Ph.D. Dissertation

Signal Recovery and Noise Reduction with Wavelets

by Jian Lu

Thayer School of Engineering
Dartmouth College
Hanover, New Hampshire

June 1993

Copyright © 1993, 1995 by Jian Lu

Signal Recovery and Noise Reduction with Wavelets

A Thesis
Submitted to the Faculty
in partial fulfillment of the requirements for the
degree of

Doctor of Philosophy

by

JIAN LU

Thayer School of Engineering
Dartmouth College
Hanover, New Hampshire

JUNE 1993

Examining Committee:

_____
Dr. Eric W. Hansen, Chairman

_____
Dr. John B. Weaver, Co-Chairman

_____
Dr. Dennis M. Healy, Jr.

_____
Dr. Keith D. Paulsen

_____
Dr. Yansun Xu

_____
Dean of Graduate Studies

_____
Author

TO MOM, DAD, AND CHESSY

Thayer School of Engineering
Dartmouth College

SIGNAL RECOVERY AND NOISE REDUCTION WITH WAVELETS

JIAN LU

Doctor of Philosophy

ABSTRACT

Signal recovery and noise reduction are closely related signal processing problems of both theoretical and practical interest. In this thesis, these classical problems are studied with a new tool—wavelets.

For a class of signal degradations where the degrading operator is linear shift-invariant and of lowpass filter type, the signal recovery problem is reformulated as finding missing data at the finest scale of a discrete wavelet transform. In this framework, the wavelet reproducing equation plays a fundamental role in determining a unique and stable recovery. It is shown that the solution to the stabilized wavelet reproducing equation is equivalent to a constrained least-squares solution. By discovering such a relationship, a new class of regularizing operators is obtained, and the wavelet-regularized solution to the signal recovery problem is derived in one and two dimensions. In experiments, the wavelet-regularized image restoration results in 75-85% reduction in mean square errors in noise-free cases, and 35-60% reduction in noisy cases.

For signal degradations caused solely by noise contamination, a new technique for noise reduction is developed based on the wavelet maxima representation. The concept of the wavelet maxima tree is formalized and algorithms for constructing such trees are developed for 1-D signals and 2-D images. Metrics are designed for the wavelet maxima tree to measure the scaling and spatial stabilities of wavelet maxima. These metrics reflect

some perceptual criteria for discriminating objects from a noisy background. A complete computational framework for denoising digital signals and images is implemented. In experiments with both simulated and real signals and images, the new technique is able to reduce noise power by more than 90% and keep the edge gradients to 80-120% of their original values. With these results, the new technique outperforms widely used Wiener and median filters in solving the tradeoff between smoothing noise and preserving edges.

Acknowledgments

I am indebted to many people who helped make this thesis possible. First, I would like to offer thanks to all members of my thesis Committee. To my research advisors Dr. John Weaver and Professor Dennis Healy, for introducing me to the fascinating world of wavelets. Their forward-looking vision initialized the research project underlying this thesis; their unwavering support and careful guidance were essential to the fruitful completion of the project. Working with both of them has been a real pleasure. To my academic advisor Professor Eric Hansen, for teaching me fundamentals of signal and image processing. He is always ready to answer my questions whenever I pop into his office. To Professor Keith Paulsen, for advising me on numerical analysis. I could go nowhere without this tool. To our former group member Dr. Yansun Xu, for sharing the excitement about wavelets and kindly serving in my Committee. His assistance and cooperation in shaping the early stage of the project helped assure its success.

I owe special thanks to the wavelet research group at Courant Institute, New York University. To Dr. Stephane Mallat, for kindly making their wavelet software available to us, and for his influence on my research pursuits. To Dr. Sifen Zhong, for advice and discussion about the wavelet maxima representation. To Wen Liang Hwang, for generously providing their images for our tests, and their results for comparison.

Special thanks also go to the staff of Feldberg Library of Dartmouth College, in particular, the former Interlibrary Loan Specialist Elizabeth Mulherrin and Librarian James Fries. They helped acquire a number of technical reports crucial to my research work years before these reports were published in journals. The excellent professional service by the staff of Feldberg Library is greatly appreciated.

Finally, but by no means the least, I thank my family for their unfailing love and support during the time I devoted to doctoral study. To my loving wife Chessy, for her

understanding, tolerance, and sacrifice. To my mother and father, for their encouragement and trust. This thesis is dedicated to my family. I love them dearly.

The work presented in this thesis is part of a research project that was supported in part by a grant from the Whitaker Foundation and a DARPA contract administered by the AFOSR as AFOSR-90-0294.

Table of Contents

List of Tables viii

List of Figures ix

PART I: PRELIMINARIES

Chapter 1: Introduction and Overview 2
1.1 Signal Recovery and Noise Reduction—The Problem 2
1.2 Wavelets and Signal Processing 3
1.3 Overview of the Thesis 6

Chapter 2: Basics of Wavelets and Wavelet Transforms 9
2.1 Continuous Wavelet Transforms 9
2.2 Dyadic Wavelet Transforms 10
2.3 Discrete Wavelet Transforms 12
2.4 Two-Dimensional Wavelet Transforms 19

PART II: SIGNAL RECOVERY AND WAVELET REPRODUCING KERNELS

Chapter 3: Wavelet-Regularized Signal Recovery in One Dimension 25
3.1 Review of Regularized Signal Recovery 25
3.1.1 Constrained Least-Squares Estimation and Regularization 25
3.1.2 Properties and Choice of Regularizing Parameter and Operator 27
3.2 Regularized Signal Recovery with Wavelet Reproducing Kernels 28
3.2.1 Reformulation of the Problem 28
3.2.2 Signal Recovery with Stabilized Wavelet Reproducing Equation 29
3.2.3 The Equivalent Least-Squares Solutions 30
3.2.4 Remarks 31
3.3 Wavelet-Regularized Signal Recovery—An Example 34

Chapter 4: Wavelet-Regularized Signal Recovery in Two Dimension 39
4.1 Extensions of One-Dimensional Results 39
4.1.1 The Two-Dimensional Signal Recovery Problem 39
4.1.2 The Solutions 40

4.1.3 Remarks 42
4.2 Wavelet-Regularized Image Restoration—Examples 43
4.2.1 Example I 44
4.2.2 Example II 45
4.2.3 Comments on Experimental Results 47

PART III: NOISE REDUCTION WITH THE WAVELET MAXIMA REPRESENTATION

Chapter 5: The Wavelet Maxima Representation 54
5.1 The One-Dimensional Wavelet Maxima Representation 54
5.1.1 Wavelet Maxima as Multiscale Edges 54
5.1.2 The One-Dimensional Analyzing Wavelet 57
5.1.3 Representation of 1-D Signals by Wavelet Maxima 60
5.1.4 Reconstruction of 1-D Signals from Wavelet Maxima 61
5.2 The Two-Dimensional Wavelet Maxima Representation 63
5.2.1 Representation of Images by Wavelet Maxima 63
5.2.2 Reconstruction of Images from Wavelet Maxima 64

Chapter 6: The Wavelet Maxima Tree and Metrics 67
6.1 Motivations 67
6.2 The Wavelet Maxima Tree of One-Dimensional Signals 71
6.2.1 Definitions and Properties 71
6.2.2 Construction of One-Dimensional Wavelet Maxima Tree 73
6.3 The Wavelet Maxima Tree of Images 77
6.3.1 Extended Definitions and Properties 77
6.3.2 Contouring Wavelet Maxima 79
6.3.3 Construction of Two-Dimensional Wavelet Maxima Tree 79
6.4 Metrics on the Wavelet Maxima Tree 83
6.4.1 Object/Noise Discrimination in Visual Perception 84
6.4.2 Design of Metrics for Object/Noise Discrimination 85

Chapter 7: Experimental Study of Noise Reduction 88
7.1 Implementation of the Noise Reduction Technique 88
7.1.1 The Computational Framework 88
7.1.2 Threshold Selection 90
7.2 Experimental Results 92
7.2.1 One-Dimensional Results 92

7.2.2 Two-Dimensional Results.....98
7.3 Comparative Analysis of Experimental Results.....114
7.3.1 Comparison of One- and Two-Dimensional Algorithms.....114
7.3.2 Comparison with Wiener and Median Filters.....115
7.3.3 Comparison with Mallat and Hwang's Method.....116

PART IV: CONCLUSIONS

Chapter 8: Conclusions and Future Directions.....120
8.1 Conclusions.....120
8.2 Directions of Future Work.....121

References.....126

List of Tables

Table 3.1. Experimental results from recovering blurred "prostate" scan lines..... 35
Table 4.1. Experimental results from restoring blurred "head" images..... 45
Table 4.2. Experimental results from restoring blurred "knee" images..... 46
Table 6.1. First ten coefficients of the discrete reproducing kernel $K_{j,j-1}$ 76
Table 7.1. Execution time percentage for each program block..... 90
Table 7.2. SNR and edge gradients of a denoised signal..... 93
Table 7.3. SNR values of noisy and denoised phantom images.....100
Table 7.4. SNR values of noisy and denoised MR head images. 100
Table 7.5. SNR values of noisy and denoised photographic images. 100

List of Figures

Figure 2.1. An example of one-dimensional DWT..... 18
Figure 2.2. An example of two-dimensional DWT. 22
Figure 3.1. A scan line from an MR image of prostate. 36
Figure 3.2. Recovery of the blurred noise-free "prostate" scan line. 37
Figure 3.3. Recovery of the blurred noisy "prostate" scan line..... 38
Figure 4.1. An MR head image..... 48
Figure 4.2. An MR knee image..... 48
Figure 4.3. Restoration of the noise-free "head" image blurred by a Fejér kernel..... 49
Figure 4.4. Restoration of the noisy "head" image blurred by a Fejér kernel..... 50
Figure 4.5. Restoration of the noise-free "knee" image blurred by a Gaussian kernel. 51
Figure 4.6. Restoration of the noisy "knee" image blurred by a Gaussian kernel..... 52
Figure 5.1. The local extrema of $|W^1_{2^j} f(x)|$ and zero-crossings of $|W^2_{2^j} f(x)|$. 56
Figure 5.2. A cubic spline scaling function and wavelet. 58
Figure 5.3. A test signal and its wavelet transform with a cubic spline wavelet..... 59
Figure 5.4. The wavelet maxima representation of a signal..... 60
Figure 5.5. A signal reconstructed from its wavelet maxima. 62
Figure 5.6. The wavelet maxima representation of an MR image 65
Figure 6.1. Evolution properties of wavelet maxima and extrema in the wavelet scale-space. 69
Figure 6.2. Typical behavior of zero-crossings in the Gaussian scale-space. 70
Figure 6.3. Possible behaviors of zero-crossings..... 70
Figure 6.4. The wavelet maxima tree of a signal. 77
Figure 6.5. The splitting and merging of zero-crossing contours in the Gaussian scale-space 78
Figure 6.6. Design of contour tracking window. 81
Figure 6.7. The valued wavelet maxima tree of a signal..... 87
Figure 7.1. The computational framework of the noise reduction technique..... 89
Figure 7.2. A histogram of P_b values for a wavelet maxima tree..... 91
Figure 7.3. The trimmed wavelet maxima tree..... 92
Figure 7.4. A denoised signal. 94
Figure 7.5. A simulated noise-free signal with different edge shapes. 94

Figure 7.6. Signal recovery from noisy observations. 95
Figure 7.7. Signal recovery from severely noisy observations..... 96
Figure 7.8. Denoising a MR image scan line. 97
Figure 7.9. A simulated noise-free phantom image and two line profiles. 101
Figure 7.10. The noisy phantom image (NP-I) and its two line profiles.....102
Figure 7.11. A denoised image from NP-I by using the 1-D algorithm.....103
Figure 7.12. A denoised image from NP-I by using the 2-D algorithm.....104
Figure 7.13. The noisy phantom image (NP-II) and its two line profiles.....105
Figure 7.14. A denoised image from NP-II by using the 1-D algorithm.....106
Figure 7.15. A denoised image from NP-II by using the 2-D algorithm.....107
Figure 7.16. The noisy and denoised MR head images (Head-I).....108
Figure 7.17. The noisy and denoised MR head images (Head-II).....109
Figure 7.18. Difference images between the noisy Head-I and denoised images.....110
Figure 7.19. Difference images between the noisy Head-II and denoised images.....111
Figure 7.20. The noisy and denoised House images.....112
Figure 7.21. The noisy and denoised Pepper images. 113

Part I

PRELIMINARIES

Chapter 1
Introduction and Overview

1.1 Signal Recovery and Noise Reduction—The Problem

Any signal acquired by electronic, electro-mechanical or electro-optical means is likely to be degraded by the sensing environment. In a large class of signal degradation problems, the sensor characteristic function is, or can be considered, linear shift-invariant and of lowpass filter type. This kind of degradation can be called "blurring" after what happens in an image. The recovery of the original signal from the blurred one, or deblurring, in continuous-time amounts to finding a solution of the Fredholm integral (convolution type) equation of the first kind. In the one-dimensional case, we solve for the original signal $x(t)$ from

$$y(t) = \int_a^b h(t-s)x(s)\,ds \tag{1.1}$$

where $y(t)$ is the degraded signal, and $h(t)$ is the convolution kernel.

Inversion of (1.1) for $x(t)$ is a notoriously ill-posed problem in the sense that a unique solution of x may not exist, or depend continuously on y. In practice, we always deal with discretized form of (1.1) and a finite, discrete observation of y. Furthermore, the observation of y is often contaminated by noise. The resulting observation model is:

$$\underline{y} = \boldsymbol{H}\underline{x} + \underline{n} \tag{1.2}$$

where $\underline{y}$ and $\underline{n}$ are n×1 vectors, $\underline{x}$ is an m×1 vector and $\boldsymbol{H}$ is an n×m Toeplitz matrix. In the discrete case, the stability of the solution is often further degraded with the ill-conditioned $\boldsymbol{H}$ because the inversion amplifies noise by a factor proportional to the condition number of $\boldsymbol{H}$.

A special case of (1.2) is when the degradation of a signal is caused solely by noise contamination, that is,

$$\underline{y} = \underline{x} + \underline{n} \tag{1.3}$$

Although the situation of (1.3) is less complicated than that of (1.2), the recovery of $\underline{x}$ from the noisy observation $\underline{y}$ is by no means a trivial problem. To comprehend the difficulty, let us think about lowpass filtering—the most frequently used method to smooth noise. The costs of doing a lowpass filtering will be: 1) fine structures of the signal are destroyed, and 2) surviving (large) structures are blurred.

Generally speaking, the signal recovery and noise reduction problems defined by (1.2) and (1.3) both involve a tradeoff between the spatial/time resolution and the signal-to-noise ratio (SNR). Ideally, one wishes to restore or preserve resolution while smoothing noise. Such dual demands are incompatible. Nevertheless, one should be able to maximize the gain and achieve an "optimal" solution by certain criteria. The topic of this thesis is to study the signal recovery and noise reduction problems by using wavelets.

1.2 Wavelets and Signal Processing

Over the last few years, there has been increasing awareness of and interest in the emerging theory and application of wavelets. In the signal processing community, wavelets have become prominent. This is reflected by the number of wavelet-related papers presented at the annual IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) in the past three years. At ICASSP'91, there were 14 papers in this category spread in regular conference sessions[1]. Last year, there were 42 along with two special sessions[2]. This year, there are 53 along with four related special sessions and one tutorial session[3]. Besides ICASSP, there have been a number of journal special issues and conferences dedicated to wavelets in the context of signal processing. Most notable are the Special Issue on Wavelet Transforms and Multiresolution

Signal Analysis published by the IEEE Transactions on Information Theory in March 1992, and the IEEE-SP International Symposium on Time-Scale and Time-Frequency Analysis held in Victoria in October 1992. Another special issue dedicated to wavelets in signal processing will be published by IEEE Transactions on Signal Processing later in this year.

Briefly speaking, wavelets are functions whose dilations and translations form a basis of the vector space of measurable, square-integrable functions. Decomposition of a function on a wavelet basis constitutes a wavelet transform. A precise definition of wavelet and the wavelet transform (WT) will be given in Chapter 2. Two features of wavelet bases are most attractive: the constructible orthogonality and good localization in both spatial/time and Fourier domains[70]. A vivid interpretation of the WT comes from signal analysis: a wavelet transformed signal is a multiresolution (or multiscale) representation of the signal. In computation, Mallat's fast algorithm for the discrete wavelet transform (DWT)[46] has made this valuable tool computationally inexpensive.

Applications of wavelets have been explored in many areas of signal processing, such as image coding and compression[6,23,84], data modeling[8,9,78], signal estimation[79], medical imaging[28,75], signal recovery[15,24,44] and noise reduction[25,43,53,74,81], to name only a few. The research work presented in this thesis focuses on using wavelets to solve the signal recovery and noise reduction problems defined by (1.2) and (1.3).

Regarding signal recovery, Bruneau *et al*[15] proposed a method of regularized image restoration with wavelets. They used the highpass filter associated with a biorthogonal DWT as a regularizing operator along with the standard regularization technique. In their experiments, the wavelet-regularized restoration produced similar results to those of Laplacian-regularized restoration. However, the authors of [15] did not give a formal analysis of the wavelet-regularized solution, nor did they elucidate the motivation for choosing a wavelet regularizing operator. This makes their approach purely synthetic. In

contrast to [15], Lu *et al*[44] took an analytic approach. They reformulated the signal recovery problem (1.2) as solving a stabilized wavelet reproducing equation. This was found equivalent to solving a constrained least-squares minimization problem, and the wavelet reproducing kernel arises as a regularizing operator. The work in [44] will be detailed in Part II of the thesis.

Very recently, Donoho[24] developed a general formalism for solving linear inverse problems which include (1.2). The method, called "wavelet-vaguelette decomposition" (WVD), resembles the method of singular value decomposition (SVD), but uses a wavelet-like basis instead of an eigenfunction basis. Donoho showed that by non-linear shrinkage of WVD coefficients, the lower bound of minimax risk can be attained. While [24] provides striking theoretical results, their implementation in one and two dimensions does not seem to be a simple task.

The early work on noise reduction with wavelets was very like lossy compression. An orthogonal WT of a noisy signal is computed. A threshold is chosen and those WT coefficients with amplitude below the threshold are set to zero. The choice of the threshold is usually empirical[74]. Most recently, Donoho[25] formalized this approach by an optimal soft-thresholding technique. He showed that this technique outperforms almost all linear smoothing techniques with regard to minimax measures. He also defined the term "denoise" that he thinks more expressive than "noise reduction". In this thesis, the two terms will be used to refer to the same signal processing task. We notice that the soft-threshold in [25] for denoising is dependent on the variance of the noise, or alternatively, the SNR. This property aligns the technique with a large class of smoothing filters that are adapted to SNR. From the philosophical point of view, it is somewhat paradoxical that a filter aimed at lifting SNR would be dependent on SNR. As a consequence, these SNR-adapted filters are incapable of denoising at low SNR. This is

not surprising because from a statistical point of view, we have no way to discriminate a weak signal that is submerged in noise.

By using a non-orthogonal WT, Xu *et al*[81] developed a noise filtration technique by direct multiplication of WT coefficients at several scales. This method is a special case of cross-scale correlation. The random nature of noise results in a weak response in correlation between noisy WT coefficients at different scales, thus providing a filtering mechanism.

Another type of noise reduction technique[43,53] employs the wavelet maxima representation developed by Mallat and Zhong[48,50]. We are among a few groups developing such techniques[41-43]. Unlike most noise reduction techniques, these are based on feature analysis in a feature space composed of wavelet maxima or "multiscale edges". Wavelet maxima are analyzed one by one to determine their likelihood to be noise-driven. Techniques of this type seem to vary only in the way that feature analysis is carried out. For example, Mallat and Hwang[53] used singularity analysis, while Lu *et al*[43] exploited perceptual criteria. The development of a noise reduction technique based on the wavelet maxima representation and perceptual criteria will be a major topic in this thesis.

1.3 Overview of the Thesis

This thesis explores the application of wavelets to solving the signal recovery and noise reduction problems (1.2) and (1.3), which still seems to be a new area of wavelet applications. With the exception of the very recent work by Donoho[24,25], the author has not seen other work on the same topic that is as extensive as what will be described in this thesis. To summarize, the main contributions of this thesis are:

1. reformulation of the signal recovery problem (1.2) as solving a stabilized wavelet reproducing equation. This represents an approach that is conceptually different from an "inverse" approach.

2. derivation of analytic solutions to the stabilized wavelet reproducing equation in one- and two-dimensions, and analysis of asymptotic properties of these solutions.
3. elucidation of a connection between constrained least-squares minimization and the stabilized wavelet reproducing equation. This led to the discovery of a new class of regularizing operators that are adapted to specific degrading operators.
4. conception and construction of the wavelet maxima tree in one and two dimensions. This is an extension to the wavelet maxima representation developed by Mallat and Zhong.
5. design of metrics on the wavelet maxima tree to measure the scaling and spatial stabilities of wavelet maxima. These metrics reflect some perceptual criteria for discriminating objects from a noisy background.
6. development of a complete computational framework and experimental study of noise reduction with the wavelet maxima tree and perception-based metrics.

This thesis divides into four parts. Part I is an introduction that contains two chapters including this one. The next one—Chapter 2 reviews the basics of wavelets and wavelet transforms. Part II of the thesis focuses on solving the signal recovery problem (1.2) with wavelet reproducing kernels. This part contains two chapters. In Chapter 3, we start with a brief introduction to the least-squares and regularization techniques used in signal recovery. Then the signal recovery problem (1.2) is reformulated within the wavelet framework. The solution is derived in one dimension, and its relations to least-squares and regularization techniques are studied. An example of wavelet-regularized recovery of a one-dimensional signal is given. Chapter 4 extends the model and results in Chapter 3 to two dimensions, and presents experimental results from wavelet-regularized image restoration. Part III of the thesis varies in theme. With three chapters Part III is devoted to solving the noise reduction problem (1.3) with the wavelet maxima representation. In Chapter 5, a comprehensive review is given to the wavelet maxima representation

developed by Mallat and Zhong. Next, in Chapter 6, the concept of a wavelet maxima tree is put forward, and the method for constructing such a tree is detailed. Additionally, several metrics that reflect perceptual criteria are designed for noise reduction with the wavelet maxima tree. Chapter 7 presents an experimental study of noise reduction with the wavelet maxima tree and perception-based metrics. First, a complete computational framework is sketched and an automated threshold selection technique is presented. Then, experiments on noise reduction are conducted with the proposed method and some other techniques, and experimental results are analyzed in a comparative way. Finally, Part IV concludes the thesis by summarizing conclusions from Part II and Part III and looking into the directions for future development.

Chapter 2

Basics of Wavelets and Wavelet Transforms

This chapter reviews the basics of wavelets and wavelet transforms (WTs). The review is not extensive. Instead, the coverage is kept as small as possible. Given the abundant literature on basic wavelet theories, one can easily find extensive presentations on the topic of interest. For example, [19,27,46] have now become classics, while [21] is an authoritative monograph covering fundamental issues of wavelets, and is becoming an increasingly popular textbook. Some material about discrete wavelet reproducing kernels is not seen in the wavelet literature but is contained in this chapter.

2.1 Continuous Wavelet Transforms

The continuous wavelet transform (CWT) was first developed by Grossmann and Morlet[27]. Let $\psi(x) \in \boldsymbol{L}^2(\boldsymbol{R})$ be an admissible mother wavelet that satisfies:

$$C_\Psi = \int_0^{+\infty} |\Psi(\omega)|^2 \frac{d\omega}{\omega} = \int_{-\infty}^{0} |\Psi(\omega)|^2 \frac{d\omega}{\omega} = 1 \tag{2.1}$$

where $\Psi(\omega)$ is the Fourier transform of $\psi(x)$. Let $\psi_s(x)$ be the dilation of $\psi(x)$ by a factor $s \in \boldsymbol{R}^+$,

$$\psi_s(x) = \frac{1}{s}\psi(\frac{x}{s})$$

The CWT of a function f at scale s and position x in scale-space plane is given by:

$$\boldsymbol{W}: \qquad W_s f(x) = \psi_s * f(x) \tag{2.2}$$

where "*" denotes convolution.

The CWT is an isometry, i.e.,

$$\| f(x) \|^2 = \int_0^{+\infty} \| W_s f(x) \|^2 \frac{ds}{s}$$

The inverse CWT can then be obtained by:

$$\boldsymbol{W}^{-1}: \quad f(x) = \int_0^{+\infty} \tilde{\psi}_s * W_s f(x) \frac{ds}{s} \tag{2.3}$$

where $\tilde{\psi}(x) = \psi(-x)$.

An arbitrary two variable function $g(s, x) \in \boldsymbol{L}^2(\boldsymbol{R}^2)$ is not necessarily the CWT of some function. For $g(s, x)$ to be in the range of $\boldsymbol{W}$, it must satisfy the reproducing equation[27],

$$g(s, x) = \int_0^{+\infty} ds' \int_{-\infty}^{+\infty} K(s, s'; x, x') g(s', x') \, dx' \tag{2.4}$$

where

$$K(s, s'; x, x') = \int_{-\infty}^{+\infty} \tilde{\psi}_s(u-x) \psi_{s'}(x'-u) \, du \tag{2.5}$$

is the wavelet reproducing kernel (WRK).

2.2 Dyadic Wavelet Transforms

For practical applications, the scale-space must be discretized. The scale parameter s is often discretized to a dyadic sequence, $\{2^j\}_{j \in \mathbf{Z}}$. If we further impose that

$$\sum_{j=-\infty}^{+\infty} | \Psi(2^j \omega) |^2 = 1 \tag{2.6}$$

the dyadic WT pair corresponding to (2.2) and (2.3) is given by:

$$\boldsymbol{W}: \qquad Wf(2^j, x) = \psi_{2^j} * f(x) \tag{2.7}$$

$$\boldsymbol{W}^{-1}: \quad f(x) = \sum_{j=-\infty}^{+\infty} \widetilde{\psi}_{2^j} \ * \ W_{2^j} f(x) \tag{2.8}$$

(2.6) is an over-restrictive condition on ψ for dyadic WTs. Daubechies[20] has shown that this condition can be relaxed. As long as there exist two constants, $A > 0$ and $B < \infty$, such that,

$$A \le \sum_{j=-\infty}^{+\infty} | \ \Psi(2^j \omega) \ |^2 \le B \tag{2.9}$$

the dyadic WT defined by (2.7) will be invertible. In general, the synthesis wavelets in (2.8) will not be $\{ \widetilde{\psi}_{2^j} \}$. Instead, they are some other dyadic functions $\{ \xi_{2^j} \}$that satisfy

$$\sum_{j=-\infty}^{+\infty} \Xi(2^j \omega) \Psi(2^j \omega) \ = 1 \tag{2.10}$$

where

$$\Xi(\omega) \ = \frac{\overline{\Psi}(\omega)}{\sum_{j=-\infty}^{+\infty} | \ \Psi(2^j \omega) \ |^2} \tag{2.11}$$

is the Fourier transform of $\xi(x)$. In such cases, the analyzing wavelets $\{ \psi_{2^j} \}$ are said to constitute a frame in $\boldsymbol{L}^2$, with A and B being the frame bounds. (2.6) is then a special case of (2.9) where $\{ \psi_{2^j} \}$ form a "tight frame"[20] with the frame bounds $A = B = 1$.

Similar to the continuous case, for a sequence $\{g_j(\mathrm{x})\}_{j \in \boldsymbol{Z}}$ to be the dyadic WT of some function, $\{g_j(\mathrm{x})\}$ must satisfy the reproducing equation[48]:

$$\forall j \in \boldsymbol{Z}, \quad g_j(x) = \sum_{i=-\infty}^{+\infty} g_i \ * \ K_{i,j}(x) \tag{2.12}$$

with the WRK defined by,

$$K_{i,j}(x) = \widetilde{\psi}_{2^i} * \psi_{2^j}(x) \tag{2.13}$$

Dyadic wavelet transforms have been studied and used extensively. We use only dyadic wavelet transforms throughout this thesis. From now on, we shall use "wavelet transform" or "WT" for the dyadic wavelet transform in most circumstances.

2.3 Discrete Wavelet Transforms

In practice, we can compute a WT only over finitely many scales. This is because the observed data is limited between a non-zero small (fine) scale and a finite large (coarse) scale. According to Mallat[46], one can normalize the observable finest scale to $2^0=1$, and the coarsest scale to 2^J where J is dependent on the sample size of the data.

Without loss of generality, we can assume that a wavelet $\psi(x)$ satisfies (2.6). Now let us introduce a scaling function $\phi(x)$ and its Fourier transform $\Phi(\omega)$ that satisfies almost everywhere:

$$| \Phi(\omega) |^2 = \sum_{j=1}^{+\infty} | \Psi(2^j\omega) |^2 \tag{2.14}$$

Similar to the dilation of a wavelet, the dilated scaling function and its Fourier transform are given by:

$$\phi_{2^j}(x) = 2^{-j}\phi(2^{-j}x)$$

$$\Phi_{2^j}(\omega) = \Phi(2^j\omega)$$

The approximation to a function $f(x)$ at scale 2^j and the Fourier transform of the approximation are given by,

$$S_{2^j}f(x) = \phi_{2^j} * f(x) \tag{2.15}$$

$$S_{2^j}F(\omega) = \Phi(2^j\omega)F(\omega) \tag{2.16}$$

Consider a finite-scale representation of (2.14),

$$| \Phi(\omega) |^2 = \sum_{j=1}^{J} | \Psi(2^j\omega) |^2 + | \Phi(2^J\omega) |^2 \tag{2.17}$$

By Parseval's theorem it can be shown that:

$$\| S_{2^0}f(x) \|^2 = \sum_{j=1}^{J} \| W_{2^j}f(x) \|^2 + \| S_{2^J}f(x) \|^2 \tag{2.18}$$

(2.18) shows that a finite-scale WT represented by $\{(W_{2^j}f(x))_{j=1,2,\ldots,J}, S_{2^J}f(x)\}$ is an isometry. It is interpreted as a multiresolution representation. The information lost in smoothing $S_{2^j}f$ to $S_{2^{j+1}}f$ is stored in $W_{2^j}f$, called the "detail signal"[46]. The original signal $S_{2^0}f$ can be recovered from $S_{2^J}f$ by adding details recursively at each scale.

Now consider a sequence $\{s_n\}$ of finite energy, i.e.

$$\sum_{n=-\infty}^{+\infty} | s_n |^2 < \infty.$$

Mallat[47] has shown that if there exist two constants, $C_2 > C_1 > 0$, such that,

$$\forall \omega \in \boldsymbol{R}, \qquad C_1 \le \sum_{k=-\infty}^{+\infty} | \Phi(\omega + 2k\pi) |^2 \le C_2,$$

there exists a function $f(x) \in \boldsymbol{L}^2(\boldsymbol{R})$ such that

$$\forall n \in \boldsymbol{Z}, \qquad S_{2^0}f(n) = s_n \tag{2.19}$$

(2.19) means that a sequence $\{s_n\}$ can be viewed as uniform sampling of $S_{2^0}f(x)$, the approximation of some function $f(x)$ at scale $2^0 = 1$. To generalize this connection, the uniform sampling sequences of a finite-scale WT of $f(x)$,

$$\{ (W_{2^j}f(n + \upsilon))_{j=1,2,\ldots,J}, \quad S_{2^J}f(n + \tau) \} \tag{2.20}$$

are called a discrete wavelet transform (DWT) of the sequence $\{s_n\}$. Here υ and τ are sampling shifts depending on the particular scaling function and wavelet.

The computation of the DWT of a sequence relies on two discrete filters, *H* and *G*. They are related to the scaling function and wavelets by[48],

$$\Phi(\omega) = H(\frac{\omega}{2})\Phi(\frac{\omega}{2}) \tag{2.21}$$

$$\Psi(\omega) = G(\frac{\omega}{2})\Phi(\frac{\omega}{2}) \tag{2.22}$$

where $H(\omega)$ and $G(\omega)$ are 2π-periodic and satisfy,

$$H(0) = 1 \tag{2.23}$$

$$| H(\omega) |^2 + | H(\omega+\pi) |^2 \leq 1 \tag{2.24}$$

$$| H(\omega) |^2 + | G(\omega) |^2 = 1 \tag{2.25}$$

(2.21) and (2.22) indicate that $H(\omega)$ and $G(\omega)$ are two-scale transfer functions. They correspond to sequences, $\{h_n\}$ and $\{g_n\}$, which are impulse responses of two discrete filters.

A converse view of (2.21)-(2.25) is valuable to the study in Part II of the thesis. If a discrete filter *H* satisfies (2.23) and (2.24), one can construct another filter *G* by (2.25) and the associated scaling function and wavelets by iterating (2.21) and (2.22). In particular, if the equality holds in (2.24), *H* is associated with a wavelet orthonormal basis[46,47]. A discrete filter *H* satisfying (2.23) and (2.24) is called a *regular scaling filter*[65]. It is worth noting that (2.23) and (2.24) implies that $H(\pi)=0$, an alternative condition that is often used to define a regular scaling filter.

The filters *H* and *G* play a crucial role in the fast recursive computation of a DWT. The DWT of a sequence $\{s_n\}$ from scale $j=1$ to *J* can be computed in the Fourier domain by,

$$\begin{bmatrix} W_{2^1}F(\omega) \\ W_{2^2}F(\omega) \\ \vdots \\ W_{2^J}F(\omega) \\ S_{2^J}F(\omega) \end{bmatrix} = \begin{bmatrix} G(\omega)S_1F(\omega) \\ G(2\omega)S_{2^1}F(\omega) \\ \vdots \\ G(2^{J-1}\omega)S_{2^{J-1}}F(\omega) \\ H(2^{J-1}\omega)S_{2^{J-1}}F(\omega) \end{bmatrix} = \begin{bmatrix} G(\omega) \\ G(2\omega)H(\omega) \\ \vdots \\ G(2^{J-1}\omega)\cdot\prod_{p=0}^{J-2} H(2^p\omega) \\ \prod_{p=0}^{J-1} H(2^p\omega) \end{bmatrix} S_1F(\omega) \tag{2.26}$$

where $W_{2^j}F(\omega)$ is the Fourier transform of the DWT coefficients of $\{s_n\}$ at scale 2^j, and $S_{2^j}F(\omega) = H(2^{j-1}\omega)S_{2^{j-1}}F(\omega)$ is the Fourier transform of the approximated $\{s_n\}$ at scale 2^j. Here we have normalized scale by using (2.19).

The inverse DWT is given by,

$$S_1F(\omega) = \begin{bmatrix} \overline{G}(\omega) \\ \overline{G}(2\omega)\overline{H}(\omega) \\ \vdots \\ \overline{G}(2^{J-1}\omega)\cdot\prod_{p=0}^{J-2} \overline{H}(2^p\omega) \\ \prod_{p=0}^{J-1} \overline{H}(2^p\omega) \end{bmatrix}^t \begin{bmatrix} W_{2^1}F(\omega) \\ W_{2^2}F(\omega) \\ \vdots \\ W_{2^J}F(\omega) \\ S_{2^J}F(\omega) \end{bmatrix} \tag{2.27}$$

where $\overline{H}$ and $\overline{G}$ denote the conjugate of H and G, respectively. Replacing (2.27) in (2.26) we get the discrete wavelet reproducing equation:

$$\begin{bmatrix} W_{2^1}F(\omega) \\ W_{2^2}F(\omega) \\ \vdots \\ W_{2^J}F(\omega) \\ S_{2^J}F(\omega) \end{bmatrix} = \begin{bmatrix} K_{1,1}(\omega) & K_{1,2}(\omega) & \cdots & K_{1,J}(\omega) & K_{1,J+1}(\omega) \\ K_{2,1}(\omega) & K_{2,2}(\omega) & \cdots & K_{2,J}(\omega) & K_{2,J+1}(\omega) \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ K_{J,1}(\omega) & K_{J,2}(\omega) & \cdots & K_{J,J}(\omega) & K_{J,J+1}(\omega) \\ K_{J+1,1}(\omega) & K_{J+1,2}(\omega) & \cdots & K_{J+1,J}(\omega) & K_{J+1,J+1}(\omega) \end{bmatrix} \begin{bmatrix} W_{2^1}F(\omega) \\ W_{2^2}F(\omega) \\ \vdots \\ W_{2^J}F(\omega) \\ S_{2^J}F(\omega) \end{bmatrix} \tag{2.28}$$

where $K_{ij}(\omega)$ is the Fourier transform of the discrete WRK and given by,

$$K_{i,j}(\omega) = \Omega_{i-1}(G(2^{i-1}\omega)\cdot\Omega_{j-1}(\overline{G}(2^{j-1}\omega) \cdot\Omega_{i-2}(\prod_{p=0}^{i-2} H(2^p\omega)) \cdot \Omega_{j-2}(\prod_{p=0}^{j-2} \overline{H}(2^p\omega))$$

$$1 \le i, j \le J+1$$

The function Ω_q is defined as follows:

$$\Omega_q(x) = \begin{cases} x & 0 \le q \le J\text{-}1 \\ 1 & q < 0 \text{ or } q > J\text{-}1 \end{cases}$$

One reproducing kernel of particular interest is $K_{1,1}(\omega)$. It can be shown that

$$K_{1,1}(\omega) = G(\omega)\overline{G}(\omega) = |\, G(\omega)\,|^2 \qquad (2.29)$$

The discrete reproducing equation (2.28) has a few important distinctions from the continuous reproducing equation (2.4) and the infinite-scale dyadic reproducing equation (2.12). First, one can only "reproduce" at scale j=1 to J. Second, while the coarse-scale (low-frequency) signal energy is contained in $S_{2^J} f$ as indicated by (2.18), there are no energy contributions from scales smaller than 1.

The computational structure of the DWT is identical to an octave-band filter bank transform[65]. Given a pair of filters, H and G satisfying (2.25), one can build a perfect reconstruction filter bank by using H and G as analysis filters and $\overline{H}$ and $\overline{G}$ as synthesis filters. (2.26)-(2.28) will hold without reference to wavelet transforms. Indeed, a DWT differs from an octave-band filter-bank transform only in that the H in the DWT is a regular scaling filter satisfying (2.23) and (2.24) and is therefore associated with a scaling function and wavelets.

Figure 2.1 is a magnetic resonance (MR) image scan line and its DWT from scale 2^1 to 2^8. The particular wavelet ψ and the discrete filters *H* and *G* used for computing the DWT are described in Chapter 7.

Figure 2.1. An example of one-dimensional DWT. (a) An MR image scan line; (b) its DWT at 8 scales.

2.4 Two-Dimensional Wavelet Transforms

The one-dimensional (1-D) wavelet transforms defined in § 2.1-2.3 can be generalized to higher dimensions. The most general case has been studied by Lawton and Resnikoff[36]. In this section, we consider only two-dimensional (2-D) DWTs using separable 2-D wavelets which are constructed from 1-D wavelets. In practice, separable 2-D wavelets are most frequently used to reduce computational complexity.

Consider that we have two 1-D regular scaling filters, H_1 and H_2. Define $\{G_i(\omega), i=1,2\}$ by (2.25), and $\{L_i(\omega), i=1,2\}$ by

$$| L_i(\omega) |^2 = \frac{1 + | H_i(\omega) |^2}{2} \tag{2.30}$$

One can easily show that

$$H_1(\omega_1)H_2(\omega_2) + G_1(\omega_1)L_2(\omega_2) + L_1(\omega_1)G_2(\omega_2) = 1 \tag{2.31}$$

One class of the 2-D scaling function and wavelets can be constructed by $\{H_i, L_i, G_i, i = 1, 2\}$ in the Fourier domain as follows[48]:

$$\Phi(\omega_1, \omega_2) = \Phi_1(\omega_1)\Phi_2(\omega_2) = \prod_{k=1}^{+\infty} H_1(2^{-k}\omega_1)H_2(2^{-k}\omega_2) \tag{2.32}$$

$$\Psi^1(\omega_1, \omega_2) = \Psi_1(\omega_1)\Phi_2(\omega_2) = G_1(\frac{\omega_1}{2})L_2(\frac{\omega_2}{2})\Phi(\frac{\omega_1}{2}, \frac{\omega_2}{2}) \tag{2.33}$$

$$\Psi^2(\omega_1, \omega_2) = \Phi_1(\omega_1)\Psi_2(\omega_2) = L_1(\frac{\omega_1}{2})G_2(\frac{\omega_2}{2})\Phi(\frac{\omega_1}{2}, \frac{\omega_2}{2}) \tag{2.34}$$

where $\Phi(\omega_1, \omega_2)$ is the Fourier transform of the 2-D scaling function, $\phi(x,y) = \phi_1(x)\phi_2(y)$, and $\Psi^1(\omega_1, \omega_2)$ and $\Psi^2(\omega_1, \omega_2)$ are Fourier transforms of two 2-D wavelets, $\psi^1(x,y) = \psi_1(x)\xi_2(y)$ and $\psi^2(x,y) = \xi_1(x)\psi_2(y)$, respectively. Here the

Fourier plane is partitioned by two wavelets in perpendicular directions. In general, K partitions are possible as long as those directional wavelets satisfy[49]:

$$\sum_{k=1}^{K} \sum_{j=-\infty}^{+\infty} | \Psi^k(2^j\omega_1, 2^j\omega_2) |^2 = 1 \tag{2.35}$$

However, the computational burden becomes formidable with large K. We shall consider only $K = 2$ throughout this thesis.

Like (2.6) for 1-D wavelets, (2.35) is an over-restrictive condition for 2-D wavelets. It imposes $\{ \psi_{2^j}^k(x,y),\ k = 1,2,...,K \}_{j\in \mathbf{Z}}$ to constitute a tight frame. In [85], Zhong considered 2-D wavelets that constitute snug frames by relaxing (2.35).

Define the dyadic dilation of the 2-D wavelets by,

$$\psi_{2^j}^k(x, y) = 2^{-2j}\, \psi^k(2^{-j}x, 2^{-j}y), \quad k=1,2 \tag{2.36}$$

The 2-D WT of a function $f(x,y)\in \boldsymbol{L}^2(\boldsymbol{R}^2)$ at scale 2^j, at position (x,y), and in orientation k is given by,

$$W_{2^j}^k f(x, y) = f * \psi_{2^j}^k(x, y), \quad k=1,2 \tag{2.37}$$

With (2.35) being satisfied, by Parseval's theorem it can be shown that the 2-D WT is an isometry, i.e.

$$\| f \|^2 = \sum_{k=1}^{2} \sum_{j=-\infty}^{+\infty} \| W_{2^j}^k f \|^2.$$

the function $f(x,y)$ can be reconstructed by the 2-D inverse WT,

$$f(x, y) = \sum_{k=1}^{2} \sum_{j=-\infty}^{+\infty} W_{2^j}^k f * \tilde{\psi}_{2^j}^k(x, y) \tag{2.38}$$

Similar to the 1-D DWT, the computation of the 2-D DWT uses only the filter set $\{H_i, L_i, G_i, i = 1,2\}$. Given a 2-D sequence $\{s_{m,n}\} = S_1 f(m,n)$, the DWT can be computed recursively from scale 2^1 to 2^J in the Fourier domain[48]:

$$\begin{cases} W_{2^j}^1 F(\omega_1,\omega_2) = G_1(2^{j-1}\omega_1)L_2(2^{j-1}\omega_2)S_{2^{j-1}}F(\omega_1,\omega_2) \\ W_{2^j}^2 F(\omega_1,\omega_2) = L_1(2^{j-1}\omega_1)G_2(2^{j-1}\omega_2)S_{2^{j-1}}F(\omega_1,\omega_2) \\ S_{2^j} F(\omega_1,\omega_2) = H_1(2^{j-1}\omega_1)H_2(2^{j-1}\omega_2)S_{2^{j-1}}F(\omega_1,\omega_2) \end{cases} \qquad (2.39)$$
$$j = 1,2,...,J.$$

where $W_{2^j}^1 F(\omega_1,\omega_2)$ and $W_{2^j}^2 F(\omega_1,\omega_2)$ are the Fourier transform of the 2-D DWT of $\{s_{m,n}\}$ at scale 2^j, and $S_{2^j} F(\omega_1,\omega_2)$ is the Fourier transform of the approximated $\{s_{m,n}\}$ at scale 2^j. Here like in (2.19), we have normalized scale by defining $\{s_{m,n}\} = S_1 f(m,n)$ to be at the finest scale, 2^0=1. Since G_1 and G_2 are highpass filters, $W_{2^j}^1 F(\omega_1,\omega_2)$ and $W_{2^j}^2 F(\omega_1,\omega_2)$ provide highpassed version (details) of $S_{2^j} F(\omega_1,\omega_2)$ in ω_1 and ω_2 directions, respectively. An example of the 2-D DWT of an image at three scales is shown in *Figure 2.2*.

The 2-D inverse DWT is given recursively by,

$$\begin{aligned} S_{2^{j-1}} F(\omega_1,\omega_2) = {} & \overline{G}_1(2^{j-1}\omega_1)\overline{L}_2(2^{j-1}\omega_2)W_{2^j}^1 F(\omega_1,\omega_2) \\ & + \overline{L}_1(2^{j-1}\omega_1)\overline{G}_2(2^{j-1}\omega_2)W_{2^j}^2 F(\omega_1,\omega_2) \\ & + \overline{H}_1(2^{j-1}\omega_1)\overline{H}_2(2^{j-1}\omega_2)S_{2^j} F(\omega_1,\omega_2) \\ & j = J,\ J\text{-}1,\ J\text{-}2,\ ...,1. \end{aligned} \qquad (2.40)$$

where $\overline{H}_i$, $\overline{L}_i$, and $\overline{G}_i$ are the complex conjugates of H_i, L_i, and G_i, respectively. Replacing (2.40) in (2.39) results in the 2-D discrete reproducing equation at scale 2^j:

$$\begin{bmatrix} W_{2^j}^1 F(\omega_1,\omega_2) \\ W_{2^j}^2 F(\omega_1,\omega_2) \\ S_{2^j} F(\omega_1,\omega_2) \end{bmatrix} = \begin{bmatrix} K_{1,1}(\omega) & K_{1,2}(\omega) & K_{1,3}(\omega) \\ K_{2,1}(\omega) & K_{2,2}(\omega) & K_{2,3}(\omega) \\ K_{3,1}(\omega) & K_{3,2}(\omega) & K_{3,3}(\omega) \end{bmatrix} \begin{bmatrix} W_{2^j}^1 F(\omega_1,\omega_2) \\ W_{2^j}^2 F(\omega_1,\omega_2) \\ S_{2^j} F(\omega_1,\omega_2) \end{bmatrix} \qquad (2.41)$$

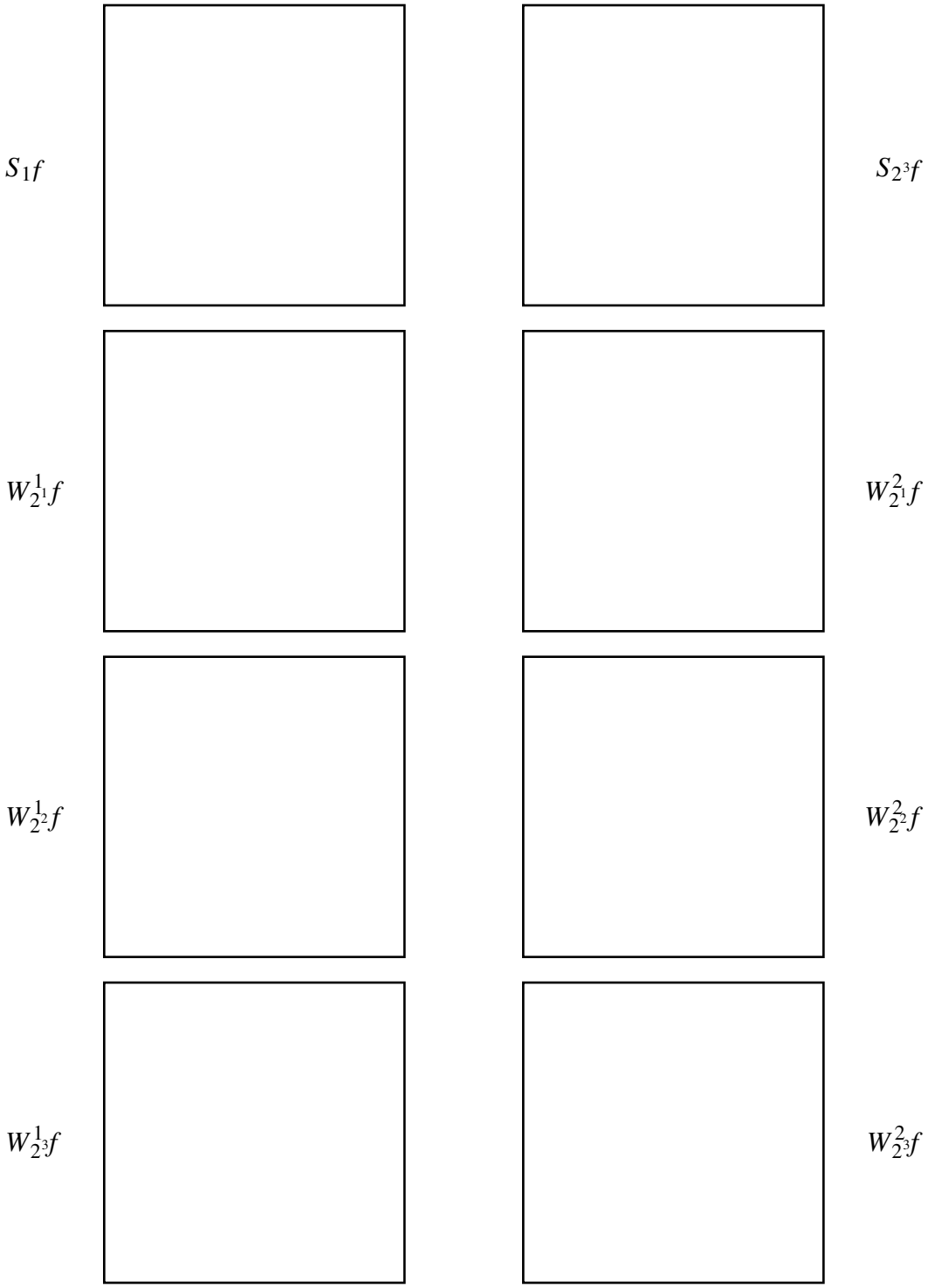
where $[K_{i,j}]$ is the Fourier WRK matrix; the subscripts in $[K_{i,j}]$ are matrix indices. $[K_{i,j}]$ is Hermitian and its upper triangular entries are:

$$K_{1,1}(\omega_1, \omega_2) = | G_1(2^{j-1}\omega_1) |^2 \ | L_2(2^{j-1}\omega_2) |^2 ;$$
$$K_{1,2}(\omega_1, \omega_2) = \overline{L}_1(2^{j-1}\omega_1)G_1(2^{j-1}\omega_1)\overline{G}_2(2^{j-1}\omega_2)L_2(2^{j-1}\omega_2) ;$$
$$K_{1,3}(\omega_1, \omega_2) = \overline{H}_1(2^{j-1}\omega_1)G_1(2^{j-1}\omega_1)\overline{H}_2(2^{j-1}\omega_2)L_2(2^{j-1}\omega_2) ;$$
$$K_{2,2}(\omega_1, \omega_2) = | L_1(2^{j-1}\omega_1) |^2 \ | G_2(2^{j-1}\omega_2) |^2 ;$$
$$K_{2,3}(\omega_1, \omega_2) = \overline{H}_1(2^{j-1}\omega_1)L_1(2^{j-1}\omega_1)\overline{H}_2(2^{j-1}\omega_2)G_2(2^{j-1}\omega_2) ;$$
$$K_{3,3}(\omega_1, \omega_2) = | \overline{H}_1(2^{j-1}\omega_1) |^2 \ | \overline{H}_2(2^{j-1}\omega_2) |^2 .$$

The 2-D discrete reproducing equation for the DWT from scale 2^1 to 2^J can be obtained recursively from (2.41). Starting from $j = 1$, we split $S_{2^j}F(\omega_1,\omega_2)$ by (2.39) recursively, and add rows representing the DWT at scale 2^{j+1} to the DWT vector in (2.41). We also add rows and columns to $[K_{i,j}]$ by using (2.41) and (2.40) and the property that $[K_{i,j}]$ is Hermitian.

For wavelets tuned in two orientations, the 2-D WRKs can be divided into four sequences, $\{K^{1,1}_{i,j}\}$, $\{K^{2,2}_{i,j}\}$, $\{K^{1,2}_{i,j}\}$, and $\{K^{2,1}_{i,j}\}$. Here the superscripts are for orientations, and subscripts for scales. The first two sequences of WRKs characterize intra-orientation correlations of 2-D DWTs, while the last two characterize inter-orientation correlations. Although the resulting reproducing equation has a complicated structure, it is completely determined by the filter set $\{H_i, L_i, G_i, i=1,2\}$.

Figure 2.2. (next page) An example of two-dimensional DWT. The top one in the left column is an MR image of head; the top right is the coarse approximation at scale 2^3 and the rest are DWT images at 3 scales.



(*Captions on previous page*)

Part II

SIGNAL RECOVERY AND WAVELET REPRODUCING KERNELS

Chapter 3

Wavelet-Regularized Signal Recovery in One Dimension

3.1 Review of Regularized Signal Recovery

Regularization is a mathematical technique for treating ill-posed problems which includes (1.2) as a special case. While the general framework of regularization was established and became well-known in the 1970s[60,72], the technique had been developed by Philips[63] ten years before under the name "constrained least-squares estimation". In early 1970s, Hunt[30] extended Philips' method to a successful application in image restoration—a typical 2-D signal recovery problem. In this section, we review the common structure of the least-squares and regularization techniques used in solving the signal recovery problem (1.2).

3.1.1 Constrained Least-Squares Estimation and Regularization

Nashed[59-61] summarized some intuitive ideas motivating the mathematical formulation of regularization methods. These ideas are:

1) change of the notion of a solution;
2) restriction on the data;
3) change of the space and/or topologies;
4) modification of the operator equation or the problem itself;
5) the concept of regularization operators;
6) well-posed stochastic extensions of ill-posed problems.

Now instead of solving equation (1.2), let us use some of Nashed's ideas; e.g. 1), 4), and 5). Consider a functional minimization subject to some constraints[72]:

$$\text{Min}\{\Gamma_\lambda = \| \underline{y} - \boldsymbol{H}\underline{x} \|^2 + \lambda \| \boldsymbol{C}\underline{x} \|^2 \}, \quad \text{with respect to } \underline{x}. \tag{3.1}$$

where $\boldsymbol{C}$ is a regularizing operator, and $\lambda>0$ is a parameter to be determined by matching specified constraints. (3.1) is a constrained least-squares (CLS) minimization[30] and its solution must satisfy the normal equation,

$$(\boldsymbol{H}^*\boldsymbol{H} + \lambda\, \boldsymbol{C}^*\boldsymbol{C})\, \underline{x} = \boldsymbol{H}^*\, \underline{y} \tag{3.2}$$

where $\boldsymbol{H}^*$ and $\boldsymbol{C}^*$ denote the Hermitian transpose of $\boldsymbol{H}$ and $\boldsymbol{C}$, respectively. With λ being determined and $|\, \boldsymbol{H}^*\boldsymbol{H} + \lambda\, \boldsymbol{C}^*\boldsymbol{C}\, | \neq 0$, the CLS minimization of (3.1) has a solution,

$$\underline{\hat{x}}_C = (\boldsymbol{H}^*\boldsymbol{H} + \lambda\, \boldsymbol{C}^*\boldsymbol{C})^{-1}\, \boldsymbol{H}^*\, \underline{y} \tag{3.3}$$

It is worth noting that when $\lambda\rightarrow 0$, (3.3) becomes the solution to an unconstrained least-squares (UCLS) minimization problem,

$$\text{Min}\{\Gamma = \| \underline{y} - \boldsymbol{H}\underline{x} \|^2\}, \quad \text{with respect to } \underline{x}. \tag{3.4}$$

Numerous constraints have been used along with (3.1) for various types of signal recovery problems. Two typical constraints are:

$$\| \boldsymbol{C}\underline{x} \|^2 \leq E^2 \tag{3.5}$$

and

$$\| \boldsymbol{H}\underline{x} - \underline{y} \| = \| \underline{n} \|^2 \leq \varepsilon^2 \tag{3.6}$$

In general, the regularizing operator $\boldsymbol{C}$ is chosen to be a highpass filter. (3.5) states that the high frequency energy in the signal is bounded. The second term in the cost functional of (3.1) prescribes a penalty for excessive high frequency energy in the solution. The role of λ is to control the level of penalty and hence the degree of smoothness of the solution. The first term in (3.1) is a measure of consistency of the solution to the observed data. Equation (3.6) states that the noise energy, or the consistency measure, is also

subject to a bound. Clearly, the two terms in the cost functional of (3.1) are incompatible in demand. There will be a tradeoff between consistency and smoothness of the solution. This tradeoff is affected by the choice of λ and $\boldsymbol{C}$.

3.1.2 Properties and Choice of Regularizing Parameter and Operator

When (3.5) and (3.6) have to be both satisfied, it can be shown[58] that

$$\lambda = (\varepsilon/E)^2 \tag{3.7}$$

This implies that, in the absence of explicit constraints, adjusting λ prescribes a ratio of noise energy to the high frequency energy in the signal. We have mentioned that when $\lambda \to 0$, (3.3) becomes a UCLS solution without any regularization. In view of (3.7), we have in effect assumed no noise. At the other extreme, when $\lambda \to \infty$, (3.3) tends to zero; we have assumed over-dominance of noise. The two extreme solutions are sometimes called "ultra-rough" and "ultra-smooth" solutions[34] in regularized signal recovery. A balancing point for smoothing and deblurring is $\lambda=1$.

While the parameter λ controls the degree of regularization (smoothing), the regularizing operator $\boldsymbol{C}$ defines the manner in which the solution is to be smoothed. The proper choice of the regularizing operator $\boldsymbol{C}$ has also been studied to some extent. Basically, two types of regularizing operators have been proposed. Those of the first type are dependent on the signal and noise. For example, a parametric Wiener filter has a regularizing operator based on the second-order statistics of the true signal and noise[5]; Reeves and Mersereau[64] proposed another operator based on the generalized cross-validation using only the observed signal. Regularizing operators of the second type are independent of specific signals and noise. For example, Tikhonov and Arsenin[72] proposed to use a p-th order differential operator. In practice, the discrete Laplacian is one of most widely used operators of this type[5,30].

In contrast to the two types of regularizing operators mentioned above, there have been few attempts to construct regularizing operators that depend on the degrading operator $\boldsymbol{H}$. In the following sections, we shall reformulate the signal recovery problem (1.2) as finding missing data at the finest scale of a DWT. The wavelet reproducing equation then plays a fundamental role in determining a unique and stable recovery. We shall show that the solution from the wavelet reproducing equation is equivalent to a least-squares solution, and the wavelet reproducing kernel defines a family of regularizing operators adapted to the degrading operator.

3.2 Regularized Signal Recovery with Wavelet Reproducing Kernels

3.2.1 Reformulation of the Problem

We now reformulate the signal recovery problem (1.2) within a wavelet framework. Assume the convolution kernel in (1.1) is bandlimited and properly discretized to a sequence $\{h_n\}$. Without loss of generality, we further assume that,

$$\Delta = \sum_{n=-\infty}^{+\infty} h_n = 1 \tag{3.8}$$

Indeed, if $\Delta \neq 1$, we can scale the signal by Δ since $\{h_n\}$ is a lowpass filter and has necessarily $\Delta \neq 0$. Equation (1.2) then becomes:

$$\underline{y} = \boldsymbol{H}' \, (\Delta \underline{x}) + \underline{n}$$

with $\Delta' = \sum h_n' = 1$. This means that we can scale $\{h_n\}$ by Δ^{-1} at the beginning. After we have obtained an estimate of $\widehat{\underline{x}}$, we rescale the result by Δ.

Let $H(\omega)$ be the Fourier transform of $\{h_n\}$; $H(\omega)$ is 2π-periodic. Since the continuous kernel is bandlimited, it is necessary that H satisfies (2.24) for a sufficiently sampled sequence $\{h_n\}$. Furthermore, $H(0) = \sum h_n = 1$. According to the definitions in §2.3, H is therefore a regular scaling filter associated with a scaling function and wavelets. One can

construct a highpass filter G by (2.25) and compute the DWT of a sequence by (2.26) and (2.27) with H and G.

Consider the original signal recovery problem (1.2) with $\underline{n}$=0. Define $\underline{x}$=$S_1\,\underline{x}$ to be at the finest scale and take the degrading operator $\boldsymbol{H}$ as the scaling filter. In view of (2.26), (1.2) can be interpreted as follows. The observation, $\underline{y} = \boldsymbol{H}\underline{x} = S_{2^1}\,\underline{x}$, is one part of the DWT of $\underline{x}$ for J=1 . The other part, $W_{2^1}\,\underline{x} = \boldsymbol{G}\underline{x}$, has been lost as a result of blurring. By (2.26), $\underline{y} = S_{2^1}\,\underline{x}$ can be further decomposed to some scale 2^J for J>1. The signal recovery problem can be then reformulated as:

$$\begin{aligned} &\text{Given:} \quad W_{2^2}\,\underline{x},\ W_{2^3}\,\underline{x},\ \ldots,\ W_{2^J}\,\underline{x},\ S_{2^J}\,\underline{x}, \\ &\text{Find:} \quad\ W_{2^1}\,\underline{x}. \end{aligned} \tag{3.9}$$

With $W_{2^1}\,\underline{x}$ being recovered as well as possible, an estimate of the original signal $\underline{x}$ can be synthesized by (2.27).

3.2.2 Signal Recovery with Stabilized Wavelet Reproducing Equation

We now study the solution to (3.9). Denote the DWT representation $\{W_{2^1}\underline{x},\ W_{2^2}\underline{x},\ \ldots,\ W_{2^J}\underline{x},\ S_{2^J}\underline{x}\ \}$ by $\{\underline{g}_1,\ \underline{g}_2,\ \ldots,\ \underline{g}_J,\ \underline{g}_{J+1}\}$ for simplicity. Rewrite the reproducing equation (2.28) in vector form with respect to the unknown $\underline{g}_1$,

$$\underline{g}_1 = \boldsymbol{K}_{1,1}\,\underline{g}_1 + \sum_{n=2}^{J+1} \boldsymbol{K}_{1,n}\,\underline{g}_n \tag{3.10}$$

where $\boldsymbol{K}_{1,n}$ is a Toeplitz matrix, $\boldsymbol{K}_{1,n}[i,j] = k_{1,n}(i-j)$.

(3.10) is the discrete version of a non-homogeneous Fredholm integral equation of the second kind with a self-adjoint operator[22]. The second term in (3.10) involves WRKs and the DWT of $\underline{x}$ from scale 2^2 to 2^J, which are completely determined by filters H and G, and the observed signal $\underline{y}$. The solution to this type of integral equation has been well

studied with the so-called Fredholm Alternative[22]. It states that (3.10) has a unique solution if and only if the corresponding homogeneous equation,

$$\underline{g}_1 = \boldsymbol{K}_{1,1}\, \underline{g}_1$$

has a non-trivial solution. This condition is equivalent to requiring that 1 not be an eigenvalue of $\boldsymbol{K}_{1,1}$. In practice, the lowpass nature of the degrading operator $\boldsymbol{H}$ often makes $\boldsymbol{K}_{1,1}$ have eigenvalues equal or close to 1. We turn to solving a stabilized equation,

$$\underline{g}_1 = \alpha\, \boldsymbol{K}_{1,1}\, \underline{g}_1 + \sum_{n=2}^{J+1} \boldsymbol{K}_{1,\boldsymbol{n}}\, \underline{g}_n \tag{3.11}$$

where α is a real number and is adjusted so that $\alpha\boldsymbol{K}_{1,1}$ does not have eigenvalues close to 1. It is possible to solve for $\underline{g}_1$ by iterating (3.11) with a tighter constraint on the choice of α. Define the L_2 norm of an operator $\boldsymbol{T}$ by

$$\| \boldsymbol{T} \| = \sup_{x \in L_2} \frac{\| \boldsymbol{T}x \|}{\| x \|} .$$

By the fixed point theorem[22], for $1 \le \| \boldsymbol{K}_{1,1} \| < \infty$, there exists a sufficiently small real number such that $|\alpha| \, \| \boldsymbol{K}_{1,1} \| < 1$ and (3.11) has a unique solution. In this case, the unique solution can be obtained by iterating (3.11) with an initial value.

3.2.3 The Equivalent Least-Squares Solutions

Recovering the original signal $\underline{x}$ from (3.11) or (3.9) has close relations to the least-squares and regularization techniques. We now show that the solution to (3.11) is equivalent to a CLS solution and that to (3.9) is equivalent to a UCLS solution.

Let us rearrange (3.11) to the form,

$$(\boldsymbol{I} - \alpha \boldsymbol{K}_{1,1})\, \underline{g}_1 = \sum_{n=2}^{J+1} \boldsymbol{K}_{1,\boldsymbol{n}}\, \underline{g}_n \tag{3.12}$$

Changing variable by α=1-λ, setting J=1, and using (2.25), (2.28), (2.29) and the fact that $\underline{g}_2=S_2{}^I\underline{x}=\underline{y}$, (3.12) can be rewritten,

$$(\boldsymbol{H}\boldsymbol{H}^* + \lambda \boldsymbol{G}\boldsymbol{G}^*)\underline{g}_1 = \boldsymbol{G}\boldsymbol{H}^*\underline{y} \tag{3.13}$$

If $\boldsymbol{G}\boldsymbol{H}^*=\boldsymbol{H}\boldsymbol{G}^*$, (3.13) is the normal equation for the CLS minimization,

$$\text{Min}\{\Gamma_\lambda = \| \boldsymbol{H}^*\underline{g}_1 - \boldsymbol{G}^*\underline{y} \|^2 + \lambda \| \boldsymbol{G}^*\underline{g}_1 \|^2 \}, \text{ with respect to } \underline{g}_1. \tag{3.14}$$

With λ properly adjusted and $| \boldsymbol{H}\boldsymbol{H}^* + \lambda\boldsymbol{G}\boldsymbol{G}^* | \neq 0$, the CLS solution for $\underline{g}_1$ is readily obtained from (3.14),

$$\hat{\underline{g}}_1 = (\boldsymbol{H}\boldsymbol{H}^* + \lambda \boldsymbol{G}\boldsymbol{G}^*)^{-1}\boldsymbol{G}\boldsymbol{H}^*\underline{y} \tag{3.15}$$

The solution for $\underline{x}$ is synthesized via the inverse DWT given by (2.27) for J=1. That is,

$$\begin{aligned} \hat{\underline{x}}_G &= \boldsymbol{G}^*\hat{\underline{g}}_1 + \boldsymbol{H}^*\underline{y} \\ &= \{ \boldsymbol{G}^*(\boldsymbol{H}\boldsymbol{H}^* + \lambda \boldsymbol{G}\boldsymbol{G}^*)^{-1}\boldsymbol{G} + \boldsymbol{I} \}\boldsymbol{H}^*\underline{y} \end{aligned} \tag{3.16}$$

When λ=0 or α=1, (3.13) becomes the normal equation of a UCLS minimization, which is equivalent to (3.9). Consequently, (3.15) becomes a UCLS solution for $\underline{g}_1$ and (3.16) gives the corresponding solution for $\underline{x}$.

Finally, we give the Fourier-domain form of (3.16) for easy implementation.

$$\hat{X}_G(\omega) = \left[\frac{| G(\omega) |^2}{| H(\omega) |^2 + \lambda | G(\omega) |^2} + 1 \right] \overline{H}(\omega)Y(\omega) \tag{3.17}$$

3.2.4 Remarks

1. The CLS minimization of (3.14) can be rewritten in terms of $\underline{x}$:

$$\text{Min}\{\Gamma_\lambda = \| \boldsymbol{H}^*\boldsymbol{G}\underline{x} - \boldsymbol{G}^*\underline{y} \|^2 + \lambda \| \boldsymbol{G}^*\boldsymbol{G}\underline{x} \|^2 \} \quad \text{with respect to } \underline{x}. \tag{3.18}$$

This indicates that $\boldsymbol{G}^*\boldsymbol{G}$ is a regularizing operator. From (2.29), the wavelet reproducing kernel $\boldsymbol{K}_{1,1} = \boldsymbol{G}\boldsymbol{G}^*$. If $\boldsymbol{G}$ is circulant, $\boldsymbol{G}^*\boldsymbol{G} = \boldsymbol{G}\boldsymbol{G}^*$; the wavelet reproducing kernel is exactly the regularizing operator in (3.18). In practice, a circulant $\boldsymbol{G}$ is often implied in the computation of a DWT by imposing periodicity on the finite data. An observation from the Fourier domain also shows that the wavelet reproducing kernel $K_{1,1}(\omega) = G(\omega)\overline{G}(\omega) = \overline{G}(\omega)G(\omega) = |\, G(\omega)\,|^2$ is the regularizing operator in (3.18). For simplicity, we shall call (3.16) a wavelet-regularized solution of $\underline{x}$. It is interesting to compare (3.14) or (3.18) with (3.1). Since $\boldsymbol{H}$ and $\boldsymbol{G}$ are a pair of lowpass and highpass filters, and $\underline{g}_1 = W_{2^1}\underline{x} = \boldsymbol{G}\underline{x}$, $\underline{y} = S_{2^1}\underline{x} = \boldsymbol{H}\underline{x}$, $\underline{g}_1$ and $\underline{y}$ can be viewed respectively as "high-band" and "low-band" components of $\underline{x}$. Then the first term in (3.14) is a "mid-band" consistency measure of solution $\underline{x}$ to the observation $\underline{y}$, while the first term in (3.1) represents a "low-band" consistency measure. Nevertheless, given (3.16), the "low-band" consistency measure of the wavelet-regularized solution can be easily computed. It is a function of λ,

$$\begin{aligned} \varepsilon_G(\lambda) &= \| \boldsymbol{H}\hat{\underline{x}}_G - \underline{y} \| \\ &\le \| \boldsymbol{H}\boldsymbol{G}^*(\boldsymbol{H}\boldsymbol{H}^* + \lambda\boldsymbol{G}\boldsymbol{G}^*)^{-1}\boldsymbol{G}\boldsymbol{H}^* + \boldsymbol{H}\boldsymbol{H}^* - \boldsymbol{I} \| \cdot \| \underline{y} \| \end{aligned} \tag{3.19}$$

which can be compared with the first term in (3.1) with $\hat{\underline{x}}_c$ given by (3.3),

$$\varepsilon_C(\lambda) \le \| \boldsymbol{H}(\boldsymbol{H}^*\boldsymbol{H} + \lambda\boldsymbol{C}^*\boldsymbol{C})^{-1}\boldsymbol{H}^* - \boldsymbol{I} \| \cdot \| \underline{y} \| \tag{3.20}$$

2. The role of λ in (3.14) is similar to that in (3.1). When $\lambda \to 0$, both (3.14) and (3.1) reduce to unconstrained least-squares minimization problems; no smoothing is done and the "low-band" consistency functionals ε_G and ε_C are minimized. When $\lambda \to 1$, smoothing and deblurring are balanced because two contributing terms in both (3.14) and (3.1) have equal weights. When $\lambda \to \infty$, smoothing is dominant and that forces $\hat{\underline{g}}_1 \to 0$ in (3.15) and $\hat{\underline{x}}_c \to 0$ in (3.3). However, the final solution of $\underline{x}$ given by (3.16) does not vanish for $\lambda \to \infty$;

$$\lim_{\lambda \to \infty} \hat{\underline{x}}_G = H^* \underline{y} \tag{3.21}$$

(3.21) gives a solution that is as smooth as the blurred original signal, $\boldsymbol{H}\underline{x}$. This reflects a conceptual difference of the wavelet approach from an "inverse" approach. In the reformulated signal recovery problem (3.9), we do not seek an inversion of the blurring process. Instead, we look for the information lost in the blurring process and resynthesize it with blurred signal by (2.27). Therefore, the worst situation in terms of deblurring happens when $\lambda \to \infty$ and the solution is given by (3.21); we gain no more bandwidth than blurred original signal. On the other hand, we lose no more bandwidth. We simply give up deblurring and smooth the noise to the same degree of smoothness as the blurred original signal. One must see the two sides of the wavelet-regularized signal recovery as revealed by (3.21). On one side this scheme prevents over-regularized signal recovery; on the other side it does not turn to a pure smoothing filter as λ becomes large. Thus wavelet regularization is perhaps not as useful as standard regularization if one's task is simply to smooth noise.

3. The condition $\boldsymbol{GH}^* = \boldsymbol{HG}^*$ is satisfied with $\boldsymbol{H}$ and $\boldsymbol{G}$ being circulant and self-adjoint. For example, circulant $\boldsymbol{H}$ and $\boldsymbol{G}$ are implied in the computation of DWT by imposing periodicity on the finite data, and many degrading kernels result in real symmetric $\boldsymbol{H}$ and $\boldsymbol{G}$. Nevertheless, when $\boldsymbol{GH}^* \neq \boldsymbol{HG}^*$, (3.13) is not a normal equation of (3.14). However, the interpretation in the frequency domain still applies because the spectral moduli of $\boldsymbol{GH}^*$ and $\boldsymbol{HG}^*$ are identical.

4. If the continuous degrading kernel in (1.1) is not bandlimited, the discretized kernel may not satisfy (2.24). In such a situation, (2.26) and (2.27) are not a DWT pair in the regular sense. However, as we mentioned in §2.3, (2.26) and (2.27) can be regarded as an octave-band filter bank transform pair and the reproducing equation (2.28) still holds. Therefore, all results in this chapter are valid regardless of H being regular or irregular.

5. Recently, Bruneau *et al*[15] took a synthetic approach to regularized image restoration with biorthogonal wavelets. In a discrete biorthogonal wavelet transform, two pairs of discrete filters, $\{H_1, G_1\}$ and $\{H_2, G_2\}$, are used for decomposition and reconstruction. These two pairs of filters correspond to two dual wavelet bases[18] and satisfy,

$$H_1(\omega)H_2(\omega) + G_1(\omega)G_2(\omega) = 1 \tag{3.22}$$

The authors in [15] took H_1 to be the discretized degrading operator, H, and constructed G_1, H_2 and G_2 by (3.22) and biorthogonal relations between these filters. They further chose G_1 as a regularizing operator for the CLS minimization in (3.1). Their formulation can be rewritten in accordance with the symbols used in (3.14) as follows:

$$\text{Min}\{\Gamma_\lambda = \| \boldsymbol{H_1 G_2}\, \underline{g}_1 - \boldsymbol{G_1 G_2}\, \underline{y} \|^2 + \lambda \| \underline{g}_1 \|^2 \} \quad \text{with respect to } \underline{g}_1. \tag{3.23}$$

(3.23) is similar to (3.14) in that the first term in (3.23) is also a "mid-band" consistency measure and the second term is a measure of high frequency energy of the solution $\underline{x}$. However, unlike (3.14), (3.23) does not relate to recovering $\underline{g}_1$ from the wavelet reproducing equation. The solution to (3.23) is not equivalent to that of (3.11).

3.3 Wavelet-Regularized Signal Recovery—An Example

In this section, we present a synthetic example of signal recovery with a wavelet reproducing kernel. The discretized degrading kernel is given below on the interval $[-\pi, \pi]$ in the Fourier domain,

$$\begin{aligned} H(\omega) &= \sum_{k=-\infty}^{\infty} \frac{\sin\,(2(\omega - k\pi))}{2(\omega - k\pi)} \\ &= \frac{1}{4} + \frac{1}{2}\cos\,(\omega) + \frac{1}{4}\cos\,(2\omega) \end{aligned} \tag{3.24}$$

This kernel models a special case of one-dimensional motion blur[31]. The continuous kernel corresponding to (3.24) is an unit-area rectangular window centered at 0 in time domain and has a length of 4; it is discretized at unit sampling rate. Clearly, *H* satisfies (2.23) and (2.24) and is therefore a regular scaling filter. The associated scaling function and wavelet can be constructed by (2.21) and (2.22). To compute a DWT, the associated highpass filter *G* is constructed from (2.25).

Figure 3.1 is a scan line from an MR image of prostate. It is synthetically blurred by *H*; the blurred one is shown in *Figure 3.2(a)*. *Figure 3.2(b)* shows the recovered scan line by using (3.17) and the minimum mean square error (MMSE) criterion. Next, Gaussian white noise is added to the blurred scan line in *Figure 3.2(a)*; the blurred signal-to-noise ratio (BSNR) is controlled at 30 dB. The resulting blurred noisy scan line is shown in *Figure 3.3(a)*, and the wavelet-regularized recovery in *Figure 3.3(b)*. Numerical results including MMSEs and optimal regularizing parameters for *Figures 3.2* and *3.3* are listed in *Table 3.1*.

Table 3.1. Experimental results from recovering blurred "prostate" scan lines (original "prostate" scan line: mean = 316.684, variance = 26395.3)

Signal	MMSE	λ
Fig.3.1(a): 1-D motion blur, BSNR=∞	998.058	—
Fig. 3.1(b): Wavelet-regularized recovery	383.205	0.014
Fig.3.2(a): 1-D motion blur, BSNR=30 dB	1023.33	—
Fig. 3.2(b): Wavelet-regularized recovery	500.785	0.030

Figure 3.1. A scan line from an MR image of prostate.

(a)

(b)

Figure 3.2. Recovery of the blurred noise-free "prostate" scan line. *(a)* motion-blurred scan line; *(b)* wavelet-regularized recovery. (See *Table 3.1* for numerical results).

(a)

(b)

Figure 3.3. Recovery of the blurred noisy "prostate" scan line. *(a)* motion-blurred scan line with noise, BSNR=30 dB; *(b)* wavelet-regularized recovery. (See *Table 3.1* for numerical results).

Chapter 4

Wavelet-Regularized Signal Recovery in Two Dimensions

4.1 Extensions of One-Dimensional Results

In the last Chapter we have been concerned with solving the 1-D signal recovery problem (1.2) in the wavelet framework. In general, a 2-D signal recovery problem is hard to treat in the same way. Nevertheless, if the 2-D degrading operator is considered separable in the two dimensions, the method and results in the last Chapter have straightforward 2-D extensions. We now present these results in the Fourier domain. We shall omit many mathematical derivations that are repetitive in structure to their 1-D counterparts.

4.1.1 The Two-Dimensional Signal Recovery Problem

The 2-D signal recovery problem analogous to (1.2) can be stated in the Fourier domain as follows,

$$Y(\omega_1,\omega_2) = H(\omega_1,\omega_2)X(\omega_1,\omega_2) + N(\omega_1,\omega_2) \tag{4.1}$$

where H is the Fourier transform of the 2-D degrading kernel, $h(m,n)$, and N, Y, and X are the Fourier transform of the noise, observed and original signals, respectively. We shall only consider separable $h(m,n)$, i.e.

$$h(m,n) = h_1(m)h_2(n) \quad \text{and} \quad H(\omega_1,\omega_2) = H_1(\omega_1)H_2(\omega_2) \tag{4.2}$$

If both H_1 and H_2 satisfy (2.23) and (2.24), i.e., they are both regular, one can construct a separable 2-D scaling function and wavelets from H_1 and H_2 by using (2.32)-(2.34). A DWT of a 2-D sequence can be computed by (2.39) and (2.40) with a

filter set $\{H_i, L_i, G_i, i=1,2\}$, where $\{L_i\}$ and $\{G_i\}$ are defined by (2.30) and (2.25), respectively.

Consider a noise-free case of (4.1). In view of (2.39), the observed signal Y is one part of the 2-D DWT of X for J=1. The other two parts, $W_{2^1}^1 F$ and $W_{2^1}^2 F$, have been lost as a result of degradation. Here as discussed in §2.4, we have normalized scale by defining $X(\omega_1,\omega_2)=S_1F(\omega_1,\omega_2)$ to be at the finest scale 2^0=1. The 2-D signal recovery problem (4.1) can be reformulated in the way similar to (3.9):

$$\begin{aligned} &\text{Given:} && \{W_{2^j}^1 F(\omega_1,\omega_2),\ W_{2^j}^2 F(\omega_1,\omega_2),\ j = 2,...,J\},\ S_{2^J}F(\omega_1,\omega_2) \\ &\text{Find:} && W_{2^1}^1 F(\omega_1,\omega_2) \text{ and } W_{2^1}^2 F(\omega_1,\omega_2) \end{aligned} \tag{4.3}$$

4.1.2 The Solutions

Again, we resort to the wavelet reproducing equation. It suffices to consider J=1 in (2.41). In this case, the first line of (4.3) reduces to $S_{2^1}F(\omega_1,\omega_2)$. The first two equations in (2.41) are of our interest. They can be made independent of each other by elimination of one unknown,

$$W_{2^1}^1 F(\omega_1,\omega_2) = \\ |P(\omega_1,\omega_2)|^2\, W_{2^1}^1 F(\omega_1,\omega_2) + \overline{H}_1(\omega_1)G_1(\omega_1)\overline{H}_2(\omega_2)L_2(\omega_2)S_{2^1}F(\omega_1,\omega_2) \tag{4.4}$$

$$W_{2^1}^2 F(\omega_1,\omega_2) = \\ |P(\omega_1,\omega_2)|^2\, W_{2^1}^2 F(\omega_1,\omega_2) + \overline{H}_1(\omega_1)L_1(\omega_1)\overline{H}_2(\omega_2)G_2(\omega_2)S_{2^1}F(\omega_1,\omega_2) \tag{4.5}$$

where

$$|P(\omega_1,\omega_2)|^2 = |G_1(\omega_1)|^2\,|L_2(\omega_2)|^2 + |L_1(\omega_1)|^2\,|G_2(\omega_2)|^2 \tag{4.6}$$

The existence and uniqueness of the solutions to (4.4) and (4.5) can be analyzed in the same way as in §3.2.2. In particular, because the operator in (4.4) and (4.5) is finite, i.e.,

$$\int_{-\pi}^{\pi} | P(\omega_1,\omega_2) |^4 \, d\omega_1 d\omega_1 < \infty,$$

the fixed-point theorem[22] states that there exists a sufficiently small real number $|\alpha|>0$, such that the two stabilized equations,

$$W_{2^1}^{1} F(\omega_1,\omega_2) = \alpha| P(\omega_1,\omega_2) |^2 \, W_{2^1}^{1} F(\omega_1,\omega_2) + \overline{H}_1(\omega_1)G_1(\omega_1)\overline{H}_2(\omega_2)L_2(\omega_2)S_{2^1} F(\omega_1,\omega_2) \tag{4.7}$$

$$W_{2^1}^{2} F(\omega_1,\omega_2) = \alpha| P(\omega_1,\omega_2) |^2 \, W_{2^1}^{2} F(\omega_1,\omega_2) + \overline{H}_1(\omega_1)L_1(\omega_1)\overline{H}_2(\omega_2)G_2(\omega_2)S_{2^1} F(\omega_1,\omega_2) \tag{4.8}$$

have unique solutions. In such a case, the two solutions can be obtained by iterating (4.7) and (4.8) with two initial values.

By an approach similar to that in §3.2.3, and using (2.31), we can show that (4.7) and (4.8) are the normal equations of two CLS minimization problems. Define:

$$|Q_{\lambda}^{1}(\omega_1,\omega_2)|^2 = | \overline{H}_1(\omega_1)\overline{H}_2(\omega_2)W_{2^1}^{1} F(\omega_1, \omega_2) - \overline{G}_1(\omega_1)\overline{L}_2(\omega_2)S_{2^1} F(\omega_1, \omega_2) |^2 + \lambda |\overline{P}(\omega_1, \omega_2)W_{2^1}^{1} F(\omega_1, \omega_2)|^2 \tag{4.9}$$

and

$$|Q_{\lambda}^{2}(\omega_1,\omega_2)|^2 = | \overline{H}_1(\omega_1)\overline{H}_2(\omega_2)W_{2^1}^{2} F(\omega_1, \omega_2) - \overline{L}_1(\omega_1)\overline{G}_2(\omega_2)S_{2^1} F(\omega_1, \omega_2) |^2 + \lambda |\overline{P}(\omega_1, \omega_2)W_{2^1}^{2} F(\omega_1, \omega_2)|^2 \tag{4.10}$$

where λ=1-α. The two CLS minimization problems corresponding to (4.7) and (4.8) are:

$$\text{Min } \{\Gamma_\lambda^1 = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |Q_\lambda^1(\omega_1, \omega_2)|^2 d\omega_1 d\omega_2 \}, \text{ with respect to } W_{2^1}^1 F(\omega_1,\omega_2) \tag{4.11}$$

and

$$\text{Min } \{\Gamma_\lambda^2 = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |Q_\lambda^2(\omega_1, \omega_2)|^2 d\omega_1 d\omega_2 \}, \text{ with respect to } W_{2^1}^2 F(\omega_1,\omega_2) \tag{4.12}$$

We have obtained the CLS estimates for $W_{2^1}^1 F(\omega_1,\omega_2)$ and $W_{2^1}^2 F(\omega_1,\omega_2)$:

$$\widehat{W_{2^1}^1 F}(\omega_1,\omega_2) = \frac{\overline{H}_1(\omega_1)G_1(\omega_1)\overline{H}_2(\omega_2)L_2(\omega_2)}{|H_1(\omega_1)|^2 |H_2(\omega_2)|^2 + \lambda |P(\omega_1,\omega_2)|^2} S_{2^1} F(\omega_1,\omega_2) \tag{4.13}$$

$$\widehat{W_{2^1}^2 F}(\omega_1,\omega_2) = \frac{\overline{H}_1(\omega_1)L_1(\omega_1)\overline{H}_2(\omega_2)G_2(\omega_2)}{|H_1(\omega_1)|^2 |H_2(\omega_2)|^2 + \lambda |P(\omega_1,\omega_2)|^2} S_{2^1} F(\omega_1,\omega_2) \tag{4.14}$$

where $S_{2^1} F(\omega_1,\omega_2) = Y(\omega_1,\omega_2)$. The wavelet-regularized solution for the original signal is synthesized from (4.13) and (4.14) via the 2-D inverse DWT given by (2.40). Finally, we have,

$$\widehat{X}(\omega_1,\omega_2) = \left[\frac{|P(\omega_1,\omega_2)|^2}{|H_1(\omega_1)|^2 |H_2(\omega_2)|^2 + \lambda |P(\omega_1,\omega_2)|^2} + 1 \right] \overline{H}_1(\omega_1)\overline{H}_2(\omega_2)Y(\omega_1,\omega_2) \tag{4.15}$$

4.1.3 Remarks

1. The two CLS minimization problems (4.11) and (4.12) can be combined into one with respect to $X(\omega_1,\omega_2)$. Computing the DWT by (2.39) for J=1, we have,

$$W_{2^1}^{1} F(\omega_1, \omega_2) = G_1(\omega_1)L_2(\omega_2)X(\omega_1, \omega_2) \tag{4.16}$$

$$W_{2^1}^{2} F(\omega_1, \omega_2) = L_1(\omega_1)G_2(\omega_2)X(\omega_1, \omega_2) \tag{4.17}$$

Replacing (4.16) and (4.17) in (4.9) and (4.10), and combining the latter two by using (4.6), we have,

$$\begin{aligned} |Q_\lambda(\omega_1,\omega_2)|^2 = {} & | \overline{H}_1(\omega_1)\overline{H}_2(\omega_2)\overline{P}(\omega_1,\omega_2)X(\omega_1,\omega_2) - P(\omega_1,\omega_2)Y(\omega_1,\omega_2) |^2 \\ & + \lambda | \overline{P}(\omega_1,\omega_2)P(\omega_1,\omega_2)X(\omega_1,\omega_2) |^2 \end{aligned} \tag{4.18}$$

and the combined CLS minimization problem,

$$\text{Min } \{ \Gamma_\lambda = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |Q_\lambda(\omega_1, \omega_2)|^2 d\omega_1 d\omega_2 \}, \text{ with respect to } X(\omega_1,\omega_2). \tag{4.19}$$

Clearly, the reproducing kernel $|P(\omega_1,\omega_2)|^2$ in (4.4) and (4.5) is exactly the regularizing operator in the CLS minimization problem (4.19).

2. The 2-D solution (4.15) is seen to have the same form as the 1-D solution (3.17). Indeed, they share the same asymptotic properties as were discussed in §3.2.4.

4.2 Wavelet-Regularized Image Restoration—Examples

We now present experimental results of the wavelet-regularized image restoration with two examples. We synthetically blurred two clinical MR images (*Figures 4.1* and *4.2*) with two frequently encountered degrading kernels—Fejér and Gaussian. Noise was added to the blurred images and the BSNR was controlled at 25 dB. The blurred noise-free and noisy images were restored by using (4.15) and the MMSE criterion. For comparison, we also did mean-square-error (MSE) restoration by using the 2-D version of (3.3) with a discrete 2-D Laplacian as the regularizing operator.

4.2.1 Example I

The discrete 2-D degrading kernel is separable in the spatial and Fourier domains:

$$h1(m,n) = u(m)u(n) \quad \Leftrightarrow \quad H1(\omega_1,\omega_2) = U(\omega_1)U(\omega_2) \tag{4.20}$$

with $u(k)$ and its Fourier transform $U(\omega)$ on $[-\pi,\pi]$ given by,

$$u(k) = \frac{2sin^2(\omega_c k/2)}{\pi\,\omega_c\,k^2} \quad \Leftrightarrow \quad U(\omega) = \begin{cases} 1 - |\frac{\omega}{\omega_c}| & |\omega| < \omega_c \\ 0 & \omega_c \le |\omega| \le \pi \end{cases} \tag{4.21}$$

where ω_c is the cut-off frequency and $U(\omega)$ is 2π-periodic. $u(k)$ is the discrete Fejér kernel and has a triangular spectrum. It is the discretized point-spread function (PSF) of a diffraction-limited incoherent imaging system with a rectangular aperture[17]. Clearly, $U(\omega)$ satisfies (2.23) and (2.24), and is therefore associated with a scaling function and wavelets.

The MR head image in *Figure 4.1* is synthetically blurred by *H1* for $\omega_c = 0.5\pi$; the blurred noise-free image is shown in *Figure 4.3(a)*. *Figures 4.3(b)* and *4.3(c)* are restored images that attain MMSEs with the wavelet and Laplacian regularizing operators, respectively. Next, the blurred image in *Figure 4.3(a)* is added with Gaussian white noise; the BSNR is 25 dB and the resulting image is shown in *Figure 4.4(a)*. The restored images are shown in *Figures 4.4(b)* and *4.4(c)* . Numerical results including the MMSE and the optimal regularizing parameter λ for restored images in *Figures 4.3* and *4.4* are listed in *Table 4.1*.

Table 4.1. Experimental results from restoring blurred "head" images (original "head" image: mean = 125.48, variance = 17583.65)

Blurred/Deblurred Images	MMSE	λ
Fig. 4.3(a): blurred by Fejér kernel, ω_c=0.5π, BSNR=∞	2229.09	—
Fig. 4.3(b): wavelet-regularized restoration	333.077	3.0E-5
Fig. 4.3(c): Laplacian-regularized restoration	333.513	5.0E-5
Fig.4.4(a): blurred by Fejér kernel, ω_c=0.5π, BSNR=25 dB	2753.02	—
Fig. 4.4(b): Wavelet-regularized restoration	1178.09	0.010
Fig. 4.4(c): Laplacian-regularized restoration	1165.33	0.050

4.2.2 Example II

The discrete 2-D degrading kernel is again separable in the spatial and Fourier domains:

$$h2(m,n) = v(m)v(n) \quad \Leftrightarrow \quad H2(\omega_1,\omega_2) = V(\omega_1)V(\omega_2) \tag{4.22}$$

with $v(k)$ and its Fourier transform $V(\omega)$ given by,

$$v(k) = \frac{1}{\sigma\sqrt{2\pi}} exp(-\frac{k^2}{2\sigma^2}) \quad \Leftrightarrow \quad V(\omega) = \sum_{j=-\infty}^{+\infty} exp[-\frac{\sigma^2}{2}(\omega + 2j\pi)^2] \tag{4.23}$$

where σ is a positive integer. $v(k)$ is the discrete Gaussian kernel and $V(\omega)$ on $[-\pi,\pi]$ is very close to a Gaussian for large σ. *H2* is a model often used to approximate a wide range of degradation such as blurring caused by atmospheric turbulence[31]. Unlike the $U(\omega)$ in the last example, $V(\omega)$ does not satisfy (2.24) because a continuous Gaussian is not

bandlimited and $V(\pi)\neq 0$. There are no scaling functions or wavelets in the regular sense to be associated with $V(\omega)$. Nevertheless, our computational framework will not collapse. The difference is in terminology: the appropriate interpretation will now be in terms of octave-band filter banks, instead of wavelets. However, we would still use the term "wavelet" for simplicity.

The MR knee image in *Figure 4.2* is synthetically blurred by *H2* for $\sigma=2$. The resulting noise-free image is shown in *Figure 4.5(a)* , and the restored images are in *Figures 4.5(b)* and *4.5(c)*. Next, Gaussian white noise is added and the BSNR is controlled at 25 dB. The resulting blurred noisy image and the restorations are shown in *Figures 4.6(a-c)* . Numerical results including the MMSE and the optimal regularizing parameter λ for restored images in *Figures 4.5* and *4.6* are listed in *Table 4.2*.

Table 4.2. Experimental results from restoring blurred "knee" images (original "knee" image: mean = 191.00, variance = 26005.44)

Blurred/Deblurred Images	MMSE	λ
Fig.4.5(a): blurred by Gaussian kernel, $\sigma=2$, BSNR=∞	914.933	—
Fig. 4.5(b): Wavelet-regularized restoration	214.147	0.0001
Fig. 4.5(c): Laplacian-regularized restoration	214.913	0.0001
Fig.4.6(a): blurred by Gaussian kernel, $\sigma=2$, BSNR=25 dB	1368.98	—
Fig. 4.6(b): Wavelet-regularized restoration	892.559	0.015
Fig. 4.6(c): Laplacian-regularized restoration	887.148	0.19

4.2.3 Comments on Experimental Results

Theoretical analyses in §3.2.4 suggest that the wavelet regularizing operator has two advantages: it adapts to the degrading operator and avoids over-smoothing. Our experimental results in *Tables 4.1 and 4.2* indicate that the wavelet regularizing operator is indeed very effective. However, wavelet-regularized solutions do not seem to outmatch Laplacian-regularized ones in terms of MMSEs. In fact, the two methods attain similar MMSEs at different λ.

In noise-free restorations, the role of regularization is to prevent amplification of roundoff errors. A small λ is sufficient. Graphical results show that restored images are very close to the originals. This observation is supplemented by the numerical results showing that MSEs were reduced by 75-85% of those before restorations.

In a noisy environment, regularization plays a key role in suppressing noise and ringing artifacts. Consequently, our deblurring capability is greatly compromised. In both wavelet- and Laplacian-regularized restorations at a SNR=25 dB, the MSE reduction is limited between 35-60% of that before restorations, depending on the degree of blur. The apparent difference between wavelet- and Laplacian-regularizations seems to be that the wavelet-regularization attains the MMSE at a smaller λ.

Perhaps this is the best we can do with a deterministic regularizing operator, i.e., the first type of regularizing operators that we reviewed in §3.1.2. One has reasons to expect that the second kind of regularizing operators may do better because they employ information about specific signals and noise. However, the level of noise is still going to determine the quality of deblurring. A good example is Wiener filtering. The next part of this thesis is devoted exclusively to the noise reduction problem (1.3). We will still use wavelets, but by a quite different approach.

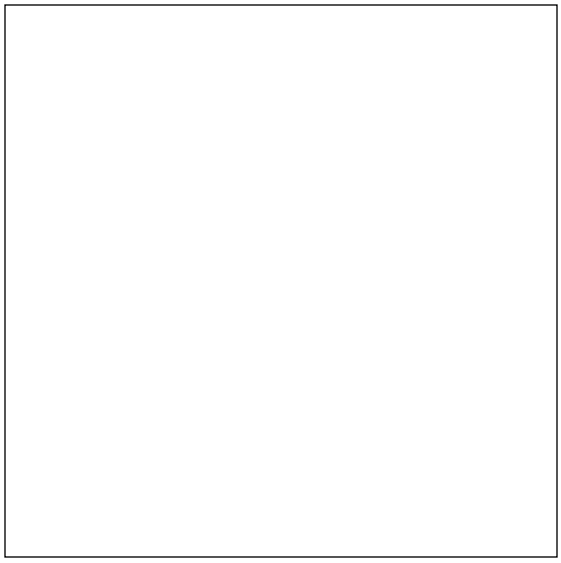


Figure 4.1. An MR head image.

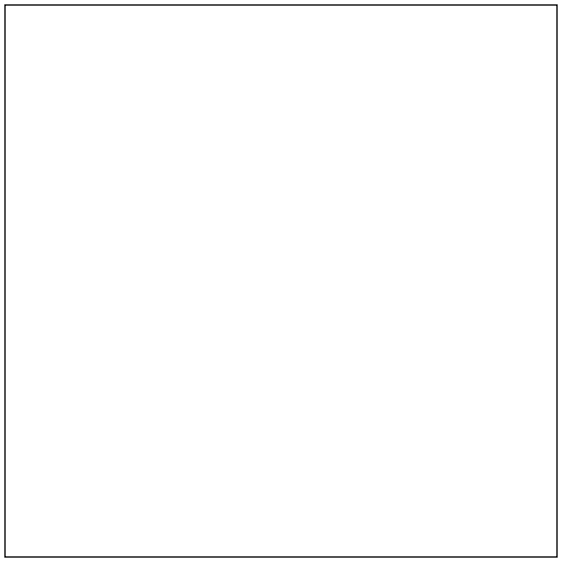
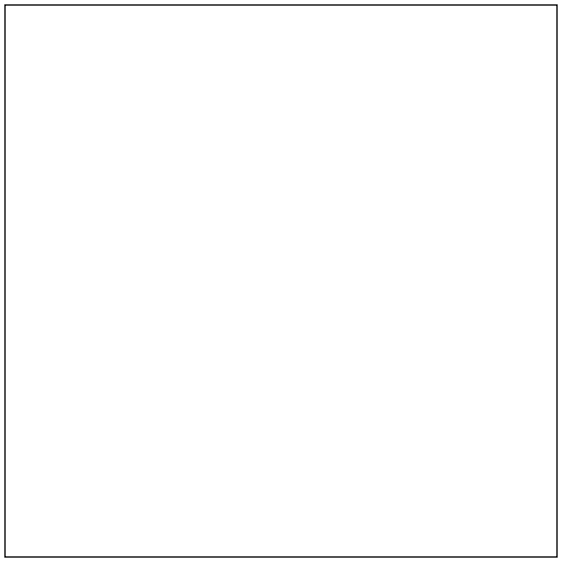
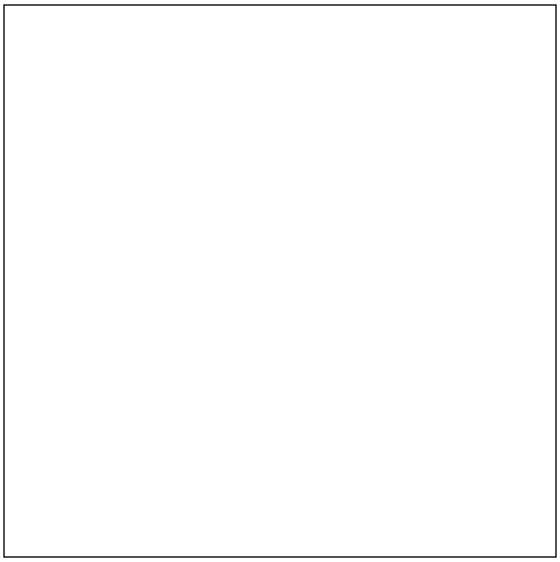
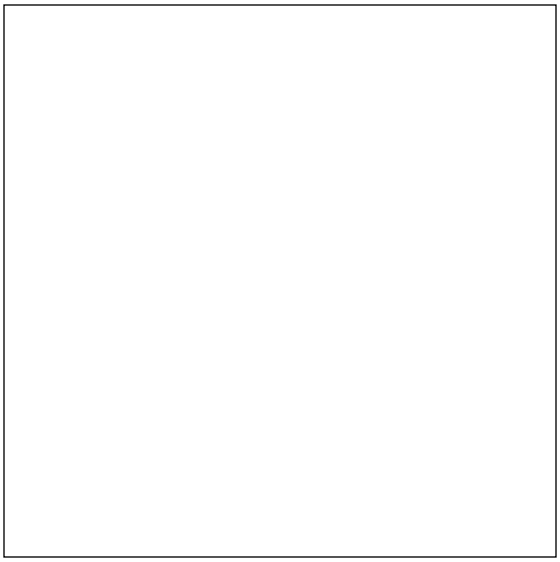


Figure 4.2. An MR knee image.



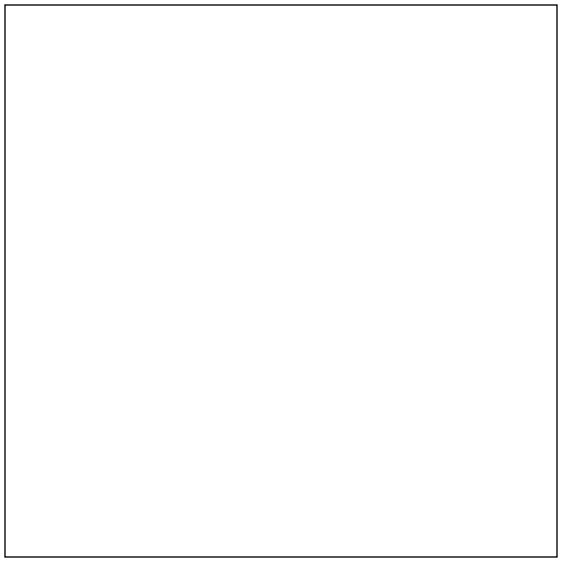
(a)



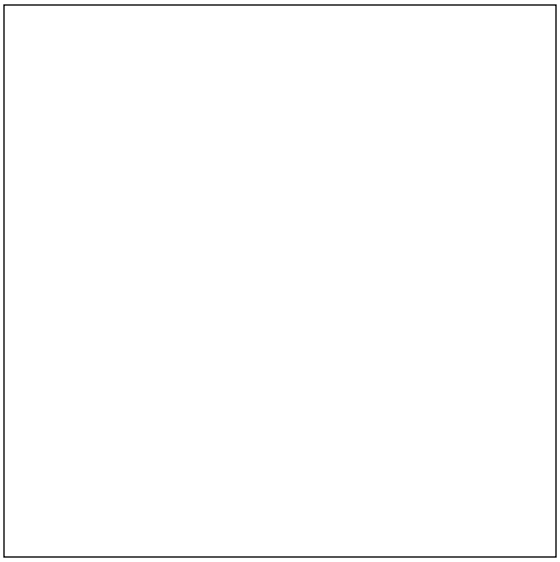


(b) *(c)*

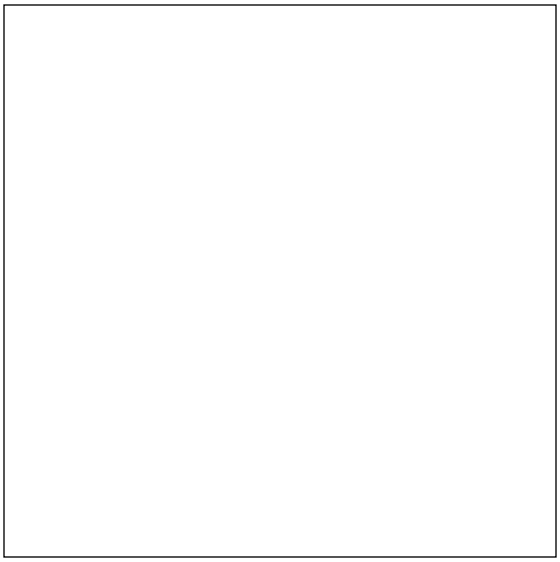
Figure 4.3. Restoration of the noise-free "head" image blurred by a Fejér kernel. *(a)* blurred by the Fejér kernel, ω_c=0.5π; *(b)* restored with wavelet regularization; *(c)* restored with Laplacian regularization. (See *Table 4.1* for numerical results).



(a)

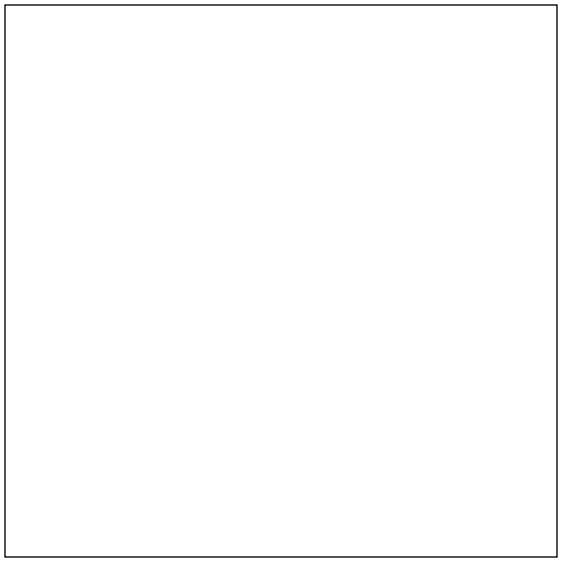


(b)

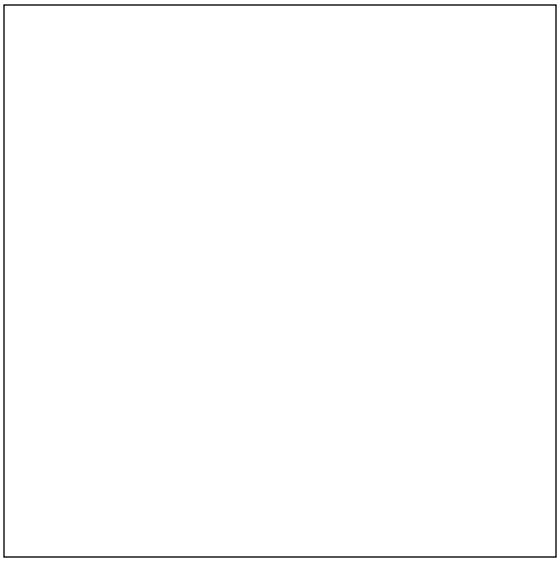


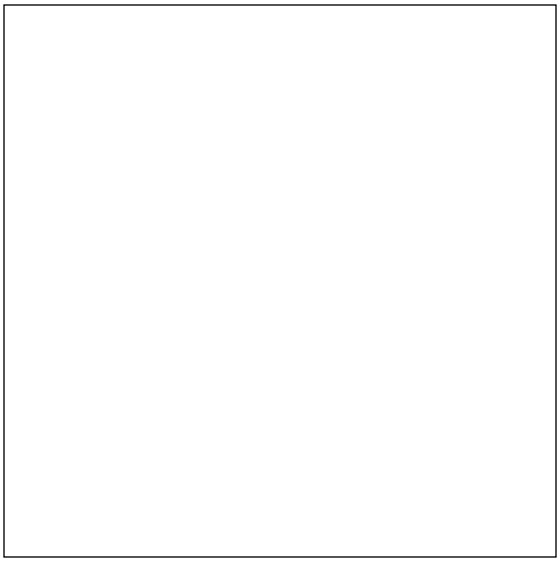
(c)

Figure 4.4. Restoration of the noisy "head" image blurred by a Fejér kernel. *(a)* blurred by the Fejér kernel, ω_c=0.5π, BSNR=25 dB; *(b)* restored with wavelet regularization; *(c)* restored with Laplacian regularization. (See *Table 4.1* for numerical results).



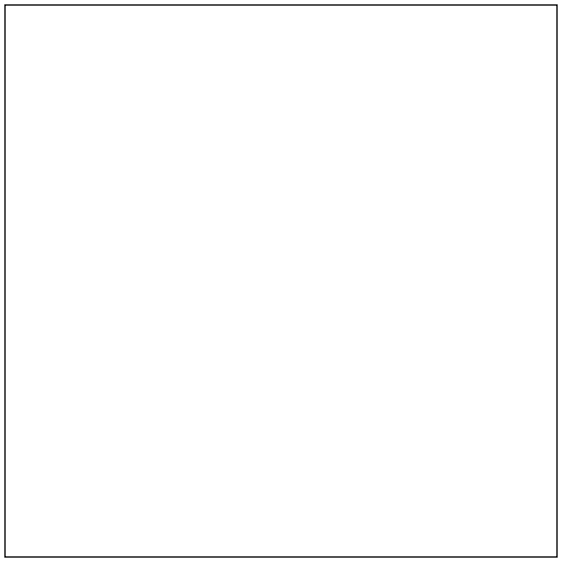
(a)



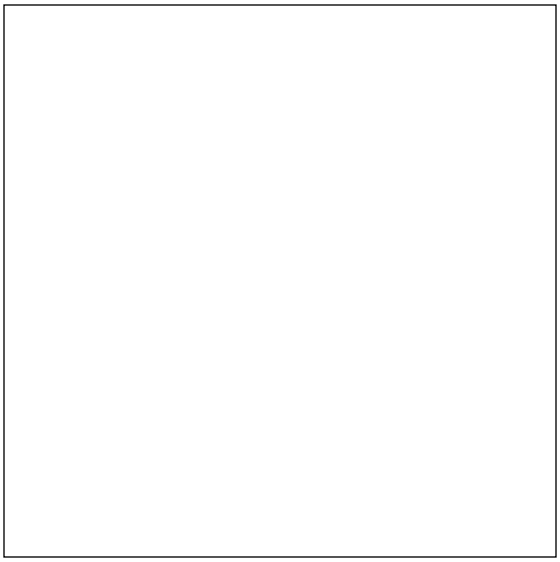


(b) *(c)*

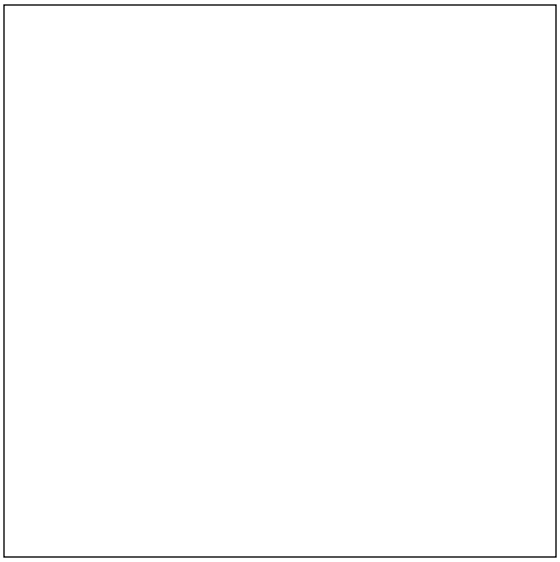
Figure 4.5. Restoration of the noise-free "knee" image blurred by a Gaussian kernel. *(a)* blurred by the Gaussian kernel, σ=2; *(b)* restored with wavelet regularization; *(c)* restored with Laplacian regularization. (See *Table 4.2* for numerical results).



(a)



(b)



(c)

Figure 4.6. Restoration of the noisy "knee" image blurred by a Gaussian kernel. *(a)* blurred by the Gaussian kernel, σ=2, BSNR=25 dB; *(b)* restored with wavelet regularization; *(c)* restored with Laplacian regularization. (See *Table 4.1* for numerical results).

Part III

NOISE REDUCTION WITH THE WAVELET MAXIMA REPRESENTATION

Chapter 5
The Wavelet Maxima Representation

In the development of wavelet theories, orthogonal wavelet transforms (OWTs) have received more attention than others. Many applications, such as image compression[23] and numerical analysis[10], use OWTs. However, OWTs have a serious weakness: shift-variance. This property is particularly undesirable to signal analysis applications. This issue was first raised by Mallat[49] and Strang[70]. In the 1992 MIT Wavelet Seminar Series[4], Simoncelli and his colleagues gave a talk on "What's Wrong with Orthogonal Wavelet Bases for Signal/Image Analysis?" They concluded that in order to achieve shiftability, they had to abandon the elegance of orthogonality. This resulted in different design philosophies[69] for wavelets and associated filters.

Nevertheless, the appeal of the OWT lies not only in its mathematical elegance, but in its compactness as well. For example, the latter property is essential for transform coding. One may ask: is there a compact multiscale representation that preserves shiftability? The multiscale edge representations developed by Mallat provide a positive answer to this question. In this Chapter, we shall review a particular class of multiscale edge representations—the wavelet maxima representation—developed by Mallat and Zhong[48,50]. For the purpose of this thesis, the review will emphasize the choice of wavelets and reconstruction algorithms, but extensive references are cited in case additional details are desired.

5.1 The One-Dimensional Wavelet Maxima Representation

5.1.1 Wavelet Maxima as Multiscale Edges

The zero-crossings and local extrema of any WT, as shown by Mallat[49,52], are shift-invariant. This means that a translation in a signal will cause the zero-crossings and local

extrema in the WT of that signal to be translated by the same amount. Now what do zero-crossings and local extrema tell us about the signal? For an arbitrary wavelet, there may not be a clear answer to this question, but for a particular class of wavelets, the interpretation is interesting and insightful.

Consider a wavelet constructed in such a way that it is the *p*-th order derivative of a smoothing function. In particular, if the wavelet is the first order derivative of a smoothing function, the local extrema of the WT by such a wavelet characterize sharp variations in a signal at multiple scales. Equivalently, if the wavelet is the second order derivative of the smoothing function, the zero-crossings of the WT convey the same information. Such observations can be made with mathematical rigor. Let $\psi_1(x)$ and $\psi_2(x)$ be, respectively, the first and second order derivatives of $\phi(x)$, i.e.,

$$\psi_1(x) = \frac{d\phi(x)}{dx}, \tag{5.1}$$

and

$$\psi_2(x) = \frac{d^2\phi(x)}{dx^2} \tag{5.2}$$

where $\phi(x)$ is a smoothing function. The dyadic WTs of a function $f(x)$ with these two wavelets are given by (2.7):

$$W^1_{2^j} f(x) = f * (2^j \frac{d\phi_{2^j}}{dx})(x), \tag{5.3}$$

and

$$W^2_{2^j} f(x) = f * (2^{2j} \frac{d^2\phi_{2^j}}{dx^2})(x) \tag{5.4}$$

Because differentiation and convolution commute, (5.3) and (5.4) can be rewritten as,

$$W^1_{2^j} f(x) = 2^j \frac{d}{dx}(f * \phi_{2^j})(x) = (2^j \frac{df}{dx}) * \phi_{2^j}(x) \tag{5.5}$$

and

$$W_{2^j}^2 f(x) = 2^j \frac{d^2}{dx^2}(f * \phi_{2^j})(x) = (2^j \frac{d^2 f}{dx^2}) * \phi_{2^j}(x) \tag{5.6}$$

(5.5) and (5.6) can be explained as differentiation of the signal followed by smoothing at multiple scales. This is the design philosophy of many edge detectors in computational vision. For example, the Canny edge detector[16] is the first order derivative of a Gaussian, while the Marr-Hildreth edge detector[54] is the Laplacian of a Gaussian. In both cases, the size of the Gaussian plays the role of a scale parameter.

Figure 5.1. The local extrema of $|W_{2^j}^1 f(x)|$ and zero-crossings of $|W_{2^j}^2 f(x)|$. x_0 and x_2 are local maximum points and x_1 is a local minimum of $|W_{2^j}^1 f(x)|$. They are all zero-crossing points of $|W_{2^j}^2 f(x)|$. However, x_0 and x_2 correspond to sharp variation points in the signal; while x_1 corresponds to an inflection point.

Mallat and Zhong[48,50] further pointed out that some local minima and zero-crossings of the WTs given by (5.5) and (5.6) do not correspond to sharp variation points in the signal; they are inflection points of $f*\phi_{2^j}(x)$. This situation is illustrated by *Figure 5.1* with a simulated signal. On the other hand, the local maxima of $|W^1_{2^j} f(x)|$ are sharp variation points in the signal observed at multiple scales. The local maxima of $|W^1_{2^j} f(x)|$ are called modulus maxima of the WT and interpreted as multiscale edges[48,50]. For simplicity, we shall call them "wavelet maxima" in this thesis. We also use the term "multiscale edges" interchangeably for "wavelet maxima".

5.1.2 The One-Dimensional Analyzing Wavelet

It has been clear from the last section that in order to make wavelet maxima be multiscale edges, the wavelet must be the first order derivative of a smoothing function. In fact, we could further require that the smoothing function from which the wavelet is constructed is simply the scaling function. In such a case, the scaling function and wavelet, as we discussed in §2.3, depend on two discrete filters H and G.

Mallat and Zhong[50,85] designed a family of wavelets that have desired properties for a multiscale edge representation. The filters H and G are given as follows[85]:

$$H(\omega) = \begin{cases} \cos^m(\frac{\omega}{2}) & \text{if } m \text{ is even} \\ \cos^m(\frac{\omega}{2})\ \exp(\frac{i\omega}{2}) & \text{if } m \text{ is odd.} \end{cases} \tag{5.7}$$

for $m = 1,\ 2,\ 3,\ ...,$ and

$$G(\omega) = i\ \sin(\frac{\omega}{2})\ \exp(\frac{i\omega}{2}) \tag{5.8}$$

The scaling function and wavelet are constructed from (2.21) and (2.22). They are:

$$\Phi(\omega) = \left[\frac{\sin(\omega/2)}{\omega/2}\right]^{m+1} \tag{5.9}$$

and

$$\Psi(\omega) = i(\omega/2)\left[\frac{\sin(\omega/2)}{\omega/2}\right]^{m+1} \tag{5.10}$$

(5.10) defines a family of piecewise spline wavelets. The larger the *m*, the smoother the wavelet at the expense of broader support. Clearly, $\psi(x)$ is the first order derivative of $\phi(x)/2$, which is desired for a multiscale edge representation. In this thesis we use the cubic spline wavelet that corresponds to m=4. It is plotted in *Figure 5.2* along with the scaling function. *Figure 5.3* shows a signal and its WT for J=5 with this cubic spline wavelet . The signal is composed of three boxcars with two small ones riding on top of a bigger one and all embedded in random noise. It is very clear that those sharp variation points in the signal are characterized by the wavelet maxima.

Figure 5.2. A cubic spline scaling function and wavelet.

We must point out that the filter *G* here is not constructed from (2.25). Instead, *G* is related to *H* by,

$$|H(\omega)|^2 + G(\omega)D(\omega) = 1 \tag{5.11}$$

where

$$D(\omega) = \frac{1 - |H(\omega)|^2}{G(\omega)} \tag{5.12}$$

(5.11) corresponds to general wavelet frames defined by (2.10). In such a case, the analyzing wavelet is still constructed from H and G by (2.22), but the synthesis wavelet is constructed from $\overline{H}$ and D, where G and D do not have conjugate symmetry. Accordingly, one must replace $\overline{G}$ by D in the inverse DWT formula (2.27) and all subsequent expressions. An advantage of using (5.12) and (5.13) instead of (2.25) is that one has some flexibility in choosing G, and hence the analyzing wavelet.

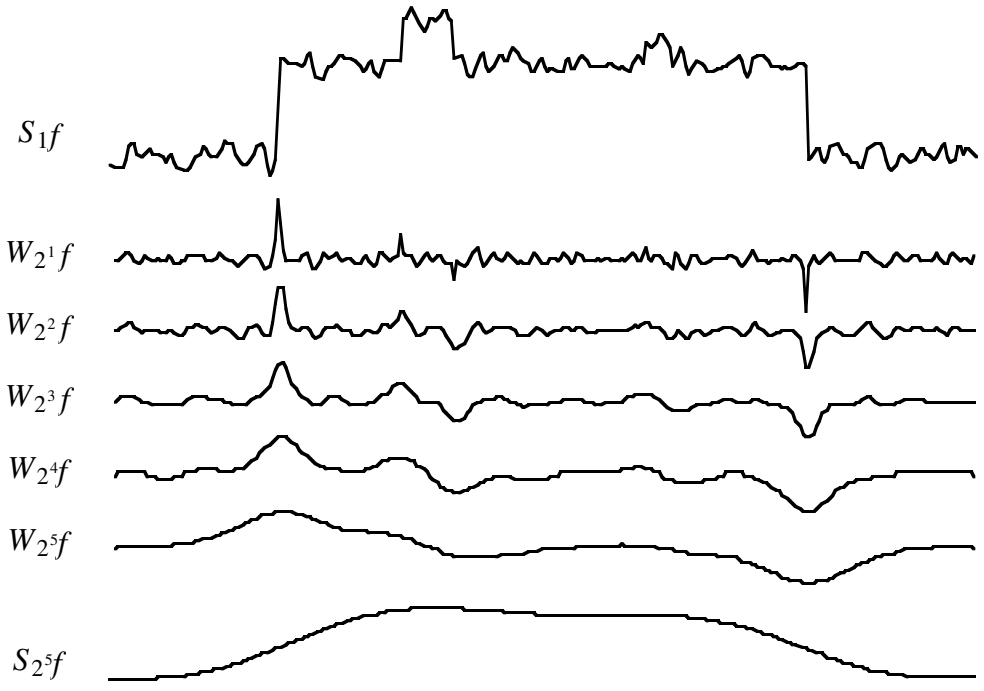


Figure 5.3. A test signal and its wavelet transform with a cubic spline wavelet.

5.1.3 Representation of 1-D Signals by Wavelet Maxima

With the analyzing wavelet defined in §5.1.2, we can build a wavelet maxima representation of a signal. Following Zhong[85], define:

$$A_{2^j} = \{ \ (x_i, W_{2^j} f(x_i)) : |W_{2^j} f(x)| \text{ has local maximum at } x = x_i \ \} \tag{5.13}$$

to be the wavelet maxima set, and $\boldsymbol{M}$ a local maxima detection operator such that,

$$\boldsymbol{M}[W_{2^j} f(x)]_{j \in \mathbf{Z}} = [A_{2^j}]_{j \in \mathbf{Z}} \tag{5.14}$$

For a J-level WT, the collection,

$$\{ S_{2^J} f(x), \ [A_{2^j}]_{1 \le j \le J} \} \tag{5.15}$$

is called the wavelet maxima representation of a signal. In words, the J-level wavelet maxima representation of a signal contains positions and values of $W_{2^j} f(x)$ at all levels where $|W_{2^j} f(x)|$ attains a local maximum, plus the smoothed signal at the coarsest scale. The coarse signal is included to ensure the completeness of a finite-scale representation. *Figure 5.4* is the wavelet maxima representation for the signal in *Figure 5.3*.

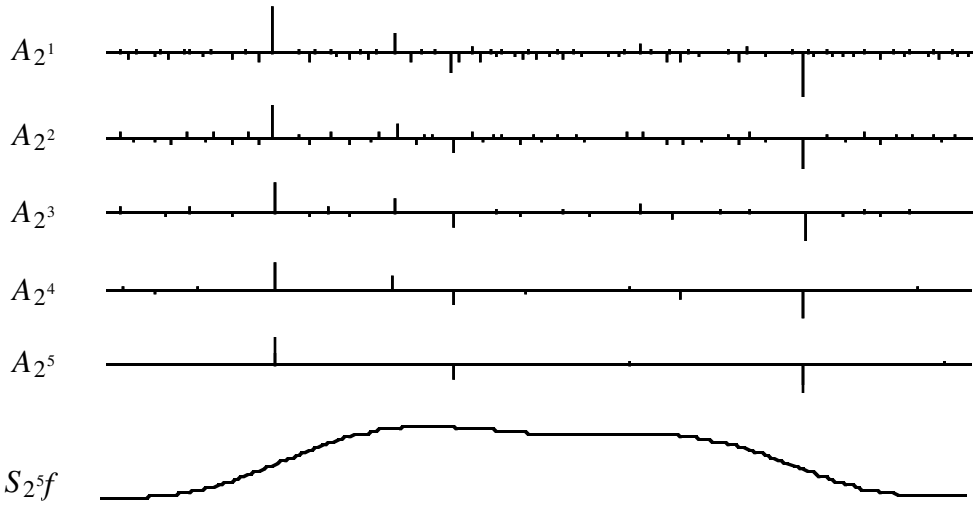


Figure 5.4. The wavelet maxima representation of a signal. The original signal is in *Figure 5.3*.

5.1.4 Reconstruction of 1-D Signals from Wavelet Maxima

That a signal can be completely characterized by its multiscale edges was a conjecture of the late David Marr[55,56]. We know that a signal can be reconstructed from its WT by the inverse WT. The challenging question is how to reconstruct the WT from wavelet maxima or multiscale edges. There are really three issues of concern in solving such a problem: the existence of a solution, the uniqueness of the solution, and the stability of algorithms for achieving the solution.

From experimental studies with a large class of signals, Mallat and Zhong[48] first suggested positive answers to all of the above questions. Later, the uniqueness issue was studied extensively, and counter-examples were given by Meyer[57] and Berman[12,13]. These results show that the uniqueness of a wavelet maxima representation is not guaranteed in general. However, for most signals in the real world, a wavelet maxima representation is precise and stable, and thus suitable for practical signal and image processing applications[50]. In our own experience, reconstruction of signals from their wavelet maxima by the algorithm developed by Mallat and Zhong is very stable, and reconstructed signals are very close to the originals. We give below the framework of Mallat and Zhong's algorithm. This algorithm will be used in the thesis for reconstructing signals from their wavelet maxima.

In reconstructing the WT from wavelet maxima, one looks for a sequence of functions $\{g_j(x)\}_{j\in\mathbf{Z}}$ satisfying the following two constraints:

1) $\{g_j(x)\}_{j\in\mathbf{Z}}$ has the same local maxima set as the given one, $[A_{2^j}]_{j\in\mathbf{Z}}$, i.e.,

$$\boldsymbol{M}[g_j(x)]_{j\in\mathbf{Z}} = [A_{2^j}]_{j\in\mathbf{Z}} \tag{5.16}$$

2) $\{g_j(x)\}_{j\in\mathbf{Z}}$ is the WT of some signal, that is, $\{g_j(x)\}_{j\in\mathbf{Z}}$ satisfies the reproducing equation (2.12). We rewrite it below for convenience:

$$\forall\, j \in \mathbf{Z}, \quad g_j = \sum_{i \in \mathbf{Z}} g_i * K_{i,j} \tag{5.17}$$

These two constraints define two subsets of functions in $\mathbf{L}^2$. Denote by Γ the first subset containing all sequences of functions satisfying (5.16), and $\mathbf{V}$ the second subset containing all sequences of functions satisfying (5.17). Clearly, a legitimate solution of reconstruction must lie in the intersection of Γ and $\mathbf{V}$, denoted by $\Gamma \cap \mathbf{V}$.

To compute an element in $\Gamma \cap \mathbf{V}$, Mallat and Zhong[48] used an alternating projection algorithm originally developed by Youla and Webb[82] for computing the intersection of convex sets. Define $\mathbf{P}_\Gamma$ to be an orthogonal projection operator on Γ, and $\mathbf{P}_\mathbf{V}$ an orthogonal projection operator on $\mathbf{V}$. A reconstruction of the WT is obtained by using $\mathbf{P}_\Gamma$ and $\mathbf{P}_\mathbf{V}$ alternately with an initial guess $[g_j(x)]^0_{j \in \mathbf{Z}}$. Although here Γ is not a convex set, there have been no reports on divergence in alternating projections onto Γ and $\mathbf{V}$.

The implementation of $\mathbf{P}_\mathbf{V}$ is quite straightforward by noticing $\mathbf{P}_\mathbf{V} = (\mathbf{W})(\mathbf{W}^{-1})$. It is simply an inverse WT followed by a WT. For $\mathbf{P}_\Gamma$, Mallat and Zhong detailed an efficient implementation in [48].

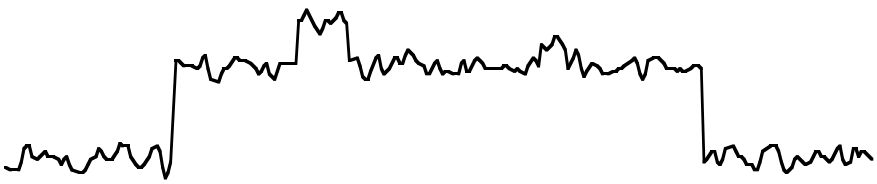


Figure 5.5. A signal reconstructed from its wavelet maxima. The original signal is in *Figure 5.3*.

Figure 5.5 is a reconstructed signal from the wavelet maxima in *Figure 5.4* by the algorithm described above. We computed 10 iterations of alternating projections. The initial guess for the reconstructed sequences of functions is a cubic spline interpolated function from the wavelet maxima set in *Figure 5.4*. Comparing the reconstructed signal in

Figure 5.5 with the original in *Figure 5.3*, one can hardly see the difference. Indeed, the relative MSE of the reconstructed signal to the original is only 0.5%.

5.2 The Two-Dimensional Wavelet Maxima Representation

The 1-D wavelet maxima representation introduced in §5.2 can be generalized to representing 2-D signals. In this section, we consider such generalizations with a special class of 2-D signals—images.

5.2.1 Representation of Images by Wavelet Maxima

As in §2.4, we consider separable 2-D wavelets in two orientations. In such a case, the two wavelets can be constructed from a set of discrete filters by using (2.33) and (2.34). For simplicity, we use $\phi_1=\phi_2$ and $\psi_1=\psi_2$ in (2.33) and (2.34). In order to build a multiscale edge representation of images, we further require that the two wavelets are the partial derivatives of the 2-D scaling function, that is,

$$\psi^1(x,y) = \frac{\partial}{\partial x}\phi(x,y), \text{ and } \psi^2(x,y) = \frac{\partial}{\partial y}\phi(x,y).$$

Then the 2-D WT defined by (2.37) becomes the gradient of a function smoothed by the scaling function $\phi(x,y)$ at multiple scales; we shall call it "multiscale gradient":

$$\nabla_{2^j} f(x,y) = \left\langle W^1_{2^j} f(x,y),\ W^2_{2^j} f(x,y) \right\rangle = \frac{1}{2^{2j}} (\nabla f) * \phi(x,y) \tag{5.18}$$

A multiscale gradient can also be characterized by the magnitude and angle. They are:

$$\rho_{2^j} f(x,\ y) = \sqrt{(W^1_{2^j} f(x,\ y))^2 + (W^2_{2^j} f(x,\ y))^2} \tag{5.19}$$

and

$$\theta_{2^j} f(x,\ y) = \tan^{-1}\left[\frac{W^2_{2^j} f(x,\ y)}{W^1_{2^j} f(x,\ y)}\right] \tag{5.20}$$

The Cartesian components of a multiscale gradient can be recovered from the magnitude-angle representation by,

$$W_{2^j}^1 f(x, y) = \rho_{2^j} f(x, y) \cos(\theta_{2^j} f(x, y)) \tag{5.21}$$

$$W_{2^j}^2 f(x, y) = \rho_{2^j} f(x, y) \sin(\theta_{2^j} f(x, y)) \tag{5.22}$$

Consider that the 2-D function $f(x,y)$ is an image. A multiscale edge point is defined as a maximum point of $\rho_{2^j} f$ along the direction of $\theta_{2^j} f$. Similar to the 1-D notation, define the gradient maxima set of scale j by,

$$\begin{aligned} A_{2^j} = \{ & [x_i, y_i, \rho_{2^j} f(x_i, y_i), \theta_{2^j} f(x_i, y_i)]: \rho_{2^j} f(x_i, y_i) \text{ has local maximum} \\ & \text{at } (x_i, y_i) \text{ along the direction of } \theta_{2^j} f(x_i, y_i) \}. \end{aligned} \tag{5.23}$$

For a J-level 2-D WT, the collection,

$$\{ S_{2^J} f(x, y), [A_{2^j}]_{1 \le j \le J} \} \tag{5.24}$$

is called the wavelet maxima representation of an image. *Figure 5.6* is a 3-scale wavelet maxima representation of the MR head image in *Figure 2.2*. The multiscale gradient magnitudes and angles are computed from Cartesian components by (5.19) and (5.20).

5.2.2 Reconstruction of Images from Wavelet Maxima

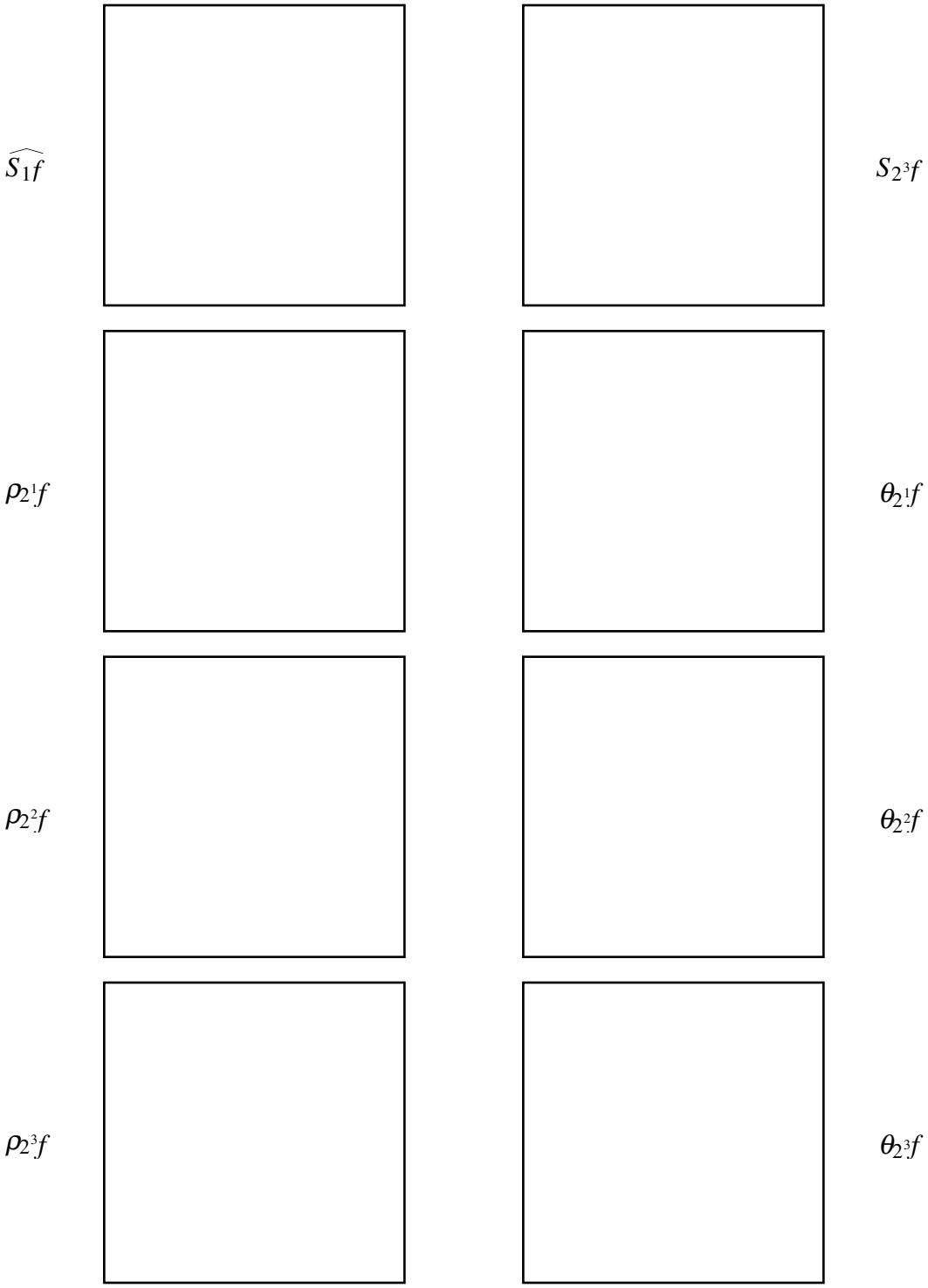
Similar to the 1-D reconstruction problem, we need to reconstruct the 2-D WT from wavelet maxima. The 2-D reconstruction algorithm is a direct generalization of the 1-D algorithm.

The solution to the reconstruction problem will be a sequence of two functions, $\{g_j^1, g_j^2\}_{j \in \mathbf{Z}}$. They must have the same local maxima set as the given one, $[A_{2^j}]_{j \in \mathbf{Z}}$, and satisfy the 2-D wavelet reproducing equation. These two constraints, as in the one-dimensional case, define two subsets in $\mathbf{L}^2(\mathbf{R}^2)$. Denote by Γ the subset satisfying the

local maxima constraint, and **V** the subset satisfying the reproducing equation. With an initial guess, the WT of an image is reconstructed from wavelet maxima by projecting alternately onto Γ and **V**.

Again, the projection onto the subset **V** can be implemented by doing an inverse 2-D WT followed by a WT. An efficient implementation of projection onto the subset Γ is detailed in [48]. As an example, the image on the top of the left column in *Figure 5.6* is reconstructed from wavelet maxima for 10 iterations. It can be compared with the original in *Figure 2.2*.

Figure 5.6. (next page) The wavelet maxima representation of an MR image. Counting clockwise from top right, the first seven images form the wavelet maxima representation; the top one in the left column is reconstructed from wavelet maxima. The original image is in *Figure 2.2*.



(*Captions on previous page*)

Chapter 6
The Wavelet Maxima Tree and Metrics

6.1 Motivations

The wavelet maxima representation reviewed in the last chapter is sufficient for coding purposes. Such applications have already been shown by Mallat and Zhong[51]. For other signal processing and analysis tasks, the raw wavelet maxima are too loosely structured. We need a higher level representation that characterizes scale-spatial relations between wavelet maxima explicitly. To explain specifically what we want and what we can possibly get, let us examine the evolution property of wavelet maxima with scale parameters varied continuously.

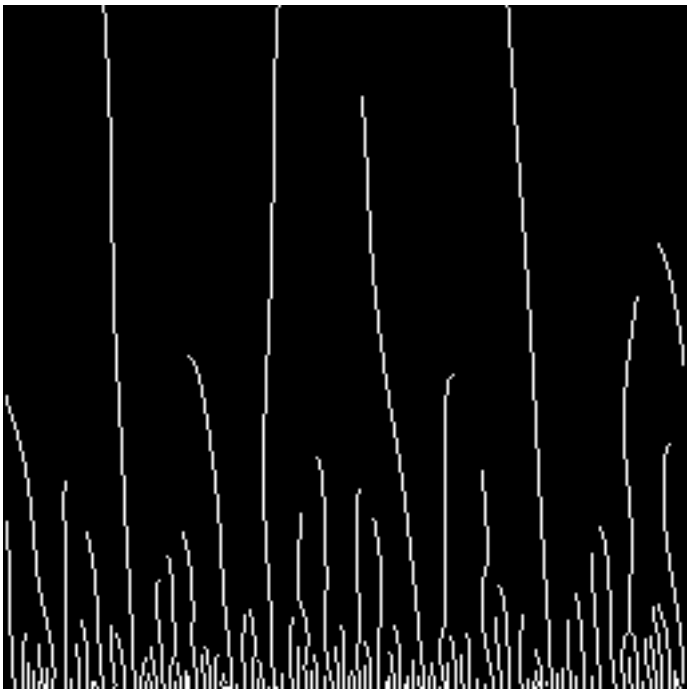
Figure 6.1 shows the traces of wavelet extrema and modulus maxima of a simulated 1-D signal in the wavelet scale-space. The original signal is in *Figure 5.3*. In accordance with the conventional plot of the Gaussian scale-space which is to be addressed later, we have reversed the direction of scale axis so that the finest scale is at the bottom and the coarsest scale is on top. The wavelet scale-space is computed through a densely sampled CWT with the cubic spline wavelet introduced in §5.1.2.

The evolution property of wavelet extrema in the wavelet scale-space closely resembles that of zero-crossings or extrema in the Gaussian scale-space. A Gaussian scale-space is computed by convolving a signal against a Gaussian (or derivatives of a Gaussian) with the size of the Gaussian varied continuously. The scaling properties of zero-crossings or extrema in the Gaussian scale-space have been studied extensively by Witkin[77], Koenderink[35], Babaud *et al*[7], and Yuille and Poggio[83]. We summarize some of their key results below with respect to zero-crossings of 1-D signals in the Gaussian scale-space. It is understood that these results apply to extrema as well.

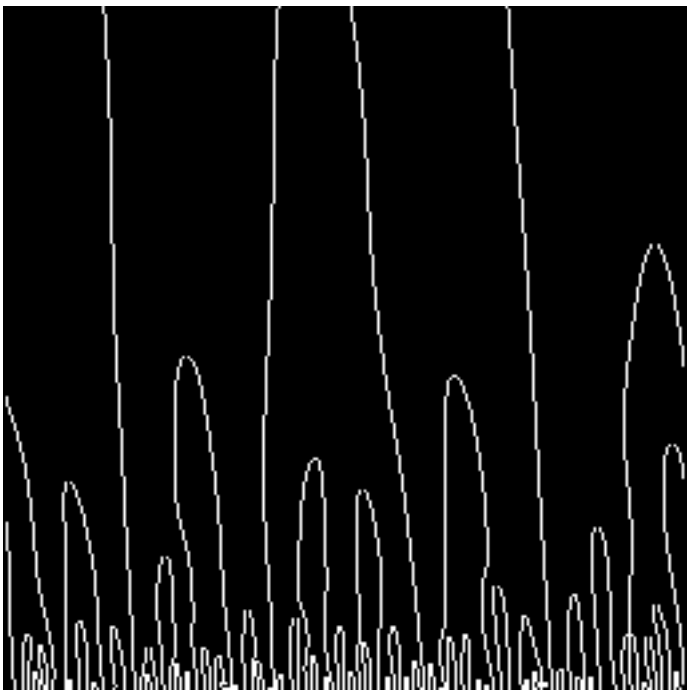
1) Zero-crossing traces are either lines from small to very large scales or closed "arches". The two arms of each arch form a complementary pair originating at small scales. As the scale goes coarser (large), the pair approach each other at the apex of the arch; they collide and annihilate[77]. See *Figure 6.2*.

2) Some possible behaviors of zero-crossings are never empirically observed in the Gaussian scale-space[83]. These cases are shown in *Figure 6.3*.

3) The above observations reflect a more general property of the Gaussian scale-space, namely the causality property shown by Koenderink[35] and Babaud *et al*[7]. It states that any feature (e.g. zero-crossings, extrema, etc.) at a coarse scale must have its cause in the finer scale. Existing features can be kept or destroyed but new features can never be created as one moves from fine to coarse scales.



(a)



(b)

Figure 6.1. Evolution properties of wavelet maxima and extrema in the wavelet scale-space. *(a)* the maxima traces of $|W_s f|$; *(b)* the extrema traces of $W_s f$. The original signal is in *Figure 5.3*.

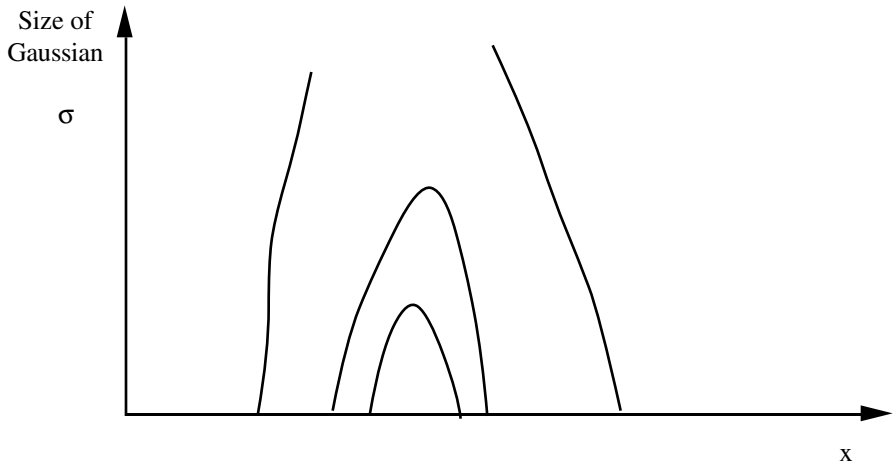


Figure 6.2. Typical behavior of zero-crossings in the Gaussian scale-space.

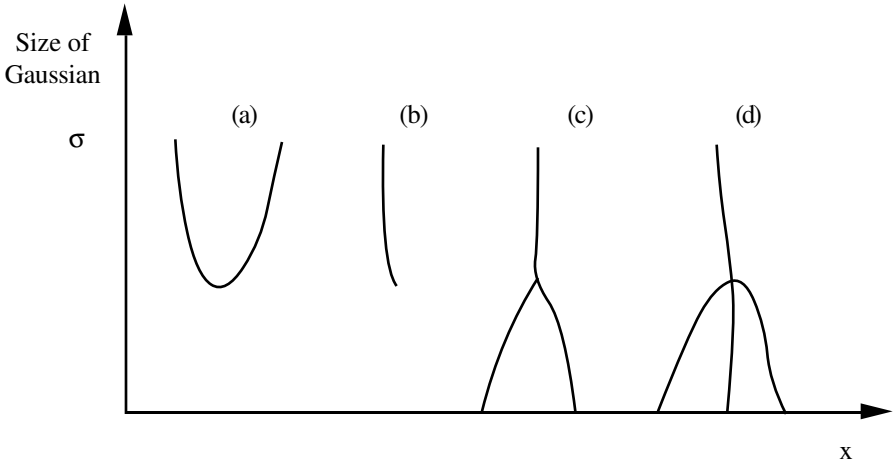


Figure 6.3. Possible behaviors of zero-crossings (from Yuille and Poggio [83]). Case (a) can never occur for a Gaussian filter; Cases (b) and (c) cannot occur for any filters. Case (d) only occurs for Gaussian filters in special conditions, but not with practical signals.

Generally speaking, these nice evolution properties of zero-crossings and extrema are not guaranteed when the kernel used for generating the scale-space is not a Gaussian or its derivatives. However, the particular scaling function and wavelet we are using is a good approximation to a Gaussian and its first-order derivative (see *Figure 5.2*). It is reasonable to assume that the scale-space generated with such a wavelet possesses the causality

property. Indeed, we have not observed any violations of causality for a large class of signals with which we have experimented. Zhong suggested the same assumption in [85]. As an example, the wavelet extrema traces in *Figure 6.1(b)* show the same pattern as zero-crossing traces in the Gaussian scale-space in *Figure 6.2*. In the modulus wavelet maxima traces of *Figure 6.1(a)*, we observe that some maxima of $|W_s f|$ do not form a closed arch. This is because these maxima evolve to minima before annihilation.

The wavelet maxima we detect in a dyadic WT can be considered as samplings of wavelet maxima traces in the scale-space. Unfortunately, the sampled wavelet maxima as a representation have lost the structural information that is present in the wavelet maxima traces. First, some wavelet maxima at different dyadic scales arise from the same edge in the signal; they belong to the same wavelet maxima trace. Second, the waveforms of the signal have some hierarchy that is shown by the nested arches in the wavelet maxima traces. For example, in *Figure 6.1*, those small arches enclosed in bigger ones correspond to small waveforms riding on top of the bigger ones in the signal.

Our objective now is to retrieve the structural information that is contained in the wavelet maxima traces from their dyadic samplings. By including the structural information with the wavelet maxima, we expand Mallat and Zhong's wavelet maxima representation. The expanded representation will be more suitable for applications of signal processing and analysis. From here on, we shall use "raw wavelet maxima representation" in some circumstances to refer to Mallat and Zhong's method in contrast to the expanded one we are developing.

6.2 The Wavelet Maxima Tree of One-Dimensional Signals

6.2.1 Definitions and Properties

A wavelet maxima tree ("tree" hereafter) is a hierarchical data structure that organizes wavelet maxima across scales. We define below some notation.

- *Level:* Each level of a maxima tree corresponds to one level of resolution, or scale. The coarsest scale in the DWT defines the *root level* of the tree, and the finest scale defines the *tip-end level*.
- *Height:* The total number of levels in a tree defines the height of the tree.
- *Node:* The wavelet maxima at a certain scale are nodes at that level of the tree.
- *Parent/Child :* Two nodes that are linked at neighboring levels form a parent-child pair. The node at the coarser scale is called the parent node and the one at the finer scale is called the child node. A parent-child pair is identified if two nodes at neighboring levels have strong interactions. A quantitative measure of interactions between nodes at neighboring levels is to be defined in the next section. By the causality assumption, a parent node must have at least one child node.
- *Branch/Main Branch:* A branch is a path from a node to the tip end of the tree. By the causality assumption, from any node there will be one and only one branch that consists of those nodes belonging to the same wavelet maxima trace in the wavelet scale-space. This branch is called the main branch from that node. Intuitively, the main branch from a node is an edge that is present at all levels from that node to the tip end; other branches are smaller edges emerging at finer scales that are superimposed with the major one.

A wavelet maxima tree is a data structure. Each node in the tree has several attributes: *level*, *position*, *value*, *parent node*, *child nodes*, and the *main branch to which it belongs*. The first three attributes have been included in the raw wavelet maxima representation. The other three attributes are unique for characterizing the relation between wavelet maxima.

It is easy to see that given a wavelet maxima tree by a specified node set, the wavelet maxima and their cross-scale relations are completely characterized. The tree branches convey important information that is absent in the raw wavelet maxima representation. A main branch corresponds to a single edge observed at different scales. It is an estimate of a wavelet maxima trace in the scale-space from its dyadic samplings. The total number of

main branches accounts for the number of non-identical edges in the signal. Other branches characterize the hierarchy and interaction of edges. A non-main branch from a node is the fine detail of the main branch that is going to emerge at small scales. As one travels from the tip-end to root levels of a wavelet maxima tree, the number of nodes decreases, which agrees with the causality assumption.

It is worth pointing out that those branches passing a node in the tree represent wavelet maxima traces which do not really intersect in the ideal wavelet scale-space. At any node (a wavelet maximum) in the wavelet scale-space, there is one and only one wavelet maxima trace passing through the node (see *Figure 6.1*). The main branch from a node in the tree is a discrete estimate of such a unique wavelet maxima trace. That other branches are associated with the same node is merely a representation of hierarchy of the branches.

6.2.2 Construction of One-Dimensional Wavelet Maxima Tree

The virtue of causality in scale-space motivates an approach called "coarse-to-fine tracking" which was first proposed by Witkin[77]. As stated in §6.1, any feature existing at a coarse scale must have continued presence at finer scales; new features can emerge as scale decreases; traces of different features do not intersect. One can therefore identify a feature at a coarse scale, and track it down to a fine scale along the trace of the feature in the scale-space. Based on this idea, Bergholm[11] developed a technique for "focusing" edges by slowly varying the scale parameter from coarse to fine in the Gaussian scale-space.

The same technique can be employed in the wavelet scale-space to track wavelet maxima if we use a CWT. Our situations and strategies are somewhat different in constructing a wavelet maxima tree from discrete wavelet maxima sampled at dyadic scales. This is because the sampling step-size increasing in the dyadic sequence may be too large to make a stable tracking and for determining a main branch in the tree. The uncertainties caused by discretization change our problem from coarse-to-fine "tracking" to coarse-to-

fine "correspondence". In order to make a correspondence between nodes at neighboring scales, we define a measure of interaction between two nodes at neighboring scales.

We start with the raw wavelet maxima representation. Denote by N_j the number of wavelet maxima at scale 2^j. As defined by (5.13), the wavelet maxima set for scale 2^j is,

$$A_{2^j} = \{ m_p = (x_p, W_{2^j} f(x_p)),\ p = 1, 2, 3, ..., N_j \}. \tag{6.1}$$

Likewise, the wavelet maxima set for the neighboring finer scale 2^{j-1} is,

$$A_{2^{j-1}} = \{ m_q = (x_q, W_{2^{j-1}} f(x_q)),\ q = 1, 2, 3, ..., N_{j-1} \} \tag{6.2}$$

The wavelet reproducing equation (2.12) provides a measure of interaction between an arbitrary pair of wavelet maxima, $\{m_p , m_q\}$, at two neighboring scales. Denote by $T_{p,q}$ the measure of influence of m_p on m_q . $T_{p,q}$ is computed by evaluating (2.12) at m_q at scale 2^{j-1} with respect to m_p at scale 2^j,

$$T_{p,q} = W_{2^j} f(x_p) K_{j,\, j-1}(x_q - x_p) \tag{6.3}$$

where $K_{j,\, j-1}(x)$ is the WRK from scale 2^j to 2^{j-1}. Likewise, the measure of influence of m_q on m_p , $T_{q,p}$, is given by,

$$T_{q,p} = W_{2^{j-1}} f(x_q) K_{j-1,\, j}(x_p - x_q) \tag{6.4}$$

We developed an algorithm for constructing the wavelet maxima tree recursively from coarse to fine scales. For a pair of neighboring scales, let us identify one of them as “the coarser” and the other as “the finer”. The algorithm has four basic steps to establish connections between wavelet maxima at the coarser and finer scales.

The first step is to design a tracking window. The size of the tracking window is proportional to the effective size (ref. *Table 6.1*) of the WRK between the coarser and finer scales.

The second step is tracking. The tracking window is centered at each wavelet maximum at the coarser scale and those wavelet maxima inside the tracking window at the finer scale are recorded.

The third step is to determine the hierarchy. A wavelet maximum at the finer scale may fall into several tracking windows associated with different wavelet maxima at the coarser scale. These are all candidates to be the parent node of the wavelet maximum at the finer scale. Among these candidates, the one that maximizes $T_{p,q}$ is determined as the parent node and connected to its child.

The fourth step is to determine the main branch. After the last step, a wavelet maximum at the coarser scale may have several nodes connected to it from the finer scale. The main branch is drawn through the child node that maximizes $T_{q,p}$.

With notation defined in (6.1)-(6.4), we give an algorithmic sketch of the above steps for constructing a wavelet maxima tree.

Algorithm 6.1. Construction of 1-D Wavelet Maxima Tree.

For $j = J, J\text{-}1, \ldots, 2$, build the tree between scales 2^j and 2^{j-1} by Steps 1-4:

Step 1. Design a tracking window B_j of size $\beta\Delta_j$, where Δ_j is the effective size of the WRK between scales 2^j and 2^{j-1}, and β is a constant.

Step 2. For $p = 1, 2, \ldots, N_j$, pick m_p from A_{2^j} : With B_j^p , a tracking window centered at x_p, track down to scale 2^{j-1}. If no $m_q \in A_{2^{j-1}}$ is found inside B_j^p such that $W_{2^{j-1}} f(x_q)$ has the same sign as $W_{2^j} f(x_p)$, increase the size of B_j^p until at least one such $m_q \in A_{2^{j-1}}$ is found inside B_j^p .

Step 3. For $q = 1, 2, \ldots, N_{j-1}$: select $U_{2^j}^q$, a subset of A_{2^j} that has N_j' elements, $U_{2^j}^q = \{ [m_r = (x_r, W_{2^j} f(x_r)) \in A_{2^j}, r = 1, 2, \ldots, N_j']: m_q \in A_{2^{j-1}}$ is inside $B_j^r \}$. The parent node of m_q is determined as $\eta_q \in U_{2^j}^q$ such that $\{ |T_{r,p}|, r = 1, 2, \ldots, N_j' \}$ is maximized.

Step 4. For $p = 1, 2, ..., N_j$: select $V_{2^{j-1}}^{p} = \{ [m_s = (x_s, W_{2^{j-1}} f(x_s)) \in A_{2^{j-1}}]$: all child nodes of $x_p\}$. The main branch from x_p is drawn through the child node $\mu_p \in V_{2^{j-1}}^{p}$, such that $W_{2^j} f(x_p)$ has the same sign as $W_{2^{j-1}} f(\mu_p)$ and $\{ |T_{s,p}|$, s = all child nodes of $x_p\}$ is maximized.

Table 6.1. First ten coefficients of the discrete reproducing kernel $K_{j,j-1}$

n	$K_{6,5}(n)$	$K_{5,4}(n)$	$K_{4,3}(n)$	$K_{3,2}(n)$	$K_{2,1}(n)$
1	7.952E-3	1.597E-2	3.244E-2	6.922E-2	1.579E-1
2	7.937E-3	1.584E-2	3.142E-2	6.042E-2	2.371E-3
3	7.891E-3	1.547E-2	2.849E-2	3.766E-2	-7.446E-2
4	7.814E-3	1.487E-2	2.394E-2	1.020E-2	-4.895E-2
5	7.707E-3	1.404E-2	1.824E-2	-1.237E-2	-2.186E-2
6	7.571E-3	1.302E-2	1.197E-2	-2.492E-2	-9.795E-3
7	7.406E-3	1.183E-2	5.694E-3	-2.761E-2	-3.733E-3
8	7.215E-3	1.049E-2	-7.595E-5	-2.395E-2	-1.300E-3
9	6.997E-3	9.050E-3	-4.952E-3	-1.181E-2	-2.058E-4
10	6.754E-3	7.536E-3	-8.701E-3	-1.127E-2	4.275E-6

In implementation, Step 1 can be pre-computed for a given wavelet. The WRKs used to evaluate (6.3) and (6.4) can also be pre-computed and stored. For the wavelet we are using, $K_{j,\,j-1} = K_{j-1,\,j}$ and is symmetrical. *Table 6.1* lists the first ten coefficients of $K_{j,\,j-1}$ for $j = 6, 5, 4, 3, 2$. It is observed that the reproducing kernels drop rapidly in magnitude at small scales. Consequently, $|T_{p,q}|$ and $|T_{q,p}|$ defined by (6.3) and (6.4) drop rapidly at small scales as $|x_p - x_q|$ becomes large. In this perspective, *Algorithm 6.1* determines the

tree hierarchy and main branches by virtue of proximity in wavelet maxima. This agrees with man's habitual approach to correspondence and grouping problems. *Figure 6.4* is a wavelet maxima tree constructed by the above algorithm from the raw wavelet maxima in *Figure 5.4*.

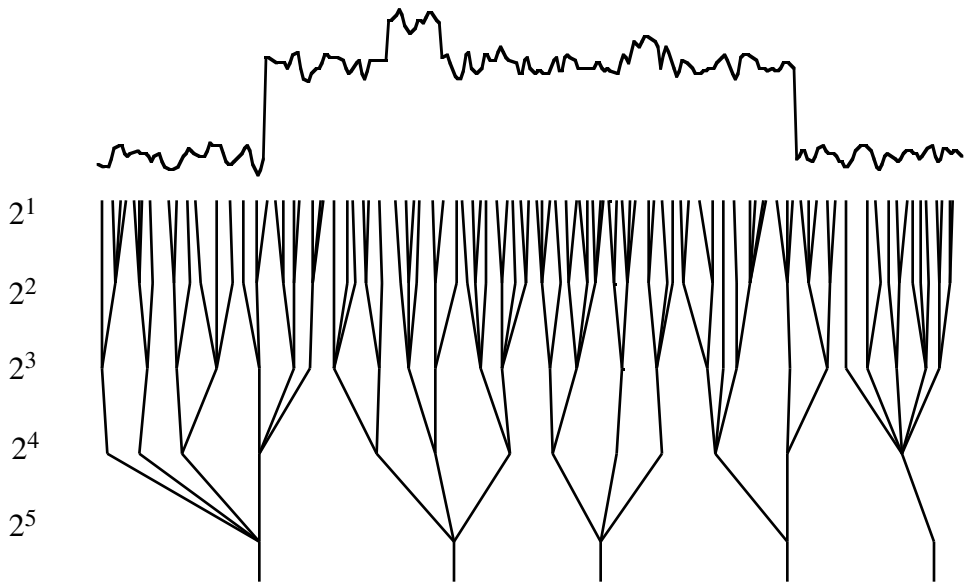


Figure 6.4. The wavelet maxima tree of a signal. The raw wavelet maxima representation of this signal is shown in *Figure 5.4*.

6.3 The Wavelet Maxima Tree of Images

6.3.1 Extended Definitions and Properties

The 1-D wavelet maxima tree developed in §6.2 can be extended to two-dimensions to represent images. Most definitions and properties of the 1-D wavelet maxima tree summarized in §6.2.1 remain the same in two-dimensions. There are two distinctions that we shall address below.

Node. In images, individual wavelet maxima are often not independent features; they are points belonging to certain lines or curves observed at multiple scales. In computer vision, an edge point is sometimes called "edgel" (edge element) to be distinguished from

an "edge" which is a chain of edge elements. To adopt this notation we shall call individual wavelet maxima in two dimensions "multiscale edgels" while reserving "multiscale edges" for contours of wavelet maxima. Obviously, a more expressive primitive for representing images would be edges instead of edgels. Mallat and Zhong[48] have discovered that thresholding and coding wavelet maxima contours are more efficient and robust than doing the same with individual wavelet maxima. Considering these, we shall use wavelet maxima contours as nodes in the 2-D wavelet maxima tree. Methods for contouring wavelet maxima are described in the next section.

Branch/Main Branch. A main branch in a 1-D wavelet maxima tree represents the trace of a single edge observed at several scales. We are able to determine the main branch by *Algorithm 6.1* because it does not intersect with others. This is however not true generally for wavelet maxima contours in two-dimensions. Babaud *et al*[7] showed that zero-crossing contours in the Gaussian scale-space, though they may not vanish, are free to split and merge at increasingly fine scales. There is no guaranteed one-to-one correspondence of contours at different scales. An example of such a phenomenon is shown in *Figure 6.5*. These results suggest that we must abandon the assumption of a *unique* main branch from a node. As we track from any node down to the finer scale, we could have either one main branch that is a continued (non-split) wavelet maxima contour, or several main branches that are contours split from the one at the coarser scale. An algorithm for the 2-D tree construction should consider these cases accordingly.

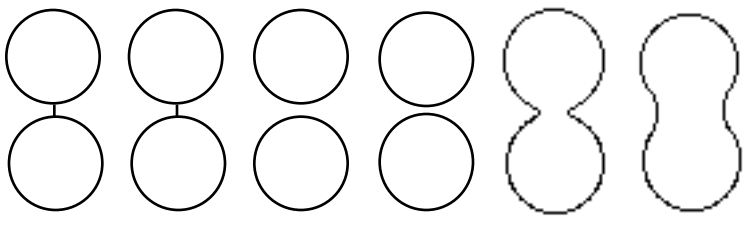


Figure 6.5. The splitting and merging of zero-crossing contours in the Gaussian scale-space (from Babaud *et al* [7]).

6.3.2 Contouring Wavelet Maxima

Contouring edgels at a fixed scale has been studied extensively and many practical algorithms have been developed. Despite the algorithmic difference in searching for a solution, one or more of the following measures are often used to decide if two edgels should be linked: (1) geometric compatibility, (2) positional proximity, and (3) amplitude similarity. Accordingly, edgels to be linked must be close to each other and have similar gradient angles and magnitudes. These measures may be combined to form a unified figure of merit[71].

For contouring wavelet maxima, it is possible to exploit information at multiple scales. Again we follow a coarse-to-fine strategy. Starting at the coarsest scale, we contour wavelet maxima by setting thresholds for the above three measures. As we move down to fine scales, the positional proximity requirement is relaxed to include proximity to the contours already constructed at the immediately coarser scale. For example, suppose two edgels are separated by a gap that could be rejected by the proximity threshold; they would still be linked if there is a contour going through these two positions at the immediately coarser scale. This approach follows Rosenfeld's idea of using features at the coarse scale as context for fine scales[67].

6.3.3 Construction of Two-Dimensional Wavelet Maxima Tree

In this section, we extend Algorithm 6.1 developed in §6.2.2 to the construction of a tree of wavelet maxima contours. For simplicity, we shall still use the term "wavelet maxima tree". It is understood that the nodes in the tree now are really wavelet maxima contours.

Similar to the approach in §6.2.2, we define a measure of interaction between two nodes at neighboring scales. Denote by $C_{2^j}^p$ a wavelet maxima contour containing n_p wavelet maxima at scale 2^j,

$$C_{2^j}^p = \{ \cup m_k = \cup[x_k, y_k, W_{2^j}^1 f(x_k, y_k), W_{2^j}^2 f(x_k, y_k)],\ m_k \text{ is a wavelet maximum on the contour } C_{2^j}^p,\ k = 1, 2, ..., n_p \}. \quad (6.5)$$

Likewise, a wavelet maxima contour containing n_q wavelet maxima at the neighboring finer scale 2^{j-1} is,

$$C_{2^{j-1}}^q = \{ \cup m_l = \cup[x_l, y_l, W_{2^{j-1}}^1 f(x_l, y_l), W_{2^{j-1}}^2 f(x_l, y_l)],\ m_l \text{ is a wavelet maximum on the contour } C_{2^{j-1}}^q,\ l = 1, 2, ..., n_q \}. \quad (6.6)$$

As we mentioned in §2.4, there are four sequences of 2-D WRKs resulting from wavelets in two orientations. Denote by $\{K_{i,j}^{1,1}\}$ and $\{K_{i,j}^{2,2}\}$ the sequences of WRKs for intra-orientations, and $\{K_{i,j}^{1,2}\}$ and $\{K_{i,j}^{2,1}\}$ the sequences of WRKs for inter-orientations. A measure of interaction between wavelet maxima contours at two neighboring scales can be defined by generalizing (6.3) and (6.4) with these WRKs. Adopting the same notation of $T_{p,q}$ as in (6.3), the influence of $C_{2^j}^p$ on $C_{2^{j-1}}^q$ is measured by,

$$T_{p,q} = \sum_{l=1}^{n_q} \sqrt{(\xi_l)^2 + (\zeta_l)^2} \quad (6.7)$$

where

$$\xi_l = \sum_{k=1}^{n_p} \left[W_{2^j}^1 f(x_k, y_k) K_{j,j-1}^{1,1}(x_l - x_k, y_l - y_k) + W_{2^j}^2 f(x_k, y_k) K_{j,j-1}^{1,2}(x_l - x_k, y_l - y_k) \right]$$

$$\zeta_l = \sum_{k=1}^{n_p} \left[W_{2^j}^1 f(x_k, y_k) K_{j,j-1}^{2,1}(x_l - x_k, y_l - y_k) + W_{2^j}^2 f(x_k, y_k) K_{j,j-1}^{2,2}(x_l - x_k, y_l - y_k) \right]$$

Similarly, the influence of $C_{2^{j-1}}^q$ on $C_{2^j}^p$ is measured by,

$$T_{q,p} = \sum_{k=1}^{n_p} \sqrt{(\xi_k)^2 + (\zeta_k)^2} \quad (6.8)$$

where

$$\xi_k = \sum_{l=1}^{n_q} \left[W_{2^{j-1}}^{1} f(x_l, y_l) K_{j-1,j}^{1,1}(x_k - x_l, y_k - y_l) + W_{2^{j-1}}^{2} f(x_l, y_l) K_{j-1,j}^{1,2}(x_k - x_l, y_k - y_l) \right]$$

$$\zeta_k = \sum_{l=1}^{n_q} \left[W_{2^{j-1}}^{1} f(x_l, y_l) K_{j-1,j}^{2,1}(x_k - x_l, y_k - y_l) + W_{2^{j-1}}^{2} f(x_l, y_l) K_{j-1,j}^{2,2}(x_k - x_l, y_k - y_l) \right]$$

We also design a window for tracking wavelet maxima contours from coarse to fine scales. Denote by Δ_j the size of a square window bounding the regions of support of the four WRKs from scale 2^j to 2^{j-1}. Define a *point tracking window*, b_j of size $\beta\Delta_j$, where β is a constant. For an arbitrary wavelet maxima contour $C_{2^j}^p$, put a point tracking window b_j on each of the wavelet maxima on $C_{2^j}^p$ with the window center right on the wavelet maximum. As shown in *Figure 6.6*, there are two choices for the *contour tracking window*, B_j^p . One is the envelope bounding all b_j's, and the other is the smallest rectangular window enclosing all b_j's. We choose the latter for computational efficiency.

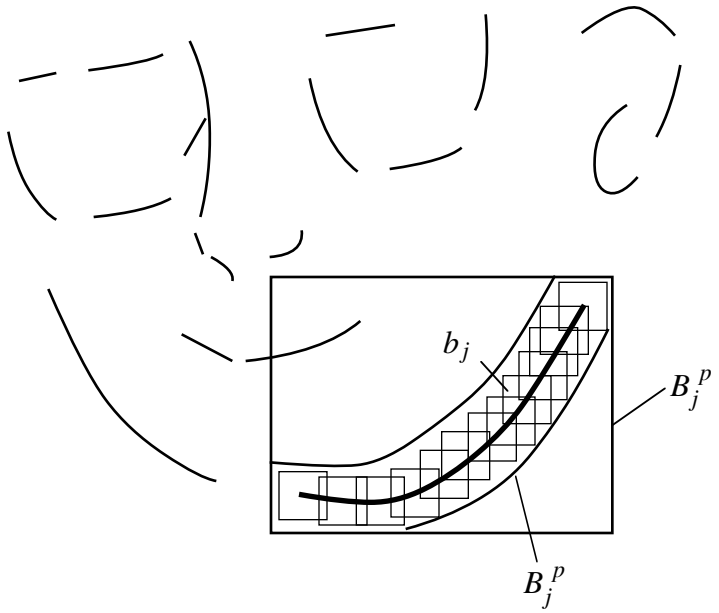


Figure 6.6. Design of contour tracking window. One choice is the envelope bounding all b_j's, and the other is the smallest rectangular window enclosing all b_j's.

Although it is theoretically possible for wavelet maxima contours to split and merge in both directions of scale changes, we shall impose a one-way causality to reduce algorithmic complexity. We assume that wavelet maxima contours can only split but not merge as scale decreases, and only merge but not split as scale increases. In practice, this assumption basically agrees with our visual experience that features often break into details at fine scales, but join together at coarse scales. Furthermore, it is reasonable to assume that a wavelet maxima contour split from another one should be in close geometric proximity to its originator at the coarser scale. To be specific, the planar distance between the two contours should not be much greater than the average length of the two contours; the planar angle between the two contours should be rather small. Denote by $d_{p,q}$ and $\alpha_{p,q}$ the distance and angle between two wavelet maxima contours $C_{2^j}^p$ and $C_{2^{j-1}}^q$ at the neighboring scales. We accept $C_{2^{j-1}}^q$ as a split component from $C_{2^j}^p$ if

$$d_{p,q} < \gamma \sqrt{n_p n_q} \tag{6.9}$$

and

$$\alpha_{p,q} < \alpha_0 \tag{6.10}$$

where γ and α_0 are two constants. We have used $\gamma = 0.5$ and $\alpha_0 = 0.5\pi$ in implementing our algorithm.

There are a number of ways to define the distance and the angle between two planar curves. We propose a "matched filter" approach: Given a pair of planar curves, fix one of them as reference and apply affine transformation (i.e. translation and rotation) to the other. The translation and rotation parameters that maximize the correlation between the two curves are defined as the distance and the angle between the two curves.

Denote by N_j the number of wavelet maxima contours at scale 2^j. With all notations defined above and (6.7)-(6.10), the algorithm for constructing a 2-D wavelet maxima tree is outlined below. It follows four steps similar to those in Algorithm 6.1.

Algorithm 6.2. Construction of 2-D Wavelet Maxima Tree.

For $j = J, J\text{-}1, ..., 2$, build the tree between scales 2^j and 2^{j-1} by Steps 1-4:

Step 1. Determine the size of the point tracking window: $b_j = \beta\Delta_j$. For $p = 1, 2, ..., N_j$, determine the contour tracking window B_j^p as shown in *Figure 6.6*.

Step 2. For $p = 1, 2, ..., N_j$, pick $C_{2^j}^p$: track down to scale 2^{j-1} with B_j^p . If no contours at scale 2^{j-1} are found inside B_j^p , increase the size of b_j and update B_j^p until at least one contour at scale 2^{j-1} is found inside B_j^p .

Step 3. For $q = 1, 2, ..., N_{j-1}$: select $U_{2^j}^q = \{ [C_{2^j}^r, r = 1, 2, ..., N_j']: C_{2^{j-1}}^q$ is inside $B_j^r \}$. The parent node of $C_{2^{j-1}}^q$ is determined as $\eta_q \in U_{2^j}^q$ such that $\{ |T_{r,p}|, r = 1, 2, ..., N_j' \}$ is maximized.

Step 4. For $p = 1, 2, ..., N_j$: select $V_{2^{j-1}}^p = \{ C_{2^{j-1}}^s$: all child nodes of $C_{2^j}^p \}$. The main branches from $C_{2^j}^p$ are drawn through those child nodes such that $\{ |T_{s,p}|, s =$ all child nodes of $C_{2^j}^p \}$ is maximized or (6.9) and (6.10) are satisfied.

6.4 Metrics on the Wavelet Maxima Tree

A tree as a graphical representation is often associated with some metrics. Such a tree is called a "valued tree" or "weighted tree"[26]. The values or weights one assigns to the nodes and branches in the tree come from the physical structure the tree is representing, and are for the problem one is trying to solve. Although a wavelet maxima tree may be used for different tasks of signal analysis, our goal here is to exploit its application in noise reduction. We shall design metrics for object/noise discrimination on a wavelet maxima tree. Our design is motivated by some observations from visual perception.

6.4.1 Object/Noise Discrimination in Visual Perception

The human visual system (HVS) performs amazingly well in discriminating meaningful features or objects from a noise background. This capability can be interpreted by the perceptual organization theory which has been developed and extensively studied by Gestalt psychologists since the 1910s[66]. Four Gestalt laws of grouping are relevant in this context. They are proximity, similarity, continuation, and closure[76]. The first three principles have been used in §6.3.2 to contour wavelet maxima. The last one suggests that closed contours are even more favorable in grouping as they convey information about shapes. According to Gestalt psychologists, the HVS spontaneously uses these principles among others to derive relevant groupings and structures from an image without prior knowledge of its contents.

The fact that some elements in the image can be grouped into contours or even closed contours while others are not is no accident, as argued by Lowe[40]. This argument is based essentially on conditional probabilities. Given the infinite number of possible configurations of spatial points, the probability of forming a long or closed contour by accident is very small. Therefore, when we observe such contours, they manifest themselves as meaningful features and stand out from the noise background that consists of random and disjoint elements. Broadly speaking, features having stable *spatial* presence are significant, no matter if they are lines, contours, or shapes. This can be said a perceptual criterion for discriminating objects from noise; we call it *spatial stability* for short.

Early Gestalt psychologists did not embed scale in their theories explicitly. Obviously, scale is an important factor that could affect the outcome of perceptual organization. In the above contouring problem, the resulting contours can be very different if we do contouring at different scales. For example, sparse dots having the same magnitude can be perceived

as a contour by proximity and similarity only at a sufficiently large scale; such perception will stay stable for larger scales.

There is convincing psychophysical evidence that the HVS performs "multichannel decomposition"[37] on the visual input signals. These "channels" corresponding to different scales are retinal receptive fields of different sizes. It is then conceivable that features causing responses in multiple channels are likely to catch our attention. Witkin observed in the Gaussian scale-space that "(features which) survive over a broad range of scales tend to leap out at the eye, while the most ephemeral are not perceived at all."[77] Again, by conditional probabilities, those perceived features are not accidental. This suggests another perceptual criterion for object/noise discrimination, namely *scaling stability*—the stability of a feature over scale changes.

6.4.2 Design of Metrics for Object/Noise Discrimination

With the Gaussian scale-space, Witkin[77] defined a measure of scaling stability as the *lifespan* of a scale-space feature along the scale dimension. Later, Bischof and Caelli justified the use of such a measure by showing that the measure is a function of cross-scale correlation and will have a weak and randomly distributed response to white noise[14].

In a wavelet maxima tree, the main branches correspond to non-identical features. Three measures for scaling stability can be defined with respect to main branches. They are: (1) length of a main branch; (2) average wavelet maxima amplitude over a main branch; (3) rate of change of maxima amplitude along a main branch.

The first measure is similar to Witkin's in the Gaussian scale-space. The second is important in discriminating isolated noise in smooth regions. Lindeberg and Eklundh[39] showed that an isolated noise spike in smooth regions can survive over a fairly large range of scales but has small average amplitude. The third measure can be used to delineate different types of edges. Mallat and Hwang[53] showed that upon appropriate scaling of wavelet maxima amplitudes, the third measure is an estimate of the Lipschitz regularity of

singularities. We now define a metric of main branches in the wavelet maxima tree, called the *branch metric*,

$$P_b = (\textit{length})*(\textit{average amplitude}) \text{ of a main branch} \tag{6.11}$$

where the *length* is the number of nodes on a main branch and the *average amplitude* is calculated over all these nodes.

In a 2-D wavelet maxima tree, a node represents a wavelet maxima contour. Three measures of spatial stability can be defined with respect to wavelet maxima contours. They are: (1) length of the contour; (2) average amplitude of the contour; (3) closure of the contour. As discussed in §6.4.1, contours due to noise are often short, of small average amplitude, and do not self-close. Similar to the branch metric, we define a metric of wavelet maxima contours, called the *contour metric*,

$$P_c = (\textit{length})*(\textit{average amplitude}) \text{ of a wavelet maxima contour} \tag{6.12}$$

where the *length* is the number of wavelet maxima on a contour and the *average amplitude* is calculated over all wavelet maxima on the contour. This metric was used by Zhong[85] for compressing wavelet maxima contours at independent individual scales. In the 2-D wavelet maxima tree, P_c defines the amplitude of each node.

It is easy to see that P_b and P_c are dual in that the former measures across the scales for a specified feature while the latter measures across the image plane at a given scale. The advantage of using a multiscale representation is that we can benefit from employing both scaling stability and spatial stability in discriminating objects from noise. The perceptual criterion can then be re-stated: features having stable *scale-spatial* presence are significant; we call it *scaling-spatial stability*, which is measured by (6.11) when the nodes in the tree are wavelet maxima contours.

Figure 6.7 shows the valued wavelet maxima tree of the 1-D signal in *Figure 5.3*. The brightness of each main branch is proportional to the average amplitude of wavelet maxima on that branch. Clearly, the valued tree delivers more information than the non-valued tree in *Figure 6.4*. In particular, we notice that those long and bright main branches (with high P_b values) correspond to edges of the three boxcars that we want to keep. This observation suggests that P_b is a good quantitative measure of perceptual criteria for object/noise discrimination.

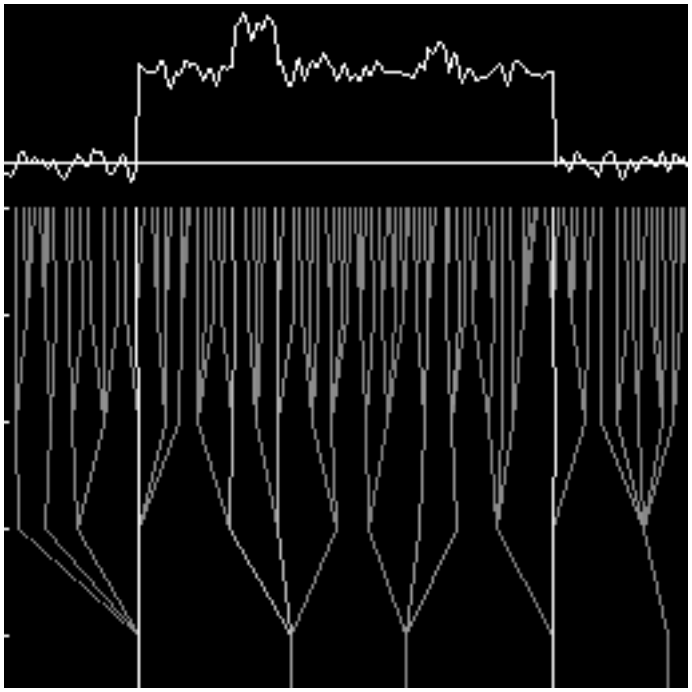


Figure 6.7. The valued wavelet maxima tree of a signal. The raw wavelet maxima representation of this signal is shown in *Figure 5.4*.

Chapter 7
Experimental Study of Noise Reduction

7.1 Implementation of the Noise Reduction Technique

7.1.1 The Computational Framework

The wavelet maxima tree and associated metrics developed in Chapter 6 have an immediate application in noise reduction for digital signals and images. The basic idea is that we build a wavelet maxima tree for a noisy signal or image, identify meaningful features with the help of the metrics we have designed, trim the noisy branches off the tree and reconstruct a denoised signal or image. The computational framework of this noise reduction technique is sketched by the block diagram in *Figure 7.1*. Methods and algorithms involved in all blocks except for the threshold selection in block #5 have been presented in Chapters 5 and 6. In the next section we give a practical method for selecting a threshold.

We have implemented the noise reduction algorithm of *Figure 7.1* in one and two-dimensions on a Sun SparcStation II. We found in all our experiments that with 10 iterations we were able to reconstruct WTs very well from wavelet maxima. Therefore, we set 10 to be the maximum number of iterations in implementing the alternating projection algorithm.

The overall execution time of the noise reduction algorithm is about 1 minute for a 256-point signal and 90 minutes for a 256×256 image. The percentage time used on each block of *Figure 7.1* is estimated and listed in *Table 7.1*. The most time-consuming computation is found to be reconstruction of WTs from wavelet maxima followed by 2-D tree construction, and computation of WTs. No efforts have been made to optimize the program overall for faster processing. However, this is possible according to our recent study on the parallelization of 2-D WT algorithms[45].

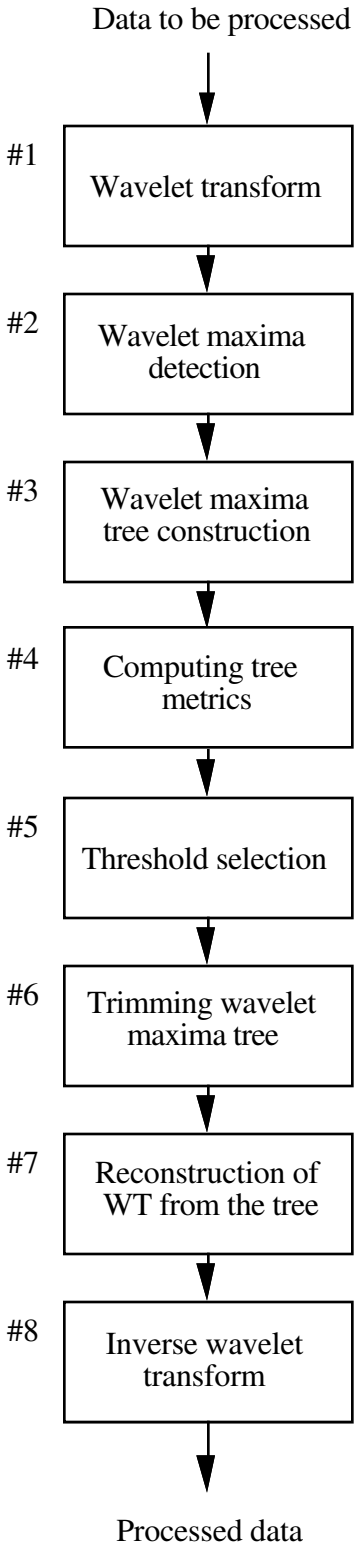


Figure 7.1. The computational framework of the noise reduction technique.

Table 7.1. Execution time percentage for each program block.

Re: *Figure 7.1.*		#1	#2	#3	#4	#5	#6	#7	#8
size	time	(%)							
256	1 min	4	1	3	1	1	1	84	4
256x256	90 min	4	1	7	1	1	1	81	4

7.1.2 Threshold Selection

In our noise reduction technique, denoising is accomplished by trimming off branches in the wavelet maxima tree that have small P_b values. A threshold must be set up for this operation like in many other techniques. It is hoped that the threshold selection should be easy and reasonable. Furthermore, the threshold should be relatively robust in the sense that varying the threshold slightly will not cause drastic changes in the thresholded signal.

The valued wavelet maxima tree in *Figure 6.7* gives us a visual impression that it is not very hard to identify the six branches corresponding to the three boxcars in noise. In particular, the four branches of the two larger boxcars are quite apparent by their length and brightness, while noisy branches are short and dim. This observation suggests that the branch metric P_b has made it easy for us to select a threshold.

Figure 7.2 is a histogram of P_b values of all main branches in the wavelet maxima tree of *Figure 6.7*. It can be seen that noisy branches form a large cluster of low P_b values. This is a typical P_b histogram of noisy signals. Without any prior knowledge about the signal and noise, a reasonable choice of a threshold would be the indicated valley to the right of the large low P_b cluster. In implementation, we first smooth the P_b histogram by a histospline[68], and then detect the first peak and the valley to the right of the first peak on the smoothed histogram. Denote by P_0 the value of P_b at the valley point. The threshold is chosen to be

$$TH = P_0 \pm \delta, \tag{7.1}$$

where δ is a positive constant.

Figure 7.3 is the wavelet maxima tree from *Figure 6.7* trimmed by a threshold computed from (7.1) with δ=0. The result is perfect: we got rid of all noisy branches but preserved all signal branches. This threshold is also very robust. For example, varying δ up to 20% of P_0, the result is invariant; further varying δ to 50% of P_0 could cause loss of the smallest boxcar, or take in 12% of noisy branches.

In general, a resulting wavelet maxima tree after trimming with a threshold determined by (7.1) will not be always as ideal as in *Figure 7.2*. In particular, we have found that for signals having low contrast-to-noise ratio (CNR), a good threshold is not easy to find with the 1-D algorithm. This is because there is no clear P_0 on the histogram of P_b when the CNR is low. In the two-dimensional case, using the contour metric will make the situation better[42,43]. Our experimental results in §7.2.1 will provide an example in this regard followed by relevant discussions in §7.3.1.

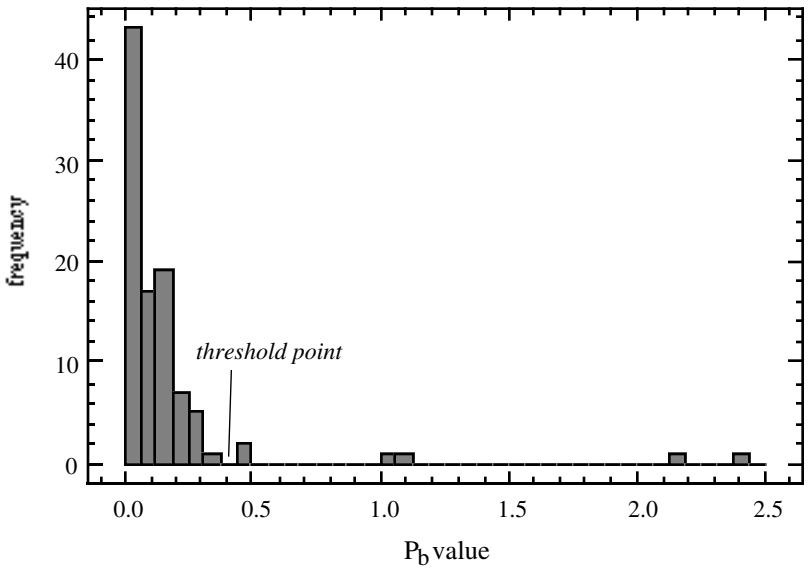


Figure 7.2. A histogram of P_b values for a wavelet maxima tree. The corresponding tree is shown in *Figure 6.7*.

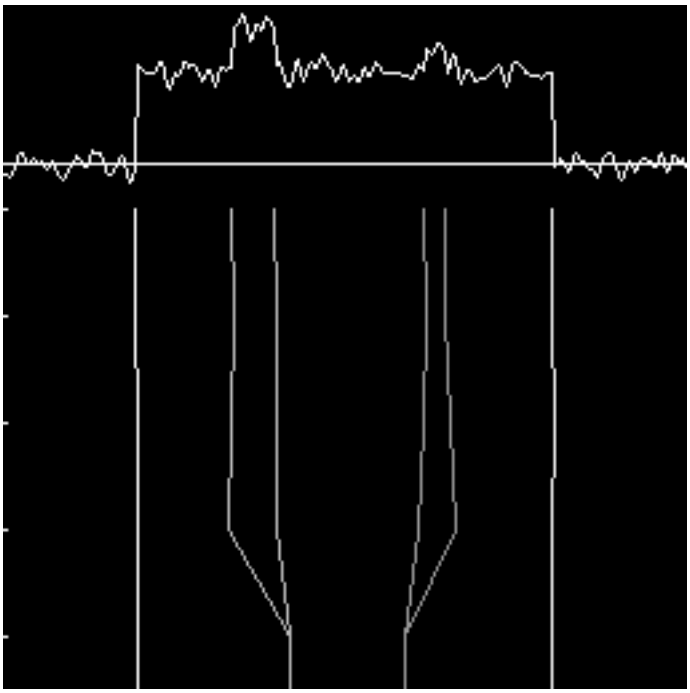


Figure 7.3. The trimmed wavelet maxima tree. The original tree is shown in *Figure 6.7*.

7.2 Experimental Results

In this section we present experimental results of noise reduction in one and two dimensions. Comparative analysis of experimental results will be given in §7.3.

7.2.1 One-Dimensional Results

Figure 7.4 is a denoised version of the signal that we have used before for demonstration. To review how the denoising technique works, one could look back to *Figures 5.3*, *5.4*, *6.7*, *7.2* and *7.3*. These graphs and pictures, together with *Figure 7.4*, give a complete graphic illustration of our denoising technique outlined in *Figure 7.1*.

In *Figure 7.4*, we also give a Gaussian filtered version ($\sigma = 4$ points) of the signal for comparison. Quantitative measures of SNR improvement and edge degradation or enhancement are given in *Table 7.2*. In *Table 7.2*, the six edges of three boxcars are numbered from left to right by the order of appearance in the graph.

Table 7.2. SNR and edge gradients of a denoised signal.

Re: *Figure 7.4.*		Edge Gradient Changes					
Signals	SNR (dB)	#1	#2	#3	#4	#5	#6
Noisy	15	1.092	0.460	-0.356	0.248	-0.200	-1.007
Denoised by wavelet	36.2 (+21.2)	1.114 (+2.0%)	0.447 (-2.8%)	-0.397 (+11.5%)	0.245 (-1.2%)	-0.186 (-7.0%)	-1.021 (+1.4%)
Filtered by Gaussian	31.5 (+16.5)	0.104 (-90.5%)	0.051 (-88.9%)	-0.055 (-84.6%)	0.025 (-89.9%)	-0.023 (-88.5%)	-0.095 (-90.6%)

Figure 7.5 is a simulated noise-free signal with different types of edges. On the left are step and ramp; the center consists of cosines of different frequencies; and on the right is an exponential multiplied by a ramp.

Figures 7.6(a) and *7.7(a)* are the signal from *Figure 7.5* with additive Gaussian white noise; the SNRs are 20 dB and 10 dB, respectively. The denoised signals are shown in *Figures 7.6(b)* and *7.7(b)*. In both cases, the SNR was lifted by about 13 dB.

Figure 7.8(a) is a line profile from a noisy MR image of human neck. The SNR is 16.3 dB. The denoised scan line, shown in *Figure 7.8(b)*, has a SNR of 28.6 dB.

In all experiments with both simulated and real data, we observed that the gradients at most edges in the denoised signal are 80% to 120% of their original values.

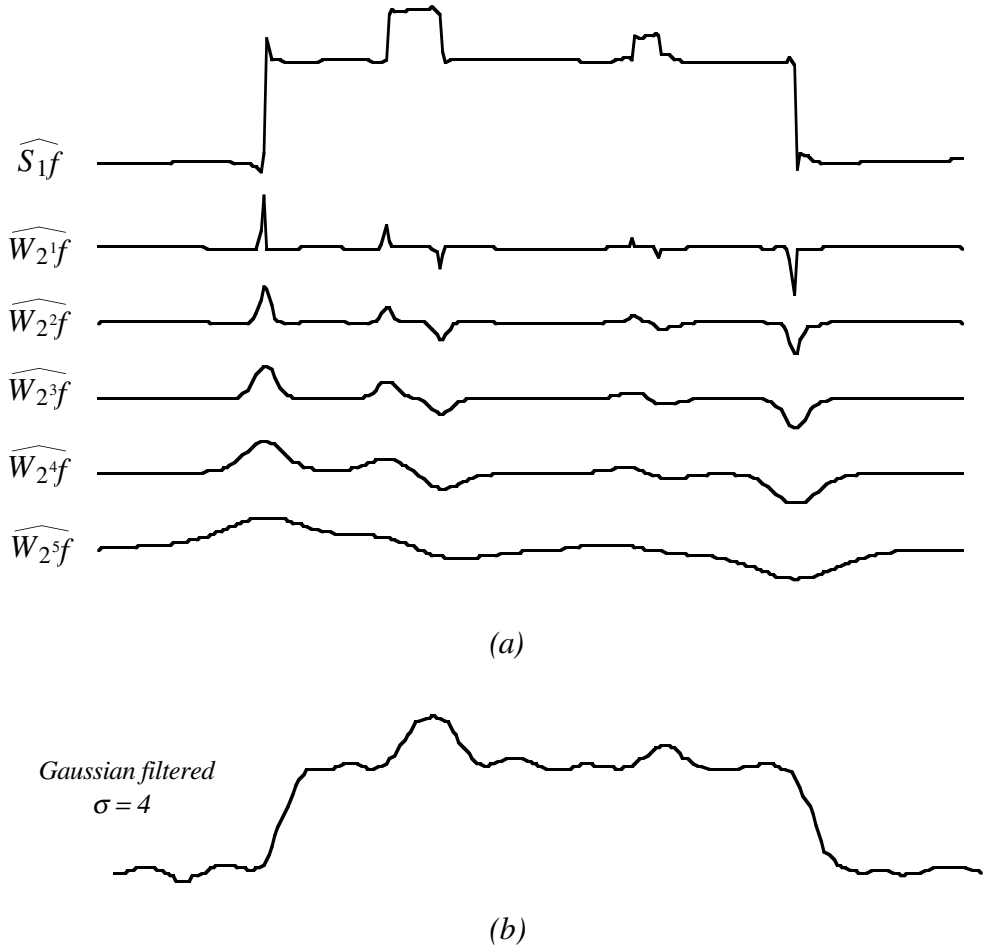


Figure 7.4. A denoised signal. *(a)* The denoised signal and its wavelet transform reconstructed from the trimmed wavelet maxima tree in Figure 7.3; *(b)* The signal filtered by a four-point wide Gaussian filter.

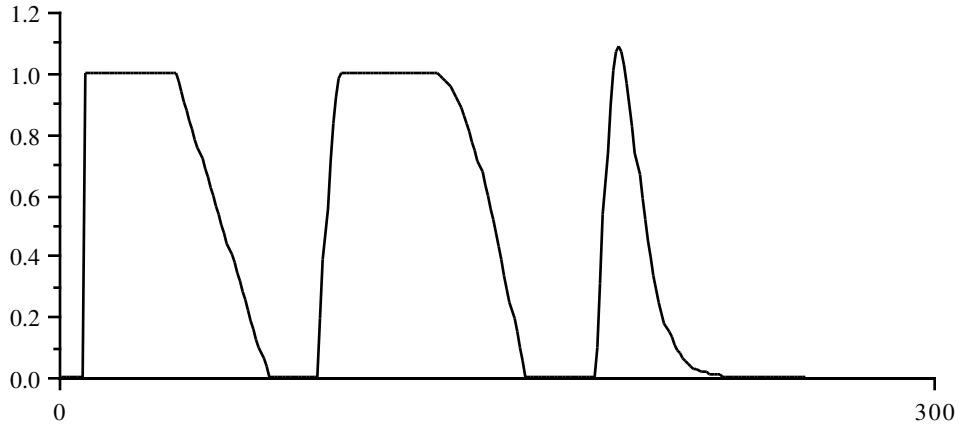
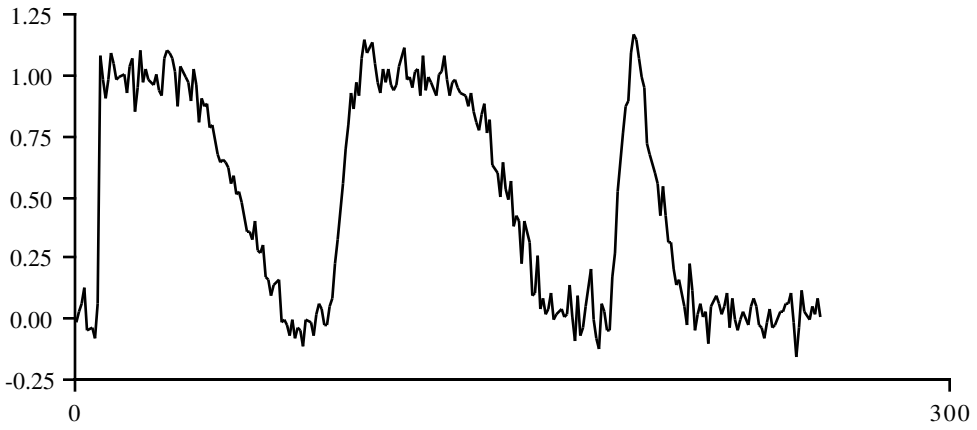
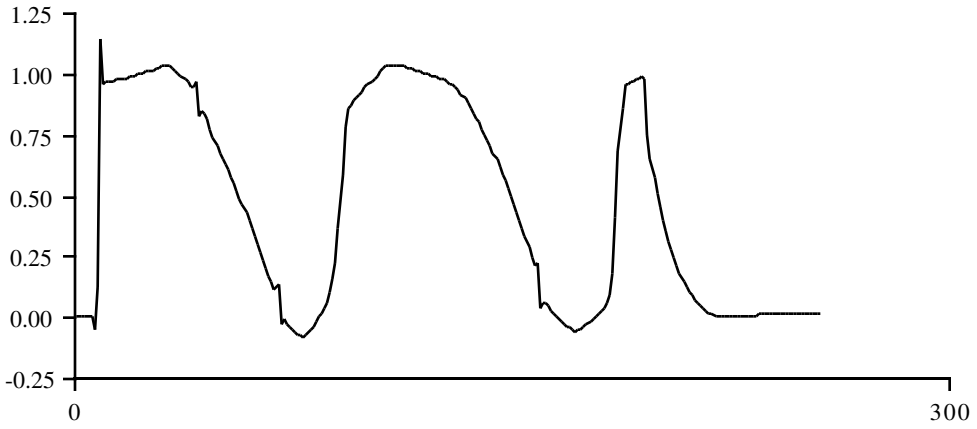


Figure 7.5. A simulated noise-free signal with different edge shapes.

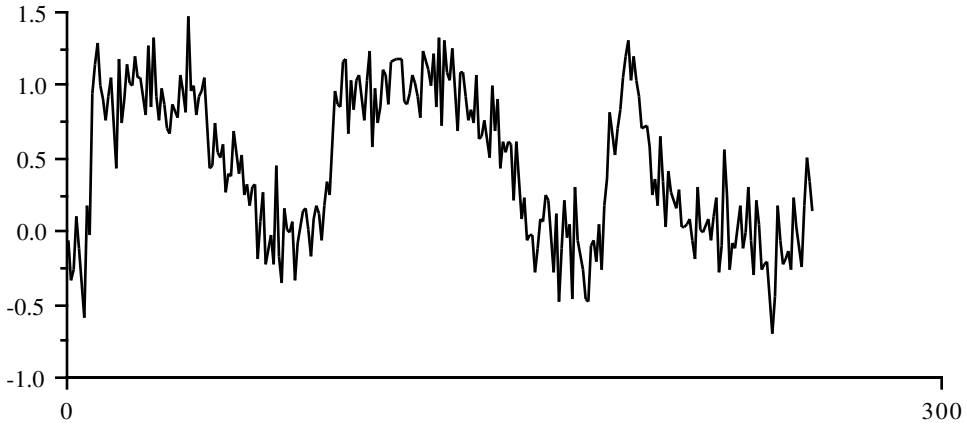


(a)

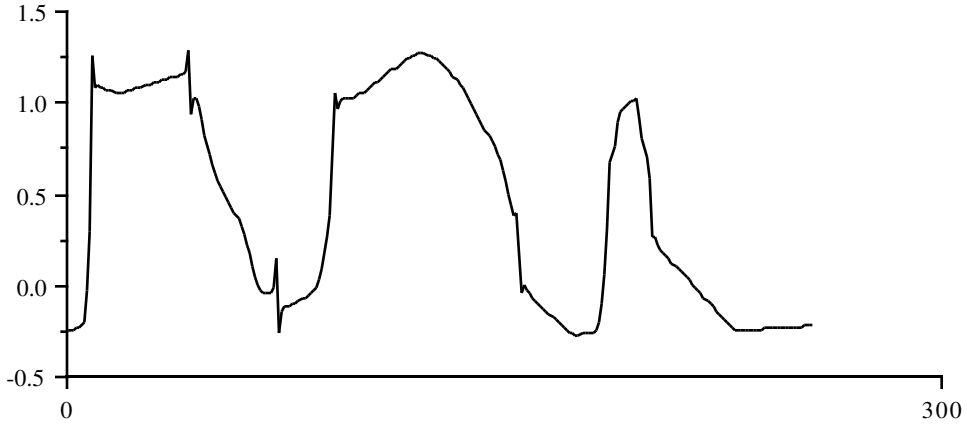


(b)

Figure 7.6. Signal recovery from noisy observations. *(a)* the simulated noise-free signal in *Figure 7.5* is added with Gaussian white noise, SNR=20 dB; *(b)* the denoised signal, SNR=33 dB.

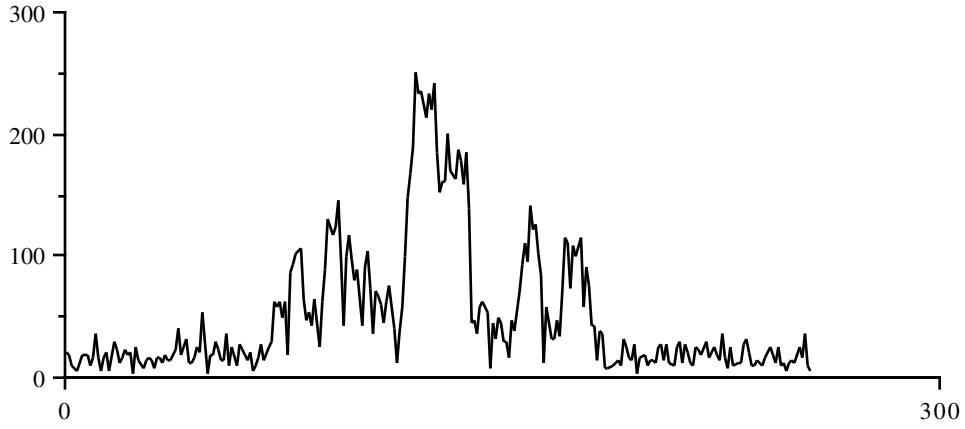


(a)

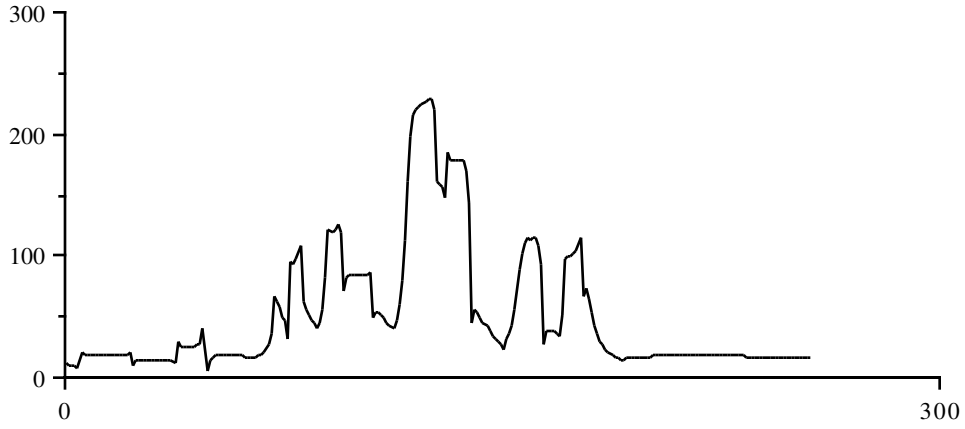


(b)

Figure 7.7. Signal recovery from severely noisy observations. *(a)* the simulated noise-free signal in *Figure 7.5* is added with Gaussian white noise, SNR=10 dB; *(b)* the denoised signal, SNR=23 dB.



(a)



(b)

Figure 7.8. *Denoising a MR image scan line.* (a) *a line profile from a noisy MR image of human neck, SNR=16.3 dB;* (b) *the denoised image scan line, SNR=28.6 dB.*

7.2.2 Two-Dimensional Results

We organized our 2-D experiments by three groups. In the first group, we used both 1-D and 2-D algorithms for denoising images. When using the 1-D algorithm, the rows and columns of the image were processed independently. In the second group, we did denoising with some conventional techniques. In particular, we used a Wiener filter and a median filter on the same set of images that we denoised by the 2-D wavelet method. In the third group, we applied our technique on two images that were previously tested with Mallat and Hwang's singularity analysis method[53]. Comparative analysis of these experimental results will be given in §7.3.

Group I:

Figure 7.9 is a simulated noise-free phantom image. The small dots on the phantom are designed to vary in size horizontally, and in contrast vertically. Two line profiles from the top two rows of the phantom dots are given. In *Figure 7.10*, Gaussian white noise is added to the phantom image; the SNR of the resulting noisy image is 12 dB. The local SNR at the top row of the image is about 5 dB. This image was processed by both 1-D and 2-D algorithms. The denoised images are shown in *Figures 7.11* and *7.12*. The SNR was increased to 26.8 dB by the 1-D algorithm, and 26.4 dB by the 2-D algorithm. A few representative line profiles are included in *Figures 7.11* and *7.12* for fine assessment.

Figure 7.13 shows a noisier simulated phantom image (SNR = 6 dB); the local SNR at the top row of the image is about 0 dB. This image was also processed by both 1-D and 2-D algorithms and the results are shown in *Figures 7.14* and *7.15*, respectively. The improved SNR is 22 dB by the 1-D algorithm and 24.1 dB by the 2-D algorithm. Again, a few representative line profiles are included in *Figures 7.14* and *7.15*. The numerical results of noise reduction are summarized in *Table 7.3*.

Group II:

Figures 7.16(a) and *7.17(a)* are two MR images of a human head. Both images were acquired from a male volunteer by a General Electric Signa 1.5 Tesla whole body MR scanner. The SNR is 14.5 dB in *Figure 7.16(a)* and 9.7 dB in *Figure 7.17(a)*. Both images were processed with our 2-D algorithm, a Wiener filter, and a median filter. The processed images are numbered *(b)-(d)* in *Figures 7.16* and *7.17*. The numerical results are given in *Table 7.4*. For evaluation of edge degradation in each of the three methods, we also computed difference images between the original noisy image and the processed ones. These difference images are shown in *Figures 7.18* and *7.19*.

In implementing the Wiener filter, we assumed the noise was white and its variance was estimated at the high frequency band of the image spectrum. An estimate of the signal spectrum was then obtained by "spectral subtraction"[38,73]. In implementing the median filter, either a cross or square window[32] was used, depending on whichever produced better results in terms of edge preservation. The size of the window was chosen to be 5×5.

Group III:

Figures 7.20(a) and *(b)*, and *7.21(a)* and *(b)* are two test images used by Mallat and Hwang in [53]. The SNR is 2.9 dB in *Figure 7.20(b)* and 5.3 dB in *Figure 7.21(b)*. In [53], Mallat and Hwang estimated the Lipschitz regularity of singularities (edges) from wavelet maxima at multiple scales. They discriminated noisy wavelet maxima by the criterion that a noise process has all singularities of negative Lipschitz exponents. Their results of denoising are shown in *Figures 7.20(d)* and *7.21(d)*. The results from using our method are shown in *Figures 7.20(c)* and *7.21(c)*. All numerical results are listed in *Table 7.5*.

Table 7.3. SNR values of noisy and denoised phantom images.

Images	Noisy	Denoised	
		1-D algorithm	2-D algorithm
NP-I	12 dB (Figure 7.10)	26.8 dB (Figure 7.11)	26.4 dB (Figure 7.12)
NP-II	6 dB (Figure 7.13)	22.0 dB (Figure 7.14)	24.1 dB (Figure 7.15)

Table 7.4. SNR values of noisy and denoised MR head images.

Images	Noisy	Denoised		
		2-D algorithm	Wiener filter	median filter
Head-I (Figure 7.16)	14.5 dB (a)	30.1 dB (b)	21.9 dB (c)	21.6 dB (d)
Head-II (Figure 7.17)	9.7 dB (a)	20.1 dB (b)	18.1 dB (c)	18.0 dB (d)

Table 7.5. SNR values of noisy and denoised photographic images.

Images	Noisy	Denoised	
		2-D algorithm	Mallat and Hwang's
House (Figure 7.20)	2.9 dB (b)	12.1 dB (c)	14.2 dB (d)
Pepper (Figure 7.21)	5.5 dB (b)	12.9 dB (c)	13.7 dB (d)

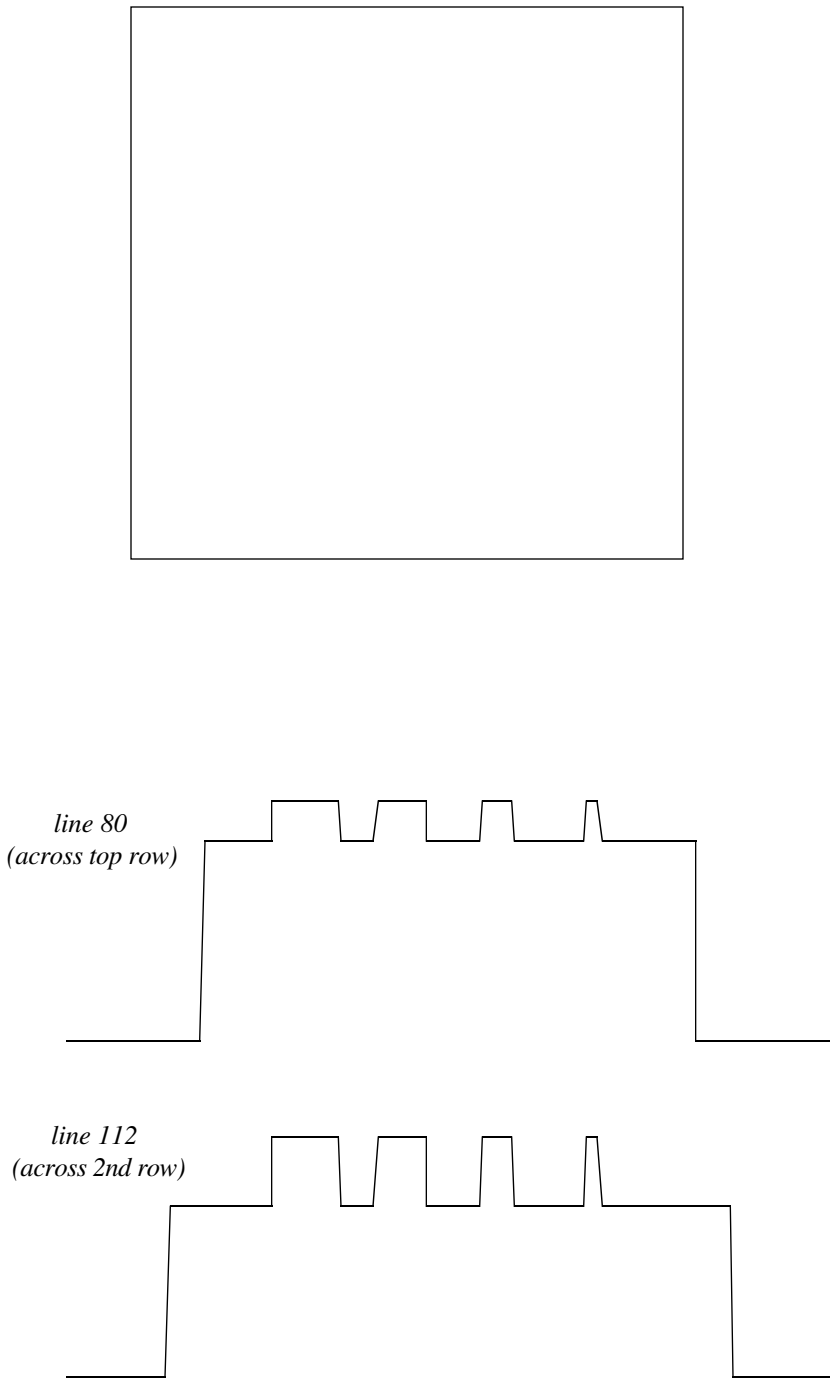


Figure 7.9. A simulated noise-free phantom image and two line profiles.

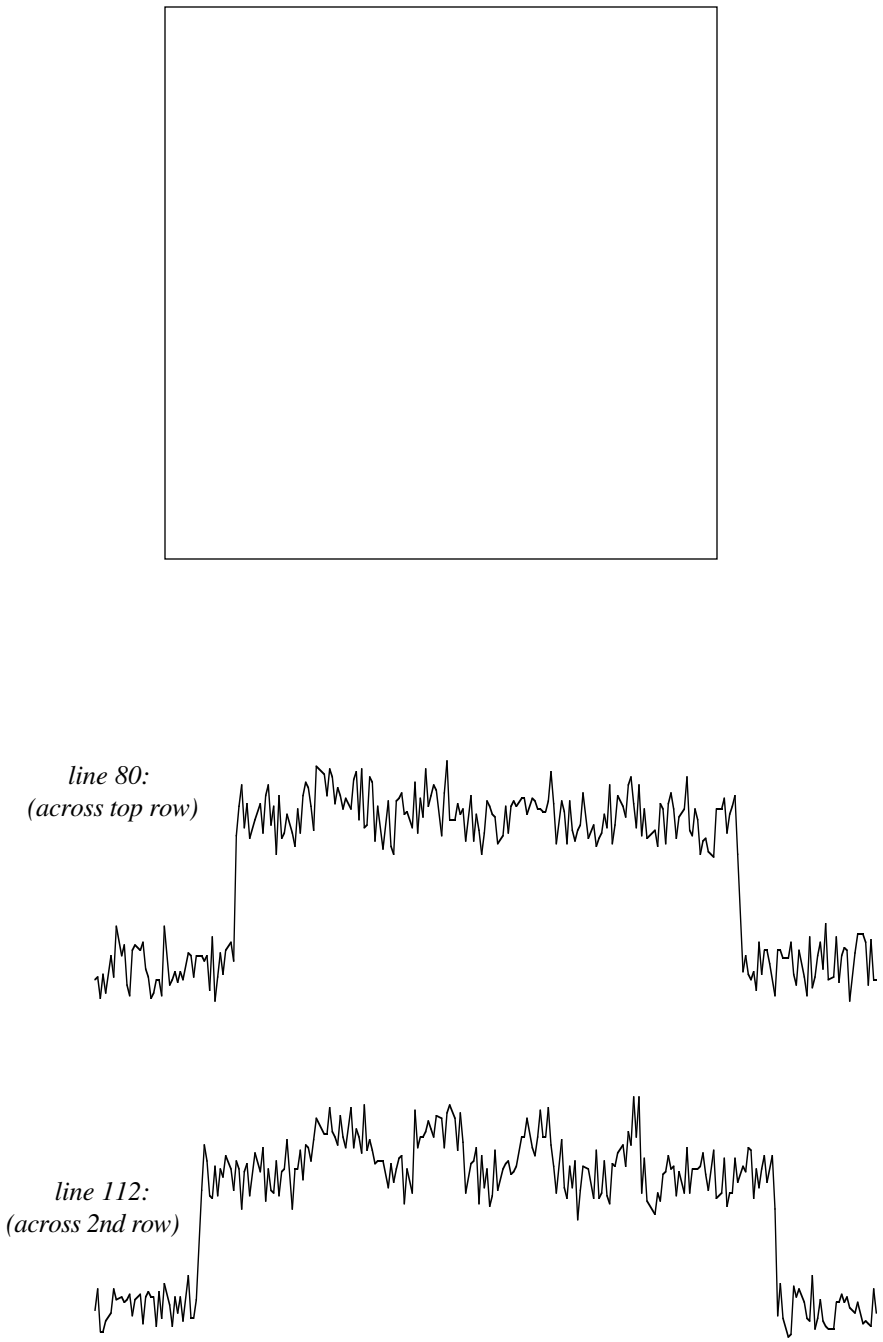


Figure 7.10. The noisy phantom image (NP-I) and its two line profiles, SNR=12 dB.

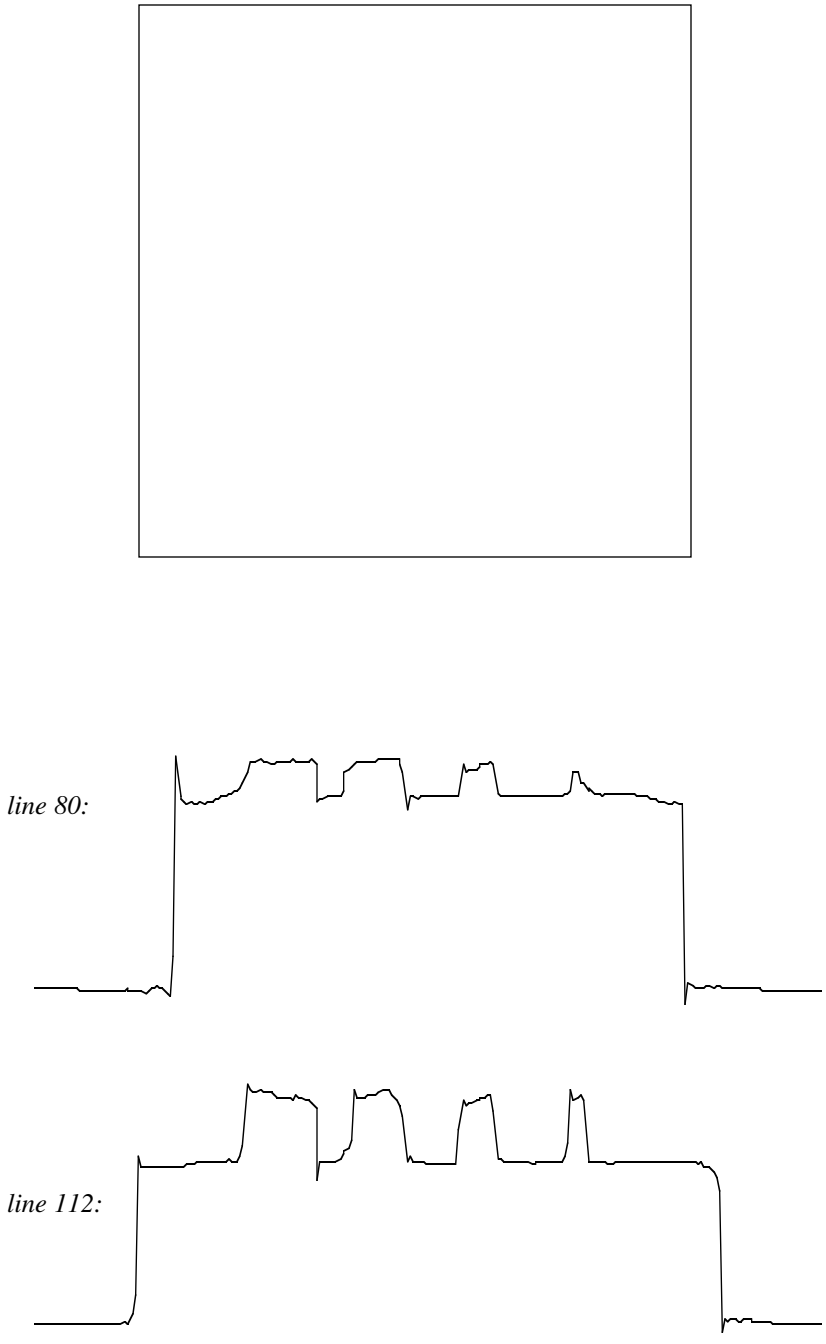


Figure 7.11. A denoised image from NP-I by using the 1-D algorithm, SNR=26.8 dB.

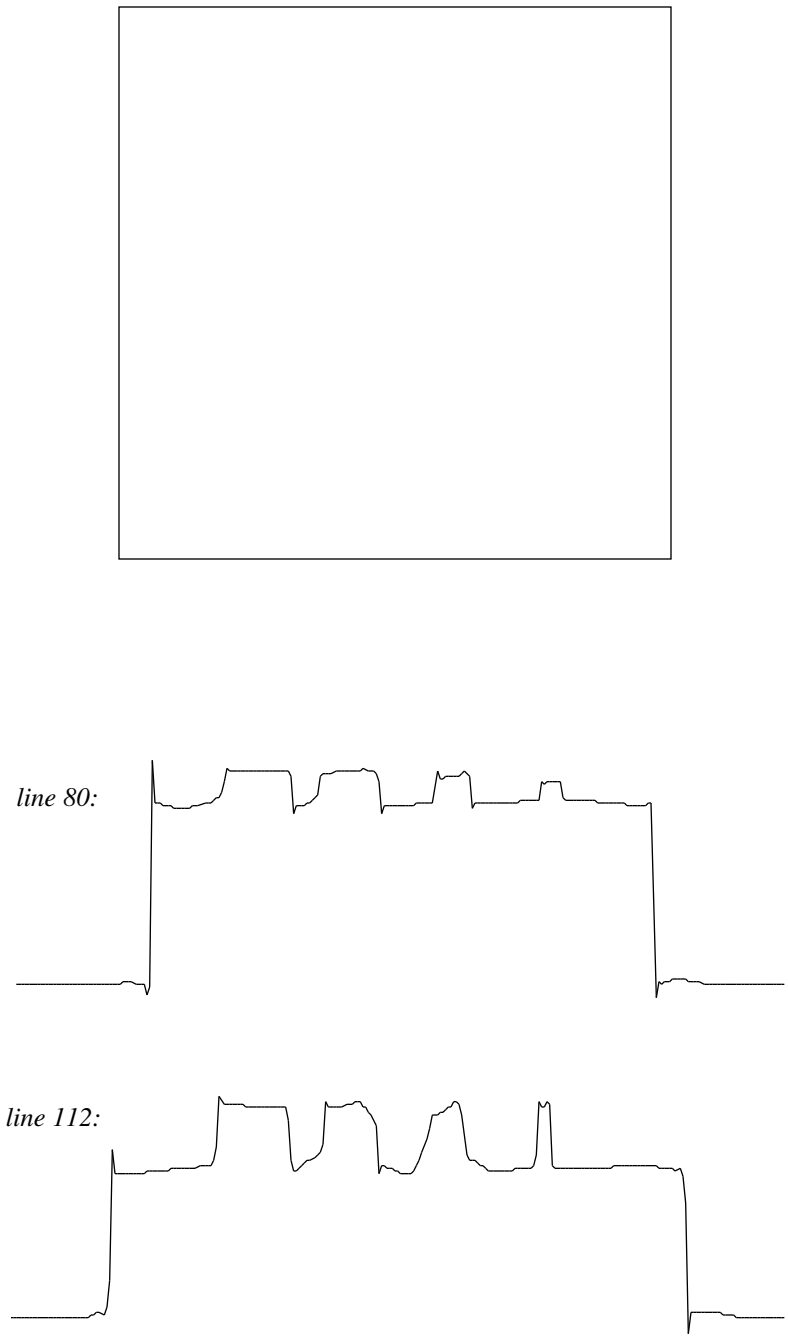


Figure 7.12. A denoised image from NP-I by using the 2-D algorithm, SNR=26.4 dB.

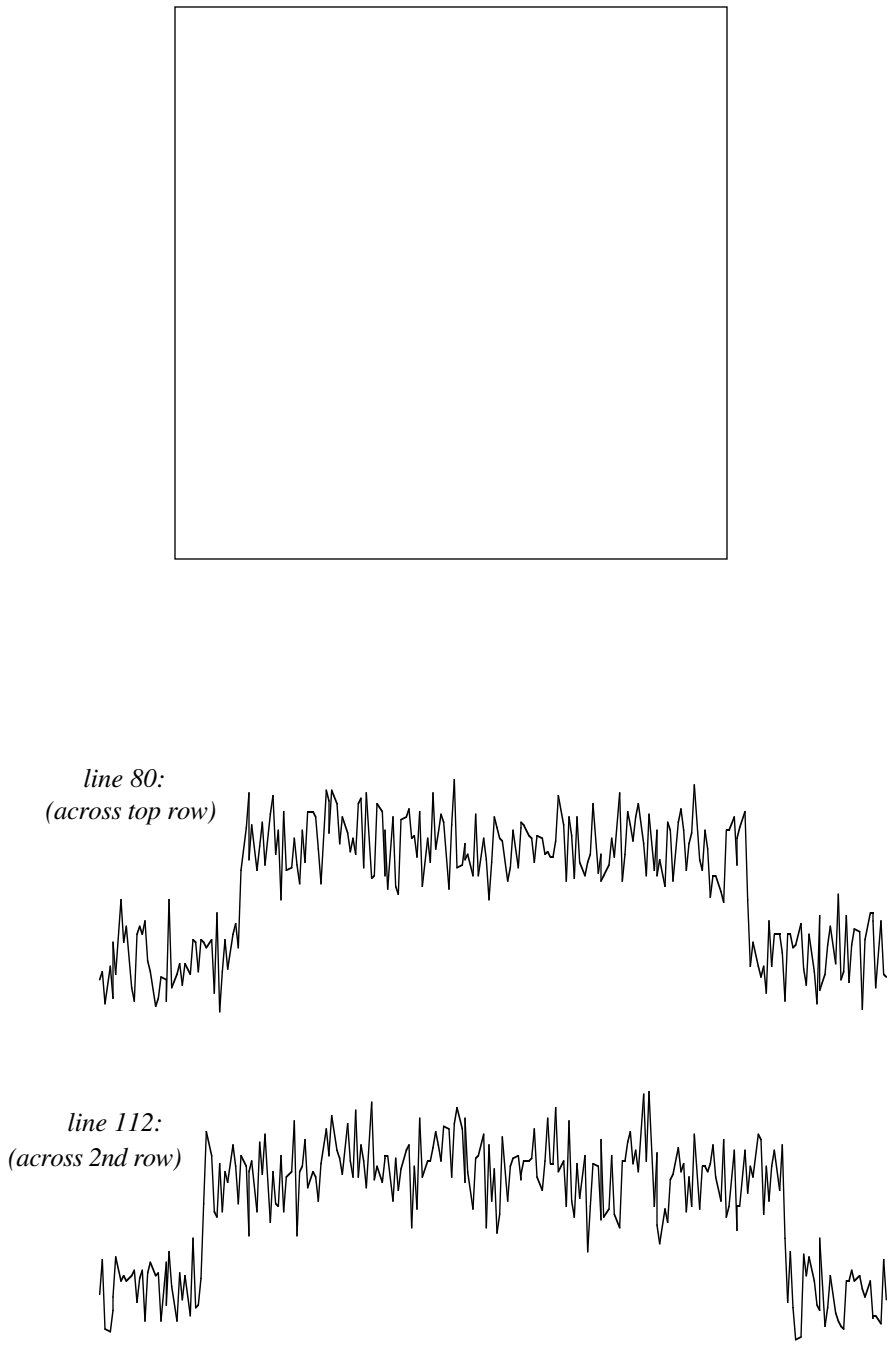


Figure 7.13. The noisy phantom image (NP-II) and its two line profiles, SNR=6 dB.

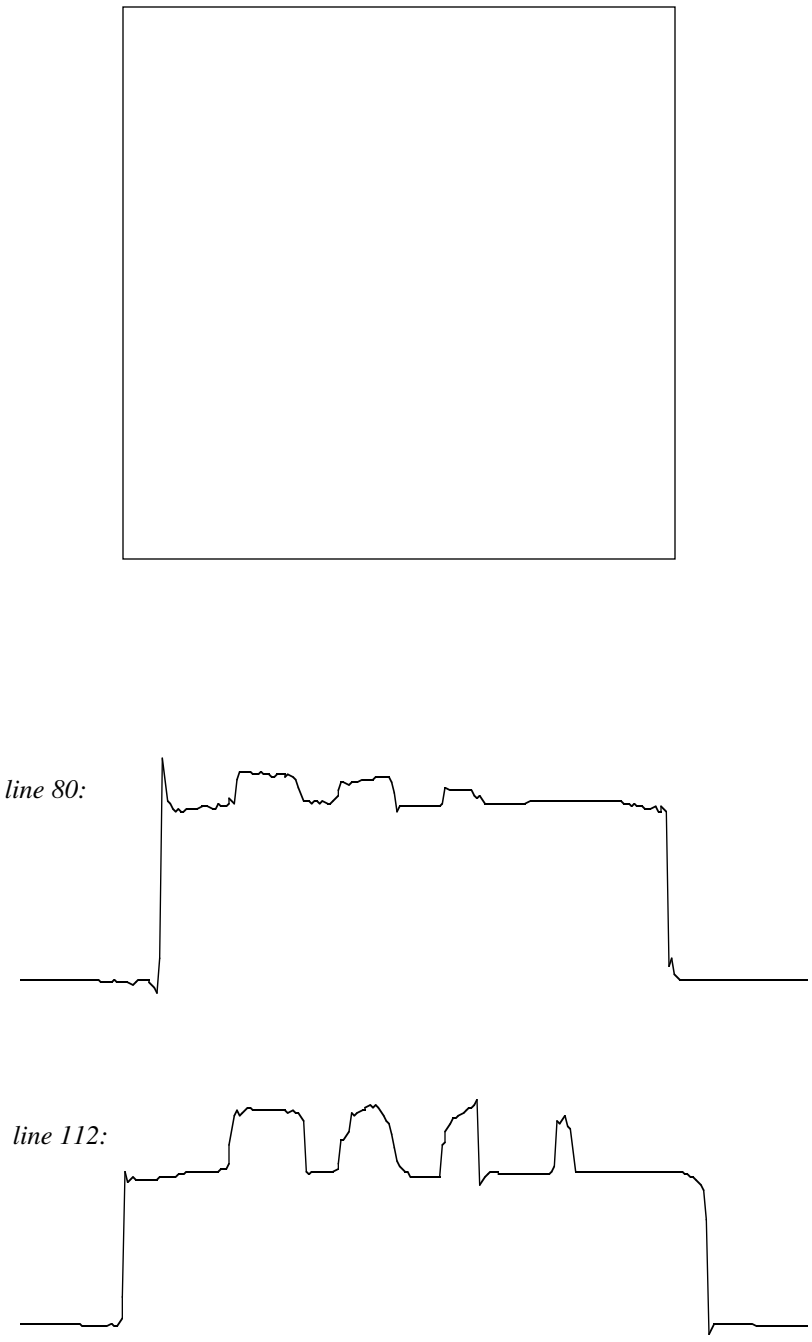


Figure 7.14. A denoised image from NP-II by using the 1-D algorithm, SNR=22 dB.

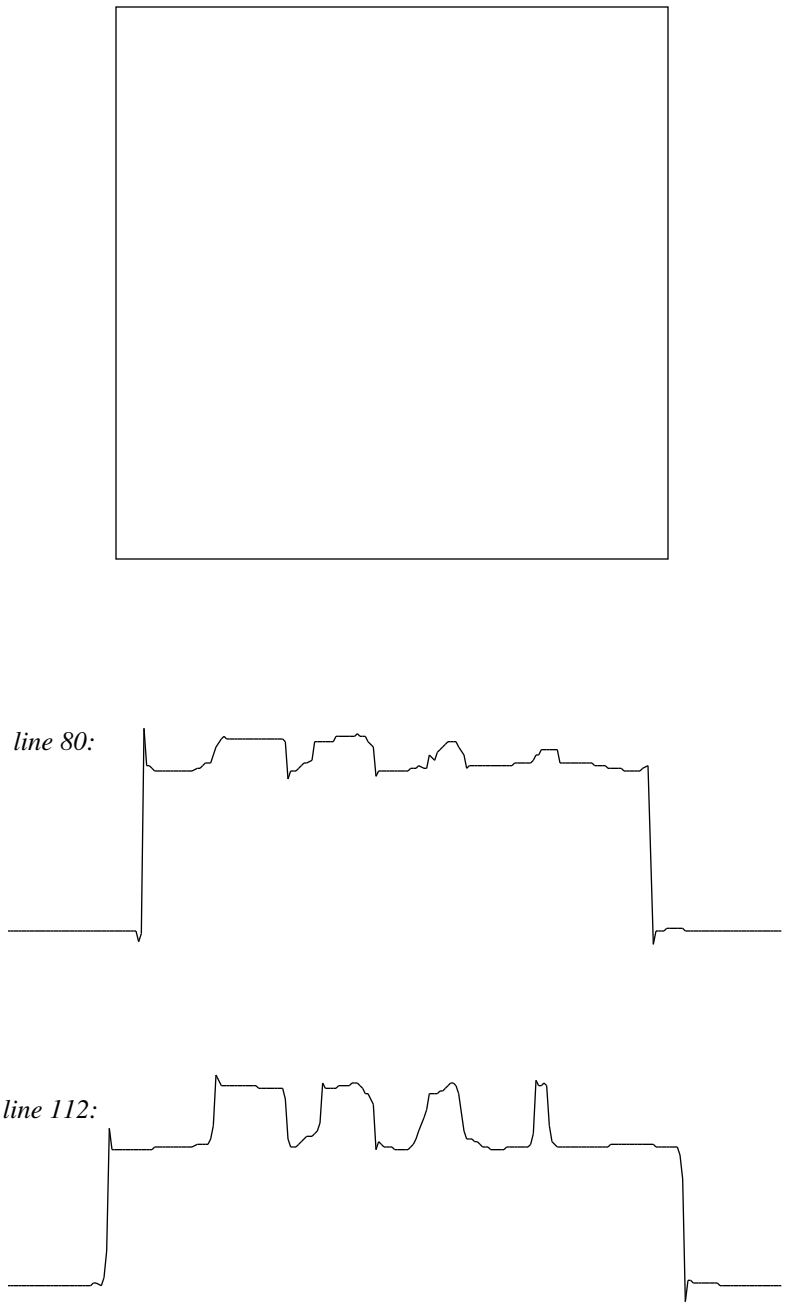
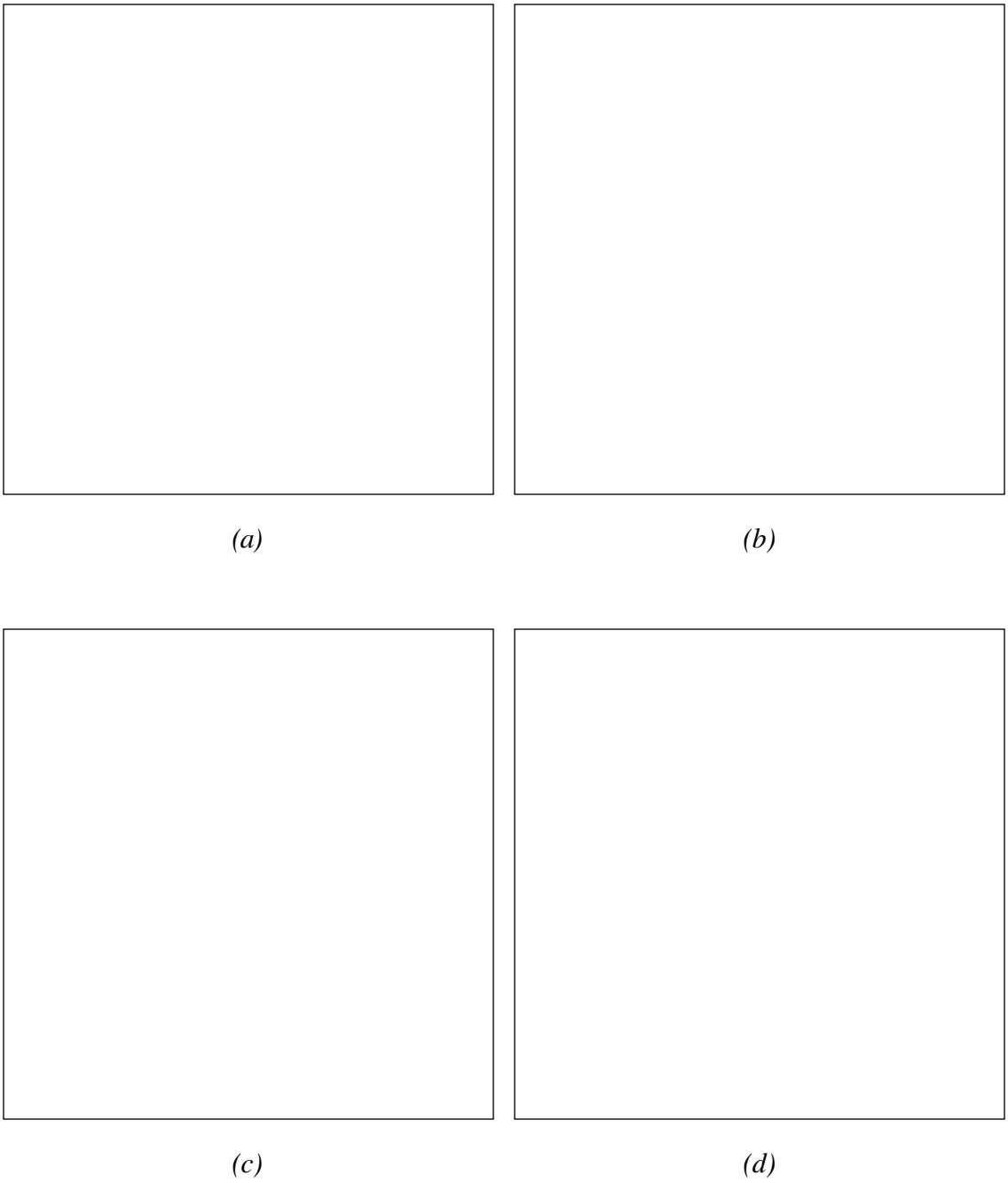


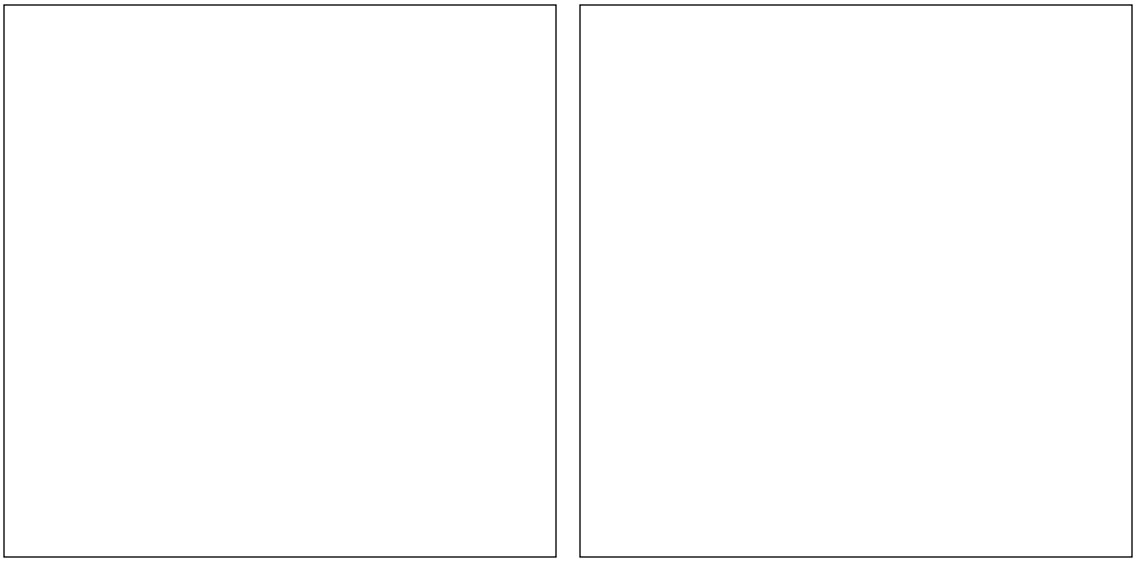
Figure 7.15. A denoised image from NP-II by using the 2-D algorithm, SNR=24.1 dB.



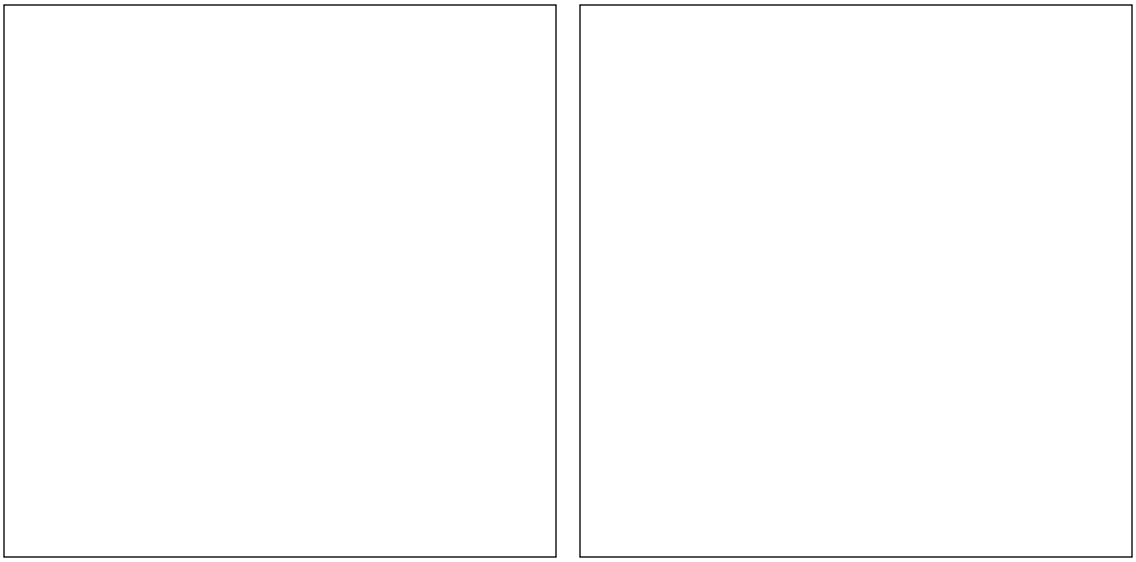
(a) *(b)*

(c) *(d)*

Figure 7.16. The noisy and denoised MR head images (Head-I). *(a)* the noisy Head-I; *(b)* denoised by the method developed in this thesis; *(c)* smoothed by a Wiener filter; *(d)* smoothed by a median filter. (See *Table 7.4* for numerical results).

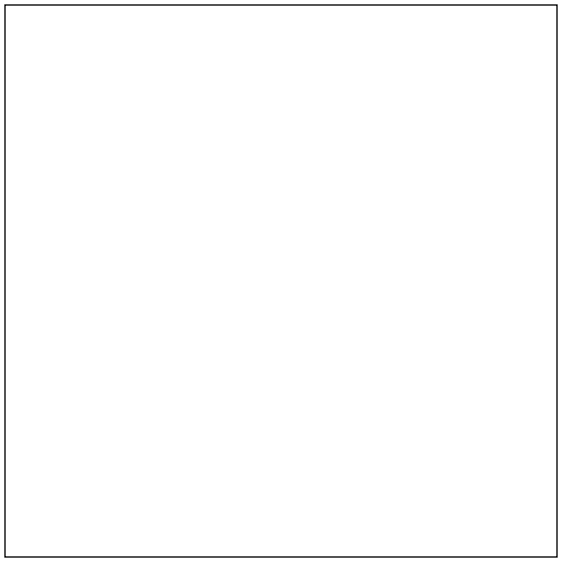


(a) *(b)*

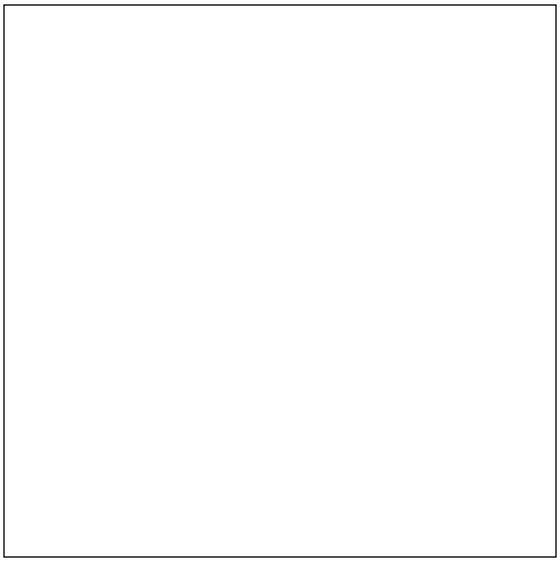


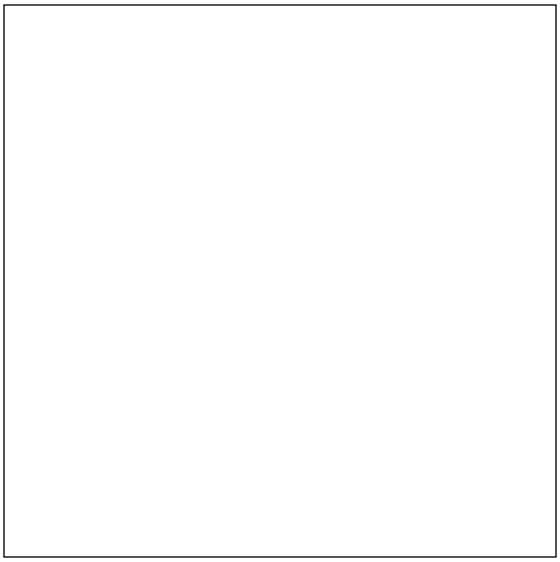
(c) *(d)*

Figure 7.17. The noisy and denoised MR head images (Head-II). *(a)* the noisy Head-II; *(b)* denoised by the method developed in this thesis; *(c)* smoothed by a Wiener filter; *(d)* smoothed by a median filter. (See *Table 7.4* for numerical results).



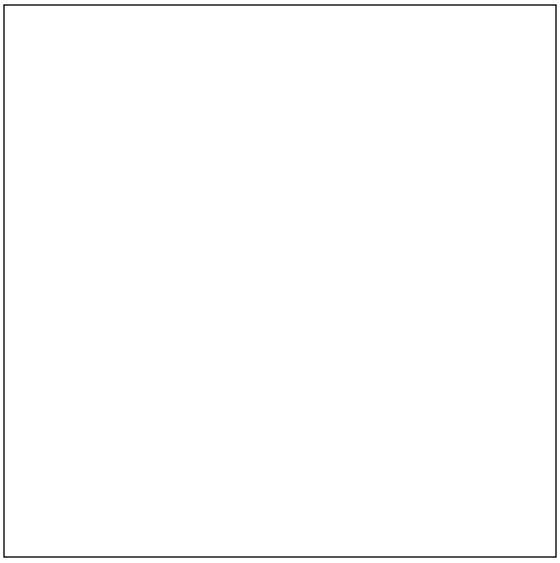
(a)



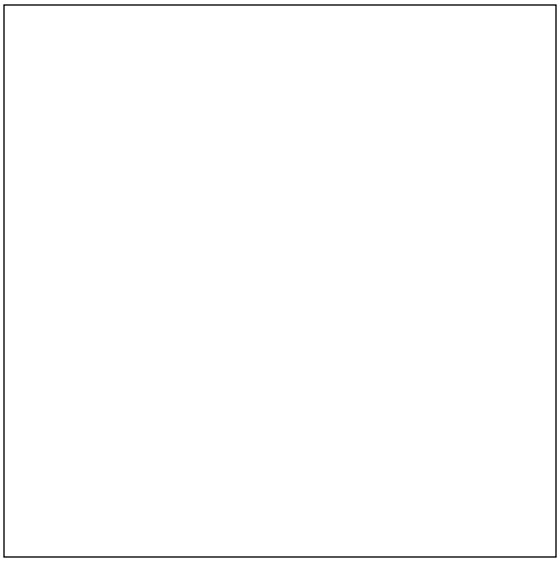


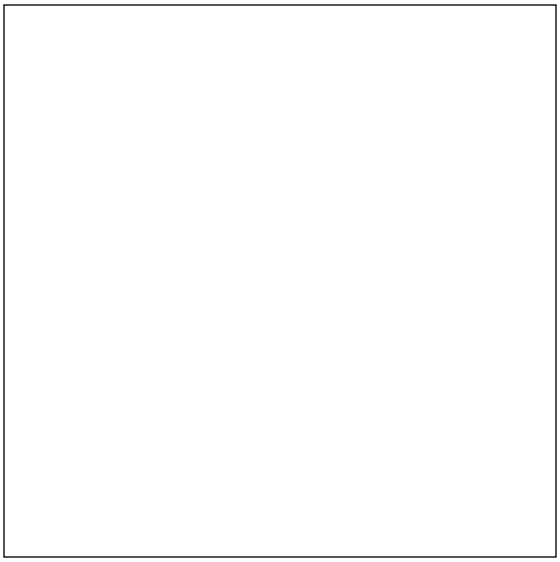
(b) *(c)*

Figure 7.18. Difference images between the noisy Head-I and denoised images. *(a)* wavelet denoised; *(b)* Wiener filtered; *(c)* median filtered.



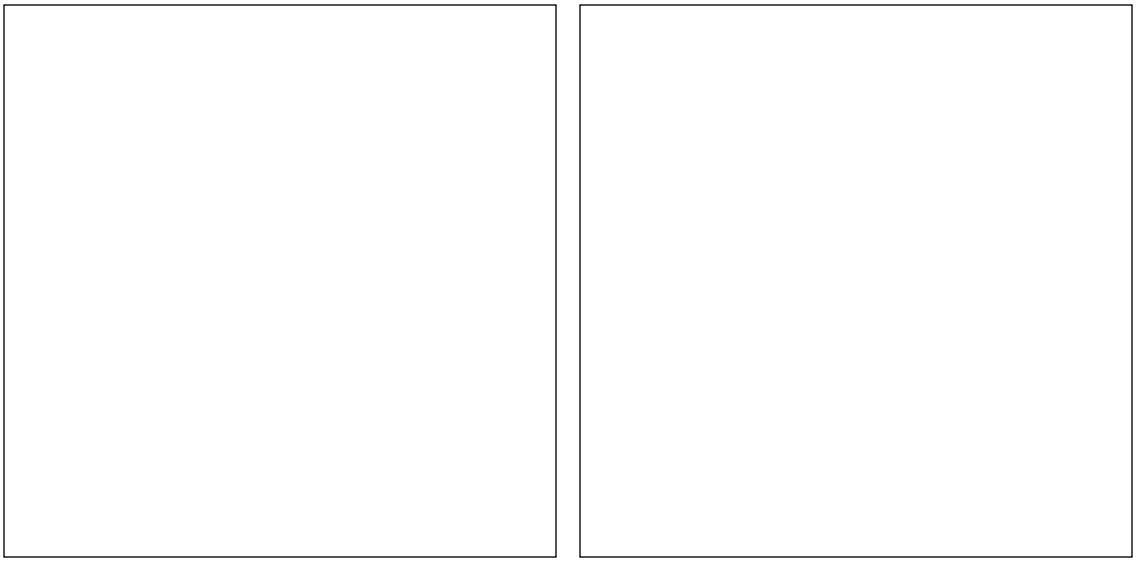
(a)



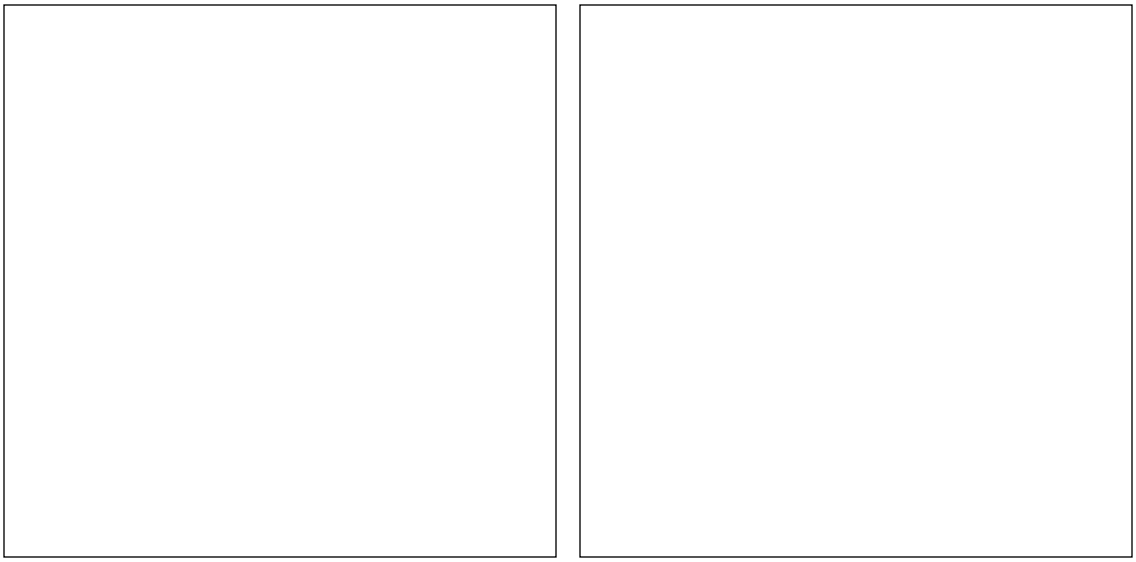


(b) *(c)*

Figure 7.19. Difference images between the noisy Head-II and denoised images. *(a)* wavelet denoised; *(b)* Wiener filtered; *(c)* median filtered.

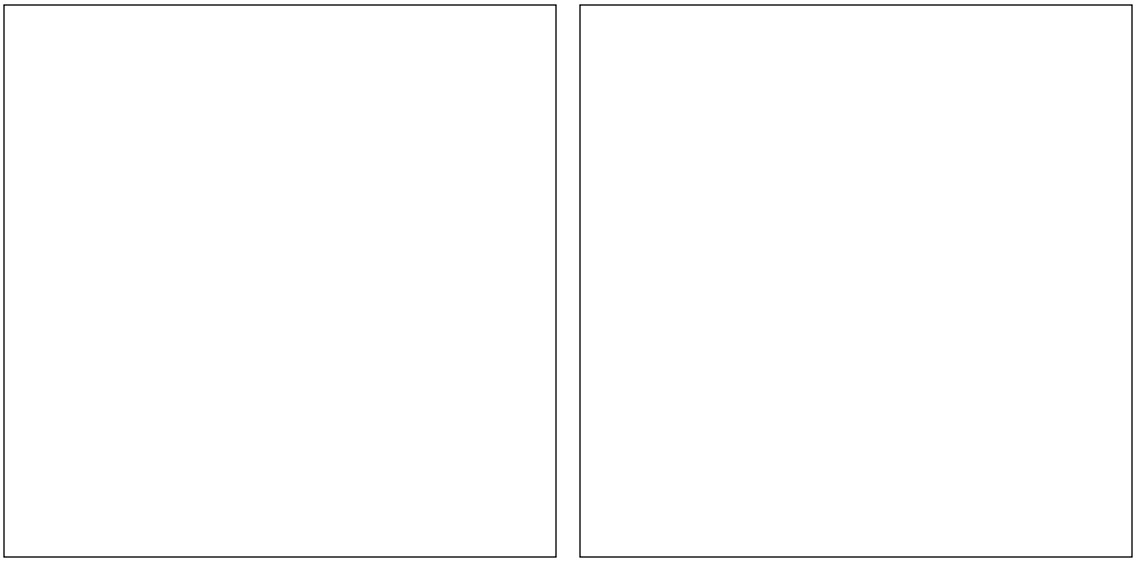


(a) *(b)*

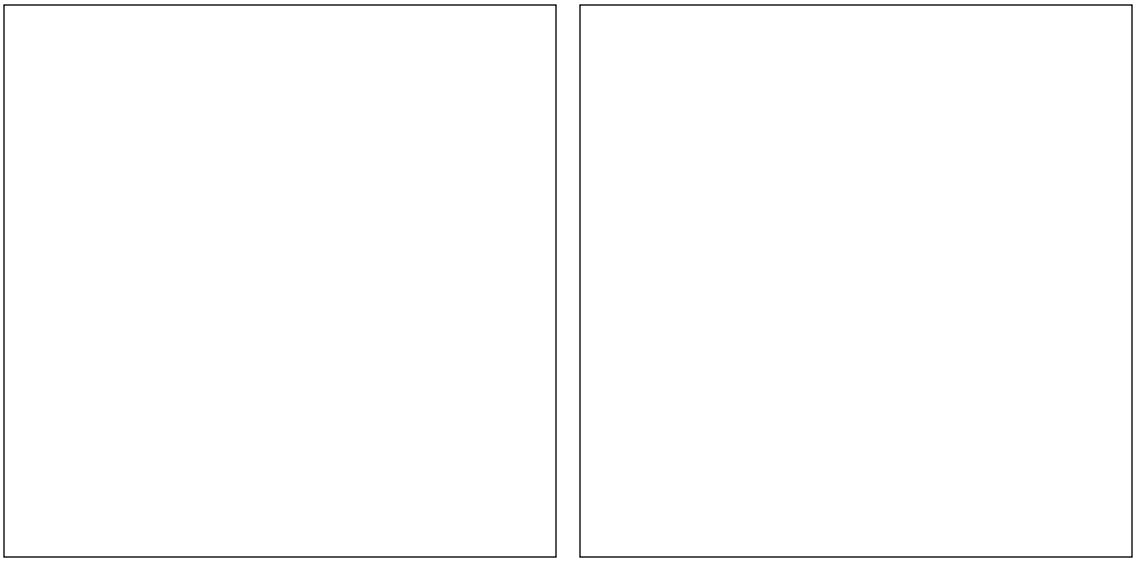


(c) *(d)*

Figure 7.20. The noisy and denoised House images. *(a)* the original house image; *(b)* the noisy House image; *(c)* denoised by the method developed in this thesis; *(d)* denoised by Mallat and Hwang's method. (See *Table 7.5* for numerical results).



(a) *(b)*



(c) *(d)*

Figure 7.21. The noisy and denoised Pepper images. *(a)* the original Pepper image; *(b)* the noisy Pepper image; *(c)* denoised by the method developed in this thesis; *(d)* denoised by Mallat and Hwang's method. (See *Table 7.5* for numerical results).

7.3 Comparative Analysis of Experimental Results

7.3.1 Comparison of One- and Two-Dimensional Algorithms

Both 1-D and 2-D algorithms appear excellent in boosting SNRs and preserving the sharpness of edges. For all noise reduction experiments in §7.2, the noise power is reduced by more than 90% and the gradients at edges, as we measured, are 80%-120% of their original values.

Some overshoots and undershoots appear at major edges of denoised signals. Examples can be found for both 1-D and 2-D algorithms; e.g. *Figures 7.6* and *7.15*. For edge enhancement, this phenomenon may be desired; it creates an effect similar to "unsharp masking" which is used for sharpening edges[31]. However, we should understand the underlying causes for this phenomenon so that it can be minimized when we desire an accurate recovery of the signal. At present, we understand two possible causes for this phenomenon. First, when we trim the noisy paths from the wavelet maxima tree, we do not deduct the influence of these noisy events on the signal. As scale increases, the energy of the noisy maxima diffuses to the major edges by merging. If the noisy maxima are trimmed while their diffused energy is left undeducted, this amount of energy will be forced into major edges by the reconstruction process. We believe that this is the major cause of the overshoot or undershoot. Second, even if the diffused energy from noise is eliminated, the overshoot may still happen as a phenomenon similar to Gibbs ringing near step edges. Fortunately, unlike Gibbs ringing, these overshoots are very localized because the wavelet used has a local support.

We mentioned in §7.1.2 the vulnerability of the 1-D algorithm to low CNR. This can be observed through *Figures 7.11* and *7.14*. In both images, the top row of dots has the lowest CNR compared with the rows under it. In *Figure 7.11*, all four dots in the top row are picked out with some distortions after denoising. In *Figure 7.14* where the CNR is

even lower, three dots are picked out but the smallest one is totally lost and the image appears more distorted. Looking at the line profiles across the top rows in *Figures 7.11* and *7.14*, one can hardly identify "meaningful features" from the noise background perceptually without any prior knowledge about the signal. In the P_b histogram of these line profiles, the wavelet maxima of the signal join the cluster of noisy maxima because they have similar P_b values. It is therefore difficult to do denoising by setting a threshold on P_b values.

The situation is improved with the 2-D algorithm because of the use of the contour metric P_c. In *Figure 7.15*, the smallest dot in the top row, though distorted, was picked out from severe noise by the 2-D algorithm. Perceptually, we see that the dots are more visible in images than in line profiles. This is because the HVS uses perceptual grouping across positions in viewing images. It is therefore not surprising that the contour metric P_c could help preserve weak features at low CNR.

7.3.2 Comparison with Wiener and Median Filters

A simple lowpass filtering will severely blur edges. This can be comprehended by visual inspection of a Gaussian lowpass filtered signal in *Figure 7.4* and by examination of *Table 7.2* for edge gradient changes in the filtered signal. In practice, Wiener and median filters are often used to replace the simple lowpass filter.

The Wiener filter is designed to adapt to SNR. It opens to pass more signal at high SNR but tends to close at low SNR. Because noise usually concentrates in the high frequency band, resulting in low SNR in that band, the high frequency components of the signal are often severely attenuated in Wiener filtering. Consequently, edges which also are rich in high frequency components are likely to be damaged. The median filter, on the other hand, is not adapted to SNR. It is considered a popular edge-preserving smoothing filter. However, the choice for the size of the window in the median filter presents a

tradeoff between noise smoothing and edge-preserving. For a too small window, the noise will not be smoothed effectively; for a too large window, edges will be blurred [32].

Figures 7.16-7.19 can be used to compare the Wiener and median filters and the method developed in this thesis. In *Figures 7.16* and *7.17*, direct visual assessment can be made. In these images, the new method results in sharper edges than the Wiener and median filters, with a similar or higher gain in SNR. This observation is confirmed by examining the difference images between the noisy and denoised images. From *Figures 7.18* and *7.19*, we can see that the pixel difference is high (bright) around edges in the difference images between the noisy and Wiener and median filtered ones. On the other hand, in the difference image between the noisy and wavelet denoised images, the pixel difference is uniform or even small (dark) at edges, suggesting that the edges are preserved well or enhanced.

These results show that the new method is superior to Wiener and median filters in preserving edges for the same SNR improvement. This superior performance comes from the wavelet maxima representation and the reconstruction mechanism. When we reconstruct an image from its wavelet maxima tree, we actually smooth the interior regions between two wavelet maxima contours or edges, but no smoothing happens *on* these edges. This kind of smoothing has been investigated by Perona and Malik[62] with the diffusion equation. They called it "anisotropic diffusion".

7.3.3 Comparison with Mallat and Hwang's Method

Donoho[25] noted that the success of wavelet-based denoising methods is due to (1) the use of wavelet basis, and (2) a non-linear treatment, particularly thresholding, in the wavelet-transform domain. From this point of view, our method and Mallat and Hwang's share the same advantages. Furthermore, both methods use wavelet maxima as primitives in the representation of signals. This makes the two methods even closer to each other and draws a distinction from others such as Donoho's[25]. The difference between ours and

Mallat and Hwang's lies in the way of analyzing the wavelet maxima and singling out those believed to be from noise. We use perception-based metrics on a wavelet maxima tree while Mallat and Hwang compute Lipschitz regularity of singularities (edges) from wavelet maxima at several scales.

By showing[53] that the Lipschitz exponents of singularities are always non-negative in signal and negative in noise, Mallat and Hwang established a mathematically rigorous criterion to discriminate signal from noise. In practice, however, the estimate of Lipschitz exponents of the signal is very much affected by the strong presence of noise at small scales. In [53], Mallat and Hwang had to exclude the wavelet maxima at the smallest scale from their estimation of Lipschitz exponents. Unfortunately, these wavelet maxima are very informative about edge locations and small features in the image. In our method, each wavelet maximum at the smallest scale belongs to a contour that is associated with another contour at the coarser scale by the tree hierarchy. The decision about keeping or dismissing a single wavelet maximum is based on the extent of its contour in the scale-space. Thus we exploit the scale-spatial structure of a wavelet maximum against noise disturbance.

From *Figures 7.20* and *7.21*, and *Table 7.5*, we see that the two methods achieved similar denoised images. In [53], Mallat and Hwang used a special way to compute SNR of denoised images. The "noise" in their computation is the difference between the original and denoised images[86]. This is not a conventional measure of smoothness because the "noise" so computed does not have a random pattern. It is more like the MSE that measures the quality of signal recovery. However, we comply with this measure in *Table 7.5* for the convenience of comparison. It seems that the two images denoised by Mallat and Hwang's method are over-smoothed. This may result from, as we guess, a non-optimum threshold and/or the exclusion of all wavelet maxima from the finest scale in the selection process. On the other hand, our method gave less smooth but less distorted images. In addition, it appears that our method did a little better in preserving small

features in the image. For example, in the House image, a few small features on the roof and wall of the house are lost or severely damaged in Mallat and Hwang's denoised image but quite visible in ours.

Part IV

CONCLUSIONS

Chapter 8
Conclusions and Future Directions

8.1 Conclusions

Signal recovery and noise reduction are typical signal processing problems of both theoretical and practical interest. We studied these problems in their general forms (1.2) and (1.3) with a new tool—wavelets. Both theoretical and experimental results in this thesis suggest that wavelets provide new perspectives and better solutions to the above problems which have been studied extensively with other methods.

In Part II of the thesis, we reformulated the signal recovery problem (1.2) as finding missing data at the finest scale of a discrete wavelet transform. In such a framework, the wavelet reproducing equation plays a fundamental role in determining a unique and stable recovery of the signal. Interestingly, the solution to the wavelet reproducing equation was found to be equivalent to a CLS solution. The corresponding CLS functional has a "mid-band" consistency term in place of the "low-band" consistency term of the usual cost functional in regularization. By linking the wavelet reproducing equation to the CLS minimization, we were able to derive the wavelet-regularized recovery of the signal in one and two dimensions. Additionally, we obtained a new family of regularizing operators that are adapted to the specific degrading operator in the signal recovery problem.

Experimental results show that the wavelet-regularized image restoration achieves 75-85% reduction in mean square errors in noise-free cases, and 35-60% reduction in noisy cases. These results are as good as Laplacian-regularized solutions. However, as with the Laplacian operator, the quality of deblurring with a wavelet regularizing operator is very dependent on SNR. Furthermore, our model of wavelet-regularized signal recovery does not seem as useful as standard regularization if one's task is simply to smooth noise.

We provided an in-depth treatment of the noise reduction problem (1.3) in Part III. We adopted and expanded the wavelet maxima representation developed by Mallat and Zhong. A key to our work is the development of the concept of the wavelet maxima tree, and algorithms for the tree construction. We then designed tree metrics that reflect some perceptual criteria used in discriminating signals from a noisy background. A complete technique for noise reduction using the wavelet maxima tree and perception-based metrics was presented and implemented in one and two dimensions.

We conducted a number of noise reduction experiments by using the new technique with 1-D signals and 2-D images. The new technique appears excellent in improving SNR and preserving edges. In all our experiments, noise power was reduced by more than 90% while the gradients at edges were 80% to 120% of their original values. In the comparative experiments, the new technique outperformed Wiener and median filters in preserving edges for the same or higher gain in SNR improvement. However, the new technique seems to be vulnerable to CNR, especially when it is implemented in one-dimension. Consequently, small features in the signal which are similar in size and strength to noise elements are likely to be removed as noise by the denoising algorithms. While this is a drawback, we observed that the resulting damage to the signal was local because of the representation we chose.

8.2 Directions of Future Work

Multiscale Wiener Filter?

We believe that the wavelet regularizing operator, as a deterministic regularizing operator, has achieved the best possible result regarding deblurring. For even better results, regularization must take into account the information contained in observations. The classic Wiener filter provides such an example at a fixed scale. It is natural to think of

designing a multiscale Wiener filter. Such a filter is expected to outperform the classic Wiener filter in preserving edges and more generally, signal resolution.

The classic Wiener filter assumes the second-order statistics of the signal and noise are known. In the Fourier domain, this assumption is equivalent to having the power spectra of the signal and noise, provided that both signal and noise are wide-sense stationary. In practice, the exact characteristics of the signal and noise are often unknown, and hence have to be estimated from an observed noisy signal. The quality of these estimates greatly affects the outcome of Wiener filtering. Furthermore, even if the power spectra of signal and noise can be estimated precisely, they lack temporal or spatial localization. Because of this a Wiener filter will attenuate high frequency components regardless of whether they originate from noise or edges.

Progress has already been made in spectral estimation with wavelets[29]. With wavelets, not only spectra can be estimated more accurately, but also time-frequency localization can be achieved. It appears possible to design a multiscale Wiener filter on the time-frequency or space-frequency plane. We hope to design a filter to treat the high frequency components differently based on their time/spatial positions so that edges are preserved.

Statistical Properties of Wavelet Maxima

As we know, the wavelet maxima representation has been applied to coding and denoising. Understandings of the statistical properties of wavelet maxima is very important to the success and efficiency in both applications. For the coding application, it can help us design an optimal quantizer; for the denoising application, it can guide threshold selections. Unfortunately, there has not been a complete study on the statistical properties of wavelet maxima of noise, and of a wide class of signals.

In a preliminary study, we estimated at each scale the mean and variance of wavelet maxima amplitudes of Gaussian white noise. It appears that both the mean and the variance

are linear functions of the variance of the Gaussian white noise process, and they are inversely proportional to the scale parameter. These linear functions, if determined, can certainly help us design the threshold for the wavelet maxima with the knowledge of noise variance.

We also hope that the study of statistical properties of wavelet maxima could lead to improvement of the current denoising algorithms at low CNRs. We found that in many cases where the algorithms fail to pick up weak signals in the low CNR regions, it is also hard for humans having no special training to recognize the desired features. This suggests that the discriminating power of perceptual criteria is limited. A statistical analysis will be helpful in uncovering complicated signal and noise characteristics that can not be seen directly though the eye.

Better Way to Construct a Wavelet Maxima Tree?

The hierarchy of wavelet maxima is determined by the wavelet maxima tree. This hierarchy affects the analysis and pruning of wavelet maxima, and hence the quality of denoising and texture preservation. We have mentioned that the sampling of wavelet maxima traces at dyadic scales introduces uncertainties about the relationship between these wavelet maxima.

Currently, we use a measure of interaction between wavelet maxima at neighboring scales with wavelet reproducing kernels and link the two wavelet maxima that maximize such a measure. We find that this method of tree construction works well in general, but may lead to erroneous connections in the regions where wavelet maxima are very dense. More sophisticated methods such as statistical decision-making can be used to refine the tree construction in these cases. For example, one could associate a probability measure to each direction of tree growing in the coarse-to-fine tracking. The tree construction then becomes an optimal estimation problem that can be combined with tree pruning.

Performance Sensitivity to Parameters

The denoising algorithm has a number of parameters that could affect the outcome of the algorithm. Besides the threshold for the branch metric that was discussed in §7.1.2, there are three parameters in Algorithms 6.1 and 6.2 relating to the tree construction. One parameter is associated with the tracking window, and the other two are with the measures of geometric compatibility between planar curves. In our implementation of the denoising algorithm, we took a heuristic approach to selecting "reasonable" parameters. For example, if a single curve is observed at two neighboring scales, the planar distance between the two curves observed at the two scales is unlikely to be greater than the average length of the two curve. This led us to set $\gamma = 0.5$ in (6.9) for Algorithm 6.2.

So far we have some empirical knowledge about the sensitivity of denoising algorithm to the above parameters from a small number of experiments. The sensitivity of the threshold for the branch metric P_b has been discussed in §7.1.2. For parameters in the tree construction algorithm, in general, large features in a signal or image are invariant to the changes in these parameters. Small features, seen as short and/or weak branches in the wavelet maxima tree, are relatively sensitive. Changing these parameters affects mostly the structures of those small branches in the tree. If those small tree branches were all from noise, we would not care how they structured. Unfortunately, there are some desired weak signals embedded in these small tree branches. Therefore, parameter sensitivity presents some uncertainty about the algorithm's ability to extract weak signals from noise. The situation becomes worse at low CNRs.

It is necessary to test systematically the program's performance sensitivity against the changing parameters. To do this, we need a set of carefully designed test signals with different features and controlled CNR variations. By counting the hit and miss of small features with different parameter settings, we shall be able to get a better understanding

about how those parameters affect the preservation of small features, and hopefully an optimum setting.

References

[1] Proceedings of 1991 International Conference on Acoustics, Speech ,and Signal Processing, Toronto, May 1991.

[2] Proceedings of 1992 International Conference on Acoustics, Speech ,and Signal Processing, San Francisco, March 1992.

[3] Advance Programs of 1993 International Conference on Acoustics, Speech ,and Signal Processing, Minneapolis, April 1993.

[4] 1992 MIT Spring Wavelet Seminar Series, C. Heil, Organizer, Department of Mathematics, Massachusetts Institute of Technology.

[5] H.C. Andrews and B. R. Hunt, *Digital Image Restoration*, Englewood Cliffs, NJ: Prentice-Hall, 1977.

[6] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using vector quantization in the wavelet transform domain," *IEEE Trans. Image Processing*, vol. 1, pp. 205-220, April 1992.

[7] J. Babaud, A.P. Witkin, M. Baudin and R.O. Duda, "Uniqueness of the Gaussian kernel for scale-space filtering," *IEEE Trans. Pattern Anal. and Machine Intell.,* vol. PAMI-8, pp. 26-33, Jan. 1986.

[8] M. Basseville, A. Benveniste, and A.S. Willsky, "Multiscale autoregressive processes, Part I: Schur-Levinson parametrizations," *IEEE Trans. Signal Processing*, vol. 40, pp. 1915-1934, Aug. 1992.

[9] M. Basseville, A. Benveniste, and A.S. Willsky, "Multiscale autoregressive processes, Part II: Lattice structures for whitening and modeling," *IEEE Trans. Signal Processing*, vol. 40, pp. 1935-1954, Aug. 1992.

[10] G. Beylkin, R. Coifman and V. Rokhlin, "Fast wavelet transforms and numerical algorithms I," *Comm. Pure Appl. Math.*, vol. 44, pp.141-183, 1991.

[11] F. Bergholm, "Edge focusing," *IEEE Trans. Pattern Anal. and Machine Intell.,* vol. PAMI-9, pp. 726-741, Nov. 1987.

[12] Z. Berman, "The uniqueness question of discrete wavelet maxima representation," Technical Report, TR 91-48, System Research Center, University of Maryland, College Park, April 1991.

[13] Z. Berman, "Generalizations and properties of the multiscale maxima and zero-crossings representations," Ph.D. Thesis, University of Maryland, College Park, Sept. 1992.

[14] W.F. Bischof and T. Caelli, "Parsing scale-space and spatial stability analysis," *Computer Vision, Graphics, and Image Processing*, vol. 42, pp. 192-205, 1988.

[15] J.M. Bruneau, M. Barlaud and P. Mathieu, "Image restoration using biorthogonal wavelet transform," in *Visual Communications and Image Processing: Proceedings of SPIE*, vol. 1360, pp. 1404-1415, 1990.

[16] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. and Machine Intell.,* vol. PAMI-8, pp. 679-698, Nov. 1986.

[17] K.R. Castleman, *Digital Image Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1979.

[18] A. Cohen, I. Daubechies and J. C. Feauveau, "Biorthogonal bases of compactly supported wavelets," *preprint*, AT&T Bell Laboratories, 1990.

[19] I. Daubechies, "Orthogonal bases of compactly supported wavelets," *Comm. Pure Appl. Math.*, vol. 41, pp. 909-996, 1988.

[20] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Inform. Theory*, vol. IT-36, pp. 961-1005, Sept. 1990.

[21] I. Daubechies, *Ten Lectures on Wavelets*, Philadelphia: SIAM, 1992.

[22] L. Debnath and P. Mikusinski, *Introduction to Hilbert Spaces with Applications*, San Diego, CA: Academic Press, Inc., 1990.

[23] R.A. DeVore, B. Jawerth, and B.J. Lucier, "Image compression through wavelet transform coding," *IEEE Trans. Inform. Theory*, vol. IT-38, No. 2, pp. 719-746, March, 1992.

[24] D.L. Donoho, "Non-linear solution of linear inverse problems by wavelet-vaguelette decomposition," Technical Report No. 403, Dept. of Statistics, Stanford University, July 1992.

[25] D.L. Donoho, "Denoising by soft-thresholding," Technical Report No. 409, Dept. of Statistics, Stanford University, Dec. 1992.

[26] S. Even, *Graph Algorithms*, Potomac, MD: Computer Science Press, 1979.

[27] A. Grossmann and J. Morlet, "Decomposition of Hardy functions into square integrable wavelets of constant shape," *SIAM J. Math*., vol. 15, pp. 723-736, 1984.

[28] D.M. Healy, Jr., and J.B. Weaver, "Two applications of wavelet transform in magnetic resonance imaging," *IEEE Trans. Inform. Theory*, vol. IT-38, No. 2, pp. 840-860, March, 1992.

[29] F. Hlawatsch and G.F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations," *IEEE Signal Processing Magazine*, vol. 9, No. 2, pp. 21-67, April 1992.

[30] B.R. Hunt, "The application of constrained least squares estimation to image restoration by digital computer," *IEEE Trans. Computer*, vol. C-22, pp. 805-812, Sept. 1973.

[31] A.K. Jain, *Fundamentals of Digital Image Processing*, Englewood Cliffs, NJ: Prentice Hall, 1989.

[32] B.I. Justusson, "Median filtering: Statistical properties," in *Two-Dimensional Digital Signal Processing II: Transforms and Median Filters*, T.S. Huang, Ed., pp. 161-196, Berlin Heidelberg: Springer-Verlag, 1981.

[33] A.K. Katsaggelos, "Iterative image restoration algorithms," *Optical Engineering*, vol. 28(7), pp. 735-748, 1989.

[34] A.K. Katsaggelos, (Ed.,) *Digital Image Restoration*, Berlin: Springer-Verlag, 1991.

[35] J.J. Koenderink, "The structure of images," *Biol. Cybern.*, vol. 50, pp. 363-370, 1984.

[36] W.M. Lawton and H.L. Resnikoff, "Multidimensional wavelet bases," Technical Report, Aware, Inc., Cambridge, MA, 1991.

[37] M.D. Levine, *Vision in Man and Machine*, New York: McGraw-Hill, 1985.

[38] J.S. Lim, *Two-Dimensional Signal and Image Processing*, Englewood Cliffs, NJ: Prentice Hall, 1990.

[39] T. Lindeberg and J. -O. Eklundh, "On the computation of a scale-space primal sketch," *J. Visual Comm. Image Proc.*, vol. 2, No. 2, pp. 55-78, 1991.

[40] D.G. Lowe, *Perceptual Organization and Visual Recognition*, Boston: Kluwer Academic Publishers, 1985.

[41] J. Lu, Y. Xu, J.B. Weaver and D.M. Healy, Jr. "Noise reduction by constrained reconstructions in the wavelet-transform domain," in *Proceedings of the 7th Multidimensional Signal Processing Workshop*, IEEE Signal Processing Society, Lake Placid, NY, Sept. 1991.

[42] J. Lu, J.B. Weaver, D.M. Healy, Jr., and Y. Xu, "MR image filtering with multiscale edge representation," in *Proceedings of the 11th Annual Meeting, Society of Magnetic Resonance in Medicine*, Berlin, Aug. 1992.

[43] J. Lu, J.B. Weaver, D.M. Healy, Jr., and Y. Xu, "Noise reduction with a multiscale edge representation and perceptual criteria," in *Proceedings of IEEE International Symposium on Time-Frequency and Time-Scale Analysis*, Victoria, B.C., Oct. 1992, pp. 555-558.

[44] J. Lu, D.M. Healy, Jr., and J.B. Weaver, "Signal recovery and wavelet reproducing kernels," in *Proceedings of IEEE International Symposium on Time-Frequency and Time-Scale Analysis*, Victoria, B.C., Oct. 1992, pp. 177-179. Also submitted to *IEEE Trans. Signal Processing*.

[45] J. Lu, "Parallelizing Mallat algorithm for 2-D wavelet transforms," *Information Processing Letters*, vol. 45, pp. 255-259, 1993.

[46] S.G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. PAMI-11, pp. 674-693, July 1989.

[47] S.G. Mallat, "Multiresolution approximations and wavelet orthogonal bases of $L^2(R)$," *Trans. Amer. Math. Soc.*, vol. 315, pp. 69-87, 1989.

[48] S.G. Mallat, "A complete signal representation from multiscale edges," Technical Report, No. 483, New York University, December, 1989.

[49] S.G. Mallat, "Multifrequency channel decomposition of images and wavelet models," *IEEE Trans Acoust. Speech and Signal Proc.*, vol. ASSP-37, pp. 2091-2110, Dec. 1989.

[50] S.G. Mallat and S. Zhong, "Characterization of signals from multiscale edges," NYU Technical Report No. 592, Nov. 1991. Also in *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. PAMI-14, pp. 710-732, July 1992.

[51] S.G. Mallat and S. Zhong, "Compact image coding from edges," in *Proc. Int. Conf. Acoust. Speech and Signal Proc.,* Toronto, May 1991.

[52] S. G. Mallat, "Zero-crossings of a wavelet transform," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1019-1033, July 1991.

[53] S.G. Mallat and W.L. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. Inform. Theory*, vol. 38, No. 2, pp. 617-643, March 1992.

[54] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. R. Soc. Lond.* vol. B-207, pp. 187-217, 1980.

[55] D. Marr and S. Ullman, "Bandpass channels, zero-crossings, and early visual information processing," *J. Opt. Soc. Am.*, vol. 69, No. 6, June 1979.

[56] D. Marr, *Vision*, San Francisco, CA: W. H. Freeman, 1982.

[57] Y. Meyer, "Un contre-example a la conjecture de Marr et a celle de S. Mallat," preprint, 1991.

[58] K. Miller, "Least-squares method for ill-posed problems with a prescribed bound," *SIAM J. Math. Anal.*, vol. 1, pp. 52-74, 1970.

[59] M.Z. Nashed, "Operator-theoretic and computational approaches to ill-posed problems with applications to antenna theory," *IEEE Trans. Antennas Propagat.*, vol. AP-29, pp.220-231, 1981.

[60] M.Z. Nashed, "Approximate regularized solutions to improperly posed linear integral and operator equations," in *Constructive and Computational Methods for Differential and Integral Equations*, D.L. Colton and R.D. Gilbert, Eds, New York: Springer, 1974, pp. 289-322.

[61] M.Z. Nashed, Ed., *Generalized Inverse and Applications*, New York: Academic, 1976.

[62] P. Perona and J. Malik, "Scale space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. and Machine Intell.,* vol. PAMI-12, pp. 629-639, July 1990.

[63] D.L. Phillips, "A technique for the numerical solution of certain integral equations of the first kind," *J. Assoc. Comp. Mach.*, vol. 9, pp. 84-97, 1962.

[64] S.J. Reeves and R.M. Mersereau, "Optimal estimation of the regularization parameter and stabilizing functional for regularized image restoration," *Optical Engineering*, vol. 29(5), pp. 446-454, 1990.

[65] O. Rioul and M. Vetterli, "Wavelets and signal processing," *IEEE Signal Processing Magazine, vol. 8, No. 4, pp. 14-38, Oct. 1991.*

[66] I. Rock and S. Palmer, "The legacy of Gestalt psychology," *Scientific American*, Dec. 1990, pp. 84-90.

[67] A. Rosenfeld, "Some useful properties of pyramids," in *Multiresolution Image Processing and Analysis*, A. Rosenfeld, ed., Berlin: Springer-Verlag, 1984.

[68] I.J. Schoenberg, *Cardinal Spline Interpolation*, Philadelphia, PA: SIAM, 1973.

[69] E.P. Simoncelli, W.T. Freeman, E.H. Adelson, and D.J. Heeger, "Shiftable multiscale transforms," *IEEE Trans. Inform. Theory*, vol. 38, No. 2, pp. 587-607, March 1992.

[70] G. Strang, "Wavelets and dilation equations: A brief introduction," *SIAM Review*, vol. 31, pp. 614-627, Dec. 1989.

[71] M. Tavakoli and A. Rosenfeld, "Edge segment linking based on gray level and geometrical compatibilities," *Pattern Recognition*, vol. 15, pp. 369-377, May, 1982.

[72] A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems*, Washington, DC: Winston, 1977.

[73] F.M. Wahl, *Digital Image Processing*, Norwood, MA: Artech House, 1987.

[74] J.B. Weaver, Y. Xu, D.M. Healy, Jr. and L.D. Cromwell, "Filtering noise from images with wavelet transforms" *Magnetic Resonance in Medicine*, vol. 21, pp. 288-295, 1991.

[75] J.B. Weaver, Y. Xu, D.M. Healy, Jr. and J.R. Driscoll, "Wavelet encoded MR imaging," *Magnetic Resonance in Medicine*, vol. 24, pp. 275-287, 1992.

[76] M. Wertheimer, "Principles of perceptual organization," in *Readings in Perception*, D. C. Beardslee and M. Wertheimer, Eds., Princeton, NJ: Van Nostrand, 1958.

[77] A.P. Witkin, "Scale space filtering," in *Proc. Int. Joint Conf. Artificial Intell.*, 1983.

[78] G.W. Wornell, "A Karhunen-Loeve-like expansion for 1/f processes via wavelets," *IEEE Trans. Inform Theory*, vol. IT-36, pp. 859-861, July 1990.

[79] G.W. Wornell and A.V. Oppenheim, "Estimation of fractal signals from noisy measurements using wavelets," *IEEE Trans Signal Processing,* vol. 40, pp. 611-623, March 1992.

[80] Y. Xu, J. Lu, J.B. Weaver and D.M. Healy, Jr. "Filtering MR images with wavelet transforms," in *Proceedings of 10th Annual Meeting*, Society of Magnetic Resonance in Medicine, San Francisco, CA, Aug. 1991.

[81] Y. Xu, J.B. Weaver, D.M. Healy, Jr. and J. Lu, "Wavelet transform domain filters: A spatial selective noise filtration technique," submitted to *IEEE Trans. Image Proc.*, 1992.

[82] D. Youla and H. Webb, "Image restoration by the method of convex projections," *IEEE Trans. Med. Imaging*, vol. MI-1, pp. 81-101, Oct. 1982.

[83] A.L. Yuille and T.A. Poggio, "Scaling theorems for zero crossing," *IEEE Trans. Pattern Anal. and Machine Intell.,* vol. PAMI-8, pp. 15-25, Jan. 1986.

[84] W. R. Zettler, J. Huffman and D.C.P. Linden, "Application of compactly supported wavelets to image compression," in *Image Processing Algorithms and Techniques*, Proceedings of SPIE, vol. 1244, pp. 150-160, 1990.

[85] S. Zhong, "Edge representation from wavelet transform maxima," Ph.D. Dissertation, New York University, September 1990.

[86] W.L. Hwang, *private communications*, 1993.