New Approaches to Ground Moving Target Indicator Radar

Dissertation

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

By

Michael Richard Riedl, B.S., M.S.

Graduate Program in Electrical and Computer Engineering

The Ohio State University

2016

Dissertation Committee:

Dr. Lee C. Potter, Advisor
Dr. Emre Ertin
Dr. Philip Schniter
Abstract

Increasing the resolution of radar imaging and ground moving target indicator (GMTI) systems puts a stress on both hardware and processing limitations. Hardware must be able to handle the transfer of the large amounts of data generated. Additionally, the processing must be robust to any heterogeneity of the data that is introduced by collecting returns from large swaths. This dissertation presents system architectures and knowledge-aided processing techniques to combat the large data rates and data heterogeneity.

Ground moving target indicator radar techniques for airborne platforms require spatial and Doppler signal diversity for separating the returns of moving targets from the returns of ground clutter. The traditional use of multiple receive antennas for jointly imaging a scene and detecting moving objects is prohibited by the system bottleneck at the data down-link. We present a frequency division multiple access, multiple-transmit single-receive radar architecture, with associated waveform design and data processing procedure. The proposed approach is demonstrated to jointly provide imaging and GMTI modalities while maintaining the data rate to that of a single antenna imaging system.

Heterogeneity of the radar backscatter data degrades detection performance by biasing statistical parameters estimated from the data. A GMTI processing technique, known as space-time adaptive processing (STAP), requires estimation of the
space-time covariance of the clutter for use in a generalized likelihood ratio test. Consequently, the performance of STAP is related to the quality of the estimated clutter covariance matrix; however, in practice it is common for the data to be limited, contaminated, and heterogeneous. In this dissertation, we introduce and evaluate two estimators for the clutter covariance and a purely Bayesian detection scheme. A Bayesian model is postulated for the angle/Doppler scene to incorporate approximate prior knowledge of the terrain height and the platform kinematics. Posterior probabilities computed using the model are then used to either estimate a covariance matrix or directly report posterior probabilities of the presence of a target. The approach is a novel means for incorporating operational knowledge into GMTI processing and admits low-complexity algorithmic implementation via recent advances in Bayesian message passing algorithms. In the second covariance estimator, a regularized shrinkage approach is proposed, whereby prior knowledge is expressed through an elastic net regularization penalty on a minimum expected squared error estimation cost. The regularized shrinkage estimator is shown to coincide with a minimax robust covariance estimator and offers simplicity in modeling and computation that may facilitate use by practitioners. In the third approach, the Bayesian model is augmented to jointly estimate calibration parameters for unknown antenna phases and detect moving targets.

The performances of the proposed estimators and detectors are evaluated using the KASSPER I dataset. We conclude that the proposed approaches extend the state-of-the-art to provide reliable detection performance when the training data is limited to a number of range bins less than the rank of the true covariance matrix. Further, when presented no training data, the Bayesian approach is shown to maintain
performance using only the data under test. Finally, the purely Bayesian detection
approach, when combined with antenna calibration, is observed to provide enhanced
resolution, allowing reliable detection of multiple targets within a single range bin
not achievable with traditional STAP.
This is dedicated to my parents Rich and Sue, and sister Amanda who inspire me to grow everyday.
Acknowledgments

I would first like to thank my advisor Dr. Lee Potter; without him none of this would have been possible. I would also like to thank my fellow graduate students, especially Adam Rich and Jeremy Vila who always made themselves available as an additional resource whenever the need arose. The advice and support of the members of AFRL and the Center for Surveillance Research (CSR), especially Edmund Zelnio, Mike Minardi, Jason Parker, and Fred Garber, also played a large part in helping me complete this work.

I also need to acknowledge the support of my family and friends. My friends, especially my girlfriend Theresa have been a constant source of support, even if from a distance. Additionally, I could not have done this were it not for the support of my parents and sister. My father is the one who inspired me to be an engineer and always makes sure I have what I need to succeed. My mother is the one who listens to my problems and always encourages me to forge ahead even when it seems impossible. Finally, my sister is the one who always makes me laugh even when I’m in the worst mood. I am eternally grateful for these people in my life and can never thank them enough for their unwavering support.
Vita

July 7, 1989 .......................... Born - Milwaukee, Wisconsin, USA

May 2011 ............................. Bachelor of Electrical Engineering and
B.S. in Computer Engineering, University of Dayton

December 2015 ........................ M.S. in Electrical and Computer Engineering, The Ohio State University

September 2011-present .................. Graduate Research Assistant, The Ohio State University.

Publications

M. Riedl, L. Potter, C. Bryant, and E. Ertin, “Joint Synthetic Aperture Radar and
Space-time Adaptive Processing on a Single Receive Channel,” *IEEE Transactions

M. Riedl, L. Potter, and E. Ertin, “Augmenting Synthetic Aperture Radar with

SPIE*, (Baltimore, MD), Apr. 2015.


Fields of Study

Major Field: Electrical and Computer Engineering
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>v</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>vi</td>
</tr>
<tr>
<td>Vita</td>
<td>vii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xiii</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Problem Statement</td>
<td>3</td>
</tr>
<tr>
<td>1.2 Proposed Approaches</td>
<td>4</td>
</tr>
<tr>
<td>1.2.1 Joint SAR/STAP Radar</td>
<td>4</td>
</tr>
<tr>
<td>1.2.2 Structured and Knowledge-Aided STAP</td>
<td>5</td>
</tr>
<tr>
<td>1.3 Summary of Contributions</td>
<td>6</td>
</tr>
<tr>
<td>1.3.1 Augmenting SAR with STAP</td>
<td>6</td>
</tr>
<tr>
<td>1.3.2 Incorporating Approximate Prior Knowledge</td>
<td>7</td>
</tr>
<tr>
<td>1.3.3 Target Rejection</td>
<td>8</td>
</tr>
<tr>
<td>1.3.4 Resolving Closely Spaced Targets</td>
<td>8</td>
</tr>
<tr>
<td>1.3.5 Sensitivity of Knowledge-Aided Approaches and Joint Detection/Calibration</td>
<td>8</td>
</tr>
<tr>
<td>1.4 Notation</td>
<td>8</td>
</tr>
<tr>
<td>2. Space-Time Adaptive Processing Overview</td>
<td>10</td>
</tr>
<tr>
<td>2.1 Geometry and Signal Model</td>
<td>10</td>
</tr>
<tr>
<td>2.2 Clutter Model</td>
<td>13</td>
</tr>
<tr>
<td>2.3</td>
<td>Detectors</td>
</tr>
<tr>
<td>-----</td>
<td>-----------</td>
</tr>
<tr>
<td>2.3.1</td>
<td>Target Detectors</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Non-Homogeneity Detectors</td>
</tr>
<tr>
<td>2.4</td>
<td>Existing Estimators</td>
</tr>
<tr>
<td>2.4.1</td>
<td>Constrained Maximum Likelihood Estimators</td>
</tr>
<tr>
<td>2.4.2</td>
<td>Knowledge-Aided Estimators</td>
</tr>
<tr>
<td>2.4.3</td>
<td>Shrinkage Estimators</td>
</tr>
<tr>
<td>2.4.4</td>
<td>Direct Data Domain Estimators</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3.</th>
<th>Joint SAR and STAP Radar</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>25</td>
</tr>
<tr>
<td>3.2</td>
<td>Scheduled FDMA Waveforms</td>
<td>27</td>
</tr>
<tr>
<td>3.3</td>
<td>FDMA MISO Extensions for SAR/STAP</td>
<td>30</td>
</tr>
<tr>
<td>3.3.1</td>
<td>FDMA MISO SAR Model</td>
<td>30</td>
</tr>
<tr>
<td>3.3.2</td>
<td>FDMA MISO STAP Model</td>
<td>31</td>
</tr>
<tr>
<td>3.4</td>
<td>FDMA MISO Clutter Covariance</td>
<td>34</td>
</tr>
<tr>
<td>3.5</td>
<td>Numerical Results</td>
<td>38</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Metrics</td>
<td>38</td>
</tr>
<tr>
<td>3.5.2</td>
<td>SAR Simulation</td>
<td>38</td>
</tr>
<tr>
<td>3.5.3</td>
<td>STAP Simulation</td>
<td>40</td>
</tr>
<tr>
<td>3.5.4</td>
<td>STAP Emulation: Gotcha GMTI Dataset</td>
<td>43</td>
</tr>
<tr>
<td>3.6</td>
<td>Conclusions</td>
<td>48</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4.</th>
<th>Knowledge-Aided Bayesian Space-Time Adaptive Processing</th>
<th>53</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>53</td>
</tr>
<tr>
<td>4.2</td>
<td>Bayesian Model</td>
<td>54</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Linear Mixing Model</td>
<td>54</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Calculation of Clutter Prior</td>
<td>58</td>
</tr>
<tr>
<td>4.3</td>
<td>Detection Approaches</td>
<td>60</td>
</tr>
<tr>
<td>4.3.1</td>
<td>MMSE Covariance Estimation</td>
<td>62</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Bayesian Hypothesis Testing</td>
<td>63</td>
</tr>
<tr>
<td>4.4</td>
<td>Numerical Results</td>
<td>63</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Metrics</td>
<td>65</td>
</tr>
<tr>
<td>4.4.2</td>
<td>KASSPER I Dataset</td>
<td>66</td>
</tr>
<tr>
<td>4.5</td>
<td>Conclusions</td>
<td>71</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5.</th>
<th>Multi-Model Shrinkage for Knowledge-Aided Space-Time Adaptive Processing</th>
<th>72</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>72</td>
</tr>
</tbody>
</table>
5.2 Shrinkage Estimator Overview ........................................ 73
5.2.1 Single Model Shrinkage Estimators ............................... 74
5.2.2 Multi-Model Shrinkage Estimators ............................... 74
5.3 Multi-Model Shrinkage with Regularization ......................... 75
5.3.1 Non-Negative Complex Valued MMS .............................. 75
5.3.2 Regularization ..................................................... 77
5.3.3 Regularization: Robust Shrinkage ............................... 78
5.3.4 Extensions: Outlier Excision ...................................... 80
5.4 Multi-Model Shrinkage with Elastic Net Regularization for STAP . 82
5.5 Numerical Results ..................................................... 84
5.5.1 Metrics ............................................................ 84
5.5.2 Simulation ........................................................ 86
5.5.3 KASSPER I Dataset ................................................ 87
5.6 Conclusions .......................................................... 90

6. Effects of Calibration on Knowledge-Aided STAP Approaches ........ 92
6.1 Introduction .......................................................... 92
6.2 Calibration on Clutter ............................................... 92
6.2.1 Channel Pair Detrending ......................................... 93
6.2.2 Parametric Bilinear Model ....................................... 94
6.3 Numerical Results ..................................................... 96
6.3.1 Metrics ............................................................ 96
6.3.2 KASSPER I Dataset ................................................ 97
6.4 Conclusions .......................................................... 102

7. Conclusions ........................................................... 104
7.1 Summary .............................................................. 104
7.2 Future Work .......................................................... 105
7.2.1 Optimize Bayesian Model for Target Detection ............... 105
7.2.2 Finite Sample Error Bounds ...................................... 107
7.2.3 MAP Covariance Estimation ..................................... 108
7.2.4 Fast Computation of MMS-EN: Parallel Architecture .......... 109
7.2.5 Evaluation of Techniques with Measured Data ............... 110
7.2.6 Empirical Bayes SAR GMTI .................................... 110
7.2.7 Joint SAR GMTI/Refocusing .................................... 110

Bibliography .............................................................. 111
# Appendices

## A. Constrained Maximum Likelihood Estimators

A.1 Maximum Likelihood Estimate (Sample Covariance) .................................................. 119  
A.2 Convex Covariance Constraints ..................................................................................... 121  
A.3 Thermal Noise and a Low-Rank Constraint .................................................................... 122  
A.4 Clutter Subspace Constraint .......................................................................................... 124  
A.5 Convex Approach for Subspace Knowledge .................................................................. 126  
A.6 Prior Knowledge via a Nominal Covariance ................................................................. 127  
A.7 Sparse Methods ............................................................................................................ 128

## B. Message Passing Derivations

B.1 Messages for GAMP Iteration ......................................................................................... 131  
B.2 Messages Leaving after GAMP Iteration ....................................................................... 132  
B.3 Clutter Indicator Posterior Calculation ......................................................................... 134

## C. Expectation Maximization Update Derivations

C.1 Overview ....................................................................................................................... 136  
C.2 Moving Target Sparsity Parameter Update .................................................................... 139  
C.3 Clutter Weight Parameter Update ................................................................................ 140

## D. Bayesian Detector Derivations

D.1 Whitened Matched Filter ............................................................................................... 143  
D.1.1 GAMP Approximate Posterior Calculation ............................................................... 143  
D.1.2 MMSE Clutter Estimate ............................................................................................ 145  
D.2 Bayesian Hypothesis Testing ......................................................................................... 147

## E. Multi-Model Shrinkage Derivations

E.1 Convex Combination ....................................................................................................... 150  
E.2 Non-Negative Combination .......................................................................................... 152
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Parameters for the KASSPER I dataset.</td>
<td>64</td>
</tr>
<tr>
<td>A.1 Constrained ML estimators for covariance</td>
<td>130</td>
</tr>
</tbody>
</table>
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Three dimensional STAP scene geometry</td>
<td>11</td>
</tr>
<tr>
<td>2.2 Overhead STAP scene geometry</td>
<td>12</td>
</tr>
<tr>
<td>3.1 Waveform scheduling comparison</td>
<td>29</td>
</tr>
<tr>
<td>3.2 FDMA MISO SAR point spread function: range</td>
<td>39</td>
</tr>
<tr>
<td>3.3 FDMA MISO SAR point spread function: cross-range</td>
<td>40</td>
</tr>
<tr>
<td>3.4 Clutter eigenspectrum for FDMA MISO STAP simulation</td>
<td>42</td>
</tr>
<tr>
<td>3.5 Adapted pattern for FDMA MISO STAP simulation</td>
<td>43</td>
</tr>
<tr>
<td>3.6 Comparison of SINR losses for FDMA MISO STAP simulation</td>
<td>44</td>
</tr>
<tr>
<td>3.7 Phase ramp across Doppler: antenna spacing estimate</td>
<td>46</td>
</tr>
<tr>
<td>3.8 Antenna spacing estimates for entire dataset</td>
<td>47</td>
</tr>
<tr>
<td>3.9 Gain imbalance estimates</td>
<td>48</td>
</tr>
<tr>
<td>3.10 SINR comparisons for FDMA MISO STAP</td>
<td>49</td>
</tr>
<tr>
<td>3.11 Sample SAR image for 2006 Gotcha dataset</td>
<td>50</td>
</tr>
<tr>
<td>3.12 Example SIMO STAP results</td>
<td>51</td>
</tr>
<tr>
<td>3.13 Example FDMA MISO STAP emulation results</td>
<td>52</td>
</tr>
</tbody>
</table>
B.3 Posterior calculation ................................. 135
Chapter 1: Introduction

Radar uses electromagnetic waves to measure distances from the transmitter to an object and back to the receiver. This range measurement can also be done coherently which allows for the measurement of the rate of change. The measurements of both range and rate of change in range can be used for a variety of applications. An application of interest is ground moving target indicator (GMTI).

Ground moving target indicator is a radar application that attempts to detect moving targets on the ground. A difficulty with GMTI is separating returns of potential targets from the ground reflections, often termed clutter. In most scenarios, it is impossible to detect targets amongst the strong returns of clutter using simple range processing. Instead, GMTI approaches leverage both the range and rate of change in range measurements, sometimes across more than one antenna. One type of GMTI processing is video synthetic aperture radar (SAR). Video SAR uses the radar measurements to form SAR images at different apertures, effectively forming frames of a video. Moving targets can then be detected by taking the difference of consecutive video frames [58]. A problem with using video SAR for GMTI is the loss of target tracks when the Doppler-shifted target energy is collocated with clutter having strong returns [54].
Space-time adaptive processing (STAP) is a GMTI processing technique that exploits measurements across antennas (space) and pulses (time). The spatial measurements allow returns to be separated by angle of arrival while the temporal measurements allow returns to be separated by Doppler. The advantage of the additional degrees of freedom provided by the multiple antennas is the ability to separate moving targets from clutter in the normalized spatial (or angle)/Doppler frequency domain [48,86]. The approach in STAP is to linearly combine the returns from \( M \) pulses and \( N \) antennas in a coherent processing interval (CPI) to maximize the signal-to-interference-pulse-noise ratio (SINR) using a weight vector of complex values, \( w \). It can be shown that the weight vector, \( w \), that maximizes SINR is also the solution to the hypothesis test of a known signal with unknown amplitude in a noise environment, specifically:

\[
H_0 : \quad n \\
H_1 : \quad \alpha a(\vartheta, \bar{\omega}) + n \\
n \sim \mathcal{CN}(n; 0, R)
\]

(1.1)

(1.2)

where \( n \), the interference, is composed of clutter, jamming, and thermal noise, and \( \mathcal{CN}(x; \mu, R) \) is the complex multivariate Gaussian distribution with mean vector \( \mu \) and covariance matrix \( R \) [14,65]. The weight vector that is the solution to both maximizing SINR and the Gauss-Gauss hypothesis test is given as a whitened matched filter,

\[
w = \kappa R^{-1} a,
\]

(1.3)

where \( \kappa \) is a nonzero scaling factor [14,32,48,65,86]. In practice the covariance matrix, \( R \), is not known a priori. Therefore, the adaptivity of STAP comes from estimating
the covariance or inverse covariance matrix, $R^{-1}$. This work addresses some of the challenges associated with GMTI radar.

1.1 Problem Statement

We consider several challenges confronting GMTI radar. First, can high-resolution SAR be augmented to provide a GMTI capability? In this challenge, a primary system constraint is the limited data rate of the communication down-link from an airborne platform. Second, target detection requires estimation of the clutter statistics, typically summarized by the covariance matrix. Estimation of the clutter covariance, and hence detection performance, is hampered by limited or heterogeneous training data. Can the state-of-the-art detection performance be improved by incorporating approximate knowledge of the scene elevation map and platform kinematics? Third, the presence of targets in the data creates bias in covariance estimates, and detection performance degrades. Can the effect of targets in the training data be ameliorated in covariance estimation? A fourth challenge is to resolve two moving targets that are closely spaced in range, angle, and Doppler. Can nonlinear processing provide a capability not provided by the conventional whitened matched filter? Finally, a fifth challenge is presented by array calibration, which is modeled as an unknown gain and phase at each array element. What sensitivities do knowledge-aided approaches have to calibration error, and can the calibration be accomplished jointly with the detection?
1.2 Proposed Approaches

In this dissertation we propose new processing approaches to address the five challenges in radar GMTI outlined above. In chapter 3, we address the challenge of augmenting a traditional single antenna SAR system with additional degrees of freedom required for STAP while maintaining the data rate of the system to that of the original single antenna system. In chapters 4 and 5 we address the challenge of estimating the interference covariance matrix when the available training data is limited, heterogeneous, and contaminated by incorporating prior knowledge about the scene. In chapter 6 we explore the sensitivity of knowledge-aided approaches to calibration error and explore the joint calibration and detection approach. Our proposed approaches to these challenges are summarized in the following subsections.

1.2.1 Joint SAR/STAP Radar

We propose using a multiple transmit single receive, or multiple input single output (MISO), system architecture for augmenting a SAR imaging system to provide enhanced GMTI capabilities. The MISO system architecture can maintain the data rate of the original single antenna SAR system. This approach can ameliorate the bottleneck of the down link data rate when augmenting a high resolution SAR system with STAP capabilities. We use a frequency division multiple access (FDMA) approach to separate the spatial channels at the receiver; the frequency channels span the entire bandwidth of the original single antenna SAR imaging system. This approach maintains the data rate and resolution of the original single antenna SAR imaging system while adding the degrees of freedom required for GMTI processing approaches. This comes at a cost of increased computational complexity and a lower
range resolution (as a factor of the number of spatial channels) in the GMTI modality relative to the SAR modality.

1.2.2 Structured and Knowledge-Aided STAP

In order to reduce the detection performance losses from non-stationary and limited training data, we propose two approaches that exploit the structure of clutter in the normalized angle/Doppler scene. This structure can be known approximately from platform kinematics (platform velocity, crab angle, and antenna spacings) and an approximate digital elevation map (DEM). This approximate structure can then be combined with the sparsity of moving targets in the scene to estimate the interference covariance matrix. Incorporating this approximate prior knowledge into the covariance estimator can ameliorate the problem of non-stationary and limited training data.

The first approach we consider for incorporating prior knowledge into covariance estimation for STAP utilizes a Bayesian linear mixing model. The Bayesian model provides a natural mechanism for incorporating prior knowledge and allows for covariance estimation without the need for secondary training data. Additionally, targets can be detected directly from the approximate posterior probabilities computed using the Bayesian model. Finally, low complexity computation with the Bayesian model is enabled by recent algorithm developments for fast inference on linear mixing models; the lower complexity is achieved by decoupling the mixing problem using Gaussian assumptions and Taylor series approximations. The algorithms can also be accelerated by exploiting fast linear operators such as the fast Fourier transform (FFT).
The second approach we consider for incorporating prior knowledge into covariance estimation for STAP introduces a regularized multi-model shrinkage estimator. The estimator admits a non-negative linear combination of the unbiased sample covariance and biased shrinkage models. The scaling factors that minimize the expected squared estimation error can be approximated by solving a quadratic program. We also present a model extension for outlier rejection amongst the training data. The proposed approach has complexity dominated by construction of the matrix in the quadratic program, which requires $O(K^2M^2N^2)$ multiplications, where $K$, $M$, and $N$ are the number of shrinkage models, the number of pulses in the CPI, and the number of antennas in the CPI, respectively.

Finally, we explore the effects of calibration on the two proposed knowledge-aided approaches. Existing calibration on clutter techniques are utilized for comparison to the uncalibrated approach, and extensions to the Bayesian model are given for posing the joint estimation and calibration as a bilinear problem.

1.3 Summary of Contributions

Here we summarize the contributions of our work grouped by the five challenges presented in the problem statement.

1.3.1 Augmenting SAR with STAP

- Demonstrated a system architecture and transmit waveform design that permits simultaneous operation of both SAR and STAP using a single receive channel.

- Presented FDMA MISO processing procedures for the combined SAR/STAP mode of operation.
• Shown both analytically and numerically that the FDMA MISO rotating waveform design maintains low-rank structure of clutter that permits target detection.

• Demonstrated the ability of FDMA MISO rotating waveform scheme to perform both high resolution imaging and space-time adaptive processing.
  – Demonstrated with simulation.
  – Demonstrated with emulation on collected data set.

1.3.2 Incorporating Approximate Prior Knowledge

• Proposed and experimentally verified a Bayesian framework that provides knowledge-aided covariance estimation without secondary data and without clairvoyant knowledge of clutter support in the normalized angle/Doppler scene.

• Demonstrated that incorporating approximate prior knowledge in a Bayesian framework can outperform other single cell approaches as well as constrained maximum likelihood covariance estimators using scores of range bins for training data.

• Proposed and experimentally demonstrated a regularized multi-model shrinkage estimator for covariance estimation in STAP. The approach was shown to outperform the existing state of the art when the amount of available training data is less than the rank of the true covariance.
1.3.3 Target Rejection

- Demonstrated that processing using the proposed Bayesian model can effectively reject targets in the data when estimating a covariance matrix.

- Demonstrated that the target rejection extension for the regularized multi-model shrinkage estimator can mitigate the detection performance loss caused by targets in the training data.

1.3.4 Resolving Closely Spaced Targets

- Demonstrated that the non-linear processing using the Bayesian model can resolve closely spaced movers that are not detected with the traditional whitened matched filter approach.

1.3.5 Sensitivity of Knowledge-Aided Approaches and Joint Detection/Calibration

- Demonstrated that calibration improves the performance of the proposed knowledge-aided approaches.

- Proposed and experimentally demonstrated a joint calibration/detection approach in the Bayesian framework. Processing with the bilinear model can improve the detection performance of the Bayesian hypothesis detection scheme.

1.4 Notation

In this dissertation the following notation is used: Boldface lowercase letters such as $a$ represent vectors while boldface uppercase letters such as $A$ represent matrices. The operators $\otimes$ and $\odot$ are the Kronecker and Hadamard products, respectively.
The superscript operators \( (\cdot)^T \), \( (\cdot)^H \), and \( (\cdot)^* \) are the transpose, conjugate transpose, and conjugation operators, respectively. The operation \( \text{tr}(\cdot) \) represents the trace of a matrix, \( \| \cdot \|^2 \) is the squared Frobenius norm, \( E\{ \cdot \} \) is the expectation of a random variable, and \( \text{Var}(\cdot) \) is the variance of a random variable. Finally, the notation \( A_{[j,k]} \) denotes the element in the \( j^{\text{th}} \) row and \( k^{\text{th}} \) column of matrix \( A \).
Chapter 2: Space-Time Adaptive Processing Overview

In this chapter, we present a brief overview of the space-time adaptive processing geometry, signal model, clutter model, and detectors.

2.1 Geometry and Signal Model

The rest of the chapters in this dissertation assume the following geometry and signal model. We consider an airborne radar platform that emits $M$ pulses using a constant pulse repetition frequency with a uniform linear array of $N$ antennas and a velocity $v_p$. The scene geometry is shown in Fig. 2.1 and Fig. 2.2. These figures have been adapted from the figures in [86]. In Fig. 2.1 and Fig. 2.2, the antennas are represented by blue asterisks, and the velocity vector is represented by a green arrow. The figures also show the angles used to represent a spatial steering vector to a patch, in red, in the scene. The spatial steering vector is characterized by an azimuth angle $\phi$ and depression angle $\theta$.

Given the scene geometry, we can then define the signal model. Using a uniform linear array (ULA) and a constant pulse repetition frequency (PRF), scatterers in the scene can be described by spatial and Doppler frequencies. A scatterer’s Doppler frequency is given by

$$f_t = \frac{2v_t}{\lambda_c} \quad (2.1)$$
Figure 2.1: Three dimensional scene geometry for STAP. The blue asterisks represent antenna locations, the green arrow represents the platform velocity vector, and the spot in the scene is defined by azimuth angle $\phi$, depression angle $\theta$, and slant range $r$. The variable $h$ is the height of the terrain below the platform.

where $v_t$ is the radial velocity of the scatterer and $\lambda_c$ is the carrier frequency of the narrowband system. The normalized Doppler frequency is then defined as

$$\bar{\omega}_t = \frac{f_t}{f_r}$$

(2.2)

where $f_r$ is the PRF. The delay from the scatterer to the $n^{th}$ element of the ULA consists of two components

$$\tau_n = \tau_t + \tilde{\tau}_n$$

(2.3)

where $\tau_t$ is the round trip delay ($\tau_t = 2r/c$), $\tilde{\tau}_n$ is the relative delay from the reference element to the $n^{th}$ element, and $c$ is the speed of light. Then, given that $\theta_t$, $\phi_t$, and $d$ are the depression angle to the scatterer, the azimuth angle to the scatterer, and the
element spacing of the array, respectively, the relative delay is
\[ \tilde{\tau}_n = -n \frac{d}{c} \cos \theta_t \sin \phi_t. \] (2.4)

It is then helpful to define the scatterer’s spatial frequency as
\[ \vartheta_t = \frac{d}{\lambda_c} \cos \theta_t \sin \phi_t. \] (2.5)

It is important to note that a scatterer’s response at a given range can be completely described by its normalized Doppler frequency and spatial frequency.

Since we can characterize a scatterer’s response by its Doppler and spatial frequencies, the signal from the scatterer will have a specific structure as it propagates across the array of elements. Given that the scatterer has a spatial frequency \( \vartheta \), the
response across the \( N \) elements of the array will take the form

\[
a_s(\vartheta) = \left[ 1; e^{j2\pi \vartheta}; \ldots; e^{j(N-1)2\pi \vartheta} \right],
\]

(2.6)

which is referred to as a \( N \times 1 \) spatial steering vector. Likewise for \( M \) pulses at a single element of the array, the response of a scatterer with a normalized Doppler frequency of \( \bar{\omega} \) will take the form

\[
a_t(\bar{\omega}) = \left[ 1; e^{j2\pi \bar{\omega}}; \ldots; e^{j(M-1)2\pi \bar{\omega}} \right],
\]

(2.7)

which is referred to as a \( M \times 1 \) temporal steering vector. The response for a single scatterer for this \( M \) pulse and \( N \) antenna CPI can then be characterized by the \( MN \times 1 \) space-time steering vector defined as

\[
a(\vartheta, \bar{\omega}) = a_t(\bar{\omega}) \otimes a_s(\vartheta).
\]

(2.8)

For the remainder of the dissertation we will only consider the joint space-time steering vector and may omit its notational dependency on \( \bar{\omega} \) and \( \vartheta \).

### 2.2 Clutter Model

As mentioned previously, the interference is modeled as complex-valued colored Gaussian noise with covariance \( R \). This covariance then admits the following decomposition,

\[
R = R_c + R_j + R_n
\]

(2.9)

where \( R_c \) is the clutter component, \( R_j \) is the jamming component, and \( R_n \) is the thermal noise component [86]. In this work we do not consider the jamming component and model the thermal noise component as

\[
R_n = \sigma_n^2 I,
\]

(2.10)
where $\sigma_n^2$ is the noise variance and $I$ is the identity matrix.

The clutter component is then modeled as

$$R_c = \sum_{k=1}^{N_c} \alpha_k^2 a_k a_k^H,$$

(2.11)

where $\alpha_k^2$ is the power of the $k$th clutter patch, $a_k$ is the steering vector to the $k$th clutter patch, and $N_c$ is the number of clutter patches with $N_c > MN$ [86]. The steering vectors for the clutter have a special relationship between normalized Doppler and spatial frequency where

$$\bar{\omega} = \beta \vartheta \cos \phi_{\text{crab}} \pm \sqrt{\vartheta^2 \beta^2 (\cos^2 \phi_{\text{crab}} - 1) + \frac{r^2 - h^2}{r^2} \left( \frac{2v_p}{f_r \lambda_c} \right)^2 \sin^2 \phi_{\text{crab}}},$$

(2.12)

$\beta = \frac{2v_p}{f_r d}$, $r$ is the slant range of the clutter patch, $h$ is the height of the clutter patch below the platform, and $\phi_{\text{crab}}$ is the crab angle (misalignment of antenna array and platform velocity vector). It can be seen from Eq. 2.12 that in the absence of crabbing ($\phi_{\text{crab}} = 0$), the normalized Doppler frequency is then $\bar{\omega} = \beta \vartheta$. However, when a crab angle is present, the normalized Doppler frequency depends on the range and height of the clutter, and the $\pm$ in Eq. 2.12 represents the front- and back-lobe clutter components [86].

In the special case where the crab angle is zero and $\beta$ is equal to an integer, it can be shown that the clutter component, $R_c$, has a low rank. The rank, $r_c$, can be calculated as

$$r_c \approx \lfloor N + \beta (M - 1) \rfloor,$$

(2.13)

where $\lfloor x \rfloor$ denotes largest integer not greater than $x$. This is known as the Brennan rule [86]. This low-rank structure can be proven analytically for integer $\beta$ and has also been demonstrated to be approximate with numerical simulations for non-integer


values as well. It is important to note that the Brennan rule does not hold when either a crab angle or internal clutter motion effects are present.

2.3 Detectors

In this section we will discuss both target detectors and non-homogeneity detectors. The target detectors are formulated to provide constant false alarm rate (CFAR) properties to aide in the choice of a detection threshold. Non-homogeneity detectors are employed to prune the secondary training to ameliorate the performance losses incurred by covariance estimation with non-homogeneous (non-stationary) training data [46, 47].

2.3.1 Target Detectors

In this subsection we present and derive target detectors for when the interference covariance is both known and unknown.

Known Interference Covariance

Given that we wish to detect targets using linear filtering of the form

\[ \gamma = w^H y. \]  

(2.14)

for data vector \( y \), we must find the optimal weight vector \( w_* \). We assume the data vector admits the decomposition

\[ y = a + n, \]  

(2.15)
where \( a \) in a known signal of interest and \( n \) is zero mean colored Gaussian noise. The optimal weight vector \( w \) can then be found by maximizing the signal-to-interference-plus-noise ratio (SINR),

\[
\text{SINR} = \frac{|w^H a|^2}{E \{ |w^H n|^2 \}}. \tag{2.16}
\]

Since \( w \) is not random, the denominator of Eq. 2.16 is found to be

\[
E \{ |w^H n|^2 \} = w^H E \{ nn^H \} w = w^H R w, \tag{2.17}
\]

where \( R \) is the covariance of the zero mean colored Gaussian noise. The optimal weight vector can then be found as

\[
w_\ast = \arg \max_{s.t. \ w^H w = 1} \frac{|w^H a|^2}{w^H R w}. \tag{2.18}
\]

Noting that \( R^{1/2} R^{-1/2} = I \) and applying the Cauchy-Schwarz inequality we can find that

\[
\frac{|w^H a|^2}{w^H R w} \leq \frac{(w^H R w) (a^H R^{-1} a)}{w^H R w} = (a^H R^{-1} a) \tag{2.19}
\]

therefore,

\[
\frac{|w^H a|^2}{w^H R w} \leq (a^H R^{-1} a). \tag{2.20}
\]

Equality is then achieved when

\[
w = \kappa R^{-1} a, \tag{2.21}
\]

where \( \kappa = 1/(a^H R^{-1} a) \) to satisfy the unity constraint.

The target detection problem can also be formulated as the following binary hypothesis test

\[
H_0 : \ y \sim \mathcal{CN}(0, R) \tag{2.22}
\]

\[
H_1 : \ y \sim \mathcal{CN}(\alpha a, R)
\]
where again \( a \) is a known signal. The unknown amplitude, \( \alpha \in \mathbb{C} \), is considered deterministic; the Generalized Likelihood Ratio Test (GLRT) is obtained by replacing \( \alpha \), conditioned on \( H_1 \), with its maximum likelihood estimate,

\[
\max_{\alpha} \frac{f_1(y|\alpha, R)}{f_0(y|R)} \xrightarrow{H_1 \geq H_0} \tau, \tag{2.23}
\]

where \( f_1 \) is the probability density function for observation vector \( y \) under hypothesis \( H_1 \) and \( f_0 \) is the probability density function under hypothesis \( H_0 \). The resulting threshold test is given by

\[
\left| a^H \hat{R}^{-1} y \right|^2 \xrightarrow{H_1 \geq H_0} \tau \tag{2.24}
\]

and provides a constant false alarm rate (CFAR) [21, 68]. We can now see that the two approaches in Eq. 2.18 and Eq. 2.23 for finding the weight vector are equivalent.

**Unknown Interference Covariance**

More generally, the covariance is likewise unknown, and a limited amount of auxiliary training data ("secondary data"), \( y_1, \ldots, y_K \), is available for estimating the covariance. A so-called "homogeneous" model is when the test data ("primary data"), \( y \), and the training data, \( \{y_1, y_2, \ldots, y_K\} \), share the same covariance matrix. Kelly’s detector [36] is derived as the GLRT for unknown \( \{\alpha, R\} \); as such, the maximum likelihood estimates of the parameters are found, under each hypothesis, and substituted into the likelihood ratio. The sample covariance matrix is the maximum likelihood estimate, and the optimization for \( \alpha \) is facilitated by working with the \((K + 1)\)st root of the likelihood ratio. The resulting GLRT takes the form

\[
\frac{\left| a^H \hat{R}^{-1} y \right|^2}{\left(1 + \frac{1}{K} y^H \hat{R}^{-1} y \right) \left(a^H \hat{R}^{-1} a\right)} \xrightarrow{H_1 \geq H_0} \tau \tag{2.25}
\]
where the sample covariance on secondary data is used for $\hat{R}$. Kelly’s test statistic converges, in probability, to the GLRT for known covariance as the number of secondary data vectors becomes very large.

A so-called “partially homogeneous” case is when the test data (“primary data”), $\mathbf{y}$, and the training data, $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_K\}$, share the same covariance matrix, except for an unknown scaling, $\lambda$:

$$
H_0: \quad \mathbf{y} \sim \mathcal{CN}(0, \lambda \mathbf{R}) \\
H_1: \quad \mathbf{y} \sim \mathcal{CN}(\alpha \mathbf{a}, \lambda \mathbf{R})
$$

(2.26)

For unknown amplitude and scaling factor, the composite test again admits a simple GLRT. Solving for the MLE of $\alpha$ under hypothesis $H_1$ and the MLEs of the scaling, $\lambda$, under each hypothesis (assuming known covariance, $\mathbf{R}$), one finds

$$
\alpha = \frac{\mathbf{a}^H \mathbf{R}^{-1} \mathbf{y}}{\mathbf{a}^H \mathbf{R}^{-1} \mathbf{a}}, \quad \lambda_0 = \frac{1}{K} (\mathbf{y}^H \mathbf{R}^{-1} \mathbf{y}), \quad \lambda_1 = \frac{1}{K} \left( \mathbf{y}^H \mathbf{R}^{-1} \mathbf{y} - \frac{|\mathbf{a}^H \mathbf{R}^{-1} \mathbf{y}|^2}{\mathbf{a}^H \mathbf{R}^{-1} \mathbf{a}} \right).
$$

(2.28)

The resulting generalized log-likelihood ratio test takes the form

$$
\frac{|\mathbf{a}^H \mathbf{R}^{-1} \mathbf{y}|^2}{(\mathbf{y}^H \mathbf{R}^{-1} \mathbf{y}) (\mathbf{a}^H \mathbf{R}^{-1} \mathbf{a})} \xrightarrow{\mathcal{H}_1} \frac{\mathcal{H}_1}{\mathcal{H}_0} \tau.
$$

(2.29)

This composite binary hypothesis test, for unknown $\{\alpha, \lambda\}$ and known $\mathbf{R}$, has several nice properties, such as uniformly most powerful scale invariant hypothesis test [37].

But, in reality the covariance is unknown. The true GLRT requires the class-conditional ML estimates of the clutter covariance given both primary and secondary data; the likelihood maximization must be performed jointly over the parameters $\{\alpha, \lambda, \mathbf{R}\}$,

$$
\frac{\max_{\{\alpha, \lambda, \mathbf{R}\}} p_1(\mathbf{y}, \mathbf{y}_1, \ldots, \mathbf{y}_K | \alpha, \lambda, \mathbf{R})}{\max_{\{\lambda, \mathbf{R}\}} p_1(\mathbf{y}, \mathbf{y}_1, \ldots, \mathbf{y}_K | \lambda, \mathbf{R})} \xrightarrow{\mathcal{H}_1} \frac{\mathcal{H}_1}{\mathcal{H}_0} \tau.
$$

(2.30)
A common practice is to derive the GLRT assuming a known covariance, $R$, then substitute for the unknown matrix using the ML estimate based on secondary data only. This is a simple procedure and very closely approximates the true GLRT for $K$ large. The resulting quasi-GLRTs in Eq. 2.24 and Eq. 2.29, respectively, have been presented in the context of radar STAP as the “adaptive matched filter” (AMF) by Robey et al. in 1992 [68] and the “adaptive coherence estimator” (ACE) by McWhorter and Scharf in 1996 [42]. Clearly, AMF is a special case of ACE for known $\lambda = 1$. Expressions (not closed form) for the probabilities of detection and false alarm have been derived in the literature for ACE and AMF using the sample covariance matrix, and using the projection of the sample covariance matrix to the set of persymmetric matrices [28]. Integrals for error probabilities can likewise be evaluated via Monte Carlo.

2.3.2 Non-Homogeneity Detectors

A non-homogeneity detector can be utilized to reduce the performance loss from heterogeneous and/or contaminated training data. The approach in [46,47] attempts to prune the training data set to remove the contaminated data using the generalized inner-product (GIP). The GIP takes the form

$$y^H R^{-1} y,$$  \hspace{1cm} (2.31)

where $R^{-1}$ must be estimated from the data and $y$ is the vectorized data for a single range bin of the training set. This statistic is calculated for all of the training data and then ordered by value. The GIP then iteratively whitens and thresholds the data vectors to prune the training set. The approach has been rigorously analyzed and multiple techniques exist for setting the threshold [29,64,80].
Rank-ordering of the GIP statistic across range bins results in a nearly linear central region. Fitting a line to this center region then allows for thresholding of the GIP based on deviation from this fitted line. Using ordered statistic theory, the threshold value can then be chosen as a function of the number of degrees of freedom in the system and the desired false alarm rate (rejecting a target-free range bin).

2.4 Existing Estimators

In this section we will present the details of existing covariance estimators employed in STAP and group them by type. We will also consider their performance by the number of training samples each requires.

The most basic estimator is the sample covariance estimate. The sample covariance matrix is given as

$$\hat{R} = \frac{1}{K} \sum_{k=1}^{K} y_k y_k^H,$$

where $K$ is the number of training samples and $y_k$ is the vectorized data for a single range bin of the training data. While the sample covariance matrix is an unbiased and consistent estimator, it yields poor performance for limited training data. It is known that at least $2MN$ target-free and homogeneous training samples are required to limit the SINR loss induced by using the sample covariance estimate to $3\text{dB}$; this is known as the Reed-Mallett-Brennan (RMB) rule [66]. In practice, it is difficult if not impossible to meet this condition as oftentimes the secondary training data is limited, heterogeneous, and/or contaminated with targets. The estimators that are presented in the following subsections attempt to lessen the amount of training data required to limit the SINR loss to $3\text{dB}$. 

20
2.4.1 Constrained Maximum Likelihood Estimators

To reduce the required number of training samples, the STAP literature presents a wealth of covariance estimators based on a constrained maximum likelihood criterion. Constraints on the unknown covariance matrix include: Toeplitz structure [27], circulant structure [23], persymmetry [53], and closeness to a nominal covariance [24]. Additionally, estimators have been proposed for arbitrary clutter plus thermal noise

\[ R = R_c + \sigma_{\text{noise}}^2 I, \]  

(2.33)

where \( \sigma_{\text{noise}}^2 \) denotes thermal noise power and \( I \) is the identity matrix. Constraints adopted with Eq. 2.33 include: known lower bound on \( \sigma_{\text{noise}}^2 \) [77], an interval bound on \( \sigma_{\text{noise}}^2 \) [24], an upper bound on the rank of \( R_c \) [35], and combined rank and thermal noise bounds [35]. Further, the maximum likelihood estimator has been proposed for \( R_c \) constrained to lie in a known low-dimensional subspace [15, 25]. By working with either \( R \) or the precision matrix, \( R^{-1} \), all of these constrained maximum likelihood estimators may be expressed as convex optimization tasks. Details and derivations of some constrained MLEs are given in Appendix A. The contribution of these structured estimation techniques is to reduce the SINR loss caused by the estimated covariance when the number of available secondary training data is greater than the rank of the clutter component but less than the \( 2MN \) required by the RMB rule.

2.4.2 Knowledge-Aided Estimators

Another set of covariance estimators is known as knowledge-aided estimators. These approaches can be divided into two categories: direct and indirect knowledge-aided approaches. Indirect knowledge-aided approaches use prior knowledge to select
data used in a sample covariance estimate. For example, a land usage map can inform selection of training data [8,20]. Direct knowledge-aided approaches use prior knowledge to calculate the clutter component, $R_c$, of the covariance estimate. For example, a high-resolution digital elevation map (DEM) and two-way antenna pattern are used to estimate the deterministic clutter component of the covariance estimate [9]. The clutter covariance estimate from the DEM can also be used in a pre-whitening step [45], reducing the amount of required secondary training data. An interesting aspect of the direct knowledge-aided approaches is that some of them can be expressed in a form similar to a shrinkage estimator [9,44,78].

2.4.3 Shrinkage Estimators

Shrinkage estimators linearly combine an unbiased sample covariance estimate with possibly multiple biased shrinkage models [6,22,34,39,40,71,76,78]. The strength of these techniques lies in the reduction of estimation error through a bias/variance trade off [6]. For example, in [78] the minimum expected squared error estimator is computed using as a single-model shrinkage model the prior knowledge of a nominal covariance matrix similar to the cell under test. The difficulty with these approaches is the choice of shrinkage models. Given a good choice of shrinkage models, the main contribution of these techniques is to reduce the SINR loss caused by the estimated covariance when the number of available secondary training data is less than the rank of the clutter component.

2.4.4 Direct Data Domain Estimators

Direct data domain (D3) methods avoid heterogeneity by using only the cell under test (CUT). Some of these approaches reduce the degrees of freedom of the original
coherent processing interval (CPI) and then process the reduced data, using the rest of the original CPI as training samples [1, 2, 18, 69]. A subset of the direct data domain approaches consists of sparse reconstruction approaches. These approaches leverage the fact that the clutter and potential moving targets only fill a small part of the normalized angle/Doppler space at each range bin. One example approach assumes an estimate of the interference covariance, $R$, and uses a sparse Bayesian model to help resolve closely spaced moving targets [57]. Another example reconstructs a high-resolution angle/Doppler image at each range using a nonparametric iterative approach using weighted least squares; targets are then detected using a median CFAR detector across range [41]. Other sparse approaches exploit structure of the clutter support in the normalized angle/Doppler scene. The structure exploited is: known support of the clutter cells [75, 79], spatially grouped clutter support for a single range [85], and similar clutter support across multiple range bins [88]. The sparse reconstruction techniques in [85] and [88] then use the estimated clutter support and amplitudes to calculate a covariance matrix from few or no training samples. Similarly, the technique in [79] estimates the complex reflectivity of the entire normalized angle/Doppler scene and estimates a covariance using the known support and estimated amplitude of the clutter cells; this approach only requires the CUT with no additional training samples but relies on user-defined parameters and exact prior knowledge of the clutter support.

The sparse technique in [87] extends the work of [45] in an attempt to reduce the computational complexity of using a large set of filter banks to select the clutter support. In [87], uncertainty in both the clutter ridge location and internal clutter
motion is represented by selecting a small neighborhood of uniformly gridded normal-
ized Doppler cells, rather than a single normalized Doppler value per normalized spatial frequency. For each of three neighborhood sizes, clutter power at all gridded cells is estimated from the CPI data using the orthogonal matching pursuit (OMP) algorithm rather than least-squares. Selection among candidate neighborhood sizes is chosen using the AMF on mainbeam clutter, as in [45]. Both [45] and [87] note performance loss when the CUT is not target free.

The main contribution of these D3 techniques is to reduce the SINR loss of the estimated covariance using no or few secondary training samples. They have also shown an ability to be robust to targets in the data [67]. The major disadvantage of these techniques is the performance loss seen as a result of the discrete gridding of the normalized angle/Doppler scene used for the reconstruction of the scene and estimation of the covariance.
3.1 Introduction

Persistent surveillance synthetic aperture radar (SAR) imagery can provide for tracking of moving targets using changes in video frames; however, a moving target’s energy will occasionally be shifted in Doppler to become co-located with strong clutter returns [54,58]. This limitation is particularly present in urban operating environments. An additional challenge in wide-area persistent surveillance SAR systems is the large data rate that must be supported by a down-link or on-board processing. Motivated by these two challenges, we propose the use of multiple transmitters and a single receiver to provide enhanced ground moving target indicator (GMTI) functionality while maintaining the same data rate as a single-antenna SAR system.

The multiple antennas allow for space-time adaptive processing (STAP) [48,86]. Traditional STAP uses a single transmitter and multiple receivers (single-input, multiple-output, or SIMO). By sampling in both space and time, the returns at a given range can be separated in both angle and Doppler, thereby permitting suppression of stationary clutter. The system-level impediment to traditional STAP in this hybrid SAR/STAP context is that the down-link data rate or on-board processing must scale with the number of receive antennas. Therefore, we propose to exploit reciprocity and
quasi-orthogonal waveforms to create the degrees of freedom required to jointly perform space-time adaptive processing with synthetic aperture imaging. Specifically, we employ multiple transmit antennas and quasi-orthogonal waveforms to create virtual receive channels that are analogous to the traditional STAP approach. In order to both maintain full SAR resolution and preserve coherence on clutter, we adopt frequency division multiple access (FDMA) waveforms with time-varying sub-carrier frequencies.

While traditional SIMO STAP has a rich and long-standing literature [48,86], only recently has ground moving target indicator radar been addressed for the multiple-input, multiple-output (MIMO) and MISO architectures. A major focus in the MIMO case is the design of orthogonal waveforms, and initial studies found that common communications waveforms, such as FDMA, are generally unsuited for clutter cancellation in GMTI [10, 60–62]. For example, Rabideau [60, 61] derived the MISO cancellation ratio as a figure of merit for clutter cancellation performance, and demonstrated that fixed-subcarrier FDMA waveforms and other communications-inspired waveforms cannot maintain the clutter coherence necessary for adaptive nulling. Instead, new waveforms were proposed, including “displaced” FDMA and CDMA waveforms [60] and Doppler division multiple access (DDMA) waveforms [10, 43, 60]. DDMA, in which a slow-time modulation provides a Doppler frequency shift at each transmitter, suffers from Doppler ambiguities, so the pulse repetition frequency must be increased by a factor equal to the number of transmit antennas. A dithered DDMA approach – with pseudo-random slow-time dither of both frequency and phase – has been proposed to combat the Doppler ambiguities [61, 62]. While viable approaches for GMTI, neither DDMA nor FDMA/CDMA have been combined with a SAR modality.
In this chapter we explore how the displaced FDMA waveforms can be used for both SAR and STAP modalities. We also explore the STAP performance of the displaced FDMA waveforms in greater detail.

### 3.2 Scheduled FDMA Waveforms

For quasi-orthogonal waveforms, we employ traditional linear frequency modulated (LFM) pulses with separate sub-carriers and disjoint frequency sub-bands. Thus, in the proposed FDMA signaling scheme the $n^{th}$ antenna transmits the following LFM waveform at the $m^{th}$ pulse:

\[
s_{n,m}(t) = \exp\left\{ j2\pi \left( f_c + f_{n,m} + \frac{\alpha}{2} t \right) t \right\}, \quad -T/2 < t < T/2, \tag{3.1}
\]

where $\alpha$, $(f_c + f_{n,m})$, $T$, $N$, and $M$ are the LFM rate, $n^{th}$ sub-band carrier at the $m^{th}$ pulse, pulse duration, number of antennas, and number of pulses, respectively. In keeping with this definition, the carrier frequency offset for each transmit antenna is calculated as

\[
f_{n,m} = \frac{2I_{n,m} - N - 1}{2} \left( \frac{BW}{N} \right), \tag{3.2}
\]

where $I_{n,m}$ is an indexing variable defined below, and $BW$ is the combined bandwidth of all the transmitted pulses.

In order for STAP to adaptively null clutter, clutter coherence must be maintained from pulse to pulse. A potential problem for a FDMA-based scheme is the loss of coherence across disjoint frequency sub-bands. This decorrelation is quantified, for example, in the MISO cancellation ratio (MCR) [60, 61]. To ameliorate the clutter coherence issue we propose scheduling the waveforms such that at a given aperture
position, a similar frequency sub-band is used to illuminate the scene from pulse to pulse. Thus, the transmit waveform on a given antenna becomes a function of the platform velocity, antenna spacings, and pulse repetition frequency (PRF).

Let $\beta$ denote the number of half inter-element spacings traveled by the platform between each pulse,

$$\beta = \frac{2v_p}{f_r d},$$

(3.3)

where $v_p$, $d$, and $f_r$ are the platform velocity, antenna element spacing, and PRF, respectively. For the simple case where $\beta$ is equal to one, the virtual phase centers are aligned pulse to pulse; and, the sub-carrier frequencies are scheduled so that the same frequency is used at the same location in space, regardless of which phase center occupies that location. This is illustrated in Figure 3.1 which compares the non-scheduled waveform scheme to the scheduled one. In the figure, the case of $N = 3$ transmitters is depicted: colors denote frequency sub-bands, phase center location is indicated left-to-right, and $M = 4$ pulses are depicted down the four rows of the graphic. We note that, while [60] correctly dismisses fixed sub-carrier FDMA as a viable waveform choice, the paper shows a slow-time modulation method of “displaced” or “rotating” waveforms will work for GMTI; we adopt the rotating waveform philosophy for our MISO FDMA waveform scheduling [16].

For the general case in which $\beta$ might not be an integer, we choose to transmit carrier frequencies from antenna $n$ at pulse $m$ using Eq. 3.2 and an index function $I_{n,m}$ given as

$$I_{n,m} = 1 + (n + \lfloor m\beta \rfloor) \mod N$$

(3.4)

$$n = 0, 1, \ldots, N - 1, \quad m = 1, 2, \ldots, M - 1,$$

where $\lfloor x \rfloor$ denotes largest integer not greater than $x$. For the first pulse, $m = 0$, we define $I_{n,0} = n + 1$. For $\beta < 1$, Eq. 3.4 may be interpreted as a hold-then-rotate
Figure 3.1: Waveform scheduling comparison for a three-element linear array. Colored squares represent phase center locations, with color corresponding to a unique carrier frequency. The vertical axis corresponds to pulse number, while the horizontal axis represents the phase center location along the flight path. The left set of squares shows the non-rotating scheme, while the right set shows the proposed rotating scheme for \( \beta = 1 \); specifically, the antenna moves one half antenna spacing per pulse, resulting in phase centers that align from pulse to pulse.

scheme in which each antenna transmits the same sub-band for several pulses until the antenna phase center has been translated at least one antenna phase center spacing.

As demonstrated below, these constant-envelope FDMA waveforms provide full SAR resolution, simple SAR processing, coherence on clutter, and STAP with maximal unaliased interval of radial velocities; the quasi-orthogonality comes at the cost of \( N \) times coarser range resolution in the space-time adaptive processing, in comparison to the concurrent SAR modality.
3.3 FDMA MISO Extensions for SAR/STAP

In this section we present the extensions required for both the SAR and STAP modalities using a FDMA MISO architecture. The SAR extensions are similar to traditional SAR processing and allows formation of high-resolution images. The STAP extensions are more computationally complex than traditional STAP and results in GTMI processing with $N$ times lower range resolution in comparison to the SAR modality.

3.3.1 FDMA MISO SAR Model

The coherent construction of full-resolution SAR images in the MISO architecture requires very little departure from the single channel case. First, the MISO system with displaced transmitters and receiver can be treated like the co-located single transmitter and receiver system for most imaging geometries. For example, for side-looking X-band operation at $\lambda = 0.03\,\text{m}$, 10 km altitude, 30 degree depression angle, and $5\lambda$ antenna separation, the monostatic approximation introduces less than 0.027 degree phase error for a 2 km diameter scene. At each virtual receive channel, stretch processing is performed resulting in I/Q spatial-frequency samples. The full resolution image can then be formed by coherently summing backprojected images computed from each channel [4, 49]. For example, using three antennas, the full resolution SAR image, $Y$, can be calculated as:

$$
Y = B_{1,2}X_1 + B_{2,2}X_2 + B_{3,2}X_3
$$

(3.5)

where $B_{i,j}$ is the backprojection operator for transmitter $i$ and receive antenna $j$, and $X_i$ is the phase history for transmitter $i$. Each backprojection operator assumes
a monostatic transmitter/receiver pair located at the phase center of the offset pair. The receiver can be chosen to be any of the antennas, but for the side-looking operation, using the center antenna as the receiver reduces the total phase error. This operation can be extended to various imaging geometries such as strip map with large scan (squint) angles. For a single coherent processing interval (CPI) in the side-looking configuration, SAR resolution degrades at high scan angles. However, for the stationary scene sought in SAR processing, multiple adjacent CPIs are jointly and coherently processed to achieve a more uniform resolution across the entire scene. Although many imaging geometries exist, circular side-looking imaging will be the focus for the remainder of the chapter.

### 3.3.2 FDMA MISO STAP Model

The system model derivation for STAP follows that of Ward [86] with modifications for the scheduled FDMA waveforms and stretch processing. Invoking superposition, consider the received signal due to a single target in the far-field of the array and a narrow-band transmitted waveform from antenna $n$:

$$\tilde{r}_{n,m}(t) = a \exp\{j\psi\} \exp\left\{j2\pi \frac{\alpha}{2} (t - \tau_n)^2\right\} \times \exp\left\{j2\pi (f_c + f_{n,m} + \tilde{f}_{n,m})(t - \tau_n)\right\},$$

$$-T/2 < t - \tau_n < T/2 \quad (3.6)$$

where $a$ is the echo amplitude, $\alpha$ is the LFM rate, $\tau_n$ is the delay to the target, $\lambda_{n,m} = c/(f_c + f_{n,m})$ is the center wavelength for transmitter $n$ at pulse $m$, and

$$\tilde{f}_{n,m} = \frac{2v}{\lambda_{n,m}} \quad (3.7)$$

is the target’s Doppler frequency due to radial velocity $v$. For simplicity, we will drop the pulse number subscript, $m$, from the notation, although the frequency $f_n$, and
any variable dependent on it, varies as a function of both antenna and pulse number. The rest of the derivation follows closely to that of [86], noting that \( \vartheta_n = \frac{d}{x_n} \cos \theta \sin \phi \) defines the spatial-frequency of the target where \( d \) is the antenna spacing, \( \theta \) is the depression angle, and \( \phi \) is the azimuth angle. After stretch processing, given that the range window of the scene is small relative to the carrier frequency, the time samples are actually frequency samples with frequencies corresponding to the transmitted waveform [50]. Therefore, each virtual receive antenna will observe disjoint frequency samples corresponding to the same ranges:

\[
    r_n(t) = r_n(f) \exp\{jn2\pi\vartheta_n\} \exp\{j2\pi\tilde{f}_nt\},
\]

\[
f \in (f_c + f_a) \pm \frac{BW}{2},
\]

where \( BW = \alpha T \) is the bandwidth of the LFM pulse. These disjoint frequency samples that correspond to the same range samples create a range dependent phase wrap that differs from antenna to antenna. The range-dependent phase wrap (see, e.g., [3]) distinguishes this received signal model from traditional STAP.

To process the data into range samples, an inverse Fourier transform can be computed, and the range dependent phase wrap can then be removed by multiplying the resulting range bins by \( \exp\{j2\pi f_n \Delta_r\} \) where \( \Delta_r \) is the differential delay between the range bin and the scene center. The data must next be transformed to angle and velocity samples for each range bin. In traditional STAP, this can be done quickly by performing a two-dimensional Fourier transform across pulses and antennas for each range bin. In contrast, here we must use a direct implementation of the matched filter due to the variation in center frequency across the virtual channels. The matched filter is implemented using target steering vectors computed over the unaliased target spatial and Doppler frequencies. The steering vectors are derived as follows: first we
define the normalized Doppler frequency as
\[
\tilde{\omega}_n = \frac{\tilde{f}_n}{f_r}.
\] (3.9)

The \(N \times 1\) spatial steering vector is then defined as
\[
a_s(\vartheta_n) = \left[1; e^{j2\pi\vartheta_n}; \ldots; e^{j(N-1)2\pi\vartheta_n}\right],
\] (3.10)

and the \(M \times 1\) temporal steering vector is
\[
a_t(\tilde{\omega}_n) = \left[1; e^{j2\pi\tilde{\omega}_n}; \ldots; e^{j(M-1)2\pi\tilde{\omega}_n}\right].
\] (3.11)

Finally the \(MN \times 1\) space-time steering vector is calculated as the Kronecker product of the temporal steering vector and the spatial steering vector,
\[
a(\vartheta_n, \tilde{\omega}_n) = a_t(\tilde{\omega}_n) \otimes a_s(\vartheta_n).
\] (3.12)

Given the definition of the space-time steering vector, if \(x\) is the \(MN \times 1\) data snapshot for a given range bin, the angle and velocity samples can be calculated as
\[
z(\vartheta_n, \tilde{\omega}_n) = a^H x.
\] (3.13)

For SAR processing, it is well known that the allowable PRF is lower bounded by the velocity of the platform, \(v_p\), and the azimuth beamwidth of the antenna, \(\theta_{az}\), and upper bounded by the depression angle, \(\theta\), and the range width, \(R_w\), being imaged:
\[
\frac{2v_p\theta_{az}}{\lambda} \leq \text{PRF} \leq \frac{c}{2R_w \cos \theta}.
\] (3.14)

Given these restrictions for the SAR modality, most airborne platforms will require PRFs from several hundred hertz to a few kilohertz. For the STAP modality, the
unambiguous range of Doppler frequencies is $-1/2 \text{PRF} \leq \tilde{f} \leq 1/2 \text{PRF}$. Thus, the joint SAR/STAP operation allows unambiguous target Doppler frequencies up to

$$|\tilde{f}_{\text{max}}| \leq \frac{c}{4R_w \cos \theta};$$

which is easily translated to radial velocities,

$$|v_{\text{max}}| \leq \frac{c\lambda}{8R_w \cos \theta}.$$ 

Thus, the compatible PRF requirements for the two modalities permit joint operation.

### 3.4 FDMA MISO Clutter Covariance

In this section we derive the rank of the clutter covariance matrix for a MISO system employing FDMA-based waveforms. We consider integer $\beta$, with extension to arbitrary $\beta$ considered in section 3.5.3. First, the Doppler frequency of the clutter induced by the platform velocity is defined as

$$\tilde{f}_I = \frac{2v_p}{\lambda_I} \sin \phi$$

for clutter at an azimuth angle $\phi$. To simplify notation, the subscript $n, m$ has been substituted with the subscript $I$ so that the transmitted frequency is a function of the index number which is in turn a function of antenna and pulse number. From Eqs. 3.3 and 3.9, the normalized Doppler frequency is then found as

$$\bar{\omega}_I = \frac{\tilde{f}_I}{f_r} = \frac{2v_p}{f_r\lambda_I} \sin \phi = \beta \vartheta_I.$$ 

To proceed, define the base quantity

$$x_{I,k} = \exp\{j2\pi \vartheta_{I,k}\}$$
for center frequency index $I$ and clutter patch $k$. The linear relationship between clutter Doppler and spatial frequencies implies that, for integer $\beta$, the elements of the space-time steering vector can be written as integer powers of $x_{I,k}$ [86].

Continuing, the clutter covariance matrix admits the decomposition

$$R_c = V_c \Sigma_c V_c^H$$

(3.20)

where $V_c$ is an $MN \times N_cN$ matrix of the space-time steering vectors to each of the $N_c$ clutter patches for each of the $N$ transmit waveforms. The term $\Sigma_c$ is a $N_cN \times N_cN$ diagonal matrix of the power from each clutter patch at each sub-carrier frequency.

Since $\Sigma_c$ is positive definite,

$$r_c = \text{rank}(R_c) = \text{rank}(V_c).$$

(3.21)

Therefore, we must only consider the rank of $V_c$. We next write the structure of $V_c$ using the definitions of the space-time steering vector. For $N_c > M$,

$$V_c = \begin{bmatrix} U_1 & U_2 & \cdots & U_{N_c} \end{bmatrix}$$

(3.22)

where $U_k$ is a $MN \times N$ matrix of steering vectors for the $N$ transmit waveforms to clutter patch $k$. Due to the different transmit frequencies, each channel will have an independent observation of the clutter. Therefore, the $i^{\text{th}}$ row and $j^{\text{th}}$ column of matrix $U_k$ will take the general form

$$U_k(i, j) = \delta(I_n, m - j)x_{I,k}^{(m-1)\beta + n - 1}.$$

(3.23)
More specifically, considering the case where the waveforms are not rotated, \( U_k \) will take the form:

\[
U_k =
\begin{bmatrix}
1 & 0 & \cdots & 0 \\
0 & x_{2,k} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & x_{N,k}^{-}\beta \cdot 1 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & x_{N,k}^{-}\beta \cdot N \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & x_{N,k}^{-}(M-1) \cdot 1 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & x_{N,k}^{-}(M-1) \cdot N \\
\end{bmatrix}
\]

(3.24)

It can be seen from Eq. 3.24 that the rows can be rearranged and grouped by carrier frequency \( I \). Each carrier block in the matrix then has a Vandermonde form. Because \( U_k \) is distinct for each unique clutter patch \( k \), \( V_c \) is expected to be full rank \((NM)\). The full rank indicates that clutter suppression may not be possible since the clutter fills the entire space. These results are consistent with the MISO cancellation ratio analysis in [61].

Next, we compare this to the rank of the clutter covariance matrix that results using the rotating waveform scheme. We will focus specifically on the case where \( \beta = 1 \), as depicted on the right side of Figure 3.1. For the rotating scheme, \( U_k \) takes
the form:

\[
U_k = \begin{bmatrix}
1 & 0 & \cdots & 0 \\
0 & x_{2,k} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & x_{N,k}^{N-1} \\
\vdots & \vdots & \ddots & \vdots \\
0 & x_{2,k} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & x_{N,k}^{N-1} \\
x_{1,k} & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
\ddots & 0 & \cdots & 0 \\
x_{N,k}^{N} & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & x_{N,k}^{N-1} \\
\end{bmatrix}
\] (3.25)

We observe that for \( \beta = 1 \), the rotation scheme causes multiple rows in the matrix \( U_k \), and consequently \( V_c \), to be repeated. Only the 0th through the \((N + M - 2)\) powers of \( x_{I,k} \) appear; the rest of the entries are repeats. Indeed, the first \( N \times N \) block of \( U_k \) (see Eq. 3.25) is diagonal, and each subsequent \( N \times N \) block of \( U_k \) introduces only one non-repeated row. By row elimination, \( V_c \) has the same rank as \( \tilde{V}_c = \begin{bmatrix} \tilde{U}_1 & \cdots & \tilde{U}_{N_c} \end{bmatrix} \) where \( \tilde{U}_k \) has \( N + M - 1 \) rows and

\[
\tilde{U}_k = \begin{bmatrix}
1 & 0 & 0 & \cdots & 0 \\
0 & x_{2,k} & 0 & \cdots & 0 \\
0 & 0 & x_{3,k}^{2} & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & x_{N,k}^{N-1} \\
x_{1,k}^{N} & 0 & 0 & \cdots & 0 \\
0 & x_{2,k}^{N+1} & 0 & \cdots & 0 \\
0 & 0 & x_{3,k}^{N+2} & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & x_{N,k}^{2N-1} \\
\vdots & 0 & 0 & \cdots & 0 \\
x_{N+M-2,k} & 0 & \cdots & 0 \\
0 & x_{N,k}^{N+M-2} & 0 & \cdots & 0 \\
\end{bmatrix}
\] (3.26)
for $I' = (M - 1) - \left\lfloor \frac{M-1}{N} \right\rfloor N$. Thus, the rank of $V_c$ is no greater than $N + M - 1$.

Finally, it suffices that $1 + \left\lceil \frac{M-1}{N} \right\rceil$ clutter patches have distinct $x_{I,k}$ for all $I$ for $V_c$ to have rank $N + M - 1$. Therefore, we have shown the rank of $V_c$ is $\min(N + M - 1, N_c N)$ or more simply $N + M - 1$. Thus for $\beta = 1$, the rotating waveform scheme produces a clutter covariance rank that is specified by the Brennan rule [86].

### 3.5 Numerical Results

In this section we present numerical results for the proposed approach. The performance metrics are given followed by simulation and emulation results. The proposed approach is emulated with the 2006 Gotcha dataset. We give details as to how the system is emulated as well as calibration approaches that were required for the collected data.

#### 3.5.1 Metrics

For the SAR modality we look at the point spread function in both range and cross-range. For the STAP modality we look at the eigenspectrum of the interference covariance matrix, adapted pattern, SINR loss, SINR, and STAP output. The eigenspectrum, adapted pattern, and SINR loss are the metrics used for performance evaluation with the STAP simulation while the SINR and STAP output metrics are used for performance evaluation with the STAP emulation.

#### 3.5.2 SAR Simulation

To demonstrate that the MISO scheme can provide full resolution SAR images, a computer simulation is run to compare the point spread function of the traditional single-antenna SAR approach to the proposed FDMA approach. The time-domain
Figure 3.2: Point spread function comparison in range. Normalized slice through the backprojected image at a cross-range of zero meters for the proposed FDMA MISO scheme (solid line) and the traditional single antenna (SISO) approach (dashed line).

simulation includes demodulation, filtering, and stretch processing in an attempt to capture system-level effects. We consider a 3.58-degree circular aperture at constant altitude, ground range to scene center of 8000 m, center frequency $f_c = 9.6$ GHz, total bandwidth 640 MHz, and $N = 3$ channels. For a single ideal point reflector at scene center, a backprojection image is computed without aperture or frequency tapering. Down-range and cross-range slices through the resulting point spread function are shown in Figures 3.2 and 3.3, where they are compared to the traditional single-channel (SISO) SAR case. We observe that the point spread function for the proposed MISO scheme is essentially identical to traditional SISO case in range and very similar in cross range; we attribute the small differences in cross-range sidelobes...
Figure 3.3: Point spread function comparison in cross-range. Normalized slice through the backprojected image at a range of zero meters for the proposed FDMA MISO scheme (solid line) and the traditional single antenna (SISO) approach (dashed line).

to the fact that the $N = 3$ LFM pulses are not strictly orthogonal. Thus, the proposed FDMA MISO scheme using quasi-orthogonal LFM pulses can faithfully provide full-resolution SAR imagery.

3.5.3 STAP Simulation

In order to test the proposed FDMA MISO scheme we conduct a time-domain simulation to verify the ability of the system to null clutter. For clutter returns, 20 constant radar cross-section (RCS) scatterers are randomly distributed in each range-angle bin. We use $N = 3$ transmit antennas with Gaussian beam patterns, and we follow the waveform scheduling in Eqs. 3.2 and 3.4 for a platform velocity of 50 m/s and $\beta$ of 1, 0.5, 0.25, or 0.125. The PRF for the $\beta = 1$ case is 6410.2 Hz,
and the PRF is increased to yield the other values of $\beta$. Received echoes from all scatterers and all transmit waveforms are coherently summed at the receiver, then demodulated, filtered, and stretch processed to produce three virtual receive channels. The simulation is run for $M = 10$ pulses and repeated for 150 clutter realizations, from which a sample covariance matrix is estimated.

Figure 3.4 contrasts the clutter covariance rank with and without the waveform scheduling; FDMA waveforms with a fixed sub-carrier per antenna do not maintain coherence on clutter, as predicted in [60]. In contrast, slow-time scheduling of the FDMA waveforms results in a covariance matrix eigen-spectrum consistent with the Brennan rule [86]. These results conform to the derivation in section 3.4 which calculated the expected rank of the clutter covariance matrix for the non-rotating and rotating schemes.

For the hold-then-rotate waveform schedule in Eq. 3.2, the adapted pattern computed via sample matrix inversion is shown for $\beta = 1$ in Figure 3.5 for 45 degree depression angle. The ideal behavior would be a clutter null on a line across the angle/Doppler plane with a slope equal to $\beta$, and correspondingly a deep null at zero velocity in the zero-angle cut. The signal-to-interference-plus-noise ratio (SINR) losses for the traditional SIMO approach and the proposed MISO scheme are displayed in Figure 3.6 for zero angle cuts. From comparison of the SINR plots, we observe similar behavior for the SIMO and MISO approaches, with slightly higher losses in the MISO scheme at low values of $\beta$. The coarse angular resolution resulting from only three antennas causes a widening of the clutter null at low $\beta$ in the SIMO case, and the widening is up to 1 dB worse in the MISO case. This loss in minimal detectable velocity can be ameliorated by use of more antennas or closer antenna
spacings. Additionally, we note that the low values of $\beta$ may imply range ambiguities due to a high PRF. Thus, MISO systems designed to operate with $0.25 < \beta \leq 1$ can be expected to perform as well as the traditional SIMO architecture. The simulation results verify that the FDMA waveforms, with simple slow-time scheduling, can maintain the coherence on clutter in the MISO architecture and provide an unaliased interval of radial velocities matching the SIMO case with no change in PRF.
Figure 3.5: Adapted pattern for integer-valued $\beta = 1$ using the rotating waveform scheme. The null shows that the clutter appears in the angle-Doppler space as a line with slope of $\beta$.

### 3.5.4 STAP Emulation: Gotcha GMTI Dataset

We next demonstrate feasibility of the proposed joint SAR/STAP MISO architecture using the Air Force Research Laboratory’s Gotcha GMTI dataset [70]. The publicly released field data are de-ramped frequency samples from a three-antenna X-band system and ground-truth positions of a single test vehicle, which is in the presence of other, undocumented, movers. From this multiple receiver system, we post-process each receive channel to emulate the proposed multiple-transmitter scheme.

**MISO Emulation**

The Gotcha data set was collected using the traditional STAP approach with $N = 3$ receive antennas, one of which served also as the transmitter. The raw
Figure 3.6: Comparison of SINR losses for (a) traditional SIMO and (b) proposed MISO processing.

The phase history consists of the stretch processed data from each of the three receive channels. These data samples are post-processed to emulate MISO operation with FDMA waveforms. For a single reflector, each de-ramped channel takes the form

\[
\begin{align*}
  r(t) &= \exp\{-j2\pi\alpha(t - \tau_o)\Delta_r\} \times \\
  &\quad \exp\{j2\pi\frac{\alpha}{2} (\Delta_r)^2\} \exp\{-j2\pi f_c \Delta_r\}, \quad -T/2 < t - \tau < T/2, \\
  \text{(3.27)}
\end{align*}
\]

where \( \tau, \tau_o, \) and \( \Delta_r \) are the delay to the target, delay to the scene center, and the differential delay, respectively. To construct a FDMA sub-band consider, for example, the highest one-third of the frequency samples:

\[
\begin{align*}
  r(t) &= \exp\{-j2\pi\alpha(t - \tau_o)\Delta_r\} \times \\
  &\quad \exp\{j2\pi\frac{\alpha}{2} (\Delta_r)^2\} \exp\{-j2\pi f_c \Delta_r\}, \quad -T/6 < t - \tau - T/3 < T/6. \\
  \text{(3.29)}
\end{align*}
\]
By change of variables, let $t' = t - T/3$, and note $\alpha T = BW$ to learn
\begin{align*}
r(t) &= \exp\{-j2\pi\alpha(t' - \tau_o)\Delta_r\} \times \\
&\quad \exp\{j2\pi\frac{\alpha}{2}(\Delta_r)^2\} \times \\
&\quad \exp\{-j2\pi(f_c + \frac{BW}{3})\Delta_r\}, \\
&\quad -T/6 < t' - \tau < T/6.
\end{align*}
(3.30)

From Eq. 3.30, we observe that simple modified partitioning of the collected data samples allows an effective emulation of the proposed FDMA MISO scheme using an existing airborne data collection.

**Calibration**

In order to perform space-time adaptive processing, we require antenna spacings, channel gains, and channel phases. As these calibration parameters are not included in the released data, they are estimated directly from the SIMO data.

To estimate the antenna spacings we adopt the method of Gierull [30]. The technique exploits the phase difference between channels across Doppler, which is a phase ramp with slope determined by the known velocity of the platform and the antenna spacings. Figure 3.7 shows a linear fit to the phase ramp across Doppler for both antenna pairs for one CPI of the data set. Using this technique, the antenna spacings are estimated for the entire 71 second scene, and the results are shown in Figure 3.8. The average spacings are used to implement STAP.

Given the estimated antenna spacings, we then calculate remaining imbalances among channel gains and phases [84]. The technique computes estimates as a function of range by essentially averaging across Doppler, as illustrated in Figure 3.9. The middle row depicts the phase imbalances after removal of the phase ramp determined
Figure 3.7: Phase ramp across Doppler for antenna pairs 1-2 and 1-3 at a single CPI. A least-squares estimate of the line is performed to calculate the unknown antenna spacing for this 1 second CPI.

by antenna spacings. The bottom row shows the negligible phase differences after removal of channel imbalances.

Results

Results of the proposed FDMA MISO scheme are compared to the traditional SIMO approach. First, we compare the signal-to-interference-plus-noise ratio (SINR) of the two schemes. This is done using an estimated clutter covariance matrix for a single CPI of 60 pulses, as shown in Figure 3.10. The similarity of the SINR curves illustrates the viability of the MISO approach to adaptively null clutter, yet maintain a data rate equal to a single-channel SAR system.
Figure 3.8: Antenna spacing estimates per CPI over the entire 71 second scene. The average spacing, in meters, for the entire data set is listed in the superimposed text.

The single documented mover in the urban scene allows for additional, yet anecdotal, verification of the proposed method using airborne data. Due to the large (∼15λ) antenna spacings, the clutter is ambiguous in angle. This is demonstrated in the SAR scene in Figure 3.11 which is computed from several CPIs composing a synthetic aperture of 0.72 degrees. Here, the blue line represents the scene center, the green lines represent the unambiguous angles for the STAP, and the red line and dot show the true location of the moving vehicle. Adaptive clutter suppression was implemented via sample matrix inversion STAP [48] using both the proposed FDMA MISO scheme and the traditional SIMO approach. The results are shown in Figures 3.12 and 3.13. The truth for the vehicle at this CPI is -8.01 m/s and -1.28
Figure 3.9: Gain imbalance estimates for a sample CPI. The first row shows the phase difference, in radians, between the pairs of antennas. The second row shows the remaining phase differences after removing the phase wrap across Doppler caused by the antenna spacings. Ideally, the second row would show zero mean, but range-dependent non-zero mean phase is observed. The third row shows the phase differences once the final calibration has been completed.

degrees. Both schemes detect the target, and the emulated FDMA MISO scheme is able to properly suppress clutter.

3.6 Conclusions

We have proposed a multiple-transmitter, single-receiver approach for simultaneously providing both SAR and STAP data products. The slow-time scheduling of FDMA MISO waveforms maintains full SAR resolution, provides coherence on
Figure 3.10: SINR comparisons for the traditional SIMO approach and the proposed FDMA MISO approach. The corresponding nulls between the two lines illustrate the ability of the proposed scheme to null clutter.

clutter, and limits the data rate to a single receive channel, while providing the degrees of freedom required to perform STAP. The viability of the approach has been demonstrated using both computer simulation and airborne multi-channel data.
Figure 3.11: Sample SAR image from the 2006 Gotcha GMTI dataset. The image was formed at time 40 of the 71 second scene with a resolution of 0.23m in range and 1.10m in cross-range. The blue line represents the line of sight from the array to the scene center. The red line and dot show the angle to the vehicle and the location of the vehicle, respectively. The green lines show the unambiguous angles that can be discerned in the space-time adaptive processing.
Figure 3.12: STAP results for traditional SIMO approach. The target is clearly visible close to the truth angle and velocity. Additionally, the clutter nulls can be seen as diagonal lines across the plot.
Figure 3.13: STAP results for proposed FDMA MISO approach. The target is clearly visible close to the truth angle and velocity. Additionally, the clutter nulls can be seen as diagonal lines across the plot. This example illustrates the ability of the proposed scheme to effectively null the clutter and detect the target.
Chapter 4: Knowledge-Aided Bayesian Space-Time Adaptive Processing

4.1 Introduction

Ground moving target indicator (GMTI) radar processing attempts to distinguish between radar returns emanating from moving targets and stationary ground clutter. The task is confounded by the relative motion between the radar platform and the scene, as well as by the strength of clutter returns. Techniques such as space-time adaptive processing require an unknown interference covariance describing clutter, jammers, and thermal noise. The covariance is estimated from training data not under test. But, heterogeneous, contaminated, or limited training data degrade the covariance estimate and reduce detection performance. State-of-the-art techniques for interference covariance estimation reduce the required amount of training data by imposing assumed structure on the covariance matrix. Here, a Bayesian signal model is adopted for jointly estimating targets and clutter in a single cell under test, allowing GMTI processing without training data. The approach incorporates the knowledge of an approximate digital elevation map, platform kinematics (velocity, crab angle, and antenna spacings), and the belief that moving targets are sparse in the scene. Low-complexity computation with the Bayesian model is enabled by recent algorithm
developments for fast inference on linear mixing models. Results from the KASSPER I dataset show improved detection performance compared to existing techniques using scores or even hundreds of training bins.

4.2 Bayesian Model

In this section we present the Bayesian model for the proposed approach. We adopt a linear mixing model which is outlined in the first subsection. The second subsection then details the calculation of the priors.

4.2.1 Linear Mixing Model

Consider a uniform linear array of \( N \) antennas and a coherent processing interval of \( M \) pulses. After pulse compression, the vectorized data at range bin \( k \) can be modeled as a noisy linear mixing model

\[
y_k = Ax_k + n_k. \tag{4.1}
\]

Throughout, boldface lowercase letters denote complex-valued column vectors, and boldface uppercase letters denote complex-valued matrices. At range bin \( k \), the \( P \) entries in \( x_k \) discretize the scene into \( P \) cells that cover the un-aliased normalized spatial and Doppler frequencies. The \( i \)th column of the \( MN \)-by-\( P \) matrix \( A \) is a steering vector \[86\] with normalized spatial (angle) frequency \( \vartheta_i \) and normalized Doppler frequency \( \bar{\omega}_i \) and is given by

\[
a_s(\vartheta_i) = \left[1; \ e^{j2\pi \vartheta_i}; \ldots; \ e^{j(N-1)2\pi \vartheta_i}\right]^	op
\]

\[
a_t(\bar{\omega}_i) = \left[1; \ e^{j2\pi \bar{\omega}_i}; \ldots; \ e^{j(M-1)2\pi \bar{\omega}_i}\right]^	op \tag{4.2}
\]

\[
a(\vartheta_i, \bar{\omega}_i) = a_t(\bar{\omega}_i) \otimes a_s(\vartheta_i).
\]
The steering vector $\mathbf{a}_i = \mathbf{a}(\vartheta_i, \bar{\omega}_i)$ is the response to a unit amplitude reflector at normalized angle/Doppler pair $(\vartheta_i, \bar{\omega}_i)$. Oversampling of the scene such that $MN < P$ yields an under-determined system which we wish to invert to estimate $\mathbf{x}_k$ from the noisy measurements $\mathbf{y}_k$. We can solve this inverse problem by exploiting the fact that the signal $\mathbf{x}_k$ exhibits structured sparsity with fewer than $MN$ degrees of freedom. Specifically, clutter appears in the normalized angle/Doppler scene as a function of the topography of the terrain, while moving targets are sparse and occur at few of the remaining frequencies. Then, given a DEM, platform kinematics, and the belief that moving targets are sparse, we can pose this inverse problem in the context of a Bayesian framework.

We begin the Bayesian modeling by partitioning the scene into clutter and non-clutter cells, utilizing an unknown binary clutter indicator variable, $c_i$, for the $i^{th}$ cell. Conditioned on the clutter indicator variable, the complex reflectivity of the scene at every cell can be modeled with one of two priors,

$$ p(x_i|c_i) = (1 - c_i)p_0(x_i) + c_i p_1(x_i). \quad (4.3) $$

Given that the cell is not clutter, $c_i = 0$, we model the scene reflectivity with a Bernoulli-Gaussian (BG) prior, $p_0(x_i)$. The BG prior enforces our assumption of sparse moving targets in the scene. Additionally, if the cell is clutter, $c_i = 1$, the scene reflectivity is modeled with a Gaussian mixture model, $p_1(x_i)$. We choose to use a Gaussian mixture model in order to be robust to the weighting of the scene reflectivity by the unknown two-way beam pattern of the system and to flexibly
approximate non-Gaussian clutter behavior. Thus, the two priors are

\[
p_0(x_i) = (1 - \lambda)\delta(x_i) + \lambda \mathcal{CN}(x_i; \mu_{\text{mov}}, \sigma^2_{\text{mov}})
\]

\[
p_1(x_i) = \sum_{l=1}^{L} w_l \mathcal{CN}(x_i; \mu_l, \sigma^2_l)
\] (4.4)

where \(\delta(x_i)\) is the probability mass function with all probability at zero, \(\lambda \in (0, 1)\) is the sparsity weight, \(w_l\) are the Gaussian mixture weights, and \(\mathcal{CN}(x; \mu, \sigma^2)\) denotes a circular complex Gaussian density with mean \(\mu\) and variance \(\sigma^2\). To complete the Bayesian model we must next introduce a prior on the clutter indicator variable, \(c_i\). The clutter indicator prior takes a Bernoulli distribution given as,

\[
p(c_i) = \begin{cases} 
1 & \text{w.p. } \alpha_i \\
0 & \text{w.p. } (1 - \alpha_i)
\end{cases}
\] (4.5)

where the abbreviation w.p. stands for with probability. The values \(\alpha_i\) are calculated from the DEM and platform kinematics. For notational simplicity, we drop the explicit index for range bin \(k\). The Bayesian data model admits the factorization,

\[
p(x, c | y) \propto p(y | x) \prod_i p(x_i | c_i) p(c_i),
\] (4.6)

where \(p(y | x) \sim \mathcal{CN}(y; Ax, \sigma^2_{\text{noise}} I)\), \(\sigma^2_{\text{noise}}\) is the variance of the additive white Gaussian model for thermal noise at the receiver, and \(I\) is the identity matrix. Note that this modeling choice assumes independence of the cell reflectivities. While this may not be true given the oversampling of the scene and the radar parameters, we assume independence knowing the model may sacrifice additional regularization that could be gained by exploiting the correlation among the cell amplitudes. This sacrifice admits a more computationally efficient model.

The Bayesian model in Eq. 4.6 may be visualized as a factor graph shown in Fig. 4.1. Inference is then performed on the graph using message passing. The
Figure 4.1: Bayesian model visualized as a factor graph. Dense, black portion can be accelerated using fast inference algorithms. Red portion represents incorporation of prior knowledge from digital elevation map and platform kinematics. Blue portion represents incorporation of prior knowledge that moving targets are sparse in the scene.

The traditional Sum-Product message passing algorithm [38] for Bayesian inference is intractable on these dense linear mixing graphs due to the coupling of the elements of $\mathbf{x}$ through the mixing matrix $\mathbf{A}$. The message passing can be made tractable using Generalized Approximate Message Passing (GAMP) [56, 63]. GAMP uses Gaussian assumptions and Taylor series approximations to decouple the mixing problem. Using a Laplacian prior and a maximum a posteriori (MAP) estimate with these message passing Bayesian approaches can be shown to be equivalent to more traditional L1 optimization problems [57]. Use of GAMP also introduces more tuning parameters than the traditional L1 optimization problems. However, we can exploit an expectation-maximization (EM) procedure [83] to iteratively learn the nuisance parameters of the
proposed model, making our processing fully self-tuning. The EM procedure is made possible through GAMP, which conveniently provides computation of the expectation step. We collect the nuisance parameters in the vector $q = \{\lambda, \omega_l, \sigma_{\text{mov}}^2, \sigma_l^2\}$ where we have chosen to set the means of the priors to zero and the number of Gaussian components in the mixture term to $L = 3$. This choice of $L$ was determined experimentally, but $L$ may likewise be adaptively learned from the data [83]. Finally, we assume that the thermal noise, $\sigma_{\text{noise}}^2$, is known but it also can be automatically tuned [82,83].

4.2.2 Calculation of Clutter Prior

Our proposed model incorporates prior knowledge of an approximate DEM and platform kinematics by calculating the prior probability, $\alpha_i$, that cell $i$ contains stationary clutter. The high-resolution DEM for the KASSPER I dataset is shown in Fig. 4.2.

We first consider the relationship between the normalized Doppler frequency and the normalized spatial frequency of clutter using the scene geometry, as described in [86]. Given a platform crab angle of $\phi_{\text{crab}}$, the normalized Doppler frequency of clutter $\bar{\omega}$ is then a function of normalized spatial frequency, $\vartheta$:

$$
\bar{\omega} = \beta \vartheta \cos \phi_{\text{crab}} \pm \sqrt{\vartheta^2 \beta^2 (\cos^2 \phi_{\text{crab}} - 1) + \frac{r^2 - h^2}{r^2} \left( \frac{2v_p}{f_r \lambda_c} \right)^2 \sin^2 \phi_{\text{crab}}} \quad (4.7)
$$

where $\beta = \frac{2v_p}{f_r d}$, $d$ is the antenna spacing, $v_p$ is the platform velocity, $f_r$ is the pulse repetition frequency, $r$ is the slant range of the clutter patch, $h$ is the height of the clutter patch below the platform, and $\lambda_c$ is the wavelength of the carrier frequency. It can be seen from Eq. 4.7 that in the absence of crabbing ($\phi_{\text{crab}} = 0$), the normalized Doppler frequency is then $\bar{\omega} = \beta \vartheta$. However, when a crab angle is present, the
Figure 4.2: High-resolution digital elevation map provided with the KASSPER I dataset. The color scale represents the height of the terrain below the platform in meters. Therefore, from the perspective of height above sea level, the blue areas represent higher elevations while yellow represents lower elevations.

The normalized Doppler frequency depends on the range and height of the clutter, and the ± in Eq. 4.7 represents the front- and back-lobe clutter components.

The work in [87] extends the work of [45] to reduce the computational complexity of using a filter bank that provides a robustness to errors in the prior knowledge of the DEM and platform kinematics. Instead of discretizing only the normalized spatial frequency and using Eq. 4.7 to find the Doppler of the clutter as in [45], the approach in [87] considers a uniformly oversampled normalized angle/Doppler grid using the same oversampling factor as [45]. The approach in [87] then choses the Doppler cell at each spatial frequency closest to the value from Eq. 4.7 to represent clutter. To be robust to errors in the prior knowledge, the approach also considers
choosing the normalized angle/Doppler cells surrounding the one chosen with Eq. 4.7. This neighbor selection is done with square windows of size $3 \times 3$ and $5 \times 5$.

Our proposed approach also assumes a uniformly oversampled normalized angle/Doppler grid but attempts to incorporate the errors in the prior knowledge with the Bayesian framework. We assume that for a given range and normalized spatial frequency, the normalized Doppler frequency has a distribution,

$$p(\bar{\omega}|r, \vartheta) \sim N(\bar{\omega}; \mu_{\bar{\omega}}, \sigma_{\bar{\omega}}^2). \quad (4.8)$$

The prior probability of clutter can then be calculated for each cell as

$$p(c_i = 1) = \alpha_i = \int_{\bar{\omega}_1}^{\bar{\omega}_2} N(\bar{\omega}; \mu_{\bar{\omega}}, \sigma_{\bar{\omega}}^2) d\bar{\omega} \quad (4.9)$$

where the interval $(\bar{\omega}_1, \bar{\omega}_2)$ is the normalized Doppler bounds of the $i^{th}$ cell.

To complete the discussion of the clutter prior, we consider how to calculate the mean $\mu_{\bar{\omega}}$ and variance $\sigma_{\bar{\omega}}^2$. While many heuristic approaches exist, we consider treating the parameters $h$, $\phi_{crab}$, and $v_p$ in Eq. 4.7 as random variables with a known mean and variance. With this approach, we can then use the non-linear relationship in Eq. 4.7 and a first-order Taylor series approximation to estimate $\mu_{\bar{\omega}}$ and $\sigma_{\bar{omega}}^2$. Fig. 4.3 shows an example clutter prior where we have considered both random clutter heights and random crab angle. For the remaining part of the chapter we consider only random heights where the entire DEM is characterized by a single mean and variance calculated from the high-resolution DEM provided with the KASSPER I dataset.

### 4.3 Detection Approaches

We propose two detection approaches: using the AMF with an estimated covariance matrix from the Bayesian model and using a Bayesian hypothesis test from the
approximate posterior outputs of the Bayesian model. These threshold tests (shown in Eq. 4.10 and Eq. 4.11) offer dissimilar strengths. On the one hand, the test in Eq. 4.11 derived from the Bayesian signal model can resolve closely spaced moving targets, due to the sparse prior. The Bayesian hypothesis test also provides a measure of confidence about whether a normalized angle/Doppler cell represents a moving target. However, the test performs poorly for low amplitude moving targets because the estimated moving target variance, $\hat{\sigma}_{\text{mov}}^2$, is obtained from a very small sample size. Consequently, a clutter component in the Gaussian mixture model may have lower variance than $\hat{\sigma}_{\text{mov}}^2$, hence confusing the estimated model likelihoods appearing in Eq. 4.12. On the other hand, the AMF test using the MMSE covariance estimate, $\hat{R}$, is robust to mixture modeling errors and can reliably detect low amplitude moving
targets. However, the AMF test may suffer SINR losses in clutter sidelobes due to imperfect nulling. These relative strengths and weaknesses are discussed further in section 4.4.

4.3.1 MMSE Covariance Estimation

Given the priors in Eqs. 4.3-4.5 and the data snapshot, \( \mathbf{y} \), in Eq. 4.1 for a cell under test, the message passing algorithm produces a posterior probability of clutter, \( p(c_i = 1|\mathbf{y}) \), and a posterior probability of scene reflectivity, \( p(x_i|\mathbf{y}, c_i = 1) \). From these, we can directly compute a minimum mean squared error (MMSE) estimate of the unknown interference covariance matrix. Let \( \mu_i \) and \( \sigma_i^2 \) denote the mean and variance of the posterior scene reflectivity given clutter, \( c_i = 1 \), with the calculation detailed in Appendix D. Then, the MMSE estimate of the interference covariance is

\[
\hat{R} = E\{R|\mathbf{y}\} = E\{R_c|\mathbf{y}\} + E\{R_n|\mathbf{y}\} = \sum_{i=1}^{P} p(c_i = 1|\mathbf{y})(|\mu_i|^2 + \sigma_i^2)\mathbf{a}_i \mathbf{a}_i^H + \sigma^2_{\text{noise}} \mathbf{I},
\]

where \( E \) denotes expectation, \( R_c \) is the clutter covariance, and \( \sigma^2_{\text{noise}} \mathbf{I} \) is the thermal noise covariance. Note that the term \( p(c_i = 1|\mathbf{y})(|\mu_i|^2 + \sigma_i^2) \) is unique to our approach compared to existing approaches. The proposed approach estimates the clutter covariance using the reconstruction results of the entire scene weighted by the posterior probability each cell contains clutter. In contrast, the approach in [79], for example, requires known clutter support in order to perform this estimation. The estimated interference covariance is then used in the AMF threshold test [21,68]

\begin{equation}
\frac{|\mathbf{a}_i^H \hat{R}^{-1} \mathbf{y}|^2}{\mathbf{a}_i^H \hat{R}^{-1} \mathbf{a}_i} \overset{H_1}{\underset{H_0}{\gtrless}} \tau.
\end{equation}

62
4.3.2 Bayesian Hypothesis Testing

Alternatively, the Bayesian model in Eq. 4.6 may be employed to directly construct a non-Gaussian Bayesian hypothesis test. The test compares the posterior probability of the model \( M_T \), for target present, to the posterior probability of the model \( M_0 \), for target not present. The Bayesian hypothesis test is readily constructed from the posterior probabilities produced from message passing and takes the form

\[
\frac{p(M_T | \mathbf{y})}{p(M_0 | \mathbf{y})} \geq \tau
\]

where,

\[
p(M_T | \mathbf{y}) = \frac{(1 - \alpha_i)\lambda S_{T,i}}{(1 - \gamma_i)\mathcal{C}N(0; \hat{r}_i, \mu_r^i) + \alpha_i \sum_{i=1}^{L} w_l S_{l,i}}.
\]

Parameters appearing in Eq. 4.12 are defined and derived in Appendix D.

4.4 Numerical Results

GMTI processing using the proposed Bayesian signal model is tested using KASSPER data [7]. The KASSPER I dataset is a synthetic dataset based on a land region located in the western United States. KASSPER includes a DEM as well as road and urban maps to simulate a realistic scenario. The pulse compressed data include no angle-independent steering vector errors, but do have angle-independent steering vector errors and simulated internal clutter motion (ICM) effects not present in our Bayesian signal model. Parameters for the KASSPER data are given in Table 4.1.

Performance is compared to two direct data domain approaches: the maximum likelihood estimation detector (MLED) [2] and the sparse reconstruction (SR) technique presented in [87]. The MLED approach is chosen because it has shown to have CFAR properties [1] similar to the AMF and does not require user-defined tuning.
Table 4.1: Parameters for the KASSPER I dataset.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>32</td>
</tr>
<tr>
<td>$N$</td>
<td>11</td>
</tr>
<tr>
<td>Carrier frequency</td>
<td>1240 MHz</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>10 MHz</td>
</tr>
<tr>
<td>PRF</td>
<td>1984 Hz</td>
</tr>
<tr>
<td>$\phi_{\text{crab}}$</td>
<td>3 degrees</td>
</tr>
<tr>
<td>$v_p$</td>
<td>100 m/s</td>
</tr>
</tbody>
</table>

parameters. The MLED partitions the CPI into lower-resolution subsets in order to perform averaging without secondary training data; we have chosen to partition the 11-by-32 CPI into sliding windows of 5 antennas by 13 pulses. This choice ensures that there are $2MN$ training samples [2]. Performance is also compared to the SR technique presented in [87]. For this approach we oversample the normalized angle/Doppler scene by 6 in both normalized spatial and Doppler frequencies and set the sparsity level of the OMP algorithm to 45, which is a little higher than the rank of the true covariance, as suggested in [87]. The estimated covariance is then used in the AMF threshold test. We also compare performance to three other existing GMTI techniques that use an estimated covariance estimate in the AMF threshold test: the sample covariance matrix, the rank-constrained maximum likelihood (RCML) estimate [35], and the knowledge-based (KB) structured maximum likelihood estimate [25]. The RCML assumes a known clutter rank, $r_c$, based on the Brennan rule [86]; using a waterfilling argument the RCML estimate is elegantly computed in closed form from the sample covariance. The KB estimator assumes a
covariance of the form

\[ R = U^H B U + \sigma^2_{\text{noise}} I \]  

(4.13)

with known \( r_c \)-dimensional clutter subspace \( \text{range}(U) \) spanned by the columns of the \( MN \)-by-\( r_c \) matrix, \( U \). For non-integer \( \beta \), as is the case in the KASSPER data, the KB estimator uses the principle \( r_c \) eigenvectors of the sample covariance to construct the matrix \( U \). Using the Brennan rule to calculate the rank, we use \( r_c = 42 \) for both the RCML and KB estimation approaches. We also assume a known thermal noise power, \( \sigma^2_{\text{noise}} \), for the proposed approach as well as for SR, RCML, and KB.

The proposed approach estimates the complex scene reflectivity for a discretized scene at each range bin. The scene is gridded with uniform spacing over the unaliased normalized spatial and Doppler frequencies using approximately 60\( N \) normalized spatial frequency bins and 20\( M \) normalized Doppler bins. We assume only the squint angle and 6 dB beamwidth of the system in order to reconstruct the subset of cells that either reside in the mainbeam or have an appreciable prior probability of containing front- or back-lobe clutter. The choice to include only these cells in Eq. 4.1 assumes there is no moving target with radar cross-section (RCS) sufficiently large to be detectable by the weak illumination in the antenna pattern sidelobes.

### 4.4.1 Metrics

We evaluate algorithm performance using two widely employed performance metrics: average SINR loss and receiver operating characteristic (ROC) curve. The ROC curve reveals detection behavior potentially obscured by the scalar SINR loss. SINR loss is computed for each normalized angle/Doppler cell, \( i \), and each range bin, \( k \),
using the true covariance, $R_k$, and the estimated covariance, $\tilde{R}_k$,

$$\text{SINR Loss} = \frac{|a_i^H \tilde{R}_k^{-1} \tilde{a}_i|^2}{|a_i^H \tilde{R}_k^{-1} R_k \tilde{R}_k^{-1} a_i||a_i^H R_k^{-1} a_i|}.$$ (4.14)

The resulting loss is averaged across both $i$ and $k$ to yield a single scalar SINR metric and is reported in decibels.

An empirical ROC curve is computed using the true locations of moving targets present in the mainbeam; the true locations are provided in the KASSPER I dataset. We declare a potential detection to be a correct detection if a true target lies within one Fourier bin. The ROC curve is then constructed as probability of detection ($P_D$) versus probability of false alarm ($P_{FA}$) where [59]

$$P_D = \frac{\# \text{ of correct detections}}{\# \text{ of correct detections} + \# \text{ of missed detections}},$$

$$P_{FA} = \frac{\# \text{ of false positives}}{\# \text{ of false positives} + \# \text{ of correct omissions}}.$$ (4.15)

The curve is traced by varying the detection test threshold value.

### 4.4.2 KASSPER I Dataset

Performance of the proposed GMTI processing is compared to existing techniques using the SINR loss and ROC curve metrics. First, the average SINR loss is shown in Fig. 4.4 for the proposed MMSE estimate of the covariance, the RCML estimate, and the sample covariance matrix (SMI). The MMSE estimate, which uses only the range cell under test, yields an average SINR loss comparable to techniques that require $r_c$ or more secondary training data range bins.

Second, in Fig. 4.5 we compare ROC curves for eight detectors. A clairvoyant detector uses the known KASSPER covariance matrix in the AMF threshold test of Eq. 4.10; the AMF test is also computed using the proposed MMSE covariance
Figure 4.4: SINR loss averaged across range bins, normalized spatial frequencies, and normalized Doppler frequencies. The proposed approach does not use any secondary training data; therefore SINR loss is displayed for a single training bin, and the value is extended as a dashed line for visual comparison.

The proposed approach using the MMSE estimate of the covariance matrix from only the CUT outperforms the other six non-clairvoyant tests considered. We attribute this performance gain to our model’s ability to reject targets in the data from contaminating the estimate of the covariance matrix. We also attribute the lower
Figure 4.5: Receiver operating characteristic (ROC) curves comparing proposed detection schemes to sample matrix inversion (SMI), rank-constrained maximum likelihood (RCML), knowledge-based (KB), maximum likelihood estimation detector (MLED), and sparse recovery (SR) approaches. Since the KASSPER I dataset is simulated, we can also compare to the performance of the clairvoyant detector, shown in the solid black line.

detection performance of the SR technique to the fact that the clutter power estimates are biased due to both targets in the CUT and discretization of the normalized spatial and Doppler frequencies. It has been observed from experimentation that a uniformly sampled normalized angle/Doppler scene requires sufficient oversampling to achieve the best results. It has also been observed that non-uniform oversampling can increase detection performance but comes at a cost of computational complexity and memory since a fast Fourier transform can no longer be utilized. As noted in [2], covariance estimation in MLED suffers from correlation among the windows taken
from the single CPI and from the non-stationarity of statistics due to ICM. Thus, the data does not conform to the assumptions underlying MLED covariance estimation.

As mentioned previously, we also attribute the lower performance of the Bayesian hypothesis test to its inability to distinguish a low amplitude moving target from low amplitude clutter emanating from one of the Gaussian components with low variance and a high weight. This can be seen in Eq. 4.12: the likelihood term representing a mover, \( S_{T,i} \), is weighted by \( \lambda \) which is very small depending on the sparsity of the moving targets. Therefore, a moving target with weak reflectivity can have \( S_{T,i} \approx S_{l,i} \); given \( \lambda \ll \alpha \) and \( w_l \approx 1 \), the test will not detect the moving target with weak reflectivity.

Fig. 4.6 shows for the proposed MMSE and RCML estimates the SINR loss at each normalized angle/Doppler pair averaged across range. We can see from the SINR loss comparison that the proposed approach has losses in the clutter sidelobes resulting from imperfect nulling due to the discretization of the scene (which could potentially be ameliorated with a data adaptive step similar to [9, 44, 45]), while the RCML approach has losses where there are moving targets in the training data. The targets in the training data cause a bias in the RCML covariance estimate. The SINR losses incurred by the RCML approach limit its ability to detect targets in the KASSPER scenario due to the clustered nature of the targets in the normalized angle/Doppler scene. We attribute the similar ROC curves of the RCML and KB approaches to the fact that they both estimate the clutter subspace from the sample covariance matrix; therefore, the KB and RCML estimation approaches have similar SINR losses due to targets in the training data.
Figure 4.6: SINR losses at normalized angle/Doppler pairs averaged across range. Figures (a) and (b) show the losses for the proposed MMSE covariance estimate with and without the KASSPER target locations overlaid (shown as white circles). Figures (c) and (d) show the losses for the RCML covariance estimate with and without the KASSPER target locations overlaid. The vertical dashed lines show the bounds of the 6 dB mainbeam of the system which has been electronically steered away from broadside.

We now discuss the computational complexity of the proposed approach. The maximum number of iterations of the message passing is limited to 50 iterations and the EM outer loop is limited to a maximum of 30 iterations. The complexity of each iteration is dominated by two fast Fourier transforms (FFTs); thus, the computational
complexity is $O(P \log P)$. We observe the worst-case computation time for jointly imaging the moving targets and clutter to be approximately 3000 times the cost of an FFT. Because each range bin is processed independently, parallelization is trivial.

4.5 Conclusions

Covariance estimation for space-time adaptive processing is hampered by limited, heterogeneous, or contaminated training data. Existing approaches try to exploit structure of the covariance to reduce the number of training samples required. We have proposed a Bayesian framework that incorporates an approximate digital elevation map, approximate platform kinematics, and the belief that moving targets are sparse to jointly estimate target and clutter reflectivity range-by-range. The proposed model yields a covariance matrix estimate without the use of secondary training data and is fully adaptive without the need for hand picking regularization parameters or choosing a rank of the clutter covariance. The results demonstrate that the proposed technique is able to outperform existing techniques when applied to the KASSPER I data.
Chapter 5: Multi-Model Shrinkage for Knowledge-Aided
Space-Time Adaptive Processing

5.1 Introduction

Space-time adaptive processing (STAP) gets its adaptivity from estimating the second order statistics of the clutter from secondary data. In practice, the amount of available secondary data may not be sufficient for the sample covariance estimate to be accurate. Existing techniques exploit known structure of the clutter covariance, such as rank, to reduce the need for secondary training data. More recently in the radar literature, single-model shrinkage estimators have been shown to reduce the reliance on training data given that a good a priori model for the covariance is available. We extend shrinkage estimation to a regularized multi-model approach that incorporates inexact knowledge of a digital elevation map (DEM) to reduce the need for large amounts of secondary data. Using both simulation and the KASSPER I dataset, performance of the proposed approach is compared to low-rank estimation and various single-model shrinkage estimation approaches. The proposed approach offers the practical ability to reliably estimate the clutter covariance from a number of training data less than the rank of the covariance matrix.
Figure 5.1: Visualization of shrinkage estimators. The single-model shrinkage estimators, \( \tilde{R}_{\text{SMS}} \), have a smaller squared error than the unbiased estimate or shrinkage model by itself. Using a multi-model shrinkage estimator, \( \tilde{R}_{\text{MMS}} \), reduces the squared error even further.

5.2 Shrinkage Estimator Overview

Shrinkage estimators are a class of estimators that combine a low-variance, biased model with the high-variance unbiased sample estimate to achieve reduced mean squared estimation error. For estimation in two or more dimensions, the shrinkage estimate is always better than the sample estimate [76]. Shrinkage was first applied to covariance matrices by [40], where the scaling factors are calculated by minimizing the expected squared error. This idea has also been extended from single-model shrinkage to multi-model shrinkage [6,39,71]. Single-model shrinkage and multi-model shrinkage can be intuitively visualized in Fig. 5.1 which is adapted from [6]. Fig. 5.1 shows how the single-model shrinkage estimate, \( \tilde{R}_{\text{SMS}} \), results in a smaller error (shown as a thin black line) than either the unbiased estimate, \( \hat{R} \), or the shrinkage model by itself.
Additionally, the multi-model shrinkage estimate, $\hat{R}_{\text{MMS}}$, has the ability to further reduce the error, shown as the thick black line in Fig. 5.1.

5.2.1 Single Model Shrinkage Estimators

A single-model shrinkage estimator has been proposed for use with STAP in [78]. The single-model shrinkage estimator presumes prior knowledge of an approximate covariance matrix and constructs either the convex combination or the general linear combination of the sample covariance matrix and shrinkage model. Minimizing the expected squared error results in a closed-form solution for the optimal scaling factors; this principled approach for calculating the scaling factors extends the previous direct knowledge-aided work [9] that has a similar estimator with no principled approach for selecting scaling factors. Given that there is no physical reason for using the convex combination in the radar context, the authors in [78] opt for using the general linear combination. As the number of samples increases, the estimator converges to the sample covariance matrix, which is a consistent estimator.

5.2.2 Multi-Model Shrinkage Estimators

A convex combination multi-model shrinkage estimator has been proposed in [6,39] for real-valued covariances. The expected squared error can be reformulated into a quadratic program, which is then solved to calculate the scaling factors. Below, a multi-model shrinkage estimator is adopted for complex-valued covariance estimation in STAP; moreover, a regularized shrinkage is proposed as a novel heuristic to incorporate physical knowledge available in the application.
5.3 Multi-Model Shrinkage with Regularization

In this section we extend the multi-model shrinkage estimator presented in [6,39] to the non-negative combination of many shrinkage models for a possibly complex-valued covariance matrix. The resulting convex program is regularized, in the manner of the elastic net [89], to enable incorporation of physical knowledge available in the STAP application. We also present possible model extensions.

5.3.1 Non-Negative Complex Valued MMS

For the multi-model shrinkage estimator, the covariance estimate, $\tilde{R}$, is given by the linear combination

$$\tilde{R} = \lambda_1 \hat{R} + \sum_{k=2}^{K+1} \lambda_k M_k$$

(5.1)

for $K$ shrinkage models, $M_k$. Previous work in [6,39] considers the convex combination of Eq. 5.1 where $\lambda_1 = 1 - \sum_{k=2}^{K+1} \lambda_k$ for real-valued data. We first consider the straightforward extension to a non-negative combination for complex-valued data. The scaling factors, $\lambda^T = \{\lambda_1, \ldots, \lambda_{K+1}\}$ are found by minimizing the expected squared error (ESE) between the estimated and true covariance shown as,

$$\text{ESE} = \mathbb{E}_y \left\{ \left\| \tilde{R} - R \right\|^2 \right\},$$

(5.2)

where $R$ is the true (unknown) covariance. The expectation in Eq. 5.2 is over the $n$ independent snapshots, $y_i \sim \mathcal{CN}(0, R)$. This cost is a heuristic widely adopted in shrinkage covariance estimators (e.g., [22, 39, 40, 78]). The resulting estimator is consistent [40], but the cost depends on the unknown covariance itself, providing no previously published performance optimality. Substituting Eq. 5.1 for the estimated
covariance, \( \hat{R} \), and rearranging terms results in the following quadratic form,

\[
ESE = \lambda^T A \lambda - 2b^T \lambda + \text{const} \tag{5.3}
\]

where const = \( \|R\|^2 \). The values of the matrix \( A \) and vector \( b \) are given in Eq. 5.4 and Eq. 5.5, respectively.

\[
A_{[1,1]} = E \left\{ \left\| \hat{R} - R \right\|^2 \right\} + \|R\|^2
\]

\[
A_{[1,k=2:K+1]} = \text{Re} \{ \text{tr} \left[ RM_k^H \right] \}
\]

\[
A_{[j=2:K+1,1]} = \text{Re} \{ \text{tr} \left[ RM_j^H \right] \}
\]

\[
A_{[j=2:K+1,k=2:K+1]} = E \left\{ \text{tr} \left[ \left( \hat{R} - M_j \right) \left( \hat{R} - M_k \right)^H \right] \right\} + \text{Re} \{ \text{tr} \left[ RM_j^H \right] \} + \text{Re} \{ \text{tr} \left[ RM_k^H \right] \}
\]

\[
- E \left\{ \left\| \hat{R} \right\|^2 \right\}
\]

\[
b_{[1]} = \|R\|^2
\]

\[
b_{[j=2:K+1]} = \text{Re} \{ \text{tr} \left[ RM_j^H \right] \}
\]

It can be seen from Eq. 5.4 and Eq. 5.5 that the matrix \( A \) and vector \( b \) rely on the true covariance matrix, \( R \), which is not known in practice; these values are replaced by sampled-based estimates. Let \( y(l) \) denote the \( l \)th data vector from a total of \( L \) training vectors, \( y(l) \in \mathbb{C}^{MN} \). Following [6,78] we adopt the estimators:

\[
E \left\{ \| \hat{R} - R \|^2 \right\} \approx \frac{1}{L^2} \sum_{l=1}^L \|y(l)\|^4 - \frac{1}{L} \left\| \hat{R} \right\|^2 \tag{5.6}
\]

\[
E \left\{ \text{tr} \left[ \left( \hat{R} - M_j \right) \left( \hat{R} - M_k \right)^H \right] \right\} \approx \text{tr} \left[ \left( \hat{R} - M_j \right) \left( \hat{R} - M_k \right)^H \right] \tag{5.7}
\]
where Eq. 5.6 comes from [78] and Eq. 5.7 comes from [6]. The remaining terms in $R$ are replaced by their sample estimates:

\[
\|R\|^2 \approx \|\hat{R}\|^2 \\
\text{tr} \left[ RM_j^H \right] \approx \text{tr} \left[ \hat{R}M_j^H \right] \\
E \left\{ \|\hat{R}\|^2 \right\} \approx \|\hat{R}\|^2.
\]

(5.8)

Using these estimated values in $A$ and $b$, the ESE is approximated:

\[
\text{ESE} \approx \lambda^T \tilde{A} \lambda - 2\tilde{b}^T \lambda + \text{const}.
\]

(5.9)

Finding the scaling factors, $\lambda_*$, is then accomplished by solving the quadratic program,

\[
\lambda_* = \arg \min_{\lambda \geq 0} \lambda^T \tilde{A} \lambda - 2\tilde{b}^T \lambda.
\]

(5.10)

The multi-model shrinkage estimate of the covariance matrix is then found by substituting $\lambda_*$ into Eq. 5.1.

5.3.2 Regularization

Here, we consider regularization of the multi-model shrinkage estimator previously outlined. The motivation for regularization comes from situations where a large set of models exists and the true covariance matrix consists of a sparse combination of the model set. We opt to use elastic net regularization [89]; this type of regularization has shown to work well for sparse regularization with correlated dictionaries and also can help with the conditioning of the $\tilde{A}$ matrix, improving the convergence of the quadratic program for finding the optimal scaling factors. The multi-model shrinkage estimator with elastic net regularization (MMS-EN) has the same form for the covariance estimate, $\tilde{R}$, in Eq. 5.1 but has a modified expected squared error.
(mESE) function for finding the optimal scaling factors. The mESE is given as,

$$mESE = E \left\{ \| \tilde{R} - R \|^2 \right\} + 2 \mu_1 \sum_{k=1}^{K+1} |c_k \lambda_k| + \mu_2 \| C \lambda \|^2_2$$  \hspace{1cm} (5.11)$$

where $0 \leq c_k \leq 1$ are scaling factors in the L1 norm of $\lambda$ that allow for the incorporation of application-specific knowledge that may exist about whether $M_k$ should be included in the estimate. Here, $C = \text{diag}(c_1, c_2, \ldots, c_{K+1})$. The values $\{\mu_1, \mu_2\}$ are two tunable regularization parameters. Similar to the previous subsection, the mESE can be represented in a quadratic form given as,

$$mESE \approx \lambda^T \tilde{A}_{reg} \lambda - 2 \tilde{b}_{reg}^T \lambda + \text{const},$$  \hspace{1cm} (5.12)$$

where $\tilde{A}_{reg}$ and $\tilde{b}_{reg}$ have absorbed the elastic net regularization terms:

$$\tilde{A}_{reg} = \tilde{A} + \mu_2 C$$  \hspace{1cm} (5.13)$$

$$\tilde{b}_{reg} = \tilde{b} - \mu_1 \lambda.$$  \hspace{1cm} (5.14)$$

Eq. 5.13 and Eq. 5.14 show how easily the elastic net regularization can be incorporated in the multi-model shrinkage estimator. As previously, $\tilde{A}_{reg}$ and $\tilde{b}_{reg}$ actually depend on the unknown true covariance, $R$; the values are approximated using sample averages in Eqs. 5.4-5.8.

### 5.3.3 Regularization: Robust Shrinkage

We can also interpret the proposed regularized shrinkage estimator in terms of a robust estimator. Because the true covariance $R$ is unknown, $A$ and $b$ must be estimated from the data and are necessarily inexact. Accordingly, we can reformulate the ESE criterion as robust estimation task. To proceed, we follow Ho et al. [33] to
assume element-wise uncertainty balls on the estimates \( \tilde{A} \) and \( \tilde{b} \):

\[
\mathcal{A} = \left\{ A : A_{i,j} = \tilde{A}_{i,j} + e_{i,j}; |e_{i,j}| \leq \Delta_{i,j} \right\} \quad (5.15)
\]

\[
\mathcal{B} = \left\{ b : b[i] = \tilde{b}[i] + d_i; |d_i| \leq \beta[i] \right\} \quad (5.16)
\]

where \( \beta[i] \geq 0 \) and the \( K \)-by-\( K \) matrix \( \Delta \) is symmetric and diagonally dominant, with \( \Delta_{i,j} \geq 0 \) for all \( i, j \); this implies that all matrices in \( \mathcal{A} \) are positive semi-definite.

Adopting a conservative minimax approach, the worst-case ESE is optimized over these uncertainty sets, yielding the robust optimization task

\[
\lambda_* = \arg \min_{\lambda \geq 0} \max_{A \in \mathcal{A}, b \in \mathcal{B}} \lambda^T A \lambda - 2 \lambda^T b. \quad (5.17)
\]

Performing the inner maximization with respect to \( b \) reduces to [33]

\[
\lambda_* = \arg \min_{\lambda \geq 0} \max_{A \in \mathcal{A}} \lambda^T A \lambda + 2 \sum_{k=1}^{K} \left( -\tilde{b}[k] + \beta[k] \text{sgn}(\lambda_k) \right) \lambda_k \quad (5.18)
\]

where

\[
\text{sgn}(\lambda_i) = \begin{cases} 
\lambda_i/|\lambda_i|, & \lambda_i \neq 0 \\
0, & \text{else.}
\end{cases} \quad (5.19)
\]

Eq. 5.18 may be rewritten

\[
\lambda_* = \arg \min_{\lambda \geq 0} \max_{A \in \mathcal{A}} \text{tr} \left( A \lambda \lambda^T \right) - 2 \lambda^T \tilde{b} + 2 \| \lambda \|_{1,\beta} \quad (5.20)
\]

where the weighted L1 norm is written

\[
\| \lambda \|_{1,\beta} = \sum_{k=1}^{K} \beta[k] |\lambda_k|. \quad (5.21)
\]

We will also employ below a weighted L2 semi-norm, with positive semi-definite \( \Delta \), written

\[
\| \lambda \|_{2,\Delta} = \sqrt{\lambda^T \Delta \lambda}. \quad (5.22)
\]
Next, the inner maximization may be performed in closed form to obtain

$$\lambda_* = \arg \min_{\lambda \geq 0} \lambda^T \tilde{A} \lambda - 2\lambda^T \tilde{b} + |\lambda|^T \Delta |\lambda| + 2\|\lambda\|_{1,\beta}. \quad (5.23)$$

Observing that the shrinkage weights are constrained to be nonnegative, we have the simplification

$$\lambda_* = \arg \min_{\lambda \geq 0} \lambda^T (\tilde{A} + \Delta) \lambda - 2\lambda^T (\tilde{b} - \beta). \quad (5.24)$$

Thus, a minimax robust covariance estimator motivates a weighted elastic net regularization, generalized to non-diagonal $L_2$ weights $\Delta$, and results in a simple quadratic program. This minimax interpretation was previously noted in the context of portfolio design [33].

The weighted elastic net was used in the context of variables selection [90] with $\Delta = I$. The regularized cost in Eq. 5.24 was employing for portfolio design [33] with diagonal $\Delta$. Here, the non-negativity of $\lambda$ permits additional simplification. The mESE is one heuristic cost function; two related cost functions are given in chapter 7. In section 7.2.2, an M-estimator [51] is proposed and admits finite sample error bounds. In section 7.2.3, a maximum a posteriori (MAP) estimate of the covariance is proposed, where a prior on clutter power is adopted to exploit coarse belief regarding a digital elevation map and platform kinematics.

### 5.3.4 Extensions: Outlier Excision

Here we present extensions for performing outlier excision on the training data directly with the shrinkage model. Assuming there exists a reasonably good unbiased sample estimate, $\hat{\mathbf{R}}$, using a small subset of the entire available training data set [80],
we can then modify the shrinkage estimator so that it adaptively selects the best of the remaining training data for inclusion in the estimate. The shrinkage estimator for outlier excision is then formulated as:

\[
\tilde{R} = \lambda_1 \hat{R} + \sum_{k=2}^{K+1} \lambda_k M_k + \frac{1}{L} \sum_{l=K+2}^{Z} \lambda_l y(l)y(l)^\mathsf{H}
\] (5.26)

where \(Z = K + 2 + (L - P) - 1\), \(K\) is the number of shrinkage models, \(L\) is the total number of training samples available, and \(P\) is the number of training samples used to estimate the unbiased sample covariance. The normalization \(L^{-1}\) in Eq. 5.26 is the weight that the remaining training data, \(y(l)\), \(l = K + 2, \ldots, Z\), would receive if all were included in the sample covariance matrix; a consequence is that the unitless scaling factors \(\lambda_l\) conveniently fall in the range \((0, 1)\). The updated mESE for outlier excision is given as,

\[
mESE = E \left\{ \left\| \tilde{R} - \hat{R} \right\|^2 \right\} + 2\mu_1 \sum_{k=1}^{K+1} \left| c_k \lambda_k \right| + \mu_2 \| (C + S) \lambda \|_2^2 + 2\mu_3 \sum_{l=K+2}^{Z} |s_l \lambda_l|
\] (5.27)

where,

\[
C = \text{diag}(c_1, c_2, \ldots, c_{K+1}, 0, \ldots, 0),
\]

\[
S = \text{diag}(0, \ldots, 0, s_{K+2}, \ldots, s_Z),
\] (5.28)

and we have added a weighted L1 penalty for the remaining training data which allows for the model to choose a subset of the remaining training data, rejecting outliers amongst the entire set. We have also included the scaling term \(s_l\) in case there is prior knowledge about the similarity of the training data to the covariance being estimated. In STAP, such information could be gathered, for example, from land usage maps. In the following section we do not consider the inclusion of the scaling factors \(s_l\) and set them all equal to one.
5.4 Multi-Model Shrinkage with Elastic Net Regularization for STAP

To apply the proposed MMS-EN covariance estimator to STAP we must first describe the choice of shrinkage models. The true interference covariance (without jammers present) is typically modeled with Eq. 2.33 and

$$R_c = \sum_{n=1}^{N_c} \zeta_n^2 a_n a_n^H$$

where $N_c$ is a large number of clutter patches, $a_n$ is the unit-length steering vector to the $n^{th}$ clutter patch, and $\zeta_n^2$ is the power of the $n^{th}$ clutter patch. Existing knowledge-aided approaches estimate $R_c$ from a DEM and knowledge of the two-way antenna beam pattern [9]. The covariance for this approach is then estimated as

$$\tilde{R} = \hat{R} + \alpha \hat{R}_c + \beta I,$$

where $\alpha$ and $\beta$ are tuned with a heuristic approach and $\hat{R}_c$ is the estimate of $R_c$ using the DEM and two-way antenna pattern. The idea behind this estimator for whitening is a combination of unknown nulling estimated from the data with deterministic nulling from prior information. This could easily be reformatted to fit the multi-model shrinkage framework by making the estimator

$$\tilde{R} = \lambda_1 \hat{R} + \lambda_2 \hat{R}_c + \lambda_3 I,$$

where the optimal scaling factors could be found by minimizing the ESE. Instead of using the shrinkage estimator in Eq. 5.31 we focus on the case where $R_c$ cannot be easily estimated from prior knowledge (i.e. DEM is not of sufficient resolution for radar resolution; DEM is only known approximately; or, two-way antenna pattern
is unknown; etc.). We therefore propose the following estimate of the interference covariance matrix for the MMS-EN approach,

$$\tilde{R} = \lambda_1 \hat{R} + \sum_{k=3}^{K+2} \lambda_k a_k a_k^H + \lambda_2 \sigma_{\text{min}}^2 I,$$  \hspace{1cm} (5.32)

where $\sigma_{\text{min}}^2$ is a minimum bound on the thermal noise floor and $a_k$ is the steering vector to the $k$th cell in a uniformly oversampled angle/Doppler grid consisting of $K$ total cells. Note that the scaling factors $\lambda_3, \ldots, \lambda_{K+2}$ in Eq. 5.32 account for both the unknown clutter power and the unknown two-way antenna pattern. The optimal non-negative scaling factors, $\lambda_*$, are then found by minimizing the mESE in Eq. 5.12 with the additional constraint of $\lambda_2 \geq 1$ to ensure the minimum thermal noise bound.

The scaling factors, $c$, in the regularized cost function, mESE, are used to incorporate prior knowledge of an approximate digital elevation map (DEM) and platform kinematics (antenna spacings, velocity, and crab angle). Exploiting the geometry of the scene, the approximate DEM, and platform kinematics, we calculate the prior probability that a cell in the discretized angle/Doppler scene contains clutter. See [67] for details of this calculation. Defining the prior probability of clutter for angle/Doppler cell $k$ as $\alpha_k$, the scaling factors in Eq. 5.11 are set such that

$$c_1 = c_2 = 0$$

$$c_k = (1 - \alpha_k) \text{ for } k = 3, 4, \ldots, K + 2.$$  \hspace{1cm} (5.33)

This choice of scaling factors, $c$, ensures inclusion of the sample covariance and noise model while scaling the effect of the L1 regularization term based on the prior belief that shrinkage model $k$ represents clutter. The performance of this proposed approach is measured by application to both simulated data and the more realistic KASSPER I dataset [7].
5.5 Numerical Results

The first simulation is used to evaluate the performance of the proposed approach when the data matches the model exactly. The simulation uses $M = 32$ pulses and $N = 11$ antennas along with a prior probability of clutter, $\alpha_k$, calculated for each of the $K$ cell locations of a two-times oversampled angle/Doppler grid using the DEM from the KASSPER I dataset. Each realization calculates a unique interference covariance by drawing active clutter locations from the prior. The active clutter locations then have a random amplitude drawn from a complex-Gaussian distribution. The covariance is then calculated using Eq. 2.33 and Eq. 5.29; the clutter-to-noise ratio (CNR) is 37dB and $\sigma^2_{\text{noise}} = 1$. A set number of training samples is then drawn from the covariance and used to estimate the covariance and calculate the performance metric; this is repeated 50 times for each number of training samples.

We also apply the proposed approach to the KASSPER I dataset. The KASSPER I data is simulated based on a DEM from a mountainous region in the western United States and includes steering vector errors as well as internal clutter motion effects [7]. Since the data is simulated, we have the true covariance of the interference as well as the true location of moving targets in the scene.

5.5.1 Metrics

The performance metric we use for comparison is average signal-to-interference-plus-noise ratio (SINR) loss shown in Eq. 5.34.

$$\text{SINR Loss} = \frac{|a_k^H \tilde{R}^{-1} a_k|^2}{|a_k^H \tilde{R}^{-1} \tilde{R} R^{-1} a_k| |a_k^H R^{-1} a_k|}. \quad (5.34)$$

The SINR loss in Eq. 5.34 is then averaged over all $K$ angle/Doppler cells where the $k^{th}$ cell has steering vector $a_k$. We compute the performance of the proposed
approach that does not use scaling factors $c_k$ (MMS-EN-NP) and the one using the prior knowledge encoded in these scaling factors (MMS-EN-P). These performances are compared to the rank-constrained maximum likelihood estimate with a lower bound on the thermal noise (RCML-LB) [35], the single-model shrinkage estimator for an identity matrix shrinkage model (SMS-I) [40], the single-model shrinkage estimator for a diagonal shrinkage model (SMS-D) [34], the single-model shrinkage estimator for the true covariance shrinkage model (SMS-C), and the sample covariance matrix (SMI). For the RCML-LB approach, we use the exact rank of the covariance for each realization noting that it is not the same for each realization, and we set the lower bound on the noise floor to the exact value of the noise floor.

For the KASSPER I dataset we also use receiver operating characteristic (ROC) curves as a performance metric. The empirical ROC curves are calculated as follows

$$P_D = \frac{\text{# of correct detections}}{\text{# of correct detections} + \text{# of missed detections}}$$

$$P_{FA} = \frac{\text{# of false positives}}{\text{# of false positives} + \text{# of correct omissions}}$$

where a detection is determined to be correct if it is within one Fourier cell in angle and Doppler of the true target location. We compare the performance of the proposed MMS-EN-P approach to the same existing approaches considered in the simulation experiment. The prior is calculated from the DEM as in [67] using a ten-times oversampled angle/Doppler grid. In order to reduce the computational complexity of the proposed algorithm, only the angle/Doppler cells with an appreciable prior probability of containing clutter (i.e., $\alpha_k > 10^{-6}$) are included as shrinkage models.
5.5.2 Simulation

First, we present the performance results of the two-times oversampled simulation described in the previous sub-section. The results of the average SINR loss are shown in Fig. 5.2. Comparing the SINR loss in Fig. 5.2, it can be seen that the techniques with the lowest SINR losses are the SMS-C approach and proposed approaches with and without the prior scaling parameters. While the SMS-C has the lowest SINR loss, this is not attainable in practice since it uses the true covariance as its shrinkage model. Observe that use of the prior scaling, $c_k = 1 - \alpha_k$, provides an additional performance gain in SINR. We believe that the gain is small in this case due to the fact that the simulation uses only two-times oversampled gridding, so the steering vectors are nearly orthogonal; the scaling is not very impactful when the oversampling
is small. The largest performance gap between the proposed approach and existing approaches is in the region where the number of available training samples is less than the rank of the true covariance. For this simulation the average rank of the covariances was 41.

5.5.3 KASSPER I Dataset

Next, we examine the results of processing the KASSPER I dataset. Again, we look at the average SINR loss shown in Fig. 5.3. Just as in the simulation, we observe that the proposed approach gives the largest performance gain when the number of training samples is less than the rank. Using the Brennan rule [86] the rank of the true covariances in the KASSPER I dataset is approximately 42. Note the decreasing performance of the SMS-C as the number of training samples increases;
as more samples are used, the variance of the sample covariance is reduced causing a larger weighting of it, compared to the clairvoyantly known true covariance used as the shrinkage model. Since the KASSPER I dataset is heterogeneous, this results in performance losses since the samples of the sample covariance do not match the true covariance. Next, we examine the ROC curves, which are generated using 5, 20, and 45 training samples shown in Fig. 5.4, Fig. 5.5, and Fig. 5.6, respectively. We can see from the ROC curves that the proposed approach is able to outperform the others when the number of training samples is less than the rank. We have also included the clairvoyant STAP results using the true covariance in order to show the optimal performance. Again the SMS-C approach is not attainable in practice and its detection performance degrades with more training samples.
Figure 5.5: ROC curve results for the KASSPER I dataset using 20 training samples.

The results of the target excision experiment using the proposed model extension are shown in Fig. 5.7. We observe that the target excision extension is able to improve the detection performance of the MMS-EN shrinkage estimator. The simple experiment in Fig. 5.7 illustrates the flexibility to extend the proposed regularized multi-model shrinkage estimator to perform outlier excision or to incorporate a land usage map.

We conclude results with a comment on computational complexity. The proposed MMS-EN-P approach has complexity dominated by construction of the $\tilde{\mathbf{A}}_{\text{reg}}$ matrix in Eq. 5.12, which requires $\mathcal{O}(K^2M^2N^2)$ multiplications, where $K$ is the number of oversampled angle/Doppler cells. In contrast, the RCML-LB [35], for example,
Figure 5.6: ROC curve results for the KASSPER I dataset using 45 training samples.

requires an eigendecomposition with complexity $O(M^3N^3)$ [31]. Thus, with $K > \sqrt{MN}$, the proposed approach will have higher complexity.

5.6 Conclusions

We have proposed an elastic net regularized multi-model shrinkage covariance estimator for use with space-time adaptive processing. The estimator has the ability to incorporate available knowledge of an approximate digital elevation map; an extension has also been proposed that can excise contaminated and heterogeneous training data and can incorporate knowledge of land usage maps. Performance comparisons using simulated data and the KASSPER I dataset show that the proposed approach is able to outperform existing approaches when the number of available training data range cells is less than the rank of the true covariance matrix.
Figure 5.7: ROC curves for the KASSPER I dataset using the proposed model extension for target excision.
Chapter 6: Effects of Calibration on Knowledge-Aided STAP Approaches

6.1 Introduction

In this chapter we explore the effects of calibration on our proposed knowledge-aided STAP approaches. There are many types of calibration errors that exist in radar, but we choose to focus on angle-independent errors. These calibration errors are modeled as a per-channel gain and phase. We also explore a parametric bilinear model for target detection; the non-linear model is motivated by the need for resolving closely spaced moving targets in the angle/Doppler scene in the same range bin. Performance comparisons are made using the KASSPER I data set; we compare amongst: no calibration, known calibration, and estimated calibration.

6.2 Calibration on Clutter

Existing calibration techniques that we choose to focus on are sometimes referred to as “calibration on clutter” (cal-on-clutter) techniques. These approaches estimate a per-channel gain and phase from training data that contain clutter. Four cal-on-clutter techniques are outlined in [45]. We employ the channel pair detrending technique since it gives the best results when applied to the KASSPER I dataset.
In this section we will also introduce and demonstrate a parametric bilinear model for joint calibration and detection. This joint non-linear processing is shown, using the KASSPER I dataset, to reduce the detection performance losses compared to the other Bayesian hypothesis testing results.

### 6.2.1 Channel Pair Detrending

Channel pair detrending is outlined in [30] and [45]. The approach exploits the known phase difference induced by clutter in the range/Doppler scene between pairs of antennas to estimate pair-wise channel gain and phase errors. Here we provide a summary of the processing details.

The channel pair detrending begins by taking the CPI and grouping the data into matrices, \( Y_m \), for the \( m \)th channel where each column is the range bin data for a single pulse. Performing a Fourier transform on the rows of \( Y_m \) then produces a range-Doppler map, \( F_m \), for the \( m \)th channel. We can then estimate the relative phase and gain imbalances from the cross-correlation of channel pairs.

The cross-correlation is defined as

\[
C_m = F_m \odot F_{m-1}^* \tag{6.1}
\]

where we perform the cross-correlation only for adjacent pairs beginning with \( m = 2 \). Since noise dominates the range-Doppler map except in the Doppler region where mainbeam clutter resides, we limit our calibration estimation to these mainbeam clutter regions. The last step before estimating the calibration coefficients is removing the known phase ramp in \( C_m \) that results from the antenna spacings. This is performed as follows:

\[
C'_m = C_m \odot \Phi_m^* \tag{6.2}
\]
where
\[ \Phi_m(r, \bar{\omega}) = \frac{\bar{\omega}}{\beta} \cos \phi_{crab} + \sqrt{\eta^2 \sin^2 \phi_{crab} + \left(\frac{\bar{\omega}}{\beta}\right)^2 (\cos^2 \phi_{crab} - 1)} \]

\[ \beta = \frac{2v_p}{f_r d} \]

\[ \eta = \frac{d\sqrt{r^2 - h^2}}{r\lambda_c}, \]

and for implementation we replace \( h \) with the mean height of the ground, \( \mu_h \).

The relative channel errors can then be found as
\[ \hat{\epsilon}_{0,rel}(m) = \frac{\|F_m(:)\|_2}{\|F_{m-1}(:)\|_2} \exp \left\{ j \arg \left( \frac{1}{\tilde{L}\tilde{M}} \sum_i \sum_j C'_m(i,j) \right) \right\} \]
for \( m = 2, 3, \ldots, M \)

(6.4)

where \( \hat{\epsilon}_{0,rel}(1) = 1 \), \( \tilde{L} \) is the number of range bins used in the estimate, and \( \tilde{M} \) is the number of Doppler bins used in the estimate. The absolute channel errors can then be calculated as
\[ \hat{\epsilon}_0(m) = \hat{\epsilon}_{0,rel}(m) \hat{\epsilon}_0(m - 1) \]
for \( m = 2, 3, \ldots, M \)

(6.5)

where \( \hat{\epsilon}_0(m) = 1 \) for \( m = 1 \).

### 6.2.2 Parametric Bilinear Model

Here we present a parametric bilinear model for jointly performing calibration and detection for space-time adaptive processing. We hypothesize that the joint formulation will improve the performance of the Bayesian hypothesis test given in chapter 4. The non-linear processing of the test in Eq. 4.12 is shown to have the ability to resolve closely spaced moving targets in angle and Doppler at a single range bin.

The parametric bilinear model for “self-calibration” is given in [55] as
\[ z = \text{diag}(H\epsilon)Ax \]

(6.6)
Figure 6.1: Bayesian model visualized as a factor graph. Dense, black portion can be accelerated using fast inference algorithms. Red portion represents incorporation of prior knowledge from digital elevation map and platform kinematics. Blue portion represents incorporation of prior knowledge that moving targets are sparse in the scene.

for a known $H$ and $A$. In this setup the vector $\mathbf{x}$ is measured through the known matrix $A$; $A$ is then perturbed by structured but unknown gains of the form $\text{diag}(H\epsilon)$. This model can easily be adopted for modeling the unknown per-channel gains and phases present in STAP; per-channel gains with a sub-sampled Fourier transform measurement matrix similar to our STAP formulation is given as an example in [56]. Assuming that the components of the random vectors $\epsilon$ and $\mathbf{x}$ are independent and the likelihood function of the noiseless measurements $z$ is separable, the Bayesian model can be visualized as the factor graph shown in Fig. 6.1.

To complete the model we must choose the prior distributions for the random elements of $\epsilon$ and $\mathbf{x}$. For simplicity we keep the same priors for $\mathbf{x}$ as in chapter 4; we
will discuss this choice further in the section 6.3. For the prior on the elements of $\epsilon$ we use the approach presented in [73] which models an unknown phase. This allows us to use an informative prior for the gain and an uninformative prior for the phase of the per-channel calibration coefficients. We discuss the choice of values further in the following section.

6.3 Numerical Results

In this section we present numerical results that demonstrate the effect of calibration on our proposed knowledge-aided approaches. First we explore the effects of calibration on the regularized multi-model shrinkage (MMS-EN) approach. Then we explore the effects of calibration on the knowledge-aided Bayesian approach for estimating an interference covariance matrix. Finally, we explore the effects of calibration on the knowledge-aided Bayesian hypothesis testing for target detection and compare the results to the proposed parametric bilinear approach.

6.3.1 Metrics

We evaluate algorithm performance using two widely employed performance metrics: average SINR loss and receiver operating characteristic (ROC) curve. The ROC curve reveals detection behavior potentially obscured by the scalar SINR loss. SINR loss is computed for each normalized spatial/Doppler cell, $i$, and each range bin, $k$, using the true covariance, $R_k$, and the estimated covariance, $\hat{R}_k$,

$$\text{SINR Loss} = \frac{|a_i^H \hat{R}_k^{-1} a_i|^2}{|a_i^H \hat{R}_k^{-1} R_k \hat{R}_k^{-1} a_i||a_i^H R_k^{-1} a_i|}$$.  \hspace{1cm} (6.7)

The resulting loss is averaged across both $i$ and $k$ to yield a single scalar SINR metric and is reported in decibels.
An empirical ROC curve is computed using the true locations of moving targets present in the mainbeam; the true locations are provided in the KASSPER I dataset. We declare a potential detection to be a correct detection if a true target lies within one Fourier bin. We choose this procedure to form the ROC curves despite the non-uniformity of procedures in the literature; for example, some declare a true detection if there is merely a detection at the same range as a true target. The ROC curve is then constructed as probability of detection \( P_D \) versus probability of false alarm \( P_{FA} \) where \([59]\)

\[
P_D = \frac{\# \text{ of correct detections}}{\# \text{ of correct detections} + \# \text{ of missed detections}}
\]

\[
P_{FA} = \frac{\# \text{ of false positives}}{\# \text{ of false positives} + \# \text{ of correct omissions}}.
\]

The curve is traced by varying the detection test threshold value.

### 6.3.2 KASSPER I Dataset

The KASSPER I dataset includes both angle-dependent and angle-independent calibration errors; the angle-independent errors are known and documented. Further, the clairvoyant interference covariances provided with the data includes both the angle-dependent and -independent parameters. The angle-dependent errors are caused by the differences among the receive antenna patterns. Despite the presence of angle-dependent errors we still model the net calibration error as a single per-channel phase. We do not include a channel gain since for the KASSPER I dataset the gain is approximately one. We compare results for both the Bayesian and MMS-EN approaches using: no calibration, known angle-independent errors, and channel pair detrending estimated errors. For the Bayesian hypothesis testing approach we
also compare to the proposed parametric bilinear approach. The parametric bilinear approach does not perform well for estimating an interference covariance matrix; therefore, such a GLRT approach is excluded from the results.

We begin by looking at the effects of calibration on the MMS-EN approach. Results are shown in Fig. 6.2 for SINR loss versus number of training samples (panel (a)) and ROC curves for three sizes of training data (panels (b)-(d)). We can observe from Fig. 6.2 that the two calibration techniques outperform using no calibration at all. Further, the estimated calibration using the channel pair detrending technique results in lower SINR loss than using only the known channel errors, but the ROC curve results are nearly indistinguishable. We note that the detection performance of the proposed MMS-EN does not improve as more training data is used. We attribute this behavior to the heterogeneity and contamination of the training data. The MMS-EN estimate should asymptotically converge to the sample covariance matrix which, due to the heterogeneous and contaminated training data, is not a good estimate.

Next we explore the effects of calibration on the proposed Bayesian approach. First we look at the SINR loss and ROC curve performance when the Bayesian model is used to estimate an interference covariance matrix as in Eq. 4.10. These results are shown in Fig. 6.3. We can observe from Fig. 6.3 that the two calibration techniques reduce SINR losses compared to using no calibration at all, and the estimated calibration using the channel pair detrending technique gives the best SINR results.

Next we present the ROC curve results for the Bayesian approach using both the interference estimate in Eq. 4.10 and Bayesian hypothesis testing in Eq. 4.12 for detections; the results are also compared to the proposed parametric bilinear approach using the Bayesian hypothesis testing for detection. The first set of ROC curves shown
Figure 6.2: Average SINR loss and ROC curves for the proposed MMS-EN-P approach. Figures (a) shows the average SINR loss for the MMS-EN-P approach with the three different calibration techniques. Figures (b), (c), and (d) show the ROC curves for the MMS-EN-P approach with the three different calibration techniques using 5, 20, and 45 training samples, respectively.

in Fig. 6.4 limits the number of detections per range bin to one since the traditional linear whitened matched filter STAP cannot resolve closely spaced targets in angle and Doppler due to sidelobes of strong targets obscuring other potential targets; for comparison, the proposed non-linear Bayesian hypothesis testing is also constrained to only one detection per range bin even though it has the ability to resolve closely spaced movers. We can observe from Fig. 6.4 that the estimated calibration using
Figure 6.3: Average SINR loss and SINR losses at normalized spatial/Doppler pairs averaged across range for Bayesian covariance estimate using only the CUT. Figure (a) shows the average SINR loss for the Bayesian approach using the three calibration techniques. Figures (b), (c), and (d) show the SINR losses at normalized spatial/Doppler pairs averaged across range for the Bayesian covariance estimate using no calibration, known calibration, and estimated calibration coefficients, respectively. The horizontal lines in Figure (a) are the average value from Figures (b)-(d).

the channel pair detrending technique improves the detection performance of both the traditional whitened matched filter STAP approach as well as the non-linear Bayesian hypothesis testing. Further, the parametric bilinear approach has better detection performance than the other non-linear Bayesian hypothesis testing approach with
any type of calibration. We attribute this performance gain to the fact that the bilinear approach is able to absorb any modeling mismatches into the calibration parameters. The parametric bilinear approach is able to reduce the number of strong spurious peaks that results from model mismatches and hinder detection performance when the processing is limited to one target detection per range.

If we then consider allowing every angle/Doppler cell per range to be a potential detection instead of solely the largest response in the scene at each range, we get the results in Fig. 6.5. We can see from Fig. 6.5 that the non-linear approaches outperform the traditional linear approaches when we allow for multiple detections per range bin. However, it is important to note that the parametric bilinear approach does not perform as well in this scenario. We attribute this to the fact that our measurement
matrix, \( A \), is highly correlated (not i.i.d Gaussian) and the bilinear nature of the problem does not provide accurate approximate posterior variances [74]. Because the Bayesian hypothesis test relies on these approximate posteriors, the detection performance is reduced as a consequence of poor approximate posterior variance estimates.

### 6.4 Conclusions

In this chapter, we have explored the role of angle-independent calibration in moving target detection. The Bayesian approach in chapter 4 was extended to a parametric bilinear approach to jointly estimate a per-channel phase. Empirical results with the KASSPER I dataset suggest that an existing ad hoc preprocessing technique, channel pair detrending, provides superior detection results compared to
the bilinear Bayesian calibration. This performance comparison likely is a result of unreliable posterior variances caused by correlation in the system model matrix. Calibration preprocessing improved STAP detection results for all covariance estimators and target detectors considered. Additionally, the Bayesian detection approach was observed to provide super-resolution of moving targets within a single range, in contrast to the AMF detection test using any covariance matrix estimate.
Chapter 7: Conclusions

7.1 Summary

In this dissertation we have presented and demonstrated new approaches to ground moving target indicator radar that meet some of the challenges facing today’s radar systems. The proposed FDMA MISO architecture is able to provide joint SAR/STAP operation while constraining the data rate to that of a single receive antenna. The proposed knowledge-aided Bayesian STAP approach is able to incorporate approximate prior knowledge, reject targets in the training data, and resolve closely spaced movers using no training data. The proposed multi-model shrinkage estimator is also able to incorporate approximate prior knowledge while using amounts of training data less than the rank of the true covariance. Finally, we show the sensitivity of the proposed knowledge-aided approaches to calibration and demonstrate that a calibrated Bayesian model can provide enhanced resolution, allowing detection of multiple moving targets present within a single range bin.
7.2 Future Work

In this section we will outline some possible avenues for future work. While little preliminary work has been done for these ideas, experiences from previous work lead us to believe these ideas contain substance.

7.2.1 Optimize Bayesian Model for Target Detection

For our previous work on a Bayesian model for space-time adaptive processing, the model prior selection was optimized for estimation of an interference covariance matrix. It is possible that a modified model would result in better performance of the Bayesian hypothesis testing. For improved performance we propose the following modified model shown a factor graph in Fig. 7.1. The priors are then chosen as

\[
p(x_i, c_i, t_i) = p(x_i | c_i, t_i)p(c_i)p(t_i)
\]

(7.1)

Figure 7.1: Modified Bayesian model visualized as a factor graph.
\[
p(x_i|c_i, t_i) = \begin{cases} 
p(x_i|c_i = 0, t_i = 0) = \mathcal{CN}(x_i; \mu_0, \sigma_0^2) \\
p(x_i|c_i = 1, t_i = 0) = \mathcal{CN}(x_i; \mu_1, \sigma_1^2) \\
p(x_i|c_i = 0, t_i = 1) = \mathcal{CN}(x_i; \mu_0, \sigma_0^2) \\
p(x_i|c_i = 1, t_i = 1) = \mathcal{CN}(x_i; \mu_1, \sigma_1^2) \end{cases} \tag{7.2}
\]

\[
p_{0,0}(x_i) = p(x_i|c_i = 0, t_i = 0) = p(x_i|c_i = 0, t_i = 1)
\]

\[
p_{1,0}(x_i) = p(x_i|c_i = 1, t_i = 0)
= p(x_i|c_i = 0, t_i = 1) = p(x_i|c_i = 1, t_i = 1)
\tag{7.3}
\]

\[
p(x_i|c_i, t_i) = (1 - c_i)(1 - t_i)p_{0,0}(x_i) + c_i(1 - t_i)p_{1,0}(x_i) + (1 - c_i)t_ip_{1,0}(x_i) + c_i(t_i)p_{1,0}(x_i)
= (1 - c_i)(1 - t_i)p_{0,0}(x_i) + [c_i(1 - t_i) + t_i(1 - c_i) + c_i(t_i)]p_{1,0}(x_i)
= (1 - c_i)(1 - t_i)p_{0,0}(x_i) + (c_i + t_i - c_i t_i)p_{1,0}(x_i)
\tag{7.4}
\]

\[
p(c_i) = \begin{cases} 
1 \quad \text{w.p.} \quad \alpha_i \\
0 \quad \text{w.p.} \quad (1 - \alpha_i) \end{cases} \tag{7.5}
\]

\[
p(t_i) = \begin{cases} 
1 \quad \text{w.p.} \quad \lambda \\
0 \quad \text{w.p.} \quad (1 - \lambda) \end{cases} \tag{7.6}
\]

The final prior passed into GAMP is then calculated as follows

\[
m_1 = (1 - \alpha_i)(1 - \lambda_i)\mathcal{CN}(x_i; \mu_0, \sigma_0^2) + (\alpha_i + \lambda - \alpha_i \lambda)\mathcal{CN}(x_i; \mu_1, \sigma_1^2)
\tag{7.7}
\]

\[
\gamma_i = \alpha_i + \lambda - \alpha_i \lambda
\tag{7.8}
\]

\[
m_1 = (1 - \gamma_i)\mathcal{CN}(x_i; \mu_0, \sigma_0^2) + \gamma_i\mathcal{CN}(x_i; \mu_1, \sigma_1^2)
\tag{7.9}
\]

This form is desirable since it takes the form of a Gaussian mixture, and the GAMP algorithm has existing tools to handle this input prior. We believe that using a single Gaussian for both clutter and targets can ameliorate the confusion of the Bayesian hypothesis test when multiple Gaussian components are used to describe the clutter. Further, using a low variance Gaussian instead of a mass at zero to describe the zero components of the angle/Doppler scene may reduce the false alarms due to model mismatches.
7.2.2 Finite Sample Error Bounds

Our previous work on multi-model shrinkage estimators for STAP has led us to a different but similar estimator that may admit provable finite sample error bounds. Instead of solving for the optimal scaling factors by minimizing the expected squared error, we consider the following cost

\[
\hat{\lambda} = \arg\min_{\lambda \geq 0} \|A\lambda - s\|_F^2 + \|\beta \odot \lambda\|_1
\]

(7.10)

where \(A\) is a matrix whose columns are the shrinkage models, \(s\) is the vectorized sample covariance matrix, and the vector \(\beta\) consists of regularizing weights. The form of this cost may allow us to use the “M-estimator” framework presented in [51] to prove error bounds on the estimated scaling factors \(\hat{\lambda}\). Suppose that the true covariance is expressible as \(A\lambda\) with \(\lambda \in \mathbb{C}^p\) drawn from independent Bernoulli-Gauss random variables; let \((1 - \alpha_i)\) be the probability that \(\lambda_i\) is non-zero, and let \(\bar{\lambda}\) be the conditional mean of \(\lambda_i\). We conjecture that, with probability exceeding \(1 - 5/n\), the error satisfies

\[
\|\hat{\lambda} - \lambda^*\|_2 \leq c_1 \sqrt{\frac{\log n p}{n}\|\alpha\|_2 \left\{ \frac{m}{n} \sigma_T^2 + (p - \|\alpha\|_1)\bar{\lambda} \right\}.
\]

(7.11)

Here, \(\sigma_T\) is the mean thermal noise power, \(n\) is the number of snapshots, and the covariance matrix is \(m \times m\). Interestingly, this finite-sample bound using structure in \(R\) is loose, offering no improvement in the finite sample bound on the sample covariance as a function of the effective rank, \(\text{tr}(R)/\|R\|_2\) [17]. The possibility of obtaining a tighter bound remains an open question.
7.2.3 MAP Covariance Estimation

Suppose $R$ is invertible, as is guaranteed by non-zero thermal noise, and consider i.i.d. snapshots,

$$Y_i \sim \mathcal{CN}(0, R). \quad (7.12)$$

Let $P$ denote the precision matrix, $P = R^{-1}$, and $S$ denote the sample covariance matrix, $S = \frac{1}{n}YY^H$. As in the multi-model shrinkage approach in chapter 5, we model $R$ as a linear combination of thermal noise and rank-1 components,

$$R = \theta_1 + \sum_{k=2}^{p} \theta_k a_k a_k^H, \quad (7.13)$$

where $\theta \in \mathbb{R}^p$ are unknown non-negative parameters. To express prior belief of an approximate digital elevation map and platform kinematics (velocity, crab angle, antenna spacings) we follow the approaches proposed in chapters 4 and 5. Prior probabilities, $\phi_k(\theta_k)$, are independently placed on the thermal noise power, $\theta_1$, and the radar cross section, $\{\theta_k\}$, at angle/Doppler location $k = 2, \ldots, p$. For tractability, the prior probabilities are modeled as independent, gaining simplicity at the potential cost of forfeited model structure. The resulting likelihood ratio test for target detection, as in Eq. 2.24, in this case includes integration over $\theta$; for example, the numerator term is written

$$p(H_1|Y, \theta)p(\theta|Y)d\theta. \quad (7.14)$$

In the spirit of a generalized likelihood ratio test, we simplify the LRT by replacing $\theta$ with its maximum aposterior probability (MAP) estimate. Substitution of $\theta_{\text{map}}$ into
the LRT returns the problem to the Gauss-Gauss test. Thus, we seek

\[ \hat{\theta}_{\text{map}} = \arg \max_{\theta} f_{Y|\Theta}(y|\theta) g_{\Theta}(\theta) \] (7.15)

\[ = \arg \min_{\theta} - \log f_{Y|\Theta}(y|\theta) g_{\Theta}(\theta) \] (7.16)

\[ = \arg \min_{\theta} \text{tr}(P_{\theta} S) - \log \det(P_{\theta}) + \left\{ \sum_{k=1}^{p} - \log \phi_k(\theta_k) \right\} \] (7.17)

\[ = \arg \min_{\theta} \text{tr}(R_{\theta}^{-1} S) + \log \det(R_{\theta}) + \rho(\theta) \] (7.18)

where

\[ R_{\theta} = \theta_1 + \sum_{k=2}^{p} \theta_k a_k a_k^H \] (linear model) (7.19)

\[ P_{\theta} = R_{\theta}^{-1} \] (precision matrix) (7.20)

\[ \rho_{\theta} = \sum_{k=1}^{p} - \log \phi_k(\theta_k) \] (regularizer) (7.21)

and Eq. 7.17 follows from Eqs. A1-A6. For example, \( \rho(\theta) \) may be an L-p penalty on \( \theta \).

### 7.2.4 Fast Computation of MMS-EN: Parallel Architecture

Given the preliminary success of our proposed MMS-EN approach for STAP we wish to accelerate the estimator through parallelization; fast computation and implementation is desirable for real time applications. The most computationally burdensome aspect of the estimator is the calculation of the \( \tilde{A} \) matrix. The calculation of \( \tilde{A} \) is dominated by repeated vector inner products. Computation can easily be accelerated through clever implementation in C code making sure to prioritize speed over memory usage (many values can be stored and used again instead of being recomputed). Additionally, these vector inner products can be sped up with hardware acceleration and parallelization using GPUs. Further work can evaluate the power
burden of such processing if there is a power limit constraint of the system implementing the approach.

7.2.5 Evaluation of Techniques with Measured Data

The work in this dissertation shows the potential of knowledge-aided approaches to outperform existing state-of-the-art techniques, but the proposed approaches have only been validated using the KASSPER I dataset. A contribution to the STAP community could come from performing the same comparisons in this dissertation using different data sets, preferably measured data sets. This would further validate our proposed techniques as well as shed light on what techniques are best suited for various operating conditions.

7.2.6 Empirical Bayes SAR GMTI

Existing Bayesian models for SAR GMTI employ a hierarchical model which is computationally complex [52]. Using the tool set of generalized approximate message passing it seems possible to rework parts of the hierarchical Bayesian model in [52] into an empirical Bayes model which can be accelerated with GAMP. The empirical Bayes model would be functionally similar to the change detection approach in [5].

7.2.7 Joint SAR GMTI/Refocusing

Building on the proposed empirical Bayes SAR GMTI technique, it may be possible to use a parametric bilinear model to jointly detect moving targets and refocus them in the SAR image. We anticipate that the key challenge will come from target acceleration during an extended CPI.
Bibliography


Appendix A: Constrained Maximum Likelihood Estimators

In this appendix we consider maximum likelihood estimators (MLEs) of the unknown covariance matrix from only secondary data under several assumed constraints on the positive semi-definite covariance matrix, $R$ or the precision matrix, $P = R^{-1}$. Note that the MLE provides a principled estimation approach with good asymptotic properties, but no guarantees for finite sample estimation performance; further, the aim in any case is detection performance. With these caveats noted, we proceed to consider a suite of covariance estimators, summarized in Table A.1.

A.1 Maximum Likelihood Estimate (Sample Covariance)

Consider maximum likelihood estimators of an unknown covariance matrix from independent and identically distributed (i.i.d.) samples. Starting from the likelihood of jointly complex circular Gaussian zero-mean data vectors, we have

$$f_Y(y) = \pi^{-p} \det(P) \exp\{-y^H P y\}$$  \hspace{1cm} (A.1)
where $y \in \mathbb{C}^p$ and $P > 0$ is the precision matrix. For i.i.d. draws $Y = [y_1, y_2, \ldots, y_n]$, the likelihood is

$$\pi^{-np}|P|^n \exp\{-\text{tr}(Y^H PY)\}$$

(A.2)

$$= \pi^{-np}|P|^n \exp\{-\text{tr}(PYY^H)\}$$

(A.3)

$$= \pi^{-np}|P|^n \exp\{-n\text{tr}(PS)\}$$

(A.4)

where $S = \frac{1}{n}YY^H$ is the sample covariance matrix, which has Wishart distribution with $n - 1$ degrees of freedom. The negative log likelihood is then

$$-np \log(\pi) - n \log \det P + n\text{tr}(PS).$$

(A.5)

Thus, we seek

$$\arg \inf_{P > 0} \text{tr}(PS) - \log \det P.$$

(A.6)

The trace is affine, and the negative log det is convex; also, the set of positive definite $P$ is convex (but neither closed nor bounded); so, the optimization is convex. This objective is convex over the open and unbounded convex set, $P > 0$.

Continuing, use the eigen decompositions following [35]:

$$S = VDV^H$$

(A.7)

$$P = \Phi \Lambda \Phi^H$$

(A.8)

where $V$ and $\Phi$ are unitary, and $D, \Lambda$ are diagonal non-negative. So, we have

$$\text{tr}(PS) - \log \det P$$

(A.9)

$$= \text{tr}(\Phi \Lambda \Phi^H VDV^H) - \log \det \{\Phi \Lambda \Phi^H\}$$

(A.10)

$$= \text{tr}(DV^H \Phi \Lambda \Phi^H V) - \log \det \{\Lambda\}$$

(A.11)

$$\geq d^T \lambda - 1^T \log \lambda$$

(A.12)
where $D = \text{diag}(d)$ and $\Lambda = \text{diag}(\lambda)$. Further, the equality holds if and only if $\Phi = V$ (by von Neumann’s trace inequality).

For $n \geq p$, the sample covariance is non-singular with probability one, and the unconstrained solution is $P_\ast = S^{-1}$, where $S$ is the sample covariance matrix. For this unconstrained case, the expected SINR loss is less than 3 dB for $n > 2p$ training bins (RMB rule) \cite{66}, where

$$\text{Loss}_{\text{SINR}}(R, \hat{R}, a) = \frac{|a^H \hat{R}^{-1} a|^2}{a^H \hat{R}^{-1} \hat{R}^{\ast -1} a a^H R^{\ast -1} a}.$$ \hfill (A.13)

Note that for recursive “on-line” computation, the matrix inversion lemma provides for the $O(p^2)$ rank-one update of the inverse of the sample covariance matrix for $n \geq p$:

$$S_n = \frac{1}{n} \sum_{i=1}^{n} y_i y_i^H$$ \hfill (A.14)

$$P_{n+1} = S_{n}^{-1} = \text{inv} \left( \frac{n}{n+1} S_n + \frac{1}{n+1} y_{n+1} y_{n+1}^H \right)$$ \hfill (A.15)

$$= (n+1) \text{inv} \left( nS_n + yy^H \right)$$ \hfill (A.16)

$$= (n+1) \left\{ \frac{1}{n} P_n - \frac{1}{n+1} P_n y_{n+1} y_{n+1}^H P_n \right\}.$$ \hfill (A.17)

### A.2 Convex Covariance Constraints

Considering the items 1-5 in Table A.1, we can minimize the convex cost in Eq. A.6 over a convex constraint set. Each constraint in items 2-5 is a closed convex set. Toeplitz (item 2) \cite{27}, circulant (item 3) \cite{23}, and persymmetric (item 4) \cite{53} constraint sets each admit a simple nearest point (in Frobenius norm) projection operator. Van Trees \cite{81} gives the nearest persymmetric matrix to $S$ as
Re\{TST^H\} = \frac{1}{2}(S + JS^*J), where \( T \) is a unitary matrix. The MLE for a persymmetric covariance matrix was given, for example, by Nitzberg [53].

### A.3 Thermal Noise and a Low-Rank Constraint

Now consider the structured covariance

\[
R = \sigma^2 I + R_c, \quad (A.18)
\]

with \( R_c \geq 0 \) and \( \sigma^2 \geq \sigma^2_{min} > 0 \). The interference is due to both thermal noise, represented by the scaled identity, and an unknown colored interference, represented by the unknown positive semi-definite \( R_c \). Further, the Brennan rule [86] approximates the rank of \( R_c \):

\[
\text{rank}(R_c) \approx \min\{N_c, \left\lceil (M - 1)\beta + N \right\rceil\}, \quad (A.19)
\]

where \( M \) is the number of pulses, \( N \) is the number of antennas, and \( N_c \) is the number of contributing clutter cells. Here, \( \beta = 2v_pT/d \) is the slope of the clutter ridge, with \( v_p \) denoting platform speed, \( T \) denoting the pulse repetition period, and \( d \) denoting the inter-element spacing. For integer \( \beta \), the rule arises from the repeated measurements obtained from different array elements.

Thus, an upper bound may be imposed on the rank of the interference: \( \text{rank}(R_c) \leq r \). Observe from Eq. A.12 that the covariance structure and rank condition combine to yield

\[
\lambda_1 \leq \lambda_1 \leq \cdots \leq \lambda_r \leq \frac{1}{\sigma^2} \quad \text{and} \quad \lambda_{r+1} = \cdots = \lambda_p = \frac{1}{\sigma^2} \quad (A.20)
\]

Following [35], express these constraints as linear equalities or inequalities, with the aim that the non-convex rank constraint on minimizing the negative log likelihood nonetheless admits a convex problem sharing the same minimum.
1. Monotone ordering:

$$
\begin{bmatrix}
1 & -1 & 0 & \cdots & 0 \\
0 & 1 & -1 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 1 & -1 \\
0 & 0 & \cdots & 0 & 0 & -1 \\
\end{bmatrix} \lambda = U \lambda \leq 0 \quad (A.21)
$$

2. Bounded variance:

$$
\epsilon \leq \lambda \iff -\lambda \leq -\epsilon 
$$

3. Equality among noise-space eigenvalues:

$$
\begin{bmatrix}
0 & 0 \\
0 & I_{p-r} \\
\end{bmatrix} \lambda = E \lambda = \sigma^2 \begin{bmatrix}
0 \\
1_{p-r} \\
\end{bmatrix} \quad (A.23)
$$

4. Noise floor on eigenvalues:

$$
\lambda \leq \frac{1}{\sigma^2} 1_p \quad (A.24)
$$

5. Lower bound on noise floor:

$$
\sigma^2 \geq \sigma^2_{\text{min}}. \quad (A.25)
$$
Combine these constraints for the constrained MLE for unknown noise floor and unknown precision matrix:

\[
\arg \min_{\{P, \sigma^2\}} \quad d^T \lambda - 1_p^T \log \lambda
\]

subject to

\[
\begin{bmatrix} U^T - I & I \end{bmatrix}^T P \lambda \leq \begin{bmatrix} 0 \\ -\epsilon 1_p \\ \frac{1}{\sigma^2} 1_p \end{bmatrix}
\]

\[
E \lambda = \sigma^2 \begin{bmatrix} 0 \\ 1_p \end{bmatrix}
\]

\[
\sigma^2 \geq \sigma_{\text{min}}^2.
\]

This is, again, a convex problem. Here, Kang et al. [35] used the Karush-Kuhn-Tucker (KKT) conditions to arrive at a closed-form solution to Eqs. A.26-A.29:

\[
\lambda_* = \begin{cases} \min \left( \frac{1}{\sigma_{\text{min}}^2}, \frac{1}{d_i} \right) & i = 1, 2, \ldots, r \\ \frac{1}{\sigma_{\text{min}}^2} & i = r + 1, r + 2, \ldots, p \end{cases}
\]

where

\[
\sigma_*^2 = \begin{cases} \frac{(d_{r+1} + \cdots + d_p)}{p-r} & \text{if } \geq \sigma_{\text{min}}^2 \\ \sigma_{\text{min}}^2 & \text{else} \end{cases}
\]

From Eq. A.12, the eigenvectors of \( P_* \) are the eigenvectors of \( S \). The solution is reminiscent of water filling.

### A.4 Clutter Subspace Constraint

Bresler’s approach [15] assumes \( R = ABA^H + \sigma^2 I \), for known \( N \times r \) matrix \( A \). Bresler derives the joint MLE of both \( R_c \) and \( \sigma^2 \), applying ML invariance principle to arrive at \( \hat{R} \). Thus, specification of \( A \) is equivalent to knowledge of the clutter subspace, \( \text{Range}(R_c) \). Brelser works with \( R \), and the non-convex optimization requires merely a discrete one-parameter search.
Note that Böhme’s approach [11],

\[
\hat{\sigma}_{Bö}^2 = \frac{1}{N-r} \text{tr} \left( P_A^{-1} S \right) \tag{A.32}
\]

\[
\hat{R}_{Bö} = \text{pinv}(A)(S - \hat{\sigma}_{Bö}^2 I)\text{pinv}(A)^H, \tag{A.33}
\]
is simpler than Bresler’s optimization steps (which include a search over a single integer parameter) and gives Bresler’s result if \( \hat{R}_{Bö} \geq 0 \). Also, Schmidt’s approach [72], while ad hoc, is aimed at the same goal and is simply given by

\[
\hat{R}_{Sch} = \text{pinv}(A)S_r\text{pinv}(A)^H, \tag{A.34}
\]

where \( S_r \) is the (Frobenius norm) projection of the sample covariance matrix to rank \( r \).

For integer \( \beta \), constraint item 14 in Table A.1 is used in KB-GLRT [25], where the range of \( A \) is found using any \( r \) steering vectors, where \( r \) is determined by \( \beta \) in Eq. A.19. But, for non-integer \( \beta \), the MLE is apparently intractable [25]; a simple surrogate is to use the clutter space from the sample covariance matrix. ¹

The KB-GLRT exploits knowledge of the clutter ridge positions [25]. The KB-GLRT is a quasi-GLRT in that the MLE is not jointly computed over the parameters; instead, the estimated covariance is separately computed using only the secondary data and is substituted into the log-likelihood ratio test derived for known covariance. But, the receiver operating characteristic (ROC) curve for the true GLRT is empirically observed to be nearly identical to the quasi-GLRT at \( K \) given by Brennan rule, and less than 1 dB separates the two at one-half Brennan rule. The KB-GLRT is seen

¹Oddly, in [25], the Monte Carlo result is actually better for \( \beta = 0.9 \) and \( \beta = 1.1 \) using the sample matrix clutter subspace than for \( \beta = 1 \) using known steering vector samples to span clutter subspace. This is curious; compare Figures 5 and 6 to Figure 4 in [25]. This behavior may be simply due to the single, specific choice of boresight angle and normalized velocity 0.01, in relation to the sidelobe structure.
in simulation to be only 1 to 2 dB better than sample matrix inversion at \( n \) equal to the Brennan rule [25]; however, for \( n \) much less than Brennan rule, very large gains are observed.

### A.5 Convex Approach for Subspace Knowledge

The non-convex approach in [15] can be replaced by an equivalent convex program.

Let \( \mathbf{R} = \mathbf{ABA}^H + \sigma^2 \mathbf{I} \), as previously, where \( \mathbf{A} \) is \( p \)-by-\( r \) with known range. By the matrix inversion lemma,

\[
\mathbf{P} = (\mathbf{ABA}^H + \sigma^2 \mathbf{I})^{-1} = \frac{1}{\sigma^2} \mathbf{I} + \frac{1}{\sigma^2} \mathbf{A}(\mathbf{A}^H \mathbf{A} + \sigma^2 \mathbf{B}^{-1})^{-1} \mathbf{A}^H. \tag{A.35}
\]

Thus, the assumption \( \mathbf{R} = \mathbf{ABA}^H + \sigma^2 \mathbf{I} \) implies that \( \frac{1}{\sigma^2} \mathbf{I} - \mathbf{P} \) is in the range of \( \mathbf{A} \). Let \( \mathbf{P}_A \) be the orthogonal projection on the range of \( \mathbf{A} \). Then, we have that the following convex constraints are equivalent to \( \mathbf{R} = \mathbf{ABA}^H + \sigma^2 \mathbf{I} \):

\[
\mathbf{P}_A \left( \frac{1}{\sigma^2} \mathbf{I} - \mathbf{P} \right) = 0 \tag{A.36}
\]

\[
\frac{1}{\sigma^2} \mathbf{I} - \mathbf{P} \geq 0 \tag{A.37}
\]

\[
\mathbf{P} > 0 \tag{A.38}
\]

Thus, we have the convex program,

\[
\arg \min_{\mathbf{P}, \sigma^2} \quad \text{tr} \{ \mathbf{PS} \} - \log \det \mathbf{P} \tag{A.39}
\]

subject to

\[
\mathbf{P}_A \left( \frac{1}{\sigma^2} \mathbf{I} - \mathbf{P} \right) = 0 \tag{A.40}
\]

\[
\frac{1}{\sigma^2} \mathbf{I} - \mathbf{P} \geq 0 \tag{A.41}
\]

\[
\mathbf{P} > 0. \tag{A.42}
\]

This convex optimization task is a maxdet problem [12, 13].
A.6 Prior Knowledge via a Nominal Covariance

Let $R_0$ be some nominal covariance matrix based on prior belief about the problem; and, constrain $P$ to approximately invert $R_0$:

$$\|PR_0 - I\| \leq \delta, \quad P > 0.$$  \hfill (A.43)

Here, $PR_0$ is constrained to be in the spectral-norm $\delta$-ball of the identity matrix; recall that the spectral norm is the largest singular value. De Maio et al. [24] expressed Eq. A.43 as a linear matrix inequality and combined with the convex cost in Eq. A.6 to pose the convex optimization they dubbed the knowledge-aided unstructured estimator (KAUE):

$$\arg\min_{\{P,\sigma^2\}} \{ \text{tr}(PS) - \log \det P \}$$  \hfill (A.44)

subject to

$$\begin{bmatrix} I & Q \\ Q^H & I \end{bmatrix} \succeq 0, \quad Q = \frac{1}{\delta}(PR_0 - I)$$  \hfill (A.45)

$$P > 0.$$  \hfill (A.46)

Note that practical choice for $\delta$ in the spectral norm ball constraint, Eq. A.43, is entirely unclear.

The authors in [24] further inserted two additional convex constraints to model the structure of a thermal noise term, $R = R_c + \sigma^2 I$, with upper and lower bounds on the noise power:

$$\frac{1}{\sigma^2} I - P \succeq 0$$  \hfill (A.47)

$$\sigma^2_{\min} \leq \sigma^2 \leq \sigma^2_{\max}$$  \hfill (A.48)
The maxdet cost with the LMI constraints and these thermal noise constraints was dubbed the knowledge-aided structured estimator (KASE) in [25, Eq. 12] and solved with the SDPT3 toolbox\(^3\). For known \(\sigma^2\) and \(R = R_c + \sigma^2 I\), the optimization of Eq. A.6 is the so-called fast maximum likelihood (FML) [77].

Obviously, this convex programming formulation can be extended to include an additional constraint that \(P \in \mathcal{K}\) for any convex set, \(\mathcal{K}\), such as: Toeplitz, persymmetric, circulant, or bounded condition number. Then, we have the convex program

$$\arg \min_{\{P, \sigma^2\}} \left\{ \text{tr}(PS) - \log \det P \right\}$$

subject to

\[
\begin{bmatrix}
I & Q \\
Q^H & I
\end{bmatrix} \succeq 0, \quad Q = \frac{1}{\delta}(PR_0 - I) \tag{A.50}
\]

\[
\frac{1}{\sigma^2} I - P \succeq 0 \tag{A.51}
\]

\[
\sigma_{\min}^2 \leq \sigma^2 \leq \sigma_{\max}^2 \tag{A.52}
\]

\[
P_{\perp} \left( \frac{1}{\sigma^2} I - P \right) = 0 \tag{A.53}
\]

\[
P > 0 \tag{A.54}
\]

\[
P \in \mathcal{K}. \tag{A.55}
\]

A.7 Sparse Methods

Graphical lasso (GLASSO) [26] regularizes the ML cost for the precision matrix via an \(\ell_1\)-norm penalty, promoting sparsity on the off diagonal entries of the precision matrix

\[
\hat{P} = \arg \min_P \left\{ \text{tr}(PS) - \log \det P + \alpha \|P\|_1 \right\}. \tag{A.56}
\]
Here, the matrix 1-norm is the entry-wise sum of absolute values. Matlab code available\(^4\).

The constrained \(\ell_1\)-minimization for inverse matrix estimation (CLIME) \(^{19}\) is a simpler convex program, with lower computational complexity and provable performance results:

\[
\hat{P} = \arg \min_P \|P\|_1 \text{ subject to } \|PS-I\|_\infty \leq \delta, \quad \delta = C \sqrt{(\log p)/n}, \quad C \propto \|P\|_\infty
\]

(A.57)

where the infinity norm is the entry of maximum absolute value. Compared to GLASSO, the likelihood cost is replaced by the infinity norm on \((PS-I)\). The spectral-norm convergence rate is \(s \sqrt{(\log p)/n}\) for \(s\)-sparse precision matrix. C code available\(^5\) (with R wrapper) and R code is available\(^6\).

\(^4\)http://statweb.stanford.edu/~tibs/glasso/ or http://www.ece.ubc.ca/~xiaohuic/code/glasso/glasso.htm

\(^5\)http://www.princeton.edu/~hanliu/software.html

Table A.1: Constrained ML Estimators for Covariance

<table>
<thead>
<tr>
<th>Item</th>
<th>Constraint</th>
<th>Acronym</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unstructured, $R$ positive semi-definite</strong></td>
<td>$R &gt; 0$ (“unconstrained”)</td>
<td>SMI</td>
<td>RMB, 1974</td>
</tr>
<tr>
<td>1</td>
<td>Toeplitz</td>
<td>Furhmann, 1991</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>circulant</td>
<td>CLR, 1998</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>persymmetric</td>
<td>Nitzberg, 1980</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>condition number</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>close to prior</td>
<td>KAUE</td>
<td>DDLF, 2009</td>
</tr>
<tr>
<td>6</td>
<td>$|R^{-1}R_0 - I|_{\ell_2} \leq \delta$</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Structured, $R = R_c + \sigma^2 I = P^{-1}$</strong></td>
<td>lower-bound thermal noise</td>
<td>FML</td>
<td>Steiner &amp; Gerlach, 2000</td>
</tr>
<tr>
<td>7</td>
<td>$\sigma^2 \geq \sigma^2_{\min}$</td>
<td>G-KASE</td>
<td>DDLF, 2009</td>
</tr>
<tr>
<td>8</td>
<td>bounded thermal noise</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>upper-bounded clutter rank</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>bounded clutter rank and known $\sigma^2$</td>
<td>RCML</td>
<td>KMR, 2014</td>
</tr>
<tr>
<td>11</td>
<td>bounds on clutter rank and thermal noise</td>
<td>RCML-LB</td>
<td>KMR, 2014</td>
</tr>
<tr>
<td>12</td>
<td>bounded thermal noise and close to prior</td>
<td>G-KASE</td>
<td>DDLF, 2009</td>
</tr>
<tr>
<td>13</td>
<td>convex constraint; bounded thermal noise; prior $R \in \mathcal{K}$ (any sets 2-5 above)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>clutter subspace</td>
<td>KB</td>
<td>Bresler, 1988</td>
</tr>
<tr>
<td>15</td>
<td>clutter subspace; bounded noise; prior $R_c = ABA^H$, $A$ is N-by-r $\sigma^2 \geq 0$</td>
<td>DFW, 2005</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>clutter subspace; bounded noise; prior constraint $R_c = ABA^H$, $A$ is N-by-r</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix B: Message Passing Derivations

In this appendix we derive the messages being passed around the proposed factor graph shown in Fig. 4.1. This will be necessary for implementing the proposed Bayesian approach. The messages entering the left side of the graph through the conditional prior will be the priors used to initialize the GAMP iteration. The messages entering the right side of the graph coming from the GAMP iteration should take the form of a Bernoulli random variable. These messages will then be used to calculate the posterior for the clutter indicator variables.

B.1 Messages for GAMP Iteration

This section will be used to derive the messages entering the GAMP block. The following figure will be used for reference in the derivation.

\[ \begin{array}{c}
\text{x}_i \\
\text{p(x}_i|c_i) \\
\text{m}_2 \\
m_1 \\
\text{c} \\
p(c) \\
\end{array} \]

Figure B.1: Messages input into the GAMP block.

The messages that we will be deriving are \( m_1 \) and \( m_2 \) shown in the figure above in blue. Using the sum-product algorithm it is seen that the message \( m_1 \) is just the pdf
represented by the factor \( p(c_i) \). Since that pdf is a Bernoulli distribution, we will pass just one variable that completely describes the distribution. Therefore the message is as follows:

\[
m_1 = \alpha_i, \tag{B.1}\]

The message \( m_2 \) is then calculated as follows:

\[
m_2 = (1 - \alpha_i)p_0(x_i) + \alpha_i p_1(x_i) \tag{B.2}\]

\[
= (1 - \alpha_i) [(1 - \lambda)\delta(x_i) + \lambda \mathcal{CN}(x_i; \mu_{\text{mov}}, \sigma_{\text{mov}}^2)] + \alpha_i \sum_{n=1}^{3} w_n \mathcal{CN}(x_i; \mu_n, \sigma_n^2). \tag{B.3}\]

Then by rearranging terms and redefining variables,

\[
= (1 - \alpha_i)(1 - \lambda)\delta(x_i) + \alpha_i \sum_{n=1}^{3} w_n \mathcal{CN}(x_i; \mu_n, \sigma_n^2) + (1 - \alpha_i)\lambda \mathcal{CN}(x_i; \mu_{\text{mov}}, \sigma_{\text{mov}}^2) \tag{B.4}\]

\[
= (1 - \gamma_i)\delta(x_i) + \gamma_i \sum_{n=1}^{4} q_n \mathcal{CN}(x_i; \mu_n, \sigma_n^2) \tag{B.5}\]

where,

\[
\gamma_i = \lambda + \alpha_i - \lambda \alpha_i \quad q_n = \frac{\alpha_i w_n}{\gamma_i} \text{ for } n = 1, 2, 3 \tag{B.6}\]

\[
q_4 = \frac{(1 - \alpha_i)\lambda}{\gamma_i}, \quad \mu_4 = \mu_{\text{mov}}, \text{ and } \sigma_4^2 = \sigma_{\text{mov}}^2.
\]

This form is useful because we can use it as the prior for the input of GAMP. We will then let GAMP run until it has converged and then calculate the messages leaving this iteration. The derivations of the messages leaving the GAMP side of the factor graph are presented in the following section.

**B.2 Messages Leaving after GAMP Iteration**

In this section we will derive the messages leaving the GAMP portion of the factor graph. These messages can be used to calculate posteriors on the clutter indicator.
variables $c_i$. We will derive the messages $m_3$ and $m_4$ which are shown in the figure below.

![Diagram](image.png)

Figure B.2: Messages output from the GAMP block.

We will first find the message $m_3$ which is leaving the GAMP block. This message is easily found using the definitions in the GAMP summary. Therefore, the message is as follows:

$$m_3 = \mathcal{CN}(x_i; \hat{r}, \mu^r). \quad \text{(B.7)}$$

The message $m_4$ is then calculated as follows:

$$m_4 = \int x_i \mathcal{CN}(x_i; \hat{r}, \mu^r) \left[ (1 - c_i) p_0(x_i) + c_i p_1(x_i) \right] \quad \text{(B.8)}$$

$$= (1 - c_i) \int x_i \mathcal{CN}(x_i; \hat{r}, \mu^r) \left[ (1 - \lambda) \delta(x_i) + \lambda \mathcal{CN}(x_i; \mu_{\text{mov}}, \sigma^2_{\text{mov}}) \right] +$$

$$c_i \int x_i \mathcal{CN}(x_i; \hat{r}, \mu^r) \sum_{n=1}^{3} w_n \mathcal{CN}(x_i; \mu_n, \sigma^2_n) \quad \text{(B.9)}$$

$$= (1 - c_i) \left[ (1 - \lambda) \mathcal{CN}(0; \hat{r}, \mu^r) + \lambda \int x_i \mathcal{CN}(x_i; \hat{r}, \mu^r) \mathcal{CN}(x_i; \mu_{\text{mov}}, \sigma^2_{\text{mov}}) \right] +$$

$$c_i \sum_{n=1}^{3} w_n \int x_i \mathcal{CN}(x_i; \hat{r}, \mu^r) \mathcal{CN}(x_i; \mu_n, \sigma^2_n). \quad \text{(B.10)}$$

We will focus on one term at a time for simplicity sake; first, using the derivation of the product of two complex Gaussian distributions we find that:

$$\lambda \int x_i \mathcal{CN}(x_i; \hat{r}, \mu^r) \mathcal{CN}(x_i; \mu_{\text{mov}}, \sigma^2_{\text{mov}}) = \lambda S_0 \int x_i \mathcal{CN}(x_i; \mu_0, \Sigma_0) = \lambda S_0 \quad \text{(B.11)}$$
\[ S_0 = \frac{1}{\pi(\mu^r + \sigma_{\text{mov}}^2)} \times \exp \left\{ -\frac{1}{\mu^r + \sigma_{\text{mov}}^2} [(\text{Re}(\hat{r}) - \text{Re}(\mu_{\text{mov}}))^2 + (\text{Im}(\hat{r}) - \text{Im}(\mu_{\text{mov}}))^2] \right\}. \] (B.12)

Now focusing on the next term,
\[
c_i \sum_{n=1}^{3} w_n \int_x \mathcal{CN}(x_i; \hat{r}, \mu^r) \mathcal{CN}(x_i; \mu_n, \sigma_n^2) = c_i \sum_{n=1}^{3} w_n S_n \int_x \mathcal{CN}(x_i; \hat{\mu}_n, \hat{\Sigma}_n)
= c_i \sum_{n=1}^{3} w_n S_n
\] (B.13)

\[
S_n = \frac{1}{\pi(\mu^r + \sigma_n^2)} \exp \left\{ -\frac{1}{\mu^r + \sigma_n^2} [(\text{Re}(\hat{r}) - \text{Re}(\mu_n))^2 + (\text{Im}(\hat{r}) - \text{Im}(\mu_n))^2] \right\}. \] (B.14)

Now using these simplifications we can rewrite the message we are calculating as:
\[
m_4 = (1 - c_i) [(1 - \lambda)\mathcal{CN}(0; \hat{r}, \mu^r) + \lambda S_0] + c_i \sum_{n=1}^{3} w_n S_n
= (1 - c_i)\rho_0 + c_i\rho_1
\] (B.15)
\[
\rho_0 = (1 - \lambda)\mathcal{CN}(0; \hat{r}, \mu^r) + \lambda S_0
\]
\[
\rho_1 = \sum_{n=1}^{3} w_n S_n. \] (B.16)

Finally, the message can be simplified to a Bernoulli distribution which is completely described by one variable so we will just pass this one variable given as
\[
m_4 = \begin{cases} 1 & \text{w.p. } \rho \\ 0 & \text{w.p. } (1 - \rho) \end{cases} = \rho = \frac{\rho_1}{\rho_0 + \rho_1}. \] (B.17)

The equation above results from normalizing the calculated coefficients to make it a valid pdf.

### B.3 Clutter Indicator Posterior Calculation

Using the messages leaving the GAMP iteration, we can then calculate the approximate posterior of the clutter indicator variables. The approximate posterior calculation is illustrated in the figure below.
Using the figure above, the approximate posterior for the clutter indicator variable is then calculated as:

\[
p(c_i|y) = \begin{cases} 
1 \quad \text{w.p.} \quad \hat{\rho}_i \\
0 \quad \text{w.p.} \quad (1 - \hat{\rho}_i) 
\end{cases} \tag{B.19}
\]

where,

\[
\hat{\rho}_i = \frac{\rho \alpha_i}{\rho \alpha_i + (1 - \rho)(1 - \alpha_i)} \tag{B.20}
\]
Appendix C: Expectation Maximization Update Derivations

In this appendix we will re-derive parts of the EM algorithm presented by Vila and Schniter [83] so that it can be easily modified for our application.

C.1 Overview

We begin by defining the vector of parameters that the prior depends on:

\[ q = \{\lambda, \omega, \theta, \phi\} \]

and we can then define the prior as:

\[ p_X(x; q) = (1 - \lambda)\delta(x) + \lambda \sum_{l=1}^{L} \omega_l \mathcal{N}(x; \theta_l, \phi_l). \]

Using this prior, we can then calculate the approximate posterior as:

\[ p_X|Y(x|y; \hat{r}, \mu^r) = \frac{p_X(x)\mathcal{N}(x; \hat{r}, \mu^r)}{\int_x p_X(x')\mathcal{N}(x'; \hat{r}, \mu^r)}. \]

Finally, we must define the \( g_{in} \) function which is used in the GAMP algorithm. The function is calculated as follows:

\[ g_{in} = \int_x x p_X|Y(x|y; \hat{r}, \mu^r) = \mathbf{E}\{X|Y\}. \]

Using these definitions, we will now calculate the posterior \( p_X|Y(x|y; \hat{r}, \mu^r) \) and the \( g_{in} \) function which we will use later in the EM calculations.
First we will calculate the posterior. To make these calculations easier, it is important to note the property of Gaussian multiplication. The following equality will be helpful throughout:

\[ \mathcal{N}(x; a, A) \mathcal{N}(x; b, B) = \mathcal{N} \left( x; \frac{a/A + b/B}{1/A + 1/B}, \frac{1}{1/A + 1/B} \right) \mathcal{N}(0; a - b, A + B). \] (C.5)

We will now calculate the numerator of the fraction in the posterior equation above. The calculation is as follows:

\[ p_X(x) \mathcal{N}(x; \hat{r}, \mu^r) = \left[ (1 - \lambda) \delta(x) + \lambda \sum_{l=1}^{L} \omega_l \mathcal{N}(x; \theta_l, \phi_l) \right] \mathcal{N}(x; \hat{r}, \mu^r) \] (C.6)

Next we will calculate the denominator of the fraction in the posterior equation. This calculation is as follows:

\[ \int p_X(x) \mathcal{N}(x; \hat{r}, \mu^r) = \int (1 - \lambda) \mathcal{N}(x; \hat{r}, \mu^r) \delta(x) + \int \lambda \sum_{l=1}^{L} \omega_l \mathcal{N}(x; \omega_l, \phi_l) \mathcal{N}(x; \hat{r}, \mu^r) \] (C.7)

and performing each integral separately,

\[ \int (1 - \lambda) \mathcal{N}(x; \hat{r}, \mu^r) \delta(x) = (1 - \lambda) \mathcal{N}(0; \hat{r}, \mu^r) \] (C.8)

\[ \int \lambda \sum_{l=1}^{L} \omega_l \mathcal{N}(x; \theta_l, \phi_l) \mathcal{N}(x; \hat{r}, \mu^r) = \lambda \sum_{l=1}^{L} \omega_l \mathcal{N}(0; \theta_l - \hat{r}, \phi_l + \mu^r) \int \mathcal{N}(x; \gamma_l, v_l) \] (C.9)

where,

\[ \gamma_l(\hat{r}, \mu^r; q) = \frac{\hat{r}/\mu^r + \theta_l/\phi_l}{1/\mu^r + 1/\phi_l} \] (C.10)

\[ v_l(\hat{r}, \mu^r; q) = \frac{1}{1/\mu^r + 1/\phi_l}. \] (C.11)

Finally, letting

\[ \beta_l(\hat{r}, \mu^r; q) = \lambda \omega_l \mathcal{N}(0; \theta_l - \hat{r}, \phi_l + \mu^r) \] (C.12)
then:

\[
\int \lambda \sum_{l=1}^{L} \omega_l \mathcal{N}(x; \theta_l, \phi_l) \mathcal{N}(x; \hat{r}, \mu^r) = \sum_{l=1}^{L} \beta_l(\hat{r}, \mu^r; q). \tag{C.13}
\]

The denominator of the posterior calculation can now be simplified to

\[
\int p_X(x) \mathcal{N}(x; \hat{r}, \mu^r) = (1 - \lambda) \mathcal{N}(0; \hat{r}, \mu^r) + \sum_{l=1}^{L} \beta_l(\hat{r}, \mu^r; q). \tag{C.14}
\]

By letting

\[
\zeta(\hat{r}, \mu^r; q) = (1 - \lambda) \mathcal{N}(0; \hat{r}, \mu^r) + \sum_{l=1}^{L} \beta_l(\hat{r}, \mu^r; q), \tag{C.15}
\]

the approximate posterior can be written as

\[
p_{X|Y}(x|y; \hat{r}, \mu^r) = \frac{(1 - \lambda) \mathcal{N}(x; \hat{r}, \mu^r) \delta(x) + \lambda \sum_{l=1}^{L} \omega_l \mathcal{N}(x; \omega_l, \phi_l) \mathcal{N}(x; \hat{r}, \mu^r)}{\zeta(\hat{r}, \mu^r; q)}
\]

\[= (1 - \lambda) \mathcal{N}(x; \hat{r}, \mu^r) \delta(x) + \sum_{l=1}^{L} \beta_l(\hat{r}, \mu^r; q) \mathcal{N}(x; \gamma_l, \nu_l) \zeta(\hat{r}, \mu^r; q). \tag{C.16}
\]

Now that we have calculated the posterior, it is now possible to calculate the \(g_{in}\) function which is used by GAMP.

We will now use the posterior that we calculated to calculate the \(g_{in}\) function which is used in GAMP. The function is calculated as follows:

\[
g_{in}(\hat{r}, \mu^r; q) = \frac{\sum_{l=1}^{L} \beta_l(\hat{r}, \mu^r; q) \gamma_l(\hat{r}, \mu^r; q)}{\zeta(\hat{r}, \mu^r; q)}. \tag{C.17}
\]

This function is necessary for the implementation of GAMP and some of the definitions used in its derivation will be needed later.

Now that we have calculated the approximate posterior and \(g_{in}\) function, we will derive the EM update rule for the weights of the Gaussian mixture component of the prior. First, it will be helpful to calculate and define the posterior activity probabilities. We begin by letting

\[
\bar{\beta}_l = \frac{\beta_l}{\sum_{k=1}^{L} \beta_k}, \tag{C.18}
\]

138
then the posterior probability \( x \neq 0 \) is calculated as:

\[
\Pr\{x \neq 0|y; q\} = \int_{x \neq 0} p_{X|Y}(x|\hat{r}, \mu^r)
\]

\[
= \int_{x \neq 0} \frac{(1 - \lambda)\mathcal{N}(x; \hat{r}, \mu^r)\delta(x) + \sum_{l=1}^L \beta_l(\hat{r}, \mu^r; q)\mathcal{N}(x; \gamma_l, v_l)}{\zeta(\hat{r}, \mu^r; q)}
\]

\[
= \int_{x \neq 0} \frac{\sum_{l=1}^L \beta_l(\hat{r}, \mu^r; q)\mathcal{N}(x; \gamma_l, v_l)}{\zeta(\hat{r}, \mu^r; q)}
\]

\[
= \sum_{l=1}^L \beta_l \frac{(1 - \lambda)\mathcal{N}(0; \hat{r}, \mu^r) + \sum_{l=1}^L \beta_l(\hat{r}, \mu^r; q)}{1 + \frac{(1 - \lambda)\mathcal{N}(0; \hat{r}, \mu^r)}{\sum_{l=1}^L \beta_l(\hat{r}, \mu^r; q)}}.
\]

For future use we will define this calculated posterior support probability as follows:

\[
\pi(\hat{r}, \mu^r; q) = \Pr\{x \neq 0|y; q\} = \frac{1}{1 + \frac{(1 - \lambda)\mathcal{N}(0; \hat{r}, \mu^r)}{\sum_{l=1}^L \beta_l(\hat{r}, \mu^r; q)}}. 
\]

(C.20)

It will also be useful to define the joint activity/mixture probabilities as

\[
\Pr\{x_n \neq 0, k_n = k|y, q^i\} = \frac{\beta_k(\hat{r}, \mu^r; q^i)}{(1 - \lambda)\mathcal{N}(0; \hat{r}, \mu^r) + \sum_{l=1}^L \beta_l(\hat{r}, \mu^r; q^i)},
\]

(C.21)

where “\( k_n = k \)” represents the event that \( x_n \) was generated by mixture component \( k \).

C.2 Moving Target Sparsity Parameter Update

In this section we will derive the EM update rule for the moving target sparsity rate, \( \lambda \). It is useful to begin with the signal prior in the following form

\[
p_X(x_n) = (1 - \alpha_n) [(1 - \lambda)\delta(x_n) + \lambda \mathcal{CN}(x_n; \mu_{\text{mov}}, \sigma_{\text{mov}}^2)] + \alpha_n \sum_{l=1}^L w_l \mathcal{CN}(x_n; \mu_l, \sigma_l^2).
\]

(C.22)

The update is then found by solving

\[
\lambda^{i+1} = \arg \max_{\lambda \in (0,1)} \sum_{n=1}^N \mathbb{E} \left\{ \ln p_X(x_n; \lambda, q^i(y; q^i) \right\}.
\]

(C.23)
Taking the derivative and setting it to zero results in

$$\sum_{n=1}^{N} \int_{x_n} p_{X|Y}(x_n|y; q^i) \frac{d}{d\lambda} \ln p_{X}(x_n; \lambda, q^i_{\lambda}) = 0. \quad (C.24)$$

Then focusing on the derivative:

$$\frac{d}{d\lambda} \ln p_{X}(x_n; \lambda, q^i_{\lambda}) = \frac{(1 - \alpha_n) \left[ \mathcal{CN}(x_n; \mu_{mov}, \sigma^2_{mov}) - \delta(x_n) \right]}{p_X(x_n; \lambda, q^i_{\lambda})}$$

$$= \frac{-(1 - \alpha_n)}{(1 - \alpha_n)(1 - \lambda)} = \frac{-1}{(1 - \lambda)} \quad (C.25)$$

where the second equality follows the same argument of [83]. Substituting the derivative back into Eq. C.24 results in

$$\sum_{n=1}^{N} \int_{x_n \in B_c} p_{X}(x_n; q^i) \mathcal{CN}(x; \hat{r}, \mu^r) \frac{1 - \alpha_n \left[ \mathcal{CN}(x_n; \mu_{mov}, \sigma^2_{mov}) - \delta(x_n) \right]}{p_X(x_n; \lambda, q^i_{\lambda})} \zeta(\hat{r}, \mu^r; q^i)$$

$$= \frac{1}{1 - \lambda} \sum_{n=1}^{N} \int_{x_n \in B_c} p_{X|Y}(x_n|y; q^i), \quad (C.26)$$

where the integral has been separated as in [83]. Solving the integrals results in

$$\sum_{n=1}^{N} \frac{(1 - \alpha_n) \lambda \xi S_0}{\zeta(\hat{r}, \mu^r; q^i)} = \frac{\lambda}{1 - \lambda} \sum_{n=1}^{N} (1 - \pi_n), \quad (C.27)$$

where if we then substitute these values with their interpreted meaning we arrive at

$$\sum_{n=1}^{N} \Pr\{x_n \neq 0, k_n = \text{mov}|y, q^i\} = \frac{\lambda}{1 - \lambda} \sum_{n=1}^{N} \Pr\{x_n = 0|y, q^i\}. \quad (C.28)$$

Finally, the update rule is calculated as

$$\lambda_{i+1} = \frac{\sum_{n=1}^{N} \Pr\{x_n \neq 0, k_n = \text{mov}|y, q^i\}}{\sum_{n=1}^{N} \Pr\{x_n \neq 0, k_n = \text{mov}|y, q^i\} + \sum_{n=1}^{N} \Pr\{x_n = 0|y, q^i\}}. \quad (C.29)$$

**C.3 Clutter Weight Parameter Update**

In this section we will modify the EM update rules of [83] to fit our prior signal model. Our model also uses a Bernoulli-Gaussian mixture model, but in our case we
know some of the values and want to keep them set while learning the others. In our model, the prior is:

\[
p_X(x_n) = (1 - \gamma_n)\delta(x_n) + \gamma_n \sum_{k=1}^{4} q_k \mathcal{N}(x_n; \theta_k, \phi_k)
\]  

(C.30)

\[
q_k = \frac{\alpha_n\omega_k}{\gamma_n} \quad \text{for } k = 1, 2, 3
\]  

(C.31)

\[
q_4 = \frac{(1 - \alpha_n)\lambda_n}{\gamma_n}, \quad \theta_4 = \theta_{\text{mov}}, \quad \text{and} \quad \phi_4 = \phi_{\text{mov}}.
\]  

(C.32)

With this prior we can use existing EM rules by Vila and Schniter [83] to update the means and variances of the complex normal distribution. However, our model departs from the traditional Bernoulli-Gaussian mixture model in that we know a part of the sparsity rate \(\gamma_n\) and a part of one of the weights of the mixture model, \(q_4\). We will therefore modify the EM update rule to learn the three unknown weight variables in our prior model.

We will now modify the EM update rule for the weight vector as follows:

\[
\arg \max_{\omega, \xi} \sum_{n=1}^{N} \int_{x_n} \ln p_X(x_n; \omega, q_i(\omega)p_X|Y(x_n|y, \hat{r}_n, \mu_{r_n}) = \xi \left( \sum_{l=1}^{3} \omega_l - 1 \right).
\]  

(C.33)

Next, taking the derivative and setting it equal to zero

\[
\sum_{n=1}^{N} \frac{\alpha_n}{\gamma_n} \int_{x_n} p_X(x_n; q^i) \mathcal{N}(x_n; \hat{r}_n, \mu_{r_n}) \gamma_n \mathcal{N}(x_n; \theta_i, \phi_i) \frac{\zeta(\hat{r}_n, \mu_{r_n}; q^i)}{p_X(x_n; \omega, q_i(\omega))} = \xi,
\]  

(C.34)

and following the steps from Vila and Schniter approximating \(\omega \approx \omega^i\) we arrive at

\[
\xi = \sum_{n=1}^{N} \frac{\alpha_n}{\gamma_n} \int_{x_n} \frac{\gamma_n \mathcal{N}(x_n; \theta^i_k, \phi^i_k) \mathcal{N}(x_n; \hat{r}_n, \mu_{r_n})}{\zeta(\hat{r}_n, \mu_{r_n}; q^i)} = \sum_{n=1}^{N} \frac{\alpha_n}{\gamma_n} \Pr\{x_n \neq 0|y; q^i\}.
\]  

(C.35)

It can then be seen from the equation above that our update rule is similar to that of Vila and Schniter [83] except for the additional factor \(\alpha_n/\gamma_n\). Showing the first step
leading to the update rule

\[
\sum_{n=1}^{N} \frac{\alpha_n}{\gamma_n} \Pr\{x_n \neq 0|y; q^i\} = \sum_{n=1}^{N} \frac{\alpha_n}{\gamma_n} \int_{x_n} \frac{\gamma_n \mathcal{N}(x_n; \theta_k^i, \phi_k^i)\mathcal{N}(x_n; \hat{r}_n, \mu_n^r)}{\zeta(\hat{r}_n, \mu_n^r; q^i)}, \quad (C.36)
\]

we then arrive at our EM update rule for the clutter weights:

\[
\omega_k^{i+1} = \frac{\sum_{n=1}^{N} \Pr\{x_n \neq 0, k_n = k|y, q^i\}}{\sum_{n=1}^{N} \frac{\alpha_n}{\gamma_n} \Pr\{x_n \neq 0|y; q^i\}}.
\quad (C.37)
\]
Appendix D: Bayesian Detector Derivations

In this appendix we derive various detection schemes that can be used with the proposed Bayesian model in chapter 4.

D.1 Whitened Matched Filter

One of the two approaches for detection is to use the output of the Bayesian model to estimate a clutter covariance matrix. Using this estimated covariance matrix, the detections are then made from the traditional whitened matched filter using the adaptive matched filter test statistic.

D.1.1 GAMP Approximate Posterior Calculation

In this subsection it is useful to note the following forms of the approximate posterior and signal prior

\[
p_{X|Y}(x_i|y; q_i, \hat{r}_i, \mu_i^r) \propto p_{X|Q}(x_i|q_i)CN(x_i; \hat{r}_i, \mu_i^r) \tag{D.1}
\]

\[
p_{X|Q}(x_i|q_i) = (1 - \alpha_i) [(1 - \lambda)\delta(x_i) + \lambda CN(x_i; \mu_{mov}, \sigma^2_{mov})] + \alpha_i \sum_{n=1}^{3} w_n CN(x_i; \mu_n, \sigma^2_n). \tag{D.2}
\]

We wish to find \( p(x_i|y, c_i = 1) \) from \( p_{X|Y}(x_i|y; q_i, \hat{r}_i, \mu_i^r) \) using the following factorization:

\[
p_{X|Y}(x_i|y) = \int p(x_i, c_i|y) \, dc_i \tag{D.3}
\]
\[
\int p(x_i|y, c_i)p(c_i|y) \, dc_i = p(x_i|y, c_i = 0)p(c_i = 0|y) + p(x_i|y, c_i = 1)p(c_i = 1|y). \tag{D.4}
\]

Knowing that the posterior \(p_{X|Y}(x_i|y)\) takes the form shown in Eq. D.5, we can find \(p(x_i|y, c_i = 1)\) by calculating the GAMP approximate posterior using Eq. D.1. We begin by looking at the second term of the GAMP approximate posterior that results from the multiplication of the prior, \(p_{X|Q}(x_i|q_i)\), with \(\text{CN}(x_i; \hat{r}_i, \mu_i^r)\):

\[
p(x_i|y, c_i = 1) \propto \alpha_i \sum_{n=1}^{3} w_n \text{CN}(x_i; \hat{\mu}_{n,i}, \hat{\sigma}^2_{n,i}) \tag{D.6}
\]

\[
= \alpha_i \left( \sum_{n=1}^{3} w_n S_{n,i} \right) \sum_{n=1}^{3} \frac{w_n S_{n,i}}{\sum_{l=1}^{3} w_l S_{l,i}} \text{CN}(x_i; \hat{\mu}_{n,i}, \hat{\sigma}^2_{n,i}). \tag{D.7}
\]

Looking at Eq. D.8 we can see that \(p(c_i = 1|y)\) is proportional to \(\alpha_i \left( \sum_{n=1}^{3} w_n S_{n,i} \right)\), therefore,

\[
p(x_i|y, c_i = 1) = \sum_{n=1}^{3} \frac{w_n S_{n,i}}{\sum_{l=1}^{3} w_l S_{l,i}} \text{CN}(x_i; \hat{\mu}_{n,i}, \hat{\sigma}^2_{n,i}) \tag{D.9}
\]

\[
\hat{\mu}_{n,i} = \frac{\sigma_n^2 \text{Re}(\mu_n) + \mu_i^r \text{Re}(\hat{r}_i)}{\sigma_n^2 + \mu_i^r} + i \frac{\sigma_n^2 \text{Im}(\mu_n) + \mu_i^r \text{Im}(\hat{r}_i)}{\sigma_n^2 + \mu_i^r} \tag{D.10}
\]

\[
\hat{\sigma}^2_{n,i} = \frac{\sigma_n^2 \mu_i^r}{\sigma_n^2 + \mu_i^r} \tag{D.11}
\]

and since the weights of the Gaussian mixture sum to one, it is a valid PDF. It is also important to note that

\[
\mu_i = \mathbb{E} [p(x_i|y, c_i = 1)] = \sum_{n=1}^{3} \frac{w_n S_{n,i}}{\sum_{l=1}^{3} w_l S_{l,i}} \hat{\mu}_{n,i} \tag{D.12}
\]

\[
\sigma_i^2 = \text{Var} (p(x_i|y, c_i = 1)) = \sum_{n=1}^{3} \frac{w_n S_{n,i}}{\sum_{l=1}^{3} w_l S_{l,i}} (|\hat{\mu}_{n,i} - \mu_i|^2 + \hat{\sigma}^2_{n,i}), \tag{D.13}
\]

and this will be used in the following sections.
D.1.2 MMSE Clutter Estimate

We wish to estimate the clutter covariance matrix from our Bayesian model and will do so as follows:

\[ \hat{R} = E \{ R | y \} \]  \hspace{1cm} (D.14)
\[ = E \{ R_c | y \} + E \{ R_n | y \} \]  \hspace{1cm} (D.15)
\[ = E \{ R_c | y \} + \sigma_{\text{noise}}^2 I \]  \hspace{1cm} (D.16)
\[ = \hat{R}_c + \hat{R}_n \]  \hspace{1cm} (D.17)

where we assume we know the variance of the thermal noise component, \( \sigma_{\text{noise}}^2 \). The clutter model presented by Ward [86] uses only the power from clutter patches to calculate the covariance matrix; however, our Bayesian approach estimates the complex reflectivity of the entire scene at a given range, not just the clutter reflectivity. Therefore we modify the model as such:

\[ R_c = A \text{diag}(c \odot |x|^2) A^H \]  \hspace{1cm} (D.18)

where \( x \) now contains the complex reflectivity of the entire scene at a given range, \( c \) is a vector of binary indicator variables that masks away non-clutter components in the vector \( x \), and \( \odot \) is the Hadamard product which is a pointwise multiplication. Our estimate of the clutter component of the interference matrix is then:

\[ \hat{R}_c = E \{ A \text{diag}(c \odot |x|^2) A^H | y \} . \]  \hspace{1cm} (D.19)

Then continuing,

\[ E \{ A \text{diag}(c \odot |x|^2) A^H | y \} = AE \{ \text{diag}(c \odot |x|^2) | y \} A^H \]  \hspace{1cm} (D.20)
\[ = AE \{ A | y \} A^H \]  \hspace{1cm} (D.21)
where $\Lambda_{i,j} = c_i |x_i|^2$ for $i = j$ and 0 otherwise. Next, we calculate $E \{ \Lambda | y \}$ as follows:

$$E \{ \Lambda | y \}_{i,j} = E \{ \Lambda_{i,j} | y \}$$

$$= 0 \quad \text{for } i \neq j \quad \text{(D.23)}$$

$$= \int \int c_i |x_i|^2 p(x_i, c_i | y) \, dc_i \, dx_i \quad \text{for } i = j \quad \text{(D.24)}$$

where we need only calculate the diagonal terms. Before calculating the integral, we should first look at the factorization of the joint probability distribution. We will be using the fact that

$$p(x_i, c_i | y) = p(x_i | y, c_i)p(c_i | y). \quad \text{(D.25)}$$

The diagonal terms of $E \{ \Lambda_{i,j} | y \}$ can then be calculated as:

$$E \{ \Lambda_{i,j} | y \} = \int \int c_i |x_i|^2 p(x_i, c_i | y) \, dc_i \, dx_i \quad \text{(D.26)}$$

$$= \int \sum_{c_i=0}^1 c_i |x_i|^2 p(x_i | y, c_i)p(c_i | y) \, dx_i \quad \text{(D.27)}$$

$$= p(c_i = 1 | y) \int |x_i|^2 p(x_i | y, c_i = 1) \, dx_i \quad \text{(D.28)}$$

$$= p(c_i = 1 | y)(\sigma_i^2 + |\mu_i|^2) \quad \text{(D.29)}$$

where $\mu_i$ and $\sigma_i^2$ are the mean and variance of $p(x_i | y, c_i = 1)$, respectively. Finally, combining the results gives the following estimated interference covariance matrix:

$$\hat{R} = AE \{ \Lambda | y \} A^H + \sigma_{\text{noise}}^2 I \quad \text{(D.30)}$$

$$E \{ \Lambda | y \}_{i,j} = E \{ \Lambda_{i,j} | y \} = \begin{cases} 0 & \text{for } i \neq j \\ p(c_i = 1 | y)(\sigma_i^2 + |\mu_i|^2) & \text{for } i = j \end{cases}. \quad \text{(D.31)}$$
D.2 Bayesian Hypothesis Testing

The second of the two detection schemes we will look at is Bayesian hypothesis testing. This approach will leverage the Bayesian model formulation in order to perform a hypothesis test. We begin by looking at the principles of Bayesian hypothesis testing.

Bayesian hypothesis testing compares the likelihood that the data observed comes from one model or another. Specifically, the test takes the form

\[
p(M_1|\mathbf{y}) \bigg/ p(M_2|\mathbf{y}) \geq 1,
\]

which arises from minimizing the expected posterior loss. The ratio is then calculated as

\[
\frac{p(M_1|\mathbf{y})}{p(M_2|\mathbf{y})} = \frac{p(\mathbf{y}|M_1)p(M_1)/p(\mathbf{y})}{p(\mathbf{y}|M_2)p(M_2)/p(\mathbf{y})}
= \left(\frac{p(M_1)}{p(M_2)}\right) \left(\frac{p(\mathbf{y}|M_1)}{p(\mathbf{y}|M_2)}\right),
\]

where the first term contains the prior model odds and the second term is known as the Bayes factor. The Bayes factor quantifies the evidence the data provides for selecting one model over the other. The numerator and denominator of the Bayes factor can be calculated as

\[
p(\mathbf{y}|M) = \int_x p(x|M)p(\mathbf{y}|x,M) \, dx,
\]

where, using the GAMP framework, we must substitute \(p(\mathbf{y}|x,M)\) with the GAMP equivalent approximate likelihood \(\mathcal{CN}(x; \hat{r}, \mu^r)\). The final step is to find the conditional prior \(p(x|M)\) and the prior model odds \(p(M_1)/p(M_2)\).
To find the conditional prior and prior model odds we must look at the prior we use in the proposed model. The prior in the model is

\[ p(x_i) = (1 - \gamma_i)\delta(x_i) + \gamma_i \sum_{n=1}^{4} q_n \mathcal{CN}(x; \mu_n, \sigma^2_n), \tag{D.36} \]

which was derived in a previous section. We now wish to rearrange the prior so that it can be separated into the moving target and non-moving target model prior. Letting \( \xi_i = (1 - \alpha_i)\lambda \), the prior in Eq. D.36 can be rearranged as

\[ (1 - \xi_i) \left[ \left( \frac{1 - \gamma_i}{1 - \xi_i} \right) \delta(x_i) + \frac{\alpha_i}{1 - \xi_i} \sum_{n=1}^{3} w_n \mathcal{CN}(x_i; \mu_n, \sigma^2_n) \right] + \xi_i \mathcal{CN}(x_i; \mu_{\text{mov}}, \sigma^2_{\text{mov}}), \tag{D.37} \]

where we can interpret \( \xi_i \) and \( 1 - \xi_i \) as the prior model odds and the other terms are the conditional priors for the mover and non-mover models. Specifically,

\[ p(M_1) = \xi_i \]

\[ p(x|M_1) = \mathcal{CN}(x_i; \mu_{\text{mov}}, \sigma^2_{\text{mov}}) \]

\[ p(M_2) = (1 - \xi_i) \]

\[ p(x|M_2) = \left( \frac{1 - \gamma_i}{1 - \xi_i} \right) \delta(x_i) + \frac{\alpha_i}{1 - \xi_i} \sum_{n=1}^{3} w_n \mathcal{CN}(x_i; \mu_n, \sigma^2_n). \tag{D.39} \]

Given these definitions, we can then calculate the Bayes factor and the ratio we need for the Bayesian hypothesis test. First, we calculate the numerator and denominator of the Bayes factor as

\[ p(y|M_1) = \int_{x_i} \mathcal{CN}(x_i; \mu_{\text{mov}}, \sigma^2_{\text{mov}}) \mathcal{CN}(x_i; \hat{r}, \mu') \]

\[ = S_0 \int_{x_i} \mathcal{CN}(x_i; \hat{\mu}_0, \hat{\sigma}^2_0) \]

\[ = S_0 \]

148
\[
p(y|M_2) = \int_{x_i} \left[ \frac{(1 - \gamma_i)}{1 - \xi_i} \delta(x_i) + \frac{\alpha_i}{1 - \xi_i} \sum_{n=1}^3 w_n \mathcal{CN}(x_i; \mu_n, \sigma_n^2) \right] \mathcal{CN}(x_i; \hat{r}, \mu^r)
\]
\[
= \frac{1 - \gamma_i}{1 - \xi_i} \int_{x_i} \delta(x_i) \mathcal{CN}(x_i; \hat{r}, \mu^r) + \frac{\alpha_i}{1 - \xi_i} \sum_{n=1}^3 w_n S_n \int_{x_i} \mathcal{CN}(x_i; \hat{\mu}_n, \hat{\sigma}_n^2)
\]
\[
= \frac{1 - \gamma_i}{1 - \xi_i} \mathcal{CN}(0; \hat{r}, \mu^r) + \frac{\alpha_i}{1 - \xi_i} \sum_{n=1}^3 w_n S_n.
\]

(D.41)

The Bayes factor is then found to be

\[
\frac{p(y|M_1)}{p(y|M_2)} = \frac{(1 - \xi_i)S_0}{(1 - \gamma_i)\mathcal{CN}(0; \hat{r}, \mu^r) + \alpha_i \sum_{n=1}^3 w_n S_n}
\]

(D.42)

and the test statistic is

\[
\frac{p(M_1|y)}{p(M_2|y)} = \frac{p(M_1)}{p(M_2)} \frac{p(y|M_1)}{p(y|M_2)}
\]

\[
= \frac{(\xi_i)}{1 - \xi_i} \frac{(1 - \xi_i)S_0}{(1 - \gamma_i)\mathcal{CN}(0; \hat{r}, \mu^r) + \alpha_i \sum_{n=1}^3 w_n S_n}
\]

(D.43)

This test statistic is then used to choose which model is more likely as shown in Eq. D.32.
Appendix E: Multi-Model Shrinkage Derivations

In this appendix we will derive the convex combination and non-negative combination multi-model shrinkage estimators for possibly complex valued covariance.

E.1 Convex Combination

The multi-model estimator is given as follows

\[ \hat{R} = \left( 1 - \sum_{k=1}^{K} \lambda_k \right) \hat{R} + \sum_{k=1}^{K} \lambda_k T_k \]  

(E.1)

where we assume the convex combination of models that require \( \sum_{k=1}^{K} \lambda_k \leq 1 \). Further, \( \hat{R} \) is assumed to be an unbiased estimator; \( T_k \) is the \( k^{th} \) shrinkage model. We then define the expected squared error as

\[ \text{ESE} = E \left\{ ||\hat{R} - R||^2 \right\} . \]  

(E.2)

We can then expand the ESE as follows:

\[ \text{ESE} = E \left\{ \left\| \left( 1 - \sum_{k=1}^{K} \lambda_k \right) \hat{R} + \sum_{k=1}^{K} \lambda_k T_k - R \right\|^2 \right\} \]

\[ = E \left\{ \text{tr} \left\{ \left( 1 - \sum_{k=1}^{K} \lambda_k \right) \hat{R} + \sum_{k=1}^{K} \lambda_k T_k - R \right\} \times \left( 1 - \sum_{k=1}^{K} \lambda_k \right) \hat{R} + \sum_{k=1}^{K} \lambda_k T_k - R \right\} \right\} \]  

(E.3)
and focusing on the inner terms

\[ \left[ \left( 1 - \sum_{k=1}^{K} \lambda_k \right) \hat{R} + \sum_{k=1}^{K} \lambda_k T_k - R \right] \left[ \left( 1 - \sum_{k=1}^{K} \lambda_k \right) \hat{R} + \sum_{k=1}^{K} \lambda_k T_k - R \right]^H \]

\[ = \left( 1 - \sum_{k=1}^{K} \lambda_k \right)^2 \hat{R} \hat{R}^H + \left( 1 - \sum_{k=1}^{K} \lambda_k \right) \sum_{k=1}^{K} \lambda_k \hat{R} T_k^H - \left( 1 - \sum_{k=1}^{K} \lambda_k \right) \hat{R} R^H \]

\[ + \left( 1 - \sum_{k=1}^{K} \lambda_k \right) \sum_{k=1}^{K} \lambda_k T_k \hat{R}^H + \left( \sum_{j=1}^{K} \lambda_j T_j \right) \left( \sum_{k=1}^{K} \lambda_k T_k \right)^H - \sum_{k=1}^{K} \lambda_k T_k R_k^H \]

\[ - \left( 1 - \sum_{k=1}^{K} \lambda_k \right) R R^H - \sum_{k=1}^{K} \lambda_k R T_k^H + R R^H. \]

Next, we expand the terms above to arrive at

\[ \left[ 1 - 2 \sum_{k=1}^{K} \lambda_k + \sum_{j=1}^{K} \sum_{k=1}^{K} \lambda_j \lambda_k \right] \hat{R} \hat{R}^H + \left[ \sum_{k=1}^{K} \lambda_k - \sum_{j=1}^{K} \sum_{k=1}^{K} \lambda_j \lambda_k \right] \hat{R} T_k^H \]

\[ + \left[ \sum_{k=1}^{K} \lambda_k - 1 \right] \hat{R} R^H + \left[ \sum_{k=1}^{K} \lambda_k - \sum_{j=1}^{K} \sum_{k=1}^{K} \lambda_j \lambda_k \right] T_j \hat{R}^H + \left[ \sum_{j=1}^{K} \sum_{k=1}^{K} \lambda_j \lambda_k \right] T_j T_k^H \]

\[ - \sum_{k=1}^{K} \lambda_k T_k R_k^H + \left[ \sum_{k=1}^{K} \lambda_k - 1 \right] R R^H - \sum_{k=1}^{K} \lambda_k R T_k^H + R R^H. \]

Rearranging the terms,

\[ \sum_{j=1}^{K} \sum_{k=1}^{K} \lambda_j \lambda_k \left[ \hat{R} \hat{R}^H - \hat{R} T_k^H - T_j \hat{R}^H + T_j T_k^H \right] \]

\[ - 2 \sum_{k=1}^{K} \lambda_k \left[ \hat{R} \hat{R}^H - \frac{1}{2} \hat{R} R^H \right] + \sum_{k=1}^{K} \lambda_k \left[ \hat{R} T_k^H + T_k \hat{R}^H - R T_k^H - T_k R_k^H \right] \]

\[ + \hat{R} R^H - \hat{R} R^H - R R^H + R R^H \]

and then exploiting the fact that \( \hat{R} \) is an unbiased estimate and \( T_k \) is deterministic we arrive at

\[ \sum_{j=1}^{K} \sum_{k=1}^{K} \lambda_j \lambda_k \left( \hat{R} - T_j \right) \left( \hat{R} - T_k \right)^H - 2 \sum_{k=1}^{K} \lambda_k \left( \hat{R} - R \right) \left( \hat{R} - R \right)^H + \]

\[ \left( \hat{R} - R \right) \left( \hat{R} - R \right)^H. \]
Substituting this back into the equation for the ESE we get

\[
ESE = \sum_{j=1}^{K} \sum_{k=1}^{K} \lambda_j \lambda_k E \left\{ \text{tr} \left[ (\hat{R} - T_j) \left( \hat{R} - T_k \right)^H \right] \right\} - 2 \sum_{k=1}^{K} \lambda_k E \left\{ ||\hat{R} - R||^2 \right\} + \text{const},
\]

where \( \text{const} = E \left\{ \text{tr} \left[ (\hat{R} - R)(\hat{R} - R)^H \right] \right\} \). This can then be stated as the quadratic

\[
ESE = \lambda^T A \lambda - 2b^T \lambda + \text{const},
\]

where

\[
A_{j,k} = E \left\{ \text{tr} \left[ (\hat{R} - T_j) \left( \hat{R} - T_k \right)^H \right] \right\} \quad \text{(E.10)}
\]

\[
b_k = b = E \left\{ ||\hat{R} - R||^2 \right\}.
\]

The result in Eq. E.10 matches the form of [6, Eq. 9] where the differences result from the fact that we have assumed a possibly complex-valued covariance matrix.

### E.2 Non-Negative Combination

The multi-model estimator is given as follows

\[
\hat{R} = \lambda_1 \hat{R} + \sum_{k=2}^{K+1} \lambda_k T_k.
\]

We then define the expected squared error again as

\[
ESE = E \left\{ ||\hat{R} - R||^2 \right\}.
\]

We can then expand the ESE as follows:

\[
ESE = E \left\{ \left\| \lambda_1 \hat{R} + \sum_{k=2}^{K+1} \lambda_k T_k - R \right\|^2 \right\} = E \left\{ \text{tr} \left[ \left( \lambda_1 \hat{R} + \sum_{k=2}^{K+1} \lambda_k T_k - R \right) \right] \right\} \left\{ \lambda_1 \hat{R} + \sum_{k=2}^{K+1} \lambda_k T_k - R \right\}^H \right\}.
\]
and focusing on the inner terms

\[
\begin{align*}
\left[ \lambda_1 \hat{R} + \sum_{k=2}^{K+1} \lambda_k T_k - R \right] \left[ \lambda_1 \hat{R} + \sum_{k=2}^{K+1} \lambda_k T_k - R \right]^H \\
= \lambda_1^2 \hat{R} \hat{R}^H + \lambda_1 \sum_{k=2}^{K+1} \lambda_k \hat{T}_k^H - \lambda_1 \hat{R} \hat{R}^H \\
+ \lambda_1 \sum_{k=2}^{K+1} \lambda_k \hat{T}_k \hat{T}_k^H + \sum_{j=2}^{K+1} \sum_{k=2}^{K+1} \lambda_j \lambda_k \hat{T}_j \hat{T}_k^H - \sum_{k=2}^{K+1} \lambda_k \hat{T}_k \hat{R}^H \\
- \lambda_1 \hat{R} \hat{R}^H - \sum_{k=2}^{K+1} \lambda_k \hat{T}_k \hat{T}_k^H + \hat{R} \hat{R}^H.
\end{align*}
\]

(E.14)

Then rearranging the terms for clarity:

\[
\begin{align*}
\lambda_1 \sum_{k=2}^{K+1} \lambda_k \left[ \hat{T}_k^H + T_k \hat{T}_k^H \right] \\
+ \lambda_1^2 \left[ \hat{R} \hat{R}^H \right] \\
+ \sum_{j=2}^{K+1} \sum_{k=2}^{K+1} \lambda_j \lambda_k \left[ T_j \hat{T}_k^H \right] \\
- \lambda_1 \left[ \hat{R} \hat{R}^H + \hat{R} \hat{R}^H \right] \\
- \sum_{k=2}^{K+1} \lambda_k \left[ \hat{T}_k \hat{R}^H + \hat{T}_k \hat{R}^H \right] \\
+ \hat{R} \hat{R}^H.
\end{align*}
\]

(E.15)

Then focusing on one term at a time and exploiting the linearity of the trace and expectation operators:

\[
\begin{align*}
E \left\{ \text{tr} \left[ \lambda_1 \sum_{k=2}^{K+1} \lambda_k \left( \hat{T}_k^H + T_k \hat{T}_k^H \right) \right] \right\} &= 2 \lambda_1 \sum_{k=2}^{K+1} \lambda_k \text{Re} \left\{ \text{tr} \left[ \hat{T}_k^H \right] \right\} \\
E \left\{ \text{tr} \left[ \lambda_1^2 \left( \hat{R} \hat{R}^H \right) \right] \right\} \\
= E \left\{ \text{tr} \left[ \lambda_1^2 \left( \hat{R} \hat{R}^H - \hat{R} \hat{R}^H - \hat{R} \hat{R}^H + \hat{R} \hat{R}^H + \hat{R} \hat{R}^H - \hat{R} \hat{R}^H \right) \right] \right\} \\
= E \left\{ \text{tr} \left[ \lambda_1^2 \left( \hat{R} - R \right) \left( \hat{R} - R \right)^H + \hat{R} \hat{R}^H + \hat{R} \hat{R}^H - \hat{R} \hat{R}^H \right) \right] \right\} \\
= \lambda_1^2 E \left\{ \| \hat{R} - R \|^2 \right\} + \lambda_1^2 \| R \|^2
\end{align*}
\]

(E.17)
\[
E \left\{ \operatorname{tr} \left[ \sum_{j=2}^{K+1} \sum_{k=2}^{K+1} \lambda_j \lambda_k T_j T_k^H \right] \right\}
\]
\[
= E \left\{ \operatorname{tr} \left[ \sum_{j=2}^{K+1} \sum_{k=2}^{K+1} \lambda_j \lambda_k \left\{ \hat{R} \hat{R}^H - \hat{R} T_k^H k - T_j \hat{R}^H + T_j T_k^H + \hat{R} T_k^H + T_j \hat{R}^H - \hat{R} \hat{R}^H \right\} \right] \right\}
\]
\[
= \sum_{j=2}^{K+1} \sum_{k=2}^{K+1} \lambda_j \lambda_k \left\{ E \left\{ \operatorname{tr} \left[ \left( \hat{R} - T_j \right) \left( \hat{R} - T_j \right)^H \right] \right\} + \right.
\]
\[
\left. \operatorname{Re} \left\{ \operatorname{tr} \left[ R T_j^H \right] \right\} + \operatorname{Re} \left\{ \operatorname{tr} \left[ R T_k^H \right] \right\} - E \left\{ \left\| \hat{R} \right\|^2 \right\} \right\}
\]

(E.18)

\[
E \left\{ \operatorname{tr} \left[ \lambda_1 \left( \hat{R} R^H + \hat{R} R^H \right) \right] \right\} = E \left\{ \operatorname{tr} \left[ 2 \lambda_1 \operatorname{Re} \left( R \hat{R}^H \right) \right] \right\} = 2 \lambda_1 \left\| R \right\|^2
\]

(E.19)

\[
E \left\{ \operatorname{tr} \left[ \sum_{k=2}^{K+1} \lambda_k \left( T_k R^H + R T_k^H \right) \right] \right\} = E \left\{ \operatorname{tr} \left[ \sum_{k=2}^{K+1} 2 \lambda_k \operatorname{Re} \left( R T_k^H \right) \right] \right\}
\]
\[
= 2 \sum_{k=2}^{K+1} \lambda_k \operatorname{Re} \left\{ \operatorname{tr} \left[ R T_k^H \right] \right\}
\]
\[
E \left\{ \operatorname{tr} \left[ R R^H \right] \right\} = \left\| R \right\|^2.
\]

(E.20)

We can then put this formulation into the form of a quadratic program

\[
\text{ESE} = \lambda^T A \lambda - 2b^T \lambda + \text{const},
\]

where \( \text{const} = \left\| R \right\|^2 \). The matrix \( A \) and vector \( b \) are then defined as follows

\[
A_{[1,1]} = E \left\{ \left\| \hat{R} - R \right\|^2 \right\} + \left\| R \right\|^2
\]

\[
A_{[1,k=2:K+1]} = \operatorname{Re} \left\{ \operatorname{tr} \left[ R T_k^H \right] \right\}
\]

\[
A_{[j=2:K+1,1]} = \operatorname{Re} \left\{ \operatorname{tr} \left[ R T_k^H \right] \right\}
\]

\[
A_{[j=2:K+1,k=2:K+1]} = E \left\{ \operatorname{tr} \left[ \left( \hat{R} - T_j \right) \left( \hat{R} - T_k \right)^H \right] \right\}
\]
\[
+ \operatorname{Re} \left\{ \operatorname{tr} \left[ R T_j^H \right] \right\} + \operatorname{Re} \left\{ \operatorname{tr} \left[ R T_k^H \right] \right\} - E \left\{ \left\| \hat{R} \right\|^2 \right\}
\]

\[
b_{[1]} = \left\| R \right\|^2
\]

\[
b_{[k=2:K+1]} = \operatorname{Re} \left\{ \operatorname{tr} \left[ R T_k^H \right] \right\}.
\]

154
We can then estimate \( E \left\{ \| \hat{R} - R \|^2 \right\} \) using the results of [78] Eq. 18, the other terms we can estimate as \([6, 78]\)
\[
\| R \|^2 = \| \hat{R} \|^2
\]
\[
\text{tr} \left[ RT_k^H \right] = \text{tr} \left[ \hat{R} T_k^H \right]
\]
\[
E \left\{ \text{tr} \left[ (\hat{R} - T_j)(\hat{R} - T_k)^H \right] \right\} = \text{tr} \left[ (\hat{R} - T_j)(\hat{R} - T_k)^H \right]
\]
\[
E \left\{ \| \hat{R} \|^2 \right\} = \| \hat{R} \|^2.
\]
(E.24)

Using the quadratic program and the estimates above for \( A \) and \( b \), we can then find \( \lambda \) to calculate the MMS estimate of the covariance matrix.

It is important to note that we wish to have a real-valued and symmetric matrix \( A \). It can be seen from Eq. E.23 that the only term that could possibly be complex-valued is \( \text{tr} \left[ (\hat{R} - T_j)(\hat{R} - T_k)^H \right] \). However, we can show that this term should always be real-valued, ensuring that the \( A \) matrix will always be real-valued. First, it is important to note that \( \hat{R}, T_j, \) and \( T_k \) are all Hermitian where \( T_j \neq T_k \) except for when \( j = k \). Then setting \( C = (\hat{R} - T_j) \) and \( D^H = (\hat{R} - T_k)^H \) we can see that \( C \) and \( D \) are also Hermitian. Using the circular property of the trace operator we can see that
\[
\text{tr} \left[ CD^H \right] = \text{tr} \left[ D^H C \right].
\]
(E.25)

Next, using the complex conjugate property of the trace operator we can show
\[
\text{tr} \left[ CD^H \right] = \text{tr} \left[ \overline{DC^H} \right]
\]
\[
= \text{tr} \left[ D^H C \right],
\]
(E.26)

where the second equality uses the fact that \( C \) and \( D \) are Hermitian. Then combining Eq. E.25 and Eq. E.26 it can be seen that
\[
\text{tr} \left[ D^H C \right] = \text{tr} \left[ \overline{D^H C} \right],
\]
(E.27)
which is only possible if $\text{tr}[D^H C]$ is real. Therefore, we have shown that
\[
\text{tr}
\left[
\left(\hat{\mathbf{R}} - \mathbf{T}_j\right)
\left(\hat{\mathbf{R}} - \mathbf{T}_k\right)^H
\right]
\]
will always be real-valued and therefore, $A$ will always be real-valued.