum_{n=0}^{N/2-1} c[n] \cos\left[\left(N-1-2n+\frac{1}{2}\right)\left(k+\frac{1}{2}\right)\frac{\pi}{N}\right] \\ &= \sum_{n=0}^{N/2-1} b[n] \cos\left[\left(n+\frac{1}{4}\right)\left(k+\frac{1}{2}\right)\frac{2\pi}{N}\right] + \\ &\quad (-1)^k \sum_{n=0}^{N/2-1} c[n] \sin\left[\left(n+\frac{1}{4}\right)\left(k+\frac{1}{2}\right)\frac{2\pi}{N}\right]. \end{aligned}$$

The even frequency indices can thus be expressed as a real part

$$\begin{aligned} &\hat{f}_{IV}[2k] = \\ &\text{Real}\left\{\exp\left[\tfrac{-i\pi k}{N}\right] \textstyle\sum_{n=0}^{N/2-1} (b[n]+ic[n]) \exp\left[-i\left(n+\tfrac{1}{4}\right)\tfrac{\pi}{N}\right] \exp\left[\tfrac{-i2\pi kn}{N/2}\right]\right\}, \end{aligned} \tag{8.59}$$

whereas the odd coefficients correspond to an imaginary part

$$\begin{aligned} &\hat{f}_{IV}[N-2k-1] = \\ &-\text{Im}\left\{\exp\left[\tfrac{-i\pi k}{N}\right] \textstyle\sum_{n=0}^{N/2-1} (b[n]+ic[n]) \exp\left[-i\left(n+\tfrac{1}{4}\right)\tfrac{\pi}{N}\right] \exp\left[\tfrac{-i2\pi kn}{N/2}\right]\right\}. \end{aligned} \tag{8.60}$$

For $0 \leq n < N/2$, we denote

$$g[n] = (b[n]+ic[n]) \exp\left[-i\left(n+\frac{1}{4}\right)\frac{\pi}{N}\right].$$

The DFT $\hat{g}[k]$ of $g[n]$ is computed with an FFT of size $N/2$. Equations (8.59) and (8.60) prove that

$$\hat{f}_{IV}[2k] = \text{Real}\left\{\exp\left[\frac{-i\pi k}{N}\right] \hat{g}[k]\right\},$$

and

$$\hat{f}_{IV}[N-2k-1] = -\text{Im}\left\{\exp\left[\frac{-i\pi k}{N}\right] \hat{g}[k]\right\}.$$

The DCT-IV coefficients $\hat{f}_{IV}[k]$ are thus obtained with one FFT of size $N/2$ plus $O(N)$ operations, which makes a total of $O(N \log_2 N)$ operations. To normalize the DCT-IV, the resulting coefficients must be multiplied by $\sqrt{\frac{2}{N}}$. An efficient implementation of the DCT-IV with a split-radix FFT requires [42]

$$\mu_{DCT-IV}(N) = \frac{N}{2} \log_2 N + N, \tag{8.61}$$

real multiplications and

$$\alpha_{DCT-IV}(N) = \frac{3N}{2} \log_2 N \tag{8.62}$$

additions.

The inverse DCT-IV of $\hat{f}_{IV}$ is given by (8.57). Up to the proportionality constant $2/N$, this sum is the same as (8.58), where $\hat{f}_{IV}$ and f are interchanged. This proves that the inverse DCT-IV is computed with the same fast algorithm as the forward DCT-IV.

Fast DCT-I A DCT-I is calculated with an induction relation that involves the DCT-IV. Regrouping the terms $f[n]$ and $f[N-1-n]$ of a DCT-I (8.56) yields

$$\hat{f}_I[2k] = \lambda_k \sum_{n=0}^{N/2-1} (f[n] + f[N-1-n]) \cos\left[\frac{\pi k}{N/2}\left(n+\frac{1}{2}\right)\right], \tag{8.63}$$

$$\hat{f}_I[2k+1] = \sum_{n=0}^{N/2-1} (f[n] - f[N-1-n]) \cos\left[\frac{\pi (k+1/2)}{N/2}\left(n+\frac{1}{2}\right)\right]. \tag{8.64}$$

The even index coefficients of the DCT-I are thus equal to the DCT-I of the signal $f[n] + f[N-1-n]$ of length $N/2$. The odd coefficients are equal to the DCT-IV of the signal $f[n] - f[N-1-n]$ of length $N/2$. The number of multiplications of a DCT-I is thus related to the number of multiplications of a DCT-IV by the induction relation

$$\mu_{DCT-I}(N) = \mu_{DCT-I}(N/2) + \mu_{DCT-IV}(N/2), \tag{8.65}$$

while the number of additions is

$$\alpha_{DCT-I}(N) = \alpha_{DCT-I}(N/2) + \alpha_{DCT-IV}(N/2) + N. \tag{8.66}$$

Since the number of multiplications and additions of a DCT-IV is $O(N \log_2 N)$ this induction relation proves that the number of multiplications and additions of this algorithm is also $O(N \log_2 N)$.

If the DCT-IV is implemented with a split-radix FFT, inserting (8.61) and (8.62) in the recurrence equations (8.65) and (8.66), we derive that the number of multiplications and additions to compute a DCT-I of size N is

$$\mu_{DCT-I}(N) = \frac{N}{2} \log_2 N + 1, \tag{8.67}$$

and

$$\alpha_{DCT\text{-}I}(N) = \frac{3N}{2}\log_2 N - N + 1. \tag{8.68}$$

The inverse DCT-I is computed with a similar recursive algorithm. Applied to $\hat{f}_I$, it is obtained by computing the inverse DCT-IV of the odd index coefficients $\hat{f}_I[2k+1]$ with (8.64) and an inverse DCT-I of a size $N/2$ applied to the even coefficients $\hat{f}_I[2k]$ with (8.63). From the values $f[n] + f[N-1-n]$ and $f[n] - f[N-1-n]$, we recover $f[n]$ and $f[N-1-n]$. The inverse DCT-IV is identical to the forward DCT-IV up to a multiplicative constant. The inverse DCT-I thus requires the same number of operations as the forward DCT-I.

8.4 LAPPED ORTHOGONAL TRANSFORMS [2]

Cosine and Fourier block bases are computed with discontinuous rectangular windows that divide the real line in disjoint intervals. Multiplying a signal with a rectangular window creates discontinuities that produce large amplitude coefficients at high frequencies. To avoid these discontinuity artifacts, it is necessary to use smooth windows.

The Balian-Low Theorem 5.6 proves that for any u_0 and ξ_0, there exists no differentiable window g of compact support such that

$$\left\{ g(t - nu_0) \exp(ik\xi_0 t) \right\}_{(n,k)\in\mathbb{Z}^2}$$

is an orthonormal basis of $\mathbf{L}^2(\mathbb{R})$. This negative result discouraged any research in this direction, until Malvar discovered in discrete signal processing that one could create orthogonal bases with smooth windows modulated by a cosine IV basis [262, 263]. This result was independently rediscovered for continuous time functions by Coifman and Meyer [138], with a different approach that we shall follow here. The roots of these new orthogonal bases are lapped projectors, which split signals in orthogonal components with overlapping supports [46]. Section 8.4.1 introduces these lapped projectors; the construction of continuous time and discrete lapped orthogonal bases is explained in the following sections. The particular case of local cosine bases is studied in more detail.

8.4.1 Lapped Projectors

Block transforms compute the restriction of f to consecutive intervals $[a_p, a_{p+1}]$ and decompose this restriction in an orthogonal basis of $[a_p, a_{p+1}]$. Formally, the restriction of f to $[a_p, a_{p+1}]$ is an orthogonal projection on the space $\mathbf{W}^p$ of functions with a support included in $[a_p, a_{p+1}]$. To avoid the discontinuities introduced by this projection, we introduce new orthogonal projectors that perform a smooth deformation of f.

Projectors on Half Lines Let us first construct two orthogonal projectors that decompose any $f \in \mathbf{L}^2(\mathbb{R})$ in two orthogonal components $P^+ f$ and $P^- f$ whose

supports are respectively $[-1,+\infty)$ and $(-\infty,1]$. For this purpose we consider a monotone increasing profile function β such that

$$\beta(t)=\begin{cases}0 & \text{if } t<-1\\ 1 & \text{if } t>1\end{cases} \tag{8.69}$$

and

$$\forall t\in[-1,1] \quad , \quad \beta^2(t)+\beta^2(-t)=1. \tag{8.70}$$

A naive definition

$$P^+f(t)=\beta^2(t)\,f(t) \quad \text{and} \quad P^-f(t)=\beta^2(-t)\,f(t)$$

satisfies the support conditions but does not define orthogonal functions. Since the supports of $P^+f(t)$ and $P^-f(t)$ overlap only on $[-1,1]$, the orthogonality is obtained by creating functions having a different symmetry with respect to 0 on $[-1,1]$:

$$P^+f(t)=\beta(t)\,[\beta(t)\,f(t)+\beta(-t)\,f(-t)]=\beta(t)\,p(t)\,, \tag{8.71}$$

and

$$P^-f(t)=\beta(-t)\,[\beta(-t)\,f(t)-\beta(t)\,f(-t)]=\beta(-t)\,q(t)\,. \tag{8.72}$$

The functions $p(t)$ and $q(t)$ are respectively even and odd, and since $\beta(t)\beta(-t)$ is even it follows that

$$\langle P^+f,P^-f\rangle=\int_{-1}^{1}\beta(t)\,\beta(-t)\,p(t)\,q^*(t)\,dt=0. \tag{8.73}$$

Clearly P^+f belongs to the space $\mathbf{W}^+$ of functions $f\in\mathbf{L}^2(\mathbb{R})$ such that there exists $p(t)=p(-t)$ with

$$f(t)=\begin{cases}0 & \text{if } t<-1\\ \beta(t)\,p(t) & \text{if } t\in[-1,1]\end{cases}.$$

Similarly P^-f is in the space $\mathbf{W}^-$ composed of $f\in\mathbf{L}^2(\mathbb{R})$ such that there exists $q(t)=-q(-t)$ with

$$f(t)=\begin{cases}0 & \text{if } t>1\\ \beta(-t)\,q(t) & \text{if } t\in[-1,1]\end{cases}.$$

Functions in $\mathbf{W}^+$ and $\mathbf{W}^-$ may have an arbitrary behavior on $[1,+\infty)$ and $(-\infty,-1]$ respectively . The following theorem characterizes P^+ and P^-. We denote by *Id* the identity operator.

Theorem 8.9 (COIFMAN, MEYER) *The operators P^+ and P^- are orthogonal projectors respectively on $\mathbf{W}^+$ and $\mathbf{W}^-$. The spaces $\mathbf{W}^+$ and $\mathbf{W}^-$ are orthogonal and*

$$P^++P^-=Id. \tag{8.74}$$

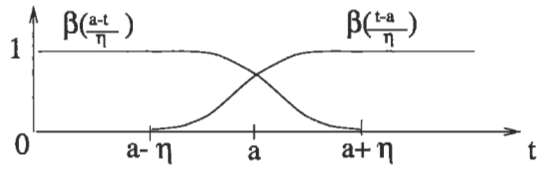


FIGURE 8.16 A multiplication with $\beta(\frac{t-a}{\eta})$ and $\beta(\frac{a-t}{\eta})$ restricts the support of functions to $[a-\eta,+\infty)$ and $(-\infty,a+\eta]$.

Proof[2]. To verify that P^+ is a projector we show that any $f \in \mathbf{W}^+$ satisfies $P^+f = f$. If $t < -1$ then $P^+f(t) = f(t) = 0$ and if $t > 1$ then $P^+f(t) = f(t) = 1$. If $t \in [-1,1]$ then $f(t) = \beta(t)\, p_0(t)$ and inserting (8.71) yields

$$P^+f(t) = \beta(t)\,[\beta^2(t)\, p_0(t) + \beta^2(-t)\, p_0(-t)] = \beta(t)\, p_0(t),$$

because $p_0(t)$ is even and $\beta(t)$ satisfies (8.70). The projector P^+ is proved to be orthogonal by showing that it is self-adjoint:

$$\langle P^+f, g\rangle = \int_{-1}^{1} \beta^2(t)\, f(t)\, g^*(t)\, dt + \int_{-1}^{1} \beta(t)\,\beta(-t)\, f(-t)\, g^*(t)\, dt + \int_{1}^{+\infty} f(t)\, g^*(t)\, dt.$$

A change of variable $t' = -t$ in the second integral verifies that this formula is symmetric in f and g and hence $\langle P^+f, g\rangle = \langle f, P^+g\rangle$. Identical derivations prove that P^- is an orthogonal projector on $\mathbf{W}^-$.

The orthogonality of $\mathbf{W}^-$ and $\mathbf{W}^+$ is proved in (8.73). To verify (8.74), for $f \in \mathbf{L}^2(\mathbb{R})$ we compute

$$P^+f(t) + P^-f(t) = f(t)\,[\beta^2(t) + \beta^2(-t)] = f(t).$$

■

These half-line projectors are generalized by decomposing signals in two orthogonal components whose supports are respectively $[a-\eta,+\infty)$ and $(-\infty,a+\eta]$. For this purpose, we scale and translate the profile function $\beta(\frac{t-a}{\eta})$, so that it increases from 0 to 1 on $[a-\eta,a+\eta]$, as illustrated in Figure 8.16. The symmetry with respect to 0, which transforms $f(t)$ in $f(-t)$, becomes a symmetry with respect to a, which transforms $f(t)$ in $f(2a-t)$. The resulting projectors are

$$P^+_{a,\eta}f(t) = \beta\Big(\frac{t-a}{\eta}\Big)\left[\beta\Big(\frac{t-a}{\eta}\Big)\, f(t) + \beta\Big(\frac{a-t}{\eta}\Big)\, f(2a-t)\right] \tag{8.75}$$

and

$$P^-_{a,\eta}f(t) = \beta\Big(\frac{a-t}{\eta}\Big)\left[\beta\Big(\frac{a-t}{\eta}\Big)\, f(t) - \beta\Big(\frac{t-a}{\eta}\Big)\, f(2a-t)\right]. \tag{8.76}$$

A straightforward extension of Theorem 8.9 proves that $P^+_{a,\eta}$ is an orthogonal projector on the space $\mathbf{W}^+_{a,\eta}$ of functions $f \in \mathbf{L}^2(\mathbb{R})$ such that there exists $p(t) = p(2a-t)$ with

$$f(t) = \begin{cases} 0 & \text{if } t < a-\eta \\ \beta(\eta^{-1}(t-a))\, p(t) & \text{if } t \in [a-\eta, a+\eta] \end{cases}. \tag{8.77}$$

Similarly $P^-_{a,\eta}$ is an orthogonal projector on the space $\mathbf{W}^-_{a,\eta}$ composed of $f \in \mathbf{L}^2(\mathbb{R})$ such that there exists $q(t) = -q(2a-t)$ with

$$f(t) = \begin{cases} 0 & \text{if } t < -1 \\ \beta(\eta^{-1}(a-t))\, q(t) & \text{if } t \in [a-\eta, a+\eta] \end{cases}. \tag{8.78}$$

The spaces $\mathbf{W}^+_{a,\eta}$ and $\mathbf{W}^-_{a,\eta}$ are orthogonal and

$$P^+_{a,\eta} + P^-_{a,\eta} = Id. \tag{8.79}$$

Projectors on Intervals A lapped projector splits a signal in two orthogonal components that overlap on $[a-\eta, a+\eta]$. Repeating such projections at different locations performs a signal decomposition into orthogonal pieces whose supports overlap. Let us divide the time axis in overlapping intervals:

$$I_p = [a_p - \eta_p, a_{p+1} + \eta_{p+1}]$$

with

$$\lim_{p \to -\infty} a_p = -\infty \quad \text{and} \quad \lim_{p \to +\infty} a_p = +\infty. \tag{8.80}$$

To ensure that I_{p-1} and I_{p+1} do not intersect for any $p \in \mathbb{Z}$, we impose that

$$a_{p+1} - \eta_{p+1} \geq a_p + \eta_p,$$

and hence

$$l_p = a_{p+1} - a_p \geq \eta_{p+1} + \eta_p. \tag{8.81}$$

The support of f is restricted to I_p by the operator

$$P_p = P^+_{a_p,\eta_p} P^-_{a_{p+1},\eta_{p+1}}. \tag{8.82}$$

Since $P^+_{a_p,\eta_p}$ and $P^-_{a_{p+1},\eta_{p+1}}$ are orthogonal projections on $\mathbf{W}^+_{a_p,\eta_p}$ and $\mathbf{W}^-_{a_{p+1},\eta_{p+1}}$, it follows that P_p is an orthogonal projector on

$$\mathbf{W}^p = \mathbf{W}^+_{a_p,\eta_p} \cap \mathbf{W}^-_{a_{p+1},\eta_{p+1}}. \tag{8.83}$$

Let us divide I_p in two overlapping intervals O_p, O_{p+1} and a central interval C_p:

$$I_p = [a_p - \eta_p, a_{p+1} + \eta_{p+1}] = O_p \cup C_p \cup O_{p+1} \tag{8.84}$$

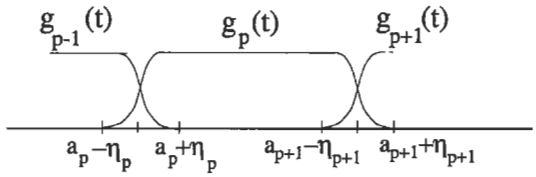


FIGURE 8.17 Each window g_p has a support $[a_p - \eta_p, a_{p+1} + \eta_{p+1}]$ with an increasing profile and a decreasing profile over $[a_p - \eta_p, a_p + \eta_p]$ and $[a_{p+1} - \eta_{p+1}, a_{p+1} + \eta_{p+1}]$.

with

$$O_p = [a_p - \eta_p, a_p + \eta_p] \text{ and } C_p = [a_p + \eta_p, a_{p+1} - \eta_{p+1}].$$

The space $\mathbf{W}^p$ is characterized by introducing a window g_p whose support is I_p, and which has a raising profile on O_p and a decaying profile on O_{p+1}:

$$g_p(t) = \begin{cases} 0 & \text{if } t \notin I_p \\ \beta(\eta_p^{-1}(t - a_p)) & \text{if } t \in O_p \\ 1 & \text{if } t \in C_p \\ \beta(\eta_{p+1}^{-1}(a_{p+1} - t)) & \text{if } t \in O_{p+1} \end{cases} \tag{8.85}$$

This window is illustrated in Figure 8.17. It follows from (8.77), (8.78) and (8.83) that $\mathbf{W}^p$ is the space of functions $f \in \mathbf{L}^2(\mathbb{R})$ that can be written

$$f(t) = g_p(t)\, h(t) \quad \text{with} \quad h(t) = \begin{cases} h(2a_p - t) & \text{if } t \in O_p \\ -h(2a_{p+1} - t) & \text{if } t \in O_{p+1} \end{cases} \tag{8.86}$$

The function h is symmetric with respect to a_p and antisymmetric with respect to a_{p+1}, with an arbitrary behavior in C_p. The projector P_p on $\mathbf{W}^p$ defined in (8.82) can be rewritten

$$P_p f(t) = \begin{cases} P^-_{a_p, \eta_p} f(t) & \text{if } t \in O_p \\ f(t) & \text{if } t \in C_p \\ P^+_{a_{p+1}, \eta_{p+1}} f(t) & \text{if } t \in O_{p+1} \end{cases} = g_p(t)\, h_p(t), \tag{8.87}$$

where $h_p(t)$ is calculated by inserting (8.75) and (8.76):

$$h_p(t) = \begin{cases} g_p(t)\, f(t) + g_p(2a_p - t)\, f(2a_p - t) & \text{if } t \in O_p \\ f(t) & \text{if } t \in C_p \\ g_p(t)\, f(t) - g_p(2a_{p+1} - t)\, f(2a_{p+1} - t) & \text{if } t \in O_{p+1} \end{cases}. \tag{8.88}$$

The following proposition derives a decomposition of the identity.

Proposition 8.6 *The operator P_p is an orthogonal projector on $\mathbf{W}^p$. If $p \neq q$ then $\mathbf{W}^p$ is orthogonal to $\mathbf{W}^q$ and*

$$\sum_{p=-\infty}^{+\infty} P_p = Id. \tag{8.89}$$

Proof[2]. If $p \neq q$ and $|p-q| > 1$ then functions in $\mathbf{W}^p$ and $\mathbf{W}^q$ have supports that do not overlap so these spaces are orthogonal. If $q = p+1$ then

$$\mathbf{W}^p = \mathbf{W}^+_{a_p,\eta_p} \cap \mathbf{W}^-_{a_{p+1},\eta_{p+1}} \quad \text{and} \quad \mathbf{W}^{p+1} = \mathbf{W}^+_{a_{p+1},\eta_{p+1}} \cap \mathbf{W}^-_{a_{p+2},\eta_{p+2}}.$$

Since $\mathbf{W}^-_{a_{p+1},\eta_{p+1}}$ is orthogonal to $\mathbf{W}^+_{a_{p+1},\eta_{p+1}}$ it follows that $\mathbf{W}^p$ is orthogonal to $\mathbf{W}^{p+1}$. To prove (8.89), we first verify that

$$P_p + P_{p+1} = P^+_{a_p,\eta_p} P^-_{a_{p+2},\eta_{p+2}}. \tag{8.90}$$

This is shown by decomposing P_p and P_{p+1} with (8.87) and inserting

$$P^+_{a_{p+1},\eta_{p+1}} + P^-_{a_{p+1},\eta_{p+1}} = Id.$$

As a consequence

$$\sum_{p=n}^{m} P_p = P^+_{a_n,\eta_n} P^-_{a_m,\eta_m}. \tag{8.91}$$

For any $f \in \mathbf{L}^2(\mathbb{R})$,

$$\|f - P^+_{a_n,\eta_n} P^-_{a_m,\eta_m} f\|^2 \geq \int_{-\infty}^{a_n+\eta_n} |f(t)|^2 dt + \int_{a_m-\eta_m}^{+\infty} |f(t)|^2 dt$$

and inserting (8.80) proves that

$$\lim_{\substack{n\to-\infty \\ m\to+\infty}} \|f - P^+_{a_n,\eta_n} P^-_{a_m,\eta_m} f\|^2 = 0.$$

The summation (8.91) implies (8.89). ■

Discretized Projectors Projectors P_p that restrict the signal support to $[a_p - \eta_p, a_{p+1} + \eta_{p+1}]$ are easily extended for discrete signals. Suppose that $\{a_p\}_{p\in\mathbb{Z}}$ are half integers, which means that $a_p + 1/2 \in \mathbb{Z}$. The windows $g_p(t)$ defined in (8.85) are uniformly sampled $g_p[n] = g_p(n)$. As in (8.86) we define the space $\mathbf{W}^p \subset \mathbf{l}^2(\mathbb{Z})$ of discrete signals

$$f[n] = g_p[n]\, h[n] \quad \text{with} \quad h[n] = \begin{cases} h[2a_p - n] & \text{if } n \in O_p \\ -h[2a_{p+1} - n] & \text{if } n \in O_{p+1} \end{cases}. \tag{8.92}$$

The orthogonal projector P_p on $\mathbf{W}^p$ is defined by an expression identical to (8.87, 8.88):

$$P_p f[n] = g_p[n]\, h_p[n] \tag{8.93}$$

with

$$h_p[n] = \begin{cases} g_p[n]\, f[n] + g_p[2a_p - n]\, f[2a_p - n] & \text{if } n \in O_p \\ f[n] & \text{if } n \in C_p \\ g_p[n]\, f[n] - g_p[2a_{p+1} - n]\, f[2a_{p+1} - n] & \text{if } n \in O_{p+1} \end{cases}. \tag{8.94}$$

Finally we prove as in Proposition 8.6 that if $p \neq q$, then $\mathbf{W}^p$ is orthogonal to $\mathbf{W}^q$ and

$$\sum_{p=-\infty}^{+\infty} P_p = Id. \tag{8.95}$$

8.4.2 Lapped Orthogonal Bases

An orthogonal basis of $\mathbf{L}^2(\mathbb{R})$ is defined from a basis $\{e_k\}_{k\in\mathbb{N}}$ of $\mathbf{L}^2[0,1]$ by multiplying a translation and dilation of each vector with a smooth window g_p defined in (8.85). A local cosine basis of $\mathbf{L}^2(\mathbb{R})$ is derived from a cosine-IV basis of $\mathbf{L}^2[0,1]$.

The support of g_p is $[a_p-\eta_p, a_{p+1}+\eta_{p+1}]$, with $l_p = a_{p+1}-a_p$, as illustrated in Figure 8.17. The design of these windows also implies symmetry and quadrature properties on overlapping intervals:

$$g_p(t) = g_{p+1}(2a_{p+1}-t) \text{ for } t \in [a_{p+1}-\eta_{p+1}, a_{p+1}+\eta_{p+1}] \tag{8.96}$$

and

$$g_p^2(t)+g_{p+1}^2(t) = 1 \text{ for } t \in [a_{p+1}-\eta_{p+1}, a_{p+1}+\eta_{p+1}].$$

Each $e_k \in \mathbf{L}^2[0,1]$ is extended over $\mathbb{R}$ into a function $\tilde{e}_k$ that is symmetric with respect to 0 and antisymmetric with respect to 1. The resulting $\tilde{e}_k$ has period 4 and is defined over $[-2,2]$ by

$$\tilde{e}_k(t) = \begin{cases} e_k(t) & \text{if } t \in [0,1] \\ e_k(-t) & \text{if } t \in (-1,0) \\ -e_k(2-t) & \text{if } t \in [1,2) \\ -e_k(2+t) & \text{if } t \in [-1,-2) \end{cases}.$$

The following theorem derives an orthonormal basis of $\mathbf{L}^2(\mathbb{R})$.

Theorem 8.10 (COIFMAN, MALVAR, MEYER) *Let $\{e_k\}_{k\in\mathbb{N}}$ be an orthonormal basis of $\mathbf{L}^2[0,1]$. The family*

$$\left\{ g_{p,k}(t) = g_p(t)\frac{1}{\sqrt{l_p}}\tilde{e}_k\left(\frac{t-a_p}{l_p}\right)\right\}_{k\in\mathbb{N},\, p\in\mathbb{Z}} \tag{8.97}$$

is an orthonormal basis of $\mathbf{L}^2(\mathbb{R})$.

Proof[2]. Since $\tilde{e}_k(l_p^{-1}(t-a_p))$ is symmetric with respect to a_p and antisymmetric with respect to a_{p+1} it follows from (8.86) that $g_{p,k} \in \mathbf{W}^p$ for all $k \in \mathbb{N}$. Proposition 8.6 proves that the spaces $\mathbf{W}^p$ and $\mathbf{W}^q$ are orthogonal for $p \neq q$ and that $\mathbf{L}^2(\mathbb{R}) = \ominus_{p=-\infty}^{+\infty} \mathbf{W}^p$. To prove that (8.97) is an orthonormal basis of $\mathbf{L}^2(\mathbb{R})$ we thus need to show that

$$\left\{ g_{p,k}(t) = g_p(t)\frac{1}{\sqrt{l_p}}\tilde{e}_k\left(\frac{t-a_p}{l_p}\right)\right\}_{k\in\mathbb{N},p\in\mathbb{Z}} \tag{8.98}$$

is an orthonormal basis of $\mathbf{W}^p$.

Let us prove first that any $f \in \mathbf{W}^p$ can be decomposed over this family. Such a function can be written $f(t) = g_p(t)h(t)$ where the restriction of h to $[a_p, a_{p+1}]$ is arbitrary, and h is respectively symmetric and antisymmetric with respect to a_p and a_{p+1}. Since $\{\tilde{e}_k\}_{k\in\mathbb{N}}$ is an orthonormal basis of $\mathbf{L}^2[0,1]$, clearly

$$\left\{ \frac{1}{\sqrt{l_p}}\tilde{e}_k\left(\frac{t-a_p}{l_p}\right)\right\}_{k\in\mathbb{N}} \tag{8.99}$$

is an orthonormal basis of $\mathbf{L}^2[a_p, a_{p+1}]$. The restriction of h to $[a_p, a_{p+1}]$ can therefore be decomposed in this basis. This decomposition remains valid for all $t \in [a_p - \eta_p, a_{p+1} + \eta_{p+1}]$ since $h(t)$ and the $l_p^{-1/2} \tilde{e}_k(l_p^{-1}(t - a_p))$ have the same symmetry with respect to a_p and a_{p+1}. Therefore $f(t) = h(t) g_p(t)$ can be decomposed over the family (8.98). The following lemma finishes the proof by showing that the orthogonality of functions in (8.98) is a consequence of the orthogonality of (8.99) in $\mathbf{L}^2[a_p, a_{p+1}]$.

Lemma 8.1 *If* $f_b(t) = h_b(t)\, g_p(t) \in \mathbf{W}^p$ *and* $f_c(t) = h_c(t)\, g_p(t) \in \mathbf{W}^p$, *then*

$$\langle f_b, f_c \rangle = \int_{a_p - \eta_p}^{a_{p+1} + \eta_{p+1}} f_b(t)\, f_c^*(t)\, dt = \int_{a_p}^{a_{p+1}} h_b(t)\, h_c^*(t)\, dt. \tag{8.100}$$

Let us evaluate

$$\langle f_b, f_c \rangle = \int_{a_p - \eta_p}^{a_{p+1} + \eta_{p+1}} h_b(t)\, h_c^*(t)\, g_p^2(t)\, dt. \tag{8.101}$$

We know that $h_b(t)$ and $h_c(t)$ are symmetric with respect to a_p so

$$\int_{a_p - \eta_p}^{a_p + \eta_p} h_b(t)\, h_c^*(t)\, g_p^2(t)\, dt = \int_{a_p}^{a_p + \eta_p} h_b(t)\, h_c^*(t)\, [g_p^2(t) + g_p^2(2a_p - t)]\, dt.$$

Since $g_p^2(t) + g_p^2(2a_{p+1} - t) = 1$ over this interval, we obtain

$$\int_{a_p - \eta_p}^{a_p + \eta_p} h_b(t)\, h_c^*(t)\, g_p^2(t)\, dt = \int_{a_p}^{a_p + \eta_p} h_b(t)\, h_c(t)\, dt. \tag{8.102}$$

The functions $h_b(t)$ and $h_c(t)$ are antisymmetric with respect to a_{p+1} so $h_b(t) h_c^*(t)$ is symmetric about a_{p+1}. We thus prove similarly that

$$\int_{a_{p+1} - \eta_{p+1}}^{a_{p+1} + \eta_{p+1}} h_b(t)\, h_c^*(t)\, g_{p+1}^2(t)\, dt = \int_{a_{p+1} - \eta_{p+1}}^{a_{p+1}} h_b(t)\, h_c^*(t)\, dt. \tag{8.103}$$

Since $g_p(t) = 1$ for $t \in [a_p + \eta_p, a_{p+1} - \eta_{p+1}]$, inserting (8.102) and (8.103) in (8.101) proves the lemma property (8.100). ■

Theorem 8.10 is similar to the block basis Theorem 8.3 but it has the advantage of using smooth windows g_p as opposed to the rectangular windows that are indicator functions of $[a_p, a_{p+1}]$. It yields smooth functions $g_{p,k}$ only if the extension $\tilde{e}_k$ of e_k is a smooth function. This is the case for the cosine IV basis $\{e_k(t) = \sqrt{2} \cos[(k + 1/2)\pi t]\}_{k \in \mathbb{N}}$ of $\mathbf{L}^2[0,1]$ defined in Theorem 8.6. Indeed $\cos[(k + 1/2)\pi t]$ has a natural symmetric and antisymmetric extension with respect to 0 and 1 over $\mathbb{R}$. The following corollary derives a local cosine basis.

Corollary 8.1 *The family of local cosine functions*

$$\left\{ g_{p,k}(t) = g_p(t) \sqrt{\frac{2}{l_p}} \cos\left[\pi \left(k + \frac{1}{2} \right) \frac{t - a_p}{l_p} \right] \right\}_{k \in \mathbb{N},\, p \in \mathbb{Z}} \tag{8.104}$$

is an orthonormal basis of $\mathbf{L}^2(\mathbb{R})$.

Cosine-Sine I Basis Other bases can be constructed with functions having a different symmetry. To maintain the orthogonality of the windowed basis, we must ensure that consecutive windows g_p and g_{p+1} are multiplied by functions that have an opposite symmetry with respect to a_{p+1}. For example, we can multiply g_{2p} with functions that are symmetric with respect to both ends a_{2p} and a_{2p+1}, and multiply g_{2p+1} with functions that are antisymmetric with respect to a_{2p+1} and a_{2p+2}. Such bases can be constructed with the cosine I basis $\{\sqrt{2}\lambda_k \cos(\pi k t)\}_{k\in\mathbb{Z}}$ defined in Theorem 8.5, with $\lambda_0 = 2^{-1/2}$ and $\lambda_k = 1$ for $k \neq 0$, and with the sine I family $\{\sqrt{2}\sin(\pi k t)\}_{k\in\mathbb{N}^*}$, which is also an orthonormal basis of $\mathbf{L}^2[0,1]$. The reader can verify that if

$$g_{2p,k}(t) = g_{2p}(t)\sqrt{\frac{2}{l_{2p}}}\,\lambda_k \cos\left[\pi k \frac{t-a_{2p}}{l_{2p}}\right]$$

$$g_{2p+1,k}(t) = g_{2p+1}(t)\sqrt{\frac{2}{l_{2p+1}}}\,\sin\left[\pi k \frac{t-a_{2p+1}}{l_{2p+1}}\right]$$

then $\{g_{p,k}\}_{k\in\mathbb{N},p\in\mathbb{Z}}$ is an orthonormal basis of $\mathbf{L}^2(\mathbb{R})$.

Lapped Transforms in Frequency Lapped orthogonal projectors can also divide the frequency axis in separate overlapping intervals. This is done by decomposing the Fourier transform $\hat{f}(\omega)$ of $f(t)$ over a local cosine basis defined on the frequency axis $\{g_{p,k}(\omega)\}_{p\in\mathbb{Z},k\in\mathbb{N}}$. This is also equivalent to decomposing $f(t)$ on its inverse Fourier transform $\{\frac{1}{2\pi}\hat{g}_{p,k}(-t)\}_{p\in\mathbb{Z},k\in\mathbb{N}}$. As opposed to wavelet packets, which decompose signals in dyadic frequency bands, this approach offers complete flexibility on the size of the frequency intervals $[a_p - \eta_p, a_{p+1} + \eta_{p+1}]$.

A signal decomposition in a Meyer wavelet or wavelet packet basis can be calculated with a lapped orthogonal transform applied in the Fourier domain. Indeed, the Fourier transform (7.92) of a Meyer wavelet has a compact support and $\{|\hat{\psi}(2^j\omega)|\}_{j\in\mathbb{Z}}$ can be considered as a family asymmetric windows, whose supports only overlap with adjacent windows with appropriate symmetry properties. These windows cover the whole frequency axis: $\sum_{j=-\infty}^{+\infty} |\hat{\psi}(2^j\omega)|^2 = 1$. As a result, the Meyer wavelet transform can be viewed as a lapped orthogonal transform applied in the Fourier domain. It can thus be efficiently implemented with the folding algorithm of Section 8.4.4.

8.4.3 Local Cosine Bases

The local cosine basis defined in (8.104) is composed of functions

$$g_{p,k}(t) = g_p(t)\sqrt{\frac{2}{l_p}}\cos\left[\pi\left(k+\frac{1}{2}\right)\frac{t-a_p}{l_p}\right]$$

with a compact support $[a_p - \eta_p, a_{p+1} + \eta_{p+1}]$. The energy of their Fourier transforms is also well concentrated. Let $\hat{g}_p$ be the Fourier transform of g_p,

$$\hat{g}_{p,k}(\omega) = \frac{\exp(-ia_p\xi_{p,k})}{2}\sqrt{\frac{2}{l_p}}\Big(\hat{g}_p(\omega-\xi_{p,k}) + \hat{g}_p(\omega+\xi_{p,k})\Big),$$

where

$$\xi_{p,k} = \frac{\pi(k+1/2)}{l_p}.$$

The bandwidth of $\hat{g}_{p,k}$ around $\xi_{p,k}$ and $-\xi_{p,k}$ is equal to the bandwidth of $\hat{g}_p$. If the sizes η_p and η_{p+1} of the variation intervals of g_p are proportional to l_p, then this bandwidth is proportional to l_p^{-1}.

For smooth functions f, we want to guarantee that the inner products $\langle f, g_{p,k}\rangle$ have a fast decay when the center frequency $\xi_{p,k}$ increases. The Parseval formula proves that

$$\langle f, g_{p,k}\rangle = \frac{\exp(ia_p\xi_{p,k})}{2\pi}\sqrt{\frac{2}{l_p}}\int_{-\infty}^{+\infty}\hat{f}(\omega)\left(\hat{g}_p^*(\omega-\xi_{p,k}) + \hat{g}_p^*(\omega+\xi_{p,k})\right)d\omega.$$

The smoothness of f implies that $|\hat{f}(\omega)|$ has a fast decay at large frequencies ω. This integral will therefore become small when $\xi_{p,k}$ increases if g_p is a smooth window, because $|\hat{g}_p(\omega)|$ has a fast decay.

Window Design The regularity of g_p depends on the regularity of the profile β which defines it in (8.85). This profile must satisfy

$$\beta^2(t) + \beta^2(-t) = 1 \quad \text{for } t \in [-1,1], \tag{8.105}$$

plus $\beta(t) = 0$ if $t < -1$ and $\beta(t) = 1$ if $t > 1$. One example is

$$\beta_0(t) = \sin\Big(\frac{\pi}{4}(1+t)\Big) \quad \text{for } t \in [-1,1],$$

but its derivative at $t = \pm 1$ is non-zero so β is not differentiable at ± 1. Windows of higher regularity are constructed with a profile β_k defined by induction for $k \geq 0$ by

$$\beta_{k+1}(t) = \beta_k\Big(\sin\frac{\pi t}{2}\Big) \quad \text{for } t \in [-1,1].$$

For any $k \geq 0$, one can verify that β_k satisfies (8.105) and has $2^k - 1$ vanishing derivatives at $t = \pm 1$. The resulting β and g_p are therefore $2^k - 1$ times continuously differentiable.

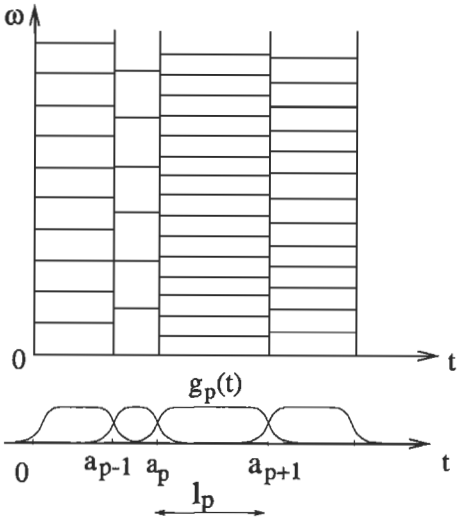


FIGURE 8.18 The Heisenberg boxes of local cosine vectors define a regular grid over the time-frequency plane.

Heisenberg Box A local cosine basis can be symbolically represented as an exact paving of the time-frequency plane. The time and frequency region of high energy concentration for each local cosine vector $g_{p,k}$ is approximated by a Heisenberg rectangle

$$[a_p, a_{p+1}] \times \left[\xi_{p,k} - \frac{\pi}{2l_p}, \xi_{p,k} + \frac{\pi}{2l_p}\right],$$

as illustrated in Figure 8.18. A local cosine basis $\{g_{p,k}\}_{k\in\mathbb{N},p\in\mathbb{Z}}$ corresponds to a time-frequency grid whose size varies in time.

Figure 8.19(a) shows the decomposition of a digital recording of the sound "grea" coming from the word "greasy". The window sizes are adapted to the signal structures with the best basis algorithm described in Section 9.4.2. High amplitude coefficients are along spectral lines in the time-frequency plane, which correspond to different harmonics. Most Heisenberg boxes appear in white, which indicates that the corresponding inner product is nearly zero. This signal can thus be approximated with a few non-zero local cosine vectors. Figure 8.19(b) decomposes the same signal in a local cosine basis composed of small windows of constant size. The signal time-frequency structures do not appear as well as in Figure 8.19(a).

Translation and Phase Cosine modulations as opposed to complex exponentials do not provide easy access to phase information. The translation of a signal can induce important modifications of its decomposition coefficients in a cosine basis.

Consider for example

$$f(t) = g_{p,k}(t) = g_p(t)\sqrt{\frac{2}{l_p}}\cos\left[\pi\left(k+\frac{1}{2}\right)\frac{t-a_p}{l_p}\right].$$

Since the basis is orthogonal, $\langle f, g_{p,k}\rangle = 1$, and all other inner products are zero. After a translation by $\tau = l_p/(2k+1)$

$$f_\tau(t) = f\left(t-\frac{l_p}{2k+1}\right) = g_p(t)\sqrt{\frac{2}{l_p}}\sin\left[\pi\left(k+\frac{1}{2}\right)\frac{t-a_p}{l_k}\right].$$

The opposite parity of sine and cosine implies that $\langle f_\tau, g_{p,k}\rangle \approx 0$. In contrast, $\langle f_\tau, g_{p,k-1}\rangle$ and $\langle f_\tau, g_{p,k+1}\rangle$ become non-zero. After translation, a signal component initially represented by a cosine of frequency $\pi(k+1/2)/l_p$ is therefore spread over cosine vectors of different frequencies.

This example shows that the local cosine coefficients of a pattern are severely modified by any translation. We are facing the same translation distortions as observed in Section 5.4 for wavelets and time-frequency frames. This lack of translation invariance makes it difficult to use these bases for pattern recognition.

8.4.4 Discrete Lapped Transforms

Lapped orthogonal bases are discretized by replacing the orthogonal basis of $\mathbf{L}^2[0,1]$ with a discrete basis of $\mathbb{C}^N$, and uniformly sampling the windows g_p. Discrete local cosine bases are derived with discrete cosine-IV bases.

Let $\{a_p\}_{p\in\mathbb{Z}}$ be a sequence of half integers, $a_p + 1/2 \in \mathbb{Z}$ with

$$\lim_{p\to-\infty} a_p = -\infty \quad \text{and} \quad \lim_{p\to+\infty} a_p = +\infty.$$

A discrete lapped orthogonal basis is constructed with the discrete projectors P_p defined in (8.93). These operators are implemented with the sampled windows $g_p[n] = g_p(n)$. Suppose that $\{e_{k,l}[n]\}_{0\le k<l}$ is an orthogonal basis of signals defined for $0 \le n < l$. These vectors are extended over $\mathbb{Z}$ with a symmetry with respect to $-1/2$ and an antisymmetry with respect to $l-1/2$. The resulting extensions have a period $4l$ and are defined over $[-2l, 2l-1]$ by

$$\tilde{e}_{l,k}[n] = \begin{cases} e_{l,k}[n] & \text{if } n \in [0, l-1] \\ e_{l,k}[-1-n] & \text{if } n \in [-l, -1] \\ -e_k[2l-1-n] & \text{if } n \in [l, 2l-1] \\ -e_k[2l+n] & \text{if } n \in [-2l, -l-1] \end{cases}.$$

The following theorem proves that multiplying these vectors with the discrete windows $g_p[n]$ yields an orthonormal basis of $\mathbf{l}^2(\mathbb{Z})$.

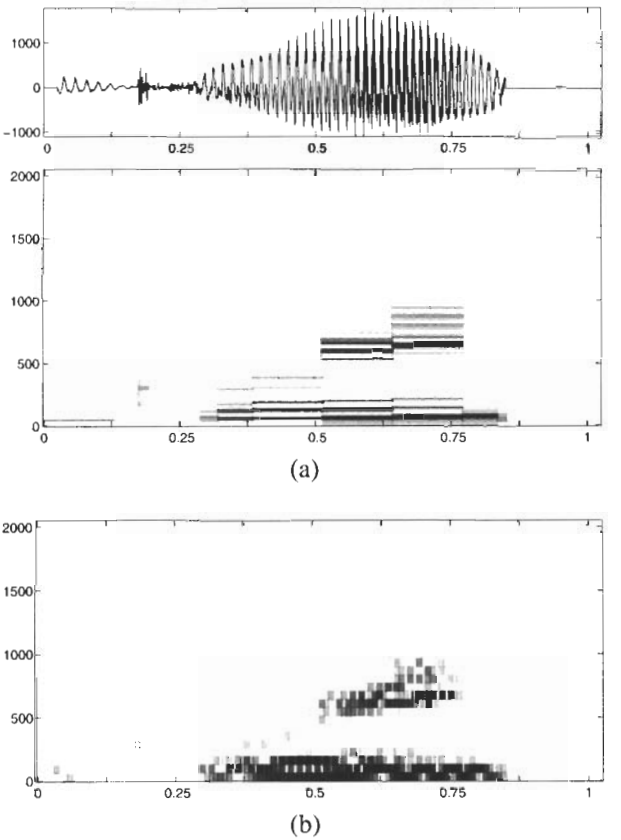


FIGURE 8.19 (a): The signal at the top is a recording of the sound "grea" in the word "greasy". This signal is decomposed in a local cosine basis with windows of varying sizes. The larger the amplitude of $|\langle f, g_{p,k}\rangle|$ the darker the gray level of the Heisenberg box. (b): Decomposition in a local cosine basis with small windows of constant size.

Theorem 8.11 (COIFMAN, MALVAR, MEYER) *Suppose that* $\{e_{k,l}\}_{0\leq k<l}$ *is an orthogonal basis of* $\mathbb{C}^l$, *for any* $l > 0$. *The family*

$$\left\{ g_{p,k}[n] = g_p[n]\, \tilde{e}_{k,l_p}[n - a_p] \right\}_{0\leq k<l_p, p\in\mathbb{Z}} \tag{8.106}$$

is a lapped orthonormal basis of $\mathbf{l}^2(\mathbb{Z})$.

The proof of this theorem is identical to the proof of Theorem 8.10 since we have a discrete equivalent of the spaces $\mathbf{W}^p$ and their projectors. It is also based on a discrete equivalent of Lemma 8.1, which is verified with the same derivations.

Beyond the proof of Theorem 8.11, we shall see that this lemma is important for quickly computing the decomposition coefficients $\langle f, g_{p,k} \rangle$.

Lemma 8.2 *Any* $f_b[n] = g_p[n]\, h_b[n] \in \mathbf{W}^p$ *and* $f_c[n] = g_p[n]\, h_c[n] \in \mathbf{W}^p$ *satisfy*

$$\langle f_b, f_c \rangle = \sum_{a_p - \eta_p < n < a_{p+1} + \eta_{p+1}} f_b[n]\, f_c^*[n] = \sum_{a_p < n < a_{p+1}} h_b[n]\, h_c^*[n]. \tag{8.107}$$

Theorem 8.11 is similar to the discrete block basis Theorem 8.4 but constructs an orthogonal basis with smooth discrete windows $g_p[n]$. The discrete cosine IV bases

$$\left\{ e_{l,k}[n] = \sqrt{\frac{2}{l}} \cos\left[\frac{\pi}{l} \left(k + \frac{1}{2}\right) \left(n + \frac{1}{2}\right) \right] \right\}_{0 \le k < l}$$

have the advantage of including vectors that have a natural symmetric and anti-symmetric extension with respect to $-1/2$ and $l - 1/2$. This produces a discrete local cosine basis of $\mathbf{l}^2(\mathbb{Z})$.

Corollary 8.2 *The family*

$$\left\{ g_{p,k}[n] = g_p[n] \sqrt{\frac{2}{l_p}} \cos\left[\pi \left(k + \frac{1}{2}\right) \frac{n - a_p}{l_p} \right] \right\}_{0 \le k < l_p, p \in \mathbb{Z}} \tag{8.108}$$

is an orthonormal basis of $\mathbf{l}^2(\mathbb{Z})$.

Fast Lapped Orthogonal Transform A fast algorithm introduced by Malvar [42] replaces the calculations of $\langle f, g_{p,k} \rangle$ by a computation of inner products in the original bases $\{e_{l,k}\}_{0 \le k < l}$, with a folding procedure. In a discrete local cosine basis, these inner products are calculated with the fast DCT-IV algorithm.

To simplify notations, as in Section 8.4.1 we decompose $I_p = [a_p - \eta_p, a_{p+1} + \eta_{p+1}]$ into $I_p = O_p \cup C_p \cup O_{p+1}$ with

$$O_p = [a_p - \eta_p, a_p + \eta_p] \quad \text{and} \quad C_p = [a_p + \eta_p, a_{p+1} - \eta_{p+1}].$$

The orthogonal projector P_p on the space $\mathbf{W}^p$ generated by $\{g_{p,k}\}_{0 \le k < l_p}$ was calculated in (8.93):

$$P_p f[n] = g_p[n]\, h_p[n],$$

where h_p is a folded version of f:

$$h_p[n] = \begin{cases} g_p[n]\, f[n] + g_p[2a_p - n]\, f[2a_p - n] & \text{if } n \in O_p \\ f[n] & \text{if } n \in C_p \\ g_p[n]\, f[n] - g_p[2a_{p+1} - n]\, f[2a_{p+1} - n] & \text{if } n \in O_{p+1} \end{cases}. \tag{8.109}$$

Since $g_{p,k} \in \mathbf{W}^p$,

$$\langle f, g_{p,k} \rangle = \langle P_p f, g_{p,k} \rangle = \langle g_p h_p, g_p \tilde{e}_{l_p,k} \rangle.$$

Since $\tilde{e}_{l_p,k}[n] = e_{l_p,k}[n]$ for $n \in [a_p, a_{p+1}]$, Lemma 8.2 derives that

$$\langle f, g_{p,k} \rangle = \sum_{a_p < n < a_{p+1}} h_p[n]\, e_{l_p,k}[n] = \langle h_p, e_{l_p,k} \rangle_{[a_p, a_{p+1}]}. \tag{8.110}$$

This proves that the decomposition coefficients $\langle f, g_{p,k} \rangle$ can be calculated by folding f into h_p and computing the inner product with the orthogonal basis $\{e_{l_p,k}\}_{0 \le k < l_p}$ defined over $[a_p, a_{p+1}]$.

For a discrete cosine basis, the DCT-IV coefficients

$$\langle h_p, e_{l_p,k} \rangle_{[a_p, a_{p+1}]} = \sum_{a_p < n < a_{p+1}} h_p[n] \sqrt{\frac{2}{l_p}} \cos\left[\pi\left(k + \frac{1}{2}\right)\frac{n - a_p}{l_p}\right] \tag{8.111}$$

are computed with the fast DCT-IV algorithm of Section 8.3.4, which requires $O(l_p \log_2 l_p)$ operations. The inverse lapped transform recovers $h_p[n]$ over $[a_p, a_{p+1}]$ from the l_p inner products $\{\langle h_p, e_{l_p,k} \rangle_{[a_p, a_{p+1}]}\}_{0 \le k < l_p}$. In a local cosine IV basis, this is done with the fast inverse DCT-IV, which is identical to the forward DCT-IV and requires $O(l_p \log_2 l_p)$ operations. The reconstruction of f is done by applying (8.95) which proves that

$$f[n] = \sum_{p=-\infty}^{+\infty} P_p f[n] = \sum_{p=-\infty}^{+\infty} g_p[n]\, h_p[n]. \tag{8.112}$$

Let us denote $O_p^- = [a_p - \eta_p, a_p]$ and $O_p^+ = [a_p, a_p + \eta_p]$. The restriction of (8.112) to $[a_p, a_{p+1}]$ gives

$$f[n] = \begin{cases} g_p[n]\, h_p[n] + g_{p-1}[n]\, h_{p-1}[n] & \text{if } n \in O_p^+ \\ h_p[n] & \text{if } n \in C_p \\ g_p[n]\, h_p[n] + g_{p+1}[n]\, h_{p+1}[n] & \text{if } n \in O_{p+1}^- \end{cases}$$

The symmetry of the windows guarantees that $g_{p-1}[n] = g_p[2a_p - n]$ and $g_{p+1}[n] = g_p[2a_{p+1} - n]$. Since $h_{p-1}[n]$ is antisymmetric with respect to a_p and $h_{p+1}[n]$ is symmetric with respect to a_{p+1}, we can recover $f[n]$ on $[a_p, a_{p+1}]$ from the values of $h_{p-1}[n]$, $h_p[n]$ and $h_{p+1}[n]$ computed respectively on $[a_{p-1}, a_p]$, $[a_p, a_{p+1}]$, and $[a_{p+1}, a_{p+2}]$:

$$f[n] = \begin{cases} g_p[n]\, h_p[n] - g_p[2a_p - n]\, h_{p-1}[2a_p - n] & \text{if } n \in O_p^+ \\ h_p[n] & \text{if } n \in C_p \\ g_p[n]\, h_p[n] + g_p[2a_{p+1} - n]\, h_{p+1}[2a_{p+1} - n] & \text{if } n \in O_{p+1}^- \end{cases} \tag{8.113}$$

This unfolding formula is implemented with $O(l_p)$ calculations. The inverse local cosine transform thus requires $O(l_p \log_2 l_p)$ operations to recover $f[n]$ on each interval $[a_p, a_{p+1}]$ of length l_p.

Finite Signals If $f[n]$ is defined for $0 \leq n < N$, the extremities of the first and last interval must be $a_0 = -1/2$ and $a_q = N - 1/2$. A fast local cosine algorithm needs $O(l_p \log_2 l_p)$ additions and multiplications to decompose or reconstruct the signal on each interval of length l_p. On the whole signal of length N, it thus needs a total of $O(N \log_2 L)$ operations, where $L = \sup_{0 \leq p < q} l_p$.

Since we do not know the values of $f[n]$ for $n < 0$, at the left border we set $\eta_0 = 0$. This means that $g_0[n]$ jumps from 0 to 1 at $n = 0$. The resulting transform on the left boundary is equivalent to a straight DCT-IV. Section 8.3.2 shows that since cosine IV vectors are even on the left boundary, the DCT-IV is equivalent to a symmetric signal extension followed by a discrete Fourier transform. This avoids creating discontinuity artifacts at the left border.

At the right border, we also set $\eta_q = 0$ to limit the support of g_{q-1} to $[0, N-1]$. Section 8.4.4 explains that since cosine IV vectors are odd on the right boundary, the DCT-IV is equivalent to an antisymmetric signal extension. If $f[N-1] \neq 0$, this extension introduces a sharp signal transition that creates artificial high frequencies. To reduce this border effect, we replace the cosine IV modulation

$$g_{q-1,k}[n] = g_{q-1}[n] \sqrt{\frac{2}{l_{q-1}}} \cos\left[\pi\left(k + \frac{1}{2}\right)\frac{n - a_{q-1}}{l_{q-1}}\right]$$

by a cosine I modulation

$$g_{q-1,k}[n] = g_{q-1}[n] \sqrt{\frac{2}{l_{q-1}}}\, \lambda_k \cos\left[\pi k \frac{n - a_{q-1}}{l_{q-1}}\right].$$

The orthogonality with the other elements of the basis is maintained because these cosine I vectors, like cosine IV vectors, are even with respect to a_{q-1}. Since $\cos[\pi k n - a_{q-1}/l_{q-1}]$ is also symmetric with respect to $a_q = N - 1/2$, computing a DCT-I is equivalent to performing a symmetric signal extension at the right boundary, which avoids discontinuities. In the fast local cosine transform, we thus compute a DCT-I of the last folded signal h_{q-1} instead of a DCT-IV. The reconstruction algorithm uses an inverse DCT-I to recover h_{q-1} from these coefficients.

8.5 LOCAL COSINE TREES [2]

Corollary 8.1 constructs local cosine bases for any segmentation of the time axis into intervals $[a_p, a_{p+1}]$ of arbitrary lengths. This result is more general than the construction of wavelet packet bases that can only divide the frequency axis into dyadic intervals, whose length are proportional to powers of 2. However, Coifman and Meyer [138] showed that restricting the intervals to dyadic sizes has the advantage of creating a tree structure similar to a wavelet packet tree. "Best" local cosine bases can then be adaptively chosen with the fast dynamical programming algorithm described in Section 9.4.2.

8.5.1 Binary Tree of Cosine Bases

A local cosine tree includes orthogonal bases that segment the time axis in dyadic intervals. For any $j \geq 0$, the interval $[0,1]$ is divided in 2^j intervals of length 2^{-j} by setting

$$a_{p,j} = p2^{-j} \text{ for } 0 \leq p \leq 2^j.$$

These intervals are covered by windows $g_{p,j}$ defined by (8.85) with a support $[a_{p,j} - \eta, a_{p+1,j} + \eta]$:

$$g_{p,j}(t) = \begin{cases} \beta(\eta^{-1}(t - a_{p,j})) & \text{if } t \in [a_{p,j} - \eta, a_{p,j} + \eta] \\ 1 & \text{if } t \in [a_{p,j} + \eta, a_{p+1,j} - \eta] \\ \beta(\eta^{-1}(a_{p+1,j} - t)) & \text{if } t \in [a_{p+1,j} - \eta, a_{p+1,j} + \eta] \\ 0 & \text{otherwise} \end{cases} \tag{8.114}$$

To ensure that the support of $g_{p,j}$ is in $[0,1]$ for $p = 0$ and $p = 2^j - 1$, we modify respectively the left and right sides of these windows by setting $g_{0,j}(t) = 1$ if $t \in [0, \eta]$, and $g_{2^j-1,j}(t) = 1$ if $t \in [1 - \eta, 1]$. It follows that $g_{0,0} = \mathbf{1}_{[0,1]}$. The size η of the raising and decaying profiles of $g_{p,j}$ is independent of j. To guarantee that windows overlap only with their two neighbors, the length $a_{p+1,j} - a_{p,j} = 2^{-j}$ must be larger than the size 2η of the overlapping intervals and hence

$$\eta \leq 2^{-j-1}. \tag{8.115}$$

Similarly to wavelet packet trees, a local cosine tree is constructed by recursively dividing spaces built with local cosine bases. A tree node at a depth j and a position p is associated to a space $\mathbf{W}_j^p$ generated by the local cosine family

$$\mathcal{B}_j^p = \left\{ g_{p,j}(t) \sqrt{\frac{2}{2^{-j}}} \cos\left[\pi \left(k + \frac{1}{2} \right) \frac{t - a_{p,j}}{2^{-j}} \right] \right\}_{k \in \mathbb{Z}}. \tag{8.116}$$

Any $f \in \mathbf{W}_p^j$ has a support in $[a_{p,j} - \eta, a_{p+1,j} + \eta]$ and can be written $f(t) = g_{p,j}(t)\, h(t)$ where $h(t)$ is respectively symmetric and antisymmetric with respect to $a_{p,j}$ and $a_{p+1,j}$. The following proposition shows that $\mathbf{W}_j^p$ is divided in two orthogonal spaces $\mathbf{W}_{j+1}^{2p}$ and $\mathbf{W}_{j+1}^{2p+1}$ that are built over the two half intervals.

Proposition 8.7 (COIFMAN, MEYER) *For any $j \geq 0$ and $p < 2^j$, the spaces $\mathbf{W}_{j+1}^{2p}$ and $\mathbf{W}_{j+1}^{2p+1}$ are orthogonal and*

$$\mathbf{W}_j^p = \mathbf{W}_{j+1}^{2p} \oplus \mathbf{W}_{j+1}^{2p+1}. \tag{8.117}$$

Proof[2]. The orthogonality of $\mathbf{W}_{j+1}^{2p}$ and $\mathbf{W}_{j+1}^{2p+1}$ is proved by Proposition 8.6. We denote $P_{p,j}$ the orthogonal projector on $\mathbf{W}_j^p$. With the notation of Section 8.4.1, this projector is decomposed into two splitting projectors at $a_{p,j}$ and $a_{p+1,j}$:

$$P_{p,j} = P^+_{a_{p,j},\eta} P^-_{a_{p+1,j},\eta}.$$

Equation (8.90) proves that

$$P_{2p,j+1} + P_{2p+1,j+1} = P^{+}_{a_{2p,j+1},\eta} P^{-}_{a_{2p+2,j+1},\eta} = P^{+}_{a_{p,j},\eta} P^{-}_{a_{p+1,j},\eta} = P_{p,j}.$$

This equality on orthogonal projectors implies (8.117). ■

The space $\mathbf{W}_j^p$ located at the node (j, p) of a local cosine tree is therefore the sum of the two spaces $\mathbf{W}_{j+1}^{2p}$ and $\mathbf{W}_{j+1}^{2p+1}$ located at the children nodes. Since $g_{0,0} = \mathbf{1}_{[0,1]}$ it follows that $\mathbf{W}_0^0 = \mathbf{L}^2[0,1]$. The maximum depth J of the binary tree is limited by the support condition $\eta \leq 2^{-J-1}$, and hence

$$J \leq -\log_2(2\eta). \tag{8.118}$$

Admissible Local Cosine Bases As in a wavelet packet binary tree, many local cosine orthogonal bases are constructed from this local cosine tree. We call *admissible binary tree* any subtree of the local cosine tree whose nodes have either 0 or 2 children. Let $\{j_i, p_i\}_{1 \leq i \leq I}$ be the indices at the leaves of a particular admissible binary tree. Applying the splitting property (8.117) along the branches of this subtree proves that

$$\mathbf{L}^2[0,1] = \mathbf{W}_0^0 = \oplus_{i=1}^{I} \mathbf{W}_{j_i}^{p_i}.$$

Hence, the union of local cosine bases $\cup_{i=1}^{I} \mathcal{B}_{j_i}^{p_i}$ is an orthogonal basis of $\mathbf{L}^2[0,1]$. This can also be interpreted as a division of the time axis into windows of various length, as illustrated by Figure 8.20.

The number B_J of different dyadic local cosine bases is equal to the number of different admissible subtrees of depth at most J. For $J = -\log_2(2\eta)$, Proposition 8.1 proves that

$$2^{1/(4\eta)} \leq B_J \leq 2^{3/(8\eta)}.$$

Figure 8.19 shows the decomposition of a sound recording in two dyadic local cosine bases selected from the binary tree. The basis in (a) is calculated with the best basis algorithm of Section 9.4.2.

Choice of η At all scales 2^j, the windows $g_{p,j}$ of a local cosine tree have raising and decaying profiles of the same size η. These windows can thus be recombined independently from their scale. If η is small compared to the interval size 2^{-j} then $g_{p,j}$ has a relatively sharp variation at its borders compared to the size of its support. Since η is not proportional to 2^{-j}, the energy concentration of $\hat{g}_{p,j}$ is not improved when the window size 2^{-j} increases. Even though f may be very smooth over $[a_{p,j}, a_{p+1,j}]$, the border variations of the window create relatively large coefficients up to a frequency of the order of π/η.

To reduce the number of large coefficients we must increase η, but this also increases the minimum window size in the tree, which is $2^{-J} = 2\eta$. The choice of η is therefore the result of a trade-off between window regularity and the maximum resolution of the time subdivision. There is no equivalent limitation in the construction of wavelet packet bases.

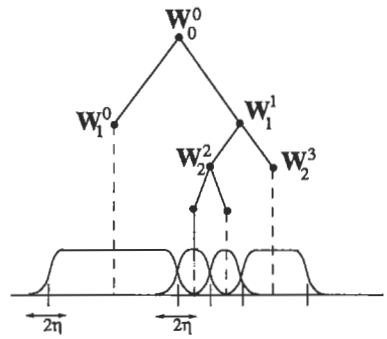


FIGURE 8.20 An admissible binary tree of local cosine spaces divides the time axis in windows of dyadic lengths.

8.5.2 Tree of Discrete Bases

For discrete signals of size N, a binary tree of discrete cosine bases is constructed like a binary tree of continuous time cosine bases. To simplify notations, the sampling distance is normalized to 1. If it is equal to N^{-1} then frequency parameters must be multiplied by N.

The subdivision points are located at half integers:

$$a_{p,j} = pN2^{-j} - 1/2 \quad \text{for } 0 \leq p \leq 2^j.$$

The discrete windows are obtained by sampling the windows $g_p(t)$ defined in (8.114), $g_{p,j}[n] = g_{p,j}(n)$. The same border modification is used to ensure that the support of all $g_{p,j}[n]$ is in $[0, N-1]$.

A node at depth j and position p in the binary tree corresponds to the space $\mathbf{W}_j^p$ generated by the discrete local cosine family

$$\mathcal{B}_j^p = \left\{ g_{p,j}[n] \sqrt{\frac{2}{2^{-j}N}} \cos\left[\pi\left(k+\frac{1}{2}\right)\frac{n-a_{p,j}}{2^{-j}N}\right] \right\}_{0 \leq k < N2^{-j}}.$$

Since $g_{0,0} = \mathbf{1}_{[0,N-1]}$, the space $\mathbf{W}_0^0$ at the root of the tree includes any signal defined over $0 \leq n < N$, so $\mathbf{W}_0^0 = \mathbb{C}^N$. As in Proposition 8.7 we verify that $\mathbf{W}_j^p$ is orthogonal to $\mathbf{W}_j^q$ for $p \neq q$ and that

$$\mathbf{W}_j^p = \mathbf{W}_{j+1}^{2p} \oplus \mathbf{W}_{j+1}^{2p+1}. \tag{8.119}$$

The splitting property (8.119) implies that the union of local cosine families $\mathcal{B}_j^p$ located at the leaves of an admissible subtree is an orthogonal basis of $\mathbf{W}_0^0 = \mathbb{C}^N$. The minimum window size is limited by $2\eta \leq 2^{-j}N$ so the maximum depth of this binary tree is $J = \log_2 \frac{N}{2\eta}$. One can thus construct more than $2^{2^{J-1}} = 2^{N/(4\eta)}$ different discrete local cosine bases within this binary tree.

Fast Calculations The fast local cosine transform algorithm described in Section 8.4.4 requires $O(2^{-j}N\log_2(2^{-j}N))$ operations to compute the inner products of f with the $2^{-j}N$ vectors in the local cosine family $\mathcal{B}_j^p$. The total number of operations to perform these computations at all nodes (j,p) of the tree, for $0 \le p < 2^j$ and $0 \le j \le J$, is therefore $O(NJ\log_2 N)$. The local cosine decompositions in Figure 8.19 are calculated with this fast algorithm. To improve the right border treatment, Section 8.4.4 explains that the last DCT-IV should be replaced by a DCT-I, at each scale 2^j. The signal f is recovered from the local cosine coefficients at the leaves of any admissible binary tree, with the fast local cosine reconstruction algorithm, which needs $O(N\log_2 N)$ operations.

8.5.3 Image Cosine Quad-Tree

A local cosine binary tree is extended in two dimensions into a quad-tree, which recursively divides square image windows into four smaller windows. This separable approach is similar to the extension of wavelet packet bases in two dimensions, described in Section 8.2.

Let us consider images of N^2 pixels. A node of the quad-tree is labeled by its depth j and two indices p and q. Let $g_{p,j}[n]$ be the discrete one-dimensional window defined in Section 8.5.2. At the depth j, a node (p,q) corresponds to a separable space

$$\mathbf{W}_j^{p,q} = \mathbf{W}_j^p \otimes \mathbf{W}_j^q, \tag{8.120}$$

which is generated by a separable local cosine basis of $2^{-2j}N^2$ vectors

$$\mathcal{B}_j^{p,q} = \left\{ g_{p,j}[n_1]\, g_{q,j}[n_2] \frac{2}{2^{-j}N} \quad \cos\left[\pi\left(k_1+\frac{1}{2}\right)\frac{n_1-a_{p,j}}{2^{-j}N}\right] \right.$$
$$\left. \cos\left[\pi\left(k_2+\frac{1}{2}\right)\frac{n_2-a_{q,j}}{2^{-j}N}\right] \right\}_{0\le k_1,k_2<2^{-j}N}$$

We know from (8.119) that

$$\mathbf{W}_j^p = \mathbf{W}_{j+1}^{2p} \oplus \mathbf{W}_{j+1}^{2p+1} \quad \text{and} \quad \mathbf{W}_j^q = \mathbf{W}_{j+1}^{2q} \oplus \mathbf{W}_{j+1}^{2q+1}.$$

Inserting these equations in (8.120) proves that $\mathbf{W}_j^{p,q}$ is the direct sum of four orthogonal subspaces:

$$\mathbf{W}_j^{p,q} = \mathbf{W}_{j+1}^{2p,2q} \oplus \mathbf{W}_{j+1}^{2p+1,2q} \oplus \mathbf{W}_{j+1}^{2p,2q+1} \oplus \mathbf{W}_{j+1}^{2p+1,2q+1}. \tag{8.121}$$

A space $\mathbf{W}_j^{p,q}$ at a node (j,p,q) is therefore decomposed in the four subspaces located at the four children nodes of the quad-tree. This decomposition can also be interpreted as a division of the square window $g_{p,j}[n_1]g_{q,j}[n_2]$ into four subwindows of equal sizes, as illustrated in Figure 8.21. The space located at the root of the tree is

$$\mathbf{W}_0^{0,0} = \mathbf{W}_0^0 \otimes \mathbf{W}_0^0. \tag{8.122}$$

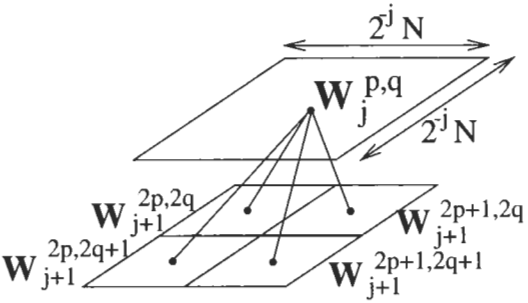


FIGURE 8.21 Functions in $\mathbf{W}_j^{p,q}$ have a support located in a square region of the image. It is divided into four subspaces that cover smaller squares in the image.

FIGURE 8.22 The grid shows the support of the windows $g_{j,p}[n_1]\, g_{j,q}[n_2]$ of a "best" local cosine basis selected in the local cosine quad-tree.

It includes all images of N^2 pixels. The size η of the raising and decaying profiles of the one-dimensional windows defines the maximum depth $J = \log_2 \frac{N}{2\eta}$ of the quad-tree.

Admissible Quad-Trees An admissible subtree of this local cosine quad-tree has nodes that have either 0 or four children. Applying the decomposition property (8.121) along the branches of an admissible quad-tree proves that the spaces $\mathbf{W}_{j_i}^{p_i,q_i}$ located at the leaves decompose $\mathbf{W}_0^{0,0}$ in orthogonal subspaces. The union of the corresponding two-dimensional local cosine bases $\mathcal{B}_{j_i}^{p_i,q_i}$ is therefore an orthogonal basis of $\mathbf{W}_0^{0,0}$. We proved in (8.42) that there are more than $2^{4^{J-1}} = 2^{N^2/16\eta^2}$

different admissible trees of maximum depth $J = \log_2 \frac{N}{2\eta}$. These bases divide the image plane into squares of varying sizes. Figure 8.22 gives an example of image decomposition in a local cosine basis corresponding to an admissible quad-tree. This local cosine basis is selected with the best basis algorithm of Section 9.4.2.

Fast Calculations The decomposition of an image $f[n]$ over a separable local cosine family $\mathcal{B}_j^{p,q}$ requires $O(2^{-2j}N^2 \log_2(2^{-j}N))$ operations, with a separable implementation of the fast one-dimensional local cosine transform. For a full local cosine quad-tree of depth J, these calculations are performed for $0 \leq p, q < 2^j$ and $0 \leq j \leq J$, which requires $O(N^2 J \log_2 N)$ multiplications and additions. The original image is recovered from the local cosine coefficients at the leaves of any admissible subtree with $O(N^2 \log_2 N)$ computations.

8.6 PROBLEMS

8.1. [1] Prove the discrete splitting Theorem 8.2.

8.2. [2] Meyer wavelet packets are calculated with a Meyer conjugate mirror filter (7.89). Compute the size of the frequency support of $\hat{\psi}_j^p$ as a function of 2^j. Study the convergence of $\psi_{j,n}(t)$ when the scale 2^j goes to $+\infty$.

8.3. [1] Extend the separable wavelet packet tree of Section 8.2.2 for discrete p-dimensional signals. Verify that the wavelet packet tree of a p-dimensional discrete signal of N^p samples includes $O(N^p \log_2 N)$ wavelet packet coefficients that are calculated with $O(K N^p \log_2 N)$ operations if the conjugate mirror filter h has K non-zero coefficients.

8.4. [1] Anisotropic wavelet packets $\psi_j^p[a - 2^{L-j}n_1]\,\psi_l^q[b - 2^{L-l}n_2]$ may have different scales 2^j and 2^l along the rows and columns. A decomposition over such wavelet packets is calculated with a filter bank that filters and subsamples the image rows $j - L$ times whereas the columns are filtered and subsampled $l - L$ times. For an image $f[n]$ of N^2 pixels, show that a dictionary of anisotropic wavelet packets includes $O(N^2[\log_2 N]^2)$ different vectors. Compute the number of operations needed to decompose f in this dictionary.

8.5. [1] *Hartley transform* Let $\mathrm{cas}(t) = \cos(t) + \sin(t)$. We define

$$\mathcal{B} = \left\{ g_k[n] = \frac{1}{\sqrt{N}} \mathrm{cas}\left(\frac{2\pi nk}{N}\right) \right\}_{0 \leq k < N}.$$

(a) Prove that $\mathcal{B}$ is an orthonormal basis of $\mathbb{C}^N$.

(b) For any signal $f[n]$ of size N, find a fast Hartley transform algorithm based on the FFT, which computes $\{\langle f, g_k \rangle\}_{0 \leq k < N}$ with $O(N \log_2 N)$ operations.

8.6. [1] Prove that $\{\sqrt{2}\sin[(k+1/2)\pi t]\}_{k \in \mathbf{Z}}$ is an orthonormal basis of $\mathbf{L}^2[0,1]$. Find a corresponding discrete orthonormal basis of $\mathbb{C}^N$.

8.7. [1] Prove that $\{\sqrt{2}\sin(k\pi t)\}_{k \in \mathbf{Z}}$ is an orthonormal basis of $\mathbf{L}^2[0,1]$. Find a corresponding discrete orthonormal basis of $\mathbb{C}^N$.

8.8. [1] *Lapped Fourier basis*

(a) Construct a lapped orthogonal basis $\{\tilde{g}_{p,k}\}_{(p,k) \in \mathbf{Z}}$ of $\mathbf{L}^2(\mathbb{R})$ from the Fourier basis $\{\exp(i2\pi kt)\}_{k \in \mathbf{Z}}$ of $\mathbf{L}^2[0,1]$.

(b) Explain why this local Fourier basis does not contradict the Balian-Low Theorem 5.6.

(c) Let $f \in \mathbf{L}^2(\mathbb{R})$ be such that $|\hat{f}(\omega)| = O((1+|\omega|^p)^{-1})$ for some $p > 0$. Compute the rate of decay of $|\langle f, \tilde{g}_{p,k}\rangle|$ when the frequency index $|k|$ increases. Compare it with the rate of decay of $|\langle f, g_{p,k}\rangle|$, where $g_{p,k}$ is a local cosine vector (8.104). How do the two bases compare for signal processing applications?

8.9. [1] Describe a fast algorithm to compute the Meyer orthogonal wavelet transform with a lapped transform applied in the Fourier domain. Calculate the numerical complexity of this algorithm for periodic signals of size N. Compare this result with the numerical complexity of the standard fast wavelet transform algorithm, where the convolutions with Meyer conjugate mirror filters are calculated with an FFT.

8.10. [2] *Arbitrary Walsh tilings*

(a) Prove that two Walsh wavelet packets $\psi^p_{j,n}$ and $\psi^{p'}_{j',n'}$ are orthogonal if their Heisenberg boxes defined in Section 8.1.2 do not intersect in the time-frequency plane [76].

(b) A dyadic tiling of the time-frequency plane is an exact cover $\{[2^j n, 2^j(n+1)] \times [k\pi 2^{-j}, (k+1)\pi 2^{-j}]\}_{(j,n,p)\in I}$, where the index set I is adjusted to guarantee that the time-frequency boxes do not intersect and that they leave no hole. Prove that any such tiling corresponds to a Walsh orthonormal basis of $\mathbf{L}^2(\mathbb{R})$ $\{\psi^p_{j,n}\}_{(p,j,n)\in I}$.

8.11. [2] *Double tree* We want to construct a dictionary of block wavelet packet bases, which has the freedom to segment both the time and frequency axes. For this purpose, as in a local cosine basis dictionary, we construct a binary tree, which divides $[0,1]$ in 2^j intervals $[p2^{-j}, (p+1)2^{-j}]$, that correspond to nodes indexed by p at the depth j of the tree. At each of these nodes, we construct another tree of wavelet packet orthonormal bases of $\mathbf{L}^2[p2^{-j}, (p+1)2^{-j}]$ [208].

(a) Define admissible sub-trees in this double tree, whose leaves correspond to orthonormal bases of $\mathbf{L}^2[0,1]$. Give an example of an admissible tree and draw the resulting tiling of the time-frequency plane.

(b) Give a recursive equation that relates the number of admissible sub-trees of depth $J+1$ and of depth J. Give an upper bound and a lower bound for the total number of orthogonal bases in this double tree dictionary.

(c) Can one find a basis in a double tree that is well adapted to implement an efficient transform code for audio signals? Justify your answer.

8.12. [2] An anisotropic local cosine basis for images is constructed with rectangular windows that have a width 2^j that may be different from their height 2^l. Similarly to a local cosine tree, such bases are calculated by progressively dividing windows, but the horizontal and vertical divisions of these windows is done independently. Show that a dictionary of anisotropic local cosine bases can be represented as a graph. Implement in WAVELAB an algorithm that decomposes images in a graph of anisotropic local cosine bases.

IX

AN APPROXIMATION TOUR

It is time to wonder why are we constructing so many different orthonormal bases. In signal processing, orthogonal bases are of interest because they can efficiently approximate certain types of signals with just a few vectors. Two examples of such applications are image compression and the estimation of noisy signals, which are studied in Chapters 10 and 11.

Approximation theory studies the error produced by different approximation schemes in an orthonormal basis. A linear approximation projects the signal over M vectors chosen a priori. In Fourier or wavelet bases, this linear approximation is particularly precise for uniformly regular signals. However, better approximations are obtained by choosing the M basis vectors depending on the signal. Signals with isolated singularities are well approximated in a wavelet basis with this non-linear procedure.

A further degree of freedom is introduced by choosing the basis adaptively, depending on the signal properties. From families of wavelet packet bases and local cosine bases, a fast dynamical programming algorithm is used to select the "best" basis that minimizes a Schur concave cost function. The approximation vectors chosen from this "best" basis outline the important signal structures, and characterize their time-frequency properties. Pursuit algorithms generalize these adaptive approximations by selecting the approximation vectors from redundant dictionaries of time-frequency atoms, with no orthogonality constraint.

9.1 LINEAR APPROXIMATIONS [1]

A signal can be represented with M parameters in an orthonormal basis by keeping M inner products with vectors chosen a priori. In Fourier and wavelet bases, Sections 9.1.2 and 9.1.3 show that such a linear approximation is efficient only if the signal is uniformly regular. Linear approximations of random vectors are studied and optimized in Section 9.1.4.

9.1.1 Linear Approximation Error

Let $\mathcal{B} = \{g_m\}_{m\in\mathbb{N}}$ be an orthonormal basis of a Hilbert space $\mathbf{H}$. Any $f \in \mathbf{H}$ can be decomposed in this basis:

$$f = \sum_{m=0}^{+\infty} \langle f, g_m \rangle \, g_m .$$

If instead of representing f by all inner products $\{\langle f, g_m\rangle\}_{m\in\mathbb{N}}$ we use only the first M, we get the approximation

$$f_M = \sum_{m=0}^{M-1} \langle f, g_m \rangle \, g_m .$$

This approximation is the orthogonal projection of f over the space $\mathbf{V}_M$ generated by $\{g_m\}_{0\le m<M}$. Since

$$f - f_M = \sum_{m=M}^{+\infty} \langle f, g_m \rangle \, g_m ,$$

the approximation error is

$$\epsilon_l[M] = \|f - f_M\|^2 = \sum_{m=M}^{+\infty} |\langle f, g_m \rangle|^2 . \tag{9.1}$$

The fact that $\|f\|^2 = \sum_{m=0}^{+\infty} |\langle f, g_m\rangle|^2 < +\infty$ implies that the error decays to zero:

$$\lim_{M\to+\infty} \epsilon_l[M] = 0.$$

However, the decay rate of $\epsilon_l[M]$ as M increases depends on the decay of $|\langle f, g_m\rangle|$ as m increases. The following theorem gives equivalent conditions on the decay of $\epsilon_l[M]$ and $|\langle f, g_m\rangle|$.

Theorem 9.1 *For any $s > 1/2$, there exists $A, B > 0$ such that if $\sum_{m=0}^{+\infty} |m|^{2s} |\langle f, g_m\rangle|^2 < +\infty$ then*

$$A \sum_{m=0}^{+\infty} m^{2s} |\langle f, g_m \rangle|^2 \le \sum_{M=0}^{+\infty} M^{2s-1} \epsilon_l[M] \le B \sum_{m=0}^{+\infty} m^{2s} |\langle f, g_m \rangle|^2 \tag{9.2}$$

and hence $\epsilon_l[M] = o(M^{-2s})$.

Proof[1]. By inserting (9.1), we compute

$$\sum_{M=0}^{+\infty} M^{2s-1}\epsilon_l[M] = \sum_{M=0}^{+\infty}\sum_{m=M}^{+\infty} M^{2s-1}|\langle f, g_m\rangle|^2 = \sum_{m=0}^{+\infty}|\langle f, g_m\rangle|^2 \sum_{M=0}^{m} M^{2s-1}.$$

For any $s > 1/2$

$$\int_0^m x^{2s-1}\,dx \leq \sum_{M=0}^{m} M^{2s-1} \leq \int_1^{m+1} x^{2s-1}\,dx$$

which implies that $\sum_{M=0}^{m} M^{2s-1} \sim m^{2s}$ and hence proves (9.2).

To verify that $\epsilon_l[M] = o(M^{-2s})$, observe that $\epsilon_l[m] \geq \epsilon_l[M]$ for $m \leq M$, so

$$\epsilon_l[M]\sum_{m=M/2}^{M-1} m^{2s-1} \leq \sum_{m=M/2}^{M-1} m^{2s-1}\epsilon_l[m] \leq \sum_{m=M/2}^{+\infty} m^{2s-1}\epsilon_l[m]. \tag{9.3}$$

Since $\sum_{m=1}^{+\infty} m^{2s-1}\epsilon_l[m] < +\infty$ it follows that

$$\lim_{M\to+\infty}\sum_{m=M/2}^{+\infty} m^{2s-1}\epsilon_l[m] = 0.$$

Moreover, there exists $C > 0$ such that $\sum_{m=M/2}^{M-1} m^{2s-1} \geq CM^{2s}$, so (9.3) implies that $\lim_{M\to+\infty}\epsilon_l[M]M^{2s} = 0$. ■

This theorem proves that the linear approximation error of f in the basis $\mathcal{B}$ decays faster than M^{-2s} if f belongs to the space

$$\mathbf{W}_{\mathcal{B},s} = \left\{ f \in \mathbf{H} \,:\, \sum_{m=0}^{+\infty} m^{2s}|\langle f, g_m\rangle|^2 < +\infty \right\}.$$

The next sections prove that if $\mathcal{B}$ is a Fourier or wavelet basis, then $\mathbf{W}_{\mathcal{B},s}$ is a Sobolev space. Observe that the linear approximation of f from the first M vectors of $\mathcal{B}$ is not always precise because these vectors are not necessarily the best ones with which to approximate f. Non-linear approximations calculated with vectors chosen adaptively depending upon f are studied in Section 9.2.

9.1.2 Linear Fourier Approximations

The Fourier basis can approximate uniformly regular signals with few low-frequency sinusoïdal waves. The approximation error is related to the Sobolev differentiability. It is also calculated for discontinuous signals having a bounded total variation.

Sobolev Differentiability The smoothness of f can be measured by the number of times it is differentiable. However, to distinguish the regularity of functions that are $n-1$ times, but not n times, continuously differentiable, we must extend the notion of differentiability to non-integers. This can be done in the Fourier domain.

Recall that the Fourier transform of the derivative $f'(t)$ is $i\omega\hat{f}(\omega)$. The Plancherel formula proves that $f' \in \mathbf{L}^2(\mathbb{R})$ if

$$\int_{-\infty}^{+\infty} |\omega|^2 |\hat{f}(\omega)|^2 d\omega = 2\pi \int_{-\infty}^{+\infty} |f'(t)|^2 dt < +\infty.$$

This suggests replacing the usual pointwise definition of the derivative by a definition based on the Fourier transform. We say that $f \in \mathbf{L}^2(\mathbb{R})$ is differentiable in the *sense of Sobolev* if

$$\int_{-\infty}^{+\infty} |\omega|^2 |\hat{f}(\omega)|^2 d\omega < +\infty. \tag{9.4}$$

This integral imposes that $|\hat{f}(\omega)|$ must have a sufficiently fast decay when the frequency ω goes to $+\infty$. As in Section 2.3.1, the regularity of f is measured from the asymptotic decay of its Fourier transform.

This definition is generalized for any $s > 0$. The space $\mathbf{W}^s(\mathbb{R})$ of Sobolev functions that are s times differentiable is the space of functions $f \in \mathbf{L}^2(\mathbb{R})$ whose Fourier transforms satisfy [72]

$$\int_{-\infty}^{+\infty} |\omega|^{2s} |\hat{f}(\omega)|^2 d\omega < +\infty. \tag{9.5}$$

If $s > n + 1/2$, then one can verify (Problem 9.2) that f is n times continuously differentiable. We define the space $\mathbf{W}^s[0,1]$ of functions on $[0,1]$ that are s times differentiable in the sense of Sobolev as the space of functions $f \in \mathbf{L}^2[0,1]$ that can be extended outside $[0,1]$ into a function $f \in \mathbf{W}^s(\mathbb{R})$.

Fourier Approximations Theorem 3.2 proves (modulo a change of variable) that $\{e^{i2\pi mt}\}_{m\in\mathbb{Z}}$ is an orthonormal basis of $\mathbf{L}^2[0,1]$. We can thus decompose $f \in \mathbf{L}^2[0,1]$ in the Fourier series

$$f(t) = \sum_{m=-\infty}^{+\infty} \langle f(u), e^{i2\pi mu} \rangle e^{i2\pi mt} \tag{9.6}$$

with

$$\langle f(u), e^{i2\pi mu} \rangle = \int_0^1 f(u) e^{-i2\pi mu} du.$$

The decomposition (9.6) defines a periodic extension of f for all $t \in \mathbb{R}$. The decay of the Fourier coefficients $|\langle f(u), e^{i2\pi mu} \rangle|$ as m increases depends on the regularity of this periodic extension. To avoid creating singularities at $t = 0$ or at $t = 1$ with this periodization, we suppose that the support of f is strictly included in $(0,1)$. One can then prove (not trivial) that if $f \in \mathbf{L}^2[0,1]$ is a function whose support is included in $(0,1)$, then $f \in \mathbf{W}^s[0,1]$ if and only if

$$\sum_{m=-\infty}^{+\infty} |m|^{2s} |\langle f(u), e^{i2\pi mu} \rangle|^2 < +\infty. \tag{9.7}$$

The linear approximation of $f \in \mathbf{L}^2[0,1]$ by the M sinusoïds of lower frequencies is

$$f_M(t) = \sum_{|m| \leq M/2} \langle f(u), e^{i2\pi mu} \rangle e^{i2\pi mt}.$$

For differentiable functions in the sense of Sobolev, the following proposition computes the approximation error

$$\epsilon_l[M] = \|f - f_M\|^2 = \int_0^1 |f(t) - f_M(t)|^2 \, dt = \sum_{|m| > M/2} |\langle f(u), e^{i2\pi mu} \rangle|^2 . \quad (9.8)$$

Proposition 9.1 *Let $f \in \mathbf{L}^2[0,1]$ be a function whose support is included in $(0,1)$. Then $f \in \mathbf{W}^s[0,1]$ if and only if*

$$\sum_{M=1}^{+\infty} M^{2s} \frac{\epsilon_l[M]}{M} < +\infty, \quad (9.9)$$

which implies $\epsilon_l[M] = o(M^{-2s})$.

Functions in $\mathbf{W}^s[0,1]$ with a support in $(0,1)$ are characterized by (9.7). This proposition is therefore a consequence of Theorem 9.1. The linear Fourier approximation thus decays quickly if and only if f has a large regularity exponent s in the sense of Sobolev.

Discontinuities and Bounded Variation If f is discontinuous, then $f \notin \mathbf{W}^s[0,1]$ for any $s > 1/2$. Proposition 9.1 thus proves that $\epsilon_l[M]$ can decay like $M^{-\alpha}$ only if $\alpha \leq 1$. For bounded variation functions, which are introduced in Section 2.3.3, the following proposition proves that $\epsilon_l[M] = O(M^{-1})$. A function has a bounded variation if

$$\|f\|_V = \int_0^1 |f'(t)| \, dt < +\infty .$$

The derivative must be taken in the sense of distributions because f may be discontinuous, as is the case for $f = \mathbf{1}_{[0,1/2]}$. Recall that $a[M] \sim b[M]$ if $a[M] = O(b[M])$ and $b[M] = O(a[M])$.

Proposition 9.2
- *If $\|f\|_V < +\infty$ then $\epsilon_l[M] = O(\|f\|_V^2 M^{-1})$.*
- *If $f = C\,\mathbf{1}_{[0,1/2]}$ then $\epsilon_l[M] \sim \|f\|_V^2 M^{-1}$.*

Proof[2]. If $\|f\|_V < +\infty$ then

$$\begin{aligned} |\langle f(u), \exp(i2m\pi u) \rangle| &= \left| \int_0^1 f(u) \exp(-i2m\pi u) \, du \right| \\ &= \left| \int_0^1 f'(u) \frac{\exp(-i2m\pi u)}{-i2m\pi} \, dt \right| \leq \frac{\|f\|_V}{2|m|\pi}. \end{aligned}$$

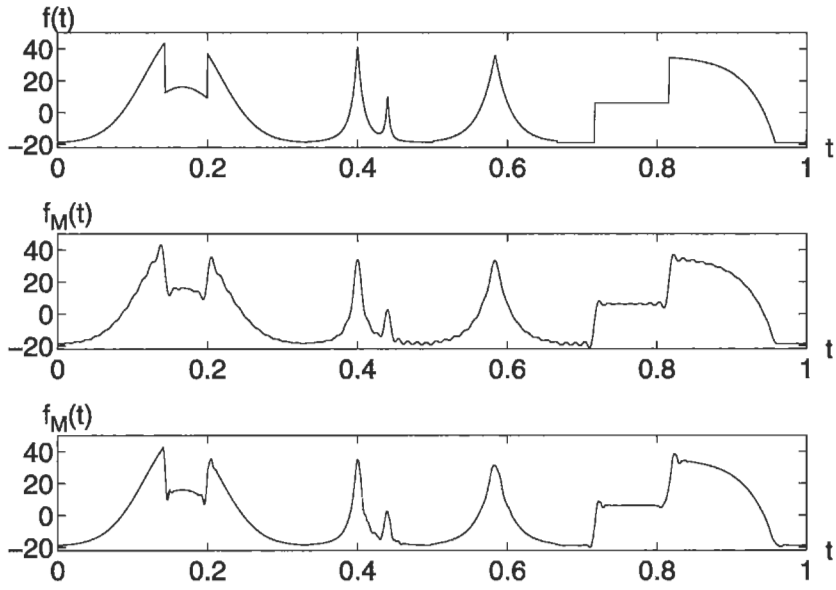


FIGURE 9.1 Top: Original signal f. Middle: Signal f_M approximated from lower frequency Fourier coefficients, with $M/N = 0.15$ and $\|f - f_M\|/\|f\| = 8.63\,10^{-2}$. Bottom: Signal f_M approximated from larger scale Daubechies 4 wavelet coefficients, with $M/N = 0.15$ and $\|f - f_M\|/\|f\| = 8.58\,10^{-2}$.

Hence

$$\epsilon_l[M] = \sum_{|m|>M/2} |\langle f(u), \exp(i2m\pi u)\rangle|^2 \leq \frac{\|f\|_V^2}{4\pi^2} \sum_{|m|>M/2} \frac{1}{m^2} = O(\|f\|_V^2 M^{-1}).$$

If $f = C\,\mathbf{1}_{[0,1/2]}$ then $\|f\|_V = 2C$ and

$$|\langle f(u), \exp(i2m\pi u)\rangle| = \begin{cases} 0 & \text{if } m \neq 0 \text{ is even} \\ C/(\pi |m|) & \text{if } m \text{ is odd,} \end{cases}$$

so $\epsilon_l[M] \sim C^2 M^{-1}$. ■

This proposition shows that when f is discontinuous with bounded variations, then $\epsilon_l[M]$ decays typically like M^{-1}. Figure 9.1(b) shows a bounded variation signal approximated by Fourier coefficients of lower frequencies. The approximation error is concentrated in the neighborhood of discontinuities where the removal of high frequencies creates Gibbs oscillations (see Section 2.3.1).

Localized Approximations To localize Fourier series approximations over intervals, we multiply f by smooth windows that cover each of these intervals. The Balian-Low Theorem 5.6 proves that one cannot build local Fourier bases with

smooth windows of compact support. However, Section 8.4.2 constructs orthonormal bases by replacing complex exponentials by cosine functions. For appropriate windows g_p of compact support $[a_p - \eta_p, a_{p+1} + \eta_{p+1}]$, Corollary 8.1 constructs an orthonormal basis of $\mathbf{L}^2(\mathbb{R})$:

$$\left\{ g_{p,k}(t) = g_p(t) \sqrt{\frac{2}{l_p}} \cos\left[\pi \left(k + \frac{1}{2}\right) \frac{t - a_p}{l_p} \right] \right\}_{k \in \mathbb{N}, p \in \mathbb{Z}}.$$

Writing f in this local cosine basis is equivalent to segmenting it into several windowed components $f_p(t) = f(t)\, g_p(t)$, which are decomposed in a cosine IV basis. If g_p is $\mathbf{C}^\infty$, the regularity of $g_p(t) f(t)$ is the same as the regularity of f over $[a_p - \eta_p, a_{p+1} + \eta_{p+1}]$. Section 8.3.2 relates cosine IV coefficients to Fourier series coefficients. It follows from Proposition 9.1 that if $f_p \in \mathbf{W}^s(\mathbb{R})$, then the approximation

$$f_{p,M} = \sum_{k=0}^{M-1} \langle f, g_{p,k} \rangle\, g_{p,k}$$

yields an error

$$\epsilon_{p,l}[M] = \| f_p - f_{p,M} \|^2 = o(M^{-2s}).$$

The approximation error in a local cosine basis thus depends on the local regularity of f over each window support.

9.1.3 Linear Multiresolution Approximations

Linear approximations of f from large scale wavelet coefficients are equivalent to finite element approximations over uniform grids. The approximation error depends on the uniform regularity of f. In a periodic orthogonal wavelet basis, this approximation behaves like a Fourier series approximation. In both cases, it is necessary to impose that f have a support inside $(0, 1)$ to avoid border discontinuities created by the periodization. This result is improved by the adapted wavelet basis of Section 7.5.3, whose border wavelets keep their vanishing moments. These wavelet bases efficiently approximate any function that is uniformly regular over $[0, 1]$, as well as the restriction to $[0, 1]$ of regular functions having a larger support.

Uniform Approximation Grid Section 7.5 explains how to design wavelet orthonormal bases of $\mathbf{L}^2[0, 1]$, with a maximum scale $2^J < 1$:

$$\left[\{\phi_{J,n}\}_{0 \le n < 2^{-J}}, \{\psi_{j,n}\}_{-\infty < j \le J, 0 \le n < 2^{-j}} \right]. \tag{9.10}$$

We suppose that the wavelets $\psi_{j,n}$ are in $\mathbf{C}^q$ and have q vanishing moments. The $M = 2^{-l}$ scaling functions and wavelets at scales $2^j > 2^l$ define an orthonormal basis of the approximation space $\mathbf{V}_l$:

$$\left[\{\phi_{J,n}\}_{0 \le n < 2^{-J}}, \{\psi_{j,n}\}_{l < j \le J, 0 \le n < 2^{-j}} \right]. \tag{9.11}$$

The approximation of f over the M first wavelets and scaling functions is an orthogonal projection on $\mathbf{V}_l$:

$$f_M = P_{\mathbf{V}_l} f = \sum_{j=l+1}^{J} \sum_{n=0}^{2^{-j}-1} \langle f, \psi_{j,n} \rangle \psi_{j,n} + \sum_{n=0}^{2^{-J}-1} \langle f, \phi_{J,n} \rangle \phi_{J,n}. \tag{9.12}$$

Since $\mathbf{V}_l$ also admits an orthonormal basis of $M = 2^{-l}$ scaling functions $\{\phi_{l,n}\}_{0 \leq n < 2^{-l}}$, this projection can be rewritten:

$$f_M = P_{\mathbf{V}_l} f = \sum_{n=0}^{2^{-l}-1} \langle f, \phi_{l,n} \rangle \phi_{l,n}. \tag{9.13}$$

This summation is an approximation of f with 2^{-l} finite elements $\phi_{l,n}(t) = \phi_l(t - 2^l n)$ translated over a uniform grid. The approximation error is the energy of wavelet coefficients at scales finer than 2^l:

$$\epsilon_l[M] = \|f - f_M\|^2 = \sum_{j=-\infty}^{l} \sum_{n=0}^{2^{-j}-1} |\langle f, \psi_{j,n} \rangle|^2. \tag{9.14}$$

If $2^{-l} < M < 2^{-l+1}$, one must include in the approximations (9.12) and (9.13) the coefficients of the $M - 2^{-l}$ wavelets $\{\psi_{l-1,n}\}_{0 \leq n < M - 2^{-l}}$ at the scale 2^{l-1}.

Approximation error Like a Fourier basis, a wavelet basis provides an efficient approximation of functions that are s times differentiable in the sense of Sobolev over $[0,1]$ (i.e., functions of $\mathbf{W}^s[0,1]$). If ψ has q vanishing moments then (6.11) proves that the wavelet transform is a multiscale differential operator of order q. To test the differentiability of f up to order s we thus need $q > s$. The following theorem gives a necessary and sufficient condition on the wavelet coefficients so that $f \in \mathbf{W}^s[0,1]$.

Theorem 9.2 *Let $0 < s < q$ be a Sobolev exponent. A function $f \in \mathbf{L}^2[0,1]$ is in $\mathbf{W}^s[0,1]$ if and only if*

$$\sum_{j=-\infty}^{J} \sum_{n=0}^{2^{-j}-1} 2^{-2sj} |\langle f, \psi_{j,n} \rangle|^2 < +\infty. \tag{9.15}$$

Proof[2]. We give an intuitive justification but not a proof of this result. To simplify, we suppose that the support of f is included in $(0,1)$. If we extend f by zeros outside $[0,1]$ then $f \in \mathbf{W}^s(\mathbb{R})$, which means that

$$\int_{-\infty}^{+\infty} |\omega|^{2s} |\hat{f}(\omega)|^2 \, d\omega < +\infty. \tag{9.16}$$

The low frequency part of this integral always remains finite because $f \in \mathbf{L}^2(\mathbb{R})$:

$$\int_{|\omega| \leq 2^{-J}\pi} |\omega|^{2s} |\hat{f}(\omega)|^2 \, d\omega \leq 2^{-2sJ} \pi^{2s} \int_{|\omega| \leq \pi} |\hat{f}(\omega)|^2 \, d\omega \leq 2^{-2sJ} \pi^{2s} \|f\|^2.$$

The energy of $\hat{\psi}_{j,n}$ is essentially concentrated in the intervals $[-2^{-j}2\pi, -2^{-j}\pi] \cup [2^{-j}\pi, 2^{-j}2\pi]$. As a consequence

$$\sum_{n=0}^{2^{-j}-1} |\langle f, \psi_{j,n}\rangle|^2 \sim \int_{2^{-j}\pi \leq |\omega| \leq 2^{-j+1}\pi} |\hat{f}(\omega)|^2 \, d\omega.$$

Over this interval $|\omega| \sim 2^{-j}$, so

$$\sum_{n=0}^{2^{-j}-1} 2^{-2sj} |\langle f, \psi_{j,n}\rangle|^2 \sim \int_{2^{-j}\pi \leq |\omega| \leq 2^{-j+1}\pi} |\omega|^{2s} |\hat{f}(\omega)|^2 \, d\omega.$$

It follows that

$$\sum_{j=-\infty}^{J} \sum_{n=0}^{2^{-j}-1} 2^{-2sj} |\langle f, \psi_{j,n}\rangle|^2 \sim \int_{|\omega| \geq 2^{-J}\pi} |\omega|^{2s} |\hat{f}(\omega)|^2 \, d\omega,$$

which explains why (9.16) is equivalent to (9.15). ■

This theorem proves that the Sobolev regularity of f is equivalent to a fast decay of the wavelet coefficients $|\langle f, \psi_{j,n}\rangle|$ when the scale 2^j decreases. If ψ has q vanishing moments but is not q times continuously differentiable, then $f \in \mathbf{W}^s[0,1]$ implies (9.15), but the opposite implication is not true. The following proposition uses the decay condition (9.15) to compute the approximation error with M wavelets.

Proposition 9.3 *Let $0 < s < q$ be a Sobolev exponent. A function $f \in \mathbf{L}^2[0,1]$ is in $\mathbf{W}^s[0,1]$ if and only if*

$$\sum_{M=1}^{+\infty} M^{2s} \frac{\epsilon_l[M]}{M} < +\infty, \tag{9.17}$$

which implies $\epsilon_l[M] = o(M^{-2s})$.

Proof[2]. Let us write the wavelets $\psi_{j,n} = g_m$ with $m = 2^{-j} + n$. One can verify that the Sobolev condition (9.15) is equivalent to

$$\sum_{m=0}^{+\infty} |m|^{2s} |\langle f, g_m\rangle|^2 < +\infty.$$

The proof ends by applying Theorem 9.1. ■

Proposition 9.3 proves that $f \in \mathbf{W}^s[0,1]$ if and only if the approximation error $\epsilon_l[M]$ decays slightly faster than M^{-2s}. The wavelet approximation error is of the same order as the Fourier approximation error calculated in (9.9). If the wavelet has q vanishing moments but is not q times continuously differentiable, then $f \in \mathbf{W}^s[0,1]$ implies (9.17) but the opposite implication is false.

If f has a discontinuity in $(0,1)$ then $f \notin \mathbf{W}^s[0,1]$ for $s > 1/2$ so Proposition 9.3 proves that we cannot have $\epsilon_l[M] = O(M^{-\alpha})$ for $\alpha > 1$. If f has a bounded total variation norm $\|f\|_V$ then one can verify (Problem 9.4) that $\epsilon_l[M] = O(\|f\|_V^2 M^{-1})$.

For example, if $f = C\mathbf{1}_{[0,1/2]}$ then $\epsilon_l[M] \sim \|f\|_V^2 M^{-1}$. This result is identical to Proposition 9.2, obtained in a Fourier basis.

Figure 9.1 gives an example of discontinuous signal with bounded variation, which is approximated by its larger scale wavelet coefficients. The largest amplitude errors are in the neighborhood of singularities, where the scale should be refined. The relative approximation error $\|f - f_M\|/\|f\| = 8.56\,10^{-2}$ is almost the same as in a Fourier basis.

9.1.4 Karhunen-Loève Approximations [2]

Let us consider a whole class of signals that we approximate with the first M vectors of a basis. These signals are modeled as realizations of a random vector $F[n]$ of size N. We show that the basis that minimizes the average linear approximation error is the Karhunen-Loève basis (principal components).

Appendix A.6 reviews the covariance properties of random vectors. If $F[n]$ does not have a zero mean, we subtract the expected value $\mathrm{E}\{F[n]\}$ from $F[n]$ to get a zero mean. The random vector F can be decomposed in an orthogonal basis $\{g_m\}_{0 \le m < N}$:

$$F = \sum_{m=0}^{N-1} \langle F, g_m \rangle \, g_m \, .$$

Each coefficient

$$\langle F, g_m \rangle = \sum_{n=0}^{N-1} F[n] \, g_m^*[n]$$

is a random variable (see Appendix A.6). The approximation from the first M vectors of the basis is

$$F_M = \sum_{m=0}^{M-1} \langle F, g_m \rangle \, g_m .$$

The resulting mean-square error is

$$\epsilon_l[M] = \mathrm{E}\left\{ \|F - F_M\|^2 \right\} = \sum_{m=M}^{N-1} \mathrm{E}\left\{ |\langle F, g_m \rangle|^2 \right\}.$$

This error is related to the covariance of F defined by

$$R[n,m] = \mathrm{E}\{F[n] \, F^*[m]\}.$$

Let K be the covariance operator represented by this matrix. For any vector $x[n]$,

$$\begin{aligned} \mathrm{E}\left\{ |\langle F, x \rangle|^2 \right\} &= \mathrm{E}\left\{ \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} F[n] \, F[m] \, x[n] \, x^*[m] \right\} \\ &= \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} R[n,m] \, x[n] \, x^*[m] \\ &= \langle Kx, x \rangle . \end{aligned}$$

The error $\epsilon_l[M]$ is therefore a sum of the last $N-M$ diagonal coefficients of the covariance operator

$$\epsilon_l[M] = \sum_{m=M}^{N-1} \langle Kg_m, g_m \rangle.$$

The covariance operator K is Hermitian and positive and is thus diagonalized in an orthogonal basis called a Karhunen-Loève basis. This basis is not unique if several eigenvalues are equal. The following theorem proves that a Karhunen-Loève basis is optimal for linear approximations.

Theorem 9.3 *Let K be a covariance operator. For all $M \geq 1$, the approximation error*

$$\epsilon_l[M] = \sum_{m=M}^{N-1} \langle Kg_m, g_m \rangle$$

is minimum if and only if $\{g_m\}_{0 \leq m < N}$ is a Karhunen-Loève basis whose vectors are ordered by decreasing eigenvalues

$$\langle Kg_m, g_m \rangle \geq \langle Kg_{m+1}, g_{m+1} \rangle \quad \text{for } 0 \leq m < N-1.$$

Proof[3]. Let us consider an arbitrary orthonormal basis $\{h_m\}_{0 \leq m < N}$. The trace $\mathrm{tr}(K)$ of K is independent of the basis:

$$\mathrm{tr}(K) = \sum_{m=0}^{N-1} \langle Kh_m, h_m \rangle.$$

The basis that minimizes $\sum_{m=M}^{N-1} \langle Kh_m, h_m \rangle$ thus maximizes $\sum_{m=0}^{M-1} \langle Kh_m, h_m \rangle$.

Let $\{g_m\}_{0 \leq m < N}$ be a basis that diagonalizes K:

$$Kg_m = \sigma_m^2 g_m \quad \text{with } \sigma_m^2 \geq \sigma_{m+1}^2 \text{ for } 0 \leq m < N-1.$$

The theorem is proved by verifying that for all $M \geq 0$,

$$\sum_{m=0}^{M-1} \langle Kh_m, h_m \rangle \leq \sum_{m=0}^{M-1} \langle Kg_m, g_m \rangle = \sum_{m=0}^{M-1} \sigma_m^2.$$

To relate $\langle Kh_m, h_m \rangle$ to the eigenvalues $\{\sigma_i^2\}_{0 \leq i < N}$, we expand h_m in the basis $\{g_i\}_{0 \leq i < N}$:

$$\langle Kh_m, h_m \rangle = \sum_{i=0}^{N-1} |\langle h_m, g_i \rangle|^2 \sigma_i^2. \tag{9.18}$$

Hence

$$\sum_{m=0}^{M-1} \langle Kh_m, h_m \rangle = \sum_{m=0}^{M-1} \sum_{i=0}^{N-1} |\langle h_m, g_i \rangle|^2 \sigma_i^2 = \sum_{i=0}^{N-1} q_i \sigma_i^2$$

with

$$0 \leq q_i = \sum_{m=0}^{M-1} |\langle h_m, g_i \rangle|^2 \leq 1 \quad \text{and} \quad \sum_{i=0}^{N-1} q_i = M.$$

We evaluate

$$\begin{aligned}\sum_{m=0}^{M-1}\langle Kh_m,h_m\rangle &- \sum_{i=0}^{M-1}\sigma_i^2=\sum_{i=0}^{N-1}q_i\sigma_i^2-\sum_{i=0}^{M-1}\sigma_i^2\\ &= \sum_{i=0}^{N-1}q_i\sigma_i^2-\sum_{i=0}^{M-1}\sigma_i^2+\sigma_{M-1}^2\left(M-\sum_{i=0}^{N-1}q_i\right)\\ &= \sum_{i=0}^{M-1}(\sigma_i^2-\sigma_{M-1}^2)\,(q_i-1)+\sum_{i=M}^{N-1}q_i\,(\sigma_i^2-\sigma_{M-1}^2).\end{aligned}$$

Since the eigenvalues are listed in order of decreasing amplitude, it follows that

$$\sum_{m=0}^{M-1}\langle Kh_m,h_m\rangle-\sum_{m=0}^{M-1}\sigma_m^2\ \le\ 0.$$

Suppose that this last inequality is an equality. We finish the proof by showing that $\{h_m\}_{0\le m<N}$ must be a Karhunen-Loève basis. If $i<M$, then $\sigma_i^2\neq\sigma_{M-1}^2$ implies $q_i=1$. If $i\ge M$, then $\sigma_i^2\neq\sigma_{M-1}^2$ implies $q_i=0$. This is valid for all $M\ge 0$ if $\langle h_m,g_i\rangle\neq 0$ only when $\sigma_i^2=\sigma_m^2$. This means that the change of basis is performed inside each eigenspace of K so $\{h_m\}_{0\le m<N}$ also diagonalizes K. ■

Theorem 9.3 proves that a Karhunen-Loève basis yields the smallest average error when approximating a class of signals by their projection on M orthogonal vectors, chosen a priori. This result has a simple geometrical interpretation. The realizations of F define a cloud of points in $\mathbb{C}^N$. The density of this cloud specifies the probability distribution of F. The vectors g_m of the Karhunen-Loève basis give the directions of the principal axes of the cloud. Large eigenvalues σ_m^2 correspond to directions g_m along which the cloud is highly elongated. Theorem 9.3 proves that projecting the realizations of F on these principal components yields the smallest average error. If F is a Gaussian random vector, the probability density is uniform along ellipsoids whose axes are proportional to σ_m in the direction of g_m. These principal directions are thus truly the preferred directions of the process.

Random Shift Processes If the process is not Gaussian, its probability distribution can have a complex geometry, and a linear approximation along the principal axes may not be efficient. As an example, we consider a random vector $F[n]$ of size N that is a random shift modulo N of a deterministic signal $f[n]$ of zero mean, $\sum_{n=0}^{N-1} f[n]=0$:

$$F[n]=f[(n-P)\bmod N]. \tag{9.19}$$

The shift P is an integer random variable whose probability distribution is uniform on $[0,N-1]$:

$$\Pr(P=p)=\frac{1}{N}\ \text{ for } 0\le p<N.$$

This process has a zero mean:

$$E\{F[n]\}=\frac{1}{N}\sum_{p=0}^{N-1}f[(n-p)\bmod N]=0,$$

and its covariance is

$$R[n,k] = \mathsf{E}\{F[n]F[k]\} = \frac{1}{N}\sum_{p=0}^{N-1} f[(n-p) \bmod N]\, f[(k-p) \bmod N]$$

$$= \frac{1}{N} f \circledast \bar{f}[n-k] \quad \text{with} \quad \bar{f}[n] = f[-n]\,. \tag{9.20}$$

Hence $R[n,k] = R_F[n-k]$ with

$$R_F[k] = \frac{1}{N} f \circledast \bar{f}[k].$$

Since R_F is N periodic, F is a circular stationary random vector, as defined in Appendix A.6. The covariance operator K is a circular convolution with R_F and is therefore diagonalized in the discrete Fourier Karhunen-Loève basis $\{\frac{1}{\sqrt{N}} \exp\left(\frac{i2\pi mn}{N}\right)\}_{0 \le m < N}$. The eigenvalues are given by the Fourier transform of R_F:

$$\sigma_m^2 = \hat{R}_F[m] = \frac{1}{N}|\hat{f}[m]|^2. \tag{9.21}$$

Theorem 9.3 proves that a linear approximation that projects F on M vectors selected a priori is optimized in this Fourier basis. To better understand this result, let us consider an extreme case where $f[n] = \delta[n] - \delta[n-1]$. Theorem 9.3 guarantees that the Fourier Karhunen-Loève basis produces a smaller expected approximation error than does a canonical basis of Diracs $\{g_m[n] = \delta[n-m]\}_{0 \le m < N}$. Indeed, we do not know a priori the abscissa of the non-zero coefficients of F, so there is no particular Dirac that is better adapted to perform the approximation. Since the Fourier vectors cover the whole support of F, they always absorb part of the signal energy:

$$\mathsf{E}\left\{\left|\left\langle F[n], \frac{1}{\sqrt{N}} \exp\left(\frac{i2\pi mn}{N}\right)\right\rangle\right|^2\right\} = \hat{R}_F[m] = \frac{4}{N}\sin^2\left(\frac{\pi k}{N}\right).$$

Selecting M higher frequency Fourier coefficients thus yields a better mean-square approximation than choosing a priori M Dirac vectors to perform the approximation.

The linear approximation of F in a Fourier basis is not efficient because all the eigenvalues $\hat{R}_F[m]$ have the same order of magnitude. A simple non-linear algorithm can improve this approximation. In a Dirac basis, F is exactly reproduced by selecting the two Diracs corresponding to the largest amplitude coefficients, whose positions P and $P-1$ depend on each realization of F. A non-linear algorithm that selects the largest amplitude coefficient for each realization of F is not efficient in a Fourier basis. Indeed, the realizations of F do not have their energy concentrated over a few large amplitude Fourier coefficients. This example shows that when F is not a Gaussian process, a non-linear approximation may be much more precise than a linear approximation, and the Karhunen-Loève basis is no longer optimal.

9.2 NON-LINEAR APPROXIMATIONS [1]

Linear approximations project the signal on M vectors selected a priori. This approximation is improved by choosing the M vectors depending on each signal. The next section analyzes the performance of these non-linear approximations. These results are then applied to wavelet bases.

9.2.1 Non-Linear Approximation Error

A signal $f \in \mathbf{H}$ is approximated with M vectors selected adaptively in an orthonormal basis $\mathcal{B} = \{g_m\}_{m\in\mathbb{N}}$ of $\mathbf{H}$. Let f_M be the projection of f over M vectors whose indices are in I_M :

$$f_M = \sum_{m\in I_M} \langle f, g_m\rangle\, g_m.$$

The approximation error is the sum of the remaining coefficients:

$$\epsilon[M] = \|f - f_M\|^2 = \sum_{m\notin I_M} |\langle f, g_m\rangle|^2. \tag{9.22}$$

To minimize this error, the indices in I_M must correspond to the M vectors having the largest inner product amplitude $|\langle f, g_m\rangle|$. These are the vectors that best correlate f. They can thus be interpreted as the "main" features of f. The resulting $\epsilon_n[M]$ is necessarily smaller than the error of a linear approximation (9.1), which selects the M approximation vectors independently of f.

Let us sort $\{|\langle f, g_m\rangle|\}_{m\in\mathbb{N}}$ in decreasing order. We denote $f^r_{\mathcal{B}}[k] = \langle f, g_{m_k}\rangle$ the coefficient of rank k:

$$|f^r_{\mathcal{B}}[k]| \geq |f^r_{\mathcal{B}}[k+1]| \quad \text{with } k > 0.$$

The best non-linear approximation is

$$f_M = \sum_{k=1}^{M} f^r_{\mathcal{B}}[k]\, g_{m_k}. \tag{9.23}$$

It can also be calculated by applying the thresholding function

$$\theta_T(x) = \begin{cases} x & \text{if } |x| \geq T \\ 0 & \text{if } |x| < T \end{cases} \tag{9.24}$$

with a threshold T such that $f^r_{\mathcal{B}}[M+1] < T \leq f^r_{\mathcal{B}}[M]$:

$$f_M = \sum_{m=0}^{+\infty} \theta_T(\langle f, g_m\rangle)\, g_m\,. \tag{9.25}$$

The minimum non-linear approximation error is

$$\epsilon_n[M] = \|f - f_M\|^2 = \sum_{k=M+1}^{+\infty} |f_B^r[k]|^2.$$

The following theorem relates the decay of this approximation error as M increases to the decay of $|f_B^r[k]|$ as k increases.

Theorem 9.4 *Let $s > 1/2$. If there exists $C > 0$ such that $|f_B^r[k]| \leq C k^{-s}$ then*

$$\epsilon_n[M] \leq \frac{C^2}{2s-1} M^{1-2s}. \tag{9.26}$$

Conversely, if $\epsilon_n[M]$ satisfies (9.26) then

$$|f_B^r[k]| \leq \left(1 - \frac{1}{2s}\right)^{-s} C k^{-s}. \tag{9.27}$$

Proof[2]. Since

$$\epsilon_n[M] = \sum_{k=M+1}^{+\infty} |f_B^r[k]|^2 \leq C^2 \sum_{k=M+1}^{+\infty} k^{-2s},$$

and

$$\sum_{k=M+1}^{+\infty} k^{-2s} \leq \int_M^{+\infty} x^{-2s}\, dx = \frac{M^{1-2s}}{2s-1} \tag{9.28}$$

we derive (9.26).

Conversely, let $\alpha < 1$,

$$\epsilon_n[\alpha M] \geq \sum_{k=\alpha M+1}^{M} |f_B^r[k]|^2 \geq (1-\alpha) M |f_B^r[M]|^2.$$

So if (9.26) is satisfied

$$|f_B^r[M]|^2 \leq \frac{\epsilon_n[\alpha M]}{1-\alpha} M^{-1} \leq \frac{C^2}{2s-1} \frac{\alpha^{1-2s}}{1-\alpha} M^{-2s}.$$

For $\alpha = 1 - 1/2s$ we get (9.27) for $k = M$. ■

The decay of sorted inner products can be evaluated from the $\mathbf{l}^p$ norm of these inner products:

$$\|f\|_{B,p} = \left(\sum_{m=0}^{+\infty} |\langle f, g_m \rangle|^p \right)^{1/p}.$$

The following theorem relates the decay of $\epsilon_n[M]$ to $\|f\|_{B,p}$.

Theorem 9.5 *Let $p < 2$. If $\|f\|_{B,p} < +\infty$ then*

$$|f_B^r[k]| \leq \|f\|_{B,p}\, k^{-1/p} \tag{9.29}$$

and $\epsilon_n[M] = o(M^{1-2/p})$.

Proof[2]. We prove (9.29) by observing that

$$\|f\|_{\mathcal{B},p}^p = \sum_{n=1}^{+\infty} |f_{\mathcal{B}}^r[n]|^p \geq \sum_{n=1}^{k} |f_{\mathcal{B}}^r[n]|^p \geq k\,|f_{\mathcal{B}}^r[k]|^p .$$

To show that $\epsilon_n[M] = o(M^{1-2/p})$, we set

$$S[k] = \sum_{n=k}^{2k-1} |f_{\mathcal{B}}^r[n]|^p \geq k\,|f_{\mathcal{B}}^r[2k]|^p .$$

Hence

$$\begin{aligned} \epsilon_n[M] &= \sum_{k=M+1}^{+\infty} |f_{\mathcal{B}}^r[k]|^2 \leq \sum_{k=M+1}^{+\infty} S[k/2]^{2/p}\,(k/2)^{-2/p} \\ &\leq \sup_{k>M/2} |S[k]|^{2/p} \sum_{k=M+1}^{+\infty} (k/2)^{-2/p} . \end{aligned}$$

Since $\|f\|_{\mathcal{B},p}^p = \sum_{n=1}^{+\infty} |f_{\mathcal{B}}^r[n]|^p < +\infty$, it follows that $\lim_{k\to+\infty} \sup_{k>M/2} |S[k]| = 0$. We thus derive from (9.28) that $\epsilon_n[M] = o(M^{1-2/p})$. ■

This theorem specifies spaces of functions that are well approximated by a few vectors of an orthogonal basis $\mathcal{B}$. We denote

$$\mathbf{B}_{\mathcal{B},p} = \left\{ f \in \mathbf{H} \;:\; \|f\|_{\mathcal{B},p} < +\infty \right\}. \tag{9.30}$$

If $f \in \mathbf{B}_{\mathcal{B},p}$ then Theorem 9.5 proves that $\epsilon_n[M] = o(M^{1-2/p})$. This is called a *Jackson inequality* [22]. Conversely, if $\epsilon_n[M] = O(M^{1-2/p})$ then the *Bernstein inequality* (9.27) for $s = 1/p$ shows that $f \in \mathbf{B}_{\mathcal{B},q}$ for any $q > p$. Section 9.2.3 studies the properties of the spaces $\mathbf{B}_{\mathcal{B},p}$ for wavelet bases.

9.2.2 Wavelet Adaptive Grids

A non-linear approximation in a wavelet orthonormal basis defines an adaptive grid that refines the approximation scale in the neighborhood of the signal singularities. Theorem 9.5 proves that this non-linear approximation introduces a small error if the sorted wavelet coefficients have a fast decay, which can be related to *Besov spaces* [157]. We study the performance of such wavelet approximations for bounded variation functions and piecewise regular functions.

We consider a wavelet basis adapted to $\mathbf{L}^2[0,1]$, constructed in Section 7.5.3 with compactly supported wavelets that are $\mathbf{C}^q$ with q vanishing moments:

$$\mathcal{B} = \left[\{\phi_{J,n}\}_{0\leq n<2^{-J}} , \{\psi_{j,n}\}_{-\infty<j\leq J, 0\leq n<2^{-j}} \right].$$

To simplify the notation we write $\phi_{J,n} = \psi_{J+1,n}$. The best non-linear approximation of $f \in \mathbf{L}^2[0,1]$ from M wavelets is

$$f_M = \sum_{(j,n)\in I_M} \langle f, \psi_{j,n} \rangle\, \psi_{j,n} ,$$

where I_M is the index set of the M wavelet coefficients having the largest amplitude $|\langle f, \psi_{j,n}\rangle|$. The approximation error is

$$\epsilon_n[M] = \|f - f_M\|^2 = \sum_{(j,n)\notin I_M} |\langle f, \psi_{j,n}\rangle|^2.$$

Let $f_B^r[k] = \langle f, \psi_{j_k,n_k}\rangle$ be the coefficient of rank k: $|f_B^r[k]| \geq |f_B^r[k+1]|$ for $k \geq 1$. Theorem 9.4 proves that $|f_B^r[k]| = O(k^{-s})$ if and only if $\epsilon_n[M] = O(M^{-2s})$. The error $\epsilon_n[M]$ is always smaller than the linear approximation error $\epsilon_l[M]$ studied in Section 9.1.3, but we must understand under which condition this improvement is important.

Piecewise Regularity If f is piecewise regular then we show that $\epsilon_n[M]$ has a fast decay as M increases. Few wavelet coefficients are affected by isolated discontinuities and the error decay depends on the uniform regularity between these discontinuities.

Proposition 9.4 *If f has a finite number of discontinuities on $[0,1]$ and is uniformly Lipschitz $\alpha < q$ between these discontinuities, then $\epsilon_n[M] = O(M^{-2\alpha})$.*

Proof[2]. We prove that $\epsilon_n[M] = O(M^{-2\alpha})$ by verifying that $f_B^r[k] = O(k^{-\alpha-1/2})$ and applying inequality (9.26) of Theorem 9.4. We distinguish type 1 wavelets $\psi_{j,n}$, whose support includes an abscissa where f is discontinuous, from type 2 wavelets, whose support is included in a domain where f is uniformly Lipschitz α. Let $f_{B,1}^r[k]$ and $f_{B,2}^r[k]$ be the values of the wavelet coefficient of rank k among type 1 and type 2 wavelets. We show that $f_B^r[k] = O(k^{-\alpha-1/2})$ by verifying that $f_{B,1}^r[k] = O(k^{-\alpha-1/2})$ and that $f_{B,2}^r[k] = O(k^{-\alpha-1/2})$.

If f is uniformly Lipschitz α on the support of $\psi_{j,n}$ then there exists A such that

$$|\langle f, \psi_{j,n}\rangle| \leq A\, 2^{j(\alpha+1/2)}. \tag{9.31}$$

Indeed, orthogonal wavelet coefficients are samples of the continuous wavelet transform $\langle f, \psi_{j,n}\rangle = Wf(2^j n, 2^j)$, so (9.31) is a consequence of (6.17).

For any $l > 0$, there are at most 2^l type 2 coefficients at scales $2^j > 2^{-l}$. Moreover, (9.31) shows that all type 2 coefficients at scales $2^j \leq 2^{-l}$ are smaller than $A\, 2^{l(\alpha+1/2)}$, so

$$f_{B,2}^r[2^l] \leq A 2^{-l(\alpha+1/2)}.$$

It follows that $f_{B,2}^r[k] = O(k^{-\alpha-1/2})$, for all $k > 0$.

Let us now consider the type 1 wavelets. There exists $K > 0$ such that each wavelet $\psi_{j,n}$ has its support included in $[2^j n - 2^j K/2, 2^j n + 2^j nK/2]$. At each scale 2^j, there are thus at most K wavelets whose support includes a given abscissa v. This implies that there are at most KD wavelets $\psi_{j,n}$ whose support includes at least one of the D discontinuities of f. Since f is uniformly Lipschitz $\alpha > 0$ outside these points, f is necessarily uniformly bounded on $[0,1]$ and thus uniformly Lipschitz 0. Hence (9.31) shows that there exists A such that $|\langle f, \psi_{j,n}\rangle| \leq A 2^{j/2}$. Since there are at most lKD type 1 coefficients at scales $2^j > 2^{-l}$ and since all type 1 coefficients at scales $2^j \leq 2^{-l}$ are smaller than $A 2^{-l/2}$ we get

$$f_{B,1}^r[lKD] \leq A 2^{-l/2}.$$

This implies that $f^r_{B,1}[k] = O(k^{-\beta-1/2})$ for any $\beta > 0$, which ends the proof. ∎

If $\alpha > 1/2$, then $\epsilon_n[M]$ decays faster than $\epsilon_l[M]$ since we saw in Section 9.1.3 that the presence of discontinuities implies that $\epsilon_l[M]$ decays like M^{-1}. The more regular f is between its discontinuities, the larger the improvement of non-linear approximations with respect to linear approximations.

Adaptive Grids Isolated singularities create large amplitude wavelet coefficients but there are few of them. The approximation f_M calculated from the M largest amplitude wavelet coefficients can be interpreted as an adaptive grid approximation, where the approximation scale is refined in the neighborhood of singularities.

A non-linear approximation keeps all coefficients $|\langle f, \psi_{j,n}\rangle| \geq T$, for a threshold $f^r_B[M] \geq T > f^r_B[M+1]$. In a region where f is uniformly Lipschitz α, since $|\langle f, \psi_{j,n}\rangle| \sim A\,2^{j(\alpha+1/2)}$ the coefficients above T are typically at scales

$$2^j > 2^l = \left(\frac{T}{A}\right)^{2/(2\alpha+1)}.$$

Setting to zero all wavelet coefficients below the scale 2^l is equivalent to computing a local approximation of f at the scale 2^l. The smaller the local Lipschitz regularity α, the finer the approximation scale 2^l.

Figure 9.2 shows the non-linear wavelet approximation of a piecewise regular signal. Observe that the largest amplitude wavelet coefficients are in the cone of influence of each singularity. Since the approximation scale is refined in the neighborhood of each singularity, they are much better restored than in the fixed scale linear approximation shown in Figure 9.1. The non-linear approximation error in this case is 17 times smaller than the linear approximation error.

Non-linear wavelet approximations are nearly optimal compared to adaptive spline approximations. A spline approximation f^s_M is calculated by choosing K nodes $t_1 < t_2 < \cdots < t_K$ inside $[0, 1]$. Over each interval $[t_k, t_{k+1}]$, f is approximated by the closest polynomial of degree r. This polynomial spline f^s_M is specified by $M = K(r+2)$ parameters, which are the node locations $\{t_k\}_{1\leq k\leq K}$ plus the $K(r+1)$ parameters of the K polynomials of degree r. To reduce $\|f - f^s_M\|$, the nodes must be closely spaced when f is irregular and farther apart when f is smooth. However, finding the M parameters that minimize $\|f - f^s_M\|$ is a difficult non-linear optimization.

A non-linear approximation with wavelets having $q = r + 1$ vanishing moments is much faster to compute than an optimized spline approximation. A spline wavelet basis of Battle-Lemarié gives non-linear approximations that are also splines functions, but the nodes t_k are restricted to dyadic locations $2^j n$, with a scale 2^j that is locally adapted to the signal regularity. For large classes of signals, including the balls of Besov spaces, the maximum approximation errors with wavelets or with optimized splines have the same decay rate when M increases [158]. The computational overhead of an optimized spline approximation is therefore not worth it.

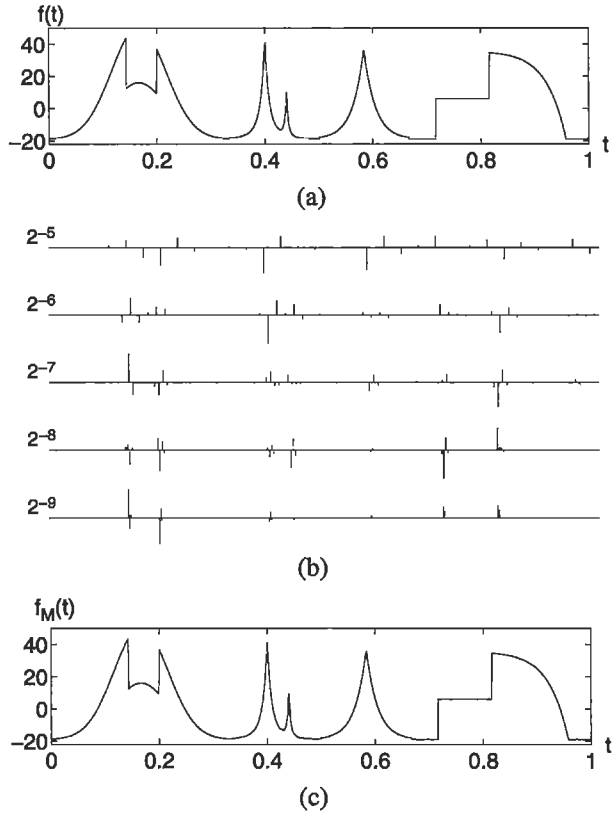


FIGURE 9.2 (a): Original signal f. (b): Larger $M = 0.15N$ wavelet coefficients calculated with a Symmlet 4. (c): Non-linear approximation f_M recovered from the M wavelet coefficients shown above, $\|f - f_M\|/\|f\| = 5.1\,10^{-3}$.

9.2.3 Besov Spaces [3]

Studying the performance of non-linear wavelet approximations more precisely requires introducing a new space. As previously, we write $\phi_{J,n} = \psi_{J+1,n}$. The *Besov* space $\mathbf{B}^s_{\beta,\gamma}[0,1]$ is the set of functions $f \in \mathbf{L}^2[0,1]$ such that

$$\|f\|_{s,\beta,\gamma} = \left(\sum_{j=-\infty}^{J+1} \left[2^{-j(s+1/2-1/\beta)} \left(\sum_{n=0}^{2^{-j}-1} |\langle f, \psi_{j,n} \rangle|^\beta \right)^{1/\beta} \right]^\gamma \right)^{1/\gamma} < +\infty \,. \tag{9.32}$$

Frazier, Jawerth [182] and Meyer [270] proved that $\mathbf{B}^s_{\beta,\gamma}[0,1]$ does not depend on the particular choice of wavelet basis, as long as the wavelets in the basis have $q > s$ vanishing moments and are in $\mathbf{C}^q$. The space $\mathbf{B}^s_{\beta,\gamma}[0,1]$ corresponds typically

to functions that have a "derivative of order s" that is in $\mathbf{L}^\beta[0,1]$. The index γ is a fine tuning parameter, which is less important. We need $q > s$ because a wavelet with q vanishing moments can test the differentiability of a signal only up to the order q.

If $\beta \geq 2$, then functions in $\mathbf{B}^s_{\beta,\gamma}[0,1]$ have a uniform regularity of order s. For $\beta = \gamma = 2$, Theorem 9.2 proves that $\mathbf{B}^s_{2,2}[0,1] = \mathbf{W}^s[0,1]$ is the space of s times differentiable functions in the sense of Sobolev. Proposition 9.3 proves that this space is characterized by the decay of the linear approximation error $\epsilon_l[M]$ and that $\epsilon_l[M] = o(M^{-2s})$. Since $\epsilon_n[M] \leq \epsilon_l[M]$ clearly $\epsilon_n[M] = o(M^{-s})$. One can verify (Problem 9.7) that for a large class of functions inside $\mathbf{W}^s[0,1]$, the non-linear approximation error has the same decay rate as the linear approximation error. It is therefore not useful to use non-linear approximations in a Sobolev space.

For $\beta < 2$, functions in $\mathbf{B}^s_{\beta,\gamma}[0,1]$ are not necessarily uniformly regular. The adaptativity of non-linear approximations then improves the decay rate of the error significantly. In particular, if $p = \beta = \gamma$ and $s = 1/2 + 1/p$, then the Besov norm is a simple $\mathbf{l}^p$ norm:

$$\|f\|_{s,\beta,\gamma} = \left(\sum_{j=-\infty}^{J+1} \sum_{n=0}^{2^{-j}-1} |\langle f, \psi_{j,n} \rangle|^p \right)^{1/p}.$$

Theorem 9.5 proves that if $f \in \mathbf{B}^s_{\beta,\gamma}[0,1]$, then $\epsilon_n[M] = o(M^{1-2/p})$. The smaller p, the faster the error decay. The proof of Proposition 9.4 shows that although f may be discontinuous, if the number of discontinuities is finite and f is uniformly Lipschitz α between these discontinuities, then its sorted wavelet coefficients satisfy $|f^r_B[k]| = O(k^{-\alpha-1/2})$, so $f \in \mathbf{B}^s_{\beta,\gamma}[0,1]$ for $1/p < \alpha + 1/2$. This shows that these spaces include functions that are not s times differentiable at all points. The linear approximation error $\epsilon_l[M]$ for $f \in \mathbf{B}^s_{\beta,\gamma}[0,1]$ can decrease arbitrarily slowly because the M wavelet coefficients at the largest scales may be arbitrarily small. A non-linear approximation is much more efficient in these spaces.

Bounded Variation Bounded variation functions are important examples of signals for which a non-linear approximation yields a much smaller error than a linear approximation. The total variation norm is defined in (2.57) by

$$\|f\|_V = \int_0^1 |f'(t)|\, dt .$$

The derivative f' must be understood in the sense of distributions, in order to include discontinuous functions. The following theorem computes an upper and a lower bound of $\|f\|_V$ from the modulus of wavelet coefficients. Since $\|f\|_V$ does not change when a constant is added to f, the scaling coefficients of f are controlled with the sup norm $\|f\|_\infty = \sup_{t \in \mathbf{R}} |f(t)|$.

Theorem 9.6 *Consider a wavelet basis constructed with ψ such that $\|\psi\|_V < +\infty$. There exist $A, B > 0$ such that for all $f \in \mathbf{L}^2[0,1]$*

$$\|f\|_V + \|f\|_\infty \leq B \sum_{j=-\infty}^{J+1} \sum_{n=0}^{2^{-j}-1} 2^{-j/2} |\langle f, \psi_{j,n} \rangle| = B \|f\|_{1,1,1} , \tag{9.33}$$

and

$$\|f\|_V + \|f\|_\infty \geq A \sup_{j \leq J+1} \left(\sum_{n=0}^{2^{-j}-1} 2^{-j/2} |\langle f, \psi_{j,n} \rangle| \right) = A \|f\|_{1,1,\infty} . \tag{9.34}$$

Proof[2]. By decomposing f in the wavelet basis

$$f = \sum_{j=-\infty}^{J} \sum_{n=0}^{2^{-j}-1} \langle f, \psi_{j,n} \rangle \psi_{j,n} + \sum_{n=0}^{2^{-J}-1} \langle f, \phi_{J,n} \rangle \phi_{J,n},$$

we get

$$\begin{aligned} \|f\|_V + \|f\|_\infty \quad \leq \quad & \sum_{j=-\infty}^{J} \sum_{n=0}^{2^{-j}-1} |\langle f, \psi_{j,n} \rangle| \left(\|\psi_{j,n}\|_V + \|\psi_{j,n}\|_\infty \right) \\ & + \sum_{n=0}^{2^{-J}-1} |\langle f, \phi_{J,n} \rangle| \left(\|\phi_{J,n}\|_V + \|\psi_{J,n}\|_\infty \right). \end{aligned} \tag{9.35}$$

The wavelet basis includes wavelets whose support are inside $(0,1)$ and border wavelets, which are obtained by dilating and translating a finite number of mother wavelets. To simplify notations we write the basis as if there were a single mother wavelet: $\psi_{j,n}(t) = 2^{-j/2} \psi(2^{-j}t - n)$. Hence, we verify with a change of variable that

$$\begin{aligned} \|\psi_{j,n}\|_V + \|\psi_{j,n}\|_\infty \quad &= \quad \int_0^1 2^{-j/2} 2^{-j} |\psi'(2^{-j}t - n)| \, dt \\ & \quad + 2^{-j/2} \sup_{t \in [0,1]} |\psi(2^{-j}t - n)| \\ &= \quad 2^{-j/2} \left(\|\psi\|_V + \|\psi\|_\infty \right) . \end{aligned}$$

Since $\phi_{J,n}(t) = 2^{-J/2} \phi(2^{-J}t - n)$ we also prove that

$$\|\phi_{J,n}\|_V + \|\phi_{J,n}\|_\infty = 2^{-J/2} \left(\|\phi\|_V + \|\phi\|_\infty \right) .$$

The inequality (9.33) is thus derived from (9.35).

Since ψ has at least one vanishing moment, its primitive θ is a function with the same support, which we suppose included in $[-K/2, K/2]$. To prove (9.34), for $j \leq J$

we make an integration by parts:

$$\begin{aligned}\sum_{n=0}^{2^{-j}-1} |\langle f, \psi_{j,n}\rangle| &= \sum_{n=0}^{2^{-j}-1} \left| \int_0^1 f(t)\, 2^{-j/2} \psi(2^{-j}t - n)\, dt \right| \\ &= \sum_{n=0}^{2^{-j}-1} \left| \int_0^1 f'(t)\, 2^{j/2} \theta(2^{-j}t - n)\, dt \right| \\ &\leq 2^{j/2} \sum_{n=0}^{2^{-j}-1} \int_0^1 |f'(t)|\, |\theta(2^{-j}t - n)|\, dt\ .\end{aligned}$$

Since θ has a support in $[-K/2, K/2]$,

$$\sum_{n=0}^{2^{-j}-1} |\langle f, \psi_{j,n}\rangle| \leq 2^{j/2} K \sup_{t\in\mathbb{R}} |\theta(t)| \int_0^1 |f'(t)|\, dt \leq A^{-1} 2^{j/2} \|f\|_V\ . \tag{9.36}$$

The largest scale 2^J is a fixed constant and hence

$$\begin{aligned}\sum_{n=0}^{2^{-J}-1} |\langle f, \phi_{J,n}\rangle| &\leq 2^{-3J/2} \sup_{t\in[0,1]} |f(t)| \int_0^1 |\phi_{J,n}(t)|dt \\ &\leq 2^{-J/2} \|f\|_\infty \int_0^1 |\phi(t)|dt \leq A^{-1} 2^{J/2} \|f\|_\infty\ .\end{aligned}$$

This inequality and (9.36) prove (9.34). ■

This theorem shows that the total variation norm is bounded by two Besov norms:

$$A\, \|f\|_{1,1,\infty} \leq \|f\|_V + \|f\|_\infty \leq B\, \|f\|_{1,1,1}\ .$$

One can verify that if $\|f\|_V < +\infty$, then $\|f\|_\infty < +\infty$ (Problem 9.1), but we do not control the value of $\|f\|_\infty$ from $\|f\|_V$ because the addition of a constant changes $\|f\|_\infty$ but does not modify $\|f\|_V$. The space $\mathbf{BV}[0,1]$ of bounded variation functions is therefore embedded in the corresponding Besov spaces:

$$\mathbf{B}^1_{1,1}[0,1] \subset \mathbf{BV}[0,1] \subset \mathbf{B}^1_{1,\infty}[0,1]\ .$$

If $f \in \mathbf{BV}[0,1]$ has discontinuities, then the linear approximation error $\epsilon_l[M]$ does not decay faster than M^{-1}. The following theorem proves that $\epsilon_n[M]$ has a faster decay.

Proposition 9.5 *There exists B such that for all $f \in \mathbf{BV}[0,1]$*

$$\epsilon_n[M] \leq B\, \|f\|_V^2\, M^{-2}\ . \tag{9.37}$$

Proof[2]. We denote by $f^r_B[k]$ the wavelet coefficient of rank k, excluding all the scaling coefficients $\langle f, \phi_{J,n}\rangle$, since we cannot control their value with $\|f\|_V$. We first show that there exists B_0 such that for all $f \in \mathbf{BV}[0,1]$

$$|f^r_B[k]| \leq B_0\, \|f\|_V\, k^{-3/2}. \tag{9.38}$$

To take into account the fact that (9.38) does not apply to the 2^J scaling coefficients $\langle f, \phi_{J,n} \rangle$, we observe that in the worst case they are selected by the non-linear approximation so

$$\epsilon_n[M] \leq \sum_{k=M-2^J+1}^{+\infty} |f_B^r[k]|^2 . \tag{9.39}$$

Since 2^J is a constant, inserting (9.38) proves (9.37).

The upper bound (9.38) is proved by computing an upper bound of the number of coefficients larger than an arbitrary threshold T. At scale 2^j, we denote by $f_B^r[j,k]$ the coefficient of rank k among $\{\langle f, \psi_{j,n} \rangle\}_{0 \leq n \leq 2^{-j}}$. The inequality (9.36) proves that for all $j \leq J$

$$\sum_{n=0}^{2^{-j}-1} |\langle f, \psi_{j,n} \rangle| \leq A^{-1} 2^{j/2} \|f\|_V .$$

It thus follows from (9.29) that

$$f_B^r[j,k] \leq A^{-1} 2^{j/2} \|f\|_V k^{-1} = C 2^{j/2} k^{-1} .$$

At scale 2^j, the number k_j of coefficients larger than T thus satisfies

$$k_j \leq \min(2^{-j}, 2^{j/2} C T^{-1}) .$$

The total number k of coefficients larger than T is

$$\begin{aligned} k &= \sum_{j=-\infty}^{J} k_j \leq \sum_{2^j \geq (C^{-1}T)^{2/3}} 2^{-j} + \sum_{2^j > (C^{-1}T)^{2/3}} 2^{j/2} C T^{-1} \\ &\leq 6 (C T^{-1})^{2/3} . \end{aligned}$$

By choosing $T = |f_B^r[k]|$, since $C = A^{-1} \|f\|_V$, we get

$$|f_B^r[k]| \leq 6^{3/2} A^{-1} \|f\|_V k^{-3/2} ,$$

which proves (9.38). ■

The error decay rate M^{-2} obtained with wavelets for all bounded variation functions cannot be improved either by optimal spline approximations or by any non-linear approximation calculated in an orthonormal basis [160]. In this sense, wavelets are optimal for approximating bounded variation functions.

9.3 IMAGE APPROXIMATIONS WITH WAVELETS [1]

Linear and non-linear approximations of functions in $\mathbf{L}^2[0,1]^d$ can be calculated in separable wavelet bases. In multiple dimensions, wavelet approximations are often not optimal because they cannot be adapted to the geometry of the signal singularities. We concentrate on the two-dimensional case for image processing.

Section 7.7.4 constructs a separable wavelet basis of $\mathbf{L}^2[0,1]^d$ from a wavelet basis of $\mathbf{L}^2[0,1]$, with separable products of wavelets and scaling functions. We suppose that all wavelets of the basis of $\mathbf{L}^2[0,1]$ are $\mathbf{C}^q$ with q vanishing moments. The wavelet basis of $\mathbf{L}^2[0,1]^2$ includes three elementary wavelets $\{\psi^l\}_{1 \leq l \leq 3}$ that

are dilated by 2^j and translated over a square grid of interval 2^j in $[0,1]^2$. Modulo modifications near the borders, these wavelets can be written

$$\psi^l_{j,n}(x) = \frac{1}{2^j}\psi^l\left(\frac{x_1 - 2^j n_1}{2^j}, \frac{x_2 - 2^j n_2}{2^j}\right). \tag{9.40}$$

If we limit the scales to $2^j \leq 2^J$, we must complete the wavelet family with two-dimensional scaling functions

$$\phi^2_{J,n}(x) = \frac{1}{2^J}\phi^2\left(\frac{x_1 - 2^J n_1}{2^J}, \frac{x_2 - 2^J n_2}{2^J}\right)$$

to obtain the orthonormal basis

$$\mathcal{B} = \left(\{\phi^2_{J,n}\}_{2^J n \in [0,1)^2} \cup \{\psi^l_{j,n}\}_{j \leq J, 2^j n \in [0,1)^2, 1 \leq l \leq 3}\right).$$

Bounded Variation Images Bounded variation functions provide good models for large classes of images, which do not have irregular textures. The total variation of f is defined in Section 2.3.3 by

$$\|f\|_V = \int_0^1 \int_0^1 |\vec{\nabla} f(x_1, x_2)|\, dx_1\, dx_2 . \tag{9.41}$$

The partial derivatives of $\vec{\nabla} f$ must be taken in the general sense of distributions in order to include discontinuous functions. Let $\partial\Omega_t$ be the level set defined as the boundary of

$$\Omega_t = \{(x_1, x_2) \in \mathbb{R}^2 \ : \ f(x_1, x_2) > t\} .$$

Theorem 2.7 proves that the total variation depends on the length $H^1(\partial\Omega_t)$ of level sets:

$$\int_0^1 \int_0^1 |\vec{\nabla} f(x_1, x_2)|\, dx_1\, dx_2 = \int_{-\infty}^{+\infty} H^1(\partial\Omega_t)\, dt. \tag{9.42}$$

The following theorem gives upper and lower bounds for $\|f\|_V$ from wavelet coefficients. We suppose that the separable wavelet basis has been calculated from a one-dimensional wavelet with bounded variation.

Theorem 9.7 *There exist $A, B > 0$ such that if $\|f\|_V < +\infty$ then*

$$A\|f\|_V \leq \sum_{j=-\infty}^{J} \sum_{l=1}^{3} \sum_{2^j n \in [0,1]^2} |\langle f, \psi^l_{j,n}\rangle| + \sum_{2^J n \in [0,1]^2} |\langle f, \phi^2_{J,n}\rangle| , \tag{9.43}$$

and

$$B\|f\|_V \geq \sup_{\substack{-\infty < j \leq J \\ 1 \leq l \leq 3}} \left(\sum_{2^j n \in [0,1]^2} |\langle f, \psi^l_{j,n}\rangle| \right) . \tag{9.44}$$

Proof[2]. The inequalities (9.43) and (9.44) are proved with the same proof as in Theorem 9.6 for one-dimensional bounded variation functions, given that $\|\psi^l_{j,n}\|_V = \|\psi^l\|_V$ and $\|\phi^2_{J,n}\|_V = \|\phi^2\|_V$. ■

Linear Approximations A linear approximation of $f \in \mathbf{L}^2[0,1]^2$ is computed by keeping only the $M = 2^{-2m}$ wavelet coefficients at scales $2^j > 2^m$. This recovers the projection of f in the multiresolution space $\mathbf{V}_m^2$. A function with finite total variation does not necessarily have a bounded amplitude, but images do have a bounded amplitude. The following theorem derives an upper bound for the linear approximation error $\epsilon_l[M]$.

Proposition 9.6 *There exists $B > 0$ such that if $\|f\|_V < +\infty$ and $\|f\|_\infty < +\infty$ then*

$$\epsilon_l[M] \leq B \|f\|_V \|f\|_\infty M^{-1/2} . \tag{9.45}$$

Proof[2]. The linear approximation error from $M = 2^{-2m}$ wavelets is

$$\epsilon_l[2^{-2m}] = \sum_{j=-\infty}^{m} \sum_{l=1}^{3} \sum_{2^j n \in [0,1]^2} |\langle f, \psi_{j,n}^l \rangle|^2 . \tag{9.46}$$

We shall verify that there exists $B_1 > 0$ such that for all j and l

$$\sum_{2^j n \in [0,1]^2} |\langle f, \psi_{j,n}^l \rangle|^2 \leq B_1 \|f\|_V \|f\|_\infty 2^j . \tag{9.47}$$

Applying this upper bound to the sum (9.46) proves that

$$\epsilon_l[2^{-2m}] \leq 6 B_1 \|f\|_V \|f\|_\infty 2^m ,$$

from which (9.45) is obtained for any $M > 1$.

The upper bound (9.47) is calculated with (9.44), which shows that there exists $B_2 > 0$ such that for all j and l

$$\sum_{2^j n \in [0,1]^2} |\langle f, \psi_{j,n}^l \rangle| \leq B_2 \|f\|_V . \tag{9.48}$$

The amplitude of a wavelet coefficient can also be bounded:

$$|\langle f, \psi_{j,n}^l \rangle| \leq \|f\|_\infty \|\psi_{j,n}^l\|_1 = \|f\|_\infty 2^j \|\psi^l\|_1 ,$$

where $\|\psi^l\|_1$ is the $\mathbf{L}^1[0,1]^2$ norm of ψ^l. If $B_3 = \max_{1 \leq l \leq 3} \|\psi^l\|_1$ this yields

$$|\langle f, \psi_{j,n}^l \rangle| \leq B_3 2^j \|f\|_\infty . \tag{9.49}$$

Since $\sum_n |a_n|^2 \leq \sup_n |a_n| \sum_n |a_n|$, we get (9.47) from (9.48) and (9.49) for $B_1 = B_2 B_3$. ■

If f has a finite total variation but is not bounded then the linear approximation error $\epsilon_l[M]$ may decay arbitrarily slowly (Problem 9.6). The indicator function $f = C \mathbf{1}_\Omega$ of a set Ω whose boundary $\partial\Omega$ has a finite length is an example of bounded variation function with a bounded amplitude. One can verify (Problem 9.5) that in this case $\epsilon_l[M] \sim \|f\|_V \|f\|_\infty M^{-1/2}$. In general, if f is discontinuous along contour of non-zero length, and if it is bounded with a bounded total variation, then $\epsilon_l[M]$ decays like $M^{-1/2}$. This is the case of many images such as Lena in Figure 9.3(a). Figure 9.3(b) is a linear approximation calculated with the $M = N^2/16$ largest scale wavelet coefficients. This approximation produces a uniform blur and creates Gibbs oscillations in the neighborhood of contours.

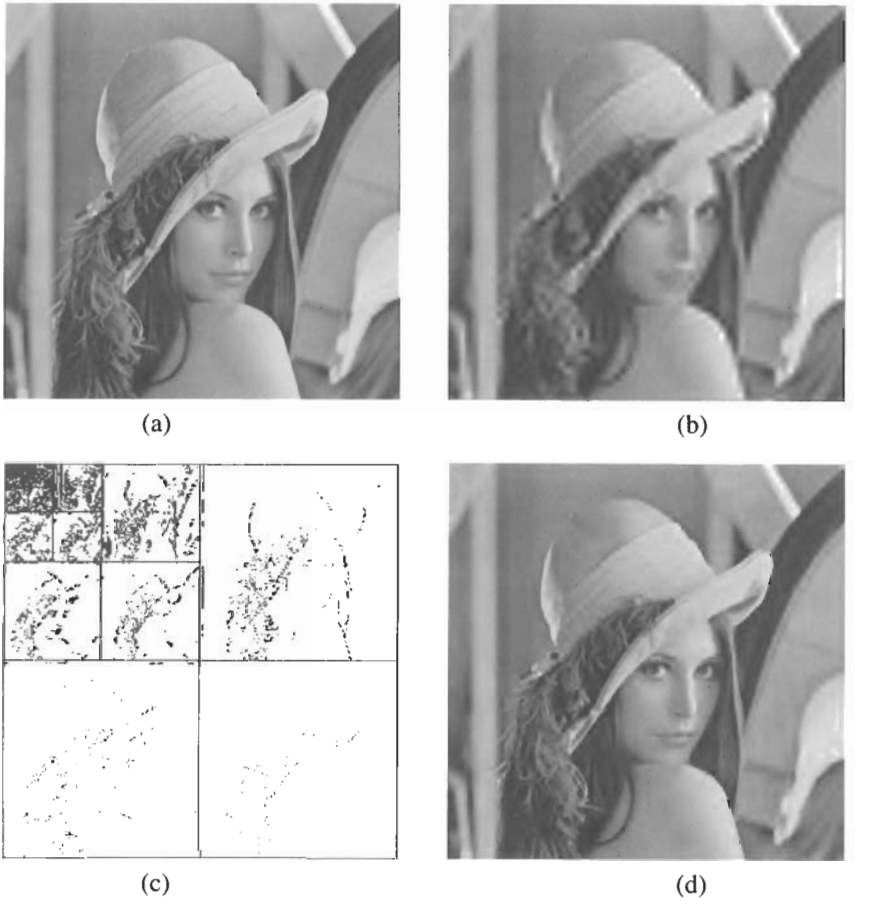


FIGURE 9.3 (a): Original Lena f of $N^2 = 256^2$ pixels. (b): Linear approximations f_M calculated from the $M = N^2/16$ Symmlet 4 wavelet coefficients at the largest scales: $\|f - f_M\|/\|f\| = 0.036$. (c): The positions of the $M = N^2/16$ largest amplitude wavelet coefficients are shown in black. (d): Non-linear approximation f_M calculated from the M largest amplitude wavelet coefficients: $\|f - f_M\|/\|f\| = 0.011$.

Non-linear Approximations A non-linear approximation f_M is constructed from the M wavelet coefficients of largest amplitude. Figure 9.3(c) shows the position of these $M = N^2/16$ wavelet coefficients for Lena. The large amplitude coefficients are located in the area where the image intensity varies sharply, in particular along the edges. The corresponding approximation f_M is shown in Figure 9.3(d). The non-linear approximation error is much smaller than the linear approximation error: $\epsilon_n[M] \leq \epsilon_l[M]/10$. As in one dimension, the non-linear wavelet approximation can be interpreted as an adaptive grid approximation. By keeping wavelet coefficients at fine scales we refine the approximation along the image contours.

We denote by $f_B^r[k]$ the rank k wavelet coefficient of f, without including the 2^{2J} scaling coefficients $\langle f, \phi_{J,n}^2 \rangle$. Indeed, scaling coefficients depend upon the average of f over $[0,1]^2$, which can be controlled from $\|f\|_V$. However, there are few scaling coefficients, which therefore play a negligible role in the non-linear approximation error. The following theorem proves that the non-linear approximation error $\epsilon_n[M]$ of a bounded variation image decays at least like M^{-1}, whereas $\epsilon_l[M]$ decays like $M^{-1/2}$ for discontinuous functions.

Theorem 9.8 (COHEN, DEVORE, PERTRUSHEV, XU) *There exist B_1, B_2 such that if $\|f\|_V < +\infty$ then*

$$|f_B^r[k]| \leq B_1 \|f\|_V \, k^{-1} , \tag{9.50}$$

and

$$\epsilon_n[M] \leq B_2 \|f\|_V^2 \, M^{-1} . \tag{9.51}$$

Proof[2]. The proof of (9.50) is quite technical and can be found in [133]. To take into account the exclusion of the 2^{2J} scaling coefficients $\langle f, \phi_{J,n}^2 \rangle$ in (9.50), we observe as in (9.39) that $\epsilon_n[M] \leq \sum_{k=M-2^{2J}+1}^{+\infty} |f_B^r[k]|^2$. Since 2^{2J} is a fixed number, we derive (9.51) from (9.50). ■

The inequality (9.51) proves that if $\|f\|_V < +\infty$ then $|f_B^r[k]| = O(k^{-1})$. Lena is a bounded variation image in the sense of (2.70), and Figure 9.4 shows that indeed $\log_2 |f_B^r[k]|$ decays with a slope that reaches -1 as $\log_2 k$ increases. In contrast, the Mandrill image shown in Figure 11.6 does not have a bounded total variation because of the fur texture, and indeed $\log_2 |f_B^r[k]|$ decays with a slope that reaches $-0.65 > -1$.

Piecewise Regular Images In one dimension, Proposition 9.4 proves that a finite number of discontinuities does not influence the decay rate of sorted wavelet coefficients $|f_B^r[k]|$, which depends on the uniform signal regularity outside the discontinuities. Piecewise regular functions are thus better approximated than functions for which we only know that they have a bounded variation. A piecewise regular image has discontinuities along curves of dimension 1, which create a non-negligible number of high amplitude wavelet coefficients. The following proposition verifies with a simple example of piecewise regular image, that the sorted wavelet coefficients $|f_B^r[k]|$ do not decay faster than k^{-1}. As in Theorem

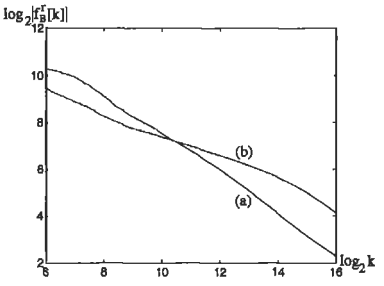


FIGURE 9.4 Sorted wavelet coefficients $\log_2 |f_B^r[k]|$ as a function of $\log_2 k$ for two images. (a): Lena image shown in Figure 9.3(a). (b): Mandrill image shown in Figure 11.6.

9.8, the 2^{2J} scaling coefficients $\langle f, \phi_{J,n}^2 \rangle$ are not included among the sorted wavelet coefficients.

Proposition 9.7 *If $f = C\,\mathbf{1}_\Omega$ is the indicator function of a set Ω whose border $\partial\Omega$ has a finite length, then*

$$|f_B^r[k]| \sim \|f\|_V\, k^{-1}\,, \tag{9.52}$$

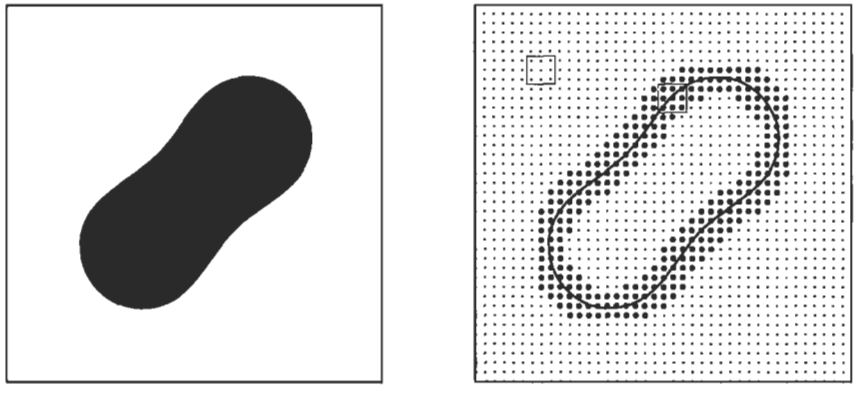
and hence

$$\epsilon_n[M] \sim \|f\|_V^2\, M^{-1}\,. \tag{9.53}$$

Proof[2]. The main idea of the proof is given without detail. If the support of $\psi_{j,n}^l$ does not intersect the border $\partial\Omega$, then $\langle f, \psi_{j,n}^l \rangle = 0$ because f is constant over the support of $\psi_{j,n}^l$. The wavelets $\psi_{j,n}^l$ have a square support of size proportional to 2^j, which is translated on a grid of interval 2^j. Since $\partial\Omega$ has a finite length L, there are on the order of $L\,2^{-j}$ wavelets whose support intersects $\partial\Omega$. Figure 9.5(b) illustrates the position of these coefficients.

Along the border, we verify like in (9.49) that $|\langle f, \psi_{j,n}^l \rangle| \sim C\,2^j$. Since the amplitude of these coefficients decreases as the scale 2^j decreases and since there are on the order of $L\,2^{-j}$ non-zero coefficients at scales larger than 2^j, the wavelet coefficient $f_B^r[k]$ of rank k is located at a scale 2^j such that $k \sim L\,2^{-j}$. Hence $|f_B^r[k]| \sim C\,2^j \sim C\,L\,k^{-1}$. The co-area (9.41) formula proves that $\|f\|_V = C\,L$, so $|f_B^r[k]| \sim \|f\|_V\, k^{-1}$, which proves (9.52). As in the proof of Theorem 9.8, (9.53) is derived from (9.52). ■

This proposition shows that the sorted wavelet coefficients of $f = C\,\mathbf{1}_\Omega$ do not decay any faster than the sorted wavelet coefficients of any bounded variation function, for which (9.50) proves that $|f_B^r[k]| = O(\|f\|_V\, k^{-1})$. This property can be extended to piecewise regular functions that have a discontinuity of amplitude larger than $C > 0$ along a contour of length $L > 0$. The non-linear approximation errors $\epsilon_n[M]$ of general bounded variation images and piecewise regular images have essentially the same decay.



(a) (b)

FIGURE 9.5 (a): Image $f = \mathbf{1}_\Omega$. (b): At the scale 2^j, the wavelets $\psi^l_{j,n}$ have a square support of width proportional to 2^j. This support is translated on a grid of interval 2^j, which is indicated by the smaller dots. The darker dots correspond to wavelets whose support intersects the frontier of Ω, for which $\langle f, \psi^l_{j,n} \rangle \neq 0$.

Approximation with Adaptive Geometry Supposing that an image has bounded variations is equivalent to imposing that its level set have a finite average length, but it does not impose geometrical regularity conditions on these level sets. The level sets and "edges" of many images such as Lena are often curves with a regular geometry, which is a prior information that the approximation scheme should be able to use.

In two dimensions, wavelets cannot use the regularity of level sets because they have a square support that is not adapted to the image geometry. More efficient non-linear approximations may be constructed using functions whose support has a shape that can be adapted to the regularity of the image contours. For example, one may construct piecewise linear approximations with adapted triangulations [178, 293].

A function $f \in \mathbf{L^2}[0,1]^2$ is approximated with a triangulation composed of M triangles by a function f_M that is linear on each triangle and which minimizes $\|f - f_M\|$. This is a two-dimensional extension of the spline approximations studied in Section 9.2.2. The difficulty is to optimize the geometry of the triangulation to reduce the error $\|f - f_M\|$. Let us consider the case where $f = \mathbf{1}_\Omega$, with a border $\partial\Omega$ which is a differentiable curve of finite length and bounded curvature. The triangles inside and outside Ω may have a large support since f is constant and therefore linear on these triangles. On the other hand, the triangles that intersect $\partial\Omega$ must be narrow in order to minimize the approximation error in the direction where f is discontinuous. One can use $M/2$ triangles for the inside and $M/2$ for the outside of Ω. Since $\partial\Omega$ has a finite length, this border can be covered by $M/2$ triangles which have a length on the order of M^{-1} in the direction of the tangent

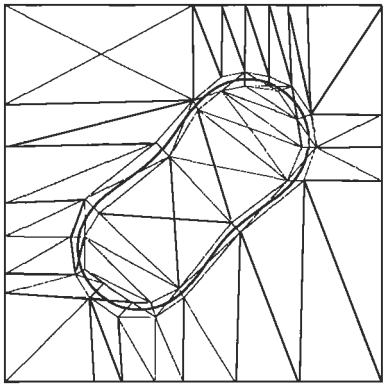


FIGURE 9.6 A piecewise linear approximation of $f = 1_\Omega$ is optimized with a triangulation whose triangles are narrow in the direction where f is discontinuous, along the border $\partial\Omega$.

$\vec{\tau}$ of $\partial\Omega$. Since the curvature of $\partial\Omega$ is bounded, one can verify that the width of these triangles can be on the order of M^{-2} in the direction perpendicular to $\vec{\tau}$. The border triangles are thus very narrow, as illustrated by Figure 9.6. One can now easily show that there exists a function f_M that is linear on each triangle of this triangulation and such that $\|f - f_M\|^2 \sim M^{-2}$. This error thus decays more rapidly than the non-linear wavelet approximation error $\epsilon_n[M] \sim M^{-1}$. The adaptive triangulation yields a smaller error because it follows the geometrical regularity of the image contours.

Donoho studies the optimal approximation of particular classes of indicator functions with elongated wavelets called *wedglets* [165]. However, at present there exists no algorithm for computing quasi-optimal approximations adapted to the geometry of complex images such as Lena. Solving this problem would improve image compression and denoising algorithms.

9.4 ADAPTIVE BASIS SELECTION [2]

To optimize non-linear signal approximations, one can adaptively choose the basis depending on the signal. Section 9.4.1 explains how to select a "best" basis from a dictionary of bases, by minimizing a concave cost function. Wavelet packet and local cosine bases are large families of orthogonal bases that include different types of time-frequency atoms. A best wavelet packet basis or a best local cosine basis decomposes the signal over atoms that are adapted to the signal time-frequency structures. Section 9.4.2 introduces a fast best basis selection algorithm. The performance of a best basis approximation is evaluated in Section 9.4.3 through particular examples.

9.4.1 Best Basis and Schur Concavity

We consider a dictionary $\mathcal{D}$ that is a union of orthonormal bases in a signal space of finite dimension N:

$$\mathcal{D} = \bigcup_{\lambda \in \Lambda} \mathcal{B}^\lambda.$$

Each orthonormal basis is a family of N vectors

$$\mathcal{B}^\lambda = \{g_m^\lambda\}_{1 \leq m \leq N}.$$

Wavelet packets and local cosine trees are examples of dictionaries where the bases share some common vectors.

Comparison of Bases We want to optimize the non-linear approximation of f by choosing a best basis in $\mathcal{D}$. Let I_M^λ be the index set of the M vectors of $\mathcal{B}^\lambda$ that maximize $|\langle f, g_m^\lambda \rangle|$. The best non-linear approximation of f in $\mathcal{B}^\lambda$ is

$$f_M^\lambda = \sum_{m \in I_M^\lambda} \langle f, g_m^\lambda \rangle \, g_m^\lambda.$$

The approximation error is

$$\epsilon^\lambda[M] = \sum_{m \notin I_M^\lambda} |\langle f, g_m^\lambda \rangle|^2 = \|f\|^2 - \sum_{m \in I_M^\lambda} |\langle f, g_m^\lambda \rangle|^2. \tag{9.54}$$

Definition 9.1 *We say that $\mathcal{B}^\alpha = \{g_m^\alpha\}_{1 \leq m \leq N}$ is a better basis than $\mathcal{B}^\gamma = \{g_m^\gamma\}_{1 \leq m \leq N}$ for approximating f if for all $M \geq 1$*

$$\epsilon^\alpha[M] \leq \epsilon^\gamma[M]. \tag{9.55}$$

This basis comparison is a partial order relation between bases in $\mathcal{D}$. Neither $\mathcal{B}^\alpha$ nor $\mathcal{B}^\gamma$ is better if there exist M_0 and M_1 such that

$$\epsilon^\alpha[M_0] < \epsilon^\gamma[M_0] \text{ and } \epsilon^\alpha[M_1] > \epsilon^\gamma[M_1]. \tag{9.56}$$

Inserting (9.54) proves that the better basis condition (9.55) is equivalent to:

$$\forall M \geq 1 \ , \ \sum_{m \in I_M^\alpha} |\langle f, g_m^\alpha \rangle|^2 \geq \sum_{m \in I_M^\gamma} |\langle f, g_m^\gamma \rangle|^2. \tag{9.57}$$

The following theorem derives a criteria based on Schur concave cost functions.

Theorem 9.9 *A basis $\mathcal{B}^\alpha$ is a better basis than $\mathcal{B}^\gamma$ to approximate f if and only if for all concave functions $\Phi(u)$*

$$\sum_{m=1}^{N} \Phi\left(\frac{|\langle f, g_m^\alpha \rangle|^2}{\|f\|^2}\right) \leq \sum_{m=1}^{N} \Phi\left(\frac{|\langle f, g_m^\gamma \rangle|^2}{\|f\|^2}\right). \tag{9.58}$$

Proof[3]. The proof of this theorem is based on the following classical result in the theory of majorization [45].

Lemma 9.1 (HARDY, LITTLEWOOD, PÓLYA) *Let $x[m] \geq 0$ and $y[m] \geq 0$ be two positive sequences of size N, with*

$$x[m] \geq x[m+1] \text{ and } y[m] \geq y[m+1] \text{ for } 1 \leq m \leq N\,, \tag{9.59}$$

and $\sum_{m=1}^{N} x[m] = \sum_{m=1}^{N} y[m]$. For all $M \leq N$ these sequences satisfy

$$\sum_{m=1}^{M} x[m] \geq \sum_{m=1}^{M} y[m] \tag{9.60}$$

if and only if for all concave functions $\Phi(u)$

$$\sum_{m=1}^{N} \Phi(x[m]) \leq \sum_{m=1}^{N} \Phi(y[m]). \tag{9.61}$$

We first prove that (9.60) implies (9.61). Let Φ be a concave function. We denote by $\mathbf{H}$ the set of vectors z of dimension N such that

$$z[1] \geq \cdots \geq z[N].$$

For any $z \in \mathbf{H}$, we write the partial sum

$$S_z[M] = \sum_{m=1}^{M} z[m].$$

We denote by Θ the multivariable function

$$\begin{aligned} \Theta(S_z[1], S_z[2], \ldots, S_z[N]) &= \sum_{m=1}^{N} \Phi(z[m]) \\ &= \Phi(S_z[1]) + \sum_{m=2}^{N} \Phi(S_z[m] - S_z[m-1]) \end{aligned}$$

The sorting hypothesis (9.59) implies that $x \in \mathbf{H}$ and $y \in \mathbf{H}$, and we know that they have the same sum $S_x[N] = S_y[N]$. Condition (9.60) can be rewritten $S_x[M] \geq S_y[M]$ for $1 \leq M < N$. To prove (9.61) is thus equivalent to showing that Θ is a decreasing function with respect to each of its arguments $S_z[k]$ as long as z remains in $\mathbf{H}$. In other words, we must prove that for any $1 \leq k \leq N$

$$\Theta(S_z[1], S_z[2], \ldots, S_z[N]) \geq \Theta(S_z[1], \ldots, S_z[k-1], S_z[k] + \eta, S_z[k+1], \ldots, S_z[N]),$$

which means that

$$\sum_{m=1}^{N} \Phi(z[m]) \geq \sum_{m=1}^{k-1} \Phi(z[m]) + \Phi(z[k] + \eta) + \Phi(z[k+1] - \eta) + \sum_{m=k+2}^{N} \Phi(z[m]). \tag{9.62}$$

To guarantee that we remain in $\mathbf{H}$ despite the addition of η, its value must satisfy

$$z[k-1] \geq z[k] + \eta \geq z[k+1] - \eta \geq z[k+2].$$

The inequality (9.62) amounts to proving that

$$\Phi(z[k])+\Phi(z[k+1]) \geq \Phi(z[k]+\eta)+\Phi(z[k+1]-\eta). \tag{9.63}$$

Let us show that this is a consequence of the concavity of Φ.

By definition, Φ is concave if for any (x, y) and $0 \leq \alpha \leq 1$

$$\Phi(\alpha x+(1-\alpha) y) \geq \alpha \Phi(x)+(1-\alpha) \Phi(y). \tag{9.64}$$

Let us decompose

$$z[k]=\alpha(z[k]+\eta)+(1-\alpha)(z[k+1]-\eta)$$

and

$$z[k+1]=(1-\alpha)(z[k]+\eta)+\alpha(z[k+1]-\eta)$$

with

$$0 \leq \alpha=\frac{z[k]-z[k+1]+\eta}{z[k]-z[k+1]+2\eta} \leq 1.$$

Computing $\Phi(z[k])+\Phi(z[k+1])$ and applying the concavity (9.64) yields (9.63). This finishes the proof of (9.61).

We now verify that (9.60) is true if (9.61) is valid for a particular family of concave thresholding functions defined by

$$\Phi_M(u)=\begin{cases} x[M]-u & \text{if } u \geq x[M] \\ 0 & \text{otherwise} \end{cases}.$$

Let us evaluate

$$\sum_{m=1}^{N} \Phi_M(x[m])=M x[M]-\sum_{m=1}^{M} x[m].$$

The hypothesis (9.61) implies that $\sum_{m=1}^{N} \Phi_M(x[m]) \leq \sum_{m=1}^{N} \Phi_M(y[m])$. Moreover $\Phi(u) \leq 0$ and $\Phi(u) \leq x[M]-u$ so

$$M x[M]-\sum_{m=1}^{M} x[m] \leq \sum_{m=1}^{N} \Phi_M(y[m]) \leq \sum_{m=1}^{M} \Phi_M(y[m]) \leq M x[M]-\sum_{m=1}^{M} y[m],$$

which proves (9.60) and thus Lemma 9.1.

The statement of the theorem is a direct consequence of Lemma 9.1. For any basis $\mathcal{B}^{\lambda}$, we sort the inner products $|\langle f, g_m^{\lambda}\rangle|$ and denote

$$x^{\lambda}[k]=\frac{|\langle f, g_{m_k}^{\lambda}\rangle|^2}{\|f\|^2} \geq x^{\lambda}[k+1]=\frac{|\langle f, g_{m_{k+1}}^{\lambda}\rangle|^2}{\|f\|^2}.$$

The energy conservation in an orthogonal basis implies $\sum_{k=1}^{N} x^{\lambda}[k]=1$. Condition (9.57) proves that a basis $\mathcal{B}^{\alpha}$ is better than a basis $\mathcal{B}^{\gamma}$ if and only if for all $M \geq 1$

$$\sum_{k=1}^{M} x^{\alpha}[k] \geq \sum_{k=1}^{M} x^{\gamma}[k].$$

Lemma 9.1 proves that this is equivalent to imposing that for all concave functions Φ,

$$\sum_{k=1}^{N} \Phi(x^{\alpha}[k]) \leq \sum_{k=1}^{N} \Phi(x^{\gamma}[k]),$$

which is identical to (9.58). ■

In practice, two bases are compared using a single concave function $\Phi(u)$. The cost of approximating f in a basis $\mathcal{B}^\lambda$ is defined by the *Schur concave* sum

$$C(f,\mathcal{B}^\lambda) = \sum_{m=1}^{N} \Phi\left(\frac{|\langle f, g_m^\lambda\rangle|^2}{\|f\|^2}\right).$$

Theorem 9.9 proves that if $\mathcal{B}^\alpha$ is a better basis than $\mathcal{B}^\gamma$ for approximating f then

$$C(f,\mathcal{B}^\alpha) \leq C(f,\mathcal{B}^\beta). \tag{9.65}$$

This condition is necessary but not sufficient to guarantee that $\mathcal{B}^\alpha$ is better than $\mathcal{B}^\gamma$ since we test a single concave function. Coifman and Wickerhauser [140] find a *best basis* $\mathcal{B}^\alpha$ in $\mathcal{D}$ by minimizing the cost of f:

$$C(f,\mathcal{B}^\alpha) = \min_{\lambda\in\Lambda} C(f,\mathcal{B}^\lambda).$$

There exists no better basis in $\mathcal{D}$ to approximate f. However, there are often other bases in $\mathcal{D}$ that are equivalent in the sense of (9.56). In this case, the choice of the best basis depends on the particular concave function Φ.

Ideal and Diffusing Bases An ideal basis $\mathcal{B}$ for approximating f has one of its vectors proportional to f, say $g_m = \eta f$ with $\eta \in \mathbb{C}$. Clearly f can then be recovered with a single basis vector. If $\Phi(0) = 0$ then the cost of f in this basis is $C(f,\mathcal{B}) = \Phi(1)$. In contrast, a worst basis for approximating f is a basis $\mathcal{B}$ that diffuses uniformly the energy of f across all vectors:

$$|\langle f, g_m\rangle|^2 = \frac{\|f\|^2}{N} \quad \text{for } 0 \leq m < N\ .$$

The cost of f in a diffusing basis is $C(f,\mathcal{B}) = N\Phi(N^{-1})$.

Proposition 9.8 *Any basis $\mathcal{B}$ is worse than an ideal basis and better than a diffusing basis for approximating f. If $\Phi(0) = 0$ then*

$$\Phi(1) \leq C(f,\mathcal{B}) \leq N\,\Phi\left(\frac{1}{N}\right). \tag{9.66}$$

Proof[2]. An ideal basis is clearly better than any other basis in the sense of Definition 9.1, since it produces a zero error for $M \geq 1$. The approximation error from M vectors in a diffusing basis is $\|f\|^2(N-M)/N$. To prove that any basis $\mathcal{B}$ is better than a diffusing basis, observe that if m is not in the index set I_M corresponding to the M largest inner products then

$$|\langle f, g_m^\lambda\rangle|^2 \leq \frac{1}{M}\sum_{n\in I_M} |\langle f, g_n^\lambda\rangle|^2 \leq \frac{\|f\|^2}{M}. \tag{9.67}$$

The approximation error from M vectors thus satisfies

$$\epsilon[M] = \sum_{m \notin I_M} |\langle f, g_m^\lambda \rangle|^2 \leq \|f\|^2 \frac{N-M}{M},$$

which proves that it is smaller than the approximation error in a diffusing basis. The costs of ideal and diffusing bases are respectively $\Phi(1)$ and $N\Phi(N^{-1})$. We thus derive (9.66) from (9.65). ■

Examples of Cost Functions As mentioned earlier, if there exists no basis that is better than all other bases in $\mathcal{D}$, the "best" basis that minimizes $C(f, \mathcal{B}^\lambda)$ depends on the choice of Φ.

Entropy The entropy $\Phi(x) = -x \log_e x$ is concave for $x \geq 0$. The corresponding cost is called the entropy of the energy distribution

$$C(f, \mathcal{B}) = -\sum_{m=1}^{N} \frac{|\langle f, g_m \rangle|^2}{\|f\|^2} \log_e \left(\frac{|\langle f, g_m \rangle|^2}{\|f\|^2} \right). \tag{9.68}$$

Proposition 9.8 proves that

$$0 \leq C(f, \mathcal{B}) \leq \log_e N. \tag{9.69}$$

It reaches the upper bound $\log_e N$ for a diffusing basis.

Let us emphasize that this entropy is a priori not related to the number of bits required to encode the inner products $\langle f, g_m \rangle$. The Shannon Theorem 11.1 proves that a lower bound for the number of bits to encode individually each $\langle f, g_m \rangle$ is the entropy of the *probability distribution* of the values taken by $\langle f, g_m \rangle$. This probability distribution might be very different from the distribution of the normalized energies $|\langle f, g_m \rangle|^2 / \|f\|^2$. For example, if $\langle f, g_m \rangle = A$ for $0 \leq m < N$ then $|\langle f, g_m \rangle|^2 / \|f\|^2 = N^{-1}$ and the cost $C(f, \mathcal{B}) = \log_e N$ is maximum. In contrast, the probability distribution of the inner product is a discrete Dirac located at A and its entropy is therefore minimum and equal to 0.

l^p Cost For $p < 2$, $\Phi(x) = x^{p/2}$ is concave for $x \geq 0$. The resulting cost is

$$C(f, \mathcal{B}) = \sum_{m=1}^{N} \frac{|\langle f, g_m \rangle|^p}{\|f\|^p}. \tag{9.70}$$

Proposition 9.8 proves that it is always bounded by

$$1 \leq C(f, \mathcal{B}) \leq N^{1-p/2}.$$

This cost measures the l^p norm of the coefficients of f in $\mathcal{B}$:

$$C^{1/p}(f, \mathcal{B}) = \frac{\|f\|_{\mathcal{B},p}}{\|f\|}.$$

We derive from (9.26) that the approximation error $\epsilon[M]$ is bounded by

$$\epsilon[M] \leq \frac{\|f\|^2 C^{2/p}(f,\mathcal{B})}{2/p-1} \frac{1}{M^{2/p-1}}.$$

The minimization of this $\mathbf{l^p}$ cost can thus also be interpreted as a reduction of the decay factor C such that

$$\epsilon[M] \leq \frac{C}{M^{2/p-1}}.$$

9.4.2 Fast Best Basis Search in Trees

A best wavelet packet or local cosine basis divides the time-frequency plane into elementary atoms that are best adapted to approximate a particular signal. The construction of dictionaries of wavelet packet and local cosine bases is explained in Sections 8.1 and 8.4. For signals of size N, these dictionaries include more than $2^{N/2}$ bases. The best basis associated to f minimizes the cost

$$C(f,\mathcal{B}^\lambda) = \sum_{m=0}^{N-1} \Phi\left(\frac{|\langle f, g_m^\lambda\rangle|^2}{\|f\|^2}\right). \tag{9.71}$$

Finding this minimum by a brute force comparison of the cost of all wavelet packet or local cosine bases would require more than $N2^{N/2}$ operations, which is computationally prohibitive. The fast dynamic programming algorithm of Coifman and Wickerhauser [140] finds the best basis with $O(N\log_2 N)$ operations, by taking advantage of the tree structure of these dictionaries.

Dynamic Programming In wavelet packet and local cosine binary trees, each node corresponds to a space $\mathbf{W}_j^p$, which admits an orthonormal basis $\mathcal{B}_j^p$ of wavelet packets or local cosines. This space is divided in two orthogonal subspaces located at the children nodes:

$$\mathbf{W}_j^p = \mathbf{W}_{j+1}^{2p} \oplus \mathbf{W}_{j+1}^{2p+1}.$$

In addition to $\mathcal{B}_j^p$ we can thus construct an orthogonal basis of $\mathbf{W}_j^p$ with a union of orthogonal bases of $\mathbf{W}_{j+1}^{2p}$ and $\mathbf{W}_{j+1}^{2p+1}$. The root of the tree corresponds to a space of dimension N, which is $\mathbf{W}_0^0$ for local cosine bases and $\mathbf{W}_L^0$ with $2^L = N^{-1}$ for wavelet packet bases.

The cost of f in a family of $M \leq N$ orthonormal vectors $\mathcal{B} = \{g_m\}_{0\leq m<M}$ is defined by the partial sum

$$C(f,\mathcal{B}) = \sum_{m=0}^{M-1} \Phi\left(\frac{|\langle f, g_m\rangle|^2}{\|f\|^2}\right). \tag{9.72}$$

This cost is *additive* in the sense that for any orthonormal bases $\mathcal{B}^0$ and $\mathcal{B}^1$ of two orthogonal spaces

$$C(f,\mathcal{B}^0 \cup \mathcal{B}^1) = C(f,\mathcal{B}^0) + C(f,\mathcal{B}^1). \tag{9.73}$$

The best basis $\mathcal{O}_j^p$ of $\mathbf{W}_j^p$ is the basis that minimizes the cost (9.72), among all the bases of $\mathbf{W}_j^p$ that can be constructed from the vectors in the tree. The following proposition gives a recursive construction of best bases, from bottom up along the tree branches.

Proposition 9.9 (COIFMAN, WICKERHAUSER) *If C is an additive cost function then*

$$\mathcal{O}_j^p = \begin{cases} \mathcal{O}_{j+1}^{2p} \cup \mathcal{O}_{j+1}^{2p+1} & \text{if } C(f,\mathcal{O}_{j+1}^{2p}) + C(f,\mathcal{O}_{j+1}^{2p+1}) < C(f,\mathcal{B}_j^p) \\ \mathcal{B}_j^p & \text{if } C(f,\mathcal{O}_{j+1}^{2p}) + C(f,\mathcal{O}_{j+1}^{2p+1}) \geq C(f,\mathcal{B}_j^p) \end{cases} \tag{9.74}$$

Proof[2]. The best basis $\mathcal{O}_j^p$ is either equal to $\mathcal{B}_p^j$ or to the union $\mathcal{B}^0 \cup \mathcal{B}^1$ of two bases of $\mathbf{W}_{j+1}^{2p}$ and $\mathbf{W}_{j+1}^{2p+1}$. In this second case, the additivity property (9.73) implies that the cost of f in $\mathcal{O}_j^p$ is minimum if $\mathcal{B}^0$ and $\mathcal{B}^1$ minimize the cost of f in $\mathbf{W}_{j+1}^{2p}$ and $\mathbf{W}_{j+1}^{2p+1}$. Hence $\mathcal{B}^0 = \mathcal{O}_{j+1}^{2p}$ and $\mathcal{B}^1 = \mathcal{O}_{j+1}^{2p+1}$. This proves that $\mathcal{O}_j^p$ is either $\mathcal{B}_p^j$ or $\mathcal{O}_{j+1}^{2p} \cup \mathcal{O}_{j+1}^{2p+1}$. The best basis is obtained by comparing the cost of these two possibilities. ■

The best basis of the space at the root of the tree is obtained by finding the best bases of all spaces $\mathbf{W}_j^p$ in the tree, with a bottom-up progression. At the bottom of the tree, each $\mathbf{W}_J^p$ is not subdecomposed. The best basis of $\mathbf{W}_J^p$ is thus the only basis available: $\mathcal{O}_J^p = \mathcal{B}_J^p$. The best bases of the spaces $\{\mathbf{W}_j^p\}_p$ are then recursively computed from the best bases of the spaces $\{\mathbf{W}_{j+1}^p\}_p$ with the aggregation relation (9.74). Repeating this for $j > J$ until the root gives the best basis of f in $\mathbf{W}_0^0$ for local cosine bases and in $\mathbf{W}_L^0$ for wavelet packet bases.

The fast wavelet packet or local cosine algorithms compute the inner product of f with all the vectors in the tree with respectively $O(N \log_2 N)$ and $O(N(\log_2 N)^2)$ operations. At a level of the tree indexed by j, there is a total of N vectors in the orthogonal bases $\{\mathcal{B}_j^p\}_p$. The costs $\{C(f,\mathcal{B}_j^p)\}_p$ are thus calculated with $O(N)$ operations by summing (9.72). The computation of the best basis of all the spaces $\{\mathbf{W}_j^p\}_p$ from the best bases of $\{\mathbf{W}_{j+1}^p\}_p$ via (9.74) thus requires $O(N)$ operations. Since the depth of the tree is smaller than $\log_2 N$, the best basis of the space at the root is selected with $O(N \log_2 N)$ operations.

Best Bases of Images Wavelet packet and local cosine bases of images are organized in quad-trees described in Sections 8.2.1 and 8.5.3. Each node of the quad-tree is associated to a space $\mathbf{W}_j^{p,q}$, which admits a separable basis $\mathcal{B}_j^{p,q}$ of wavelet packet or local cosine vectors. This space is divided into four subspaces located at the four children nodes:

$$\mathbf{W}_j^{p,q} = \mathbf{W}_{j+1}^{2p,2q} \oplus \mathbf{W}_{j+1}^{2p+1,2q} \oplus \mathbf{W}_{j+1}^{2p,2q+1} \oplus \mathbf{W}_{j+1}^{2p+1,2q+1}.$$

The union of orthogonal bases of the four children spaces thus defines an orthogonal basis of $\mathbf{W}_j^{p,q}$. At the root of the quad-tree is a space of dimension N^2, which corresponds to $\mathbf{W}_0^{0,0}$ for local cosine bases and to $\mathbf{W}_L^{0,0}$ with $2^L = N^{-1}$ for wavelet packet bases.

Let $\mathcal{O}_j^{p,q}$ be the best basis $\mathbf{W}_j^{p,q}$ for a signal f. Like Proposition 9.9 the following proposition relates the best basis of $\mathbf{W}_j^{p,q}$ to the best bases of its children. It is proved with the same derivations.

Proposition 9.10 (COIFMAN, WICKERHAUSER) *Suppose that C is an additive cost function. If*

$$\begin{aligned} C(f, \mathcal{B}_j^{p,q}) \;<\; & C(f, \mathcal{O}_{j+1}^{2p,2q}) + C(f, \mathcal{O}_{j+1}^{2p+1,2q}) + \\ & C(f, \mathcal{O}_{j+1}^{2p,2q+1}) + C(f, \mathcal{O}_{j+1}^{2p+1,2q+1}) \end{aligned}$$

then

$$\mathcal{O}_j^{p,q} = \mathcal{B}_j^{p,q}$$

otherwise

$$\mathcal{O}_j^{p,q} = \mathcal{O}_{j+1}^{2p,2q} \cup \mathcal{O}_{j+1}^{2p+1,2q} \cup \mathcal{O}_{j+1}^{2p,2q+1} \cup \mathcal{O}_{j+1}^{2p+1,2q+1}.$$

This recursive relation computes the best basis of $\{\mathbf{W}_j^{p,q}\}_{p,q}$ from the best bases of the spaces $\{\mathbf{W}_{j+1}^{p,q}\}_{p,q}$, with $O(N^2)$ operations. Iterating this procedure from the bottom of the tree to the top finds the best basis of f with $O(N^2 \log_2 N)$ calculations.

9.4.3 Wavelet Packet and Local Cosine Best Bases

The performance of best wavelet packet and best local cosine approximations depends on the time-frequency properties of f. We evaluate these approximations through examples that also reveal their limitations.

Best Wavelet Packet Bases A wavelet packet basis divides the frequency axis into intervals of varying sizes. Each frequency interval is covered by a wavelet packet function that is translated uniformly in time. A best wavelet packet basis can thus be interpreted as a "best" frequency segmentation.

A signal is well approximated by a best wavelet packet basis if in any frequency interval, the high energy structures have a similar time-frequency spread. The time translation of the wavelet packet that covers this frequency interval is then well adapted to approximating all the signal structures in this frequency range that appear at different times. Figure 9.7 gives the best wavelet packet basis computed with the entropy $\Phi(u) = -u \log_e u$, for a signal composed of two hyperbolic chirps. The wavelet packet tree was calculated with the Symmlet 8 conjugate mirror filter. The time-support of the wavelet packets is reduced at high frequencies to adapt itself to the rapid modification of the chirps' frequency content. The energy distribution revealed by the wavelet packet Heisenberg boxes is similar to the scalogram shown in Figure 4.17. Figure 8.6 gives another example of a best wavelet packet basis, for a different multi-chirp signal. Let us mention that the application of best wavelet packet bases to pattern recognition remains difficult because they are not translation invariant. If the signal is translated, its wavelet packet coefficients are

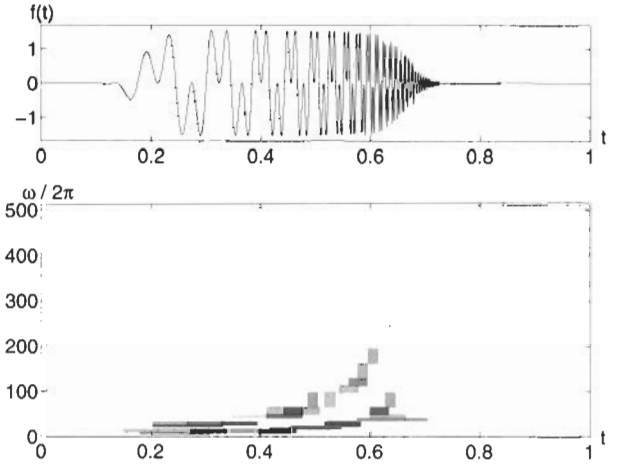


FIGURE 9.7 The top signal includes two hyperbolic chirps. The Heisenberg boxes of the best wavelet packet basis are shown below. The darkness of each rectangle is proportional to the amplitude of the wavelet packet coefficient.

severely modified and the minimization of the cost function may yield a different basis. This remark applies to local cosine bases as well.

If the signal includes *different types* of high energy structures, located at different times but in the same frequency interval, there is no wavelet packet basis that is well adapted to all of them. Consider, for example a sum of four transients centered respectively at u_0 and u_1, at two different frequencies ξ_0 and ξ_1:

$$f(t) = \frac{K_0}{\sqrt{s_0}} g\left(\frac{t-u_0}{s_0}\right) \exp(i\xi_0 t) + \frac{K_1}{\sqrt{s_1}} g\left(\frac{t-u_1}{s_1}\right) \exp(i\xi_0 t) \quad (9.75)$$
$$+ \frac{K_2}{\sqrt{s_1}} g\left(\frac{t-u_0}{s_1}\right) \exp(i\xi_1 t) + \frac{K_3}{\sqrt{s_0}} g\left(\frac{t-u_1}{s_0}\right) \exp(i\xi_1 t).$$

The smooth window g has a Fourier transform $\hat{g}$ whose energy is concentrated at low frequencies. The Fourier transform of the four transients have their energy concentrated in frequency bands centered respectively at ξ_0 and ξ_1:

$$\begin{aligned}\hat{f}(\omega) = {} & K_0 \sqrt{s_0}\, \hat{g}\Big(s_0(\omega-\xi_0)\Big) \exp(-iu_0[\omega-\xi_0]) \\ & + K_1 \sqrt{s_1}\, \hat{g}\Big(s_1(\omega-\xi_0)\Big) \exp(-iu_1[\omega-\xi_0]) \\ & + K_2 \sqrt{s_1}\, \hat{g}\Big(s_1(\omega-\xi_1)\Big) \exp(-iu_0[\omega-\xi_1]) \\ & + K_3 \sqrt{s_0}\, \hat{g}\Big(s_0(\omega-\xi_1)\Big) \exp(-iu_1[\omega-\xi_1]).\end{aligned}$$

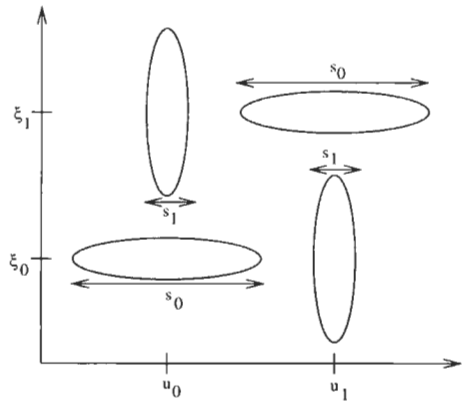


FIGURE 9.8 Time-frequency energy distribution of the four elementary atoms in (9.75).

If s_0 and s_1 have different values, the time and frequency spread of these transients is different, which is illustrated in Figure 9.8. In the best wavelet packet basis selection, the first transient $K_0 s_0^{-1/2} g(s_0^{-1}(t-u_0))\exp(i\xi_0 t)$ "votes" for a wavelet packet whose scale 2^j is of the order s_0 at the frequency ξ_0 whereas $K_1 s_1^{-1/2} g(s_1^{-1}(t-u_1))\exp(i\xi_0 t)$ "votes" for a wavelet packet whose scale 2^j is close to s_1 at the same frequency. The "best" wavelet packet is adapted to the transient of highest energy, which yields the strongest vote in the cost (9.71). The energy of the smaller transient is then spread across many "best" wavelet packets. The same thing happens for the second pair of transients located in the frequency neighborhood of ξ_1.

Speech recordings are examples of signals whose properties change rapidly in time. At two different instants, in the same frequency neighborhood, the signal may have a totally different energy distributions. A best wavelet packet is not adapted to this time variation and gives poor non-linear approximations.

As in one dimension, an image is well approximated in a best wavelet packet basis if its structures within a given frequency band have similar properties across the whole image. For natural scene images, the best wavelet packet often does not provide much better non-linear approximations than the wavelet basis included in this wavelet packet dictionary. For specific classes of images such as fingerprints, one may find wavelet packet bases that outperform significantly the wavelet basis [103].

Best Local Cosine Bases A local cosine basis divides the time axis into intervals of varying sizes. A best local cosine basis thus adapts the time segmentation to the variations of the signal time-frequency structures. In comparison with wavelet packets, we gain time adaptation but we lose frequency flexibility. A best local

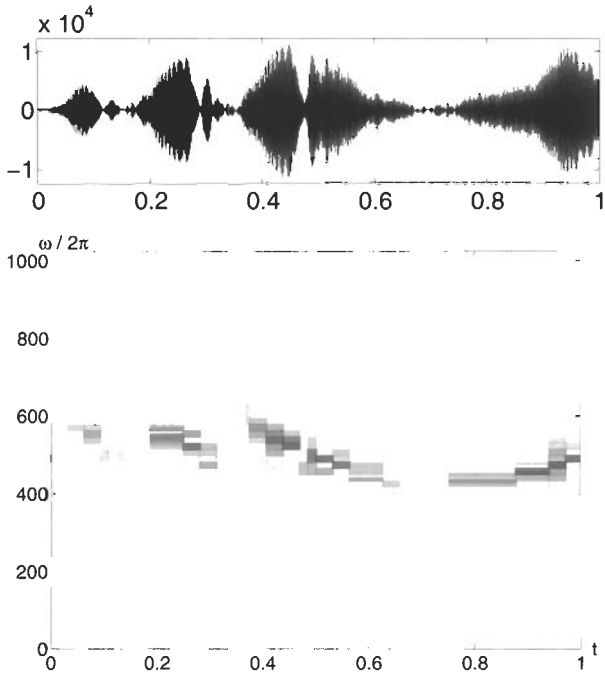


FIGURE 9.9 Recording of bird song. The Heisenberg boxes of the best local cosine basis are shown below. The darkness of each rectangle is proportional to the amplitude of the local cosine coefficient.

cosine basis is therefore well adapted to approximating signals whose properties may vary in time, but which do not include structures of very different time and frequency spread at any given time. Figure 9.9 shows the Heisenberg boxes of the best local cosine basis for the recording of a bird song, computed with an entropy cost. Figure 8.19 shows the best local cosine basis for a speech recording.

The sum of four transients (9.75) is not efficiently represented in a wavelet packet basis but neither is it well approximated in a best local cosine basis. Indeed, if the scales s_0 and s_1 are very different, at u_0 and u_1 this signal includes two transients at the frequency ξ_0 and ξ_1 that have a very different time-frequency spread. In each time neighborhood, the size of the window is adapted to the transient of highest energy. The energy of the second transient is spread across many local cosine vectors. Efficient approximations of such signals require using larger dictionaries of bases, which can simultaneously divide the time and frequency axes in intervals of various sizes [208].

In two dimensions, a best local cosine basis divides an image into square windows whose sizes are adapted to the spatial variations of local image structures. Figure 9.10 shows the best basis segmentation of the Barbara image, computed

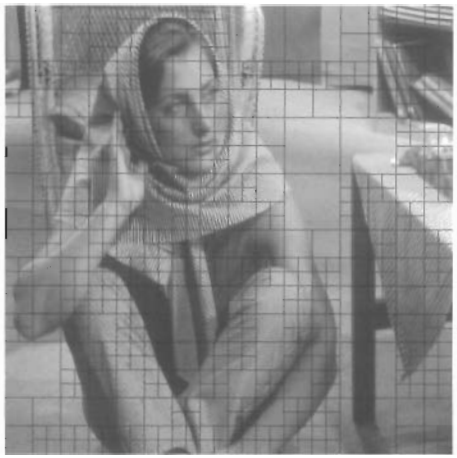


FIGURE 9.10 The grid shows the approximate support of square overlapping windows in the best local cosine basis, computed with an $\mathbf{l}^1$ cost.

with an $\mathbf{l}^1$ cost calculated with $\Phi(u) = u^{1/2}$. The squares are bigger in regions where the image structures remain nearly the same. Figure 8.22 shows another example of image segmentation with a best local cosine basis computed with the same cost function. As in one dimension, a best local cosine basis is an efficient representation if the image does not include very different frequency structures in the same spatial region.

9.5 APPROXIMATIONS WITH PURSUITS [3]

A music recording often includes notes of different durations at the same time, which means that such a signal is not well represented in a best local cosine basis. The same musical note may also have different durations when played at different times, in which case a best wavelet packet basis is also not well adapted to represent this sound. To approximate musical signals efficiently, the decomposition must have the same flexibility as the composer, who can freely choose the time-frequency atoms (notes) that are best adapted to represent a sound.

Wavelet packet and local cosine dictionaries include $P = N \log_2 N$ different vectors. The set of orthogonal bases is much smaller than the set of non-orthogonal bases that could be constructed by choosing N linearly independent vectors from these P. To improve the approximation of complex signals such as music recordings, we study general non-orthogonal signal decompositions.

Consider the space of signals of size N. Let $\mathcal{D} = \{g_p\}_{0 \leq p < P}$ be a redundant dictionary of $P > N$ vectors, which includes at least N linearly independent vectors. For any $M \geq 1$, an approximation f_M of f may be calculated with a linear

combination of any M dictionary vectors:

$$f_M = \sum_{m=0}^{M-1} a[p_m]\, g_{p_m}.$$

The freedom of choice opens the door to a considerable combinatorial explosion. For general dictionaries of $P > N$ vectors, computing the approximation f_M that minimizes $\|f - f_M\|$ is an *NP hard problem* [151]. This means that there is no known polynomial time algorithm that can solve this optimization.

Pursuit algorithms reduce the computational complexity by searching for efficient but non-optimal approximations. A basis pursuit formulates the search as a linear programming problem, providing remarkably good approximations with $O(N^{3.5} \log_2^{3.5} N)$ operations. For large signals, this remains prohibitive. Matching pursuits are faster greedy algorithms whose applications to large time-frequency dictionaries is described in Section 9.5.2. An orthogonalized pursuit is presented in Section 9.5.3.

9.5.1 Basis Pursuit

We study the construction of a "best" basis $\mathcal{B}$, not necessarily orthogonal, for efficiently approximating a signal f. The N vectors of $\mathcal{B} = \{g_{p_m}\}_{0 \le m < N}$ are selected from a redundant dictionary $\mathcal{D} = \{g_p\}_{0 \le p < P}$ with a pursuit elaborated by Chen and Donoho [119]. Let us decompose f in this basis:

$$f = \sum_{m=0}^{N-1} a[p_m]\, g_{p_m}. \tag{9.76}$$

If we had restricted ourselves to orthogonal bases, Section 9.4.1 explains that the basis choice would be optimized by minimizing

$$C(f, \mathcal{B}) = \sum_{m=0}^{N-1} \Phi\left(\frac{|a[p_m]|^2}{\|f\|^2}\right), \tag{9.77}$$

where $\Phi(u)$ is concave. For non-orthogonal bases, this result does not hold in general.

Despite the absence of orthogonality, a basis pursuit searches for a "best" basis that minimizes (9.77) for $\Phi(u) = u^{1/2}$:

$$C(f, \mathcal{B}) = \frac{1}{\|f\|} \sum_{m=0}^{N-1} |a[p_m]|. \tag{9.78}$$

Minimizing the $\mathbf{l}^1$ norm of the decomposition coefficients avoids diffusing the energy of f among many vectors. It reduces cancellations between the vectors $a[p_m] g_{p_m}$ that decompose f, because such cancellations increase $|a[p_m]|$ and thus increase the cost (9.78). The minimization of an $\mathbf{l}^1$ norm is also related to linear programming, which leads to fast computational algorithms.

Linear Programming Instead of immediately isolating subsets of N vectors in the dictionary $\mathcal{D}$, a linear system of size P is written with all dictionary vectors

$$\sum_{p=0}^{P-1} a[p]\, g_p[n] = f[n], \tag{9.79}$$

while trying to minimize

$$\sum_{p=0}^{P-1} |a[p]|. \tag{9.80}$$

The system (9.79) can be expressed in matrix form with the $P \times N$ matrix $G = \{g_p[n]\}_{0 \le n < N, 0 \le p < P}$

$$Ga = f. \tag{9.81}$$

Although the minimization of (9.80) is nonlinear, it can be reformulated as a linear programming problem.

A standard-form linear programming problem [28] is a constrained optimization over positive vectors of size L. Let $b[n]$ be a vector of size $N < L$, $c[p]$ a non-zero vector of size L and $A[n, p]$ an $L \times N$ matrix. We must find $x[p] \in \mathbb{R}^L$ such that $x[p] \ge 0$, while minimizing

$$\sum_{p=0}^{L-1} x[p]\, c[p] \tag{9.82}$$

subject to

$$Ax = b.$$

To reformulate the minimization of (9.80) subject to (9.81) as a linear programming problem, we introduce "slack variables" $u[p] \ge 0$ and $v[p] \ge 0$ such that

$$a[p] = u[p] - v[p].$$

As a result

$$Ga = Gu - Gv = f \tag{9.83}$$

and

$$\sum_{p=0}^{P-1} |a[p]| = \sum_{p=0}^{P-1} u[p] + \sum_{p=0}^{P-1} v[p]. \tag{9.84}$$

We thus obtain a standard form linear programming of size $L = 2P$ with

$$A = (G, -G) \;\; , \;\; x = \begin{pmatrix} u \\ v \end{pmatrix} \;\; , \;\; b = f \;\; , \;\; c = 1.$$

The matrix A of size $N \times L$ has rank N because the dictionary $\mathcal{D}$ includes N linearly independent vectors. A standard result of linear programming [28] proves

that the vector x has at most N non-zero coefficients. One can also verify that if $a[p] > 0$ then $a[p] = u[p]$ and $v[p] = 0$ whereas if $a[p] \leq 0$ then $a[p] = v[p]$ and $u[p] = 0$. In the non-degenerate case, which is most often encountered, the non-zero coefficients of $x[p]$ thus correspond to N indices $\{p_m\}_{0 \leq m < N}$ such that $\{g_{p_m}\}_{0 \leq m < N}$ are linearly independent. This is the best basis of $\mathbb{R}^N$ that minimizes the cost (9.78).

Linear Programming Computations The collection of feasible points $\{x : Ax = b\ ,\ x \geq 0\}$ is a convex polyhedron in $\mathbb{R}^L$. The vertices of this polyhedron are solutions $x[p]$ having at most N non-zero coefficients. The linear cost (9.82) can be minimum only at a vertex of this polyhedron. In the non-degenerate case, the N non-zero coefficients correspond to N column vectors $\mathcal{B} = \{g_{p_m}\}_{0 \leq m < N}$ that form a basis.

One can also prove [28] that if the cost is not minimum at a given vertex then there exists an adjacent vertex whose cost is smaller. The simplex algorithm takes advantage of this property by jumping from one vertex to an adjacent vertex while reducing the cost (9.82). Going to an adjacent vertex means that one of the zero coefficients of $x[p]$ becomes non-zero while one non-zero coefficient is set to zero. This is equivalent to modifying the basis $\mathcal{B}$ by replacing one vector by another vector of $\mathcal{D}$. The simplex algorithm thus progressively improves the basis by appropriate modifications of its vectors, one at a time. In the worst case, all vertices of the polyhedron will be visited before finding the solution, but the average case is much more favorable.

Since the 1980's, more effective interior point procedures have been developed. Karmarkar's interior point algorithm [234] begins in the middle of the polyhedron and converges by iterative steps towards the vertex solution, while remaining inside the convex polyhedron. For finite precision calculations, when the algorithm has converged close enough to a vertex, it jumps directly to the corresponding vertex, which is guaranteed to be the solution. The middle of the polyhedron corresponds to a decomposition of f over all vectors of $\mathcal{D}$, typically with $P > N$ non-zero coefficients. When moving towards a vertex some coefficients progressively decrease while others increase to improve the cost (9.82). If only N decomposition coefficients are significant, jumping to the vertex is equivalent to setting all other coefficients to zero. Each step requires computing the solution of a linear system. If A is an $N \times L$ matrix then Karmarkar's algorithm terminates with $O(L^{3.5})$ operations. Mathematical work on interior point methods has led to a large variety of approaches that are summarized in [252]. The basis pursuit of Chen and Donoho [119] is implemented in WaveLab with a "Log-barrier" method [252], which converges more quickly than Karmarkar's original algorithm

Wavelet Packet and Local Cosine Dictionaries These dictionaries have $P = N \log_2 N$ time-frequency atoms. A straightforward implementation of interior point algorithms thus requires $O(N^{3.5} \log_2^{3.5} N)$ operations. By using the fast wavelet packet and local cosine transforms together with heuristic computational

rules, the number of operations is considerably reduced [119]. The algorithm still remains relatively slow and the computations become prohibitive for $N \geq 1000$.

Figure 9.11 decomposes a synthetic signal that has two high frequency transients followed by two lower frequency transients and two Diracs for $t < 0.2$. The signal then includes two linear chirps that cross each other and which are superimposed with localized sinusoidal waves. In a dictionary of wavelet packet bases calculated with a Daubechies 8 filter, the best basis shown in Figure 9.11(c) optimizes the division of the frequency axis, but it has no flexibility in time. It is therefore not adapted to the time evolution of the signal components. A basis pursuit algorithm adapts the wavelet packet choice to the local signal structures; Figure 9.11(d) shows that it better reveals its time-frequency properties.

9.5.2 Matching Pursuit

Despite the linear programming approach, a basis pursuit is computationally expensive because it minimizes a global cost function over all dictionary vectors. The matching pursuit introduced by Mallat and Zhang [259] reduces the computational complexity with a greedy strategy. It is closely related to projection pursuit algorithms used in statistics [184] and to shape-gain vector quantizations [27]. Vectors are selected one by one from the dictionary, while optimizing the signal approximation at each step.

Let $\mathcal{D} = \{g_\gamma\}_{\gamma \in \Gamma}$ be a dictionary of $P > N$ vectors, having a unit norm. This dictionary includes N linearly independent vectors that define a basis of the space $\mathbb{C}^N$ of signals of size N. A matching pursuit begins by projecting f on a vector $g_{\gamma_0} \in \mathcal{D}$ and computing the residue Rf:

$$f = \langle f, g_{\gamma_0} \rangle \, g_{\gamma_0} + Rf. \tag{9.85}$$

Since Rf is orthogonal to g_{γ_0}

$$\|f\|^2 = |\langle f, g_{\gamma_0} \rangle|^2 + \|Rf\|^2. \tag{9.86}$$

To minimize $\|Rf\|$ we must choose $g_{\gamma_0} \in \mathcal{D}$ such that $|\langle f, g_{\gamma_0} \rangle|$ is maximum. In some cases, it is computationally more efficient to find a vector g_{γ_0} that is almost optimal:

$$|\langle f, g_{\gamma_0} \rangle| \geq \alpha \sup_{\gamma \in \Gamma} |\langle f, g_\gamma \rangle|, \tag{9.87}$$

where $\alpha \in (0, 1]$ is an optimality factor. The pursuit iterates this procedure by subdecomposing the residue. Let $R^0 f = f$. Suppose that the m^{th} order residue $R^m f$ is already computed, for $m \geq 0$. The next iteration chooses $g_{\gamma_m} \in \mathcal{D}$ such that

$$|\langle R^m f, g_{\gamma_m} \rangle| \geq \alpha \sup_{\gamma \in \Gamma} |\langle R^m f, g_\gamma \rangle|, \tag{9.88}$$

and projects $R^m f$ on g_{γ_m}:

$$R^m f = \langle R^m f, g_{\gamma_m} \rangle \, g_{\gamma_m} + R^{m+1} f. \tag{9.89}$$

The orthogonality of $R^{m+1}f$ and g_{γ_m} implies

$$\|R^m f\|^2 = |\langle R^m f, g_{\gamma_m}\rangle|^2 + \|R^{m+1} f\|^2. \tag{9.90}$$

Summing (9.89) from m between 0 and $M-1$ yields

$$f = \sum_{m=0}^{M-1} \langle R^m f, g_{\gamma_m}\rangle\, g_{\gamma_m} + R^M f. \tag{9.91}$$

Similarly, summing (9.90) from m between 0 and $M-1$ gives

$$\|f\|^2 = \sum_{m=0}^{M-1} |\langle R^m f, g_{\gamma_m}\rangle|^2 + \|R^M f\|^2. \tag{9.92}$$

The following theorem proves that $\|R^m f\|$ converges exponentially to 0 when m tends to infinity.

Theorem 9.10 *There exists $\lambda > 0$ such that for all $m \geq 0$*

$$\|R^m f\| \leq 2^{-\lambda m}\, \|f\|. \tag{9.93}$$

As a consequence

$$f = \sum_{m=0}^{+\infty} \langle R^m f, g_{\gamma_m}\rangle\, g_{\gamma_m}, \tag{9.94}$$

and

$$\|f\|^2 = \sum_{m=0}^{+\infty} |\langle R^m f, g_{\gamma_m}\rangle|^2. \tag{9.95}$$

Proof[3]. Let us first verify that there exists $\beta > 0$ such that for any $f \in \mathbb{C}^N$

$$\sup_{\gamma\in\Gamma} |\langle f, g_\gamma\rangle| \geq \beta \|f\|. \tag{9.96}$$

Suppose that it is not possible to find such a β. This means that we can construct $\{f_m\}_{m\in\mathbb{N}}$ with $\|f_m\| = 1$ and

$$\lim_{m\to+\infty} \sup_{\gamma\in\Gamma} |\langle f_m, g_\gamma\rangle| = 0. \tag{9.97}$$

Since the unit sphere of $\mathbb{C}^N$ is compact, there exists a sub-sequence $\{f_{m_k}\}_{k\in\mathbb{N}}$ that converges to a unit vector $f \in \mathbb{C}^N$. It follows that

$$\sup_{\gamma\in\Gamma} |\langle f, g_\gamma\rangle| = \lim_{k\to+\infty} \sup_{\gamma\in\Gamma} |\langle f_{m_k}, g_\gamma\rangle| = 0 \tag{9.98}$$

so $\langle f, g_\gamma\rangle = 0$ for all $g_\gamma \in \mathcal{D}$. Since $\mathcal{D}$ contains a basis of $\mathbb{C}^N$, necessarily $f = 0$ which is not possible because $\|f\| = 1$. This proves that our initial assumption is wrong, and hence there exists β such that (9.96) holds.

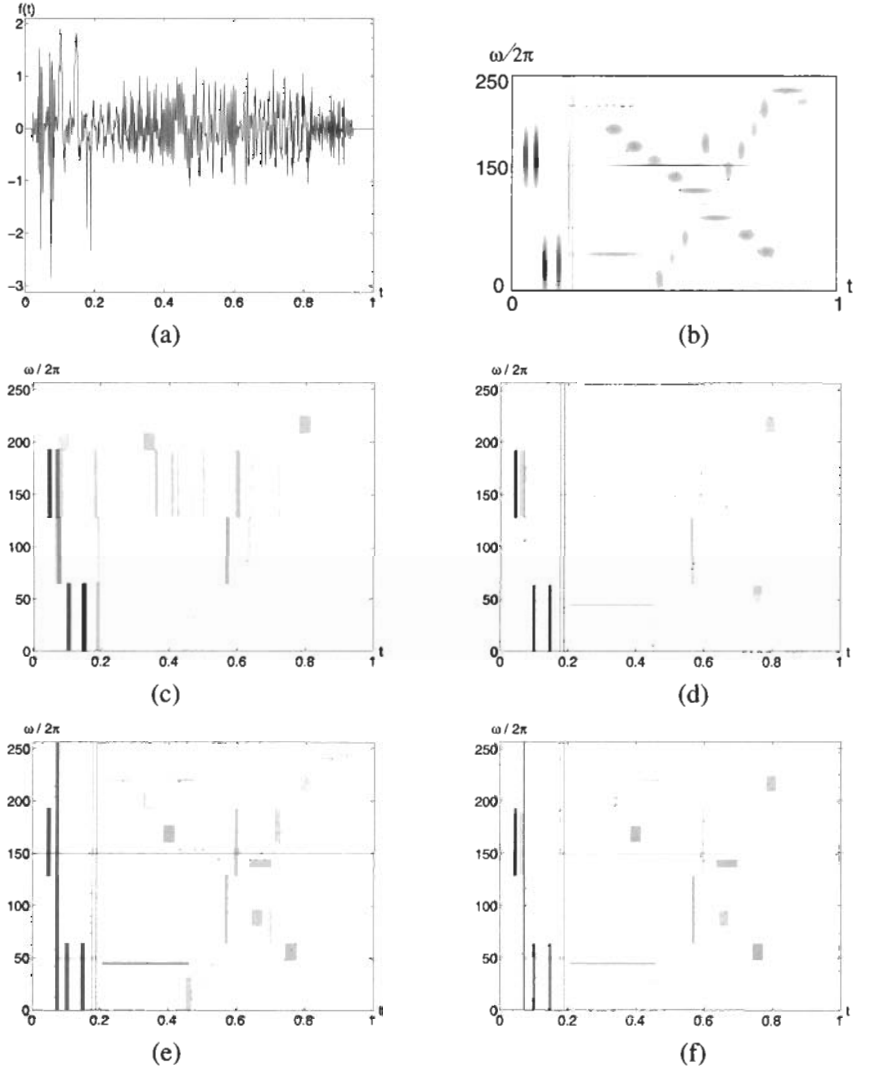


FIGURE 9.11 (a): Signal synthesized with a sum of chirps, truncated sinusoids, short time transients and Diracs. The time-frequency images display the atoms selected by different adaptive time-frequency transforms. The darkness is proportional to the coefficient amplitude. (b): Gabor matching pursuit. Each dark blob is the Wigner-Ville distribution of a selected Gabor atom. (c): Heisenberg boxes of a best wavelet packet basis calculated with Daubechies 8 filter. (d): Wavelet packet basis pursuit. (e): Wavelet packet matching pursuit. (f): Wavelet packet orthogonal matching pursuit.

The decay condition (9.93) is derived from the energy conservation

$$\|R^{m+1}f\|^2 = \|R^m f\|^2 - |\langle R^m f, g_{p_m}\rangle|^2.$$

The matching pursuit chooses g_{γ_m} that satisfies

$$|\langle R^m f, g_{\gamma_m}\rangle| \geq \alpha \sup_{\gamma\in\Gamma} |\langle R^m f, g_\gamma\rangle|, \tag{9.99}$$

and (9.96) implies that $|\langle R^m f, g_{\gamma_m}\rangle| \geq \alpha\beta\|R^m f\|$. So

$$\|R^{m+1}f\| \leq \|R^m f\|\,(1-\alpha^2\beta^2)^{1/2}, \tag{9.100}$$

which verifies (9.93) for

$$2^{-\lambda} = (1-\alpha^2\beta^2)^{1/2} < 1.$$

This also proves that $\lim_{m\to+\infty} \|R^m f\| = 0$. Equation (9.94) and (9.95) are thus derived from (9.91) and (9.92). ■

The convergence rate λ decreases when the size N of the signal space increases. In the limit of infinite dimensional spaces, Jones' theorem proves that the algorithm still converges but the convergence is not exponential [230, 259]. The asymptotic behavior of a matching pursuit is further studied in Section 10.5.2. Observe that even in finite dimensions, an infinite number of iterations is necessary to completely reduce the residue. In most signal processing applications, this is not an issue because many fewer than N iterations are needed to obtain sufficiently precise signal approximations. Section 9.5.3 describes an orthogonalized matching pursuit that converges in fewer than N iterations.

Fast Network Calculations A matching pursuit is implemented with a fast algorithm that computes $\langle R^{m+1}f, g_\gamma\rangle$ from $\langle R^m f, g_\gamma\rangle$ with a simple *updating* formula. Taking an inner product with g_γ on each side of (9.89) yields

$$\langle R^{m+1}f, g_\gamma\rangle = \langle R^m f, g_\gamma\rangle - \langle R^m f, g_{\gamma_m}\rangle \langle g_{\gamma_m}, g_\gamma\rangle. \tag{9.101}$$

In neural network language, this is an inhibition of $\langle R^m f, g_\gamma\rangle$ by the selected pattern g_{γ_m} with a weight $\langle g_{\gamma_m}, g_\gamma\rangle$ that measures its correlation with g_γ. To reduce the computational load, it is necessary to construct dictionaries with vectors having a sparse interaction. This means that each $g_\gamma \in \mathcal{D}$ has non-zero inner products with only a small fraction of all other dictionary vectors. It can also be viewed as a network that is not fully connected. Dictionaries are designed so that non-zero weights $\langle g_\alpha, g_\gamma\rangle$ can be retrieved from memory or computed with $O(1)$ operations. A matching pursuit with a relative precision ϵ is implemented with the following steps.

1. *Initialization* Set $m = 0$ and compute $\{\langle f, g_\gamma\rangle\}_{\gamma\in\Gamma}$.

2. *Best match* Find $g_{\gamma_m} \in \mathcal{D}$ such that

$$|\langle R^m f, g_{\gamma_m}\rangle| \geq \alpha \sup_{\gamma\in\Gamma} |\langle R^m f, g_\gamma\rangle|. \tag{9.102}$$

3. *Update* For all $g_\gamma \in \mathcal{D}$ with $\langle g_{\gamma_m}, g_\gamma \rangle \neq 0$

$$\langle R^{m+1} f, g_\gamma \rangle = \langle R^m f, g_\gamma \rangle - \langle R^m f, g_{\gamma_m} \rangle \langle g_{\gamma_m}, g_\gamma \rangle. \tag{9.103}$$

4. *Stopping rule* If

$$\|R^{m+1} f\|^2 = \|R^m f\|^2 - |\langle R^m f, g_{\gamma_m} \rangle|^2 \leq \epsilon^2 \|f\|^2$$

then stop. Otherwise $m = m + 1$ and go to 2.

If $\mathcal{D}$ is very redundant, computations at steps 2 and 3 are reduced by performing the calculations in a sub-dictionary $\mathcal{D}_s = \{g_\gamma\}_{\gamma \in \Gamma_s} \subset \mathcal{D}$. The sub-dictionary $\mathcal{D}_s$ is constructed so that if $g_{\tilde{\gamma}_m} \in \mathcal{D}_s$ maximizes $|\langle f, g_\gamma \rangle|$ in $\mathcal{D}_s$ then there exists $g_{\gamma_m} \in \mathcal{D}$ which satisfies (9.102) and whose index γ_m is "close" to $\tilde{\gamma}_m$. The index γ_m is found with a local search. This is done in time-frequency dictionaries where a sub-dictionary can be sufficient to indicate a time-frequency region where an almost best match is located. The updating (9.103) is then restricted to vectors $g_\gamma \in \mathcal{D}_s$.

The particular choice of a dictionary $\mathcal{D}$ depends upon the application. Specific dictionaries for inverse electro-magnetic problems, face recognition and data compression are constructed in [268, 229, 279]. In the following, we concentrate on dictionaries of local time-frequency atoms.

Wavelet Packets and Local Cosines Wavelet packet and local cosine trees constructed in Sections 8.2.1 and 8.5.3 are dictionaries containing $P = N \log_2 N$ vectors. They have a sparse interaction and non-zero inner products of dictionary vectors can be stored in tables. Each matching pursuit iteration then requires $O(N \log_2 N)$ operations.

Figure 9.11(e) is an example of a matching pursuit decomposition calculated in a wavelet packet dictionary. Compared to the best wavelet packet basis shown in Figure 9.11(c), it appears that the flexibility of the matching pursuit selects wavelet packet vectors that give a more compact approximation, which reveals better the signal time-frequency structures. However, a matching pursuit requires more computations than a best basis selection.

In this example, matching pursuit and basis pursuit algorithms give similar results. In some cases, a matching pursuit does not perform as well as a basis pursuit because the greedy strategy selects decomposition vectors one by one [159]. Choosing decomposition vectors by optimizing a correlation inner product can produce a partial loss of time and frequency resolution [119]. High resolution pursuits avoid the loss of resolution in time by using non-linear correlation measures [195, 223] but the greediness can still have adverse effects.

Translation Invariance Section 5.4 explains that decompositions in orthogonal bases lack translation invariance and are thus difficult to use for pattern recognition. Matching pursuits are translation invariant if calculated in translation invariant

dictionaries. A dictionary $\mathcal{D}$ is *translation invariant* if for any $g_\gamma \in \mathcal{D}$ then $g_\gamma[n-p] \in \mathcal{D}$ for $0 \leq p < N$. Suppose that the matching decomposition of f in $\mathcal{D}$ is

$$f[n] = \sum_{m=0}^{M-1} \langle R^m f, g_{\gamma_m} \rangle \, g_{\gamma_m}[n] + R^M f[n]. \tag{9.104}$$

One can verify [151] that the matching pursuit of $f_p[n] = f[n-p]$ selects a translation by p of the same vectors g_{γ_m} with the same decomposition coefficients

$$f_p[n] = \sum_{m=0}^{M-1} \langle R^m f, g_{\gamma_m} \rangle \, g_{\gamma_m}[n-p] + R^M f_p[n].$$

Patterns can thus be characterized independently of their position. The same translation invariance property is valid for a basis pursuit. However, translation invariant dictionaries are necessarily very large, which often leads to prohibitive calculations. Wavelet packet and local cosine dictionaries are not translation invariant because at each scale 2^j the waveforms are translated only by $k2^j$ with $k \in \mathbb{Z}$.

Translation invariance is generalized as an invariance with respect to any group action [151]. A frequency translation is another example of a group operation. If the dictionary is invariant under the action of a group then the pursuit remains invariant under the action of the same group.

Gabor Dictionary A time and frequency translation invariant Gabor dictionary is constructed by Qian and Chen [287] as well as Mallat and Zhong [259], by scaling, translating and modulating a Gaussian window. Gaussian windows are used because of their optimal time and frequency energy concentration, proved by the uncertainty Theorem 2.5.

For each scale 2^j, a discrete window of period N is designed by sampling and periodizing a Gaussian $g(t) = 2^{1/4}\exp(-\pi t^2)$:

$$g_j[n] = K_j \sum_{p=-\infty}^{+\infty} g\left(\frac{n-pN}{2^j}\right).$$

The constant K_j is adjusted so that $\|g_j\| = 1$. This window is then translated in time and frequency. Let Γ be the set of indexes $\gamma = (p,k,2^j)$ for $(p,k) \in [0,N-1]^2$ and $j \in [0, \log_2 N]$. A discrete Gabor atom is

$$g_\gamma[n] = g_j[n-p] \exp\left(\frac{i2\pi kn}{N}\right). \tag{9.105}$$

The resulting Gabor dictionary $\mathcal{D} = \{g_\gamma\}_{\gamma\in\Gamma}$ is time and frequency translation invariant modulo N. A matching pursuit decomposes real signals in this dictionary by grouping atoms g_{γ^+} and g_{γ^-} with $\gamma^\pm = (p, \pm k, 2^j)$. At each iteration, instead

of projecting $R^m f$ over an atom g_γ, the matching pursuit computes its projection on the plane generated by $(g_{\gamma^+}, g_{\gamma^-})$. Since $R^m f[n]$ is real, one can verify that this is equivalent to projecting $R^m f$ on a real vector that can be written

$$g_\gamma^\phi[n] = K_{j,\phi}\, g_j[n-p] \cos\left(\frac{2\pi kn}{N} + \phi\right).$$

The constant $K_{j,\phi}$ sets the norm of this vector to 1 and the phase ϕ is optimized to maximize the inner product with $R^m f$. Matching pursuit iterations yield

$$f = \sum_{m=0}^{+\infty} \langle R^m f, g_{\gamma_m}^{\phi_m} \rangle\, g_{\gamma_m}^{\phi_m}. \tag{9.106}$$

This decomposition is represented by a time-frequency energy distribution obtained by summing the Wigner-Ville distribution $P_V g_{\gamma_m}[n,k]$ of the complex atoms g_{γ_m}:

$$P_M f[n,k] = \sum_{m=0}^{+\infty} |\langle R^m f, g_{\gamma_m}^{\phi_m} \rangle|^2\, P_V g_{\gamma_m}[n,k]. \tag{9.107}$$

Since the window is Gaussian, if $\gamma_m = (p_m, k_m, 2^{j_m})$ then $P_V g_{\gamma_m}$ is a two-dimensional Gaussian blob centered at (p_m, k_m) in the time-frequency plane. It is scaled by 2^{j_m} in time and $N2^{-j_m}$ in frequency.

Example 9.1 Figure 9.11(b) gives the matching pursuit energy distribution $P_M f[n,k]$ of a synthetic signal. The inner structures of this signal appear more clearly than with a wavelet packet matching pursuit because Gabor atoms have a better time-frequency localization than wavelet packets, and they are translated over a finer time-frequency grid.

Example 9.2 Figure 9.12 shows the Gabor matching pursuit decomposition of the word "greasy", sampled at 16 kHz. The time-frequency energy distribution shows the low-frequency component of the "g" and the quick burst transition to the "ea". The "ea" has many harmonics that are lined up. The "s" is noise whose time-frequency energy is spread over a high-frequency interval. Most of the signal energy is characterized by a few time-frequency atoms. For $m = 250$ atoms, $\|R^m f\|/\|f\| = .169$, although the signal has 5782 samples, and the sound recovered from these atoms is of excellent audio-quality.

Matching pursuit calculations in a Gabor dictionary are performed with a subdictionary $\mathcal{D}_s$. At each scale 2^j, the time-frequency indexes (p,k) are subsampled at intervals $a2^j$ and $aN2^{-j}$ where the sampling factor $a < 1$ is small enough to detect the time-frequency regions where the signal has high energy components. The step 2 of the matching pursuit iteration (9.102) finds the Gabor atom in $g_{\tilde{\gamma}_m} \in \mathcal{D}_s$ which best matches the signal residue. This match is then improved by searching

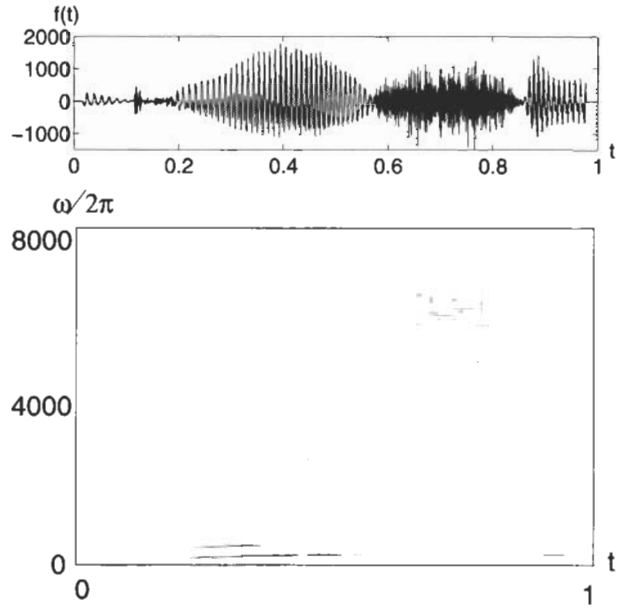


FIGURE 9.12 Speech recording of the word "greasy" sampled at $16kHz$. In the time-frequency image, the dark blobs of various sizes are the Wigner-Ville distributions of a Gabor functions selected by the matching pursuit.

for an atom $g_{\gamma_m} \in \mathcal{D}$ whose index γ_m is close to $\tilde{\gamma}_m$ and which locally maximizes the correlation with the signal residue. The updating formula (9.103) is calculated for $g_\gamma \in \mathcal{D}_s$. Inner products between two Gabor atoms are computed with an analytic formula [259]. Since $\mathcal{D}_s$ has $O(N \log_2 N)$ vectors, one can verify that each matching pursuit iteration is implemented with $O(N \log_2 N)$ calculations.

9.5.3 Orthogonal Matching Pursuit

The approximations of a matching pursuit are improved by orthogonalizing the directions of projection, with a Gram-Schmidt procedure proposed by Pati et al. [280] and Davis et al. [152]. The resulting orthogonal pursuit converges with a finite number of iterations, which is not the case for a non-orthogonal pursuit. The price to be paid is the important computational cost of the Gram-Schmidt orthogonalization.

The vector g_{γ_m} selected by the matching algorithm is a priori not orthogonal to the previously selected vectors $\{g_{\gamma_p}\}_{0 \leq p < m}$. When subtracting the projection of $R^m f$ over g_{γ_m} the algorithm reintroduces new components in the directions of $\{g_{\gamma_p}\}_{0 \leq p < m}$. This is avoided by projecting the residues on an orthogonal family $\{u_p\}_{0 \leq p < m}$ computed from $\{g_{\gamma_p}\}_{0 \leq p < m}$.

Let us initialize $u_0 = g_{\gamma_0}$. For $m \geq 0$, an orthogonal matching pursuit selects

g_{γ_m} that satisfies

$$|\langle R^m f, g_{\gamma_m}\rangle| \geq \alpha \sup_{\gamma\in\Gamma} |\langle R^m f, g_\gamma\rangle|. \tag{9.108}$$

The Gram-Schmidt algorithm orthogonalizes g_{γ_m} with respect to $\{g_{\gamma_p}\}_{0\leq p<m}$ and defines

$$u_m = g_{\gamma_m} - \sum_{p=0}^{m-1} \frac{\langle g_{\gamma_m}, u_p\rangle}{\|u_p\|^2} u_p. \tag{9.109}$$

The residue $R^m f$ is projected on u_m instead of g_{γ_m}:

$$R^m f = \frac{\langle R^m f, u_m\rangle}{\|u_m\|^2} u_m + R^{m+1} f. \tag{9.110}$$

Summing this equation for $0 \leq m < k$ yields

$$\begin{aligned} f &= \sum_{m=0}^{k-1} \frac{\langle R^m f, u_m\rangle}{\|u_m\|^2} u_m + R^k f \\ &= P_{\mathbf{V}_k} f + R^k f, \end{aligned} \tag{9.111}$$

where $P_{\mathbf{V}_k}$ is the orthogonal projector on the space $\mathbf{V}_k$ generated by $\{u_m\}_{0\leq m<k}$. The Gram-Schmidt algorithm ensures that $\{g_{\gamma_m}\}_{0\leq m<k}$ is also a basis of $\mathbf{V}_k$. For any $k \geq 0$ the residue $R^k f$ is the component of f that is orthogonal to $\mathbf{V}_k$. For $m = k$ (9.109) implies that

$$\langle R^m f, u_m\rangle = \langle R^m f, g_{\gamma_m}\rangle. \tag{9.112}$$

Since $\mathbf{V}_k$ has dimension k there exists $M \leq N$ such that $f \in \mathbf{V}_M$, so $R^M f = 0$ and inserting (9.112) in (9.111) for $k = M$ yields

$$f = \sum_{m=0}^{M-1} \frac{\langle R^m f, g_{\gamma_m}\rangle}{\|u_m\|^2} u_m. \tag{9.113}$$

The convergence is obtained with a finite number M of iterations. This is a decomposition in a family of orthogonal vectors so

$$\|f\|^2 = \sum_{m=0}^{M-1} \frac{|\langle R^m f, g_{\gamma_m}\rangle|^2}{\|u_m\|^2}. \tag{9.114}$$

To expand f over the original dictionary vectors $\{g_{\gamma_m}\}_{0\leq m<M}$, we must perform a change of basis. The triangular Gram-Schmidt relations (9.109) are inverted to expand u_m in $\{g_{\gamma_p}\}_{0\leq p\leq m}$:

$$u_m = \sum_{p=0}^{m} b[p,m]\, g_{\gamma_p}. \tag{9.115}$$

Inserting this expression into (9.113) gives

$$f = \sum_{p=0}^{M-1} a[\gamma_p]\, g_{\gamma_p} \tag{9.116}$$

with

$$a[\gamma_p] = \sum_{m=p}^{M-1} b[p,m]\, \frac{\langle R^m f, g_{\gamma_m}\rangle}{\|u_m\|^2}.$$

During the first few iterations, the pursuit often selects nearly orthogonal vectors, so the Gram-Schmidt orthogonalization is not needed. The orthogonal and non-orthogonal pursuits are then nearly the same. When the number of iterations increases and gets close to N, the residues of an orthogonal pursuit have norms that decrease faster than for a non-orthogonal pursuit.

Figure 9.11(f) displays the wavelet packets selected by an orthogonal matching pursuit. A comparison with Figure 9.11(e) shows that the orthogonal and non-orthogonal pursuits selects nearly the same wavelet packets having a high amplitude inner product. These wavelet packets are selected during the first few iterations, and since they are nearly orthogonal the Gram-Schmidt orthogonalization does not modify much the pursuit. The difference between the two algorithms becomes significant when selected wavelet packet vectors have non-negligible inner products, which happens when the number of iterations is large.

The Gram-Schmidt summation (9.109) must be carefully implemented to avoid numerical instabilities [29]. Orthogonalizing M vectors requires $O(NM^2)$ operations. In wavelet packet, local cosine and Gabor dictionaries, M matching pursuit iterations are calculated with $O(MN\log_2 N)$ operations. For M large, the Gram-Schmidt orthogonalization increases very significantly the computational complexity of the pursuit. The non-orthogonal pursuit is thus more often used for large signals.

9.6 PROBLEMS

9.1. [1] Prove that for any $f \in \mathbf{L}^2[0,1]$, if $\|f\|_V < +\infty$ then $\|f\|_\infty < +\infty$. Verify that one can find an image $f \in \mathbf{L}^2[0,1]^2$ such that $\|f\|_V < +\infty$ and $\|f\|_\infty = +\infty$.

9.2. [1] Prove that if $f \in \mathbf{W}^s(\mathbb{R})$ with $s > p + 1/2$ then $f \in \mathbf{C}^p$.

9.3. [1] The family of discrete polynomials $\{p_k[n] = n^k\}_{0\le k<N}$ is a basis of $\mathbb{C}^N$.

(a) Implement in WAVELAB a Gram-Schmidt algorithm that orthogonalizes $\{p_k\}_{0\le k<N}$.

(b) Let f be a signal of size N. Compute the polynomial f_k of degree k which minimizes $\|f - f_k\|$. Perform numerical experiments on signals f that are uniformly smooth and piecewise smooth. Compare the approximation error with the error obtained by approximating f with the k lower frequency Fourier coefficients.

9.4. [1] If f has a finite total variation $\|f\|_V$ on $[0,1]$, prove that its linear approximation in a wavelet basis satisfies $\epsilon_l[M] = O(\|f\|_V^2 M^{-1})$ (Hint: use Theorem 9.6). Verify that $\epsilon_l[M] \sim \|f\|_V^2 M^{-1}$ if $f = C\,\mathbf{1}_{[0,1/2]}$.

9.5. [1] Let $f = C\mathbf{1}_\Omega$ where Ω is a subset of $[0,1]^2$ with a regular boundary $\partial\Omega$ of finite length $L > 0$. Prove that the linear approximation error in a wavelet basis satisfies $\epsilon_l[M] \sim \|f\|_V \|f\|_\infty M^{-1/2}$.

9.6. [2] Let $\alpha[M]$ be a decreasing sequence such that $\lim_{M\to+\infty} \alpha[M] = 0$. By using (9.43) prove that there exists a bounded variation function $f \in \mathbf{L}^2[0,1]^2$ such that $\epsilon_l[M] \geq \alpha[M]$ (the amplitude of f is not bounded).

9.7. [1] Consider a wavelet basis of $\mathbf{L}^2[0,1]$ constructed with wavelets having $q > s$ vanishing moments and which are $\mathbf{C}^q$. Construct functions $f \in \mathbf{W}^s[0,1]$ for which the linear and non-linear approximation errors in this basis are identical: $\epsilon_l[M] = \epsilon_n[M]$ for any $M \geq 0$.

9.8. [1] *Color images* A color pixel is represented by red, green and blue components (r,g,b), which are considered as orthogonal coordinates in a three dimensional color space. The red $r[n_1,n_2]$, green $g[n_1,n_2]$ and blue $b[n_1,n_2]$ image pixels are modeled as values taken by respectively three random variables R, G and B, that are the three coordinates of a color vector. Estimate numerically the 3 by 3 covariance matrix of this color random vector from several images and compute the Karhunen-Loève basis that diagonalizes it. Compare the color images reconstructed from the two Karhunen-Loève color channels of highest variance with a reconstruction from the red and green channels.

9.9. [1] Let us define $\|x\|_p = \left(\sum_{n=-\infty}^{+\infty} |x[n]|^p\right)^{1/p}$. Prove that $\|x\|_q \leq \|x\|_p$ if $q \geq p$.

9.10. [1] Let $f(t)$ be a piecewise polynomial signal of degree 3 defined on $[0,1]$, with K discontinuities. We denote by f_K and $\tilde{f}_K$ respectively the linear and non-linear approximations of f from K vectors chosen from a Daubechies wavelet basis of $\mathbf{L}^2[0,1]$, with $p+1$ vanishing moments.

(a) Give upper bounds as a function of K and p of $\|f - f_K\|$ and $\|f - \tilde{f}_K\|$.

(b) The Piece-Polynomial signal f in WAVELAB is piecewise polynomial with degree 3. Decompose it in a Daubechies wavelet basis with four vanishing moments, and compute $\|f - f_K\|$ and $\|f - \tilde{f}_K\|$ as a function of K. Verify your analytic formula.

9.11. [2] Let $f[n]$ be defined over $[0,N]$. We denote by $f_{p,k}[n]$ the signal that is piecewise constant on $[0,k]$, takes at most p different values, and minimizes

$$\epsilon_{p,k} = \|f - f_{p,k}\|^2_{[0,k]} = \sum_{n=0}^{k} |f[n] - f_{p,k}[n]|^2 .$$

(a) Compute as a function of $f[n]$ the value $a_{l,k}$ that minimizes $c_{l,k} = \sum_{n=l}^{k} |f[n] - a_{l,k}|^2$.

(b) Prove that

$$\epsilon_{p,k} = \min_{l\in[0,k-1]} \{\epsilon_{p-1,l} + c_{l,k}\}.$$

Derive a bottom up algorithm that computes progressively $f_{p,k}$ for $0 \leq k \leq N$ and $1 \leq p \leq K$, and obtains $f_{K,N}$ with $O(KN^2)$ operations. Implement this algorithm in WAVELAB.

(c) Compute the non-linear approximation of f with the K largest amplitude Haar wavelet coefficients, and the resulting approximation error. Compare

this error with $\|f - f_{K,N}\|$ as a function of K, for the Lady and the Piece-Polynomial signals in WAVELAB. Explain your results.

9.12. [2] *Approximation of oscillatory functions*

(a) Let $f(t) = a(t)\exp[i\phi(t)]$. If $a(t)$ and $\phi'(t)$ remain nearly constant on the support of $\psi_{j,n}$ then show with an approximate calculation that

$$\langle f, \psi_{j,n}\rangle \approx a(2^j n)\sqrt{2^j}\,\hat{\psi}\left(2^j \phi'(2^j n)\right) . \tag{9.117}$$

(b) Let $f(t) = \sin t^{-1} 1_{[-1/\pi,1/\pi]}(t)$. Show that the $\mathbf{l}^p$ norm of the wavelet coefficients of f is finite if and only if $p < 1$. Use the approximate formula (9.117).

(c) Compute an upper bound of the non-linear approximation error $\epsilon[M]$ of $\sin t^{-1}$ from M wavelet coefficients. Verify your theoretical estimate with a numerical calculation in WAVELAB.

9.13. [1] Let f be a signal of size N and T a given threshold. Describe a fast algorithm that searches in a wavelet packet or a local cosine dictionary for the best basis $\mathcal{B} = \{g_m\}_{0 \le m < N}$ that minimizes the number of inner products such that $|\langle f, g_m\rangle| \ge T$.

9.14. [1] *Best translated basis* Let $\{\psi_{j,m}[n]\}_{j,m}$ be a discrete wavelet orthonormal basis of signals of period N, computed with a conjugate mirror filter h with K non-zero coefficients. Let $\psi^k_{j,m}[n] = \psi_{j,m}[n-k]$ and $\mathcal{B}^k = \{\psi^k_{j,m}[n]\}_{j,m}$ be the translated basis, for any $0 \le k < N$.

(a) Describe an algorithm that decomposes f over all wavelets $\psi^k_{j,m}$ with $O(KN\log_2 N)$ operations.

(b) Let $C(f,\mathcal{B}^k) = \sum_{j,m} \Phi\left(|\langle f, \psi^k_{j,m}\rangle|^2/\|f\|^2\right)$. Describe an algorithm that finds the best shift l such that $C(f,\mathcal{B}^l) = \min_{0 \le k < N} C(f,\mathcal{B}^k)$, with $O(N\log_2 N)$ operations [281].

9.15. [1] *Best wavelet packet and local cosine approximations*

(a) Synthesize a discrete signal that is well approximated by few vectors in a best wavelet packet basis, but which requires many more vectors to obtain an equivalent approximation in a best local cosine basis. Test your signal in WAVELAB.

(b) Design a signal that is well approximated in a best local cosine basis but requires many more vectors to approximate it efficient in a best wavelet packet basis. Verify your result in WAVELAB.

9.16. [1] In two dimensions, a wavelet packet quad-tree of an image of size N^2 requires a storage of $N^2\log_2 N$ numbers. Describe an algorithm that finds the best wavelet packet basis with a storage of $4N^2/3$, by constructing the wavelet packet tree and computing the cost function in a depth-first preorder [76].

9.17. [2] A double tree of block wavelet packet bases is defined in Problem 8.11.

(a) Describe a fast best basis algorithm which requires $O(N(\log_2 N)^2)$ operations to find the block wavelet packet basis that minimizes an additive cost (9.71) [208].

(b) Implement the double tree decomposition and the best basis search in WAVELAB. Program a display that shows the time-frequency tiling of the

best basis and the amplitude of the decomposition coefficients. How does the best block wavelet packet basis compare with a best local cosine basis for the Greasy and Tweet signals?

9.18. [2] Let $\mathcal{D} = \{\delta[n-k], \exp(i2\pi kn/N)\}_{0\le k<N}$ be a Dirac-Fourier dictionary to decompose N periodic signals.

(a) Prove that a matching pursuit residue calculated with an optimality factor $\alpha = 1$ satisfies $\|R^m f\| \le \|f\| \exp(-m/(2N))$.

(b) Implement the matching pursuit in this Dirac-Fourier dictionary and decompose $f[n] = \exp(-i2\pi n^2/N)$. Compare the decay rate of the residue with the upper bound that was calculated. Suggest a better dictionary to decompose this signal.

9.19. [2] Let f be a piecewise constant image defined over $[0,N]^2$. Suppose that f is constant over regions $\{\Omega_i\}_{1\le k\le K}$ whose borders are differentiable curves with a bounded curvature. It may be discontinuous along the borders of the Ω_i. Prove that there exists $K > 0$ such that for any $M > 0$ one can construct f_M which is constant on the M triangles of a triangulation of $[0,N]^2$ and which satisfies $\|f - f_M\| \le K M^{-2}$. Design and implement in WAVELAB an algorithm which computes f_M for any piecewise constant function f. Compare the performance of your algorithm with an approximation with M vectors selected from a two-dimensional Haar wavelet basis.

9.20. [3] Let $\theta(t)$ be a cubic box spline centered at $t = 0$. We define a dictionary of N periodic cubic splines:

$$\mathcal{D} = \left\{ \theta_j[(n-k) \bmod N] \right\}_{0\le j\le \log_2 N, 0\le k<N},$$

where $\theta_j[n] = K_j \theta(2^{-j}n)$ for $j \ge 1$, and $\theta_0[n] = \delta[n]$.

(a) Implement a matching pursuit in this dictionary.

(b) Show that if $f[n] = \theta_j[n] + \theta_j[n-k]$ where k is on the order of 2^j, then the greediness of the matching pursuit may lead to a highly non-optimal decomposition. Explain why. Would a basis pursuit decomposition do better?

(c) If $f[n] \ge 0$, explain how to improve the matching pursuit by imposing that $R^m f[n] \ge 0$ for any $m \ge 0$.

X

ESTIMATIONS ARE APPROXIMATIONS

In a background noise of French conversations, it is easier to carry on a personal discussion in English. The estimation of signals in additive noise is similarly optimized by finding a representation that discriminates the signal from the noise.

An estimation is calculated by an operator that attenuates the noise while preserving the signal. Linear operators have long predominated because of their simplicity, despite their limited performance. It is possible to keep the simplicity while improving the performance with non-linearities in a sparse representation. Thresholding estimators are studied in wavelet and wavelet packet bases, where they are used to suppress additive noises and restore signals degraded by low-pass filters. Non-linear estimations from sparse representations are also studied for operators, with an application to power spectrum estimation.

Optimizing an estimator requires taking advantage of prior information. Bayes theory uses a probabilistic signal model to derive estimators that minimize the average risk. These models are often not available for complex signals such as natural images. An alternative is offered by the minimax approach, which only requires knowing a prior set where the signal is guaranteed to be. The quasi-minimax optimality of wavelet thresholding estimators is proved for piecewise regular signals and images.

10.1 BAYES VERSUS MINIMAX [2]

A signal $f[n]$ of size N is contaminated by the addition of a noise. This noise is modeled as the realization of a random process $W[n]$, whose probability distribution is known. The measured data are

$$X[n] = f[n] + W[n] .$$

The signal f is estimated by transforming the noisy data X with a *decision operator* D. The resulting estimator is

$$\tilde{F} = DX .$$

Our goal is to minimize the error of the estimation, which is measured by a *loss function*. For speech or images, the loss function should measure the audio and visual degradation, which is often difficult to model. A mean-square distance is certainly not a perfect model of perceptual degradations, but it is mathematically simple and sufficiently precise in most applications. Throughout this chapter, the loss function is thus chosen to be a square Euclidean norm. The risk of the estimator $\tilde{F}$ of f is the average loss, calculated with respect to the probability distribution of the noise W:

$$r(D,f) = \mathsf{E}\{\|f - DX\|^2\} . \tag{10.1}$$

The optimization of the decision operator D depends on prior information that is available about the signal. The Bayes framework supposes that we know the probability distribution of the signal and optimizes D to minimize the expected risk. The main difficulty is to acquire enough information to define this prior probability distribution, which is often not possible for complex signals. The minimax framework uses a simpler model which says that signals remain in a prior set Θ. The goal is then to minimize the maximum risk over Θ. Section 10.1.2 relates minimax and Bayes estimators through the minimax theorem.

10.1.1 Bayes Estimation

The Bayes principle supposes that signals f are realizations of a random vector F whose probability distribution π is known a priori. This probability distribution is called the *prior distribution*. The noisy data are thus rewritten

$$X[n] = F[n] + W[n] .$$

We suppose that the noise values $W[k]$ are independent from the signal $F[n]$ for any $0 \leq k, n < N$. The joint distribution of F and W is the product of the distributions of F and W. It specifies the conditional probability distribution of F given the observed data X, also called the *posterior distribution*. This posterior distribution can be used to construct a decision operator D that computes an estimation $\tilde{F} = DX$ of F from the data X.

The *Bayes risk* is the expected risk calculated with respect to the prior probability distribution π of the signal:

$$r(D,\pi) = \mathsf{E}_\pi\{r(D,F)\} .$$

By inserting (10.1), it can be rewritten as an expected value relative to the joint probability distribution of the signal and the noise:

$$r(D,\pi) = \mathsf{E}\{\|F-\tilde{F}\|^2\} = \sum_{n=0}^{N-1} \mathsf{E}\{|F[n]-\tilde{F}[n]|^2\}.$$

Let $\mathcal{O}_n$ be the set of all operators (linear and non-linear) from $\mathbb{C}^N$ to $\mathbb{C}^N$. Optimizing D yields the *minimum Bayes risk*:

$$r_n(\pi) = \inf_{D\in\mathcal{O}_n} r(D,\pi)\ .$$

The following theorem proves that there exist a *Bayes decision operator* D and a corresponding *Bayes estimator* $\tilde{F}$ that achieve this minimum risk.

Theorem 10.1 *The Bayes estimator $\tilde{F}$ that yields the minimum Bayes risk $r_n(\pi)$ is the conditional expectation*

$$\tilde{F}[n] = \mathsf{E}\{F[n] \mid X[0],X[1],\ldots,X[N-1]\}. \tag{10.2}$$

Proof[2]. Let $\pi_n(y)$ be the probability distribution of the value y of $F[n]$. The minimum risk is obtained by finding $\tilde{F}[n] = D_n(X)$ that minimizes $r(D_n,\pi_n) = \mathsf{E}\{|F[n]-\tilde{F}[n]|^2\}$, for each $0 \le n < N$. This risk depends on the conditional distribution $P_n(x|y)$ of the data $X = x$, given $F[n] = y$:

$$r(D_n,\pi_n) = \int\int (D_n(x)-y)^2\, dP_n(x|y)\, d\pi_n(y)\ .$$

Let $P(x) = \int P_n(x|y)\, d\pi_n(y)$ be the marginal distribution of X and $\pi_n(y|x)$ be the posterior distribution of $F[n]$ given X. The Bayes formula gives

$$r(D_n,\pi_n) = \int \left[\int (D_n(x)-y)^2\, d\pi_n(y|x)\right] dP(x)\ .$$

The double integral is minimized by minimizing the inside integral for each x. This quadratic form is minimum when its derivative vanishes:

$$\frac{\partial}{\partial D_n(x)} \int (D_n(x)-y)^2\, d\pi_n(y|x) = 2\int (D_n(x)-y)\, d\pi_n(y|x) = 0$$

which implies that

$$D_n(x) = \int y\, d\pi_n(y|x) = \mathsf{E}\{F[n] \mid X = x\}\ ,$$

so $D_n(X) = \mathsf{E}\{F[n] \mid X\}$. ■

Linear Estimation The conditional expectation (10.2) is generally a complicated non-linear function of the data $\{X[k]\}_{0\leq k<N}$, and is difficult to evaluate. To simplify this problem, we restrict the decision operator D to be linear. Let $\mathcal{O}_l$ be the set of all linear operators from $\mathbb{C}^N$ to $\mathbb{C}^N$. The *linear minimum Bayes risk* is:

$$r_l(\pi) = \inf_{D\in\mathcal{O}_l} r(D,\pi) \ .$$

The linear estimator $\tilde{F} = DF$ that achieves this minimum risk is called the *Wiener estimator*. The following proposition gives a necessary and sufficient condition that specifies this estimator. We suppose that $\mathsf{E}\{F[n]\} = 0$, which can be enforced by subtracting $\mathsf{E}\{F[n]\}$ from $X[n]$ to obtain a zero-mean signal.

Proposition 10.1 *A linear estimator $\tilde{F}$ is a Wiener estimator if and only if*

$$\mathsf{E}\{(F[n]-\tilde{F}[n])X[k]\} = 0 \ \textit{for}\ 0\leq k,n<N \ . \tag{10.3}$$

Proof[2]. For each $0\leq n<N$, we must find a linear estimation

$$\tilde{F}[n] = D_nX = \sum_{k=0}^{N-1} h[n,k]X[k]$$

which minimizes

$$r(D_n,\pi_n) = \mathsf{E}\left\{\left(F[n]-\sum_{k=0}^{N-1}h[n,k]X[k]\right)\left(F[n]-\sum_{k=0}^{N-1}h[n,k]X[k]\right)\right\}. \tag{10.4}$$

The minimum of this quadratic form is reached if and only if for each $0\leq k<N$,

$$\frac{\partial r(D_n,\pi_n)}{\partial h[n,k]} = -2\,\mathsf{E}\left\{\left(F[n]-\sum_{l=0}^{N-1}h[n,l]X[l]\right)X[k]\right\} = 0,$$

which verifies (10.3). ■

If F and W are independent Gaussian random vectors, then the linear optimal estimator is also optimal among non-linear estimators. Indeed, two jointly Gaussian random vectors are independent if they are non-correlated [56]. Since $F[n]-\tilde{F}[n]$ is jointly Gaussian with $X[k]$, the non-correlation (10.3) implies that $F[n]-\tilde{F}[n]$ and $X[k]$ are independent for any $0\leq k,n<N$. In this case, we can verify that $\tilde{F}$ is the Bayes estimator (10.2): $\tilde{F}[n] = \mathsf{E}\{F[n] \mid X\}$.

Estimation in a Karhunen-Loève Basis The following theorem proves that if the covariance matrices of the signal F and of the noise W are diagonal in the same Karhunen-Loève basis $\mathcal{B} = \{g_m\}_{0\leq m<N}$ then the optimal linear estimator is diagonal in this basis. We write

$$X_{\mathcal{B}}[m] = \langle X, g_m\rangle \ , \ \tilde{F}_{\mathcal{B}}[m] = \langle \tilde{F}, g_m\rangle \ ,$$
$$F_{\mathcal{B}}[m] = \langle F, g_m\rangle \ , \ W_{\mathcal{B}}[m] = \langle W, g_m\rangle \ ,$$
$$\beta_m^2 = \mathsf{E}\{|F_{\mathcal{B}}[m]|^2\} \ , \ \sigma_m^2 = \mathsf{E}\{|W_{\mathcal{B}}[m]|^2\} \ .$$

Theorem 10.2 (WIENER) *If there exists a Karhunen-Loève basis $\mathcal{B} = \{g_m\}_{0 \le m < N}$ that diagonalizes the covariance matrices of both F and W, then the Wiener estimator is*

$$\tilde{F} = \sum_{m=0}^{N-1} \frac{\beta_m^2}{\beta_m^2 + \sigma_m^2} X_{\mathcal{B}}[m]\, g_m \tag{10.5}$$

and the resulting minimum linear Bayes risk is

$$r_l(\pi) = \sum_{m=0}^{N-1} \frac{\beta_m^2 \sigma_m^2}{\beta_m^2 + \sigma_m^2}. \tag{10.6}$$

Proof[2]. Let $\tilde{F}[n]$ be a linear estimator of $F[n]$:

$$\tilde{F}[n] = \sum_{l=0}^{N-1} h[n,l]\, X[l]. \tag{10.7}$$

This equation can be rewritten as a matrix multiplication by introducing the $N \times N$ matrix $H = (h[n,l])_{0 \le n,l < N}$:

$$\tilde{F} = HX. \tag{10.8}$$

The non-correlation condition (10.3) implies that for $0 \le n,k < N$

$$\mathsf{E}\{F[n]X[k]\} = \mathsf{E}\{\tilde{F}[n]X[k]\} = \sum_{l=0}^{N-1} h[n,l]\, \mathsf{E}\{X[l]X[k]\}.$$

Since $X[k] = F[k] + W[k]$ and $\mathsf{E}\{F[n]\,W[k]\} = 0$, we derive that

$$\mathsf{E}\{F[n]\,F[k]\} = \sum_{l=0}^{N-1} h[n,l] \left(\mathsf{E}\{F[l]\,F[k]\} + \mathsf{E}\{W[l]\,W[k]\} \right). \tag{10.9}$$

Let R_F and R_W be the covariance matrices of F and W, whose entries are respectively $\mathsf{E}\{F[n]\,F[k]\}$ and $\mathsf{E}\{W[n]\,W[k]\}$. Equation (10.9) can be rewritten as a matrix equation:

$$R_F = H\,(R_F + R_W).$$

Inverting this equation gives

$$H = R_F\,(R_F + R_W)^{-1}.$$

Since R_F and R_W are diagonal in the basis $\mathcal{B}$ with diagonal values respectively equal to β_m^2 and σ_m^2, the matrix H is also diagonal in $\mathcal{B}$ with diagonal values equal to $\beta_m^2(\beta_m^2 + \sigma_m^2)^{-1}$. So (10.8) shows that the decomposition coefficients of $\tilde{F}$ and X in $\mathcal{B}$ satisfy

$$\tilde{F}_{\mathcal{B}}[m] = \frac{\beta_m^2}{\beta_m^2 + \sigma_m^2} X_{\mathcal{B}}[m], \tag{10.10}$$

which implies (10.5).

The resulting risk is

$$\mathsf{E}\{\|F - \tilde{F}\|^2\} = \sum_{m=0}^{N-1} \mathsf{E}\left\{ |F_{\mathcal{B}}[m] - \tilde{F}_{\mathcal{B}}[m]|^2 \right\}. \tag{10.11}$$

Inserting (10.10) in (10.11) knowing that $X_{\mathcal{B}}[m] = F_{\mathcal{B}}[m] + W_{\mathcal{B}}[m]$ where $F_{\mathcal{B}}[m]$ and $W_{\mathcal{B}}[m]$ are independent yields (10.6). ■

This theorem proves that the Wiener estimator is implemented with a diagonal attenuation of each data coefficient $X_B[m]$ by a factor that depends on the signal to noise ratio β_m^2/σ_m^2 in the direction of g_m. The smaller the signal to noise ratio, the more attenuation is required. If F and W are Gaussian processes, then the Wiener estimator is optimal among linear and non-linear estimators of F.

If W is a white noise then its coefficients are uncorrelated with the same variance

$$\mathsf{E}\{W[n]\,W[k]\} = \sigma^2\,\delta[n-k]\ .$$

Its covariance matrix is therefore $R_W = \sigma^2 Id$. It is diagonal in all orthonormal bases and in particular in the Karhunen-Loève basis of F. Theorem 10.2 can thus be applied and $\sigma_m = \sigma$ for $0 \leq m < N$.

Frequency Filtering Suppose that F and W are zero-mean, wide-sense circular stationary random vectors. The properties of such processes are reviewed in Appendix A.6. Their covariance satisfies

$$\mathsf{E}\{F[n]\,F[k]\} = R_F[n-k]\ ,\ \ \mathsf{E}\{W[n]\,W[k]\} = R_W[n-k]\ ,$$

where $R_F[n]$ and $R_W[n]$ are N periodic. These matrices correspond to circular convolution operators and are therefore diagonal in the discrete Fourier basis

$$\left\{ g_m[n] = \frac{1}{\sqrt{N}} \exp\left(\frac{i2m\pi n}{N}\right) \right\}_{0 \leq m < N}\ .$$

The eigenvalues β_m^2 and σ_m^2 are the discrete Fourier transforms of $R_F[n]$ and $R_W[n]$, also called *power spectra*:

$$\beta_m^2 = \sum_{n=0}^{N-1} R_F[n] \exp\left(\frac{-i2m\pi n}{N}\right) = \hat{R}_F[m]\ ,$$

$$\sigma_m^2 = \sum_{n=0}^{N-1} R_W[n] \exp\left(\frac{-i2m\pi n}{N}\right) = \hat{R}_W[m].$$

The Wiener estimator (10.5) is a diagonal operator in the discrete Fourier basis, computed with the frequency filter:

$$\hat{h}[m] = \frac{\hat{R}_F[m]}{\hat{R}_F[m] + \hat{R}_W[m]}\ . \tag{10.12}$$

It is therefore a circular convolution:

$$\tilde{F}[n] = DX = X \circledast h[n].$$

The resulting risk is calculated with (10.6):

$$r_l(\pi) = \mathsf{E}\{\|F - \tilde{F}\|^2\} = \sum_{m=0}^{N-1} \frac{\hat{R}_F[m]\,\hat{R}_W[m]}{\hat{R}_F[m] + \hat{R}_W[m]}\ . \tag{10.13}$$

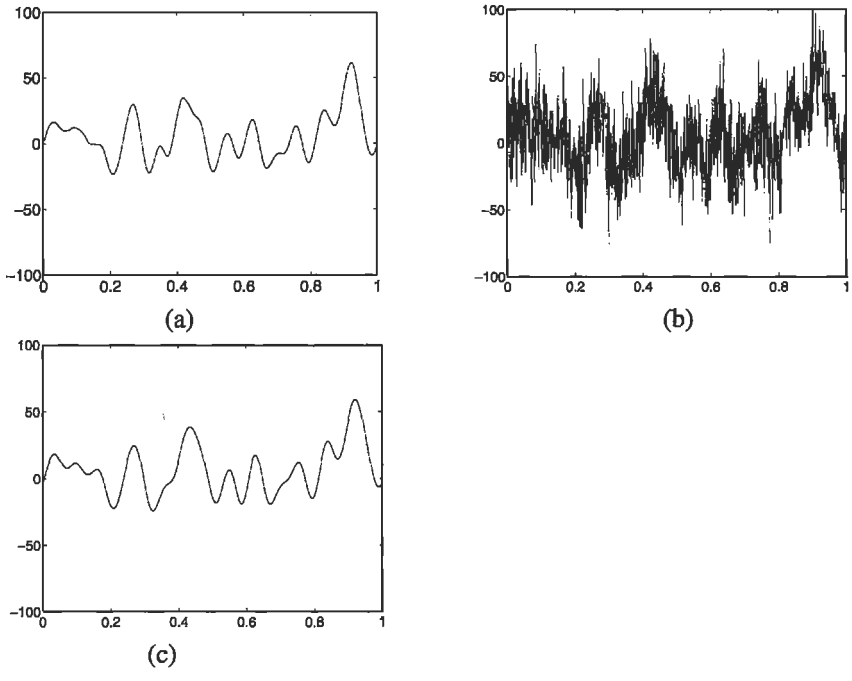


FIGURE 10.1 (a): Realization of a Gaussian process F. (b): Noisy signal obtained by adding a Gaussian white noise (SNR $= -0.48$ db). (c): Wiener estimation $\tilde{F}$ (SNR $= 15.2$ db).

The numerical value of the risk is often specified by the *Signal to Noise Ratio*, which is measured in decibels

$$SNR_{\text{db}} = 10 \log_{10} \left(\frac{\mathsf{E}\{\|F\|^2\}}{\mathsf{E}\{\|F - \tilde{F}\|^2\}} \right). \tag{10.14}$$

Example 10.1 Figure 10.1(a) shows a realization of a Gaussian process F obtained as a convolution of a Gaussian white noise B of variance β^2 with a low-pass filter g:

$$F[n] = B \circledast g[n],$$

with

$$g[n] = C \cos^2 \left(\frac{\pi n}{2K} \right) \mathbf{1}_{[-K,K]}[n].$$

Theorem A.4 proves that

$$\hat{R}_F[m] = \hat{R}_B[m] \, |\hat{g}[m]|^2 = \beta^2 \, |\hat{g}[m]|^2.$$

The noisy signal X shown in Figure 10.1(b) is contaminated by a Gaussian white noise W of variance σ^2, so $\hat{R}_W[m] = \sigma^2$. The Wiener estimation $\tilde{F}$ is calculated with the frequency filter (10.12)

$$\hat{h}[m] = \frac{\beta^2 |\hat{g}[m]|^2}{\beta^2 |\hat{g}[m]|^2 + \sigma^2}.$$

This linear estimator is also an optimal non-linear estimator because F and W are jointly Gaussian random vectors.

Piecewise Regular The limitations of linear estimators appear clearly for processes whose realizations are piecewise regular signals. A simple example is a random shift process F constructed by translating randomly a piecewise regular signal $f[n]$ of zero mean, $\sum_{n=0}^{N-1} f[n] = 0$:

$$F[n] = f[(n-P) \bmod N] \,. \tag{10.15}$$

The shift P is an integer random variable whose probability distribution is uniform on $[0, N-1]$. It is proved in (9.20) that F is a circular wide-sense stationary process whose power spectrum is calculated in (9.21):

$$\hat{R}_F[m] = \frac{1}{N} |\hat{f}[m]|^2. \tag{10.16}$$

Figure 10.2 shows an example of a piecewise polynomial signal f of degree $d = 3$ contaminated by a Gaussian white noise W of variance σ^2. Assuming that we know $|\hat{f}[m]|^2$, the Wiener estimator $\tilde{F}$ is calculated as a circular convolution with the filter whose transfer function is (10.12). This Wiener filter is a low-pass filter that averages the noisy data to attenuate the noise in regions where the realization of F is regular, but this averaging is limited to avoid degrading the discontinuities too much. As a result, some noise is left in the smooth regions and the discontinuities are averaged a little. The risk calculated in (10.13) is normalized by the total noise energy $\mathrm{E}\{\|W\|^2\} = N\sigma^2$:

$$\frac{r_l(\pi)}{N\sigma^2} = \sum_{m=0}^{N-1} \frac{N^{-1} |\hat{f}[m]|^2}{|\hat{f}[m]|^2 + N\sigma^2} \,. \tag{10.17}$$

Suppose that f has discontinuities of amplitude on the order of $C \geq \sigma$ and that the noise energy is not negligible: $N\sigma^2 \geq C^2$. Using the fact that $|\hat{f}[m]|$ decays typically like CNm^{-1}, a direct calculation of the risk (10.17) gives

$$\frac{r_l(\pi)}{N\sigma^2} \sim \frac{C}{\sigma N^{1/2}} \,. \tag{10.18}$$

The equivalence $\sim$ means that upper and lower bounds of the left-hand side are obtained by multiplying the right-hand side by two constants $A, B > 0$ that are independent of C, σ and N.

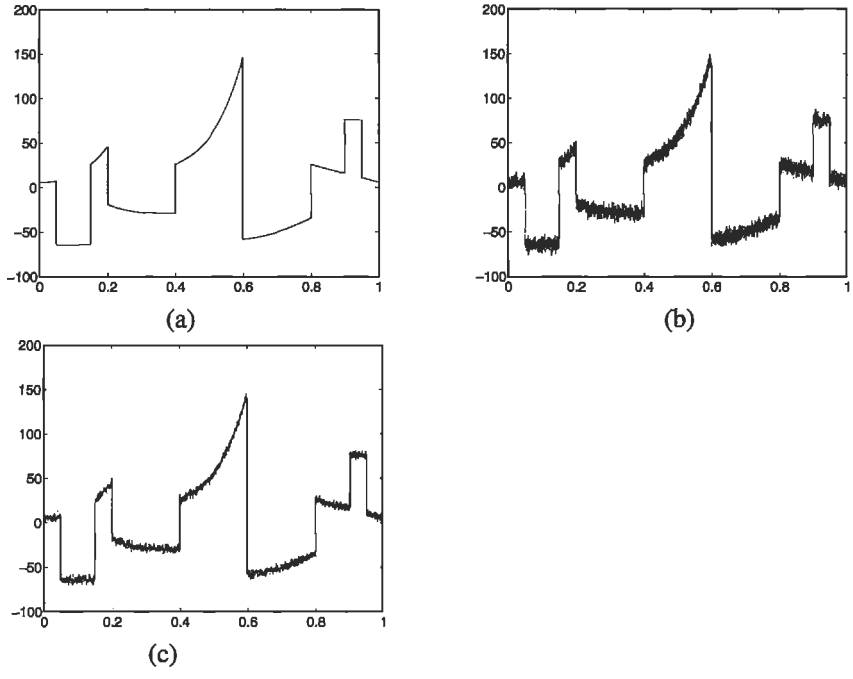


FIGURE 10.2 (a): Piecewise polynomial of degree 3. (b): Noisy signal degraded by a Gaussian white noise (SNR = 21.9 db). (c): Wiener estimation (SNR= 25.9 db).

The estimation of F can be improved by non-linear operators, which average the data X over large domains where F is regular but do not make any averaging where F is discontinuous. Many estimators have been studied [183, 276] that estimate the position of the discontinuities of f in order to adapt the data averaging. These algorithms have long remained *ad hoc* implementations of intuitively appealing ideas. Wavelet thresholding estimators perform such an adaptive smoothing and Section 10.3.3 proves that the normalized risk decays like N^{-1} as opposed to $N^{-1/2}$ in (10.18).

10.1.2 Minimax Estimation

Although we may have some prior information, it is rare that we know the probability distribution of complex signals. This prior information often defines a set Θ to which signals are guaranteed to belong, without specifying their probability distribution in Θ. The more prior information, the smaller the set Θ. For example, we may know that a signal has at most K discontinuities, with bounded derivatives outside these discontinuities. This defines a particular prior set Θ. Presently, there exists no stochastic model that takes into account the diversity of natural images.

However, many images, such as the one in Figure 2.2, have some form of piecewise regularity, with a bounded total variation. This also specifies a prior set Θ.

The problem is to estimate $f \in \Theta$ from the noisy data

$$X[n] = f[n] + W[n] .$$

The risk of an estimation $\tilde{F} = DX$ is $r(D, f) = \mathsf{E}\{\|DX - f\|^2\}$. The expected risk over Θ cannot be computed because we do not know the probability distribution of signals in Θ. To control the risk for any $f \in \Theta$, we thus try to minimize the maximum risk:

$$r(D, \Theta) = \sup_{f \in \Theta} \mathsf{E}\{\|DX - f\|^2\}.$$

The *minimax risk* is the lower bound computed over all linear and non-linear operators D:

$$r_n(\Theta) = \inf_{D \in \mathcal{O}_n} r(D, \Theta).$$

In practice, we must find a decision operator D that is simple to implement and such that $r(D, \Theta)$ is close to the minimax risk $r_n(\Theta)$.

As a first step, as for Wiener estimators in the Bayes framework, we can simplify the problem by restricting D to be a linear operator. The *linear minimax risk* over Θ is the lower bound:

$$r_l(\Theta) = \inf_{D \in \mathcal{O}_l} r(D, \Theta).$$

This strategy is efficient only if $r_l(\Theta)$ is of the same order as $r_n(\Theta)$.

Bayes Priors A Bayes estimator supposes that we know the prior probability distribution π of signals in Θ. If available, this supplement of information can only improve the signal estimation. The central result of game and decision theory shows that minimax estimations are Bayes estimations for a "least favorable" prior distribution.

Let F be the signal random vector, whose probability distribution is given by the prior π. For a decision operator D, the expected risk is $r(D, \pi) = \mathsf{E}_\pi\{r(D, F)\}$. The minimum Bayes risks for linear and non-linear operators are defined by:

$$r_l(\pi) = \inf_{D \in \mathcal{O}_l} r(D, \pi) \text{ and } r_n(\pi) = \inf_{D \in \mathcal{O}_n} r(D, \pi) .$$

Let Θ^* be the set of all probability distributions of random vectors whose realizations are in Θ. The minimax theorem relates a minimax risk and the maximum Bayes risk calculated for priors in Θ^*.

Theorem 10.3 (MINIMAX) *For any subset Θ of $\mathbb{C}^N$*

$$r_l(\Theta) = \sup_{\pi \in \Theta^*} r_l(\pi) \text{ and } r_n(\Theta) = \sup_{\pi \in \Theta^*} r_n(\pi) . \tag{10.19}$$

Proof[2]. For any $\pi \in \Theta^*$

$$r(D,\pi) \leq r(D,\Theta) \tag{10.20}$$

because $r(D,\pi)$ is an average risk over realizations of F that are in Θ, whereas $r(D,\Theta)$ is the maximum risk over Θ. Let $\mathcal{O}$ be a convex set of operators (either $\mathcal{O}_l$ or $\mathcal{O}_n$). The inequality (10.20) implies that

$$\sup_{\pi\in\Theta^*} r(\pi) = \sup_{\pi\in\Theta^*} \inf_{D\in\mathcal{O}} r(D,\pi) \leq \inf_{D\in\mathcal{O}} r(D,\Theta) = r(\Theta)\,. \tag{10.21}$$

The main difficulty is to prove the reverse inequality: $r(\Theta) \leq \sup_{\pi\in\Theta^*} r(\pi)$. When Θ is a finite set, the proof gives an important geometrical interpretation of the minimum Bayes risk and the minimax risk. The extension to an infinite set Θ is sketched.

Suppose that $\Theta = \{f_i\}_{1\leq i\leq p}$ is a finite set of signals. We define a risk set:

$$R = \{(y_1,\ldots,y_p) \in \mathbb{C}^p \;:\; \exists D \in \mathcal{O} \text{ with } y_i = r(D,f_i) \text{ for } 1 \leq i \leq p\}\,.$$

This set is convex in $\mathbb{C}^p$ because $\mathcal{O}$ is convex. We begin by giving geometrical interpretations to the Bayes risk and the minimax risk.

A prior $\pi \in \Theta^*$ is a vector of discrete probabilities $(\pi_1,\ldots,\pi_p)$ and

$$r(\pi,D) = \sum_{i=1}^{p} \pi_i\, r(D,f_i)\,. \tag{10.22}$$

The equation $\sum_{i=1}^{p} \pi_i y_i = b$ defines a hyperplane P_b in $\mathbb{C}^p$. Computing $r(\pi) = \inf_{D\in\mathcal{O}} r(D,\pi)$ is equivalent to finding the infimum $b_0 = r(\pi)$ of all b for which P_b intersects R. The plane P_{b_0} is tangent to R as shown in Figure 10.3.

The minimax risk $r(\Theta)$ has a different geometrical interpretation. Let $Q_c = \{(y_1,\ldots,y_p) \in \mathbb{C}^p \;:\; y_i \leq c\}$ One can verify that $r(\Theta) = \inf_{D\in\mathcal{O}} \sup_{f_i\in\Theta} r(D,f_i)$ is the infimum $c_0 = r(\Theta)$ of all c such that Q_c intersects R.

To prove that $r(\Theta) \leq \sup_{\pi\in\Theta^*} r(\pi)$ we look for a prior distribution $\tau \in \Theta^*$ such that $r(\tau) = r(\Theta)$. Let $\tilde{Q}_{c_0}$ be the interior of Q_{c_0}. Since $\tilde{Q}_{c_0} \cap R = \emptyset$ and both $\tilde{Q}_{c_0}$ and R are convex sets, the hyperplane separation theorem says that there exists a hyperplane of equation

$$\sum_{i=1}^{p} \tau_i y_i = \tau \,.\, y = b\,, \tag{10.23}$$

with $\tau \,.\, y \leq b$ for $y \in \tilde{Q}_{c_0}$ and $\tau \,.\, y \geq b$ for $y \in R$. Each $\tau_i \geq 0$, for if $\tau_j < 0$ then for $y \in \tilde{Q}_{c_0}$ we obtain a contradiction by taking y_j to $-\infty$ with the other coordinates being fixed. Indeed, $\tau \,.\, y$ goes to $+\infty$ and since y remains in $\tilde{Q}_{c_0}$ it contradicts the fact that $\tau \,.\, y \leq b$. We can normalize $\sum_{i=1}^{p} \tau_i = 1$ by dividing each side of (10.23) by $\sum_{i=1}^{p} \tau_i > 0$. So τ corresponds to a probability distribution. By letting $y \in \tilde{Q}_{c_0}$ converge to the corner point $(c_0,\ldots,c_0)$, since $y \,.\, \tau \leq b$ we derive that $c_0 \leq b$. Moreover, since $\tau \,.\, y \geq b$ for all $y \in R$,

$$r(\tau) = \inf_{D\in\mathcal{O}} \sum_{i=1}^{p} \tau_i\, r(D,f_i) \geq c \geq c_0 = r(\Theta)\,.$$

So $r(\Theta) \leq \sup_{\pi\in\Theta^*} r(\pi)$ which, together with (10.21), proves that $r(\Theta) = \sup_{\pi\in\Theta^*} r(\pi)$.

The extension of this result to an infinite set of signals Θ is done with a compactness argument. When $\mathcal{O} = \mathcal{O}_l$ or $\mathcal{O} = \mathcal{O}_n$, for any prior $\pi \in \Theta^*$ we know from Theorem 10.1

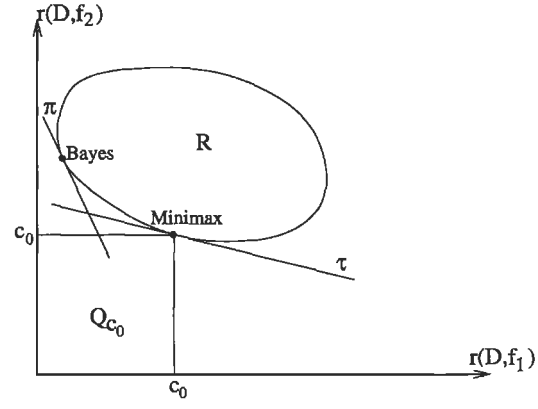


FIGURE 10.3 At the Bayes point, a hyperplane defined by the prior π is tangent to the risk set R. The least favorable prior τ defines a hyperplane that is tangential to R at the minimax point.

and Proposition 10.1 that $\inf_{D\in\mathcal{O}} r(D,\pi)$ is reached by some Bayes decision operator $D \in \mathcal{O}$. One can verify that there exists a subset of operators $\mathcal{C}$ that includes the Bayes operator for any prior $\pi \in \Theta^*$, and such that $\mathcal{C}$ is compact for an appropriate topology. When $\mathcal{O} = \mathcal{O}_l$, one can choose $\mathcal{C}$ to be the set of linear operators of norm smaller than 1, which is compact because it belongs to a finite dimensional space of linear operators. Moreover, the risk $r(f,D)$ can be shown to be continuous in this topology with respect to $D \in \mathcal{C}$.

Let $c < r(\Theta)$. For any $f \in \Theta$ we consider the set of operators $\mathcal{S}_f = \{D \in \mathcal{C} : r(D,f) > c\}$. The continuity of r implies that $\mathcal{S}_f$ is an open set. For each $D \in \mathcal{C}$ there exists $f \in \Theta$ such that $D \in \mathcal{S}_f$, so $\mathcal{C} = \cup_{f\in\Theta}\mathcal{S}_f$. Since $\mathcal{C}$ is compact there exists a finite subcovering $\mathcal{C} = \cup_{1\le i\le p}\mathcal{S}_{f_i}$. The minimax risk over $\Theta_c = \{f_i\}_{1\le i\le p}$ satisfies

$$r(\Theta_c) = \inf_{D\in\mathcal{O}} \sup_{1\le i\le p} r(D,f_i) \ge c\ .$$

Since Θ_c is a finite set, we proved that there exists $\tau_c \in \Theta_c^* \subset \Theta^*$ such that $r(\tau_c) = r(\Theta_c)$. But $r(\Theta_c) \ge c$ so letting c go to $r(\Theta)$ implies that $\sup_{\pi\in\Theta^*} r(\pi) \ge r(\Theta)$. Together with (10.21) this shows that $\inf_{\tau\in\Theta^*} r(\tau) = r(\Theta)$. ■

A distribution $\tau \in \Theta^*$ such that $r(\tau) = \inf_{\pi\in\Theta^*} r(\pi)$ is called a *least favorable* prior distribution. The minimax theorem proves that the minimax risk is the minimum Bayes risk for a least favorable prior.

In signal processing, minimax calculations are often hidden behind apparently orthodox Bayes estimations. Let us consider an example involving images. It has been observed that histograms of the wavelet coefficients of "natural" images can be modeled with generalized Gaussian distributions [255, 311]. This means that natural images belong to a certain set Θ, but it does not specify a prior distribution over this set. To compensate for the lack of knowledge about the dependency of wavelet coefficients spatially and across scales, one may be tempted to create a

"simple probabilistic model" where all wavelet coefficients are considered to be independent. This model is clearly wrong since images have geometrical structures that create strong dependencies both spatially and across scales (see Figure 7.26). However, calculating a Bayes estimator with this inaccurate prior model may give valuable results when estimating images. Why? Because this "simple" prior is often close to a least favorable prior. The resulting estimator and risk are thus good approximations of the minimax optimum. If not chosen carefully, a "simple" prior may yield an optimistic risk evaluation that is not valid for real signals. Understanding the robustness of uncertain priors is what minimax calculations are often about.

10.2 DIAGONAL ESTIMATION IN A BASIS [2]

It is generally not possible to compute the optimal Bayes or minimax estimator that minimizes the risk among all possible operators. To manage this complexity, the most classical strategy limits the choice of operators among linear operators. This comes at a cost, because the minimum risk among linear estimators may be well above the minimum risk obtained with non-linear estimators. Figure 10.2 is an example where the linear Wiener estimation can be considerably improved with a non-linear averaging. This section studies a particular class of non-linear estimators that are diagonal in a basis $\mathcal{B}$. If the basis $\mathcal{B}$ defines a sparse signal representation, then such diagonal estimators are nearly optimal among all non-linear estimators.

Section 10.2.1 computes a lower bound for the risk when estimating an arbitrary signal f with a diagonal operator. Donoho and Johnstone [167] made a fundamental breakthrough by showing that thresholding estimators have a risk that is close to this lower bound. The general properties of thresholding estimators are introduced in Sections 10.2.2 and 10.2.3. Thresholding estimators in wavelet bases are studied in Section 10.2.4. They implement an adaptive signal averaging that is much more efficient than linear operators to estimate piecewise regular signals. Section 10.2.5 ends this section by explaining how to search for a best basis that minimizes the risk. The minimax optimality of diagonal operators for estimating signals in a prior set Θ is studied in Section 10.3.

10.2.1 Diagonal Estimation with Oracles

We consider estimators computed with a diagonal operator in an orthonormal basis $\mathcal{B} = \{g_m\}_{0 \leq m < N}$. Lower bounds for the risk are computed with "oracles," which simplify the estimation by providing information about the signal that is normally not available. These lower bounds are closely related to errors when approximating signals from a few vectors selected in $\mathcal{B}$.

The noisy data

$$X = f + W \tag{10.24}$$

is decomposed in $\mathcal{B}$. We write

$$X_{\mathcal{B}}[m] = \langle X, g_m \rangle \ , \ f_{\mathcal{B}}[m] = \langle f, g_m \rangle \ \text{ and } \ W_{\mathcal{B}}[m] = \langle W, g_m \rangle \ .$$

The inner product of (10.24) with g_m gives

$$X_{\mathcal{B}}[m] = f_{\mathcal{B}}[m] + W_{\mathcal{B}}[m].$$

We suppose that W is a zero-mean *white noise* of variance σ^2, which means

$$\mathsf{E}\{W[n]\,W[k]\} = \sigma^2\,\delta[n-k]\,.$$

The noise coefficients

$$W_{\mathcal{B}}[m] = \sum_{n=0}^{N-1} W[n]\,g_m^*[n]$$

also define a white noise of variance σ^2. Indeed,

$$\begin{aligned}\mathsf{E}\{W_{\mathcal{B}}[m]\,W_{\mathcal{B}}[p]\} &= \sum_{n=0}^{N-1}\sum_{k=0}^{N-1} g_m[n]\,g_p[k]\,\mathsf{E}\{W[n]\,W[k]\}\\ &= \sigma^2\,\langle g_p, g_m\rangle = \sigma^2\,\delta[p-m].\end{aligned}$$

Since the noise remains white in all bases, it does not influence the choice of basis. When the noise is not white, which is the case for the inverse problems of Section 10.4, the noise can have an important impact on the basis choice.

A *diagonal operator* estimates independently each $f_{\mathcal{B}}[m]$ from $X_{\mathcal{B}}[m]$ with a function $d_m(x)$. The resulting estimator is

$$\tilde{F} = DX = \sum_{m=0}^{N-1} d_m(X_{\mathcal{B}}[m])\,g_m\,. \tag{10.25}$$

The class of signals that are considered is supposed to be centered at 0, so we set $D0 = 0$ and hence $d_m(0) = 0$. As a result, we can write

$$d_m(X_{\mathcal{B}}[m]) = a[m]\,X_{\mathcal{B}}[m] \quad \text{for } 0 \le m < N,$$

where $a[m]$ depends on $X_{\mathcal{B}}[m]$. The operator D is linear when $a[m]$ is a constant independent of $X_{\mathcal{B}}[m]$. We shall see that a smaller risk is obtained with $|a[m]| \le 1$, which means that the diagonal operator D attenuates the noisy coefficients.

Attenuation With Oracle Let us find the $a[m]$ that minimizes the risk $r(D,f)$ of the estimator (10.25):

$$r(D,f) = \mathsf{E}\left\{\|f-\tilde{F}\|^2\right\} = \sum_{m=0}^{N-1} \mathsf{E}\{|f_{\mathcal{B}}[m] - X_{\mathcal{B}}[m]\,a[m]|^2\}\,. \tag{10.26}$$

Since $X_{\mathcal{B}} = f_{\mathcal{B}} + W_{\mathcal{B}}$ and $\mathsf{E}\{|W_{\mathcal{B}}[m]|^2\} = \sigma^2$ it follows that

$$\mathsf{E}\{|f_{\mathcal{B}}[m] - X_{\mathcal{B}}[m]\,a[m]|^2\} = |f_{\mathcal{B}}[m]|^2\,(1-a[m])^2 + \sigma^2\,a[m]^2. \tag{10.27}$$

This risk is minimum for

$$a[m] = \frac{|f_B[m]|^2}{|f_B[m]|^2 + \sigma^2}, \tag{10.28}$$

in which case

$$r_{\inf}(f) = \mathsf{E}\{\|f - \tilde{F}\|^2\} = \sum_{m=0}^{N-1} \frac{|f_B[m]|^2 \sigma^2}{|f_B[m]|^2 + \sigma^2}. \tag{10.29}$$

In practice, the attenuation factor $a[m]$ in (10.28) cannot be computed since it depends on $|f_B[m]|$, whose value is not known. The risk $r_{\inf}(f)$ is therefore a lower bound which is not reachable. This risk is obtained with an *oracle* that provides information that is normally not available. Section 10.2.2 shows that one can get close to $r_{\inf}(f)$ with a simple thresholding.

Linear Projection The analysis of diagonal estimators can be simplified by restricting $a[m] \in \{0, 1\}$. When $a[m] = 1$, the estimator $\tilde{F} = DX$ selects the coefficient $X_B[m]$, and it removes it if $a[m] = 0$.

If each $a[m]$ is a constant, then D is a linear orthonormal projection on the space generated by the M vectors g_m such that $a[m] = 1$. Suppose that $a[m] = 1$ for $0 \leq m < M$. The risk (10.26) becomes

$$r(D, f) = \sum_{m=M}^{N-1} |f_B[m]|^2 + M\sigma^2 = \epsilon_l[M] + M\sigma^2, \tag{10.30}$$

where $\epsilon_l[M]$ is the linear approximation error computed in (9.1). The two terms $\epsilon_l[M]$ and $M\sigma^2$ are respectively the bias and the variance components of the estimator. To minimize $r(D, f)$, the parameter M is adjusted so that the bias is of the same order as the variance. When the noise variance σ^2 decreases, the following proposition proves that the decay rate of $r(D, f)$ depends on the decay rate of $\epsilon_l[M]$ as M increases.

Proposition 10.2 *If* $\epsilon_l[M] \sim C^2 M^{1-2s}$ *with* $1 \leq C/\sigma \leq N^s$ *then*

$$\min_M r(D, f) \sim C^{1/s} \sigma^{2-1/s}. \tag{10.31}$$

Proof[2]. Let M_0 be defined by:

$$(M_0 + 1)\sigma^2 \geq \epsilon_l[M_0] \geq M_0 \sigma^2.$$

Since $\epsilon_l[M] \sim C^2 M^{1-2s}$ we get $M_0 \sim C^s/\sigma^s$. The condition $1 \leq C/\sigma \leq N^s$ ensures that $1 \leq M_0 \leq N$. The risk (10.30) satisfies

$$M_0 \sigma^2 \leq \min_M r(D, f) \leq (2M_0 + 1)\sigma^2, \tag{10.32}$$

and $M_0 \sim C^s/\sigma^s$ implies (10.32). ■

Projection With Oracle The non-linear projector that minimizes the risk (10.27) is defined by

$$a[m] = \begin{cases} 1 & \text{if } |f_{\mathcal{B}}[m]| \geq \sigma \\ 0 & \text{if } |f_{\mathcal{B}}[m]| < \sigma \end{cases} . \tag{10.33}$$

This projector cannot be implemented because $a[m]$ depends on $|f_{\mathcal{B}}[m]|$ instead of $X_{\mathcal{B}}[m]$. It uses an "oracle" that keeps the coefficients $f_{\mathcal{B}}[m]$ that are above the noise. The risk of this oracle projector is computed with (10.27):

$$r_p(f) = \mathsf{E}\{\|f - \tilde{F}\|^2\} = \sum_{m=0}^{N-1} \min(|f_{\mathcal{B}}[m]|^2, \sigma^2). \tag{10.34}$$

Since for any x, y

$$\min(x, y) \geq \frac{xy}{x+y} \geq \frac{1}{2}\min(x, y)$$

the risk of the oracle projector (10.34) is of the same order as the risk of an oracle attenuation (10.29):

$$r_p(f) \geq r_{\inf}(f) \geq \frac{1}{2} r_p(f) . \tag{10.35}$$

As in the linear case, the risk of an oracle projector can be related to the approximation error of f in the basis $\mathcal{B}$. Let M be the number of coefficients such that $|f_{\mathcal{B}}[m]| \geq \sigma$. The optimal non-linear approximation of f by these M larger amplitude coefficients is

$$f_M = \sum_{|f_{\mathcal{B}}[m]| \geq \sigma} f_{\mathcal{B}}[m]\, g_m .$$

The approximation error is studied in Section 9.2:

$$\epsilon_n[M] = \|f - f_M\|^2 = \sum_{|f_{\mathcal{B}}[m]| < \sigma} |f_{\mathcal{B}}[m]|^2.$$

The risk (10.34) of an oracle projection can thus be rewritten

$$r_p(f) = \sum_{m=0}^{N-1} \min(|f_{\mathcal{B}}[m]|^2, \sigma^2) = \epsilon_n[M] + M\sigma^2. \tag{10.36}$$

The following proposition proves that when σ decreases, the decay of this risk depends on the decay of $\epsilon_n[M]$ as M increases.

Proposition 10.3 *If* $\epsilon_n[M] \sim C^2 M^{1-2s}$ *with* $1 \leq C/\sigma \leq N^s$, *then*

$$r_p(f) \sim C^{1/s} \sigma^{2-1/s} . \tag{10.37}$$

Proof[2]. Observe that

$$r_p(f) = \min_{0 \le m \le N} \left(\epsilon_n[m] + m\sigma^2 \right)$$

and hence

$$r_p(f) \sim \min_{0 \le m \le N} \left(C^2 m^{1-2s} + m\sigma^2 \right).$$

Let m_0 be such that $C^2 m_0^{1-2s} = m_0 \sigma^2$,

$$r_p(f) \sim m_0 \sigma^2 = \frac{C^{1/s}}{\sigma^{1/s}} \sigma^2$$

which proves (10.37). The hypothesis $1 \le C/\sigma \le N^s$ is required to make sure that $1 \le m_0 \le N$. ■

Propositions 10.2 and 10.3 prove that the performance of linear and oracle projection estimators depends respectively on the precision of linear and non-linear approximations in the basis $\mathcal{B}$. Having an approximation error that decreases quickly means that one can then construct a sparse and precise signal representation with only a few vectors in $\mathcal{B}$. Section 9.2 shows that non-linear approximations can be much more precise, in which case the risk of a non-linear oracle projection is much smaller than the risk of a linear projection.

10.2.2 Thresholding Estimation

In a basis $\mathcal{B} = \{g_m\}_{0 \le m < N}$, a diagonal estimator of f from $X = f + W$ can be written

$$\tilde{F} = DX = \sum_{m=0}^{N-1} d_m(X_{\mathcal{B}}[m])\, g_m . \tag{10.38}$$

We suppose that W is a Gaussian white noise of variance σ^2. When d_m are thresholding functions, the risk of this estimator is shown to be close to the lower bounds obtained with oracle estimators.

Hard thresholding A hard thresholding estimator is implemented with

$$d_m(x) = \rho_T(x) = \begin{cases} x & \text{if } |x| > T \\ 0 & \text{if } |x| \le T \end{cases} . \tag{10.39}$$

The operator D in (10.38) is then a non-linear projector in the basis $\mathcal{B}$. The risk of this thresholding is

$$r_t(f) = r(D, f) = \sum_{m=0}^{N-1} \mathrm{E}\{ |f_{\mathcal{B}}[m] - \rho_T(X_{\mathcal{B}}[m])|^2 \}.$$

Since $X_{\mathcal{B}}[m] = f_{\mathcal{B}}[m] + W_{\mathcal{B}}[m]$,

$$|f_{\mathcal{B}}[m] - \rho_T(X_{\mathcal{B}}[m])|^2 = \begin{cases} |W_{\mathcal{B}}[m]|^2 & \text{if } |X_{\mathcal{B}}[m]| > T \\ |f_{\mathcal{B}}[m]|^2 & \text{if } |X_{\mathcal{B}}[m]| \le T \end{cases} .$$

A thresholding is a projector whose risk is therefore larger than the risk (10.34) of an oracle projector:

$$r_t(f) \geq r_p(f) = \sum_{m=0}^{N-1} \min(|f_{\mathcal{B}}[m]|^2, \sigma^2).$$

Soft Thresholding An oracle attenuation (10.28) yields a risk $r_{\inf}(f)$ that is smaller than the risk $r_p(f)$ of an oracle projection, by slightly decreasing the amplitude of all coefficients in order to reduce the added noise. A similar attenuation, although non-optimal, is implemented by a soft thresholding, which decreases by T the amplitude of all noisy coefficients. The resulting diagonal estimator $\tilde{F}$ in (10.38) is calculated with the soft thresholding function

$$d_m(x) = \rho_T(x) = \begin{cases} x - T & \text{if } x \geq T \\ x + T & \text{if } x \leq -T \\ 0 & \text{if } |x| \leq T \end{cases}. \tag{10.40}$$

This soft thresholding is the solution that minimizes a quadratic distance to the data, penalized by an $\mathbf{l}^1$ norm. Given the data $x[m]$, the vector $y[m]$ which minimizes

$$\sum_{m=1}^{N-1} |y[m] - x[m]|^2 + 2T \sum_{m=1}^{N-1} |y[m]|$$

is $y[m] = \rho_T(x[m])$.

The threshold T is generally chosen so that there is a high probability that it is just above the maximum level of the noise coefficients $|W_{\mathcal{B}}[m]|$. Reducing by T the amplitude of all noisy coefficients thus ensures that the amplitude of an estimated coefficient is smaller than the amplitude of the original one:

$$|\rho_T(X_{\mathcal{B}}[m])| \leq |f_{\mathcal{B}}[m]|. \tag{10.41}$$

In a wavelet basis where large amplitude coefficients correspond to transient signal variations, this means that the estimation keeps only transients coming from the original signal, without adding others due to the noise.

Thresholding Risk The following theorem [167] proves that for an appropriate choice of T, the risk of a thresholding is close to the risk of an oracle projector $r_p(f) = \sum_{m=0}^{N-1} \min(|f_{\mathcal{B}}[m]|^2, \sigma^2)$. We denote by $\mathcal{O}_d$ the set of all operators that are diagonal in $\mathcal{B}$, and which can thus be written as in (10.38).

Theorem 10.4 (DONOHO, JOHNSTONE) *Let $T = \sigma\sqrt{2\log_e N}$. The risk $r_t(f)$ of a hard or a soft thresholding estimator satisfies for all $N \geq 4$*

$$r_t(f) \leq (2\log_e N + 1)\left(\sigma^2 + r_p(f)\right). \tag{10.42}$$

The factor $2\log_e N$ *is optimal among diagonal estimators in* $\mathcal{B}$*:*

$$\lim_{N\to+\infty} \inf_{D\in\mathcal{O}_d} \sup_{f\in\mathbb{C}^N} \frac{\mathsf{E}\{\|f-\tilde{F}\|^2\}}{\sigma^2+r_p(f)} \frac{1}{2\log_e N} = 1\,. \tag{10.43}$$

Proof[2]. The proof of (10.42) is given for a soft thresholding. For a hard thresholding, the proof is similar although slightly more complicated. For a threshold λ, a soft thresholding is computed with

$$\rho_\lambda(x) = (x-\lambda\,\mathrm{sign}(x))\,\mathbf{1}_{|x|>\lambda}\,.$$

Let X be a Gaussian random variable of mean μ and variance 1. The risk when estimating μ with a soft thresholding of X is

$$r(\lambda,\mu) = \mathsf{E}\{|\rho_\lambda(X)-\mu|^2\} = \mathsf{E}\{|(X-\lambda\,\mathrm{sign}(X))\,\mathbf{1}_{|X|>\lambda}-\mu|^2\}\,. \tag{10.44}$$

If X has a variance σ^2 and a mean μ then by considering $\tilde{X} = X/\sigma$ we verify that

$$\mathsf{E}\{|\rho_\lambda(X)-\mu|^2\} = \sigma^2 r\left(\frac{\lambda}{\sigma},\frac{\mu}{\sigma}\right).$$

Since $f_B[m]$ is a constant, $X_B[m] = f_B[m]+W_B[m]$ is a Gaussian random variable of mean $f_B[m]$ and variance σ^2. The risk of the soft thresholding estimator $\tilde{F}$ with a threshold T is thus

$$r_t(f) = \sigma^2 \sum_{m=0}^{N-1} r\left(\frac{T}{\sigma},\frac{f_B[m]}{\sigma}\right). \tag{10.45}$$

An upper bound of this risk is calculated with the following lemma.

Lemma 10.1 *If* $\mu \geq 0$ *then*

$$r(\lambda,\mu) \leq r(\lambda,0) + \min(\mu^2, 1+\lambda^2). \tag{10.46}$$

To prove (10.46), we first verify that if $\mu \geq 0$ then

$$0 \leq \frac{\partial r(\lambda,\mu)}{\partial \mu} = 2\mu \int_{-\lambda+\mu}^{\lambda-\mu} \phi(x)\,dx \leq 2\mu\,, \tag{10.47}$$

where $\phi(x)$ is the normalized Gaussian probability density

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Indeed (10.44) shows that

$$r(\lambda,\mu) = \mu^2 \int_{-\lambda+\mu}^{\lambda-\mu} \phi(x)\,dx + \int_{\lambda-\mu}^{+\infty} (x-\lambda)^2\,\phi(x)\,dx + \int_{-\infty}^{-\lambda+\mu} (x+\lambda)^2\,\phi(x)\,dx. \tag{10.48}$$

We obtain (10.47) by differentiating with respect to μ.

Since $\int_{-\infty}^{+\infty} \phi(x)\,dx = \int_{-\infty}^{+\infty} x^2\,\phi(x)\,dx = 1$ and $\frac{\partial r(\lambda,\mu)}{\partial \mu} \geq 0$, necessarily

$$r(\lambda,\mu) \leq \lim_{\mu\to+\infty} r(\lambda,\mu) = 1+\lambda^2. \tag{10.49}$$

Moreover, since $\frac{\partial r(\lambda,s)}{\partial s} \leq 2s$

$$r(\lambda,\mu) - r(\lambda,0) = \int_0^\mu \frac{\partial r(\lambda,s)}{\partial s}\, ds \leq \mu^2. \tag{10.50}$$

The inequality (10.46) of the lemma is finally derived from (10.49) and (10.50):

$$r(\lambda,\mu) \leq \min(r(\lambda,0)+\mu^2, 1+\lambda^2) \leq r(\lambda,0) + \min(\mu^2, 1+\lambda^2).$$

By inserting the inequality (10.46) of the lemma in (10.45), we get

$$r_t(f) \leq N\sigma^2 r\left(\frac{T}{\sigma},0\right) + \sigma^2 \sum_{m=0}^{N-1} \min\left(\frac{T^2}{\sigma^2}, \frac{|f_B[m]|^2}{\sigma^2}\right). \tag{10.51}$$

The expression (10.48) shows that $r(\lambda,0) = 2\int_0^{+\infty} x^2 \phi(x+\lambda)\, dx$. For $T = \sigma\sqrt{2\log_e N}$ and $N \geq 4$, one can verify that

$$N r\left(\frac{T}{\sigma},0\right) \leq 2\log_e N + 1\,. \tag{10.52}$$

Moreover,

$$\begin{aligned} \sigma^2 \min\left(\frac{T^2}{\sigma^2}, \frac{|f_B[m]|^2}{\sigma^2}\right) &= \min(2\sigma^2 \log_e N, |f_B[m]|^2) \\ &\leq (2\log_e N + 1)\min(\sigma^2, |f_B[m]|^2). \end{aligned} \tag{10.53}$$

Inserting (10.52) and (10.53) in (10.51) proves (10.42).

Since the soft and hard thresholding estimators are particular instances of diagonal estimators, the inequality (10.42) implies that

$$\lim_{N\to+\infty} \inf_{D\in\mathcal{O}_d} \sup_{f\in\mathbb{C}^N} \frac{\mathrm{E}\{\|f-\tilde{F}\|^2\}}{\sigma^2 + r_p(f)} \frac{1}{2\log_e N} \leq 1\,. \tag{10.54}$$

To prove that the limit is equal to 1, for N fixed we compute a lower bound by replacing the sup over all signals f by an expected value over the distribution of a particular signal process F. The coefficients $F_B[m]$ are chosen to define a very sparse sequence. They are independent random variables having a high probability $1-\alpha_N$ to be equal to 0 and a low probability α_N to be equal to a value μ_N that is on the order of $\sigma\sqrt{2\log_e N}$, but smaller. By adjusting μ_N and α_N, Donoho and Johnstone [167] prove that the Bayes estimator $\tilde{F}$ of F tends to zero as N increases and they derive a lower bound of the left-hand side of (10.54) that tends to 1. ■

The upper bound (10.42) proves that the risk $r_t(f)$ of a thresholding estimator is at most $2\log_e N$ times larger than the risk $r_p(f)$ of an oracle projector. Moreover, (10.43) proves that the $2\log_e N$ factor cannot be improved by any other diagonal estimator. For $r_p(f)$ to be small, (10.36) shows that f must be well approximated by a few vectors in $\mathcal{B}$. One can verify [167] that the theorem remains valid if $r_p(f)$ is replaced by the risk $r_{\inf}(f)$ of an oracle attenuation, which is smaller.

Choice of Threshold The threshold T must be chosen just above the maximum level of the noise. Indeed, if $f = 0$ and thus $X_B = W_B$, then to ensure that $\tilde{F} \approx 0$ the noise coefficients $|W_B[m]|$ must have a high probability of being below T. However, if $f \neq 0$ then T must not be too large, so that we do not set to zero too many coefficients such that $|f_B[m]| \geq \sigma$. Since W_B is a vector of N independent Gaussian random variables of variance σ^2, one can prove [9] that the maximum amplitude of the noise has a very high probability of being just below $T = \sigma\sqrt{2\log_e N}$:

$$\lim_{N \to +\infty} \Pr\left(T - \frac{\sigma \log_e \log_e N}{\log_e N} \leq \max_{0 \leq m < N} |W_B[m]| \leq T\right) = 1. \tag{10.55}$$

This explains why the theorem chooses this value. That the threshold T increases with N may seem counterintuitive. This is due to the tail of the Gaussian distribution, which creates larger and larger amplitude noise coefficients when the sample size increases. The threshold $T = \sigma\sqrt{2\log_e N}$ is not optimal and in general a lower threshold reduces the risk. One can however prove that when N tends to $+\infty$, the optimal value of T grows like $\sigma\sqrt{2\log_e N}$.

Upper-Bound Interpretation Despite the technicality of the proof, the factor $2\log_e N$ of the upper bound (10.42) can be easily explained. The ideal coefficient selection (10.33) sets $X_B[m]$ to zero if and only if $|f_B[m]| \leq \sigma$, whereas a hard thresholding sets $X_B[m]$ to zero when $|X_B[m]| \leq T$. If $|f_B[m]|| \leq \sigma$ then it is very likely that $|X_B[m]| \leq T$, because T is above the noise level. In this case the hard thresholding sets $X_B[m]$ to zero as the oracle projector (10.33) does. If $|f_B[m]| \geq 2T$ then it is likely that $|X_B[m]| \geq T$ because $|W_B[m]| \leq T$. In this case the hard thresholding and the oracle projector retain $X_B[m]$.

The hard thresholding may behave differently from the ideal coefficient selection when $|f_B[m]|$ is on the order of T. The ideal selection yields a risk: $\min(\sigma^2, |f_B[m]|^2) = \sigma^2$. If we are unlucky and $|X_B[m]| \leq T$, then the thresholding sets $X_B[m]$ to zero, which produces a risk

$$|f_B[m]|^2 \sim T^2 = 2\log_e N \sigma^2.$$

In this worst case, the thresholding risk is $2\log_e N$ times larger than the ideal selection risk. Since the proportion of coefficients $|f_B[m]|$ on the order of T is often small, the ratio between the hard thresholding risk and the oracle projection risk is generally significantly smaller than $2\log_e N$.

Colored Noise Thresholding estimators can be adapted when the noise W is not white. We suppose that $\mathrm{E}\{W[n]\} = 0$. Since W is not white, $\sigma_m^2 = \mathrm{E}\{|W_B[m]|^2\}$ depends on each vector g_m of the basis. As in (10.33) and (10.34), we verify that an oracle projector which keeps all coefficients such that $|f_B[m]| \geq \sigma_m$ and sets to zero all others has a risk

$$r_p(f) = \sum_{m=0}^{N-1} \min(|f_B[m]|^2, \sigma_m^2) .$$

Any linear or non-linear projector in the basis $\mathcal{B}$ has a risk larger than $r_p(f)$.

Since the noise variance depends on m, a thresholding estimator must vary the threshold T_m as a function of m. Such a hard or soft thresholding estimator can be written

$$\tilde{F} = DX = \sum_{m=0}^{N-1} \rho_{T_m}(X_{\mathcal{B}}[m])\, g_m . \tag{10.56}$$

The following proposition generalizes Theorem 10.4 to compute the thresholding risk $r_t(f) = \mathrm{E}\{\|f - \tilde{F}\|^2\}$.

Proposition 10.4 (DONOHO, JOHNSTONE) *Let $\tilde{F}$ be a hard or soft thresholding estimator with*

$$T_m = \sigma_m \sqrt{2 \log_e N} \quad \textit{for } 0 \leq m < N .$$

Let $\bar{\sigma}^2 = N^{-1} \sum_{m=0}^{N-1} \sigma_m^2$. For any $N \geq 4$

$$r_t(f) \leq (2 \log_e N + 1)\left(\bar{\sigma}^2 + r_p(f)\right). \tag{10.57}$$

The proof of (10.57) is identical to the proof of (10.42). The thresholds T_m are chosen to be just above the amplitude of each noisy coefficient $W_{\mathcal{B}}[m]$. Section 10.4.2 studies an application to the restoration of blurred signals.

10.2.3 Thresholding Refinements [3]

We mentioned that the thresholding risk can be reduced by choosing a threshold smaller than $\sigma\sqrt{2 \log_e N}$. A threshold adapted to the data is calculated by minimizing an estimation of the risk. This section finishes with an important improvement of thresholding estimators, obtained with a translation invariant algorithm.

SURE Thresholds To study the impact of the threshold on the risk, we denote by $r_t(f,T)$ the risk of a soft thresholding estimator calculated with a threshold T. An estimate $\tilde{r}_t(f,T)$ of $r_t(f,T)$ is calculated from the noisy data X, and T is optimized by minimizing $\tilde{r}_t(f,T)$.

To estimate the risk $r_t(f,T)$, observe that if $|X_{\mathcal{B}}[m]| < T$ then the soft thresholding sets this coefficient to zero, which produces a risk equal to $|f_{\mathcal{B}}[m]|^2$. Since

$$\mathrm{E}\{|X_{\mathcal{B}}[m]|^2\} = |f_{\mathcal{B}}[m]|^2 + \sigma^2,$$

one can estimate $|f_{\mathcal{B}}[m]|^2$ with $|X_{\mathcal{B}}[m]|^2 - \sigma^2$. If $|X_{\mathcal{B}}[m]| \geq T$, the soft thresholding subtracts T from the amplitude of $X_{\mathcal{B}}[m]$. The expected risk is the sum of the noise energy plus the bias introduced by the reduction of the amplitude of $X_{\mathcal{B}}[m]$ by T. It is estimated by $\sigma^2 + T^2$. The resulting estimator of $r_t(f,T)$ is

$$\tilde{r}_t(f,T) = \sum_{m=0}^{N-1} \Phi(|X_{\mathcal{B}}[m]|^2) \tag{10.58}$$

with

$$\Phi(u) = \begin{cases} u-\sigma^2 & \text{if } u \leq T^2 \\ \sigma^2+T^2 & \text{if } u > T \end{cases} . \tag{10.59}$$

The following theorem [169] proves that $\tilde{r}_t(f,T)$ is a Stein Unbiased Risk Estimator (SURE) [319].

Theorem 10.5 (DONOHO, JOHNSTONE) *For a soft thresholding, the risk estimator $\tilde{r}_t(f,T)$ is unbiased:*

$$E\{\tilde{r}_t(f,T)\} = r_t(f,T). \tag{10.60}$$

Proof[3]. As in (10.45), we prove that the risk of a soft thresholding can be written

$$r_t(f,T) = \mathsf{E}\{\|f-\tilde{F}\|^2\} = \sigma^2 \sum_{m=0}^{N-1} r(T, f_B[m], \sigma),$$

with

$$r(\lambda,\mu,\sigma) = \mathsf{E}\{|\rho_\lambda(X)-\mu|^2\} = \mathsf{E}\{|(X-\lambda \operatorname{sign}(X))\mathbf{1}_{|X|>\lambda}-\mu|^2\}\,, \tag{10.61}$$

where X is a Gaussian random variable with mean μ and variance σ^2. The equality (10.60) is proved by verifying that

$$r(T,\mu,\sigma) = \mathsf{E}\Big\{\Phi(|X|^2)\Big\} = (\sigma^2+T^2)\,\mathsf{E}\{\mathbf{1}_{|X|\geq T}\} + \mathsf{E}\Big\{(|X|^2-\sigma^2)\,\mathbf{1}_{|X|\leq T}\Big\}. \tag{10.62}$$

Following the calculations of Stein [319], we rewrite

$$r(T,\mu,\sigma) = \mathsf{E}\{(X-g(X)-\mu)^2\}, \tag{10.63}$$

where $g(x) = T\operatorname{sign}(x) + (x - T\operatorname{sign}(x))\,\mathbf{1}_{|x|<T}$ is a differentiable function. Developing (10.63) gives

$$r(T,\mu,\sigma) = \mathsf{E}\{(X-\mu)^2\} + \mathsf{E}\{|g(X)|^2\} - 2\mathsf{E}\{(X-\mu)\,g(X)\}. \tag{10.64}$$

The probability density of X is the Gaussian $\phi_\sigma(y-\mu)$. The change of variable $x = y-\mu$ shows that

$$\mathsf{E}\{(X-\mu)\,g(X)\} = \int_{-\infty}^{+\infty} x\,g(x+\mu)\,\phi_\sigma(x)\,dx.$$

Since $x\,\phi_\sigma(x) = -\sigma^2\,\phi'_\sigma(x)$, an integration by parts gives

$$\begin{aligned} \mathsf{E}\{(X-\mu)g(X)\} &= -\sigma^2 \int_{-\infty}^{+\infty} g(x+\mu)\,\phi'_\sigma(x)\,dx \\ &= \sigma^2 \int_{-\infty}^{+\infty} g'(x+\mu)\,\phi_\sigma(x)\,dx\,. \end{aligned}$$

Since $g'(x) = \mathbf{1}_{|x|\leq T}$,

$$\mathsf{E}\{(X-\mu)\,g(X)\} = \sigma^2\,\mathsf{E}\{\mathbf{1}_{|X|\leq T}\}.$$

Inserting this expression in (10.64) yields

$$r(T,\mu,\sigma) = \sigma^2 + \mathsf{E}\{|g(X)|^2\} - 2\sigma^2\,\mathsf{E}\{\mathbf{1}_{|X|\leq T}\}.$$

But $|g(x)|^2 = |x|^2\mathbf{1}_{|x|<T} + T^2\mathbf{1}_{|x|\geq T}$ and $\mathsf{E}\{\mathbf{1}_{|X|\geq T}\} + \mathsf{E}\{\mathbf{1}_{|X|<T}\} = 1$, so

$$r(T,\mu,\sigma) = (\sigma^2+T^2)\,\mathsf{E}\{\mathbf{1}_{|X|\geq T}\} + \mathsf{E}\Big\{(|X|^2-\sigma^2)\mathbf{1}_{|X|\leq T}\Big\},$$

which proves (10.62) and hence (10.60). ■

To find the $\tilde{T}$ that minimizes the SURE estimator $\tilde{r}_t(f,T)$, the N data coefficients $X_{\mathcal{B}}[m]$ are sorted in decreasing amplitude order with $O(N\log_2 N)$ operations. Let $X^r_{\mathcal{B}}[k] = X_{\mathcal{B}}[m_k]$ be the coefficient of rank k: $|X^r_{\mathcal{B}}[k]| \geq |X^r_{\mathcal{B}}[k+1]|$ for $1 \leq k < N$. Let l be the index such that $|X^r_{\mathcal{B}}[l]| \leq T < |X^r_{\mathcal{B}}[l+1]|$. We can rewrite (10.58):

$$\tilde{r}_t(f,T) = \sum_{k=l}^{N} |X^r_{\mathcal{B}}[k]|^2 - (N-l)\sigma^2 + l(\sigma^2+T^2). \tag{10.65}$$

To minimize $\tilde{r}_t(f,T)$ we must choose $T = |X^r_{\mathcal{B}}[l]|$ because $r_t(f,T)$ is increasing in T. To find the $\tilde{T}$ that minimizes $\tilde{r}_t(f,T)$ it is therefore sufficient to compare the N possible values $\{|X^r_{\mathcal{B}}[k]|\}_{1\leq k\leq N}$, that requires $O(N)$ operations if we progressively recompute the formula (10.65). The calculation of $\tilde{T}$ is thus performed with $O(N\log_2 N)$ operations.

Although the estimator $\tilde{r}_t(f,T)$ of $r_t(f,T)$ is unbiased, its variance may induce errors leading to a threshold $\tilde{T}$ that is too small. This happens if the signal energy is small relative to the noise energy: $\|f\|^2 \ll \mathsf{E}\{\|W\|^2\} = N\sigma^2$. In this case, one must impose $T = \sigma\sqrt{2\log_e N}$ in order to remove all the noise. Since $\mathsf{E}\{\|X\|^2\} = \|f\|^2 + N\sigma^2$, we estimate $\|f\|^2$ with $\|X\|^2 - N\sigma^2$ and compare this value with a minimum energy level $\epsilon_N = \sigma^2 N^{1/2}(\log_e N)^{3/2}$. The resulting SURE threshold is

$$T = \begin{cases} \sigma\sqrt{2\log_e N} & \text{if } \|X\|^2 - N\sigma^2 \leq \epsilon_N \\ \tilde{T} & \text{if } \|X\|^2 - N\sigma^2 > \epsilon_N \end{cases}. \tag{10.66}$$

Let Θ be a signal set and $\min_T r_t(\Theta)$ be the minimax risk of a soft thresholding obtained by optimizing the choice of T depending on Θ. Donoho and Johnstone [169] prove that the threshold computed empirically with (10.66) yields a risk $r_t(\Theta)$ equal to $\min_T r_t(\Theta)$ plus a corrective term that decreases rapidly when N increases, if $\epsilon_N = \sigma^2 N^{1/2}(\log_e N)^{3/2}$.

Problem 10.9 studies a similar risk estimator for hard thresholding. However, this risk estimator is biased. We thus cannot guarantee that the threshold that minimizes the estimated risk is nearly optimal for hard thresholding estimations.

Translation Invariant Thresholding An improved thresholding estimator is calculated by averaging estimators for translated versions of the signal. Let us consider signals of period N. Section 5.4 explains that the representation of f in a basis $\mathcal{B}$ is

not translation invariant, unless $\mathcal{B}$ is a Dirac or a Fourier basis. Let $f^p[n] = f[n-p]$. The vectors of coefficients $f_{\mathcal{B}}$ and $f^p_{\mathcal{B}}$ are not simply translated or permuted. They may be extremely different. Indeed

$$f^p_{\mathcal{B}}[m] = \langle f[n-p], g_m[n]\rangle = \langle f[n], g_m[n+p]\rangle,$$

and not all the vectors $g_m[n+p]$ belong to the basis $\mathcal{B}$, for $0 \leq p < N$. As a consequence, the signal recovered by thresholding the coefficients $f^p_{\mathcal{B}}[m]$ is not a translation of the signal reconstructed after thresholding $f_{\mathcal{B}}[m]$.

The translation invariant algorithm of Coifman and Donoho [137] estimates all translations of f and averages them after a reverse translation. For all $0 \leq p < N$, the estimator $\tilde{F}^p$ of f^p is computed by thresholding the translated data $X^p[n] = X[n-p]$:

$$\tilde{F}^p = \sum_{m=0}^{N-1} \rho_T(X^p_{\mathcal{B}}[m])\, g_m\,,$$

where $\rho_T(x)$ is a hard or soft thresholding function. The translation invariant estimator is obtained by shifting back and averaging these estimates:

$$\tilde{F}[n] = \frac{1}{N}\sum_{p=0}^{N-1} \tilde{F}^p[n+p]. \tag{10.67}$$

In general, this requires N times more calculations than for a standard thresholding estimator. In wavelet and wavelet packet bases, which are partially translation invariant, the number of operations is only multiplied by $\log_2 N$, and the translation invariance reduces the risk significantly.

10.2.4 Wavelet Thresholding

A wavelet thresholding is equivalent to estimating the signal by averaging it with a kernel that is locally adapted to the signal regularity [4]. This section justifies the numerical results with heuristic arguments. Section 10.3.3 proves that the wavelet thresholding risk is nearly minimax for signals and images with bounded variation.

A filter bank of conjugate mirror filters decomposes a discrete signal in a discrete orthogonal wavelet basis defined in Section 7.3.3. The discrete wavelets $\psi_{j,m}[n] = \psi_j[n - N2^j m]$ are translated modulo modifications near the boundaries, which are explained in Section 7.5. The support of the signal is normalized to $[0,1]$ and has N samples spaced by N^{-1}. The scale parameter 2^j thus varies from $2^L = N^{-1}$ up to $2^J < 1$:

$$\mathcal{B} = \left[\{\psi_{j,m}[n]\}_{L<j\leq J, 0\leq m<2^{-j}}\,,\ \{\phi_{J,m}[n]\}_{0\leq m<2^{-J}}\right]. \tag{10.68}$$

A thresholding estimator in this wavelet basis can be written

$$\tilde{F} = \sum_{j=L+1}^{J}\sum_{m=0}^{2^{-j}} \rho_T\Big(\langle X, \psi_{j,m}\rangle\Big)\,\psi_{j,m} + \sum_{m=0}^{2^{-J}} \rho_T\Big(\langle X, \phi_{J,m}\rangle\Big)\,\phi_{J,m}\,, \tag{10.69}$$

where ρ_T is a hard thresholding (10.39) or a soft thresholding (10.40). The upper bound (10.42) proves that the estimation risk is small if the energy of f is absorbed by a few wavelet coefficients.

Adaptive Smoothing The thresholding sets to zero all coefficients $|\langle X, \psi_{j,m}\rangle| \leq T$. This performs an adaptive smoothing that depends on the regularity of the signal f. Since T is above the maximum amplitude of the noise coefficients $|\langle W, \psi_{j,m}\rangle|$, if

$$|\langle X, \psi_{j,m}\rangle| = |\langle f, \psi_{j,m}\rangle + \langle W, \psi_{j,m}\rangle| \geq T,$$

then $|\langle f, \psi_{j,m}\rangle|$ has a high probability of being at least of the order T. At fine scales 2^j, these coefficients are in the neighborhood of sharp signal transitions, as shown by Figure 10.4(b). By keeping them, we avoid smoothing these sharp variations. In the regions where $|\langle X, \psi_{j,m}\rangle| < T$, the coefficients $\langle f, \psi_{j,m}\rangle$ are likely to be small, which means that f is locally regular. Setting wavelet coefficients to zero is equivalent to locally averaging the noisy data X, which is done only if the underlying signal f appears to be regular.

Noise Variance Estimation To estimate the variance σ^2 of the noise $W[n]$ from the data $X[n] = W[n] + f[n]$, we need to suppress the influence of $f[n]$. When f is piecewise smooth, a robust estimator is calculated from the median of the finest scale wavelet coefficients [167].

The signal X of size N has $N/2$ wavelet coefficients $\{\langle X, \psi_{l,m}\rangle\}_{0 \leq m < N/2}$ at the finest scale $2^l = 2N^{-1}$. The coefficient $|\langle f, \psi_{l,m}\rangle|$ is small if f is smooth over the support of $\psi_{l,m}$, in which case $\langle X, \psi_{l,m}\rangle \approx \langle W, \psi_{l,m}\rangle$. In contrast, $|\langle f, \psi_{l,m}\rangle|$ is large if f has a sharp transition in the support of $\psi_{l,m}$. A piecewise regular signal has few sharp transitions, and hence produces a number of large coefficients that is small compared to $N/2$. At the finest scale, the signal f thus influences the value of a small portion of large amplitude coefficients $\langle X, \psi_{l,m}\rangle$ that are considered to be "outliers." All others are approximately equal to $\langle W, \psi_{l,m}\rangle$, which are independent Gaussian random variables of variance σ^2.

A robust estimator of σ^2 is calculated from the median of $\{\langle X, \psi_{l,m}\rangle\}_{0 \leq m < N/2}$. The median of P coefficients $\mathrm{Med}(\alpha_p)_{0 \leq p < P}$ is the value of the middle coefficient α_{n_0} of rank $P/2$. As opposed to an average, it does not depend on the specific values of coefficients $\alpha_p > \alpha_{n_0}$. If M is the median of the absolute value of P independent Gaussian random variables of zero-mean and variance σ_0^2, then one can show that

$$\mathrm{E}\{M\} \approx 0.6745\, \sigma_0.$$

The variance σ^2 of the noise W is estimated from the median M_X of $\{|\langle X, \psi_{l,m}\rangle|\}_{0 \leq m < N/2}$ by neglecting the influence of f:

$$\tilde{\sigma} = \frac{M_X}{0.6745}. \tag{10.70}$$

Indeed f is responsible for few large amplitude outliers, and these have little impact on M_X.

Hard or Soft Thresholding If we choose the threshold $T = \sigma\sqrt{2\log_e N}$ of Theorem 10.4, we saw in (10.41) that a soft thresholding guarantees with a high probability that

$$|\langle \tilde{F}, \psi_{j,m}\rangle| = |\rho_T(\langle X, \psi_{j,m}\rangle)| \leq |\langle f, \psi_{j,m}\rangle| .$$

The estimator $\tilde{F}$ is at least as regular as f because its wavelet coefficients have a smaller amplitude. This is not true for the hard thresholding estimator, which leaves unchanged the coefficients above T, and which can therefore be larger than those of f because of the additive noise component.

Figure 10.4(a) shows a piecewise polynomial signal of degree at most 3, whose wavelet coefficients are calculated with a Symmlet 4. Figure 10.4(c) gives an estimation computed with a hard thresholding of the noisy wavelet coefficients in Figure 10.4(b). An estimator $\tilde{\sigma}^2$ of the noise variance σ^2 is calculated with the median (10.70) and the threshold is set to $T = \tilde{\sigma}\sqrt{2\log_e N}$. Thresholding wavelet coefficients removes the noise in the domain where f is regular but some traces of the noise remain in the neighborhood of singularities. The resulting SNR is 30.8db. The soft thresholding estimation of Figure 10.4(d) attenuates the noise effect at the discontinuities but the reduction by T of the coefficient amplitude is much too strong, which reduces the SNR to 23.8db. As already explained, to obtain comparable SNR values, the threshold of the soft thresholding must be about half the size of the hard thresholding one. In this example, reducing by two the threshold increases the SNR of the soft thresholding to 28.6db.

Multiscale SURE Thresholds Piecewise regular signals have a proportion of large coefficients $|\langle f, \psi_{j,m}\rangle|$ that increases when the scale 2^j increases. Indeed, a singularity creates the same number of large coefficients at each scale, whereas the total number of wavelet coefficients increases when the scale decreases. To use this prior information, one can adapt the threshold choice to the scale 2^j. At large scale 2^j the threshold T_j should be smaller in order to avoid setting to zero too many large amplitude signal coefficients, which would increase the risk. Section 10.2.3 explains how to compute the threshold value for a soft thresholding, from the coefficients of the noisy data. We first compute an estimate $\tilde{\sigma}^2$ of the noise variance σ^2 with the median formula (10.70) at the finest scale. At each scale 2^j, a different threshold is calculated from the 2^{-j} noisy coefficients $\{\langle X, \psi_{j,m}\rangle\}_{0 \leq m < 2^{-j}}$ with the algorithm of Section 10.2.3. A SURE threshold T_j is calculated by minimizing an estimation (10.65) of the risk at the scale 2^j. The soft thresholding is then performed at each scale 2^j with the threshold T_j. For a hard thresholding, we have no reliable formula with which to estimate the risk and hence compute the adapted threshold with a minimization. One possibility is simply to multiply by 2 the SURE threshold calculated for a soft thresholding.

Figure 10.5(c) is a hard thresholding estimation calculated with the same threshold $T = \tilde{\sigma}\sqrt{2\log_e N}$ at all scales 2^j. The SNR is 23.3db. Figure 10.5(d) is obtained by a soft thresholding with SURE thresholds T_j adapted at each scale 2^j. The SNR is 24.1db. A soft thresholding with the threshold $T = \tilde{\sigma}/2\sqrt{2\log_e N}$

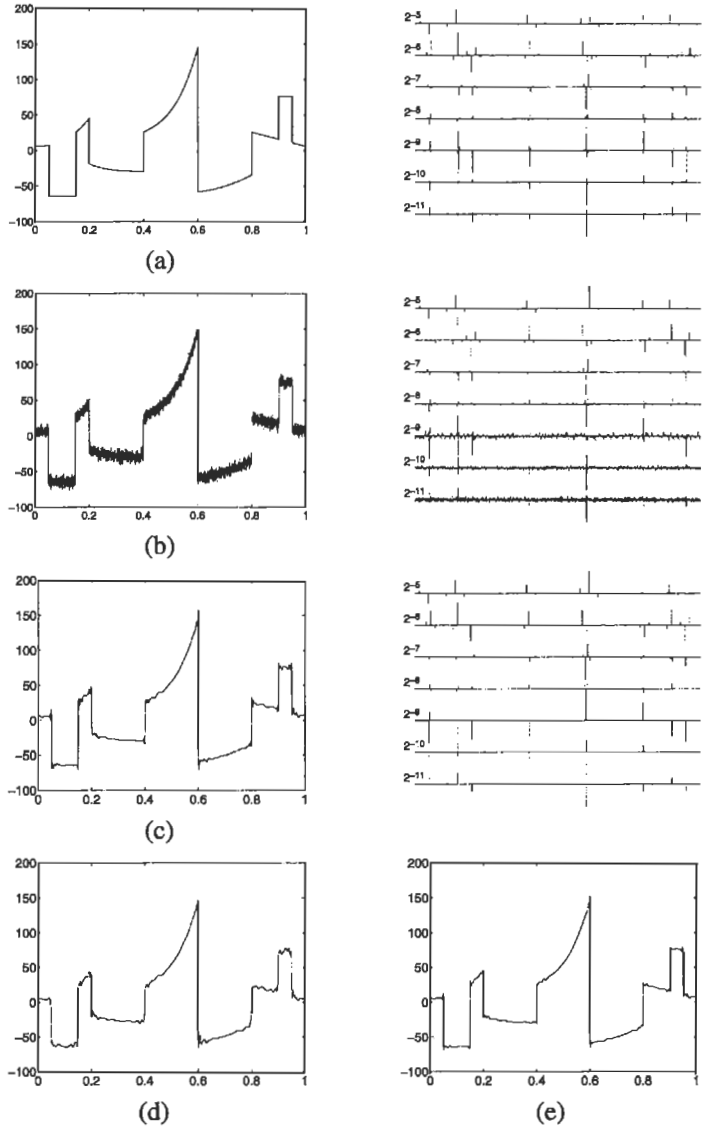


FIGURE 10.4 (a): Piecewise polynomial signal and its wavelet transform on the right. (b): Noisy signal (SNR = 21.9 db) and its wavelet transform. (c): Estimation reconstructed from the wavelet coefficients above threshold, shown on the right (SNR = 30.8 db). (d): Estimation with a wavelet soft thresholding (SNR = 23.8 db). (e): Estimation with a translation invariant hard thresholding (SNR = 33.7 db).

at all scales gives a smaller SNR equal to 21.7 db. The adaptive calculation of thresholds clearly improves the estimation.

Translation Invariance Thresholding noisy wavelet coefficients creates small ripples near discontinuities, as seen in Figures 10.4(c,d) and 10.5(c,d). Indeed, setting to zero a coefficient $\langle f, \psi_{j,m} \rangle$ subtracts $\langle f, \psi_{j,m} \rangle \psi_{j,m}$ from f, which introduces oscillations whenever $\langle f, \psi_{j,m} \rangle$ is non-negligible. Figure 10.4(e) and Figures 10.5(e,f) show that these oscillations are attenuated by a translation invariant estimation (10.67), significantly improving the SNR. Thresholding wavelet coefficients of translated signals and translating back the reconstructed signals yields shifted oscillations created by shifted wavelets that are set to zero. The averaging partially cancels these oscillations, reducing their amplitude.

When computing the translation invariant estimation, instead of shifting the signal, one can shift the wavelets in the opposite direction:

$$\langle f[n-p], \psi_{j,m}[n] \rangle = \langle f[n], \psi_{j,m}[n+p] \rangle = \langle f[n], \psi_j[n - N2^j m + p] \rangle.$$

If f and all wavelets ψ_j are N periodic then all these inner products are provided by the dyadic wavelet transform defined in Section 5.5:

$$Wf[2^j, p] = \langle f[n], \psi_j[n-p] \rangle \text{ for } 0 \leq p < N.$$

The "algorithme à trous" of Section 5.5.2 computes these $N \log_2 N$ coefficients for $L < j \leq 0$ with $O(N \log_2 N)$ operations. One can verify (Problem 10.10) that the translation invariant wavelet estimator (10.67) can be calculated by thresholding the dyadic wavelet coefficients $\langle X[n], \psi_j[n-p] \rangle$ and by reconstructing a signal with the inverse dyadic wavelet transform.

Image Estimation in Wavelet Bases Piecewise regular images are particularly well estimated by thresholding their wavelet coefficients. The image $f[n_1, n_2]$ contaminated by a white noise is decomposed in a separable two-dimensional wavelet basis. Figure 10.6(c) is computed with a hard thresholding in a Symmlet 4 wavelet basis. For images of $N^2 = 512^2$ pixels, the threshold is set to $T = 3\sigma$ instead of $T = \sigma\sqrt{2 \log_e N^2}$, because this improves the SNR significantly. This estimation restores smooth image components and discontinuities, but the visual quality of edges is affected by the Gibbs-like oscillations that also appear in the one-dimensional estimations in Figure 10.4(c) and Figure 10.5(c). Figure 10.6(c) is obtained with a wavelet soft thresholding calculated with a threshold half as large $T = 3/2\sigma$. When using a different SURE threshold T_j calculated with (10.66) at each scale 2^j, the SNR increases to 33.1 db but the visual image quality is not improved. As in one dimension, the Figures 10.6(e,f) calculated with translation invariant thresholdings have a higher SNR and better visual quality. A translation invariant soft thresholding, with SURE thresholds, gives an SNR of 34.2 db.

Section 10.3.3 proves that a thresholding in a wavelet basis has a nearly minimax risk for bounded variation images. Irregular textures are badly estimated

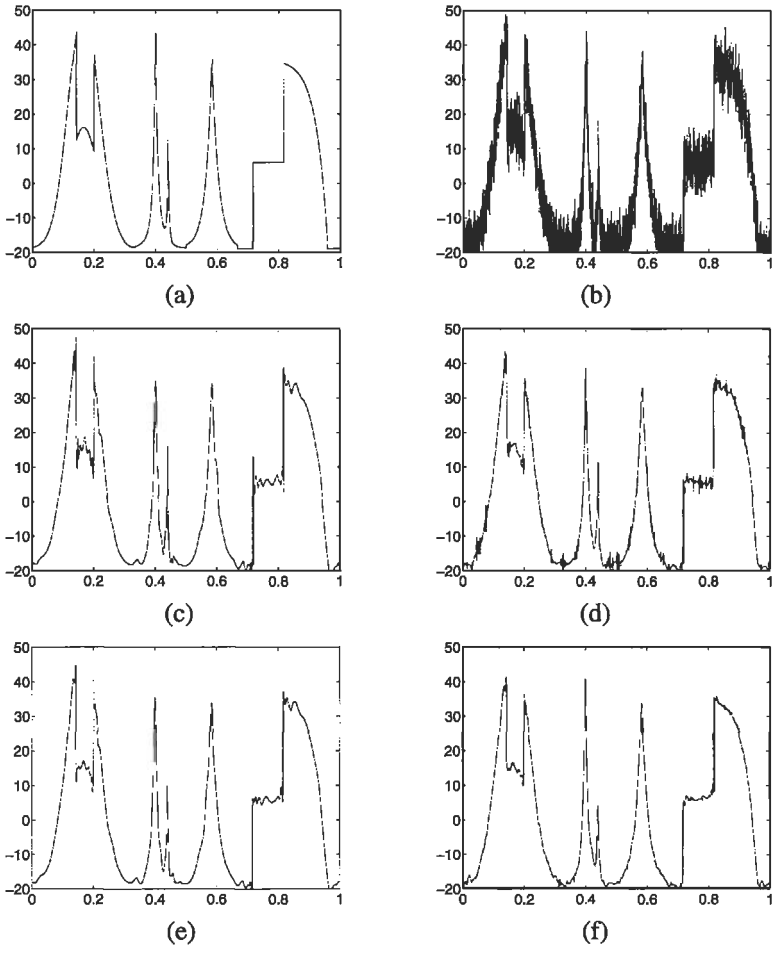


FIGURE 10.5 (a): Original signal. (b): Noisy signal (SNR = 13.1 db). (c): Estimation by a hard thresholding in a wavelet basis (Symmlet 4), with $T = \tilde{\sigma}\sqrt{2\log_e N}$ (SNR = 23.3 db). (d): Soft thresholding calculated with SURE thresholds T_j adapted to each scale 2^j (SNR = 24.5 db). (e): Translation invariant hard thresholding with $T = \tilde{\sigma}\sqrt{2\log_e N}$ (SNR = 25.7 db). (f): Translation invariant soft thresholding with SURE thresholds (SNR = 25.6 db).

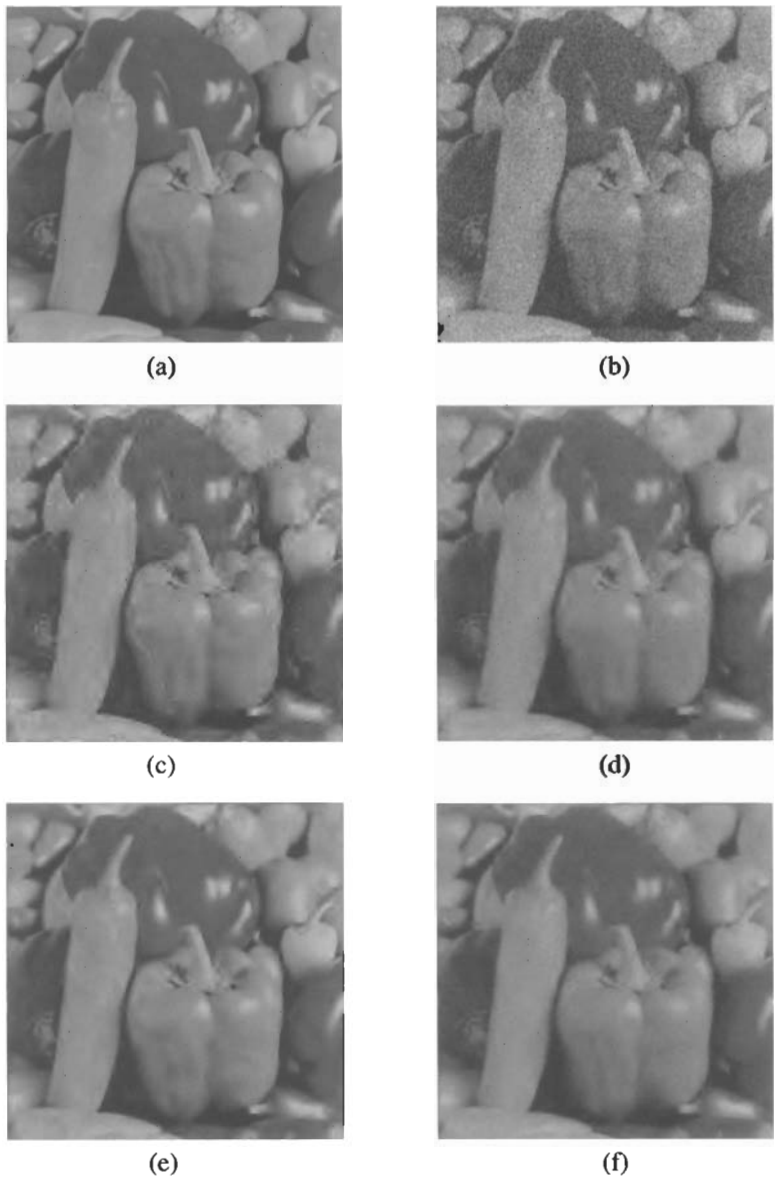


FIGURE 10.6 (a): Original image. (b): Noisy image (SNR = 28.6db). (c): Estimation with a hard thresholding in a separable wavelet basis (Symmlet 4), (SNR = 31.8db). (d): Soft thresholding (SNR = 31.1db). (e): Translation invariant hard thresholding (SNR = 34.3db). (f): Translation invariant soft thresholding (SNR = 31.7db).

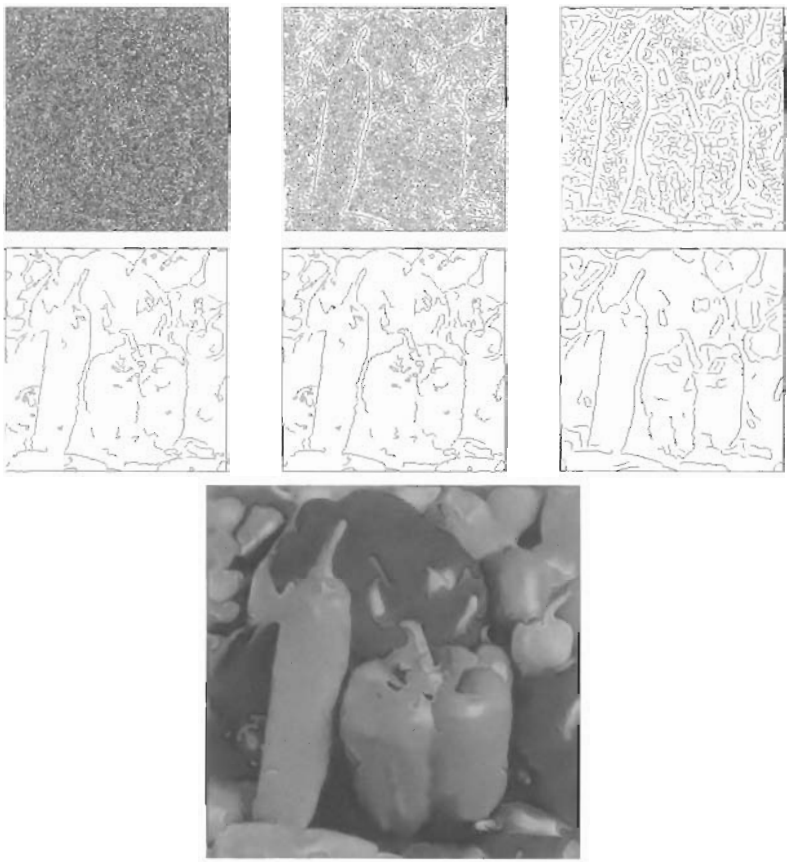


FIGURE 10.7 The first row shows the wavelet modulus maxima of the noisy image 10.6(b). The scale increases from left to right, from 2^{-7} to 2^{-5}. The chains of modulus maxima selected by the thresholding procedure are shown below. The bottom image is reconstructed from the selected modulus maxima at all scales.

because they produce many coefficients whose amplitudes are at the same level as the noise. To restore textures, it is often necessary to use stochastic models of textures, with parameters that can be estimated in presence of noise. Constructing such models is however difficult, as explained in Section 5.5.3.

Multiscale Edge Estimation Section 9.3 explains that wavelet bases are not optimal for approximating images because they do not take advantage of the geometrical regularity of edges. Understanding how to use the geometrical image regularity to enhance wavelet estimations is a difficult open issue. One approach implemented by Hwang and Mallat [258] is to regularize the multiscale edge

representation of Section 6.3. In many images, discontinuities belong to regular geometrical curves that are the edges of important structures. Along an edge, the wavelet coefficients change slowly and their estimation can thus be improved with an averaging.

The image is decomposed with a two-dimensional dyadic wavelet transform, whose modulus maxima locate the multiscale edges. At each scale 2^j, the chaining algorithm of Section 6.3.1 links the wavelet maxima to build edge curves. Instead of thresholding each wavelet maxima independently, the thresholding is performed over contours. An edge curve is removed if the average wavelet maxima amplitude is below $T = 3\sigma$. Prior geometrical information can also be used to refine the edge selection. Important image structures may generate long contours, which suggests removing short edge curves that are likely to be created by noise. The first line of Figure 10.7 shows the modulus maxima of the noisy image. The edges selected by the thresholding are shown below. At the finest scale shown on the left, the noise is masking the image structures. Edges are therefore selected by using the position of contours at the previous scale.

The thresholded wavelet maxima are regularized along the edges with an averaging. A restored image is recovered from the resulting wavelet maxima, using the reconstruction algorithm of Section 6.2.2. Figure 10.7 shows an example of an image restored from regularized multiscale edges. Edges are visually well recovered but textures and fine structures are removed by the thresholding based on the amplitude and length of the maxima chains. This produces a cartoon-like image.

10.2.5 Best Basis Thresholding [3]

When the additive noise W is white, the performance of a thresholding estimation depends on its ability to efficiently approximate the signal f with few basis vectors. Section 9.4 explains that a single basis is often not able to approximate well all signals of a large class. It is then necessary to adapt the basis to the signal [242]. We study applications of adaptive signal decompositions to thresholding estimation.

Best Orthogonal Basis Sections 8.1 and 8.5 construct dictionaries $\mathcal{D} = \cup_{\lambda \in \Lambda} \mathcal{B}^\lambda$ where each $\mathcal{B}^\lambda = \{g_m^\lambda\}_{0 \leq m < N}$ is a wavelet packet or a local cosine orthogonal basis. These dictionaries have $P = N \log_2 N$ distinct vectors but include more than $2^{N/2}$ different orthogonal bases by recombining these vectors.

An estimation of f from the noisy measurements $X = f + W$ is obtained by thresholding the decomposition of X in $\mathcal{B}^\lambda$:

$$\tilde{F}^\lambda = \sum_{m=0}^{N-1} \rho_T(\langle X, g_m^\lambda \rangle)\, g_m^\lambda.$$

The ideal basis $\mathcal{B}^\alpha$ is the one that minimizes the average estimation error

$$\mathsf{E}\{\|f - \tilde{F}^\alpha\|^2\} = \min_{\lambda \in \Lambda} \mathsf{E}\{\|f - \tilde{F}^\lambda\|^2\}. \tag{10.71}$$

In practice, we cannot find this ideal basis since we do not know f. Instead, we estimate the risk $\mathrm{E}\{\|f-\tilde{F}^{\lambda}\|^2\}$ in each basis $\mathcal{B}^{\lambda}$, and choose the best empirical basis that minimizes the estimated risk.

Threshold Value If we wish to choose a basis adaptively, we must use a higher threshold T than the threshold value $\sigma\sqrt{2\log_e N}$ used when the basis is set in advance. Indeed, an adaptive basis choice may also find vectors that better correlate the noise components. Let us consider the particular case $f=0$. To ensure that the estimated signal is close to zero, since $X=W$, we must choose a threshold T that has a high probability of being above all the inner products $|\langle W, g_m^{\lambda}\rangle|$ with all vectors in the dictionary $\mathcal{D}$. For a dictionary including P distinct vectors, for P large there is a negligible probability for the noise coefficients to be above

$$T=\sigma\sqrt{2\log_e P}. \tag{10.72}$$

This threshold is however not optimal and smaller values can improve the risk.

Basis Choice For a soft thresholding, (10.58) defines an estimator $\tilde{r}_t^{\lambda}(f,T)$ of the risk $r_t^{\lambda}(f,T)=\mathrm{E}\{\|f-\tilde{F}^{\lambda}\|^2\}$:

$$\tilde{r}_t^{\lambda}(T,f)=\sum_{m=1}^{N-1}\Phi(|\langle X, g_m^{\lambda}\rangle|^2), \tag{10.73}$$

with

$$\Phi(u)=\begin{cases} u-\sigma^2 & \text{if } u\le T^2 \\ \sigma^2+T^2 & \text{if } u>T^2 \end{cases}. \tag{10.74}$$

Theorem 10.5 proves that this estimator is unbiased.

The empirical best basis $\mathcal{B}^{\tilde{\alpha}}$ for estimating f is obtained by minimizing the estimated risk

$$\tilde{r}_t^{\tilde{\alpha}}(T,f)=\min_{\lambda\in\Lambda}\tilde{r}_t^{\lambda}(T,f)\,. \tag{10.75}$$

The estimated risk is calculated in (10.73) as an additive cost function over the noisy coefficients. The fast algorithm of Section 9.4.2 can thus find the best basis $\mathcal{B}^{\tilde{\alpha}}$ in wavelet packet or local cosine dictionaries, with $O(N\log_2 N)$ operations. Figure 10.8(d) shows the estimation of a sound recording "grea" in the presence of a white noise with an SNR of 8.7db. A best empirical local cosine basis is chosen by the minimization (10.75) and is used to decompose the noisy signal. This best basis is composed of local cosine vectors having a time and a frequency resolution adapted to the transients and harmonic structures of the signal. A hard thresholding is performed and the Heisenberg boxes of the remaining coefficients are shown in Figure 10.8(c).

Donoho and Johnstone [166] prove that for $T=\sigma\sqrt{2\log_e P}$ the risk $\mathrm{E}\{\|f-\tilde{F}^{\tilde{\alpha}}\|^2\}$ in the empirical best basis $\mathcal{B}^{\tilde{\alpha}}$ is within a $\log_e N$ factor of the minimum risk $\mathrm{E}\{\|f-\tilde{F}^{\alpha}\|^2\}$ in the ideal best basis $\mathcal{B}^{\alpha}$. In that sense, the best basis algorithm is guaranteed to find a nearly optimal basis.

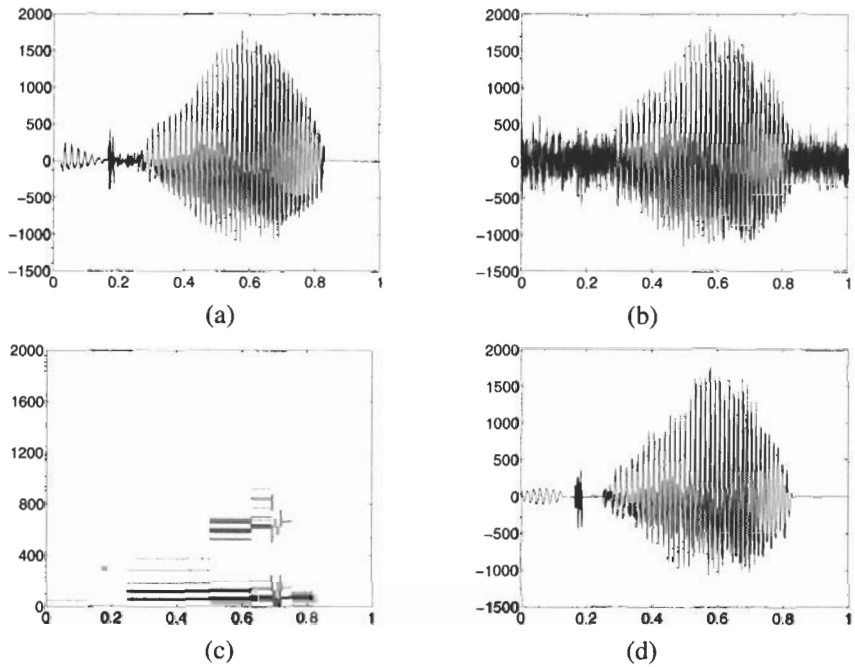


FIGURE 10.8 (a): Speech recording of "grea." (b): Noisy signal (SNR = 8.7db). (c): Heisenberg boxes of the local coefficients above the threshold in the best basis. (d): Estimated signal recovered from the thresholded local cosine coefficients (SNR = 10.9db).

Cost of Adaptivity An approximation in a basis that is adaptively selected is necessarily more precise than an approximation in a basis chosen a priori. However, in the presence of noise, estimations by thresholding may not be improved by an adaptive basis choice. Indeed, using a dictionary of several orthonormal bases requires raising the threshold, because the larger number of dictionary vectors produces a higher correlation peak with the noise. The higher threshold removes more signal components, unless it is compensated by the adaptivity, which can better concentrate the signal energy over few coefficients. The same issue appears in parametrized estimations, where increasing the number of parameters may fit the noise and thus degrade the estimation.

For example, if the original signal is piecewise smooth, then a best wavelet packet basis does not concentrate the signal energy much more efficiently than a wavelet basis. In the presence of noise, in regions where the noise dominates the signal, the best basis algorithm may optimize the basis to fit the noise. This is why the threshold value must be increased. Hence, the resulting best basis estimation is not as precise as a thresholding in a fixed wavelet basis with a lower threshold.

However, for oscillatory signals such as the speech recording in Figure 10.8(a), a best local cosine basis concentrates the signal energy over much fewer coefficients than a wavelet basis, and thus provides a better estimation.

10.3 MINIMAX OPTIMALITY [3]

We consider the noisy data $X = f + W$, where W is a Gaussian white noise of variance σ^2. An estimation $\tilde{F} = DX$ of f has a risk $r(D, f) = \mathrm{E}\{\|DX - f\|^2\}$. If some prior information tells us that the signal we estimate is in a set Θ, then we must construct estimators whose maximum risk over Θ is as small as possible. Let $r(D, \Theta) = \sup_{f \in \Theta} r(D, f)$ be the maximum risk over Θ. The *linear minimax risk* and *non-linear minimax risk* are respectively defined by

$$r_l(\Theta) = \inf_{D \in \mathcal{O}_l} r(D, \Theta) \text{ and } r_n(\Theta) = \inf_{D \in \mathcal{O}_n} r(D, \Theta),$$

where $\mathcal{O}_l$ is the set of all linear operators from $\mathbb{C}^N$ to $\mathbb{C}^N$ and $\mathcal{O}_n$ is the set of all linear and non-linear operators from $\mathbb{C}^N$ to $\mathbb{C}^N$. We study operators D that are diagonal in an orthonormal basis $\mathcal{B} = \{g_m\}_{0 \le m < N}$:

$$\tilde{F} = DX = \sum_{m=0}^{N-1} d_m(X_{\mathcal{B}}[m])\, g_m ,$$

and find conditions to achieve a maximum risk over Θ that is close to the minimax risk. The values of $r_l(\Theta)$ and $r_n(\Theta)$ are compared, so that we can judge whether it is worthwhile using non-linear operators.

Section 10.3.1 begins by studying linear diagonal operators. For orthosymmetric sets, Section 10.3.2 proves that the linear and non-linear minimax risks are nearly achieved by diagonal operators. As a consequence, thresholding estimators in a wavelet basis are proved to be nearly optimal for signals and images having a bounded variation. Readers more interested by algorithms and numerical applications may skip this section, which is mathematically more involved.

10.3.1 Linear Diagonal Minimax Estimation

An estimator that is linear and diagonal in the basis $\mathcal{B}$ can be written

$$\tilde{F} = DX = \sum_{m=0}^{N-1} a[m]\, X_{\mathcal{B}}[m]\, g_m , \tag{10.76}$$

where each $a[m]$ is a constant. Let $\mathcal{O}_{l,d}$ be the set of all such linear diagonal operators D. Since $\mathcal{O}_{l,d} \subset \mathcal{O}_l$, the *linear diagonal minimax risk* is larger than the linear minimax risk

$$r_{l,d}(\Theta) = \inf_{D \in \mathcal{O}_{l,d}} r(D, \Theta) \ge r_l(\Theta) .$$

We characterize diagonal estimators that achieve the minimax risk $r_{l,d}(\Theta)$. If Θ is translation invariant, we prove that $r_{l,d}(\Theta) = r_l(\Theta)$ in a discrete Fourier basis. This risk is computed for bounded variation signals.

Quadratic Convex Hull The "square" of a set Θ in the basis $\mathcal{B}$ is defined by

$$(\Theta)_{\mathcal{B}}^2 = \{\tilde{f} \ : \ \tilde{f} = \sum_{m=0}^{N-1} |f_{\mathcal{B}}[m]|^2 \, g_m \ \text{ with } f \in \Theta\} \, . \tag{10.77}$$

We say that Θ is *quadratically convex* in $\mathcal{B}$ if $(\Theta)_{\mathcal{B}}^2$ is a convex set. A hyperrectangle $\mathcal{R}_x$ in $\mathcal{B}$ of vertex $x \in \mathbb{C}^N$ is a simple example of quadratically convex set defined by

$$\mathcal{R}_x = \left\{ f \ : \ |f_{\mathcal{B}}[m]| \leq |x_{\mathcal{B}}[m]| \text{ for } 0 \leq m < N \right\} .$$

The *quadratic convex hull* $\mathrm{QH}[\Theta]$ of Θ in the basis $\mathcal{B}$ is defined by

$$\mathrm{QH}[\Theta] = \left\{ f \ : \ \sum_{m=0}^{N-1} |f_{\mathcal{B}}[m]|^2 \text{ is in the convex hull of } (\Theta)_{\mathcal{B}}^2 \right\} . \tag{10.78}$$

It is the largest set whose square $(\mathrm{QH}[\Theta])_{\mathcal{B}}^2$ is equal to the convex hull of $(\Theta)_{\mathcal{B}}^2$.

The risk of an oracle attenuation (10.28) gives a lower bound of the minimax linear diagonal risk $r_{l,d}(\Theta)$:

$$r_{l,d}(\Theta) \geq r_{\inf}(\Theta) = \sup_{f \in \Theta} \sum_{m=0}^{N-1} \frac{\sigma^2 \, |f_{\mathcal{B}}[m]|^2}{\sigma^2 + |f_{\mathcal{B}}[m]|^2} \ . \tag{10.79}$$

The following theorem proves that this inequality is an equality if Θ is quadratically convex.

Theorem 10.6 *If Θ is a bounded and closed set, then there exists $x \in \mathrm{QH}[\Theta]$ such that $r_{\inf}(x) = r_{\inf}(\mathrm{QH}[\Theta])$ in the basis $\mathcal{B}$. Moreover, the linear diagonal operator D defined by*

$$a[m] = \frac{|x_{\mathcal{B}}[m]|^2}{\sigma^2 + |x_{\mathcal{B}}[m]|^2} \, , \tag{10.80}$$

achieves the linear diagonal minimax risk

$$r(D, \Theta) = r_{l,d}(\Theta) = r_{\inf}\left(\mathrm{QH}[\Theta]\right) . \tag{10.81}$$

Proof[3]. The risk $r(D, f)$ of the diagonal operator (10.76) is

$$r(D, f) = \sum_{m=0}^{N-1} \left(\sigma^2 \, |a[m]|^2 + |1 - a[m]|^2 \, |f_{\mathcal{B}}[m]|^2 \right) . \tag{10.82}$$

Since it is a linear function of $|f_{\mathcal{B}}[m]|^2$, it reaches the same maximum in Θ and in $\mathrm{QH}[\Theta]$. This proves that $r(D, \Theta) = r(D, \mathrm{QH}[\Theta])$ and hence that $r_{l,d}(\Theta) = r_{l,d}\left(\mathrm{QH}[\Theta]\right)$.

To verify that $r_{l,d}(\Theta) = r_{\inf}\left(\mathrm{QH}[\Theta]\right)$ we prove that $r_{l,d}(\mathrm{QH}[\Theta]) = r_{\inf}(\mathrm{QH}[\Theta])$. Since (10.79) shows that $r_{\inf}(\mathrm{QH}[\Theta]) \leq r_{l,d}(\mathrm{QH}[\Theta])$ to get the reverse inequality, it is

sufficient to prove that the linear estimator defined by (10.80) satisfies $r(D, \mathrm{QH}[\Theta]) \leq r_{\inf}(\mathrm{QH}[\Theta])$. Since Θ is bounded and closed, $\mathrm{QH}[\Theta]$ is also bounded and closed and thus compact, which guarantees the existence of $x \in \mathrm{QH}[\Theta]$ such that $r_{\inf}(x) = r_{\inf}(\mathrm{QH}[\Theta])$. The risk of this estimator is calculated with (10.82):

$$\begin{aligned} r(D,f) &= \sum_{m=0}^{N-1} \frac{|f_{\mathcal{B}}[m]|^2\sigma^4 + \sigma^2|x_{\mathcal{B}}[m]|^4}{(\sigma^2+|x_{\mathcal{B}}[m]|^2)^2} \\ &= \sum_{m=0}^{N-1} \frac{\sigma^2|x_{\mathcal{B}}[m]|^2}{\sigma^2+|x_{\mathcal{B}}[m]|^2} + \sigma^4 \sum_{m=0}^{N-1} \frac{|f_{\mathcal{B}}[m]|^2 - |x_{\mathcal{B}}[m]|^2}{(\sigma^2+|x_{\mathcal{B}}[m]|^2)^2} . \end{aligned}$$

To show that $r(D,f) \leq r_{\inf}(\mathrm{QH}[\Theta])$, we verify that the second summation is negative. Let $0 \leq \eta \leq 1$ and y be a vector whose decomposition coefficients in $\mathcal{B}$ satisfy

$$|y_{\mathcal{B}}[m]|^2 = (1-\eta)\,|x_{\mathcal{B}}[m]|^2 + \eta\,|f_{\mathcal{B}}[m]|^2 .$$

Since $\mathrm{QH}[\Theta]$ is quadratically convex, necessarily $y \in \mathrm{QH}[\Theta]$ so

$$J(\eta) = \sum_{m=0}^{N-1} \frac{\sigma^2\,|y_{\mathcal{B}}[m]|^2}{\sigma^2+|y_{\mathcal{B}}[m]|^2} \leq \sum_{m=0}^{N-1} \frac{\sigma^2\,|x_{\mathcal{B}}[m]|^2}{\sigma^2+|x_{\mathcal{B}}[m]|^2} = J(0).$$

Since the maximum of $J(\eta)$ is at $\eta = 0$,

$$J'(0) = \sum_{m=0}^{N-1} \frac{|f_{\mathcal{B}}[m]|^2 - |x_{\mathcal{B}}[m]|^2}{(\sigma^2+|x_{\mathcal{B}}[m]|^2)^2} \leq 0,$$

which finishes the proof. ■

This theorem implies that $r_{l,d}(\Theta) = r_{l,d}(\mathrm{QH}[\Theta])$. To take advantage of the fact that Θ may be much smaller than its quadratic convex hull, it is necessary to use non-linear diagonal estimators.

Translation Invariant Set Signals such as sounds or images are often arbitrarily translated in time or in space, depending on the beginning of the recording or the position of the camera. To simplify border effects, we consider signals of period N. We say that Θ is *translation invariant* if for any $f[n] \in \Theta$ then $f[n-p] \in \Theta$ for all $0 \leq p < N$.

If the set is translation invariant and the noise is stationary, then we show that the best linear estimator is also translation invariant, which means that it is a convolution. Such an operator is diagonal in the discrete Fourier basis $\mathcal{B} = \{g_m[n] = \frac{1}{\sqrt{N}} \exp(i2\pi mn/N)\}_{0 \leq m < N}$. The decomposition coefficients of f in this basis are proportional to its discrete Fourier transform:

$$f_{\mathcal{B}}[m] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} f[n] \exp\left(\frac{-i2\pi mn}{N}\right) = \frac{\hat{f}[m]}{\sqrt{N}} .$$

For a set Θ, the lower bound $r_{\inf}(\Theta)$ in (10.79) becomes

$$r_{\inf}(\Theta) = \sup_{f \in \Theta} \sum_{m=0}^{N-1} \frac{\sigma^2 N^{-1} |\hat{f}[m]|^2}{\sigma^2 + N^{-1} |\hat{f}[m]|^2} .$$

The following theorem proves that diagonal operators in the discrete Fourier basis achieve the linear minimax risk.

Theorem 10.7 *Let Θ be a closed and bounded set. Let $x \in \mathrm{QH}[\Theta]$ be such that $r_{\inf}(x) = r_{\inf}(\mathrm{QH}[\Theta])$ and*

$$\hat{h}[m] = \frac{|\hat{x}[m]|^2}{N\sigma^2 + |\hat{x}[m]|^2} . \tag{10.83}$$

If Θ is translation invariant then $\tilde{F} = DX = X \circledast h$ achieves the linear minimax risk

$$r_l(\Theta) = r(D, \Theta) = r_{\inf}(\mathrm{QH}[\Theta]) . \tag{10.84}$$

Proof[3]. Since $r_l(\Theta) \leq r_{l,d}(\Theta)$ Theorem 10.6 proves in (10.81) that

$$r_l(\Theta) \leq r_{\inf}(\mathrm{QH}[\Theta]) .$$

Moreover, the risk $r_{\inf}(\mathrm{QH}[\Theta])$ is achieved by the diagonal estimator (10.80). In the discrete Fourier basis it corresponds to a circular convolution whose transfer function is given by (10.83).

We show that $r_l(\Theta) \geq r_{\inf}(\mathrm{QH}[\Theta])$ by using particular Bayes priors. If $f \in \mathrm{QH}[\Theta]$ then there exists a family $\{f_i\}_i$ of elements in Θ such that for any $0 \leq m < N$,

$$|\hat{f}[m]|^2 = \sum_i p_i |\hat{f}_i[m]|^2 \quad \text{with} \quad \sum_i p_i = 1 .$$

To each $f_i \in \Theta$ we associate a random shift vector $F_i[n] = f_i[n - P_i]$ as in (9.19). Each $F_i[n]$ is circular stationary, and its power spectrum is computed in (9.21): $\hat{R}_{F_i}[m] = N^{-1}|\hat{f}_i[m]|^2$. Let F be a random vector that has a probability p_i to be equal to F_i. It is circular stationary and its power spectrum is $\hat{R}_F[m] = N^{-1}|\hat{f}[m]|^2$. We denote by π_f the probability distribution of F. The risk $r_l(\pi_f)$ of the Wiener filter is calculated in (10.13):

$$r_l(\pi_f) = \sum_{m=0}^{N-1} \frac{\hat{R}_F[m]\hat{R}_W[m]}{\hat{R}_F[m] + \hat{R}_W[m]} = \sum_{m=0}^{N-1} \frac{N^{-1}|\hat{f}[m]|^2 \sigma^2}{N^{-1}|\hat{f}[m]|^2 + \sigma^2} . \tag{10.85}$$

Since Θ is translation invariant, the realizations of F are in Θ, so $\pi_f \in \Theta^*$. The minimax Theorem 10.3 proves in (10.19) that $r_l(\pi_f) \leq r_l(\Theta)$. Since this is true for any $f \in \mathrm{QH}[\Theta]$, taking a sup with respect to f in (10.85) proves that $r_l(\mathrm{QH}[\Theta]) \leq r_l(\Theta)$, which finishes the proof. ■

Bounded Variation Signals The total variation defined in (2.60) measures the amplitude of all signal oscillations. Bounded variation signals may include sharp transitions such as discontinuities. A set Θ_V of bounded variation signals of period N is defined by

$$\Theta_V = \{f \,:\, \|f\|_V = \sum_{n=0}^{N-1} |f[n] - f[n-1]| \leq C\} \,. \tag{10.86}$$

Since Θ_V is translation invariant, the linear minimax estimator is diagonal in the discrete Fourier basis. The following proposition computes the minimax linear risk, which is renormalized by the noise energy $\mathrm{E}\{\|W\|^2\} = N\sigma^2$.

Proposition 10.5 *If* $1 \leq C/\sigma \leq N^{1/2}$ *then*

$$\frac{r_l(\Theta_V)}{N\sigma^2} \sim \frac{C}{\sigma N^{1/2}}. \tag{10.87}$$

Proof[3]. The set Θ_V is translation invariant but it is not bounded because we do not control the average of a bounded variation signal. However, one can verify with a limit argument that the equality $r_l(\Theta_V) = r_{\inf}(\mathrm{QH}[\Theta_V])$ of Theorem 10.7 is still valid. To compute $r_{\inf}(\mathrm{QH}[\Theta_V])$ we show that Θ_V is included in a hyperrectangle $\mathcal{R}_x = \{f \,:\, |\hat{f}[m]| \leq |\hat{x}[m]|\}$, by computing an upper bound of $|\hat{f}[m]|$ for each $f \in \Theta_V$. Let $g[n] = f[n] - f[n-1]$. Its discrete Fourier transform satisfies

$$|\hat{g}[m]| = |\hat{f}[m]| \left|1 - \exp\left(\frac{-i2\pi m}{N}\right)\right| = 2|\hat{f}[m]| \left|\sin\frac{\pi m}{N}\right| \,. \tag{10.88}$$

Since $\sum_{n=0}^{N-1} |g[n]| \leq C$, necessarily $|\hat{g}[m]| \leq C$ so

$$|\hat{f}[m]|^2 \leq \frac{C^2}{4|\sin(\pi m/N)|^2} = |\hat{x}[m]|^2 \,, \tag{10.89}$$

which proves that $\Theta_V \subset \mathcal{R}_x$. The value $|\hat{x}[0]| = \infty$ is formally treated like all others. Since $\mathcal{R}_x$ is quadratically convex, $\mathrm{QH}[\Theta_V] \subset \mathcal{R}_x$. Hence

$$r_{\inf}(\mathrm{QH}[\Theta_V]) \leq r_{\inf}(\mathcal{R}_x) = \sum_{m=0}^{N-1} \frac{\sigma^2 N^{-1} |\hat{x}[m]|^2}{\sigma^2 + N^{-1} |\hat{x}[m]|^2} \,,$$

with $\sigma^2 N^{-1} |\hat{x}[0]|^2 (\sigma^2 + N^{-1} |\hat{x}[0]|^2)^{-1} = \sigma^2$. Since $|\hat{x}[m]| \sim CN\,|m|^{-1}$ and $1 \leq C/\sigma \leq N^{1/2}$, a direct calculation shows that

$$r_{\inf}(\mathrm{QH}[\Theta_V]) \leq r_{\inf}(\mathcal{R}_x) \sim CN^{1/2}\sigma \,. \tag{10.90}$$

To compute a lower bound for $r_{\inf}(\mathrm{QH}[\Theta_V])$ we consider the two signals in Θ_V defined by

$$f_1 = \frac{C}{2}\mathbf{1}_{[0,N/2-1]} - \frac{C}{2} \quad \text{and} \quad f_2 = \frac{C}{4}\mathbf{1}_{[0,N/2-1]} - \frac{C}{4}\mathbf{1}_{[N/2,N-1]} \,.$$

Let $f \in \mathrm{QH}[\Theta_V]$ such that

$$|\hat{f}[m]|^2 = \frac{1}{2}(|\hat{f}_1[m]|^2 + |\hat{f}_2[m]|^2).$$

A simple calculation shows that for $m \neq 0$

$$|\hat{f}[m]|^2 = \frac{C^2}{8|\sin(\pi m/N)|^2} \sim C^2 N^2 |m|^{-2}$$

so

$$r_{\inf}(\mathrm{QH}[\Theta_V]) \geq r_{\inf}(f) \sim C N^{1/2} \sigma .$$

Together with (10.90) this proves (10.87). ■

This theorem proves that a linear estimator reduces the energy of the noise by a factor that increases like $N^{1/2}$. The minimax filter averages the noisy data to remove part of the white noise, without degrading too much the potential discontinuities of f. Figure 10.2(c) shows a linear Wiener estimation calculated by supposing that $|\hat{f}[m]|^2$ is known. The resulting risk (10.17) is in fact the minimax risk over the translation invariant set $\Theta_f = \{g : g[n] = f[n-p] \text{ with } p \in \mathbb{Z}\}$. If f has a discontinuity whose amplitude is on the order of C then although the set Θ_f is much smaller than Θ_V, the minimax linear risks $r_l(\Theta_f)$ and $r_l(\Theta_V)$ are of the same order.

10.3.2 Orthosymmetric Sets

We study geometrical conditions on Θ that allow us to nearly reach the non-linear minimax risk $r_n(\Theta)$ with estimators that are diagonal in a basis $\mathcal{B} = \{g_m\}_{0 \leq m < N}$. The maximum risk on Θ of any linear or non-linear diagonal estimator has a lower bound calculated with the oracle diagonal attenuation (10.28):

$$r_{\inf}(\Theta) = \sup_{f \in \Theta} \sum_{m=0}^{N-1} \frac{\sigma^2 |f_{\mathcal{B}}[m]|^2}{\sigma^2 + |f_{\mathcal{B}}[m]|^2} .$$

Thresholding estimators have a maximum risk that is close to this lower bound. We thus need to understand under what conditions $r_n(\Theta)$ is on the order of $r_{\inf}(\Theta)$ and how it compares with $r_l(\Theta)$.

Hyperrectangle The study begins with hyperrectangles which are building blocks for computing the minimax risk over any set Θ. A hyperrectangle

$$\mathcal{R}_x = \{f : |f_{\mathcal{B}}[m]| \leq |x_{\mathcal{B}}[m]| \text{ for } 0 \leq m < N\}$$

is a separable set along the basis directions g_m. The risk lower bound for diagonal estimators is

$$r_{\inf}(\mathcal{R}_x) = \sum_{m=0}^{N-1} \frac{\sigma^2 |x_{\mathcal{B}}[m]|^2}{\sigma^2 + |x_{\mathcal{B}}[m]|^2} .$$

The following theorem proves that for a hyperrectangle, the non-linear minimax risk is very close to the linear minimax risk.

Theorem 10.8 *On a hyperrectangle $\mathcal{R}_x$ the linear and non-linear minimax risks are reached by diagonal estimators. They satisfy*

$$r_l(\mathcal{R}_x) = r_{\inf}(\mathcal{R}_x) \ , \tag{10.91}$$

and

$$\mu\, r_{\inf}(\mathcal{R}_x) \leq r_n(\mathcal{R}_x) \leq r_{\inf}(\mathcal{R}_x) \ \textit{with}\ \mu \leq 1/1.25 \ . \tag{10.92}$$

Proof[3]. We first show that a linear minimax estimator is necessarily diagonal in $\mathcal{B}$. Let $\tilde{F} = DX$ be the estimator obtained with a linear operator D represented by the matrix A in $\mathcal{B}$:

$$\tilde{F}_{\mathcal{B}} = A X_{\mathcal{B}} \ .$$

Let $\mathrm{tr}A$ be the trace of A, and A^* be its complex transpose. Since $X = f + W$ where W is a white noise of variance σ^2 , a direct calculation shows that

$$r(D,f) = \mathsf{E}\{\|\tilde{F} - f\|^2\} = \sigma^2\, \mathrm{tr}AA^* + (Af_{\mathcal{B}} - f_{\mathcal{B}})^*\,(Af_{\mathcal{B}} - f_{\mathcal{B}}). \tag{10.93}$$

If D_d is the diagonal operator whose coefficients are $a[m] = a_{m,m}$ the risk is then

$$r(D_d,f) = \sum_{m=0}^{N-1} \left(\sigma^2\, |a_{m,m}|^2 + |1 - a_{m,m}|^2\, |f_{\mathcal{B}}[m]|^2 \right) . \tag{10.94}$$

To prove that the maximum risk over $\mathcal{R}_x$ is minimized when A is diagonal, we show that $r(D_d,\mathcal{R}_x) \leq r(D,\mathcal{R}_x)$. For this purpose, we use a prior probability distribution $\pi \in \mathcal{R}_x^*$ corresponding to a random vector F whose realizations are in $\mathcal{R}_x$:

$$F_{\mathcal{B}}[m] = S[m]\, x_{\mathcal{B}}[m] \ . \tag{10.95}$$

The random variables $S[m]$ are independent and equal to 1 or -1 with probability $1/2$. The expected risk $r(D,\pi) = \mathsf{E}\{\|F - \tilde{F}\|^2\}$ is derived from (10.93) by replacing f by F and taking the expected value with respect to the probability distribution π of F. If $m \neq p$ then $\mathsf{E}\{F_{\mathcal{B}}[m]\, F_{\mathcal{B}}[p]\} = 0$ so we get

$$\begin{aligned} r(D,\pi) &= \sigma^2 \sum_{m=0}^{N-1} |a_{m,m}|^2 + \sum_{m=0}^{N-1} |x_{\mathcal{B}}[m]|^2 \left[|a_{m,m} - 1|^2 + \sum_{\substack{p=0 \\ p \neq m}}^{N-1} |a_{m,p}|^2 \right] \\ &\geq \sigma^2 \sum_{m=0}^{N-1} |a_{m,m}|^2 + \sum_{m=0}^{N-1} |1 - a_{m,m}|^2 |x_{\mathcal{B}}[m]|^2 \ = r(D_d,x). \end{aligned} \tag{10.96}$$

Since the realizations of F are in $\mathcal{R}_x$, (10.20) implies that $r(D,\mathcal{R}_x) \geq r(D,\pi)$, so $r(D,\mathcal{R}_x) \geq r(D_d,x)$. To prove that $r(D,\mathcal{R}_x) \geq r(D_d,\mathcal{R}_x)$ it is now sufficient to verify that $r(D_d,\mathcal{R}_x) = r(D_d,x)$. To minimize $r(D_d,f)$, (10.94) proves that necessarily $a_{m,m} \in [0,1]$. In this case (10.94) implies

$$r(D_d,\mathcal{R}_x) = \sup_{f \in \mathcal{R}_x} r(D_d,f) = r(D_d,x) \ .$$

Now that we know that the minimax risk is achieved by a diagonal operator, we apply Theorem 10.6 which proves in (10.81) that the minimax risk among linear diagonal operator is $r_{\inf}(\mathcal{R}_x)$ because $\mathcal{R}_x$ is quadratically convex. So $r_l(\mathcal{R}_x) = r_{\inf}(\mathcal{R}_x)$.

To prove that the non-linear minimax risk is also obtained with a diagonal operator we use the minimax Theorem 10.3 which proves that

$$r_n(\mathcal{R}_x) = \sup_{\pi \in \mathcal{R}_x^*} \inf_{D \in \mathcal{O}_n} r(D, \pi) \, . \tag{10.97}$$

The set $\mathcal{R}_x$ can be written as a product of intervals along each direction g_m. As a consequence, to any prior $\pi \in \mathcal{R}_x^*$ corresponding to a random vector F we associate a prior $\pi' \in \mathcal{R}_x^*$ corresponding to F' such that $F'_{\mathcal{B}}[m]$ has the same distribution as $F_{\mathcal{B}}[m]$ but with $F'_{\mathcal{B}}[m]$ independent from $F'_{\mathcal{B}}[p]$ for $p \neq m$. We then verify that for any operator D, $r(D, \pi) \leq r(D, \pi')$. The sup over $\mathcal{R}_x^*$ in (10.97) can thus be restricted to processes that have independent coordinates. This independence also implies that the Bayes estimator that minimizes $r(D, \pi)$ is diagonal in $\mathcal{B}$. The minimax theorem proves that the minimax risk is reached by diagonal estimators.

Since $r_n(\mathcal{R}_x) \leq r_l(\mathcal{R}_x)$ we derive the upper bound in (10.92) from the fact that $r_l(\mathcal{R}_x) = \mathcal{R}_{\inf}(\mathcal{R}_x)$. The lower bound (10.92) is obtained by computing the Bayes risk $r_n(\pi) = \inf_{D \in \mathcal{O}_n} r(D, \pi)$ for the prior π corresponding to F defined in (10.95), and verifying that $r_n(\pi) \geq \mu r_{\inf}(\mathcal{R}_x)$. We see from (10.97) that $r_n(\mathcal{R}_x) \geq r_n(\pi)$, which implies (10.92). ■

The bound $\mu > 0$ was proved by Ibragimov and Khas'minskii [219] but the essentially sharp bound $1/1.25$ was obtained by Donoho, Liu and MacGibbon [172]. They showed that μ depends on the variance σ^2 of the noise and that if σ^2 tends to 0 or to $+\infty$ then μ tends to 1. Linear estimators are thus asymptotically optimal compared to non-linear estimators.

Orthosymmetric set To differentiate the properties of linear and non-linear estimators, we consider more complex sets that can be written as unions of hyperrectangles. We say that Θ is *orthosymmetric* in $\mathcal{B}$ if for any $f \in \Theta$ and for any $a[m]$ with $|a[m]| \leq 1$ then

$$\sum_{m=0}^{N-1} a[m] \, f_{\mathcal{B}}[m] \, g_m \in \Theta \, .$$

Such a set can be written as a union of hyperrectangles:

$$\Theta = \bigcup_{f \in \Theta} \mathcal{R}_f \, . \tag{10.98}$$

An upper bound of $r_n(\Theta)$ is obtained with the maximum risk $r_l(\Theta) = \sup_{f \in \Theta} r_l(f)$ of a hard or soft thresholding estimator in the basis $\mathcal{B}$, with a threshold $T = \sigma\sqrt{2\log_e N}$.

Proposition 10.6 *If Θ is orthosymmetric in $\mathcal{B}$ then the linear minimax estimator is reached by linear diagonal estimators and*

$$r_l(\Theta) = r_{\inf}(\mathrm{QH}[\Theta]) \, . \tag{10.99}$$

The non-linear minimax risk satisfies

$$\frac{1}{1.25} r_{\inf}(\Theta) \leq r_n(\Theta) \leq r_t(\Theta) \leq (2\log_e N + 1)\left(\sigma^2 + r_{\inf}(\Theta)\right). \tag{10.100}$$

Proof[2]. Since Θ is orthosymmetric, $\Theta = \cup_{f\in\Theta}\mathcal{R}_f$. On each hyperrectangle $\mathcal{R}_f$, we showed in (10.96) that the maximum risk of a linear estimator is reduced by letting it be diagonal in $\mathcal{B}$. The minimax linear estimation in Θ is therefore diagonal: $r_l(\Theta) = r_{l,d}(\Theta)$. Theorem 10.6 proves in (10.81) that $r_{l,d}(\Theta) = r_{\inf}(\mathrm{QH}[\Theta])$ which implies (10.99).

Since $\Theta = \cup_{f\in\Theta}\mathcal{R}_f$ we also derive that $r_n(\Theta) \geq \sup_{f\in\Theta} r_n(\mathcal{R}_f)$. So (10.92) implies that

$$r_n(\Theta) \geq \frac{1}{1.25} r_{\inf}(\Theta) .$$

Theorem 10.42 proves in (10.4) that the thresholding risk satisfies

$$r_t(f) \leq (2\log_e N + 1)\left(\sigma^2 + r_p(f)\right) .$$

A modification of the proof shows that this upper bound remains valid if $r_p(f)$ is replaced by $r_{\inf}(f)$ [167]. Taking a sup over all $f \in \Theta$ proves the upper bound (10.100), given that $r_n(\Theta) \leq r_t(\Theta)$. ■

This proposition shows that $r_n(\Theta)$ always remains within a factor $2\log_e N$ of the lower bound $r_{\inf}(\Theta)$ and that the thresholding risk $r_t(\Theta)$ is at most $2\log_e N$ times larger than $r_n(\Theta)$. In some cases, the factor $2\log_e N$ can even be reduced to a constant independent of N.

Unlike the nonlinear risk $r_n(\Theta)$, the linear minimax risk $r_l(\Theta)$ may be much larger than $r_{\inf}(\Theta)$. This depends on the convexity of Θ. If Θ is quadratically convex then $\Theta = \mathrm{QH}[\Theta]$ so (10.99) implies that $r_l(\Theta) = r_{\inf}(\Theta)$. Since $r_n(\Theta) \geq r_{\inf}(\Theta)/1.25$, the risk of linear and non-linear minimax estimators are of the same order. In this case, there is no reason for working with non-linear as opposed to linear estimators. When Θ is an orthosymmetric ellipsoid, Problem 10.14 computes the minimax linear estimator of Pinsker [282] and the resulting risk.

If Θ is not quadratically convex then its hull $\mathrm{QH}[\Theta]$ may be much bigger than Θ. This is the case when Θ has a star shape that is elongated in the directions of the basis vectors g_m, as illustrated in Figure 10.9. The linear risk $r_l(\Theta) = r_{\inf}(\mathrm{QH}[\Theta])$ may then be much larger than $r_{\inf}(\Theta)$. Since $r_n(\Theta)$ and $r_t(\Theta)$ are on the order of $r_{\inf}(\Theta)$, they are then much smaller than $r_l(\Theta)$. A thresholding estimator thus brings an important improvement over any linear estimator.

Example 10.2 Let Θ be an $\mathbf{l^p}$ ball defined by

$$\Theta = \{f \;:\; \sum_{m=0}^{N-1} \beta_m^p |f_{\mathcal{B}}[m]|^p \leq C^p\}. \tag{10.101}$$

It is an orthosymmetric set. Its square is

$$(\Theta)_{\mathcal{B}}^2 = \{f \;:\; \sum_{m=0}^{N-1} \beta_m^p |f_{\mathcal{B}}[m]|^{p/2} \leq C^p\}.$$

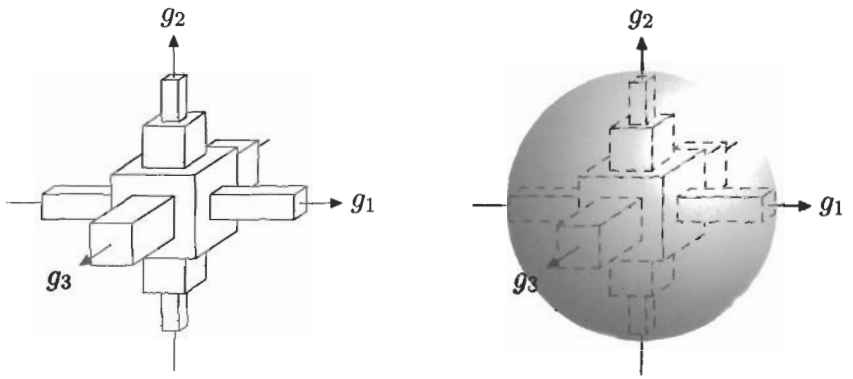


FIGURE 10.9 (a): Example of orthosymmetric set Θ in three dimensions. (b): The quadratically convex hull $\mathrm{QH}[\Theta]$ is a larger ellipsoid including Θ.

If $p \geq 2$ then $(\Theta)^2_{\mathcal{B}}$ is convex so Θ is quadratically convex. If $p < 2$, the convex hull of $(\Theta)^2_{\mathcal{B}}$ is $\{f : \sum_{m=0}^{N-1} \beta_m^2 |f_{\mathcal{B}}[m]| \leq C^2\}$ so the quadratic convex hull of Θ is

$$\mathrm{QH}[\Theta] = \{f : \sum_{m=0}^{N-1} \beta_m^2 |f_{\mathcal{B}}[m]|^2 \leq C^2\}. \tag{10.102}$$

The smaller p, the larger the difference between Θ and $\mathrm{QH}[\Theta]$.

Risk calculation The following proposition proves that $r_{\inf}(\Theta)$ depends on the error $\epsilon_n[M]$ of non-linear approximations of signals $f \in \Theta$ with M vectors selected from the basis $\mathcal{B}$.

Proposition 10.7 *Let $s > 1/2$ and C be such that $1 \leq C/\sigma \leq N^s$. If for each $f \in \Theta$ we have $\epsilon_n[M] \leq C^2 M^{1-2s}$ then*

$$r_{\inf}(\Theta) \leq 2 C^{1/s} \sigma^{2-1/s}. \tag{10.103}$$

In particular, if

$$\Theta_{C,s} = \left\{ f : \left(\sum_{m=0}^{N-1} |f_{\mathcal{B}}[m]|^{1/s} \right)^s \leq C \right\}, \tag{10.104}$$

then $r_{\inf}(\Theta_{C,s}) \sim C^{1/s} \sigma^{2-1/s}$.

Proof[2]. The same derivations as in the proof of Proposition 10.3 show that $r_p(f) \leq 2C^{1/s}\sigma^{2-1/s}$ for any $f \in \Theta$. Since $r_{\inf}(f) \leq r_p(f)$ we get (10.103) by taking a sup over Θ.

Theorem 9.5 together with Theorem 9.4 proves that any $f \in \Theta_{C,s}$ satisfies $\epsilon_n[M] \leq C^2 M^{1-2s}/(2s-1)$, which implies that $r_{\inf}(\Theta_{C,s}) = O(C^{1/s}\sigma^{2-1/s})$. To get a reverse

inequality we consider $f \in \Theta_{C,s}$ such that $|f_B[m]| = \sigma$ for $0 \leq m < \lfloor (C/\sigma)^{1/s} \rfloor$ and $f_B[m] = 0$ for $m \geq \lfloor (C/\sigma)^{1/s} \rfloor$. In this case

$$r_p(f) = \lfloor (C/\sigma)^{1/s} \rfloor \sigma^2 \sim C^{1/s} \sigma^{2-1/s} .$$

Since $r_{\inf}(\Theta_{C,s}) \geq r_p(f)/2$, it follows that $r_{\inf}(\Theta_{C,s}) \sim \sigma^{2-1/s} C^{1/s}$. ■

The hypothesis $C/\sigma \geq 1$ guarantees that the largest signal coefficient is not dominated by the noise, whereas $C/\sigma \leq N^s$ indicates that the smallest coefficient has an amplitude smaller than the noise. This is typically the domain of application for noise removal algorithms. If s is large, then $r_{\inf}(\Theta)$ is almost on the order of σ^2. This risk is much smaller than the noise energy $\mathrm{E}\{\|W\|^2\} = N\sigma^2$, which means that the estimation removed most of the noise.

10.3.3 Nearly Minimax with Wavelets

A thresholding estimator in a wavelet basis has a nearly minimax risk for sets of piecewise regular signals.This result is proved for piecewise polynomial signals, which have key characteristics that explain the efficiency of wavelet thresholding estimators. The more general case of bounded variation signals and images is studied.

Piecewise Polynomials Piecewise polynomials are among the most difficult bounded variation signals to estimate with a linear operator. Indeed, the proof of Proposition 10.5 shows that the maximum risk of an optimal linear estimator is nearly reached by piecewise constant signals.

The estimation of a piecewise polynomial f is improved by non-linear operators that average the noisy data $X = f + W$ over large domains where f is regular, but which avoid averaging X across the discontinuities of f. These adaptive smoothing algorithms require estimating the positions of the discontinuities of f from X. Let $\Theta_{K,d}$ be the set of piecewise polynomial signals on $[0, N-1]$, with at most K polynomial components of degree d or smaller. Figure 10.2 gives an example with $d = 3$ and $K = 9$. The following proposition computes a lower bound of the minimax risk $r_n(\Theta_{K,d})$.

Proposition 10.8 *If $\Theta_{K,d}$ is a set of piecewise polynomial signals then*

$$\frac{r_n(\Theta_{K,d})}{N\sigma^2} \geq K(d+1)N^{-1} . \tag{10.105}$$

Proof[2]. We consider $f \in \Theta_{K,d}$ which is equal to polynomials of degree d on a partition of $[0, N-1]$ composed of K sub-intervals $\{[\tau_k, \tau_{k+1} - 1]\}_{0 \leq k \leq K}$. To compute a lower bound of $r_n(\Theta_{K,d})$, we create an oracle estimator that knows in advance the position of each interval $[\tau_k, \tau_{k+1} - 1]$. On $[\tau_k, \tau_{k+1} - 1]$, f is equal to a polynomial p_k of degree d, which is characterized by $d+1$ parameters. Problem 10.3 shows that the minimum risk when estimating p_k on $[\tau_k, \tau_{k+1} - 1]$ from $X = f + W$ is obtained with an orthogonal projection on the space of polynomials of degree d over $[\tau_k, \tau_{k+1} - 1]$.

The resulting risk is $(d+1)\sigma^2$. Since $r_n(\Theta_{K,d})$ is larger than the sum of these risks on the K intervals,

$$r_n(\Theta_{K,d}) \geq K(d+1)\sigma^2.$$

■

The lower bound (10.105) is calculated with an oracle estimator that knows in advance the positions of the signal discontinuities. One can prove [227] that the need to estimate the position of the signal discontinuities introduces another $\log_2 N$ factor in the non-linear minimax risk:

$$\frac{r_n(\Theta_{K,d})}{N\sigma^2} \sim K(d+1)\frac{\log_e N}{N}.$$

It is much smaller than the normalized linear minimax risk (10.87), which decays like $N^{-1/2}$.

The inner product of a wavelet with $d+1$ vanishing moments and a polynomial of degree d is equal to zero. A wavelet basis thus gives a sparse representation of piecewise polynomials, with non-zero coefficients located in the neighborhood of their discontinuities. Figure 10.4(a) gives an example. The following theorem derives that a thresholding estimator in a wavelet basis has a risk that is close to the non-linear minimax.

Proposition 10.9 *Let $T = \sigma\sqrt{2\log_e N}$. The risk of a hard or a soft thresholding in a Daubechies wavelet basis with $d+1$ vanishing moments satisfies*

$$\frac{r_t(\Theta_{K,d})}{N\sigma^2} \leq \frac{4K(d+1)}{\log_e 2}\frac{\log_e^2 N}{N}(1+o(1)) \tag{10.106}$$

when N tends to $+\infty$.

Proof[2]. On $[0, N-1]$, the discrete wavelets $\psi_{j,m}[n]$ of a Daubechies basis with $d+1$ vanishing moments have a support of size $N2^j(2d+2)$. Let $f \in \Theta_{K,d}$. If the support of $\psi_{j,m}$ is included inside one of the polynomial components of f, then $\langle f, \psi_{j,m}\rangle = 0$. At each scale 2^j, there are at most $K(2d+2)$ wavelets $\psi_{j,m}$ whose support includes one of the K transition points of f. On at most $\log_2 N$ scales, the number M of non-zero coefficients thus satisfies

$$M \leq K(2d+2)\log_2 N. \tag{10.107}$$

Since $\min(|\langle f, \psi_{j,m}\rangle|^2, \sigma^2) \leq \sigma^2$ and $\min(|\langle f, \psi_{j,m}\rangle|^2, \sigma^2) = 0$ if $\langle f, \psi_{j,m}\rangle = 0$, we derive from (10.42) that the thresholding risk satisfies

$$r_t(f) \leq (2\log_e N + 1)(M+1)\sigma^2.$$

Inserting (10.107) yields

$$r_t(\Theta) \leq (1 + 2K(d+1)\log_2 N)(2\log_e N + 1)\sigma^2.$$

Extracting the dominating term for N large gives (10.106). ■

The wavelet thresholding risk $r_t(\Theta_{K,d})$ is thus larger than $r_n(\Theta_{K,d})$ by at most a $\log_e N$ factor. This loss comes from a non-optimal choice of the threshold $T = \sigma\sqrt{2\log_e N}$. If a different threshold T_j is used to threshold the wavelet coefficients at each scale 2^j, then one can prove [227] that the $\log_e N$ factor disappears:

$$\frac{\min_{(T_j)_j} r_t(\Theta_{K,d})}{N\sigma^2} \sim K\,(d+1)\,\frac{\log_e N}{N}. \tag{10.108}$$

For a soft thresholding, nearly optimal values T_j are calculated from the noisy data with the SURE estimator (10.66), and the resulting risk $r_t(\Theta_{K,d})$ has an asymptotic decay equivalent to (10.108) [169].

Bounded Variation Let Θ_V be the set of signals having a total variation bounded by C:

$$\Theta_V = \left\{ f \;:\; \|f\|_V = \sum_{n=0}^{N-1} |f[n] - f[n-1]| \le C \right\}.$$

To prove that a thresholding estimator in a wavelet basis has nearly a minimax risk, we show that Θ_V can be embedded in two sets that are orthosymmetric in the wavelet basis. This embedding is derived from the following proposition that computes an upper bound and a lower bound of $\|f\|_V$ from the wavelet coefficients of f. To simplify notations we write the scaling vectors of the wavelet basis: $\phi_{J,m} = \psi_{J+1,m}$. Recall that the minimum scale is $2^L = N^{-1}$.

Proposition 10.10 *There exist $A, B > 0$ such that for all $N > 0$*

$$\|f\|_V \le B N^{-1/2} \sum_{j=L+1}^{J+1} \sum_{m=0}^{2^{-j}-1} 2^{-j/2} |\langle f, \psi_{j,m}\rangle|\,, \tag{10.109}$$

and

$$\|f\|_V \ge A N^{-1/2} \sup_{j \le J} \left(\sum_{m=0}^{2^{-j}-1} 2^{-j/2} |\langle f, \psi_{j,m}\rangle| \right). \tag{10.110}$$

The proof is identical to the proof of Theorem 9.6, replacing integrals by discrete sums. The factor $N^{-1/2} 2^{-j/2}$ comes from the fact that $\|\psi_{j,m}\|_V \sim N^{-1/2} 2^{-j/2}$. The upper bound (10.109) and the lower bound (10.110) correspond to Besov norms (9.32), calculated at scales $2^j > N^{-1}$. The two Besov balls

$$\Theta_{1,1,1} = \left\{ f \;:\; \sum_{j=L+1}^{J+1} \sum_{m=0}^{2^{-j}-1} 2^{-j/2} |\langle f, \psi_{j,m}\rangle| \le B^{-1} C N^{1/2} \right\}$$

and

$$\Theta_{1,1,\infty} = \left\{ f \;:\; \sup_{j \le J} \left(\sum_{m=0}^{2^{-j}-1} 2^{-j/2} |\langle f, \psi_{j,m}\rangle| \right) \le A^{-1} C N^{1/2} \right\}$$

are clearly orthosymmetric in the wavelet basis. Proposition 10.10 proves that

$$\Theta_{1,1,1} \subset \Theta_V \subset \Theta_{1,1,\infty} . \tag{10.111}$$

Proposition 10.6 shows that a thresholding risk is nearly minimax over orthosymmetric sets. The following theorem derives a similar result over Θ_V by using the orthosymmetric embedding (10.111).

Theorem 10.9 (DONOHO, JOHNSTONE) *Let* $T = \sigma\sqrt{2\log_e N}$. *There exist* $A_1, B_1 > 0$ *such that if* $1 \leq C/\sigma \leq N$ *then*

$$A_1 \left(\frac{C}{\sigma}\right)^{2/3} \frac{1}{N^{2/3}} \leq \frac{r_n(\Theta_V)}{N\sigma^2} \leq \frac{r_t(\Theta_V)}{N\sigma^2} \leq B_1 \left(\frac{C}{\sigma}\right)^{2/3} \frac{\log_e N}{N^{2/3}}. \tag{10.112}$$

Proof[3]. Since $\Theta_{1,1,1}$ and $\Theta_{1,\infty,1}$ are orthosymmetric, Proposition 10.6 proves that

$$\frac{1}{1.25} r_{\inf}(\Theta_{1,1,1}) \leq r_n(\Theta_{1,1,1})$$

and

$$r_t(\Theta_{1,1,\infty}) \leq (2\log_e N + 1)\left(\sigma^2 + r_{\inf}(\Theta_{1,1,\infty})\right).$$

But $\Theta_{1,1,1} \subset \Theta_V \subset \Theta_{1,1,\infty}$ so

$$\frac{1}{1.25} r_{\inf}(\Theta_{1,1,1}) \leq r_n(\Theta_V) \leq r_t(\Theta_V) \leq (2\log_e N + 1)\left(\sigma^2 + r_{\inf}(\Theta_{1,1,\infty})\right).$$

The double inequality (10.112) is proved by verifying that

$$\frac{r_{\inf}(\Theta_{1,1,1})}{N\sigma^2} \sim \frac{r_{\inf}(\Theta_{1,1,\infty})}{N\sigma^2} \sim \left(\frac{C}{\sigma}\right)^{2/3} \frac{1}{N^{2/3}} . \tag{10.113}$$

The same proof as in Proposition 9.5 verifies that there exists B_2 such that for each $f \in \Theta_{1,1,\infty}$ the non-linear approximation error satisfies

$$\epsilon_n[M] \leq B_2 C^2 N M^{-2} .$$

Applying Proposition 10.7 for $s = 3/2$ shows that

$$r_{\inf}(\Theta_{1,1,\infty}) = O\left((CN^{1/2})^{2/3}\sigma^{2-2/3}\right) . \tag{10.114}$$

Since $r_{\inf}(\Theta_{1,1,1}) \leq r_{\inf}(\Theta_{1,1,\infty})$, it is now sufficient to compute a similar lower bound for $r_{\inf}(\Theta_{1,1,1})$. Let $\Theta_l \subset \Theta_{1,1,1}$ be the set of signals f such that $\langle f, \psi_{j,m}\rangle = 0$ for $j \neq l$, and which satisfy

$$\sum_{m=0}^{2^{-l}} |\langle f, \psi_{l,m}\rangle| \leq B^{-1} C N^{1/2} 2^{l/2} = C_l .$$

Over these 2^{-l} non-zero wavelet coefficients, Θ_l is identical to the set $\Theta_{C_l,s}$ defined in (10.104), for $s = 1$. Proposition 10.7 proves that

$$r_{\inf}(\Theta_l) \sim C N^{1/2} 2^{l/2} \sigma . \tag{10.115}$$

Since $1 \leq C/\sigma \leq N$ one can choose $1 \leq 2^{-l} \leq N$ such that

$$2^{-l} < \left(\frac{CN^{1/2}}{\sigma}\right)^{2/3} < 2^{-l+1}.$$

So

$$r_{\inf}(\Theta_{1,1,1}) \geq r_{\inf}(\Theta_l) \sim (CN^{1/2})^{2/3}\sigma^{2-2/3}.$$

■

This theorem proves that for bounded variation signals, the thresholding risk in a wavelet basis is close to the minimax risk $r_n(\Theta_V)$. The theorem proof can be refined [168] to show that

$$\frac{r_n(\Theta_V)}{N\sigma^2} \sim \left(\frac{C}{\sigma}\right)^{2/3} \frac{1}{N^{2/3}} \quad \text{and} \quad \frac{r_t(\Theta_V)}{N\sigma^2} \sim \left(\frac{C}{\sigma}\right)^{2/3} \frac{\log_e N}{N^{2/3}}.$$

The loss of a factor $\log_e N$ in the thresholding risk is due to a threshold choice $T = \sigma\sqrt{2\log_e N}$ that is too high at large scales. If the wavelet coefficients are thresholded with different thresholds T_j that are optimized for scale 2^j then the $\log_e N$ factor disappears [170, 227]. In this case, when N increases, $r_t(\Theta_V)$ and $r_n(\Theta_V)$ have equivalent decay. For a soft thresholding, the thresholds T_j can be calculated with the SURE estimator (10.66). We restrict our study to bounded variation signals because they have a simple characterization, but the minimax and thresholding risks can also be calculated in balls of any Besov space, leading to similar near-optimality results [170].

Bounded Variation Images We now study the estimation of bounded variation images of N^2 pixels, which we will assume to be periodic to avoid border problems. The discrete total variation is defined in (2.70):

$$\|f\|_V = \frac{1}{N} \sum_{n_1,n_2=0}^{N-1} \left(\left| f[n_1,n_2] - f[n_1-1,n_2] \right|^2 + \left| f[n_1,n_2] - f[n_1,n_2-1] \right|^2 \right)^{1/2}.$$

Images also have a bounded amplitude, which has an important influence on linear estimation. If we subtract 128 from the image intensity, its amplitude is bounded by $C_\infty = 128$. Let $\Theta_{V,\infty}$ be the set of images that have a total variation and an amplitude bounded respectively by C_V and C_∞:

$$\Theta_{V,\infty} = \left\{ f \,:\, \|f\|_V \leq C_V \,,\, \|f\|_\infty \leq C_\infty \right\}.$$

In one dimension, Theorem 10.7 proves that if a set Θ is translation invariant, then the linear minimax risk $r_l(\Theta)$ is reached by an estimator that is diagonal in the discrete Fourier basis, and thus corresponds to a circular convolution. This result remains valid for images, and is proved similarly. Since $\Theta_{V,\infty}$ is translation invariant, the minimax linear estimator can be written as a circular convolution. The next theorem computes the linear minimax risk $r_l(\Theta_{V,\infty})$.

Theorem 10.10 *If* $N^{-2} \leq C_V C_\infty / \sigma^2 \leq N$ *then*

$$\frac{r_l(\Theta_{V,\infty})}{N^2\sigma^2} \sim \left(\frac{C_V C_\infty}{\sigma^2}\right)^{2/3} \frac{1}{N^{2/3}} . \tag{10.116}$$

Proof[3]. An upper bound of $r_l(\Theta_{V,\infty})$ is calculated in a separable discrete wavelet basis. As in Theorem 9.7 one can prove that there exists B_1 such that for all $1 \leq l \leq 3$ and $N^{-1} < 2^j < 1$ the discrete wavelet coefficients satisfy

$$\sum_{2^j m \in [0,1]^2} |\langle f, \psi^l_{j,m}\rangle| \leq B_1 N \|f\|_V .$$

The factor N comes from the fact that two-dimensional discrete wavelets satisfy $\|\psi^l_{j,m}\|_V \sim N^{-1}$. Since the $\mathbf{l}^1$ norm of $\psi^l_{j,m}$ is proportional to $N 2^j$ we also prove as in (9.49) that there exists B_2 such that

$$|\langle f, \psi^l_{j,m}\rangle| \leq B_2 N 2^j \|f\|_\infty .$$

Since $\sum_n |a_n|^2 \leq \sup_n |a_n| \sum_n |a_n|$ and $\|f\|_V \|f\|_\infty \leq C_V C_\infty$ we get

$$\sum_{2^j m \in [0,1]^2} |\langle f, \psi^l_{j,m}\rangle|^2 \leq B_1 B_2 N^2 C_V C_\infty 2^j . \tag{10.117}$$

Let Θ_2 be the set of signals f that satisfy (10.117). We just proved that $\Theta_{V,\infty} \subset \Theta_2$, which implies that $r_l(\Theta_{V,\infty}) \leq r_l(\Theta_2)$. Moreover, since Θ_2 is orthosymmetric and quadratically convex in the wavelet basis, Theorem 10.6 proves that $r_l(\Theta_2) = r_{\inf}(\Theta_2)$. We now use Proposition 10.7 to compute an upper bound of $r_{\inf}(\Theta_2)$. For any $f \in \Theta_2$, as in Proposition 9.6 we verify that there exists B_3 such that for any $f \in \Theta_2$ the linear approximation error satisfies

$$\epsilon_l[M] \leq B_3 C_V C_\infty N^2 M^{-1/2} .$$

Since $\epsilon_n[M] \leq \epsilon_l[M]$ we can apply (10.103), which proves for $s = 3/4$ and for $C = (C_V C_\infty)^{1/2} N$ that

$$r_l(\Theta_V) \leq r_l(\Theta_2) = r_{\inf}(\Theta_2) = O\left(C_V^{2/3} C_\infty^{2/3} N^{4/3} \sigma^{2/3}\right) . \tag{10.118}$$

This is valid only for $1 \leq C/\sigma \leq N^{2s}$, hence the theorem hypothesis.

To compute a lower bound of $r_l(\Theta_{V,\infty})$ we use the fact that the minimax linear operator is diagonal in the discrete Fourier basis and hence that in the Fourier basis

$$r_l(\Theta_{V,\infty}) \geq r_{\inf}(\Theta_{V,\infty}) = \sup_{f \in \Theta_{V,\infty}} \sum_{k_1,k_2=0}^{N-1} \frac{\sigma^2 N^{-2} |\hat{f}[k_1,k_2]|^2}{\sigma^2 + N^{-2} |\hat{f}[k_1,k_2]|^2} .$$

If $f = C_\infty \mathbf{1}_{[0,L]^2}$ with $L = C_V N/(4 C_\infty)$ then $f \in \Theta_{V,\infty}$. A direct calculation shows that there exists B_4 such that

$$r_l(\Theta_{V,\infty}) \geq \sum_{k_1,k_2=0}^{N-1} \frac{\sigma^2 N^{-2} |\hat{f}[k_1,k_2]|^2}{\sigma^2 + N^{-2} |\hat{f}[k_1,k_2]|^2} \geq B_4 C_V^{2/3} C_\infty^{2/3} N^{4/3} \sigma^{2/3} .$$

Together with (10.118) it proves (10.116). ■

The hypothesis of bounded amplitude is very important for linear estimation. Let us consider the set of images having a uniformly bounded total variation, with no constraint on their amplitude:

$$\Theta_V = \left\{ f \ : \ \|f\|_V \leq C_V \right\}.$$

Clearly $\Theta_{V,\infty} \subset \Theta_V$. One can prove (Problem 10.15) that $r_l(\Theta_V)/(N\sigma^2) \sim 1$, which means that the linear minimax risk over Θ_V is much larger than over $\Theta_{V,\infty}$.

For non-linear estimation, the hypothesis of bounded amplitude plays a minor role. The following theorem computes the non-linear minimax risk $r_n(\Theta_V)$ as well as the thresholding risk $r_t(\Theta_V)$ in a separable wavelet basis.

Theorem 10.11 *Let $T = \sigma\sqrt{2\log_e N^2}$. There exist $A_1, B_1 > 0$ such that if $N^{-1} \leq C_V/\sigma \leq N$ then*

$$A_1 \frac{C_V}{\sigma}\frac{1}{N} \leq \frac{r_n(\Theta_V)}{N^2\sigma^2} \leq \frac{r_t(\Theta_V)}{N^2\sigma^2} \leq B_1 \frac{C_V}{\sigma}\frac{\log_e N}{N}. \tag{10.119}$$

Proof[3]. An upper bound and a lower bound are calculated by embedding Θ_V in two sets that are orthosymmetric in a separable wavelet basis. These sets are derived from upper and lower bounds of $\|f\|_V$ calculated with the wavelet coefficients of f.

As in Theorem 9.7, one can prove that there exist $A, B > 0$ such that for all $N > 0$ the discrete wavelet coefficients satisfy

$$AN\|f\|_V \leq \sum_{j=L+1}^{J} \sum_{l=1}^{3} \sum_{2^j m \in [0,1]^2} |\langle f, \psi_{j,m}^l \rangle| + \sum_{2^J m \in [0,1]^2} |\langle f, \phi_{J,m}^2 \rangle| , \tag{10.120}$$

with $L = -\log_2 N$. The factor N comes from the fact discrete two-dimensional wavelets satisfy $\|\psi_{j,m}^l\|_V \sim N^{-1}$. Let us define

$$\Theta_1 = \left\{ f \ : \ \sum_{j=L+1}^{J} \sum_{l=1}^{3} \sum_{2^j m \in [0,1]^2} |\langle f, \psi_{j,m}^l \rangle| + \sum_{2^J m \in [0,1]^2} |\langle f, \phi_{J,m}^2 \rangle| \leq AN C_V \right\}.$$

Clearly $\Theta_1 \subset \Theta_V$.

Let $f_B^r[k]$ be the sorted wavelet coefficients in decreasing amplitude order. This sorting excludes the 2^{2J} scaling coefficients $\langle f, \phi_{J,m}^2 \rangle$. As in Theorem 9.8, one can verify that

$$BN\|f\|_V \geq k|f_B^r[k]| . \tag{10.121}$$

Hence the set $\Theta_2 = \{ f \ : \ |f_B^r[k]| \leq C_V N B k^{-1} \}$ includes Θ_V:

$$\Theta_1 \subset \Theta_V \subset \Theta_2 . \tag{10.122}$$

Since Θ_1 and Θ_2 are orthosymmetric in the wavelet basis, Proposition 10.6 implies that

$$\frac{1}{1.25} r_{\inf}(\Theta_1) \leq r_n(\Theta_V) \leq r_t(\Theta_V) \leq (2\log_e N + 1)\left(\sigma^2 + r_{\inf}(\Theta_2)\right). \tag{10.123}$$

Any signal $f \in \Theta_2$ has a non-linear approximation error which satisfies $\epsilon_n[M] = O(C_V^2 N^2 M^{-1})$. Applying Proposition 10.7 for $s = 1$ and $C = C_V N$ proves that $r_{\inf}(\Theta_2) = O(C_V N \sigma)$ for $N^{-1} \leq C_V/\sigma \leq N$. Since Θ_1 is defined with an upper bound $C = AN C_V$ on the $\mathbf{l}^1$ norm of the wavelet coefficients, Proposition 10.7 also proves that $r_{\inf}(\Theta_1) \sim C_V N \sigma$. But $r_{\inf}(\Theta_1) \leq r_{\inf}(\Theta_2)$ so

$$r_{\inf}(\Theta_1) \sim r_{\inf}(\Theta_2) \sim C_V N \sigma \,. \tag{10.124}$$

We thus derive (10.119) from (10.123) and (10.124). ■

As in one dimension, if the threshold $T = \sigma\sqrt{2\log_e N}$ is replaced by thresholds T_j that are optimized at each scale, then the $\log_e N$ term disappears [170, 227] and

$$\frac{r_t(\Theta_V)}{N\sigma^2} \sim \frac{r_n(\Theta_V)}{N\sigma^2} \sim \frac{C_V}{\sigma}\frac{1}{N}. \tag{10.125}$$

For a soft thresholding, the thresholds T_j can be calculated with the SURE estimator (10.66). If Θ_V is replaced by the smaller set $\Theta_{V,\infty}$ of images with a bounded amplitude, the decay of the minimax risk and of the thresholding risk is not improved.

When N increases, a thresholding estimator has a risk which decays like N^{-1} as opposed to $N^{-2/3}$ for an optimal linear estimator. This numerical improvement remains limited for images where typically $N \leq 512$. However, thresholding estimators bring important visual improvements by suppressing totally the noise in regions where the image intensity is highly regular.

10.4 RESTORATION [3]

Measurement devices can introduce important distortions and add noise to the original signal. Inverting the degradation is often numerically unstable and thus amplifies the noise considerably. The signal estimation must be performed with a high amplitude noise that is not white. Deconvolutions are generic examples of such unstable inverse problems.

Section 10.4.1 studies the estimation of signals contaminated by non white Gaussian noises. It shows that thresholding estimators are quasi-minimax optimal if the basis nearly diagonalizes the covariance of the noise and provides sparse signal representations. Inverse problems and deconvolutions are studied in Section 10.4.2, with an application to the removal of blur in satellite images.

10.4.1 Estimation in Arbitrary Gaussian Noise

The signal f is contaminated by an additive Gaussian noise Z:

$$X = f + Z \,.$$

The random vector Z is characterized by its covariance operator K, and we suppose that $\mathrm{E}\{Z[n]\} = 0$. When this noise is white, Section 10.3.2 proves that diagonal estimators in an orthonormal basis $\mathcal{B} = \{g_m\}_{0 \leq m < N}$ are nearly minimax optimal

if the basis provides a sparse signal representation. When the noise is not white, the coefficients of the noise have a variance that depends on each g_m:

$$\sigma_m^2 = \mathrm{E}\{|Z_{\mathcal{B}}[m]|^2\} = \langle K g_m, g_m \rangle .$$

The basis choice must therefore depend on the covariance K.

Diagonal Estimation We study the risk of estimators that are diagonal in $\mathcal{B}$:

$$\tilde{F} = DX = \sum_{m=0}^{N-1} d_m(X_{\mathcal{B}}[m])\, g_m . \tag{10.126}$$

If $d_m(X_{\mathcal{B}}[m]) = a[m] X_{\mathcal{B}}[m]$, we verify as in (10.28) that the minimum risk $\mathrm{E}\{\|\tilde{F} - f\|^2\}$ is achieved by an oracle attenuation:

$$a[m] = \frac{|f_{\mathcal{B}}[m]|^2}{|f_{\mathcal{B}}[m]|^2 + \sigma_m^2}, \tag{10.127}$$

and

$$\mathrm{E}\{\|\tilde{F} - f\|^2\} = r_{\mathrm{inf}}(f) = \sum_{m=0}^{N-1} \frac{\sigma_m^2 |f_{\mathcal{B}}[m]|^2}{\sigma_m^2 + |f_{\mathcal{B}}[m]|^2} . \tag{10.128}$$

Over a signal set Θ, the maximum risk of an oracle attenuation is $r_{\mathrm{inf}}(\Theta) = \sup_{f \in \Theta} r_{\mathrm{inf}}(f)$. An oracle attenuation cannot be implemented because $a[m]$ depends on $|f_{\mathcal{B}}[m]|$ which is not known, so $r_{\mathrm{inf}}(\Theta)$ is only a lower bound for the minimax risk of diagonal estimators. However, a simple thresholding estimator has a maximum risk that is close to $r_{\mathrm{inf}}(\Theta)$. We begin by studying linear diagonal estimators D, where each $a[m]$ is a constant. The following proposition computes an upper bound of the minimax linear risk. The quadratic convex hull $\mathrm{QH}[\Theta]$ of Θ is defined in (10.78).

Proposition 10.11 *Let Θ be a closed and bounded set. There exists $x \in \mathrm{QH}[\Theta]$ such that $r_{\mathrm{inf}}(x) = r_{\mathrm{inf}}(\mathrm{QH}[\Theta])$. If D is the linear operator defined by*

$$a[m] = \frac{|x_{\mathcal{B}}[m]|^2}{\sigma_m^2 + |x_{\mathcal{B}}[m]|^2} , \tag{10.129}$$

then

$$r_l(\Theta) \leq r(D, \Theta) = r_{\mathrm{inf}}(\mathrm{QH}[\Theta]) . \tag{10.130}$$

Proof[2]. Let $r_{l,d}(\Theta)$ be the minimax risk obtained over linear operators that are diagonal in $\mathcal{B}$. Clearly $r_l(\Theta) \leq r_{l,d}(\Theta)$. The same derivations as in Theorem 10.6 prove that the diagonal operator defined by (10.129) satisfies

$$r(D, \Theta) = r_{l,d}(\Theta) = r_{\mathrm{inf}}(\mathrm{QH}[\Theta]) .$$

Hence (10.130). ∎

Among non-linear diagonal estimators, we concentrate on thresholding estimators:

$$\tilde{F} = \sum_{m=0}^{N-1} \rho_{T_m}(X_B[m])\, g_m \,, \tag{10.131}$$

where $\rho_T(x)$ is a hard or soft thresholding function. The threshold T_m is adapted to the noise variance σ_m^2 in the direction of g_m. Proposition 10.4 computes an upper bound of the risk $r_t(f)$ when $T_m = \sigma_m \sqrt{2\log_e N}$. If the signals belong to a set Θ, the threshold values are improved by considering the maximum of signal coefficients:

$$s_B[m] = \sup_{f \in \Theta} |f_B[m]| \,.$$

If $s_B[m] \leq \sigma_m$ then setting $X_B[m]$ to zero yields a risk $|f_B[m]|^2$ that is always smaller than the risk σ_m^2 of keeping it. This is done by choosing $T_m = \infty$ to guarantee that $\rho_{T_m}(X_B[m]) = 0$. Thresholds are therefore defined by

$$T_m = \begin{cases} \sigma_m \sqrt{2\log_e N} & \text{if } \sigma_m < s_B[m] \\ \infty & \text{if } \sigma_m \geq s_B[m] \end{cases} \,. \tag{10.132}$$

Proposition 10.12 *For the thresholds (10.132), the risk of a thresholding estimator satisfies for* $N \geq 4$

$$r_t(\Theta) \leq (2\log_e N + 1)\left(\bar{\sigma}^2 + r_{\inf}(\Theta)\right) \tag{10.133}$$

with $\bar{\sigma}^2 = \frac{1}{N}\sum_{\sigma_m < s_B[m]} \sigma_m^2$.

Proof[2]. The thresholding risk $r_t(f)$ is calculated by considering separately the case $T_m = \infty$, which produces a risk of $|f_B[m]|^2$, from the case $T_m < \infty$

$$r_t(f) = \sum_{\sigma_m \geq s_B[m]} |f_B[m]|^2 + \sum_{\sigma_m < s_B[m]} \mathrm{E}\{|f_B[m] - \rho_{T_m}(X_B[m])|^2\} \,. \tag{10.134}$$

A slight modification [167] of the proof of Theorem 10.4 shows that

$$\mathrm{E}\{|f_B[m] - \rho_{T_m}(X_B[m])|^2\} \leq (2\log_e N + 1)\left(\frac{\sigma_m^2}{N} + \frac{\sigma_m^2 |f_B[m]|^2}{\sigma_m^2 + |f_B[m]|^2}\right). \tag{10.135}$$

If $\sigma_m \geq s_B[m]$ then $|f_B[m]|^2 \leq 2\sigma_m^2 |f_B[m]|^2 (\sigma_m^2 + |f_B[m]|^2)^{-1}$, so inserting (10.135) in (10.134) proves (10.133). ■

This proposition proves that the risk of a thresholding estimator is not much above $r_{\inf}(\Theta)$. It now remains to understand under what conditions the minimax risk $r_n(\Theta)$ is also on the order of $r_{\inf}(\Theta)$.

Nearly Diagonal Covariance To estimate efficiently a signal with a diagonal operator, the basis $\mathcal{B}$ must provide a sparse representation of signals in Θ but it must also transform the noise into "nearly" independent coefficients. Since the noise Z is Gaussian, it is sufficient to have "nearly" uncorrelated coefficients, which means that the covariance K of Z is "nearly" diagonal in $\mathcal{B}$. This approximate diagonalization is measured by preconditioning K with its diagonal. We denote by K_d the diagonal operator in the basis $\mathcal{B}$, whose diagonal coefficients are equal to the diagonal coefficients σ_m^2 of K. We suppose that K has no eigenvalue equal to zero, because the noise would then be zero in this direction, in which case the estimation is trivial. Let K^{-1} be the inverse of K, and $K_d^{1/2}$ be the diagonal matrix whose coefficients are the square root of the diagonal coefficients of K_d. The following theorem computes lower bounds of the minimax risks with a preconditioning factor defined with the operator sup norm $\|\,.\,\|_S$ introduced in (A.16).

Theorem 10.12 (DONOHO, KALIFA, MALLAT) *The preconditioning factor satisfies*

$$\lambda_{\mathcal{B}} = \|K_d^{1/2} K^{-1} K_d^{1/2}\|_S \geq 1 \,. \tag{10.136}$$

If Θ is orthosymmetric in $\mathcal{B}$ then

$$r_l(\Theta) \geq \frac{1}{\lambda_{\mathcal{B}}}\, r_{\inf}(\mathrm{QH}[\Theta]) \tag{10.137}$$

and

$$r_n(\Theta) \geq \frac{1}{1.25\,\lambda_{\mathcal{B}}}\, r_{\inf}(\Theta) \,. \tag{10.138}$$

Proof[3]. The proof considers first the particular case where K is diagonal. If K is diagonal in $\mathcal{B}$ then the coefficients $Z_{\mathcal{B}}[m]$ are independent Gaussian random variables of variance σ_m^2. Estimating $f \in \Theta$ from $X = f + Z$ is equivalent to estimating f_0 from $X_0 = f_0 + Z_0$ where

$$Z_0 = \sum_{m=0}^{N-1} \frac{Z_{\mathcal{B}}[m]}{\sigma_m}\, g_m \;,\; X_0 = \sum_{m=0}^{N-1} \frac{X_{\mathcal{B}}[m]}{\sigma_m}\, g_m \;,\; f_0 = \sum_{m=0}^{N-1} \frac{f_{\mathcal{B}}[m]}{\sigma_m}\, g_m \,. \tag{10.139}$$

The signal f_0 belongs to an orthosymmetric set Θ_0 and the renormalized noise Z_0 is a Gaussian white noise of variance 1. Proposition 10.6 applies to the estimation problem $X_0 = f_0 + Z_0$. By reinserting the value of the renormalized noise and signal coefficients, we derive that

$$r_n(\Theta) \geq \frac{1}{1.25}\, r_{\inf}(\Theta) \;\text{ and }\; r_l(\Theta) = r_{\inf}(\mathrm{QH}[\Theta]) \,. \tag{10.140}$$

To prove the general case we use inequalities over symmetrical matrices. If A and B are two symmetric matrices, we write $A \geq B$ if the eigenvalues of $A - B$ are positive, which means that $\langle Af, f\rangle \geq \langle Bf, f\rangle$ for all $f \in \mathbb{C}^N$. Since $\lambda_{\mathcal{B}}$ is the largest eigenvalue of $K_d^{1/2} K^{-1} K_d^{1/2}$, the inverse $\lambda_{\mathcal{B}}^{-1}$ is the smallest eigenvalue of the inverse $K_d^{-1/2} K K_d^{-1/2}$. It follows that $\langle K_d^{-1/2} K K_d^{-1/2} f, f\rangle \geq \lambda_{\mathcal{B}}^{-1} \langle f, f\rangle$. By setting $g =$

$K_d^{-1/2} f$ we get $\langle Kg, g\rangle \geq \lambda_{\mathcal{B}}^{-1} \langle K_d^{1/2} g, K_d^{1/2} g\rangle$. Since this is valid for all $g \in \mathbb{C}^N$, we derive that

$$K \geq \lambda_{\mathcal{B}}^{-1} K_d \,. \tag{10.141}$$

Observe that $\lambda_{\mathcal{B}} \geq 1$ because $\langle Kg_m, g_m\rangle = \langle K_d g_m, g_m\rangle$. Lower bounds for the minimax risks are proved as a consequence of the following lemma.

Lemma 10.2 *Consider the two estimation problems $X_i = f + Z_i$ for $i = 1, 2$, where K_i is the covariance of the Gaussian noise Z_i. We denote by $r_{i,n}(\Theta)$ and $r_{i,l}(\Theta)$ the non-linear and linear minimax risks for each estimation problem $i = 1, 2$. If $K_1 \geq K_2$ then*

$$r_{1,n}(\Theta) \geq r_{2,n}(\Theta) \;\; \textit{and} \;\; r_{1,l}(\Theta) \geq r_{2,l}(\Theta) \,. \tag{10.142}$$

Since $K_1 \geq K_2$ one can write $Z_1 = Z_2 + Z_3$ where Z_2 and Z_3 are two independent Gaussian random vectors and the covariance of Z_3 is $K_3 = K_1 - K_2 \geq 0$. We denote by π_i the Gaussian probability distribution of Z_i. To any estimator $\tilde{F}_1 = D_1 X_1$ of f from X_1 we can associate an estimator $\tilde{F}_2$, calculated by augmenting the noise with Z_3 and computing the average with respect to its probability distribution:

$$\tilde{F}_2 = D_2 X_2 = E_{\pi_3}\{D_1(X_2 + Z_3)\} = E_{\pi_3}\{D_1 X_1\} \,.$$

The risk is

$$\begin{aligned} E_{\pi_2}\{|D_2 X_2 - f|^2\} &= E_{\pi_2}\{|E_{\pi_3}\{D_1 X_1\} - f|^2\} \\ &\leq E_{\pi_2}\{E_{\pi_3}\{|D_1 X_1 - f|^2\}\} = E_{\pi_1}\{|D_1 X_1 - f|^2\} \,. \end{aligned}$$

To any estimator $\tilde{F}_1 = D_1 X_1$ we can thus associate an estimator $\tilde{F}_2 = D_2 X_2$ of lower risk for all $f \in \Theta$. Taking a sup over all $f \in \Theta$ and the infimum over linear or non-linear operators proves (10.142).

Since $K \geq \lambda_{\mathcal{B}}^{-1} K_d$, Lemma 10.2 proves that the estimation problem with the noise Z of covariance K has a minimax risk that is larger than the minimax risk of the estimation problem with a noise of covariance $\lambda_{\mathcal{B}}^{-1} K_d$. But since this covariances is diagonal we can apply (10.140). The definition of $r_{\inf}(\Theta)$ is the same for a noise of covariance K and for a noise of covariance K_d because $\sigma_m^2 = \langle Kg_m, g_m\rangle = \langle K_d g_m, g_m\rangle$. When multiplying K_d by a constant $\lambda_{\mathcal{B}}^{-1} \leq 1$, the value $r_{\inf}(\Theta)$ that appears in (10.140) is modified into $r'_{\inf}(\Theta)$ with $r'_{\inf}(\Theta) \geq \lambda_{\mathcal{B}}^{-1} r_{\inf}(\Theta)$. We thus derive (10.138) and (10.137). ■

One can verify that $\lambda_{\mathcal{B}} = 1$ if and only if $K = K_d$ and hence that K is diagonal in $\mathcal{B}$. The closer $\lambda_{\mathcal{B}}$ is to 1 the more diagonal K. The main difficulty is to find a basis $\mathcal{B}$ that nearly diagonalizes the covariance of the noise and provides sparse signal representations so that Θ is orthosymmetric or can be embedded in two close orthosymmetric sets.

An upper bound of $r_l(\Theta)$ is computed in (10.130) with a linear diagonal operator, and together with (10.137) we get

$$\frac{1}{\lambda_{\mathcal{B}}} r_{\inf}(\mathrm{QH}[\Theta]) \leq r_l(\Theta) \leq r_{\inf}(\mathrm{QH}[\Theta]) \,. \tag{10.143}$$

Similarly, an upper bound of $r_n(\Theta)$ is calculated with the thresholding risk calculated by Proposition 10.12. With the lower bound (10.138) we obtain

$$\frac{1}{1.25\,\lambda_{\mathcal{B}}} r_{\inf}(\Theta) \leq r_n(\Theta) \leq r_t(\Theta) \leq (2\log_e N + 1)\left(\bar{\sigma}^2 + r_{\inf}(\Theta)\right). \quad (10.144)$$

If the basis $\mathcal{B}$ nearly diagonalizes K so that $\lambda_{\mathcal{B}}$ is on the order of 1 then $r_l(\Theta)$ is on the order of $r_{\inf}(\mathrm{QH}[\Theta])$, whereas $r_n(\Theta)$ and $r_t(\Theta)$ are on the order of $r_{\inf}(\Theta)$. If Θ is quadratically convex then $\Theta = \mathrm{QH}[\Theta]$ so the linear and non-linear minimax risks are close. If Θ is not quadratically convex then a thresholding estimation in $\mathcal{B}$ may significantly outperform an optimal linear estimation.

10.4.2 Inverse Problems and Deconvolution

The measurement of a discrete signal f of size N is degraded by a linear operator U and a Gaussian white noise W of variance σ^2 is added:

$$Y = Uf + W. \quad (10.145)$$

We suppose that U and σ^2 have been calculated through a calibration procedure. The restoration problem is transformed into a denoising problem by inverting the degradation. We can then apply linear or non-linear diagonal estimators studied in the previous section. When the inverse U^{-1} is not bounded, the noise is amplified by a factor that tends to infinity. This is called an *ill-posed* inverse problem [96, 323] The case where U is a convolution operator is studied in more detail with an application to satellite images.

Pseudo Inverse The degradation U is inverted with the pseudo-inverse defined in Section 5.1.2. Let $\mathbf{V} = \mathbf{Im}\mathrm{U}$ be the image of U and $\mathbf{V}^{\perp}$ be its orthogonal complement. The pseudo-inverse $\tilde{U}^{-1}$ of U is the left inverse whose restriction to $\mathbf{V}^{\perp}$ is zero. The restoration is said to be unstable if

$$\lim_{N \to +\infty} \|\tilde{U}^{-1}\|_S^2 = +\infty\,.$$

Estimating f from Y is equivalent to estimating it from

$$X = \tilde{U}^{-1}Y = \tilde{U}^{-1}Uf + \tilde{U}^{-1}W. \quad (10.146)$$

The operator $\tilde{U}^{-1}U = P_{\mathbf{V}}$ is an orthogonal projection on $\mathbf{V}$ so

$$X = P_{\mathbf{V}}f + Z \text{ with } Z = \tilde{U}^{-1}W\,. \quad (10.147)$$

The noise Z is not white but remains Gaussian because $\tilde{U}^{-1}$ is linear. It is considerably amplified when the problem is unstable. The covariance operator K of Z is

$$K = \sigma^2\,\tilde{U}^{-1}\,\tilde{U}^{-1*}\,, \quad (10.148)$$

where A^* is the adjoint of an operator A.

To simplify notation, we formally rewrite (10.147) as a standard denoising problem:

$$X = f + Z\,, \tag{10.149}$$

while considering that the projection of Z in $\mathbf{V}^\perp$ is a noise of infinite energy to express the loss of all information concerning the projection of f in $\mathbf{V}^\perp$. It is equivalent to write formally $Z = U^{-1}W$.

Let $\mathcal{B} = \{g_m\}_{0\leq m<N}$ be an orthonormal basis such that a subset of its vectors defines a basis of $\mathbf{V} = \mathbf{Im}\mathrm{U}$. The coefficients of the noise have a variance $\sigma_m^2 = \mathrm{E}\{|Z_{\mathcal{B}}[m]|^2\}$, and we set $\sigma_m = \infty$ if $g_m \in \mathbf{V}^\perp$. An oracle attenuation (10.127) yields a lower bound for the risk

$$r_{\inf}(f) = \sum_{m=0}^{N-1} \frac{\sigma_m^2\,|f_{\mathcal{B}}[m]|^2}{\sigma_m^2 + |f_{\mathcal{B}}[m]|^2}\,. \tag{10.150}$$

The loss of the projection of f in $\mathbf{V}^\perp$ appears in the terms

$$\frac{\sigma_m^2\,|f_{\mathcal{B}}[m]|^2}{\sigma_m^2 + |f_{\mathcal{B}}[m]|^2} = |f_{\mathcal{B}}[m]|^2 \quad \text{if } \sigma_m = \infty\,.$$

Proposition 10.12 proves that a thresholding estimator in $\mathcal{B}$ yields a risk that is above $r_{\inf}(\Theta)$ by a factor $2\log_e N$. Theorem 10.12 relates linear and non-linear minimax risk to $r_{\inf}(\Theta)$. Let K_d be the diagonal operator in $\mathcal{B}$, equal to the diagonal of the covariance K defined in (10.148). The inverse of K is replaced by its pseudo inverse $K^{-1} = \sigma^{-2}U^*U$ and the preconditioning number is

$$\lambda_{\mathcal{B}} = \|K_d^{1/2}K^{-1}K_d^{1/2}\|_S = \sigma^{-2}\,\|K_d^{1/2}U\|_S^2\,.$$

Thresholding estimators have a risk $r_t(\Theta)$ that is close to $r_n(\Theta)$ if Θ is nearly orthosymmetric in $\mathcal{B}$ and if $\lambda_{\mathcal{B}}$ is on the order of 1. The main difficulty is to find such a basis $\mathcal{B}$.

The thresholds (10.132) define a projector that is non-zero only in the space $\mathbf{V}_0 \subset \mathbf{V}$ generated by the vectors $\{g_m\}_{\sigma_m < s_{\mathcal{B}}[m]}$. This means that the calculation of $X = \tilde{U}^{-1}Y$ in (10.146) can be replaced by a regularized inverse $X = P_{\mathbf{V}_0}\tilde{U}^{-1}Y$, to avoid numerical instabilities.

Deconvolution The restoration of signals degraded by a convolution operator U is a generic inverse problem that is often encountered in signal processing. The convolution is supposed to be circular to avoid border problems. The goal is to estimate f from

$$Y = f \circledast u + W\,.$$

The circular convolution is diagonal in the discrete Fourier basis $\mathcal{B} = \left\{g_m[n] = N^{-1/2}\exp\left(i2\pi mn/N\right)\right\}_{0\leq m<N}$. The eigenvalues are equal to the discrete

Fourier transform $\hat{u}[m]$, so $\mathbf{V} = \mathbf{Im}\mathbf{U}$ is the space generated by the sinusoids g_m such that $\hat{u}[m] \neq 0$. The pseudo inverse of U is $\tilde{U}^{-1} f = f \circledast \tilde{u}^{-1}$ where the discrete Fourier transform of $\tilde{u}^{-1}$ is

$$\widehat{\tilde{u}^{-1}}[m] = \begin{cases} 1/\hat{u}[m] & \text{if } \hat{u}[m] \neq 0 \\ 0 & \text{if } \hat{u}[m] = 0 \end{cases}.$$

The deconvolved data are

$$X = \tilde{U}^{-1} Y = Y \circledast \tilde{u}^{-1} .$$

The noise $Z = \tilde{U}^{-1} W$ is circular stationary. Its covariance K is a circular convolution with $\sigma^2 \tilde{u}^{-1} \circledast \overline{\tilde{u}}^{-1}$, where $\overline{\tilde{u}}^{-1}[n] = \tilde{u}^{-1}[-n]$. The Karhunen-Loève basis that diagonalizes K is therefore the discrete Fourier basis $\mathcal{B}$. The eigenvalues of K are $\sigma_m^2 = \sigma^2 |\hat{u}[m]|^{-2}$. When $\hat{u}[m] = 0$ we formally set $\sigma_m^2 = \infty$.

When the convolution filter is a low-pass filter with a zero at a high frequency, the deconvolution problem is highly unstable. Suppose that $\hat{u}[m]$ has a zero of order $p \geq 1$ at the highest frequency $m = \pm N/2$:

$$|\hat{u}[m]| \sim \left| \frac{2m}{N} - 1 \right|^p . \tag{10.151}$$

The noise variance σ_m^2 has a hyperbolic growth when the frequency m is in the neighborhood of $\pm N/2$. This is called a *hyperbolic deconvolution* problem of degree p.

Linear Estimation In many deconvolution problems the set Θ is translation invariant, which means that if $g \in \Theta$ then any translation of g modulo N also belongs to Θ. Since the amplified noise Z is circular stationary the whole estimation problem is translation invariant. In this case, the following theorem proves that the linear estimator that achieves the minimax linear risk is diagonal in the discrete Fourier basis. It is therefore a circular convolution. In the discrete Fourier basis,

$$r_{\inf}(f) = \sum_{m=0}^{N-1} \frac{\sigma_m^2 N^{-1} |\hat{f}[m]|^2}{\sigma_m^2 + N^{-1} |\hat{f}[m]|^2} . \tag{10.152}$$

We denote by $\mathrm{QH}[\Theta]$ the quadratic convex hull of Θ in the discrete Fourier basis.

Theorem 10.13 *Let Θ be a translation invariant set. The minimax linear risk is reached by circular convolutions and*

$$r_l(\Theta) = r_{\inf}(\mathrm{QH}[\Theta]) . \tag{10.153}$$

Proof[3]. Proposition 10.11 proves that the linear minimax risk when estimating $f \in \Theta$ from the deconvolved noisy data X satisfies $r_l(\Theta) \leq r_{\inf}(\mathrm{QH}[\Theta])$. The reverse inequality is obtained with the same derivations as in the proof of Theorem 10.7. The risk $r_{\inf}(\mathrm{QH}[\Theta])$ is reached by estimators that are diagonal in the discrete Fourier basis. ■

If Θ is closed and bounded, then there exists $x \in \mathrm{QH}[\Theta]$ such that $r_{\inf}(x) = r_{\inf}(\mathrm{QH}[\Theta])$. The minimax risk is then achieved by a filter whose transfer function $\hat{d}_1[m]$ is specified by (10.129). The resulting estimator is

$$\tilde{F} = D_1 X = d_1 \circledast X = d_1 \circledast \tilde{u}^{-1} \circledast Y \,.$$

So $\tilde{F} = DY = d \circledast Y$, and one can verify (Problem 10.16) that

$$\hat{d}[m] = \frac{N^{-1} |\hat{x}[m]|^2 \hat{u}^*[m]}{\sigma^2 + N^{-1} |\hat{x}[m]|^2 |\hat{u}[m]|^2} \,. \tag{10.154}$$

If $\sigma_m^2 = \sigma^2 |\hat{u}[m]|^{-2} \ll N^{-1} |\hat{x}[m]|^2$ then $\hat{d}[m] \approx \hat{u}^{-1}[m]$, but if $\sigma_m^2 \gg N^{-1} |\hat{x}[m]|^2$ then $\hat{d}[m] \approx 0$. The filter $\hat{d}$ is thus a regularized inverse of u.

Theorem 10.13 can be applied to a set of signals with bounded total variation

$$\Theta_V = \left\{ f \,:\, \|f\|_V = \sum_{n=0}^{N-1} \Big| f[n] - f[n-1] \Big| \leq C \right\} \,. \tag{10.155}$$

The set Θ_V is indeed translation invariant.

Proposition 10.13 *For a hyperbolic deconvolution of degree p, if* $N^{1/2} \leq C/\sigma \leq N^{r+1/2}$ *then*

$$\frac{r_l(\Theta_V)}{N\sigma^2} \sim \left(\frac{C}{N^{1/2}\sigma} \right)^{2-1/p} \,. \tag{10.156}$$

Proof[2]. Since Θ_V is translation invariant, Theorem 10.13 proves that $r_l(\Theta_V) = r_{\inf}(\mathrm{QH}[\Theta_V])$. Proposition 10.5 shows in (10.89) that all $f \in \Theta_V$ have a discrete Fourier transform that satisfies

$$|\hat{f}[m]|^2 \leq \frac{C^2}{4 |\sin \frac{\pi m}{N}|^2} = |\hat{x}[m]|^2 \,. \tag{10.157}$$

Hence Θ_V is included in the hyperrectangle $\mathcal{R}_x$. The convex hull $\mathrm{QH}[\Theta_V]$ is thus also included in $\mathcal{R}_x$ which is quadratically convex, and one can verify that

$$r_{\inf}(\mathrm{QH}[\Theta]_V) \leq r_{\inf}(\mathcal{R}_x) \leq 2 r_{\inf}(\mathrm{QH}[\Theta]_V) \,. \tag{10.158}$$

The value $r_{\inf}(\mathcal{R}_x)$ is calculated by inserting (10.157) with $\sigma_m^{-2} = \sigma^{-2} |\hat{u}[m]|^2$ in (10.152):

$$r_{\inf}(\mathcal{R}_x) = \sum_{m=0}^{N-1} \frac{N^{-1} C^2 \sigma^2}{4\sigma^2 |\sin \frac{\pi m}{N}|^2 + N^{-1} C^2 |\hat{u}[m]|^2} \,. \tag{10.159}$$

For $|\hat{u}[m]| \sim |2mN^{-1} - 1|^p$, if $1 \leq C/\sigma \leq N$ then an algebraic calculation gives $r_{\inf}(\mathcal{R}_x) \sim (CN^{-1/2}\sigma^{-1})^{(2p-1)/p}$. So $r_l(\Theta_V) = r_{\inf}(\mathrm{QH}[\Theta_V])$ satisfies (10.156). ■

The condition $N^{1/2} \leq C/\sigma \leq N^{r+1/2}$ imposes that the noise variation is sufficiently large so that the problem is indeed unstable, but not too large so that some high frequencies can be restored. The larger the number p of zeroes of the low-pass filter $\hat{u}[k]$ at $k = \pm N/2$, the larger the risk.

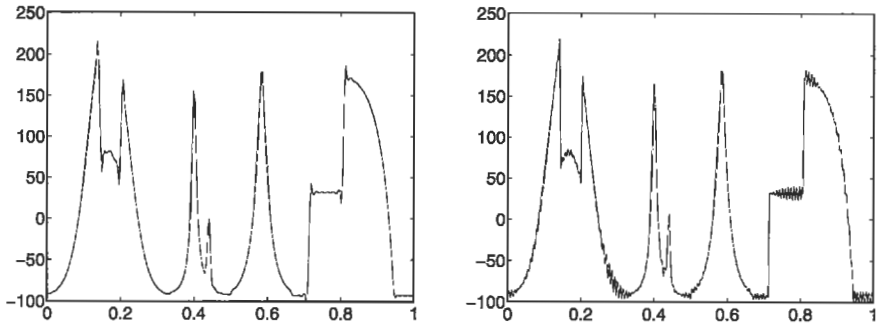


FIGURE 10.10 (a): Degraded data Y, blurred with the filter (10.160) and contaminated by a Gaussian white noise (SNR = 25.0 db). (b): Deconvolution calculated with a circular convolution estimator whose risk is close to the linear minimax risk over bounded variation signals (SNR = 25.8 db).

Example 10.3 Figure 10.10(a) is a signal Y obtained by smoothing a signal f with the low-pass filter

$$\hat{u}[m] = \cos^2\left(\frac{\pi m}{N}\right) . \tag{10.160}$$

This filter has a zero of order $p = 2$ at $\pm N/2$. Figure 10.10(b) shows the estimation $\tilde{F} = Y \circledast d$ calculated with the transfer function $\hat{d}[m]$ obtained by inserting (10.157) in (10.154). The maximum risk over Θ_V of this estimator is within a factor 2 of the linear minimax risk $r_l(\Theta_V)$.

Thresholding Deconvolution An efficient thresholding estimator is implemented in a basis $\mathcal{B}$ that defines a sparse representation of signals in Θ_V and which nearly diagonalizes K. This approach was introduced by Donoho [163] to study inverse problems such as inverse Radon transforms. We concentrate on more unstable hyperbolic deconvolutions.

The covariance operator K is diagonalized in the discrete Fourier basis and its eigenvalues are

$$\sigma_k^2 = \frac{\sigma^2}{|\hat{u}[k]|^2} \sim \sigma^2 \left|\frac{2k}{N} - 1\right|^{-2p} . \tag{10.161}$$

Yet the discrete Fourier basis is not appropriate for the thresholding algorithm because it does not provide efficient approximations of bounded variation signals. In contrast, periodic wavelet bases provide efficient approximations of such signals. We denote by $\psi_{0,0}[n] = N^{-1/2}$. A discrete and periodic orthonormal wavelet basis can be written

$$\mathcal{B} = \{\psi_{j,m}\}_{L < j \leq 0, 0 \leq m < 2^{-j}} . \tag{10.162}$$

However, we shall see that this basis fails to approximately diagonalize K.

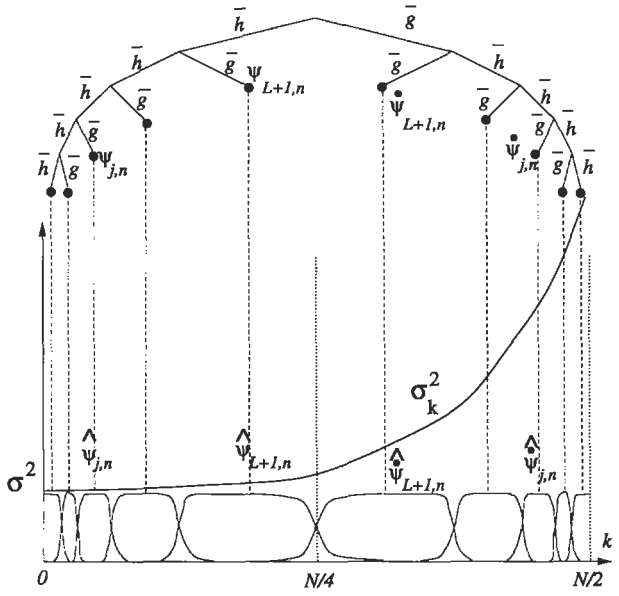


FIGURE 10.11 Wavelets and mirror wavelets are computed with a wavelet packet filter bank tree, where each branch corresponds to a convolution with a filter $\bar{h}$ or $\bar{g}$ followed by a subsampling. The graphs of the discrete Fourier transforms $|\hat{\psi}_{j,n}[k]|$ and $|\hat{\dot{\psi}}_{j,n}[k]|$ are shown below the tree. The variance σ_k^2 of the noise has a hyperbolic growth but varies by a bounded factor on the frequency support of each mirror wavelet.

The discrete Fourier transform $\hat{\psi}_{j,m}[k]$ of a wavelet has an energy mostly concentrated in the interval $[2^{-j-1}, 2^{-j}]$, as illustrated by Figure 10.11. If $2^j < 2N^{-1}$ then over this frequency interval (10.161) shows that the eigenvalues σ_k^2 remain on the order of σ^2. These wavelets are therefore approximate eigenvectors of K. At the finest scale $2^l = 2N^{-1}$, $|\hat{\psi}_{l,m}[k]|$ has an energy mainly concentrated in the higher frequency band $[N/4, N/2]$, where σ_k^2 varies by a huge factor on the order of N^{2r}. These fine scale wavelets are thus far from approximating eigenvectors of K.

To construct a basis of approximate eigenvectors of K, the finest scale wavelets must be replaced by wavelet packets that have a Fourier transform concentrated in subintervals of $[N/4, N/2]$ where σ_k^2 varies by a factor that does not grow with N. In order to efficiently approximate piecewise regular signals, these wavelet packets must also have the smallest possible spatial support, and hence the largest possible frequency support. The optimal trade-off is obtained with wavelet packets that we denote $\dot{\psi}_{j,m}$, which have a discrete Fourier transform $\hat{\dot{\psi}}_{j,m}[k]$ mostly concentrated in $[N/2 - 2^{-j}, N/2 - 2^{-j-1}]$, as illustrated by Figure 10.11. This basis is constructed

with a wavelet packet filtering tree that subdecomposes the space of the finest scale wavelets. These particular wavelet packets introduced by Kalifa and Mallat [232, 233] are called *mirror wavelets* because

$$|\hat{\dot{\psi}}_{j,m}[k]| = |\hat{\psi}_{j,m}[N/2-k]| \,.$$

Let $L = -\log_2 N$. A mirror wavelet basis is a wavelet packet basis composed of wavelets $\psi_{j,m}$ at scales $2^j < 2^{L-1}$ and mirror wavelets to replace the finest scale wavelets 2^{L-1}:

$$\mathcal{B} = \left\{ \psi_{j,m} \,,\, \dot{\psi}_{j,m} \right\}_{0 \le m < 2^{-j}, L-1 < j \le 0} .$$

To prove that the covariance K is "almost diagonalized" in $\mathcal{B}$ for all N, the asymptotic behavior of the discrete wavelets and mirror wavelets must be controlled. The following theorem thus supposes that these wavelets and wavelet packets are constructed with a conjugate mirror filter that yields a continuous time wavelet $\psi(t)$ with $q > p$ vanishing moments and which is $\mathbf{C}^q$. The near diagonalization is verified to prove that a thresholding estimator in a mirror wavelet basis has a risk whose decay is equivalent to the non-linear minimax risk.

Theorem 10.14 (KALIFA, MALLAT) *Let $\mathcal{B}$ be a mirror wavelet basis constructed with a conjugate mirror filter that defines a wavelet that is $\mathbf{C}^q$ with q vanishing moments. For a hyperbolic deconvolution of degree $p < q$ and $p > 1/2$, if $N^{1/2} \le C/\sigma \le N^{p+1/2}$ then*

$$\frac{r_n(\Theta_V)}{r_l(\Theta_V)} \sim \frac{r_t(\Theta_V)}{r_l(\Theta_V)} \sim \left(\frac{C^{1/p} \log_e N}{\sigma^{1/p} N^{1+1/2p}} \right)^{1/(2p+1)} . \tag{10.163}$$

Proof[3]. The main ideas of the proof are outlined. We must first verify that there exists λ such that for all $N > 0$

$$\| K_d^{1/2} K^{-1} K_d^{1/2} \|_S \le \lambda \,. \tag{10.164}$$

The operator $K^{-1} = \sigma^{-2} U^* U$ is a circular convolution whose transfer function is $\sigma^{-2} |\hat{u}[m]|^2 \sim \sigma^2 |2m/N - 1|^{2p}$. The matrix of this operator in the mirror wavelet basis is identical to the matrix in the discrete wavelet basis of a different circular convolution whose transfer function satisfies $\sigma^{-2} |\hat{u}[m+N/2]|^2 \sim \sigma^{-2} |2m/N|^{2p}$. This last operator is a discretized and periodized version of a convolution operator in $\mathbf{L}^2(\mathbb{R})$ of transfer function $\hat{u}(\omega) \sim \sigma^{-2} N^{-2p} |\omega|^{2p}$. One can prove [47, 221] that this operator is preconditioned by its diagonal in a wavelet basis of $\mathbf{L}^2(\mathbb{R})$ if the wavelet has $q > p$ vanishing moments and is $\mathbf{C}^q$. We can thus derive that in the finite case, when N grows, $\| K_d^{1/2} K^{-1} K_d^{1/2} \|_S$ remains bounded.

The minimax and thresholding risk cannot be calculated directly with the inequalities (10.144) because the set of bounded variation signals Θ_V is not orthosymmetric in the mirror wavelet basis $\mathcal{B}$. The proof proceeds as in Theorem 10.9. We first show that we can compute an upper bound and a lower bound of $\|f\|_V$ from the absolute value of the decomposition coefficients of f in the mirror wavelet basis $\mathcal{B}$. The resulting inequalities are similar to the wavelet ones in Proposition 10.10. This constructs two orthosymmetric sets Θ_1 and Θ_2 such that $\Theta_1 \subset \Theta_V \subset \Theta_2$. A refinement of the

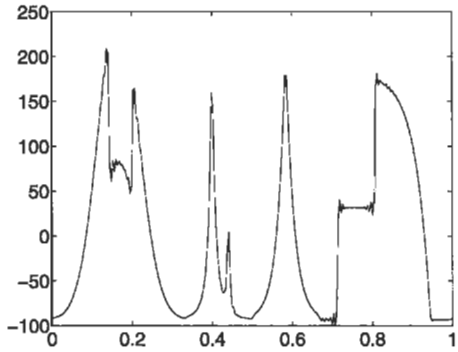


FIGURE 10.12 Deconvolution of the signal in Figure 10.10(a) with a thresholding in a mirror wavelet basis (SNR = 29.2 db).

inequalities (10.144) shows that over these sets the minimax and thresholding risks are equivalent, with no loss of a $\log_e N$ factor. The risk over Θ_1 and Θ_2 is calculated by evaluating $r_{\text{inf}}(\Theta_1)$ and $r_{\text{inf}}(\Theta_2)$, from which we derive (10.163), by using the expression (10.156) for $r_l(\Theta_V)$. ■

This theorem proves that a thresholding estimator in a mirror wavelet basis yields a quasi-minimax deconvolution estimator for bounded variation signals: $r_n(\Theta_V) \sim r_t(\Theta_V)$. Moreover, it shows that the thresholding risk $r_t(\Theta_V)$ is much smaller than the linear minimax risk $r_l(\Theta_V)$ as long as there is enough noise so that the inverse problem is indeed unstable. If $C/\sigma \sim N^{1/2}$ then

$$\frac{r_t(\Theta_V)}{r_l(\Theta_V)} \sim \left(\frac{\log_e N}{N}\right)^{1/(2p+1)}. \tag{10.165}$$

Example 10.4 Figure 10.10(a) shows a signal Y degraded by a convolution with a low-pass filter $\hat{u}[k] = \cos^2(\pi k/N)$. The result of the deconvolution and denoising with a thresholding in the mirror wavelet basis is shown in Figure 10.12. A translation invariant thresholding is performed to reduce the risk. The SNR is 29.2 db, whereas it was 25.8 db in the linear restoration of Figure 10.10(b).

Deconvolution of Satellite Images Nearly optimal deconvolution of bounded variation images can be calculated with a separable extension of the deconvolution estimator in a mirror wavelet basis. Such a restoration algorithm is used by the French spatial agency (CNES) for the production of satellite images.

The exposition time of the satellite photoreceptors cannot be reduced too much because the light intensity reaching the satellite is small and must not be dominated by electronic noises. The satellite movement thus produces a blur, which is aggravated by the imperfection of the optics. The electronics of the photoreceptors adds a Gaussian white noise. The image 10.14(b), provided by the CNES,

is a simulated satellite image calculated from the airplane image shown in Figure 10.14(a). The total variation of satellite images is often bounded, and since their amplitude is also bounded, they belong to a set

$$\Theta_{V,\infty} = \left\{ f \;:\; \|f\|_V \leq C_V \;,\; \|f\|_\infty \leq C_\infty \right\} .$$

When the satellite is in orbit, a calibration procedure measures the impulse response u of the blur and the noise variance σ^2. The impulse response is a separable low-pass filter:

$$Uf[n_1,n_2] = f \circledast u[n_1,n_2] \;\text{ with } u[n_1,n_2] = u_1[n_1]\,u_2[n_2] .$$

The discrete Fourier transform of u_1 and u_2 have respectively a zero of order p_1 and p_2 at $\pm N/2$:

$$\hat{u}_1[k_1] \sim \left| \frac{2k_1}{N} - 1 \right|^{p_1} \;\text{ and }\; \hat{u}_2[k_2] \sim \left| \frac{2k_2}{N} - 1 \right|^{p_2} .$$

The deconvolved noise has a covariance K that is diagonalized in a two-dimensional discrete Fourier basis. The eigenvalues are

$$\sigma^2_{k_1,k_2} = \frac{\sigma^2}{|\hat{u}_1[k_1]|^2\,|\hat{u}_2[k_2]|^2} \sim \sigma^2 \left| \frac{2k_1}{N} - 1 \right|^{-2p_1} \left| \frac{2k_2}{N} - 1 \right|^{-2p_2} . \tag{10.166}$$

Most satellite images are well modeled by bounded variation images. The main difficulty is again to find an orthonormal basis that provides a sparse representation of bounded variation images and which nearly diagonalizes the noise covariance K. Each vector of such a basis should have a Fourier transform whose energy is concentrated in a frequency domain where the eigenvectors $\sigma^2_{k_1,k_2}$ vary at most by a constant factor. Rougé [299, 300] has demonstrated numerically that efficient deconvolution estimations can be performed with a thresholding in a wavelet packet basis.

At low frequencies $(k_1,k_2) \in [0,N/4]^2$ the eigenvalues remain approximately constant: $\sigma^2_{k_1,k_2} \sim \sigma^2$. This frequency square can thus be covered with two-dimensional wavelets $\psi^I_{j,m}$. The remaining high frequency annulus is covered by two-dimensional mirror wavelets that are separable products of two one-dimensional mirror wavelets. One can verify that the union of these two families defines an orthonormal basis of images of N^2 pixels:

$$\mathcal{B} = \left(\left\{ \psi^I_{j,m}[n_1,n_2] \right\}_{j,m,I} , \left\{ \dot{\psi}_{j,m}[n_1]\,\dot{\psi}_{j',m'}[n_2] \right\}_{j,j',m,m'} \right) . \tag{10.167}$$

This two-dimensional mirror wavelet basis segments the Fourier plane as illustrated in Figure 10.13. It is an anisotropic wavelet packet basis as defined in Problem 8.4. Decomposing a signal in this basis with a filter bank requires $O(N^2)$ operations.

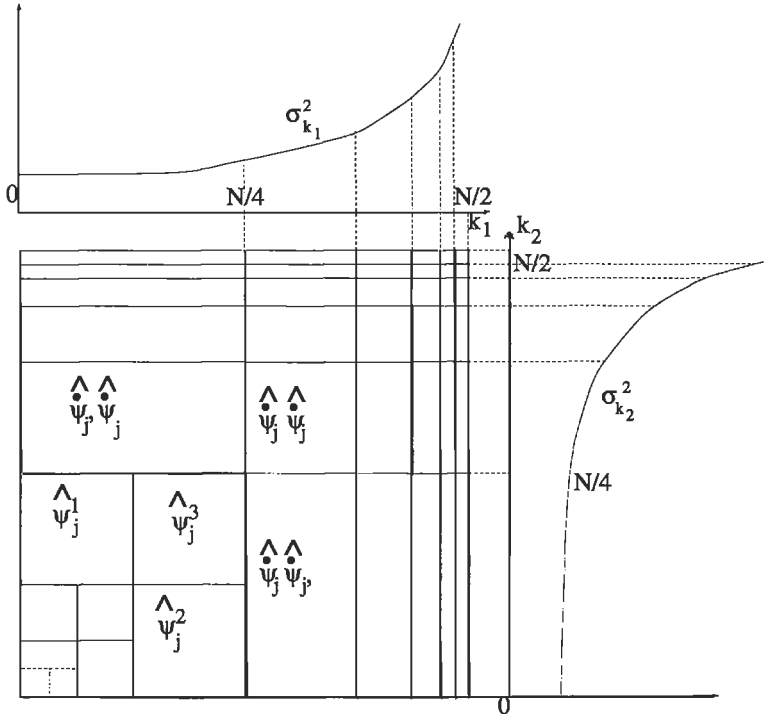


FIGURE 10.13 The mirror wavelet basis (10.167) segments the frequency plane (k_1, k_2) into rectangles over which the noise variance $\sigma^2_{k_1,k_2} = \sigma^2_{k_1}\,\sigma^2_{k_2}$ varies by a bounded factor. The lower frequencies are covered by separable wavelets ψ^k_j, and the higher frequencies are covered by separable mirror wavelets $\dot{\psi}_j \dot{\psi}_{j'}$.

To formally prove that a thresholding estimator in $\mathcal{B}$ has a risk $r_t(\Theta_{V,\infty})$ that is close to the non-linear minimax risk $r_n(\Theta_{V,\infty})$, one must prove that there exists λ such that $\|K_d^{1/2} K^{-1} K_d^{1/2}\|_S \leq \lambda$ and that $\Theta_{V,\infty}$ can be embedded in two close sets that are orthosymmetric in $\mathcal{B}$. Since this deconvolution problem is essentially separable, one can prove [232] that the minimax linear and non-linear risks as well as the thresholding risk are about N times larger than the risks calculated for one-dimensional signals in the set Θ_V defined in (10.155). Proposition 10.13 and Theorem 10.14 compute these risks. In two dimensions, it is however crucial to incorporate the fact that images have a bounded amplitude. The constant factors depend upon C_V and C_∞. The improvement of a thresholding estimator over an optimal linear estimator is then of the same order for a one-dimensional signal of size N and an image of N^2 pixels.

Figure 10.14(c) shows an example of deconvolution calculated in the mirror wavelet basis. The thresholding is performed with a translation invariant algo-

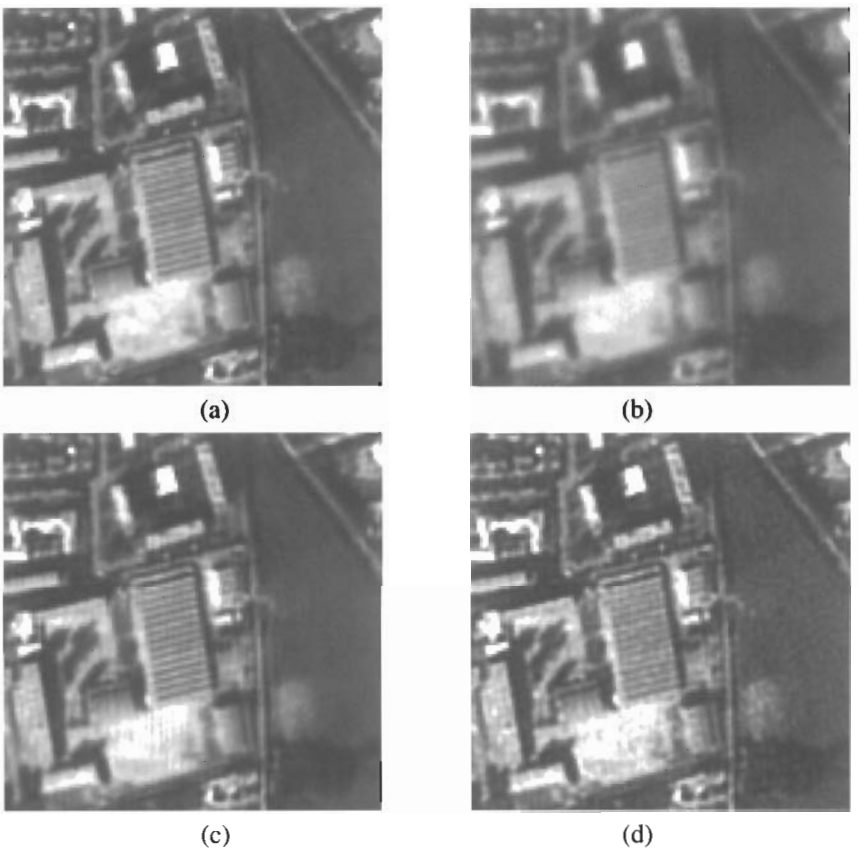


FIGURE 10.14 (a): Original airplane image. (b): Simulation of a satellite image provided by the CNES (SNR = 31.1db). (c): Deconvolution with a translation invariant thresholding in a mirror wavelet basis (SNR = 34.1db). (d): Deconvolution calculated with a circular convolution, which yields a nearly minimax risk for bounded variation images (SNR = 32.7db).

rithm. This can be compared with the linear estimation in Figure 10.14(d), calculated with a circular convolution estimator whose maximum risk over bounded variation images is close to the minimax linear risk. As in one dimension, the linear deconvolution sharpens the image but leaves a visible noise in the regular parts of the image. The thresholding algorithm completely removes the noise in these regions while improving the restoration of edges and oscillatory parts.

10.5 COHERENT ESTIMATION [3]

If we cannot interpret the information carried by a signal component, it is often misconstrued as noise. In a crowd speaking a foreign language, we perceive sur-

rounding conversations as background noise. In contrast, our attention is easily attracted by a remote conversation spoken in a known language. What is important here is not the information content but whether this information is in a coherent format with respect to our system of interpretation. The decomposition of a signal in a dictionary of vectors can similarly be considered as a signal interpretation [259]. Noises are then defined as signal components that do not have strong correlation with any vector of the dictionary. In the absence of any knowledge concerning the noise, a signal is estimated by isolating the coherent structures which have a high correlation with vectors in the dictionary. If the noise is not Gaussian, computing the estimation risk is much more difficult. This section introduces algorithms that can be justified intuitively, but which lack a firm mathematical foundation.

10.5.1 Coherent Basis Thresholding

Let $\mathcal{B} = \{g_m\}_{0 \leq m < N}$ be an orthonormal basis. If $W[n]$ is a Gaussian white process of size N and variance σ^2, then $\mathsf{E}\{\|W\|^2\} = N\sigma^2$ and the coefficients $\langle W, g_m \rangle$ are independent Gaussian random variables. When N increases there is a probability converging towards 1 that [9]

$$\frac{\max_{0 \leq m < N} |\langle W, g_m \rangle|}{\|W\|} \leq \frac{\sqrt{2 \log_e N}\, \sigma}{\sqrt{N}\, \sigma} = \frac{\sqrt{2 \log_e N}}{\sqrt{N}} = C_N. \tag{10.168}$$

The factor C_N is the maximum normalized correlation of a Gaussian white noise of size N.

The *correlation* of a signal f with the basis $\mathcal{B}$ is defined by

$$C(f) = \frac{\sup_{0 \leq m < N} |\langle f, g_m \rangle|}{\|f\|}.$$

We say that f is a noise with respect to $\mathcal{B}$ if it does not correlate vectors in $\mathcal{B}$ any better than a Gaussian white noise: $C(f) \leq C_N$. For example, $f[n] = e^{i\xi n}$ is a noise in a basis of discrete Diracs $g_m[n] = \delta[n-m]$, because

$$\frac{\sup_{0 \leq m < N} |f[m]|}{\|f\|} = \frac{1}{\sqrt{N}} < C_N.$$

Coherent Structures Let Z be an unknown noise. To estimate a signal f from $X = f + Z$, we progressively extract the vectors of $\mathcal{B}$ that best correlate X. Let us sort the inner products $\langle X, g_m \rangle$:

$$|\langle X, g_{m_k} \rangle| \geq |\langle X, g_{m_{k+1}} \rangle| \quad \text{for } 1 \leq k < N-1.$$

The data X is not reduced to a noise if

$$C(X) = \frac{|\langle X, g_{m_1} \rangle|}{\|X\|} > C_N.$$

The vector g_{m_1} is then interpreted as a *coherent structure*.

For any $k \geq 1$, we consider

$$R^k X = X - \sum_{p=1}^{k} \langle X, g_{m_p} \rangle \, g_{m_p} = \sum_{p=k+1}^{N} \langle X, g_{m_p} \rangle \, g_{m_p} .$$

The residue $R^k X$ is the orthogonal projection of X in a space of dimension $N-k$. The normalized correlation of this residue with vectors in $\mathcal{B}$ is compared with the normalized correlation of a Gaussian white noise of size $N-k$. This residue is not a noise if

$$C^2(R^k X) = \frac{|\langle X, g_{m_k} \rangle|^2}{\sum_{p=k+1}^{N} |\langle X, g_{m_p} \rangle|^2} > C_{N-k}^2 = \frac{2 \log_e (N-k)}{N-k} .$$

The vector g_{m_k} is then also a coherent structure.

Let M be the minimum index such that

$$C(R^M X) \leq C_{N-M} . \tag{10.169}$$

Observe that M is a random variable whose values depend on each realization of X. The signal f is estimated by the sum of the $M-1$ coherent structures:

$$\tilde{F} = \sum_{p=1}^{M-1} \langle X, g_{m_p} \rangle \, g_{m_p} .$$

This estimator is also obtained by thresholding the coefficients $\langle X, g_m \rangle$ with the threshold value

$$T = C_{N-M} \left(\sum_{p=M}^{N-1} |\langle X, g_{m_p} \rangle|^2 \right)^{1/2} . \tag{10.170}$$

The extraction of coherent structures can thus be interpreted as a calculation of an appropriate threshold for estimating f, in the absence of any knowledge about the noise. This algorithm estimates f efficiently only if most of its energy is concentrated in the direction of few vectors g_m in $\mathcal{B}$. For example, $f = \sum_{m=0}^{N-1} g_m$ has no coherent structures because $C(f) = N^{-1/2} < C_N$. Even though $Z = 0$, the extraction of coherent structures applied to $X = f$ yields $\tilde{F} = 0$. This indicates that the basis representation is not well adapted to the signal.

Figure 10.15(a) shows a piecewise regular signal contaminated by the addition of a complex noise, which happens to be an old musical recording of Enrico Caruso. Suppose that we want to remove this "musical noise." The coherent structures are extracted using a wavelet basis, which approximates piecewise smooth functions efficiently but does not correlate well with high frequency oscillations. The estimation in Figure 10.15(b) shows that few elements of the musical noise are coherent structures relative to the wavelet basis. If instead of this musical noise

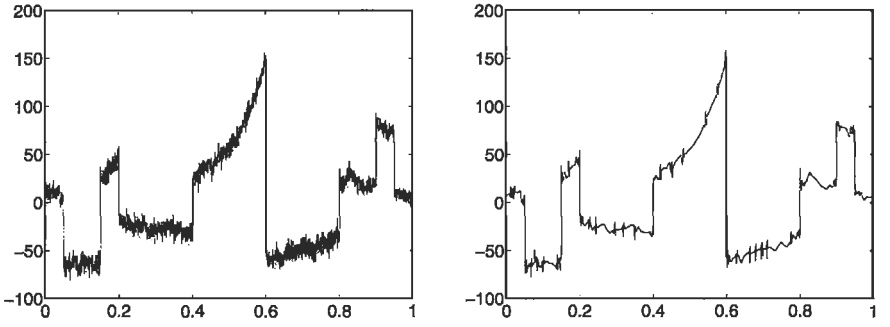


FIGURE 10.15 (a): The same signal as in Figure 10.4 to which is added a noisy musical signal (SNR = 19.3 db). (b): Estimation by extracting coherent structures in a Daubechies 4 wavelet basis (SNR = 23.0 db).

a Gaussian white noise of variance σ is added to this piecewise smooth signal, then the coherent structure algorithm computes an estimated threshold (10.170) that is within 10% of the threshold $T = \sigma\sqrt{2\log_e N}$ used for white noises. The estimation is therefore very similar to the hard thresholding estimation in Figure 10.4(c).

Pursuit of Bases No single basis can provide a "coherent" interpretation of complex signals such as music recordings. To remove noise from historical recordings, Berger, Coifman and Goldberg [92] introduced an orthogonal basis pursuit algorithm that searches a succession of "best bases." Excellent results have been obtained on the restoration the recording of Enrico Caruso. In this case, we must extract coherent structures corresponding to the original musical sound as opposed to the degradations of the recording. The coherent extraction shown in Figure 10.15(b) demonstrates that hardly any component of this recording is highly coherent in the Daubechies 4 wavelet basis. It is therefore necessary to search for other bases that match the signal properties.

Let $\mathcal{D} = \cup_{\lambda\in\Lambda}\mathcal{B}^\lambda$ be a dictionary of orthonormal bases. To find a basis in $\mathcal{D}$ that approximates a signal f efficiently, Section 9.4.1 selects a best basis $\mathcal{B}^\alpha$ that minimizes a Schur concave cost function

$$C(f,\mathcal{B}^\lambda) = \sum_{m=1}^{N-1} \Phi\left(\frac{|\langle f, g_m^\lambda\rangle|^2}{\|f\|^2}\right),$$

where $\Phi(x)$ is a concave function, possibly an entropy (9.68) or an $\mathbf{l}^p$ norm (9.70). A pursuit of orthogonal bases extracts coherent structures from noisy data X with an iterative procedure that computes successive residues that we denote X_p:

1. *Initialization* $X_0 = X$.

2. *Basis search* A best basis $\mathcal{B}^{\alpha_p}$ is selected in $\mathcal{D}$ by minimizing a cost:

$$C(X_p, \mathcal{B}^{\alpha_p}) = \min_{\lambda \in \Lambda} C(X_p, \mathcal{B}^{\lambda}) .$$

3. *Coherent calculation* Coherent structures are extracted as long as $C(R^k X_p) > C_{N-k}$ in $\mathcal{B}^{\alpha_p}$. Let M_p be the number of coherent structures defined by $C(R^{M_p} X_p) \leq C_{N-M_p}$. The remainder is

$$X_{p+1} = R^{M_p} X_p .$$

4. *Stopping rule* If $M_p = 0$, stop. Otherwise, go to step 2.

For musical signals [92], the pursuit of bases is performed in a general dictionary that is the union of a dictionary of local cosine bases and a dictionary of wavelet packet bases, introduced respectively in Sections 8.1 and 8.5. In each dictionary, a best basis is calculated with an entropy function $\Phi(x)$ and is selected by the fast algorithm of Section 9.4.2. The best of these two "best" bases is retained. To take into account some prior knowledge about the noise and the properties of musical recordings, the correlation $C(f)$ used to extract coherent structures can be modified, and further ad-hoc refinements can be added [92].

10.5.2 Coherent Matching Pursuit

A matching pursuit offers the flexibility of searching for coherent structures in arbitrarily large dictionaries of patterns $\mathcal{D} = \{g_\gamma\}_{\gamma \in \Gamma}$, which can be designed depending on the properties of the signal. No orthogonal condition is imposed. The notions of coherent structure and noise are redefined by analyzing the asymptotic properties of the matching pursuit residues.

Dictionary Noise A matching pursuit decomposes f over selected dictionary vectors with the greedy strategy described in Section 9.5.2. Theorem 9.10 proves that the residue $R^m f$ calculated after m iterations of the pursuit satisfies $\lim_{m \to +\infty} \|R^m f\| = 0$.

The matching pursuit behaves like a non-linear chaotic map, and it has been proved by Davis, Mallat and Avelaneda [151] that for particular dictionaries, the normalized residues $R^m f \, \|R^m f\|^{-1}$ converge to an attractor. This attractor is a set of signals h that do not correlate well with any $g_\gamma \in \mathcal{D}$ because all coherent structures of f in $\mathcal{D}$ are removed by the pursuit. The *correlation* of a signal f with the dictionary $\mathcal{D}$ is defined by

$$C(f) = \frac{\sup_{\gamma \in \Gamma} |\langle f, g_\gamma \rangle|}{\|f\|} .$$

For signals in the attractor, this correlation has a small amplitude that remains nearly equal to a constant $C_{\mathcal{D}}$, which depends on the dictionary $\mathcal{D}$ [151]. Such signals do not correlate well with any dictionary vector and are thus considered as noise with respect to $\mathcal{D}$.

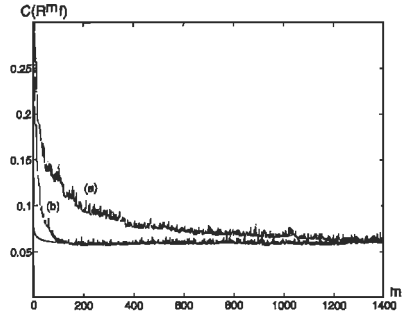


FIGURE 10.16 Decay of the correlation $C(R^m f)$ as a function of the number of iterations m, for two signals decomposed in a Gabor dictionary. (a): f is the recording of "greasy" shown in Figure 10.17(a). (b): f is the noisy "greasy" signal shown in Figure 10.17(b).

The convergence of the pursuit to the attractor implies that after a sufficiently large number M of iterations the residue $R^M f$ has a correlation $C(R^M f)$ that is nearly equal to $C_{\mathcal{D}}$. Figure 10.16 gives the decay of $C(R^m f)$ as a function of m, for two signals decomposed in a Gabor dictionary. After respectively $M = 1400$ and $M = 76$ iterations, both curves reach the attractor level $C_{\mathcal{D}} = 0.06$.

Coherent Pursuit Coherent structures are progressively extracted to estimate f from $X = f + Z$. These coherent structures are dictionary vectors selected by the pursuit, and which are above the noise level $C_{\mathcal{D}}$. For any $m \geq 0$, the matching pursuit projects the residue $R^k X$ on a vector $g_{\gamma_k} \in \mathcal{D}$ such that

$$|\langle R^k X, g_{\gamma_k} \rangle| = \sup_{\gamma \in \Gamma} |\langle R^k X, g_\gamma \rangle|.$$

The vector g_{γ_k} is a coherent structure of $R^k X$ if

$$C(R^k f) = \frac{|\langle R^k X, g_{\gamma_k} \rangle|}{\|R^k X\|} > C_{\mathcal{D}}.$$

Let M be the minimum integer such that $C(R^M f) \leq C_{\mathcal{D}}$. The residue $R^M X$ has reached the noise level and is therefore not further decomposed. The signal is estimated from the M coherent structures:

$$\tilde{F} = \sum_{p=0}^{M-1} \langle R^p X, g_{\gamma_p} \rangle \, g_{\gamma_p}.$$

This estimator can also be interpreted as a thresholding of the matching pursuit of X with a threshold that is adaptively adjusted to

$$T = C_{\mathcal{D}} \, \|R^M X\|.$$

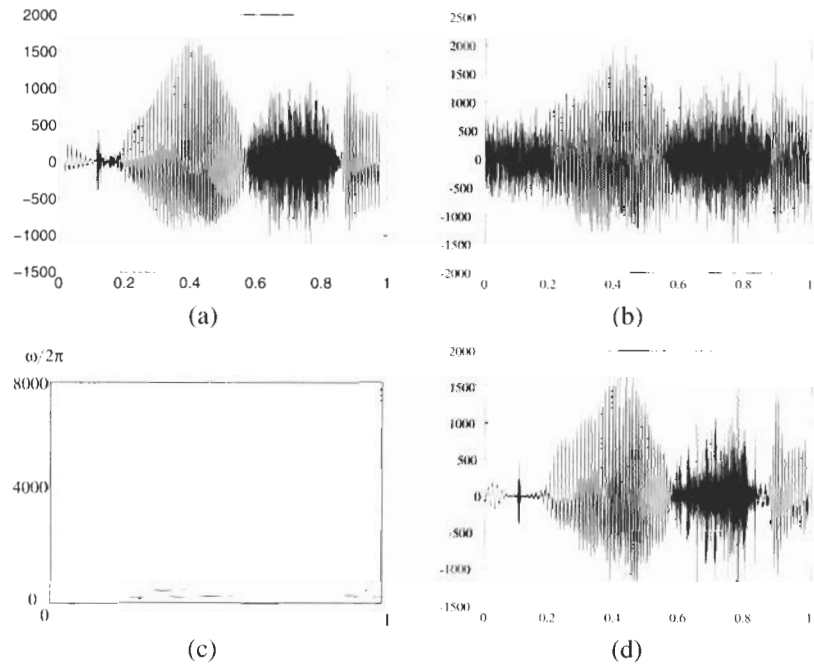


FIGURE 10.17 (a): Speech recording of "greasy." (b): Recording of "greasy" plus a Gaussian white noise (SNR = 1.5 db). (c): Time-frequency distribution of the $M = 76$ coherent Gabor structures. (d): Estimation $\tilde{F}$ reconstructed from the 76 coherent structures (SNR = 6.8 db).

Example 10.5 Figure 10.17(b) from [259] shows the speech recording of "greasy" contaminated with a Gaussian white noise, with an SNR of 1.5 db. The curve (b) of Figure 10.16 shows that the correlation $C(R^m f)$ reaches $C_{\mathcal{D}}$ after $m = M = 76$ iterations. The time-frequency energy distribution of these 76 Gabor atoms is shown in Figure 10.16(c). The estimation $\tilde{F}$ calculated from the 76 coherent structures is shown in Figure 10.17(d). The SNR of this estimation is 6.8 db. The white noise has been removed and the restored speech signal has a good intelligibility because its main time-frequency components are retained.

10.6 SPECTRUM ESTIMATION [2]

A zero-mean Gaussian process X of size N is characterized by its covariance matrix. For example, unvoiced speech sounds such as "ch" or "s" can be considered as realizations of Gaussian processes, which allows one to reproduce intelligible sounds if the covariance is known. The estimation of covariance matrices is difficult because we generally have few realizations, and hence few data points,

compared to the N^2 covariance coefficients that must be estimated. If parametrized models are available, which is the case for speech recordings [61], then a direct estimation of the parameters can give an accurate estimation of the covariance [60]. This is however not the case for complex processes such as general sounds or seismic and underwater signals. We thus follow a non-parametrized approach that applies to non-stationary processes.

When the Karhunen-Loève basis is known in advance, one can reduce the estimation to the N diagonal coefficients in this basis, which define the *power spectrum*. This is the case for stationary processes, where the Karhunen-Loève basis is known to be the Fourier basis. For non-stationary processes, the Karhunen-Loève basis is not known, but it can be approximated by searching for a "best basis" in a large dictionary of orthogonal bases. This approach is illustrated with locally stationary processes, where the basis search is performed in a dictionary of local cosine bases.

10.6.1 Power Spectrum

We want to estimate the covariance matrix of a zero-mean random vector X of size N from L realizations $\{X_k\}_{0\leq k<L}$. Let $\mathcal{B} = \{g_m\}_{0\leq m<N}$ be an orthonormal basis. The N^2 covariance coefficients of the covariance operator K are

$$a[l,m] = \langle Kg_l, g_m\rangle = \mathrm{E}\{\langle X, g_l\rangle \langle X, g_m\rangle^*\}.$$

When L is much smaller than N, which is most often the case, a naive estimation of these N^2 covariances gives disastrous results. In signal processing, the estimation must often be done with only $L = 1$ realization.

Naive Estimation Let us try to estimate the covariance coefficients with sample mean estimators

$$\bar{A}[l,m] = \frac{1}{L}\sum_{k=1}^{L} \langle X_k, g_l\rangle \langle X_k, g_m\rangle^*. \tag{10.171}$$

We denote by $\bar{K}$ the estimated covariance matrix whose coefficients are the $\bar{A}[l,m]$. The estimation error is measured with a Hilbert-Schmidt norm. The squared Hilbert-Schmidt norm of an operator K is the sum of its squared matrix coefficients, which is also equal to the trace of the product of K and its complex transpose K^*:

$$\|K\|_H^2 = \sum_{l,m=0}^{N-1} |a[l,m]|^2 = \mathrm{tr}(KK^*).$$

The Hilbert-Schmidt error of the covariance estimation is

$$\|K - \bar{K}\|_H^2 = \sum_{l,m=0}^{N-1} |a[l,m] - \bar{A}[l,m]|^2.$$

The following proposition computes its expected value when X is a Gaussian random vector.

Proposition 10.14 *If X is a Gaussian random vector then*

$$\mathsf{E}\left\{|\bar{A}[l,m]-a[l,m]|^2\right\}=\frac{1}{L}\Big(|a[l,m]|^2+a[l,l]\,a[m,m]\Big), \tag{10.172}$$

and

$$\mathsf{E}\{\|\bar{K}-K\|_H^2\}=\frac{\|K\|_H^2}{L}+\frac{\mathsf{E}^2\{\|X\|^2\}}{L}. \tag{10.173}$$

Proof[2]. The sample mean-estimator (10.171) is unbiased:

$$\mathsf{E}\{\bar{A}[l,m]\}=a[l,m],$$

so

$$\mathsf{E}\left\{|\bar{A}[l,m]-a[l,m]|^2\right\}=\mathsf{E}\left\{|\bar{A}[l,m]|^2\right\}-|a[l,m]|^2. \tag{10.174}$$

Let us compute $\mathsf{E}\{|\bar{A}[l,m]|^2\}$.

$$\begin{aligned}\mathsf{E}\{|\bar{A}[l,m]|^2\} &= \mathsf{E}\left\{\left|\frac{1}{L}\sum_{k=1}^{L}\langle X_k,g_l\rangle\,\langle X_k,g_m\rangle^*\right|^2\right\}\\ &= \frac{1}{L^2}\sum_{k=1}^{L}\mathsf{E}\left\{|\langle X_k,g_l\rangle|^2\,|\langle X_k,g_m\rangle|^2\right\}+\\ &\quad \frac{1}{L^2}\sum_{\substack{k,j=1\\k\neq j}}^{L}\mathsf{E}\left\{\langle X_k,g_l\rangle\,\langle X_k,g_m\rangle^*\right\}\mathsf{E}\left\{\langle X_j,g_l\rangle^*\,\langle X_j,g_m\rangle\right\}.\end{aligned} \tag{10.175}$$

Each $\langle X_k,g_l\rangle$ is a Gaussian random variable and for all k

$$\mathsf{E}\left\{\langle X_k,g_l\rangle\,\langle X_k,g_m\rangle^*\right\}=a[l,m].$$

If A_1,A_2,A_3,A_4 are jointly Gaussian random variables, one can verify that

$$\mathsf{E}\{A_1A_2A_3A_4\}=\mathsf{E}\{A_1A_2\}\mathsf{E}\{A_3A_4\}+\mathsf{E}\{A_1A_3\}\mathsf{E}\{A_2A_4\}+\mathsf{E}\{A_1A_4\}\mathsf{E}\{A_2A_3\}.$$

Applying this result to (10.175) yields

$$\mathsf{E}\left\{|\bar{A}[l,m]|^2\right\}=\frac{1}{L^2}L\left(a[l,l]\,a[m,m]+2|a[l,m]|^2\right)+\frac{1}{L^2}(L^2-L)\,|a[l,m]|^2,$$

so

$$\mathsf{E}\left\{|\bar{A}[l,m]|^2\right\}=\left(1+\frac{1}{L}\right)|a[l,m]|^2+\frac{1}{L}a[l,l]\,a[m,m].$$

We thus derive (10.172) from (10.174).

The Hilbert-Schmidt norm is

$$\begin{aligned}\mathsf{E}\{\|\bar{K}-K\|_H^2\} &= \sum_{l,m=0}^{N-1}\mathsf{E}\{|a[l,m]-\bar{A}[l,m]|^2\}\\ &= \frac{1}{L}\sum_{l,m=0}^{N-1}|a[l,m]|^2+\frac{1}{L}\sum_{l,m=0}^{N-1}a[l,l]\,a[m,m].\end{aligned}$$

Observe that

$$\mathsf{E}\{\|X\|^2\} = \sum_{m=0}^{N-1} \mathsf{E}\{|\langle X, g_m\rangle|^2\} = \sum_{m=0}^{N-1} a[m,m].$$

Inserting this in the previous equation gives (10.173). ■

The error calculation (10.172) proves that $\mathsf{E}\{|\bar{A}[l,m]-a[l,m]|^2\}$ depends not only on $|a[l,m]|^2$ but also on the amplitude of the diagonal coefficients $a[l,l]$ and $a[m,m]$. Even though $a[l,m]$ may be small, the error of the sample mean estimator is large if the diagonal coefficients are large:

$$\mathsf{E}\left\{|\bar{A}[l,m]-a[l,m]|^2\right\} \geq \frac{a[l,l]\,a[m,m]}{L}. \tag{10.176}$$

The error produced by estimating small amplitude covariance coefficients accumulates and produces a large Hilbert-Schmidt error (10.173).

Example 10.6 Suppose that X is a random vector such that $\mathsf{E}\{|X[n]|^2\}$ is on the order of σ^2 but that $\mathsf{E}\{X[n]X[m]\}$ decreases quickly when $|n-m|$ increases. The Hilbert-Schmidt norm of K can be calculated in a Dirac basis $g_m[n] = \delta[n-m]$, which gives

$$\|K\|_H^2 = \sum_{l,m=0}^{N-1} |\mathsf{E}\{X[l]X[m]\}|^2 \sim N\sigma^2,$$

and

$$\mathsf{E}\{\|X\|^2\} = \sum_{n=0}^{N-1} \mathsf{E}\left\{|X[n]|^2\right\} \sim N\sigma^2.$$

As a consequence, for $N \gg L$,

$$\mathsf{E}\left\{\|K-\bar{K}\|_H^2\right\} \geq \frac{\mathsf{E}^2\{\|X\|^2\}}{L} \sim \frac{\sigma^4 N^2}{L} \gg \|K\|_H^2.$$

The estimation error is huge; a better result is obtained by simply setting $\bar{K}=0$.

Power Spectrum If we know in advance the Karhunen-Loève basis that diagonalizes the covariance operator, we can avoid estimating off-diagonal covariance coefficients by working in this basis. The N diagonal coefficients $p[m] = a[m,m]$ are the eigenvalues of K, and are called its *power spectrum*.

We denote by $\bar{P}[m] = \bar{A}[m,m]$ the sample mean estimator along the diagonal. The sample mean error is computed with (10.172):

$$\mathsf{E}\left\{|\bar{P}[m]-p[m]|^2\right\} = \frac{2|p[m]|^2}{L}. \tag{10.177}$$

Since the covariance is diagonal,

$$\|K\|_H^2 = \sum_{m=0}^{N-1} |p[m]|^2 = \|p\|^2. \tag{10.178}$$

The estimated diagonal operator $\bar{K}$ with diagonal coefficients $\bar{P}[m]$ has therefore an expected error

$$\mathsf{E}\{\|\bar{K}-K\|_H^2\} = \mathsf{E}\{\|\bar{P}-p\|^2\} = \sum_{m=0}^{N-1} \frac{2\,|p[m]|^2}{L} = \frac{2\,\|K\|_H^2}{L}. \tag{10.179}$$

The relative error $\mathsf{E}\{\|\bar{K}-K\|_H^2\}/\|K\|_H^2$ decreases when L increases but it is independent of N. To improve this result, we must regularize the estimation of the power spectrum $p[m]$.

Regularization Sample mean estimations $\bar{P}[m]$ can be regularized if $p[m]$ varies slowly when m varies along the diagonal. These random coefficients can be interpreted as "noisy" measurements of $p[m]$:

$$\bar{P}[m] = p[m]\,(1+W[m]).$$

Since $\bar{P}[m]$ is unbiased, $\mathsf{E}\{W[m]\} = 0$. To transform the multiplicative noise into an additive noise, we compute

$$\log_e \bar{P}[m] = \log_e p[m] + \log_e(1+W[m]). \tag{10.180}$$

If $X[n]$ is Gaussian, then $W[m]$ has a χ_2^2 distribution [40], and (10.177) proves that

$$\mathsf{E}\{|W[m]|^2\} = \frac{2}{L}.$$

The coefficients $\{\langle X, g_m\rangle\}_{0\le m<N}$ of a Gaussian process in a Karhunen-Loève basis are independent variables, so $\bar{P}[m]$ and $\bar{P}[l]$ and hence $W[m]$ and $W[l]$ are independent for $l \neq m$. As a consequence, $W[m]$ and $\log_e(1+W[m])$ are non-Gaussian white noises.

In the Gaussian case, computing a regularized estimate $\tilde{P}[m]$ of $p[m]$ from (10.180) is a white noise removal problem. Let $\tilde{K}$ be the diagonal matrix whose diagonal coefficients are $\tilde{P}[m]$. This matrix is said to be a *consistent* estimator of K if

$$\lim_{N\to+\infty} \frac{\mathsf{E}\{\|K-\tilde{K}\|_H^2\}}{\|K\|_H^2} = \lim_{N\to+\infty} \frac{\mathsf{E}\{\|p-\tilde{P}\|^2\}}{\|p\|^2} = 0.$$

Linear estimations and Wiener type filters perform a weighted average with a kernel whose support covers a domain where $\log_e p[m]$ is expected to have small variations. This is particularly effective if $p[m]$ is uniformly regular.

If $p[m]$ is piecewise regular, then wavelet thresholding estimators improve the regularization of linear smoothings [188]. Following the algorithm of Section 10.2.4, the wavelet coefficients of $\log_e \bar{P}[m]$ are thresholded. Despite the fact that $\log_e(1+W[m])$ is not Gaussian, if $X[n]$ is Gaussian then results similar to Theorem 10.4 are proved [344] by verifying that wavelet coefficients have asymptotic Gaussian properties.

Stationary Processes If X is circular wide-sense stationary, then its covariance operator is a circular convolution that is diagonalized in the discrete Fourier basis

$$\left\{ g_m[n] = \frac{1}{\sqrt{N}} \exp\left(\frac{i2\pi mn}{N}\right) \right\}_{0 \le m < N} .$$

The power spectrum is the discrete Fourier transform of the covariance $R_X[l] = \mathrm{E}\{X[n]X[n-l]\}$:

$$\hat{R}_X[m] = \sum_{l=0}^{N-1} R_X[l] \exp\left(\frac{-i2m\pi l}{N}\right) = \mathrm{E}\{|\langle X, g_m\rangle|^2\} .$$

It is estimated with only $L = 1$ realization by computing $\bar{P}[m]$, which is called a *periodogram* [60]:

$$\bar{P}[m] = |\langle X, g_m\rangle|^2 = \frac{1}{N} \left| \sum_{n=0}^{N-1} X[n] \exp\left(\frac{-i2\pi mn}{N}\right) \right|^2 . \tag{10.181}$$

Most often, the stationarity of X is not circular and we only know the restriction of its realizations to $[0, N-1]$. The discrete Fourier basis is thus only an approximation of the true Karhunen-Loève basis, and this approximation introduces a bias in the spectrum estimation. This bias is reduced by pre-multiplying $X[n]$ with a smooth window $g[n]$ of size N, which removes the discontinuities introduced by the Fourier periodization. Such discrete windows are obtained by scaling and sampling one of the continuous time windows $g(t)$ studied in Section 4.2.2. This windowing technique can be improved by introducing several orthogonal windows whose design is optimized in [331].

To obtain a consistent estimator from the periodogram $\bar{P}[m]$, it is necessary to perform a regularization, as previously explained. If the spectrum is uniformly regular, then a linear filtering can yield a consistent estimator [60]. Figure 10.18(c) shows a regularized log periodogram calculated with such a linear filtering. The random fluctuations are attenuated but the power spectrum peaks are smoothed. A linear filtering of the spectra is more often implemented with a time windowing procedure, described in Problem 10.19. The interval $[0, N-1]$ is divided in M subintervals with windows of size N/M. A periodogram is computed over each interval and a regularized estimator of the power spectrum is obtained by averaging these M periodograms. Wavelet thresholdings can also be used to regularize piecewise smooth spectra [344].

10.6.2 Approximate Karhunen-Loève Search [3]

If X is non-stationary, we generally do not know in advance its Karhunen-Loève basis. But we may have prior information that makes it possible to design a dictionary of orthonormal bases guaranteed to contain at least one basis that closely

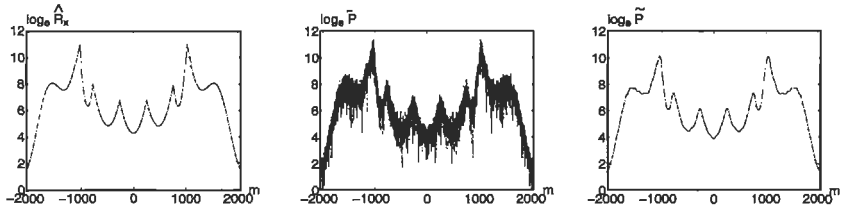


FIGURE 10.18 (a): Log power spectrum $\log_e \hat{R}_X[m]$ of a stationary process $X[n]$. (b): Log periodogram $\log_e \bar{P}[m]$ computed from $L = 1$ realization. (c): Linearly regularized estimator $\log_e \tilde{P}[m]$.

approximates the Karhunen-Loève basis. Locally stationary processes are examples where an approximate Karhunen-Loève basis can be found in a dictionary of local cosine bases. The algorithm of Mallat, Papanicolaou and Zhang [260] estimates this best basis by minimizing a negative quadratic sum. This is generalized to other Schur concave cost functions, including the entropy used by Wickerhauser [76].

Diagonal Estimation Proposition 10.14 proves that an estimation of all covariance coefficients produces a tremendous estimation error. Even though a basis $\mathcal{B}$ is not a Karhunen-Loève basis, it is often preferable to estimate the covariance K with a diagonal matrix $\tilde{K}$, which is equivalent to setting the off-diagonal coefficients to zero. The N diagonal coefficients $\tilde{P}[m]$ are computed by regularizing the sample mean-estimators (10.171). They approximate the spectrum of K.

The Hilbert-Schmidt error is the sum of the diagonal estimation errors plus the energy of the off-diagonal coefficients:

$$\|\tilde{K} - K\|_H^2 = \sum_{m=0}^{N-1} |\tilde{P}[m] - p[m]|^2 + \sum_{\substack{l,m=0 \\ l \neq m}}^{N-1} |a[l,m]|^2.$$

Since

$$\|K\|_H^2 = \sum_{l,m=0}^{N-1} |a[l,m]|^2 = \sum_{m=0}^{N-1} |p[m]|^2 + \sum_{\substack{l,m=0 \\ l \neq m}}^{N-1} |a[l,m]|^2,$$

we have

$$\|\tilde{K} - K\|_H^2 = \sum_{m=0}^{N-1} |\tilde{P}[m] - p[m]|^2 + \|K\|_H^2 - \sum_{m=0}^{N-1} |p[m]|^2. \tag{10.182}$$

Let us denote

$$C(K, \mathcal{B}) = -\sum_{m=0}^{N-1} |p[m]|^2. \tag{10.183}$$

Clearly $C(K,\mathcal{B}) \geq -\|K\|_H$ and this sum is minimum in a Karhunen-Loève basis $\mathcal{B}^{KL}$ where $C(K,\mathcal{B}^{KL}) = -\|K\|_H^2$. The error (10.182) can thus be rewritten

$$\|\tilde{K} - K\|_H^2 = \|\tilde{P} - p\|^2 + C(K,\mathcal{B}) - C(K,\mathcal{B}^{KL}). \tag{10.184}$$

Best Basis Let $\mathcal{D} = \{\mathcal{B}^\gamma\}_{\gamma\in\Gamma}$ be a dictionary of orthonormal bases $\mathcal{B}^\gamma = \{g_m^\gamma\}_{0\leq m<N}$. The error formulation (10.184) suggests defining a "best" Karhunen-Loève approximation as the basis that minimizes $C(K,\mathcal{B})$. Since we do not know the true diagonal coefficients $p[m]$, this cost is estimated with the regularized sample mean coefficients:

$$\tilde{C}(K,\mathcal{B}) = -\sum_{m=0}^{N-1} |\tilde{P}[m]|^2. \tag{10.185}$$

The covariance estimation thus proceeds as follows.

1. *Sample means* For each vector $g_m^\gamma \in \mathcal{D}$, we compute the sample mean estimator of the variance in the direction of each $g_m^\gamma \in \mathcal{D}$:

$$\bar{P}^\gamma[m] = \frac{1}{L}\sum_{k=1}^{L} |\langle X_k, g_m^\gamma\rangle|^2. \tag{10.186}$$

2. *Regularization* Regularized estimators $\tilde{P}^\gamma[m]$ are calculated with a local averaging or a wavelet thresholding among a particular group of dictionary vectors.
3. *Basis choice* The cost of K is estimated in each basis $\mathcal{B}^\gamma$ by

$$\tilde{C}(K,\mathcal{B}^\gamma) = -\sum_{m=0}^{N-1} |\tilde{P}^\gamma[m]|^2, \tag{10.187}$$

and we search for the best basis $\mathcal{B}^\alpha$ that minimizes these costs:

$$\tilde{C}(K,\mathcal{B}^\alpha) = \inf_{\gamma\in\Gamma} \tilde{C}(K,\mathcal{B}^\gamma). \tag{10.188}$$

4. *Estimation* The covariance K is estimated by the operator $\tilde{K}^\alpha$ that is diagonal in $\mathcal{B}^\alpha$, with diagonal coefficients equal to $\tilde{P}^\alpha[m]$.

Since $C(K,\mathcal{B}^{KL}) = -\|K\|_H^2$ and $\|K\|_H^2 \geq \|p^\alpha\|^2$, to evaluate the consistency of this algorithm, we derive from (10.184) that

$$\frac{\|\tilde{K}^\alpha - K\|_H^2}{\|K\|_H^2} \leq \frac{\|\tilde{P}^\alpha - p^\alpha\|^2}{\|p^\alpha\|^2} + \frac{C(K,\mathcal{B}^{KL}) - C(K,\mathcal{B}^\alpha)}{C(K,\mathcal{B}^{KL})}.$$

This covariance estimator is therefore consistent if there is a probability converging to 1 that

$$\frac{C(K,\mathcal{B}^{KL}) - C(K,\mathcal{B}^\alpha)}{C(K,\mathcal{B}^{KL})} \to 0 \text{ when } N \to +\infty \tag{10.189}$$

and

$$\frac{\|\tilde{P}^{\alpha} - p^{\alpha}\|^2}{\|p^{\alpha}\|^2} \to 0 \text{ when } N \to +\infty . \tag{10.190}$$

This means that the estimated best basis tends to the Karhunen-Loève basis and the estimated diagonal coefficients converge to the power spectrum. The next section establishes such a result for locally stationary processes in a dictionary of local cosine bases.

Generalized Basis Search The quadratic cost $C(K, \mathcal{B})$ defined in (10.183) yields a positive pseudo-distance between any $\mathcal{B}$ and $\mathcal{B}^{KL}$:

$$d(\mathcal{B}, \mathcal{B}^{KL}) = C(K, \mathcal{B}) - C(K, \mathcal{B}^{KL}), \tag{10.191}$$

which is zero if and only if $\mathcal{B}$ is a Karhunen-Loève basis. The following theorem proves that any Schur concave cost function satisfies this property.

Theorem 10.15 *Let K be a covariance operator and $\mathcal{B} = \{g_m\}_{0 \le m < N}$ be an orthonormal basis. If $\Phi(x)$ is strictly concave then*

$$C(K, \mathcal{B}) = \sum_{m=0}^{N-1} \Phi(\langle K g_m, g_m \rangle)$$

is minimum if and only if K is diagonal in $\mathcal{B}$.

Proof[3]. Let $\{h_m\}_{0 \le m < N}$ be a Karhunen-Loève basis that diagonalizes K. As in (9.18), by decomposing g_m in the basis $\{h_i\}_{0 \le i < N}$ we obtain

$$\langle K g_m, g_m \rangle = \sum_{i=0}^{N-1} |\langle g_m, h_i \rangle|^2 \langle K h_i, h_i \rangle. \tag{10.192}$$

Since $\sum_{i=0}^{N-1} |\langle g_m, h_i \rangle|^2 = 1$, applying the Jensen inequality (A.2) to the concave function $\Phi(x)$ proves that

$$\Phi(\langle K g_m, g_m \rangle) \ge \sum_{i=0}^{N-1} |\langle g_m, h_i \rangle|^2 \Phi(\langle K h_i, h_i \rangle). \tag{10.193}$$

Hence

$$\sum_{m=0}^{N-1} \Phi(\langle K g_m, g_m \rangle) \ge \sum_{m=0}^{N-1} \sum_{i=0}^{N-1} |\langle g_m, h_i \rangle|^2 \Phi(\langle K h_i, h_i \rangle).$$

Since $\sum_{m=0}^{N-1} |\langle g_m, h_i \rangle|^2 = 1$, we derive that

$$\sum_{m=0}^{N-1} \Phi(\langle K g_m, g_m \rangle) \ge \sum_{i=0}^{N-1} \Phi(\langle K h_i, h_i \rangle).$$

This inequality is an equality if and only if for all m (10.193) is an equality. Since $\Phi(x)$ is strictly concave, this is possible only if all values $\langle K h_i, h_i \rangle$ are equal as long as $\langle g_m, h_i \rangle \neq 0$. We thus derive that g_m belongs to an eigenspace of K and is thus also an eigenvector of K. Hence, $\{g_m\}_{0 \le m < N}$ diagonalizes K as well. ■

The pseudo-distance (10.191) is mathematically not a true distance since it does not satisfy the triangle inequality. The choice of a particular cost depends on the evaluation of the error when estimating the covariance K. If $\Phi(x) = -x^2$, then minimizing the pseudo-distance (10.191) is equivalent to minimizing the Hilbert-Schmidt norm of the estimation error (10.184). Other costs minimize other error measurements, whose properties are often more complex. The cost associated to $\Phi(x) = -\log_e x$ can be related to the Kullback-Liebler discriminant information [173]. The entropy $\Phi(x) = -x \log_e x$ has been used in image processing to search for approximate Karhunen-Loève bases for face recognition [76].

10.6.3 Locally Stationary Processes [3]

Locally stationary processes appear in many physical systems, where random fluctuations are produced by a mechanism that changes slowly in time or which has few abrupt transitions. Such processes can be approximated locally by stationary processes. Speech signals are locally stationary. Over short time intervals, the throat behaves like a steady resonator that is excited by a stationary source. For a vowel the time of stationarity is about 10^{-1} seconds, but it may be reduced to 10^{-2} seconds for a consonant. The resulting process is therefore locally stationary over time intervals of various sizes.

A *locally stationary process* X is defined qualitatively as a process that is approximately stationary over small enough intervals, and whose values are uncorrelated outside these intervals of stationarity. A number of mathematical characterizations of these processes have been proposed [143, 260, 266, 267, 286].

Donoho, Mallat and von Sachs [172] give an asymptotic definition of local stationarity for a sequence of random vectors having N samples, with N increasing to $+\infty$. The random vector $X_N[n]$ has N samples over an interval normalized to $[0,1]$. Its covariance is $R_N[n,m] = \mathrm{E}\{X_N[n]\,X_N[m]\}$ and we write

$$C_N[n,\tau] = R_N[n,n+\tau]\ .$$

The decorrelation property of locally stationary processes is imposed by a uniform decay condition along τ for all n. There exist Q_1 and $\delta_1 > 1/2$ independent of N such that

$$\forall n\ ,\ \sum_{\tau} (1+2|\tau|^{\delta_1})\,|C_N[n,\tau]|^2 \leq Q_1\ . \tag{10.194}$$

If X_N is stationary, then $C_N[n,\tau] = C_N[\tau]$. A local approximation of X_N with stationary processes is obtained by approximating $C_N[n,\tau]$ over consecutive intervals with functions that depend only on τ. Such approximations are precise if $C_N[n,\tau]$ has slow variations in n in each approximation interval. This occurs when the average total variation of $C_N[n,\tau]$ decreases quickly enough as N increases. Since $X_N[n]$ are samples separated by $1/N$ on $[0,1]$, we suppose that there exist Q_2 and $0 \leq \delta_2 \leq 1$ independent of N such that

$$\forall h\ ,\ \frac{1}{N-h} \sum_{n=0}^{N-1-h} \|C_N[n+h,.] - C_N[n,.]\| \leq Q_2\,\left|hN^{-1}\right|^{\delta_2}\ , \tag{10.195}$$

with

$$\|C_N[n+h,.]-C_N[n,.]\|^2=\sum_{\tau}|C_N[n+h,\tau]-C_N[n,\tau]|^2 \ .$$

Processes that belong to a sequence $\{X_N\}_{N\in\mathbb{N}}$ that satisfies (10.194) and (10.195) are said to be locally stationary.

Example 10.7 Simple locally stationary processes are obtained by blending together a collection of unrelated stationary processes. Let $\{X_{l,N}[n]\}_{1\le l\le I}$ be a collection of mutually independent Gaussian stationary processes whose covariances $R_{l,N}[n,n+\tau]=C_{l,N}[\tau]$ satisfy for $\delta_1>1$

$$\sum_{\tau}(1+2|\tau|^{\delta_1})\,|C_{l,N}[\tau]|^2\le Q_1 \ .$$

Let $\{w_l[n]\}_{1\le l\le I}$ be a family of windows $w_l[n]\ge 0$ with $\sum_{l=1}^{I} w_l[n]\le 1$. Define the blended process

$$X_N[n]=\sum_{l=1}^{I} w_l[n]\,X_{l,N}[n] \ . \tag{10.196}$$

One can then verify [173] that X_N satisfies the local stationarity properties (10.194) and (10.195), with $\delta_2=1$.

If the windows w_l are indicator functions of intervals $[a_l,a_{l+1})$ in $[0,N-1]$, then the blend process has I abrupt transitions. The process X_N remains locally stationary because the number of abrupt transitions does not increase with N. Figure 10.19(a) gives an example.

Best Local Cosine Basis The covariance of a circular stationary process is a circular convolution whose eigenvectors are the Fourier vectors $\exp(i2\pi mn/N)$. Since the eigenvalues are the same at the frequencies $2\pi m/N$ and $-2\pi m/N$, we derive that $\cos(2\pi mn/N+\phi)$ is also an eigenvector for any phase ϕ. A locally stationary process can be locally approximated by stationary processes on appropriate intervals $\{[a_l,a_{l+1})\}_l$ of sizes $b_l=a_{l+1}-a_l$. One can thus expect that its covariance is "almost" diagonalized in a local cosine basis constructed on these intervals of approximate stationarity. Corollary 8.108 constructs orthonormal bases of local cosine vectors over any family of such intervals:

$$\left\{g_l[n]\sqrt{\frac{2}{b_l}}\cos\left[\pi\left(k+\frac{1}{2}\right)\frac{n-a_l}{b_l}\right]\right\}_{0\le k<b_l,\,1\le l\le I} \ . \tag{10.197}$$

Local cosine bases are therefore good candidates for building approximate Karhunen-Loève bases.

When estimating the covariance of a locally stationary process, the position and sizes of the approximate stationarity intervals are generally not known in advance. It is therefore necessary to search for an approximate Karhunen-Loève basis among

a dictionary of local cosine bases, with windows of varying sizes. For this purpose, the best basis search algorithm of Section 10.6.2 is implemented in the dictionary $\mathcal{D}$ of local cosine bases defined in Section 8.5.2. This dictionary is organized as a tree. A family $\mathcal{B}_j^p$ of $N2^{-j}$ orthonormal cosine vectors is stored at depth j and position p. The support of these vectors cover an interval $[a_l, a_l + 2^{-j}N]$ with $a_l = qN2^{-j} - 1/2$:

$$\mathcal{B}_j^q = \left\{ g_{q,k,j}[n] = g_l[n] \sqrt{\frac{2}{2^{-j}N}} \cos\left[\pi\left(k+\frac{1}{2}\right)\frac{n-a_l}{2^{-j}N}\right] \right\}_{0 \le k < N2^{-j}}.$$

The maximum depth is $j \le \log_2 N$, so the dictionary includes fewer than $N \log_2 N$ local cosine vectors. The decomposition of a signal of size N over all these vectors requires $O(N \log_2^2 N)$ operations. The power spectrum estimation from L realizations of a locally stationary process X_N proceeds in four steps:

1. *Sample means* The local cosine coefficients $\langle X_N, g_{q,k,j} \rangle$ of the L realizations are computed. The sample mean estimators $\bar{P}[q,k,j]$ of their variances are calculated with (10.186). This requires $O(LN \log_2^2 N)$ operations.
2. *Regularization* The regularization of $\bar{P}[q,k,j]$ is computed in each family $\mathcal{B}_j^p$ of $2^{-j}N$ cosine vectors corresponding to $0 \le k < 2^{-j}N$. A regularized estimate $\tilde{P}[q,k,j]$ is obtained either with a local averaging along k of $\bar{P}[q,k,j]$, or by thresholding the wavelet coefficients of $\bar{P}[q,k,j]$ in a wavelet basis of size $2^{-j}N$. Over the whole dictionary, this regularization is calculated with $O(N \log_2 N)$ operations.
3. *Basis choice* The cost $\tilde{C}(K, \mathcal{B}^\gamma)$ of each local cosine basis $\mathcal{B}^\gamma$ in (10.187) is an additive function of $|\tilde{P}[q,k,j]|^2$ for the cosine vectors $g_{q,k,j}$ in the basis $\mathcal{B}^\gamma$. The algorithm of Section 9.4.2 finds the best basis $\mathcal{B}^\alpha$ that minimizes this cost with $O(N \log_2 N)$ operations.
4. *Estimation* The local cosine power spectrum is estimated by the coefficients $\tilde{P}[q,k,j]$ for $g_{q,k,j}$ in the best basis $\mathcal{B}^\alpha$.

This best basis algorithm requires $O(LN \log_2^2 N)$ operations to compute a diagonal estimator $\tilde{K}_N^\alpha$ of the covariance K_N. If the regularization of the local cosine coefficients is performed with a wavelet thresholding, using a conservative threshold that is proportional to the maximum eigenvalue of the process, Donoho, Mallat and von Sachs [172] prove that this covariance estimation is consistent for locally stationary processes. As N goes to $+\infty$, the best local cosine basis converges to the Karhunen-Loève basis and the regularized variance estimators converge to the power spectrum. As a result, $\|K_N - \tilde{K}_N^\alpha\|_H$ decreases to 0 with a probability that converges to 1 as N goes to $+\infty$.

Example 10.8 Let X_N be a locally stationary process constructed in (10.196) by aggregating independent Gaussian stationary processes with three windows $w_l[n]$ that are indicator functions of the intervals $[0, 0.2]$, $[0.2, 0.78]$ and $[0.78, 1]$.

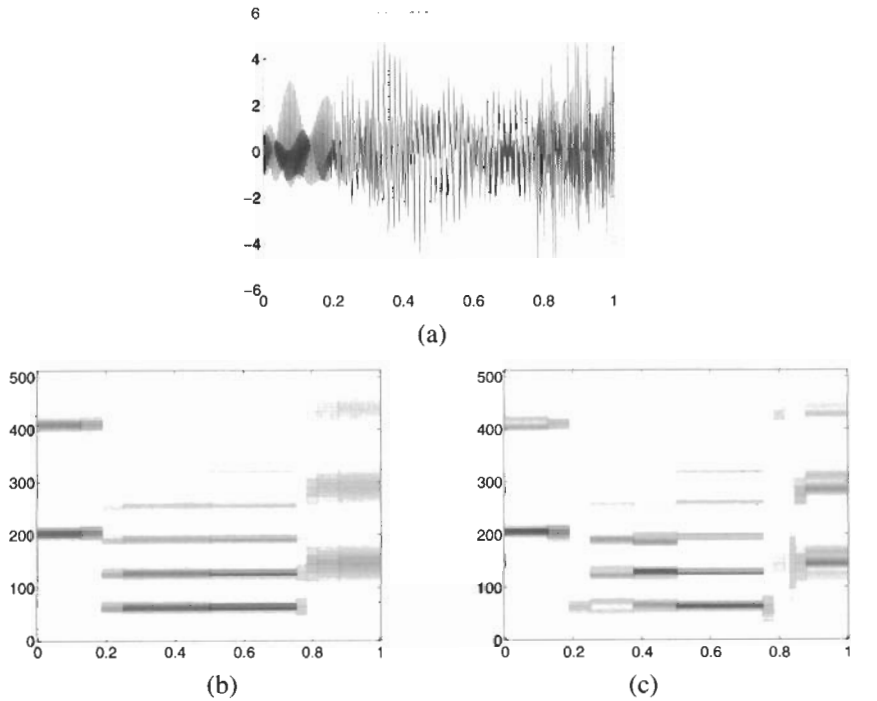


FIGURE 10.19 (a): One realization of a process X_N that is stationary on $[0, 0.2]$, $[0.2, 0.78]$ and $[0.78, 1]$, with N = 1024. (b): Heisenberg boxes of the best local cosine basis computed with $L = 500$ realizations of this locally stationary process. Grey levels are proportional to the estimated spectrum. (c): Best local cosine basis calculated with $L = 3$ realizations.

In each time interval, the power spectrum of the stationary process is composed of harmonics whose amplitude decreases when the frequency increases. Figure 10.19(a) shows one realization of this locally stationary process.

A diagonal covariance is calculated in a best local cosine basis. For a large number $L = 500$ of realizations, the regularized estimator $\tilde{P}[q, k, j]$ gives a precise estimation of the variance $\mathrm{E}\{|\langle X_N, g_{q,k,j}\rangle|^2\}$. The time-frequency tiling of the estimated best basis is shown in Figure 10.19(b). Each rectangle is the Heisenberg box of a local cosine vector $g_{q,k,j}$ of the best basis $\mathcal{B}^{\alpha}$. Its grey level is proportional to $\tilde{P}[q, k, j]$. As expected, short windows are selected in the neighborhood of the transition points at 0.2 and 0.78, and larger windows are selected where the process is stationary. Figure 10.19(c) gives the time-frequency tiling of the best basis computed with only $L = 3$ realizations. The estimators $\tilde{P}[q, k, j]$ are not as precise and the estimated best basis $\mathcal{B}^{\tilde{\alpha}}$ has window sizes that are not optimally adapted to the stationarity intervals of X_N.

10.7 PROBLEMS

10.1. [1] *Linear prediction* Let $F[n]$ be a zero-mean, wide-sense stationary random vector whose covariance is $R_F[k]$. We predict the future $F[n+l]$ from past values $\{F[n-k]\}_{0\leq k<N}$ with $\tilde{F}[n+l]=\sum_{k=0}^{N-1} a_k F[n-k]$.

(a) Prove that $r=\mathrm{E}\{|F[n+l]-\tilde{F}[n+l]|^2\}$ is minimum if and only if

$$\sum_{k=0}^{N-1} a_k R_F[q-k] = R_F[q+l] \quad \text{for } 0\leq q<N.$$

Verify that $r=R_F[0]-\sum_{k=0}^{N-1} a_k R_F[k+l]$ is the resulting minimum error. Hint: use Proposition 10.1.

(b) Suppose that $R_F[n]=\rho^{|n|}$ with $|\rho|<1$. Compute $\tilde{F}[n+l]$ and r.

10.2. [1] Let $X=F+W$ where the signal F and the noise W are zero-mean, wide-sense circular stationary random vectors. Let $\tilde{F}[n]=X\circledast h[n]$ and $r(D,\pi)=\mathrm{E}\{\|F-\tilde{F}\|^2\}$. The minimum risk $r_l(\pi)$ is obtained with the Wiener filter (10.12). A frequency selective filter h has a discrete Fourier transform $\hat{h}[m]$ which can only take the values 0 or 1. Find the frequency selective filter that minimizes $r(D,\pi)$. Prove that $r_l(\pi)\leq r(D,\pi)\leq 2r_l(\pi)$.

10.3. [1] Let $\{g_m\}_{0\leq m<N}$ be an orthonormal basis. We consider the space $\mathbf{V}_p$ of signals generated by the first p vectors $\{g_m\}_{0\leq m<p}$. We want to estimate $f\in\mathbf{V}_p$ from $X=f+W$, where W is a white Gaussian noise of variance σ^2.

(a) Let $\tilde{F}=DX$ be the orthogonal projection of X in $\mathbf{V}_p$. Prove that the resulting risk is minimax:

$$r(D,\mathbf{V}_p)=r_n(\mathbf{V}_p)=p\sigma^2\,.$$

(b) Find the minimax estimator over the space of discrete polynomial signals of size N and degree d. Compute the minimax risk.

10.4. [1] Let $F=f[(n-P) \bmod N]$ be the random shift process (10.15) obtained with a Dirac doublet $f[n]=\delta[n]-\delta[n-1]$. We want to estimate F from $X=F+W$ where W is a Gaussian white noise of variance $\sigma^2=4N^{-1}$.

(a) Specify the Wiener filter $\tilde{F}$ and prove that the resulting risk satisfies $r_l(\pi)=E\{\|F-\tilde{F}\|^2\}\geq 1$.

(b) Show that one can define a thresholding estimator $\tilde{F}$ whose expected risk satisfies

$$E\{\|F-\tilde{F}\|^2\}\leq 12\,(2\log_e N+1)N^{-1}\,.$$

10.5. [1] Let $f=\mathbf{1}_{[0,P-1]}$ be a discrete signal of $N>P$ samples. Let $F=f[(n-P) \bmod N]$ be the random shift process defined in (10.15). We measure $X=F+W$ where W is a Gaussian white noise of variance σ^2.

(a) Suppose that $\tilde{F}=F\circledast h$. Compute the transfer function $\hat{h}[m]$ of the Wiener filter and resulting risk $r_l(\pi)=\mathrm{E}\{\|F-\tilde{F}\|^2\}$.

(b) Let $\tilde{F}$ be the estimator obtained by thresholding the decomposition coefficients of each realization of F in a Haar basis, with $T=\sigma\sqrt{2\log_2 N}$. Prove that $\mathrm{E}\{\|F-\tilde{F}\|^2\}\leq\sigma^2(2\log_e N+1)^2$.

(c) Compare the Wiener and Haar thresholding estimators when N is large.

10.6. [1] Let $|\langle f, g_{m_k}\rangle| \geq |\langle f, g_{m_{k+1}}\rangle|$ for $k \geq 1$ be the sorted decomposition coefficients of f in $\mathcal{B} = \{g_m\}_{0 \leq m < N}$. We want to estimate f from $X = f + W$, where W is a Gaussian white noise of variance σ^2. If $|\langle f, g_{m_k}\rangle| = 2^{-k/2}$, compute the oracle projection risk r_p in (10.34) as a function of σ^2 and N. Give an upper bound on the estimation error ϵ if we threshold at $T = \sigma\sqrt{2\log_e N}$ the decomposition coefficients of X. Same question if $|\langle f, g_{m_k}\rangle| = k^{-1}$. Explain why the estimation is more precise in one case than in the other.

10.7. [1] Compare the SNR and the visual quality of translation invariant hard and soft thresholding estimators in a wavelet basis, for images contaminated by an additive Gaussian white noise. Perform numerical experiments on the Lena, Barbara and Peppers images in WAVELAB. Find the best threshold values T as a function of the noise variance. How does the choice of wavelet (support, number of vanishing moments, symmetry) affect the result?

10.8. [2] Let $g(t)$ be a Gaussian of variance 1. Let $g_s[n] = K_s\, g(n/s)$ where K_s is adjusted so that $\sum_n \rho_s[n] = 1$. An adaptive smoothing of $X = f + W$ is calculated by adapting the scale s as a function of the abscissa:

$$\tilde{F}[l] = \sum_{n=0}^{N-1} X[n]\, g_{s(l)}[l-n]\,. \tag{10.198}$$

The scale $s(l)$ should be large where the signal f seems to be regular, whereas it should be small if we guess that f may have a sharp transition in the neighborhood of l.

(a) Find an algorithm that adapts $s(l)$ depending on the noisy data $X[n]$, and implement the adaptive smoothing (10.198). Test your algorithm on the Piece-Polynomial and Piece-Regular signals in WAVELAB, as a function of the noise variance σ^2.

(b) Compare your numerical results with a translation invariant hard wavelet thresholding. Analyze the similarities between your algorithm that computes $s(l)$ and the strategy used by the wavelet thresholding to smooth or not to smooth certain parts of the noisy signal.

10.9. [3] Let $r_t(f, T)$ be the risk of an estimator of f obtained by hard thresholding with a threshold T the decomposition coefficient of $X = f + W$ in a basis $\mathcal{B}$. The noise W is Gaussian white with a variance σ^2. This risk is estimated by

$$\tilde{r}_t(f, T) = \sum_{m=0}^{N-1} \Phi(|X_B[m]|^2)$$

with

$$\Phi(u) = \begin{cases} u - \sigma^2 & \text{if } u \leq T^2 \\ \sigma^2 & \text{if } u > T^2 \end{cases}.$$

(a) Justify intuitively the definition of this estimator as was done for (10.59) in the case of a soft thresholding estimator.

(b) Let $\phi_\sigma(x) = (2\pi\sigma^2)^{-1/2} \exp(-x^2/(2\sigma^2))$. With calculations similar to the proof of Theorem 10.5, show that

$$r_t(T) - E\{\tilde{r}_t(T)\} = 2T\sigma^2 \sum_{m=0}^{N-1} \Big[\phi_\sigma(T - f_B[m]) + \phi_\sigma(T + f_B[m])\Big].$$

(c) Implement in MATLAB an algorithm in $O(N \log_2 N)$ which finds $\tilde{T}$ that minimizes $\tilde{r}_t(T, f)$. Study numerically the performance of $\tilde{T}$ to estimate noisy signals with a hard thresholding in a wavelet basis.

10.10. [1] Let $\mathcal{B}$ be an orthonormal wavelet basis of the space of discrete signals of period N. Let $\mathcal{D}$ be the family that regroups all translations of wavelets in $\mathcal{B}$.

(a) Prove that $\mathcal{D}$ is a dyadic wavelet frame for signals of period N.

(b) Show that an estimation by thresholding decomposition coefficients in the dyadic wavelet family $\mathcal{D}$ implements a translation invariant thresholding estimation in the basis $\mathcal{B}$.

10.11. [3] A translation invariant wavelet thresholding is equivalent to thresholding a wavelet frame that is not subsampled. For images, elaborate and implement an algorithm that performs a spatial averaging of the wavelet coefficients above the threshold, by using the geometrical information provided by multiscale edges. Coefficients should not be averaged across edges.

10.12. [2] Let $X = f + W$ where f is piecewise regular. A best basis thresholding estimator is calculated with the cost function (10.74) in a wavelet packet dictionary. Compare numerically the results with a simpler wavelet thresholding estimator, on the Piece-Polynomial and Piece-Regular signals in WAVELAB. Find a signal f for which a best wavelet packet thresholding yields a smaller estimation error than a wavelet thresholding.

10.13. [2] Among signals $f[n]$ of size N we consider $\Theta_V = \{f \ : \ \|f\|_V \leq C\}$. Let $X = f + W$ where W is a Gaussian white noise of variance σ^2. We define a linear estimator $DX[n] = X \star h[n]$ with

$$\hat{h}[m] = \frac{C^2}{C^2 + 4\sigma^2 N |\sin(\pi m/N)|^2} \ . \tag{10.199}$$

Prove that the maximum risk of this estimator is close to the minimax linear risk:

$$r_l(\Theta_V) \leq r(D, \Theta_V) \leq 2\, r_l(\Theta_V) \ .$$

Hint: follow the approach of the proof of Proposition 10.5.

10.14. [2] We want to estimate a signal f that belongs to an ellipsoid

$$\Theta = \left\{ f \ : \ \sum_{m=0}^{N-1} \beta_m^2 |f_B[m]|^2 \leq C^2 \right\}$$

from $X = f + W$, where W is a Gaussian white noise of variance σ^2. We denote $x_+ = \max(x, 0)$.

(a) Using Proposition 10.6 prove that the minimax linear risk on Θ satisfies

$$r_l(\Theta) = \sigma^2 \sum_{m=0}^{N-1} a[m] \tag{10.200}$$

with $a[m] = (\frac{\lambda}{\beta_m} - 1)_+$, where λ is a Lagrange multiplier calculated with

$$\sum_{m=0}^{N-1} \beta_m (\frac{\lambda}{\beta_m} - 1)_+ = \frac{C^2}{\sigma^2}. \tag{10.201}$$

(b) By analogy to Sobolev spaces, the Θ of signals having a discrete derivative of order s whose energy is bounded by C^2 is defined from the discrete Fourier transform:

$$\Theta = \{f : \sum_{m=-N/2+1}^{N/2} |m|^{2s} N^{-1} |\hat{f}[m]|^2 \leq C^2\}. \tag{10.202}$$

Show that the minimax linear estimator D in Θ is a circular convolution $DX = X \circledast h$. Explain how to compute the transfer function $\hat{h}[m]$.

(c) Show that the minimax linear risk satisfies

$$r_l(\Theta) \sim C^{2/(2s+1)} \sigma^{2-2/(2s+1)}.$$

10.15. [3] Let $\Theta_V = \{f : \|f\|_V \leq C\}$ be a set of bounded variation discrete images of N^2 pixels. Prove that for signals in Θ_V contaminated by an additive Gaussian white noise of variance σ^2, if $1 \leq C/\sigma \leq N$ then the linear minimax risk satisfies $r_l(\Theta_V) \sim N^2 \sigma^2$. Hint: compute a lower bound by finding an appropriate subset of Θ_V which is orthosymmetric in a wavelet basis.

10.16. [2] We want to estimate $f \in \Theta$ from $Y = f \circledast u + W$ where W is a white noise of variance σ^2. Suppose that Θ is closed and bounded. We consider the quadratic convex hull $\mathrm{QH}[\Theta]$ in the discrete Fourier basis and $x \in \mathrm{QH}[\Theta]$ such that $r(x) = r_{\inf}(\mathrm{QH}[\Theta])$. Prove that the linear estimator that achieves the minimax linear risk $r_l(\Theta)$ in Theorem 10.13 is $\tilde{F} = Y \circledast h$ with

$$\hat{h}[m] = \frac{N^{-1} |\hat{x}[m]|^2 \hat{u}^*[m]}{\sigma^2 + N^{-1} |\hat{x}[m]|^2 |\hat{u}[m]|^2}.$$

Hint: use the diagonal estimator in Proposition 10.11.

10.17. [2] Implement in WAVELAB the algorithm of Section 10.5.1 that extracts coherent structures with a pursuit of bases. Use a dictionary that is a union of a wavelet packet and a local cosine dictionary. Apply this algorithm to the restoration of the Caruso signal in WAVELAB. Find stopping rules to improve the auditory quality of the restored signal [92].

10.18. [1] *Stationary spectrum estimation* Let $X[n]$ be a zero-mean, infinite size process that is wide-sense stationary. The power spectrum $\hat{R}_X(\omega)$ is the Fourier transform of the covariance $R_X[p] = \mathrm{E}\{X[n]X[n-p]\}$. Let $\tilde{R}_X[p] = \frac{1}{N} \sum_{n=0}^{N-1-|p|} X[n]X[n+|p|]$ be an estimation of $R_X[k]$ from a single realization of $X[n]$.

(a) Show that $\mathrm{E}\{\tilde{R}_X[p]\} = \frac{N-|p|}{N} R_X[p]$ for $|p| \leq N$.

(b) Verify that the discrete Fourier transform of $\tilde{R}_X[p]$ is the periodogram $\bar{P}[m]$ in (10.181).

(c) Let $\hat{h}(\omega) = \frac{1}{N} \left(\frac{\sin(N\omega/2)}{\sin(\omega/2)} \right)^2$. Prove that

$$\mathrm{E}\{\bar{P}[m]\} = \frac{1}{2\pi} \hat{R}_X \star \hat{h}\left(\frac{2\pi m}{N}\right) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \hat{R}_X(\omega) \hat{h}\left(\frac{2\pi m}{N} - \omega\right) d\omega.$$

(d) Let $g[n]$ be a discrete window whose support is $[0, N-1]$ and let $h(\omega) = |\hat{g}(\omega)|^2$. The periodogram of the windowed data is

$$\bar{P}_g[m] = \frac{1}{N}\left|\sum_{n=0}^{N-1} g[n]\, X[n] \exp\left(\frac{-i2\pi mn}{N}\right)\right|^2 . \tag{10.203}$$

Prove that $p[m] = \mathrm{E}\{\bar{P}_g[m]\} = \frac{1}{2\pi}\hat{R}_X \star \hat{h}(\frac{2\pi m}{N})$. How should we design $g[n]$ in order to reduce the bias of this estimator of $\hat{R}_X(\omega)$?

(e) Verify that the variance is: $\mathrm{E}\{|\bar{P}_g[k] - p[k]|^2\} = 2\,|d[k]|^2$. Hint: use Proposition 10.14.

10.19. [1] *Lapped spectrum estimation* Let $X[n]$ be a zero-mean, infinite size process that is Gaussian and wide-sense stationary. Let $\hat{R}_X(\omega)$ be the Fourier series of its covariance $R_X[k]$. We suppose that one realization of $X[n]$ is known on $[-\eta, N+\eta-1]$. To reduce the variance of the spectrogram (10.203), we divide $[0, N-1]$ in Q intervals $[a_q, a_{q+1}]$ of size M, with $a_q = qM - 1/2$ for $0 \leq p < Q$. We denote by $\{g_{q,k}\}_{q,k}$ the discrete local cosine vectors (8.108) constructed with windows g_q having a support $[a_q - \eta, a_{q+1} + \eta]$, with raising and decaying profiles of constant size 2η. Since all windows are identical but translated, $|\hat{g}_q(\omega)|^2 = h(\omega)$.

(a) Let $\bar{P}_q[k] = |\langle X, g_{q,k}\rangle|^2$ and $\tilde{P}[k] = \frac{1}{L}\sum_{l=0}^{L-1} \bar{P}_l[k]$. Verify that

$$p[k] = \mathrm{E}\{\tilde{P}[k]\} = \frac{1}{2\pi}\hat{R}_X \star h\left(\frac{\pi}{M}\left(k + \frac{1}{2}\right)\right).$$

(b) Suppose that $X[n]$ has a correlation length smaller than M so that its values on different intervals $[a_q, a_{q+1}]$ can be considered as independent. Show that $\mathrm{E}\{|\tilde{P}[k] - p[k]|^2\} = 2\,|p[k]|^2 L^{-1}$. Discuss the trade-off between bias and variance in the choice of L.

(c) Implement this power spectrum estimator in WAVELAB.

10.20. [3] *Adaptive spectrum estimation* Problem 10.19 estimates the power spectrum and hence the covariance K of a stationary Gaussian process $X[n]$ with a diagonal operator $\tilde{K}$ in a local cosine basis. The diagonal values of $\tilde{K}$ are the regularized coefficients $\tilde{P}[k] = \frac{1}{L}\sum_{l=0}^{L-1} \bar{P}_l[k]$.

(a) Verify with (10.182) that

$$\mathrm{E}\left\{\|\tilde{K} - K\|_H^2\right\} = L\sum_{k=1}^{M} \mathrm{E}\left\{|\tilde{P}[k] - p[k]|^2\right\} + \|K\|_H^2 - L\sum_{k=1}^{M} |p[k]|^2. \tag{10.204}$$

(b) Find a best basis algorithm that chooses the optimal window size $M = 2^j$ by minimizing an estimator of the error (10.204). Approximate $p[k]$ with $\tilde{P}[k]$ and find a procedure for estimating $\mathrm{E}\{|\tilde{P}[k] - p[k]|^2\}$ from the data values $\{\bar{P}_l[k]\}_{0 \leq l < L}$. Remember that when they are independent $\mathrm{E}\{|\tilde{P}[k] - p[k]|^2\} = 2\,|p[k]|^2 L$.

XI

TRANSFORM CODING

Reducing a liter of orange juice to a few grams of concentrated powder is what lossy compression is about. The taste of the restored beverage is similar to the taste of orange juice but has often lost some subtlety. We are more interested in sounds and images, but we face the same trade-off between quality and compression. Major applications are data storage and transmission through channels with a limited bandwidth.

A transform coder decomposes a signal in an orthogonal basis and quantizes the decomposition coefficients. The distortion of the restored signal is minimized by optimizing the quantization, the basis and the bit allocation. The basic information theory necessary for understanding quantization properties is introduced. Distortion rate theory is first studied in a Bayes framework, where signals are realizations of a random vector whose probability distribution is known a priori. This applies to audio coding, where signals are often modeled with Gaussian processes.

Since no appropriate stochastic model exists for images, a minimax approach is used to compute the distortion rate of transform coding. Image compression algorithms in wavelet bases and cosine block bases are described. These transform codes are improved by embedding strategies that first provide a coarse image approximation, then progressively refine the quality by adding more bits. The compression of video sequences with motion compensation and transform coding is also explained.

11.1 SIGNAL COMPRESSION [2]

11.1.1 State of the Art

Speech Speech coding is used for telephony, where it may be of limited quality but good intelligibility, and for higher quality teleconferencing. Telephone speech is limited to the frequency band 200-3400 Hz and is sampled at 8 kHz. A Pulse Code Modulation (PCM) that quantizes each sample on 8 bits produces a code with 64 kb/s ($64\,10^3$ bits per second). This can be considerably reduced by removing some of the speech redundancy.

The production of speech signals is well understood. Model based analysis-synthesis codes give intelligible speech at 2.4 kb/s. This is widely used for defense telecommunications [225, 333]. Digital cellular telephony uses 8 kb/s to reproduce more natural voices. Linear Predictive Codes (LPC) restore speech signals by filtering white noise or a pulse train with linear filters whose parameters are estimated and coded. For higher bit rates, the quality of LPC speech production is enhanced by exciting the linear filters with waveforms chosen from a larger family. These Code-Excited Linear Prediction (CELP) codes provide nearly perfect telephone quality at 16 kb/s.

Audio Audio signals include speech but also music and all types of sounds. On a compact disk, the audio signal is limited to a maximum frequency of 20 kHz. It is sampled at 44.1 kHz and each sample is coded on 16 bits. The bit rate of the resulting PCM code is 706 kb/s. For compact disks and digital audio tapes, signals must be coded with hardly any noticeable distortion. This is also true for multimedia CD-ROM and digital television sounds.

No models are available for general audio signals. At present, the best compression is achieved by transform coders that decompose the signal in a local time-frequency basis. To reduce perceived distortion, perceptual coders [226] adapt the quantization of time-frequency coefficients to our hearing sensitivity. Compact disk quality sounds are restored with 128 kb/s. Nearly perfect audio signals are obtained with 64 kb/s.

Images A grey-level image typically has 512 by 512 pixels, each coded with 8 bits. Like audio signals, images include many types of structures that are difficult to model. Currently, the best image compression algorithms are also transform codes, with cosine bases or wavelet bases. The efficiency of these bases comes from their ability to construct precise non-linear image approximations with few non-zero vectors. With fewer than 1 bit/pixel, visually perfect images are reconstructed. At 0.25 bit/pixel, the image remains of good quality.

Video Applications of digital video range from low quality videophones and teleconferencing to high resolution television. The most effective compression algorithms remove the time redundancy with a motion compensation. Local image displacements are measured from one frame to the next, and are coded as

motion vectors. Each frame is predicted from the previous one by compensating for the motion. An error image is calculated and compressed with a transform code. The MPEG standards described in Section 11.5.2 are based on this motion compensation [248].

For teleconferencing, color images have only 360 by 288 pixels. A maximum of 30 images per second are transmitted, but more often 10 or 15. If the images do not include too much motion, a decent quality video is obtained at 128kb/s, which can be transmitted in real time through a digital telephone line.

The High Definition Television (HDTV) format has color images of 1280 by 720 pixels, and 60 images per second. The resulting bit rate is on the order of 10^3 Mb/s. To transmit the HDTV through channels used by current television technology, the challenge is to reduce the bit rate to 20 Mb/s, without any loss of quality.

11.1.2 Compression in Orthonormal Bases

A transform coder decomposes signals in an orthonormal basis $\mathcal{B} = \{g_m\}_{0 \leq m < N}$ and optimizes the compression of the decomposition coefficients. The performance of such a transform code is studied from a Bayes point of view, by supposing that the signal is the realization of a random process $F[n]$ of size N, whose probability distribution is known a priori.

Let us decompose F over $\mathcal{B}$:

$$F = \sum_{m=0}^{N-1} F_{\mathcal{B}}[m]\, g_m .$$

Each coefficient $F_{\mathcal{B}}[m]$ is a random variable defined by

$$F_{\mathcal{B}}[m] = \langle F, g_m \rangle = \sum_{n=0}^{N-1} F[n]\, g_m^*[n].$$

To center the variations of $F_{\mathcal{B}}[m]$ at zero, we code $F_{\mathcal{B}}[m] - \mathsf{E}\{F_{\mathcal{B}}[m]\}$ and store the mean value $\mathsf{E}\{F_{\mathcal{B}}[m]\}$. This is equivalent to supposing that $F_{\mathcal{B}}[m]$ has a zero mean.

Quantization To construct a finite code, each coefficient $F_{\mathcal{B}}[m]$ is approximated by a quantized variable $\tilde{F}_{\mathcal{B}}[m]$, which takes its values over a finite set of real numbers. A scalar quantization approximates each $F_{\mathcal{B}}[m]$ independently. If the coefficients $F_{\mathcal{B}}[m]$ are highly dependent, quantizer performance is improved by vector quantizers that approximate together the vector of N coefficients $\{F_{\mathcal{B}}[m]\}_{0 \leq m < N}$ [27]. Scalar quantizers require fewer computations and are thus more often used. If the basis $\{g_m\}_{0 \leq m < N}$ can be chosen so that the coefficients $F_{\mathcal{B}}[m]$ are nearly independent, the improvement of a vector quantizer becomes marginal. After

quantization, the reconstructed signal is

$$\tilde{F} = \sum_{m=0}^{N-1} \tilde{F}_B[m]\, g_m.$$

Distortion Rate A major issue is to evaluate the distortion introduced by this quantization. Ultimately, we want to restore a signal that is perceived as nearly identical to the original signal. Perceptual transform codes are optimized with respect to our sensitivity to degradations in audio signals and images [226]. However, distances that evaluate perceptual errors are highly non-linear and thus difficult to manipulate mathematically. A mean-square norm often does not properly quantify the perceived distortion, but reducing a mean-square distortion generally enhances the coder performance. Weighted mean-square distances can provide better measurements of perceived errors and are optimized like a standard mean-square norm.

In the following, we try to minimize the average coding distortion, evaluated with a mean-square norm. Since the basis is orthogonal, this distortion can be written

$$d = \mathsf{E}\{\|F - \tilde{F}\|^2\} = \sum_{m=0}^{N-1} \mathsf{E}\{|F_B[m] - \tilde{F}_B[m]|^2\}.$$

The average number of bits allocated to encode a quantized coefficient $\tilde{F}_B[m]$ is denoted R_m. For a given R_m, a scalar quantizer is designed to minimize $\mathsf{E}\{|F_B[m] - \tilde{F}_B[m]|^2\}$. The total mean-square distortion d depends on the average total bit budget

$$R = \sum_{m=0}^{N-1} R_m.$$

The function $d(R)$ is called the *distortion rate*. For a given R, the bit allocation $\{R_m\}_{0 \le m < N}$ must be adjusted in order to minimize $d(R)$.

Choice of Basis The distortion rate of an optimized transform code depends on the orthonormal basis $\mathcal{B}$. We see in Section 11.3.2 that the Karhunen-Loève basis minimizes $d(R)$ for high resolution quantizations of signals that are realizations of a Gaussian process. This is not true when the process is non-Gaussian.

To achieve a high compression rate, the transform code must produce many zero quantized coefficients whose positions are efficiently recorded. Section 11.4 shows that $d(R)$ then depends on the precision of non-linear approximations in the basis $\mathcal{B}$.

11.2 DISTORTION RATE OF QUANTIZATION [2]

Quantized coefficients take their values over a finite set and can thus be coded with a finite number of bits. Section 11.2.1 reviews entropy codes of random sources. Section 11.2.2 studies the optimization of scalar quantizers in order to reduce the mean-square error for a given bit allocation.

11.2.1 Entropy Coding

Let X be a random source that takes its values among a finite alphabet of K symbols $\mathcal{A} = \{x_k\}_{1 \leq k \leq K}$. The goal is to minimize the average bit rate needed to store the values of X. We consider codes that associate to each symbol x_k a binary word w_k of length l_k. A sequence of values produced by the source X is coded by aggregating the corresponding binary words.

All symbols x_k can be coded with binary words of the same size $l_k = \lceil \log_2 K \rceil$ bits. However, the average code length may be reduced with a *variable length code* using smaller binary words for symbols that occur frequently. Let us denote by p_k the probability of occurrence of a symbol x_k:

$$p_k = \Pr\{X = x_k\}.$$

The average bit rate to code each symbol emitted by the source X is

$$R_X = \sum_{k=1}^{K} l_k \, p_k. \tag{11.1}$$

We want to optimize the code words $\{w_k\}_{1 \leq k \leq K}$ in order to minimize R_X.

Prefix Code Codes with words of varying lengths are not always uniquely decodable. Let us consider the code that associates to $\{x_k\}_{1 \leq k \leq 4}$ the code words

$$\{w_1 = 0\ ,\ w_2 = 10\ ,\ w_3 = 110\ ,\ w_4 = 101\}. \tag{11.2}$$

The sequence 1010 can either correspond to $w_2 w_2$ or to $w_4 w_1$. To guarantee that any aggregation of code words is uniquely decodable, the *prefix* condition imposes that no code word may be the prefix (beginning) of another one. The code (11.2) does not satisfy this condition since w_2 is the prefix of w_4. The following code

$$\{w_1 = 0\ ,\ w_2 = 10\ ,\ w_3 = 110\ ,\ w_4 = 111\}$$

satisfies this prefix condition. Any code that satisfies the prefix condition is clearly uniquely decodable.

A prefix code is characterized by a binary tree that has K leaves corresponding to the symbols $\{x_k\}_{1 \leq k \leq K}$. Figure 11.1 shows an example for a prefix code of $K = 6$ symbols. The left and right branches of the binary tree are respectively coded by 0 and 1. The binary code word w_k associated to x_k is the succession of 0 and 1 corresponding to the left and right branches along the path from the root to the leaf x_k. The binary code produced by such a binary tree is always a prefix code. Indeed, w_m is a prefix of w_k if and only if x_m is an ancestor of x_k in the binary tree. This is not possible since both symbols correspond to a leaf of the tree. Conversely, we can verify that any prefix code can be represented by such a binary tree.

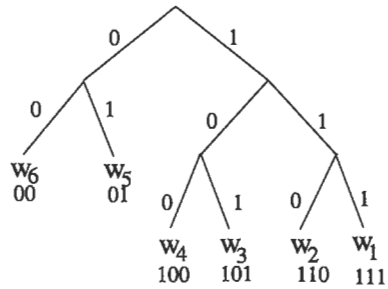


FIGURE 11.1 Prefix tree corresponding to a code with six symbols. The code word w_k of each leaf is indicated below it.

The length l_k of the code word w_k is the depth in the binary tree of the corresponding leaf. The optimization of a prefix code is thus equivalent to the construction of an optimal binary tree that distributes the depth of the leaves in order to minimize

$$R_X = \sum_{k=1}^{K} l_k \, p_k. \tag{11.3}$$

Higher probability symbols should therefore correspond to leaves higher in the tree.

Shannon Entropy The Shannon theorem [306] proves that entropy is a lower bound for the average bit rate R_X of any prefix code.

Theorem 11.1 (SHANNON) *Let X be a source whose symbols $\{x_k\}_{1 \leq k \leq K}$ occur with probabilities $\{p_k\}_{1 \leq k \leq K}$. The average bit rate R_X of a prefix code satisfies*

$$R_X \geq \mathcal{H}(X) = -\sum_{k=1}^{K} p_k \log_2 p_k. \tag{11.4}$$

Moreover, there exists a prefix code such that

$$R_X \leq \mathcal{H}(X) + 1. \tag{11.5}$$

The sum $\mathcal{H}(X)$ is called the entropy of X.

Proof[2]. This theorem is based on the Kraft inequality given by the following lemma.

Lemma 11.1 (KRAFT) *Any prefix code satisfies*

$$\sum_{k=1}^{K} 2^{-l_k} \leq 1. \tag{11.6}$$

Conversely, if $\{l_k\}_{1 \leq k \leq K}$ is a positive sequence that satisfies (11.6), then there exists a sequence of binary words $\{w_k\}_{1 \leq k \leq K}$ of length $\{l_k\}_{1 \leq k \leq K}$ that satisfies the prefix condition.

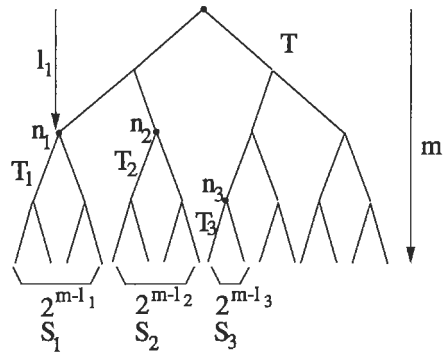


FIGURE 11.2 The leaves at the depth m of the tree T are regrouped as sets S_k of 2^{m-l_k} nodes that are the leaves of a tree T_k whose root n_k is at the depth l_k. Here $m = 4$ and $l_1 = 2$ so S_1 has 2^2 nodes.

To prove (11.6), we construct a full binary tree T whose leaves are at the depth $m = \max\{l_1, l_2, \ldots, l_K\}$. Inside this tree, we can locate the node n_k at the depth l_k that codes the binary word w_k. We denote T_k the subtree whose root is n_k, as illustrated in Figure 11.2. This subtree has a depth $m - l_k$ and thus contains 2^{m-l_k} nodes at the level m of T. There are 2^m nodes at the depth m of T and the prefix condition implies that the subtrees $T_1, \ldots, T_K$ have no node in common, so

$$\sum_{k=1}^{K} 2^{m-l_k} \leq 2^m,$$

which proves (11.6).

Conversely, we consider $\{l_k\}_{1 \leq k \leq K}$ that satisfies (11.6), with $l_1 \leq l_2 \leq \cdots \leq l_K$ and $m = \max\{l_1, l_2, \ldots, l_K\}$. Again we construct a full binary tree T whose leaves are at the depth m. Let S_1 be the 2^{m-l_1} first nodes at the level m , and S_2 be the next 2^{m-l_2} nodes, and so on, as illustrated by Figure 11.2. Since $\sum_{k=1}^{K} 2^{m-l_k} \leq 2^m$, the sets $\{S_k\}_{1 \leq k \leq K}$ have fewer than 2^m elements and can thus be constructed at the level m of the tree. The nodes of a set S_k are the leaves of a subtree T_k of T. The root n_k of T_k is at the depth l_k and corresponds to a binary word w_k. By construction, all these subtrees T_k are distinct, so $\{w_k\}_{1 \leq k \leq K}$ is a prefix code where each code word w_k has a length l_k. This finishes the lemma proof.

To prove the two inequalities (11.4) and (11.5) of the theorem, we consider the minimization of

$$R_X = \sum_{k=1}^{K} p_k l_k$$

under the Kraft inequality constraint

$$\sum_{k=1}^{K} 2^{-l_k} \leq 1.$$

If we admit non-integer values for l_k, we can verify with Lagrange multipliers that the minimum is reached for $l_k = -\log_2 p_k$. The value of this minimum is the entropy lower bound:

$$R_X = \sum_{k=1}^{K} p_k l_k = -\sum_{k=1}^{K} p_k \log_2 p_k = \mathcal{H}(X),$$

which proves (11.4).

To guarantee that l_k is an integer, the Shannon code is defined by

$$l_k = \lceil -\log_2 p_k \rceil, \tag{11.7}$$

where $\lceil x \rceil$ is the smallest integer larger than x. Since $l_k \geq -\log_2 p_k$, the Kraft inequality is satisfied:

$$\sum_{k=1}^{K} 2^{-l_k} \leq \sum_{k=1}^{K} 2^{\log_2 p_k} = 1.$$

Lemma 11.1 proves that there exists a prefix code whose binary words w_k have length w_k. For this code,

$$R_X = \sum_{k=1}^{K} p_k l_k \leq \sum_{k=1}^{K} p_k(-\log_2 p_k + 1) = \mathcal{H}(X) + 1,$$

which proves (11.5). ■

The entropy $\mathcal{H}(X)$ measures the uncertainty as to the outcome of the random variable X. As in (9.69), we prove that

$$0 \leq \mathcal{H}(X) \leq \log_2 K.$$

The maximum value $\log_2 K$ corresponds to a sequence with a uniform probability distribution $p_k = 1/K$, for $1 \leq k \leq K$. Since no value is more probable than any other, the uncertainty as to the outcome of X is maximum. The minimum entropy value $\mathcal{H}(X) = 0$ corresponds to a source where one symbol x_k occurs with probability 1. There is no uncertainty as to the outcome of X because we know in advance that it will be equal to x_k.

Huffman Code The lower entropy bound $\mathcal{H}(X)$ is nearly reachable with an optimized prefix code. The *Huffman algorithm* is a dynamical programming algorithm that constructs a binary tree that minimizes the average bit rate $R_X = \sum_{k=1}^{K} p_k l_k$. This tree is called an optimal prefix code tree. The following proposition gives an induction rule that constructs the tree from bottom up by aggregating lower probability symbols.

Proposition 11.1 (HUFFMAN) *Let us consider K symbols with their probability of occurrence sorted in increasing order $p_k \leq p_{k+1}$:*

$$\{(x_1, p_1), (x_2, p_2), (x_3, p_3), \ldots, (x_K, p_K)\}. \tag{11.8}$$

We aggregate the two lower probability symbols x_1 and x_2 in a single symbol $x_{1,2}$ of probability

$$p_{1,2} = p_1 + p_2.$$

An optimal prefix tree for the K symbols (11.8) is obtained by constructing an optimal prefix tree for the $K-1$ symbols

$$\{(x_{1,2}, p_{1,2}), (x_3, p_3), \ldots, (x_K, p_K)\}, \tag{11.9}$$

and by dividing the leaf $x_{1,2}$ into two children nodes corresponding to x_1 and x_2.

The proof of this proposition [27, 216] is left to the reader. The Huffman rule reduces the construction of an optimal prefix code of K symbols (11.8) to the construction of an optimal code of $K-1$ symbols (11.9) plus an elementary operation. The Huffman algorithm iterates this regrouping $K-1$ times to grow a prefix code tree progressively from bottom to top. The Shannon Theorem 11.1 proves that the average bit rate of the optimal Huffman prefix code satisfies

$$\mathcal{H}(X) \leq R_X \leq \mathcal{H}(X) + 1 . \tag{11.10}$$

As explained in the proof of Theorem 11.1, the bit rate may be up to one bit more than the entropy lower bound because this lower bound is obtained with $l_k = -\log_2 p_k$, which is generally not possible since l_k must be an integer. In particular, lower bit rates are achieved when one symbol has a probability close to 1.

Example 11.1 We construct the Huffman code of six symbols $\{x_k\}_{1 \leq k \leq 6}$ whose probabilities are

$$\{p_1 = 0.05\ ,\ p_2 = 0.1\ ,\ p_3 = 0.1\ ,\ p_4 = 0.15\ ,\ p_5 = 0.2\ ,\ p_6 = 0.4\}.$$

The symbols x_1 and x_2 are the lower probability symbols, which are regrouped in a symbol $x_{1,2}$ whose probability is $p_{1,2} = p_1 + p_2 = 0.15$. At the next iteration, the lower probabilities are $p_3 = 0.1$ and $p_{1,2} = 0.15$, so we regroup $x_{1,2}$ and x_3 in a symbol $x_{1,2,3}$ whose probability is 0.25. The next two lower probability symbols are x_4 and x_5, which are regrouped in a symbol $x_{4,5}$ of probability 0.35. We then group $x_{4,5}$ and $x_{1,2,3}$ which yields $x_{1,2,3,4,5}$ of probability 0.6, which is finally aggregated with x_6. This finishes the tree, as illustrated in Figure 11.3. The resulting average bit rate (11.3) is $R_X = 2.35$ whereas the entropy is $\mathcal{H}(X) = 2.28$. This Huffman code is better than the prefix code of Figure 11.1, whose average bit rate is $R_X = 2.4$.

Block coding As mentioned above, the inequality (11.10) shows that a Huffman code may require one bit above the entropy because the length l_k of each binary word must be an integer, whereas the optimal value $-\log_2 p_k$ is generally a real

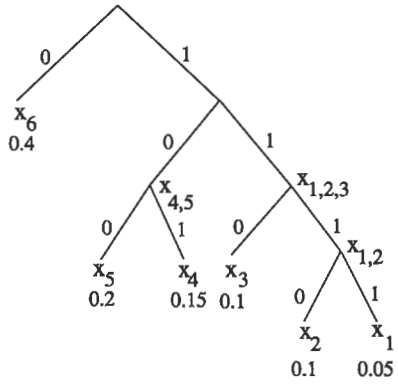


FIGURE 11.3 Prefix tree grown with the Huffman algorithm for a set of $K = 6$ symbols x_k whose probabilities p_k are indicated at the leaves of the tree.

number. To reduce this overhead the symbols are coded together in blocks of size n.

Let us consider the block of n independent random variables $\vec{X} = X_1, \ldots, X_n$, where each X_k takes its values in the alphabet $\mathcal{A} = \{x_k\}_{1 \leq k \leq K}$ with the same probability distribution as X. The vector $\vec{X}$ can be considered as a random variable taking its values in the alphabet $\mathcal{A}^n$ of size K^n. To each block of symbols $\vec{s} \in \mathcal{A}^n$ we associate a binary word of length $l(\vec{s})$. The average number of bits per symbol for such a code is

$$R_X = \frac{1}{n} \sum_{\vec{s} \in \mathcal{A}^n} p(\vec{s})\, l(\vec{s}).$$

The following proposition proves that the resulting Huffman code has a bit rate that converges to the entropy of X as n increases.

Proposition 11.2 *The Huffman code for a block of size n requires an average number of bits per symbol that satisfies*

$$\mathcal{H}(X) \leq R_X \leq \mathcal{H}(X) + \frac{1}{n}. \tag{11.11}$$

Proof[2]. The entropy of $\vec{X}$ considered as a random variable is

$$\mathcal{H}(\vec{X}) = \sum_{\vec{s} \in \mathcal{A}^n} p(\vec{s}) \log_2 p(\vec{s}).$$

Denote by $R_{\vec{X}}$ the average number of bits to code each block $\vec{X}$. Applying (11.10) shows that with a Huffman code, $R_{\vec{X}}$ satisfies

$$\mathcal{H}(\vec{X}) \leq \vec{R}_{\vec{X}} \leq \mathcal{H}(\vec{X}) + 1. \tag{11.12}$$

Since the random variables X_i that compose $\vec{X}$ are independent,

$$p(\vec{s}) = p(s_1, \ldots, s_n) = \prod_{i=1}^{n} p(s_i) .$$

We thus derive that $\mathcal{H}(\vec{X}) = n\mathcal{H}(X)$ and since $R = \vec{R}/n$, we obtain (11.11) from (11.12). ■

Coding together the symbols in blocks is equivalent to coding each symbol x_k with an average number of bits l_k that is not an integer. This explains why block coding can nearly reach the entropy lower bound. The Huffman code can also be adaptively modified for long sequences in which the probability of occurrence of the symbols may vary [20]. The probability distribution is computed from the histogram (cumulative distribution) of the N most recent symbols that were decoded. The next N symbols are coded with a new Huffman code calculated from the updated probability distribution. However, recomputing the Huffman code after updating the probability distribution is computationally expensive. Arithmetic codes have a causality structure that makes it easier to adapt the code to a varying probability distribution.

Arithmetic Code Like a block Huffman code, an arithmetic code [294] records in blocks the symbols $\{x_k\}_{1 \leq k \leq K}$ to be coded. However, an arithmetic code is more structured. It constructs progressively the code of a whole block as each symbol is taken into account. When the probability p_k of each symbol x_k is not known, an adaptive arithmetic code progressively learns the probability distribution of the source and adapts the encoding.

We consider a block of symbols $\vec{s} = s_1, s_2, \ldots, s_n$ produced by a random vector $\vec{X} = X_1, .., X_n$ of n independent random variables. Each X_k has the same probability distribution $p(x)$ as the source X, with $p(x_j) = p_j$. An arithmetic code represents each $\vec{s}$ by an interval $[a_n, a_n + b_n]$ included in $[0, 1]$, whose length is equal to the probability of occurrence of this sequence:

$$b_n = \prod_{k=1}^{n} p(s_k) .$$

This interval is defined by induction as follows. We initialize $a_0 = 0$ and $b_0 = 1$. Let $[a_i, a_i + b_i]$ be the interval corresponding to the first i symbols $s_1, \ldots, s_i$. Suppose that the next symbol s_{i+1} is equal to x_j so that $p(s_{i+1}) = p_j$. The new interval $[a_{i+1}, a_{i+1} + b_{i+1}]$ is a sub-interval of $[a_i, a_i + b_i]$ whose size is reduced by p_j:

$$a_{i+1} = a_i + b_i \sum_{k=1}^{j-1} p_k \quad \text{and} \quad b_{i+1} = b_i \, p_j .$$

The final interval $[a_n, a_n + b_n]$ characterizes the sequence $s_1, \ldots, s_n$ unambiguously because the K^n different blocks of symbols $\vec{s}$ correspond to K^n different

intervals that make a partition of $[0,1]$. Since these intervals are non-overlapping, $[a_n, a_n + b_n]$ is characterized by coding in binary form a number $c_n \in [a_n, a_n + b_n]$. The binary expression of the chosen numbers c_n for each of the K^n intervals defines a prefix code, so that a sequence of such numbers is uniquely decodable. The value of c_n is progressively calculated by adding refinement bits when $[a_i, a_i + b_i]$ is reduced in the next sub-interval $[a_{i+1}, a_{i+1} + b_{i+1}]$ until $[a_n, a_n + b_n]$. There are efficient implementations that avoid numerical errors caused by the finite precision of arithmetic calculations when calculating c_n [355]. The resulting binary number c_n has d_n digits with

$$-\lceil \log_2 b_n \rceil \leq d_n \leq -\lceil \log_2 b_n \rceil + 2 \, .$$

Since $\log_2 b_n = \sum_{i=1}^{n} \log_2 p(s_i)$ and $\mathcal{H}(X) = \mathsf{E}\{\log_2 X\}$, one can verify that the average number of bits per symbol of this arithmetic code satisfies

$$\mathcal{H}(X) \leq R_X \leq \mathcal{H}(X) + \frac{2}{n} \, . \tag{11.13}$$

When the successive values X_k of the blocks are not independent, the upper and lower bounds (11.13) remain valid because the successive symbols are encoded as if they were independent.

An arithmetic code has a causal structure in the sense that the first i symbols of a sequence $s_1, \ldots, s_i, s_{i+1}, \ldots, s_n$ are specified by an interval $[a_i, a_i + b_i]$ that does not depend on the value of the last $n - i$ symbols. Since the sequence is progressively coded and decoded, one can implement an adaptive version which progressively learns the probability distribution $p(x)$ [273, 295]. When coding s_{i+1}, this probability distribution can be approximated by the histogram (cumulative distribution) $p_i(x)$ of the first i symbols. The sub-interval of $[a_i, a_i + b_i]$ associated to s_{i+1} is calculated with this estimated probability distribution. Suppose that $s_{i+1} = x_j$. We denote $p_i(x_j) = p_{i,j}$. The new interval is defined by

$$a_{i+1} = a_i + b_i \sum_{k=1}^{j-1} p_{i,k} \quad \text{and} \quad b_{i+1} = b_i \, p_{i,j} \, . \tag{11.14}$$

The decoder is able to recover s_{i+1} by recovering the first i symbols of the sequence and computing the cumulative probability distribution $p_i(x)$ of these symbols. The interval $[a_{i+1}, a_{i+1} + b_{i+1}]$ is then calculated from $[a_i, a_i + b_i]$ with (11.14). The initial distribution $p_0(x)$ can be set to be uniform.

If the symbols of the block are produced by independent random variables, then as i increases the estimated probability distribution $p_i(x)$ converges to the probability distribution $p(x)$ of the source. As the total block size n increases to $+\infty$ one can prove that the average bit rate of this adaptive arithmetic code converges to the entropy of the source. Under weaker Markov random chain hypotheses this result remains also valid [295].

Noise Sensitivity Huffman and arithmetic codes are more compact than a simple fixed length code of size $\log_2 K$, but they are also more sensitive to errors. For a constant length code, a single bit error modifies the value of only one symbol. In contrast, a single bit error in a variable length code may modify the whole symbol sequence. In noisy transmissions where such errors might occur, it is necessary to use an error correction code that introduces a slight redundancy in order to suppress the transmission errors [20].

11.2.2 Scalar Quantization

If the source X has arbitrary real values, it cannot be coded with a finite number of bits. A scalar quantizer Q approximates X by $\tilde{X} = Q(X)$, which takes its values over a finite set. We study the optimization of such a quantizer in order to minimize the number of bits needed to code $\tilde{X}$ for a given mean-square error

$$d = \mathrm{E}\{(X - \tilde{X})^2\}.$$

Suppose that X takes its values in $[a,b]$, which may correspond to the whole real axis. We decompose $[a,b]$ in K intervals $\{(y_{k-1}, y_k]\}_{1 \leq k \leq K}$ of variable length, with $y_0 = a$ and $y_K = b$. A scalar quantizer approximates all $x \in (y_{k-1}, y_k]$ by x_k:

$$\forall x \in (y_{k-1}, y_k] \ , \ Q(x) = x_k \ .$$

The intervals $(y_{k-1}, y_k]$ are called *quantization bins*. Rounding off integers is a simple example where the quantization bins $(y_{k-1}, y_k] = (k - \frac{1}{2}, k + \frac{1}{2}]$ have size 1 and $x_k = k$ for any $k \in \mathbb{Z}$.

High-Resolution Quantizer Let $p(x)$ be the probability density of the random source X. The mean-square quantization error is

$$d = \mathrm{E}\{(X - \tilde{X})^2\} = \int_{-\infty}^{+\infty} \Big(x - Q(x)\Big)^2 p(x)\, dx. \tag{11.15}$$

A quantizer is said to have a *high resolution* if $p(x)$ is approximately constant on each quantization bin $(y_{k-1}, y_k]$ of size $\Delta_k = y_k - y_{k-1}$. This is the case if the sizes Δ_k are sufficiently small relative to the rate of variation of $p(x)$, so that one can neglect these variations in each quantization bin. We then have

$$p(x) = \frac{p_k}{\Delta_k} \ \text{ for } x \in (y_{k-1}, y_k], \tag{11.16}$$

where

$$p_k = \Pr\{X \in (y_{k-1}, y_k]\}.$$

The next proposition computes the mean-square error under this high resolution hypothesis.

Proposition 11.3 *For a high resolution quantizer, the mean-square error d is minimized when $x_k = (y_k + y_{k-1})/2$, which yields*

$$d = \frac{1}{12} \sum_{k=1}^{K} p_k \Delta_k^2. \tag{11.17}$$

Proof[2]. The quantization error (11.15) can be rewritten

$$d = \sum_{k=1}^{K} \int_{y_{k-1}}^{y_k} (x - x_k)^2 \, p(x) \, dx.$$

Replacing $p(x)$ by its expression (11.16) gives

$$d = \sum_{k=1}^{K} \frac{p_k}{\Delta_k} \int_{y_{k-1}}^{y_k} (x - x_k)^2 \, dx. \tag{11.18}$$

One can verify that each integral is minimum for $x_k = (y_k + y_{k-1})/2$, which yields (11.17). ■

Uniform Quantizer The uniform quantizer is an important special case where all quantization bins have the same size

$$y_k - y_{k-1} = \Delta \quad \text{for } 1 \leq k \leq K.$$

For a high resolution uniform quantizer, the average quadratic distortion (11.17) becomes

$$d = \frac{\Delta^2}{12} \sum_{k=1}^{K} p_k = \frac{\Delta^2}{12}. \tag{11.19}$$

It is independent of the probability density $p(x)$ of the source.

Entropy Constrained Quantizer We want to minimize the number of bits required to code the quantized values $\tilde{X} = Q(X)$ for a fixed distortion $d = \mathrm{E}\{(X - \tilde{X})^2\}$. The Shannon Theorem 11.1 proves that the minimum average number of bits to code $\tilde{X}$ is the entropy $\mathcal{H}(\tilde{X})$. Huffman or an arithmetic codes produce bit rates close to this entropy lower bound. We thus design a quantizer that minimizes $\mathcal{H}(\tilde{X})$.

The quantized source $\tilde{X}$ takes K possible values $\{x_k\}_{1 \leq k \leq K}$ with probabilities

$$p_k = \Pr(\tilde{X} = x_k) = \Pr(X \in (y_{k-1}, y_k]) = \int_{y_{k-1}}^{y_k} p(x) \, dx.$$

Its entropy is

$$\mathcal{H}(\tilde{X}) = -\sum_{k=1}^{K} p_k \log_2 p_k.$$

For a high resolution quantizer, the following theorem of Gish and Pierce [191] relates $\mathcal{H}(\tilde{X})$ to the *differential entropy* of X defined by

$$\mathcal{H}_d(X) = -\int_{-\infty}^{+\infty} p(x)\log_2 p(x)\,dx. \tag{11.20}$$

Theorem 11.2 (GISH, PIERCE) *If Q is a high resolution quantizer with respect to $p(x)$, then*

$$\mathcal{H}(\tilde{X}) \geq \mathcal{H}_d(X) - \frac{1}{2}\log_2(12d). \tag{11.21}$$

This inequality is an equality if and only if Q is a uniform quantizer.

Proof[2]. By definition, a high resolution quantizer satisfies (11.16), so $p_k = p(x)\Delta_k$ for $x \in (y_{k-1}, y_k]$. Hence

$$\begin{aligned}\mathcal{H}(\tilde{X}) &= -\sum_{k=1}^{K} p_k \log_2 p_k \\ &= -\sum_{k=1}^{K}\int_{y_{k-1}}^{y_k} p(x)\log_2 p(x)\,dx - \sum_{k=1}^{K} p_k \log_2 \Delta_k \\ &= \mathcal{H}_d(X) - \frac{1}{2}\sum_{k=1}^{K} p_k \log_2 \Delta_k^2. \end{aligned} \tag{11.22}$$

The Jensen inequality for a concave function $\phi(x)$ proves that if $p_k \geq 0$ with $\sum_{k=1}^{K} p_k = 1$, then for any $\{a_k\}_{1\leq k\leq K}$

$$\sum_{k=1}^{K} p_k \phi(a_k) \leq \phi\left(\sum_{k=1}^{K} p_k a_k\right). \tag{11.23}$$

If $\phi(x)$ is strictly concave, the inequality is an equality if and only if all a_k are equal when $p_k \neq 0$. Since $\log_2(x)$ is strictly concave, we derive from (11.17) and (11.23) that

$$\frac{1}{2}\sum_{k=1}^{K} p_k \log_2(\Delta_k^2) \leq \frac{1}{2}\log_2\left(\sum_{k=1}^{K} p_k \Delta_k^2\right) = \frac{1}{2}\log_2(12d).$$

Inserting this in (11.22) proves that

$$\mathcal{H}(\tilde{X}) \geq \mathcal{H}_d(X) - \frac{1}{2}\log_2(12d).$$

This inequality is an equality if and only if all Δ_k are equal, which corresponds to a uniform quantizer. ■

This theorem proves that for a high resolution quantizer, the minimum average bit rate $R_X = \mathcal{H}(\tilde{X})$ is achieved by a uniform quantizer and

$$R_X = \mathcal{H}_d(X) - \frac{1}{2}\log_2(12d). \tag{11.24}$$

In this case $d = \Delta^2/12$ so

$$R_X = \mathcal{H}_d(X) - \log_2 \Delta. \tag{11.25}$$

The distortion rate is obtained by taking the inverse of (11.24):

$$d(R_X) = \frac{1}{12} 2^{2\mathcal{H}_d(X)} 2^{-2R_X}. \tag{11.26}$$

11.3 HIGH BIT RATE COMPRESSION [2]

Section 11.3.1 studies the distortion rate performance of a transform coding computed with high resolution quantizers. These results are illustrated with a wavelet transform image coder. For Gaussian processes, Section 11.3.2 proves that the optimal basis is the Karhunen-Loève basis. An application to audio compression is studied in Section 11.3.3.

11.3.1 Bit Allocation

Let us optimize the transform code of a random vector $F[n]$ decomposed in an orthonormal basis $\{g_m\}_{0 \leq m < N}$:

$$F = \sum_{m=0}^{N-1} F_B[m]\, g_m.$$

Each $F_B[m]$ is a zero-mean source that is quantized into $\tilde{F}_B[m]$ with an average bit budget R_m. For a high resolution quantization, Theorem 11.2 proves that the error $d_m = \mathrm{E}\{|F_B[m] - \tilde{F}_B[m]|^2\}$ is minimized with a uniform scalar quantization, and $R_m = \mathcal{H}_d(X) - \log_2 \Delta_m$ where Δ_m is the bin size. It now remains to optimize the choice of $\{\Delta_m\}_{0 \leq m < N}$ in order to minimize the average total number of bits

$$R = \sum_{m=0}^{N-1} R_m$$

for a specified total error

$$d = \sum_{m=0}^{N-1} d_m.$$

Let $\bar{R} = R/N$ be the average number of bits per sample. The following bit allocation theorem proves that the transform code is optimized when all Δ_m are equal.

Theorem 11.3 *For high resolution quantizations and a fixed total distortion d the number of bits R is minimum if*

$$\Delta_m^2 = \frac{12\, d}{N} \quad \textit{for } 0 \leq m < N \tag{11.27}$$

and

$$d(\bar{R}) = \frac{N}{12} 2^{2\overline{\mathcal{H}}_d} 2^{-2\bar{R}}, \tag{11.28}$$

where $\overline{\mathcal{H}_d}$ is the averaged differential entropy

$$\overline{\mathcal{H}_d} = \frac{1}{N}\sum_{m=0}^{N-1} \mathcal{H}_d(F_{\mathcal{B}}[m]).$$

Proof. For uniform high resolution quantizations, (11.25) proves that

$$R_m = \mathcal{H}_d(F_{\mathcal{B}}[m]) - \frac{1}{2}\log_2(12\,d_m).$$

So

$$R = \sum_{m=0}^{N-1} R_m = \sum_{m=0}^{N-1} \mathcal{H}_d(F_{\mathcal{B}}[m]) - \sum_{m=0}^{N-1} \frac{1}{2}\log_2(12\,d_m). \tag{11.29}$$

Minimizing R is equivalent to maximizing $\sum_{m=0}^{N-1}\log_2(12d_m)$. Applying the Jensen inequality (11.23) to the concave function $\phi(x) = \log_2(x)$ and $p_k = 1/N$ proves that

$$\frac{1}{N}\sum_{m=0}^{N-1}\log_2(12\,d_m) \leq \log_2\left(\frac{12}{N}\sum_{m=0}^{N-1} d_m\right) = \log_2\left(\frac{12\,d}{N}\right).$$

This inequality is an equality if and only if all d_m are equal. Hence $\Delta_m^2/12 = d_m = d/N$, which proves (11.27). We also derive from (11.29) that

$$R = \sum_{m=0}^{N-1} \mathcal{H}_d(F_{\mathcal{B}}[m]) - \frac{N}{2}\log_2\left(\frac{12\,d}{N}\right)$$

which implies (11.28). ■

This theorem shows that the transform code is optimized if it introduces the same expected error $d_m = \Delta_m^2/12 = d/N$ along each direction g_m of the basis $\mathcal{B}$. The number of bits R_m used to encode $F_{\mathcal{B}}[m]$ then depends only on its differential entropy:

$$R_m = \mathcal{H}_d(F_{\mathcal{B}}[m]) - \frac{1}{2}\log_2\left(\frac{12d}{N}\right). \tag{11.30}$$

Let σ_m^2 be the variance of $F_{\mathcal{B}}[m]$, and let $\tilde{F}_{\mathcal{B}}[m] = F_{\mathcal{B}}[m]/\sigma_m$ be the normalized random variable of variance 1. A simple calculation shows that

$$\mathcal{H}_d(F_{\mathcal{B}}[m]) = \mathcal{H}_d(\tilde{F}_{\mathcal{B}}[m]) + \log_2\sigma_m\,.$$

The "optimal bit allocation" R_m in (11.30) may become negative if the variance σ_m is too small, which is clearly not an admissible solution. In practice, R_m must be a positive integer but the resulting optimal solution has no simple analytic expression (Problem 11.7). If we neglect the integral bit constraint, (11.30) gives the optimal bit allocation as long as $R_m \geq 0$.

Weighted Mean-Square Error We mentioned that a mean-square error often does not measure the perceived distortion of images or audio signals very well. When the vectors g_m are localized in time and frequency, a mean-square norm sums the errors at all times and frequencies with equal weights. It thus hides the temporal and frequency properties of the error $F - \tilde{F}$. Better norms can be constructed by emphasizing certain frequencies more than others, in order to match our audio or visual sensitivity, which varies with the signal frequency. A weighted mean-square norm is defined by

$$d = \sum_{m=0}^{N-1} \frac{d_m}{w_m^2}, \tag{11.31}$$

where $\{w_m^2\}_{0 \leq m < N}$ are constant weights.

We can apply Theorem 11.3 to weighted mean-square errors by observing that

$$d = \sum_{m=0}^{N-1} d_m^w,$$

where $d_m^w = d_m / w_m^2$ is the quantization error of $F_B^w[m] = F_B[m]/w_m$. Theorem 11.3 proves that bit allocation is optimized by quantizing uniformly all $F_B^w[m]$ with the same bin size Δ. This implies that the coefficients $F_B[m]$ are uniformly quantized with a bin size $\Delta_m = \Delta w_m$; it follows that $d_m = w_m^2 d/N$. As expected, larger weights increase the error in the corresponding direction. The uniform quantization Q_{Δ_m} with bins of size Δ_m is often computed with a quantizer Q that associates to any real number its closest integer:

$$Q_{\Delta_m}\left(F_B[m]\right) = \Delta_m Q\left(\frac{F_B[m]}{\Delta_m}\right) = \Delta w_m Q\left(\frac{F_B[m]}{\Delta w_m}\right). \tag{11.32}$$

11.3.2 Optimal Basis and Karhunen-Loève

Transform code performance depends on the choice of an orthonormal basis $\mathcal{B}$. For high resolution quantizations, (11.28) proves that the distortion rate $d(\bar{R})$ is optimized by choosing a basis $\mathcal{B}$ that minimizes the average differential entropy

$$\overline{\mathcal{H}}_d = \frac{1}{N} \sum_{m=0}^{N-1} \mathcal{H}_d(F_B[m]).$$

In general, we do not know how to compute this optimal basis because the probability density of the $F_B[m] = \langle F, g_m \rangle$ may depend on g_m in a complicated way.

Gaussian Process If F is a Gaussian random vector then the coefficients $F_B[m]$ are Gaussian random variables in any basis. In this case, the probability density of $F_B[m]$ depends only on the variance σ_m^2:

$$p_m(x) = \frac{1}{\sigma_m \sqrt{2\pi}} \exp\left(\frac{-x^2}{2\sigma_m^2}\right).$$

With a direct integration, we verify that

$$\mathcal{H}_d\left(F_{\mathcal{B}}[m]\right) = -\int_{-\infty}^{+\infty} p_m(x) \log_2 p_m(x)\, dx = \log_2 \sigma_m + \log_2 \sqrt{2\pi \mathrm{e}}.$$

Inserting this expression in (11.28) yields

$$d(\bar{R}) = N \frac{\pi \mathrm{e}}{6} \rho^2 2^{-2\bar{R}}, \tag{11.33}$$

where ρ^2 is the geometrical mean of all variances:

$$\rho^2 = \left(\prod_{m=0}^{N-1} \sigma_m^2 \right)^{1/N}.$$

The basis must therefore be chosen in order to minimize ρ^2.

Proposition 11.4 *The geometrical mean variance ρ^2 is minimized in a Karhunen-Loève basis of F.*

Proof[2]. Let K be the covariance operator of F,

$$\sigma_m^2 = \langle K g_m, g_m \rangle.$$

Observe that

$$\log_2 \rho^2 = \frac{1}{N} \sum_{m=0}^{N-1} \log_2 (\langle K g_m, g_m \rangle). \tag{11.34}$$

Theorem 10.15 proves that if $\Phi(x)$ is strictly concave then

$$\sum_{m=0}^{N-1} \Phi(\langle K g_m, g_m \rangle)$$

is minimum if and only if $\{g_m\}_{0 \le m < N}$ diagonalizes K. Since $\log_2(x)$ is strictly concave, we derive that ρ^2 is minimum if and only if $\mathcal{B}$ is a Karhunen-Loève basis. ■

Together with the distortion rate (11.33), this result proves that a high bit rate transform code of a Gaussian process is optimized in a Karhunen-Loève basis. The Karhunen-Loève basis diagonalizes the covariance matrix, which means that the decomposition coefficients $F_{\mathcal{B}}[m] = \langle F, g_m \rangle$ are uncorrelated. If F is a Gaussian random vector, then the coefficients $F_{\mathcal{B}}[m]$ are jointly Gaussian. In this case, being uncorrelated implies that they are independent. The optimality of a Karhunen-Loève basis is therefore quite intuitive since it produces coefficients $F_{\mathcal{B}}[m]$ that are independent. The independence of the coefficients justifies using a scalar quantization rather than a vector quantization.

Coding Gain The Karhunen-Loève basis $\{g_m\}_{0 \leq m < N}$ of F is a priori not well structured. The decomposition coefficients $\{\langle f, g_m \rangle\}_{0 \leq m < N}$ of a signal f are thus computed with N^2 multiplications and additions, which is often too expensive in real time coding applications. Transform codes often approximate this Karhunen-Loève basis by a more structured basis that admits a faster decomposition algorithm. The performance of a basis is evaluated by the coding gain [35]

$$G = \frac{\mathrm{E}\{\|F\|^2\}}{N \rho^2} = \frac{\sum_{m=0}^{N-1} \sigma_m^2}{N \left(\prod_{m=0}^{N-1} \sigma_m^2 \right)^{1/N}}. \tag{11.35}$$

Proposition 11.4 proves that G is maximum in a Karhunen-Loève basis.

Non-Gaussian Processes When F is not Gaussian, the coding gain G no longer measures the coding performance of the basis. Indeed, the distortion rate (11.28) depends on the average differential entropy factor $2^{2\overline{H}_d}$, which is not proportional to ρ^2. The Karhunen-Loève basis is therefore not optimal.

Circular stationary processes with piecewise smooth realizations are examples of non-Gaussian processes that are not well compressed in their Karhunen-Loève basis, which is the discrete Fourier basis. Section 11.4 shows that wavelet bases yield better distortion rates because they can approximate these signals with few non-zero coefficients.

11.3.3 Transparent Audio Code

The compact disk standard samples high quality audio signals at 44.1 kHz. Samples are quantized with 16 bits, producing a Pulse Code Modulation of 706 kb/s. Audio codes must be "transparent," which means that they should not introduce errors that can be heard by an "average" listener.

Sounds are often modeled as realizations of Gaussian processes. This justifies the use of a Karhunen-Loève basis to minimize the distortion rate of transform codes. To approximate the Karhunen-Loève basis, we observe that many audio signals are locally stationary over a sufficiently small time interval. This means that over this time interval, the signal can be approximated by a realization of a stationary process. Section 10.6.2 explains that the Karhunen-Loève basis of locally stationary processes is well approximated by a local cosine basis with appropriate window sizes. The local stationarity hypothesis is not always valid, especially for attacks of musical instruments, but bases of local time-frequency atoms remain efficient for most audio segments.

Bases of time-frequency atoms are also well adapted to matching the quantization errors with our hearing sensitivity. Instead of optimizing a mean-square error as in Theorem 11.3, perceptual coders [226] adapt the quantization so that errors fall below an auditory threshold, which depends on each time-frequency atom g_m.

Audio Masking A large amplitude stimulus often makes us less sensitive to smaller stimuli of a similar nature. This is called a masking effect. In a sound, a

small amplitude quantization error may not be heard if it is added to a strong signal component in the same frequency neighborhood. Audio masking takes place in critical frequency bands $[\omega_c - \Delta\omega/2, \omega_c + \Delta\omega/2]$ that have been measured with psychophysical experiments [304]. A strong narrow band signal whose frequency energy is in the interval $[\omega_c - \Delta\omega/2, \omega_c + \Delta\omega/2]$ decreases the hearing sensitivity within this frequency interval. However, it does not influence the sensitivity outside this frequency range. In the frequency interval $[0, 20\text{kHz}]$, there are approximately 25 critical bands. Below 700 Hz, the bandwidths of critical bands are on the order of 100 Hz. Above 700 Hz the bandwidths increase proportionally to the center frequency ω_c:

$$\Delta\omega \approx \begin{cases} 100 \text{ for } \omega_c \leq 700 \\ 0.15\omega_c \text{ for } 700 \leq \omega_c \leq 20,000 \end{cases} \tag{11.36}$$

The masking effect also depends on the nature of the sound, particularly its tonality. A tone is a signal with a narrow frequency support as opposed to a noise-like signal whose frequency spectrum is spread out. A tone has a different masking influence than a noise-type signal; this difference must be taken into account [314].

Adaptive Quantization To take advantage of audio masking, transform codes are implemented in orthogonal bases of local time-frequency atoms $\{g_m\}_{0 \leq m < N}$, whose frequency supports are inside critical bands. To measure the effect of audio masking at different times, the signal energy is computed in each critical band. This is done with an FFT over short time intervals, on the order of 10ms, where signals are considered to be approximately stationary. The signal tonality is estimated by measuring the spread of its Fourier transform. The maximum admissible quantization error in each critical band is estimated depending on both the total signal energy in the band and the signal tonality. This estimation is done with approximate formulas that are established with psychophysical experiments [240]. For each vector g_m whose Fourier transform is inside a given critical band, the inner product $\langle f, g_m \rangle$ is uniformly quantized according to the maximum admissible error. Quantized coefficients are then entropy coded.

Although the SNR may be as low as 13 db, such an algorithm produces a nearly transparent audio code because the quantization error is below the perceptual threshold in each critical band. The most important degradations introduced by such transform codes are *pre-echoes*. During a silence, the signal remains zero, but it can suddenly reach a large amplitude due to a beginning speech or a musical attack. In a short time interval containing this attack, the signal energy may be quite large in each critical band. By quantizing the coefficients $\langle f, g_m \rangle$ we introduce an error both in the silent part and in the attack. The error is not masked during the silence and will clearly be heard. It is perceived as a "pre-echo" of the attack. This pre-echo is due to the temporal variation of the signal, which does not respect the local stationarity hypothesis. It can however be detected and removed with post-processings.

Choice of Basis The MUSICAM (Masking-pattern Universal Subband Integrated Coding and Multiplexing) coder [153] used in the MPEG-I standard [102] is the simplest perceptual subband coder. It decomposes the signal in 32 equal frequency bands of 750 Hz bandwidth, with a filter bank constructed with frequency modulated windows of 512 samples. This decomposition is similar to a signal expansion in a local cosine basis. The quantization levels are adapted in each frequency band, to take into account the masking properties of the signal. Quantized coefficients are not entropy coded. This system compresses audio signals up to 128 kb/s without audible impairment. It is often used for digital radio transmissions where small defects are admissible.

The AC-systems produced by Dolby decompose the signal in a local cosine basis, and adapt the window sizes to the local signal content. They also perform a perceptual quantization followed by a Huffman entropy coding. These coders operate on a variety of bit rates from 64 kb/s to 192 kb/s.

In order to best match human perception, transform code algorithms have been developed in wavelet packet bases, whose frequency resolution match the critical frequency bands [350]. Sinha and Tewfik [314] propose the wavelet packet basis shown in Figure 11.4, which is an $M = 4$ wavelet basis. The properties of M-band wavelet bases are explained in Section 8.1.3. These four wavelets have a bandwidth of $1/4$, $1/5$, $1/6$ and $1/7$ octaves respectively. The lower frequency interval $[0, 700]$ is decomposed with eight wavelet packets of the same bandwidth, to match the critical frequency bands (11.36). These wavelet packet coefficients are quantized with perceptual models and are entropy coded. Nearly transparent audio codes are obtained at 70 kb/s.

Wavelets produce smaller pre-echo distortions compared to local cosine bases. At the sound attack, the largest wavelet coefficients appear at fine scales. Because fine scale wavelets have a short support, a quantization error creates a distortion that is concentrated near the attack. However, these bases have the disadvantage of introducing a bigger coding delay than local cosine bases. The coding delay is approximately equal to half the maximum time support of the vector used in the basis. It is typically larger for wavelets and wavelet packets than for local cosine vectors.

Choice of Filter Wavelet and wavelet packet bases are constructed with a filter bank of conjugate mirror filters. For perceptual audio coding, the Fourier transform of each wavelet or wavelet packet must have its energy well concentrated in a single critical band. Second order lobes that may appear in other frequency bands should have a negligible amplitude. Indeed, a narrow frequency tone creates large amplitude coefficients for all wavelets whose frequency support covers this tone, as shown in Figure 11.5. Quantizing the wavelet coefficients is equivalent to adding small wavelets with amplitude equal to the quantization error. If the wavelets excited by the tone have important second order lobes in other frequency intervals, the quantization errors introduces some energy in these frequency intervals that is

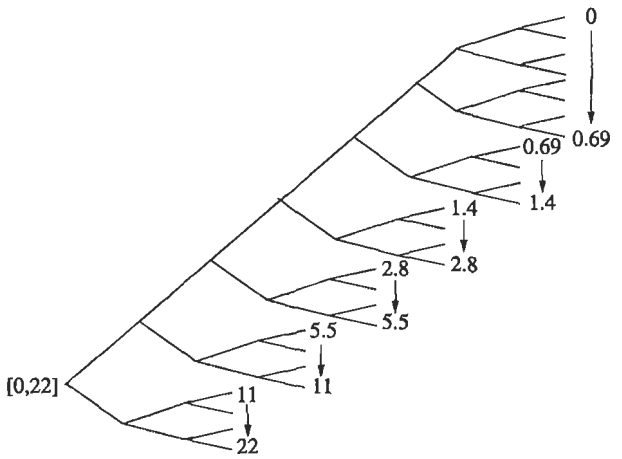


FIGURE 11.4 Wavelet packet tree that decomposes the frequency interval [0, 22kHz] in 24 frequency bands covered by $M = 4$ wavelets dilated over six octaves, plus eight low frequency bands of the same bandwidth. The frequency bands are indicated at the leaves in kHz.

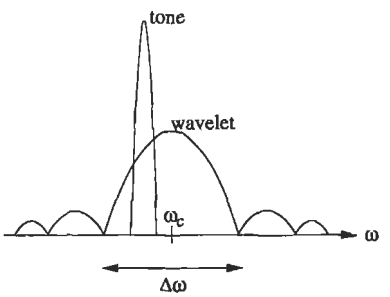


FIGURE 11.5 A high energy narrow frequency tone can excite a wavelet whose Fourier transform has second order lobes outside the critical band of width $\Delta\omega$. The quantization then creates audible distortion.

not masked by the energy of the tone, introducing audible distortion.

To create wavelets and wavelet packets with small second order frequency lobes, the transfer function of the corresponding conjugate mirror filter $\hat{h}(\omega)$ must have a zero of high order at $\omega = \pi$. Theorem 7.5 proves that conjugate mirror filters with p zeros at $\omega = \pi$ have at least $2p$ non-zero coefficients, and correspond to wavelets of size $2p - 1$. Increasing p thus produces a longer coding delay. Numerical experiments [314] show that increasing p up to 30 can enhance the perceptual quality of the audio code, but the resulting filters have at least 60 non-zero coefficients.

11.4 IMAGE COMPRESSION [2]

So far, we have studied the performance of transform codes from a Bayes point of view, by considering signals as realizations of a random vector whose probability distribution is known. However, there is no stochastic model that incorporates the diversity of image structures such as non-stationary textures and edges. In particular, Gaussian processes and homogeneous Markov random fields are not appropriate. The distortion rate formulas were also calculated with a high resolution quantization hypothesis, which is not valid for image transform codes.

Section 11.4.1 introduces a different framework where the distortion rate is computed by considering images as deterministic signals. Image transform codes in orthonormal wavelet bases and block cosine bases are studied in Sections 11.4.2 and 11.4.3. Embedding strategies to improve wavelet transform codes are introduced in Section 11.4.4.

Any prior information about the class of images to be compressed can be used to specify a set Θ that includes this class. For example, large classes of images are included in sets of bounded variation signals. In the absence of probabilistic models, we cannot calculate the expected coding distortion over Θ, which is replaced by the maximum distortion. Minimizing this maximum distortion leads to the notion of Kolmogorov ϵ-entropy. Section 11.4.5 gives conditions for reaching a minimax distortion rate with a transform coding.

11.4.1 Deterministic Distortion Rate

An image is considered as a deterministic signal f that is decomposed in an orthonormal basis $\mathcal{B} = \{g_m\}_{0 \leq m < N^2}$:

$$f = \sum_{m=0}^{N^2-1} f_{\mathcal{B}}[m]\, g_m \quad \text{with } f_{\mathcal{B}}[m] = \langle f, g_m \rangle .$$

A transform code quantizes all coefficients and reconstructs

$$\tilde{f} = \sum_{m=0}^{N^2-1} Q(f_{\mathcal{B}}[m])\, g_m . \tag{11.37}$$

Let R be the number of bits used to code the N^2 quantized coefficients $Q(f_{\mathcal{B}}[m])$. The coding distortion is

$$d(R, f) = \|f - \tilde{f}\|^2 = \sum_{m=0}^{N^2-1} |f_{\mathcal{B}}[m] - Q(f_{\mathcal{B}}[m])|^2 . \tag{11.38}$$

We denote by $p(x)$ the histogram of the N^2 coefficients $f_{\mathcal{B}}[m]$, normalized so that $\int p(x)\, dx = 1$. The quantizer approximates each $x \in (y_{k-1}, y_k]$ by $Q(x) = x_k$. The proportion of quantized coefficients equal to x_k is

$$p_k = \int_{y_{k-1}}^{y_k} p(x)\, dx . \tag{11.39}$$

Suppose that the quantized values of f can take at most K different quantized values x_k. A variable length code represents the quantized values equal to x_k with an average of l_k bits, where the lengths l_k are specified independently from f. It is implemented with a prefix code or an arithmetic code, over blocks of quantized values that are large enough so that the l_k can be assumed to take any real values that satisfy the Kraft inequality (11.6) $\sum_{k=1}^{K} 2^{-l_k} \leq 1$. Encoding a signal with K symbols requires a total number of bits

$$R = N^2 \sum_{k=1}^{K} p_k l_k \, . \tag{11.40}$$

A *constant size code* corresponds to $l_k = \log_2 K$, in which case $R = N^2 \log_2 K$. The bit budget R reaches its minimum for $l_k = -\log_2 p_k$ and hence

$$R \geq \mathcal{H} = -N^2 \sum_{k=1}^{K} p_k \log_2 p_k \, . \tag{11.41}$$

We denote by $d_{\mathcal{H}}(R, f)$ the distortion obtained with $R = \mathcal{H}$. Minimizing R for a given quantizer produces a minimum distortion for a fixed R. So $d_{\mathcal{H}}(R, f)$ is a lower bound of the distortion rate $d(R, f)$ obtained with a prefix or an arithmetic code. In practice, we do not know in advance the values of p_k, which depend on the signal f. The *oracle distortion rate* $d_{\mathcal{H}}(R, f)$ is obtained by an *oracle coder* that uses extra information that is normally not available.

An *adaptive variable length code* takes a different approach, as explained in Section 11.2.1. Instead of fixing a priori the values $\{l_k\}_{1 \leq k \leq K}$, such a code estimates the distribution p_k as the coding progresses and adapts the lengths l_k. It produces a bit budget R that is often close to $\mathcal{H}$, but it can be smaller when the sequence of quantized coefficients is not homogeneous. For example, the wavelet coefficients of an image often have a larger amplitude at large scales. An adaptive arithmetic code adapts the encoding to the probability distribution which is different depending on the scale. It thus produces a total bit budget that is smaller than the entropy $\mathcal{H}$ obtained with a fixed code optimized for the N^2 wavelet coefficients.

High Resolution Quantization The high resolution assumption supposes that $p(x)$ is approximately constant over each quantization interval $(y_{k-1}, y_k]$. The following proposition computes the distortion rate under this assumption.

Proposition 11.5 *Suppose that the high resolution quantization assumption is valid for* $\bar{R} = R/N^2$.

- *The oracle distortion rate is minimum if and only if Q is a uniform quantizer and*

$$d_{\mathcal{H}}(\bar{R}, f) = \frac{N^2}{12} 2^{2\mathcal{H}_d(f)} 2^{-2\bar{R}}, \tag{11.42}$$

with $\mathcal{H}_d(f) = -\int p(x) \log_2 p(x) \, dx$.

- *Suppose that there exists C such that* $\sup_{0\leq m<N^2} |f_B[m]| \leq C$. *If the quantized coefficients are coded with constant size binary words then the distortion rate is*

$$d(\bar{R}, f) = \frac{N^2}{3} C^2 2^{-2\bar{R}}. \tag{11.43}$$

Proof[2]. Let X be a random variable whose probability distribution is the histogram $p(x)$. The distortion defined in (11.38) can be rewritten

$$d_{\mathcal{H}}(\bar{R}, f) = N^2 \mathrm{E}\{|X - Q(X)|^2\} .$$

The minimum bit budget (11.41) is equal to the entropy $R = \mathcal{H}(Q(X))$. Under the high resolution assumption, Theorem 11.2 proves that $\mathrm{E}\{|X-Q(X)|^2\}$ is minimum if and only if Q is a uniform quantizer, and (11.21) implies (11.42).

A uniform high resolution quantization with bin size Δ has a distortion calculated in (11.19): $d(\bar{R}, f) = N^2 \Delta^2/12$. The number of quantization bins is $K = 2C/\Delta$ and the total number of bits is $\bar{R} = \log_2 K$, from which we derive (11.43). ■

The high resolution quantization assumption is valid if the quantization bins are small enough, which means that the bit rate $\bar{R}$ is sufficiently large. In this case, the distortion rate decays exponentially.

Wavelet Image Code A simple wavelet image code is introduced to illustrate the properties of transform coding. The image is decomposed in a separable wavelet basis. All wavelet coefficients are quantized with a uniform quantizer

$$Q(x) = \begin{cases} 0 & \text{if } |x| < \Delta/2 \\ \mathrm{sign}(x)\, k\, \Delta & \text{if } (k-1/2)\Delta \leq |x| < (k+1/2)\Delta \end{cases} . \tag{11.44}$$

The quantized coefficients are coded with an adaptive arithmetic code. The particular choice of wavelet basis and the implementation details are discussed in the next section. Figure 11.6 shows examples of coded images with $\bar{R} = 0.5$ bit/pixel. Mandrill is the only image where one can see a slight degradation, in the fur.

The Peak Signal to Noise Ratio (PSNR) is defined by

$$PSNR(\bar{R}, f) = 10 \log_{10} \frac{N^2 255^2}{d(\bar{R}, f)}.$$

The distortion rate formula (11.42) predicts that there exists a constant K such that

$$PSNR(\bar{R}, f) = (20 \log_{10} 2)\, \bar{R} + K .$$

Figure 11.7 shows that $PSNR(\bar{R}, f)$ has indeed a linear growth for $\bar{R} \geq 1$, but not for $\bar{R} < 1$. At low bit rates $\bar{R} \leq 1$, the quantization interval Δ is relatively large. The normalized histogram $p(x)$ of wavelet coefficients in Figure 11.8 has a narrow peak in the neighborhood of $x = 0$. Hence $p(x)$ is poorly approximated by a constant in the zero bin $[-\Delta/2, \Delta/2]$, where $Q(x) = 0$. The high resolution quantization hypothesis is not valid in this zero bin, which explains why the distortion rate formula (11.42) is wrong. For Mandrill, the high resolution hypothesis remains valid up to $\bar{R} = 0.5$ because the histogram of its wavelet coefficients is wider in the neighborhood of $x = 0$.

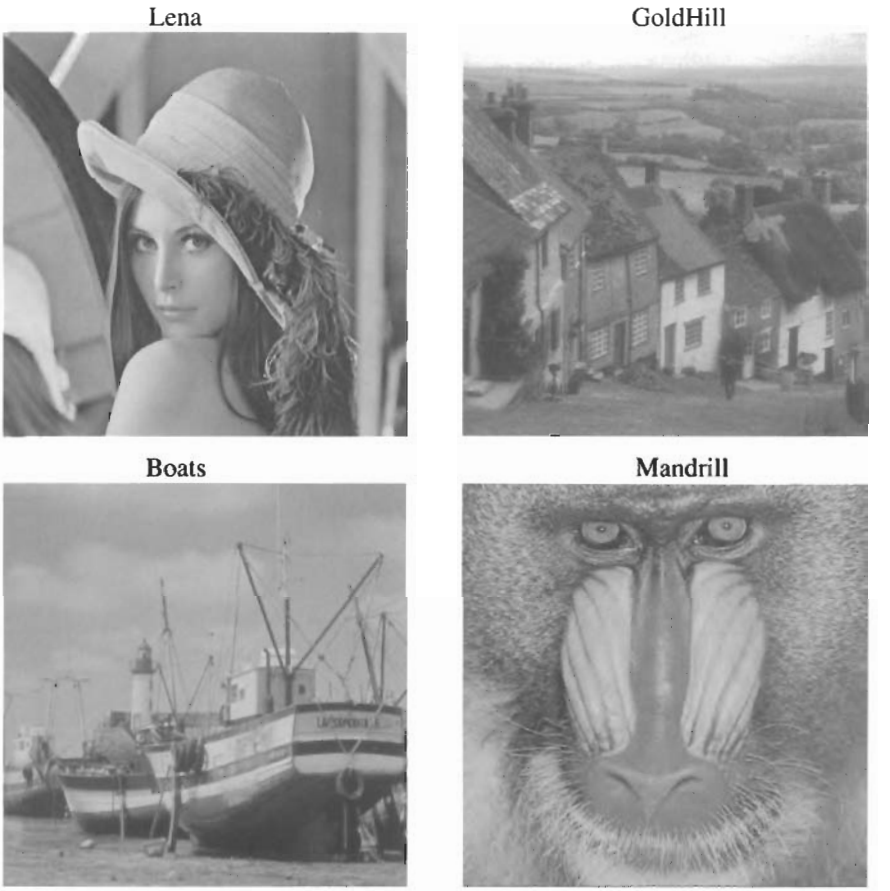


FIGURE 11.6 These images of $N^2 = 512^2$ pixels are coded with $\bar{R} = 0.5$ bit/pixel, by a wavelet transform coding.

Low Resolution Quantization If the basis $\mathcal{B}$ is chosen so that many coefficients $f_{\mathcal{B}}[m] = \langle f, g_m \rangle$ are close to zero, then the histogram $p(x)$ has a sharp high amplitude peak at $x = 0$, as in the wavelet histograms shown in Figure 11.8. At low bit rates R the distortion $d(R, f)$ must therefore be computed without using a high resolution quantization assumption.

The bit budget R can be calculated by considering separately the *significant coefficients* $f_{\mathcal{B}}[m]$ such that $Q(f_{\mathcal{B}}[m]) \neq 0$. The positions of these significant coefficients are recorded by a binary *significance map*

$$b[m] = \begin{cases} 0 & \text{if } Q(f_{\mathcal{B}}[m]) = 0 \\ 1 & \text{if } Q(f_{\mathcal{B}}[m]) \neq 0 \end{cases} . \tag{11.45}$$

Let M be the number of significant coefficients. The proportions of 0 and 1 in

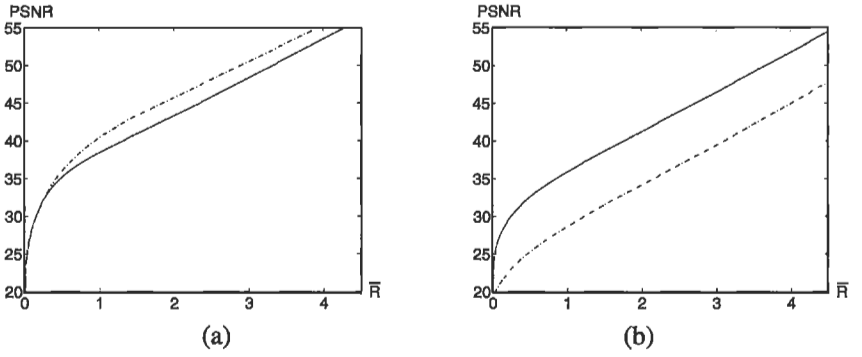


FIGURE 11.7 PSNR as a function of $\bar{R}$. (a): Lena (solid line) and Boats (dotted line). (b): GoldHill (solid line) and Mandrill (dotted line)

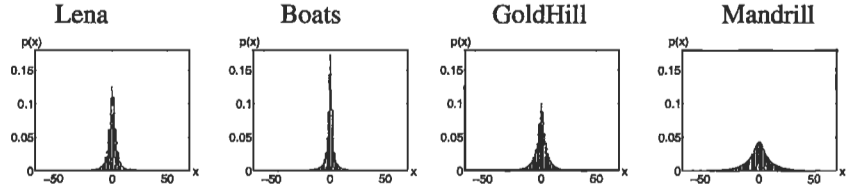


FIGURE 11.8 Normalized histograms of orthogonal wavelet coefficients for each image.

the significance map are respectively $p_0 = (N^2 - M)/N^2$ and $p_1 = M/N^2$. The number of bits R_0 needed to code the significance map with a variable length code has a lower bound calculated with $l_0 = -\log_2 p_0$ and $l_1 = -\log_2 p_1$:

$$R_0 \geq -N^2 \left(p_0 \log_2 p_0 + p_1 \log_2 p_1 \right) . \tag{11.46}$$

An adaptive variable length code can nearly reach this lower bound. The number $M \leq N^2$ of significant coefficients is first computed and coded on $\log_2 N^2$ bits. The values of p_0 and p_1 are derived from M and the significance map is coded with $l_0 = -\log_2 p_0$ and $l_1 = -\log_2 p_1$. This adaptive code has an overhead of $\log_2 N^2$ bits relative to the lower bound (11.46).

Figure 11.9 shows the significance maps of the quantized wavelet coefficients that code the four images in Figure 11.6. The total bit budget R to code all quantized coefficients is

$$R = R_0 + R_1 ,$$

where R_1 is the number of bits coding the quantized values of the significant coefficients, with a variable length code.

The distortion $d(R, f)$ is calculated by separating in (11.38) the coefficients

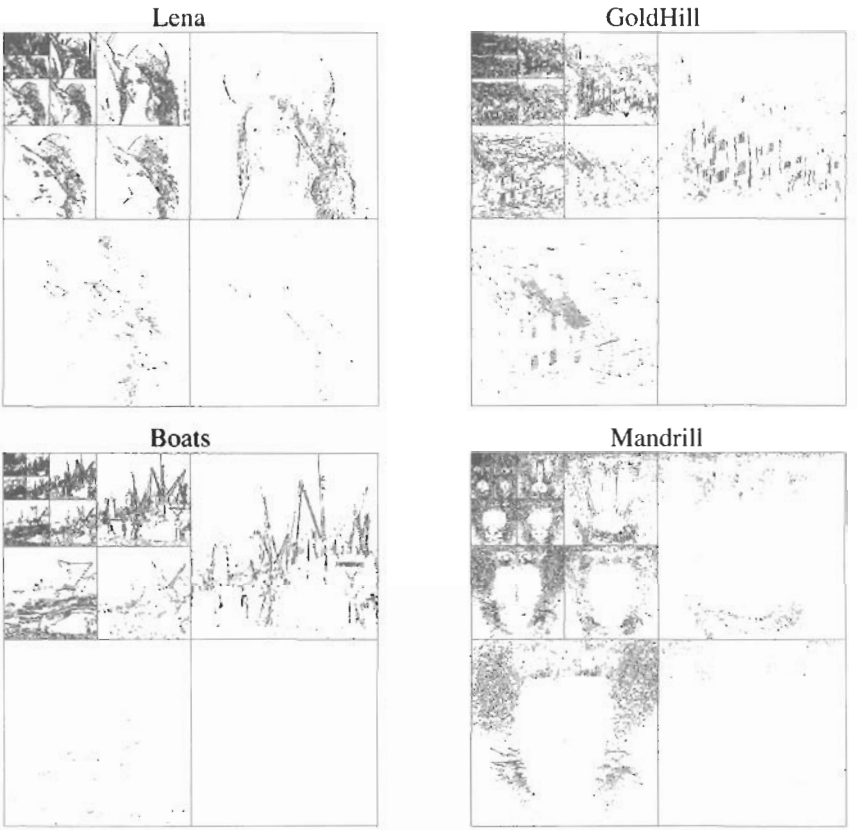


FIGURE 11.9 Significance map of quantized wavelet coefficients for images coded with $\bar{R} = 0.5$ bit/pixel.

that are in the zero bin $[-\Delta/2, \Delta/2]$ from the significant coefficients:

$$d(R,f) = \sum_{|f_{\mathcal{B}}[m]|<\Delta/2} |f_{\mathcal{B}}[m]|^2 + \sum_{|f_{\mathcal{B}}[m]|\geq\Delta/2} |f_{\mathcal{B}}[m] - Q(f_{\mathcal{B}}[m])|^2 . \quad (11.47)$$

Let f_M be the non-linear approximation of f from the M significant coefficients:

$$f_M = \sum_{|f_{\mathcal{B}}[m]|\geq\Delta/2} f_{\mathcal{B}}[m]\, g_m .$$

The first sum of $d(R,f)$ can be rewritten as a non-linear approximation error:

$$\|f - f_M\|^2 = \sum_{|f_{\mathcal{B}}[m]|<\Delta/2} |f_{\mathcal{B}}[m]|^2 .$$

DeVore, Jawerth and Lucier [156] observed that this approximation error often dominates the value of $d(R,f)$.

The following theorem computes $d(R,f)$ depending on the decay rate of the sorted decomposition of f in the basis $\mathcal{B}$. We denote by $f_{\mathcal{B}}^r[k] = f_{\mathcal{B}}[m_k]$ the coefficient of rank k, defined by $|f_{\mathcal{B}}^r[k]| \geq |f_{\mathcal{B}}^r[k+1]|$ for $1 \leq k \leq N^2$. We write $|f_{\mathcal{B}}^r[k]| \sim Ck^{-s}$ if there exist two constants $A, B > 0$ independent of C, k and N such that $ACk^{-s} \leq |f_{\mathcal{B}}^r[k]| \leq BCk^{-s}$.

Theorem 11.4 (FALZON, MALLAT) *Let Q be a uniform quantizer. There exists an adaptive variable length code such that for all $s > 1/2$ and $C > 0$, if $|f_{\mathcal{B}}^r[k]| \sim Ck^{-s}$ then for $R \leq N^2$*

$$d(R,f) \sim d_{\mathcal{H}}(R,f) \sim C^2 R^{1-2s} \left(1 + \log_2 \frac{N^2}{R}\right)^{2s-1} . \tag{11.48}$$

Proof[3]. Let Δ be the quantization step of the uniform quantizer. Since $0 \leq |x - Q(x)| \leq \Delta/2$, (11.47) implies

$$\|f - f_M\| \leq d(R,f) \leq \|f - f_M\|^2 + M\frac{\Delta^2}{4} , \tag{11.49}$$

where M is the number of coefficients such that $|f_{\mathcal{B}}[m]| \geq \Delta/2$. Since the sorted coefficients satisfy $|f_{\mathcal{B}}^r[k]| \sim Ck^{-s}$ we derive that

$$M \sim C^{1/s} \Delta^{-1/s} . \tag{11.50}$$

We shall see that $R < N^2$ implies $M < N^2/2$. Since $s > 1/2$, the approximation error is

$$\|f - f_M\|^2 = \sum_{k=M+1}^{N^2} |f_{\mathcal{B}}^r[k]|^2 \sim \sum_{k=M+1}^{N^2} C^2 k^{-2s} \sim C^2 M^{1-2s}. \tag{11.51}$$

But (11.50) shows that $M\Delta^2 \sim C^2 M^{1-2s}$ so (11.49) yields

$$d(R,f) \sim C^2 M^{1-2s}. \tag{11.52}$$

Let us now evaluate the bit budget $R = R_0 + R_1$. We construct an adaptive variable length code that requires a number of bits that is of the same order as the number of bits obtained with an oracle code. The proportion of significant and insignificant coefficients is respectively $p_1 = M/N^2$ and $p_0 = (N^2 - M)/N^2$. An oracle codes the significance map with a bit budget

$$\begin{aligned} \mathcal{H}_0 &= -N^2 \left(p_0 \log_2 p_0 + p_1 \log_2 p_1\right) \\ &= M\left(\log_2 \frac{N^2}{M} + \log_2 \mathrm{e} + O\left(\frac{M}{N^2}\right)\right) . \end{aligned} \tag{11.53}$$

An adaptive variable length code adds an overhead of $\log_2 N^2$ bits to store the value of M. This does not modify the order of magnitude of R_0:

$$R_0 \sim \mathcal{H}_0 \sim M\left(\log_2 \frac{N^2}{M} + 1\right) . \tag{11.54}$$

We also decompose $R_1 = R_a + R_s$, where R_a is the number of bits that code the amplitude of the M significant coefficients of f, and R_s is the number of bits that code their sign, given that their amplitude is already coded. Clearly $0 \leq R_s \leq M$. Let p_j be the fraction of significant coefficients whose amplitude is quantized to $j\Delta$. An oracle codes the amplitude with a variable length code defined by $l_j = -\log_2 p_j$. The resulting bit budget is

$$\mathcal{H}_a = -M \sum_{j=1}^{+\infty} p_j \log_2 p_j \,. \tag{11.55}$$

Let $n_j = M p_j$ be the number of coefficients such that $|Q(f_B^0[k])| = j\Delta$, which means that $|f_B^r[k]| \in [(j-1/2)\Delta, (j+1/2)\Delta)$. Since $|f_B^r[k]| \sim C k^{-s}$

$$n_j \sim C^{1/s} \Delta^{-1/s} (j-1/2)^{-1/s} - C^{1/s} \Delta^{-1/s} (j+1/2)^{-1/s} \,. \tag{11.56}$$

Together with (11.50) we get

$$p_j = \frac{n_j}{M} \sim (j-1/2)^{-1/s} - (j+1/2)^{-1/s}$$

so (11.55) proves that $\mathcal{H}_a \sim M$.

The value of s is not known a priori, but one may choose a variable length code optimized for $s = 1/2$ by setting

$$l_j = \log_2 \Big((j-1/2)^{-2} - (j+1/2)^{-2} \Big) \,.$$

We can verify that for all $s > 1/2$, the resulting bit budget satisfies

$$R_a = -M \sum_{j=1}^{+\infty} p_j l_j \sim M \sim \mathcal{H}_a \,.$$

As a result $R_1 = R_s + R_a \sim M$. Together with (11.54) this proves that

$$R = R_0 + R_1 \sim M \left(1 + \log_2 \frac{N^2}{M} \right) \sim \mathcal{H} \,, \tag{11.57}$$

with $\mathcal{H} = -N^2 \sum_{k=1}^{K} p_k \log_2 p_k$.

One can also verify that $R_0 + R_1 \geq 2M$ so that $R \leq N^2$ implies that $M \leq N^2/2$, which was used to prove that $d(R,f) \sim C^2 M^{1-2s}$. Inverting equation (11.57) gives

$$M \sim R \left(1 + \log_2 \frac{N^2}{R} \right)^{-1} \,,$$

and $d(R,f) \sim C^2 M^{1-2s}$ implies (11.48). ■

The equivalence sign $\sim$ means that lower and upper bounds of $d(R,f)$ and $d_{\mathcal{H}}(R,f)$ are obtained by multiplying the right expression of (11.48) by two constants $A, B > 0$ that are independent of C, R and N. It thus specifies the decay of $d(R,f)$ and $d_{\mathcal{H}}(R,f)$ as R increases. Theorem 11.4 proves that at low bit rates, the distortion is proportional to R^{1-2s}, as opposed to $2^{-2R/N^2}$ as in the high bit rate distortion

formula of Proposition 11.5. At low bit rates, to minimize the distortion one must find a basis $\mathcal{B}$ where the sorted coefficients of f have a fast decay that maximizes the value of s. The notion of minimax distortion rate and results obtained in optimized bases are studied in Section 11.4.5.

The theorem supposes that the transform code uses a uniform quantizer. When the high resolution quantization hypothesis holds, a uniform quantizer is optimal, but this is not the case at low bit rates. The proof shows that coefficients quantized to zero are mostly responsible for the behavior of the distortion rate. Modifying the size of other quantization bins has a marginal effect. The distortion rate equivalence (11.48) thus remains valid for a non-uniform optimal quantizer.

Adaptive Basis Choice For signals having complicated structures, the decay of sorted coefficients can be significantly improved by choosing the basis adaptively. Let $\mathcal{D} = \{\mathcal{B}^\lambda\}_{\lambda \in \Lambda}$ be a family of orthonormal bases $\mathcal{B}^\lambda = \{g_m^\lambda\}_{0 \leq m < N^2}$. The decay of sorted coefficients can be controlled indirectly by minimizing the number M^λ of significant coefficients (not quantized to zero). This number can be written as a cost function:

$$M^\lambda = \sum_{m=0}^{N^2-1} \Phi(\langle f, g_m^\lambda \rangle) \tag{11.58}$$

with

$$\Phi(x) = \begin{cases} 0 & \text{if } Q(x) = 0 \\ 1 & \text{if } Q(x) \neq 0 \end{cases} . \tag{11.59}$$

The proof of Theorem 11.4 shows that the bit budget R^λ of the transform coding in the basis $\mathcal{B}^\lambda$ is almost proportional to M^λ.

Dictionaries of wavelet packets and local cosine bases include more than $2^{N^2/4}$ different orthonormal bases. Since the cost function (11.58) is additive, the algorithm of Section 9.4.2 finds the best basis that minimizes M^λ with $O(N^2 \log_2 N)$ operations. It is also necessary to code which basis is selected [288, 269], because this basis depends on f. In wavelet packets and local cosine dictionaries, a constant size code requires more than $N^2/2$ bits. This overhead can more than offset the potential gain obtained by optimizing the basis choice. If we know the probability distribution of the bases that are chosen, then a variable length code reduces the average value of the overhead.

More flexible adaptive decompositions with matching pursuits can also improve the distortion rate [193]. Section 9.5.2 discusses matching pursuit algorithms, which decompose signals as a sum of vectors selected in a dictionary $\mathcal{D} = \{g_\gamma\}_{\gamma \in \Gamma}$ plus a residue:

$$f = \sum_{p=1}^{M} \alpha_p g_{\gamma_p} + R^M f .$$

The transform code neglects the residue $R^M f$, quantizes the coefficients α_p and records the indices $\gamma_p \in \Gamma$. Coding these indices is equivalent to storing a significance map defined by $b[\gamma_p] = 1$ for $1 \leq p \leq M$ and $b[\gamma] = 0$ for other $\gamma \in \Gamma$.

Choosing dictionaries that are larger than orthogonal bases results in a more precise approximation of f using M dictionary vectors but requires more bits to code the significance map, which is larger. Optimal dictionary sizes are therefore obtained through a trade-off between the approximation improvement and the bit rate increase of the significance map code. Efficient matching pursuit image codes have been designed to code the errors of motion compensation algorithms in video compression [279, 343].

11.4.2 Wavelet Image Coding

Implementation At low bit rates, a uniform quantizer does not minimize the distortion rate of a wavelet transform code. One can verify both numerically and mathematically [257] that doubling the size of the zero bin improves the distortion rate for large classes of images. This reduces the proportion of significant coefficients and thus improves the bit budget by a factor that is not offset by the increase of the quantization error. A larger zero bin increases the quantization error too much, degrading the overall distortion rate. The quantizer thus becomes

$$Q(x) = \begin{cases} 0 & \text{if } |x| < \Delta \\ \text{sign}(x)\,(\lfloor x/\Delta \rfloor + 1/2)\,\Delta & \text{if } |x| \geq \Delta \end{cases}. \tag{11.60}$$

The histogram of wavelet coefficients often has a slower decay for $|x| \geq \Delta$ than for $|x| > \Delta$, as shown by Figure 11.8. The high resolution quantization hypothesis is approximately valid for $|x| \geq \Delta$ with $1/8 \leq \bar{R} \leq 1$, which means that a uniform quantizer is nearly optimal.

Figure 11.9 displays several significance maps of wavelet coefficients. There are more significant coefficients at larger scales 2^j because wavelet coefficients have a tendency to decrease across scales. This is further explained in Section 11.4.4. The compression package of Davis in LastWave (Appendix B.2) encodes these significance maps and the values of non-zero quantized coefficients with adaptive arithmetic codes. Each subimage of wavelet coefficients is scanned in zig-zag order, and all wavelet subimages are successively visited from large to fine scales, as illustrated by Figure 11.10. An adaptive arithmetic code takes advantage of the fact that higher amplitude wavelet coefficients tend to be located at large scales. The resulting code depends on the distribution of wavelet coefficients at each scale.

Visual distortions introduced by quantization errors of wavelet coefficients depend on the scale 2^j. Errors at large scales are more visible than at fine scales [347]. This can be taken into account by quantizing the wavelet coefficients with intervals $\Delta_j = \Delta w_j$ that depend on the scale 2^j. For $\bar{R} \leq 1$ bit/pixel, $w_j = 2^{-j}$ is appropriate for the three finest scales. As shown in (11.31), choosing such weights is equivalent to minimizing a weighted mean-square error. For simplicity, in the following we set $w_j = 1$.

Bounded Variation Images Section 2.3.3 explains that large classes of images have a bounded total variation because the average length of their contours is

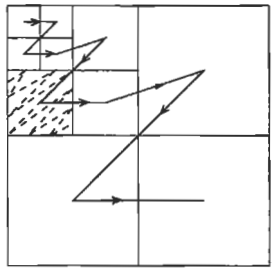


FIGURE 11.10 Each binary significance map of wavelet coefficients is scanned in zig-zag order, illustrated with a dotted line. All wavelet subimages are successively visited from coarse to fine scales in the order indicated by the arrows.

bounded independent of the resolution N. The total variation norm $\|f\|_V$ is related to the wavelet coefficients of f by an upper bound (10.120) and a lower bound (10.121). If f has discontinuities along edges, then its sorted wavelet coefficients $|f_B^r[k]|$ satisfy

$$|f_B^r[k]| \sim N \|f\|_V k^{-1} .$$

This decay property is verified by the wavelet coefficients of the Lena and Boat images. We derive from Theorem 11.4 that if $\bar{R} = R/N^2 \leq 1$ then

$$d(\bar{R}, f) \sim d_{\mathcal{H}}(\bar{R}, f) \sim \|f\|_V^2 \bar{R}^{-1} (1 - \log_2 \bar{R}) . \tag{11.61}$$

For general bounded variation images, Section 11.4.5 proves that the decay $\bar{R}^{-1}$ cannot be improved by any other signal coder. In that sense, a wavelet transform coding is optimal for bounded variation images. The resulting PSNR is

$$PSNR(\bar{R}, f) \approx 10 \log_{10} 2 \left[\log_2 \bar{R} - \log_2 \left(1 - \log_2 \bar{R} \right) - K \right] , \tag{11.62}$$

where K is a constant. Figure 11.11(a) shows the PSNR computed numerically from the wavelet transform code of the Lena and Boat images. As expected from (11.62), the PSNR increases almost linearly as a function of $\log_{10} \bar{R}$, with a slope of $10 \log_{10} 2 \approx 3$ db/bit.

More Irregular Images Mandrill and GoldHill are examples of images that do not have a bounded variation. This appears in the fact that their sorted wavelet coefficients satisfy $|f_B^r[k]| \sim C k^{-s}$ for $s < 1$. The PSNR calculated from the distortion rate formula (11.48) is

$$PSNR(\bar{R}, f) \approx (2s - 1) 10 \log_{10} 2 \left[\log_2 \bar{R} - \log_{10} (1 - \log_2 \bar{R}) - K \right] ,$$

where K is a constant. For GoldHill, $s \approx 0.8$, so the PSNR increases by 1.8 db/bit. Mandrill is even more irregular, with $s \approx 2/3$, so at low bit rates $\bar{R} < 1/4$ the PSNR

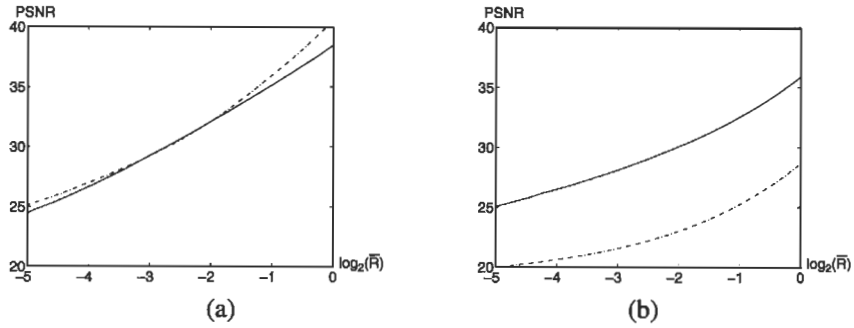


FIGURE 11.11 PSNR as a function of $\log_2(\bar{R})$. (a): Lena (solid line) and boats (dotted line) (b): GoldHill (solid line) and Mandrill (dotted line)

increases by only 1 db/bit. Such images can be modeled as elements of Besov spaces whose regularity index $s/2+1/2$ is smaller than 1. The distortion rate of transform coding in general Besov spaces is studied in [132].

For natural images, the competition organized for the JPEG-2000 image compression standard shows that wavelet image transform codes give the best distortion rate and best visual quality among all existing real time image coders. The adaptive arithmetic coder is quite efficient but not optimal. Section 11.4.4 shows that embedded wavelet transform codes produce a slightly larger PSNR, by better taking into account the distribution of large versus small wavelet coefficients across scales.

For specialized images such as fingerprints [103] or seismic images, other bases can outperform wavelets. This is particularly true when the image includes high frequency oscillatory textures, which create many significant fine scale wavelet coefficients. These images are better compressed in local cosine or wavelet packet bases, whose vectors approximate high frequency oscillations more efficiently. Block cosine bases used by the JPEG standard may also outperform wavelets for such images.

Choice of Wavelet To optimize the transform code one must choose a wavelet basis that produces as many zero quantized coefficients as possible. A two-dimensional separable wavelet basis is constructed from a one-dimensional wavelet basis generated by a mother wavelet ψ. Three criteria may influence the choice of ψ: number of vanishing moments, support size and regularity.

High amplitude coefficients occur when the supports of the wavelets overlap a brutal transition like an edge. The number of high amplitude wavelet coefficients created by an edge is proportional to the width of the wavelet support, which should thus be as small as possible. Over smooth regions, wavelet coefficients are small at fine scales if the wavelet has enough vanishing moments to take advantage of the image regularity. However, Proposition 7.4 shows that the support size of ψ

increases proportionally to the number of vanishing moments. The choice of an optimal wavelet is therefore a trade-off between the number of vanishing moments and support size. If the image is discontinuous then the wavelet choice does not modify the asymptotic behavior of the distortion rate (11.61) but it influences the multiplicative constant.

Wavelet regularity is important in reducing the visibility of artifacts. A quantization error adds to the image a wavelet multiplied by the amplitude of the quantized error. If the wavelet is irregular, the artifact is more visible because it looks like an edge or a texture element [81]. This is the case for Haar wavelets. Continuously differentiable wavelets produce errors that are less visible, but more regularity often does not improve visual quality.

To avoid creating large amplitude coefficients at the image border, it is best to use the folding technique of Section 7.5.2, which is much more efficient than the periodic extension of Section 7.5.1. However, it requires using wavelets that are symmetric or antisymmetric. Besides Haar, there is no symmetric or antisymmetric wavelet of compact support which generates an orthonormal basis. Biorthogonal wavelet bases that are nearly orthogonal can be constructed with symmetric or antisymmetric wavelets. They are therefore more often used for image compression.

Overall, many numerical studies have shown that the 7-9 biorthogonal wavelets of Figure 7.15 give the best distortion rate performance for wavelet image transform codes. They provide an appropriate trade-off between the vanishing moments, support and regularity requirements. This biorthogonal wavelet basis is nearly orthogonal and thus introduces no numerical instability. The compression examples of Figure 11.6 are calculated in this basis.

Geometric Regularity For the set of all images having a total variation bounded by a constant C, Section 11.4.5 proves that a wavelet transform coding has a distortion rate that is close to the minimax lower bound. The total variation of an image is equal to the average length of its level sets, which may be highly irregular curves. If we only consider images having level sets that are piecewise regular curves then one can improve the distortion rate of a wavelet transform coding, by taking into account the geometric regularity. This is case for the Lena and Boats images.

The inefficiency of a wavelet transform code for Lena and Boat appears in the significance maps of Figure 11.9. They are coded with a zig-zag scanning that does not take advantage of the geometric regularity of edges where significant coefficients are located. The use of three oriented wavelets translated on dyadic grids also partially destroys the geometry.

Using edges for image coding was originally proposed by Carlsson [115]. Figure 11.12 shows the larger amplitude wavelet maxima at the finest scales, calculated with the algorithm of Section 6.3.1. A compact code has been designed [261, 186] to use the geometric regularity of these maxima chains and the slow variation of wavelet coefficients along these chains. Other edge coding strategies have

FIGURE 11.12 Edge chains calculated from the larger amplitude wavelet transform modulus maxima at the finest scale.

also been implemented [121, 243]. However, these geometry-oriented codes are mathematically not well understood, and whether they can significantly improve the distortion rate for interesting classes of images has yet to be determined.

11.4.3 Block Cosine Image Coding

The JPEG image compression standard [345] is a transform coding in a block cosine-I basis. Theorem 8.7 proves that the following cosine-I family is an orthogonal basis of an image block of L by L pixels:

$$\left\{ g_{k,j}[n,m] = \lambda_k \lambda_j \frac{2}{L} \cos\left[\frac{k\pi}{L}\left(n+\frac{1}{2}\right)\right] \cos\left[\frac{j\pi}{L}\left(m+\frac{1}{2}\right)\right] \right\}_{0 \le k,j < L} \tag{11.63}$$

with

$$\lambda_p = \begin{cases} 1/\sqrt{2} & \text{if } p = 0 \\ 1 & \text{otherwise} \end{cases} . \tag{11.64}$$

In the JPEG standard, images of N^2 pixels are divided in $N^2/64$ blocks of 8 by 8 pixels. Each image block is expanded in this separable cosine basis with a fast separable DCT-I transform.

JPEG quantizes the block cosine coefficients uniformly. In each block of 64 pixels, a significance map gives the position of zero versus non-zero quantized coefficients. Lower frequency coefficients are located in the upper right of each block, whereas high frequency coefficients are in the lower right, as illustrated in Figure 11.13. Many image blocks have significant coefficients only at low frequencies and thus in the upper left of each block. To take advantage of this prior knowledge, JPEG codes the significance map with a run-length code. Each block

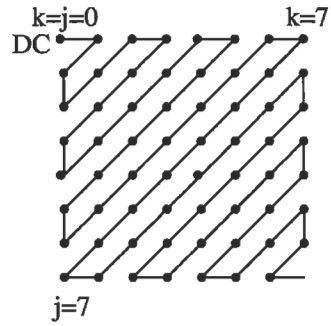


FIGURE 11.13 A block of 64 cosine coefficients has the zero frequency (DC) coefficient at the upper left. The run-length makes a zig-zag scan from low to high frequencies.

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	108	103	77
24	35	55	64	81	194	113	92
49	64	78	87	103	121	120	101
72	92	95	98	121	100	103	99

Table 11.1 Matrix of weights $w_{k,j}$ used to quantize the block cosine coefficient corresponding to each cosine vector $g_{k,j}$ [74]. The order is the same as in Figure 11.13.

of 64 coefficients is scanned in zig-zag order as indicated in Figure 11.13. In this scanning order, JPEG registers the size of the successive runs of coefficients quantized to zero, which are efficiently coded together with the values of the following non-zero quantized coefficients. Insignificant high frequency coefficients often produce a long sequence of zeros at the end of the block, which is coded with an End Of Block (EOB) symbol.

In each block i, there is one cosine vector $g^i_{0,0}[n,m]$ of frequency zero, which is equal to $1/8$ over the block and 0 outside. The inner product $\langle f, g^i_{0,0}\rangle$ is proportional to the average of the image over the block. Let $DC^i = Q(\langle f, g^i_{0,0}\rangle)$ be the quantized zero-frequency coefficient. Since the blocks are small, the averages are often close for adjacent blocks, and JPEG codes the differences $DC^i - DC^{i-1}$.

Weighted Quantization Our visual sensitivity depends on the frequency of the image content. We are typically less sensitive to high frequency oscillatory patterns than to low frequency variations. To minimize the visual degradation of the coded images, JPEG performs a quantization with intervals that are proportional to weights specified in a table, whose values are not imposed by the standard. This is equivalent to optimizing a weighted mean-square error (11.31). Table 11.1 is an example of an 8 by 8 weight matrix that is used in JPEG [345]. The weights at the lowest frequencies, corresponding to the upper left of Table 11.1, are roughly 10 times smaller than at the highest frequencies, corresponding to the bottom right.

Distortion Rate At 0.25-0.5 bit/pixel, the quality of JPEG images is moderate. At 0.2 bit/pixel, Figure 11.14 shows that there are blocking effects due to the discontinuities of the square windows. At 0.75-1 bit/pixel, images coded with the JPEG standard are of excellent quality. Above 1 bit/pixel, the visual image quality is perfect. The JPEG standard is often used for $\bar{R} \in [0.5, 1]$.

At low bit rates, the artifacts at the block borders are reduced by replacing the block cosine basis by a local cosine basis [42, 80], designed in Section 8.4.4. If the image is smooth over a block, a local cosine basis creates lower amplitude high frequency coefficients, which slightly improves the coder performance. The quantization errors for smoothly overlapping windows also produce more regular grey level image fluctuations at the block borders. However, the improvement has not been significant enough to motivate replacing the block cosine basis by a local cosine basis in the JPEG standard.

Figure 11.15 compares the PSNR of JPEG and of the wavelet transform code for two images. The wavelet transform code gives an improvement of approximately 2-3db. For $\bar{R} \leq 2^{-4}$ the performance of JPEG deteriorates because it needs to keep at least $N^2/64$ zero-frequency coefficients in order to recover an estimate of image intensity everywhere.

Implementation of JPEG The baseline JPEG standard [345] uses an intermediate representation that combines run-length and amplitude values. In each block, the 63 (non-zero frequency) quantized coefficients indicated in Figure 11.13 are integers that are scanned in zig-zag order. A JPEG code is a succession of symbols $S_1 = (L, B)$ of eight bits followed by symbols S_2. The L variable is the length of a consecutive run of zeros, coded on four bits. Its value is thus limited to the interval $[0, 15]$. Actual zero-runs can have a length greater than 15. The symbol $S_1 = (15, 0)$ is interpreted as a run-length of size 16 followed by another run-length. When the run of zeros includes the last 63^{rd} coefficient of the block, a special End Of Block symbol $S_1 = (0, 0)$ is used, which terminates the coding of the block. For high compression rates, the last run of zeros may be very long. The EOB symbol stops the coding at the beginning of this last run of zeros.

The B variable of S_1 is coded on four bits and gives the number of bits used to code the value of the next non-zero coefficient. Since the image grey level values are in the interval $[0, 2^8]$, one can verify that the amplitude of the block cosine

FIGURE 11.14 Image compression with JPEG.

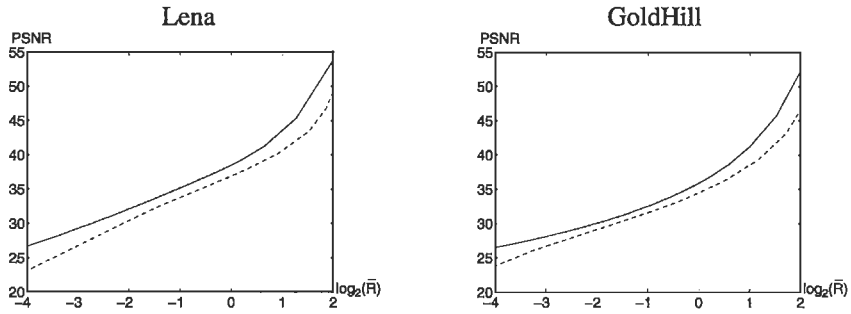


FIGURE 11.15 Comparison of the PSNR obtained with JPEG (dotted line) and the wavelet transform code (solid line) for Lena and GoldHill.

B	Range of values
1	$-1,1$
2	$-3,-2,2,3$
3	$-7 \ldots -4,4 \ldots 7$
4	$-15 \ldots -8,8 \ldots 15$
5	$-31 \ldots -16,16 \ldots 31$
6	$-63 \ldots -32,32 \ldots 63$
7	$-127 \ldots -64,64 \ldots 127$
8	$-255 \ldots -128,128 \ldots 255$
9	$-511 \ldots -256,256 \ldots 511$
10	$-1023 \ldots -512,512 \ldots 1023$

Table 11.2 The value of coefficients coded on B bits belongs to a set of 2^B values that is indicated in the second column.

coefficients remains in $[-2^{10}, 2^{10}-1]$. For any integers in this interval, Table 11.2 gives the number of bits used by the code. For example, 70 is coded on $B = 7$ bits. There are 2^7 different numbers that are coded with seven bits. If B is non-zero, after the symbol S_1 the symbol S_2 of length B specifies the amplitude of the following non-zero coefficient. This variable length code is a simplified entropy code. High amplitude coefficients appear less often and are thus coded with more bits.

For DC coefficients (zero frequency), the differential values $DC^i - DC^{i-1}$ remain in the interval $[-2^{11}, 2^{11}-1]$. They are also coded with a succession of two symbols. In this case, S_1 is reduced to the variable B which gives the number of bits of the next symbol S_2 which codes $DC^i - DC^{i-1}$.

For both the DC and the other coefficients, the S_1 symbols are encoded with a Huffman entropy code. JPEG does not impose the Huffman tables, which may vary depending on the type of image. An arithmetic entropy code can also be used. For coefficients that are not zero frequency, the L and the B variables are lumped

together because their values are often correlated, and the entropy code of S_1 takes advantage of this correlation.

11.4.4 Embedded Transform Coding

For rapid transmission or fast image browsing from a data base, a coarse signal approximation should be quickly provided, and progressively enhanced as more bits are transmitted. Embedded coders offer this flexibility by grouping the bits in order of significance. The decomposition coefficients are sorted and the first bits of the largest coefficients are sent first. An image approximation can be reconstructed at any time, from the bits already transmitted.

Embedded coders can take advantage of any prior information about the location of large versus small coefficients. Such prior information is available for natural images decomposed on wavelet bases. As a result, an implementation with zero-trees designed by Shapiro [307] yields better compression rates than classical wavelet transform coders.

Embedding The decomposition coefficients $f_B[m] = \langle f, g_m \rangle$ are partially ordered by grouping them in index sets $\mathcal{S}_k$ defined for any $k \in \mathbb{Z}$ by

$$\mathcal{S}_k = \{m \ : \ 2^k \leq |f_B[m]| < 2^{k+1}\}.$$

The set $\mathcal{S}_k$ is coded with a binary significance map $b_k[m]$:

$$b_k[m] = \begin{cases} 0 & \text{if } m \notin \mathcal{S}_k \\ 1 & \text{if } m \in \mathcal{S}_k \end{cases} . \tag{11.65}$$

An embedded algorithm quantizes $f_B[m]$ uniformly with a quantization step (bin size) $\Delta = 2^n$ that is progressively reduced. Let $m \in \mathcal{S}_k$ with $k \geq n$. The amplitude $|Q(f_B[m])|$ of the quantized number is represented in base 2 by a binary string with non-zero digits between the bit k and the bit n. The bit k is necessarily 1 because $2^k \leq |Q(f_B[m])| < 2^{k+1}$. Hence, $k-n$ bits are sufficient to specify this amplitude, to which is added one bit for the sign.

The embedded coding is initiated with the largest quantization step that produces at least one non-zero quantized coefficient. In the loop, to refine the quantization step from 2^{n+1} to 2^n, the algorithm records the significance map $b_n[m]$ and the sign of $f_B[m]$ for $m \in \mathcal{S}_n$. This can be done by directly recording the sign of significant coefficients with a variable incorporated in the significance map $b_n[m]$. Afterwards, the code stores the bit n of all amplitudes $|Q(f_B[m])|$ for $m \in \mathcal{S}_k$ with $k > n$. If necessary, the coding precision is improved by decreasing n and continuing the encoding. The different steps of the algorithm can be summarized as follows [301]:

1. *Initialization* Store the index n of the first non-empty set $\mathcal{S}_n$

$$n = \left\lfloor \sup_m \log_2 |f_B[m]| \right\rfloor . \tag{11.66}$$

2. *Significance map* Store the significance map $b_n[m]$ and the sign of $f_B[m]$ for $m \in \mathcal{S}_n$.
3. *Quantization refinement* Store the n^{th} bit of all coefficients $|f_B[m]| > 2^{n+1}$. These are coefficients that belong to some set $\mathcal{S}_k$ for $k > n$, whose coordinates were already stored. Their n^{th} bit is stored in the order in which their position was recorded in the previous passes.
4. *Precision refinement* Decrease n by 1 and go to step 2.

Distortion Rate This algorithm may be stopped at any time in the loop, providing a code for any specified number of bits. The distortion rate is analyzed when the algorithm is stopped at the step 4. All coefficients above $\Delta = 2^n$ are uniformly quantized with a bin size $\Delta = 2^n$. The zero quantization bin $[-\Delta, \Delta]$ is therefore twice as big as the other quantization bins, which was shown to be efficient for wavelet image coders.

Once the algorithm stops, we denote by M the number of significant coefficients above $\Delta = 2^n$. The total number of bits of the embedded code is

$$R = R_0^e + R_1^e,$$

where R_0^e is the number of bits needed to code all significance maps $b_k[m]$ for $k \geq n$, and R_1^e the number of bits used to code the amplitude of the quantized significant coefficients $Q(f_B[m])$, knowing that $m \in \mathcal{S}_k$ for $k > n$.

To appreciate the efficiency of this embedding strategy, let us compare the bit budget $R_0^e + R_1^e$ to the number of bits $R_0 + R_1$ used by the direct transform code of Section 11.4.1. The value R_0 is the number of bits that code the overall significance map

$$b[m] = \begin{cases} 0 & \text{if } |f_B[m]| \leq \Delta \\ 1 & \text{if } |f_B[m]| > \Delta \end{cases} \tag{11.67}$$

and R_1 is the number of bits that code the quantized significant coefficients.

An embedded strategy codes $Q(f_B[m])$ knowing that $m \in \mathcal{S}_k$ and hence that $2^k \leq |Q(f_B[m])| < 2^{k+1}$, whereas a direct transform code knows only that $|Q(f_B[m])| > \Delta = 2^n$. Fewer bits are thus needed for embedded codes:

$$R_1^e \leq R_1. \tag{11.68}$$

However, this improvement may be offset by the supplement of bits needed to code the significance maps $\{b_k[m]\}_{k>n}$ of the sets $\{\mathcal{S}_k\}_{k>n}$. A direct transform code records a single significance map $b[m]$, which specifies $\cup_{k \geq n} \mathcal{S}_k$. It provides less information and is therefore coded with fewer bits:

$$R_0^e \geq R_0. \tag{11.69}$$

An embedded code brings an improvement over a direct transform code if

$$R_0^e + R_1^e \leq R_0 + R_1.$$

This can happen if we have some prior information about the position of large coefficients $|f_B[m]|$ versus smaller ones. This allows us to reduce the number of bits needed to encode the partial sorting of all coefficients provided by the significance maps $\{b_k[m]\}_{k>n}$. The use of such prior information produces an overhead of R_0^e relative to R_0 that is smaller than the gain of R_1^e relative to R_1. This is the case both for embedded transform codes implemented in wavelet bases and for the block cosine I basis used by JPEG [358].

Wavelet Embedded Code Wavelet coefficients have a large amplitude where the signal has sharp transitions. If an image f is uniformly Lipschitz α in the neighborhood of (x_0, y_0), then (6.62) proves that for wavelets $\psi^l_{j,p,q}$ located in this neighborhood there exists $A \geq 0$ such that

$$|\langle f, \psi^l_{j,p,q} \rangle| \leq A 2^{j(\alpha+1)}.$$

The worst singularities are often discontinuities, so $\alpha \geq 0$. This means that in the neighborhood of singularities without oscillations, the amplitude of wavelet coefficients decreases when the scale 2^j decreases. This property is not valid for oscillatory patterns. High frequency oscillations create coefficients at large scales 2^j that are typically smaller than at the fine scale which matches the period of oscillation. We thus consider images where such oscillatory patterns are relatively rare.

Wavelet zero-trees introduced by Lewis and Knowles [250] take advantage of the decay of wavelet coefficients by relating these coefficients across scales with quad-trees. Shapiro [307] used this zero-tree structure to code the embedded significance maps of wavelet coefficients. The numerical examples are computed with the algorithm of Said and Pearlman [301], which improves Shapiro's zero-tree code with a set partitioning technique.

Good visual quality images are obtained in Figure 11.16 with 0.2 bit/pixel, which considerably improves the JPEG compression results shown in Figure 11.14. At 0.05 bit/pixel the wavelet embedded code recovers a decent approximation, which is not possible with JPEG. Figure 11.17 compares the PSNR of the wavelet embedded code with the PSNR of the direct wavelet transform code described in Section 11.4.2. For any quantization step both transform codes yield the same distortion but the embedded code reduces the bit budget:

$$R_0^e + R_1^e \leq R_0 + R_1.$$

As a consequence the PSNR curve of the embedded code is a translation to the left of the PSNR of the direct transform code. For a set $\mathcal{S}_V$ of bounded variation images, one can show that the zero-tree algorithm can at most reduce the bit budget by a constant. However, for particular classes of images where the significant coefficients are well aligned across scales, the $\log_2(N^2/R)$ term that appears in the distortion rate (11.61) can disappear [132].

FIGURE 11.16 Embedded wavelet transform coding.

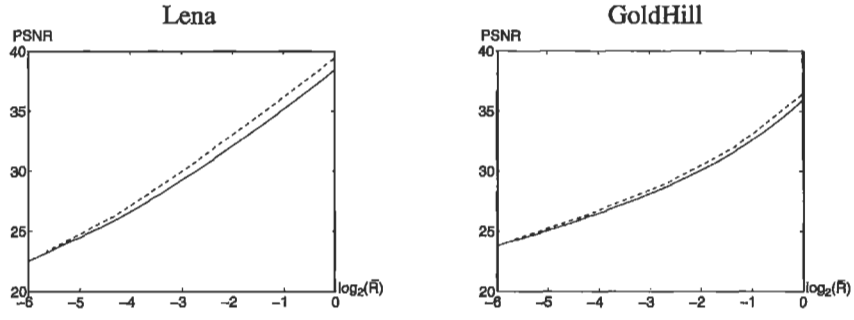


FIGURE 11.17 Comparison of the PSNR obtained with an embedded wavelet transform code (dotted line) and a direct wavelet transform code (solid line).

Zero-Tree Implementation The significance maps of a wavelet embedded code are stored in zero-trees with an algorithm introduced by Shapiro [307]. These zero-trees also incorporate the sign of non-zero coefficients, which is therefore not coded with the amplitude. The signed significance map of a set $\mathcal{S}_n$ has the same structure as an array of wavelet coefficients defined for $l = 1,2,3$ by

$$b_j^l[p,q] = \begin{cases} 1 & \text{if } 2^n \leq \langle f, \psi_{j,p,q}^l \rangle < 2^{n+1} \\ -1 & \text{if } -2^{n+1} < \langle f, \psi_{j,p,q}^l \rangle \leq -2^n \\ 0 & \text{otherwise} \end{cases} . \tag{11.70}$$

At the largest scale 2^J, there is also a significance map $b_J^0[p,q]$ computed from the scaling coefficients.

Wavelet zero-trees encode these significance maps with *quad-trees*. For each orientation $l = 1,2,3$, we create quad-trees by recursively relating each coefficient $b_j^l[p,q]$ to the following four children at the next finer scale 2^{j-1}:

$$b_{j-1}^l[2p,2q] \ , \ b_{j-1}^l[2p+1,2q] \ , \ b_{j-1}^l[2p,2q+1] \ , \ b_{j-1}^l[2p+1,2q+1].$$

The values of a wavelet coefficient and its four children depend on the variations of the image grey level in the same spatial area. At the largest scale 2^J, the children of $b_J^0[p,q]$ are defined to be the three wavelet coefficients at the same scale and location: $b_J^1[p,q]$, $b_J^2[p,q]$ and $b_J^3[p,q]$. The construction of these trees is illustrated in Figure 11.18.

If $b_j^l[p,q] = 0$, we say that this coefficient belongs to a zero-tree if all its descendants in the quad-tree are also zero. This happens if its descendants have wavelet coefficients of smaller amplitude, which is likely. The position of all the zero values inside a zero-tree are specified by the position (p,q), orientation l and scale 2^j of the *zero-tree root*, which is labeled by R. This encoding is particularly effective if R is located at a large scale because the zero-tree includes more zero coefficients. If $b_j^l[p,q] = 0$ but one of its descendants in the quad-tree is non-zero, then this coefficient is called an *isolated zero*, denoted by I. The coefficients

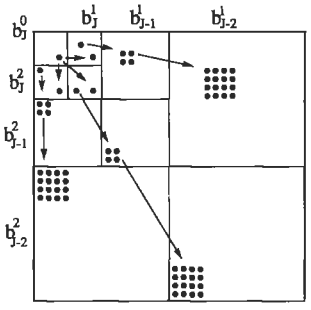


FIGURE 11.18 At the coarsest scale 2^J, the children of each pixel in b_J^0 are the three pixels of the same position in b_J^1, b_J^2 and b_J^3. At all other scales, in each direction l, quad-trees are constructed by relating a pixel of b_j^l to its four children in b_{j-1}^l.

$b_j^l[p,q] = 1$ or $b_j^l[p,q] = -1$ are represented respectively by the symbols P (positive) and N (negative). The wavelet table of symbols (R,I,P,N) corresponding to the significance map (11.70) is scanned in the zig-zag order illustrated in Figure 11.10. The resulting sequence of symbols is then coded with an adaptive arithmetic code [307].

Let us mention that zero-tree wavelet encoding are closely related to fractal compression algorithms implemented with Iterated Function Systems [7]. Davis [150] shows that these wavelet embedded codes bring significant improvements over existing fractal codes [84].

Example 11.2 Figure 11.19 gives an example of wavelet coefficients for an image of 64 pixels [307]. The amplitude of the largest coefficient is 63 so $\mathcal{S}_5$ is the first non-empty set $\mathcal{S}_n$. The coefficients in $\mathcal{S}_5$ have an amplitude included in $[32, 64)$. The array of symbols is on the right of Figure 11.19. The dots correspond to coefficients inside a zero-tree. A zig-zag scanning in the order indicated in Figure 11.10 yields the sequence: P,N,I,R,P,R,R,R,R,I,R,R,I,P,I,I.

11.4.5 Minimax Distortion Rate [3]

A compression algorithm is not optimized for a single signal f but for a whole class. The coder must be adapted to the prior information available about these signals. We rarely have a probabilistic model of complex signals such as images, but we can define a prior set Θ that includes our signal class. The model Θ is a set of functions in $\mathbf{L}^2[0,1]^2$, which are approximated and discretized by the coder. To control the coder distortion for all signals in Θ we want to minimize the maximum

63	-34	49	10	7	13	-12	7
-31	23	14	-13	3	4	6	-1
15	14	3	-12	5	-7	3	9
-9	-7	-14	8	4	-2	3	2
-5	9	-1	47	4	6	-2	2
3	0	-3	2	3	-2	0	4
2	-3	6	-4	3	6	3	6
5	11	5	6	0	3	-4	4

P	N	P	R	•	•	•	•
I	R	R	R	•	•	•	•
R	I	•	•	•	•	•	•
R	R	•	•	•	•	•	•
•	•	I	P	•	•	•	•
•	•	I	I	•	•	•	•
•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•

FIGURE 11.19 The left table is an array of 64 wavelet coefficients. The set $\mathcal{S}_5$ corresponding to coefficients in $[32, 64)$ has a significance map whose zero-tree symbols are shown on the right.

distortion over Θ:

$$d(R, \Theta) = \sup_{f \in \Theta} d(R, f) .$$

The definition of a minimax distortion rate is closely related to Kolmogorov ϵ-entropy [174]. As in the estimation problems of Chapter 10, it is necessary to approach this minimax distortion rate with a signal coder that is fast and simple to implement. If the basis provides a sparse representation of signals in Θ then a transform code can be nearly optimal among all possible coders.

Kolmogorov ϵ-Entropy In its most general form, a coding algorithm is specified by an operator D that approximates any $f \in \Theta$ by $\tilde{f} = Df$ which belongs to the *approximation net*

$$\Theta_D = \{\tilde{f} \, : \, \exists f \in \Theta \, , \, \tilde{f} = Df\} .$$

Let $\text{Card}(\Theta_D)$ be the cardinal of the net Θ_D, i.e., the number of elements in this set. The number of bits required to specify each coded signal $\tilde{f}$ is

$$R = \lceil \log_2 \text{Card}(\Theta_D) \rceil .$$

The maximum distortion over Θ is

$$d_D(\Theta) = \sup_{f \in \Theta} \|f - Df\| .$$

Let $\mathcal{O}_\epsilon$ be the set of all coders D such that $d_D(\Theta) \leq \epsilon$. An optimal coder $D \in \mathcal{O}_\epsilon$ has an approximation net Θ_D of minimum size. The corresponding *Kolmogorov ϵ-entropy* [174] is defined by

$$\mathcal{H}_\epsilon(\Theta) = \log_2 \Big(\min_{D \in \mathcal{O}_\epsilon} \text{Card}(\Theta_D) \Big) . \tag{11.71}$$

The best coder $D \in \mathcal{O}_\epsilon$ has a bit budget $R = \lceil \mathcal{H}_\epsilon(\Theta) \rceil$. For sets Θ of functions in infinite dimensional spaces, few cases are known where the Kolmogorov ϵ-entropy can be computed exactly [174].

For a fixed bit budget R, to minimize the distortion we consider the set of all coders $D \in \mathcal{C}_R$ whose approximation net has a size $\text{Card}(\Theta_D) \leq 2^R$. The *minimax distortion rate* is defined by

$$d_{\min}(R, \Theta) = \inf_{D \in \mathcal{C}_R} d_D(\Theta) \,. \tag{11.72}$$

When R varies, the distortion rate of a particular family of coders $D_R \in \mathcal{C}_R$ is defined by

$$\forall R > 0 \;,\; d(R, \Theta) = d_{D_R}(\Theta) \,.$$

Transform coders are examples where R depends upon the size of the quantization bins. The *decay exponent* of $d(R, \Theta)$ for large R is

$$\beta(\Theta) = \sup \{\beta \;:\; \exists \lambda > 0 \,,\, d(R, \Theta) \leq \lambda R^{-\beta}\} \,.$$

For the minimax distortion rate $d(R, \Theta) = d_{\min}(R, \Theta)$, the decay exponent $\beta(\Theta) = \beta_{\max}(\Theta)$ is maximum. In practice, we need to find coders that are simple to implement, and such that $d(R, \Theta)$ has a decay exponent close to or equal to $\beta_{\max}(\Theta)$.

Transform Coding Let Θ be a set of images in $\mathbf{L}^2[0,1]^2$. A transform coding in an orthonormal basis $\mathcal{B} = \{g_m\}_{m \in \mathbb{N}}$ sets to zero all coefficients for $m \geq N^2$ and quantizes all the others with R bits:

$$\tilde{f} = Df = \sum_{m=0}^{N^2-1} Q(f_{\mathcal{B}}[m])\, g_m \,. \tag{11.73}$$

We suppose that Q is a uniform quantizer, and the quantized coefficients $Q(f_{\mathcal{B}}[m])$ are coded with a variable length code. The distortion can be decomposed as a sum of two errors:

$$d(R, f) = d_N(R, f) + \epsilon_l(N^2, f) \,,$$

where

$$d_N(R, f) = \sum_{m=0}^{N^2-1} |f_{\mathcal{B}}[m] - Q(f_{\mathcal{B}}[m])|^2$$

is the distortion rate of the transform coding on the first N^2 coefficients and

$$\epsilon_l(N^2, f) = \sum_{m=N^2}^{+\infty} |f_{\mathcal{B}}[m]|^2$$

is the linear approximation error produced by setting to zero all coefficients $\langle f, g_m \rangle$ for $m \geq N^2$.

In practice, it is the camera which restricts the signal to its lower frequencies, and sets to zero all inner products $\langle f, g_m \rangle$ for $m \geq N^2$. The bit budget R is generally

adjusted so that the coding distortion is larger than the error introduced by the camera:

$$d_N(R,f) \geq \epsilon_l(N^2,f) .$$

To control $\epsilon_l(N^2,f)$, we suppose that Θ has a *compact tail*, which means that there exists $B>0$ and $\alpha>1/2$ such that

$$\forall f \in \Theta \ , \ |\langle f, g_m \rangle| \leq B m^{-\alpha} .$$

This implies that $\epsilon_l(N^2,f) = O(B^2 N^{-\gamma})$ with $\gamma = 2(2\alpha-1) > 0$.

Theorem 11.4 proves that the distortion rate of a transform code depends essentially on how the sorted coefficients decay. Let $f^r_{\mathcal{B}}[k]$ be the decomposition coefficient of f in $\mathcal{B}$ whose amplitude has rank k. The decay of signal coefficients over a set Θ is characterized by

$$s(\Theta) = \sup\{s \ : \ \exists C>0 \ \forall f \in \Theta \ , \ |f^r_{\mathcal{B}}[k]| \leq C k^{-s}\} .$$

The following theorem proves that a transform coding is asymptotically optimal over orthosymmetric sets, in the sense that the distortion rate has an optimal decay exponent. We recall from Section 10.3.2 that a set Θ is *orthosymmetric* in $\mathcal{B} = \{g_m\}_{m\in\mathbb{N}}$ if for all $f \in \Theta$ and all $|a[m]| \leq 1$ then

$$\sum_{m=0}^{+\infty} a[m] \, f_{\mathcal{B}}[m] \, g_m \in \Theta .$$

Theorem 11.5 (DONOHO) *Let Θ be an orthosymmetric set in $\mathcal{B}$, with a compact tail. If $s(\Theta) > 1/2$ then*

$$\beta_{\max}(\Theta) = 2s(\Theta) - 1 . \tag{11.74}$$

The decay exponent $\beta_{\max}(\Theta)$ is reached by a transform coding in $\mathcal{B}$ with a uniform quantizer and an adaptive variable length code.

Proof[3]. Since Θ has a compact tail, there exist B and $\alpha > 1/2$ such that $\Theta \subset \Lambda_{B,\alpha}$ with

$$\Lambda_{B,\alpha} = \left\{ f \ : \ |f_{\mathcal{B}}[m]| \leq B m^{-\alpha} \right\} .$$

Moreover, for any $s < s(\Theta)$ there exists $C>0$ such that $\Theta \subset \Theta_{C,s}$ with

$$\Theta_{C,s} = \left\{ f \ : \ |f^r_{\mathcal{B}}[k]| \leq C k^{-s} \right\} . \tag{11.75}$$

So $\Theta \subset \Theta_{C,s} \cap \Lambda_{B,\alpha}$. The following lemma computes the distortion rate of a transform coding in $\mathcal{B}$.

Lemma 11.2 *If $s > \alpha > 1/2$ then for large R a transform coding in $\mathcal{B}$ with a uniform quantizer and an optimized variable length code satisfies*

$$d(R, \Theta_{C,s} \cap \Lambda_{B,\alpha}) \sim C^2 \left(\frac{\log_2 R}{R} \right)^{2s-1} . \tag{11.76}$$

The main steps of the proof are given without details. First, we verify that

$$\sup_{f\in\Theta_{C,s}\cap\Lambda_{B,\alpha}} d(R,f) \sim \sup_{f\in\Theta_{C,s}} d_N(R,f) + \sup_{f\in\Lambda_{B,\alpha}} \epsilon_l(N^2,f) .$$

The same derivations as in Theorem 11.4 show that an optimized transform coding has a distortion rate such that

$$\sup_{f\in\Theta_{C,s}} d_N(R,f) \sim C^2 R^{1-2s} \left(1+\log_2 \frac{N^2}{R}\right)^{2s-1} .$$

Moreover,

$$\sup_{f\in\Lambda_{B,\alpha}} \epsilon_l(N^2,f) \sim B^2 N^{2(1-2\alpha)} .$$

Hence

$$\sup_{f\in\Theta_{C,s}\cap\Lambda_{B,\alpha}} d(R,f) \sim C^2 R^{1-2s} \left(1+\log_2 \frac{N^2}{R}\right)^{2s-1} + B^2 N^{2(1-2\alpha)} .$$

The values of R and N are adjusted to reach the minimum. If $B^2 N^{2(1-2\alpha)} = C^2 R^{1-2s}$ the minimum is nearly reached and

$$d(R,\Theta_{C,s}\cap\Lambda_{B,\alpha}) \sim C^2 \left(\frac{R}{\log_2 R}\right)^{1-2s} .$$

This lemma proves that the distortion of a transform coding in $\mathcal{B}$ satisfies

$$\sup\{\beta \,:\, \exists \lambda > 0 \; d(R,\Theta_{C,s}\cap\Lambda_{B,\alpha}) \leq \lambda R^{-\beta}\} = 2s-1 .$$

For any $s < s(\Theta)$, since $\Theta \subset \Theta_{C,s}\cap\Lambda_{B,\alpha}$, it follows that

$$d_{\min}(R,\Theta) \leq d(R,\Theta) \leq d(R,\Theta_{C,s}\cap\Lambda_{B,\alpha}) ,$$

and hence that $\beta_{\max}(\Theta) \geq 2s-1$. Increasing s up to $s(\Theta)$ proves that $\beta_{\max}(\Theta) \geq 2s(\Theta)-1$.

The proof that $\beta_{\max}(\Theta) \leq 2s(\Theta)-1$ requires computing a lower bound of the Kolmogorov ϵ-entropy over orthosymmetric sets. This difficult part of the proof is given in [164]. ■

Bounded Variation Images The total variation $\|f\|_V$ of $f \in \mathbf{L}^2[0,1]^2$ is defined in (2.65). We consider a set of bounded variation images that have a bounded amplitude:

$$\Theta_{V,\infty} = \{f \in \mathbf{L}^2[0,1]^2 \,:\, \|f\|_V \leq C \text{ and } \|f\|_\infty \leq C\} .$$

The following proposition computes the distortion rate of a wavelet transform coding and proves that its decay is optimal.

Proposition 11.6 *In a separable wavelet basis, for large R the distortion rate of a transform coding with a uniform quantization satisfies*

$$d(R, \Theta_{V,\infty}) \sim C^2 \frac{\log_2 R}{R} . \tag{11.77}$$

The resulting decay exponent is equal to the minimax decay exponent

$$\beta_{\max}(\Theta_{V,\infty}) = 1 .$$

Proof[3]. This proposition is derived from Theorem 11.5 by finding two sets Θ_1 and Θ_2 that are orthosymmetric in a wavelet basis and such that $\Theta_1 \subset \Theta_{V,\infty} \subset \Theta_2$. The main steps are given without details. One can verify that there exists $B_1 > 0$ such that

$$B_1 \min(\|f\|_V, \|f\|_\infty) \geq \sum_{j=-\infty}^{J} \sum_{l=1}^{3} \sum_{2^j n \in [0,1]^2} 2^{-j} |\langle f, \psi^l_{j,n} \rangle| + \sum_{2^J n \in [0,1]^2} 2^{-J} |\langle f, \phi^2_{J,n} \rangle| .$$

The set Θ_1 of functions such that

$$\sum_{j=-\infty}^{J} \sum_{l=1}^{3} \sum_{2^j n \in [0,1]^2} 2^{-j} |\langle f, \psi^l_{j,n} \rangle| + \sum_{2^J n \in [0,1]^2} 2^{-J} |\langle f, \phi^2_{J,n} \rangle| \leq B_1 C$$

is thus an orthosymmetric set included in $\Theta_{V,\infty}$.

Theorem 9.8 proves that there exists B_2 such that the sorted wavelet coefficients satisfy $|f^r_B[k]| \leq B_2 \|f\|_V \, k^{-1}$. Moreover, we saw in (9.49) that there exists B_3 such that

$$|\langle f, \psi^l_{j,n} \rangle| \leq B_3 \, 2^j \, \|f\|_\infty .$$

The set Θ_2 of all f such that $|\langle f, \psi^l_{j,n} \rangle| \leq B_3 \, 2^j \, C$ and $|f^r_B[k]| \leq B_2 \, C \, k^{-1}$ is therefore an orthosymmetric set that includes $\Theta_{V,\infty}$.

Since $\Theta_1 \subset \Theta_{V,\infty} \subset \Theta_2$ we have $\beta_{\max}(\Theta_1) \geq \beta_{\max}(\Theta_{V,\infty}) \geq \beta_{\max}(\Theta_2)$. One can verify that the sets Θ_1 and Θ_2 have a compact tail and that $s(\Theta_1) = s(\Theta_2) = 1$. Theorem 11.5 thus implies that

$$\beta_{\max}(\Theta_1) = \beta_{\max}(\Theta_2) = 1$$

and hence that $\beta_{\max}(\Theta_{V,\infty}) = 1$.

For a transform coding in a wavelet basis,

$$d(R, \Theta_1) \leq d(R, \Theta_{V,\infty}) \leq d(R, \Theta_2) .$$

Lemma 11.2 shows that an optimized transform code in a wavelet basis satisfies

$$d(R, \Theta_2) \sim C^2 \frac{\log_2 R}{R} .$$

Over Θ_1, one can select a family of signals whose distortion rate reaches the same decay, which finishes the proof of (11.77). ■

This proposition proves that for a general set of bounded variation images with a bounded amplitude, one cannot improve the decay exponent R^{-1} obtained in a wavelet basis. The hypothesis of bounded amplitude is important to guarantee that this set has a compact tail. Of course, for particular subsets of bounded variation images, one may construct coders with a distortion rate having a faster decay. Section 11.4.2 explains that this can be the case for images having edges with a regular geometry, if the coder can take advantage of this geometric regularity.

11.5 VIDEO SIGNALS [2]

Video compression is currently the most challenging coding problem, with considerable commercial applications. A video signal is a time sequence of images, with 30 images per second in the NTSC television standard. Time could be viewed as just another dimension, which would suggest decomposing video signals in an orthonormal basis for three dimensional signals. However, such a representation is inefficient because it ignores the fact that most image modifications in time are due to the relative motion of the scene with respect to the camera. This induces a displacement of grey level points in the image, which is called *optical flow*.

Section 11.5.2 describes MPEG video compression algorithms that use motion compensation to predict an image from the previous one. In addition to image compression, measuring optical flow has major applications in computer vision, for example in tracking objects and recovering depth information. Section 11.5.1 explains how to compute optical flow with multiscale approaches.

11.5.1 Optical Flow

The movement of a point in a three dimensional scene is observed through its projection in the image plane. If the image intensity is not uniform in the neighborhood of this point, the motion in the image plane appears as a displacement of the grey levels, which is the optical flow. Let $f(x,t)$ be the image intensity as a function of space and time. Measuring optical flow is an ill-posed problem. At one extreme, suppose that $f(x,t)$ is constant. Are the image points moving or not? They could be, but you cannot tell, because all grey levels have the same value. At the other extreme, the motion of a single white point in a black background is uniquely defined. The image grey level $f(x,t)$ must vary sufficiently along x in order to compute the grey level motion in time.

Most often, optical flow is measured using matching techniques, which are based on intuitive ideas that have relatively simple hardware implementations. Yet, comparative studies [87] show that the most accurate results are obtained with differential computations. Both approaches process images at several scales.

Block Matching The displacement vector $\tau_p(x)$ of the grey level $f_p(x)$ from time $t = p\Delta$ to $(p+1)\Delta$ satisfies $f_p(x) = f_{p+1}(x - \tau_p(x))$. The corresponding velocity vector is $v_p(x) = \Delta^{-1}\tau_p(x)$. If the velocity varies slowly in the neighborhood of

u, then for s sufficiently small,

$$f_p(x) \approx f_{p+1}(x - \tau_p(u)) \quad \text{if } |x-u| < s.$$

Under this assumption, the matching strategy estimates the displacement vector $\tau_p(u)$ by minimizing the norm of the difference $f_p(x) - f_{p+1}(x-\tau)$ over a square block of size s,

$$\epsilon(u,\tau) = \int\int_{|x-u|\leq s} \left| f_p(x_1,x_2) - f_{p+1}(x_1-\tau_1, x_2-\tau_2) \right|^2 dx_1\, dx_2 \,. \qquad (11.78)$$

The optical flow displacement $\tau_p(u)$ is approximated by the translation τ_u,ß which minimizes the error: $\epsilon(u,\tau_u) = \min_\tau \epsilon(u,\tau)$.

A major difficulty is to optimize the choice of s. If s is too small, then $f_p(x)$ may not have enough details in the neighborhood of u to match a unique neighborhood in $f_{p+1}(x)$. In this case, $\epsilon(u,\tau)$ has several local minima of similar amplitudes for different τ. Choosing the global minimum can lead to a wrong estimate of $\tau_p(u)$ in the presence of noise. If s is too large, then the velocity $v_p(x)$ may not remain constant for $|x-u| \leq s$. The neighborhood is not just translated but also deformed, which increases the differences between the two blocks. The minimization of $\epsilon(u,\tau)$ may thus also yield a wrong estimate of the displacement.

For a discrete image $f_p[n]$, the norm (11.78) is replaced by a discrete norm over a block of s^2 pixels. A sub-pixel estimation of the displacement $\tau_p(u)$ is obtained by calculating $f_{p+1}[n-\tau]$ with an interpolation when τ is not an integer. Computing the discretized norm (11.78) over a block of s^2 pixels requires s^2 additions and multiplications. For each of the N^2 pixels of $f_p[n]$, a brute force algorithm would compute the matching error of its neighborhood with the N^2 blocks corresponding to all potential displacement vectors τ. This requires s^2N^4 additions and multiplications, which is prohibitive.

Faster matching algorithms are implemented with multiscale strategies that vary the width s of the matching neighborhood [77]. Section 7.7.1 explains how to compute multiscale approximations a_j of an image f of N^2 pixels. At each scale $N^{-1} \leq 2^j \leq 1$, the approximation a_j includes 2^{-2j} pixels, as illustrated by Figure 11.20. Figure 7.23 gives an example. Let a_j^p and a_j^{p+1} be respectively the approximations of f_p and f_{p+1}. A *coarse to fine* algorithm matches the approximations a_j^p and a_j^{p+1} at a large scale $2^j = 2^J$, which is then progressively reduced. A coarse estimate of the displacement field is computed by matching each pixel of a_J^p with a pixel of a_J^{p+1} by minimizing a distance calculated over blocks of s^2 pixels, where s is typically equal to 5. This requires $s^2 2^{-4J}$ operations. A refined estimation of the velocity is then calculated at the next finer scale 2^{J-1}.

At any scale 2^j, each pixel $a_j^p[n]$ is an averaging of f_p around $2^j n$, over a neighborhood of size proportional to $N2^j$. At the same location, a displacement vector $\tau_{j+1}(2^j n)$ was previously calculated at the scale 2^{j+1}, by finding the best match in a_{j+1}^{p+1} for the block around $a_{j+1}^p[n/2]$. This displacement vector is used as

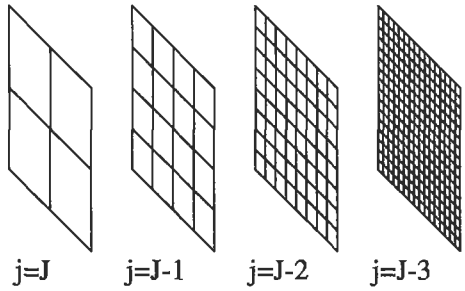


FIGURE 11.20 The pixels of an image approximation a_j correspond to averages of the image intensity over blocks of width proportional to $N2^j$.

an initial guess to find the block of a_j^{p+1} that best matches the block around $a_j^p[n]$. Among the K^2 displacement vectors τ of integers such that $|\tau - 2\tau_{j+1}(2^j n)| \leq K$, the best match $\tau_j(2^j n)$ is calculated by minimizing the difference between the blocks around $a_j^p[n]$ and $a_j^{p+1}[n-\tau]$. The 2^{-2j} displacement vectors $\tau_j(2^j n)$ for all pixels $a_j^p[n]$ are thus obtained with $O(K^2 s^2 2^{-2j})$ operations. The width s of the matching neighborhood remains unchanged across scales. A block of width s in a_j^p corresponds to a block of width $sN2^{2j}$ in f_p. This multiscale algorithm thus computes matching errors over neighborhoods of large sizes at first, and then reduces the size as the scale is refined. The total number of operations at all scales $2^J \leq 2^j \leq N^{-1}$ is $O(K^2 s^2 N^2)$, as opposed to $O(s^2 N^4)$ with a fixed scale algorithm.

Optical Flow Equation Motion vectors can be calculated with a totally different approach that relates the motion to the time and space derivatives of the image. Suppose that $x(t) = (x_1(t), x_2(t))$ is the coordinate of a grey level point that moves in time. By definition the grey level value $f(x(t), t)$ remains constant, so

$$\frac{df(x(t),t)}{dt} = \frac{\partial f(x,t)}{\partial x_1} x_1'(t) + \frac{\partial f(x,t)}{\partial x_2} x_2'(t) + \frac{\partial f(x,t)}{\partial t} = 0. \tag{11.79}$$

The velocity vector is $v = (v_1, v_2) = (x_1', x_2')$. Let $\vec{\nabla} f = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y})$ be the gradient of f. The optical flow equation derived from (11.79) relates v and f at any point x at time t:

$$\vec{\nabla} f \,.\, v = \frac{\partial f}{\partial x_1} v_1 + \frac{\partial f}{\partial x_2} v_2 = -\frac{\partial f}{\partial t}\,. \tag{11.80}$$

The optical flow equation specifies the projection of $v(x,t)$ over $\vec{\nabla} f(x,t)$ but gives no information about the orthogonal component of $v(x,t)$. This is commonly known as the *aperture problem*. Taking a pointwise derivative is similar to observing an edge through an aperture so small that the edge looks straight, as illustrated in Figure 11.21. The velocity parallel to the gradient can be calculated but the other

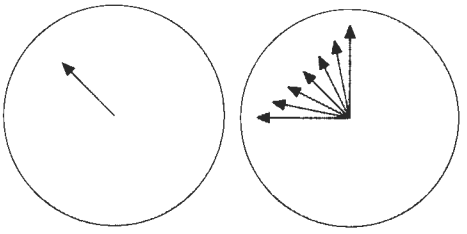


FIGURE 11.21 Seen through a small enough aperture, an edge looks straight. The gradient vector $\vec{\nabla} f$ shown on the left is perpendicular to the edge. The motion vector on the right can only be measured in the direction of $\vec{\nabla} f$.

component remains unknown. To circumvent this aperture problem one needs to make assumptions about the regularity in x of the velocity vector $v(x,t)$ [214].

Wavelet Flow Suppose that $v(x,t)$ is a smooth function of x, which means that it can be approximated by a constant over a sufficiently small domain. Weber and Malik [348] as well as Simoncelli [309] have shown that a precise estimate of $v(x,t)$ can be calculated by projecting the optical flow equation over multiscale wavelets. We describe the fast algorithm of Bernard [95], which computes the optical flow in a complex wavelet frame.

Let us consider a family of K complex mother wavelets $\{\psi^k\}_{0 \le k < K}$ of compact support that are dilated and translated:

$$\psi^k_{u,s}(x) = \frac{1}{s}\psi^k\left(\frac{x_1 - u_1}{s}, \frac{x_2 - u_2}{s}\right).$$

Suppose that s is small enough so that $v(x,t) \approx v(u,t)$ over the support of $\psi^k_{u,s}$. For any $1 \le k \le K$, computing a spatial inner product of the optical flow equation (11.80) with $\psi^k_{u,s}$ and performing an integration by parts gives

$$\left\langle f, \frac{\partial \psi^k_{u,s}}{\partial x_1} \right\rangle v_1(u,t) + \left\langle f, \frac{\partial \psi^k_{u,s}}{\partial x_2} \right\rangle v_2(u,t) = \frac{\partial}{\partial t}\left\langle f, \psi^k_{u,s}\right\rangle + \epsilon_s(u,t). \quad (11.81)$$

The error term $\epsilon_s(u,t)$ is due to the approximation of $v(x,t)$ by $v(u,t)$ over the wavelet support. The coefficients $\langle f, \frac{\partial \psi^k_{u,s}}{\partial x_1}\rangle$ and $\langle f, \frac{\partial \psi^k_{u,s}}{\partial x_2}\rangle$ are the wavelet coefficients of f at time t, calculated with new wavelets that are partial derivatives of the original ones. The wavelet flow equation (11.81) is a weak form of the original optical flow equation, which does not require that f be differentiable. If $v(x,t)$ is twice continuously differentiable and f has a singularity at u that is Lipschitz $\alpha < 1$, then Bernard [95] proves that $\epsilon_s(u,t)$ becomes negligible when s goes to zero.

Time Aliasing A video sequence is composed of images at intervals Δ in time: $f_p(x) = f(x, p\Delta)$. A second order estimate of $\frac{\partial}{\partial t}\langle f, \psi^k_{u,s}\rangle$ at $t = (p+1/2)\Delta$ is calculated with a finite difference:

$$\frac{\partial}{\partial t}\langle f, \psi^k_{u,s}\rangle = \frac{1}{\Delta}\langle f_{p+1} - f_p, \psi^k_{u,s}\rangle + \tilde{\epsilon}_a(u,t)\ . \tag{11.82}$$

If $v(x,t)$ is twice differentiable, we obtain a second order error term

$$|\tilde{\epsilon}_a(u,t)| = O\left(|v|^2\,\Delta^2\, s^{-2}\right)\ .$$

This error is small if $|v|\,\Delta \ll s$, which means that the image displacement during the time interval Δ is small compared to the wavelet support. Since $f(x,t)$ at $t = (p+1/2)\Delta$ is not known, it is approximated by $[f_p(x) + f_{p+1}(x)]/2$. Inserting (11.82) in (11.81) gives a wavelet flow equation at $t = (p+\frac{1}{2})\Delta$:

$$\left\langle \frac{f_p + f_{p+1}}{2}, \frac{\partial \psi^k_{u,s}}{\partial x_1} \right\rangle v_1(u,t) + \left\langle \frac{f_p + f_{p+1}}{2}, \frac{\partial \psi^k_{u,s}}{\partial x_2} \right\rangle v_2(u,t) \quad = \quad \frac{1}{\Delta}\langle f_{p+1} - f_p, \psi^k_{u,s}\rangle + \epsilon_s(u,t) + \epsilon_a(u,t) \quad . \tag{11.83}$$

If $v(x,t)$ is twice differentiable, then $|\epsilon_a(u,t)| = O(|v|^2\,\Delta^2\, s^{-2})$. The two error terms ϵ_s and ϵ_a are small if $|v| \ll s\Delta^{-1}$ and $v(x,t)$ is nearly constant over the support of $\psi^k_{u,s}$ whose size is proportional to s. The choice of s is a trade-off between these conflicting conditions.

The velocity v cannot be calculated if the wavelet coefficients in the left of (11.83) are zero. This happens when the image is constant over the wavelet support. Complex wavelets are used to avoid having wavelet coefficients that vanish at locations where the image is not locally constant. The sign of real wavelet coefficients can change in such domains, which means that real coefficients vanish regularly.

Optical Flow System If we neglect the error $\epsilon_s + \epsilon_a$ then (11.83) defines a system of K complex equations

$$W_{u,s}\, v_{u,s} = D_{u,s}\ . \tag{11.84}$$

The K by 2 matrix $W_{u,s}$ gives the inner products of $\frac{1}{2}(f_{p+1} + f_p)$ with partial derivatives of complex wavelets, whereas $D_{u,s}$ is the matrix of wavelet coefficients of $(f_{p+1} - f_p)/\Delta$, and $v_{u,s}$ is an estimate of the velocity vector at u. Let $\mathrm{Real}(M)$ be the matrix whose coefficients are the real parts of the coefficients of a complex matrix M. Since the velocity v is real, we compute the real least square solution of the overdetermined system (11.84), which is the solution of

$$\mathrm{Real}(W^*_{u,s}\, W_{u,s})\, v_{u,s} = \mathrm{Real}(W^*_{u,s} D_{u,s})\,. \tag{11.85}$$

There are three possible cases.

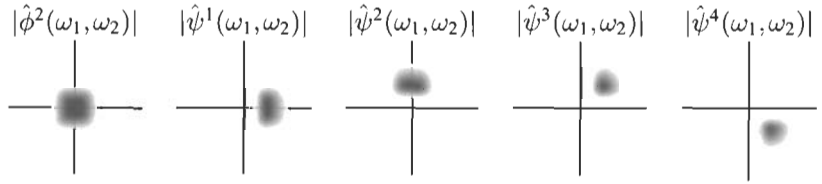


FIGURE 11.22 The image at far left shows the Fourier transform modulus $|\hat{\phi}^2(\omega_1,\omega_2)|$ of a scaling function in the Fourier plane. Black amplitude values are close to 1. The next four figures show $|\hat{\psi}^k(\omega_1,\omega_2)|$ for four analytic wavelets which generate a frame.

1. The least square solution yields a larger error $\|W_{u,s}v_{u,s} - D_{u,s}\|$. This means that the error terms $\epsilon_s + \epsilon_a$ cannot be neglected. Either the velocity is not approximately constant over the wavelet support, or its amplitude is too large.
2. The smallest eigenvalue of $\text{Real}(W^*_{u,s} W_{u,s})$ is comparable to the variance of the image noise. This happens when there is an aperture problem over the wavelet support. In this neighborhood of u, the image has local variations along a single direction, or is uniformly constant.
3. Otherwise the solution $v_{u,s}$ gives a precise estimate of $v(u,t)$ at $t = (p + 1/2)\Delta$.

Multiscale Calculation For fast calculations, we use the frame of complex wavelets with compact support calculated in Problem 7.15 from a separable wavelet basis. Figure 11.22 shows the Fourier transforms of these analytic wavelets. Each of them has an energy concentrated in one quadrant of the Fourier plane. There are four complex mother wavelets $\psi^k(x)$ which generate a frame of the space of *real* functions in $\mathbf{L}^2(\mathbb{R}^2)$

$$\left\{\psi^k_{j,n}(x) = 2^{-j}\psi^k(2^{-j}x - n)\right\}_{j\in\mathbb{Z},n\in\mathbb{Z}^2,1\le k\le 4}.$$

To the four complex wavelets, we add the real scaling function of the wavelet basis: $\psi^0_{j,n} = \phi^2_{j,n}$, which is considered as a wavelet in the following. The Fourier transform of ψ^0 is concentrated at low frequencies, as illustrated by Figure 11.22.

The optical flow system (11.83) is calculated with the five wavelets $\psi^k_{j,n}$ centered at $u = 2^j n$. The error ϵ_a is small at the scale 2^j if $|v|\,\Delta \ll 2^j$. This constraint is avoided by a multiscale algorithm that computes an estimate of the flow at a coarse scale 2^J, and progressively refines this estimate while performing a motion compensation.

Suppose that an estimate $v_{j+1,n/2}$ of $v(u,t)$ at $u = 2^{j+1}n/2 = 2^j n$ is already computed with a wavelet system at the scale 2^{j+1}. To reduce the amplitude of the displacement, a motion compensation is performed by translating $f_{p+1}(x)$ by

$-2^j\tau$ where $\tau \in \mathbb{Z}^2$ is chosen to minimize $|2^j\tau - \Delta v_{j+1,n/2}|$. Since the translation is proportional to 2^j, the indices of the wavelet coefficients of f_{p+1} at the scale 2^j are just translated by $-\tau$. This translation subtracts $2^j\tau\Delta^{-1}$ from the motion vector. At the scale 2^j, the estimate $v_{j,n}$ of the motion at $u = 2^j n$ is decomposed into

$$v_{j,n} = 2^j\tau\Delta^{-1} + v^r_{j,n},$$

where the residual motion $v^r_{j,n} = (v^r_1, v^r_2)$ is a solution of the motion-compensated wavelet flow equation calculated with wavelet coefficients translated by $-\tau$:

$$\frac{1}{2}\left(\left\langle f_p, \frac{\partial\psi^k_{j,n}}{\partial x_1}\right\rangle + \left\langle f_{p+1}, \frac{\partial\psi^k_{j,n+\tau}}{\partial x_1}\right\rangle\right) v^r_1 +$$
$$\frac{1}{2}\left(\left\langle f_p, \frac{\partial\psi^k_{j,n}}{\partial x_2}\right\rangle + \left\langle f_{p+1}, \frac{\partial\psi^k_{j,n+\tau}}{\partial x_2}\right\rangle\right) v^r_2 =$$
$$\frac{1}{\Delta}\left(\langle f_{p+1}, \psi^k_{j,n+\tau}\rangle - \langle f_p, \psi^k_{j,n}\rangle\right). \quad (11.86)$$

This motion compensation is similar to the multiscale matching idea, which takes advantage of a coarse scale estimate of the velocity to limit the matching search to a narrow domain at the next finer scale. The system (11.86) has five complex equations, for $0 \le k \le 4$. A real least square solution $v^r_{j,n}$ is calculated as in (11.85). Even when the velocity amplitude $|v|$ is large, the motion compensation avoids creating a large error when calculating the time derivatives of wavelet coefficients at fine scales, because the residual motions $v^r_{j,n}$ are small. Reducing the scale 2^j gives estimates of the motion that are denser and more precise.

For a discrete image of N^2 pixels, a fast filter bank algorithm requires $O(N^2)$ operations to compute the wavelet coefficients of the system (11.86) at all scales $N^{-1} \le 2^j < 1$ and locations $2^j n$ (Problem 11.18). Computing the least-square solutions of the motion compensated systems at all scales also requires $O(N^2)$ operations. The overall complexity of this algorithm is thus $O(N^2)$ [95]. A MATLAB code is available at `http://wave.cmap.polytechnique.fr/soft/OF/`.

Figure 11.23(b) shows the optical flow calculation for a Rubik cube on a turntable, and a street scene where three cars are moving. The arrows indicate the direction and amplitude of the motion vectors. A point corresponds to zero velocity. The algorithm does not compute the optical flow in areas where the image intensity is locally constant, or at the border of moving objects, where the motion vectors are discontinuous. If the motion is discontinuous, the assumption of having a nearly constant velocity over the wavelet support is violated at all scales. Figure 11.23(c) shows that the difference between two consecutive images of a video sequence has a large amplitude in regions where the image grey level has fine scale variations. This is particularly visible along edges. To reduce this error, a motion compensation predicts one image from a previous one by translating the image pixels with the displacement vectors derived from the motion vectors in Figure 11.23(b). Along sharp edges, small errors on the motion vectors produce

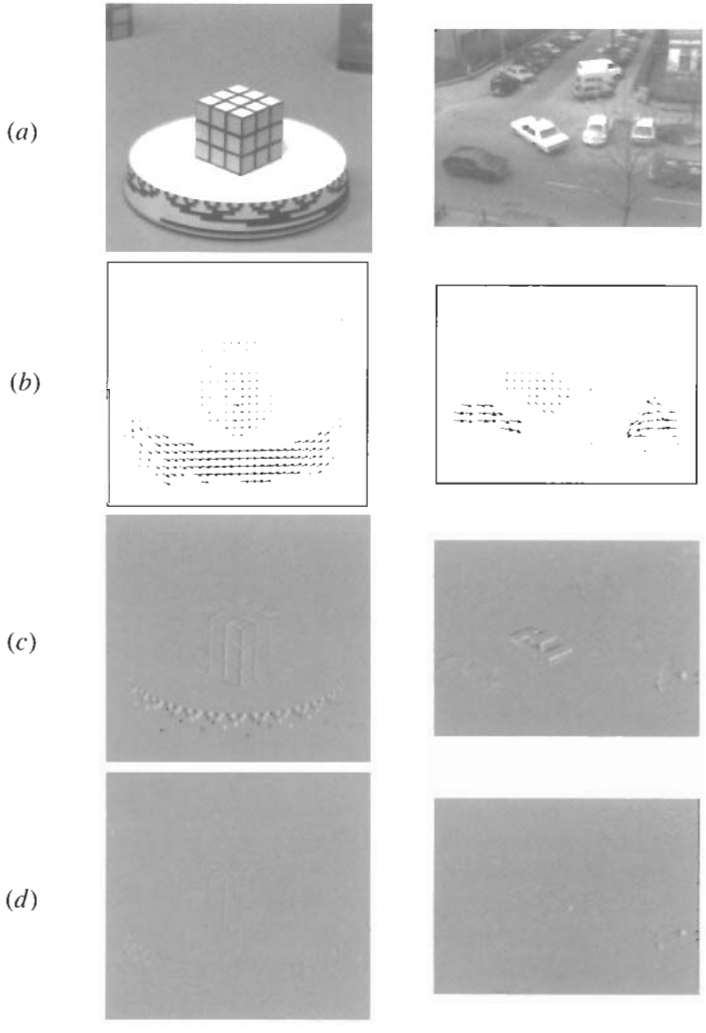


FIGURE 11.23 (a): Images of two video sequences: at left, a Rubik cube on a turntable; at right, three cars moving in a street. (b): Each arrow indicates the direction and amplitude of the local motion vector. (c): Difference between two consecutive images of the video sequence. Black, grey and white pixels correspond respectively to negative, zero and positive values. (d): Difference between an image and a prediction calculated from a previous image, with a motion compensation using the optical flow vectors in (b).

prediction errors, which are visible in Figure 11.23(d). However, these errors are much smaller than with the simple difference shown in Figure 11.23(c). The largest amplitude errors are along occlusion boundaries where the motion is discontinuous. Warping algorithms can compute motion discontinuities reliably, but they require more calculations [98].

11.5.2 MPEG Video Compression

In the MPEG video compression standards, motion vectors are coded to predict an image from a previous one with a motion compensation [120, 315]. MPEG-1 is devoted to lower quality video with applications to CD-ROM and video telecommunication. The bit rate is below 1.5 Mbits/s, with decent quality for entertainment video at 1.2 Mbits/s. It can handle video whose spatial and temporal resolution goes up to the NTSC television standard, with a degraded quality. MPEG-2 is designed for higher quality video, not lower than the NTSC television and up to High Definition Television (HDTV). It uses the same motion compensation algorithm as MPEG-1, and can handle interlaced video. MPEG-2 also offers scalability features, to provide a layered video bit stream that can be decoded and used by video supports having different spatial and temporal resolution. This is important for video browsing and for producing video signals used both for HDTV and the NTSC television formats.

MPEG-1 divides a video sequence in Groups Of Pictures (GOP) of typically 15 frames (half a second of video). The first image of a GOP is called an Intra frame. It is coded with the block cosine JPEG algorithm described in Section 11.4.3. In a GOP there are typically four P pictures, each coded from the previous one with a prediction algorithm using a motion compensation. They are divided in blocks of 16^2 pixels. For each block, MPEG-1 codes a displacement vector that specifies a matching block in the previous image. The difference between the two blocks is coded with JPEG, which uses a cosine transform over blocks of 8 by 8 pixels.

MPEG-1 does not specify the algorithm that computes the displacement vectors. To minimize the number of bits required by the JPEG code, we want to minimize the square norm of the difference between blocks. The displacement vector is thus often calculated with a matching algorithm that finds the displacement vector by minimizing the square norm of the error. The multiscale algorithm of the previous section can be implemented with various flavors. The motion compensation error is often concentrated along edges where the image has a sharp transition, and in particular occlusion contours. Figure 11.23(d) shows two examples of motion compensation errors calculated with a higher resolution optical flow, obtained with the wavelet flow algorithm.

Between P pictures, there are typically two B pictures that are coded with a bidirectional motion compensation, as illustrated in Figure 11.24. If the video jumps to a different scene or if occlusions are present, the precision of a block prediction may be very different if performed from a frame before or a frame after. For each block of a B picture, we find the blocks that have a minimum distance in the I or P picture that is just before and in the P picture that is just after. From these

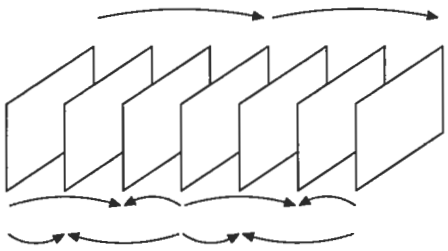


FIGURE 11.24 The P pictures are coded from the previous ones with a motion compensation, as indicated by the arrows. The B pictures are coded from the previous and next I or P pictures, in the order indicated by the numbers below.

two, the block with the smallest matching error is selected, and the difference with the original block of the B picture is coded with a block cosine JPEG.

To store or transmit video at a constant bit rate, it is necessary to buffer the variable bitstream generated over time by the encoder. A rate control algorithm adjusts the quantizers of the JPEG compression, depending on the video content and activity. It must ensure that the video buffer does not overflow while trying to maintain it as full as possible to maximize the image quality.

Future trends The MPEG-1 and MPEG-2 standards use low resolution motion vectors associated to blocks of 16^2 pixels. The design of these standards is limited by the constraints of real-time calculations. Calculating the motion vectors dominates the overall computational complexity of video coding. Real time calculation of higher resolution optical flow is also possible with the wavelet flow algorithm described in the previous section. However, higher resolution optical flow can improve video coding only if the array of motion vectors is efficiently coded and if the bit allocation between motion vectors and prediction errors is optimized.

To reach very low bit rates (4-64 kbits/s), prediction errors of motion compensation must be coded with a bit budget that is so reduced that transform codings in block cosine bases and wavelet bases introduce important degradations. Better results are obtained with adaptive representations such as the matching pursuit expansion of Section 9.5.2 [279]. The dictionary of two-dimensional vectors is optimized to match the structures of motion compensation errors [343], and the resulting decomposition coefficients are quantized and entropy coded.

The ongoing MPEG-4 standardization offers a more structured, "content based" approach to video coding at very low bit rates. Besides compression performance, MPEG-4 is adapted to the requirements of interactive video. A video scene is represented through *media objects*. These objects may correspond to elements of a natural scene such as a moving head in a video-conference, or computer graphics structures such as a written text. Natural images must be segmented in regions. Each region is a media object, which can be characterized by its con-

tour, motion and texture parameters. This approach is clearly promising but very difficult. It brings image compression into the world of computer vision [70].

11.6 PROBLEMS

11.1. [1] Let X be a random variable which takes its values in $\{x_k\}_{1 \leq k \leq 7}$ with probabilities $\{0.49, 0.26, 0.12, 0.04, 0.04, 0.03, 0.02\}$.

(a) Compute the entropy $\mathcal{H}(X)$. Construct a binary Huffman code and calculate the average bit rate R_X.

(b) Suppose that the symbols are coded with digits that may take three values: $-1, 0, 1$ instead of two as in a bit representation. Variable length ternary prefix codes can be represented with ternary trees. Extend the Huffman algorithm to compute a ternary prefix code for X that has a minimal average length.

11.2. [1] Let x_1 be the symbol of highest probability of a random variable X, and l_1 the length of its binary word in a Huffman code. Show that if $p_1 > 2/5$ then $l_1 = 1$. Verify that if $p_1 < 1/3$ then $l_1 \geq 2$.

11.3. [1] Let X be a random variable equal to x_1 or x_2 with probabilities $p_1 = 1 - \epsilon$ and $p_2 = \epsilon$. Verify that $\mathcal{H}(X)$ converges to 0 when ϵ goes to 0. Show that the Huffman code has an average number of bits that converges to 1 when ϵ goes to 0.

11.4. [2] Prove the Huffman code Proposition 11.1.

11.5. [1] Let X be a random variable with a probability density $p(x)$. Let Q be a quantizer whose bin sizes are $\{(y_{k-1}, y_k]\}_{1 \leq k \leq K}$.

(a) Prove that $\mathrm{E}\{|X - Q(X)|^2\}$ is minimum if and only if

$$Q(x) = x_k = \frac{\int_{y_{k-1}}^{y_k} x\, p(x)\, dx}{\int_{y_{k-1}}^{y_k} p(x)\, dx} \quad \text{for } x \in (y_{k-1}, y_k].$$

(b) Suppose that $p(x)$ is a Gaussian with variance σ^2. Find x_0 and x_1 for a "1 bit" quantizer defined by $y_0 = -\infty$, $y_1 = 0$ and $y_2 = +\infty$.

11.6. [1] Consider a pulse code modulation that quantizes each sample of a Gaussian random vector $F[n]$ and codes it with an entropy code that uses the same number of bits for each n. If the high resolution quantization hypothesis is satisfied, prove that the distortion rate is

$$d(\bar{R}) = \frac{\pi \mathrm{e}}{6} \mathrm{E}\{\|F\|^2\} 2^{-2\bar{R}}.$$

11.7. [1] Let $d = \sum_{m=0}^{N-1} d_m$ be the total distortion of a transform code. We suppose that the distortion rate $d_m(r)$ for coding the m^{th} coefficient is convex. Let $R = \sum_{m=0}^{N-1} R_m$ be the total number of bits.

(a) Prove that there exists a unique bit allocation that minimizes $d(R)$ for R fixed, and that it satisfies $\frac{\partial d_m(R_m)}{\partial r} = -\lambda$ where λ is a constant that depends on R. Hint: use Lagrange multipliers.

(b) Derive a new proof of Theorem 11.3.

(c) To impose that each R_m is a positive integer, we use a greedy iterative algorithm that allocates the bits one by one. Let $\{R_{m,p}\}_{0\leq m<N}$ be the bit allocation after p iterations, which means that a total of p bits have been allocated. The next bit is added to $R_{k,p}$ such that $\left|\frac{\partial d_k(R_{k,p})}{\partial r}\right| = \max_{0\leq m<N}\left|\frac{\partial d_m(R_{m,p})}{\partial r}\right|$. Justify this strategy. Prove that this algorithm gives an optimal solution if all curves $d_m(r)$ are convex and if $d_m(n+1)-d_m(n)\approx\frac{\partial d_m(n)}{\partial r}$ for all $n\in\mathbb{N}$.

11.8. [2] Let $X[m]$ be a binary first order Markov chain, which is specified by the transition probabilities $p_{01}=\Pr\{X[m]=1\,|\,X[m-1]=0\}$, $p_{00}=1-p_{01}$, $p_{10}=\Pr\{X[m]=0\,|\,X[m-1]=1\}$ and $p_{11}=1-p_{10}$.

(a) Prove that $p_0=\Pr\{X[m]=0\}=p_{10}/(p_{10}+p_{01})$ and that $p_1=\Pr\{X[m]=1\}=p_{01}/(p_{10}+p_{01})$.

(b) A run-length code records the length Z of successive runs of 0 values of $X[m]$ and the length I of successive runs of 1. Show that if Z and I are entropy coded, the average number of bits per sample of the run-length code, denoted $\bar{R}$, satisfies

$$\bar{R}\geq\bar{R}_{\min}=p_0\frac{\mathcal{H}(Z)}{\mathsf{E}\{Z\}}+p_1\frac{\mathcal{H}(I)}{\mathsf{E}\{I\}}.$$

(c) Let $\mathcal{H}_0=-p_{01}\log_2 p_{01}-(1-p_{01})\log_2(1-p_{01})$ and $\mathcal{H}_1=-p_{10}\log_2 p_{10}-(1-p_{10})\log_2(1-p_{10})$. Prove that

$$\bar{R}_{\min}=\mathcal{H}(X)=p_0\mathcal{H}_0+p_1\mathcal{H}_1\,,$$

which is the average information gained by moving one step ahead in the Markov chain.

(d) Suppose that the binary significance map of the transform code of a signal of size N is a realization of a first order Markov chain. We denote $\alpha=1/\mathsf{E}\{Z\}+1/\mathsf{E}\{I\}$. Let M be the number of significant coefficients (equal to 1). If $M\ll N$ then show that

$$\bar{R}_{\min}\approx\frac{M}{N}\left(\alpha\log_2\frac{N}{M}+\beta\right) \tag{11.87}$$

with $\beta=\alpha\log_2 e-2\alpha\log_2\alpha-(1-\alpha)\log_2(1-\alpha)$.

(e) Implement a run-length code for the binary significance maps of wavelet image coefficients $d_j^l[n,m]=\langle f,\psi_{j,n,m}^l\rangle$, for j and l fixed. See whether (11.87) approximates the bit rate $\bar{R}$ calculated numerically as a function of N/M for the Lena and Barbara images. How does α vary depending on the scale 2^j and the orientation $l=1,2,3$?

11.9. [3] Implement in WAVELAB a transform code that can compress an image in any basis of a separable wavelet packet dictionary. Perform numerical experiments on the Lena, Barbara and Peppers images. Compute the bit rate $\bar{R}$ in the "best basis" that minimizes the two cost functions (9.68) and (11.59). Compare the results. Is it more efficient to code these images with one of these best basis algorithm compared to a fixed wavelet basis?

11.10. [1] Implement the JPEG compression algorithm and replace the DCT-I by an orthogonal local cosine transform over blocks of the same size. Compare the compression rates in DCT-I and local cosine bases, as well as the visual image quality for $\bar{R} \in [0.2, 1]$.

11.11. [1] Implement a zero-tree embedded wavelet code for one-dimensional signals.

11.12. [3] Implement an adaptive image coder that selects a best basis by minimizing the cost function (11.58) in a wavelet packet dictionary. To optimize your transform code, you can either restrict the size of the wavelet packet dictionary, or elaborate an entropy code to specify the chosen basis within the dictionary. Compare this transform code with a wavelet transform code.

11.13. [3] Elaborate and implement a wavelet transform code for color images. Transform the red, green and blue channels in the color Karhunen-Loève basis calculated in Problem 9.8. Find an efficient algorithm that encodes together the embedded significance maps of these three channels, which are uncorrelated but highly dependent. Take advantage of the fact that the amplitude of grey level variations typically decreases from the Karhunen-Loève channel of highest variance to that of lowest variance.

11.14. [3] For most images, the amplitudes of DCT-I coefficients used in JPEG have a tendency to decrease when the frequency of the cosine vectors increases. Develop an embedded DCT-I transform code that takes advantage of this property by using zero-trees to record the position of significant coefficients in each block of 64 DCT-I coefficients [358].

11.15. [3] Develop and implement an algorithm that computes the optical flow of an image sequence with the coarse to fine multiscale matching strategy described in Section 11.5.1.

11.16. [3] Develop a video compression algorithm in a three dimensional wavelet basis [341]. In the time direction, choose a Haar wavelet in order to minimize the coding delay. This yields zero coefficients at locations where there is no movement in the image sequence. Implement a separable three-dimensional wavelet transform and design an efficient algorithm that records the positions of coefficients quantized to zero. How does your compression scheme compare to a motion compensation algorithm?

11.17. [2] Let $x(t)$ be the trajectory in the image of the projection of a point that moves in a scene. Suppose that the illumination of a scene changes in time by a factor $l(t)$.

(a) Explain why the image intensity satisfies $f(x(t),t) = \lambda l(t)$ where λ is a constant.

(b) Write a modified optical flow equation that adds a term $l'(t)/l(t)$ to the optical flow equation (11.79).

(c) Modify the wavelet flow algorithm of Section 11.5.1 to recover both the motion vectors and the illumination change.

11.18. [3] Let f_p and f_{p+1} be two consecutive images of N^2 pixels in a video sequence. With the results of Problem 7.14 and Problem 7.15, design a fast filter bank algorithm that requires $O(N^2)$ operations to compute all the inner products that appear in equation (11.86), for $4N^{-1} \leq 2^j < 1$ and $2^j n \in [0,1]^2$. Compute the motion vectors as a least square solution of these wavelet optical flow

systems. Compare your implementation with the Matlab code available at `http://wave.cmap.polytechnique.fr/soft/OF/`.

APPENDIX A

MATHEMATICAL COMPLEMENTS

Important mathematical concepts are reviewed without proof. Sections A.1–A.5 present results of real and complex analysis, including fundamental properties of Hilbert spaces, bases and linear operators [63]. Random vectors and Dirac distributions are covered in the last two sections.

A.1 FUNCTIONS AND INTEGRATION

Analog signals are modeled by measurable functions. We first give the main theorems of Lebesgue integration. A function f is said to be *integrable* if $\int_{-\infty}^{+\infty} |f(t)|\,dt < +\infty$. The space of integrable functions is written $\mathbf{L}^1(\mathbb{R})$. Two functions f_1 and f_2 are equal in $\mathbf{L}^1(\mathbb{R})$ if

$$\int_{-\infty}^{+\infty} |f_1(t) - f_2(t)|\,dt = 0.$$

This means that $f_1(t)$ and $f_2(t)$ can differ only on a set of points of measure 0. We say that they are *almost everywhere* equal.

The Fatou lemma gives an inequality when taking a limit under the Lebesgue integral of positive functions.

Lemma A.1 (FATOU) *Let $\{f_n\}_{n\in\mathbb{N}}$ be a family of positive functions $f_n(t) \geq 0$. If $\lim_{n\to+\infty} f_n(t) = f(t)$ almost everywhere then*

$$\int_{-\infty}^{+\infty} f(t)\,dt \leq \lim_{n\to+\infty} \int_{-\infty}^{+\infty} f_n(t)\,dt.$$

The dominated convergence theorem supposes the existence of an integrable upper bound to obtain an equality when taking a limit under a Lebesgue integral.

Theorem A.1 (DOMINATED CONVERGENCE) *Let* $\{f_n\}_{n\in\mathbb{N}}$ *be a family such that* $\lim_{n\to+\infty} f_n(t) = f(t)$ *almost everywhere. If*

$$\forall n \in \mathbb{N} \ |f_n(t)| \leq g(t) \quad \textit{and} \quad \int_{-\infty}^{+\infty} g(t)\,dt < +\infty \tag{A.1}$$

then f *is integrable and*

$$\int_{-\infty}^{+\infty} f(t)\,dt = \lim_{n\to+\infty} \int_{-\infty}^{+\infty} f_n(t)\,dt.$$

The Fubini theorem gives a sufficient condition for inverting the order of integrals in multidimensional integrations.

Theorem A.2 (FUBINI) *If* $\int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} |f(x_1,x_2)|dx_1 \right) dx_2 < +\infty$ *then*

$$\begin{aligned} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x_1,x_2)\,dx_1\,dx_2 &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(x_1,x_2)\,dx_1 \right) dx_2 \\ &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(x_1,x_2)\,dx_2 \right) dx_1 \,. \end{aligned}$$

Convexity A function $f(t)$ is said to be *convex* if for all $p_1, p_2 > 0$ with $p_1 + p_2 = 1$ and all $(t_1,t_2) \in \mathbb{R}^2$,

$$f(p_1t_1 + p_2t_2) \leq p_1 f(t_1) + p_2 f(t_2) \,.$$

The function $-f$ satisfies the reverse inequality and is said to be *concave*. If f is convex then the Jensen inequality generalizes this property for any $p_k \geq 0$ with $\sum_{k=1}^{K} p_k = 1$ and any $t_k \in \mathbb{R}$:

$$f\left(\sum_{k=1}^{K} p_k t_k \right) \leq \sum_{k=1}^{K} p_k f(t_k) \,. \tag{A.2}$$

The following proposition relates the convexity to the sign of the second order derivative.

Proposition A.1 *If* f *is twice differentiable, then* f *is convex if and only if* $f''(t) \geq 0$ *for all* $t \in \mathbb{R}$.

The notion of convexity also applies to sets $\Omega \subset \mathbb{R}^n$. This set is convex if for all $p_1, p_2 > 0$ with $p_1 + p_2 = 1$ and all $(x_1,x_2) \in \Omega^2$, then $p_1x_1 + p_2x_2 \in \Omega$. If Ω is not convex then its convex hull is defined as the smallest convex set that includes Ω.

A.2 BANACH AND HILBERT SPACES

Banach Space Signals are often considered as vectors. To define a distance, we work within a vector space $\mathbf{H}$ that admits a norm. A norm satisfies the following properties:

$$\forall f \in \mathbf{H} \ , \ \|f\| \geq 0 \text{ and } \|f\| = 0 \ \Leftrightarrow \ f = 0, \tag{A.3}$$

$$\forall \lambda \in \mathbb{C} \ \|\lambda f\| = |\lambda| \, \|f\|, \tag{A.4}$$

$$\forall f, g \in \mathbf{H} \ , \ \|f+g\| \leq \|f\| + \|g\|. \tag{A.5}$$

With such a norm, the convergence of $\{f_n\}_{n\in\mathbb{N}}$ to f in $\mathbf{H}$ means that

$$\lim_{n\to+\infty} f_n = f \ \Leftrightarrow \ \lim_{n\to+\infty} \|f_n - f\| = 0.$$

To guarantee that we remain in $\mathbf{H}$ when taking such limits, we impose a completeness property, using the notion of *Cauchy sequences*. A sequence $\{f_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence if for any $\epsilon > 0$, if n and p are large enough, then $\|f_n - f_p\| < \epsilon$. The space $\mathbf{H}$ is said to be *complete* if every Cauchy sequence in $\mathbf{H}$ converges to an element of $\mathbf{H}$.

Example A.1 For any integer $p > 0$ we define over discrete sequences $f[n]$

$$\|f\|_p = \left(\sum_{n=-\infty}^{+\infty} |f[n]|^p \right)^{1/p}.$$

The space $\mathbf{l^p} = \{f \ : \ \|f\|_p < +\infty\}$ is a Banach space with the norm $\|f\|_p$.

Example A.2 The space $\mathbf{L^p}(\mathbb{R})$ is composed of the measurable functions f on $\mathbb{R}$ for which

$$\|f\|_p = \left(\int_{-\infty}^{+\infty} |f(t)|^p \, dt \right)^{1/p} < +\infty.$$

This integral defines a norm and $\mathbf{L^p}(\mathbb{R})$ is a Banach space, provided one identifies functions that are equal almost everywhere.

Hilbert Space Whenever possible, we work in a space that has an inner product to define angles and orthogonality. A *Hilbert space* $\mathbf{H}$ is a Banach space with an inner product. The inner product of two vectors $\langle f, g \rangle$ is linear with respect to its first argument:

$$\forall \lambda_1, \lambda_2 \in \mathbb{C} \ , \ \langle \lambda_1 f_1 + \lambda_2 f_2, g \rangle = \lambda_1 \langle f_1, g \rangle + \lambda_2 \langle f_2, g \rangle. \tag{A.6}$$

It has an Hermitian symmetry:

$$\langle f, g \rangle = \langle g, f \rangle^*.$$

Moreover

$$\langle f, f\rangle \geq 0 \text{ and } \langle f, f\rangle = 0 \Leftrightarrow f = 0.$$

One can verify that $\|f\| = \langle f, f\rangle^{1/2}$ is a norm. The positivity (A.3) implies the Cauchy-Schwarz inequality:

$$|\langle f, g\rangle| \leq \|f\|\,\|g\|, \tag{A.7}$$

which is an equality if and only if f and g are linearly dependent.

Example A.3 An inner product between discrete signals $f[n]$ and $g[n]$ can be defined by

$$\langle f, g\rangle = \sum_{n=-\infty}^{+\infty} f[n]\, g^*[n].$$

It corresponds to an $\mathbf{l}^2(\mathbb{Z})$ norm:

$$\|f\|^2 = \langle f, f\rangle = \sum_{n=-\infty}^{+\infty} |f[n]|^2.$$

The space $\mathbf{l}^2(\mathbb{Z})$ of finite energy sequences is therefore a Hilbert space. The Cauchy-Schwarz inequality (A.7) proves that

$$\left| \sum_{n=-\infty}^{+\infty} f[n]\, g^*[n] \right| \leq \left(\sum_{n=-\infty}^{+\infty} |f[n]|^2 \right)^{1/2} \left(\sum_{n=-\infty}^{+\infty} |g[n]|^2 \right)^{1/2}.$$

Example A.4 Over analog signals $f(t)$ and $g(t)$, an inner product can be defined by

$$\langle f, g\rangle = \int_{-\infty}^{+\infty} f(t)\, g^*(t)\, dt.$$

The resulting norm is

$$\|f\| = \left(\int_{-\infty}^{+\infty} |f(t)|^2\, dt \right)^{1/2}.$$

The space $\mathbf{L}^2(\mathbb{R})$ of finite energy functions is thus also a Hilbert space. In $\mathbf{L}^2(\mathbb{R})$, the Cauchy-Schwarz inequality (A.7) is

$$\left| \int_{-\infty}^{+\infty} f(t)\, g^*(t)\, dt \right| \leq \left(\int_{-\infty}^{+\infty} |f(t)|^2\, dt \right)^{1/2} \left(\int_{-\infty}^{+\infty} |g(t)|^2\, dt \right)^{1/2}.$$

Two functions f_1 and f_2 are equal in $\mathbf{L}^2(\mathbb{R})$ if

$$\|f_1 - f_1\|^2 = \int_{-\infty}^{+\infty} |f_1(t) - f_2(t)|^2\, dt = 0,$$

which means that $f_1(t) = f_2(t)$ for almost all $t \in \mathbb{R}$.

A.3 BASES OF HILBERT SPACES

Orthonormal Basis A family $\{e_n\}_{n\in\mathbb{N}}$ of a Hilbert space $\mathbf{H}$ is orthogonal if for $n \neq p$

$$\langle e_n, e_p \rangle = 0.$$

If for $f \in \mathbf{H}$ there exists a sequence $\lambda[n]$ such that

$$\lim_{N\to+\infty} \|f - \sum_{n=0}^{N} \lambda[n]\, e_n\| = 0,$$

then $\{e_n\}_{n\in\mathbb{N}}$ is said to be an *orthogonal basis* of $\mathbf{H}$. The orthogonality implies that necessarily $\lambda[n] = \langle f, e_n \rangle / \|e_n\|^2$ and we write

$$f = \sum_{n=0}^{+\infty} \frac{\langle f, e_n \rangle}{\|e_n\|^2}\, e_n. \tag{A.8}$$

A Hilbert space that admits an orthogonal basis is said to be *separable*.

The basis is *orthonormal* if $\|e_n\| = 1$ for all $n \in \mathbb{N}$. Computing the inner product of $g \in \mathbf{H}$ with each side of (A.8) yields a Parseval equation for orthonormal bases:

$$\langle f, g \rangle = \sum_{n=0}^{+\infty} \langle f, e_n \rangle \langle g, e_n \rangle^*. \tag{A.9}$$

When $g = f$, we get an energy conservation called the *Plancherel formula*:

$$\|f\|^2 = \sum_{n=0}^{+\infty} |\langle f, e_n \rangle|^2. \tag{A.10}$$

The Hilbert spaces $\mathbf{l}^2(\mathbb{Z})$ and $\mathbf{L}^2(\mathbb{R})$ are separable. For example, the family of translated Diracs $\{e_n[k] = \delta[k-n]\}_{n\in\mathbb{Z}}$ is an orthonormal basis of $\mathbf{l}^2(\mathbb{Z})$. Chapter 7 and Chapter 8 construct orthonormal bases of $\mathbf{L}^2(\mathbb{R})$ with wavelets, wavelet packets and local cosine functions.

Riesz Bases In an infinite dimensional space, if we loosen up the orthogonality requirement, we must still impose a partial energy equivalence to guarantee the stability of the basis. A family of vectors $\{e_n\}_{n\in\mathbb{N}}$ is said to be a Riesz basis of $\mathbf{H}$ if it is linearly independent and there exist $A > 0$ and $B > 0$ such that for any $f \in \mathbf{H}$ one can find $\lambda[n]$ with

$$f = \sum_{n=0}^{+\infty} \lambda[n]\, e_n, \tag{A.11}$$

which satisfies

$$\frac{1}{B}\|f\|^2 \leq \sum_{n} |\lambda[n]|^2 \leq \frac{1}{A}\|f\|^2. \tag{A.12}$$

The Riesz representation theorem proves that there exist $\tilde{e}_n$ such that $\lambda[n] = \langle f, \tilde{e}_n \rangle$, and (A.12) implies that

$$\frac{1}{B}\|f\|^2 \leq \sum_n |\langle f, \tilde{e}_n \rangle|^2 \leq \frac{1}{A}\|f\|^2. \tag{A.13}$$

Theorem 5.2 derives that for all $f \in \mathbf{H}$,

$$A\|f\|^2 \leq \sum_n |\langle f, e_n \rangle|^2 \leq B\|f\|^2, \tag{A.14}$$

and

$$f = \sum_{n=0}^{+\infty} \langle f, \tilde{e}_n \rangle e_n = \sum_{n=0}^{+\infty} \langle f, e_n \rangle \tilde{e}_n.$$

The dual family $\{\tilde{e}_n\}_{n\in\mathbb{N}}$ is linearly independent and is also a Riesz basis. The case $f = e_p$ yields $e_p = \sum_{n=0}^{+\infty} \langle e_p, \tilde{e}_n \rangle e_n$. The linear independence of $\{e_n\}_{n\in\mathbb{N}}$ thus implies a biorthogonality relationship between dual bases, which are called *biorthogonal bases*:

$$\langle e_n, \tilde{e}_p \rangle = \delta[n-p]. \tag{A.15}$$

A.4 LINEAR OPERATORS

Classical signal processing algorithms are mostly based on linear operators. An operator T from a Hilbert space $\mathbf{H}_1$ to another Hilbert space $\mathbf{H}_2$ is linear if

$$\forall \lambda_1, \lambda_2 \in \mathbb{C}\ ,\ \forall f_1, f_2 \in \mathbf{H}\ ,\ T(\lambda_1 f_1 + \lambda_2 f_2) = \lambda_1 T(f_1) + \lambda_2 T(f_2).$$

Sup Norm The sup operator norm of T is defined by

$$\|T\|_S = \sup_{f\in\mathbf{H}_1} \frac{\|Tf\|}{\|f\|}. \tag{A.16}$$

If this norm is finite, then T is continuous. Indeed, $\|Tf - Tg\|$ becomes arbitrarily small if $\|f - g\|$ is sufficiently small.

Adjoint The *adjoint* of T is the operator T^* from $\mathbf{H}_2$ to $\mathbf{H}_1$ such that for any $f \in \mathbf{H}_1$ and $g \in \mathbf{H}_2$

$$\langle Tf, g \rangle = \langle f, T^* g \rangle.$$

When T is defined from $\mathbf{H}$ into itself, it is *self-adjoint* if $T = T^*$.

A non-zero vector $f \in \mathbf{H}$ is a called an *eigenvector* if there exists an *eigenvalue* $\lambda \in \mathbb{C}$ such that

$$Tf = \lambda f.$$

In a finite dimensional Hilbert space (Euclidean space), a self-adjoint operator is always diagonalized by an orthogonal basis $\{e_n\}_{0\leq n<N}$ of eigenvectors

$$Te_n = \lambda_n e_n.$$

When T is self-adjoint the eigenvalues λ_n are real. For any $f \in \mathbf{H}$,

$$Tf = \sum_{n=0}^{N-1} \langle Tf, e_n \rangle e_n = \sum_{n=0}^{N-1} \lambda_n \langle f, e_n \rangle e_n.$$

In an infinite dimensional Hilbert space, this result can be generalized by introducing the spectrum of the operator, which must be manipulated more carefully.

Orthogonal Projector Let $\mathbf{V}$ be a subspace of $\mathbf{H}$. A *projector* $P_{\mathbf{V}}$ on $\mathbf{V}$ is a linear operator that satisfies

$$\forall f \in \mathbf{H} \ , \ P_{\mathbf{V}} f \in \mathbf{V} \text{ and } \forall f \in \mathbf{V} \ , \ P_{\mathbf{V}} f = f.$$

The projector $P_{\mathbf{V}}$ is *orthogonal* if

$$\forall f \in \mathbf{H} \ , \ \forall g \in \mathbf{V} \ , \ \langle f - P_{\mathbf{V}} f, g \rangle = 0.$$

The following properties are often used in this book.

Proposition A.2 *If $P_{\mathbf{V}}$ is a projector on $\mathbf{V}$ then the following statements are equivalent:*

(i) $P_{\mathbf{V}}$ *is orthogonal.*
(ii) $P_{\mathbf{V}}$ *is self-adjoint.*
(iii) $\|P_{\mathbf{V}}\|_S = 1$.
(iv) $\forall f \in \mathbf{H}$, $\|f - P_{\mathbf{V}} f\| = \min_{g \in \mathbf{V}} \|f - g\|$.

If $\{e_n\}_{n \in \mathbb{N}}$ is an orthogonal basis of $\mathbf{V}$ then

$$P_{\mathbf{V}} f = \sum_{n=0}^{+\infty} \frac{\langle f, e_n \rangle}{\|e_n\|^2} e_n. \tag{A.17}$$

If $\{e_n\}_{n \in \mathbb{N}}$ is a Riesz basis of $\mathbf{V}$ and $\{\tilde{e}_n\}_{n \in \mathbb{N}}$ is the biorthogonal basis then

$$P_{\mathbf{V}} f = \sum_{n=0}^{+\infty} \langle f, e_n \rangle \tilde{e}_n = \sum_{n=0}^{+\infty} \langle f, \tilde{e}_n \rangle e_n \ . \tag{A.18}$$

Limit and Density Argument Let $\{T_n\}_{n \in \mathbb{N}}$ be a sequence of linear operators from $\mathbf{H}$ to $\mathbf{H}$. Such a sequence *converges weakly* to a linear operator T_∞ if

$$\forall f \in \mathbf{H} \ , \ \lim_{n \to +\infty} \|T_n f - T_\infty f\| = 0.$$

To find the limit of operators it is often preferable to work in a well chosen subspace $\mathbf{V} \subset \mathbf{H}$ which is dense. A space $\mathbf{V}$ is *dense* in $\mathbf{H}$ if for any $f \in \mathbf{H}$ there exist $\{f_m\}_{m \in \mathbb{N}}$ with $f_m \in \mathbf{V}$ such that

$$\lim_{m \to +\infty} \|f - f_m\| = 0.$$

The following proposition justifies this approach.

Proposition A.3 (DENSITY) *Let* $\mathbf{V}$ *be a dense subspace of* $\mathbf{H}$. *Suppose that there exists* C *such that* $\|T_n\|_S \leq C$ *for all* $n \in \mathbb{N}$. *If*

$$\forall f \in \mathbf{V} \ , \ \lim_{n \to +\infty} \|T_n f - T_\infty f\| = 0 \ ,$$

then

$$\forall f \in \mathbf{H} \ , \ \lim_{n \to +\infty} \|T_n f - T_\infty f\| = 0.$$

A.5 SEPARABLE SPACES AND BASES

Tensor Product Tensor products are used to extend spaces of one-dimensional signals into spaces of multiple dimensional signals. A tensor product $f_1 \otimes f_2$ between vectors of two Hilbert spaces $\mathbf{H}_1$ and $\mathbf{H}_2$ satisfies the following properties:

Linearity

$$\forall \lambda \in \mathbb{C} \ , \ \lambda (f_1 \otimes f_2) = (\lambda f_1) \otimes f_2 = f_1 \otimes (\lambda f_2). \tag{A.19}$$

Distributivity

$$(f_1 + g_1) \otimes (f_2 + g_2) = (f_1 \otimes f_2) + (f_1 \otimes g_2) + (g_1 \otimes f_2) + (g_1 \otimes g_2). \tag{A.20}$$

This tensor product yields a new Hilbert space $\mathbf{H} = \mathbf{H}_1 \otimes \mathbf{H}_2$ that includes all vectors of the form $f_1 \otimes f_2$ where $f_1 \in \mathbf{H}_1$ and $f_2 \in \mathbf{H}_2$, as well as linear combinations of such vectors. An inner product in $\mathbf{H}$ is derived from inner products in $\mathbf{H}_1$ and $\mathbf{H}_2$ by

$$\langle f_1 \otimes f_2 , g_1 \otimes g_2 \rangle_{\mathbf{H}} = \langle f_1, g_1 \rangle_{\mathbf{H}_1} \langle f_2, g_2 \rangle_{\mathbf{H}_2}. \tag{A.21}$$

Separable Bases The following theorem proves that orthonormal bases of tensor product spaces are obtained with separable products of two orthonormal bases. It provides a simple procedure for transforming bases for one-dimensional signals into separable bases for multidimensional signals.

Theorem A.3 *Let* $\mathbf{H} = \mathbf{H}_1 \otimes \mathbf{H}_2$. *If* $\{e_n^1\}_{n \in \mathbb{N}}$ *and* $\{e_n^2\}_{n \in \mathbb{N}}$ *are two Riesz bases respectively of* $\mathbf{H}_1$ *and* $\mathbf{H}_2$ *then* $\{e_n^1 \otimes e_m^2\}_{(n,m) \in \mathbb{N}^2}$ *is a Riesz basis of* $\mathbf{H}$. *If the two bases are orthonormal then the tensor product basis is also orthonormal.*

Example A.5 A product of functions $f \in \mathbf{L}^2(\mathbb{R})$ and $g \in \mathbf{L}^2(\mathbb{R})$ defines a tensor product:

$$f(x_1)\, g(x_2) = f \otimes g(x_1, x_2).$$

Let $\mathbf{L}^2(\mathbb{R}^2)$ be the space of $h(x_1, x_2)$ such that

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |h(x_1, x_2)|^2 \, dx_1 \, dx_2 < +\infty.$$

One can verify that $\mathbf{L}^2(\mathbb{R}^2) = \mathbf{L}^2(\mathbb{R}) \otimes \mathbf{L}^2(\mathbb{R})$. Theorem A.3 proves that if $\{\psi_n(t)\}_{n \in \mathbb{N}}$ is an orthonormal basis of $\mathbf{L}^2(\mathbb{R})$, then $\{\psi_{n_1}(x_1)\, \psi_{n_2}(x_2)\}_{(n_1, n_2) \in \mathbb{N}^2}$ is an orthonormal basis of $\mathbf{L}^2(\mathbb{R}^2)$.

Example A.6 A product of discrete signals $f \in \mathbf{l}^2(\mathbb{Z})$ and $g \in \mathbf{l}^2(\mathbb{Z})$ also defines a tensor product:

$$f[n_1]\, g[n_2] = f \otimes g[n_1, n_2].$$

The space $\mathbf{l}^2(\mathbb{Z}^2)$ of images $h[n_1, n_2]$ such that

$$\sum_{n_1=-\infty}^{+\infty} \sum_{n_2=-\infty}^{+\infty} |h[n_1, n_2]|^2 < +\infty$$

is also decomposed as a tensor product $\mathbf{l}^2(\mathbb{Z}^2) = \mathbf{l}^2(\mathbb{Z}) \otimes \mathbf{l}^2(\mathbb{Z})$. Orthonormal bases can thus be constructed with separable products.

A.6 RANDOM VECTORS AND COVARIANCE OPERATORS

A class of signals can be modeled by a random process (random vector) whose realizations are the signals in the class. Finite discrete signals f are represented by a random vector Y, where $Y[n]$ is a random variable for each $0 \leq n < N$. For a review of elementary probability theory for signal processing, the reader may consult [56, 59].

Covariance Operator The average of a random variable X is $\mathsf{E}\{X\}$. The covariance of two random variables X_1 and X_2 is

$$\mathrm{Cov}(X_1, X_2) = \mathsf{E}\Big\{ \Big(X_1 - \mathsf{E}\{X_1\}\Big) \Big(X_2 - \mathsf{E}\{X_2\}\Big)^* \Big\}. \tag{A.22}$$

The covariance matrix of a random vector Y is composed of the N^2 covariance values

$$R[n, m] = \mathrm{Cov}\Big(Y[n], Y[m]\Big).$$

It defines the covariance operator K which transforms any $h[n]$ into

$$Kh[n] = \sum_{m=0}^{N-1} R[n, m]\, h[m].$$

For any h and g

$$\langle Y, h \rangle = \sum_{n=0}^{N-1} Y[n]\, h^*[n] \quad \text{and} \quad \langle Y, g \rangle = \sum_{n=0}^{N-1} Y[n]\, g^*[n]$$

are random variables and

$$\mathrm{Cov}\Big(\langle Y, h \rangle, \langle Y, g \rangle\Big) = \langle Kg, h \rangle. \tag{A.23}$$

The covariance operator thus specifies the covariance of linear combinations of the process values.

Karhunen-Loève Basis The covariance operator K is self-adjoint because $R[n,m] = R^*[m,n]$ and positive because

$$\langle Kh, h\rangle = \mathsf{E}\{|\langle Y, h\rangle|^2\} \geq 0. \tag{A.24}$$

This guarantees the existence of an orthogonal basis $\{g_k\}_{0\leq k<N}$ that diagonalizes K:

$$Kg_k = \sigma_k^2 g_k.$$

This basis is called a *Karhunen-Loève basis* of Y, and the vectors g_k are the *principal directions*. The eigenvalues are the variances

$$\sigma_k^2 = \langle Kg_k, g_k\rangle = \mathsf{E}\{|\langle Y, g_k\rangle|^2\}. \tag{A.25}$$

Wide-Sense Stationarity We say that Y is *wide-sense stationary* if

$$\mathsf{E}\{Y[n]\, Y^*[m]\} = R[n,m] = R_Y[n-m]. \tag{A.26}$$

The correlation at two points depends only on the distance between these points. The operator K is then a convolution whose kernel $R_Y[k]$ is defined for $-N < k < N$. A wide-sense stationary process is *circular stationary* if $R_Y[n]$ is N periodic:

$$R_Y[n] = R_Y[N+n] \quad \text{for } -N \leq n \leq 0. \tag{A.27}$$

This condition implies that a periodic extension of $Y[n]$ on $\mathbb{Z}$ remains wide-sense stationary on $\mathbb{Z}$. The covariance operator K of a circular stationary process is a discrete circular convolution. Section 3.3.1 proves that the eigenvectors of circular convolutions are the discrete Fourier vectors

$$\left\{ g_k[n] = \frac{1}{\sqrt{N}} \exp\left(\frac{i2\pi kn}{N}\right)\right\}_{0\leq k<N}.$$

The discrete Fourier basis is therefore the Karhunen-Loève basis of circular stationary processes. The eigenvalues (A.25) of K are the discrete Fourier transform of R_Y and are called the *power spectrum*

$$\sigma_k^2 = \hat{R}_Y[k] = \sum_{n=0}^{N-1} R_Y[n] \exp\left(\frac{-i2k\pi n}{N}\right). \tag{A.28}$$

The following theorem computes the power spectrum after a circular convolution.

Theorem A.4 *Let Z be a wide-sense circular stationary random vector. The random vector $Y[n] = Z \circledast h[n]$ is also wide-sense circular stationary and its power spectrum is*

$$\hat{R}_Y[k] = \hat{R}_Z[k]\, |\hat{h}[k]|^2. \tag{A.29}$$

A.7 DIRACS

Diracs are useful in making the transition from functions of a real variable to discrete sequences. Symbolic calculations with Diracs simplify computations, without worrying about convergence issues. This is justified by the theory of distributions [66, 69]. A Dirac δ has a support reduced to $t = 0$ and associates to any continuous function ϕ its value at $t = 0$

$$\int_{-\infty}^{+\infty} \delta(t)\,\phi(t)\,dt = \phi(0). \tag{A.30}$$

Weak Convergence A Dirac can be obtained by squeezing an integrable function g such that $\int_{-\infty}^{+\infty} g(t)\,dt = 1$. Let $g_s(t) = \frac{1}{s} g(\frac{t}{s})$. For any continuous function ϕ

$$\lim_{s \to 0} \int_{-\infty}^{+\infty} g_s(t)\,\phi(t)\,dt = \phi(0) = \int_{-\infty}^{+\infty} \delta(t)\,\phi(t)\,dt. \tag{A.31}$$

A Dirac can thus formally be defined as the limit $\delta = \lim_{s \to 0} g_s$, which must be understood in the sense of (A.31). This is called *weak convergence*. A Dirac is not a function since it is zero at $t \neq 0$ although its "integral" is equal to 1. The integral at the right of (A.31) is only a symbolic notation which means that a Dirac applied to a continuous function ϕ associates its value at $t = 0$.

General distributions are defined over the space $\mathbf{C}_0^\infty$ of *test functions* which are infinitely continuously differentiable with a compact support. A distribution d is a linear form that associates to any $\phi \in \mathbf{C}_0^\infty$ a value that is written $\int_{-\infty}^{+\infty} d(t)\phi(t)dt$. It must also satisfy some weak continuity properties [66, 69] that we do not discuss here, and which are satisfied by a Dirac. Two distributions d_1 and d_2 are equal if

$$\forall \phi \in \mathbf{C}_0^\infty \;,\; \int_{-\infty}^{+\infty} d_1(t)\,\phi(t)\,dt = \int_{-\infty}^{+\infty} d_2(t)\,\phi(t)\,dt. \tag{A.32}$$

Symbolic Calculations The symbolic integral over a Dirac is a useful notation because it has the same properties as a usual integral, including change of variables and integration by parts. A translated Dirac $\delta_\tau(t) = \delta(t - \tau)$ has a mass concentrated at τ and

$$\int_{-\infty}^{+\infty} \phi(t)\,\delta(t-u)\,dt = \int_{-\infty}^{+\infty} \phi(t)\,\delta(u-t)\,dt = \phi(u).$$

This means that $\phi \star \delta(u) = \phi(u)$. Similarly $\phi \star \delta_\tau(u) = \phi(u - \tau)$.

A Dirac can also be multiplied by a continuous function ϕ and since $\delta(t - \tau)$ is zero outside $t = \tau$, it follows that

$$\phi(t)\,\delta(t-\tau) = \phi(\tau)\,\delta(t-\tau).$$

The derivative of a Dirac is defined with an integration by parts. If ϕ is continuously differentiable then

$$\int_{-\infty}^{+\infty} \phi(t)\,\delta'(t)\,dt = -\int_{-\infty}^{+\infty} \phi'(t)\,\delta(t)\,dt = -\phi'(0).$$

The k^{th} derivative of δ is similarly obtained with k integrations by parts. It is a distribution that associates to $\phi \in \mathbf{C}^k$

$$\int_{-\infty}^{+\infty} \phi(t)\,\delta^{(k)}(t)\,dt = (-1)^k\,\phi^{(k)}(0).$$

The Fourier transform of δ associates to any $e^{-i\omega t}$ its value at $t = 0$:

$$\hat{\delta}(\omega) = \int_{-\infty}^{+\infty} \delta(t)\,e^{-i\omega t}\,dt = 1,$$

and after translation $\hat{\delta}_\tau(\omega) = e^{-i\tau\omega}$. The Fourier transform of the Dirac comb $c(t) = \sum_{n=-\infty}^{+\infty} \delta(t - nT)$ is therefore $\hat{c}(\omega) = \sum_{n=-\infty}^{+\infty} e^{-inT\omega}$. The Poisson formula (2.4) proves that

$$\hat{c}(\omega) = \frac{2\pi}{T} \sum_{k=-\infty}^{+\infty} \delta\left(\omega - \frac{2\pi k}{T}\right).$$

This distribution equality must be understood in the sense (A.32).

APPENDIX B

SOFTWARE TOOLBOXES

The book algorithms are implemented in WaveLab and LastWave, which are freeware softwares that can be retrieved through the Internet. Nearly all the computational figures of the book are reproduced as demos. Other freeware toolboxes are listed in Section B.3. Pointers to new software and information concerning the Wavelet Digest newsletter is available at

`http://www.wavelet.org`.

B.1 WaveLab

WaveLab is a library of Matlab routines for wavelets and related time-frequency transforms. It is improved and maintained at Stanford University by David Donoho with contributions to earlier versions by John Buckheit, Shaobing Chen, Xiaoming Huo, Iain Johnstone, Eric Kolaczyk, Jeffrey Scargle, and Thomas Yu [105]. It requires buying Matlab, which offers an interactive environment for numerical computations and visualizations. Matlab is a product of The Mathworks company based in Natick, Massachusetts. The WaveLab version 0.800 has more than 800 files including programs, data, documentation and scripts, which can be retrieved at:

`http://www-stat.stanford.edu/~wavelab`.

Versions are available for Unix workstations, Linux, Macintosh, and PC (Windows).

A partial list of directories inside `WaveLab` is provided (in bold). For each directory, we give the names of the main computational subroutines, followed by the sections that describe the algorithms and the figures that use them.

Datasets Synthetic and real signals.

`ReadSignal` Reads a Signal from a data set of one-dimensional signals. Figures 4.7, 6.7, 8.19, 9.9 and 9.11.

`ReadImage` Reads an Image from an image data set. Figure 9.10.

`MakeSignal` Makes a synthetic one-dimensional Signal. Figures 2.1, 4.3, 4.14, 4.13, 4.18, 6.3, 6.6, 9.1, 10.1, 10.5.

`MakeImage` Makes a synthetic Image. Figure 7.26.

`MakeProcess` Makes a realization of a stochastic Process. Section 10.6.3. Figure 10.19.

`MakeBrownian` Makes a realization of a fractional Brownian motion. Section 6.4.3. Figure 6.20.

`MakeCantor` Makes a generalized Cantor measure. Section 6.4.1. Figures 6.16 and 6.18.

Continuous Continuous wavelet transform tools.

`RWT` Real Wavelet Transform. Sections 4.3.1 and 4.3.3. Figures 4.7, 6.1, 6.3, 6.5, 6.6, 6.16 and 6.20.

`IRWT` Inverse Real Wavelet Transform. Sections 4.3.1 and 4.3.3.

`MM_RWT` Modulus Maxima of a Real Wavelet Transform. Section 6.2. Figures 6.5, 6.6, 6.7, 6.16 and 6.20.

`SkelMap` Skeleton Map of maxima curves. Section 6.2. Figures 6.5, 6.6, 6.7 and 6.16.

`AWT` Analytic Wavelet Transform. Sections 4.3.2 and 4.3.3. Figures 4.11, 4.16 and 4.17.

`IAWT` Inverse Analytic Wavelet Transform. Sections 4.3.2 and 4.3.3.

`Ridge_AWT` Ridges of an Analytic Wavelet Transform. Section 4.4.2. Figures 4.15, 4.16 and 4.17.

Fractals Fractal computations.

`FracPartition` Fractal Partition function based on wavelet modulus maxima. Section 6.4.2. Figure 6.18.

`FracScalExp` Fractal Scaling Exponent of the partition function. Section 6.4.2. Figures 6.18 and 6.20.

`FracSingSpect` Fractal Singularity Spectrum. Section 6.4.2. Figures 6.18 and 6.20.

TimeFrequency Time-frequency distributions.

`WindowFT` Windowed Fourier Transform. Section 4.2. Figures 4.3, 4.13 and 4.14.

`IWindowFT` Inverse Windowed Fourier Transform. Sections 4.2.1 and 4.2.3.

`Ridge_WindowFT` Ridges of a Windowed Fourier Transform. Section 4.4.1. Figures 4.12, 4.13 and 4.14.

`WignerDist` Wigner-Ville Distribution. Sections 4.5.1 and 4.5.4. Figures 4.18 and 4.19.

`CohenDist` Cohen class time-frequency Distributions. Sections 4.5.3 and 4.5.4. Figures 4.20 and 4.21.

Orthogonal Periodic Orthogonal wavelet transforms.

`FWT_PO` Forward Wavelet Transform, Periodized and Orthogonal. Sections 7.3.1 and 7.5.1. Figures 7.7 and 9.2.

`IWT_PO` Inverse Wavelet Transform, Periodized and Orthogonal. Sections 7.3.1 and 7.5.1. Figure 9.2.

`FWT_IO` Forward Wavelet Transform, on the Interval and Orthogonal. Sections 7.3.1 and 7.5.3. Figure 9.2.

`IWT_IO` Inverse Wavelet Transform, on the Interval and Orthogonal. Sections 7.3.1 and 7.5.3. Figure 9.2.

`FWT2_PO` Forward Wavelet Transform of images, Periodized and Orthogonal. Section 7.7.3. Figure 7.26.

`IWT2_PO` Inverse Wavelet Transform of images, Periodized and Orthogonal. Sections 7.7.3.

`MakeONFilter` Makes Orthogonal conjugate mirror Filters for Daubechies, Coiflets, Symmlets, Haar and Battle-Lemarié wavelets. Sections 7.1.3 and 7.2.3. Figure 7.4.

`MakeOBFilter` Makes Orthogonal Boundary conjugate mirror Filters for Cohen-Daubechies-Vial wavelets. Section 7.5.3.

`MakeWavelet` Makes graph of orthogonal Wavelets and scaling functions. Section 7.3.1. Figures 7.2, 7.5, 7.9 and 7.10.

Meyer Meyer orthogonal and periodic wavelet transforms.

`FWT_YM` Forward Wavelet Transform with Yves Meyer wavelets. Sections 7.2.2, 8.4.2 and 8.4.4.

`IWT_YM` Inverse Wavelet Transform with Yves Meyer wavelets. Sections 7.2.2, 8.4.2 and 8.4.4.

`FWT2_YM` Forward Wavelet Transform of images with Yves Meyer wavelets. Sections 7.7.2, 8.4.2 and 8.4.4.

`IWT2_YM` Inverse Wavelet Transform of images with Yves Meyer wavelets. Sections 7.7.2, 8.4.2 and 8.4.4.

Biorthogonal Biorthogonal wavelet transforms.

`FWT_PB` Forward Wavelet Transform, Periodized and Biorthogonal. Sections 7.3.2 and 7.4.

`IWT_PB` Inverse Wavelet Transform, Periodized and Biorthogonal. Sections 7.3.2 and 7.4.

`FWT2_PB` Forward Wavelet Transform of images, Periodized and Biorthogonal. Section 7.7.3.

`IWT2_PB` Inverse Wavelet Transform of images, Periodized and Biorthogonal. Section 7.7.3.

`MakeBSFilter` Makes perfect reconstruction Biorthogonal Symmetric Filters. Section 7.4.3.

`MakeBSWavelet` Makes graph of Biorthogonal Symmetric Wavelets and scaling functions. Figures 7.14 and 7.15.

Interpolating Multiscale interpolations.

`FWT_DD` Forward interpolating Wavelet Transform calculated with Deslauriers-Dubuc filters. Section 7.6.2.

`IWT_DD` Inverse interpolating Wavelet Transform calculated with Deslauriers-Dubuc filters. Section 7.6.2.

Invariant Translation invariant wavelet transforms.

`FWT_ATrou` Forward dyadic Wavelet Transform calculated with the Algorithme à Trous. Section 5.5. Figures 5.5 and 6.7.

`IWT_ATrou` Inverse dyadic Wavelet Transform calculated with the Algorithme à Trous. Sections 5.5 and 6.7.

`FWT_Stat` Forward dyadic Wavelet Transform calculated with Stationary shifts of the signal. Section 10.2.4. Figures 10.4 and 10.5.

`IWT_Stat` Inverse dyadic Wavelet Transform calculated with Stationary shifts of the signal. Section 10.2.4. Figures 10.4 and 10.5.

`MM_DWT` Modulus Maxima of a Dyadic Wavelet Transform. Section 6.2.2. Figure 6.7.

`IMM_DWT` Inverse reconstruction of signals from Modulus Maxima of a Dyadic Wavelet Transform. Section 6.2.2. Figure 6.8.

`FWT2_ATrou` Forward dyadic Wavelet Transform of images calculated with the Algorithme à Trous. Section 6.3.2. Figures 6.9 and 6.10.

`MM2_DWT` Modulus Maxima of an image Dyadic Wavelet Transform. Section 6.3.2. Figures 6.9 and 6.10.

`IMM2_DWT` Inverse reconstruction of an image from Modulus Maxima of a Dyadic Wavelet Transform. Section 6.3. Figure 6.11.

Packets Best wavelet packet and local cosine bases.

One-D For one-dimensional signals.

`WPTour` WavePacket tree decomposition and best basis selection. Sections 8.1 and 9.4. Figures 8.6 and 8.8.

`MakeWaveletPacket` Makes graph of WavePacket functions. Section 8.1. Figures 8.2 and 8.4.

`CPTour` Local Cosine Packet tree decomposition and best basis selection. Sections 8.5 and 9.4. Figures 8.19, 9.9 and 9.11.

`KLinCP` Karhunen-Loève basis estimation in a Cosine Packet tree. Section 10.6.2. Figure 10.19.

Two-D For two-dimensional signals.

`WP2Tour` WavePacket 2-dimensional decomposition and best basis selection. Sections 8.2 and 9.4.2.

`CP2Tour` Local Cosine Packet 2-dimensional decomposition and best basis selection. Sections 8.5.3 and 9.4.2. Figures 8.22 and 9.10.

Pursuit Basis and matching pursuits.

`WPBPursuitTour` WavePacket dictionary for Basis Pursuits. Section 9.5.1. Figure 9.11.

`CPBPursuitTour` Cosine Packet dictionary for Basis Pursuits. Section 9.5.1.

`WPMPursuitTour` WavePacket dictionary for Matching Pursuits. Section 9.5. Figures 9.11 and 9.12.

`CPMPursuitTour` Cosine Packet dictionary for Matching Pursuits. Section 9.5.

`GaborPursuitTour` Gabor dictionary for Matching Pursuits. Section 9.5.2. Figures 9.11(b) and 9.12.

DeNoising Removal of additive noises.

`ThreshWave` Thresholds orthogonal Wavelet coefficients. Section 10.2.4. Figures 10.4 and 10.5.

`ThreshWave2` Thresholds orthogonal Wavelet coefficients of images. Section 10.2.4. Figure 10.6.

`ThreshWP` Thresholds coefficients of a best WavePacket basis. Section 10.2.5.

`ThreshCP` Thresholds coefficients of a best Cosine Packet basis. Section 10.2.5. Figure 10.8.

`CohWave` Coherent threshold of orthogonal Wavelet coefficients. Section 10.5.1. Figure 10.15.

Figure Demonstration The `Wavelab` directory has a folder called `WaveTour`. It contains a subdirectory for each chapter (`WTCh01`, `WTCh02`,...); these subdirectories include all the files needed to reproduce the computational figures. Each directory has a demo file. For example, the figures of Chapter 4 are reproduced by invoking the file `WTCh04Demo` in MATLAB. A menu bar appears on the screen, listing all computational figures of Chapter 4. When a figure number is activated by a mouse-click, the calculations are reproduced and the resulting graphs are displayed in a separate window. The command window gives a narrative explaining the results. The file `WTCh04Demo.m` is in the directory `WTCh04`. The MATLAB source code that computes Figure 4.X is in the file `wt04figX.m` in that same directory. Equivalent names are used for all other chapters.

B.2 LastWave

LastWave is a wavelet signal and image processing environment, written in C for X11/Unix and Macintosh computers. This stand-alone freeware does not require any additional commercial package, and can be retrieved through the Internet at:

```
http://wave.cmap.polytechnique.fr/soft/LastWave/ .
```

LastWave was created and is maintained by Emmanuel Bacry, at École Polytechnique in France. It includes a command line language, and a high level object-oriented graphic language for displaying simple objects (buttons, strings,...) and more complex ones (signals, images, wavelet transforms, time-frequency planes...). The computational subroutines and commands are regrouped in independent packages. An extensive on-line documentation is available. New commands are added with the command language or as C subroutines. This software is rapidly evolving with packages provided by users across the Internet. The current contributors include Benjamin Audit, Geoff Davis, Nicolas Decoster, Jérôme Fraleu, Rémi Gribonval, Wen-Liang Hwang, Stéphane Mallat, Jean François Muzy and Sifen Zhong. The following gives a list of current packages (in bold) with their main computational commands, and the sections they relate to.

Signal Section 3.3.

`s=` Arithmetic calculations over signals.

`fft` Forward and inverse fast Fourier transforms.

`conv` Fast convolution.

Wavelet Transform (1d) Sections 4.3 and 7.3.

`cwt` Continuous wavelet transform.

`owtd, owtr` Orthogonal and biorthogonal wavelet transforms, forward and reverse.

`wthresh` Wavelet coefficient thresholding.

Wavelet Transform Maxima (1d) Section 6.2.

`extrema, chain` Computes the maxima of a continuous wavelet transform, and chains them through scales.

Wavelet Transform Modulus Maxima Method (1d) Section 6.4.

`pf` Computes the partition functions and singularity spectra of multifractal signals.

Matching Pursuit Sections 4.2 and 9.5.2.

`stftd` Short time windowed Fourier transform.

`mp, mpr` Matching pursuit in a Gabor dictionary, forward and reverse.

Image Section 3.4.

`i=` Arithmetic operations over images.

Orthogonal Wavelet Transform (2d) Section 7.7.2.

`owt2d, owt2r` Orthogonal and biorthogonal wavelet transforms of images, forward and reverse.

Dyadic Wavelet Transform (2d) Section 6.3.

`dwt2d, dwt2r` Dyadic wavelet decomposition of images, forward and reverse.

`extrema2, extrecons2` Computes the modulus maxima of a dyadic wavelet transform, and reconstructs the image from these maxima.

`chain2` Computes the chains of modulus maxima corresponding to edges.

`denoise2` Denoising by thresholding the chains of modulus maxima.

Compression (2d) Section 11.4.2.

`code2, decode2` Image compression with a wavelet transform code, and reconstruction of the coded image.

B.3 FREEWARE WAVELET TOOLBOXES

We give a partial list of freeware toolboxes for wavelet signal processing that can retrieved over the Internet.

EMBEDDED IMAGE COMPRESSION is a C++ software for wavelet image compression (Amir Said and William Pearlman):

`http://ipl.rpi.edu/SPIHT`.

FRACLAB is wavelet fractal analysis toolbox developed at INRIA (Christophe Canus, Paulo Goncalvès, Bertrand Guiheneuf and Jacques Lévy Véhel):

`http://www-syntim.inria.fr/fractales/`.

MEGAWAVE is a collection of command line C subroutines under Unix for wavelet, wavelet packet and local cosine processing, with sound and image processing applications (Jacques Froment):

`http://www.ceremade.dauphine.fr/~mw`.

RICE WAVELET TOOLBOX is a wavelet Matlab toolbox with orthogonal and biorthogonal transforms and applications to denoising (DSP group at Rice university):

`http://www-dsp.rice.edu/software/RWT`.

SWAVE is an S+ tool box with continuous wavelet transforms and windowed Fourier transforms, including detection of ridges (René Carmona, Wen-Liang Hwang and Bruno Torrésani):

`http://chelsea.princeton.edu/~rcarmona/TFbook/`.

TIME-FREQUENCY is a Matlab toolbox for the analysis of non-stationary signals with quadratic time-frequency distributions (Francois Auger, Patrick Flandrin, Olivier Lemoine and Paulo Goncalvès):

`http://www.physique.ens-lyon.fr/ts/tftb.html`.

XWPL, WPLIB, DENOISE are libraries of subroutines that implement orthogonal signal decompositions in dictionaries of wavelet packet and local cosine bases, with applications to noise removal and signal compression (wavelet group at Yale University):

`http://pascal.math.yale.edu/pub/wavelets/software/`.

WAVELET TOOLBOX IN KHOROS includes orthogonal and biorthogonal wavelet transforms for multidimensional signals (Jonio Cavalcanti and Ramiro Jordon):

`http://www.khoral.com/obtain/contrib.html`.

Bibliography

BOOKS

[1] A. Akansu and R. Haddad. *Multiresolution Signal Decomposition*. Academic Press, 1993.

[2] A. Akansu and M. J. Smith, editors. *Subband and Wavelet Transforms*. Kluwer, 1995.

[3] A. Aldroubi and M. Unser, editors. *Wavelets in Medicine and Biology*. CRC Press, 1996.

[4] A. Antoniadis and G. Oppenheim, editors. *Wavelets and Statistics*. Springer, 1995.

[5] A. Arneodo, F. Argoul, E. Bacry, J. Elezgaray, and J. F. Muzy. *Ondelettes, Multifractales et Turbulences*. Diderot editeur, Paris, 1995.

[6] M. Barlaud, editor. *Wavelets in Image Communication*. Elsevier, 1995.

[7] M. F. Barnsley and L. P. Hurd. *Fractal Image Compression*. A K Peters, Wellesley, MA, 1993.

[8] J. J. Benedetto and M. W. Frazier, editors. *Wavelets: Mathematics and Applications*. CRC Press, Boca Raton, Ann Arbor, London, Tokyo, 1994.

[9] S. M. Berman. *Sojournes and Extremes of Stochastic Processes*. Wadsworth, Reading, MA, 1989.

[10] B. Boashash, editor. *Time-frequency Signal Analysis*. Wiley Halsted Press, 1992.

[11] C. S. Burrus and T. W. Parks. *DFT/FFT and Convolution Algorithms: Theory and Implementation*. John Wiley and Sons, New York, 1985.

[12] M. Cannone. *Ondelettes, Paraproduits et Navier-Stokes*. Diderot, Paris, 1995.

[13] R. Carmona, W. L. Hwang, and R. Torrésani. *Practical Time-Frequency Analysis*. Academic Press, New York, 1998.

[14] W. K. Chen. *Passive and Active Filters*. John Wiley and Sons, New York, 1986.

[15] C. K. Chui. *An Introduction to Wavelets*. Academic Press, New York, 1992.

[16] C. K. Chui, editor. *Wavelets: A Tutorial in Theory and Applications*. Academic Press, New York, 1992.

[17] A. Cohen and R. D. Ryan. *Wavelets and Multiscale Signal Processing*. Chapman and Hall, London, 1995.

[18] L. Cohen. *Time-frequency Analysis*. Prentice Hall, Englewoods Cliffs, 1995.

[19] J. M. Combes, A. Grossmann, and P. Tchamitchian, editors. *Wavelets time-frequency methods and phase space*. Springer-Verlag, Berlin, 1989.

[20] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley Interscience, 1991.

[21] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, Philadelphia, PA, 1992.

[22] R. DeVore and G. Lorentz. *Constructive Approximation*, volume 303 of *Comprehensive Studies in Mathematics*. Springer-Verlag, 1993.

[23] D. E. Dudgeon and R. M. Mersereau. *Multidimensional Digital Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ, 1984.

[24] H. Dym and H. P. McKean. *Fourier Series and Integrals*. Academic Press, New York, 1972.

[25] F. Feder. *Fractals*. Pergamon, New York, 1988.

[26] P. Flandrin. *Temps-Fréquence*. Hermes, Paris, 1993.

[27] A. Gersho and R. M. Gray. *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, Boston, 1992.

[28] P. E. Gill, W. Murray, and M. H. Wright. *Numerical Linear Algebra and Optimization*. Addison Wesley, Redwood City, CA, 1991.

[29] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins Univ. Press, 1989.

[30] E. Hernández and G. Weiss. *A First Course on Wavelets*. CRC Press, New York, 1996.

[31] M. Holschneider. *Wavelets: An Analysis Tool*. Oxford Mathematical Monographs, Clarendon Press, Oxford, 1995.

[32] B. Burke Hubbard. *The World According to Wavelets*. A K Peters, Wellesley, MA, 1996.

[33] S. Jaffard and Y. Meyer. *Wavelet Methods for Pointwise Regularity and Local Oscillations of Functions*, volume 123. American Mathematical Society, Providence, RI, 1996.

[34] A. K. Jain. *Fundamentals of Digital Image Processing*. Prentice-Hall, Englewood Cliffs, NJ, 1989.

[35] N. J. Jayant and P. Noll. *Digital Coding of Waveforms*. Prentice-Hall, Englewood-Cliffs, NJ, 1984.

[36] F. John. *Partial Differential Equations*. Springer-Verlag, New York, 1975.

[37] B. Joseph and R. L. Motard. *Wavelet Applications in Chemical Engineering*. Kluwer Academic Publishers, Boston, 1994.

[38] G. Kaiser. *A Friendly Guide to Wavelets*. Birkhäuser, 1994.

[39] G. Kanizsa. *Organization in Vision*. Praeger Scientific, New York, 1979.

[40] S. M. Kay. *Fundamentals of Statistical Signal Processing*. Prentice-Hall, Englewood Cliffs, 1993.

[41] P. G. Lemarié, editor. *Les Ondelettes en 1989*. Lecture Notes in Mathematics no. 1438. Springer-Verlag, Berlin, 1990.

[42] H. S. Malvar. *Signal Processing with Lapped Transforms*. Artech House, Norwood, MA, 1992.

[43] B. B. Mandelbrot. *The Fractal Geometry of Nature*. W. H. Freeman and Co., San Fransisco, 1982.

[44] D. Marr. *Vision*. W.H. Freeman and Co., San Fransisco, 1982.

[45] A. W. Marshall and I. Olkin. *Inequalities: Theory of Majorization and its Applications*. Academic Press, Boston, 1979.

[46] Y. Meyer. *Ondelettes et Algorithmes Concurrents*. Hermann, Paris, 1992.

[47] Y. Meyer. *Wavelets and Operators*. Advanced mathematics. Cambridge University Press, 1992.

[48] Y. Meyer. *Wavelets: Algorithms and Applications*. SIAM, 1993. Translated and revised by R. D. Ryan.

[49] Y. Meyer. *Wavelets, Vibrations and Scalings*. CRM, Université de Montréal, Montréal, 1997. Cours de la chaire Aisenstadt.

[50] Y. Meyer and S. Roques, editors. *Progress in Wavelet Analysis and Applications*. Frontières, 1993.

[51] H. J. Nussbaumer. *Fast Fourier Transform and Convolution Algorithms*. Springer-Verlag, Berlin, 1982.

[52] T. Ogden. *Essential Wavelets for Statistical Applications and Data Analysis*. Birkhauser, Boston, 1996.

[53] T. Qian and C. Chen. *Joint Time-Frequency Analysis: Method and Application*. Prentice Hall, Englewoods Cliffs, 1996.

[54] A. V. Oppenheim, A. S. Willsky, and I. T. Young. *Signals and Systems*. Prentice-Hall, Englewood Cliffs, NJ, 1997.

[55] A. V. Oppenheim and R. W. Shafer. *Discrete-Time Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ, 1989.

[56] A. Papoulis. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, New York, NY, second edition, 1984.

[57] A. Papoulis. *The Fourier Integral and its Applications*. McGraw-Hill, New York, NY, second edition, 1987.

[58] A. Papoulis. *Signal Analysis*. McGraw-Hill, New York, NY, 1988.

[59] B. Porat. *Digital Processing of Random Signals: Theory and Method*. Prentice-Hall, Englewood Cliffs, NJ, 1994.

[60] M. B. Priestley. *Spectral Analysis and Time Series*. Academic Press, Boston, 1981.

[61] L. R. Rabiner and R. W. Shafer. *Digital Signal Processing of Speech Signals*. Englewood Cliffs, NJ, 1978.

[62] A. Rosenfeld, editor. *Multiresolution Techniques in Computer Vision*. Springer-Verlag, New York, 1984.

[63] W. Rudin. *Real and Complex Analysis*. Mc Graw Hill, 1987.

[64] M. B. Ruskai *et al.*, editor. *Wavelets and their Applications*. Jones and Bartlett, Boston, 1992.

[65] D. J. Sakrison. *Communicaton Theory: Transmission of Waveforms and Digital Information*. John Wiley, New York, 1968.

[66] L. Schwartz. *Théorie Des Distributions*. Hermann, Paris, 1970.

[67] J. J. Slotine and W. Li. *Applied Nonlinear Control*. Prentice-Hall, Englewood Cliffs, NJ, 1991.

[68] G. Strang and T. Nguyen. *Wavelets and Filter Banks*. Wellesley-Cambridge Press, Boston, 1996.

[69] R. Strichartz. *A Guide to Distribution Theory and Fourier Transforms*. CRC Press, Boca Raton, 1994.

[70] L. Torres and M. Kunt, editors. *Video Coding - The Second Generation Approach.* Kluwer, 1996.

[71] B. Torrésani. *Analyse Continue par Ondelettes.* CNRS Editions, Paris, 1995.

[72] H. Triebel. *Theory of Function Spaces.* Birkhäuser Verlag, Boston, 1992.

[73] P. P. Vaidyanathan. *Multirate Systems and Filter Banks.* Prentice-Hall, Englewood Cliffs, NJ, 1993.

[74] M. Vetterli and J. Kovacevic. *Wavelets and Subband Coding.* Prentice-Hall, Englewood Cliffs, NJ, 1995.

[75] H. Weyl. *The Theory of Groups and Quantum Mechanics.* Dutton, New York, 1931.

[76] M. V. Wickerhauser. *Adapted Wavelet Analysis from Theory to Software.* AK Peters, 1994.

[77] J. W. Woods, editor. *Subband Image Coding.* Kluwer, Boston, MA, 1991.

[78] G. W. Wornell. *Signal Processing with Fractals: A Wavelet-Based Approach.* Prentice-Hall, 1995.

[79] W. P. Ziemer. *Weakly Differentiable Functions.* Springer-Verlag, 1989.

ARTICLES

[72] E. H. Adelson, E. Simoncelli, and R. Hingorani. Orthogonal pyramid transforms for image coding. In *Proc. SPIE*, volume 845, pages 50–58, Cambridge, MA, October 1987.

[73] A. N. Akansu, R. A. Haddad, and H. Caglar. The binomial QMF-wavelet transform for multiresolution signal decomposition. *IEEE Trans. on Signal Processing*, SP-40, 1992.

[74] A. Aldroubi and H. Feichtinger. Complete iterative reconstruction algorithms for irregularly sampled data in spline-like spaces. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, Munich, Germany, April 1997.

[75] A. Aldroubi and M. Unser. Families of multiresolution and wavelet spaces with optimal properties. *Numer. Functional Anal. and Optimization*, 14:417–446, 1993.

[76] J. Aloimonos and A. Rosenfeld. Computer vision. *Science*, 253:1249–1253, 1991.

[77] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *Int. J. Comp. Vision*, 1(2):283–310, 1989.

[78] R. Ansari and C. Guillemont. Exact reconstruction filter banks using diamond FIR filters. In *Proc. Bilkent Intl. Conf.*, pages 1412–1424, July 1990.

[79] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies. Image coding using wavelet transform. *IEEE Trans. Image Proc.*, 1(2):205–220, April 1992.

[80] A. Averbuch, G. Aharoni, R. Coifman, and M. Israeli. Local cosine transform – a method for the reduction of the blocking effect in JPEG. *J. of Math. Imaging and Vision*, 3:7–38, 1993.

[81] A. Averbuch, D. Lazar, and M. Israeli. Image compression using wavelet decomposition. *IEEE Trans. Image Proc.*, 5(1):4–15, 1996.

[82] P. M. Aziz, H. V. Sorensen, and J. Van Der Spiegel. An overview of sigma-delta converters. *IEEE Sig. Proc. Mag.*, 13(1):61–84, January 1996.

[83] E. Bacry, J. F. Muzy, and A. Arneodo. Singularity spectrum of fractal signals: exact results. *J. of Stat. Phys.*, 70(3/4):635–674, 1993.

[84] Z. Baharav, H. Krupnik, D. Malah, and E. Karnin. A multi-resolution framework for fractal image representation and its applications. Technical report, Electr. Eng., Technion, Israel, Haifa, 1996.

[85] R. Bajcsy. Computer description of textured surfaces. In *JCAI*, Stanford, CA, August 1973.

[86] R. Balian. Un principle d'incertitude en théorie du signal ou en mécanique quantique. *C. R. Acad. Sci. Paris*, 292, 1981. Série 2.

[87] J. L. Barron, D. J. Fleet and S. S. Beauchemin. Performance of optical flow techniques. *International Jour. on Computer Vision*, 12(1):43–77, 1994.

[88] M. Basseville and A. Benveniste A. S. Willsky. Multiscale autoregressive processes: Shur-Levinson parametrizations. *IEEE Trans. Signal Proc.*, 1992.

[89] G. Battle. A block spin construction of ondelettes. Part I: Lemarié functions. *Comm. Math. Phys.*, 110:601–615, 1987.

[90] M. Bayram and R. Baraniuk. Multiple window time-frequency analysis. In *Proc. of Time-Freq. and Time-Scale Symp.*, Paris, July 1996.

[91] J. J. Benedetto. Irregular sampling and frames. In C. K. Chui, editor, *Wavelets: A Tutorial in Theory and Applications*. Academic Press, New York, 1992.

[92] J. Berger, R. Coifman, and M. Goldberg. Removing noise from music using local trigonometric bases and wavelet packets. *J. Audio Eng. Soci.*, 42(10):808–818, October 1994.

[93] T. Berger, J. O. Stromberg. Exact reconstruction algorithms for the discrete wavelet tranform using spline wavelets. *J. of Appl. and Comput. Harmonic Analysis*, 2:392–397, 1995.

[94] Z. Berman and J. S. Baras. Properties of the multiscale maxima and zero-crossings representations. *IEEE Trans. Signal Proc.*, 41(12):3216–3231, December 1993.

[95] C. Bernard. Discrete wavelet analysis: a new framework for fast optical flow computations. subm. *IEEE Trans. Image Proc.*, 1999.

[96] M. Bertero. Linear inverse and ill-posed problems. In *Advances in Electronics and Electron Physics.*, Academic Press, NY 1989.

[97] G. Beylkin, R. Coifman, and V. Rokhlin. Fast wavelet transforms and numerical algorithms. *Comm. on Pure and Appl. Math.*, 44:141–183, 1991.

[98] M. J. Blake, P. P. Anandran. The robust estimation of multiple motions: parametric and piecewise smooth flow fields. *Computer Vision and Image Understanding*, 63(1):75–104, 1996.

[99] J. M. Bony. Two-microlocalization and propagation of singularities for semilinear hyperbolic equations. In *Proc. of Tanaguchi Symp.*, pages 11–49, HERT. Katata, October 1984.

[100] A. C. Bovik, N. Gopal, T. Emmoth, and A. Restrepo. Localized measurement of emergent image frequencies by Gabor wavelets. *IEEE Trans. Info. Theory*, 38(2):691–712, March 1992.

[101] A. C. Bovik, P. Maragos, and T. F. Quatieri. AM-FM energy detection and separation in noise using multiband energy operators. *IEEE Trans. Signal Proc.*, 41(12):3245–3265, December 1993.

[102] K. Brandenburg, G. Stoll, F. Dehery, and J. D. Johnstone. The ISO-MPEG-Audio codec: A generic-standard for coding of high quality digital audio. *J. Audio Eng. Soc.*, 42(10):780–792, October 1994.

[103] C. Brislawn. Fingerprints go digital. *Notices of the AMS*, 42(11):1278–1283, November 1995.

[104] A. Bruce, D. Donoho, and H. Y. Gao. Wavelet analysis. *IEEE Spectrum*, pages 26–35, October 1996.

[105] J. B. Buckheit and D. L. Donoho. Wavelab and reproducible research. In *Wavelets and Statistics*, pages 53–81. Springer-Verlag, Berlin, 1995. A. Antoniadis and G. Oppenheim editors.

[106] T. Burns, S. Rogers, D. Ruck, and M. Oxley. Discrete, spatio-temporal, wavelet multiresolution analysis method for computing optimal flow. *Optical Eng.*, 33(7):2236–2247, July 1994.

[107] P. J. Burt. Smart sensing within a pyramid vision machine. *Proc. IEEE*, 76(8):1006–1015, August 1988.

[108] P. J. Burt and E. H. Adelson. The Laplacian pyramid as a compact image code. *IEEE Trans. Commun.*, 31(4):532–540, April 1983.

[109] C. A. Cabrelli and U. M. Molter. Wavelet transform of the dilation equation. *J. of the Australian Math. Soc.*, 37, 1996.

[110] C. A. Cabrelli and U. M. Molter. Generalized self-similarity *J. of Math. Anal. and Appl.*, 230:251–260, 1999.

[111] A. P. Calderón. Intermediate spaces and interpolation, the complex method. *Stud. Math.*, 24:113–190, 1964.

[112] M. Cannon and J. J. Slotine. Space-frequency localized basis function networks for nonlinear system estimation and control. *Neurocomputing*, 9(3), 1995.

[113] J. Canny. A computational approach to edge detection. *IEEE Trans. Patt. Anal. and Mach. Intell.*, 36:961–1005, September 1986.

[114] L. Carleson. On the convergence and growth of partial sums of Fourier series. *Acta Math.*, 116:135–157, 1966.

[115] S Carlsson. Sketch based coding of grey level images. *Signal Processing*, 57–83, July 1988.

[116] R. Carmona. Extrema reconstruction and spline smoothing: variations on an algorithm of Mallat and Zhong. In *Wavelets and Statistics*, pages 96–108. Springer-Verlag, Berlin, 1995. A. Antoniadis and G. Oppenheim editors.

[117] R. Carmona, W. L. Hwang, and B. Torrésani. Identification of chirps with continuous wavelet transform. In *Wavelets and Statistics*, pages 96–108. Springer-Verlag, Berlin, 1995. A. Antoniadis and G. Oppenheim editors.

[118] A. S. Cavaretta, W. Dahmen, and C. Micchelli. Stationary subdivision. *Mem. Amer. Math. Soc.*, 93:1–186, 1991.

[119] S. Chen and D. Donoho. Atomic decomposition by basis pursuit. In *SPIE International Conference on Wavelets*, San Diego, July 1995.

[120] C. T. Chen. Video compression: standards and applications. *Jour. of Visual Communication and Image Representation*, 4(2):103–111, June 1993.

[121] Y. W. Choi and N. T. Thao. Implicit Coding for Very Low Bit Rate Image Compression. In *Proc. IEEE Int. Conf. Image Processing*, I:783–786, October 1998.

[122] H. I. Choi and W. J. Williams. Improved time-frequency representation of muticomponent signals using exponential kernels. *IEEE Trans. Acoust., Speech, and Signal Proc.*, 37(6):862–871, 1989.

[123] C. K. Chui and X. Shi. Characterization of fundamental scaling functions and wavelets. *Approx. Theory and its Appl.*, 1993.

[124] C. K. Chui and X. Shi. Inequalities of Littlewood-Paley type for frames and wavelets. *SIAM J. Math. Anal.*, 24(1):263–277, January 1993.

[125] C. K. Chui and J. Z. Wang. A cardinal spline approach to wavelets. *Proc. Amer. Math. Soc.*, 113:785–793, 1991.

[126] T. C. Claasen and W. F. Mecklenbrauker. The aliasing problem in discrete-time Wigner distribution. *IEEE Trans. Acoust., Speech, and Signal Proc.*, 31:1067–1072, 1983.

[127] J. Claerbout. Hypertext documents about reproducible research, 1994. http://sepwww.stanford.edu.

[128] A. Cohen. Ondelettes, analyses multirésolutions et filtres miroir en quadrature. *Ann. Inst. H. Poincaré, Anal. Non Linéaire*, 7:439–459, 1990.

[129] A. Cohen and J. P. Conze. Régularité des bases d'ondelettes et mesures ergodiques. Technical report, CEREMADE, Université Paris Dauphine, 1991.

[130] A. Cohen and I. Daubechies. On the instability of arbitrary biorthogonal wavelet packets. *SIAM J. of Math. Anal.*, 24(5):1340–1354, 1993.

[131] A. Cohen, I. Daubechies, and J.-C. Feauveau. Biorthogonal bases of compactly supported wavelets. *Commun. on Pure and Appl. Math.*, 45:485–560, 1992.

[132] A. Cohen, I. Daubechies, O. Guleryuz, and M. Orchard. On the importance of combining wavelet-based non-linear approximation with coding strategies. Submitted to *IEEE Trans. Info. Theory*, 1999.

[133] A. Cohen, R. DeVore, P. Pertrushev, and H. Xu. Non-linear approximation and the space $BV(\mathbb{R}^2)$. *American J. of Math.*, 1998.

[134] A. Cohen, I. Daubechies, and P. Vial. Wavelet bases on the interval and fast algorithms. *J. of Appl. and Comput. Harmonic Analysis*, 1:54–81, 1993.

[135] L. Cohen. Generalized phase-space distribution functions. *J. Math. Phys.*, 7(5):781–786, 1966.

[136] L. Cohen. Time-frequency distributions: A review. *Proc. IEEE*, 77(7):941–981, July 1989.

[137] R. R. Coifman and D. Donoho. Translation invariant de-noising. Technical Report 475, Dept. of Statistics, Stanford University, May 1995.

[138] R. R. Coifman and Y. Meyer. Remarques sur l'analyse de Fourier a fenêtre. *C.R. Acad. Sci.*, pages 259–261, 1991.

[139] R. R. Coifman, Y. Meyer, and M. V. Wickerhauser. Wavelet analysis and signal processing. In *Wavelets and their Applications*, pages 153–178, Boston, 1992. Jones and Barlett. B. Ruskai *et al.* editors.

[140] R. R. Coifman and M. V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Trans. Info. Theory*, 38(2):713–718, March 1992.

[141] A. Croisier, D. Esteban, and C. Galand. Perfect channel splitting by use of interpolation/-decimation/tree decomposition techniques. In *Int. Conf. on Info. Sciences and Systems*, pages 443–446, Patras, Greece, August 1976.

[142] Z. Cvetkovic and M. Vetterli. Consistent reconstruction of signals from wavelet extrema/zero crossings representation. *IEEE Trans. Signal Proc.*, March 1995.

[143] R. Dahlhaus. On the Kullback-Leibler information divergence of locally stationary processes. *Stoch. Proc. Appl.*, 1995.

[144] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Commun. on Pure and Appl. Math.*, 41:909–996, November 1988.

[145] I. Daubechies. The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans. Info. Theory*, 36(5):961–1005, September 1990.

[146] I. Daubechies, A. Grossmann, and Y. Meyer. Painless nonorthogonal expansions. *J. Math. Phys.*, 27:1271–1283, 1986.

[147] I. Daubechies and J. Lagarias. Two-scale difference equations: II. Local regularity, infinite products of matrices and fractals. *SIAM J. of Math. Anal.*, 24, 1992.

[148] I. Daubechies and W. Sweldens. Factoring wavelet transforms into lifting steps, 1996. *J. Fourier Analysis and Applications*, 4(3):245–267, 1998.

[149] J. G. Daugmann. Two-dimensional spectral analysis of cortical receptive field profile. *Vision Research*, 20:847–856, 1980.

[150] G. M. Davis. A wavelet-based analysis of fractal image compression. *IEEE Trans. on Image Proc.*, February 1998.

[151] G. M. Davis, S. Mallat, and M. Avelanedo. Greedy adaptive approximations. *J. of Constr. Approx.*, 13:57–98, 1997.

[152] G. M. Davis, S. Mallat, and Z. Zhang. Adaptive time-frequency decompositions. *SPIE J. of Opt. Engin.*, 33(7):2183–2191, July 1994.

[153] Y. F. Dehery, M. Lever, and P. Urcum. A MUSICAM source codec for digital audio broadcasting and storage. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, pages 3605–3608, Toronto, Canada, May 1991.

[154] N. Delprat, B. Escudié, P. Guillemain, R. Kronland-Martinet, P. Tchamitchian, and B. Torrésani. Asymptotic wavelet and Gabor analysis: extraction of instantaneous frequencies. *IEEE Trans. Info. Theory*, 38(2):644–664, March 1992.

[155] G. Deslauriers and S. Dubuc. Symmetric iterative interpolation. *Constr. Approx.*, 5:49–68, 1989.

[156] R. A. DeVore, B. Jawerth, and B. J. Lucier. Image compression through wavelet transform coding. *IEEE Trans. Info. Theory*, 38(2):719–746, March 1992.

[157] R. A. DeVore, B. Jawerth, and V. Popov. Compression of wavelet decompositions. *Americ. J. of Math.*, 114:737–785, 1992.

[158] R. A. DeVore, G. Kyriazis, and D. Leviatan. Wavelet compression and nonlinear n-widths. *Advances in Comput. Math.*, 1:197–214, 1993.

[159] R. A. DeVore and V. N. Temlyakov. Some remarks on greedy algorithms. *Advances in Comput. Math.*, 5:173–187, 1996.

[160] R. A. DeVore. Nonlinear approximation. *Acta Numerica*, 51–150, 1998.

[161] D. Donoho. Unconditional bases are optimal bases for data compression and for statistical estimation. *J. of Appl. and Comput. Harmonic Analysis*, 1:100–115, 1993.

[162] D. Donoho. Interpolating wavelet transforms. *J. of Appl. and Comput. Harmonic Analysis*, 1994.

[163] D. Donoho. Nonlinear solution of linear inverse problems by wavelet-vaguelette deconposition. *J. of Appl. and Comput. Harmonic Analysis*, 2(2):101–127, 1995.

[164] D. Donoho. Unconditional bases and bit-level compression. *J. of Appl. and Comput. Harmonic Analysis*, 3:388–392, 1996.

[165] D. Donoho. Wedgelets: nearly-minimax estimation of edges. Tech. Report. Statist. Depart., Stanford University, October 1997.

[166] D. Donoho and I. Johnstone. Ideal denoising in an orthonormal basis chosen from a library of bases. *C.R. Acad. Sci. Paris, Série I*, 319:1317–1322, 1994.

[167] D. Donoho and I. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81:425–455, December 1994.

[168] D. Donoho, I. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage: asymptopia? *J. of Royal Stat. Soc. B.*, 57(2):301–369, 1995.

[169] D. Donoho and I. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *J. American Statist. Assoc.*, 90:1200–1224, 1995.

[170] D. Donoho and I. Johnstone. Minimax estimation via wavelet shrinkage. *Annals of Statistics*, 1998.

[171] D. Donoho, R. C. Liu, and K. B. MacGibbon. Minimax risk over hyperrectangles, and implications. *Annals of Statistics*, 18:1416-1437, 1990.

[172] D. Donoho, S. Mallat, and R. von Sachs. Estimating covariances of locally stationary processes: consistency of best basis methods. In *Proc. of Time-Freq. and Time-Scale Symp.*, Paris, July 1996.

[173] D. Donoho, S. Mallat, and R. von Sachs. Estimating covariances of locally stationary processes: rates of convergence of best basis methods. submitted to *Annals of Statistics*, 1998.

[174] D. Donoho, M. Vetterli, R. A. DeVore, and I. Daubechies. Data compression and harmonic analysis. *IEEE Trans. Info. Theory*, 44(6):2435–2476, October 1998.

[175] R. J. Duffin and A. C. Schaeffer. A class of nonharmonic Fourier series. *Trans. Amer. Math. Soc.*, 72:341–366, 1952.

[176] P. Duhamel, Y. Mahieux, and J. Petit. A fast algorithm for the implementation of filter banks based on time domain aliasing cancellation. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, pages 2209–2212, Toronto, Canada, May 1991.

[177] P. Duhamel and M. Vetterli. Fast Fourier transforms: a tutorial review and a state of the art. *Signal Proc.*, 19(4):259–299, April 1990.

[178] N. Dyn and S. Rippa. Data-dependent triangulations for scattered data interpolation and finite element approximation. *Applied Num. Math.*, 12:89–105, 1993.

[179] M. Farge and M. Holschneider. Interpretation of two-dimensional turbulence spectrum in terms of singularity in the vortex cores. *Europhys. Lett.*, 15(7):737–743, 1990.

[180] M. Farge, N. Kevlahan, V. Perrier, and E. Goirand. Wavelets and turbulence. *Proc. IEEE*, 84(4):639–669, April 1996.

[181] P. Flandrin. Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Trans. Info. Theory*, 38(2):910–916, March 1992.

[182] M. Frazier and B. Jawerth. Decomposition of Besov spaces. *Indiana Univ. Math. J.*, 34:777–789, 1985.

[183] J. H. Friedman. Multivariate adaptive regression splines. *Annals of Stat.*, 19(1):1–141, 1991.

[184] J. H. Friedman and W. Stuetzle. Projection pursuit regression. *J. of Amer. Stat. Assoc.*, 76:817–823, 1981.

[185] U. Frisch and G. Parisi. *Turbulence and predictability in geophysical fluid dynamics and climate dynamics*, chapter Fully developed turbulence and intermittency, page 84. North-Holland, Amsterdam, 1985. M. Ghil, R. Benzi, and G. Parisi, editors.

[186] J. Froment and S. Mallat. Second generation compact image coding with wavelets. In C. K. Chui, editor, *Wavelets: A Tutorial in Theory and Applications*. Academic Press, New York, 1992.

[187] D. Gabor. Theory of communication. *J. IEE*, 93:429–457, 1946.

[188] H.Y. Gao. *Wavelet estimation of spectral densities in time series analysis*. PhD thesis, University of California, Berkeley, 1993.

[189] D. Geiger and K. Kumaran. Visual organization of illusory surfaces. In *4th European Conf. in Comp. Vision*, Cambridge, UK, 1996.

[190] J. Geronimo, D. Hardin, and P. R. Massupust. Fractal functions and wavelet expansions based on several functions. *J. of Approx. Theory*, 78:373–401, 1994.

[191] H. Gish and J. Pierce. Asymptotically efficient quantizing. *IEEE Trans. on Info. Theory*, 14:676–683, September 1968.

[192] P. Goupillaud, A. Grossman, and J. Morlet. Cycle-octave and related transforms in seismic signal analysis. *Geoexploration*, 23:85–102, 1984/85. Elsevier Science Pub.

[193] V. Goyal, M. Vetterli, and T. Nugyen. Quantized overcomplete expansions in $\mathbb{R}^N$: analysis, synthesis and algorithms. *IEEE Trans. on Info. Theory*, 44(1):16–31, January 1988.

[194] R. M. Gray. Quantization noise spectra. *IEEE Trans. Info. Theory*, pages 1220–1240, June 1990.

[195] R. Gribonval, P. Depalle, X. Rodet, E. Bacry, and S. Mallat. Sound signals decomposition using a high resolution matching pursuit. In *Proc. Int. Computer Music Conf. (ICMC'96)*, pages 293–296, August 1996.

[196] W. Grimson. Computational experiments with a feature based stereo algorithm. *IEEE Trans. Patt. Anal. and Mach. Intell.*, 7:17–34, January 1985.

[197] K. Gröchenig. *NATO ASI 1991 on Probabilistic and Stochastic Methods in Analysis and Applications*, chapter Sharp results on random sampling of band-limited function. Kluwer, 1992. J.S. Byrnes ed.

[198] K. Gröchenig. Acceleration of the frame algorithm. *IEEE Trans. Signal Proc.*, 41(12):3331–3340, December 1993.

[199] K. Gröchenig. Irregular sampling of wavelet and short-time Fourier transforms. *Constr. Approx.*, 9:283–297, 1993.

[200] A. Grossmann and J. Morlet. Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM J. of Math. Anal.*, 15(4):723–736, July 1984.

[201] P. Guillemain and R. Kronland-Martinet. Characterization of acoustic signals through continuous linear time-frequency representations. *Proc. IEEE*, 84(2):561–585, April 1996.

[202] A. Haar. Zur theorie der orthogonalen funktionensysteme. *Math. Annal.*, 69:331–371, 1910.

[203] T. Halsey, M. Jensen, L. Kadanoff, I. Procaccia, and B. Shraiman. Fractal measures and their singularities: The characterization of strange sets. *Phys. Rev. A*, 33(2):1141–1151, 1986.

[204] F. J. Harris. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proc. IEEE*, pages 11–33, January 1978.

[205] D. M. Healy and J. B. Weaver. Two applications of wavelet transforms in magnetic resonance imaging. *IEEE Trans. Info. Theory*, 38(2):840–860, March 1992.

[206] D. M. Heeger. Optical flow using spatio-temporal filters. *Internat. J. of Computer Vision*, 1:279–302, 1988.

[207] C. Heil and D. Walnut. Continuous and discrete wavelet transforms. *SIAM Rev.*, 31:628–666, 1989.

[208] C. Herley, J. Kovačević, K. Ramchandran, and M. Vetterli. Tilings of the time-frequency plane: construction of arbitrary orthogonal bases and fast tiling algorithms. *IEEE Trans. Signal Proc.*, 41(12):3341–3359, December 1993.

[209] C. Herley and M. Vetterli. Wavelets and recursive filter banks. *IEEE Trans. Signal Proc.*, 41(8):2536–2556, August 1993.

[210] F. Hlawatsch and F. Boudreaux-Bartels. Linear and quadratic time-frequency signal representations. *IEEE Sig. Proc. Mag.*, 9(2):21–67, April 1992.

[211] F. Hlawatsch and P. Flandrin. *The Wigner Distribution-Theory and Applications in Signal Processing*, chapter The interference structure of the Wigner distribution and related time-frequency signal representations. Elsevier, Amsterdam, 1993. W.F.G. Mecklenbrauker ed.

[212] M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian. *Wavelets, Time-Frequency Methods and Phase Space*, chapter A Real-Time Algorithm for Signal Analysis with the Help of the Wavelet Transform, pages 289–297. Springer-Verlag, Berlin, 1989.

[213] M. Holschneider and P. Tchamitchian. Pointwise analysis of Riemann's nondifferentiable function. *Inventiones Mathematicae*, 105:157–176, 1991.

[214] B. K. P. Horn and B. G. Schunk. Determining optical flow. *Aritificial Intelligence*, 17:185–204, 1981.

[215] D. Hubel and T. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. of Physiology*, 160, 1962.

[216] D. Huffman. A method for the construction of minimum redundancy codes. *Proc. of IRE*, 40:1098–1101, 1952.

[217] B. Hummel and R. Moniot. Reconstruction from zero-crossings in scale-space. *IEEE Trans. Acoust., Speech, and Signal Proc.*, 37(12), December 1989.

[218] W. L. Hwang and S. Mallat. Characterization of self-similar multifractals with wavelet maxima. *J. of Appl. and Comput. Harmonic Analysis*, 1:316–328, 1994.

[219] I. A. Ibragimov and R. Z. Khas'minskii. Bounds for the risks of non parametric regression estimates. *Theory of Probability and its Applications*, 27:84–99, 1984.

[220] S. Jaffard. Pointwise smoothness, two-microlocalization and wavelet coefficients. *Publicacions Matemàtiques*, 35:155–168, 1991.

[221] S. Jaffard. Wavelet methods for fast resolution of elliptic problems. *SIAM J. of Numerical Analysis*, 29(4):965–986, August 1992.

[222] S. Jaffard. Multifractal formalism for functions parts I and II. *SIAM J. of Mathematical Analysis*, 28(4):944–998, 1997.

[223] S. Jaggi, W. C. Karl, S. Mallat, and A. S. Willsky. High resolution pursuit for feature extraction. *J. of Appl. and Comput. Harmonic Analysis*, 5:428–449, 1998.

[224] A. K. Jain and F. Farrokhnia. Unsupervised texture segmentation using Gabor filters. *Patt. Recogn.*, 24(12):1167–1186, 1991.

[225] N. Jayant. Signal compression: technology targets and research directions. *IEEE J. on Sel. Areas in Commun.*, 10(5):796–818, June 1992.

[226] N. J. Jayant, J. Johnstone, and B. Safranek. Signal compression based on models of human perception. *Proc. IEEE*, 81(10):1385–1422, October 1993.

[227] I. M. Johnstone. Function estimation and wavelets. Lecture Notes, Dept. of Statistics, Stanford University, 1999.

[228] I. M. Johnstone and B. W. Silverman. Wavelet threshold estimators for data with correlated noise. Technical report, Dept. of Statistics, Stanford University, December 1995.

[229] P. Jonathon Phillips. Matching Pursuit filters Applied to Face Identification. *IEEE Trans. Image Proc.*, 7(8):1150–1164, August 1998.

[230] L. K. Jones. On a conjecture of Huber concerning the convergence of projection pursuit regression. *Annals of Stat.*, 15(2):880–882, 1987.

[231] B. Julesz. Textons, the elements of texture perception and their interactions. *Nature*, 290, March 1981.

[232] J. Kalifa and S. Mallat. Minimax deconvolutions in mirror wavelet bases. Tech. Rep., CMAP, Ecole Polytechnique, 1999.

[233] J. Kalifa and S. Mallat. Minimax restoration and deconvolution. In P. Muller and B. Vidakovic, editors, *Bayesian Inference in Wavelet Based Models*. Springer-Verlag, 1999.

[234] N. K. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4, 1984.

[235] C. J. Kicey and C. J. Lennard. Unique reconstruction of band-limited signals by a Mallat-Zhong wavelet transform algorithm. *Fourier Analysis and Appl.*, 3(1):63–82, 1997.

[236] J. J. Koenderink. *Biological Cybernetics*, chapter The structure of images, pages 360–370. Springer-Verlag, New York, 1984. Y. Meyer, ed.

[237] A. N. Kolmogorov. The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *C.R. Acad. Sc. USSR*, 31(4):538–540, 1941.

[238] A. N. Kolmogorov. A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number. *J. Fluid Mech.*, 13:82–85, 1962.

[239] J. Kovačević and M. Vetterli. Nonseparable multidimensional perfect reconstruction filter banks and wavelet bases for $\mathcal{R}^n$. *IEEE Trans. Info. Theory*, 38(2):533–555, March 1992.

[240] M. A. Krasner. The critical band coder-digital encoding of speech signals based on the perceptual requirements of the auditor system. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, pages 327–331, Denver, CO, 1990.

[241] H. Krim, D. Donoho, and S. Mallat. Best basis algorithm for signal enhancement. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, 1995.

[242] H. Krim and J. C. Pesquet. On the statistics of best bases criteria. In *Wavelets and Statistics*, pages 193–205. Springer-Verlag, Berlin, 1995. A. Antoniadis and G. Oppenheim editors.

[243] M. Kunt, A. Ikonomopoulos, and M. Kocher. Second generation image coding techniques. *Proceedings of the IEEE*, 73(4):549–575, April 1985.

[244] A. Laine and J. Fan. Frame representations for texture segmentation. *IEEE Trans. Image Proc.*, 5(5):771–780, 1996.

[245] J. Laroche, Y. Stylianos, and E. Moulines. HNS: speech modification based on a harmonic plus noise model. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, Minneapolis, Minnesota, April 1993.

[246] W. Lawton. Tight frames of compactly supported wavelets. *J. Math. Phys.*, 31:1898–1901, 1990.

[247] W. Lawton. Necessary and sufficient conditions for constructing orthonormal wavelet bases. *J. Math. Phys.*, 32:57–61, 1991.

[248] D. LeGall. MPEG: a video compression standard for multimedia applications. *Communications of the ACM*, 34(4):46–58, April 1991.

[249] P. G. Lemarié. Ondelettes à localisation exponentielle. *J. Math. Pures et Appl.*, 67:227–236, 1988.

[250] A. S. Lewis and G. Knowles. Image compression using the 2-D wavelet transform. *IEEE Trans. Image Proc.*, 1(2):244–250, April 1992.

[251] J. M. Lina and M. Mayrand. Complex Daubechies wavelets. *J. of Appl. and Comput. Harmonic Analysis*, 2:219–229, 1995.

[252] I. J. Lustig, R. E. Marsten, and D. F. Shanno. Interior point methods for linear programming: computational state of the art. *ORSA J. on Comput.*, 6(1):1–14, 1994.

[253] S. Mallat. An efficient image representation for multiscale analysis. In *Proc. of Machine Vision Conference*, Lake Taho, February 1987.

[254] S. Mallat. Multiresolution approximations and wavelet orthonormal bases of $\mathbf{L}^2(\mathbb{R})$. *Trans. Amer. Math. Soc.*, 315:69–87, September 1989.

[255] S. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Patt. Anal. and Mach. Intell.*, 11(7):674–693, July 1989.

[256] S. Mallat. Minimax distortion rate for image tranform coding. Submitted to *IEEE Trans. Image Proc.*, 1999.

[257] S. Mallat and F. Falzon. Analysis of low bit rate image transform coding. *IEEE Trans. Signal Proc.*, 46(4), April 1998.

[258] S. Mallat and W. L. Hwang. Singularity detection and processing with wavelets. *IEEE Trans. Info. Theory*, 38(2):617–643, March 1992.

[259] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, December 1993.

[260] S. Mallat, Z. Zhang, and G. Papanicolaou. Adaptive covariance estimation of locally stationary processes. *Annals of Stat.*, 26(1):1–47, 1998.

[261] S. Mallat and S. Zhong. Characterization of signals from multiscale edges. *IEEE Trans. Patt. Anal. and Mach. Intell.*, 14(7):710–732, July 1992.

[262] H. S. Malvar. The LOT: A link between block transform coding and multirate filter banks. In *Proc. IEEE Int. Symp. Circ. and Syst.*, pages 835–838, Espoo, Finland, June 1988.

[263] H. S. Malvar and D. H. Staelin. The LOT: transform coding without blocking effects. *IEEE Trans. Acoust., Speech, and Signal Proc.*, 37(4):553–559, April 1989.

[264] B. B. Mandelbrot. Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of carrier. *J. Fluid. Mech.*, 62:331–358, 1975.

[265] B. B. Mandelbrot and J. W. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Rev.*, 10:422–437, 1968.

[266] W. Martin and P. Flandrin. Wigner-Ville spectral analysis of non-stationary processes. *IEEE Trans. Acoust., Speech, and Signal Proc.*, 33(6):1461–1470, December 1985.

[267] G. Matz, F. Hlawatsch, and W. Kozek. Generalized evolutionary spectral analysis and the Weyl spectrum of nonstationary random processes, 1995. Tech. Rep. 95-04, Dept. of Communications, Vienna University of Tech.

[268] M. R. McClure and L. Carin. Matching Pursuits with a Wave-Based Dictionary. *IEEE Trans. Signal Proc.*, 45(12):2912–2927, December 1997.

[269] F. Meyer, A. Averbuch, and R. Coifman. Multilayered image representation: application to image compression. Submitted to *IEEE Trans. Image Proc.*, 1998.

[270] Y. Meyer. Principe d'incertitude, bases hilbertiennes et algèbres d'opérateurs. In *Séminaire Bourbaki*, volume 662, Paris, 1986.

[271] E. Miller and A. S. Willsky. A multiscale approach to sensor fusion and the solution of linear inverse problems. *J. of Appl. and Comput. Harmonic Analysis*, 2(2):127–147, 1995.

[272] F. Mintzer. Filters for distortion-free two-band multirate filter banks. *IEEE Trans. Acoust., Speech, and Signal Proc.*, 33(3):626–630, June 1985.

[273] A. Moffat. Linear time adaptive arithmetic coding. *IEEE Trans. Info. Theory*, 36(2):401–406, March 1990.

[274] P. Moulin. A wavelet regularization method for diffuse radar target imaging and speckle noise reduction. *J. Math. Imaging and Vision*, pages 123–134, 1993.

[275] J. E. Moyal. Quantum mechanics as a statistical theory. In *Proc. Cambridge Phi. Soci.*, volume 45, pages 99–124, 1949.

[276] H. G. Müller and U. Stadtmuller. Variable bandwidth kernel estimators of regression curves. *Annals of Stat.*, 15:182–201, 1987.

[277] S. C. Zhu, and D. Mumford. Prior learning and Gibbs reaction-diffusion. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 19:1236–1250, November 1997.

[278] J. F. Muzy, E. Bacry, and A. Arneodo. The multifractal formalism revisited with wavelets. *Int. J. of Bifurcation and Chaos*, 4:245, 1994.

[279] R. Neff and A. Zakhor. Very low bit-rate video coding based on matching pursuit. *IEEE Trans. on Circuit Syst. for Video Tech.*, 7(1):158–171, February 1997.

[280] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In 27^{th} *Asilomar Conf. on Signals, Systems and Comput.*, November 1993.

[281] J. C. Pesquet, H. Krim, H. Carfantan, and J. G. Proakis. Estimation of noisy signals using time-invariant wavelet packets. In *Asilomar Conf. on Signals, Systems and Comput.*, November 1993.

[282] M. Pinsker. Optimal filtering of square integrable signals in Gaussian white noise. *Problems in Information Transmission.*, 16:120-133, 1980.

[283] D. A. Pollen and S. F. Ronner. Visual cortical neurons as localized spatial frequency filter. *IEEE Trans. Syst., Man, Cybern.*, 13, September 1983.

[284] M. Porat and Y. Zeevi. The generalized Gabor scheme of image representation in biological and maching vision. *IEEE Trans. Patt. Anal. and Mach. Intell.*, 10(4):452–468, July 1988.

[285] M. Porat and Y. Zeevi. Localized texture processing in vision: analysis and synthesis in Gaborian space. *IEEE Trans. Biomed. Eng.*, 36(1):115–129, January 1989.

[286] M. B. Priestley. Evolutionary spectra and non-stationary processes. *J. Roy. Statist. Soc. Ser. B*, 27:204–229, 1965.

[287] S. Qian and D. Chen. Signal representation via adaptive normalized Gaussian functions. *IEEE Trans. Signal Proc.*, 36(1), January 1994.

[288] K. Ramchandran and M. Vetterli. Best wavelet packet bases in a rate-distortion sense. *IEEE Trans. Image Proc.*, 2(2):160–175, April 1993.

[289] K. S. Riedel. Optimal data-based kernel estimation of evolutionary spectra. *IEEE Trans. Signal Proc.*, 41(7):2439–2447, July 1993.

[290] O. Rioul. Regular wavelets: A discrete-time approach. *IEEE Trans. on Signal Proc.*, 41(12):3572–3578, December 1993.

[291] O. Rioul and P. Duhamel. Fast algorithms for discrete and continous wavelet transforms. *IEEE Trans. Info. Theory*, 38(2):569–586, March 1992.

[292] O. Rioul and M. Vetterli. Wavelets and signal processing. *IEEE Sig. Proc. Mag.*, 8(4):14–38, October 1991.

[293] S. Rippa. Long and thin triangles can be good for linear interpolation. *SIAM J. Numer. Anal.*, 29(1):257–270, February 1992.

[294] J. Rissanen and G. Langdon. Arithmetic coding. *IBM Journal of Research and Development*, 23(2):149–162, 1979.

[295] J. Rissanen and G. Langdon. Universal modeling and coding. *IEEE Trans. Info. Theory*, 27(1):12–23, January 1981.

[296] X. Rodet. *Spoken Language Generation and Understanding*, chapter Time-domain formant-wave function synthesis. D. Reidel Publishing, 1980.

[297] X. Rodet and P. Depalle. A new additive synthesis method using inverse fourier transform and spectral envelopes. In *Proc. ICMC*, San Jose, October 1992.

[298] A. Rosenfeld and M. Thurston. Edge and curve detection for visual scene analysis. *IEEE Trans. Comput.*, C-29, 1971.

[299] B. Rougé. Remarks about space-frequency and space-scale representations to clean and restore noisy images in satellite frameworks. In Y. Meyer and S. Roques, editors, *Progress in wavelet analysis and applications*. Frontières, 1993.

[300] B. Rougé. *Théorie de la Chaine Image Optique et Restauration*. Université Paris-Dauphine, 1997. Thèse d'habilitation.

[301] A. Said and W. A. Pearlman. A new, fast, and efficient image codec based on set partitioning in hierarchical trees. *IEEE Trans. on Circ. and Syst. for Video Tech.*, 6(3):243–250, June 1996.

[302] N. Saito and G. Beylkin. Multiresolution representation using the auto-correlation functions of compactly supported wavelets. *IEEE Trans. on Signal Proc.*, 41(12):3584–3590, December 1993.

[303] A. M. Sayeed and D. L. Jones. Optimal kernels for nonstationary spectral estimation. *IEEE Transactions on Signal Processing*, 43(2):478–491, February 1995.

[304] B. Scharf. Critical bands. In *Foundations in Modern Auditory Theory*, pages 150–202. Academic, New York, 1970.

[305] E. Schwartz. Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to perceptual coding. *Vision Research*, 20:665, 1980.

[306] C. E. Shannon. Communications in the presence of noise. In *Proc. of the IRE*, volume 37, pages 10–21, January 1949.

[307] J. M. Shapiro. Embedded image coding using zerotrees of wavelet coefficients. *IEEE Trans. Signal Proc.*, 41(12):3445–3462, December 1993.

[308] M. J. Shensa. The discrete wavelet transform: Wedding the à trous and Mallat algorithms. *IEEE Trans. Signal Proc.*, 40(10):2464–2482, October 1992.

[309] E. P. Simoncelli. Bayesian multiscale differential optical flow. In Hassecker, Jahne and Geissler, editors, *Handbook of computer vision and applications*. Academic Press, 1998.

[310] E. P. Simoncelli and E. H. Adelson. Nonseparable extensions of quadrature mirror filters to multiple dimensions. *Proc. IEEE*, 78(4):652–664, April 1990.

[311] E. P. Simoncelli and R. W. Buccigrossi. Embedded wavelet image compression based on joint probability model. In *Proc. IEEE Int. Conf. Image Proc.*, October 1997.

[312] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger. Shiftable multiscale transforms. *IEEE Trans. Info. Theory*, 38(2):587–607, March 1992.

[313] E. Simoncelli., and J. Portilla. Texture characterization via second-order statistics of wavelet coefficient amplitudes. *Proc. of 5th IEEE Int. Conf. on Image Proc.*, Chicago, October 1998.

[314] D. Sinha and A. H. Tewfik. Low bit rate transparent audio compression using adapted wavelets. *IEEE Trans. Signal Proc.*, 41(12):3463–3479, December 1993.

[315] T. Sikora. MPEG digital video coding standards. In R. Jurgens editor, *Digital Electronics Consumer Handbook*, MacGraw Hill Company, 1997.

[316] M. J. Smith and T. P. Barnwell III. A procedure for designing exact reconstruction filter banks for tree structured sub-band coders. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, San Diego, CA, March 1984.

[317] M. J. Smith and T. P. Barnwell III. Exact reconstruction for tree-structured subband coders. *IEEE Trans. Acoust., Speech, and Signal Proc.*, 34(3):431–441, June 1986.

[318] J. L. Starck and A. Bijaoui. Filtering and deconvolution by the wavelet transform. *Signal Processing*, 35:195–211, 1994.

[319] C. Stein. Estimation of the mean of a multivariate normal distribution. *Annals of Statistics*, 9:1135–1151, 1981.

[320] G. Strang and G. Fix. A Fourier analysis of the finite element variational method. *Construct. Aspects of Funct. Anal.*, pages 796–830, 1971.

[321] G. Strang and V. Strela. Short wavelets and matrix dilation equations. *IEEE Trans. Signal Proc.*, 43:108–115, 1995.

[322] J. O. Strömberg. A modified Franklin system and higher order spline systems on $\mathbb{R}^n$ as unconditional bases for Hardy spaces. In W. Beckner, A. P. Calderón, R. Fefferman, and P. W. Jones, editors, *Proc. Conf. in Honor of Antoni Zygmund*, volume II, pages 475–493, New York, 1981. Wadsworth.

[323] F. O'Sullivan. A statistical perspective on ill-posed inverse problems. *Statist. Sci.*, 1:502–527, 1986.

[324] W. Sweldens. The lifting scheme: a custom-design construction of biorthogonal wavelets. *J. of Appl. and Comput. Harmonic Analysis*, 3(2):186–200, 1996.

[325] W. Sweldens. The lifting scheme: a construction of second generation wavelets. *SIAM J. of Math. Analysis*, 29(2):511–546, 1997.

[326] W. Sweldens and P. Schröder. Spherical wavelets: efficiently representing functions on the sphere. In *Computer Graphics Proc. (SIGGRAPH 95)*, 161–172, 1995.

[327] P. Tchamitchian. Biorthogonalité et théorie des opérateurs. *Revista Mathemática Iberoamericana*, 3(2):163–189, 1987.

[328] P. Tchamitchian and B. Torrésani. Ridge and skeleton extraction from the wavelet transform. In *Wavelets and their Applications*, pages 123–151. Jones and Bartlett, Boston, 1992. B. Ruskai *et al.* editors.

[329] A. Teolis and J. Benedetto. Local frames and noise reduction. *Signal Processing*, 45:369–387, 1995.

[330] N. T. Thao and M. Vetterli. Deterministic analysis of oversampled a/d conversion and decoding improvement based on consistent estimates. *IEEE Trans. Signal Proc.*, 42:519–531, March 1994.

[331] D. J. Thomson. Spectrum estimation and harmonic analysis. *Proc. IEEE*, 70:1055–1096, 1982.

[332] B. Torrésani. Wavelets associated with representations of the affine Weil-Heisenberg group. *J. Math. Physics*, 32:1273, 1991.

[333] T. Tremain. The government standard linear predictive coding algorithm: LPC-10. *Speech Technol.*, 1(2):40–49, April 1982.

[334] M. Unser. Texture classification and segmentation using wavelet frames. *IEEE Trans. Image Proc.*, 4(11):1549–1560, November 1995.

[335] M. Unser and A. Aldroubi. A general sampling theory for nonideal acquisition device. *IEEE Trans. Signal Proc.*, 42(11):2915–2925, November 1994.

[336] P. P. Vaidyanathan. Quadrature mirror filter banks, M-band extensions and perfect reconstruction techniques. *IEEE ASSP Mag.*, 4(3):4–20, July 1987.

[337] P. P. Vaidyanathan and P. Q. Hoang. Lattice structures for optimal design and robust implementation of two-channel perfect reconstruction filter banks. *IEEE Trans. Acoust., Speech, and Signal Proc.*, 36(1):81–94, January 1988.

[338] M. Vetterli. Splitting a signal into subsampled channels allowing perfect reconstruction. In *Proc. IASTED Conf. on Appl. Sig. Proc. and Dig. Filt.*, Paris, June 1985.

[339] M. Vetterli. Filter banks allowing perfect reconstruction. *Signal Proc.*, 10(3):219–244, April 1986.

[340] M. Vetterli and C. Herley. Wavelets and filter banks: Theory and design. *IEEE Trans. Signal Proc.*, 40(9):2207–2232, September 1992.

[341] M. Vetterli and K.M. Uz. Multiresolution coding techniques for digital video: a review. *Video Signals, Multidimensional Systems and Signal Processing*, 3:161–187, 1992.

[342] J. Ville. Theorie et applications de la notion de signal analytique. *Cables et Transm.*, 2A(1):61–74, 1948.

[343] C. De Vleeschouwer and B. Macq. Subband dictionaries for low cost matching pursuits of video residues. To appear in *IEEE Trans. on Circuit Syst. for Video Tech.*, 1999.

[344] R. von Sachs and K. Schneider. Wavelet smoothing of evolutionary spectra by nonlinear thresholding. *J. of Appl. and Comput. Harmonic Analysis*, 3(3):268–282, July 1996.

[345] G. K. Wallace. The JPEG still picture compression standard. *Communications of the ACM*, 34(4):30–44, April 1991.

[346] A. Wang. Fast algorithms for the discrete wavelet transform and for the discrete Fourier transform. *IEEE Trans. Acoust., Speech, and Signal Proc.*, 32:803–816, August 1984.

[347] A. Watson, G. Yang, J. Soloman and J. Villasenor. Visual thresholds for wavelet quantization error. *Proceedings of the SPIE*, 2657:382–392, 1996.

[348] J. Weber and J. Malik. Robust computation of optical flow in a multi-scale differential framework. *International Jour. of Computer Vision*, 14(1):5–19, 1995.

[349] J. Whittaker. Interpolatory function theory. *Cambridge Tracts in Math. and Math. Physics*, 33, 1935.

[350] M. V. Wickerhauser. Acoustic signal compression with wavelet packets. In C. K. Chui, editor, *Wavelets: A Tutorial in Theory and Applications*. Academic Press, New York, 1992.

[351] E. P. Wigner. On the quantum correction for thermodynamic equilibrium. *Phys. Rev.*, 40:749–759, 1932.

[352] E. P. Wigner. *Perspective in Quantum Theory*, chapter Quantum-mechanical distribution functions revisited. Dover, Boston, MA, 1971. W. Yourgrau, A. van der Merwe, editors.

[353] K. G. Wilson. Generalized Wannier functions. Technical report, Cornell University, 1987.

[354] A. Witkin. Scale space filtering. In *Proc. Int. Joint. Conf. Artificial Intell.*, Espoo, Finland, June 1983.

[355] I. Witten, R. Neal, and J. Cleary. Arithmetic coding for data compression. *Comm. of the ACM*, 30(6):519–540, 1987.

[356] J. W. Woods and S. D. O'Neil. Sub-band coding of images. *IEEE Trans. Acoust., Speech, and Signal Proc.*, 34(5):1278–1288, May 1986.

[357] G. W. Wornell and A. V. Oppenheim. Wavelet-based representations for a class of self-similar signals with application to fractal modulation. *IEEE Trans. Info. Theory*, 38(2):785–800, March 1992.

[358] Z. X. Xiong, O. Guleryuz, and M. T. Orchard. Embedded image coding based on DCT. In *VCIP in Europeen Image Proc. Conf.*, 1997.

[359] A. Yuille and T. Poggio. Scaling theorems for zero-crossings. *IEEE Trans. Patt. Anal. and Mach. Intell.*, 8, January 1986.

[360] M. Zibulski, V. Segalescu, N. Cohen, and Y. Zeevi. Frame analysis of irregular periodic sampling of signals and their derivatives. *J. of Fourier Analysis and Appl.*, 42:453–471, 1996.

[361] M. Zibulski and Y. Zeevi. Frame analysis of the discrete Gabor-scheme analysis. *IEEE Trans. Signal Proc.*, 42:942–945, April 1994.

Index

Shelving: Applied Mathematics/Engineering

a wavelet tour of signal processing

This book is intended to serve as an invaluable reference for anyone concerned with the application of wavelets to signal processing. It has evolved from material used to teach "wavelet signal processing" courses in electrical engineering departments at Massachusetts Institute of Technology and Tel Aviv University, as well as applied mathematics departments at the Courant Institute of New York University and École Polytechnique in Paris.

New in the second edition

- Optical flow calculation and video compression algorithms.
- Image models with bounded variation functions.
- Bayes and Minimax theories for signal estimation.
- 200 pages rewritten and most illustrations redrawn.
- More problems and topics for a graduate course in wavelet signal processing, in electrical engineering and applied mathematics.

"Mallat has not only written a treatise, but also an excellent graduate text for students in computer science, electrical engineering, and mathematics" *Professor John J. Benedetto, SIAM review.*

- Provides a broad perspective on the principles and applications of transient signal processing with wavelets.
- Emphasizes intuitive understanding, while providing the mathematical foundations and description of fast algorithms.
- Numerous examples of real applications to noise removal, deconvolution, audio and image compression, singularity and edge detection, multifractal analysis, and time-varying frequency measurements.
- Algorithms and numerical examples are implemented in Wavelab, which is a Matlab toolbox freely available over the Internet.
- Content is accessible on several levels of complexity, depending on the individual reader's needs.
  - Reviews Fourier analysis and elementary signal processing
  - Introduces windowed Fourier transforms, continuous wavelet transforms, and Wigner-Ville transforms
  - Explains the construction of frames, wavelet orthogonal and biorthogonal bases, wavelet packet and local cosine bases
  - Covers basic approximation theory with applications to signal estimation and transform coding.

About the Author

Stèphane Mallat is a Professor in the Computer Science Department of the Courant Institute of Mathematical Sciences at New York University and a Professor in the Applied Mathematics Department at École Polytechnique, Paris, France. He has been a visiting professor in the Electrical Engineering Department at Massachusetts Institute of Technology and in the Applied Mathematics Department at the University of Tel Aviv.

Dr Mallat received the 1990 IEEE Signal Processing Society's paper award, the 1993 Alfred Sloan fellowship in Mathematics, the 1997 Outstanding Achievement Award from the SPIE Optical Engineering Society, and the 1997 Blaise Pascal Prize in applied mathematics, from the French Academy of Sciences.

ACADEMIC PRESS

A Harcourt Science and Technology Company

San Diego San Francisco New York Boston London Sydney Tokyo

ISBN 0-12-466606-X

90000

6 08628 66606 89 780124 666061