THESIS

UNDERWATER TARGET DETECTION USING MULTIPLE DISPARATE SONAR PLATFORMS

Submitted by

Nicholas Harold Klausner

Department of Electrical and Computer Engineering

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2010

Master's Committee:

Department Chair: Anthony A. Maciejewski

Advisor: Mahmood R. Azimi-Sadjadi

Ali Pezeshki
Dan Cooley

ABSTRACT

UNDERWATER TARGET DETECTION USING MULTIPLE DISPARATE SONAR PLATFORMS

The detection of underwater objects from sonar imagery presents 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. Additionally, collecting data from multiple platforms can present more challenging questions such as "how should I collaboratively perform detection to achieve optimal performance?","how many platforms should be employed?", "when do we reach a point of diminishing return when adding platforms?", or more importantly "when does adding an additional platform not help at all?". To perform multi-platform detection and answer these questions we use the *coherent* information among all disparate sources of information and perform detection on the premise that the amount of coherent information will be greater in situations where a target is present in a region of interest within an image versus a situation where our observation strictly consists of background clutter.

To exploit the coherent information among the different sources, we recast the standard Neyman-Pearson, Gauss-Gauss detector into the Multi-Channel Coherence Analysis (MCA) framework. The MCA framework allows one to optimally decompose the multi-channel data into a new appropriate coordinate system in order to analyze

their linear dependence or coherence in a more meaningful fashion. To do this, new expressions for the log-likelihood ratio and J-divergence are formulated in this multi-channel coordinate system. Using the MCA framework, the data of each channel is first whitened individually, hence removing the second-order information from each channel. Then, a set of linear mapping matrices are obtained which maximizes the sum of the cross-correlations among the channels in the mapped domain. To perform detection in the coordinate system provided by MCA, we first of all construct a model suited to this multiple sensor platform problem and subsequently represent observations in their MCA coordinates associated with the H_1 hypothesis. Performing detection in the MCA framework results in a log-likelihood ratio that is written in terms of the MCA correlations and mapping vectors as well as a local signal-to-noise ratio matrix. In this coordinate system, the J-divergence, which is a measure of the difference in means of the likelihood ratio, can effectively be represented in terms of the multi-channel correlations and mapping vectors. Using this J-divergence expression, one can get a more clear picture of the amount of discriminatory information available for detection by analyzing the amount of coherent information present among the channels.

New analytical and experimental results are also presented to provide better insight on the effects of adding a new piece of data to the multi-channel Gauss-Gauss detector represented in the MCA framework. To answer questions like those posed in the first paragraph, one must carefully analyze the amount of discriminatory information that is brought to the detection process when adding observations from an additional channel. Rather than attempting to observe the increase (or lack thereof) from the full detection problem it is advantageous to look at the change incrementally. To accomplish this goal, new updating equations for the likelihood ratio are derived that involve linearly estimating the new data from the old (already existing) and updating the likelihood ratio accordingly. In this case, the change in J-divergence

can be written in terms of error covariance matrices under each hypothesis. We then derive a change in coordinate system that can be used to perform dimensionality reduction. This especially becomes useful when the data we wish to add exists in high-dimensional space. To demonstrate the usefulness of log-likelihood updating, we conduct two simulation studies. The first simulation corresponds to detecting the presence of dynamical structure in data we have observed and corresponds to a *temporal* updating scheme. The second is concerned with detecting the presence of a single narrow-band source using multiple linear sensor arrays in which case we consider a *platform* (or *channel*) updating scheme.

A comprehensive study is carried out on the MCA-based detector on three data sets acquired from the Naval Surface Warfare Center (NSWC) in Panama City, FL. The first data set consists of one high frequency (HF) and three broadband (BB) side-looking sonar imagery coregistered over the same region on the sea floor captured from an Autonomous Underwater Vehicle (AUV) platform. For this data set we consider three different detection schemes using different combinations of these sonar channels. The next data set consists of one HF and only one BB beamformed sonar imagery again coregistered over the same region on the sea floor. This data set consists of not only target objects but also lobster traps giving us experimental intuition as how the multi-channel correlations change for different object compositions. The use of multiple disparate sonar images, e.g., a high frequency, high resolution sonar with good target definition and a multitude of lower resolution broadband sonar with good clutter suppression ability significantly improves the detection and false alarm rates comparing to situations where only single sonar is utilized. Finally, a data set consisting of synthetically generated images of targets with differing degrees of disparity such as signal-to-noise ratio (SNR), aspect angle, resolution, etc., is used to conduct a thorough sensitivity analysis in order to study the effects of different SNR, target types, and disparateness in aspect angle.

ACKNOWLEDGEMENTS

I would first like to thank my adviser, Dr. Mahmood Azimi, for his invaluable support and guidance throughout the course of this research. He has taught me the value of developing the intellectual reasoning needed to conduct thorough and well-developed research. His guidance and time is greatly appreciated throughout the course of my graduate education.

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

I would also like to thank Dr. Louis Scharf for his useful suggestions throughout the course of this research.

I would like to thank the Office of Naval Research and the Naval Surface Warfare Center-Panama City, Florida for providing the funding and the data used in this project. This project was funded by the Office of Naval Research (ONR-3210E) under contract number N00014-08-1-0142. Without the funding from ONR and the data from NSWC, this work would have never been completed.

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, Mike M., Mike K., Yinghui, Gergely, Katalin, Sonia, and Tomas. Special thanks to Derek Tucker for his much needed help and advice.

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

To my parents, Joe and Kathy, and my brother, Chris.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Problem Statement and Motivations

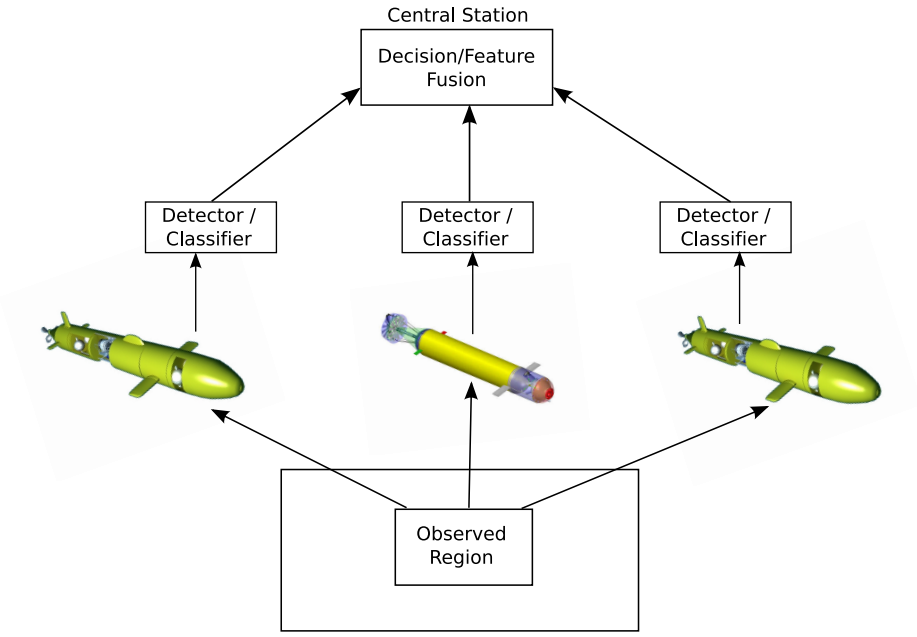
The problem of underwater object detection and classification in sonar imagery has recently attracted a substantial amount of attention [1]. This problem is complicated due to various factors such as variations in operating and environmental conditions, presence of spatially varying clutter, variations 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, a robust detection system should be able to quantify changes between the returns from the bottom and any target activity in sonar images, while at the same time extract useful features for subsequent classification. Thus, a system designed without the need to perform separate detection and feature extraction is highly desirable.

Normally, a single sensor (sonar, lidar, etc.) is used to detect and classify the objects based upon observations taken from the environment. From these observations, the sensor will either make a local decision and transmit it to a central station or record the entire sonar image for post mission analysis (PMA) at the central station. The issues faced with detection based upon one sensor is that the detection process is limited to only one field of view. This makes the detection of weak targets particularly challenging. Moreover, the structure of targets within an image vary as a function of aspect, grazing angle, and range from the sonar which makes detection difficult, especially if the target is in a disadvantaged position in relationship to the sensor, e.g. partially obscured targets. Therefore, any improvement in the detection results becomes hindered due the limited amount of data and observations from the

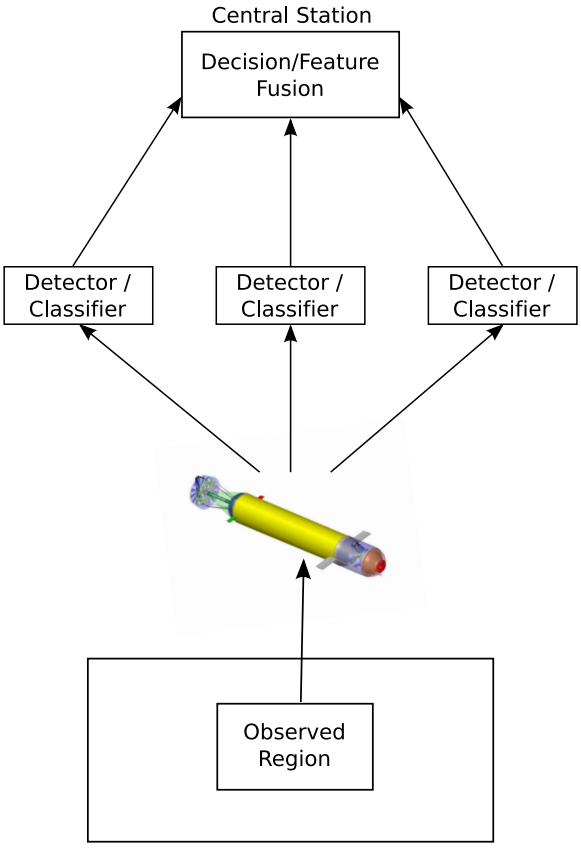
environment.

Distributed sensor networks offer a solution to overcome the shortcomings of the single sensor situations. The use of multiple sensors allows for significantly better capture of the target characteristics due to the fact that the targets are viewed from different aspects, grazing angles, ranges, frequencies, sensor modalities, etc. In a surveillance area there could be multiple autonomous underwater vehicles (AUV's) each equipped with a wide variety of sensors including different types of sonar, magnetics, or electro-optical systems (see Figure 1.1(a)) or a single AUV equipped with multiple sensors (see Figure 1.1(b)). Preliminary detection, feature extraction, and object classification can be performed based upon the data collected using every sensory system on one or multiple vehicles. A final decision-making usually takes places at the central station, either in the standard PMA method or real-time network-centric sensor analysis (NSA) using some type of decision-level or feature-level fusion. However, decision-making based upon individual sensory data typically leads to incomplete, degraded or biased local (sensor-level) decisions hence resulting in an unacceptable final detection and classification performance at the fusion center as the coherent information shared among the sensory systems is ignored.

To allow collaborative decision-making among multiple sonar platforms, it is essential to detect and further scrutinize the information bearing parts of the data collected by the various sensory systems. This involves detecting, isolating, and representing; in terms of some pertinent attributes, the *coherent*, or mutual information among one or multiple data sets. This is an extremely challenging problem due the disparate nature of the problem and therefore to develop such a solution, new methodologies are needed to: (a) collaboratively detect and agree on threats occurring within the field of view of the sensors, (b) perform feature extraction to capture common target attributes from multiple sensor platforms, (c) perform object classification and identification, (d) and finally develop a single integrated target assessment picture based



(a) Multi-Platform, Single-Sensory Fusion



(b) Single-Platform, Multi-Sensory Fusion

Figure 1.1: Two Different Paradigms for Distributed Decision-Making.

upon the detected, localized and classified targets from one or multiple disparate sensors.

Performing detection with multiple sources of information from any number of sensory systems should undoubtedly increase our confidence in the detection decisions we have made but at the same time opens up a new line of questions. Detection performance can vary depending on disparateness in sensor platform location, target composition, frequency, resolution, clutter density, aspect angle, etc. Thus, it becomes pertinent to analyze the behavior of the detection system to these variables of disparity. An equally important factor in any multiple sensor platform problem is the optimal number of sensing systems used to perform detection. Naturally, the number of sensor platforms used in any practical application directly relates to cost, computational, as well as bandwidth and latency constraints particularly in network-centric operations. Therefore, it becomes pertinent to analyze all such trade-offs when adding an additional sensory system into the decision-making process. Measuring the performance of the detection system with respect to variability in disparateness and the number of sensor platforms should obviously be stated in terms of probability of detection but a sufficient characterization of the distribution of the test statistic can be difficult in some situations. Alternatively, one can look at any number of information measures to get a good picture of the sensitivity of the performance of the detector, one of them being the J-divergence which simply gives a first-order characterization of the distribution of the likelihood ratio under both hypotheses. With this, one can hope to gain insight into what effects disparity and the number of platforms can have on the multiple platform detection problem in different situations.

In this work we design a coherence-based detector based on the Multi-Channel Coherence Analysis (MCA) framework [2] - [6] where the coherent information is found among multiple sensory channels and is used to detect objects from sonar imagery captured from one or multiple sonar systems with a high degree of confidence.

Additionally, we take a closer look at what effect adding an additional piece of data has on the detection statistics and find update equations which can be used to describe the increase in detection performance solely attributed to the new data.

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 a detailed review of such work the reader is referred to [1]. However, recently 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 [7], [8], 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. This type of decision-making, which is based upon individual sensory data, typically leads to incomplete, degraded or biased decisions. However, when the information from the individual sensors is used collaboratively and simultaneously in the decision-making a more reliable decision about the observation can be made.

Multi-channel detection has recently been looked at in [9] - [11]. In [9], a non-parametric (in the sense that no assumptions on the distribution of the data are made) approach to multi-channel detection is proposed by defining the *generalized coherence* (GC) estimate which measures the linear relationships among N channels by forming an N-dimensional normalized Grammian matrix from the N vectors of measurements. This measure can then be used to test whether the multi-channel data contains independent realizations of noise under H_0 versus its complement under H_1. Since no explicit *a priori* assumptions are made about the H_1 hypothesis, the estimate applies to a wide range of different signal models. Then making the explicit assumption that the measurements from all N channels are white Gaussian random vectors, the H_0 distribution of the *Magnitude-Squared Coherence* (MSC) estimate (the two-channel version of the GC estimate) is derived which can effectively be used to find thresholds corresponding to a constant false alarm rate. Under the same assumptions the authors then derive the distribution of the three-channel GC estimate under H_0 which leads to a recursive formulation of the distribution as one adds an additional channel thus in some sense generalizing the distribution to any number of channels. The authors note, however, that the GC estimate ignores temporally correlated random processes and so one drawback to the detector is the fact that propagation delays and doppler shifts must be estimated and accounted for prior to detection. Multi-channel detection for uncalibrated sensor elements in an array is considered in [10] by forming a *Generalized Likelihood Ratio Test* (GLRT). Based on the assumption that observations from the array are zero-mean, complex Gaussian distributed random vectors, the authors form a sample covariance matrix estimated from a window's worth of data. The GLRT then involves testing whether the sample covariance matrix has diagonal structure under H_0 versus any arbitrary, positive-definite (PD) covariance structure under H_1. Based on this detection problem, the GLRT test statistic becomes a Hadamard ratio [12] involving the sample covariance

matrix. Because the diagonal noise structure is not assumed to be known, the authors note that the GLRT presents an advantageous method for performing detection when sensor calibration data is absent. This work is then somewhat generalized in [11] by considering a problem where one is given multiple independent copies of a finite-length, vector-valued time series. Vectorizing each copy into a larger dimensional, spatio-temporal random vector and forming a data matrix from all the independent copies of this vector, the sample covariance matrix is estimated across the independent copies. The GLRT then involves testing whether the sample covariance matrix is block-diagonal under H_0 versus any arbitrary, positive-definite (PD) covariance structure under H_1. With these assumptions, the GLRT test statistic is written in terms of a generalized Hadamard ratio involving the sample covariance matrices or the determinant of the estimated coherence matrix. Making the asymptotic assumption that both the number of temporal measurements and the number of independent copies grow large, the test statistic is then written in terms of the log of the Hadamard ratio involving the estimated power spectral density matrices for each channel, integrated over the Nyquist band. So even though the analysis begins as a finite-dimensional result, the authors extend the statistic into infinite dimensions and note that it is equivalent to integrating the "power cepstrum" to form the test statistic. Properties of the proposed detection statistic are given, its relationship to mutual information shown, and a low-correlation regime approximation is then developed.

Recently, Canonical Correlation Analysis (CCA) [13] - [15] has been looked at for coherent detection and feature extraction. The canonical coordinate decomposition method not only determines linear dependence [13] 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 [16] - [19]. 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 [17] were shown to have high discriminatory power for both detection and classification. Another study, [20] extended this coherence analysis to the frequency domain by measuring coherence between the same frequency band in two sonar pings. Comparing to the time domain coherence-based features, these features provided substantially better results on the BOSS data sets as well as the ability to offer a more rigorous way of generating acoustic-color for possible target identification from multiple sonar pings. Previous to the work of this thesis [21], [22] - [24], CCA was utilized to form 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 [1] and is shown to exhibit high performance with a low false alarm rate and high probability of detection.

Multi-Channel Coherence Analysis (MCA) [2] - [5] can be seen as a natural extension of CCA to more than two channels. MCA-based decomposition similarly looks for sets of one-dimensional mappings that maximize the sum of the cross-correlations among any number of channels. In [3], [6] an iterative procedure for performing the MCA decomposition is proposed and applied to data from the BOSS system as a feature extraction method for classification with the three channels formed from three sonar returns with different ping separations. In [4], MCA is applied to Landsat imagery to quantify the amount of coherent information from multiple spectral bands and across different images in time. Finally, MCA is cast as a non-linear dynamical system in [5] and several cost functions for solving the problem are proposed. The

non-linear dynamical system associated with each cost function is then shown to be stable via Lyapunov functions and are shown to converge to the true value as time grows large.

1.3 Research Objectives

The goal of this present work is to develop and test an efficient and robust coherence-based detection system for multi-platform sonar imagery that maintains good underwater target detection performance in varying operating and environmental conditions. Specifically, we would like to develop and test a detection system that can be applied to multiple disparate sonar systems using the data provided by the Naval Surface Warfare Center, Panama City (NSWC-PC). For this system, our detection hypothesis is that the presence of a target in coregistered ROIs from multiple sonar images will exhibit higher level of coherence than when those same ROIs contained only background. The MCA method [2] - [5] provides an excellent framework for quantifying changes between the returns from the bottom and the returns from a target by determining the linear dependence (or coherence) between multiple data channels. It is also the objective of this analysis to study the set of features that naturally arise from MCA, namely the sum of correlations, and analyze their properties and suitability for target classification.

Based on the detection hypothesis just briefly described above, we construct a simple multi-channel detection problem which is then solved via MCA and cast into the standard Gauss-Gauss detection framework. The work presented here will develop the log-likelihood ratio and J-divergence for the composite multi-channel hypothesis test in the MCA framework and explore the link to the standard Gauss-Gauss detector presented in [25]. Test results on three different data sets are presented. The first data set contains one high frequency (HF) and three broadband (BB) sonar imagery coregistered over the sea floor. This data set is used to examine different combinations

of these sonar images, construct three different multi-channel detectors with the HF image and one more BB images, and compare detection performance as a function of the number of disparate sonar images used. The second data set consists of only one HF and one BB sonar imagery, again coregistered over the sea floor. This data set contains both target and non-target objects (lobster traps) which can give us some intuition of the discriminatory power of the extracted multi-channel features for target/non-target classification. Finally, we present sensitivity analysis results on a data set consisting of simulated target objects in simulated background generated for various choices of SNR, target aspect angle, image resolution, target shape, etc. The hope of the sensitivity analysis is to give one some idea of the proposed system's response to different variables that can be expected in multiple disparate platform detection problems.

Finally, we take a closer look at how adding an additional channel's worth of data effects the multi-channel Gauss-Gauss detector in terms of the increase in J-divergence as a consequence of this (time or channel) augmentation. In this case, we find that adding additional data to the detection problem involves linearly estimating the new data from the old and updating the old likelihood ratio through simple addition. The change in J-divergence can then be written in terms of error covariance matrices when filtering with a smoother that is matched/miss-matched to the given hypothesis. The change in J-divergence can then be used to decide whether or not to add that channel's worth of data or, when given the opportunity, to choose among several channels and decide which one to add. If it is decided to add that portion of the data to the detection problem, then a similar argument as in [25] is taken to find a coordinate system where error vectors can be approximated in a lower dimensional space. This coordinate system becomes especially useful in applications where the data to be added is high dimensional which can result in high computation savings without a significant decrease in the change in J-divergence. To demonstrate situations where

updating in this fashion may be useful we provide two simulations. The first example is concerned with detecting the presence of dynamical structure in our observation and adding data amounts to *time updating* the likelihood ratio. Because of the special structure of the problem, this time updating scheme uses a Kalman filter [26] to propagate the necessary error and error variances required for updating. The second example is concerned with detecting the presence of a single narrow-band source in the sensing environment of multiple uniform linear arrays (ULAs). A rank-one detector is then built for one platform alone where after the likelihood ratio is updated via a rank-one *channel update* to successively account for the observations from all other platforms. In this case, we exhibit how the change in J-divergence can be used as an information measure in deciding when adding additional sensor platforms reach a point of diminishing return.

1.4 Organization of the Thesis

This thesis is organized as follows: Chapter 2 gives a detailed review of classical Bayesian and Neyman-Pearson detection, the development of the full and reduced-rank Gauss-Gauss detector, and a review of the Gauss-Gauss detector cast in the CCA framework. The development of the multi-channel detector is also given in this chapter. Chapter 4 introduces the three sonar imagery data sets used in this study, reviews the preprocessing and data description for each, and gives a comprehensive study of the effectiveness of the proposed detection system by presenting the results from each data set. Chapter 5 describes the formulations behind Gauss-Gauss likelihood updating when adding more data, presents the low-rank version of the updating, and provides results from two simulations covering both time and channel updating. Finally, Chapter 6 concludes the studies carried out in this research and discusses the goals for future work.

CHAPTER 2

CLASSICAL BINARY HYPOTHESIS DETECTION

2.1 Introduction

Detection problems can be simply described as deciding which of a set of models has most likely generated the realization that we have observed. This is easily cast into the framework of statistical hypothesis testing the most basic being the binary hypothesis test where we must decide only among two models. The binary hypothesis test is described as a decision between either the "true" hypothesis H_1, or the alternative "null" hypothesis H_0. In signal processing applications, one typically encounters tests where observations are assumed to be either a deterministic or stochastic signal additively corrupted by noise under the alternative hypothesis versus that of noise alone under H_0.

The most fundamental building block for hypothesis testing is the Bayesian framework [27] - [29] where the problem is to minimize the expected Bayesian risk involved with making a decision. This leads to a solution involving the comparison of a likelihood ratio with a threshold that is dependent on the costs and *a priori* probabilities. When these costs and probabilities are available, Bayesian detection is optimal, but in general it is difficult to ascertain these free parameters and in practical applications it often comes down to forming educated guesses. The Neyman-Pearson criterion [27] - [29] offers an alternative to the standard Bayesian framework by formulating the hypothesis test as a constrained optimization problem where it is our objective to maximize the probability of detection subject to the constraint that the probability

of false alarm equals a particular value. Solving the optimization problem leads to a likelihood ratio test and a threshold that does not depend on the *a priori* probabilities or costs of the problem. If the assumption is made that under both hypotheses the data is Gaussian distributed, the likelihood ratio test then becomes the standard Gauss-Gauss detector [25].

In developing a detection framework for sonar imagery, the *a priori* probabilities are impossible to ascertain as this would require knowledge of how densely targets are spaced on the ocean floor even before sensors are deployed to the field. Additionally, assigning costs to the detection problem is heuristic and user-dependent. Thus, we choose to restrict our attention to the Neyman-Pearson detection framework and more specifically the Gauss-Gauss framework. One thing that must be defined irregardless of the framework are the conditional probability densities under both hypotheses. Although it is well-known [30] that the data resulting from coherent beamforming is not described well with a Gaussian distribution, this assumption for the sonar data is made here due to the ease in theoretical and computational implementation of the Neyman-Pearson criterion.

The outline of this chapter is as follows. Section 2.2 starts from the ground up with a brief review of Bayesian detection and the Neyman-Pearson criterion. Section 2.3 then develops both the full-rank and reduced-rank Gauss-Gauss detector [25] and presents a review of the Gauss-Gauss detector cast in the coordinate system provided by two-channel Canonical Correlation Analysis (CCA) [15]. Concluding remarks are given in Section 2.4.

2.2 Classical Detection Review

2.2.1 Bayesian Detection

Consider the classical detection problem of choosing between two hypotheses [27] where each hypothesis relates to a point in the observation space denoted as Z.

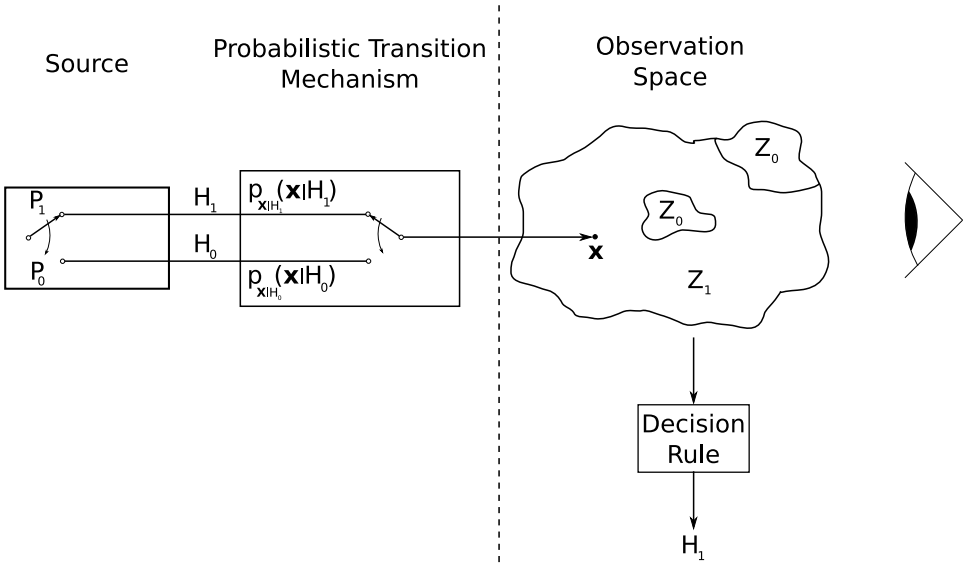


Figure 2.1: Classical Detection.

From an abstract point of view, we assume that an unobserved source generates one of two possible random variables, denoted as H_0 and H_1, that influences the observation that we make according to a conditional probability statement. Based on this measurement we would then like to infer which random variable was generated by the source to the best of our ability. Figure 2.1 gives a graphical overview of the classical detection problem. Clearly, each time we conduct the test there are four possible outcomes. Those are: (a) H_0 is true and we choose H_0, (b) H_0 is true and we choose H_1, (c) H_1 is true and we choose H_1, and (d) H_1 is true but we choose H_0. The first and third outcomes lead to correct decisions while the second and fourth outcomes lead to erroneous decisions. The Bayes test is based on two assumptions. First, the two hypotheses, H_0 and H_1, are generated according to a binary probability law represented by P_0 and P_1, respectively. These probabilities represent the prior observer's information about the hypotheses before the detection is conducted. The second assumption is that there is a cost associated with each of the four courses of action described above. These costs are denoted by, C_{00}, C_{10}, C_{11}, and C_{01}, for outcomes 1-4, respectively. It is (for obvious reasons) assumed that the cost of a

wrong decision is higher than the cost of a correct decision, i.e. $C_{10} > C_{00}$ and $C_{01} > C_{11}$. The goal of the Bayes test is to design a decision rule so that the expected cost of a decision is as small as possible, which subsequently leads to minimizing the Bayesian risk when making the decision. If we denote the expected value of the cost as the risk $\mathcal{R}$, we can then write [27]

$$\begin{aligned} \mathcal{R} &= C_{00}P_0Pr(\text{decide } H_0|H_0) \\ &+ C_{10}P_0Pr(\text{decide } H_1|H_0) \\ &+ C_{11}P_1Pr(\text{decide } H_1|H_1) \\ &+ C_{01}P_1Pr(\text{decide } H_0|H_1) \end{aligned} \tag{2.1}$$

Since the decision rule is binary, i.e. there are only two possible decisions, we can view the rule as a division of the observation space into two parts Z_0 and Z_1. In other words, if the observation is found in the region Z_0 the hypothesis H_0 is declared true and if the observation is found in the region Z_1 the hypothesis H_1 is declared true. By viewing the problem in this manner we can now express the risk in terms of the decision regions and probabilities as,

$$\begin{aligned} \mathcal{R} &= C_{00}P_0 \int_{Z_0} p_{\mathbf{X}|H_0}(\mathbf{x}|H_0)\, d\mathbf{x} \\ &+ C_{10}P_0 \int_{Z_1} p_{\mathbf{X}|H_0}(\mathbf{x}|H_0)\, d\mathbf{x} \\ &+ C_{11}P_1 \int_{Z_1} p_{\mathbf{X}|H_1}(\mathbf{x}|H_1)\, d\mathbf{x} \\ &+ C_{01}P_1 \int_{Z_0} p_{\mathbf{X}|H_1}(\mathbf{x}|H_1)\, d\mathbf{x}. \end{aligned} \tag{2.2}$$

Because each element of $\mathbf{x}$ must be assigned to either the Z_0 or Z_1 in the observation space Z, we can say that $Z = Z_0 \cup Z_1$ and $Z_0 \cap Z_1 = \emptyset$. Noting that integrating the conditional densities over the entire observation space is defined to be one, we can

rewrite (2.2) as

$$\begin{aligned}\mathcal{R} &= P_0C_{10} + P_1C_{11} \\ &+ \int_{Z_0} \left[P_1(C_{01} - C_{11})p_{\mathbf{X}|H_1}(\mathbf{x}|H_1) - P_0(C_{10} - C_{00})p_{\mathbf{X}|H_0}(\mathbf{x}|H_0)\right] d\mathbf{x} \quad (2.3)\end{aligned}$$

The first two terms in (2.3) represent the fixed cost and the integral represents the cost controlled by the points in the observation space, Z that are assigned to Z_0. All values of $\mathbf{x}$ in which the second term in the integral is larger than the first should be assigned to Z_0 as they negatively contribute to the expected cost. Likewise, all values of $\mathbf{x}$ in which the first term is larger than the second should be assigned to Z_1 as they contribute positively to the risk. This statement can alternatively be written as [27]

$$l(\mathbf{x}) = \frac{p_{\mathbf{X}|H_1}(\mathbf{x}|H_1)}{p_{\mathbf{X}|H_0}(\mathbf{x}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})}. \quad (2.4)$$

where $l(\mathbf{x})$ is called the *likelihood ratio* and the term on the right is simply the threshold of the test which is dependent on the *a priori* probabilities and costs associated with the detection problem. Thus, Bayes criterion simply leads to a likelihood ratio test (LRT).

2.2.2 Neyman-Pearson Criterion

As stated earlier in the introduction, it is impossible to realistically assign costs and *a priori* probabilities to our detection problem (as it typically is with any practical problem) and thus difficult to find a suitable threshold for the Bayesian detector. The Neyman-Pearson criteria [27], [31] overcomes this by restating the problem in terms of the two conditional probabilities P_D and P_F representing the probability of detection and false alarm, respectively. We would like to make P_F as small as possible and P_D as large as possible but these are often times two conflicting statements. Thus, we constrain the probability of false alarm to be a particular value, $P_F = \alpha$, and look to maximize P_D or alternatively minimize $P_M = 1 - P_D$, the probability of

miss-detection. To solve the problem we construct the Lagrangian

$$\begin{aligned}\mathcal{L} &= P_M + \lambda\left(P_F - \alpha\right) \\ &= \int_{Z_0} p_{\mathbf{X}|H_1}\left(\mathbf{x}|H_1\right) d\mathbf{x} + \lambda\left(\int_{Z_1} p_{\mathbf{X}|H_0}\left(\mathbf{x}|H_0\right) d\mathbf{x} - \alpha\right)\end{aligned}$$

which, using a similar argument from before, can be rewritten as

$$\mathcal{L} = \lambda(1-\alpha) + \int_{Z_0}\left[p_{\mathbf{X}|H_1}\left(\mathbf{x}|H_1\right) - \lambda p_{\mathbf{X}|H_0}\left(\mathbf{x}|H_0\right)\right] d\mathbf{x} \tag{2.5}$$

For any positive Lagrangian multiplier, we can see that $\mathcal{L}$ is minimized when the term inside the integral is negative thus resulting in the LRT [27]

$$l(\mathbf{x}) = \frac{p_{\mathbf{X}|H_1}(\mathbf{x}|H_1)}{p_{\mathbf{X}|H_0}(\mathbf{x}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda. \tag{2.6}$$

Denoting the probability density of the likelihood ratio as $p_L(l)$, the threshold of the LRT is chosen to satisfy the equation

$$P_F = \int_{\lambda}^{\infty} p_{L|H_0}(l|H_0) dl = \alpha \tag{2.7}$$

which is independent of the costs and *a priori* probabilities defined before. So even though the problem is stated in a slightly different manner, we find that both the Bayesian and Neyman-Pearson criteria involve the use of a likelihood ratio, $l(\mathbf{x})$, to map observations to a scalar, real-valued statistic which is compared to a threshold to form our detection decision. The only difference among the two methods is the threshold that is chosen which ultimately determines how one partitions the observation space to form decisions.

2.3 Gauss-Gauss Detection

2.3.1 Preliminaries

The next question is how to characterize the conditional densities describing our data under both hypotheses to facilitate the computation of the likelihood ratio. For certain classes of distributions, it is possible to obtain closed-form solutions and easily

compute the likelihood function. An example of such a class is the multivariate Gaussian distribution which serves as a low-level model of what actually exists. It must be mentioned that one cannot put an extreme amount of faith in their modeling capabilities as often times creating models with an increasing amount of accuracy amounts to an increase in the degree of complexity, hence hindering its capability to be practically implementable. For the detection problem at hand, we assume that the statistical nature of our problem can be completely characterized by second-order covariance matrices for each hypothesis. More specifically, we assume that realizations of our random vector are circular symmetric complex Gaussian distributed [32] with zero-mean and covariance matrix $E_{H_1}\left[\mathbf{x}\mathbf{x}^H\right] = R_1$ under hypothesis H_1 versus $E_{H_0}\left[\mathbf{x}\mathbf{x}^H\right] = R_0$ under hypothesis H_0. Note that E_{H_1} and E_{H_0} denote conditional expectation under H_1 and H_0, respectively, and assuming a circular symmetric complex Gaussian distribution implies that our observation is proper under both hypotheses so that $E_{H_1}\left[\mathbf{x}\mathbf{x}^T\right] = E_{H_0}\left[\mathbf{x}\mathbf{x}^T\right] = \mathbf{O}$. No specific structure is assumed for either covariance matrix with the only restriction that they be positive-definite (PD) under both hypotheses. More compactly, we write the detection statement as

$$\begin{aligned} H_1 &: \mathbf{x} \sim \mathcal{CN}(\mathbf{0}, R_1) \\ H_0 &: \mathbf{x} \sim \mathcal{CN}(\mathbf{0}, R_0) \end{aligned}$$

With this assumption, we have the following conditional densities

$$\begin{aligned} p_{\mathbf{x}|H_0}(\mathbf{x}|H_0) &= \frac{1}{\pi^n |\det R_0|} e^{-\mathbf{x}^H R_0^{-1} \mathbf{x}} \\ p_{\mathbf{x}|H_1}(\mathbf{x}|H_1) &= \frac{1}{\pi^n |\det R_1|} e^{-\mathbf{x}^H R_1^{-1} \mathbf{x}} \end{aligned}$$

where $(\cdot)^H$ denotes the Hermitian transpose and $\det R$ is the determinant of the matrix R. Realizing that the application of any monotonically increasing function of the likelihood ratio (along with a suitable correction for the threshold) leaves our decision unchanged and that the likelihood ratio is always positive, it becomes desirable in

this situation to look at the log of the likelihood ratio (which we will still denote as $l(\mathbf{x})$)

$$l(\mathbf{x}) = \log \frac{|\det R_0|}{|\det R_1|} + \mathbf{x}^H \left(R_0^{-1} - R_1^{-1}\right) \mathbf{x}$$

Simplifying the log-likelihood ratio (LLR) even further by noting that the first term in this equation is independent of the observation and can be taken into account by adjusting the threshold, we obtain the final expression for the LLR

$$l(\mathbf{x}) = \mathbf{x}^H \left(R_0^{-1} - R_1^{-1}\right) \mathbf{x} \tag{2.8}$$

This leads to the log-likelihood ratio test (LLRT)

$$l(\mathbf{x}) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda.$$

where again the threshold λ is chosen to achieve a desirable false alarm rate.

2.3.2 Full-Rank Gauss-Gauss Detection

With the functional form of the LLR established previously and given the observation $\mathbf{x} \in \mathbb{C}^n$, we can rewrite the test statistic in terms of the whitened version of our data as

$$l(\mathbf{y}) = \mathbf{x}^H R_0^{-H/2} \left(I - S^{-1}\right) R_0^{-1/2} \mathbf{x} = \mathbf{y}^H \left(I - S^{-1}\right) \mathbf{y}$$

where $S = R_0^{-1/2} R_1 R_0^{-H/2}$ is a "signal-to-noise ratio" matrix and the whitened vector $\mathbf{y} = R_0^{-1/2}\mathbf{x}$ has covariance structure

$$\begin{aligned} E_{H_0}\left[\mathbf{y}\mathbf{y}^H\right] &= I \\ E_{H_1}\left[\mathbf{y}\mathbf{y}^H\right] &= S \end{aligned}$$

We then go a step further by taking the eigenvalue decomposition of the matrix S so that

$$\begin{aligned} S &= U \Lambda U^H \\ U^H U &= U U^H = I \end{aligned}$$

and

$$\Lambda = \text{diag}\left[\lambda_1, \ldots, \lambda_n\right]$$

Now, we can rewrite the LLR as

$$l(\mathbf{z}) = \mathbf{y}^H U \left(I - \Lambda^{-1}\right) U^H \mathbf{y} = \mathbf{z}^H \left(I - \Lambda^{-1}\right) \mathbf{z} \tag{2.9}$$

where the vector $\mathbf{z} = U^H \mathbf{y}$ has covariance structure

$$E_{H_0}\left[\mathbf{z}\mathbf{z}^H\right] = I$$
$$E_{H_1}\left[\mathbf{z}\mathbf{z}^H\right] = \Lambda$$

Measuring the performance of the detector should obviously be portrayed in terms of probability of detection. However, this requires an accurate characterization of the distribution of (2.9) under hypothesis H_1 which can be difficult to do in general. With that said, an alternative is to simply characterize the moments of the log-likelihood ratio. The J-divergence is just that and measures the difference in the means of the log-likelihood ratio under H_1 and H_0. Taking advantage of the cyclic property of the trace operation, we note that the first moment of the LLR can be written as

$$\begin{aligned} E[l(\mathbf{x})] &= E\left[\mathbf{x}^H \left(R_0^{-1} - R_1^{-1}\right) \mathbf{x}\right] \\ &= \text{tr}\left(\left(R_0^{-1} - R_1^{-1}\right) E\left[\mathbf{x}\mathbf{x}^H\right]\right) \end{aligned}$$

and the divergence of the detector in this coordinate system becomes

$$\begin{aligned} J &= E_{H_1}[l(\mathbf{x})] - E_{H_0}[l(\mathbf{x})] \\ &= \text{tr}\left(-2I + R_0^{-1}R_1 + R_0 R_1^{-1}\right) \\ &= \text{tr}\left(-2I + S + S^{-1}\right) \\ &= \text{tr}\left(-2I + \Lambda + \Lambda^{-1}\right) \\ &= \sum_{i=1}^{n} \left(-2 + \lambda_i + \lambda_i^{-1}\right) \end{aligned} \tag{2.10}$$

Therefore, we find that it is the not necessarily the largest eigenvalues of the S matrix that are important for detection but rather the term $\left(\lambda_i + \lambda_i^{-1}\right)$. One might be quick to note that this analysis shares many things in common to standard Principle Components Analysis (PCA) where one retains only the largest eigenvalues of the covariance matrix. However, choosing to retain λ_j and discard λ_i because $\lambda_i < \lambda_j$ when in fact it is the case that $\left(\lambda_i + \lambda_i^{-1}\right) > \left(\lambda_j + \lambda_j^{-1}\right)$ can unintentionally result in a drop in detection performance as the smallest eigenvalues can sometimes contribute more to divergence than those that are larger.

2.3.3 Reduced-Rank Gauss-Gauss Detection

When performing rank-p detection, we sort the eigenvalues of the S matrix in a descending fashion such that

$$\left(\lambda_1 + \lambda_1^{-1}\right) > \cdots > \left(\lambda_n + \lambda_n^{-1}\right)$$

We then partition the coordinate system according to the following statements

$$\begin{aligned}
U &= [U_p \; U_{p+1}] \\
U_p &= [\mathbf{u}_1 \; \cdots \; \mathbf{u}_p] \\
U_{p+1} &= [\mathbf{u}_{p+1} \; \cdots \; \mathbf{u}_n] \\
\Lambda &= \begin{bmatrix} \Lambda_p & \mathbf{O} \\ \mathbf{O} & \Lambda_{p+1} \end{bmatrix} \\
\Lambda_p &= \operatorname{diag}[\lambda_1 \; \cdots \; \lambda_p] \\
\Lambda_{p+1} &= \operatorname{diag}[\lambda_{p+1} \; \cdots \; \lambda_n]
\end{aligned}$$

and form the reduced-rank approximation of the observation vector $\tilde{\mathbf{z}} = U_p^H R_0^{-1/2} \mathbf{x} \in \mathbb{C}^p$. The log-likelihood ratio can then be written as

$$l_p(\tilde{\mathbf{z}}) = \tilde{\mathbf{z}}^H \left(I - \Lambda_p^{-1}\right) \tilde{\mathbf{z}}$$

with an associated J-divergence

$$J_p = \sum_{i=1}^{p} \left(-2 + \lambda_i + \lambda_i^{-1}\right)$$

There are many possible methods for choosing a suitable value for p. Similar to PCA, one might be interested in solving the optimization problem

$$\min \{q : J_q/J \geq 1 - \epsilon\}$$

where typically ϵ is very small. In other words, we choose the smallest set of coordinates that retain a large percentage of the J-divergence. Note that in the special case that $R_1 = R_s + R_0$, i.e. a signal-plus-noise model, the matrix S becomes

$$S = R_0^{-1/2} R_1 R_0^{-H/2} = I + R_0^{-1/2} R_s R_0^{-H/2}$$

which will undoubtedly have eigenvalues that are all greater than one. Thus, in such a situation it becomes optimal to pick the largest eigenvalues when performing reduced-rank detection as this orthogonally projects the data onto one-dimensional basis vectors with high per-mode signal-to-noise ratio (SNR).

2.3.4 CCA-Based Detection

The Gauss-Gauss detector in the previous two sections can also be cast in the CCA framework in which the detection test and detection criterion are formed in terms of the canonical coordinates and canonical correlations. For a review of CCA the reader is referred to Appendix A. Canonical correlation analysis offers an ideal framework for coherent-based detection and feature extraction [13] - [15]. The extracted canonical correlations for the two data channels provide a coherence (or incoherence) measure that can be used to determine if a target is present (or absent). Therefore, only the dominant correlations need to be retained to build a rank-p detector that maximizes the divergence. In this case, we restrict our attention to a signal-plus-noise model and consider the message vector or signal of interest $\mathbf{x} \in \mathbb{C}^n$ which is a zero-mean, complex

Gaussian random vector with covariance matrix R_{xx}. Given the measurement or observation $\mathbf{y} \in \mathbb{C}^n$, we consider the detection problem

$$\begin{aligned} H_1 &: \mathbf{y} = \mathbf{x} + \mathbf{n} \\ H_0 &: \mathbf{y} = \mathbf{n} \end{aligned}$$

where $\mathbf{n} \in \mathbb{C}^n$ is a zero-mean, complex Gaussian random vector with covariance matrix $R_{nn} = R_0$ and is uncorrelated with the signal which has covariance matrix $R_{xx} = R_s$. The detection problem is displayed diagrammatically in Figure 2.2. In the CCA framework posed by this problem, the matrix S can be rewritten as

$$S = R_{nn}^{-1/2}(R_{xx} + R_{nn})R_{nn}^{-H/2} \tag{2.11}$$

and we define the *coherence matrix* [15]

$$C = R_{xx}^{-1/2}R_{xy}R_{yy}^{-H/2} = R_{xx}^{H/2}R_{yy}^{-H/2} = R_{xx}^{H/2}\left(R_{xx} + R_{nn}\right)^{-H/2} \tag{2.12}$$

With the quadratic detector described earlier, we can then write the log-likelihood ratio in terms of the squared coherence matrix, CC^H, as

$$\begin{aligned} l(\mathbf{y}) &= \mathbf{y}^H R_{nn}^{-H/2}\left(I - S^{-1}\right)R_{nn}^{-1/2}\mathbf{y} \\ &= \mathbf{y}^H R_{xx}^{-H/2}\left(\left[\left(CC^H\right)^{-1} - I\right]^{-1} - CC^H\right)R_{xx}^{-1/2}\mathbf{y} \end{aligned}$$

We then take the Singular Value Decomposition (SVD) of the coherence matrix so that $C = FKG^H$ with $FF^H = F^HF = GG^H = G^HG = I$ and $K = \mathrm{diag}\left[k_1, \ldots, k_n\right]$. From here it is easy to show that $CC^H = FK^2F^H$ and $F^HR_{xx}^{-1/2} = K^{-1}G^HR_{yy}^{-1/2}$ so that we may rewrite the log-likelihood ratio as

$$\begin{aligned} l(\mathbf{y}) &= \mathbf{y}^H R_{xx}^{-H/2}F\left(\left[\left(K^2\right)^{-1} - I\right]^{-1} - K^2\right)F^HR_{xx}^{-1/2}\mathbf{y} \\ &= \mathbf{y}^H R_{yy}^{-H/2}G\left(\left[I - K^2\right]^{-1} - I\right)G^HR_{yy}^{-1/2}\mathbf{y} \\ &= \sum_{i=1}^{n}\left|\mathbf{g}_i^H R_{yy}^{-1/2}\mathbf{y}\right|^2\left(\frac{k_i^2}{1-k_i^2}\right) \end{aligned} \tag{2.13}$$

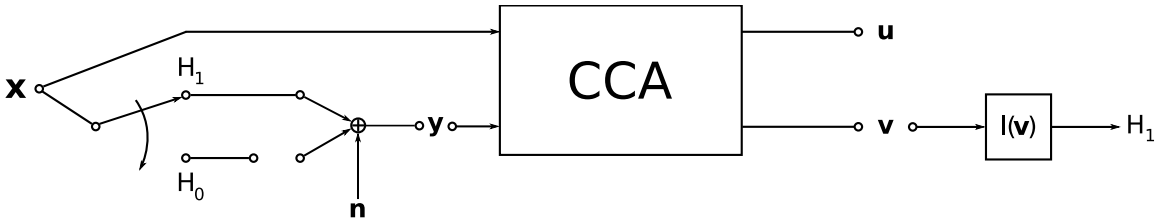


Figure 2.2: Signal-Plus-Noise Detection in CCA Coordinates

This is the standard Gauss-Gauss log-likelihood ratio, expressed in the coordinates $G^H R_{yy}^{-1/2}\mathbf{y}$ which under H_1 are the canonical coordinates of $\mathbf{y}$.

It can be shown [15] that the expected value of the log-likelihood ratio under H_0 can be written as

$$E_{H_0}\left[l(\mathbf{y})\right] = \mathrm{tr}\left(K^2\right) = \sum_{i=1}^{n} k_i^2$$

and likewise under H_1

$$E_{H_1}\left[l(\mathbf{y})\right] = \mathrm{tr}\left(\left[I - K^2\right]^{-1} - I\right) = \sum_{i=1}^{n} \frac{k_i^2}{1-k_i^2}$$

Therefore, the J-divergence in the CCA coordinate system can be written as

$$J = E_{H_1}\left[l(\mathbf{y})\right] - E_{H_0}\left[l(\mathbf{y})\right] = \sum_{i=1}^{n} \frac{k_i^4}{1-k_i^2} \tag{2.14}$$

The function $k_i^4/(1-k_i^2)$ is a monotonically increasing function of $k_i \in (0,1]$ and therefore, if the canonical correlations are sorted in a descending fashion, the divergence is a monotonically decreasing function of the coordinate index i. Consequently, the rank-p detector that maximizes divergence is the detector that uses those coordinates associated with the largest p canonical correlations where p can be chosen as described previously. It is noted in [15] that the canonical correlations and the eigenvalues of the matrix S can be related through the identities

$$\lambda_i = \frac{1}{1-k_i^2}$$
$$k_i = \sqrt{1-\lambda_i^{-1}}$$

We can see that, since k_i is bounded between zero and one, all the eigenvalues of the matrix S must be greater than one confirming an observation we have made previously. We can also see that the larger the value of λ_i, the closer k_i approaches its upper bound of one. Thus, representing the data in the coordinate system associated with the largest canonical correlations can again be interpreted as orthogonally projecting the data onto one-dimensional basis vectors with high per-mode SNR. This leads us to conclude that the low-rank detector built in the previous section can alternatively be built using CCA when the underlying model is a signal-plus-noise model.

2.4 Conclusion

In this chapter, optimum Bayesian detection and the Neyman-Pearson criterion were reviewed. Bayesian detection is optimal when the decision costs and prior probabilities are known. However, in general it is difficult to generate these values due to lack of *a priori* information about the environment. The Neyman-Pearson criterion provides a decision rule for testing hypothesis in which the decision costs and prior probabilities are not required. In the case where data is Gaussian distributed under both the null and alternative hypotheses, the Gauss-Gauss detector is obtained. In this method the likelihood ratio and J-divergence can be reformulated in terms of the eigenvalues and eigenvectors of the "signal-to-noise ratio" matrix, where rank reduction can be performed by maximizing the divergence with a rank constraint.

This reduced-rank Gauss-Gauss detector is then cast in the CCA framework in the special case of a signal-plus-noise model. In this case, we form the CCA coherence matrix between the signal and the observation under H_1. Taking the SVD of this coherence matrix, the data is then filtered into its canonical coordinates under H_1 and subsequently applied to a quadratic detector. The J-divergence in this coordinate system then becomes written in terms of the sum of squared canonical correlations. As a consequence, it is concluded that low-rank detection can alternatively be performed

in this coordinate system by retaining the largest canonical correlations which again maximizes the divergence subject to a rank constraint.

CHAPTER 3

MULTI-CHANNEL COHERENCE ANALYSIS (MCA) DETECTION

3.1 Introduction

In this chapter, we begin by reviewing the Multi-Channel Coherence Analysis (MCA) framework. The two-channel CCA-based detector presented in [1] looks for coherence among two sonar images. To extend and generalize this idea to multiple sources of information, one must define a coordinate system that finds the linear relationships among every pair of sensory data channels. MCA discovers this coherence structure by whitening the observations from each channel to remove the second-order information pertaining to that channel alone. The MCA-based detector then finds sets of linear mapping functions that maximize the sum of the cross-correlations among the channels thus "discovering" the coherence structure across the platforms. Solving this optimization problem boils down to simply performing a generalized eigenvalue decomposition from which the mapping vectors and multi-channel correlations are obtained.

Our *detection hypothesis* for this multi-sonar detection problem is that the presence of a target in the disparate platform data will lead to a higher level of coherence versus when our set of observations consist of background clutter arising from returns off the seafloor clutter. Based on this detection hypothesis, we then recast the standard Gauss-Gauss detector into the MCA coordinate system under hypothesis H_1. Doing so results in new formulations for the log-likelihood ratio which can be

written as a quadratic term in the MCA variates under H_1. The J-divergence then becomes decomposed in terms of expressions measuring the coherence shared between data channels under H_1 and the SNR local to each individual channel. With the assumption of low SNR in each individual data channel, we then approximate the log-likelihood ratio and J-divergence in terms of a detector that only takes advantage of coherent information among the channels under H_1. We finally consider a simple example to illuminate the inner workings of the proposed MCA-based detector which is then compared to the detector presented in [25].

The outline of this chapter is as follows. Section 3.2 reviews MCA and highlights its properties and relationship to CCA. Section 3.3 presents the MCA-based detection method. A simple example is then given to illuminate the effectiveness of the MCA-based detector for multi-channel detection problems. Concluding remarks are then given in Section 3.4.

3.2 Multi-Channel Coherence Analysis (MCA)

Consider N zero mean random vectors, $\mathbf{x}_1$, $\mathbf{x}_2$,..., and $\mathbf{x}_N$, representing multiple data channels comprising the composite data channel $\mathbf{z} = [\mathbf{x}_1^H \mathbf{x}_2^H \cdots \mathbf{x}_N^H]^H \in \mathbb{C}^{d\times 1}$. For the sake of simplicity and without loss of generality, we will assume that all random vectors are zero mean throughout this analysis. Let each channel $\mathbf{x}_j \in \mathbb{C}^{d_j \times 1}$ be of dimension d_j and we denote $d = \sum_{j=1}^{N} d_j$. The $d \times d$ dimensional covariance matrix of the composite data channel $\mathbf{z}$ is given by

$$R_{\mathbf{zz}} = E\left[\mathbf{z}\mathbf{z}^H\right] = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1N} \\ R_{21} & R_{22} & \cdots & R_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ R_{N1} & R_{N2} & \cdots & R_{NN} \end{bmatrix}, \tag{3.1}$$

where $R_{jk} = E[\mathbf{x}_j \mathbf{x}_k^H]$ is the auto-covariance ($j = k$) or cross-covariance ($j \neq k$) matrices of data channels $\mathbf{x}_j$ and $\mathbf{x}_k$ and clearly we have $R_{jk} = R_{kj}^H$. No specific

structure is assumed for $R_{\mathbf{zz}}$, e.g. a Toeplitz structure resulting from stationary assumptions, with the only assumption that it be positive definite (PD).

Similar to two-channel CCA [14], [33] the ith multi-channel coordinate of the jth channel is found by searching for the ith coordinate mapping vector, $\boldsymbol{\alpha}_{i,j}$, of data channel $\mathbf{x}_j$. This linear transformation produces the ith multi-channel coordinate for the jth channel,

$$v_{i,j} = \boldsymbol{\alpha}_{i,j}^H \mathbf{x}_j. \tag{3.2}$$

If the ith coordinate mapping vectors are found for all N channels, we obtain the *composite coordinate mapping* vector $\mathbf{a}_i = \begin{bmatrix} \boldsymbol{\alpha}_{i,1}^H & \boldsymbol{\alpha}_{i,2}^H & \cdots & \boldsymbol{\alpha}_{i,N}^H \end{bmatrix}^H$. This is then used to find the *composite coordinate* vector $\mathbf{v}_i = \begin{bmatrix} v_{i,1} & v_{i,2} & \cdots & v_{i,N} \end{bmatrix}^T$ which consists of the ith multi-channel coordinate of every channel. The associated covariance matrix of $\mathbf{v}_i$ is given by

$$R_{\mathbf{v}_i\mathbf{v}_i} = \begin{bmatrix} \boldsymbol{\alpha}_{i,1}^H R_{11} \boldsymbol{\alpha}_{i,1} & \cdots & \boldsymbol{\alpha}_{i,1}^H R_{1N} \boldsymbol{\alpha}_{i,N} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\alpha}_{i,N}^H R_{N1} \boldsymbol{\alpha}_{i,1} & \cdots & \boldsymbol{\alpha}_{i,N}^H R_{NN} \boldsymbol{\alpha}_{i,N} \end{bmatrix}. \tag{3.3}$$

Recall that in the two-channel CCA [14], [33] the correlations between the mapped coordinates are maximized subject to the constraint that the transformed coordinates have unit variance. In the multi-channel case, however, the analysis is not as well-defined as all correlations between all possible pairs of channels must be maximized simultaneously. To accomplish this, one approach [6] is to maximize the sum of all correlations subject to the unit trace constraint of matrix $R_{\mathbf{v}_i\mathbf{v}_i}$. Thus, the optimization problem for finding the i^{th} composite coordinate mapping vector $\mathbf{a}_i$ using the objective function and constraint just described becomes

$$\begin{aligned} \mathbf{a}_i &= \arg\max_{\mathbf{a}_i} \sum_{j=1}^{N} \sum_{k=1}^{N} \boldsymbol{\alpha}_{i,j}^H R_{j,k} \boldsymbol{\alpha}_{i,k} \\ &= \arg\max_{\mathbf{a}_i} \sum_{j=1}^{N} \sum_{k=1}^{N} \left[R_{\mathbf{v}_i\mathbf{v}_i} \right]_{j,k} \end{aligned}$$

subject to the constraint

$$\mathrm{tr}\left(R_{\mathbf{v}_i\mathbf{v}_i}\right) = \sum_{j=1}^{N} \boldsymbol{\alpha}_{i,j}^H R_{jj} \boldsymbol{\alpha}_{i,j} = \sum_{j=1}^{N} \left[R_{\mathbf{v}_i\mathbf{v}_i}\right]_{j,j} = 1,$$

It is shown [6] that the constrained optimization problem for finding the set of mapping vectors for the i^{th} coordinate, $\boldsymbol{\alpha}_{i,j}$ using a Lagrange multiplier method leads to

$$\sum_{k=1}^{N} R_{jk} \boldsymbol{\alpha}_{i,k} = \lambda_i R_{jj} \boldsymbol{\alpha}_{i,j},$$

or in matrix notation as

$$R_{\mathbf{zz}} \mathbf{a}_i = \lambda_i D \mathbf{a}_i, \tag{3.4}$$

where D is a block diagonal matrix with diagonal blocks R_{jj}, $\forall\ j \in [1, N]$, i.e. $D = \mathrm{diag}\left[R_{11}, R_{22}, \ldots, R_{NN}\right]$. Simply left-multiplying (3.4) by $\mathbf{a}_i^H$ and recalling that $\mathbf{a}_i^H D \mathbf{a}_i$ is constrained to equal 1 yields $\lambda_i = \mathbf{a}_i^H R_{\mathbf{zz}} \mathbf{a}_i = \sum_{j=1}^{N} \sum_{k=1}^{N} E\left[v_{i,j} v_{i,k}^*\right]$ which implies that each Lagrange multiplier is the sum of the correlations among all mapped variates.

The result of (3.4) represents a generalized eigenvalue problem for which standard methods of solution are well-known [34]. We will then consider the simultaneous solution to the problem by assuming the user to have access to all mapping vectors $\mathbf{a}_i$'s, $i \in [1, d]$ and write (3.4) as $R_{\mathbf{zz}} A = D A \Lambda$ where $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_d \end{bmatrix}$ consists of all d coordinate mapping vectors, and $\Lambda = \mathrm{diag}\left[\lambda_1, \lambda_2, \ldots, \lambda_d\right]$ consists of all d eigenvalues. Since both $R_{\mathbf{zz}}$ and D represent symmetric covariance matrices and further D is positive definite (PD), the solution can then be rewritten in terms of a standard eigenvalue problem

$$\begin{aligned} R_{\mathbf{zz}} A &= D^{\frac{1}{2}} D^{\frac{H}{2}} A \Lambda \\ D^{-\frac{1}{2}} R_{\mathbf{zz}} A &= D^{\frac{H}{2}} A \Lambda \\ E P &= P \Lambda \end{aligned}$$

where $E = D^{-\frac{1}{2}} R_{\mathbf{zz}} D^{-\frac{H}{2}}$ and $P = D^{\frac{H}{2}} A$. Since E, which we refer to as the coherence

matrix[1], is also symmetric all eigenvalues are guaranteed [34] to be real-valued and P forms a linearly independent, orthonormal basis of eigenvectors, i.e. $P^H P = PP^H = I$. Clearly, we may then extract the matrix A via $A = D^{-\frac{H}{2}} P$.

Inspection of matrix E shows that it is simply the composite covariance matrix of the whitened version of $\mathbf{z} = [\ \mathbf{x}_1^H \ \cdots \ \mathbf{x}_N^H\ ]^H$. That is, if we define this whitened version of the composite data channel vector by $\mathbf{w} = [\ \mathbf{w}_1^H \ \cdots \ \mathbf{w}_N^H\ ]^H = D^{-\frac{1}{2}}\mathbf{z}$ where $E\left[\mathbf{w}_j \mathbf{w}_j^H\right] = I_{d_j}$ and $E\left[\mathbf{w}_j \mathbf{w}_k^H\right] = R_{jj}^{-\frac{1}{2}} R_{jk} R_{kk}^{-\frac{H}{2}}$, then the whitened composite vector $\mathbf{w}$ has correlation matrix $E\left[\mathbf{w}\mathbf{w}^H\right] = D^{-\frac{1}{2}} R_{\mathbf{zz}} D^{-\frac{H}{2}} = E$. Matrix P is then used to map the whitened channels to their multi-channel coordinates. In order to find mapping vectors corresponding to the principal coordinates [6], we only consider the $r = \min_j \{d_j\}$ coordinates such that $\lambda_1 > \lambda_2 > \ldots > \lambda_r$. Thus, $\Lambda = \mathrm{diag}\left[\lambda_1, \lambda_2, \ldots, \lambda_r\right]$ will become a $r \times r$ diagonal matrix composed of the dominant eigenvalues and P will become a $d \times r$ matrix composed of the eigenvectors corresponding to r dominant eigenvalues.

To find the mapped coordinate vector, $\mathbf{v}$, that contains all mapped coordinates for all N channels, we will first define matrix Ψ_j (dimension $d_j \times r$) to contain those dominant r eigenvectors $\mathbf{p}_{i,j}$, $\forall\ i \in [1, r]$ of the mapping matrix P that correspond to the j^{th} channel

$$\Psi_j = \left[\ \mathbf{p}_{1,j} \quad \mathbf{p}_{2,j} \quad \cdots \quad \mathbf{p}_{r,j}\ \right],\ \forall\ j \in [1, N]\,. \tag{3.5}$$

Clearly, the connection between P and Ψ_j is evident

$$P = \begin{bmatrix} \Psi_1 \\ \Psi_2 \\ \vdots \\ \Psi_N \end{bmatrix}_{d \times r}. \tag{3.6}$$

All of the mapped coordinates of the jth channel can then be found by

[1]Note that in the two-channel CCA, the off-diagonal blocks of this matrix become the coherence matrix [13], [15]

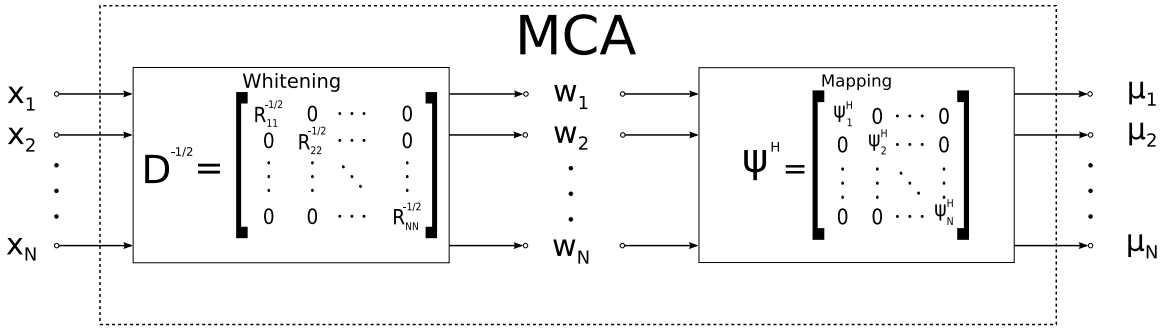


Figure 3.1: MCA Processing Block Diagram.

$$\boldsymbol{\mu}_j = \Psi_j^H R_{jj}^{-\frac{1}{2}} \mathbf{x}_j, \ \forall\ j \in [1, N], \tag{3.7}$$

where $\boldsymbol{\mu}_j = [\ v_{1,j} \quad v_{2,j} \quad \cdots \quad v_{r,j}\ ]^T$. With this definition in mind, we have the following two properties

$$\begin{aligned}
\sum_{j=1}^{N} E\left[\boldsymbol{\mu}_j \boldsymbol{\mu}_j^H\right] &= \sum_{j=1}^{N} \Psi_j^H \Psi_j \\
&= P^H P = A^H D A = I_r \\
\sum_{j=1}^{N}\sum_{k=1}^{N} E\left[\boldsymbol{\mu}_j \boldsymbol{\mu}_k^H\right] &= \sum_{j=1}^{N}\sum_{k=1}^{N} \Psi_j^H R_{jj}^{-\frac{1}{2}} R_{jk} R_{kk}^{-\frac{H}{2}} \Psi_k \\
&= P^H E P = A^H R_{\mathbf{zz}} A = \Lambda.
\end{aligned} \tag{3.8}$$

Remark 1 *Note the simple property that the sum of all d eigenvalues must equal d itself, i.e.*

$$\begin{aligned}
\sum_{i=1}^{d} \lambda_i &= tr\left(P^H E P\right) = tr(E) = tr\left(D^{-1} R_{\mathbf{zz}}\right) \\
&= \sum_{j=1}^{N} tr\left(I_{d_j}\right) = \sum_{j=1}^{N} d_j = d
\end{aligned} \tag{3.9}$$

If we define block diagonal matrix $\boldsymbol{\Psi}$ that contains the Ψ_j matrices along its diagonal blocks, i.e. $\boldsymbol{\Psi} = \text{diag}\left[\Psi_1, \Psi_2, \ldots, \Psi_N\right]$, then we can resolve all N channels into their multi-channel coordinates using

$$\mathbf{v} = \left[\boldsymbol{\mu}_1^H \ \ \boldsymbol{\mu}_2^H \ \ \cdots \ \ \boldsymbol{\mu}_N^H\right]^H = \boldsymbol{\Psi}^H \mathbf{w} = \boldsymbol{\Psi}^H D^{-\frac{1}{2}} \mathbf{z}. \tag{3.10}$$

Figure 3.1 displays the process behind the MCA analysis filter. As can be seen, similar to CCA [14], [33], all channels are whitened in order to remove the auto-correlation contributions from each individual component thereby allowing one to analyze the linear dependence shared among one another using matrix $\mathbf{\Psi}$.

3.3 MCA-Based Detection

We now turn our attention to MCA-based detection. For this multi-sonar problem, we assume that the observations from all N channels consist of uncorrelated realizations of background noise under H_0 verse the hypothesis that our observations contain correlated signal components corrupted by additive noise. More specifically, for the j^{th} platform we consider the signal-plus-noise model

$$\begin{aligned} H_1 &: \mathbf{x}_j = \mathbf{s}_j + \mathbf{n}_j \\ H_0 &: \mathbf{x}_j = \mathbf{n}_j \end{aligned}$$

where $\mathbf{n}_j \in \mathbb{C}^{d_j}$ and $\mathbf{s}_j \in \mathbb{C}^{d_j}$ are both zero-mean, proper complex Gaussian random vectors with the following auto and cross-covariance matrices

$$\begin{aligned} E\left[\mathbf{n}_j \mathbf{n}_k^H\right] &= \delta_{j-k} R_{\mathbf{n}_j} \\ E\left[\mathbf{s}_j \mathbf{s}_k^H\right] &= R_{\mathbf{s}_j \mathbf{s}_k} \\ E\left[\mathbf{s}_j \mathbf{n}_k^H\right] &= \mathbf{O} \end{aligned}$$

for all $j, k = 1, \ldots, N$. Figure 3.2 shows the graphical setup of the problem under consideration. We then take the observations from all channels and concatenate them into the composite random vector $\mathbf{z} = \left[\mathbf{x}_1^H \ \mathbf{x}_2^H \ \cdots \ \mathbf{x}_N^H\right]^H \in \mathbb{C}^d$ with $d = \sum_{j=1}^N d_j$. The composite random vector $\mathbf{z}$ then has a composite covariance structure

$$E\left[\mathbf{z}\mathbf{z}^H\right] = R_{\mathbf{z}\mathbf{z}} = \begin{bmatrix} R_{\mathbf{x}_1\mathbf{x}_1} & R_{\mathbf{x}_1\mathbf{x}_2} & \cdots & R_{\mathbf{x}_1\mathbf{x}_N} \\ R_{\mathbf{x}_2\mathbf{x}_1} & R_{\mathbf{x}_2\mathbf{x}_2} & \cdots & R_{\mathbf{x}_2\mathbf{x}_N} \\ \vdots & \vdots & \ddots & \vdots \\ R_{\mathbf{x}_N\mathbf{x}_1} & R_{\mathbf{x}_N\mathbf{x}_2} & \cdots & R_{\mathbf{x}_N\mathbf{x}_N} \end{bmatrix}$$

and, of course, $R_{\mathbf{zz}}$ exhibits Hermitian symmetry such that $R_{\mathbf{x}_j\mathbf{x}_k} = R^H_{\mathbf{x}_k\mathbf{x}_j}$.

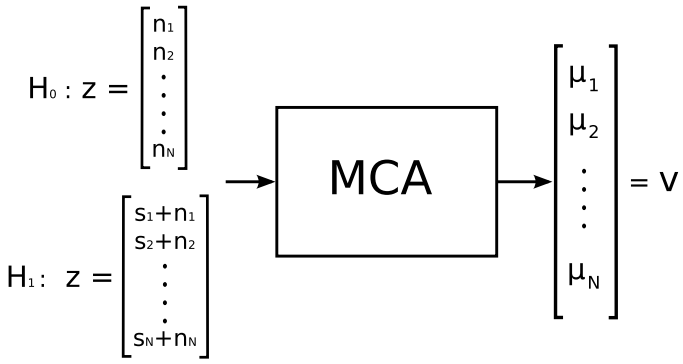


Figure 3.2: Graphical Representation of the Detection Problem

Because we have assumed realizations of noise to be uncorrelated across different channels, both composite covariance matrices will become block-diagonal under the null hypothesis, i.e.

$$R_{\mathbf{zz}_0} = D_0 = \text{diag}\left[R_{\mathbf{n}_1}, R_{\mathbf{n}_2}, \ldots, R_{\mathbf{n}_N}\right]$$

Note that the subscript notation refers to the hypothesis being considered.

Under H_1 and using the stated assumptions, the corresponding composite covariance matrices become

$$\begin{aligned} R_{\mathbf{zz}_1} &= \begin{bmatrix} R_{\mathbf{s}_{11}} + R_{\mathbf{n}_1} & . & . & . & R_{\mathbf{s}_{1N}} \\ R_{\mathbf{s}_{21}} & . & . & . & R_{\mathbf{s}_{2N}} \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ R_{\mathbf{s}_{N1}} & . & . & . & R_{\mathbf{s}_{NN}} + R_{\mathbf{n}_N} \end{bmatrix} \\ D_1 &= \text{diag}\left[R_{\mathbf{s}_{11}} + R_{\mathbf{n}_1}, \ldots, R_{\mathbf{s}_{NN}} + R_{\mathbf{n}_N}\right] \end{aligned} \tag{3.11}$$

Recalling (3.4), this leads to the following eigenvalue decomposition for the H_1 hypothesis.

$$R_{\mathbf{zz}_1} A_1 = D_1 A_1 \Lambda_1 \tag{3.12}$$

Noting that we can alternatively write the inverse of the covariance matrix under H_1 as

$$R_{\mathbf{zz}_1}^{-1} = D_1^{-H/2} P_1 \Lambda_1^{-1} P_1^H D_1^{-1/2}$$

the log-likelihood ratio can be written as

$$\begin{aligned} l(\mathbf{z}) &= \mathbf{z}^H \left(R_{\mathbf{zz}_0}^{-1} - R_{\mathbf{zz}_1}^{-1} \right) \mathbf{z} \\ &= \mathbf{z}^H \left(D_0^{-1} - D_1^{-H/2} P_1 \Lambda_1^{-1} P_1^H D_1^{-1/2} \right) \mathbf{z} \end{aligned}$$

We then remove the second-order information associated with the H_1 hypothesis from each individual channel by "whitening" with the filter $D_1^{-1/2}$ so that

$$\begin{aligned} \mathbf{z} &\rightarrow \mathbf{w} = D_1^{-1/2} \mathbf{z} \\ E_{H_0}\left[\mathbf{w}\mathbf{w}^H\right] &= D_1^{-1/2} D_0 D_1^{-H/2} = \Sigma^{-1} \\ E_{H_1}\left[\mathbf{w}\mathbf{w}^H\right] &= D_1^{-1/2} R_{\mathbf{zz}_1} D_1^{-H/2} = P_1 \Lambda_1 P_1^H \end{aligned}$$

where the matrix Σ is in some sense a *local* SNR matrix with the j^{th} diagonal block equal to

$$\Sigma_j = \left(R_{\mathbf{s}_{jj}} + R_{\mathbf{n}_j} \right)^{H/2} R_{\mathbf{n}_j}^{-1} \left(R_{\mathbf{s}_{jj}} + R_{\mathbf{n}_j} \right)^{1/2}$$

The log-likelihood ratio in this new coordinate system then becomes

$$l(\mathbf{z}) = \mathbf{w}^H \left(\Sigma - P_1 \Lambda_1^{-1} P_1^H \right) \mathbf{w}$$

Finally, we map our data into the MCA coordinate system (under H_1) through the filter P_1 so that

$$\begin{aligned} \mathbf{w} &\rightarrow \bar{\mathbf{v}} = P_1^H \mathbf{w} \\ E_{H_0}\left[\bar{\mathbf{v}}\bar{\mathbf{v}}^H\right] &= P_1^H \Sigma^{-1} P_1 \\ E_{H_1}\left[\bar{\mathbf{v}}\bar{\mathbf{v}}^H\right] &= \Lambda_1 \end{aligned}$$

where Λ_1 is a matrix with the sum of the correlations among the mapped data under H_1 along its diagonal. We can then rewrite the log-likelihood ratio as

$$l(\mathbf{z}) = \bar{\mathbf{v}}^H \left(P_1^H \Sigma P_1 - \Lambda_1^{-1} \right) \bar{\mathbf{v}} \tag{3.13}$$

where $\bar{\mathbf{v}} = \left[\sum_{j=1}^{N} v_{1,j} \ \cdots \ \sum_{j=1}^{N} v_{d,j}\right]^T$ is a vector of the sum of the MCA coordinates under H_1. Again, this is still the standard Gauss-Gauss log-likelihood ratio, but in the coordinates $P_1^H D_1^{-1/2}\mathbf{z}$. The CCA-based detector reviewed in Section 2.3 looks for coherence among our measurement and the underlying signal that composes our observation under H_1. However, the detector built here searches for coherence structure among all pairwise combinations of channels under H_1. MCA is then used to "discover" the coherence structure among the channels by solving a generalized eigenvalue problem. The amount of coherence in each MCA coordinate can then be interpreted and analyzed through the generalized eigenvalue, λ_i.

With this, it easy to see that the J-divergence can be written as

$$\begin{aligned} J &= E_{H_1}\left[l(\mathbf{z})\right] - E_{H_0}\left[l(\mathbf{z})\right] \\ &= \operatorname{tr}\left[\left(P_1^H \Sigma P_1 - \Lambda_1^{-1}\right)\Lambda_1 - \left(P_1^H \Sigma P_1 - \Lambda_1^{-1}\right) P_1^H S^{-1} P_1\right] \\ &= \operatorname{tr}\left(-2I + \Lambda_1 P_1^H \Sigma P_1 + \Lambda_1^{-1} P_1^H \Sigma^{-1} P_1\right) \\ &= \sum_{i=1}^{d}\left(-2 + \mathbf{p}_i^H\left[\lambda_i \Sigma + (\lambda_i \Sigma)^{-1}\right]\mathbf{p}_i\right) \end{aligned} \tag{3.14}$$

Therefore, we find that the divergence in the MCA coordinate system becomes decomposed in terms of the MCA generalized eigenvalue, λ_i, and also the two quadratic terms, $\mathbf{p}_i^H \Sigma \mathbf{p}_i$ and $\mathbf{p}_i^H \Sigma^{-1} \mathbf{p}_i$. The quadratic term $\mathbf{p}_i^H \Sigma \mathbf{p}_i$ in some sense gives us a scalar measurement of the sum of the local signal-to-noise ratios in the one-dimensional subspace spanned by $\mathbf{p}_i$. Thus, it appears that writing the J-divergence in this manner decomposes the information needed for detection into the coherence shared between data channels (λ_i) and the coherent information among the individual channels themselves ($\mathbf{p}_i^H \Sigma \mathbf{p}_i$).

Note that in a situation where the local signal-to-noise ratios are very small but the coherence shared between pairs of channels is significant for detection (a situation we will later motivate), we can approximate the matrix Σ with the identity matrix ($\Sigma \approx I$). Such a situation arises when the distribution of the data associated with any

particular channel is similar under both H_0 and H_1 and yet there exists a sufficient amount of cross-correlation between data channels under H_1 to perform detection. In such cases, the log-likelihood ratio can be approximated by the equation

$$l(\mathbf{z}) \approx \bar{\mathbf{v}}^H \left(I - \Lambda^{-1}\right) \bar{\mathbf{v}}$$

with an associated J-divergence

$$J \approx \sum_{i=1}^{d} \left(-2 + \lambda_i + \lambda_i^{-1}\right) \tag{3.15}$$

Therefore, in such a situation we disregard the coherent information among each individual channel and focus our attention around detecting the presence of coherence among the data channels.

3.3.1 Example

We will conclude by studying a very simple example in which we assume that our observation consists of two scalars with each scalar representing an independent channel. To be specific, we consider the detection problem

$$H_1 \quad : \quad \begin{cases} x_1 &= s_1 + n_1 \\ x_2 &= s_2 + n_2 \end{cases}$$

$$H_0 \quad : \quad \begin{cases} x_1 &= n_1 \\ x_2 &= n_2 \end{cases}$$

where $E\left[n_j n_k\right] = \delta_{j-k}\sigma_n^2$, $E\left[s_j^2\right] = \sigma_{s_j}^2$, and $E\left[s_1 s_2\right] = \rho\sigma_{s_1}\sigma_{s_2}$ for $j, k = 1, 2$. With these assumptions, it easy to see that the local SNR matrix becomes

$$\Sigma = D_1^{H/2} D_0^{-1} D_1^{1/2} = \begin{bmatrix} 1+\eta_1 & 0 \\ 0 & 1+\eta_2 \end{bmatrix}$$

where $\eta_j = \sigma_{s_j}^2/\sigma_n^2$ is the SNR for the j^{th} channel. Likewise, the coherence matrix for the H_1 hypothesis becomes

$$E = D_1^{-1/2} R_{\mathbf{z}\mathbf{z}_1} D_1^{-H/2} = \begin{bmatrix} 1 & \frac{\rho\sigma_{s_1}\sigma_{s_2}}{\sqrt{\left(\sigma_{s_1}^2+\sigma_n^2\right)\left(\sigma_{s_2}^2+\sigma_n^2\right)}} \\ \frac{\rho\sigma_{s_1}\sigma_{s_2}}{\sqrt{\left(\sigma_{s_1}^2+\sigma_n^2\right)\left(\sigma_{s_2}^2+\sigma_n^2\right)}} & 1 \end{bmatrix} = \begin{bmatrix} 1 & \xi \\ \xi & 1 \end{bmatrix}$$

where ξ represents the cross-correlation among the whitened versions of both channels. Solving the eigenvalue problem $EP_1 = \Lambda_1 P_1$ results in the sum of correlations and mapping matrices

$$\begin{aligned} \Lambda_1 &= \begin{bmatrix} 1 + \frac{\rho\sigma_{s_1}\sigma_{s_2}}{\sqrt{\left(\sigma_{s_1}^2+\sigma_n^2\right)\left(\sigma_{s_2}^2+\sigma_n^2\right)}} & 0 \\ 0 & 1 - \frac{\rho\sigma_{s_1}\sigma_{s_2}}{\sqrt{\left(\sigma_{s_1}^2+\sigma_n^2\right)\left(\sigma_{s_2}^2+\sigma_n^2\right)}} \end{bmatrix} = \begin{bmatrix} 1+\xi & 0 \\ 0 & 1-\xi \end{bmatrix} \\ P_1 &= [\mathbf{p}_1\ \mathbf{p}_2] = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \end{aligned}$$

The one observed in both eigenvalues is clearly an artifact of the unit-trace constraint and represents the sum of the auto-correlations for each channel in the mapped domain. We can also observe that the mapping matrix P becomes a 2-D, discrete Haar transform [35] which is used to map the whitened versions of the channels, i.e. $\mathbf{w} = D_1^{-1/2}\mathbf{z}$, to their MCA coordinates

$$\begin{aligned} \bar{\mathbf{v}} &= \begin{bmatrix} v_{1,1} + v_{1,2} \\ v_{2,1} + v_{2,2} \end{bmatrix} = P_1^H \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \\ &= \begin{bmatrix} 1/\sqrt{2}\left[\left(\sigma_{s_1}^2 + \sigma_n^2\right)^{-1/2} x_1 + \left(\sigma_{s_2}^2 + \sigma_n^2\right)^{-1/2} x_2\right] \\ 1/\sqrt{2}\left[\left(\sigma_{s_1}^2 + \sigma_n^2\right)^{-1/2} x_1 - \left(\sigma_{s_2}^2 + \sigma_n^2\right)^{-1/2} x_2\right] \end{bmatrix} \end{aligned}$$

Thus, in this simple case, mapping the whitened data into the MCA coordinate system can be interpreted as implementing a rudimentary analysis filter bank where the first coordinate represents the low frequency approximation to the data and the second the high frequency details. This justifies disregarding the sub-dominant coordinates as we can now see that in doing so, we are simply choosing to perform detection with

low frequency approximations of the data. As described earlier in Section 3.2, when performing detection we use the principal $r = d_1 = 1$ coordinate associated with λ_1 as $\lambda_1 > \lambda_2$. Thus, disregarding the second coordinate associated with λ_2, we find that

$$\mathbf{p}_1^H \Sigma \mathbf{p}_1 = 1 + 1/2\,(\eta_1 + \eta_2)$$

which is one plus the average of the SNRs of each channel. The log-likelihood ratio in (3.13) can then be written as

$$\begin{aligned} l\,(x_1, x_2) &= \left(\mathbf{p}_1^H \Sigma \mathbf{p}_1 - \lambda_i^{-1}\right)\left(v_{1,1} + v_{1,2}\right)^2 \\ &= \left(\frac{\frac{1}{4}\,(\eta_1 + \eta_2) + \frac{1}{2}\xi + \frac{1}{4}\xi\,(\eta_1 + \eta_2)}{1 + \xi}\right)\left[\left(\sigma_{s_1}^2 + \sigma_n^2\right)^{-1/2} x_1 + \left(\sigma_{s_2}^2 + \sigma_n^2\right) x_2\right]^2 \end{aligned}$$

Also, the J-divergence in (3.14) becomes

$$\begin{aligned} J &= -2 + \lambda_1 \mathbf{p}_1^H \Sigma \mathbf{p}_1 + \lambda_1^{-1} \mathbf{p}_1^H \Sigma^{-1} \mathbf{p}_1 \\ &= -2 + (1 + \xi) + \frac{1}{2}(1 + \xi)(\eta_1 + \eta_2) + \frac{1}{2}\frac{1}{(\eta_1 + 1)(1 + \xi)} + \frac{1}{2}\frac{1}{(\eta_2 + 1)(1 + \xi)} \end{aligned}$$

Figures 3.3 and 3.4 display plots of the (rank-one) log-likelihood ratio at two different SNRs for channel 1 namely $\eta_1 = 0$ and 20 dB, an SNR of $\eta_2 = 0$ dB for channel 2, and a correlation coefficient of $\rho = 0.8$. Note that the lines in each figure denote equilikelihood contours for different pairs (x_1, x_2). From the figures we can see that the more the SNR in the first channel begins to dominate that of the second, the more the equilikelihood contours of the detector become vertically oriented. Thus, we can see that the detector begins to disregard the second channel and focus its attention more to the observations that are brought by channel one.

We next look at the low-rank detector presented in Section 2.3 to gain some insights into the relationships and differences among the two detectors. Recall that we begin by forming the "signal-to-noise" ratio matrix

$$S = R_{\mathbf{z}\mathbf{z}_0}^{-1/2} R_{\mathbf{z}\mathbf{z}_1} R_{\mathbf{z}\mathbf{z}_0}^{-H/2} = \begin{bmatrix} 1 + \eta_1 & \frac{\rho \sigma_{s_1} \sigma_{s_2}}{\sigma_n^2} \\ \frac{\rho \sigma_{s_1} \sigma_{s_2}}{\sigma_n^2} & 1 + \eta_2 \end{bmatrix}$$

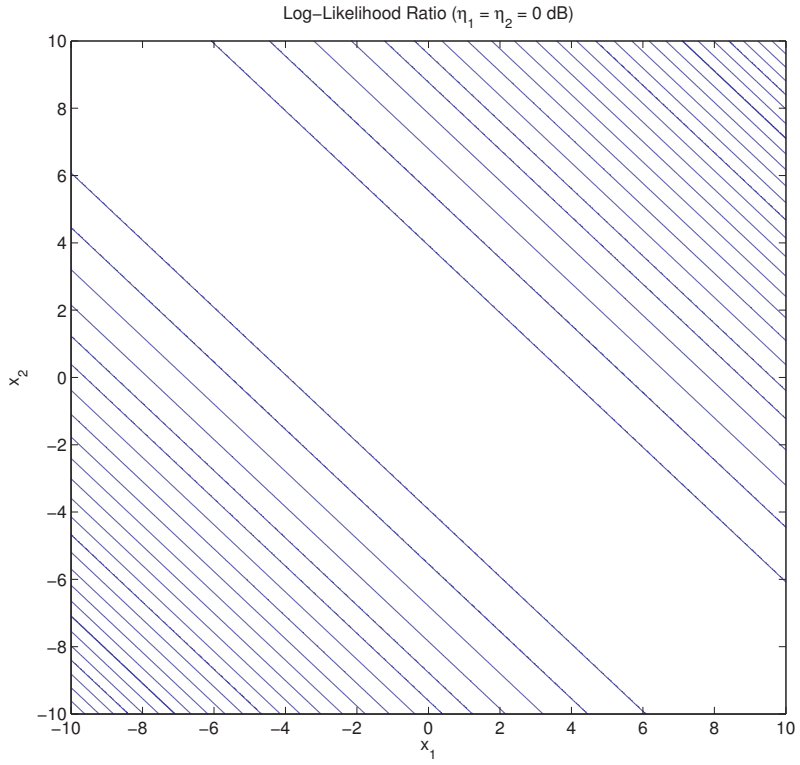


Figure 3.3: Log-Likelihood Contours at $\eta_1 = \eta_2 = 0\ dB$.

Solving the eigenvalue problem $SU = \Lambda U$ results in the per-mode SNR and mapping matrix

$$\begin{aligned} \lambda_i &= 1 + \frac{1}{2}(\eta_1 + \eta_2) \pm \frac{1}{2}\sqrt{(\eta_1 - \eta_2)^2 + 4\rho^2\eta_1\eta_2} \quad i = 1, 2 \\ U &= [\mathbf{u}_1\ \mathbf{u}_2] = \frac{1}{\sqrt{\frac{\rho^2\eta_1\eta_2}{(\lambda_1 - 1 - \eta_1)^2} + 1}} \begin{bmatrix} \frac{\rho\sigma_{s_1}\sigma_{s_2}}{\sigma_n^2(\lambda_1 - 1 - \eta_1)} & 1 \\ 1 & -\frac{\rho\sigma_{s_1}\sigma_{s_2}}{\sigma_n^2(\lambda_1 - 1 - \eta_1)} \end{bmatrix} \end{aligned}$$

Note that in the case where the SNRs of both channels are equal ($\eta_1 = \eta_2 = \eta$), we find the per-mode signal to noise ratio

$$\lambda_i = 1 + \eta \pm \frac{\rho\sigma_{s_1}\sigma_{s_2}}{\sigma_n^2}$$

and it can be shown that the mapping matrix U again becomes the 2-D Haar basis. This points out one disadvantage of the MCA-based detector for this example: the

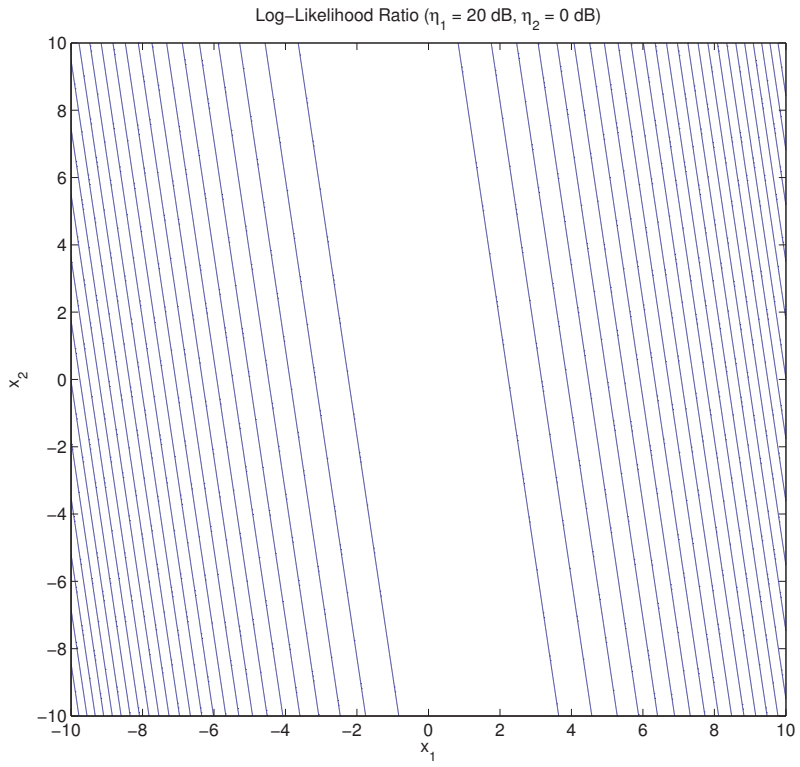


Figure 3.4: Log-Likelihood Contours at $\eta_1 = 20\ dB$, $\eta_2 = 0\ dB$.

MCA detector always represents the data in the Haar basis regardless of the characteristics of the channels. On the other hand, the detector proposed in [25] takes advantage of the characteristics of the channel to build a coordinate system more suited to the problem. The exception to this statement being when the local SNRs for each channel are equal. Because of the properties of the data mentioned earlier, we are interested in knowing what happens to either detector when there is low SNR in both channels but a sufficient amount of coherence among the two channels under H_1. Figure 3.5 displays the J-divergence for both the proposed MCA-based detector as well as that presented in [25] (denoted simply by SNR) for an equal SNR for both channels of $\eta_1 = \eta_2 = 0\ dB$. Note that both detectors are rank-one. We see that the proposed detection system exhibits poorer performance at low correlation coefficients but overtakes that of the SNR detector at correlation coefficients greater

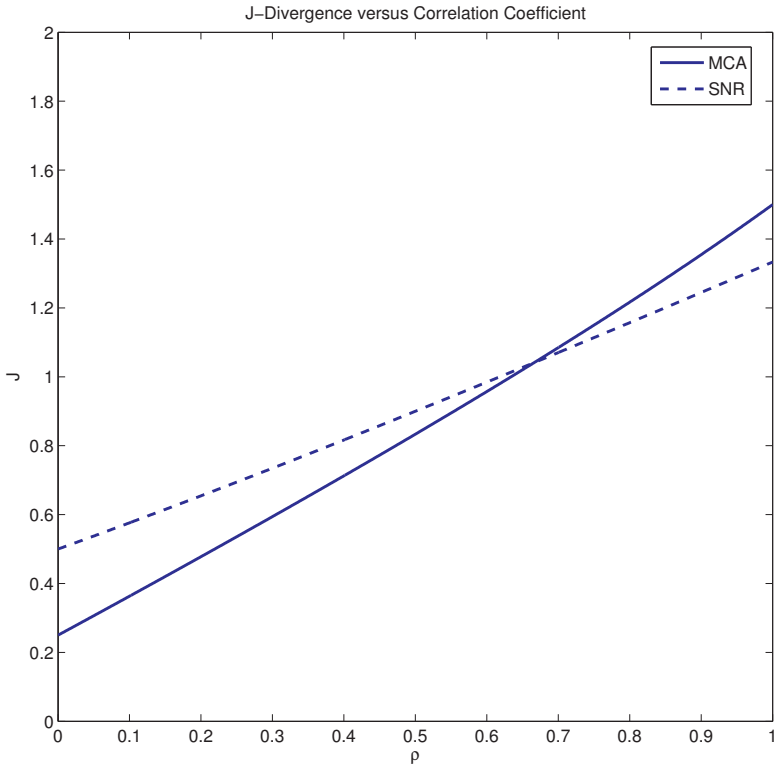


Figure 3.5: J-Divergence versus Correlation Coefficient ($\eta_1 = \eta_2 = 0\ dB$).

than approximately 0.65. Thus, we can see that this simple example seems to hint that the MCA-based detector proposed in this chapter may be better suited to conditions where the SNRs local to each individual channel are low yet there exists a large amount of coherence among the channels under hypothesis H_1.

3.4 Conclusion

In this chapter, we reviewed the MCA framework which can be seen as a natural extension of CCA to more than two channels. Similar to two-channel CCA, the objective of MCA is to discover the coherence among $N \geq 2$ channels of data by searching for one-dimensional mapping vectors that simultaneously maximize the cross correlation among all channels. One solution that has been proposed [4] is to maximize the sum of the cross-correlations among all pairwise combinations of channels subject

to the constraint that the sum of the auto-correlations equals one. This leads to a generalized eigenvalue problem involving the composite covariance matrix. Properties of the coordinates produced by MCA are then reviewed which share many similarities to the canonical variates produced by CCA, namely that the sum of the auto-covariance matrices of the mapped variates equals identity and the double sum of the cross-covariance matrices of the mapped variates equals a matrix with the MCA correlations along its diagonal.

The two-channel CCA-based detector presented in [1] looks for coherence among two sonar images. To extend this idea to multi-sonar detection, one must define a coordinate system that finds the linear relationships among every pair of sensory data channels. MCA is suited for this problem as it begins by removing the auto-correlation contributions from each channel via a whitening procedure and then finds mapping vectors that discover the coherence structure among the channels. We then presented a multi-channel signal-plus-noise model that fits well with the problem at hand. The data is then represented in the MCA coordinate system under the H_1 hypothesis and subsequently applied to a quadratic detector built for this multi-channel coordinate system. In this case, the log-likelihood ratio involves the sum of the correlations matrix, Λ_1, the orthonormal mapping matrix, P_1, and the *local* signal-to-noise ratio matrix, Σ, a block-diagonal matrix composed of the SNR matrices for each channel. The J-divergence for this MCA-based detector again becomes decomposed in terms of the same matrices. With the representations for the log-likelihood and J-divergence given, it seems that the detector decomposes the information needed for detection into the coherence shared between data channels under H_1, namely λ_i in 3.12, and the SNR information among the individual channels themselves given by $\mathbf{p}_i^H \Sigma \mathbf{p}_i$ in 3.14.

With the assumption of low SNR in each individual data channel, we approximated the log-likelihood ratio and J-divergence in terms of a detector that only takes advantage of the sum of the cross-correlations of the data in the mapped domain

under the H_1 hypothesis. This results in a similar expression as in [25]. We then considered a simple two-channel detection problem to gain some insight into the inner workings of both the proposed MCA-based detector and that presented in [25]. In this simple case, we found that the MCA decomposition boils down to representing the data in a 2-D Haar basis regardless of the characteristics of the channel. On the other hand, the SNR detector finds mapping vectors that depend on the SNRs in each channel and the correlation coefficient among them under H_1. However, in situations where there is low SNR in both channels and yet there exists high coherence among them under the H_1 hypothesis, we showed through this simple example that the proposed MCA-based detector is better suited, comparing to the detector of [25], for such situations as it yields higher J-divergence for a sufficiently large correlation coefficient.

CHAPTER 4

DATA DESCRIPTION AND EXPERIMENTAL RESULTS

4.1 Introduction

Detection of underwater objects in sonar imagery is a complicated problem 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 object. Current detection methods normally use a single sonar image to detect potential targets [36], [37]. However, detection based off one image can lead to unacceptable results as the information is limited to the field of view of only one sonar which could be dependent on the relative position of the sonar platform to the target. This motivates the use of multiple disparate sonar platforms, where disparateness could be in frequency, resolution, location, etc., to better capture the target characteristics. Using multiple disparate sonar types allows one to use a high resolution sonar with good target definition and the clutter suppressing abilities of a low resolution sonar co-registered over the same region and build detection systems that take advantage of all the information available at once to make a detection call.

In this chapter we will begin by describing the preprocessing methods used prior to the MCA-based detector and then present the results on three different sonar data sets. In each case, we are given multiple (≥ 2) sonar images that can be disparate in different ways. We then partition the sonar images into Regions of Interest (ROIs) of different sizes. Each ROI is then partitioned into blocks and used to form a composite

ensemble data matrix from which the necessary MCA mapping vectors and sum of correlation matrices are extracted to build the log-likelihood ratio. Each composite realization is then applied to the log-likelihood ratio to form a detection decision for that set of blocks. If a majority of the blocks pass the likelihood ratio test, a call to the H_1 hypothesis is made for that set of ROIs. Our detection hypothesis for this multiple sonar image detection scheme is that the presence of target structure in all or a partial subset of the ROIs will result in higher level of coherence versus when the ROIs contain background clutter alone.

Test results on three different data sets are presented. All data sets were provided by the Naval Surface Warfare Center (NSWC) - Panama City, Florida. The first data set contains one high frequency (HF) and three broadband (BB) sonar imagery coregistered over the sea floor. This data set is used to examine different combinations of these sonar images, construct three different multi-channel detectors with the HF image and one more BB images, and compare detection performance as a function of the number of disparate sonar images used. The second data set consists of only one HF and one BB sonar imagery, again coregistered over the sea floor. This data set contains both target and non-target objects (lobster traps) which can give us some intuition of the discriminatory power of the extracted multi-channel features for target/non-target classification. Finally, we present sensitivity analysis results on a data set consisting of simulated target objects in simulated background generated for various choices of SNR, target aspect angle, image resolution, target shape, etc. The hope of the sensitivity analysis is to give one some idea of the proposed system's response to different variables that can be expected in multiple disparate sonar detection problems.

The outline of this chapter is as follows. Section 4.2 describes the preprocessing method used to prepare the data for the MCA-based detector. Sections 4.3 and 4.4 give a detailed description of the properties of the multiple sonar image data sets

and presents the results of the proposed MCA-based detection system when applied to each of the two real sonar data sets. Section 4.5 describes the simulated data set and presents sensitivity results with respect to SNR, target type, and aspect angle separation. Finally, concluding remarks are made in Section 4.6.

4.2 Sonar Image Description and Pre-Processing

In order to better understand the difficulty of underwater object detection from sonar imagery it is important to understand the formation of a target signature in a sonar image. Figure 4.1 shows how the signature of a target is formed in a sonar image. In this figure, region A-B corresponds to the *highlight* or a strong sonar return off of the object, region B-C is known as the *dead zone* where no sonar return can occur due to the return being blocked by the object, and region C-D is known as the *shadow*. The size of each of these regions greatly depends on the type of target, range from the sonar, height of the target, aspect, and grazing angle of the sonar with respect to the object. More specifically, the signature of a particular target can vary greatly as a function of range from the sonar. If the target is further out in range from the sonar, the shadow region (C-D) will become longer. Additionally, the overall signature becomes less defined with less definition between the highlight, dead zone, and shadow regions. This creates a particular problem for most detection methods which rely on detecting the specific structure of the target, i.e. matched filter-based approaches.

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 [38], [39]. Each impinging sound wave on the receiver array elements of the sonar is converted to magnitude and phase. The delay and sum beamforming algorithm [40] 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

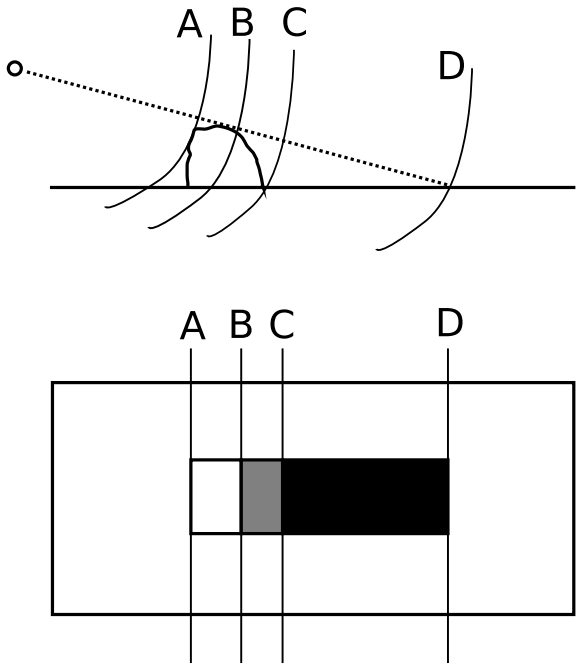


Figure 4.1: Formation of a Target in a Sonar Image.

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 [41]. 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.

When processing the images in the data sets for the MCA-based detector, each set of N images is first partitioned into coregistered ROIs with 50% overlap in both the vertical and horizontal directions. 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 target structure. Based on the average target size, ROIs pertaining to HF images are chosen to be 72 pixels tall by 112 pixels wide. Because of differences in beamwidth in HF and BB sonar leading to disparateness in image resolution, the ROIs pertaining to BB images are not the same

size and are chosen to be 24 pixels tall by 224 pixels tall. This choice of ROI sizes ensures correspondance among the HF and BB images, i.e. whatever the dimension of the HF ROI, the BB ROI must be three times smaller in the vertical dimension and twice as large in the horizontal dimension.

Once the set of N coregistered ROIs has been extracted from each of the N disparate sonar images, each ROI is partitioned into non-overlapping blocks of size 6×4 for HF images and 2×8 for BB. Again, the difference in block size for each sonar type is a byproduct of their disparateness in resolution. Corresponding blocks in the ROIs are then reshaped into vectors and concatenated to form the composite observation vector $\mathbf{z}$ defined in Section 3.2. This is done until all blocks in the ROIs have been accounted for. An ensemble set is then formed from all 336 blocks $\left(\frac{72\times112}{6\times4} = \frac{24\times224}{2\times8} = 336\right)$ in each ROI and is subsequently used to form an estimate of the composite covariance matrix $R_{\mathbf{zz}}$. This composite covariance matrix is then decomposed via MCA and the sum of correlations and MCA mapping matrices, Λ and A respectively, are extracted using (3.12) to form the log-likelihood ratio test statistic. Each observation vector from the ensemble set is then applied to the log-likelihood ratio in (3.13) and compared to a threshold to form a decision for that set of N blocks. If 50% or more of the blocks within an ROI set pass the log-likelihood ratio test, it is concluded that that ROI contains a target. Figure 4.2 gives a graphical overview of the processing steps just described. From an implementation point of view, forming detection decisions on individual blocks of the ROI is desirable in several aspects. First of all, partitioning the ROI into smaller blocks yields observation vectors that sit in a low dimensional space. This, in turn facilitates the use of multiple sonar images as the smaller the block sizes, the more sonar images we can add to the detection problem without processing an extremely high dimensional composite observation vector. Second, partitioning the ROI into blocks gives us multiple independent observations of the information contained in each ROI allowing us

to make detection decisions with much higher confidence.

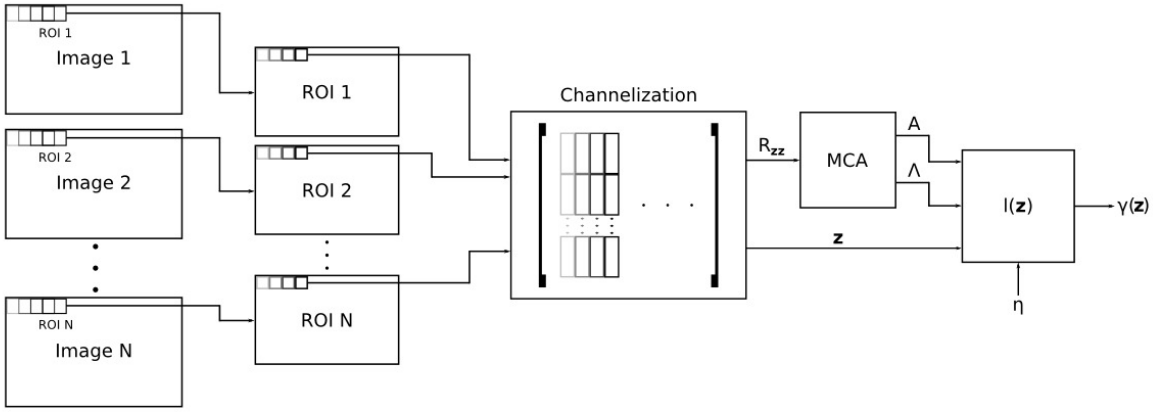


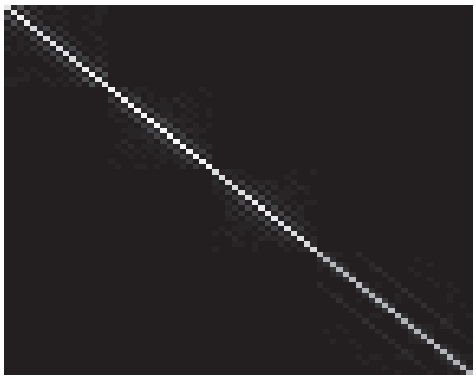
Figure 4.2: Mutiple Sonar Detection System.

As previously mentioned at the end of Section 3.3, we can approximate the MCA-based detector by a quadratic detector that only takes advantage of the sum of correlation, Λ_1, and MCA mapping matrices, P_1, in situations where the local SNRs of each channel are very small. To investigate this, we extract both target and background features from the multiple sonar image data set corresponding to one HF image and three BB images. The block-diagonal covariance matrices D_1 and D_0 are estimated from these features and used to form the matrix $\Sigma = D_1^{H/2} D_0^{-1} D_1^{1/2}$ in (3.13). Figure 4.3(a) displays an image of this matrix and Figure 4.3(b) displays the distribution of its eigenvalues. Thus, we conclude that, for all intents and purposes, the local SNR matrix is diagonal with diagonal elements equal to some small perturbation from unity. As a result, when implementing the MCA detector we approximate the log-likelihood ratio with the quadratic function

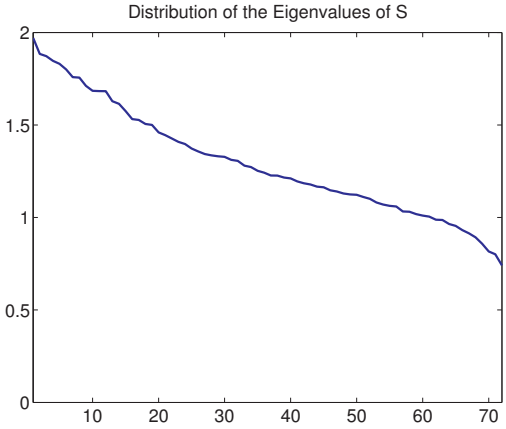
$$l(\mathbf{z}) = \mathbf{z}^H A \left(I - \Lambda^{-1} \right) A^H \mathbf{z} \tag{4.1}$$

It was previously shown in [42] and [43] that this approximation can effectively be used when performing detection with multiple sonar images.

The reasons for low SNR could be many. One explanation may be in the structure of the target. Small highlight and (zero pixel) shadow structures will undoubtedly



(a) Image of Σ



(b) Distribution of the Eigenvalues of Σ

Figure 4.3: Local Signal-to-Noise Ratio Matrix Σ

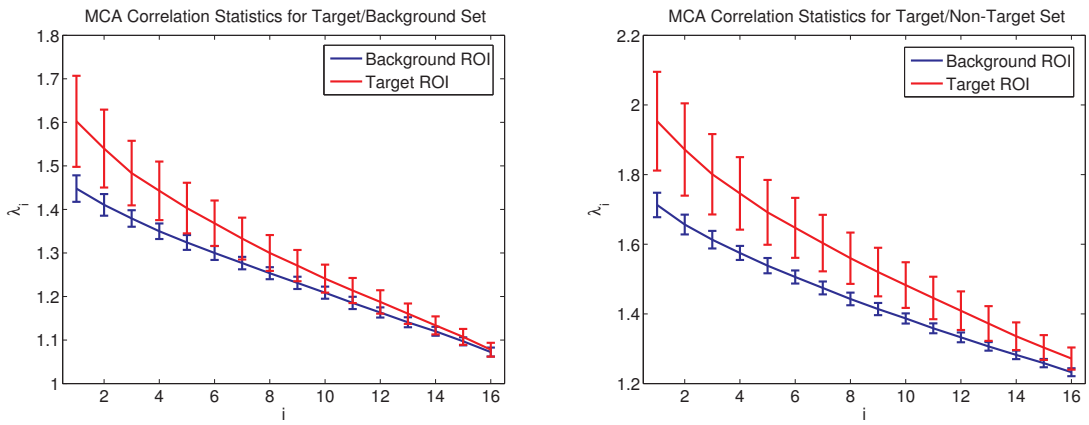
lead to low signal power causing poor SNR. Another explanation for poor SNR may be in both the structure of the target and the way the ROIs are processed. Targets are generally composed of two distinct structures, namely highlight with high variance and shadow with low variance. Therefore, partitioning the ROI into equally sized blocks will result in observations with completely different statistical properties. That is, some with high variance, some with low variance, and others that will obviously be somewhere in between. We then stochastically average over all these realizations, blending the information together to form a second-order statistical characterization

that does not look dissimilar to that associated with background, thus leading to low SNR.

4.3 Multiple Sonar Detection Results

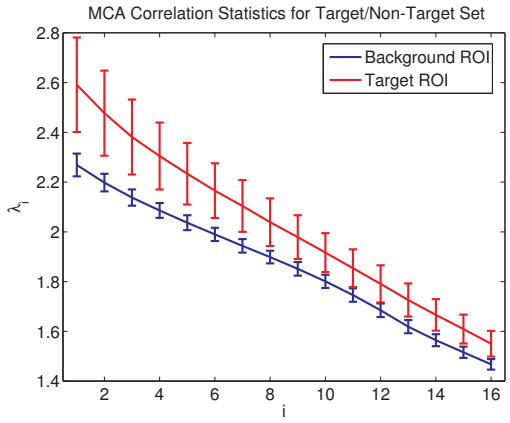
The MCA-based coherence detector is first applied to a four-sonar data set consisting of one HF high-resolution side-scan sonar image as well as three BB sonar images. Each image is complex-valued and is a direct result of the beamforming methods described earlier in Section 4.2. The image database used in this study contains 59 co-registered sonar images with each image consisting of both port and starboard-side images. The database contains 53 targets with some images containing more than one target. Because the HF sonar provides higher spatial resolution and better ability to capture target details and characteristics while the BB sonar offers much better clutter suppression ability with lower spatial resolution, detectors were run using HF images along with one or more of the BB sonars to ensure a high probability of detection with a low false alarm rate. Three different cases were implemented, a two-channel detector with the HF sonar along with one of three BB sonar images (referred to as HF-BB_1), a three-channel detector with the HF sonar, the same BB sonar, and another different BB sonar (referred to as HF-BB_1-BB_2), and finally a four-channel with the HF sonarr and all three BB sonar images (referred to as HF-BB_1-BB_2-BB_3). The goal of this study is to determine the impact different combinations and numbers of HF and BB sonar systems have on the detection performance and establish the point of diminishing returns.

To show the separability of the principal multi-channel correlations between ROIs that contain targets immersed in background and those that solely contain background, a test was conducted on the entire set of 53 target ROI's and a same size randomly selected set of ROI's containing only background clutter. Note that as a result of the 50% overlap in ROI formation, there are essentially 4 ROIs that contain



(a) HF-BB_1 Detector

(b) HF-BB_1-BB_2 Detector



(c) HF-BB_1-BB_2-BB_3 Detector

Figure 4.4: Statistics of the MCA Correlations for all Detection Systems.

one target leading to a total of 212 target ROIs and 212 randomly selected background ROIs. Figures 4.4(a)-(c) exhibit plots of the mean and standard deviation of the dominant 16 multi-channel correlations, $\lambda_i, i = 1, \ \ldots, 16$, of ROIs containing targets and those containing only background for the HF-BB_1, HF-BB_1-BB_2, and HF-BB_1-BB_2-BB_3 detectors, respectively. Mean values for each λ_i are shown by the solid line whereas the length of the bar denotes its corresponding standard deviation. As can be seen, there is suitable separation among the principal correlation values pertaining to targets versus those pertaining to background alone as there is a significant difference in mean values. We can also observe that the more channels

included for detection the higher the MCA correlation values. However, this does not necessarily correspond to a subsequent increase in the separability among target and background MCA correlation values. Next, for the detection process, the log-likelihood ratio expression in (4.1) was found for each block within an ROI set. A detection score was then created based on the percent of the log-likelihood measurements within an ROI set that fall above the detection threshold, η. A detection score of $\geq 50\%$ signifies the presence of a target within that set of ROI's. Using the entire set of 212 target ROIs and an equally sized set of background ROIs, an optimum threshold was experimentally determined to be 10.2 for all three detection cases.

All three detection systems are then implemented on the entire NSWC multi-sonar imagery data set using the predetermined threshold mentioned above. Table 4.1 lists the results of all three detectors. As one can see, the two-channel (HF-BB_1) detector performs marginally well with 51 out of 53 targets being detected and less than 8 false alarms per image. The three (HF-BB_1-BB_2) and four-channel (HF-BB_1-BB_2-BB_3) detectors perform better as they both detect all but one of the targets while still maintaining a low number of false alarms.

Table 4.1: Multi-Platform Detection Results

Detector	Targets Detected (Out of 53 Targets)	Average False Detections per Image
HF-BB_1	51	7.48
HF-BB_1-BB_2	52	8.93
HF-BB_1-BB_2-BB_3	52	9.32

The ROC curves for all three detectors are presented in Figure 4.5. Again, we can see that the three-channel detector provided an increase in performance over that of the two-channel detector as the two-channel detector exhibits $P_d = 96\%$ and $P_{fa} = 4\%$ at the knee point of the ROC curve (where $P_d + P_{fa} = 1$) whereas that of the three-channel detector gives $P_d = 98\%$ and $P_{fa} = 2\%$. However, the performance of the four-channel detector actually decreases with $P_d = 96\%$ and $P_{fa} =$

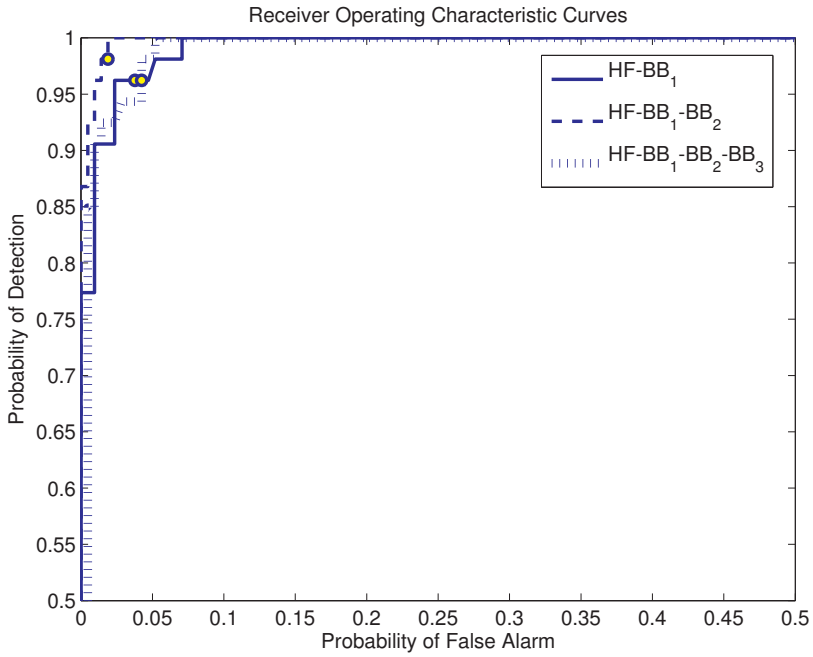


Figure 4.5: ROC Curves for all Three Detectors.

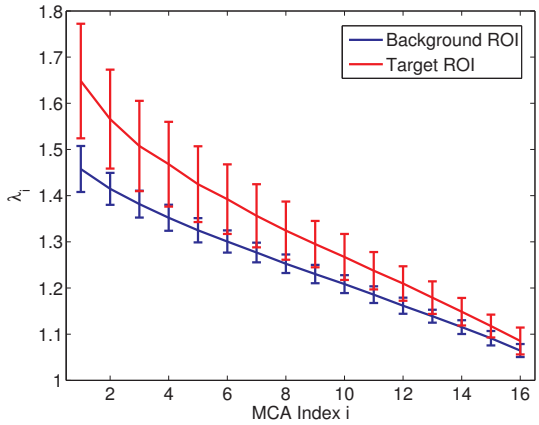
4% at the knee point of the ROC curve. This decrease in detection performance when increasing the number of channels from three to four could possibly suggest that three sonar images is the point of diminishing return for this data set as the BB_3 sonar did not bring any new pertinent information of the targets and actually worsened the performance. This could be due to the fact that increasing the number of BB sonars that essentially contain similar target information smears the overall coherence as the correlations become less representative and more deficient. However, because of the small number of targets present within this data set, it is hard to say with any confidence whether this is the actual underlying response of the detector or not. Both the HF-BB_1-BB_2 and HF-BB_1-BB_2-BB_3 detectors missed the same target. However, the HF-BB_1 detector missed two other targets, neither of which are the same target as that missed by the three and four channel detectors. The targets missed by these detectors were faint in signature and hard to visually discern in all sonar images hence leading to low coherence and subsequent misdetection. Overall, all the

detection systems tested performed extremely well given that the detection threshold was formed from 212 ROIs corresponding to only 53 targets and 212 background ROIs.

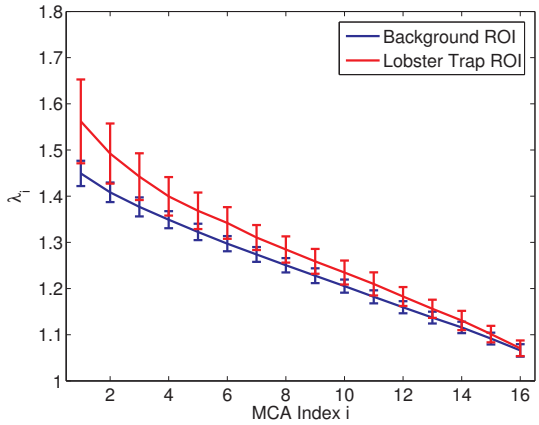
4.4 Dual-Sonar Detection Results

The MCA-based coherence detector is then applied to a dual-platform sonar data set consisting of one HF high-resolution side-scan sonar image as well as one BB sonar image. Each image is real-valued, envelope data which is the result of quantizing the amplitude (magnitude) of the complex-valued images resulting from the beamforming methods described earlier in Section 4.2. The image database used in this study contains over 1200 co-registered sonar images with each image consisting of both port and starboard-side images. The database contains 99 objects of interest, 49 target and 50 lobster trap objects, with some images containing more than one object. For this data set, only one case is studied: a two-channel detector with the HF sensor along with the BB sensor. The difficulties of the data set, compared to that in Section 4.3, are the inclusion of man-made objects (lobster traps) which can undoubtedly increase the false alarm rate when interested in detecting target objects alone. However, note that for this study, we are interested in detecting both targets and lobster traps. Another difficulty stems from taking the magnitude of the data and the inherent Gaussian assumptions that are made for the detection methods given in this thesis. Even if the data were truly Gaussian to begin with, it is a known fact [32] that real and imaginary *iid* Gaussian random variables result in a Rayleigh distributed magnitude leading to a heavy-tailed distribution that deviates from Gaussianity assumptions. Not to mention, performing detection solely with the magnitude completely disregards the information carried in the phase of the images.

Again, to show the separability of the principal multi-channel correlations between ROIs that contain objects of interest immersed in background and those that



(a) Target Objects



(b) Lobster Trap Objects

Figure 4.6: Statistics of MCA Correlations for Sample Target/Background Set.

solely contain background, a test was conducted on the entire target and lobster trap set ROIs and a same size randomly selected set of ROIs containing only background clutter. Again, because of the 50% overlap in ROI formation, this results in 196 ROIs corresponding to targets, 200 ROIs corresponding to lobster traps, and a total of 396 ROIs corresponding to background. Figures 4.6(a) and (b) exhibit plots of the mean (denoted by the solid line in each figure) and standard deviation (denoted by the length of the bar for each λ_i) of the dominant 16 multi-channel correlations of ROI's

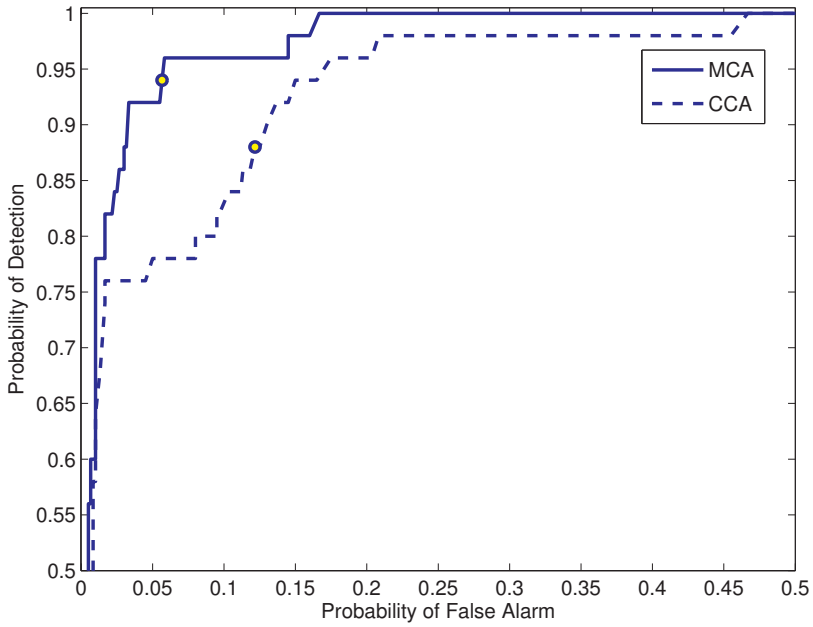


Figure 4.7: ROC Curve of Dual-Sonar "Training" Set

containing targets and lobster traps, respectively, versus those solely containing background for this dual-image detection problem. We can see that Figure 4.6(a) shares many similarities to the statistics of the MCA correlations shown in Figure 4.4(a) for the two-channel detector looked at previously though they correspond to two completely different data sets. We also see a noticeable difference among the statistics of the MCA correlations pertaining to targets and lobster traps as the separation among target and background features seems to be larger than that among lobster trap and background features. These figures seem to suggest that the MCA correlations may provide a useful set of features for discriminating among target and man-made objects for the purposes of classification. However, this is merely an observation and will not be discussed further in this work. From the entire dual-sonar data set, a partial subset of images containing 50 objects of interest (25 targets, 25 lobster traps) corresponding to 200 ROIs in total is extracted. This is done to observe the response of the detection system to a threshold that is determined from one set of objects and

is then tested on a completely novel set of objects. Using the 50 objects of interest (200 ROIs) and a same size set of background ROIs, a threshold of $\eta = 0.5212$ was experimentally chosen based on the knee-point of the ROC curve generated from this "training" set. The ROC curve is displayed in Figure 4.7 and exhibits $P_d = 94\%$ and $P_{fa} = 6\%$ at the knee-point. Also shown in the figure is the ROC curve associated with the CCA-based detection system proposed in [1] and is generated from the same set of target, lobster trap, and background ROIs. This detector exhibits $P_d = 88\%$ and $P_{fa} = 12\%$ at the knee-point of the ROC curve.

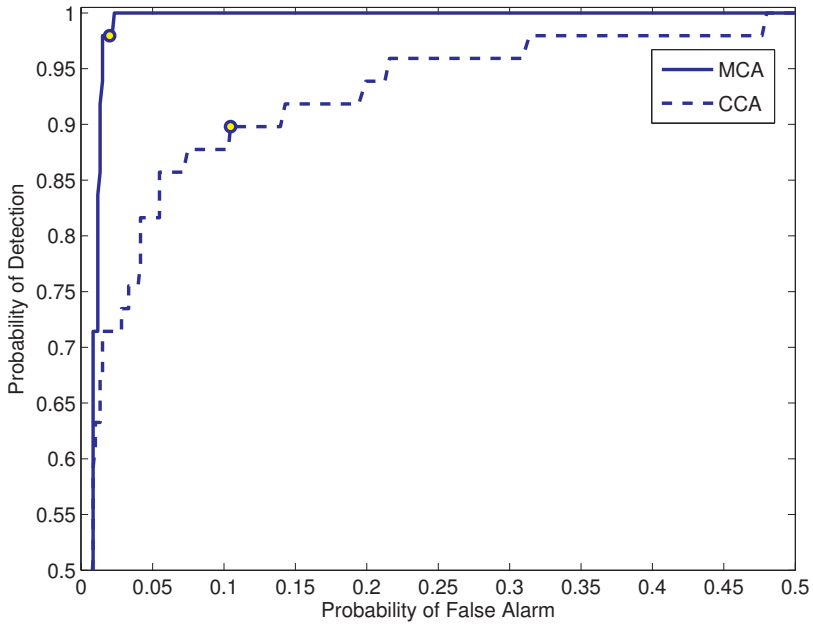


Figure 4.8: ROC Curve of Dual-Sonar Detection System.

The dual-sonar detection system is then implemented on the remaining images containing 49 objects of interest (24 targets, 50 lobster traps) using the predetermined threshold mentioned above. Recalling that there are 4 ROIs for every object of interest, there are a total of 196 ROIs corresponding to targets and lobster traps. This dual-sonar detection system detects all 49 targets with an average of 7 false alarms per image. Based on all 196 ROIs corresponding to 49 objects of interest and

a set of 196 randomly selected background ROIs, a ROC curve is generated for this detection system and is displayed in Figure 4.8. The detector exhibits $P_d = 98\%$ and $P_{fa} = 2\%$ at the knee-point of the ROC curve. Again, also shown in the figure is the ROC curve of the CCA-based detection method of [1] generated from the same set of target, lobster trap, and background ROIs. We can see that the proposed MCA-based detector provides substantial improvement over the CCA-based method which exhibits $P_d = 90\%$ and $P_{fa} = 10\%$ at the knee-point of its ROC curve.

4.5 Sensitivity Analysis Results

In this section, we apply the N-channel coherence-based detection method to a data set consisting of synthetically generated sonar images (snippets) of both targets and non-targets of different geometrical shapes embedded in synthetically generated background. The sonar snippets were generated with different resolutions, SNR values, range, and aspect angles mimicking different realistic operating conditions. For this study, two different resolutions, namely $1in$ and $3in$, were considered. Additionally, SNR ranged from $0dB$ to $15dB$ in increments of $3dB$, range values spanned from $10m$ to $120m$ in increments of $1m$, and aspect angle ranged anywhere from 0° to 360° in increments of 1°. A subset of 1610 snippets corresponding to targets was used to represent the H_1 hypothesis while background snippets were used to represent the H_0 hypothesis. Thus, all non-target snippets were excluded. The processing involved for this data set is the same as that depicted in Figure 4.2 with the exception of ROI partitioning of the images. For this data set, there are no images and each sonar snippet plays the role of an ROI. The subset of target snippets (1610 snippets) was further partitioned into 3 different parts forming 138 cone-shape targets, 736 cylinder-shape targets, and 736 trapezoidal-shape targets. When performing detection, each of the N images is partitioned into blocks of size dependent on the resolution.

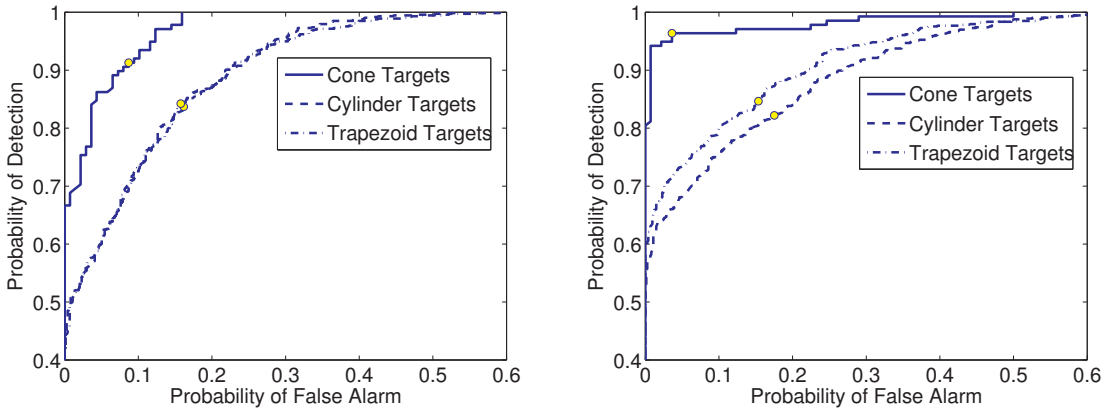
4.5.1 Sensitivity Analysis for Dual Resolution Disparate Detection

In some disparate detection applications, each platform may carry multiple sensing systems with different spatial and spectral characteristics in order to highlight different attributes of the target. To simulate such a situation, a two-channel detection problem was constructed where each channel consisted of snippets of targets of the same type at the same range, aspect angle, and SNR. However, the two channels differed in resolution, i.e. one snippet-image of high resolution ($1in$) and the other of a lower resolution ($3in$). When performing detection, a 4×4 block size was used for the high resolution images and a 2×1 block size for those of lower resolution. This setup was then run for all 1610 images at various ranges and aspect angles and the results partitioned on the basis of target type and SNR.

Table 4.2: Probability of Detection (%) vs. SNR

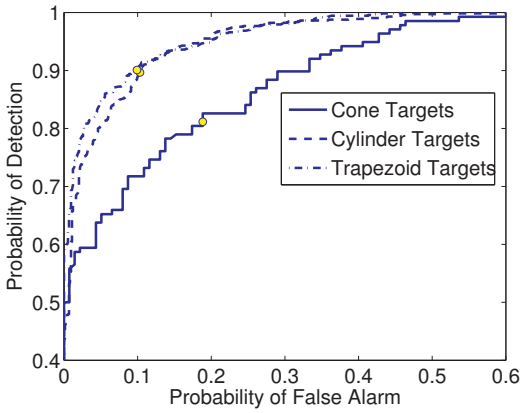
Target Type	$0dB$	$3dB$	$6dB$	$9dB$	$12dB$	$15dB$
Cone	91.30	94.93	96.38	89.86	81.16	87.68
Cylinder	83.70	85.19	82.20	85.33	89.67	94.02
Trapezoid	84.24	84.51	84.65	85.73	90.08	93.75

Figures 4.10(a)-(c) display the ROC curves for conical, cylindrical, and trapezoidal shape targets, respectively, at three different SNR values of $3dB$, $9dB$, and $15dB$. Figures 4.9(a)-(c), on the other hand, display sensitivity to target types for SNR values of $0dB$, $6dB$, and $12dB$, respectively. Table 4.2 gives the probability of detection at the knee point of the ROC for all target types and SNR values. As can be observed from the results in Table 4.2, it is apparent that for cylindrical and trapezoidal targets the detection performance generally improves as a function of SNR as one would expect. However, the performance of the detector for the cone targets does not follow the same behavior. This may be attributed, in part, to the fact that only a small number of cone targets were available for this study. Another explanation for this observation may be due to the fact that the gray-scale resolution (8-bit) of the images



(a) $0dB$ SNR

(b) $6dB$ SNR



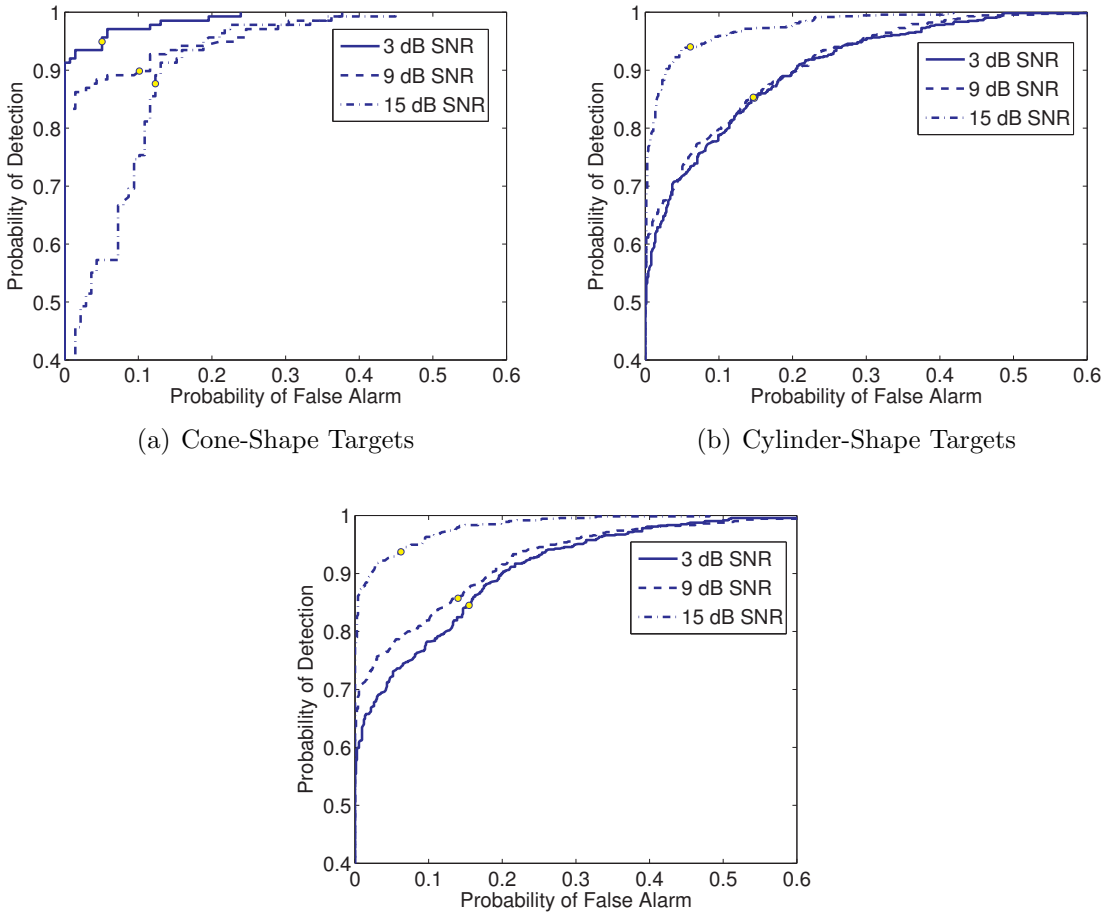
(c) $12dB$ SNR

Figure 4.9: Detector Performance for Different SNR vs. Target Type.

was insufficient to capture the large sonar returns from cone type targets, leading to saturation and clipping of the highlight.

4.5.2 Sensitivity Analysis for Dual Aspect Angle Separation Disparate Detection

Another question that may arise in disparate detection problems is that of sensor locations. That is, for different types of targets how does the detection performance change as a function of disparities in location of the sensor platforms. This clearly relates to the target's aspect/orientation with respect to the sensor platform as well as the range of the platform from the target. To determine the answer to this question,



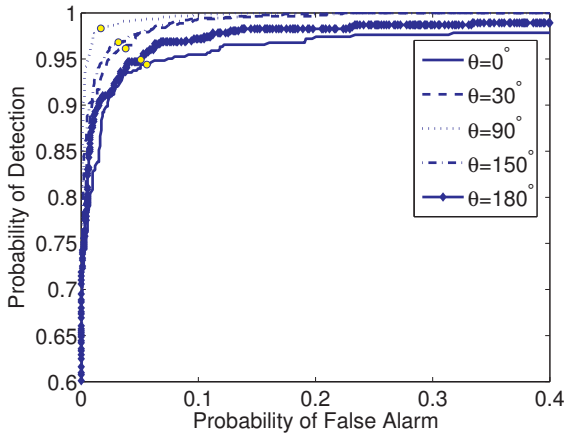
(a) Cone-Shape Targets

(b) Cylinder-Shape Targets

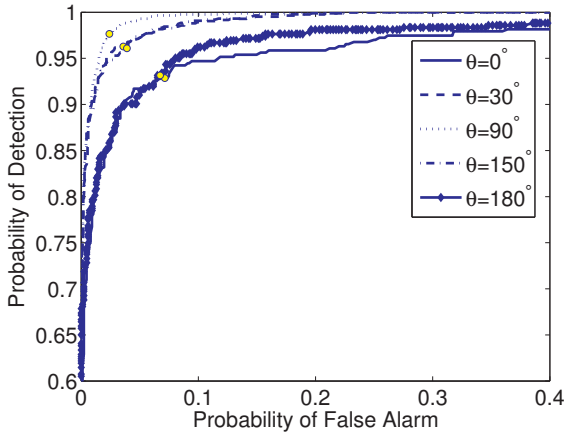
(c) Trapezoid-Shape Targets

Figure 4.10: Detector Performance for Different Target Types vs. SNR.

a study is carried out where both channels consisted of images of the same resolution $(1in)$, at ranges within $\pm 1m$ of one another, and at an identical SNR of $9dB$, while the disparateness was with respect to aspect angle separation. More specifically, the two channels correspond to sonar snippet-images from the same target at two aspect angles with separation angle θ such that if ϕ_1 and ϕ_2 represent the aspect angles associated with their respective image then pairs of images were chosen such that $|\phi_1 - \phi_2| \in [\theta - \delta, \theta + \delta]$, where δ represents the perturbation from the separation angle, θ, due to non-uniform motion of the vehicle. Here, the value of δ was chosen to be $10°$.



(a) Cylinder-Shape Targets



(b) Trapezoid-Shape Targets

Figure 4.11: Detector Performance vs. Aspect Angle Separation.

The aspect angle separation θ was then varied from 0° to 180° in increments of 30° and its affect on the performance of the detector was studied. Figures 4.11(a) and (b) display the ROC curves for cylindrical and trapezoidal targets, respectively, for several values of θ. Note that all images in the database corresponding to cone targets were generated at an angle of 0° and thus excluded from this study. Table 4.3 gives the probability of detection at the knee point of the ROC versus aspect angle separation, θ, and target type. From both Figures 4.11(a) and (b) and the results in Table 4.3, it can be concluded that the performance of the detector is fairly robust to

disparateness in aspect angle separation as the probability of detection at the knee point of the ROC never falls below 92%.

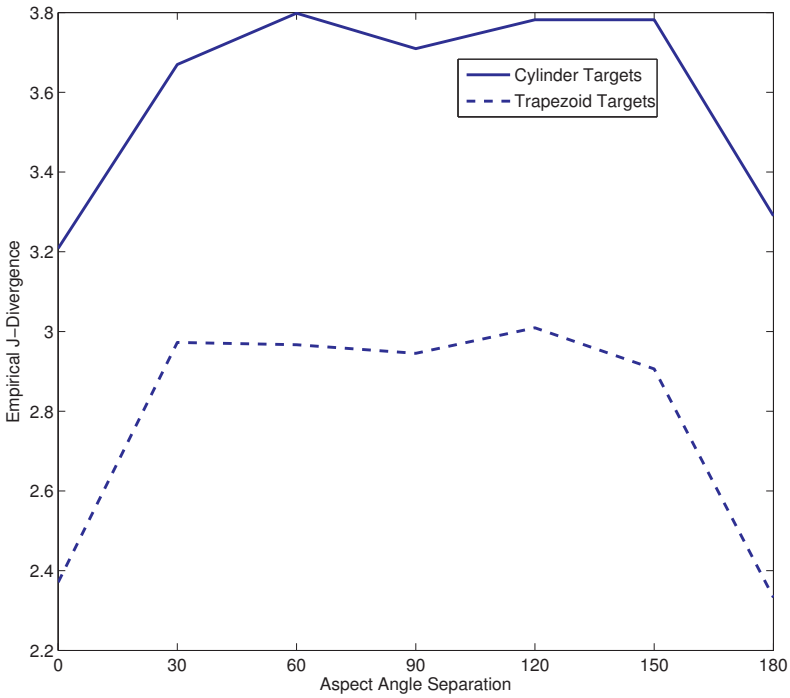


Figure 4.12: J-Divergence vs. Aspect Angle Separation.

Table 4.3: Probability of Detection (%) vs. Aspect Angle Separation (θ)

Target Type	0°	30°	60°	90°	120°	150°	180°
Cylinder	94.40	96.14	97.83	98.33	97.47	96.82	94.92
Trapezoid	92.86	96.29	97.21	97.66	97.57	96.09	93.14

Figure 4.12 displays the plots of empirical J-divergence as a function of aspect angle separation, θ, for both cylindrical and trapezoidal targets. Both curves were generated by empirically estimating the difference in means of the log-likelihood ratio among target (H_1) and noise (H_0) snippets and averaging over all such pairs of images that match the criteria explained previously (i.e. same SNR, same resolution, ranges within $1m$, aspect separation within some range of a particular angle). The results in Figure 4.12 match what was observed in Figures 4.11(a) and (b) as the performance

of the detector seems to improve as the separation in aspect angle deviates from 0° and 180° and approaches an aspect angle separation of $\theta = 90°$. Again we can draw the same conclusion that the detector is fairly robust to separation in aspect angles as the difference between maximum and minimum J-divergence never grows larger than approximately 0.6. Additionally, the J-divergence values for cylindrical targets are higher than those of trapezoidal targets which may be attributed to the large-pixel highlight characteristics of cylindrically shaped targets. This result is also evident in the ROC plots of Figures 4.11(a) and (b).

4.6 Conclusion

In this chapter, the proposed multiple sonar image detection system is applied to two real data sets consisting of one HF and one to three BB sonar images. The images in both data sets are the result of the k-space beamforming algorithm. The first data set consists of complex-valued images and are a direct result of this beamforming algorithm. The second data set consists of real-valued envelope data which the quantized version of the magnitude of the raw complex-valued images from the beamforming algorithm. Comparing the images of the second data set to that of the first, performing detection with the magnitude of the data presents difficulties as it more than likely results in a deviation from the Gaussian assumptions made throughout this work and disregards the phase information present in each of the images. To take advantage of the high target definition capabilities of HF sonar and the clutter suppression ability of BB sonar, detection cases were always constructed with one HF and at least one BB sonar image. A sensitivity analysis is then conducted using a data set consisting of snippets of simulated target and non-target shapes embedded in synthetically generated background. Snippets are generated at 8-bit gray-scale resolution with different target and non-target types and at various SNR values, image resolutions, ranges, aspect angles, etc.

Using the first data set consisting of one HF and three BB sonar images, three detection systems are designed with the HF image along with different combinations of the BB images to investigate the detector's performance with respect to the number of disparate sonar images. For this study, we considered two (HF-BB_1), three (HF-BB_1-BB_2), and four (HF-BB_1-BB_2-BB_3) sonar detection systems. Each of the three detection systems is then tested on the data set using the same threshold of 10.2. The two-sonar detector performs well detecting 51 of 53 targets with 7.48 false alarms per image. The three-sonar detector improves upon this by detecting 52 targets with 8.93 false alarms per image. However, the four-sonar detector does not seem to provide as great an increase in performance as it detects the same 52 targets but incurs a higher false alarm rate of 9.32 false alarms per image. The ROC curves corresponding to all three detection systems convey the same message as the two-sonar detector exhibits $P_d = 96\%$ at the knee-point which is then improved upon when going to the three-sonar detector with $P_d = 98\%$ at the knee-point. However, the probability of detection for the four-sonar detector falls to $P_d = 96\%$. When going from the three-channel to the four-channel detector, this lack of significant increase in performance suggests that the point of diminishing return is a system with one HF and two BB sonar images.

For the second data set, a two-channel MCA-based detection system was then implemented consisting of one HF image and one BB sonar image. In this case, the data set contains two different types of objects, targets and lobster traps. Studying the multi-channel correlations pertaining to each object, we observed differences among the features generated from each object exemplifying the power of the multi-channel correlations for target versus man-made object discrimination. A threshold was then chosen based on the knee-point of the ROC curve generated from a partial subset of images and was subsequently tested on the remaining images in this data set. We found that the detector performed well with all targets detected successfully

while maintaining 7 false alarms per image on average. A ROC curve generated from the images used in the test also suggests high performance as the knee-point exhibits $P_d = 98\%$ and $P_{fa} = 2\%$. We found that the proposed MCA-based detector outperforms that of the CCA-based detector given in [1] as this detection method exhibits only $P_d = 90\%$ and $P_{fa} = 10\%$ at the knee-point of its ROC curve. This degradation in performance may be attributed to the fact that it was assumed in [1] that both channels contain a signal component with the same covariance structure. However, given the disparateness in HF and BB images, this is more than likely a poor model of the true statistical nature of the signal components in both of the images. We found that the proposed MCA-based detector performed well given the fact that the data lacked any phase information and was more than likely non-Gaussian in nature resulting from taking the magnitude of the complex-valued data.

Finally, the detection method was applied to data set consisting of simulated target and non-target shapes embedded in simulated background. The images are generated with different target types and at various SNR values, image resolutions, ranges, aspect angles, etc. To mimic detection in multi-sonar situations, we built a two-channel MCA-based detector where each image contained the same target type at the same SNR, range, and aspect angle. However, the two images differed in resolution: one at $1in$ and the other at $3in$ resolution. The results were then partitioned on the basis of target type and SNR and ROC curves were plotted. These results indicated the robustness of the detection method to different target structures at various SNR values. Next, to mimic multi-platform, single-sonar situations, we again built a two-channel MCA-based detector where each image contained the same target type at the same SNR ($9dB$), range, and image resolution ($1in$). However, this time the two channels differ in aspect angle and pairs of images were chosen such that the difference in aspect angle was within the range of a particular aspect separation. ROC curves were then plotted for two different target types at various aspect separations. These

results demonstrated the detection method's robustness to aspect separation. Overall, we found that the detection method developed in Chapter 3 performed extremely well when given multiple sources of information and remained fairly robust to variables that are encountered in realistic multi-channel target detection problems.

CHAPTER 5

GAUSS-GAUSS LIKELIHOOD UPDATING

5.1 Introduction

Next, we will investigate the effects of incrementally adding additional data to the Gauss-Gauss detector. More specifically, we show that updating the log-likelihood ratio involves linearly estimating the *new* data we wish to add from the *old* data that we have already measured and adjusting the likelihood ratio accordingly. Updating in this fashion can have many practical applications such as situations where we would rather build a detector for a small subset of observations and iteratively update to account for the rest of our measurement. Or there could exist situations where we have already built a detector to handle an observation and we wish to add more to our vector of measurements. In either case, log-likelihood ratio updating can be utilized to solve both problems provided that we correctly estimate new observations from our previous measurements. For our sonar target detection problem, the latter corresponds to situations when new platforms join the decision-making, e.g. in collaborative AUVs, while the former corresponds to adding more data samples, e.g. pings.

We also look at the increase in J-divergence as a consequence of this incremental data augmentation. In this case, we find that the change in divergence can be written in terms of error covariance matrices when filtering with a smoother that is matched/miss-matched to the given hypothesis. The change in J-divergence in this framework gives us some insight as to how the performance of the detector incrementally changes as we add additional measurements to our observation. When adding measurements from disparate sources of information, such as multiple sensor

platforms, the change in J-divergence can also give us some intuition as to which platform's observation we should add or when adding observations from platforms reaches a point of diminishing return. A similar argument as in [25] is taken to find a coordinate system where error vectors can be approximated in a lower dimensional space. This coordinate system becomes especially useful in applications where the data to be added is high dimensional. This can result in high computation savings without a significant decrease in the change in J-divergence.

We then provide results from two simulations to demonstrate the usefulness of log-likelihood ratio updating. The first example is concerned with detecting the presence of dynamical structure in our observation and adding data amounts to *temporally updating* the likelihood ratio. The second example is concerned with detecting the presence of a single narrow-band source in the sensing environment of multiple uniform linear arrays (ULAs). A situation is constructed where a detector is built for one individual platform alone and where after the likelihood ratio is updated via a *channel updating* to successively account for the observations from other available platforms. We finally consider a situation where there are multiple platforms in the sensing environment of the source but each detector can only take advantage of a subset of the total number of observations being made. In this case, we show that the change in J-divergence can be an effective tool for deciding which set of observations should be added for increased detection performance.

The outline of this chapter is as follows. Section 5.2 presents general and Gauss-Gauss likelihood ratio updating as well as reduced-rank Gauss-Gauss updating. Section 5.3 presents the two simulation studies and gives the results from each. Concluding remarks are then given in Section 5.4.

5.2 Log-Likelihood Updating

5.2.1 General Log-Likelihood Updating

Consider the observation vector $\mathbf{z}_k \in \mathbb{C}^m$ which we assume arises from one of two possible hypotheses. With this observation vector, we consider the detection problem

$$\begin{aligned} H_1 &: \mathbf{z}_k \sim f_{\mathbf{Z}_k}(\mathbf{z}_k|H_1) \\ H_0 &: \mathbf{z}_k \sim f_{\mathbf{Z}_k}(\mathbf{z}_k|H_0) \end{aligned}$$

where $f_{\mathbf{Z}_k} : \mathbb{C}^m \to [0, \infty]$ is the multivariate marginal density function of our observation. With this, we construct the log-likelihood ratio

$$l(\mathbf{z}_k) = \log \frac{f_{\mathbf{Z}_k}(\mathbf{z}_k|H_1)}{f_{\mathbf{Z}_k}(\mathbf{z}_k|H_0)}$$

We then add the vector of measurements $\mathbf{x}_{k+1} \in \mathbb{C}^n$ to the detection problem (with no effect on the previous observation we have already made) to form the new augmented observation $\mathbf{z}_{k+1} = \left[\mathbf{z}_k^H \ \mathbf{x}_{k+1}^H\right]^H$. Noting the definition of conditional probability density, namely

$$f_{\mathbf{X}_{k+1}|\mathbf{Z}_k}(\mathbf{x}_{k+1}|\mathbf{z}_k) = \frac{f_{\mathbf{Z}_k,\mathbf{X}_{k+1}}(\mathbf{z}_k, \mathbf{x}_{k+1})}{f_{\mathbf{Z}_k}(\mathbf{z}_k)}$$

we can write the log-likelihood ratio of the augmented observation as [44], [45]

$$\begin{aligned} l(\mathbf{z}_{k+1}) &= \log \frac{f_{\mathbf{Z}_{k+1}}(\mathbf{z}_{k+1}|H_1)}{f_{\mathbf{Z}_{k+1}}(\mathbf{z}_{k+1}|H_0)} \\ &= \log \frac{f_{\mathbf{Z}_k,\mathbf{X}_{k+1}}(\mathbf{z}_k, \mathbf{x}_{k+1}|H_1)}{f_{\mathbf{Z}_k,\mathbf{X}_{k+1}}(\mathbf{z}_k, \mathbf{x}_{k+1}|H_0)} \\ &= \log \frac{f_{\mathbf{Z}_k}(\mathbf{z}_k|H_1)\, f_{\mathbf{X}_{k+1}|\mathbf{Z}_k}(\mathbf{x}_{k+1}|\mathbf{z}_k, H_1)}{f_{\mathbf{Z}_k}(\mathbf{z}_k|H_0)\, f_{\mathbf{X}_{k+1}|\mathbf{Z}_k}(\mathbf{x}_{k+1}|\mathbf{z}_k, H_0)} \\ &= l(\mathbf{z}_k) + \log \frac{f_{\mathbf{X}_{k+1}|\mathbf{Z}_k}(\mathbf{x}_{k+1}|\mathbf{z}_k, H_1)}{f_{\mathbf{X}_{k+1}|\mathbf{Z}_k}(\mathbf{x}_{k+1}|\mathbf{z}_k, H_0)} \end{aligned} \tag{5.1}$$

Thus, in this very general setting, we can see that the log-likelihood ratio can be updated through simple addition provided that we can accurately characterize the distribution of our new observation, conditioned upon the old measurement, under both hypotheses. In general, this updating factor is not a likelihood ratio itself, the

exception to this statement being situations where the new observation is conditionally independent of the old measurement under both hypotheses in which case the update simply becomes a likelihood ratio for the new data, i.e.

$$l(\mathbf{z}_{k+1}) = l(\mathbf{z}_k) + \log \frac{f_{\mathbf{X}_{k+1}}(\mathbf{x}_{k+1}|H_1)}{f_{\mathbf{X}_{k+1}}(\mathbf{x}_{k+1}|H_0)}$$

5.2.2 Full-Rank Gauss-Gauss Updating

Recall the framework of the Gauss-Gauss detector described in Section 2.3. From the previous remarks given above, the ability to update the Gauss-Gauss detector by simply adding an additional term should come as no surprise. Just as we have done in the previous section, we begin by making the assumption that we have measured the (zero-mean) Gaussian random vector $\mathbf{z}_k \in \mathbb{C}^m$ and form the likelihood ratio $l(\mathbf{z}_k)$, i.e.

$$l(\mathbf{z}_k) = \mathbf{z}_k^H \left(R^{-1}_{\mathbf{z}_k \mathbf{z}_{k_0}} - R^{-1}_{\mathbf{z}_k \mathbf{z}_{k_1}} \right) \mathbf{z}_k$$

where

$$R_{\mathbf{z}_k \mathbf{z}_{k_0}} = E_{H_0}\left[\mathbf{z}_k \mathbf{z}_k^H\right], \quad R_{\mathbf{z}_k \mathbf{z}_{k_1}} = E_{H_1}\left[\mathbf{z}_k \mathbf{z}_k^H\right]$$

and $\mathbf{z}_k = \left[\mathbf{x}_1^H \ \cdots \ \mathbf{x}_k^H\right]^H$. Note that the binary subscripts $(0, 1)$ denote which hypothesis is being conditioned upon. We then add a new observation vector, $\mathbf{x}_{k+1} \in \mathbb{C}^n$, to the measurements to obtain the augmented observation vector

$$\mathbf{z}_{k+1} = \left[\mathbf{z}_k^H \ \mathbf{x}_{k+1}^H\right]^H \in \mathbb{C}^{m+n}$$

Thus, the log-likelihood ratio of the augmented observation vector becomes

$$l(\mathbf{z}_{k+1}) = \mathbf{z}_{k+1}^H \left(R^{-1}_{\mathbf{z}_{k+1} \mathbf{z}_{k+1_0}} - R^{-1}_{\mathbf{z}_{k+1} \mathbf{z}_{k+1_1}} \right) \mathbf{z}_{k+1}$$

where

$$R_{\mathbf{z}_{k+1} \mathbf{z}_{k+1_0}} = E_{H_0}\left[\mathbf{z}_{k+1} \mathbf{z}_{k+1}^H\right], \quad R_{\mathbf{z}_{k+1} \mathbf{z}_{k+1_1}} = E_{H_1}\left[\mathbf{z}_{k+1} \mathbf{z}_{k+1}^H\right]$$

Noting the structure that is inherent in the augmented observation we can use a block matrix inversion identity [26] and write

$$R^{-1}_{\mathbf{z}_{k+1}\mathbf{z}_{k+1}} = \begin{bmatrix} R^{-1}_{\mathbf{z}_k\mathbf{z}_k} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} + \begin{bmatrix} -W^H \\ I \end{bmatrix} Q^{-1} \begin{bmatrix} -W & I \end{bmatrix}$$

where

$$\begin{aligned} W &= R_{\mathbf{x}_{k+1}\mathbf{z}_k} R^{-1}_{\mathbf{z}_k\mathbf{z}_k} \\ Q &= R_{\mathbf{x}_{k+1}\mathbf{x}_{k+1}} - R_{\mathbf{x}_{k+1}\mathbf{z}_k} R^{-1}_{\mathbf{z}_k\mathbf{z}_k} R^H_{\mathbf{x}_{k+1}\mathbf{z}_k} \end{aligned}$$

It is interesting to note [26] that W is a discrete Wiener smoothing matrix that estimates the new data from the old and Q is its associated error covariance matrix. With this in mind, it is easy to observe that the change in the likelihood ratio becomes

$$\begin{aligned} \Delta l\left(\mathbf{z}_{k+1}, \mathbf{z}_k\right) &= l\left(\mathbf{z}_{k+1}\right) - l\left(\mathbf{z}_k\right) \\ &= \begin{bmatrix} \mathbf{z}_k^H & \mathbf{x}_{k+1}^H \end{bmatrix} \left(\begin{bmatrix} -W_0^H \\ I \end{bmatrix} Q_0^{-1} \begin{bmatrix} -W_0 & I \end{bmatrix} \right. \\ &\quad \left. - \begin{bmatrix} -W_1^H \\ I \end{bmatrix} Q_1^{-1} \begin{bmatrix} -W_1 & I \end{bmatrix} \right) \begin{bmatrix} \mathbf{z}_k \\ \mathbf{x}_{k+1} \end{bmatrix} \\ &= \mathbf{e}_0^H Q_0^{-1} \mathbf{e}_0 - \mathbf{e}_1^H Q_1^{-1} \mathbf{e}_1 \end{aligned} \tag{5.2}$$

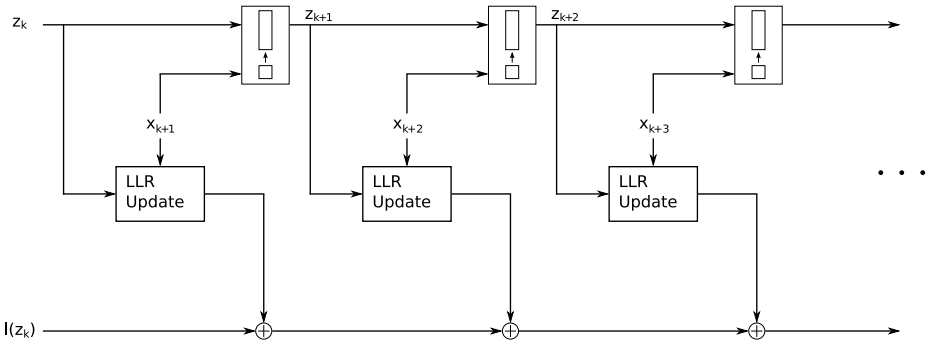
where $W_0 = R_{\mathbf{x}_{k+1}\mathbf{z}_{k_0}} R^{-1}_{\mathbf{z}_k\mathbf{z}_{k_0}}$ and $W_1 = R_{\mathbf{x}_{k+1}\mathbf{z}_{k_1}} R^{-1}_{\mathbf{z}_k\mathbf{z}_{k_1}}$ are Wiener filters conditioned upon H_0 and H_1, respectively, and $\mathbf{e}_0 = \mathbf{x}_{k+1} - W_0\mathbf{z}_k$ and $\mathbf{e}_1 = \mathbf{x}_{k+1} - W_1\mathbf{z}_k$ are the error vectors produced by these smoothing matrices with covariance matrices

$$Q_0 = E_{H_0}\left[\mathbf{e}_0\mathbf{e}_0^H\right] = R_{\mathbf{x}_{k+1}\mathbf{x}_{k+1_0}} - R_{\mathbf{x}_{k+1}\mathbf{z}_{k_0}} R^{-1}_{\mathbf{z}_k\mathbf{z}_{k_0}} R^H_{\mathbf{x}_{k+1}\mathbf{z}_{k_0}}$$

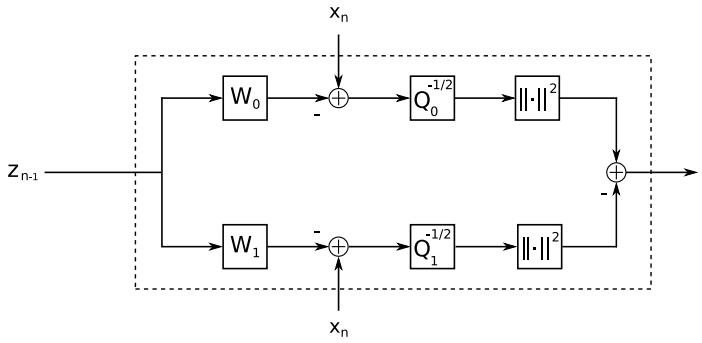
and

$$Q_1 = E_{H_1}\left[\mathbf{e}_1\mathbf{e}_1^H\right] = R_{\mathbf{x}_{k+1}\mathbf{x}_{k+1_1}} - R_{\mathbf{x}_{k+1}\mathbf{z}_{k_1}} R^{-1}_{\mathbf{z}_k\mathbf{z}_{k_1}} R^H_{\mathbf{x}_{k+1}\mathbf{z}_{k_1}},$$

respectively. Clearly, due to the orthogonality principle [26], we have the following



(a) LLR Updating Structure.



(b) LLR Update Block.

Figure 5.1: Log-Likelihood Ratio Updating.

properties

$$E_{H_0}\left[\mathbf{e}_0\mathbf{z}_k^H\right] = \mathbf{O}$$

$$E_{H_1}\left[\mathbf{e}_1\mathbf{z}_k^H\right] = \mathbf{O}$$

However, the error vectors produced by these filters do not share the same property when conditioned upon the wrong hypothesis, i.e.

$$E_{H_0}\left[\mathbf{e}_1\mathbf{z}_k^H\right] \neq \mathbf{O}$$

$$E_{H_1}\left[\mathbf{e}_0\mathbf{z}_k^H\right] \neq \mathbf{O}$$

Thus, we can perform single-channel updating as depicted in Figures 5.1(a) and 5.1(b) provided that we have the correct smoothing and error covariance matrices. From a practical standpoint this figure shows us that we do not have to rebuild structure from

scratch when we wish to add an additional channel; we simply add an additional piece that estimates the new from the old and update the likelihood measure accordingly.

Because of the linearity property of expectation, it is easily seen that the change in J-divergence becomes

$$\begin{aligned}
\Delta J\left(\mathbf{z}_{k+1}, \mathbf{z}_k\right) &= J\left(\mathbf{z}_{k+1}\right) - J\left(\mathbf{z}_k\right) \\
&= E_{H_1}\left[\Delta l\left(\mathbf{z}_{k+1}, \mathbf{z}_k\right)\right] - E_{H_0}\left[\Delta l\left(\mathbf{z}_{k+1}, \mathbf{z}_k\right)\right] \\
&= \operatorname{tr}\left(-Q_0^{-1} E_{H_0}\left[\mathbf{e}_0 \mathbf{e}_0^H\right] - Q_1^{-1} E_{H_1}\left[\mathbf{e}_1 \mathbf{e}_1^H\right]\right. \\
&\qquad \left. + Q_0^{-1} E_{H_1}\left[\mathbf{e}_0 \mathbf{e}_0^H\right] + Q_1^{-1} E_{H_0}\left[\mathbf{e}_1 \mathbf{e}_1^H\right]\right) \\
&= \operatorname{tr}\left(-2I + Q_0^{-1} Q_{10} + Q_1^{-1} Q_{01}\right) \qquad (5.3)
\end{aligned}$$

where

$$\begin{aligned}
Q_{10} &= E_{H_1}\left[\mathbf{e}_0 \mathbf{e}_0^H\right] \\
&= R_{\mathbf{x}_{k+1}\mathbf{x}_{k+1_1}} - W_0 R_{\mathbf{z}_k \mathbf{x}_{k+1_1}} - R_{\mathbf{x}_{k+1}\mathbf{z}_{k_1}} W_0^H + W_0 R_{\mathbf{z}_k \mathbf{z}_{k_1}} W_0^H \\
Q_{01} &= E_{H_0}\left[\mathbf{e}_1 \mathbf{e}_1^H\right] \\
&= R_{\mathbf{x}_{k+1}\mathbf{x}_{k+1_0}} - W_1 R_{\mathbf{z}_k \mathbf{x}_{k+1_0}} - R_{\mathbf{x}_{k+1}\mathbf{z}_{k_0}} W_1^H + W_1 R_{\mathbf{z}_k \mathbf{z}_{k_0}} W_1^H
\end{aligned}$$

are cross terms representing the error covariance when using the wrong smoothing filter. That is, Q_{10} is the error covariance matrix incurred when filtering with W_0 given that it is actually the H_1 model that produced the data and vice versa for Q_{01}. Since we are filtering with a sub-optimal smoother in such situations, we can make the following two statements

$$\begin{aligned}
\mathbf{x}^H Q_{10} \mathbf{x} &\geq \mathbf{x}^H Q_1 \mathbf{x} \\
\mathbf{x}^H Q_{01} \mathbf{x} &\geq \mathbf{x}^H Q_0 \mathbf{x} \qquad (5.4)
\end{aligned}$$

for any $\mathbf{x} \in \mathbb{C}^n$ not equal to the empty vector. These two inequalities will be used later to lower bound the change in J-divergence.

5.2.3 Reduced-Rank Gauss-Gauss Updating

In detection applications where the data sits in a high dimensional space, one may be interested in finding a low-rank approximation of the data to save processing time and would like to do so without sacrificing a great deal of performance in terms of detection. Again, we assume the same structure as in the previous section i.e., adding the new observation $\mathbf{x}_{k+1} \in \mathbb{C}^n$ to the old observation $\mathbf{z}_k$. Along the same lines as that presented in [25], we begin by removing the contribution from the H_0 hypothesis through a whitening transformation. To be specific, we begin by rewriting the change in log-likelihood as follows

$$\begin{aligned} \Delta l\left(\mathbf{z}_{k+1}, \mathbf{z}_k\right) &= \mathbf{e}_0^H Q_0^{-H/2} Q_0^{-1/2} \mathbf{e}_0 - \mathbf{e}_1^H Q_0^{-H/2} Q_0^{H/2} Q_1^{-1} Q_0^{1/2} Q_0^{-1/2} \mathbf{e}_1 \\ &= \mathbf{w}_0^H \mathbf{w}_0 - \mathbf{w}_1^H \Gamma^{-1} \mathbf{w}_1 \end{aligned} \tag{5.5}$$

where the vectors $\mathbf{w}_0 = Q_0^{-1/2}\mathbf{e}_0$ and $\mathbf{w}_1 = Q_0^{-1/2}\mathbf{e}_1$ have the following covariance structure under their respective hypotheses

$$E_{H_0}\left[\mathbf{w}_0 \mathbf{w}_0^H\right] = I$$
$$E_{H_1}\left[\mathbf{w}_1 \mathbf{w}_1^H\right] = \Gamma$$

and the matrix $\Gamma = Q_0^{-1/2} Q_1 Q_0^{-H/2}$ is simply a normalized version of Q_1. We then use the eigenvalue decomposition of this normalized error covariance matrix, Γ, so that

$$\Gamma = U \Sigma U^H$$

where $U^H U = U U^H = I$ and $\Sigma = \operatorname{diag}\left[\sigma_1, \ldots, \sigma_n\right]$. With this, we rewrite (5.5) as

$$\begin{aligned} \Delta l\left(\mathbf{z}_{k+1}, \mathbf{z}_k\right) &= \mathbf{w}_0^H U U^H \mathbf{w}_0 - \mathbf{w}_1^H U \Sigma^{-1} U^H \mathbf{w}_1 \\ &= \mathbf{y}_0^H \mathbf{y}_0 - \mathbf{y}_1^H \Sigma^{-1} \mathbf{y}_1 \end{aligned} \tag{5.6}$$

where the vectors $\mathbf{y}_0 = U^H \mathbf{w}_0 = U^H Q_0^{-1/2} \mathbf{e}_0$ and $\mathbf{y}_1 = U^H \mathbf{w}_1 = U^H Q_0^{-1/2} \mathbf{e}_1$ now have covariance structure

$$E_{H_0}\left[\mathbf{y}_0 \mathbf{y}_0^H\right] = I$$
$$E_{H_1}\left[\mathbf{y}_1 \mathbf{y}_1^H\right] = \Sigma$$

To find the corresponding expressions for the change in J-divergence, we define the following two matrices

$$\begin{aligned}
\Gamma_{10} &= E_{H_1}\left[Q_0^{-\frac{1}{2}} \mathbf{e}_0 \mathbf{e}_0^H Q_0^{-\frac{H}{2}}\right] = Q_0^{-\frac{1}{2}} Q_{10} Q_0^{-\frac{H}{2}} \\
\Gamma_{01} &= E_{H_0}\left[Q_0^{-\frac{1}{2}} \mathbf{e}_1 \mathbf{e}_1^H Q_0^{-\frac{H}{2}}\right] = Q_0^{-\frac{1}{2}} Q_{01} Q_0^{-\frac{H}{2}}
\end{aligned}$$

Now, using (5.3) the change in J-divergence can be written as

$$\begin{aligned}
\Delta J\left(\mathbf{z}_{k+1}, \mathbf{z}_k\right) &= \operatorname{tr}\left(-2I + Q_0^{-\frac{1}{2}} Q_{10} Q_0^{-\frac{H}{2}} + Q_0^{\frac{H}{2}} Q_1^{-1} Q_0^{\frac{1}{2}} Q_0^{-\frac{1}{2}} Q_{01} Q_0^{-\frac{H}{2}}\right) \\
&= \operatorname{tr}\left(-2I + \Gamma_{10} + \Gamma^{-1} \Gamma_{01}\right) \\
&= \operatorname{tr}\left(-2I + U^H \Gamma_{10} U + \Sigma^{-1} U^H \Gamma_{01} U\right) \\
&= \sum_{i=1}^{n} -2 + \mathbf{u}_i^H \left(\Gamma_{10} + \sigma_i^{-1} \Gamma_{01}\right) \mathbf{u}_i
\end{aligned} \tag{5.7}$$

Thus, we find that the change in J-divergence seen when adding an additional observation becomes decomposed in terms of the eigenvalues of Γ, i.e. σ_i's, and the two quadratic terms $\mathbf{u}_i^H \Gamma_{10} \mathbf{u}_i$ and $\mathbf{u}_i^H \Gamma_{01} \mathbf{u}_i$. Now the eigenvalues, σ_i, give us a canonical measure of the mean-squared error (due to Q_1) incurred when estimating data generated from the H_1 hypothesis with the filter W_1 normalized by the mean-squared error incurred when estimating data generated from the H_0 hypothesis with the filter W_0 (due to $Q_0^{-1/2}$). On the other hand, the quadratic term $\mathbf{u}_i^H \Gamma_{10} \mathbf{u}_i$ measures the mean-squared error incurred when estimating data generated from the H_1 hypothesis with the filter W_0 in the one-dimensional subspace spanned by the vector $\mathbf{u}_i$, again normalized by Q_0. A similar argument can be made about $\mathbf{u}_i^H \Gamma_{01} \mathbf{u}_i$. The effect of this decomposition on covariance ellipses is shown in Figure 5.2. Therefore,

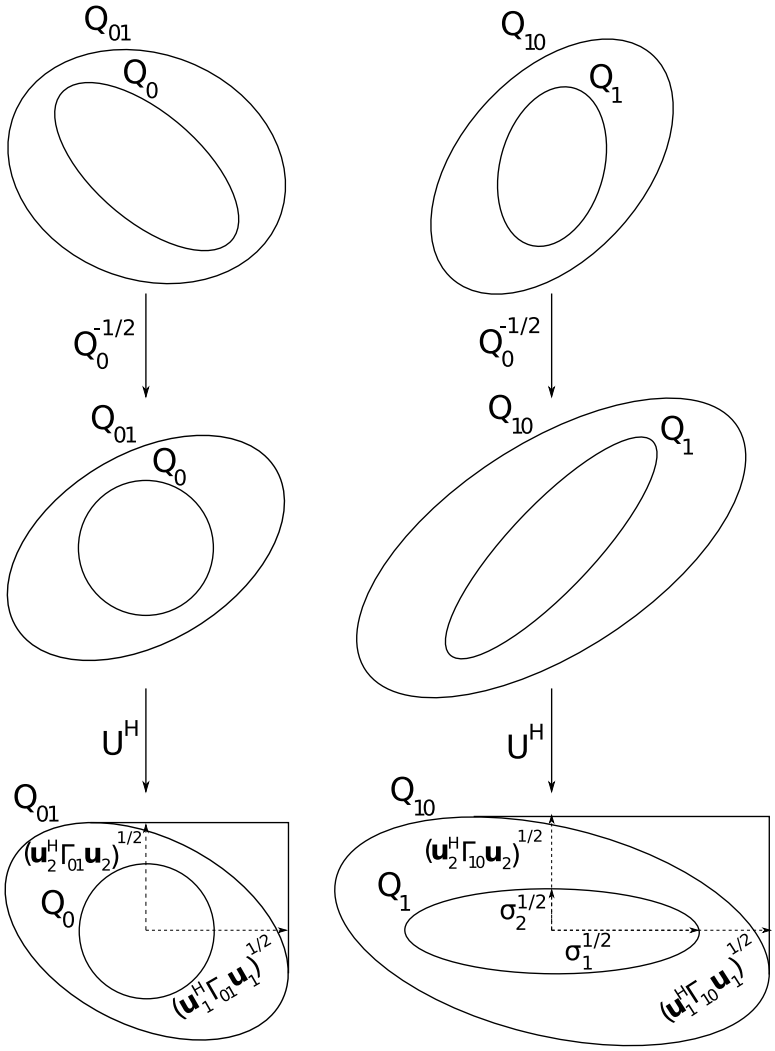


Figure 5.2: Error Covariance Decomposition

we first, in some sense, "normalize" the problem by the H_0 hypothesis so that, with high probability, error vectors generated by filtering with W_0 under H_0 lie somewhere within a sphere of radius one. The picture is then rotated through U which resolves the data into a coordinate system associated with the principal axes of the ellipsoid corresponding to Γ. The square of the radii along these principal axes correspond to the eigenvalues of Γ and the squared distance from the origin to that line which is orthogonal to the i^{th} basis vector, $\mathbf{u}_i$, but also tangent to the appropriate covariance ellipse corresponds to the two quadratic terms, $\mathbf{u}_i^H \Gamma_{10} \mathbf{u}_i$ and $\mathbf{u}_i^H \Gamma_{01} \mathbf{u}_i$.

Recalling the two inequalities given in (5.4), we can find a lower bound on the

change in J-divergence via the following sequence of statements

$$\begin{aligned}
\Delta J\left(\mathbf{z}_{k+1}, \mathbf{z}_k\right) &= \sum_{i=1}^{n} -2 + \mathbf{u}_i^H \Gamma_{10} \mathbf{u}_i + \sigma_i^{-1} \mathbf{u}_i^H \Gamma_{01} \mathbf{u}_i \\
&= \sum_{i=1}^{n} -2 + \mathbf{u}_i^H Q_0^{-1/2} Q_{10} Q_0^{-H/2} \mathbf{u}_i + \sigma_i^{-1} \mathbf{u}_i^H Q_0^{-1/2} Q_{01} Q_0^{-H/2} \mathbf{u}_i \\
&\geq \sum_{i=1}^{n} -2 + \mathbf{u}_i^H Q_0^{-1/2} Q_1 Q_0^{-H/2} \mathbf{u}_i + \sigma_i^{-1} \mathbf{u}_i^H Q_0^{-1/2} Q_0 Q_0^{-H/2} \mathbf{u}_i \\
&= \sum_{i=1}^{n} -2 + \mathbf{u}_i^H \Gamma \mathbf{u}_i + \sigma_i^{-1} \mathbf{u}_i^H I \mathbf{u}_i
\end{aligned}$$

Since $\mathbf{u}_i$ is orthogonal and diagonalizes matrix Γ, it then follows that

$$\Delta J\left(\mathbf{z}_{k+1}, \mathbf{z}_k\right) \geq \sum_{i=1}^{n} -2 + \sigma_i + \sigma_i^{-1} \tag{5.8}$$

More importantly, we can then observe that the function $-2 + \sigma_i + \sigma_i^{-1} \geq 0$ for any $\sigma_i \geq 0$ leading us to the conclusion that $\Delta J\left(\mathbf{z}_{k+1}, \mathbf{z}_k\right) \geq 0$. As stated before, J-divergence is not a definitive measure of the performance of the detector which should be truly portrayed in terms of probability of detection. However, strictly from a first-order moment point-of-view, this inequality seems to suggest that supplying an additional observation to the detector can *never* deteriorate its performance.

When performing rank-p updating ($p < n$), we re-order the coordinates in a descending fashion so that the pair $(\mathbf{u}_{1^*}, \sigma_{1^*})$ satisfies

$$(\mathbf{u}_{1^*}, \sigma_{1^*}) = \arg\max_{\mathbf{u}_i, \sigma_i} \left\{ \mathbf{u}_i^H \left(\Gamma_{10} + \sigma_i^{-1} \Gamma_{01} \right) \mathbf{u}_i : i \in [1, n] \right\}$$

Likewise, the pair $(\mathbf{u}_{2^*}, \sigma_{2^*})$ satisfies

$$(\mathbf{u}_{2^*}, \sigma_{2^*}) = \arg\max_{\mathbf{u}_i, \sigma_i} \left\{ \mathbf{u}_i^H \left(\Gamma_{10} + \sigma_i^{-1} \Gamma_{01} \right) \mathbf{u}_i : i \in [1, n], i \neq 1^* \right\}$$

and so on until all coordinates have been accounted for. We would like to see coordinates where σ_i is very small and the two quadratic terms, $\mathbf{u}_i^H \Gamma_{10} \mathbf{u}_i$ and $\mathbf{u}_i^H \Gamma_{01} \mathbf{u}_i$, very large. As σ_i represents the mean-squared error in estimating data generated from H_1 with the right smoothing filter W_1, it is natural that we would want small mean-squared error in such a situation. Small mean-squared error in this situation implies

a significant amount of cross-correlation information among our new observation and old measurement under H_1 allowing us to accurately estimate the new data from the old. Asking for large mean-squared error with $\mathbf{u}_i^H \Gamma_{10} \mathbf{u}_i$ and $\mathbf{u}_i^H \Gamma_{01} \mathbf{u}_i$ is a subtle point and somewhat counter-intuitive as typically when one speaks of estimation problems, the smaller the mean-squared error the better. However, if we can generate data from one hypothesis and accurately estimate it with the smoothing filter from the other, then that gives us some indication that there is not much difference among the models of the two hypotheses (or their distributions for that matter). Thus, in such situations the *larger* the mean-squared error the better.

With the coordinates suitably ordered, we decompose the coordinate system as follows

$$U = \begin{bmatrix} U_p & U_{p+1} \end{bmatrix}$$

where $U_p = \begin{bmatrix} \mathbf{u}_{1^*} & \cdots & \mathbf{u}_{p^*} \end{bmatrix}$ and $U_{p+1} = \begin{bmatrix} \mathbf{u}_{(p+1)^*} & \cdots & \mathbf{u}_{n^*} \end{bmatrix}$. Also,

$$\Sigma = \begin{bmatrix} \Sigma_p & \mathbf{O} \\ \mathbf{O} & \Sigma_{p+1} \end{bmatrix}$$

where $\Sigma_p = \operatorname{diag}\left[\sigma_{1^*} \ \cdots \ \sigma_{p^*}\right]$ and $\Sigma_{p+1} = \operatorname{diag}\left[\sigma_{(p+1)^*} \ \cdots \ \sigma_{n^*}\right]$. These yield the low-rank approximations of the error vectors as

$$\begin{aligned} \tilde{\mathbf{y}}_0 &= U_p^H Q_0^{-\frac{1}{2}} \mathbf{e}_0 \ \in \ \mathbb{C}^p \\ \tilde{\mathbf{y}}_1 &= U_p^H Q_0^{-\frac{1}{2}} \mathbf{e}_1 \ \in \ \mathbb{C}^p \end{aligned}$$

Using (5.6), the change in log-likelihood in this reduced-rank subspace becomes

$$\begin{aligned} \Delta l_p\left(\mathbf{z}_{k+1}, \mathbf{z}_k\right) &= \tilde{\mathbf{y}}_0^H \tilde{\mathbf{y}}_0 - \tilde{\mathbf{y}}_1^H \Sigma_p^{-1} \tilde{\mathbf{y}}_1 \\ &= \mathbf{e}_0^H Q_0^{-\frac{H}{2}} U_p U_p^H Q_0^{-\frac{1}{2}} \mathbf{e}_0 - \mathbf{e}_1^H Q_0^{-\frac{H}{2}} U_p \Sigma_p^{-1} U_p^H Q_0^{-\frac{1}{2}} \mathbf{e}_1 \end{aligned}$$

The reduced-rank version of the LLR updating block shown in Figure 5.1(b) is now

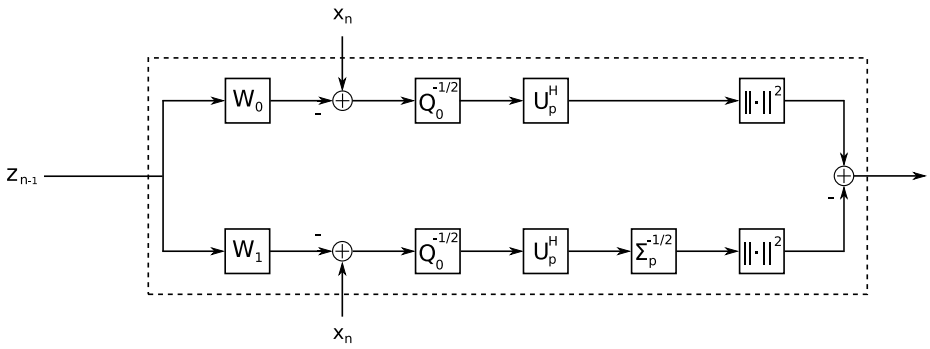


Figure 5.3: Reduced-Rank LLR Update Block.

displayed in Figure 5.3 for this reduced-rank update. In each branch, the updating involves filtering the normalized (by $Q_0^{-1/2}$) error vectors into a lower-dimensional vector after which the energy is computed to give the required log-likelihood ratio update. This is not necessarily the same solution one would achieve when implementing the reduced-rank detectors of [25] for the augmented measurement $\mathbf{z}_{k+1}$. However, in situations where a detector (full or reduced-rank) has already been built for $\mathbf{z}_k$ and we wish to add the new observation $\mathbf{x}_{k+1}$, we do not have to rebuild a low-rank detector to handle this new observation vector and can simply add a low-rank update.

Using (5.3), the associated change in J-divergence can finally be written as

$$\begin{aligned}\Delta J_p\left(\mathbf{z}_{k+1}, \mathbf{z}_k\right) &= \operatorname{tr}\left(-2I + U_p^H \Gamma_{10} U_p + \Sigma_p^{-1} U_p^H \Gamma_{01} U_p\right) \\ &= \sum_{i=1}^{p} -2 + \mathbf{u}_i^H \left(\Gamma_{10} + \sigma_i^{-1} \Gamma_{01}\right) \mathbf{u}_i\end{aligned}$$

5.3 Simulation Studies

To demonstrate situations where log-likelihood updating may be useful, we next consider two simulations. The first simulation example is concerned with detecting the presence of dynamical structure in data that we have observed. In a situation such as this, one does not want to perform detection with the entire time series as each observation (time series) can sit in a high-dimensional space. To get past this computational obstruction, we take advantage of the structure of the hypothesis test and

implement a Kalman filter [26] which is used to propagate the necessary errors and error variances needed for log-likelihood ratio updating yielding a detection system that can be implemented in real-time. The second simulation example is concerned with detecting the presence of a single, narrow-band source in the sensing environment of multiple Uniform Linear Arrays (ULAs). Here, we investigate the usefulness of the incremental change in J-divergence in deciding if/when adding new platforms reaches a point of diminishing returns and also its usefulness in deciding which platforms to use for an increase in detection performance.

5.3.1 Detection of Dynamical Structure-Time Updating

Consider the scalar time series, $\{y[k]\}_{k=0}^{\infty}$, which we assume arises from a white process being passed through a known LTI system and corrupted by noise under one hypothesis versus noise alone under the other. To be specific, we consider the detection problem

$$\begin{aligned} H_1 &: y[k] = \sum_{l=0}^{\infty} h[k-l]u[l] + n[k] \\ H_0 &: y[k] = n[k] \end{aligned}$$

where $h[k]$ is the impulse response of an arbitrary infinite-impulse response (IIR) system. We also assume that the discrete-time processes $u[k]$ and $n[k]$ are individually Gaussian wide-sense stationary (GWSS) and jointly GWSS with auto and cross-correlation sequences

$$\begin{aligned} R_u[l] &= E\left[u[k+l]u[k]\right] = \sigma_u^2\delta[l] \\ R_n[l] &= E\left[n[k+l]n[k]\right] = \sigma_n^2\delta[l] \\ R_{un}[l] &= E\left[u[k+l]n[k]\right] = 0 \ \forall \ l \end{aligned}$$

If the system is rational, then we may represent the LTI system in terms of a state space equation and can alternatively write the detection statement as

$$\begin{aligned} H_1 \quad : \quad & \begin{cases} \mathbf{x}[k+1] &= A\mathbf{x}[k] + \mathbf{b}u[k] \\ y[k] &= \mathbf{c}^T\mathbf{x}[k] + du[k] + n[k] \end{cases} \\ H_0 \quad : \quad & y[k] = n[k] \end{aligned}$$

where $\mathbf{x}[k]$ is the state vector and A, $\mathbf{b}$, $\mathbf{c}$, and d are matrices of appropriate dimensions [26]. Now, to put this problem in the contexts of this paper, we form the vector of measurements up to time k, $\mathbf{z}_k = [y[0]\ y[1]\ \cdots\ y[k]]^T$, and we wish to add a new measurement, $y[k+1]$. This example illustrates the time updating process as described before. We know that incrementally updating the likelihood ratio requires finding errors from each hypothesis as well as their variances every time instance. Because of the nature of the time updating problem we are considering, under H_1 the error corresponds to the difference in the observation, $y[k+1]$, and the best estimate of our measurement given observations up to time k, $\hat{y}[k+1|k]$, i.e. the innovations sequence $e[k+1] = y[k+1] - \hat{y}[k+1|k]$. So we can run a Kalman filter [26] in parallel with the log-likelihood ratio updating to provide this information (the innovations and the innovations variance) for the H_1 hypothesis as the Kalman filter reduces to that of the Wiener filter when steady-state conditions have been reached [26]. Also, since our observation is white under H_0, the best estimate of $y[k+1]$ given observations up to time k is zero. Thus, the innovations process in this case amounts to the observation itself which, under H_0, has variance σ_n^2. A block diagram of the time updating problem is shown in Figure 5.4. We assume that there exists an unobserved white process which is colored through $H(z)$ and corrupted by noise under H_1 versus noise alone under H_0. The observation is then applied to a Kalman filter, which is built from the state equation describing the dynamics of our observations under H_1, and extract the innovations and innovations variance. The innovation and its variance, as well as the observation itself, are then applied to the log-likelihood

ratio update to form the test statistic at time instant k. If we denote the variance of the innovation process by $\sigma^2_{e_{k+1}}$, then the change in the log-likelihood ratio in (5.2) becomes

$$\Delta l\left(\mathbf{z}_{k+1}, \mathbf{z}_k\right) = l\left(\mathbf{z}_{k+1}\right) - l\left(\mathbf{z}_k\right) = \sigma_n^{-2} y[k+1]^2 - \sigma_{e_{k+1}}^{-2} e[k+1]^2 \tag{5.9}$$

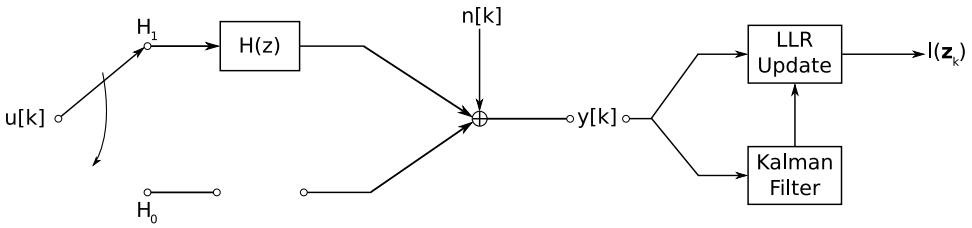
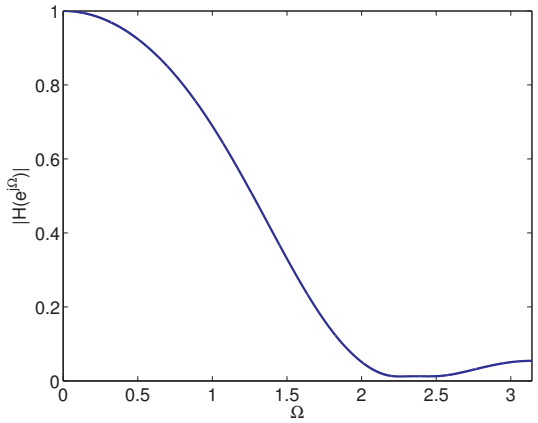


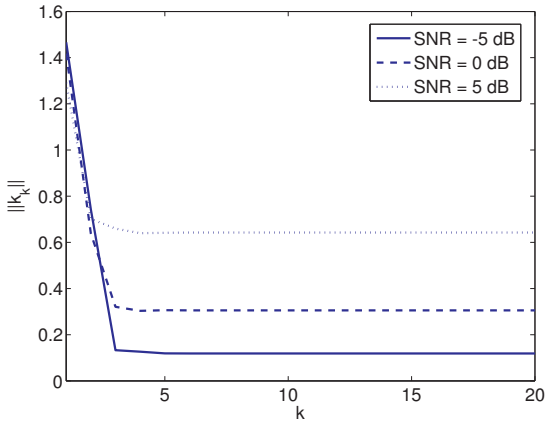
Figure 5.4: Block Diagram of Time Updating Detection System.

For simulation, we generated an arbitrary 4^{th}-order, proper LTI system by choosing random zero/pole locations with poles that lie inside the unit circle (stable system). The magnitude response of this filter is displayed in Figure 5.5(a). With σ_n^2 set to unity, Figure 5.5(b) shows the norm of the Kalman gain [26] vector, $\mathbf{k}_k$, as function of time for both the H_0 and H_1 hypotheses for different choices of σ_u^2 to achieve SNR values of -5, 0, and 5 dB. Note that the curve for H_1 associated with an SNR of -5 dB cannot be seen as it is identical to that of the H_0 hypothesis. From the figure we can see that the system has a small settling time as the norm of the Kalman gain vector reaches steady state after 10 iterations. Figure 5.6(a) displays $l\left(\mathbf{z}_k\right)$ versus k for both H_0 and H_1 at SNR values of -5, 0, and 5 dB. We see that, even under H_0, the log-likelihood ratio continually increases as time updating progresses. Nonetheless, the distance between H_1 and H_0 also increases. This presents somewhat of an inconvenience from a practical standpoint as it is obvious that to make detection decisions we must define a threshold that also continually increases over time. Failing to do so will result in an ever increasing probability of false alarm.

We then generated a finite sequence of data, $y[k];\ k = 0, \ldots, N-1$ with $N = 300$, from the H_0 hypothesis which is run through the system shown in Figure 5.4.



(a) LTI Magnitude Response

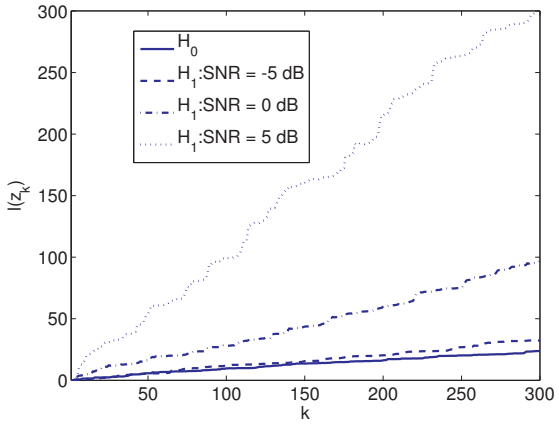


(b) Norm of the Kalman Gain Vector at SNR = -5 dB, 0 dB, and 5 dB

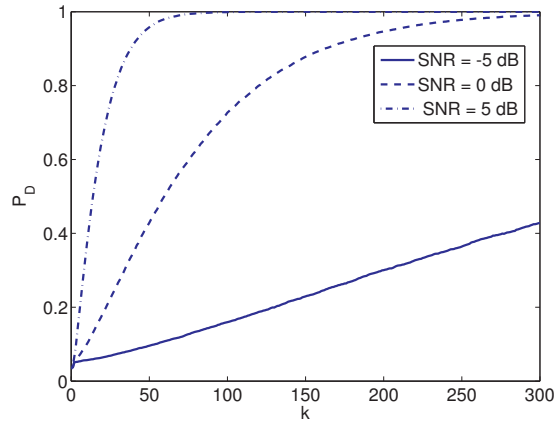
Figure 5.5: LTI Magnitude Response and Kalman Gain Vector.

When initializing the Kalman filter recursions, a state error covariance matrix of $P_{0|0} = E\left[(\mathbf{x}[0] - \hat{\mathbf{x}}[0|0])(\mathbf{x}[0] - \hat{\mathbf{x}}[0|0])^T\right] = 10I$ is always used. From this, the initial innovation variance is easily seen to be

$$\begin{aligned}
\sigma_{e_0}^2 &= E_{H_1}\left[(y[0] - \hat{y}[0|0])^2\right] \\
&= E_{H_1}\left[\left(\mathbf{c}^T(\mathbf{x}[0] - \hat{\mathbf{x}}[0|0]) + du[0] + n[0]\right)^2\right] \\
&= \mathbf{c}^T P_{0|0}\mathbf{c} + d^2\sigma_u^2 + \sigma_n^2
\end{aligned}$$



(a) Log-Likelihood Values at SNR = -5 dB, 0 dB, and 5 dB



(b) P_D versus time for SNR = -5 dB, 0 dB, and 5 dB

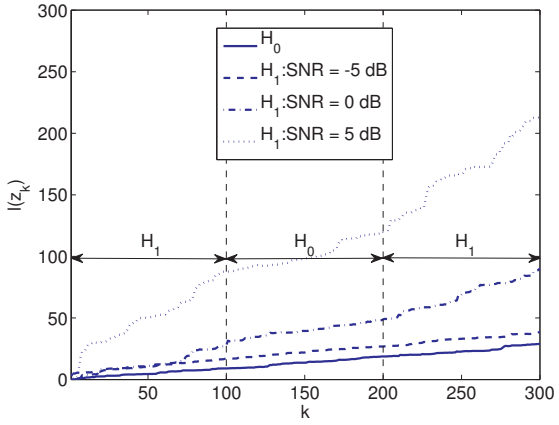
Figure 5.6: Log-Likelihood Values and Probability of Detection.

and therefore, the log-likelihood ratio is initialized at the value

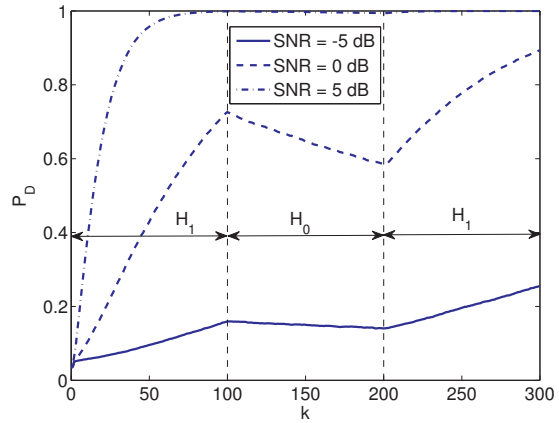
$$l\left(\mathbf{z}_0\right) = \left[\sigma_n^{-2} - \left(\mathbf{c}^T P_{0|0}\mathbf{c} + d^2\sigma_u^2 + \sigma_n^2\right)^{-1}\right] y[0]^2$$

This experiment is conducted 1500 times using a Monte Carlo study and a (time-dependent) threshold is determined corresponding to a fixed false alarm rate of 5%. Likewise, data is then generated from the H_1 hypothesis in a similar fashion and the probability of detection corresponding to this threshold is measured. The result of the simulation is shown in Figure 5.6(b). As we expect, the probability of detection

gets better the longer the system is time updated. We can also see that the smaller the SNR, the longer the system has to run before we can say with any confidence that hypothesis H_1 is in force. From Figure 5.6(a) this becomes clear as we can see that the larger the SNR, the faster the log-likelihood ratio grows. Thus, the distance between H_1 and H_0 increases at a faster rate leading to better detection performance at any instance in time.



(a) Log-Likelihood Values at SNR = -5 dB, 0 dB, and 5 dB



(b) P_D versus time for SNR = -5 dB, 0 dB, and 5 dB

Figure 5.7: LLR and P_D for Switched Hypothesis Model.

Now, the results of this simulation make the assumption that only one of the two hypotheses is always in force throughout the course of the experiment. To observe

what can happen when this assumption falls apart, we construct a similar situation where we generate data from the H_0 hypothesis and find a threshold corresponding to a false alarm probability of 5%. However, when looking at the H_1 hypothesis, we generate the first 100 samples (out of 300 samples) from the H_1 model and then the system is subsequently "switched" to the H_0 model for the next 100 realizations and finally the system is "switched" back to the H_1 model for the last 100 samples. Figure 5.7(a) displays the likelihood ratio for H_0 and H_1 at SNR values of -5, 0, and 5 dB. Likewise, Figure 5.7(b) gives P_D versus time for this simulation. As we can see, the plot looks identical to that shown in Figure 5.6(b) in the first 100 samples. However, for the next 100 samples our confidence in the H_1 hypothesis begins to degrade. Then, for the last 100 samples, our confidence begins to increase once more. We can also see from the figure that this degradation is relatively large for an SNR value of 0 dB but smaller for -5 dB and little if any at all degradation occurs at higher SNRs. Looking at Figure 5.7(a), this becomes clear as switching between hypotheses has little effect on the likelihood values associated with an SNR of -5 dB. This switching does have a large impact on the likelihood values at an SNR of 5 dB. However, the likelihood values at this SNR are so large after 100 samples of the time updating under hypothesis H_1 that switching the model has little effect on the difference among the LLR values under H_1 and those under H_0 thus having little effect on P_D.

5.3.2 Narrow-band, Single Source Detection with Multiple Disparate Linear Arrays-Channel Updating

Next, we turn our attention to the detection of a single, narrow-band source using multiple uniform linear arryas (ULAs) that mimic underwater target detection using multiple sonar platforms. We assume that each platform individually pings the environment and collects far-field measurements via an L-element ULA at half-wavelength spacing. For simulation, we consider a situation where 20 ULAs are oriented in the

same direction and are all above the seafloor at an elevation of $5m$. The platforms are uniformly spaced across a $40m$ distance in the cross-track direction and each platform is located $1m$ behind the platform to the left. We assume that the source is located at the origin and all platforms are moving in the same along-track direction with a speed of $1.5m/s$. Figure 5.8 gives a three-dimensional perspective of the problem being considered. The platform farthest to the left, which is labeled as $Platform$ 1, is taken to be our reference platform and the element farthest to the left in each ULA is taken to be the corresponding reference element. The narrow-band frequency is taken to be $10kHz$ and the speed of sound $1500m/s$ leading to an inter-element sensor spacing of $7.5cm$. There are 16 elements in each array which are sampling the environment at the Nyquist rate of $20kHz$.

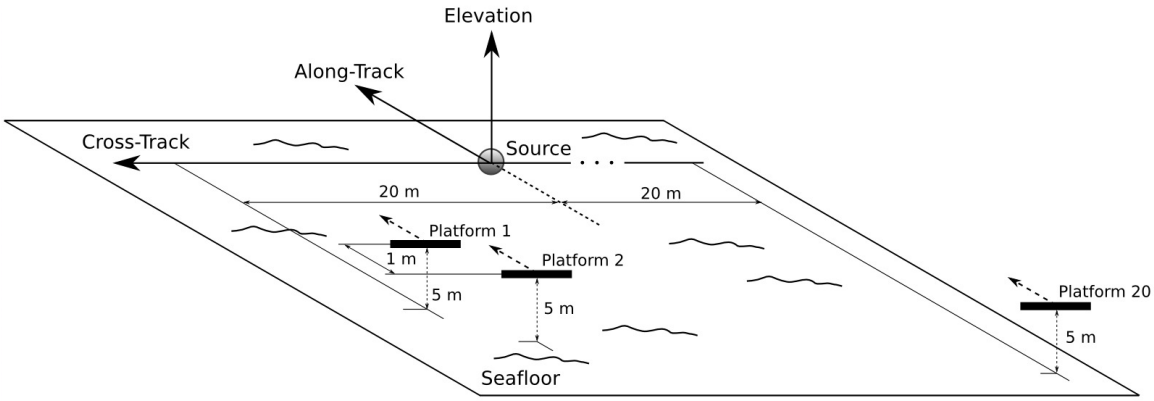


Figure 5.8: Multi-Platform Simulation Setup.

Assuming that each platform synchronously pings the environment with the same transmit signal via a global clock and each array knows the location of the others, then it may be possible in such a situation to account and equalize the effects of direct path propagation. We assume that such a situation exists and, for the k^{th} observation from the l^{th} array, consider the detection problem

$$\begin{aligned} H_1 &: \mathbf{y}_l[k] = \mathbf{h}\left(\theta_l\right)\boldsymbol{\alpha}_l^T\mathbf{s}[k] + \mathbf{n}_l[k] \\ H_0 &: \mathbf{y}_l[k] = \mathbf{n}_l[k] \end{aligned}$$

where $\mathbf{n}_l[k] \in \mathbb{C}^L$ is a zero-mean complex Gaussian random vector with covariance structure $E\left[\mathbf{n}_l[k]\mathbf{n}_i[k]^H\right] = \delta_{l-i}\sigma_n^2 I$ and

$$\mathbf{h}\left(\theta_l\right) = \begin{bmatrix} 1 \\ e^{-j\pi\cos(\theta_l)} \\ e^{-j2\pi\cos(\theta_l)} \\ \vdots \\ e^{-j(L-1)\pi\cos(\theta_l)} \end{bmatrix} \in \mathbb{C}^L$$

is the steering vector of the ULA at DOA θ_l and at half-wavelength spacing ($d = \lambda/2$). Assuming that there are N platforms in the sensing environment, the vector $\boldsymbol{\alpha}_l \in \mathbb{R}^N$ is given to be

$$\boldsymbol{\alpha}_l = \begin{bmatrix} \left(||\mathbf{r}_l|| + ||\mathbf{r}_1||\right)^{-1} \\ \left(||\mathbf{r}_l|| + ||\mathbf{r}_2||\right)^{-1} \\ \vdots \\ \left(||\mathbf{r}_l|| + ||\mathbf{r}_N||\right)^{-1} \end{bmatrix}$$

where each element is an attenuation (or fading) weighting for the signal emanated from the source and received by the l^{th} platform and $\mathbf{r}_i$, $i = 1, \ldots, N$, is a three-dimensional vector describing the platform's location with respect to the source. We assume for this simulation that all temporal delays are negligible. Finally, the vector $\mathbf{s}[k] \in \mathbb{R}^N$ is a zero-mean Gaussian random vector containing the source-signals from each platform at the k^{th} observation and has covariance matrix $R_{\mathbf{s}} = E\left[\mathbf{s}[k]\mathbf{s}[k]^T\right]$ given by

$$R_{\mathbf{s}} = \sigma_s^2 \begin{bmatrix} 1 & \rho(\phi_1 - \phi_2) & \cdots & \rho(\phi_1 - \phi_N) \\ \rho(\phi_2 - \phi_1) & 1 & \cdots & \rho(\phi_2 - \phi_N) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(\phi_N - \phi_1) & \rho(\phi_N - \phi_2) & \cdots & 1 \end{bmatrix}$$

where ϕ_l is the aspect angle among the source and the l^{th} platform and $\rho(\cdot)$ is a

correlation coefficient assumed to be a Gaussian function of the form

$$\rho(\phi_l - \phi_i) = e^{-\frac{(\phi_l - \phi_i)^2}{\Omega}}$$

where Ω is a parameter loosely describing the target structure. For example, if the target is spherical in nature and measurements are relatively invariant to aspect then one might choose a large value for the parameter Ω to reflect this. The geometrical setup of the model under H_1 is shown in Figure 5.9 for one platform.

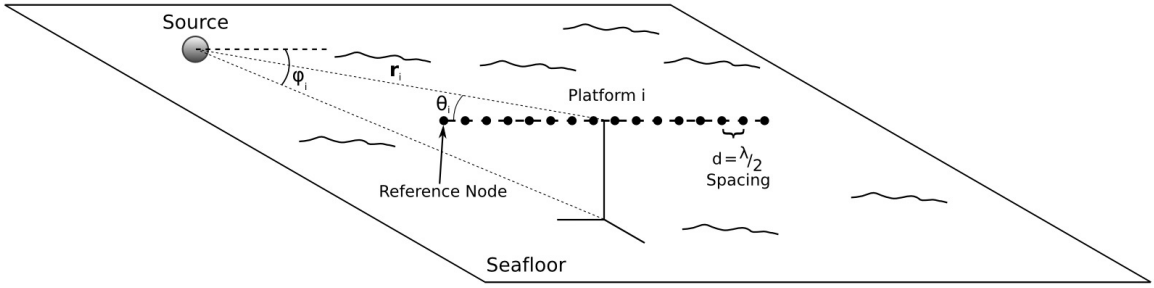


Figure 5.9: Geometry of Multi-Platform Model.

To perform channel updating as described in this chapter, we measure the observations from m arrays and form the composite observation $\mathbf{z}_m[k] = \left[\mathbf{y}_1[k]^T \ \mathbf{y}_2[k]^T \ \cdots \ \mathbf{y}_m[k]^T\right]^T$. We then wish to update the test statistic using the observation $\mathbf{y}_{m+1}[k]$. If we let H denote the matrix

$$H = \left[\boldsymbol{\alpha}_1 \mathbf{h}\left(\theta_1\right)^H \ \boldsymbol{\alpha}_2 \mathbf{h}\left(\theta_2\right)^H \ \cdots \ \boldsymbol{\alpha}_m \mathbf{h}\left(\theta_m\right)^H\right]^H$$

then the smoothing filters needed to update the log-likelihood ratio are given to be

$$\begin{aligned} W_0 &= \mathbf{O} \\ W_1 &= \mathbf{h}\left(\theta_{m+1}\right) \boldsymbol{\alpha}_{m+1}^T R_{\mathbf{s}} H \left(H R_{\mathbf{s}} H^H + \sigma_n^2 I\right)^{-1} \end{aligned}$$

The error covariance matrices associated with these smoothing filters are then given to be

$$\begin{aligned} Q_0 &= \sigma_n^2 I \\ Q_1 &= \sigma_n^2 I + \mathbf{h}\left(\theta_{m+1}\right) \boldsymbol{\alpha}_{m+1}^T \left(R_{\mathbf{s}} - R_{\mathbf{s}} H^H \left(H R_{\mathbf{s}} H^H + \sigma_n^2 I\right)^{-1} H R_{\mathbf{s}}\right) \boldsymbol{\alpha}_{m+1} \mathbf{h}\left(\theta_{m+1}\right)^H \end{aligned}$$

With this, the change in log-likelihood in (5.2) as a result of adding a new platform $m+1$ becomes

$$\Delta l\left(\mathbf{z}_{m+1},\mathbf{z}_m\right)=\left(1/\sigma_n^2\right)\mathbf{y}_{m+1}^H\mathbf{y}_{m+1}-\left(\mathbf{y}_{m+1}-W_1\mathbf{z}_m\right)^H Q_1^{-1}\left(\mathbf{y}_{m+1}-W_1\mathbf{z}_m\right)$$

Likewise, the error covariance matrices associated with filtering with a smoother that is mismatched to the given hypothesis are given to be

$$\begin{aligned}Q_{10} &= \sigma_n^2 I+\mathbf{h}\left(\theta_{m+1}\right)\boldsymbol{\alpha}_{m+1}^T R_{\mathbf{s}}\boldsymbol{\alpha}_{m+1}\mathbf{h}\left(\theta_{m+1}\right)\\ Q_{01} &= \sigma_n^2 I+\sigma_n^2\mathbf{h}\left(\theta_{m+1}\right)\boldsymbol{\alpha}_{m+1}^T R_{\mathbf{s}}H^H\left(HR_{\mathbf{s}}H^H+\sigma_n^2 I\right)^{-2}HR_{\mathbf{s}}\boldsymbol{\alpha}_{m+1}\mathbf{h}\left(\theta_{m+1}\right)^H\end{aligned}$$

The corresponding change in J-divergence given in (5.3) is then

$$\Delta J\left(\mathbf{z}_{m+1},\mathbf{z}_m\right)=\operatorname{tr}\left(-2I+Q_0^{-1}Q_{10}+Q_1^{-1}Q_{01}\right)$$

Assuming that our knowledge of the signal and noise power and the structure of the target (Ω) are known *a priori*, updating the likelihood ratio requires estimation of the direction of arrival DOA (θ_l), the aspect angle (ϕ_l), and the range ($||\mathbf{r}_l||$) from each platform to the target (see Figure 5.9). For this simulation example, the DOA is estimated across a 750-observation window using the conventional MUSIC algorithm [46] with a spectral resolution of $0.0628 rad$. We do not explicitly perform target localization for this simulation but instead model the estimate of the range as a random perturbation from its true value. More specifically, if $||\hat{\mathbf{r}}_l||$ is the estimate of the range and $\hat{\theta}_l$ is the DOA estimate from MUSIC, then $||\hat{\mathbf{r}}_l|| = ||\mathbf{r}_l|| + \epsilon$ where $\epsilon \sim N\left(0, ||\mathbf{r}_l||\left(\theta_l-\hat{\theta}_l\right)^2\right)$. Thus, the farther the array is from the source and the more error in the DOA estimate, the more the average error in the estimate of the range. Once the range from the source is estimated and assuming each platform knows its elevation from the seafloor, which we will denote as l, simple trigonometry and close inspection of Figure 5.9 shows that the aspect angle can then be found via the equation

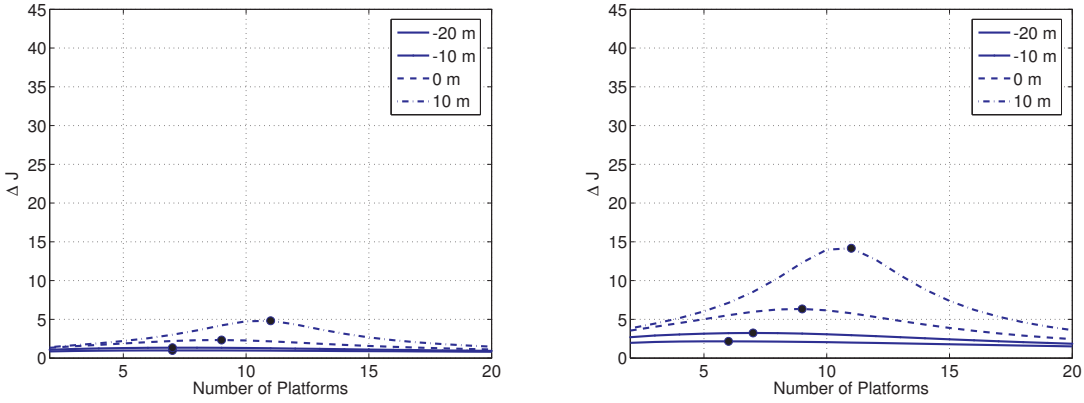
$$\hat{\phi}_i=\sin^{-1}\left(\frac{\sqrt{||\hat{\mathbf{r}}_i||^2\sin^2\hat{\theta}_i-l^2}}{\sqrt{||\hat{\mathbf{r}}_i||^2-l^2}}\right)$$

It is clear that these parameters will not remain unchanged over the 750-observation window used to estimate them. However, with the far-field condition along with a high sampling rate and low vehicle velocity, it is justifiable to assume that these parameters essentially remain the same over 750 measurements. Finally, the SNR value for this simulation example is taken to be σ_s^2/σ_n^2.

5.3.2.1 Systematic Channel Updating

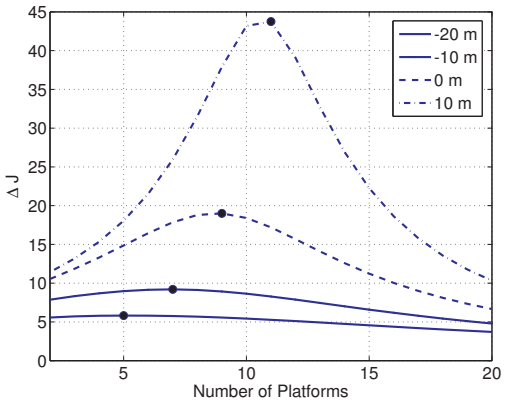
We begin by considering a simulation where we build a detector to handle data from $Platform$ 1 and subsequently perform updating to account for the addition of the data from $Platform$ 2 and so on until the data from all 20 arrays have been taken into account. Figures 5.10(a) and 5.11(a) display ΔJ and P_D corresponding to a false alarm rate of 5% at an SNR value of -5 dB when $Platform$ 1 is located at -20, -10, 0, and 10 m in the along-track direction. For these figures, 7500 Monte Carlo simulations were performed with different realizations of $\mathbf{s}[k]$ and $\mathbf{n}_l[k]$, $l = 1, \ldots, 20$. Recall that all platforms are traveling in the same direction with the same speed and so are always one meter behind the platform to the left as displayed in Figure 5.8.

Figures 5.10(b) and 5.11(b) as well as Figures 5.10(c) and 5.11(c) likewise display the same for SNR values of 0 and 5 dB, respectively. As can be observed, the performance of the detector always increases the closer the platforms approach the target irrespective of SNR. Additionally, at higher SNR values the gap between P_D plots at different along-track locations diminishes, especially as the number of platforms gets large. The change in J-divergence (i.e. discrimination ability) also always increases the closer we approach the target and at higher SNR values but does not always increase with the number of platforms. The point at which the change in J-divergence reaches its maximum value (denoted by a black dot in Figures 5.10(a)-(c) and Figures 5.11(a)-(c)) signifies the point of diminishing return as it is after this point that the J-divergence increases but at a *decreasing* rate. Comparing change in divergence



(a) SNR = -5 dB

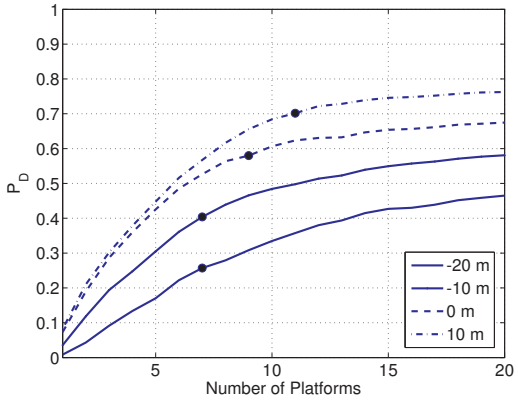
(b) SNR = 0 dB



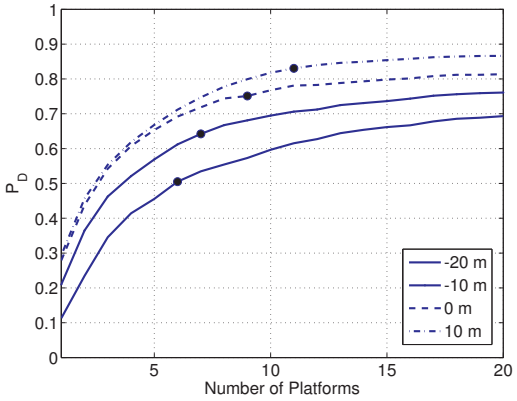
(c) SNR = 5 dB

Figure 5.10: Change in J-Divergence versus Number of Platforms.

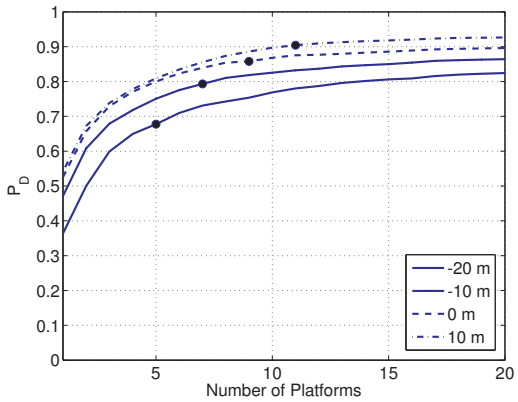
and P_D plots, one can see that the point at which the change in J-divergence is maximum generally correlates with the point where the increase in detection performance (i.e. P_D) begins to slow down. Comparing the plots in Figures 5.10(a) and 5.11(a) at -20 and -10 m to that shown in Figures 5.10(c) and 5.11(c) at 10 m, one can also observe that the faster the change in J-divergence diminishes, the faster the increase in probability of detection slows down. Therefore, when adding channels (platforms), we can see that the change in J-divergence gives us an effective measure for recognizing when adding an additional platform would not bring tangible improvement in detection performance and hence deciding when "enough is enough".



(a) SNR = -5 dB



(b) SNR = 0 dB



(c) SNR = 5 dB

Figure 5.11: Probability of Detection versus Number of Platforms.

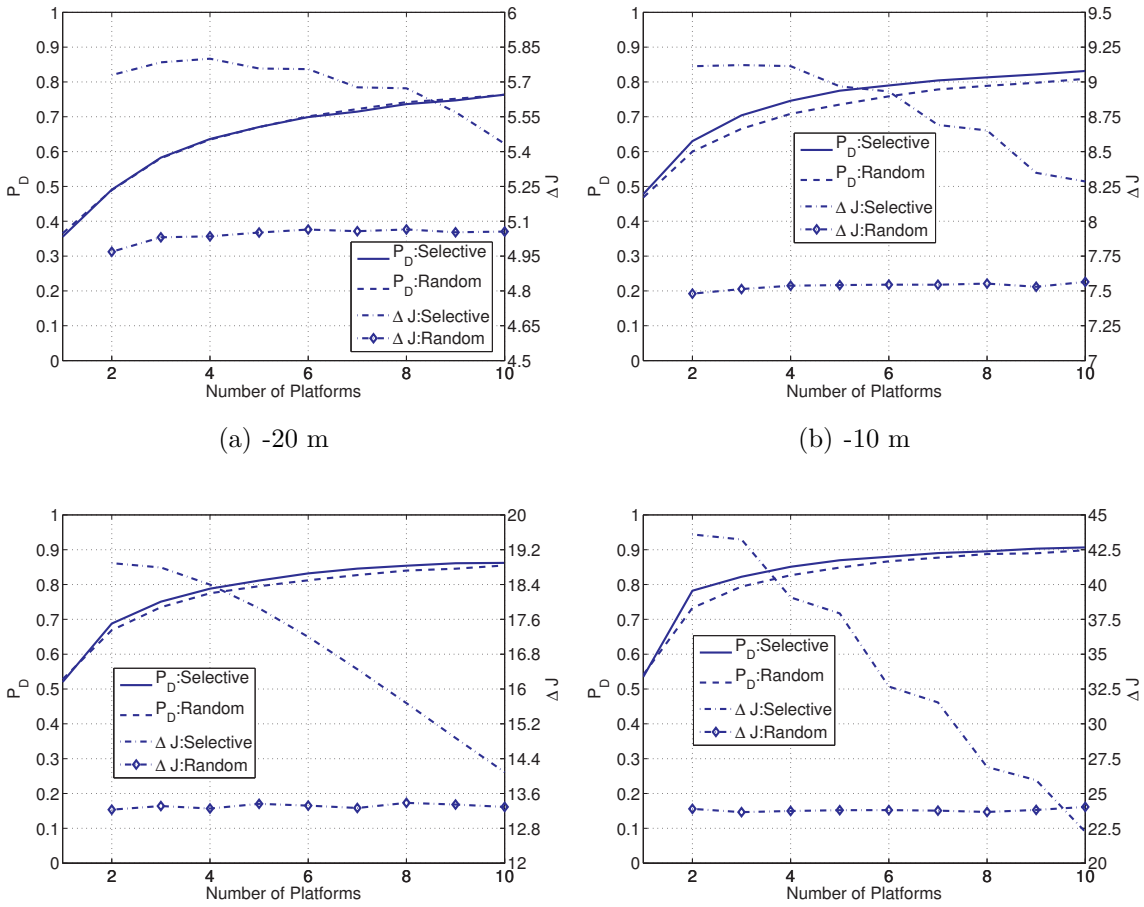
5.3.2.2 Selective Channel Updating

The previous simulation assumed that we can do detection with every platform in the operating environment of the source. Suppose now that we are not granted this luxury and can only use a subset of the total number of platforms available. However, we are given the opportunity to choose which platforms to use when performing detection. To decide which set of platforms to use, one could simply recursively search through all the platforms and choose to add that set of observations that yields the largest increase in J-divergence.

To simulate this situation, we consider the same setup shown in Figure 5.8 and

focus our attention on the detection system associated with $Platform$ 1. Our assumption is that the system can only handle observations from 10 platforms, the observation from the platform itself and nine others. Again, the system is initialized by building a detector to handle data from $Platform$ 1. We then search through all the other 19 platforms, measuring the increase in J-divergence that would be seen if we were to add the observation from that platform. We then choose the one that gives the largest increase in divergence. Then, the likelihood ratio is channel updated accordingly by augmenting that observation to the observation of $Platform$ 1. We then search through all the remaining 18 platforms and measure the increase in divergence that would be seen if we were to add the observation from that platform given the new augmented measurement. Again, we choose the one that gives the largest increase in J-divergence, incrementally update the likelihood ratio, and stack the observation from that platform with the augmented measurement from the previous iteration. This is continued until we have added the observations from nine other platforms. So the recursive nature of the problem is still the same as the previous simulation but, rather than channel update the likelihood ratio by adding the observation from the platform to the immediate right, we are now selectively choosing which observations are to be added. This selective platform-allocation scheme is compared to a scheme where platforms are chosen at random, i.e. we simply select 9 integers ranging from 2 to 20 which are uniformly sampled at random without replacement. We only consider the SNR = 5 dB case for this simulation.

Figures 5.12(a)-(d) display the results of this simulation in terms of P_D and ΔJ versus the number of platforms when $Platform$ 1 is located at -20, -10, 0, and 10 m in the along-track direction, respectively. Again the results shown in the figures are the average of 7500 Monte Carlo simulations where, again, different realizations of $\mathbf{s}[k]$ and $\mathbf{n}_l[k]$ are taken but also the choice of platforms when choosing at random also changes from simulation to simulation. We can see that, when selectively (according



(a) -20 m (b) -10 m

(c) 0 m (d) 10 m

Figure 5.12: ΔJ and P_D versus Number of Platforms - Different $Platform$ 1 locations in along-track.

to ΔJ) adding platforms, the change in J-divergence starts out large and subsequently decreases at a faster rate compared to that when we choose at random which begins at a lower change in J-divergence and remains fairly constant throughout. The constancy of the change in J-divergence for random platform selection is clearly the result of averaging over all Monte Carlo simulations and gives us an idea of what we can *expect* when choosing platforms in a random fashion. We can observe in Figure 5.12(a) that it makes little difference whether platforms are chosen selectively or at random when the platforms are far from the source. As the platforms approach the source in

Figures 5.12(b)-(d), we can see that the performance of the detector associated with selective allocation always exceeds that of random platform selection. We can also see that the difference in P_D among selective and random platform allocation diminishes as the number of platforms grows large. Thus, if we can only take advantage of a small number of platforms, the selective allocation scheme can give a significant increase in detection performance compared to a situation where platforms are arbitrarily chosen, i.e. at random. For a given P_D, we can also see that selectively choosing platforms generally requires a smaller number of platforms than when we choose at random. For example, at an along-track location of -10 m, Figure 5.12(b) shows us that we only require 7 platforms to achieve $P_D = 80\%$ whereas we need 9 when choosing platforms at random. Therefore, we again see that the recursive framework of log-likelihood time and channel updating and the corresponding change in J-divergence can be effective tools for deciding which platforms we wish to use when performing detection in such a framework.

5.4 Conclusion

In this chapter, we first reviewed a well known property of the log-likelihood ratio when adding data in the detection process. This led to an updating process in which the old likelihood ratio is incremented by adding an additional term involving probability densities conditioned upon past measurements. Noting this property, we then investigated what this updating term looks like for the Gauss-Gauss detector and found that it involves linearly estimating the new data from the old and forming quadratic terms in the error of the estimates under both hypotheses. Updating the likelihood ratio in this fashion opens up a wide range of practical possibilities, two of which are temporal and channel updating schemes. In temporal updating, we incrementally form the likelihood ratio as time progresses, forgoing the hassles of finding the test statistic for the time series in its entirety, and can systematically propagate

the errors and error covariances needed for the incremental update. In channel updating, we build a detector to handle observations from one channel and incrementally update the likelihood ratio with the observations from all the other channels. The change in J-divergence seen when adding new data can then be written in terms of error covariance matrices when filtering with a smoother that is matched/miss-matched to the given hypothesis. A reduced-rank version of the log-likelihood ratio update is then developed which can be of great advantage when the data we wish to add lies in a high dimensional space and it is desirable to perform updating with a lower-dimensional approximation without sacrificing performance. In this new coordinate system built for low-rank updating, we found that the change in J-divergence becomes decomposed in terms of a scalar term representing the per-mode mean-squared error under H_1 when using the right smoothing filter and two quadratic terms representing the mean-squared errors in that coordinate under both the H_1 and H_0 hypotheses when using the wrong smoothing filter. Taking advantage of the J-divergence in this coordinate system and a pair of inequalities resulting from sub-optimal filtering statements, we found a lower bound for the change in J-divergence which allows us to conclude that adding data can never cause a decrease in J-divergence.

To give a situation where the time log-likelihood ratio updating may be useful, we presented an example involving the detection of an underlying dynamical system in our measurements. Because of the structure of the problem, we take advantage of the Kalman filter recursions for propagating the necessary innovation and innovation variance needed for this time updating under hypothesis H_1 as the Kalman filter reduces to that of the Wiener filter when steady-state conditions have been reached. Thus, we do not have to find the necessary smoothing filters and error covariances at each instance in time as the Kalman filter produces them at each iteration and hence the detection system can easily be implemented in real-time. We then generated an arbitrary LTI system and provided simulations for three different SNRs. We observed

that, at lower SNR values it takes longer to achieve a desired P_D at a given false alarm rate. We then constructed a situation where the model is switched between the H_1 and H_0 hypothesis and showed what can happen to the performance of the detector when one model does not exclusively generate the measurements that we have used.

We next looked at another simulation for the channel updating of the log-likelihood ratio where we are interested in detecting the presence of a single narrow-band source in the sensing environment of multiple ULAs. We then presented the physical model which generated our measurements and formed our multi-platform detection hypothesis. For the simulation, we built a detector for one platform and subsequently performed channel updating to successively account for the observations from all the existing platforms. It was observed that adding platforms in this fashion results in a change in J-divergence which at first increases but then begins to decrease. This exemplifies the power of this information measure as an effective tool for helping us decide when to stop employing additional platforms to aid the detector. We next considered a case where we can perform detection based only on the observations from a subset of the total number of platforms. We again built a detector for one platform but then performed the channel updating by finding the platform whose observation gives the largest increase in J-divergence after data augmentation. This is compared to an alternative scenario where platforms are chosen at random. In this case, we observed that, for a small number of platforms at close range to the source, selectively choosing which platforms to use outperforms that of choosing at random. For a given P_D, we also found that selectively choosing platforms generally requires a smaller number of platforms than when we choose at random. This simulation again exemplifies the effectiveness of the change in J-divergence for allocating sensing resources when performing detection with multiple disparate platforms.

CHAPTER 6

CONCLUSIONS AND SUGGESTIONS FOR FUTURE WORK

6.1 Conclusions and Discussions

The problem of multi-sensory underwater object detection 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. To allow decision-making among multiple sonar platforms, it is essential to detect and further scrutinize the information bearing parts of the data collected by the various sensory systems. This work introduced a solution to this problem by proposing a new multi-channel, multi-sensory binary hypothesis detection system using the Multi-Channel Coherence Analysis (MCA) framework which is a natural extension of Canonical Correlation Analysis (CCA) to more than two channels. Similar to two-channel CCA, the objective of MCA is to discover the coherence among N channels of data by searching for one-dimensional mapping vectors that maximizes the sum of the cross correlations among every pairwise combination of channels. Solving the optimization problem proposed by MCA leads to a generalized eigenvalue problem involving the composite covariance matrix of the N channels and its corresponding block-diagonal matrix.

We then presented a multi-channel signal-plus-noise model that fits well with the problem at hand. The standard Gauss-Gauss detector of Section 2.3 is then recast into the MCA framework by representing the data in the MCA coordinate system under the H_1 hypothesis. The MCA-based detector led to a new log-likelihood function

involving the sum of the correlations matrix, Λ_1, the MCA orthonormal mapping matrix, P_1, and the *local* signal-to-noise ratio matrix, Σ, a block-diagonal matrix composed of the SNR matrices for each channel. The J-divergence for this MCA-based detector again becomes decomposed in terms of the same matrices. In the MCA coordinate system, it seems that the detector decomposes the information needed for detection into the coherence shared between data channels under H_1 and the SNR information among the individual channels themselves.

With the assumption of poor SNR among each individual data channel, which is found to be valid using actual data from multiple sonar, we implemented a detector that only takes advantage of the sum of the cross-correlations of the data in the mapped domain under the H_1 hypothesis. This detector is then applied to two real data sets consisting of one high frequency (HF) and one to three broadband (BB) sonar images and a data set consisting of simulated target and non-target shapes embedded in simulated background.

For the multiple sonar image data set consisting of one HF and three BB sonar images, three detection systems are configured with the HF image along with different combinations of the BB images to take advantage of the high target definition capabilities of HF sonar and the clutter suppression ability of BB sonar. The two-sonar detection system performs well detecting 51 out of 53 total targets with a false alarm rate of 7.48 false detections per image. The three-sonar detector improves upon this result, detecting 52 of the targets while maintaining 8.93 false alarms per image. The four-sonar detection system likewise detects 52 of the 53 targets but exhibits a higher false alarm rate of 9.32 false alarms per image. This lack of tangible increase in detection performance when going from three to four sonar images suggested that the point of diminishing return is a system with one HF image along with two BB images as the addition of the third BB observation did not bring any new information to the detector.

For the dual sonar image data set, a two-channel MCA-based detection system was configured with one HF image along with one BB image. This data set consists of both target and man-made (lobster trap) non-target objects. A test conducted on all the target and non-target objects, as well as an equally sized set of background data, showed promise for using the multi-channel correlations for the purposes of classifying target from non-target. Based on the target and non-target objects as well as background features from a partial subset of images, a suitable detection threshold is determined and used on the remaining data set. The detection threshold led to detecting all 49 target and non-target objects in the tested data set while maintaining 7 false alarms per image, which is an excellent detection performance given the fact that the data lacked phase information and was more than likely non-Gaussian in nature.

Finally, the detection method is applied to a simulated data set to study the MCA-based detector's sensitivity to variables that are encountered in multi-channel target detection problems. The images are generated with different target types and at various SNR values, image resolutions, ranges, aspect angles, etc. To mimic detection in single-platform, multi-sonar situations, we built a two-channel detector where each image contains the same target type at the same SNR, range, and aspect angle. However, the two images differed in resolution: one at $1in$ and the other at $3in$ resolution. The results were then partitioned on the basis of target type and SNR and ROC curves were plotted. The results indicated robustness of the detection method to different target structures at various SNR values. Next, to mimic multi-platform, single-sonar situations, we again built a two-channel detector where each image consisted of the same target type at the same SNR ($9dB$), range, and image resolution ($1in$). However, this time the two channels differed in aspect angle and pairs of images were chosen such that the difference in aspect angle was within the range of a particular aspect separation. ROC curves and the empirical J-divergence

were then plotted for two different target types at various aspect separations. The results again demonstrated the detection method's robustness to aspect separation.

Finally, we investigated how adding an additional observation to the Gauss-Gauss detector changes the log-likelihood ratio and found that the likelihood function can be updated by adding a term involving linearly estimating the new data from the old and forming quadratic terms in the error of the estimates under both hypotheses. Updating the likelihood ratio in this fashion opens up a wide range of practical possibilities, two of which are temporal and channel updating schemes. The change in J-divergence seen when adding new data can then be written in terms of error covariance matrices when filtering with a smoother that is matched/miss-matched to the given hypothesis. A reduced-rank version of the log-likelihood ratio update is then developed which can be of great advantage when the data we wish to add lies in a high dimensional space and it is desirable to perform updating with a lower-dimensional approximation without sacrificing performance. In this new coordinate system built for low-rank updating, we found that the change in J-divergence becomes decomposed in terms of a scalar term representing the per-mode mean-squared error under H_1 when using the right smoothing filter and two quadratic terms representing the mean-squared errors in that coordinate under both the H_1 and H_0 hypotheses when using the wrong smoothing filter. Taking advantage of the J-divergence in this coordinate system and a pair of inequalities resulting from sub-optimal filtering statements, we found a lower bound for the change in J-divergence which led us to the conclusion that adding data can never cause a decrease in J-divergence.

We then provided two simulations to exhibit situations where log-likelihood ratio updating may be useful. The first simulation involved the detection of an underlying dynamical system in our measurements and exhibited a time updating scenario. Knowing that the Kalman filter reduces to that of the Wiener filter at steady state,

we took advantage of the Kalman filter recursions for propagating the necessary innovation and innovation variance needed for time updating under hypothesis H_1. We then generated an arbitrary LTI system for the simulation and observed that, at lower SNR values, it takes longer to achieve a desired P_D at a given false alarm rate. We then constructed a situation where the model is switched between the H_1 and H_0 hypotheses and showed what can happen to the performance of the detector when one model does not exclusively generate the measurements that we have used. The second simulation is interested in detecting the presence of a single narrow-band source in the sensing environment of multiple Uniform Linear Arrays (ULAs) and exhibited a channel updating scenario. We then built a detector for one platform and recursively updated the likelihood ratio with the observation from the platform to the immediate right until the observations from all the existing platforms had been taken into account. It was observed that adding platforms in this fashion results in a change in J-divergence which at first increases but then begins to decrease. This exemplifies the power of this information measure as an effective tool for helping us decide when to stop employing additional platforms to aid the detector. We then considered a case where we could perform detection based only on the observations from a subset of the total number of platforms. We again built a detector for one platform but then performed the channel updating by finding the platform whose observation gave the largest increase in J-divergence after data augmentation. This was compared to an alternative scenario where platforms were chosen at random. In this case, we observed that, for a small number of platforms, selectively choosing which platforms to use outperformed that of choosing at random, especially at close ranges from the source. For a given P_D, we also found that selectively choosing platforms generally required a smaller number of platforms than when we choose at random. This simulation again exemplified the effectiveness of the change in J-divergence for allocating sensing resources when performing detection with multiple disparate platforms.

6.2 Future Work

Although, the MCA-based detector proposed in this thesis offers a powerful tool for detection of underwater targets from multiple disparate sonar platforms, there are several important areas and extensions that can be pursued in the future. These include, but are not limited to:

- The data used in this study was limited to only a few runs and types of underwater targets. Ideally, the next step in the development of the MCA-based detector would be to test the performance on more data to prove the usefulness of the detection systems developed in this thesis. The testing on more difficult data sets provided by the NSWC as well those including more man-made non-targets will 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 probability of detection and false alarm rate would be insightful and help to illustrate the real effectiveness of the detector for realistic underwater target detection problems.

- The main development of this thesis was focused on the detection of underwater targets from sonar imagery. Although it was observed that there may be suitable discriminatory information in the multi-channel correlation features among targets and non-targets (in this case lobster traps), no specific study on this observation was conducted. Another potential extension of this research would be to study the use of multi-channel correlation features for classification of targets and non-targets. If successful, this will allow us to carry out simultaneous detection and classification using only the extracted MCA coordinates and sum of correlations without requiring a separate feature extraction system.

- The MCA-based detector developed in this thesis is applicable not only to sonar image detection, but could be used on other disparate sensory systems,

i.e. magnetic, infrared, and optical. A study of its usefulness on these types of sensing modalities would be highly valuable. By finding the coherence information between more than one type of sensors, the detection and classification performance could be improved.

- Extension of the coherence-based detection method to account for the multi-hypothesis testing in the disparate sensor detection problem. For the multi-platform detection problem, additional hypotheses must be added in the formulations to detect the targets that don't appear very well in either the high-frequency or broadband sonar images. This requires extending the Gauss-Gauss detector to M-hypothesis testing [47] problem and relating the corresponding J-divergence detectability measure to the multi-channel correlations of the sonar imagery data.

- Similar to that presented in [1], investigate distributed detection systems which include collaboration between local decision makers to yield global detection decisions with higher confidence. More specifically, develop a new collaborative distributed detection methodology that takes into account (a) limited communication bandwidth for relaying essential target information among multiple disparate AUVs and the mother ship; (b) computational limitations of the processing systems on each AUV platform; (c) real-time decision requirements; (d) practically feasible, versatile, and robust implementation of the of MCA-bsed detection method developed in this thesis.

- More investigation is needed for the application of updating the log-likelihood ratio specifically for temporal and channel updating procedures. For the time updating problem, the validity of using the Kalman filter as a means of propagating the necessary information needed for updating must be proven. If it is not the case, methods for properly and efficiently propagating the necessary Wiener

smoothing and error covariance matrices must be developed to yield a recursive system that can be implemented in real-time. For the channel updating problem, a more in-depth study on the change in J-divergence seen when adding sonar images with varying degrees of disparity must be conducted. Orthogonal projection updating [48] for both temporal and channel updating procedures should also be investigated as a means of effectively incrementing the likelihood ratio in both situations.

- An extension of the Generalized Likelihood Ratio Test (GLRT) detector proposed in [11] to multi-variate Gaussian time series should be investigated to generalize the idea for multiple sonar array applications. The test statistic should then be extended to the frequency domain involving a composite matrix composed of the auto and cross power spectral density matrices associated with all N channels and its properties explored. Accordingly, an in-depth study of the response of this GLRT to varying degrees of disparity must also be conducted.

REFERENCES

[1] J. D. Tucker, "Coherence-based underwater target detection for side-scan sonar imagery," *Master's Thesis, Colorado State University*, 2009.

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

[3] B. Thompson, "Multi-channel coherence analysis for feature extraction with application to buried underwater target classification," *Master's Thesis, Colorado State University*, 2007.

[4] A. Neilsen, "Multiset canonical correlations analysis and multispectral, truly multitemporal remote sensing data," *IEEE Transactions on Image Processing*, vol. 11, pp. 293–305, March 2002.

[5] M. A. Hasan, "On multi-set canonical correlation analysis," *Proceedings of International Joint Conference on Neural Networks*, pp. 1128–1133, 2009.

[6] B. D. Thompson and M. R. Azimi-Sadjadi, "Iterative multi-channel coherence analysis with applications," *Neural Networks*, vol. 21, no. 2-3, pp. 493–501, 2008.

[7] 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.

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

[9] 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.

[10] 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.

[11] D. Ramirez, J. Via, I. Santamaria, and L. Scharf, "Detection of spatially correlated gaussian time series," *Submitted to IEEE Transactions on Signal Processing*, 2010.

[12] J. Dixon, "How good is hadamard's inequality for determinants?" *Canadian Mathematical Bulletin*, vol. 27, no. 3, pp. 260–264, 1984.

[13] L. Scharf and C. Mullis, "Canonical coordinates and the geometry of inference, rate, and capacity," *IEEE Transactions on Signal Processing*, vol. 48, no. 3, pp. 824–891, March 2000.

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

[15] 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.

[16] A. Pezeshki, M. Azimi-Sadjadi, L. Scharf, and M. Robinson, "Underwater target classification using canonical correlations," *Proc. of MTS/IEEE Oceans 2003 Conference*, vol. 4, pp. 1906–1911, Sept 2003.

[17] 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.

[18] 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.

[19] J. Cartmill, N. Wachowski, and M. R. Azimi-Sadjadi, "Buried underwater object classification using a collaborative multi-aspect classifier," *IEEE Journal of Oceanic Engr.*, vol. 34, no. 1, pp. 32–44, 2009.

[20] N. Wachowski and M. R. Azimi-Sadjadi, "Buried underwater target classification using frequency subband coherence analysis," *Proc. of MTS/IEEE Oceans 2008 Conference*, to be published.

[21] M. R. Azimi-Sadjadi and J. D. Tucker, "Target detection from dual disparate sonar platforms using canonical correlations," *Proc. SPIE*, vol. 6953, pp. J1 – J10, March 2008.

[22] J. D. Tucker, M. R. Azimi-Sadjadi, and G. J. Dobeck, "Coherent-based method for detection of underwater objects from sonar imagery," *Proc. SPIE*, vol. 6553, pp. U1–U8, April 2007.

[23] J. D. Tucker, M. R. Azimi-Sadjadi, and G. J. Dobeck, "Canonical coordinates for detection and classification of underwater objects from sonar imagery," *Proc. of IEEE OCEANS 2007 Conference Europe*, pp. 1–6, June 2007.

[24] J. D. Tucker, N. Klausner, and M. R. Azimi-Sadjadi, "Target detection in m-disparate sonar platforms using multichannel hypothesis testing," *Proc. of MTS/IEEE Oceans 2008 Conference*, to be published.

[25] L. L. Scharf and B. D. Van Veen, "Low rank detectors for Gaussian random vectors," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 11, pp. 1579–1582, Nov 1987.

[26] T. Kailath, A. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall Information and System Sciences Series, 2000.

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

[28] P. K. Varshney, *Distributed Detection and Data Fusion*. Springer-Verlag New York, 1996.

[29] M. L. Eaton, *Multivariate Statistics: A Vector Space Approach*. New York: Wiley, 1983, ch. 10.

[30] J. T. Cobb, K. C. Slatton, and G. J. Dobeck, "A parametric model for characterizing seabed textures in synthetic aperture sonar images," *To Appear in IEEE Journal of Oceanic Engineering*, 2010.

[31] E. L. Lehman, *Testing Statistical Hypotheses*. New York: Wiley, 1986.

[32] J. Gubner, *Probability and Random Processes for Electrical and Computer Engineers*. Cambridge University Press, 2006.

[33] A. Rencher, *Methods of Multivariate Analysis*, 2nd ed. Wiley-Interscience, 2002.

[34] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Third ed. John Hopkins University Press, 1996.

[35] R. Gonzalez and R. Woods, *Digital Image Processing*, 3rd ed. Pearson Prentice Hall, 2008.

[36] 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.

[37] G. J. Dobeck, "Fusing sonar images for mine detection and classification," *Proc. SPIE*, vol. 3710, pp. 602–614, April 1999.

[38] 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.

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

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

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

[42] N. Klausner, M. R. Azimi-Sadjadi, and J. D. Tucker, "Underwater target detection from multi-platform sonar imagery using multi-channel coherence analysis," *Proc. of SMC 2009 Conference*, to be published.

[43] N. Klausner, J. Tucker, and M. Azimi, "Multi-platform target detection using multi-channel coherence analysis and robustness to the effects of disparity," *Proc. of MTS/IEEE Oceans 2008 Conference*, 2009.

[44] L. Scharf and L. Nolte, "Likelihood ratios for sequential hypothesis testing on markov sequences," *IEEE Transactions on Information Theory*, vol. 23, no. 1, pp. 101–109, 1977.

[45] F. Schweppe, "Evaluation of likelihood functions for gaussian signals," *IEEE Transactions on Information Theory*, pp. 61–70, 1965.

[46] M. Hayes, *Statistical Digital Signal Processing and Modeling.* John Wiley and Sons, Inc., 1996.

[47] E. Drakopoulos, J. J. Chao, and C. C. Lee, "A two-level distributed multiple hypothesis decision system," *IEEE Trans. on Automatic Control*, vol. 37, no. 3, pp. 380–384, Mar 1992.

[48] S. Alexander, *Adaptive Signal Processing: Theory and Applications*. Springer-Verlag, 1986.

[49] A. Pezeshki, L. L. Scharf, M. R. Azimi-Sadjadi, and Y. Hua, "Two-channel constrained least squares problems: Solutions using power methods and connections with canonical coordiates," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 121–135, Jan 2005.

APPENDIX A

CANONICAL CORRELATION ANALYSIS (CCA) REVIEW

In this section, we provide a review of the CCA method in which a set of basis vectors is found for two sets of multidimensional variables such that correlations between the projections onto these basis vectors are mutually maximized. CCA was proposed by Hotelling [14] for the analysis of linear dependence between two data channels. CCA decomposes the linear dependence between the original channels into the linear dependence between the canonical coordinates of the channels, where this linear dependence is easily determined by the corresponding canonical correlations.

Consider the composite data vector $\mathbf{z}$ consisting of two random vectors $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^n$, i.e.

$$\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \in \mathbb{R}^{(m+n)}. \tag{A-1}$$

For the remainder of the derivations, it is assumed that $m \geq n$, also the notation $(\cdot)^H$ represents the Hermitian operation. Assume that $\mathbf{x}$ and $\mathbf{y}$ have zero means and share the composite covariance matrix

$$R_{zz} = E[\mathbf{z}\,\mathbf{z}^H] = E\left[\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \begin{pmatrix} \mathbf{x}^H & \mathbf{y}^H \end{pmatrix} \right] = \begin{bmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{bmatrix}. \tag{A-2}$$

If $\mathbf{x}$ and $\mathbf{y}$ are now replaced by their corresponding whitened vectors, then the composite vector $\boldsymbol{\xi}$ is

$$\boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{\zeta} \\ \boldsymbol{\nu} \end{bmatrix} = \begin{bmatrix} R_{xx}^{-1/2} & \mathbf{0} \\ \mathbf{0} & R_{yy}^{-1/2} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}, \tag{A-3}$$

where $R_{xx}^{1/2}$ is a square-root of R_{xx} with $R_{xx}^{1/2}R_{xx}^{H/2} = R_{xx}$ and $R_{xx}^{-1/2}R_{xx}R_{xx}^{-H/2} = \mathrm{I}$. The covariance matrix of $\boldsymbol{\xi}$ may be written as

$$R_{\xi\xi} = E[\boldsymbol{\xi}\,\boldsymbol{\xi}^H] = E\left[\begin{pmatrix}\boldsymbol{\zeta}\\ \boldsymbol{\nu}\end{pmatrix}\begin{pmatrix}\boldsymbol{\zeta}^T & \boldsymbol{\nu}^H\end{pmatrix}\right] = \begin{bmatrix} R_{\zeta\zeta} & R_{\zeta\nu}\\ R_{\nu\zeta} & R_{\nu\nu}\end{bmatrix} = \begin{bmatrix}\mathrm{I} & \mathrm{C}\\ \mathrm{C}^H & \mathrm{I}\end{bmatrix}, \tag{A-4}$$

where

$$\mathrm{C} = E[\boldsymbol{\zeta}\boldsymbol{\nu}^T] = E[(\mathrm{R}_{xx}^{-1/2}\mathbf{x})(\mathrm{R}_{yy}^{-1/2}\mathbf{y})^H] = \mathrm{R}_{xx}^{-1/2}\mathrm{R}_{xy}\mathrm{R}_{yy}^{-H/2} \tag{A-5}$$

is called the *coherence matrix* of $\mathbf{x}$ and $\mathbf{y}$ [13], [49]. Therefore, the coherence matrix C is the cross-covariance matrix between the whitened versions of $\mathbf{x}$ and $\mathbf{y}$. Correspondingly, the coordinates $\boldsymbol{\zeta}$ and $\boldsymbol{\nu}$ are called the *coherence coordinates*. Now it is possible to determine the singular value decomposition (SVD) of the coherence matrix, namely

$$\begin{aligned}\mathrm{C} &= R_{xx}^{-1/2}R_{xy}R_{yy}^{-H/2} = \mathrm{FKG}^H \quad \text{and}\\ \mathrm{F}^H\mathrm{CG} &= \mathrm{F}^H\mathrm{R}_{xx}^{-1/2}R_{xy}R_{yy}^{-T/2}\mathrm{G} = \mathrm{K},\end{aligned} \tag{A-6}$$

where $\mathrm{F} \in \mathbb{R}^{m\times m}$ and $\mathrm{G} \in \mathbb{R}^{n\times n}$ are orthogonal matrices [34], i.e.

$$\mathrm{F}^H\mathrm{F} = \mathrm{FF}^H = \mathrm{I}(m) \quad \text{and} \quad \mathrm{G}^H\mathrm{G} = \mathrm{GG}^H = \mathrm{I}(n), \tag{A-7}$$

and

$$\mathrm{K} = \begin{bmatrix}\mathrm{K}(n)\\ \mathbf{0}\end{bmatrix} \in \mathbb{R}^{m\times n} \tag{A-8}$$

is a diagonal singular value matrix, with $\mathrm{K}(n) = \mathrm{diag}[k_1, k_2, \ldots, k_n]$ and $1 \geq k_1 \geq k_2 \geq \ldots \geq k_n > 0$.

We then use the orthogonal matrices F and G to transform the whitened composite vector $\boldsymbol{\xi}$ into the canonical composite vector $\mathbf{w}$,

$$\mathbf{w} = \begin{bmatrix}\mathbf{u}\\ \mathbf{v}\end{bmatrix} = \begin{bmatrix}\mathrm{F}^H & \mathbf{0}\\ \mathbf{0} & \mathrm{G}^H\end{bmatrix}\begin{bmatrix}\boldsymbol{\zeta}\\ \boldsymbol{\nu}\end{bmatrix} = \begin{bmatrix}\mathrm{F}^H & \mathbf{0}\\ \mathbf{0} & \mathrm{G}^H\end{bmatrix}\begin{bmatrix}R_{xx}^{-1/2} & \mathbf{0}\\ \mathbf{0} & R_{yy}^{-1/2}\end{bmatrix}\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix}. \tag{A-9}$$

Then, the covariance matrix for the canonical composite vector $\mathbf{w}$ is obtained as

$$R_{ww} = E[\mathbf{w}\mathbf{w}^H] = E\left[\begin{pmatrix}\mathbf{u}\\ \mathbf{v}\end{pmatrix}\begin{pmatrix}\mathbf{u}^H & \mathbf{v}^H\end{pmatrix}\right] = \begin{bmatrix}\mathrm{R}_{uu} & R_{uv}\\ R_{vu} & R_{vv}\end{bmatrix} = \begin{bmatrix}\mathrm{I} & \mathrm{K}\\ \mathrm{K}^H & \mathrm{I}\end{bmatrix}. \tag{A-10}$$

The elements of $\mathbf{u} = [u_i]_{i=1}^m \in \mathbb{R}^m$ are referred to as the *canonical coordinates* of $\mathbf{x}$ and the elements of $\mathbf{v} = [v_i]_{i=1}^n \in \mathbb{R}^n$ are the canonical coordinates of $\mathbf{y}$. The diagonal cross-correlation matrix K,

$$K = E[\mathbf{u}\mathbf{v}^H] = E[(\mathrm{F}^H\mathrm{R}_{xx}^{-1/2}\mathbf{x})(\mathrm{G}^H\mathrm{R}_{yy}^{-1/2}\mathbf{y})^H] = \mathrm{F}^H\mathrm{CG} \tag{A-11}$$

is called the *canonical correlation matrix* of *canonical correlations* k_i, with $1 \geq k_1 \geq k_2 \geq \cdots \geq k_n > 0$. Thus, the canonical correlations measure the correlations between pairs of corresponding canonical coordinates. That is, $E[u_i v_j] = k_i\delta_{ij};\ i \in [1,n]$, $j \in [1,m]$, with δ_{ij} being the Kronecker delta. The canonical correlations k_i are also the singular values of the coherence matrix C. Correspondingly, KK^H is the squared canonical correlation matrix of the squared canonical correlations k_i^2. Since F and G are orthogonal matrices, we may write the squared coherence matrix CC^H as

$$\begin{aligned}\mathrm{CC}^H &= R_{xx}^{-1/2}R_{xy}R_{yy}^{-1}R_{yx}R_{xx}^{-H/2}\\ &= \mathrm{FKG}^H\mathrm{GK}^H\mathrm{F}^H = \mathrm{FKK}^H\mathrm{F}^H.\end{aligned} \tag{A-12}$$

This shows that the squared canonical correlations k_i^2 are the eigenvalues of the squared coherence matrix CC^H, or equivalently, of the matrix $R_{xx}^{-H/2}\mathrm{CC}^H R_{xx}^{H/2} = R_{xx}^{-1}\mathrm{R}_{xy}\mathrm{R}_{yy}^{-1}\mathrm{R}_{yx}$. It is interesting to note that these eigenvalues are invariant to the choice of a square-root for R_{xx}.

Figure A-1 illustrates the transformation from standard coordinates $\mathbf{x}$ and $\mathbf{y}$ to coherence coordinates $\boldsymbol{\zeta}$ and $\boldsymbol{\nu}$ and then to canonical coordinates $\mathbf{u}$ and $\mathbf{v}$. It can be noted that the transformation from standard coordinates $\mathbf{x}$ and $\mathbf{y}$ to canonical coordinates $\mathbf{u}$ and $\mathbf{v}$ can be represented by $\mathbf{u} = \mathrm{W}^H\mathbf{x}$ and $\mathbf{v} = \mathrm{D}^H\mathbf{y}$ where $\mathrm{W}^H = \mathrm{F}^H R_{xx}^{-1/2}$ and $\mathrm{D}^H = \mathrm{G}^H R_{yy}^{-1/2}$. In this case, W and D are known as the canonical mapping matrices.

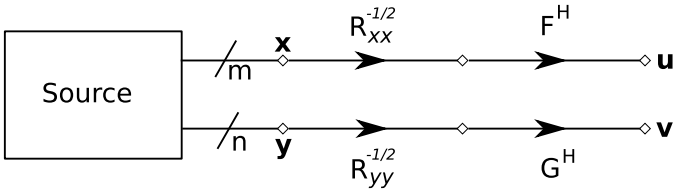


Figure A-1: Transformation from standard coordinates $\mathbf{x}$ and $\mathbf{y}$ to canonical coordinates $\mathbf{u}$ and $\mathbf{v}$.

The canonical correlations k_i are invariant to block-diagonal transformations of R_{zz} of form

$$\mathrm{T}R_{zz}\mathrm{T}^H = \begin{bmatrix} \mathrm{T}_1 & \mathbf{0} \\ \mathbf{0} & \mathrm{T}_2 \end{bmatrix} \begin{bmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{bmatrix} \begin{bmatrix} \mathrm{T}_1^H & \mathbf{0} \\ \mathbf{0} & \mathrm{T}_2^H \end{bmatrix}, \tag{A-13}$$

where $\mathrm{T}_1 \in \mathbb{R}^{m\times m}$ and $\mathrm{T}_2 \in \mathbb{R}^{n\times n}$ are nonsingular matrices [29]. This may easily be proved by showing that the coherence matrix of the transformed data $\mathrm{T}_1\mathbf{x}$ and $\mathrm{T}_2\mathbf{y}$ is the same as that of $\mathbf{x}$ and $\mathbf{y}$.

In fact, the canonical correlations k_i form a *complete* or *maximal* set of invariants [29] for the composite covariance matrix $\mathrm{R}_{zz} = E[\mathbf{z}\mathbf{z}^T]$, under the linear transformation group

$$\mathcal{T} = \left\{ \mathbf{T} = \begin{bmatrix} \mathrm{T}_1 & \mathbf{0} \\ \mathbf{0} & \mathrm{T}_2 \end{bmatrix}, \det\{\mathrm{T}\} \neq 0 \right\}, \tag{A-14}$$

with group action $R_{zz} \rightarrow \mathrm{T}R_{zz}\mathrm{T}^T$ [29]. That is, any function of R_{zz} that is invariant under the transformation group $\mathcal{T}$ is a function of K. This is the reason that the correlations k_i and coordinates $\mathbf{u} = [u_i]_{i=1}^m$ and $\mathbf{v} = [v_i]_{i=1}^n$ are called canonical [13].

A.1 Linear Dependence and Coherence

The standard measure of linear dependence for the composite data vector $\mathbf{z} = [\mathbf{x}^H \quad \mathbf{y}^H]^H$ is the Hadamard ratio, inside the inequality

$$0 \leq \frac{\det\{\mathrm{R}_{zz}\}}{\prod_{i=1}^{m+n}[\mathrm{R}_{zz}]_{ii}} \leq 1, \tag{A-15}$$

where $[R_{zz}]_{ii}$'s, $i \in [1, m+n]$ are the diagonal elements of R_{zz}. This ratio takes the value 0 iff there is linear dependence among elements of $\mathbf{z}$; it takes the value 1 iff the elements of $\mathbf{z}$ are mutually uncorrelated.

By introducing a block Cholesky factorization [13, 34] for R_{zz} of the form

$$R_{zz} = \begin{bmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{bmatrix} = \begin{bmatrix} \mathrm{I} & R_{xy}R_{yy}^{-1} \\ \mathbf{0} & \mathrm{I} \end{bmatrix} \begin{bmatrix} \mathrm{Q}_{xx} & \mathbf{0} \\ \mathbf{0} & R_{yy} \end{bmatrix} \begin{bmatrix} \mathrm{I} & \mathbf{0} \\ R_{yy}^{-1}R_{yx} & \mathrm{I} \end{bmatrix}, \quad \text{(A-16)}$$

where $\mathrm{Q}_{xx} = R_{xx} - R_{xy}R_{yy}^{-1}R_{yx}$ and is known as the error covariance matrix. It is then possible to write $\det\{R_{zz}\}$ as

$$\begin{aligned} \det\{R_{zz}\} &= \det\{\mathrm{Q}_{xx}\}\det\{R_{yy}\} \\ &= \det\{R_{xx}\}\tfrac{\det\{\mathrm{Q}_{xx}\}}{\det\{R_{xx}\}}\det\{R_{yy}\}, \end{aligned} \quad \text{(A-17)}$$

yielding the following decomposition of the Hadamard ratio:

$$\frac{\det\{R_{zz}\}}{\prod_{i=1}^{m+n}[R_{zz}]_{ii}} = \frac{\det\{R_{xx}\}}{\prod_{i=1}^{m}[R_{xx}]_{ii}}\det\{\mathrm{I} - \mathrm{K}\mathrm{K}^H\}\frac{\det\{R_{yy}\}}{\prod_{i=1}^{n}[R_{yy}]_{ii}}. \quad \text{(A-18)}$$

The first and third terms on the right hand side of (A-18) measure the linear dependence *among* the elements of $\mathbf{x}$ and $\mathbf{y}$, respectively, while the middle term,

$$L = \det(\mathrm{I} - \mathrm{K}\mathrm{K}^H) = \prod_{i=1}^{n}(1 - k_i^2); \quad 0 \le L \le 1, \quad \text{(A-19)}$$

measures the linear dependence *between* the elements of $\mathbf{x}$ and $\mathbf{y}$. The measure L takes the value 0 iff there is perfect linear dependence between elements of $\mathbf{x}$ and $\mathbf{y}$; it takes the value 1 iff the elements of $\mathbf{x}$ and $\mathbf{y}$ are independent. The ith term of the product on the right hand side of (A-19), i.e. $(1 - k_i^2)$, measures the linear dependence between the ith canonical coordinate of $\mathbf{x}$ and the ith canonical coordinate of $\mathbf{y}$. This implies that the linear dependence between $\mathbf{x}$ and $\mathbf{y}$ is decomposed into the linear dependence between their canonical coordinates, and is measured only by their canonical correlations or principal cosines.

Correspondingly, we may define the coherence measure between the elements of $\mathbf{x}$ and $\mathbf{y}$ as

$$H = 1 - L = 1 - \det(\mathrm{I} - \mathrm{K}\mathrm{K}^H) = 1 - \prod_{i=1}^{n}(1 - k_i^2); \quad 0 \leq H \leq 1. \tag{A-20}$$

The elements of $\mathbf{x}$ and $\mathbf{y}$ are perfectly coherent iff $H = 1$; and non-coherent iff $H = 0$.