April, 1991

LIDS- TH 2033

Research Supported By:

AFOSR grant 88-0032C
NSF grant MIP-9015281
ARO grant DAAL03-86-K-0171(CICS)

Identifying Multiscale Statistical Models Using the Wavelet Transform*

Golden, S.A.

April 1991 LIDS-TH-2033

IDENTIFYING MULTISCALE STATISTICAL MODELS USING THE WAVELET TRANSFORM

by

S.A. Golden

This report is based on the unaltered thesis of Stuart A. Golden submitted to the Department of Electrical Engineering and Computer Science in partial fulfillment of the requirements for the degrees of Electrical Engineer and Master of Science at the Massachusetts Institute of Technology in June 1991. This research was conducted at the M.I.T. Laboratory for Information and Decision Systems with partial support provided by the Air Force Office of Scientific Research under grant AFOSR-88-0032C, the National Science Foundation under grant NSF 9015281-MIP and the Army Research Office under grant ARO DAAL03-86-K-0171 (Center for Intelligent Control Systems).

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA 02139

IDENTIFYING MULTISCALE STATISTICAL MODELS USING THE WAVELET TRANSFORM

by

Stuart A. Golden

B.S. Electrical Engineering and Management Science,

University California San Diego

(1988)

Submitted to the Department of
Electrical Engineering and Computer Science
in Partial Fulfillment of the Requirements for the Degrees of

ELECTRICAL ENGINEER

and

MASTER OF SCIENCE

at the

Massachusetts Institute of Technology

June 1991

Signature of Author Stuart Golden
Department of Electrical Engineering and Computer Science
Tuesday, April 23, 1991

Certified by Alan S. Willsky
Alan S. Willsky
Professor of Electrical Engineering
Thesis Supervisor

Accepted by _____
Arthur C. Smith
Department Committee on Graduate Studies

IDENTIFYING MULTISCALE STATISTICAL MODELS USING THE WAVELET TRANSFORM

by

Stuart A. Golden

Submitted to the Department of Electrical Engineering
on Tuesday, April 23, 1991 in partial fulfillment of the
requirements for the Degrees of Electrical Engineer
and Master of Science in Electrical Engineering

ABSTRACT

Recently much attention has been focused on methods for performing multiple resolution decompositions of signals based on wavelet transforms. In this thesis we develop an algorithm to determine optimal wavelet transforms based upon the statistical characterization of the signal being analyzed. Our criterion is based upon the desire to find the optimal wavelet transform approximation of a Karhunen-Loeve expansion, i.e. we would like the transformed coefficients to be as close to white as possible. We determine the optimal wavelet transform in a level-by-level procedure. Using Vaidyanathan and Hoang's paramaterization [1] of quadrature mirror filters (QMFs), we chose the QMF pair such that the 1-level coarser approximation of the signal and the wavelet coefficients at that level are as close to being statistically uncorrelated with each other as possible. This procedure is then repeated for an arbitrary number of levels. We examine the ability of the transform to achieve an approximate Karhunen-Loeve expansion by considering several examples. The examples that we consider are a first-order Gauss-Markov process, a second-order under-damped process, and fractional Brownian motions.

Thesis Supervisor: Dr. Alan S. Willsky

Title: Professor of Electrical Engineering

Acknowledgements

I would like to thank my advisor Professor Alan Willsky for his continual guidance and support throughout this endeavor. I feel that a large portion of my learning at MIT is a direct result of working with Prof. Willsky. I am also very grateful for the support and assistance of my fellow graduate students. In particular, I greatly appreciate the times that Ken Chou and Darrin Taylor have spent in helping me answer some of my many questions.

1 **Motivation** 8

2 **Multiscale Signal Processing** 14

The Wavelet Transform 14

Perfect Reconstruction of QMF Filters 19

Parametrization of QMF Filters 24

The Discrete-Time Wavelet Transform for Finite Length Sequences 27

3 **Multiscale Statistical Modeling** 35

Multiscale Stochastic Processes 35

Solving the Minimization Problem 40

4 **Analyzing Processes Using the Wavelet Transform** 44

Stationary Processes 44

The First-Order Gauss-Markov Process 48

Under-Damped Second-Order Processes 70

Fractional Brownian Motions 86

5 **Modeling Processes Using the Wavelet Transform** 106

Modeling Processes 106

The First-Order Gauss-Markov Process 114

Fractional Brownian Motions 126

6 **Conclusion** 132

A **Periodicity Properties of the Objective Function** 140

B **Optimal two-tap QMF Filters for Stationary Processes** 144

References 148

List of Figures

2.1 Block-diagram of low-pass filtering followed by 2:1 decimation. 19

2.2 Block-diagram of sub-band coder and decoder. 20

4. 1 Normalized autocovariance of a Gauss-Markov process. 57

4. 2 Gray-scale mapping. 58

4. 3 The covariances of the autocovariance of the one-level Haar transform of a Gauss-Markov process viewed separately 59

4. 4 The autocovariance of the one-level Haar transform of a Gauss-Markov process. 60

4. 5 The normalized autocovariance of one to four-level Haar transform of a Gauss-Markov process. 61

4. 6 Cross-section of normalized autocovariance of one to four-level Haar transform of a Gauss-Markov process. 62

4. 7 The variances of the transformed coefficients of one to four-level Haar transform of a Gauss-Markov process. 63

4. 8 The power spectrum of a Gauss-Markov process. 64

4. 9 The normalized autocovariance of the three-level wavelet transform of a Gauss-Markov process using optimal QMFs. 65

4. 10 Cross-section of normalized autocovariance of the three-level wavelet transform of a Gauss-Markov process using optimal QMFs. 66

4. 11 The variances of the transformed coefficients of the three-level wavelet transform of a Gauss-Markov process using optimal QMFs. 67

4. 12 The normalized autocovariance and variances of the three-level wavelet transform of a Gauss-Markov process using QMFs designed by Daubechies. 68

4. 13 Cross-section of the normalized autocovariance of a Gauss-Markov process using optimal QMFs and QMFs derived by Daubechies. 69

4. 14 The power spectrum of a second-order under-damped process. 75

4. 15 Normalized autocovariance of a second-order under-damped process. 76

4. 16 The magnitude of the normalized autocovariance of one to four-level Haar transform of a second-order under-damped process. 77

4. 17 The positive values of the normalized autocovariance of the one to four-level Haar transform of a second-order under-damped process. 78

4. 18 Correlation structure of the 100th transformed coefficient of a second-order under-damped process when using a one to four-level Haar transform. 79

4. 19 The variances of the transformed coefficients of a one to four-level Haar transform of a second-order under-damped process. 80

4. 20 Block diagram of two-level wavelet and complete wave packet transforms. 81

4. 21 The magnitude of the normalized autocovariance of the one to four-level complete wave packet transform of a second-order under-damped process. 82

4. 22 The positive values of the normalized autocovariance of the one to four-level complete wave packet transform of a second-order under-damped process. 83

4. 23 Correlation structure of the 100th transformed coefficient of a second-order under-damped process when using a one to four-level complete wave packet transform. 84

4. 24 The variances of the transformed coefficients of one to four-level complete wave packet transform of a second-order under-damped process. 85

4. 25 Autocovariance of Brownian motion. 93

4. 26 The magnitude of the normalized autocovariance of the one to four and seven-level Haar transform of Brownian motion. 94

4. 27 Correlation structure of the 100th transformed coefficient of Brownian motion when using a one to four and seven-level Haar transform. . . . 95

4. 28 The variances of the transformed coefficients of one to four and seven-level Haar transform of Brownian motion. 96

4. 29 The normalized autocovariance of the three-level wavelet transform of Brownian motion using optimal QMFs. 97

4. 30 Correlation structure of the 100th transformed coefficient of the three-level wavelet transform of Brownian motion using optimal QMFs . . . 98

4. 31 The variances of the transformed coefficients of the three-level wavelet transform of Brownian motion using optimal QMFs. 99

4. 32 The normalized autocovariance and the variances of the three-level wavelet transform of Brownian motion using QMFs derived by Daubechies. 100

4. 33 Cross-section of the autocovariance of Brownian motion using optimal QMFs and QMFs derived by Daubechies. 101

4. 34 Autocovariance of an fBm process with parameter H=0.25. 102

4. 35 The normalized autocovariance, the variances, and a cross-section of the full Haar transform of an fBm with H=0.25. 103

4. 36 Autocovariance of an fBm process with parameter H=0.75. 104

4. 37 The normalized autocovariance, the variances, and a cross-section of the full Haar transform of an fBm with H=0.75. 105

5. 1 Approximations of the autocovariance of the first-order Gauss-Markov process. 118

5. 2 The errors in the approximations of the autocovariance of the first-order Gauss-Markov process. 119

5. 3 The variances of the approximated coefficients of the first-order Gauss-Markov process. 120

5. 4 Bhattacharyya distance and probability of error between the first-order Gauss-Markov process and its approximations. 121

5. 5 The mapping used to map the Bhattacharyya distance to the equivalent Gauss-Markov parameter. 122

5. 6 Bhattacharyya distance between the first-order Gauss-Markov process and its approximation in terms of the equivalent Gauss-Markov parameter. 123

5. 7 Sample path of the first-order Gauss-Markov process. 124

5. 8 Sample paths of the approximations of the first-order Gauss-Markov process. 125

5. 9 Approximations of the autocovariance of Brownian motion. 129

5. 10 The errors in the approximations of the autocovariance of Brownian motion. 130

5. 11 Bhattacharyya distance and probability of error between Brownian motion and its approximation. 131

6. 1 Normalized autocovariance and the variances of the three-level wavelet transform coefficients using QMFs that were obtained from a modified optimization algorithm. 137

6. 2 Cross-section of the normalized autocovariance of a Gauss-Markov process using QMFs obtained from a modified optimization algorithm and QMFs derived by Daubechies. 138

6. 3 Impulse responses of the QMFs obtained from a modified optimization algorithm and QMFs derived by Daubechies. 139

CHAPTER 1

MOTIVATION

In recent years there has been considerable interest and activity in the signal and image processing community in developing multi-resolution processing algorithms. Among the reasons for this are the apparent or claimed computational advantages of such methods and the fact that representing signals or images at multiple scales is an evocative notion-- it seems like a "natural" thing to do. One of the more recent areas of investigation in multiscale analysis has been the emerging theory of multiscale representations of signals and wavelet transforms [2-12]. This theory has sparkled an impressive flurry of activity in a wide variety of technical areas, at least in part because it offers a common, unifying language and perspective and perhaps the promise of a framework in which a rational methodology can be developed for multiscale signal processing, complete with a theoretical structure that pinpoints when multiresolution methods might be useful and why.

It is important to realize, however, that the wavelet transform by itself is not the only element needed to develop a methodology for signal analysis. To understand this one need only look to another orthonormal transform, namely the Fourier transform which decomposes signals into its frequency components rather than its components at different resolutions. The reason that such a transform is useful is that its use simplifies the description of physically meaningful classes of signals and important classes of transformations of those signals. In particular stationary stochastic processes are whitened by the Fourier transform so that individual frequency components of such a process are statistically uncorrelated.

Not only does this greatly simplify their analysis, but, it also allows us to deduce that frequency-domain operations such as Wiener or matched filtering (or their time domain realizations as linear shift-invariant systems) are not just convenient things to do. They are in fact the right, i.e. the statistically optimal, things to do. In analogy, what is needed to complement wavelet transforms for the construction of a rational framework for multi-resolution signal analysis is the identification of a rich class of signals and phenomena whose description is simplified by wavelet transforms. Having this, we then have the basis for developing a methodology for scale domain filtering and signal processing, for deducing that such operations are indeed the right ones to use, and for developing a new and potentially powerful set of insights and perspectives on signal and image analysis that are complementary to those that are the heritage of Fourier.

When approaching a signal or image analysis problem, there are three distinct ways that a theory of multiresolution statistical signal processing may be particularly beneficial. First, the phenomenon under investigation may possess features and physically significant effects at multiple scales. For example, fractal models have often been suggested for the description of natural scenes, topography, ocean wave height textures, etc [13-17]. Also, anomalous broadband transient events or spatially-localized features can naturally be thought of as the superposition of finer resolution features on a more coarsely varying background. As we will see, the modeling framework we describe is rich enough to capture such phenomena. For example, we will see that 1/f -like stochastic processes as in [18,19], fractional Brownian motions [20], and many other useful models are captured in our framework.

Secondly, whether the underlying phenomenon has multi-resolution features or not, it may be the case that the data that has been collected is at several different resolutions. For example the resolutions of remote sensing devices operating in

different bands -- such as IR, microwave, and radar -- may differ. Furthermore, even if only one sensor type is involved, measurement geometry may lead to resolution differences (for example, if zoomed and un-zoomed data are to be fused or if data is collected at different sensor-to-scene distances). In [21], we formulate an optimal estimation problem for the purpose of fusing these measurements at multiple resolutions by using the wavelet transform.

Finally, whether the phenomenon or data have multi-resolution features or not, the signal analysis algorithm may have such features in order to minimize the complexity of many signal processing problems. One such use of a multiresolution algorithm is to combat the computational demands of signal processing problems by solving coarse (and therefore computationally simpler) versions and using these to guide (and hopefully speed up) their higher resolution counterparts. Multigrid relaxation algorithms for solving partial differential equations [22-24] are of this type as are a variety of computer vision algorithms. Another such use of a multiresolution algorithm stems from the fact that a multiresolution formalism allows one to exercise very direct control over "greed" in signal and image reconstruction. In particular, many imaging problems are, in principle, ill-posed in that they require reconstructing more degrees of freedom then one has elements of data. In such cases one must "regularize" the problem in some manner, thereby guaranteeing accuracy of the reconstruction at the cost of some resolution. Since the usual intuition is precisely that one should have higher confidence in the reconstruction of lower resolution features, we are led directly to the idea of reconstruction at multiple scales, allowing the resolution-accuracy tradeoff to be confronted directly.

These issues of complexity are addressed in [25,26] where efficient multiresolution signal processing algorithms are obtained by modeling stochastic processes on dyadic trees. Specifically, a generalization of the Rauch-Tung-Striebel smoothing algorithm [27] is developed in [26] by performing a Kalman filtering

sweep from fine-to-coarse resolution signals followed by a smoothing sweep that proceeds from coarse-to-fine resolution signals. Algorithms such as these arise from using a multiscale framework that provide both the analytical tools for assessing resolution versus accuracy and for correctly accounting for fine scale fluctuations as a source of "noise" in coarser scale reconstructions.

In this thesis we concentrate on the question of when the signal itself possesses such multi-resolution features. If the process does possess these multiresolution features then various signal analysis and processing tasks -- such as filtering, smoothing, detection, classifications, segmentation, etc. -- can be performed quite quickly by using the wavelet transform. Specifically, if the transformed coefficients are uncorrelated, i.e. white, then smoothing algorithms, e.g. a Rauch-Tung-Striebel smoother, can be implemented efficiently in the wavelet transform domain. Even though this implementation would require transforming the data as well as inverse transforming the estimate, the overall procedure would be very efficient due to the fact that the wavelet transform is a very efficient operation (for a discussion on the efficiency of the wavelet transform see Section 2.4).

In order to obtain a greater insight as to what processes have these multiscale features, consider constructing a multiscale stochastic process. The construction of a multiscale stochastic process can most naturally be thought of in terms of the synthesis of a signal via the wavelet transform. That is, as one proceeds from coarse to fine scales, additional detail is added at each scale, yielding a coarse-to-fine dynamic evolution of finer and finer approximations to a signal. In particular the class of coarse-to-fine Markov processes, in which fine level detail is uncorrelated with coarser-level approximations, has substantial structure which can be exploited to construct optimal algorithms.

Specifically, if we perform a wavelet transform analysis of such a signal, we produce a fine-to-coarse sequence of signals, where at each stage, we extract the

signal components at that scale and produce a coarser level approximation to be used at the next stage. If the signal is in fact a multiscale Markov process and if the correct wavelet transform is used, the sequence of extracted signal components will be uncorrelated with each other. This suggests an approach to the identification of multiscale Markov models via the determination of an optimal wavelet transform, where optimality is specified in terms of the degree of decorrelation of the resulting sequence of multiscale signal components. In standard signal processing terminology, this problem corresponds to the specification of a pair of optimal Quadrature Mirror Filters (QMFs) which produce optimally decorrelated outputs. In more classical statistical terms, we are seeking the best approximations to a Karhunen-Loeve expansion in terms of a set of approximate eigenfunctions specified by these filters.

In Chapter 2, we discuss the existing theory of multiresolution signal processing. We show that the wavelet transform coefficients can be computed recursively without explicit computation of the continuous-time basis functions but with only a knowledge of the Quadrature Mirror Filters (QMFs). In order for the wavelet transform to be an orthogonal transformation, we show in Section 2.2 that the QMFs must satisfy the perfect reconstruction condition [9]. In Section 2.3 we show a recursion for parametrizing all QMFs that satisfy this condition [1].

In Chapter 3, we develop an algorithm that is based upon the desire that the optimal wavelet transform maximally decorrelates the transformed coefficients of a given stochastic process. In order to reduce the computational complexity of determining the optimal wavelet transform, we determine the optimal wavelet transform using a level-by-level procedure. Specifically, consider a one-level wavelet transform, in this case, our desire is to find the best QMFs that decorrelate the coarse approximation from the wavelet coefficients at that level when the autocovariance of the signal is known. We find this QMF pair by minimizing the

norm of the cross-covariance between the coarse approximation and the wavelet coefficients at that level. This minimization yields the optimal wavelet transform for the first-level which is used to determine the autocovariance of the coarse-approximation of the signal. The procedure is then continued for an arbitrary number of levels.

In Chapter 4, we investigate, in general, the ability of the wavelet transform to achieve a Karhunen-Loeve expansion of the transformed coefficients of particular classes of well-known processes. Specifically, we show images of the autocovariance matrices of the following processes transformed by their optimal wavelet transforms: a first-order Gauss-Markov process, a second-order under-damped process, and fractional Brownian motions. As we will see these transformed autocovariances have a very special structure. Specifically, there are bands of high correlation in the transformed autocovariance that we refer to as "fingers." The procedure we will describe attenuates these "fingers" in attempting to diagonalize the transformed autocovariance.

Chapter 5 deals with modeling processes by assuming that the transformed coefficients are indeed white. Specifically, we consider approximating the first-order Gauss-Markov process and fractional Brownian motions. We show images of the processes approximated in this manner and quantify how close the approximation is to the true process by using the Bhattacharyya distance [28].

The final chapter summarizes many of the key issues we have addressed in this thesis. We also discuss modifications that can be made to improve the performance of our optimization algorithm.

CHAPTER 2

MULTISCALE SIGNAL PROCESSING

2.1 THE WAVELET TRANSFORM

As developed in [5,6] the multi-scale representation of a continuous-time signal x(t) consists of a sequence of approximations of that signal on finer and finer subspaces of functions. The entire representation is completely specified by a single function $\phi(t)$, where the approximation at the mth scale is given by:

$$x_m(t)=\sum_{n=-\infty}^{+\infty} x[m,n]\phi_{m,n}(t) \qquad (2.1)$$

$$\phi_{m,n}(t)=2^{m/2}\phi(2^m t-n) \qquad (2.2)$$

Thus as $m\rightarrow\infty$ the approximation consists of a sum of many highly compressed, weighted, and shifted versions of $\phi(t)$. This function is far from arbitrary in its choice. First of all, one desirable property is that the approximations of Equation (2. 1) improve for larger values of m. For this to be true we want the space spanned by $\{\phi_{m,n}(t) \mid -\infty < n < \infty\}$ to be a subspace of the space spanned by $\{\phi_{m+1,n}(t) \mid -\infty < n < \infty\}$. Secondly the coefficients of the expansion, x[m,n], will be particularly simple to compute

if for each m the functions $\{\phi_{m,n}(t) \mid -\infty < n < \infty\}$ are orthonormal. It is easy to show that if the functions $\{\phi_{m,n}(t) \mid -\infty < n < \infty\}$ are orthonormal for any particular m, such as m=0, then the functions $\{\phi_{m,n}(t) \mid -\infty < n < \infty\}$ are orthonormal for any choice of m.

In the case that $\phi_{m,n}(t)$ forms an orthonormal basis for fixed m, the coefficents, x[m,n], can be written in terms of an inner product:

$$x[m,n] = \langle \phi_{m,n}(t),\ x_m(t) \rangle = \int_{t=-\infty}^{+\infty} \phi_{m,n}(t) x_m(t) dt \qquad (2.3)$$

A property of this choice of orthogonal functions is that the inner product of the ϕ's differing only by one level of m are not a function of m and can be expressed as:

$$h[2k-n] = \langle \phi_{m-1,k}(t),\ \phi_{m,n}(t) \rangle \qquad (2.4)$$

The representation just described is closely related to the wavelet transform. The wavelet transform is based on a single function $\psi(t)$. In a similar fashion to $\phi(t)$, the function $\psi(t)$ has the property that for fixed m, the set of functions $\psi_{m,n}(t)=2^{m/2}\psi(2^m t-n)$ are orthonormal for all n. Furthermore, $\psi(t)$ is also chosen such that the space spanned by $\{\phi_{m+1,n}(t) \mid -\infty < n < \infty\}$ is exactly equal to the subspace spanned by

$\{\phi_{m,n}(t) \mid -\infty < n < \infty\}$ plus an orthogonal subspace spanned by $\{\psi_{m,n}(t) \mid -\infty < n < \infty\}$. In order for these two subspaces to be orthogonal we require that $\phi_{m,n}(t)$ be orthogonal to $\psi_{m,k}(t)$ for all n and k. If $\psi(t)$ is chosen in this fashion, then [29,30,2] refer to $\{\psi_{m,n}\}$ as forming an orthonormal wavelet basis since in fact $\psi_{m_1, n_1}(t)$ and $\psi_{m_2, n_2}(t)$ are orthogonal whenever $(m_1, n_1) \neq (m_2, n_2)$.

We relate the approximation of x(t) at the mth scale to the approximation at the m-1st scale:

$$x_m(t)=x_{m-1}(t)+u_{m-1}(t) \tag{2. 5}$$

$$u_m(t)=\sum_{n=-\infty}^{+\infty} u[m,n]\psi_{m,n}(t) \tag{2. 6}$$

Taking the inner product of (2. 5) with $\phi_{m-1,k}(t)$ and using the orthogonality of $\phi_{m-1,k}(t)$ and $u_{m-1}(t)$ we obtain a dynamical relationship between the coefficents x[m,n] at one scale to those at its coarser approximation x[m-1,k] in the following manner:

$$x[m-1,k] = \sum_n h[2k-n]x[m,n] \tag{2. 7}$$

Similarly we can take the inner product of (2. 5) with $\psi_{m-1,k}(t)$ and use the orthogonality

of $\psi_{m-1,k}(t)$ and $x_{m-1}(t)$ to obtain:[1]

$$u[m-1,k] = \sum_n g[2k-n]x[m,n] \tag{2. 8}$$

$$g[2k-n] = < \psi_{m-1,k}(t), \phi_{m,n}(t) > \tag{2. 9}$$

As well as obtaining coarser approximations from finer ones, we can also obtain finer approximations from coarser ones. If we were to take the inner product of (2. 5) with $\phi_{m,n}(t)$ we would obtain:

$$x[m,n] = \sum_k h[2k-n]x[m-1,k] + \sum_k g[2k-n]u[m-1,k] \tag{2. 10}$$

Equations (2. 7), (2. 8), and (2. 10) show that we can achieve coarser or finer approximations of $x_m(t)$ without explicitly specifying $\phi(t)$ and $\psi(t)$ but with only knowledge of the sequences h[n] and g[n] defined in (2. 4) and (2. 9).

Ingrid Daubechies in [2] introduces an additional constraint referred to as the regularity condition that will not be discussed in this thesis. The constraint arises in order to insure that the continuous-time signal $x_m(t)$ converges to x(t) as $m \to \infty$. In this thesis we will concentrate on analyzing a discrete-time signal. A discrete-time signal, x[n], can be analyzed in the same framework as its continuous-time counterpart, x(t). Simply think of the process starting at some finite level m instead of $m=\infty$. The signal x[n] can then be represented as the coefficents x[m,n]. It is important to realize that if the signal is discrete

1. Note g[2k-n], like h[2k-n], is not a function of the scale m.

in nature then convergence to the continuous-time signal is not an issue.

In order to specify a wavelet transform, we can determine $\phi(t)$ and $\psi(t)$ by obtaining the functions that satisfy the necessary properties discussed earlier. However there has been a great deal of interest in determining the sequences h[n] and g[n] without ever explicitly determining $\phi(t)$ and $\psi(t)$ since we can achieve coarser or finer approximations of the signal without this knowledge. The literature [9,31] that attacks the problem from this perspective typically does not mention the wavelet transform but refers to the sequences h[n] and g[n] as Quadrature Mirror Filters (QMF). In the next section we discuss the properties that the QMF sequences must satisfy for them to be used in the multiscale analysis discussed in this section.

2.2 PERFECT RECONSTRUCTION OF QMF FILTERS

In this section we investigate the constraints on the sequences h[n] and g[n] rather than the functions $\phi(t)$ and $\psi(t)$. Let us consider the three major equations from the last section that allow us to achieve finer or coarser approximations:

$$x[m-1,k] = \sum_n h[2k-n]x[m,n] \tag{2. 11}$$

$$u[m-1,k] = \sum_n g[2k-n]x[m,n] \tag{2. 12}$$

$$x[m,n] = \sum_k h[2k-n]x[m-1,k] + \sum_k g[2k-n]u[m-1,k] \tag{2. 13}$$

Clearly if we substitute (2. 11) and (2. 12) into the right-hand side of (2. 13) we obtain the following equation:

$$x[m,n] = \sum_{k,p} h[2k-n]h[2k-p]x[m,p] + \sum_{k,p} g[2k-n]g[2k-p]x[m,p] \tag{2. 14}$$

The above equation shows that we need to choose h[n] and g[n] in such a manner that an arbitrary input x[m,n] is mapped back to itself. Since (2. 14) must hold for all choices of x[m,n], we can express a constraint on the operator that maps the sequence x[m,n] back to itself:

$$\delta_{n,p} = \sum_{k} (h[2k-n]h[2k-p] + g[2k-n]g[2k-p]) \qquad (2.\ 15)$$

The above constraint is referred in the QMF literature [31] as the perfect reconstruction condition. In order to gain additional insight we now derive this constraint in a slightly different manner.

In order to analyze the constraints on the sequences h[n] and g[n], let us consider the operator acting upon the signal x[m,n] of Equation (2. 11):

$$\sum_{n} h[2k-n] \qquad (2.\ 16)$$

This operator can be performed in two separate steps. The steps being convolution followed by 2:1 decimation. Since both of these steps are easily represented using z-transforms, we will depict these steps as:

Fig. 2. 1. A key operation, convolution followed by 2:1 decimation, in the computation of the wavelet transform of a signal.

Using the idea that we can express the operator of Equation (2. 16) pictorially as Fig. 2. 1, we display Equations (2. 11) and (2. 12) in block diagram form, as shown on the left-hand side of Fig. 2. 2. The division of a signal in this manner is typically referred to in the signal processing community as a sub-band coder [9]. The block diagram for (2. 13) is depicted on the right-hand side of Fig. 2. 2 and referred to as a sub-band

decoder.

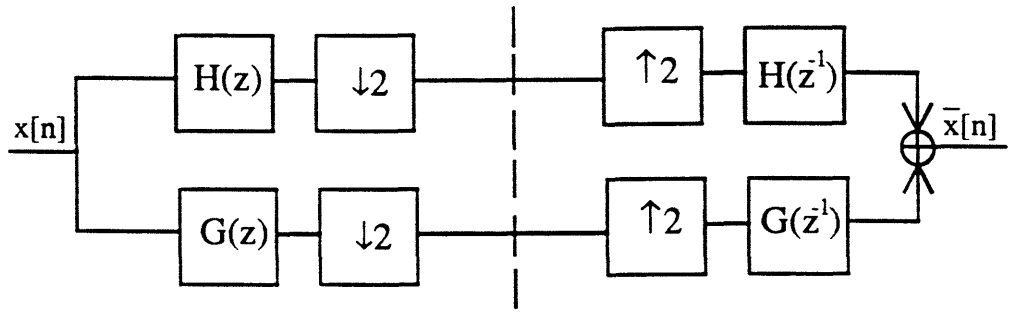


Fig. 2. 2. Sub-band coder and decoder

If x[n] is the input to Fig. 2. 2, then the z-transform of the output $\bar{x}[n]$ is:[2]

$$\bar{X}(z) = \frac{1}{2}\left[H(z)H(z^{-1}) + G(z)G(z^{-1})\right] X(z) + \frac{1}{2}\left[H(-z)H(z^{-1}) + G(-z)G(z^{-1})\right] X(-z) \tag{2. 17}$$

In order to achieve perfect reconstruction we require that $\bar{X}(z) = X(z)$, and to achieve this we would like the second term (the aliasing component) of (2. 17) to be identically zero, as expressed below:

$$\frac{1}{2}[H(z)H(-z^{-1})+G(z)G(-z^{-1})] = 0 \tag{2. 18}$$

We also require that the first term of (2. 17) to be identically X(z). This requirement becomes:

2. Note $\bar{X}(z)$, X(z), G(z), and H(z) are the z-transforms of $\bar{x}[n]$, x[n], g[n], and h[n] respectively.

$$\frac{1}{2}[H(z)H(z^{-1})+G(z)G(z^{-1})] = 1 \qquad (2.19)$$

Note that Equations (2. 18) and (2. 19) provide exactly the same information as the perfect reconstruction condition, Equation (2. 15).

A compact notation [31] for expressing Equations (2. 18) and (2. 19) is by requiring that the matrix:

$$\mathbf{H}(z) = \frac{1}{\sqrt{2}}\begin{bmatrix} H(z) & G(z) \\ H(-z) & G(-z) \end{bmatrix} \qquad (2.20)$$

be orthogonal[3] [32]. Therefore the prefect reconstruction condition can be stated as requiring $\mathbf{H}(z)$ to be an orthogonal matrix.

Although many choices of G(z) will satisfy (2. 18) as discussed in [2,9], the one that seems particularly important at this time is the following:

$$G(z) = z^{-(W-1)}H(-z^{-1}) \qquad (2.21)$$

or in the time domain:

$$g[n] = (-1)^{n+1}h[W-1-n] \qquad (2.22)$$

It is easy to verify that Equation (2. 21) satisfies Equation (2. 18) for any W that is an even integer. Although h[n] and g[n] can have infinite impulse responses (IIR), we will limit our attention to the case when they are finite length sequences (FIR). We will typically choose W to be the support of the filter h[n]. This guarantees that the filter g[n] is non-zero in the same region as h[n]. For example, if we choose h[n] to be the causal filter being non-zero between 0 and W-1, then Equation (2. 21) will generate a filter g[n]

3. The reader should note that since z is being evaluated on the unit circle, the adjoint operation of $\mathbf{H}(z)$ is $\mathbf{H}^{T}(z^{-1})$.

that is also causal and non-zero between 0 and W-1. Note that there is no loss of generality in assuming that the support of the filter h[n] is even. As shown later in this thesis, the only filters that satisfy Equation (2. 19) with the substitution of Equation (2. 21) are filters with even support.

If we substitute Equation (2. 21) into Equation (2. 19), then we obtain the following equation:

$$\frac{1}{2}[H(z)H(z^{-1})+H(-z)H(-z^{-1})] = 1 \qquad (2.\ 23)$$

The constraint of (2. 23) typically appears in the context of sub-band coding. The constraint is commonly referred to as the Quadrature Mirror Filter (QMF) Condition [9]. An important point to realize is that if the choice of the FIR filter h[n] is made, the filter g[n] can be determined by using (2. 22). Thus the procedure is completely defined, once h[n] satisfying the QMF condition is specified.

2.3 PARAMETRIZATION OF QMF FILTERS

The purpose of this section is to parametrize the class of FIR filters that obey the QMF condition. We will show that we can construct a parametrization, originally developed by Vaidyanathan [31], such that for any choice of the parameters we are confined to stay within the set of QMF filters. Furthermore as stated in [33] the parametrization we will describe actually characterizes all FIR filters that obey the QMF condition.

To begin let us consider the cascade of the following matrices:

$$\frac{1}{\sqrt{2}}\begin{bmatrix} H(z) & G(z) \\ H(-z) & G(-z) \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & z^{-2} \end{bmatrix}\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \tag{2.23}$$

If the sequences h[n] and g[n] were chosen such that the first matrix is an orthogonal matrix, then the cascade of these matrices is also orthogonal, since the second and third matrices are orthogonal. Also note that the matrix that results in the cascade of these matrices will have the desired property that its second row is equivalent to its first row with z replaced by -z. Thus we are led to believe that we can develop a recursion for creating filters that satisfy the QMF constraint.

This recursion can be stated as follows, where we introduce a superscript to denote the level of the recursion:

$$\mathbf{H}^{i+1}(z) = \mathbf{H}^{i}(z)\begin{bmatrix} 1 & 0 \\ 0 & z^{-2} \end{bmatrix}\begin{bmatrix} \cos\theta_{i+1} & \sin\theta_{i+1} \\ -\sin\theta_{i+1} & \cos\theta_{i+1} \end{bmatrix} \tag{2.25}$$

$$\mathbf{H}^{i}(z) = \frac{1}{\sqrt{2}} \begin{bmatrix} H^{i}(z) & G^{i}(z) \\ H^{i}(-z) & G^{i}(-z) \end{bmatrix} \tag{2. 26}$$

Thus we obtain a recursive method for creating QMF filters in the z-transform domain:

$$H^{i+1}(z) = \cos \theta_{i+1} \, H^{i}(z) - \sin \theta_{i+1} \, z^{-2} \, G^{i}(z) \tag{2. 27}$$

$$G^{i+1}(z) = \sin \theta_{i+1} \, H^{i}(z) + \cos \theta_{i+1} \, z^{-2} \, G^{i}(z) \tag{2. 28}$$

or equivalently, in the time domain:

$$h^{i+1}[n] = \cos \theta_{i+1} \, h^{i}[n] - \sin \theta_{i+1} \, g^{i}[n-2] \tag{2. 29}$$

$$g^{i+1}[n] = \sin \theta_{i+1} \, h^{i}[n] + \cos \theta_{i+1} \, g^{i}[n-2] \tag{2. 30}$$

Note that if $h^{i}[n]$ and $g^{i}[n]$ are FIR with support W then $h^{i+1}[n]$ and $g^{i+1}[n]$ will be FIR with support W+2. Thus a method for obtaining even length filters[4] that satisfy the perfect reconstruction property is to use the recursive relationships of Equations (2. 29) and (2. 30) with some type of initialization. The initialization that is the most relevant involves the parametric description of all two-tap filters. Then using the above recursion we can describe all two-tap, four-tap, six-tap, ... filters. The two-tap initialization [34] can be expressed as follows:

4. The support of perfect reconstruction filters is necessarily even as discussed in the following section.

$$h^1[n] = \begin{cases} \cos\theta_1 & n = 0 \\ \sin\theta_1 & n = 1 \\ 0 & \text{Otherwise} \end{cases} \qquad (2.31)$$

$$g^1[n] = \begin{cases} -\sin\theta_1 & n = 0 \\ \cos\theta_1 & n = 1 \\ 0 & \text{Otherwise} \end{cases} \qquad (2.32)$$

where $g^1[n]$ was obtained from Equation (2. 22).

It is easily shown that these initializing sequences (2. 31) and (2. 32) satisfy the perfect reconstruction property for all values of θ_1. Furthermore, we can show that these sequences actually characterize all two-tap filters that are QMF. The sequences produce a particularly interesting example when $\theta_1 = \frac{\pi}{4}$. In this case the sequences are the ones that are used in the well known **Haar Transform** [2].

Thus we have described a procedure to generate W-tap filters expressed as a function of L = W/2 free parameters $\{\theta_1, \theta_2, \cdots, \theta_L\}$. For notational simplicity, we will subsequently denote this set of parameters as $\theta = \{\theta_1, \theta_2, \cdots, \theta_L\}$. We have shown that we can construct filters such that any choice of θ corresponds to a sequence h[n] that satisfies the QMF condition. Also note that we can generate g[n] simultaneously as we construct h[n] from the recursive relationships or, equivalently, we can construct g[n] via the one-to-one relationship between h[n] and g[n] as expressed in Equation (2. 22).

2.4 THE DISCRETE-TIME WAVELET TRANSFORM FOR FINITE LENGTH SEQUENCES

In this section we address the issues that arise when dealing with signals that have finite length. If a signal has finite length and undergoes linear convolution and 2:1 decimation, then, in general, the resulting sequence will have support greater than half of the support of the original sequence. However, we desire the resulting sequence's support to be exactly half of the support of the original sequence. This requirement will be useful to develop a wavelet transform that preserves the number of degrees of freedom in going from the time domain to the wavelet domain.

Suppose that we wish to perform an M-level wavelet transformation of our original signal. In this case our original signal is denoted by x[M,n], and its coarser scale transforms by x[m,n], $0 \leq m \leq M-1$. We assume also than x[M,n] has finite support of at most length N, i.e., x[M,n] = 0 if $n \leq 0$ or $n > N$. Also to preserve the number of degrees of freedom, we perform the convolution step on the original fine-scale signal as an N-point circular convolution. Furthermore we assume that N is a multiple of 2^M, i.e. $N = C\,2^M$, $C \in \{1,2,3,\ldots\}$ which guarantees that the support of x[m,n], for $m \leq M$, is at most $C\,2^m$. Note that since we are using cyclic convolution, we may encounter edge distortion due to cyclic wrap-around. Typically the support of the QMF filters h[n] and g[n] is quite small compared to N so this effect is only seen in small boundary regions. In our context, in which we are considering the transformation of stochastic processes and the (hoped-for) approximate diagonalizations of their covariance matrices, the use of cyclic convolution will introduce some edge effects. While these edge effects would need to be recognized and dealt with in practice, we will not concern ourselves with them here

in order to focus our attention on the new issues concerning the use of the wavelet transform.

Note further that all of the previous analysis using z-transforms is still valid even though we are using cyclic convolution. For example, suppose that we consider performing the analysis on an approximation of the signal x[m,n]. In this case, we simply consider the z-transforms (H(z), G(z), X(z), $\overline{X}(z)$) to have the property that z^K is identically equal to one, where $K = N/2^{M-m}$.

A useful method for describing the analysis and synthesis operations on finite length sequences involves the use of matrix notation. Let $X_m \in \mathbb{R}^K$ denote the column vector whose entries are the values of x[m,n] over its support. That is,

$$X_m = \begin{bmatrix} x[m, 1] \\ x[m, 2] \\ \vdots \\ x[m, K] \end{bmatrix} \tag{2. 33}$$

Since we are using circular convolution and have chosen K to be even, we know that the coarser approximation $X_{m-1} \in \mathbb{R}^{K/2}$. The matrix H_{m-1} that performs the mapping of X_m to X_{m-1} must have dimension $H_{m-1} \in \mathbb{R}^{(K/2,\ K)}$. In a similar manner, we define G_{m-1} as the matrix that maps X_m to the detail at its coarser level U_{m-1}. We rewrite Equations (2. 11)(2. 12) and(2. 13) in equivalent matrix equations:

$$X_{m-1} = H_{m-1} X_m \tag{2. 34}$$

$$U_{m-1}=G_{m-1}X_m \tag{2.35}$$

$$X_m = H_{m-1}^T X_{m-1} + G_{m-1}^T U_{m-1} \tag{2.36}$$

Note that Equation (2. 36) uses the transpose of the matrices that are used in Equations (2. 34) and (2. 35). This is due to the fact that the operator in (2. 11) is the adjoint [2] of the one in (2. 13).

We illustrate the particular form that the matrices H_{m-1} and G_{m-1} must take on by considering an example when K=8 and the support of the filter is four. In this case the matrices H_{m-1} and G_{m-1} are:

$$H_{m-1} = \begin{bmatrix} h[1] & h[0] & 0 & 0 & 0 & 0 & h[3] & h[2] \\ h[3] & h[2] & h[1] & h[0] & 0 & 0 & 0 & 0 \\ 0 & 0 & h[3] & h[2] & h[1] & h[0] & 0 & 0 \\ 0 & 0 & 0 & 0 & h[3] & h[2] & h[1] & h[0] \end{bmatrix} \tag{2.37}$$

$$G_{m-1} = \begin{bmatrix} g[1] & g[0] & 0 & 0 & 0 & 0 & g[3] & g[2] \\ g[3] & g[2] & g[1] & g[0] & 0 & 0 & 0 & 0 \\ 0 & 0 & g[3] & g[2] & g[1] & g[0] & 0 & 0 \\ 0 & 0 & 0 & 0 & g[3] & g[2] & g[1] & g[0] \end{bmatrix} \tag{2.38}$$

Note the use of cyclic convolution has introduced the wrap-around effect we have mentioned earlier. In this example the first coefficient of X_{m-1} has a contribution from both the last two coefficients of X_m as well as the first two coefficients. Further note that there is no wrap-around effect when the support of the filter is two.

Using this matrix notation we can rewrite the perfect reconstruction condition of Equation (2. 15) in the following form, which could also be obtained by substituting Equations (2. 34) and (2. 35) into Equation (2. 36):

$$H_{m-1}^T H_{m-1} + G_{m-1}^T G_{m-1} = I \tag{2. 39}$$

Furthermore, if we define the matrix $A_{m-1} \in \mathbb{R}^{(K, K)}$ as follows then Equation (2. 39) has a very simple form

$$A_{m-1} = \begin{pmatrix} H_{m-1} \\ G_{m-1} \end{pmatrix} \tag{2. 40}$$

so that (2. 39) becomes:

$$A_{m-1}^T A_{m-1} = I \tag{2. 41}$$

Equation (2. 41) and the fact that A_{m-1} is a square matrix implies that A_{m-1} is an orthogonal matrix so that we also have:

$$A_{m-1} A_{m-1}^T = I \tag{2. 42}$$

Equation (2. 42) expresses four equations that are functions of the matrices H_{m-1} and G_{m-1}. These equations are stated as follows:

$$\begin{pmatrix} H_{m-1}H_{m-1}^T & H_{m-1}G_{m-1}^T \\ G_{m-1}H_{m-1}^T & G_{m-1}G_{m-1}^T \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} \tag{2. 43}$$

We can express these matrix equations using the index notation discussed earlier. The first of these equations:

$$H_{m-1}H_{m-1}^{T} = I \tag{2. 44}$$

could be described using index notation as:

$$\sum_{n} h[n]h[n-2k] = \delta_{k0} \tag{2. 45}$$

It is easily verified that Equation (2. 45) is equivalent to Equation (2. 44).

We can use Equation (2. 45) to show that if the QMF filter has finite support (and the support of the QMF filter is greater than one) then the support of the QMF filter must be even. To do this we need only show that if the support of h[n] is less than or equal to W, for W odd (and greater than one), then the support of h[n] must actually be less than or equal to W-1. This claim is proven by choosing k=(W-1)/2 in Equation (2. 45). For this choice of k, Equation (2. 45) implies that h[0]h[W-1] = 0. Thus since either h[0] or h[W-1] must be zero, we see that the support of h[n] must indeed be less than or equal to W-1.

Consider now an M-level circulant wavelet transformation. Specifically, we consider transforming a fine-scale scale signal X_M to a coarser approximation X_{M-1} and the detail at that level U_{M-1}. The approximation X_{M-1} is then transformed into a coarser approximation X_{M-2} and the detail at that level U_{M-2}. This procedure is continued for all M-levels obtaining the transformed coefficients $\{X_0, U_0, U_1, \cdots, U_{M-1}\}$. Let $Y_M \in \mathbb{R}^N$ denote the column vector that is formed from the transformed coefficients in the following manner:

$$Y_M = \begin{bmatrix} X_0 \\ U_0 \\ U_1 \\ \vdots \\ U_{M-2} \\ U_{M-1} \end{bmatrix} \tag{2.46}$$

Note that the dimensionality of the vector Y_M is exactly equal to the dimensionality of the original signal X_M. This is due to the fact that the circular wavelet transform preserves the number of degrees of freedom in going from one domain to the other. Furthermore the transformation T_M from the signal X_M to its transformed coefficients Y_M is an orthogonal transformation. Specifically, we use Equations (2. 34) and (2. 35) to express T_M as:

$$T_M = \begin{bmatrix} H_0H_1 \ldots . H_{M-1} \\ G_0H_1 \ldots . H_{M-1} \\ G_1H_2 \ldots . H_{M-1} \\ \vdots \\ G_{M-2}H_{M-1} \\ G_{M-1} \end{bmatrix} \tag{2.47}$$

To show that the matrix T_M is an orthogonal matrix, we simply express Equation (2. 47) as the cascade of M matrices:

$$T_M = \cdots \begin{bmatrix} H_{M-2} & \vdots & 0 \\ G_{M-2} & \vdots & 0 \\ \cdots & \vdots & \cdots \\ 0 & \vdots & I \end{bmatrix} \begin{bmatrix} H_{M-1} \\ G_{M-1} \end{bmatrix} \tag{2.48}$$

We know that the right-most matrix of Equation (2. 48) is an orthogonal matrix. The matrices to its left are all orthogonal since they are composed of orthogonal matrices along their diagonal. Using the fact that the cascade of orthogonal matrices is orthogonal, we conclude that the matrix T_M is an orthogonal matrix.

With these definitions, we know that the signal and its wavelet coefficients are related via the following equations:

$$Y_M = T_M X_M \tag{2.49}$$

$$X_M = T_M^T Y_M \tag{2.50}$$

Furthermore, the computation of the wavelet transform and its inverse, i.e. Equations (2. 49) and (2. 50), are very efficient. Specifically, the transform, or its inverse, can be computed using a direct implementation that requires on O(WN) multiplies, where O(·) represents the order of the argument. This result is easily obtained by counting the number of multiplies that are required to multiply the orthogonal matrix in Equation (2. 48) by a signal, X_M, that has support N. The right-most matrix of Equation (2. 48) multiplied by X_M would require NW multiplies since A_{M-1} has at most W non-zero entries per row and N rows. The second matrix from the right in Equation (2. 48) operating on $A_{M-1}X_M$ would require WN/2 multiplies since multiplying by the identity does not require any additional computation. Continuing in this manner we obtain a

simple upper bound for the number of multiplies in computing the transform:

$$WN + WN/2 + WN/4 + \ldots \leq 2WN \tag{2.51}$$

thus leading to the statement that a direct implementation of the wavelet transform requires O(WN) multiplies. A more sophisticated implementation of the transform is given in [8] where the authors show that the computation of the wavelet transform only requires performing $O(N\log_2 W)$ multiplies. Since, typically, W is much smaller then N, the wavelet transform is a very efficient operation.

CHAPTER 3

MULTISCALE STATISTICAL MODELING

3.1 MULTISCALE STOCHASTIC PROCESSES

Let us consider constructing a multiscale stochastic process by synthesizing it using the wavelet transform. More precisely, by beginning with a coarse (stochastically specified) approximation X_0 of a signal we can add via Equation (3. 1), additional stochastic detail U_0 that is independent of the signal X_0 to obtain a finer level signal, X_1. We can then continue with this process by adding an additional detail U_1 that is independent of X_1 to obtain an even finer scale, X_2. This method can then be repeated to an arbitrary number of scales. The general recursion then is a stochastic dynamic system in scale:

$$X_m = H_{m-1}^T X_{m-1} + G_{m-1}^T U_{m-1} \tag{3. 1}$$

The problem we are trying to solve can be stated as follows. Let us suppose that we have a statistical characterization of a fine level signal. We wish to model the fine level signal as being synthesized from the wavelet transform as prescribed in the above manner, that is, we wish to model our signal as being X_M for some value of M. The major question to be answered is what are the best QMF filters h[n] and g[n] such that the cross-covariance between the approximation of the process X_m, at any coarser level m, and the detail at that level U_m is as small as possible for all m.

In other words, given the autocovariance $\Lambda_{X_M X_M}$ of the process X_M we would like to determine the QMF filters h[n] and g[n] such that the following equations are satisfied as closely as possible (in a manner to be made precise):

$$\mathrm{Cov}\{X_m, U_m\} = 0 \qquad \forall\ m = 0, 1, 2, \ldots, M\text{-}1 \tag{3.2}$$

where $\mathrm{Cov}\{\cdot\ ,\cdot\ \}$ represents the covariance between the two arguments.

This problem, however, is very difficult to solve since the coefficients of the matrix of equation (3. 2) is a high-order nonlinear function of the coefficents of h[n]. This is demonstrated very easily. Recall that we can express X_m and U_m as a function of the finer scale signal X_{m+1}:

$$X_m = H_m X_{m+1} \tag{3.3}$$

$$U_m = G_m X_{m+1} \tag{3.4}$$

By substituting Equations (3. 3) and (3. 4) into the left hand side of Equation (3. 2) we obtain:

$$\mathrm{Cov}\{X_m, U_m\} = H_m \mathrm{Cov}\{X_{m+1}, X_{m+1}\}\, G_m^T \tag{3.5}$$

$$= H_m\, \Lambda_{X_{m+1}X_{m+1}}\, G_m^T \tag{3.6}$$

By using Equation (3. 3) we obtain the autocovariance of the mth scale signal from the autocovariance of the m+1st scale signal using the following equation:

$$\Lambda_{X_m X_m} = H_m \Lambda_{X_{m+1}X_{m+1}} H_m^T \tag{3.7}$$

So by using Equations (3. 6) and (3. 7), we express the cross-covariance between X_m and

U_m, for any m, as a function of the known autocovariance $\Lambda_{X_M X_M}$. For example when m=M-2, the cross-covariance can be stated as:

$$\mathrm{Cov}\{X_{M-2}, U_{M-2}\} = H_{M-2} H_{M-1} \Lambda_{X_M X_M} H_{M-1}^T G_{M-2}^T \qquad (3.8)$$

It is clear for the example m=M-2 that the coefficients of the above matrix have in general a fourth-order dependency on the coefficients of h[n]. Thus, in general, the cross-covariance between the coarsest signal approximation X_0 and the detail at that level U_0 will be a nonlinear function of the coefficients of h[n] of order 2M.

Problems of this high nonlinearity are typically very difficult and computationally intensive to solve. So in order to simplify the computation we will not require that the same QMF filters be used at each level. We will now continue to use the same matrix notation H_m and G_m but there matrices are now constructed from the sequences $h_m[n]$ and $g_m[n]$, respectively, which may be different for different values of m. As is evident from the subscript, the QMF sequences can be different for different levels. With this assumption in mind let us proceed in formulating a criterion that captures our objective.

To begin let us recall that we can parametrize H_m and G_m to stay within the set of perfect reconstruction filters. To denote the parametrization of H_m and G_m explicitly, we will write these matrices as $H_m(\theta_m)$ and $G_m(\theta_m)$, respectively. Thus we restate our problem as trying to find the θ_m for all m that best satisfies the following equations:

$$H_m(\theta_m)\, \Lambda_{X_{m+1} X_{m+1}}\, G_m^T(\theta_m) = 0 \qquad \forall\, m = 0, 1, 2, \ldots, M\text{-}1 \qquad (3.9)$$

where $\Lambda_{X_M X_M}$ is known and $\Lambda_{X_m X_m}$ for $m < M$ can be computed recursively via the

following equations:

$$\Lambda_{X_m X_m} = H_m(\theta_m)\Lambda_{X_{m+1}X_{m+1}}H_m^T(\theta_m) \qquad \forall\ m = 0, 1, 2, \ldots, M\text{-}1 \qquad (3.10)$$

Ideally we would like every element of the matrices in the equations of (3. 9) to be zero, so the criterion we use is to try to find the values of θ_m, for all m, such that a weighted sum of the norm-squares of the matrices in the equations of (3. 9) is as small as possible. Thus we state this criterion formally as:

$$\min_{\theta_0,\theta_1,\ldots,\theta_{M-1}} \sum_{m=1}^{M} \alpha_m \left\| H_{m-1}(\theta_{m-1})\, \Lambda_{X_m X_m}\, G_{m-1}^T(\theta_{m-1}) \right\|_F^2 \qquad (3.11)$$

where $\|\cdot\|_F$ represents the Frobenius norm[5] [35] and α_m is the weight associated with the norm of the mth equation of (3. 9).

This minimization problem is difficult to solve since we are trying to perform the minimization of a highly nonlinear criterion over such a large number of arguments. In particular for each m, $\Lambda_{X_m X_m}$ is an implicit nonlinear function of $\theta_m, \theta_{m+1}, \ldots, \theta_{M-1}$. Thus the optimal choice of θ_{M-1} is determined not only by the desire to decorrelate X_{M-1} and U_{M-1} but also by the influence this choice has on subsequent levels of the decomposition. If we neglect this secondary influence on the choice of each θ_m, we obtain a further simplification in which we perform M successive minimization problems for each θ_m in turn. Specifically, we begin by performing the following optimization:

5. The Frobenius norm of a matrix is the square root of the sum of the squares of the elements of the matrix.

$$\min_{\theta_{M-1}} \left\| H_{M-1}(\theta_{M-1})\, \Lambda_{X_M X_M}\, G^T_{M-1}(\theta_{M-1}) \right\|^2_F \qquad (3.12)$$

We can then use the optimal θ_{M-1} to obtain the first level quadrature mirror filter pair and to compute:

$$\Lambda_{X_{M-1}X_{M-1}} = H_{M-1}(\theta_{M-1})\Lambda_{X_M X_M} H^T_{M-1}(\theta_{M-1}) \qquad (3.13)$$

This procedure can then be repeated for all M levels to find the best QMF pair at each level. Although (3. 12) is computationally easier to solve than Equation (3. 11), it is still highly nonlinear and deserves additional investigation as to the methodology used for solving this optimization problem. In the next section we develop an efficient method for obtaining the solution of Equation (3. 12).

3.2 SOLVING THE MINIMIZATION PROBLEM

The purpose of this section is to describe an efficient and reliable method for solving the following optimization problem:

$$\min_{\theta} \left\| H(\theta)\, \Lambda_{XX}\, G^{T}(\theta) \right\|_{F}^{2} \tag{3.14}$$

The optimization in Equation (3. 14) is equivalent to the optimization described in Equation (3. 12). For notational simplicity we have simply dropped the subscript notation that denotes the scale.

First let us note that Equation (3. 14) is a problem in nonlinear programming that we can solve using standard methods [36]. The nonlinear optimization routines that solve this problem typically work by finding the closest local minimum to the initial guess that was passed to the routine by the user. The user must determine the locations of the initial guesses for which the local minima will be computed. Then from this set of local minima the smallest of these is chosen as the solution to the minimization problem.

The problem of designing an optimization routine is further complicated if the optimization is over a vector of variables, as multidimensional searches for the local and global minima are in general very time-consuming. In some problems, however, there are natural ways in which to perform optimizations over a sequence of lower-dimensional subspaces using the lower dimensional subspaces to guide the higher-dimensional ones. One methodology for doing this is to first fix all but one of the variables and minimize the function over the one free variable. Then as a second step one fixes all but two of the variables and minimizes over the two free variables. The solution of the first minimization is typically used as the initial guess for the second minimization. This procedure can be continued for all of the variables of the minimization.

The problem here, of course, is finding a natural way in which to "fix" some of the variables at each stage. Fortunately in our context there is an extremely natural way in which to do that. Using the notation of Section 2.3 let us consider QMF filters that have support $W = 2L$. Thus θ will consist of the L variables $\theta = \{\theta_1, \theta_2, \theta_3, \ldots, \theta_L\}$. The recursive description in Section 2.3 for constructing a QMF from such a vector then directly leads to the structure of the optimization procedure that we use. Specifically, we begin by choosing θ_1 to yield the best QMF of support two. This provides the starting point for choosing θ_1, θ_2 to provide the best QMF of support four. This procedure can then be continued until we have reached the final desired support. To make this precise, let us begin with the first step, where the two-tap QMF pair has the form:

$$h_\theta^1[n] = \left\{ \begin{array}{cc} \cos\theta_1 & n = 0 \\ \sin\theta_1 & n = 1 \\ 0 & \text{Otherwise} \end{array} \right\} \qquad (3.15)$$

$$g_\theta^1[n] = \left\{ \begin{array}{cc} -\sin\theta_1 & n = 0 \\ \cos\theta_1 & n = 1 \\ 0 & \text{Otherwise} \end{array} \right\} \qquad (3.16)$$

The subscript on the filters h[n] and g[n] is used to make clear that the filters are functions of θ, and the superscript is used to denote the number of components of θ on which these filters depend. The first step of our minimization problem then is to solve the following one dimensional minimization:

$$\min_{\theta_1} \left\| H(\theta_1)\, \Lambda_{XX}\, G^T(\theta_1) \right\|_F^2 \qquad (3.17)$$

where $H(\theta_1)$ and $G(\theta_1)$ are computed from $h_\theta^1[n]$ and $g_\theta^1[n]$ as described previously.

Since the objective function of Equation (3. 17) has a periodicity of $\pi/2$ (see Appendix A), we can perform the minimization of (3. 17) by choosing a number of initial guesses evenly spaced between 0 and $\pi/2$ to obtain several local minima (using any nonlinear programming routine) that can be compared to determine the global minimum. It is important to realize that even though the objective function has a periodicity of $\pi/2$, the filters do not. Therefore if θ_1^1 denotes the optimal solution to the one-parameter minimization problem then both θ_1^1 and $\theta_1^1 + \pi/2$ are possible solutions to the minimization. An interesting point about these solutions is that the two solutions correspond to the two alternatives specifying which filter is h[n] and which is g[n] (see Appendix A).

The above comments regarding the periodicity of the function motivates our procedure for the following two-parameter minimization problem:

$$\min_{\theta_0,\theta_1} \left\| H(\theta_0,\theta_1)\, \Lambda_{XX}\, G^T(\theta_0,\theta_1) \right\|_F^2 \qquad (3.18)$$

A reasonable set of initial guesses for the two-parameter minimization problem would then be to use θ_1^1 as the guess for θ_1 and to choose several guesses for θ_2 between 0 and $\pi/2$. However $\theta_1^1 + \pi/2$ could also be a solution to the one-parameter minimization problem so we should also try initial guesses of $\theta_1^1 + \pi/2$ and again vary θ_2 between 0 and $\pi/2$. Out of both sets of local minima resulting from these two sets of initial guesses,

we obtain the global minimum (θ_1^2,θ_2^2) for the two-parameter solution. Again using the periodicity result of Appendix A, we find that both (θ_1^2,θ_2^2) and $(\theta_1^2,\theta_2^2+\pi/2)$ are solutions to the 2-parameter minimization problem.

We can continue this procedure for all L parameters. We will then have two possible solutions to the minimization problem of Equation (3. 12): $(\theta_1^L, \theta_2^L, \ldots, \theta_L^L)$ and $(\theta_1^L, \theta_2^L, \ldots, \theta_L^L+\pi/2)$. Either solution would yield the same global minimum but we would like to choose the one that makes h[n] the low-pass filter and g[n] the high-pass filter. Thus we pick the solution such that h[n] has the greatest DC content or equivalently the solution that satisfies the following equation:

$$\left|\sum_n h[n]\right| \geq \left|\sum_n g[n]\right| \qquad (3.19)$$

CHAPTER 4

ANALYZING PROCESSES USING THE WAVELET TRANSFORM

4.1 STATIONARY PROCESSES

In this chapter, we examine the ability of our models to approximate the statistical description of several processes. In particular we examine the wavelet transformation of the statistics of these processes and illustrate the ability of this transform to achieve approximate block diagonalization of the covariance matrix of the transformed coefficients. The first three sections of this chapter deal with analyzing stationary processes. Specifically, we consider the first-order Gauss-Markov process and under-damped second-order oscillatory processes. The final section of this chapter considers a class of non-stationary processes known as fractional Brownian motions.

We will denote the covariance between the coefficients x[m,i] and x[m,j] as $\Lambda_{X_m X_m}(i, j) = \mathrm{Cov}\{x[m,i], x[m,j]\}$. Of course with this notation, the covariance $\Lambda_{X_m X_m}(i, j)$ is the element in the ith row and jth column of the autocovariance matrix $\Lambda_{X_m X_m}$. For the case of stationary processes, $\Lambda_{X_m X_m}(i, j)$ does not explicitly depend upon both i and j but only upon the difference i-j. In the case of these processes we write:

$$\Lambda_{X_m X_m}(i, j) = \Lambda_{X_m X_m}(i-j, 0) \tag{4. 1}$$

We will abuse notation and drop the explicit dependence upon the origin to express the autocovariance of the stochastic process as:

$$\Lambda_{X_m X_m}(i, j) = \Lambda_{X_m X_m}(i-j) \tag{4. 2}$$

A useful measure that we will use in analyzing the degree of correlation between two samples of a stochastic process is the normalized autocovariance (correlation coefficient). We define the ith row and jth column of the normalized autocovariance matrix $P_{X_mX_m}$ as:

$$P_{X_mX_m}(i, j) = \frac{\Lambda_{X_mX_m}(i, j)}{\sqrt{\Lambda_{X_mX_m}(i, i)\,\Lambda_{X_mX_m}(j, j)}} \tag{4. 3}$$

Some of the properties that the normalized autocovariance possess are $|P_{X_mX_m}(i, j)| \leq 1$ and $P_{X_mX_m}(i, i) = 1$. Note that for stationary processes the normalized autocovariance $P_{X_mX_m}(i, j)$ also depends only on the difference i-j and can be expressed as $P_{X_mX_m}(i, j) = P_{X_mX_m}(i-j)$.

We have already seen that the correlation structure of X_{m-1} and U_{m-1} depends upon the QMFs h[n] and g[n] and the autocovariance $\Lambda_{X_mX_m}$. The correlation between any two coefficients of X_{m-1} and U_{m-1} (when X_m are the samples of a stationary process) is easily determined by using the relationships:

$$x[m-1,k] = \sum_n h[n]x[m,2k-n] \tag{4. 4}$$

$$u[m-1,k] = \sum_n g[n]x[m,2k-n] \tag{4. 5}$$

Using the above two equations we express the covariances of the one-level transformed coefficients as:

$$\Lambda_{X_{m-1}X_{m-1}}(s,t) = \sum_{i,j} h[i]\Lambda_{X_mX_m}(2s - i,2t - j)h[j] \tag{4. 6}$$

$$\Lambda_{U_{m-1}U_{m-1}}(s,t) = \sum_{i,j} g[i]\Lambda_{X_mX_m}(2s - i,2t - j)g[j] \tag{4. 7}$$

$$\Lambda_{X_{m-1}U_{m-1}}(s,t) = \sum_{i,j} h[i]\Lambda_{X_mX_m}(2s - i,2t - j)g[j] \tag{4. 8}$$

If X_m is a stationary process then Equations (4. 6) to (4. 8) imply that X_{m-1} and U_{m-1} are jointly stationary and have the following covariances:

$$\Lambda_{X_{m-1}X_{m-1}}(\tau) = \sum_{i,j} h[i]\Lambda_{X_mX_m}(2\tau - i + j)h[j] \tag{4. 9}$$

$$\Lambda_{U_{m-1}U_{m-1}}(\tau) = \sum_{i,j} g[i]\Lambda_{X_mX_m}(2\tau - i + j)g[j] \tag{4. 10}$$

$$\Lambda_{X_{m-1}U_{m-1}}(\tau) = \sum_{i,j} h[i]\Lambda_{X_mX_m}(2\tau - i + j)g[j] \tag{4. 11}$$

Therefore if the sampled process X_M is stationary, then X_{M-1} and U_{M-1} will be jointly stationary. Since X_{M-1} is stationary, we can use the same argument to show that X_{M-2} and U_{M-2} are jointly stationary. We deduce that if the sampled process is stationary then the coefficients at any particular level, e.g. $\{X_m, U_m\}$, are jointly stationary; however, the transformed coefficients, e.g. $\{U_{M-1},\cdots,U_0, X_0\}$, are not jointly stationary. The transformed coefficients are not jointly stationary since the covariance of wavelet coefficients at different levels do not depend simply on the time difference but they do have the following dependence for fixed k and m:

$$\Lambda_{U_mU_k}(s, t) = \Lambda_{U_mU_k}(s - 2^{-k+m}t) \tag{4. 12}$$

Equation (4. 12) is easily proven for arbitrary m and k by using the continuous-time wavelet notation from Chapter 2. Since it is more difficult to show Equation (4. 12) via the discrete-time wavelet notation we have been using, we will simply verify Equation (4. 12) when m=M-1 and k=M-2. Of course u[M-1, s] and u[M-2, t] are:

$$u[M-1, s] = \sum_{k_1} g(k_1)x[M, 2s - k_1] \tag{4. 13}$$

$$u[M-2, t] = \sum_{k_3} \sum_{k_2} g(k_3)h(k_2)x[M, 4t - 2k_3 - k_2] \tag{4. 14}$$

Using the fact that X_M is stationary, we obtain:

$$\Lambda_{U_{M-1}U_{M-2}}(s, t) = \sum_{k_3} \sum_{k_2} \sum_{k_1} g(k_3)h(k_2)g(k_1)\Lambda_{X_MX_M}(2(s - 2t) - 2k_3 - k_2 - k_1) \tag{4. 15}$$

Therefore as in Equation (4. 12), we will use the notation that:

$$\Lambda_{U_{M-1}U_{M-2}}(s, t) = \Lambda_{U_{M-1}U_{M-2}}(s - 2t) \tag{4. 16}$$

In Appendix B we show that if the sampled process X_M is stationary then the optimal two-tap one-level transformation is the Haar Transform (where optimality is defined by the criterion of Chapter 3). We can combine this fact with our knowledge that X_m for $m < M$ is stationary to deduce that the optimal two-tap filters for all levels will be the filters used in the Haar Transform.

In the next section we look at a particular stationary process, the first-order Gauss-Markov process, and examine in more detail the statistics of the wavelet transform of this process.

4.2 THE FIRST-ORDER GAUSS-MARKOV PROCESS

The first-order continuous-time model that generates a first-order Gauss-Markov process is of the form:

$$\dot{x}(t) = -\beta x(t) + w(t) \tag{4. 17}$$

In the numerical examples that follow we use a discretized version of Equation (4. 17). In particular we sample the process at a sufficiently high rate to minimize any aliasing effects. Of course this criterion means that the sampling rate, $2\pi/T$ where T is the sampling interval, must be equal to twice the effective cutoff frequency ω_0. We choose the effective cutoff frequency to be the frequency such that

$$S_{xx}(\omega_0) = 0.00222 \tag{4. 18}$$

where $S_{xx}(\omega)$ is the power spectral density of x(t). The power-spectral density of Equation (4. 17) is

$$S_{xx}(\omega) = \frac{2\beta}{\beta^2 + \omega^2} \tag{4. 19}$$

when the variance of x(t) is unity. When β is normalized to unity, Equations (4. 19) and (4. 18) imply that $\omega_0 \approx 30$. This analysis yields the following discretized model that will be used throughout this thesis:

$$x[n+1] = \alpha x[n] + w[n] \tag{4. 20}$$

$$\alpha = e^{-\beta T} \approx .9006 \tag{4. 21}$$

where w[n] is a constant variance, zero-mean white noise process and x[n] has unit variance. Since x[n] is stationary it is easily seen that the variance of w[n] must equals $(1 - \alpha^2)$. Assuming that (4. 20) is in steady-state, the autocovariance (which in this case is also the normalized autocovariance) for the process is

$$P_{XX}(\tau) = \alpha^{|\tau|} \tag{4. 22}$$

In Fig. 4. 1 we display the normalized autocovariance, as given in Equation (4. 22), in several different ways. In Fig. 4. 1 (a) the normalized autocovariance is shown by plotting $P_{XX}(\tau)$ versus τ which is the method that is commonly used to view autocovariances of stationary processes. A mesh plot of $P_{XX}(s, t) = P_{XX}(s-t)$ is depicted in Fig. 4. 1 (b). In this plot $P_{XX}(s, t)$ is plotted versus s and t. In Fig. 4. 1 (c) the same plot is shown as Fig. 4. 1 (a) and (b) but plotted as a 2-D image. In this plot, the intensity of the magnitude of $P_{XX}(s, t)$ is plotted versus s and t. The plots in (b) and (c) are useful for viewing autocovariances that are not stationary.

The mapping of the gray-scale to the correlation values that are used in all 2-D images (unless shown otherwise) is shown in Fig. 4. 2. In order to show how the correlation structure decays we will plot the magnitude of the covariance for all 2-D images throughout this section. We use the word "decay" to denote a decrease in the magnitude of the correlation between any two coefficients as the interval between those coefficients increase and one of those coefficients remains fixed.

We now consider taking a one-level wavelet transform of a set of samples of the Gauss-Markov process, which we again collect as a vector denoted by X_M. The transformed coefficients have the property that they decay at a rate faster than the original process. Specifically, let us consider the cross-covariance between the processes X_{M-1} and U_{M-1} for a Gauss-Markov process:

$$\Lambda_{X_{M-1}U_{M-1}}(\tau) = \sum_{i,j} h[i]\,\alpha^{|2\tau - i + j|}\,g[j] \tag{4. 23}$$

An upper bound for the above covariance is obtained by the summation of the absolute value of each term in Equation (4. 23).

$$\Lambda_{X_{M-1}U_{M-1}}(\tau) \le \sum_{i,j} |h[i]|\,|\alpha|^{|2\tau - i + j|}\,|g[j]| \tag{4. 24}$$

The above expression is further bounded by realizing that

$$|\alpha|^{|2\tau - i + j|} \le \begin{cases} |\alpha|^{2|\tau|} - W + 1 & 2|\tau| \ge W - 1 \\ 1 & 2|\tau| < W - 1 \end{cases} \tag{4. 25}$$

for any i and j, where W is the support of the QMFs. Since the right-hand side of Equation (4. 25) is not a function of i and j, we pull it out of the summation to obtain the bound:

$$\Lambda_{X_{M-1}U_{M-1}}(\tau) \le \begin{cases} \|h\|_{l_1}^2\,|\alpha|^{2|\tau|} - W + 1 & 2|\tau| \ge W - 1 \\ \|h\|_{l_1}^2 & 2|\tau| < W - 1 \end{cases} \tag{4. 26}$$

where $\|.\|_{l_1}$ represents the l_1 norm[1] [35]. Note that we have used the fact that the l_1 norm of h[n] is equal to the l_1 norm of g[n] which is easily verified by using Equation (2. 22). It is trivial to show that the bound in Equation (4. 26) will also bound the covariances of the processes X_{M-1} and U_{M-1}.

The key point that we note about this bound is that the covariances of the wavelet coefficients (computed by taking a one-level wavelet transform of a first-order Gauss-Markov process) decays at a much faster rate than the original Gauss-Markov process.

1. The l_1 norm of a sequence is the sum of the absolute value of each element of the sequence.

Specifically, if the autocovariance of the original Gauss-Markov process decays as $O(\alpha^{|\tau|})$, (where $O(\cdot)$ represents the order of) then using the bound in Equation (4. 26) the covariances that result from a one-level wavelet transform ($\Lambda_{X_{M-1}U_{M-1}}(\tau)$, $\Lambda_{X_{M-1}X_{M-1}}(\tau)$, and $\Lambda_{U_{M-1}U_{M-1}}(\tau)$) decay as $O(\alpha^{2|\tau|})$. We direct the reader to Beylkin, Coifman, and Rokhlin [10] for a further discussion of obtaining bounds on the 2-D wavelet transform coefficients when the kernel is in the class of Calderon-Zygmund operators. Although an exponential decay is within this class of operators, the bound that Beylkin *et al.* determined is much weaker than the bound in Equation (4. 26).

Let us now consider taking a one-level Haar Transform of X_M. Using Equation (4. 26) we know that the covariances that result from a one-level Haar transform of the samples of the first-order Gauss-Markov process are upper bounded by:

$$\begin{cases} 2|\alpha|^{2|\tau|} - 1 & \tau \neq 0 \\ 2 & \tau = 0 \end{cases} \tag{4. 27}$$

Specifically, the covariances are:

$$\Lambda_{X_{M-1}X_{M-1}}(\tau) = \begin{cases} \left(\frac{1}{2} + \alpha + \frac{1}{2}\alpha^2\right)\alpha^{2|\tau|-1} & \tau \neq 0 \\ 1 + \alpha & \tau = 0 \end{cases} \tag{4. 28}$$

$$\Lambda_{U_{M-1}U_{M-1}}(\tau) = \begin{cases} \left(-\frac{1}{2} + \alpha - \frac{1}{2}\alpha^2\right)\alpha^{2|\tau|-1} & \tau \neq 0 \\ 1-\alpha & \tau = 0 \end{cases} \tag{4.29}$$

$$\Lambda_{X_{M-1}U_{M-1}}(\tau) = \begin{cases} -\frac{1}{2}\,\mathrm{sgn}(\tau)\left(1-\alpha^2\right)\alpha^{2|\tau|-1} & \tau \neq 0 \\ 0 & \tau = 0 \end{cases} \tag{4.30}$$

where:

$$\mathrm{sgn}(\tau) = \begin{cases} 1 & \tau \geq 0 \\ -1 & \tau < 0 \end{cases} \tag{4.31}$$

The above covariances, for $\alpha = 0.9006$, are depicted in Fig. 4. 3 (a) to (d). The same information is shown using a single mesh plot Fig. 4. 4 (a) and a 2-D image Fig. 4. 4 (b). The gray-scale mapping that is used in Fig. 4. 4 (b) is different than the mapping used in other images so the gray-scale is shown in Fig. 4. 4 (c). Note that the coarse scale information is displayed on the far left-hand side of the mesh plot, while this information is displayed in the upper-left-hand corner of the 2-D image. We will adhere to this convention for future 2-D images. The mesh plots were introduced as a method for the reader to have a better understanding of what we are plotting. Hopefully at this point the reader understands these plots and only needs to visualize one of these forms. Therefore in future plots we will only show the 2-D images.

Of course the normalized autocovariance decays at the same rate as the autocovariance. The normalized autocovariance of the one-level Haar transform of X_M is given by:

$$P_{X_{M-1}X_{M-1}}(\tau) = \begin{cases} \frac{1+\alpha}{2}\alpha^{2|\tau|-1} & \tau \neq 0 \\ 1 & \tau = 0 \end{cases} \tag{4.32}$$

$$P_{U_{M-1}U_{M-1}}(\tau) = \begin{cases} \frac{1-\alpha}{2}\alpha^{2|\tau|-1} & \tau \neq 0 \\ 1 & \tau = 0 \end{cases} \tag{4.33}$$

$$P_{X_{M-1}U_{M-1}}(\tau) = \begin{cases} -\frac{1}{2}\,\mathrm{sgn}(\tau)\sqrt{1-\alpha^2}\,\alpha^{2|\tau|-1} & \tau \neq 0 \\ 1 & \tau = 0 \end{cases} \tag{4.34}$$

The above normalized covariances are depicted in Fig. 4. 5 (a) by showing the 2-D image of the one-level normalized autocovariance of the Haar transform of a first-order Gauss-Markov process. An important point about this plot is the band of correlation due to the cross-covariance $\Lambda_{X_{M-1}U_{M-1}}$. We refer to this band of correlation as a "finger" of the transformed autocovariance. Note that the autocovariance of an M-level wavelet transform of a first-order Gauss-Markov process will create M-fingers in the transformed autocovariance. This fact is illustrated in Fig. 4. 5 (b) to Fig. 4. 5 (d) where we show the normalized autocovariances of the first-order Gauss-Markov process transformed by higher level Haar transforms (two-level to four-level), respectively.

In Fig. 4. 6 (a) to (d) we show a representative cross-section (chosen carefully so that it does not exhibit edge or windowing effects) of Fig. 4. 5 (a) to (d), respectively. Specifically, we consider the correlation structure of the 100th transformed coefficient when the transformed coefficients are stacked in a vector as shown in Equation (2. 46), i.e.

the correlation structure of the 36th element of U_{M-1}. Although these plots do not display the entire autocovariance, like the 2-D images, these plots do allow the reader to observe the actual values of a subset of the normalized autocovariance. Observe from these plots that the magnitude of the second finger of the 100th transformed coefficient is smaller than the magnitude of the first and third fingers; however, certain neighboring coefficients, e.g. the 99th transformed coefficient, do not have this property. This property is a result of the fact that the covariance between the wavelet coefficients at different scales are not stationary but have the structure shown in Equation (4. 12).

In Fig. 4. 7 (a) to (d) we show the variances of the one to four-level Haar transform of the samples of the first-order Gauss-Markov process. Since for Gaussian random variables information is proportional to the logarithm of the variance of the random variable [37], we observe from these plots that this process has predominantly coarse-scale information. Of course, this conclusion is obvious for the first-order Gauss-Markov process since it has a power spectrum concentrated at low-frequency Fig. 4. 8. We make this observation as a means of determining if a process is well-suited to an approximate Karhunen-Loeve expansion, i.e., approximately diagonal normalized autocovariance, using the wavelet transform. We will observe in future sections of this chapter that processes with predominantly coarse-scale information are well-suited to this type of diagonalization. This observation should not be surprising since the wavelet transform effectively "zooms" in on low-frequency signals by doing successive stages of low-pass filtering and decimation.

We now consider a three-level wavelet transform where the filters that are used are not necessarily the filters used in the Haar transform. We will make a comparison between QMFs that are designed by using the criterion from Chapter 3 and QMFs designed by Daubechies [2].

Recall that the criterion from Chapter 3 is motivated by our desire to obtain an

approximate Karhunen-Loeve expansion of the transformed coefficients, i.e., we desire the autocovariances $P_{X_{M-1}X_{M-1}}$ and $P_{U_{M-1}U_{M-1}}$ to decay as quickly as possible and the cross-covariance $P_{X_{M-1}U_{M-1}}$ to be as small as possible. We show here the normalized autocovariances of the wavelet transformed coefficients using QMFs that were designed using the criterion from Chapter 3 in order to visualize how well our optimization procedure does in achieving an approximate Karhunen-Loeve expansion. The normalized autocovariance of a three-level wavelet transform of samples of the first-order Gauss-Markov process are shown in Fig. 4. 9 (a) to (d) as we increase the support of the optimal QMFs from two-taps to eight-taps, respectively. A representative cross-section of Fig. 4. 9 is shown in Fig. 4. 10, i.e. the correlation structure of the 100th transformed coefficient. The variances of the transformed coefficients (i.e. the values along the diagonals of Fig. 4. 9) are plotted in Fig. 4. 11.

The QMFs derived by Daubechies have received considerable attention. Filters derived by Daubechies with 2P-taps possess the property that the continuous-time wavelet has P vanishing moments, i.e. such a wavelet $\psi(t)$ satisfies

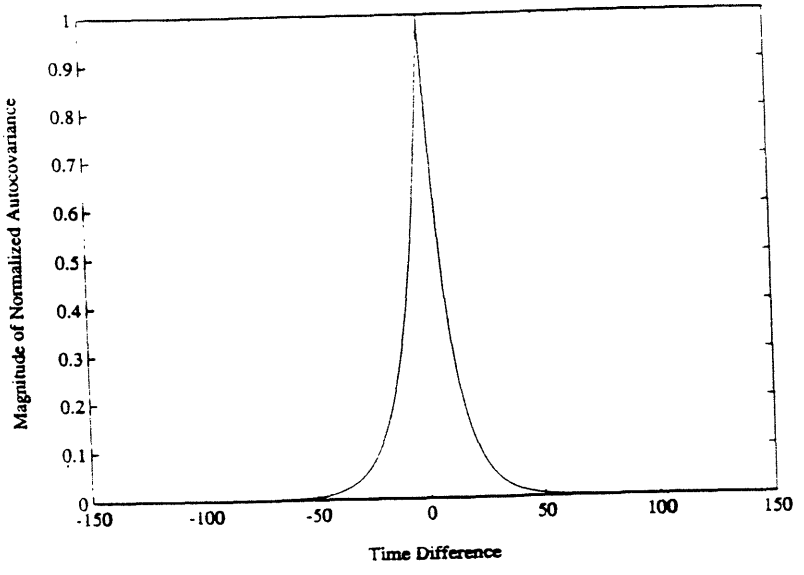
$$\int_{-\infty}^{\infty} t^p \psi(t)dt = 0 \qquad \forall p \in \{0,1,\ldots,P-1\} \qquad (4.35)$$

For the purposes of comparison, we consider a three-level wavelet transform using eight-tap QMFs. The normalized autocovariance of the transformed coefficients when the optimal eight-tap QMFs are used is shown in Fig. 4. 9 (d), while the normalized autocovariance of the transformed coefficients using eight-tap QMFs that were designed by Daubechies is shown in Fig. 4. 12. The variances for these coefficients are shown in Fig. 4. 11 (d) and Fig. 4. 12 (b). The correlation structure of the 100th transformed coefficient is shown when using both of these methods on a linear scale Fig. 4. 13 (a) and

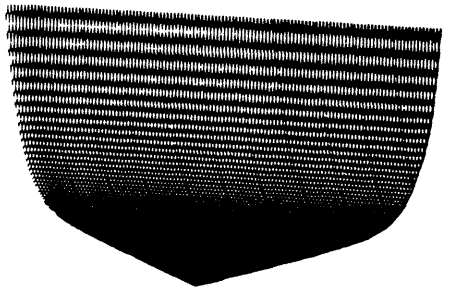
a semi-logarithmic scale Fig. 4. 13 (b). We observe from these plots that the two methods yield results that are very similar.

Summarizing our results, we have made the following observations. As we increase the number of levels of transformation that we perform on the Gauss-Markov process, we substantially increase the rate of decay of the transformed covariances. However, we also introduce a new (undesired) "finger" of the covariance of the transformed coefficients for each level of transformation that we perform on the sampled process. These fingers may be undesirable since they represent an additional band of correlation that may need to be dealt with. One method that can be used to lessen the correlation in these regions is to choose QMFs that were derived using the criterion of Chapter 3. Although we have concentrated on the first-order Gauss-Markov process, these results are valid for most processes that have autocovariances that decay monotonically.

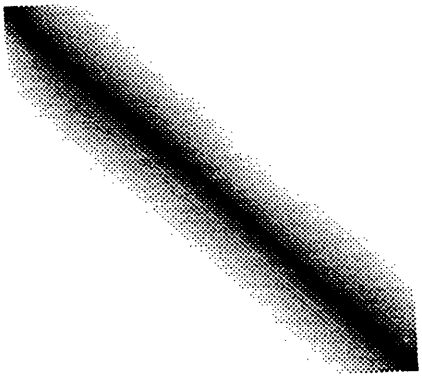
Further we have introduced the idea of examining the variances of the Haar transform of a process in order to determine if the process is well-suited to an approximate Karhunen-Loeve expansion using the wavelet transform. Note other wavelet transforms can be used other than Haar, but the Haar is computationally efficient and does not introduce undesired edge, windowing, or wrap-around effects.



(a)



(b)



(c)

Fig. 4. 1. Plots of the magnitude of the normalized autocovariance of the first-order Gauss-Markov process in steady state when $\alpha = 0.9006$. The normalized autocovariance is viewed as: (a) function of the time difference that is commonly used to view autocovariances of stationary processes, (b) 3-D mesh plot, (c) 2-D image. The "time difference" in (a) and in other plots corresponds to the relative time lag between the elements of the quantities whose covariance function is being displayed. The plots shown in (b) and (c) are useful for viewing autocovariances that are both stationary and nonstationary.

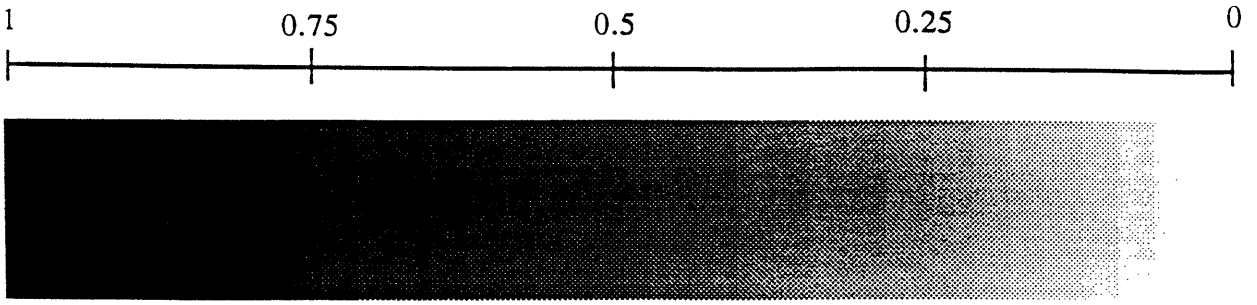


Fig. 4. 2. Mapping of the gray-scale to the correlation values that are used in all 2-D images (unless shown otherwise).

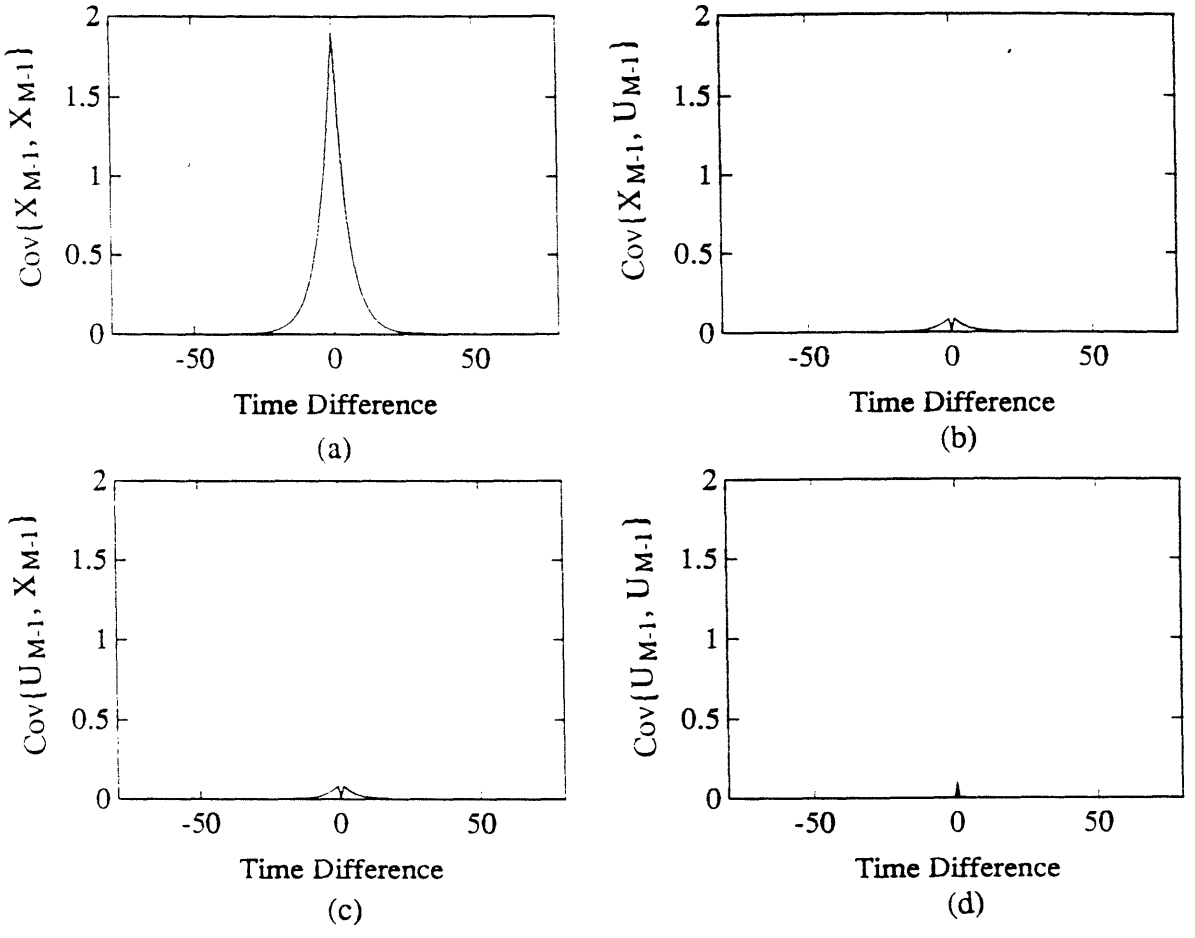
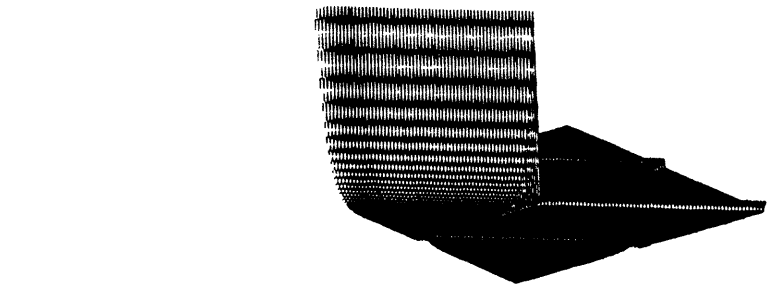
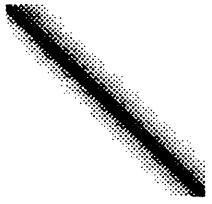


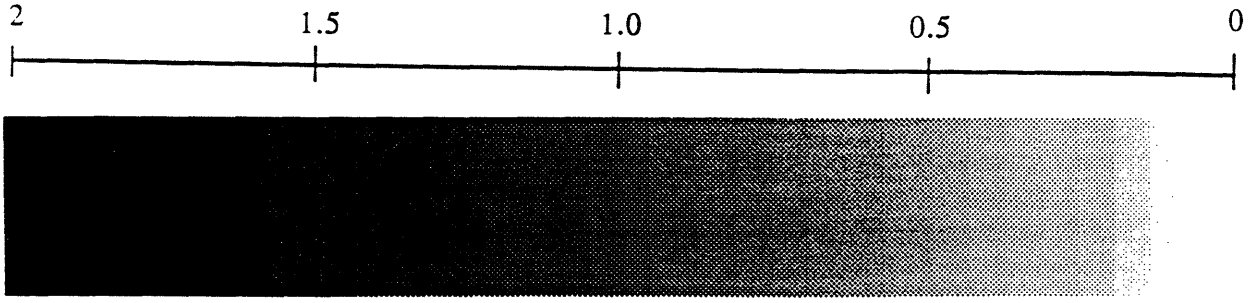
Fig. 4. 3. Plots of the four covariances that are used in the magnitude of the autocovariance of the one-level Haar transform of the first-order Gauss-Markov process. The covariances are viewed separately as: (a) $\mathrm{Cov}\{X_{M-1}, X_{M-1}\}$, (b) $\mathrm{Cov}\{X_{M-1}, U_{M-1}\}$, (c) $\mathrm{Cov}\{U_{M-1}, X_{M-1}\}$, (d) $\mathrm{Cov}\{U_{M-1}, U_{M-1}\}$. The "time difference" in these plots correspond to the relative time lag between the elements of the quantities whose covariance function is being displayed.



(a)



(b)



(c)

Fig. 4. 4. Plots of the magnitude of the autocovariance of the one-level Haar transform of the first-order Gauss-Markov process, viewed as: (a) 3-D mesh plot, (b) 2-D image. The additional band of correlation in the transformed autocovariance is due to the cross-correlation between different transformed sequences, i.e. X_{M-1} and U_{M-1}. We refer to this band of correlation as a "finger". Since the 2-D image in (b) is not normalized, the gray-scale mapping is different then other images. The gray-scale mapping for this image is shown in (c).

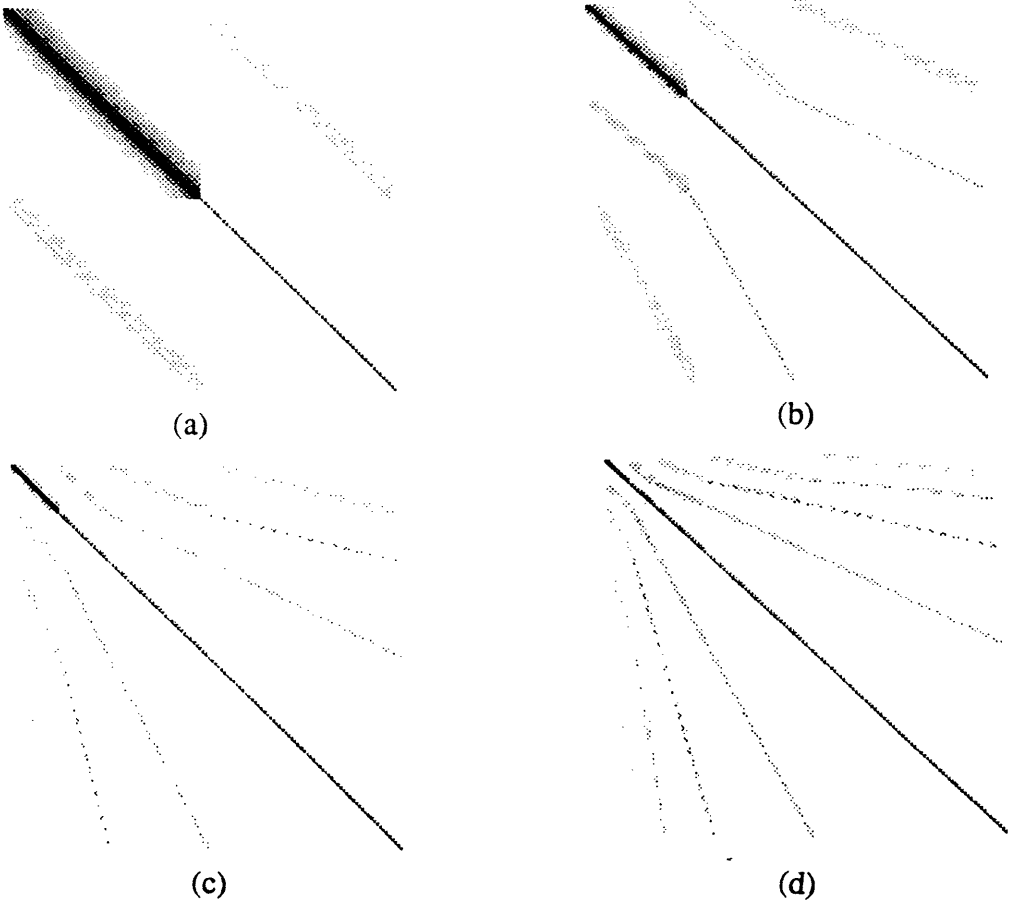


Fig. 4. 5. Plots of the magnitude of the normalized autocovariance of the L-level Haar transform of the first-order Gauss-Markov process, where: (a) L=1, (b) L=2, (c) L=3, (d) L=4. The plots show that an L-level transform will produce L fingers in the transformed autocovariance.

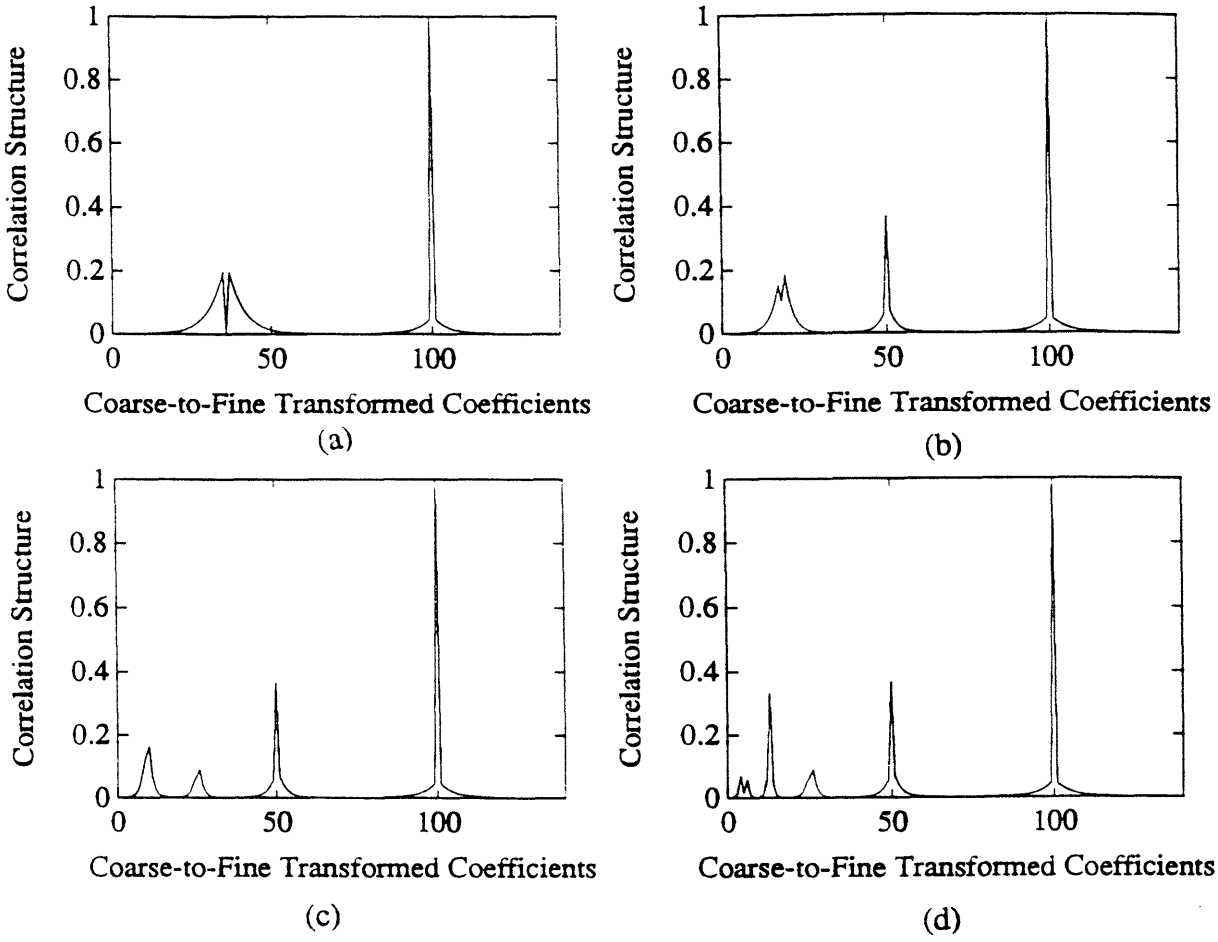


Fig. 4. 6. Plots of the correlation structure of the 100th transformed coefficient of the first-order Gauss-Markov process. The transform being used is an L-level Haar transform where: (a) L=1, (b) L=2, (c) L=3, (d) L=4. These plots show a representative cross-section of Fig. 4. 5 that does not exhibit edge or windowing effects.

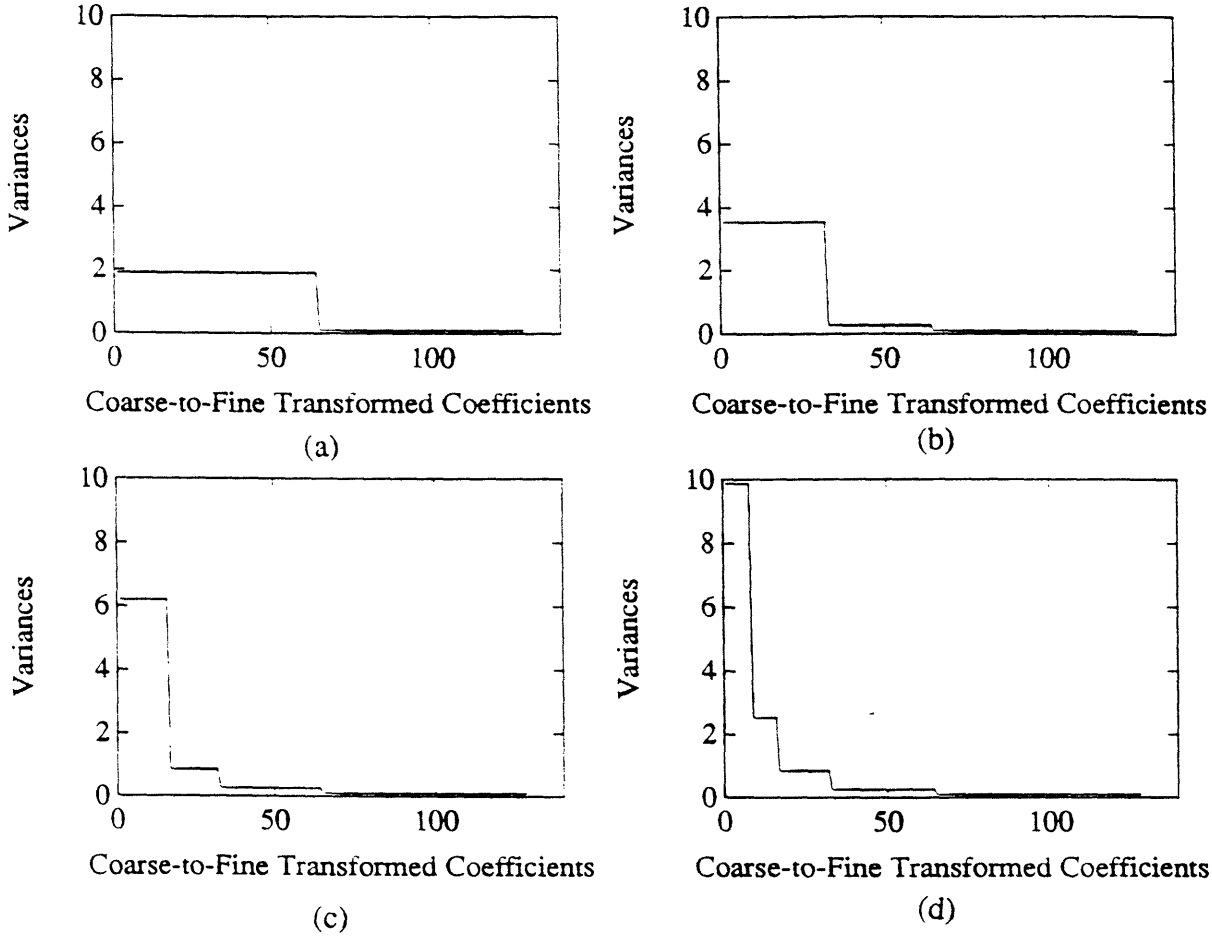


Fig. 4. 7. Plots of the variances of the coefficients obtained by taking an L-level Haar transform of the first-order Gauss-Markov process, where: (a) L=1, (b) L=2, (c) L=3, (d) L=4. The plots show that for the first-order Gauss-Markov process the variances of the transformed coefficients are larger at coarser levels.

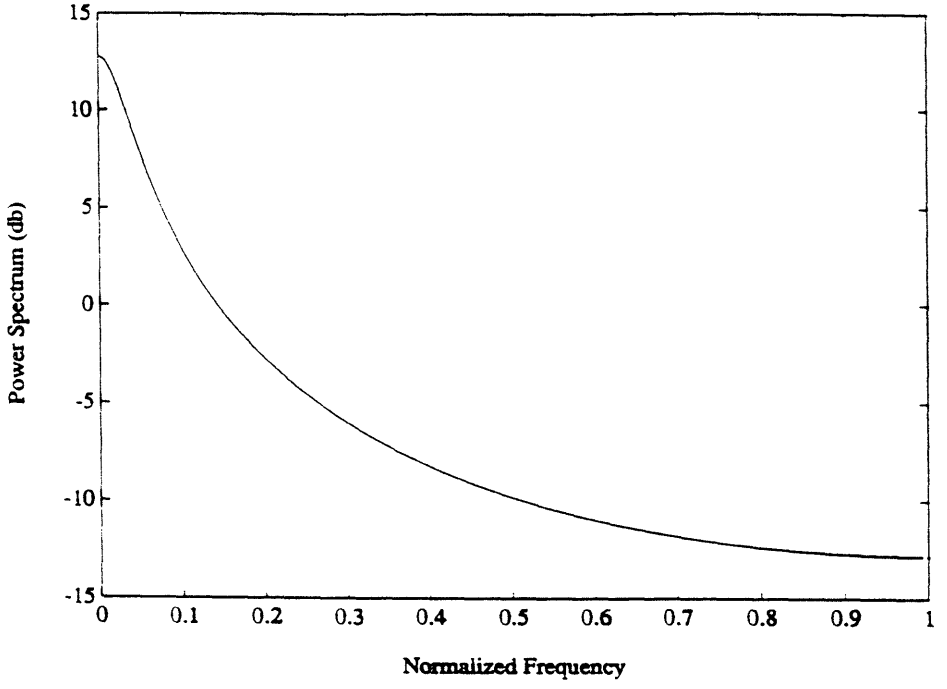


Fig. 4. 8. Plot of the power spectrum of the first-order Gauss-Markov process. This plot shows that the first-order Gauss-Markov process has most of the process's energy concentrated at low-frequency.

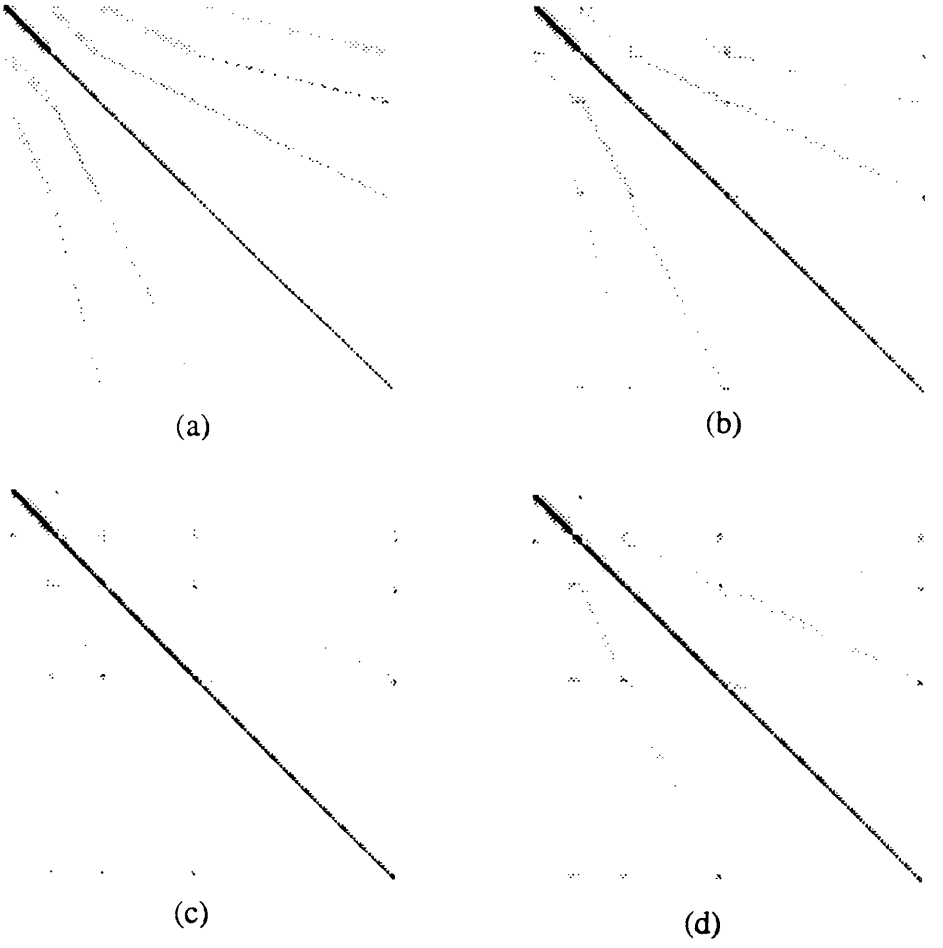


Fig. 4. 9. Plots of the normalized autocovariance of the three-level wavelet transform of the first-order Gauss-Markov process using QMFs that were derived by using the optimality criterion from Chapter 3. The QMFs have support equaling W, where (a) W=2, (b) W=4, (c) W=6, (d) W=8. The plots indicate, as expected, that as the support of the QMFs increase the transformed autocovariance more closely resembles a Karhunen-Loeve expansion.

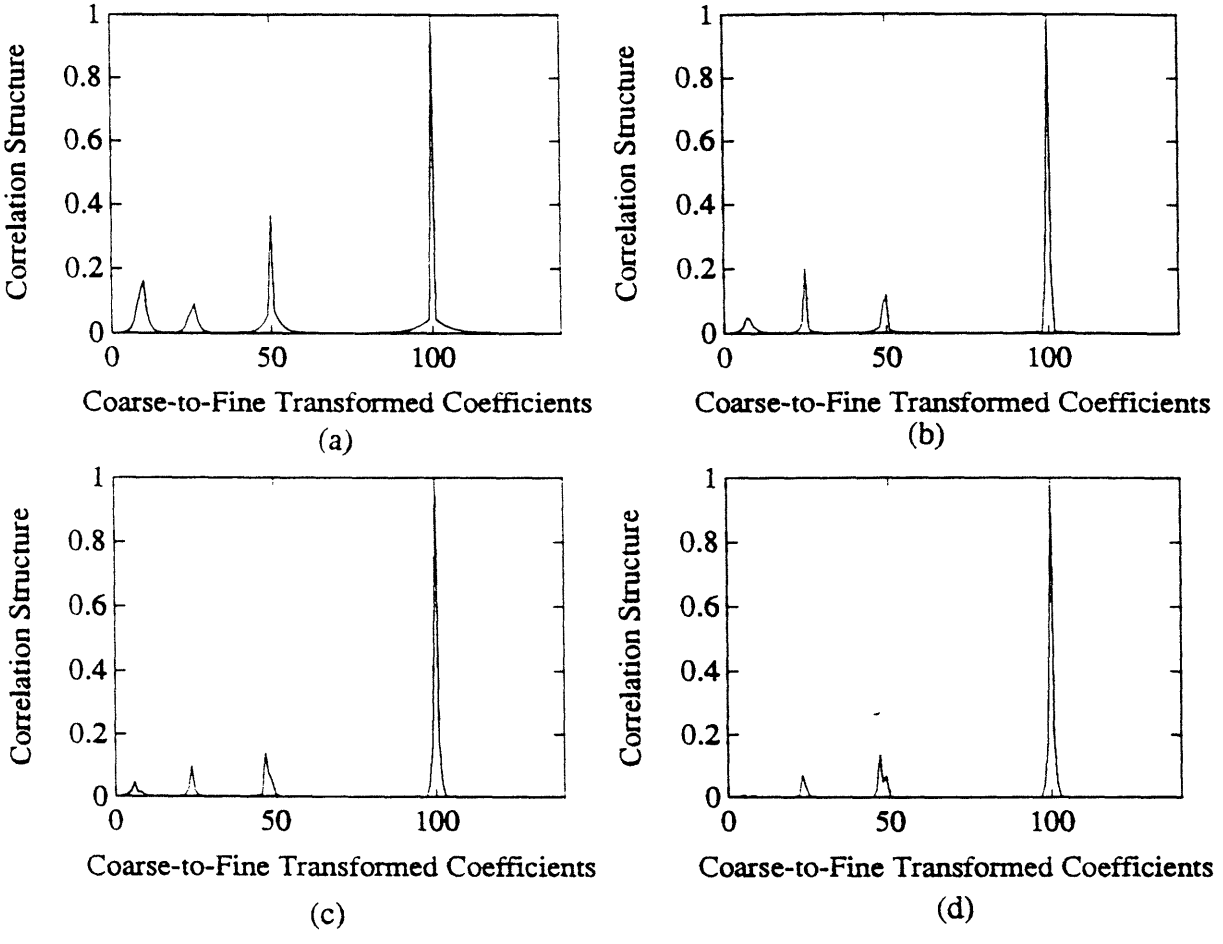


Fig. 4. 10. Plots of the correlation structure of the 100th transformed coefficient of the first-order Gauss-Markov process. The transform being used is a 3-level wavelet transform using optimal QMFS with support W, where: (a) W=2, (b) W=4, (c) W=6, (d) W=8.

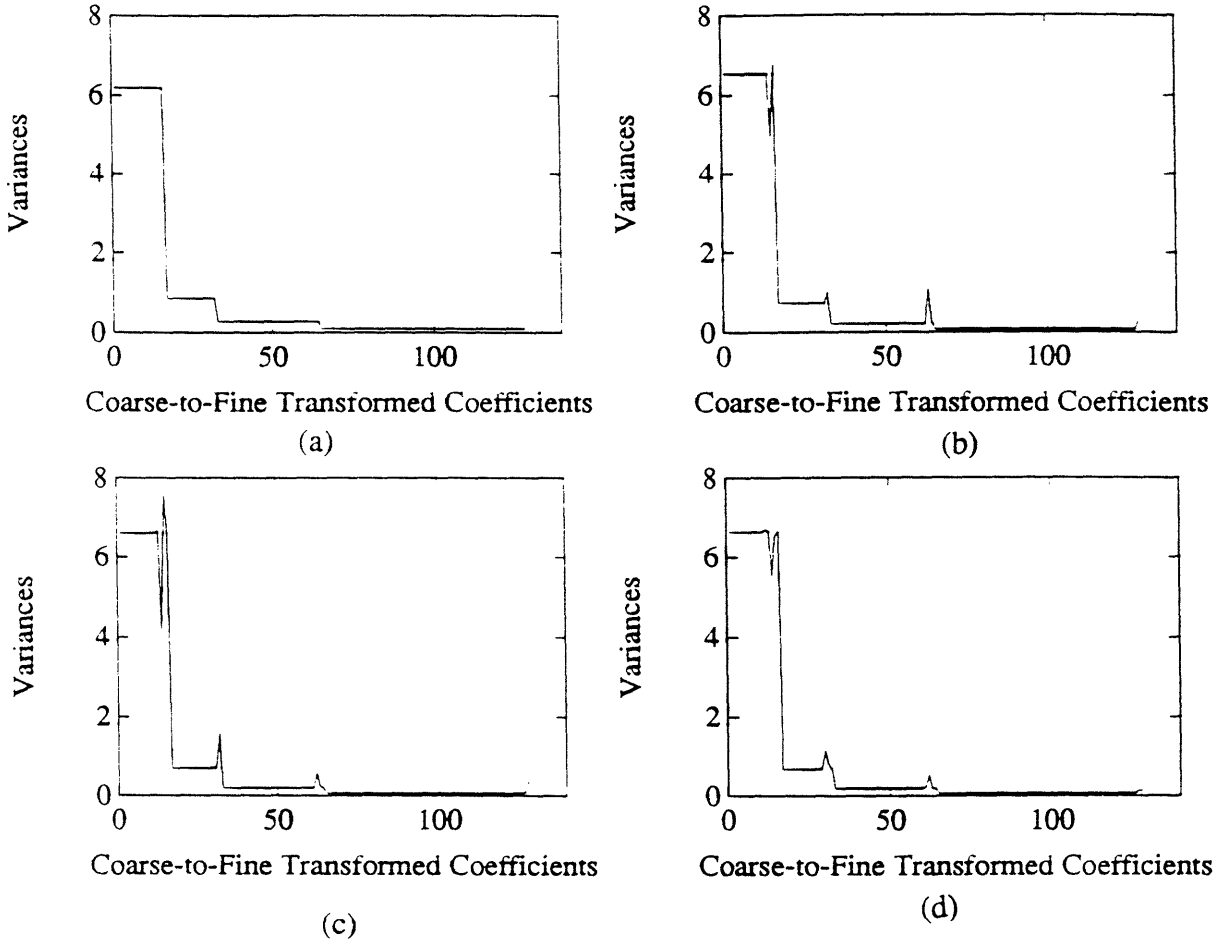
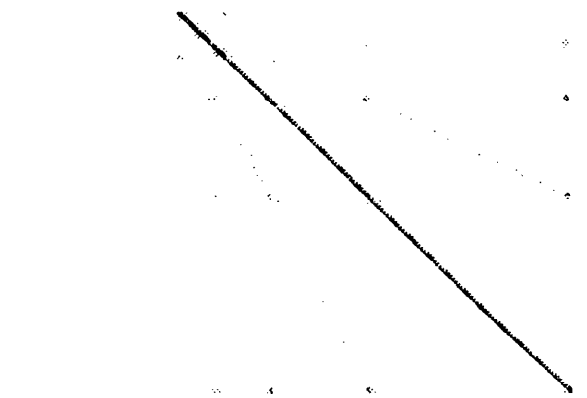
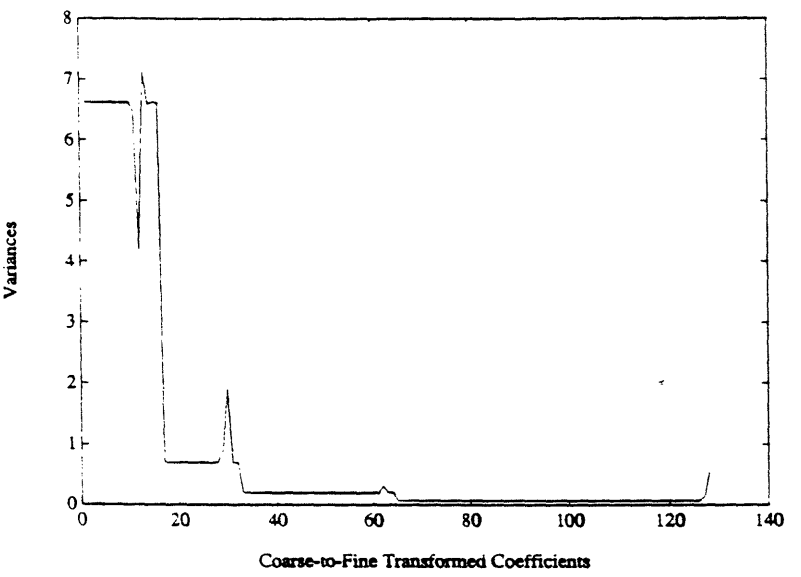


Fig. 4. 11. Plots of the variances of the coefficients obtained by taking a three-level wavelet transform of the first-order Gauss-Markov process. The QMFs have support equaling W, where (a) W=2, (b) W=4, (c) W=6, (d) W=8. The edge effects are due to using circular convolution.

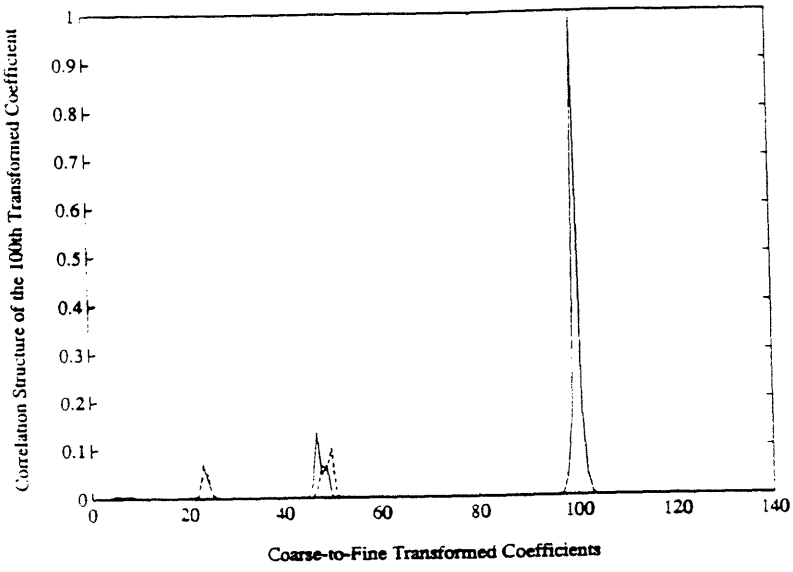


(a)

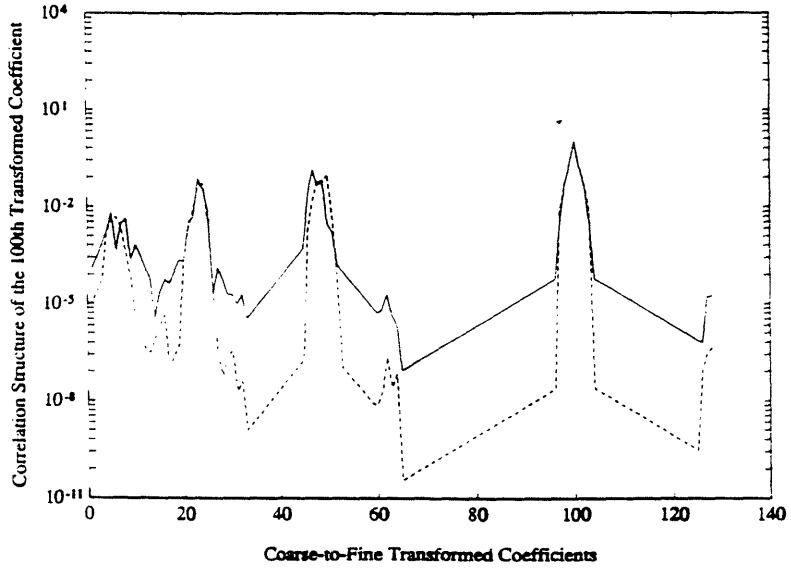


(b)

Fig. 4. 12. Plot (a) shows the normalized autocovariance of the three-level wavelet transform of the first-order Gauss-Markov process using eight-tap QMFs that were derived by Daubechies. Plot (b) shows the variances of these transformed coefficients.



(a)



(b)

Fig. 4. 13. A representative cross-section (that does not exhibit edge effects) of the normalized autocovariance of the three-level wavelet transform of the first-order Gauss-Markov process using eight-tap QMFs. The plots show the correlation structure of the 100th wavelet transformed coefficient using QMFs derived by the optimality criterion from Chapter 3 (solid line) and QMFs derived by Daubechies (dashed line) where the correlation (y-axis) is plotted on: (a) linear scale, (b) logarithmic scale.

4.3 UNDER-DAMPED SECOND-ORDER PROCESSES

In the preceding section we have discussed the first-order Gauss-Markov process. The first-order Gauss-Markov process is an example of a process that has a power spectrum concentrated at low frequency. We now consider a class of processes that do not have this property. Specifically, we consider the following second-order oscillatory model in steady-state:

$$x[n] = 2\alpha \cos\theta\, x[n-1] - \alpha^2 x[n-2] + w[n] \tag{4. 36}$$

or in the z-transform domain:

$$\frac{X(z)}{W(z)} = \frac{1}{(1-\alpha e^{j\theta}z^{-1})(1-\alpha e^{-j\theta}z^{-1})} \tag{4. 37}$$

where w[n] is iid white noise, and X(z) and W(z) are the z-transforms of x[n] and w[n], respectively.

Again we collect a set of samples of this process into a vector and denote the vector as X_M. The power spectrum of the process is trivially obtained by using Equation (4. 37) and is plotted in Fig. 4. 14 when $\alpha = .9006$ and $\theta = \pi/4$. The normalized autocovariance is easily obtained by inverse Fourier transforming the power spectrum and is shown in Fig. 4. 15. In Fig. 4. 15 (a) we show the normalized autocovariance as a plot that is commonly used for visualizing autocovariances of stationary processes. The magnitude of this normalized autocovariance is plotted as a 2-D image in Fig. 4. 15 (b). In addition, we also plot only the positive values of the normalized autocovariance in Fig. 4. 15 (c) so that we can clearly illustrate the oscillations of the process. Note that neither Fig. 4. 15 (b) nor

Fig. 4. 15 (c) viewed by itself illustrates the autocovariance of the process, but by viewing both plots the reader has an accurate description of the autocovariance.

The period of oscillation of the autocovariance for a process in the form of Equation (4. 36) is $2\pi/\theta$. The period of oscillation when $\theta = \pi/4$ is simply 2^3. Therefore the cross-correlation between any coefficient above the M-3 level, i.e. an element of $\{U_{M-1}, U_{M-2}\}$, and any other transformed coefficient will show some form of oscillation. This fact is illustrated by observing the 2-D images of the transformed autocovariances.

In Fig. 4. 16 (a) to (d), we plot the magnitude of the normalized autocovariances of a one to four-level Haar Transform of this process. Again in order to display the oscillations of the process, we only plot the positive values of the normalized autocovariances in Fig. 4. 17 (a) to (d). Note the 2-D images of the positive values of the normalized autocovariances depict the oscillations as "patterns" in the plot of the transformed autocovariance. In Fig. 4. 17 (c) and (d) the covariances of the transformed coefficients at or below the M-3 level do not have any of these "patterns", i.e., there exists only regions that have a solid shading that is either completely dark or completely white. Since some of the transformed coefficients have covariances that oscillate, we conclude that the sampled process has not been transformed into a sufficiently simple form, i.e. approximate Karhunen-Loeve expansion.

In Fig. 4. 18, we again plot the correlation structure of the 100th transformed coefficient; however, these plots are not strictly cross-sections of Fig. 4. 16 or Fig. 4. 17. The plots in Fig. 4. 18 show both positive and negative correlation values. Of course, cross-sections of Fig. 4. 16 are obtained by considering the magnitude of the correlation of the plots in Fig. 4. 18. The cross-sections of Fig. 4. 17 are obtained by displaying only the nonnegative correlation values of the plots in Fig. 4. 18, i.e. by clipping negative values.

The variances of the transformed coefficients are shown in Fig. 4. 19 (a) to (d). Note that the transformed coefficients with the largest variances are the coefficients of

U_{M-3} which were obtained when using a decimation factor that is equal to the period of oscillation. Of course, if the magnitude of α is small then Equation (4. 36) can be approximated as a first-order Gauss-Markov process and we would observe that a plot of the variances of the transformed coefficients would be similar to the plots of the variances in the preceding section. If θ was chosen such that the period of oscillation was not a power of two, e.g. six, then we would observe that the transformed coefficients with the largest variances would be the wavelet coefficients that were obtained when using decimation factors that were closest to the period of oscillation. For example, if the period of oscillation was six then we would expect that the variances of U_{M-2} and U_{M-3} to be large compared with the variances of the other transformed coefficients.

Recently, Coifman *et al.* [11] have investigated the use of a modified version of the wavelet transform that is called the wave packet transform. As pointed out in [11], the wave packet transform is especially well-suited to kernels that have oscillatory behavior. The wavelet transform effectively "zooms" into the low frequency component of the signal; however, the wave packet transform "zooms" into desired frequencies. Coifman and Wickerhauser [12] describe several techniques for choosing the wave packet transformation that "zooms" into specified frequencies. We will refer to the complete wave packet transform as a transformation that "zooms" into all frequencies and this transform is described in the following manner.

Recall that we expressed the wavelet transformation matrix T_M as a cascade of M orthogonal matrices:

$$T_M = \cdots \begin{bmatrix} A_{M-2} & \vdots & 0 \\ \cdots & \vdots & \cdots \\ 0 & \vdots & I \end{bmatrix} [A_{M-1}] \qquad (4.38)$$

The complete wave packet transformation matrix C_M is specified by the cascade of M orthogonal matrices where the jth orthogonal matrix from the right is constructed by repeating A_{M-j} along the diagonal. Specifically, the transformation matrix C_M is given by:

$$C_M = \cdots \begin{bmatrix} A_{M-2} & \vdots & 0 \\ \cdots & \vdots & \cdots \\ 0 & \vdots & A_{M-2} \end{bmatrix} [A_{M-1}] \tag{4. 39}$$

and X_M transformed by a M-level complete wave packet transform is denoted:

$$Z_M = C_M X_M \tag{4. 40}$$

Of course, the autocovariance of the coefficients of the complete wave packet transform of X_M is:

$$\Lambda_{Z_M Z_M} = C_M \Lambda_{X_M X_M} C_M^T \tag{4. 41}$$

To illustrate the transformation more clearly, Fig. 4. 20 (a) shows the block-diagram of the two-level wavelet transform, while Fig. 4. 20 (b) shows the block-diagram of the two-level complete wave packet transform.

We now consider performing a complete wave packet transform of X_M, where X_M consists of the samples of the under-damped second-order oscillatory process. In Fig. 4. 21 (a) to (d) we show the magnitude of the normalized autocovariances of the complete wave packet transformed coefficients of X_M when $\alpha = .9006$ and $\theta = \pi/4$. Again these illustrations depict the normalized autocovariances when Haar QMFs are used and the number of levels of transformation performed on the process range from one to four-levels.

These plots illustrate that if a process has a decaying autocovariance then an M-level complete wave packet transform of the samples of that process will have $2^M - 1$ "fingers".

From the plots of the positive values of the normalized autocovariance of these same transformed coefficients Fig. 4. 22 (a) to (d), we observe that the autocovariances obtained from a three-level (or higher level) complete wave packet transform of X_M do not have any covariances that oscillate. The correlation structure of the 100th transformed coefficient is shown in Fig. 4. 23. The variances of the transformed coefficients (obtained by the complete one to four-level wave packet transform of X_M when Haar filters are used) are shown in Fig. 4. 24 (a) to (d).

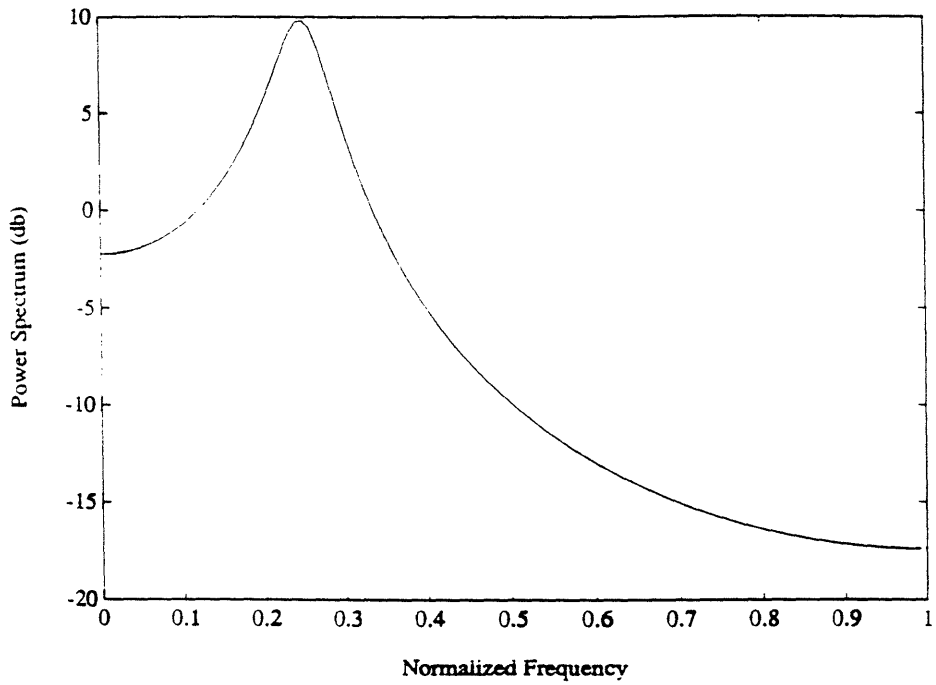
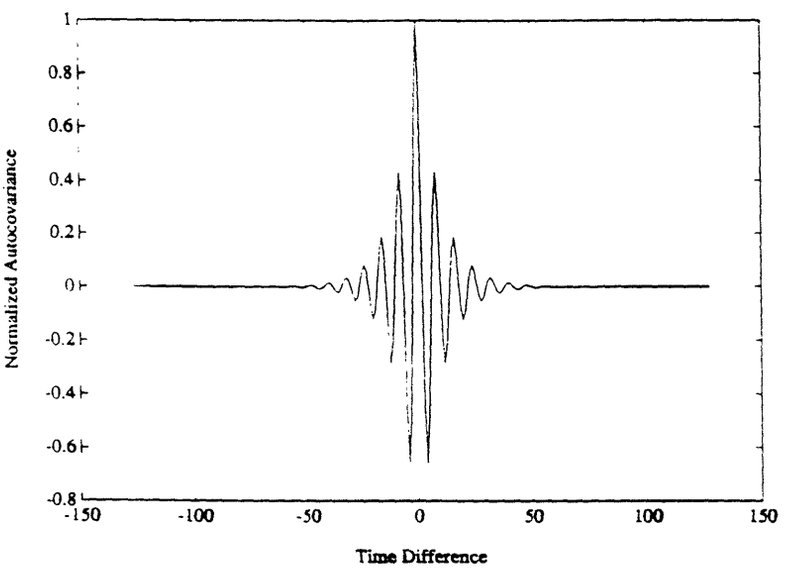
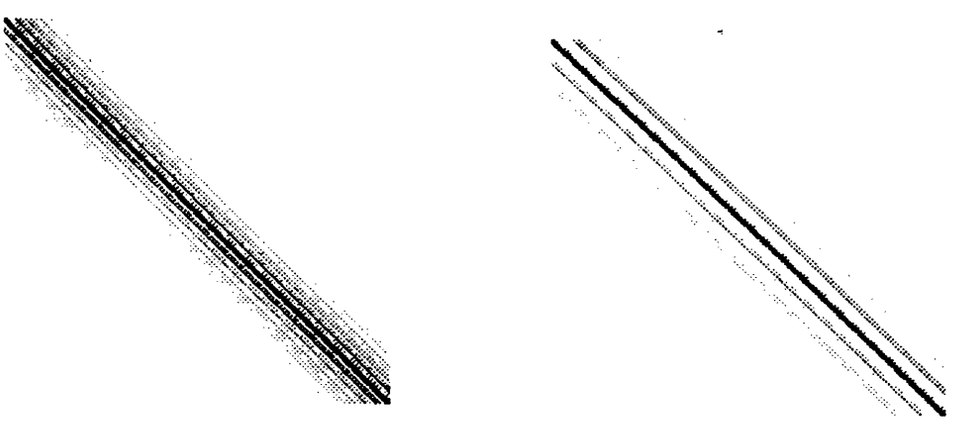


Fig. 4. 14. Plot of the power spectrum of a second-order under-damped process with $\alpha = 0.9006$ and $\theta = \pi/4$. This process does not have most of the process' energy concentrated at low-frequency.



(a)



(b) (c)

Fig. 4. 15. The normalized autocovariance of a second-order under-damped process with $\alpha = 0.9006$ and $\theta = \pi/4$ is shown by: (a) plotting the autocovariance as a function of the time difference, which is commonly used for showing the autocovariance of stationary processes (b) plotting the magnitude of the autocovariance as a 2-D image, which is useful for observing how the autocovariance decays, (c) plotting only the positive values of the autocovariance as a 2-D image, which is useful for observing the oscillations of the process.

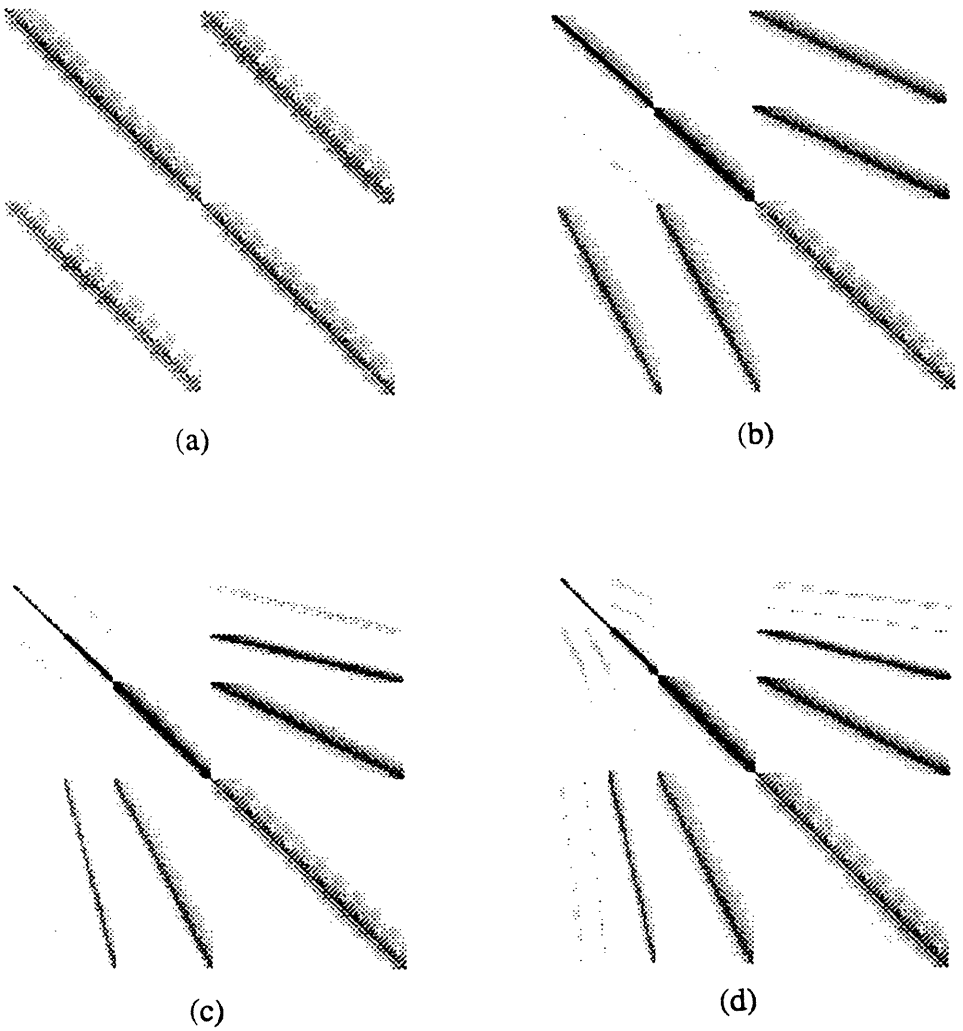


Fig. 4. 16. Plots of the magnitude of the normalized autocovariance of the L-level Haar transform of a second-order under-damped process, where (a) L=1, (b) L=2, (c) L=3, (d) L=4. The autocovariances have the same "finger" structure as the first-order Gauss-Markov process.

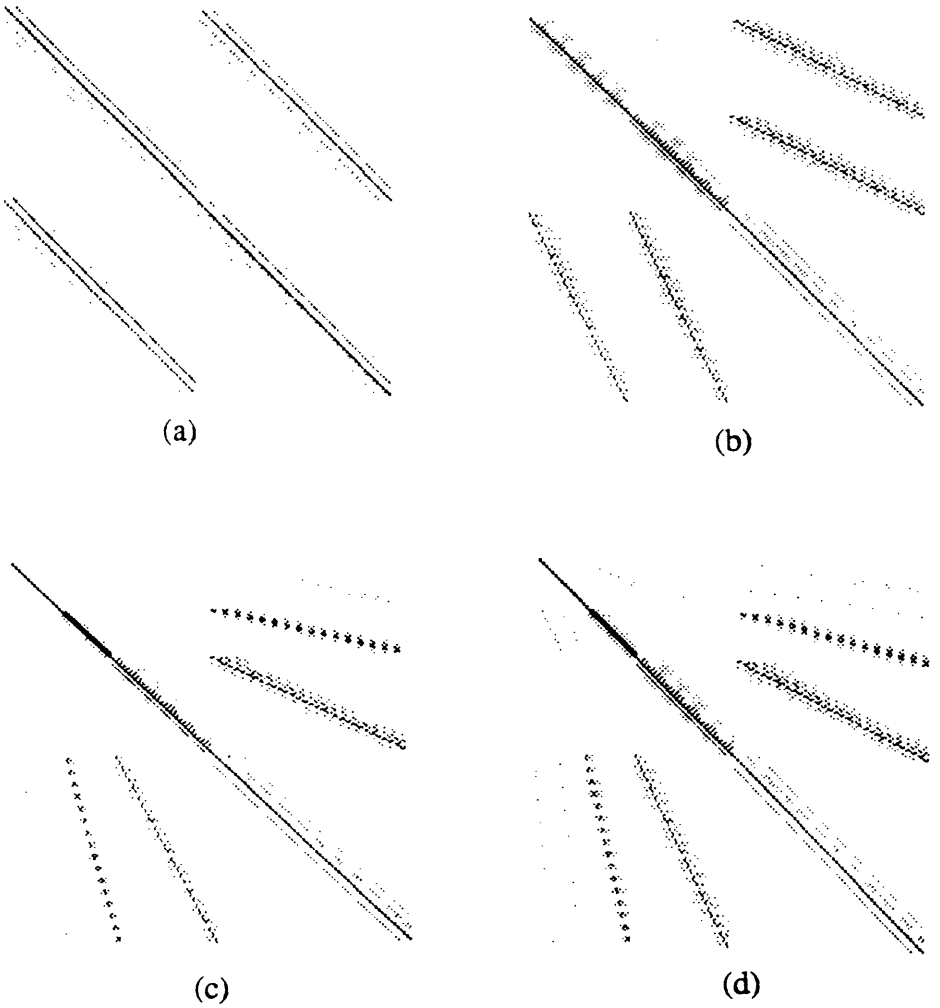


Fig. 4. 17. Plots of the positive values of the normalized autocovariance of the L-level Haar transform of a second-order under-damped process, where (a) L=1, (b) L=2, (c) L=3, (d) L=4. Covariances that oscillate are depicted as "patterns" in the transformed autocovariances.

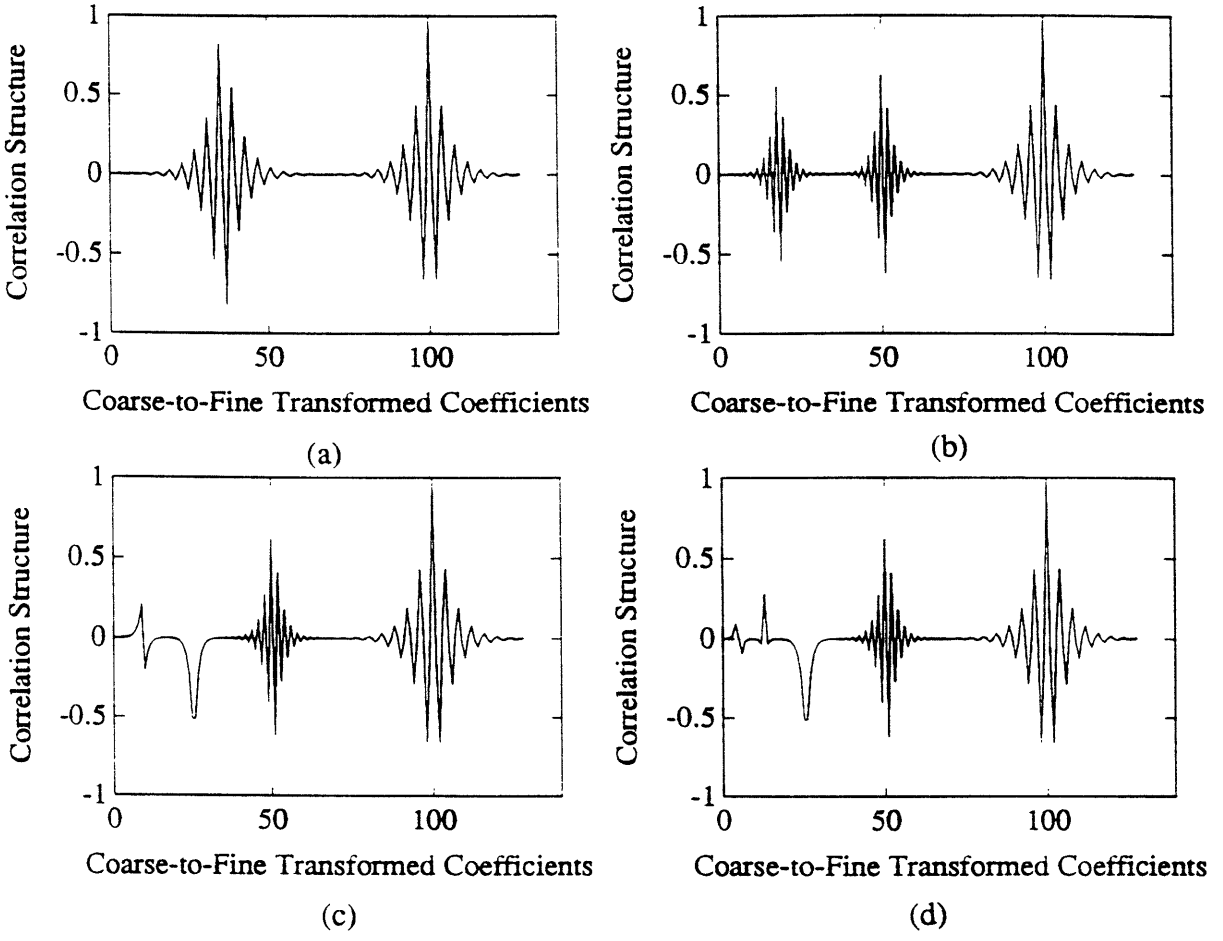


Fig. 4. 18. Plots of the correlation structure of the 100th transformed coefficient of a second-order under-damped process. The transform being used is an L-level Haar transform where: (a) L=1, (b) L=2, (c) L=3, (d) L=4.

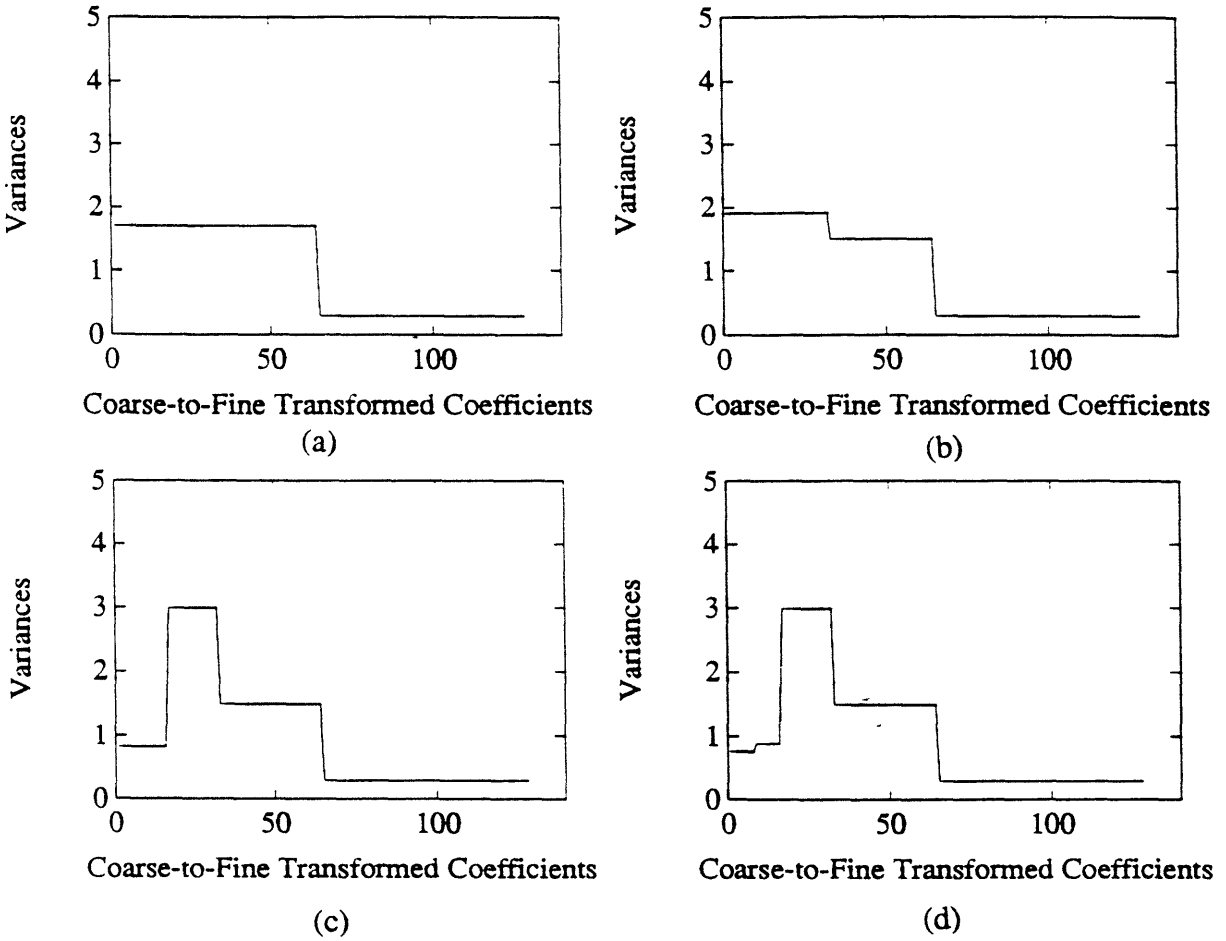
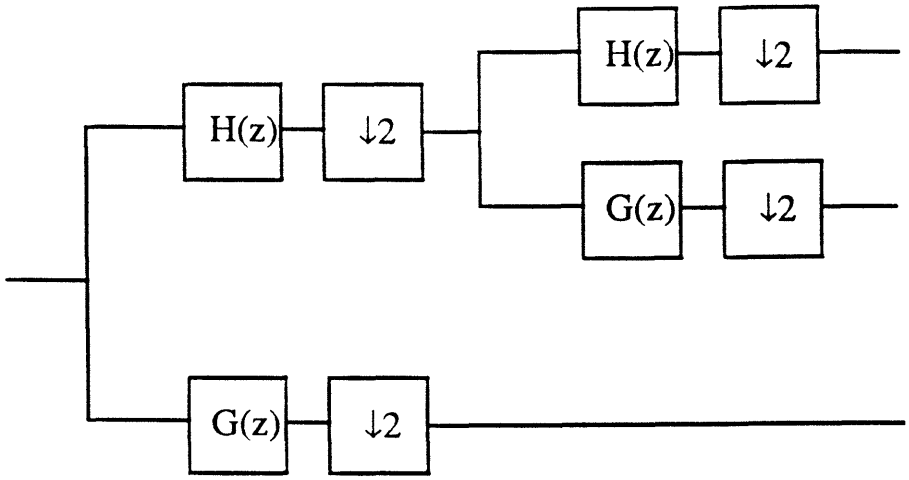
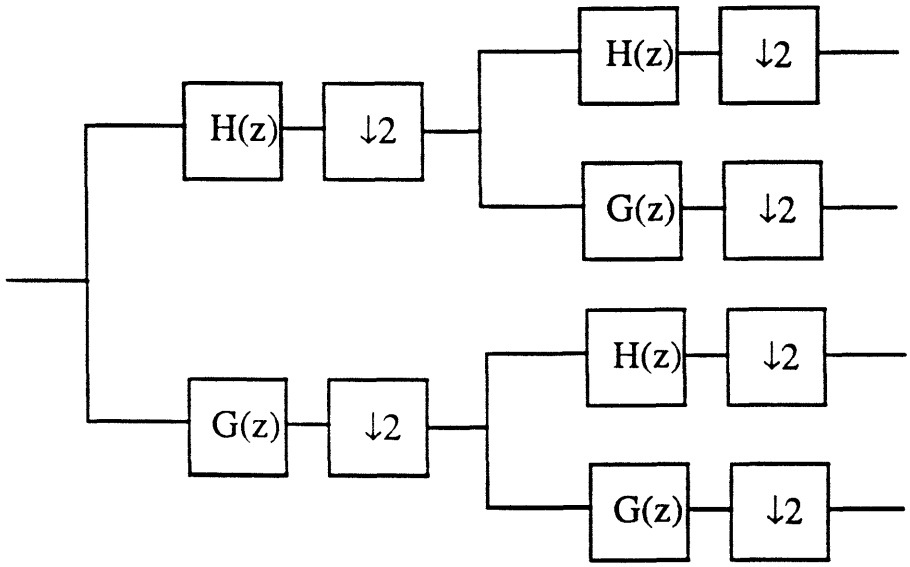


Fig. 4. 19. Plots of the variances of the transformed coefficients that are obtained by taking the L-level Haar transform of a second-order under-damped process, where (a) L=1, (b) L=2, (c) L=3, (d) L=4.



(a)



(b)

Fig. 4. 20. Block diagrams of the two-level (a) wavelet transform and (b) complete wave packet transform.

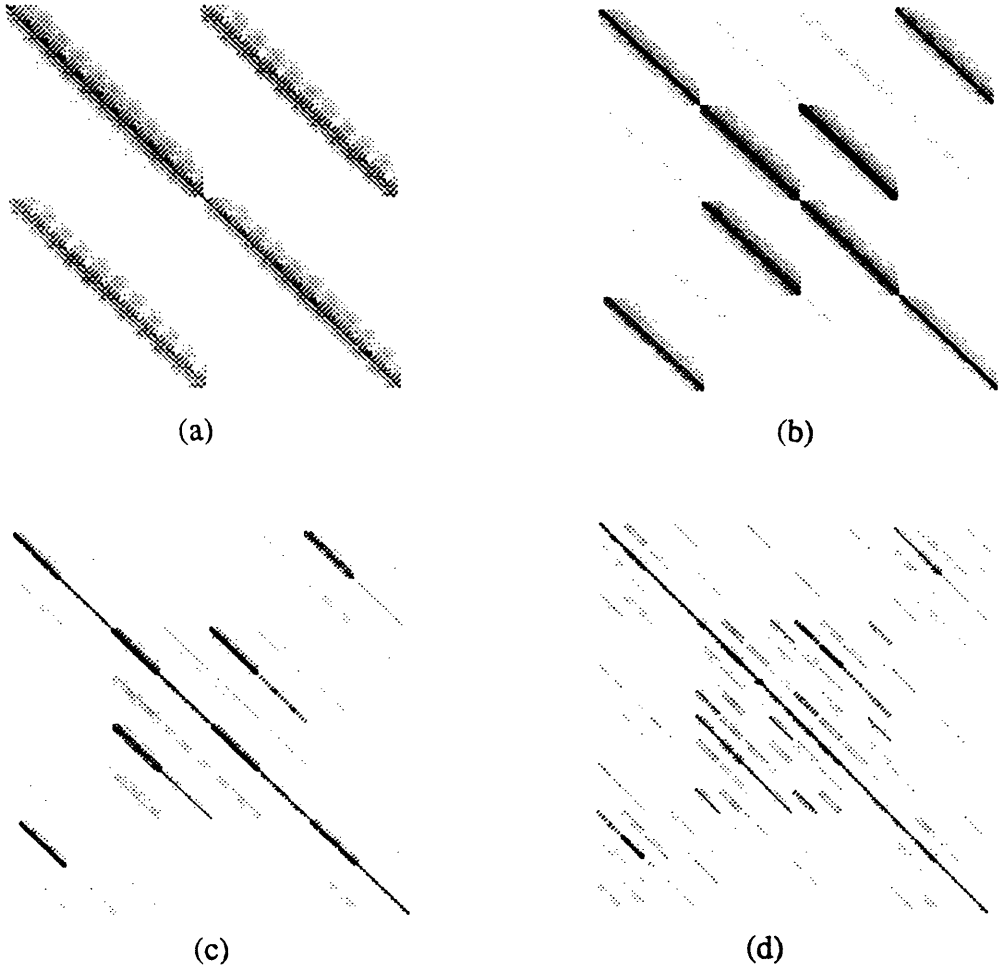


Fig. 4. 21. Plots of the magnitude of the normalized autocovariance of the L-level complete wave packet transform, using Haar filters, of a second-order under-damped process, where (a) L=1, (b) L=2, (c) L=3, (d) L=4. These autocovariances have 2^L-1 "fingers."

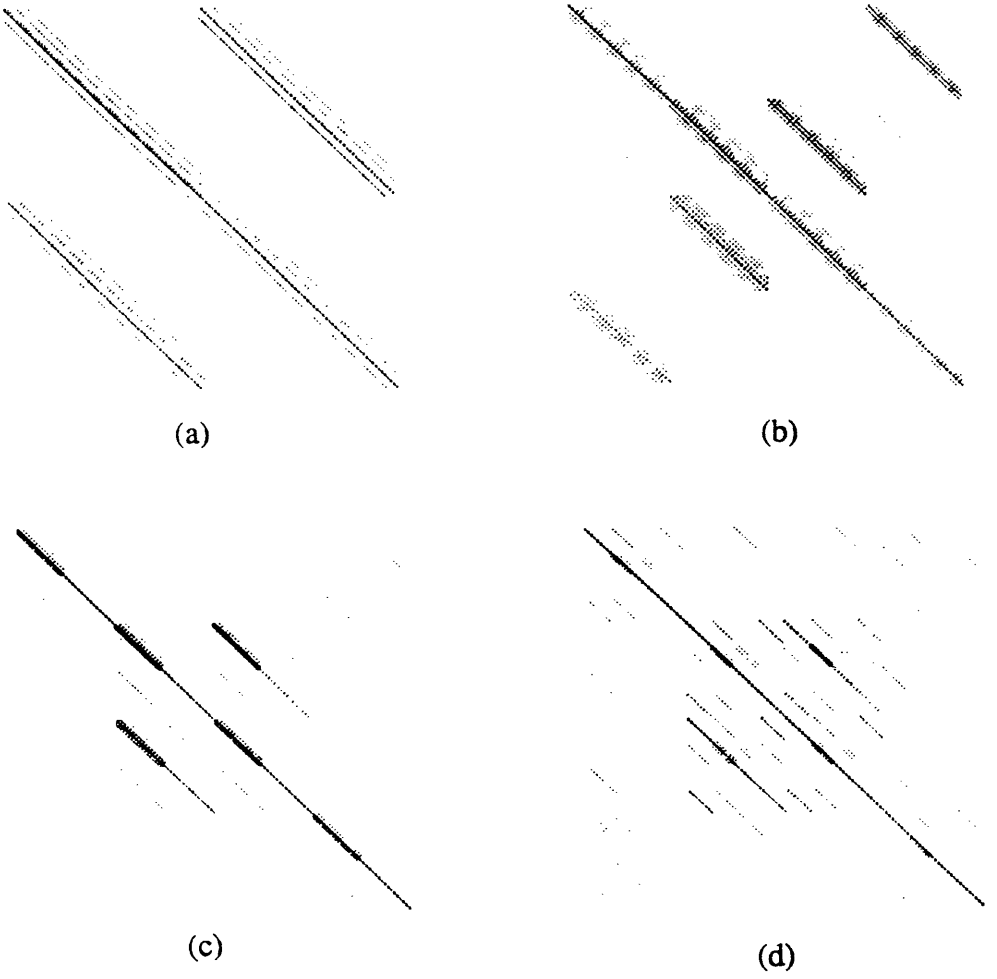


Fig. 4. 22. Plots of the positive values of the normalized autocovariance of the L-level complete wave packet transform, using Haar filters, of a second-order under-damped process, where (a) L=1, (b) L=2, (c) L=3, (d) L=4. The autocovariances shown in (c) and (d) do not have any covariances that oscillate.

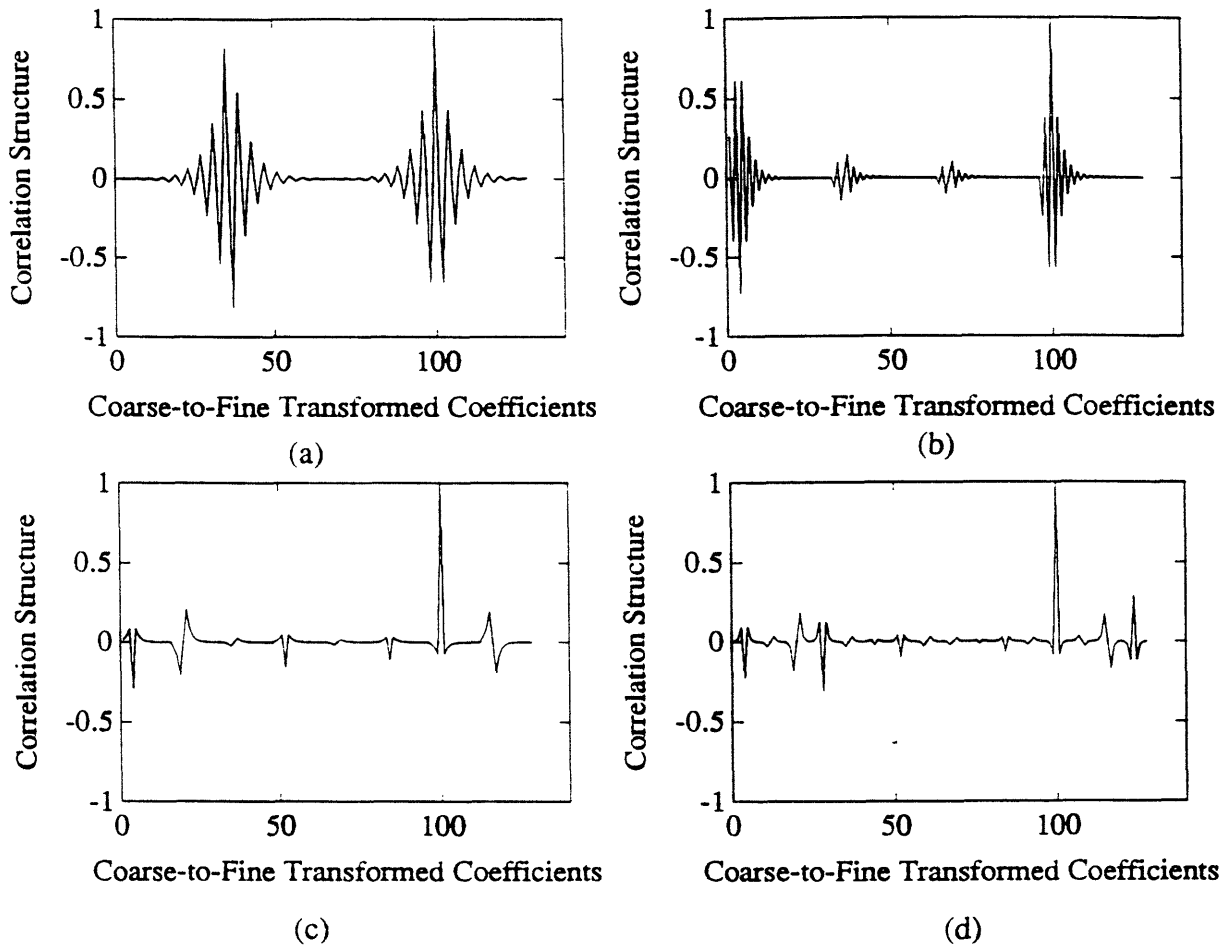


Fig. 4. 23. Plots of the correlation structure of the 100th transformed coefficient of a second-order under-damped process. The transform being used is an of the L-level complete wave packet transform where: (a) L=1, (b) L=2, (c) L=3, (d) L=4. The plots in (c) and (d) show that the correlation does not oscillate.

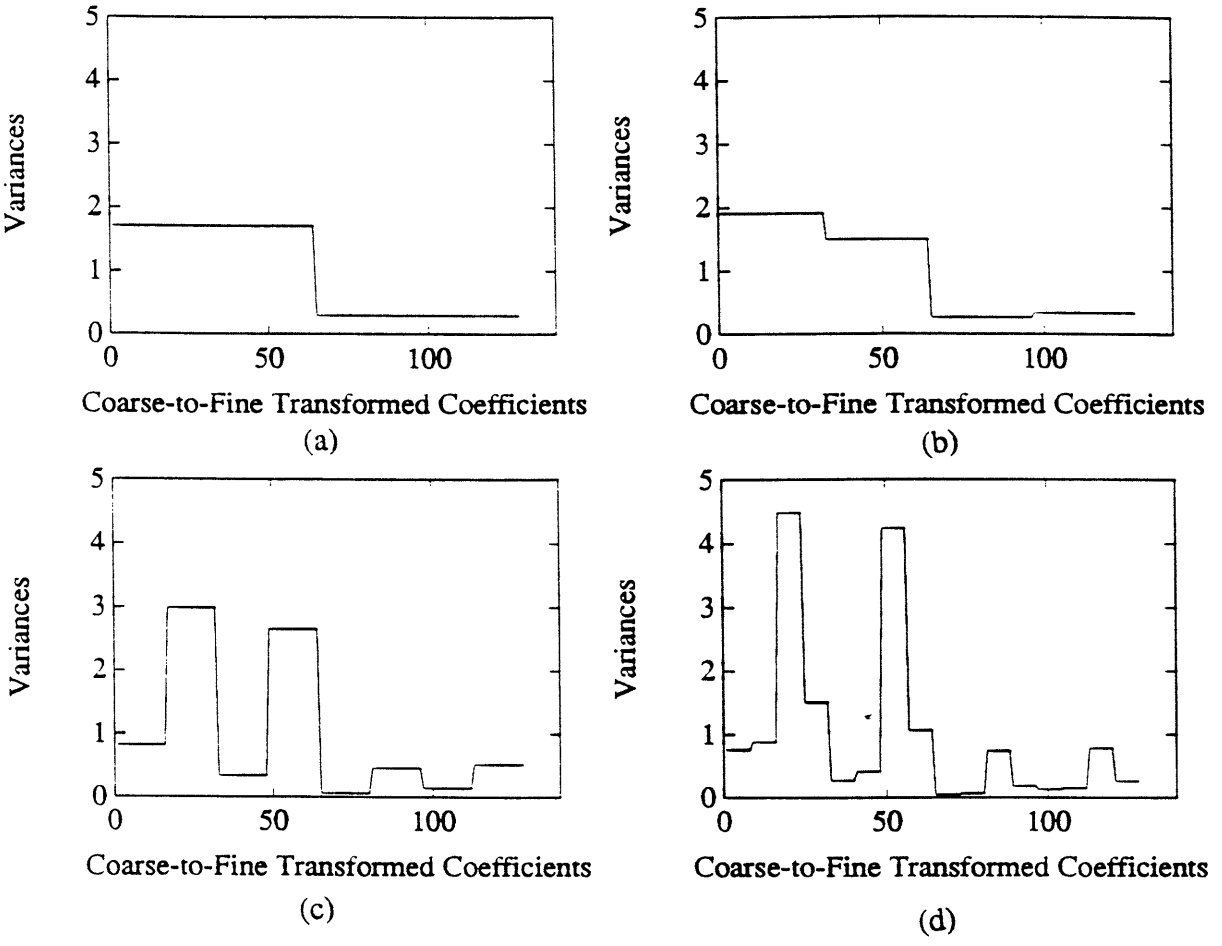


Fig. 4. 24. Plots of the variances of the transformed coefficients that are obtained by taking the L-level complete wave packet transform, using Haar filters, of a second-order under-damped process, where (a) L=1, (b) L=2, (c) L=3, (d) L=4.

4.4 FRACTIONAL BROWNIAN MOTIONS

Recently there has been a great deal of attention [17-20] focused on analyzing the class of nonstationary processes known as fractional Brownian motions (fBm's) by using the wavelet transform. In this section we perform optimal wavelet transforms on samples of a fractional Brownian motion process (again optimality is defined by the criterion of Chapter 3). We compare optimal wavelet transforms to wavelet transforms that use QMFs that were derived by Daubechies for a sampled Brownian motion process. Further by considering the class of fractional Brownian motions we investigate, in general, the ability of the wavelet transform to achieve an approximate diagonalization of the normalized autocovariance of the transformed coefficients.

The fBm, denoted by $B_H(t)$, is a nonstationary zero-mean Gaussian random function. It is defined as [15]:

$$B_H(t) = \frac{1}{\Gamma(H+0.5)}\left\{\int_{-\infty}^{0}(t-s)^{H-0.5} - (-s)^{H-0.5}\,dB(s) + \int_{0}^{t}(t-s)^{H-0.5}\,dB(s)\right\} \tag{4.42}$$

where $0 < H < 1$, $B_H(0) = 0$, $B(t)$ is ordinary Brownian motion, and the gamma function is defined by:

$$\Gamma(n) = \int_{0}^{\infty} t^{n-1}e^{-t}dt \tag{4.43}$$

which yields the recursion $\Gamma(n+1) = n\,\Gamma(n)$. The autocovariance of $B_H(t)$ is given by:

$$\Lambda_{B_H}(s,t) = \frac{V_H}{2}\left(|s|^{2H} + |t|^{2H} - |t-s|^{2H}\right) \qquad s, t \in \mathbb{R} \qquad (4.44)$$

where

$$V_H = \begin{cases} \dfrac{-\Gamma(2-2H)\cos(\pi H)}{(2H-1)\pi H} & 0 < H < .5 \\ & .5 < H < 1 \\ 1 & H=0.5 \end{cases} \qquad (4.45)$$

We let X_M be the column vector formed from N uniformly spaced samples of a segment of $B_H(t)$ where (without loss of generality) the sampling interval is normalized to unity, i.e.:

$$X_M = \begin{bmatrix} B_H(1) \\ B_H(2) \\ B_H(3) \\ \vdots \\ B_H(N) \end{bmatrix} \qquad (4.46)$$

Of course, the autocovariance of X_M is:

$$\Lambda_{X_M X_M}(s,t) = \frac{V_H}{2}\left(|s|^{2H} + |t|^{2H} - |t-s|^{2H}\right) \qquad s, t \in \{1,2, \ldots, N\} \qquad (4.47)$$

Equation (4. 42) becomes ordinary Brownian motion when H=0.5 and the autocovariance

of X_M, in this case, can be expressed as:

$$\Lambda_{X_M X_M}(s,t) = \min(s, t) \tag{4. 48}$$

This autocovariance is shown in Fig. 4. 25.

The Haar transform of the samples of a segment of Brownian motion has a particularly simple form. The covariances that result from a one-level Haar transform of Brownian motion are obtained by substituting Equation (4. 48) into Equations (4. 6) to (4. 8):

$$\Lambda_{X_{M-1} X_{M-1}}(s,t) = 4 \min(s, t) - 1 - \frac{1}{2}\delta(s-t) \tag{4. 49}$$

$$\Lambda_{U_{M-1} U_{M-1}}(s,t) = \frac{1}{2}\delta(s-t) \tag{4. 50}$$

$$\Lambda_{X_{M-1} U_{M-1}}(s,t) = \begin{cases} -1 & s > t \\ \frac{1}{2} & s = t \\ 0 & s < t \end{cases} \tag{4. 51}$$

A key fact about these equations is that the wavelet coefficients U_{M-1} are white and have a constant variance. In fact by considering higher-level Haar transforms we can show that the wavelet coefficients are white and have a variance that obeys the following relationship:

$$\mathrm{Var}(u[m-1,k]) = 4\,\mathrm{Var}(u[m,j]) - \frac{1}{2} \tag{4. 52}$$

When the sampling interval is not unity but T_s, Equation (4. 52) becomes:

$$\mathrm{Var}(u[m-1,k]) = 4\,\mathrm{Var}(u[m,j]) - \frac{1}{2}T_s \tag{4.53}$$

A more general statement of Equation (4. 53) (for continuous-time wavelet transforms, i.e. in the limit when T_s approaches 0) is presented in [20]. Tewfik and Kim show that the coefficients obtained from a continuous-time wavelet transform of fractional Brownian motions are stationary and have the following variance:

$$\mathrm{Var}(u[m-1,k]) = 2^{(2H+1)}\,\mathrm{Var}(u[m,j]) \tag{4.54}$$

In Fig. 4. 26 (a) to (d), we plot the magnitude of the normalized autocovariances of a one to four-level Haar transform of Brownian motion. In Fig. 4. 26 (e), we show the seven-level Haar transform of the sampled Brownian motion process. This is the highest level transformation of this process that we can consider since X_M is constructed from N samples, where $N = 2^7$. A representative cross-section of these plots is shown in Fig. 4. 27. The variances for the one to four-level and seven-level Haar transformed coefficients are shown in Fig. 4. 28 (a) to (d) and (e), respectively. The Haar transform has localized the correlation between coefficients to "fingers" of the transformed autocovariance. The values of the autocovariance of the transformed coefficients between these fingers is equal to zero.

Also in [20], the authors show that a bound for the covariances of the transformed wavelet coefficients, e.g. $\left\{\Lambda_{U_{M-1}U_{M-1}}, \Lambda_{U_{M-1}U_{M-2}}, \cdots\right\}$, decay more rapidly as the number of vanishing moments of the wavelet increases. Recall that in order to achieve an approximate Karhunen-Loeve expansion, we desire not only that the covariances of the wavelet coefficients at any fixed level, e.g.

$\left\{\Lambda_{U_{M-1}U_{M-1}}, \Lambda_{U_{M-2}U_{M-2}}, \ldots\right\}$, to decay rapidly but also that the covariances of wavelet coefficients between levels, e.g. $\left\{\Lambda_{U_{M-1}U_{M-2}}, \Lambda_{U_{M-1}U_{M-3}}, \ldots\right\}$, be as small as possible. Since the QMFs with 2P-taps derived by Daubechies have the property that the continuous-time wavelet has P vanishing moments, we will compare the autocovariance of the wavelet transformed coefficients using QMFs constructed by Daubechies and QMFs obtained by using the criterion from Chapter 3.

The normalized autocovariance of a three-level wavelet transform of samples of ordinary Brownian motion is shown in Fig. 4. 29 (a) to (d) using optimal two-tap to eight-tap QMFs. A cross-section of these plots is shown in Fig. 4. 30. The variances for the transformed coefficients are shown in Fig. 4. 31. We compare the normalized autocovariance of samples of ordinary Brownian motion using the optimal eight-tap QMFs Fig. 4. 29 (d) to the normalized autocovariance of samples of ordinary Brownian motion using eight-tap QMFs designed by Daubechies Fig. 4. 32 (a). The variances of the coefficients for these two methods are shown in Fig. 4. 31 (d) and Fig. 4. 32 (b). Again we consider a representative cross-section of these normalized autocovariances (that do not exhibit edge effects) by showing the correlation structure of the 100th transformed coefficient. The correlation structure obtained when using both of these methods is shown on a linear scale Fig. 4. 33 (a) and a semi-logarithmic scale Fig. 4. 33 (b).

From these plots we observe that QMFs designed by Daubechies do an excellent job in minimizing the "fingers" of the transformed autocovariance; however, the transformed autocovariance shows a substantial amount of correlation due to edge effects. This should not be surprising since these QMFs have been designed for signals that do not have finite support. Recall that our QMFs are derived by minimizing the cross-covariance, in a mean-square sense, between the coarse approximation and the wavelet coefficients. Thus we are minimizing both the "fingers" and the edge effects that are within this cross-

covariance.

Note that these edge effects are much more prominent for Brownian motion than the first-order Gauss-Markov model. This fact is due to the non-stationarity of Brownian motion. Recall that in order to perform a K-point circular convolution, K-samples of a segment of the process are periodically replicated. For stationary processes, this periodic replication will not introduce a substantial amount of undesired edge effects since all of the samples of the segment of the process have the same distribution. However, since fBm's have variances that increase the coefficients near the edges of the segment will have a substantial amount of undesired correlation. For example, the coefficient with the largest variance will have a substantial amount of correlation with the coefficient that has the smallest variance since, due to the periodic replication, these are adjoining coefficients.

We will concern ourselves with the issues that arise in minimizing the "fingers" rather than the effects that are due to cyclic wrap-around. We therefore make the observation that QMFs designed by Daubechies do a considerably better job than the filters designed by the criterion from Chapter 3. However these observations do suggest a method that we can use to modify the criterion from Chapter 3. Rather than minimizing the norm of the cross-covariance between the coarse-approximation and the wavelet coefficients $\left\|\Lambda_{X_{M-1}U_{M-1}}\right\|_F$, we would minimize the norm of only the coefficients of the cross-covariance (between the coarse-approximation and the wavelet coefficients) that do not exhibit the effects due to cyclic wrap-around.

We now examine the autocovariance of the wavelet transform of samples of processes that have an fBm parameter that is either greater than or less than the fBm parameter that represents Brownian motion (H=0.5). In Fig. 4. 34 we show the autocovariance of the fBm when H=0.25. The autocovariance of the full seven-level Haar transform of this process is shown in Fig. 4. 35 (a). A cross-section of Fig. 4. 35 (a) is

shown in Fig. 4. 35 (b). The variances of the transformed coefficients are shown in Fig. 4. 35 (c). In Fig. 4. 36 we show the autocovariance of the fBm when H=0.75. And again we show the full seven-level Haar transform of the fBm when H=0.75 in Fig. 4. 37 (a). A cross-section of Fig. 4. 37 (a) is shown in Fig. 4. 37 (b). The variances of the transformed coefficients are shown in Fig. 4. 37 (c).

We observe from these plots that, even though the autocovariances of the Haar transform of these fBm's all have the same finger structure, only Brownian motion has exactly zero correlation in between these fingers. We also observe from these plots that the magnitude of the bands of cross-correlation between coefficients at different levels, i.e. the fingers, increase as the fBm parameter H increases. Therefore fBm's with smaller H parameters, e.g. $0 < H \leq 0.5$, seem to be especially well-suited to an approximate Karhunen-Loeve expansion using the wavelet transform.

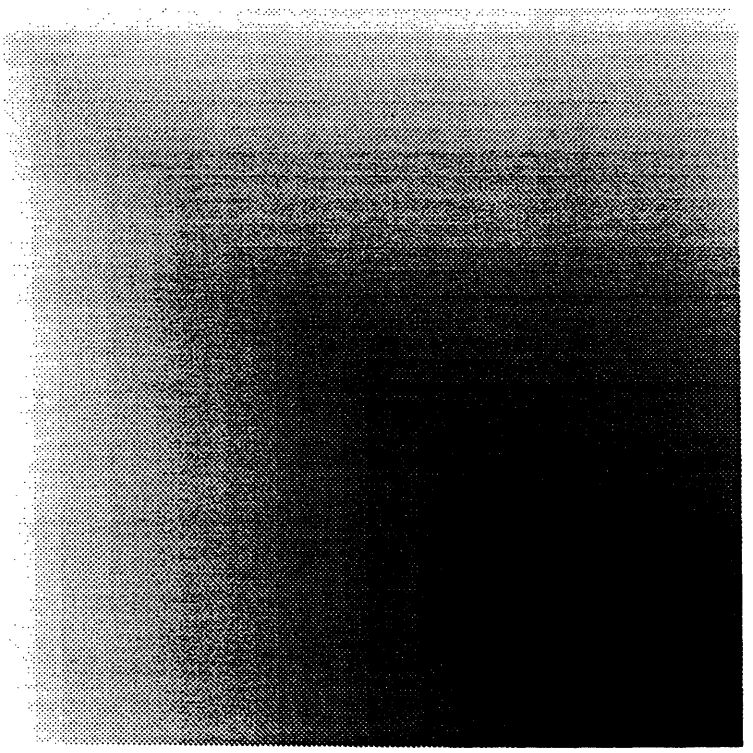


Fig. 4. 25. The autocovariance of samples of ordinary Brownian motion shown as a 2-D image.

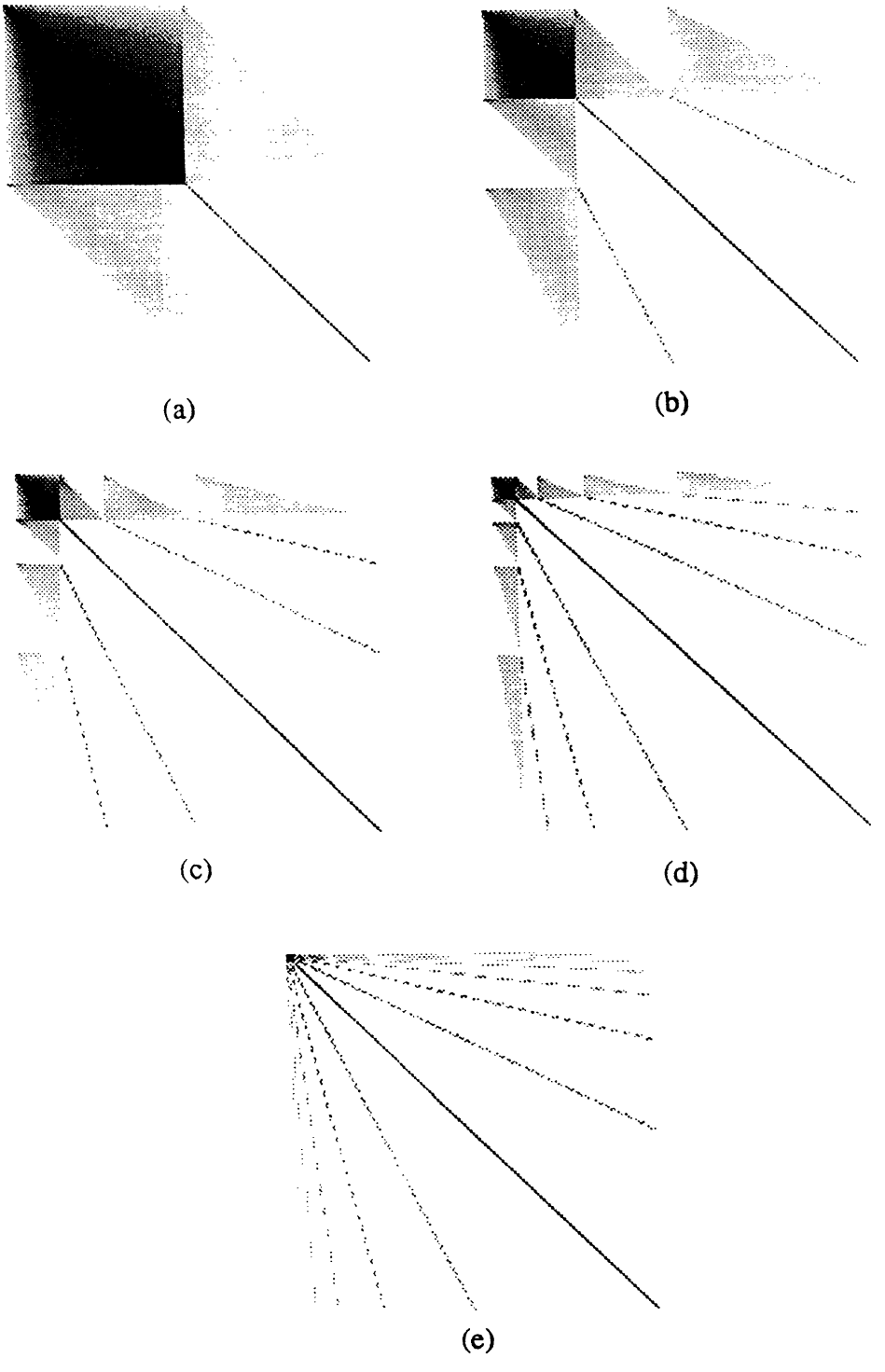


Fig. 4. 26. Plots of the magnitude of the normalized autocovariance of the L-level Haar transform of samples of Brownian motion, where (a) L=1, (b) L=2, (c) L=3, (d) L=4, (e) L=7 (largest level possible). The correlation between fingers is equal to zero.

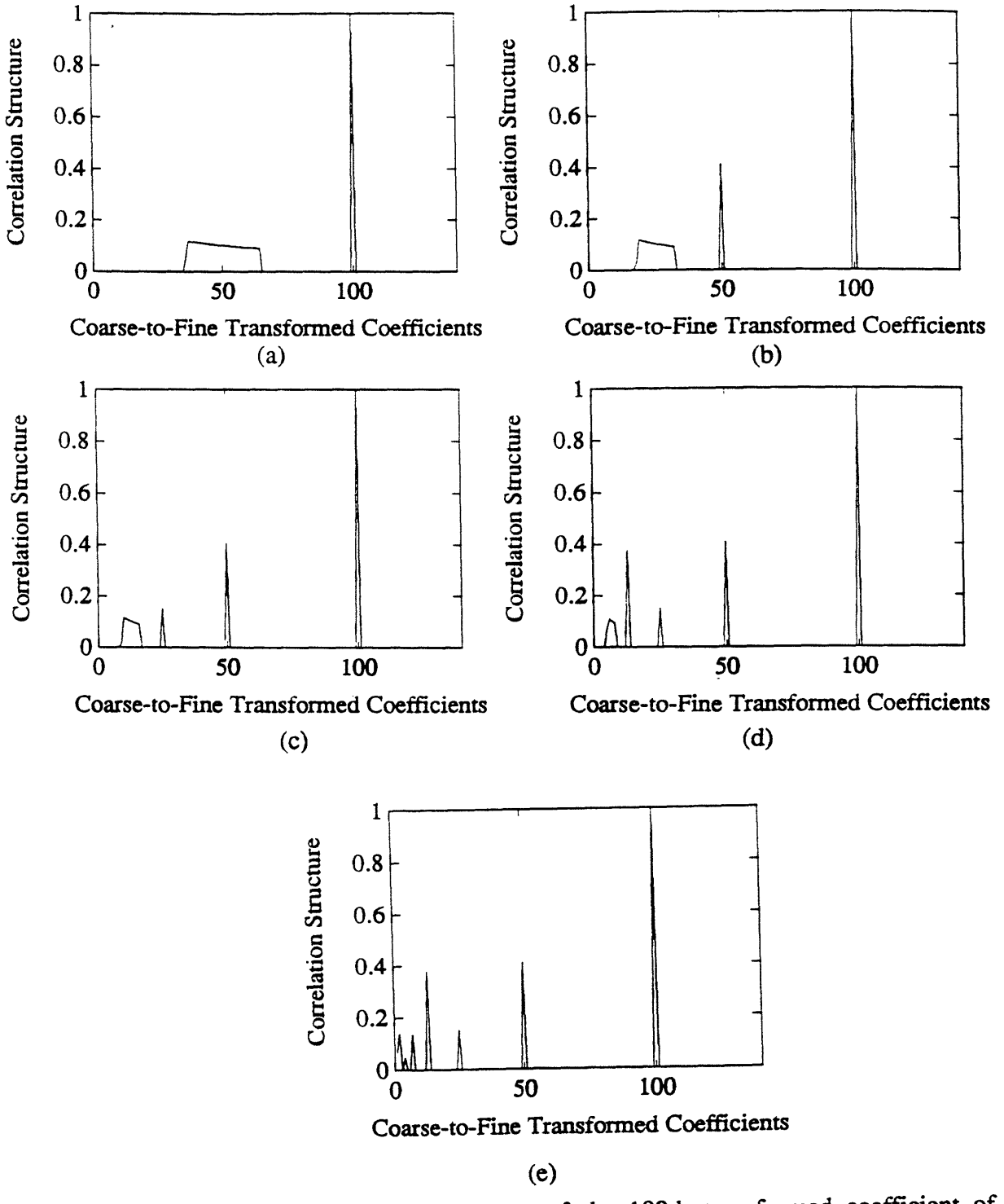


Fig. 4. 27. Plots of the correlation structure of the 100th transformed coefficient of Brownian motion. The transform being used is an of the L-level Haar transform where: (a) L=1, (b) L=2, (c) L=3, (d) L=4, (e) L=7 (largest level possible).

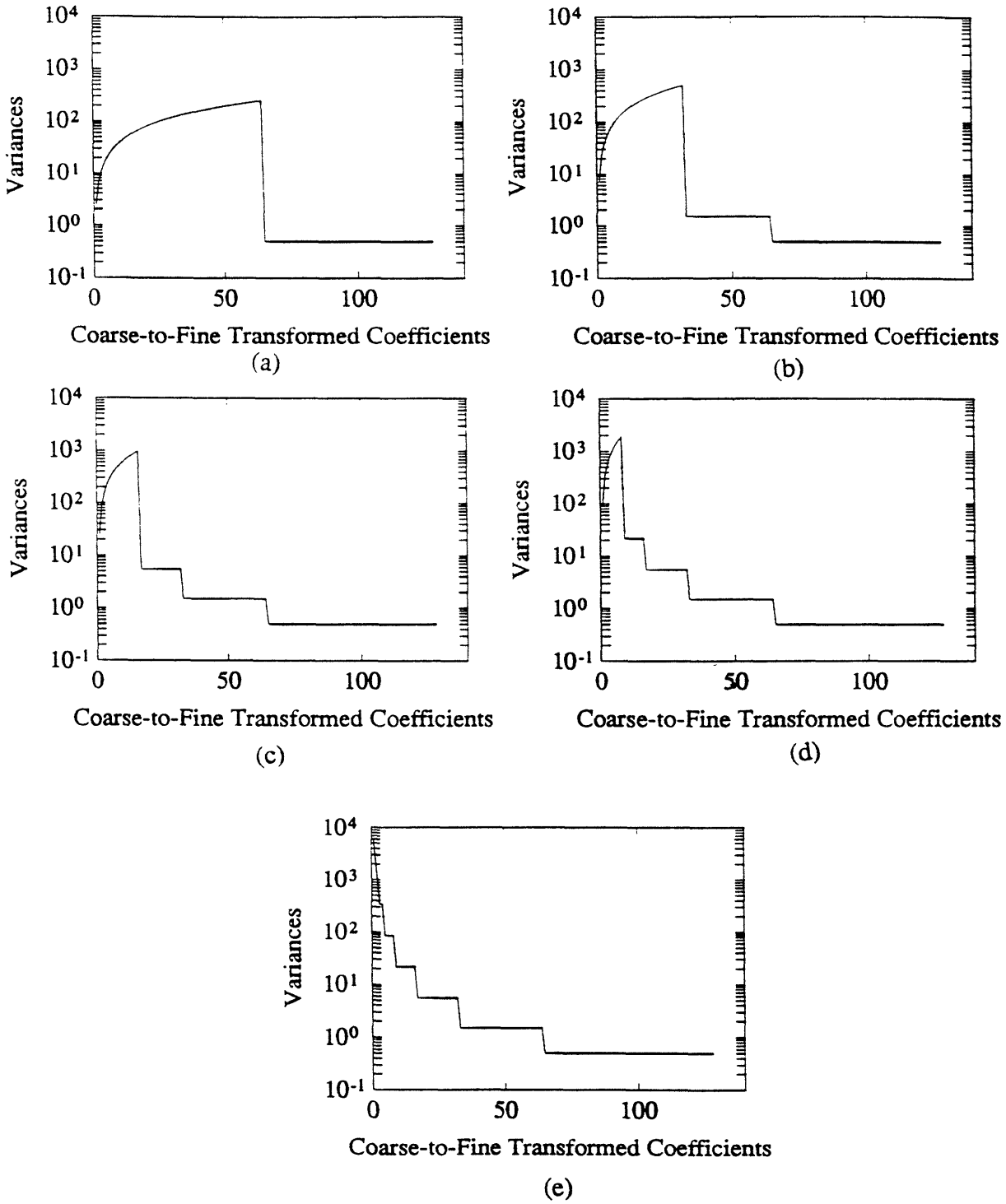


Fig. 4. 28. Plots of the variances of the transformed coefficients that are obtained by taking the L-level Haar transform of samples of Brownian motion, where (a) L=1, (b) L=2, (c) L=3, (d) L=4, (e) L=7 (largest level possible). The plots show that the wavelet coefficients at any level are stationary.

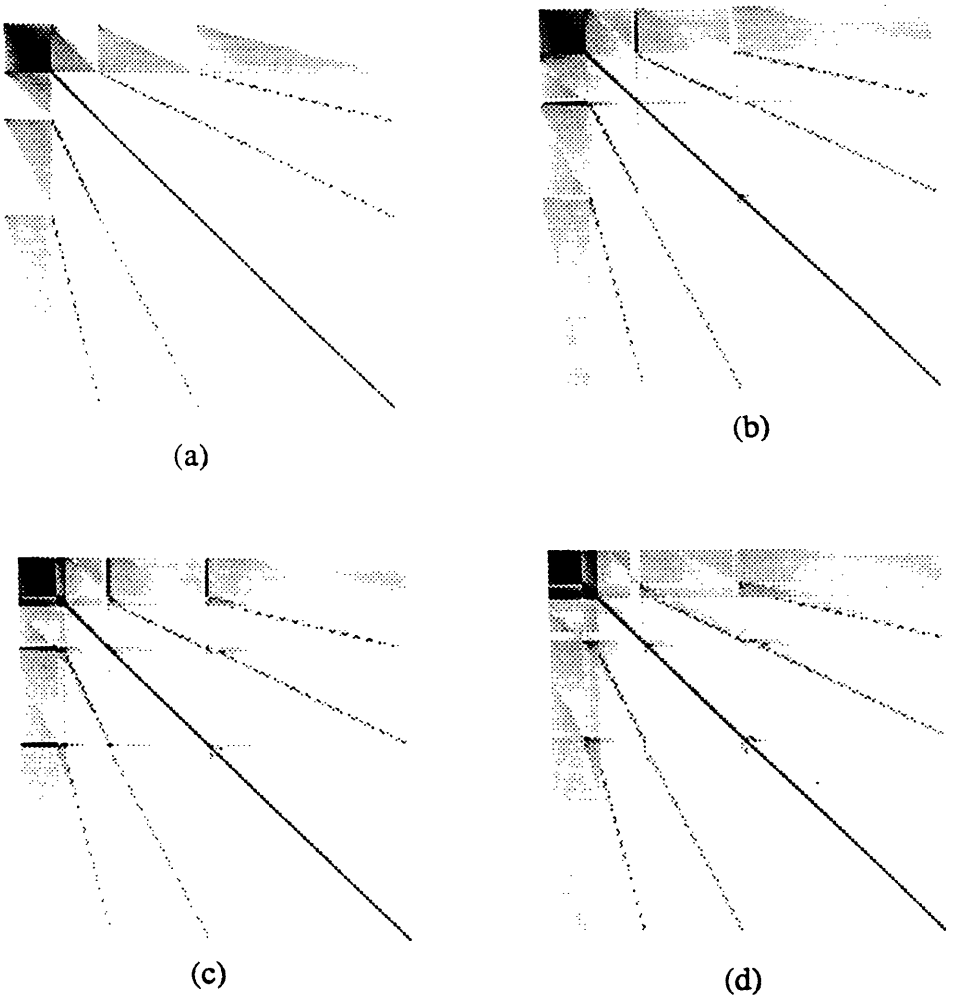


Fig. 4. 29. Plots of the magnitude of the normalized autocovariance of the three-level wavelet transform of samples of Brownian motion using QMFs that were derived by the optimality criterion from Chapter 3. The QMFs have support equaling W, where: (a) W=2, (b) W=4, (c) W=6, (d) W=8.

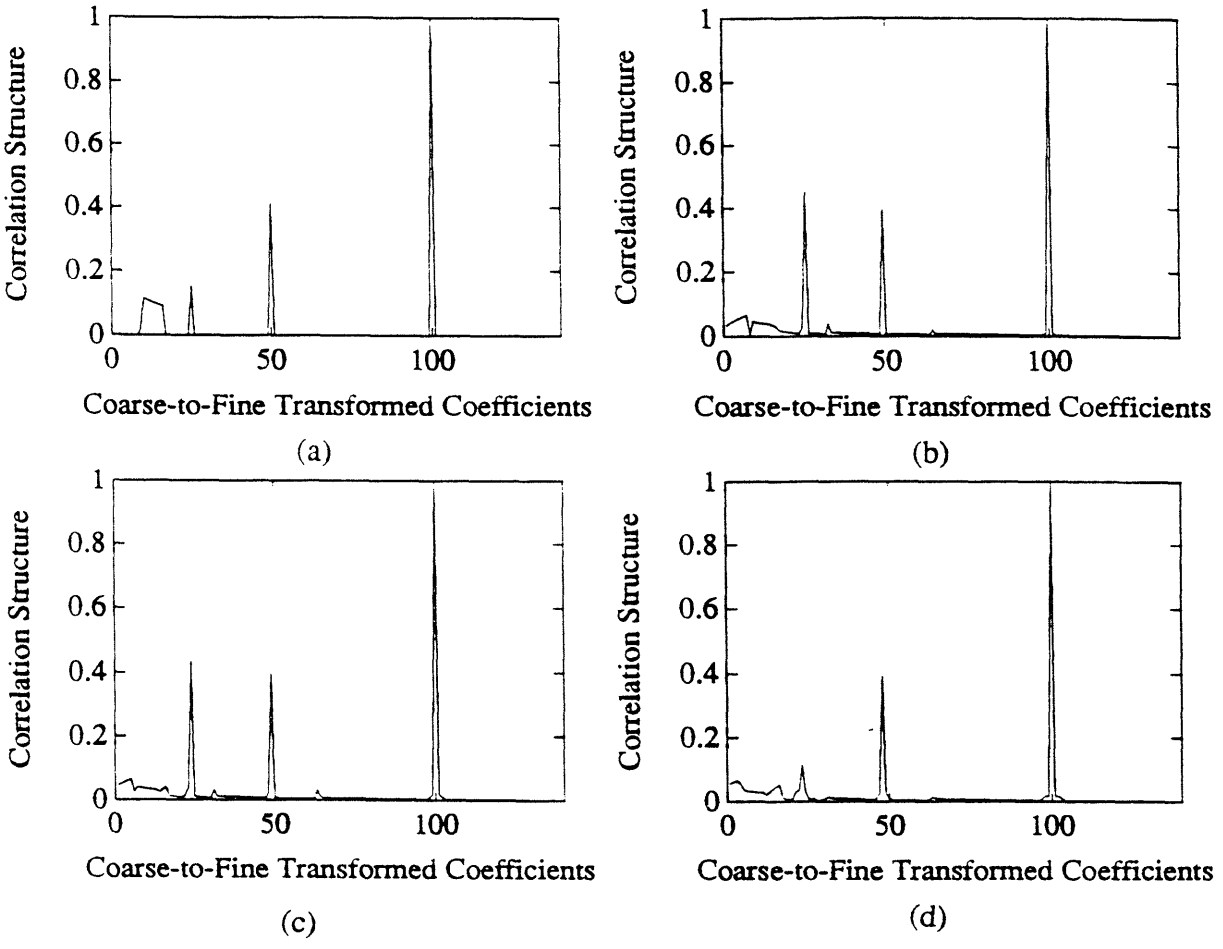


Fig. 4. 30. Plots of the correlation structure of the 100th transformed coefficient of Brownian motion. The transform being used is a three-level wavelet transform using optimal QMFS with support W, where: (a) W=2, (b) W=4, (c) W=6, (d) W=8.

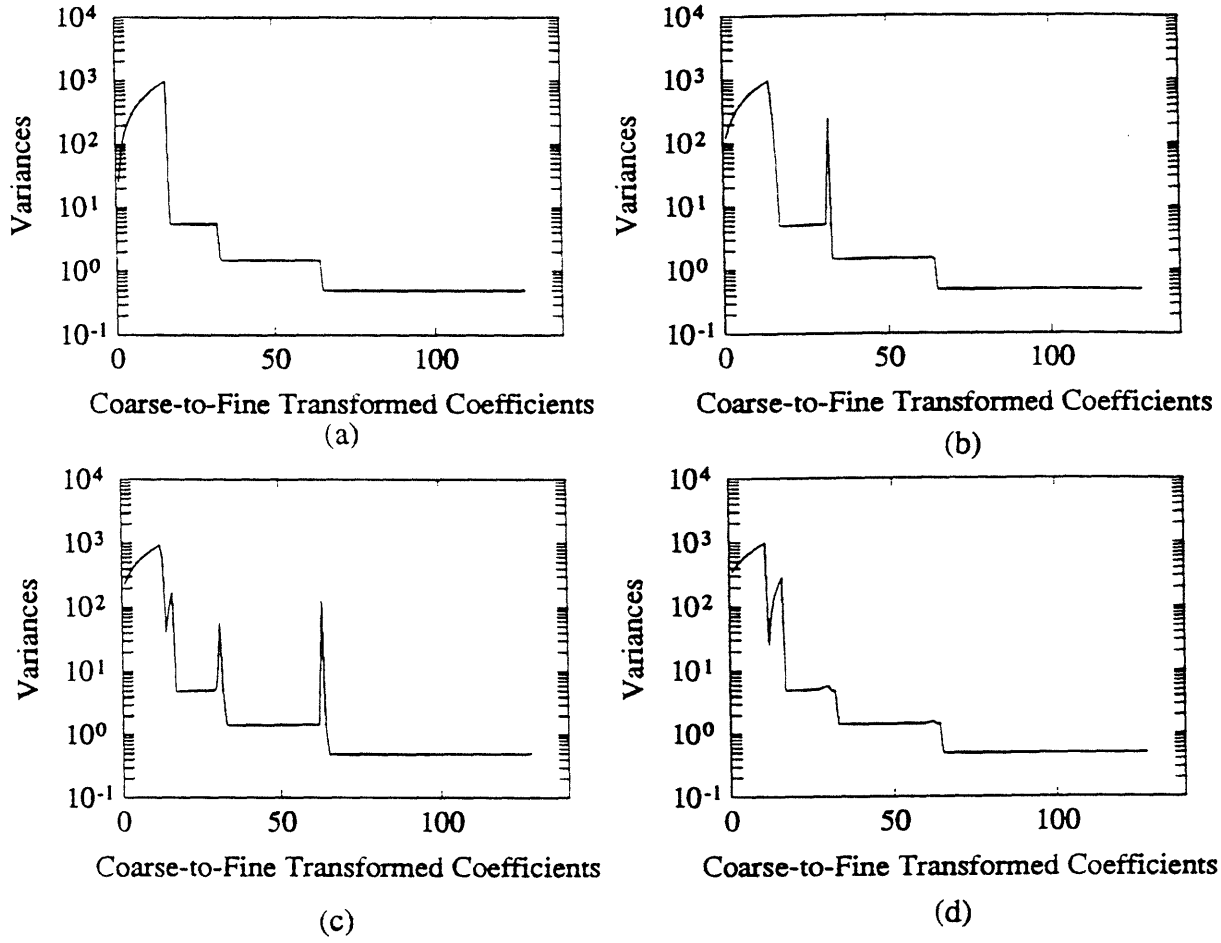
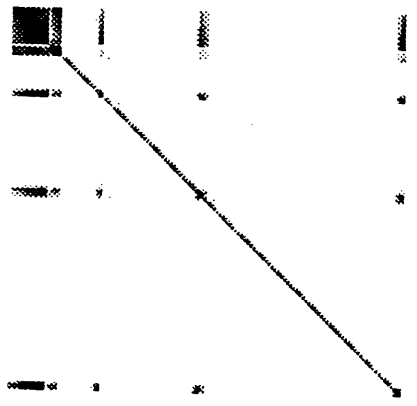
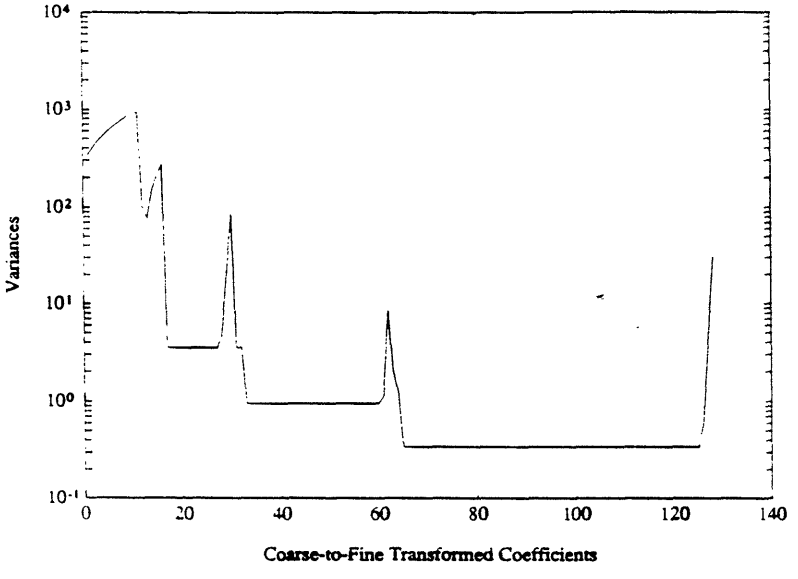


Fig. 4. 31. Plots of the variances of the coefficients obtained by taking a three-level wavelet transform of samples of Brownian motion. The QMFs have support equaling W, where: (a) W=2, (b) W=4, (c) W=6, (d) W=8.

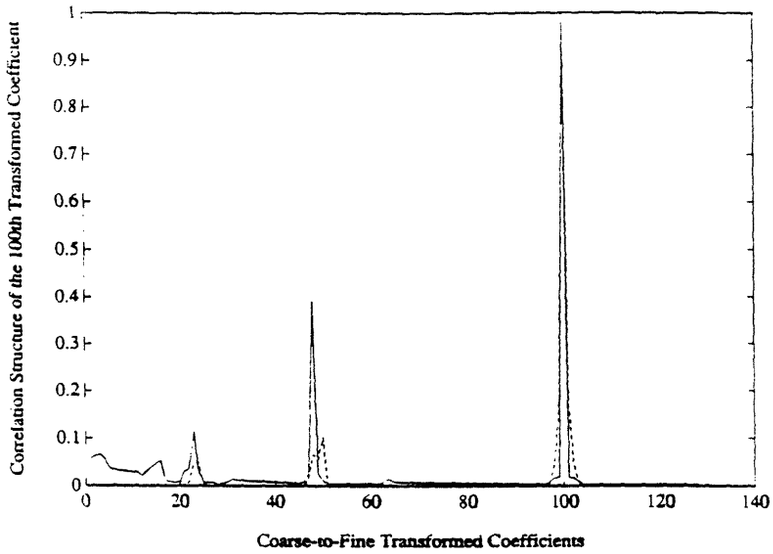


(a)

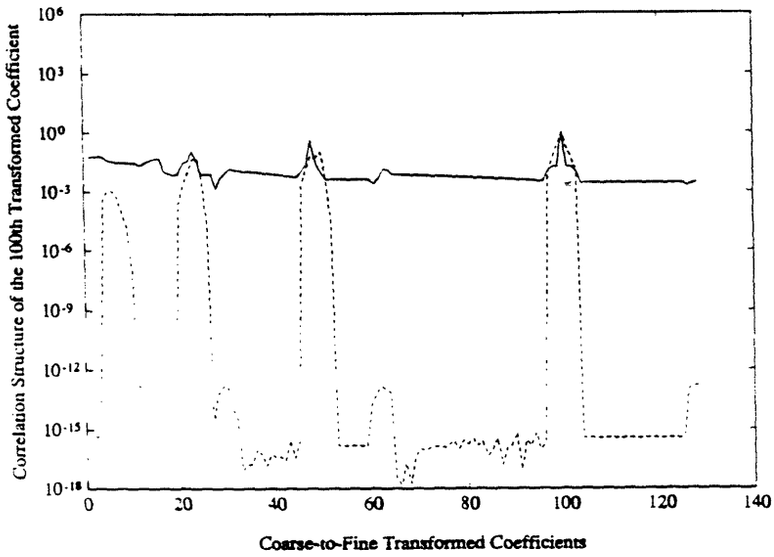


(b)

Fig. 4. 32. Plot (a) shows the normalized autocovariance of the three-level wavelet transform of samples of Brownian motion using eight-tap QMFs. Plot (b) shows the variances of these coefficients. The plot shows that edge effects are much more significant using QMFs derived by Daubechies.



(a)



(b)

Fig. 4. 33. Plots of a representative cross-section (that does not exhibit edge effects) of the normalized autocovariance of the three-level wavelet transform of samples of Brownian motion using eight-tap QMFs. The plots show the correlation structure of the 100th wavelet transformed coefficient using QMFs derived by the optimality criterion from Chapter 3 (solid line) and QMFs derived by Daubechies (dashed line) where the correlation (y-axis) is plotted on: (a) linear scale, (b) logarithmic scale. The QMFs derived by Daubechies did better for the correlation structure of coefficients that do not exhibit edge effects.

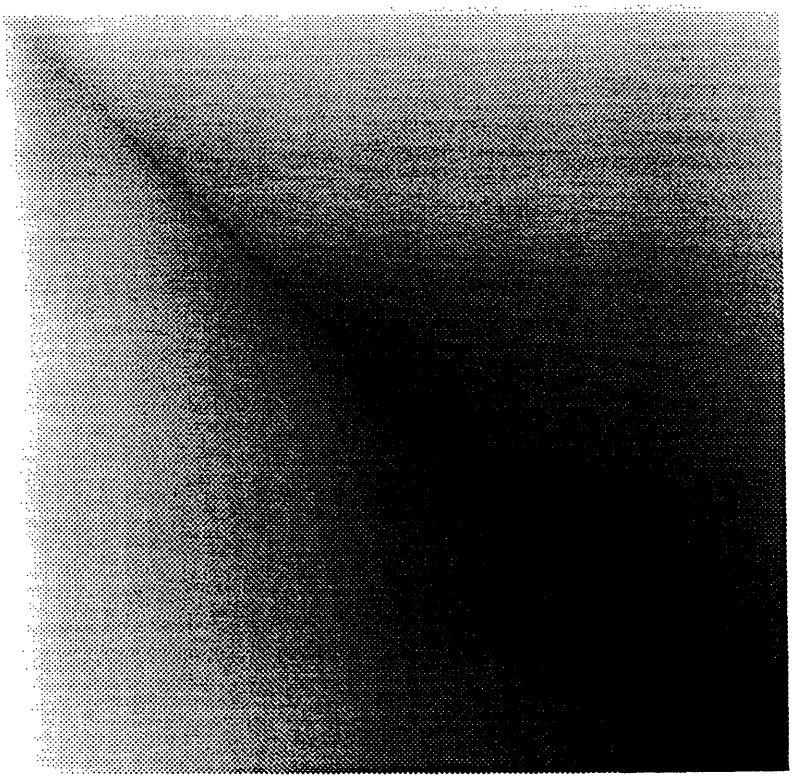
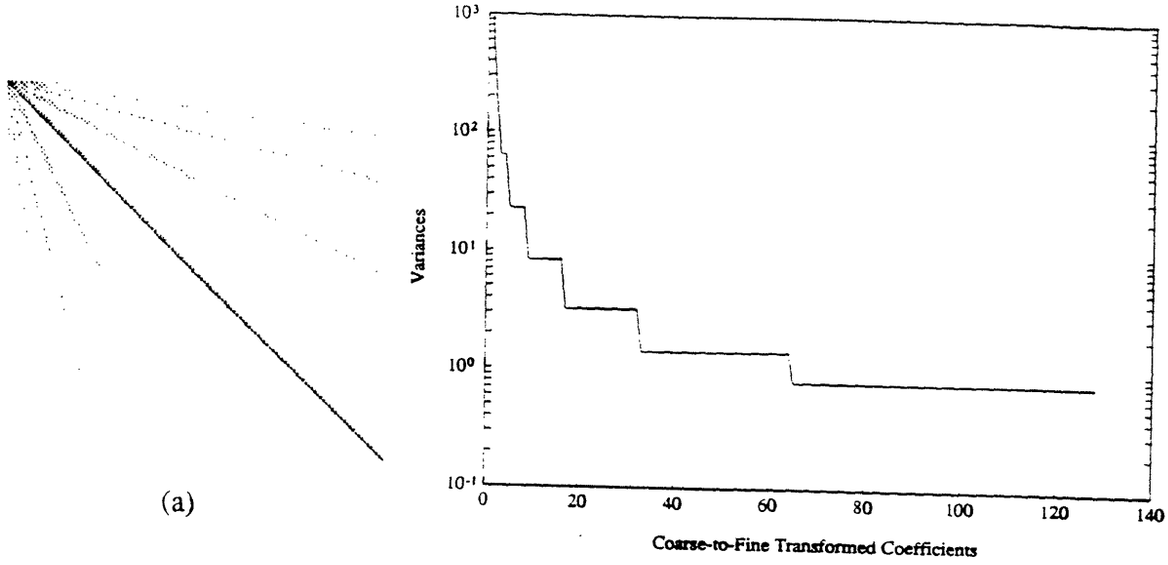
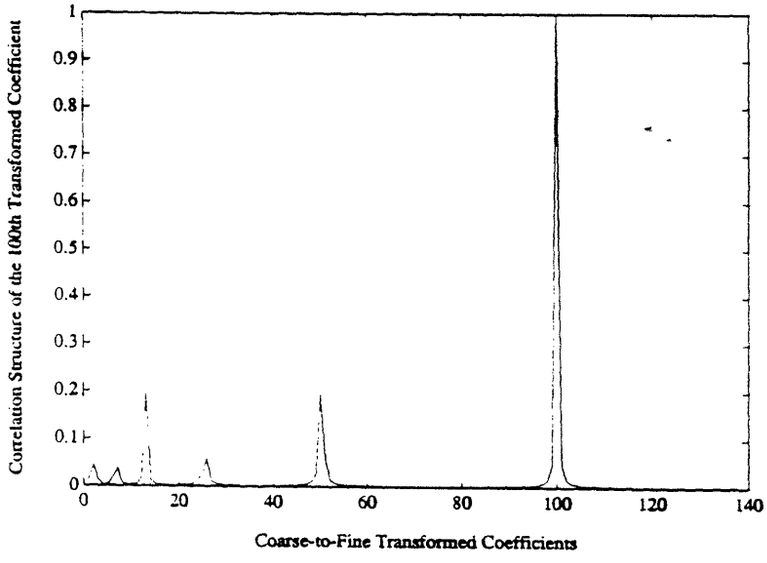


Fig. 4. 34. Plot of the autocovariance of samples of an fBm process with parameter H=0.25.





(c)

Fig. 4. 35. (a) Plot of the autocovariance of the full (seven-level) Haar transform of an fBm process with parameter H=0.25. Plot (b) shows a cross-section of the plot in (a). Plot (c) shows the variances of the transformed coefficients.

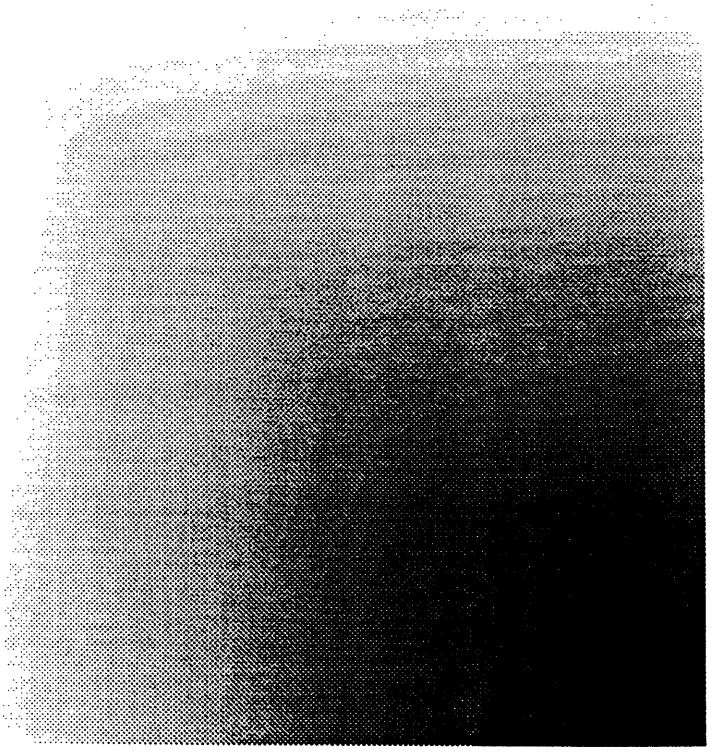
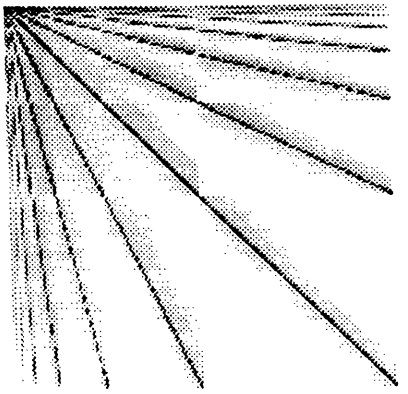
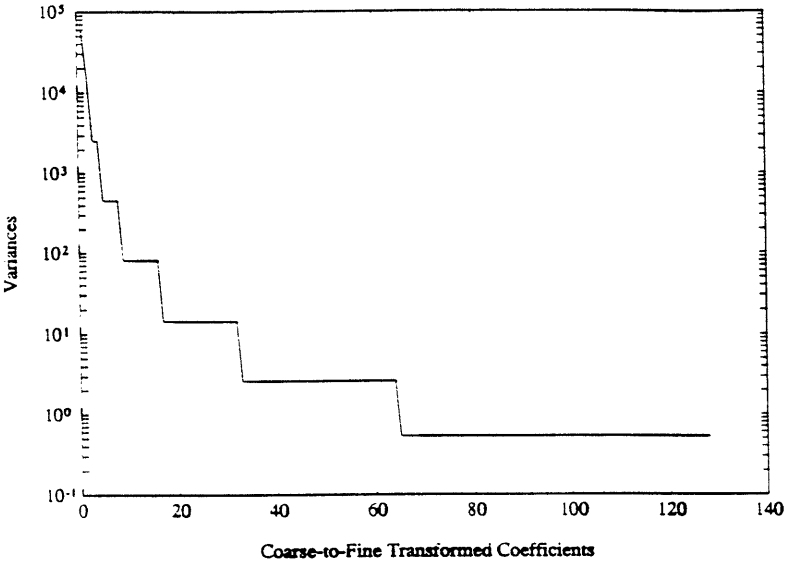


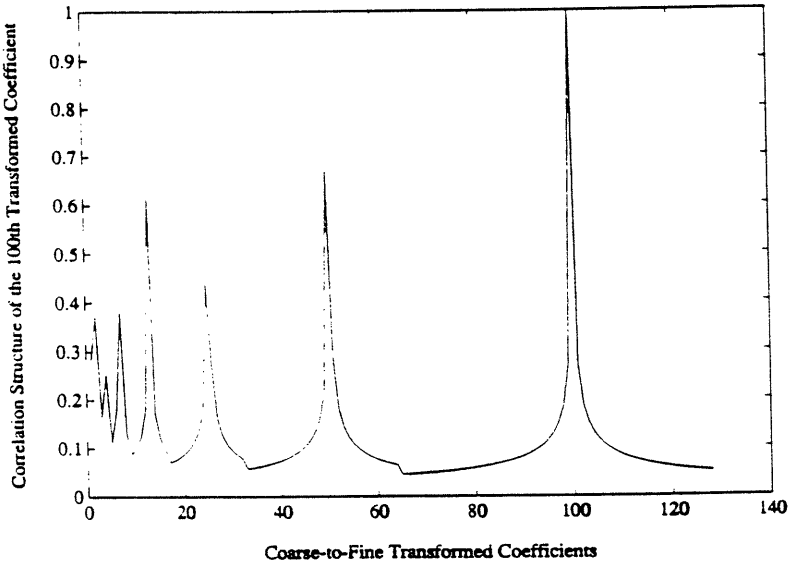
Fig. 4. 36. Plot of the autocovariance of samples of an fBm process with parameter H=0.75.



(a)



(b)



(c)

Fig. 4. 37. (a) Plot of the autocovariance of the full (seven-level) Haar transform of an fBm process with parameter H=0.75. Plot (b) shows a cross-section of the plot in (a). Plot (c) shows the variances of the transformed coefficients.

CHAPTER 5

MODELING PROCESSES USING THE WAVELET TRANSFORM

5.1 MODELING PROCESSES

As we have discussed, there are very efficient signal processing algorithms that are based on the assumption that the wavelet transform coefficients of the samples of the process are uncorrelated. This in fact provided the motivation for our investigation of methods to achieve maximal decorrelation via the wavelet transform. In this chapter we take another look at these methods by examining the implications on the modeling of processes by assuming that the wavelet transform coefficients are indeed white. We compare the statistics of the samples of a given process with those of a process that approximates the given (or true) process by neglecting the correlation between wavelet coefficients. Specifically, the autocovariance of the approximated process is obtained, in a manner to be made precise, by transforming the autocovariance of the samples of the process into the wavelet transform domain, the transformed autocovariance is then approximated by assuming whiteness among coefficients, and then the transformed autocovariance is transformed back to the domain of the original process. First, we will compare the approximation of the process to the true process by showing the 2-D images of the autocovariance of these processes. Second, we will be able to quantify how close the approximated process is to the true process by using the Bhattacharyya distance [28]. Third we will show that sample paths generated from the approximated processes become increasingly similar to the true process as the support of the QMFs increase.

We investigate the effect of making such an approximation by using the following

analysis. We collect the samples of a process into a vector and denote the vector as X_M. We again denote the autocovariance of this process as $\Lambda_{X_M X_M}$ and the autocovariance of the wavelet transform of X_M as $\Lambda_{Y_M Y_M}$. Of course the autocovariance of X_M and the autocovariance of the wavelet transform of X_M are related via the following equation:

$$\Lambda_{X_M X_M} = T_M^T \Lambda_{Y_M Y_M} T_M \tag{5. 1}$$

The autocovariance of the wavelet transformed coefficients can be represented as the sum of two terms:

$$\Lambda_{Y_M Y_M} = D_{Y_M Y_M} + F_{Y_M Y_M} \tag{5. 2}$$

where the first term $D_{Y_M Y_M}$ is constructed from the diagonal elements of $\Lambda_{Y_M Y_M}$, i.e. the variances of the transformed coefficients, and all off-diagonal elements are set to zero. The second term $F_{Y_M Y_M}$ is constructed from the off-diagonal elements of $\Lambda_{Y_M Y_M}$ with the diagonal elements being set to zero. Of course using this decomposition, the elements that constitute the "fingers" of the transformed autocovariance are contained in the matrix $F_{Y_M Y_M}$.

Substituting Equation (5. 2) into Equation (5. 1) we obtain the following relationship:

$$\Lambda_{X_M X_M} = T_M^T D_{Y_M Y_M} T_M + T_M^T F_{Y_M Y_M} T_M \tag{5. 3}$$

In order to refer to these two terms in a more concise manner, we let:

$$A_M = T_M^T D_{Y_M Y_M} T_M \tag{5. 4}$$

$$E_M = T_M^T F_{Y_M Y_M} T_M \tag{5. 5}$$

The matrix A_M is the autocovariance of the Karhunen-Loeve approximation using the

wavelet transform. The matrix E_M represents the error in the approximation or, in other words, the contribution of $\Lambda_{X_M X_M}$ that we would ideally like to be equal to zero. In order to investigate how-well A_M approximates $\Lambda_{X_M X_M}$, we examine the information that is lost when we make this approximation, i.e. E_M.

For simplicity, in our analysis and in order to isolate the key aspects of the approximation errors, let us examine a one-level wavelet transform, i.e. where we transform X_M to $\{X_{M-1}, U_{M-1}\}$ and then approximate the statistical description at this stage. Using the method we have described, our method for approximating processes using a one-level wavelet transform would be to only keep the variances of the elements of X_{M-1} and U_{M-1}, i.e. to force the covariance of the vector Y_M (which is formed from the vectors X_{M-1} and U_{M-1}) to be diagonal. As we have seen in Chapter 4, this would seem to be a reasonable approximation for U_{M-1} since $\Lambda_{U_{M-1}U_{M-1}}$ is typically nearly diagonal. However, we also know from Chapter 4 that it is typically unreasonable to approximate $\Lambda_{X_{M-1}X_{M-1}}$ by a diagonal matrix. It makes more sense to keep the entire matrix $\Lambda_{X_{M-1}X_{M-1}}$ and approximate $\Lambda_{U_{M-1}U_{M-1}}$ as a diagonal matrix or to perform a higher-level transform that further decomposes and whitens X_{M-1}. In the following analysis, we will show that, whether we approximate $\Lambda_{X_{M-1}X_{M-1}}$ as a diagonal matrix or keep the entire matrix, the autocovariance of the approximation of X_M will have a periodic structure along its diagonals.

To begin this analysis let us consider transforming the covariance $\Lambda_{X_{M-1}U_{M-1}}$ back into the domain of the original signal. This transformed covariance will be one of the terms

that makes up the error in the approximation, i.e. one term of E_M, and as we have argued in the preceding paragraph, this term and its transpose should be the dominant portion of E_M (since we will keep all of $\Lambda_{X_{M-1}X_{M-1}}$, and $\Lambda_{U_{M-1}U_{M-1}}$ is nearly diagonal). Specifically, this transformed covariance is:

$$\Lambda_{\tilde{X}_{M-1}\tilde{U}_{M-1}} = H_{M-1}^T \Lambda_{X_{M-1}U_{M-1}} G_{M-1} \tag{5. 6}$$

where $\tilde{X}_{M-1}$ and $\tilde{U}_{M-1}$ are the projections of X_{M-1} and U_{M-1} back into the domain of the original signal. Note that $\Lambda_{\tilde{X}_{M-1}\tilde{U}_{M-1}}$ should not be interpreted as a covariance matrix since this matrix is not positive definite; however, $\Lambda_{\tilde{X}_{M-1}\tilde{U}_{M-1}}$ is one of several terms that constitute E_M. The sequences $\tilde{X}_{M-1}$ and $\tilde{U}_{M-1}$ can be obtained by using the operators in Equation (2. 13); however, an equivalent way of obtaining $\tilde{X}_{M-1}$ and $\tilde{U}_{M-1}$ is more useful for our analysis:

$$\tilde{x}[M-1,s] = \sum_{i \in Q_s} h[i]x[M-1,\tfrac{s+i}{2}] \tag{5. 7}$$

$$\tilde{u}[M-1,t] = \sum_{j \in Q_t} g[j]u[M-1,\tfrac{t+j}{2}] \tag{5. 8}$$

where
$$Q_s = \begin{cases} \text{even integers} & \text{if s is even} \\ \text{odd integers} & \text{if s is odd} \end{cases} \tag{5. 9}$$

These equations yield the following relationship between $\Lambda_{\tilde{X}_{M-1}\tilde{U}_{M-1}}$ and $\Lambda_{X_{M-1}U_{M-1}}$:

$$\Lambda_{\tilde{X}_{M-1}\tilde{U}_{M-1}}(s,t) = \sum_{i\in Q_s}\sum_{j\in Q_t} h[i]g[j]\Lambda_{X_{M-1}U_{M-1}}(\frac{s+i}{2}, \frac{t+j}{2}) \quad (5.10)$$

We have shown in Chapter 4 that if X_M is stationary then $\Lambda_{X_{M-1}U_{M-1}}$ will depend only upon the time difference. In this case, Equation (5. 10) becomes:

$$\Lambda_{\tilde{X}_{M-1}\tilde{U}_{M-1}}(s,t) = \sum_{i\in Q_s}\sum_{j\in Q_t} h[i]g[j]\Lambda_{X_{M-1}U_{M-1}}(\frac{s-t+i-j}{2}) \quad (5.11)$$

Since the values that we are summing over in Equation (5. 11) depend upon s and t, $\Lambda_{\tilde{X}_{M-1}\tilde{U}_{M-1}}$ is not strictly a function of the time difference s-t. However $\Lambda_{\tilde{X}_{M-1}\tilde{U}_{M-1}}$ does have a very special structure. Specifically, let us consider the diagonals of $\Lambda_{\tilde{X}_{M-1}\tilde{U}_{M-1}}$, i.e. when s - t is a constant. We observe that the diagonals of the matrix are periodic along the diagonal with period two. This fact becomes clear by considering two different alternatives when s - t is a constant. The first alternative is that s - t is even: in this case, the summation in Equation (5. 11) would be summed over all even integers i and j or all odd integers i and j. The important point being that along the diagonal there are only two possible values and these values alternate along the diagonal. The second alternative is that s - t is odd: in this case, the summation is over all even integers i and all odd integers j or all odd integers i and all even integers j. And again the two possible values alternate along any given diagonal.

The analysis using Equation (5. 11) is sufficient to show that the diagonals of E_M must be periodic with period two. This statement is shown by realizing that for a one-level transform E_M is the sum of four terms. The first and second of these terms are $\Lambda_{\tilde{X}_{M-1}\tilde{U}_{M-1}}$ and its transpose, which we know to have this periodic structure. The third term is

constructed from $\Lambda_{\tilde{U}_{M-1}\tilde{U}_{M-1}}$ with the main diagonal of $\Lambda_{X_{M-1}X_{M-1}}$ set to zero. Thus the computation of $\Lambda_{\tilde{U}_{M-1}\tilde{U}_{M-1}}$ is analogous to the computation of $\Lambda_{\tilde{X}_{M-1}\tilde{U}_{M-1}}$. The only differences are that h[n] is replaced by g[n] and a different matrix that also has a Toeplitz structure is used instead of $\Lambda_{X_{M-1}U_{M-1}}$, i.e. $\Lambda_{X_{M-1}X_{M-1}}$ with its main diagonal set to zero. Thus this third term of E_M must have the same periodic structure as the first two terms. If the entire matrix $\Lambda_{X_{M-1}X_{M-1}}$ is kept in the approximation of X_M then there would not be a fourth term; however, if only the diagonal of $\Lambda_{X_{M-1}X_{M-1}}$ is kept then the fourth term would be constructed in a manner analogous to the construction of the third term. Thus E_M will have the afore mentioned periodic structure, and since $\Lambda_{X_MX_M}$ is a Toeplitz matrix the approximation A_M will also have this periodic structure.

There is also a periodic structure in A_M and E_M when we carry out the approximation to more than a single level. Specifically, the approximation A_M of a stationary process when using an L-level wavelet transform will be periodic along the diagonals with the period being equal to 2^L. (Of course there does not necessarily have to be any periodicity along the diagonals if the period, 2^L, was equal to the length of the segment.) Due to this periodicity, we believe that processes that are cyclostationary [38] are especially well-suited to this method of modeling, since cyclostationary processes have autocovariances with periodic diagonals. Specifically, the autocovariance of a cyclostationary process has the property [38]:

$$\Lambda_{XX}(i + nT, j + nT) = \Lambda_{XX}(i, j) \tag{5. 12}$$

where n is any integer and T is the period.

In this chapter, we use the Bhattacharyya distance as a means of measuring how

close the approximation is to the true sampled process. The Bhattacharyya distance allows us to determine an upper bound on the probability of error of deciding whether a sample path originated from the true process or its approximation.

The Bhattacharyya distance is defined as [28]:

$$B = -\ln \int [p_1(z)p_2(z)]^{1/2} dz \tag{5. 13}$$

where $p_1(z)$ and $p_2(z)$ are the probability density functions of the two random vectors under consideration. Using the assumptions that the two random vectors under consideration have equal means, unequal covariances (i.e. the covariances be Λ_1 and Λ_2), and the random vectors consist of Gaussian random variables then Equation (5. 13) becomes [28]:

$$B = \frac{1}{2} \ln \left| \frac{1}{2}(\Lambda_1 + \Lambda_2) \right| - \frac{1}{4} \ln |\Lambda_1 \Lambda_2| \tag{5. 14}$$

where $|\cdot|$ represents the determinant of the argument. An upper bound on the probability of error of deciding whether a sample path originated from two equally likely processes is obtained via the Bhattacharyya distance [28]:

$$P_e \leq \frac{1}{2} e^{-B} \tag{5. 15}$$

Also in this chapter, we briefly consider the situation in which the process in question is observed in the presence of additive white Gaussian noise. Specifically, we consider observations of the form:

$$\hat{X}_M = X_M + W_M \tag{5. 16}$$

where X_M is the vector of samples of either a first-order Gauss-Markov process (Section 5.2) or an fBm (Section 5.3), and W_M denotes the vector of samples of white Gaussian

noise that is independent of X_M. The variance of the samples of W_M are chosen so that the Signal-To-Noise ratio (SNR) of X_M to W_M can be controlled and varied. In particular, we consider SNR values of infinity (no noise), five, two, and one, where the SNR of the ith coefficient of $\hat{X}_M$ is calculated as the standard deviation of the ith coefficient of X_M divided by the standard deviation of the ith coefficient of W_M.

5.2 MODELING THE FIRST-ORDER GAUSS-MARKOV PROCESS

Recall the first-order Gauss-Markov process that we considered in Chapter 4. Specifically, the process is of the form:

$$x[n] = \alpha x[n-1] + w[n] \tag{5. 17}$$

We again collect the samples of this process into a vector denoted by X_M. The autocovariance of this process when α is equal to 0.9006 is shown as a 2-D image in Fig. 4. 1 (c).

We achieve better approximations of the true process by using higher-level wavelet transforms. Recall that the L-level wavelet transform of X_M yields the transformed coefficients (U_{M-1}, U_{M-2}, . . ., U_{M-L}, X_{M-L}). As we have seen in the preceding chapter, the covariance of X_{M-L} typically exhibits a great deal of correlation; that is the sequence X_{M-L} is a coarse approximation of the original process, and thus X_{M-L} is typically not well-approximated by a white sequence. Of course by increasing L the number of elements, i.e. the support, of X_{M-L} will decrease which will typically reduce the error term, E_M. Thus we would expect that the best approximate Karhunen-Loeve expansion of a process using the wavelet transform will consist of using an L-level wavelet transform where L is as large as possible. Ideally we would like to choose the maximum value of L, i.e. $L = \lfloor \log_2(N) \rfloor$, where N is the length of the segment and $\lfloor \cdot \rfloor$ represents the greatest integer that is less than or equal to the argument; however, a problem arises when L approaches this value and the support of the QMFs are greater than two. Recall that a K-point circular convolution is performed at the mth level of the wavelet transform where $K = N/2^{M-m+1}$. The problem arises as to decide how to perform the convolution step when the support of the QMFs, W,

is greater than K. We avoid this issue by simply choosing the level of the wavelet transform, L, such that W is always less than or equal to K. That is we choose L via the equation:

$$L = \lfloor \log_2(\frac{2N}{W}) \rfloor \tag{5. 18}$$

In Fig. 5. 1 (a) to (d), we show the autocovariances of the approximation of the first-order Gauss-Markov process obtained from a Karhunen-Loeve expansion using the wavelet transform, i.e. A_M, with QMFs that were derived by Daubechies and have support ranging from two to eight-taps. The number of levels of the wavelet transform that were used in each of these plots was determined using Equation (5. 18). That is since N=128, we used a seven-level transform for two-tap QMFs, a six-level transform for four-tap QMFs, and a five-level transform for both six-tap and eight-tap QMFs. The errors in these approximations, E_M, is shown in Fig. 5. 2 (a) to (d).

As we can see from the 2-D images, the main difference between the autocovariance of the original Gauss-Markov process and the autocovariance of the approximation is the periodicity along the diagonals. We examine the periodicity along the main diagonal by observing the plots of the variances of the approximations, Fig. 5. 3 (a) to (d). Note from these plots that the variances are periodic and, in some cases, the period is even smaller than 2^L.

We use the Bhattacharyya distance to quantify how close the sampled Gauss-Markov process is to the approximation of the Gauss-Markov process when using the wavelet transform. In Fig. 5. 4 (a) we show the Bhattacharyya distance, calculated by using Equation (5. 14), between the Gauss-Markov process and the approximation of the Gauss-Markov process when using the wavelet transform. In this plot the distance is shown as the support of the QMFs increase from two to eight-taps. We also consider, in this plot, that the process being modeled is not simply a first-order Gauss-Markov process but a first-

order Gauss-Markov process in the presence of additive white Gaussian noise. The Bhattacharyya distance is shown when the variance of the samples of W_M are chosen so that the Signal-To-Noise ratio (SNR) of X_M to W_M is five, two, and one. Equation (5. 15) is used to determine an upper bound on the probability of error of deciding whether a sample path originated from the first-order Gauss-Markov process in the presence of white noise or from its approximation. This probability of error is shown in Fig. 5. 4 (b).

We observe, as expected, that the Bhattacharyya distance decreases as the support of the QMFs increase, i.e. the Gauss-Markov process and the approximation of the Gauss-Markov process becomes more similar as the support of the QMFs increase. Also, as expected, the upper bound on the probability of error of deciding whether a sample path originated from the Gauss-Markov process or the approximation of the Gauss-Markov process increases as the support of the QMFs increase. That is since the Gauss-Markov process and the approximation of the Gauss-Markov process are becoming more similar (as the support of the QMFs increase), the ability of determining which process the sample path originated from is becoming more difficult.

To provide a point of comparison it is useful to relate the Bhattacharyya distance between the true and the approximated process to the value of the Gauss-Markov parameter that corresponds to a Gauss-Markov process that is at an equivalent distance from the true process. Specifically, consider Fig. 5. 5 where we show the Bhattacharyya distances between Gauss-Markov processes with various parameters α and the true Gauss-Markov process with parameter $\alpha = 0.9006$. Suppose that the approximated process is at a distance d from the true process. We can use Fig. 5. 5 to find the parameter of the Gauss-Markov process that is a distance d from the true process. In Fig. 5. 6 (a) we plot the equivalent (in the sense that the distance between the true and approximated process is the same as the distance between the true process and another Gauss-Markov process) Gauss-Markov parameter of the approximated process when we use the portion of the mapping

for $\alpha \leq 0.9006$. In Fig. 5. 6 (b), we show the same plot but using the portion of the mapping for $\alpha > 0.9006$.

From Fig. 5. 6, we observe that the equivalent parameter of the Gauss-Markov process is very close to the true process. For example consider approximating a Gauss-Markov process with parameter $\alpha = 0.9006$ using a wavelet transform with eight-tap QMFs when the SNR equals one. The distance between the approximation and the true process is the same as the distance between a Gauss-Markov process with parameter $\alpha = 0.9006$ and a Gauss-Markov process with either a parameter $\alpha = 0.882$ or a parameter $\alpha = 0.916$. In many applications simply trying to estimate the Gauss-Markov parameter will result in errors of this amount implying that this difference is not particularly significant.

Finally, in Fig. 5. 8 (a) to (d) we show that the sample paths generated from the approximated processes become increasingly similar to a sample path generated from the true process, Fig. 5. 7, as the support of the QMFs increase. The sample path in Fig. 5. 7 was generated by collecting 128 identically distributed samples of white noise into a vector W_M and then multiplying this vector by the square root of the autocovariance of a first-order Gauss-Markov process. That is

$$X_M = \Lambda_{X_M X_M}^{1/2} W_M \tag{5. 19}$$

where $\Lambda_{X_M X_M}^{1/2}$ was obtained by a Cholesky factorization [35] of $\Lambda_{X_M X_M}$. Similarly, the sample paths generated from the approximated processes were obtained by replacing $\Lambda_{X_M X_M}^{1/2}$ in Equation (5. 19) with $A_M^{1/2}$ and using the same samples of white noise that were used in the generation of the Gauss-Markov sample path.

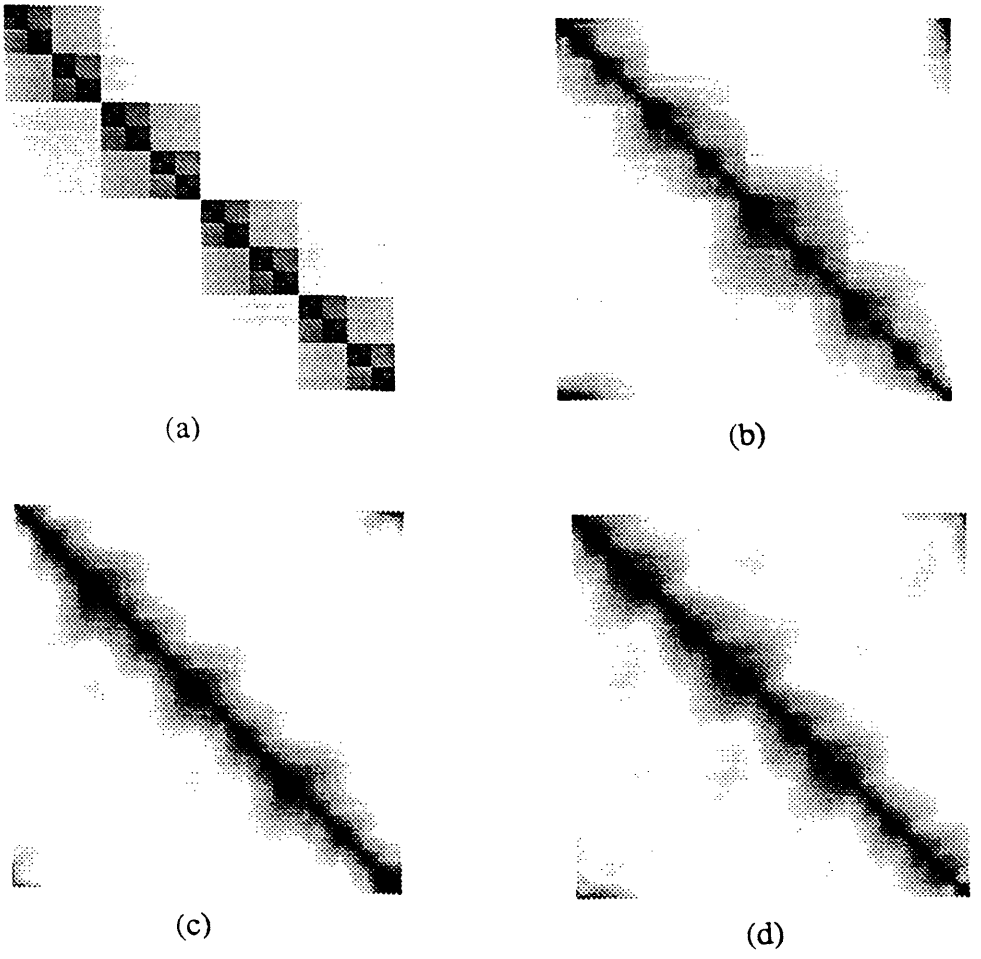


Fig. 5. 1. Plots of A_M, the autocovariances of the approximations of the first-order Gauss-Markov process, when using QMFs that have support equaling W, where: (a) W=2, (b) W=4, (c) W=6, (d) W=8. The approximations are periodic along the diagonal illustrating that the modeled process is cyclostationary.

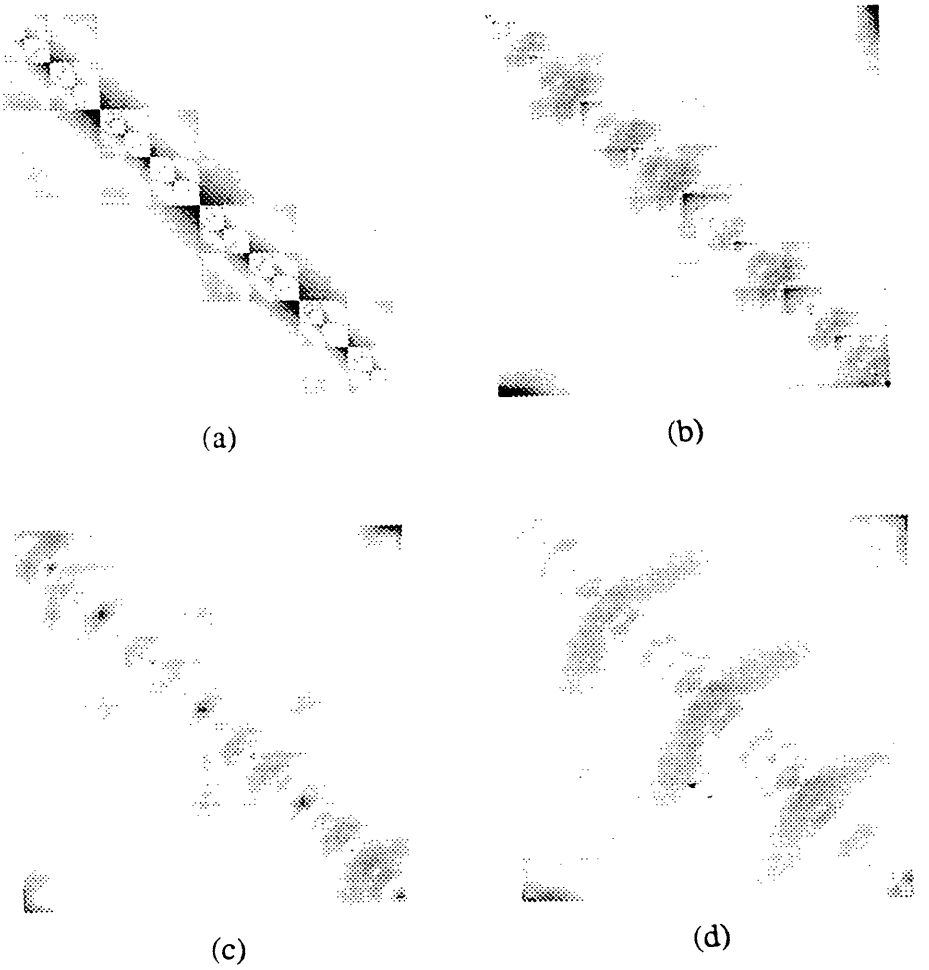


Fig. 5. 2. Plots of E_M, the errors of the approximations of the first-order Gauss-Markov process, when using QMFs that have support equaling W, where: (a) W=2, (b) W=4, (c) W=6, (d) W=8. The errors are periodic along the diagonals.

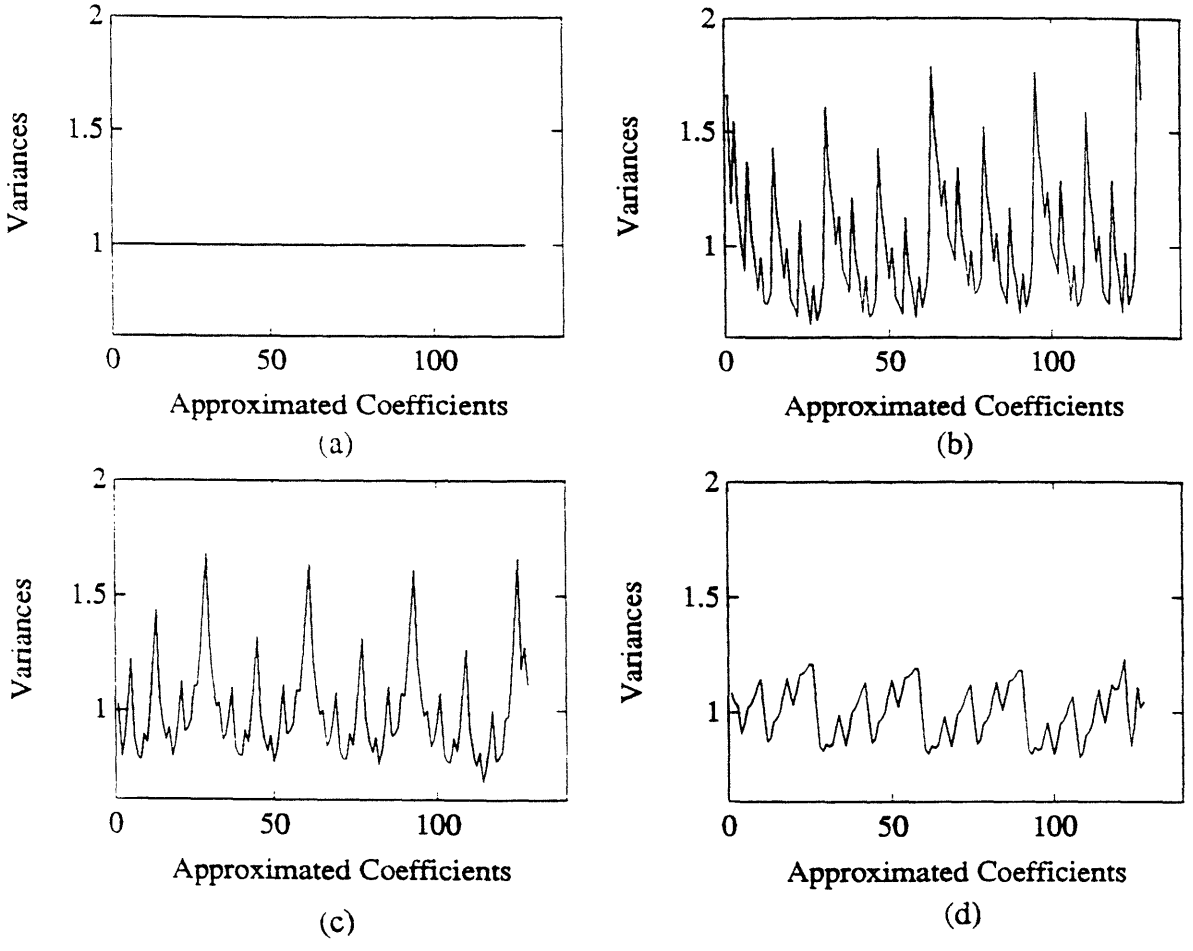


Fig. 5. 3. Plots of the variances of the approximated coefficients of the first-order Gauss-Markov process, when using QMFs that have support equaling W, where: (a) W=2, (b) W=4, (c) W=6, (d) W=8. The plots show that the variances are periodic and, in some cases, the period is even smaller then 2^L.

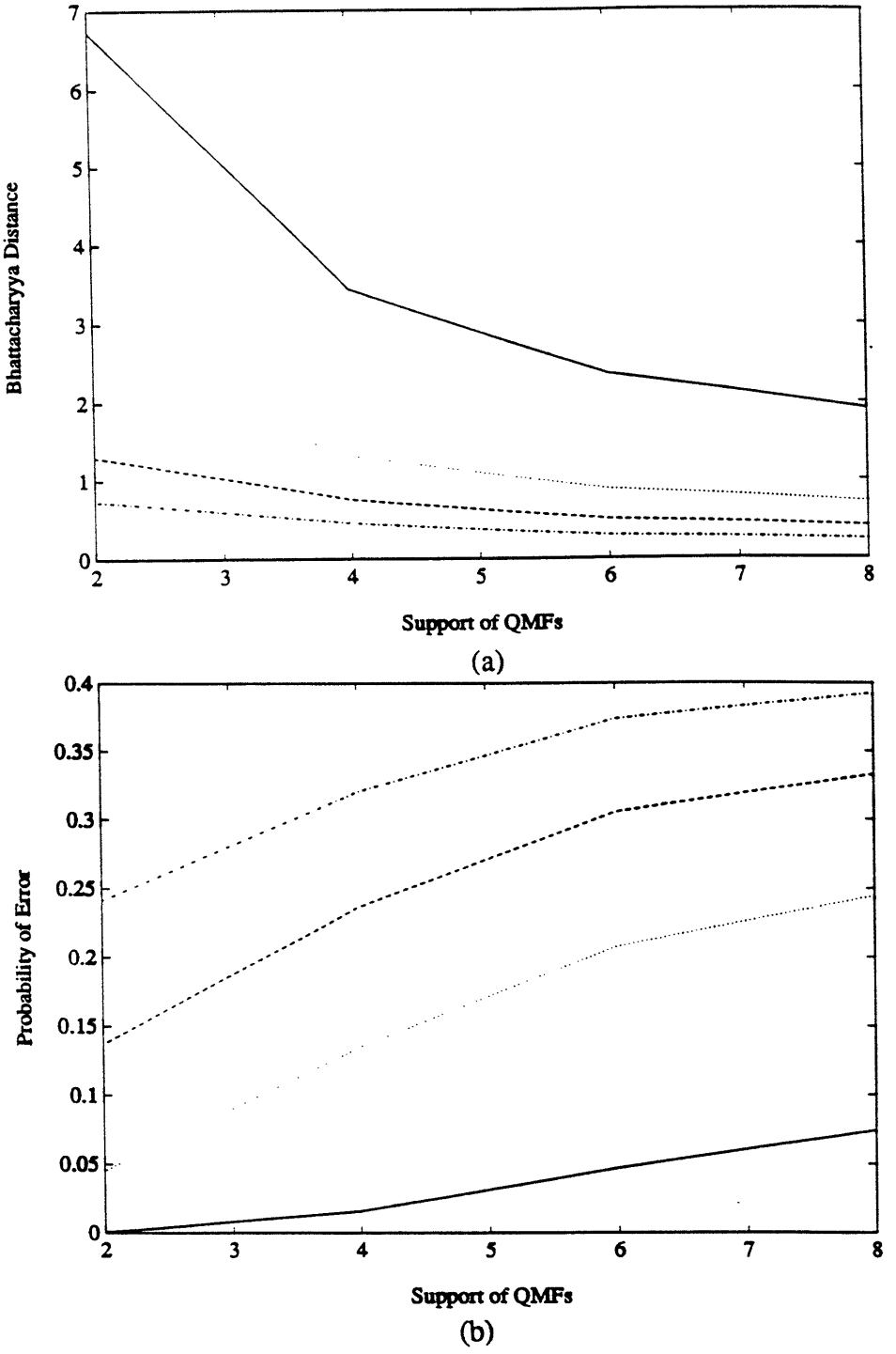


Fig. 5. 4. (a) Plot of the Bhattacharyya distance between the first-order Gauss-Markov process and its approximation using the wavelet transform for SNR values of infinity (—), five (···), two (---), and one (-·-·-). (b) Plot of the upper bound on the probability of error of deciding whether a sample path originated from the first-order Gauss-Markov process or its approximation for SNR values of infinity (—), five (···), two (---), and one (-·-·-). The plots quantify how the approximation improves as the support of the QMFs increase.

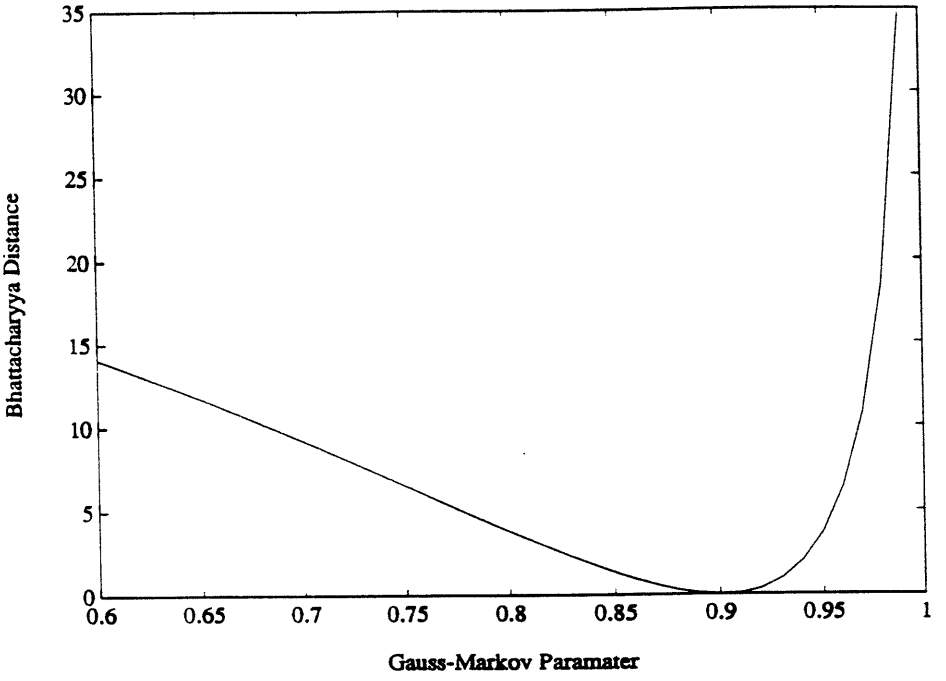
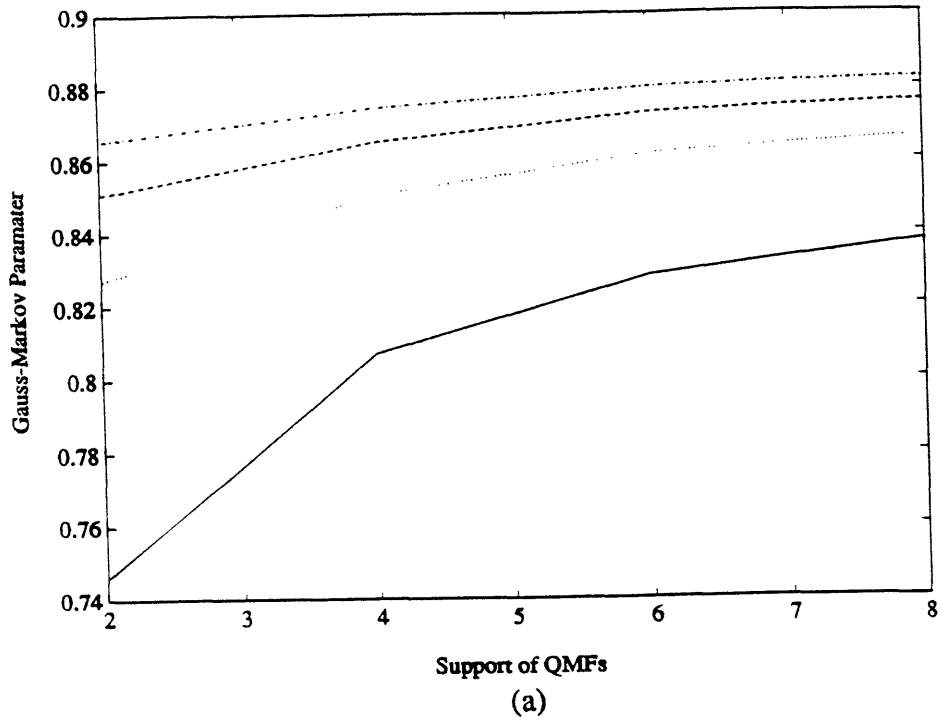
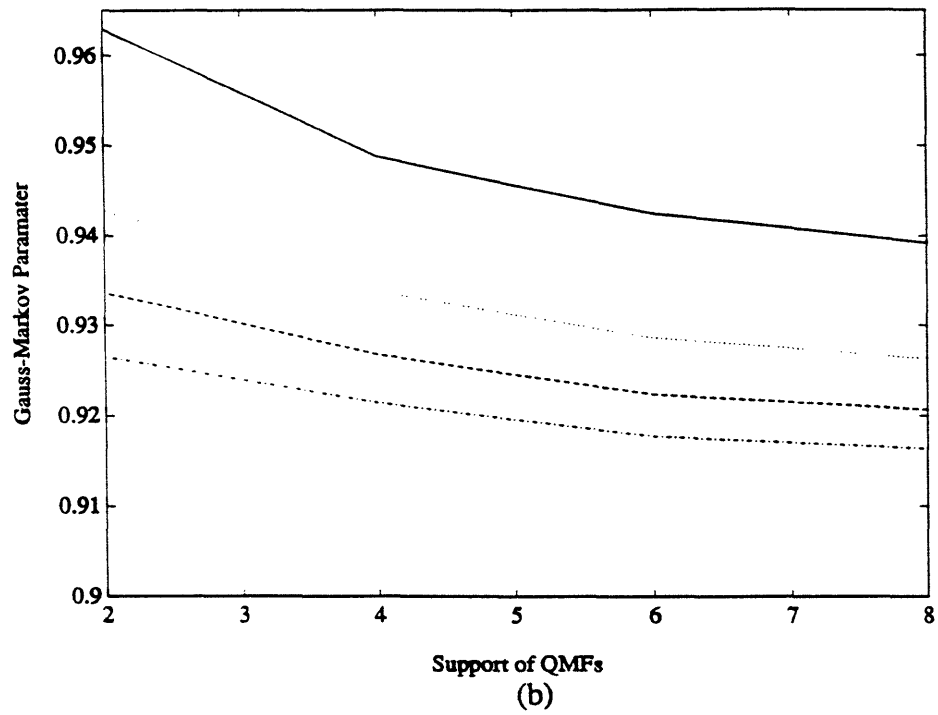


Fig. 5. 5. The x-axis shows the parameter for the Gauss-Markov process that is at the distance (as shown on the y-axis) from the Gauss-Markov process with parameter $\alpha = 0.9006$.



(a)



(b)

Fig. 5. 6. These plots show the Bhattacharyya distance, after undergoing the transformation shown in Fig. 5. 5, between the first-order Gauss-Markov process and its approximation for SNR values of infinity (—), five (···), two (---), and one (-·-·-). We used the portion of the plot shown in Fig. 5. 5 for $\alpha \leq 0.9006$ to obtain plot (a) and the portion of the plot for $\alpha > 0.9006$ to obtain plot (b).

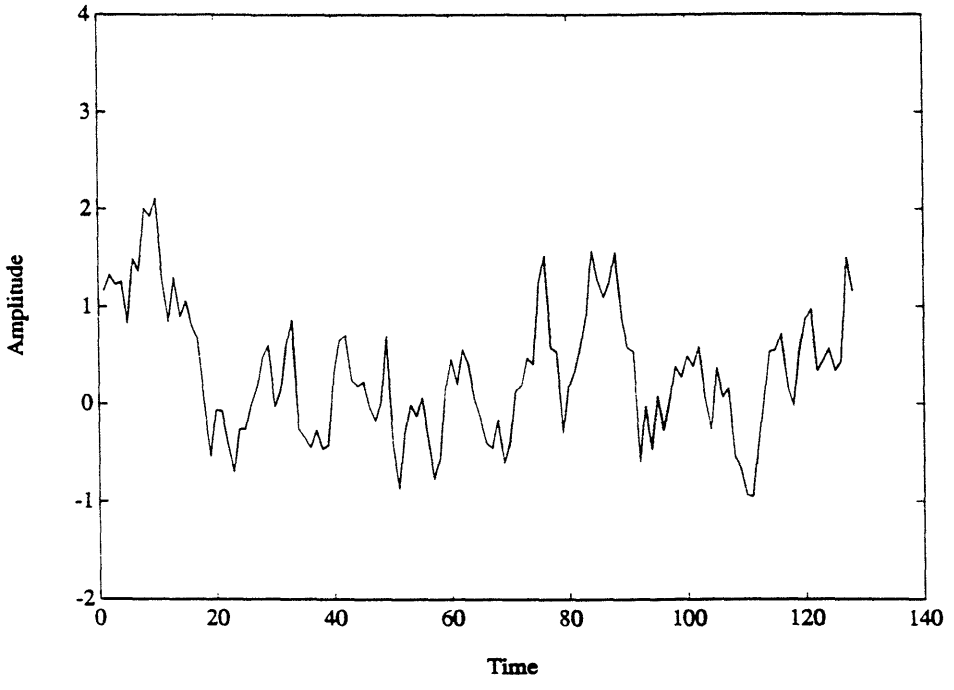


Fig. 5. 7. Plot of a sample path of the first-order Gauss-Markov process.

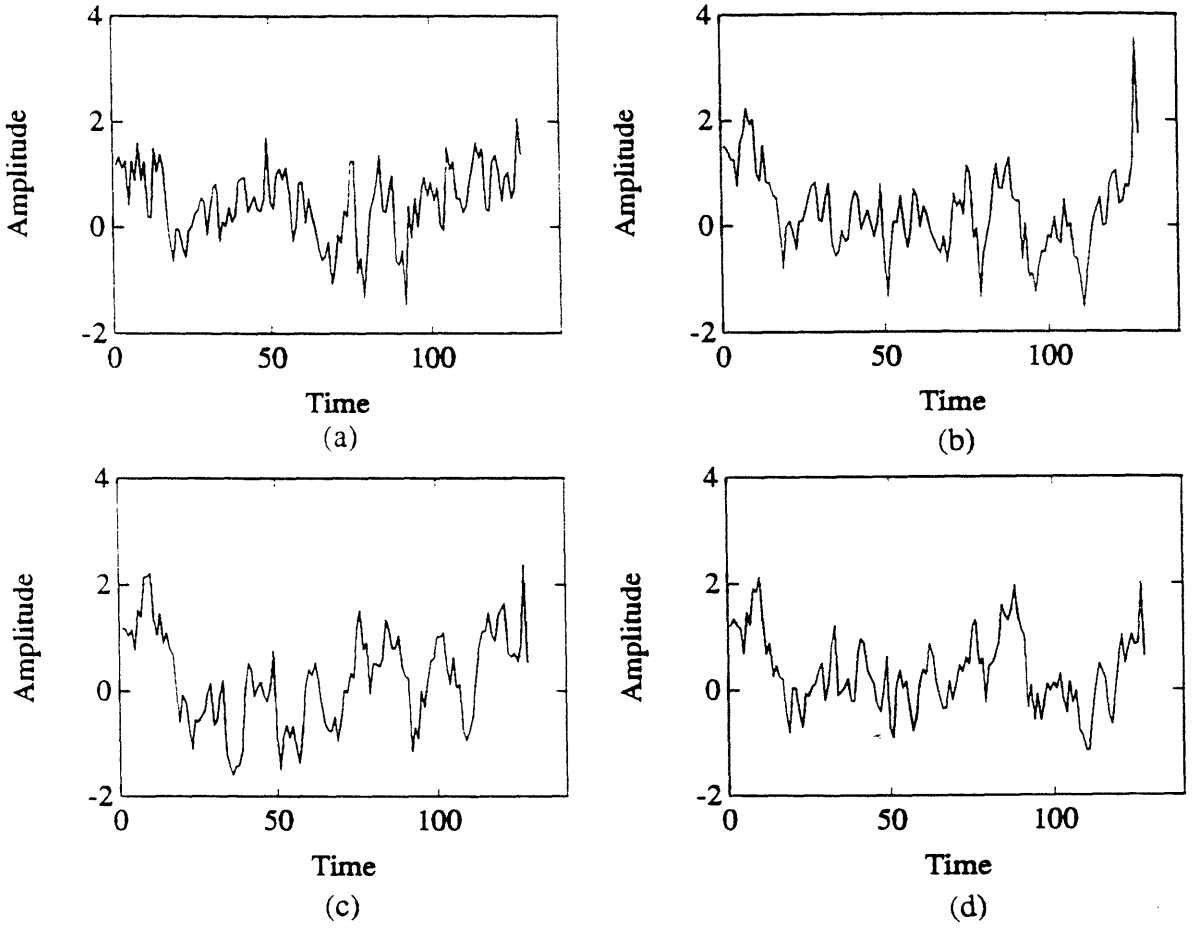


Fig. 5. 8. Plot of sample paths of the approximations of the first-order Gauss-Markov process when using QMFs that have support equaling W, where: (a) W=2, (b) W=4, (c) W=6, (d) W=8. The plots show that the sample paths are becoming more similar to the Gauss-Markov process as the support of the QMFs increase.

5.3 MODELING FRACTIONAL BROWNIAN MOTIONS

In this section we examine modeling a particular fBm, ordinary Brownian motion, by using the wavelet transform. That is we consider modeling the process described by Equation (4. 42) when H=0.5.

The 2-D image of the autocovariance of samples of ordinary Brownian motion is shown in Fig. 4. 25. We compare this image with the 2-D images of A_M in Fig. 5. 9, where A_M is the autocovariance of the approximation of samples of ordinary Brownian motion that are obtained when using wavelet transforms that have QMFs with support ranging from two to eight-taps. The level of the wavelet transform, L, that is being used is obtained from Equation (5. 18). The errors, E_M, are shown in Fig. 5. 10.

We again use the Bhattacharyya distance as a means to quantify how close the autocovariance of our approximation of Brownian motion, A_M, is to the autocovariance of the true Brownian motion process. In Fig. 5. 11 (a), we show the Bhattacharyya distance when using an SNR of infinity, five, two, and one. An upper bound on the probability of error of deciding whether a sample path originated from the Brownian motion process or the approximation of the Brownian motion process is shown in Fig. 5. 11 (b) using the same SNR values.

From these plots we observe that there is a significant difference between the approximation and the true process. Comparing the plot of the Bhattacharyya distance for Brownian motion Fig. 5. 11 (a) to the Bhattacharyya distance for the Gauss-Markov process Fig. 5. 4 (a), we observe that the distances between Brownian motion and its approximations are greater than twice the distances that were obtained between the first-order Gauss Markov process and its approximations. The fact that these distances have more than doubled imply that the probability of error for Brownian motion is less than the square of the probability of error that was obtained for the Gauss-Markov process (cf.

Fig. 5. 4 (b) and Fig. 5. 11 (b)). Therefore, we come to the conclusion that the fingers of the transformed autocovariance should not be neglected in modeling ordinary Brownian motion.

Although the technique we have described does a poor job at modeling fBm's, we believe that certain processes (that may be more representative of real phenomena) are well-modeled by using the wavelet transform. Specifically, Wornell [18] considers a related class of models that generate "almost" 1/f processes. A 1/f process is defined as a process whose empirical power spectrum S(w) is proportional to

$$S(w) \sim \frac{1}{|w|^{\gamma}} \tag{5. 20}$$

where γ is some parameter in the range $0 < \gamma < 2$. Note that these processes are not stationary since the process has infinite variance. For the range $1 < \gamma < 2$, fBm's are often thought of as 1/f - like processes, where the parameter H is

$$H = \frac{\gamma - 1}{2} \tag{5. 21}$$

However, fBm's have the undesirable and unrealistic property of having growing variances with time. Consequently it is of considerable interest to develop models with less drastic nonstationary features. In [18], Wornell develops models of almost 1/f processes by assuming that the coefficients in the wavelet transform domain are uncorrelated. The wavelet transform coefficients U_m are modeled as being stationary at every level with the following variance

$$\mathrm{Var}\{U_{m-1}\} = 2^{\gamma}\mathrm{Var}\{U_m\} \tag{5. 22}$$

Note that this model is exactly of the type that we have described in this thesis. Specifically, given the autocovariance of a 1/f process, the method that we have described

will find both the optimal wavelet transform and the correct value of the variances of the corresponding wavelet coefficients. For the processes described in [18] the coefficients would be exactly decorrelated and the coefficients would have variances satisfying Equation (5. 22).

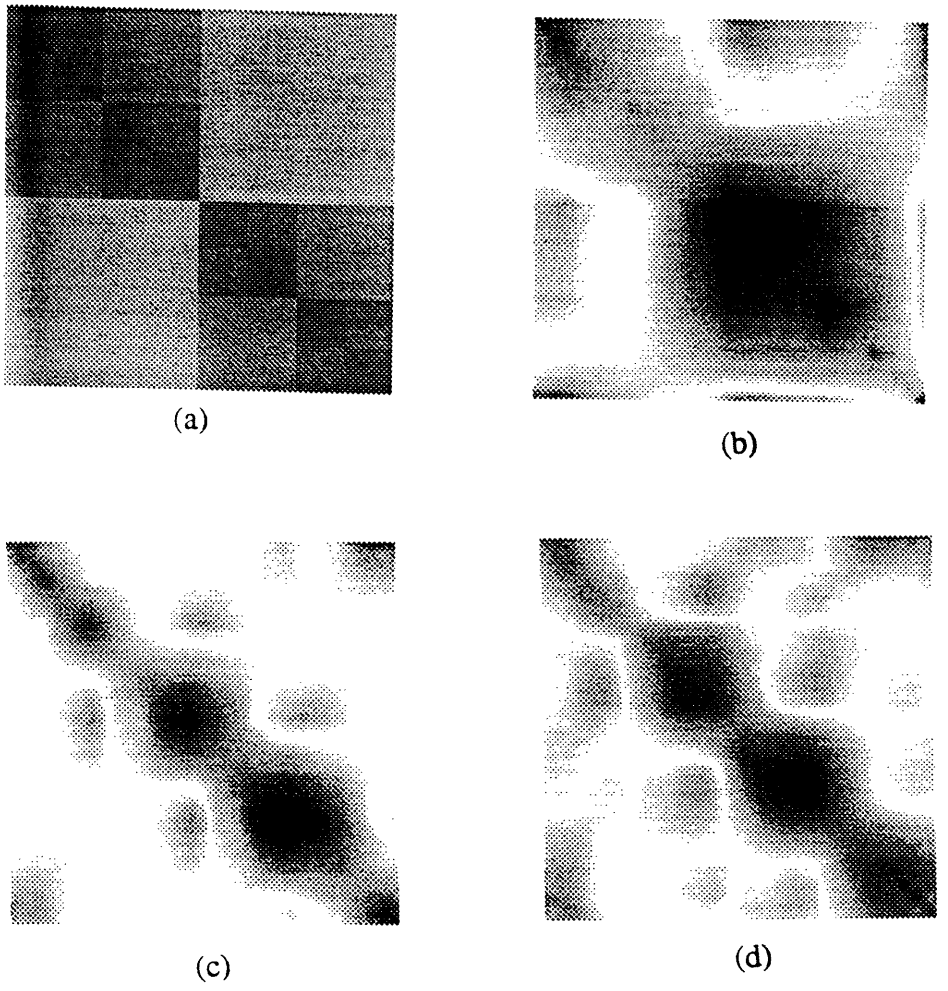


Fig. 5. 9. Plots of A_M, the autocovariances of the approximations of samples of Brownian motion, when using QMFs that have support equaling W, where: (a) W=2, (b) W=4, (c) W=6, (d) W=8. The plots show that approximations do a poor job of modeling ordinary Brownian motion.

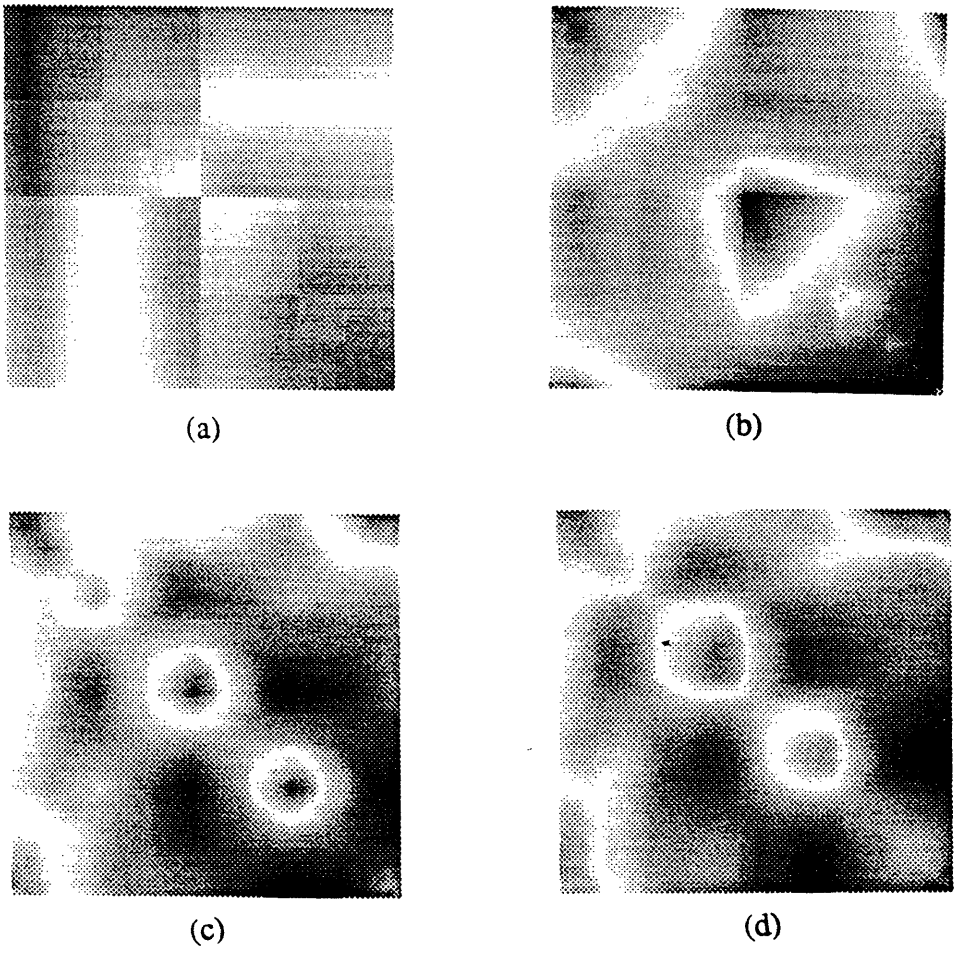


Fig. 5. 10. Plots of E_M, the errors of the approximations of samples of Brownian motion, when using QMFs that have support equaling W, where: (a) W=2, (b) W=4, (c) W=6, (d) W=8.

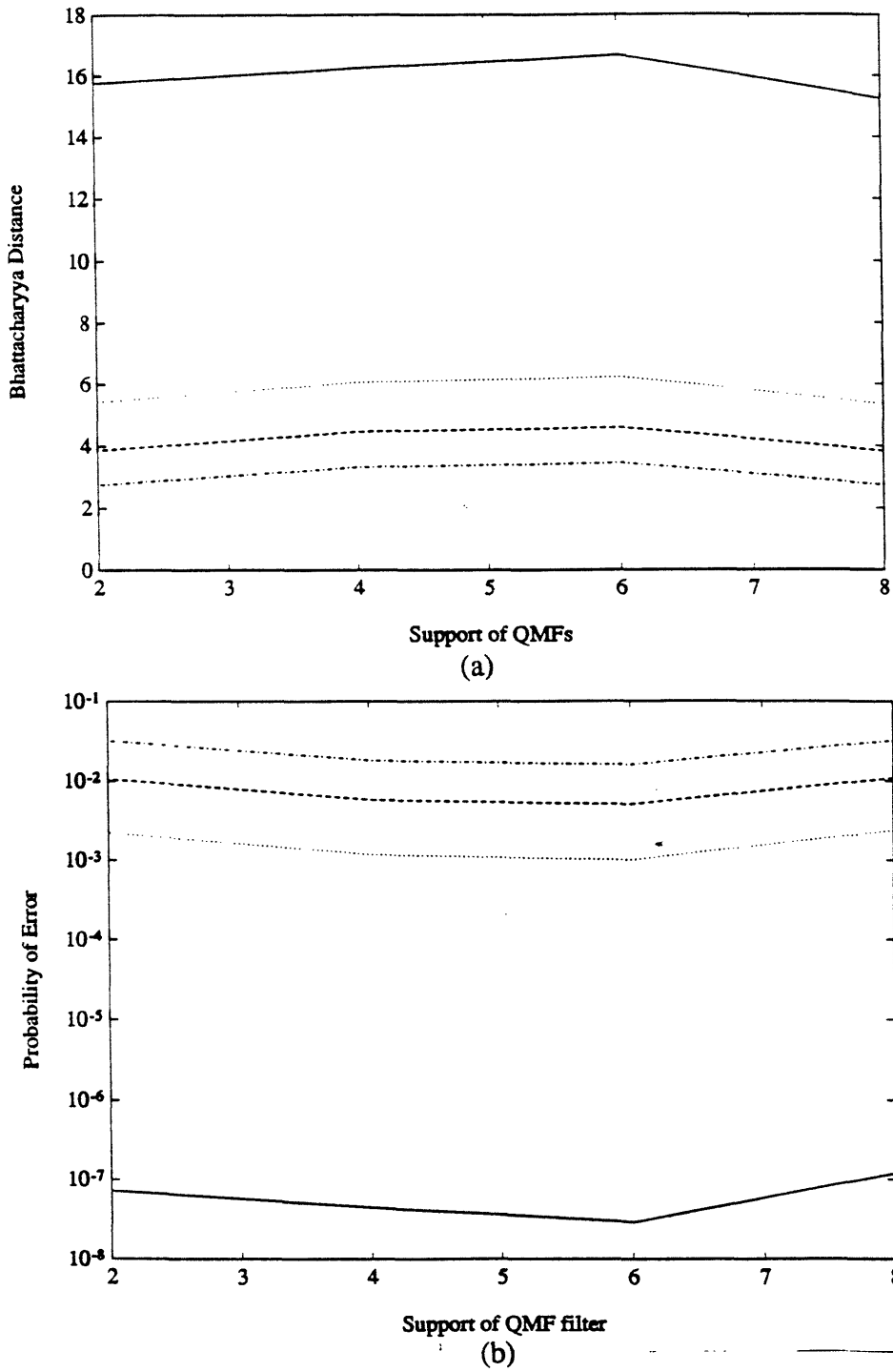


Fig. 5. 11. (a) Plot of the Bhattacharyya distance between the samples of Brownian motion and its approximation using the wavelet transform for SNR values of infinity (—), five (···), two (---), and one (-·-·-). (b) Plot of the upper bound on the probability of error of deciding whether a sample path originated from the samples of Brownian motion or its approximation for SNR values of infinity (—), five (···), two (---), and one (-·-·-). The plots show that the approximations are not doing a good job of modeling Brownian motion.

CHAPTER 6

CONCLUSION

In this thesis we have investigated the use of the wavelet transform for the scale-to-scale decorrelation of stochastic processes. We have seen that we can determine the optimal wavelet transform in a level-by-level procedure. Specifically, at each level we obtain the QMFs that minimizes the correlation between the coarse approximation and the wavelet coefficients at that level. We obtain these QMFs by using a paramaterization that stays within the set of perfect reconstruction filters. The construction of the paramaterization led us to a natural way of performing the minimization. Specifically, we solved the problem using QMFs with smaller support to allow these solutions to guide us in solving the more complex problems that use QMFs with larger support.

We investigated the ability, in general, for the wavelet transform to achieve a Karhunen-Loeve expansion by considering several classes of processes. The first such class of processes that we examined were stationary processes. We examined the statistical properties of the transformed coefficients when the process being analyzed was stationary. We saw that the covariance between the coarse approximation and the wavelet coefficients at any particular level were also stationary. Although the covariance of the wavelet coefficients at different levels were not simply a function of the time-difference, the covariance did have a very special structure. Specifically, the cross-covariance is only a function of the time-difference after the difference in sampling rates are taken into consideration. This special structure leads us to believe that efficient algorithms can be created that will model processes without assuming that the "fingers" in the transformed autocovariance are zero.

We then examined two particular classes of stationary processes: the first-order

Gauss-Markov process and a second-order under-damped process. By using images to depict the transformed autocovariances of these processes we showed that the wavelet transform is more appropriate for modeling the first-order Gauss-Markov process than the second-order under-damped process. The reason for this was that the first-order Gauss-Markov process is an example of a process whose power spectrum is concentrated at low-frequency. The wavelet transform does a good job of modeling processes of this type because the transform "zooms" into the low-frequency components of the signal. We have also seen that the wave packet transform is more appropriate for the second-order under-damped process since the wave packet transform can be chosen to "zoom" into the desired frequency range.

We also examined using the wavelet transform to model a class of nonstationary processes known as fractional Brownian motions. The autocovariance of the Haar transform of ordinary Brownian motion has a very special structure. The transformed autocovariance is comprised of "fingers" where the values between the "fingers" is exactly zero. When we used QMFs that have support greater than two, we saw that edge effects (due to windowing and cyclic convolution) were quite substantial in the transformed autocovariance. The edge effects were more prominent for Brownian motion than the first-order Gauss-Markov process because the variances of Brownian motion increase over time. Therefore, due to the fact that we are using circular convolution, the coefficients with the largest and smallest variances will exhibit a substantial degree of correlation.

Finally, we considered approximating processes by assuming that the transformed coefficients were indeed white. We compared the approximation of processes to the original processes by showing images of the approximated autocovariances. We used the Bhattacharyya distance to quantify the ability of being able to use the wavelet transform to approximate processes. Our results showed that the first-order Gauss-Markov process is well-approximated by assuming that the transformed coefficients are white; however, we

also observed that when modeling Brownian motion with the wavelet transform that the "fingers" can not be neglected, although there are other, more realistic models for 1/f processes where the "fingers" are either negligible or identically zero.

The Bhattacharyya distance can also be used to determine the support of the QMF that will be used in the transform. Specifically, when given a design constraint (such as the desire that the approximation is to be within a certain distance of the true process) the figures we have presented can be used to determine the support of the QMF that should be used to achieve this desired performance level.

When we compared our results of determining the appropriate QMFs with the QMFs that was derived by Daubechies, we were surprised to find out that the QMFs derived by Daubechies yielded results (on several occasions) that were superior to ours. We attributed this problem to two difficulties with our algorithm. First, minimizing the entire cross-covariance matrix between the coarse approximation and the wavelet coefficients at a particular level implies that attention is being paid to terms distorted by edge effects. Since other methods (such as windowing) can be used to minimize the edge effects, we might alternatively consider minimizing the norm of that portion of the cross-covariance matrix that is not subject to edge effects. Secondly, the efficient method that we devised to search the local minima for the global minimum results in a solution that is a local minimum (typically a very good one) but may not be the desired global minimum.

In order to improve our results, we investigated improving the performance of our algorithm by making two modifications to correct for these difficulties. The first modification is to minimize the norm of the portion of the cross-covariance $\Lambda_{X_m U_m}$ that does not include the elements of the matrix that exhibit edge effects. Specifically, this cross-covariance $\Lambda_{X_m U_m}$ will have W/2 - 1 rows and W/2 - 1 columns that do exhibit edge effects, where W is the support of the QMFs. Depending upon the index of the first

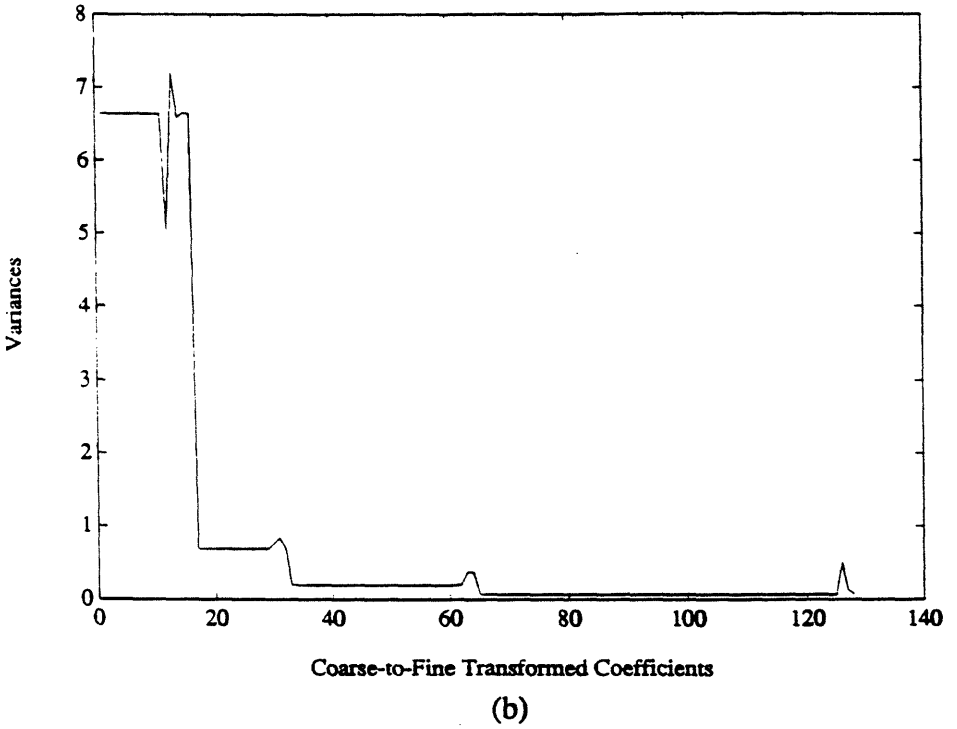
coefficient of the signal being analyzed, these W/2 - 1 rows (or columns) may be either the first or the last rows (or columns) of the cross-covariance or a combination of both. Wherever these rows (or columns) are located, the optimization procedure is improved by ignoring them in the minimization. As we have seen, the rows that exhibit edge effects are typically much more prominent than the columns since the rows that exhibit edge effects are due to cyclic wrap-around of X_m while the columns that exhibit edge effects are due to cyclic wrap-around of U_m. That is consider the elements $\Lambda_{X_m U_m}(s, t)$ of the matrix $\Lambda_{X_m U_m}$ when the coefficient u[m, t] does not exhibit edge effects and the coefficient x[m, s] does exhibit edge effects, in this case, the row of the matrix $\Lambda_{X_m U_m}$ that corresponds to x[m, s] will be corrupted by edge effects. The second modification would be to use a better search technique to locate the global minimum: that is make a better choice for the initial guess to obtain smaller local minima. In some further tests we used the QMFs derived by Daubechies as the initial guesses for our minimization. These experiments led to superior decorrelation performance indicating, in particular, that the QMFs derived by Daubechies are not local minima of our criterion. Thus we can obtain QMFs that will do at least as well as our previous solution and better than the QMFs derived by Daubechies.

In certain applications the efficiency of the optimization is not as important as finding the true global minimum. We obtain the true global minimum of the norm of the cross-covariance $\Lambda_{X_{M-1} U_{M-1}}$, where X_M is the first-order Gauss-Markov process described in Section 4.2, by comparing the local minima that result from many initial guesses. Our set of initial guesses are obtained by simply taking all possible combinations of a uniform sampling of each unknown parameter. In Fig. 6. 1 (a), we show the normalized autocovariance of the three-level transformed coefficients of the first-order Gauss-Markov process using the eight-tap QMFs that correspond to the global minimum. The variances of

these transformed coefficients are shown in Fig. 6. 1 (b). The normalized autocovariance that we show in Fig. 6. 1 can be compared to the normalized autocovariance shown in Fig. 4. 12, where the autocovariance of the same transformed coefficients are shown but when using QMFs derived by Daubechies. In Fig. 6. 2, we show a comparison between Fig. 6. 1 and Fig. 4. 12 by showing the correlation structure of the 100th transformed coefficient. We also compare the particular values of h[n] that we obtained in this experiment to the one derived by Daubechies by plotting the impulse responses obtained from both of these methods in the time domain Fig. 6. 3 (a) and in the frequency domain Fig. 6. 3 (b).

From these plots we observe that our solution yields better results than the QMFs derived by Daubechies for the first-order Gauss-Markov process. Since our algorithm is based upon the particular signal's characteristics, we believe that our algorithm can obtain even more dramatic results when applied to other processes.

(a)



(b)

Fig. 6. 1. Plot (a) is the normalized autocovariance of the three-level wavelet transform of the first-order Gauss-Markov process using eight-tap QMFs obtained by modifying the optimization algorithm presented in Chapter 3. Plot (b) shows the variances of the transformed coefficients.

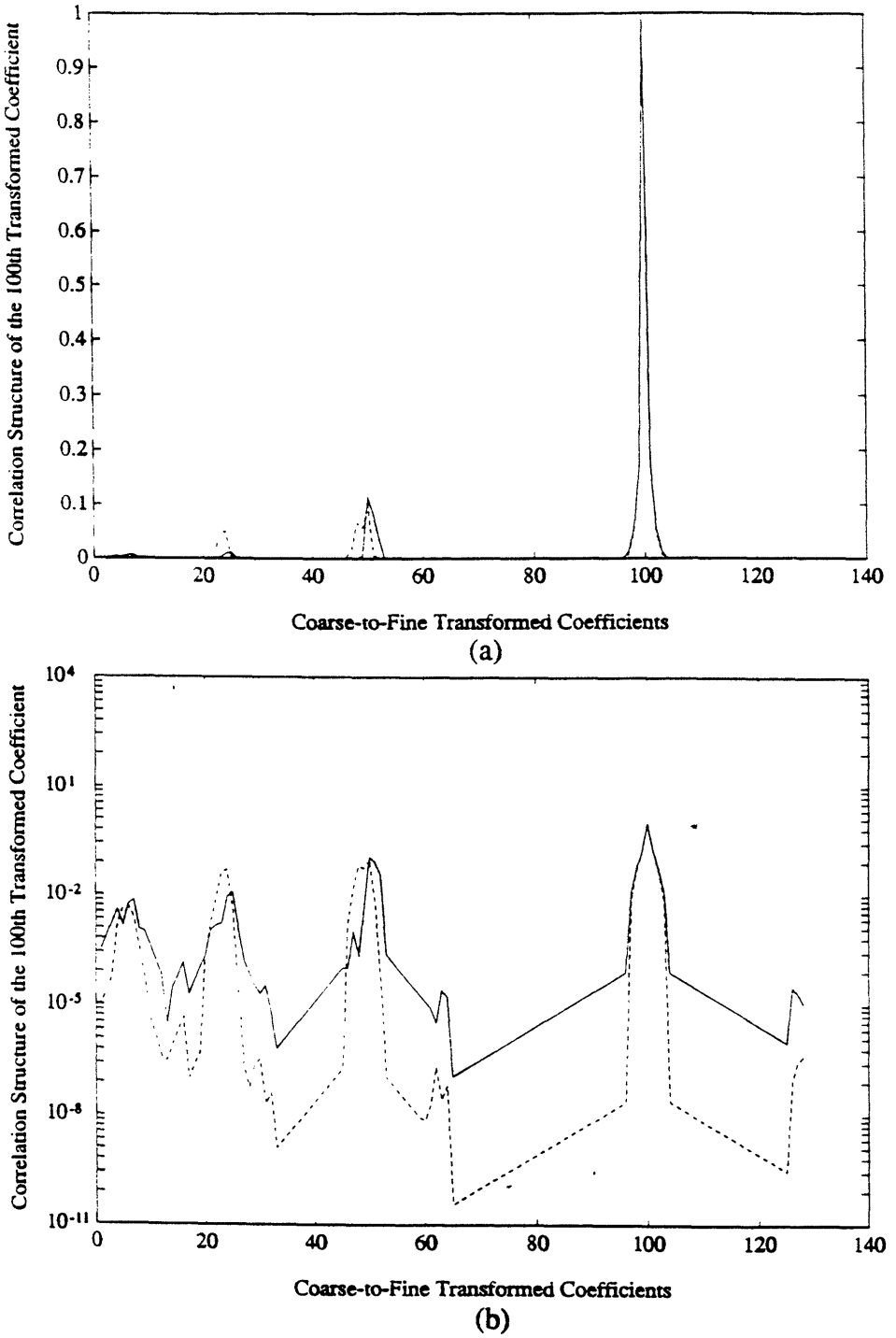


Fig. 6. 2. Cross-sections of the normalized autocovariance of the one-level wavelet transform of the first-order Gauss-Markov process using eight-tap QMFs derived by modifying the optimization algorithm presented in Chapter 3 (solid line) and derived by Daubechies (dashed line). The correlation (y-axis) is plotted on: (a) linear scale, (b) logarithmic scale.

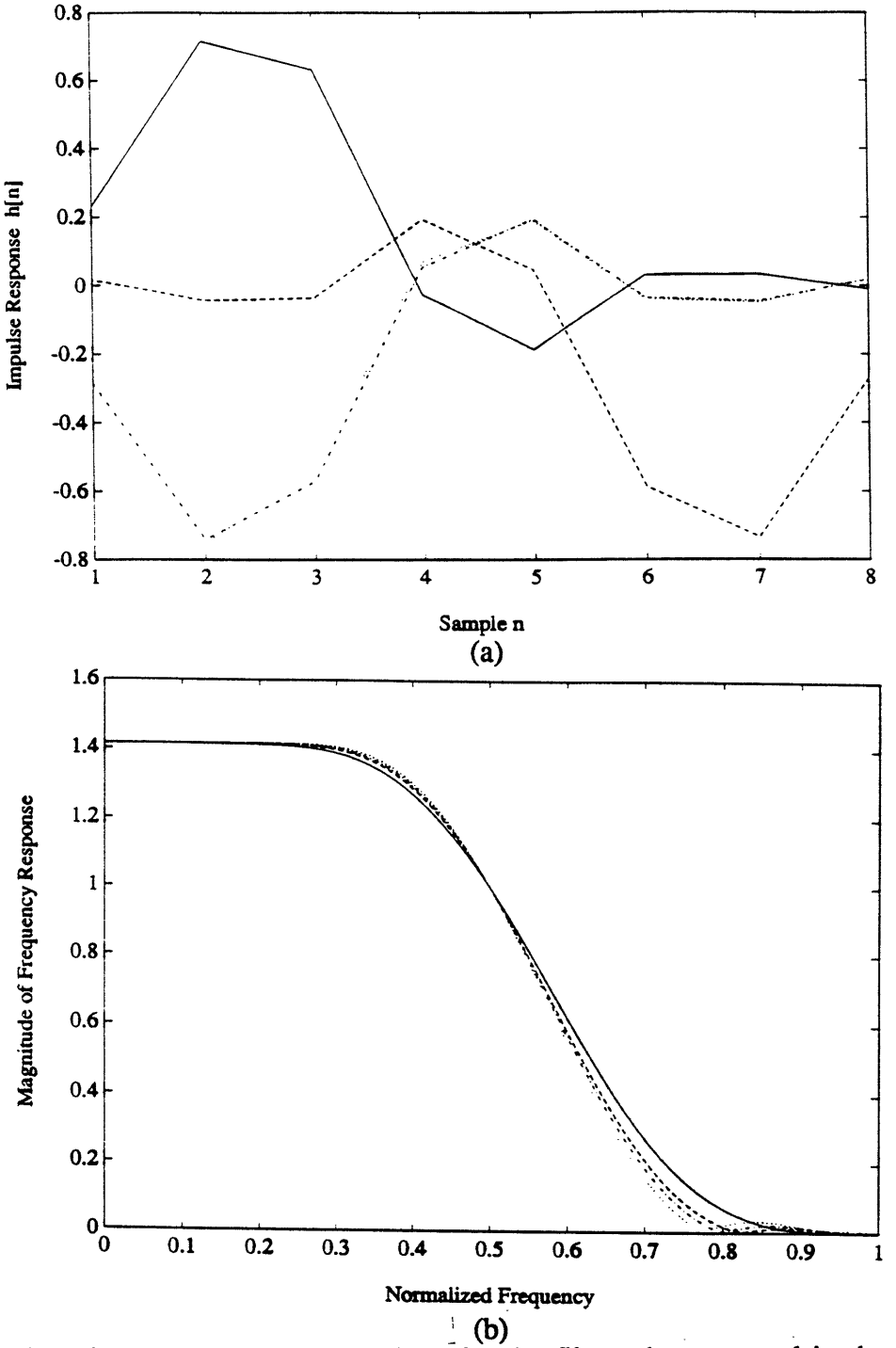


Fig. 6. 3. Plot of the impulse responses h[n] for the filters that are used in the optimal three-level wavelet transform of the first-order Gauss-Markov process, where optimality is defined by the modified optimization algorithm of Chapter 3. The optimal filters are represented by: (---) for the first level, (-·-·-) for the second level, and (···) for the third level. These filters are compared to the eight-tap filter derived by Daubechies (—). The impulse responses are shown in the (a) time domain and (b) frequency domain.

APPENDIX A

PERIODICITY PROPERTIES OF THE OBJECTIVE FUNCTION

The purpose of this Appendix is to prove two statements about the periodicity of the objective function from Equation (3. 14):

$$f(\theta) = \left|\left|H(\theta)\, \Lambda_X\, G^T(\theta)\right|\right|_F^2 \qquad \text{(A. 1)}$$

Recall that $\theta = \{\theta_1, \theta_2, \ldots, \theta_L\} \in \mathbb{R}^L$, and let $\mathbf{e}_j \in \mathbb{R}^L$ be the sequence composed of all zeros except for a one in the jth position. For example $\theta + \pi\mathbf{e}_j$ would be expressed as $\theta + \pi\mathbf{e}_j = \{\theta_1, \theta_2, \cdots, \theta_j+\pi, \cdots, \theta_L\}$.

We will show the following two properties regarding $f(\theta)$:

(1) That $f(\theta)$ has a periodicity of π in the θ_j variable for any j.

$$f(\theta) = f(\theta + \pi\mathbf{e}_j) \qquad \text{(A. 2)}$$

(2) That $f(\theta)$ has a periodicity of $\pi/2$ in the θ_L variable.

$$f(\theta) = f(\theta + \pi/2\mathbf{e}_L) \qquad \text{(A. 3)}$$

To prove (1) we first show the following equations are true using induction in L.

$$h_{\theta}^{L}[n] = - h_{\theta}^{L} + \pi e_j[n] \tag{A. 4}$$

$$g_{\theta}^{L}[n] = - g_{\theta}^{L} + \pi e_j[n] \tag{A. 5}$$

Recall that the superscript on the filters are used to denote the number of elements in θ. To prove the above statements by induction we must first prove them for the case L=1.

For the case L=1 we can express $\theta + \pi e_j$ as simply $\theta_1 + \pi$. We prove (A. 4) and (A. 5) for L=1 by simply replacing $\theta_1 + \pi$ for θ_1 in Equations (3. 15) and (3. 16). Using well-known trigonometric identities it is trivial to show that the L=1 case is satisfied.

To prove the general case we will assume:

$$h_{\theta}^{L-1}[n] = - h_{\theta}^{L-1} + \pi e_j[n] \tag{A. 6}$$

$$g_{\theta}^{L-1}[n] = - g_{\theta}^{L-1} + \pi e_j[n] \tag{A. 7}$$

and recall the recursive equations of (2. 29) and (2. 30):

$$h_{\theta}^{L}[n] = \cos \theta_L \, h_{\theta}^{L-1}[n] - \sin \theta_L \, g_{\theta}^{L-1}[n-2] \tag{A. 8}$$

$$g_{\theta}^{L}[n] = \sin \theta_L \, h_{\theta}^{L-1}[n] + \cos \theta_L \, g_{\theta}^{L-1}[n-2] \tag{A. 9}$$

In order to prove Equations (A. 4) and (A. 5) in the general case we substitute $\theta + \pi e_j$ for

θ in Equations (A. 8) and (A. 9). If we consider $j \leq L-1$ we verify Equations (A. 4) and (A. 5) by using the assumptions (A. 6) and (A. 7). If j=L then we verify Equations (A. 4) and (A. 5) directly from the recursive equations.

Relying on the fact that $H(\theta)$ and $G(\theta)$ are composed of the elements of the filters $h_\theta^L[n]$ and $g_\theta^L[n]$ respectively, we note that (A. 4) and (A. 5) imply:

$$H(\theta) = - H(\theta + \pi e_j) \tag{A. 10}$$

$$G(\theta) = - G(\theta + \pi e_j) \tag{A. 11}$$

Thus using (A. 10) and (A. 11) we infer:

$$H(\theta)\Lambda_X G^T(\theta) = H(\theta + \pi e_j)\Lambda_X G^T(\theta + \pi e_j) \tag{A. 12}$$

Equation (A. 12) proves Equation (A. 2) showing that $f(\theta)$ is periodic by π for any θ_j variable.

To prove Equation (A. 3) we will substitute $\theta + \pi/2e_L$ for θ in Equations (A. 8) and (A. 9) to yield the following results:

$$h_{\theta + \pi/2e_L}^L[n] = - g_\theta^L[n] \tag{A. 13}$$

$$g_{\theta + \pi/2e_L}^L[n] = h_\theta^L[n] \tag{A. 14}$$

Equations (A. 13) and (A. 14) imply:

$$H(\theta + \pi/2e_L)\Lambda_X G^T(\theta + \pi/2e_L) = - G(\theta)\Lambda_X H^T(\theta) \qquad \text{(A. 15)}$$

We make use of the facts that Λ_X is symmetric and the norm of a matrix is equal to the norm of its negative transpose to conclude that the function that we are minimizing is periodic with period $\pi/2$ in the θ_L variable.

APPENDIX B

OPTIMAL TWO-TAP QMF FILTERS FOR STATIONARY PROCESSES

The purpose of this appendix is to show that for the class of stationary processes the optimal two-tap QMF filters are the filters used in the Haar Transform. Optimality is defined by the criterion of Chapter 3, specifically to minimize the following function over all two-tap QMF filters:

$$f(\theta) = \left|\left| H_{m-1}(\theta)\, \Lambda_{X_m X_m}\, G^{T}_{m-1}(\theta) \right|\right|^2_F \qquad \text{(B. 1)}$$

To begin let us write $f(\theta)$ as a sum of squares of the elements of $\Lambda_{X_{m-1}U_{m-1}}$:

$$f(\theta) = \sum_{i=1}^{N} \sum_{j=1}^{N} \Lambda^2_{X_{m-1}U_{m-1}}(i-j\,;\,\theta) \qquad \text{(B. 2)}$$

where the second index is introduced to denote explicitly that $\Lambda_{X_{m-1}U_{m-1}}$ is a function of θ. Exploiting the Toeplitz structure of the matrix $\Lambda_{X_{m-1}U_{m-1}}$, we rewrite Equation (B. 2) in the following manner:

$$f(\theta) = N\Lambda^2_{X_{m-1}U_{m-1}}(0\,;\,\theta) + \sum_{\tau=1}^{N-1} (N-\tau)\left[\Lambda^2_{X_{m-1}U_{m-1}}(\tau\,;\,\theta) + \Lambda^2_{X_{m-1}U_{m-1}}(-\tau\,;\,\theta)\right] \qquad \text{(B. 3)}$$

By substituting the parametrization of all two-tap QMF filters of Equations (2. 31) and (2. 32) into Equation (4. 11), we can describe $\Lambda_{X_{m-1}U_{m-1}}(\tau\,;\,\theta)$ as a function of θ:

$$\Lambda_{X_{m-1}U_{m-1}}(\tau\,;\,\theta) = \cos^2\theta\ \Lambda_{X_mX_m}(2\tau\text{-}1) - \sin^2\theta\ \Lambda_{X_mX_m}(2\tau\text{+}1) \tag{B. 4}$$

For $\tau = 0$, Equation (B. 4) becomes:

$$\Lambda_{X_{m-1}U_{m-1}}(0\,;\,\theta) = \cos 2\theta\ \Lambda_{X_mX_m}(1) \tag{B. 5}$$

For $\tau \geq 1$ the following equation holds:

$$\Lambda^2_{X_{m-1}U_{m-1}}(\tau\,;\,\theta) + \Lambda^2_{X_{m-1}U_{m-1}}(-\tau\,;\,\theta) = \left(\cos^4\theta + \sin^4\theta\right)\left(\Lambda^2_{X_mX_m}(2\tau\text{+}1) + \Lambda^2_{X_mX_m}(2\tau\text{-}1)\right)$$

$$- 4\sin^2\theta\cos^2\theta\ \Lambda_{X_mX_m}(2\tau\text{+}1)\,\Lambda_{X_mX_m}(2\tau\text{-}1) \tag{B. 6}$$

For notational simplicity, we define the following constants that are not a function of θ:

$$C_1 = \sum_{\tau=1}^{N-1} (N\text{-}\tau)\left[\Lambda^2_{X_mX_m}(2\tau\text{+}1) + \Lambda^2_{X_mX_m}(2\tau\text{-}1)\right] \tag{B. 7}$$

$$C_2 = 2\sum_{\tau=1}^{N-1} (N\text{-}\tau)\Lambda_{X_mX_m}(2\tau\text{+}1)\Lambda_{X_mX_m}(2\tau\text{-}1) \tag{B. 8}$$

We now write $f(\theta)$ as:

$$f(\theta) = N\,\Lambda^2_{X_m X_m}(1)\cos^2 2\theta + C_1\left(1-\frac{1}{2}\sin^2 2\theta\right) - C_2\frac{1}{2}\sin^2 2\theta \qquad \text{(B. 9)}$$

In order to obtain the minimum of $f(\theta)$ we determine its derivative with respect to θ.

$$\frac{\partial f(\theta)}{\partial \theta} = -\left(2N\,\Lambda^2_{X_m X_m}(1) + C_1 + C_2\right)\sin 4\theta \qquad \text{(B. 10)}$$

Since we know from Appendix A that $f(\theta)$ has a period of $\pi/2$, the only values of θ we need to consider are values of θ between 0 and $\pi/2$. The values of θ that satisfy the first derivative test in this range are $\theta = 0$ and $\theta = \pi/4$. The second derivative test and the fact that $2N\,\Lambda^2_{X_m X_m}(1) + C_1 + C_2 \geq 0$ imply that $\theta = 0$ is a maximum and $\theta = \pi/4$ is a minimum. Recall that the filters are not periodic with period $\pi/2$, so we want to choose from the pair of values $(\pi/4, 3\pi/4)$ the value of θ such that $h[n]$ is the low-pass filter, i.e. we take $\theta = \pi/4$ as the value of θ that minimizes $f(\theta)$. When $\theta = \pi/4$ the filters are the ones used in the Haar Transform:

$$h[n] = \left\{ \begin{array}{ll} \frac{1}{\sqrt{2}} & n = 0 \\ \frac{1}{\sqrt{2}} & n = 1 \\ 0 & \text{Otherwise} \end{array} \right\} \qquad \text{(B. 11)}$$

$$g[n] = \left\{ \begin{array}{cc} -\frac{1}{\sqrt{2}} & n = 0 \\ \frac{1}{\sqrt{2}} & n = 1 \\ 0 & \text{Otherwise} \end{array} \right\} \qquad \text{(B. 12)}$$

[1] Vaidyanathan, P., Hoang, P.,"Lattice structures for optimal design and robust implementation of two-channel perfect reconstruction QMF banks," *IEEE Trans. Acoust. Speech Signal Proc.*, Vol. 36, pp. 81-94, Jan. 1988.

[2] Daubechies, I., "Orthonormal Bases of Compactly Supported Wavelets", *Communications on Pure and Applied Mathematics*, Vol. XLI 909-996 (1988).

[3] Daubechies, I., "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. on Information Theory*, Vol. 35, pp. 197-199, 1989.

[4] Daubechies, I., Grossman, A., Meyer, Y., "Painless non-orthogonal expansions," *J. Math. Phys.* 27, pp. 1271-1283, 1986.

[5] Mallat, S., "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Transactions on Pattern Anal. and Mach. Intel.*, Vol. PAMI-11, pp. 674-693, July 1989.

[6] Mallat, S., "Multifrequency Channel Decompositions of Images and Wavelet Models," *IEEE Transactions on ASSP*, Vol. 37, pp. 2091-2110, Dec. 1989.

[7] Meyer, Y., "L'analyse par ondelettes," *Pour la Science*, Sept. 1987.

[8] Vitterly, M., and Herley, C., "Wavelet and Filter Banks: Relationships and New Results," *Proceedings of the ICASSP*, Albuquerque, NM, 1990.

[9] Smith, M.J. and Barnwell, T.P., "Exact Reconstruction Techniques for Tree-Structured Sub-Band Coders", *IEEE trans. on ASSP*, vol. 34, pp. 434-441, Jun 1986.

[10] Beylkin, G., Coifman, R., Rokhlin, V. "Fast Wavelet Transforms and Numerical Algorithms I," to appear in *Comm. Pure and Appl. Math.*

[11] Coifman, R., Meyer, Y., Quake, S., Wickerhauser, M., "Signal Processing and Compression with Wave Packets," preprint, April 1990.

[12] Coifman, R., Wickerhauser, "Best-adapted wave packet bases," Yale U., Feb. 1990 (To be published).

[13] Barnsley, M., *Fractals Everywhere*, Academic Press, San Diego, CA, 1988.

[14] Mandelbrot, B., *The Fractal Geometry of Nature*, Freeman, New York, 1982.

[15] Mandelbrot, B., Van Ness, H., "Fractional Brownian Motions, Fractional Noises and Applications," *SIAM Review*, Vol. 10,pp. 422-436, October 1968.

[16] Pentland, A., "Fractal-Based Description of Natural Scenes," *IEEE Transactions on Patt. Anal. and Mach. Intel.*, Vol. PAMI-6, November 1989.

[17] Flandrin, P., "On the Spectrum of Fractional Brownian Motions," *IEEE Transactions on*

Information Theory, Jan. 1989.

[18] Wornell, G., "A Karhunen-Loeve-like Expansion for 1/f Processes via Wavelets," *IEEE Transactions on Information Theory*, Vol. 36, No. 9, July 1990.

[19] Wornell, G., Oppenheim, A., "Estimation of Fractal Signs from Noisy Measurements Using Wavelets," submitted to *IEEE Transactions on ASSP*.

[20] Tewfik, A., Kim, M., "Correlation Structure of the Discrete Wavelet Coefficients of Fractional Brownian Motions," Submitted to *IEEE Transactions on Information Theory*, Jan. 1992.

[21] Chou, K., Golden S., Willsky, A., "Modeling and Estimation of Multiscale Stochastic Processes," *International Conference on Acoustics, Speech, and Signal Processing*, Toronto, May 1991.

[22] Brandt, A., "Multi-level adaptive solutions to boundary value problems," Math. Comp. Vol. 13, 1977, pp. 333-390.

[23] Hackbusch, W. and Trottenberg, U., Eds., *Multigrid Methods*, Springer-Verlag, N.Y., N.Y., 1982.

[24] Paddon, D., and Holstein, H., Eds., *Multigrid Methods for Integral and Differential Equations*, Clarendon Press, Oxford, England, 1985.

[25] Basseville, M., Benveniste, A., Chou, K., Golden, S., Nikoukhah R., Willsky, A., "Modeling and Estimation of Multiresolution Stochastic Processes," submitted to *IEEE Transactions on Information Theory*, Jan. 1992.

[26] Chou, K., *A Stochastic Modeling Approach to Multiscale Signal Processing*, MIT, Department of Electrical Engineering and Computer Science, Ph.D. Thesis, (in preparation).

[27] Rauch, E., Tung, F., Striebel, C., "Maximum Likelihood Estimates of Linear Dynamic Systems," *AIAA Journal*, Vol. 3, No. 8, Aug. 1965.

[28] Schweppe, F., *Uncertain Dynamic Systems*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1973.

[29] Mallat, S. G., "A theory for multiresolution signal decomposition: the wavelet representation," Dept. of Computer and Info. Science - U. of Penn., MS-CIS-87-22, GRASP LAB 103, May 1987.

[30] Mallat, S., "Multiresolution Approximation and Wavelets". Dept of Computer and Information Science-Univ. of Penn., MS-CIS-87-87, GRASP Lab 80, Sept. 1987.

[31] Vaidyanathan, P. P., "Theory and design of m-channel maximally decimated quadrature

mirror filters with arbitrary M, having the perfect reconstruction property, *IEEE Trans. Acoust. Speech Signal Process.*", 35 (1987).

[32] Strang, G., *Linear Algebra and Its Applications*, Second Ed., Academic Press, N.Y., N.Y., 1980.

[33] Vaidyanathan, P. P., "Multirate Digital Filters, Filter Banks, Polyphase Networks, and Applications: A Tutorial," *Proceedings of the IEEE*, Vol. 78, No. 1, January 1990.

[34] Simoncelli, E., *Orthogonal Sub-band Image Transforms*, M.S. Thesis, M.I.T., 1988.

[35] Golub, G. H., Van Loan, C. F., *Matrix Computations*, The Johns Hopkins University Press, Baltimore, Maryland, 1983.

[36] Pierre, Donald A., *Optimization Theory with Applications*, Dover Publications, Inc., New York, 1986.

[37] Gallager, R., *Information Theory and Reliable Communications*, John Wiley & Sons, Inc., New York, 1968.

[38] Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, Third Ed., McGraw-Hill, Inc., New York, 1991.