Dissertation

Detection of Multiple Correlated Time Series and its Application in Synthetic Aperture Sonar Imagery

Submitted by

Nicholas Harold Klausner

Department of Electrical and Computer Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2014

Doctoral Committee:

Advisor: Mahmood R. Azimi-Sadjadi
Co-Advisor: Louis L. Scharf

Ali Pezeshki
Dan Cooley

Abstract

Detection of Multiple Correlated Time Series and its Application in Synthetic Aperture Sonar Imagery

Detecting the presence of a common but unknown signal among two or more data channels is a problem that finds its uses in many applications, including collaborative sensor networks, geological monitoring of seismic activity, radar, and sonar. Some detection systems in such situations use decision fusion to combine individual detection decisions into one global decision. However, this detection paradigm can be sub-optimal as local decisions are based on the perspective of a single sensory system. Thus, methods that capture the coherent or mutual information among multiple data sets are needed. This work considers the problem of testing for the independence among multiple (≥ 2) random vectors. The solution is attained by considering a Generalized Likelihood Ratio Test (GLRT) that tests the null hypothesis that the composite covariance matrix of the channels, a matrix containing all inter and intra-channel second-order information, is block-diagonal. The test statistic becomes a generalized Hadamard ratio given by the ratio of the determinant of the estimate of this composite covariance matrix over the product of the determinant of its diagonal blocks.

One important question in the practical application of any likelihood ratio test is the values of the test statistic needed to achieve sufficient evidence in support of the decision to reject the null hypothesis. To gain some understanding of the false alarm probability or size of the test for the generalized Hadamard ratio, we employ the theory of Gram determinants to show that the likelihood ratio can be written as a product of ratios of the squared residual from two linear prediction problems. This expression for the likelihood ratio leads quite simply to the fact that the generalized Hadamard ratio is stochastically equivalent to a product

of independently distributed beta random variables under the null hypothesis. Asymptotically, the scaled logarithm of the generalized Hadamard ratio converges in distribution to a chi-squared random variable as the number of samples used to estimate the composite covariance matrix grows large. The degrees of freedom for this chi-squared distribution are closely related to the dimensions of the parameter spaces considered in the development of the GLRT. Studies of this asymptotic distribution seem to indicate, however, that the rate of convergence is particularly slow for all but the simplest of problems and may therefore lack practicality. For this reason, we consider the use of saddlepoint approximations as a practical alternative for this problem. This leads to methods that can be used to determine the threshold needed to approximately achieve a desired false alarm probability.

We next turn our attention to an alternative implementation of the generalized Hadamard ratio for 2-dimensional wide-sense stationary random processes. Although the true GLRT for this problem would impose a Toeplitz structure (more specifically, a Toeplitz-block-Toeplitz structure) on the estimate of the composite covariance matrix, an intractable problem with no closed-form solution, the asymptotic theory of large Toeplitz matrices shows that the generalized Hadamard ratio converges to a broadband coherence statistic as the size of the composite covariance matrix grows large. Although an asymptotic result, simulations of several applications show that even finite dimensional implementations of the broadband coherence statistic can provide a significant improvement in detection performance. This improvement in performance is most likely attributed to the fact that, by constraining the model to incorporate stationarity, we have alleviated some of the difficulties associated with estimating highly parameterized models. Although more generally applicable, the unconstrained covariance estimates used in the generalized Hadamard ratio require the estimation of a much larger number of parameters.

These methods are then applied to the detection of underwater targets in pairs of high frequency and broadband sonar images coregistered over the seafloor. This is a difficult problem due to various factors such as variations in the operating and environmental conditions, presence of spatially varying clutter, and variations in target shapes, compositions, and orientation. A comprehensive study of these methods is conducted using three sonar imagery datasets. The first two datasets are actual images of objects lying on the seafloor and are collected at different geographical locations with the environments from each presenting unique challenges. These two datasets will be used to demonstrate the usefulness of results pertaining to the null distribution of the generalized Hadamard ratio and to study the effects different clutter environments can have on its applicability. They are also used to compare the performance of the broadband coherence detector to several alternative detection techniques. The third dataset used in these studies contains actual images of the seafloor with synthetically generated targets of different geometrical shapes inserted into the images. The primary purpose of this dataset is to study the proposed detection technique's robustness to deviations from coregistration which may occur in practice due to the disparities in high frequency and broadband sonar. Using the results of this section, we will show that the fundamental principle of detecting underwater targets using coherence-based approaches is itself a very useful solution for this problem and that the broadband coherence statistic is adequately adept at achieving this.

Acknowledgements

I would first like to thank my adviser, Dr. Mahmood Azimi, for his invaluable support and guidance throughout the course of my graduate education. I would also like to thank my co-advisor Dr. Louis Scharf for his invaluable guidance throughout the course of this work. Both Dr. Azimi and Dr. Scharf have taught me the value of developing the intellectual reasoning needed to conduct thorough and well-developed research and their support is greatly appreciated.

I would like to thank my committee members, Dr. Ali Pezeshki and Dr. Dan Cooley, for their time and assistance.

I would like to thank the Office of Naval Research for providing the funding for this project. This project was funded by the Office of Naval Research under contracts N00014-12-C-0017 and N00014-12-1-0154. Without the funding from ONR, this work would have never been completed. I would also like to thank the Naval Surface Warfare Center – Panama City, FL for providing the logistical and data support needed for this project.

I would like to thank my colleagues in the Signal and Image Processing Lab. They have provided a great environment for discussing work and providing help when most needed. Thanks to Neil, Soheil, Jarrod, and Justin.

Finally, I would like to thank my family for their support and guidance throughout my education.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

1.1. Problem Statement and Motivations

Detecting the presence of a common but unknown signal among two or more data channels is a problem that finds its uses in many applications, including collaborative sensor networks [1], geological monitoring of seismic activity [2], radar [3], and sonar [4]. Distributed sensor networks consisting of spatially distributed sensors to monitor physical or environmental conditions offer a solution to overcome the shortcomings possibly encountered in single sensor situations by collecting observations at several distinct locations. Some detection systems in such situations use decision fusion [5] to combine individual detection decisions into one global decision. However, this detection paradigm can be sub-optimal as local decisions are based on the perspective of a single sensory system and fail to jointly incorporate the information from multiple sensors. Thus, methods designed without the need to perform separate detection by capturing the coherent or mutual information among multiple data sets are needed.

Nonparametric multichannel detection has been recently discussed in [6] and [7]. The detection methods considered in these works are nonparametric in the sense that they do not make any *a priori* assumptions about the signal that may be present in the observations of each sensor but rather simply looks for high levels of coherence or linear dependence among the observations from all channels. Here, the assumption is that the presence of a common signal among all channels will provide a substantial increase in coherence compared to a situation where each channel contains independent sensor noise.

An important application where the nonparametric multichannel detection paradigm discussed in [6] and [7] proves very useful is the detection of underwater targets in multiple sonar images [8]. This problem is complicated due to various factors such as variations in operating and environmental conditions, presence of spatially varying clutter, variations in target shapes, compositions and orientation. Moreover, bottom features such as coral reefs, sand formations, and vegetation may totally obscure a target or confuse the detection process. Consequently, there is a need for robust detection methods that jointly incorporates the information from multiple sonar images, maintains high performance in varying operating and environmental conditions, and works well given the great variety in target conditions that can be observed for this problem. Given the wide variation in both target and environmental conditions, detection methods such as those developed in [6] and [7] which simply look for high levels of coherence among multiple sonar images, as opposed to those that rely on specific models (e.g. matched filtering), can in some cases be desirable.

In this dissertation, we address several aspects of the detection problem considered in [7] and apply the results to the problem of underwater target detection in pairs of high frequency (HF) and broadband (BB) sonar images coregistered over the seafloor. Although the detection problem in [7] leads to a very simple test statistic that is easily computable, often times the most difficult part of hypothesis testing lies not so much in deriving the appropriate criteria but rather in finding its exact distribution when the hypotheses are true and identifying the threshold needed to achieve a given false alarm probability. For this reason, one of the main subjects of this dissertation is the null distribution of the test statistic that arises from this analysis and the development of several methods that can be used to determine such a threshold. This extends the result in [6] to the case of temporally correlated time series and makes it possible to set thresholds for false alarm control. The second problem considered in this dissertation is the development of efficient test statistics for 2-dimensional wide-sense stationarity (WSS). Here, the goal is the development of methods that exploit the inherent Toeplitz-block-Toeplitz structure of a composite covariance matrix when the data is truly WSS. This extends the frequency domain technique developed in [7] to the frequency/wavenumber domain when each observer employs an array of sensors. The application of both of these results to underwater target detection among multiple sonar images is then demonstrated using several real and synthetic sonar imagery datasets.

1.2. Literature Review on Multi-Channel Detection

Considerable research has been devoted to the development of different detection and classification methodologies to detect and classify underwater objects from single sonar imagery. For instance, in [9] – [11] the authors utilize a matched filter to identify regions in the image that match a template designed to capture the general behavior of targets lying on the seafloor. Recently, however, multi-sensor detection and classification has been considered for this problem. One such work that has looked at underwater target classification from multiple sonar images is given in [12] and [13], where three different sonar images with varying frequency and bandwidth characteristics were used. The classification on each image is done using a multistage classification approach, which entails a repeated application of a classifier. During the training stage, it is determined how many times to apply the classifier and an optimal subset of features are extracted. Each stage of the classifier results in a reduction in the number of false alarms. The final classification decision is made by a fusion of the three classification results from the three different sonar images. Although this work uses disparate sonar systems (with disparateness in the operating frequency of the sonar), the classifier of this method processes each image individually and does not use the information contained in the three images simultaneously to make classification calls.

Many of the methods considered in this work can be related in one way or another to Canonical Correlation Analysis (CCA) [14] which is a well-established method of analyzing the linear dependence among two datasets. The canonical coordinate decomposition method not only determines linear dependence or coherence between two data channels but also extracts, via the canonical coordinates, a subset of the most coherent features for detection and classification purposes. The CCA method has shown great promise in underwater target classification problems using sonar backscatter using data collected by the buried object scanning sonar (BOSS) system [15], [16]. The work in these references presented a multi-ping classification system that extracts coherence-based features from blocks of range cells of time series associated with two sonar returns with single ping separation. These coherence patterns were shown to be different for pairs of pings that contain mine-like objects than those that contain non-mine-like objects. The canonical correlations that capture the coherence patterns [15] were shown to have high discriminatory power for both detection and classification.

The CCA technique was employed in [17] as an alternative to the decision fusion techniques developed in [12] and [13] by forming a dual disparate detector in which detection decisions are based on the amount of coherent information shared among pairs of coregistered Regions of Interest (ROIs) from two different sonar images. This dual disparate detector is then applied to a distributed detection framework [17] and is shown to exhibit high performance with a low false alarm rate and high probability of detection. However, this work considered the fusion of several independent dual-channel detection problems to incorporate multiple sonar images. To fully incorporate the mutual information from multiple ($\geq$ 2) sonar images, the dual-channel CCA-based detector was extended in [8] to multiple sonar images by employing Multichannel Coherence Analysis (MCA) [18] which can be seen as a natural extension of CCA to more than two channels. In this work, the standard Gauss-Gauss detector was cast into the MCA coordinate system and used to detect targets by looking at the amount of coherent information shared among coregistered ROIs in sonar images corresponding to different frequency bands. In [8], the performance of the MCA-based detector to a different choice in the number of sonar images was studied and the detector shown to be an adequate solution for this problem.

Multichannel detection has been considered in [6], [7], and [19]. In [6], a geometric approach to multi-channel detection is proposed by defining the *generalized coherence* (GC) among multiple channels, which is shown to be a natural extension of the *magnitude-squared coherence* (MSC) for more than two channels. Under the assumption that the observations from each channel contain white, complex normal noise, the authors derived closed-form expressions for the null distributions of both the MSC and three-channel GC measures. This leads to a recursive formulation for finding the null distribution as one adds additional channels. In [4], the GC detector was applied to the problem of detecting a common signal among multiple channels containing deep ocean noise. A similar technique was considered in [19] by forming a *Generalized Likelihood Ratio Test* (GLRT) [20] and using the assumption that observations are zero-mean, complex normal random vectors. Given multiple independent realizations of this random vector, the GLRT involves testing whether the sample covariance matrix has diagonal structure under the null hypothesis versus any arbitrary, positive-definite (PD) covariance structure under the alternative. In both [6] and [19], the detection statistic applies to *temporally white but spatially correlated* Gaussian sequences, and is given by the determinant of the sample covariance matrix over the product of its diagonal elements, i.e. a Hadamard ratio.

The work in [6] and [19] was recently extended in [7] by considering the detection of both *spatially and temporally correlated* time series. Given multiple independent realizations of a vector-valued time series, the GLRT of [7] tests whether or not the space-time covariance matrix is block-diagonal. The GLRT is a *generalized* Hadamard ratio involving the sample covariance matrix. Assuming temporally wide-sense stationary processes, and allowing the length of each time series to grow large, the test statistic is shown to be a function of frequency-dependent Hadamard ratios for narrowband cross spectral matrices. At each frequency this Hadamard ratio is a narrowband coherence statistic that measures linear dependence among the time series at that frequency. The log of each such narrowband coherence is integrated over the Nyquist band to produce the broadband coherence statistic. This GLRT is shown to exhibit many appealing properties including invariance to channel-by-channel filtering, a connection to mutual information for WSS Gaussian random processes, as well as providing a generalization of the MSC spectrum for more than two channels [7].

1.3. Research Objectives

The goal of the research in this dissertation is to address two aspects of a GLRT that tests for the independence among multiple random vectors and to study its application to the problem of underwater target detection in pairs of HF and BB images coregistered over the seafloor. The ultimate goal of this work is the development of a detection technique that jointly incorporates the information from multiple sonar images, maintains high performance in varying operating and environmental conditions, and works well given the great variability in target conditions that can be observed for this problem. Due to this wide variation in both target and environmental conditions, detection methods that take advantage of general discriminative features in the data, as opposed to those that rely on specific models, can in some cases be desirable. Such is the case for the solution presented here where the detection principle simply relies on the assumption that the presence of targets in coregistered sonar images will lead to a higher degree of coherence than when those images contain background alone. Posing the problem as a test of independence, the GLRT relies on the computation of a test statistic known as a generalized Hadamard ratio [7].

However, one of the questions that arises in the practical implementation of any likelihood ratio test is the appropriate selection of thresholds. When setting this threshold to achieve a given false alarm probability, an understanding of the likelihood ratio's probabilistic behavior under the null hypothesis is needed and this is thus one of the main subjects of this dissertation. Using the theory of Gram determinants [21], we will show that the generalized Hadamard ratio can be written as a product of ratios of the squared residual from two linear least-squares problems, each of which is independently beta distributed under the null hypothesis. Once this stochastic representation is established, it becomes straightforward to derive various attributes of the null distribution of this test statistic including its moments, characteristic function, and cumulant generating function. Asymptotically, this random variable's characteristic function is shown to converge to that of a chi-squared random variable giving one a very simple way of determining thresholds that approximately achieve a given false alarm probability when the situation applies. Results immediately suggest, however, that the distribution is slow to converge. For this reason, we consider the use of saddlepoint approximations [22] which prove to be a practical alternative to finding thresholds that achieve a given false alarm probability in small sample support scenarios.

The second topic of this dissertation is an alternative implementation of the generalized Hadamard ratio for wide-sense stationary (WSS) random processes. The goal of this portion of the dissertation is the development of methods that exploit the inherent Toeplitz-block-Toeplitz structure of a composite covariance matrix when the data is truly WSS. Although the true GLRT for this problem would impose this Toeplitz structure on the estimate of the composite covariance matrix, this is an intractable problem with no closed-form solution. The asymptotic theory of large Toeplitz matrices, however, is well understood [23] and leads to very tractable results involving the eigenvalues, multiplication, and inversion of large Toeplitz matrices. Using this asymptotic theory, the generalized Hadamard ratio converges to the broadband integral of a narrowband Hadamard ratio of cross-spectral matrices, a test statistic referred to as broadband coherence. Although an asymptotic result, simulations of several applications where this test statistic may apply demonstrate that even finite-dimensional implementations of the broadband coherence statistic can bring substantial improvements in performance.

The final objective of this dissertation is to test the methods developed here on three datasets of sonar images containing both real and synthetically generated targets. The first two datasets were collected at different geographical locations and consist of different environments with each presenting unique challenges and difficulties. These two datasets, which contain images of actual objects lying on the seafloor, are first used to not only demonstrate the usefulness of the null distribution of the GLRT to this application but also to study how different environments affect its applicability. The third dataset contains actual images of the seafloor with synthetically generated targets of different geometrical shapes inserted into the image. This dataset is used to study how a lack of coregistration among the HF and BB images can affect the detectability of the proposed coherence-based methods. The two real sonar datasets are then once again used to evaluate the performance of the broadband coherence statistic and to compare its performance with several alternatives including the generalized Hadamard ratio and a matched subspace detector designed to specifically look for the highlight and shadow characteristics typically associated with targets lying on the seafloor.

1.4. Organization of the Dissertation

This dissertation is organized as follows. Chapter 2 gives a brief review of binary hypothesis testing and detection using the Generalized Likelihood Ratio Test. Using this theory, we then review the development of the GLRT for multichannel detection and discuss several of its properties and implications. Chapter 3 presents the null distribution of the multichannel GLRT and discusses its asymptotic form as the number of samples used to form maximum likelihood estimators grows large. Chapter 4 presents an alternative method of finding closed-form approximations of the likelihood ratio's null distribution using the saddlepoint technique. Chapter 5 then extends the multichannel GLRT to 2-dimensional wide-sense stationary observations and presents results using several simulations. Chapter 6 applies the theoretical developments presented in the dissertation to the problem of detecting the presence of underwater targets in pairs of high-frequency and broadband sonar images and gives a comprehensive study of the effectiveness of the proposed methods by presenting results on several sonar imagery datasets. Finally, Chapter 7 concludes the studies carried out in this research and discusses the goals for future work.

CHAPTER 2

Development of the GLRT for Multichannel Detection

2.1. Introduction

Detection problems can be simply described as deciding which of a set of candidate models most accurately describes or is most consistent with a set of measurements we have collected. This is easily cast into the framework of statistical hypothesis testing, the most basic being the binary hypothesis test where one must decide only among two competing models. This decision paradigm establishes a dichotomy in the parametric description of our observation in the form of a null (typically noise alone) and alternative (typically signal plus noise) hypothesis. If this parametric characterization is completely specified under both models, the Neyman-Pearson lemma [21] confirms intuitive reasoning that the optimal decision criterion is obtained by comparing likelihoods through the use of a likelihood ratio test.

In many practical scenarios, however, it can be difficult to ascertain a suitable parametric description under either one or both hypotheses. Several techniques have been suggested [20] for such a situation with each being different in how they approach the unknown set of parameters. The first technique is the Bayesian approach [20] which treats the parameters as random variables with known probabilistic properties and decides which hypothesis is in force through the use of Bayes factors. The more common approach is the generalized likelihood ratio test (GLRT) [20] which treats the parameters as unknown but deterministic quantities, replaces them with their maximum likelihood (ML) estimates under each model, and selects the hypothesis by conducting a likelihood ratio test.

The GLRT is subsequently employed for the detection of coherence among $L \geq 2$ temporally correlated time series. The methods developed in this section will be used later for the purposes of detecting underwater targets in pairs ($L = 2$) of coregistered sonar images. Under the assumption that these time series are zero mean, complex normal random vectors, their joint distribution is completely specified by a space-time covariance matrix that captures all spatiotemporal second-order information. The objective of the analysis is to determine whether the data channels contain a common but unknown signal by looking for high levels of spatial correlation. Posing the problem as a test of independence among all L channels boils down to testing the null hypothesis that the space-time covariance matrix is block-diagonal through the use of a generalized Hadamard ratio [7]. This likelihood ratio exhibits many appealing properties including invariance to different classes of linear transformations and connections to Canonical Correlation Analysis (CCA).

The outline of this chapter is as follows. Section 2.2 gives a brief review of binary hypothesis testing to provide some background to the methods employed in the later sections of this chapter. Section 2.3 subsequently discusses the hypothesis test considered for this multichannel detection problem, develops the likelihood ratio used in the GLRT, and discusses several properties and implications of the likelihood ratio statistic. Concluding remarks are then given in Section 2.4.

2.2. Binary Hypothesis Testing - A Review

In this section, a brief review of detection theory (or hypothesis testing) is given so as to provide some background into the methods developed in Section 2.3. A more detailed treatment of this subject can be found in [20] – [24]. A binary hypothesis test is a method of deciding which of two candidate models most accurately describes a set of collected measurements. The measurement $\mathbf{x} \in \mathcal{X}$ is taken from the vector space $\mathcal{X}$ and is assumed to be a real or complex-valued random vector distributed according to the probability density function (PDF) $f(\mathbf{x}; \boldsymbol{\theta})$. The values in the vector $\boldsymbol{\theta}$ parameterize this distribution and are assumed to be taken from some parameter space Θ. This space is partitioned into two disjoint subsets $\Theta = \Theta_0 \cup \Theta_1$ with Θ_0 containing the parameters associated with the null hypothesis $\mathcal{H}_0$ and Θ_1 containing those associated with the alternative hypothesis $\mathcal{H}_1$. If the set Θ_i is a singleton set containing only one element, the hypothesis $\mathcal{H}_i$ is said to be simple otherwise $\mathcal{H}_i$ is said to be composite. Our goal is simply to decide which of these two subsets the unknown parameter vector $\boldsymbol{\theta}$ belongs to given the measurement $\mathbf{x}$. Stated succinctly, we consider the binary hypothesis test

$$\begin{aligned} \mathcal{H}_0 &: \boldsymbol{\theta} \in \Theta_0 \\ \mathcal{H}_1 &: \boldsymbol{\theta} \in \Theta_1 \end{aligned}$$

To make these formal statements more concrete, suppose that the vector $\mathbf{x} = [x_1 \ \cdots \ x_M]^T \in \mathbb{R}^M$ is a vector of *iid* samples of a scalar normal distribution with unknown mean μ and known variance σ^2, i.e. $x_i \overset{iid}{\sim} N(\mu, \sigma^2)$. With $\boldsymbol{\theta} = \mu$ the unknown parameter in this example, the vector $\mathbf{x}$ has PDF

$$f(\mathbf{x}; \mu) = \frac{1}{(2\pi\sigma^2)^{M/2}} \exp\left\{ -\frac{1}{2\sigma^2} ||\mathbf{x} - \mu \mathbf{1}_M||^2 \right\}$$

with $\mathbf{1}_M$ denoting an M-dimensional vector of ones. Table 2.1 gives several examples of different classes of binary hypothesis tests that might be considered under this model and the parameter space Θ associated with each.

TABLE 2.1. Several examples of different classes of binary hypothesis tests for a scalar distribution with unknown mean.

Simple vs. Simple	Simple vs. Composite	Composite vs. Composite
$\mathcal{H}_0 \ : \ \mu = 0$ $\mathcal{H}_1 \ : \ \mu = 1$	$\mathcal{H}_0 \ : \ \mu = 0$ $\mathcal{H}_1 \ : \ \mu \neq 0$	$\mathcal{H}_0 \ : \ \mu \leq 0$ $\mathcal{H}_1 \ : \ \mu > 0$
$\Theta = \{0, 1\}$ $\Theta_0 = \{0\}$ $\Theta_1 = \{1\}$	$\Theta = \mathbb{R}$ $\Theta_0 = \{0\}$ $\Theta_1 = (-\infty, 0) \cup (0, \infty)$	$\Theta = \mathbb{R}$ $\Theta_0 = (-\infty, 0]$ $\Theta_1 = (0, \infty)$

To make a decision as to which hypothesis is in force, we can construct the arbitrary decision function $\phi(\mathbf{x})$

$$\phi(\mathbf{x}) = \begin{cases} 1 \sim \mathcal{H}_1 & \mathbf{x} \in R \\ 0 \sim \mathcal{H}_0 & \mathbf{x} \in A \end{cases}$$

with $\mathcal{X} = A \cup R$. Stated in words, our decision is to accept the null hypothesis $\mathcal{H}_0$ if our measurement falls within the "acceptance" region A or to reject it in favor of the alternative $\mathcal{H}_1$ if the measurement falls within the "rejection" region R. The trick is then to define these regions according to an appropriate optimization criterion.

If both the null and alternative hypotheses are simple so that $\Theta = \{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1\}$, our intuition tells us to reject the null hypothesis in favor of $\mathcal{H}_1$ if the *likelihood* of observing $\mathbf{x}$ under $\mathcal{H}_1$, $f(\mathbf{x}; \boldsymbol{\theta}_1)$, is large relative to that under $\mathcal{H}_0$, $f(\mathbf{x}; \boldsymbol{\theta}_0)$. In other words, we should base our decision on the *likelihood ratio test*

$$\Lambda(\mathbf{x}) = \frac{f(\mathbf{x}; \boldsymbol{\theta}_1)}{f(\mathbf{x}; \boldsymbol{\theta}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda \tag{1}$$

for some $\lambda \geq 0$. The Neyman-Pearson lemma [24], [21] establishes that this is in fact optimal in the sense that, of all the test functions $\phi(\cdot)$ that exhibit the same false alarm probability (probability of incorrectly accepting $\mathcal{H}_1$), the likelihood ratio test is the one that maximizes the probability of detection (probability of correctly accepting $\mathcal{H}_1$).

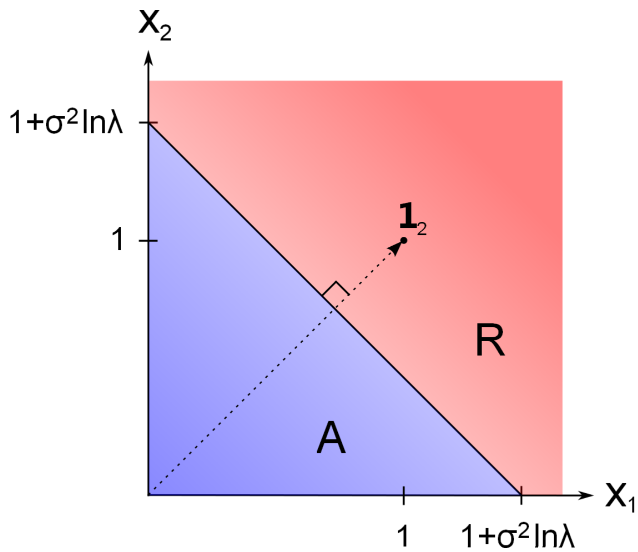


FIGURE 2.1. Acceptance and rejection regions when testing $\mu = 0$ versus $\mu = 1$ with two *iid* normal random variables.

The likelihood ratio test is also very convenient in that it *implicitly* tells us how to construct the regions A and R in the measurement space. To see this, let's go back to the example considered earlier in this section involving *iid* scalar normal random variables along with the hypothesis test given in the left column of Table 2.1. Noting that the logarithm is monotone increasing and will thus not change the outcome of the test, we choose to reject the null hypothesis in favor of $\mathcal{H}_1$ whenever the following inequality is satisfied

$$\ln \Lambda(\mathbf{x}) = \ln \frac{f(\mathbf{x};\mu=1)}{f(\mathbf{x};\mu=0)} = \frac{1}{2\sigma^2}\left(||\mathbf{x}||^2 - ||\mathbf{x} - \mathbf{1}_M||^2\right) > \ln \lambda$$

If $M = 2$ so that $\mathbf{x} = [x_1\ x_2]^T \in \mathbb{R}^2$, the inequality can be rewritten as

$$x_1 + x_2 > 1 + \sigma^2 \ln \lambda$$

which describes the set of points lying to the right of a linear hyperplane that lies orthogonal to the vector $\mathbf{1}_2$ as depicted in Figure 2.1.

Although a very powerful and convenient result, it is often times difficult to implement the Neyman-Pearson lemma in practice as the PDF $f(\mathbf{x};\boldsymbol{\theta})$ may not be completely known under one or both hypotheses due to uncertainties in $\boldsymbol{\theta}$. Such is the case in radar/sonar where the return from a target will be delayed and attenuated as the signal propagates through the medium resulting in an unknown arrival time and amplitude. Composite hypothesis tests therefore arise as a manifestation of our lack of *a priori* knowledge in the vector $\boldsymbol{\theta}$.

Two general approaches have been considered when dealing with composite hypothesis testing with each differing in how they philosophically treat the unknown vector $\boldsymbol{\theta}$. The first approach relies on Bayesian inference [20] by treating the unknown vector $\boldsymbol{\theta}_i$ as a random vector distributed according to the prior PDF $f(\boldsymbol{\theta}_i)$ which probabilistically captures our uncertainty in $\boldsymbol{\theta}_i$. With these assumptions, the marginal PDF of the measurement

$$f(\mathbf{x}) = \int f(\mathbf{x}|\boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta}$$

is completely specified, no longer dependent on the vector $\boldsymbol{\theta}$. In accordance with the Neyman-Pearson lemma, the optimal test is

$$\Lambda(\mathbf{x}) = \frac{\displaystyle\int f(\mathbf{x}|\boldsymbol{\theta}_1)f(\boldsymbol{\theta}_1)d\boldsymbol{\theta}_1}{\displaystyle\int f(\mathbf{x}|\boldsymbol{\theta}_0)f(\boldsymbol{\theta}_0)d\boldsymbol{\theta}_0} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda$$

which is simply a ratio of likelihoods averaged over the appropriate prior PDF $f(\boldsymbol{\theta}_i)$. The ratio given in this expression is often times referred to as the *Bayes factor* [25]. The Bayesian approach, however, can still be fairly difficult to implement in practice as the probabilistic characteristics of the parameters must be known *a priori* and the need for multidimensional integration in the construction of the likelihood ratio for vector-valued parameters.

The more common approach to composite hypothesis testing is the *generalized likelihood ratio test* (GLRT) [20] which treats $\boldsymbol{\theta}$ as deterministic but unknown and replaces it with the value that is most likely under each hypothesis. That is, the GLRT asks us to produce ML estimates of the parameters under each hypothesis

$$\begin{aligned}\hat{\boldsymbol{\theta}}_1 &= \arg\max_{\boldsymbol{\theta}\in\Theta_1} f(\mathbf{x};\boldsymbol{\theta}) \\ \hat{\boldsymbol{\theta}}_0 &= \arg\max_{\boldsymbol{\theta}\in\Theta_0} f(\mathbf{x};\boldsymbol{\theta})\end{aligned}$$

and construct the following likelihood ratio test

$$\Lambda_G(\mathbf{x}) = \frac{\max_{\boldsymbol{\theta}\in\Theta_1} f(\mathbf{x};\boldsymbol{\theta})}{\max_{\boldsymbol{\theta}\in\Theta_0} f(\mathbf{x};\boldsymbol{\theta})} = \frac{f(\mathbf{x};\hat{\boldsymbol{\theta}}_1)}{f(\mathbf{x};\hat{\boldsymbol{\theta}}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda$$

Note that if Θ_1 and Θ_0 are both singleton sets corresponding to simple hypotheses, then the expression given above reverts back the test given in (1), hence the term "generalized" in GLRT. While no notion of optimality can be ascribed to the GLRT in general, it seems to perform fairly well in practice. It is also conceptually very intuitive and simple to implement, the only possible difficulty arising from finding closed-form solutions for the ML estimator. Although subtly different from the expression of the GLRT given above, it is very common [26] to express the problem as a nested set of hypotheses and to consider the following likelihood ratio test as an alternative

$$\Lambda_G(\mathbf{x}) = \frac{\max_{\boldsymbol{\theta}\in\Theta_0} f(\mathbf{x};\boldsymbol{\theta})}{\max_{\boldsymbol{\theta}\in\Theta} f(\mathbf{x};\boldsymbol{\theta})} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \gamma \tag{2}$$

for some $0 \leq \gamma \leq 1$. In this expression, the likelihood ratio is simply the likelihood of the best model under the null hypothesis (maximizes likelihood over Θ_0) normalized by the likelihood of the best model overall (maximizes likelihood over $\Theta = \Theta_1 \cup \Theta_0$). As a consequence, it

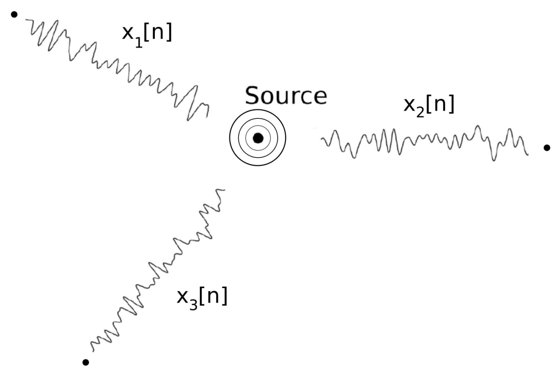


FIGURE 2.2. In many cases, the random vector $\mathbf{x}_i$ represents the collection of a length N time series at one sensor location.

intuitively follows that the likelihood ratio given in (2) satisfies the following inequality

$$0 \leq \frac{\max_{\boldsymbol{\theta} \in \Theta_0} f(\mathbf{x}; \boldsymbol{\theta})}{\max_{\boldsymbol{\theta} \in \Theta} f(\mathbf{x}; \boldsymbol{\theta})} \leq 1$$

The test statistic given in (2) will be employed in the next section for the purposes of multichannel detection.

2.3. Multichannel GLRT

Detecting the presence of a common but unknown signal among multiple channels is a problem that finds its uses in many applications, including collaborative sensor networks [1], geological monitoring of seismic activity [2], radar [3], and sonar [4]. Consider the setup shown in Figure 2.2 consisting of L spatially distributed sensors and define the random vector $\mathbf{x}_i = [x_i[0] \ \cdots \ x_i[N-1]]^T \in \mathbb{C}^N$ to be the length N time series captured at sensor i. Assuming the collection of random vectors $\{\mathbf{x}_i\}_{i=1}^{L}$ to be zero mean, the composite vector

$\mathbf{z} = \left[\mathbf{x}_1^T \ \cdots \ \mathbf{x}_L^T\right]^T \in \mathbb{C}^{LN}$ has covariance matrix

$$R = E\left[\mathbf{z}\mathbf{z}^H\right] = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1L} \\ R_{12}^H & R_{22} & \cdots & R_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ R_{1L}^H & R_{2L}^H & \cdots & R_{LL} \end{bmatrix} \in \mathbb{C}^{LN \times LN} \tag{3}$$

with $R_{ik} = R_{ki}^H = E\left[\mathbf{x}_i\mathbf{x}_k^H\right] \in \mathbb{C}^{N \times N}$. This matrix not only characterizes the second-order information for each channel individually via R_{ii} but also captures the interdependence between every pair of channels via R_{ik} for all $i \neq k$.

Without making any *a priori* assumption about the signal observed by each sensor, one very intuitive way of determining if a common signal exists among all L channels is to test for deviations from statistical independence. If the set of random vectors $\{\mathbf{x}_i\}_{i=1}^L$ is jointly proper complex normal [27], testing for independence among all L channels boils down to testing whether or not the covariance matrix R is block-diagonal. Casting this problem into the standard inference framework, we consider the hypothesis test,

$$\begin{aligned} \mathcal{H}_0 &: \quad R \in \mathcal{R}_0 \\ \mathcal{H}_1 &: \quad R \in \mathcal{R}, \end{aligned} \tag{4}$$

with $\mathcal{R}$ denoting the set of all PD Hermitian matrices and $\mathcal{R}_0$ denoting the set of all matrices in $\mathcal{R}$ which are block-diagonal. Thus, we wish to test the null hypothesis that all L channels are spatially uncorrelated yet possibly temporally correlated versus the alternative that they are both spatially and temporally correlated. In the context of Section 2.2, the parameter vector $\boldsymbol{\theta}$ represents the covariance matrix R, the parameter space Θ corresponds to the set $\mathcal{R}$, and Θ_0 represents the set $\mathcal{R}_0$.

We now assume we are given an experiment producing $M \geq LN$ *iid* realizations $\{\mathbf{x}_i[m]\}_{m=1}^{M}$ of the random vector from each channel i, where

$$\mathbf{x}_i[m] = [x_i[0,m] \ \cdots \ x_i[N-1,m]]^T \in \mathbb{C}^N$$

The composite vectors $\mathbf{z}[m] = \left[\mathbf{x}_1^T[m] \ \cdots \mathbf{x}_L^T[m]\right]^T$ are organized into a data matrix $\mathcal{Z}$:

$$\mathcal{Z} = [\mathbf{z}[1] \ \cdots \ \mathbf{z}[M]] = \begin{bmatrix} \mathbf{x}_1[1] & \cdots & \mathbf{x}_1[M] \\ \vdots & \ddots & \vdots \\ \mathbf{x}_L[1] & \cdots & \mathbf{x}_L[M] \end{bmatrix} \in \mathbb{C}^{LN \times M} \tag{5}$$

The probability density function (PDF) of $\mathcal{Z}$ is

$$f(\mathcal{Z}; R) = \prod_{m=1}^{M} f\left(\mathbf{z}[m]; R\right) = \frac{1}{\pi^{LNM} \det\left(R\right)^M} \exp\left\{-M \mathrm{tr}\left(R^{-1}\hat{R}\right)\right\}$$

In this expression $\hat{R}$ is the estimated composite covariance matrix

$$\hat{R} = \frac{1}{M}\mathcal{Z}\mathcal{Z}^H = \frac{1}{M}\sum_{m=1}^{M} \mathbf{z}[m]\mathbf{z}^H[m] = \begin{bmatrix} \hat{R}_{11} & \hat{R}_{12} & \cdots & \hat{R}_{1L} \\ \hat{R}_{12}^H & \hat{R}_{22} & \cdots & \hat{R}_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{R}_{1L}^H & \hat{R}_{2L}^H & \cdots & \hat{R}_{LL} \end{bmatrix}$$

and $\hat{R}_{ik}$ is an M sample estimate of the $N \times N$ cross-covariance matrix R_{ik}.

As described in Section 2.2, the first step in computing the GLRT is to derive ML estimators for the matrix R. Noting that the function $\frac{1}{M}\ln x + LN \ln \pi$ is monotonically increasing and will not affect our answer, one arrives at the following log-likelihood function

$$\ell(R) = \frac{1}{M} \ln f(\mathcal{Z}; R) + LN \ln \pi = -\ln\det(R) - \mathrm{tr}(R^{-1}\hat{R})$$

Using identities concerning the derivative of the determinant and trace of a PD Hermitian matrix, namely $\frac{d \ln\det(R)}{dR} = R^{-1}$ and $\frac{d\,\mathrm{tr}(R^{-1}\hat{R})}{dR} = -R^{-1}\hat{R}R^{-1}$, taking the derivative of this expression with respect to R and setting it equal to zero yields

$$\frac{d\ell(R)}{dR} = -R^{-1} + R^{-1}\hat{R}R^{-1} = \mathbf{O}_{LN} \;\Rightarrow\; R = \hat{R}$$

where $\mathbf{O}_{LN}$ represents an $LN \times LN$ matrix of zeros. In other words, the unconstrained ML estimate of R is simply the estimated composite covariance matrix $\hat{R}$

$$\hat{R} = \arg\max_{R \in \mathcal{R}} f(\mathcal{Z}; R)$$

Under the constraint that $R = D = \mathrm{blkdiag}\{R_{11}, \ldots, R_{LL}\} \in \mathcal{R}_0$, the log-likelihood function becomes

$$\ell(D) = -\sum_{i=1}^{L} \ln\det(R_{ii}) - \sum_{i=1}^{L} \mathrm{tr}(R_{ii}^{-1}\hat{R}_{ii})$$

In a manner very similar to before, we may take the partial derivative of this expression with respect to R_{ii} and set it equal to zero

$$\frac{\partial\ell(D)}{\partial R_{ii}} = -R_{ii}^{-1} + R_{ii}^{-1}\hat{R}_{ii}R_{ii}^{-1} = \mathbf{O}_N \;\Rightarrow\; R_{ii} = \hat{R}_{ii}$$

In other words, the ML estimate of the $N \times N$ matrix R_{ii} is simply the corresponding $N \times N$ diagonal block of $\hat{R}$ so that

$$\hat{D} = \text{blkdiag}\left\{\hat{R}_{11}, \ldots, \hat{R}_{LL}\right\} = \arg\max_{R \in \mathcal{R}_0} f(\mathcal{Z}; R)$$

Using these ML estimators, the M^{th}-root of the likelihood ratio given in (2) becomes

$$\begin{aligned}\Lambda &= \left(\frac{\max_{R \in \mathcal{R}_0} f(\mathcal{Z}; R)}{\max_{R \in \mathcal{R}} f(\mathcal{Z}; R)}\right)^{1/M} = \left(\frac{f(\mathcal{Z}; \hat{D})}{f(\mathcal{Z}; \hat{R})}\right)^{1/M} \\ &= \frac{\det \hat{R}}{\det \hat{D}} \exp\left\{\text{tr}\left[\hat{R}(\hat{R}^{-1} - \hat{D}^{-1})\right]\right\} \\ &= \frac{\det \hat{R}}{\det \hat{D}} = \frac{\det \hat{R}}{\prod_{i=1}^{L} \det \hat{R}_{ii}} = \det \hat{C} \end{aligned} \tag{6}$$

where $\hat{C} = \hat{D}^{-1/2} \hat{R} \hat{D}^{-H/2}$ is referred to as the coherence matrix [7] as it measures the cross-correlation between the "whitened" random vectors $\hat{R}_{ii}^{-1/2} \mathbf{x}_i$ for $i = 1, \ldots, L$. Note that in developing the expression given in (6), we have used the fact that

$$\exp\left\{\text{tr}\left[\hat{R}\left(\hat{R}^{-1} - \hat{D}^{-1}\right)\right]\right\} = \exp\{LN - LN\} = 1$$

The likelihood ratio given in (6) is referred to as a *generalized* Hadamard ratio in [7] to differentiate it from the Hadamard ratio [28] representing the determinant of a covariance matrix over the product of its diagonal elements.

2.3.1. Invariance Properties of the Multichannel GLRT. Many inference problems in statistics possess inherent symmetry or invariance properties that quite naturally impose restrictions on the possible procedures that should be used. As a simple motivating example, suppose that $\mathbf{x} = [x_1 \; x_2]^T \in \mathbb{C}^2$ is a zero-mean, proper complex normal random vector. Given M samples of this vector, $\mathbf{x}[m]$ for $m = 1, \ldots, M$, consider the problem of estimating the population correlation coefficient

$$\rho = \frac{E[x_1 x_2^*]}{\sqrt{E[|x_1|^2]E[|x_2|^2]}}$$

The correlation coefficient ρ is itself unchanged by, or is invariant under, the transformation

$$\tilde{\mathbf{x}} = \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix} = \begin{bmatrix} t_1 & 0 \\ 0 & t_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

for any real-valued, strictly positive values of t_1 and t_2. If the statistic $\phi\left(\mathbf{x}[1], \ldots, \mathbf{x}[M]\right)$ is to be used as an estimate of ρ, then it makes sense to restrict our attention to those estimators that exhibit the same invariance property, i.e. the set of estimators $\phi(\cdot)$ such that $\phi\left(\mathbf{x}[1], \ldots, \mathbf{x}[M]\right) = \phi\left(\tilde{\mathbf{x}}[1], \ldots, \tilde{\mathbf{x}}[M]\right)$, since both sides of this equality are estimating the same quantity. The sample correlation coefficient

$$\hat{\rho} = \frac{\sum_{m=1}^{M} x_1[m] x_2^*[m]}{\sqrt{\left(\sum_{m=1}^{M} |x_1[m]|^2\right)\left(\sum_{m=1}^{M} |x_2[m]|^2\right)}}$$

is an example of such an estimator.

The same argument holds true for many hypothesis testing problems where there is often a natural group of transformations with respect to which a specific testing problem is invariant. It is sensible in these situations to restrict our attention to the class of invariant tests, i.e. tests based on statistics that are invariant under this group of transformations. Turning our attention to the multichannel problem once again, the hypothesis testing problem in (4) and the likelihood ratio statistic in (6) are invariant under the transformation $\tilde{\mathbf{z}} = T\mathbf{z}$ for certain classes of matrix T. If we denote the estimated composite covariance matrix of the original data vector $\mathbf{z}$ as $\hat{R}^{(\mathbf{z})}$, then it is easy to see that the new data vector $\tilde{\mathbf{z}}$ will have the estimate

$$\hat{R}^{(\tilde{\mathbf{z}})} = \frac{1}{M}\sum_{m=1}^{M} T\mathbf{z}[m]\mathbf{z}^H[m]T^H = T\left(\frac{1}{M}\sum_{m=1}^{M}\mathbf{z}[m]\mathbf{z}^H[m]\right)T^H = T\hat{R}^{(\mathbf{z})}T^H$$

Two examples of such linear transformations are given below.

(i) Invertible Block-Diagonal Matrices

The set of all matrices T such that $T = \text{blkdiag}\{T_1, \ldots, T_L\}$ with T_i any $N \times N$ invertible matrix. Substituting the matrix $\hat{R}^{(\tilde{\mathbf{z}})}$ into the expression given in (6) and recalling properties of the determinant, it is straightforward to see that the likelihood ratio remains invariant

$$\frac{\det\left(\hat{R}^{(\tilde{\mathbf{z}})}\right)}{\prod_{i=1}^{L}\det\left(\hat{R}_{ii}^{(\tilde{\mathbf{z}})}\right)} = \frac{\det\left(\hat{R}^{(\mathbf{z})}\right)\prod_{i=1}^{L}\det\left(T_i\right)^2}{\left[\prod_{i=1}^{L}\det\left(\hat{R}_{ii}^{(\mathbf{z})}\right)\right]\left[\prod_{i=1}^{L}\det\left(T_i\right)^2\right]} = \frac{\det\left(\hat{R}^{(\mathbf{z})}\right)}{\prod_{i=1}^{L}\det\left(\hat{R}_{ii}^{(\mathbf{z})}\right)}$$

This invariance property was noted in [7] and shows us that there exists no channel-by-channel invertible linear transformation, including scaling and filtering, that moves a covariance from $\mathcal{H}_0$ to $\mathcal{H}_1$ or vice versa. One of the practical implications of this property is that it guarantees that the result of the test will be independent of the basis or coordinate system used to measure and interpret the data. This fact will be used in Section 3.3 where we will analyze the data in a coordinate system where each channel is white by setting $T_i = R_{ii}^{-1/2}$. It will also be used in Section 5.3 where we will analyze the data in the frequency domain by setting $T_i = F_N$ with F_N denoting an $N \times N$ DFT matrix.

(*ii*) Block-Permutation Matrices

The set of all matrices T such that $T = P \otimes I_N$ with P any $L \times L$ permutation matrix. To see that the likelihood ratio is invariant to this type of linear transformation, we can first of all note that, for any $n \times n$ matrix A and any $m \times m$ matrix B, the matrix $A \otimes B$ has determinant $\det(A \otimes B) = \det(A)^m \det(B)^n$. Using this property of the Kronecker product along with the fact that permutation matrices are orthogonal ($\det(P) = \pm 1$), it is easy to see that the determinant in the numerator of the likelihood ratio remains unchanged

$$\det\left(\hat{R}^{(\tilde{\mathbf{z}})}\right) = \det(P)^{2N} \det(I_N)^{2L} \det\left(\hat{R}^{(\mathbf{z})}\right) = \det\left(\hat{R}^{(\mathbf{z})}\right)$$

In the denominator of the likelihood ratio, the linear operator $T = P \otimes I_N$ only serves to change the order in which terms are multiplied. However, as multiplication is commutative, it follows that the denominator of the likelihood ratio remains unchanged as well. The importance of this invariance property is that it gaurantees that the ordering in channel index will have no influence on the result of the test.

2.3.2. Geometry of the GLRT. To get some notion of the geometry of the parameter spaces $\mathcal{R}$ and $\mathcal{R}_0$ for this problem, we can consider the special case of real data with $L = 2$ channels and a length $N = 1$ time series, i.e. $\mathbf{z} = [x_1 \ x_2]^T \in \mathbb{R}^2$. In this case, the covariance matrix R can be described as a point $\mathbf{r} \in \mathbb{R}^3$

$$R = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{12}^2 & \sigma_{22}^2 \end{bmatrix} \Leftrightarrow \mathbf{r} = \begin{bmatrix} \sigma_{11}^2 \\ \sigma_{22}^2 \\ \sigma_{12}^2 \end{bmatrix}$$

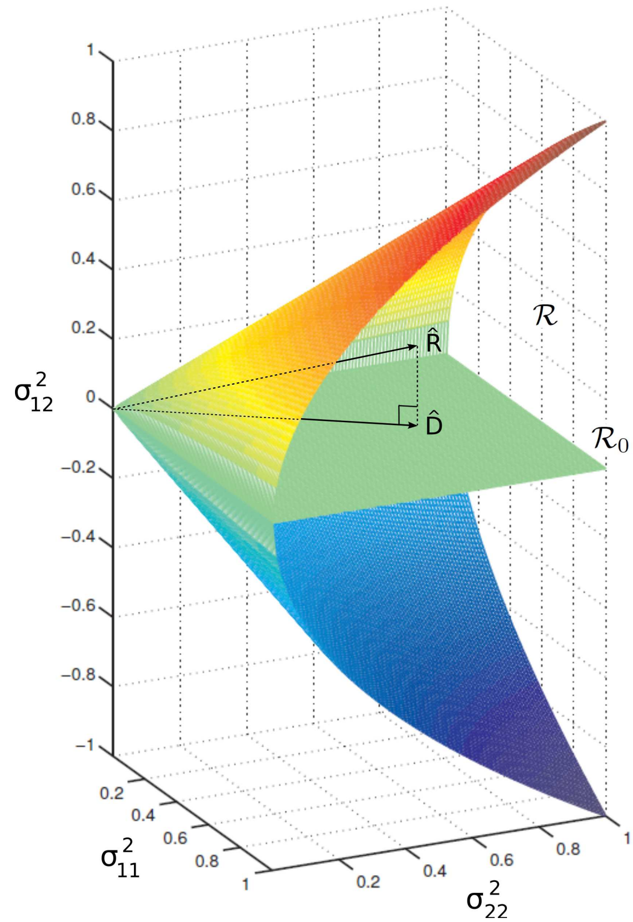


FIGURE 2.3. Geometry of the parameter spaces $\mathcal{R}$ and $\mathcal{R}_0$ in $\mathbb{R}^3$.

Note that we are interested in $\mathbb{R}^3$ and not $\mathbb{R}^4$ as the matrix R is symmetric, i.e. one of the two off-diagonal variables is redundant. The set of points with $\sigma^2_{11}, \sigma^2_{22} > 0$ and $|\sigma^2_{12}| = \sqrt{\sigma^2_{11}\sigma^2_{22}}$ describes the convex cone of positive semidefinite matrices. The parameter space $\mathcal{R}$ therefore corresponds to the set of points that strictly lie within this cone

$$\mathcal{R} = \left\{(\sigma^2_{11}, \sigma^2_{22}, \sigma^2_{12}) : \sigma^2_{11} > 0, \sigma^2_{22} > 0, \left|\sigma^2_{12}\right| < \sqrt{\sigma^2_{11}\sigma^2_{22}}\right\}$$

while $\mathcal{R}_0$ is a 2 dimensional plane that bisects this cone

$$\mathcal{R}_0 = \left\{(\sigma^2_{11}, \sigma^2_{22}, \sigma^2_{12}) : \sigma^2_{11} > 0, \sigma^2_{22} > 0, \sigma^2_{12} = 0\right\}$$

The geometry of these two sets is depicted in Figure 2.3.

Upon identifying the most likely estimate of the composite covariance matrix $\hat{R}$, the GLRT tells us to find $\hat{D}$ by simply taking the matrix $\hat{R}$ and setting it's off-diagonal elements equal to zero, i.e. $\hat{\sigma}_{12}^2 = 0$. As depicted in Figure 2.3, this process corresponds to orthogonally projecting the point $\hat{R}$ into the subspace $\mathcal{R}_0$. In fact, it isn't difficult to see that, in general, $\hat{D}$ satisfies the principle of orthogonality. That is, for any $D = \text{blkdiag}\{R_{11}, \ldots, R_{LL}\} \in \mathcal{R}_0$

$$\begin{aligned}
\left\langle \hat{R} - \hat{D}, D \right\rangle &= \text{tr}\left[D\left(\hat{R} - \hat{D}\right)\right] \\
&= \text{tr}\left(\begin{bmatrix} R_{11} & 0 & \cdots & 0 \\ 0 & R_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R_{LL} \end{bmatrix}\begin{bmatrix} 0 & \hat{R}_{12} & \cdots & \hat{R}_{1L} \\ \hat{R}_{12}^H & 0 & \cdots & \hat{R}_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{R}_{1L}^H & \hat{R}_{2L}^H & \cdots & 0 \end{bmatrix}\right) \\
&= 0
\end{aligned}$$

So, by simply nulling the off-diagonal blocks of $\hat{R}$, we have effectively solved the optimization problem

$$\hat{D} = \arg\min_{D \in \mathcal{R}_0} \left|\left|\hat{R} - D\right|\right|_F^2$$

In addition to being the ML estimator under the null hypothesis, $\hat{D}$ is also *the* block-diagonal matrix, D, which minimizes the Euclidean distance from $\hat{R}$, i.e. minimizes the Frobenius norm of the error matrix $E = \hat{R} - D$.

2.3.3. Connection with Adaptive Canonical Coordinates. Canonical correlation analysis (CCA) [14] is a well-established method of analyzing and interpreting the linear relationships among two sets of random variables. CCA achieves this by finding linear combinations of each pair of random variables that produces maximum correlation between them.

Many classical problems in detection [29] and estimation [28] can be posed in the coordinate system produced by CCA. In fact, the author in [30] notes that "virtually all of the commonly encountered parametric tests of significance can be treated as special cases of canonical correlation analysis, which is the general procedure for investigating the relationships between two sets of variables," and the problem considered here is no exception.

For the time being, consider the two-channel composite observation $\mathbf{z} = \left[\mathbf{y}^T \; \mathbf{x}^T\right]^T$ with $\mathbf{y} \in \mathbb{C}^m$ and $\mathbf{x} \in \mathbb{C}^N$. Without any loss in generality, we will assume that $N \leq m$. The covariance matrix of the composite vector $\mathbf{z}$ may be decomposed into the product of a lower-triangular matrix by a block-diagonal matrix by an upper-triangular matrix as follows

$$R_{zz} = E\left[\mathbf{z}\mathbf{z}^H\right] = \begin{bmatrix} R_{yy} & R_{xy}^H \\ R_{xy} & R_{xx} \end{bmatrix} = \begin{bmatrix} I & 0 \\ W & I \end{bmatrix} \begin{bmatrix} R_{yy} & 0 \\ 0 & Q_{xx} \end{bmatrix} \begin{bmatrix} I & W^H \\ 0 & I \end{bmatrix}$$

The matrices W and Q_{xx} in this expression correspond to the Wiener filter and error covariance matrices [28], respectively, when linearly estimating $\mathbf{x}$ from $\mathbf{y}$

$$W = R_{xy} R_{yy}^{-1} \in \mathbb{C}^{N \times m}$$

and

$$Q_{xx} = E\left[(\mathbf{x} - W\mathbf{y})(\mathbf{x} - W\mathbf{y})^H\right] = R_{xx} - R_{xy} R_{yy}^{-1} R_{xy}^H \in \mathbb{C}^{N \times N}$$

From this decomposition, it becomes self-evident that the generalized Hadamard ratio of R_{zz} can be written

$$\frac{\det R_{zz}}{(\det R_{yy})\,(\det R_{xx})} = \frac{\det R_{yy}}{\det R_{yy}} \frac{\det Q_{xx}}{\det R_{xx}} = \frac{\det Q_{xx}}{\det R_{xx}}$$

This ratio is referred to as a relative filtering volume [28] in that it compares the volume of the concentration ellipse associated with the a posteriori error $\hat{\mathbf{e}}_x = \mathbf{x} - W\mathbf{y}$ to that of the prior "error" $\mathbf{x}$. Defining the two-channel coherence matrix $C = R_{xx}^{-1/2} R_{xy} R_{yy}^{-H/2}$, one may take its singular value decomposition

$$\begin{aligned} C &= F\tilde{K}G^H;\ FF^H = I_N;\ GG^H = I_m \\ \tilde{K} &= [K\ 0];\ K = \operatorname{diag}\{k_1, \ldots, k_N\} \end{aligned}$$

It was shown in [28] that this relative filtering volume can be expressed solely in terms of the squared canonical correlations, k_n^2 for $n = 1, \ldots, N$, according to the relationship

$$\frac{\det Q_{xx}}{\det R_{xx}} = \prod_{n=1}^{N} \left(1 - k_n^2\right) \tag{7}$$

Given M *iid* realizations of the vectors $\mathbf{x}$ and $\mathbf{y}$, these results extend to the case in which R_{zz} is replaced by its sample estimate $\hat{R}_{zz}$, resulting in a framework referred to in [28] as *adaptive* canonical coordinates.

Returning to the multichannel detection problem, let the vector $\mathbf{x}$ represent the time series for the L^{th} channel, $\mathbf{x} = \mathbf{x}_L \in \mathbb{C}^N$, and the vector $\mathbf{y}$ represent the time series corresponding to channels 1 through $L-1$, $\mathbf{y} = \mathbf{z}_{L-1} = \begin{bmatrix}\mathbf{x}_1^T & \cdots & \mathbf{x}_{L-1}^T\end{bmatrix}^T \in \mathbb{C}^{(L-1)N}$. According to the arguments given above, the determinant of the space-time covariance matrix R in (3) can be written

$$\det R = (\det R_{\mathbf{z}_{L-1}\mathbf{z}_{L-1}})(\det Q_{LL})$$

where $R_{\mathbf{z}_{L-1}\mathbf{z}_{L-1}}$ contains all spatiotemporal second-order information for the first $L-1$ channels

$$R_{\mathbf{z}_{L-1}\mathbf{z}_{L-1}} = \begin{bmatrix} R_{11} & \cdots & R_{1,L-1} \\ \vdots & \ddots & \vdots \\ R_{1,L-1}^H & \cdots & R_{L-1,L-1} \end{bmatrix}$$

and Q_{LL} is the error covariance matrix associated with linearly estimating $\mathbf{x}_L$ from $\mathbf{x}_1, \ldots, \mathbf{x}_{L-1}$ collectively. Similarly, one may then define $\mathbf{x} = \mathbf{x}_{L-1} \in \mathbb{C}^N$ and $\mathbf{y} = \mathbf{z}_{L-2} = \begin{bmatrix} \mathbf{x}_1^T & \cdots & \mathbf{x}_{L-1}^T \end{bmatrix}^T \in \mathbb{C}^{(L-2)N}$ to decompose the determinant of $R_{\mathbf{z}_{L-1}\mathbf{z}_{L-1}}$ as

$$\det R_{\mathbf{z}_{L-1}\mathbf{z}_{L-1}} = (\det R_{\mathbf{z}_{L-2}\mathbf{z}_{L-2}})(\det Q_{L-1,L-1})$$

where $Q_{L-1,L-1}$ is the error covariance matrix when estimating $\mathbf{x}_{L-1}$ from $\mathbf{x}_1, \ldots, \mathbf{x}_{L-2}$. Proceeding in this manner, it then follows that

$$\det R = \det R_{11} \prod_{i=2}^{L} \det Q_{ii}$$

with Q_{ii} denoting the error covariance matrix when linearly estimating the time series $\mathbf{x}_i$ from $\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}$. Replacing these matrices with their sample estimates and using (7), this course-grained filtering procedure shows us that the generalized Hadamard ratio given in (6) can be written as

$$\frac{\det \hat{R}}{\prod_{i=1}^{L} \det \hat{R}_{ii}} = \frac{\det \hat{R}_{11}}{\det \hat{R}_{11}} \prod_{i=2}^{L} \frac{\det \hat{Q}_{ii}}{\det \hat{R}_{ii}} = \prod_{i=2}^{L} \prod_{n=1}^{N} \left[1 - \hat{k}_n^2(i)\right]$$

where the adaptive canonical correlations $\hat{k}_n^2(i)$, $n = 1, \ldots, N$, are found by performing a two-channel canonical correlation analysis using all M *iid* copies of the vector $\mathbf{x}_i$ with the *iid* copies of the vector $\mathbf{z}_{i-1} = \begin{bmatrix} \mathbf{x}_1^T & \cdots & \mathbf{x}_{i-1}^T \end{bmatrix}^T$. The realization of the generalized Hadamard ratio as an iterated sequence of two-channel CCA problems is depicted in Figure 2.4. The process starts by measuring the linear dependence between the vectors $\mathbf{z}_1 = \mathbf{x}_1$ and $\mathbf{x}_2$ through their squared canonical correlations, $\hat{k}_n^2(2)$. These vectors are concatenated to form the vector $\mathbf{z}_2 = \begin{bmatrix} \mathbf{x}_1^T \ \mathbf{x}_2^T \end{bmatrix}^T$ and its linear dependence with the new vector $\mathbf{x}_3$ is likewise measured through the squared canonical correlations, $\hat{k}_n^2(3)$. Repeating this process until all

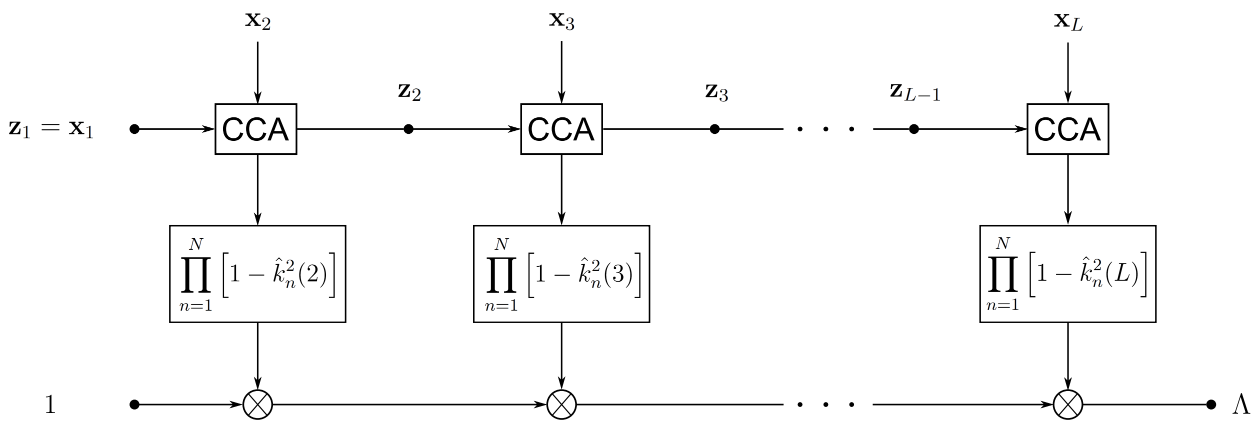


FIGURE 2.4. The generalized Hadamard ratio in (6) can be realized as an iterated sequence of two-channel adaptive CCA problems.

L channels have been incorporated while multiplying the results of each CCA problem all the way through yields the test statistic Λ. Thus, one can see that the generalized Hadamard ratio has close connections to measuring coherence using two-channel CCA.

2.4. Conclusion

In this chapter, a general overview of binary hypothesis testing was given. Detection problems involving binary hypotheses is simply the problem of deciding which of a pair of competing models is most consistent with a set of collected measurements. If the parametric description of this measurement is completely specified under both hypotheses, the optimal test is given by a likelihood ratio test. The GLRT is an extension of this principle to combat situations where this parametric description may not be completely known under one or both hypotheses and accomplishes this by substituting parameters with their ML estimate.

The GLRT is then employed for the detection of multiple temporally correlated time series. The joint probabilistic characteristics of these time series are solely determined by a space-time covariance matrix which captures all spatiotemporal second-order information. Testing the null hypothesis that this space-time covariance matrix is block-diagonal, implementing the GLRT results in a likelihood ratio that becomes a generalized Hadamard ratio.

This likelihood ratio exhibits many appealing properties including invariance to different classes of linear transformations and connections to Canonical Correlation Analysis (CCA).

The generalized Hadamard ratio will be the subject of much of the upcoming chapters in this work. In Chapters 3 and 4 we will derive the null distribution and saddlepoint approximations, respectively, of this test statistic for the purposes of defining thresholds for achieving a specific false alarm rate. In Chapter 5 we will extend the likelihood ratio to situations in which the random vector from each channel arises from a 2D wide sense stationary random process. Finally, in Chapter 6 these methods will be applied for the purpose of the detection of underwater targets in pairs of synthetic aperture sonar images.

CHAPTER 3

Null Distribution of the GLRT for Multichannel Detection

3.1. Introduction

In Chapter 2 we discussed the generalized Hadamard ratio and its manifestation as the likelihood ratio of a GLRT that tests for linear dependence among a collection of spatially distributed time series. One important question that arises in the practical application of this likelihood ratio (and any likelihood ratio test for that matter) is "what values of this test statistic constitute sufficient evidence in support of the decision to reject the null hypothesis?" Although the GLRT may be very easy to implement, often times the most difficult part of hypothesis testing lies not so much in deriving the appropriate criteria but rather in finding its exact distribution when the hypotheses are true and identifying the threshold needed to achieve a given false alarm probability. To be able to characterize the false alarm rate, however, an explicit understanding of the probabilistic behavior of the test statistic under the null hypothesis is needed.

In this chapter we start by taking a second look at the generalized Hadamard ratio. Using the theory of Gram determinants, it is shown that this test statistic can be written as a product of ratios of the squared residual from two linear prediction problems. Geometrical insights into the structure of these ratios leads to the conclusion that, under the null hypothesis that the space-time covariance matrix R is truly block-diagonal, the generalized Hadamard ratio is stochastically equivalent to a product of independent but *not* identically distributed beta random variables. This stochastic representation makes it clear that the null distribution of this test is solely dependent on the number of channels (L), the length of each time series (N), and the number of samples (M) used to construct the estimated covariance matrix $\hat{R}$ but in no way dependent on the second-order temporal characteristics of each individual channel.

We then turn our attention to applying this result for the purposes of determining thresholds which approximately achieve a desired false alarm probability. Knowing that the null distribution of the test is statistically equivalent to a product of betas makes it very straightforward to derive certain characteristics of this random variable including its moments as well as the characteristic function of its logarithm. Employing results concerning the asymptotic behavior of the log-gamma function, we begin by showing that, for large M, the log-likelihood ratio converges in distribution to a chi-squared random variable. Moreover, the degrees of freedom of this chi-squared distribution can be interpreted in terms of the dimensions of the parameter spaces $\mathcal{R}$ and $\mathcal{R}_0$ considered in the construction of the likelihood ratio. Both of these results (the fact that the null distribution is asymptotically chi-squared as well as the interpretation for the degrees of freedom) are in complete agreement with a well-established result [31] concerning the null distribution of the GLRT in general.

The remainder of this chapter is organized as follows. In Section 3.2, we use the theory of Gram determinants to demonstrate that the generalized Hadamard ratio can alternatively be expressed as a product of ratios involving the error from two linear prediction problems. Building on this result, Section 3.3 shows that the null distribution of the test is statistically equivalent to a product of beta random variables and demonstrates the result using a Monte Carlo simulation. This stochastic representation is subsequently employed in Section 3.4 for determining the asymptotic form of the likelihood ratio's probability distribution. Concluding remarks are finally given in Section 3.5.

3.2. The GLRT Revisited

As stated in the preceding section, the goal of this chapter is to develop an understanding of how the generalized Hadamard ratio given in (6) behaves probabilistically under the null hypothesis. In this section, we will show that the generalized Hadamard ratio can equivalently be expressed as a product of ratios of squared residuals from two linear least squares problems. This fact will be used in the next section to show that the likelihood ratio in (6) is stochastically equivalent to a product of independent beta random variables under the null hypothesis.

We'll begin by noting that, for any $i \geq 2$ and any $n = 0, \ldots, N-1$, the data matrix $\mathcal{Z}$ given in (5) can be partitioned as follows

$$\mathcal{Z} = \begin{bmatrix} Z_i \\ X_{in} \\ \mathbf{x}_{in}^H \\ \vdots \end{bmatrix}$$

where the matrix $Z_i \in \mathbb{C}^{(i-1)N \times M}$ contains all M realizations of the time-series $\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}$ (sensor channels 1 to $i-1$)

$$Z_i = \begin{bmatrix} \mathbf{x}_1[1] & \mathbf{x}_1[2] & \cdots & \mathbf{x}_1[M] \\ \mathbf{x}_2[1] & \mathbf{x}_2[2] & \cdots & \mathbf{x}_2[M] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{i-1}[1] & \mathbf{x}_{i-1}[2] & \cdots & \mathbf{x}_{i-1}[M] \end{bmatrix}$$

the matrix $X_{in} \in \mathbb{C}^{n \times M}$ contains all M realizations of the i^{th} time-series up to temporal sample $n-1$ (first n temporal samples of channel i)

$$X_{in} = \begin{bmatrix} x_i[0,1] & x_i[0,2] & \cdots & x_i[0,M] \\ x_i[1,1] & x_i[1,2] & \cdots & x_i[1,M] \\ \vdots & \vdots & \ddots & \vdots \\ x_i[n-1,1] & x_i[n-1,2] & \cdots & x_i[n-1,M] \end{bmatrix}$$

and the vector $\mathbf{x}_{in} = [x_i[n,1]\ x_i[n,2]\ \cdots\ x_i[n,M]]^H \in \mathbb{C}^M$ contains all M realizations of random variable $x_i[n]$ (n^{th} temporal sample of channel i). With this partition in the data matrix, the northwest corner of the Gram matrix $\mathcal{Z}\mathcal{Z}^H$ has the structure

$$\mathcal{Z}\mathcal{Z}^H = \begin{bmatrix} \hat{R}_{ZZ} & \hat{R}_{ZX} & \hat{\mathbf{r}}_{Z\mathbf{x}} & \\ \hat{R}_{ZX}^H & \hat{R}_{XX} & \hat{\mathbf{r}}_{X\mathbf{x}} & \cdots \\ \hat{\mathbf{r}}_{Z\mathbf{x}}^H & \hat{\mathbf{r}}_{X\mathbf{x}}^H & \hat{r}_{\mathbf{xx}} & \\ & \vdots & & \ddots \end{bmatrix},$$

with entries defined as follows

$$\hat{R}_{ZZ} = Z_i Z_i^H,\ \hat{R}_{ZX} = Z_i X_{in}^H,\ \hat{R}_{XX} = X_{in} X_{in}^H$$

$$\hat{\mathbf{r}}_{Z\mathbf{x}} = Z_i \mathbf{x}_{in},\ \hat{\mathbf{r}}_{X\mathbf{x}} = X_{in} \mathbf{x}_{in}$$

$$\hat{r}_{\mathbf{xx}} = \mathbf{x}_{in}^H \mathbf{x}_{in}$$

Note that the "ˆ" notation is used here simply to remind one of the connection these terms share with estimated covariance matrices. With $L = 3$ and $N = 5$, Figure 3.1 gives a color coded demonstration of this partition when $i = 3$ and $n = 2$.

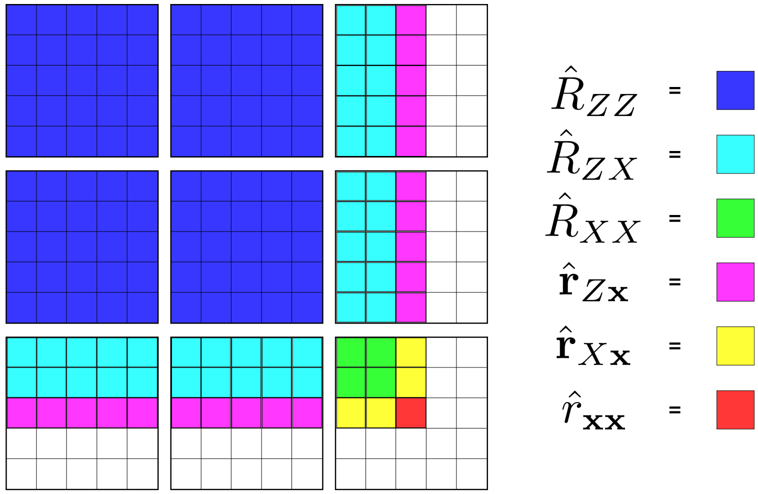


FIGURE 3.1. Partitioning of the northwest corner of matrix $\mathcal{Z}\mathcal{Z}^H$.

Using results concerning the determinants of Gram matrices discussed in Appendix A, it is straightforward to show using (A-2) that the determinant of the estimated composite covariance matrix can be decomposed into a product of scalars as follows

$$M^{LN} \det \hat{R} = \det\left(\mathcal{Z}\mathcal{Z}^H\right) = \det\left(X_{1N}X_{1N}^H\right) \prod_{i=2}^{L} \prod_{n=0}^{N-1} \sigma_{in}^2(\hat{R})$$

where

$$\sigma_{in}^2(\hat{R}) = \hat{r}_{\mathbf{xx}} - \begin{bmatrix} \hat{\mathbf{r}}_{Z\mathbf{x}}^H & \hat{\mathbf{r}}_{X\mathbf{x}}^H \end{bmatrix} \begin{bmatrix} \hat{R}_{ZZ} & \hat{R}_{ZX} \\ \hat{R}_{ZX}^H & \hat{R}_{XX} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{r}}_{Z\mathbf{x}} \\ \hat{\mathbf{r}}_{X\mathbf{x}} \end{bmatrix}$$

Using the definition of these matrices given above, this term can be written

$$\sigma_{in}^2(\hat{R}) = \mathbf{x}_{in}^H \left(I - \left[Z_i^H X_{in}^H \right] \begin{bmatrix} \hat{R}_{ZZ} & \hat{R}_{ZX} \\ \hat{R}_{ZX}^H & \hat{R}_{XX} \end{bmatrix}^{-1} \begin{bmatrix} Z_i \\ X_{in} \end{bmatrix} \right) \mathbf{x}_{in}$$

$$= \mathbf{x}_{in}^H \left(I - P_{ZX}\right) \mathbf{x}_{in} = \mathbf{x}_{in}^H P_{ZX}^{\perp} \mathbf{x}_{in}$$

where P_{ZX} denotes the projection onto the $(i-1)N+n$ dimensional subspace $\langle ZX \rangle$ spanned by the columns of matrix $[Z_i^H \; X_{in}^H]$. Moreover, using the block-wise inversion identity

$$\begin{bmatrix} \hat{R}_{ZZ} & \hat{R}_{ZX} \\ \hat{R}_{ZX}^H & \hat{R}_{XX} \end{bmatrix}^{-1} = \begin{bmatrix} \Delta^{-1} & -\Delta^{-1}\hat{R}_{ZX}\hat{R}_{XX}^{-1} \\ -\hat{R}_{XX}^{-1}\hat{R}_{ZX}^H\Delta^{-1} & \hat{R}_{XX}^{-1} + \hat{R}_{XX}^{-1}\hat{R}_{ZX}^H\Delta^{-1}\hat{R}_{ZX}\hat{R}_{XX}^{-1} \end{bmatrix}$$

with the Schur complement $\Delta = \hat{R}_{ZZ} - \hat{R}_{ZX}\hat{R}_{XX}^{-1}\hat{R}_{ZX}^H$, one can derive yet another equivalent expression for this term

$$\begin{aligned} \sigma_{in}^2(\hat{R}) &= \mathbf{x}_{in}^H P_{ZX}^{\perp} \mathbf{x}_{in} \\ &= \mathbf{x}_{in}^H \left(P_X^{\perp} - P_X^{\perp} Z_i^H \left(Z_i P_X^{\perp} Z_i^H \right)^{-1} Z_i P_X^{\perp} \right) \mathbf{x}_{in} \\ &= \mathbf{x}_{in}^H P_X^{\perp} \mathbf{x}_{in} - \mathbf{x}_{in}^H P_{P_X^{\perp} Z} \mathbf{x}_{in} \end{aligned} \tag{8}$$

where $P_X = X_{in}^H R_{XX}^{-1} X_{in}$ and $P_{P_X^{\perp} Z}$ denote the projection onto the n dimensional subspace spanned by the columns of matrix X_{in}^H and the projection onto the $(i-1)N$ dimensional subspace spanned by the columns of matrix $P_X^{\perp} Z_i^H$, respectively.

To compute the determinant of the block-diagonal matrix $\hat{D}$ in the denominator of the likelihood ratio in (6), one can take a very similar approach to show that

$$M^N \det \hat{R}_{ii} = \det \left(X_{iN} X_{iN}^H \right) = \prod_{n=0}^{N-1} \sigma_{in}^2(\hat{R}_{ii})$$

where

$$\begin{aligned} \sigma_{in}^2(\hat{R}_{ii}) &= \hat{r}_{\mathbf{xx}} - \hat{\mathbf{r}}_{X\mathbf{x}}^H \hat{R}_{XX}^{-1} \hat{\mathbf{r}}_{X\mathbf{x}} \\ &= \mathbf{x}_{in}^H \left(I - X_{in}^H \left(X_{in} X_{in}^H \right)^{-1} X_{in} \right) \mathbf{x}_{in} = \mathbf{x}_{in}^H P_X^{\perp} \mathbf{x}_{in} \end{aligned}$$

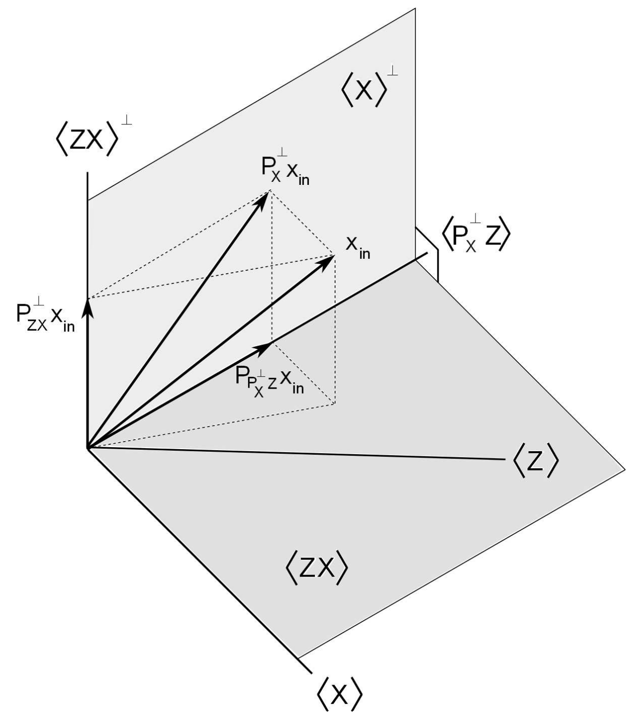


FIGURE 3.2. Orthogonal decomposition of the projection $P_X^{\perp}\mathbf{x}_{in}$ into $P_{ZX}^{\perp}\mathbf{x}_{in}$ and $P_{P_X^{\perp}Z}\mathbf{x}_{in}$.

Recalling the relationship given in (8), note that this term can alternatively be expressed as follows:

$$\sigma_{in}^2(\hat{R}_{ii}) = \mathbf{x}_{in}^H P_X^{\perp} \mathbf{x}_{in} = \mathbf{x}_{in}^H P_{ZX}^{\perp} \mathbf{x}_{in} + \mathbf{x}_{in}^H P_{P_X^{\perp}Z} \mathbf{x}_{in} \tag{9}$$

The geometry of this decomposition is depicted in Figure 3.2.

Using the decompositions of the determinants of these matrices, it is then straightforward to see that the likelihood ratio given in (6) can finally be written

$$\begin{aligned}
\Lambda = \frac{\det \hat{R}}{\det \hat{D}} &= \frac{\det\left(\mathcal{Z}\mathcal{Z}^H\right)}{\prod_{i=1}^{L} \det\left(X_{iN}X_{iN}^H\right)} = \frac{\det\left(X_{1N}X_{1N}^H\right)}{\det\left(X_{1N}X_{1N}^H\right)} \prod_{i=2}^{L} \frac{\prod_{n=0}^{N-1} \sigma_{in}^2(\hat{R})}{\det\left(X_{iN}X_{iN}^H\right)} \\
&= \prod_{i=2}^{L} \prod_{n=0}^{N-1} \frac{\sigma_{in}^2(\hat{R})}{\sigma_{in}^2(\hat{R}_{ii})} = \prod_{i=2}^{L} \prod_{n=0}^{N-1} \frac{\mathbf{x}_{in}^H P_{ZX}^{\perp} \mathbf{x}_{in}}{\mathbf{x}_{in}^H P_{ZX}^{\perp} \mathbf{x}_{in} + \mathbf{x}_{in}^H P_{P_X^{\perp}Z} \mathbf{x}_{in}}
\end{aligned} \tag{10}$$

Each term within the product of this expression represents the ratio of the estimated variance of a residual from two different linear least squares problems: the numerator, $\sigma_{in}^2(\hat{R})$, is found by regressing $x_i[n]$ onto all the random variables previous to it while the denominator, $\sigma^2_{in}(\hat{R}_{ii})$, is found by regressing $x_i[n]$ onto those random variables associated with channel i only. Recalling the second invariance property in Section 2.3.1, one must keep in mind that the order in channel index used when constructing this sequence of estimation problems ultimately has no effect on the likelihood ratio. That is, switching orders of channels will lead to different estimation problems generally leading to different individual ratios within the product in (10). However, the product as a whole is invariant to order.

Figure 3.2 depicts the geometry of the projections $P^{\perp}_{ZX}\mathbf{x}_{in}$ and $P_{P^{\perp}_X Z}\mathbf{x}_{in}$ used to construct each term of the likelihood ratio in (10). As demonstrated by the relationship given in (9), the squared length of the vector $P_{P^{\perp}_X Z}\mathbf{x}_{in}$ in essence represents the increase in mean-squared error incurred by excluding channels $\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}$ from the estimation problem. The smaller the length of this vector relative to the length of $P^{\perp}_{ZX}\mathbf{x}_{in}$, the more evidence in support of the null hypothesis of independence as it indicates that one can just as accurately estimate $x_i[n]$ by ignoring the previous channels. Although the expression of the likelihood ratio given in (10) is no easier to compute than that in (6), it shows that (6) can be written as a product of scalar random variables whose null distribution is the subject of the upcoming section.

3.3. Stochastic Representation under the Null Hypothesis

To characterize the distribution of the likelihood ratio under the null hypothesis, we begin by imposing the assumption that the composite vector $\mathbf{z} = \begin{bmatrix}\mathbf{x}_1^T & \cdots & \mathbf{x}_L^T\end{bmatrix}^T \sim \mathcal{CN}(\mathbf{0}, D)$ for any $D = \operatorname{blkdiag}\{R_{11}, \ldots, R_{LL}\} \in \mathcal{R}_0$. Recalling the first invariance property in Section 2.3.1, we note that under these circumstances we can always apply the linear transformation

$$T = D^{-1/2} = \operatorname{blkdiag}\left\{R_{11}^{-1/2}, \ldots, R_{LL}^{-1/2}\right\},$$

a pre-whitener, to the random vector $\mathbf{z}$ without any consequence to the likelihood ratio. Thus, there is no loss in generality to assume that $D = I_{LN}$ (note that $I_{LN} \in \mathcal{R}_0$) or equivalently that $\mathbf{x}_{in} \overset{iid}{\sim} \mathcal{CN}(\mathbf{0}, I_M)$.

Looking again at Figure 3.2, it is clear that the two projections $P^{\perp}_{ZX}\mathbf{x}_{in}$ and $P_{P^{\perp}_{X}Z}\mathbf{x}_{in}$ lie in two orthogonal subspaces of $\mathbb{C}^M$, i.e. $\langle P^{\perp}_{ZX}\mathbf{x}_{in}, P_{P^{\perp}_{X}Z}\mathbf{x}_{in}\rangle = 0$. A straightforward application of Cochran's Theorem [32] then shows that $2\mathbf{x}^H_{in}P^{\perp}_{ZX}\mathbf{x}_{in}$ and $2\mathbf{x}^H_{in}P_{P^{\perp}_{X}Z}\mathbf{x}_{in}$ are statistically independent chi-squared random variables with degrees of freedom $2\ rank\left(P^{\perp}_{ZX}\right) = 2\alpha_{in}$ and $2\ rank\left(P_{P^{\perp}_{X}Z}\right) = 2\beta_i$, respectively, where

$$
\begin{aligned}
\alpha_{in} &= M - (i-1)N - n \\
\beta_i &= (i-1)N
\end{aligned}
$$

Similar to the arguments given at the end of Appendix A concerning the distribution of the determinant of a complex Wishart distributed matrix, we can also see that, although the random variables $\sigma^2_{in}(\hat{R})$ and $\sigma^2_{in}(\hat{R}_{ii})$ given in (10) are *functionally* dependent on the data matrix $\left[Z^H_i\ X^H_{in}\right]$ through the construction of the projection matrices $P^{\perp}_{ZX}$ and $P^{\perp}_{X}$, respectively, they are in fact *statistically* independent making the sequence of squared residuals

$$
\frac{\sigma^2_{in}(\hat{R})}{\sigma^2_{in}(\hat{R}_{ii})} = \frac{\mathbf{x}^H_{in}P^{\perp}_{ZX}\mathbf{x}_{in}}{\mathbf{x}^H_{in}P^{\perp}_{ZX}\mathbf{x}_{in} + \mathbf{x}^H_{in}P_{P^{\perp}_{X}Z}\mathbf{x}_{in}};\ i = 2, \ldots, L
$$

a set of mutually independent random variables. Note that the two quadratic forms $2\mathbf{x}^H_{in}P^{\perp}_{ZX}\mathbf{x}_{in}$ and $2\mathbf{x}^H_{in}P_{P^{\perp}_{X}Z}\mathbf{x}_{in}$ will be chi-squared distributed with degrees of freedom $2\alpha_{in}$ and $2\beta_i$, respectively, whenever the random vector $\mathbf{x}_{in}$ is distributed according to any spherically symmetric distribution [26]. That is, this result holds for any random vector whose probability density depends on the data only through the squared-norm $||\mathbf{x}_{in}||^2$. However, the normal distribution is the only example of an elliptically symmetric distribution [26] which can be made spherically symmetric using a linear transformation.

As discussed in Appendix B, if X and Y represent two independent chi-squared random variables with degrees of freedom ν_X and ν_Y, respectively, then the random variable $\frac{X}{X+Y}$ is distributed according to a beta distribution with parameters $\nu_X/2$ and $\nu_Y/2$. This fact leads us to conclude that [33], [34]

$$\Lambda|\mathcal{H}_0 \stackrel{d}{=} \prod_{i=2}^{L}\prod_{n=0}^{N-1} Y_{in} \tag{11}$$

where Y_{in} denotes a random variable with distribution $Y_{in} \sim \text{Beta}(\alpha_{in}, \beta_i)$, all distributed independently of one another. Equation (11) says, "under the null hypothesis, the likelihood ratio statistic is distributed as the product of beta random variables, $\text{Beta}(\alpha_{in}, \beta_i)$". Note that if the assumption of a complex normal distribution for the data channels is replaced with a real-valued multivariate normal, we can modify the above statements accordingly by simply halving the parameters of these beta random variables, i.e. $Y_{in} \sim \text{Beta}(\alpha_{in}/2, \beta_i/2)$.

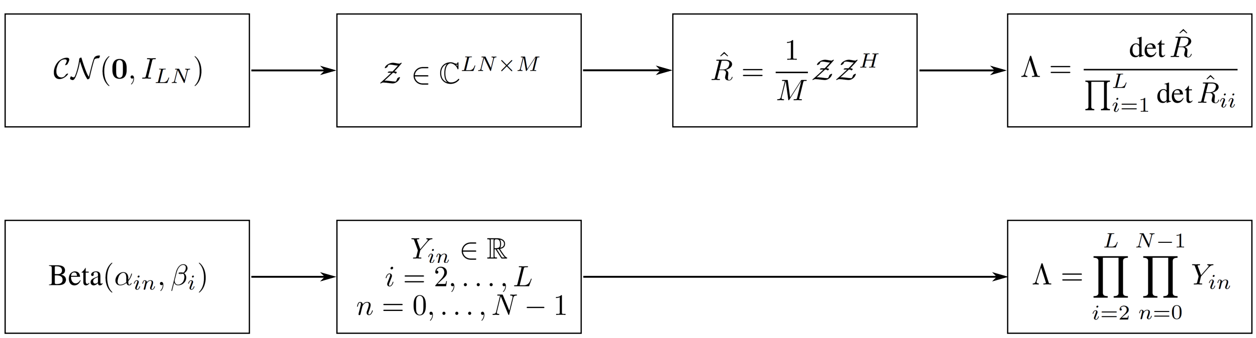


FIGURE 3.3. Two statistically equivalent realizations of (6) under the null hypothesis.

With the result given in (11), Figure 3.3 demonstrates two statistically equivalent means of simulating the test statistic under the null hypothesis. For both the real and complex-valued versions of the GLR and with $L = 3$, $N = 24$, and $M = 100$, Figure 3.4 displays

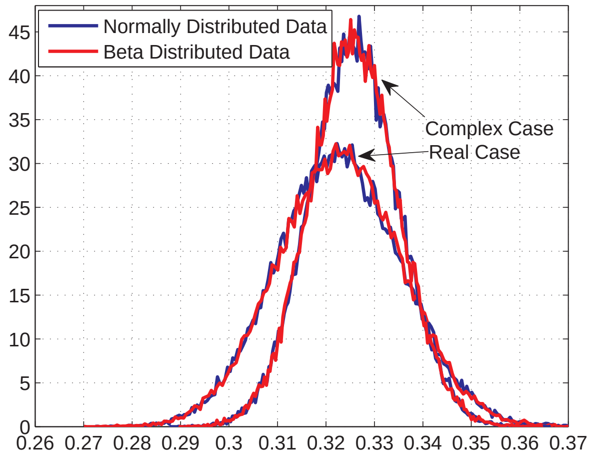


FIGURE 3.4. Histogram Comparison with $L = 3$, $N = 24$, and $M = 100$.

histograms of Monte Carlo trials generated according to these two methods. The blue plots in Figure 3.4 are formed by generating the data matrix $\mathcal{Z}$ using the complex normal distribution, forming the sample covariance matrix, and computing the ratio of determinants given in (6) as depicted in the top of Figure 3.3. Likewise, the red plots are formed by generating the independent but not identically distributed scalars Y_{in} using the beta distribution and forming the product given in (11) as depicted in the bottom of Figure 3.3. These figures show good agreement in the histograms, illustrating stochastically the mathematical fact that the null distribution of (6) can be generated by drawing independent beta random variables and forming their product.

3.4. Asymptotic Null Distribution

Although an interesting conclusion in its own right, the result given in (11) fails to get us any closer to attaining a closed-form expression for the PDF of the likelihood ratio under the null hypothesis. One of the more convenient and beautiful results concerning the GLRT, however, is the distribution that it takes as the number of samples, M, used to form ML estimators goes to infinity. Let $\{\mathbf{x}_i\}_{i=1}^{M}$ be a collection of *iid* samples from the PDF

$f(\mathbf{x};\boldsymbol{\theta})$ where $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^{r+s}$. Furthermore, assume that the unknown parameter vector is partitioned $\boldsymbol{\theta} = \left[\boldsymbol{\eta}^T \; \boldsymbol{\xi}^T\right]^T$ with $\boldsymbol{\eta} \in \mathbb{R}^r$, $\boldsymbol{\xi} \in \mathbb{R}^s$ representing a vector of nuisance parameters, and consider the s-dimensional subspace $\Theta_0 = \{(\boldsymbol{\eta}, \boldsymbol{\xi}) \in \Theta : \boldsymbol{\eta} = \boldsymbol{\eta}_0\}$. Then, under the null hypothesis $\mathcal{H}_0 : \boldsymbol{\eta} = \boldsymbol{\eta}_0$, it follows that the log-likelihood ratio converges in distribution to a chi-squared random variable [31]

$$-2\ln\left(\frac{\max\limits_{\boldsymbol{\theta}\in\Theta_0}\prod\limits_{i=1}^{M} f(\mathbf{x}_i;\boldsymbol{\theta})}{\max\limits_{\boldsymbol{\theta}\in\Theta}\prod\limits_{i=1}^{M} f(\mathbf{x}_i;\boldsymbol{\theta})}\right) \xrightarrow{M\to\infty} \chi^2_\nu \tag{12}$$

with degrees of freedom $\nu = r = (r+s) - s = \dim(\Theta) - \dim(\Theta_0)$. In this section, we will use the fact that the generalized Hadamard ratio in (6) is stochastically equivalent to a product of independent beta random variables to derive this result for this specific test statistic. Moreover, we will find that it is possible to scale the generalized Hadamard ratio to improve the rate at which the test statistic converges to its asymptotic chi-squared distribution.

Given the definition of the beta function $B(x, y)$ and its connection with the gamma function $\Gamma(z)$

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} = \int_0^1 z^{x-1}(1-z)^{y-1}dz \tag{13}$$

every random variable within the product given in (11) has k^{th} raw moment

$$\begin{aligned}
E\left[Y_{in}^k\right] &= \int_0^1 x^k f\left(x;\alpha_{in},\beta_i\right)dx = \frac{1}{B(\alpha_{in},\beta_i)}\int_0^1 x^{(\alpha_{in}+k)-1}(1-x)^{\beta_i-1}dx \\
&= \frac{B(\alpha_{in}+k,\beta_i)}{B(\alpha_{in},\beta_i)} = \frac{\Gamma(\beta_i)}{\Gamma(\beta_i)}\,\frac{\Gamma(\alpha_{in}+k)\Gamma(\alpha_{in}+\beta_i)}{\Gamma(\alpha_{in})\Gamma(\alpha_{in}+\beta_i+k)} \\
&= \frac{\Gamma(\alpha_{in}+k)\Gamma(\alpha_{in}+\beta_i)}{\Gamma(\alpha_{in})\Gamma(\alpha_{in}+\beta_i+k)}
\end{aligned} \tag{14}$$

where $f(x;\alpha_{in},\beta_i)$ denotes the PDF of a beta random variable with parameters α_{in} and β_i. Using (14) along with the fact that these random variables are independent, one can see that the likelihood ratio has the following k^{th} order moment under the null hypothesis

$$\mu_\Lambda(k) = E\left[\Lambda^k|\mathcal{H}_0\right] = \prod_{i=2}^{L}\prod_{n=0}^{N-1} E\left[Y_{in}^k\right] = \prod_{i=2}^{L}\prod_{n=0}^{N-1} \frac{\Gamma(M-n)\Gamma(M-(i-1)N-n+k)}{\Gamma(M-(i-1)N-n)\Gamma(M-n+k)} \tag{15}$$

If we define $\xi_n = (1-\rho)M-n$ with $0 \le \rho \le 1$ an arbitrary real number, the characteristic function of the random variable $Z = -2\rho M \ln\Lambda$ can be written

$$\begin{aligned}\phi_Z(jt) &= E\left[e^{jtZ}|\mathcal{H}_0\right] = E\left[\Lambda^{-2j\rho M t}|\mathcal{H}_0\right] = \mu_\Lambda(-2j\rho M t)\\ &= \prod_{i=2}^{L}\prod_{n=0}^{N-1} \frac{\Gamma(\rho M+\xi_n)\Gamma(\rho M(1-2jt)+\xi_n-(i-1)N)}{\Gamma(\rho M+\xi_n-(i-1)N)\Gamma(\rho M(1-2jt)+\xi_n)}\end{aligned}$$

Its cumulant generating function is

$$\begin{aligned}\psi_Z(jt) &= \ln\phi_Z(jt)\\ &= \sum_{i=2}^{L}\sum_{n=0}^{N-1}\left[\ln\Gamma(\rho M+\xi_n) - \ln\Gamma\left(\rho M+\xi_n-(i-1)N\right)\right]\\ &+ \sum_{i=2}^{L}\sum_{n=0}^{N-1}\left[\ln\Gamma\left(\rho M(1-2jt)+\xi_n-(i-1)N\right) - \ln\Gamma\left(\rho M(1-2jt)+\xi_n\right)\right]\end{aligned} \tag{16}$$

To investigate the properties of this cumulant generating function for large M, one may employ the following asymptotic expansion [26], [35]

$$\ln\Gamma(z+a) = \frac{1}{2}\ln 2\pi + (z+a-\frac{1}{2})\ln z - z + \sum_{n=1}^{\infty}(-1)^{n+1}\frac{B_{n+1}(a)}{n(n+1)z^n} \tag{17}$$

where $B_n(x)$ denotes an n^{th} order Bernoulli polynomial

$$B_n(x) = \sum_{k=0}^{n} \binom{n}{k} b_{n-k} x^k$$

and b_n for $n \geq 0$ are Bernoulli numbers [36]. The series given in (17) converges as $|z| \to \infty$ provided that $|\arg z| < \pi$. Assuming that the variable ρ does not go to zero as M becomes large (in fact we'll find that $(1-\rho)$ will be chosen to be $\mathcal{O}(M^{-1})$), then the expression given above can be used to expand the log-gamma functions in the cumulant generating function for large M. As described in detail in Appendix C, this expansion of the log-gamma function allows one to obtain the asymptotic expression

$$\psi_Z(jt) = -\frac{\nu}{2} \ln(1-2jt) + \omega_1(\rho) \left[(1-2jt)^{-1} - 1\right] + \mathcal{O}(M^{-2}) \tag{18}$$

where

$$\nu = 2 \sum_{i=2}^{L} \sum_{n=0}^{N-1} (i-1)N = L^2N^2 - LN^2$$

$$\omega_1(\rho) = \frac{1}{2\rho M} \sum_{i=2}^{L} \sum_{n=0}^{N-1} \left[(i-1)^2 N^2 + (1-2\xi_n)(i-1)N\right] = \frac{1}{2\rho} \left(-\nu(1-\rho) + \frac{L(L^2-1)N^3}{6M} \right)$$

The purpose behind the variable ρ in this story is to manipulate the higher order terms in the expansion so that one obtains a more accurate approximation [37]. Namely, it is clear that if we choose the following value for ρ

$$\rho^* = 1 - \frac{L(L^2-1)N^3}{6M\nu} = 1 - \frac{(L+1)N}{6M},$$

then the first order term in the asymptotic expansion of $\psi_Z(t)$ can be made to vanish as $\omega_1(\rho^*) = 0$. This effectively produces an approximation whose error is $\mathcal{O}(M^{-2})$ compared to an error that is $\mathcal{O}(M^{-1})$ if one were to set $\rho = 1$. Letting M tend to infinity and exponentiating the resulting cumulant generating function, we find that

$$\phi_Z(jt) = e^{\psi_Z(jt)} \stackrel{M\to\infty}{\longrightarrow} (1-2jt)^{-\nu/2}$$

which one recognizes as the characteristic function of a chi-squared random variable with ν degrees of freedom. Thus, for large M it follows that [33]

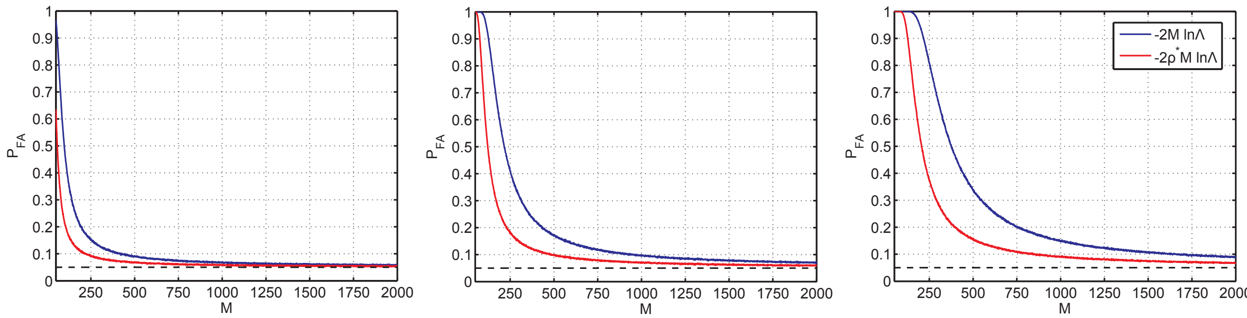
$$\begin{aligned} P\left[-2\rho^* M \ln\Lambda \le x\right] &= P\left[\left(\frac{1}{3}(L+1)N - 2M\right)\ln\Lambda \le x\right] \\ &\stackrel{M\to\infty}{\longrightarrow} P\left[\chi^2_\nu \le x\right];\ \nu = L^2N^2 - LN^2 \end{aligned}$$

Carefully counting the number of independent real parameters that must be estimated under the alternative and null hypotheses, respectively, one finds that $\dim(\mathcal{R}) = L^2N^2$ and $\dim(\mathcal{R}_0) = LN^2$. Thus, it is clear that the degrees of freedom ν associated with this chi-squared distribution can be related to the dimensions of these two spaces

$$\nu = \dim(\mathcal{R}) - \dim(\mathcal{R}_0)$$

in accordance with the result given in (12).

With $N = 12$, Figures 3.5 (a)-(c) display empirical false alarm probabilities versus M for the random variable $-2\rho M \ln\Lambda$ when $\rho = 1$ (shown by a darker line) and when $\rho = \rho^*$ (shown by a lighter line) for $L = 2,\ 3,$ and 4, respectively. Here the threshold is chosen from the asymptotic chi-squared distribution and set to achieve a desired false alarm rate of $P_{FA} = 0.05$ (shown by a dashed line). From the figure we can see that, by incorporating a scaling and allowing it to deviate from unity, one can achieve false alarm rates that are closer to the desired value for any finite M by selecting the appropriate value for ρ. Even



(a) $L = 2$ (b) $L = 3$ (c) $L = 4$

FIGURE 3.5. Asymptotic Empirical False Alarm Probabilities, $P\left[-2\rho M \ln \Lambda > \eta\right]$ for $\rho = 1$ and $\rho = \rho^*$.

for moderate values of L and N, however, one can also see that this approximation requires very large values for M to achieve convergence of the false alarm probability to the desired value of 0.05.

3.5. Conclusion

This chapter discusses the probabilistic behavior of the generalized Hadamard ratio under the null hypothesis when all L channels are truly independent. Using the theory of Gram determinants presented in Appendix A, it was shown that this test statistic can be written as a product of scalars regardless of what hypothesis is truly in force. Each term within this product represents the ratio of the squared residual associated with two linear least-squares problems which look to predict the random variable $x_i[n]$ from the sensory channels $\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}$ and the first n temporal samples of channel i, $x_i[0], \ldots, x_i[n-1]$. The numerator of the ratio, $\sigma_{in}^2(\hat{R})$, uses both $\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}$ and $x_i[0], \ldots, x_i[n-1]$ while the denominator, $\sigma_{in}^2(\hat{R}_{ii})$, uses only $x_i[0], \ldots, x_i[n-1]$. This result leads directly to the conclusion that, under the null hypothesis, the generalized Hadamard ratio is stochastically equivalent to a product of independent beta random variables - a stochastic representation which is solely dependent on the number of channels (L), the length of each time series (N), and the number of samples (M).

Although an appealing result, this stochastic representation fails to bring us any closer to finding simple, closed-form expressions for the PDF of the generalized Hadamard ratio under the null hypothesis. This is critical if a multichannel detector that uses this generalized Hadamard ratio is to achieve a desired false alarm probability. One of the more convenient results concerning the GLRT in general, however, is the asymptotic distribution it converges to as M grows large. Using the fact that the null distribution of the test is statistically equivalent to a product of betas makes it very straightforward to show that its characteristic function can be written as a product of ratios of gamma functions. Employing results concerning the asymptotic behavior of the log-gamma function, we showed that the log-likelihood ratio converges in distribution to a chi-squared random variable. Moreover, the degrees of freedom of this chi-squared distribution can be interpreted in terms of the dimensions of the parameter spaces $\mathcal{R}$ and $\mathcal{R}_0$ considered in the construction of the likelihood ratio. This result is significant as it allows one to set thresholds to achieve a desired false alarm in situations where M is large. Even with the inclusion of a scaling designed to improve the rate of convergence, however, results show that the approximation is slow to converge for even moderate values for L and N. For this reason, we will next turn our attention to the use of the saddlepoint method for approximating the PDF of the likelihood ratio under the null hypothesis.

CHAPTER 4

Approximating the Null Distribution using the Saddlepoint Method

4.1. Introduction

In the previous chapter, we discussed the null distribution of the GLRT developed in Chapter 2 and discovered that the likelihood ratio is stochastically equivalent to a product of independent beta random variables. Asymptotically, the log-likelihood ratio converges in distribution for any value of L and N to a chi-squared random variable as M grows large. Although theoretically interesting, results immediately show that, even for moderate values of L and N, one requires an overwhelming number of samples to achieve convergence to this asymptotic chi-squared distribution. For this reason, we turn our attention to the use of saddlepoint approximations for estimating this random variable's probability distribution when M is not so large.

Saddlepoint approximations [22] – [39] are powerful tools for obtaining accurate approximations for densities and distribution functions. The saddlepoint method accomplishes this by approximating the inverse Fourier transform of a random variable's known characteristic function. In this chapter, we will begin by giving a general overview of the saddlepoint approximation starting from simple Taylor series expansions and the Laplace approximation of integrals [39]. This general theory is subsequently applied to the null distribution of the GLRT using the characteristic function derived in Section 3.4. The ability of the saddlepoint approximation in achieving a desired false alarm rate as a function of M is demonstrated for various choices of L and N using Monte Carlo simulations and compared to the asymptotic result derived in Chapter 3.

The remainder of this chapter is organized as follows. In Section 4.2, we briefly review the theory of saddlepoint approximations and give several pedagogical examples. Section 4.3 applies this theory to the null distribution of the multichannel GLRT and compares its ability to match a desired false alarm rate to the asymptotic chi-squared distribution. Concluding remarks are finally given in Section 4.4.

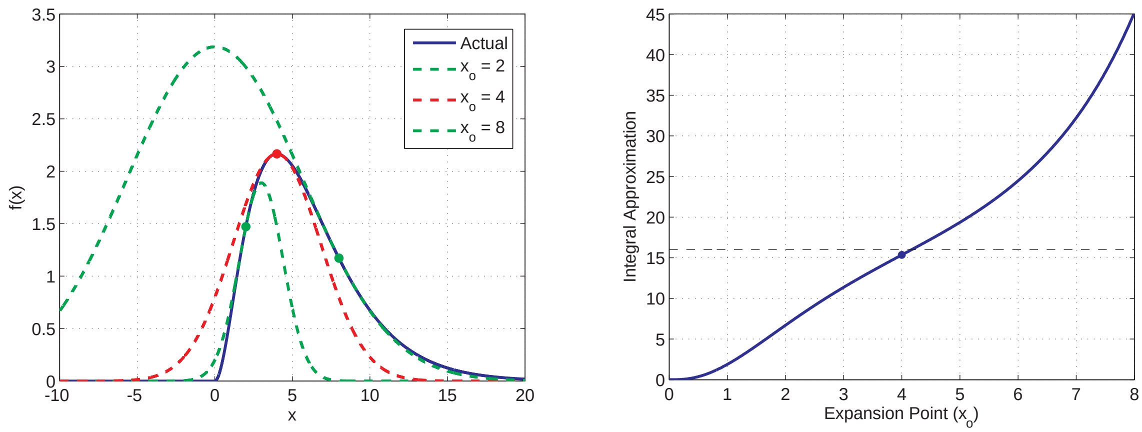
4.2. Saddlepoint Approximations

In this section, we will give a brief review of the saddlepoint method starting with fundamental Taylor series expansions and the Laplace approximation for integrals. Let $f(x)$ denote some twice continuously differentiable function with $f(x) \geq 0$. We can always approximate the value of this function in the neighborhood of the point x_o by retaining the first few terms (up to the 2^{nd} derivative) of the Taylor series expansion of its logarithm, $g(x) = \ln f(x)$

$$f(x) \approx \exp\left\{g(x_o) + g'(x_o)(x - x_o) + \frac{1}{2}g''(x_o)(x - x_o)^2\right\}$$

where $g'(x)$ and $g''(x)$ denote the first and second-order derivatives of $g(x)$, respectively. This expression simplifies even further if we choose $x_o = \hat{x}$ with $\hat{x}$ a stationary point of $g(x)$ such that $g'(\hat{x}) = 0$

$$f(x) \approx \exp\left\{g(\hat{x}) + \frac{1}{2}g''(\hat{x})(x - \hat{x})^2\right\}$$



(a) Function Approximation (b) Integral Approximation

FIGURE 4.1. Demonstration of the Laplace approximation with $f(x) = x^2 e^{-\frac{1}{2}x}$.

In addition to having well-established applications for function approximation, this idea is particularly useful for approximating the integral of $f(x)$ over the real line

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)dx &\approx \int_{-\infty}^{\infty} \exp\left\{g(\hat{x}) + \frac{1}{2}g''(\hat{x})(x-\hat{x})^2\right\} dx \\ &= \exp\{g(\hat{x})\} \int_{-\infty}^{\infty} \exp\left\{\frac{1}{2}g''(\hat{x})(x-\hat{x})^2\right\} dx \end{aligned} \tag{19}$$

If we further assume that $\hat{x}$ is a unique global maximum of $g(x)$ (and hence a unique global maximum of $f(x)$ as the logarithm is monotonic) with $g''(\hat{x}) < 0$, then one can see that the integral given in (19) is the integral of a Gaussian function with mean $\hat{x}$ and variance $-\frac{1}{g''(\hat{x})}$ so that

$$\int_{-\infty}^{\infty} f(x)dx \approx \sqrt{-\frac{2\pi}{g''(\hat{x})}} \exp\{g(\hat{x})\} \tag{20}$$

Thus to gain an approximation of the integral of the function $f(x)$, one simply requires the value of the function $g(\cdot)$ and its curvature $g''(\cdot)$ at the stationary point $\hat{x}$. This technique for approximating the integral of a function is commonly referred to as the *Laplace approximation* [39]. For the function $f(x) = x^2 e^{-\frac{1}{2}x}$, Figure 4.1 (a) displays a graph of the function along with several of its quadratic approximations with different choices in the expansion point x_o, one of them being the stationary point $x_o = \hat{x} = 4$. Figure 4.1 (b) likewise shows the area under each of these quadratic approximations as a function of the expansion point x_o and compares it to the true value of the integral which is shown with a dashed horizontal line. From these figures we can see that, by choosing the expansion point to correspond with this function's maximum, the approximation given in (20) gives an answer that is just shy of its true value.

In addition to being useful for approximating the integral of a function, this idea also finds its uses in approximating functions of the form

$$f(x) = \int_{-\infty}^{\infty} m(x,t)dt$$

where $m(x,t) \geq 0$. Assuming that, for any fixed value of x, the function $m(x,t)$ exhibits a unique global maximum in the dummy variable t and defining $k(x,t) = \ln m(x,t)$, one may take a very similar approach as before to derive the approximation

$$\begin{aligned} f(x) &\approx \int_{-\infty}^{\infty} \exp\left\{ k\left(x, \hat{t}(x)\right) + \frac{1}{2} \left.\frac{\partial^2 k(x,t)}{\partial t^2}\right|_{\hat{t}(x)} \left(t - \hat{t}(x)\right)^2 \right\} dt \\ &= \sqrt{-\frac{2\pi}{\left.\frac{\partial^2 k(x,t)}{\partial t^2}\right|_{\hat{t}(x)}}} \exp\left\{ k\left(x, \hat{t}(x)\right) \right\} \end{aligned} \qquad (21)$$

where $\hat{t}(x)$ is the value of t such that $\left.\frac{\partial k(x,t)}{\partial t}\right|_{\hat{t}(x)} = 0$ and $\left.\frac{\partial^2 k(x,t)}{\partial t^2}\right|_{\hat{t}(x)} < 0$. This expression is commonly referred to as the *saddlepoint approximation* [22], [38] and the value $\hat{t}(x)$ is referred to as the *saddlepoint*.

One application where the approximation in (21) finds its uses is in the approximation of the inverse Fourier transform of a random variable's known characteristic function [40].

Given the random variable X with moment generating function $\phi_X(t)$, one may obtain the random variable's PDF $f_X(x)$ by using the inversion formula

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_X(jt)\exp\{-jtx\}\,dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\{\psi_X(jt) - jtx\}\,dt \tag{22}$$

The function $\psi_X(t) = \ln \phi_X(t)$ is the cumulant generating function for the random variable X with the property that, similar to the moment generating function, one may obtain the random variables cumulants, κ_n for $n = 0, 1, \ldots$, by differentiating and evaluating at zero

$$\left.\frac{d^n \psi_X(t)}{dt^n}\right|_{t=0} = \kappa_n$$

Making a change of variable ($t' = jt$), one may express the inversion formula (22) as

$$f_X(x) = \frac{1}{j2\pi}\int_{\tau-j\infty}^{\tau+j\infty} \exp\{\psi_X(t') - t'x\}\,dt' \tag{23}$$

for $\tau = 0$. Let $\psi_X'(t) = \frac{d\psi_X(t)}{dt}$ and $\psi_X''(t) = \frac{d^2\psi_X(t)}{dt^2}$ denote the first and second-order derivative of the cumulant generating function, respectively. Then using the approximation given in (21) with $k(x,t) = \psi_X(t) - tx$, we first find the saddlepoint, the point $\hat{t}(x)$ such that $\psi_X'\left(\hat{t}(x)\right) = x$. Expanding the exponent of (23) around $\hat{t}(x)$ we have

$$k(x,t) \approx \psi_X\left(\hat{t}(x)\right) - \hat{t}(x)x - \frac{1}{2}\psi_X''\left(\hat{t}(x)\right)\left(t - \hat{t}(x)\right)^2$$

We then substitute this quadratic approximation into (23) and integrate along the line that runs parallel to the imaginary axis and passes through the saddlepoint by setting $\tau = \hat{t}(x)$ in the limits of the integral [38]. Applying expression (21), one finally obtains the saddlepoint approximation of the PDF $f_X(x)$

$$\begin{aligned} f_X(x) &\approx \frac{1}{j2\pi}\left(j\sqrt{\frac{2\pi}{\psi_X''\left(\hat{t}(x)\right)}}\exp\left\{\psi_X\left(\hat{t}(x)\right)-\hat{t}(x)x\right\}\right) \\ &= \frac{1}{\sqrt{2\pi\psi_X''\left(\hat{t}(x)\right)}}\exp\left\{\psi_X\left(\hat{t}(x)\right)-\hat{t}(x)x\right\} \end{aligned} \tag{24}$$

where, because of the complex integration, we now require $\hat{t}(x)$ to be a minimum with $\psi_X''\left(\hat{t}(x)\right) > 0$. We'll now take a look at several simple examples to demonstrate this approximation before moving on to the null distribution of the multichannel GLRT.

4.2.1. Normal Distribution. Suppose that $X \sim \mathcal{N}(\mu, \sigma^2)$ with PDF and cumulant generating function

$$\begin{aligned} f_X(x) &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\} \\ \psi_X(t) &= \mu t + \frac{1}{2}\sigma^2 t^2 \ ; \ t \in \mathbb{R} \end{aligned}$$

This cumulant generating function has the following first and second-order derivatives $\psi_X'(t) = \mu + \sigma^2 t$ and $\psi_X''(t) = \sigma^2$, respectively, from which we obtain the unique saddlepoint

$$\psi_X'(t) = x \ \Rightarrow \ \hat{t}(x) = \frac{x-\mu}{\sigma^2} \tag{25}$$

Evaluating the cumulant generating function and its second derivative at the saddlepoint

$$\begin{aligned} \psi_X\left(\hat{t}(x)\right) &= \mu\left(\frac{x-\mu}{\sigma^2}\right) + \frac{1}{2}\frac{(x-\mu)^2}{\sigma^2} \\ \psi_X''\left(\hat{t}(x)\right) &= \sigma^2 \end{aligned}$$

and substituting these expressions into (24), the saddlepoint approximation, which we will denote $\hat{f}_X(x)$, becomes

$$
\begin{aligned}
\hat{f}_X(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\mu\left(\frac{x-\mu}{\sigma^2}\right) + \frac{1}{2}\frac{(x-\mu)^2}{\sigma^2} - x\left(\frac{x-\mu}{\sigma^2}\right)\right\} \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\} \\
&= f_X(x)
\end{aligned}
$$

So, for a normal distribution the saddlepoint approximation is exact which is to be expected as the cumulant generating function in this case is a quadratic function of t.

4.2.2. Gamma Distribution. Suppose now that $X \sim \Gamma(k,\theta)$ with PDF and cumulant generating function

$$
\begin{aligned}
f_X(x) &= \frac{1}{\theta^k \Gamma(k)} x^{k-1} \exp\left\{-\frac{x}{\theta}\right\} \\
\psi_X(t) &= -k\ln(1-\theta t)\ ;\ t < \frac{1}{\theta}
\end{aligned}
$$

Simple calculus shows that the cumulant generating function for this example has the derivatives

$$
\begin{aligned}
\psi'_X(t) &= \frac{k\theta}{1-\theta t} \\
\psi''_X(t) &= \frac{k\theta^2}{(1-\theta t)^2}
\end{aligned}
$$

which leads to the unique saddlepoint

$$
\psi'_X(t) = x \ \Rightarrow\ \hat{t}(x) = \frac{1}{\theta} - \frac{k}{x}
$$

At the saddlepoint the cumulant generating function and its second derivative become

$$\begin{aligned}\psi_X\left(\hat{t}(x)\right) &= -k\ln\left(\frac{k\theta}{x}\right)\\ \psi_X''\left(\hat{t}(x)\right) &= \frac{x^2}{k}\end{aligned}$$

which yield the saddlepoint approximation

$$\begin{aligned}\hat{f}_X(x) &= \sqrt{\frac{k}{2\pi x^2}}\exp\left\{-k\ln\left(\frac{k\theta}{x}\right)-\left(\frac{1}{\theta}-\frac{k}{x}\right)x\right\}\\ &= \frac{1}{\theta^k\left(\sqrt{2\pi}k^{k-\frac{1}{2}}\exp\{-k\}\right)}x^{k-1}\exp\left\{-\frac{x}{\theta}\right\}\end{aligned}$$

Comparing this approximation to the expression for the actual PDF $f_X(x)$, one can see that the two are exact with the exception of the scaling term $\Gamma(k)$ in the actual expression being replaced by $\hat{\Gamma}(k) = \sqrt{2\pi}k^{k-1/2}\exp\{-k\}$. The scaling $\hat{\Gamma}(k)$ is in fact an approximation of $\Gamma(k)$ that improves with increasing k and is very closely related to the asymptotic expansion given in (17). Indeed, ignoring the sum containing Bernoulli polynomials, setting $z = k$, $a = 0$, and exponentiating (17), one obtains

$$\exp\{\ln\Gamma(k)\} \approx \exp\left\{\frac{1}{2}\ln 2\pi + (k-\frac{1}{2})\ln k - k\right\} = \sqrt{2\pi}k^{k-\frac{1}{2}}\exp\{-k\} = \hat{\Gamma}(k)$$

With $\theta = 1$ and differnt choices in k, Figure 4.2 plots the saddlepoint density for several gamma distributions and compares each to the actual PDF. For k equal to 1 and 2, one can see slight differences between each density and its approximation but as k increases to 4 one can see that the approximation is nearly exact.

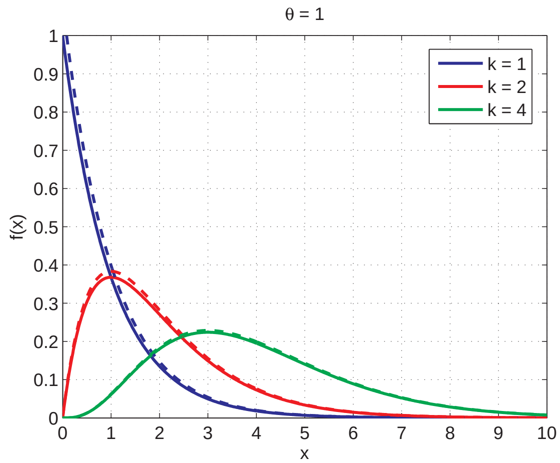


FIGURE 4.2. Actual (solid) and saddlepoint approximation (dashed) for several gamma densities.

4.2.3. LAPLACE DISTRIBUTION. Finally, we'll consider the case that $X \sim \text{Laplace}(\mu, \lambda)$ with PDF and cumulant generating function

$$
\begin{aligned}
f_X(x) &= \frac{1}{2\lambda} \exp\left\{-\frac{1}{\lambda}|x-\mu|\right\} \\
\psi_X(t) &= \mu t - \ln\left(1 - \lambda^2 t^2\right) \ ; \ |t| < \frac{1}{\lambda}
\end{aligned}
$$

As before, the first step involves taking the first and second-order derivative of this cumulant generating function leading to the expressions

$$
\begin{aligned}
\psi'_X(t) &= \mu + \frac{2\lambda^2 t}{1 - \lambda^2 t^2} \\
\psi''_X(t) &= 2\lambda^2 \frac{1 + \lambda^2 t^2}{\left(1 - \lambda^2 t^2\right)^2}
\end{aligned}
$$

Solving the saddlepoint equation in this example is equivalent to finding the roots of a second-order polynomial in t

$$
\psi'_X(t) = x \ \Rightarrow \ -(x-\mu)\lambda^2 t^2 - 2\lambda^2 t + (x-\mu) = 0
$$

in which case one finds the pair of solutions

$$\hat{t}(x) = -\frac{1}{x-\mu}\left(1 \pm \frac{\sqrt{\lambda^2 + (x-\mu)^2}}{\lambda}\right)$$

At first glance it would appear that we have come across an example where there exists no unique solution for the saddlepoint. However, recalling the fact that the cumulant generating function for this example exists only for arguments satisfying $|t| < \frac{1}{\lambda}$, one can use this condition to set a constraint on the saddlepoint equation and determine which solution is truly valid. We first of all note that the condition that $|\hat{t}(x)| < \frac{1}{\lambda}$ is equivalent to the condition that $\left|\lambda \pm \sqrt{\lambda^2 + (x-\mu)^2}\right| < |x-\mu|$. Using the subadditivity property of the square root ($\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$), one obtains the pair of inequalities

$$\begin{aligned} \left|\lambda + \sqrt{\lambda^2 + (x-\mu)^2}\right| &\leq 2\lambda + |x-\mu| \\ \left|\lambda - \sqrt{\lambda^2 + (x-\mu)^2}\right| &\leq |x-\mu| \end{aligned}$$

with equality when $x = \mu$. From these inequalities, it becomes clear that the valid saddlepoint is the one that involves subtraction

$$\hat{t}(x) = -\frac{1}{x-\mu}\left(1 - \frac{\sqrt{\lambda^2 + (x-\mu)^2}}{\lambda}\right)$$

Substituting this solution into (24) yields the saddlepoint approximation

$$\hat{f}_X(x) = \frac{1}{2\lambda}\sqrt{\frac{(x-\mu)^2}{2\pi\left(\lambda^2 + (x-\mu)^2 - \lambda\sqrt{\lambda^2 + (x-\mu)^2}\right)}}\exp\left\{1 - \frac{\sqrt{\lambda^2 + (x-\mu)^2}}{\lambda}\right\}$$

which bears little resemblance to the true expression for the PDF $f_X(x)$. For several choices in the parameters μ and λ, Figure 4.3 compares this saddlepoint approximation to the

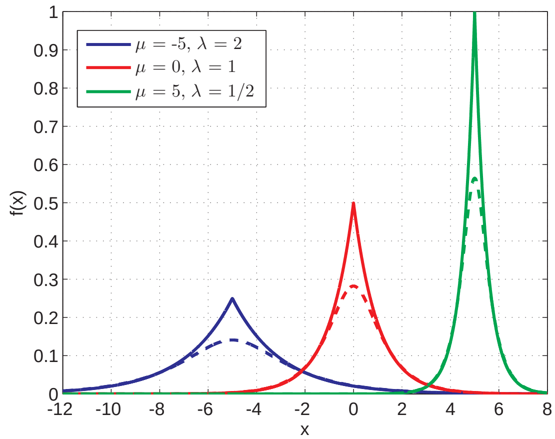


FIGURE 4.3. Actual (solid) and saddlepoint approximation (dashed) for several Laplace densities.

actual PDF. From these plots we can see that the saddlepoint approximation in this case has poor accuracy in a neighborhood of the mean but excels at approximating the tails of the distribution.

4.3. Application to the Multichannel GLRT

In this section, we discuss the saddlepoint approximation as it applies to the null distribution of the GLRT developed in Chapter 3. Recalling the arguments leading to expression (15), the generalized Hadamard ratio, Λ, has the following moments under the null hypothesis

$$\mu_\Lambda(k) = E\left[\Lambda^k | \mathcal{H}_0\right] = \prod_{i=2}^{L} \prod_{n=0}^{N-1} \frac{\Gamma(M-n)\Gamma(M-(i-1)N-n+k)}{\Gamma(M-(i-1)N-n)\Gamma(M-n+k)}$$

For any real-valued scalar c, the random variable $Z = c \ln \Lambda | \mathcal{H}_0$ has the moment generating function $\phi_Z(t)$

$$\phi_Z(t) = E\left[e^{tZ} | \mathcal{H}_0\right] = E\left[\Lambda^{ct} | \mathcal{H}_0\right] = \mu_\Lambda(ct) = \prod_{i=2}^{L} \prod_{n=0}^{N-1} \frac{\Gamma(M-n)\Gamma(M-(i-1)N-n+ct)}{\Gamma(M-(i-1)N-n)\Gamma(m-n+ct)}$$

from which we obtain the cumulant generating function

$$
\begin{aligned}
\psi_Z(t) &= \ln \phi_Z(t) \\
&= \sum_{i=2}^{L} \sum_{n=0}^{N-1} \left[\ln \Gamma(M-n) - \ln \Gamma(M-(i-1)N-n)\right] \\
&\quad + \sum_{i=2}^{L} \sum_{n=0}^{N-1} \left[\ln \Gamma(M-(i-1)N-n+ct) - \ln \Gamma(M-n+ct)\right]
\end{aligned}
$$

Letting $\gamma_k(x)$ denote the polygamma function [36] of order k

$$
\gamma_k(x) = \frac{d^{(k+1)} \ln \Gamma(x)}{dx^{(k+1)}}
$$

this cumulant generating function has k^{th}-order derivative

$$
\frac{d^k \psi_Z(t)}{dt^k} = c^k \sum_{i=2}^{L} \sum_{n=0}^{N-1} \left[\gamma_{k-1}\left(M-(i-1)N-n+ct\right) - \gamma_{k-1}\left(M-n+ct\right)\right]
$$

As described in Section 4.2, the first step is to find the saddlepoint, the value of t such that

$$
\psi'_Z(t) = c \sum_{i=2}^{L} \sum_{n=0}^{N-1} \left[\gamma_0\left(M-(i-1)N-n+ct\right) - \gamma_0\left(M-n+ct\right)\right] = x \tag{26}
$$

Unfortunately, for all but one particular value for x, there is no simple closed-form solution to this problem. That specific value of x corresponds to the approximation of the density at the mean of the distribution. Knowing that the first-order cumulant is the mean, i.e. $\psi'_Z(0) = E[Z] = \mu_Z$, at $x = \mu_Z$ it follows that the saddlepoint must be zero, $\hat{t}(\mu_Z) = 0$. However, this fact is not specific to this distribution and must be true in general.

In lieu of the fact that there is no closed-form solution to the saddlepoint equation, one must resort to numerical root-finding algorithms to find the saddlepoint. One very simple method for accomplishing this is the Newton–Raphson method where, for every value of x, one calculates the sequence of iterations, $\hat{t}_n(x)$ for $n = 0, 1, \ldots$, satisfying the recursive relationship

$$\hat{t}_{n+1}(x) = \hat{t}_n(x) - \frac{\psi_Z'\left(\hat{t}_n(x)\right) - x}{\psi_Z''\left(\hat{t}_n(x)\right)}$$

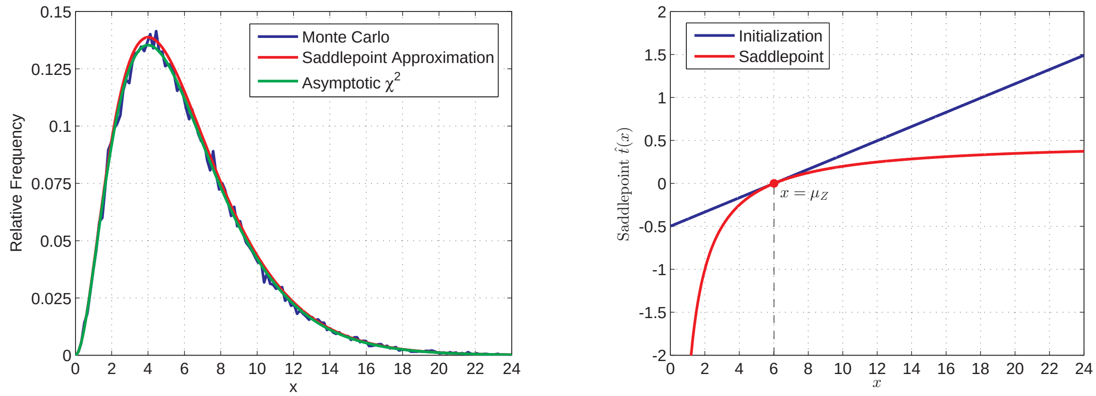
One of the things required by the Newton–Raphson method is an initial guess at what the saddlepoint may be. One possible choice for this situation is to use the saddlepoint associated with a normal distribution with known mean and variance to initialize the search. Using properties of the cumulant generating function, the random variable Z has mean and variance

$$\begin{aligned}
\mu_Z &= \psi_Z'(0) = c \sum_{i=2}^{L} \sum_{n=0}^{N-1} \left[\gamma_0(M - (i-1)N - n) - \gamma_0(M - n)\right] \\
\sigma_Z^2 &= \psi_Z''(0) = c^2 \sum_{i=2}^{L} \sum_{n=0}^{N-1} \left[\gamma_1(M - (i-1)N - n) - \gamma_1(M - n)\right]
\end{aligned}$$

Recalling the result given in (25), one can then initialize the saddlepoint using the following affine function of x

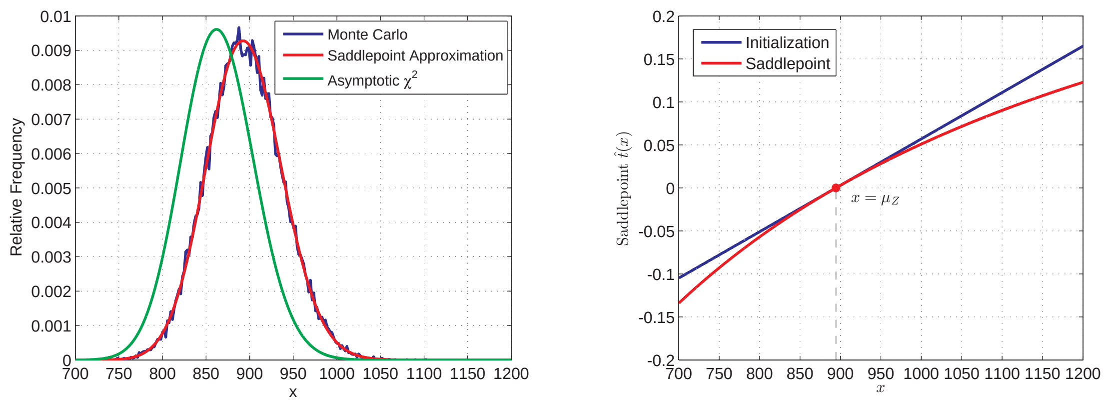
$$\hat{t}_0(x) = \frac{x - \mu_Z}{\sigma_Z^2}$$

Scaling the log-likelihood ratio by the factor $c = \frac{1}{3}(L+1)N - 2M$ to compare the results of the saddlepoint approximation to those of the asymptotic distribution developed in Section 3.4, Figure 4.4 (a) plots the saddlepoint approximation and compares it to the asymptotic chi-squared distribution and a Monte Carlo simulation using (11) when $L = 3$, $N = 1$ and $M = 250$. From this plot, one can see that the asymptotic chi-squared distribution accurately captures the distribution of the random variable Z while one sees a slight deviation in the saddlepoint approximation. Figure 4.4 (b) shows the initial saddlepoint $\hat{t}_0(x)$ and the actual



(a) Density Approximation (b) Saddlepoint

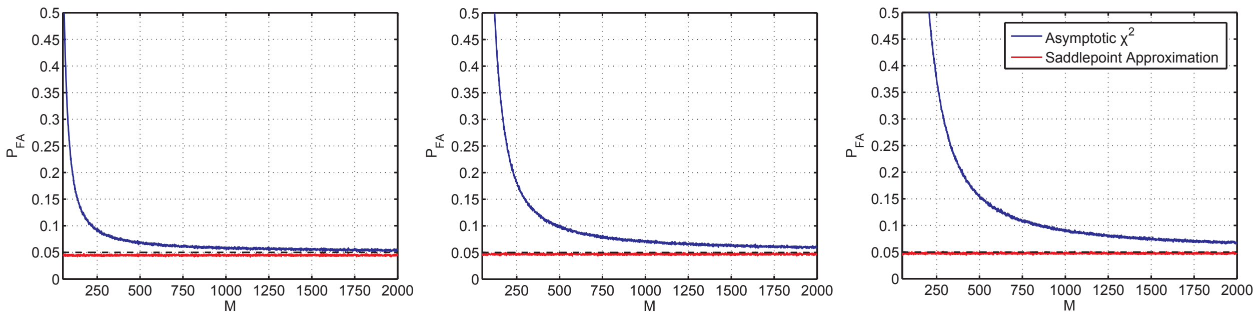
FIGURE 4.4. Saddlepoint approximation and the saddlepoint with $L = 3$, $N = 1$, and $M = 250$.



(a) Density Approximation (b) Saddlepoint

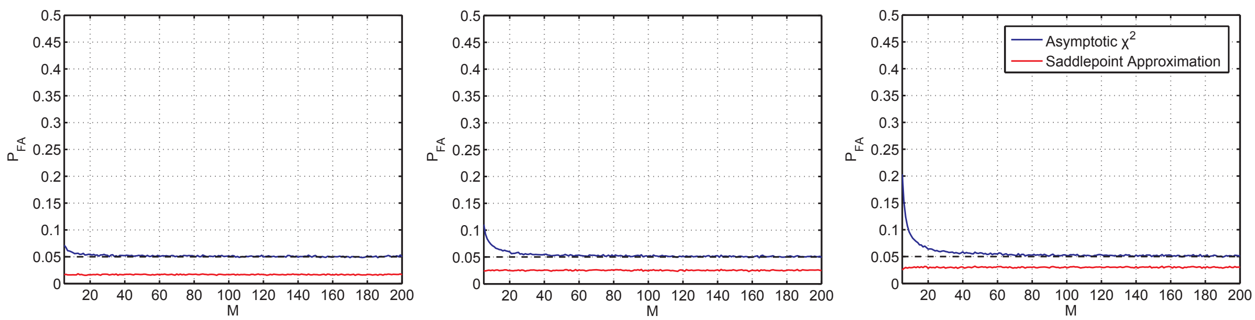
FIGURE 4.5. Saddlepoint approximation and the saddlepoint with $L = 3$, $N = 12$, and $M = 250$.

saddlepoint found using the Newton–Raphson recursions. From this plot one can see that the initialization $\hat{t}_0(x)$ is not a bad guess of the saddlepoint when considering values of x that are within a neighborhood of the mean μ_Z (shown by a vertical dashed line in the figure). However, the farther one gets into the tails of the distribution, the larger the deviation between the initialization and the true saddlepoint. For comparison, Figure 4.5 displays the same plots as Figure 4.4 but with a time series of length $N = 12$. From this figure we can see that the saddlepoint approximation is much more capable of capturing the distribution of the log-likelihood ratio compared to the asymptotic chi-squared distribution in this scenario.



(a) $L = 2$ (b) $L = 3$ (c) $L = 4$

FIGURE 4.6. Asymptotic and saddlepoint density false alarm probabilities with $N = 12$.



(a) $L = 2$ (b) $L = 3$ (c) $L = 4$

FIGURE 4.7. Asymptotic and saddlepoint density false alarm probabilities with $N = 1$.

Using the saddlepoint approximation to determine the threshold needed to achieve a false alarm rate of $P_{FA} = 0.05$, Figure 4.6 plots empirical false alarm probabilities versus M with $N = 12$ and several choices in L. These plots are compared to the same results pertaining to the asymptotic chi-squared distribution with $\rho = \rho^* = 1 - \frac{(L+1)N}{6M}$ that were given in Figure 3.5. In these three plots, it is clear that choosing a threshold using the saddlepoint approximation is much more capable of achieving the desired false alarm probability. Finally, Figure 4.7 shows the same plots as those shown in Figure 4.6 but with $N = 1$. In this situation, choosing a threshold based on the asymptotic chi-squared distribution is much more practical and, for all but the smallest values for M, tends to produce a false alarm rate that is closer to what is desired.

4.4. Conclusion

In this chapter, we extended the results of Chapter 3 by investigating the saddlepoint approximation and its application to the null distribution of the generalized Hadamard ratio. The saddlepoint approximation can be interpreted as a straightforward extension of the Laplace method for approximating integrals. In statistics, one of the applications of the saddlepoint method is the approximation of the inverse Fourier transform of a random variable's known characteristic function. Determining the saddlepoint approximation requires finding the unique, real root of the saddlepoint equation, an equation involving the derivative of the cumulant generating function and the point x at which we wish to approximate the PDF.

The general theory of the saddlepoint approximation is then applied to the multichannel GLRT in Section 2.3. Using the fact that the generalized Hadamard ratio is stochastically equivalent to a product of independent beta random variables under the null hypothesis, it becomes very straightforward to derive the cumulant generating function of any scale multiple of the log-likelihood ratio. This results in an expression for the cumulant generating function which contains a sum of log-gamma functions and derivatives that consist of a sum of polygamma functions. These expressions lead to a saddlepoint equation which unfortunately cannot be solved in closed-form. Numerical root-finding algorithms such as the Newton-Raphson method must then be employed instead. The saddlepoint approximation is compared to the asymptotic chi-squared distribution developed in Section 3.4 to what improvement it brings in terms of matching a desired false alarm probability. These results show that the saddlepoint approximation is a very useful alternative for capturing the distribution of the log-likelihood ratio under the null hypothesis, especially in situations where L or N is large. If both L and N are small, however, results show that the asymptotic chi-squared distribution is probably more practical in these scenarios and tends to produce false alarm rates closer to what is desired.

CHAPTER 5

Multichannel Detection for 2D WSS Processes

5.1. Introduction

In this chapter, we turn our attention away from the null distribution of the generalized Hadamard ratio and discuss an alternative implementation of this test statistic for wide-sense stationary (WSS) random processes. Recall from Section 2.3 that, in many multichannel detection problems, $L \geq 2$ sensors are used to collect a length-N time series at multiple distinct spatial locations. To test for independence among all L channels, the generalized Hadamard ratio in (6) relies on an unconstrained ML estimate of the space-time covariance matrix R, i.e. a matrix with no constraints imposed on its structure other than Hermitian symmetry and positive definiteness. While making as few assumptions about the structure of R can be advantageous, a clear disadvantage is the large number of parameters that must be estimated as a result. In fact, we noted in Section 3.4 that $\dim(\mathcal{R}) = L^2N^2$, i.e. the number of parameters in the set of PD Hermitian matrices $\mathcal{R}$ grows quadratically in both the number of channels (L) and the length of each channel's time series (N). When the collection of time series $\{\mathbf{x}_i\}_{i=1}^{L}$ are jointly WSS, every $N \times N$ block $R_{ik} = E\left[\mathbf{x}_i\mathbf{x}_k^H\right]$ of matrix R will assume the form of a Toeplitz matrix [21]. Thus, in such situations it makes sense to develop methods that exploit this structure for multichannel detection.

Unfortunately, ML estimation of Toeplitz matrices [41] is an intractable problem with no closed-form solution. However, the asymptotic behavior of Toeplitz matrices as N grows large is well understood and leads to very tractable results involving the eigenvalues, multiplication, and inversion of large Toeplitz matrices [23]. When all L time series are jointly WSS, this asymptotic theory, and more specifically its extension to block-Toeplitz matrices, shows that the generalized Hadamard ratio developed in Section 2.3 converges to the broadband integral of a narrowband Hadamard ratio, a test statistic referred to as *broadband coherence* [7]. We then turn our attention to the problem of multichannel detection for 2-dimensional WSS processes. In this case, every covariance matrix $R_{ik} = E\left[\mathbf{x}_i \mathbf{x}_k^H\right]$ will assume the form of a Toeplitz-block-Toeplitz matrix, i.e. a matrix with constant diagonal blocks with the blocks themselves being Toeplitz. As the number of blocks in each of these matrices as well as the size of the block itself goes to infinity, the same asymptotic theory on large Toeplitz matrices shows that the generalized Hadamard ratio can again be expressed as a *broadband coherence* statistic, although in this case computed over a 2-dimensional frequency spectrum. Several applications where this test statistic may apply, namely multichannel detection in a network of sensor arrays and in coregistered images, are then demonstrated through simulation.

The remainder of this chapter is organized as follows. In Section 5.2, we briefly review the generalized Hadamard ratio in the frequency domain and the development of the broadband coherence statistic. Section 5.3 extends the broadband coherence detector to the case of 2-dimensional WSS processes and demonstrates its performance through simulation in Section 5.4. Concluding remarks are finally given in Section 5.5.

5.2. Generalized Hadamard Ratio in the Frequency Domain

In this section we give a review of the generalized Hadamard ratio given in (6) implemented in the frequency domain and the development of the broadband coherence statistic as presented in [7]. Recall the problem setup considered in Section 2.3 consisting of L spatially distributed sensors with each sensor collecting the length-N time series $\mathbf{x}_i = [x_i[0] \ \cdots \ x_i[N-1]]^T \in \mathbb{C}^N$ as depicted in Figure 5.1. The goal of this analysis is the development of an alternative detection technique that exploits the inherent Toeplitz structure

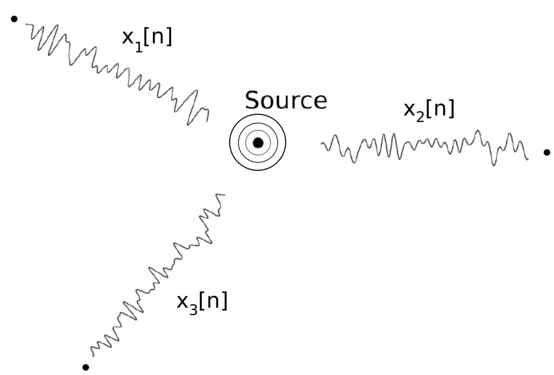


FIGURE 5.1. The collection of multiple time series at several distinct locations.

that each cross-covariance matrix $R_{ik} = E\left[\mathbf{x}_i \mathbf{x}_k^H\right]$ will assume when the collection of time series $\{\mathbf{x}_i\}_{i=1}^L$ are jointly WSS.

As described in [7], the extension of the likelihood ratio given in (6) can be accomplished by applying the linear transformation $T = F_N$ to the data from each channel. Here, the matrix F_N denotes an $N \times N$ Discrete Fourier Transform (DFT) matrix with entries $[F_N]_{\ell,k} = \frac{1}{\sqrt{N}} e^{-j2\pi \ell k/N}$. Recalling the first invariance property discussed in Section 2.3.1, it follows that both sets of signals, $\{\mathbf{x}_i\}$ and $\{F_N \mathbf{x}_i\}$, share the same likelihood ratio so that (6) can be written

$$\Lambda = \det\left((F_N \otimes I_L)\hat{C}(F_N \otimes I_L)^H\right)$$

where $\hat{C} = \hat{D}^{-1/2}\hat{R}\hat{D}^{-H/2}$ is the coherence matrix defined in Section 2.3. Introducing a simple permutation to the rows and columns of the matrix inside the determinant of this matrix, the likelihood ratio may be rewritten as $\Lambda = \det \tilde{C}$ where

$$\tilde{C} = \begin{bmatrix} \hat{C}(e^{j\theta_0}) & \hat{C}(e^{j\theta_0}, e^{j\theta_1}) & \cdots & \hat{C}(e^{j\theta_0}, e^{j\theta_{N-1}}) \\ \hat{C}(e^{j\theta_1}, e^{j\theta_0}) & \hat{C}(e^{j\theta_1}) & \cdots & \hat{C}(e^{j\theta_1}, e^{j\theta_{N-1}}) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{C}(e^{j\theta_{N-1}}, e^{j\theta_0}) & \hat{C}(e^{j\theta_{N-1}}, e^{j\theta_1}) & \cdots & \hat{C}(e^{j\theta_{N-1}}) \end{bmatrix} \tag{27}$$

is a global coherence matrix in the frequency domain with frequency $\theta_\ell = \frac{2\pi\ell}{N}$ for $\ell = 0, \ldots, N-1$ and we have used the convention $\hat{C}(e^{j\theta_\ell}) = \hat{C}(e^{j\theta_\ell}, e^{j\theta_\ell})$. The elements of each $L \times L$ block $\hat{C}(e^{j\theta_\ell}, e^{j\theta_m})$ of this global coherence matrix are given by

$$\left[\hat{C}(e^{j\theta_\ell}, e^{j\theta_m})\right]_{i,k} = \mathbf{f}_N^H(e^{j\theta_\ell}) \hat{R}_{ii}^{-1/2} \hat{R}_{ik} \hat{R}_{kk}^{-H/2} \mathbf{f}(e^{j\theta_m})$$

where $\mathbf{f}_N(e^{j\theta})$ denotes a length-N DFT vector

$$\mathbf{f}_N(e^{j\theta}) = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & e^{j\theta} & e^{j2\theta} & \cdots & e^{j(N-1)\theta} \end{bmatrix}^T$$

That is, if we define the DFT of the "whitened" random vector of each channel

$$w_i^{(m)}(e^{j\theta_\ell}) = \mathbf{f}_N^H(e^{j\theta_\ell}) \hat{R}_{ii}^{-1/2} \mathbf{x}_i[m]$$

then the $(i,k)^{\text{th}}$ element of matrix $\hat{C}(e^{j\theta_\ell}, e^{j\theta_m})$ represents the sample cross-covariance between the random variable for channel i at frequency θ_ℓ and the random variable for channel k at frequency θ_m

$$\left[\hat{C}(e^{j\theta_\ell}, e^{j\theta_m})\right]_{i,k} = \frac{1}{M} \sum_{m=1}^{M} w_i^{(m)}(e^{j\theta_\ell}) \left(w_k^{(m)}(e^{j\theta_m}) \right)^*$$

So far, taking the determinant of the global coherence matrix in (27) is no different than computing the generalized Hadamard ratio in (6). We now assume that the collection of time series $\{\mathbf{x}_i\}_{i=1}^{L}$ is jointly WSS so that, for any pair of channels i and k, there exists a complex-valued sequence $\{\gamma_{ik}[\ell]\}$ such that

$$E\left[x_i[n+\ell]x_k^*[n]\right] = \gamma_{ik}[\ell] \in \mathbb{C}$$

in which case the cross-covariance matrix $R_{ik} = E\left[\mathbf{x}_i\mathbf{x}_k^H\right] \in \mathbb{C}^{N\times N}$ is Toeplitz. Invoking results on large block-Toeplitz matrices and their asymptotic equivalence with block-circulant matrices [42], the global coherence matrix in (27) is asymptotically block-diagonal

$$\tilde{C} \to \text{blkdiag}\left\{\hat{C}(e^{j\theta_0}), \ldots, \hat{C}(e^{j\theta_{N-1}})\right\}$$

so that as N and M grow large but L remains fixed the generalized Hadamard ratio in (6) converges to the *broadband coherence* statistic [7]

$$\Lambda^{1/N} \overset{N\to\infty}{\longrightarrow} \exp\left\{\int_{-\pi}^{\pi} \ln\det\hat{C}(e^{j\theta})\frac{d\theta}{2\pi}\right\} = \exp\left\{\int_{-\pi}^{\pi} \ln\frac{\det\hat{S}(e^{j\theta})}{\prod_{i=1}^{L}\hat{S}_{ii}(e^{j\theta})}\frac{d\theta}{2\pi}\right\} \tag{28}$$

with $\hat{S}(e^{j\theta})$, $-\pi < \theta \leq \pi$, an estimated cross-spectral matrix

$$\hat{S}(e^{j\theta}) = \begin{bmatrix} \hat{S}_{11}(e^{j\theta}) & \hat{S}_{12}(e^{j\theta}) & \cdots & \hat{S}_{1L}(e^{j\theta}) \\ \hat{S}_{12}^*(e^{j\theta}) & \hat{S}_{22}(e^{j\theta}) & \cdots & \hat{S}_{2L}(e^{j\theta}) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{S}_{1L}^*(e^{j\theta}) & \hat{S}_{2L}^*(e^{j\theta}) & \cdots & \hat{S}_{LL}(e^{j\theta}) \end{bmatrix} \in \mathbb{C}^{L\times L}$$

Here, $\hat{S}_{ik}(e^{j\theta}) = \mathbf{f}_N^H(e^{j\theta})\hat{R}_{ik}\mathbf{f}_N(e^{j\theta})$ is a quadratic estimator of the cross power spectrum between channels i and k at frequency θ.

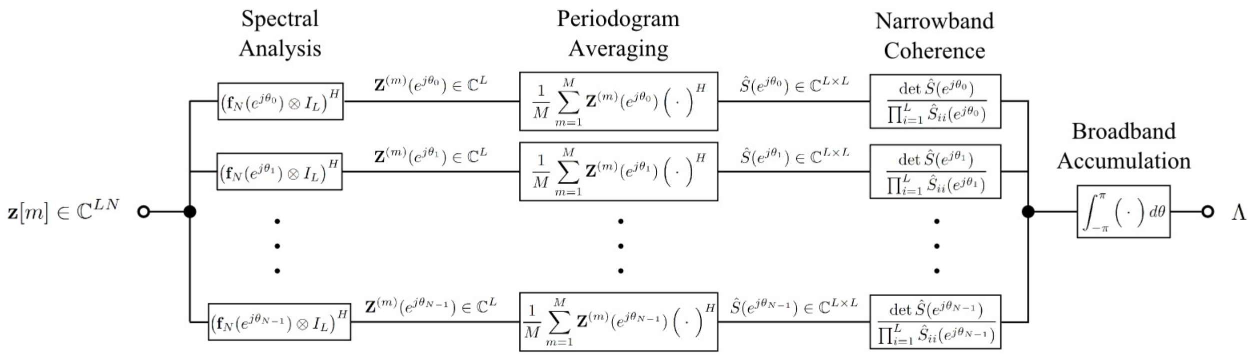


FIGURE 5.2. Formation of the broadband coherence statistic given in (28).

Figure 5.2 gives a block diagram implementation of the broadband coherence statistic. Here, each sample of the composite vector $\mathbf{z}[m]$ is passed through a bank of analysis filters, each producing the composite vector $\mathbf{Z}^{(m)}(e^{j\theta_\ell})$ at frequency $\theta_\ell = \frac{2\pi\ell}{N}$

$$\begin{aligned}
\mathbf{Z}^{(m)}(e^{j\theta_\ell}) &= \left(\mathbf{f}_N(e^{j\theta_\ell}) \otimes I_L\right)^H \mathbf{z}[m] = \left[X_1^{(m)}(e^{j\theta_\ell}) \ \cdots \ X_L^{(m)}(e^{j\theta_\ell})\right]^T \in \mathbb{C}^L \\
X_i^{(m)}(e^{j\theta_\ell}) &= \mathbf{f}_N^H(e^{j\theta_\ell})\mathbf{x}_i[m]
\end{aligned}$$

The periodograms for all M samples are subsequently averaged to produce the estimated cross-spectral matrix $\hat{S}(e^{j\theta_\ell})$

$$\hat{S}(e^{j\theta_\ell}) = \frac{1}{M}\sum_{m=1}^{M} \mathbf{Z}^{(m)}(e^{j\theta_\ell}) \left(\mathbf{Z}^{(m)}(e^{j\theta_\ell})\right)^H \in \mathbb{C}^{L\times L}$$

The narrowband coherence at each frequency is then measured using the Hadamard ratio $\det \hat{S}(e^{j\theta_\ell}) / \prod_{i=1}^{L} \hat{S}_{ii}(e^{j\theta_\ell})$ and accumulated through broadband integration to yield the test statistic given in (28). Although the true GLRT for this problem would enforce a Toeplitz structure on the ML estimates of each $N \times N$ block R_{ik} of matrix R, this is again a nontrivial problem with no closed-form solution. On the other hand, the broadband coherence statistic is computationally efficient and easy to implement. Moreover, by using only the diagonal

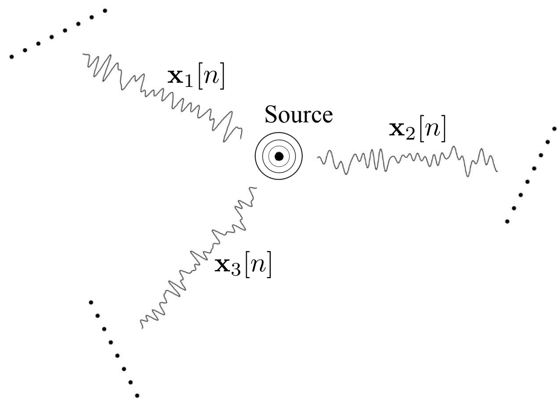


FIGURE 5.3. Detection of a common source using several distributed sensor arrays.

blocks of the coherence matrix given in (27), the broadband coherence statistic only requires the estimation of L^2N parameters (L^2 parameters for each cross-spectral matrix $\hat{S}(e^{j\theta})$ computed at N discrete frequencies) versus the L^2N^2 parameters required by the unconstrained covariance estimators used in the generalized Hadamard ratio of (6).

5.3. Extensions to 2D WSS Processes

In certain examples of multi-channel detection applications, one may have the opportunity of collecting multiple time series from each channel when deciding if a phenomenon common to all channels exists. One such example is a problem where several platforms each employ an array of sensors to take advantage of the spatial diversity such a sensing paradigm might have [43]. Consider the setup depicted in Figure 5.3 consisting of L spatially distributed sensor arrays observing a common event. In this case, we assume that the random vector from each channel $\mathbf{x}_i = \left[\mathbf{x}_i^T[0] \ \cdots \ \mathbf{x}_i^T[N-1]\right]^T \in \mathbb{C}^{NP}$ now contains N samples of a P-dimensional vector-valued time series $\mathbf{x}_i[n] = [x_i[n,0] \ \cdots \ x_i[n,P-1]]^T \in \mathbb{C}^P$ where $x_i[n,p]$ corresponds to the n^{th} temporal sample collected by the p^{th} sensor of the i^{th} array. As in Chapter 2, the goal is to test for the independence among the random vectors $\{\mathbf{x}_i\}_{i=1}^L$ using the same likelihood ratio given in (6). The only difference is that each block of the composite covariance matrix, R_{ik}, is now $NP \times NP$ rather than $N \times N$. In the next two sections we will discuss several extensions of the broadband coherence statistic (28) for this problem.

5.3.1. Frequency Domain GLRT. We will first begin by discussing the direct extension of the likelihood ratio given in (28) for the type of data collection scenario considered in Figure 5.3. Similar to that described in Section 5.2, the GLRT can be extended to the frequency domain by first independently applying the linear transformation $T = F_N \otimes I_P$, with F_N again denoting an $N \times N$ DFT matrix, to the data from each channel. Defining the matrix $F_N(e^{j\theta_\ell}) = \mathbf{f}_N(e^{j\theta_\ell}) \otimes I_P \in \mathbb{C}^{NP \times P}$, note that the linear transformation

$$F_N^H(e^{j\theta_\ell})\mathbf{x}_i = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} e^{-j\theta_\ell n}\mathbf{x}_i[n]$$

simply corresponds to a unitary DFT analysis at frequency $\theta_\ell = \frac{2\pi\ell}{N}$, $\ell = 0, \ldots, N-1$, applied temporally to all P sensors of the i^{th} channel. As the linear transformation $T = F_N \otimes I_P$ is an invertible matrix, the first invariance property in Section 2.3.1 guarantees that both sets of signals, $\{\mathbf{x}_i\}$ and $\{T\mathbf{x}_i\}$, share the same likelihood ratio in which case (6) can be written

$$\Lambda = \det\left((I_L \otimes T)\hat{C}(I_L \otimes T)^H\right)$$

where $\hat{C}$ is the coherence matrix defined in (6). As in Section 5.2, a permutation to the rows and columns of the matrix inside this determinant may be introduced so that the GLRT can be written

$$\Lambda = \det \tilde{C}$$

where

$$\tilde{C} = \begin{bmatrix} \hat{C}(e^{j\theta_0}) & \cdots & \hat{C}(e^{j\theta_0}, e^{j\theta_{N-1}}) \\ \vdots & \ddots & \vdots \\ \hat{C}(e^{j\theta_{N-1}}, e^{j\theta_0}) & \cdots & \hat{C}(e^{j\theta_{N-1}}) \end{bmatrix}$$

is a global coherence matrix in the frequency domain similar to that defined in (27). However, in this case the matrix $\hat{C}(e^{j\theta_\ell}, e^{j\theta_m}) \in \mathbb{C}^{PL\times PL}$ is an $L \times L$ block matrix consisting of $P \times P$ blocks of the form

$$\left\{\hat{C}(e^{j\theta_\ell}, e^{j\theta_m})\right\}_{i,k} = F_N^H(e^{j\theta_\ell})\hat{C}_{ik}F_N(e^{j\theta_m})$$

where $\hat{C}_{ik} = \hat{R}_{ii}^{-1/2}\hat{R}_{ik}\hat{R}_{kk}^{-H/2}$ and we use the convention $\hat{C}(e^{j\theta_\ell}) = \hat{C}(e^{j\theta_\ell}, e^{j\theta_\ell})$.

We now assume that all channels are temporally WSS in the sense that, for any pair of channels $\mathbf{x}_i$ and $\mathbf{x}_k$, there exists a matrix-valued covariance sequence, $\{\Gamma_{ik}[\ell]\}$, such that

$$E\left[\mathbf{x}_i[n+\ell]\mathbf{x}_k^H[n]\right] = \Gamma_{ik}[\ell] \in \mathbb{C}^{P\times P}$$

with no assumptions made about the structure of these matrices. Again using results on large block-Toeplitz matrices [42], the matrix $\tilde{C}$ becomes asymptotically equivalent to the block-diagonal matrix

$$\tilde{C} \to \operatorname{blkdiag}\left\{\hat{C}(e^{j\theta_0}), \ldots, \hat{C}(e^{j\theta_{N-1}})\right\}$$

so that as N and M grow large but P and L remain fixed the GLRT becomes

$$\begin{aligned} \Lambda^{1/N} &\xrightarrow{N\to\infty} \exp\left\{\int_{-\pi}^{\pi} \ln\det\hat{C}(e^{j\theta})\frac{d\theta}{2\pi}\right\} \\ &= \exp\left\{\int_{-\pi}^{\pi} \ln\frac{\det\hat{S}(e^{j\theta})}{\prod_{i=1}^{L}\det\hat{S}_{ii}(e^{j\theta})}\frac{d\theta}{2\pi}\right\} \end{aligned} \tag{29}$$

Here, the matrix $\hat{S}(e^{j\theta}) \in \mathbb{C}^{PL\times PL}$ is an $L \times L$ block-structured matrix consisting of $P \times P$ submatrices of the form

$$\left\{\hat{S}(e^{j\theta})\right\}_{i,k} = F_N^H(e^{j\theta})\hat{R}_{ik}F_N(e^{j\theta})$$

which is a quadratic estimate of the cross-power spectral density matrix between channels i and k at frequency θ and we use the convention $\hat{S}_{ii}(e^{j\theta}) = \left\{\hat{S}(e^{j\theta})\right\}_{i,i}$. In other words, the matrix $\hat{S} = \left\{\hat{S}_{ik}\right\}_{i,k}$ for $i, k = 1, \ldots, L$ is a cross-spectral matrix of cross-spectral matrices, with $\hat{S}_{ik}$ the cross-spectral matrix $\left[\hat{S}_{ik}^{\ell m}\right]_{\ell,m}$ for $\ell, m = 1, \ldots, P$.

5.3.2. Frequency/Wavenumber Domain GLRT. The likelihood ratio given in (29) is a direct extension of the results in [7] to account for the situation being considered here and is not a particularly interesting result in that it simply corresponds to replacing every scalar-valued power spectral density estimate in (28) with a $P \times P$ matrix. Although this result is perfectly general and nothing has been assumed about these vector-valued time series other than that they are temporally WSS, we proceed under the context of multiple-array detection in which case a notion of space can be ascribed to the time series of each channel.

To take advantage of the spatiotemporal properties of the problem, we now consider independently applying the linear transformation $T = F_N \otimes F_P$ to each channel (instead of the matrix $T = F_N \otimes I_P$ considered earlier) with F_P denoting a $P \times P$ DFT matrix. Note that pre-multiplying the vector $\mathbf{x}_i$ by the matrix T simply corresponds to the application of a 2-dimensional DFT, one applied temporally and the other spatially as opposed to previously where the DFT was only applied temporally. For any frequency θ, we can then introduce a permutation of the rows and columns of the previously defined matrix $\hat{C}(e^{j\theta})$ so that

$$\det \hat{C}(e^{j\theta}) = \det \tilde{C}(e^{j\theta}) \tag{30}$$

where

$$\tilde{C}(e^{j\theta}) = \begin{bmatrix} \hat{C}(e^{j\theta}, e^{j\phi_0}) & \cdots & \hat{C}(e^{j\theta}, e^{j\phi_0}, e^{j\phi_{P-1}}) \\ \vdots & \ddots & \vdots \\ \hat{C}(e^{j\theta}, e^{j\phi_{P-1}}, e^{j\phi_0}) & \cdots & \hat{C}(e^{j\theta}, e^{j\phi_{P-1}}) \end{bmatrix} \tag{31}$$

and $\hat{C}(e^{j\theta}, e^{j\phi_\ell}) = \hat{C}(e^{j\theta}, e^{j\phi_\ell}, e^{j\phi_\ell})$. Define the length-$P$ DFT vector at frequency $\phi_\ell = \frac{2\pi\ell}{P}$ for $\ell = 0, \ldots, P-1$ as follows

$$\mathbf{f}_P(e^{j\phi_\ell}) = \frac{1}{\sqrt{P}} \left[1 \ e^{j\phi_\ell} \ e^{j2\phi_\ell} \ \cdots \ e^{j(P-1)\phi_\ell}\right]^T.$$

Then the matrix $\hat{C}(e^{j\theta}, e^{j\phi_\ell}, e^{j\phi_m}) \in \mathbb{C}^{L\times L}$ has entries of the form

$$\left[\hat{C}(e^{j\theta}, e^{j\phi_\ell}, e^{j\phi_m})\right]_{i,k} = \mathbf{f}_P^H(e^{j\phi_\ell}) F_N^H(e^{j\theta}) \hat{C}_{ik} F_N(e^{j\theta}) \mathbf{f}_P(e^{j\phi_m}) \tag{32}$$

When the entries of $\mathbf{x}_i[n]$ correspond to time series at different spatial locations, the frequency variable ϕ is often referred to as the *wavenumber* and, to avoid confusion with the frequency variable θ, we will adopt this terminology. Similar to 5.2, if we define the 2D DFT of the "whitened" random vector at frequency θ and wavenumber ϕ

$$w_i^{(m)}(e^{j\theta}, e^{j\phi}) = \mathbf{f}_P^H(e^{j\phi}) F_N^H(e^{j\theta}) \hat{R}_{ii}^{-1/2} \mathbf{x}_i[m]$$

then (32) represents the sample cross-covariance

$$\left[\hat{C}(e^{j\theta}, e^{j\phi_\ell}, e^{j\phi_m})\right]_{i,k} = \frac{1}{M} \sum_{m=1}^{M} w_i^{(m)}(e^{j\theta}, e^{j\phi_\ell}) \left(w_k^{(m)}(e^{j\theta}, e^{j\phi_m})\right)^*$$

We now impose additional structure on the problem at hand by assuming that all channels are not only temporally WSS but spatially WSS as well so that the multivariate covariance function $\Gamma_{ik}[\ell] = E\left[\mathbf{x}_i[n+\ell]\mathbf{x}_k^H[n]\right]$, $\ell = 0, \ldots, N-1$, considered in Section 5.3.1 now corresponds to a sequence of Toeplitz matrices. That is, for any pair of channels $\mathbf{x}_i$ and $\mathbf{x}_k$, we now assume that there exists a two-dimensional covariance sequence, $\{\gamma_{ik}[\ell, m]\}$, such that

$$E\left[x_i[n+\ell, p+m]x_k^*[n,p]\right] = \gamma_{ik}[\ell, m] \in \mathbb{C}$$

with ℓ a temporal lag and m a spatial lag. An example of when this model would hold would be a set of L sensor suites, laid out in an arbitrary geometry, but with their respective P-element arrays laid out co-linearly as shown in Figure 5.4.

Given the assumption that the data is not only temporally WSS but spatially as well, results on large block-Toeplitz matrices [42] again show that the matrix $\tilde{C}(e^{j\theta})$ is asymptotically equivalent with the block-diagonal matrix

$$\tilde{C}(e^{j\theta}) \rightarrow \text{blkdiag}\left\{\hat{C}(e^{j\theta}, e^{j\phi_0}), \ldots, \hat{C}(e^{j\theta}, e^{j\phi_{P-1}})\right\}$$

so that as M, N, and P grow large but L remains fixed the GLRT becomes [33], [44]

$$\begin{aligned}
\Lambda^{\frac{1}{NP}} &= \det\left((I_L \otimes T)\hat{C}(I_L \otimes T)^H\right)^{\frac{1}{NP}} = \det\left(\tilde{C}\right)^{\frac{1}{NP}} \\
&\stackrel{N\to\infty}{\to} \exp\left\{\int_{-\pi}^{\pi} \ln\det\left(\hat{C}(e^{j\theta})\right)^{\frac{1}{P}} \frac{d\theta}{2\pi}\right\} \\
&= \exp\left\{\int_{-\pi}^{\pi} \ln\det\left(\tilde{C}(e^{j\theta})\right)^{\frac{1}{P}} \frac{d\theta}{2\pi}\right\} \\
&\stackrel{P\to\infty}{\to} \exp\left\{\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \ln\det\hat{C}(e^{j\theta}, e^{j\phi}) \frac{d\theta d\phi}{4\pi^2}\right\} \\
&= \exp\left\{\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \ln\frac{\det\hat{S}(e^{j\theta}, e^{j\phi})}{\prod_{i=1}^{L}\hat{S}_{ii}(e^{j\theta}, e^{j\phi})} \frac{d\theta d\phi}{4\pi^2}\right\}
\end{aligned} \tag{33}$$

The matrix $\hat{S}(e^{j\theta}, e^{j\phi}) \in \mathbb{C}^{L \times L}$ has elements

$$\left[\hat{S}(e^{j\theta}, e^{j\phi})\right]_{i,k=1}^{L} = \mathbf{f}_P^H(e^{j\phi}) F_N^H(e^{j\theta}) \hat{R}_{ik} F_N(e^{j\theta}) \mathbf{f}_P(e^{j\phi})$$

which is a quadratic estimate of the cross power spectral density between channels i and k in the frequency/wavenumber domain. Thus, similar to the expression in (28), the asymptotic form of the generalized Hadamard ratio in (6) involves the computation of a Hadamard ratio at each frequency/wavenumber pair (θ, ϕ), followed by broadband integration of its logarithm. As we will see in the next section, even finite-dimensional implementations of this statistic can bring a significant improvement in performance by taking advantage of the WSS assumption and its manifestation in a Toeplitz structure for the $NP \times NP$ blocks of R.

5.4. Simulation Results

In this section we provide simulation results to demonstrate several situations where the likelihood ratio in (33) applies and to demonstrate the improvement in detection performance that can be achieved in such applications. The first example is the detection of spatially correlated time series in a network of sensor arrays where the observation $x_i[n, p]$ represents the time series collected at the p^{th} sensor of the i^{th} array. The second simulation example is the detection of correlation among two or more coregistered images. This simulation is more in line with the underwater target detection application that will be discussed in Chapter 6 where $x_i[n, p]$ will represent the pixel location at a particular along-track/range location in the i^{th} sonar image. In this case, index i will correspond to one of the two sonar images, either high frequency or broadband, employed in this dual-channel detection problem.

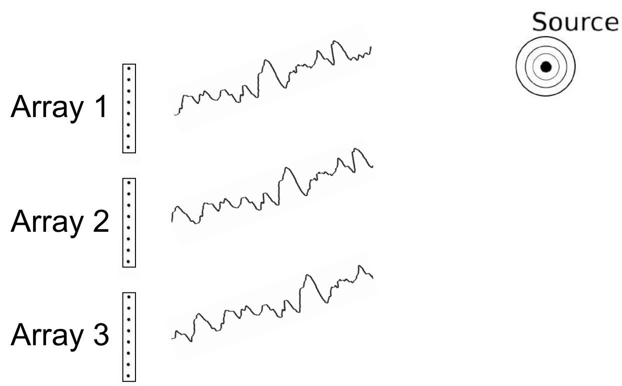


FIGURE 5.4. Detection of a Source using Multiple Linear Arrays.

5.4.1. MULTICHANNEL DETECTION IN A NETWORK OF SENSOR ARRAYS. For this simulation, we consider a network of $L = 3$ sensor arrays, each of which is a uniform linear array (ULA) of $P = 16$ sensor elements. Our aim is to generate a quite arbitrary field and to propagate this field to all three sensor arrays as depicted in Figure 5.4. The propagating signal $s[n]$ for $n = 0, \ldots, MN - 1$ produced by the source is assumed to be a zero-mean WSS random process. As such, it follows that there exists an orthogonal increment process $\{\psi(\theta), -\pi < \theta \leq \pi\}$ such that [45]

$$s[n] = \int_{-\pi}^{\pi} e^{jn\theta} d\psi(\theta)$$

The random measure $d\psi(\theta)$, which may be treated as a narrowband component of the signal $s[n]$ at the instantaneous frequency θ, is a normal random variable with covariance

$$E\left[d\psi(\theta)d\psi^*(\omega)\right] = \delta(\theta - \omega)\sigma_s^2(e^{j\theta})d\theta$$

and $\sigma_s^2(e^{j\theta})$ is the power spectral density

$$\sigma_s^2(e^{j\theta}) = \frac{1}{2\pi} \sum_{l=-\infty}^{\infty} e^{-jl\theta} E\left[s[n]s^*[n+l]\right]$$

This signal is then propagated to each sensor array so that the signal measured at the i^{th} sensor array, $s_i[n]$, may be written

$$s_i[n] = \int_{-\pi}^{\pi} e^{jn\theta} e^{-j\theta T_i} d\psi(\theta)$$

with T_i a bulk propagation delay representing the time taken for the signal to reach the i^{th} array. For this simulation, it is assumed that $s[n]$ arises from a first-order autoregressive process with coefficient a and driven by a white noise sequence with variance σ^2 so that the spectral density $\sigma_s^2(e^{j\theta})$ may be written [45]

$$\sigma_s^2(e^{j\theta}) = \frac{\sigma^2}{2\pi} \frac{1}{|1 - ae^{-j\theta}|^2}$$

The signal received at each array is then propagated as a planewave among its elements. At each P-element array a temporally colored nonpropagating noise component is added independently of all sensors so that the observation at the i^{th} array, $\mathbf{x}_i[n] \in \mathbb{C}^P$, may be written

$$\mathbf{x}_i[n] = \int_{-\pi}^{\pi} e^{jn\theta} e^{-j\theta T_i} \mathbf{a}(e^{j\theta}) d\psi(\theta) \, + \, \mathbf{w}_i[n]$$

where the noise vector $\mathbf{w}_i[n]$ has the cross spectral matrix

$$\frac{1}{2\pi} \sum_{l=-\infty}^{\infty} e^{-jl\theta} E\left[\mathbf{w}_i[n]\mathbf{w}_i^H[n+l]\right] = \sigma_w^2(e^{j\theta}) I_P$$

Also, the vector $\mathbf{a}(e^{j\theta})$ denotes the array response or steering vector

$$\mathbf{a}(e^{j\theta}) = \left[1 \ e^{-j\theta\tau} \ \cdots \ e^{-j(P-1)\theta\tau}\right]^T$$

with τ a propagation delay dependent on the properties of the medium, the distance between sensor elements, and the Direction-of-Arrival (DOA) of the far-field source. Consequently, each $P \times P$ block of the frequency-dependent spectral density matrix of the composite observation can then be written as follows

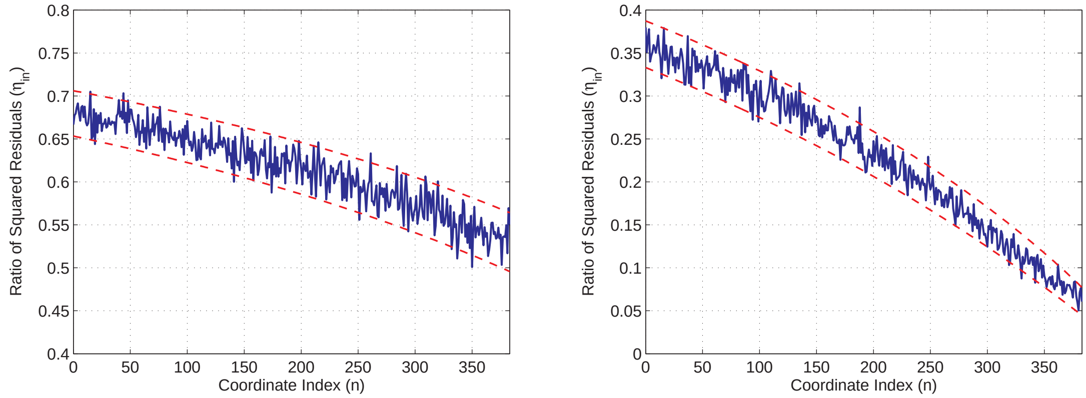
$$\left\{S(e^{j\theta})\right\}_{i,k} = \begin{cases} \sigma_s^2(e^{j\theta})\mathbf{a}(e^{j\theta})\mathbf{a}^H(e^{j\theta}) + \sigma_w^2(e^{j\theta})I_P & i = k \\ \\ \sigma_s^2(e^{j\theta})e^{j\theta(T_k - T_i)}\mathbf{a}(e^{j\theta})\mathbf{a}^H(e^{j\theta}) & i \neq k \end{cases}$$

For this simulation, the sensor noise is generated by passing unit-variance white noise through a 5^{th}-order FIR filter with weights $b_0, \ldots, b_5$ so that the noise spectral density may be written

$$\sigma_w^2(e^{j\theta}) = \frac{1}{2\pi}\left|\sum_{k=0}^{5} b_k e^{-jk\theta}\right|^2$$

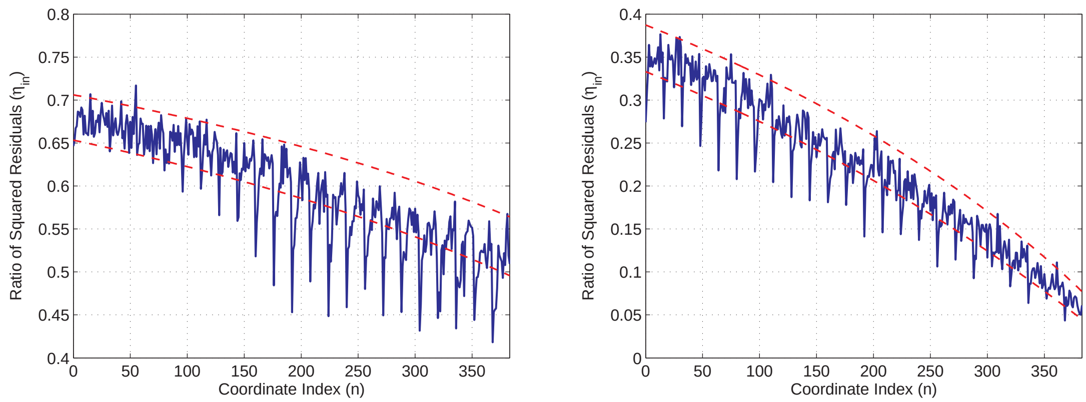
Upon collecting all MN measurements at each sensor element, the data record is temporally partitioned into M non-overlapping copies of a time series of length $N = 24$. The likelihood ratio given in (33), referred to as "Frequency/Wavenumber Domain GLRT", is then used to discriminate situations where a source is present from those in which each sensor array observes its own correlated noise field only. The performance of this detector will be compared to the classical likelihood ratio given in (6), referred to as "Time Domain GLRT", as well as its frequency domain version given in (29) which is referred to as "Frequency Domain GLRT".

As mentioned briefly at the beginning of Section 5.3, all of the results of Section 3 generalize to the situation considered here by simply replacing N with NP in which case



(a) Channel Index $i = 2$ (b) Channel Index $i = 3$

FIGURE 5.5. Ratio of squared residuals under $\mathcal{H}_0$.



(a) Channel Index $i = 2$ (b) Channel Index $i = 3$

FIGURE 5.6. Ratio of squared residuals under $\mathcal{H}_1$.

the stochastic representation given in (11) becomes

$$\Lambda|\mathcal{H}_0 \stackrel{d}{=} \prod_{i=2}^{L} \prod_{n=0}^{NP-1} Y_{in} \tag{34}$$

$$Y_{in} \sim \text{Beta}\left(M-(i-1)NP-n,(i-1)NP\right)$$

With $M = 1200$ realizations of the composite vector $\mathbf{z}[m]$ under the null hypothesis, Figure 5.5 plots the ratio of squared residuals, $\sigma_{in}^2(\hat{R})$ to $\sigma_{in}^2(\hat{R}_{ii})$, given in (10)

$$\eta_{in} = \frac{\sigma_{in}^2(\hat{R})}{\sigma_{in}^2(\hat{R}_{ii})} = \frac{\mathbf{x}_{in}^H P_{ZX}^{\perp} \mathbf{x}_{in}}{\mathbf{x}_{in}^H P_{ZX}^{\perp} \mathbf{x}_{in} + \mathbf{x}_{in}^H P_{P_X^{\perp} Z} \mathbf{x}_{in}}$$

for $n = 0, \ldots, NP - 1$ when (a) the channel index of the product given in (34) is $i = 2$ and (b) when it is $i = 3$. The dashed lines in both of these plots show a 95% confidence interval for each beta random variable Y_{in}, i.e. the interval $[a, b]$ such that $P\left[Y_{in} \leq a\right] = P\left[Y_{in} \geq b\right] = 0.025$. Recall from Section 3.2 that what differentiates the values observed in these two plots is that in Figure 5.5 (a) $\mathbf{x}_1$ is used to predict $\mathbf{x}_2$ while in Figure 5.5 (b) both $\mathbf{x}_1$ and $\mathbf{x}_2$ are being used to predict $\mathbf{x}_3$. Likewise, with the presence of a source with signal-to-noise ratio $SNR = 10\log_{10}\sigma^2 = -3$ dB, Figure 5.6 displays the same for $M = 1200$ realizations from the alternative hypothesis. Comparing these two figures, it is clear that this interval traps η_{in} with high probability under the null hypothesis, but does not under the alternative hypothesis where many values fall below the interval, signaling a deviation from independence.

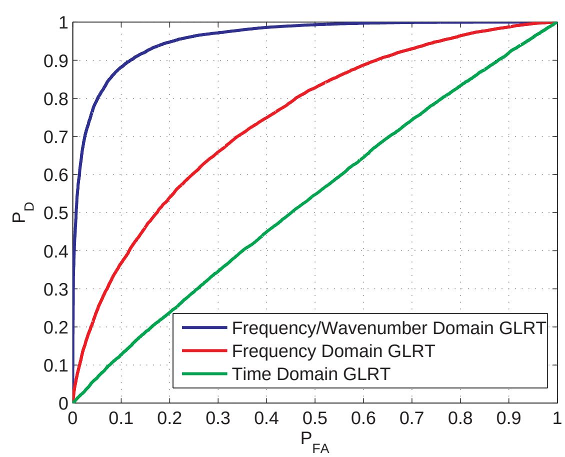
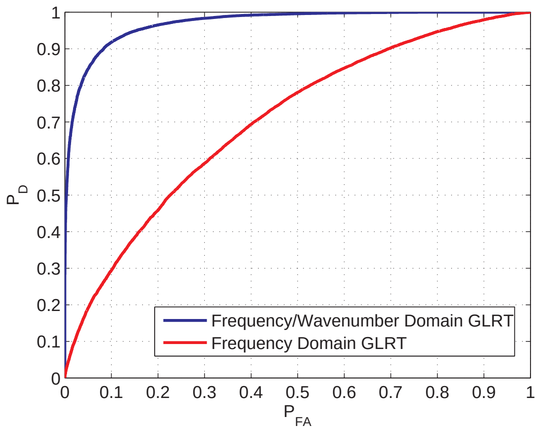


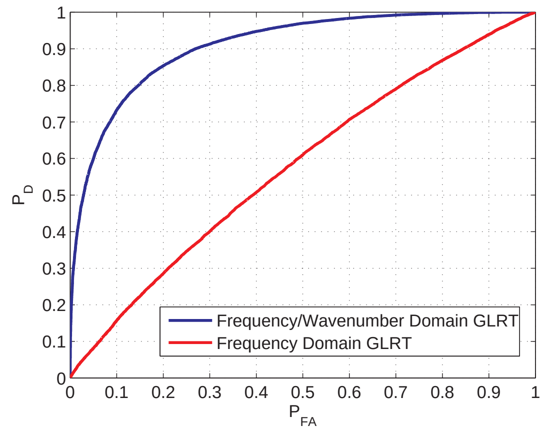
FIGURE 5.7. Detection performance with $M = 1200$ and $SNR = -30$ dB.

With $M = 1200$ and a $SNR = -30$ dB source, Figure 5.7 displays the Receiver Operating Characteristic (ROC) curves for all three detection methods considered here. From Figure 5.7 we can see that the Frequency/Wavenumber Domain GLRT exhibits a performance that exceeds that of the Frequency Domain GLRT when discriminating these two hypotheses while the performance of the Time Domain GLRT is particularly poor. This is most likely due to the fact that the Time Domain GLRT does not exploit the WSS assumption and its manifestation in a Toeplitz structure for the $N \times N$ or the $NP \times NP$ blocks of R. A true GLRT for this case would use a ML estimate for Toeplitz matrices, an intractable problem with no analytical solution. So the time-domain GLRT, while generally applicable, is actually mis-matched to the WSS problem. On the other hand, the frequency-domain and frequency-wavenumber domain GLRTs, while not as generally applicable, are better matched to the WSS case. Moreover, these forms estimate cross-spectral matrices, which are approximately block-diagonal in the WSS case, and use only their block-diagonals. In other words, they exploit the assumed wide-sense stationarity by using only diagonal blocks of the cross spectral matrix. Asymptotically, this approaches a GLRT that is faithful to the assumptions of wide-sense stationarity.

Finally, Figures 5.8 (a) and (b) compare the performances of the Frequency/Wavenumber and Frequency Domain GLRTs for a successively smaller number of copies but with a higher power source. Note that the Time Domain GLRT has been excluded from these two studies because of insufficient sample support, i.e. M is too small to construct positive-definite covariance estimates. Similar to the results of Figure 5.7, we can again see that the Frequency/Wavenumber Domain GLRT outperforms in each case. Again, this is likely due to the fact that the likelihood ratio given in (33) is better matched to the (spatially) WSS case versus its alternative given in (29) which, while more generally applicable, does not



(a) $M = 250$, $SNR = -26$ dB



(b) $M = 50$, $SNR = -20$ dB

FIGURE 5.8. Detection performance with fewer samples but a higher power source.

fully exploit wide-sense stationarity, i.e. the cross-spectral matrices $\hat{S}(e^{j\theta}) \in \mathbb{C}^{PL \times PL}$ in (33) do not exploit the spatial WSS problem. Thus, by taking advantage of the spatiotemporal properties of the problem at hand, we can see that the GLRT given in (33) presents an appealing likelihood ratio that exhibits improved detection performance when compared to the two alternatives considered in (6) and (29).

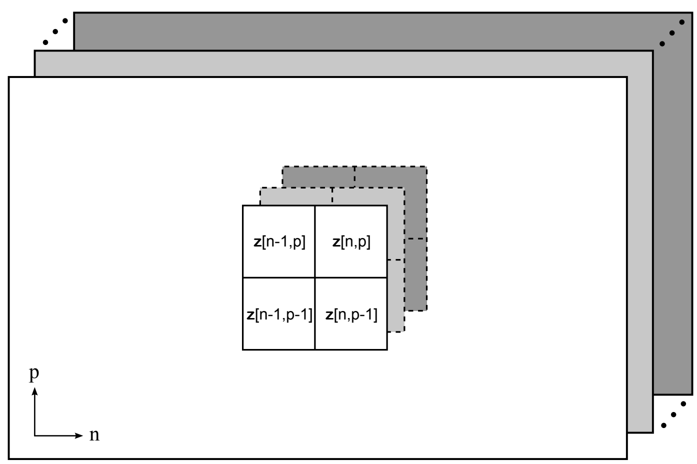


FIGURE 5.9. The support region for the 2D multivariate AR process.

5.4.2. MULTICHANNEL DETECTION IN COREGISTERED IMAGES. Another application where the use of (33) might be useful is the detection of coherence among two or more coregistered images. This idea will be employed in Chapter 6 for the purposes of detecting underwater targets in pairs of high frequency and broadband sonar images. In this simulation, we assume a pair ($L = 2$) of real-valued images $\{x_i[n,p]\}$ whose composite vector $\mathbf{z}[n,p] = [x_1[n,p]\ x_2[n,p]]^T \in \mathbb{R}^2$ follows the 2D multivariate autoregressive (AR) process

$$\mathbf{z}[n,p] = A_{10}\mathbf{z}[n-1,p] + A_{01}\mathbf{z}[n,p-1] + A_{11}\mathbf{z}[n-1,p-1] + \mathbf{w}[n,p] \tag{35}$$

with coefficient matrices

$$A_{jk} = \begin{bmatrix} a_{jk}^{11} & a_{jk}^{12} \\ a_{jk}^{21} & a_{jk}^{22} \end{bmatrix}, \ j,k = 0,1$$

and composite white noise sequence $\mathbf{w}[n,p] \overset{iid}{\sim} \mathcal{N}(0, \sigma^2 I_2)$. The support region and generation of the composite vector $\mathbf{z}[n,p]$ using this multivariate AR model for an arbitrary number of images is depicted in Figure 5.9. For this simulation, we wish to test the null hypothesis that $A_{jk} = \text{diag}\left(a_{jk}^{11}, a_{jk}^{22}\right)$ for all $j,k = 0,1$, i.e. we consider the hypothesis test

$$\begin{aligned} \mathcal{H}_0 &: a_{jk}^{i\ell} = 0 \ \forall\ j,k = 0,1 \text{ and } i \neq \ell \\ \mathcal{H}_1 &: a_{jk}^{i\ell} \neq 0 \ \forall\ j,k = 0,1 \text{ and } i \neq \ell \end{aligned}$$

Using the autoregressive model in (35) with $\sigma^2 = 0.25$ and coefficient matrices

$$A_{10} = \begin{bmatrix} 0.4 & 0.005 \\ -0.05 & 0.25 \end{bmatrix}, \quad A_{10} = \begin{bmatrix} 0.25 & 0.005 \\ 0.05 & 0.3 \end{bmatrix}, \quad A_{11} = \begin{bmatrix} 0.3 & 0.005 \\ 0.01 & 0.25 \end{bmatrix}$$

windows of size 128×128 are generated and each pair is partitioned into non-overlapping blocks of size $N \times P$. Each corresponding block from the images of both channels is then vectorized and concatenated to form the composite vector $\mathbf{z}[m] = \left[\mathbf{x}_1[m]^T\ \mathbf{x}_2[m]^T\right]^T \in \mathbb{R}^{2NP}$.

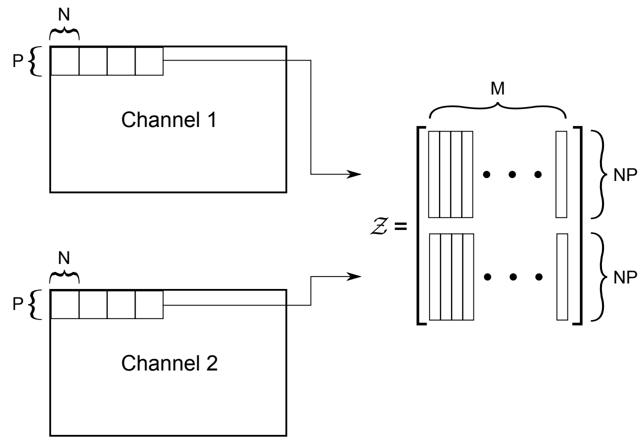
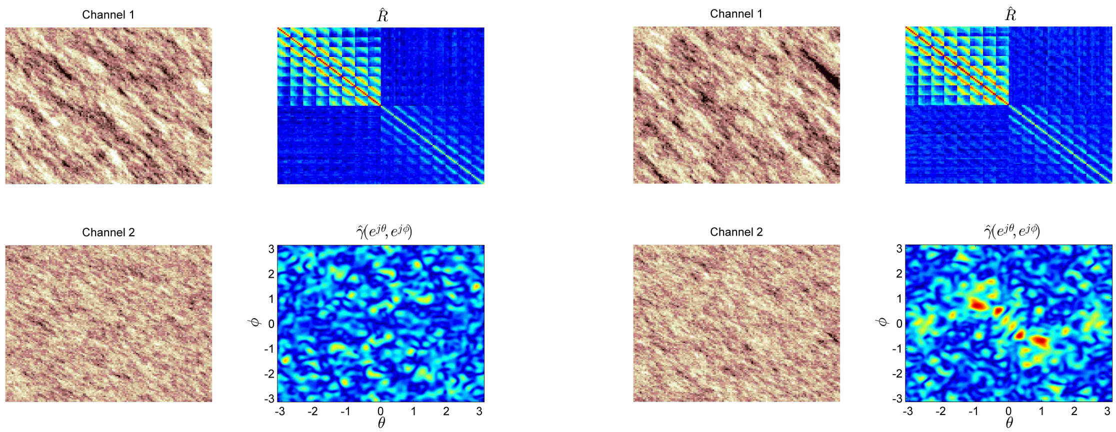


FIGURE 5.10. Formation of the data matrix $\mathcal{Z}$ used to compute the generalized Hadamard ratio Λ.

All block are subsequently accumulated into the data matrix $\mathcal{Z}$ given in (5) as is shown in Figure 5.10. With a $N = P = 8$ block size and $M = 128^2/8^2 = 256$ samples, the data matrix $\mathcal{Z}$ is used to compute the generalized Hadamard ratio given in (6). This detection strategy is compared to the broadband coherence detector given in (33) using a larger block size of $N = P = 32$. For this detection strategy, the cross-spectral matrix $\hat{S}(e^{j\theta}, e^{j\phi})$ is estimated using Welch's method [46] by partitioning each pair of images into blocks as in Figure 5.10 but with a 50% overlap in both dimensions. Each block is subsequently windowed using a separable Hamming window, i.e. the 2D window function can be written $w[n,p] = w[n]w[p]$ with $w[n]$ and $w[p]$ representing two 1D Hamming windows [47]. A 2D FFT is subsequently applied and the corresponding periodograms for each block are averaged to produce the estimate $\hat{S}(e^{j\theta}, e^{j\phi})$.

Figure 5.11 gives examples of the images from both channels, the unconstrained covariance estimate $\hat{R}$ used in the computation of the generalized Hadamard ratio, and the estimated magnitude coherence $\hat{\gamma}(e^{j\theta}, e^{j\phi})$ defined as

$$\hat{\gamma}(e^{j\theta}, e^{j\phi}) = \frac{|\hat{S}_{12}(e^{j\theta}, e^{j\phi})|}{\sqrt{\hat{S}_{11}(e^{j\theta}, e^{j\phi})\hat{S}_{22}(e^{j\theta}, e^{j\phi})}} \tag{36}$$



(a) Uncorrelated ($\mathcal{H}_0$) (b) Correlated ($\mathcal{H}_1$)

FIGURE 5.11. Examples of the images for each channel, the estimated covariance matrix $\hat{R}$, and the estimated magnitude coherence $\hat{\gamma}(e^{j\theta}, e^{j\phi})$.

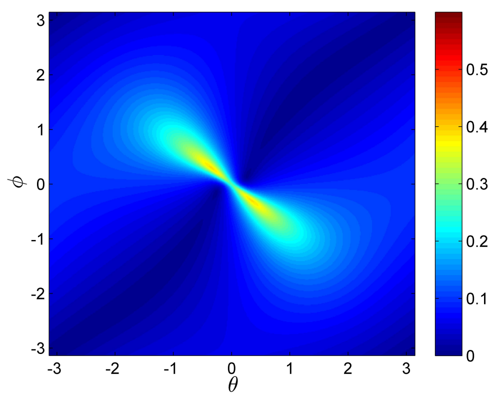


FIGURE 5.12. The actual magnitude coherence $\gamma(e^{j\theta}, e^{j\phi})$ under $\mathcal{H}_1$.

when (a) both channels are uncorrelated ($\mathcal{H}_0$) and (b) when they are correlated ($\mathcal{H}_1$). Note that when $L = 2$, the narrowband Hadamard ratio in the integrand of (33) can be expressed in terms of $\hat{\gamma}(e^{j\theta}, e^{j\phi})$ via

$$\frac{\det \hat{S}(e^{j\theta}, e^{j\phi})}{\prod_{i=1}^{L} \hat{S}_{ii}(e^{j\theta}, e^{j\phi})} = 1 - \hat{\gamma}^2(e^{j\theta}, e^{j\phi})$$

Comparing the results shown in Figure 5.11, one can see a very minor difference between the covariance matrices $\hat{R}$ in each case. The difference between the magnitude coherence

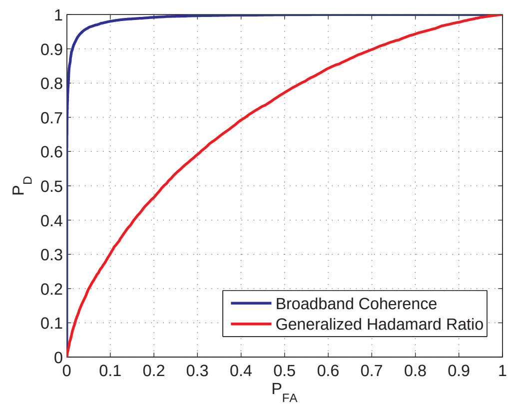


FIGURE 5.13. Comparison of the ROC curves for the generalized Hadamard ratio and broadband coherence detectors.

$\hat{\gamma}(e^{j\theta}, e^{j\phi})$, however, is very noticeable as one sees much larger values when the data is correlated in Figure 5.11 (b) versus when it is uncorrelated in Figure 5.11 (a). Under the alternative hypothesis $\mathcal{H}_1$, Figure 5.12 plots the true coherence

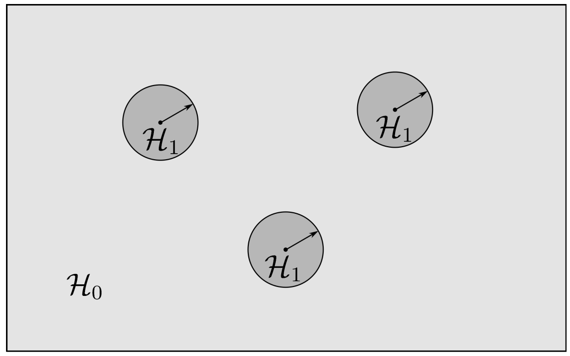
$$\gamma(e^{j\theta}, e^{j\phi}) = \frac{|S_{12}(e^{j\theta}, e^{j\phi})|}{\sqrt{S_{11}(e^{j\theta}, e^{j\phi})S_{22}(e^{j\theta}, e^{j\phi})}}$$

with cross-spectral matrix $\left[S(e^{j\theta}, e^{j\phi})\right]_{i,k=1}^{2} = S_{ik}(e^{j\theta}, e^{j\phi})$

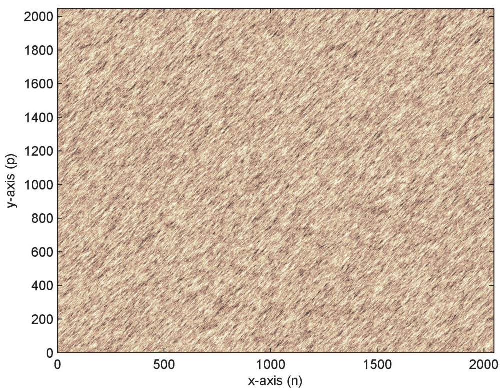
$$\begin{aligned} S(e^{j\theta}, e^{j\phi}) &= \sigma^2 A^{-1}(e^{j\theta}, e^{j\phi}) A^{-H}(e^{j\theta}, e^{j\phi}) \\ A(e^{j\theta}, e^{j\phi}) &= I_2 - A_{10}e^{-j\theta} - A_{01}e^{-j\phi} - A_{11}e^{-j(\theta+\phi)} \end{aligned}$$

in which case one can see a fairly good correspondence between this image and the example of its estimate given in the bottom right of Figure 5.11 (b). Note that, under the null hypothesis $\mathcal{H}_0$, the true magnitude coherence is zero, i.e. $\gamma(e^{j\theta}, e^{j\phi}) = 0$ for all $-\pi < \theta, \phi \leq \pi$.

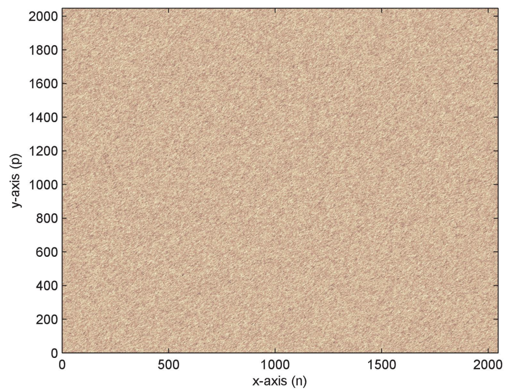
Figure 5.13 compares the ROC curves of the generalized Hadamard ratio detector in (6) and broadband coherence detector in (33) in which case one can see that the broadband coherence detector is clearly better in performance. As a final demonstration, a pair of



(a) Image Generation



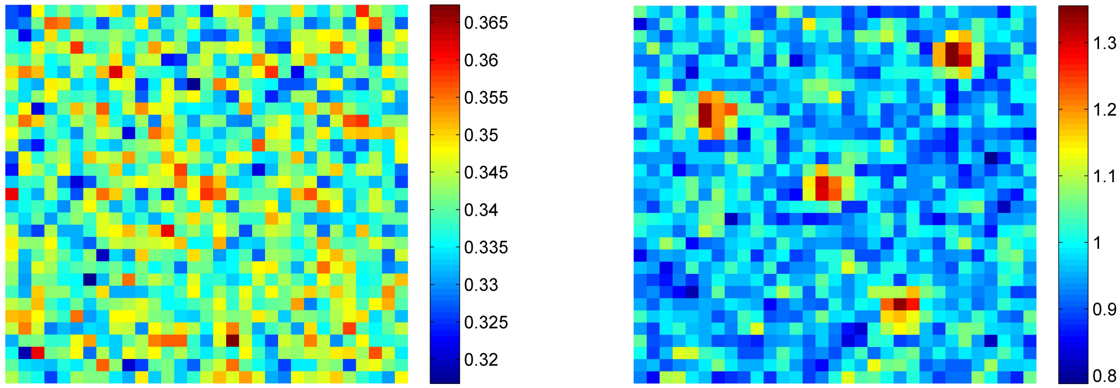
(b) Channel 1



(c) Channel 2

FIGURE 5.14. Generation of two images which are uncorrelated except at several arbitrarily chosen locations as illustrated in (a).

images of size 2048×2048 were generated using the model in (35) which, by-and-large, are uncorrelated by setting $a_{jk}^{i\ell} = 0$ for all $i \neq \ell$. However, for arbitrarily selected locations within these images the data was generated to be correlated by switching $a_{jk}^{i\ell}$ to its non-zero value for all pixels within a 128×128 region of each location. The idea behind this setup is depicted in Figure 5.14 (a) with the images for both channels shown in Figures 5.14 (b) and (c). Trying to visually discern where these locations have been hidden in this pair of images is difficult. A window of size 128×128 is then scanned through each image and the data applied to both detectors. Figure 5.15 (a) displays the negative log-likelihood ratio for each window when employing the generalized Hadamard ratio in which case it is clear that it too fails to determine where linear dependence has been enforced. However, the negative



(a) Generalized Hadamard Ratio

(b) Broadband Coherence

FIGURE 5.15. Likelihood ratio values for the pair of images shown in Figure 5.14 (b) and (c).

log-likelihood ratio values for the broadband coherence detector shown in Figure 5.15 (b) make it obviously clear where these locations have been placed as indicated by the relatively large values observed in each region. Thus, by exploiting the inherent Toeplitz-block-Toeplitz behavior in each block of matrix R, one can again see that the broadband coherence detector presents an appealing likelihood ratio that can improve the performance over the generalized Hadamard ratio.

5.5. Conclusion

In this chapter we considered an extension of the generalized Hadamard ratio given in (6) which exploits the inherent Toeplitz-block-Toeplitz structure that each block R_{ik} of matrix R will take when the set of observations $\{x_i[n,p]\}_{i=1}^{L}$ correspond to 2D WSS processes. Examples of where this model may apply is the temporally and spatially WSS observations collected in a network of sensor arrays or the spatially WSS data observed among multiple coregistered images. Although the true GLRT for this problem would impose this Toeplitz structure on the ML estimate of R, this is an intractable problem with no closed-form solution. However, the asymptotic behavior of Toeplitz matrices is well understood and leads to very tractable results. Taking advantage of this asymptotic theory, the GLRT is expressed as the broadband integral of a narrowband Hadamard ratio over the 2D frequency plane, a test statistic referred to as *broadband coherence*.

The broadband coherence detector is then demonstrated using simulations of two different multichannel detection applications. These applications involve detection in a network of sensor arrays and among multiple coregistered images. In the first simulation, we considered the problem of detecting a single wideband source using several ULAs laid out in a linear geometry. For this problem, the broadband coherence detector was shown to provide substantial improvement over the generalized Hadamard ratio and a likelihood ratio that is analogous to that developed in [7]. In this case, the broadband coherence detector involves the analysis of coherence in a frequency/wavenumber domain. For the second simulation, we considered the problem of detecting correlation among a pair of images generated using a 2D multivariate autoregressive model. In this situation it was shown that the broadband coherence detector again provides substantial improvement in detection performance over the generalized Hadamard ratio. For this application, the broadband coherence detector involves the analysis of coherence in the 2D frequency domain. In the next chapter, the broadband coherence detector will be applied to pairs of coregistered high frequency and broadband sonar images for the detection of underwater targets on the seafloor.

CHAPTER 6

Application to Sonar Imagery

6.1. Introduction

In this chapter we will apply the results and methods developed in previous chapters to the problem of automatic target detection in pairs of coregistered sonar images of the seafloor. The detection of underwater objects in sonar imagery is a difficult problem due to many factors such as variations in operating and environmental conditions, presence of spatially varying clutter, variations in target shapes, compositions and orientation, as well as different degrees with which a target may be buried in sediment. Moreover, bottom features such as coral reefs, sand formations, and vegetation may obscure a target and can be a source of high false alarm rates. Due to the wide variation in target conditions that can be observed for this problem, detection methods that take advantage of general discriminative features in the data, as opposed to those that rely on specific target models, can in some cases be desirable. Such is the case for the solution presented here where the detection principle simply relies on the assumption that the presence of targets in coregistered high frequency (HF) and broadband (BB) images will lead to a higher degree of coherence than when those images contain background alone. In other words, we expect the matrix $\hat{R}$ to be better approximated by a block-diagonal matrix when the image snippets taken from this pair of images only contains returns from the seafloor compared to a situation where they also contain a target. It is this deviation from a block-diagonal matrix as measured by the generalized Hadamard ratio and broadband coherence detectors which we will use to determine if and where targets are located within each pair of images.

To demonstrate the methods developed in this dissertation, tests are conducted on three datasets consisting of pairs of HF and BB sonar images collected using the Small Synthetic Aperture Minehunter (SSAM) system [48]. Two of the datasets used in the studies of this chapter contain images collected at different geographical locations with each environment presenting different difficulty levels in clutter density. These datasets will be used to investigate the use of the null distribution developed in Chapters 3 and Chapters 4 and to demonstrate the use of the broadband coherence detector developed in Chapter 5. The other dataset used in this chapter consists of actual HF and BB sonar images of the seafloor with simulated targets of different geometrical shapes inserted into the image. This dataset will be primarily used to study the proposed method's robustness to deviation from coregistration, an issue which can arise in dual frequency band sonar imagery.

The remainder of this chapter is organized as follows. In Section 6.2, we give a brief review of the three datasets used in the results of this chapter as well as the processing methods used for these coherence-based detection algorithms. Section 6.3 investigates the use of the null distribution of the generalized Hadamard ratio to sonar imagery and the accuracy with which the distribution can capture the likelihood ratio of background for various clutter difficulties. In Section 6.4 we discuss the proposed method's robustness to coregistration. We then present results of the broadband coherence detector when applied to the two datasets containing actual targets and compare its performance to several alternative detectors. Concluding remarks are finally given in Section 6.5.

6.2. Data Description and Processing

To test and compare the results presented in this dissertation, the methods were applied to several dual-sonar data sets collected using the Small Synthetic Aperture Minehunter

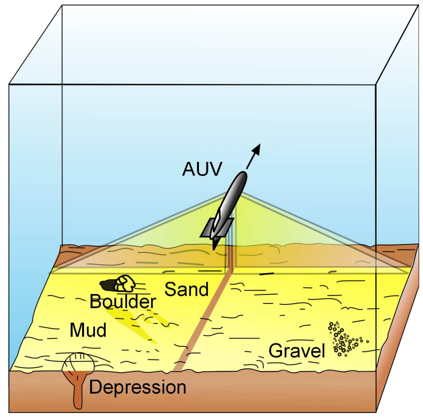


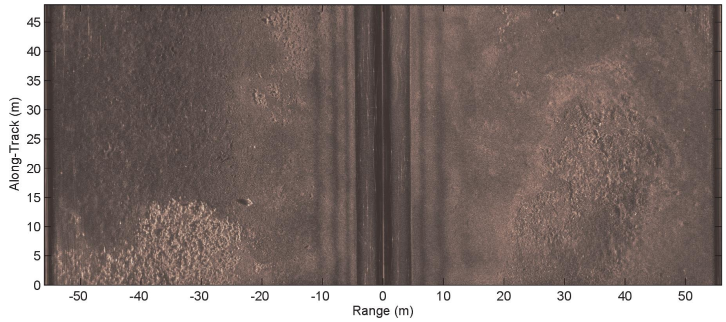
FIGURE 6.1. Collection of sonar data using synthetic aperture processing.

(SSAM) [48]. SSAM is a dual frequency band Synthetic Aperture Sonar (SAS) system capable of producing high resolution images of the seafloor. Although the real aperture mounted on the autonomous underwater vehicle (AUV) only consists of 8 receiver elements, high resolution images are produced by synthetic aperture processing and coherently combining a number of acoustic pings as the AUV moves in along-track. A depiction of this data collection scenario is given in Figure 6.1. With the SSAM system's use of dual frequency bands, each is used to construct one HF high-resolution sonar image as well as one BB sonar image coregistered over the same region on the sea-floor. One of the advantages of jointly employing both HF and BB sonar is the ability of HF sonar to provide higher spatial resolution and a better ability to capture target details and characteristics while BB sonar offers much better clutter suppression ability with lower spatial resolution. So although the multichannel methods outlined in this dissertation apply to any number of channels L, the results of this chapter only pertain to the use of $L = 2$ channels. Coregistration in these images is easily achieved as the two sonar systems are mounted on the same autonomous underwater vehicle (AUV) and use the same receive hydrophone array. The pinging for both HF and BB systems is done simultaneously as they are sufficiently far apart in frequency such that their returns are easily separable.

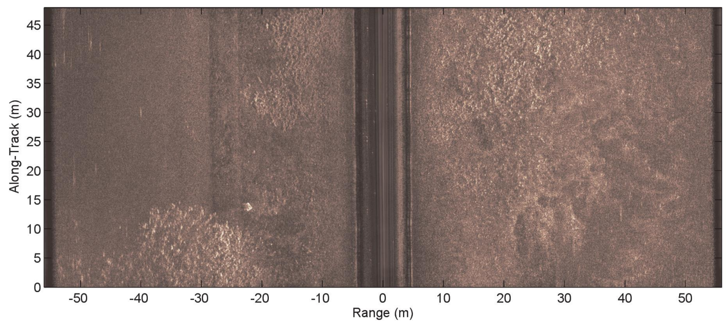
The sonar images used in this work are generated at the output of a coherent processor, in this case the k-space or wavenumber beamformer [49], [50]. Each impinging sound wave on the receiver array elements of the sonar is converted to magnitude and phase. The delay and sum beamforming algorithm [51] attempts to coherently combine the sound waves in a way that resolves the echo returns into a complex-valued pixel. More specifically, the k-space or wavenumber algorithm computes the 2-D Fourier transform of the raw or range-compressed sonar data in the delay-time/aperture domain. This converts the data into the spatial frequency/wavenumber (ω, k)-domain where it is multiplied by the power spectrum of the transmitted wavefront. A change of variables is done by Stolt interpolation [52]. This change of variables maps the frequency/wavenumber (ω, k)-domain into the wavenumber domain (k_x, k_y). The inverse 2-D Fourier transform is then taken of the mapped data to form the complex image.

Figure 6.2 gives an example of a corresponding pair of HF and BB images produced using the SSAM system and beamforming procedures described above. Note that the images shown in this figure only display the magnitude of each complex-valued pixel. When observing a target lying proud on the seafloor, one typically notices two defining characteristics: a highlight region corresponding to a strong sonar return from the target itself followed by a shadow region as the target blocks the sonar return immediately behind it. An example of a target that exhibits these characteristics can be seen in the images of Figure 6.2 at approximately 14 m in along-track and -22 m in range. This highlight/shadow relationship for targets will be exploited in Section 6.4.3 to construct a matched subspace detector which will be used to benchmark the performance of the proposed coherence-based techniques.

The results of this chapter were generated using three sonar imagery datasets provided by NSWC-Panama City. The first two datasets consist of sonar images containing actual



(a) HF Image



(b) BB Image

FIGURE 6.2. An example of the HF and BB images produced by the SSAM sonar.

TABLE 6.1. Characteristics of Both Real Sonar Datasets

		Clutter Difficulty	Number of Images	Number of Targets
Dataset 1		Medium/Hard	122	77
Dataset 2	Day 1	Easy	180	4
	Day 2	Hard	136	17
	Day 3	Easy	142	47
	Total	–	458	68
Dataset 3		Easy	145	580

targets lying on the seafloor. The first of these two datasets, from now on referred to as "Dataset 1", was collected from one geographical location with an environment consisting of medium to hard clutter difficulty. The second dataset, referred to as "Dataset 2", was collected at a completely different geographical location from Dataset 1. Dataset 2 is further partitioned into three days of data collection with each day consisting of different target fields. Day 1 and day 3 of this dataset have minimal clutter density while day 2 contains very difficult clutter. Table 6.1 lists the total number of images as well as the number of targets for each dataset. These two datasets will not only be used to test the validity of the null distribution of the generalized Hadamard ratio in Section 6.3 but they will also be used to evaluate the performance of the broadband coherence detector in Section 6.4.

The third dataset used in this chapter, referred to as “Dataset 3”, contains actual SAS images of the seafloor with synthetically generated targets inserted into the images [53]. Using physics-based target scattering models, the images in this dataset are generated by inserting the simulated sonar returns from the target into actual returns from the seafloor prior to the beamforming process. This simulated sonar imagery dataset consists of 145 pairs of HF and BB sonar images with each pair corresponding to a particular data collection scenario characterized by its background and object configuration. Each image pair uses one of 29 different backgrounds that are real images of the seafloor and contains the synthetically generated signatures of four different geometrical shapes that are inserted into the image according to one of 5 different configurations that define the orientations and relative positions of the objects with respect to the sonar. The total number of images and targets in this dataset are given at the bottom of Table 6.1. Each pair of HF and BB images contains one of each of the following four target types: block, cone, sphere, and cylinder. This dataset will be used in Section 6.4 to investigate the sensitivity of the proposed methods to deviations from coregistration as mentioned before.

When processing the images in the data sets for the detection methods considered in this chapter, each pair of HF and BB images is first partitioned into coregistered regions of interest (ROIs) with 50% overlap in both the range and along-track dimensions. ROIs are formed in an overlapping fashion to ensure that the target will not be split among different ROIs. Thus, if an ROI contains a target, it will encompass the entirety of the

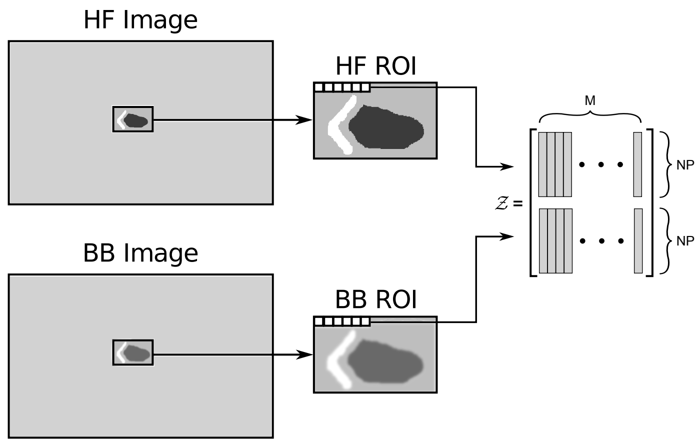


FIGURE 6.3. Partitioning of the HF and BB images into coregistered ROIs and formation of the data matrix $\mathcal{Z}$.

target structure. Based on the average target size, ROIs for both the HF and BB images are chosen to be 80 pixels tall by 144 pixels wide. Once each pair of coregistered ROIs has been extracted from the sonar images, the processing steps are exactly the same as that used in the simulations of Section 5.4.2: each pair of ROIs is in turn partitioned into non-overlapping blocks of size $N \times P$, the blocks are vectorized and concatenated to form the composite vector $\mathbf{z}[m] = \left[\mathbf{x}_1[m]^T \; \mathbf{x}_2[m]^T\right]^T \in \mathbb{R}^{2NP}$, and all blocks are subsequently accumulated into the data matrix $\mathcal{Z}$ given in (5). The size of each block (N and P) as well as the number of samples (M) are different depending on the detection method being considered. This processing procedure is depicted in Figure 6.3. The processing for the broadband coherence detector is also the same as that used in Section 5.4.2: ROIs are again partitioned into blocks but with a 50% overlap in both the along-track and range dimensions, each block is windowed using a 2D Hamming window, a 2D FFT is applied, and the periodograms for each block are averaged to estimate the cross-spectral matrix $\hat{S}(e^{j\theta}, e^{j\phi})$. As the realizations (blocks) of the composite data vector $\mathbf{z}[m]$ used to form second-order estimates of the covariance matrix $\hat{R}$ and cross-spectral matrix $\hat{S}$ are spatially distributed over the area of the ROI, it is difficult to argue that these realizations are independent and identically distributed. As we

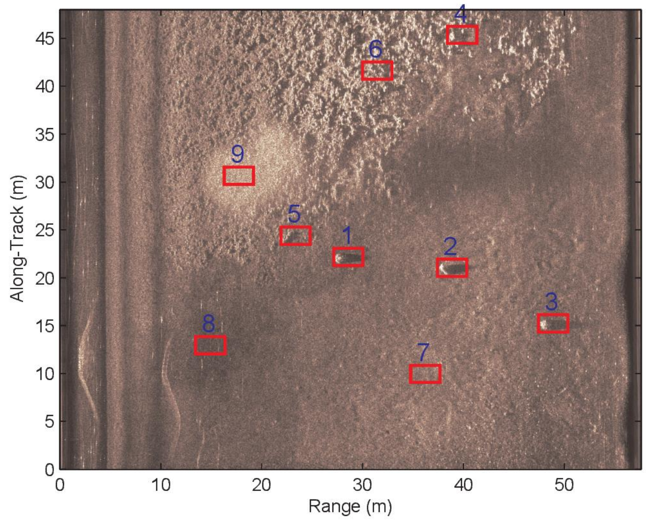


FIGURE 6.4. An example of one HF image with several locations corresponding to both target and background chosen throughout the image.

shall see in the subsequent sections of this chapter, however, this data processing strategy is still capable of producing estimates that can sufficiently discriminate between target and background.

6.3. False Alarm Studies

One of the main subjects of this dissertation was the null distribution of the generalized Hadamard ratio developed in Chapter 3 and the use of these results in defining thresholds that match a desired false alarm rate. In this section, we will demonstrate the use of these methods on the two datasets containing actual targets lying on the seafloor when employing the generalized Hadamard ratio. To keep the number of samples relatively large compared to the dimension of each channel, a fairly small block size of $N = P = 8$ was used resulting in a total of $M = 180$ samples per every pair of ROIs. Using both Dataset 1 and Dataset 2, we will first demonstrate the stochastic behavior of the ratio of squared residuals given in (10) and then test the use of the saddlepoint approximation in the different environments presented by these two datasets.

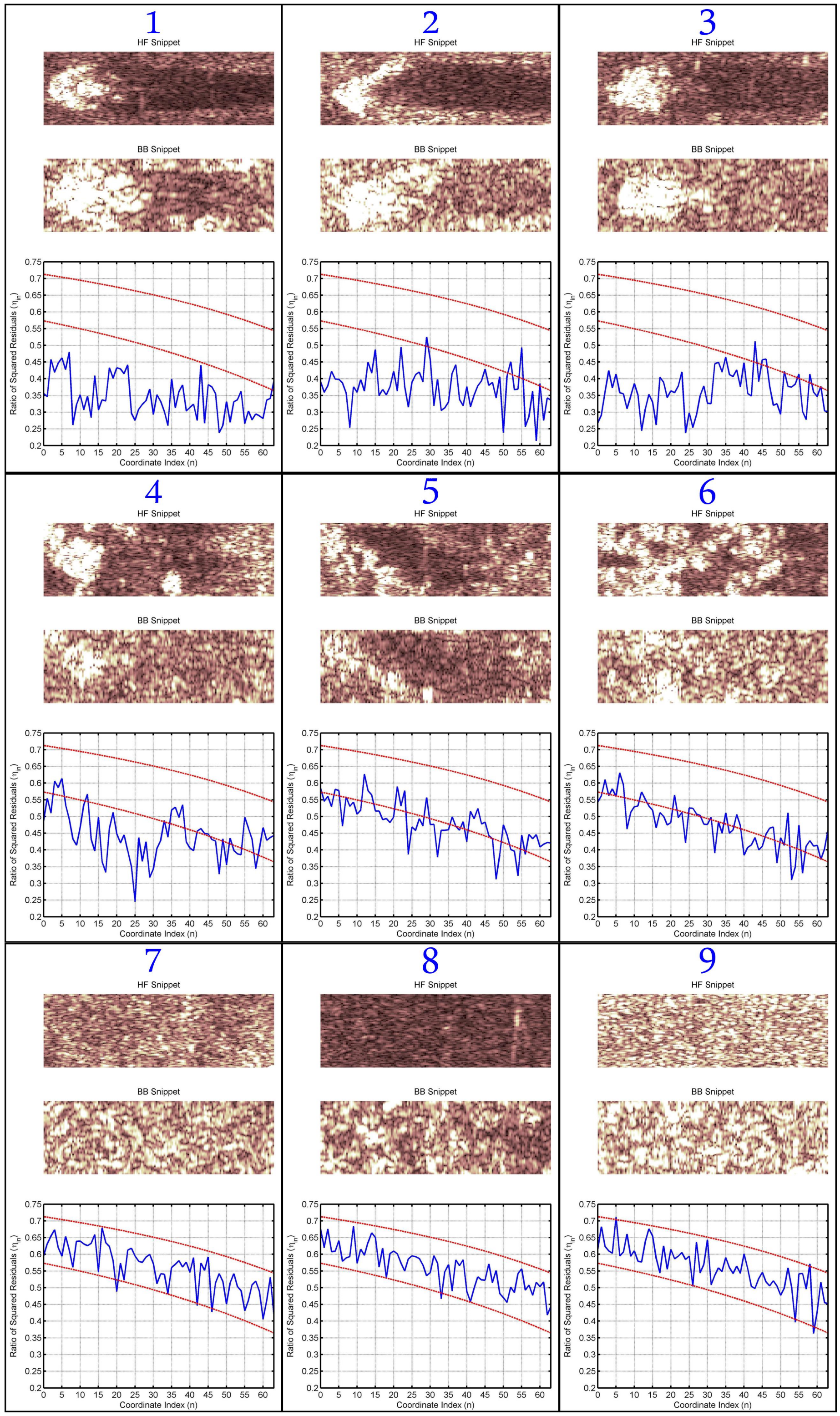


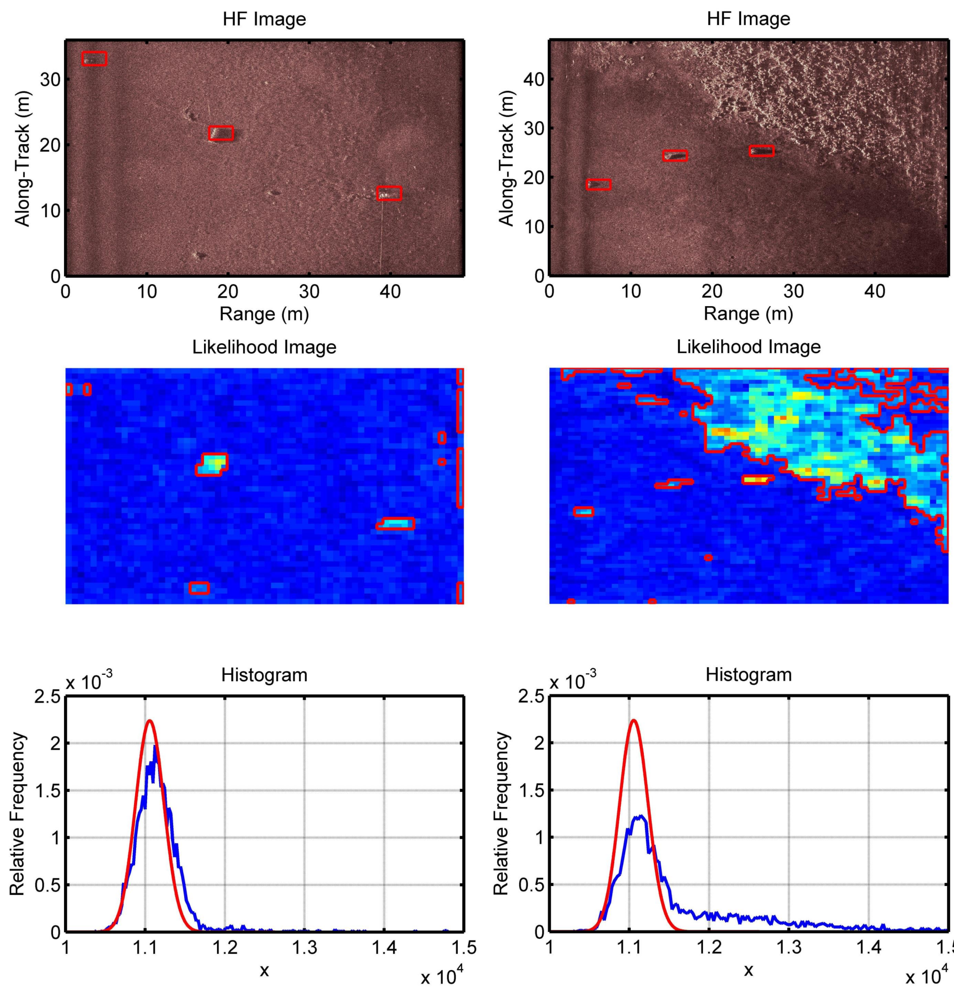
FIGURE 6.5. Ratio of squared residuals and their 95% confidence interval for each target/non-target window shown in Figure 6.4

Figure 6.4 gives an example of one HF image from Dataset 1 with several ROIs chosen throughout the image and highlighted with a red box. Windows 1-3 in this image correspond to targets, whereas windows 4-5 represent ROIs taken from the structured clutter field observed in the upper-left portion of the image, and windows 7-9 correspond to ROIs containing unstructured background. For each highlighted window in Figure 6.4, Figure 6.5 displays both the HF and BB ROIs as well as plots of the ratio of squared residuals given in (10)

$$\eta_{in} = \frac{\sigma_{in}^2(\hat{R})}{\sigma_{in}^2(\hat{R}_{ii})} = \frac{\mathbf{x}_{in}^H P_{ZX}^{\perp} \mathbf{x}_{in}}{\mathbf{x}_{in}^H P_{ZX}^{\perp} \mathbf{x}_{in} + \mathbf{x}_{in}^H P_{P_X^{\perp} Z} \mathbf{x}_{in}}$$

for $i = 2$ and $n = 0, \ldots, 63$. Recall that the product of these ratios produces the generalized Hadamard ratio Λ as demonstrated by the result given in (10). Exactly similar to the plots shown in Figures 5.5 and 5.6, the red lines in each of these plots represent a 95% confidence interval for these ratios under the null hypothesis and are constructed using the appropriate beta distribution. For windows 1-3 each containing a target, one can see a significant deviation from the interval indicating a strong deviation from independence. In contrast, for windows 7-9 containing unstructured clutter one can see that the interval accurately bounds these ratios giving a good indication that each pair of these ROIs are indeed uncorrelated. However, for windows 4-6 containing difficult structured clutter, one can see that the interval fails to bound the ratios and that there is a slight deviation from independence for these ROIs as well. In other words, the assumption that each η_{in} follows a beta distribution holds very well when the data contains no seafloor clutter but tends to be less applicable when clutter is present.

As the number of samples M used here is fairly small relative to the dimensions of each channel, the saddlepoint approximation was used to find the threshold needed to achieve a false alarm rate of $P_{FA} = 0.01$ for the scaled logarithm of the generalized Hadamard ratio,

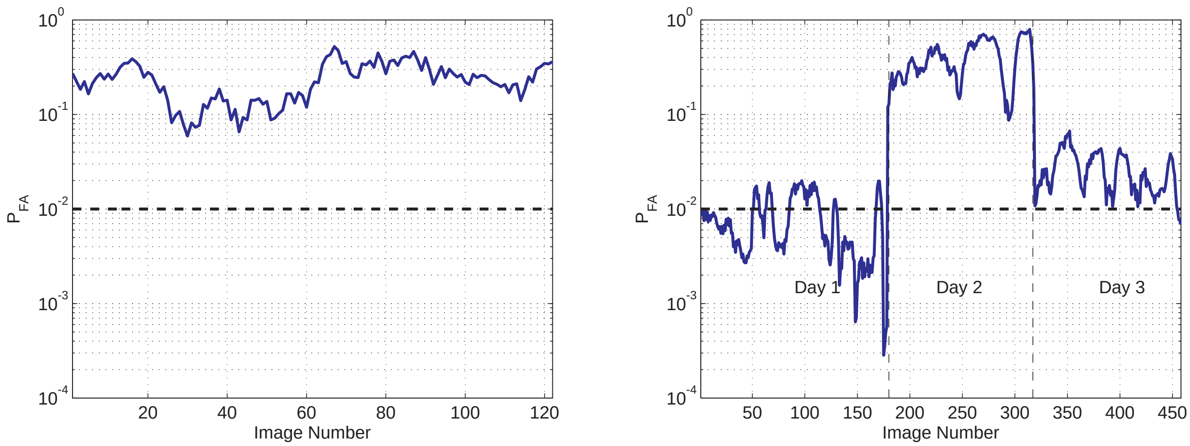


(a) Easy Background (b) Cluttered Background

FIGURE 6.6. Example of two HF images, the regions in each likelihood image that fall above the threshold, and the histograms of the likelihood ratio.

$\tilde{\Lambda} = \left(\frac{1}{3}(L+1)NP - 2M\right) \ln \Lambda$, given in (6). Recalling the discussion given in Section 4.3, this saddlepoint density is constructed by solving the saddlepoint equation given in (26) and substituting it into the density approximation given in (24). The threshold needed to achieve a false alarm rate of $P_{FA} = 0.01$ is then determined using this saddlepoint density function. Using this threshold, this modified likelihood ratio was applied to both Dataset 1 and Dataset 2 by partitioning every image into ROIs with 50% overlap and applying the data from each corresponding pair of HF and BB ROIs to the likelihood ratio. Figure 6.6 shows two examples of the results of this test for (a) a HF sonar image from Dataset 2 with no clutter and (b) a HF sonar image from Dataset 1 with difficult clutter in part of the image. The top image in both of these plots shows the HF image with several instances of target highlighted by a red box. The middle image displays the log-likelihood ratio $\tilde{\Lambda}$ for each ROI of this image and collections of ROIs that fall above the threshold have been outlined in red. Finally, the bottom plot displays the histogram of the log-likelihood values for that image in blue and compares it to the saddlepoint approximation of the density function which is plotted in red. From these two figures, one can see that the detector approximately achieves the desired false alarm rate in Figure 6.6 (a) and in the bottom-left of Figure 6.6 (b). However, the observed false alarm rate in the top-right of Figure 6.6 (b) is clearly much higher than what is desired and this clutter field's effect on the histogram of the log-likelihood ratio is clear.

Finally, Figures 6.7 (a) and (b) plot the empirical false alarm probability, i.e. the percentage of log-likelihood values for ROIs containing only background that fall above the threshold, for each image in both Dataset 1 and Dataset 2, respectively. Again, we can conclude that the methods developed in Chapters 3 and 4 can be used to determine thresholds that produce false alarm rates that are relatively close to what is desired in environments that lack any significant amount of clutter, such as day 1 and day 3 in Dataset 2. However, in environments where the clutter difficulty is medium to hard such as that observed in Dataset 1 and day 2 of Dataset 2, then the actual false alarm rate of the detector will be significantly higher than what is desired. The wide range of values observed in Figure 6.7 demonstrates just how difficult it can be to maintain a constant false alarm rate given the great variety of different environments and bottom conditions that can be encountered in this problem.



(a) Dataset 1

(b) Dataset 2

FIGURE 6.7. Probability of false alarm (P_{FA}) for each image in Dataset1 and Dataset 2 using the saddlepoint-based threshold.

6.4. SONAR IMAGERY DETECTION RESULTS

In this section, we apply the broadband coherence detector developed in Section 5.3 to the two datasets containing actual targets on the seafloor and compare its performance to the generalized Hadamard ratio given in (6) as well as a matched subspace detector [54] which specifically searches for highlight and shadow characteristics in the magnitude of the HF image only. Additionally, we will discuss the effects time delays and translation can have on these coherence-based methods and demonstrate these effects using the dataset containing simulated objects of different geometrical shapes.

6.4.1. APPLICATION OF BROADBAND COHERENCE TO SONAR IMAGERY. In this section, we give a preliminary demonstration of the broadband coherence detector by looking at several examples of the estimated covariance matrix $\hat{R}$ and the estimated magnitude coherence spectrum $\hat{\gamma}(e^{j\theta}, e^{j\phi})$ for both targets and background. The problem considered here is very similar to the simulations considered in Section 5.4.2. In this case, $x_1[n, p]$ represents the pixel of the HF image in the along-track/range coordinate system of the image

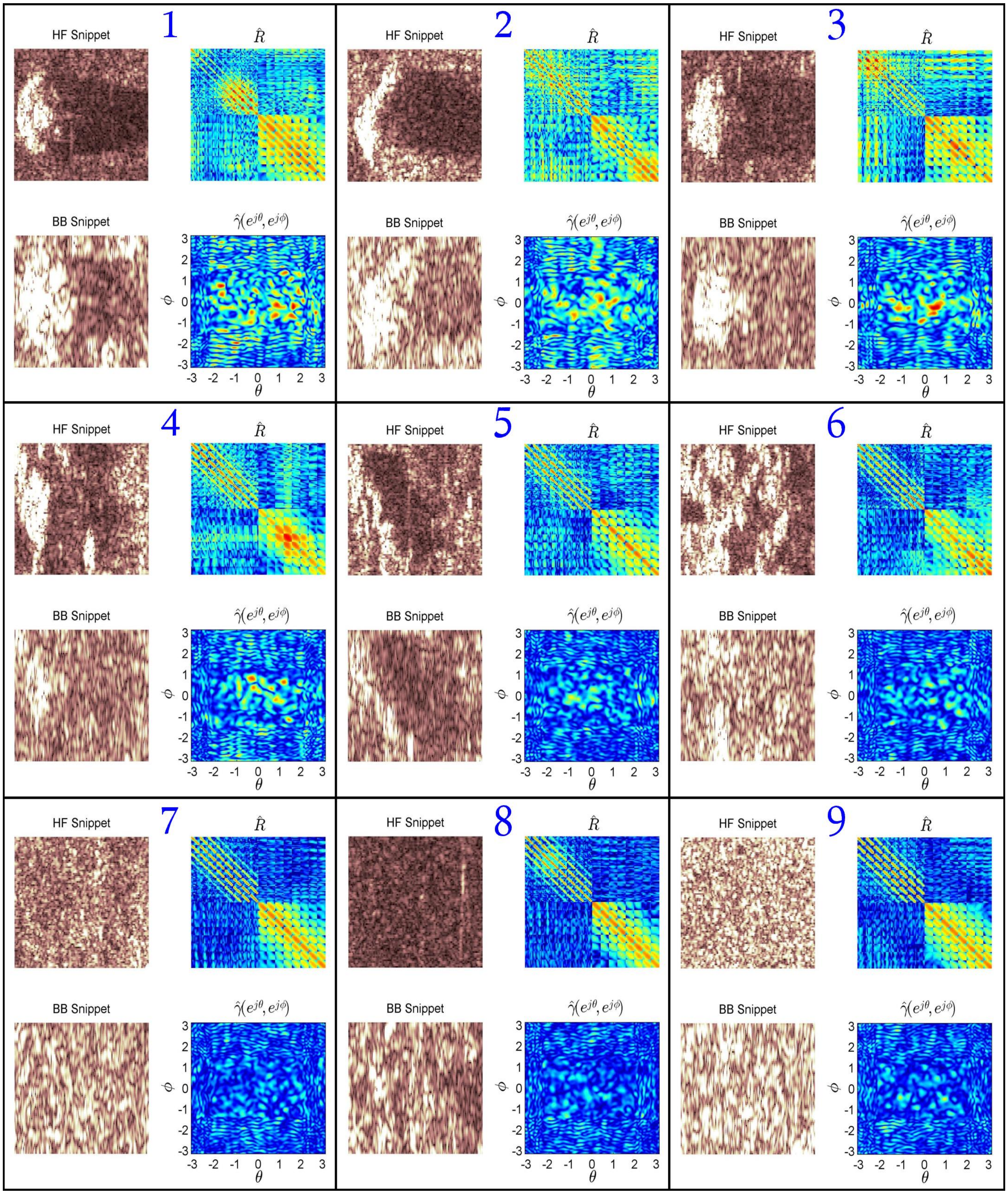


FIGURE 6.8. The HF and BB snippets, estimated covariance matrix $\hat{R}$, and estimated magnitude coherence $\hat{\gamma}(e^{j\theta}, e^{j\phi})$ for each window shown in Figure 6.4.

and $x_2[n, p]$ is the corresponding pixel for the BB image. Recalling the second invariance property discussed in Section 2.3.1, it's important to note that the results presented here would remain the same had we chosen to let x_1 represent the BB image and x_2 the HF image. In applying the broadband coherence detector given in (33), we assume that the data in this pair of images is spatially WSS, i.e. that there exists a sequence $\gamma_{ik}[l, m]$ for $l = 0, \dots, N-1$ and $m = 0, \dots, P-1$ such that

$$E\left[x_i[n+l, p+m]x_k^*[n, p]\right] = \gamma_{ik}[l, m] \in \mathbb{C} \;\forall\; i, k = 1, 2$$

As discussed in Section 5.3, this assumption will, in theory, manifest itself in the form of Toeplitz-block-Toeplitz structured matrices in each block R_{ik} of matrix R.

Very similar to the plots shown in Figure 5.11, Figure 6.8 displays the HF and BB snippets for each ROI delineated with a red box in the HF image shown in Figure 6.4 taken from Dataset 1. Using the processing steps described in Section 6.2, the top right of each window in this figure shows the estimated covariance matrix $\hat{R}$ using a block size of $N = P = 8$ for that pair of HF and BB ROIs. Likewise, the bottom right of each window shows the estimated magnitude coherence spectrum $\hat{\gamma}(e^{j\theta}, e^{j\phi})$ for each HF and BB pair using Welch's method with a $N = P = 32$ block size. Note that a smaller block size is required by the generalized Hadamard so as to produce a sufficiently large number of samples (M) such that the unconstrained covariance estimate $\hat{R} \in \mathbb{C}^{LNP \times LNP}$ will be PD, i.e. large block sizes easily result in sample poor conditions. On the other hand, the broadband coherence detector only requires the estimation of the cross-spectral matrix $\hat{S}(e^{j\theta}, e^{j\phi}) \in \mathbb{C}^{L \times L}$, a matrix of much lower dimension, and is thus not nearly as constrained when it comes to an appropriate block size.

Looking at the estimated covariance matrix for each example shown in Figure 6.8, one can for the most part observe the Toeplitz-block-Toeplitz structure associated with WSS data in the autocovariance matrices for both channels. This is especially true for the examples shown in windows 7-9 containing unstructured clutter. However, the cross-covariance matrix $\hat{R}_{12} = \hat{R}_{21}^H$ generally lacks this Toeplitz-block-Toeplitz structure. In other words, the assumption that each channel is individually WSS seems to be a fairly applicable assumption for this problem but the assumption that they are jointly WSS is harder to argue. Be that as it may, one can see that the magnitude coherence tends to be significantly higher for windows 1-3 containing targets than when those windows contain background and clutter. Moreover, one can also see that the coherence patterns themselves tend to be different when comparing target to structured clutter. Hence, although the WSS assumptions behind the broadband coherence detector may not directly apply for this problem, the results shown in Figure 6.8 make it clear that broadband coherence is still a useful tool for discriminating between those ROIs that contain target from those that only contain background (structured or unstructured).

6.4.2. Sensitivity to Coregistration. As the proposed coherence-based detection techniques require the use of pairs of HF and BB images that are coregistered over the seafloor, one important question that arises in the practical application of these methods is the effect relative translation and deviations from coregistration can have on target detection. This question is very much equivalent to asking how relative time delays among two or more length-N time series affects the ability of these methods to adequately determine whether or not these time series are linearly dependent. To answer this question, we will begin by investigating the effect of time delay on a particular performance criteria associated with the generalized Hadmard ratio.

The performance criteria which we are interested in studying is the absolute separation between the mean of the log of the generalized Hadamard ratio under the alternative and null hypotheses. To develop this criteria, we will begin by making the same assumption used at the beginning of Section 3.3 and that is, under the null hypothesis, the data matrix $\mathcal{Z}$ given in (5) contains *iid* realizations of $\mathcal{CN}(0,1)$ random variables. What differentiates the alternative hypothesis from the null hypothesis is that this data matrix will be colored with the non-block-diagonal transformation $R^{1/2}$ so that the absolute difference in mean, δ, can be expressed as

$$\begin{aligned}
\delta &= \left| E\left[\ln \Lambda | \mathcal{H}_1\right] - E\left[\ln \Lambda | \mathcal{H}_0\right] \right| \\
&= \left| E\left[\ln \frac{\det\left(R^{1/2}\mathcal{Z}\mathcal{Z}^H R^{H/2}\right)}{\prod_{i=1}^{L} \det\left(R_{ii}^{1/2} X_{iN} X_{iN}^H R_{ii}^{H/2}\right)}\right] - E\left[\ln \frac{\det\left(\mathcal{Z}\mathcal{Z}^H\right)}{\prod_{i=1}^{L} \det\left(X_{iN} X_{iN}^H\right)}\right] \right| \\
&= \left| \ln \frac{\det R}{\prod_{i=1}^{L} \det R_{ii}} + E\left[\ln \frac{\det\left(\mathcal{Z}\mathcal{Z}^H\right)}{\prod_{i=1}^{L} \det\left(X_{iN} X_{iN}^H\right)}\right] - E\left[\ln \frac{\det\left(\mathcal{Z}\mathcal{Z}^H\right)}{\prod_{i=1}^{L} \det\left(X_{iN} X_{iN}^H\right)}\right] \right| \\
&= -\ln \frac{\det R}{\prod_{i=1}^{L} \det R_{ii}}
\end{aligned} \tag{37}$$

Here, we've used the same notation used in Section 3.3 where the data matrix X_{iN} represents the collection of all M samples of the time series $\mathbf{x}_i = \left[x_i[0] \ \cdots \ x_i[N-1]\right]^T$, i.e.

$$X_{iN} = \begin{bmatrix} x_i[0,1] & x_i[0,2] & \cdots & x_i[0,M] \\ x_i[1,1] & x_i[1,2] & \cdots & x_i[1,M] \\ \vdots & \vdots & \ddots & \vdots \\ x_i[N-1,1] & x_i[N-1,2] & \cdots & x_i[N-1,M] \end{bmatrix}$$

Thus, the magnitude of the difference in mean of the log-likelihood ratio between these two hypotheses is simply the negative log of the generalized Hadamard ratio of the true model

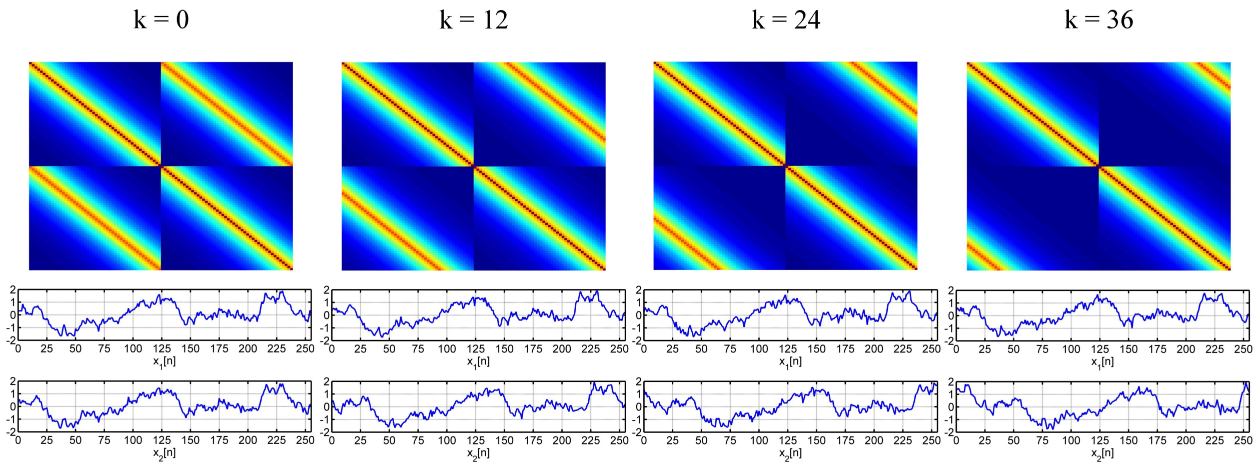


FIGURE 6.9. Delaying the signal from one channel produces a shift in the cross-covariance matrix $R_{12} = R_{21}^H$.

represented by the matrix R under the alternative hypothesis. This measure is completely independent of the number of samples M used in the formation of the generalized Hadamard ratio.

To study the effects time delays have on value of δ, consider a situation where $L = 2$ channels observe a common autoregressive signal $s[n]$ plus white Gaussian noise $w_i[n]$ where the signal received by the second channel is delayed by an integer number of k samples. That is, we consider the following model under the alternative hypothesis

$$\begin{aligned} x_1[n] &= s[n] + w_1[n] \\ x_2[n] &= s[n-k] + w_2[n] \end{aligned}$$

where $w_i[n] \overset{iid}{\sim} \mathcal{CN}(0,1)$ for $i = 1, 2$ and $n = 0, \ldots, N-1$ and $s[n]$ is a first-order autoregressive process with coefficient a and white noise variance σ^2. With $a = 0.9$ and $\sigma^2 = 2$, Figure 6.9 displays the true covariance matrix R under this model with an increasing value of the delay k as well as simulated examples of the time series for both channels in each case. As the time delay grows large, one can see that the high correlation peak in the cross-covariance

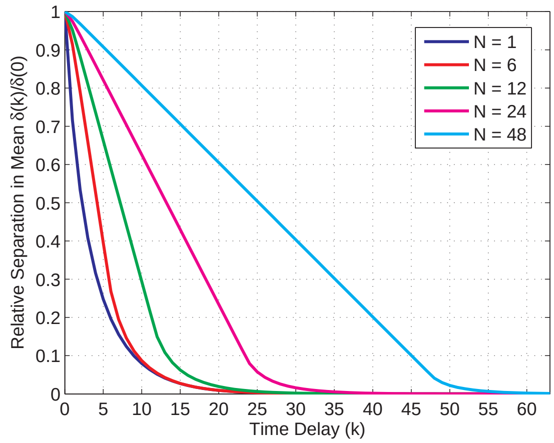
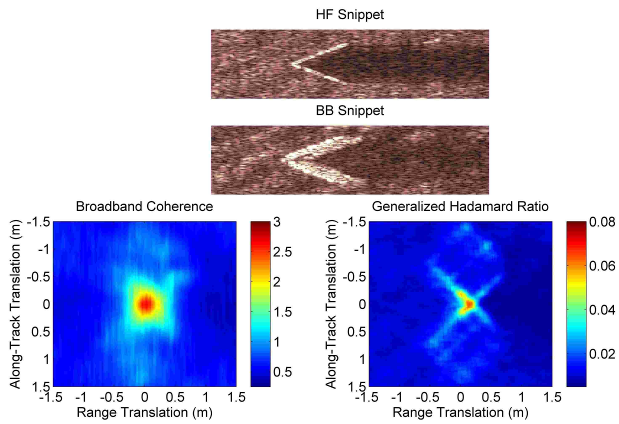


FIGURE 6.10. The relative separation in mean $\delta(k)/\delta(0)$ as a function of increasing time delay.

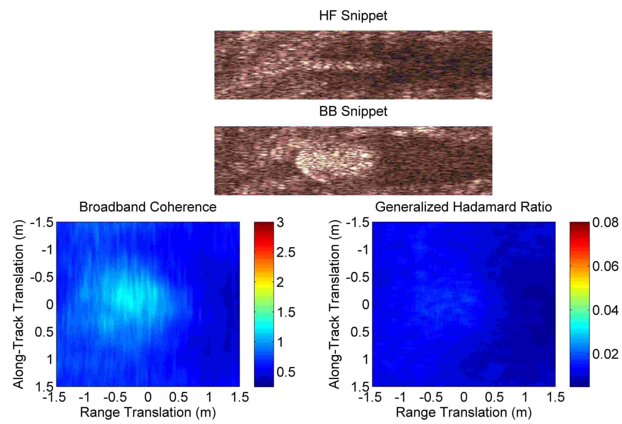
matrix $R_{12} = R_{21}^H$ gradually shifts out of the picture. The farther the shift in this correlation peak, the more and more block-diagonal the matrix R becomes making it more difficult to discriminate the alternative hypothesis from the null hypothesis.

For several choices in N, Figure 6.10 plots the relative separation in mean $\delta(k)/\delta(0)$, i.e. the value of δ given in (37) with time delay k normalized by its maximum value at $k = 0$. From this figure, one can see that the separation in mean generally decreases with an increasing delay in time. However, one also sees that the detector is generally more robust to time delays with the use of a longer time series for each channel, i.e. a larger value for N. For example, the generalized Hadamard ratio with $N = 24$ exhibits a 20% decrease in separation with every delay of 5 samples while there is only a 10% decrease when $N = 48$. Thus, when it comes to the effects of relative time delays on the separation in mean of this test statistic, the larger the value of N the more robust the detector.

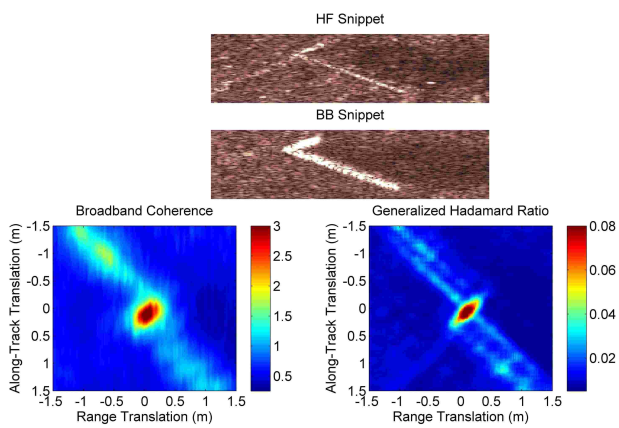
Now, to demonstrate the effects of spatial translation on the detection of targets among two HF and BB images, a test was conducted using Dataset 3 containing actual images of the seafloor with synthetically generated targets inserted into the image. Figures 6.11 (a)-(c) shows examples of the HF and BB snippets for the block, cone, cylinder, and sphere target



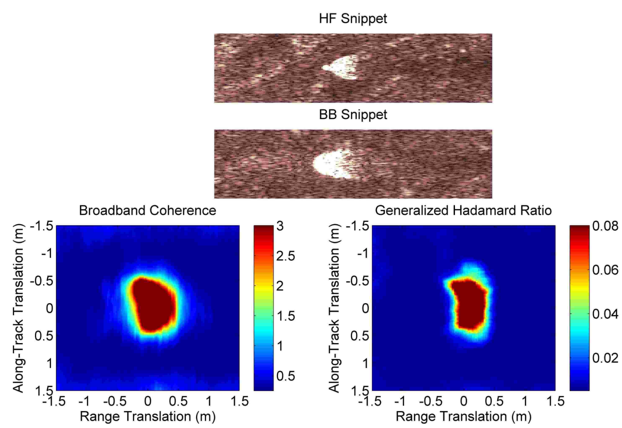
(a) Block



(b) Cone



(c) Cylinder



(d) Sphere

FIGURE 6.11. The likelihood ratio for both broadband coherence and the generalized Hadamard ratio as a function of translation in along-track and range.

types, respectively. For each of these four targets, the BB image was then translated in both the range and along-track directions while the HF image was left fixed to mimic deviations from coregistration. The plot in the bottom left of each of these figures displays the log of the likelihood ratio given in (33) with a $N = P = 32$ block size as a function of spatial translation in meters. The plot in the bottom right of each figure likewise shows the log of the likelihood ratio given in (6) with a $N = P = 8$ block size. Although the likelihood values for the cone target in Figure 6.11 (b) are relatively low, one can again see that both detectors maintain relatively high values for low spatial tranlation. Looking at the results in Figures

6.10 and 6.11 one can conclude that, while these coherence-based detection methods are not completely invariant to the effects of time delays and spatial translation, one can afford some deviation from coregistration before completely losing the ability to detect a signal common to multiple channels. In fact, the results in Figure 6.11 suggest that the broadband coherence detector can withstand approximately 0.5 m of translation in either along-track or range before losing the ability to detect these four targets. The results of Figure 6.11 also make it clear that the broadband coherence detector's use of a larger 32×32 block size results in a detector that is generally more robust to deviations from coregistration compared to the generalized Hadamard ratio using a smaller 8×8 block size. This conclusion is consistent with the results given in Figure 6.10 which show that higher dimensional observations tend to be more robust to the effects of relative translation.

6.4.3. Comparison Studies and Detection Results. In this section, we apply the broadband coherence detector in (33) and compare its performance to the generalized Hadamard ratio in (6) as well as a matched subspace detector [54] using Dataset 1 and Dataset 2 in Table 6.1. The processing steps used to compute the likelihood ratio for both the broadband coherence and generalized Hadamard ratio are the same as those mentioned in Section 6.2. The detection methods studied in this section will be compared using performance metrics such as Receiver Operating Characteristic (ROC) curve characteristics, probability of detection (P_D), and average number of false alarms per image (FA/Image).

In addition to studying the performance of the coherence-based detection methods given in (6) and (33), we will also benchmark the performance of these methods with a detection technique that specifically looks for the highlight and shadow characteristics typically associated with targets. For this technique, let the vector $\mathbf{x} \in \mathbb{R}^M$ denote the vector formed by vectorizing the magnitude of the pixels in the HF image only where $M = 80 \times 144 = 11520$

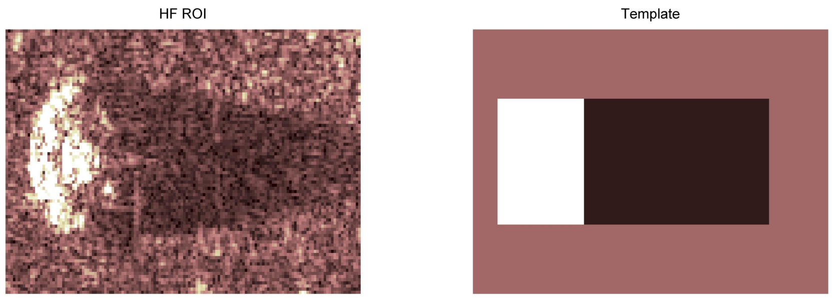


FIGURE 6.12. The template $\mathbf{h}$ used in the matched subspace detector is constructed to mimic the highlight/shadow attributes associated with targets lying proud on the seafloor.

corresponds to the total number of pixels in each ROI. One technique commonly employed for the detection of targets lying proud on the seafloor is to construct a template $\mathbf{h} \in \mathbb{R}^M$ that mimics the highlight and shadow characteristics of targets and to find regions of the seafloor that best match that template. Here, $\mathbf{h}$ is the vector formed by vectorizing the idealistic template given on the right of Figure 6.12. Comparing this template to the example of a HF ROI containing a target on the left-hand side of Figure 6.12, the resemblance with the template is not exact but it generally captures the behavior of the target.

As this detection technique relies solely on the magnitude of the HF image, we assume that the observation $\mathbf{x} \in \mathbb{R}^M$ follows the linear model

$$\mathbf{x} = \mu\mathbf{h} + \phi\mathbf{1} + \mathbf{w}$$

where $\mathbf{w}$ is a vector of white noise, $\mathbf{w} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 I\right)$, with unknown variance σ^2 and $\mathbf{1}$ denotes an M-dimensional vector of ones. Thus, the data for any ROI is assumed to consist of signal, $\mu\mathbf{h}$, plus an unknown bias representing an unknown average pixel intensity, $\phi\mathbf{1}$, plus noise of unknown variance, $\mathbf{w}$. Given this model, the matched subspace detector [54] tests the null hypothesis that the signal (target) is absent ($\mu = 0$) versus the alternative that it is present with a strictly positive scaling ($\mu > 0$), i.e. we consider the hypothesis test

$$\begin{aligned} \mathcal{H}_0 &: \mu = 0 \\ \mathcal{H}_1 &: \mu > 0 \end{aligned}$$

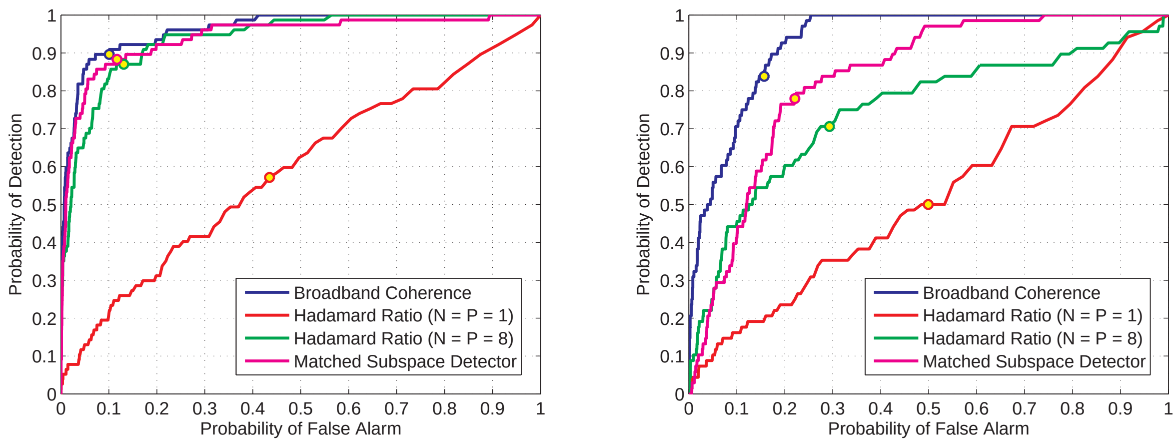
Note that if the scaling μ associated with the signal were less than zero then the template shown in Figure 6.12 would exhibit a region of relatively low intensity followed by a region of relatively high intensity. However, the constraint that $\mu > 0$ under the alternative hypothesis is enforced as targets always exhibit a high intensity highlight followed by a low intensity shadow. Letting $\boldsymbol{\theta} = [\mu\ \phi\ \sigma^2]^T$ denote the vector of unknown parameters, employing the GLRT given in (2) yields the likelihood ratio [54]

$$\Lambda = \begin{cases} 0 & \hat{\mu} \leq 0 \\ \dfrac{\mathbf{x}^T P_{\mathbf{1}}^{\perp} P_{\mathbf{g}} P_{\mathbf{1}}^{\perp} \mathbf{x}}{\mathbf{x}^T P_{\mathbf{1}}^{\perp} P_{\mathbf{g}}^{\perp} P_{\mathbf{1}}^{\perp} \mathbf{x}} & \hat{\mu} > 0 \end{cases} \tag{38}$$

where $\hat{\mu} = \frac{\mathbf{h}^T P_{\mathbf{1}}^{\perp} \mathbf{x}}{\mathbf{h}^T P_{\mathbf{1}}^{\perp} \mathbf{h}}$ is the ML estimate of μ and $P_{\mathbf{g}}$ is the projector $P_{\mathbf{g}} = P_{P_{\mathbf{1}}^{\perp}\mathbf{h}}$. Note that the projection matrix $P_{\mathbf{1}}^{\perp} = I - \frac{1}{M}\mathbf{1}\mathbf{1}^T$ is an $M \times M$ centering matrix which simply subtracts the average value from any vector since if $\mathbf{x} = [x_1\ \cdots\ x_M]^T$ then

$$P_{\mathbf{1}}^{\perp}\mathbf{x} = \mathbf{x} - \frac{1}{M}\mathbf{1}\mathbf{1}^T\mathbf{x} = \mathbf{x} - \left(\frac{1}{M}\sum_{i=1}^{M} x_i\right)\mathbf{1}$$

When applying this detection technique to the images of Dataset 1 and Dataset 2 the images are again partitioned to ROIs and applied to the likelihood ratio given in (38). As the location of the target in the ROI is important, however, a 90% overlap in both the along-track and range dimensions is introduced as opposed to the 50% overlap employed for the broadband coherence and generalized Hadamard ratio detectors. Again, the underlying principle behind this detection technique is fundamentally different from the coherence-based methods given



(a) Dataset 1

(b) Dataset 2

FIGURE 6.13. Comparison of the ROC curves for each detector.

in (6) and (33). While the likelihood ratios of (6) and (33) are used to find regions where the complex-valued data within the HF and BB images is highly correlated, the likelihood ratio given in (38) is designed to find regions in the magnitude of the HF image which provide a good match with the template shown in Figure 6.12.

The detection methods developed in this work were then applied to the two datasets containing actual targets lying on the seafloor. For the studies of this section, four different detection methods were considered: the broadband coherence detector given in (33) with a $N = P = 32$ block size, the generalized Hadamard ratio given in (6) with a $N = P = 8$ block size, the likelihood ratio given in (6) but with $N = P = 1$ so that the pixels in each pair of ROIs are treated as independent samples and the likelihood ratio is a Hadamard ratio [28], and finally the matched subspace detector given in (38). Using all the targets in both Dataset 1 and Dataset 2 as well as a randomly selected set of background ROIs extracted from the images of both datasets, Figures 6.13 (a) and (b) compare the ROC curves of all four detectors for Dataset 1 and Dataset 2, respectively. For reference, the knee-point of each ROC curve, i.e. the point where $P_D + P_{FA} = 1$, is shown as a small circle in both of these figures. Looking at the results in Figure 6.13 (a), one can see that the broadband coherence, generalized Hadamard ratio, and matched subspace detectors all perform fairly similarly on Dataset 1. However, the broadband coherence detector clearly outperforms all the alternatives considered here for Dataset 2 as can be seen in Figure 6.13 (b) while matched subspace detector tends to perform better than the generalized Hadamard ratio, especially for high false alarm probabilities. Also, one can see from these figures that the Hadamard ratio fails to discriminate target from background for both datasets. Recalling the arguments made in Section 6.4.2 and the results shown in Figure 6.10, this poor performance may be attributed to small deviations in coregistration between the HF and BB images of both datasets which is particularly noticeable when using small block sizes.

Using a small subset of images from each dataset, a threshold was then selected for each detector and empirically set based on the likelihood values from this subset of images to achieve a false alarm rate of $P_{FA} = 10^{-2}$. Using these selected thresholds, all four detection methods were then applied to both Dataset 1 and Dataset 2 with this subset of images reinserted into both datasets. For each detection method, all overlapping ROIs that produce a likelihood ratio that exceeds its corresponding threshold were accumulated into one detection. If the location of that detection is within 2.5 m of the known ground truth location of a target, the detection is labeled a correct detection else it is reported as a false alarm. Using this performance criteria, Table 6.2 gives the detection rates for each method on both Dataset 1 and each day (environment) of Dataset 2. Similar to what was observed in Figure 6.13, the Hadamard ratio performs poorly on these datasets, detecting only about a third of the total number of targets. The other three alternatives perform better, however, with overall detection rates ranging from a low of about 70% to a high of nearly 90%. From this table, one can also see that the broadband coherence detector performs well with

TABLE 6.2. Comparison of the Detection Rates (P_D) for Each Method

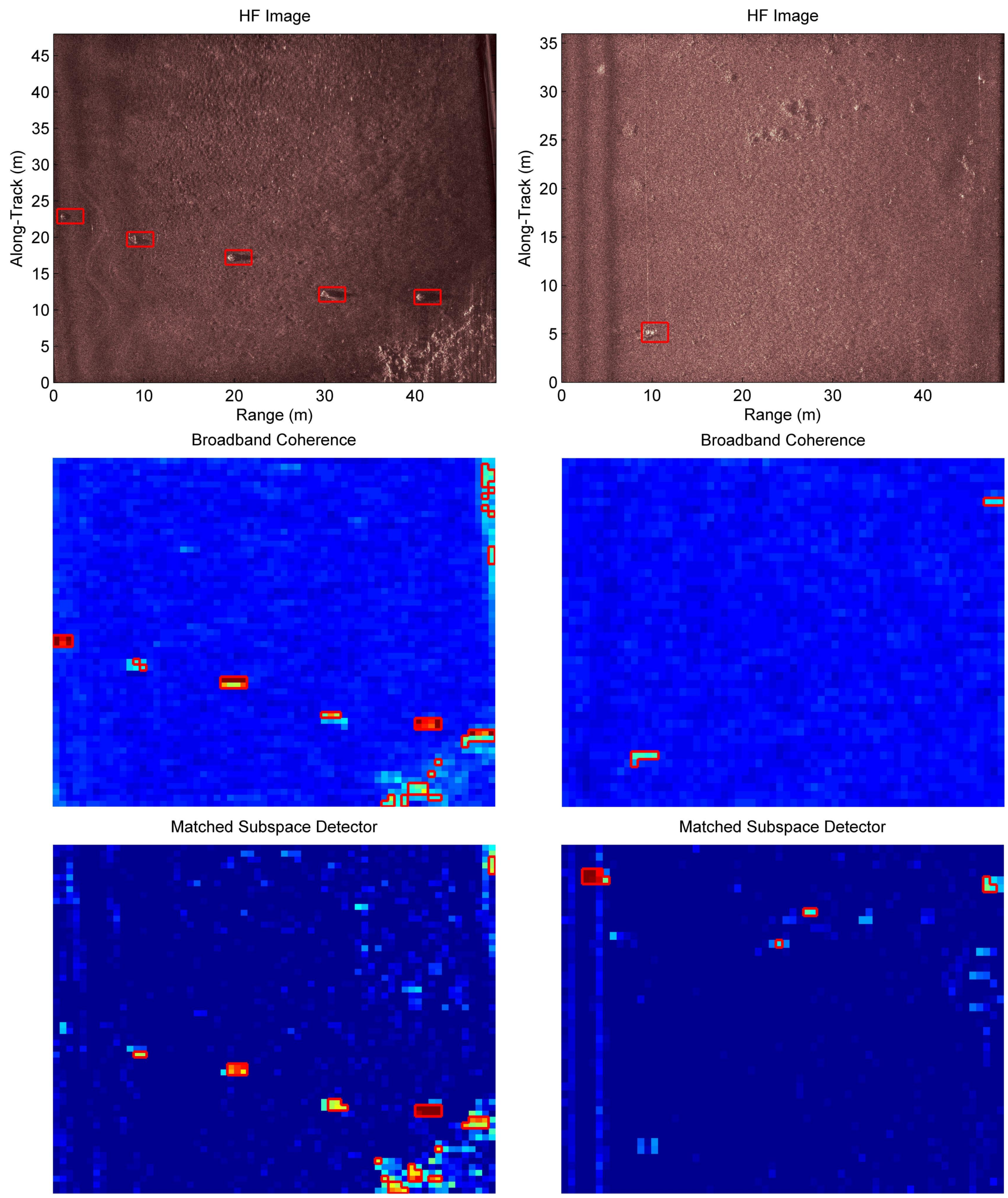
Detection Method	Dataset 1	Dataset 2			
	–	Day1	Day 2	Day 3	Total
Broadband Coherence	68 (88%)	4 (100%)	16 (94%)	39 (83%)	59 (87%)
Hadamard Ratio ($N = P = 1$)	29 (38%)	2 (50%)	7 (41%)	8 (17%)	17 (25%)
Gen. Hadamard Ratio ($N = P = 8$)	61 (79%)	4 (100%)	14 (82%)	31 (66%)	49 (72%)
Matched Subspace Detector	67 (87%)	4 (100%)	14 (82%)	38 (81%)	56 (82%)

TABLE 6.3. Comparison of the Average Number of False Alarms per Image (FA/Image) for Each Method

Detection Method	Dataset 1	Dataset 2			
	–	Day1	Day 2	Day 3	Total
Broadband Coherence	12.6	0.5	26.0	2.4	8.64
Hadamard Ratio ($N = P = 1$)	12.4	8.7	14.3	8.0	10.13
Gen. Hadamard Ratio ($N = P = 8$)	10.2	1.1	24.6	3.4	8.83
Matched Subspace Detector	13.2	2.3	40.4	4.8	14.4

detection rates of 88% and 87% on Dataset 1 and Dataset 2, respectively, compared to its closest competitor, the matched subspace detector, which achieves corresponding detection rates of 87% and 82%. Finally, Table 6.3 gives the average number of false alarms per image for each method when applied to both datasets. From this table it is clear that all methods exhibit a relatively high false alarm rate for the difficult environments presented by Dataset 1 and day 2 of Dataset 2 while the false detection rates are significantly smaller for day 1 and 3 of Dataset 2. One can also see that, while the broadband coherence detector exhibits the second highest false alarm rate on Dataset 1, it achieves the lowest overall false alarm rate for Dataset 2 among all the alternatives considered here.

As a final comparison between the broadband coherence and matched subspace detectors, the top two images of Figures 6.14 (a) and (b) give examples of two HF images with the targets in each highlighted with a red box. The plot in the middle of each figure displays the likelihood ratio of the broadband coherence detector for each ROI in their respective image while the bottom two plots show the same for the matched subspace detector. Groups of overlapping ROIs that fall above the empirically chosen threshold for each method are



(a) Example Image 1

(b) Example Image 2

FIGURE 6.14. Two HF images with targets and comparison of the areas detected by the broadband coherence and matched subspace detectors.

surrounded with a red box in these plots. The matched subspace detector fails to detect the target at the far left of Figure 6.14 (a) which exhibits a low intensity highlight and short shadow as the target is close to the AUV in range. The matched subspace detector also fails to detect the target in Figure 6.14 (b) which exhibits no shadow region behind the highlight. In both cases, the template in Figure 6.12 fails to adequately model these targets and produces low likelihood ratio values as a result. However, the broadband coherence detector, which successfully detects both targets, does not rely on any specific model of the target and simply looks for high levels coherence among the HF and BB data. These two examples demonstrate why this simple principle can in some cases be an advantage for this problem.

6.5. Conclusion

In this chapter we applied the results of Chapters 3, 4, and 5 to the detection of underwater targets in pairs of coregistered HF and BB SAS images formed from the data collected using the SSAM system. The underlying detection principle behind these coherence-based approaches is that the presence of a target in any pair of ROIs produces a higher level of coherence than when those ROIs contain background alone. To demonstrate the results of this dissertation the methods were applied to three sonar imagery datasets, two consisting of actual targets lying the seafloor with each being collected in different environments with unique difficulties and one dataset consisting of actual SAS images of the seafloor with synthetically generated targets of different geometrical shapes inserted into the images.

First, the usefulness of the results concerning the null distribution of the generalized Hadamard ratio discussed in Chapters 3 and 4 was studied. Using one image containing targets and both cluttered and smooth background as an example, the ratio of squared residuals used in the construction of the likelihood ratio were compared to a 95% confidence interval knowing the fact that each ratio is beta distributed under the null hypothesis. From the few examples of ROIs chosen in this study, it was found that this bound tends to capture the behavior of these ratios when that ROI contains background with no clutter. However, the same bound fails to capture their behavior when that ROI contains difficult clutter density. A saddlepoint approximation was then used to find the threshold needed to achieve a desired false alarm probability and the actual false alarm rate for that threshold was measured when applied to both real datasets. From the results of this study it was observed that these methods are capable of approximately achieving the desired false alarm rate in environments with low or no clutter density. However, in environments with difficult clutter density, the realized false alarm rates will be much higher than what is desired.

The broadband coherence detector developed in Chapter 5 was then applied to all three datasets. Looking at the separation in mean of the generalized Hadamard ratio, we first studied the effects time delays can have on detectability and observed that the robustness of this performance criteria tends to increase with increasing time series length. This principle was then demonstrated by comparing the likelihood ratio values produced by the broadband coherence detector and generalized Hadamard ratio for several simulated targets of different geometrical shapes as the data of the BB image is spatially translated relative to the HF image. As a result of these studies we concluded that, while these coherence-based detection methods are not completely invariant to spatial translation, the proposed technique can withstand approximately 0.5 m of translation in either along-track or range before completely losing the ability to detect each target. Moreover, we found that the broadband coherence detector's use of a larger block size results in a detector that is generally more robust to deviations from coregistration compared to the smaller block sizes used for the generalized Hadamard ratio. If the total amount of data, i.e. ROI size, remains fixed, however, the disadvantage of using larger block sizes is that it reduces the number of samples used to form estimates of the second-order parameters required by these coherence-based techniques.

The performance of the broadband coherence detector on the two real sonar datasets was then compared to the generalized Hadamard ratio using both pixel and block-based observations as well as a matched subspace detector designed to find regions in the magnitude of the HF image that exhibit the highlight and shadow characteristics typical of targets lying proud on the seafloor. A comparison of the results for each method shows that the broadband coherence, generalized Hadamard ratio, and matched subspace detectors all perform fairly similarly on Dataset 1 but the broadband coherence detector clearly outperforms its alternatives on Dataset 2. The Hadamard ratio performs poorly on both datasets which may be attributed to a lack of robustness to even slight deviations from coregistration. Through the results presented in this section, one can see that the fundamental principle of detecting underwater targets using coherence-based approaches is itself a very useful solution for this problem and that the broadband coherence statistic is adequately adept at achieving this.

CHAPTER 7

Conclusions and Suggestions for Future Work

7.1. Conclusions and Discussions

The problem of underwater target detection in multiple sonar images is complicated due to various factors such as variations in operating and environmental conditions, presence of spatially varying clutter, variations in target shapes, compositions and orientation. Moreover, bottom features such as coral reefs, sand formations, and vegetation may obscure a target and can be a source of high false alarm rates. Due to the wide variation in target conditions that can be observed for this problem, detection methods that take advantage of general discriminative features in the data, as opposed to those that rely on specific target models, can in some cases be desirable. Such is the case for the solution presented here where the detection principle simply relies on the assumption that the presence of targets in coregistered HF and BB images will lead to a higher degree of coherence than when those images contain background alone. Posing the problem as a test of independence among multiple data channels, the Generalized Likelihood Ratio Test (GLRT) is a generalized Hadamard ratio of a composite covariance matrix estimated from multiple *iid* samples.

The first subject discussed in this work was the characterization of the null distribution of the generalized Hadamard ratio for finding thresholds that achieve a desired false alarm probability. Using the theory of Gram determinants, it was shown in Chapter 3 that this test statistic can be written as a product of ratios of the squared residual from two linear prediction problems. Geometrical insights into these ratios leads to the conclusion that, under the null hypothesis, the generalized Hadamard ratio is stochastically equivalent to a product of independent beta random variables. This stochastic representation makes it very straightforward to derive various attributes of the null distribution of this test statistic including its moments, characteristic function, and cumulant generating function. Asymptotically, the characteristic function of this random variable was shown to converge to a chi-squared random variable as the number of *iid* samples used to form the estimated composite covariance matrix grows large. However, even with the inclusion of a scaling designed to improve the rate of convergence, results show that the distribution is slow to converge. For this reason, we turned our attention to the use of saddlepoint approximations [22], [39] which, with the help of numerical root-finding algorithms to find the saddlepoint, gives a practical alternative for determining the threshold needed to approximately achieve a desired false alarm rate. Simulation results show that the saddlepoint approximation is a very useful alternative for capturing the distribution of the log-likelihood ratio under the null hypothesis, especially in situations where L or N is large. If both L and N are small, however, results show that the asymptotic chi-squared distribution is probably more practical in these scenarios and tends to produce false alarm rates closer to what is desired.

The second subject of this dissertation was an alternative implementation of the generalized Hadamard ratio which exploits the inherent Toeplitz-block-Toeplitz structure of the composite covariance matrix when the data collected by each channel is jointly wide-sense stationary (WSS) so as to produce detection methods that are better matched to the problem under the stationarity assumption. Although the true GLRT for this problem would impose this Toeplitz structure on the ML estimate of the composite covariance matrix this is an intractable problem with no closed-form solution. However, the asymptotic behavior of Toeplitz matrices is well understood and leads to very tractable results involving various matrix operations for Toeplitz matrices. Using this asymptotic theory, the GLRT is expressed as the broadband integral of a narrowband Hadamard ratio over the 2D frequency plane in Chapter 5, a test statistic referred to as broadband coherence. The broadband coherence detector was then demonstrated using simulations of two different multichannel detection applications. In the first simulation, we considered the problem of detecting a single wideband source using several ULAs laid out in a linear geometry. For the second simulation, we considered the problem of detecting correlation among a pair of images generated using a 2D multivariate autoregressive model. In both cases, this broadband coherence detector is shown to provide substantial improvements in performance compared to alternative techniques which, while more generally applicable, do not exploit stationarity and its manifistation in a Toeplitz structure for the composite covariance matrix.

These results were then applied in Chapter 6 to the problem of underwater target detection in pairs of high frequency (HF) and broadband (BB) sonar images coregistered over the seafloor. Here, the detection hypothesis is that the presence of a target in this pair of images will lead to a higher level of coherence compared to situations where they contain background alone and that the difference in coherence in these two situations will be sufficiently high to adequately discriminate one from the other. To demonstrate the results of this dissertation the methods were applied to three datasets, two consisting of actual targets lying the seafloor with each being collected in different environments with unique difficulties and one dataset consisting of actual sonar images of the seafloor with synthetically generated targets of different geometrical shapes inserted into the images.

Using the two real sonar datasets, the first study involved the usefulness of the results concerning the null distribution of the generalized Hadamard ratio developed in Chapters 3 and 4 for this problem. Using one pair of HF and BB sonar images as an example, the ratio of squared residuals that comprise the generalized Hadamard ratio were compared to a confidence interval designed to capture their behavior under the null hypothesis. From these few examples, it was observed that the bound accurately captures the behavior of these ratios for uncluttered background but fails when the background contains significant clutter density. As the number of samples used to form maximum likelihood (ML) estimates of the composite covariance matrix were relatively small for this application, a saddlepoint approximation was used to find the threshold needed to achieve a false alarm rate of $P_{FA} = 0.01$. The actual false alarm rate subsequently measured when applied to all the different environments encompassed by the two real sonar datasets. These results seemed to suggest that the saddlepoint approximation, and more importantly the fact that the likelihood ratio is distributed as a product of independent betas under the null hypothesis, is capable of producing thresholds which approximately achieve the desired false alarm rate in environments with low clutter density. However, in environments with high clutter density, the realized false alarm rates can be much higher than what is desired.

All three sonar imagery datasets were then used to demonstrate the broadband coherence detector developed in Chapter 5. Before analyzing the performance of the proposed detector, we first studied the effects of deviation in coregistration among the HF and BB images. This principle was then demonstrated by comparing the likelihood ratio values produced by the broadband coherence detector and generalized Hadamard ratio for several simulated targets of different geometrical shapes as a function of spatial translation. The results of these studies showed that the broadband coherence detector can withstand approximately 0.5 m of translation in either along-track or range before completely losing the ability to discriminate target from background. Another conclusion drawn from these results was that the use of larger block sizes with the broadband coherence detector makes it more robust to spatial translation compared to the smaller block sizes used with the generalized Hadamard ratio. This result is consistent with arguments made in the beginning of Section 6.4.2 which show that the separation among the null and alternative hypotheses tends to be more robust with the use of higher dimensional observations.

Finally, the performance of the broadband coherence detector on the two real sonar datasets was then compared to the generalized Hadamard ratio using both pixel and block-based observations as well as a matched subspace detector [54] designed to find regions in the magnitude of the HF image that exhibit the highlight and shadow characteristics typical of targets lying proud on the seafloor. A comparison of the results for each method shows that the broadband coherence, generalized Hadamard ratio, and matched subspace detectors all perform fairly similarly on the first real sonar dataset. However on the second dataset, the broadband coherence detector provides a 5% improvement in P_D along with a significant reduction in false alarm rate compared to its closest competitor, the matched subspace detector. The Hadamard ratio performed poorly on both datasets which may be attributed to a lack of robustness to even slight deviations from coregistration. Through the results presented in this dissertation, one can see that the fundamental principle of detecting underwater targets using coherence-based approaches is itself a very useful solution for this problem. Moreover, while not the true GLRT for WSS random processes, results show that the broadband coherence detector is a computationally efficient method which is adequately adept at achieving this.

7.2. Future Work

Although, the coherence-based techniques discussed in this dissertation offers powerful tools for detection of underwater targets from multiple sonar images, there are several important areas and extensions that can be pursued in the future. These include, but are not limited to:

- As discussed in Chapter 5, the broadband coherence detector given in (33) is not the true GLRT for WSS random processes in that it does not impose a Toeplitz structure on the ML estimate of the composite covariance matrix. Instead, the methods discussed in Chapter 5 all relied on asymptotic versions of the generalized Hadamard ratio as results concerning large Toeplitz matrices are very practical. Thus, one line of future work could involve the development of efficient means of computing ML estimates of Toeplitz matrices that would be valid even for finite dimensions. In addition to the problem considered here, the development of such methods would find its uses in a wide range of applications.
- One of the conclusions drawn in the investigation of the geometry of the GLRT in Section 2.3.2 was that the block-diagonal ML estimator $\hat{D}$ used in (6) is also a least-squares estimator in that it minimizes the Euclidean distance from the unconstrained ML estimator $\hat{R}$. Along similar lines of the previous item, one line of future research could be to investigate whether this principle extends beyond the ML problem considered here. That is, given a set of *iid* samples $\mathbf{x}_i \in \mathbb{C}^n$ for $i = 1, \ldots, M$ and some convex subspace $\mathcal{R}_0 \subset \mathcal{R}$ of the set of all PD Hermitian matrices $\mathcal{R}$, the goal of this work would be to establish necessary and sufficient conditions for the

set $\mathcal{R}_0$ such that the ML estimator $R^* \in \mathbb{C}^{n\times n}$ satisfies

$$R^* = \arg\max_{R\in\mathcal{R}_0} \prod_{i=1}^{M} f\left(\mathbf{x}_i; R\right) = \arg\min_{R\in\mathcal{R}_0} \left|\left|\hat{R} - R\right|\right|_F^2$$

where $f(\mathbf{x}_i; R)$ denotes the probability density of a zero-mean complex normal random vector with covariance matrix R and $\hat{R} = \frac{1}{M}\sum_{i=1}^{M} \mathbf{x}_i\mathbf{x}_i^H$. Such a result could not only find its uses in statistical hypothesis testing but in the ML estimation of structured covariance matrices in general.

- The data used in this study was limited to only a few environments and types of underwater targets. Ideally, the next step in the development of the coherence-based detection developed in this dissertation would be to test the performance on more data to prove the usefulness of the detection systems developed here. The testing on more difficult data sets as well as those including more man-made non-targets could be done in the future. More specifically, a study on the effect of different bottom types, target orientations, sonar aspect, resolution, and SNR on the performance of the detector would be insightful and help to illustrate the real effectiveness of these techniques for realistic underwater target detection problems.
- The coherence-based detection methods developed in this thesis are applicable not only to sonar image detection, but could be used on other disparate sensory systems, i.e. radar, infrared, and optical. A study of its usefulness on these types of sensing modalities would be highly valuable. By finding the coherent information between more than one type of sensor, the detection and classification performance could be improved.
- Another line of possible future work could be the investigation of not only using broadband coherence for detection but also using the coherence patterns themselves

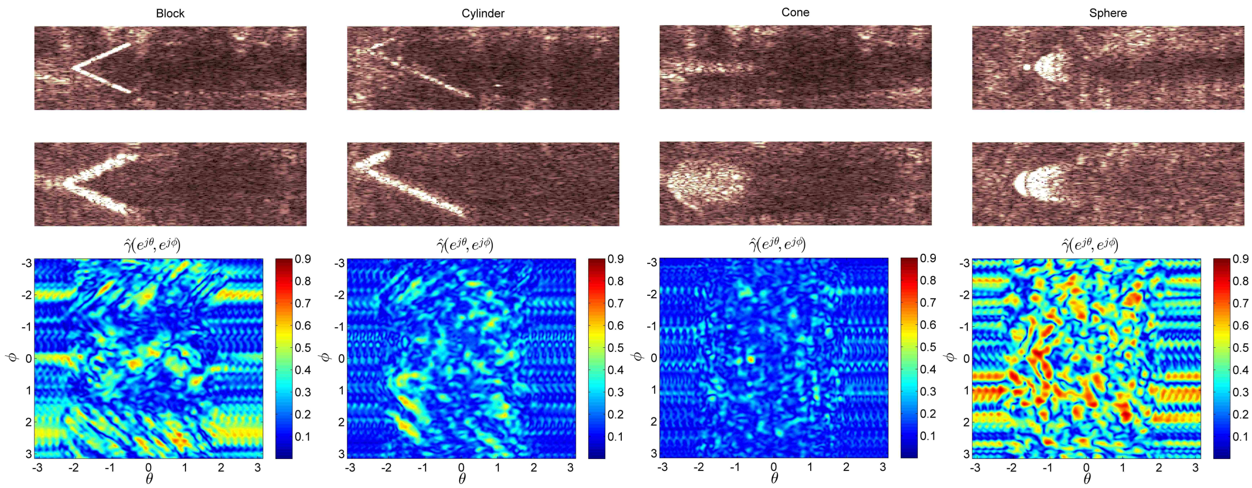


FIGURE 7.1. Several examples of the HF and BB snippets of different simulated targets and their coherence patterns.

as a classification feature for discriminating one target type from another. Figure 7.1 gives examples of four simulated targets from Dataset 3 used in Chapter 6 as well as the estimated magnitude coherence $\hat{\gamma}(e^{j\theta}, e^{j\phi})$ defined in (36) for each pair of HF and BB snippets. One can see from this figure that each target type, representing block, cylinder, cone, and sphere respectively, does in fact exhibit a distinct coherence pattern which may be used to discriminate one target type from the other. Work along this line would involve studies to gain an understanding of how different variables such as aspect angle and range effect the coherence patterns of each target type as well as an investigation of how robust the coherence patterns are to different data collection scenarios, e.g. proud versus partially buried. Work would also involve the development of methods that exploit these patterns for classification.

- One of the conclusions that may be drawn from the false alarm results presented in Chapter 6 is the difficulty in designing detectors that achieve a constant false alarm rate given the wide variation in environmental and seafloor conditions that can be encountered in this problem. Thus, one line of future work could involve the development of robust methods which adaptively adjust the threshold when new environments are encountered in the hopes of achieving a constant false alarm rate. Moreover, these methods might take advantage of environmental context or *a priori* knowledge about the environment in which data is being collected to determine an appropriate threshold.

- Beyond being a method well-suited to measuring coherence among multiple channels, one of the advantages of the broadband coherence detector in (33) is that it makes it very straightforward to be selective in terms of what frequencies are and are not used in the construction of the likelihood ratio. Thus, along the lines of the previous item, one possible area of future research could involve the development of frequency-selective detection techniques designed to reduce the effects of clutter in the computation of the likelihood ratio. Here, the objective would be to remove or suppress those frequencies typically associated with clutter while at the same time retaining as much of the coherent information for targets as possible.

Bibliography

[1] A. Nasipui and K. Li, "Multisensor collaboration in wireless sensor networks for detection of spatially correlated signals," *International Journal of Mobile Network Design and Innovation*, vol. 1, pp. 215–223, 2006.

[2] G. Wagner and T. Owens, "Signal detection using multi-channel seismic data," *Bulletin of the Seismological Society of America*, vol. 86, pp. 221–231, 1996.

[3] B. Liu, B. Chen, and J. Michels, "A GLRT for multichannel radar detection in the presence of both compound Gaussian clutter and additive white Gaussian noise," *Digital Signal Processing*, vol. 15, no. 5, pp. 437 – 454, 2005.

[4] A. Clausen and D. Cochran, "Non-parametric multiple channel detection in deep ocean noise," *Proceedings of the Thirty-First Asilomar Conference on Signals, Systems, and Computers*, pp. 850–854, November 1997.

[5] R. Viswanathan and P. Varshney, "Distributed detection with multiple sensors i. fundamentals," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 54–63, January 1997.

[6] D. Cochran, H. Gish, and D. Sinno, "A geometric approach to multiple-channel signal detection," *IEEE Transactions on Signal Processing*, vol. 43, pp. 2049–2057, September 1995.

[7] D. Ramirez, J. Via, I. Santamaria, and L. Scharf, "Detection of spatially correlated Gaussian time series," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5006–5015, 2010.

[8] N. Klausner and M. R. Azimi-Sadjadi, "Detection in multiple disparate systems using multi-channel coherence analysis," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, pp. 3554–3566, October 2012.

[9] J. C. Hyland and G. J. Dobeck, "Sea mine detection and classification using side-looking sonar," *Proc. SPIE*, vol. 2496, pp. 442–453, April 1995.

[10] G. J. Dobeck, J. Hyland, and L. Smedley, "Automated detection/classification of sea mines in sonar imagery," *Proc. SPIE*, vol. 3079, pp. 90–110, April 1997.

[11] V. Chandran, S. Elgar, and A. Nguyen, "Detection of mines in acoustic images using higher order spectral features," *IEEE Journal of Oceanic Engineering*, vol. 27, no. 3, pp. 610–618, July 2002.

[12] T. Aridgides and M. F. Fernandez, "Automated target classification in high resolution dual frequency sonar imagery," *Proc. SPIE*, vol. 6553, pp. 1–12, April 2007.

[13] T. Aridgides and M. Fernandez, "Enhanced ATR algorithm for high resolution multi-band sonar imagery," *Proc. SPIE*, vol. 6953, pp. 0–1, March 2008.

[14] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, pp. 321–377, 1936.

[15] A. Pezeshki, M. R. Azimi-Sadjadi, L. L. Scharf, and M. Robinson, "Underwater target classification using canonical correlations," *Proc. of MTS/IEEE Oceans 2003 Conference*, pp. 1906–1911, September 2003.

[16] A. Pezeshki, M. R. Azimi-Sadjadi, and L. L. Scharf, "Undersea target classification using canonical correlation analysis," *IEEE Journal of Oceanic Engineering*, vol. 32, no. 4, pp. 948–955, Oct. 2007.

[17] J. Tucker and M. Azimi-Sadjadi, "Coherence-based underwater target detection from multiple disparate sonar platforms," *Oceanic Engineering, IEEE Journal of*, vol. 36, no. 1, pp. 38–52, January 2011.

[18] J. R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, vol. 58, pp. 433–451, 1971.

[19] A. Leshem and A. J. Van der Veen, "Multichannel detection of Gaussian signals with uncalibrated receivers," *IEEE Signal Processing Letters*, vol. 8, no. 4, pp. 120–122, 2001.

[20] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory.* Prentice Hall, 1991.

[21] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis.* Addison-Wesley, 1991.

[22] R. Butler, *Saddlepoint Approximations with Applications.* Cambridge University Press, New York, 2007.

[23] R. M. Gray, "Toeplitz and circulant matrices: A review," Stanford University, Department of Electrical Engineering, Tech. Rep., March 2000.

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

[25] R. E. Kass and A. E. Raftery, "Bayes factors," *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 773–795, 1995.

[26] R. Muirhead, *Aspects of Multivariate Statistical Theory.* John Wiley and Sons, New Jersey, 2005.

[27] N. R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction)," *Annals of Mathematical Statistics*, vol. 34, no. 1, pp. 152–177, 1963.

[28] L. L. Scharf and J. K. Thomas, "Wiener filters in canonical coordinates for transform coding, filtering, and quantizing," *IEEE Transactions on Signal Processing*, vol. 46,

no. 3, pp. 647–654, 1998.

[29] A. Pezeshki, L. L. Scharf, J. K. Thomas, and B. D. Van Veen, "Canonical coordiantes are the right coordinates for low-rank Gauss-Gauss detection and estimation," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4817–4820, Dec 2006.

[30] T. R. Knapp, "Canonical correlation analysis: A general parametric significance testing system," *Psychological Bulletin*, vol. 85, no. 2, pp. 410–416, 1978.

[31] S. S. Wilks, "The large-sample distribution of the likelihood ratio for testing composite hypotheses," *The Annals of Mathematical Statistics*, vol. 9, no. 1, pp. 60–62, Mar. 1938.

[32] W. Cochran, "The distribution of quadratic forms in a normal system, with applications to the analysis of covariance," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 30, no. 2, pp. 178–191, 1934.

[33] N. Klausner, M. Azimi-Sadjadi, and L. Scharf, "Detection of spatially correlated time series from a network of sensor arrays," *Signal Processing, IEEE Transactions on*, vol. 62, no. 6, pp. 1396–1407, March 2014.

[34] N. Klausner, M. Azimi-Sadjadi, L. Scharf, and D. Cochran, "Space-time coherence and its exact null distribution," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, May 2013, pp. 3919–3923.

[35] A. Erdélyi, W. Magnus, F. Obergettinger, and F. Tricomi, *Higher Transcendental Functions*. McGraw-Hill, New York, 1953, vol. 1.

[36] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*. Dover Publications, New York, 1965.

[37] G. Box, "A general distribution theory for a class of likelihood criteria," *Biometrika*, vol. 36, no. 4, pp. 317–346, 1949.

[38] H. Daniels, "Saddlepoint approximations in statistics," *The Annals of Mathematical Statistics*, vol. 25, no. 4, pp. 631–650, Dec. 1954.

[39] C. Goutis and G. Casella, "Explaining the saddlepoint approximation," *The American Statistician*, vol. 53, no. 3, pp. 216–224, Aug. 1999.

[40] G. Grimmett and D. Stirzaker, *Probability and Random Processes*, 3rd ed. Oxford University Press, 2001.

[41] L. Davis, R. Evans, and E. Polak, "Maximum likelihood estimation of positive definite Hermitian Toeplitz matrices using outer approximations," in *Statistical Signal and Array Processing, 1998. Proceedings., Ninth IEEE SP Workshop on*, September 1998, pp. 49–52.

[42] J. Gutierrez-Gutierrez and P. Crespo, "Asymptotically equivalent sequences of matrices and Hermitian block Toeplitz matrices with continuous symbols: Applications to MIMO systems," *IEEE Transactions on Information Theory*, vol. 54, no. 12, pp. 5671–5680, 2008.

[43] R. Kozick and B. Sadler, "Source localization with distributed sensor arrays and partial spatial coherence," *Signal Processing, IEEE Transactions on*, vol. 52, no. 3, pp. 601–616, March 2004.

[44] N. Klausner, M. Azimi-Sadjadi, and L. Scharf, "Detection of correlated time series in a network of sensor arrays," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, May 2014, pp. 1–5.

[45] P. Brockwell and R. Davis, *Time Series: Theory and Methods*, 2nd ed. Springer, 2006.

[46] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, June 1967.

[47] S. K. Mitra, *Digital Signal Processing: A Computer-Based Approach*, Third ed. McGraw-Hill, 2006.

[48] D. Brown, D. Cook, and J. Fernandez, "Results from a small synthetic aperture sonar," in *OCEANS 2006*, September 2006, pp. 1–6.

[49] R. Bamler, "A comparison of range-Doppler and wavenumber domain SAR focusing algorithms," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 30, no. 4, pp. 706–713, 1992.

[50] M. Soumekh, *Fourier Array Imaging.* Prentice Hall, Englewood Cliffs, NJ, 1994.

[51] H. L. Van Trees, *Optimum Array Processing.* Wiley-Interscience, 2002.

[52] R. H. Stolt, "Migration by Fourier transform," *Geophysics*, vol. 43, no. 1, pp. 23–48, 1978.

[53] G. Sammelmann, "High-frequency images of proud and buried 3D-targets," in *OCEANS 2003*, vol. 1, Sept 2003, pp. 266–272.

[54] L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Transactions on Signal Processing*, vol. 42, no. 8, pp. 2146–2157, August 1994.

[55] N. R. Goodman, "The distribution of the determinant of a complex Wishart distributed matrix," *Annals of Mathematical Statistics*, vol. 34, no. 1, pp. 178–180, 1963.

APPENDIX A

Gram Determinants and Complex Wishart Matrices

In this appendix, we will briefly review the theory of Gram determinants and its connection to the distribution of the determinant of complex Wishart matrices [27]. The results of this chapter will be used in Sections 3.2 and 3.3 when discussing the null distribution of the generalized Hadamard ratio.

Consider the arbitrary set of vectors $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{N-1}$ with $\mathbf{x}_n \in \mathbb{C}^M$. This collection of vectors may be organized into the data matrix $X = [\mathbf{x}_0 \ \mathbf{x}_1 \ \cdots \ \mathbf{x}_{N-1}] \in \mathbb{C}^{M\times N}$. The Gram matrix for X is the $N \times N$ Hermitian matrix containing all pair-wise inner products for this set of vectors

$$G = X^H X = \begin{bmatrix} \langle \mathbf{x}_0, \mathbf{x}_0 \rangle & \langle \mathbf{x}_1, \mathbf{x}_0 \rangle & \cdots & \langle \mathbf{x}_{N-1}, \mathbf{x}_0 \rangle \\ \langle \mathbf{x}_0, \mathbf{x}_1 \rangle & \langle \mathbf{x}_1, \mathbf{x}_1 \rangle & \cdots & \langle \mathbf{x}_{N-1}, \mathbf{x}_1 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{x}_0, \mathbf{x}_{N-1} \rangle & \langle \mathbf{x}_1, \mathbf{x}_{N-1} \rangle & \cdots & \langle \mathbf{x}_{N-1}, \mathbf{x}_{N-1} \rangle \end{bmatrix}$$

where $\langle \mathbf{x}_i, \mathbf{x}_k \rangle = \mathbf{x}_k^H \mathbf{x}_i$. As discussed in [21], this set of vectors is linearly independent if and only if the real-valued scalar $\det G$, referred to as the Gram determinant, is nonzero.

The fact that a nonzero Gram determinant constitutes a necessary and sufficient condition for linear independence gives us a procedure for sequentially testing each of these N vectors to determine if $\mathbf{x}_n$ is linearly independent of $\mathbf{x}_0, \ldots, \mathbf{x}_{n-1}$ for any $n = 1, \ldots, N-1$. The data matrix $X_n = [X_{n-1} \ \mathbf{x}_n]$ with $X_{n-1} = [\mathbf{x}_0 \ \cdots \ \mathbf{x}_{n-1}]$ has the structured Gram matrix

$$G_n = X_n^H X_n = \begin{bmatrix} G_{n-1} & \mathbf{h}_n \\ \mathbf{h}_n^H & g_{nn} \end{bmatrix}$$

In this expression, $G_{n-1} = X_{n-1}^H X_{n-1}$ is the Gram matrix of X_{n-1}, $\mathbf{h}_n = X_{n-1}^H \mathbf{x}_n$ is a vector of inner products between $\mathbf{x}_i$ and $\mathbf{x}_n$ for $i = 0, \ldots, n-1$, and $g_{nn} = \mathbf{x}_n^H \mathbf{x}_n$ is the squared-norm of $\mathbf{x}_n$. Using results concerning the determinant of block matrices, one can show that the Gram determinant with $\mathbf{x}_n$ included is the Gram determinant without $\mathbf{x}_n$ multiplied by the *Schur complement* σ_n^2

$$\begin{aligned} \det G_n &= \sigma_n^2 \det G_{n-1} \qquad \text{(A-1)} \\ \sigma_n^2 &= g_{nn} - \mathbf{h}_n^H G_{n-1}^{-1} \mathbf{h}_n \end{aligned}$$

This recursive expression may be iterated to express the Gram determinant as a product of σ_n^2 for $n = 0, \ldots, N-1$

$$\det G = \left(\sigma_{N-1}^2\right) \det G_{(N-1)-1} = \left(\sigma_{N-1}^2\right) \left(\sigma_{N-2}^2\right) \det G_{(N-2)-1} = \cdots = \prod_{n=0}^{N-1} \sigma_n^2 \qquad \text{(A-2)}$$

with $\sigma_0^2 = g_{00} = \mathbf{x}_0^H \mathbf{x}_0$.

If the vector $\mathbf{x}_n$ is linearly independent of $\mathbf{x}_0, \ldots, \mathbf{x}_{n-1}$, it is clear from the expression given in (A-1) that $\det G_n$ will be nonzero if and only if σ_n^2 is nonzero. Recalling the definitions of G_{n-1}, $\mathbf{h}_n$, and g_{nn} given above, we can rewrite the scalar σ_n^2 as a sole function of X_{n-1} and $\mathbf{x}_n$

$$\begin{aligned} \sigma_n^2 &= \mathbf{x}_n^H \mathbf{x}_n - \mathbf{x}_n^H X_{n-1} \left(X_{n-1}^H X_{n-1}\right)^{-1} X_{n-1}^H \mathbf{x}_n \\ &= \mathbf{x}_n^H \left(I - X_{n-1} \left(X_{n-1}^H X_{n-1}\right)^{-1} X_{n-1}^H\right) \mathbf{x}_n \\ &= \mathbf{x}_n^H P_X^{\perp} \mathbf{x}_n = ||P_X^{\perp} \mathbf{x}_n||^2 \end{aligned}$$

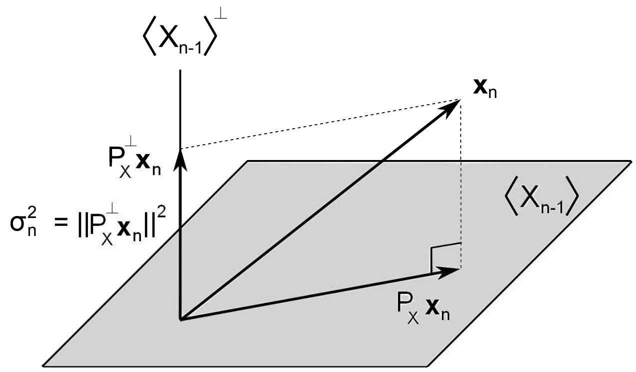


FIGURE A-1. The incremental Gram determinant σ_n^2 is the squared-length of the vector $\mathbf{x}_n$ projected onto the $M-n$ dimensional subspace $\langle X\rangle^\perp$.

where $P_X = X_{n-1}\left(X_{n-1}^H X_{n-1}\right)^{-1} X_{n-1}^H$ denotes the orthogonal projection onto the n dimensional subspace $\langle X_{n-1}\rangle$ spanned by the columns of matrix X_{n-1}. The geometry of this result is depicted in Figure A-1. Thus, one can see that the scalar σ_n^2 represents the squared-length of the vector $\mathbf{x}_n$ projected onto the orthogonal complement $\langle X_{n-1}\rangle^\perp$ and its connection with linear dependence is clear: if the vector $\mathbf{x}_n$ is linearly dependent on the columns of matrix X_{n-1} then there exists a vector $\mathbf{a} \in \mathbb{C}^n$ such that $\mathbf{x}_n = X_{n-1}\mathbf{a}$, the projection $P_X^\perp \mathbf{x}_n$ is the null vector

$$\begin{aligned}
P_X^\perp \mathbf{x}_n &= \left(I - X_{n-1}\left(X_{n-1}^H X_{n-1}\right)^{-1} X_{n-1}^H\right) X_{n-1}\mathbf{a} \\
&= X_{n-1}\mathbf{a} - X_{n-1}\left(X_{n-1}^H X_{n-1}\right)^{-1}\left(X_{n-1}^H X_{n-1}\right)\mathbf{a} \\
&= X_{n-1}\mathbf{a} - X_{n-1}\mathbf{a} = \mathbf{0},
\end{aligned}$$

i.e. $\mathbf{x}_n$ is in the null space of $P_X^\perp$, and hence $\sigma_n^2 = 0$.

Although the arguments given above are true for a general choice in the vector $\mathbf{x}_n$, let's consider the case where $\mathbf{x}_n = [x_n[1]\ x_n[2]\ \cdots\ x_n[M]]^H$ with $x_n[m] \overset{iid}{\sim} \mathcal{CN}(0,1)$ for $n = 0, \ldots, N-1$ and $m = 1, \ldots, M$. We will use this collection of vectors, $\{\mathbf{x}_n\}_{n=0}^{N-1}$, to build the rows of the random data matrix $X \in \mathbb{C}^{N \times M}$ with $M \geq N$

$$X = \begin{bmatrix} x_0[1] & x_0[2] & \cdots & x_0[M] \\ x_1[1] & x_1[2] & \cdots & x_1[M] \\ \vdots & \vdots & \ddots & \vdots \\ x_{N-1}[1] & x_{N-1}[2] & \cdots & x_{N-1}[M] \end{bmatrix}$$

One can think of the n^{th} row of this matrix as M independent samples of the standard complex normal random variable x_n. These random variables have the sample covariance matrix $\hat{R} = \frac{1}{M} X X^H$. The random matrix $M\hat{R} = XX^H$, which is a Gram matrix for the rows of X, is said to be distributed according to the complex Wishart distribution [27] with identity scaling matrix and M degrees of freedom, denoted $XX^H \sim \mathcal{CW}_N(I, M)$. Recalling the discussion on sequential Gram determinants given above, we may write the determinant of the Gram matrix as $\det\left(XX^H\right) = \prod_{n=0}^{N-1} \sigma_n^2$ where

$$\sigma_n^2 = \mathbf{x}_n^H \left(I - X_{n-1}^H (X_{n-1} X_{n-1}^H)^{-1} X_{n-1}\right) \mathbf{x}_n = \mathbf{x}_n^H P_X^{\perp} \mathbf{x}_n$$

is again the squared length of the projection onto the orthogonal complement of the space spanned by the columns of data matrix $X_{n-1}^H = [\mathbf{x}_0, \ldots, \mathbf{x}_{n-1}]$, i.e. the set of random variables $x_0, \ldots, x_{n-1}$.

Conditioned on the data matrix $X_{n-1} \in \mathbb{C}^{n \times M}$, the random scalar $y_n = 2\mathbf{x}_n^H P_X^{\perp} \mathbf{x}_n = 2\sigma_n^2$ is a quadratic involving standard complex normal random variables with a deterministic idempotent matrix, i.e. if X_{n-1} is known then $P_X = X_{n-1}^H (X_{n-1} X_{n-1}^H)^{-1} X_{n-1}$ is deterministic, so that $y_n \mid X_{n-1} \sim \chi^2_{2(M-n)}$. This conditional probability distribution is dependent on the number of rows (n) and columns (M) of X_{n-1} but in no way dependent on what value this data matrix actually takes making y_n statistically independent of X_{n-1}. As the sequence of preceding random variables $y_0, \ldots, y_{n-1}$ are all a function of X_{n-1}, this also implies pair-wise independence between y_n and $y_0, \ldots, y_{n-1}$. By induction on n, it then follows that the entire sequence of random variables $y_0, \ldots, y_{N-1}$ are mutually independent so that the scaled determinant of the complex Wishart distributed matrix XX^H is probabilistically equivalent to a product of independent chi-squared random variables

$$2^N \det\left(XX^H\right) \stackrel{d}{=} \prod_{n=0}^{N-1} \chi^2_{2(M-n)}$$

A more rigorous proof of this fact involving the characteristic function of the complex Wishart distribution can be found in [55]. Thus, in addition to giving one a method for sequentially determining if a set of vectors are linearly independent, Gram determinants are also a useful tool for analyzing the probabilistic behavior of determinants of sample covariance matrices.

APPENDIX B

Chi-Squared and Beta Random Variables

In this appendix, we establish a well-known relationship between chi-squared and beta random variables using the change of variables technique [40]. The purpose of this appendix is to prove the statement given in Section 3.3 that "if X and Y represent two independent chi-squared random variables with degrees of freedom ν_X and ν_Y, respectively, then the random variable $\frac{X}{X+Y}$ is distributed according to a beta distribution with parameters $\nu_X/2$ and $\nu_Y/2$."

Let $\mathbf{X} = [X_1 \ \cdots \ X_n]^T \in \mathcal{D} \subseteq \mathbb{R}^n$ be a random vector with joint PDF $f_{\mathbf{X}}(\mathbf{x})$ over the domain $\mathcal{D} = \{\mathbf{x} : f_{\mathbf{X}}(\mathbf{x}) > 0\}$. Let the transformation $T : \mathcal{D} \to \mathcal{R}$ denote some one-to-one, continuously differentiable function that maps the domain $\mathcal{D}$ to the range $\mathcal{R}$ and set $\mathbf{Y} = T(\mathbf{X})$. Then, it is well known that the joint PDF of the random vector $\mathbf{Y} = [Y_1 \ \cdots \ Y_n]^T$ can be related to that of the random vector $\mathbf{X}$ through the equality

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}\left[T^{-1}(\mathbf{y})\right] |\det J_{T^{-1}}| \, \mathbf{1}_{\mathcal{R}}(\mathbf{y})$$

where $T^{-1}(\cdot)$ denotes the inverse transformation, $J_{T^{-1}}$ is the Jacobian matrix of the inverse transformation

$$J_{T^{-1}} = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{bmatrix}$$

and $\mathbf{1}_{\mathcal{R}}(\mathbf{y})$ denotes the indicator function for the range $\mathcal{R}$

$$\mathbf{1}_{\mathcal{R}}(\mathbf{y}) = \begin{cases} 1 & \mathbf{y} \in \mathcal{R} \\ \\ 0 & \mathbf{y} \notin \mathcal{R} \end{cases}$$

Now, suppose that $\mathbf{X} = [X_1 \ X_2]^T$ with $X_1 \sim \chi^2_{\nu_1}$, $X_2 \sim \chi^2_{\nu_2}$ and X_1 is independent of X_2. This pair of random variables has joint PDF

$$f_{X_1X_2}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2) = \frac{1}{2^{\frac{\nu_1+\nu_2}{2}}\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} x_1^{\frac{\nu_1}{2}-1} x_2^{\frac{\nu_2}{2}-1} e^{-\frac{x_1+x_2}{2}} \ ; \ x_1, x_2 \geq 0$$

Define the transform/inverse-transform pair

$$\begin{aligned} T \ &: \ Y_1 = X_1 + X_2; \ Y_2 = \frac{X_1}{X_1 + X_2} \\ T^{-1} \ &: \ X_1 = Y_1Y_2; \ X_2 = Y_1(1 - Y_2) \end{aligned}$$

with range $\mathcal{R} = (0, \infty) \times (0, 1)$ and Jacobian matrix

$$\begin{aligned} J_{T^{-1}} &= \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{bmatrix} = \begin{bmatrix} y_2 & y_1 \\ \\ 1 - y_2 & -y_1 \end{bmatrix} \\ |\det J_{T^{-1}}| &= |-y_1y_2 - y_1(1 - y_2)| = y_1 \end{aligned}$$

With this, the pair of random variables Y_1 and Y_2 has joint PDF

$$
\begin{aligned}
f_{Y_1Y_2}(y_1,y_2) &= f_{X_1X_2}\left(y_1y_2, y_1(1-y_2)\right) y_1 \mathbf{1}_{(0,\infty)}(y_1)\mathbf{1}_{(0,1)}(y_2) \\
&= \left(\frac{1}{2^{\frac{\nu_1+\nu_2}{2}}\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} (y_1y_2)^{\frac{\nu_1}{2}-1} \left[y_1(1-y_2)\right]^{\frac{\nu_2}{2}-1} e^{-\frac{y_1y_2+y_1(1-y_2)}{2}} \right) y_1 \mathbf{1}_{(0,\infty)}(y_1)\mathbf{1}_{(0,1)}(y_2) \\
&= \left(\frac{1}{2^{\frac{\nu_1+\nu_2}{2}}} y_1^{\frac{\nu_1+\nu_2}{2}-1} e^{-\frac{y_1}{2}} \mathbf{1}_{(0,\infty)}(y_1) \right) \left(\frac{1}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})} y_2^{\frac{\nu_1}{2}-1}(1-y_2)^{\frac{\nu_2}{2}-1} \mathbf{1}_{(0,1)}(y_2) \right)
\end{aligned}
$$

Multiplying and dividing by $\Gamma(\frac{\nu_1+\nu_2}{2})$ and using the relationship given in (13), this joint PDF can finally be written

$$
\begin{aligned}
f_{Y_1Y_2}(y_1,y_2) &= f_{Y_1}(y_1)f_{Y_2}(y_2) \\
&= \left(\frac{1}{2^{\frac{\nu_1+\nu_2}{2}}\Gamma(\frac{\nu_1+\nu_2}{2})} y_1^{\frac{\nu_1+\nu_2}{2}-1} e^{-\frac{y_1}{2}} \mathbf{1}_{(0,\infty)}(y_1) \right) \left(\frac{1}{B(\frac{\nu_1}{2},\frac{\nu_2}{2})} y_2^{\frac{\nu_1}{2}-1}(1-y_2)^{\frac{\nu_2}{2}-1} \mathbf{1}_{(0,1)}(y_2) \right)
\end{aligned}
$$

From this expression for the joint density $f_{Y_1Y_2}(y_1,y_2)$, three consequences become self-evident:

(1) If $X_1 \sim \chi^2_{\nu_1}$, $X_2 \sim \chi^2_{\nu_2}$, and $X_1 \perp\!\!\!\perp X_2$ then the random variable X_1+X_2 is also chi-squared with degrees of freedom $\nu_1+\nu_2$ as evidenced by the expression for $f_{Y_1}(y_1)$.

(2) If $X_1 \sim \chi^2_{\nu_1}$, $X_2 \sim \chi^2_{\nu_2}$, and $X_1 \perp\!\!\!\perp X_2$ then the random variable $\frac{X_1}{X_1+X_2}$ is beta distributed with parameters $\nu_1/2$ and $\nu_2/2$ as evidenced by the expression for $f_{Y_2}(y_2)$.

(3) If $X_1 \sim \chi^2_{\nu_1}$, $X_2 \sim \chi^2_{\nu_2}$, and $X_1 \perp\!\!\!\perp X_2$ then the random variables X_1+X_2 and $\frac{X_1}{X_1+X_2}$ are independent as evidenced by the fact that $f_{Y_1Y_2}(y_1,y_2) = f_{Y_1}(y_1)f_{Y_2}(y_2)$.

APPENDIX C

Asymptotic Characteristic Function of the Generalized Hadamard Ratio

In this appendix, we discuss in more detail the process of going from the cumulant generating function of the random variable $Z = -2\rho M \ln \Lambda$ given in (16) to the asymptotic expression given in (18) as discussed in Section 3.4. Recall from the discussion in Chapter 3 that, as a consequence of knowing that the null distribution of the generalized Hadamard ratio is stochastically equivalent to a product of independent beta random variables, the cumulant generating function of the random variable Z given in (16) may be written as a sum involving log-gamma functions with various arguments

$$\psi_Z(jt) = \sum_{i=2}^{L}\sum_{n=0}^{N-1} S_{in}^{(1)} - S_{in}^{(2)} + S_{in}^{(3)} - S_{in}^{(4)} \qquad \text{(C-1)}$$

$$S_{in}^{(1)} = \ln\Gamma\left(\rho M + \xi_n\right) \qquad \text{(C-2)}$$

$$S_{in}^{(2)} = \ln\Gamma\left(\rho M + \xi_n - (i-1)N\right) \qquad \text{(C-3)}$$

$$S_{in}^{(3)} = \ln\Gamma\left(\rho M(1-2jt) + \xi_n - (i-1)N\right) \qquad \text{(C-4)}$$

$$S_{in}^{(4)} = \ln\Gamma\left(\rho M(1-2jt) + \xi_n\right) \qquad \text{(C-5)}$$

where $\xi_n = (1-\rho)M - n$. Also recall that the log-gamma function exhibits the asymptotic expansion given in (17) which, using the 2^{nd} order Bernoulli polynomial $B_2(x) = x^2 - x + \frac{1}{6}$, can be written

$$\ln\Gamma(z+a) = \frac{1}{2}\ln(2\pi) + (z + a - \frac{1}{2})\ln z - z + \frac{1}{2z}\left(a^2 - a + \frac{1}{6}\right) + \mathcal{O}(|z|^{-2})$$

Letting the variables z and a in this expansion represent the following terms for each log-gamma function given in (C-2) – (C-5)

$$S_{in}^{(1)} \;:\; z = \rho M,\; a = \xi_n$$

$$S_{in}^{(2)} \;:\; z = \rho M,\; a = \xi_n - (i-1)N$$

$$S_{in}^{(3)} \;:\; z = \rho M(1-2jt),\; a = \xi_n - (i-1)N$$

$$S_{in}^{(4)} \;:\; z = \rho M(1-2jt),\; a = \xi_n$$

and ignoring higher order terms, each term $S_{in}^{(k)}$ for $k = 1, \ldots, 4$ in (C-2) – (C-5) may be replaced with its respective approximation $\tilde{S}_{in}^{(k)}$

$$\begin{aligned}
\tilde{S}_{in}^{(1)} &= \frac{1}{2}\ln(2\pi) + \left(\rho M + \xi_n - \frac{1}{2}\right)\ln(\rho M) - \rho M + \frac{1}{2}\left(\xi_n^2 - \xi_n + \frac{1}{6}\right)(\rho M)^{-1} \\
\tilde{S}_{in}^{(2)} &= \frac{1}{2}\ln(2\pi) + \left(\rho M + \xi_n - (i-1)N - \frac{1}{2}\right)\ln(\rho M) - \rho M \\
&\quad + \frac{1}{2}\left(\xi_n^2 - 2(i-1)N\xi_n + (i-1)^2N^2 - \xi_n + (i-1)N + \frac{1}{6}\right)(\rho M)^{-1} \\
\tilde{S}_{in}^{(3)} &= \frac{1}{2}\ln(2\pi) + \left(\rho M(1-2jt) + \xi_n - (i-1)N - \frac{1}{2}\right)\ln\left[\rho M(1-2jt)\right] - \rho M(1-2jt) \\
&\quad + \frac{1}{2}\left(\xi_n^2 - 2(i-1)N\xi_n + (i-1)^2N^2 - \xi_n + (i-1)N + \frac{1}{6}\right)\left[\rho M(1-2jt)\right]^{-1} \\
\tilde{S}_{in}^{(4)} &= \frac{1}{2}\ln(2\pi) + \left(\rho M(1-2jt) + \xi_n - \frac{1}{2}\right)\ln\left[\rho M(1-2jt)\right] - \rho M(1-2jt) \\
&\quad + \frac{1}{2}\left(\xi_n^2 - \xi_n + \frac{1}{6}\right)\left[\rho M(1-2jt)\right]^{-1}
\end{aligned}$$

These expressions are very unwieldy but, upon adding and subtracting them together, many terms cancel yielding the fairly simple expression

$$\begin{aligned}\sum_{k=1}^{4}(-1)^{k+1}\tilde{S}_{in}^{(k)} &= -(i-1)N\ln(1-2jt) \\ &\quad +\frac{1}{2\rho M}\left[(i-1)^2N^2+(1-2\xi_n)(i-1)N\right]\left[(1-2jt)^{-1}-1\right] \quad \text{(C-6)}\end{aligned}$$

To find the approximation of the cumulant generating function in (C-1), we must then sum the expression given in (C-6) for all $i = 2, \ldots, L$ and $n = 0, \ldots, N-1$. As this requires sums of various powers of integers, one can employ the two series

$$\begin{aligned}\sum_{k=1}^{m} k &= \frac{m(m+1)}{2} \\ \sum_{k=1}^{m} k^2 &= \frac{m(m+1)(2m+1)}{6}\end{aligned}$$

to derive the following sums involving the number of channels (L), the length of each time series (N), the number of samples (M), and the scaling factor (ρ)

$$\begin{aligned}\sum_{i=2}^{L}\sum_{n=0}^{N-1}(i-1)N &= N^2\sum_{i=1}^{L-1} i = \frac{1}{2}(L^2-L)N^2 \\ \sum_{i=2}^{L}\sum_{n=0}^{N-1}(i-1)^2N^2 &= N^3\sum_{i=1}^{L-1} i^2 = \frac{1}{3}\left(L^3-\frac{3}{2}L^2+\frac{1}{2}L\right)N^3 \\ 2\sum_{i=2}^{L}\sum_{n=0}^{N-1}(i-1)N\xi_n &= 2(1-\rho)M\sum_{i=2}^{L}\sum_{n=0}^{N-1}(i-1)N - 2\sum_{i=2}^{L}\sum_{n=0}^{N-1}(i-1)Nn \\ &= (1-\rho)M(L^2-L)N^2-(N^3-N^2)\sum_{i=1}^{L-1} i \\ &= (1-\rho)M(L^2-L)N^2-\frac{1}{2}(L^2-L)(N^3-N^2)\end{aligned}$$

With these three results, the sum of the latter term in (C-6) becomes

$$
\begin{aligned}
\sum_{i=2}^{L}\sum_{n=0}^{N-1}\left[(i-1)^2N^2+(1-2\xi_n)(i-1)N\right] &= \frac{1}{3}\left(L^3-\frac{3}{2}L^2+\frac{1}{2}L\right)N^3+\frac{1}{2}(L^2-L)N^2 \\
&\quad -(1-\rho)M(L^2-L)N^2+\frac{1}{2}(L^2-L)(N^3-N^2) \\
&= -(1-\rho)M(L^2-L)N^2+\frac{1}{6}L(L^2-1)N^3
\end{aligned}
$$

which can be used to finally obtain an asymptotic expression of the cumulant generating function $\psi_Z(jt)$

$$
\begin{aligned}
\psi_Z(jt) &= \sum_{i=2}^{L}\sum_{n=0}^{N-1}\tilde{S}_{in}^{(1)}-\tilde{S}_{in}^{(2)}+\tilde{S}_{in}^{(3)}-\tilde{S}_{in}^{(4)}+\mathcal{O}(M^{-2}) \\
&= -\ln(1-2jt)\sum_{i=2}^{L}\sum_{n=0}^{N-1}(i-1)N \\
&\quad +\frac{1}{2\rho M}\left[(1-2jt)^{-1}-1\right]\sum_{i=2}^{L}\sum_{n=0}^{N-1}\left[(i-1)^2N^2+(1-2\xi_n)(i-1)N\right]+\mathcal{O}(M^{-2}) \\
&= -\frac{1}{2}(L^2-L)N^2\ln(1-2jt) \\
&\quad +\frac{1}{2\rho M}\left(-(1-\rho)M(L^2-L)N^2+\frac{1}{6}L(L^2-1)N^3\right)\left[(1-2jt)^{-1}-1\right]+\mathcal{O}(M^{-2})
\end{aligned}
$$

Defining the two expressions for ν and $\omega_1(\rho)$

$$
\begin{aligned}
\nu &= L^2N^2-LN^2 \\
\omega_1(\rho) &= \frac{1}{2\rho}\left(-\nu(1-\rho)+\frac{L(L^2-1)N^3}{6M}\right),
\end{aligned}
$$

one obtains the asymptotic expression of the cumulant generating function as it appears in (18)

$$
\psi_Z(jt) = -\frac{\nu}{2}\ln(1-2jt)+\omega_1(\rho)\left[(1-2jt)^{-1}-1\right]+\mathcal{O}(M^{-2})
$$