Detection of Low Order Nonstationary Gaussian Random Processes

A THESIS

Presented to

The Academic Faculty

By

Wayne Thomas Padgett

In Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy in Electrical Engineering

Georgia Institute of Technology

March, 1994

Detection of Low Order Nonstationary Gaussian Random Processes

Approved:

Douglas B. Williams, Chairman

James H. McClellan

Russell M. Mersereau

Date approved by Chairman _____

To Tomi

Acknowledgments

First of all, I thank Tomi for her love, continuous encouragement and support. Her efforts and assistance have been a great blessing and without her help I could not have persevered. I am also very thankful to both my parents for their tremendous confidence in me and for their great examples in life. Their love and discipline and continuing support is invaluable.

For his efforts, advice, contributions and support I thank Dr. Williams. He has helped me and educated me in a variety of ways, from financial support through his research to his generosity with his time and insight. I would also like to thank Dr. McClellan and Dr. Mersereau for reading and commenting on my proposal and thesis. Their thoughts and comments not only helped me improve my work, but also improved my ability to view my thesis in the broader context of the DSP literature. I can only hope that all my future professional friendships can be as pleasant and beneficial. I am pleased to have had Dr. DeWeerth, Dr. Heil and Dr. Ingram on various committees, their efforts in reviewing my work and their comments were very valuable.

During my graduate education, I have been fortunate enough to be supported by the Office of Naval Research, the National Science Foundation, and Georgia Tech Research Institute. I am appreciative for all of the support, and I am particularly grateful to Jeff Holder, Martha Willis and Jim Echard for their investment in me. I would also like to thank Bill Marshall and Keith Vaughn for their excellent support and friendliness.

I would like to thank the staff at Atlanta Signal Processors, Inc. for the support and experience I received while working there. Thanks to Craig Richardson, Larry Heck, Doug Reynolds, Tom Gardos and Kwan Truong for making ASPI a great place to work. Thanks also to Dr. Schafer, Dr. Barnwell and Dr. Mersereau for starting ASPI and for all your support at Georgia Tech. A special thanks to Maricia Kimbrell for taking over a job that would keep ten people busy.

Without a friendly, helpful and extremely competent system support staff, my research would have taken years longer. Thanks to Lonnie Harvel, Dave Webb, Sam Smith and Steven McGrath the network has run like a top — I am most grateful. No acknowlegement would be complete without a mention of the two best secretaries in the world: Kay Gilstrap and Stacy Schultz. I owe them both a million small favors and owe my sanity on stressful days to their humor and friendliness.

To the friends I have made here in the DSP group, thank you for being there, for great technical conversations both about my work and things I don't know anything about. I would have learned much less and grown much less at Georgia Tech if not for the contributions of the other students. I especially thank Kate Maloney, Dan Drake, Soonjoo Hwang, Steve Martucci, Ragnar Jonsson, Johnathan Su, and Stan Reeves for their friendship, support, and advice.

I would like to thank several of the Georgia Tech faculty and staff for their special encouragement and friendship. Thanks to Glenna Thomas, who has been a bright spot with her helpfulness and friendliness ever since I have been at Georgia Tech. I also appreciate the unselfishness of Jack Mangham and Paul Wright who were a great source of encouragement during some of my darkest days at Tech.

Finally, I would like to thank God for providing me with the opportunity, the ability, and the support of a wonderful family and group of friends to allow me to complete this work. I cannot boast over any of these things because they are all gifts from Him.

Contents

List of Figures

Conventions

scalar a

vector $\underline{a}$

matrix $\mathbf{A}$

transpose $\mathbf{A}^T$

Hermitian transpose $\mathbf{A}'$

conjugate a'

spectral domain $\widetilde{\mathbf{A}}$

Abbreviations

AR(q) Autoregressive with model order q

BLRT Block Likelihood Ratio Test (detector)

BLRTsd Block Likelihood Ratio Test with spectral diagonalization (detector)

CDF Cumulative Distribution Function

CFAR Constant False Alarm Rate

CGRV Complex Gaussian Random Vector (Variable if scalar)

EGG ElectroGastroGraph

FFT Fast Fourier Transform

FIR Finite Impulse Response

GBLRTsd Generalized Block Likelihood Ratio Test with spectral diagonalization (detector)

GLRT Generalized Likelihood Ratio Test (detector)

GRV Gaussian Random Vector (Variable if scalar)

HMM Hidden Markov Model

IID Independent Identically Distributed

K-L Karhunen - Loeve

LRT Likelihood Ratio Test (detector)

MA Moving Average

MLE Maximum Likelihood Estimate

MKM Modified Kay Method (detector)

PDF Probability Density Function

ROC Receiver Operating Characteristic

SNR Signal to Noise Ratio

STFT Short-Time Fourier Transform

SUF Stationary Unknown Frequency (detector)

WGN White Gaussian Noise

WVD Wigner-Ville Distribution

XWVD Cross Wigner-Ville Distribution

Summary

This thesis develops a new method for detecting low order Gaussian random signals in noise when the parameters of the signal are unknown and the frequency of the signal is nonstationary and slowly varying. A low order signal is defined here to to be a signal that can be modeled with a small number of autoregressive (AR) model coefficients and has only one significant spectral peak. The new detector is appropriate for applications where the signal is wideband and the time-bandwidth product is large such as passive sonar, radar and biomedical signal detection.

Based on the low order signal restriction, the signal is modeled as the output of a first order AR system with a time varying pole angle. This signal model is applied to the computation of the optimal likelihood ratio test when the signal parameters are known and extended to an approximation of the generalized likelihood ratio test (GLRT) when the signal parameters must be estimated. The resulting detector makes use of several types of computationally efficient approximation, including block processing and a spectral diagonalization of the data covariance matrix. The stack algorithm and the Viterbi algorithm are investigated as maximization techniques for producing the maximum likelihood estimates required by the GLRT.

The statistical behavior of the approximations is determined analytically for both a single block of the data using numerical characteristic function inversion, and for the full observation using a distribution approximation. Simulations are used to verify and extend the analytical results. Using comparisons with a near-optimum detector as an upper bound and the energy detector as a lower bound, the new detection algorithm is shown to give nearly optimal performance for a variety of nonstationary and non-AR signals. Recent methods from the literature are examined

for applicability to this problem, and the most suitable method is compared to the new detector.

The new detection algorithm is found to be computationally efficient with near optimal performance, and is robust to model inaccuracy. The thesis concludes with a discussion of extensions of the theoretical framework of the approximate GLRT to other signal models and applications.

CHAPTER 1

Introduction

The objective of this thesis is to determine a method for detecting a low order Gaussian random signal in noise when the parameters of the signal are unknown and the frequency of the signal is nonstationary. These signals are difficult to detect because the nonstationarity produces a large number of free signal parameters. For this investigation it will be assumed that the noise covariance is known and the signal frequency varies slowly. A low order signal is defined here to be a signal which can be modeled with a small number of autoregressive model coefficients and has only one significant spectral peak. The goal of this thesis is to develop a detector possessing several desirable characteristics: computational efficiency, near-optimal performance, and robustness to model inaccuracy.

There are a number of application areas that require detection of signals which exhibit nonstationary frequency. These include such divergent areas as

- biomedical signal analysis (EGG, etc.) [1],
- spread spectrum communication,
- wear monitoring and fault detection,
- active radar and sonar [2],
- and passive sonar [3].

Therefore, this thesis should be applicable to a number of current research problems. The specific problem stated in Section 1.1 involves detection of low SNR wideband signals with nonstationary frequency in a noise environment with stationary harmonic

components. Although this thesis is designed to address this problem thoroughly, the variations of the ideas presented below could be applied to detection of a variety of signals with a nonstationary parameter. Figure 1.1 shows an example of the type of signal and noise environment to which this work has been applied. The figure is a spectrogram of an AR(1) synthetic nonstationary signal in harmonic noise. The harmonic noise is actually measured sonar data.

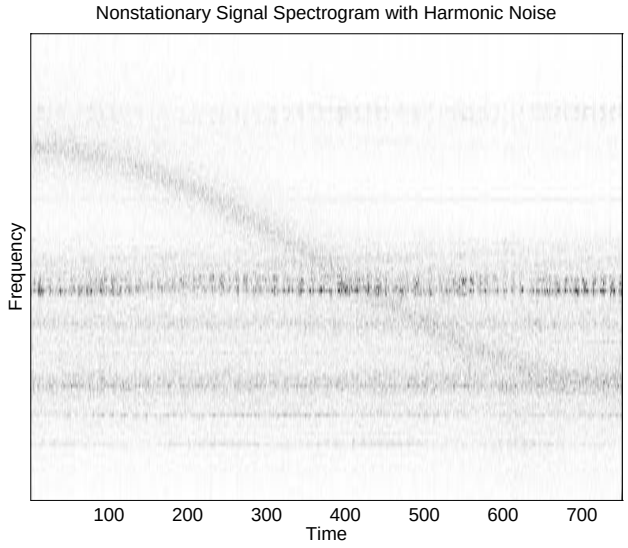

Figure 1.1: The signal is AR(1) with a pole radius of $\sqrt{(}0.9)$ and a sinusoidal frequency track.

This chapter deals with the main issues of the detection problem and gives a preview of the entire thesis. Section 1.1 describes the problem and gives a basic mathematical framework. Section 1.2 discusses the major issues that are dealt with in the thesis and how these issues are addressed. Sections 1.3 through 1.6 give some background material from the literature and Section 1.7 discusses the shortcomings of earlier approaches. Section 1.8 gives a chapter by chapter overview of the thesis.

1.1 Problem Statement

Consider the mathematical model that is used for the random signal detection problem. The signal and noise are zero mean independent discrete-time random processes $s(m)$ and $n(m)$, respectively. The received vector $\underline{x}$ is defined as

$$\underline{x} = [x(0), x(1), \ldots, x(M-1)]^T \tag{1.1}$$

where ^{T} represents the simple transpose. There are two hypotheses, H_0 and H_1, describing the possible structure of $\underline{x}$

$$\begin{aligned} H_0: &\quad x(m) = n(m) \\ H_1: &\quad x(m) = s(m) + n(m). \end{aligned} \tag{1.2}$$

Since both the signal and noise processes are zero mean, the covariance matrix is a complete description of the received vector's characteristics. Therefore, a sufficient detector statistic [4] can be written entirely in terms of the appropriate covariance matrices. The hypotheses may be restated in terms of the covariance matrices, so that

$$\begin{aligned} H_0: &\quad \mathbf{K}_x = \mathbf{K}_n = E[\underline{n}\underline{n}'] \\ H_1: &\quad \mathbf{K}_x = \mathbf{K}_s + \mathbf{K}_n = E[\underline{s}\underline{s}'] + E[\underline{n}\underline{n}'] \end{aligned} \tag{1.3}$$

where $'$ represents the Hermitian transpose. Since the signal, $s(m)$, is assumed to be nonstationary, $\mathbf{K}_s$ is generally *not* Toeplitz. Recall that $\mathbf{K}_s$ is unknown, while $\mathbf{K}_n$ is known and Toeplitz.

Because the spectral characteristics of $s(m)$ vary with time, it is intuitive to consider a time-frequency decomposition as part of the detector. Detector performance can be improved by using such a decomposition to form an estimate of $\mathbf{K}_s$. It is possible to implement the detector in the frequency domain since $\mathbf{K}_s$ has a simple relationship to $\widetilde{\mathbf{K}}_s$, the covariance matrix of the FFT of $s(m)$. In fact, several of the methods given in the literature are based on a time-frequency decomposition.

1.2 Approach

The approach of this thesis to the detection problem is to estimate $\mathbf{K}_s$ by using a signal model whose parameters are estimated from the data. The detector is based on the Generalized Likelihood Ratio Test (GLRT) which requires the Maximum Likelihood Estimate (MLE) of the model parameters. The details of the GLRT are discussed Section 1.3. A first order autoregressive, or AR(1), signal model is employed to restrict the number of parameters to be searched. The AR(1) model is seen to be robust, so that the detector can be applied to non-AR signals which resemble the low order Gaussian signal type described above. Frequently the MLEs are determined by searching over a predetermined set of allowed model parameter values. The speed of the search is increased by using large parameter step-sizes and efficient search methods. This approach is only possible if the statistical behavior of the likelihood ratio is well understood. Since the frequency parameter is nonstationary, detection is carried out in the time-frequency domain where the frequency path of the signal is easy to determine and to visualize.

1.3 Background

One of the main goals of this research is to develop a detector with near-optimal performance. Since optimality of the detector is important, consider the likelihood ratio test (LRT), which is optimal in the sense that the probability of error is minimized. The log-likelihood function (with constants removed) of a zero mean stationary Gaussian random vector $\underline{x}$ is given by

$$l_i(\underline{x}) = \ln \mathcal{P}_i - \frac{1}{2}\ln|\mathbf{K}_i| - \frac{\underline{x}'\mathbf{K}_i^{-1}\underline{x}}{2} \tag{1.4}$$

where the $\mathcal{P}_i$ is the probability of H_i occurring, and $\mathbf{K}_i$ is the value of $\mathbf{K}_x$ when H_i is true. The likelihood function is defined as the probability density function (PDF) of the random vector evaluated at the value of the observed data, so that a Gaussian log-likelihood function is simply the log of the Gaussian PDF. The LRT is then given by

$$l_1(\underline{x}) - l_0(\underline{x}) \underset{H_0}{\overset{H_1}{\gtrless}} 0 \tag{1.5}$$

which is represented as the difference of two logs instead of the "ratio" which gives the test its name.

The GLRT is known to perform extremely well in a number of applications, but its optimality properties are not simple to prove. The GLRT can be shown to be the detector that minimizes the probability of error when the unknown parameters are chosen to maximize the probability of error—the "minimax" solution. It is therefore the test that minimizes

$$\max_{\underline{u}} Pr(\text{error}|\underline{u}) \tag{1.6}$$

where $\underline{u}$ is the vector of unknown parameters [5]. When the unknown parameters are first chosen to maximize the probability of error given the observation, then the test which minimizes the resulting error is the GLRT. This is simply one criteria under which the GLRT is optimal, for a detailed explanation of the optimality of the GLRT

see [6]. In the general case, the GLRT is given by

$$\max_{\underline{u}}[l_1(\underline{x})] - \max_{\underline{u}}[l_0(\underline{x})] \underset{H_0}{\overset{H_1}{\gtrless}} 0. \tag{1.7}$$

Specifically, if the received signal $\underline{x}$ has unknown covariance $\mathbf{K}_1$ under H_1, and known covariance $\mathbf{K}_0$ under H_0, the GLRT can be written as

$$\max_{\mathbf{K}_1} l_1(\underline{x}|\mathbf{K}_1) - l_0(\underline{x}) \underset{H_0}{\overset{H_1}{\gtrless}} 0 \tag{1.8}$$

so that $\mathbf{K}_1$ is chosen to maximize $l_1(\underline{x})$.

Although covariance matrices have a few special properties such as Hermitian symmetry and Toeplitz structure for stationary processes, it is clear that $\mathbf{K}_1$ contains many parameters, and an exhaustive search for the best choice is out of the question. Since the Toeplitz property does not apply to nonstationary processes, the search is even less tractable when the signal is not stationary.

One way to reduce the number of parameters, and therefore the computational complexity of the search, is to use a signal model. One of the simplest models that can be applied to a nonstationary signal is the AR(1) model, which inherently requires that the signal have a single spectral peak and that the signal be a complex quantity. This model choice conveniently avoids the difficulties associated with conjugate pairs for the poles of a real AR(2) model, where the signal power and bandwidth vary with the frequency of the poles. Real observations can be converted to complex data to match the AR(1) model by using the Hilbert transform, or by converting a bandpass frequency range of the data to a complex baseband signal. To model a nonstationary frequency, the pole location of the AR(1) model can be allowed to vary with time. In such cases, only a single parameter, the pole angle, must be varied to change the signal frequency.

If $\underline{s}$ is a complex AR(1) process, it is equivalent to the output of a single pole AR filter whose input is unit variance complex white Gaussian noise. The autocorrelation function $R(k)$ of the system output can be written

$$R(k) = \frac{G^2}{1-|a|^2} r^{|k|} e^{jk\theta} \tag{1.9}$$

where G is the gain of the AR filter, a is the pole of the filter, and r and θ are the radius and angle, respectively, of that pole so that $a = re^{j\theta}$. Since the (i,j)th element of the Toeplitz covariance matrix $\mathbf{K}_s$ is just $R(j-i)$, the covariance matrix $\mathbf{K}_1$ can be written entirely in terms of of the signal parameters, G, r, θ, and the known noise covariance, $\mathbf{K}_n$:

$$\mathbf{K}_1 = \mathbf{K}_s(G, r, \theta) + \mathbf{K}_n. \tag{1.10}$$

Therefore, when the signal is AR(1), the LRT in (1.5) can be written in terms of the signal model parameters and the noise covariance.

Suppose a signal is modeled as an AR(1) signal with unknown parameter values. The GLRT requires the maximization of the likelihood of $\underline{x}$ over the parameters G, r and θ for H_1:

$$\max_{G,r,\theta} l_1(\underline{x}|G, r, \theta). \tag{1.11}$$

Since the noise covariance is known, there are no unknown parameters for H_0, and the GLRT is computed as

$$\max_{G,r,\theta} l_1(\underline{x}|G, r, \theta) - l_0(\underline{x}) \underset{H_0}{\overset{H_1}{\gtrless}} 0. \tag{1.12}$$

The parameter search is over only three variables. However, if the frequency is allowed to be nonstationary, there may be a different frequency parameter (pole angle) for each time sample in the observation $\underline{x}$, and θ becomes $\theta(m)$. The addition of the nonstationary frequency parameter increases the size of the maximization search to the point of intractability. As a result, other methods have come under consideration in the literature. The first method to be discussed has been proposed purely for stationary signals, but can be applied to this more general case. The reviewed methods do not assume the same model given above, but are directed at similar signal types, i.e., those with a single spectral peak.

1.4 Kay's AR Modeling Method

A detector for broadband signals with unknown parameters has been proposed by S. Kay [7]. He proposed replacing the GLRT's search for MLE parameters with an AR model parameter estimation technique such as the linear prediction method. Although Kay chooses a detection problem slightly different than the one presented in Section 1.3, most of his approach can be adapted to the context of the current problem.

Kay's hypotheses are not equivalent to those given in (1.3) since he assumes white Gaussian noise (WGN) with unknown power, and models both the signal and the noise under H_1 with the AR model rather than treating the noise separately. This method gives the AR(q) model for $x(m)$ under H_1

$$x(m) = -\sum_{i=1}^{q} a_i x(m-i) + n_s(m) \tag{1.13}$$

where the AR model order is q and $n_s(m)$ is a WGN process independent of the previously mentioned noise process $n(m)$. Kay's hypotheses for an AR(1) model can be stated as

$$\begin{aligned} \mathbf{K}_0 &= \quad \sigma\mathbf{I} \\ \mathbf{K}_1 &= \quad \mathbf{K}_s(r,\theta) \end{aligned} \tag{1.14}$$

where the G parameter is unnecessary due to the use of σ.

In this form Kay's detector requires the estimation of parameters σ, r, and θ for the AR(1) case and in the case of higher order models, all the autoregressive coefficients must be estimated. If the estimates are the maximum likelihood estimates, the detector is the GLRT just as in (1.12) but with the addition of the unknown noise parameter. However, Kay uses a linear prediction method to obtain the $\widehat{a}_i$'s. He also suggests the following approximation to the likelihood ratio which eliminates the σ and depends only on the observed data and the $\widehat{a}_i$'s

$$\Lambda_{\text{Kay}} = M \ln\left[\frac{1}{M}\sum_{m=q}^{M-1} x^2(m)\right] - M \ln \frac{1}{M}\sum_{m=q}^{M-1}\left[\sum_{t=0}^{q} \widehat{a}_t x(m-t)\right]^2. \tag{1.15}$$

This method has several of the desired characteristics for a detector operating on a stationary signal. It uses the AR model with unknown parameters and uses a computationally efficient method, not an exhaustive search, for parameter estimation. Kay's analysis shows that the performance of this detector is up to 9dB above that of the energy detector, with particularly good performance for low order signals.

However, the classical linear prediction method is not well suited to problems where the noise contains harmonics, and only approaches optimal performance when the observation vector is long. Since the linear prediction estimator assumes the signal is stationary, it can only be applied sequentially to short segments of a nonstationary signal. Therefore, Kay's method would require a tradeoff in block length, with better performance for longer blocks at the expense of limiting the model parameter's rate of change. Kay's work suggests the detector only for the stationary signal application and does not deal with the nonstationary frequency problem. The division of the received vector $\underline{x}$ into subsections also introduces a loss of performance since samples in different blocks may be correlated, but are treated as independent when processed separately. This issue will be dealt with in more detail later.

1.5 Hidden Markov Model (HMM) Method

The hidden Markov model (HMM) method is one of two methods to be mentioned that deal with nonstationary signals. When the signal is allowed to be nonstationary, the GLRT given in (1.12) must be modified to allow a sequence of values for θ, the frequency parameter. The other parameters will be assumed stationary. The sequence of frequency parameters can be associated with the sample index m, or with some larger block index p, if the frequency varies slowly enough to be considered stationary over the length of each block. In either case, the sequence of all values of θ can be represented as a vector $\underline{\theta}$. This thesis will use the block definition of $\underline{\theta}$ almost

exclusively. With a nonstationary frequency, (1.12) can be rewritten as

$$\max_{G,r,\underline{\theta}} l_1(\underline{x}|G,r,\underline{\theta}) - l_0(\underline{x}) \underset{H_0}{\overset{H_1}{\gtrless}} 0 \tag{1.16}$$

where the additional number of parameters in $\underline{\theta}$ require substantially more computation for maximization.

The nonoverlapping STFT can be used to divide the signal into a discrete two dimensional time-frequency space where the FFT length defines the size of the time blocks and the frequency blocks. Using the STFT structure, the problem of finding a maximum likelihood estimate of the signal's frequency path can be approached using an HMM. This problem is equivalent to the maximization of (1.16) over $\underline{\theta}$.

The HMM model is easily explained by considering the signal's presence in each FFT bin to be the possible states, although it is not restricted to this formulation. The likelihood of the signal being present in an FFT bin is determined, and the transition probability to the next state may either be derived from training, or assumed to be Gaussian with the mean centered on the most recent state. In the Gaussian transition case, the variance of the Gaussian must still be assumed or found by training.

Given the HMM established above and a measured signal, the maximum likelihood state sequence can be determined by the Viterbi algorithm. The Viterbi algorithm keeps track of the maximum likelihood path to each state from the last time sample. At the end of the sequence, the path can be reconstructed in reverse to give the maximum likelihood path over the whole sequence. Because the likelihood of transition from each past state to each next state must be calculated for every time block, the Viterbi algorithm is an $O(V^2P)$ computation, where V is the number of states (corresponding to frequency resolution), and P is the number of time blocks (corresponding to the number of nonoverlapping FFTs required). The Viterbi algorithm is an efficient method but it has two significant disadvantages. The first is that both computation and storage increase as the square of the desired frequency resolution, limiting the range of resolution available. The second is that the computation is the same for either low or high SNR cases. Even if the problem is easier, no gain

in speed is realized.

The algorithm proposed in [8] uses an HMM structure to track nonstationary sinusoidal signals by using the Viterbi algorithm to find the path with the maximum score. This work proposes an ad-hoc approximation to the likelihood function based simply on the FFT magnitude in each bin. Streit and Barrett [8] suggest that the method could be made more effective by using the FFT phase information, but there is no mention of other signal models.

In a more recent article, Barrett and Holdsworth describe an extension of the method given in [8] that takes advantage of both the amplitude and phase information to calculate the likelihood of a sinusoidal signal being present in a particular frequency bin [9]. The new method is considered not only as a frequency tracker, but also as a detector. Although the method of [9] improves detection performance by about 4dB, the method is once again applied only to sinusoidal signals and therefore cannot be directly compared to the work of this thesis.

1.6 Iterative Wigner Method

The Wigner-Ville distribution (WVD) is a popular alternative to the STFT as a time-frequency decomposition. Although it has problems with cross-terms and is somewhat more difficult to compute, the WVD has a higher frequency resolution than the STFT. Since the estimate of the signal frequency is critical to the computation of (1.16), the WVD has been considered as a method for obtaining an estimate for $\underline{\theta}$. Once the estimate has been obtained, detection can be performed by building a covariance matrix consistent with the AR model parameters and computing (1.16).

The Wigner-based method to be considered has been proposed by O'Shea and Boashash. This method uses the cross Wigner-Ville distribution (XWVD) to iteratively improve the estimate of the nonstationary signal's instantaneous frequency. The instantaneous frequency of the signal is estimated to be the largest peak in the

XWVD. Each time an estimate is obtained, a signal corresponding to the frequency estimate is constructed and the XWVD is computed between the original signal and the constructed one. Given an instantaneous frequency estimate $f_i(t)$, the corresponding signal estimate is

$$\widehat{s}(t) = \text{Real}\left[\exp\left(j2\pi \int_{-T/2}^{t} f_i(\alpha)d\alpha\right)\right]. \tag{1.17}$$

The improved estimate is used to isolate the signal and remove noise for improved energy detection [10]. This use is not equivalent to the potential use of the estimate in a GLRT-based detector, which would be more comparable to the work of this thesis. Figure 1.2 [10] shows how a signal may be enhanced by removing noise outside the signal region.

This iterative procedure was shown to produce excellent estimates and improved detector performance [10], but it has significant shortcomings. The iterative procedure requires an initial estimate, which must be generated by some other method. The initial estimate affects the performance of the algorithm and can be most difficult to determine in low SNR cases. Colored noise or multiple signals cause significant alternative peaks as well as cross-term peaks in the XWVD, which rapidly degrade the detector's performance. Therefore, the iterative XWVD method is difficult to modify for use in multi-signal or harmonic noise environments. The usefulness of this method is diminished when the signal is wideband, since the peak in frequency becomes more difficult to distinguish and is poorly modeled by the narrowband signal estimate in (1.17).

1.7 Motivation

Each of the methods discussed above relate to the problem investigated in this thesis, but each fails to address the entire problem. Kay's AR modeling method deals with the unknown frequency problem, but not with the nonstationary frequency or colored noise elements of the problem. The HMM method uses an approach to the frequency

Iterative Wigner-Ville Method

(a) (b)

(c) (d)

Figure 1.2: Iterative Wigner-Ville method: (a) is the original signal; (b) is the signal with noise; (c) is the instantaneous frequency estimate; (d) shows the improvement over the energy detector in probability of detection vs. SNR [10]

estimation problem which is very similar to the approach used in this thesis, but is only applied to nonstationary sinusoids, not to random signals. The iterative Wigner method represents an attempt to solve the nonstationary frequency estimation problem for an unknown signal, but does not approach the subject of using that estimate for detection purposes. Also, the iterative Wigner method uses a narrowband signal model which would likely break down if applied to wideband signals.

None of the methods in the literature deal with the full problem that this thesis investigates. In addition, some of these methods are based on ad-hoc approaches to the detection problem. The development of a detection method which is clearly based on the optimal theory of detection using the likelihood ratio test is more desirable.

This thesis is presented to deal with the nonstationary random signal detection problem directly based on a theoretically sound method, the GLRT. Although this thesis deals with a subset of the problem, i.e. low order signals, it is intended to provide not only a practical solution, but also a theoretical framework based on the GLRT for the extension of this work so that more general problems can be solved in the future.

1.8 Overview of Thesis

Chapter 2 goes into detail concerning the implementation of the full observation detector. Chapter 3 discusses the effectiveness of the frequency domain approximation when applied as a stationary signal quadratic detector. Chapter 4 examines the performance of the full observation detector for a variety of signal types and compares it to various other detection schemes. Chapter 5 presents the conclusions of this thesis and discusses some possible future extensions. The appendix gives a derivation of the covariance matrix of a nonstationary signal based on the output of a time-varying filter.

CHAPTER 2

Detection of Nonstationary Random Signals

The detection algorithm developed for this thesis is based on the the GLRT as given in (1.12), which assumes a complex AR(1) signal model. However, in order to clearly discuss some of the issues involved in computing (1.12) it is convenient to first deal with the computation of the LRT. Temporarily ignoring the multiparameter maximization will simplify the notation and discussion of the approximations to be implemented for practical computational reasons.

2.1 Likelihood Ratio Test (LRT)

Consider the LRT for a full observation represented in terms of its sufficient statistic. The LRT for the known signal covariance and known noise covariance case is given by

$$\underline{x}'[\mathbf{K}_0^{-1} - \mathbf{K}_1^{-1}]\underline{x} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \tag{2.1}$$

where the subscript of $\mathbf{K}$ refers to the hypothesis, and the γ term includes constants due to the energy and probability of each hypothesis. Since $\underline{x}$ is a complex Gaussian random vector (CGRV), the detector statistic given in (2.1) is a general quadratic form random variable and its distribution is useful for performance analysis. Although there is much more to be said in Section 2.5.1 about finding the distribution of a general quadratic form such as the detector statistic, the computation of the statistic itself is the focus of this discussion.

When the observation size M is large, that is, when the number of samples in $\underline{x}$

is more than a few hundred, complexity problems arise in the computation of (2.1). The storage required for the matrix $[\mathbf{K}_0^{-1} - \mathbf{K}_1^{-1}]$ is M^2 elements and the computation of the statistic is $O(M^2)$. The matrix in (2.1) is generally not Toeplitz, and the matrix inverse required to compute it is $O(M^3)$. The cost of computing the test matrix becomes more important when $\mathbf{K}_1$ is not known in advance. When $\underline{x}$ contains thousands of samples, the direct computation of (2.1) is impractical for almost any sort of real application. A more practical approach is to divide the observation into small segments or blocks, process them independently, and then combine the results.

2.2 Block Likelihood Ratio Test (BLRT)

Consider an alternative mathematical model of the problem: let the signal and noise be the zero mean, independent, discrete-time random processes $s(m)$ and $n(m)$ as before. Then let $\underline{x}(p)$ be defined as a subset or block of $\underline{x}$,

$$\underline{x}(p) = [x(pL), x(pL+1), \ldots, x(pL+(L-1))]^T \quad \text{for} \quad 0 \leq p < P-1 \tag{2.2}$$

where L is the number of samples per block, and P is the number of blocks in the observation.

Recall that $\mathbf{K}_i = E[\underline{x}\underline{x}']$ is the $M \times M$ covariance matrix of the entire observation, and let $\mathbf{K}_i(p,q) = E[\underline{x}(p)\underline{x}'(q)]$, the $L \times L$ cross covariance matrix of two blocks of $\underline{x}$. Then $\mathbf{K}_i$ may be written

$$\mathbf{K}_i = \begin{bmatrix} \boxed{\mathbf{K}_i(0,0)} & \mathbf{K}_i(0,1) & \cdots & \mathbf{K}_i(0,P-1) \\ \mathbf{K}_i(1,0) & \boxed{\mathbf{K}_i(1,1)} & \cdots & \mathbf{K}_i(1,P-1) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{K}_i(P-1,0) & \mathbf{K}_i(P-1,1) & \cdots & \boxed{\mathbf{K}_i(P-1,P-1)} \end{bmatrix} \tag{2.3}$$

where P is as defined above. Instead of calculating the optimum detection statistic Λ_{LRT} for the full observation

$$\Lambda_{LRT} = \underline{x}'[\mathbf{K}_0^{-1} - \mathbf{K}_1^{-1}]\underline{x}, \tag{2.4}$$

the block processing method calculates the block LRT (BLRT) statistic

$$\Lambda_{BLRT} = \sum_{p=0}^{P-1} \underline{x}'(p)[\mathbf{K}_0^{-1}(p,p) - \mathbf{K}_1^{-1}(p,p)]\underline{x}(p) \tag{2.5}$$

so that $\Lambda_{LRT} = \Lambda_{BLRT}$ only if (2.3) is block diagonal. When the signal is stationary, this method is simply a form of time-averaging, however the method retains its relationship to the optimal detector even when the signal (and therefore $\mathbf{K}_i(p,p)$) is nonstationary.

Therefore, performing block processing is optimal only if the blocks are independent. Although the BLRT is sub-optimal, it is a practical method for implementing a near-optimal detector for large observations. It implicitly makes the approximation that the blocks are statistically independent, i.e. that $\mathbf{K}_i(p,q) = 0$ when $p \neq q$ and $\mathbf{K}_i$ is approximated by

$$\mathbf{K}_i \approx \begin{bmatrix} \boxed{\mathbf{K}_i(0,0)} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boxed{\mathbf{K}_i(1,1)} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boxed{\mathbf{K}_i(P-1,P-1)} \end{bmatrix}. \tag{2.6}$$

This approximation and another major approximation will be used in the final detection algorithm to make the problem more tractable. The effects of both of these approximations will be evaluated later.

It is possible to show that the block processing method is a more accurate approximation for wideband signals than for narrowband signals. Covariance matrices of stationary processes are Toeplitz, and the only Toeplitz matrix which is block diagonal is the trivial case, a diagonal matrix. A process with a diagonal covariance matrix is white, and a white signal does not meet the original low order process assumption. Therefore, the block processing procedure will not be optimum for the signals under consideration. If the signal is wideband, however, its covariance matrix will resemble a diagonal matrix, since there will be little energy in the off-diagonal elements. Therefore, the block processing method is most appropriate for wideband signals.

2.3 BLRT with Spectral Diagonalization

After the application of the block processing method, the detector's complexity is related to block length L rather than the observation length M, and clearly L will be chosen much smaller than M. However, L cannot be made too small since this would degrade the accuracy of the block independence approximation. The detector still depends on the square of L in terms of computation.

A particular method of detector simplification which can reduce the computational complexity has been suggested by Johnson and Dudgeon [5]. This method involves taking the FFT of $\underline{x}$ and then using only the diagonal values of the resulting covariance matrix. To convert (2.5) to the spectral domain, let

$$\underline{\tilde{x}}(p) = \mathbf{W}\underline{x}(p) \tag{2.7}$$

and let

$$\widetilde{\mathbf{K}}_i(p,p) = E[\underline{\tilde{x}}(p)\underline{\tilde{x}}'(p)] = \mathbf{W}E[\underline{x}(p)\underline{x}'(p)]\mathbf{W}' = \mathbf{W}\mathbf{K}_i(p,p)\mathbf{W}' \tag{2.8}$$

where $\mathbf{W}$ is the FFT matrix and $\tilde{}$ indicates a spectral domain quantity. Then (2.5) becomes

$$\Lambda_{BLRT} = \sum_{p=0}^{P-1} \underline{\tilde{x}}'(p)[\widetilde{\mathbf{K}}_0^{-1}(p,p) - \widetilde{\mathbf{K}}_1^{-1}(p,p)]\underline{\tilde{x}}(p) \tag{2.9}$$

which has the same complexity as before, not including the FFT. In fact, the expression of (2.9) gives exactly the same result as (2.5). However, because the FFT transformation tends to diagonalize $\widetilde{\mathbf{K}}_i$, it is possible to approximate (2.9) with the BLRT with spectral diagonalization (BLRTsd), given by

$$\Lambda_{BLRTsd} = \sum_{p=0}^{P-1} \underline{\tilde{x}}'(p)[\mathcal{D}(\widetilde{\mathbf{K}}_0(p,p))^{-1} - \mathcal{D}(\widetilde{\mathbf{K}}_1(p,p))^{-1}]\underline{\tilde{x}}(p) \tag{2.10}$$

where $\mathcal{D}()$ is defined as the operator which sets the off-diagonal elements to zero. Since the inverse is trivial to compute for a diagonal matrix, the diagonalization is implemented before the matrix inverse. Now (2.10) can be evaluated with two dot products instead of a vector-matrix-vector multiply for a significant computational

savings. On the other hand, (2.10) is no longer equivalent to (2.5). Therefore, the accuracy of yet another approximation must be evaluated.

2.4 GBLRT with Spectral Diagonalization

Now that the development of the BLRTsd is complete, it is possible to discuss its generalization to the case of unknown parameters. Just as the LRT can be generalized by using maximum likelihood estimates of unknown parameters to become the GLRT, the BLRTsd can also be generalized to become the GBLRTsd. If the complex AR(1) model described earlier is applied, there are only three possible unknown parameters, G, r and θ. It would be straightforward to implement the GBLRTsd using a multidimensional optimization technique to find the parameter values which would maximize $\Lambda_{GBLRTsd}$ under H_1. Since the noise covariance is assumed known, no search would be necessary under H_0. However, since the parameters are unknown, $\Lambda_{GBLRTsd}$ would have to be recomputed for each possible combination of parameters that were tested. Clearly, the parameter search would multiply the computational cost of detection many times over what would be required for the known parameter case. The detection statistic can be stated as

$$\Lambda_{GBLRTsd} = \max_{G,r,\theta} \sum_{p=0}^{P-1} \underline{\tilde{x}}'(p)[\mathcal{D}(\widetilde{\mathbf{K}}_0(p,p))^{-1} - \mathcal{D}(\widetilde{\mathbf{K}}_1(p,p;G,r,\theta))^{-1}]\underline{\tilde{x}}(p) \qquad (2.11)$$

where the notation $\widetilde{\mathbf{K}}_1(p,p;G,r,\theta)$ simply denotes the dependence on the AR parameters.

2.4.1 Nonstationary Frequency

Up to this point in the discussion, the detector has only been general enough to handle stationary signals. The AR parameters have been fixed over the length of the observation. However, the stated goal of this thesis is to develop a detector for random signals with nonstationary frequency. Therefore, suppose the frequency

parameter θ is allowed to vary with the observation sample index, as in $\theta(m)$. This once again introduces a very large number of free parameters into the maximization problem and renders the problem intractable. Due to the original motivation for this work, the parameters r and G are presumed to be stationary. Although the theoretical development does not require this restriction, computational resources make it necessary.

Since the signal was originally assumed to be slowly varying in frequency, it is possible to make the assumption that the signal is stationary over a single block. In this case, θ is only allowed to vary with the block index p. Since there are many fewer blocks in the observation than samples, this assumption produces a large gain in computational efficiency, albeit at the expense of a less general set of signal applications. Since the signal is known to be slowly varying in frequency, a further restriction can be made on the value of $\theta(p)$. Each successive value of $\theta(p+1)$ would normally be expected to be close to the value of $\theta(p)$, and this fact can be translated into a range restriction on the value of $\theta(p+1)$. This range restriction can be used to further reduce the requirements of the likelihood maximization search. The nonstationary signal detection statistic can be stated as

$$\Lambda_{GBLRTsd} = \max_{G,r,\underline{\theta}} \sum_{p=0}^{P-1} \underline{\tilde{x}}'(p)[\mathcal{D}(\widetilde{\mathbf{K}}_0(p,p))^{-1} - \mathcal{D}(\widetilde{\mathbf{K}}_1(p,p;G,r,\theta(p)))^{-1}]\underline{\tilde{x}}(p) \qquad (2.12)$$

where $\underline{\theta}$ denotes the vector of all allowable $\theta(p)$ values. With the range restriction on successive $\theta(p)$ values applied, $\underline{\theta}$ is a member of a much smaller set of possible vectors.

2.4.2 Parameter Quantization

The algorithm used to accomplish the maximization of (2.12) is a major part of the detection process. There are a number of available algorithms from the literature that could be applied to the continuous maximization problem, but most deal poorly with problems which exhibit many local maxima. Experience has shown that like-

lihood maximization problems can involve extremely complex surfaces, often with large numbers of local maxima. However, the continuous optimization methods actually offer more precision than is required for this problem. In addition, constraints such as the range restriction on $\theta(p+1)$ can slow a continuous optimization down rather than speed it up. To avoid some of these problems, this work has been focused on discretizing the problem by quantizing the AR parameter values. The quantized values can be represented as $\{G_1, \cdots, G_T\}, \{r_1, \cdots, r_U\}, \{\theta_1, \cdots, \theta_V\}$ where T, U, and V are the number of values to be used for each parameter, respectively. The triple of index values (t, u, v) completely specifies an AR model, and since v is allowed to vary with the block index, $(t, u, v(p))$ refers to the AR model for block p.

If only certain values of the AR parameters are evaluated, a search over many possible values becomes more tractable. In addition, the spacing between allowed parameter values can be chosen in such a way as to avoid redundancy without significant loss in performance. The details of the analysis required to make these choices are given later. Another advantage of quantizing the parameter values is that the search for the best parameters can be organized as a tree search. Each level of the tree is a time segment of the observation, and each branch represents one possible combination of parameters for that time segment. Searching a tree structure to maximize likelihood is very similar to the decoding problem for convolutional codes. There is a large body of literature for the decoding problem which includes such efficient methods as the Viterbi algorithm and the stack algorithm, which is an approximation to the Viterbi algorithm. The HMM method of [8] takes advantage of this similarity to apply the Viterbi algorithm to estimation and detection, but only for nonstationary sinusoids.

2.4.3 Maximum Likelihood: the Stack Algorithm

The stack algorithm is an iterative procedure which has been applied to the problem of decoding bitstreams for error correction purposes [11]. In coding problems, the

goal is to choose the best symbol for each successive received word. In the GBLRTsd, the stack algorithm can be used to choose the best set of model parameters to apply to each time segment of received data. Both problems require maximization of the likelihood function and use a restricted set of choices at each block based on the previous choices.

The stack algorithm as implemented in this thesis begins with a list of starting values for the first block. The list contains a number of possible AR parameter values each specified as a triple of quantized parameter indices, $(t, u, v(0))$, representing the quantized model parameters $(G_t, r_u, \theta_{v(0)})$, as discussed in the previous section. To obtain reasonable performance, the initial list need not include all possible starting values of each parameter, only a representative sampling. Each parameter triple is evaluated with respect to the received signal block $\underline{x}(0)$ in terms of the approximate likelihood.

The approximate likelihood is equivalent to a single term in the summation of (2.12). This term becomes part of a cumulative score assigned to each parameter set in the list. The list of parameters is then ordered according to the scores forming the "stack" which gives the algorithm its name. The stack is limited to a length of 100 parameter sets, or paths, and is iteratively updated by extending the topmost (by score) path to the next time block as illustrated in Figure 2.1. After the stack has been initialized for time segment 0, the topmost path can be extended to time segment 1. The new longer paths are represented by the parameter set $(t, u, v(0), v(1))$.

Since the frequency is slowly varying, $v(1)$ must fall within a certain range of $v(0)$, which is the range restriction discussed earlier. For this implementation, the range restriction chosen is that only the two nearest frequency values or the same frequency value will be allowed for successive blocks. Therefore $v(1)$ must be a member of $\{v(0)+1, v(0), v(0)-1\}$, and $1 < v(1) < V$, where V is the total number of quantized frequency values.

Each time a path is extended, three new paths are created, and the old topmost

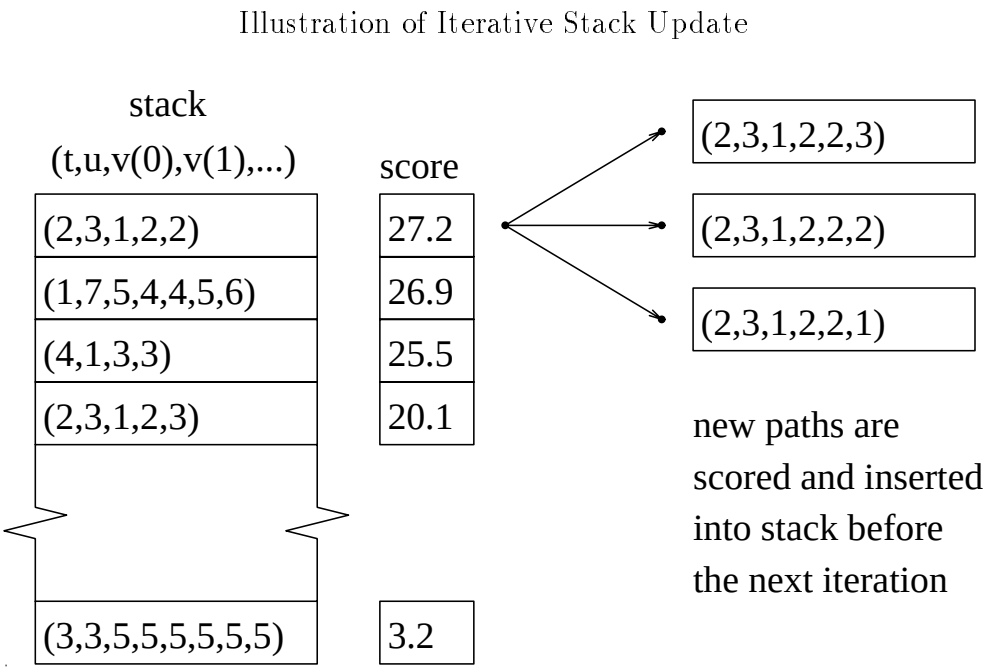

Figure 2.1: The stack update procedure is performed repeatedly until the top path reaches the end of the observation.

path is deleted. This process extends the stack by two, and if necessary, the two lowest paths are discarded. After the topmost path is extended, the three new paths are placed in the stack according to their score, and the process is repeated until the topmost path is extended to the last block. The score of each path is based on the sum of the terms represented in (2.12) minus a length penalty so that paths of unequal length can be compared. The length penalty is computed as the number of blocks in the path multiplied by an adjustable penalty value.

Two parameters of this algorithm have a strong effect on its performance and computational efficiency: the stack length limit and the penalty value. The stack length limit dictates the storage requirements of the algorithm and determines how far backward (to lower scoring paths) the algorithm can go when poor results are found. The stack length limit has been chosen somewhat arbitrarily, based on experience and computational resources. The penalty value has an even stronger effect on performance, since it determines how far a path can be extended before shorter paths must be considered. The penalty value for this implementation is determined using an average of a random sample of single block scores for the observation. Therefore, a perfectly average path will continue to be extended, but whenever the score drops below average, shorter paths should exist with higher scores.

The stack algorithm provides an effective method for searching for the set of parameters which will maximize $\Lambda_{GBLRTsd}$. It has a number of appealing characteristics including adjustable storage and computational requirements. By adjusting the stack length limit and the penalty value, the method can make best use of the computational resources available. When the signal to noise ratio (SNR) of the received data is high, the stack method is likely to finish quickly, requiring less time. For low SNR cases, the stack method will tend to require more time to finish, since more paths will have to be evaluated. Although the stack method does not give optimal results, it does an effective job of rejecting paths which are unlikely to give good results, and generally produces a result near the maximum.

2.4.4 Maximum Likelihood: the Viterbi Algorithm

Unlike the stack algorithm discussed above, the Viterbi algorithm guarantees an exact maximum likelihood solution to the tree search. In fact, the stack algorithm is just an approximation to the Viterbi algorithm. They may both be applied to similar problems, but they have different computational characteristics and different performance.

The Viterbi algorithm computes the likelihood of a particular parameter sequence based on two likelihoods, the likelihood of the observed data $\underline{x}(p)$ being generated by the given parameters, and by the likelihood of the transition from $\theta_{v(p-1)}$ to $\theta_{v(p)}$. This structure makes the frequency estimation procedure equivalent to a HMM state sequence search, since the likelihood of transition to a state depends only on the previous state. Since the likelihood of a particular block of data being produced by a given set of model parameters can be calculated easily, only the frequency to frequency transition values need to be defined.

Since nothing has been assumed about the nonstationary frequency except that it is slowly varying, it is impossible to define an optimal set of state transition likelihoods. However, a reasonable set of likelihoods can be assumed with good results. First of all, the likelihood of starting at any particular frequency is assumed to be equal to any other frequency, since the initial frequency is entirely unknown. Next, a set of frequency transition likelihoods is defined according to the distance of the frequency in block $p-1$ from the frequency in block p. Since the index v represents the allowed frequency values, the transition likelihood L_t is just a function of $d = |v(p) - v(p-1)|$. Once the function $L_t(d)$ is defined, implementation of the Viterbi algorithm is straightforward.

The goal is to find a sequence of model parameters which maximize the function

$$L_V = \prod_{p=0}^{P-1} L(\underline{x}(p); G_t, r_u, \theta_{v(p)}) L_t(|v(p) - v(p-1)|) \tag{2.13}$$

which can be written in terms of log likelihood as

$$l_V = \sum_{p=0}^{P-1} l(\underline{x}(p); G_t, r_u, \theta_{v(p)}) + l_t(d) \tag{2.14}$$

where $l_t(d) = \ln L_t(d)$. The maximization is based on the principle that the path with the largest likelihood leading to a particular state must be extended if the final path passes through that state. This principle is applied successively as p increases to $P-1$ so that at the end of the process, the maximum likelihood value leading to each state is known. The procedure can be most easily explained using MATLAB pseudocode to represent the steps which must be carried out.

Suppose that the values of G and r are fixed, and the maximum of l_V must be calculated for all possible sets of $v(p)$ with $0 < p < P-1$. Let the matrix `S(p,v)` be defined to contain intermediate scores associated with the corresponding p and v. Let `l_V(v,p)` $= l(\underline{x}(p); G, r, \theta_v)$ and assume `S(0,v)` has been initialized as `S(0,v) = l_V(0,v)`. Finally, let `l_t(d)` $= l_t(d)$ The Viterbi algorithm can be written as

```
for p = 1:(P-1)
    for v = 0:(V-1)
        S(v,p) = -Inf
        for q = 0:(V-1)
            S(v,p) = max(S(v,p),l_V(v,p) + l_t(abs(v-q)))
            % This loop compares the initial -Inf with every
            % possible replacement and keeps the maximum value
        end % for q
    end % for v
end % for p .
```

This procedure simply keeps a running total of the likelihood of the maximal path which could have led to a particular frequency in each time block. At the end

of the observation, the frequency associated with the maximum likelihood path is the frequency at which the path ended. This procedure does not determine the path which generated the maximum likelihood score since only the score itself is needed for detection, but path information can be obtained by working backward through the score matrix `S(v,p)` if it is needed. It is clear that the inner loop is executed $O(V^2P)$ times. The factor of V^2 is one of the significant disadvantages of the Viterbi method, since it tends to limit the frequency resolution that can be applied due to computational expense. This is an important problem since, unlike the stack algorithm, the Viterbi algorithm must be recomputed for every allowed combination of G_t and r_u, in order to maximize over these parameters as well. The factor of V^2 is introduced because every possible state transition must be considered and the value of $l_t(d)$ determined.

Streit and Barrett used a Gaussian $l_V(d)$ to track nonstationary sinusoids [8]. Since this function is chosen arbitrarily, it is more practical to apply a distance restriction like the one applied above in the stack algorithm. In this case, the transition likelihood function can be simply defined as

$$L_t(d) = \begin{cases} \frac{1}{3} & d \leq 1 \\ 0 & \text{otherwise} \end{cases} . \tag{2.15}$$

Of course, larger uniform likelihood regions could be defined and applied, but the $d \leq 1$ restriction guarantees continuity and also minimizes the computational cost of applying the Viterbi algorithm. For this application, it reduces the number of inner loop executions required to $3VP$, since only 3 transitions need to be evaluated for each frequency.

Just as the $d \leq 1$ restriction made the stack algorithm practical to compute, it also increases the efficiency of the Viterbi algorithm, so that it is possible to compute the exact maximum likelihood value without an overwhelming computational cost.

2.4.5 Relative Importance of G, r and θ

An in depth study, discussed in Chapter 3 and [12], of the sensitivity of Λ_{LRT} to the parameters G, r and θ has shown that detection performance is much less dependent on G and r than on θ. Since G and r are fixed for a given frequency path in the algorithm described above, a number of studies have been performed assuming known G and r. This assumption allows for reduced complexity in the design of the simulation experiments and substantial savings in time to complete the studies while testing the most complex part of the detection algorithm, the nonstationary frequency estimator. Studies in the final part of this thesis include all unknown parameters, but only after the usefulness of the known G and r assumption has been exhausted.

2.5 Analytical Evaluation

Simulations can be used to observe the performance of the algorithm for specific cases, but are rarely generalizable. To gain a more general insight into the behavior of the algorithm, it is desirable to consider some analytical results. As stated earlier, the detector statistic is a random variable generated by the quadratic form of a Gaussian random vector. The exact solution for the distribution does not have a closed form although it may be computed numerically. Another way to gain insight is to approximate the distribution and avoid the computation of the exact distribution function.

2.5.1 Numerical Characteristic Function Inversion

It is not perfectly straightforward to find the distribution function by performing numerical inversion of the characteristic function of a quadratic form of a complex GRV. First, let's consider the case of the real GRV, and then the extend the procedure to deal with the complex case. The quadratic random variable has the form

$$Z = \underline{r}'\mathbf{A}\underline{r} \quad \text{where} \quad E[\underline{r}\underline{r}'] = \mathbf{V} \tag{2.16}$$

where $\underline{r}$ is a zero mean GRV. Then the characteristic function of Z is given by

$$\phi_z(w) = \frac{1}{|\mathbf{I} + 2jw\mathbf{AV}|^{1/2}}. \tag{2.17}$$

Since all the matrices of interest for this discussion are covariance matrices or their inverses, they are known to have real eigenvalues, and the following simplification is possible [13]:

$$\phi_z(w) = \frac{1}{\prod_{n=1}^{N} (1 + 2jw\lambda_n)^{1/2}} \tag{2.18}$$

where the λ_n are the N eigenvalues of $\mathbf{AV}$. To find the PDF, compute the inverse integral

$$p_z(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_z(w) e^{-jwx}\, dw \tag{2.19}$$

where $p_z(x)$ is the PDF of Z. The cumulative density function (CDF) can also be obtained from the characteristic function.

Now consider the case of the CGRV with only one element. The unit variance CGRV, c, can be written in terms of two independent unit variance real GRVs as follows:

$$c = \frac{(r_r + jr_i)}{\sqrt{2}}. \tag{2.20}$$

This has an important implication for quadratic forms: there are twice as many degrees of freedom involved when the underlying GRV is complex as when it is real.

In the case of the simplest quadratic form, the central χ^2_n, a random variable Z with the χ^2_1 distribution can be formed by taking $Z_r = r'r$ (recall that the $'$ indicates the Hermitian transpose, or the complex conjugate in this scalar case). If, however, $Z_c = c'c$, the resulting distribution is $\chi^2_2/2$. Therefore, if the original GRV is complex, the characteristic function described in (2.18) must be modified to take the underlying CGRVs into account. This is easily accomplished by letting

$$\widehat{\lambda}_{2n-1} = \widehat{\lambda}_{2n} = \lambda_n/2 \tag{2.21}$$

where $n = 1$ to N and the $\widehat{\ }$ denotes the complex underlying GRV written in terms of real GRVs. This procedure doubles the degrees of freedom and accomplishes the

scaling simultaneously. Now (2.18) can be rewritten for the CGRV as

$$\widehat{\phi}_z(w) = \frac{1}{\prod_{n=1}^{2N}(1+2jw\widehat{\lambda}_n)^{1/2}} = \frac{1}{\prod_{n=1}^{N}(1+jw\lambda_n)} \tag{2.22}$$

where the right hand side is the characteristic function written in terms of CGRVs. It is important to consider the writing of a complex GRV in terms of real random vectors because the literature on numerical inversion deals almost exclusively with real GRVs. Therefore the transformation given in (2.21) is assumed to have been performed before using any of the following equations, and the $2N$ and $\widehat{\ }$ notation are dropped.

If the numerical integration of (2.18), or equivalently (2.22), were simply a matter of using an available numerical routine, there would be no need for further discussion. Unfortunately, the typical characteristic function is not well-behaved. Since the imaginary part cancels out and the real part is even, only the real part of (2.18) needs to be evaluated from 0 to ∞. However, the real part of (2.18) often has discontinuities which make it impossible for a numerical integration routine to converge. This undesirable property also holds for the numerical integration of the characteristic function to obtain the CDF of many quadratic forms.

Fortunately, the problem can be reformulated in terms of a continuous function which is simpler to integrate, although these alternate forms often retain poor behaviors such as rapid oscillation. Johnson and Kotz [14] give a formula for numerical inversion of the characteristic function to obtain the CDF due to Feiveson and Delaney [15]. However, the formula given in both contains an error. The corrected formula for the CDF of Z is

$$F_z(x) = \frac{2}{\pi}\int_0^\infty \frac{\sin(tx/2)}{t}\cos\left(\frac{-tx}{2}+\frac{1}{2}\sum_{n=1}^{N}\tan^{-1}2\lambda_n t\right)\prod_{n=1}^{N}(1+4\lambda_n^2t^2)^{-1/4}\,dt \tag{2.23}$$

which converges quickly in most cases. To determine a maximum value of t to sub-

stitute for ∞, let

$$b = \frac{1}{2}\left(\frac{4}{\epsilon\pi N\prod_{n=1}^{N}\lambda_n^{1/2}}\right)^{2/N} \tag{2.24}$$

where b is the upper limit of integration which is guaranteed to cause an error less than ϵ [15].

Typically, a step size must be chosen for the numerical integration, but a simpler method is to reduce the step size recursively by half until the integration value converges. Interestingly, all the sources agree that low order trapezoidal rule integration is most efficient for this problem, requiring fewer steps than higher order methods such as Simpson's or quadrature.

Other procedures based on various infinite series approximations are also available in the literature [16, 17, 18]. Although the literature deals with the general problem of numerical inversion often, only a few of the sources mentioned are specifically related to the problems associated with the characteristic functions of quadratic forms. Several make use of infinite series approximations instead of numerical integration. Specifically, Johnson and Kotz give a FORTRAN program to find values of the CDF and PDF of quadratic forms using Laguerre series [16]. This method is less desirable, both in terms of speed and consistent convergence, compared to the numerical inversion method described above.

However, the corrected method given in (2.23) is not the only numerical inversion based procedure to be found in the literature for this problem. Davies and Imhof describe a similar procedure that seems to be an alternative [19, 20], but has not been implemented in this research. Davies' and Imhof's methods do not appear to be restricted to positive eigenvalues as is (2.24), although there was no need to deal with negative eigenvalues in this analysis. The method given by Rice [21] appears to converge much more quickly than the methods above, but at the expense of requiring some initial values that must be tabulated in advance.

2.5.2 Distribution Approximation

There are many possible distribution approximations suggested in the literature [14, 22, 23]. Many involve expressing the quadratic form as an infinite sum of simpler distributions and truncating the sum to form an approximation. This approach usually requires a large number of the moments of the quadratic distribution to be calculated and tends to be too complex to lend much insight.

A simpler approach is to approximate the entire quadratic distribution by a single distribution such as the F, Γ, χ^2, or noncentral χ^2 distributions. The parameters of the simple distribution can be chosen to force its first few moments to match those of the quadratic distribution. This type of approximation is well suited to the purpose of gaining insight into the behavior of the generalized quadratic form distribution function. The first two moments of a *real* generalized quadratic form $\underline{r}'\mathbf{A}\underline{r}$ when $\mathbf{V} = E[\underline{r}\underline{r}']$ are given by $\mu = \mathrm{tr}[\mathbf{Q}]$ and $\sigma^2 = 2\,\mathrm{tr}[\mathbf{Q}^2]$ where $\mathbf{Q} = \mathbf{AV}$ [13]. Note that unlike the calculation of the exact distribution, calculation of these moments does not require calculation of the eigenvalues of $\mathbf{Q}$, making the process even more computationally attractive.

For a quadratic form in a CGRV, the resulting distribution has twice as many degrees of freedom as for the real case because of the "two-dimensional" nature of the CGRV. Then the moment formulae given above do not apply. Fortunately, the complex quadratic form can be written in terms of a real quadratic form, as shown in (2.21). Let $\mathbf{Q}_{re}$ be the matrix whose eigenvalues are the replacements for those of $\mathbf{Q}$, noting that $\mathbf{Q}_{re}$ is twice the dimension of $\mathbf{Q}$. The eigenvalues of $\mathbf{Q}$ and $\mathbf{Q}_{re}$ do not need to be calculated. Using the property of the trace function,

$$\mathrm{tr}[\mathbf{Q}] = \sum_i \lambda_i \tag{2.25}$$

it is easy to see that $\mathrm{tr}[\mathbf{Q}_{re}] = \mathrm{tr}[\mathbf{Q}]$ and that $\mathrm{tr}[\mathbf{Q}_{re}^2] = \mathrm{tr}[\mathbf{Q}^2]/2$, so only trivial calculations are required to convert between the two first moments.

One possible distribution approximation is the χ^2 approximation. This method

can be applied by considering a random variable distributed as $c\chi^2_f$ where c is a constant scale factor and f is the number of degrees of freedom. The values of c and f are set so that the moments of the resulting distribution are the same as those of the original distribution. The appropriate values are

$$c = \frac{\mathrm{tr}[\mathbf{Q}^2]}{\mathrm{tr}[\mathbf{Q}]}; \qquad f = \frac{(\mathrm{tr}[\mathbf{Q}])^2}{\mathrm{tr}[\mathbf{Q}]} \tag{2.26}$$

where $\mathbf{Q}$ is for a quadratic form in *complex* Gaussian random vectors, but c and f are in terms of the standard *real* distribution $c\chi^2_f$. The eigenvalue conversion has already been taken into account. This method has the advantage of being simpler to compute than the methods based on the F or noncentral χ^2 distributions. However, the Γ approximation is slightly more appealing since both its PDF and CDF are readily available in numerical libraries.

The Γ approximation to the generalized quadratic form distribution has been investigated and found to be reasonably accurate when the first two moments of the Γ distribution match those of the quadratic distribution [15]. The approximation was found to improve with increasing observation size, with observation sizes ranging from 4 to 21 samples [15]. The work of this thesis will normally use observation sizes from 8 to 64. The Γ distribution is given by

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} \tag{2.27}$$

where the appropriate parameters for approximating the generalized quadratic form are

$$\alpha = \frac{(\mathrm{tr}[\mathbf{Q}])^2}{2\,\mathrm{tr}[\mathbf{Q}^2]}; \qquad \beta = \frac{2\,\mathrm{tr}[\mathbf{Q}^2]}{\mathrm{tr}[\mathbf{Q}]}. \tag{2.28}$$

Again, the $\mathbf{Q}$ given above is for the complex Gaussian case, while the Γ distribution parameters are in terms of the standard real Γ function. The PDF can be computed easily since the Γ function is well known.

Fortunately, the CDF can be written in terms of the incomplete gamma function which is also a commonly available special function. The incomplete gamma function

is given as

$$\gamma(x,\alpha) = \int_0^x \frac{t^{\alpha-1}e^{-t}}{\Gamma(\alpha)}dt \tag{2.29}$$

and the CDF can simply be written as

$$F(x) = \gamma(x/\beta,\alpha). \tag{2.30}$$

The simplicity of computation for the Γ density makes it an excellent choice for investigation of the behavior of the detection algorithm.

The Γ approximation can be applied to the single block detector discussed in Chapter 3. The same approximation will work for the detector either in the time domain or in the spectral domain. It is interesting to note that diagonalizing in the spectral domain has no effect on the value of $\mathrm{tr}[\mathbf{Q}]$ but does have an effect on $\mathrm{tr}[\mathbf{Q}^2]$.

It is also possible to use the Γ approximation to evaluate full observation detectors such as the LRT, BLRT, and BLRTsd. In each case, the detector can easily be written in terms of a quadratic form, and the eigenvalues need not be calculated. The most difficult computation required is simply the squaring of the large $\mathbf{Q}$ matrix involved with the full observation detectors, which must only be performed once for each SNR scenario. Unfortunately, generating $\mathbf{Q}$ itself can be quite strenuous, since it has the dimension of the observation and in the worst case requires the inversion of both full observation covariance matrices.

The Γ approximation can be applied to the BLRT by simply forcing the block diagonal structure given in (2.6) onto the $\mathbf{A}$ matrix so that $\mathbf{A}$ becomes $\mathbf{A}_{bd}$, but not onto the $\mathbf{V}$ matrix. The resulting weight matrix would be

$$\mathbf{Q}_{BLRT} = \mathbf{A}_{bd}\mathbf{V}. \tag{2.31}$$

This same procedure can be extended to the BLRTsd method by performing a block diagonal spectral transformation on both $\mathbf{A}_{bd}$ and $\mathbf{V}$ and then forcing $\mathbf{A}$ to be diagonal. To specify the block diagonal spectral transformation matrix, let $\mathbf{W}_L$ be the FFT matrix in the dimension of the block size L. Then the block diagonal

spectral transformation matrix can be written as

$$\mathbf{W}_M = \begin{bmatrix} \mathbf{W}_L & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_L & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{W}_L \end{bmatrix}. \tag{2.32}$$

where M is the dimension of $\mathbf{W}_M$ and of the observation. The resulting weight matrix is equivalent to the result obtained when performing the BLRTsd detector,

$$\mathbf{Q}_{BLRTsd} = \mathcal{D}(\mathbf{W}_M \mathbf{A}_{bd} \mathbf{W}'_M)(\mathbf{W}_M \mathbf{V} \mathbf{W}'_M) \tag{2.33}$$

where the $\mathcal{D}()$ operator once again represents forced diagonalization.

If the Γ approximation is to be useful in the analysis of the GBLRT, it must be applied to the detection algorithm where some or all of the parameters are unknown. When there are unknown parameters, a maximization technique must be applied to maximize the likelihood of the estimated parameter value. Maximization of a group of random variables makes the resulting distribution difficult to find. However, if the random variables are independent and identically distributed (IID), the method of order statistics can be applied [24]. Suppose there is an IID group of n random variables Z_i and $y_n = \max Z_i$. If the distribution of Z_i is given by $f_z(x)$, and the CDF is given by $F_z(x)$, then the PDF of the y_n is given by

$$g_n(y_n) = n[F(y_n)]^{n-1} f(y_n). \tag{2.34}$$

This result can be applied to the Γ approximation since both the CDF and PDF of the Γ distribution are readily available. For performance evaluations, the CDF of y_n can be obtained by numerical integration. Since each Z_i represents a set of values of the unknown parameter set, the independence of these variables is dependent on the spacing between the values considered (parameter quantization). Even if independence is not an issue, the question of identical distributions is a significant problem. Although order statistics are a convenient method for dealing with the

maximization process in the detector, the assumption that all of the RV's under consideration are IID is a poor one which is not true for a number of cases. It will be shown later that an argument can be made for use of the IID assumption and order statistics when only the frequency is unknown and H_0 is true. Although this restriction will not allow a complete analysis of the detection algorithm, it is enough to allow for the computation of threshold values for a given probability of false alarm.

2.5.3 Verification of Distribution Approximation

The results obtained from the Γ approximation are compared with the simulated results for the LRT, BLRT, and BLRTsd detectors using the Receiver Operating Characteristic (ROC) plot figures below. Figure 2.2 is an example ROC plot of the Γ approximation of the LRT. The figure simply shows the probability of detection for a given probability of false alarm. Each line corresponds to a single SNR value. ROC plots are particularly useful for characterizing detectors with variable false alarm rates. The format of Figure 2.2 is based on a "normal probability paper" plot, so that all the probabilities are plotted with distances transformed by the inverse of a normal RV's CDF, giving an effect similar to a bi-directional log plot. This "normal probability paper" transformation makes the ROC curves into nearly straight lines, and displays detail in the regions near zero and one which would be invisible in a linear ROC plot.

As the SNR increases, the ROC curve moves toward the upper left hand corner—this property will hold for all the ROC curves shown below, so that the curves need not be individually labeled. The dashed line in Figure 2.2 represents the performance of a detector which gains no information from the signal, i.e. the distribution under both hypotheses is the same. In this case, the probability of false alarm and the probability of detection are the same, which is the worst case performance for a detector. The curves never fall in the lower right region of the plot, since a detector whose performance was in this region could be improved by reversing its guesses,

bringing the curve into the upper left region.

To verify the accuracy of the Γ approximation, Figure 2.3 shows an ROC plot for the LRT detector. The solid lines are obtained from the Γ approximation, and the dashed lines are from simulations. The simulations were performed with 200 iterations per hypothesis at each SNR plotted. The figure clearly shows that the approximations are giving results similar to the simulated data. The ends of each simulated line are often less accurate than the middle of the line because the data points tend to cluster near the mean of the distribution. This effect also explains why the simulated lines end near the 0.99 or 0.01 tic marks on the plot. Since there are only 200 iterations, the simulated probability estimates never range past 0.005 and 0.995, and have a very high variance when they approach these regions. Figures 2.4 and 2.5 demonstrate the accuracy of the Γ approximation for the BLRT and BLRTsd detectors. Since the simulated values on the ROC plots are difficult to generate accurately near zero and one, it is helpful to consider an alternative to Monte Carlo sampling which reduces the variance for estimates of small probabilities.

2.6 Importance Sampling

When an analytical approach to determining detector performance is not possible, simulations are often the only alternative. Simple Monte Carlo simulation is very effective for determining the transition region of the detector, and it is very accurate and inexpensive when the detection or false alarm event probability is near $\frac{1}{2}$. However, when the event probability is near zero or one, the variance of the estimate from Monte Carlo simulation increases dramatically. Without loss of generality, consider the case when the event probability is near zero. To achieve 95% confidence in the interval $[2\widehat{P}_e/5, 8\widehat{P}_e/5]$, Monte Carlo simulation normally requires $\frac{10}{P_e}$ trials where P_e is the actual event probability [25]. Clearly, as P_e becomes small, the required number of trials becomes prohibitive. Importance sampling, somewhat of a hybrid between

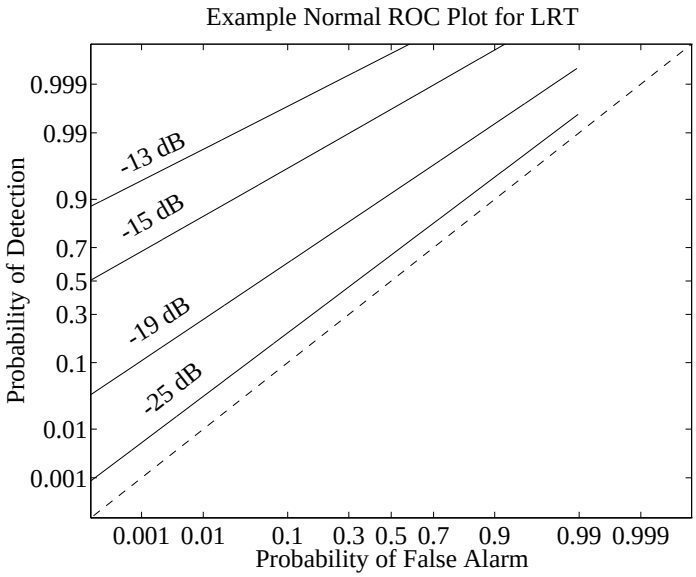

Figure 2.2: An example ROC curve at four SNR values of -25, -19, -15, and -13dB. The worst case dashed line is shown to illustrate the performance of a detector which gains no information from the signal—hence the probability of false alarm is equal to the probability of detection.

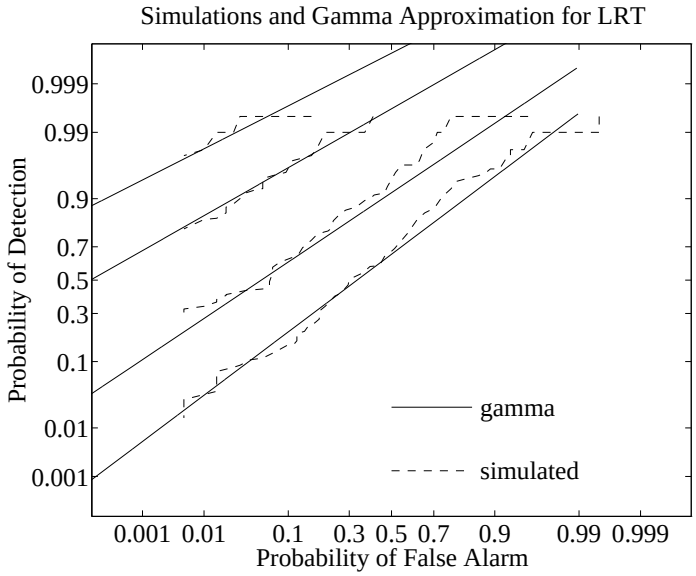

Figure 2.3: The simulated and Γ approximation curves match closely for the LRT with a stationary signal at four SNR values of -25, -19, -15, and -13dB. Each simulated curve is the result of 200 iterations per hypothesis.

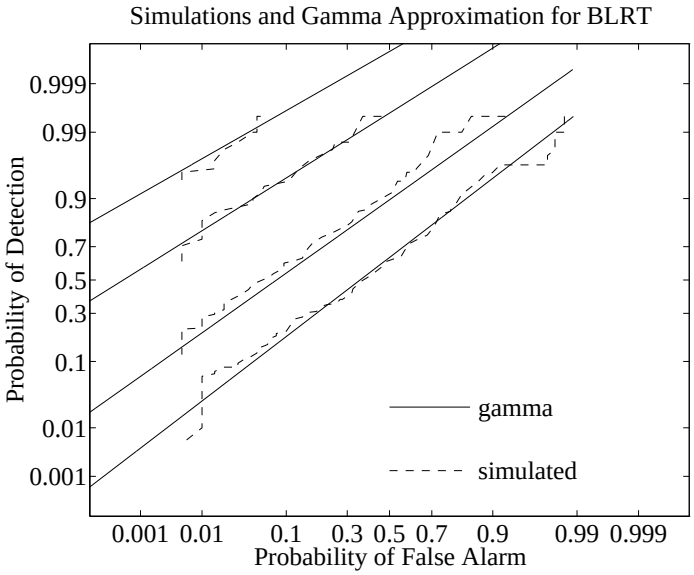

Figure 2.4: The simulated and Γ approximation curves match closely for the BLRT with a stationary signal at four SNR values of -25, -19, -15, and -13dB. Each simulated curve is the result of 200 iterations per hypothesis.

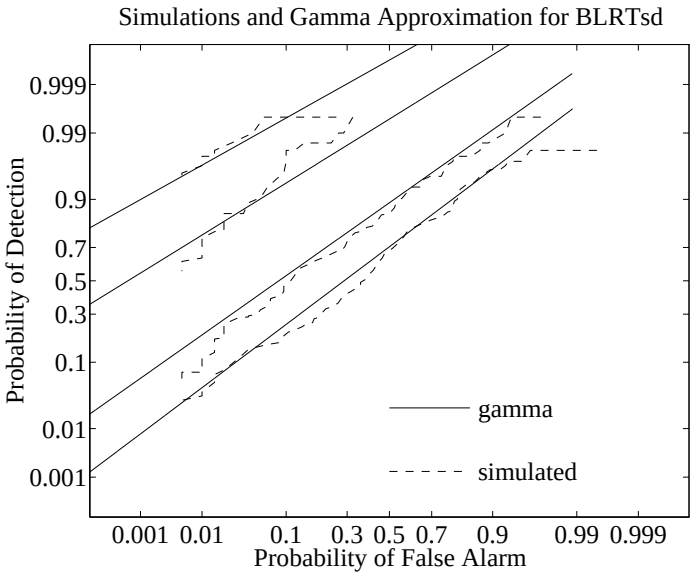

Figure 2.5: The simulated and Γ approximation curves match closely for the BLRTsd with a stationary signal at four SNR values of -23, -19, -15, and -13dB. Each simulated curve is the result of 200 iterations per hypothesis.

brute force simulation and analytical methods, can be used to reduce the variance of the estimate of P_e, so that many fewer trials are required.

Importance sampling requires that the PDF of the input data to the detector be biased in a way that the event is more likely, and then the events are assigned a weight according to the bias of the PDF. The Monte Carlo estimation formula for N trials is

$$\hat{P}_e = \frac{1}{N} \sum_{i=1}^{N} I_{\mathcal{Z}_e}(\underline{r}_i) \tag{2.35}$$

where

$$I_{\mathcal{Z}_e}(\underline{r}) = \begin{cases} 1 & , \quad \underline{r} \in \mathcal{Z}_e \\ 0 & , \quad \underline{r} \in \mathcal{Z}_e^c \end{cases} \tag{2.36}$$

so that (2.35) simply counts the number of events that occur with $\underline{r} \in \mathcal{Z}_e$. The importance sampling estimation formula is

$$P_e^* = \frac{1}{N} \sum_{i=1}^{N} W(\underline{r}_i^*) I_{\mathcal{Z}_e}(\underline{r}_i^*) \tag{2.37}$$

where $W(\underline{r}_i^*)$ is the weighting function with the form of a likelihood ratio

$$W(\underline{r}_i^*) = \frac{\frac{1}{(2\pi)^{M/2}|\mathbf{K}|^{1/2}} \exp\left(-\frac{1}{2}\underline{r}_i^{*T}\mathbf{K}^{-1}\underline{r}_i^*\right)}{\frac{1}{(2\pi)^{M/2}|\mathbf{K}_*|^{1/2}} \exp\left(-\frac{1}{2}\underline{r}_i^{*T}\mathbf{K}_*^{-1}\underline{r}_i^*\right)} \tag{2.38}$$

and the $\underline{r}_i^*$'s are generated with covariance $\mathbf{K}_*$ instead of with $\mathbf{K}$.

The main difficulty in the use of importance sampling is the choice of $\mathbf{K}_*$. For one dimensional cases, the choice is rather simple, but for multidimensional problems it can be more difficult. Orsak [26] has suggested a simple biasing density choice for detector simulations, known as the "swap" method, which dictates that rather than using $\mathbf{K}_0$ for false alarm simulations, $\mathbf{K}_* = \mathbf{K}_1$. A choice of $\mathbf{K}_*$ which is asymptotically optimal as the false alarm probability decreases has been derived more recently [27]. The new asymptotic method is a substantial improvement over the swap method for small P_e, but the asymptotic approximation breaks down for P_e near .5, so that the new method is not applicable in that region. Fortunately, this region can be evaluated accurately using the Monte Carlo method.

The importance sampling approach is useful for cases where a threshold for a very small probability of false alarm must be calculated, since the Monte Carlo method requires very large numbers of trials but importance sampling does not. One measure of the benefit of importance sampling is the importance sampling gain defined as the ratio of the number of trials required by Monte Carlo simulation to the number of trials required to obtain the same variance with importance sampling. This is defined for a fixed variance as $\Phi^* = N_{IS}/N_{MC}$.

Orsak has suggested that importance sampling gain can be estimated from the relationship [25]

$$\Phi^* = \frac{P_e^*(1 - P_e^*)}{\overline{W}^* - P_e^{*2}} \tag{2.39}$$

where P_e^* is estimated as in (2.37) and $\overline{W}^*$ is estimated by

$$\overline{W}^* = \frac{1}{N}\sum_{i=1}^{N} W^2(\underline{r}_i^*) I_{\mathcal{Z}_e}(\underline{r}_i^*). \tag{2.40}$$

Importance sampling has been shown to produce excellent results in application to the quadratic detector for random signals. The estimated importance sampling gain for false alarm probability estimation for an optimum quadratic detector (the LRT) is shown in Figure 2.6 for three possible biasing densities of the input signal. The covariance of the real input signal is a Toeplitz matrix whose first row is $[1\ 0.9\ (0.9)^2\ (0.9)^3]$ with white noise added at varying power to generate different false alarm probabilities. The dashed curve in Figure 2.6 is the swap method given by Orsak [26], the dot-dashed curve is an approximate asymptotic biasing density given by Padgett and Williams [27], and the solid curve is a more accurate approximation of the asymptotic density also given in [27].

Importance sampling is primarily required when accurate estimates of small probabilities are needed. The majority of the work in this thesis, however, only requires a knowledge of the transition region of the detector, i.e. what SNR is required for the detector to give reasonable performance for the purpose of comparing various algorithms. Also, the large observation sizes make the covariance matrices implicit in

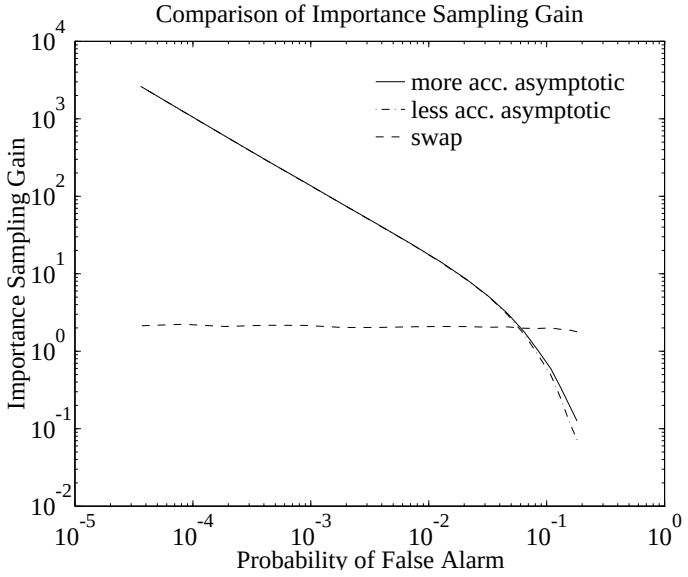

Figure 2.6: Estimated importance sampling gain for a particular detector with three of the biasing densities: the swap method, the less accurate asymptotic method, and the accurate asymptotic method.

(2.37) very laborious to compute. Consequently, importance sampling was useful for certain small observation test cases, but not for the large observation size studies in this investigation.

CHAPTER 3

Effects of Approximations on Detection

A number of approximations have been implemented as part of the GBLRTsd detector. Each individual approximation can be analyzed separately or as part of the entire detection algorithm. When the approximations are dealt with separately, it is simpler to isolate the conditions under which they perform most accurately and most poorly. This information can then be applied to the entire algorithm to gain insight into its behavior. This chapter will focus on analyzing the approximations separately and determining under what conditions they tend to perform best and worst.

3.1 Single Block Analysis

Because the GBLRTsd detector simply adds likelihood values calculated from single blocks, it is possible to evaluate the effects of several approximations in the simpler single block calculation with good confidence that the full observation detector will be similarly affected. The spectral diagonalization approximation, the quantization of parameters, and the first order model assumption can all be evaluated on the single block level. The single block likelihood value is of the form $\underline{x}'(p)\mathbf{A}\underline{x}(p)$.

Since the single block calculation is simply a quadratic form of a CGRV, it is simple to write the characteristic function of the resulting random variable as in (2.22). Unfortunately, it is much more difficult to find the distribution, since it is not one of the well known forms, but a general quadratic form. The CDF can only be obtained by numerical characteristic function inversion. There is significant literature on this procedure [14, 15, 16, 17, 18, 19, 20, 21, 28, 29, 30]. By using numerical characteristic function inversion as discussed in Section 2.5.1, it is possible to analyze

the effect of the above mentioned approximations *without simulations*. The numerical analysis has the advantage of avoiding concerns about confidence intervals, and is less computationally expensive than simulation.

3.1.1 Spectral Diagonalization Approximation

When the observation and the associated covariance matrix are converted into the spectral domain as described in Section 2.3, the resulting spectral covariance matrix $\widetilde{\mathbf{K}}_i$ tends to be diagonalized. This diagonalization can be seen by a simple derivation based on the formula for each element of $\widetilde{\mathbf{K}}_i$ [5]. The (k,l)th element of $\widetilde{\mathbf{K}}_i$ is given by

$$\widetilde{\mathbf{K}}_{i_{(k,l)}} = \sum_{m=0}^{N-1} e^{-j\frac{2\pi km}{N}} \sum_{n=-m}^{N-1-m} R(n) e^{+j\frac{2\pi l(n+m)}{N}} \tag{3.1}$$

where $R(n)$ for the AR(1) signal is given by (1.9). By reversing the summations and combining the exponentials, this can be written as

$$\widetilde{\mathbf{K}}_{i_{(k,l)}} = \sum_{n=0}^{N-1} R(n)[e^{+j\frac{2\pi kn}{N}} + e^{-j\frac{2\pi ln}{N}}] \frac{sin[\pi(N-n)(k-l)/N]}{sin[\pi(k-l)/N]}. \tag{3.2}$$

When $k = l$, the result can be further simplified to

$$\widetilde{\mathbf{K}}_{i_{(k,k)}} = \sum_{n=-(N-1)}^{N-1} (N-|n|) R(n) e^{+j\frac{2\pi nk}{N}} \tag{3.3}$$

which applies only to the diagonal elements of the spectral covariance.

Therefore, (3.2) shows that the transformation is equivalent to applying a time-domain window to the autocorrelation before evaluating the FFT of one or two frequencies to create the (k,l) element of $\widetilde{\mathbf{K}}_i$. The applied window is the Barlett window when $k = l$ and is much smaller for off-diagonal elements. Thus the spectral covariance matrix is guaranteed to be nearly diagonal [5].

Figure 3.1 shows the behavior of the spectral transformation in the time domain. The real part of an AR(1) autocorrelation function is shown in Figure 3.1(a) and three of the windows are shown in Figure 3.1(b). The solid line in (b) is the Bartlett window which is applied when a diagonal element of $\widetilde{\mathbf{K}}_i$ is computed. In this case (3.3) can be

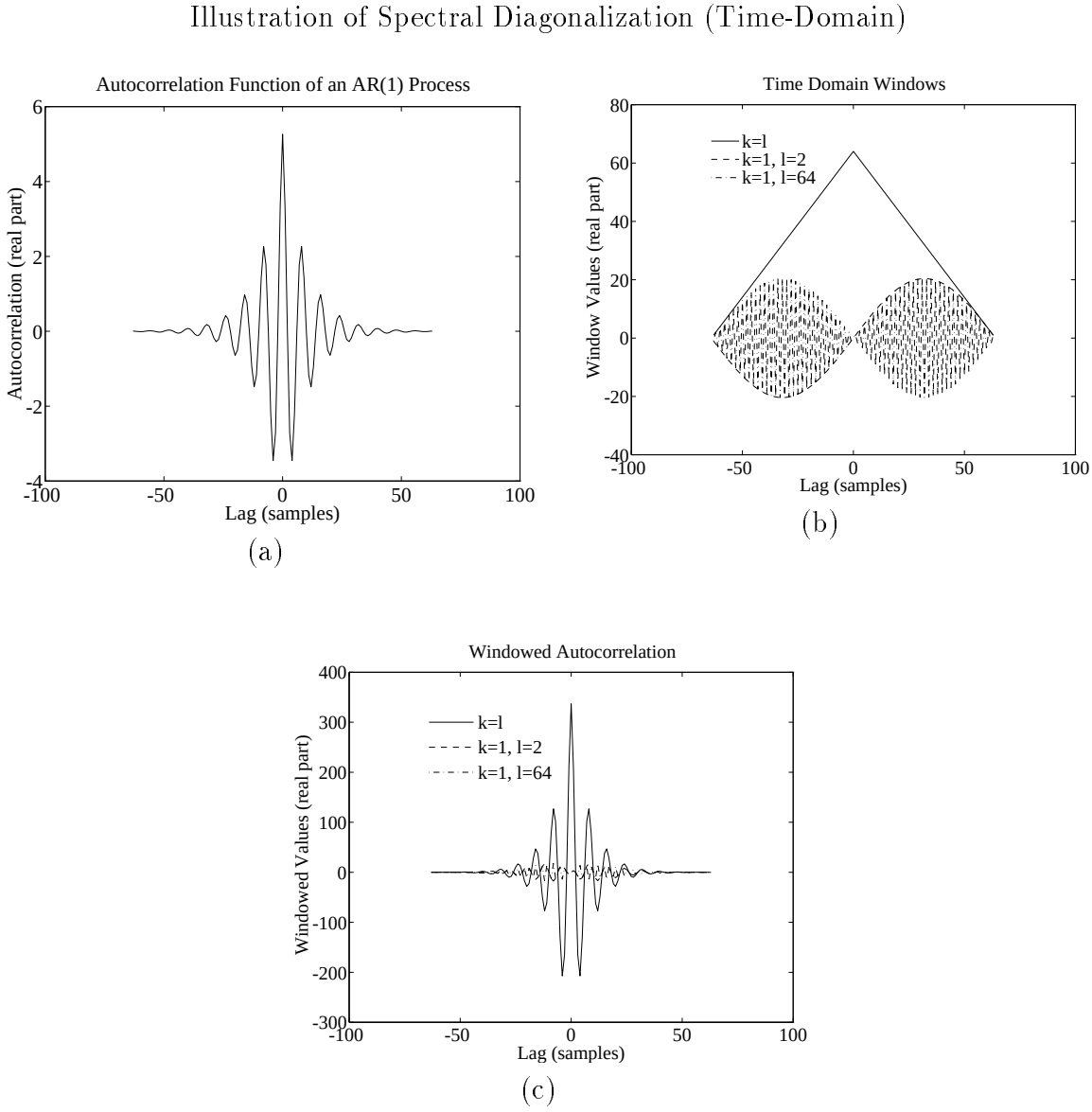

Figure 3.1: Part (a) is the real part of the autocorrelation function of an AR(1) process with pole $0.9e^{j\pi/4}$. Part (b) is a comparison of the time domain windows for a 64 sample observation. Part (c) shows the resulting windowed autocorrelation values.

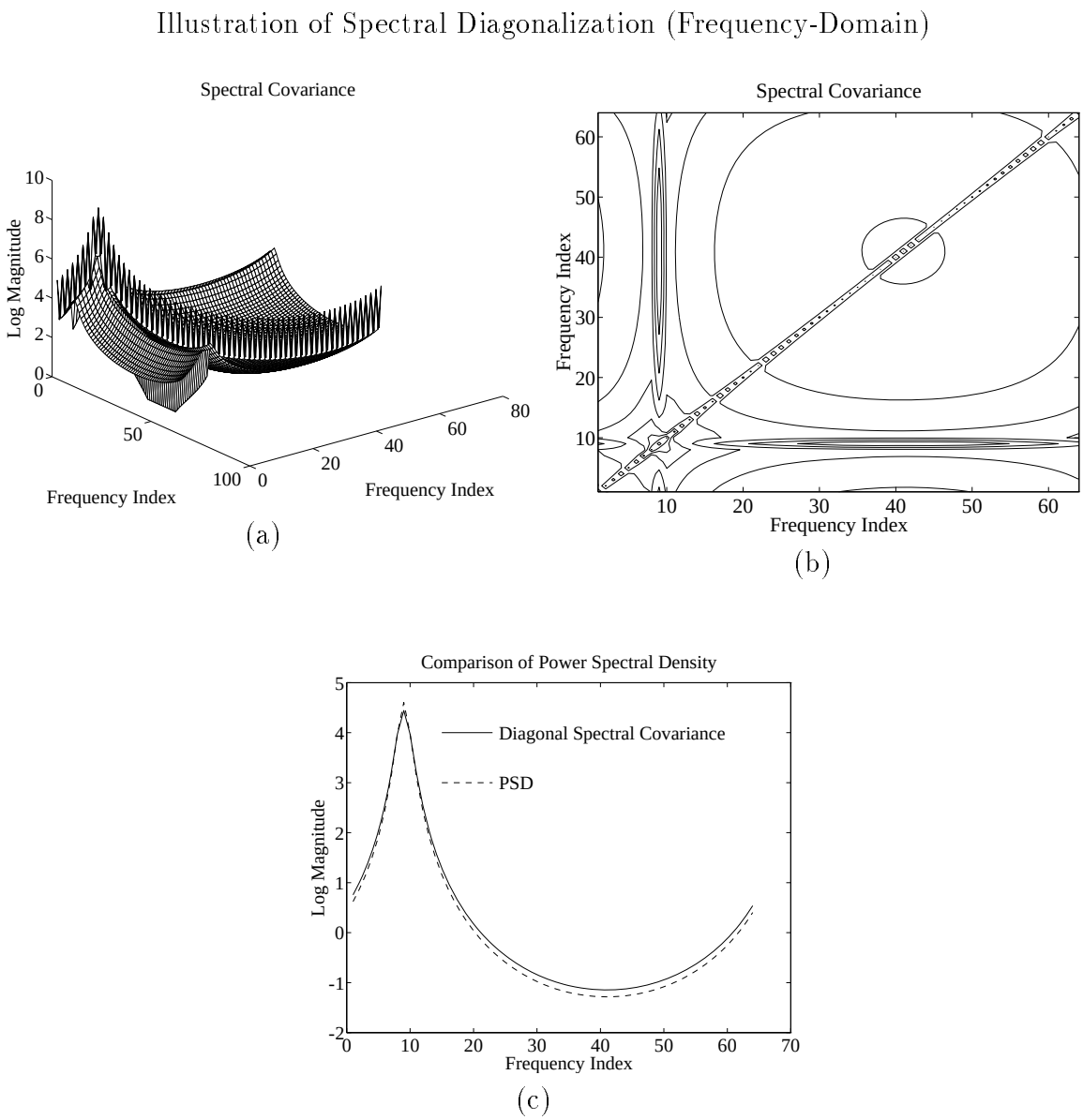

Figure 3.2: Parts (a) and (b) show the log magnitude of the spectral covariance matrix. Part (c) shows the slight difference between the scaled diagonal of the spectral covariance and the actual PSD.

interpreted as the evaluation of the FFT of the windowed $R(n)$ at frequency $-2\pi k/N$ radians. When $k \neq l$, (3.2) can be interpreted similarly as the sum of the FFT values of the windowed $R(n)$ at two frequencies. Figure 3.1(c) shows the real parts of the resulting windowed $R(n)$ functions for the same cases. Note that the value of the windowed function is always zero at zero lag for off-diagonal values. This result is due to the sinusoidal nature of the off-diagonal window, and is very important since the peak of the autocorrelation always occurs at zero lag.

Figure 3.2 shows the behavior of the spectral transformation in the spectral domain. The sharp line of spikes in Figure 3.2(a) is the diagonal of the spectral covariance matrix, which clearly stands out from the off-diagonal elements. The two ridges which correspond to the peak are caused by the interaction of the exponentials, the window frequency, and the frequency of $R(n)$. Their position can be clearly seen in (b). The centers of the ridges are not peaks but troughs, apparently due to destructive interference. Part (c) of the figure shows that the diagonal of the spectral covarinace (scaled down by 64) is essentially the same as the PSD of the signal. The differences are due to the use of the FFT instead of the discrete-time Fourier transform.

The spectral diagonalization approximation improves as the number of samples increases since the FFT becomes an orthogonal transformation when its length goes to infinity. Therefore, if an infinite number of samples were available, the FFT would diagonalize every covariance matrix and would be equivalent [4] to the Karhunen-Loeve (K-L) decomposition. Although the FFT is only an approximation to the K-L transformation for finite matrices, it happens to be a reasonably good one for many types of signals. A more thorough discussion of this property for the continuous-time Fourier transform is given by Van Trees [4].

Rather than interpret the distribution function of the quadratic form of the single block likelihood, it is simpler to treat the statistic as part of a single block detector, so the effect on detector performance can be evaluated directly. Therefore in the studies to follow, the likelihood distributions under both H_0 and H_1 were found and

the detector performance was plotted for a constant false alarm rate (CFAR) detector as probability of detection versus SNR.

The figures in Section 3.1 were calculated for CFAR detection by determining the detector threshold values required to fix the false alarm rate at each value of the SNR. Although this approach produces a simple plot, it does not reflect the fact that the quadratic detector is not inherently CFAR. Since these plots relate only to the single block analysis, it was practical to compute thresholds to fix the false alarm rate and simplify the plots. However, outside this section the detector's performance will be displayed in terms of the ROC curve, which requires less computation and reflects the variable false alarm rate inherent in the detectors.

The purpose of the first study is to measure the loss in performance of the single block detector using the spectrally diagonal approximation as compared to the optimal LRT detector. The single block detector using the spectral diagonalization approximation will be referred to as the spectrally diagonalized detector or just the diagonalized detector. The FFT does not remove any information, so using the spectral form of the detector (without diagonalization) is still optimal. By forcing diagonalization on the spectral covariance matrix, the diagonal values are being used as if they are the eigenvalues obtained by K-L transformation. The loss of performance is directly related to the difference between the eigenvalues and the diagonal values of the spectral covariance matrix. Although it is possible to write exact expressions for the difference between the eigenvalues and the diagonal values, there is no useful expression for the CDF of a general quadratic form random variable. Therefore, it is impossible to write a direct formula for the loss in detection performance.

Numerical results will be shown for a wide range of signal parameters that demonstrate only a small loss of performance due to the diagonal approximation, e.g. Figure 3.3. The performance of the detectors can be evaluated by using numerical inversion of the characteristic function under H_0 to find a threshold for the desired probability of false alarm, and then using inversion under H_1 to find the probability

of detection. In this way performance of the diagonalized detector can be compared accurately with the optimum detector.

While the FFT is a well known and highly efficient algorithm, its performance for some purposes can be improved by using a time-domain window to smooth the spectral version of the signal [31]. The FFT has the property

$$\mathcal{F}\{h(m)x(m)\} = \frac{1}{M}H(k) * X(k) \tag{3.4}$$

where $H(k) = \mathcal{F}\{h(m)\}$ is the FFT of $h(m)$, M is the number of samples, and $*$ is the circular convolution operator. Various popular windows such as the Hamming or Kaiser windows can be be substituted for $h(m)$ to smooth the spectrum of the signal, $X(k)$. This knowledge combined with the frequency domain window interpretation of the diagonalizing effect might lead one to believe that the diagonalizing property could be improved by including a time-domain window in the FFT transformation.

Adding a window to the spectral transformation of (2.7) gives

$$\underline{\tilde{x}}(p) = \mathbf{W}\mathbf{T}_h\underline{x}(p) \tag{3.5}$$

where $\mathbf{T}_h$ is a diagonal matrix with $h(m)$ as its diagonal. The resulting spectral covariance is given by

$$\widetilde{\mathbf{K}}_i(p,p) = \mathbf{W}\mathbf{T}_h\mathbf{K}_i(p,p)\mathbf{T}_h'\mathbf{W}'. \tag{3.6}$$

Intuition suggests that this structure should improve detection performance, but experience has shown that this is not the case. Since the distribution of the detector statistic under both hypotheses can be found from the numerical CDF inversion technique, the effect of a time window on the probability of detection can be evaluated and compared to the identical case without the window.

The use of a Hamming or Kaiser window for h consistently reduced the diagonalized detector performance, even for many different damping coefficients applied to the Kaiser window. Since the Kaiser window is considered the optimal trade-off for smoothing and resolution, it seems unlikely that any reasonable window would improve performance. In addition, attempts to find a better window through numerical

optimization have resulted in windows which slightly emphasized the end-points of the vector and improved the performance only by insignificant amounts attributable to roundoff error.

It appears that distorting the time-domain version of the vector at best maintains the same level of performance. It is clear that the time domain window cannot improve the detector performance in the absence of the spectral diagonalization approximation, since the LRT detector is optimal. Although the LRT cannot be improved by a time-domain window, it is somewhat surprising that no window could be found to improve the performance of the spectrally diagonalized detector.

Evaluations of the diagonalized detector have been performed for variations in both bandwidth (pole radius) and frequency (pole angle). For each parameter the detector was tested with a fixed probability of false alarm of 0.01 over a range of signal to noise ratios from -30dB to about 5dB so that the probability of detection ranged roughly from 0.01 to 1. Figures 3.3 and 3.4 show how the performance of the diagonalized detector varies as the signal bandwidth increases.

It is important to note that with increasing signal bandwidth, the performance of the optimal detector declines and approaches that of the energy detector. This degradation is because the energy detector is the optimal detector for a white signal in white noise, and, as the signal bandwidth increases, it becomes more like white noise. It should also be noted that the energy detector curve is different in Figures 3.3 and 3.4. This difference is due to the fact that for the case of a random, non-white Gaussian signal, the PDF of the energy detector output is a generalized quadratic form which is signal dependent. The same method of numerical CDF inversion was applied to this case as for the other quadratic detectors to analyze the performance.

The diagonal approximation shows a slight decrease in performance from the optimal when the signal has a distinct spectral peak. The performance loss disappears as the signal becomes more wideband in nature. Thus as the signal becomes more wideband, the diagonal approximation approaches the optimal detector which in turn

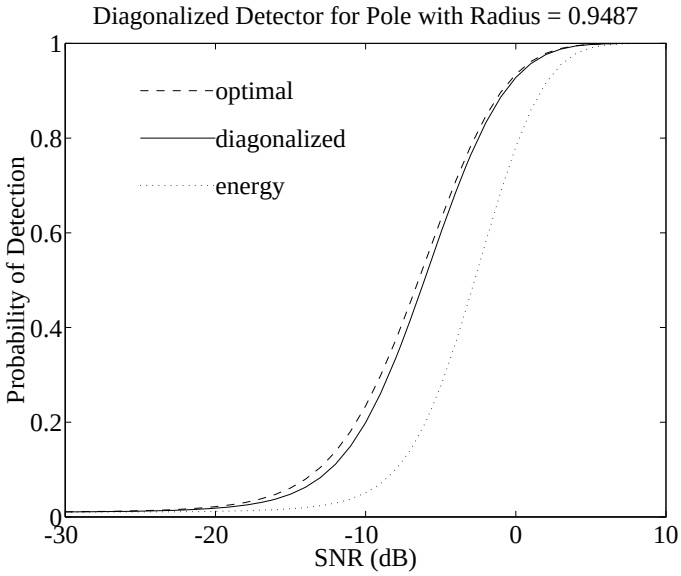

Figure 3.3: Probability of detection for three detectors: the optimal covariance detector, the diagonalized detector, and the energy detector. All three detectors were analyzed using the same input signal. The input signal is a complex first order AR signal with pole = $0.6018 + 0.7333i$, giving a radius of 0.9487. The signal gain was chosen to give the SNR value specified, and the threshold was set in each case to give a probability of false alarm of 0.01.

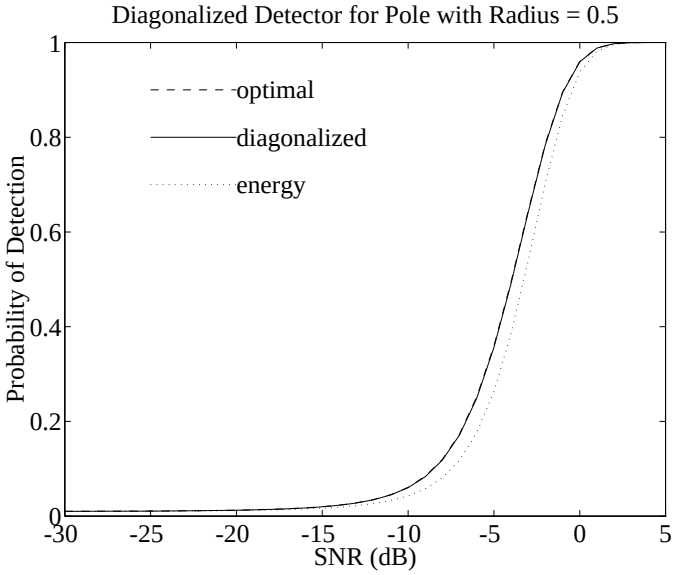

Figure 3.4: Probability of detection for three detectors: the optimal covariance detector, the diagonalized detector, and the energy detector. All three detectors were analyzed using the same input signal. The input signal is a complex first order AR signal with pole = $0.3172 + 0.3865i$, giving a radius of 0.5. The signal gain was chosen to give the SNR value specified, and the threshold was set in each case to give a probability of false alarm of 0.01.

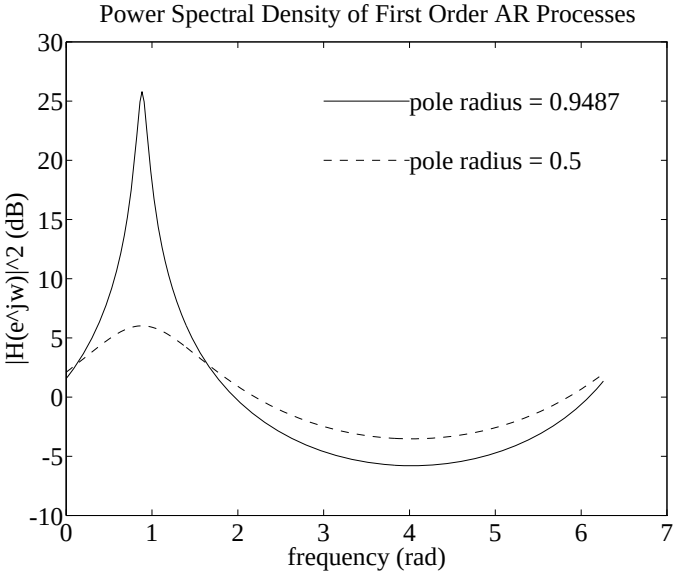

Figure 3.5: The power spectral density of the large and small bandwidth signals.

approaches the energy detector. Both the optimal and the diagonalized detector closely follow the energy detector performance for extremely wideband signals, such as those with a radius in the range of 0.1. Figure 3.5 shows the spectral peak of the PSD of the signals with pole radii of $\sqrt{0.9} = 0.9487$ and 0.5.

This algorithm has also been evaluated for performance at different frequencies. Intuitively, one would not expect any variation in behavior since the PSD is fully represented and is only shifted to obtain a new frequency. In fact, the optimal algorithm shows no meaningful change in performance for signals of constant bandwidth and varied frequency. However, the diagonal approximation has a loss in performance which is related to the proximity of the signal frequency to the FFT bin frequency. The spectral covariance matrix is more effectively diagonalized when the signal is not split between two bins. Therefore, the diagonal approximation algorithm has the greatest performance loss when the signal pole frequency is exactly between two FFT bins, and least when the signal pole corresponds exactly with an FFT bin frequency. Figure 3.6 shows the optimum detector performance compared with the diagonal approximation for an on-bin frequency, the diagonal approximation for a between-bin frequency, and the energy detector. Since the diagonalized detector shows the largest loss of performance for narrowband signals, this comparison was performed with a signal pole radius of 0.9487 to emphasize the difference between the optimal and the approximation.

3.1.2 Parameter Quantization

When a parameter is unknown, the search for a MLE of the parameter can be performed more rapidly by using a finite set, or quantized set, of values of the parameter. Although this method is faster, the exact MLE will not be one of the quantized values, so the result is only an approximate MLE. The parameter quantization approximation makes the implicit assumption that small errors in the parameter values will have little or no effect on the detector's performance. However, the size of spacing

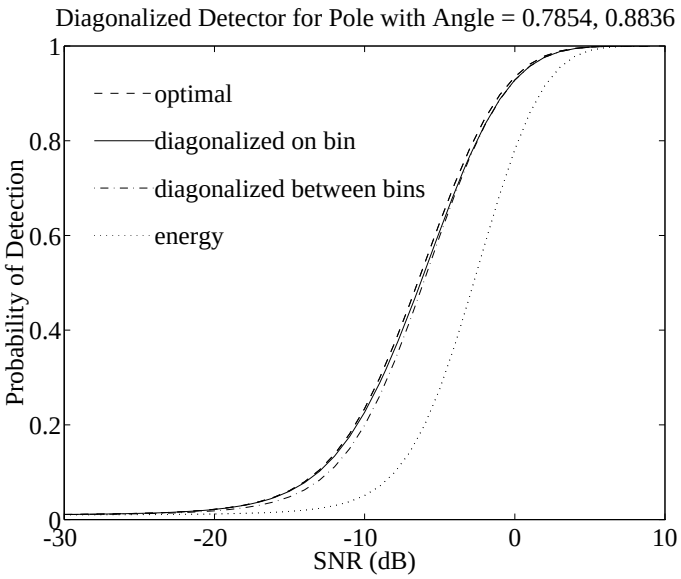

Figure 3.6: Probability of detection for the three detectors, each compared for two similar signals. One signal pole had the angle 0.7854 radians which is on-bin, or the same as the fourth of 32 FFT bins. The other signal pole had the angle 0.8836 radians which is exactly between the fourth and fifth bins. Both poles had radius 0.9487, and the probability of false alarm was set to 0.01.

between allowed values clearly has a strong effect on the degradation associated with the quantization. Therefore, it is quite important to study the behavior of the detector in the presence of small errors in the parameter values. Since the actual signal parameters are known, the study can be performed in terms of parameter mismatch, which is simply the difference between the actual signal parameter value and the value assumed by the detector.

Knowledge of the sensitivity of the detector to each parameter can be exploited to reduce search time by choosing the largest parameter step size within the limits of acceptable performance degradation. All three parameters will be evaluated: frequency mismatch, bandwidth mismatch, and gain mismatch. The energy detector performance will be even more important in this study as an established minimum. When the spectrally diagonalized detector's performance drops below that of the energy detector's, it is clear that the inaccurate model information is reducing the detector performance rather than enhancing it, indicating that the parameter mismatch is too large.

A special case arises when the detector is particularly insensitive to a parameter. If the range of reasonable parameter values is smaller than the step size, then that parameter can be estimated independently from the others, reducing the order of the search.

In order to carefully examine the behavior of the detectors in the presence of parameter mismatch, four signals have been chosen which represent the permutations of large and small bandwidth, and large and small signal power. The signal frequency is the same for each signal, since the problem is symmetric with respect to frequency. Note however, that the pole angle of $4.5(2\pi)/32$ is between FFT bins and is a worst case choice that maximizes the difference between the optimal and diagonalized detectors. The small bandwidth pole radius was chosen as 0.9487 while the large bandwidth radius was 0.5. These are the same values chosen for analysis as in Figures 3.3, 3.4, and 3.5. The large signal was given an SNR of 0dB, while the

small signal has an SNR of −10dB. These values are near the upper and lower knee of the detection curve, respectively, for both the large and small bandwidth signals.

Figure 3.7 shows a comparison of the performance of the optimal, diagonal, and energy detectors in the presence of signal power mismatch for each of the four signals described above. Each plot gives the probability of detection for the detector formed assuming the SNR value of the horizontal axis, operating on the signal with parameters described in the plot title. The vertical dotted line marks the actual signal SNR in each plot. The energy detector performance is the same for any assumed power, since no SNR assumption is used and, of course, the signal remains the same across a given plot. The two plots on the left of the figure show the performance of the same tests (based on the same pole and range of assumed SNR) on two signals, each with a different actual SNR. Likewise, the two plots on the right of the figure compare an identical set of tests performed on two signals of different SNR.

It is apparent from Figure 3.7 that, as suggested earlier, the diagonalized detector suffers only a slight loss of performance when an incorrect SNR is assumed. Even when the assumed value differs by as much as 20 to 30dB the probability of detection only drops by about 0.1 and never falls as low as the energy detector. In most cases the diagonalized detector maintains its performance slightly below that of the optimal detector. In the upper left plot of Figure 3.7, the diagonalized detector performance exceeds that of the optimal detector for low values of assumed SNR. This is an indication of the robustness of the diagonalized detector, since it exhibits less sensitivity to SNR mismatch than the optimal detector. There is no need to be concerned that the diagonalized detector outperforms the optimal one for this case, since the optimal detector is no longer the minimum probability of error solution when the wrong SNR value is assumed. In order to obtain the best performance for the worst case (a minimax type of solution), an SNR near the lower knee of the detection curve, or in this case about −10dB, should be assumed. This choice minimizes performance loss for the low SNR case without causing an extreme loss for

the high SNR case.

The plots in Figure 3.8 are a comparison of performance in the presence of mismatch of assumed and actual pole angle. As with the previous figure, the vertical dotted line marks the actual signal pole angle, and the two plots on the left of the figure show the same set of tests performed on two signals of different SNR. The two plots on the left are generated similarly, except with a pole radius that gives a larger bandwidth. From the plots in Figure 3.8 observe that once again the diagonalized detector is quite insensitive to pole angle mismatch, allowing fairly large steps to be used in a parameter searching procedure. Naturally, larger steps are acceptable for the larger bandwidth signal. The diagonalized detector is most sensitive to pole angle mismatch in the low SNR and small bandwidth cases (the lower left plot), which are intuitively the most difficult cases. Once again the diagonalized detector is more robust than the optimal detector, particularly in the small bandwidth case.

The final comparison considers the case of pole radius mismatch. The four plots in Figure 3.9 show the familiar four signals and the performance of the detectors based on the assumption of the pole radius value shown on the horizontal axis. Once again observe that the diagonalized detector is quite robust, since its performance remains above that of the energy detector throughout the range of pole radii tested, and generally degrades in performance more slowly than the optimum detector. There are two interesting features of the plots in Figure 3.9. First, the rapid drop in performance is present in each plot for the diagonalized detector when the pole radius assumed by the test is near 1. The optimal detector is unaffected by this problem, which suggesting that the diagonalization is the culprit. This is in fact the case, as the spectral covariance matrix becomes less and less diagonal in nature as the pole radius approaches 1. The non-diagonal form of the spectral covariance matrix causes the eigenvalues of the matrix to differ drastically from the diagonal elements, thus invalidating the spectrally diagonal approximation on which the diagonalized detector is based.

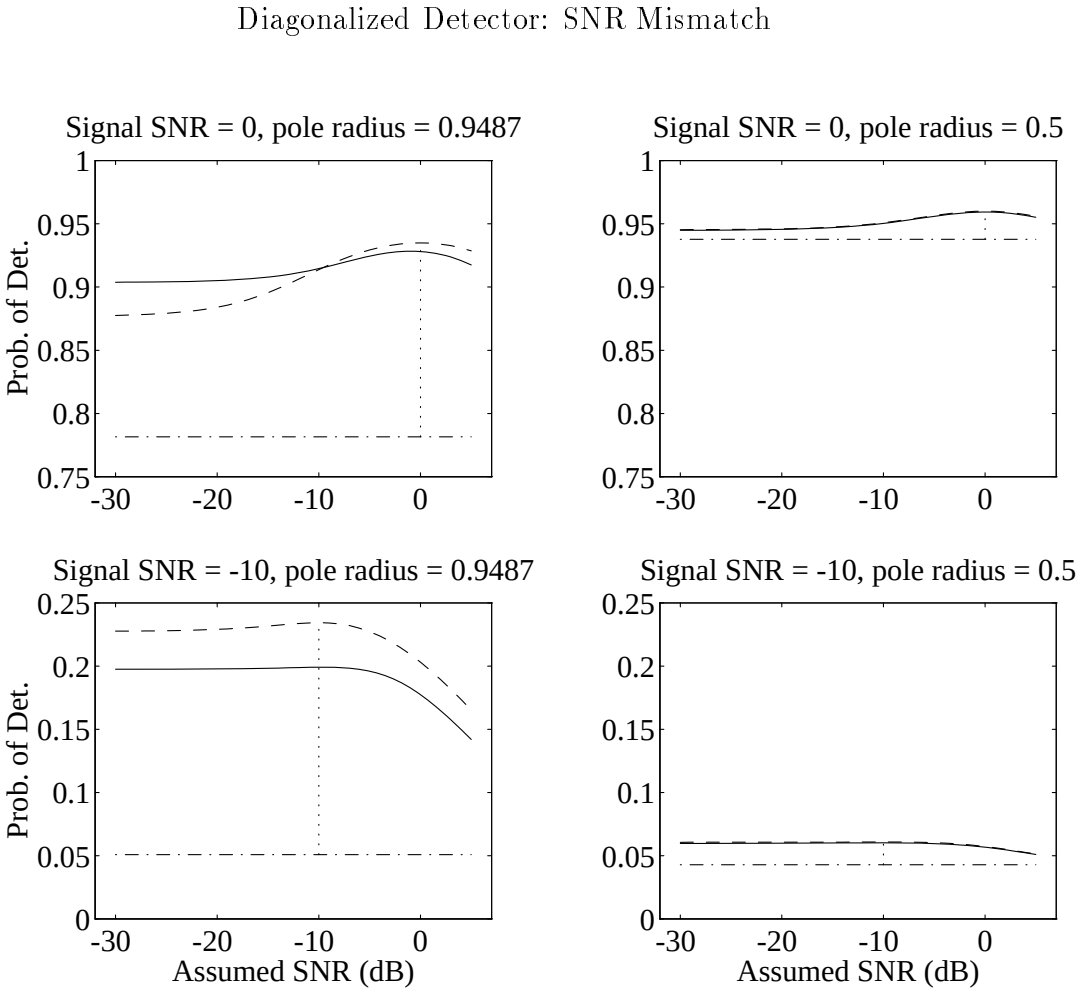

Figure 3.7: Performance comparison of three detectors with SNR mismatch on four signals. Probability of detection versus SNR assumed by the detector. The dashed line represents the optimal detector, the solid line represents the diagonalized detector, the dot-dashed line represents the energy detector, and the vertical dotted line marks the true SNR of the signal.

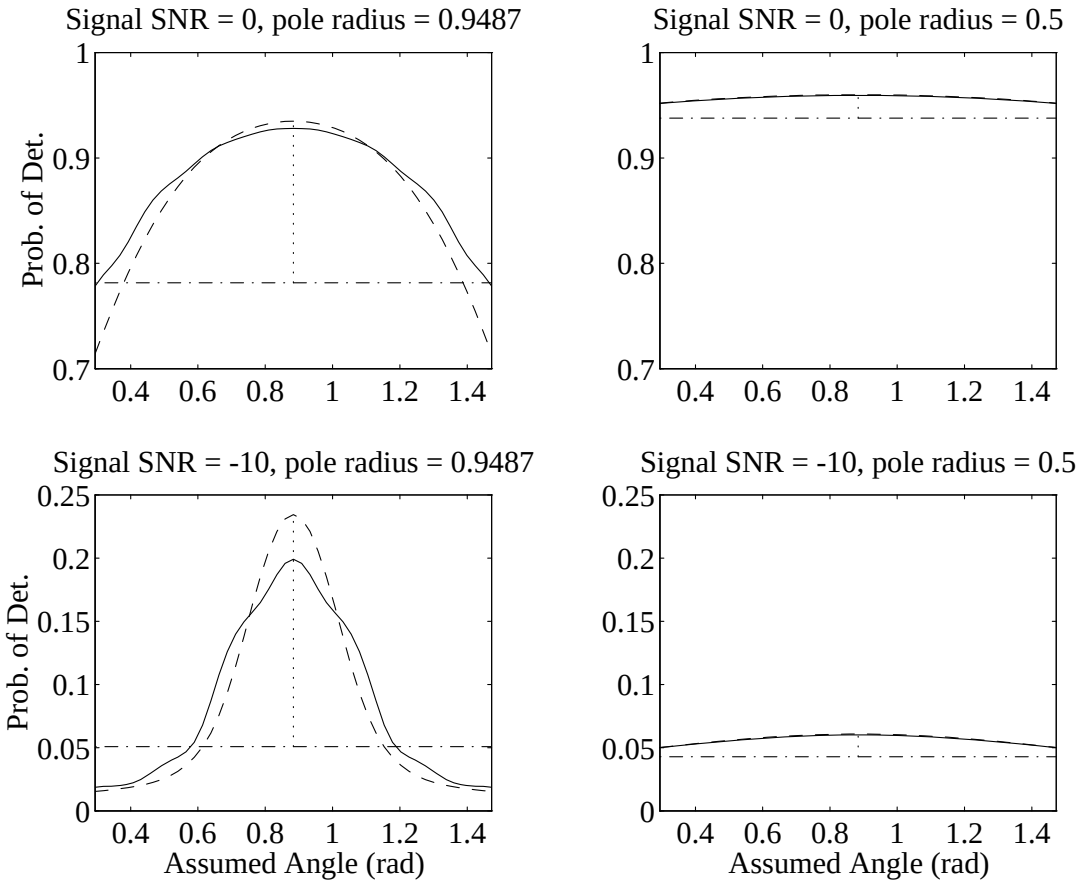

Figure 3.8: Performance comparison of three detectors with pole angle mismatch on four signals. Probability of detection versus pole angle assumed by the detector. The dashed line represents the optimal detector, the solid line represents the diagonalized detector, the dot-dashed line represents the energy detector, and the vertical dotted line marks the true pole angle of the signal.

The second interesting feature of these plots is the peak which is present in the large bandwidth plots just before the drop-off occurs near pole radius 1. This peak is also an artifact of the diagonalization, which only occurs for the larger bandwidth signal. As the pole radius approaches 1, the diagonal approximation begins to break down, and the diagonal elements of the matrix become more equally distributed instead of forming a strong peak as they should, and as the eigenvalues actually do. The evenly distributed eigenvalues are characteristic of a wideband signal, where the energy is widely distributed over the spectral bins, and thus the incorrect approximation nonetheless becomes a good detector for a wideband signal. Since the wideband signal is actually the one present in the right hand plots, we see the improved performance in this region of pole radius values.

3.1.3 Underestimated Model Order

In order to evaluate the robustness of the diagonalized detection method to model inaccuracy and colored noise, two scenarios are considered. In both cases, the signal is a second order AR process, with a pole of multiplicity two in a familiar location (angle = $4.5(2\pi)/32$, and radius = 0.9487). In case 1, the noise process is first order AR with a pole angle of $4(2\pi)/32$ and the same radius as the signal. In case 2, the pole angle is changed to $15(2\pi)/32$. To demonstrate performance in the presence of modeling errors, the diagonalized detector assuming a first order model is used. The noise in case 1 has a spectral peak at nearly the same location as the signal, making it difficult to distinguish the two. In case 2, the noise has a spectral peak much farther from the signal to demonstrate the effects of underestimating the model order more clearly.

Figure 3.10 shows the power spectral density of the signal and noise processes for both cases. The gain, and therefore the variance of each process, is adjusted to obtain the desired SNR. Figure 3.11 shows the performance of three detectors in both cases. The optimal detector shown is optimal for the second order signal and the energy

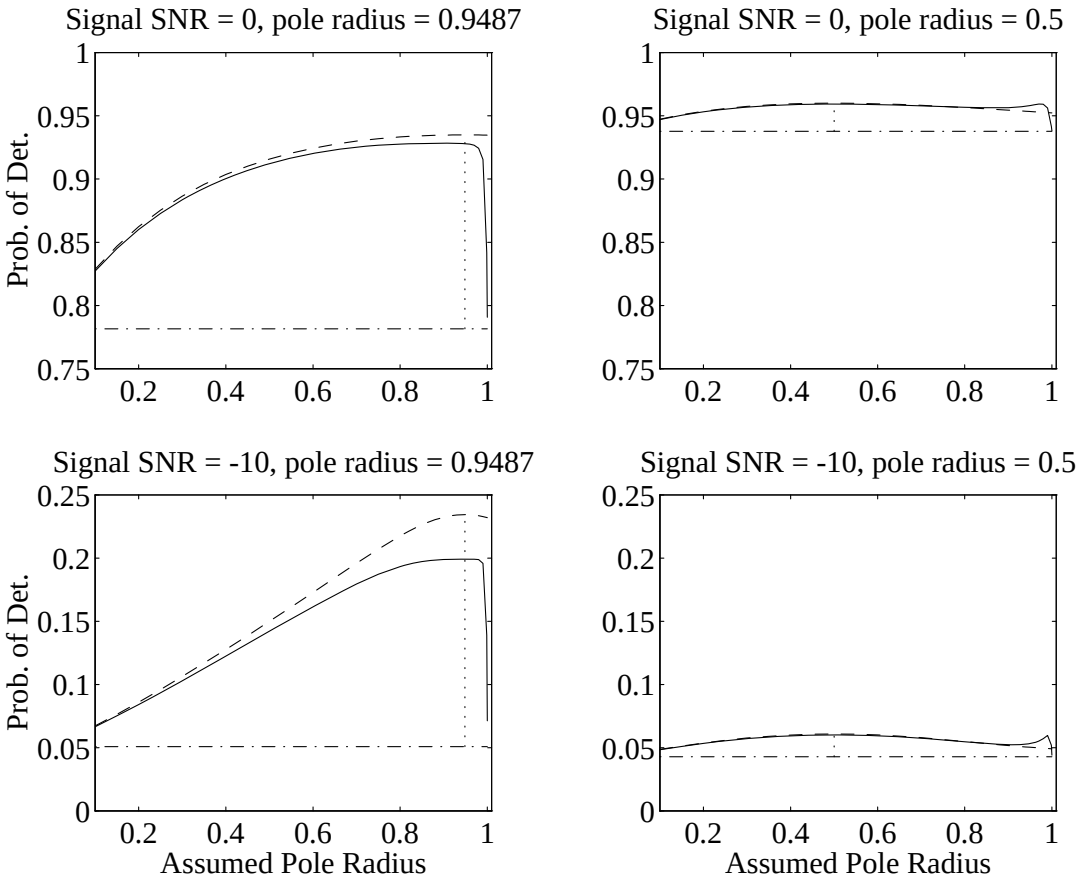

Figure 3.9: Performance comparison of three detectors with pole radius mismatch on four signals. Probability of detection versus pole radius assumed by the detector. The dashed line represents the optimal detector, the solid line represents the diagonalized detector, the dot-dashed line represents the energy detector, and the vertical dotted line marks the true pole radius of the signal.

detector is again shown for comparison.

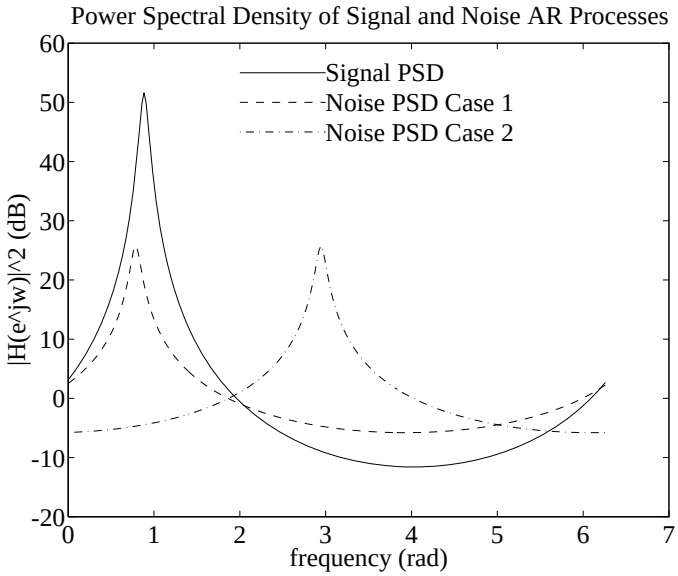

Figure 3.10: The signal PSD is second order AR with pole radius 0.9487 and angle $4.5(2\pi)/32$. The noise PSDs are first order with the same radius. The noise PSD for case 1 has a pole angle $4(2\pi)/32$, and case 2 has pole angle $15(2\pi)/32$.

Figure 3.11 shows that in case 1 the signal is difficult to detect even for the optimal detector, since the detector does not perform well at all until the SNR is above -10dB. The diagonal detector shows an improvement over the energy detector, in spite of the underestimated model order and noise which is very similar to the signal. Figure 3.11 also shows how well the diagonalized detector performs in case 2. Here, the noise is not similar to the signal, and although the energy detector's performance is not improved, both the optimal and the diagonalized detector are much more effective due to their use of spectral information. It should be noted that although some other detection and estimation techniques are significantly more complicated to perform when the noise is colored, diagonalized detection incorporates the noise covariance automatically.

Diagonalized Detector: Model Order Mismatch

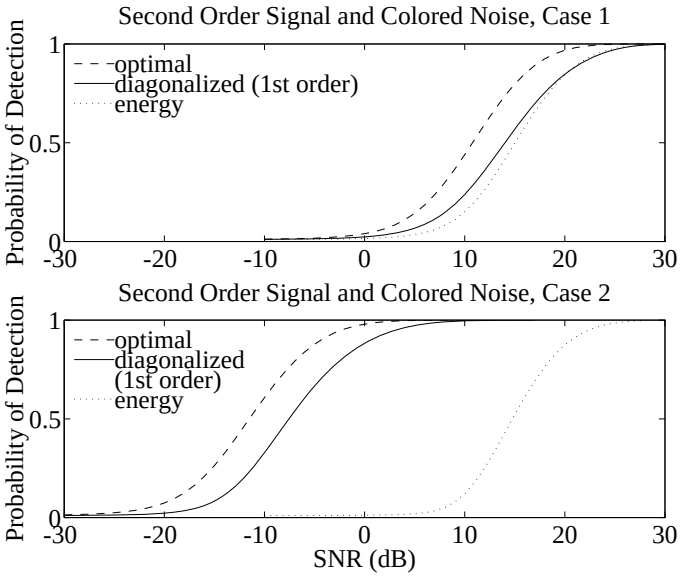

Figure 3.11: Comparison of the three detectors for the two cases of noise. In case 1 the noise pole is near the signal pole, while in case 2 the noise pole is far away in frequency.

3.2 Block Likelihood Ratio Test (BLRT)

The entire purpose of defining the BLRT is to allow the performance loss associated with treating blocks of the observation independently to be analyzed apart from the effects of the other approximations. As discussed earlier, an exact analytical approach to measuring the BLRT's performance is not available, but simulations can be carried out, and the Γ distribution approximation can also be applied. Since the Γ approximation gives reasonably accurate analytic results it is the preferable method for this application.

The BLRT detection method is discussed in detail in Section 2.2 and the detector statistic calculation is given in (2.5). Figure 3.12 shows how the performance of the BLRT decreases as the block size decreases. Just as the ROC plots shown in Section 2.5.3 show a family of curves for each detector scenario, Figure 3.12 shows three SNR values for each block size. The lowest curve for a given block size is always associated with the lowest SNR value, moving upward from -25 to -11dB. The reduction of block size clearly degrades the detector's performance, but even when the the block size is reduced from 64 to 8, the performance still does not fall as far as with a 2dB drop in SNR.

3.3 BLRT with Spectral Diagonalization

The BLRTsd detector is detailed in Section 2.3, and the detector statistic is given in (2.10). Like the BLRT, the BLRTsd is defined purely as a tool to examine the performance loss induced by the block and spectral diagonalization approximations. Since the accuracy of the spectral diagonalization approximation is at least partly dependent on block size, more degradation will occur in the BLRTsd than in the BLRT as block size decreases. Recall that the diagonalization that occurs in the spectral covariance matrix is an asymptotic effect, so when the block size is small, the approximation tends to break down.

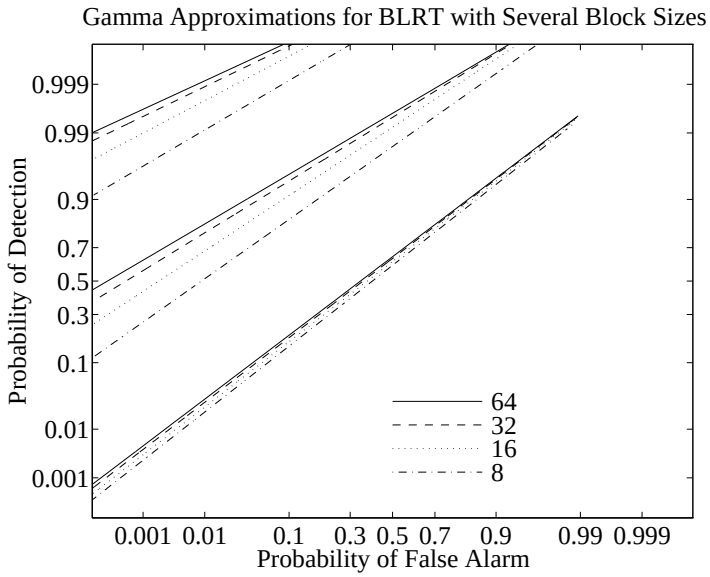

Figure 3.12: The Γ approximation for the BLRT with block sizes of 64, 32, 16, and 8, for SNR values of $-25, -15$, and -11dB.

Figure 3.13 shows how the performance of the BLRTsd detector drops with reduced block size. The SNR values and line markings shown are the same as for Figure 3.12. The BLRTsd actually appears to be less sensitive to block size than the BLRT. This is partially explained by the fact that the BLRTsd starts with lower performance. For both the BLRT and the BLRTsd, the loss in performance is significant for a block size of 8.

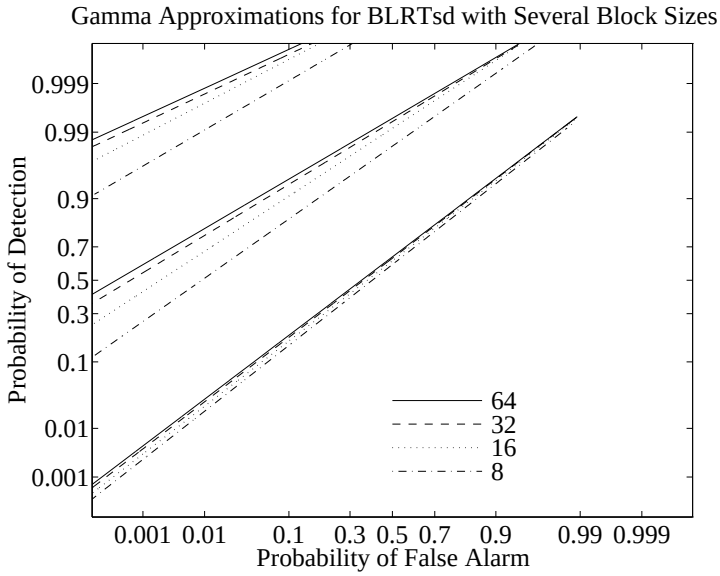

Figure 3.13: The Γ approximation for the BLRTsd with block sizes of 64, 32, 16, and 8, for SNR values of -25, -15, and -11dB.

Figure 3.14 shows a comparison of all three detectors: the LRT, the BLRT, and the BLRTsd. There is a loss associated with the approximations, but at the block size of 64 samples shown, the loss is much less than the loss from a 1dB decrease in signal strength. Since computing the BLRTsd for block sizes larger than 64 was barely practical and the BLRTsd shows performance comparable to the LRT for this block size, it was a good compromise between performance and computational load. Therefore, it was chosen as a primary size to be carried through all the simulations which are not concerned specifically with block size. At block sizes smaller than 8, the loss would be larger, and might require the elimination of the spectral diagonalization

approximation.

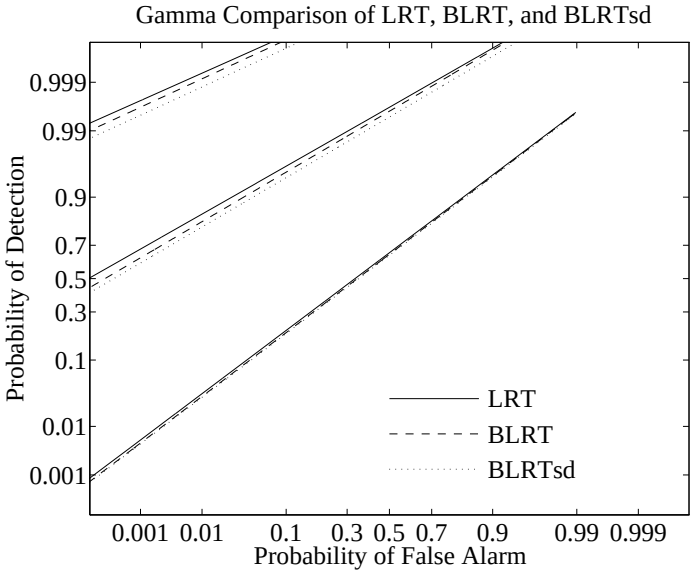

Figure 3.14: The Γ approximation for the LRT, the BLRT, and the BLRTsd detectors with SNR values of $-25, -15$, and -11dB.

CHAPTER 4

Simulation and Evaluation

This study of the new detector's performance is divided into several sections corresponding to different signal types on which the detector has been tested. Section 4.1 discusses several detectors which represent upper and lower bounds on the new detector's performance. Section 4.2 compares the performance of the GBLRTsd based on the Viterbi and the stack algorithms to that of a benchmark detector. All the simulations mentioned in this chapter are for stationary signals with white noise unless they are specifically described otherwise. Section 4.3 describes the performance of the GBLRTsd for two types of nonstationary signals with varying rates of change in frequency. These simulations use white noise, but clearly employ nonstationary signals. Section 4.4 evaluates the performance of the detector with different block sizes. Since reduced block size tends to improve the accuracy of the assumption that the signal is stationary over a block, nonstationary signals are used. In every section previous to Section 4.5 the signal gain G and pole radius r are assumed known. However, in Section 4.5, the detector's performance is evaluated when these model parameters are estimated. The pole radius used is exactly between the two values at which the search compares the likelihood, so that the result is a worst-case test. The signal is stationary and the noise is white. As an examination of the full capabilities of the GBLRTsd, this section also includes a simulation of the detector with all the model parameters unknown, a nonstationary signal, and colored noise of unknown covariance. Section section:eval:modelinacc examines the detector's robustness to model inaccuracy by considering a case where the stationary signal is actually AR(2), and another case where the stationary signal is a moving average process. Both simulations in this section use white noise. Section 4.7 compares the GBLRTsd

with a recent method from the literature which has been modified to operate on nonstationary signals.

4.1 Detector Performance Bounds

Now that the GBLRTsd detector has been defined and the approximations it incorporates analyzed, it is time to consider measuring its performance. As discussed in Chapter 1, the LRT is the optimal detector for a random signal. However, the LRT requires that signal parameters be known which are rarely known in practical cases. For instance, knowledge of the signal's strength is practically never available, and often the covariance can only be estimated in a general way, such as by using the AR(1) model suggested earlier.

The energy detector given by

$$\Lambda_E = \underline{x}'\underline{x} \tag{4.1}$$

is a very general method which assumes nothing about the signal except that it adds its energy to the noise energy. Under H_0, Λ_E is distributed as $\chi^2_{2M}/2$ if the noise is unit WGN. Under H_1, Λ_E is a generalized quadratic form which can be analyzed using the Γ approximation. Because the energy detector assumes nothing about the signal, it makes an excellent minimum standard of performance. In the search for a signal detector, if a significant improvement cannot be made over the energy detector, then the simpler solution would be preferable.

The LRT and the energy detector make good upper and lower standards of comparison for a detection algorithm. However, one may wonder whether the performance of a GLRT-based detector would fall closer to the LRT or the energy detector. To get a sharper idea of the best performance possible for a GLRT-based method, the Stationary Unknown Frequency (SUF) detector has been defined as

$$\Lambda_{SUF} = \max_{\theta} \sum_{p=0}^{P-1} \underline{\widetilde{x}}'(p)[\mathcal{D}(\widetilde{\mathbf{K}}_0(p,p))^{-1} - \mathcal{D}(\widetilde{\mathbf{K}}_1(p,p;G,r,\theta))^{-1}]\underline{\widetilde{x}}(p) \tag{4.2}$$

which is equivalent to the stationary frequency GBLRTsd statistic given in (2.11) except that G and r are assumed known. In most of the studies to follow except where specified otherwise, the GBLRTsd will be performed with known G and r to reduce computation and simplify descriptions. Also, the majority of the simulations shown below will be for a stationary signal, which simplifies calculation of the LRT.

In the studies described below, the SUF detector will have only a single advantage over the GBLRTsd methods: the *assumption* of stationary frequency. Although the frequency will be stationary where the SUF detector is involved, the GBLRTsd detectors will not make that assumption. Neither detector will have any information about the fixed frequency value, only the gain and pole radius.

The SUF detector cannot be analyzed using the Γ approximation and must be simulated. Recall from Section 2.5.2 that the use of order statistics requires that the variables being maximized, here associated with each possible frequency assumption, must be IID. Under H_0 with WGN, the values of $\Lambda_{SUF}(\theta)$ would be identically distributed, and if the values of θ were sufficiently distant, they would also be independent. Therefore, it would be possible to use the Γ approximation with order statistics to approximate Λ_{SUF} under H_0. However, under H_1, the signal would tend to contribute to some frequencies more than others, invalidating the identical distribution assumption. Therefore, the SUF detector was simulated in the studies shown below, just as were the GBLRTsd methods. The signal was an AR(1) process with a stationary pole at $\sqrt{0.9}e^{j\pi/4}$. All the detector studies in this section assume known G and r, while the LRT assumes all parameters are known. As discussed above, the SUF detector assumes the signal is stationary, while the energy detector uses no model information at all.

Figure 4.1 gives a comparison between the LRT, the SUF, and the energy detectors. The solid lines are the ROC curves for the LRT, given for SNR values of $-25, -13$ and -11dB. The dashed lines are for the SUF at the same SNR values. The dotted lines are for the energy detector, also at the same SNR values. By observing

the lowest line for each detector, it is clear that at -25dB SNR, the detectors are nearly equal, with a slight advantage for the LRT. However, observing the highest line for each detector shows that at -11dB, the LRT gives a probability of detection entirely above 0.99 for the range of false alarm probability shown, with the SUF detector slightly below. The energy detector at -11dB, however, performs much below the other two, with a probability of detection less than 0.5 for a false alarm probability of 0.01.

Figure 4.1 shows that the SUF detector falls clearly between the LRT and the energy detector. This is important because the SUF differs from the ideal GLRT only in the block and frequency approximations which were shown in Section 3.3 to have minimal effect at this block size. Therefore, the SUF represents a performance level near the upper bound for the GLRT. The GBLRTsd methods can only fall below the SUF since they assume less information about the frequency path, and in later cases, about G and r. So the SUF and the energy detectors represent reasonable upper and lower bounds for the GBLRTsd detector's performance.

4.2 GBLRT with Spectral Diagonalization

In Chapter 2, two methods for maximizing the likelihood were described: the stack algorithm and the Viterbi algorithm. The GBLRTsd detector can be implemented with either method. The two possible detectors will be referred to hereafter as the "stack detector" and the "Viterbi detector."

The simulations described below are for the case where the signal is stationary with a pole of $\sqrt{0.9}e^{j\pi/4}$ and each detector assumes the correct values for G and r, but not frequency. Figure 4.2 compares the SUF, the Viterbi, and the stack detectors. The solid lines are for the SUF at SNR values of $-25, -13$ and -11dB with 1000 trials per hypothesis. The dashed lines are for the Viterbi detector also with 1000 trials per hypothesis, and the dotted lines are for the stack detector with 400 trials

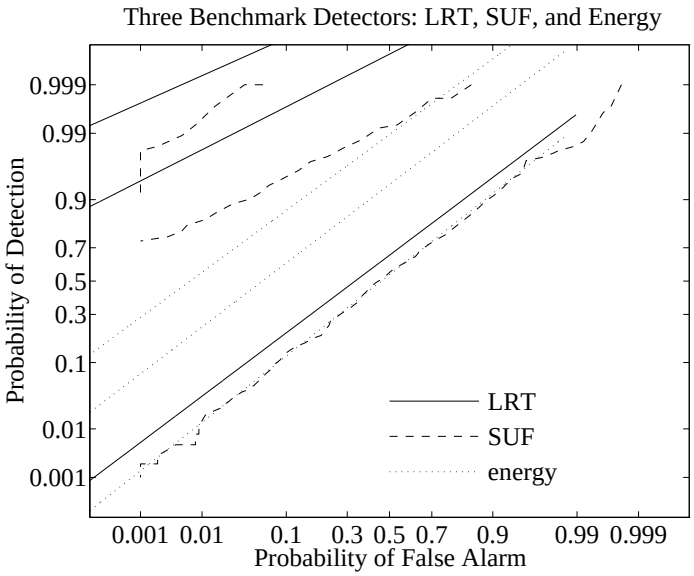

Figure 4.1: The ROC for three detectors: The likelihood ratio test, the stationary unknown frequency detector, and the energy detector. The LRT is shown at SNR values of $-25, -13$ and -11dB. The SUF is shown at the same SNR values with 1000 iterations per hypothesis. The energy detector is shown at the same SNR values.

per hypothesis, all at the same SNR values. More trials were required for the SUF and Viterbi methods because their performance curves reside at probabilities closer to one which are harder to estimate with low variance (see Section 2.6 on importance sampling). It is possible to see that the detectors are all unable to distinguish between the hypotheses at -25dB. However, at -11dB it is clear that the SUF and Viterbi detectors perform quite well, while the stack detector's performance is equivalent to the SUF at -13dB. Therefore, the performance of the stack detector -11dB is approximately 2dB poorer than the SUF method, while the Viterbi method nearly matches the SUF.

Figure 4.2 indicates that the Viterbi method for maximizing the likelihood is significantly superior to the stack method for this implementation of the stack algorithm. The stack algorithm in this case had a stack of 100 paths and a penalty value such that a path would have to have an average score 11% greater than the average path score of the same length in order to be extended. Recall that when the penalty is larger, more paths of each length tend to be examined before longer paths are accepted, since longer paths are penalized more. Although computational speed dictated the choice of these parameters, simulations have also been run with a stack size of 1000 paths and a penalty value requiring a path to be 25% above average to be extended, with no substantial gain in performance. Additionally, the Viterbi method was considerably faster to compute for the block size of 64 used for this study. One reason for the speed of the Viterbi method was that the single block likelihood values for each frequency were computed only once and the search took advantage of the single frequency transition restriction. Rather than requiring $O(V^2P)$ computations, the implementation based on the transition restriction required only $O(3VP)$ computations, a factor of $\frac{64}{3}$ improvement.

Although the stack algorithm implementation was not efficiently designed to compute all the scores once at the beginning, only its speed and not its detection performance was affected. Clearly its detection performance was inferior to that of

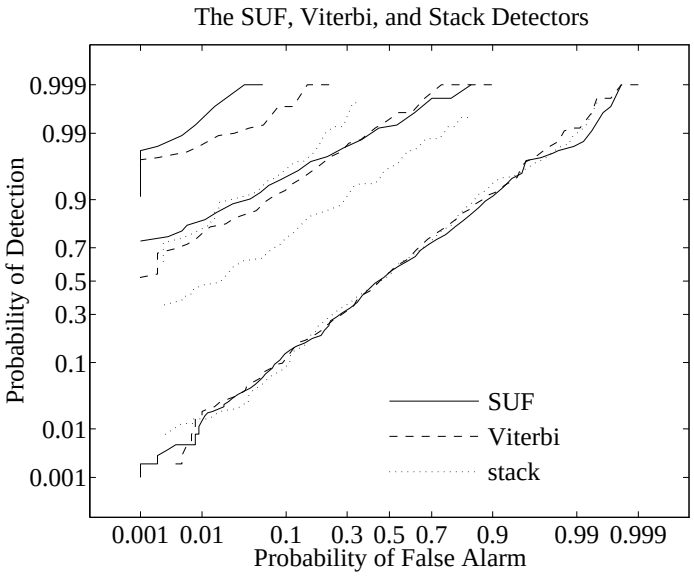

Figure 4.2: The ROC for three detectors: The SUF, the Viterbi detector, and the stack detector. The SUF and Viterbi detectors are shown at SNR values of $-25, -13$ and -11dB with 1000 trials per hypothesis. The stack detector is shown at the same SNR values and 400 trials.

the Viterbi algorithm. Early intuition in this investigation suggested that the stack algorithm would give results nearly as good as the Viterbi method (which is optimal for likelihood maximization) but would require less computational effort. Both of these ideas have been refuted. One fact which contributed to this surprise is that the transition restriction designed to improve the computational speed of the stack algorithm was at least as effective in reducing the complexity of the Viterbi algorithm. Therefore, it is prudent to take advantage of the Viterbi algorithm's optimality and its computational efficiency. The stack detector will be included in some of the figures below for completeness, but the focus of investigation will center on the Viterbi detector.

4.3 Slowly Varying Frequency Assumption

One of the original characteristics assumed in Chapter 1 for the nonstationary frequency signal was that the signal frequency must be slowly varying. This statement is intentionally open to interpretation, since the rate of frequency change allowable depends on the detector's ability to tolerate it. The studies in this section are intended to answer the question: "How is the detector's performance effected by the rate of signal frequency change?"

Two types of nonstationary signals are considered: the frequency ramp signal and the frequency step signal. First this section deals with frequency ramp signals, then the frequency step signals are considered. Each signal used in this section had a pole radius of $r = \sqrt{0.9}$ with the detectors assuming known r and G. The frequency ramp signal is a nonstationary AR(1) signal whose pole has a constant radius, but a frequency which varies linearly over the length of the signal. In each case, the signal frequency passes through π halfway through the observation, while the signal begins at a frequency half its total frequency span below π and ends at an equal frequency above π. For example, if the signal spans a frequency range of π, it would begin at

$\pi/2$ and move linearly to $3\pi/2$ at the end of the observation.

The GBLRTsd method does *not* require that the signal frequency be linear or any other particular shape, only that it vary slowly. The frequency rate of change can be specified in radians per sample. The example above has a rate of change or frequency slope of π/M, where M, the number of observation samples, is 1024 in every case in this chapter. Figure 4.3 shows the spectrogram of a frequency ramp signal with frequency slope of $\pi/1024$. The spectrogram is an unwindowed method using 64 point FFT's and no overlap, identical to the input to the GBLRTsd detectors.

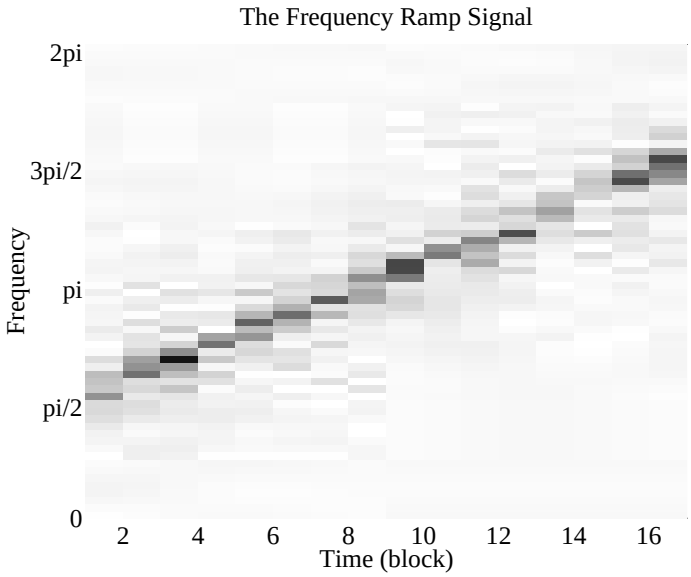

Figure 4.3: An example spectrogram of a frequency ramp signal with a frequency slope of $\pi/1024$, or a span of π.

Since the signal is assumed to be present for the entire observation, the frequency slope signal cannot be used to consider slopes larger than $2\pi/M$, since the signal would cross the 2π to 0 boundary and either be outside the sampled bandwidth or be discontinuous in frequency, depending on the interpretation. In either case, the signal would the violate basic assumptions of the problem statement and this situation is therefore avoided. In order to consider the case where the slope is infinite at a single instance, the frequency step signal is defined. The frequency step signal has the same

beginning and ending frequencies as the frequency slope signal for a given frequency span, but the frequency step signal is stationary at its beginning frequency from sample zero to sample $M/2$, where it changes frequency immediately to its ending frequency and remains there throughout the end of the observation. This signal was synthesized using an AR(1) filter with a variable pole location. The frequency step signal has a slope of zero everywhere except at sample $M/2$ where its slope is infinite. The step size, or frequency span is varied just as with the frequency ramp signal. Figure 4.4 shows a spectrogram of a frequency step signal with a span of π. The spectrogram parameters are identical to those in Figure 4.3.

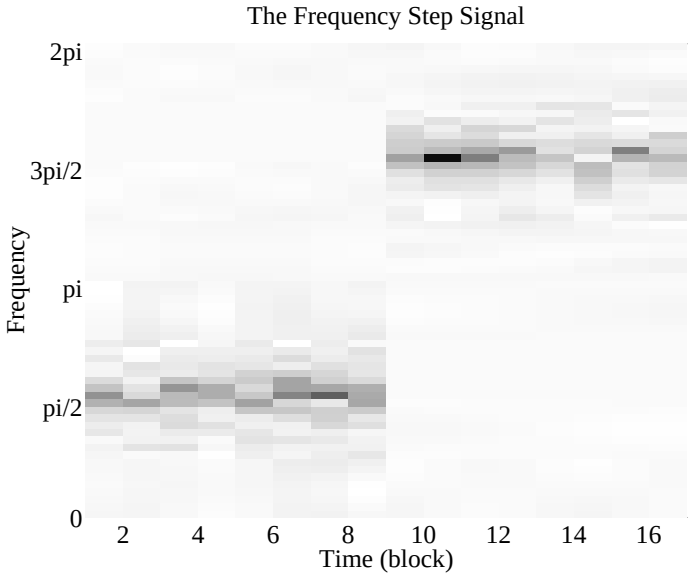

Figure 4.4: An example spectrogram of a frequency step signal with a frequency step of π.

Figure 4.5 shows the ROC curves generated by a study of the Viterbi detector operating on frequency ramp signals. The solid lines are for the signal with span of 0, or a stationary signal, at SNR values of $-25, -13$ and -11dB. The dashed lines are for a signal with a span of $\pi/4$ at the same SNR values. The dotted lines are for a signal with span $\pi/2$ at the same SNR values. The dash-dotted lines are for a signal with a span of π for the same SNR values. The detector performs very similarly for

spans of 0 to $\pi/2$, but begins to degrade when the span is π. The figure shows that the detector's performance is similar for a signal with span 0 at -13dB and a signal with span π at -11dB, which is a 2dB drop. Simulations were also performed for a span of 2π, but the level of degradation was large and detracted from the readability of the plot.

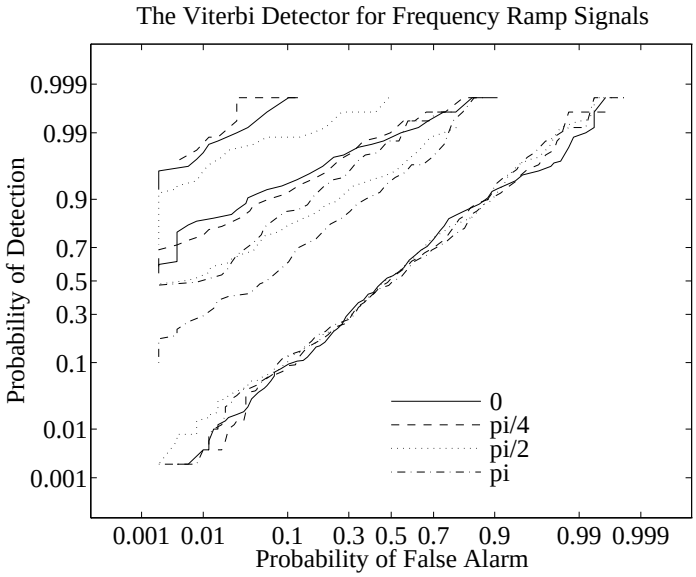

Figure 4.5: The Viterbi detector for frequency ramp signals. Frequency span values of 0, $\pi/4$, $\pi/2$, and π are shown for SNR values of $-25, -13$, and -11dB. Each curve was simulated with 500 trials per hypothesis.

Figure 4.6 shows a similar comparison for the frequency slope signal and the stack detector. The same spans and line markings are used as above but only SNR values $-25, -13$ and -9dB are shown for clarity. The -9dB group in the upper left corner shows a slight degradation in detector performance as the span increases, but not nearly as much as for the Viterbi detector above. This is because the overall performance of the stack detector is less than that of the Viterbi detector. Note that the -9dB group for the stack detector is in approximately the same position as the -11dB group for the Viterbi detector, again indicating an approximate 2dB loss associated with the stack detector.

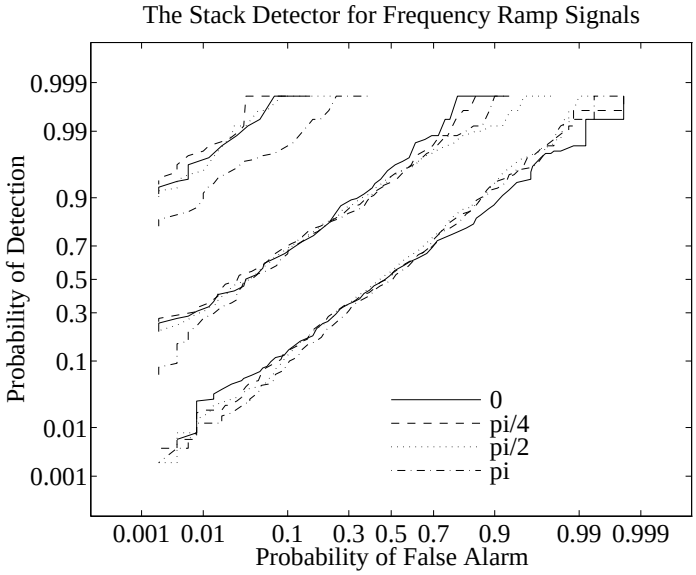

Figure 4.6: The stack detector for frequency ramp signals. Frequency span values of 0, $\pi/4$, $\pi/2$, and π are shown for SNR values of $-25, -13$, and -9dB. Each curve was simulated with 500 trials per hypothesis.

For both the Viterbi detector and the stack detector, the stationary, or 0 span, signal had slightly lower performance than that of the frequency slope signal with a span of $\pi/4$. This result is due to the nature of the Markov model used, given in (2.15). This model is clearly not ideal for a stationary signal, which should assume no possibility of transition from state to state, rather than an equal probability of trasition to any of three states. Therefore, since the slowly varying signal matches the Markov model more closely, it should be more easily detected than the stationary signal, as is in fact the case.

Figure 4.7 shows the performance of the Viterbi detector for frequency step signals. The span values and line markings are as for Figure 4.5. SNR values of $-25, -13$ and -11dB are shown for the span values of 0 to π. As expected, the Viterbi detector appears to be less tolerant of the frequency step signal than the frequency slope signal. The detector achieves good performance for the span of 0, but degrades significantly when the frequency step is introduced. The degradation is only slightly larger for a span of π than for a span of $\pi/4$, therefore the presence of the step seems to be more important than its size in this range.

The stack detector is again slightly more robust to the frequency step signals than the Viterbi detector, showing less degradation for the larger step signals, but at the expense of overall poorer performance. The span values and line markings in Figure 4.8 are the same as above, and all three SNR values are shown for each span value.

4.4 Block Length and Performance

Another major issue is the effect of block length on performance, particularly for a nonstationary signal. Ideally, a study could examine the performance of the detectors for many block sizes and many nonstationary signal slopes. However, much of this information would be redundant, and as always, resources are limited. Therefore,

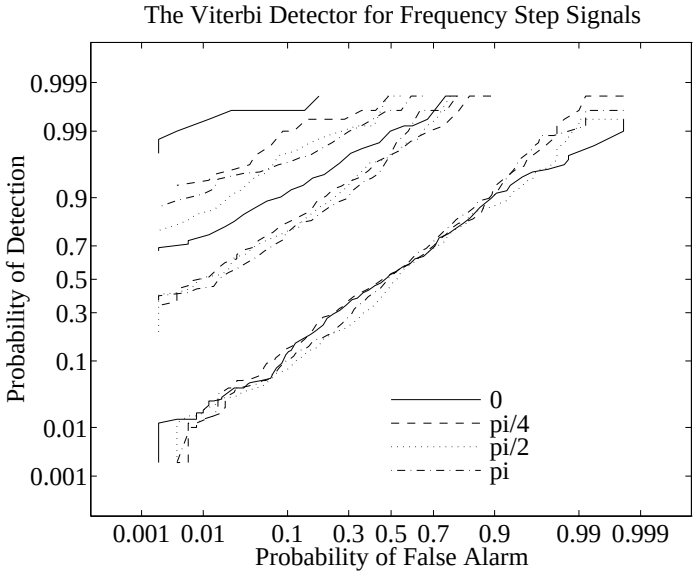

Figure 4.7: The Viterbi detector for frequency step signals. Frequency span values of 0, $\pi/4$, $\pi/2$, and π are shown for SNR values of $-25, -13$, and -11dB. Each curve was simulated with 500 trials per hypothesis.

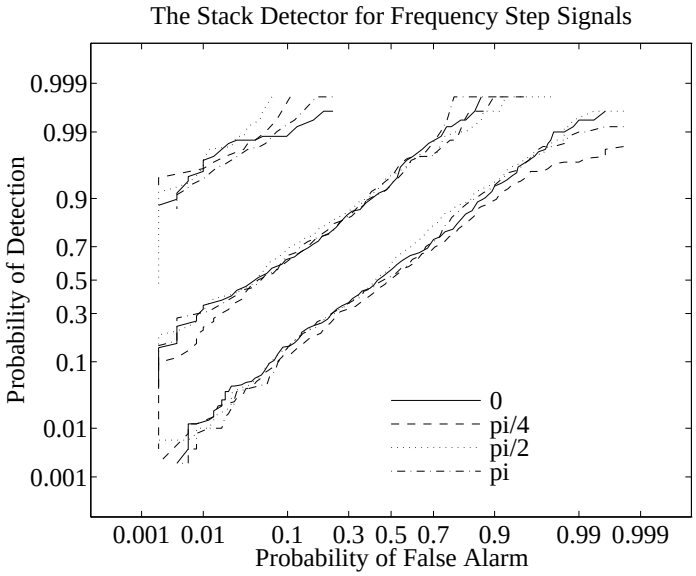

Figure 4.8: The stack detector for frequency step signals. Frequency span values of 0, $\pi/4$, $\pi/2$, and π are shown for SNR values of $-25, -13$, and -11dB. Each curve was simulated with 500 trials per hypothesis.

this study was limited to just a two signal types at several block sizes. First, recall the BLRT and BLRTsd analysis discussed in Sections 3.2 and 3.3. The Γ approximation was used there to demonstrate the effect of reduced block size on the BLRT and BLRTsd. For a stationary signal, the loss associated with these approximations would be the only loss and further discussion would be unnecessary. Unfortunately, nonstationary signals introduce another effect on performance. Since the structure of the detector assumes the signal is varying slowly enough to be considered stationary over a single block, signals with frequency slopes might be expected to benefit from shorter block sizes, since the block stationary assumption is relatively more accurate.

In these studies the signal is nonstationary with a pole radius of $r = \sqrt{0.9}$ and the detectors once again assume known G and r. The figures based on the Viterbi detector $-25, -13$, and -11dB, while figures based on the stack detector show SNR values of $-25, -13$, and -9dB. The stack detector figures use a higher SNR for the top curve because the performance is poorer and the probabilities are not as close to one. Figure 4.9 shows a comparison of the performance of the Viterbi detector for block sizes of 64, 32, 16, and 8 for a frequency slope signal with a span of $\pi/4$. Performance loss with decreasing block size is significant at the higher SNR values. The block size 8 curve at -11dB is close to the block size 64 curve at -13dB, which indicates a total 2dB loss. Figure 4.10 shows a comparison of the performance of the stack detector for block sizes of 64, 32, and 16 for a frequency slope signal with a span of $\pi/4$. The stack detector suffers much less with decreasing block size, but starts with lower performance.

Figure 4.11 shows a comparison of the performance of the Viterbi detector for block sizes of 64, 32, 16, and 8 for a frequency step signal with a span of $\pi/4$. The change with decreasing block size for the frequency step signal is much less than with the frequency ramp signal, except for the block size 8 curve. Again, the detector begins at a lower standard due to the frequency step (recall Figure 4.7). Figure 4.12 shows a comparison of the performance of the stack detector for block sizes of 64,

32, and 16 for a frequency step signal with a span of $\pi/4$. The stack detector again shows only a small effect from decreasing block size. One effect which is not shown in the figure is the increase in computational time for smaller block size. Since the observation length is the same, when the block size decreases, the number of blocks increases. Therefore, the search effort required of the iterative stack algorithm also increases. For block size 8, the computation time increased to the point that the simulations were impractical with the current implementation of the stack algorithm.

Although it was postulated that smaller block sizes would improve performance for the nonstationary signals, the loss from the block independence approximation appears to outweigh any gains produced by allowing less frequency change per block. This result might be reversed for a signal with a larger frequency slope.

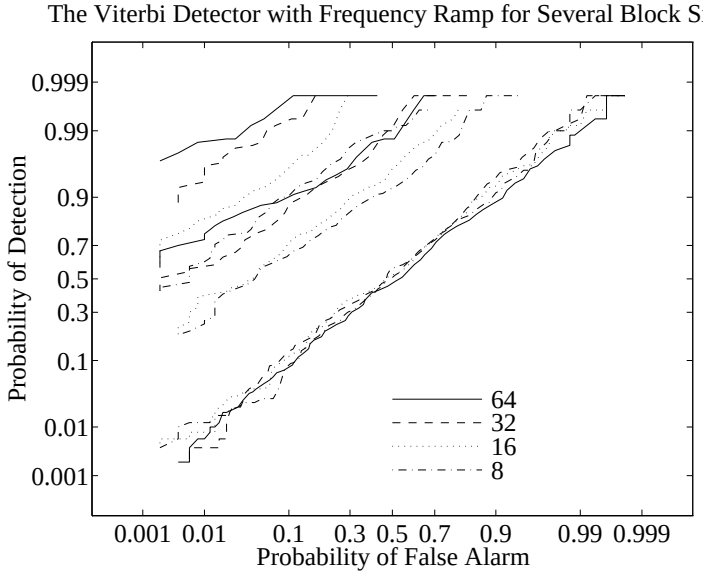

Figure 4.9: The Viterbi detector computed with block sizes of 64, 32, 16, and 8, for a frequency slope signal with a span of $\pi/4$. The SNR values shown are $-25, -13$, and -11dB. Each curve was simulated with 500 trials per hypothesis.

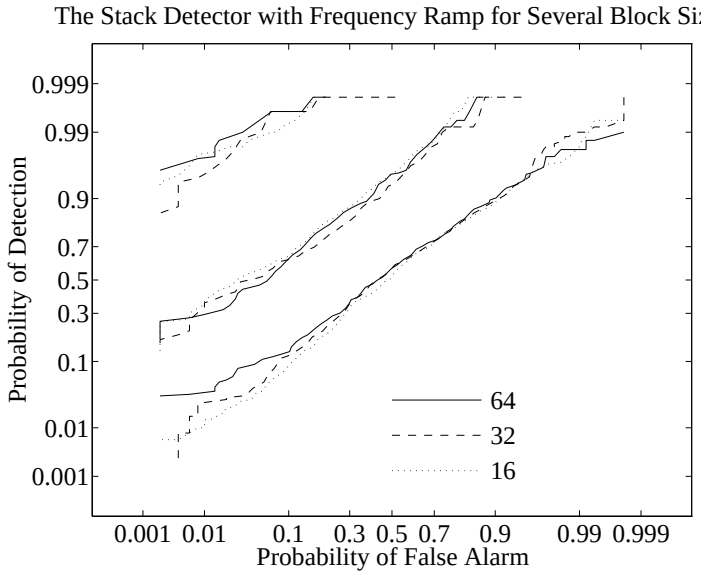

Figure 4.10: The stack detector computed with block sizes of 64, 32, and 16, for a frequency slope signal with a span of $\pi/4$. The SNR values shown are $-25, -13$, and -9dB. Each curve was simulated with 500 trials per hypothesis.

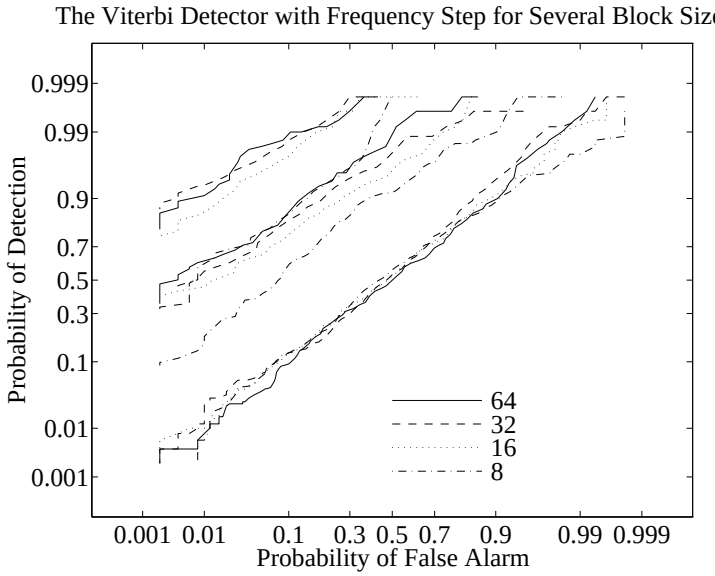

Figure 4.11: The Viterbi detector computed with block sizes of 64, 32, 16, and 8, for a frequency step signal with a span of $\pi/4$. The SNR values shown are $-25, -13$, and -11dB. Each curve was simulated with 500 trials per hypothesis.

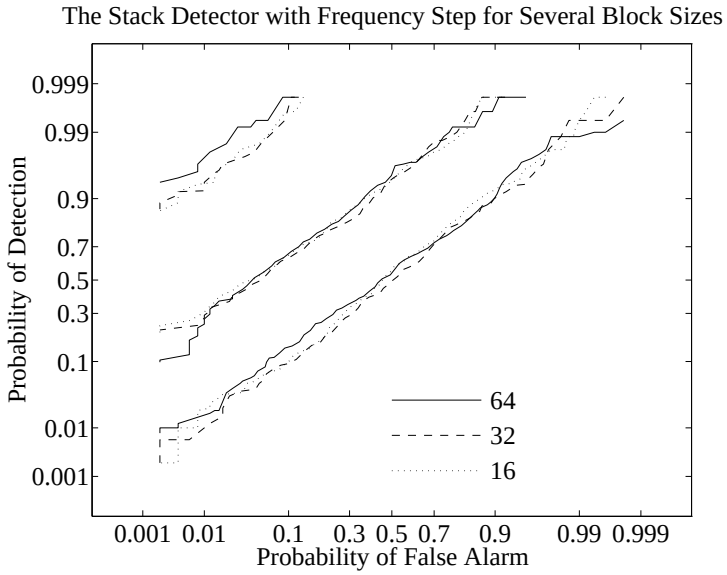

Figure 4.12: The stack detector computed with block sizes of 64, 32, and 16, for a frequency step signal with a span of $\pi/4$. The SNR values shown are $-25, -13$, and -9dB. Each curve was simulated with 500 trials per hypothesis.

4.5 Unknown G and r, colored noise

A number of the Viterbi detector's properties can be demonstrated and evaluated using known G and r but unknown $\theta(p)$. However, the detector's performance in cases where these parameters are unknown is quite important, since G is almost never known, and r may not be known. The stack detector has been shown to perform poorly compared to the Viterbi detector and therefore has not been pursued to the point of performing detection with unknown G and r values. The primary difficulty in evaluating the Viterbi detector's performance with unknown parameters is that the detector performance is strongly dependent on the value of r. When r is small, the signal is nearly white and the SUF and energy detectors become nearly equivalent. This effect is caused by the fact that the energy detector is the optimal detector for a white signal in white noise. When r is near 1, however, there is a large difference between the SUF detector and the energy detector, simply because the estimation of frequency provides more information about the signal.

Fortunately, the single block studies in Chapter 3 show that the likelihood value is not strongly affected when the estimates of G and r are inaccurate. It is possible to use this information to limit the number of values of G and r which are examined so that the likelihood search is not dramatically expanded. Since the Viterbi algorithm must be run in its entirety for each combination of G and r values that is considered, the number of values allowed has a large effect on the computational cost of the method. Experience has shown that only two values each of G and r need to be examined to produce detector performance near that of the SUF detector. This approach is equivalent to dividing the parameter space into the categories of "high" and "low" SNR and "large" and "small" pole radius. The values have been chosen based on the data shown in Section 3.1.2, to be $r_u \in \{0.45,\ 0.9\}$ and the G_t values associated with SNR's of -20 and -15dB. The G_t values are linear gains over a constant noise power of 1 and are intended to be near the lower and upper knees of the detector curve.

With two possible values of each parameter, the Viterbi algorithm must be calculated four times for the entire observation, where the final score is the maximum likelihood from the four possible combinations. Figure 4.13 shows a comparison of the SUF with known G (varying with SNR) and r, and the Viterbi algorithm using the parameter values above to estimate G and r. The actual signal had a roughly worst case r value of 0.675, which is exactly in between the two guesses for r used by the Viterbi method. The simulations show that even with no *a priori* knowledge of the model parameters, the Viterbi method performs nearly as well as the SUF detector which assumes the G and r values are known. There is more loss for the highest SNR value of -11dB than for the lower SNRs, but the decrease in performance is only about 1dB for this roughly worst case value of r.

To demonstrate the full capability of the Viterbi detector, Figure 4.14 shows the ROC curves for the Viterbi detector running on a frequency ramp AR(1) signal with unknown G and r and colored noise of unknown covariance compared with the Viterbi detector for the same signal in WGN with known r and G. Since the signal is nonstationary with span π, the noise covariance can be estimated from the observed signal based on the assumption that the noise is stationary. This estimate is not optimal, but if the signal is known to be nonstationary, it is an effective way to deal with stationary noise of unknown covariance. Figure 4.15 shows a spectrogram of an example of the observed signal plus noise at -11dB. The noise covariance is an

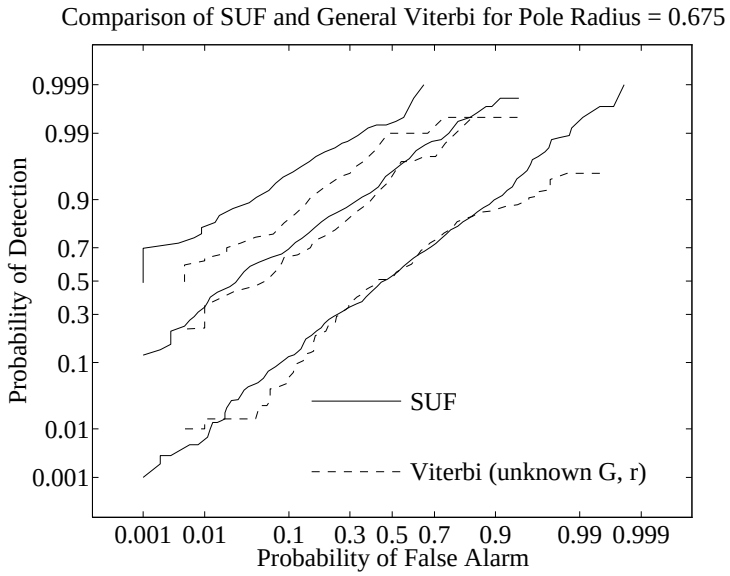

Figure 4.13: The upper bound for the smaller pole radius of 0.675 is lower than for a larger pole radius. The curves shown are for SNR values of $-25, -13$, and -11dB. The SUF curve is for 1000 iterations per hypothesis and the Viterbi curve is for 200 iterations per hypothesis.

AR(10) process with AR polynomial coefficients

$$\underline{a} = \begin{bmatrix} 1.0000 & & \\ 0.9793 & + & 0.7477i \\ -0.0246 & + & 0.6187i \\ -0.5117 & - & 0.5048i \\ -0.0291 & - & 0.8683i \\ 0.4844 & - & 0.4285i \\ 0.3927 & + & 0.1490i \\ 0.0818 & + & 0.2581i \\ -0.2237 & + & 0.1024i \\ -0.2030 & - & 0.1583i \\ -0.1118 & - & 0.1658i \end{bmatrix}. \tag{4.3}$$

These coefficients were derived from observed sonar data in an effort to simulate real-world background noise. The detector performance for unknown colored noise with estimated covariance is much better than for the same signal with white noise. This is because the noise covariance tends to concentrate the signal power in a relatively small portion of the spectrum so that the effective SNR is higher in other portions where the signal is present. However, even with this effect taken into account the detector's performance is impressive since the signal is barely visible at -11dB, where the detector performs almost perfectly.

4.6 Model Inaccuracy

One of the desired characteristics of the detection algorithm was robustness to model inaccuracy. Since the detector assumes an AR(1) model for the signal regardless of its actual origin, it is important that the performance be reasonable even when the signal is not AR(1), as will likely be the case in real applications.

A first step toward examining the detector's robustness to model inaccuracy is to

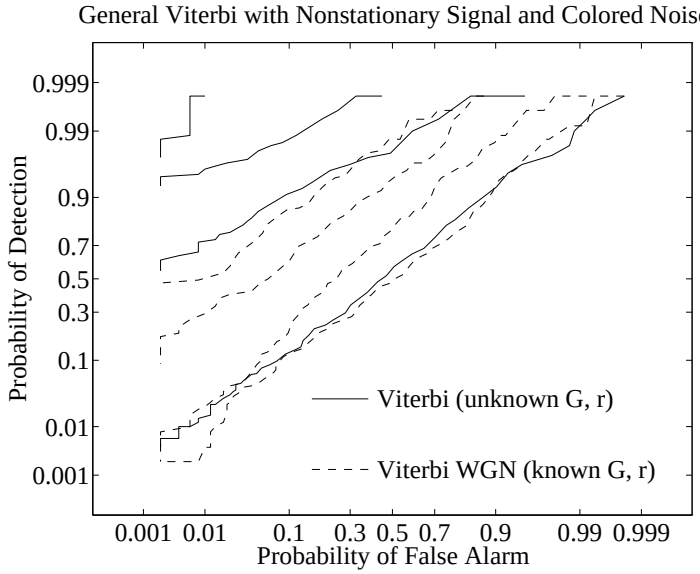

Figure 4.14: The general Viterbi detector was operated on a frequency ramp signal with a span of π and colored noise of *unknown* covariance. The covariance was estimated using the assumption that the noise was stationary and the signal was not. The Viterbi WGN curve shown is for the same signal in white Gaussian noise assuming known G and r. Both detectors were tested using 500 trials per hypothesis and are shown at SNR values of $-25, -15, -13$, and -11dB.

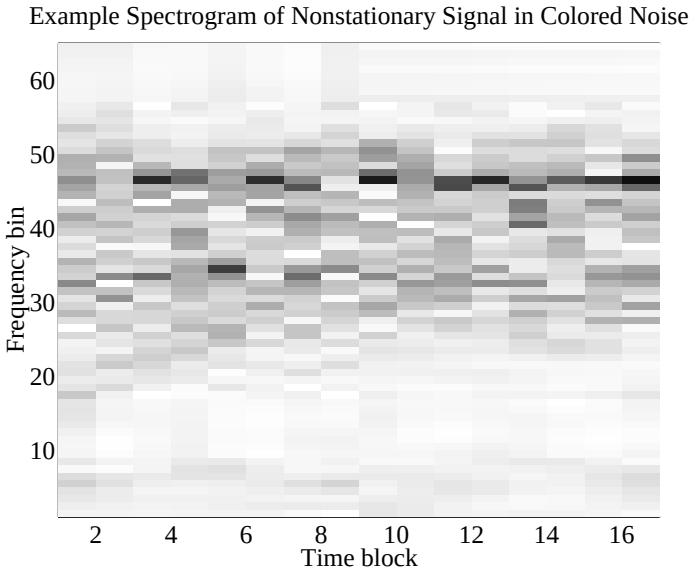

Figure 4.15: The unwindowed spectrogram of the 1024 point observation using nonoverlapping 64 point FFTs is shown for a frequency slope signal with a span of π at -11dB SNR. The signal is barely perceptible in the colored noise even though the detector performs well at this SNR. See Figure 4.3 for an example of the signal without noise, and Figure 1.1 for an example of the harmonic noise from which the AR noise spectrum shown above was derived.

present the detector with an AR(2) signal with two poles in the same location where the AR(1) pole would have occurred. This allows the continued use of "known" G and r values even though the model is no longer correct. The Viterbi detector which assumes known (although incorrect) values of G and r and the general Viterbi detector which estimates the values of G and r can be compared to each other and the SUF detector can be modified to make the correct AR(2) assumptions and again serve as a near upper bound. Figure 4.16 shows a comparison of the SUF AR(2) detector, along with the Viterbi detector assuming known G and r, and the general Viterbi detector with no parameter assumptions. The AR(2) signal in this study uses a stationary pole of multiplicity two at $\sqrt{0.9}e^{j\pi/4}$. The figure shows that the general Viterbi detector is nearly equivalent to the SUF, while the Viterbi detector making the incorrect "known" assumptions is also quite close. The SUF AR(2) required 1000 trials per hypothesis rather than the 500 required by others because the probabilities in its highest curve are very close to 1. Even though 1000 trials is only enough to accurately estimate probabilities in the 0.99 range, the estimates obtained should be reasonable. However, the confidence interval is larger for such small values than for the other curves on the plot.

A second step in examining the detector's robustness to model inaccuracy is to present the detector with a stationary signal which is a low order Gaussian signal, but derives its PSD from a moving average (MA), or finite impulse response (FIR), system. Once again, the SUF detector can be modified to represent a near upper bound for the Viterbi detector by allowing it to make use of the FIR spectral information known in advance, but without knowledge of the actual frequency.

The fact that the SUF can be modified to make use of an FIR model is an indication of how extremely flexible the GBLRTsd structure is with respect to the model involved. Given an exact or even an approximate knowledge of the FIR or AR spectral shape of the signal, the GBLRTsd can be operated on the assumption that the spectral shape is constant over frequency. Nonstationary signals that meet

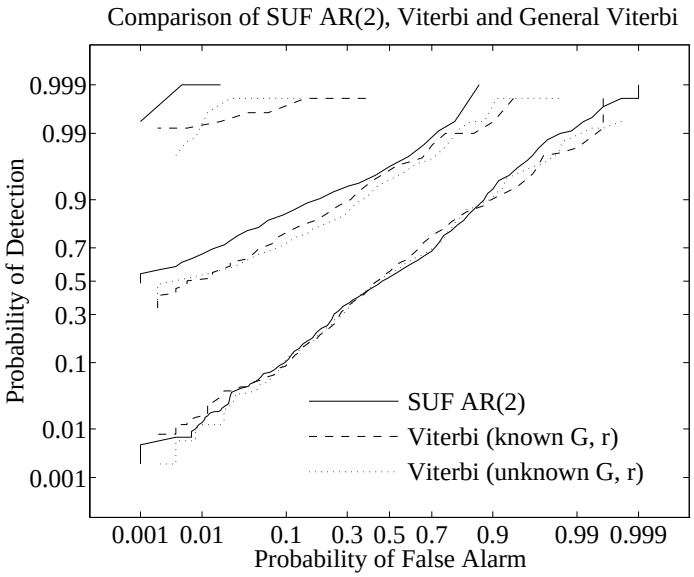

Figure 4.16: The SUF AR(2) is the near-optimum detector for a stationary AR(2) signal with unknown frequency. The Viterbi detector assuming known G and r for the AR(1) case is shown along with the stack detector making the same assumptions. The general Viterbi detector estimates the values of G and r which gives more freedom to compensate for the model inaccuracy. The SNR values shown are $-25, -15$, and -11dB. Each curve was simulated with 500 trials per hypothesis, except for the SUF AR(2), which required 1000 trials.

the assumptions can be detected nearly optimally. Even if the spectral shape is not constant over frequency, if a spectral model is known, it can be incorporated into the GBLRTsd.

The purpose of this study is not to demonstrate the GBLRTsd's flexibility, but its robustness. Therefore, the Viterbi detector's assumption was that the signal was an AR(1) process, even though it was actually FIR. The FIR filter used to generate the signal process had coefficients

$$h(k) = \begin{bmatrix} 0.0234 & - & 0.0776i \\ 0.0925 & - & 0.1479i \\ 0.2325 & - & 0.1368i \\ 0.3032 & - & 0.0000i \\ 0.2325 & + & 0.1368i \\ 0.0925 & + & 0.1479i \\ 0.0234 & + & 0.0776i \end{bmatrix}. \tag{4.4}$$

These filter coefficients were generated by creating an order 6 real FIR lowpass filter with a pass band of 0.05 of the Nyquist frequency using the "window" method. The Hilbert transform of the result was taken to create the complex filter $h(k)$ shown above. The $h(k)$ was frequency shifted to create a bandpass filter which was used to shape a WGN process to create the final signal. The autocorrelation of the WGN process filtered by $h(k)$ is given by $r(k) = h(k) * h(-k)$ where $*$ represents convolution. Figure 4.17 shows the magnitude of the 64 point FFT of $h(k)$. Clearly $h(k)$ meets the requirements of the problem statement in Chapter 1 stating that the signal must have a single dominant spectral peak.

Figure 4.18 shows the SUF and the Viterbi detector's performance using a signal with FIR autocorrelation $r(k)$. Recall that the SUF detector has been modified to make the correct assumptions about the FIR signal's covariance, while the Viterbi detector has not. Since the correct values of G and r are not defined, the general Viterbi detector was used which estimates values of G and r in addition to $\theta(p)$ to

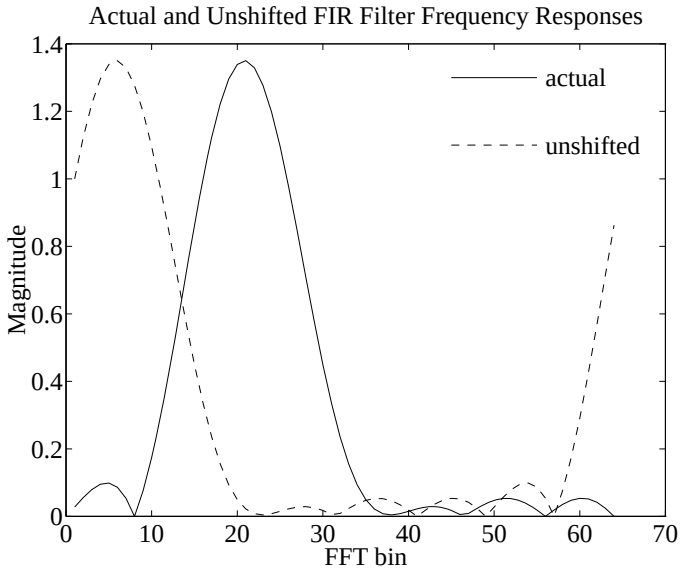

Figure 4.17: The two curves show the FFT magnitude of the actual filter used to generate the signal in Figure 4.18 and of the complex lowpass filter "unshifted" which was used to generate the bandpass filter.

maximize the likelihood. The figure makes it clear that the Viterbi detector is quite robust to the incorrect model assumptions, giving performance only slightly below that of the near-optimal SUF FIR detector. Pathological signals could be invented to further challenge the robustness of the detector, such as a bandlimited white signal, but it is clear from the example above that the AR assumption does not preclude the detector's use on signals which are not truly AR in origin.

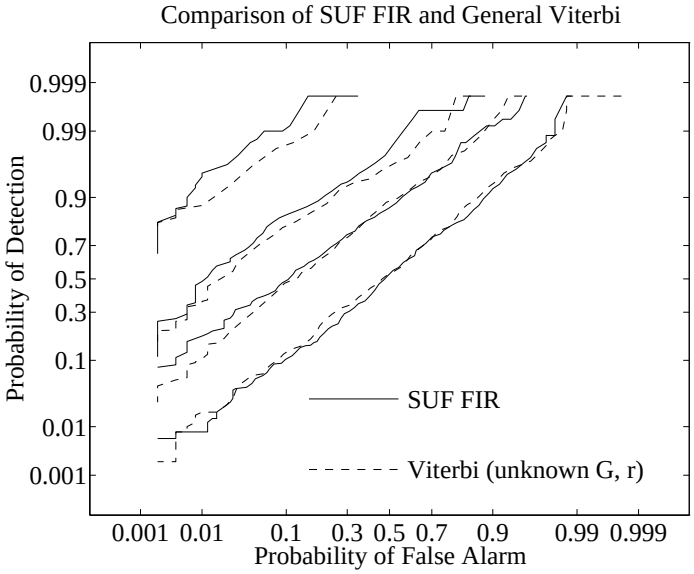

Figure 4.18: Four SNR values are shown, $-25, -15, -13$, and -11dB. Each curve was generated with 500 trials per hypothesis.

4.7 Comparison with a Modern Method

In order to discuss the relative merits of the Viterbi detector as compared to methods in the literature, it is necessary to modify these algorithms to be address the problem at hand. Kay's AR signal detector [7] discussed in Section 1.4, is the most convenient to convert to the problem addressed in this thesis. For a stationary signal, Kay's method is similar to the SUF in that it estimates a stationary frequency to calculate the likelihood. The modified version of Kay's detector to be described will be referred

to as the Modified Kay Method (MKM) detector.

Kay's method assumes that the noise is white with unknown power, and that the signal under H_1 can be written entirely as an AR process. By simply assuming the noise power is known, Kay's detector can be applied to the white noise case where the hypotheses are defined as

$$\begin{aligned} \mathbf{K}_0 &= \quad \sigma\mathbf{I} \\ \mathbf{K}_1 &= \quad \mathbf{K}_s(r,\theta) \end{aligned} \tag{4.5}$$

so that $\mathbf{K}_1$ does not have an additive white noise component. This set of hypotheses can be normalized to unit variance white noise by defining $\sigma = 1$ and allowing the model parameter r to compensate for the signal gain. This simplification makes the problem more similar to the problem defined in (1.10), although there is still no noise only component in $\mathbf{K}_1$. The resulting detector would compute

$$l_1(\underline{x}|r,\theta) - l_0(\underline{x}) \underset{H_0}{\overset{H_1}{\gtrless}} 0 \tag{4.6}$$

where r and θ are computed by the AR modeling technique, and G is unnecessary. Although this detector is no longer equivalent to the one described in [7], it makes use of the same basic ideas.

Nonstationary signals are not considered in [7]. To deal with a nonstationary signal, the block stationary assumption must be invoked. The conversion of the LRT to the BLRT allowed the frequency parameter to be vary from block to block, but be considered stationary within a block. Similarly, the MKM detector must be modified to process the signal in blocks so that the frequency model of the signal can be varied form block to block. Although block processing allows the signal to be nonstationary, it also limits the data available to the stationary model estimator which tends to reduce the accuracy of the model estimates.

Since the purpose of using an AR estimate to find the signal model for each block is to avoid evaluating every possibility, the Viterbi algorithm cannot be applied in the MKM detector, and the estimate for each block will depend entirely on the data

in that block. Since the signal will actually be an AR(1) signal, the MKM detector will be restricted to AR(1) models. The AR estimation works with the time domain signal and multiple frequencies do not need to be evaluated, so the MKM detector can operate without the spectral diagonalization approximation and its detection statistic is given by

$$\Lambda_{MKM} = \sum_{p=0}^{P-1} \underline{x}'(p)[\mathbf{K}_0(p,p)^{-1} - \mathbf{K}_1(p,p;r,\theta(p))^{-1}]\underline{x}(p) \tag{4.7}$$

where the values of r and $\theta(p)$ are determined by the autocorrelation, or maximum entropy AR modeling method. Although the simplicity of this detector is pleasing, the lack of such approximations as the spectral diagonalization approximation mean that full matrices must be used in the calculation of Λ_{MKM}. In particular, each term of 4.7 requires the inverse of $\mathbf{K}_1(r,\theta)$ and both r and θ can vary over the block index, so that it is impossible to compute $\mathbf{K}_1$ in advance. This is a substantial limitation on block size, since the matrix inversion computation is $O(L^3)$. The detector remains practical if not desirable to compute with $L = 64$ as in the current investigation. The simulations performed in this section use the familiar stationary AR(1) signal with a pole at $\sqrt{0.9}e^{j\pi/4}$.

Figure 4.19 shows a comparison of the performance of the MKM and the general Viterbi detector for a stationary signal in WGN, along with a comparison of the performance of the SUF detector. The MKM detector does a much poorer job than either GLRT-based detector. Since the general Viterbi detector performs nearly as well as the SUF detector, it is a substantial improvement over the approach based on Kay's method. This result is not surprising since Kay's method was not originally suggested for application to nonstationary signals. The MKM detector suffers mainly from the small amount of data available for estimating the AR parameters due to the necessity of block processing.

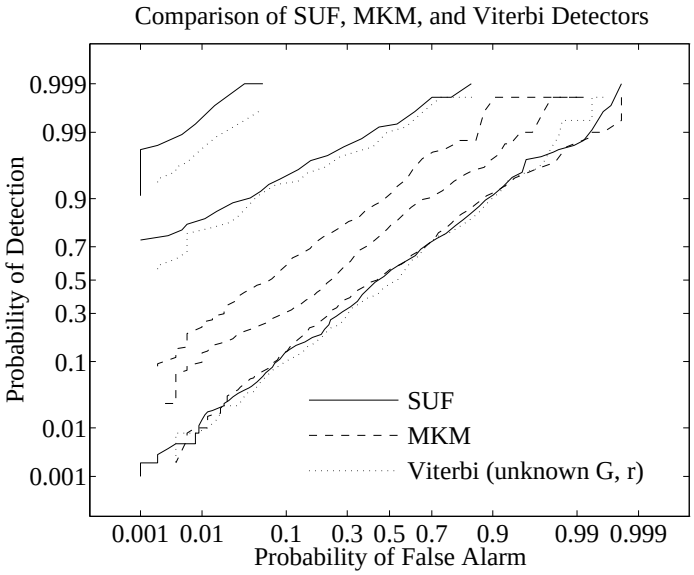

Figure 4.19: SNR values of $-25, -13$, and -11dB are shown. The SUF detector was simulated with 1000 trials per hypothesis and the MKM with AR(1) and the general Viterbi methods were simulated with 500.

CHAPTER 5

Conclusions

The goal of this thesis was to find a computationally efficient algorithm to detect nonstationary low order Gaussian random signals with unknown parameters in noise of known covariance. A secondary goal was to build a theoretical framework for the optimal detection of nonstationary signals based on maximum likelihood estimation of signal parameters. This problem was successfully approached by making a series of simple approximations to the GLRT to reduce its computational expense.

5.1 Contributions

A number of problems were dealt with in the development of this thesis, including the determination of the usefulness of the spectral diagonalization approximation and the related analysis of the robustness of the likelihood function to AR parameter mismatch using numerical characteristic function inversion. Because the computational demands of the inversion methods were so great, an approximate analytical approach was developed using the Γ distribution. This approximate approach was used to examine the usefulness of the block approximation and the spectral approximation in the form of the BLRT and BLRTsd detectors. For cases where simulation was required, an importance sampling approach was developed for improving the simulation efficiency for the random signal detector.

To deal with unknown parameters, two methods for maximizing the likelihood were examined with the result that the Viterbi algorithm was a significant improvement over the stack algorithm. A detection algorithm for the nonstationary low order Gaussian random signals was implemented and tested under variations of several pa-

rameters including signal type and block length. The Viterbi detector was found to perform nearly as well as the SUF detector in virtually all cases for those signal types which met the low order, slowly varying frequency assumptions. Finally the detector was compared to a modified version of Kay's detection algorithm [7], and found to provide superior performance.

In summary, the major contributions of this thesis are

- a numerical investigation of the spectral diagonalization approximation
- a numerical investigation of parameter mismatch
- the adaptation of the Γ distribution approximation for detector analysis
- an approximate analysis of the BLRT and BLRTsd detectors
- the successful development of the nearly optimal GBLRTsd detector based on the Viterbi algorithm for detection of low order nonstationary Gaussian random signals
- the simulation, evaluation, and comparison of the GBLRTsd detector with the MKM detector.

The GBLRTsd detector using Viterbi maximization has been shown to be computationally efficient with near-optimal performance and is quite robust to model inaccuracy. Therefore, this thesis has contributed to the state of the art a useful new solution to the nonstationary random signal detection problem.

5.2 Future Extensions

This thesis has resolved a number of issues discussed above, but it has also created a number of new problems to be investigated in the future. These new problems fall into two broad classes: investigations of potential modifications of the Viterbi detector which were outside the scope of this thesis, and extensions of the work given

here which would allow direct application to some of the applications mentioned in Chapter 1.

There are many issues which for reasons of practicality were considered but not pursued as possible modifications of the Viterbi algorithm. The possibility of applying FIR signal models or higher order AR models was mentioned in Section 4.6. Such approaches would allow the application of the GBLRTsd method to more general signal types instead of only low order signals, but would require either more specific knowledge of the signal or a much greater effort in estimating the model parameters.

A more general frequency transition model could be applied, such as the Gaussian one used in [8]. Such a model would increase the cost of computing the Viterbi algorithm, but would also allow for better performance on signals with rapid frequency transitions or large slopes. Again, such a model would require more knowledge of the signal or extensive training to obtain an accurate transition model. A possible compromise would be to employ a clipped Gaussian or a binomial distribution so that transitions outside a certain distance would have zero probability. This method would allow for some of the efficiency of the transition restriction without the oversimplification of a single bin restriction.

The stack algorithm has two parameters which can be adjusted to improve its accuracy and also increase its computational burden. The penalty factor determines how a path of a particular length is compared to paths of other lengths. When it is increased, the longer paths must have much higher scores than average to be extended. The stack length determines how many old paths are kept for possible extension. These parameters were set at values which were as large as practical for this investigation, and further work was made unnecessary by the Viterbi algorithm. However, the stack algorithm has flexibility in the search for unknown G and r parameters that the Viterbi algorithm does not have. The stack algorithm does not have to be run separately for each parameter combination but can directly compare paths with different G and r values and different lengths. If the GBLRTsd were

applied to a problem where the non-frequency parameters were more sensitive, the stack algorithm would have a distinct advantage on this basis.

The block independence assumption implicit in the BLRT does not have a strong effect on the detector's performance for a block size of 64 but the block size cannot be lowered significantly without serious performance loss. A possible tradeoff could be made to allow smaller block sizes. This tradeoff would involve calculating the detector statistic for overlapping blocks and using a few matrix identities to reconstruct a full observation covariance which was not block diagonal, but was formed from the overlapping diagonal blocks. Such a method would allow a full observation covariance which did not assume the blocks were independent, and therefore a smaller block size could potentially be tolerated. Of course, the immediate drawbacks include a doubling of the required computation due to the overlapping block structure.

Another method of avoiding the block independence assumption would be to build an estimated full observation covariance by interpolating a diagonal band from the fixed values obtained from each block [32]. While this method has the potential to be more accurate than block independent methods, it also requires the construction of a large matrix, and computations using the full observation. Other extensions of the GBLRTsd method could be suggested which are too numerous to mention, but the suggestions above should provide a number of possibilities to readers wishing to apply the method to a slightly different problem.

As for direct application of this method, it is fairly rare that the application consists of operating on a fixed block of data in which the signal is known to either be present for the entire time (H_1) or not present at all (H_0). The question which most often arises is "when did the signal begin and end?" Therefore, the issue of time of arrival should be considered. Suppose an application requires that a data stream be examined for nonstationary signals 8 blocks in length and the blocks are short enough that only block boundaries need to be considered for starting and ending points. An inefficient approach would be to calculate the detector statistic sequentially on

8 blocks at a time, shifting in the first block and shifting out the last each time. Fortunately, the Viterbi method can be applied in a much more efficient way. Since the likelihood score for each frequency is calculated from the first block to the last in a cumulative way, it would be simple to save the last 8 sets of Viterbi scores and continually subtract the score from block $(n-8)$ from the score for block n. This technique would allow the detector to run continuously, while only computing a new score for one block at a time rather than for 8 at a time. Of course, such issues as rescaling and multiple parameter combinations would also have to be considered.

In any real application, better performance could be obtained by making better use of *a priori* information, such as a known range for r for the signal of interest. Analysis of a particular class of data such as biomedical signals or passive sonar data could potentially yield such useful information and produce an even more effective detector.

APPENDIX A

The Covariance Matrix of a Nonstationary CGRV

Two nonstationary signals were used in this thesis, but the covariance matrices were not required due to the assumption of a stationary frequency for each block. The nonstationary covariance can be derived in several ways. The central assumption of almost any derivation would be that the signal was generated by filtering WGN with a filter whose frequency response was variable. A fully general derivation can be performed using a time-varying impulse response and convolution, however this method is quite involved, and the generality is not required here. A simpler method is to directly analyze the output function of a time-varying filter. Although this method is inexact in the sense that the output never reaches the exact steady state after the initial start up, the initial condition effect is one that is encountered in the signal synthesis problem. Therefore, the direct output derivation is actually more accurate for the problem at hand.

Unfortunately, neither derivation method would produce results which are particularly useful for analysis by hand or for much intuitive insight. Nevertheless, the direct derivation can at least be used to numerically verify the behavior of a nonstationary covariance matrix, and also for demonstration of a few properties such as the Hermitian symmetry and lack of Toeplitz form.

To begin the derivation, assume the signal $y(n)$ is generated by a complex AR(1) filter with a time-varying pole value, and a *causal* WGN input. Causality of the input is important to make the summations finite and therefore tractable, but it also causes

the above mentioned initial effects. The output of the filter is

$$y(n) = -a(n)y(n-1) + x(n) \tag{A.1}$$

where $x(n)$ is the WGN input, and a varies with the sample index. This is the actual signal arrangement used to synthesize the frequency slope and frequency step signals.

The output $y(n)$ can be rewritten entirely in terms of x and a as

$$y(n) = \sum_{l=0}^{n} x(l) \prod_{k=l+1}^{n} [-a(k)]. \tag{A.2}$$

The covariance matrix can be written in terms of its elements so that

$$\mathbf{K}_y(n,p) = E[y(n)y'(p)]. \tag{A.3}$$

Then

$$E[y(n)y'(p)] = E\left[\sum_{l=0}^{n} x(l) \prod_{k=l+1}^{n} [-a(k)] \sum_{q=0}^{p} x'(q) \prod_{m=q+1}^{p} [-a'(m)]\right] \tag{A.4}$$

which simplifies greatly since $E[x(n)x'(p)] = 0$ for $n \neq p$ to give

$$E[y(n)y'(p)] = E\left[\sum_{l=0}^{\min(n,p)} |x(l)|^2 \prod_{k=l+1}^{n} [-a(k)] \prod_{m=l+1}^{p} [-a'(m)]\right] \tag{A.5}$$

It is simple to show from this result that the covariance matrix is Hermitian, and it should be obvious that it will not be Toeplitz in general. The initial value effect encountered with this method can be dealt with by running the filter with a constant pole frequency over several time-constants before allowing the frequency to change. This method effectively simulates the steady state of the filter before generating usable data. The same procedure may be used to evaluate (A.5) at a steady state.

Using (A.5), the covariance of a 100 sample frequency slope signal with a span of 2π has been calculated and the imaginary part is shown in Figure A.1. The figure shows the Hermitian symmetry of the covariance as well as the fact that it is not Toeplitz. The transition at the center of the matrix is caused by the change from the positive imaginary region to the negative imaginary region.

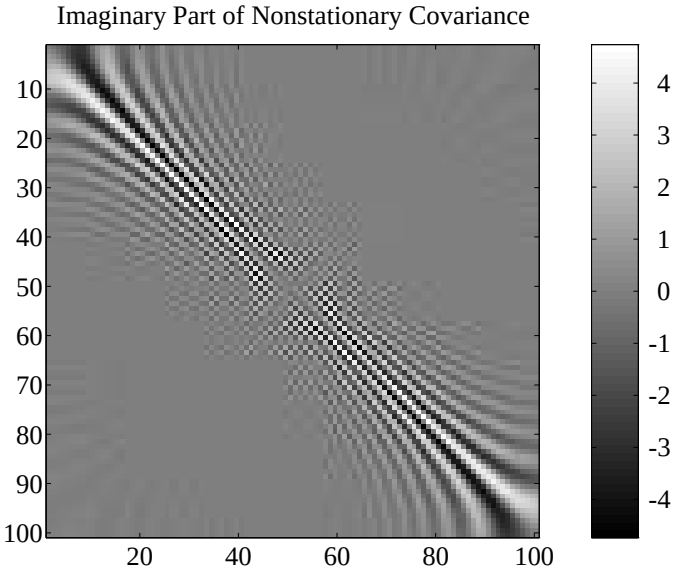

Figure A.1: The imaginary part of the nonstationary covariance matrix shown as an intensity plot. The frequency slope signal covariance for a span of 2π was calculated using a 100 sample settling time before the 100 sample covariance shown.

Bibliography

[1] J. Chen *et al.*, "Adaptive spectral analysis of cutaneous electrogastric signals using autoregressive moving average modelling," *Medical & Biological Engineering and Computing*, vol. 28, pp. 531–536, November 1990.

[2] S. Haykin and A. Steinhardt, eds., *Adaptive Radar Detection and Estimation.* New York: John Wiley and Sons, 1992.

[3] J. C. Hassab, *Underwater Signal and Data Processing.* Boca Raton, Florida: CRC Press, Inc., 1989.

[4] H. L. Van Trees, *Detection, Estimation,and Modulation Theory*, vol. 1. New York: John Wiley and Sons, 1968.

[5] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing.* Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1993.

[6] E. L. Lehmann, *Testing Statitistical Hypotheses.* New York: John Wiley and Sons, 1959.

[7] S. Kay, "Broadband detection of signals with unknown spectra," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-85*, pp. 1263–1265, 1985.

[8] R. L. Streit and R. F. Barrett, "Frequency line tracking using hidden Markov models," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, pp. 586–598, April 1990.

[9] R. F. Barrett and D. A. Holdsworth, "Frequency tracking using hidden Markov models with amplitude and phase information," *IEEE Transactions on Signal Processing*, vol. 41, pp. 2965–2976, October 1993.

[10] P. O'Shea and B. Boashash, "Instantaneous frequency estimation using the cross Wigner-Ville distribution with application to non-stationary transient detection," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Proceesing, ICASSP-90*, pp. 2887–2890, 1990.

[11] S. Lin and D. J. Costello, Jr., *Error Control Coding.* Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1983.

[12] W. T. Padgett, D. B. Williams, and E. J. Holder, "Simplified quadratic detection in the frequency domain for signals with unknown parameters." unpublished, April 1994.

[13] L. L. Scharf, *Statistical Signal Processing.* Reading, Massachusetts: Addison-Wesley, 1991.

[14] N. L. Johnson and S. Kotz, *Continuous Univariate Distributions*, vol. 2. Boston: Houghton Mifflin, 1970.

[15] A. H. Feiveson and F. C. Delaney, "The distribution and properties of a weighted sum of chi squares," Tech. Rep. TN D-4575, NASA, Houston, February 1968.

[16] N. L. Johnson and S. Kotz, "Tables of distributions of positive definite quadratic forms in central normal variables," *Sankhya: The Indian Journal of Statistics: Series B*, vol. 30, pp. 303–314, December 1968.

[17] G. G. Tziritas, "On the distribution of positive-definite gaussian quadratic forms," *IEEE Transactions on Information Theory*, vol. IT-33, pp. 895–906, November 1987.

[18] N. C. Beaulieu, "An infinite series for the computation of the complementary probability distribution function of a sum of independent random variables and its application to the sum of rayleigh random variables," *IEEE Transactions on Communications*, vol. 38, pp. 1463–1474, September 1990.

[19] R. B. Davies, "Numerical inversion of a characteristic function," *Biometrika*, vol. 60, no. 2, pp. 415–417, 1973.

[20] J. P. Imhof, "Computing the distribution of quadratic forms in normal variables," *Biometrika*, vol. 48, no. 3/4, pp. 419–426, 1961.

[21] S. O. Rice, "Distribution of quadratic forms in normal variables - evaluation by numerical integration," *SIAM J. Sci. Stat. Comput.*, vol. 1, no. 4, pp. 438–448, 1980.

[22] G. E. P. Box, "A general distribution theory for a class of likelihood criteria," *Biometrika*, vol. 36, pp. 317–346, December 1949.

[23] K. H. Biyari and W. C. Lindsey, "Statistical distributions of Hermitian quadratic forms in complex Gaussian variables," *IEEE Transactions on Information Theory*, vol. 39, pp. 1076–1082, May 1993.

[24] R. V. Hogg and A. T. Craig, *Introduction to Mathematical Statistics*. New York: Macmillan, fourth ed., 1978.

[25] G. Orsak and B. Aazhang, "On the theory of importance sampling applied to the analysis of detection systems," *IEEE Transactions on Communications*, vol. 37, pp. 332–339, April 1989.

[26] G. C. Orsak, "A note on estimating false alarm rates via importance sampling," *IEEE Transactions on Communications*, vol. 41, pp. 1275–1277, September 1993.

[27] W. T. Padgett and D. B. Williams, "A simple importance sampling strategy for random signal detectors." submitted to *IEEE Transactions on Communications*, April 1994.

[28] S. Kotz, N. L. Johnson, and D. W. Boyd, "Series representations of distributions of quadratic forms in normal variables: I. central case," *Annals of Mathematical Statistics*, vol. 38, pp. 823–837, June 1967.

[29] J. Abate and W. Whitt, "The Fourier-series method for inverting transforms of probability distributions," *Queueing Systems*, vol. 10, pp. 5–87, January 1992.

[30] R. Lugannani and S. O. Rice, "Distribution of the ratio of quadratic forms in normal variables - numerical methods," *SIAM J. Sci. Stat. Comput.*, vol. 5, no. 2, pp. 476–488, 1984.

[31] F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proceedings of the IEEE*, vol. 66, pp. 51–83, January 1978.

[32] R. Lacoss, December 1993. personal communication.

[33] M. Abramowitz and I. A. Stegun, eds., *Handbook of Mathematical Functions*. Washington, D. C.: National Bureau of Standards, 1972.

[34] R. A. Altes, "Detection, estimation, and classification with spectrograms," *Journal of the Acoustical Society of America*, vol. 67, pp. 1232–1246, April 1980.

[35] W. H. Beyer, ed., *CRC Standard Mathematical Tables*. Boca Raton, Florida: CRC Press, Inc., 28th ed., 1987.

[36] P. Flandrin, "A time-frequency formulation of optimum detection," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, pp. 1377–1384, September 1988.

[37] P. Flandrin, "Time-frequency receivers for locally optimum detection," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-88*, pp. 2725–2728, 1988.

[38] D. R. Fuhrmann and M. I. Miller, "On the existence of positive-definite maximum-likelihood estimates of structured covariance matrices," *IEEE Transactions on Information Theory*, vol. 34, pp. 722–729, July 1988.

[39] Y. Grenier, "Time-dependent ARMA modeling of nonstationary signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-31, pp. 899–911, August 1983.

[40] W. Martin and P. Flandrin, "Wigner-Ville spctral analysis of nonstationary processes," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-33, pp. 1461–1470, December 1985.

[41] R. J. Muirhead, *Aspects of Multivariate Statistical Theory.* New York: John Wiley and Sons, 1982.

[42] W. H. Press *et al.*, *Numerical Recipes in C.* Cambridge University Press, second ed., 1992.

[43] L. R. Rabiner and B. H. Juang, "An introduction to hidden Markov models," *IEEE ASSP Magazine*, pp. 4–16, January 1986.

[44] L. L. Scharf and B. D. Van Veen, "Low rank detectors for gaussian random vectors," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-35, pp. 1579–1582, November 1987.

[45] V. V. Veeravalli and H. V. Poor, "Quadratic detection of signals with drifting phase," *Journal of the Acoustical Society of America*, vol. 89, pp. 811–819, February 1991.

[46] D. B. Williams and D. H. Johnson, "Robust maximum-likelihood estimation of structured covariance matrices," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-88*, pp. 2845–2848, 1988.

Vita

Wayne Thomas Padgett was born in Syracuse, New York on September 27, 1965. He received the Bachelor of Electrical Engineering degree with highest honors from Auburn University in 1989. He also received a cooperative education certificate for completing a nine quarter work program with Intergraph Corp. in Huntsville, Alabama. After a graduate Summer internship with the Texas Instruments Image Processing Lab in Lewisville, Texas, he entered graduate school at Georgia Institute of Technology. From 1989 to 1994 he completed his graduate work, receiving support from the National Science Foundation Fellowship, and assistantships from the Office of Naval Research, and Georgia Tech Research Institute. He was also employed part-time by Atlanta Signal Processors, Inc. as an Application Engineer during this period. In 1990, he received the Master of Science degree and in 1994 he received the Doctor of Philosophy degree from the Georgia Institute of Technology. His research interests include digital and statistical signal processing, wideband and transient signals, detection and estimation, and pattern recognition.

Detection of Low Order Nonstationary Gaussian Random Processes

Wayne Thomas Padgett

115 pages
Directed by Dr. Douglas B. Williams

This thesis develops a new method for detecting low order Gaussian random signals in noise when the parameters of the signal are unknown and the frequency of the signal is nonstationary and slowly varying. A low order signal is defined here to to be a signal that can be modeled with a small number of autoregressive (AR) model coefficients and has only one significant spectral peak. The new detector is appropriate for applications where the signal is wideband and the time-bandwidth product is large such as passive sonar, radar and biomedical signal detection.

Based on the low order signal restriction, the signal is modeled as the output of a first order AR system with a time varying pole angle. This signal model is applied to the computation of the optimal likelihood ratio test when the signal parameters are known and extended to an approximation of the generalized likelihood ratio test (GLRT) when the signal parameters must be estimated. The resulting detector makes use of several types of computationally efficient approximation, including block processing and a spectral diagonalization of the data covariance matrix. The stack algorithm and the Viterbi algorithm are investigated as maximization techniques for producing the maximum likelihood estimates required by the GLRT.

The statistical behavior of the approximations is determined analytically for both a single block of the data using numerical characteristic function inversion, and for the full observation using a distribution approximation. Simulations are used to verify and extend the analytical results. Using comparisons with a near-optimum detector as an upper bound and the energy detector as a lower bound, the new

detection algorithm is shown to give nearly optimal performance for a variety of nonstationary and non-AR signals. Recent methods from the literature are examined for applicability to this problem, and the most suitable method is compared to the new detector.

The new detection algorithm is found to be computationally efficient with near optimal performance, and is robust to model inaccuracy. The thesis concludes with a discussion of extensions of the theoretical framework of the approximate GLRT to other signal models and applications.