NOVEL BAYESIAN MULTISCALE METHODS
FOR IMAGE DENOISING
USING ALPHA-STABLE DISTRIBUTIONS

By
Alin Achim

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
AT
UNIVERSITY OF PATRAS
PATRAS, GREECE
JUNE 2003

Interdepartmental Program of Postgraduate Studies in
BIOMEDICAL ENGINEERING
© Copyright by Alin Achim, 2003
ΕΠΙΤΑΜΕΛΗΣ ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ

1. Χ. Α. Μπεζεριάνος, Αναπληρωτής Καθηγητής, Τμήμα Ιατρικής, Πανεπιστήμιο Πατρών
2. Χ. Κ. Νικήτα, Αναπληρώτρια Καθηγήτρια, Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Τεχνολογιών, Εθνικό Μετσόβιο Πολυτεχνείο
3. Χ. Γ. Νασηφορίδης, Καθηγητής Ιατρικής, Καθηγητής, Τμήμα Ιατρικής, Πανεπιστήμιο Πατρών
4. Χ. Ν. Παλληκαρίδης, Καθηγητής Ιατρικής, Καθηγητής, Τμήμα Ιατρικής, Πανεπιστήμιο Πατρών
5. Χ. Γ. Παναγιωτάκης, Αναπληρωτής Καθηγητής, Τμήμα Ιατρικής, Πανεπιστήμιο Πατρών
6. Χ. Α. Στουρατής, Καθηγητής, Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Τεχνολογιών Υπολογιστών, Πανεπιστήμιο Πατρών
7. Χ. Π. Τσακωλίδης, Αναπληρωτής Καθηγητής, Τμήμα Πληροφορικής, Πανεπιστήμιο Κρήτης

ΤΡΙΜΕΛΗΣ ΣΥΜΒΟΥΛΕΥΤΙΚΗ ΕΠΙΤΡΟΠΗ

1. Χ. Α. Μπεζεριάνος, Αναπληρωτής Καθηγητής, Τμήμα Ιατρικής, Πανεπιστήμιο Πατρών, Πρόεδρος
2. Χ. Ν. Παλληκαρίδης, Καθηγητής Ιατρικής, Καθηγητής, Τμήμα Ιατρικής, Πανεπιστήμιο Πατρών, Μέλος
3. Χ. Α. Στουρατής, Καθηγητής, Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Τεχνολογιών Υπολογιστών, Πανεπιστήμιο Πατρών, Μέλος
ΠΕΡΙΛΗΨΗ ΔΙΔΑΚΤΟΡΙΚΗΣ ΔΙΑΤΡΙΒΗΣ

Ο απότερος σκοπός της έρευνας που παρουσιάζεται σε αυτή τη διδακτορική διατριβή είναι η διάθεση στην κοινότητα των κλινικών επιστημόνων μεθόδων οι οποίες να παρέχουν την καλύτερη δυνατή πληροφορία για να γίνει μια σωστή ιατρική διάγνωση. Οι εικόνες υπερήχουν προσβάλλονται ενδογενώς από θόρυβο, ο οποίος οφείλεται στην διαδικασία δημιουργίας των εικόνων μέσω ακτινοβολίας που χρησιμοποιεί σύμφωνες κυματομορφές. Είναι σημαντικό πριν τη διαδικασία ανάλυσης της εικόνας να γίνειει απάλειψη του θορύβου με κατάλληλο τρόπο ώστε να διατηρείται η υφή της εικόνας, η οποία βοηθά στην διάκριση ενός ιστού από έναν άλλο.

Κύριος στόχος της διατριβής αυτής υπήρξε η ανάπτυξη νέων μεθόδων καταστολής του θορύβου σε ιατρικές εικόνες υπερήχου στο πεδίο του μετασχηματισμού κυματιδίων. Αρχικά αποδείχθηκε μέσω εκτενών πειραμάτων μοντελοποίησης, ότι τα δεδομένα που προκύπτουν από τον διαχωρισμό των εικόνων υπερήχου σε υποπεριοχές συνεχώς περιγράφονται επακριβώς από μη-γκαουσιανές κατανομές βαρών ουρών, όπως είναι οι άλφα-ευσταθείς κατανομές. Κατόπιν, αναπτύχθηκε Μπεϋζιανούς εκτιμήσεις που αξιοποιούν αυτή τη στατιστική περιγραφή. Πιο συγκεκριμένα, χρησιμοποιήσαμε το άλφα-ευσταθείς μοντέλο για να σχεδιάσουμε εκτιμήσεις ελάχιστου απόλυτου λάθους και μέγιστης εκ των υστέρων πιθανότητας για άλφα-ευσταθή σήματα αναμετρημένα με μη-γκαουσιανό θόρυβο. Οι επεξεργαστές αφαίρεσης θορύβου που προέκυψαν επενδυργούν κατά μη-γραμμικό τρόπο στα δεδομένα και συσχετίζουν με βέλτιστο τρόπο αυτή την μη-γραμμικότητα με τον βαθμό τους τον οποίο τα δεδομένα είναι μη-γκαουσιανά. Συγκρίνουμε τις τεχνικές μας με κλασικά φίλτρα καθώς και σύγχρονες μεθόδους αναστηρούν και μαλακού κατωφλίου εφαρμοζόντας το καταγραφικές ιατρικές εικόνες υπερήχου και ποσοτικοποιήσαμε την απόδοση που επιτεύχθηκε. Τέλος, δείξαμε ότι οι προτεινόμενοι επεξεργαστές μπορούν να βρουν εφαρμογές καταγραφικές εικόνες και σε άλλες περιοχές ενδιαφέροντος και απελεύθημες ως ενδεικτικά παράδειγμα την περίπτωση εικόνων μας συνθετικής διατομής.
Părinților mei, Ion și Mariana
și surorii mele Laura
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table of Contents</td>
<td>ix</td>
</tr>
<tr>
<td>Abstract</td>
<td>xi</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>xiii</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 State of the Art</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Contributions and Publications</td>
<td>3</td>
</tr>
<tr>
<td>2 Wavelets in Image Processing</td>
<td>7</td>
</tr>
<tr>
<td>2.1 Introduction to Wavelet Theory</td>
<td>8</td>
</tr>
<tr>
<td>2.1.1 Rationale for the Use of Wavelets in Signal Processing</td>
<td>8</td>
</tr>
<tr>
<td>2.1.2 Short-Time Fourier Transform vs Wavelet Transform</td>
<td>10</td>
</tr>
<tr>
<td>2.2 Dyadic Wavelet Transform</td>
<td>15</td>
</tr>
<tr>
<td>2.2.1 Multiresolution Analysis</td>
<td>15</td>
</tr>
<tr>
<td>2.2.2 Fast Discrete Wavelet Transform Algorithm in Two Dimensions</td>
<td>19</td>
</tr>
<tr>
<td>2.2.3 Daubechies’ Family of Regular Filters and Wavelets</td>
<td>21</td>
</tr>
<tr>
<td>2.3 Wavelet Shrinkage Principles</td>
<td>26</td>
</tr>
<tr>
<td>2.3.1 Hard and Soft Thresholding</td>
<td>27</td>
</tr>
<tr>
<td>2.3.2 Bayesian Wavelet Shrinkage</td>
<td>29</td>
</tr>
<tr>
<td>3 The Alpha-Stable Family of Distributions</td>
<td>33</td>
</tr>
<tr>
<td>3.1 Basic Properties of the Alpha-Stable Family</td>
<td>34</td>
</tr>
<tr>
<td>3.2 The Class of Real $S_\alpha S$ Distributions</td>
<td>36</td>
</tr>
<tr>
<td>3.3 Bivariate Isotropic Stable Distributions</td>
<td>39</td>
</tr>
<tr>
<td>3.4 Symmetric Alpha-Stable Processes</td>
<td>40</td>
</tr>
<tr>
<td>3.5 Parameter Estimation for $S_\alpha S$ Distributions</td>
<td>43</td>
</tr>
<tr>
<td>3.5.1 Maximum Likelihood Method</td>
<td>43</td>
</tr>
</tbody>
</table>
Abstract

Before launching into ultrasound research, it is important to recall that the ultimate goal is to provide the clinician with the best possible information needed to make an accurate diagnosis. Ultrasound images are inherently affected by speckle noise, which is due to image formation under coherent waves. Thus, it appears to be sensible to reduce speckle artifacts before performing image analysis, provided that image texture that might distinguish one tissue from another is preserved.

The main goal of this thesis was the development of novel speckle suppression methods from medical ultrasound images in the multiscale wavelet domain. We started by showing, through extensive modeling, that the subband decompositions of ultrasound images have significantly non-Gaussian statistics that are best described by families of heavy-tailed distributions such as the alpha-stable. Then, we developed Bayesian estimators that exploit these statistics. We used the alpha-stable model to design both the minimum absolute error (MAE) and the maximum a posteriori (MAP) estimators for alpha-stable signal mixed in Gaussian noise. The resulting noise-removal processors perform non-linear operations on the data and we relate this non-linearity to the degree of non-gaussianity of the data. We compared our techniques to classical speckle filters and current state-of-the-art soft and hard thresholding methods applied on actual ultrasound medical images and we quantified the achieved performance improvement.

Finally, we have shown that our proposed processors can find application in other areas of interest as well, and we have chosen as an illustrative example the case of synthetic aperture radar (SAR) images.
Acknowledgements

First of all, I would like to thank Professor Anastasios Bezerianos, my main advisor, for his many suggestions and constant support during this research. The door of his office was always open for me and he always found some time to hear my problems. I am also grateful to Professor Nikolas Pallikarakis and Professor Athanasios Stouraitis for their participation in my advisory committee. In particular, Professor Pallikarakis together with Professor Giorgos Kostopoulos have also helped me in a hard episode of my private live. I will never forget their support. Professor A. Stouraitis expressed his interest in my work and provided me the reprints of some of his recent joint work with Y. Karayiannis, which hopefully constitutes the ground for continued collaboration.

Besides the members in my advisory committee, there was a person without whom this work could not have been carried out in the way it has been done: Panos Tsakalides was the one who introduced me in the alpha-stable world and always helped me to keep the hope alive. I’m sure that his advisory work put the bases for a long lasting collaboration and friendship.

Also, I should definitely mention here the great contribution that Dr. Radu Negescu from the Institute of Public Health and Professor F.M.G. Tomescu from “Politehnica” University of Bucharest have had to my formation as an young researcher in Romania.

I would like to thank Dr. C. Frank Starmer and the IT Lab at the Medical University of South Carolina for providing most of the ultrasound images used in this thesis. The image of the fetal chest used in Chapter 5 has been provided by Acuson Corporation (Mountain View, CA). Dr. Daniel E. Wahl from Sandia National Laboratories supplied me with part of the SAR imagery used in Chapter 6. I am also grateful to Dr. John P. Nolan from American University who kindly provided his STABLE program in library form.

The State Scholarships Foundation (IKY) grant that was awarded to me for the period 1999–2003, was crucial to the successful completion of this project.
I wish also to thank all my colleagues from the Biosignal Processing Lab for offering me an enjoyable working environment. Special thanks to Adi and Ovidiu for their patience to answer my computer related questions. And to all the members, past and present, of the “Romanian Mafia” for all the (mostly) good times we had together in *Fititiki Estia*. I’m thinking especially of Laura, Adi, Cosmin, Claudia, Ovidiu, Lorenzo, Claudia, Mihai, Anda, Adi, Bobby, Otilia, Cristi, Alexandra, Andrei, Liviu, Delia, Edi, Diana ...

Of course, I am grateful to my parents and my sister for their unconditional love, trust and support. With gratitude, this thesis is dedicated to them.

Finally, I wish to thank my wife, Lumi. She will probably never read this dissertation, yet with her I shared most of the last four years and she was the main witness of my struggles during its accomplishment.

Patras, Greece
June 20, 2003

Alin Achim
Chapter 1

Introduction

1.1 State of the Art

For more than two decades, ultrasonography has been considered as one of the most powerful techniques for imaging organs and soft tissue structures in the human body. Today, it is being used at an ever-increasing rate in the field of medical diagnostic technology. Ultrasonography is often preferred over other medical imaging modalities because it is noninvasive, portable, and versatile, it does not use ionizing radiations, and it is relatively low-cost. The images produced by commercial ultrasound systems are usually optimized for visual interpretation, since they are mostly used in real-time diagnostic situations. However, the main disadvantage of medical ultrasonography is the poor quality of images, which are affected by multiplicative speckle noise [45].

Imaging speckle is a phenomenon that occurs when a coherent source and a non-coherent detector are used to interrogate a medium, which is rough on the scale of the wavelength. Speckle occurs especially in images of the liver and kidney whose underlying structures are too small to be resolved by large wavelength ultrasound. The presence of speckle is undesirable since it degrades image quality and it affects
the tasks of human interpretation and diagnosis. As a result, speckle filtering is a critical pre-processing step for feature extraction, analysis, and recognition from medical imagery measurements.

Current speckle reduction methods are based on temporal averaging [1, 38], median filtering [44, 81], and Wiener filtering. The adaptive weighted median filter, first introduced in [60], can effectively suppress speckle but it fails to preserve many useful details, being merely a low-pass filter. The classical Wiener filter, which utilizes the second order statistics of the Fourier decomposition, is not adequate for removing speckle since it is designed mainly for additive noise suppression. To address the multiplicative nature of speckle noise, Jain developed a homomorphic approach, which by taking the logarithm of the image, converts the multiplicative into additive noise, and consequently applies the Wiener filter [45].

Recently, there has been considerably interest in using the wavelet transform as a powerful tool for recovering signals from noisy data [27, 37, 73, 78, 89, 106]. The main reason for the choice of multiscale bases of decompositions is that the statistics of many natural signals, when decomposed in such bases, are significantly simplified. More specifically, methods based on multiscale decompositions consist of three main steps: First, the raw data are analyzed by means of the wavelet transform, then the empirical wavelet coefficients are shrunk, and finally, the denoised signal is synthesized from the processed wavelet coefficients through the inverse wavelet transform. These methods are generally referred to as wavelet shrinkage techniques. In [106], Zong et al. use a logarithmic transform to separate the noise from the original image. They adopt regularized soft thresholding (wavelet shrinkage) to remove noise energy within the finer scales and nonlinear processing of feature energy for contrast
1.2 Contributions and Publications

enhancement. A similar approach applied to synthetic aperture radar (SAR) images is presented in [37]. The authors perform a comparative study between a complex wavelet coefficient shrinkage filter and several standard speckle filters that are largely used by SAR imaging scientists, and show that the wavelet-based approach is among the best for speckle removal.

Thresholding methods have two main drawbacks: (i) the choice of the threshold, arguably the most important design parameter, is done in an *ad hoc* manner; and (ii) the specific distributions of the signal and noise may not be well matched at different scales. To address these disadvantages, Simoncelli *et al.* developed nonlinear estimators, based on formal Bayesian theory, which outperform classical linear processors and simple thresholding estimators in removing noise from visual images [88, 89]. They used a generalized Laplacian model for the subband statistics of the signal and developed a noise-removal algorithm, which performs a “coring” operation to the data. The term “coring” refers to a widely used technique for noise suppression, which preserves high-amplitude observations while suppressing low-amplitude values from the highpass bands of a signal decomposition.

1.2 Contributions and Publications

In this thesis, we develop novel speckle suppression methods for medical ultrasound images. The proposed processors consist of two major modules: (i) a subband representation function that utilizes the wavelet transform, and (ii) a Bayesian denoising algorithm based on an alpha-stable *prior* for the signal. First, the original image is logarithmically transformed to change multiplicative speckle to additive white noise. Then, the transformed image is analyzed into a multiscale wavelet domain. We show
that the subband decompositions of actual ultrasound images have significantly non-Gaussian statistics that are best described by families of heavy-tailed distributions like the alpha-stable. Motivated by our modeling results, we design Bayesian estimators that exploits these statistics. We use the alpha-stable model to develop blind speckle-suppression processors that perform non-linear operations on the data, and we relate these non-linearities to the degree of non-Gaussianity of the data.

The thesis is organized as follows: In Chapter 2, we revise some of the basic wavelet concepts that we use for our later developments. Chapter 3 is intended to provide some necessary preliminaries on the alpha-stable statistical model that we employ to characterize the wavelet subband coefficients of images. Our main contributions are highlighted in Chapters 4 through 6. Specifically, in Chapter 4 we define the ultrasound speckle suppression problem, present results on the modeling of the subband coefficients of actual medical ultrasound images and we design the Bayesian MAE estimator based on the signal alpha-stable statistics. An alternate approach to solving the same problem is presented in Chapter 5. More exactly, we select a different cost function for the design of the Bayesian estimator, which results to a MAP filter based again on alpha-stable statistics. The results of processing ultrasound images using both algorithms are compared with the results of other state-of-the-art methods using simulated as well as real images. The improvement is quantified using different quality measures. The methods proposed in Chapters 4 and 5 can be easily adapted for the purpose of denoising images obtained by means of other imaging modalities. This is not only true for biomedical images but also for images from other fields of interest. Thus, in Chapter 6 we show application for the case of SAR images. Finally, future work directions are drawn in Section 4.5.
The research presented in this thesis contributed so far in two publications in international journals [3, 7], and four papers in Conference Proceedings [2, 4, 5, 6]. The paper “Ultrasound Image Denoising via Maximum a Posteriori Estimation of Wavelet Coefficients” [4] participated in 2001 as open finalist in the EMBS Whitaker Student Paper Competition held in Istanbul during the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society.
Chapter 2
Wavelets in Image Processing

This chapter is intended to review some of the basic wavelet concepts that will be used later for our developments. The fundamentals on wavelet theory can be found in a number of books and in many papers at different levels of exposition. Some of the standard books are [26, 62, 68, 96]. Introductory papers include [39, 91, 97], and more technical ones are [21, 61, 95]. For the purpose of this thesis, in this chapter we only present a synthetic view of the wavelet theory and show connections of the wavelet transform properties to the potential applications in image processing. We start by synthesizing the main rationales for the use of wavelets in signal processing and present their advantages over the short-time Fourier transform. Then, we review the concept of multiresolution analysis, we describe Mallat’s Discrete Wavelet Transform algorithm and Daubechies’ family of filters that we use in our developments. The rest of the chapter presents ideas of various wavelet based image denoising methods and reviews the state of the art in this field.
2.1 Introduction to Wavelet Theory

2.1.1 Rationale for the Use of Wavelets in Signal Processing

Despite the continuously growing interest in the time domain modeling of random medical signals, spectral analysis remains a fundamental approach that can provide useful information when it is applied under the assumption of stationary, linear processes. However, in many biomedical applications the assumption of stationarity fails to be true. Thus, the strong non-stationarity of several medical signals requires a proper non-stationary approach in their analysis.

The time-frequency representation of the non-stationary signals is an issue which has been increasingly discussed in the general signal processing literature [22, 79, 80]. It has become a powerful alternative for the analysis of the non-stationary signals since the classical Fourier transform gives the frequency contents of the signals without providing information about the time localization of the observed frequency components. Several techniques have been proposed such as Short-Time Fourier Transform (STFT) [36, 79], Wigner-Ville Transform [65, 98] and Wavelet Transform (WT) [62]. In the STFT transform (which is also called the window Fourier transform or the Gabor transform) the signal is multiplied by a smooth window function (typically Gaussian) and the Fourier integral is applied to the windowed signal. Thus, choosing a short analysis window may cause a poor frequency resolution, while a long analysis window introduce a poor time (space) resolution and a high risk to violate the assumption of stationarity within the window. The Wigner-Ville Transform offers the best simultaneous resolution in time and frequency. However, its main drawback is to generate some ghost frequencies which do not exist in the analyzed signal (these cross-terms are called interferences). The WT is characterized by a frequency
response logarithmically scaled along the frequency axis, as opposed to the STFT, which uses a fixed window in time domain. Thus, it provides a good time resolution at high frequencies and a good frequency resolutions at low frequencies [6], being appropriate to discriminate transient high-frequency components closely located in time and long duration components closely spaced in frequency. Despite the fact that the time-varying spectrum of the non-stationary signals was a starting point for the development of different time-frequency representation techniques, it did not remain a singular objective of wavelet transform based analysis. The joint time-frequency analysis effected by the WT provides also natural settings for well-defined statistical applications, which include estimation, detection, classification, filtering and compression. Consequently, the use of wavelets as a tool for medical signal processing has rapidly increased in recent years for both 1-D [77, 78] and 2-D signals [41, 75, 106].

The WT is presently applied mainly in two different ways: as a pseudo-continuous transformation (used for spectro-temporal analysis of 1-D medical signals), and as a multiresolution orthogonal signal decomposition (with specific applications in feature extraction, pattern recognition, filtering and compression of both 1-D and 2-D medical signals). Other important properties of the WT that make it suitable for signal and image processing applications are:

- Multiresolution - the WT offers a scale invariant representation
- Sparsity - the wavelet coefficients distribution of images is sparse
- Fast algorithms - efficient decomposition and reconstruction algorithms exist for implementing the WT
- Edge detection - small wavelet coefficients correspond to homogenous areas, while large wavelet coefficients correspond to image edges.
These properties will be discussed later in this chapter.

### 2.1.2 Short-Time Fourier Transform vs Wavelet Transform

The Short Time Fourier Transform assumes that the signal $x(t)$ is stationary within a window $g(t)$ of limited extent, centered at time location $t$. Consequently, it maps the signal $x(t)$ into a two-dimensional function in a time-frequency plane $(t, \omega)$:

$$STFT(\tau, \omega) = \int_{-\infty}^{\infty} x(t) \cdot g^*(t - \tau) \cdot e^{-j\omega t} dt$$  \hspace{1cm} (2.1.1)

The analysis here depends critically on the choice of the window $g(t)$. The resulting time-varying spectrum is displayed as a three-dimensional plot of energy versus time and frequency. Figure 2.1 (a) shows vertical stripes in the time-frequency plane, illustrating this “windowing of the signal” view of the STFT. Given a version of the signal windowed around time $t$, one computes all “frequencies” of the STFT. An alternative view is based on a filter bank interpretation of the same process. At a given frequency $\omega$, the whole signal is filtered using a bandpass filter, which has as impulse response the window function modulated to that frequency (this is shown as the horizontal stripes in Figure 2.1 (a). In this way, the STFT may be also seen as a modulated filter bank. From this dual interpretation, a possible drawback related to the time and frequency resolution can be shown. Consider the ability of the STFT to discriminate between two pure sinusoids. Given a window function $g(t)$ and its Fourier transform $G(\omega)$, define the bandwidth $\Delta \omega$ of the filter as:

$$\Delta \omega^2 = \frac{\int \omega^2 \cdot |G(\omega)|^2 d\omega}{\int |G(\omega)|^2 d\omega}$$  \hspace{1cm} (2.1.2)

where the denominator is the energy of $g(t)$. Two sinusoids will be discriminated only if they are more than $\Delta \omega$ apart (the resolution in frequency of the STFT analysis is
2.1 Introduction to Wavelet Theory

Figure 2.1: Time-frequency tilings for Short-Time Fourier Transform (a) and Wavelet Transform (b) (cf. [96]).

The Wavelet Transform provides an interesting and useful alternative to the classical STFT briefly presented before. The basic difference consists on the fact that the WT performs a multiresolution analysis of the signals, by making use of short windows at high frequencies and long windows at low frequencies. Formally, the definition of the WT is given by:

\[ CW(T_x, r) = \frac{1}{\sqrt{r}} \int x(t) h^*(t - \tau) dt \]  

(2.1.4)

where the analysis is made now in the spirit of constant relative bandwidth. The
straightforward interpretation of the equation 2.1.4 is that as the scale parameter \( r \) increases, the filter impulse response becomes spread out in time, and takes only long-time behavior (low frequencies) into consideration (see the lowest horizontal stripes of Figure 2.1 (b). Similarly, when the scale \( r \) decreases, only transient features of the signal are seen through the analysis window (see the highest vertical stripes of Figure 2.1 (b). The time resolution becomes arbitrarily good at high frequencies, while the frequency resolution becomes arbitrarily good at low frequencies. Consequently, two very close short bursts can always be separated in the analysis by going up to higher analysis frequencies in order to increase the corresponding time resolution.

The resolution in time and frequency cannot be arbitrarily small, because their product is lower bounded as given by the well-known Heisenberg inequality:

\[
\Delta \omega \cdot \Delta t \geq \frac{1}{2} \tag{2.1.5}
\]

It implies that one can only trade time resolution for frequency resolution or vice versa. To overcome the resolution limitation of the STFT, one can let the resolution \( \Delta \omega \) and \( \Delta t \) to vary in the time-frequency plane in order to obtain a multiresolution analysis. When the analysis is viewed as a filter bank [80], the time resolution must increase with the central frequency of the analysis filters. By imposing that \( \Delta \omega \) is proportional to \( \omega \), or :

\[
\frac{\Delta \omega}{\omega} = c \tag{2.1.6}
\]

where \( c \) is a constant, the analysis filter bank is then composed of band-pass filters with constant relative bandwidth. Consequently, instead of the frequency responses of the analysis filter being regularly spaced over the frequency axis (as for the STFT case), they are regularly spread in a logarithmic scale (figure 2.2 (b)). The time
and frequency resolutions still satisfy the Heisenberg inequality, but now the time resolution becomes arbitrarily good at high frequencies, while the frequency resolution becomes arbitrarily good at low frequencies.

The Continuous Wavelet Transform (CWT) exactly follows the above ideas while and, furthermore, all impulse responses of the filter bank are defined as scaled versions of the same prototype \( h(t) \):

\[
h_r(t) = \frac{1}{\sqrt{|r|}} h\left(\frac{t}{r}\right) \tag{2.1.7}
\]

where \( r \) is a scale factor and the constant \( \frac{1}{\sqrt{|r|}} \) is used for energy normalisation. Since the same prototype \( h(t) \), called the basic wavelet, is used for all the filter impulse responses, no specific scale is privileged, i.e. the wavelet analysis is self-similar at all scales. For comparison purposes, the basic wavelet \( h(t) \) could be chosen as a modulated window:

\[
h(t) = g(t) \cdot e^{-j\omega_0 t} \tag{2.1.8}
\]

Then, the frequency responses of the analysis filters indeed satisfy 2.1.6 with the identification

\[
r = \frac{\omega_0}{\omega}
\]

It should be mentioned that the local frequency \( \omega = r\omega_0 \), whose definition depends on the basic wavelet, is no longer linked to frequency modulation (as was the case for the STFT) but is now related to time-scaling. The scale is defined as in the geographical maps: since the filter bank impulse responses are dilated as the scale increases, large scales correspond to contracted signals, while small scales correspond to dilated signals.

Once a window has been chosen for the STFT, then the time-frequency resolution is fixed over the entire time-frequency plane (since the same window is used at all
Figure 2.2: Division of the frequency domain (a) for the STFT (uniform coverage) and (b) for the WT (logarithmic coverage).

frequencies) as shown in Figure 2.2 (a). Alternatively, in the case of Wavelet Transform, $\Delta \omega$ and $\Delta t$ change with the center frequency of the analysis filter. They still satisfy the Heisenberg inequality, but now, the time resolution becomes arbitrarily good at high frequencies, while the frequency resolution becomes arbitrarily good at low frequencies as shown in Figure 2.2 (b).
2.2 Dyadic Wavelet Transform

In this section we review wavelet theory starting with the concept of multiresolution analysis [63] and the implementation of Mallat’s Fast DWT algorithm. Because our particular interest is in image processing application we address directly the 2-D extension of the classical theory. Finally, we review the main properties of the Daubechies’ wavelet family and orthogonal filters, extensively used later in this thesis.

2.2.1 Multiresolution Analysis

Let’s denote by $L^2(\mathbb{R}^2)$, the Hilbert space of finite energy two-dimensional functions $f(x, y)$. The wavelet multiresolution decomposition makes use of a linear approximation operator $A_{2j}$, which transforms a function $f(x, y) \in L^2(\mathbb{R}^2)$ into approximations at different resolution levels $2^j$. The operator $A_{2j}$ is an orthogonal projection on the vector space $V_{2j} \subset L^2(\mathbb{R}^2)$ of all possible approximations at resolution $2^j$ of functions in $L^2(\mathbb{R}^2)$. A multiresolution approximation of $L^2(\mathbb{R}^2)$ is defined as a sequence of subspaces $(V_{2j})_{j \in \mathbb{Z}}$, which satisfies the following properties [62, 63]:

1. Causality: The approximation of an image at a resolution $2^{j+1}$ contains all the necessary information to compute the same image at a smaller resolution $2^j$:

$$\forall j \in \mathbb{Z}, V_{2j} \subset V_{2^{j+1}} \quad (2.2.1)$$

2. The subspaces of approximated functions should be derived from one another by scaling each approximated function by the ratio of their resolution values:

$$\forall j \in \mathbb{Z}, f(x, y) \in V_{2j} \Leftrightarrow f(2x, 2y) \in V_{2^{j+1}} \quad (2.2.2)$$
3. Discrete characterization: The approximation \( A_{2^j} f(x, y) \) of an image \( f(x, y) \) can be characterized by \( 2^j \) samples per length unit:

There exists an isomorphism \( I \) from \( V_1 \) onto \( I^2(\mathbb{Z}^2) \).  \( (2.2.3) \)

4. Translation of the approximation: When \( f(x, y) \) is translated (eventually in two directions) by some lengths proportional with \( 2^{-j} \), \( A_{2^j} f(x, y) \) is translated by the same amount and it contains the same samples that have been translated:

\[
\forall k, l \in \mathbb{Z}, A_{1} f_{k,l}(x, y) = A_{1} f(x - k, y - l), \text{ where } f_{k,l}(x, y) = f(x - k, y - l) \quad (2.2.4)
\]

5. Translation of the samples:

\[
I(A_{1} f(x, y)) = (\alpha_{m,n})_{(m,n) \in \mathbb{Z}^2} \Leftrightarrow I(A_{1} f_{k,l}(x, y)) = (\alpha_{m-k,n-l})_{(m,n) \in \mathbb{Z}^2} \quad (2.2.5)
\]

6. As the resolution increases to \(+\infty\) the image approximation should converge to the original image:

\[
\lim_{j \to +\infty} V_{2^j} = \bigcup_{j=-\infty}^{+\infty} V_{2^j} \text{ is dense in } L^2(\mathbb{R}^2) \quad (2.2.6)
\]

7. As the resolution decreases to zero, the image approximation contains less and less information and converges to zero:

\[
\lim_{j \to -\infty} V_{2^j} = \bigcap_{j=-\infty}^{+\infty} V_{2^j} = \{0\} \quad (2.2.7)
\]

In [63], Mallat showed that there exists a unique scaling function \( \phi(x, y) \) whose dilation and translation defines an orthonormal basis of each space \( V_{2^j} \). Specifically, if we set \( \phi_{2^j}(x, y) = 2^j \phi(2^j x, 2^j y) \) for \( j \in \mathbb{Z} \) (the dilation of \( \phi(x, y) \) by \( 2^j \)), then \( (2^{-j}\phi_{2^j}(x - 2^{-j}n, y - 2^{-j}m))(n,m)_{(n,m) \in \mathbb{Z}^2} \) is an orthonormal basis of \( V_{2^j} \).
2.2 Dyadic Wavelet Transform

For the particular case of separable multiresolution approximations of \( L^2(R^2) \), each vector space can be decomposed as a tensor product of two identical subspaces of \( L^2(R) \):

\[
V_{2^j} = V_{2^j}^1 \bigotimes V_{2^j}^1
\]

It can be shown that the scaling function \( \phi(x, y) \) can be written as

\[
\phi(x, y) = \phi(x) \, \phi(y)
\]

where \( \phi(x) \) is the one-dimensional scaling function of \( V_{2^j}^1 \), \( j \in \mathbb{Z} \). Consequently, the orthogonal basis of \( V_{2^j} \) is given by:

\[
(2^{-j} \Phi_{2^j}(x - 2^{-j}n, y - 2^{-j}m))_{(n,m) \in \mathbb{Z}^2} = (2^{-j} \Phi_{2^j}(x - 2^{-j}n) \Phi_{2^j}(y - 2^{-j}m))_{(n,m) \in \mathbb{Z}^2}
\]

Thus, the approximation of a signal \( f(x, y) \) at a resolution \( 2^j \) is characterized by the set of inner products:

\[
A_{2^j}^d f = (\langle f(x, y), \Phi_{2^j}(x - 2^{-j}n) \Phi_{2^j}(y - 2^{-j}m) \rangle)_{(n,m) \in \mathbb{Z}^2} \tag{2.2.8}
\]

\( A_{2^j}^d \) is called a discrete approximation of \( f(x, y) \) at resolution \( 2^j \).

The multiresolution decomposition is based on the difference of information available at two successive resolutions \( V_{2^j} \) and \( V_{2^{j+1}} \), which is called the detail signal at the resolution \( 2^j \). It can be shown that the detail signal at the resolution \( 2^j \) is given by the orthogonal projection of the original signal on the orthogonal complement of \( V_{2^j} \) in \( V_{2^{j+1}} \). Let \( O_{2^j} \) be this orthogonal complement. Mallat [63] showed that an orthonormal basis of \( O_{2^j} \) could be built by scaling and translating three wavelets functions:

\[
\psi^1(x, y) = \Phi(x) \, \psi(y), \quad \psi^2(x, y) = \psi(x) \Phi(y), \quad \psi^3(x, y) = \psi(x) \, \psi(y) \tag{2.2.9}
\]
where \( \psi(x) \) is the one-dimensional wavelet associated with the scaling function \( \phi(x) \).

The three “wavelets” \( \psi^1, \psi^2, \text{and } \psi^3 \) are such that

\[
(2^{-j} \psi^1_{2j}(x - 2^{-j}n, y - 2^{-j}m) \\
2^{-j} \psi^2_{2j}(x - 2^{-j}n, y - 2^{-j}m) \\
2^{-j} \psi^3_{2j}(x - 2^{-j}n, y - 2^{-j}m))_{(n,m) \in \mathbb{Z}^2}
\]

is an orthonormal basis of \( O_{2^j} \) and

\[
(2^{-j} \psi^1_{2j}(x - 2^{-j}n, y - 2^{-j}m) \\
2^{-j} \psi^2_{2j}(x - 2^{-j}n, y - 2^{-j}m) \\
2^{-j} \psi^3_{2j}(x - 2^{-j}n, y - 2^{-j}m))_{(n,m) \in \mathbb{Z}^2}
\]

is an orthonormal basis of \( L^2(\mathbb{R}^2) \).

The difference of information between \( A^d_{2^{j+1}} f \) and \( A^d_{2^j} f \) is equal to the orthonormal projection of \( f(x) \) on \( O_{2^j} \) and is characterized by the following sets of inner products:

\[
D^1_{2^j} f = \langle \langle f(x, y), \psi^1_{2^j}(x - 2^{-j}n, y - 2^{-j}m) \rangle \rangle_{(n,m) \in \mathbb{Z}^2}
\]  

\[
D^2_{2^j} f = \langle \langle f(x, y), \psi^2_{2^j}(x - 2^{-j}n, y - 2^{-j}m) \rangle \rangle_{(n,m) \in \mathbb{Z}^2}
\]

\[
D^3_{2^j} f = \langle \langle f(x, y), \psi^3_{2^j}(x - 2^{-j}n, y - 2^{-j}m) \rangle \rangle_{(n,m) \in \mathbb{Z}^2}
\]

\( D^1_{2^j} f, D^2_{2^j} f \) and \( D^3_{2^j} f \) are called the details image at resolution \( 2^j \). It can be shown that each of the inner products that define the approximation and the details are equal to a uniform sampling of two-dimensional convolution products:

\[
A^d_{2^j} f = ((f(x,y) * \Phi_{2^j}(-x)\Phi_{2^j}(-y))(2^{-j}n, 2^{-j}m))_{(n,m) \in \mathbb{Z}}
\]

\[
D^1_{2^j} f = ((f(x,y) * \Phi_{2^j}(-x)\psi_{2^j}(-y))(2^{-j}n, 2^{-j}m))_{(n,m) \in \mathbb{Z}}
\]

\[
D^2_{2^j} f = ((f(x,y) * \psi_{2^j}(-x)\Phi_{2^j}(-y))(2^{-j}n, 2^{-j}m))_{(n,m) \in \mathbb{Z}}
\]

\[
D^3_{2^j} f = ((f(x,y) * \psi_{2^j}(-x)\psi_{2^j}(-y))(2^{-j}n, 2^{-j}m))_{(n,m) \in \mathbb{Z}}
\]
\[ D_2^3 f = (f(x, y) * \psi_{2j}(-x)\psi_{2j}(-y))(2^{-j}n, 2^{-j}m)_{(n,m) \in \mathbb{Z}} \]  

(2.2.18)

Expressions 2.2.15- 2.2.18 show that in two dimensions the image approximation and details are computed with separable filtering of the signal along the abscissa and the ordinate. For any \( J > 0 \), an image \( A_1^d f \) is completely represented by \( 3J+1 \) subimages:

\[
(A_2^d f, (D_1^1 f)_{-J \leq j \leq -1}, (D_2^2 f)_{-J \leq j \leq -1}, (D_3^3 f)_{-J \leq j \leq -1})
\]

This set of discrete signals is called an \textit{orthogonal wavelet representation} in two dimensions, and consists of the reference signal at a coarse resolution \( (A_2^d f) \) and the detail signals \( (D_k^j f) \) for different orientations at the resolution \( 2^j_{-J \leq j \leq -1} \). It can be interpreted as a decomposition of the original signal in an orthonormal wavelet basis or as a decomposition of the signal in a set of independent, spatially oriented frequency channels.

### 2.2.2 Fast Discrete Wavelet Transform Algorithm in Two Dimensions

A fast, pyramidal filter bank algorithm was introduced by Mallat [63] for computing the coefficients of a 2-D orthogonal wavelet representation. The algorithm can be seen as a one dimensional wavelet transform applied successively along both the x and y axes and it is based on appropriately designed \textit{quadrature mirror filters}, i.e. a low-pass filter \( H \) and a high-pass filter \( G \) (Figure 2.3), and on a binary decimation operator \( D \downarrow \). The properties of the filters \( H \) and \( G \) are discussed in [26, 63]. The block diagram of the 2-D DWT algorithm is illustrated in Figure 2.3(a). At each step \( j \), the approximation \( A_{2j+1} \) is decomposed into \( A_{2j}, D_{2j}^1, D_{2j}^2, D_{2j}^3 \). First, the rows of \( A_{2j+1} \) are convolved with a one-dimensional filter and we retain every other row, then the columns of the resulting signals are convolved with another one-dimensional
Figure 2.3: (a) Block-diagram of the pyramidal decomposition algorithm used for 2-D DWT computation. (b) Block-diagram of the pyramidal reconstruction algorithm

filter and we retain every other column, and so on.

Figure 2.4 shows the wavelet representation of the classical Lena image, decomposed on 2 resolution levels. From the figure it can be seen that the DWT yields fairly decorrelated coefficients. An important observation is that the positions of large wavelet coefficients designate image edges, i.e., the DWT has an edge detection property.

The wavelet representation is complete. It can be shown [63] that the original image can also be reconstructed by means of a pyramidal algorithm. At each step, the image $A_{2j+1}^{2j}$ is reconstructed from the approximation $A_{2j}$ and the details
2.2 Dyadic Wavelet Transform

Figure 2.4: Wavelet decomposition of the Lena image on 2 resolution levels. Bright pixels designate large amplitude coefficients.

\[ D_{2}^{1}, D_{2}^{2}, D_{2}^{3}. \]

Specifically, columns of zeros are inserted between each columns of the images \( A_{2}, D_{2}^{1}, D_{2}^{2}, \) and \( D_{2}^{3} \), the rows are convolved with a one-dimensional filter, then a row of zeros is inserted between each row of the resulting image and the columns are convolved with another dimensional filter. The block diagram shown in Figure 2.3 illustrates this algorithm. The image \( A_{1} \) is reconstructed from its wavelet decomposition by repeating this process for \( -J \leq j \leq -1 \).

2.2.3 Daubechies’ Family of Regular Filters and Wavelets

In Section 2.2.1 it has been shown that we can construct orthonormal families of functions where each function is related to a single prototype wavelet through shifting and scaling. This construction is a direct continuous-time approach based on
the concept of multiresolution analysis. A different, indirect approach starts from
discrete-time filters, which can be iterated and, under certain conditions, leads to
continuous time prototype wavelet and the corresponding derived orthonormal fami-
lies of functions. This construction pioneered by Daubechies [25] provides very prac-
tical wavelet decomposition schemes, implementable with the pyramidal algorithms
described in Section 2.2.2 and based on finite-length discrete time filters. The method
of construction can be demonstrated by starting from a two-channel orthogonal filter
bank as illustrated in figure 2.5 (a). The discrete time-domain low-pass and high-pass
analysis filters are denoted by \( h[n] \) and \( g[n] \), while the synthesis filters are denoted
by \( \tilde{h}[n] \) and \( \tilde{g}[n] \), respectively. It should be mentioned that orthogonality imposes
that the impulse response of the analysis filters are the time-reversed versions of the
synthesis filters [96]. Considering that the filter bank is iterated on the branch with
the low-pass filter as shown in figure 2.5 (c), the two equivalent filters after \( i \) steps
can be expressed in the \( z \)-domain using the fact that filtering with \( \tilde{H}(z) \) followed by
upsampling by 2 is equivalent to upsampling by 2 followed by filtering with \( \tilde{H}(z^2) \),
as follows:

\[
\tilde{H}^{(i)}(z) = \prod_{k=0}^{i-1} \tilde{H}\left(z^{2^k}\right) \tag{2.2.19}
\]

\[
\varphi^{(i)}(t) = 2^{i/2} \tilde{h}^{(i)}[n] , \quad \frac{n}{2^i} \leq t < \frac{n+1}{2^i} \tag{2.2.20}
\]

\[
\tilde{G}^{(i)}(z) = \tilde{G}\left(z^{2^{i-1}}\right) \prod_{k=0}^{i-2} \tilde{G}\left(z^{2^k}\right) , \quad i = 1, 2, ... \tag{2.2.21}
\]

The discrete-time iterated filters \( \tilde{h}^{(i)}[n] \) and \( \tilde{g}^{(i)}[n] \) are associated with the continuous-
time functions \( \varphi^{(i)}(x) \), \( \psi^{(i)}(x) \) as follows:

\[
\psi^{(i)}(t) = 2^{i/2} \tilde{g}^{(i)}[n] , \quad \frac{n}{2^i} \leq t < \frac{n+1}{2^i} \tag{2.2.22}
\]
2.2 Dyadic Wavelet Transform

The elementary interval is divided by $1/2^i$ in order to ensure that the associated continuous-time functions remain compactly supported despite the fact that the length of the equivalent discrete-time filters is increasing with each iteration. The factor $2^{i/2}$ that multiplies the iterated discrete-time filters is necessary to preserve the $L^2$ norm between the discrete and continuous-time cases. Figure 2.5 illustrates the graphical function corresponding to the first four iterations of a length-4 Daubechies’
filter, indicating the piecewise constant approximation and the halving of the inter-
val. Perfect reconstruction together with orthogonality can be expressed using the
Smith-Barnwell condition as:

\[
\left| M (e^{j\omega}) \right|^2 + \left| M (e^{j(\omega+\pi)}) \right|^2 = 1 \tag{2.2.23}
\]

where \( M (e^{j\omega}) = \tilde{H} (e^{j\omega}) / \sqrt{2} \) is normalized such as \( M(1) = 1 \) and \( M(\pi) = 0 \). For reg-
ularity related purposes (a discrete-time filter is called regular if it converges through
the iteration scheme to a scaling function and to a wavelet with some degree of
regularity as piecewise smooth, continuous or derivable), the following condition is
imposed on \( M (e^{j\omega}) \) (the filters must have \( N \) zeros at \( \omega = \pi \)):

\[
M (e^{j\omega}) = \left[ \frac{1}{2} (1 + e^{j\omega}) \right]^N R (e^{j\omega}), \quad N \geq 1 \tag{2.2.24}
\]

Hence, \( |M (e^{j\omega})|^2 \) can be written as:

\[
\left| M (e^{j\omega}) \right|^2 = \left[ \cos^2 \left( \frac{\omega}{2} \right) \right]^N \left| R (e^{j\omega}) \right|^2 \tag{2.2.25}
\]

Since \( |R (e^{j\omega})|^2 = R (e^{j\omega}) \cdot R^* (e^{j\omega}) = R (e^{j\omega}) \cdot R (e^{-j\omega}) \), it can be expressed as a
polynomial in \( \cos^2 \left( \frac{\omega}{2} \right) \). Using the shorthand \( y = \cos^2 \left( \frac{\omega}{2} \right) \) and \( P (1 - y) = |R (e^{j\omega})|^2 \),
equation 2.2.23 can be written as:

\[
y^N P (1 - y) + (1 - y)^N P (y) = 1, \quad P (y) \geq 0 \text{ for } y \in [0, 1] \tag{2.2.26}
\]

Supposing that there exist a polynomial \( P(y) \) satisfying the previous condition and,
moreover:

\[
\sup_{\omega} \left| R (e^{j\omega}) \right| = \sup_{y \in [0,1]} |P(y)|^{\frac{1}{2}} < 2^{N-1} \tag{2.2.27}
\]
then there exist an orthonormal basis associated with \( \tilde{H} (e^{j\omega}) \), since the iterated filter
will converge to a continuous scaling function from which a wavelet basis can be
obtained. Daubechies [25, 26] showed that any polynomial $P$ solving equation 2.2.26 is of the form:

$$P(y) = \sum_{j=0}^{N-1} \binom{N - 1 + j}{j} y^j + y^N Q(y)$$  \hspace{1cm} (2.2.28)$$

where $Q$ is an antisymmetric polynomial. Furthermore Daubechies constructed filters of minimum order, i.e. $Q = 0$, called maximally flat filters (they have a maximum number of zeros at $\omega = \pi$). The $R$ is derived from $P$ using spectral factorization [96]. Figure 2.6 illustrates the iterated graphical scaling function and wavelet for $N=8$ (the
twelfth iteration is plotted). The corresponding filters coefficients are also illustrated (the filter coefficients for different values of $N$ are tabulated in [25, 96]).

### 2.3 Wavelet Shrinkage Principles

As already mentioned in this chapter, the joint time-frequency analysis effected by the WT provides natural settings for statistical applications, which include estimation, filtering and compression. Particularly, a considerable effort has been recently directed to develop asymptotically minimax methods based on the orthogonal wavelet transform in order to recover signals from noisy data [27, 28]. The theory underlying these methods exploits a correspondence between optimal recovery and the good compressibility characteristics of the wavelet transform. The wavelet based filtering methods basically comprise three different steps: the dyadic Wavelet Transform (WT) computation, the processing (shrinkage) of the wavelet coefficients and the reconstruction of the denoised signal from the processed wavelet coefficients through the inverse WT. As the first and the last steps have been briefly described previously, this section is intended to review the basic principles concerning the wavelet coefficients shrinkage.

We start from the following additive model (the case of multiplicative noise can be treated by the same techniques, assuming an adequate preprocessing step) of a discrete image $g$ and noise $\epsilon$:

$$f = g + \epsilon \quad (2.3.1)$$

In the equation above all terms are considered as vectors. Specifically, the image $g = [g_1, ..., g_n]$ that should be recovered is a deterministic signal, where the index refers to the spatial position exactly as is the case in raster scanning. The vector $f$ is the recorded image, while the noise $\epsilon$ is a vector of independent and identically
distributed (i.i.d.) random variables with distribution $N(0, \sigma^2)$. Due to the linearity of the wavelet transform and assuming that the data length is a power of 2, the DWT decomposes the signal into a set of wavelet coefficients:

\[ d = s + \xi, \quad (2.3.2) \]

where $d$ are the observed wavelet coefficients, $s$ are the noise free coefficients and $\xi$ is additive white noise.

### 2.3.1 Hard and Soft Thresholding

One of the most popular approach for signal denoising is wavelet thresholding, due to its simplicity. In its most basic form, this technique operates in the orthogonal wavelet domain, where the magnitude of all coefficients in the finest scale are set to zero. This is called the *projection estimator* and it can be successfully applied when the power of the target signal is concentrated in the lower frequency components (higher scales), while the noise is spread evenly across coefficients, and will dominate the high frequency components (lower scales). As this is a very restrictive assumption, more refined thresholding estimators, which selectively add in wavelet coefficients from finer scales have been designed. Specifically, within each wavelet detail, each coefficient is reduced after comparison against a threshold, i.e. if the coefficient is smaller than the threshold it set to zero, otherwise it is kept or modified. Two standard thresholding techniques exist: soft thresholding (“shrink or kill”), and hard thresholding (“keep or kill”). In both cases, the coefficients below a certain threshold are set to zero. In soft thresholding, the remaining coefficient are reduced by an amount equal to the value
Figure 2.7: Soft-thresholding estimator versus hard thresholding estimator

of the threshold:

\[ \hat{s} = T^{\text{soft}}_d(d) = \begin{cases} 
\text{sgn}(d)(|d| - t), & |d| > t \\
0, & |d| \leq t 
\end{cases} \]  
(2.3.3)

In hard-thresholding, the magnitudes of the wavelet coefficients above the threshold are left unchanged:

\[ \hat{s} = T^{\text{hard}}_d(d) = \begin{cases} 
d, & |d| > t \\
0, & |d| \leq t 
\end{cases} \]  
(2.3.4)

For comparison purposes, Figure 2.7 illustrates the two estimators described above.

Several threshold values were tested in recent studies [27, 28]. A simple choice of the threshold value \( t \) is based on the attempt to remove all wavelet coefficients that are pure noise. It uses the result of [57] which states that if the wavelet coefficients corresponding to the noise are i.i.d. samples with distribution \( N(0, \sigma^2) \), then:

\[ P(\max |\xi| \leq \sigma \sqrt{2 \log(n)}) \rightarrow 1, \]  
(2.3.5)

Hence, the universal threshold value \( t_u = \sqrt{2 \log(n)} \) proposed in [27] will set to zero...
all wavelet coefficients that contains noise and no signal. In soft thresholding, the estimates are biased: large coefficients are always reduced in magnitude; therefore the mathematical expectation of their estimates differ from the observed values. As a consequence, the reconstructed image is often oversmoothed. A smaller threshold value, which has been especially designed to adjust for some of the bias problems introduced by soft-thresholding, is the minimax-optimal threshold, proposed in [28]. However, a careful selection of the wavelet basis, thresholding procedure, and threshold value is a key-factor in each particular application.

2.3.2 Bayesian Wavelet Shrinkage

An alternate approach to the standard thresholding technique, which is less ad hoc since it relies on the knowledge of the wavelet coefficients statistics, makes use of Bayes rules. To use Bayesian methods, one depart from the classical approach to statistical estimation in which \( s \) is assumed to be a deterministic but unknown constant. Instead one assume that \( s \) is a random variable whose particular realisation one must estimate. If we have available some prior knowledge about \( s \), we can incorporate it into our estimator. The mechanism for doing this requires us to assume that \( s \) is a random variable with a given prior probability density function (PDF). The goal is to find the Bayes risk estimator \( \hat{s} \) that minimizes the conditional risk, which is the loss averaged over the conditional distribution of \( s \), given the set of wavelet coefficients, \( d \):

\[
\hat{s}(d) = \arg \min_{\hat{s}} \int L[s, \hat{s}(d)] P_{s|d}(s \mid d) \, ds
\]  

(2.3.6)

In order to minimize the above expression, the cost function \( L \) needs to be specified. Three typical cost functions are shown in Figure 2.8.

Figure 2.8 (a) illustrates the quadratic cost function \( L(s_e) = s_e^2 \). Under a quadratic
Figure 2.8: Typical cost functions: (a) quadratic error (b) absolute error, and (c) hit-or-miss error

cost function the Bayes risk estimator minimizes the mean-square error (MSE) and is given by the conditional mean of $s$, given $d$:

$$\hat{s}_{\text{mmse}}(d) = \int s \cdot P_s \mid d(s \mid d) \cdot ds$$

(2.3.7)

The quadratic cost function accentuates the effects of large errors. The resulting estimate is called the minimum mean square error (MMSE) estimate $\hat{s}_{\text{mmse}}$.

Another possible choice as cost function is the absolute error, $L(s_e) = |s_e|$, illustrated in Figure 2.8 (b). This cost function penalizes errors proportionally. The corresponding Bayesian estimator minimizes the mean absolute error and can be shown to be the conditional median of $s$, given $d$ [85].

By selecting the uniform cost function in Figure 2.8 (c):

$$L(s_e) = \begin{cases} 
0, & \text{for } |s - \hat{s}| < \delta \\
1, & \text{otherwise}
\end{cases}$$

(2.3.8)

the optimal estimator can be shown to be:

$$\hat{s}_{\text{map}}(d) = \arg \max_{\hat{s}} P_s \mid d(s \mid d)$$

(2.3.9)

The uniform cost function assigns zero cost to all errors less than $\pm \delta/2$. The estimator that minimizes the Bayes risk for the uniform (or “hit-or-miss”) cost function
is the mode (location of the maximum) of the posterior PDF. It is called the *maximum a posteriori* estimate. Taking in the account that, according to Bayes’ theorem
\[ P_{s|d}(s | d) = P_{d|s}(d | s) P_s(s) / P_d(d), \]
the estimator above can be also written as:
\[ \hat{s}_{map}(d) = \arg \max_{\hat{s}} P_{d|s}(d | s) P_s(s) \quad (2.3.10) \]

For some posterior PDFs these three estimators are identical, a notable example being the Gaussian posterior PDF. For the particular case of Gaussian signal with variance \( \sigma_s \), the Bayesian estimator can be obtained in closed form:
\[ \hat{s}(d) = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_d^2} d, \quad (2.3.11) \]

The solution is a linear rescaling of the measured values. When applied to the coefficients of a Fourier transform, this estimator corresponds to the Wiener filtering. When applied to wavelet details, the power spectral density information is averaged over each of the subbands. In the more general case when the noise distribution is Gaussian, but the signal prior is a more sharply peaked distribution, closed-form expressions for the Bayesian estimator may not be available, but numerical solutions that finally lead to non-linear shrinkage estimators may be used [20, 89, 97].
Chapter 3

The Alpha-Stable Family of Distributions

This section is intended to provide an introduction on the alpha-stable statistical model that will be used in Chapter 4 to characterize the wavelet subband coefficients of logarithmic transforms of ultrasound images, and of SAR images later in chapter 6. The model is suitable for describing signals that have highly non-Gaussian statistics and its parameters can be estimated from noisy observations. A review of the state of the art on stable processes from a statistical point of view is provided by a collection of papers edited by Cambanis, Samorodnitsky and Taqqu [18]. Several statisticians including Cambanis, Zolotarev, Weron, et al. have published extensively on the theory and applications of stable processes. They studied the properties of stable processes [13, 15, 43, 47, 49, 59, 64, 69, 86, 100, 105], their spectral representation [14, 42, 66], as well as prediction and linear filtering problems [16, 17, 48, 50, 82, 101]. Textbooks in the area were written by Samorodnitsky and Taqqu [83] and by Janicki and Weron [46]. An extensive review of stable processes from a signal processing point of view can be found in a tutorial paper by Shao and Nikias [87] as well as in a monogram written by the same authors [70].
3.1 Basic Properties of the Alpha-Stable Family

The appeal of alpha-stable distributions as a statistical model for signals derives from some important theoretical and empirical reasons.

First, stable random variables satisfy the stability property which states that linear combinations of jointly stable variables are indeed stable. Specifically, if $X$, $X_1$, and $X_2$ are $\alpha$-stable independent random variables with the same distribution and $a$ and $b$ are arbitrary constants, then there exist constants $c$ and $d$ such that

$$aX_1 + bX_2 \overset{d}{=} cX + d$$  \hspace{1cm} (3.1.1)

where the notation $\overset{d}{=}$ denotes equality in distributions. The word stable is used because the shape of the distribution is unchanged (or stable) under such linear combinations.

Second, stable processes arise as limiting processes of sums of i.i.d. random variables via the generalized central limit theorem, which states that: $X$ is $\alpha$-stable if and only if $X$ is the limit in distribution of the normalized sums of the form:

$$S_n = \frac{X_1 + X_2 + \ldots + X_n}{a_n} - b_n$$  \hspace{1cm} (3.1.2)

where $X_1, X_2, \ldots$, are i.i.d. random variables and $a_n \to \infty$.

Actually, the only possible non-trivial limit of normalized sums of i.i.d. terms is stable.

On the other hand, strong empirical evidence suggests that many data sets in several physical and economic systems exhibit heavy tail features that justify the use of stable models [8].

The $\alpha$-stable distribution is best defined by its characteristic function

$$\varphi(\omega) = \exp\{j\delta\omega - \gamma |\omega|^\alpha [1 + j\beta \text{sign}(\omega) \varpi(\omega, \alpha)]\}$$  \hspace{1cm} (3.1.3)
where
\[
\varpi(\omega, \alpha) = \begin{cases} 
\frac{\tan \frac{\alpha \pi}{2}}{2} & \text{if } \alpha \neq 1 \\
\frac{2}{\pi} \log |\omega| & \text{if } \alpha = 1 
\end{cases} 
\] (3.1.4)

and
\[
\text{sign}(\omega) = \begin{cases} 
1 & \text{for } \omega > 0 \\
0 & \text{for } \omega = 0 \\
-1 & \text{for } \omega < 0 
\end{cases} 
\] (3.1.5)

- \( \alpha \) is the characteristic exponent, taking values \( 0 < \alpha \leq 2 \)
- \( \delta \) \((-\infty < \delta < \infty)\) is the location parameter. For values of \( \alpha \) in the interval \((1, 2]\), the location parameter \( \delta \) corresponds to the mean of the distribution, while for \( 0 < \alpha \leq 1 \), \( \delta \) corresponds to its median.
- \( \gamma \) \((\gamma > 0)\) is the dispersion of the distribution and it determines the spread of the distribution around its location parameter \( \delta \), similar to the variance of the Gaussian distribution.
- \( \beta \) is the index of skewness taking values in the interval \(-1 \leq \beta \leq 1 \).

The characteristic exponent \( \alpha \) is the most important parameter of the \( \alpha \)-stable distribution and it determines the shape of the distribution. The smaller the characteristic exponent \( \alpha \) is, the heavier the tails of the stable density. This implies that random variables following alpha-stable distributions with small characteristic exponents are highly impulsive.

A stable distribution is called standard if \( \delta = 0 \) and \( \gamma = 1 \). Clearly, if a random variable \( X \) is stable with parameters \( \alpha, \beta, \gamma, \delta \), then \( (X - \delta)/\gamma^{1/\alpha} \) is standard with
The characteristic exponent $\alpha$ and skewness parameter $\beta$. The standard $\alpha$-stable density functions for a few values of the characteristic exponent $\alpha$ are shown in Figure 3.1. An important particular case of stable distributions is obtained for $\beta = 1$. In this case, the distribution is symmetric about the center $\delta$. Symmetric stable distributions with characteristic exponent $\alpha$ are called symmetric $\alpha$-stable or $S\alpha S$. Since in this thesis we are interested in modeling the distribution of wavelet coefficients of images, which are symmetric in nature, in the following we are only going to consider the $S\alpha S$ family.

### 3.2 The Class of Real $S\alpha S$ Distributions

The characteristic function of a $S\alpha S$ distribution is obtained simply by setting to zero the skewness parameter, $\beta$, in equation 3.1.3

$$\varphi(\omega) = \exp(j\delta \omega - \gamma |\omega|^\alpha),$$  \hspace{1cm} (3.2.1)
By letting $\alpha$ take the values 1 and 2, we get two important special cases of $S\alpha S$ distributions, namely, the Cauchy ($\alpha = 1$), and the Gaussian ($\alpha = 2$):

**Cauchy**

$$f_1(\gamma, \delta; x) = \frac{1}{\pi \gamma^2 + (x - \delta)^2}$$  \hspace{1cm} (3.2.2)

**Gaussian**

$$f_2(\gamma, \delta; x) = \frac{1}{\sqrt{4\pi \gamma}} \exp \left[-\frac{(x - \delta)^2}{4\gamma}\right].$$ \hspace{1cm} (3.2.3)

Unfortunately, no closed form expressions exist for general $S\alpha S$ distributions other than the Cauchy and the Gaussian. However, power series expansions can be derived for $f_\alpha(\gamma, \delta; x)$. In the following, we shall assume that all $S\alpha S$ distributions are centered at the origin, i.e., the location parameter $\delta = 0$. This is equivalent to the zero-mean assumption for Gaussian distributions. Then, the standard $S\alpha S$ density function is given by [87]

$$f_\alpha(x) = \begin{cases} 
\frac{1}{\pi x} \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k!} \Gamma(\alpha k + 1) x^{-\alpha k} \sin\left(\frac{k\pi x}{2}\right) & \text{for } 0 < \alpha < 1 \\
\frac{1}{\pi \alpha} \left(\sum_{k=0}^{\infty} \frac{(-1)^k}{2k!} \Gamma\left(\frac{2k+1}{\alpha}\right) x^{2k}\right) & \text{for } \alpha = 1 \\
\frac{1}{2\sqrt{\pi}} \exp\left[-\frac{x^2}{4}\right] & \text{for } 1 < \alpha < 2 \\
\frac{1}{2\sqrt{\pi}} \exp\left[-\frac{x^2}{4}\right] & \text{for } \alpha = 2.
\end{cases}$$ \hspace{1cm} (3.2.4)

The $S\alpha S$ density behaves approximately like a Gaussian density near the origin, however its tails decay at a lower rate than the Gaussian density tails [83]. Indeed, let $X$ be a non-Gaussian $S\alpha S$ random variable. Then, as $x \to \infty$

$$P(X > x) \sim c_\alpha x^{-\alpha}$$ \hspace{1cm} (3.2.5)

where $c_\alpha = \Gamma(\alpha)(\sin\frac{\pi \alpha}{2})/\pi$, $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t} dt$ is the Gamma function, and the statement $h(x) \sim g(x)$ as $x \to \infty$ means that $\lim_{x \to \infty} h(x)/g(x) = 1$. Hence, the tail probabilities are asymptotically power laws. In other words, while the Gaussian
density has exponential tails, the stable densities have algebraic tails. Figure 3.2 shows the tail behavior of several $S\alpha S$ densities including the Cauchy and the Gaussian. We should note that because expression (3.2.5) gives exactly the tail probability of the Pareto distribution, the term “stable Paretian laws” is used to distinguish between the fast decay of the Gaussian law and the Pareto like tail behavior when $\alpha < 2$. It is this heavy-tail characteristic that makes the $S\alpha S$ densities appropriate for modeling signals and noise or interference which are impulsive in nature.

One consequence of heavy tails is that only moments of order less than $\alpha$ exist for the non-Gaussian alpha-stable family members, i.e.,

$$E|X|^p < \infty \text{ for } p < \alpha$$

and the so-called fractional lower order moments (FLOM’s) of a $S\alpha S$ random variable
with zero location parameter and dispersion $\gamma$ are given by:

$$E|X|^p = C(p, \alpha)\gamma^{\frac{p}{\alpha}} \text{ for } 0 < p < \alpha$$  \hspace{1cm} (3.2.7)

where

$$C(p, \alpha) = \frac{2^{p+1}\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(-\frac{p}{\alpha}\right)}{\alpha\sqrt{\pi}\Gamma\left(-\frac{p}{2}\right)}$$  \hspace{1cm} (3.2.8)

and $\Gamma(\cdot)$ is the Gamma function. As a result, stable Pareto laws have infinite variance. In the past, the infinite variance property of the $S\alpha S$ family has caused skeptics to dismiss the stable model. With the same reasoning, one could argue that the routinely used Gaussian distribution, which has infinite support, should also be dismissed as a model of bounded measurements. In practice, one should remember that it is important to capture the shape of the distribution and that the variance is only one measure of the spread of a density [71].

### 3.3 Bivariate Isotropic Stable Distributions

Bivariate stable distributions, much like the univariate stable distributions, are characterized by the stability property and the generalized central limit theorem. However, they are much more difficult to describe because they form a nonparametric set [87]. An exception is the family of multidimensional isotropic stable distributions.

The characteristic function of a bivariate isotropic $\alpha$-stable distribution has the form

$$\varphi(\omega_1, \omega_2) = \exp(j(\delta_1\omega_1 + \delta_2\omega_2) - \gamma|\omega|^\alpha),$$  \hspace{1cm} (3.3.1)

where $\omega = (\omega_1, \omega_2)$ and $|\omega| = \sqrt{\omega_1^2 + \omega_2^2}$.

Again here, the parameters $\alpha$ and $\gamma$ are the characteristic exponent and the dispersion, respectively. The parameters $\delta_1, \delta_2$ are the location parameters. The distribution
is isotropic with respect to the point \((\delta_1, \delta_2)\). Note that the two marginal distributions of the isotropic stable distribution are \(S_{\alpha\gamma}\) with parameters \((\delta_1, \gamma, \alpha)\) and \((\delta_2, \gamma, \alpha)\).

In the following, we will assume that \((\delta_1, \delta_2) = (0, 0)\). The bivariate isotropic Cauchy and Gaussian distributions are special cases for \(\alpha = 1\) and \(\alpha = 2\), respectively.

As in the case of the univariate \(S_{\alpha\gamma}\) density function, when \(\alpha \neq 1\) or \(\alpha \neq 2\), no closed form expressions exist for the density function of the bivariate stable random variable. By using the polar coordinate \(r = |x| = \sqrt{x_1^2 + x_2^2}\), the density function can be written as \(f_{\alpha, \gamma}(x_1, x_2) = \chi_{\alpha, \gamma}(r)\), and can be expressed in a power series expansion form:

\[
\chi_{\alpha, \gamma}(r) = \begin{cases} 
\frac{1}{\pi^{2\gamma/\alpha}} \sum_{k=1}^{\infty} \frac{2^{\alpha k}(\Gamma(\alpha k/2 + 1))^2 \sin(k \alpha \pi/2 \gamma)}{k! (\Gamma(k/\alpha))^{\alpha k}} r^{-\alpha k-2} & \text{for } 0 < \alpha < 1 \\
\frac{1}{\pi \alpha^{2\gamma/\alpha}} \sum_{k=0}^{\infty} \frac{(-1)^k}{2^{2k+1} (k!)^2} \Gamma(2k+2/\alpha) (r^{-\alpha k})^{2k} & \text{for } 1 < \alpha < 2 \\
\frac{1}{4\pi \gamma} \exp\left(-\frac{r^2}{4\gamma}\right) & \text{for } \alpha = 2.
\end{cases}
\]

The density function \(\chi_{\alpha, \gamma}(r)\) described above is also a heavy-tailed function. An expression for the FLOM’s, similar to the one for the single-dimensional case, can be found in [92, 105]. If \(X\) is a random vector in \(\mathbb{R}^n\) having the isotropic stable distribution with dispersion \(\gamma\), then

\[
E|X|^p = C_n(p, \alpha) \gamma^{\frac{p}{\alpha}} \quad \text{for } 0 < p < \alpha
\]

where

\[
C_n(p, \alpha) = \frac{2^{p+1} \Gamma\left(\frac{p+n}{2}\right) \Gamma\left(-\frac{p}{\alpha}\right)}{\alpha \Gamma\left(\frac{n}{2}\right) \Gamma\left(-\frac{p}{\alpha}\right)}.
\]

### 3.4 Symmetric Alpha-Stable Processes

A collection of random variables \(\{X(t), \ t \in T\}\) where \(T\) is an arbitrary index set, is said to be a \(S_{\alpha\gamma}\) stochastic process if for all combinations of distinct indices
3.4 Symmetric Alpha-Stable Processes

t_1, \ldots, t_n \in T$, the random variables $X(t_1), \ldots, X(t_n)$ are jointly $S\alpha S$ with the same characteristic exponent $\alpha$. The stochastic process $\{X(t), \ t \in T\}$ is stationary if the random vectors $(X(t_1), \ldots, X(t_n))$ and $(X(t_1 + s), \ldots, X(t_n + s))$ are identically distributed for each choice of $s, t_1, \ldots, t_n \in T$. The family of stable processes has many members with mutually exclusive properties. In the following, we present three important types of stable processes that are commonly used.

1) Sub-Gaussian Processes: A stable process $\{X(t), \ t \in T\}$ is said to be an $\alpha$-sub-Gaussian process ($\alpha$-SG($R$)) if for distinct indices $t_1, \ldots, t_n \in T$, $(X(t_1), \ldots, X(t_n))$ has characteristic function given by

$$\varphi(\omega) = \exp\left(-\frac{1}{2} \sum_{l,m=1}^{n} \omega_l \omega_m R(t_l, t_m)^{\alpha/2}\right), \quad (3.4.1)$$

where $R(t, s)$ is a positive definite function, $\omega = [\omega_1, \ldots, \omega_n]^T$ and $\alpha$ takes values in $(1, 2]$. When $\alpha = 2$, $X(t)$ is a Gaussian process with zero mean and covariance function $R(t, s)$. A sub-Gaussian process is stationary if and only if $R(t, s) = R(t - s) = R(s - t)$.

Sub-Gaussian processes share many common features with Gaussian processes. In fact, sub-Gaussian processes are variance mixtures of Gaussian processes [17]. Specifically, if $X(t)$ is $\alpha$-SG($R$), then

$$X(t) = S^{1/2}Y(t) \quad (3.4.2)$$

where $S$ is a positive $\frac{\alpha}{2}$-stable random variable and independent of $Y(t)$ which is a Gaussian process with zero-mean and covariance function $R(t, s)$. An important distinction between Gaussian and sub-Gaussian processes is that, while linear spaces of Gaussian random variables may contain non-degenerate independent elements, sub-Gaussian random variables cannot be independent [19]. A sub-Gaussian process
The Alpha-Stable Family of Distributions

42

\( X(t) = S^{1/2}Y(t) \) is stationary if and only if the Gaussian process \( Y(t) \) is stationary.

2) Linear Stable Processes: Let \( \{U(n), \; n = 0, \pm 1, \pm 2, \ldots\} \) be a family of i.i.d. \( S_\alpha S \) random variables. Then,

\[
X(n) = \sum_{i = -\infty}^{+\infty} a_i U(n - i)
\]

defines a stationary \( S_\alpha S \) random process if \( \sum_{-\infty}^{+\infty} |a_i|^{\alpha - \delta} < \infty \) for some \( 0 < \delta < \alpha \) when \( 0 < \alpha < 1 \), or if \( \sum_{-\infty}^{+\infty} |a_i| < \infty \) when \( \alpha \geq 1 \). These processes are called linear stable processes or stable processes with moving-average representation. Examples of linear stable processes include finite-order autoregressive (AR), moving-average (MA) and autoregressive moving-average (ARMA) processes [87].

3) Harmonizable Stable Processes: A complex valued \( S_\alpha S \) process \( X(t) \) is called harmonizable if it can be expressed as:

\[
X(t) = \int_{-\infty}^{\infty} e^{i\tau \omega} d\xi(\omega); \quad -\infty < t < \infty,
\]

where \( d\xi(\omega) \) is a \( S_\alpha S \) process with independent increments satisfying

\[
\{E|d\xi(\omega)|^p\}^{\alpha/p} = C(p, \alpha)\phi(\omega) \; d\omega \quad \text{for all} \; \; 0 < p < \alpha,
\]

and \( C(p, \alpha) \) is a constant depending on \( p \) and \( \alpha \) and \( \phi(\omega) \) is a nonnegative function called the spectral density of \( X(t) \) [42]. The spectral density \( \phi(\omega) \) describes fully the distribution of the process \( X(t) \). In another sharp contrast with the Gaussian case, the class of \( S_\alpha S \) harmonizable processes is disjoint from the class of linear processes. In addition, sub-Gaussian processes are neither linear nor harmonizable [19].

In modeling the signals and/or noise for the Bayesian estimation problem, we need efficient methods for parameter estimation of alpha-stable distributions. In the following section, we present several methods that have been proposed in the past to achieve this task.
3.5 Parameter Estimation for $S\alpha S$ Distributions

The alpha-stable tail power law provided one of the earliest approaches in estimating the stability index $\alpha$ of real measurements [83]. The empirical distribution of the data, plotted on a log-log scale, should approach a straight line with slope $-\alpha$ if the data is stable. Another approach is based on quantiles [30]. Maximum likelihood (ML) methods developed by DuMouchel [29] and by Brorsen and Yang [12] are asymptotically efficient but were considered difficult to compute.

3.5.1 Maximum Likelihood Method

DuMouchel [29] obtained approximate maximum likelihood estimates of $\alpha$ and $\gamma$ for $\delta = 0$. The standard $S\alpha S$ density function is given by [104]:

$$f_\alpha(x) = \frac{\alpha}{|1-\alpha|\pi} x^{1/(\alpha-1)} \int_0^{\pi/2} v(\theta) e^{-x^{\alpha/(\alpha-1)} v(\theta)} d\theta$$

for $\alpha \neq 1$, $x > 0$  \( (3.5.1) \)

where

$$v(\theta) = \frac{1}{(\sin \alpha \theta)^{\alpha/(\alpha-1)}} \cos[(\alpha - 1)\theta](\cos \theta)^{1/(\alpha-1)}$$

and

$$f_1(x) = \frac{1}{\pi(1 + x^2)}$$

$$f_\alpha(0) = \frac{1}{\pi} \Gamma((\alpha + 1)/\alpha)$$

$$f_2(x) = \frac{1}{2\sqrt{\pi}} e^{-x^2/4}$$

(3.5.3)
Consequently, the parameters $\alpha$, $\delta$, $c$ can be estimated from the observations $x_1$, $x_2$, ..., $x_n$ by maximization of the following log-likelihood function:

$$
\sum_{i=1}^{n} \log[f_{\alpha}(z_i)] = n \log \alpha - n \log(\alpha - 1) + \sum_{i=1}^{n} \left( \log z_i \right)/(\alpha - 1) 
$$

$$
+ \sum_{i=1}^{n} \log \int_{0}^{\pi/2} v(\theta) e^{-z_i^{\alpha/(\alpha-1)}v(\theta)} d\theta
$$

where

$$z_i = |x_i - \delta|/c$$

Based on (3.5.4), Brorsen and Yang [12] performed Monte-Carlo simulations and obtained fairly good results. The disadvantage of this method is that it is a highly nonlinear optimization problem and no initialization and convergence analysis is available.

### 3.5.2 Method of Sample Quantiles

Fama and Roll [30] proposed a method for estimating the parameters of $S\alpha S$ distributions with $1 \leq \alpha \leq 2$, which is based on order statistics. Their suggested estimate for $c$ is given by

$$
\hat{c} = \frac{1}{1.654} [\hat{x}_{0.72} - \hat{x}_{0.28}]
$$

where $\hat{x}_f$ ($f = 0.72$, $0.28$) is the estimated f quantile of the $S\alpha S$ distribution. A consistent estimate of the f quantile, $\hat{x}_f$, is usually the $(N+1)f$ st order statistic, where N is the number of observations.

On the other hand, $\alpha$ can be estimated from the tail behavior of the distribution. Specifically, for some large $f$ (e.g. $f = 0.95$), one first calculate

$$
\hat{z}_f = \frac{\hat{x}_f - \hat{x}_{1-f}}{2\hat{c}} = 0.827 \frac{\hat{x}_f - \hat{x}_{1-f}}{\hat{x}_{0.72} - \hat{x}_{0.28}}
$$
3.5 Parameter Estimation for $S\alpha S$ Distributions

from the sample. Taking in the account that the population distribution is $S\alpha S$ with characteristic exponent $\alpha$ and dispersion $\gamma = c^\alpha$, $\hat{z}_f$ is an estimate of the $f$ quantile of the standard $S\alpha S$ distribution. Thus, an estimate $\hat{\alpha}_f$ can be obtained by searching a table of standard $S\alpha S$ distribution functions.

Since a $S\alpha S$ distribution has a finite mean for $1 < \alpha \leq 2$, the sample mean constitute a consistent estimate of the location parameter $\delta$. A more robust estimate is the truncated mean. A $p$ percent truncated sample mean is the arithmetic mean of the middle $p$ percent of the ranked observations.

Famme-Roll’s method is simple but suffers from a small asymptotic bias and is not asymptotically efficient.

3.5.3 Method of Sample Characteristic Function

The sample characteristic function is defined as

$$\hat{\Phi}(\omega) = \frac{1}{N} \sum_{k=1}^{N} \exp(j\omega x_k) \quad (3.5.7)$$

where $N$ is the sample size and $x_1, ..., x_n$ are the observations. It is a consistent estimator of the true characteristic function that uniquely determines the density function. It is also important to note that $\hat{\Phi}(\omega)$, $-\infty < \omega < \infty$ is a stochastic, non-stationary process with the property that $0 < |\hat{\Phi}(\omega)| \leq 1$. Consequently, all moments of $\hat{\Phi}(\omega)$ are finite.

Among the different estimation methods based on characteristic function that have been proposed in the literature, one should refer to the method of Paulson, Holcomb, and Leitch [74], and the regression methods of Koutrouvelis [53] and Kogon [52].

Koutrouvelis’ method is based on the following relations between the characteristic
function of a $S\alpha S$ distribution and its parameters

$$\log(-\log |\Phi(\omega)|^2) = \log(2c^\alpha) + \alpha \log |\omega|$$  \hspace{1cm} (3.5.8)

and

$$\frac{\text{Im } \Phi(\omega)}{\text{Re } \Phi(\omega)} = \tan \delta t$$  \hspace{1cm} (3.5.9)

The parameters $\alpha$ and $c$ can be estimated from the linear regression

$$y_k = \mu + \alpha w_k + \varepsilon_k, \quad k = 1, 2, ..., K$$  \hspace{1cm} (3.5.10)

where $y_k = \log(-\log |\hat{\Phi}(\omega_k)|^2)$, $\mu = \log(2c^\alpha)$, $w_k = \log(\omega_k)$, $\varepsilon_k$ is an error term and $\omega_1, ..., \omega_k$ is an appropriate set of real numbers.

In a similar way, the location parameter $\delta$ can be estimated through the linear regression

$$z_l = \delta u_l + \varepsilon_l, \quad l = 1, 2, ..., L$$  \hspace{1cm} (3.5.11)

where $z_l = \arctan(\text{Im}(\hat{\Phi}(u_l))/\text{Re}(\hat{\Phi}(u_l)))$, and $u_1, ..., u_l$ is an appropriate set of real numbers.

The regression estimators $\hat{\alpha}$, $\hat{c}$, and $\hat{\delta}$ described above are consistent and asymptotically unbiased. According to Koutrouvelis [53], the regression method is better than both DuMouchel’s approximate maximum likelihood method and Famma-Roll’s quantile method. It is a method that involves low computational cost and easy to implement.
Chapter 4

Wavelet-based Ultrasound Image Denoising using an Alpha-Stable Prior Probability Model

In this chapter, a novel speckle suppression method for medical ultrasound images is presented. First, the logarithmic transform of the original image is analyzed into the multiscale wavelet domain. We show that the subband decompositions of ultrasound images have significantly non-Gaussian statistics that are best described by families of heavy-tailed distributions such as the alpha-stable. Then, we design a Bayesian estimator that exploits these statistics. We use the alpha-stable model to develop a blind noise-removal processor that performs a non-linear operation on the data. Finally, we compare our technique to current state-of-the-art soft and hard thresholding methods applied on actual ultrasound medical images and we quantify the achieved performance improvement.
4.1 Problem Formulation

Speckle noise affects all coherent imaging systems including laser, SAR imagery, and ultrasound. Speckle may appear distinct in different imaging systems but it is always manifested in a granular pattern due to image formation under coherent waves. The basic properties of speckle are described by Goodman in [38] while the main differences between ultrasound and laser speckle are discussed in [1]. A general model for speckle noise proposed by Jain [45] was also used by Zong [106]. In the following, we formulate the ultrasound speckle removal problem starting with a brief essential overview of the speckle model.

Denote by $I(x, y)$ a noisy observation (i.e. the recorded ultrasound image) of the two-dimensional function $S(x, y)$ (i.e. the noise-free image that has to be recovered) and by $\eta_m(x, y)$ and $\eta_a(x, y)$ the corrupting multiplicative and additive speckle noise components, respectively. One can write:

$$I(x, y) = S(x, y) \cdot \eta_m(x, y) + \eta_a(x, y), \quad (x, y) \in \mathbb{Z}^2 \quad (4.1.1)$$

Generally, the effect of the additive component of the speckle in ultrasound images is less significant than the effect of the multiplicative component. Thus, ignoring the term $\eta_a(x, y)$, one can rewrite (4.1.1) as

$$I(x, y) = S(x, y) \cdot \eta_m(x, y) \quad (4.1.2)$$

To transform the multiplicative noise model into an additive one, we apply the logarithmic function on both sides of (4.1.2):

$$\log I(x, y) = \log S(x, y) + \log \eta_m(x, y). \quad (4.1.3)$$
Figure 4.1: Block diagram of the proposed multiscale homomorphic Bayesian-based algorithm for speckle suppression. Our proposed novel wavelet coefficient statistical characterization and Bayesian processing modules result in a more accurate ultrasound image reconstruction.

Expression (4.1.3) can be rewritten as

\[ f(x, y) = g(x, y) + \epsilon(x, y), \tag{4.1.4} \]

where \( f(\cdot), g(\cdot), \) and \( \epsilon(\cdot) \) are the logarithms of \( I(\cdot), S(\cdot), \) and \( \eta_m(\cdot) \), respectively. In fact, this logarithmic transform constitutes the first preprocessing step of our proposed algorithm as shown in the block diagram depicted in Figure 4.1.

At this stage, one can consider \( \epsilon(x, y) \) to be white noise and subsequently apply any conventional additive noise suppression technique, such as Wiener filtering. However, it is recognized that standard noise filtering methods often result in blurred image features. Indeed, single-scale representations of signals, either in time or in frequency, are often inadequate when attempting to separate signals from noisy data. The wavelet transform has been proposed as a useful processing tool for signal recovery [67, 25, 62].

The wavelet transform expands a signal using a set of basis functions, which are obtained from a single prototype function called the “mother wavelet.” The result of the expansion is a sequence of signal approximations at successively coarser resolutions. The so-called “detail signal” is the difference in information between approximations at two consecutive resolutions, and it can be represented by another
series expansion. If we consider an original two-dimensional signal of size $N \times N$, $N$ usually being a power of 2 ($N = 2^j$), such a decomposition scheme is mathematically referred to as the dyadic wavelet transform (DWT). In image processing applications, the above scheme is applied along both the abscissa and the ordinate. Thus, the DWT decomposes images with a multiresolution scale factor of two, providing at each resolution level one low-resolution approximation and three spatially oriented wavelet details [62, 63], which are referred as image subbands.

The wavelet transform is a linear operation. Consequently, after applying the DWT to equation (4.1.4) we get, in each of the three directions, sets of noisy wavelet coefficients written as the sum of the transformations of the signal and of the noise:

$$d_{j,k}^i = s_{j,k}^i + \xi_{j,k}^i,$$  \hspace{1cm} (4.1.5)

where $k = 0, ..., 2^{J+j} - 1$ and $-1 < j < -J$ refer to the decomposition level or scale and $i = 1, 2, 3$ refers to the three spatial orientations. In Figure 4.2, we show an example of a three-scale decomposition of an ultrasound image.

Current state-of-the-art multiscale techniques for ultrasonic speckle suppression are based on various thresholding schemes [106, 41]. These methods try to address the inability of the original soft thresholding technique to balance between speckle suppression and signal detail preservation. In principle, a successful ultrasound imaging algorithm should achieve both noise reduction and feature preservation if it takes into consideration the true statistics of the signal and noise components. Previous studies related to wavelet shrinkage using Bayesian theory have underlined the need for a prior model that accurately approximates the probability density function of the signal and noise wavelet coefficients [89, 78, 20, 97]. For example, equations (4.1.1)-(4.1.4) suggest the use of a multiplicative random field as speckle noise model. It has
been shown that, if the number of scatterers per resolution cell is large, a fully developed speckle pattern can be modeled as the magnitude of a complex Gaussian field with independent and identically distributed (i.i.d.) real and imaginary components (see [38], [24] and references therein). In order to generate spatially correlated speckle noise for use in simulations, one can lowpass filter a complex Gaussian random field and take the magnitude of the filtered output [84].

The Bayesian approach for ultrasound speckle noise removal, which we propose in this chapter, is based on the novel heavy-tailed family of distributions that we
introduced in Chapter 3. In the next section we justify its use by fitting actual ultrasonic signals.

4.2 Alpha-Stable Modeling of Ultrasound Wavelet Coefficients

In the past, several authors have pointed out that, in a subband representation of images, histograms of wavelet coefficients have heavier tails and more sharply peaked modes at zero than what is assumed by the Gaussian distribution [88, 94, 63]. In this section, we study whether the stable family provides a flexible and appropriate tool for modeling the coefficients within the framework of multiscale wavelet analysis of logarithmically transformed ultrasound images. The appearance of alpha-stable models in the context of ultrasound images has been already noticed in [55, 56], but they were used to process the ultrasound RF echoes, rather than the recorded images.

Two sets of test images, obtained from two different sources, are included in this research. The first set consists of a series of 44 abdominal ultrasound images (DICOM format) including liver, kidney, gall bladder, and pancreas images. These images were acquired from the same patient with a 4-MHz transducer frequency on a GE LOGIQ 500 system. They have been made available to us by the IT Lab at the Medical University of South Carolina. The second set of test data comes from a directory containing example DICOM image files that were donated by various vendors for the DICOM demonstrations held at the annual meetings of the Radiological Society of North America from 1993 to 1996 (ftp://wuerlim.wustl.edu/pub/dicom/images/). We considered two criteria for selecting images for our test set. First, we looked for
good quality images in order to be able to consider them as noise-free. Moreover, since speckle appears to some extent in any ultrasound image, we have first processed the actual images using the homomorphic Wiener filter [45] and considered the resulting images as reasonable approximations of the speckle free data. Also, we were interested in performing experiments on images of different organs and from various sources in order to be able to obtain modeling results, which we can claim to be general enough.

In the following we describe the modeling of eight representative images. Each image is referred with the name of the organ that was imaged. All of them have a 256 gray-level resolution.

We proceed in two steps: First, we assess whether the data deviate from the normal distribution and if they have heavy tails. To determine that, we make use of the normal probability plots. Then, we check if the data is in the stable domain of attraction by estimating the characteristic exponent, $\alpha$, directly from the data and by providing the related confidence intervals. Several methods have been proposed for estimating stable parameters. Here, we use the maximum likelihood method described by Nolan in [71], which gives reliable estimates and provides the most tight confidence intervals. As further stability diagnostics, we employ probability density plots that give a good indication of whether the $S\alpha S$ fit matches the data near the mode and on the tails of the distribution.

Figure 4.3 depicts the normal probability plot of the vertical subband at the first level of decomposition of the gallbladder image (data denoted by “GLBD_Vsbd_1lvl,” for short). The plot provides strong evidence that the underlying distribution is not normal. The circles in the plot show the empirical probability versus the data value for each point in the sample. The circles are in a curve that does not follow the
Figure 4.3: Normal probability plot of the vertical subband at the first level of decomposition of the gallbladder image (GLBD_Vsbd_1lvl data for short). Characterization of data non-Gaussianity. The “◦” marks correspond to the empirical probability density versus the data value for each point in the sample. Since the circles are in a curve that does not follow the straight Gaussian line, the normality assumption is violated for this data.

straight Gaussian line and thus, the normality assumption is violated for this data.

While non-Gaussian stable densities are heavy-tailed, not all heavy-tailed distributions are stable. Hence, in Figures 4.4 and 4.5 we assess the stability of the data. First, the characteristic exponent is estimated and the data sample is fitted with the corresponding stable distribution. For the particular case shown here, the characteristic exponent of the $S\alpha S$ distribution which best fits the data was estimated to be $\hat{\alpha} = 1.069$. The stabilized p-p $S\alpha S$ plot in Figure 4.4 shows a highly accurate stable fit for this data set.

Naturally, the real question is whether the stable fit describes the data more
4.2 Alpha-Stable Modeling of Ultrasound Wavelet Coefficients

Figure 4.4: Stabilized p-p plot for $S\alpha S$ fit of data set GLBDVsbd_1lvl. The “+” marks, denoting the empirical probability density, are in a curve that very accurately follows the straight $S\alpha S$ line corresponding to $\alpha = 1.069$.

accurately than other PDF functions proposed in the literature. Here, we compare the $S\alpha S$ fits with those provided by the generalized Laplacian density function proposed by Mallat in [63] and also used by Simoncelli in [88]:

$$f_{s,p}(c) = \frac{e^{-|c/s|^p}}{Z(s,p)}$$ (4.2.1)

where $Z(s, p) = 2 \frac{s}{p} \Gamma(\frac{1}{p})$. The parameters $s$ and $p$ can be computed from the second and fourth moments of the data:

$$\sigma^2 = \frac{s^2 \Gamma(\frac{3}{p})}{\Gamma(\frac{1}{p})}, \quad k = \frac{\Gamma(\frac{1}{p}) \Gamma(\frac{5}{p})}{\Gamma^2(\frac{3}{p})}$$ (4.2.2)

where $\sigma^2$ is the distribution variance, and $k$ is the kurtosis. Figure 4.5 shows that the $S\alpha S$ distribution is superior to the generalized Laplacian distribution because it
Figure 4.5: Modeling of the ultrasound image wavelet coefficients GLBD_Vsbd_1lvl with the $S\alpha S$ and the generalized Laplacian density functions, depicted in solid and dashed lines, respectively. The $S\alpha S$ distribution has characteristic exponent $\alpha = 1.069$ and dispersion $\gamma = 0.051$ while the generalized Laplacian has parameters $p = 0.498$ and $s = 0.022$ (cf. Equation 4.2.1). The dotted line denotes the empirical PDF. Note that the $S\alpha S$ PDF provides a better fit to both the mode and the tails of the empirical density of the actual data.

For every image we iterated three times the separable wavelet decomposition described in Section 4.1 and we modeled the coefficients of each subband by using the $S\alpha S$ family. The wavelet decomposition was done using Daubechies’ Symmlet 4 basis wavelet because we found this basis to be the most effective in decorrelating the data. The results are summarized in Table 4.1, which shows the ML estimates of the characteristic exponent $\alpha$ together with the corresponding 95% confidence intervals. It can be observed that the confidence interval depends on the size of the images and
Table 4.1: Alpha-stable modeling of wavelet subband coefficients of actual ultrasound images. Maximum likelihood parameter estimates and 95% confidence intervals for the $\alpha$ characteristic exponent, $\alpha$. The tabulated key parameter $\alpha$ defines the degree of non-Gaussianity as deviations from the value $\alpha = 2$, which corresponds to the Gaussian condition. The size of each image is given in parentheses.

<table>
<thead>
<tr>
<th>IMAGE</th>
<th>Level</th>
<th>Image Subbands</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Horizontal</td>
<td>Vertical</td>
<td>Diagonal</td>
<td></td>
</tr>
<tr>
<td>breast (361 × 361)</td>
<td>1</td>
<td>1.279 ± 0.016</td>
<td>0.965 ± 0.013</td>
<td>1.128 ± 0.015</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.380 ± 0.032</td>
<td>1.178 ± 0.028</td>
<td>1.248 ± 0.031</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.349 ± 0.059</td>
<td>1.073 ± 0.053</td>
<td>1.303 ± 0.058</td>
<td></td>
</tr>
<tr>
<td>gallbladder (256 × 256)</td>
<td>1</td>
<td>1.349 ± 0.022</td>
<td>1.069 ± 0.020</td>
<td>0.974 ± 0.019</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.516 ± 0.044</td>
<td>1.265 ± 0.043</td>
<td>1.352 ± 0.040</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.508 ± 0.081</td>
<td>1.295 ± 0.077</td>
<td>1.181 ± 0.075</td>
<td></td>
</tr>
<tr>
<td>kidney (293 × 293)</td>
<td>1</td>
<td>1.382 ± 0.019</td>
<td>1.121 ± 0.018</td>
<td>1.124 ± 0.017</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.494 ± 0.038</td>
<td>1.308 ± 0.038</td>
<td>1.427 ± 0.039</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.348 ± 0.071</td>
<td>1.126 ± 0.067</td>
<td>1.495 ± 0.074</td>
<td></td>
</tr>
<tr>
<td>liver (256 × 256)</td>
<td>1</td>
<td>1.469 ± 0.020</td>
<td>1.197 ± 0.021</td>
<td>1.269 ± 0.022</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.482 ± 0.045</td>
<td>1.254 ± 0.042</td>
<td>1.546 ± 0.045</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.112 ± 0.072</td>
<td>1.008 ± 0.068</td>
<td>1.391 ± 0.080</td>
<td></td>
</tr>
<tr>
<td>pancreas (230 × 230)</td>
<td>1</td>
<td>1.443 ± 0.023</td>
<td>1.159 ± 0.023</td>
<td>1.264 ± 0.023</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.491 ± 0.050</td>
<td>1.230 ± 0.045</td>
<td>1.523 ± 0.050</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.237 ± 0.085</td>
<td>0.936 ± 0.071</td>
<td>1.501 ± 0.091</td>
<td></td>
</tr>
<tr>
<td>spinal cord (256 × 256)</td>
<td>1</td>
<td>1.253 ± 0.022</td>
<td>1.110 ± 0.020</td>
<td>1.196 ± 0.021</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.416 ± 0.044</td>
<td>1.298 ± 0.043</td>
<td>1.528 ± 0.045</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.424 ± 0.081</td>
<td>1.116 ± 0.072</td>
<td>1.487 ± 0.082</td>
<td></td>
</tr>
<tr>
<td>urinary bladder (190 × 190)</td>
<td>1</td>
<td>1.167 ± 0.027</td>
<td>1.008 ± 0.026</td>
<td>1.002 ± 0.027</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.279 ± 0.056</td>
<td>1.167 ± 0.054</td>
<td>1.141 ± 0.053</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.408 ± 0.101</td>
<td>1.006 ± 0.088</td>
<td>0.958 ± 0.085</td>
<td></td>
</tr>
<tr>
<td>spline (256 × 256)</td>
<td>1</td>
<td>1.417 ± 0.020</td>
<td>1.104 ± 0.020</td>
<td>1.178 ± 0.021</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.470 ± 0.041</td>
<td>1.046 ± 0.038</td>
<td>1.306 ± 0.043</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.154 ± 0.073</td>
<td>0.922 ± 0.064</td>
<td>1.197 ± 0.075</td>
<td></td>
</tr>
</tbody>
</table>
on the particular level of decomposition. The confidence interval becomes larger as the size decreases and as the level increases since the number of samples used for estimating $\alpha$ gets smaller. The table demonstrates that the coefficients of different subbands and decomposition levels exhibit various degrees of non-Gaussianity. The important observation is that all subbands exhibit distinctly non-Gaussian characteristics, with values of $\alpha$ varying between 0.9 and 1.6, away from the Gaussian point of $\alpha = 2$. Our modeling results clearly point to the need for the design of Bayesian processors that take into consideration the non-Gaussian heavy-tailed character of the data to achieve close to optimal speckle mitigation performance.

### 4.3 A Bayesian Processor for Ultrasound Speckle Removal

Current state-of-the-art wavelet-based denoising and image enhancement techniques employ a combination of wavelet shrinkage by soft and hard thresholding together with a generalized adaptive gain (GAG) for feature emphasis (see [106] and references therein). In particular, Zong et al. apply soft thresholding at fine scales (levels 1 and/or 2) and hard thresholding within middle levels 3 and/or 4 to eliminate noise, followed by nonlinear processing of feature energy to enhance contrast. The regularized threshold parameter used in [106] is related to the noise level, orientation, and scale through a judiciously chosen but at the same time ad hoc linearly decreasing function. Moreover, the five parameters which determine the empirical GAG function in [106] are tuned experimentally to achieve the appropriate nonlinear stretching of wavelet coefficients that accomplishes the desired contrast enhancement.
In this section, our goal is the design of a formal Bayesian estimator that recovers the signal component of the wavelet coefficients in ultrasound images by using an alpha-stable signal prior distribution. The proposed processor is motivated by the modeling studies in the previous section, it is based on solid statistical theory, and it does not depend on the use of ad hoc thresholding and stretching parameters.

In a Bayesian framework, referring to (4.1.5), \( d_{j,k}, s_{j,k} \), and \( \xi_{j,k} \) are considered as samples of the random variables \( d \), \( s \), and \( \xi \), respectively. The signal component \( s \) is modeled according to a \( S\alpha S \) distribution with zero location parameter, while the noise component \( \xi \) is modeled as a zero-mean Gaussian random variable. Our goal is to find the Bayes risk estimator \( \hat{s} \) that minimizes the conditional risk, which is the loss averaged over the conditional distribution of \( s \), given the set of wavelet coefficients, \( d \):

\[
\hat{s}(d) = \arg \min_{\hat{s}} \int L[s, \hat{s}(d)] P_{s|d}(s |d) \, ds
\]

(4.3.1)

The Bayes risk estimator under a quadratic cost function minimizes the mean-square error (MSE) and is given by the conditional mean of \( s \), given \( d \):

\[
\hat{s}(d) = \int s P_{s|d}(s |d) \cdot ds
\]

(4.3.2)

Of course, the MSE metric is defined for random variables that possess finite second-order moments. In this work, the signal component of the wavelet coefficients is modeled as an alpha-stable random variable that does not have finite second-order statistics. Hence, we use the absolute error \(|s - \hat{s}(d)|\) as the loss function in expression (4.3.1). Under this loss function, expression (4.3.1) is well defined for all \( S\alpha S \) random variables with characteristic exponent \( \alpha \) greater than one. The Bayesian estimator that we consider minimizes the mean absolute error (MAE) and can be shown to be the conditional median of \( s \), given \( d \) [85]. But, since the conditional density...
$P_{s|d}(s|d)$ is symmetric around zero, the conditional median coincides with the conditional mean. Hence, the Bayesian MAE estimator for the absolute error cost function is again given by equation (4.3.2).

Bayes’ theorem gives the a posteriori probability density function of $s$ based on the measured set of wavelet coefficients:

$$P_{s|d}(s|d) = \frac{P_{d|s}(d|s) \cdot P_{s}(s)}{\int P_{d|s}(d|s) \cdot P_{s}(s) \cdot ds},$$

(4.3.3)

where $P_{s}(s)$ is the prior PDF of the signal component of the wavelet coefficients of the ultrasound image and $P_{d|s}(d|s)$ is the likelihood function. Substituting (4.3.3) into (4.3.2), we get:

$$\hat{s}(d) = \frac{\int P_{d|s}(d|s) \cdot P_{s}(s) \cdot ds}{\int P_{d|s}(d|s) \cdot P_{s}(s) \cdot ds} = \frac{\int P_{\xi}(d-s) \cdot P_{s}(s) \cdot ds}{\int P_{\xi}(d-s) \cdot P_{s}(s) \cdot ds} = \frac{\int P_{\xi}(\xi) \cdot P_{s}(s) \cdot ds}{\int P_{\xi}(\xi) \cdot P_{s}(s) \cdot ds},$$

(4.3.4)

where $P_{\xi}(\xi)$ is the PDF of the wavelet coefficients corresponding to the noise.

In order to be able to construct the Bayesian processor in (4.3.4), first we estimate the parameters of the prior distributions of the signal ($s$) and noise ($\xi$) components of the wavelet coefficients ($d$). Then, we use the parameters to “build” the two prior PDFs $P_{\xi}(\xi)$ and $P_{s}(s)$ and the nonlinear (in general) input-output relationship $\hat{s}(d)$. Observing (4.3.4), we note that the denominator, referred to as the “evidence” is the PDF of the noisy observation, $d$, computed as the convolution between the noise and signal PDFs. Hence, we need a parameterized model for the two PDFs that provides a good fit to the statistics of the ultrasound images. Furthermore, the distribution parameters should be estimated from the noisy observations in an efficient manner.

Motivated by our modeling results in Section 4.2, we use a two-parameter $SaS$ model for the signal component while we use a zero-mean Gaussian model for the noise component. In other words, the observed signal is a mixture of $SaS$ signal
and Gaussian noise. Moreover, we consider the signal and noise components to be independent. Because of the lack of closed-form expressions for the general $S\alpha S$ PDF, we propose a method that is based on characteristic functions. In particular, since the PDF of the measured coefficients is the convolution between the PDFs of the signal and noise components, the associated characteristic function of the measurements is given by the product of the characteristic functions of the signal and noise:

$$\Phi_d(\omega) = \Phi_s(\omega) \cdot \Phi_\xi(\omega), \quad (4.3.5)$$

where

$$\Phi_s(\omega) = \exp \left(-\gamma_s |\omega|^{\alpha_s}\right), \quad 1 < \alpha \leq 2$$

and

$$\Phi_\xi(\omega) = \exp \left(-\frac{\sigma^2}{2}|\omega|^2\right).$$

Using expression (4.3.5), we estimate the parameters $\alpha_s$, $\gamma_s$, and $\sigma$ by fitting the Fourier transform of the empirical PDF of the measured coefficients with function $\Phi_d(\omega)$ in the least-squares (LS) sense:

$$\{\hat{\alpha}_s, \hat{\gamma}_s, \hat{\sigma}\} = \arg \min_{\alpha_s, \gamma_s, \sigma} \sum_i^n [\Phi_d(\omega_i) - \Phi_{d_e}(\omega_i)]^2 \quad (4.3.6)$$

where $\Phi_{d_e}(\omega)$ denotes the empirical characteristic function. In practice, we first estimate the level of noise $\sigma$, and we optimize (4.3.6) only with respect to the $S\alpha S$ parameters $\alpha_s$ and $\gamma_s$. As proposed in [28], a robust estimate of the noise standard deviation, $\sigma$, is obtained in the finest decomposition scale by the measured wavelet coefficients as

$$\hat{\sigma} = \frac{1}{0.6745} \text{MAD}((d_{J,k}, 0 \leq k < 2^J)), \quad (4.3.7)$$

where the operator $\text{MAD}$ signifies the median absolute deviation and $J$ denotes the highest level of wavelet decomposition. We found that this method for estimating
the $S\alpha S$ parameters gives reliable estimates, it is not computationally expensive and, more importantly, it allows us to estimate the parameters from the noisy measurements. We should note here that Paulson et al. used a similar approach to estimate the parameters of alpha-stable densities and observed that LS fitting in the characteristic function domain produces estimates within two standard errors of the actual values and with a bias that is inversely proportional to the sample size [74].

As expected, in the general case the Bayesian processor described in (4.3.4) does not have a closed-form expression. Only for the case of Gaussian signal in Gaussian noise a well-known closed-form solution exists:

$$\hat{s}(d) = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_d^2}d,$$

where $\sigma_s^2$ is the Gaussian signal variance. In other words, the processing is a simple linear rescaling of the measurement. For the general non-Gaussian $S\alpha S$ signal case, we computed numerically the Bayesian processor function in (4.3.4). Figure 4.6(a) depicts the Bayesian input-output curves for four different values of the signal characteristic exponent, $\alpha$, namely, $\alpha = 2$ (Gaussian data), $\alpha = 1.95$ (slightly non-Gaussian data), $\alpha = 1.5$, and $\alpha = 1.05$ (considerably heavy-tailed data). All curves except the case $\alpha = 2$, correspond to a nonlinear “coring” operation, i.e., large-amplitude observations are essentially preserved while small-amplitude values are suppressed. This is expected since small measurement values are assumed to come from signal values close to zero.

On the other hand, Figure 4.6(b) shows how the processor nonlinearity is varied for certain signal statistics ($\alpha = 1.5$) and various noise levels. It is evident from the curves that as the noise level increases, the amount of shrinkage also increases. We should note at this point that curves similar to the ones in Figure 4.6 are chosen
Figure 4.6: Bayesian processor input - output curves for alpha-stable signal ($1 < \alpha \leq 2$) and Gaussian noise prior distributions. The straight line with ×’s indicates the identity function. (a) Bayesian curves for constant $\gamma_s/\sigma$ ratio and four different signal statistics corresponding to $\alpha = 2$ (Gaussian signal, solid line), $\alpha = 1.95$ (slightly non-Gaussian signal, dash-dotted line), $\alpha = 1.5$, and $\alpha = 1.05$ (considerably heavy-tailed signal, dashed and dotted lines, respectively). (b) Bayesian curves for $S\alpha S$ signal $\alpha = 1.5$ and four different $\gamma_s/\sigma$ ratios: 0.1 (solid), 0.12 (dash-dotted), 0.14 (dashed), and 0.2 (dotted).
adaptively by our processor since the PDF parameters are estimated by means of equations (4.3.6) and (4.3.7) at each level of wavelet decomposition and for each orientation. In actual ultrasound images, at the first decomposition levels where the wavelet coefficients arising from noise are predominant, the Bayesian shrinkage function would resemble to that corresponding to low signal-to-noise ratio (SNR). As the resolution decreases, in general the noise level decreases and the nonlinearity applied to the wavelet coefficients corresponds to the high SNR curves in Figure 4.6(b) gradually approaching the identity function as the SNR becomes very high.

4.4 Experimental Results

We tested our proposed multiscale Bayesian speckle suppressing algorithm on the ultrasound images modeled in Section 4.2. In order to obtain speckle images, we degraded the original test images by multiplying them with unit-mean random fields, as shown in expression (4.1.2). We generated spatially correlated speckle noise \( \eta_m(x, y) \) by lowpass filtering a complex Gaussian random field and taking the magnitude of the filtered output. We controlled the correlation length of the speckle by appropriately setting the size of the kernel used to introduce correlation to the underlying Gaussian noise. A kernel size of one corresponds to white noise. On the other hand, in order to allow the noise correlation to taper gradually to zero we could not set the kernel size arbitrarily large. Thus, a short-term correlation obtained with a kernel of size three was sufficient to model reality. In our experiments, we considered three different levels of simulated speckle noise.

We compared the results of our approach with other speckle reduction techniques including median filtering, homomorphic Wiener filtering, and wavelet shrinkage
4.4 Experimental Results

denoising using soft and hard thresholding. All the parameters involved in these methods were selected by trial-and-error in order to get an optimal result from each method. Specifically, for the median filter we used a $3 \times 3$ mask for the lowest level of noise and $5 \times 5$ masks for the other two levels. The homomorphic Wiener filter was implemented using a window of size $5 \times 5$ pixels for the highest level of noise and $3 \times 3$ pixels in all the other cases. For soft thresholding we used a threshold $t = 1.5 \sigma_d$, while for hard thresholding we have chosen $t = 3\sigma_d$, $\sigma_d$ being the standard deviation of the wavelet coefficients. Both wavelet shrinkage soft and hard thresholding schemes were developed using Daubechies’ Symmlet 8 mother wavelet. Denoising results using this basis wavelet have been found to be less affected by pseudo-Gibbs phenomena [106]. Moreover, in order to minimize such side effects, we have embedded all wavelet-based methods (including the Bayesian approach) into the cycle spinning algorithm [23]. This consists in averaging the result of the wavelet shrinkage method over all circulant shifts of the input image. In practice we found that a number of 8 translations is sufficient. The maximum number of wavelet decompositions we used was 5.

In order to quantify the achieved performance improvement, three different measures were computed based on the original and the denoised data. For quantitative evaluation, an extensively used measure is the MSE defined as:

$$MSE = \frac{1}{K} \sum_{i=1}^{K} (\hat{S}_i - S_i)^2$$  \hspace{1cm} (4.4.1)

where $S$ is the original image, $\hat{S}$ is the denoised image, and $K$ is the image size. The standard signal to noise ratio (SNR) is not adequate to evaluate the noise suppression in case of multiplicative noise. Instead, a common way to achieve this in coherent
imaging is to calculate the signal-to-MSE (S/MSE) ratio, defined as [37]:

\[
S/MSE = 10 \log_{10} \left( \sum_{i=1}^{K} S_i^2 / \sum_{i=1}^{K} (\hat{S}_i - S_i)^2 \right)
\] (4.4.2)

This measure corresponds to the classical SNR in the case of additive noise.

Remember that in ultrasound imaging, we are interested in suppressing speckle noise while at the same time preserving the edges of the original image that often constitute features of interest for diagnosis. Thus, in addition to the above quantitative performance measures, we also consider a qualitative measure for edge preservation. More specifically, we used a parameter \( \beta \) originally defined in [41, 84]:

\[
\beta = \frac{\Gamma(\Delta S - \overline{\Delta S}, \hat{\Delta S} - \overline{\Delta S})}{\sqrt{\Gamma(\Delta S - \overline{\Delta S}, \Delta S - \overline{\Delta S}) \cdot \Gamma(\hat{\Delta S} - \overline{\Delta S}, \hat{\Delta S} - \overline{\Delta S})}}
\] (4.4.3)

where \( \Delta S \) and \( \hat{\Delta S} \) are the high-pass filtered versions of \( S \) and \( \hat{S} \) respectively, obtained with a \( 3 \times 3 \)-pixel standard approximation of the Laplacian operator, the overline operator represents the mean value, and

\[
\Gamma(S_1, S_2) = \sum_{i=1}^{K} S_{1_i} \cdot S_{2_i}.
\] (4.4.4)

The correlation measure, \( \beta \) should be close to unity for an optimal effect of edge preservation.

The obtained values of MSE, S/MSE, and \( \beta \) for all methods applied to the kidney image are given in Table 4.2. It is evident from the table that the three wavelet-based methods (i.e., soft and hard thresholding and our proposed Bayesian denoising technique) are more successful in speckle noise suppression than median and homomorphic Wiener filtering. In terms of MSE and S/MSE, the soft thresholding scheme achieves comparable performance with the homomorphic Wiener filter, but the visual quality
Table 4.2: Image enhancement measures obtained by the 5 denoising methods tested on the kidney-ultrasound image. The S/MSE is given in dB. Values of the correlation measure, $\beta$, close to unity denote optimal edge preservation performance.

<table>
<thead>
<tr>
<th>Method</th>
<th>MSE</th>
<th>S/MSE</th>
<th>$\beta$</th>
<th>MSE</th>
<th>S/MSE</th>
<th>$\beta$</th>
<th>MSE</th>
<th>S/MSE</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noisy</td>
<td>26.052</td>
<td>5.61</td>
<td>0.287</td>
<td>16.294</td>
<td>9.69</td>
<td>0.433</td>
<td>7.211</td>
<td>16.77</td>
<td>0.735</td>
</tr>
<tr>
<td>Median</td>
<td>13.700</td>
<td>11.19</td>
<td>0.213</td>
<td>9.763</td>
<td>14.13</td>
<td>0.344</td>
<td>6.677</td>
<td>17.43</td>
<td>0.570</td>
</tr>
<tr>
<td>Wiener</td>
<td>13.838</td>
<td>11.10</td>
<td>0.177</td>
<td>8.810</td>
<td>15.03</td>
<td>0.498</td>
<td>6.394</td>
<td>17.81</td>
<td>0.628</td>
</tr>
<tr>
<td>Soft</td>
<td>13.637</td>
<td>11.23</td>
<td>0.336</td>
<td>8.924</td>
<td>14.91</td>
<td>0.588</td>
<td>6.113</td>
<td>18.20</td>
<td>0.806</td>
</tr>
<tr>
<td>Hard</td>
<td>13.500</td>
<td>11.32</td>
<td>0.316</td>
<td>8.640</td>
<td>15.20</td>
<td>0.557</td>
<td>5.560</td>
<td>19.02</td>
<td>0.756</td>
</tr>
<tr>
<td>Bayesian</td>
<td>12.739</td>
<td>11.82</td>
<td>0.455</td>
<td>8.203</td>
<td>15.65</td>
<td>0.625</td>
<td>4.886</td>
<td>20.15</td>
<td>0.824</td>
</tr>
</tbody>
</table>

of the soft threshold processed image seems to be better (cf. Figure 4.7). This is due to the fact that the soft thresholding approach is not intended to minimize the MSE, the result being an estimator which achieves a low variance at the expense of bias [28]. Observing the $\beta$ metric values, we see that the multiresolution techniques exhibit a clearly better performance in terms of edge preservation, as expected. Among them, our proposed Bayesian approach exhibits the best performance according to all three metrics.

Figure 4.7 shows a representative result from the processing of the noisy kidney image. The simulated speckle image shown in Figure 4.7(b) corresponds to a S/MSE value of 9.69dB. For this noise level, all the methods that we tested achieved a good speckle suppression performance. However, the median and homomorphic Wiener filters loose many of the signal details and the resulting images are blurred (cf. Figure 4.7(c,d)). On the other hand, the images processed by soft and hard thresholding are oversmoothed (cf. Figure 4.7(e,f)). Clearly, as it can be seen in Figure 4.7(g), our proposed Bayesian processor effectively reduces speckle, it preserves step edges, and it enhances fine signal details, better than the other methods.
Figure 4.7: Results of various speckle suppressing methods. (a) Original kidney ultrasound image. (b) Image degraded with simulated speckle noise $S/MSE = 9.69$ dB. (c) Median filtering. (d) Homomorphic Wiener filtering. (e) Soft thresholding. (f) Hard thresholding. (g) Proposed Bayesian denoising.
Indeed, the problem with the MSE, S/MSE, and $\beta$ measures, or with any other metric, is their connection to the visual interpretation of an human observer. A radiologist, in analyzing ultrasound images, does not compute any of the above measures. Hence, in order to visually study the merit of the proposed $S\alpha S$ subband coefficient modeling and the Bayesian processor, we chose a noisy ultrasound image, applied the algorithm without artificially adding noise, and visually evaluated the denoised image. The results of this experiment are shown in Figure 4.8. The figure only shows results obtained using the wavelet based schemes, which were proved to give better results for simulated speckle noise. Although qualitative evaluation in this case is highly subjective, the results of this experiment seem to be consistent with the simulation results. It appears that the Bayesian processor performs like a feature detector, retaining the features that are clearly distinguishable in the speckled data.

4.5 Summary

In this chapter, we introduced a novel multiscale nonlinear homomorphic method for speckle suppression in ultrasound images. Three are the main processing stages of our approach. First, similarly to existing multiresolution techniques [41, 106], the logarithm of the image is decomposed into several scales through a multiresolution analysis employing the 2-D wavelet transform. This step guarantees that the speckle is transformed from multiplicative into additive and its characteristics are differentiated from the signal characteristics in each decomposition level. The second and third steps differentiate our technique from existing ones. After decomposing the original image, the signal and noise components in various scales are modeled as $S\alpha S$ and Gaussian processes, respectively. The parameters of the distributions are estimated from the
Figure 4.8: (a) Noisy ultrasound image of the bladder. (b) Image denoised using translation-invariant soft thresholding. (c) Image denoised using translation-invariant hard thresholding. (d) Image enhanced using our Bayesian algorithm.
measurements by means of a LS fitting in the characteristic function domain. We showed that the class of $S\alpha S$ distributions is more effective in modeling detail image histograms than other exponentially tailed densities.

In the third step, a Bayesian processor based on a $S\alpha S$ signal prior is built at each scale for statistically optimal signal feature extraction and speckle suppression. The main advantage of our method is that the obtained input-output shrinkage functions are optimal in the Bayesian sense. As a result, a more accurate signal reconstruction is achieved in each scale.
Chapter 5

Ultrasound Image Denoising via Maximum a Posteriori Estimation of Wavelet Coefficients

In this chapter we present another technique for speckle noise removal from ultrasound images using nonlinear processing of wavelet coefficients. Compared with the algorithm in the previous chapter, the innovative aspects introduced here consist of the following: (i) In the data modeling component of our processor, we propose a new method for estimating the parameters of the alpha-stable distribution from noisy observations, which is based on Koutrouvelis’ [53] regression method; (ii) in the data filtering component of our processor, we select a uniform loss cost function for the design of the Bayes risk estimator, which results to a MAP filter based on alpha-stable statistics. Our design gives rise to a set of optimal nonlinear input-output processor curves parameterized by the degree of non-Gaussianity of the data.
5.1 $S\alpha S$ Parameters Estimation from Noisy Measurements

We refer again to the block diagram in Figure 4.1 from Chapter 4. The complete algorithm with six major steps is similar to that proposed in the previous chapter. The major changes concern the “$\alpha, \gamma, \sigma$ estimation” and “Bayesian Processor” blocks. In the following we will describe the changes leading to the modified algorithm and we will show simulation results obtained after processing an ultrasound image of a fetal chest.

Recall that at the output of the DWT block we get, at each resolution level and for all orientations, sets of noisy wavelet coefficients written as the sum of the transformations of the signal and the noise:

$$d_{j,k}^n = s_{j,k}^i + \xi_{j,k}^i,$$

where $k = 0, ..., 2^{j+1} - 1$ and $-1 < j < -J$ refer to the decomposition level or scale and $i = 1, 2, 3$ refers to the three spatial orientations.

In a Bayesian framework $d_{j,k}$, $s_{j,k}$, and $\xi_{j,k}$ are considered as samples of the random variables $d$, $s$, and $\xi$, respectively. The distribution parameters corresponding to the signal ($s$) and noise wavelet coefficients ($\xi$) should be estimated from the noisy observations ($d$) in an efficient manner. To achieve this, we start again from the characteristic function of the measurements, which is given by the product of the
characteristic functions of the signal and noise:

$$\Phi_d(\omega) = \Phi_s(\omega) \cdot \Phi_\xi(\omega). \quad (5.1.2)$$

We model the signal component of the wavelet coefficients using a two-parameter S\(\alpha\)S distribution:

$$\Phi_s(\omega) = \exp (-\gamma_s |\omega|^{\alpha_s}), \quad 0 < \alpha \leq 2$$

while a Gaussian distribution characterizes the noise component:

$$\Phi_\xi(\omega) = \exp (-\frac{\sigma^2}{2} |\omega|^2).$$

At this point, instead of directly fitting the Fourier transform of the empirical PDF of the measured coefficients with the function \(\Phi_d(\omega)\) as we did in Chapter 4, we observe that (5.1.2) implies:

$$\log[-(\log |\Phi_d(\omega)|^2 + \sigma^2 \omega^2)] = \log(2\gamma_s) + \alpha_s \log |\omega| \quad (5.1.3)$$

First, we estimate the level of noise using equation 4.3.7. Then, we find the parameters \(\alpha_s\) and \(\gamma_s\) by regressing \(y = \log[-(\log |\Phi_d(\omega)|^2 + \sigma^2 \omega^2)]\) on \(w = \log |\omega|\) in the model

$$y_k = \mu + \alpha w_k + \epsilon_k \quad (5.1.4)$$

where \(\mu = \log(2\gamma)\), \(\epsilon_k\) denotes an error term, and \((\omega_k, k = 1, \ldots, K)\) is an appropriate set of real numbers. The optimum number \(K\) of points depends on the characteristic exponent \(\alpha\) and on the sample size. Specifically, \(K\) decreases as \(\alpha\) increases and as the number of samples increases. For a more detailed discussion on choosing the optimal \(K\), we refer the reader to [53].

We found that this method for estimating the S\(\alpha\)S parameters gives reliable estimates, it is computationally efficient and it allows us to estimate the parameters from
the noisy measurements. Koutrouvelis in [53] used a similar approach to estimate the parameters of alpha-stable distributions and he showed that his regression method gives very good results in terms of consistency, bias, and efficiency.

5.2 Design of a MAP Processor for Speckle Mitigation

Having estimated the necessary signal and noise distribution parameters from the data, our goal is to design and implement a Bayes risk processor. The Bayes estimator $\hat{s}$ minimizes the conditional risk, which is the loss averaged over the conditional distribution of $s$, given the noisy observation, $d$:

$$\hat{s}(d) = \arg\min_{\hat{s}} \int L[s, \hat{s}(d)] P_{s|d}(s | d) \, ds$$

(5.2.1)

Selecting the uniform cost function:

$$L[s, \hat{s}(d)] = \begin{cases} 
0, & \text{for } |s - \hat{s}| < \epsilon \\
1, & \text{otherwise}
\end{cases}$$

(5.2.2)

the optimal estimator can be derived as follows:

$$\hat{s}(d) = \arg\min_{\hat{s}} \int_{|s - \hat{s}| \geq \epsilon} P_{s|d}(s | d) \, ds = \arg\min_{\hat{s}} \left[ 1 - \int_{|s - \hat{s}| < \epsilon} P_{s|d}(s | d) \, ds \right]$$

(5.2.3)

Thus, in order to minimize the expected cost, when $\epsilon \to 0$ one should select:

$$\hat{s}(d) = \arg\max_{\hat{s}} P_{s|d}(s | d)$$

(5.2.4)

It is important to underline at this point that under the loss function in (5.2.2), the estimator given by expression (5.2.1) is well defined for all SαS random variables (with characteristic exponent $\alpha$ taking values in the whole range $0 < \alpha \leq 2$). This
estimator is called the *maximum a posteriori* (MAP) estimator. Bayes’ theorem gives the *a posteriori* PDF of $s$ based on the measured data:

$$P_{s|d}(s \mid d) = \frac{P_{d|s}(d \mid s) P_s(s)}{P_d(d)}, \quad (5.2.5)$$

where $P_s(s)$ is the *prior* PDF of the alpha-stable modeled signal component of the measurements and $P_{d|s}(d \mid s)$ is the *likelihood* function. Substituting (5.2.5) in (5.2.4), we get:

$$\hat{s}(d) = \arg \max_{\hat{s}} P_{d|s}(d \mid s) P_s(s) = \arg \max_{\hat{s}} P_\xi(d - s) P_s(s) = \arg \max_{\hat{s}} P_\xi(\xi) P_s(s). \quad (5.2.6)$$

Only for the case of Gaussian signal and Gaussian noise does a closed-form solution exist for the processor described above:

$$\hat{s}(d) = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_d^2} d, \quad (5.2.7)$$

where $\sigma_s^2$ is the Gaussian signal variance. In other words, the processing is a simple linear rescaling of the measurement. For the general alpha-stable signal case, the Bayesian processor does not have a closed-form expression and one has to numerically compute the MAP input-output curves.

Figure 5.1 depicts the numerically computed MAP input-output curves for five different values of the signal characteristic exponent, $\alpha$, namely, $\alpha = 2$ (Gaussian data), $\alpha = 1.95$ (slightly non-Gaussian data), $\alpha = 1.5$, $\alpha = 1$, and $\alpha = 0.5$ (considerably heavy-tailed data). The figure illustrates the processor dependency on the parameter $\alpha$ of the signal prior PDF. Specifically, for a given ratio $\gamma_s/\sigma$, the amount of shrinkage decreases as $\alpha$ decreases. The intuitive explanation for this behavior is that the smaller the value of $\alpha$, the heavier the tails of the signal PDF and the greater the probability that the measured value is due to the signal.
Figure 5.1: MAP processor input-output curves for alpha-stable signal and Gaussian noise prior distributions. The straight line with x's indicates the identity function. The five different signal statistics correspond to \( \alpha = 2 \) (Gaussian signal, solid line), \( \alpha = 1.95 \) (slightly non-Gaussian signal, dashed line), \( \alpha = 1.5 \), \( \alpha = 1 \), and \( \alpha = 0.5 \) (considerably heavy-tailed signal, dotted, dash-dotted and solid with \( \circ \) lines, respectively). All the curves correspond to a same ratio \( \sigma/\gamma = 2 \).

5.3 Simulation Results

In this section, we show simulation results obtained by processing one ultrasound image, randomly chosen from our database. The original image is shown in Fig. 5.2(a) and it represents an ultrasound scan of a fetal chest.

In order to obtain speckle images, we degraded the original test image by multiplying it with unit-mean random fields, as explained in Section 4.1. We controlled the correlation length of the speckle by appropriately setting the size of the kernel used to
introduce correlation to the underlying Gaussian noise. In practice uncorrelatedness of the noise could be achieved by decimating the image to the theoretical resolution limit of the imaging device [24]. Thus, a short-term correlation obtained with a kernel of size three was sufficient to model reality. We considered three different levels of simulated speckle noise.

We compared the results of our approach with the classical median filter, and wavelet shrinkage denoising using soft thresholding. The soft thresholding scheme was developed using Daubechies’ Symmlet 8 mother wavelet as suggested in [106], while for our algorithm we used the Symmlet 4 wavelet. The maximum number of wavelet decompositions we used was five. In order to minimize the effect of pseudo-Gibbs phenomena, we have embedded both wavelet-based methods into the cycle spinning algorithm [23].

In order to quantify the achieved performance improvement of our method over the existing ones, we computed the two measures defined in Chapter 4. Specifically, for quantitative evaluation, we used the (S/MSE) ratio (Equation 4.4.2), while as measure for edge preservation we considered again the correlation measure $\beta$ (cf. Equation 4.4.3). The results are summarized in Tables 5.1 and 5.2 respectively. It can be seen that our proposed Bayesian approach exhibits the best performance according to both metrics.

Table 5.1: Quantitative image enhancement measures obtained using three denoising methods. The tabulated S/MSE metric is given in $dB$.

<table>
<thead>
<tr>
<th></th>
<th>5.63</th>
<th>9.67</th>
<th>16.68</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noisy image</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median Filtering</td>
<td>11.07</td>
<td>14.24</td>
<td>17.97</td>
</tr>
<tr>
<td>Soft Thresholding</td>
<td>10.95</td>
<td>14.70</td>
<td>18.38</td>
</tr>
<tr>
<td>Bayesian Denoising</td>
<td>11.66</td>
<td>15.48</td>
<td>19.93</td>
</tr>
</tbody>
</table>
Table 5.2: Qualitative image enhancement measures obtained using three denoising methods. Values of $\beta$ close to unity denote optimal edge preservation performance.

<table>
<thead>
<tr>
<th>Method</th>
<th>Noisy image</th>
<th>Median Filtering</th>
<th>Soft Thresholding</th>
<th>Bayesian Denoising</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.2577</td>
<td>0.1989</td>
<td>0.2806</td>
<td>0.3814</td>
</tr>
<tr>
<td></td>
<td>0.3930</td>
<td>0.4230</td>
<td>0.4985</td>
<td>0.5203</td>
</tr>
<tr>
<td></td>
<td>0.6933</td>
<td>0.5310</td>
<td>0.7391</td>
<td>0.7957</td>
</tr>
</tbody>
</table>

In Fig. 5.2 we show for visual comparison a representative result from the processing of our test image. On inspecting the images, it is obvious that the result of median filtering the image is inferior to the two wavelet-based approaches. Specifically, the result is a blurred image in which many edges have been lost (Fig. 5.2(c)). Also, although it achieves a good speckle suppression performance, the soft thresholding technique oversmooths the result (Fig. 5.2(d)). Clearly, the Bayesian processor performs like a feature detector, which retains the features that are clearly distinguishable in the speckled data but cuts out anything which is assumed to be constituted by noise (Fig. 5.2(e)).

5.4 Discussions

We have shown that a successful ultrasound imaging algorithm can achieve both noise reduction and feature preservation if it takes into consideration the true statistics of the signal and noise components. The proposed algorithm is based on solid statistical theory, and it does not depend on the use of any *ad hoc* parameter.

Our processor was tested and found to be more effective than thresholding methods, which are *ad hoc* in the sense that they do not allow for an exact matching of the signal and noise distributions at different scales and orientations. The method
Figure 5.2: Results of various speckle suppressing methods. (a) Original image. (b) Image degraded with simulated speckle noise ($S/MSE = 9.67dB$). (c) Median filtering. (d) Translation-invariant soft thresholding. (e) Bayesian denoising.
proposed in Section 5.2 for choosing the “coring” nonlinearity could be thus considered as a principled way of shrinking noisy data, relying on the true statistics of the signal and noise wavelet coefficient. For example, the curve corresponding to $\alpha = 0.5$ in Fig. 5.1 mimics quite accurately the flavor of a hard thresholding operator.

Finally, we should note that the methods proposed in Chapters 4 and 5 could be easily adapted for the purpose of denoising other types of biomedical images where the noise can be (eventually after an appropriate transformation) modeled as additive Gaussian and signal-independent.
Chapter 6

Application to SAR Image Despeckling

As explained in the Discussions of the previous chapter, the alpha-stable based Bayesian processors introduced in this thesis could be easily adapted for the purpose of denoising other types of medical images. Moreover, it could find application in other areas of interest as well, a straightforward example being the case of SAR images, which are known to be affected by the same type of noise. The purpose of this chapter is to show that the radar reflectivity wavelet coefficients exhibit also significantly non-Gaussian statistics that are accurately described by the alpha-stable family of distributions and consequently, the Bayesian estimators previously introduced can be successfully applied in this case.


6.1 Introduction

After more than a half century since its inception as an imaging system in the 1950s and 1960s, there is still a growing interest in SAR imaging on account of its importance in a variety of applications such as high-resolution remote sensing for mapping, surface surveillance, search-and-rescue, mine detection, and automatic target recognition (ATR). SAR systems are currently employed in many airborne and satellite-borne platforms, such as the E-3 AWACS (Airborne Warning and Control System) airplane devoted to target tracking, the E-8C Joint STARS (Surveillance Target Attack Radar System) airplane performing target detection and localization, and the NASA space shuttle [90]. The one most important SAR attribute leading to its gain in popularity is the ability to image large areas of terrain at very fine resolutions and in all-weather conditions.

A major issue in SAR imagery is that basic textures are generally affected by multiplicative speckle noise [38]. Speckle noise is a consequence of image formation under coherent radiation. It is not truly a noise in the typical engineering sense, since its texture often carries useful information about the scene being imaged. However, the presence of speckle is generally considered undesirable since it damages radiometric resolution and it affects the tasks of human interpretation and scene analysis. Thus, it appears sensible to reduce speckle in SAR images, provided that the structural features and textural information are not lost.

Many adaptive filters for speckle reduction have been proposed in the past. The Frost filter was designed as an adaptive Wiener filter that assumed an autoregressive (AR) exponential model for the scene reflectivity [33]. Kuan considered a multiplicative speckle model and designed a linear filter based on the minimum mean-square
error (MMSE) criterion, optimal when both the scene and the detected intensities are Gaussian distributed [54]. The Lee MMSE filter was a particular case of the Kuan filter due to a linear approximation made for the multiplicative noise model [58]. A two-dimensional Kalman filter was developed by Sadjadi and Bannour under the modeling of the image as a Markov field satisfying a causal AR model [10]. The Gamma MAP filter was based on a Bayesian analysis of the image statistics where both signal and speckle noise follow a Gamma distribution [11]. Finally, a family of six robust filters for speckle reduction was proposed by Frery et al. employing trimmed maximum likelihood, best linear unbiased, and moment-based estimation, as well as median, interquartile range, and median absolute deviation [32].

Recently, there has been considerable interest in using the wavelet transform as a powerful tool for recovering SAR images from noisy data [40, 37, 34, 35]. As in the case of ultrasound images, when multiplicative contamination is concerned, multiscale methods involve a pre-processing step consisting of a logarithmic transform to separate the noise from the original image. Then, different wavelet shrinkage approaches are employed, which are based on Donoho’s pioneering work [27]. In [37], Gagnon and Jouan perform a comparative study between a complex wavelet coefficient shrinkage filter and several standard speckle filters that are largely used by SAR imaging scientists, and show that the wavelet-based approach is among the best for speckle removal. To address the disadvantages involved by thresholding methods, Pizurica et al. [76] proposed an efficient technique for despeckling SAR images by using analytic model distributions for the noise and signal wavelet coefficients. They developed a simple local model for spatial context and they obtained a family of
adaptive shrinkage functions. Finally, Xie et al. developed a similar method by fusing the wavelet Bayesian denoising technique with Markov-random-field-based SAR image regularization [102].

As previously noted in Chapter 4 parametric Bayesian processing presupposes proper modeling for the prior probability density function of the signal. Consequently, we demonstrate through extensive modeling of real data that the subband decompositions of SAR images have also significantly non-Gaussian statistics that are best described by the alpha-stable family of distributions. Finally, we apply the Bayesian estimator described in the previous chapter to the case of SAR images and we evaluate and compare the performance of our proposed algorithm with the performance of existing SAR despeckling methods.

6.2 Modeling SAR Wavelet Coefficients with Alpha-Stable Distributions

In this section, we show results on modeling data obtained by applying the 2-D wavelet transform to a set of real SAR images. Specifically, we study whether the stable family provides a flexible and appropriate tool for modeling the coefficients within the framework of multiscale wavelet analysis of logarithmically transformed SAR images.

To achieve this goal, we modeled a series of SAR images from the MSTAR Public Clutter dataset. The data set contains X-band images with 1784 × 1476 pixels and 1 ft × 1 ft resolution at 15° depression angles. Since speckle appears inherently

\[1\] The dataset can be obtained through the Sensor Data Management System (SDMS) of Wright Laboratory at the URL http://www.mbvlab.wpafb.af.mil/public/sdms/
6.2 Modeling SAR Wavelet Coefficients

in any SAR image, we have first processed the actual images using the Gamma-
MAP filter [11] and considered the resulting images as reasonable approximations
of the speckle free radar reflectivity. Because of limited space, we only describe the
modeling of ten representative images in intensity format. All of them have a 256
gray-level resolution and constitute cropped versions (512x512 pixels) of the original
images.

We proceed by using the same methodology as in Section 4.2: First, we assess
whether the data deviate from the normal distribution and if they have heavy tails.
To determine that, we make use of normal probability plots. Then, we check if the
data is in the stable domain of attraction by estimating the characteristic exponent,
α, directly from the data and by providing the related confidence intervals. Several
methods have been proposed for estimating stable parameters. Here, we use the
maximum likelihood method described by Nolan in [71], which gives reliable estimates
and provides the most tight confidence intervals. As further stability diagnostics, we
employ probability density plots that give a good indication of whether the $S_{αS}$ fit
matches the data near the mode and at the tails of the distribution.

In Figure 6.1, we show the filtered image HB06158 from the MSTAR collection,
its log-transformed version and the corresponding three-scale decomposition. The
normal probability plot corresponding to the vertical subband at the first level of
decomposition of this image is shown in Figure 6.2. The plot provides strong evidence
that the underlying distribution is not normal. The ”+” marks in the plot show the
empirical probability versus the data value for each point in the sample. The marks
are in a curve that does not follow the straight Gaussian line and thus, the normality
assumption is violated for this data. While non-Gaussian stable densities are heavy-
tailed, not all heavy-tailed distributions are stable. Hence, in Figure 6.3 we assess
the stability of the data.

First, the characteristic exponent is estimated and the data sample is fitted with
the corresponding stable distribution. For the particular case shown here, the char-
acteristic exponent of the $S_{\alpha}S$ distribution which best fits the data was estimated
to be $\hat{\alpha} = 1.253$. The stabilized p-p $S_{\alpha}S$ plot in Figure 6.3 shows a highly accurate
Figure 6.2: Normal probability plot of the vertical subband at the first level of decomposition of filtered image HB06158 from MSTAR dataset. Characterization of data non-Gaussianity. The “+” marks correspond to the empirical probability density versus the data value for each point in the sample. Since the marks are in a curve that does not follow the straight Gaussian line, the normality assumption is violated for this data.

stable fit for this data set.

A question of interest is whether the stable model provides a better fit to the data than other heavy-tailed families of distributions. Thus, we compare the $S\alpha S$ fits with those provided by the generalized Laplacian density function (see 4). In order to model the wavelet subband coefficients of images, one can examine their histograms [89, 3, 63], which model their probability density functions or equivalently use amplitude probability density (APD) functions ($P|X| > x$). The APD can be evaluated empirically directly from the data, as well as theoretically from the density function considered. Figure 6.4 shows an example of modeling the vertical subband
Figure 6.3: Stabilized p-p plot for $S\alpha S$ fit of data set representing the vertical subband at the first level of decomposition of filtered image HB06158. The “◦” marks, denoting the empirical probability density, are in a curve that very accurately follows the straight $S\alpha S$ line corresponding to $\alpha = 1.253$.

at the first level of decomposition of the SAR image under study. A highly accurate stable fit can be observed. In particular, the figure shows that the $S\alpha S$ distribution is superior to the generalized Laplacian distribution because it provides a better fit to both the mode and the tails of the empirical density of the actual data.

For every image we iterated three times the separable wavelet decomposition and we modeled the coefficients of each subband by using the $S\alpha S$ family. The wavelet decomposition was accomplished using Daubechies’ Symmlet 8 basis wavelet. The results are summarized in Table 6.1, which shows the ML estimates of the characteristic exponent $\alpha$ together with the corresponding 95% confidence intervals. It can be
observed that the confidence interval depends on the particular level of decomposition. The confidence interval becomes wider as the level increases since the number of samples used for estimating $\alpha$ gets smaller. The table demonstrates that the coefficients of different subbands and decomposition levels exhibit various degrees of non-Gaussianity. The important observation is that all subbands exhibit distinctly non-Gaussian characteristics, with values of $\alpha$ varying between 0.7 and 1.9, away from the Gaussian point of $\alpha = 2$. Our modeling results clearly prove that the Bayesian processors designed in the previous two chapters can be successfully applied for the case of radar reflectivity to achieve close to optimal speckle mitigation performance.
Table 6.1: Alpha-stable modeling of wavelet subband coefficients of actual SAR images from the MSTAR Public Clutter dataset. Maximum likelihood parameter estimates and 95% confidence intervals for the $S_{\alpha}S$ characteristic exponent, $\alpha$. The tabulated key parameter $\alpha$ defines the degree of non-Gaussianity as deviations from the value $\alpha = 2$, which corresponds to the Gaussian condition.

<table>
<thead>
<tr>
<th>IMAGE</th>
<th>Level</th>
<th>Image Subbands</th>
<th>Horizontal</th>
<th>Vertical</th>
<th>Diagonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>HB06158</td>
<td></td>
<td>1.232 ± 0.010</td>
<td>1.253 ± 0.010</td>
<td>0.812 ± 0.009</td>
<td></td>
</tr>
<tr>
<td>HB06159</td>
<td></td>
<td>1.336 ± 0.011</td>
<td>1.362 ± 0.010</td>
<td>0.867 ± 0.008</td>
<td></td>
</tr>
<tr>
<td>HB06160</td>
<td></td>
<td>1.014 ± 0.009</td>
<td>1.042 ± 0.010</td>
<td>0.686 ± 0.007</td>
<td></td>
</tr>
<tr>
<td>HB06161</td>
<td>I</td>
<td>1.064 ± 0.009</td>
<td>1.094 ± 0.009</td>
<td>0.703 ± 0.007</td>
<td></td>
</tr>
<tr>
<td>HB06162</td>
<td></td>
<td>1.048 ± 0.009</td>
<td>1.058 ± 0.009</td>
<td>0.709 ± 0.007</td>
<td></td>
</tr>
<tr>
<td>HB06163</td>
<td></td>
<td>1.084 ± 0.010</td>
<td>1.084 ± 0.010</td>
<td>0.718 ± 0.007</td>
<td></td>
</tr>
<tr>
<td>HB06215</td>
<td></td>
<td>1.371 ± 0.011</td>
<td>1.367 ± 0.011</td>
<td>0.885 ± 0.007</td>
<td></td>
</tr>
<tr>
<td>HB06245</td>
<td></td>
<td>1.286 ± 0.010</td>
<td>1.300 ± 0.010</td>
<td>0.850 ± 0.008</td>
<td></td>
</tr>
<tr>
<td>HB06246</td>
<td></td>
<td>1.308 ± 0.011</td>
<td>1.315 ± 0.010</td>
<td>0.844 ± 0.007</td>
<td></td>
</tr>
<tr>
<td>HB06278</td>
<td></td>
<td>1.231 ± 0.011</td>
<td>1.242 ± 0.011</td>
<td>0.807 ± 0.007</td>
<td></td>
</tr>
<tr>
<td>HB06158</td>
<td>II</td>
<td>1.395 ± 0.022</td>
<td>1.442 ± 0.022</td>
<td>1.032 ± 0.019</td>
<td></td>
</tr>
<tr>
<td>HB06159</td>
<td></td>
<td>1.492 ± 0.022</td>
<td>1.527 ± 0.023</td>
<td>1.029 ± 0.018</td>
<td></td>
</tr>
<tr>
<td>HB06160</td>
<td></td>
<td>1.215 ± 0.021</td>
<td>1.253 ± 0.022</td>
<td>0.963 ± 0.020</td>
<td></td>
</tr>
<tr>
<td>HB06161</td>
<td></td>
<td>1.269 ± 0.021</td>
<td>1.293 ± 0.021</td>
<td>0.944 ± 0.019</td>
<td></td>
</tr>
<tr>
<td>HB06162</td>
<td></td>
<td>1.248 ± 0.021</td>
<td>1.277 ± 0.022</td>
<td>0.961 ± 0.019</td>
<td></td>
</tr>
<tr>
<td>HB06163</td>
<td></td>
<td>1.256 ± 0.022</td>
<td>1.295 ± 0.022</td>
<td>0.968 ± 0.020</td>
<td></td>
</tr>
<tr>
<td>HB06215</td>
<td></td>
<td>1.516 ± 0.022</td>
<td>1.525 ± 0.021</td>
<td>1.073 ± 0.019</td>
<td></td>
</tr>
<tr>
<td>HB06245</td>
<td></td>
<td>1.422 ± 0.022</td>
<td>1.437 ± 0.021</td>
<td>1.053 ± 0.019</td>
<td></td>
</tr>
<tr>
<td>HB06246</td>
<td></td>
<td>1.447 ± 0.022</td>
<td>1.468 ± 0.021</td>
<td>0.993 ± 0.017</td>
<td></td>
</tr>
<tr>
<td>HB06278</td>
<td></td>
<td>1.365 ± 0.021</td>
<td>1.394 ± 0.021</td>
<td>0.993 ± 0.018</td>
<td></td>
</tr>
<tr>
<td>HB06158</td>
<td>III</td>
<td>1.684 ± 0.045</td>
<td>1.688 ± 0.043</td>
<td>1.578 ± 0.043</td>
<td></td>
</tr>
<tr>
<td>HB06159</td>
<td></td>
<td>1.667 ± 0.045</td>
<td>1.817 ± 0.040</td>
<td>1.513 ± 0.044</td>
<td></td>
</tr>
<tr>
<td>HB06160</td>
<td></td>
<td>1.456 ± 0.045</td>
<td>1.660 ± 0.046</td>
<td>1.543 ± 0.047</td>
<td></td>
</tr>
<tr>
<td>HB06161</td>
<td></td>
<td>1.567 ± 0.047</td>
<td>1.685 ± 0.043</td>
<td>1.542 ± 0.046</td>
<td></td>
</tr>
<tr>
<td>HB06162</td>
<td></td>
<td>1.614 ± 0.044</td>
<td>1.627 ± 0.045</td>
<td>1.556 ± 0.047</td>
<td></td>
</tr>
<tr>
<td>HB06163</td>
<td></td>
<td>1.705 ± 0.047</td>
<td>1.626 ± 0.048</td>
<td>1.486 ± 0.047</td>
<td></td>
</tr>
<tr>
<td>HB06215</td>
<td></td>
<td>1.883 ± 0.034</td>
<td>1.776 ± 0.042</td>
<td>1.576 ± 0.044</td>
<td></td>
</tr>
<tr>
<td>HB06245</td>
<td></td>
<td>1.659 ± 0.046</td>
<td>1.863 ± 0.041</td>
<td>1.594 ± 0.044</td>
<td></td>
</tr>
<tr>
<td>HB06246</td>
<td></td>
<td>1.704 ± 0.043</td>
<td>1.683 ± 0.042</td>
<td>1.515 ± 0.042</td>
<td></td>
</tr>
<tr>
<td>HB06278</td>
<td></td>
<td>1.568 ± 0.045</td>
<td>1.620 ± 0.044</td>
<td>1.568 ± 0.045</td>
<td></td>
</tr>
</tbody>
</table>
The statistical properties of speckle noise $\eta_m(x, y)$ were studied by Goodman [38]. He has shown that, if the number of scatterers per resolution cell is large, a fully developed speckle pattern can be modeled as the magnitude of a complex Gaussian field with i.i.d. real and imaginary components. Arsenault and April [9] have shown that when the image intensity is logarithmically transformed, the speckle noise is approximately Gaussian additive noise, and it tends to a normal probability much faster than the intensity distribution. Xie et al. employ a distance between cumulative distributions to measure the deviation of the log-transformed speckle from Gaussianity [103]. They confirm the result in [9] and show that even for the amplitude image, although the log-transformed speckle tends to a Gaussian PDF slightly slower than the original speckle noise, the former is still statistically very close to the Gaussian PDF.

Other realistic speckle noise models include the K-distribution [72], $\mathcal{G}$-distribution [31], log-normal distribution [37], and correlated speckle pattern [3, 72]. However, since our processor employs the wavelet transform which, through the central limit theorem, drives the noise wavelet coefficients to approximate a Gaussian distribution, we use the log-normal distribution as the speckle noise model: If $X$ follows the log-normal distribution with parameters $\mu$ and $\sigma^2$, then $\ln X$ follows the normal distribution with mean $\mu$ and variance $\sigma^2$. For the log-normal distribution, the mean and variance are given respectively by:

\begin{align*}
M &= \exp (\mu + \frac{\sigma^2}{2}) \quad (6.3.1) \\
\sigma_{\ln}^2 &= \exp (2\mu + 2\sigma^2) - \exp (2\mu + \sigma^2) \quad (6.3.2)
\end{align*}

The appropriateness of the use of the log-normal model for speckle noise has also
Figure 6.5: Modified block diagram of the multiscale homomorphic Bayesian-based algorithm for speckle suppression. The additional block “adjust mean” is necessary to correct the biased mean resulting from the logarithmic transformation.

been assessed by Kaplan [51]. A log-normal random variable can be generated using:

\[ X_{\text{log-normal}} = \exp \left( X_{\text{normal}} \sqrt{2 \log \frac{M}{m} + \ln m} \right) \]  

(6.3.3)

where \( M \) and \( m \) are the mean and the median values of the distribution, respectively, and \( X_{\text{normal}} \) is a standard zero-mean, unit-variance Gaussian random variable. There is a straightforward equivalence between the Equivalent Number of Looks (ENL) in a speckle image and the parameter \( m \) in the above expression [37].

We should note here that due to the use of the logarithmic transformation, the mean of the log-transformed speckle field is biased [103]. For unit-mean log-normal distributed speckle noise, the mean of the corresponding Gaussian distribution is
6.4 Experimental Results

In this section, we present simulation results obtained by processing several test SAR images using our proposed WIN-SAR speckle suppression processor and we compare the results of our approach with other current state-of-the-art speckle filtering methods. In order to be able to quantify the improvement achieved by our method, we have first degraded three original “noiseless” images with synthetic speckle in a controlled manner. Finally, for qualitative visual evaluation, we processed various unaltered SAR images with WIN-SAR.

6.4.1 Synthetic Data Examples

We were interested in performing experiments on images of different types and with various content in order to be able to obtain results, which we could claim to be general enough (Figure 6.6). Thus, an aerial image was used for its identical content with real SAR images. This image was obtained by cropping “westaerialconcorde” found in Matlab’s Image Processing Toolbox. For testing the smoothing performance of the algorithm as well as its edge preservation potential, we also chose to apply it on the classical “boat” image. Finally, as a test for texture preservation, we generated an image containing four different textures and applied the algorithms to it. In order
to obtain speckle images, we degraded the original test images by multiplying them with unit-mean random fields, defined in expression (6.3.3). In our experiments, we considered three different levels of simulated speckle noise, with ENL = 1, 3, and 8 respectively.

We compared the results of our approach with other speckle reduction techniques including the Lee filter [58], the GMAP filter [11], and wavelet shrinkage denoising using soft thresholding [27]. We selected the parameters associated with each method by trial-and-error in order to achieve optimal results. Specifically, for the Lee filter we used a $5 \times 5$ mask, while the GMAP filter was implemented using a window of size $7 \times 7$ pixels. For soft thresholding we used a threshold $t = 1.5 \sigma_d$, $\sigma_d$ being the standard deviation of the wavelet coefficients. The wavelet shrinkage soft thresholding scheme was implemented using Daubechies' Symmlet 8 mother wavelet. In order to minimize side effects like pseudo-Gibbs phenomena, we embedded both wavelet-based methods (including our Bayesian approach) into the cycle spinning algorithm [23]. This algorithm was implemented using 8 circulant shifts of the input image. The parameters $\alpha_s$ and $\gamma_s$ in the WIN-SAR processor are estimated for each shift. The maximum number of wavelet decompositions we used was 5.

In order to quantify the achieved performance improvement, three different measures were computed based on the original and the denoised data. For quantitative evaluation, we used the MSE defined in Section 4.4. Also, in order to quantify the speckle reduction performance we computed the standard-deviation-to-mean ratio ($S/M$). This quantity is a measure of image speckle in homogeneous regions. Recalling that in SAR imaging, we are interested in suppressing speckle noise while at the same time preserving the edges of the original image, we also considered again the
6.4 Experimental Results

Table 6.2: Image enhancement measures obtained by four denoising methods applied on the “aerial,” “boat,” and “texture” test images. Three levels of noise are considered corresponding to ENL=1, 3, and 8. The measures are calculated on an average of ten noise realizations.

<table>
<thead>
<tr>
<th>Method</th>
<th>ENL = 1</th>
<th>ENL = 3</th>
<th>ENL = 8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE</td>
<td>S/M</td>
<td>β</td>
</tr>
<tr>
<td>Noisy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aerial</td>
<td>133.008</td>
<td>1.116</td>
<td>0.009</td>
</tr>
<tr>
<td>Lee</td>
<td>43.016</td>
<td>0.469</td>
<td>0.064</td>
</tr>
<tr>
<td>GMAP</td>
<td>49.744</td>
<td>0.507</td>
<td>0.046</td>
</tr>
<tr>
<td>Soft Thresh</td>
<td>26.032</td>
<td>0.363</td>
<td>0.173</td>
</tr>
<tr>
<td>WIN-SAR</td>
<td>22.691</td>
<td>0.354</td>
<td>0.297</td>
</tr>
<tr>
<td>Noisy</td>
<td>146.66</td>
<td>1.123</td>
<td>0.007</td>
</tr>
<tr>
<td>Boat</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lee</td>
<td>46.028</td>
<td>0.490</td>
<td>0.069</td>
</tr>
<tr>
<td>GMAP</td>
<td>54.293</td>
<td>0.531</td>
<td>0.058</td>
</tr>
<tr>
<td>Soft Thresh</td>
<td>27.072</td>
<td>0.399</td>
<td>0.199</td>
</tr>
<tr>
<td>WIN-SAR</td>
<td>21.903</td>
<td>0.379</td>
<td>0.393</td>
</tr>
<tr>
<td>Noisy</td>
<td>137.94</td>
<td>1.322</td>
<td>0.019</td>
</tr>
<tr>
<td>Texture</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lee</td>
<td>54.854</td>
<td>0.638</td>
<td>0.126</td>
</tr>
<tr>
<td>GMAP</td>
<td>56.732</td>
<td>0.702</td>
<td>0.178</td>
</tr>
<tr>
<td>Soft Thresh</td>
<td>40.957</td>
<td>0.543</td>
<td>0.328</td>
</tr>
<tr>
<td>WIN-SAR</td>
<td>38.593</td>
<td>0.618</td>
<td>0.395</td>
</tr>
</tbody>
</table>

The obtained values of MSE, S/M, and β for all methods applied to the three test images are given in Table 6.2. The numbers in the table represent average values obtained after repeating each experiment ten times, using the same settings but for different noise realizations. It is evident from the table that the two wavelet-based methods are more successful in speckle noise suppression than the Lee and GMAP filters in most situations. It can be seen that in general our proposed processor exhibits the best performance according to all three metrics. The soft thresholding
Figure 6.6: Results of various speckle suppressing methods. From top to bottom: original, noisy \((ENL = 8)\), GMAP filtered, soft thresholding, and Bayesian denoised images respectively. From left to right: Aerial image, Boat image, Brodatz textures.
method occasionally gives better results in terms of the $S/M$ measure but at the expense of over-smoothed images as it can be seen by comparing the $\beta$ index metric as well as by visual inspection of Figure 6.6. In terms of MSE, the soft thresholding scheme achieves comparable performance with the GMAP filter, but the visual quality of the soft threshold processed images seems to be better. This is due to the fact that the soft thresholding approach is not intended to minimize the MSE, the result being an estimator which achieves a low variance at the expense of bias [27]. Observing the $\beta$ metric values, we see that our Bayesian multiresolution technique exhibits a clearly better performance in terms of edge preservation, as expected.

6.4.2 Real SAR Imagery Examples

The problem with the MSE, $S/M$, and $\beta$ measures, or with any other metric, is associating them directly to the visual interpretation of a human observer. Hence, in order to study the merit of the proposed $S\alpha S$ subband coefficient modeling and the resulting Bayesian processor, we also chose noisy SAR images, we applied the algorithm without artificially adding noise, and we visually evaluated the denoised images. The first test image (single-look, amplitude format), shown in Figure 6.7(a), depicts an urban scene having a dense set of large cross-section targets with intermingled tree shadows. This image was provided by Dr. D. E. Wahl of Sandia National Laboratories in New Mexico, USA and it was also used in [93, 99].

We should note at this point that in situations where the image is affected by speckle with a high correlation length, algorithmic design should account for noise correlation and a whitening filter should be used. Alternatively, the data could be downsampled at the cost of reducing the spatial resolution. The use of an orthonormal
Figure 6.7: Processing of SAR image of urban scene. (a) Original SAR image. (b) Image denoised using soft thresholding. (c) Image enhanced using our algorithm.
wavelet basis guarantees that the noise component of the wavelet coefficients will be uncorrelated, provided that the noise was white in the image domain. The second test image shown in Figure 6.8(a) illustrates this idea. The image represents a rural scene from the MSTAR collection. The results shown in Figure 6.8 are obtained after downsampling the original image by a factor of 2.

For visual comparison, we show results obtained using the soft thresholding based scheme (Figures 6.7(b) and 6.8(b)) and the WIN-SAR processor (Figures 6.7(c) and 6.8(c)). Although qualitative evaluation in these cases is highly subjective, i.e., no universal quality measure for filtered SAR data exists, the results of the above two experiments seem to be consistent with the simulation results. The soft thresholding method achieves good speckle suppression performance but it over-smoothes images and thus many features are blurred. It appears that the proposed WIN-SAR processor performs like a feature detector, retaining the features that are clearly distinguishable in the speckled data but filtering out anything which is assumed to be constituted by noise.
Figure 6.8: Processing of a clutter-like scene. (a) Noisy SAR image. (b) Image denoised using soft thresholding. (c) Image enhanced using our algorithm.
6.5 Discussions

We introduced a new statistical representation for the wavelet decomposition coefficients of SAR images, based on heavy-tailed alpha-stable models. Consequently, we tested a MAP processor which relies on this representation and we found it to be more effective than traditional wavelet shrinkage methods both in terms of speckle reduction and signal detail preservation. We evaluated the results on both synthetic data and real SAR images, all coded in 8-bit. Naturally, our approach is more computationally expensive due to the fact that the prior distribution parameters need to be estimated at each decomposition scale of interest. However, this is not a serious problem for off-line processing.

It should also be noted that in this work, the parameters of the $S\alpha S$ model are estimated globally within each decomposition scale. For this reason, the shrinking functions shown in Figure 5.1 act the same for strong point target and for extended homogenous regions. According to the results, the proposed filter achieves a global compromise between smoothing and edge preservation.
Chapter 7

Future Work Directions

Currently, we are addressing several issues related to the work we presented in this thesis. One major issue is the choice of a statistical model for the speckle noise component of the wavelet coefficients that is more appropriate than the currently used Gaussian model. It is to be tested whether the $S\alpha S$ family is a good model also for the noise component. In this case, our problem will be formulated as Bayesian signal detection from measurements that are mixtures of $S\alpha S$ signal in $S\alpha S$ noise with different characteristic exponents, in general.

Statistical correlation between adjacent pixels is a result of diffraction effects in the transverse direction and intersymbol interference effects in the range direction [24]. Speckle correlation was not considered in this work. As we mentioned, this problem can be addressed by image subsampling at the expense of reduced spatial resolution. A more sophisticated approach is to consider the speckle correlation structure into the MAP function.

A fully global Bayesian estimator based on alpha-stable statistics that takes into consideration both the inter- and intra-scale dependencies of the wavelet coefficient
should be also developed. This issue could eventually be addressed by first developing the theory of \textit{alpha-stable Markov random fields}.

Following denoising, subsequent image analysis tasks should become easier to accomplish. The alpha-stable model could be further used for developing image segmentation or texture classification/synthesis algorithms.

Finally, one issue that could be addressed and subsequently applied in medical imagery concerns the optimal quantization of the general alpha-stable distribution. Up to now, only a particular member of this family of distributions, namely the Cauchy distribution, has been successfully applied to natural image coding [94]. Moreover, we propose to extend this approach within the framework of independent component analysis (ICA) bases. It should be noted that the solution of this problem would constitute an important result in the image compression literature: there has not been reported up to now any successful implementation of an image compression algorithm using overcomplete ICA bases for image windows.
Bibliography


