Polarimetric Processing and Sequential Detection for Automatic Target Recognition Systems

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the
Graduate School of The Ohio State University

By

Emre Ertin, B.S.E.E., B.S. Physics, M.S.E.E.

* * * * *

The Ohio State University

1999

Dissertation Committee:

Lee C. Potter, Adviser
Urbashi Mitra
Randolph L. Moses

Approved by

Lee C. Potter
Adviser
Department of Electrical Engineering
This dissertation consists of studies on polarimetric processing and sequential detection for automatic target recognition (ATR) systems.

In the first part, we study polarimetric processing of radar signals. We first present multi-channel techniques to compensate for effects of antenna shading and crosstalk in wideband, wide-angle full polarization radar imaging. These techniques achieve multi-channel deconvolution of the antenna response in the Doppler domain for regularized inversion of the system model. Results on measured antenna patterns are also given. Then, we present a Bayesian approach to detection and classification of scattering signals using their polarimetric signatures. Scattering centers are modeled as one of $M$ canonical reflectors with unknown amplitude, phase and orientation angle; clutter is modeled as a spherically invariant random vector. The resulting decision rule yields computationally simple implementation and posterior estimation of decision uncertainty. Performance of the proposed classifier is illustrated on measured imagery.
In the second part we study optimal design of sequential detectors under computational cost constraints. First, we study the problem of designing sequential procedures that maximize detection performance under expected experimentation cost constraints. Two different problem formulations are proposed. The first problem formulation identifies decision rules which maximize average decision risk under an average computational cost constraint. The second problem formulation is an extension of Neyman-Pearson detection to sequential scenarios. The relationship is established between solutions of these two problems and the classic Bayes solutions to sequential detection and a design procedure based on this result is presented. Next, we consider a staged detection network commonly adopted in ATR systems, where individual detectors can communicate hard decisions to the others. Each detector should decide whether to stop and make a system-wide decision or pass the job to the subsequent detector based on its local observation and on the fact that the previous detectors passed. We provide optimal design rules for staged detection networks under computational constraints. Then, we show that the optimal design can be implemented using Receiver Operating Characteristic curves of individual detectors in the network. Finally, we study the performance of staged detection networks which consists of an infinite serial-chain of identical detectors.
ACKNOWLEDGMENTS

I wish to thank my adviser Lee C. Potter for his guidance and motivation over the past five years. I also thank Urbashi Mitra and Randy Moses for several valuable comments and suggestions.
VITA

Nov 10, 1969 .......................... Born - Istanbul, Turkey

1992 ............................... B.S. Electrical Engineering,
                            B.S. Physics,
                            Bogazici University, Turkey

1993 ............................... M.Sc. Electrical Engineering,
                            Imperial College, UK

PUBLICATIONS

Research Publications


FIELDS OF STUDY

Major Field: Electrical Engineering

Studies in:
- Statistical Signal Processing
  - Prof. Lee Potter, Prof. Randy Moses
  - Prof. Urbashi Mitra

Minor: Control
- Prof. Hitay Özbay

Minor: Mathematics
- Prof. Paul Nevai, Prof. George Majda,
- Prof. Bill Davis, Prof. Joseph Ferrar
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td>Automatic Target Recognition Systems</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>Summary of Chapters and Contributions</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>Polarimetric Classification of Scattering Centers using M-ary Bayesian Decision Rules</td>
<td>9</td>
</tr>
<tr>
<td>2.1</td>
<td>Introduction</td>
<td>9</td>
</tr>
<tr>
<td>2.2</td>
<td>Physical Model</td>
<td>11</td>
</tr>
<tr>
<td>2.2.1</td>
<td>Scattering Matrix</td>
<td>11</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Polarization Basis Vectors</td>
<td>13</td>
</tr>
<tr>
<td>2.2.3</td>
<td>Incomplete Polarization Information</td>
<td>14</td>
</tr>
<tr>
<td>2.3</td>
<td>M-ary Target Classification</td>
<td>16</td>
</tr>
<tr>
<td>2.3.1</td>
<td>Polarimetric Signatures in SIRV clutter</td>
<td>16</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Bayes Detection and Classification</td>
<td>18</td>
</tr>
<tr>
<td>2.3.3</td>
<td>The GLRT Approach</td>
<td>19</td>
</tr>
<tr>
<td>2.3.4</td>
<td>Classification Significance</td>
<td>22</td>
</tr>
</tbody>
</table>
2.3.5 Geometrical Interpretation and Computational Cost .................. 23
2.4 Experimental Results .............................................. 25
   2.4.1 Simulated Data ............................................... 25
   2.4.2 Measured Data ................................................. 26
2.5 Conclusion .................................................................. 32

3. Multichannel Imaging for Wide-Band Synthetic Aperture Radar Systems 34
   3.1 Introduction .......................................................... 34
   3.2 System Model ........................................................ 35
   3.3 Multi-channel Techniques for Antenna Pattern Compensation .... 40
      3.3.1 Local Approximation Method ............................... 40
      3.3.2 Doppler Domain Technique .................................. 43
      3.3.3 Image Postprocessing Method ............................... 47
   3.4 Results .................................................................. 49

4. Sequential Decision Problems with Cost Constraints ................. 56
   4.1 Introduction .......................................................... 56
   4.2 Notation and Mathematical Model .................................. 59
   4.3 Bayesian Decision Problem ....................................... 61
      4.3.1 Case 1: I.I.D. Experiments ................................. 61
      4.3.2 Case 2: Feature Selection .................................... 63
   4.4 Sequential Decision Problems with Computational Cost Constraints 64
      4.4.1 Example 1: Feature Selection ............................... 67
   4.5 Approximate Design (Single Experiment, Binary Hypothesis) .... 70

5. Serial Distributed Detection with Computational Cost Constraints .... 76
   5.1 Introduction .......................................................... 76
   5.2 Bayesian Problem Formulation ..................................... 79
      5.2.1 Solution to Bayes Problem ................................... 81
      5.2.2 Direct Solution to Bayes Problem using ROC Curve .... 83
   5.3 Extension to Multiple Detectors and Binary Local Decisions .......... 84
      5.3.1 Multiple Detectors- Network I .............................. 84
      5.3.2 Multiple Detectors- Network II .............................. 88
   5.4 Infinite Serial Detector with Identical Sensors ..................... 93

Appendices:

A. ML Estimate ............................................................. 100
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Different bases for canonical scattering centers</td>
<td>14</td>
</tr>
<tr>
<td>2.2</td>
<td>Polarimetric signatures for canonical scattering centers in the TD basis.</td>
<td>15</td>
</tr>
<tr>
<td>2.3</td>
<td>Incomplete polarimetric signatures for canonical scattering centers using left circular transmit with both left and right circular receive polarizations.</td>
<td>16</td>
</tr>
<tr>
<td>2.4</td>
<td>Minimum distance from normalized signature to eight target cones using a forest clutter H-V basis covariance matrix</td>
<td>26</td>
</tr>
<tr>
<td>2.5</td>
<td>Simulated confusion matrix, forest clutter with SNR=20dB</td>
<td>27</td>
</tr>
<tr>
<td>2.6</td>
<td>P-3 radar system characteristics</td>
<td>29</td>
</tr>
<tr>
<td>2.7</td>
<td>Classification results for the P3 Calibration Array</td>
<td>30</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>ARL Boom SAR</td>
<td>3</td>
</tr>
<tr>
<td>1.2</td>
<td>ATR Staged Processing System</td>
<td>4</td>
</tr>
<tr>
<td>2.1</td>
<td>Probability of correct detection versus SNR using Bayes optimal detection</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>with prior distributions on unknown parameters ( (A, \rho, \psi) ) (solid</td>
<td></td>
</tr>
<tr>
<td></td>
<td>line) and a GLRT (dashed line)</td>
<td></td>
</tr>
<tr>
<td>2.2</td>
<td>Notional representation of the geometry for the GLRT detector. Projection</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>onto target cones yields ML estimates ( (\hat{A}, \hat{\psi}) ); the nearest</td>
<td></td>
</tr>
<tr>
<td></td>
<td>cone yields the class decision.</td>
<td></td>
</tr>
<tr>
<td>2.3</td>
<td>Probability of correct classification vs SNR(dB)</td>
<td>27</td>
</tr>
<tr>
<td>2.4</td>
<td>Schematic diagram of the Grayling reflector array</td>
<td>28</td>
</tr>
<tr>
<td>2.5</td>
<td>HH polarization magnitude image of the reflector array corresponding to the</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>dotted area in Figure 4</td>
<td></td>
</tr>
<tr>
<td>2.6</td>
<td>Sample Target from the MIT Lincoln Lab P-3 polarimetric imagery</td>
<td>31</td>
</tr>
<tr>
<td>3.1</td>
<td>Ground plane radar geometry</td>
<td>36</td>
</tr>
<tr>
<td>3.2</td>
<td>Effect of antenna functions on imaging: (a) four channel image with</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>antenna pattern omitted; (b) four channel image with effect of antenna</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pattern</td>
<td></td>
</tr>
<tr>
<td>3.3</td>
<td>Results for the local approximation method: four channel image</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>constructed from range lines processed using the local approximation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>algorithm</td>
<td></td>
</tr>
</tbody>
</table>
3.4 Results for the local approximation method: cross-sections from calibrated images in Fig. 3.3 (solid) and the ideal antenna response in Fig. 3.2 (dashed) ........................................... 51

3.5 Results for the Doppler domain method: four channel image constructed from range lines processed in the Doppler domain .................. 52

3.6 Results for the Doppler domain method: cross-sections from calibrated images in Fig. 3.5 (solid) and the ideal antenna response in Fig. 3.2a (dashed) ........................................... 53

3.7 Results for the image postprocessing method: four channel postprocessed imagery. ........................................... 54

4.1 Block diagram of a non-adaptive, sequential, multi-feature three stage automatic target recognition system ........................................... 57

4.2 Comparison of Bayes Risk versus average computation cost for two non-adaptive decision rules and the Bayes-optimal adaptive decision rule. ........................................... 69

4.3 NP Solution with cost constraints ........................................... 73

4.4 Bayes Solution with average cost constraint ........................................... 75

5.1 Staged Detection Network-I ........................................... 75

5.2 Staged Detection Network-II ........................................... 75

5.3 Staged Detection Network-Two detectors ........................................... 78

5.4 ROC curve for binary symmetric channel ........................................... 83

5.5 ROC curves for Network I ........................................... 84

5.6 $C_2$ Problem-Network I ........................................... 86

5.7 ROC curves for Network II ........................................... 88

5.8 $C_2$ Problem-Network II ........................................... 90
CHAPTER 1

INTRODUCTION

1.1 Automatic Target Recognition Systems

The objective of Automatic Target Recognition (ATR) systems is to process data obtained by various sensors to detect and recognize objects of interest (targets). ATR systems are primarily used in defense applications to automatically detect and recognize strategic targets with reduced risk and increased accuracy [16]. Such systems can be used for detection of buried objects, location of downed aircraft (search and rescue) [61] and mine detection [44]. ATR technology is also applicable to non-military problems such as identification of natural resources and agricultural surveying. Another example, the problem of recognizing cancer cells from medical imagery (X-ray, magnetic resonance and ultrasound imagery) falls in the scope of ATR theory.

The input of the ATR system is usually in the form of an image. The input image is obtained by processing signals from a forward-looking infrared (FLIR) camera, a synthetic aperture radar (SAR), a television camera or a laser radar, although ATR techniques can be applied to outputs of non-imaging sensors as well [16]. Current SAR systems can operate day or night and in all-weather conditions. Furthermore,
they have foliage and ground penetration capability and can produce high resolution imagery suitable for ATR applications. Synthetic Aperture Radar (SAR) systems synthesize the effect of an antenna array using the motion of a single sensor. Scattering data collected over the large aperture of the synthesized array can be coherently processed to produce high resolution imagery [9]. Additionally, the sensor may measure scattered fields in two independent polarizations to provide additional characterization of the remotely imaged objects.

Our studies on ATR systems are motivated by Army Research Laboratory’s research program for development of a low frequency, ultra-wideband (UWB) imaging radar system to detect obscured targets such as vehicles concealed by foliage and objects buried underground. ARL has designed and constructed a UWB radar that provides instrumentation-grade data for phenomenology studies and algorithm development [49]. A 50 meter telescoping boom lift allows emulation of airborne synthetic radar imaging geometry at a fraction of the cost. Figure 1.1 gives a schematic description of the BoomSAR [10]. Remote radar imaging of objects obscured by tree canopy or buried underground requires long wavelengths for propagation through the obscuring foliage or soil. Additionally, a large bandwidth is required for high spatial resolution. Together, the dual requirements on wavelength and bandwidth demand a radar sensor with bandwidth exceeding one quarter the center frequency. Moreover, image resolution is governed not only by bandwidth but also by the span of aspect angles provided by the synthetic array. For a given resolution, the required angle diversity is proportional to the center frequency; therefore, UWB SAR systems require much larger angle diversity than high frequency (e.g., X- or K-band) SAR systems. The ARL boom-SAR has a frequency range of 50MHz to 1.2 GHz, with integration
angles exceeding 90 degrees. These extreme operating conditions necessitate new approaches to imaging, detection and recognition.

The basic task of the ATR system is to detect and recognize targets of interest in an environment which includes clutter (objects which are not targets of interest) using raw data measurements from an imperfect sensor that produces noisy signals. Precise definition of target, clutter and noise depends on the particular application at hand. The high data rates and real-time processing requirements for wide area surveillance have given rise to a staged processing strategy [16]. The block diagram in Figure 1.2 shows the staged decision system employed by current ATR systems. The four main stages are: Imaging, Preprocessing, Detection and Classification. The raw data from the sensor is processed to form an image, which serves as a list of decision statistics indexed by spatial coordinates of the scene. Then at the preprocessing stage imperfections introduced by the sensor and imaging process are removed. Simple
Figure 1.2: ATR Staged Processing System

detection algorithms are applied to all sensor data to isolate small portions that are likely to contain targets. Then increasingly more complex detection and classification algorithms are employed on the selected portions of the data to reject clutter and classify targets. Ideally, only the targets pass through the system and are included in the output target list. As the data moves through the staged processing system, algorithms become more target specific and computationally costly, while the number of false alarms and the number of regions processed decreases.

In this dissertation, we study several new approaches to different processing stages, which address some of the challenges introduced by the ARL experimental fully polarimetric, wide-angle, UWB radar system:

- Polarization diversity in radar backscatter measurements can yield improved detection and characterization of man-made targets in clutter. We develop a Bayesian detection and classification algorithm for recognizing polarimetric signatures in non-Gaussian clutter.

- Any UWB antenna will have a nonideal magnitude and phase response that varies as a function of frequency, aspect angle, depression angle, and polarization of the incident energy. Thus, standard processing yields poor image quality due to the inherent variation of the antenna response – in magnitude, phase, and
polarization – across the large band of frequencies and wide range of aspect angles. We introduce multi-channel techniques to calibrate for effects of antenna shading and crosstalk in wideband, wide-angle full polarization radar imaging. The result is imagery with improved polarization purity and a more localized point spread function.

- An ATR system accesses a large volume of sensor data to perform a multiple hypothesis test with a constraint on computation. The high data rates and real-time processing requirements for wide area surveillance have given rise ad-hoc staged detection strategies. We address the problem of designing optimal sequential procedures that maximize detection performance under expected experimentation (observation) cost constraints.

- The optimal sequential procedure requires that detectors in the serial network pass soft decisions to subsequent stages. However, a common practice in the design of ATR decision networks is not to allow local decisions to propagate to later stages [16]. We combine results from sequential and distributed detection theory to analyze the design and performance of staged decision networks with limited communication capabilities.

### 1.2 Summary of Chapters and Contributions

In the second chapter, we present a Bayesian approach to detection and classification of scattering signals using their polarimetric signatures [19]. Scattering centers are modeled as one of $M$ canonical reflectors with unknown amplitude, phase and orientation angle; clutter is modeled as a spherically invariant random vector. A choice of costs in the Bayes Risk is shown to yield a two-stage classification rule. The
first stage is a Neyman-Pearson detector which rejects clutter, whereas the second stage classifies the detection in one of the $M$ target classes. The resulting decision rule yields computationally simple implementation, intuitive geometric interpretation, and posterior estimation of decision uncertainty. Performance of the proposed classifier is illustrated on real imagery from an airborne UHF-band radar, processed by MIT Lincoln Lab [5].

In the third chapter, we present multi-channel techniques to compensate for effects of antenna shading and crosstalk in wideband, wide-angle full-polarization radar imaging. The system is modeled as a two-dimensional integral operator that includes the transmit pulse function, receive and transmit antenna transfer functions, and response from scattering objects. We introduce three techniques for compensation of antenna patterns in multi-channel radar data. All techniques require measured antenna patterns over the range of frequencies and look angles of the radar. The first technique uses a local approximation to the system model, and therefore is suitable for the processing of small image chips. Its performance also depends on the smoothness of the antenna transfer function over a range of depression and azimuth angles. The second technique proposes an inversion method for the system model in a transform domain. This Doppler domain technique processes the Fourier transform of the range lines, to compensate for the effects of the antenna transfer function. The third technique is an image domain implementation of the Doppler domain method.

In the fourth chapter, we study optimal design of sequential detectors under computational cost constraints. In this chapter we provide design procedures for improved sequential decision rules under computational cost constraints. Sequential detection
procedures have been previously proposed, primarily motivated by the idea of minimizing the number of observations for a desired level of performance. However, the design of such sequential procedures does not consider the expected cost of observation as one of the desired performance criteria. In the research we address the problem of designing sequential procedures that maximize detection performance under expected experimentation (observation) cost constraints. Two different problem formulations are proposed. The first problem formulation identifies decision rules which maximize average decision risk under an average computational cost constraint. The second problem formulation is an extension of Neyman-Pearson detection to sequential scenarios. The relationship between solutions of these two problems and the classic Bayes solutions to sequential detection is established, and a design procedure based on this result is presented.

In the fifth chapter, we combine results from sequential and distributed detection theory, to analyze staged detection networks commonly adopted in ATR systems. This detection networks adopt a "sieve approach," where each detector decides to defer to the next detector or terminate the decision process and make the final system-wide decision. We provide optimal design rules for staged detection networks under computational constraints. Then, we show that the optimal design can be implemented using Receiver Operating Characteristic (ROC) curves of individual detectors in the network. These design techniques are then used to derive the optimal ROC curve for the staged detection network under computational cost constraints. Finally, we study the performance staged detection networks which consists of infinite serial-chain of identical detectors. For this infinite-serial-network we obtain necessary
and sufficient conditions for the converge of the probability of error to zero as the expected computational cost goes to infinity.
CHAPTER 2

POLARIMETRIC CLASSIFICATION OF SCATTERING CENTERS USING M-ARY BAYESIAN DECISION RULES

2.1 Introduction

Polarization diversity in radar backscatter measurements has been shown to yield improved detection and characterization of man-made targets in clutter [11, 14, 15, 42, 35, 36, 7, 53, 12, 46, 17]. Multiple polarization channels have been combined to enhance performance of energy detectors [11, 14, 15, 42, 35, 36]; similarly, polarimetric measurements have been used to characterize the shape or orientation of reflectors [7, 53, 12, 46] and to segment wide-area terrain imagery [17].

Several polarimetric radar detectors have been proposed using Huynen's parametrization of the polarimetric scattering matrix [31]. The detection problem is the binary hypothesis test of clutter versus a known polarimetric signature in clutter. DeGraaf [11] derived matched-filter and quadratic detectors. Dilsavor and Moses [14, 15] presented a generalized likelihood ratio test (GLRT) to account for unknown target orientation and K-distributed clutter; additionally, Dilsavor [14] showed that the polarimetric whitening filter (PWF), introduced for speckle reduction, is the GLRT for a completely unknown scattering matrix. Larson and Novak [36] derived a closed form expression for the GLRT detector for a dihedral with unknown orientation angle.
They also derived an adaptive, constant false alarm rate (CFAR) version of this detector and demonstrated that it outperforms PWF CFAR detection on 1 ft resolution synthetic aperture radar (SAR) imagery.

Polarization diversity is useful not only for detection but also for extracting geometrically relevant features of a detected target [12, 46]. Polarimetric features were shown to improve ATR performance in both K-band and L-band empirical studies [35]. Classification of scattering matrices was proposed by Cameron and Leung [7]; symmetric scattering matrices were diagonalized and correlated against template signatures without regard to clutter.

In this paper, we join and extend the two themes of statistical detection and physical description of reflectors in clutter. A composite target class of M reflector types and choice of costs in a Bayesian hypothesis test is shown to yield a Neyman-Pearson detection rule [19]. Further, the Bayes formulation provides a Maximum A-posteriori Probability (MAP) classification rule for detected reflectors. The decision rule is derived using the general class of Spherically Invariant Random Vectors (SIRV) to model terrain clutter, but nonetheless is shown to admit computationally simple implementation. The decision rule extends previous classification efforts to allow non-Gaussian clutter, to compute Maximum Likelihood (ML) estimates of unknown orientation angles, and to provide an estimate of uncertainty in the class decision.

In Section 2.2 we present a physical model for scattering centers and SIRV clutter; the scattering matrix is conveniently represented in the vector space $\mathbb{C}^3$. We also discuss the case of incomplete polarization information as a special case of the model. In Section 2.3 we give the Bayes formulation of the polarimetric classification problem and its solution; we adopt the Generalized Likelihood Ratio Test (GLRT)
approach and derive an approximate expression for the uncertainty associated with
the classification. We then provide geometric interpretation of the classification rule
and a computationally simple implementation. In Section 2.4 we present signature
classification results for both a synthetic data set and measured SAR imagery.

2.2 Physical Model

2.2.1 Scattering Matrix

A fully polarimetric radar measures a backscattered electric field with two linearly
independent transmit and receive polarizations. The scattering matrix $S(f, \theta)$ is
the linear transformation that maps an incident polarization state to the scattered
polarization state. In general, the monostatic scattering matrix $S$ for an object is
a function of frequency, $f$, and angle of incidence, $\theta$. For a horizontal and vertical
linear polarization basis (HV), the matrix of scattering coefficients is given by

$$
S(f, \theta) = \begin{bmatrix}
S_{hh}(f, \theta) & S_{hv}(f, \theta) \\
S_{vh}(f, \theta) & S_{vv}(f, \theta)
\end{bmatrix}
$$

(2.1)

where the $S_{ij}(f, \theta)$ represents the complex gain for the $j^{\text{th}}$ incident polarization state
and $i^{\text{th}}$ scattered polarization state.

The scattering coefficients in $S(f, \theta)$ may be obtained from complex SAR imagery
directly from single pixels or via parametric modeling. Under the assumption of a
parametric scattering model [53, 23], one can jointly estimate the location and multi-
channel amplitude of a scattering center, by fitting the model to the complex image.
The estimated multi-channel amplitudes provide an estimate of the scattering matrix.
Alternatively, the scattering coefficients may be approximated by the complex-valued
pixels in pixel-registered polarimetric imagery. When using image pixels, each resolu-
tion cell is assumed to contain a dominant scattering center whose scattering response
is averaged over the bandwidth and aperture angle of the imaging system. Further, for time-domain pulse radars, the scattering matrix can be inferred from the complex envelope of the range profiles [14].

A polarimetric representation reparametrizes the scattering matrix to a set of physically interpretable quantities characterizing the scattering center. Early work by Sinclair [50], Deschamps [13], and Kennaugh [33] introduced polarimetric representations and null polarizations. For a review of polarimetric representations see Giuli [24] and references therein. Huynen [31] presented a polarimetric representation in which parameters are physically linked to the geometry of the scatterer. The Huynen model expresses a general complex-valued scattering matrix in terms of six geometrically relevant real-valued descriptors

$$S = e^{i\rho}U^*(\psi, \tau, \nu)A \begin{bmatrix} 1 & 0 \\ 0 & \tan^2\gamma \end{bmatrix} U^H(\psi, \tau, \nu)$$  \hspace{1cm} (2.2)

where $A$ is the magnitude, $\rho$ is the absolute phase, $\psi$ is the orientation (tilt) angle, $\tau$ is the ellipticity angle, $\nu$ is the skip angle, and $\gamma$ is the characteristic angle. The parameters are physical descriptors of the target. For example, the skip angle $\nu$ is related to the number of bounces of the geometric optics reflected signal. Odd bounce mechanisms correspond to $\nu = 0$ degrees, whereas even bounce mechanisms correspond to $\nu = \pm 45$ degrees. In Eqn. (2.2) the matrix $U$ is unitary, and the decomposition $U^*DU^H$ for complex symmetric matrices is known as the Takagi factorization ([30], page 250). For a complex valued matrix $U$, we denote the complex conjugate of $U$ by $U^*$, and the Hermitian transpose of $U$ by $U^H$.

Scattering matrices for many canonical scattering centers are well known. We adopt these scattering matrices as deterministic models for polarimetric signatures measured in random clutter.
2.2.2 Polarization Basis Vectors

We adopt an alternative representation to Eqn. (2.2) of the Huynen model for convenience in computation and interpretation of decision rules. Assuming monostatic measurements in free space, we invoke reciprocity, $S_{hv} = S_{vh}$, to represent $S$ as the complex-valued triple, $S = [S_{hh} \sqrt{2} S_{hv} S_{vv}]^T$. If $S_{hv}$ and $S_{vh}$ are measured independently, they can be averaged to provide a single cross-pole measurement. The scaling by $\sqrt{2}$ simplifies a rotation operator defined below.

Consider a given canonical scattering center oriented at a tilt angle $\psi$ and backscattered field measurements in a HV basis. We choose to express the scattering vector $S$ as a normalized signature vector $b(\tau, \nu, \gamma)$ in $\mathbb{C}^3$ rotated to tilt angle $\psi$ and scaled by a complex amplitude, $A_c = A e^{j\rho}$. The parameters $(A, \rho, \psi)$ describing amplitude, absolute phase, and orientation angle are modeled as deterministic and unknown. Choice of the remaining parameters, $(\tau, \nu, \gamma)$, defines a canonical scattering signature. Thus, we obtain a scattering representation in $\mathbb{C}^3$,

$$S = A_c R_{HV}(\psi) b(\tau, \nu, \gamma)$$  \hspace{1cm} (2.3)

$R_{HV}(\psi)$ is the 3-by-3 rotation matrix in the HV coordinate system. To simplify the implementation and interpretation of the rotation matrix, it is convenient to perform the rotation in an alternate coordinate system defined by the columns of a 3-by-3 matrix, $B$. Accordingly, the measured HV scattering vector $S$ is expressed

$$S = A_c B R_B(\psi) b(\tau, \nu, \gamma)$$  \hspace{1cm} (2.4)

The rotation matrix $R_B(\psi)$ for basis $B$ is given by

$$R_B(\psi) = B^H R_{HV}(\psi) B$$  \hspace{1cm} (2.5)
<table>
<thead>
<tr>
<th>Basis</th>
<th>$B$</th>
<th>$R_B(\psi)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HV-Basis</td>
<td>$\begin{bmatrix} 1 &amp; 0 &amp; 0 \ 0 &amp; 1 &amp; 0 \ 0 &amp; 0 &amp; 1 \end{bmatrix}$</td>
<td>$\begin{bmatrix} \cos^2 \psi &amp; -\sqrt{2} \sin \psi \cos \psi \ \sqrt{2} \sin \psi \cos \psi &amp; \cos^2 \psi - \sin^2 \psi \ \sin^2 \psi &amp; \sqrt{2} \sin \psi \cos \psi \end{bmatrix}$</td>
</tr>
<tr>
<td>Trihedral-Dihedral Basis</td>
<td>$\begin{bmatrix} \frac{1}{\sqrt{2}} &amp; \frac{1}{\sqrt{2}} &amp; 0 \ 0 &amp; 0 &amp; 1 \ \frac{1}{\sqrt{2}} &amp; -\frac{1}{\sqrt{2}} &amp; 0 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 1 &amp; 0 &amp; 0 \ 0 &amp; \cos(2\psi) &amp; -\sin(2\psi) \ 0 &amp; \sin(2\psi) &amp; \cos(2\psi) \end{bmatrix}$</td>
</tr>
<tr>
<td>LR-Basis</td>
<td>$\begin{bmatrix} \frac{1}{\sqrt{2}} &amp; \frac{1}{\sqrt{2}} \ \frac{1}{\sqrt{2}} &amp; 0 &amp; \frac{1}{\sqrt{2}} \ \frac{1}{\sqrt{2}} &amp; \frac{1}{\sqrt{2}} &amp; \frac{1}{\sqrt{2}} \end{bmatrix}$</td>
<td>$\begin{bmatrix} e^{-j2\psi} &amp; 0 &amp; 0 \ 0 &amp; 1 &amp; 0 \ 0 &amp; 0 &amp; e^{j2\psi} \end{bmatrix}$</td>
</tr>
</tbody>
</table>

Table 2.1: Different bases for canonical scattering centers

Three bases and the corresponding rotation matrices are given in Table 2.2.2. In Section 3 we choose the Trihedral-Dihedral basis and rotation, $R_{TD}(\psi)$, for computational and conceptual simplicity.

For each of $M$ canonical forms we define the scattering signature, $b_i(\psi)$, at orientation (tilt) angle $\psi$.

$$b_i(\psi) = BR_B(\psi)b(\tau_i, \nu_i, \gamma_i) \quad i = 1, \ldots, M \quad (2.6)$$

In Table 2.2.2 we list $b_i(\psi)$ for eight canonical scattering matrices using the Trihedral-Dihedral coordinate system.

### 2.2.3 Incomplete Polarization Information

The classification rules presented in this paper can be directly applied to classification of polarimetric signatures when the measurements provide incomplete polarization information. One important case is a radar which transmits only left circularly polarized pulses and receives both left and right circularly polarized returns. Incomplete polarization measurements can be formulated as projections of the scattering
<table>
<thead>
<tr>
<th>i</th>
<th>Scattering Center</th>
<th>$R_{TD}(\psi)b(\tau_i, \nu_i, \gamma_i)$</th>
</tr>
</thead>
</table>
| 1 | Trihedral               | \[
\begin{bmatrix}
1 \\
0 \\
0
\end{bmatrix} \]
| 2 | Dihedral ($a = 0$)      | \[
\begin{bmatrix}
a \\
\cos(2\psi) \\
\sin(2\psi)
\end{bmatrix}
\]
| 3 | Dipole ($a = 1$)        | \[
\begin{bmatrix}
a \\
\cos(2\psi) \\
\sin(2\psi)
\end{bmatrix}
\]
| 4 | Cylinder ($a = 3$)      | \[
\begin{bmatrix}
\frac{\sqrt{2}}{2} \\
\frac{\sqrt{2}}{2} \\
\frac{\sqrt{2}}{2}
\end{bmatrix}
\]
| 5 | Narrow Diplane ($a = \frac{1}{3}$) | \[
\begin{bmatrix}
\frac{\sqrt{2}}{2} \\
\frac{\sqrt{2}}{2} \\
\frac{\sqrt{2}}{2}
\end{bmatrix}
\]
| 6 | Quarter Wave ($a = j$)  | \[
\begin{bmatrix}
\cos(2\psi) \\
\sin(2\psi)
\end{bmatrix}
\]
| 7(8)| Left (Right) Helix | $e^{2\psi} \begin{bmatrix} 0 \\ 1 \\ (-)j \end{bmatrix}$ |

Table 2.2: Polarimetric signatures for canonical scattering centers in the TD basis.

matrix $S$ in $\mathbb{C}^3$ to a subspace spanned by the incomplete polarization basis $B$. For example, for the left circular transmit and left-right circular receive measurements, the orthonormal basis $B$ is given by:

$$B = \begin{bmatrix}
\frac{1}{2} & \frac{\sqrt{2}}{2} \\
\frac{\sqrt{2}}{2} & 0 \\
\frac{1}{2} & \frac{\sqrt{2}}{2}
\end{bmatrix}$$

To obtain the corresponding canonical polarimetric signatures for incomplete polarization measurements, we first project a signature from each class onto the new basis, $B$.

$$b_B(\tau_i, \nu_i, \gamma_i) = B^H b(\tau_i, \nu_i, \gamma_i)$$

Second, we find the new tilt angle operator $R_B(\psi)$ using

$$R_B(\psi) = B^H R_{HV}(\psi) B$$

Then, the canonical classes in the new basis are given by

$$S_i = A_c R_B(\psi)b_B(\tau_i, \nu_i, \gamma_i) = A_c b_i(\psi) \quad i = 1, \ldots, M$$

15
<table>
<thead>
<tr>
<th>$i$</th>
<th>Scattering Center</th>
<th>$R_B(\psi)b_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Trihedral</td>
<td>$0$</td>
</tr>
<tr>
<td>2,7</td>
<td>Dihedral-L Helix ($a = 0$)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Dipole ($a = \sqrt{2}$)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Cylinder ($a = 3\sqrt{2}$)</td>
<td>$e^{-j\frac{2\pi}{a}}$</td>
</tr>
<tr>
<td>5</td>
<td>Narrow Diplane ($a = \frac{\sqrt{2}}{3}$)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Quarter Wave ($a = j\sqrt{2}$)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.3: Incomplete polarimetric signatures for canonical scattering centers using left circular transmit with both left and right circular receive polarizations.

where $b_i(\psi) = R_B(\psi)b_B(\tau_i, \nu_i, \gamma_i)$ is the signature of scattering center $i$ in the incomplete polarization basis $B$ rotated by a tilt angle $\psi$. In Table 2.2.3 we present eight canonical classes in a incomplete polarization basis.

Three effects arise from incomplete polarization information. First, some classes may collapse into a single class, and therefore cannot be distinguished using measurements in the reduced basis. For example, in Table 3 the dihedral and left helix collapse into a single class. Second, a class may become a null vector of the $B$ basis and hence yield zero return; such is the case for the right helix in Table 2.2.3. Third, due to the projection onto a reduced dimension subspace, in general the separation of classes is reduced.

2.3 M-ary Target Classification

2.3.1 Polarimetric Signatures in SIRV clutter

For $M$ canonical target polarization signatures we have $M$ corresponding hypotheses. We have also the null hypothesis which corresponds to clutter. By minimizing
Bayes risk, we obtain two Bayes-optimal decision rules: a Neyman-Pearson type decision rule for detection of canonical targets in clutter and a MAP rule for classification of detected canonical forms. The $M + 1$ hypotheses are

\[ H_0 : \tilde{S} = n \]  
\[ H_i : \tilde{S} = A_c b_i(\psi) + n \quad i = 1, \ldots, M \]

where $n$ is the noise that results from clutter and $\tilde{S}$ is the noisy measurement of the scattering matrix.

For the clutter signal $n$ we adopt an additive, SIRV model [47]. The complex clutter random vector is the product of a positive real amplitude $a$ with a complex Gaussian random vector $v$ having zero mean and covariance $\Sigma$. Let $f(a)$ denote the probability density function for the scalar random variable $a$. Then the probability density function of $n$ can be expressed as

\[ f_n(n) = F_n(||n||_\Sigma) \]  
\[ F_n(x) = \frac{1}{\pi^3|\Sigma|} \int_0^\infty \frac{f(a)}{a^3} \exp(-x^2/a^2)da \]

and $||n||_\Sigma$ is the weighted norm of the vector $n$ given by $n^H \Sigma^{-1} n$.

Empirical studies at L-band [14] and K-band [42] have shown that terrain clutter has a heavy-tailed distribution; special cases of SIRV product models such as the Weibull, log-normal and K-distributions have been widely adopted as models for terrain clutter. The covariance matrix has the specific structure [42]:

\[ \Sigma = \sigma^2_{HH} \begin{bmatrix} 1 & 0 & \alpha \sqrt{\beta} \\ 0 & 2\varepsilon & 0 \\ \alpha^* \sqrt{\beta} & 0 & \beta \end{bmatrix} \]
where $\sigma_{HH}^2$ is the clutter variance on the $HH$ component of the measurement vector $\hat{S}$, $\beta$ is the ratio of clutter power between co-pol channels $\sigma_{VV}^2/\sigma_{HH}^2$, $\varepsilon = \sigma_{HV}^2/\sigma_{HH}^2$, and $\alpha$ is the correlation coefficient between the clutter noise components in $HH$ and $VV$ channels. Finally, the zero elements of the covariance matrix indicate that the clutter is uncorrelated between cross-pol (HV) and co-pol channels (HH and VV). In [42] the parameters for forest clutter are estimated as $\alpha = 0.65$, $\beta = 0.89$ and $\epsilon = 0.16$, for a radar operating at 33 GHz.

### 2.3.2 Bayes Detection and Classification

The choice of the cost matrix $C_{ij} = \text{Cost}[\text{decide } H_i | H_j]$ affects the form of the Bayes classifier. Here we show that a particular choice of the cost matrix leads naturally to a two-stage classifier where the first stage is a variable false-alarm rate detector which rejects clutter, whereas the second stage classifies the detection in one of the $M$ classes.

We assign a Bayes cost to each decision as

$$C_{ij} = \text{Cost}[\text{decide } H_i | H_j] = \begin{cases} 0, & i = j \\ 1, & i \neq j \quad j \neq 0 \\ \eta, & i \neq j \quad j = 0 \end{cases} \quad (2.12)$$

Note that choice of $\eta$ serves to penalize missed detection differently than misclassifications of the $M$ polarimetric signatures. The Bayes risk is given by

$$\text{Risk} = \sum_{j=0}^{M} \sum_{i=0}^{M} P_j C_{ij} P_{ij} \quad (2.13)$$

where $P_j$ is the prior probability of $H_j$ and $P_{ij}$ is the probability of deciding $H_i$ when $H_j$ is true. The Bayes optimal decision rule partitions the measurement space to minimize the risk and can be expressed as a piecewise linear rule in an $M$-dimensional
space with coordinates given by likelihood ratios ([55], page 48).

\[
\Lambda_k(\tilde{S}) = \frac{P_{\tilde{S}|H_k}(\tilde{S}|H_k)}{P_{\tilde{S}|H_0}(\tilde{S}|H_0)} \quad k = 1, \ldots, M \tag{2.14}
\]

Two decision rules emerge as a consequence of minimizing the Bayes risk. First, we have a decision rule for the detection of canonical targets in SIRV clutter:

\[
\max_{i=1,\ldots,M} \left\{ P_i \Lambda_i(\tilde{S}) \right\} \overset{H_1 \cup \cdots \cup H_M}{\gtrsim} \overset{H_0}{\eta} P_0 \tag{2.15}
\]

That is, if the largest likelihood over all target classes is higher than a certain threshold, then a detection is declared. Second, we have a decision rule assigning a detection to one of \( M \) classes:

\[
\max_{i=1,\ldots,M} P_i \Lambda_i(\tilde{S}) \tag{2.16}
\]

The detection is classified in the target class that maximizes the likelihood of the measured scattering vector.

However, the likelihood ratios \( \Lambda_i(\tilde{S}) \) depend on the unknown target parameters \((A, \rho, \psi)\). A complete Bayes approach would be to integrate \( P(\tilde{S}|H_k, (A, \rho, \psi)) \) over these parameters using their prior distributions and use this reduced form in the likelihood ratios.

### 2.3.3 The GLRT Approach

In the absence of prior distributions, we adopt a generalized likelihood ratio test (GLRT). That is, we replace the likelihood ratios with the maximum likelihood ratios, where the unknown parameters are replaced by their ML estimates ([55], page 92).

For equal priors \( P_1 = \ldots = P_M \), the detection rule in Eqn. (2.15) is in the form of a Neyman–Pearson test:

\[
\max_{i=1,\ldots,M} \left\{ \frac{F_{\eta}(\tilde{S} - \hat{A}(\tilde{S}; \hat{\psi}))^H \Sigma^{-1}(\tilde{S} - \hat{A}(\tilde{S}; \hat{\psi}))}{F_{\eta}(\tilde{S}^H \Sigma^{-1} \tilde{S})} \right\} \overset{H_1 \cup \cdots \cup H_M}{\gtrsim} \overset{H_0}{\eta} \tag{2.17}
\]
and the classification rule is in the form of a MAP rule (since $F_n$ is monotone decreasing)

$$i = \arg \min_{i=1,\ldots,M} \left( \tilde{S} - \hat{A}_{ci} b_i(\hat{\psi}_i) \right)^H \Sigma^{-1} \left( \tilde{S} - \hat{A}_{ci} b_i(\hat{\psi}_i) \right)$$

(2.18)

where $\hat{A}_{ci}$ and $\hat{\psi}_i$ are ML estimates for the complex amplitude $A_c$ and tilt angle $\psi$ under Hypothesis $i$. In Eqn. (2.18), arg min denotes the index $i$ that maximizes the argument. From Eqn. (2.9), the ML estimates are given by:

$$\{ \hat{A}_{ci}, \hat{\psi}_i \} = \arg \min_{A_c, \psi} (\tilde{S} - A_c b_i(\psi))^H \Sigma^{-1} (\tilde{S} - A_c b_i(\psi))$$

(2.19)

In Appendix A, we explicitly compute $\hat{A}_{ci}$ and $\hat{\psi}_i$ and find

$$\hat{\psi}_i = \arg \max_{\psi} \frac{|b_i(\psi)^H \Sigma^{-1} \tilde{S}|}{\|b_i(\psi)\|_\Sigma}$$

(2.20)

$$\hat{A}_{ci} = \frac{b_i(\hat{\psi}_i)^H \Sigma^{-1} \tilde{S}}{\|b_i(\hat{\psi}_i)\|_\Sigma^2}$$

(2.21)

Inserting these ML estimates in Eqn. (2.18), the decision rule for classification is:

$$i = \arg \max_{i=1,\ldots,M} \left[ \max_{\psi} \frac{|b_i(\psi)^H \Sigma^{-1} \tilde{S}|}{\|b_i(\psi)\|_\Sigma} \right]$$

(2.22)

For the detection rule in Eqn. (15), the choice of Bayes cost $\eta$ allows regulation of the false alarm rate. The detection rule is expressed in terms of $F_n$, a decreasing function determined by the SIRV clutter model given in Eqn. (2.9). In contrast, the polarimetric signature classification rule in Eqn. (22) is invariant to the particular form of the SIRV clutter model and depends only on the clutter covariance, $\Sigma$.

The detection approach in Eqns. (17) and (22) may be applied for any set of $M$ desired polarimetric signatures, as dictated by specific application.

The GLRT approach is extensively used in statistics literature. Here we replace the likelihood ratios in the optimal Bayes decision rule with the maximum likelihood
Figure 2.1: Probability of correct detection versus SNR using Bayes optimal detection with prior distributions on unknown parameters \((A, \rho, \psi)\) (solid line) and a GLRT (dashed line).

By approximating the expectation \(E_{A, \rho, \psi}[F_n(\|\hat{S} - A\hat{c}b_i(\hat{\psi})\|_\Sigma)]\) with \(F_n(\|\hat{S} - \hat{A}_c b_i(\hat{\psi})\|_\Sigma)\). Intuitive justification is that the argument of the expectation integral is largest at \((\hat{A}, \hat{\rho}, \hat{\psi})\) and becomes smaller as \(\|\hat{S} - A_c b_i(\psi)\|_\Sigma\) takes larger values because \(F_n(\cdot)\) is monotonically decreasing. This approximation performs well for smooth a-priori probability density functions for \(A, \rho, \psi\). The approximation is adopted both because the calculation of the expectation is computationally difficult and because the a-priori distribution of parameters may not be available. Moreover, the approximation results in very little performance loss, as seen in the next example.

In Figure 2.1, we present classification results for simulated polarimetric signatures using both the optimal test in Eqn. (16) (see Appendix B) and the GLRT in Eqn. (22). To facilitate numerical integration in Eqn. (16), signatures are simulated for the special case of Rayleigh distributed \(A\), uniformly distributed \(\rho\) and \(\psi\), equal
class priors, and Gaussian clutter with $\Sigma = I$. Probability of correctly classifying the scattering center as one of the $M = 8$ canonical classes is reported versus signal to noise level at the output of a whitening filter: $SNR = \frac{E[\|S\|_2^2]}{E[\|n\|_2^2]}$. For a specified set of $M$ canonical signatures, this SNR figure uniquely determines average classification performance. For each SNR level, $\eta$ is set for $100\%$ detection and classification performance is averaged from 5000 realizations of randomly selected classes and tilt angles. Although the GLRT approximation is suboptimal, we observe that the classification performance of the GLRT rule deviates from the optimal decision rule by less than $4.3\%$ over the entire SNR range (-15dB, 15dB).

2.3.4 Classification Significance

We provide a confidence measure on the class decision by reporting the a-posteriori class probabilities $P(H_k|\tilde{S})$. Exact computation of these probabilities requires expectation over the prior distribution on the parameters $(A, \rho, \psi)$:

$$P(\tilde{S}|H_i) = E_{A_c, \psi}[F_n((\hat{S} - A_c b_i(\psi))H\Sigma^{-1}(\hat{S} - A_c b_i(\psi)))] \quad (2.23)$$

We replace the distributions with a point mass assumption at their maximum likelihood values, consistent with our use of maximum likelihood ratios in Section 3.3. Thus,

$$P(\tilde{S}|H_i) \approx F_n((\hat{S} - \hat{A}_c b_i(\psi))H\Sigma^{-1}(\hat{S} - \hat{A}_c b_i(\psi)) = F_n(d_k^\text{min}) \quad (2.24)$$

where $d_k^\text{min}$ is the minimum weighted distance of scattering center $\tilde{S}$ to the $k^{th}$ canonical target cone. The cone is described in detail in Section 3.4. Then using Bayes Rule we compute a-posteriori probabilities as:

$$P(H_k|\tilde{S}) = \frac{P_k F_n(d_k^\text{min})}{\sum_{i=1}^{M} P_i F_n(d_i^\text{min})} \quad (2.25)$$
Figure 2.2: Notional representation of the geometry for the GLRT detector. Projection onto target cones yields ML estimates \((\hat{A}_c, \hat{\psi})\); the nearest cone yields the class decision.

For the Gaussian case with equal priors, Eqn. (2.25) simplifies to

\[
P(H_k|\tilde{S}) = \frac{\exp(-d_k^{\text{min}})}{\sum_{i=1}^{M} \exp(-d_i^{\text{min}})}
\]  \hspace{1cm} (2.26)

Equation 2.25 provides a measure of the certainty in deciding target class \(k\) from a measurement \(\tilde{S}\).

2.3.5 Geometrical Interpretation and Computational Cost

The detection rules in Section 3.3 admit a simple geometric interpretation and implementation in \(\mathbb{C}^3\). From Eqn. (2.4), variation of \((A_c, \psi)\) for a canonical polarimetric signature forms a cone in \(\mathbb{C}^3\). A cone is a set of vectors such that any element of the cone remains in the cone if scaled by an arbitrary complex constant. Note that a cone need not be convex. This geometry is notionally represented in \(\mathbb{R}^3\) in Figure 2.2.
The MAP polarimetric classification rule in Eqn. (2.22) selects the target cone closest to the measured scattering matrix \( \tilde{S} \) using a distance weighted by the inverse of the clutter covariance, \( \Sigma^{-1} \). For computation, the 1-by-3 complex vectors \( \frac{b_i(\psi)^H \Sigma^{-1}}{||b_i(\psi)||_\Sigma} \) in Eqn. (2.22) can be precomputed for sampled values of \( \psi \), leaving only inner products to be computed for classification of a measured signature, \( \tilde{S} \). For the results presented in Section 4, we have arbitrarily chosen a sampling increment of 1 degree. For the eight target classes in Table 2.2.2 and a sampling increment of 5 degrees, the decision rule requires evaluation of only 165 1-by-3 complex-valued inner products, owing to target symmetries.

The largest inner product provides not only the class decision but also the maximum likelihood estimates of orientation, amplitude, and phase \( (\hat{\psi}, \hat{A}, \hat{\rho}) \) from Eqns. (2.20-2.21). For the special cases of dihedral [7, 36], trihedral and helix targets, projection onto the cone can be computed in closed form. However, the ML estimate of \( \psi \) is different for each target class, even for \( \Sigma = I \).

From the Bayes solution in Eqn. (2.22), we observe that if \( \tilde{S} \) is classified to class \( m \), then \( c\tilde{S} \) is also classified to class \( m \), where \( c \) is any complex constant. Thus, the decision regions in Eqn. (2.22) are cones in \( \mathbb{C}^3 \). Note, however, that the cones need not have interior nor volume. Also, if \( \Sigma = I \) then rotated versions of \( \tilde{S} \) are also classified to the same class as \( \tilde{S} \). However, for a general clutter covariance matrix, the decision rule is not invariant to rotation of orientation angle, \( \psi \).

In interpreting Figure 2.2, we note that distances are weighted by \( \Sigma^{-1} \); thus cones are not circularly symmetric. Also, although \( \psi \) parametrizes the circumference of the cone for each target class, rotation of \( \psi \) does not correspond to an equal angular rotation of the target cones in the figure.
The set of all amplitudes, phases, and orientation angles for a canonical target constitute the surface of a cone in $\mathbb{C}^\ell$, and a scattering matrix $\hat{S}$ is classified to the nearest cone. Therefore, misclassifications are more likely if the cone of a canonical target is close to one of the other cones. For the eight scattering centers in Table 2, we present in Table 2.4 the minimum distances between normalized signatures and each cone, using a forest clutter covariance matrix [42] to compute weighted distances. We use the same covariance matrix in our simulated data examples in Section 4. The $i, j^{th}$ entry of the table gives the weighted distance of the normalized canonical signature, $\|\hat{S}_i(\psi)\| = 1$, to the cone for signature $j$; the distance is minimized over $\psi$ and therefore indicates a worst case evaluation. From the distance matrix, we observe target pairs which are prone to misclassification when the clutter is strong. For example, in row 1 the trihedral is most easily confused as a cylinder. The classification results in Section 4 confirm this intuition. The asymmetry of the distance matrix (distance$_{ij} \neq$ distance$_{ji}$) is a consequence of the colored clutter covariance, $\Sigma \neq I$. This asymmetry is likewise observed in the experiments.

2.4 Experimental Results

2.4.1 Simulated Data

In this section we simulate polarimetric canonical target signatures from Table 2 in K-distributed forest clutter [42] with a shape parameter $\alpha = 0.7806$, a forest clutter covariance matrix and SNR $= \frac{E[|S|^2]}{E[|\nu|^2]} = 20$ dB. $A$ is Rayleigh and $\rho$ is uniform distributed. The confusion matrix of the classification is given in Table 2.4.1. True targets are given in rows, and columns denote the decisions. For each SNR value, we computed 2500 trials with uniformly distributed class and orientation angle. The
<table>
<thead>
<tr>
<th>signature</th>
<th>distance to target cone</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1 Trih.</td>
<td>0.00</td>
</tr>
<tr>
<td>2 Dih.</td>
<td>1.72</td>
</tr>
<tr>
<td>3 Dipole</td>
<td>1.22</td>
</tr>
<tr>
<td>4 Cyl.</td>
<td>0.55</td>
</tr>
<tr>
<td>5 Nar. Di.</td>
<td>1.64</td>
</tr>
<tr>
<td>6 Qtr.W.</td>
<td>1.22</td>
</tr>
<tr>
<td>7 L. Helix</td>
<td>1.75</td>
</tr>
<tr>
<td>8 R. Helix</td>
<td>1.75</td>
</tr>
</tbody>
</table>

Table 2.4: Minimum distance from normalized signature to eight target cones using a forest clutter H-V basis covariance matrix

threshold \( \eta \) is chosen such that probability of detection is 100\% to illustrate the performance of the classifier rule. From Tables 4 and 5 we confirm the intuition that misclassification between classes is more probable when the corresponding cones are closer. Figure 2.3 shows the classification performance of the proposed decision rule. For 90\% reliable classification of the eight canonical signatures with equal priors, the figure shows that SNR must exceed 18 dB.

2.4.2 Measured Data

Next, the proposed classification rule is applied to fully-polarized radar data measured at Grayling, Michigan and imaged by Lincoln Laboratories [5]. Table 2.6 summarizes the synthetic aperture radar system. The imaged scene, M19p43, includes an array of twelve reflectors of various types and orientation angles, thereby permitting evaluation of detection results from measured data. The array, and corresponding HH polarization magnitude image, are shown in Figures 2.4 and 2.5. The image shows in greyscale the top 60 dB of a 34,000\( m^2 \) region. Using a 4250\( m^2 \) region void of targets,
Figure 2.3: Probability of correct classification vs SNR(dB)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Trih.</td>
<td>0.961</td>
<td>0</td>
<td>0</td>
<td>0.035</td>
<td>0.002</td>
<td>0.02</td>
<td>0.005</td>
<td>0</td>
</tr>
<tr>
<td>2 Dih.</td>
<td>0</td>
<td>0.823</td>
<td>0.005</td>
<td>0.002</td>
<td>0.147</td>
<td>0.02</td>
<td>0.003</td>
<td>0</td>
</tr>
<tr>
<td>3 Dipole</td>
<td>0</td>
<td>0</td>
<td>0.956</td>
<td>0.005</td>
<td>0.027</td>
<td>0.007</td>
<td>0.002</td>
<td>0.003</td>
</tr>
<tr>
<td>4 Cyl.</td>
<td>0.005</td>
<td>0</td>
<td>0.003</td>
<td>0.979</td>
<td>0.002</td>
<td>0.007</td>
<td>0.002</td>
<td>0</td>
</tr>
<tr>
<td>5 Nar. Di.</td>
<td>0</td>
<td>0.090</td>
<td>0.023</td>
<td>0.003</td>
<td>0.859</td>
<td>0.023</td>
<td>0.002</td>
<td>0</td>
</tr>
<tr>
<td>6 Qtr. W.</td>
<td>0</td>
<td>0.005</td>
<td>0.012</td>
<td>0.003</td>
<td>0.015</td>
<td>0.9617</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>7 L. Helix</td>
<td>0</td>
<td>0</td>
<td>0.005</td>
<td>0.002</td>
<td>0.002</td>
<td>0.012</td>
<td>0.979</td>
<td>0</td>
</tr>
<tr>
<td>8 R. Helix</td>
<td>0</td>
<td>0.002</td>
<td>0.002</td>
<td>0.003</td>
<td>0.005</td>
<td>0.008</td>
<td>0</td>
<td>0.970</td>
</tr>
</tbody>
</table>

Table 2.5: Simulated confusion matrix, forest clutter with SNR=20dB
the estimated covariance is specified by

\[
\hat{\Sigma} = 0.0132 \begin{bmatrix}
1 & 0 & 0.115 \\
0 & 0.206 & 0 \\
0.115 & 0 & 0.452
\end{bmatrix}
\]

Details on image formation and calibration are given in [5]; only targets 3 and 6 from Table 2.4.2 were used in the calibration procedure.

Classification results using Eqn. (2.22) and \( M = 4 \) are given in Table 2.4.2. The estimated orientation angle and the estimated posterior probability from Eqn. (2.25) are also listed in the table. The proposed classifier correctly identifies 11 of 12 objects in the Grayling reflector array. Further, for the correctly identified targets, the standard deviation of error in the estimated tilt angle is 3 degrees. For the reflector array, SNR varies from 11dB for the sphere to 46dB for the 8 ft trihedrals. The SNR is estimated using the estimated covariance \( \hat{\Sigma} \), tilt \( \hat{\psi} \) and amplitude \( \hat{A} \) in the SNR definition from Section 3.3.
Figure 2.5: HH polarization magnitude image of the reflector array corresponding to the dotted area in Figure 4

Table 2.6: P-3 radar system characteristics

<table>
<thead>
<tr>
<th>Aperture</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Shape</strong></td>
<td>Linear</td>
</tr>
<tr>
<td><strong>Integration Angle</strong></td>
<td>35 degrees</td>
</tr>
<tr>
<td><strong>Range Resolution</strong></td>
<td>0.33m</td>
</tr>
<tr>
<td><strong>Cross-Range Resolution</strong></td>
<td>0.66m</td>
</tr>
<tr>
<td><strong>Waveform</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Type</strong></td>
<td>Linear Chirp</td>
</tr>
<tr>
<td>$f_{min}$</td>
<td>215 MHz</td>
</tr>
<tr>
<td>$f_{max}$</td>
<td>730 MHz</td>
</tr>
<tr>
<td>Reflector Type</td>
<td>Classification</td>
</tr>
<tr>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Trihedral (8ft)</td>
<td>1</td>
</tr>
<tr>
<td>Trihedral (8ft)</td>
<td>1</td>
</tr>
<tr>
<td>Trihedral (8ft)</td>
<td>1</td>
</tr>
<tr>
<td>Dihedral</td>
<td>2</td>
</tr>
<tr>
<td>Dihedral</td>
<td>2</td>
</tr>
<tr>
<td>Dihedral</td>
<td>2</td>
</tr>
<tr>
<td>Sphere (1m)</td>
<td>1</td>
</tr>
<tr>
<td>Trihedral (1m)</td>
<td>4</td>
</tr>
<tr>
<td>Tophat</td>
<td>2</td>
</tr>
<tr>
<td>Tophat</td>
<td>2</td>
</tr>
<tr>
<td>Tophat</td>
<td>2</td>
</tr>
<tr>
<td>Tophat</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 2.7: Classification results for the P3 Calibration Array

In addition to classifying known targets, the proposed M-ary detection rule is applied to an image region containing an unspecified vehicle. The top 40 dB of the HH polarization magnitude image is shown in Figure 2.6 (a). Choice of the vehicle scene is motivated by the conjecture that backscatter will be dominated by a virtual dihedral formed between the ground and the vehicle [1]. Figure 2.6 (b) shows pixels classified as dihedrals with with horizontal tilt angle ($\psi = 0 \pm 10$) and with a-posteriori probability exceeding 99%. Figure 2.6 (c) shows pixels classified as dihedrals with with horizontal tilt angle ($\psi = 0 \pm 10$) and with a-posteriori probability exceeding 90%. The example points to the potential value of polarimetric signatures for vehicle detection using this 109% relative bandwidth radar.

The model in Eqn. (8) approximates scattering at a single frequency and angle, whereas the Grayling SAR imagery gives a weighted average of scattering across two
Figure 2.6: Sample Target from the MIT Lincoln Lab P-3 polarimetric imagery
octaves of bandwidth and 35 degrees of aspect angle. Detection results suggest that
the simple model and correspondingly simple detector are legitimate approximations
owing to the electrical size of the reflectors. Extension to multi-pixel detector for
wideband measurements of frequency and angle dependent scattering in considered
in [18, 39].

2.5 Conclusion

We have presented a Bayes approach to detection and classification of polarimetric
radar signatures in spherically invariant random clutter. A particular choice of
the cost matrix in Bayes risk leads naturally to a two-stage classifier where the first
stage is a variable false-alarm rate detector which rejects clutter, and the second stage
classifies the detection in one of the $M$ classes. Further, despite the non-Gaussian
clutter model, the classification is performed by computationally simple projection
onto cones in $\mathbb{C}^3$; the projections also yield the maximum likelihood estimates of or-
ientation angle, amplitude and phase. Additionally, distance from the cones provides
a measure of uncertainty in the class decision. The detection approach is applicable to
both full and incomplete polarization measurements. The Bayes decision rules can be
used both for detection and for estimation of reflector types. Using fully polarimetric
measurements from an airborne SAR system, we observed 11 correct identifications
of 12 known reflectors with a standard deviation of 3 degrees in tilt angle estimation.
Analysis suggests that for an eight class problem 18 dB SNR is sufficient to provide
90% correct classification of reflector type.

The frequency and aspect independent scattering model considered here yields
simple detection rules which can be applied to single pixels in polarimetric SAR.
imagery. The Bayes approach can be extended to wideband, wide-angle scattering behaviors [18, 39], resulting in multi-pixel tests that provide a processing gain with associated increase in computational complexity.
CHAPTER 3

MULTICHANNEL IMAGING FOR WIDE-BAND SYNTHETIC APERTURE RADAR SYSTEMS

3.1 Introduction

Synthetic Aperture Radar (SAR) systems synthesize the effect of an antenna array using the motion of a single sensor. Scattering data collected over the large aperture of the synthesized array can be coherently processed to produce high resolution imagery[9]. The sensor may measure scattered fields in two independent polarizations to provide additional characterization of the remotely imaged objects[25, 19]. Polariometric SAR imagery is used for all-weather mapping, search-and-rescue[61], mine detection[44] and target recognition[42].

Remote radar imaging of objects obscured by tree canopy or buried underground requires long wavelengths for propagation through the obscuring foliage or soil. Additionally, a large bandwidth is required for high spatial resolution. Together, the dual requirements on wavelength and bandwidth demand a radar sensor with bandwidth exceeding one quarter the center frequency, i.e., an ultra wide-band (UWB) radar. Moreover, image resolution is governed not only by bandwidth but also by the span of aspect angles provided by the synthetic array. For a given resolution, the required angle diversity is proportional to the center frequency; therefore, UWB
SAR systems require much larger angle diversity than high frequency (e.g., X- or K-band) radars. Any UWB antenna will have a magnitude and phase response that varies as a function of frequency, aspect angle, depression angle, and polarization of the incident energy. Accordingly, polarimetric returns must be processed jointly as a multi-channel, wideband signal.

In the Section 3.2, we review the geometry of UWB SAR data collection and present a system model for UWB polarimetric imaging. We illustrate the artifacts in conventional single channel imaging due to the inappropriate assumption of uniform antenna response. In Section 3.3 we present multi-channel processing techniques for improved polarimetric purity and a more localized point spread function. The proposed techniques are applied in Section 3.4 using measured UWB polarimetric antenna responses.

3.2 System Model

In this section we formulate a system model for wideband, wide-angle, polarimetric SAR imaging. The imaging geometry is given in Figure 3.1. We consider the monostatic SAR geometry in which a single sensor both transmits and receives. The standard notation is from [51]. The radar illuminates a region with a large bandwidth pulse and records the resulting echo signal at the same position. The \((x, y)\) vector represents the spatial coordinates in the imaging plane, \(z = 0\). The \(x\)-axis direction is perpendicular to the aperture synthesized by the path of the radar, and the \(y\)-axis direction is termed the cross-range, or along-track, direction. The radar position \((0, u, z_u)\) is shown projected onto the imaging plane, \(z = 0\). The along-track position, \(u\), of the radar takes values, \(u_i\), along the aperture. Because the sensor velocity is
Figure 3.1: Ground plane radar geometry

much smaller than the propagation velocity $c$, we assume that the radar is stationary at $(0, u, z_u)$ for the monostatic transmit-and-receive pair at location $u$.

The radar illuminates the target area with a pulse signal $p(t)$. The imaged area is centered at $(x_0, y_0)$. The local coordinates in the imaged area are given by $(x', y')$. The round trip time-delay of the echoed signal from the radar at point $(0, u, z_u)$ to a point scatterer at local coordinates $(x', y')$ is given by

$$2 \frac{1}{c} \sqrt{(x' + x_0)^2 + (y' + y_0 - u)^2 + z_u^2}$$  \hspace{1cm} (3.1)

If we assume the scene is composed of point scatterers on the imaging plane, then the recorded signal at the radar position $u$ is given by

$$F(u, \omega) = P(\omega) \int \int \frac{1}{r^2} s(x', y', \omega, \theta) e^{-j2k \sqrt{(x' + x_0)^2 + (y' + y_0 - u)^2 + z_u^2}} dx' dy'$$  \hspace{1cm} (3.2)
where \( F(u, \omega) \) is a 2 \times 2 matrix representing the four polarimetric channels of data, \( P(\omega) \) is the Fourier Transform of the pulse function, \( k \) is given by \( \omega/c \), and \( r = \sqrt{(x' + x_0)^2 + (y' + y_0 - u)^2 + z_a^2} \). The range factor \( 1/r^2 \) is compensated by preprocessing of the range lines. \( s(x', y', \omega, \theta) \) is the scattering matrix for the scatterer at \((x', y')\) in the imaging plane, for look angle \( \theta \) and radian frequency \( \omega \). For notational simplicity, the frequency and look angle dependence of the scattering matrix will be suppressed in the remainder of the paper. The scattering matrix \( s(x', y') \) is the linear transformation that maps an incident polarization state to the scattered polarization state. For a horizontal and vertical polarization basis, the matrix of scattering coefficients is given by

\[
s(x', y') = \begin{bmatrix} s_{hh}(x', y') & s_{hv}(x', y') \\ s_{vh}(x', y') & s_{vv}(x', y') \end{bmatrix}
\]  

(3.3)

In ultra-wideband radar systems the system model must incorporate the wide-band, wide-angle response of the receiving and transmitting elements. A system model that accounts for the antenna response is given by

\[
F(u, \omega) = P(\omega) \int \int R_\phi(\omega) s(x', y') T_\phi(\omega) e^{-j2k\sqrt{(x'+x_0)^2+(y'+y_0-u)^2+z_a^2}} dx'dy'
\]

(3.4)

where the four-channel polarimetric response \( F(u, \omega) \) at aperture position \((0, u, z_a)\) and frequency \( \omega \) is obtained by the superposition of scattering contributions with scattering matrices \( s(x', y') \). The scattering matrix is filtered by receive and transmit antenna transfer function matrices \( R_\phi(\omega), T_\phi(\omega) \) at the angles \( \phi = [\phi_a, \phi_d] \)

\[
\phi_a = \tan^{-1}\left(\frac{y'+y_0-u}{x'+x_0}\right) \text{ azimuth angle}
\]

\[
\phi_d = \tan^{-1}\left(\frac{z_a}{x'+x_0}\right) \text{ depression angle}
\]

(3.5)

If the size of the imaged area is small with respect to the distance to the aperture, then the deviation of the response due to the depression and azimuth angles can be
approximated by one azimuth angle. If we assume a linear aperture, define the slant range \( x \) as \( x = \sqrt{(x' + x_0)^2 + z_\theta^2} \), the cross-range \( y \) as \( y = y' + y_0 \), and assume the scene lies on the \( xy \) plane, then the system model in Eqn. (3.4) can be rewritten without loss of generality as

\[
F(u, \omega) = \int \int \mathcal{A} \left( \omega, \tan^{-1} \left( \frac{y-u}{x} \right) \right) S(x, y) e^{-j2k\sqrt{x^2 + (y-u)^2}} \, dx \, dy
\]  

(3.6)

where the matrices \( F \) and \( S \) are vectorized so that the shifted scattering matrix is represented by a \( 4 \times 1 \) vector \( S(x, y) \). Delay-and-sum beamformer imagers produce images directly on the imaging plane \( z = 0 \) with coordinates \((x', y')\) not on slant plane represented with coordinates \((x, y)\). Therefore the techniques here can be applied to imagery on the image plane if the image is scaled in range direction using \( x = \sqrt{(x' + x_0)^2 + z_\theta^2} \). In practice we can approximate the relation by simple linear scaling \( x = x' / (\cos \frac{\theta}{x_0}) \).

The \( 4 \times 4 \) matrix \( \mathcal{A} \left( \omega, \tan^{-1} \left( \frac{y-u}{x} \right) \right) \) combines from Eqn. (3.4) the effects of receive and transmit antenna functions and the pulse function. The matrix operator \( \mathcal{A} \) is a function of both frequency and spatial coordinates \((x, y)\), due to the dependence of the antenna response on both the frequency and the incident angle of scattered energy. The combination of frequency and spatial dependence makes Eqn. (3.6) an integral equation not invertible by deconvolution.

Current imaging algorithms[38] for wideband, wide-angle SAR systems are based on Eqn. (3.2) and provide an approximate inversion of the integral operator in Eqn. (3.4), but without consideration of the effect of antenna transfer functions. This type of processing results in images with unsatisfactory polarization purity. Additionally, the energy of a scattering response in an image is dispersed due to the non-ideal phase and magnitude responses of the antennas. Standard polarimetric
calibration [58] for C-band radars adopts a system model that is independent of frequency and angle; the calibration is achieved with a matrix multiply directly applied to single pixels in complex-valued imagery.

To illustrate the effects of an uncompensated antenna pattern on UWB, wide-angle, polarimetric SAR imaging, we present a simple example. The example uses simulated data to represent the data collection geometry and antenna pattern present in the Army Research Laboratory UWB BoomSAR data collection at Aberdeen Proving Ground[40, 49]. A 10m by 10m region of interest (ROI) containing two trihedral s is imaged using a delay-and-sum algorithm. Polarimetric imagery is formed using a 90 degree integration angle and a 400 MHz bandwidth (300–700 MHz). An additive complex Gaussian white noise with a signal-to-noise ratio of 20dB is used in simulation. The imaged chip is centered at \((x_0, y_0) = (35m, 0m)\) and the antenna height is \(z = 35m\). The aperture spacing is uniform at 0.15m. The UWB, wide-angle antenna pattern was measured at the Ohio State University Electro-Science Laboratory[62].

In Figure 3.2(a), the returns from two ideal point scatterers are synthesized with ideal antenna transfer functions and imaged using a delay-and-sum beamformer. The point spread function is determined by the antenna bandwidth and integration angle. In contrast, inclusion of the measured antenna transfer functions in synthesizing the radar returns yields Figure 3.2(b) when imaged with the same imaging method. In each image, 40 dB dynamic range is depicted in grayscale. We observe that the cross-polarized image, although ideally zero for the synthesized trihedral scatterers, has a peak response 11dB below the peak response in the HH-polarized image. Also, the peak response in the VV-polarized image exceeds the peak response in the HH-polarized image by 10dB. Further, the trihedral responses in the co-polarized images
are not localized due to the nonideal phase and magnitude response of the antenna over the large range of look angles and frequencies.

3.3 Multi-channel Techniques for Antenna Pattern Compensation

In this section we introduce three techniques for compensation of antenna patterns in multi-channel radar data. All techniques require measured antenna patterns over the range of frequencies and look angles of the radar. The first technique uses a local approximation to the system model, and therefore is suitable for the processing of small image chips. Its performance also depends on the smoothness of the antenna transfer function over a range of depression and azimuth angles. The second technique proposes an inversion method for the system model in a transform domain. This Doppler domain technique processes the Fourier transform of the range lines, to compensate for the effects of the antenna transfer function. Third technique is an image domain implementation of the doppler domain method. All methods require interpolation if the antenna response functions have been measured for a discrete set of frequencies and look angles.

3.3.1 Local Approximation Method

We propose a local approximation method for inversion of the system model in Eqn. (3.6). The technique is applicable to a small region of interest (ROI) and is consistent with existing delay-and-sum imaging approaches which form a mosaic of image tiles[38]. Mosaic techniques are motivated by a desire to maintain near-constant point spread functions for near-field imaging geometries that are incompatible with plane-wave imaging assumptions.
Figure 3.2: Effect of antenna functions on imaging: (a) four channel image with antenna pattern omitted; (b) four channel image with effect of antenna pattern
For each aperture position, the recorded range line is time gated around the center pixel of the ROI to obtain $\tilde{f}(u, t)$. Then each truncated range line $\tilde{f}(u, t)$ is Fourier transformed to obtain $\hat{F}(u, \omega)$. If the region of interest is sufficiently small we can approximate the matrix function $A\left(\omega, \tan^{-1}\left(\frac{y_0 - u}{x_0}\right)\right)$ with $A\left(\omega, \tan^{-1}\left(\frac{y_0 - u}{x_0}\right)\right)$, for every $x, y$ in the ROI. That is, for all pixels in the ROI, the antenna response at location $(0, u, z_u)$ is assumed to be equal to the antenna response corresponding to the aspect angle calculated with respect to the center pixel of the ROI. This approximation is valid if the ROI is sufficiently small and if the antenna pattern is a smooth function of aspect angle. Using the approximation we can write the system model for $\hat{F}(u, \omega)$ as

$$\hat{F}(u, \omega) = A\left(\omega, \tan^{-1}\left(\frac{y_0 - u}{x_0}\right)\right) \int \int S(x, y)e^{-j2k\sqrt{x^2 + (y - u)^2}} \, dx \, dy \quad (3.7)$$

To form the image on the ROI, first $\hat{F}(u, \omega)$ is decorrelated with a $4 \times 4$ matrix

$$\hat{F}(u, \omega) = \left(A^H\left(\omega, \tan^{-1}\left(\frac{y_0 - u}{x_0}\right)\right) A\left(\omega, \tan^{-1}\left(\frac{y_0 - u}{x_0}\right)\right) + \mu(\omega, u) I\right)^{-1} A^H\left(\omega, \tan^{-1}\left(\frac{y_0 - u}{x_0}\right)\right) \hat{F}(u, \omega) \quad (3.8)$$

This requires a $4 \times 4$ matrix multiplication for each frequency sample of $\hat{F}(u, \omega)$. In the inversion, we use Tikhonov regularization[34, 26] with the regularization parameter $\mu$. The choice of the parameter $\mu(\omega, u)$ provides a tradeoff between accuracy and stability of the inversion. $\hat{F}(u, \omega)$ is the solution to the optimization problem given by

$$\min \|A\hat{F} - \tilde{F}\|^2 + \mu\|\hat{F}\|^2 \quad (3.9)$$

The choice of the regularization parameter in discrete ill-posed problems has been examined by Hanson[28, 27]. For best performance, the regularization parameter should be varied as a function of $u$ and $\omega$, according to the ill-conditioning of the
matrix operator $\mathcal{A} \left( \omega, \tan^{-1} \left( \frac{z_0}{\sqrt{x_0^2 + y_0^2}} \right) \right)$. In particular if an estimate of the noise power $\sigma^2(\omega, u)$ is available for each frequency and aperture position, then the regularization parameter $\mu(\omega, u)$ can be chosen to satisfy $\| \mathcal{A} \hat{F} - \hat{F} \|^2 = \sigma^2(\omega, u)$.

Next, the inverse Fourier transform of the frequency samples yields the compensated range line records for all four channels. Finally, a single-channel delay-and-sum imaging algorithm is applied to compensated range lines from each polarization channel to recover the scattering matrix $\mathbf{s}(x, y)$ at each pixel in the ROI. This method provides a conceptually simple, local approximation to the inversion of Eqn. (3.6).

### 3.3.2 Doppler Domain Technique

In this section we use reciprocity of the antenna and a spherical wave decomposition to derive a system model in the transform (Doppler) domain, where the effect of the antenna is multiplicative. Instead of expressing returns using imperfect receive and transmit antenna responses, we can equivalently adopt a radar model with an ideal antenna and assume that the spherical waves emanating from the scattering centers have a shading equal to the two-way antenna response. Accordingly, in this equivalent representation the scattering centers have a frequency and aspect dependent response $\mathcal{A}(\omega, \phi) \mathbf{S}(x, y)$. This radiating reflector [6] assumption leads to the following radar system model

$$
\mathbf{F}(u, \omega) = \int \int \left( \mathcal{A}(\omega, \phi) \mathbf{S}(x, y) \right) \frac{1}{r} e^{-j2k \sqrt{x^2 + (y-u)^2}} dx dy
$$

(3.10)

where $\mathcal{A}(\omega, \phi)$ is the $4 \times 4$ polarimetric antenna response matrix introduced in Eqn. (3.6). (We can again pre-process the range lines for the extra range factor).

The plane wave decomposition of spherical waves is a commonly used tool in SAR signal processing [6, 51]. Here we decompose the two dimensional spherical wave
\[
\frac{1}{\sqrt{k^2}}e^{-j2k\sqrt{x^2+(y-u)^2}} \text{ into plane waves emanating in direction } \phi(k_0) = \tan^{-1}\left(\frac{k_0}{\sqrt{(4k^2-k_0^2)}}\right)
\]
for \(k_0\) between \(-2k\) and \(2k\), where each direction is multiplied by the appropriate weight \(\frac{1}{\sqrt{(4k^2-k_0^2)}}A(\omega, \phi(k_0))\). This weighted decomposition leads to a multiplicative antenna model in the Doppler domain. We note that the same multiplicative antenna model was derived for single channel systems by Rau and McClellan [48] and Soumekh [52] using stationary phase techniques [9].

Using this decomposition we can rewrite the system model, suppressing a complex valued constant

\[
F(u, \omega) = \int_{-2k}^{2k} \int \frac{1}{\sqrt{(4k^2-k_0^2)}}A\left(\omega, \tan^{-1}\left(\frac{k_0}{\sqrt{(4k^2-k_0^2)}}\right)\right)S(x, y) \\
\exp\left\{-j\left[\sqrt{(4k^2-k_0^2)}x + k_0(y-u)\right]\right\} dx dy dk_0 \tag{3.11}
\]

Integrating with respect to \(x\) and \(y\) we obtain

\[
F(u, \omega) = \int_{-2k}^{2k} \frac{1}{\sqrt{(4k^2-k_0^2)}}A\left(\omega, \tan^{-1}\left(\frac{k_0}{\sqrt{(4k^2-k_0^2)}}\right)\right)S_{XY}(\sqrt{(4k^2-k_0^2)}, k_0) \exp(jk_0u) dk_0
\]

where \(S_{XY}\) denotes the 2D Fourier transform of \(S\). Applying the Fourier transform with respect to \(u\) on both sides we obtain

\[
F(k_u, \omega) = \frac{1}{\sqrt{(4k^2-k_u^2)}}A\left(\omega, \tan^{-1}\left(\frac{k_u}{\sqrt{(4k^2-k_u^2)}}\right)\right)S_{XY}(\sqrt{(4k^2-k_u^2)}, k_u) \tag{3.12}
\]

We note that for a radar model with idealized antenna response the same derivation will result in the system model [51]

\[
F(k_u, \omega) = \frac{1}{\sqrt{(4k^2-k_u^2)}}S_{XY}(\sqrt{(4k^2-k_u^2)}, k_u) \tag{3.13}
\]

From Eqn. (3.12), the effects of the antenna function can be simply compensated in the spatial Doppler domain by

\[
\hat{F}(k_u, \omega) = \left(A^H\left(\omega, \tan^{-1}\left(\frac{k_u}{\sqrt{(4k^2-k_u^2)}}\right)\right)A\left(\omega, \tan^{-1}\left(\frac{k_u}{\sqrt{(4k^2-k_u^2)}}\right)\right) + \mu(\omega, k_u)I \right)^{-1} \\
A^H\left(\omega, \tan^{-1}\left(\frac{k_u}{\sqrt{(4k^2-k_u^2)}}\right)\right)F(k_u, \omega) \tag{3.14}
\]
Again \( \hat{\mathbf{F}}(k_u, \omega) \) is the solution to the optimization problem given by

\[
\min \| \mathcal{A} \hat{\mathbf{F}} - \mathbf{F} \|^2 + \mu \| \hat{\mathbf{F}} \|^2
\]  \hspace{1cm} (3.15)

For best performance the choice of the parameter \( \mu(\omega, k_u) \) should be varied as a function of \( \omega \) and \( k_u \). The method described in the previous section can be applied for the choice of \( \mu(\omega, k_u) \), if an estimate of the noise power is available for each \( \omega, k_u \) pair. The polarimetric calibration in Eqn. (3.14) can be incorporated in the multiplication step of migration imaging [6].

**Statistical Estimation Approach**

If the statistics of the noise can be determined it can be used to pose the inversion problem as a statistical estimation problem. Here, we briefly discuss the case where the additive noise vector is modeled as zero mean gaussian random vector with known covariance matrix. Using the results of the previous section we can rewrite the system model as:

\[
\mathbf{F}(m) = \mathcal{A}(m)\mathbf{F}_p(m) + \mathbf{N}(m)
\]

where \( m = (k_u, \omega) \) and \( \mathbf{F}_p \) is the received signal for a system with ideal antenna and infinite SNR.

**Regularized Maximum Likelihood Estimate**

If we assume no prior distribution on \( \mathbf{F}_p(m) \) and Gaussian \( \mathbf{N}(m) \) with covariance matrix \( \mathbf{R}_N(m) \), the regularized maximum likelihood estimate can be found as the solution to the maximization problem:

\[
[(\mathbf{F}(m) - \mathcal{A}(m)\mathbf{F}_p(m))^H\mathbf{R}_N^{-1}(m)(\mathbf{F}(m) - \mathcal{A}(m)\mathbf{F}_p(m))] + \mu(m)\|\mathbf{F}_p(m)\|^2
\]  \hspace{1cm} (3.16)
\[ \hat{F}_{RML}(m) = (A^H(m)R_N^{-1}(m)A(m) + \mu(m)I)^{-1}A^H(m)R_N^{-1}(m)F(m) \]  

(3.17)

The regularization is applicable when the error covariance of the estimate \( A^{-1}F(m) \) is unacceptably high. The regularization introduces a small bias in the estimate in order to reduce the variance. The covariance matrix of the estimate \( A^{-1}F(m) \) can be computed as:

\[ (A^H(m)R_N^{-1}(m)A(m))^{-1} \]

The error covariance can become very large at frequencies where \( A(m) \) has a small singular value.

**Maximum a posteriori estimate**

It is also illustrative to derive the maximum a posteriori estimate of \( F_p(m) \) under the Gaussian distribution assumption on both \( F_p(m) \) and \( N(m) \) (with zero mean and covariance matrices \( R_{F_p}(m) \) and \( R_N(m) \), respectively). The MAP estimate is given by

\[ \hat{F}_{MAP}(m) = (A^H(m)R_N^{-1}(m)A(m) + R_{F_p}^{-1}(m))^{-1}A^H(m)R_N^{-1}(m)F(m) \]

(3.18)

If we compare the regularized ML estimate of \( F_N(m) \) given in Eqn. (3.17) with the MAP estimate given in Eqn. (3.18), we observe that the regularized ML estimate is identical to the MAP estimate if we assume Gaussian distribution with covariance matrix \( R_{F_p}(m) = \frac{1}{\mu(m)}I \) for the signal \( F_p(m) \). We also observe that the unregularized estimate \( A^{-1}F(m) \) approaches to the MAP estimate in the limit as the SNR level goes to infinity.
3.3.3 Image Postprocessing Method

We can get an equivalent implementation of the Doppler domain method discussed in the previous section on the images formed by an imaging method such as: delay-and-sum beamforming, migration or convolution backprojection. We discuss the model based on a migration method ($\omega - k$) imager [9]. However, the result of this section can be easily modified for other imaging techniques.

The model in Eqn. (3.12) can be rewritten by defining $k_x = \sqrt{(4k^2 - k_u^2)}$ and $k_y = k_u$

$$\frac{1}{k_x}A \left( \frac{c}{2} \sqrt{k_x^2 + k_y^2}, \frac{k_y}{k_x} \right) S_{XY}(k_x, k_y) = F \left( k_y, \frac{c}{2} \sqrt{k_x^2 + k_y^2} \right) \quad (3.19)$$

Now consider an ($\omega - k$) domain imager, which obtains an approximate inversion of the model in Eqn. (3.13) to get an estimate, $M_{XY}(k_x, k_y)$, of the image from range line data. That is

$$M_{XY}(k_x, k_y) \approx k_y F \left( k_y, \frac{c}{2} \sqrt{k_x^2 + k_y^2} \right) \quad (3.20)$$

In the presence of antenna effects our result in Eqn. (3.12) implies that the image will contain the effect of the antenna as :

$$M_{XY}(k_x, k_y) \approx A \left( \frac{c}{2} \sqrt{k_x^2 + k_y^2}, \tan^{-1}\left( \frac{k_y}{k_x} \right) \right) S_{XY}(k_x, k_y) \quad (3.21)$$

Rau and McClellan [48] have shown that a similar result holds true for the constant angle backprojection (CIAB) imaging using stationary phase techniques. Their model also includes a term which represents the degradation resulting from the (CIAB) imaging.
The result in Eqn. (3.21) implies that the effects of the antenna pattern can be compensated in the spatial frequency domain using

$$S_{XY}(k_x, k_y) \approx \left( A^H \left( \frac{c}{2} \sqrt{k_x^2 + k_y^2, \tan^{-1}\left( \frac{k_y}{k_x} \right)} \right) A \left( \frac{c}{2} \sqrt{k_x^2 + k_y^2, \tan^{-1}\left( \frac{k_y}{k_x} \right)} \right) + \mu(k_x, k_y) I \right)^{-1} \times A^H \left( \frac{c}{2} \sqrt{k_x^2 + k_y^2, \tan^{-1}\left( \frac{k_y}{k_x} \right)} \right) M_{XY}(k_x, k_y).$$
3.4 Results

Figure 3.3 presents the results of the local approximation method of Section 3.3.1 applied to the simulated noisy data described above. A regularization parameter $\mu = 0.001$ was used for each $(\omega, u)$ value. In each image 40 dB dynamic range is depicted in grayscale. We first note that the responses from trihedrals are more localized than the responses in the uncalibrated imagery. Additionally, the peak response in the cross-polarized image is now 33.6 dB below the peak response in the co-polarized images. Also, the peak response in the VV polarized image is only 0.35 dB above the peak response in the HH polarized image. Therefore, the polarization signature derived from the calibrated imagery is considerably closer to the ideal trihedral signature than the signature obtained from the uncalibrated images. We note that the results of the algorithm degrade at the pixels farther away from the center; this is due to the local approximation nature of the algorithm.

Figure 3.4 compares cross-sections from the images of Figure 3.4, constructed from range lines processed by the local approximation method, with cross-sections from the images of Figure 3.2a, synthesized with ideal antenna response functions. The cross-section through each image is taken at the off-center trihedral position along the cross-range direction.

Figure 3.5 presents the results of the Doppler domain method of Section 3.3.2 applied to the simulated noisy data described above. A regularization parameter $\mu = 0.001$ was used for each $(\omega, k_u)$ value. In each image 40 dB dynamic range is depicted in grayscale. For this implementation, the peak response in the VV polarized image is 0.4 dB above the peak response in the HH polarized image, and
Figure 3.3: Results for the local approximation method: four channel image constructed from range lines processed using the local approximation algorithm.
Figure 3.4: Results for the local approximation method: cross-sections from calibrated images in Fig. 3.3 (solid) and the ideal antenna response in Fig. 3.2 (dashed)
Figure 3.5: Results for the Doppler domain method: four channel image constructed from range lines processed in the Doppler domain.

The peak response in the cross-polarized image is 52.5 dB below the peak response in the co-polarized image.

Figure 3.6 compares cross-sections from the images of Figure 3.5 constructed from range lines processed by the Doppler domain method, with the cross-sections from the images of Figure 3.2a, synthesized with ideal antenna response functions. The cross-section through each image is taken at the off-center trihedral position along the cross-range direction.

Figure 3.7 presents the results of the image postprocessing applied to the simulated noisy data described above. A regularization parameter $\mu = 0.01$ was used for
Figure 3.6: Results for the Doppler domain method: cross-sections from calibrated images in Fig. 3.5 (solid) and the ideal antenna response in Fig. 3.2a (dashed)
Figure 3.7: Results for the image postprocessing method: four channel postprocessed imagery.

...each \((k_x, k_y)\) value. In each image 40 dB dynamic range is depicted in grayscale. For this implementation the peak response in the VV polarized image is 0.8 dB above the peak response in the HH polarized image, and the peak response in the co-polarized image is now 21 dB below the peak response in the cross-polarized images. The imperfect results of the algorithm may be due to the implementation of the interpolation and additional effects introduced by the imager. Improved implementation of this algorithm in spatial domains remains as a future topic.

Polarimetric imagery has been used for detection of scattering signatures [19, 42]. However, due to the imaging point spread function, single pixel detection on
full polarization ultra wide band images can be improved by extending pixel based polarization features to template based features computed locally over a small number of pixels. Moreover, the nonideal antenna response can be incorporated into these templates for improved detection on uncalibrated imagery.
CHAPTER 4

SEQUENTIAL DECISION PROBLEMS WITH COST CONSTRAINTS

4.1 Introduction

In most statistical signal processing problems, a fixed number of observations is available to the detector to make inference about the prevailing hypothesis. An alternative approach is to fix desired performance levels and allow the number of samples to vary to achieve this performance. This approach is primarily motivated by minimizing the cost of observations for a given level of performance; however, the design of such sequential procedures usually does not consider the expected cost of observation as one of the desired performance criteria. In this chapter we address the problem of designing sequential procedures that maximize detection performance under expected experimentation (observation) cost constraints.

The development is motivated by feature-based ATR systems in which several statistics, called features, are computed from imagery, such as Synthetic Aperture Radar (SAR) imagery. Within an automatic target recognition and detection algorithm, a large volume of sensor data is accessed for performing a multiple hypothesis test with a constraint on computation. A suite of detection statistics, called features, is available; each feature is computed from measured imagery and computation of
Figure 4.1: Block diagram of a non-adaptive, sequential, multi-feature three stage automatic target recognition system.

Each feature imposes a specific computational load. The high data rates and real-time processing requirements for wide area surveillance have given rise to a non-adaptive, staged decision strategy as the de facto approach to ATR. Each stage of the ATR system computes discrimination statistics to reduce false alarms while maintaining high probability of detection. Screening false alarms reduces the data rate faced by the subsequent stages; computational complexity of the discrimination statistics is allowed to increase as the data rate drops. An example of this non-adaptive decision rule [41] is depicted in Figure 4.1.

For the thousands of pixels in a SAR image and the many compound target hypotheses in even a simple ATR problem, the posterior probabilities are intractable. Instead, decision statistics are computed from the image. A feature may be, for example, the output of a constant false alarm rate (CFAR) detector, a fractal dimension, an inertia, a correlation score, etc. The observation vector is the set of available features computed from the SAR imagery. The hypothesis, or “decision class,” determines the joint distribution of the observation vector. For example, features might
have distribution according to the hypothesis “T72 tank present” or the hypothesis “clutter.”

The application of sequential tests in ATR systems is not restricted to feature selection problems. Sequential tests with repeated experimentation (data collection) are adopted in ATR systems to minimize target acquisition time, for a given set of error probabilities [32].

The use of sequential decision strategies is not constrained to radar applications. For example, consider a medical diagnosis scenario. In determining the presence (and if present, type) of heart disease, a physician may choose to request an electrocardiogram (ECG) following a physical examination and patient interview. Noninvasive measurement of left ventricular ejection fraction (LVEF) or heart rate variability might be requested next. Finally, patient risk might be aggressively assessed by invasive induction of ventricular tachyarrhythmia using programmed direct electrical stimulation of the heart tissue. However, the risk assessment, and the sequence of decisions for the patient to undergo which, if any, additional tests, is based on the test results already observed. Further, the decision is made by jointly considering previous test results. For example, an abnormal ECG or low LVEF alone is not alarming, but the combination is highly indicative of risk for sudden cardiac death [20].

In the next section we present a probabilistic framework for formally describing decision problems where the decisions are made under expected cost constraints. In section 4.3 the Bayesian decision problem is stated, and solution for two special cases of interest is reviewed. In section 4.4 two new problem formulations are proposed, for sequential detection with computational cost constraints. Then the relationship between solutions of these two problems and the Bayes solutions is established. These

58
results are illustrated on a computed example. In section 4.5 approximate design
techniques are discussed.

4.2 Notation and Mathematical Model

In this section, we present a probabilistic framework adopted from [37], for fo-
mally describing decision problems where the decisions are made under expected cost
constraints. Specific examples of this problem are binary and M-ary sequential hy-
pothesis testing problems and dynamic feature selection problems.

The elements of this sequential decision problem are

- States of Nature $\Theta = \{\theta_1, \ldots, \theta_M\}$

- Set of final decisions $\mathcal{A} = \{a_1, \ldots, a_M\}$

- The loss function $L : \mathcal{A} \times \Theta \rightarrow \mathbb{R}$

- The observation vector $X$

- The state space $\Lambda$

- Available experiment correspondence $\mathcal{E} = \Lambda \rightarrow \mathbb{N}$

- Cost function $c : \Lambda \times \Theta \rightarrow \mathbb{R}$

The observation $X$ is a stochastic process, $X = \{X_i\}$, and takes values in the sample
space $\mathcal{X}$, with associated $\sigma$-algebra $\mathcal{B}$. We also assume that for each $\theta \in \Theta$, there is
a probability distribution on $\mathcal{B}$. The decision maker is at state $\lambda = \{i_1, \ldots, i_n\}$ if the
variables $X_{i_1}, X_{i_2}, \ldots, X_{i_n}$, have been observed. The variables $\{X_{i_1}, X_{i_2}, \ldots, X_{i_n}\}$
determine on $\mathcal{X}$ the smallest $\sigma$-field $\mathcal{B}_\lambda$ with respect to which they are measurable.
(We also note that a function \( f(x) \) is measurable with respect to \( \mathcal{B}_\lambda \) if \( f \) is a function of \( \{X_i\}_{i \in \Lambda} \) only).

At each step \( \lambda \) the available set of experiments is given by the correspondence \( \mathcal{E} : \Lambda \to \mathbb{N} \). Therefore the available choices at stage \( \lambda \) are given by \( \Delta(\lambda) = \mathcal{A} \cup \mathcal{E}(\lambda) \), that is, either a final decision or an available experiment may be chosen. We assume that at each stage \( \lambda \), \( \mathcal{E}(\lambda) \) is a finite subset of \( \mathbb{N} \). This implies that the set of variables \( \Lambda_n \) that can be observed in not more than \( n \) steps is finite. We define \( \Lambda \) and \( \Delta \) as \( \Lambda = \bigcup_n \Lambda_n \) and \( \Delta = \mathcal{A} \cup \Lambda \). If the decision maker chooses to continue experimentation, then an experiment \( j \in \mathcal{E}(\lambda) \) is chosen, and the realization of the random variable \( X_j \) is observed. The distribution of the random variable is given by \( f_{X_j}(x_j | \theta, \{x_i\}_{i \in \Lambda}) \).

A sequential decision rule is a function \( d : \Lambda \times \mathcal{X} \to \Delta \) such that for each \( \lambda \in \Lambda \), \( d(\lambda, \cdot) \) is \( \mathcal{B}_\lambda \) measurable function and chooses an element of \( \Delta(\lambda) \). The set of \( X \)'s for which state \( \lambda \) is reached and experimentation terminates at state \( \lambda \) is denoted by \( T(d, \lambda) \). If the decision maker reaches a decision \( a \in \mathcal{A} \) in \( n \) steps by observing \( \lambda \in \Lambda_n \), he pays the amount \( c(\theta, \lambda) \) and also incurs the Loss \( L(a, \theta) \). The expected risk \( R(\theta, d) \) of the decision rule \( d \), when the state of nature is \( \theta \) is given as

\[
R(\theta, d) = \sum_{n=0}^{\infty} \sum_{\lambda \in \Lambda_n} \left( \int_{T(d, \lambda)} c(\theta, \lambda) + L(d(\lambda, x), \theta) dP_x(\theta) \right)
\]

\[
= C(\theta, d) + R(\theta, d)
\]

(4.1)

(4.2)

\( C(\theta, d) \) is the expected observation cost for the decision rule \( d \) when the state of nature is \( \theta \). \( R(\theta, d) \) is the expected loss when the state is \( \theta \). The set of deterministic decision rules is denoted by \( \mathcal{D} \). Similarly the set of randomized decision rules is denoted by \( \mathcal{D}^* \).
4.3 Bayesian Decision Problem

In this section, the Bayes optimization problem for minimization of the average risk function $\mathcal{R}$ is defined. Then we review the Bayes solution [21, 22] for two special cases, which are frequently encountered in ATR systems.

**Problem-B:** Find the Bayes optimal decision rule $d^B(\pi)$:

$$d^B = \arg\min_{d \in D} \sum_{\theta \in \Theta} \pi(\theta)\mathcal{R}(\theta, d) \tag{4.3}$$

4.3.1 Case 1: I.I.D. Experiments

This is the case where at each stage the set available experiments contains a single experiment ($E(\lambda) = \{|\lambda| + 1\}$), and the underlying stochastic process is the product of independent random variables $X_i$ with distribution $f(x|\theta)$. For simplicity in presentation we will consider the binary hypothesis testing problem. The sequential binary hypothesis testing problem ($\mathcal{A} = \Theta = \{\theta_0, \theta_1\}$), extensively studied by Wald [60], is the foundation of sequential statistics. We review the i.d.d. case in this subsection.

The Bayes solution to this problem is conveniently expressed in the posteriori probability space, $\{\pi \in [0, 1]\}$. The Bayes solution can be obtained by first considering the truncated problem, where the stopping time is constrained to be less than a given bound $J$ [21]. This finite horizon optimization problem can be solved by backward induction. Then, the solution of the finite horizon problem is examined for $J \to \infty$. Let $U(\pi) = \min\{L(1, 0)\pi, L(0, 1)(1 - \pi)\}$ be the posterior expected loss, if an immediate decision is made and the posterior probability of $\theta_0$ is $\pi$. The solution is completely characterized by the function $V^{(\infty)}(\pi)$ which satisfies the fundamental
equation of dynamic programming:

\[ V^{(\infty)}(\pi) = \min\{U(\pi), c + E_{X_i}[V^{(\infty)}(\pi(x))]|\pi]\]  

where \(c\) denotes the uniform cost of experimentation.

Using \(V^{(\infty)}(\pi)\) we can define the function \(V^*(\pi)\), as \(V^*(\pi) = c + E_{X_i}[V^{(\infty)}(\pi(x))]|\pi]\) which represents the minimum Bayes risk over all rules that require at least one observation. Then the test is completely specified as:

**Step 1:** Initialize \(\pi = \text{Prob}(\theta_1)\)

**Step 2:** If \(V^*(\pi) > U(\pi)\) stop and goto Step 4

**Step 3:** Collect one more observation \(x_i\) and update \(\pi\) using Bayes rule; return to Step 2

**Step 4:** If \(L(1,0)\pi < L(0,1)(1 - \pi)\) choose \(\theta_0\) otherwise choose \(\theta_1\).

The function \(V^*(\pi)\) is continuous and concave on \([0,1]\) with \(V^*(0) = V^*(1) = c\) [21]. This observation shows that alternatively the test is completely specified with the pair of nonnegative thresholds, \((\pi_L < \pi_U)\).

**Step 1:** Initialize \(\pi = \text{Prob}(\theta_1)\)

**Step 2:** If \(\pi < \pi_L\) or \(\pi > \pi_U\) stop and goto Step 4

**Step 3:** Collect one more observation \(x_i\) and update \(\pi = \frac{\pi f(x_i|\theta_1)}{\pi f(x_i|\theta_1) + (1 - \pi) f(x_i|\theta_0)}\), return to Step 2

**Step 4:** If \(\pi < \pi_L\) choose \(\theta_0\) otherwise choose \(\theta_1\).

This construction of Bayes Rules can be extended to multiple hypothesis testing problems where \(\mathcal{A} = \Theta = \{\theta_1, \ldots, \theta_M\}\). For these problems, the posteriori probability space becomes the closed simplex \(S^{M-1} = \{\pi \in \mathbb{R}_+^M | \sum \pi_i = 1\}\) and the final decision rules are summarized by the convex sets \(A_m\) (such that the \(m\)'th unit vector in \(\mathbb{R}^M\)
is an element $A_m$) [60]. The decision maker continues to collect observations and update the posteriori probability $\pi$, until $\pi$ falls in one of the sets $A_m$.

For the binary hypothesis case, the Bayes rule $d^B$ for $\pi \in (\pi_L, \pi_U)$ can be expressed in an alternative form, the sequential probability ratio test (SPRT) [60], described by the parameters $(A, B)$.

**Step 1**: Initialize $S = 0$

**Step 2**: If $S < A$ or $S > B$ stop and goto Step 4

**Step 3**: Collect one more observation $x_t$, update $S = S + \log \frac{f(x_t|\theta_1)}{f(x_t|\theta_0)}$ and return to Step 2

**Step 4**: If $S < A$ choose $\theta_0$ otherwise choose $\theta_1$.

The parameters $(A, B)$ are related to Bayes boundaries $(\pi_L, \pi_U)$ through

$$A = \log \frac{\pi \pi_L}{(1 - \pi)(1 - \pi_L)} \quad B = \log \frac{\pi \pi_U}{(1 - \pi)(1 - \pi_U)}$$

### 4.3.2 Case 2: Feature Selection

For this case there exists a finite set $E = \{1, \ldots, N\}$ of experiments. Experiments cannot be repeated by the decision maker. Therefore $E(\lambda) = E \setminus \lambda$. The observation $X$ is simply a vector $X_1, \ldots, X_N$ of a finite number of statistics and therefore the decision maker is forced to choose a final action after $N$ steps, i.e., the problem is a truncated sequential problem. The Bayes solution to this problem can be found using dynamic programming techniques [22].

The main idea is again that at each stage $\lambda$, one should compare the posterior Bayes loss of making an immediate decision $U(\lambda)(x^\lambda)$ with the expected Bayes risk that will be obtained if more observations are taken using the optimal decision rule. If stopping gives a smaller posterior Bayes loss, then the procedure will be stopped.
Once the procedure has stopped the final action is chosen to minimize the Bayes Risk, using the finite sample decision rule. Specifically, the final action decision rule is not dependent on how the observations are taken, nor on which stopping rule has been adopted. With this fact we can derive the Bayes optimal decision strategy using dynamic programming principles [4]. The expected Bayesian risk at stage \( \lambda = \{1, \ldots, N\} \) is \( V_\lambda(x^\lambda) = U_\lambda(x^\lambda) \). For all the intermediate stages \( \lambda \) the expected Bayesian risk \( V_\lambda(x^\lambda) \) can be calculated recursively using:

\[
V_\lambda(x^\lambda) = \min\{U_\lambda(x^\lambda), \{E[V_{\lambda'}(x^\lambda)|x^\lambda]\}_{\lambda' \in \mathcal{E}(\lambda)}\}
\]

Generally, the Bayes optimal decision rule can only be calculated iteratively and the rule does not have a closed form suitable for storage [3]. This complexity of the optimal solution makes its use impractical except in a few special cases.

### 4.4 Sequential Decision Problems with Computational Cost Constraints

In this section we formulate two new statistical decision problems, where a certain decision performance is maximized, under performance constraints. Then, we prove results which relate the solution of these problems to the Bayes solutions discussed in the previous section.

For the first problem, we seek to minimize the expected decision loss, under an expected computational cost constraint. Both expectations are with respect to a prior distribution \( \pi \) on \( \Theta \).

**Problem** \( C_1 \): Find the optimal decision rule \( d^* \), defined by:

\[
\begin{align*}
\text{minimize} & \quad E_\theta R(\theta, d) \\
\text{subject to} & \quad E_\theta C(\theta, d) \leq C_0
\end{align*}
\]

(4.4)
The second decision problem may be considered a generalization of the Neyman-Pearson criterion to sequential detection problems.

**Problem C₂:** Find the optimal decision rule \( d^* \), defined by:

\[
d^* = \arg \min_d \quad R(\theta_0, d) \\
\text{such that} \quad R(\theta, d) \leq \alpha(\theta) \quad \forall \theta \in \Theta \setminus \{\theta_0\} \\
C(\theta, d) \leq \beta(\theta) \quad \forall \theta \in \Theta 
\]

(4.5)

The following theorem relates the solutions of Problem-C₁ to Bayes solutions the Problem-B.

**Theorem 1**

(i) \( d^* \in \mathcal{D}^* \) is an optimal solution for Problem-C₁ if and only if \( E_\theta[C(\theta, d^*)] = C_0 \) and there exists some \( \mu \geq 0 \) such that \( d^* \) is the Bayes Solution of Problem-B, which minimizes \( E_\theta[R(\theta, d) + \mu C(\theta, d)] \).

(ii) Let \( d^* \in \mathcal{D}^* \) be an optimal solution for Problem-C₁, then there exists \( \nu \in [0, 1] \) and two nonrandomized decision rules \( d^1, d^2 \in \mathcal{D} \) such that the optimal solution of Problem-C₁ is given by \( d^* = \nu d^1 + (1 - \nu) d^2 \). Specifically, if a nonrandomized decision rule \( d \in \mathcal{D} \) is a Bayes solution to Problem-B satisfying \( E_\theta[C(\theta, d)] = C_0 \) then it is an optimal solution to Problem-C₁.

**Proof:** To establish result (i), we follow a technique given in [21]. In [21] similar ideas were used to establish completeness of Bayes strategies for fixed observation models. The same technique is used in [57] to investigate the relation between Neyman-Pearson decision rules and Bayes rules for distributed detection systems.
First we define the operating characteristic set $Q$ of deterministic decision rules in $\mathcal{D}$ as

$$Q = \{(E_\theta R(\theta, d), E_\theta C(\theta, d)) \in \mathbb{R}^2 | d \in D\}$$

Similarly we define the operating characteristic set $Q^*$ of randomized decision rules in $\mathcal{D}^*$ as:

$$Q^* = \{(E_\theta R(\theta, d), E_\theta C(\theta, d)) \in \mathbb{R}^2 | d \in D^*\}$$

It is easy to see that $Q^*$ is the convex hull of $Q$. We observe that Problem-$C_1$ is equivalent to minimizing $r$ subject to $c = C_0$ and $(r, c) \in Q^*$. Suppose $r^* \in [0, 1]$ attains the minimum subject to these constraints. It is clear that $(r^*, C_0)$ lies at the lower boundary of $Q^*$, (otherwise $(r^* - \epsilon, C_0)$ would be in $Q^*$ as well, contradicting optimality of $r^*$). Then by the supporting hyperplane theorem, there exists a hyperplane parameterized by $(\mu_1, \mu_2)$ such that $(r^*, c)$ maximizes $\mu_1 r + \mu_2 c$ over $Q^*$. We also note that the lower boundary of $Q^*$ contains $[0.5, 0]$ and is a decreasing concave function, therefore $\mu_1 > 0$ and $\mu_2 \geq 0$. Defining $\mu = \mu_2/\mu_1$, we establish the “only if” part of the desired result.

Conversely, assume $d^* \in D^*$ is a Bayes solution minimizing $E_\theta[R(\theta, d) + \mu C(\theta, d)]$ for some $\mu \geq 0$ and $E_\theta[C(\theta, d^*)] = C_0$. Then clearly $E_\theta[R(\theta, d^*)] \leq r^*$. On the other hand, there exists a $d \in D^*$, which achieves the point $(r^*, C_0)$. Therefore the Bayes solution $d^*$ should satisfy $E_\theta[R(\theta, d^*)] = r^*$, otherwise we would have $E_\theta[R(\theta, d) + \mu C(\theta, d)] > E_\theta[R(\theta, d^*) + \mu C(\theta, d^*)]$ which contradicts the Bayes optimality of $d^*$.

To establish result (ii), recall that $Q^* \subset \mathbb{R}^2$ is the convex hull of $Q$. A theorem due to Carathéodory, states that any point in $Q^* \subset \mathbb{R}^2$ can be obtained by a convex combination of finitely many points in $Q$. Therefore the point $(E_\theta[R(\theta, d^*)], C_0) \in Q^*$ can be obtained as convex combination of finitely many points $\{q_n\}$ in $Q$. Moreover, the
point \((E_\theta[R(\theta, d^*)], C_0)\) is at the boundary of the set \(Q^*\) as argued in part (i); therefore there is supporting hyperplane at \((E_\theta[R(\theta, d^*)], C_0)\). We can conclude that all the points \(\{q_n\}\) lie on the same hyperplane. Thus, a convex combination of two points \((E_\theta[R(\theta, d^i)], E_\theta[C(\theta, d^i)]) \in Q, (i = 1, 2)\) is sufficient to represent \((E_\theta[R(\theta, d^*)], C_0)\).

A similar theorem relates the solutions of Problem-\(C_2\) to Bayes solutions of Problem-\(B\).

**Theorem 2**

(i) \(d^* \in D^*\) is an optimal solution for Problem-\(C_2\) if and only if the inequality conditions in the Problem are satisfied with equality and \(d^*\) is the Bayes Solution of Problem-\(B\), for properly chosen cost function \(c(\theta, \cdot)\) and priors \(\pi\).

(ii) There exists \(\nu \in S^{2M-1}\) and \(2M\) nonrandomized decision rules \(\{d^i \in D\}_{i=1}^{2M}\) such that the optimal solution of Problem-\(C_2\) is given by \(d^* = \sum_i \nu_i d^i\)

The proof is an immediate extension of the proof of Theorem 1.

### 4.4.1 Example 1: Feature Selection

The following example illustrates how Theorem 1 can be used to achieve solutions for the Problem \(C_1\).

To illustrate adaptive rules, we consider a simple two class problem \((M = 2)\) in which the suite of \(L = 3\) features is normally distributed with means \(m_i\) and covariances \(\Sigma_i, i = 1, 2\). The two actions \((K = 2)\) are to decide class 0 or decide class 1 given a single observation, \(x\). The class means and covariances for the vector of three features are given by

\[
\Sigma_0 = \begin{bmatrix}
1 & 0.7 & 0 \\
0.7 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix} \quad m_0 = \begin{bmatrix}
0 \\
0 \\
0
\end{bmatrix}
\]
\[ \Sigma_1 = \begin{bmatrix} 1 & -0.7 & 0 \\ -0.7 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad m_1 = \begin{bmatrix} 0.2 \\ 1 \\ 1 \end{bmatrix} \]

Finally, the relative costs of computing the three features are given by

\[ c_1 = 0.025 \quad c_2 = 0.1 \quad c_3 = 0.1 \]

Feature 1 is the least costly, while features 2 and 3 each are four times as costly as feature 1.

Figure 4.2 shows the Bayes Risk as a function of expected computational cost for three rules. Consider the non-adaptive rule marked (132). This rule consists of first computing feature 1, then next computing the feature least correlated with 1. Interestingly, this intuitively appealing rule is considerably worse than an alternative non-adaptive rule (123). The optimal rule is data adaptive and is shown in the solid line.

The solid curve in Figure 4.2 is actually precisely the lower boundary of the set \( Q^* \) introduced in the proof of Theorem 1. For constraints of \( C_0 \in (0, 0.08) \), the optimal rule belongs to the set of randomized decision rules \( D^* \). These randomized rules can be obtained by mixing the decision rule \( d^1 \) (where no samples are taken and a fixed decision is made) and \( d^2 \) (the decision rule with adaptive experiment selection which corresponds to \( C_0 = 0.08 \)).

In general, adaptive selection of features is advantageous, vis-à-vis non-adaptive rules, for multi-hypothesis testing (\( M > 2 \)). Likewise, adaptive feature selection offers gains when feature covariance structures are not identical under different hypotheses, and when features are not independent.
Figure 4.2: Comparison of Bayes Risk versus average computation cost for two non-adaptive decision rules and the Bayes-optimal adaptive decision rule.
4.5 Approximate Design (Single Experiment, Binary Hypothesis)

In Section 1.3.1 it was shown that the Bayes solution to the sequential binary hypothesis problem was a SPRT for some boundary values \((A, B)\). The classical statistical approach to hypothesis testing considers the power function \(\beta(\theta)\) and the operating characteristic curve \(\alpha(\theta)\) defined by:

\[
\beta(\theta) = P_\theta(\text{deciding } a_1) = P_\theta(L_N \geq B),
\]

\[
\alpha(\theta) = P_\theta(\text{deciding } a_0) = P_\theta(L_N \leq A).
\]

Specifically, the type 1 and type 2 probabilities \(\alpha_0 = \beta(\theta_0)\) and \(\alpha_1 = \alpha(\theta_1)\) are the variables of main interest. The computational constraints to these problems involve two additional variables: the expected computational costs under \(\theta_0\) and \(\theta_1\). In this case, these costs are given by the expected stopping times \(E_{\theta_0}N\) and \(E_{\theta_1}N\), respectively. Under either hypothesis, the stopping time is bounded with probability 1 [60]. This implies:

\[
1 = P_\theta(N < \infty) = P_\theta(L_N \leq A) + P_\theta(L_N \geq B) = \alpha(\theta) + \beta(\theta)
\]

Therefore, \(\alpha(\theta) = 1 - \beta(\theta)\), and it is sufficient to estimate the power function \(\beta(\theta)\) only. Following Wald [60], we define the \(Z_i\) as the log-likelihood of the \(i\)'th observation

\[
Z_i = \log \left[ \frac{f(X_i | \theta_1)}{f(X_i | \theta_0)} \right]
\]

Then the log-likelihood function of all the observations up to the \(n\)'th step is simply the sum \(S_n = \sum_{i=1}^{n} Z_i\). The fundamental identity of sequential analysis, established by Wald, involves consideration of the moment generating function of \(Z_i\), defined by

\[
M_\theta(t) = E_\theta \left[ e^{tZ_i} \right].
\]
Theorem 3 (Wald) If \( P_0(Z_i = 0) < 1 \) and \( P_0(|Z_i| < \infty) = 1 \), then for all \( t \) for which \( M_\theta(t) \) is finite \( E_\theta \left[ \exp(tS_N)M_\theta(t)^{-N} \right] = 1 \).

Using this theorem Wald derived the following approximate expressions for the power and average sample number (ASN) function.

\[
\alpha_0 = \beta(\theta_0) \approx \frac{1 - A}{B - A}, \quad (4.6)
\]

\[
\alpha_1 = 1 - \beta(\theta_1) \approx \frac{A(B - 1)}{B - A}, \quad (4.7)
\]

\[
E_{\theta_0}N \approx \frac{1}{\mu_{\theta_0}} [(1 - \alpha_0) \log(\alpha_1/(1 - \alpha_0)) + \alpha_0 \log((1 - \alpha_1)/\alpha_0)], \quad (4.8)
\]

\[
E_{\theta_1}N \approx \frac{1}{\mu_{\theta_1}} [\alpha_1 \log(\alpha_1/(1 - \alpha_0)) + (1 - \alpha_1) \log((1 - \alpha_1)/\alpha_0)]. \quad (4.9)
\]

Now we will use these expressions for direct design of approximate algorithms for binary hypothesis sequential decision problems with computational constraints.

**Approximate Design for Problem C_2**

Consider Problem \( C_2 \), which is a generalization of the Neyman-Pearson approach to sequential problems. For binary hypothesis testing problems, this problem is equivalent to deriving a receiver operating characteristics (ROC) curve of sequential tests satisfying given computational cost constraints under \( H_0 \) and \( H_1 \).

\[
\text{Max} \quad P_{Detection} = (1 - \alpha_1)
\]

such that

\[
P_{FalseAlarm} \leq M
\]

\[
E_{\theta_0}N \leq C_0
\]

\[
E_{\theta_1}N \leq C_1
\]

The \((P_D, P_{FA})\) pairs satisfying the computational cost constraints can be obtained from the Wald approximations (4.8, 4.9).

\[
S = \left\{ (P_D, P_{FA}) \left| \begin{array}{l}
C_0 \mu_{\theta_0} \geq (1 - P_{FA}) \log((1 - P_D)/(1 - P_{FA})) + P_D \log(P_D/P_{FA}) \text{and}
C_1 \mu_{\theta_1} \geq (1 - P_D) \log((1 - P_D)/(1 - P_{FA})) + P_D \log(P_D/P_{FA})
\end{array} \right. \right\}
\]

71
The boundary of $S$ gives the ROC curve for given constraint pair $(C_0, C_1)$. Also, for a given bound $M$ on the false alarm rate, the maximum achievable detection probability $P^*_D$ is given by:

$$P^*_D = \max \{ P_D \mid (P_D, M) \in S \}$$

The boundary values $(A, B)$ for the SPRT which achieves $(P^*_D, M)$, can be approximately derived from the Wald approximations as:

$$A \approx \frac{1 - P^*_D}{1 - M} \quad B \approx \frac{P_D}{M}$$

**Example: Constant Signal in Gaussian Noise**

$H_0 : \ X_k = N_k$

$H_1 : \ X_k = N_k + K$

Let $K = 1$ and the noise process $N_k$ be assumed to be iid normal. The cost constraints are given by $(C_0, C_1) = (c, c)$. Figure 4.3 gives the ROC curves for different values of $c$.

**Approximate Design for Problem $C_1$**

In this problem we seek the decision rule which minimizes the Bayes Risk among all decision rules satisfying an average cost constraint. In the present terminology the problem can be expressed as:

Min $\pi P_{False Alarm} + (1 - \pi)(1 - P_{Detection})$

such that $\pi E_{\theta_0} N + (1 - \pi)E_{\theta_1} N \leq C$.

The $(P_D, P_{FA})$ pairs satisfying the computational cost constraint can be obtained from the Wald approximations (4.8, 4.9).

$$S = \left\{ (P_D, P_{FA}) \mid \begin{array}{c} C \geq (\frac{1}{\mu_{\theta_0}}(1 - P_{FA}) + \frac{1 - \pi}{\mu_{\theta_1}}(1 - P_D)) \log((1 - P_D)/(1 - P_{FA})) \\ + (\frac{\pi}{\mu_{\theta_0}} P_{FA} + \frac{1 - \pi}{\mu_{\theta_1}} P_D) \log(P_D/P_{FA}) \end{array} \right\}$$
Figure 4.3: NP Solution with cost constraints
Again the boundary of $S$ gives the ROC curve for a given constraint $C$. For a given bound $M$ on the false alarm rate, the maximum achievable detection probability $P_D^*$ is given by:

$$P_D^* = \max\{P_D|(M, P_D) \in S\}$$

The optimum $P_D^*, P_{FA}^*$ pair is found by:

$$(P_D^*, P_{FA}^*) = \arg \max\{\pi P_{FA} + (1 - \pi)(1 - P_D))|(P_D, P_{FA}) \in S\}$$

Once the $(P_D^*, P_{FA}^*)$ pair is determined, then the boundary values $(A, B)$ for the SPRT which achieves $(P_D^*, P_{FA}^*)$ can be approximately derived from Wald approximations as described previously.

*Example: Constant Signal in Gaussian Noise (continued)*

The cost constraint is given by $C = 1$ and $\pi = 0.25$. Figure 4.4 gives the ROC curve, and the tangent point on the ROC curve gives the optimum $(P_D^*, P_{FA}^*)$. 
Figure 4.4: Bayes Solution with average cost constraint
CHAPTER 5

SERIAL DISTRIBUTED DETECTION WITH COMPUTATIONAL COST CONSTRAINTS

5.1 Introduction

Within an automatic target recognition and detection algorithm, a large volume of sensor data is accessed for performing a multiple hypothesis test with a constraint on computation. ATR systems typically compute several statistics, called features, from radar imagery. Computation of each feature imposes a specific computational load. The high data rates and real-time processing requirements for wide area surveillance have given rise to a non-adaptive, staged decision strategy as the de facto approach to ATR [16, 35, 40, 41]. Each stage of the ATR system computes discrimination statistics to reduce false alarms while maintaining high probability of detection. Screening false alarms reduces the data rate faced by the subsequent stages; computational complexity of the discrimination statistics is allowed to increase as the data rate drops. This type staged detection network is given in Figure 5.1.

An alternative staged detection network is obtained if binary local decisions are allowed. This generalized detection network is given in Figure 5.2.
Figure 5.1: Staged Detection Network-I

Figure 5.2: Staged Detection Network-II
We studied these staged detector networks in Chapter 4, under the assumption that the detectors can communicate soft local decisions to subsequent stages. However, a common practice in the design of an ATR decision network is not to propagate local soft decisions to later stages [16, 35]. In this “sieve approach,” each detector decides whether to pass its decision to the next detector or to terminate the process and make the final system-wide decision. Because of this assumption, the dynamic-programming solution given in Chapter 4 is not valid for these detection networks. For the study of this problem, distributed detection theory is more appropriate.

Recently, distributed detection systems received increased attention in the literature [8, 43, 2, 29, 56, 59]. In typical applications, information from multiple sensors is used to make a system wide decision. Distributed detection systems consider sensors with processing capability. Hard decisions from the local detectors are combined to yield the global inference. Distributed detection can be more attractive than centralized detection for various applications due to computational cost, complexity, survivability and communication bandwidth considerations. Applications of distributed detection networks include remote sensing, surveillance, communication and medical diagnosis. Serial distributed detection networks have been studied in [59, 43, 54] where sensors connected in series make local decisions based on their observation to maximize system-wide performance. Each sensor communicates its local hard decision to the next stage and the decision of the last sensor becomes the system-wide decision. We note that no sensor can break the process and make a system-wide decision. The performance of an infinite serial chain of sensors is analyzed in [43].

The contribution of this chapter can be seen as introduction of limited information transfer to sequential detection theory or introduction of computational complexity
and intermediate system-wide decision capability to serial distributed detection networks.

In this chapter, we study staged detection networks with limited communication between the detectors. In Section 5.2 we formulate the Bayes optimal staged detector, which minimizes a weighted sum of decision risk and expected computational cost. We derive optimal detection and associated stopping rules for the two network topologies given in Figures 5.1 and 5.2. Then, we show that the optimal design can be posed as an optimization problem in terms of the operating points on the ROC curve of individual detectors. In Section 5.3, we extend the results to multiple detectors and to different network topologies. The design rules are then used to obtain ROC curves for the staged detection network under computational cost constraints. In Section 5.4, we study the performance of infinite chain of identical detectors for the staged detection network given in Figure 5.2. We characterize the attainable pairs of decision risk and expected computational cost for this limiting case. Then we address the question: When does the probability of error (decision risk) go to zero as the expected computational cost goes to infinity? We obtain necessary and sufficient conditions for the convergence of the probability of error to zero.

5.2 Bayesian Problem Formulation

We first consider the special case binary hypothesis testing with a network of two detectors as shown in Figure 5.3.

The elements of the decision problem are

- States of Nature $\Omega = \{\omega_0, \omega_1\}$

- The observation vector $(y_1, y_2)$
Figure 5.3: Staged Detection Network—Two detectors

- Prior probability $\pi_0, \pi_1$

- Decision Cost $C : \{0, 1\}^2 \rightarrow \mathbb{R}$

  $C_{ij}$ is the cost of deciding $H_i$ under $H_j$.

- Computational Cost $c : \{1, 2\} \rightarrow \mathbb{R}$

The total Bayes Risk $R$ is the sum of the classical Bayes Decision Risk $R_{dec}$ and the expected computational cost given by $C$:

$$R = R_{dec} + C$$

$$= \sum_k \int_{y_1, y_2} \left\{ p(u_1 = 0|y_1)p(y_1|\omega_k)\pi_k[c_1 + C_{0k}] 
+ \sum_j p(u_1 = 1|y_1)p(u_2 = j|u_1 = 1, y_2)p(y_1, y_2|\omega_k)\pi_k[c_1 + c_2 + C_{jk}] \right\} dy_1 dy_2$$

Since in this model the first detector always is used, we can take without loss of generality $c_1 = 0$. Then we have:
\[ \mathcal{R} = \sum_k \int_{y_1, y_2} p(u_1 = 0 | y_1)p(y_1 | \omega_k)\pi_k[C_{0k}] + \sum_j p(u_1 = 1 | y_1)p(u_2 = j | u_1 = 1, y_2)p(y_1, y_2 | \omega_k)\pi_k[c_2 + C_{jk}]dy_1dy_2 \]

### 5.2.1 Solution to Bayes Problem

Now, we first study the necessary conditions for optimality, starting with the last detector (DM-2).

#### DM-2:

The observation of the first detector \(y_1\) does not affect the decision of the second detector. However, the decision rule applied in the first detector is implicit in the term \(p(u_1 | \omega_k)\).

\[ \mathcal{R} = \sum_{jk} \int_{y_2} p(u_2 = j | u_1 = 1, y_2)p(y_1, y_2 | \omega_k)p(u_1 = 1 | \omega_k)\pi_k[c_2 + C_{jk}]dy_2 \]  

(5.1)

\[ \mathcal{R} = \ldots + \sum_k \int_{y_2} p(u_2 = 0 | y_2)p(y_2 | \omega_k)p(u_1 = 1 | \omega_k)\pi_k C_{0k}dy_2 + \int_{y_2} (1 - p(u_2 = 0 | y_2))p(y_2 | \omega_k)p(u_1 = 1 | \omega_k)\pi_k C_{1k}dy_2 \]

\[ = \ldots + \sum_k \int_{y_2} p(y_2 | \omega_k)p(u_1 = 1 | \omega_k)\pi_k C_{1k} + \int_{y_2} p(u_2 = 0 | y_2)\sum_k \pi_k(C_{0k} - C_{1k})p(u_1 = 1 | \omega_k)p(y_2 | \omega_k)dy_2 \]

We observe that \(u_2 = 1\) is the optimal decision if:

\[ \sum_k \pi_k(C_{0k} - C_{1k})p(u_1 = 1 | \omega_k)p(y_2 | \omega_k) > 0 \]
This implies that the following threshold is optimal for the second decision maker (DM-2) given the decision rule of the first detector:

$$\Lambda^2(y_2) = \frac{p(y_2 | \omega_1)}{p(y_2 | \omega_0)} \begin{cases} \frac{\pi_0 (C_{10} - C_{00}) p(u_1 = 1 | \omega_0)}{\pi_1 (C_{11} - C_{01}) p(u_1 = 1 | \omega_1)} & \text{if } H_1 \\ \frac{\pi_0 C_{10} P_{FA}^1}{\pi_1 C_{01} P_B^1} & \text{if } H_0 \end{cases}$$

(5.2)

The expression in Eqn. (5.2) simplifies if we set $C_{00} = C_{11} = 0$, and define $p(u_1 = 1 | \omega_0) = P_{FA}^1$ and $p(u_1 = 1 | \omega_1) = P_B^1$

$$\Lambda^2(y_2) = \frac{p(y_2 | \omega_1)}{p(y_2 | \omega_0)} \begin{cases} \frac{\pi_0 C_{10} P_{FA}^1}{\pi_1 C_{01} P_B^1} & \text{if } H_1 \\ \frac{\pi_0 C_{10} P_{FA}^1}{\pi_1 C_{01} P_B^1} & \text{if } H_0 \end{cases}$$

(5.3)

DM-1:

Next, we derive person-by-person optimality rules for the first detector (DM-1).

$$\mathcal{R} = \sum_k \int_{y_1} p(u_1 = 0 | y_1) p(y_1 | \omega_k) \pi_k C_{0k}$$

$$+ \sum_j \int_{y_2} p(u_1 = 1 | y_1) p(u_2 = j | y_1 = 1, y_2) p(y_2 | \omega_k) \pi_k [c_2 + C_{jk}] dy_1 dy_2$$

We observe that $\{u_1 = 1\}$ is the optimal decision if:

$$\sum_k p(y_1 | \omega_k) C_{0k} \pi_k - \sum_j p(u_2 = j | \omega_k) p(y_1 | \omega_k) \pi_k [c_2 + C_{jk}] > 0$$

$$p(y_1 | \omega_0) \pi_0 C_{00} - \sum_j p(u_2 = j | \omega_0) \pi_0$$

$$+ p(y_1 | \omega_1) \pi_1 C_{01} - \sum_j p(u_2 \neq j | \omega_1) p(y_1 | \omega_1) \pi_1 [c_2 + c_{jk}] > 0$$

This implies that the following threshold is optimal for the first decision maker (DM-1) given the decision rule of the second detector:

$$\Lambda^1(y_1) = \frac{p(y_1 | \omega_1)}{p(y_1 | \omega_0)} \begin{cases} \frac{\pi_0 (C_{10} P_{FA}^2 + c_2)}{\pi_1 (C_{01} P_B^2 - c_2)} & \text{if } H_1 \\ \frac{\pi_0 (C_{10} P_{FA}^2 + c_2)}{\pi_1 (C_{01} P_B^2 - c_2)} & \text{if } H_0 \end{cases}$$

(5.4)

The expression in Eqn. (5.4) simplifies if we set $C_{00} = C_{11} = 0$, and define $p(u_2 = 1 | \omega_0) = P_{FA}^2$ and $p(u_2 = 1 | \omega_1) = P_B^1$

$$\Lambda^1(y_1) = \frac{p(y_1 | \omega_1)}{p(y_1 | \omega_0)} \begin{cases} \frac{\pi_0 (C_{10} P_{FA}^2 + c_2)}{\pi_1 (C_{01} P_B^2 - c_2)} & \text{if } H_1 \\ \frac{\pi_0 (C_{10} P_{FA}^2 + c_2)}{\pi_1 (C_{01} P_B^2 - c_2)} & \text{if } H_0 \end{cases}$$

(5.5)
5.2.2 Direct Solution to Bayes Problem using ROC Curve

In this section we show that the Bayes Solution obtained in Section 5.2.1 by direct manipulation of the Risk function can be obtained using dynamic programming principle and ROC-curves of the two-detectors characterized by \((P_{FA}^1, P_{BD}^1, P_{FA}^2, P_{BD}^2)\).

For simplicity we set \(C_{00} = C_{11} = 0\). The second detector updates its prior probability of each hypothesis, given its local observation and the fact that the first detector “passed” the decision job:

\[
\frac{P(\omega_0)}{P(\omega_1)} = \frac{\pi_0 p(u_1 = 1|\omega_0) p(y_2|\omega_0)}{\pi_1 p(u_1 = 1|\omega_1) p(y_2|\omega_1)}
\]

Then DM-2 makes its decision to minimize the posterior Bayes Risk:

\[
P(\omega_0) \overset{H_0}{\gtrless} \frac{1}{H_1} - P(\omega_0)
\]

The resulting threshold determines \((P_{FA}^2, P_{BD}^2)\). This decision rule is equivalent to the likelihood test in Eqn. 5.2.

The first detector updates its prior probability of each hypothesis given its local observation:

\[
\frac{P(\omega_0)}{P(\omega_1)} = \frac{\pi_0 p(y_1|\omega_0)}{\pi_1 p(y_1|\omega_1)}
\]

Then DM-1 makes its decision to minimize the posterior Bayes Risk.

\[
1 - P(\omega_0) \overset{H_1}{\gtrless} c(k) + P(\omega_0) P_{FA}^2 + (1 - P(\omega_0))(1 - P_{BD}^2)
\]

Where \(c(k)\) denotes the computation cost of the \(k^{th}\) decision unit. This decision rule is equivalent to the likelihood test in Eqn. 5.5.
5.3 Extension to Multiple Detectors and Binary Local Decisions

In Section 5.2 we have illustrated that the Bayes Solution can be obtained using the ROC curves of individual detectors. In this section we extend the results of the previous sections to multiple detectors and binary local decisions. We characterize the overall performance of the decision network in terms of the operating point of individual detectors. Then we apply these results for the solution of the Problem $C_2$ formulated in Chapter 4.

5.3.1 Multiple Detectors- Network I

We consider non-identical multiple detectors connected in series in the network topology given in Figure 5.1. We note that the local decisions are restricted to declaring $H_0$.

The ROC curve of the overall network $(P_{FA}(N), P_D(N))$ and the expected computational cost under each hypothesis $(EC_0(N), EC_1(N))$ can be calculated recursively using the following relations:

\[
\begin{align*}
P_D(k) &= P_D(k-1)P_D[k] \\
P_{FA}(k) &= P_{FA}(k-1)P_{FA}[k] \\
EC_0(k) &= EC_0(k-1) + P_{FA}(k-1)c(k) \\
EC_1(k) &= EC_1(k-1) + P_D(k-1)c(k)
\end{align*}
\]

where $P_D[k]$ ans $P_{FA}[k]$ characterize the threshold used by the $k$'th detector:

\[
\begin{align*}
P_D[k] &= Pr(\Lambda^k(y_k) > t(k)|H_1) \\
P_{FA}[k] &= Pr(\Lambda^k(y_k) > t(k)|H_0)
\end{align*}
\]
We note that $P_D[k]$ can be obtained from $P_{FA}[k]$ and the ROC curve of the $k'$th detector. Therefore, only $N$ independent variables determine the performance of the serial network characterized by four performance figures $(P_D(N), P_{FA}(N), EC_0(N), EC_1(N))$

$$P_D(N) = \prod_{k=1}^{N} P_D[k]$$

$$P_{FA}(k) = \prod_{k=1}^{N} P_{FA}[k]$$

$$EC_0(N) = \sum_{k=1}^{N-1} P_{FA}(k)c(k + 1)$$

$$EC_1(N) = \sum_{k=1}^{N-1} P_D(k)c(k + 1)$$

**Example:**

We consider two identical detectors whose ROC curves are given in Figure 5.4. An example of a DM with such a ROC curve is a statistician who observes the output of a binary symmetric channel [45], with error probability $\alpha < 1$.

For the serial network of two identical detectors, Figure 5.5 shows all the ROC curves obtainable with the serial network is given by choosing various operating points for the two detectors.

Next, we consider Problem $C_2$ for this serial network:

Maximize $P_D(N)$

such that $P_{FA} \leq \alpha_0$

$EC_0(N) \leq \beta_0$

$EC_1(N) \leq \beta_1$
Figure 5.4: ROC curve for binary symmetric channel
Figure 5.5: ROC curves for Network I
Alternatively, the optimization problem can be expressed in terms of the operating points of the individual detectors \((P_{FA}[k], P_D[k])\):

Maximize \( \prod_{k=1}^{N} P_D[k] \)

such that \( \prod_{k=1}^{N} P_{FA}[k] \leq \alpha_0 \)
\( \sum_{k=1}^{N-1} P_{FA}(k)c(k+1) \leq \beta_0 \)
\( \sum_{k=1}^{N-1} P_D(k)c(k+1) \leq \beta_1 \)

**Example:**

For the serial network of two identical detectors, we choose \( \beta_0 = \beta_1 = 0.8c(2) \). The set of \( P_{FA}, P_D \) pairs achievable under this computation cost are given in Figure 5.6. We observe that the ROC curve does not reach to \( P_D = 1 \). This is due to the fact that local decision at the first detector is restricted to \( H_0 \). We will see that once binary local decisions are permitted the resulting ROC curve will range from \((0,0)\) to \((1,1)\).

### 5.3.2 Multiple Detectors- Network II

We consider non-identical multiple detectors connected in series in the network topology given in Figure 5.2. We note that the binary local decisions are permitted in this network structure. Therefore, two threshold values should be selected for each detector \((t_1(k), t_2(k))\), except the last one. The following summarize the decision rule

88
Figure 5.6: $C_2$ Problem-Network I
of the $k$th DM ($k < N$):

$$\Lambda(y_k) < t_1(k) \quad \text{decide } H_0$$

$$t_1(k) < \Lambda(y_k) < t_2(k) \quad \text{pass}$$

$$\Lambda(y_k) > t_2(k) \quad \text{decide } H_1$$

Alternatively, the two thresholds can be characterized with corresponding operating points on the DM’s ROC curve.

$$P_D^1[k] = Pr(\Lambda^k(y_k) > t_1(k)|H_1)$$

$$P_{FA}^1[k] = Pr(\Lambda^k(y_k) > t_1(k)|H_0)$$

$$P_D^2[k] = Pr(\Lambda^k(y_k) > t_2(k)|H_1)$$

$$P_{FA}^2[k] = Pr(\Lambda^k(y_k) > t_2(k)|H_0)$$

The ROC curve of the overall network ($P_{FA}(N), P_D(N)$) and the expected computational cost under each hypothesis ($EC_0(N), EC_1(N)$) can again be calculated recursively using the following relations:

For $k = 1, \ldots, N - 1$

$$P_D^1(k) = P_D^1(k-1) + (P_D^2(k-1) - P_D^1(k-1))P_D^1[k]$$

$$P_D^2(k) = P_D^2(k-1) - (P_D^2(k-1) - P_D^1(k-1))(1 - P_D^2[k])$$

$$P_{FA}^1(k) = P_{FA}^1(k-1) + (P_{FA}^2(k-1) - P_{FA}^1(k-1))P_{FA}^1[k]$$

$$P_{FA}^2(k) = P_{FA}^2(k-1) - (P_{FA}^2(k-1) - P_{FA}^1(k-1))P_{FA}^2[k]$$

$$EC_0(k) = EC_0(k-1) + (P_{FA}^2(k-1) - P_{FA}^1(k-1))c(k)$$

$$EC_1(k) = EC_1(k-1) + (P_D^2(k-1) - P_D^1(k-1))c(k)$$
and the operating point of the last detector is characterized by a single threshold or equivalently by a single point \((P_{FA}[N], P_{D}[N])\) on ROC. Then the overall performance figures \(\{P_{D}(N), P_{FA}(N), EC_{0}(N), EC_{1}(N)\}\) of the network is given:

\[
P_{D}(N) = P^1_{D}(N-1) + (P^2_{D}(N-1) - P^1_{D}(N-1))P_{D}[N]
\]

\[
P_{FA}(N) = P^1_{FA}(N-1) + (P^2_{FA}(N-1) - P^1_{FA}(N-1))P_{FA}[N]
\]

\[
EC_{0}(N) = EC_{0}(N-1) + (P^2_{FA}(N-1) - P^1_{FA}(N-1))c(N)
\]

\[
EC_{1}(N) = EC_{1}(N-1) + (P^2_{D}(N-1) - P^1_{D}(N-1))c(N)
\]

Again \(P^j_{D}[k]\) can be obtained from \(P^j_{FA}[k]\) and the ROC curve of the \(k'\)th detector.

Therefore, only \(2N-1\) independent variables determine the performance of the serial network characterized by four performance figures \(\{P_{D}(N), P_{FA}(N), EC_{0}(N), EC_{1}(N)\}\)

\[
P_{D}(N) = \sum_{k=1}^{N} P^1_{D}[k] \prod_{l=1}^{k-1} (P^2_{D}[l] - P^1_{D}[l])
\]

\[
P_{FA}(N) = \sum_{k=1}^{N} P^1_{FA}[k] \prod_{l=1}^{k-1} (P^2_{FA}[l] - P^1_{FA}[l])
\]

\[
EC_{0}(N) = \sum_{k=1}^{N-1} (P^2_{FA}[k] - P^1_{FA}[k])c(k+1)
\]

\[
EC_{1}(N) = \sum_{k=1}^{N-1} (P^2_{D}[k] - P^1_{D}[k])c(k+1)
\]

**Example:**

For the serial network of two identical detectors, Figure 5.7 shows all the ROC curves obtainable by choice of operating points for the two detectors.
Figure 5.7: ROC curves for Network II
Next, we consider Problem $C_2$ for this serial network:

Maximize \[ P_D(N) \]

such that \[ P_{FA} \leq \alpha_0 \]
\[ EC_0(N) \leq \beta_0 \]
\[ EC_1(N) \leq \beta_1 \]

Again, the optimization problem can be expressed in terms of the operating points of the individual detectors \((P_{FA}[k], P_D[k])\):

Maximize \[ \sum_{k=1}^{N} P^1_B[k] \prod_{l=1}^{k} (P^2_D[l] - P^1_B[l]) \]

such that \[ \sum_{k=1}^{N} P^1_{FA}[k] \prod_{l=1}^{k} (P^2_{FA}[l] - P^1_{FA}[l]) \leq \alpha_0 \]
\[ \sum_{k=1}^{N-1} (P^2_{FA}[k] - P^1_{FA}[k])c(k + 1) \leq \beta_0 \]
\[ \sum_{k=1}^{N-1} (P^2_D[k] - P^1_D[k])c(k + 1) \leq \beta_1 \]

Example:

For the serial network of two identical detectors, we choose $\beta_0 = \beta_1 = 0.4c(2)$. The set of $P_{FA}, P_D$ pairs achievable under this computation cost are given in Figure 5.8. We observe that the ROC curve ranges from (0,0) to (1,1).

5.4 Infinite Serial Detector with Identical Sensors

In this section, we consider the limiting case of the staged detection network given in Figure 5.2. Specifically, we analyze the performance of an infinite chain of identical detectors connected in series, performing identical tests. One possible interpretation
Figure 5.8: $C_2$ Problem-Network II
for this network is a detector without memory performing sequential tests to determine the prevailing state of nature. We are interested in the performance of the infinite network characterized by the attainable (Decision Risk, Expected Computational Cost) pairs.

We are interested in stationary strategies, namely a pair of operating points \((P_{FA}^1, P_D^1)\) and \((P_{FA}^2, P_D^2)\) (or equivalently two thresholds) that satisfy the recursive relation

\[
P_D(\infty) = P_D^1 + (P_D^2 - P_D^1)P_D(\infty)
\]
\[
P_{FA}(\infty) = P_{FA}^1 + (P_{FA}^2 - P_{FA}^1)P_{FA}(\infty)
\]

The relations assure that the performance of the detector does not change if an additional identical detector is placed in front.

Assume that the ROC curve of the detector is parametrized with a concave function \(f(\cdot)\). For simplicity in notation let \(a := P_{FA}^1\) and \(b := P_{FA}^2\), with \(b > a\). Then we have \(f(a) := P_D^1\) and \(f(b) = P_D^2\). Then the performance of the infinite serial network is characterized by four performance figures \(\{P_D(\infty), P_{FA}(\infty), EC_0(\infty), EC_1(\infty)\}\).

\[
P_D(\infty) = \frac{f(a)}{1 - f(b) + f(a)}
\]
\[
P_{FA}(\infty) = \frac{a}{1 - b + a}
\]
\[
EC_0(\infty) = \frac{c}{1 - b + a}
\]
\[
EC_1(\infty) = \frac{c}{1 - f(b) + f(a)}
\]

95
Figure 5.9: ROC curve for Infinite Serial Network

We observe that any point $a, b$ satisfies the recursive relation given above. In Figure 5.9 all possible operating points are calculated for the ROC curve of the DM observing the outputs of the BSC channel.

We are interested in the behavior of $P_{\text{error}} = \pi_0 P_{A}(\infty) + (1 - \pi_0)(1 - P_{D}(\infty))$ as the expected computation cost $E C = \pi_0 E C_0(\infty) + (1 - \pi_0)(E C_1(\infty))$ goes to infinity. We observe that for this example the probability of error $P_{\text{error}}$ does not go to zero for all prior probabilities $\pi \in (0, 1)$, as the expected computation cost goes to infinity. The following theorem completely characterizes the limiting behavior of $P_{\text{error}}$. The statement of the theorem is similar to [43], where the limiting behavior for a serial distributed detection network without intermediate decisions was established.
Theorem: Let $P_{error}$ be the probability of error for the infinite serial network of identical sensors employing identical tests. Then as $EC \to \infty$, the network will achieve zero probability of error for all prior probabilities, if and only if either the initial slope of the ROC curve of the individual DM is infinite or its final slope is zero.

Proof: Suppose that the initial slope of the DM’s ROC curve is $m_0 < \infty$ and the final slope is $m_1 > 0$. Then from the convexity of the ROC curve, the performance of the serial network is worse than a serial network consisting of sensors with BSC ROC curves for $\alpha = \max \{1/m_0, m_1\}$. Therefore, we just need to show that the probability of error for the serial network of BSC sensors is bounded away from zero, for some prior probability. For simplicity let $\pi_0 = \pi_1 = 0.5$ and assume $f(x) \in (0, 1)$ for $x \in (0, 1)$, otherwise the proof is immediate. Then the probability of error for this network is given as:

$$P_{error} = P_{FA}(\infty) + (1 - P_D(\infty)) = \frac{a}{1 - b + a} + \frac{1 - f(b)}{1 - f(b) + f(a)}$$

We also have for $\pi_0 = \pi_1 = 0.5$:

$$EC = EC_0(\infty) + EC_1(\infty) = \frac{c}{1 - b + a} + \frac{c}{1 - f(b) + f(a)}$$

There are three cases in which $EC$ can approach infinity.

(i) $a = 0, b \to 1^-$

(ii) $b = 1, a \to 0^+$

(iii) $a \to 0^+, b \to 1^-$
For cases (i)-(ii) $P_{\text{error}} = 1$, so there is nothing to prove for case (iii). We define $c := (1 - b) \to 0^+$. 

$$P_{\text{error}} = P_{FA}(\infty) + (1 - P_D(\infty))$$

$$= \frac{1}{1 + \frac{c}{a}} + \frac{\alpha^2}{\alpha^2 + \frac{a}{c}}$$

We observe that for this case $P_{\text{error}}$ depends on the ratio of $\frac{a}{c}$ and is bounded below by $\frac{2a}{a+1}$. Therefore, since $\alpha$ is nonzero by hypothesis we have shown that if the initial slope of the DM's ROC curve is $m_0 < \infty$ and the final slope is $m_1 > 0$, then the $P_{\text{error}}$ does not go to zero for all prior probabilities as $EC$ approaches infinity.

Conversely, assume that either the initial slope of the ROC curve of the individual DM is infinite or its final slope is zero. Without loss of generality we assume that the final slope is zero. We must show that for any given $\epsilon$, there exists a $(a,b)$ such that both $P_{FA}(\infty)$ and $(1 - P_D(\infty))$ are less than $\epsilon$. Alternatively

$$P_{FA}(\infty) = \frac{1}{1 + \frac{1-b}{a}} < \epsilon$$

$$(1 - P_D(\infty)) = \frac{1}{1 + \frac{f(a)}{1-f(b)}} < \epsilon$$

If we define $\delta = \frac{1-\epsilon}{\epsilon} > 1$, the above conditions are equivalent to:

$$\frac{1-b}{a} > \delta$$

$$\frac{f(a)}{1-f(b)} > \delta$$

We have, from concavity of ROC curve $f(a) \geq a$. Also, and as $b \to 1$, $\frac{1-f(b)}{1-b} \to 0$, from the slope condition. So, given $\delta$ (equivalently $\epsilon$) we can choose $b$ small enough, such that $(1 - b) \geq 4\delta^2(1 - f(b))$, then choose $a$ such that $\frac{1-b}{a} = 2\delta > \delta$. Then we have

$$4\delta^2(1 - f(b)) \leq (1 - b) = 2\delta a \leq 2\delta f(a)$$

98
implying $\frac{f(a)}{1-f(b)} \geq 2\delta > \delta$. This completes the proof.

For the ROC curve given by the function $f(x) = 2x - x^2$, the computed example in 5.10 illustrates this convergence.

Papastavrou and Athans [43] have shown that the condition of the Theorem is equivalent to requiring that the range of the likelihood function computed at the sensor is bounded away from infinity.
APPENDIX A

ML ESTIMATE

In this appendix we derive the maximum likelihood estimates in Eqns. (2.20-2.21). We simplify the ML estimation by prewhitening both the scattering measurement, \( \tilde{S} \), and basis vectors, \( B \). The whitening changes only the covariance and mean of the SIRV clutter [47]. Let \( \Sigma = \phi \Lambda \phi^H \) be the eigendecomposition of the clutter covariance matrix. Then the prewhitening is given by

\[
\begin{align*}
\tilde{S} &= \Lambda^{-1/2} \phi^H \tilde{S} \\
\tilde{b}_1(\psi) &= \Lambda^{-1/2} \phi^H b_1(\psi) \\
&= \tilde{B} R_B(\psi)b(\tau_i, \nu_i, \gamma_i)
\end{align*}
\]

First we compute the ML-estimate of \( A_c \) given \( \psi \)

\[
\hat{A}_{ci}(\psi) = \arg\min_{A_c}(\tilde{S} - A_c \tilde{b}_1(\psi)) H (\tilde{S} - A_c \tilde{b}_1(\psi))
\]

(A.1)

Applying the projection theorem, we minimize the quadratic form to obtain the ML estimate of \( A_c \)

\[
\hat{A}_{ci}(\psi) = \frac{\tilde{b}_1(\psi)^H \tilde{S}}{\|\tilde{b}_1(\psi)\|^2}
\]

(A.2)

Next we consider the ML estimate, \( \hat{\psi} \), of the orientation angle. For trihedral and helical scatterers, \( \hat{\psi} \) does not affect \( b \) and is therefore not required in the GLRT. For \( b \)
a function of $\psi$, we substitute Eqn. (A.2) in Eqn. (A.3) to learn that the ML estimate of $\psi$ is given by

$$\arg \max_{\psi} \frac{|\tilde{b}_i(\psi)^H \tilde{S}|}{||\tilde{b}_i(\psi)||}$$  \hspace{1cm} (A.4)

Eqn. (A.4) provides a simple interpretation for the ML estimate of $\psi$: *i.e.*, maximize the projection of the received vector $\tilde{S}$ on the test vector $\tilde{b}_i(\psi)$. The projection is defined by a weighted norm using $\Sigma^{-1}$. Finding the value of $\psi$ which maximizes this projection can be posed as either constrained maximization problem in $\mathbb{R}^2$ or as a one dimensional search in $\psi$. 

101
APPENDIX B

BAYES RULE

In Section 3.3 we approximated the optimal Bayes rule by replacing likelihood ratios by maximum likelihood ratios. In this section we derive a complete Bayes approach for Gaussian clutter for specific distributions of \( (A, \rho, \psi) \). Specifically, we will assume Rayleigh distribution for \( A \), uniform distribution for \( \rho \) and an arbitrary distribution \( f_\psi(\psi) \) for \( \psi \). If \( A \) is Rayleigh distributed and \( \rho \) has a uniform distribution then the complex amplitude \( A_c = Ae^{j\rho} \) will have a complex Gaussian distribution:

\[
f_{A_c}(A_c) = \frac{1}{\pi\sigma^2} \exp\{-|A_c|^2/\sigma^2\}
\]

The conditional probability of a scattering matrix under the hypothesis \( H_i \) is given by:

\[
P_{\mathcal{S}|H_i}(s|H_i) = E_{A,\rho,\psi} \left[ \frac{1}{\pi^3|\Sigma|} \exp\left\{ -\left( \mathcal{S} - Ae^{j\rho}b_i(\psi) \right)^H\Sigma^{-1}\left( \mathcal{S} - Ae^{j\rho}b_i(\psi) \right) \right\} \right]
\]

\[
= E_\psi \left[ \frac{1}{\pi^3|\Sigma|} \int dA_c \frac{1}{\pi\sigma^2} \exp\{-|A_c|^2/\sigma^2\} \exp\left\{ -\left( \mathcal{S} - Ae^{j\rho}b_i(\psi) \right)^H\Sigma^{-1}\left( \mathcal{S} - Ae^{j\rho}b_i(\psi) \right) \right\} \right]
\]

\[
= E_\psi \left[ \frac{1}{\pi^3|\Sigma|} \exp\{-\|\mathcal{S}\|^2/2\} \frac{1}{1 + \sigma^2\|b_k(\psi)\|^2} \exp\left\{ \frac{\sigma^2\|\mathcal{S}\|^2}{1 + \sigma^2\|b_k(\psi)\|^2} \right\} \right] f_\psi(\psi) d\psi
\]

\[
= \frac{1}{\pi^3|\Sigma|} \int_0^{2\pi} \frac{1}{1 + \sigma^2\|b_k(\psi)\|^2} \exp\left\{ \frac{\sigma^2\|\mathcal{S}\|^2}{1 + \sigma^2\|b_k(\psi)\|^2} \right\} f_\psi(\psi) d\psi
\]

102
The last integral is calculated using a numerical integration technique. We observe that for these prior distributions for $A, \rho$ and $\psi$, the Bayes rule in Eqn. (2.22) takes the form

$$i = \arg \max_{i=1,\ldots,M} \left[ \int_0^{2\pi} \frac{1}{1 + \sigma^2 ||b_i(\psi)||^2_2} \exp \left\{ \frac{\sigma^2 ||\hat{S}^H \Sigma^{-1} b_i(\psi)||^2}{1 + \sigma^2 ||b_i(\psi)||^2_2} \right\} f_\psi(\psi) d\psi \right]$$  \hspace{1cm} (B.5)

For $\Sigma = \sigma_n^2 I$ the expression in Eqn.(B.5) simplifies to

$$i = \arg \max_{i=1,\ldots,M} \left[ \int_0^{2\pi} \exp \left\{ \frac{(\sigma^2 / \sigma_n^2)||\hat{S}^H b_i(\psi)||^2}{\sigma_n^2 + \sigma^2} \right\} f_\psi(\psi) d\psi \right]$$ \hspace{1cm} (B.6)
BIBLIOGRAPHY


