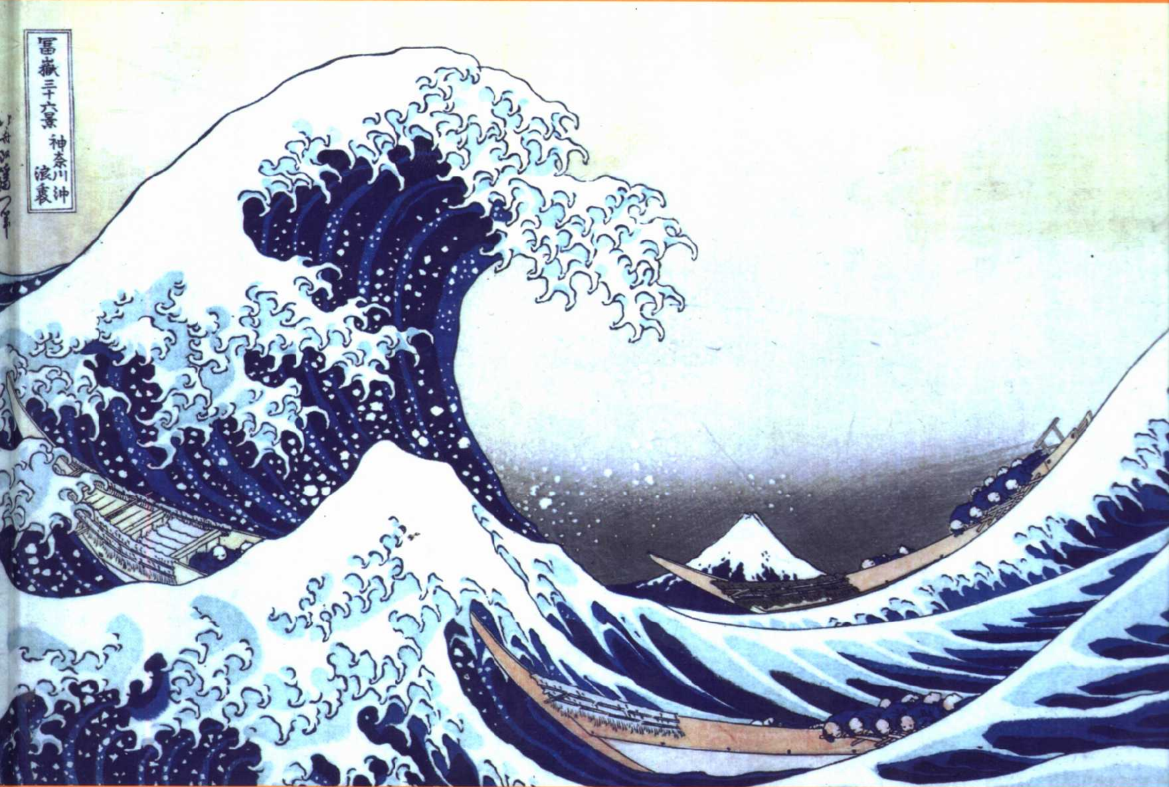
a wavelet tour of signal processing



Second Edition

Stéphane Mallat

AP

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. ♾

Reprinted 2001

ACADEMIC PRESS
A Harcourt Science and Technology Company
525 B Street, Suite 1900, San Diego, CA 92101-4495, USA
1300 Boylston Street, Chestnut Hill, MA 02167, USA
http://www.academicpress.com

Academic Press
A Harcourt Science and Technology Company
Harcourt Place, 32 Jamestown Road, London NW1 7BY, UK
http://www.academicpress.com

ISBN 0-12-466606-X

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

Library of Congress Catalog Card Number: 99-65087

Produced by HWA Text and Data Management, Tunbridge Wells
Printed in the United Kingdom at the University Press, Cambridge
01 02 03 04 CU 9 8 7 6 5 4 3 2

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).

Contents

II

FOURIER KINGDOM

III

DISCRETE REVOLUTION

IV

TIME MEETS FREQUENCY

V

FRAMES

VI
WAVELET ZOOM

VII
WAVELET BASES