Statistical Modeling by Wavelets

BRANI VIDAKOVIC
Duke University

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.
New York / Chichester / Weinheim / Brisbane / Singapore / Toronto

This Page Intentionally Left Blank

Statistical Modeling by Wavelets

A complete list of the titles in this series appears at the end of this volume.

Statistical Modeling by Wavelets

BRANI VIDAKOVIC
Duke University

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.
New York / Chichester / Weinheim / Brisbane / Singapore / Toronto

This book is printed on acid-free paper. ♾

Published simultaneously in Canada.

For ordering and customer service, call 1-800-CALL-WILEY.

Library of Congress Cataloging-in-Publication Data:

Vidakovic, Brani, 1955–
Statistical modeling by wavelets / Brani Vidakovic.
p. cm.
Includes bibliographical references and index.
ISBN 0-471-29365-2 (alk. paper)
1. Wavelets (Mathematics). 2. Mathematical statistics. I. Title.
QA403.3.V53 1999
515'.2433—dc21 98-48449

10 9 8 7 6 5 4 3

Contents

Preface xi

Acknowledgments xiii

1 *Introduction* 1
- 1.1 *Wavelet Evolution* 1
- 1.2 *Wavelet Revolution* 6
- 1.3 *Wavelets and Statistics* 16
- 1.4 *An Appetizer: California Earthquakes* 19

2 *Prerequisites* 23
- 2.1 *General* 23
- 2.2 *Hilbert Spaces* 24
  - 2.2.1 *Projection Theorem* 25
  - 2.2.2 *Orthonormal Sets* 26
  - 2.2.3 *Reproducing Kernel Hilbert Spaces* 29
- 2.3 *Fourier Transformation* 29
  - 2.3.1 *Basic Properties* 30
  - 2.3.2 *Poisson Summation Formula and Sampling Theorem* 32

2.3.3 Fourier Series 33
2.3.4 Discrete Fourier Transform 34
2.4 Heisenberg's Uncertainty Principle 35
2.5 Some Important Function Spaces 36
2.6 Fundamentals of Signal Processing 39
2.7 Exercises 40

3 Wavelets 43
3.1 Continuous Wavelet Transformation 43
3.1.1 Basic Properties 44
3.1.2 Wavelets for Continuous Transformations 48
3.2 Discretization of the Continuous Wavelet Transform 50
3.3 Multiresolution Analysis 51
3.3.1 Derivation of a Wavelet Function 57
3.4 Some Important Wavelet Bases 60
3.4.1 Haar's Wavelets 60
3.4.2 Shannon's Wavelets 63
3.4.3 Meyer's Wavelets 65
3.4.4 Franklin's Wavelets 70
3.4.5 Daubechies' Compactly Supported Wavelets 73
3.5 Some Extensions 80
3.5.1 Regularity of Wavelets 80
3.5.2 The Least Asymmetric Daubechies' Wavelets: Symmlets 86
3.5.3 Approximations and Characterizations of Functional Spaces 87
3.5.4 Daubechies-Lagarias Algorithm 89
3.5.5 Moment Conditions 91
3.5.6 Interpolating (Cardinal) Wavelets 93
3.5.7 Pollen-Type Parameterization of Wavelets 94
3.6 Exercises 96

4 Discrete Wavelet Transformations 101
4.1 Introduction 101
4.2 The Cascade Algorithm 104
4.3 The Operator Notation of DWT 109
4.3.1 Discrete Wavelet Transformations as Linear Transformations 115
4.4 Exercises 117

5 Some Generalizations 119
5.1 Coiflets 120
5.1.1 Construction of Coiflets 121
5.2 Biorthogonal Wavelets 122
5.2.1 Construction of Biorthogonal Wavelets 125
5.2.2 B-Spline Wavelets 128
5.3 Wavelet Packets 133
5.3.1 Basic Properties of Wavelet Packets 135
5.3.2 Wavelet Packet Tables 138
5.4 Best Basis Selection 140
5.4.1 Some Cost Measures and the Best Basis Algorithm 141
5.5 ϵ-Decimated and Stationary Wavelet Transformations 145
5.5.1 ϵ-Decimated Wavelet Transformation 145
5.5.2 Stationary (Non-Decimated) Wavelet Transformation 147
5.6 Periodic Wavelet Transformations 150
5.7 Multivariate Wavelet Transformations 153
5.8 Discussion 160
5.9 Exercises 163

6 Wavelet Shrinkage 167
6.1 Shrinkage Method 168
6.2 Linear Wavelet Regression Estimators 170
6.2.1 Wavelet Kernels 171
6.2.2 Local Constant Fit Estimators 172
6.3 The Simplest Non-Linear Wavelet Shrinkage: Thresholding 175
6.3.1 Variable Selection and Thresholding 177
6.3.2 Oracular Risk for Thresholding Rules 177
6.3.3 Why the Wavelet Shrinkage Works 179
6.3.4 Almost Sure Convergence of Wavelet Shrinkage Estimators 184
6.4 General Minimax Paradigm 184
6.4.1 Translation of Minimaxity Results to the Wavelet Domain 186
6.5 Thresholding Policies and Thresholding Rules 188
6.5.1 Exact Risk Analysis of Thresholding Rules 188
6.5.2 Large Sample Properties of $\hat{f}$ 189

6.5.3 Some Other Shrinkage Rules 191
6.6 How to Select a Threshold 194
6.6.1 Mallat's Model and Induced Percentile Thresholding 194
6.6.2 Universal Threshold 195
6.6.3 A Threshold Based on Stein's Unbiased Estimator of Risk 199
6.6.4 Cross-Validation 200
6.6.5 Thresholding as a Testing Problem 202
6.6.6 Lorentz Curve Thresholding 203
6.6.7 Block Thresholding Estimators 206
6.7 Other Methods and References 207
6.8 Exercises 213

7 Density Estimation 217
7.1 Orthogonal Series Density Estimators 217
7.2 Wavelet Density Estimation 219
7.2.1 δ-Sequence Density Estimators 219
7.2.2 Bias and Variance of Linear Wavelet Density Estimators 223
7.2.3 Linear Wavelet Density Estimators in a More General Setting 224
7.3 Non-Linear Wavelet Density Estimators 225
7.3.1 Global Thresholding Estimator 227
7.4 Non-Negative Density Estimators 228
7.4.1 Estimating the Square Root of a Density 229
7.4.2 Density Estimation by Non-Negative Wavelets 234
7.5 Other Methods 236
7.5.1 Multivariate Wavelet Density Estimators 236
7.5.2 Density Estimation as a Regression Problem 239
7.5.3 Cross-Validation Estimator 240
7.5.4 Multiscale Estimator 240
7.5.5 Estimation of a Derivative of a Density 241
7.6 Exercises 243

8 Bayesian Methods in Wavelets 247
8.1 Motivational Examples 247
8.2 Smooth Shrinkage 250
8.3 Bayesian Thresholding 255

8.4 MAP-Principle 257
8.5 Density Estimation Problem 259
8.6 Full Bayesian Model 262
8.7 Discussion and References 265
8.8 Exercises 267

9 Wavelets and Random Processes 271
9.1 Stationary Time Series 272
9.2 Wavelets and Stationary Processes 273
9.2.1 Wavelet Transformations of Stationary Processes 273
9.2.2 Whitening of Stationary Processes 274
9.2.3 Karhunen-Loève-Like Expansions 276
9.3 Estimation of Spectral Densities 279
9.3.1 Gao's Algorithm 280
9.3.2 Non-Gaussian Stationary Processes 284
9.4 Wavelet Spectrum 286
9.4.1 Wavelet Spectrum of a Stationary Time Series 286
9.4.2 Scalogram and Periodicities 289
9.5 Long-Memory Processes 290
9.5.1 Wavelets and Fractional Brownian Motion 290
9.5.2 Estimating Spectral Exponents in Self-Similar Processes 292
9.5.3 Quantifying the Whitening Property of Wavelet Transformations for fBm Processes 295
9.6 Discussion and References 296
9.7 Exercises 297

10 Wavelet-Based Random Variables and Densities 299
10.1 Scaling Function as a Density 299
10.2 Wavelet-Based Random Variables 300
10.3 Random Densities via Wavelets 308
10.3.1 Tree Algorithm 308
10.4 Properties of Wavelet-Based Random Densities 310
10.5 Random Densities With Constraints 312
10.5.1 Smoothness Constraints 312
10.5.2 Constraints on Symmetry 312
10.5.3 Constraints on Modality 313
10.5.4 Skewed Random Densities 313

10.6 Exercises 316

11 Miscellaneous Statistical Applications 319
11.1 Deconvolution Problems 319
11.2 Wavelet-Vaguelette Decompositions 323
11.3 Pursuit Methods 325
11.4 Moments of Order Statistics 329
11.5 Wavelets and Statistical Turbulence 333
11.5.1 K41 Theory 335
11.5.2 Townsend's Decompositions 336
11.6 Software and WWW Resources for Wavelet Analysis 338
11.6.1 Commercial Wavelet Software 338
11.6.2 Free Wavelet Software 340
11.6.3 Some WWW Resources 341
11.7 Exercises 342

References 345

Notation Index 371

Author Index 373

Subject Index 379

Preface

Just two months ago astronomers did not know about it. But now they are giving good odds that Hyakutake will be the most impressive comet since the invention of telescope 400 years ago. *(Herald Sun, Durham, NC, March 24, 1996.)*

One can trace the origins of wavelets back to the beginning of this century; however, wavelets, understood as a systematic way of producing local orthogonal bases, are a recent unification of existing theories in various fields and some important "discoveries." They are mathematical objects that have interpretation and application in many scientific fields, most notably in the fields of signal processing, nonparametric function estimation, and data compression. In the early 1990s, a series of papers by Donoho and Johnstone and their coauthors demonstrated that wavelets are appropriate tools in problems of denoising, regression, and density estimation. The subsequent burgeoning wavelet research broadened to a wide range of statistical problems.

This book is aimed at graduate students in statistics and mathematics, practicing statisticians, and statistically curious engineers. It can serve as a text for an introductory wavelet course concerned with an interface of wavelet methods and statistical inference. The necessary mathematical background is proficiency in advanced calculus and algebra; consequently, this book should be useful to advanced undergraduate students as well as to graduate students in statistics, mathematics, and engineering.

This book originated from the class notes supporting the Special Topics Course on Multiscale Methods at Duke University. The content can be divided into two parts:

an introduction to wavelets (Chapters 1–5) and statistical modeling (Chapters 6–11). An introduction and some mathematical prerequisites are presented in Chapters 1 and 2. Continuous and discrete wavelet transformations are covered in Chapters 3 and 4. Some important generalizations (coiflets, biorthogonal wavelets, wavelet packets, stationary, periodized and multivariate wavelets) are covered in Chapter 5.

Chapters 6–11 are data-oriented. Chapter 6 is the crux of the book, covering the theory and practice of wavelet shrinkage. Important theoretical aspects of wavelet density estimation are covered in Chapter 7. Chapter 8 discusses Bayesian modeling in the wavelet domain. Time series are covered in Chapter 9, while Chapter 10 contains several probabilistic and simulational properties of wavelet-based random functions and densities. Chapter 11 gives some novel and important wavelet applications in statistics.

Instead of providing appendices with data sets and programs used in the book, I opted for a more modern style. The web page:

`http://www.isds.duke.edu/~brani/wiley.html`

is associated with the book. This page contains all data sets, functions, and programs referred to.

I hope the reader will find this book useful. All comments, suggestions, updates, and critiques will be appreciated.

BRANI VIDAKOVIC

Institute of Statistics and Decision Sciences
Duke University
Durham, February 1999

Acknowledgments

This book was made possible by the support of many individuals. First, I am grateful to colleagues from the Institute of Statistics and Decision Sciences (ISDS) – Duke University for supporting a wavelet-based statistics course. The attending students were patient and understood the difficulties of transferring research papers to a working copy of course notes. Their industry and enthusiasm is treasured. Duke University helped with this project through the Grant of Arts and Sciences Council 1997, and partial support was provided by the National Science Foundation Award DMS-9626159 at Duke University. Figure 1.10 is reproduced with the permission of the Salvador Dalí Museum, Inc. in St. Petersburg, Florida.

Many colleagues contributed to this project in different ways: Anestis Antoniadis, Tony Cai, Merlise Clyde, Lubo Dechevsky, Iain Johnstone, Gabriel Katul, Eric Kolaczyk, Pedro Morettin, Peter Müller, Giovanni Parmigiani, Marianna Pensky, David Rios, Fabrizio Ruggeri, Rainer von-Sacks, Naoki Saito, Yazhen Wang, and Gilbert Walter, to list a few. Collaboration with software gurus Hong-Ye Gao [TeraLogic Inc.] and Andrew Bruce [MathSoft Inc.] was fruitful. The S+Wavelets module (for S-Plus) was used for almost all of the computer examples, figures, and calculations. I am grateful to Alison Bory, Angioline Loredo, and Steve Quigley from Wiley, for their enthusiastic assistance, and to Courtney Johnson, Michael Kozdron, and Kathy Zhou, doctoral students at Duke University, for their help in proofreading the manuscript.

And most of all, I am grateful to my family for their love and strong and continuous support.

1

Introduction

In this chapter, we give a brief overview of the history of wavelets, make a case for their use in statistics, and provide a real-life example that emphasizes specificities of wavelets in data processing problems. The wavelet method in this example is compared with its counterpart traditional approaches. The reader may encounter unfamiliar jargon or undefined objects. Some of these notions will be defined later and some are used to illustrate the general picture.

1.1 WAVELET EVOLUTION

Wavelets are developed not only from a couple of bright discoveries, but from concepts and theories that already existed in various fields. In this section, we will give a brief historic tour of some important milestones in the development of wavelets.

Functional series have a long history that can be traced back to the early nineteenth century. French mathematician (and politician!) Jean-Baptiste-Joseph Fourier [Fig. 1.1(a)] in 1807 [1] decomposed a continuous, periodic on $[-\pi, \pi]$ function $f(x)$ into

[1] Jean-Baptiste-Joseph Fourier's *Theorie analitique de la chaleur* (The Mathematical Theory of Heat) inaugurated simple methods for the solution of boundary value problems occurring in the conduction of heat.

the series

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx + b_n \sin nx,$$

where the coefficients a_n and b_n are defined as

$$\begin{aligned} a_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\, dx, \quad n = 0, 1, 2, \ldots \\ b_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\, dx, \quad n = 1, 2, \ldots . \end{aligned}$$

It is interesting that, at the time of Fourier's discovery, the notion of a function was not yet precisely defined.

(a) (b)

Fig. 1.1 (a) Jean-Baptiste-Joseph Fourier 1768-1830 and (b) Alfred Haar 1885-1933.

The first "wavelet basis" was discovered in 1910 when Alfred Haar [Fig. 1.1(b)] showed that any continuous function $f(x)$ on $[0, 1]$ can be approximated by

$$f_n(x) = \langle \xi_0, f \rangle \xi_0(x) + \langle \xi_1, f \rangle \xi_1(x) + \cdots + \langle \xi_n, f \rangle \xi_n(x), \tag{1.1}$$

and that, when $n \to \infty$, f_n converges to f uniformly ([181]). The coefficients $\langle \xi_i, f \rangle$ are given by $\int \xi_i(x) f(x) dx$. The Haar basis is very simple:

$$\begin{aligned}
\xi_0(x) &= \mathbf{1}(0 \le x \le 1), \\
\xi_1(x) &= \mathbf{1}(0 \le x \le 1/2) - \mathbf{1}(1/2 \le x \le 1), \\
\xi_2(x) &= \sqrt{2}[\mathbf{1}(0 \le x \le 1/4) - \mathbf{1}(1/4 \le x \le 1/2)], \\
&\dots \\
\xi_n(x) &= 2^{j/2}[\mathbf{1}(k \cdot 2^{-j} \le x \le (k+1/2) \cdot 2^{-j}) \\
&\quad -\mathbf{1}((k+1/2) \cdot 2^{-j} \le x \le (k+1) \cdot 2^{-j})], \\
&\dots
\end{aligned}$$

where n is uniquely decomposed as $n = 2^j + k,\ j \ge 0,\ 0 \le k \le 2^j - 1$, and $\mathbf{1}(A)$ is the indicator of a set A, i.e., $\mathbf{1}(A) = 1$, if $x \in A$, and $\mathbf{1}(A) = 0$, if $x \in A^c$.

The approximation in (1.1) is equivalent to an approximation by step functions whose values are the averages (mean values) of the function over appropriate dyadic intervals.

Fig. 1.2 gives an exemplary function, $f(x) = \sin \pi x + \cos 2\pi x + 0.6 \cdot \mathbf{1}(x > 1/2)$, and three different levels of approximation: f_3, f_{15}, and f_{63}. Basis functions ξ_1, ξ_2, ξ_{14}, and ξ_{25} are shown in Fig. 1.3. Since $\int \xi_n^2(x)\, dx = 1$ for an arbitrary n, there is a trade-off between the magnitude and the support of the basis functions in the Haar system.

Notice that for any $n \ge 1$ the basis function ξ_n can be expressed as a scale-shift transformation of a single function ξ_1,

$$\xi_n(x) = 2^{j/2}\xi_1(2^j \cdot x - k), \quad n = 2^j + k,$$

a property shared by critically sampled wavelets, as we will see later. The function $\xi_0(x)$ is different in nature than the functions ξ_n, $n \ge 1$; while the functions ξ_n, $n \ge 1$ describe the details in the decomposition, the function $\xi_0(x)$ is responsible for the "average" of the decomposed function.

The Schauder basis on $[0, 1]$ (Schauder [369]) consists of the primitives of the Haar basis functions, the triangle functions. Let $\Delta(x) = 2x\, \mathbf{1}(0 \le x \le 1/2) + 2(1 - x)\, \mathbf{1}(1/2 \le x \le 1)$, and let $\Delta_n(x) = \Delta(2^j x - k),\ n = 2^j + k,\ j \ge 0,\ 0 \le k \le 2^j - 1$. Then $\{1, \Delta(x), \Delta_1(x), \dots\}$ constitutes a Schauder basis on $[0, 1]$ and, as in the case of Haar's basis, any continuous function $f(x)$ on $[0, 1]$ can be approximated by

$$f_N(x) = a + bx + \sum_{n=1}^{N} s_n \Delta_n(x). \tag{1.2}$$

Coefficients a and b are solutions of the system $f(0) = a$ and $f(1) = a + b$, while the coefficients s_n can be obtained by the simple relation

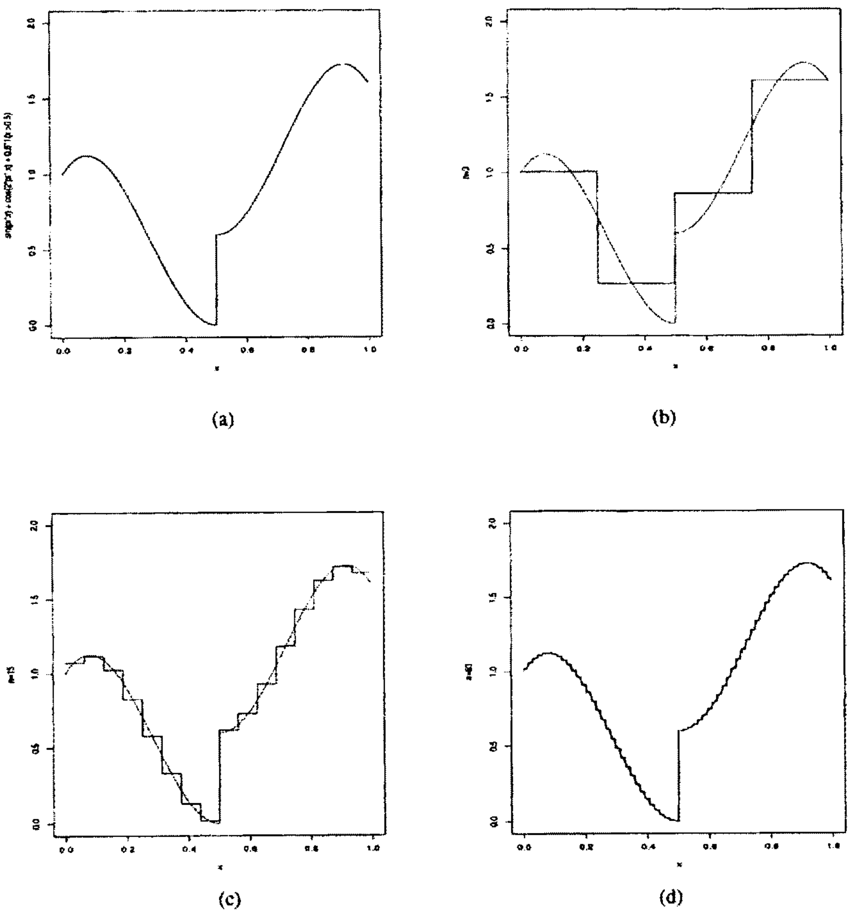


Fig. 1.2 Panels (a)-(d) show the original function $f(x) = \sin \pi x + \cos 2\pi x + 0.6 \cdot \mathbf{1}(x > \frac{1}{2})$, $0 \leq x \leq 1$, and three different levels of approximation in the Haar basis. Using the notation of (1.1), approximations f_3, f_{15}, and f_{63} are plotted.

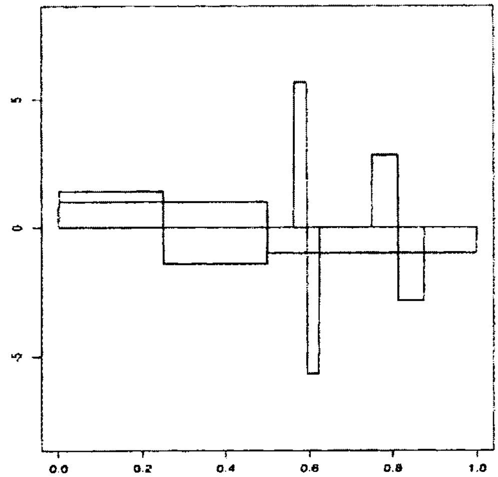


Fig. 1.3 Functions ξ_1, ξ_2, ξ_{14}, and ξ_{25} from the Haar basis of $\mathbb{L}_2([0,1])$.

$$s_n = f\left(\frac{k+1/2}{2^j}\right) - \frac{1}{2}\left[f\left(\frac{k}{2^j}\right) + f\left(\frac{k+1}{2^j}\right)\right],$$
$$n = 2^j + k,\ j \geq 0,\ 0 \leq k \leq 2^j - 1.$$

The convergence $f_N(x) \to f(x)$ is uniform and the coefficients are unique; however, the Schauder system is not orthogonal. We will see later that its orthogonalization leads to a family of wavelets, known as Franklin wavelets.

In the mid-1930s, Littlewood-Paley techniques (based on Fourier methods) [264] were broadly used in research on Fourier summability and in investigation of the behavior of analytic functions.

Prototypes of wavelets first appeared in Lusin's work in the 1930s. A standard characterization of Hardy's spaces can be given in terms of Lusin's "area" functions.

In the 1950s and 1960s, techniques by Littlewood-Paley and Lusin were developed into powerful tools for studying physical phenomena describable by solutions of differential and integral equations. Researchers realized that these techniques could be unified by the Calderón-Zygmund theory [292], now a branch of harmonic analysis.

Strömberg [390] was the first to construct an orthonormal basis of $\mathbb{L}_2(\mathbb{R})$ of the form $\{\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k),\ j,k \in \mathbb{Z}\}$, a wavelet-like basis more general than Haar's basis. Stromberg's construction uses Franklin systems which are Gram-Schmidt orthogonalized Schauder basis functions $\Delta_n(x)$.

For more information about the historical roots of wavelets, we direct the reader to monographs by Meyer [294, 295] and Daubechies [104].

1.2 WAVELET REVOLUTION

The first definitions of wavelets can be attributed to Morlet et al. [300] and Morlet and Grossmann [179] and it is given in the Fourier domain: A wavelet is an $\mathbb{L}_2(\mathbb{R})$ function for which the Fourier transformation $\Psi(\omega)$ satisfies

$$\int_0^\infty |\Psi(t\omega)|^2 \frac{dt}{t} = 1, \text{ for almost all } \omega.$$

The definition of Morlet and Grossmann is quite broad and over time the meaning of the term *wavelet* became narrower. Currently, the term wavelet is usually associated with a function $\psi \in \mathbb{L}_2(\mathbb{R})$ such that the translations and dyadic dilations of ψ,

$$\psi_{jk}(x) = 2^{j/2}\, \psi(2^j x - k),\ j, k \in \mathbb{Z} \tag{1.3}$$

constitute an orthonormal basis of $\mathbb{L}_2(\mathbb{R})$.

Calculating wavelet expansions directly is a computationally expensive task, moreover, most interesting wavelets are without a closed form. In the mid-1980s, Mallat [274, 275, 276] connected quadrature-mirror filtering and pyramidal algorithms from the signal processing theory with wavelets. He demonstrated that discrete wavelet transformation can be calculated very rapidly via cascade-like algorithm. This link was of paramount importance for the practice of wavelets. Daubechies' discovery of compactly supported wavelet bases represents another important milestone in the development of wavelet theory. Daubechies' bases are versatile in smoothness and locality and represent a starting point for much of the subsequent generalizations and theoretical advances.

Wavelet theory has developed now into a methodology used in many disciplines: mathematics, geophysics, astronomy, signal processing, numerical analysis, and statistics, to list a few. Wavelets are providing a rich source of useful and sometimes intriguing tools for applications in "time-scale" types of problems. In analyses of signals, the wavelet representations allow us to view a time-domain evolution in terms of scale components. In this respect, wavelet transformations behave similarly to Fourier transformations. The Fourier transform extracts details from the signal frequency, but all information about the location of a particular frequency within the signal is lost. Time localization can be achieved by first windowing the signal, and then by taking its Fourier transform. The problem with windowing is that slices of the processed signal are of a fixed length, which is determined by the window. Slices of the same length are used to resolve both high and low frequency components. For nonstationary signals, this lack of adaptivity may lead to a local under- or over-fitting.

Table 1.1 Coefficients of the `doppler` function in the Haar basis plotted in levels determined by the length of support of corresponding basis functions, ξ_n, $n \geq 0$.

level in Fig. 1.4	coefficients of	support	j and k in the notation: j	$n = 2^j + k$ k
	...	...	...	...
d1	ξ_{512} - ξ_{1023}	$1/2^9$	$j = 9$	$0 \leq k \leq 2^9 - 1$
d2	ξ_{256} - ξ_{511}	$1/2^8$	$j = 8$	$0 \leq k \leq 2^8 - 1$
	...	...	...	...
d8	ξ_4 - ξ_7	1/4	$j = 2$	$0 \leq k \leq 3$
d9	ξ_2 - ξ_3	1/2	$j = 1$	$0 \leq k \leq 1$
d10	ξ_1	1	$j = 0$	$k = 0$
s10	ξ_0	1		
	...	...	...	...

In contrast to windowed Fourier transforms, wavelets select widths of time slices according to the local frequency in the signal. This adaptivity property of wavelets is very important, and we will make it more precise later in the discussion of Heisenberg's uncertainty principle. Two panels in Fig. 1.11, on page 18, depict slicing the time–scale plane for a windowed Fourier (left) and a wavelet transformation (right).

Now we give several examples: The first example views the Haar decomposition as a wavelet decomposition and discusses connections between "levels" and resolutions of the decomposition. The subsequent four examples demonstrate important properties of wavelets: the ability to filter, "disbalance", and "whiten" signals as well as to detect self-similarity within a signal.

Example 1.2.1 The Haar basis as a wavelet basis. To illustrate the time and scale adaptivity of wavelets, and to introduce some necessary wavelet notations and jargon, let us consider a decomposition of the function

$$y(x) = \sqrt{x(1-x)}\ \sin\frac{2.1\pi}{x + 0.05},\ 0 \leq x \leq 1, \tag{1.4}$$

in Haar's basis. This function is known as the `doppler` test-function. Notice that frequency in the function increases as x decreases.

In Table 1.1, n is represented as $2^j + k$ where j is a *level* and k is a *shift* within the level. Notice that all functions ξ within a level have supports of the same length. The support of a function is defined as closure of the set at which the function differs from zero.

When $j \to \infty$, the number of coefficients in the level increases and the length of support of the corresponding basis functions decreases. For example, the level indexed by $j = 5$ has $2^5 = 32$ coefficients and the supports are of length 2^{-5}. The shifts within a level are indexed by k, where k ranges from 0 to $2^j - 1$. For

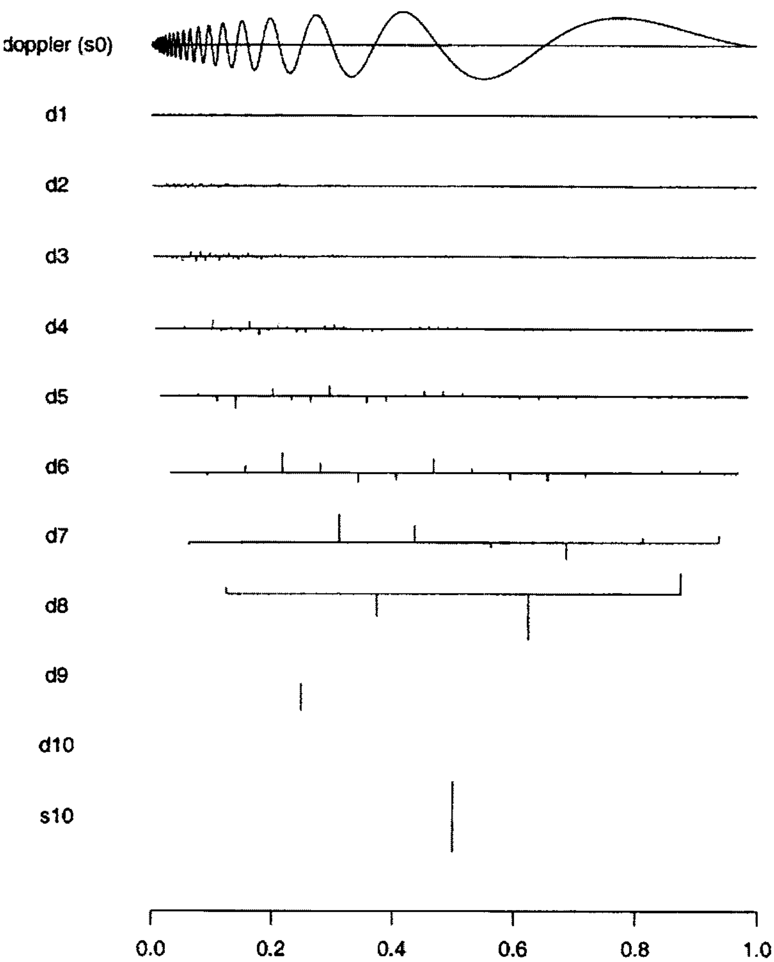


Fig. 1.4 The `doppler` function and its Haar basis decomposition.

an exact description of an arbitrary function, the number of levels is infinite. The coefficient corresponding to ξ_0 is called the "smooth" coefficient ($s0$ in Fig. 1.4) and the coefficients corresponding to ξ_n, $n > 1$ are called "detail" coefficients. Level $j = 9$ ($d1$ in Fig. 1.4) contains coefficients corresponding to "fine" details.

When dealing with functions that are given by their sampled values, it is customary to set the sampled values to be "smooth" coefficients at the level $j = J$. The subsequent "detail" levels denoted by $d1, d2, \ldots$, correspond to $j = J-1, J-2, \ldots$.

We provide four more examples that emphasize the most interesting features of wavelet transformations. Occasionally, we will use terms like "fine and coarse levels", "wavelet domain", and "energy", which have not been previously defined and will be defined in the subsequent chapters. However, the intended messages of the examples should be clear even without precise definitions of such terms.

Example 1.2.2 Wavelets generate local bases. Classical orthonormal bases (Fourier, Hermite, Legendre, etc.) have been used with great success in applied mathematics for decades. However, there is a serious limitation shared by many classical bases, which is *non-locality*. A basis is non-local when many basis functions are substantially contributing at any value of a decomposition. The convergence of non-local classical decompositions often relies on a multitude of cancellations.

Local bases are desirable since they are more adaptive and parsimonious. In 1946, Gabor [161] suggested localizing Fourier bases by modulating and translating an appropriate "window" function g. More precisely, Gabor suggested bases in the form

$$\{g_{m,n}(x) = e^{2\pi m i x} g(x-n)\},$$

where m and n are integers and g is a square-integrable function. An example of a function g that produces an orthonormal basis of $\mathbb{L}_2(\mathbb{R})$ is $\sin(\pi x)/(\pi x)$.

The Balian-Low theorem stipulates limitations of Gabor bases. If the Gabor basis is orthogonal and $\hat{g}(\omega)$ is the Fourier transformation of the window $g(x)$ then, by the Balian-Low theorem, either $\int x^2 |g(x)|^2\, dx = \infty$ or $\int \omega^2 |\hat{g}(\omega)|^2\, d\omega = \infty$. In other words, orthogonal Gabor bases are non-local either in time or in scale (frequency). Modulations and translations of the Gaussian window $g(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$ (which is well localized in both time and frequency, and for which the above integrals are finite) will not produce an orthonormal basis.

Locality of wavelet bases comes from their construction. Most of the wavelets that are used in statistics now are either compactly supported or decay exponentially. An exception are Meyer-type wavelets (with a polynomial decay) used in deconvolution problems.

Example 1.2.3 Wavelets filter data. To illustrate the action of wavelets as a filtering device, we generate two periodic functions with different frequencies, $y_1 = \sin x + \cos 2x$, and $y_2 = \frac{1}{5} \arcsin(\sin 20x)$, where $x \in [-2\pi, 2\pi]$. These are shown in panel (a) in Fig. 1.5. Our goal is to filter out the component y_2 from the given sum $y_1 + y_2$

[Fig. 1.5(b)]. Since the periods of y_1 and y_2 are different, the functions are described by wavelets with different supports (and whose coefficients belong to different levels). Fig. 1.5(c), depicts the level-wise energies (sums of squares of wavelet coefficients). The support of wavelets associated with level 1 is 32 times larger than the support of wavelets associated with level 5. This means that almost all the energy in levels 0,1, and 2 comes from signal y_1, and the energy in level 5 comes from y_2, thus allowing an easy separation. The filtered components are depicted in Fig. 1.5(d).

Example 1.2.4 Wavelets "disbalance" energy in data. The term "disbalance" is coined and it relates to an uneven distribution of energy in a signal. Disbalancing is desirable since a signal can be well described by only a few energetic components.

To illustrate the disbalancing action that is typical of wavelets, we first introduce some necessary notation. Given a vector $\underset{\sim}{a} = (a_1, a_2, \dots, a_n)$ let $||\underset{\sim}{a}||^2 = \sum_i a_i^2$ be the total energy of $\underset{\sim}{a}$ and let a_i^2 be the ith energy component. Let $a_{(1)}^2, a_{(2)}^2, \dots, a_{(n)}^2$ be increasingly ordered energy components. The standard measure of disbalance used in economics is the Lorentz curve. The Lorentz curve was introduced at the beginning of the century. It was used by economics researchers to assess inequality of distribution of wealth in a country, region, or among people within a particular population group.

One definition of the Lorentz curve, in terms of energy components, is

$$L(p) = \frac{1}{||\underset{\sim}{a}||^2} \cdot \sum_{i=1}^{\lfloor np \rfloor} a_{(i)}^2, \; p \in [0, 1],$$

where $\lfloor x \rfloor$ is the largest integer smaller than x. In Fig. 1.6(a), an observed time series (turbulence data set) is given. Below is its wavelet transformation represented in a vector form beginning with coarse coefficients. Orthogonality of the transformation preserves the total energy, $||\underset{\sim}{a}||^2$. However, the energy in the wavelet domain is more disbalanced, as indicated by the Lorentz curves in Fig. 1.6(b). Notice that 90% of energy is contained in about 6-7% of the components in the wavelet-transformed data set compared to nearly 50% of the components in the original (time) domain.

Example 1.2.5 Wavelets whiten data. In this example, we show another interesting property of wavelets. Orthogonal wavelet transformations map white noise to white noise, which is a consequence of orthogonality. However, signals that are correlated in the time domain become almost uncorrelated in the wavelet domain. Informally, the wavelet transformation acts as an approximation to the Karhunen-Loève transformation. To exemplify this statement, a time series of 256 components was generated from a random process with stationary increments, ARIMA(1,1,1) process. Such processes exhibit long-range dependence and their autocovariance functions [Fig. 1.7(a)] show slow decay. The autocovariance function of the wavelet-transformed time series exhibits very different behavior. Only the covariances at the first few lags are significant at a 5% significance level.

Related discussion can be found in Johnstone and Silverman [222], Mallat [277],

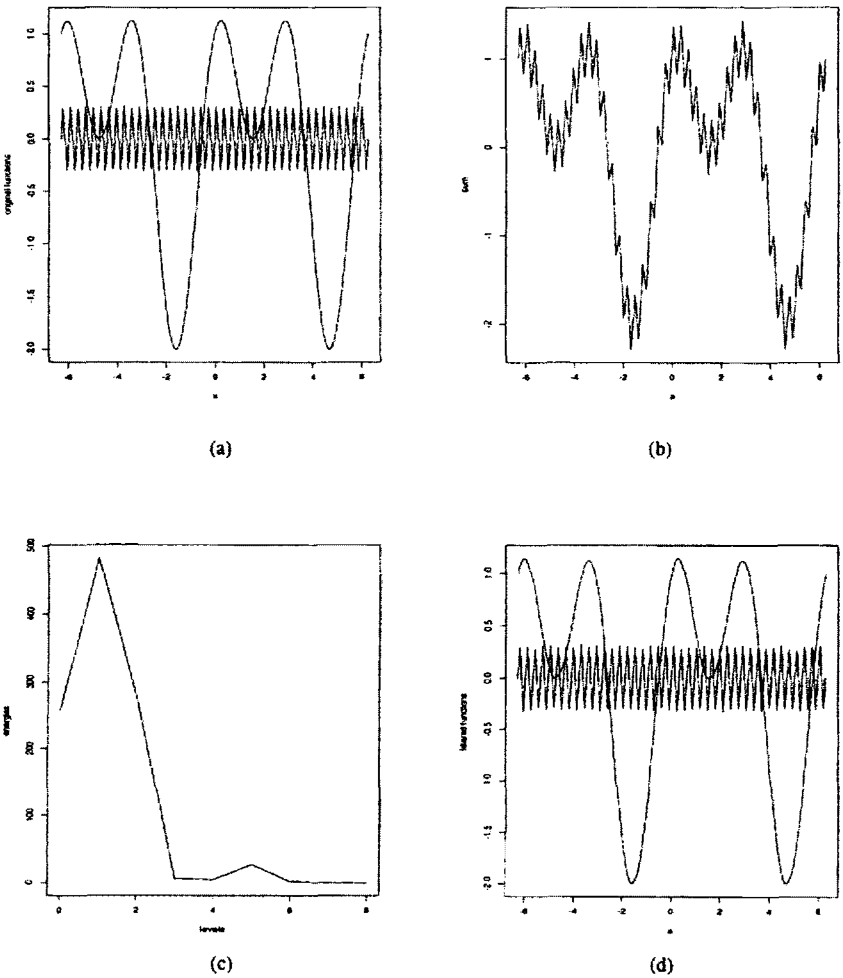


Fig. 1.5 Filtering property of wavelets. Two functions $y_1 = \sin x + \cos 2x$ and $y_2 = \frac{1}{5}\arcsin(\sin 20x)$, and their sum $y_1 + y_2$ are plotted in panels (a) and (b). Panel (c) shows the separation of "energy" to different levels in wavelet decomposition, while panel (d) shows filtered functions.

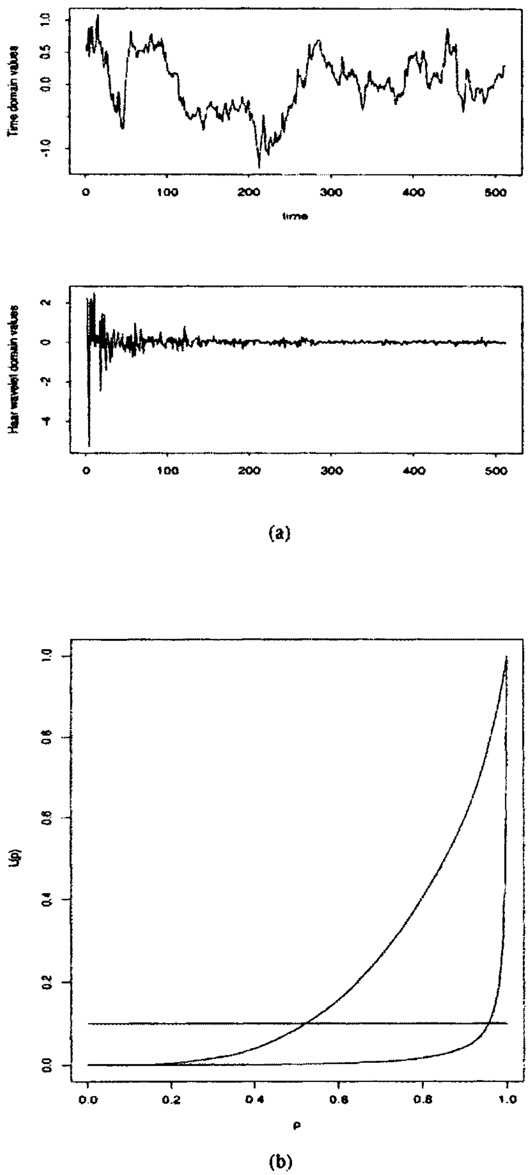


Fig. 1.6 (a) Atmospheric turbulence measurements of u velocity component (upper panel) and their wavelet transformation (lower panel). (b) Lorentz curves of the original and transformed measurements. The curve corresponding to transformed measurements has higher curvature.

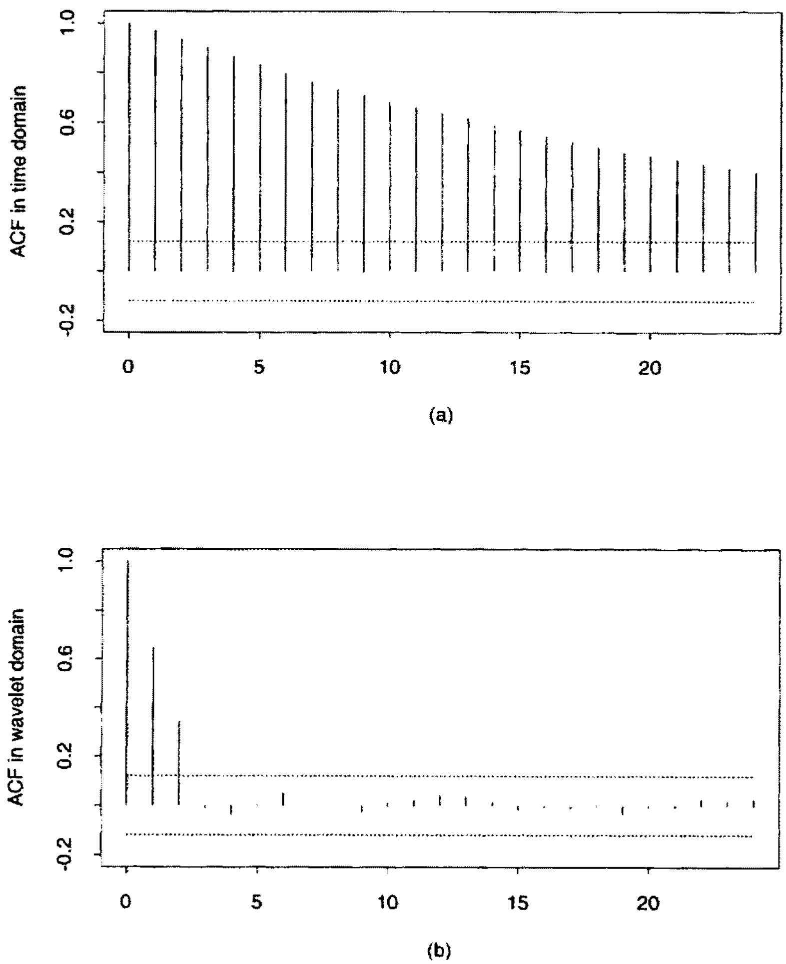


Fig. 1.7 Illustration of the whitening effect of wavelet transformations. Autocovariance function for a time series [ARIMA(1,1,1)] in the time domain [panel (a)] and the wavelet domain [panel (b)].

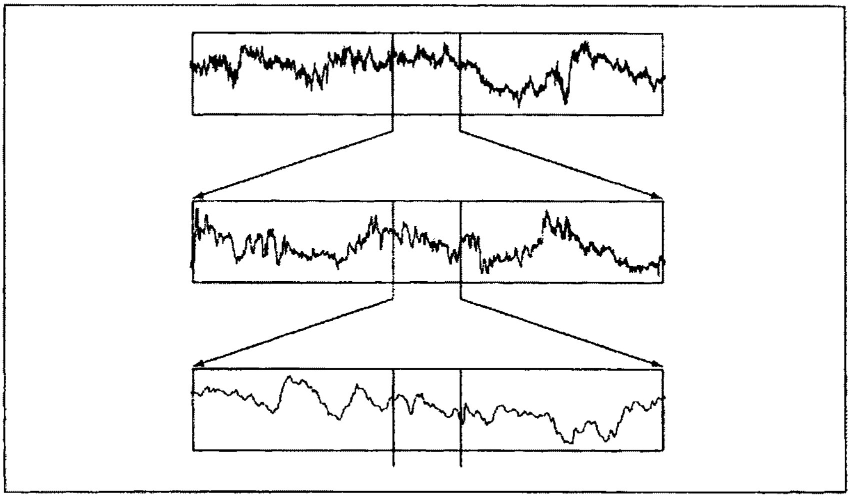


Fig. 1.8 Self-similarity of a turbulence time series.

Walter [439], and Wornell [461].

Example 1.2.6 Wavelets detect self-similar phenomena. Being self-similar themselves, wavelets are especially apt to describe phenomena exhibiting self-similarity in different scales (Fig. 1.8). Early research on wavelets was generated to address related problems in geophysics, especially in turbulence. An overview can be found in Kumar and Foufoula-Georgiou [250]. A curious phenomenon is that atmospheric turbulence measurements of different physical quantities, such as air velocities, ozone and humidity concentrations, temperature, and so on, follow identical power laws (as predicted by Kolmogorov's [242] theory). Such laws describe the energy transport in the inertial range of turbulent flows. A nice reference is a book by Frisch [160].

One of the theoretical laws is the "$-\frac{5}{3}$" law. It states that the log-power spectrum in the inertial range decreases linearly, with the slope of $-\frac{5}{3}$. Fig. 1.9 shows the wavelet-spectrum of air velocity measurements and it's near-perfect compliance with the $-\frac{5}{3}$ law.

There are problems in which wavelets should be used with caution. For instance, in the wavelet domain, the dependence structure in the transformed time series is influenced by the choice of the decomposing wavelet. In some cases, the extent of such non-robustness hinders practical generalizations. When non-robustness is of particular concern, researchers usually fix a *good* wavelet for a class of problems, as is the case with the prevalent use of the Haar and Walsh bases in processing the turbulence data.

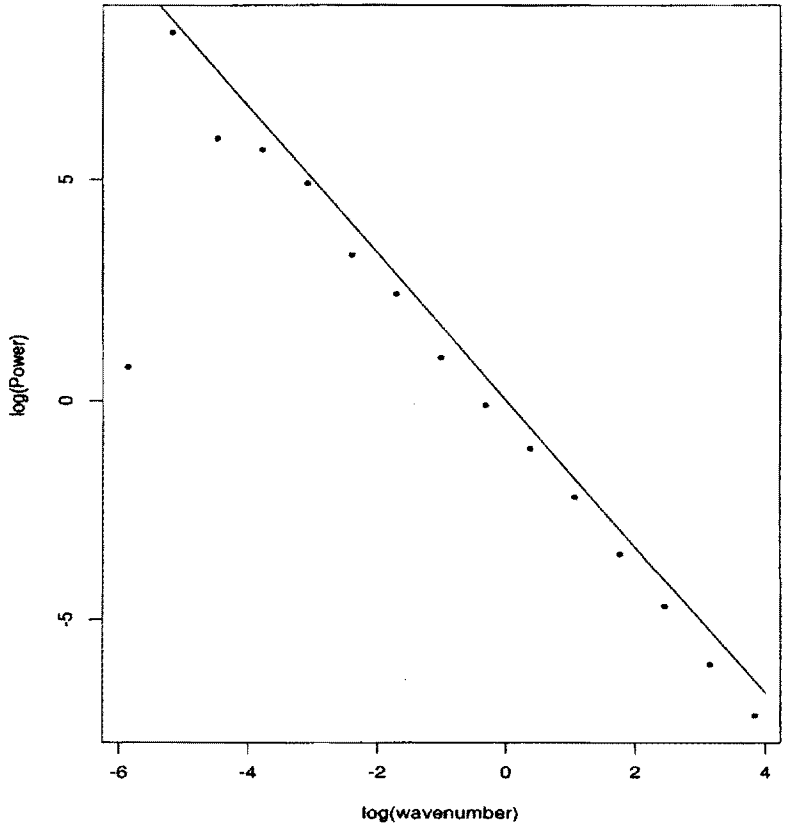


Fig. 1.9 Wavelet power-spectrum and Kolmogorov's $-\frac{5}{3}$ law. Dots represent the logarithms of the cumulative level-energies. The variable `log(wavenumber)` is linearly related to the level j.

Wavelets are not replacements for the standard Fourier methods, they are alternatives. If the signal is a linear combination of harmonics, clearly wavelets are suboptimal building blocks. For instance, if f is given as a lacunary (sparse) Fourier series like $\sum_{j=0}^{\infty} 2^{-j} \sin(2\pi 2^j x)$, then wavelets will be inferior in tasks of denoising and compression as compared to the Fourier transformation.

The whitening property, discussed in Example 1.2.5, impairs the performance of wavelet-based methods in prediction problems.

Another example in which wavelets should be cautiously used comes from image processing. In this example the interpretation of an object changes when the resolution changes. The two women in the background in Salvador Dalí's picture *Mercado de esclavos con aparicion del busto invisible de Voltaire*[2] (Fig. 1.10) at a coarser resolution level can be interpreted as a bust of Voltaire. Clearly, the meaning of the object changes in different scales.

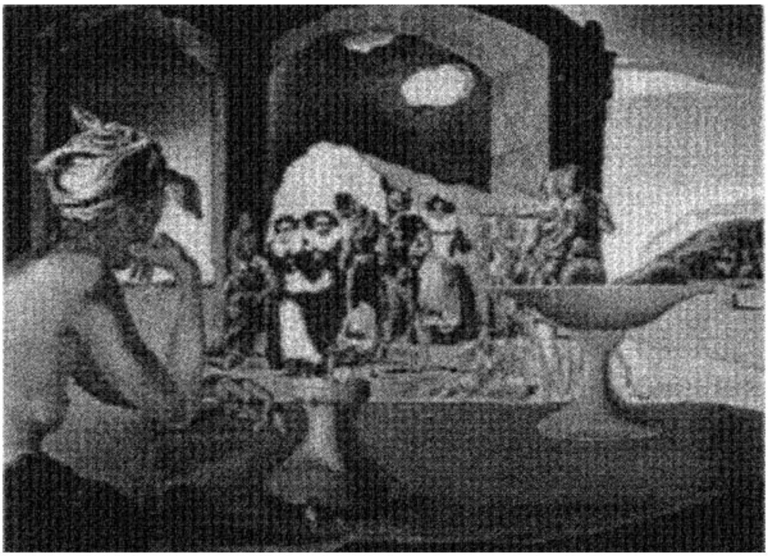


Fig. 1.10 *Mercado de esclavos con aparicion del busto invisible de Voltaire,* a 1940 painting by Salvador Dalí.

1.3 WAVELETS AND STATISTICS

Statistical multiscale modeling has, in recent years, become a burgeoning area in both theoretical and applied statistics, and is beginning to impact developments

[2] *Slave Market with the Disappearing Bust of Voltaire (1940), Oil on canvas.* $18\frac{1}{4} \times 25\frac{3}{8}$ *in.* Collection of The Salvador Dalí Museum, St. Petersburg, Florida. ©1998 Salvador Dalí Museum, Inc.

in statistical methodology as well as in various applied scientific fields. Wavelet-based methods are developing in statistics in areas such as regression, density and function estimation, factor analysis, modeling and forecasting in time series analysis, and spatial statistics. Emerging connections of Bayesian statistical modeling and wavelets are generating exciting new directions for the interface of the two research areas, with significant potential for future impact on applied work.

The attention of the statistical community was attracted when Mallat established a connection between wavelets and signal processing and Donoho and Johnstone showed that wavelet thresholding had desirable statistical optimality properties. Since then, wavelets have proved useful in many statistical disciplines, notably in nonparametric statistics and time series analysis. Bayesian concepts and modeling approaches have, more recently, been identified as providing promising contexts for wavelet-based denoising applications.

In addition to replacing traditional orthonormal bases in a variety statistical problems, wavelets brought novel techniques and invigorated some of the existing ones. Even in the cases in which the traditional orthogonal series are simply replaced by wavelet bases, wavelets often offer better localization and parsimony. For example, Čencov's [66] linear density estimator in the form of a Fourier series uses traditional orthonormal bases (Hermite, Fourier) to express its empirical Fourier coefficients. Wavelets achieve the same convergence rates and at the same time provide efficient non-linear approximations and adaptivity to unknown smoothness (via wavelet shrinkage). Wavelet shrinkage is achieved via explicit or implicit use of statistical models in the wavelet domain.

We elaborate further on the modeling in the wavelet domain and formalize some of the concepts already mentioned.

Low Entropy Modeling Environment. As we mentioned before, wavelet transformations tend to disbalance the data on input. Even though the transformations preserve the ℓ_2-norm of the data, the energy of the transformed data (an engineering term for the ℓ_2-norm) is concentrated in only a few wavelet coefficients. This concentration narrows the class of plausible statistical models and facilitates the thresholding. Different formalizations of this disbalancing property can yield a variety of criteria for the best basis selection. For more discussion, see Coifman and Wickerhauser [94], Donoho [123], and Mallat [277], among others.

Ockham's Razor Principle. Wavelets, as building blocks of models, are well localized in both time and scale (frequency). Signals with rapid local changes (signals with discontinuities, cusps, sharp spikes, etc.) can be precisely represented with just a few wavelet coefficients. Generally, this statement does not apply to other standard orthonormal bases that may require many "compensating" coefficients to describe discontinuity artifacts or to suppress Gibbs' effects. The latest "generation" of wavelets form over-complete dictionaries and provide parsimonious representations of real phenomena with complicated time and frequency behaviors.

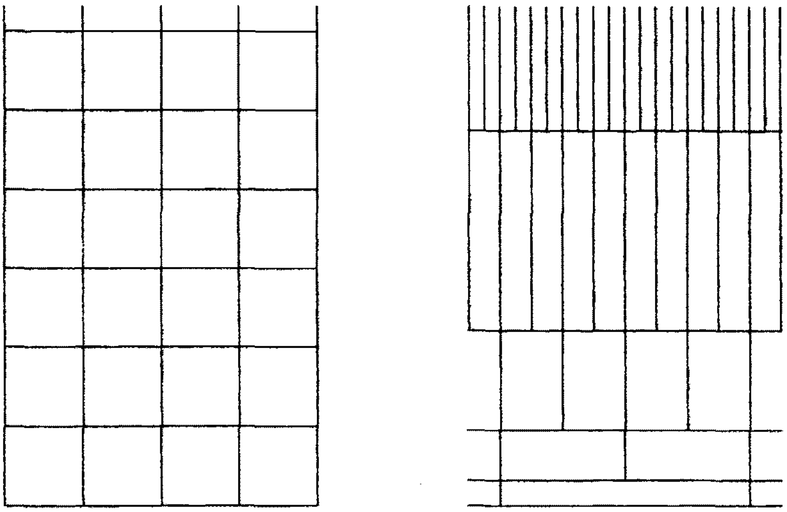


Fig. 1.11 Localized Fourier and wavelet paving of time–scale space.

By-Passing the Curse of Heisenberg. Heisenberg's principle states that in modeling time-frequency phenomena, we cannot be precise in the time domain and in the frequency domain simultaneously. In other words, squares and rectangles in the pavement of the time–scale plane (as given in Fig. 1.11) have areas bounded from below by a universal constant.

Wavelets automatically trade-off the time-frequency precision by their innate nature. The parsimony mentioned above can be ascribed to the ability of wavelets to cope with limitations of Heisenberg's principle in a data-dependent manner.

Whitening Property. There is ample theoretical and empirical evidence that wavelet transformations tend to simplify the dependence structure in the original data. It is even possible to construct a biorthogonal basis that will decorrelate a given stationary time series (a wavelet-counterpart of the Karhunen-Loève transformation). For a discussion and examples, see Walter [439].

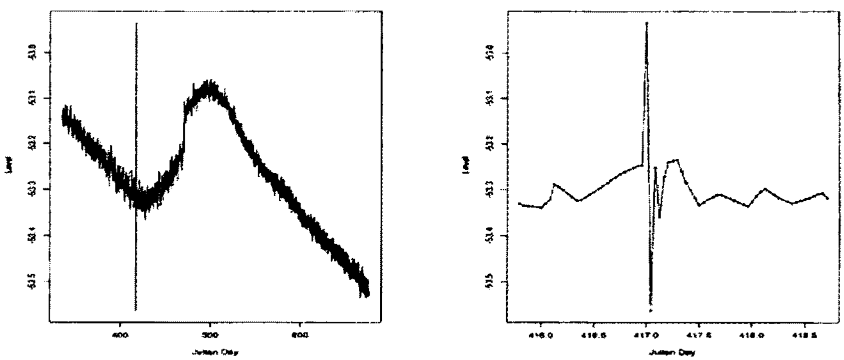
Smoothness Control. Under mild conditions wavelets provide unconditional bases for many important smoothness spaces ($\mathbb{L}_p, p > 1$; Besov Spaces $\mathbb{B}^{\sigma}_{pq}$; Hölder Spaces $\mathbb{C}^{\alpha}$). Using simpler terminology, this means that by controlling the magnitude of the coefficients in the wavelet domain one controls the smoothness of the decomposed function. This connection provides the theoretical framework for wavelet smoothing and wavelet function and density estimation.

1.4 AN APPETIZER: CALIFORNIA EARTHQUAKES

We conclude this introductory chapter with a real-life example. The example we provide emphasizes basic differences between wavelet-based and standard denoising methods. It shows the ability of wavelets to "zoom-in" and adapt their space-scale "descriptors" to the data at hand.

A researcher from the geology department at Duke University was interested in the possibility of predicting earthquakes by monitoring water–levels in the nearby wells. To do this, he obtained water level measurements from six wells located in California that were taken every hour for approximately six years. The goal was to smooth the data, eliminate the noise, and inspect the signal at pre-earthquake time. Here is some background (provided by Dr. Stuart Rojstaczer, Duke University).

> The ability of water wells to act as strain meters has been observed for centuries. The Chinese, for example, have records of water flowing from wells prior to earthquakes. Lab studies indicate that a seismic slip occurs along a fault prior to rupture. Recent work has attempted to quantify this response, in an effort to use water wells as sensitive indicators of volumetric strain. If this is possible, water wells could aid in earthquake prediction by sensing precursory earthquake strain. Water level records from six wells in southern California are collected over a six year time span. At least 13 moderate size earthquakes (Magnitude 4.0 - 6.0) occurred in close proximity to the wells during this time interval. There is a significant amount of noise in the water level record which must first be filtered out. Environmental factors such as earth tides and atmospheric pressure create noise with frequencies ranging from seasonal to semidiurnal. The amount of rainfall also affects the water level, as do surface loading, pumping, recharge (such as an increase in water level due to irrigation), and sonic booms, to name a few.



(a) (b)

Fig. 1.12 (a) California water-level data set. (b) Water-level oscillation at the earthquake time.

Once the noise is subtracted from the signal, the record can be analyzed for changes in water level, either an increase or a decrease depending upon whether the aquifer is experiencing a tensile or compressional volume strain, just prior to an earthquake.

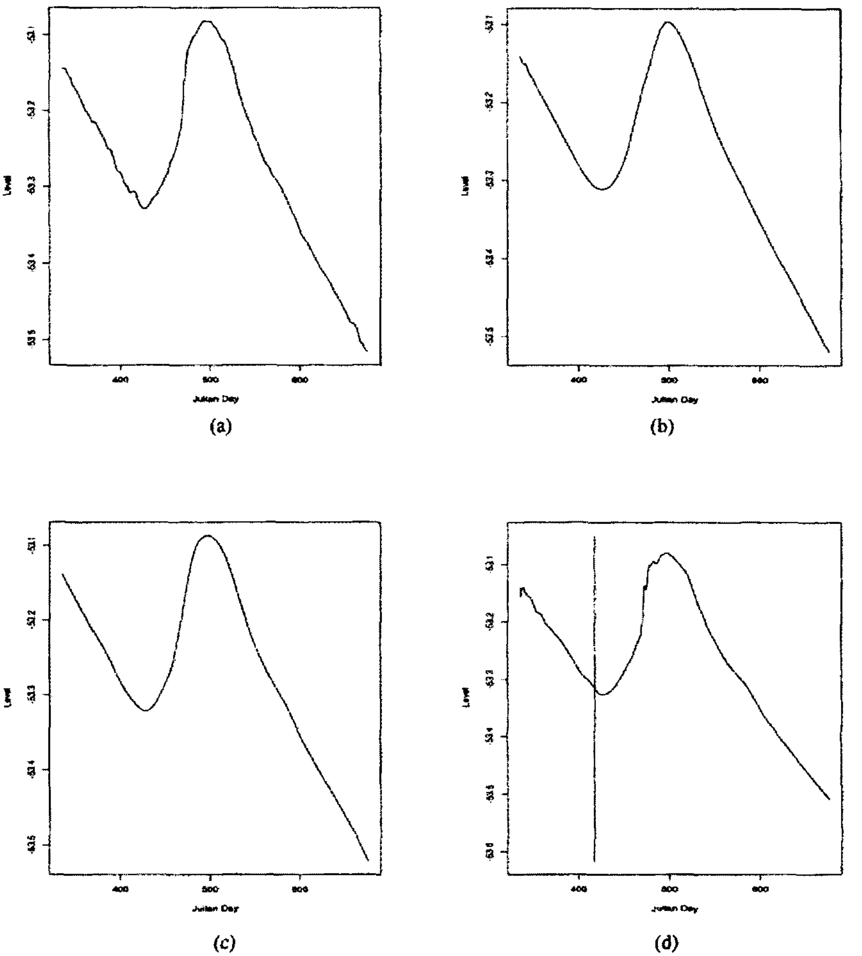


Fig. 1.13 Comparison of several smoothing methods. (a) Data smoothed by kernel method (normal window, $k = 5$); (b) Data smoothed by `loess` method; (c) Data smoothed by `supsmu` method; and (d) Wavelet smoothed data (Daubechies' wavelet with four vanishing moments).

A plot of the raw data for hourly measurements collected over one year (8192 =

2^{13} observations), is given in Fig. 1.12(a). The line-like artifact [enlarged in Fig. 1.12(b)] represents a line connecting two extreme level values at the earthquake time (Julian day of 417).

The measurements were smoothed by three traditional methods (kernel, lowess, and supsmu) and by wavelet shrinkage. Fig. 1.13(a), (b), and (c) are processed data, smoothed by the kernel method (normal kernel, bandwidth = 5), by the locally weighted regression smoother (implemented in S-Plus as `lowess`), and by the local cross-validation smoother (implemented in S-Plus as `supsmu`) method.

Rather than discussing whether the filtering indicates that the earthquake could have been predicted, we emphasize differences in the outputs of traditional and wavelet smoothing methods. Notice that in all traditional methods the artifact of interest (earthquake jump) is lost. The application of nonlinear wavelet shrinkage to the data, results in a smooth signal with the jump at earthquake time preserved. The wavelet-smoothed data are given in Fig. 1.13(d). Only 20 of 8192 (0.244%) coefficients (those 20 with largest magnitude) were used in describing the wavelet shrinkage estimator.

This Page Intentionally Left Blank

2

Prerequisites

In this chapter, we introduce notation and briefly review several mathematical concepts necessary for the definition and derivation of the basic properties of wavelets. Some fundamental concepts from the theory of Hilbert spaces, Fourier analysis, linear algebra, and signal processing will be used to define the multiresolution analysis and to develop wavelet formalism.

2.1 GENERAL

For denoting the sets of natural, integer, real, and complex numbers we use notation $\mathbb{N}, \mathbb{Z}, \mathbb{R}$, and $\mathbb{C}$. The modulus of a complex number $z \in \mathbb{C}$ will be denoted by $|z|$, and the complex conjugate by $\bar{z}$. The set of positive real numbers will be denoted by $\mathbb{R}^+$. It is tacitly assumed that all functions are measurable. The support of a function f, denoted supp(f), is the closure in $\mathbb{R}$ of the set $\{x \in \mathbb{R} : f(x) \neq 0\}$.

The indicator of a relation ρ, $\mathbf{1}(\rho)$, is defined to be 1 if the relation ρ is satisfied and 0 otherwise. The Kronecker delta $\delta_{u,v}$ can be defined using the indicator function as $\mathbf{1}(u = v)$. We also define δ_u to be $\delta_{u,0}$. Maximum and minimum of a and b are denoted by $a \vee b$ and $a \wedge b$, respectively.

Let $f_+ = f \cdot \mathbf{1}(f \geq 0) = f \vee 0$ be a positive part of a function f and $f_- = -f \cdot \mathbf{1}(f \leq 0) = -(f \wedge 0)$ be its negative part. By definition, $|f| = f_+ + f_-$ and $f = f_+ - f_-$. We will sometimes use O-notation; $a_n = O(b_n)$ would mean that a_n/b_n is asymptotically bounded away from 0 and ∞; $a_n = o(b_n)$ would mean that $a_n/b_n \to 0$.

A Lebesgue point of a function f is any point x such that

$$\lim_{r \to 0} \frac{1}{2r} \int_{-r}^{r} |f(x+t) - f(x)| \, dt = 0.$$

The Dirac function $\delta(x)$ (not to be confused with the Kronecker symbol $\delta_{k,l}$) is defined as

$$\delta(x) = \lim_{a \to 0} \frac{1}{a} \mathbf{1}(0 \le x \le a). \tag{2.1}$$

The Dirac function (2.1) satisfies the following relations

$$\int_{\mathbb{R}} \delta(x) \, dx = 1,$$
$$\int_{\mathbb{R}} f(x)\delta(x - x_0) \, dx = f(x_0).$$

It can be thought of as a generalized derivative of a Heaviside step function $H(x) = \mathbf{1}(x \ge 0)$.

2.2 HILBERT SPACES

Hilbert spaces are natural generalizations of finite dimensional Euclidean spaces $\mathbb{R}^n$.

Working with abstract Hilbert spaces is beneficial in several respects. Our geometric intuition, based on properties of Euclidean $\mathbb{R}^2$ or $\mathbb{R}^3$ spaces, can in part be easily extended to an arbitrary Hilbert space. An example is the *projection theorem* (Theorem 2.2.1). The norm in the Hilbert space is connected with a quadratic expression and the process of norm-minimization falls in the class of linear problems. All separable Hilbert spaces are (abstractly) equivalent to one another.

The Inner Product Space. A complex vector space $\mathcal{H}$ is said to be an inner product space if for any two elements $x, y \in \mathcal{H}$ there exists a complex number $\langle x, y \rangle$ (called the inner product of x and y) that satisfies

(i) $\langle x, y \rangle = \overline{\langle y, x \rangle}$
(ii) $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$, for all x, y, and $z \in \mathcal{H}$.
(iii) $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$ for $x, y \in \mathcal{H}$ and $\alpha \in \mathbb{C}$.
(iv) $\langle x, x \rangle \ge 0$, for all $x \in \mathcal{H}$.
(v) $\langle x, x \rangle = 0$, if and only if $x = 0$.

The norm $||x||$ of an element $x \in \mathcal{H}$ is defined via inner product, $||x|| = \sqrt{\langle x, x \rangle}$.

Example 2.2.1 Euclidean space $\mathbb{R}^n$.

$x = (x_1, \ldots, x_n)$
$y = (y_1, \ldots, y_n)$
$\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i, \quad \|x\| = \sqrt{\sum_{i=1}^{n} x_i^2}$

Example 2.2.2 $\mathbb{L}_2(\mathbb{R})$ space (space of all square-integrable functions).

$f \in \mathbb{L}_2(\mathbb{R})$ if $\int |f|^2 < \infty$.
$\langle f, g\rangle = \int fg, \quad \|f\| = \sqrt{\int f^2}$.
If $f, g \in \mathbb{L}_2(\mathbb{C})$, $\langle f, g\rangle = \int f\bar{g}$ and $\|f\| = \sqrt{\int f\bar{f}}$.

Example 2.2.3 ℓ_2 space (space of all square-summable sequences).

$\underset{\sim}{x} = \{x_n\} \in \ell_2$ if $\sum_{n\in\mathbb{Z}} |x_n|^2 < \infty$.
$\langle \underset{\sim}{x}, \underset{\sim}{y}\rangle = \sum_{i\in\mathbb{Z}} x_i \overline{y_i}, \quad \|x\| = \sqrt{\sum_{i\in\mathbb{Z}} |x_i|^2}$.

Example 2.2.4 A function f belongs to the *Lebesgue space* $\mathbb{L}_p(\mathbb{A})$, $1 \le p < \infty$, if

$$\|f\|_p = \left(\int_{\mathbb{A}} |f(x)|^p \, dx\right)^{1/p} < \infty, \quad \text{and}$$
$$\|f\|_\infty = \operatorname{ess}\sup_{x\in\mathbb{A}} |f(x)| < \infty.$$

To "upgrade" the linear space $\mathcal{H}$ equipped with a norm to the Hilbert space one needs the *completeness property.*

Definition 2.2.1 *The sequence $\{x_n\}_{n\in\mathbb{N}}$ is called a Cauchy sequence in $\mathcal{H}$ if and only if (iff)*

$$\|x_m - x_n\| \to 0,$$

whenever $m, n \to \infty$.

The space $\mathcal{H}$ is complete if any Cauchy sequence $\{x_n\}$ is convergent, i.e., $x_n \to x \in \mathcal{H}$.

2.2.1 Projection Theorem

A linear subspace $\mathcal{V}$ of a Hilbert space $\mathcal{H}$ is said to be a closed subspace of $\mathcal{H}$ if $\mathcal{V}$ contains all its limiting points, i.e., if $x_n \in \mathcal{V}$ and $\|x_n - x\| \to 0$, as $n \to \infty$, then $x \in \mathcal{V}$.