SUBSPACE DETECTION AND SCALE EVOLUTIONARY EIGENDECOMPOSITION

SPYROS KYPEROUNTAS

SUBSPACE DETECTION AND SCALE EVOLUTIONARY EIGENDECOMPOSITION

by

Spyros Kyperountas

A Dissertation Submitted to the Faculty of the

College of Engineering

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

Florida Atlantic University

Boca Raton, Florida

December 2001

SUBSPACE DETECTION AND SCALE

EVOLUTIONARY EIGENDECOMPOSITION

by

Spyros Kyperountas

This dissertation was prepared under the direction of the candidate's dissertation advisor, Dr. Nurgün Erdöl, Department of Electrical Engineering, and has been approved by the members of his supervisory committee. It was submitted to the faculty of the College of Engineering and was accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

SUPERVISORY COMMITTEE:

Nurgun Erdol
Dissertation Advisor

N. Erdol for R. Sudhaker

Salvatore DMorgera
Chairperson, Department of Electrical Engineering

Dean, College of Engineering

Nathan W. Dean
Vice Provost

11/26/01
Date

ABSTRACT

Author:	Spyros Kyperountas
Title:	Subspace Detection And Scale Evolutionary Eigendecomposition
Institution:	Florida Atlantic University
Thesis Advisor:	Dr. Nurgün Erdöl
Degree:	Doctor of Philosophy
Year:	2001

A measure of the potential of a receiver for detection is detectability. Detectability is a function of the signal and noise, and given any one of them the detectability is fixed. In addition, complete transforms of the signal and noise cannot change detectability. Throughout this work we show that "Subspace methods" as defined here can improve detectability in specific subspaces, resulting in improved Receiver Operating Curves (ROC) and thus better detection in arbitrary noise environments. Our method is tested and verified on various signals and noises, both simulated and real. The optimum detection of signals in noise requires the computation of noise eigenvalues and vectors (EVD). This process neither is a trivial one nor is it computationally cheap, especially for non-stationary noise and can result in numerical instabilities when the covariance matrix is large. This work addresses this problem and provides solutions that take advantage of the subspace structure through plane rotations to improve on existing algorithms for EVD by improving their convergence rate and reducing their computational expense for given thresholds.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1

INTRODUCTION

1.1 An Overview of Subspace Detection

There are many possible representations that can be employed to represent a signal in terms that will enable one to manipulate any given prior information about it. Representations range from trivial time-sample representations through Fourier coefficients and Karhunen-Loeve (KL) coefficients to various time-frequency representations like Gabor coefficients (Short Time Fourier Transform (STFT) with translated time windows) and wavelet coefficient representations. A great deal of research has been done in the last ten years on the utilization of sub-band decomposition and wavelet analysis as well as other time-frequency transforms in an attempt to improve the existing detection schemes under different scenarios of noise and signal. Transforms like the wavelet decomposition have the advantage over transforms like Fourier or KL, of being localized in both time and scale (frequency) and this fact makes them a convenient approach to certain detection problems. Gabor transform was proposed for the detection of signals with unknown arrival time even if partially overlapped [BFBP89]. Wavelet transform was also proposed and successfully implemented for a known transient signal detection of unknown scaling and arrival time in additive white Gaussian noise (AWGN) [MFHM94], as well as for unknown transient signals by exploiting prior information of the signal's relative bandwidth and its time-

bandwidth product [MFHM92], [SKGV00]. Wavelet transform is also used in the detection/characterization of non-stationarities [HKKD92], [WLHW99], an application that results from the property of the wavelet transform to introduce stationarity in its coefficients [SMAL89], [PFLA92]. Other implementations of the wavelet structure for detection worth mentioning deal with specific noise scenarios. These methods use the wavelet structure as a sub-optimum receiver replacing the whitening filter (need for sufficient statistics), but only work for a limited range of noise like impulsive noise [GFLI97] and 1/f noise [GWAO92], [GWOR90], [GWRL90], [WLHW99], and [TLAF00].

A measure of the potential of a receiver for detection is detectability. Detectability is a function of the signal and noise, and given any one of them the detectability is fixed. In addition, complete transforms of the signal and noise cannot change detectability. Throughout this work we show that "Subspace methods" as defined in this work can improve detectability in certain subspaces resulting in improved Receiver Operating Curves (ROC) and thus better detection in arbitrary noise environments. Our method is tested and verified on various signals and noises, both simulated and real. Furthermore there is a need to describe the process of detection of the signal in a multirate environment, in a more analytical way that would allow a measure of the detection goodness at any part of the process. For this task we used the Karhunen-Loeve transform (KLT) as a tool to assess and analyze the results, particularly when wavelet decomposition is used, although the results should be easily applied to any transform detection schemes.

The optimum detection of signals in Gaussian noise requires the computation of noise eigenvalues and vectors. This process is known as eigenvalue decomposition (EVD). The complete set of eigenvectors form the Karhunen-Loeve basis of the noise process. This process is neither a trivial one nor is it computationally cheap, especially for non-stationary noise. It can result in numerical instabilities when the covariance matrix is large. This work addresses this problem and provides solutions that take advantage of the subspace structure, through plane rotations, to improve on existing algorithms by improving their convergence rate and reducing their computational expense for given thresholds.

In the next section we give some background information on wavelets, multiresolution analysis and filterbanks.

1.2 Wavelets and The Wavelet Transform

The basic (dyadic) wavelet analysis represents a windowing technique with variable sized regions both in frequency (scale) and time. It allows the use of long time intervals, where we want more precise low frequency information, and shorter regions, where we want high frequency information. Figure 1.1 shows an illustration of how the wavelet transform compares with time-based, frequency-based and STFT views of a signal. Wavelet transform, just like Fourier transform, uses a series of expansion signals (basis) to represent a signal. The expansion signals are called wavelets ('small waves') that have zero mean value and, unlike Fourier basis (sinusoids), they can be concentrated in time with finite energy as shown in Figure 1.2. This allows for accurate analysis of transient, nonstationary, or time-varying phenomena. Also whereas sinusoids are predictable and smooth, the wavelets can be irregular and non symmetric. They still have the oscillating

wavelike characteristic but also the ability to allow simultaneous time and frequency analysis with a flexible mathematical foundation.

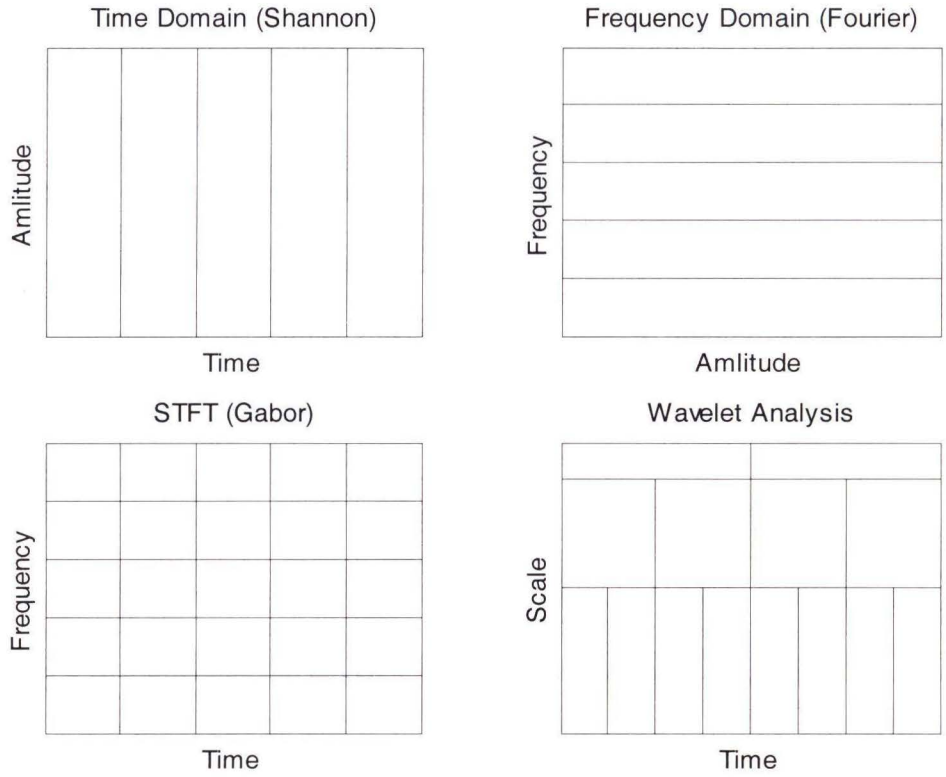
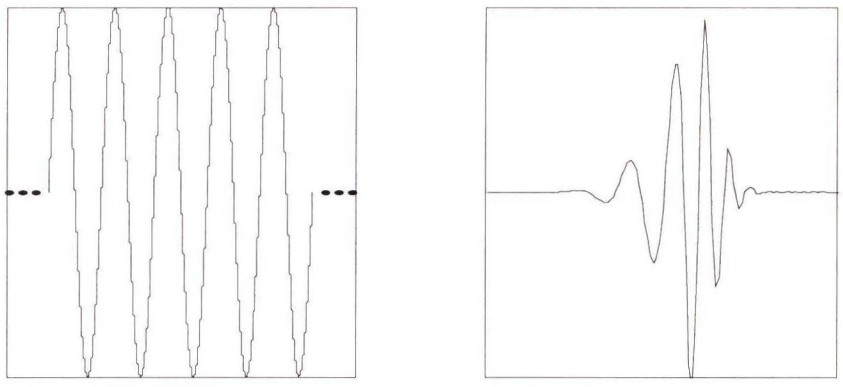


Figure 1.1: Comparison of signal views in different transforms.



a) b)

Figure 1.2: a) Sine Wave, b) Wavelet (db8)

Looking at Figure 1.2 one can see intuitively that signals with sharp edges might be analyzed with an irregular wavelet than a smooth sinusoid. It also makes sense that local features can be described better with wavelets that have finite support.

An important difference between the wavelet transform and other major transforms (like Fourier, Gabor etc.) is that the basis functions in the wavelet transform, or wavelets, are not unique i.e. for different applications the wavelets can be chosen differently. Although this aspect of the wavelet transform makes it difficult to develop general properties for the transform, it gives flexibility to the user to employ the appropriate set of basis functions for a specific application. In fact it is this flexibility that makes the wavelet transform an attractive tool for its users.

A wavelet system is a set of building blocks to construct or represent a signal or a function. It is a two-dimensional expansion set (usually a basis) for some class of one- (or higher) dimensional signals. In other words, if the wavelet is a given set of $\psi_{j,k}(t)$ for indices of j,k = ..-2,-1,0,1,2,..., a linear expansion of signal f(t) would be:

$$f(t) = \sum_k \sum_j \alpha_{j,k} \psi_{j,k}(t) \tag{1.1}$$

for some set of coefficients $\alpha_{j,k}$ [BSRG98]. All so-called first-generation wavelet systems are generated by simple scaling and translation. The two dimensional parameterization is achieved from the function (sometimes called generating wavelet or mother wavelet) $\psi(t)$ by:

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k), \qquad j, k \in \mathbf{Z} \tag{1.2}$$

where **Z** is the set of all integers and the factor $2^{j/2}$ maintains a constant norm, independent of scale j. This parameterization of the time or space location by k and the frequency or scale (logarithm of scale) by j turns out to be extraordinary effective. Figure 1.3 is a pictorial representation of the translation and scaling of a single wavelet (db4).

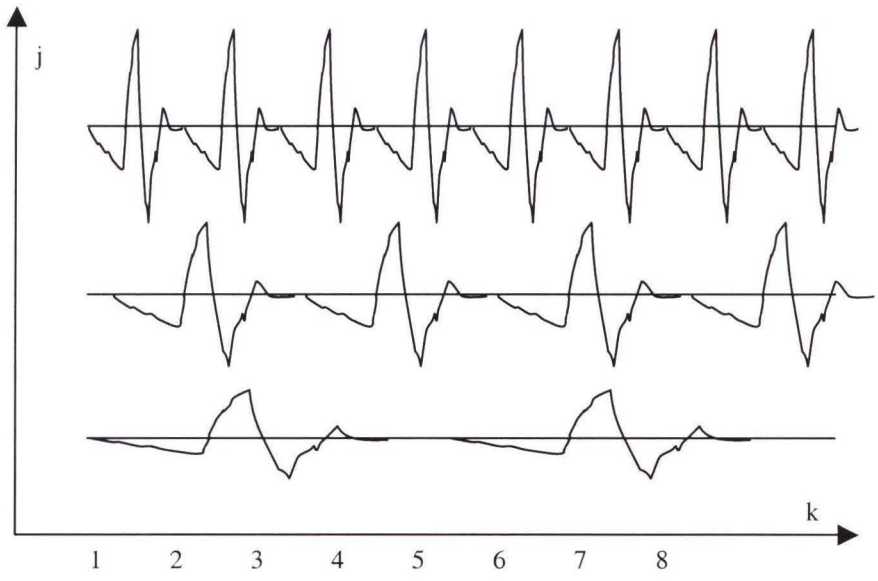


Figure 1.3: Translation (every fourth k) and Scaling of a Wavelet db4.

Substituting equation (1.2) into equation (1.1) results in:

$$f(t) = \sum_k \sum_j \alpha_{j,k} 2^{j/2} \psi(2^j t - k) \tag{1.3}$$

where the two dimensional set of coefficients $\alpha_{j,k}$ are called the discrete wavelet transform (DWT) of f(t). The coefficients $\alpha_{j,k}$ are calculated using the inner product as follows:

$$\alpha_{j,k} = \left\langle \psi_{j,k}(t), f(t) \right\rangle = \int \psi_{j,k}(t) f(t)^* dt \tag{1.4}$$

if the $\psi_{j,k}(t)$ form an orthonormal basis for the space of signals of interest [IDAU92]. If the signal itself is discrete (either samples of a continuous signal or a set of inner products), the expansion of the signal is called a discrete-time wavelet transform (DTWT). If the signal is a function of a continuous variable and a transform that is a function of two continuous variables is desired, the continuous wavelet transform (CWT) can be defined by

$$F(a,b) = \int f(t) \psi\left(\frac{t-a}{b} \right) dt \tag{1.5}$$

with an inverse transform of:

$$f(t) = \int\int F(a,b) \psi\left(\frac{t-a}{b} \right) da db \tag{1.6}$$

where $\psi(t)$ is the basic wavelet and a, b $\in$ **R** are real continuous variables [IDAU92], [CHDW89].

In the next section we introduce the multiresolution formulation of the wavelet transform, which seems to be a convenient approach in explaining the effects of changing scale both in mathematical and practical interpretations of wavelets.

1.3 Multiresolution Analysis

Multiresolution analysis, which is formulated by Mallat and Mayer [SMAL89], [YMAY90] serves as a convenient framework for the understanding of orthonormal wavelet basis as well as for the construction of new ones. Let us start by defining the scaling functions $\varphi_{j,k}(t)$ as:

$$\varphi_{j,k}(t) = 2^{j/2}\varphi(2^j t - k), \qquad \varphi \in L^2, \quad j,k \in \mathbf{Z} \tag{1.7}$$

and subspace of $L^2(\mathbf{R})$ spanned by these functions defined by:

$$\mathcal{V}_j = \overline{\operatorname{Span}_k\{\varphi_k(2^j t)\}} = \overline{\operatorname{Span}_k\{\varphi_{j,k}(t)\}} \tag{1.8}$$

for all integers $k \in \mathbf{Z}$. The over-bar denotes closure. This means that

$$f(t) = \sum_k \alpha_k \varphi(2^j t + k) \tag{1.9}$$

for any $f(t) \in \mathcal{V}_j$. Let us now formulate the basic requirements of multiresolution analysis by requiring a nesting of spanned subspaces as:

$$\cdots \subset \mathcal{V}_{-2} \subset \mathcal{V}_{-1} \subset \mathcal{V}_0 \subset \mathcal{V}_1 \subset \mathcal{V}_2 \subset \cdots \subset \mathbf{L}^2 \tag{1.10}$$

or $\mathcal{V}_j \subset \mathcal{V}_{j+1}$ for all $j \in \mathbf{Z}$, with $\mathcal{V}_{-\infty} = \mathbf{L}^2$. The space that contains the higher resolution signals will contain the lower ones also. Because of the definition of $\mathcal{V}_j$, the spaces have to satisfy a natural scaling condition

$$f(t) \in \mathcal{V}_j \Leftrightarrow f(2t) \in \mathcal{V}_{j+1} \tag{1.11}$$

which ensures elements in a scale are simply scaled versions of the elements in the next space. The relationship of the spanned is illustrated in Figure 1.4.

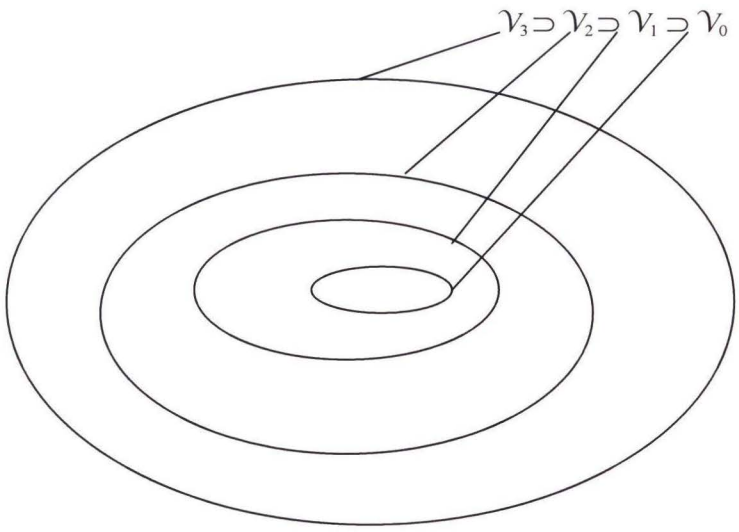


Figure 1.4: Nested Vector Spaces Spanned by the Scaling Functions.

The nesting of spans of $\varphi(2^j t-k)$, denoted by $\mathcal{V}_j$ and shown in equations (1.10) and (1.11) and graphically illustrated in Figure 1.4, is achieved by requiring that $\varphi(t) \in \mathcal{V}_1$, which means that if $\varphi(t)$ is in $\mathcal{V}_0$, it is also in $\mathcal{V}_1$, the space spaned by $\varphi(2t)$. This means $\varphi(t)$ can be expressed in terms of a weighted sum of shifted $\varphi(2t)$ as:

$$\varphi(t) = \sum_n h(n)\sqrt{2}\varphi(2t - n), \qquad n \in \mathbf{Z} \tag{1.12}$$

two. Coefficients h(n) are found by evaluating the inner product of $\varphi(t)$ and $\varphi_{1,n}(t)$. This recursive equation is the fundamental equation to the theory of the scaling functions and is called the dilation equation. As an example we show in Figure 1.5 the scaling function of Haar wavelet $\varphi(t)$ which is constructed with $\varphi(2t)$ as:

$$\varphi(t) = \varphi(2t) + \varphi(2t-1) \tag{1.13}$$

which means equation (1.12) is satisfied for coefficients $h(0)=1/\sqrt{2}$, $h(1)=1/\sqrt{2}$.

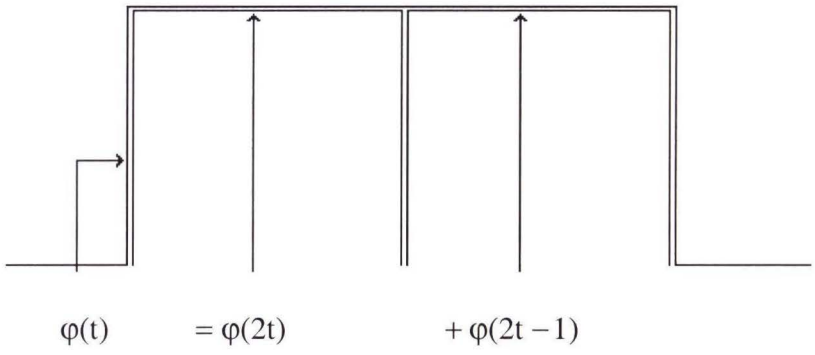


Figure 1.5: Haar Scaling Functions.

The next step in defining multiresolution subspaces is to define a set of functions $\psi_{j,k}(t)$ (the wavelets defined in the previous section) such that

$$\left\langle \varphi_{j,k}(t), \psi_{j,l}(t) \right\rangle = \int \varphi_{j,k}(t)\psi_{j,l}(t)dt = 0 \tag{1.14}$$

This means that if $\psi_{j,k}(t)$ spans subspace W_j, all members of V_j are orthogonal to members of W_j. Also W_j can be defined as the orthogonal complement of V_j in V_{j+1}. In other words, it

This means that if $\psi_{j,k}(t)$ spans subspace $\mathcal{W}_j$, all members of $\mathcal{V}_j$ are orthogonal to members of $\mathcal{W}_j$. Also $\mathcal{W}_j$ can be defined as the orthogonal complement of $\mathcal{V}_j$ in $\mathcal{V}_{j+1}$. In other words, it represents the difference between the spaces spanned by the various scales of the scaling function. As such the wavelet subspace $\mathcal{W}_0$ can be defined as:

$$\mathcal{V}_1=\mathcal{V}_0 \oplus \mathcal{W}_0 \tag{1.15}$$

or in general,

$$L^2=\mathcal{V}_0 \oplus \mathcal{W}_0 \oplus \mathcal{W}_1 \oplus \cdots \tag{1.16}$$

Figure 1.6 pictorially shows the nesting of the scaling functions $\mathcal{V}_j$ for different scales j and how the wavelet spaces are the disjoint differences or the orthogonal complements.

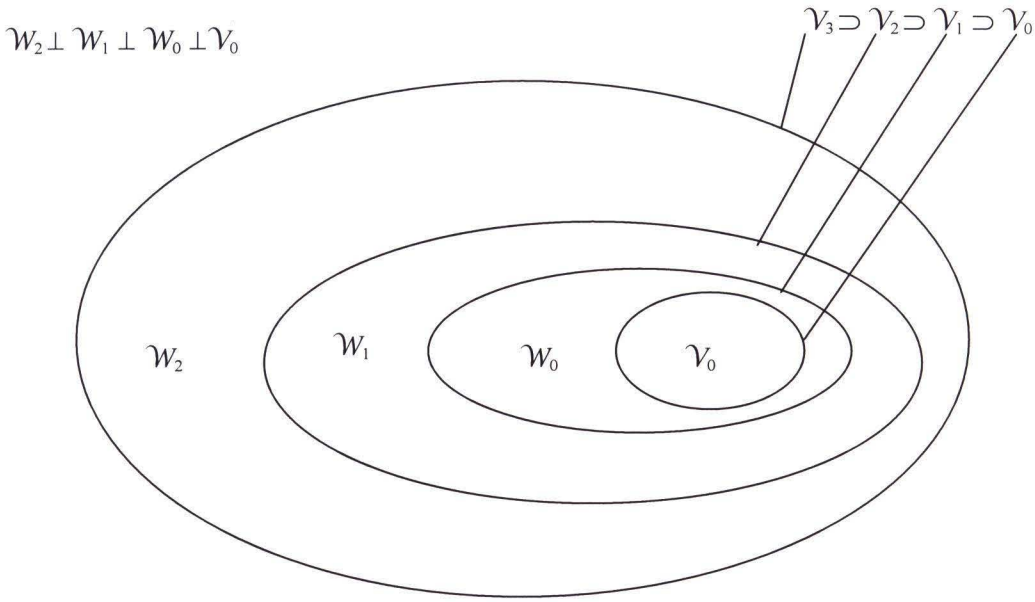


Figure 1.6: Scaling Function and Wavelet Vector Spaces.

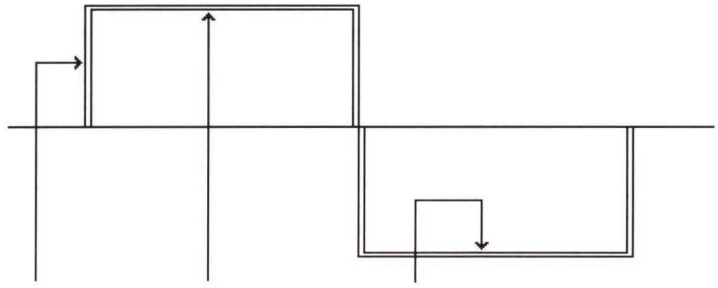
Since the wavelets reside in the space spanned by the next narrower scaling function, $W_0 \subset V_1$, they can be represented by a weighted sum of shifted scaling function $\varphi(2t)$ defined in equation (1.12) by

$$\psi(t) = \sum_n g(n)\sqrt{2}\varphi(2t - n), \qquad n \in Z \tag{1.17}$$

for some set of coefficients g(n). Coefficients g(n) are found by evaluating the inner product of $\psi(t)$ and $\varphi_{1,n}(t)$. As an example, we show in Figure 1.7 the wavelet function of Haar wavelet $\psi(t)$ which is constructed with $\varphi(2t)$ as:

$$\psi(t) = \varphi(2t) - \varphi(2t - 1) \tag{1.18}$$

which means equation (1.17) is satisfied for coefficients $g(0)=1/\sqrt{2}$, $g(1)=-1/\sqrt{2}$.



$\psi(t)$ $= \varphi(2t)$ $- \varphi(2t - 1)$

Figure 1.7: Haar Wavelets.

Since the set of functions $\psi_{j,k}(t)$ and $\varphi_k(t)$ could span all of $L^2(\mathbf{R})$, any function $f(t) \in L^2(\mathbf{R})$ could be written as

$$f(t) = \sum_k c_{j_0}(k) 2^{j_0/2} \varphi\left(2^{j_0/2} t - k\right) + \sum_k \sum_{j=j_0}^{\infty} d_j(k) 2^{j/2} \psi\left(2^{j/2} t - k\right) \tag{1.19}$$

where j_0 is the initial subspace, and coefficients $c_j(k)$ and $d_j(k)$ can be calculated as inner products of the signal f(t) and the scaling and the wavelet functions respectively as shown below

$$c_j(k) = \left\langle f(t), \varphi_{j,k}(t) \right\rangle = \int f(t) \varphi_{j,k}(t) dt \tag{1.20}$$

and

$$d_j(k) = \left\langle f(t), \psi_{j,k}(t) \right\rangle = \int f(t) \psi_{j,k}(t) dt \tag{1.21}$$

For better clarity as to what these coefficients mean and their relationships to the subspaces an example of a doppler signal decomposition with db8 wavelet is shown. Figure 1.8 shows the resulting coefficients $c_0(k)$ and $d_j(k)$ with respect to the original signal. Figures 1.9, 1.10 show the associated projections of the signal onto $\mathcal{V}$ and $\mathcal{W}$ spaces respectively.

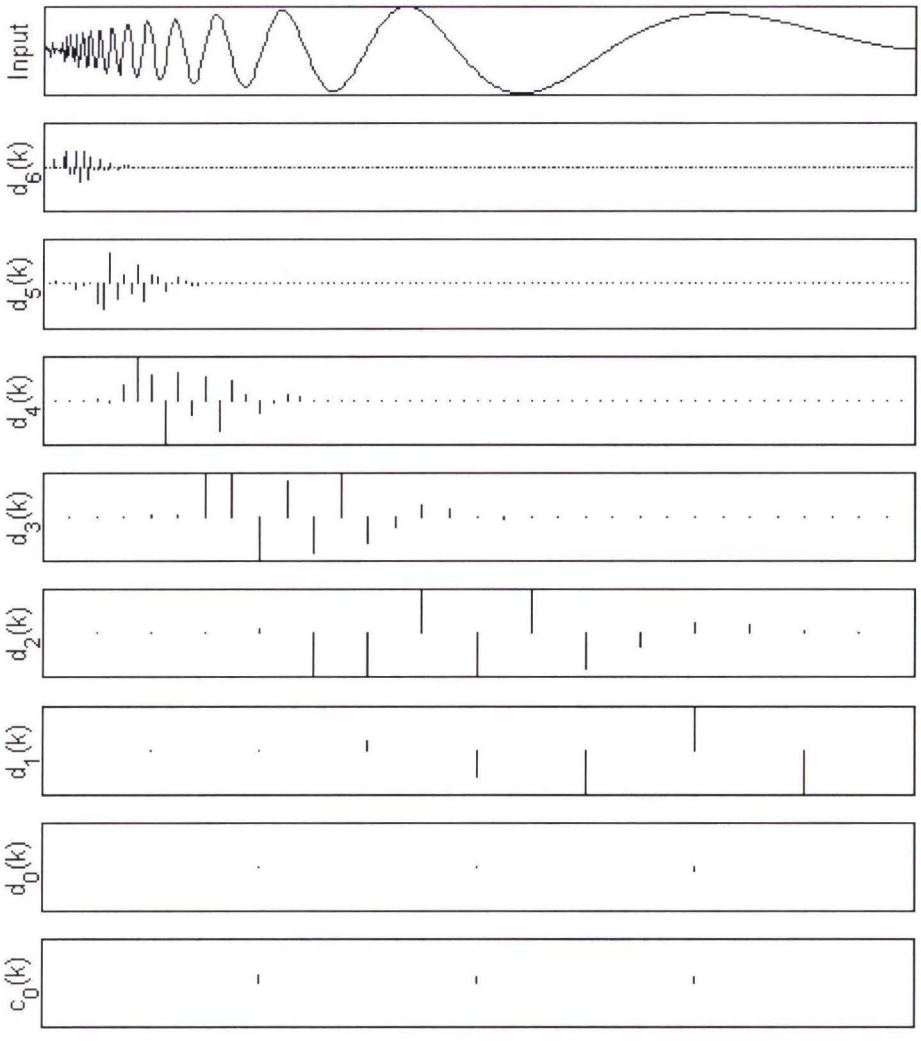


Figure 1.8: Discrete Wavelet Transform of a Doppler, using db8.

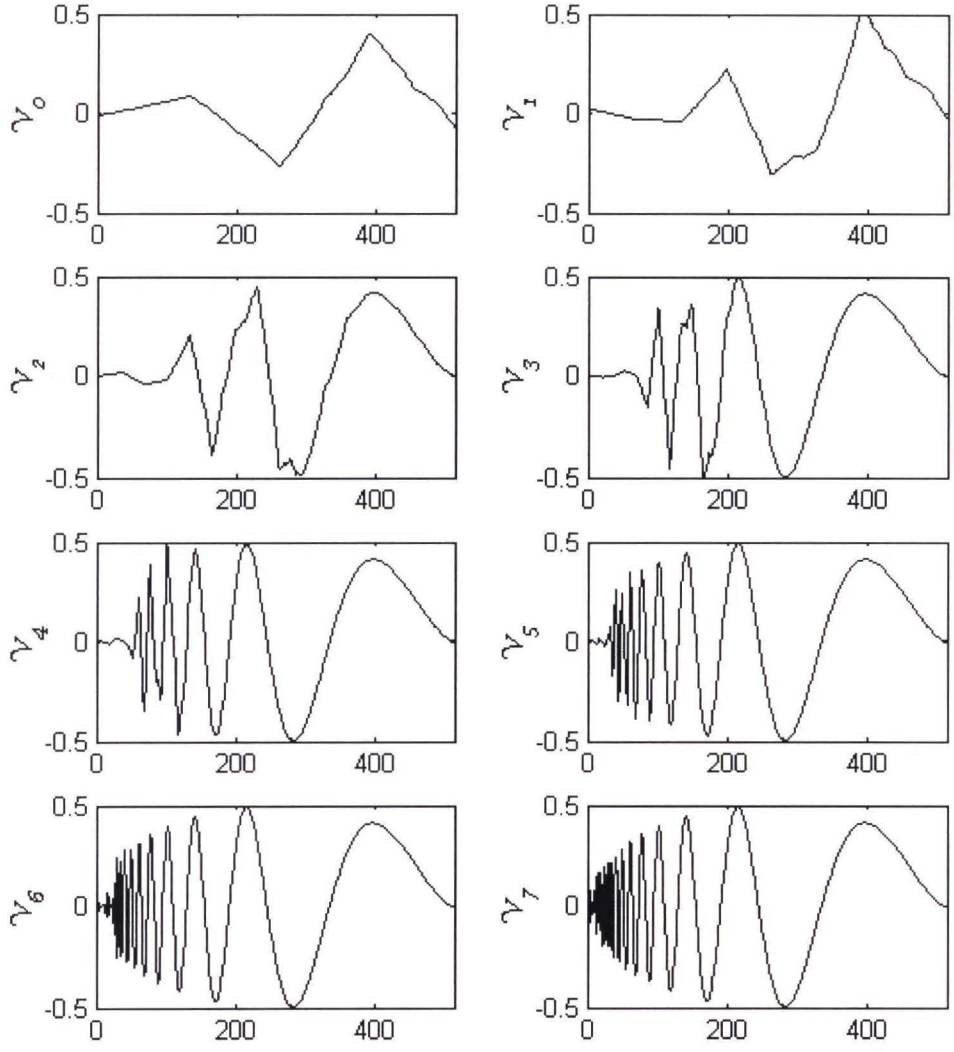


Figure 1.9: Projection of the Doppler Signal onto V Spaces using db8.

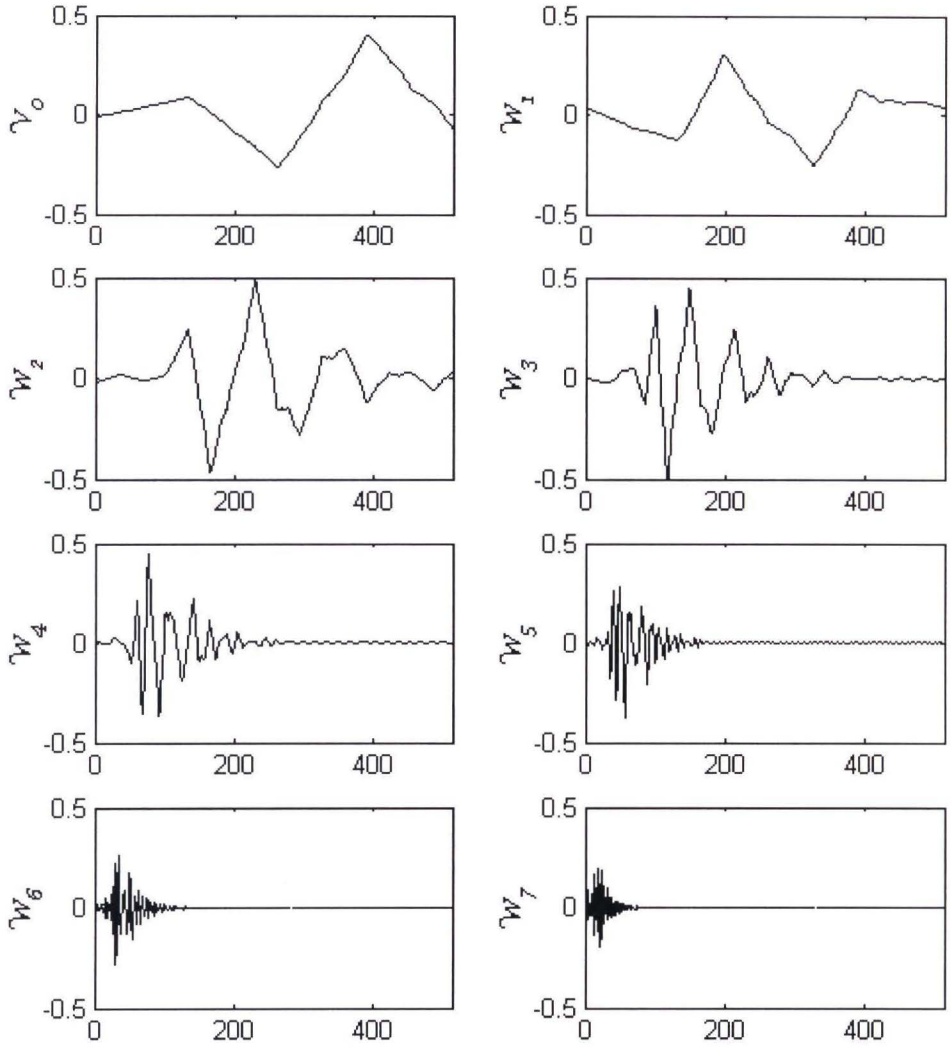


Figure 1.10: Projection of the Doppler Signal onto $\mathcal{W}$ Spaces using db8.

To end this section of Chapter 1 we would like to give expressions for the wavelet and scaling functions as well as the signal decomposition in the frequency domain. We start from expressions (1.12) and (1.17) which in frequency domain translate to

$$\Phi(\omega) = 2^{N/2} H\left(\frac{\omega}{2}\right) H\left(\frac{\omega}{4}\right) H\left(\frac{\omega}{8}\right) \cdots H\left(\frac{\omega}{2^N}\right) \Phi\left(\frac{\omega}{2^N}\right) \tag{1.22}$$

and

$$\Psi(\omega) = 2^{N/2} G\left(\frac{\omega}{2}\right) H\left(\frac{\omega}{4}\right) H\left(\frac{\omega}{8}\right) \cdots H\left(\frac{\omega}{2^N}\right) \Phi\left(\frac{\omega}{2^N}\right) \tag{1.23}$$

for any arbitrary N>0, where $\Phi(\omega)$, $\Psi(\omega)$, $H(\omega)$, and $G(\omega)$ are the Fourier transform of $\varphi(t)$, $\psi(t)$, h(n), and g(n) respectively. Finally from equation (1.19) we can find the Fourier transform of f(t), $F(\omega)$, with respect to the spectrum of $\varphi(t)$, and $\psi(t)$ as

$$F(\omega) = C_{j_0}\left(\omega 2^{-(L+j_0)}\right)\left\{2^{-(2L+3j_0/2)}\Phi\left(2^{-j_0}\omega\right)\prod_{i=1}^{L} H\left(-2^{-(i+j_0)}\omega\right) + \right.$$
$$\left. + \sum_{j=j_0}^{j_0+L-1} 2^{-(2L'+3j/2)}\Phi\left(2^{-j}\omega\right)\prod_{j=2}^{L'} H\left(-2^{-(i+j)}\omega\right) G\left(-2^{-(1+j)}\omega\right)\right\} \tag{Eq.1.24}$$

where L is the total levels of decomposition (depth) we are willing to go and $L' = L-(j-j_0)$.

In the next section we introduce the filter bank theory that allows an efficient calculation of the lower resolution coefficients from the higher resolution coefficients by a tree structure of filters.

1.4 Filter Banks and Discrete Wavelet Transform

In many applications, one never has to deal directly with the scaling functions or wavelets. The entire concept of decomposition can be described by coefficients h(n) and g(n) from equations (1.12) and (1.17), and c(k) and d(k) in expansions (1.19), (1.20) and

(1.21). In this section we are going to show how we can get these coefficients recursively using an old tool, the filter. This work is primary attributed to the work of Mallat and Daubechies [SMLL89], [SMAL89], [IDAU92].

1.4.1 Analysis – From Fine Scale to Coarse Scale

In order to work with the wavelet transform coefficients, we will derive the relationship between the expansion coefficients at a lower scale in terms of those at a higher scale. We start first by substituting equation (1.7) into equation (1.20). This results to

$$c_j(k) = \left\langle f(t), \varphi_{j,k}(t) \right\rangle = \int f(t) 2^{j/2} \varphi(2^j t - k) dt. \tag{1.25}$$

Then from equation (1.12) by substituting t with 2^jt-k we can write

$$\varphi(2^j t - k) = \sum_n h(n)\sqrt{2}\varphi\left(2\left(2^j t - k\right) - n\right) = \sum_n h(n)\sqrt{2}\varphi\left(2^{j+1} t - 2k - n\right) \tag{1.26}$$

and then we can define m = 2k + n resulting to

$$\varphi(2^j t - k) = \sum_m h(m - 2k)\sqrt{2}\varphi\left(2^{j+1} t - m\right). \tag{1.27}$$

Equation (1.27) can be then substituted to equation (1.25) and interchanging the sum and the integral, leads to

$$c_j(k) = \int f(t) 2^{j/2} \sum_m h(m - 2k)\sqrt{2}\varphi\left(2^{j+1}t - m\right)dt =$$
$$= \sum_m h(m - 2k) \underbrace{\int f(t) 2^{(j+1)/2} \varphi\left(2^{j+1}t - m\right)dt}_{c_{j+1}(m)} \tag{1.28}$$

and in a final form as

$$c_j(k) = \sum_m h(m - 2k) c_{j+1}(m). \tag{1.29}$$

The corresponding relationship for the wavelet coefficients can be found using similar procedure as above that results to

$$d_j(k) = \sum_m g(m - 2k) c_{j+1}(m). \tag{1.30}$$

From first look equations (1.29), and (1.30) remind us of a convolution procedure that results from some filtering of $c_{j+1}(k)$ with Finite Impulse Response (FIR) filters h(-2k) and g(-2k) respectively. More precisely these equations can be interpreted with the filter structure shown in Figure 1.11. The $\downarrow 2$ block represents a down-sampler or a decimator by two. In general, if the input to the decimator is an arbitrary discrete signal x(n), its output is x(2n).

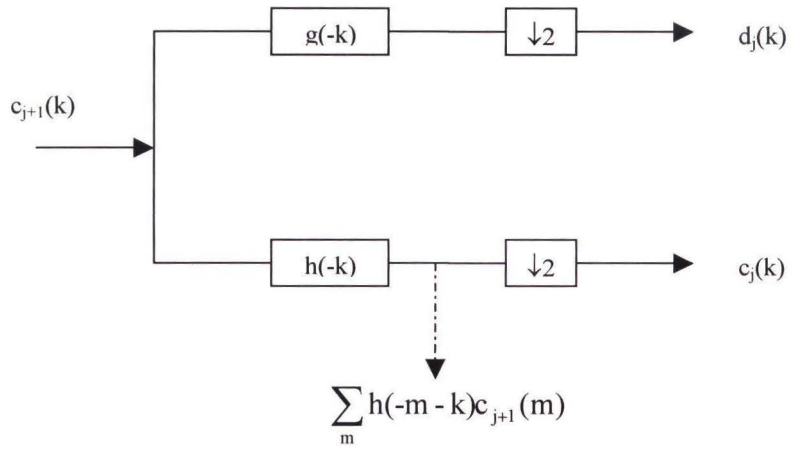


Figure 1.11: Two-Band Analysis Bank.

This procedure can be repeated on $c_j(k)$ to produce $c_{j-1}(k)$, and $d_{j-1}(k)$ and so on (iterating the filter-bank), resulting to a tree-like structure as shown in Figure 1.12.

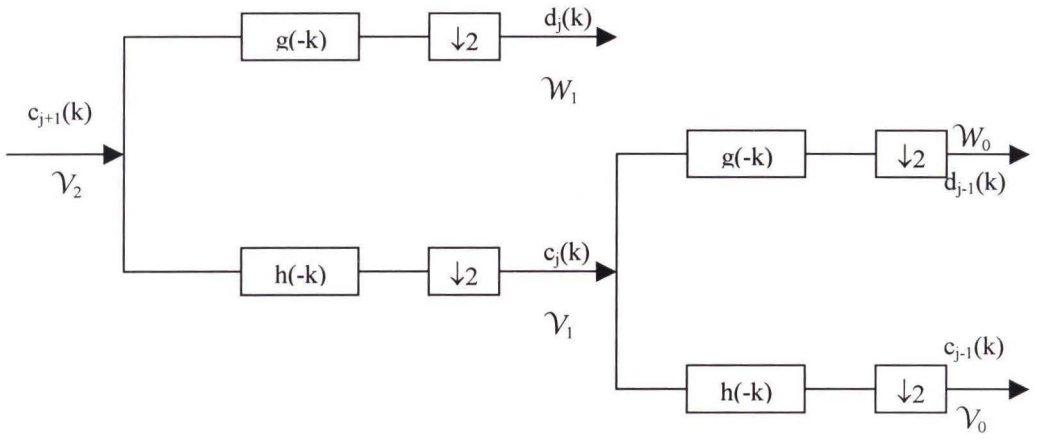


Figure 1.12: Two Stage Two-Band Analysis Tree.

The filter implemented by h(-k) and g(-k) are found to have low-pass and high-pass properties respectively. This means that in every stage output the input signal spectrum is separated to a high-pass component and a low-pass component. This is illustrated in Figure 1.13.

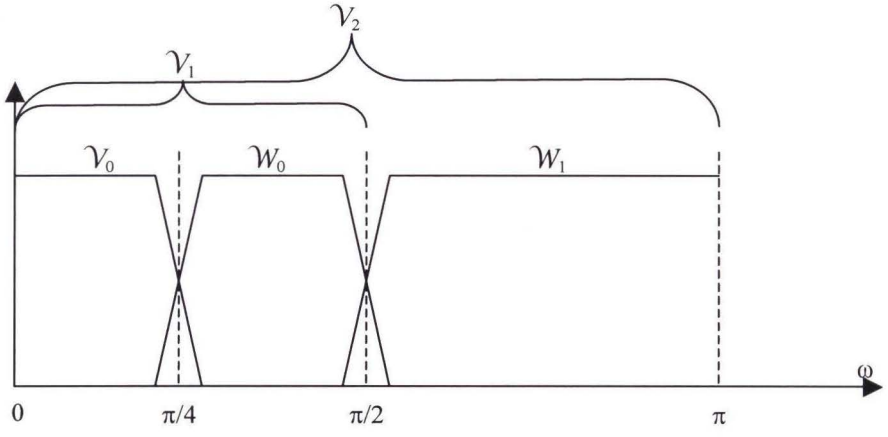


Figure 1.13: Frequency Bands for the Analysis Tree.

Note that the number of points at the input to such a structure is equal to the total number of points to the output of it. This means that there is a possibility that no information is lost and it will be possible to completely recover the original input signal. Indeed with certain restrains on the filter structure this is proven to be possible. The aliasing occurring in the upper bank can be canceled by using the signal of the lower bank. This is the idea behind perfect reconstruction in filter theory, which will be presented, more analytically in upcoming chapters.

1.4.2 Synthesis – From Coarse Scale to Fine Scale

A reconstruction of the original fine scale coefficients of the signal can be made from a combination of the scaling function and wavelet coefficients at a coarse resolution. This can be derived by considering a signal f(t) in the j+1 scaling function space f(t) $\in$ $\mathcal{V}_{j+1}$. Thus the function can be written as

$$f(t) = \sum_k c_{j+1}(k) 2^{(j+1)/2} \varphi\left(2^{j+1} t - k\right) \tag{1.31}$$

or in terms of the next scale as

$$f(t) = \sum_k c_j(k) 2^{j/2} \varphi\left(2^j t - k\right) + \sum_k d_j(k) 2^{j/2} \psi\left(2^j t - k\right). \tag{1.32}$$

Similarly to equation (1.27) we can write

$$\psi(2^j t - k) = \sum_m g(m - 2k)\sqrt{2}\varphi\left(2^{j+1} t - m\right) \tag{1.33}$$

and substitute both expressions in equation (1.32) and changing the summation order we get

$$f(t) = \sum_m 2^{(j+1)/2} \varphi\left(2^{j+1} - m\right)\left(\sum_k c_j(k) h(m - 2k) + \sum_k d_j(k) g(m - 2k) \right). \tag{1.34}$$

Finally comparing equation (1.34) to (1.31) we get

$$c_{j+1}(k) = \sum_n c_j(n) h(k - 2n) + \sum_n d_j(n) g(k - 2n). \tag{1.35}$$

Similarly to the analysis case, these equations can be realized using filters. More precisely equation (1.35) can be interpreted with the filter structure shown in Figure 1.14. The $\uparrow 2$ block represents an up-sampler or stretching by two. In general the up-sampler by two simply inserts zeros to the input sequence every other sample.

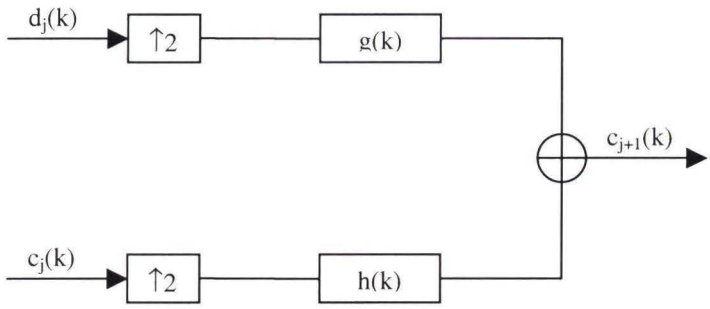


Figure 1.14: Two-Band Synthesis Bank.

This procedure can be repeated at any level by combining the appropriate wavelet coefficients. Figure 1.15 illustrates a two-stage two-band synthesis tree, which can be used to reconstruct Figure 1.12 analysis tree.

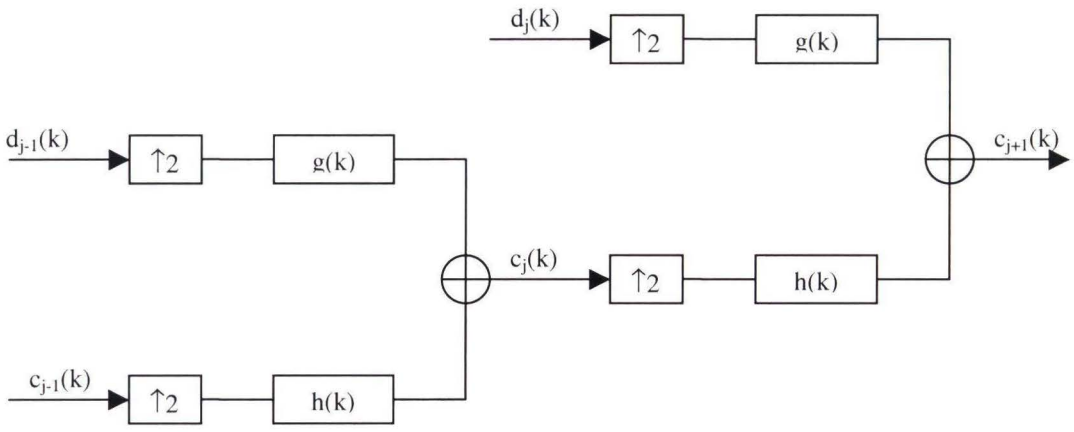


Figure 1.15: Two-Stage Two-Band Synthesis Tree.

1.5 Dissertation Contributions in Upcoming Chapters

In this Chapter we have given a brief overview of the theory of wavelets. In Chapter 2 the detection over multi-resolution analysis subspaces will be examined. Chapter 3 will deal with description of the simulations and the associated results. Chapter 4 will present the eigenvalue-eigenvector problem solution through plane rotation algorithms, and show an efficient way for these algorithms to be implemented in subspace to save computational expense. Finally Chapter 5 will consist of the conclusions of the study as well as recommendations for future studies. The discrete subspace analysis of Chapter 2 and the related simulation results presented in Chapter 3, we believe to be original work. Furthermore, the work done on EVD methods in subspaces using plane rotations, as described in Chapter 4, is also believed to be original work. More specifically the dissertation contributions are:

- Improve signal detection in any noise environment by utilizing a subspace detector.
    - Optimal detection in subspaces.
    - Evaluate detectability-Verify results using ROC curves.
    - Provide theoretical backbone for discrete signal subspace detection.
- Examine subspace EVD.
    - Derive relationships for the EVD evolution along scales.
    - Provide methods for improving existing EVD algorithms by using subspace signal structure.
    - Evaluate and quantify the advantages-disadvantages of various algorithms for different subspace signal structures.

CHAPTER 2

DETECTION OVER MULTIRESOLUTION ANALYSIS SUBSPACES

2.1 Subspace Detection using KL Basis

Detection of signals in noise has been studied independently of wavelets for nearly half a century. Its success has resulted in significant advances in the broad areas of digital communications, radar/sonar, and pattern recognition. It is also a general area that since the introduction of wavelets and multiresolution analysis (MRA) into signal processing, has attracted research attention for detection of short duration signals embedded in correlated noise, detection in non-stationary environments, and detection and estimation in environments with low signal-to-noise ratios. Optimum detection of a known signal in Gaussian noise requires the estimation of the KL basis functions or covariance estimates or other noise whitening strategies, which are in general difficult to implement since they are computationally complex. This led to the exploration of strategies for optimum detection in MRA subspaces.

Optimum detection of signals in Gaussian noise requires the computation of noise eigenvalues and vectors, which form the Karhunen-Loeve (KL) basis. This is a computationally complex operation and is subject to numerical instabilities when the size of the covariance matrix is large. Fixed transforms such as the Discrete Cosine

Transform (DCT) and Discrete Wavelet Transform (DWT) [NESK01] are acceptable approximations to the KL transform and work for specific random processes. Many natural or human generated noise environments are non-stationary. Some examples are the different types of ambient acoustic noise through which a cellular telephone [NERS00] user communicates and underwater ambient noise [ANSH01]. Under such circumstances, a fixed transform is not sufficient; and there are benefits to be gained from using the KL transform, which is based on noise statistics.

The underlying reason for obtaining eigenvalues and eigenvectors of the noise is to whiten the noise, and follow it by matched filtering. If the information signal can be designed, then the probability of error in the detection process is reduced by choosing this signal to reside in the subspace spanned by the eigenvector corresponding to the smallest eigenvalue. In such cases, matched filtering is equivalent to finding the projection of the incoming signal on the signal subspace.

In this chapter we explore optimum detection strategies of signal in noise in any subspace particularly in a multiresolution analysis (MRA) subspaces. We set the stage for detection on sets of nested subspaces: Multiresolution (MR) subspaces on the outside and KL subspaces on the inside. This process allows for optimum detection, uses small KL matrixes, and takes into consideration long and short-term autocorrelation lags. It is numerically stable and allows for flexibility in signal design that may be combined with coding and encryption.

This chapter is organized as follows. We review the optimum detection problem in Section 2.2. Section 2.3 highlights the KL transform over MR subspaces for an

infinite and finite signal dimension case. In Section 2.5, we discuss the evolution of detectability over scales.

2.2 Review of Optimum Detection [CWHE95], [HVRS68]

For clarification purposes we will start with the notation to be used in this chapter. Bold lower case letters are used to represent a vector, bold upper case letters a matrix, and calligraphic letters represent spaces.

A binary detection problem is described by two hypotheses:

$$H_1: x(t) = s(t)+n(t), \qquad (2.1.a)$$

$$H_0: x(t) = n(t) \qquad t \in T_0 \qquad (2.1.b)$$

where s(t) is the message signal of known shape and energy, n(t) is zero-mean, second-order Gaussian random noise with a continuous, square-integrable covariance function $E[n(t)n(u)]=r_n(t,u)$, and T_0 is the observation interval. It is assumed that the noise energy in the observation interval $\int_{T_0} r_n(t,t)dt$ is finite. The fundamental components of detection are decision-making and signal processing.

The two established principles of decision-making are the Bayes and the Neyman-Pearson (N-P) criteria. Both Bayes and N-P tests lead to the likelihood ratio (LR) test, which is derived to minimize the risk of making a wrong decision. In terms of transition probabilities,

$$LR(\mathbf{X}) = \frac{P(\mathbf{X}/H_1)}{P(\mathbf{X}/H_0)} \tag{2.2}$$

where $P(\mathbf{X}/H_0)$ represents the conditional probability that the signal x(t) is received given the hypothesis H_0 is true. In an ideal system, LR=∞ if x(t) contains the message signal, and LR=0 if the received signal is just noise; and detection is possible with zero probability of error. In general, a decision is made by comparing the LR against a threshold. Naturally, if LR(**X**) clusters around two points that are separated by a sufficient distance, the threshold can be chosen with a higher degree of confidence. The purpose of an ideal receiver can thus be expressed as one that, in some sense, comes close to the ideal values for the LR. Expressions for LR are more useful in terms of sufficient statistics.

A general expression for LR uses the eigenfunctions, $\theta_k(t)$, and the eigenvalues, λ_k, of the noise covariance function $r_n(t,u)$. These quantities satisfy the eigenvalue equation

$$\int_{T_0} r_n(t,u)\theta_k(u)dt = \lambda\theta_k(t) \quad t\in T_0,\ k=1,2,\ldots . \tag{2.3}$$

Equivalently, the coefficients of n(t) in the series

$$n(t) = \sum_k n_k \theta_k(t) \tag{2.4}$$

are uncorrelated, i.e. $E[n_j n_k]=\lambda_k\delta_{jk}$, where δ_{jk} is the Kronecker delta. Using Mercer's formula, its covariance can be written as

$$r_n(t,u) = \sum_k \lambda_k \theta_k(t)\theta_k(u) \tag{2.5}$$

The series of equation (2.5) is called the *Karhunen-Loéve* (KL) expansion of the noise signal. The input and the message signals also admit expansion in terms of the noise KL basis functions where the expansion coefficients are

$$x_k = \int_{T_0} x(t)\theta_k(t)dt \text{ and } s_k = \int_{T_0} s(t)\theta_k(t)dt \tag{2.6}$$

The LR can be expressed as

$$LR = \exp\left[\sum_k \frac{x_k s_k}{\lambda_k} - \frac{1}{2}\sum_k \frac{s_k^2}{\lambda_k}\right]. \tag{2.7}$$

Equivalently, we can write

$$LR = \exp\left[\sum_k \frac{n_k s_k}{\lambda_k} \pm \frac{1}{2}\sum_k \frac{s_k^2}{\lambda_k}\right] \tag{2.8}$$

with + if H_1 is true and - if H_0 is true and both LR expressions are valid only if the quantity

$$d^2 = \sum_k \frac{s_k^2}{\lambda_k} \tag{2.9}$$

is finite. Perfect detection, i.e., detection with zero probability of error, is achieved if d^2 (also known as detectability) diverges and LR becomes ∞ for H_1, and 0 for H_0. Also called singular detection, this condition can be achieved in two ways:

1. If all the eigenvalues are nonzero, then the message signal coefficients, s_k^2, can be chosen to be proportional to λ_k. If $s_k^2 = c_k^2 \lambda_k$ then $d^2 = \sum_k c_k^2$ will diverge if the rate of decay of c_k^2 is slow enough. Two examples are i) $c_k^2 = c^2$ for all k. The signal energy is proportional to $\sum_k \lambda_k$ which is finite since it is the energy of the noise in the observation interval; ii) $c_k^2 = c^2/k$ causes $d^2 = \lim_{N\to\infty} c^2 \sum_k 1/k$ to diverge while keeping signal energy $\sum_k \lambda_k/k < \sum_k \lambda_k$ finite.

2. If the noise covariance has zero eigenvalues, then singular detection can be achieved if at least one signal component, s_j, corresponding to the zero eigenvalue, λ_j, is non-zero.

In all cases of practical interest, detection is non-singular. If the noise is white, then $d^2 = \frac{1}{\lambda} \sum_k s_k^2$. This implies that detectability in white noise is proportional only to the signal-to-noise ratio (SNR) in x(t) where $\lambda = E[n_k^2]$ is the variance or expected value of the energy of each coefficient, n_k, in equation (2.4). For non-white noise, the detectability (d^2) defines a generalized concept of distance which is the inner product of two positive valued (possibly infinite length) vectors $(s_k^2\}$ and $\{1/\lambda_k\}$. In all implementations, infinite series expansions are truncated. Detectability is given by a finite sum

$$d_K^2 = \sum_{k \in K} \frac{s_k^2}{\lambda_k} \le d^2, \tag{2.10}$$

where κ denotes the K element subset $\{k_1, k_2, \ldots, k_K\}$ of the set of positive integers. For a given signal energy, $\xi_s = \sum_{k\in\kappa} s_k^2$, maximum detectability may be found by minimizing d_K^2 subject to the condition of total energy being ξ_s. Applying standard optimization theory results, we set the objective function

$$O(s, Y) = d_K^2 + Y\left(\xi_s - \sum_{k\in\kappa} s_k^2\right) \tag{2.11}$$

and differentiate it with respect to the elements, s_k, of $\mathbf{s}=(s_{k1}, s_{k2} \ldots s_{kK})$ and the Lagrange multiplier Y, d_K^2 is maximized if all the signal energy is concentrated in one coefficient corresponding to the minimum eigenvalue λ_{k1}, where it is assumed that the eigenvalues in equation (2.10) are distinct and ordered so that $\lambda_{k1} < \lambda_{k2} < \ldots < \lambda_{kK}$. We have

$$s_{k_1} = \sqrt{\xi_s}, \quad s_{k_1} = 0 \text{ if } j \neq 1, k_j \in \kappa \tag{2.12}$$

and the Lagrange multiplier $Y=1/\lambda_{k1}$. The optimum value of detectability becomes $d_K^2=\xi_s/\lambda_{k1}$. Increasing detectability also increases the probability of detection in the Neyman-Pearson sense.

Decision is made by comparison of the sufficient statistic, $\sum_k x_k s_k/\lambda_k$, to a threshold determined by the probability of false detection and d^2.

2.3 KL Transform on MRA Subspaces

The dyadic wavelet transform [SGML89] analyzes the space of square integrable (finite energy) functions, L^2, into embedded subspaces that form a multiresolution analysis. Its filter banks structure leads to efficient implementations, thus making it a commonly utilized form of the wavelet transform. It has been shown [GWRL90], [GWAO92], [PFLA92], [ATMK92], that the orthogonal wavelet transform provides a natural setting in which to analyze and synthesize fractional Brownian motion (*fBm*) and other 1/f processes. They show the extent to which wavelets approximate the KL basis functions of 1/f process. The inverse question explores the structure of processes for which wavelets are eigenfunctions. The question is motivated by the wish to characterize such processes so that we may predict the behavior of already existing models, such as *fBms*, under the wavelet transform, develop new models and use them to synthesize innovations for signaling in communication systems and develop detection strategies. In this section, we show that we cannot find the KL basis of a random process by finding the KL bases of its projection onto the MRA subspaces, but we show how these different bases are related. This is done both in the infinite (Section 2.3.1) and finite discrete dimensional case (Section 2.3.2).

When analyzing signals in multiresolution subspaces, it is useful to know how KL bases defined over subspaces $\mathcal{W}_j$, $\mathcal{V}_j$ and $L^2(\mathbf{R})$ are related. The key question is if the decorrelation of coefficients on two disjoint subspaces such as $\mathcal{W}_j$ and $\mathcal{V}_j$ implies the same for coefficients on $\mathcal{V}_{j+1}=\mathcal{W}_j \oplus \mathcal{V}_j$. Another relevant example of disjoint subspaces are $\mathcal{W}_j$ and $\mathcal{W}_{j+1}$. It is not difficult to see that KL bases found for the constituent

subspaces do not comprise a KL basis for their sum. This means that if we are interested in finding a basis that decorrelates all the coefficients $\{d_{jk}\}$, for all k and $0 \le j \le J$, then we need to be working in $\bigcup_{j=0}^{J} \mathcal{W}_j$. We will formulate these concepts for the sake of rigor as well as to gain insight into the problem. We will start with the infinite dimensional case scenario and move on to the finite dimensional afterwards.

2.3.1 Infinite Dimensional Case [NUER99]

We will work in a general setting of subspaces $\mathcal{A}=\mathcal{A}_1 \oplus \mathcal{A}_0$ in $L^2(\mathbf{R})$, where $\mathcal{A}_1$ and $\mathcal{A}_0$ are disjoint. For simplicity of notation, we will use such variables as f(t), y(t), z(t) exclusive of their previously defined meanings. We will assume $f(t) = y(t) + z(t) \in \mathcal{A}$ has orthogonal projections $y(t) \in \mathcal{A}_0$ and $z(t) \in \mathcal{A}_1$. The corresponding covariance functions $r_\bullet(t,u)$ are assumed to have the eigenvalues, $\lambda_k^\bullet$, and orthonormal eigenfunctions $\theta_k^\bullet$, where $\bullet$ means f, y or z. Defining $(\bullet)_k=\langle \bullet, \theta_k^\bullet \rangle$, we write the series expansions

$$f(t) = \sum_{k=1}^{\infty} f_k \theta_k^f(t) \tag{2.13}$$

$$y(t) = \sum_{k=1}^{\infty} y_k \theta_k^y(t) \text{ and } z(t) = \sum_{k=1}^{\infty} z_k \theta_k^z(t) \tag{2.14}$$

where y(t) and z(t) are the orthogonal projections of f(t) onto $\mathcal{A}_0$ and $\mathcal{A}_1$, respectively. Mercer's formula gives

$$r_f(t,u) = \sum_k \lambda_k^f \theta_k^f(t)\theta_k^f(u), \tag{2.15}$$

and similar expressions can be written for $r_\bullet(t,u)$ for y and z. The expansion coefficients are uncorrelated: $E[f_k f_l]=\lambda_k^f \delta_{kl}$, $E[y_k y_l]=\lambda_k^y \delta_{kl}$, and $E[z_k z_l]=\lambda_k^z \delta_{kl}$. Clearly, $\{\theta_k^y(t)\}_k$ and $\{\theta_k^z(t)\}_k$ form complete orthonormal bases, respectively, of $\mathcal{A}_0$ and $\mathcal{A}_1$, and $\{\theta_k^y(t), \theta_k^z(t)\}_k$ and $\{\theta_k^f(t)\}_k$ are two different orthonormal bases spanning $\mathcal{A}$. The covariance function above can be expressed in terms of these bases as

$$\begin{aligned} r_f(t,u) &= \sum_k \lambda_k^y \theta_k^y(t)\theta_k^y(u) + \sum_k \lambda_k^z \theta_k^z(t)\theta_k^z(u) \\ &+ \sum_{kl} E[y_k z_l]\left[\theta_k^y(t)\theta_l^z(u) + \theta_k^y(u)\theta_l^z(t)\right] \end{aligned} \tag{2.16}$$

As long as the signals y(t) and z(t) in the subspaces are correlated, the set $\{\theta_k^y(t), \theta_k^z(t)\}_k$ does not form a KL basis for f(t) on $\mathcal{A}$. Since $\mathcal{A}_0$, $\mathcal{A}_1 \subset \mathcal{A}$, then the following expansions are valid:

$$\theta_k^y(t) = \sum_n q_{fy}(n,k)\theta_n^f(t) \tag{2.17}$$

$$\theta_k^z(t) = \sum_n q_{fz}(n,k)\theta_n^f(t) \tag{2.18}$$

$$\theta_k^f(t) = \sum_n \left(q_{fy}(k,n)\theta_n^y(t) + q_{fz}(k,n)\theta_n^z(t)\right) \tag{2.19}$$

$$\sum_{n} q_{fy}(n,k)q_{fy}(n,l) = \delta_{kl}\,, \tag{2.20}$$

$$\sum_{n} q_{fz}(n,k)q_{fz}(n,l) = \delta_{kl}\,, \tag{2.21}$$

$$\sum_{k}\left(q_{fy}(n,k)q_{fy}(m,k) + q_{fz}(n,k)q_{fz}(m,k)\right) = \delta_{nm}\,, \tag{2.22}$$

$$\sum_{n} q_{fy}(n,k)q_{fz}(n,l) = 0\,. \tag{2.23}$$

Analogous KL expansions can be defined for the discrete coefficients $\{y_k\}_k$, $\{z_k\}_k$ and $\{f_k\}_k$ in terms of eigenvectors of subspaces of $l_2(R)$, the space of square summable or finite energy sequences. We first note that the discretization that generates the coefficients is not linear, as the expansions are given in terms of different bases; hence f_k is not the sum of y_k and z_k, even though f(t) is the sum of y(t) and z(t). Since y(t) and z(t) are orthogonal complements of f(t), then

$$y_k = \left\langle f(t), \theta_k^y(t)\right\rangle \text{ and } z_k = \left\langle f(t), \theta_k^z(t)\right\rangle. \tag{2.24}$$

The resolution of identity or Parseval's rule dictates conservation of energy:

$$\sum_{k} f_k^2 = \sum_{k} y_k^2 + \sum_{k} z_k^2 \tag{2.25}$$

$$\sum_{k} \lambda_k^f = \sum_{k} \lambda_k^y + \sum_{k} \lambda_k^z \tag{2.26}$$

Using equations (2.17) and (2.18) in equation (2.24), we have the decomposition relationships

$$y_k = \sum_n f_n q_{fy}(n,k)\,, \tag{2.27}$$

$$z_k = \sum_n f_n q_{fz}(n,k) \tag{2.28}$$

and the reconstruction relationship

$$f_k = \sum_n y_n q_{fy}(k,n) + \sum_n z_n q_{fz}(k,n) \tag{2.29}$$

Since the elements of each of the sequences y_k and z_k are uncorrelated, using equations (2.27) and (2.28) in $E[y_k y_l]$ and $E[z_k z_l]$, we have

$$\lambda_k^y = \sum_n \lambda_n^f q_{fy}(n,k) q_{fy}(n,k) \tag{2.30}$$

$$\lambda_k^z = \sum_n \lambda_n^f q_{fz}(n,k) q_{fz}(n,k) \tag{2.31}$$

The correlation between y_k and z_l can be expressed by

$$E[y_k z_l] = \sum_n \lambda_n^f q_{fy}(n,k) q_{fz}(n,l)\,. \tag{2.32}$$

Using equation (2.29) in $E[f_k f_l]$, we also obtain

$$\begin{aligned} \lambda_k^f \delta_{kl} &= \sum_n \lambda_n^y q_{fy}(k,n) q_{fy}(l,n) + \sum_n \lambda_n^z q_{fz}(k,n) q_{fz}(l,n) \\ &+ \sum_{nm} E[y_n z_m]\left[q_{fy}(k,n) q_{fz}(l,m) + q_{fy}(l,n) q_{fz}(k,m)\right] \end{aligned} \tag{2.33}$$

These results can be expressed using infinite dimensional matrices. We define vector $\mathbf{u}=[y_1y_2\ldots y_n,\ldots z_1z_2\ldots z_n,\ldots]^T$ and $\mathbf{f}=[f_1f_2\ldots f_n\ldots]^T$. The respective correlation matrices are $\Lambda_f=E[\mathbf{ff}^T]=\text{diag}(\lambda_1^f\lambda_2^f\ldots\lambda_n^f)$ and $\Lambda_u=E[\mathbf{uu}^T]$

$$\mathbf{R}_u=\begin{bmatrix}\mathbf{\Lambda}_y & \mathbf{R}\\ \mathbf{R}^T & \mathbf{\Lambda}_z\end{bmatrix} \tag{2.34}$$

where $\Lambda_y=\text{diag}(\lambda_1^y\lambda_2^y\ldots\lambda_n^y)$, $\Lambda_z=\text{diag}(\lambda_1^z\lambda_2^z\ldots\lambda_n^z)$, and $[\mathbf{R}]_{mn}=E[y_mz_n]$ for $m,n\in\{1,2,\ldots\}$. From equations (2.30), (2.31), (2.32), we have $\mathbf{R}_u=\mathbf{Q}\Lambda_f\mathbf{Q}^T$ where

$$\mathbf{Q}=\begin{bmatrix}
q_{fy}(1,1) & q_{fy}(1,2) & \cdots & q_{fy}(1,n) & \cdots\\
\vdots & \vdots & \ddots & \vdots & \cdots\\
q_{fy}(n,1) & q_{fy}(n,2) & \cdots & q_{fy}(n,n) & \cdots\\
\vdots & \vdots & \ddots & \vdots & \cdots\\
q_{fz}(1,1) & q_{fz}(1,2) & \cdots & q_{fz}(1,n) & \cdots\\
\vdots & \vdots & \ddots & \vdots & \cdots\\
q_{fz}(n,1) & q_{fz}(n,2) & \cdots & q_{fz}(n,n) & \cdots\\
\vdots & \vdots & \ddots & \vdots & \cdots
\end{bmatrix} \tag{2.35}$$

By equation (2,22), $\mathbf{Q}^{-1}=\mathbf{Q}^T$, i.e. $\mathbf{Q}$ represents a unitary operation thus we also have $\Lambda_f=\mathbf{Q}^T\mathbf{R}_u\mathbf{Q}$. Clearly, $\lambda_\bullet^f$ are the eigenvalues and the columns of $\mathbf{Q}$ are the eigenvectors of $\mathbf{R}_u$. The transformation $\mathbf{f}=\mathbf{Q}^T\mathbf{u}$ is the KL transform of $\mathbf{u}$. Analogous results are obtained in the finite dimensional case.

2.3.2 Finite Discrete Dimensional Case

For the finite dimensional case with a discrete input sequence projected in the subspaces, we will concentrate on the wavelet subspaces and derive relationships for the eigenvalues and eigenvectors between each space. These results can be easily applied to any set of subspaces. In real life this is the case that can be implemented or simulated and experimental results can be produced. This is the scenario we used for our simulations as described in Chapter 3.

The work will be done on the basic analysis-synthesis filter structure of the wavelet transform as shown in Figures 1.11 and 1.14, but can be easily extrapolated for any decomposition depth. We will define all quantities and operations in matrix and vector form, starting with structures in Figures 1.11 and 1.14 that can be represented as in Figure 2.1. Notice that the notations on this Figure represent discrete signals and decomposition in a matrix-vector form. There are other ways of representing wavelet

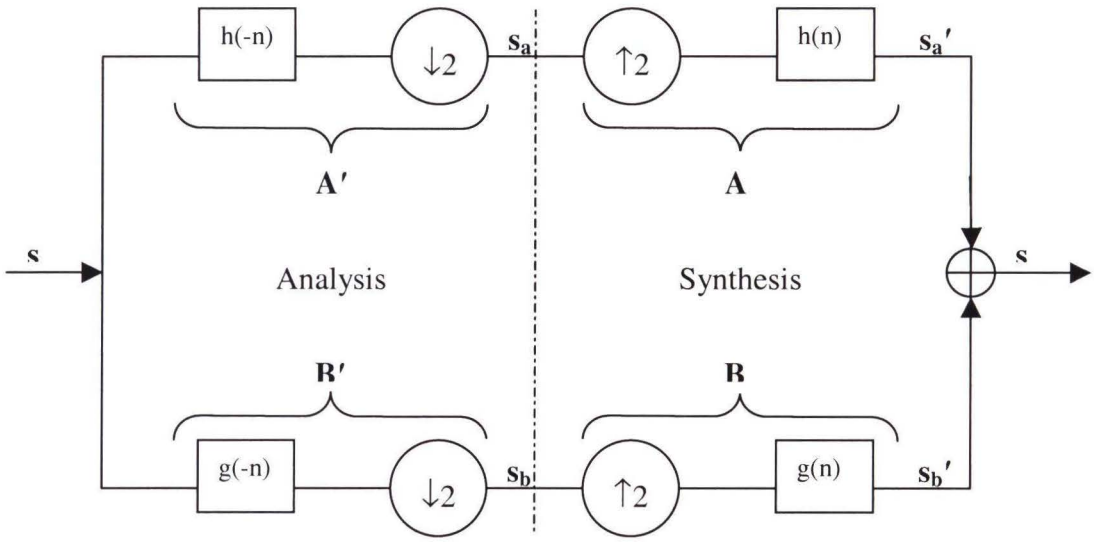


Figure 2.1: Analysis/Synthesis for a Two-Band Filter Bank.

transformations. Use of this format allows us to compute the covariance matrices of projections and coefficients, which are needed for further analysis. It also shows that it

is sufficient to limit the analysis to a two-stage transform that will be repeated as we go along the scales. With the help of Figure 2.1 let us define an input vector $\mathbf{s}=[s_1\ s_2\ \ldots\ s_N]^T$ and coefficients $\mathbf{s_a}=[s_{a1}\ s_{a2}\ \ldots\ s_{aM}]^T$, and $\mathbf{s_b}=[s_{b1}\ s_{b2}\ \ldots\ s_{bM}]^T$. $N = 2^n$ is the total input signal samples for any integer n (which is chosen depending on the decomposition levels we are required to go), and M=N/2 (due to the down-sampler). Matrices **A**, **B**, **A′**, **B′** are derived from the analysis and synthesis filter coefficients that result from the wavelet and scaling functions. Operations **A′**, **B′** represent the discrete convolution of the signal with filter coefficients h(-n) and g(-n) at their input followed by down-sampling by two. Thus their size is MxN. Operations **A**, **B** represent the up-sampling by two followed by discrete convolution of the signal with filter coefficients h(n) and g(n) at their input and the down sampling. Thus their size is NxM. Because of the h(-n), h(n), g(n), g(-n) coefficient relationships (not true for all wavelets i.e. biorthogonal wavelets) one can easily show that $\mathbf{A'}= \mathbf{A}^T$ and $\mathbf{B'}=\mathbf{B}^T$. As an example the filter coefficients for the db3 wavelet have a constant length of six and they are

$$h(-n) = [0.0352 \quad -0.0854 \quad -0.1350 \quad 0.4599 \quad 0.8069 \quad 0.3327] \tag{2.36a}$$

$$g(-n) = [-0.3327 \quad 0.8069 \quad -0.4599 \quad -0.1350 \quad 0.0854 \quad 0.0352]. \tag{2.36b}$$

For a length N=8, this results to matrices **A**, and **B** as

$$\mathbf{A}=\begin{bmatrix} 0.4599 & 0.3327 & 0 & -0.0854 \\ -0.1350 & 0.8069 & 0 & 0.0352 \\ -0.0854 & 0.4599 & 0.3327 & 0 \\ 0.0352 & -0.1350 & 0.8069 & 0 \\ 0 & -0.0854 & 0.4599 & 0.3327 \\ 0 & 0.0352 & -0.1350 & 0.8069 \\ 0.3327 & 0 & -0.0854 & 0.4599 \\ 0.8069 & 0 & 0.0352 & -0.1350 \end{bmatrix} \tag{2.37a}$$

$$\mathbf{B}=\begin{bmatrix} -0.1350 & 0.0352 & 0 & 0.8069 \\ -0.4599 & 0.0854 & 0 & -0.3327 \\ 0.8069 & -0.1350 & 0.0352 & 0 \\ -0.3327 & -0.4599 & 0.0854 & 0 \\ 0 & 0.8069 & -0.1350 & 0.0352 \\ 0 & -0.3327 & -0.4599 & 0.0854 \\ 0.0352 & 0 & 0.8069 & -0.1350 \\ 0.0854 & 0 & -0.3327 & -0.4599 \end{bmatrix}. \tag{2.37b}$$

The function for generating these matrices for any wavelet and any signal length N was created using Matlab.

The signals $\mathbf{s_a}'$ and $\mathbf{s_b}'$ are of size Nx1 and if added they should produce $\mathbf{s}$ (perfect reconstruction). We will now write the relationships between the signals and derive some filter properties that we are going to use later on. First we start with expressing the analysis signals as

$$\mathbf{s_a} = \mathbf{A}^T\mathbf{s} \tag{2.38a}$$

$$\mathbf{s_b} = \mathbf{B}^T\mathbf{s} \tag{2.38b}$$

and the synthesis signals as

$$\mathbf{s_a}' = \mathbf{A}\mathbf{s_a} = \mathbf{A}\mathbf{A}^T\mathbf{s} \tag{2.39a}$$

$$\mathbf{s_b}' = \mathbf{B}\mathbf{s_b} = \mathbf{B}\mathbf{B}^T\mathbf{s}. \tag{2.39b}$$

Also using equation (2.39) we require that

$$\mathbf{s} = \mathbf{s_a}' + \mathbf{s_b}' = (\mathbf{A}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T)\mathbf{s} \tag{2.40}$$

which means that

$$\mathbf{A}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = \mathbf{I} \tag{2.41}$$

for perfect signal reconstruction. We can also write from equation (2.38)

$$\mathbf{s_a} = \mathbf{A}^T(\mathbf{A}\mathbf{s_a} + \mathbf{B}\mathbf{s_b}) = \mathbf{A}^T\mathbf{A}\mathbf{s_a} + \mathbf{A}^T\mathbf{B}\mathbf{s_b} \tag{2.42a}$$

$$\mathbf{s_b} = \mathbf{B}^T(\mathbf{A}\mathbf{s_a} + \mathbf{B}\mathbf{s_b}) = \mathbf{B}^T\mathbf{A}\mathbf{s_a} + \mathbf{B}^T\mathbf{B}\mathbf{s_b}. \tag{2.42b}$$

Using these relationships we can write another three properties of the **A**, **B** matrices as

$$\mathbf{A}^T\mathbf{A} = \mathbf{I} \tag{2.43a}$$

$$\mathbf{B}^T\mathbf{B} = \mathbf{I} \tag{2.43b}$$

$$\mathbf{A}^T\mathbf{B} = \mathbf{0}, \mathbf{B}^T\mathbf{A} = \mathbf{0}. \tag{2.43c}$$

Let $\mathbf{R_s} = E[\mathbf{s}\mathbf{s}^T]$ be the NxN covariance matrix of **s**. We can define similarly $\mathbf{R_a} = E[\mathbf{s_a}\mathbf{s_a}^T]$, $\mathbf{R_b} = E[\mathbf{s_b}\mathbf{s_b}^T]$, $\mathbf{R_a}' = E[\mathbf{s_a}'\mathbf{s_a}'^T]$, $\mathbf{R_b}' = E[\mathbf{s_b}'\mathbf{s_b}'^T]$, $\mathbf{R_{ab}} = E[\mathbf{s_a}\mathbf{s_b}^T]$, and $\mathbf{R_{ba}} = E[\mathbf{s_b}\mathbf{s_a}^T]$. A natural assumption for the covariance matrices $\mathbf{R_s}$, $\mathbf{R_a}$, $\mathbf{R_b}$, $\mathbf{R_a}'$, and $\mathbf{R_b}'$, is that they are symmetric positive semi-definite matrices. The requirements for a matrix to be positive

semi-definite and their corresponding properties are listed in Appendix III. Substituting the signals from equations (2.38), and (2.39) to these relationships, we can write the coefficients and projection covariance matrices with respect to the covariance matrix of the input signal as follows

$$\mathbf{R}_a = E[\mathbf{A}^T\mathbf{s}(\mathbf{A}^T\mathbf{s})^T] = E[\mathbf{A}^T\mathbf{s}\mathbf{s}^T\mathbf{A})] = \mathbf{A}^T E[\mathbf{s}\mathbf{s}^T]\mathbf{A} = \mathbf{A}^T\mathbf{R}_s\mathbf{A} \quad (2.44a)$$

$$\mathbf{R}_b = E[\mathbf{B}^T\mathbf{s}(\mathbf{B}^T\mathbf{s})^T] = E[\mathbf{B}^T\mathbf{s}\mathbf{s}^T\mathbf{B})] = \mathbf{B}^T E[\mathbf{s}\mathbf{s}^T]\mathbf{B} = \mathbf{B}^T\mathbf{R}_s\mathbf{B} \quad (2.44b)$$

$$\mathbf{R}_{ab} = E[\mathbf{A}^T\mathbf{s}(\mathbf{B}^T\mathbf{s})^T] = E[\mathbf{A}^T\mathbf{s}\mathbf{s}^T\mathbf{B})] = \mathbf{A}^T E[\mathbf{s}\mathbf{s}^T]\mathbf{B} = \mathbf{A}^T\mathbf{R}_s\mathbf{B} \quad (2.44c)$$

$$\mathbf{R}_{ba} = E[\mathbf{B}^T\mathbf{s}(\mathbf{A}^T\mathbf{s})^T] = E[\mathbf{B}^T\mathbf{s}\mathbf{s}^T\mathbf{A})] = \mathbf{B}^T E[\mathbf{s}\mathbf{s}^T]\mathbf{A} = \mathbf{B}^T\mathbf{R}_s\mathbf{A} = \mathbf{R}_{ab}{}^T \quad (2.44d)$$

$$\mathbf{R}_a' = E[\mathbf{A}\mathbf{A}^T\mathbf{s}(\mathbf{A}\mathbf{A}^T\mathbf{s})^T] = E[\mathbf{A}\mathbf{A}^T\mathbf{s}\mathbf{s}^T\mathbf{A}\mathbf{A}^T)] = \mathbf{A}\mathbf{A}^T\mathbf{R}_s\mathbf{A}\mathbf{A}^T \quad (2.44e)$$

$$\mathbf{R}_b' = E[\mathbf{B}\mathbf{B}^T\mathbf{s}(\mathbf{B}\mathbf{B}^T\mathbf{s})^T] = E[\mathbf{B}\mathbf{B}^T\mathbf{s}\mathbf{s}^T\mathbf{B}\mathbf{B}^T)] = \mathbf{B}\mathbf{B}^T\mathbf{R}_s\mathbf{B}\mathbf{B}^T. \quad (2.44f)$$

More importantly though (as we will demonstrate in Chapter 4) would be to find an inverse relationship. In other words knowing the covariance matrices of the coefficients (at the next decomposition level) find the covariance matrix of the input signal. With the use of Equations (2.39) and (2.40) we can write

$$\mathbf{R}_s = E[\mathbf{s}\mathbf{s}^T] = E[(\mathbf{A}\mathbf{s}_a+\mathbf{B}\mathbf{s}_b)(\mathbf{A}\mathbf{s}_a+\mathbf{B}\mathbf{s}_b)^T] = E[(\mathbf{A}\mathbf{s}_a+\mathbf{B}\mathbf{s}_b)(\mathbf{s}_a{}^T\mathbf{A}^T+\mathbf{s}_b{}^T\mathbf{B}^T)] =$$

$$= \mathbf{A}\mathbf{R}_a\mathbf{A}^T+\mathbf{A}\mathbf{R}_{ab}\mathbf{B}^T+\mathbf{A}\mathbf{R}_{ba}\mathbf{A}^T+\mathbf{B}\mathbf{R}_b\mathbf{B}^T. \quad (2.45)$$

This Equation can be written in a matrix form as follows

$$\mathbf{R}_s = [\mathbf{A} \quad \mathbf{B}] \underbrace{\begin{bmatrix} \mathbf{R}_a & \mathbf{R}_{ab} \\ \mathbf{R}_{ab}^T & \mathbf{R}_b \end{bmatrix}}_{\mathbf{R}_1} \begin{bmatrix} \mathbf{A}^T \\ \mathbf{B}^T \end{bmatrix}. \tag{2.46}$$

Let $\mathbf{R}_s = \mathbf{Q}_s\Lambda_s\mathbf{Q}_s^T$, $\mathbf{R}_a = \mathbf{Q}_a\Lambda_a\mathbf{Q}_a^T$, and $\mathbf{R}_b = \mathbf{Q}_b\Lambda_b\mathbf{Q}_b^T$ be the KL transform of $\mathbf{R}_a$ and $\mathbf{R}_b$, where $\mathbf{Q}_\bullet$ and $\Lambda_\bullet$ are the associate eigenvectors and eigenvalues respectively. Then $\mathbf{R}_s$ can be written as

$$\mathbf{R}_s = [\mathbf{A} \quad \mathbf{B}] \begin{bmatrix} \mathbf{Q}_a\Lambda_a\mathbf{Q}_a^T & \mathbf{R}_{ab} \\ \mathbf{R}_{ab}^T & \mathbf{Q}_b\Lambda_b\mathbf{Q}_b^T \end{bmatrix} \begin{bmatrix} \mathbf{A}^T \\ \mathbf{B}^T \end{bmatrix} \tag{2.47}$$

or

$$\mathbf{R}_s = [\mathbf{A} \quad \mathbf{B}] \begin{bmatrix} \mathbf{Q}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_b \end{bmatrix} \underbrace{\begin{bmatrix} \Lambda_a & \overline{\mathbf{R}}_{ab} \\ \overline{\mathbf{R}}_{ab}^T & \Lambda_b \end{bmatrix}}_{\mathbf{R}_2} \begin{bmatrix} \mathbf{Q}_a^T & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_b^T \end{bmatrix} \begin{bmatrix} \mathbf{A}^T \\ \mathbf{B}^T \end{bmatrix} \tag{2.48}$$

where $\overline{\mathbf{R}}_{ab} = \mathbf{Q}_a^T\mathbf{R}_{ab}\mathbf{Q}_b$. Since

$$[\mathbf{A} \quad \mathbf{B}] \begin{bmatrix} \mathbf{A}^T \\ \mathbf{B}^T \end{bmatrix} = \mathbf{I} \text{ and } \begin{bmatrix} \mathbf{Q}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_b \end{bmatrix} \begin{bmatrix} \mathbf{Q}_a^T & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_b^T \end{bmatrix} = \mathbf{I} \tag{2.49}$$

or otherwise called similarity transforms, the eigenvalues of $\mathbf{R}_1$ are the same as the ones of $\mathbf{R}_2$ and $\mathbf{R}_s$. Let $\mathbf{R}_1 = \mathbf{Q}_1\Lambda_1\mathbf{Q}_1^T$, and $\mathbf{R}_2 = \mathbf{Q}_2\Lambda_2\mathbf{Q}_2^T$, then the eigenvectors of $\mathbf{R}_s$ can be evaluated as

$$\mathbf{Q}_s = [\mathbf{A} \quad \mathbf{B}]\mathbf{Q}_1 = [\mathbf{A} \quad \mathbf{B}] \begin{bmatrix} \mathbf{Q}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_b \end{bmatrix} \mathbf{Q}_2 \tag{2.50}$$

In Chapter 4 we are going to show ways of taking advantage of expressions (2.46), (2.48) and (2.50) for calculating the eigenvalues and eigenvectors of $\mathbf{R}_s$ for optimum detection.

Now will try to relate the eigenvalues of $\mathbf{R}_a'$ and $\mathbf{R}_a$. The eigenvalues of the covariance matrix $\mathbf{R}_a'$ are the roots of the polynomial

$$|\lambda \mathbf{I} - \mathbf{R}_a'| = |\lambda \mathbf{I} - \mathbf{A}\mathbf{R}_a'\mathbf{A}^T| \tag{2.51a}$$

which are the same as the roots of

$$|\mathbf{R}_a|\,|\lambda \mathbf{I} - \mathbf{A}\mathbf{R}_a\mathbf{A}^T|. \tag{2.51b}$$

Applying to this, the identity

$$|\mathbf{D}||\mathbf{E} - \mathbf{C}\mathbf{D}^{-1}\mathbf{F}| = |\mathbf{E}||\mathbf{D} - \mathbf{F}\mathbf{E}^{-1}\mathbf{C}| \tag{2.52}$$

with $\mathbf{D} = \mathbf{R}_a^{-1}$, $\mathbf{E} = \lambda\mathbf{I}$, $\mathbf{C} = \mathbf{A}$, and $\mathbf{F} = \mathbf{A}^T$, yields

$$|\mathbf{R}_a||\lambda \mathbf{I} - \mathbf{A}\mathbf{R}_a\mathbf{A}^T| = |\lambda \mathbf{I}|\left|\mathbf{R}_a^{-1} - \frac{1}{\lambda}\mathbf{A}^T\mathbf{A}\right| = |\lambda \mathbf{I}||\lambda \mathbf{R}_a^{-1} - \mathbf{A}^T\mathbf{A}| = |\lambda \mathbf{I}||\mathbf{R}_a^{-1}||\lambda \mathbf{I} - \mathbf{R}_a| \tag{2.53}$$

where use has been made of $\mathbf{A}^T\mathbf{A} = \mathbf{I}$.

This proves that the eigenvalues of $\mathbf{R}_a'$ are identical to the eigenvalues of $\mathbf{R}_a$, and the remaining N-M eigenvalues are zero. Similar results are true for matrix $\mathbf{R}_b'$ are identical to the eigenvalues of $\mathbf{R}_b$. This means that matrices $\mathbf{R}_b'$, and $\mathbf{R}_a'$ although of size NxN they are of rank M.

2.4 Detectability Over Scales

For our discussion we will consider the discrete finite signal case under wavelet subspaces. A binary detection problem [HVRS68] of a known signal $\mathbf{s}=[s[1]\ \ s[2]\ \ ...\ \ s[N]]^T$, in noise, **n,** is described by two hypotheses: $H_1 : \mathbf{x} = \mathbf{s}+\mathbf{n}$ and $H_0 : \mathbf{x} = \mathbf{n}$ where **n** is a zero-mean Gaussian random noise. The noise autocorrelation matrix **R** has the EVD decomposition $\mathbf{R}=\mathbf{Q}\Lambda\mathbf{Q}^T$, where Λ is the diagonal matrix of eigenvalues, λ_k. The likelihood ratio (LR) test is derived from probabilistic considerations to minimize the risk of making a wrong decision. Expressions for LR are useful and efficient when stated in terms of noise statistics. The difference between LR under the two hypotheses is used as a measure of detectability, given by

$$d^2 = \sum_{k=1}^{N} \frac{|s_k|^2}{\lambda_k}, \tag{2.54}$$

where $\mathbf{s_k}$ are the elements of $\mathbf{Q}^T\mathbf{s}$. The sufficient statistic required for making a decision is the first term of the exponent of LR as in equation

$$g = \sum_{k=1}^{N} \frac{x_k s_k}{\lambda_k}. \tag{2.55}$$

Decision is made by comparing g to a threshold determined by the probability of false detection and d^2. Given by Equation (2.54), detectability is a metric of distance between the vector $\mathbf{s_k}$ and $\gamma = [1/\lambda_1, 1/\lambda_2, ..., 1/\lambda_N]$ where $\mathbf{s_k}$ is the KL transform (with respect to the noise EVD) of the signal **s.** If noise is white, then d^2 is the signal-to-noise ratio,

otherwise it is a quantity between 0 and infinity and has to do with the relative distribution of the signal and eigenvalues. The relationship between detectability at scales j and j+1 is based on the principle of energy conservation. For simplicity, we will consider the one step scale decomposition of Figure 2.1. Let the KL-transform of the signal coefficients $\mathbf{s}$, $\mathbf{s_b}$ and $\mathbf{s_a}$ with respect to the noise EVD be given as $\mathbf{s_k}$, $\mathbf{s_{bk}}$ and $\mathbf{s_{ak}}$ respectively. Also, let the vectors of noise eigenvalues be given as $\boldsymbol{\lambda}, \boldsymbol{\lambda}^a, \boldsymbol{\lambda}^b$ respectively. We have

$$\sum_{k=1}^{N} \lambda_k = \sum_{k=1}^{M} \lambda_k^a + \sum_{k=1}^{M} \lambda_k^b , \tag{2.56}$$

$$\sum_{k=1}^{N} s_k^2 = \sum_{k=1}^{M} [s_{ak}^2 + s_{bk}^2] , \text{ and} \tag{2.57}$$

$$d^2 = \sum_{k=1}^{N} \frac{s_k^2}{\lambda_k} , \quad d_i^2 = \sum_{k=1}^{M} \frac{s_{ik}^2}{\lambda^i{}_k} , \text{ for i=a, b.} \tag{2.58}$$

If noise is white, the detectability in parent-children subspaces are related by

$$d^2 = \frac{\sigma_a^2}{\sigma^2} d_a{}^2 + \frac{\sigma_b^2}{\sigma^2} d_b{}^2 \tag{2.59}$$

where $\sigma^2 = \sigma_a^2 + \sigma_b^2$ are the respective noise variances. Clearly, detectability in any subspace is highly affected by the noise eigenvalues. A redistribution causing one of them to be close to zero may result in near perfect detection in the subspace. This is a very encouraging reason to do detection over multiresolution subspaces.

The detectability can also be expressed as a function of the noise autocorrelation matrix. Equation (2.54) can be written in matrix form as

$$d^2 = \mathbf{s_k}^T \Lambda^{-1} \mathbf{s_k} . \tag{2.60}$$

Substituting $\mathbf{s_k}$ in this equation we get

$$d^2 = \mathbf{s}^T \mathbf{Q} \Lambda^{-1} \mathbf{Q}^T \mathbf{s} = \mathbf{s}^T \mathbf{R}^{-1} \mathbf{s}. \tag{2.61}$$

Similarly we can write expressions for the detectability for the coefficients and the projections. Also we can substitute equations (2.38), (2.39), and (2.44) and the result is

$$d_a^2 = \mathbf{s_a}^T \mathbf{R_a}^{-1} \mathbf{s_a} = \mathbf{s}^T \mathbf{A} \mathbf{R_a}^{-1} \mathbf{A}^T \mathbf{s} = \mathbf{s}^T \mathbf{A} (\mathbf{A}^T \mathbf{R} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{s} \tag{2.59a}$$

$$d_b^2 = \mathbf{s_b}^T \mathbf{R_b}^{-1} \mathbf{s_b} = \mathbf{s}^T \mathbf{B} \mathbf{R_b}^{-1} \mathbf{B}^T \mathbf{s} = \mathbf{s}^T \mathbf{B} (\mathbf{B}^T \mathbf{R} \mathbf{B})^{-1} \mathbf{B}^T \mathbf{s} \tag{2.59b}$$

In the next Chapter we will go in depth through the different scenarios that were assumed for our simulations. Also we will present amongst other results, detectability figures for a multi level subspace detection scheme and ROC graphs to verify them.

CHAPTER 3

DETECTION OVER MULTIRESOLUTION SUBSPACES-SIMULATIONS AND RESULTS

3.1 Subspace Detection Scheme

An example of the multiresolution system to be used in our simulations is the full binary tree, shown in Figure 3.1 for two levels of decomposition. This system was proposed by Ronald Coifman. It allows decomposition at high as well as low frequencies, for a complete, evenly spaced frequency coverage, unlike the traditional, constant-Q wavelet filter bank structure arising from a half-tree decomposition which was shown in Figure 1.12. Note that the synthesis part shown in this figure is not unique. $\mathcal{V}_{ij}$ and $\mathcal{W}_{ij}$ are the various projections on the subspaces and can be visualized as shown in Figure 3.2. We note that the subscript i denotes the level of decomposition (depth in the tree) starting with the 0^{th} level at the tree input, and the subscript j denotes the branch and takes values from 0 to 2^i-1. The j = 0 and the $j=2^i$-1 branch represent the lowest and highest frequency bands, respectively. This notation of the nodes in the tree is the method we use to refer to any signal-point at the tree in the rest of the Chapter.

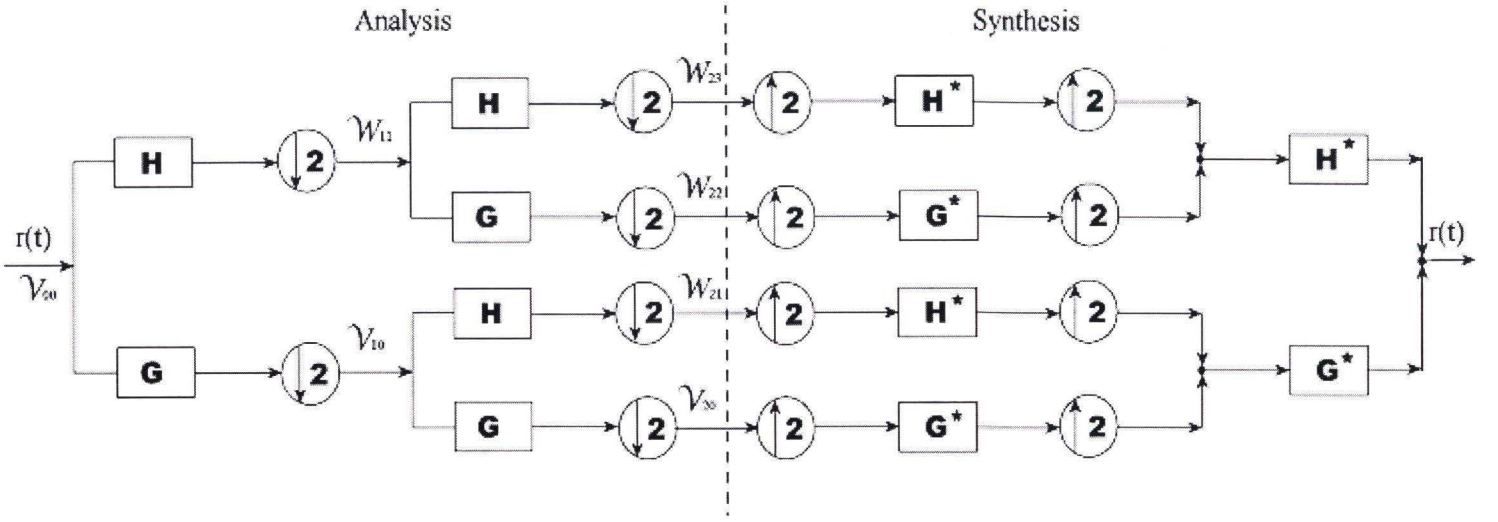


Figure 3.1: The Full Binary Tree for the Two-scale Wavelet Packet Transform.

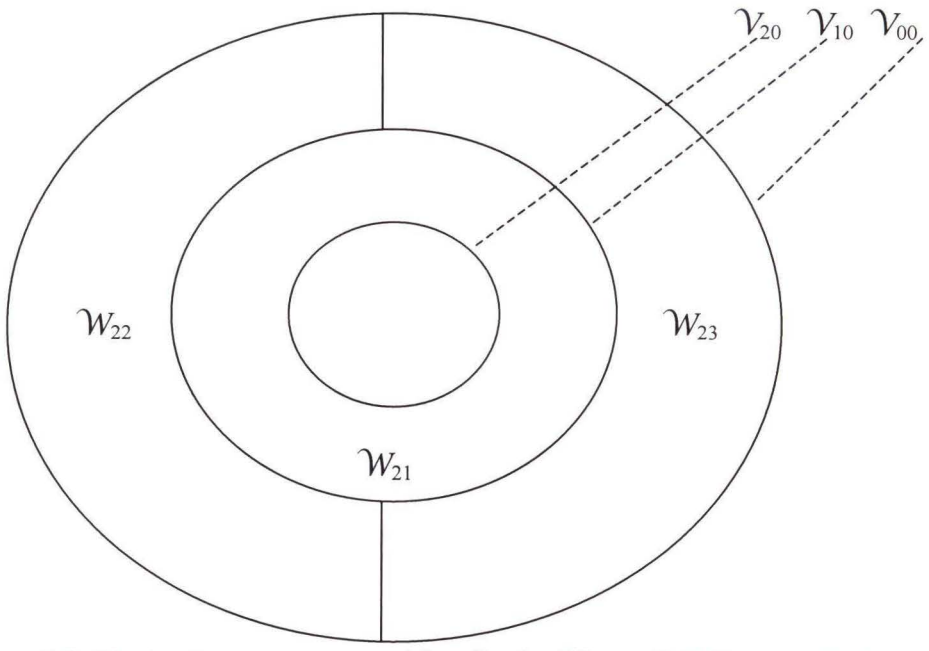


Figure 3.2: Vector Space Decomposition for the Binary Full Wavelet Packet System.

Using this decomposition scheme gives us flexibility as to what subsets we want to choose as basis, depending on the signal and noise statistics, according to some optimization criterion (e.g. entropy which results in the 'best wavelet packet basis algorithm' [BSRG98], [MMYM96]). Any pruning of this full tree could generate a valid packet basis system and would allow very flexible tiling of the time-scale plane, which is very useful especially in the analysis of non-stationary signals. In our case the best basis would be the basis with maximum detectability, and best basis set will be the one that leads to the best basis. The filters used have low-pass (g(n)) and high-pass (h(n)) characteristics and are derived in accordance with Mallat's algorithm for filter design using wavelet basis functions. As discussed in Chapter 1.4, for perfect reconstruction to occur with orthogonal wavelets, these filter relationships must be true: $h^*(n) = h(-n)$ and $g^*(n) = g(-n)$, where h(n), g(n) are filter coefficients associated with blocks marked H^*,

G^*. In an attempt to visualize the projections in the frequency domain, Figure 3.3 shows an arbitrary frequency response of the two-band wavelet packet filter bank of Figure 3.1.

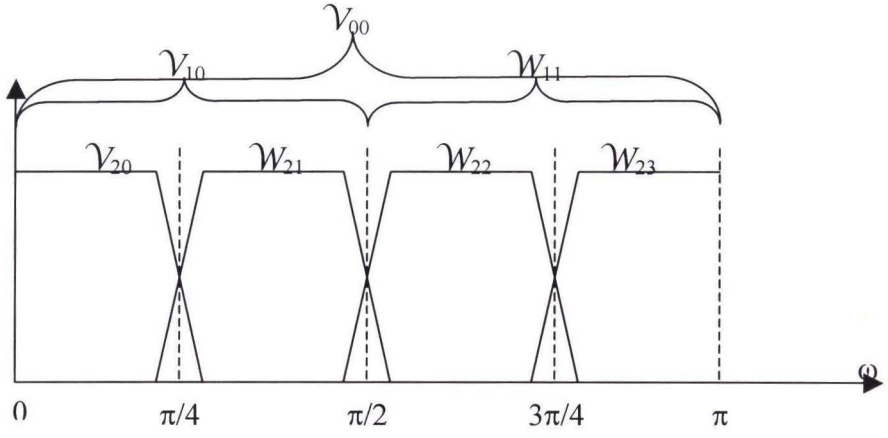


Figure 3.3: Frequency Responses for the Two-band Wavelet Packet Filter Bank.

Using this subspace decomposition structure, the final stage of the proposed detector will be a whitening filter, a correlator and the threshold stage for decision

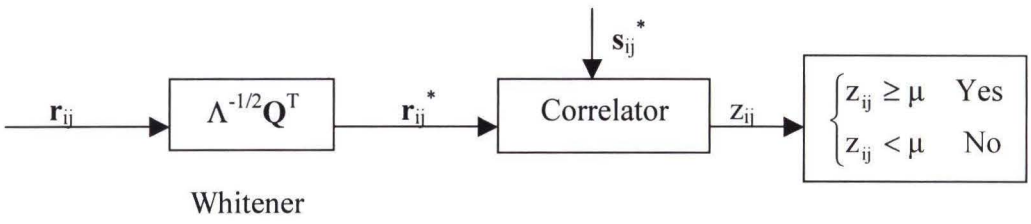


Figure 3.4: Final Stage of Wavelet Detector.

The vector $\mathbf{r}_{ij}$ is the signal plus noise vector, $\mathbf{s}_{ij} + \mathbf{n}_{ij}$, at the (i, j) node of the wavelet decomposition tree (i.e. i^{th} level j^{th} branch). The branch used for detection is the one that yields maximum detectability. The star (*) represents a whitened signal based on the noise statistics as

$$\mathbf{r}_{ij}^{*} = \Lambda_{ij}^{-1/2}\mathbf{Q}_{ij}^{T}\mathbf{r}_{ij} = \mathbf{s}_{ij}^{*} + \mathbf{n}_{ij}^{*} \tag{3.1}$$

where Λ_{ij},and $\mathbf{Q}_{ij}$ are the associated eigenvalue and eigenvector matrices of the covariance matrix of $\mathbf{n}_{ij}$ denoted by $\mathbf{R}_{ij} = E[\mathbf{n}_{ij}\mathbf{n}_{ij}^{T}]$ (i.e $\mathbf{R}_{ij} = \mathbf{Q}_{ij}\Lambda_{ij}\mathbf{Q}_{ij}^{T}$). This whitening strategy is the optimum one for any noise and it can be shown that the detectability does not change before and after whitening. This proof, as well as how the signal to noise ratio is affected from the whitening stage, are given in Appendix I. The resulting scalar, z_{ij}, at the output of the correlator is the inner product of $\mathbf{s}_{ij}^{*}$ and $\mathbf{r}_{ij}^{*}$ ($<\mathbf{s}_{ij}^{*}, \mathbf{r}_{ij}^{*}> = \mathbf{s}_{ij}^{*T}\mathbf{r}_{ij}^{*}$).

The complexity of the eigenvalue decomposition process is proportional to the size of $\mathbf{R}_{ij}$ or the length of vector $\mathbf{n}_{ij}$. Thus by going down in scale or performing optimum detection at deeper levels into the wavelet decomposition tree, the covariance matrix size becomes smaller and smaller (elements reduced by a factor of four at each consecutive scale). This results in a significant advantage, in computational complexity, in the estimation of the eigenvalues and the eigenvectors for optimum detection. Using the subspace structure this advantage can be amplified, as will be demonstrated in Chapter 4. Also the EVD of large matrices that can result in unstable results is avoided. Another advantage, in real time applications, of using the wavelet structure or any multirate system for that matter, is that performing any sort of processing in lower levels

(lower rates) requires less processing speed (at each consecutive level, the processor speed can be reduced by a factor of two) [PPVA93].

Obviously the estimation of the covariance matrix $\mathbf{R}_{ij}$ is crucial for optimum detection. The problem becomes even more complicated if no previous knowledge of the noise statistics are accessible or if the noise is non-stationary or time varying. In our simulations this is the assumed case. In this scenario the covariance matrix should be updated as every new vector $\mathbf{n}_{ij}^{(k)}$ arrives at the whiteners input (adaptively estimated), where k is a discrete time counter starting at zero. A thorough and comprehensive treatment of the different techniques for updating EVD and Singular Value Decompositions (SVD) can be found in [TMWH99]. Update techniques surveyed share a common objective: To determine the 'best' way to find the new covariance matrix given the current covariance matrix and new data. Does the rank of the update go up, down or stay the same? Increasing the rank, computing the eigenvalues and then reducing the rank as determined by the dominant eigenvalues is theoretically risk free, but computationally very expensive. In practice, increase in size of the covariance matrix subjects the EVD to numerical instabilities. On the other hand, confining the rank to be a small number, even when it is coupled by dominant subspace tracking, does not take into consideration the long-term correlations of the data. In our simulations we have assumed the initial value of $\mathbf{R}_{ij}^{(0)} = \mathbf{n}_{ij}^{(0)}\mathbf{n}_{ij}^{(0)T}$ and the method for updating the covariance matrix to be as

$$\mathbf{R}_{ij}^{(k+1)} = b\mathbf{R}_{ij}^{(k)} + (1-b)\mathbf{n}_{ij}^{(k+1)}\mathbf{n}_{ij}^{(k+1)T} \tag{3.2}$$

where $0<b<1$ and it is called the forgetting factor [SIHA96]. The slower the noise statistics vary with time the closer b is to one. We also confined the matrix rank to the length of the

where 0<b<1 and it is called the forgetting factor [SIHA96]. The slower the noise statistics vary with time the closer b is to one. We also confined the matrix rank to the length of the noise vector at each node. For every adaptation or update of the covariance matrix, the new eigenvectors and eigenvalues must be calculated. A typical example of how the eigenvalues vary with time is illustrated in Figure 3.5. This is a sample computation of the eigenvalues at node (1,0) for ocean noise. We note that the dominant eigenvalue fluctuates significantly in time at that node, and this observation stresses the need for tracking these eigenvalues over time, for this noise, to achieve optimum detection. This simulation was done with a forgetting factor b equal to 0.95. The noise vectors were of length 32 at that node, but for illustration purposes we show only the four dominant eigenvalues.

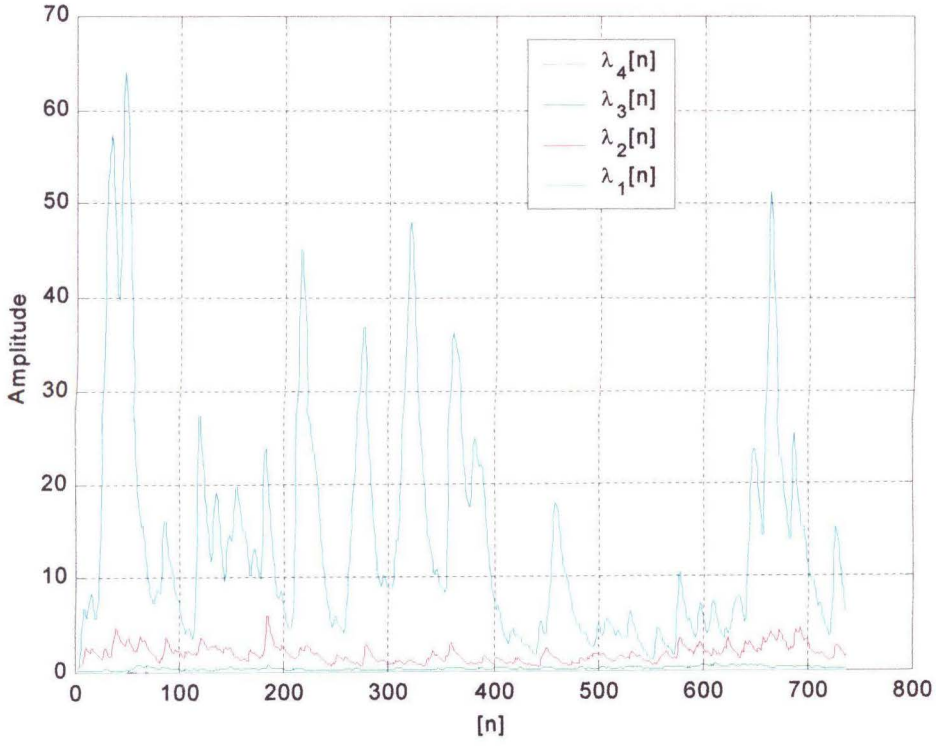


Figure 3.5: Time Evolution of Eigenvalues in Node (1,0) for ocean noise.

3.2 Signal and Noise

Our simulations were performed under different scenarios of signal shape and length, noise, and wavelet waveform. The signals used are square pulse trains, chirp signals, wavelet and scaling function bases, windowed sinusoids and exponentially decaying sinusoids. The simulations that we are going to show in this dissertation are of an exponentially decaying sinusoid as a signal. This signal is very commonly used for signaling purposes in communication and radar systems. It has the advantage of having parameters that control its energy concentration in time as well as in the frequency domain. In general the signal vector we used can be expressed as

$$s(n) = \frac{\sqrt{E_s}}{\sum_{n=0}^{N-1}\left(\cos(a\pi nT_s)e^{-bnT_s}\right)^2}\cos(a\pi nT_s)e^{-bnT_s}, \text{ for } T_s=\frac{T}{N-1}, \text{ and } n = 0,1,\ldots N-1 \quad (3.3)$$

where N is the signal vector length, E_s is the desired signal energy, T is the signal time interval, and a, b are positive valued parameters. The parameter a is the frequency shifting parameter or modulating factor. In other words the bigger a is the higher is the center frequency (always below the Nyquist frequency) of the signal spectrum. Parameter b is the spreading factor. The smaller b is, the more localized (narrow bandwidth) is the signal spectrum and the larger is its time duration. Figure 3.6 and 3.7 illustrate this point. In these figures, the signal energy $E_s = 0.2$ and $N = 32$. The parameter a changed from 4 to 8 and parameter b changed from 3 to 7. The discrete Fourier transform performed has four times the points of the signal (4N). In our simulations the signal length N was taken to be 64 points.

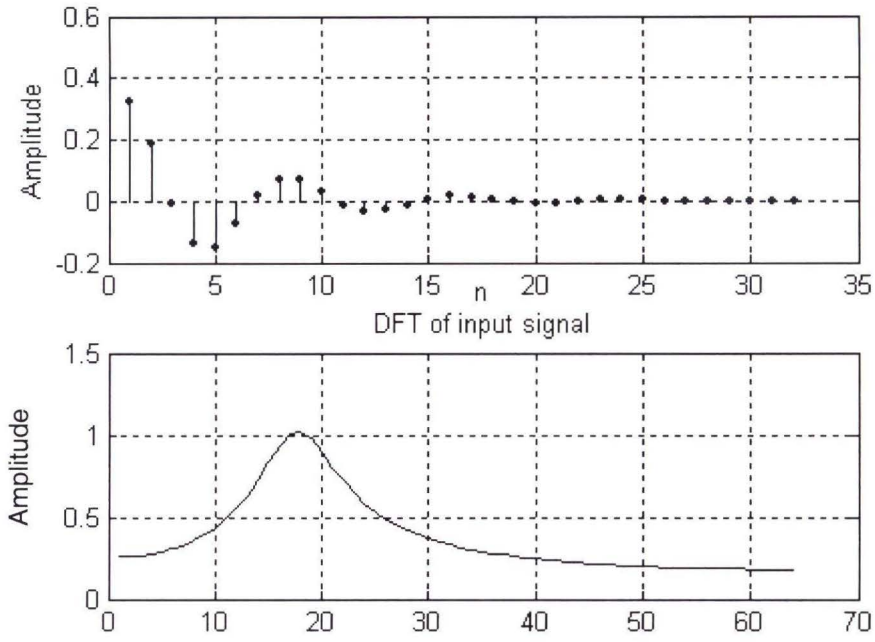


Figure 3.6: Illustration of Signal for modulation and spreading factor a=4, and b=3.

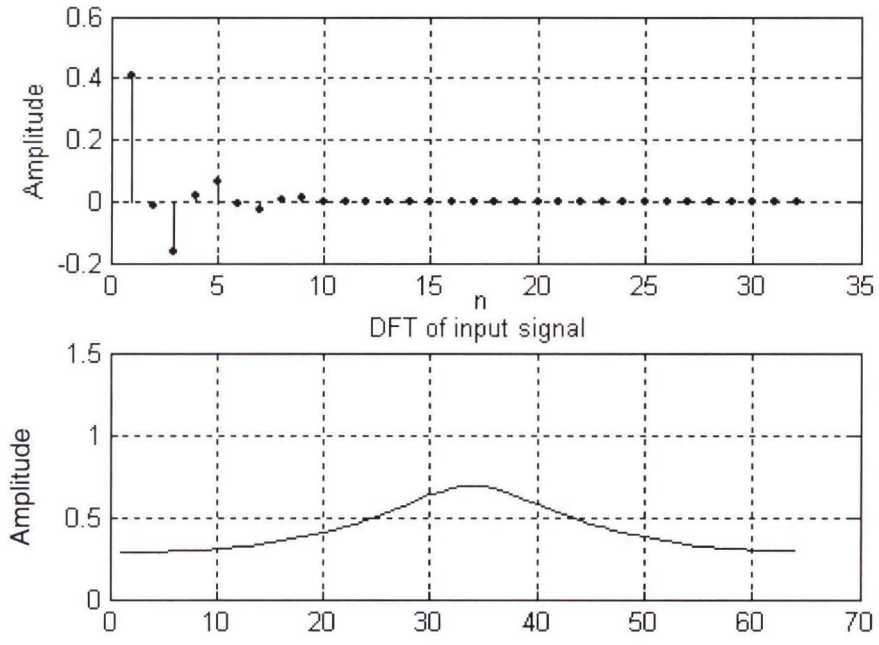


Figure 3.7: Illustration of Signal for modulation and spreading factor a=8, and b=7.

The noise used for our simulations varied from simulated noise like pink noise, for which the simulation technique is based on an autoregressive model, to noise acquired from experimental data. The first experimental noise is ambient ocean noise acquired with the help of a hydrophone, by the ocean engineering department in FAU. It approximates 1/f noise (pink noise) characteristics with additive shot noise. For more information on the characteristics and modeling of this noise refer to [ANSH01]. The second experimental noise was noise recorded in a moving vehicle through a cellular phone. The third and last experimental noise we used is called babble noise and it was recorded in a room of many people talking at the same time. The energy of these noises used in the simulations was normalized to 1 for comparison purposes.

The wavelet basis we used for decomposition were chosen from a variety of Daubechie wavelets, as well as some Coiflets, and biorthogonal wavelets, but our research focused mainly on Daubechie 3 (db3) wavelet. Since different wavelet basis exhibit different characteristics, exploration in more detail of the best wavelet choice for a given signal and noise is left for future research.

3.3 Simulations and Results

Our simulations produced two categories of results. The first was estimation of the detectability and measurement of the SNR at each analysis and synthesis branch, and the second was ROC curves to verify the results. For the detectability and SNR, a special code was written for displaying the results in a tree-like structure, which will enable the viewer to get the feel of how these quantities are evolved in the wavelet tree. The ROC curves were done using Monte Carlo Simulation (employing important sampling),

producing the probability of false alarm (P_{fa}) and detection (P_d), by sending repeatedly hypothesis H_0 and H_1 (as they are described in Chapter 2.4) respectively and checking if decision for signal presence is made. The results we are going to present involve lower than 1 SNR conditions for our experimental noise scenario. The simulated pink noise behaves very similarly to the ocean noise and thus we show only ocean noise results. The detectability figures show results down to the fourth depth level, but the ROC curves are shown for only the first two levels for visual clarity.

3.3.1 Vehicle Noise

The signal used for this noise as shown in Figure 3.8 was the one in equation (3.3) with parameters a=16, b=20, T=4, and E_s= 0.2 (i.e. SNR = 0.2). Its corresponding ROC curves before and after whitening are shown in Figure 3.9. Figures 3.10, 3.11, 3.12, show the detectability, normalized SNR (see Appendix I) before, and after whitening respectively, as they are evolved along the analysis scales. Notice that the SNR after whitening is very close to the detectability, and that is an indication of how good our whitener is. Figures 3.13, 3.14 show the ROC curves before and after whitening for the first two levels of decomposition in comparison with the ROC of the original signal. In Figures 3.10, 3.11, 3.12, and 3.15 the red branch line is the best tree path (maximum detectability). The corresponding nodes for each level are (0,0), (1,1), (2,2), (3,4), and (4,8), with maximum detectability found at node (3,4). This value is approximately 12.7 times better than the detectability at the input of our detector. Notice that the detectability does not necessarily increase as we go down the tree, but there is a point in the tree that reaches maximum value. For a more detail results on the eigenvalues (Λ_{ij}) and signal

projections onto the noise covariance matrix eigenvectors (equation (2.54)), for the first level of decomposition, refer to Appendix II.

The increase in the detectability is verified by the ROC curves for the first two decomposition levels (i.e. the more the detectability the better the probability of detection for a given probability of false alarm). For example for a probability of false alarm P_{fa} = .05, the probability of detection using an optimum detector at the input is approximately 0.528 whereas it becomes 0.993 at node (2,2) which is a significant improvement.

Similar results for the detectability and the associated ROC curves, in the synthesis branches, are shown in Figures 3.15, 3.16, and 3.17 respectively. We can observe in this case that the results are even more impressive. The maximum detectability occurs at node (3,4), which is approximately 76 times better than the detectability at the tree input.

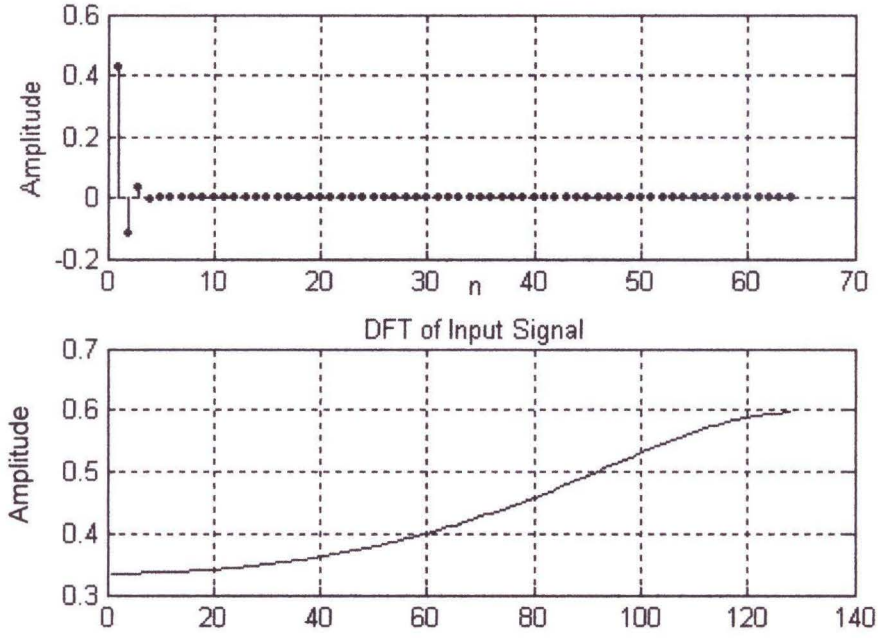


Figure 3.8: Input Signal in Time and Frequency Domain.

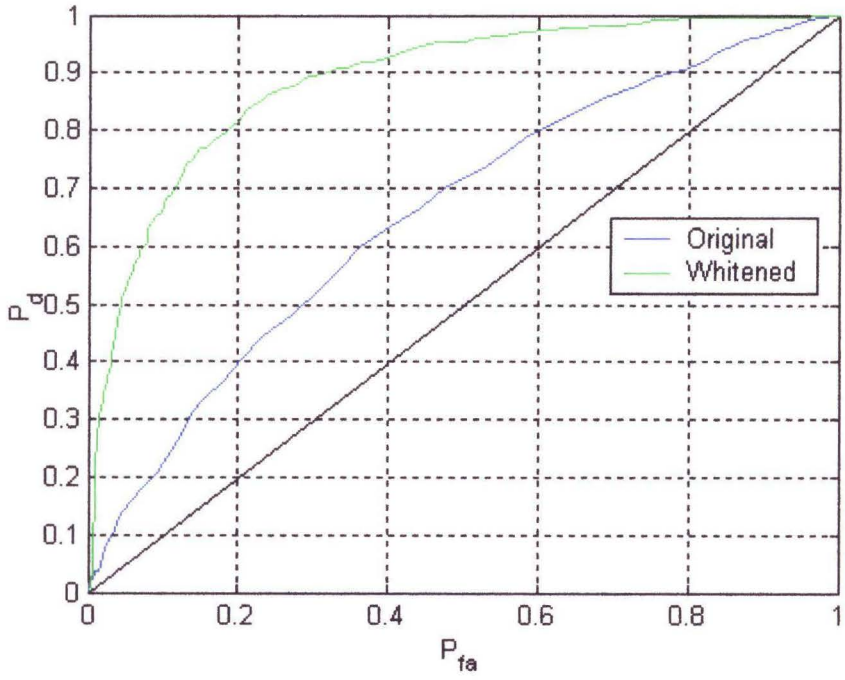


Figure 3.9: ROC of Input Signal Before and After Whitening in Vehicle Noise.

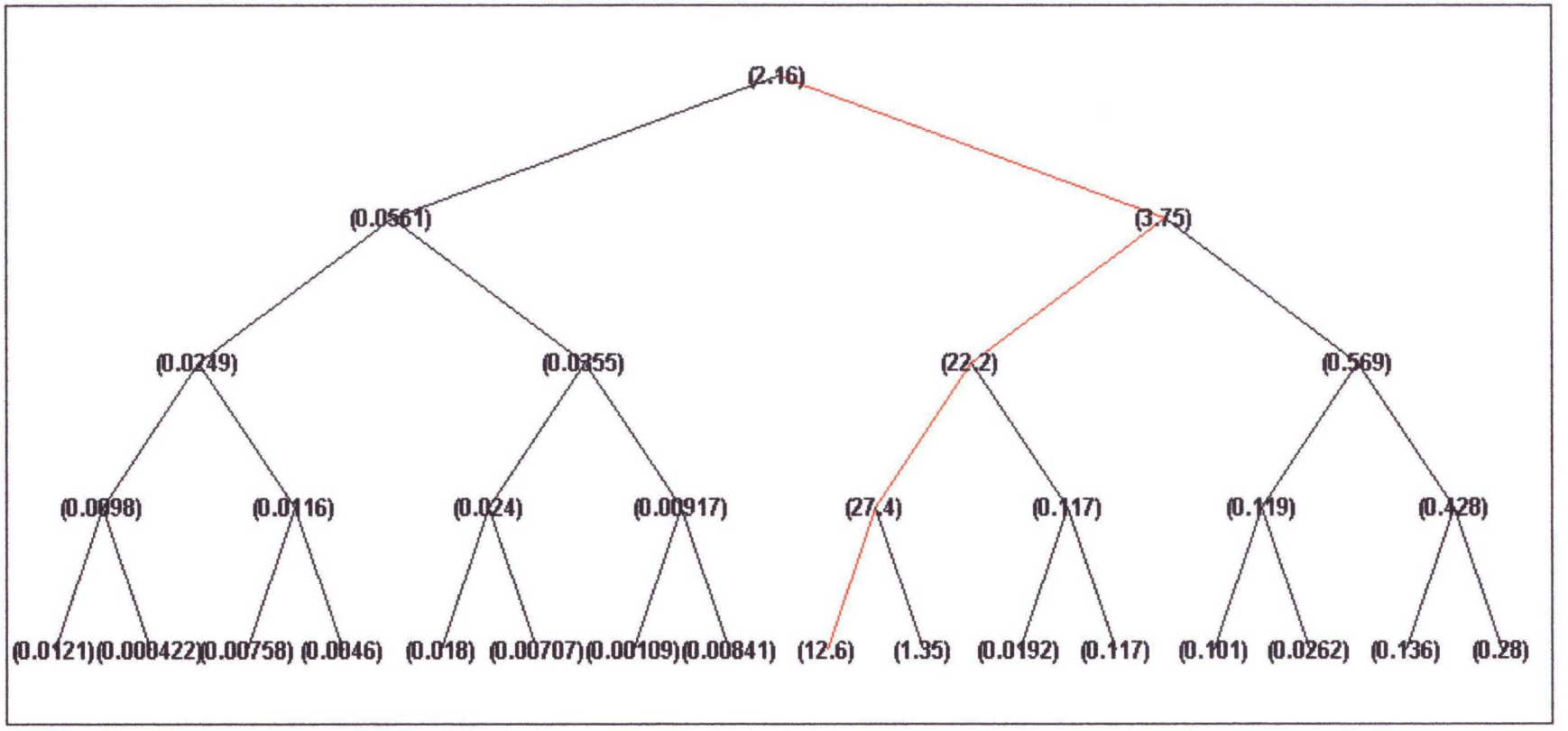


Figure 3.10: Detectability Along Scales for Vehicle Noise, Analysis.

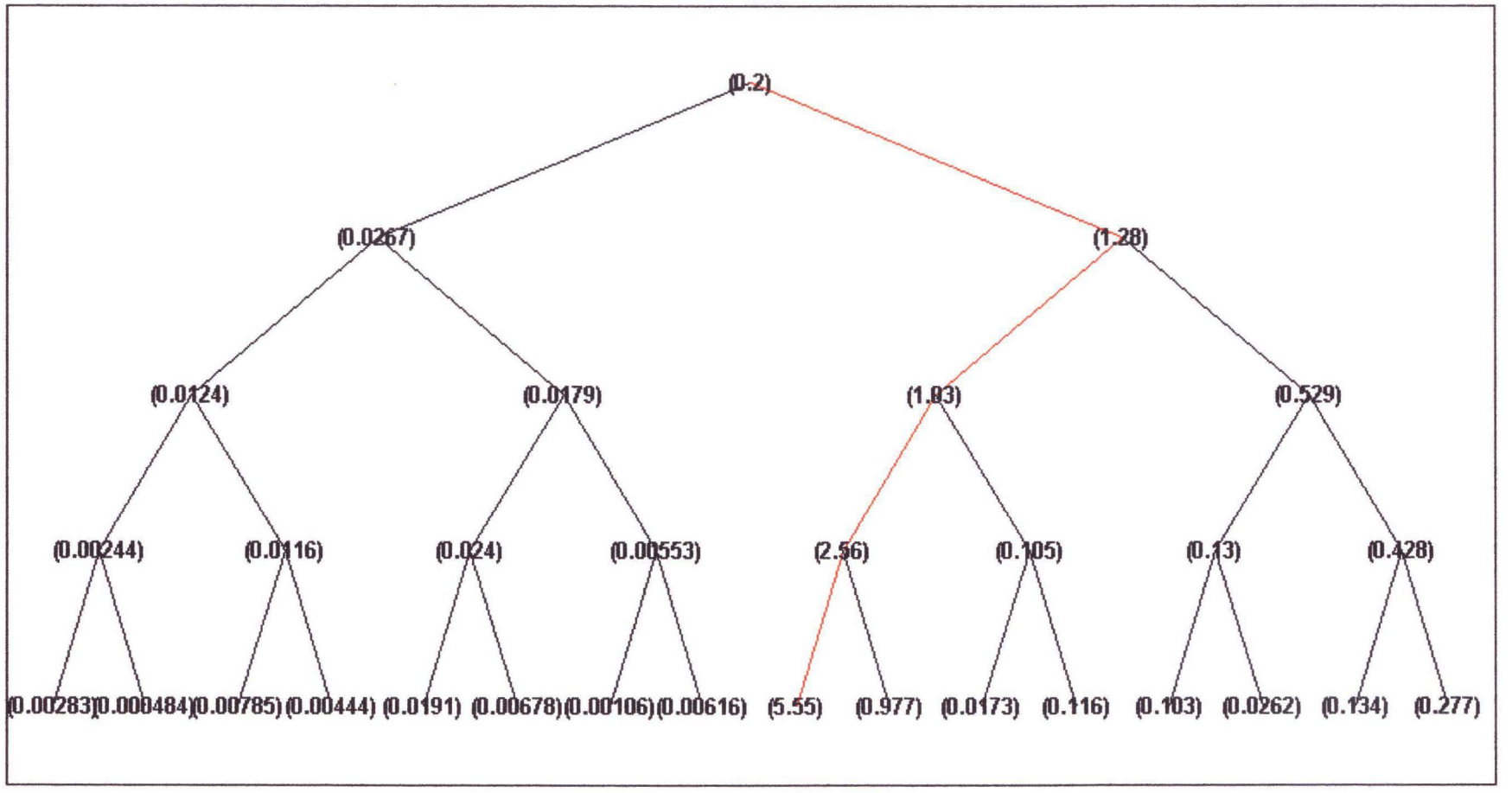


Figure 3.11: SNR Before Whitening Along Scales for Vehicle Noise, Analysis.

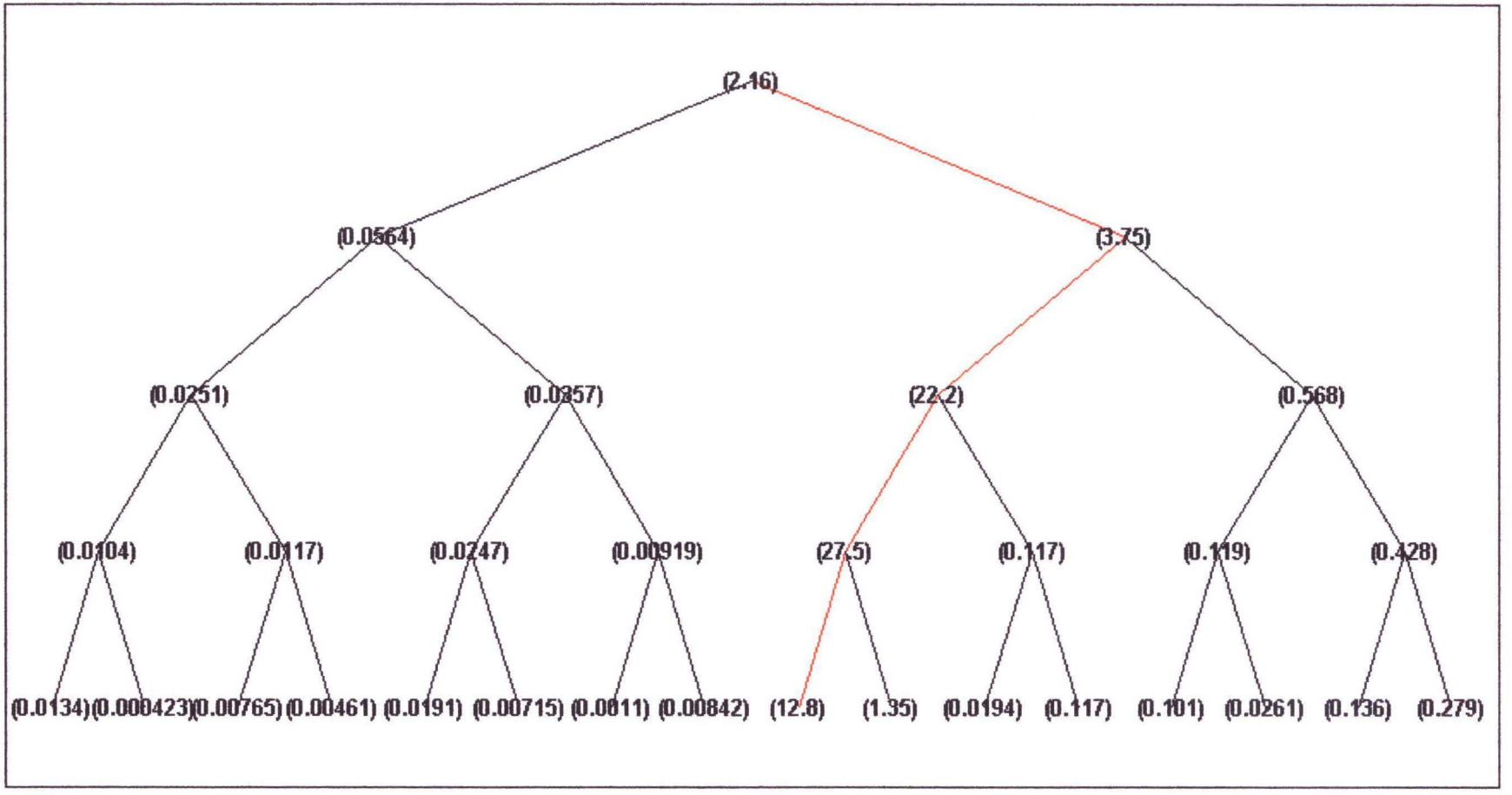


Figure 3.12: SNR After Whitening Along Scales for Vehicle Noise, Analysis.

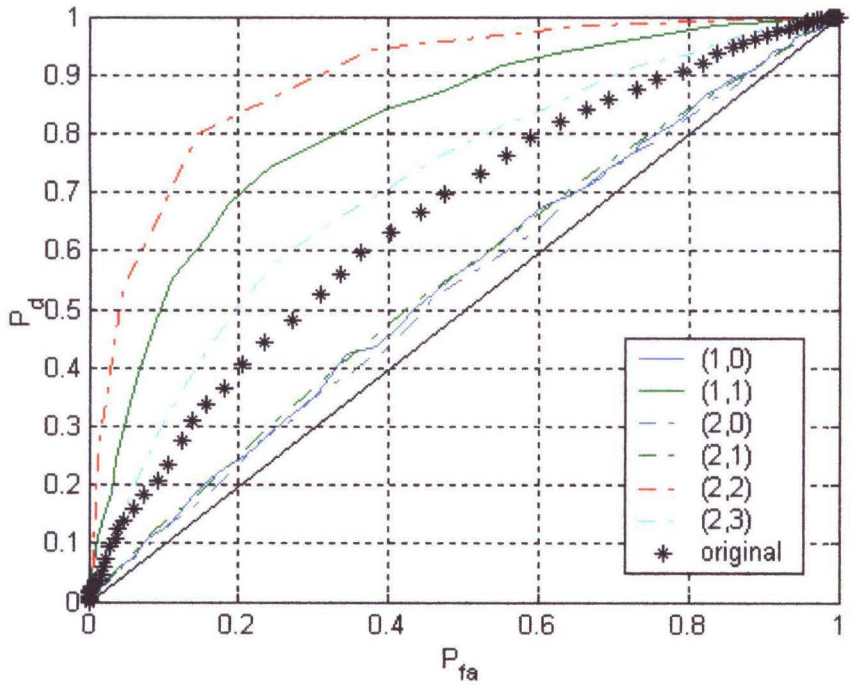


Figure 3.13: ROC Along Scales Before Whitening for Vehicle Noise, Analysis.

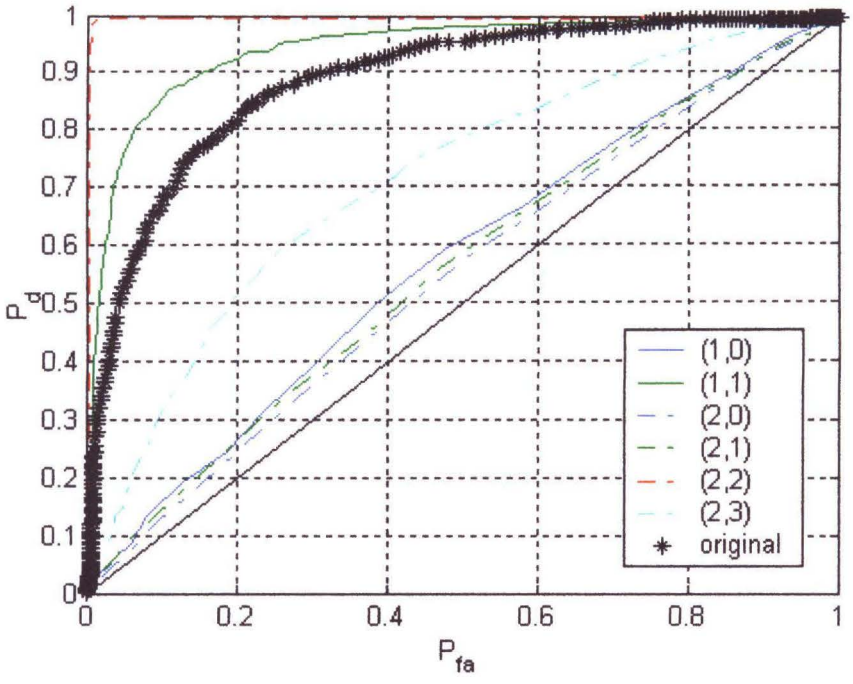


Figure 3.14: ROC Along Scales After Whitening for Vehicle Noise, Analysis.

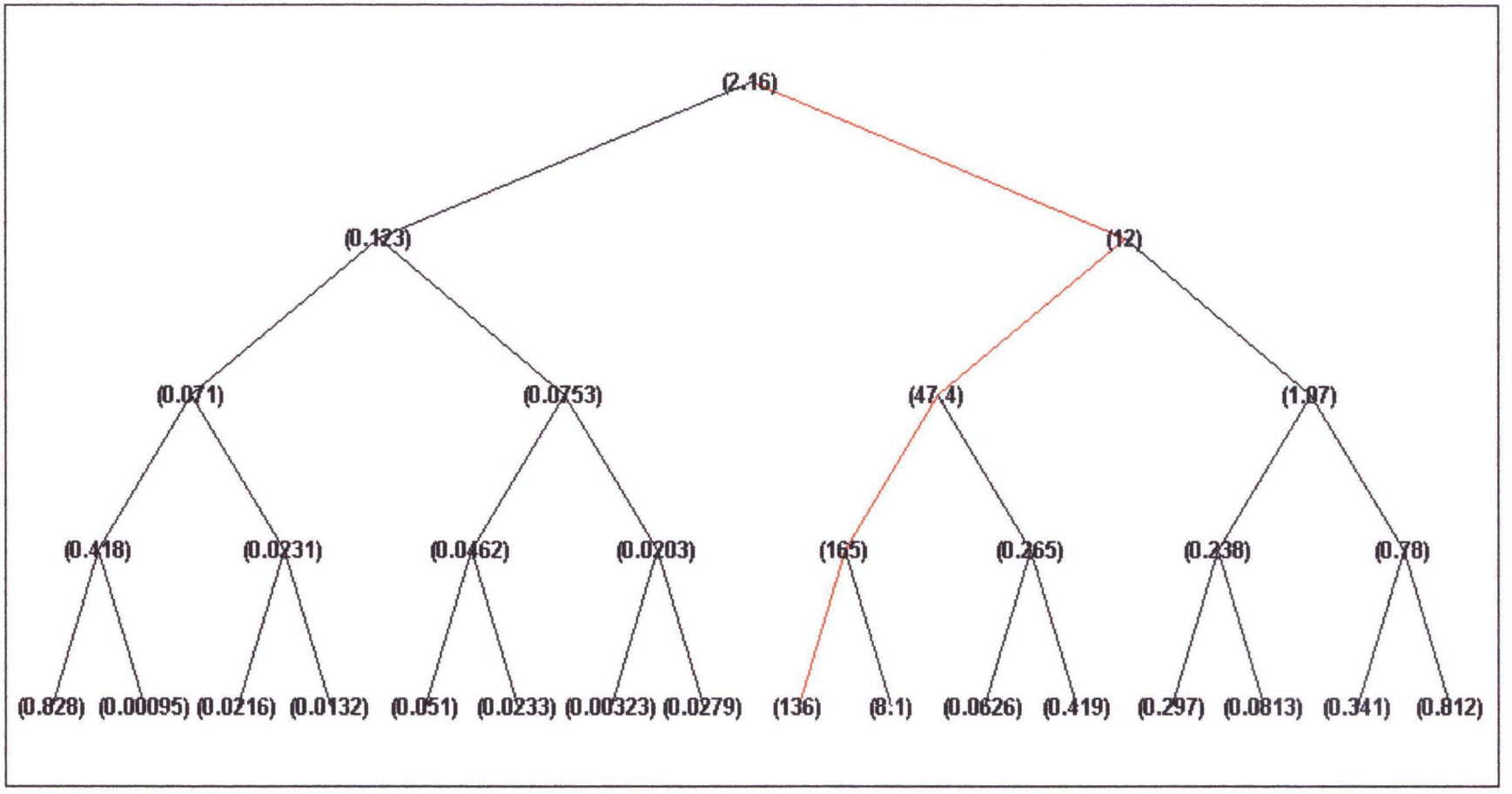


Figure 3.15: Detectability Along Scales for Vehicle Noise, synthesis.

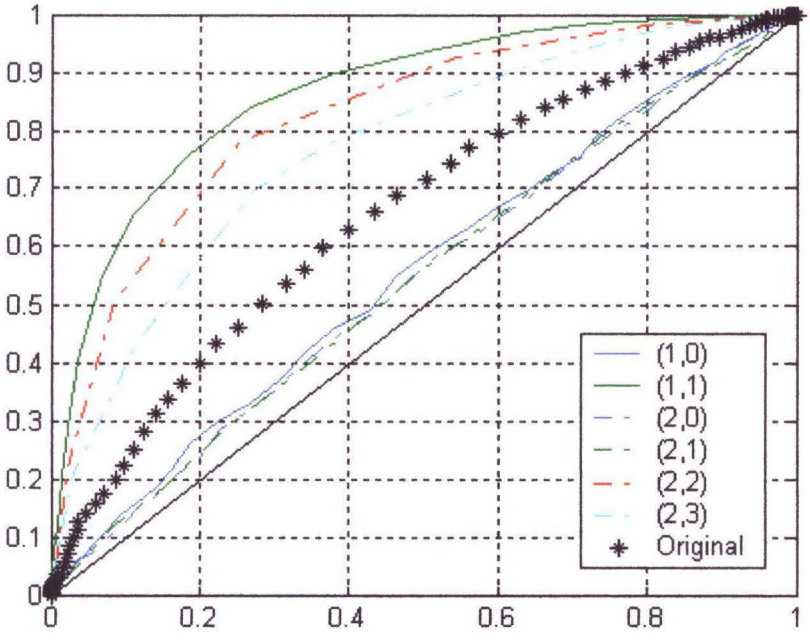


Figure 3.16: ROC Along Scales Before Whitening for Vehicle Noise, Synthesis.

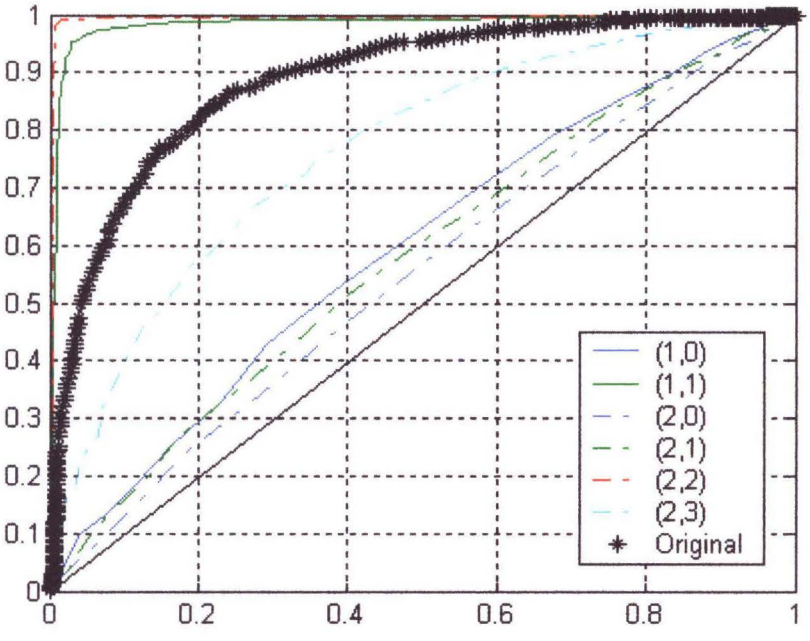


Figure 3.17: ROC Along Scales Before Whitening for Vehicle Noise, Synthesis.

3.3.2 Babble Noise

The signal used for this noise as shown in Figure 3.18, was the one in equation (3.3) with parameters a=14, b=10, T=4, and E_s= 0.2 (i.e SNR = 0.2). Its corresponding ROC curves before and after whitening are shown in Figure 3.19. Figures 3.20, 3.21, 3.22, show the detectability, SNR before, and SNR after whitening respectively, as they are evolved along the analysis scales. Figures 3.23, 3.24 show the ROC curves before and after whitening for the first two levels of decomposition in comparison with the ROC of the original signal. We can observe that the results are similar to the vehicles noise case (regarding the path of maximum detectability). The corresponding nodes for each level that detectability is maximum are (0,0), (1,1), (2,2), (3,4), and (4,8), with maximum detectability found at node (2,2) unlike the case with the vehicle noise. This detectability is approximately 11.4 times better than the detectability at the input of our detector. The improvement is also transparent in the ROC curves where for a probability of false alarm P_{fa} = .05, the probability of detection using an optimum detector at the input is approximately 0.517 whereas it becomes 0.9991 at node (2,2).

Analogous results for the detectability and the associated ROC curves, in the synthesis branches, are shown in Figures 3.25, 3.26, and 3.27 respectively. The maximum detectability occurs at node (4,8), which is approximately 60 times better than the detectability at the tree input.

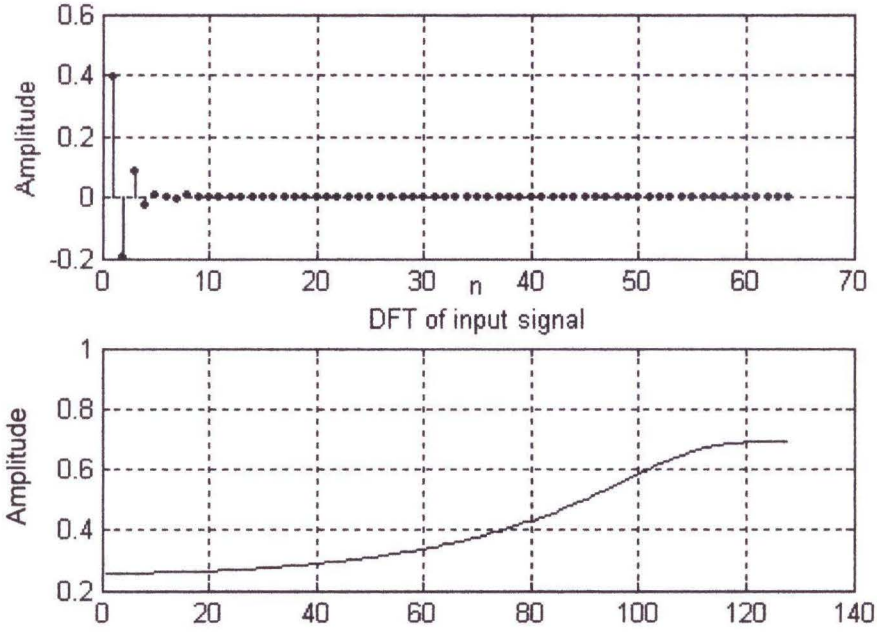


Figure 3.18: Input Signal in Time and Frequency Domain.

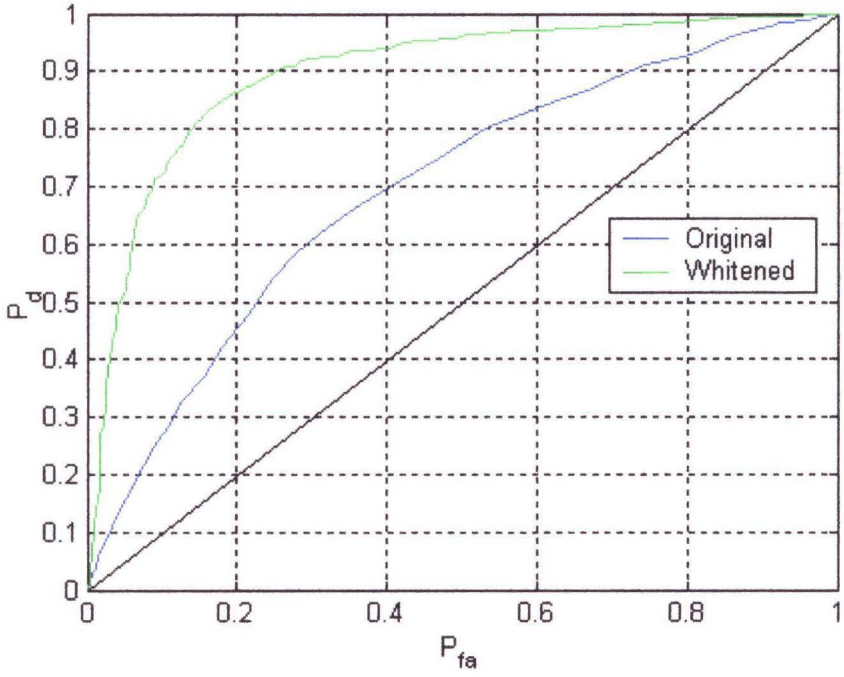


Figure 3.19: ROC of Input Signal Before and After Whitening for Babble Noise.

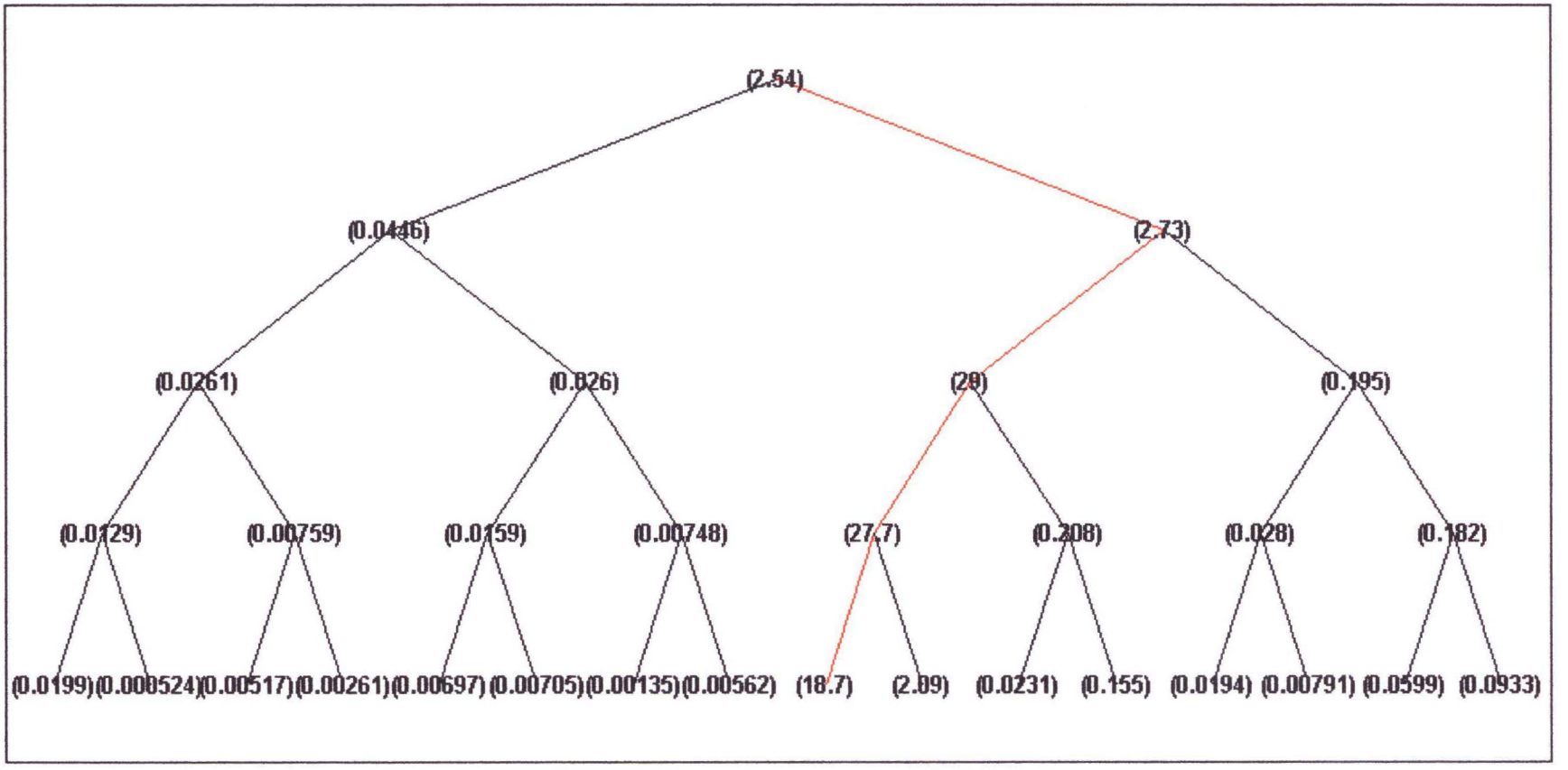


Figure 3.20: Detectability Along Scales for Babble Noise, Analysis.

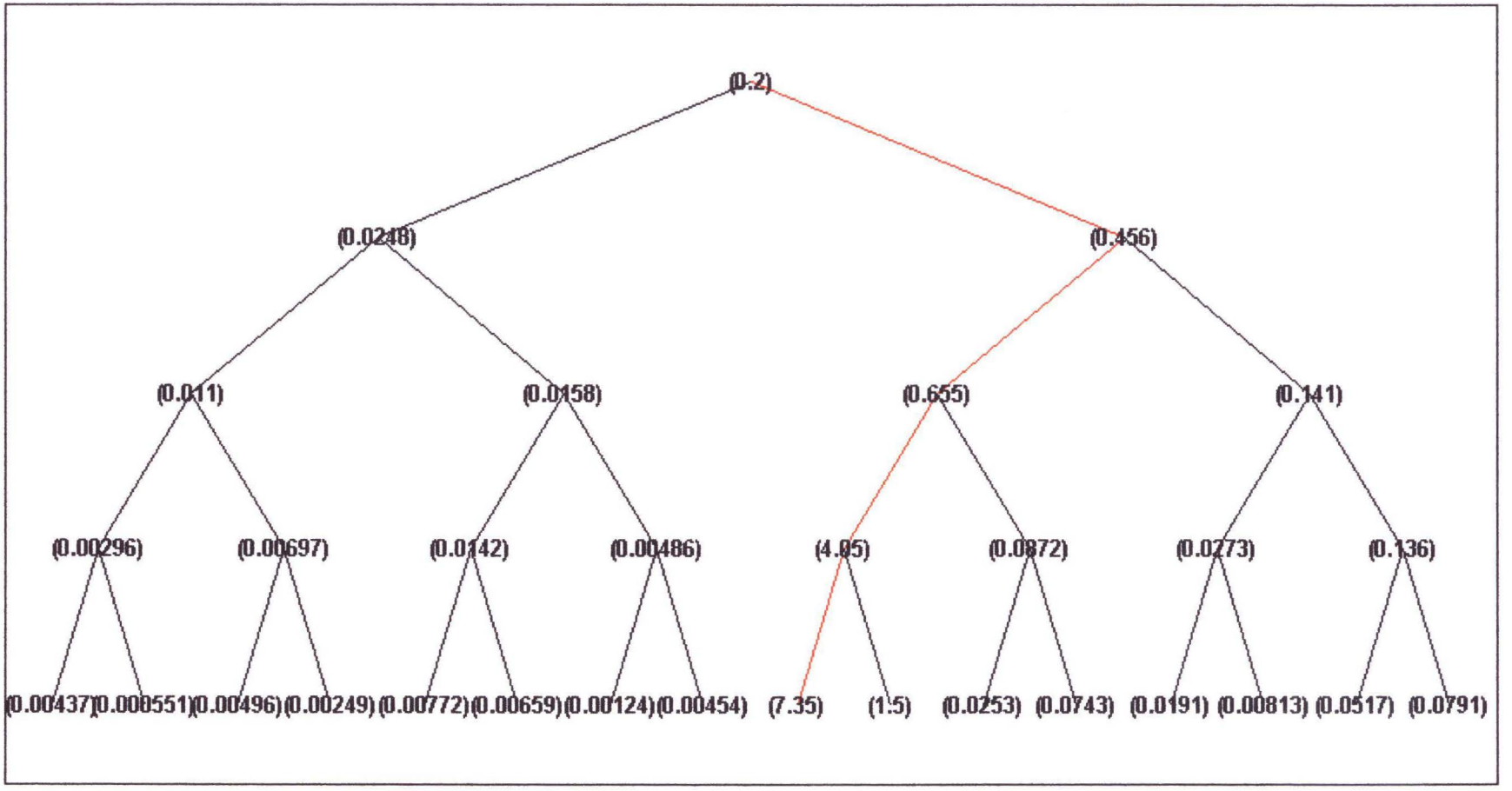


Figure 3.21: SNR Before Whitening Along Scales for Babble Noise, Analysis.

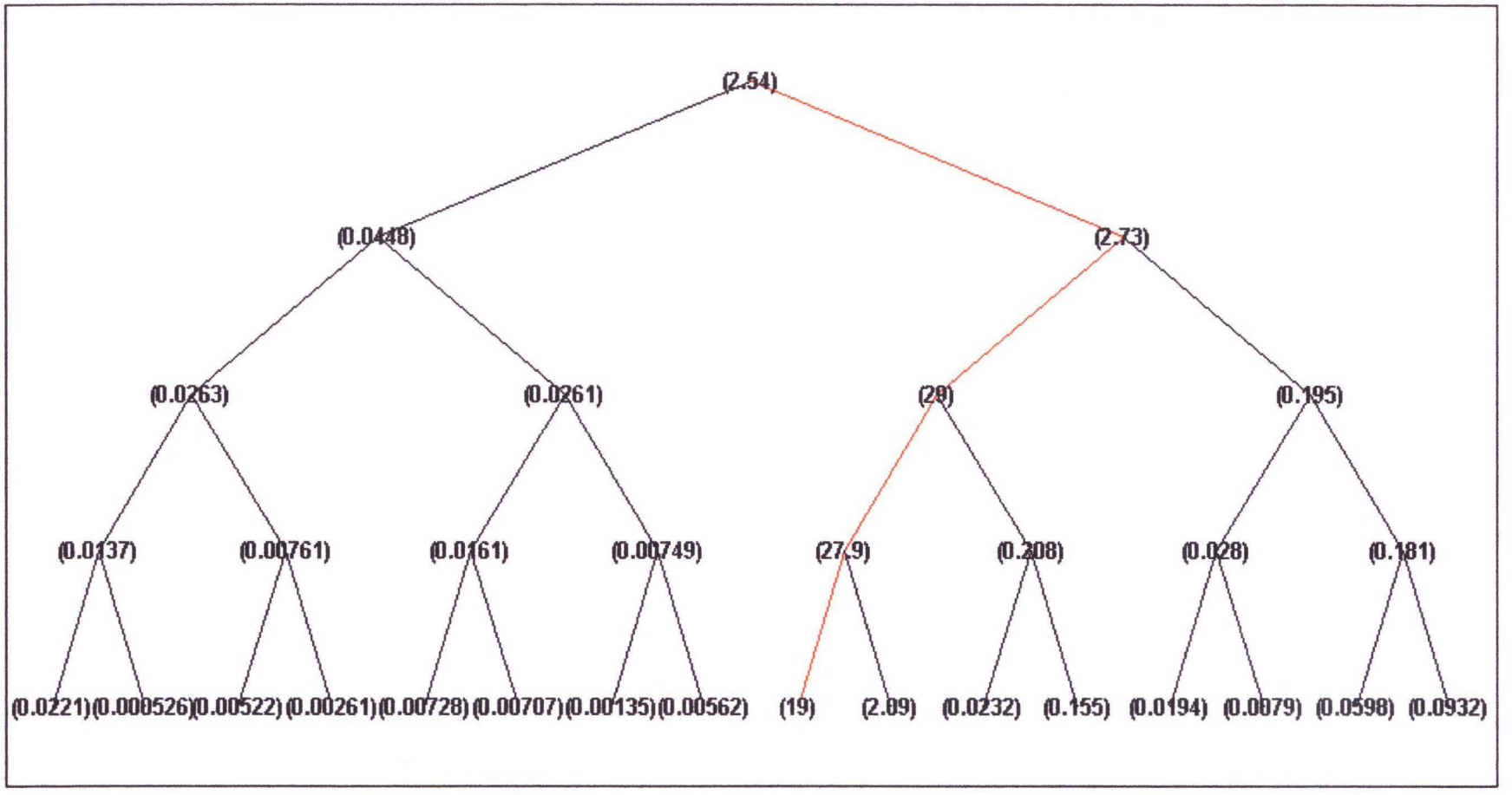


Figure 3.22: SNR After Whitening Along Scales for Babble Noise, Analysis.

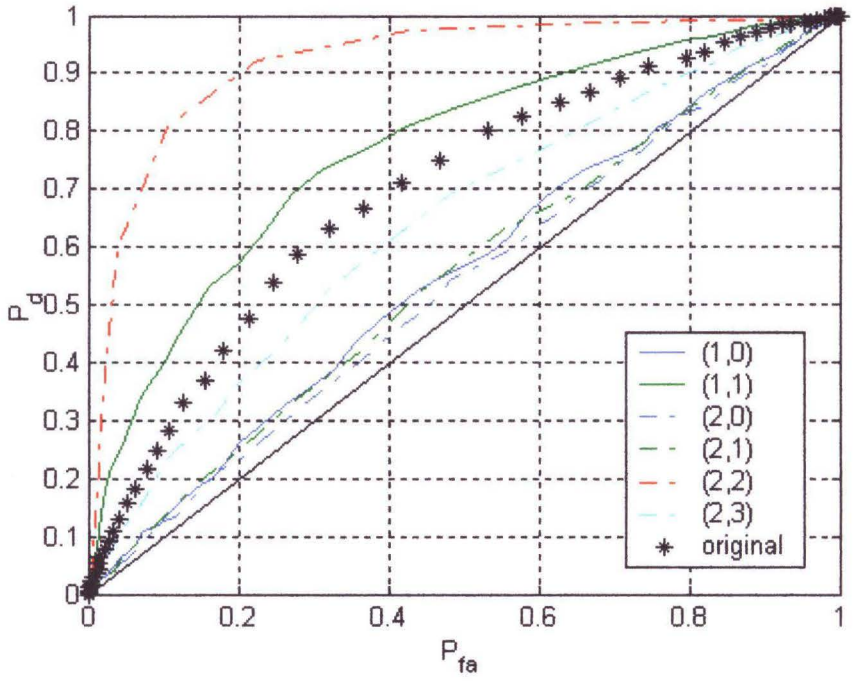


Figure 3.23: ROC Along Scales Before Whitening for Babble Noise, Analysis.

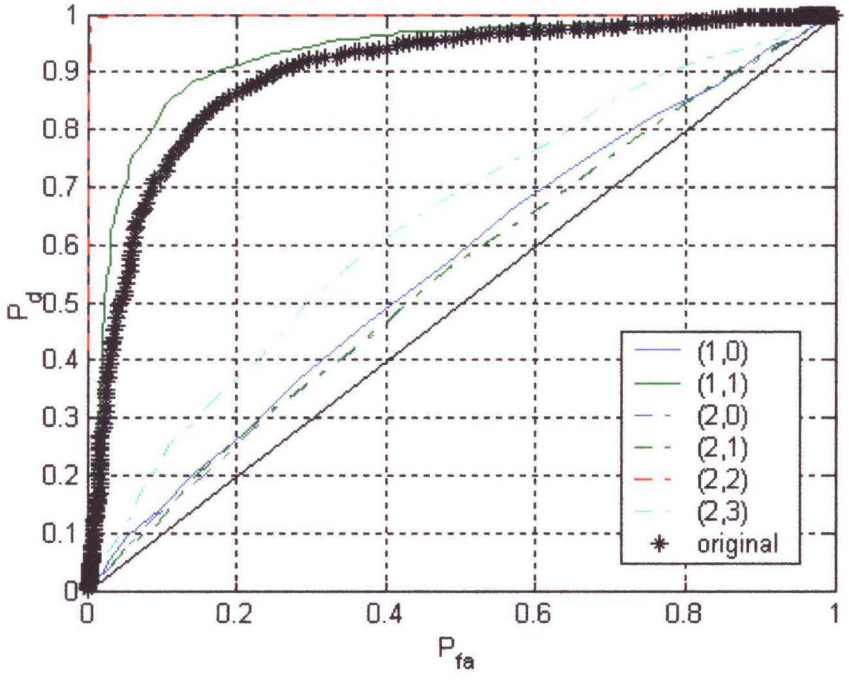


Figure 3.24: ROC Along Scales After Whitening for Babble Noise, Analysis.

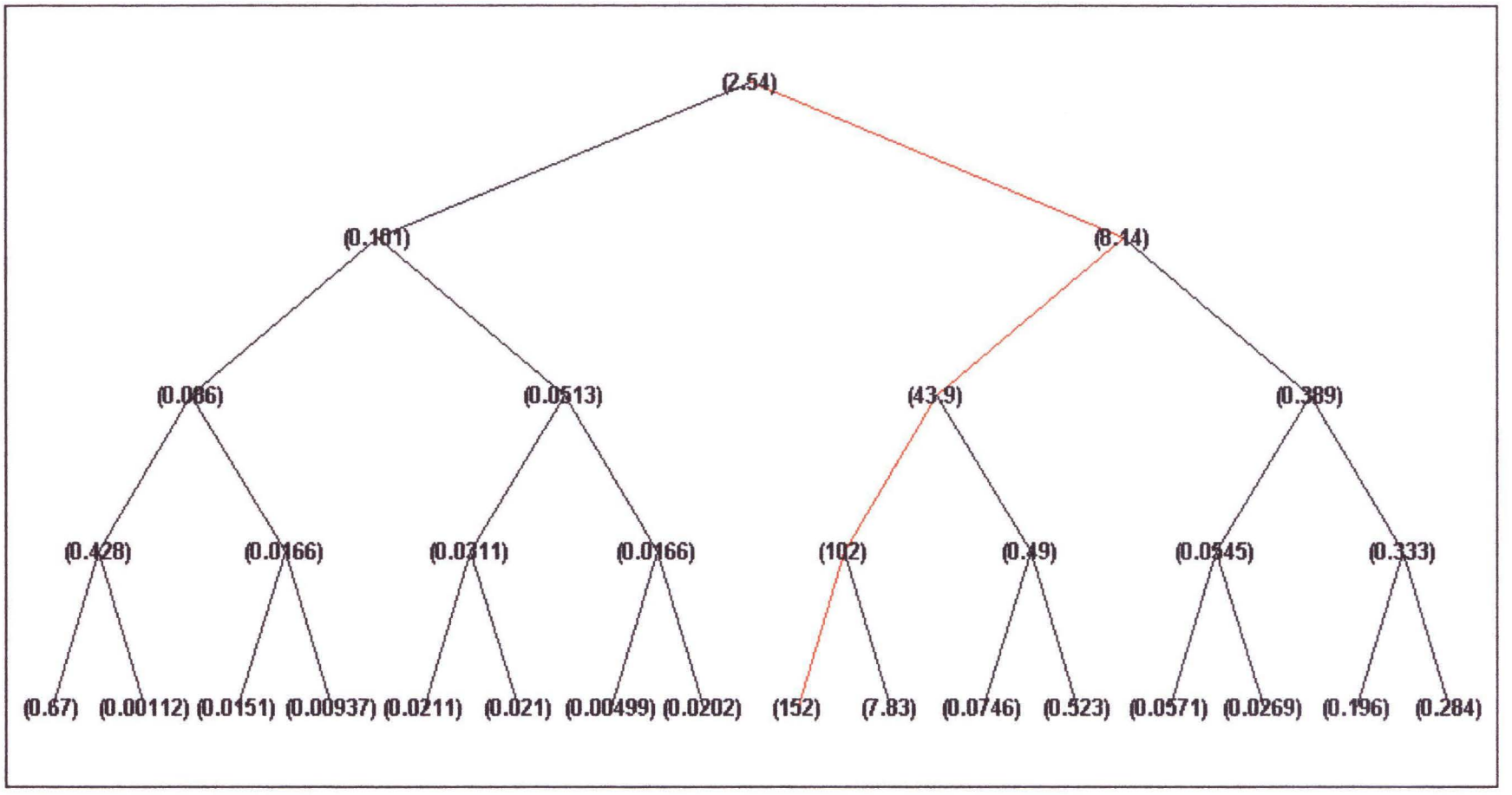


Figure 3.25: Detectability Along Scales for Babble Noise, Synthesis.

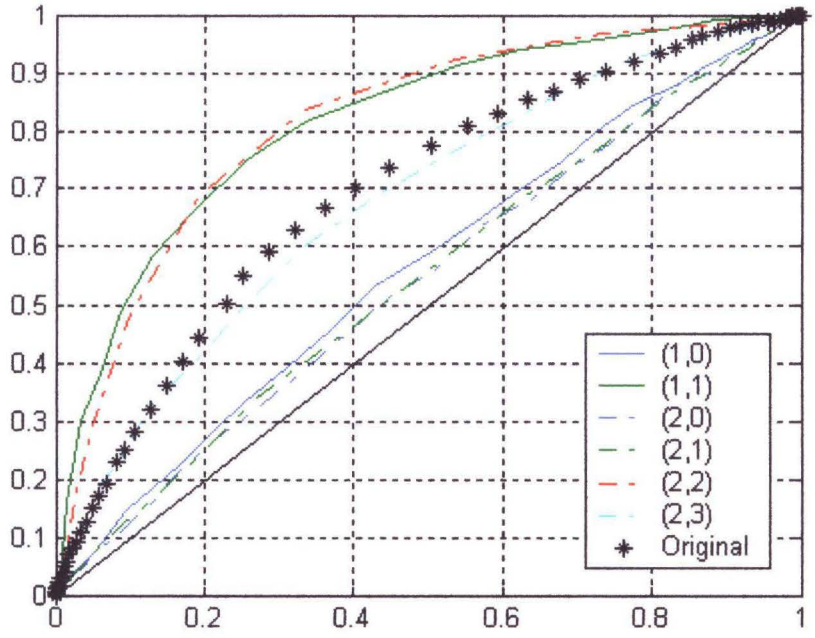


Figure 3.26: ROC Along Scales Before Whitening for Babble Noise, Synthesis.

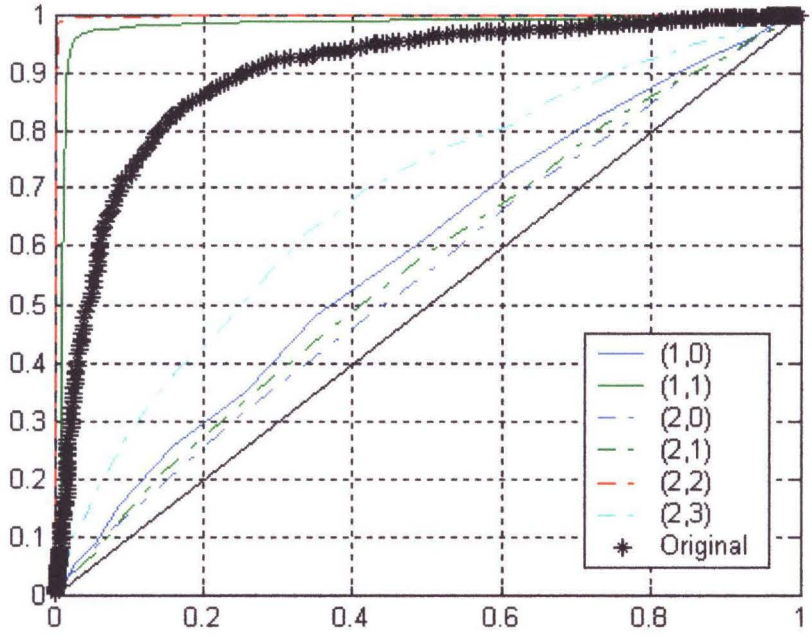


Figure 3.27: ROC Along Scales After Whitening for Babble Noise, Synthesis.

3.3.3 Ocean Noise

The signal used for this noise as shown in Figure 3.28, was the one in equation (3.3) with parameters a=1, b=18, T=4, and E_s= 0.05 (i.e SNR = 0.05). Its corresponding ROC curves before and after whitening are shown in Figure 3.29. Figures 3.30, 3.31, 3.32, show the detectability, SNR before, and SNR after whitening respectively, as they are evolved along the analysis scales. Figure 3.33, 3.34 shows the ROC curves before and after whitening for the first two levels of decomposition in comparison with the ROC of the original signal. The corresponding nodes for each level that detectability is maximum are (0,0), (1,0), (2,0), (3,0), and (4,0), with maximum detectability found at node (1,0) unlike the case with the vehicle noise. This detectability is approximately 3.1 times better than the detectability at the input of our detector. The improvement is also transparent in the ROC curves where for a probability of false alarm P_{fa} = .05, the probability of detection using an optimum detector at the input is approximately 0.367 whereas it becomes 0.733 at node (1,0).

Similar results for the detectability and the associated ROC curves, in the synthesis branches, are shown in Figures 3.35, 3.36, and 3.37 respectively. The maximum detectability occurs at node (4,0), which is approximately 3.5 times better than the detectability at the tree input.

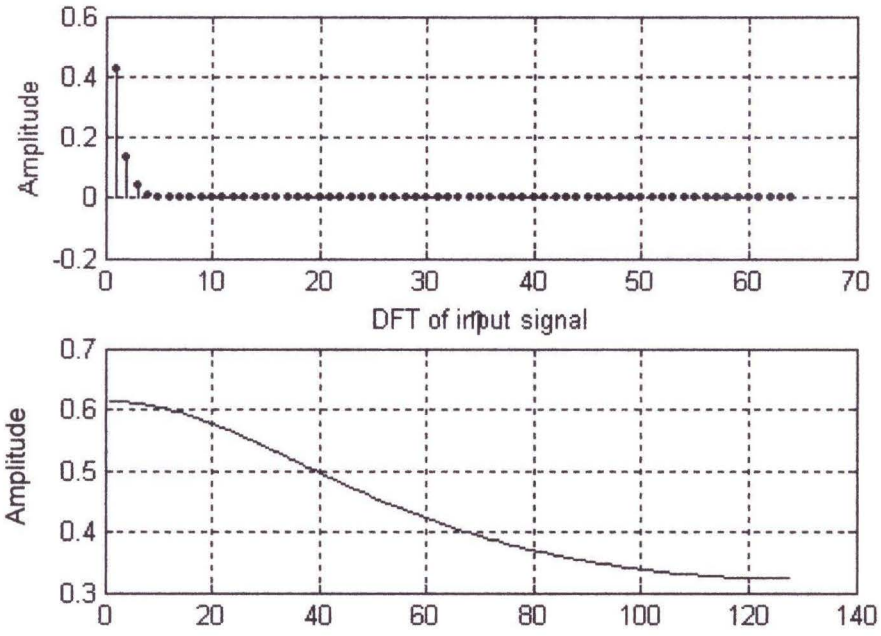


Figure 3.28: Input Signal in Time and Frequency Domain.

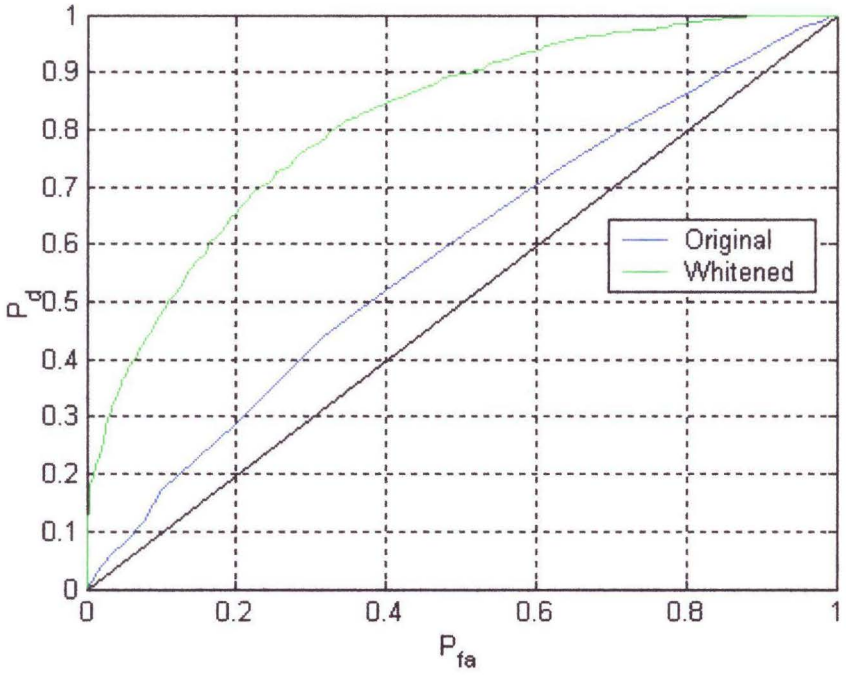


Figure 3.29: ROC of Input Signal Before and After Whitening for Ocean Noise.

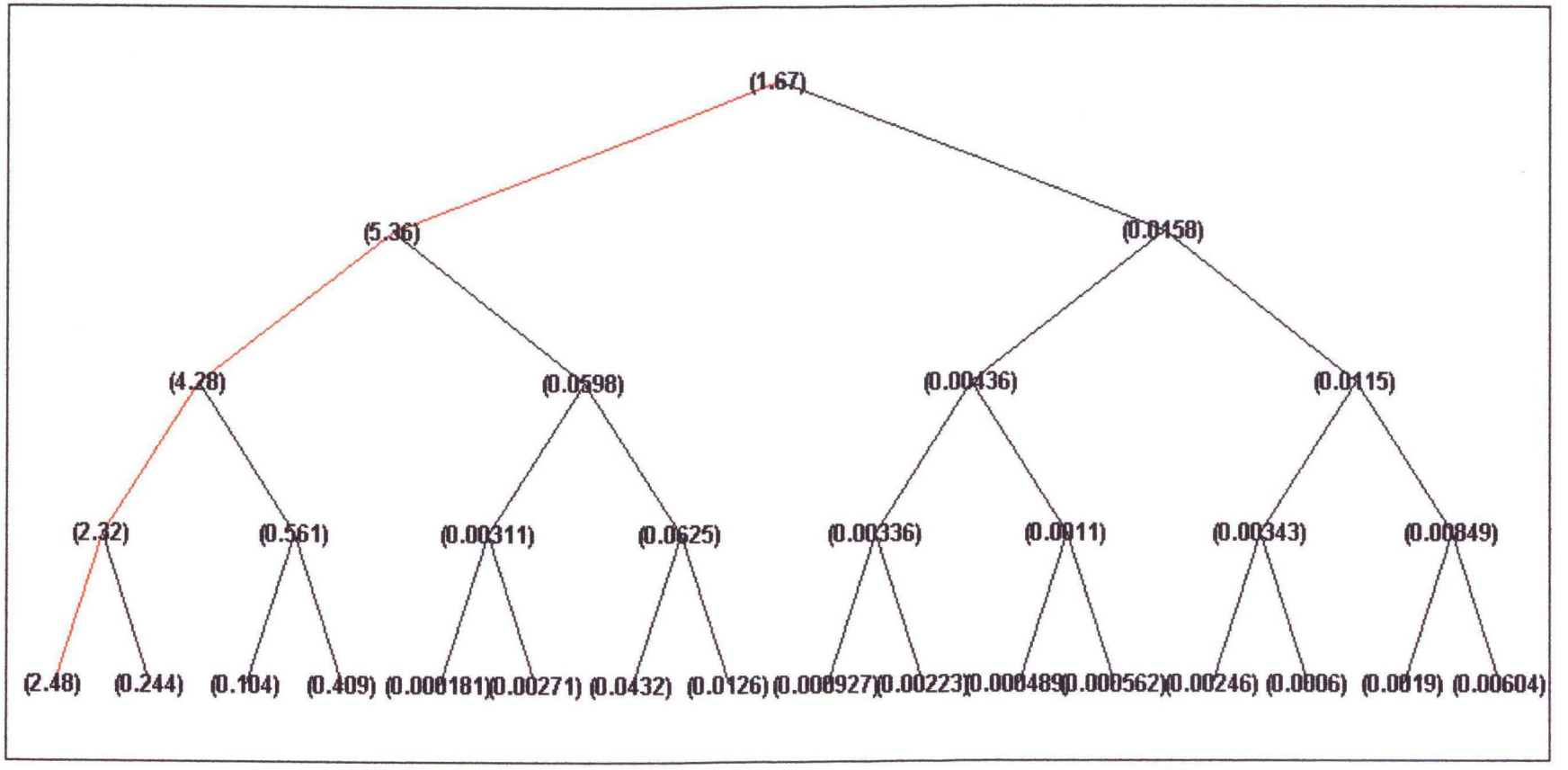


Figure 3.30: Detectability Along Scales for Ocean Noise, Analysis.

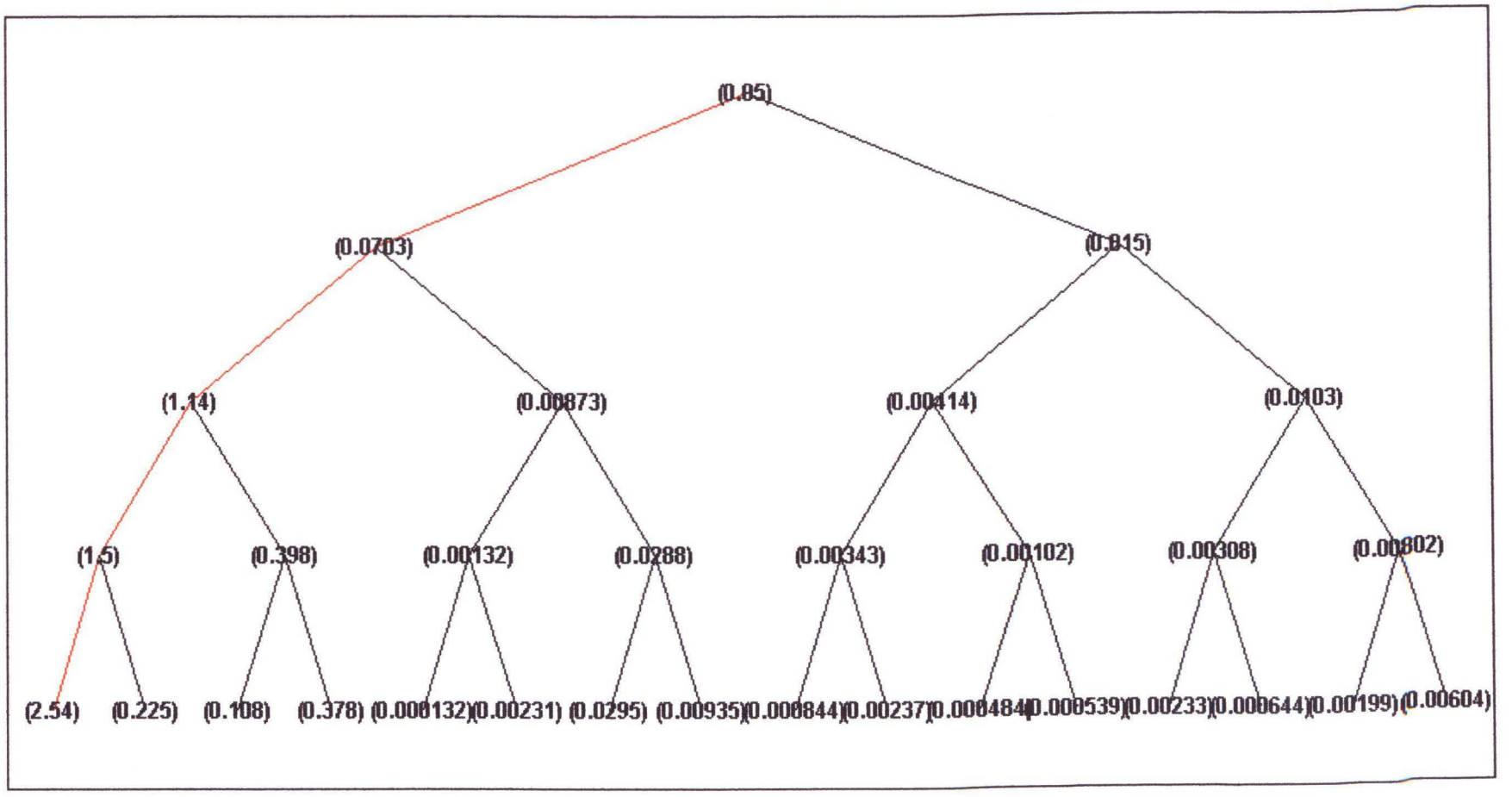


Figure 3.31: SNR Before Whitening Along Scales for Ocean Noise, Analysis.

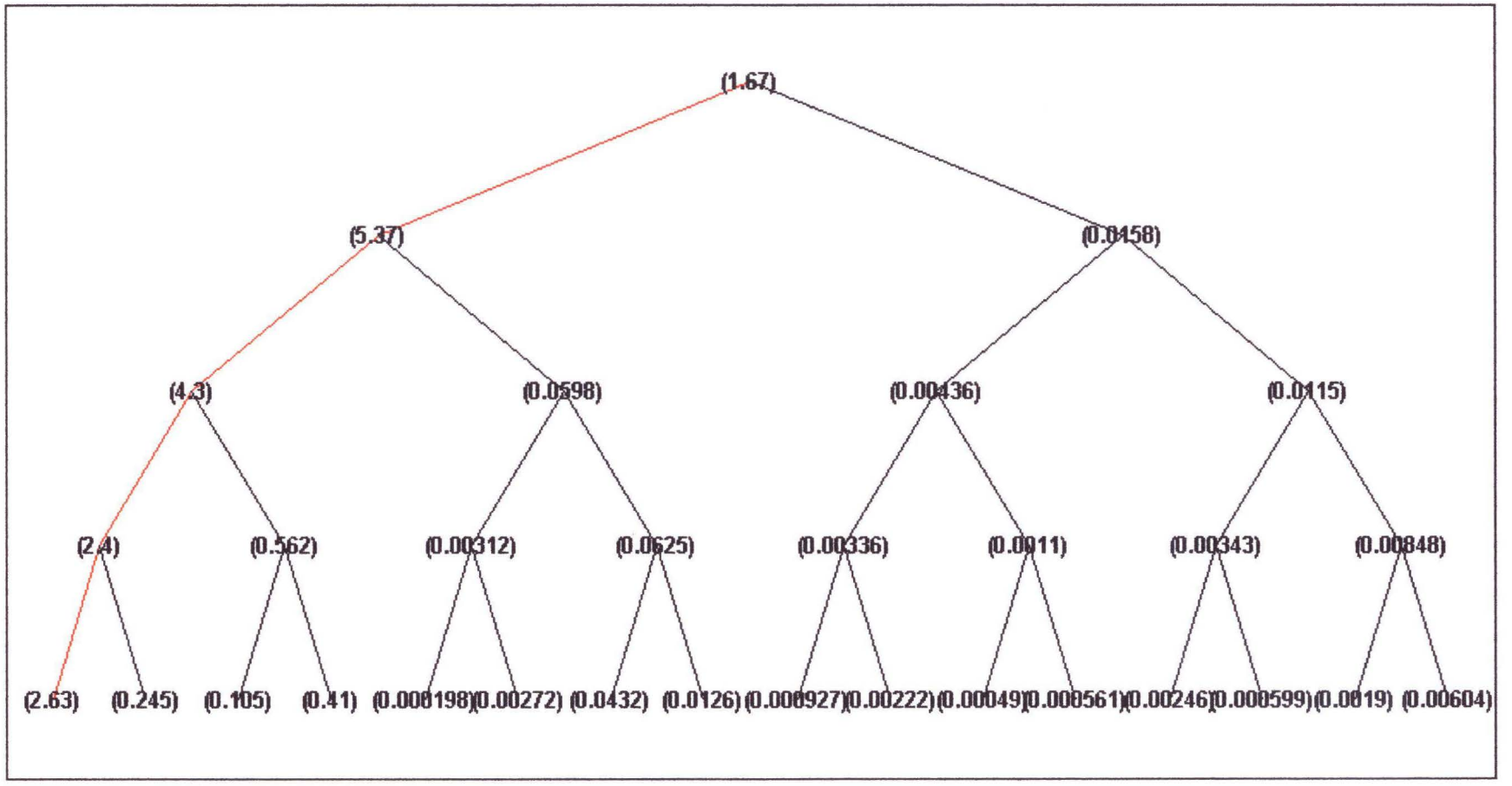


Figure 3.32: SNR After Whitening Along Scales for Ocean Noise, Analysis.

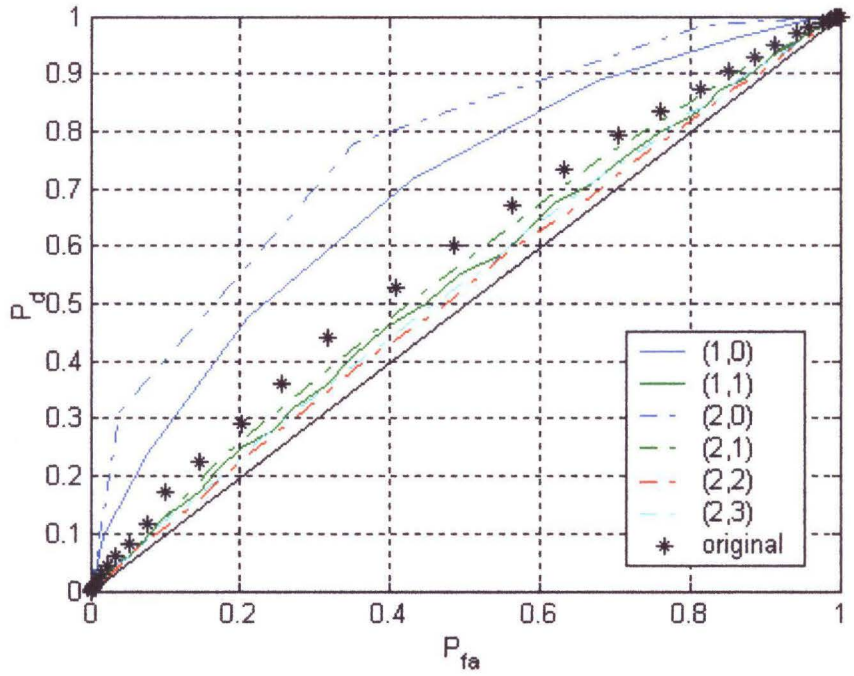


Figure 3.33: ROC Along Scales Before Whitening for Ocean Noise, Analysis.

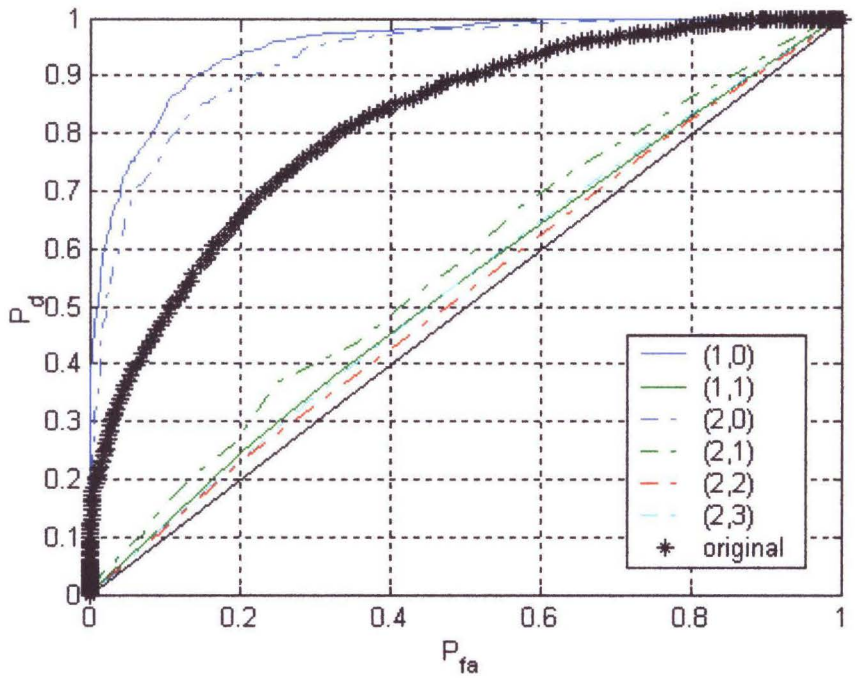


Figure 3.34: ROC Along Scales After Whitening for Ocean Noise, Analysis.

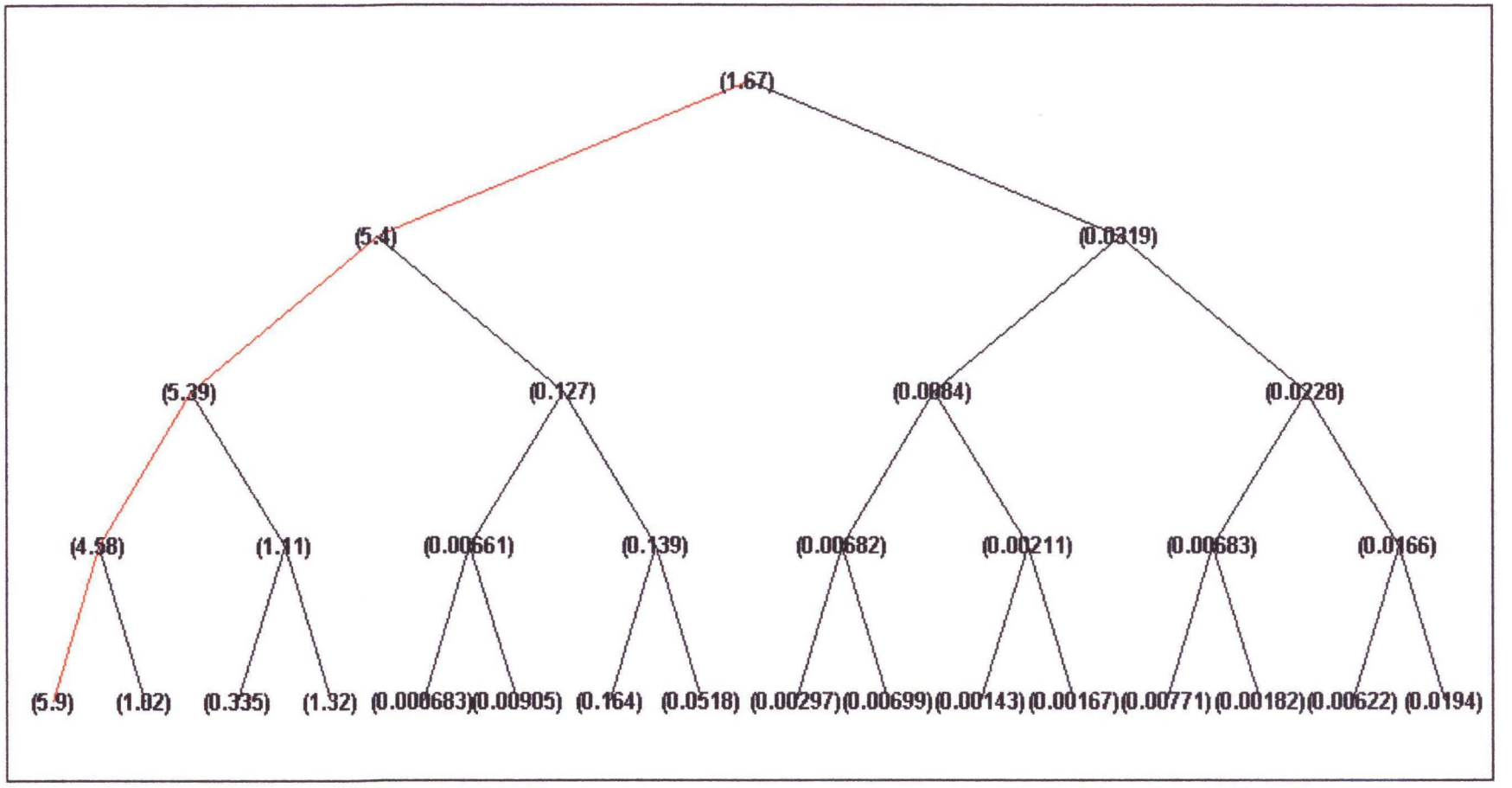


Figure 3.35: Detectability Along Scales for Ocean Noise, Synthesis.

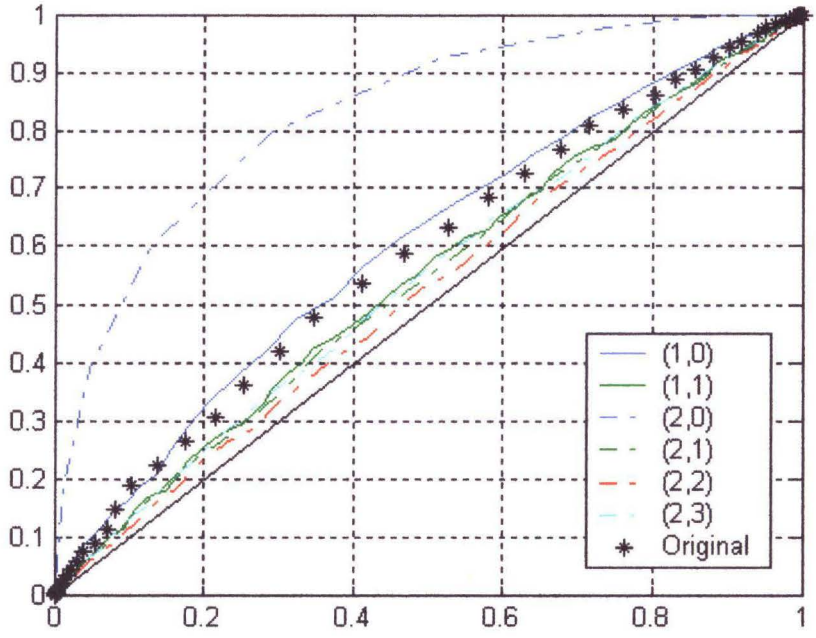


Figure 3.36: ROC Along Scales Before Whitening for Ocean Noise, Synthesis.

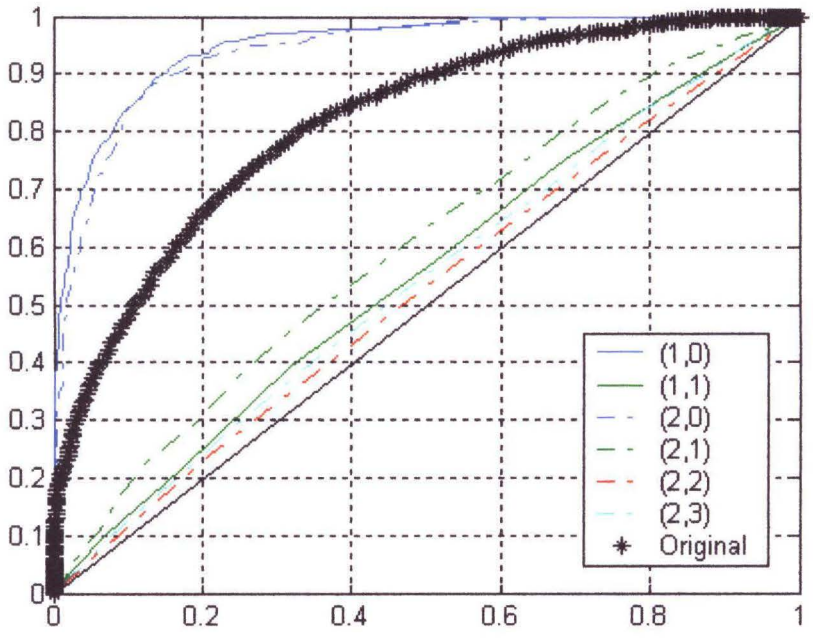


Figure 3.37: ROC Along Scales After Whitening for Ocean Noise, Synthesis.

The importance of the estimation of the noise statistics (KL transform) is essential for achieving optimum whitening and thus optimum detection. In the next Chapter we will show how we can compute them efficiently taking advantage of the wavelet structure.

CHAPTER 4

EIGENVALUE DECOMPOSITION METHODS ANALYSIS-SIMULATIONS AND RESULTS

4.1 Plane Rotations for Finding EVD

Optimum detection requires optimum whitening. This is achieved by an estimation of the statistics of the noise. More specifically the eigenvalues and eigenvectors of the covariance function of the noise have to be accurately calculated for optimum whitening as shown in Chapter 3. For the discrete and finite signal case the problem to be solved is really the diagonalization of a finite dimension symmetric matrix with no particular structure (such as Toeplitz) to be taken advantage of. This process is very computationally expensive and sometimes unstable, especially as the size of the covariance matrix increases.

Let $\mathbf{R}$ be an NxN square matrix. The problem to be solve is to find a number λ and an N dimensional vector $\mathbf{q}$ such that $\mathbf{Rq} = \lambda\mathbf{q}$, where λ is called an eigenvalue and vector $\mathbf{q}$ is its corresponding eigenvector. This equation can be rewritten as $(\mathbf{R}-\lambda\mathbf{I})\mathbf{q}=0$, where $\mathbf{I}$ is an n dimensional unity matrix. This is equivalent to asking for those values of λ that allow the resultant system of N homogeneous equations in the N unknown coordinates of $\mathbf{q}$ to be solved. Such solution exists if and only if the determinant $|\mathbf{R}-\lambda\mathbf{I}| = 0$. This results to an N^{th} degree polynomial to be solved and thus there are N eigenvalues to be found. Then for each

eigenvalue we can find the appropriate eigenvector to satisfy $(\mathbf{R}-\lambda_i\mathbf{I})\mathbf{q}_i=0$ for i=1,2 ... N. This problem can also be stated as $\mathbf{RQ}=\Lambda\mathbf{Q}$, where Λ is a diagonal matrix containing λ_i, and $\mathbf{Q}=[\mathbf{q}_1\ \mathbf{q}_2 \ldots \mathbf{q}_N]$ is an orthogonal matrix with the corresponding eigenvectors. In real life this method of solving the eigenvalue problem is not practical. This is because the search for the eigenvalues (roots of the polynomial) is not a trivial task especially in large matrices that have no particular structure (assuming one found the polynomial coefficients first).

In practical situations recursive (adaptive) methods are used for finding the eigenvalues and eigenvectors of a matrix. Since we will be concentrating on the matrices **R** that are symmetric and real (although for Hermitian matrices the methods are very similar), methods based on plane rotations are best suited for this scenario [SIHA96] since they have been widely studied and are very commonly employed in such cases. These methods are very popular because they are easily coded with relatively low computational cost and one can control the error of the final result [PAWH58]. They employ a sequence of 2x2 plane rotations known as Jacobi or Givens rotations for the diagonalization of a data matrix, in a step-by-step fashion within some prescribed numerical precision. Moreover, these methods using plane rotations result to both eigenvectors and eigenvalues with no need for separate calculation. Also a lot of theoretical work has been done to verify convergence and estimate convergence speed of such methods [PHEN58], [PHGF60], [ERHA63], and [PJEB62], and even methods for parallel processing based on plane rotations have been devised [FLHP89]. The reader can get more information about alternative methods as well as methods for the non-symmetric cases from [PAWH58].

We are going to start describing these methods by explaining how plane rotations can be used to manipulate elements from a matrix and how this concept can be used for matrix diagonalization.

4.1.1 Plane Rotations Applied on a Matrix [PAWH58]

Any plane rotation can be described by the 2x2 orthogonal matrix given by equation (4.1) as

$$\Theta = \begin{bmatrix} c & s \\ -s & c \end{bmatrix} \tag{4.1}$$

where c and s are real parameters defined by

$$c = \cos(\theta) \tag{4.2a}$$

$$s = \sin(\theta) \tag{4.2b}$$

with the trigonometric constrain

$$c^2+s^2 = 1. \tag{4.3}$$

We refer to the transformation Θ as a plane rotation because multiplication of a 2x1 data vector by Θ amounts to a plane rotation of that vector. This is true whether the data vector is pre or post multiplied by Θ. This rotation is also referred to as Jacobi rotation in honor of Jacobi who proposed the method in 1846 for a symmetric matrix diagonalization. It is also refer to as Givens rotation. To illustrate the nature of a plane rotation we will consider the case of a real 2x1 vector $\mathbf{a} = [x_a\ y_a]^T$. Then the premultiplication of vector **a** with Θ gives

$$\mathbf{b} = \Theta\mathbf{a} = \begin{bmatrix} c & s \\ -s & c \end{bmatrix}\begin{bmatrix} x_a \\ y_a \end{bmatrix} = \begin{bmatrix} cx_a + sy_a \\ -sx_a + cy_a \end{bmatrix} = \begin{bmatrix} x_b \\ y_b \end{bmatrix}. \tag{4.4}$$

Notice that using equation (4.3) the Euclidean length of vector **x** is the same as the one of **y** which is consistent with a vector rotation i.e.

$$\|\mathbf{y}\| = \sqrt{(cx_a + sy_a)^2 + (cy_a - sx_a)^2} = \sqrt{x_a^2(c^2 + s^2) + y_a^2(c^2 + s^2)} = \|\mathbf{x}\|. \tag{4.5}$$

Moreover, for an angle θ to be positive, then vector **a** is rotated in the clockwise direction as illustrated in Figure 4.1.

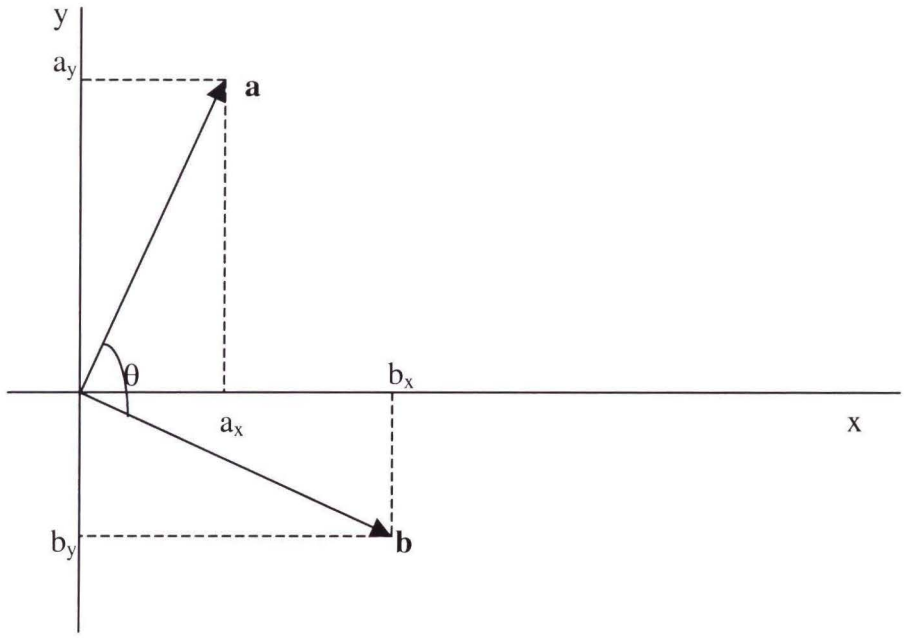


Figure 4.1: Plane Rotation of a Real 2x1 Vector.

4.1.2 Plane Rotations for Diagonalization

We are now going to show the procedure and conditions of a simple diagonalization of a real symmetric 2x2 matrix **X** using plane rotations by post and pre multiplying by the rotation matrix as:

$$\mathbf{\Theta}^T\mathbf{X}\mathbf{\Theta} = \begin{bmatrix} c & s \\ -s & c \end{bmatrix}^T \begin{bmatrix} x_{11} & x_{12} \\ x_{12} & x_{22} \end{bmatrix} \begin{bmatrix} c & s \\ -s & c \end{bmatrix} = \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix}. \tag{4.6}$$

The question we want to answer is what are the values of s, and c that equation (4.6) is true. Expanding equation (4.6) and equating with right hand side we get

$$d_1 = c^2x_{11} - 2scx_{12} + s^2x_{22} \tag{4.7a}$$

$$d_1 = s^2x_{11} + 2scx_{12} + c^2x_{22} \tag{4.7b}$$

$$0 = c^2x_{12} + sc(x_{11}+x_{22}) - s^2x_{12} = (c^2-s^2)x_{12} + sc(x_{11} - x_{22}). \tag{4.7c}$$

Essentially d_1, and d_2 are the eigenvalues of **X**, Θ^T is the set of orthogonal eigenvectors of **X** corresponding to the d eigenvalues, and equation (4.7c) gives the condition for this to happen. We can rewrite equation (4.7c) as

$$\left(\frac{c}{s} - \frac{s}{c}\right) x_{12} + (x_{11} - x_{22}) = 0 \tag{4.8}$$

and let $\frac{s}{c} = t$ we get

$$\left(\frac{1}{t} - t\right)x_{12} + (x_{11}+x_{22}) = t^2 x_{12} - t(x_{11} - x_{22}) - x_{12} = t^2 + t\frac{(x_{22} - x_{11})}{x_{12}} - 1 = 0. \qquad (4.9a)$$

Then letting $\frac{(x_{22} - x_{11})}{x_{12}} = 2b$, equation (4.9a) becomes

$$t^2+2bt-1=0 \qquad (4.9b)$$

with solution

$$t = \frac{-2b \pm \sqrt{4b^2+4}}{2} = -b \pm \sqrt{1+b^2} = \begin{cases} \dfrac{\text{sign}(b)}{|b| + \sqrt{1+b^2}} \text{ for inner rotation (+)} \\ -\text{sign}(b)\left(|b| + \sqrt{1+b^2}\right) \text{ for outer rotation (−)} \end{cases} \qquad (4.10)$$

with sign(b) = 1 for $b \geq 0$, and sign(b) = -1 for $b < 0$. Then from equation (4.3) substituting $s = \sqrt{1-c^2}$ in the definition of $t = \frac{s}{c}$ we get

$$c = \frac{1}{\sqrt{1+t^2}} \text{ and thus } s = tc. \qquad (4.11)$$

Thus the algorithm steps are a) find b, b) compute t from (4.10) c) use equation (4.11) to compute s and c, and d) Θ^T will give you the eigenvectors and d_1, d_2 are the eigenvalues.

In general this procedure done once can zero out any set of two symmetric off diagonal points for any dimension matrix **X** of size NxN. Lets define a rotation matrix $\Theta_{i,j}$ of size NxN, as the rotation matrix that zeros elements $x_{i,j}$ (element of **X** at i^{th} row, and j^{th} column or **X**(i,j)), and $x_{j,i}$ for $N \geq j > i$ as

$$\boldsymbol{\Theta}_{i,j} = \begin{bmatrix} 1 & & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots \\ 0 & & c & \cdots & s & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & & -s & \cdots & c & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots \\ 0 & & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \begin{matrix} \\ \\ \leftarrow \text{row i} \\ \\ \leftarrow \text{row j} \\ \\ \\ \end{matrix} \qquad (4.12)$$

$$\begin{matrix} & \uparrow & & \uparrow & \\ & \text{col.i} & & \text{col. j} & \end{matrix}$$

Let matrix $\mathbf{Y} = \Theta_{i,j}{}^{T}\mathbf{X}\Theta_{i,j}$. When the rotations are enforced on $\mathbf{X}$ the resulting elements of $\mathbf{Y}$ can be categorized into the affected elements and the unaffected elements. The unaffected elements are those elements of $\mathbf{Y}$ that are unchanged through the rotation process or they are the same as the elements of $\mathbf{X}$ at that same position. From the structure of matrix $\Theta_{i,j}$ one can observe that the affected elements are the elements that belong to the i^{th} and j^{th} column and the ones that belong to the i^{th} and j^{th} row of $\mathbf{Y}$. Caring out a few set of rotations one can see a pattern emerging, and categorize the new elements (of $\mathbf{Y}$) as we did and show below:

$$\mathbf{Y}(m,n) = \mathbf{X}(m,n), \text{ for } m \neq i, j, n \neq i, j \qquad (4.13a)$$

$$\mathbf{Y}(m,n) = c\mathbf{X}(i,n) - s\mathbf{X}(j,n), \text{ for } m = i, n \neq i, j \qquad (4.13b)$$

$$\mathbf{Y}(m,n) = c\mathbf{X}(m,i) - s\mathbf{X}(m,j), \text{ for } n = i, m \neq i, j \qquad (4.13c)$$

$$\mathbf{Y}(m,n) = c\mathbf{X}(j,n) + s\mathbf{X}(i,n), \text{ for } m = j, n \neq i, j \qquad (4.13d)$$

$$\mathbf{Y}(m,n) = c\mathbf{X}(m,j) + s\mathbf{X}(m,i), \text{ for } n = j, m \neq i, j \qquad (4.13e)$$

$$\mathbf{Y}(i,j) = \mathbf{Y}(j,i) = 0 \tag{4.13f}$$

$$\mathbf{Y}(i,i) = c[c\mathbf{X}(i,i) - s\mathbf{X}(j,i)] - s[c\mathbf{X}(i,j) - s\mathbf{X}(j,j)] =$$

$$= c^2\mathbf{X}(i,i) + s^2\mathbf{X}(j,j) - sc[\mathbf{X}(j,i) + \mathbf{X}(i,j)] \tag{4.13g}$$

$$\mathbf{Y}(j,j) = s[s\mathbf{X}(i,i) + c\mathbf{X}(j,i)] + c[s\mathbf{X}(i,j) + c\mathbf{X}(j,j)] =$$

$$= s^2\mathbf{X}(i,i) + c^2\mathbf{X}(j,j) + sc[\mathbf{X}(j,i) + \mathbf{X}(i,j)]. \tag{4.13h}$$

Notice that equations (4.13b) to (4.13f) give the affected off-diagonal elements, equations (4.13g) and (4.13h) give the affected diagonal elements, and equation (4.13a) the unaffected elements of the rotation process. Also one can point out that since **X** is symmetric then the elements of **Y** given by equation (4.13b) are the same as those of (4.13c). Similarly the elements of **Y** given by equation (4.13d) are the same as those of (4.13e), which suggests that the resulting matrix **Y** after the plane rotations is symmetric since **X** is symmetric. For more clarity on the changes that **X** undergoes, we will demonstrate a series of two rotations on an arbitrary matrix **X** of size 4X4. The code to do this was written in Matlab. Let **X** be a unity energy matrix given as:

$$\mathbf{X} = \begin{bmatrix} 0.4079 & 0.1632 & 0.1224 & 0.0816 \\ 0.1632 & 0.3671 & 0.2040 & 0.1632 \\ 0.1224 & 0.2040 & 0.4895 & 0.2447 \\ 0.0816 & 0.1632 & 0.2447 & 0.3263 \end{bmatrix}. \tag{4.14}$$

The off-diagonal energy (sum of the square of all the off-diagonal components) of **X** can be calculated to be 0.3527. For the first rotation lets zero points **X**(1,2) and **X**(2,1). Then $\Theta_{1,2}$ can be calculated to be:

$$\mathbf{\Theta}_{1,2} = \begin{bmatrix} 0.7497 & -0.6618 & 0 & 0 \\ 0.6618 & 0.7497 & 0 & 0 \\ 0 & 0 & 1.0000 & 0 \\ 0 & 0 & 0 & 1.0000 \end{bmatrix}. \tag{4.15}$$

The resulting $\mathbf{Y}_1$ is then given by

$$\mathbf{Y}_1 = \begin{bmatrix} 0.5519 & 0 & 0.2267 & 0.1691 \\ 0 & 0.2231 & 0.0719 & 0.0683 \\ 0.2267 & 0.0719 & 0.4895 & 0.2447 \\ 0.1691 & 0.0683 & 0.2447 & 0.3263 \end{bmatrix}. \tag{4.16}$$

Notice that the changed elements are the ones that belong to rows and columns 1 and 2. The off-diagonal energy of $\mathbf{Y}_1$ can be calculated to be 0.2995 which is the off-diagonal of $\mathbf{X}$ (i.e. 0.3527) minus $2x(0.1632)^2$, which is the energy of the zeroed elements. Also notice that this reduction in the off-diagonal energy was deposited on diagonal components at positions (1,1) and (2,2). Thus in this sense the new matrix $\mathbf{Y}_1$ is closer to a diagonal form than $\mathbf{X}$. For the second rotation operating on the resulting $\mathbf{Y}_1$ matrix, lets zero points $\mathbf{Y}_1(2,4)$, and $\mathbf{Y}_1(4,2)$. Then $\Theta_{2,4}$ can be calculated to be:

$$\mathbf{\Theta}_{2,4} = \begin{bmatrix} 1.0000 & 0 & 0 & 0 \\ 0 & 0.8952 & 0 & 0.4456 \\ 0 & 0 & 1.0000 & 0 \\ 0 & -0.4456 & 0 & 0.8952 \end{bmatrix} \tag{4.17}$$

with resulting $\mathbf{Y}_2$ to be

$$\mathbf{Y}_2 = \begin{bmatrix} 0.5519 & -0.0754 & 0.2267 & 0.1514 \\ -0.0754 & 0.1891 & -0.0447 & 0 \\ 0.2267 & -0.0447 & 0.4895 & 0.2511 \\ 0.1514 & 0 & 0.2511 & 0.3603 \end{bmatrix}, \tag{4.18}$$

with off-diagonal energy of 0.2902. Theoretical explanations and proofs for these observations will follow.

We will now show the impact of these plane rotations to the off-diagonal energy of the resulting matrix **Y**. Since the plane rotations are orthogonal or form a 'similarity' transform (i.e $\Theta_{i,j}{}^T\Theta_{i,j} = \mathbf{I}$), then the eigenvalues of **Y** (λ_{Yk} for k = 1, …, N) are the same as the eigenvalues of **X** (λ_{Xk} for k = 1, …, N). Also from matrix theory we know that for any square matrix **A** NxN, the following is true:

$$\text{Trace}(\mathbf{A}^n) = \sum_{k=1}^{N} \lambda_k^n \,. \tag{4.19}$$

Also it can be shown that if **A** is symmetric then

$$\text{Trace}(\mathbf{A}^2) = \sum_{k=1}^{N}\sum_{m=1}^{N} a_{k,m}^2 \tag{4.20}$$

or in other words the trace of $\mathbf{A}^2$ is the total energy of the matrix. This being said we then can write for matrices **X** and **Y** the following:

$$\text{Trace}(\mathbf{X}^2) = \sum_{k=1}^{N}\sum_{m=1}^{N} x_{k,m}^2 = \sum_{k=1}^{N} \lambda_{Xk}^2 = \sum_{k=1}^{N} \lambda_{Yk}^2 = \sum_{k=1}^{N} \lambda_k^2 = \sum_{k=1}^{N}\sum_{m=1}^{N} y_{k,m}^2 = \text{Trace}(\mathbf{Y}^2)\,. \tag{4.21}$$

Let us now consider the energy of the affected off-diagonal elements of **Y** as given by equations (4.13b-e). The energy contained in equations (4.13b) and (4.13d) can be written as:

$$\begin{aligned}\underset{m=i,n\neq i,j}{\mathbf{Y}(m,n)^2}+\underset{m=j,n\neq i,j}{\mathbf{Y}(m,n)^2} &= \left(c\mathbf{X}(i,n)-s\mathbf{X}(j,n)\right)^2+\left(c\mathbf{X}(j,n)+s\mathbf{X}(i,n)\right)^2 = \\ &= c^2\mathbf{X}(i,n)^2+s^2\mathbf{X}(j,n)^2+c^2\mathbf{X}(j,n)^2+s^2\mathbf{X}(i,n)^2 = \\ &= \left(c^2+s^2\right)\mathbf{X}(i,n)^2+\left(c^2+s^2\right)\mathbf{X}(j,n)^2\end{aligned} \tag{4.22}$$

Substituting equation (4.3) into (4.22) we then get:

$$\mathbf{Y}(i,n)^2+\mathbf{Y}(j,n)^2=\mathbf{X}(i,n)^2+\mathbf{X}(j,n)^2,\ \text{for } n\neq i,j. \tag{4.23}$$

Similarly the energy of elements of equations (4.13c) and (4.13e) can be found as

$$\mathbf{Y}(m,i)^2+\mathbf{Y}(m,j)^2=\mathbf{X}(m,i)^2+\mathbf{X}(m,j)^2,\ \text{for } m\neq i,j. \tag{4.24}$$

The results of equations (4.23), and (4.24) indicate that the elements of equations (4.13b-e), change values but their energy sum stays the same through the rotations. This means that the only off-diagonal energy change came from the two elements of equation (4.13f), which are the elements that were zeroed by the rotations. Since the energy before and after the rotations is the same (equation (4.21)) then the energy change was added to the two diagonal elements of equations (4.13g) and (4.13h) or one can write:

$$\mathbf{Y}(i,i)^2+\mathbf{Y}(j,j)^2=\mathbf{X}(i,i)^2+\mathbf{X}(j,j)^2+\mathbf{X}(i,j)^2+\mathbf{X}(j,i)^2. \tag{4.25}$$

This means that with every set of rotations the off-diagonal energy decreases by the amount of energy that the zeroed elements had as described by equation (4.26) as

$$OFF(\mathbf{Y}) = OFF(\mathbf{X})-\mathbf{X}(i,j)^2 - \mathbf{X}(j,i)^2 = OFF(\mathbf{X})-2\mathbf{X}(i,j)^2, \tag{4.26}$$

where

$$OFF(\mathbf{Y}) = \sum_{i}\sum_{\substack{j \\ i\neq j}} \mathbf{Y}(i,j)^2 . \tag{4.27}$$

If the elements to be eliminated are appropriately chosen and with enough repetitions of the plane rotations, then the resulting $\mathbf{Y}_k$ (at step k) matrix will be diagonalized within a certain error that we can tolerate. In other words the elements of the diagonal of the resulting matrix $\mathbf{Y}$ are going to approximate the eigenvalues of $\mathbf{X}$. Thus the multiplication of all the Θ_k^T (for k = 1, ... K, where K is the total rotation repetitions) used in repeated rotations will result to the eigenvectors of $\mathbf{X}$ as shown in equation (4.28).

$$\mathbf{X} = \mathbf{Q}\Lambda\mathbf{Q}^T = \Theta_K^T \ldots \Theta_1^T \mathbf{Y}_k \Theta_1 \ldots \Theta_K \tag{4.28a}$$

$$\mathbf{Q} \approx \Theta_K^T \ldots \Theta_1^T \tag{4.28b}$$

$$\Lambda \approx \mathbf{Y}_k \tag{4.28c}$$

This is the basis of the diagonalization procedures that we used end evaluated.

4.1.3 Plane Rotation Methods for Diagonalization

The first method we used is called the Jacobi method. This method before every set of rotations it searches for the maximum off-diagonal component and zeros it out. It has extremely fast convergence rate, the fastest of all the methods used, but the down side to it is that the search for the maximum component takes is a computationally expensive process. By convergence speed we mean the number of rotations required to converge to the desired minimum error.

The second method used is called the Serial or Cyclic Jacobi method. It is a variation of the Jacobi method that consists of annihilating systematically all of the off-diagonal components row-by-row or column-by-column, avoiding the searching process for a maximum element. Once the matrix is scanned once, the process is repeated until the minimum error is reached. This method is the one most widely used because of the relative easy programming that is required, and the low computational cost. The disadvantage of this method is that it is not as fast converging to the minimum acceptable error as the Jacobi method. Proofs for convergence of this method has been presented in [PHGF60].

The third method used is a method we thought intuitively and it can be described as a hybrid of the Jacobi and Cyclic Jacobi methods. It consists of annihilating systematically all of the off-diagonal components row-by-row or column-by-column, but skipping components that have energy below a certain threshold. In general there are many methods that use this threshold idea [PHGF60] in different forms but result to higher potential minimum error with little if any improvement in convergence speed. In our case we chose the threshold to change or adapt with every set of rotations. Adaptation usually means

significantly additional computational expense but not in our case, sine we take advantage of the plane rotation properties. More specifically adaptation is achieved by choosing the threshold to be a constant percentage γ of the total off-diagonal energy at that instance before the rotation. The off-diagonal energy of a matrix is evaluated only once before the first rotation, and adapted every time by simply subtracting the energy of the annihilated element as shown in equation (4.26). This saves the time of computing the off-diagonal energy at every rotation. This method although having a computational cost (execution time) that is close to the serial Jacobi, it performs very close to the Jacobi method convergence speed wise (required plane rotations). The downside is that the choice of threshold γ must be chosen carefully for the minimum error requirements.

In the next section we will examine how these methods perform in the subspaces, and demonstrate how we can take advantage of the subspace structure to improve even more the convergence speed for the eigenvalue-eigenvector calculation.

4.2 EVD in Subspaces

The covariance matrix $\mathbf{R}_s$ (of size NxN) at any given scale can be evaluated with respect to the covariance matrices in the next scale, and can be written in two forms (structures) as derived in Chapter 2. These expressions are (2.46) and (2.48). The matrices of interest are $\mathbf{R}_1$ and $\mathbf{R}_2$ given by:

$$\mathbf{R}_1 = \begin{bmatrix} \mathbf{R}_a & \mathbf{R}_{ab} \\ \mathbf{R}_{ab}^T & \mathbf{R}_b \end{bmatrix}, \tag{4.29a}$$

$$\mathbf{R}_2 = \begin{bmatrix} \boldsymbol{\Lambda}_a & \overline{\mathbf{R}}_{ab} \\ \overline{\mathbf{R}}_{ab}^T & \boldsymbol{\Lambda}_b \end{bmatrix}, \tag{4.29b}$$

where all variables are as defined as in Chapter 2. We showed (section 2.3.2) that the eigenvalues of $\mathbf{R}_1$ and $\mathbf{R}_2$ are the same as the ones of $\mathbf{R}_s$. Also we established relationships between the eigenvectors of $\mathbf{R}_s$, $\mathbf{R}_1$, and $\mathbf{R}_2$ (equation (2.50)). This fact makes the matrices $\mathbf{R}_1$ and $\mathbf{R}_2$ alternative structures to apply the EVD instead of $\mathbf{R}_s$, which present significant advantages, as we will demonstrate in this section.

The first advantage we immediately see is that inherently $\mathbf{R}_{ab}$ (and thus $\overline{\mathbf{R}}_{ab}$) have in general very little energy compared to $\mathbf{R}_a$, and $\mathbf{R}_b$. In fact there are noise covariance matrices $\mathbf{R}_s$ and filter matrices $\mathbf{A}$, $\mathbf{B}$ that result to $\mathbf{R}_{ab}$ to be equal to $\mathbf{0}$. If this is the case then

$$\mathbf{R}_s = \begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_b \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}_a & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_b \end{bmatrix} \begin{bmatrix} \mathbf{Q}_a^T & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_b^T \end{bmatrix} \begin{bmatrix} \mathbf{A}^T \\ \mathbf{B}^T \end{bmatrix}, \tag{4.30}$$

which means that

$$\boldsymbol{\Lambda}_s = \begin{bmatrix} \boldsymbol{\Lambda}_a & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_b \end{bmatrix} \tag{4.31a}$$

$$\mathbf{Q}_s = \begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_b \end{bmatrix}. \tag{4.31b}$$

This gives us a great advantage since instead of evaluating the EVD of an NxN matrix ($\mathbf{R}_s$), we only require to do it for two ($\mathbf{R}_a$, and $\mathbf{R}_b$) MxM matrices (and that's only considering one stage decomposition). This can save significant time in our processing.

The most obvious case where this happens, is the trivial case of white noise with variance σ^2 i.e. $\mathbf{R_s} = \sigma^2\mathbf{I}$, which makes equation (2.44c) as $\mathbf{R_{ab}} = \mathbf{A}^T\mathbf{B} = \mathbf{0}$. Another case that comes inherently from the filter properties is when $\mathbf{R_s}$ is a matrix with all its elements equal to σ^2. To show this we will start with rewriting equation (2.44c) as

$$\mathbf{R_{ab}} = \mathbf{A}^T\mathbf{R_s}\mathbf{B} = \begin{bmatrix} a_{11} & \cdots & a_{N1} \\ \vdots & \ddots & \vdots \\ a_{1M} & \cdots & a_{NM} \end{bmatrix} \begin{bmatrix} r_{11} & \cdots & r_{N1} \\ \vdots & \ddots & \vdots \\ r_{1N} & \cdots & r_{NN} \end{bmatrix} \begin{bmatrix} b_{11} & \cdots & b_{1M} \\ \vdots & \ddots & \vdots \\ b_{N1} & \cdots & b_{NM} \end{bmatrix} = \\ = \begin{bmatrix} \sum_{j=1}^{N}\sum_{i=1}^{N} a_{i1}r_{ij}b_{j1} & \cdots & \sum_{j=1}^{N}\sum_{i=1}^{N} a_{i1}r_{ij}b_{jM} \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^{N}\sum_{i=1}^{N} a_{iM}r_{ij}b_{j1} & \cdots & \sum_{j=1}^{N}\sum_{i=1}^{N} a_{iM}r_{ij}b_{jM} \end{bmatrix} \tag{4.32}$$

From the properties of wavelet filters we can write these relationships for Daubechies wavelets, but are true with different constant for all orthogonal and biorthogonal wavelets.

$$\sum_{i=1}^{N} a_{in} = \sqrt{2}, \text{ for any n,} \tag{4.33a}$$

$$\sum_{i=1}^{M} a_{ni} = \frac{1}{\sqrt{2}}, \text{ for any n,} \tag{4.33b}$$

$$\sum_{i=1}^{M} b_{ni} = \begin{cases} \frac{1}{\sqrt{2}}, \text{ for n odd} \\ -\frac{1}{\sqrt{2}}, \text{ for n even} \end{cases} \tag{4.33c}$$

$$\sum_{i=1}^{N} b_{in} = 0, \text{ for any n.} \tag{4.33d}$$

Thus

$$\sum_{j=1}^{N}\sum_{i=1}^{N} a_{ik} r_{ij} b_{jn} = c\sum_{j=1}^{N}\sum_{i=1}^{N} a_{ik} b_{jn} = c\sum_{i=1}^{N} a_{ik} \underbrace{\sum_{j=1}^{N} b_{jn}}_{0} = 0. \qquad (4.44)$$

Also a combination of these two cases leads to a $\mathbf{R_{ab}} = \mathbf{0}$, i.e. if $\mathbf{R_s}$ has diagonal components equal to one constant and the off diagonal equal to another. This is the covariance matrix of a uniformly distributed noise.

In many applications even though $\mathbf{R_{ab}}$ is not identically zero, but has negligible energy compared to the energy of $\mathbf{R_a}$ and $\mathbf{R_b}$. Then the eigenvalues and vectors for whitening can be approximated by expressions (4.31), with satisfactory results. This approximation results are shown for the experimental noises as well as for the simulated pink noise we used in Figures 4.2-4.5. The noise vectors were chosen to be of length 32. In our simulations we used db3 (Daubechies 3) wavelet decomposition. We see that the approximated eigenvalues for the most part are very close to the real ones, especially when they have small values, which would give the maximum error in the whitening process. The best result (least approximation error) we get in the simulated pink noise scenario and the worse in the experimental see noise. Another important factor for $\mathbf{R_{ab}}$ matrix to have zero or close to zero elements, is the choice of the wavelet family (the set of **A**, **B** matrices). This fact can be illustrated in Figure 4.6, which is the case of the experimental see noise with the discrete Meyer wavelet. We immediately see much improvement in the estimated eigenvalues as far as better approximating the real ones. We leave the examination of finding the optimum wavelet for a given noise for future research.

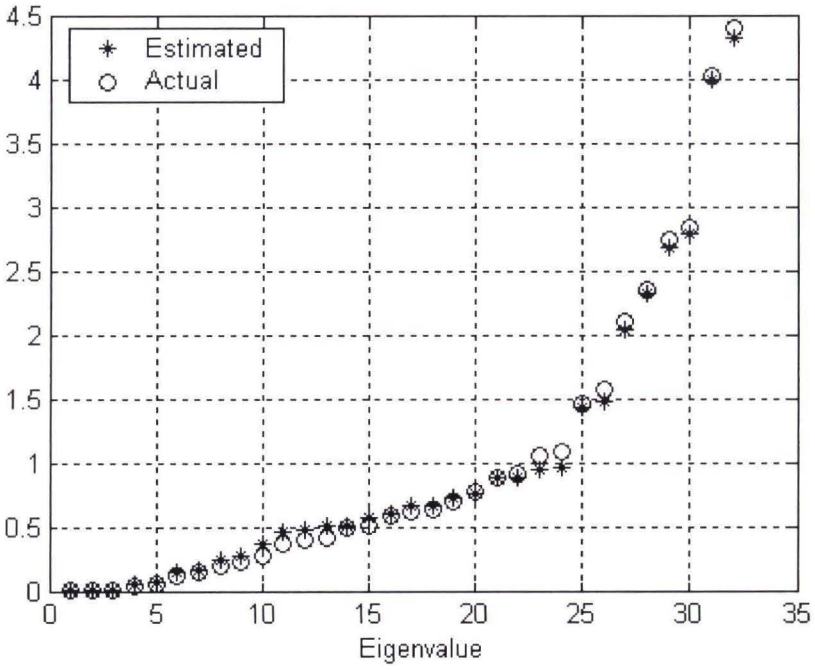


Figure 4.2: Eigenvalue Results for Babble Noise, db3.

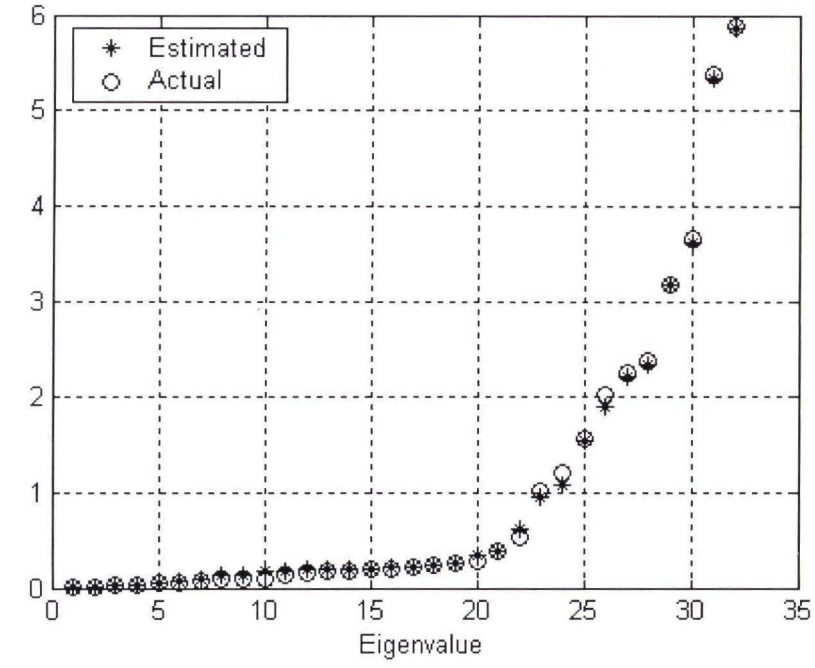


Figure 4.3: Eigenvalue Results for Vehicle Noise, db3.

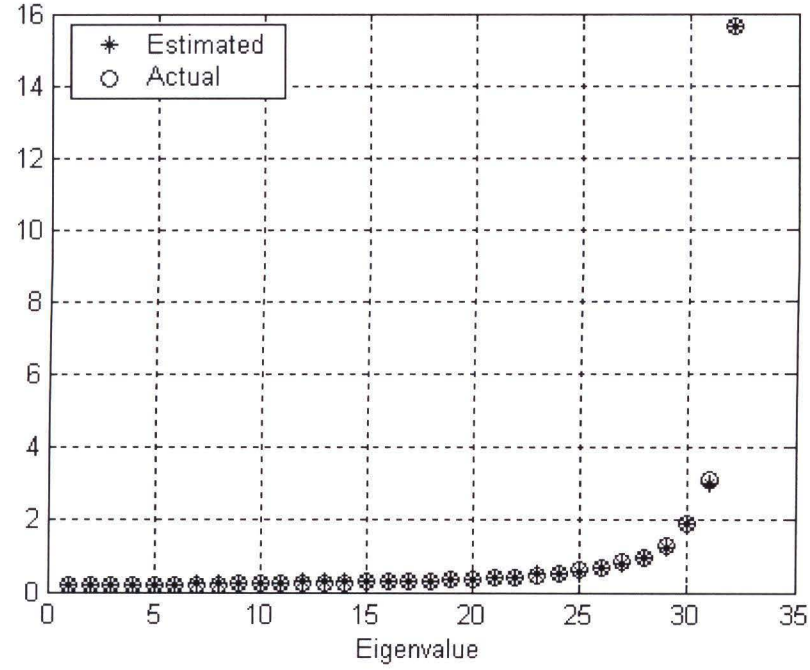


Figure 4.4: Eigenvalue Results for Pink Noise, db3.

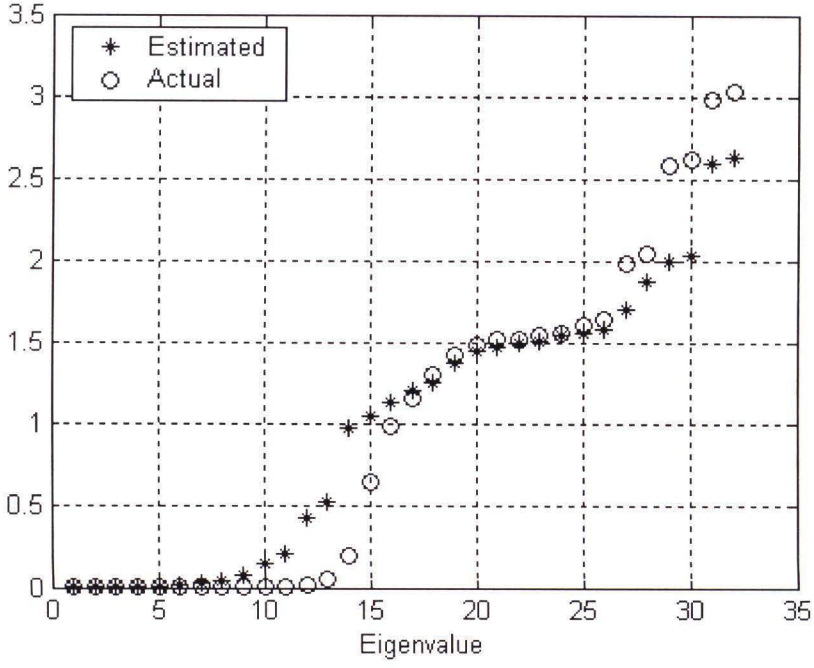


Figure 4.5: Eigenvalue Results for Sea Noise, db3.

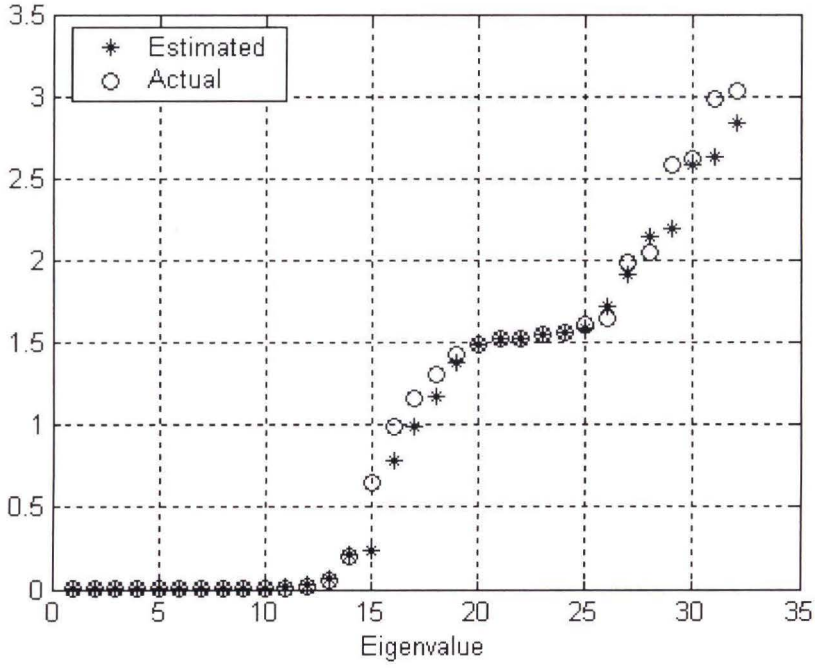


Figure 4.6: Eigenvalue Results for Sea Noise, Discrete Meyer.

4.3 Effect of Plane Rotation Methods on Subspace EVD

In many applications an approximation of the eigenvalues of a process is not acceptable. In cases like this (given that $\mathbf{R}_{ab} \neq \mathbf{0}$) one has to perform EVD on any of the three NxN size matrices $\mathbf{R}_s$, or $\mathbf{R}_1$, or $\mathbf{R}_2$. In this section we will demonstrate the clear advantages of performing EVD through plane rotations, not directly on $\mathbf{R}_s$ but on matrices $\mathbf{R}_1$ or $\mathbf{R}_2$. We will show through simulations that not only the convergence speed increases, but also the initial off-diagonal energy reduces from using $\mathbf{R}_s$ to $\mathbf{R}_1$ and even more by using $\mathbf{R}_2$, at least for the noise scenarios we dealt with. These simulations are done using all three plane rotation methods and thus we will get an opportunity to examine how the Hybrid Jacobi method performs compared to the other methods (Jacobi, and Cyclic Jacobi). The noise vectors had length 16 or each matrix had 256 elements and the wavelet used was db3.

For comparison purposes all noises are normalized to result in a covariance matrix $\mathbf{R}_s$ energy equal to one. The simulations are ended when the off diagonal energy is less than 10^{-14}. The simulations results are shown in Figures 4.7-4.10 for babble, vehicle, pink, and ocean noise respectively. For clarity of results and comparison purposes, Tables 4.1-4.4 were added to present results of the off-diagonal energy at the 60th plane rotation.

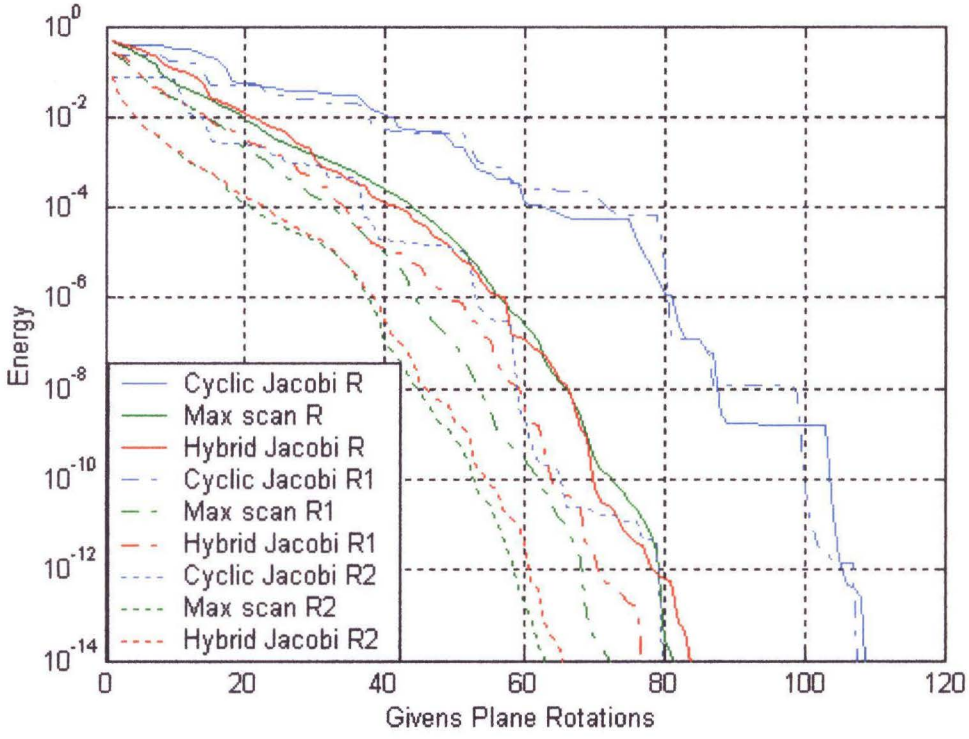


Figure 4.7: Off-Diagonal Energy for Babble Noise.

Matrix / Method	Cyclic Jacobi	Hybrid Jacobi	Jacobi
$\mathbf{R}_s$	$1.2\text{x}10^{-4}$	$1.1\text{x}10^{-7}$	$2.5\text{x}10^{-7}$
$\mathbf{R}_1$	$2.5\text{x}10^{-4}$	$3.3\text{x}10^{-9}$	$2.8\text{x}10^{-10}$
$\mathbf{R}_2$	$2.1\text{x}10^{-9}$	$4.1\text{x}10^{-12}$	$1.9\text{x}10^{-13}$

Table 4.1: Off-Diagonal Energy for Babble Noise for 60 Plane Rotations.

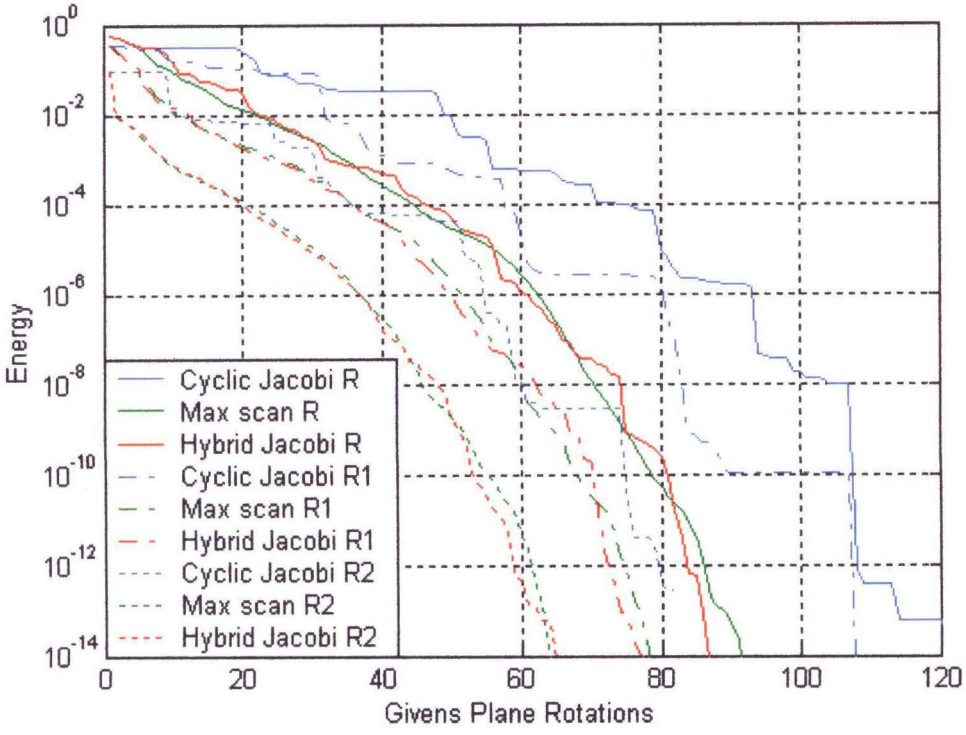


Figure 4.8: Off-Diagonal Energy for Vehicle Noise.

Matrix / Method	Cyclic Jacobi	Hybrid Jacobi	Jacobi
$\mathbf{R_s}$	$5.2\text{x}10^{-4}$	$1.02\text{x}10^{-6}$	$2.6\text{x}10^{-6}$
$\mathbf{R_1}$	$2\text{x}10^{-5}$	$1.2\text{x}10^{-8}$	$2.3\text{x}10^{-8}$
$\mathbf{R_2}$	$1\text{x}10^{-8}$	$4.5\text{x}10^{-13}$	$4.4\text{x}10^{-12}$

Table 4.2: Off-Diagonal Energy for Vehicle Noise for 60 Plane Rotations.

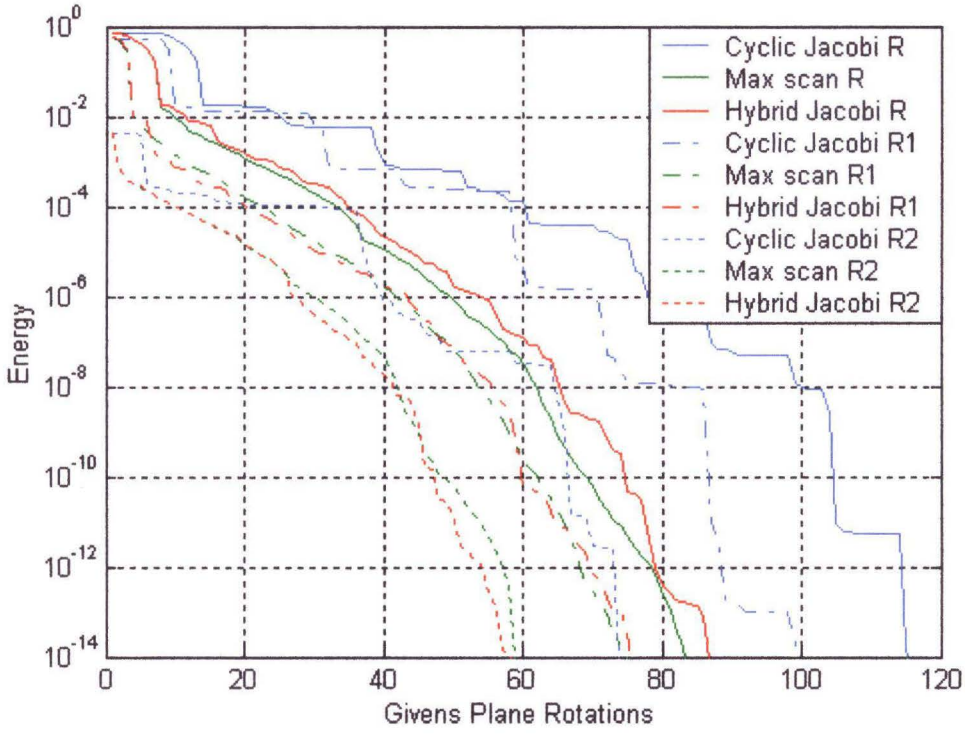


Figure 4.9: Off-Diagonal Energy for Pink Noise.

Matrix / Method	Cyclic Jacobi	Hybrid Jacobi	Jacobi
$\mathbf{R_s}$	1.01×10^{-4}	1.09×10^{-7}	2.8×10^{-8}
$\mathbf{R_1}$	3.1×10^{-6}	9.7×10^{-11}	1.12×10^{-10}
$\mathbf{R_2}$	5.1×10^{-8}	6.2×10^{-15}	6.1×10^{-15}

Table 4.3: Off-Diagonal Energy for Pink Noise for 60 Plane Rotations.

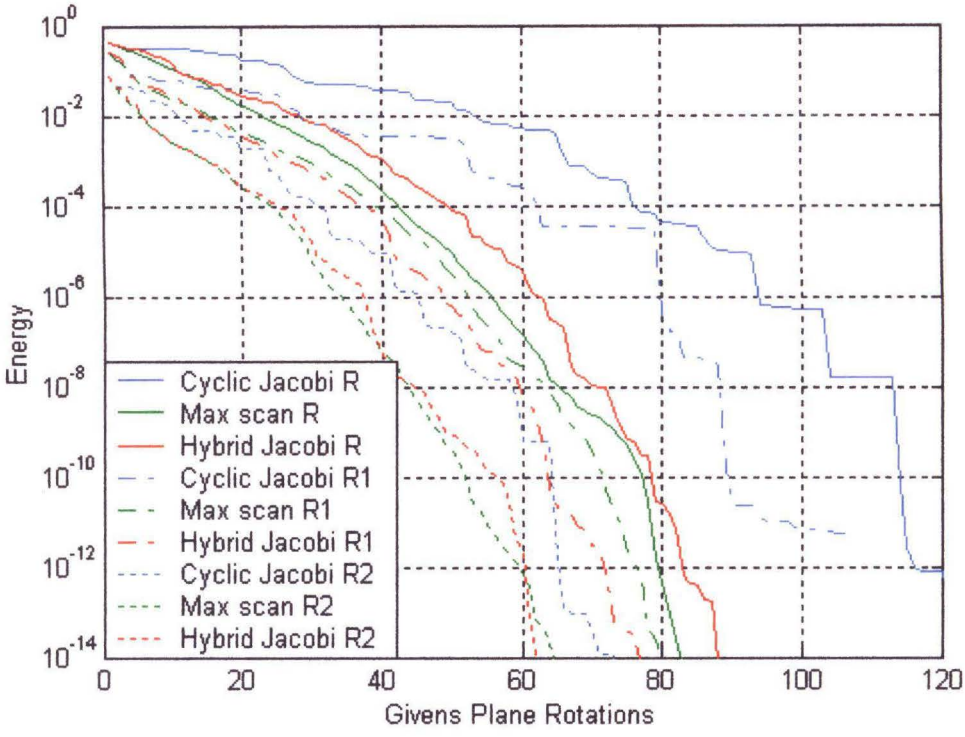


Figure 4.10: Off-Diagonal Energy for Ocean Noise.

Matrix / Method	Cyclic Jacobi	Hybrid Jacobi	Jacobi
$\mathbf{R_s}$	$5x10^{-3}$	$3.8x10^{-6}$	$1.3x10^{-7}$
$\mathbf{R_1}$	$1.5x10^{-4}$	$6x10^{-9}$	$2.4x10^{-8}$
$\mathbf{R_2}$	$6x10^{-10}$	$2.1x10^{-12}$	$7.1x10^{-13}$

Table 4.4: Off-Diagonal Energy for Ocean Noise for 60 Plane Rotations.

Looking at the results one can deduce, regarding the choice of matrix structure, that we achieve significantly better results using matrix $\mathbf{R}_1$ and even more using matrix $\mathbf{R}_2$ over matrix $\mathbf{R}_s$. To begin with, we can observe much faster convergence to the prescribed acceptable off-diagonal energy. This is true for all methods used and all the noises with the exception of the Cyclic Jacobi method in babble noise, which we did not see much improvement by using $\mathbf{R}_1$ over $\mathbf{R}_s$. As an example we point to the ocean noise case for the Hybrid Jacobi method for the sixty plane rotations mark, where the off-diagonal energy of $\mathbf{R}_1$ and $\mathbf{R}_2$ is approximately a thousand and a million times smaller respectively, than the one in matrix $\mathbf{R}_s$. Notice that the difference in the results, in all the cases, is amplified with the increase of the plane rotations. Another observation, even though we do not show the simulations, is that the improvement in the performance is even more noticeable as the covariance matrix increases in size. Furthermore, the initial off-diagonal energy is smaller in matrix $\mathbf{R}_1$ over $\mathbf{R}_s$ and even smaller in matrix $\mathbf{R}_2$. All these observations lead us to believe that the matrices structures ($\mathbf{R}_1$, and $\mathbf{R}_2$) derived from subspaces are an attractive alternative to the input matrix ($\mathbf{R}_s$) structure for the EVD problem.

Regarding the EVD methods we can conclude that the in all the cases the Jacobi and Hybrid Jacobi outperform the Cyclic Jacobi method convergence wise (required plane rotations), with the Jacobi and Hybrid Jacobi having very similar results. Considering the fact though that the Hybrid Jacobi does not have to go through a sorting procedure before every plane rotation, which is very computationally expensive, then this method seems the best choice for a given noise.

CHAPTER 5

SUMMARY, CONCLUSIONS, AND FUTURE RESEARCH

5.1 Summary and Conclusions

A measure of the potential of a receiver for detection is detectability. Detectability is a function of the signal and noise, and given any one of them the detectability is fixed. In addition, complete transforms of the signal and noise cannot change detectability. Throughout this work we demonstrated (Chapter 3) that "Subspace methods" as defined in Chapter 2 improve detectability in certain subspaces resulting in improved Receiver Operating Curves (ROC) and thus better detection in arbitrary noise environments in this subspaces. Our method is tested and verified on various signals and noises, both simulated and real. More specifically we used wavelet (db3) subspaces and evaluated signal detection potential in different scales and branches for vehicle, ocean, and babble noise. Indeed for a given signal and noise we verified that there can exist a subspace that the signal is concentrated where as the projection of the noise in that same subspace is of a very small energy. This results to detectability improvement in the order of four to sixty times depending on the noise and whether we evaluated the synthesis or analysis part of the wavelet decomposition. Consequently in these subspaces the probability of detection for a given probability of false alarm significantly improves compared to the ROC that result by performing optimum detection in on the original signal and noise.

The optimum detection of signals in noise requires the computation of noise eigenvalues and vectors (EVD). This process neither is a trivial one nor is it computationally cheap, especially for non-stationary noise and can result in numerical instabilities when the covariance matrix is large. This work addressed this problem and provided solutions in Chapter 4 that take advantage of the subspace structure through plane rotations to improve on existing algorithms for EVD by improving their convergence rate and reducing their computational expense for given thresholds. More specifically, we proposed two covariance matrix structures that result from subspace decomposition that can be used to evaluate the noise statistics (KL basis) with a clear advantage regarding convergence speed and initial off diagonal energy using various plane rotation methods. Moreover we proposed an algorithm, the hybrid Jacobi, that is almost as computationally cheap as the Cyclic Jacobi algorithm but performs similarly to the computationally expensive Jacobi algorithm. The drawback of this method is that the parameter used had to be predetermined and is in general different for any arbitrary noise environment, but for any given communication system where the noise expected does not change characteristics it is an attractive alternative to the traditional methods.

5.2 Future Research

The detectability of the signal in a given arbitrary noise in subspaces is greatly affected by the choice of the subspace transform used. The choice of the subspace transform also affects the convergence speed of the EVD algorithms. When wavelets are used then there is a great variety of wave functions to chose from, depending on the noise and signal in hand. In our study we used db3 wavelet for subspace decomposition

acknowledging the fact that for our noise and signal environments a better transform may exist. The exploration of subspace transform alternatives for a given noise environment is left for future studies. More over the subspace detection is greatly affected by the signal choice. Throughout our work an attempt was made to choose a signal for a given noise, which places the signal in subspaces were the noise projections do not reside. Describing the exact process that leads to the optimum signal choice, given the subspace transform and the noise, can advance this work. Also the alternative of multiple signaling could be explored. Finally, the advantages and disadvantages of using all the subspaces for detection could be explored.

APPENDIX I

Detectability and SNR After Whitening

In this appendix we will derive the relationships of the SNR and the detectability before and after the whitening process. Let the vector $\mathbf{r} = [r_1\ r_2\ \ldots\ r_N]^T = \mathbf{s} + \mathbf{n}$ where s is and n is the signal and noise column vectors respectively. The covariance matrix of the noise is defined as $\mathbf{R} = E[\mathbf{n}\mathbf{n}^T] = \mathbf{Q}\Lambda\mathbf{Q}^T$, where $\mathbf{Q}$ is the eigenvector matrix of $\mathbf{R}$, and Λ is a diagonal matrix containing the eigenvalues of $\mathbf{R}$. Let the whitening process be based on the noise statistics as shown in Figure AI.1 below.

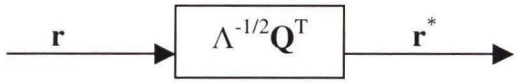


Figure AI.1: Whitening Process Using Noise Statistics.

The vector $\mathbf{r}^*$ is the whitened $\mathbf{r}$ vector which is equal to:

$$\mathbf{r}^* = \mathbf{s}^* + \mathbf{n}^* = \Lambda^{-1/2}\mathbf{Q}^T\mathbf{s} + \Lambda^{-1/2}\mathbf{Q}^T\mathbf{n}. \qquad \text{(AI.1)}$$

Evidently the resulting noise $\mathbf{n}^*$ is white i.e.

$$\mathbf{R}^* = E[\mathbf{n}^*\mathbf{n}^{*T}] = E[\Lambda^{-1/2}\mathbf{Q}^T\mathbf{n}\mathbf{n}^T\mathbf{Q}\Lambda^{-1/2}] = \Lambda^{-1/2}\mathbf{Q}^T E[\mathbf{n}\mathbf{n}^T]\mathbf{Q}\Lambda^{-1/2} =$$

$$= \Lambda^{-1/2}\mathbf{Q}^T\mathbf{R}\mathbf{Q}\Lambda^{-1/2} = \Lambda^{-1/2}\Lambda\Lambda^{-1/2} = \mathbf{I} = \mathbf{Q}^*\Lambda^*\mathbf{Q}^{*T}. \qquad \text{(AI.2)}$$

The detectability (equation 2.58) and SNR before the whitening process are given as

$$d_{in}^2 = \mathbf{s}^T\mathbf{R}^{-1}\mathbf{s} \tag{AI.3a}$$

$$SNR_{in} = \frac{E_s}{E_n} = \frac{\sum_{i=1}^{N} s(i)^2}{\sum_{i=1}^{N} \lambda_i} = \frac{\mathbf{s}^T\mathbf{s}}{\sum_{i=1}^{N} \lambda_i} \tag{AI.3b}$$

Similarly we can derive the detectability and SNR at the output of the whitening process as:

$$d_{out}^2 = \mathbf{s}^{*T}\mathbf{R}^{*-1}\mathbf{s}^* = \mathbf{s}^T\mathbf{Q}\Lambda^{-1/2}\Lambda^{-1/2}\mathbf{Q}^T\mathbf{s} = \mathbf{s}^T\mathbf{Q}\Lambda^{-1}\mathbf{Q}^T\mathbf{s} = \mathbf{s}^T\mathbf{R}^{-1}\mathbf{s} \tag{AI.4a}$$

$$SNR_{out} = \frac{E_{s^*}}{E_{n^*}} = \frac{\sum_{i=1}^{N} s^*(i)^2}{\sum_{i=1}^{N} \lambda_i} = \frac{\mathbf{s}^{*T}\mathbf{s}^*}{N} = \frac{\mathbf{s}^T\mathbf{R}^{-1}\mathbf{s}}{N}. \tag{AI.4b}$$

From these equations we conclude that the detectability does not change through the whitening process but the SNR does. Actually, form equation (AI.4b) we conclude that if the signal is whitened then the SNR is directly proportional to the detectability. In Chapter 3 we gave simulations results of the normalized SNR, and all that means is that we multiply the SNR by the observation vector length N, so that the detectability and the SNR have the same value along scales.

APPENDIX II

Eigenvalue and Detectability Evolution for Vehicle Noise

In this appendix we will show a numerical example on how the eigenvalues of the vehicle noise evolve in the wavelet subspace structure described in Chapter 3 (db3). Furthermore, we will show the projections of the signal onto the eigenvectors of the noise covariance matrix, as well as the detectability, evolution along scales. The signal used is the one in equation (3.3) with parameters N=16, a=14, b=17, T=2, and E_s= 0.2. The results are shown in Tables AII.1 – AII.3 for the input node (0,0), and the first level of subspace decomposition nodes (1,0) and (1,1) respectively. Moreover, Figure AII.1 shows the eigenvalues for these nodes. All variables used are as defined in section 2.4.

Analysis at Level (0,0)			
No	Eigenvalues (λ)	Signal Projections (s_k)	s_k^2/λ
1	0.0005	-0.0157	0.49298
2	0.0045	0.0451	0.45200
3	0.0214	-0.0796	0.29608
4	0.0313	-0.006	0.00115
5	0.0483	0.0303	0.01901
6	0.0515	-0.0497	0.04796
7	0.0655	-0.1413	0.30482
8	0.0815	0.1166	0.16682
9	0.0893	-0.0416	0.01938
10	0.1002	0.1326	0.17548
11	0.1377	0.1005	0.07335
12	0.157	0.0266	0.00451
13	0.176	-0.0803	0.03664
14	0.1814	0.057	0.01791
15	0.1998	-0.0824	0.03398
16	0.201	-0.0163	0.00132
17	0.2107	-0.1235	0.07239
18	0.2245	0.1446	0.09314
19	0.2551	0.0631	0.01561
20	0.2726	0.0385	0.00544
21	0.3858	-0.0324	0.00272
22	0.5088	-0.1562	0.04795
23	1.0132	-0.0788	0.00613
24	1.1911	-0.0754	0.00477
25	1.5541	-0.1062	0.00726
26	2.0219	-0.0822	0.00334
27	2.2225	-0.0712	0.00228
28	2.4168	-0.0092	0.00004
29	3.1934	-0.0336	0.00035
30	3.6347	0.0145	0.00006
31	5.3573	-0.0236	0.00010
32	5.9898	0.0587	0.00058
Detectability = $d^2 = \sum_{n=1}^{32} \frac{\lvert s_k(n) \rvert^2}{\lambda_n}$			2.40554

$$\sum_{n=1}^{32} |s_k(n)|^2 = 0.199975$$

$$\sum_{n=1}^{32} \lambda_n = 31.9992$$

Table AII.1: Results for node (0,0).

Analysis at Level (1,0)			
No	Eigenvalues (λ)	Signal Projections (s_k)	s_k^2/λ
1	0.1935	-0.0006	0.00000
2	0.2092	-0.0169	0.00137
3	0.2366	-0.031	0.00406
4	0.2446	-0.0429	0.00752
5	0.3833	-0.0751	0.01471
6	0.4619	0.0791	0.01355
7	0.9267	0.075	0.00607
8	1.1106	0.0407	0.00149
9	1.5776	0.0791	0.00397
10	1.8508	-0.0258	0.00036
11	2.1913	0.0201	0.00018
12	2.4426	-0.0787	0.00254
13	3.0779	-0.0812	0.00214
14	3.6201	-0.0003	0.00000
15	5.3617	0.0183	0.00006
16	6.2363	0.075	0.00090
Detectability = $d^2 = \sum_{n=1}^{16} \frac{\lvert s_k(n) \rvert^2}{\lambda_n}$			0.05893

$$\sum_{n=1}^{16} \lvert s_k(n) \rvert^2 = 0.048339$$

$$\sum_{n=1}^{16} \lambda_n = 30.1247$$

Table AII.2: Results for node (1,0).

Analysis at Level (1,1)			
No	Eigenvalues (λ)	Signal Projections (s_k)	s_k^2/λ
1	0.0005	0.0236	1.11392
2	0.0045	0.0657	0.95922
3	0.0211	0.11	0.57346
4	0.0365	0.0134	0.00492
5	0.0547	0.098	0.17558
6	0.0631	0.1421	0.32001
7	0.0756	0.185	0.45271
8	0.1196	0.1574	0.20715
9	0.1379	0.1007	0.07354
10	0.157	0.1521	0.14735
11	0.1713	-0.0228	0.00303
12	0.1805	-0.0862	0.04117
13	0.1926	0.0358	0.00665
14	0.2022	-0.0348	0.00599
15	0.2219	-0.0443	0.00884
16	0.2295	-0.0061	0.00016
Detectability = $d^2 = \sum_{n=1}^{16} \frac{\lvert s_k(n) \rvert^2}{\lambda_n}$			4.09370

$$\sum_{n=1}^{16} \lvert s_k(n) \rvert^2 = 0.151667$$

$$\sum_{n=1}^{16} \lambda_n = 1.8685$$

Table AII.3: Results for node (1,1).

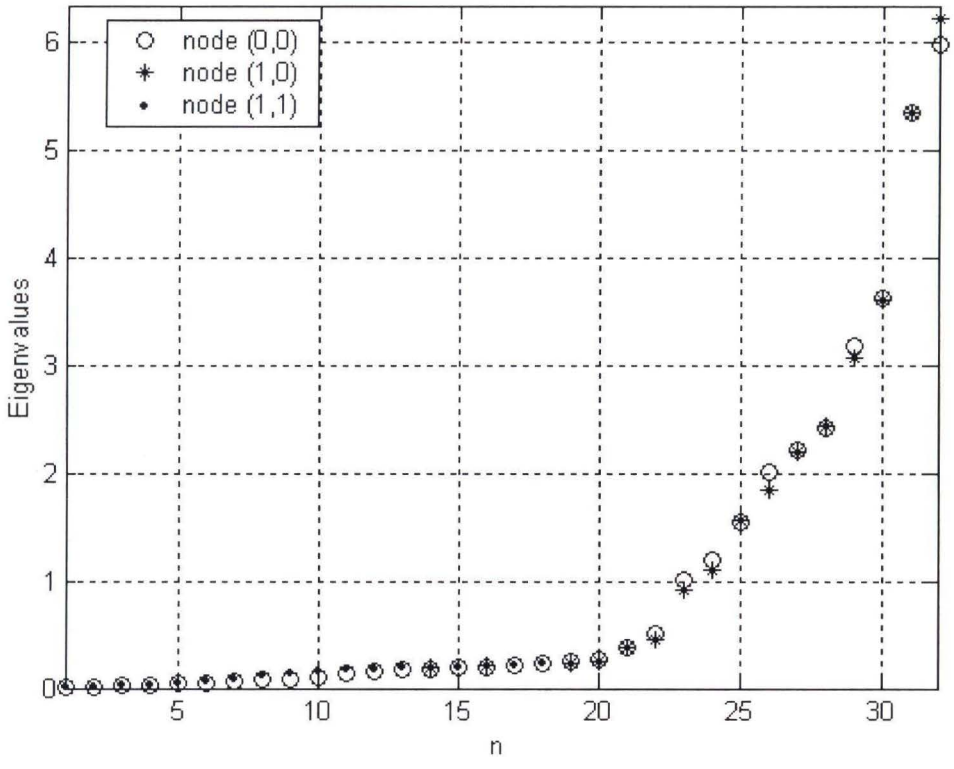


Figure AII.1: Eigenvalues for Parent – Children Nodes.

APPENDIX III

Positive Definite Matrices

In this appendix we will describe the criterion for a matrix to be positive definite and their corresponding properties. A conjugate symmetric matrix **R** of size MxM is called positive definite if and only if the Hermitian form [DMVI00]:

$$\mathbf{x}^H\mathbf{R}\mathbf{x} = \sum_{i,j}^{M} r_{ij} x_i^* x_j > 0 \text{ , for } \mathbf{x} \neq \mathbf{0}. \qquad \text{(AIII.1)}$$

This is not the only criterion though that establishes a positive definite property of a matrix. More precisely an MxM matrix **R** is positive definite if and only if it satisfies any one of the following criteria:

1. $\mathbf{x}^H\mathbf{R}\mathbf{x} > 0$ for all nonzero vectors **x**.

2. All eigenvalues of **R** are positive.

3. All principal submatrices $\mathbf{R}_m$ (comprised of the first m columns and rows of **R**), $1 \leq m \leq M$, have positive determinants.

4. There exists an LxM, M > L, matrix **S** with linearly independent columns such that $\mathbf{R} = \mathbf{S}^H\mathbf{S}$ (i.e. **S** is of rank M).

5. There exists a non-singular MxM matrix **W** such that $\mathbf{R} = \mathbf{W}^H\mathbf{W}$ (i.e. **W** could be an orthogonal matrix obtained from the eigenvectors of **R**, or a triangular matrix obtained from Cholesky's decomposition).

6. There exists a non-singular MxM matrix **P** such that $\mathbf{P}^H\mathbf{RP}$ is positive definite.

Using these criteria some properties of positive definite matrices are listed below:

1. The diagonal elements of **R** are positive.

2. $r_{ii}r_{jj} > |\, r_{ij} \,|^2$, $(i \neq j)$.

3. The element of **R** with the largest absolute value lies on the diagonal.

4. The determinant of $\mathbf{R} > 0$. Hence **R** is non-singular.

5. The inverse matrix $\mathbf{R}^{-1}$ is positive definite.

6. The matrix obtained by deleting a row and the corresponding column from **R** is positive definite.

BIBLIOGRAPHY

[ANSH01] Ambient Noise Sonar Homepage, *http://www.oe.fau.edu/~acoustics/.*

[ATMK92] A. H. Tewfik, and M. Kim, 'Correlation Structure of the Discrete Wavelet Coefficients of Fractional Brownian Motions', *IEEE Transactions on Information Theory*, vol.38, no. 2, pp. 904-909, March 1992.

[BFBP89] B. Friedlander and Boaz Porat, 'Detection of Transient Signals by the Gabor Representation', *IEEE. Transactions on Acoustics Speech and Signal Processing*, Feb. 1989.

[BSRG98] C. Sidney Burrus, Ramesh A. Gopinath, and Haito Guo, *Intoduction to wavelets and wavelet transforms*, Prentice Hall, 1998.

[CHDW89] C. E. Heil and D. F. Walnut, 'Continues and discrete wavelet transforms', *SIAM review*, December 1989.

[CWHE95] C.W. Helstorm, *Elements of Signal Detection and Estimation*, Prentice Hall, 1995.

[DMVI00] Dimitris G. Manolakis, Vinay K. Ingle, and Stephen M. Kogon, *Statistical and adaptive signal processing*, Mc Graw Hill, 2000.

[DSHL78] D. Slepian and H. J. Landau, 'A note on the eigenvalues of Hermitian matrices', *SIAM\ Journal of Math. Anal.*, Vol. 9, No.2, April 1978.

[ERHA63] Eldon R. Hansen, 'On Cyclic Jacobi Methods', *Journal of the Society for Industrial and Applied Mathematics*, Vol. 11, No. 2, pp. 448-459, Jun., 1963.

[FLHP89] Franklin T. Luk, 'A Proof of Convergence for Two Parallel Jacobi SVD Algorithms', *IEEE. Transactions on Computers*, vol. 38. No. 6, June 1989.

[GFLI97] Garcia, F.M. Lourtie, I.M.G., 'Detection of Gaussian band pass transients under impulsive noise: a wavelet transform approach', *Acoustics, Speech, and Signal Processing*, 1997. ICASSP-97, 1997 IEEE International Conference on, Volume: 1, 1997.

[GWAO92] Wornell, G.W. Oppenheim, A.V., 'Estimation of fractal signals from noisy measurements using wavelets', *Signal Processing, IEEE Transactions on*, Volume: 40 Issue: 3, March 1992.

[GWOR90] Wornell, G.W., 'A Karhunen-Loeve-like expansion for 1/f processes via wavelets', *Information Theory, IEEE Transactions on*, Volume: 36 Issue: 4, July 1990.

[GWRL90] Wornell, G.W, 'Wavelet-based representations for the 1/f family of fractal processes', *Proceedings of the IEEE*, Volume: 81 Issue: 10, Oct. 1993.

[HKKD92] Krim, H. Drouiche, K.; Pesquet, J.C., 'Multiscale detection of nonstationary signals, Time-Frequency and Time-Scale Analysis', *Proceedings of the IEEE-SP International Symposium*, 1992.

[HVRS68] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*, John Wiley and Sons, 1968.

[IDAU92] Ingrid Daubechies, 'Ten lectures on wavelets', *SIAM Philadelphia*, PA, 1992.

[MFHM94] Frisch, M. Messer, H., 'Detection of a known transient signal of unknown scaling and arrival time', *Signal Processing, IEEE Transactions on*, Volume: 42 Issue: 7, July 1994.

[MFHM92] Frisch, M. Messer, H., 'The use of the wavelet transform in the detection of an unknown transient signal', *Information Theory, IEEE Transactions on*, Volume: 38 Issue: 2 Part: 2, March 1992.

[MMYM96] Michel Misiti, Yves Misiti, George Oppenheim, Jean-Michel Poggi, *Wavelet Toolbox*, March 1996.

[NERS00] Erdol, N. and R.S. Schultz, 'Voice Activity Detection Over Multiresolution Subspaces', *Proceedings of the IEEE Sensor Array Multichannel Signal Processing Workshop.*, pp.217-220, 2000.

[NESK01] Nurgun Erdol, Spyros Kyperountas and Branko Petljanski, 'Time and Scale Evolutionary EVD and Detection', *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2001.

[NUER99] N. Erdol, *Multiscale Detection in Wavelet and Subband Transforms in Communications and Multimedia*, Editors: Ali N. Akansu and Michael J. Medley. Kluwer Books, 148-167, 1999.

[PAWH58] Paul A. White, 'The Computation of Eigenvalues and Eigenvectors of a Matrix', *Journal of the Society for Industrial and Applied Mathematics*, Vol. 6, No. 4, pp. 393-437, Dec., 1958.

[PFLA92] Flandrin, P., 'Wavelet analysis and synthesis of fractional Brownian motion', *Information Theory, IEEE Transactions on*, Volume: 38 Issue: 2 Part: 2, March 1992.

[PHEN58] Peter Henrici, 'On the Speed of Convergence of Cyclic and Quasicyclic Jacobi Methods for Computing Eigenvalues of Hermitian Matrices', *Journal of the Society for Industrial and Applied Mathematics*, Vol. 6, No. 2, pp. 144-162, Jun., 1958.

[PHGF60] G. E. Forsythe, P. Henrici, 'The Cyclic Jacobi Method for Computing the Principal Values of a Complex Matrix', *Transactions of the American Mathematical Society*, Vol. 94, No. 1, pp. 1-23, Jan., 1960.

[PJEB62] P. J. Eberlein ,'A Jacobi-like Method for the Automatic Computation of Eigenvalues and Eigenvectors of an Arbitrary Matrix', *Journal of the Society for Industrial and Applied Mathematics*, Vol. 10, No. 1, pp. 74-88, Mar., 1962.

[PPVA93] P.P Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice Hall Signal Processing Series, 1993.

[SGML89] S. G. Mallat, 'Multiresolution Approximations and Wavelet Orthonormal Bases of L2(R)', *Trans. Am. Math. Soc. 315*, 1989.

[SIHA96] Simon Haykin, *Adaptive Filter Theory*, Prentice Hall, 1996.

[SKGV00] Kulkarni, S. Gadre, V.M.; Bellary, S.V., 'Nonuniform M-band wavepackets for transient signal detection', *Signal Processing, IEEE Transactions on*, Volume: 48 Issue: 6, June 2000.

[SMAL89] S. Mallat, 'A theory of Multiresolution Signal Decomposition: The Wavelet Representation', *IEEE Trans. Pattern Analysis and Machine Inteligence*, Vol. 11, 1989.

[SMLL89] S. Mallat, 'Multi-resolution approximation and wavelet orthonormal bases of L^{2}', *Trans. of the American Math. Society*,1989.

[TLAF00] Liu, T.T.; Fraser-Smith, A.C., 'Detection of transients in 1/f noise with the undecimated discrete wavelet transform', *Signal Processing, IEEE Transactions on*, Volume: 48 Issue: 5, May 2000.

[TMWH99] Todd McWhorter, *Updating the EVD of Rank-One Modified Hermitian Matrices*, Unpublished Report, December 14, 1999.

[WLHW99] Wen-Liang Hwang, 'Estimation of fractional Brownian motion embedded in a noisy environment using nonorthogonal wavelets', *Signal Processing, IEEE Transactions on*, Volume: 47 Issue: 8, Aug. 1999.

[YMAY90] Y. Mayer, *Ondelettes et Operateurs,* Tome I: Ondelettes, Herman, Paris, 1990.