

The Pennsylvania State University
The Graduate School

ROBUST COVARIANCE MATRIX ESTIMATION FOR RADAR
SPACE-TIME ADAPTIVE PROCESSING (STAP)

A Dissertation in
Electrical Engineering
by
Bosung Kang

© 2015 Bosung Kang

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2015

The dissertation of Bosung Kang was reviewed and approved* by the following:

Vishal Monga
Associate Professor of Electrical Engineering
Dissertation Advisor, Chair of Committee

Constantino M. Lagoa
Professor of Electrical Engineering

David J. Miller
Professor of Electrical Engineering

Jesse L. Barlow
Professor of Computer Science and Engineering

Kultegin Aydin
Professor of Electrical Engineering
Head of the Department of Electrical Engineering

*Signatures are on file in the Graduate School.

Abstract

Estimating the disturbance or clutter covariance is a centrally important problem in radar space time adaptive processing (STAP) since estimation of the disturbance or interference covariance matrix plays a central role on radar target detection in the presence of clutter, noise and a jammer. The disturbance covariance matrix should be inferred from training sample observations in practice. Traditional maximum likelihood (ML) estimators are effective when homogeneous (target free) training data is abundant but lead to poor estimates, degraded false alarm rates, and detection loss in the regime of limited training. However, large number of homogeneous training samples are generally not available because of difficulty of guaranteeing target free disturbance observation, practical limitations imposed by the spatio-temporal nonstationarity, and system considerations. The problem has been exacerbated by recent advances that have led to more antenna elements (J) and higher temporal resolution (P) time epochs resulting in a large dimension ($N = JP$).

In this dissertation, we look to address the aforementioned challenges by exploiting physically inspired constraints into ML estimation. While adding constraints is beneficial to achieve satisfactory performance in the practical regime of limited training, it leads to a challenging problem. Unlike unconstrained estimators, a vast majority of constrained radar STAP estimators are iterative and expensive numerically, which prohibits practical deployment. We focus on breaking this classical trade-off between computational tractability and desirable performance measures, particularly in training starved regimes. In particular, we exploit both the structure of the disturbance covariance and importantly the knowledge of the clutter rank to yield a new rank constrained maximum likelihood (RCML) estimator of clutter/disturbance covariance. We demonstrate that the rank-constrained estimation problem can in fact be cast in the framework of a tractable convex optimization problem, and derive closed form expressions for the estimated covariance matrix. In addition, we derive a new covariance estimator for STAP that jointly considers a Toeplitz structure and a rank constraint on the clutter component. Past work has shown that in the regime of low training, even handling each constraint individually is hard and techniques often resort to slow numerically based so-

lutions. Our proposed solution leverages the rank constrained ML estimator (RCML) of structured covariances to build a computationally friendly approximation that involves a cascade of two closed form solutions. Performance analysis using the KASSPER data set (where ground truth covariance is made available) shows that the proposed RCML estimator vastly outperforms state-of-the art alternatives even for low training including the notoriously difficult regime of $K \leq N$ training regimes and for the experiments considering real-world scenarios such as target detection performance and the case that some of training samples are corrupted by target information.

Finally, we address the problem of working with inexact physical radar parameters under a practical radar environment. As shown in this dissertation, employing practical constraints such as a rank of the clutter subspace and a condition number of disturbance covariance leads to a practically powerful estimator as well as a closed form solution. While the rank and the condition number are very effective constraints, often practical non-ideality makes it difficult to be known precisely using physical models. We propose a robust covariance estimation method via an expected likelihood (EL) approach. We analyze covariance estimation algorithms under three different cases of imperfect constraints: 1) only rank constraint, 2) both rank and noise power constraint, and 3) condition number constraint. For each case, we formulate estimation of the constraint as an optimization problem with the expected likelihood criterion and formally derive and prove a significant analytical result such as uniqueness of the solution. Through experimental results from a simulation model and the KASSPER data set, we show the estimator with optimal constraints obtained by the EL approach outperforms alternatives in the sense of a normalized signal-to-interference and noise ratio (SINR).

Table of Contents

List of Figures	viii
List of Tables	x
Acknowledgments	xi
Chapter 1	
Introduction	1
1.1 Motivation	1
1.2 Overview of Dissertation Contributions	3
1.2.1 Estimation of Structured Covariance Matrices	3
1.2.2 Estimation of Constraints under Imperfect Knowledge	4
1.3 Organization	4
Chapter 2	
Background	7
2.1 Space-Time Adaptive Processing (STAP)	7
2.2 Estimation of Covariance Matrices	9
2.2.1 Structured Covariance Estimation	9
2.2.2 Toeplitz Covariance Estimation	13
2.3 Expected Likelihood Approach	19
Chapter 3	
Rank Constrained Maximum Likelihood Estimation	21
3.1 Introduction	21
3.2 Motivation and Review	23
3.3 Rank Constrained ML Estimation of Structured Covariance Matrices . .	24
3.3.1 Overview of contribution	24
3.3.2 ML Estimation	25
3.3.2.1 Known Noise Level Case	28

3.3.2.2	Unknown Noise Level Case	33
3.4	Experimental Results	38
3.4.1	Experimental Setup and Methods Compared	38
3.4.2	Experimental Evaluation	41
3.4.2.1	Eigenspectrum	41
3.4.2.2	Normalized SINR vs. angle and Doppler	42
3.4.2.3	Performance vs. number of training samples	43
3.4.2.4	Rank Sensitivity	44
3.4.2.5	Probability of Detection Vs. SNR	46
3.4.2.6	Complexity Comparison	49
3.4.2.7	Robustness to Nonhomogeneous Training Samples	50
3.4.3	RCML vs. RCML _{LB} and Wax and Kailath Estimator	51
3.5	Conclusion	53

Chapter 4

	Computationally Efficient Toeplitz Approximation under a Rank Constraint	54
4.1	Introduction	54
4.1.1	Motivation and Challenges	55
4.1.2	Our contributions	56
4.2	Efficient Approximation of Structured Covariance	58
4.2.1	Conditions for Eigenvalues of Toeplitz Covariance	59
4.2.2	Exact Toeplitz Solution	61
4.2.3	Toeplitz Approximation	62
4.3	Experimental Investigation	63
4.3.1	Experimental Setup and Methods Compared	63
4.3.2	Whiteness Test	65
4.3.3	Normalized SINR	68
4.3.4	Probability of Detection vs. SNR	69
4.3.5	Complexity Comparison	69
4.4	Conclusion	71

Chapter 5

	Robust Covariance Estimation under Imperfect Constraints using Expected Likelihood Approach	73
5.1	Introduction	73
5.2	Constraints selection method via Expected Likelihood Approach	76
5.2.1	Imperfect rank constraint	76
5.2.2	Imperfect rank and noise power constraints	79
5.2.3	Imperfect condition number constraint	84
5.3	Experimental Validation	94
5.3.1	Experimental setup	94
5.3.2	Imperfect rank constraint	96

5.3.3	Imperfect rank and noise power constraints	99
5.3.4	Imperfect condition number constraint	100
5.4	Conclusion	103
Chapter 6		
	Conclusion	106
6.1	Summary of Contributions	106
6.2	Future Research	107
6.2.1	Newer Constraint: Sparsity or Block Sparsity	107
6.2.2	Non-Homogeneity Detection	107
6.2.3	Expansion of the EL Approach	108
Bibliography		109

List of Figures

1.1	Knowledge-base processing algorithms	1
2.1	The target and interference scenario in an airborne radar.	8
2.2	The procedure of the ITAM algorithm.	16
3.1	Eigenspectra of the estimated covariance matrices.	41
3.2	Normalized SINR vs. normalized azimuthal angle and doppler frequency.	43
3.3	Normalized SINR vs. normalized azimuthal angle and doppler frequency.	44
3.4	Normalized SINR vs. the number of training samples.	45
3.5	Normalized SINR of rank constrained maximum likelihood (RCML) for various rank information	46
3.6	Probability of detection vs. SNR.	47
3.7	Normalized SINR vs. target intensity α and percentage corruption P_t	49
3.8	TRD($\hat{\mathbf{R}}$) vs. target intensity α and percentage corruption P_t	52
3.9	Normalized SINR vs. normalized azimuthal angle.	52
4.1	Trace deviation measure vs. the number of training samples.	66
4.2	Normalized SINR versus the number of training samples for the simulation model	66
4.3	Normalized SINR versus azimuthal angle Doppler frequency for KASSPER data set.	67
4.4	Probability of detection vs. SNR for simulation model via normalized matched filter (NMF) test.	70
4.5	Probability of detection vs. SNR for KASSPER data set via normalized matched filter (NMF) test.	71
5.1	$\left(\log \left(\text{LR}(\mathbf{R}_{\text{RCML}}(r), \mathbf{Z}) / \text{LR}_0 \right) \right)^2$ versus r for KASSPER dataset ($K = 2N = 704$)	76
5.2	The value of the coefficient $\frac{\sigma^2}{\lambda_{r+1}} \cdot \exp(\frac{\lambda_{r+1}}{\sigma^2} - 1)$	79
5.3	The LR value versus σ^2 for the simulation model, $N = 20$, $K = 40$, $r = 5$	84

5.4	Normalized SINR in dB versus number of training samples K ($N = 20$) for the simulation model.	96
5.5	Normalized SINR versus azimuthal angle and Doppler frequency for the KASSPER data set.	97
5.6	Normalized SINR versus azimuthal angle and Doppler frequency for the KASSPER data set.	98
5.7	Normalized SINR in dB versus number of training samples K ($N = 20$) for the simulation model.	100
5.8	Normalized SINR versus azimuthal angle and Doppler frequency for the KASSPER data set.	101
5.9	Normalized SINR versus azimuthal angle and Doppler frequency for the KASSPER data set. (a) and (b) for $K = N = 352$, (c) and (d) for $K = 1.5N = 528$, and (e) and (f) for $K = 2N = 704$	102

List of Tables

3.1	KASSPER Dataset-1 parameters	39
3.2	Running time (sec) for various estimators for KASSPER dataset. FML, RCML, and EigC involve eigenvalue decomposition which takes 0.0497 sec.	48
4.1	Running time (sec) for various estimators for simulation model	72
5.1	Normalized SINR for various values of parameters for the simulation model.	104
5.2	Normalized SINR for various values of parameters for the simulation model.	105

Acknowledgments

I would like to thank my advisor, Professor Vishal Monga, for supporting me during these past four years. He is one of the smartest people I know. He always focuses on successful careers of his students and does his best for it. His passion for research has been a driving force of my successful doctoral program. He has taken care of as many things as he can such as research, technical writing, effective presentation, and even personal problems. His influence really has had me grow up not only as a successful researcher but also as a maturer person. He is one of my greatest academic advisors as well as one of my best friends.

I also want to thank other members in my PhD committee: Prof. Constantino Lagoa, Prof. David Miller, and Prof. Jesse Barlow. Their valuable suggestions and comments are helpful to improve the quality of this thesis. I am also thankful to my collaborator who supported me during my PhD program as a project manager, Dr. Muralidhar Rangaswamy, U.S. Air Force Research Laboratory. I thank him for his support and encouragement.

Next, I want also to thank my wonderful lab mates in the Information Processing and Information Laboratory (iPAL): Umamahesh Srinivas, Mu Li, Xuan Mo, Hojjat Mousavi, Yuelong Li, Tiep Huu Vu, Chul Lee, Omar Aldayel, John McKay, and Tiantong Guo for the insightful discussions on many things; I'm really proud to work with them.

I want to thank my faithful fellows in State College Korean Church for their prayers. We had graceful fellowship and shared our troubles and encouraged one another. I will never be able to forget the memories with them. I am indebted to Pastor Jonathan Gwangil Kim for his prayer and helping me keep my faith.

I would like to thank my family. My wife, Jandy, is always dedicated to support me. Her devotion is the most important factor that makes this thesis successfully completed. My mother and parents-in-law give me tremendous support, not only in my graduate study, but also in every aspect of my life. I also thank my sons, Chanyu and Nathan, who make me laugh even when I am frustrated are also energizers for my life.

I can not help thanking God who is my creator. He alone is my rock, my salvation, my fortress. He guides my life and I can do nothing without Him. I confess that this thesis is done by God the Father not by my wisdom.

Dedication

I dedicate this thesis to
my family, my wife, Jandy, and my loving kids, Chanyu and Nathan
for their constant support and unconditional love.
I love you all dearly.

Introduction

1.1 Motivation

Space-time adaptive processing (STAP), joint adaptive processing in the spatial and temporal domains [1–3], is the cornerstone of radar signal processing and creates the ability to suppress interfering signals while simultaneously preserving gain on the desire signal. However, for STAP to be successful, interference statistics, specifically interference covariance matrix, must be estimated from observed signals. In the absence of prior knowledge about the interference environment, a large number of target free disturbance training samples are required to obtain accurate estimates. A compelling challenge emerges because such a large number of homogeneous training samples are generally not available in practice [4]. Therefore, recent research in radar STAP has focused on overcoming this practical issue. Knowledge-based processing which uses a priori information about the radar environment is one of the most popular approaches to this problem. Figure 1.1 shows previous works related to the knowledge-based processing.

A subset of these techniques includes intelligent training selection [5] and spatio-

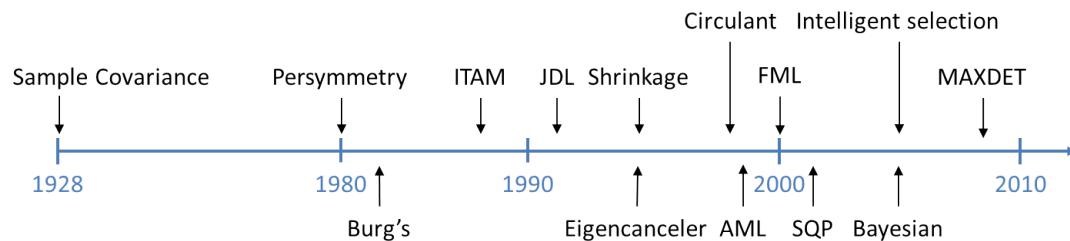


Figure 1.1. Knowledge-base processing algorithms

temporal degree reduction [6–8]. Enforcing structure on covariance matrices, which is the main focus of this dissertation, has also been pursued. Examples of structure include persymmetry [9], Toeplitz structure [10–14], circulant structure [15], and eigenstructure [16–22]. The fast maximum likelihood (FML) method [20] which enforces a special eigenstructure is shown to be the most competitive technique experimentally [8, 23]. In particular, the disturbance covariance matrix represents the clutter matrix which is a low rank and positive semidefinite plus an identity matrix scaled by noise power which represents noise subspace. The FML method ensures that the estimated covariance matrix has eigenvalues all greater than the noise level. Under ideal conditions which means there are no mutual coupling between array elements and no internal clutter motion, the rank of the clutter matrix is easily calculated by the Brennan rule [24]. Even if there is mutual coupling in practice, it is well known that the rank of the clutter matrix is much less than the spatio-temporal degree (the dimension of the covariance matrix).

Since the covariance matrix from a stationary stochastic signal is Hermitian and Toeplitz, estimation of Toeplitz covariance benefits many applications such as array processing and time series analysis. Though various estimation and approximation techniques of Toeplitz covariance matrices have been proposed [25–28], it is well-known that there is no closed-form solution of the ML estimation of a Toeplitz covariance matrix. Therefore, many Toeplitz covariance estimation methods require large number of training samples for computational tractability [11], otherwise it involves iterations or numerical approaches [12]. In particular, the iterated Toeplitz approximation method (ITAM) [14] and the iterative approach by Forster *et al.* [27] deal with both the clutter rank and Toeplitz structure jointly. However, both approaches are based on a computationally expensive iterative procedure. The ITAM estimator has been shown to be effective under very low training, but does not show performance improvements for realistic training regime [29].

Most covariance estimation techniques which exploit the knowledge of radar environment assume that the constraint used in the estimation problem is known or can be estimated from radar physics under ideal condition. For example, the Brennan rule says the rank is easily calculated. However, in practice the clutter rank departs from the Brennan rule prediction due to antenna errors and internal clutter motion. In this case, the rank is not known precisely and needs to be determined. Though determination of the number of signals is a classical eigenvalue problem, it is important to note that the problem does not have a simple and unique solution. A noise level and a condition num-

ber of a disturbance covariance matrix should be estimated as well if they are unknown or not precisely known in practice. The expected likelihood (EL) approach [30] proposed by Abramovich *et al.* provides a novel criterion to determine a regularization parameter based on the statistical invariance property of the likelihood ratio (LR) values. The regularization parameters are selected so that the LR value of the estimate agree as closely as possible with the statistics of the LR value of the true covariance matrix which does not depend on the true covariance itself.

1.2 Overview of Dissertation Contributions

In view of the aforementioned observations, we develop structured covariance estimation methods which exploit practical and powerful constraints, that is, the rank of the clutter matrix and Toeplitz structure. Furthermore, we also develop covariance estimation methods which automatically and adaptively determine the values of practical constraints via an expected likelihood approach. The main contributions of this dissertation are described in detail in the following sections.

1.2.1 Estimation of Structured Covariance Matrices

We first set up the optimization problem to estimate the disturbance covariance matrix with a structural constraint and the rank constraint. The estimation problem is unfortunately not convex, because neither the cost function nor the rank constraint are convex. However, we show that using a transformation of variables, reduction to a convex form is possible. Furthermore, by invoking Karush-Kuhn-Tucker (KKT) [31] for the resulting convex problems, it is possible to derive a closed-form solution. Akin to the FML, we initially assume that the noise power is known. Then we extend our results to the case of unknown noise variance. In that case, we assume that only a lower bound on the noise power is available.

Our second contribution aims to break the classical trade-off between performance under low training regimes and computational complexity. We develop a computationally efficient approximation of structured Toeplitz covariance under a rank (EASTR) constraint. EASTR satisfies both the Toeplitz structure property (at least approximately) and the rank information of the clutter subspace at the same time. Decades of research has shown that enforcing even each constraint individually can be quite onerous. We propose to decouple the rank and Toeplitz constraints, which lends analytical tractability. Crucially, the EASTR solution does not need iterative steps, that is, in-

volves a cascade of two steps in which a closed form solution is available in each step. First, a closed form solution using ML employing the rank constraint, is obtained from the RCML estimator. Next we propose a new method to perturb the eigenvalues of the RCML estimator in a rank preserving manner so as to impose the Toeplitz structure. The merits of the RCML estimator and EASTR are also verified experimentally over both simulated data and the knowledge-aided sensor signal processing expert reasoning (KASSPER) data set.

1.2.2 Estimation of Constraints under Imperfect Knowledge

We develop covariance estimation methods which automatically and adaptively determine the values of practical constraints via an expected likelihood approach for practical radar STAP. The proposed methods guide the selection of the constraints via the expected likelihood criteria in the case that the knowledge of the constraints is imperfectly known in practice. We consider three different cases of the constraints: 1) only the clutter rank constraint, 2) both the clutter rank and the noise power constraints, and 3) the condition number constraint. For each case, we develop significant analytical results. First we formally prove that the rank selection problem has a unique solution. Second, we derive a closed form solution of the optimal noise power for a given rank, which means an iterative or numerical method is not required to find the optimal noise power and enables fast implementation. Finally we also prove there exists the unique condition number for the condition number estimation method. Experimental investigation on a simulation model and on the KASSPER data set shows that the proposed methods for three different cases outperform alternatives such as the FML, leading rank selection methods in radar literature and statistics, and the ML estimation of the condition number constraint.

1.3 Organization

A snapshot of the main contributions of this dissertation is presented next. Publications related to the contribution in each chapter are also listed where applicable.

Chapter 2 discusses background of radar space time adaptive processing (STAP) and introduce previous works for knowledge-based processing in radar applications. First, significance and benefit of radar STAP which is main application of the proposed covariance estimation methods are presented. We also discuss the practical limitations of radar STAP and approaches to overcome the limitation. One of the approaches is to

exploit knowledge of radar environment. In particular, many researchers have enforced structure on covariance and shrinkage estimation techniques have also been considered. A more thorough review of these techniques is provided. We introduce an expected likelihood approach which is useful to determine parameters in optimization problem. Our third contribution, the method of automatic selection of constraints, is based on this approach.

In **Chapter 3**, we propose a novel covariance estimation method which exploits the rank of the clutter subspace, rank constrained maximum likelihood (RCML) estimation. We formulate the RCML estimation problem, and subsequent derivation is provided. The solutions of the estimation and optimization problems are derived for the two cases of both known and unknown (known lower bound) noise levels. Experimental validation reports on the performance of the proposed estimator and compares it against well-known existing methods in terms of signal to interference and noise ratio. In addition, we discuss practical merits of the proposed method, such as probability of detection and whiteness test. The material related to this method was presented at the 2012 IEEE International Radar Conference [32] and the 2013 IEEE International Radar Conference [33] and was published in the IEEE Transactions on Aerospace and Electronic Systems [34].

Our second contribution presented in **Chapter 4** is towards breaking a classical trade-off between computational complexity and performance in low or realistic training regimes for estimation problem of Toeplitz disturbance covariance matrix. We develop a computationally efficient approximation of structured Toeplitz covariance under a rank constraint (EASTR). We consider two cases: 1) when the Toeplitz constraint is satisfied exactly, we obtain the exact Toeplitz estimate, satisfying the rank constraint and Toeplitz property and 2) when the Toeplitz constraint is not exactly satisfied, we make slight modification on the Toeplitz and obtain approximately the Toeplitz estimate. Experimental validation of the proposed method is also provided to compare the proposed method against widely used existing radar STAP covariance estimators. This material was presented at the 2013 IEEE Asilomar Conference on Signals, Systems and Computers [35], the 2013 IEEE 5th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP) [36], and the 2014 IEEE International Radar Conference [37] and was published in the IEEE Transactions on Aerospace and Electronic Systems [29] and in the IEEE Aerospace and Electronic Systems Magazine [38].

Chapter 5 presents our third contribution considering a robust covariance estimation method under imperfectly known constraints. We propose a robust covariance estimation method via an expected likelihood approach and solutions of the constraint estimation

problems for three different cases of the constraints. Experimental validation of our proposed methods is provided to report the performance of the proposed methods and compare them against existing parameter selection such as rank selection and maximum likelihood estimators on both a simulation model and the KASSPER data set. The material was presented at the 2015 IEEE International Radar Conference [39] and has been submitted to the IEEE Transactions on Aerospace and Electronic Systems in May 2015 [40].

Background

2.1 Space-Time Adaptive Processing (STAP)

Signal detection using adaptive processing in both spatial and temporal domains offers significant benefits in a variety of applications including radar, sonar, satellite communications, and seismic systems [1]. Specifically, the space-time adaptive processing (STAP) enables to suppress interference while preserving gain on desired signal. Consider the airborne array radar with J elements. The radar transmits a pulse in a certain direction and the transmitted signal is reflected by various objects such as buildings, land, water, vegetation, and one or more targets of interest. On receive, the radar samples this reflected wave at a high rate, with each of the K samples corresponding directly to reflections from a specific range. The received signal may include not only desired signals but also undesired interfering effects from extraneous objects (clutter) and electronic counter-measures (ECM), such as jamming. In addition, the background white noise caused by the radar receiver circuitry as well as man-made sources and machinery is also included in the received returns. This process is repeated for P pulses transmitted. Therefore the entire received data can be $J \times P \times K$ data cube.

More precisely, the radar receiver front end consists of an array of J antenna elements, which receives signals from targets, clutter, and jammers. These reflections induce a voltage at each element of the antenna array, which constitutes the measured array data at a given time instant. Snapshots of the measured data collected at P successive time epochs give rise to the spatio-temporal nature of the received radar data. The spatio-temporal product $JP = N$ is defined to be the system dimensionality. Figure 2.1 uses the angle-Doppler space to illustrate the need for space-time adaptive processing (STAP). A

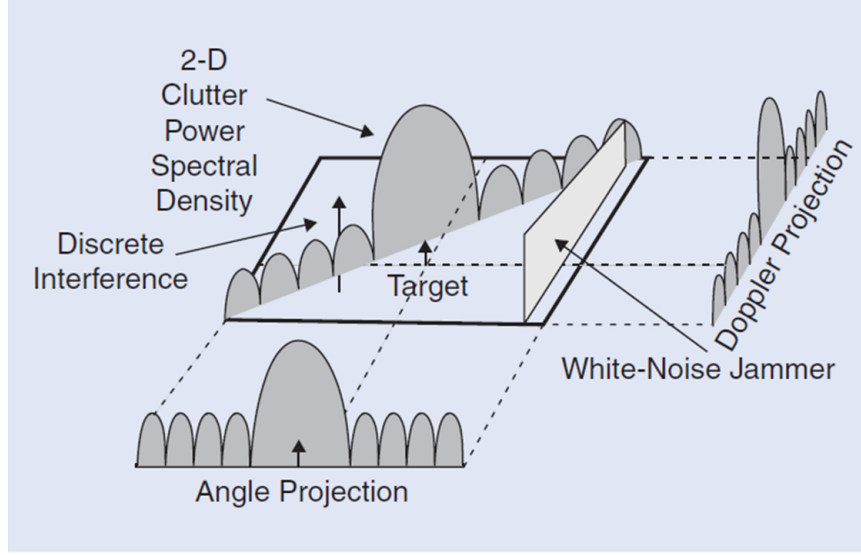


Figure 2.1. The target and interference scenario in an airborne radar.

target at a specific angle and traveling at a specific velocity (corresponding to a Doppler frequency) occupies a single point in this space. A jammer originates from a particular angle but is temporally (in Doppler domain) white and the clutter occupies a ridge in this 2-D space. Consequently the target signal is masked by white jammer in Doppler domain and the clutter in spatial domain and therefore, carrying out merely temporal (Doppler domain) or spatial (angle domain) processing fails to separate the target from the interference [24]. On the other hand, joint domain processing in angle and Doppler enables target detection as shown in Figure 2.1.

The target detection problem can be cast in the framework of a statistical hypothesis test of the form

$$H_0 : \mathbf{x} = \mathbf{d} = \mathbf{c} + \mathbf{j} + \mathbf{n} \quad (2.1)$$

$$H_1 : \mathbf{x} = \alpha \mathbf{s}(\theta_t, f_t) + \mathbf{d} = \alpha \mathbf{s}(\theta_t, f_t) + \mathbf{c} + \mathbf{j} + \mathbf{n} \quad (2.2)$$

where $\mathbf{x} \in \mathbb{C}^{JP \times 1}$ is a vector form of the received data under either hypothesis, \mathbf{d} represents the overall disturbance which is the sum of \mathbf{c} , clutter, \mathbf{j} , jammers, and \mathbf{n} , the background white noise. The vector \mathbf{s} is a known spatio-temporal steering vector that represents the signal returned from the target for a specific angle and Doppler and α is the unknown target complex amplitude. The steering vector \mathbf{s} is defined in the case of

an equally spaced linear array by

$$\mathbf{s} = \mathbf{s}_t \otimes \mathbf{s}_s \quad (2.3)$$

$$\mathbf{s}_t = [1 \ z_t \ z_t^2 \ \dots \ z_t^{(N-1)}]^T, \quad (2.4)$$

$$\mathbf{s}_s = [1 \ z_s \ z_s^2 \ \dots \ z_s^{(J-1)}]^T, \quad (2.5)$$

$$\mathbf{z}_s = e^{j2\pi f_s} = e^{(j2\pi \frac{d}{\lambda} \sin \phi_t)}, \quad z_t = e^{j2\pi f_t / f_R}, \quad (2.6)$$

where ϕ_t and f_t are the angle and Doppler frequency respectively, and \otimes denotes the Kronecker vector product, f_R the pulse repetition frequency (PRF), and λ the wavelength of operation. Finally, \mathbf{s}_t and \mathbf{s}_s represent the temporal and spatial steering vectors, respectively.

Now, a test statistic Λ using a weight vector \mathbf{w} is

$$\Lambda = |\mathbf{w}^H \mathbf{x}|^2 \underset{H_0}{\overset{H_1}{\gtrless}} \lambda, \quad (2.7)$$

where H denotes the Hermitian transpose and λ a *threshold* which is determined by the probability of false alarm. The whiten-and-match filter (MF) for detecting a rank-1 signal is the optimum processing method for Gaussian interference statistics. It is given by [41]

$$\mathbf{w} = \frac{\mathbf{R}_d^{-1} \mathbf{s}}{\sqrt{\mathbf{s}^H \mathbf{R}_d^{-1} \mathbf{s}}} \Rightarrow \Lambda_{MF} = \frac{|\mathbf{s}^H \mathbf{R}_d^{-1} \mathbf{x}|^2}{\mathbf{s}^H \mathbf{R}_d^{-1} \mathbf{s}} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda_{MF} \quad (2.8)$$

where \mathbf{R}_d is a known interference covariance matrix. Eq. (2.8) represents the matched filtering of the whitened data $\check{\mathbf{x}} = \mathbf{R}_d^{-1/2} \mathbf{x}$ and whitened steering vector $\check{\mathbf{s}} = \mathbf{R}_d^{-1/2} \mathbf{s}$. From Eq. (2.8), it turns out that the interference covariance matrix \mathbf{R}_d is crucial in the detection statistic.

2.2 Estimation of Covariance Matrices

2.2.1 Structured Covariance Estimation

To overcome the practical issue of lack of generous training data, researchers have developed approaches using signal processing and statistical learning techniques, covariance matrix estimation techniques that enforce and exploit particular structure have been pursued. In this section, we discuss representative algorithms, the fast maximum likelihood (FML), shrinkage estimation methods, and eigencanceler.

Steiner and Gerlach proposed a maximum likelihood (ML) solution for a structured

covariance matrix that has the form of the identity matrix plus an unknown positive semi-definite Hermitian (PSDH) matrix under the assumption that the input interference is Gaussian [20]. In particular, the disturbance covariance matrix \mathbf{R} represents the exhibits the following structure

$$\mathbf{R} = \sigma^2 \mathbf{I} + \mathbf{R}_c \quad (2.9)$$

where \mathbf{R}_c denotes the clutter matrix which has a low rank and is positive semi-definite and \mathbf{I} is an identity matrix. This covariance matrix form is often valid in realistic interference scenarios for radar and communication systems.

The ML estimate for the covariance matrix is given by

$$\hat{\mathbf{R}} = \arg \min_{\mathbf{R}_c} [tr\{\mathbf{R}^{-1}\mathbf{S}\} + \log(|\mathbf{R}|)] \quad (2.10)$$

Consider an eigenvalue decomposition of the sample covariance matrix (SCM), $\mathbf{S} = \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^H$, where $\mathbf{\Lambda}$ is an $N \times N$ diagonal matrix with diagonal entries $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ that are eigenvalues of \mathbf{S} and $\mathbf{\Phi}$ is an $N \times N$ unitary eigenmatrix. Then the ML solution for \mathbf{R} is given by

$$\hat{\mathbf{R}} = \mathbf{\Phi}\mathbf{\Lambda}_0\mathbf{\Phi}^H \quad (2.11)$$

where $\mathbf{\Phi}$ is the unitary eigenmatrix associated with the eigenvalue decomposition of the SCM,

$$\mathbf{\Lambda}_0 \equiv \text{Diag}(\lambda_1, \lambda_2, \dots, \lambda_M, \sigma^2, \sigma^2, \dots, \sigma^2) \quad (2.12)$$

is an $N \times N$ diagonal matrix, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ are the eigenvalues of the SCM, and M is the number of eigenvalues greater than σ^2 . Therefore, the FML technique ensures that the estimated covariance matrix has eigenvalues all greater than σ^2 by assuming that its value is known, which is sometimes unrealistic.

Shrinkage estimators are commonly used approach to the problem of covariance estimation for high dimensional data [42]. In general, shrinkage estimators shrink the sample covariance matrix to a target structure

$$\hat{\mathbf{R}} = \alpha \mathbf{D} + (1 - \alpha) \mathbf{S} \quad (2.13)$$

where \mathbf{D} and \mathbf{S} are a positive definite matrix and the sample covariance matrix, respectively.

There are several methods of shrinkage estimators according to how to define the positive definite matrix \mathbf{D} . For instance, \mathbf{D} is considered as the identity matrix in some

papers, that is,

$$\hat{\mathbf{R}} = \alpha \mathbf{I} + (1 - \alpha) \mathbf{S} \quad (2.14)$$

This approach is usually used in Ridge regression and Tikhonov regularization. Second, others refer \mathbf{D} as an scaled identity matrix by the average of the eigenvalues,

$$\hat{\mathbf{R}} = \alpha \left(\frac{\text{tr}(\mathbf{S})}{N} \right) \mathbf{I} + (1 - \alpha) \mathbf{S} \quad (2.15)$$

Finally, \mathbf{D} can be set to the diagonal of \mathbf{S} ,

$$\hat{\mathbf{R}} = \alpha \text{diag}(\mathbf{S}) + (1 - \alpha) \mathbf{S} \quad (2.16)$$

Here, the shrinkage intensity parameter α is chosen by the leave-one-out likelihood (LOOC), specifically, it is chosen so that the average likelihood of omitted samples is maximized as suggested in [16].

The eigencanceler [19] provides simultaneous rejection of both clutter and directional interferences by adaptive processing in the spatial and Doppler domains. The eigencanceler uses eigendata to suppress clutter and directional interferences while minimizing noise contributions and maintaining specified beam pattern constraints. They first consider the space-time correlation matrix as the form of

$$\mathbf{R} = E\{\mathbf{x}\mathbf{x}^H\} \quad (2.17)$$

$$= \mathbf{R}_J + \mathbf{R}_c + \mathbf{M}_\nu \quad (2.18)$$

where \mathbf{R}_J , \mathbf{R}_c , \mathbf{R}_ν are the correlation matrices of the jammers, clutter, noise, respectively and \mathbf{x} are the received signals. Each of the correlation matrices can be then written by

$$\mathbf{R}_J = \sum_{\theta_i} \int_{B_\nu} S_{J,i}(\nu) \mathbf{v}(\theta_i, \nu) \mathbf{v}^H(\theta_i, \nu) d\nu \quad (2.19)$$

$$\mathbf{R}_C = \int_{\Theta} \int_{B_\nu} S_C(\theta, \nu) \mathbf{v}(\theta, \nu) \mathbf{v}^H(\theta, \nu) d\nu d\theta \quad (2.20)$$

$$\mathbf{R}_\nu = \sigma_\nu^2 \mathbf{I} \quad (2.21)$$

where \mathbf{v} is the position vector on the angle θ_i and Doppler frequency ν and S is the power spectral density.

The eigencanceler is based on the eigenanalysis which suggests a small number of eigenvalues contain all the information about interferences (jammers and clutter), and

therefore, the span of the eigenvectors associated with these significant eigenvalues includes all the position vectors that comprise the interference signals. Under assumption that r dominant eigenvalues and eigenvectors represent interferences, the covariance matrix can be expressed by

$$\mathbf{R} = \sum_{i=1}^r p_i \mathbf{v}_i \mathbf{v}_i^H + \sigma^2 \mathbf{I} \quad (2.22)$$

where p_i and \mathbf{v}_i are the clutter power and the eigenvector corresponding to r dominant eigenvalues, respectively.

Aubry *et al.* proposed the method of a structured covariance matrix under a condition number upper-bound constraint [22]. The initial non-convex optimization problem is

$$\left\{ \begin{array}{l} \max_{\mathbf{R}} \quad f(\mathbf{Z}) = \frac{1}{\pi^{NK} |\mathbf{R}|^K} \exp(-\text{tr}\{\mathbf{Z}^H \mathbf{R}^{-1} \mathbf{Z}\}) \\ s.t. \quad \mathbf{R} = \sigma^2 \mathbf{I} + \mathbf{R}_c \\ \quad \frac{\lambda_{\max}(\mathbf{R})}{\lambda_{\min}(\mathbf{R})} \leq K_{\max} \\ \quad \mathbf{R}_c \succeq \mathbf{0} \\ \quad \sigma^2 \geq c \end{array} \right. \quad (2.23)$$

The authors showed that the optimization problem falls within the class of MAXDET problems [43,44] and developed an efficient procedure for its solution in closed form which is given by

$$\mathbf{R}^* = \mathbf{V} \mathbf{\Lambda}^{*-1} \mathbf{V}^H \quad (2.24)$$

where

$$\mathbf{\Lambda}^* = \text{diag}(\boldsymbol{\lambda}^*(\bar{u})), \quad (2.25)$$

$\boldsymbol{\lambda}^*(\bar{u}) = [\lambda_1^*(\bar{u}), \dots, \lambda_N^*(\bar{u})]$ with

$$\lambda_i^*(\bar{u}) = \min \left(\min(K_{\max} \bar{u}, 1), \max \left(\bar{u}, \frac{1}{\bar{d}_i} \right) \right), \quad (2.26)$$

K_{\max} is a condition number constraint, and \bar{u} is an optimal solution of the following optimization problem,

$$\left\{ \begin{array}{l} \min_u \quad \sum_{i=1}^N G_i(u) \\ s.t. \quad 0 < u \leq 1 \end{array} \right. \quad (2.27)$$

where

$$G_i(u) = \begin{cases} \log K_{\max} - \log u + K_{\max} \bar{d}_i u & \text{if } 0 < u \leq \frac{1}{K_{\max}} \\ \bar{d}_i & \text{if } \frac{1}{K_{\max}} \leq u \leq 1 \end{cases} \quad (2.28)$$

for $\bar{d}_i \leq 1$, and

$$G_i(u) = \begin{cases} \log K_{\max} - \log u + K_{\max} \bar{d}_i u & \text{if } 0 < u \leq \frac{1}{K_{\max} \bar{d}_i} \\ \log \bar{d}_i + 1 & \text{if } \frac{1}{K_{\max} \bar{d}_i} < u \leq \frac{1}{\bar{d}_i} \\ \frac{1}{\bar{d}_i} + \bar{d}_i u & \text{if } \frac{1}{\bar{d}_i} \leq u \leq 1 \end{cases}$$

for $\bar{d}_i > 1$. Similarly to the RCML estimator, the ML solution of the eigenvalue is a function of \bar{d}_i 's and the condition number K_{\max} .

2.2.2 Toeplitz Covariance Estimation

Many Toeplitz covariance estimation methods have been published in the literature on structured covariance matrix estimation. An important technique is the maximum likelihood (ML) approach. However, it is well known that the ML estimation of a Hermitian Toeplitz covariance matrix has no closed-form solution [45]. Some previous works consider only the Toeplitz matrix constraint [10, 11, 25]. Li *et al.* proposed asymptotic maximum likelihood (AML) estimation and its closed form solution [11]. However, they did not consider the rank constraint and the closed form solution is applied only when the number of sample data is large enough. On the other hand, Al-Homidan proposed an algorithm to find the nearest Toeplitz matrix with rank r given a certain matrix by using LDL^T decomposition [12]. ITAM (Iterated Toeplitz Approximation Method) [14] was proposed to find the approximate rank r Toeplitz matrix under structural constraint but it assumes the signal is harmonic with complex exponentials which we do not assume necessarily. We briefly discuss Toeplitz covariance estimation algorithms in this section.

Burg *et al.* proposed a method for estimating a covariance matrix of specified structure from vector samples of the random process [10]. It appears to be the first to study the ML method in its full generality. They assumed that the random process is zero-mean multivariate Gaussian and found the maximum-likelihood covariance matrix that has the specified structure. They used the fact that the gradient must be orthogonal to variations in a feasible set. To derive the necessary conditions, they define the variation of \mathbf{R} to be

$$\partial \mathbf{R} = \begin{pmatrix} \partial \mathbf{R}(1,1) & \partial \mathbf{R}(1,2) & \cdots & \partial \mathbf{R}(1,N) \\ \partial \mathbf{R}(2,1) & \partial \mathbf{R}(2,2) & \cdots & \partial \mathbf{R}(2,N) \\ \vdots & \vdots & \ddots & \vdots \\ \partial \mathbf{R}(N,1) & \partial \mathbf{R}(N,2) & \cdots & \partial \mathbf{R}(N,N) \end{pmatrix} \quad (2.29)$$

Then, the variation of the determinant of \mathbf{R} in terms of the variation of \mathbf{R} will be

$$\partial|\mathbf{R}| = |\mathbf{R}|\text{tr}(\mathbf{R}^{-1}\partial\mathbf{R}) \quad (2.30)$$

From Eq. (2.30),

$$\partial \log |\mathbf{R}| = \text{tr}(\mathbf{R}^{-1}\partial\mathbf{R}). \quad (2.31)$$

We have another useful matrix theorem. From $\mathbf{R}\mathbf{R}^{-1} = \mathbf{I}$,

$$\partial\mathbf{R}^{-1} = -\mathbf{R}^{-1}\partial\mathbf{R}\mathbf{R}^{-1} \quad (2.32)$$

Now the variation of the cost function is derived by

$$\begin{aligned} \partial \log(|\mathbf{R}|) - \partial \text{tr}(\mathbf{R}^{-1}\mathbf{S}) &= -\text{tr}(\mathbf{R}^{-1}\partial\mathbf{R}) - \text{tr}[\partial(\mathbf{R}^{-1})\mathbf{S}] \\ &= -\text{tr}(\mathbf{R}^{-1}\partial\mathbf{R} - \mathbf{R}^{-1}\partial\mathbf{R}\mathbf{R}^{-1}\mathbf{S}) \\ &= -\text{tr}(\mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1}\partial\mathbf{R} - \mathbf{R}^{-1}\partial\mathbf{R}) \end{aligned}$$

The expression is thus neatly written as

$$\text{tr}[(\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1})\partial\mathbf{R}] \quad (2.33)$$

The condition for minimization is that the variation of the cost function is zero for any feasible variation of \mathbf{R} . Thus the equation to be solved is

$$\text{tr}[(\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1})\partial\mathbf{R}] = 0 \quad (2.34)$$

Then they used the inverse iteration algorithm to find the estimate of \mathbf{R} . However, we have to find the basis in the feasible set which is the set of all of Hermitian Toeplitz matrices with rank r in order to perform the iteration.

In the case of a signal comprised only of complex exponentials, the true $N \times N$ covariance matrix \mathbf{R}_S will have the additional property that its rank is equal to r , the number of complex exponentials in the signal. If white noise having a variance of σ^2 is added to this signal, the covariance matrix becomes

$$\mathbf{R} = \mathbf{R}_S + \mathbf{R}_N \quad (2.35)$$

where $\mathbf{R}_N = \sigma^2\mathbf{I}$. ITAM is a heuristically appealing technique, that uses repeated application of the singular value decomposition (SVD) to produce a rank r Hermitian

Toeplitz estimate of \mathbf{R}_S .

For the harmonic retrieval problem, the observed signal is assumed to have the form of a sum of r complex exponentials in white noise

$$x(n) = \sum_{i=1}^r A_i e^{jw_i n} + w(n) \quad (2.36)$$

The noise $w(n)$ is assumed to be zero mean and have a variance of σ^2 . The goal of Toeplitz approximation is to find the best model $\hat{\mathbf{R}}$ for \mathbf{R} where the model is constrained to have the form

$$\hat{\mathbf{R}} = \mathbf{R}_S + \sigma^2 \mathbf{I} \quad (2.37)$$

where

$$\mathbf{R}_S = \sum_{i=1}^r P_i \mathbf{q}_i \mathbf{q}_i^H \quad (2.38)$$

$$\mathbf{q} = [1 \ e^{j\hat{w}_i} \ e^{j2\hat{w}_i} \ \dots \ e^{j(N-1)\hat{w}_i}]^T \quad (2.39)$$

First estimate $\hat{\mathbf{R}}_S$ from the observed covariance matrix \mathbf{R} .

$$\mathbf{R} = \mathbf{U} \mathbf{\Sigma} \mathbf{U}^H \quad (2.40)$$

$$\mathbf{U} = [\mathbf{U}_1 \ \mathbf{U}_2] \quad (2.41)$$

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_2 \end{bmatrix} \quad (2.42)$$

$$\hat{\mathbf{R}}_S = \mathbf{U}_1 (\mathbf{\Sigma}_1 - \lambda_{av} \mathbf{I}) \mathbf{U}_1^H \quad (2.43)$$

where

$$\lambda_{av} = \frac{1}{M-r} \sum_{i=r+1}^N \lambda_i \quad (2.44)$$

Now, since $\hat{\mathbf{R}}_S$ is not in general a Toeplitz matrix, the next step is to find the Toeplitz matrix, \mathbf{T} , that is closest to $\hat{\mathbf{R}}_S$. This is done by averaging along the diagonals of $\hat{\mathbf{R}}_S$ and using these average values to define the corresponding diagonals of the new Toeplitz matrix, $\hat{\mathbf{T}}$. The problem now is that $\hat{\mathbf{T}}$ does not have a rank of p , so the next iteration begins by finding the rank p approximation to $\hat{\mathbf{T}}$. This establishes the ITAM algorithm [14] which is illustrated in Figure 2.2.

Li *et al.* proposed a computationally efficient method that provides asymptotic (for large samples) maximum likelihood (AML) estimation for structured covariance matrix-

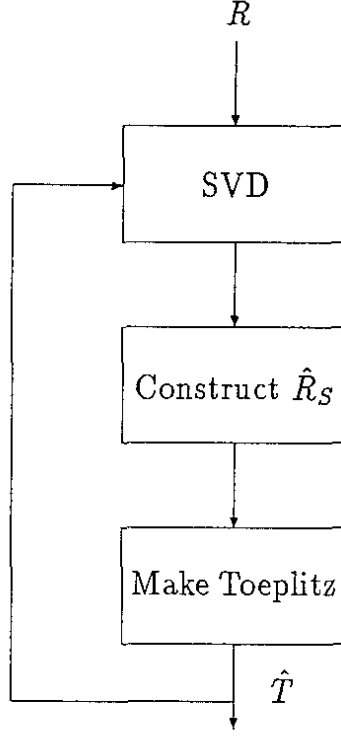


Figure 2.2. The procedure of the ITAM algorithm.

ces [11]. They provided a closed-form formula for estimating Hermitian Toeplitz covariance matrices by invoking the extended invariance principle (EXIP) [46] for parameter estimation. They denote Hermitian Toeplitz covariance matrix $\mathbf{R}(\phi)$ that is a known function of an unknown parameter vector $\phi \in \mathbb{R}^{(2N-1) \times 1}$. Usually, a Hermitian Toeplitz matrix is parameterized such that ϕ consists of the real and imaginary parts of the first column (or row) of \mathbf{R} .

The exact ML estimate $\hat{\phi}$ of ϕ is obtained by maximizing the likelihood function, which is equivalent to

$$\hat{\phi} = \arg \min_{\phi \in \mathbb{R}^{(2N-1) \times 1}} L_{\phi}(\phi) = \ln |\mathbf{R}(\phi)| + \text{tr}[\mathbf{R}^{-1}(\phi)\mathbf{S}] \quad (2.45)$$

Let $\mathbf{r} = \text{vec}(\mathbf{R}) \in \mathbb{C}^{N^2 \times 1}$ and $\gamma \in D_{\gamma} \subset \mathbb{R}^{N^2 \times 1}$ denote the vector made from the real and imaginary parts of the elements of \mathbf{R} both above and on the main diagonal. Evidently, there is an $N^2 \times N^2$ matrix \mathbf{F} such that

$$\gamma = \mathbf{F}\mathbf{r} \quad (2.46)$$

Note that γ represents a reparameterization of the covariance matrix \mathbf{R} originally parameterized by ϕ . Hence there exists a mapping from ϕ to γ , i.e.,

$$\gamma = f(\phi) \in D_\gamma, \quad \forall \phi \in D_\phi \quad (2.47)$$

such that

$$L_\gamma(f(\phi)) = L_\phi(\phi), \quad \forall \phi \in D_\phi \quad (2.48)$$

Let $\tilde{\gamma} = \mathbf{F}\text{vec}(\mathbf{S})$. Then

$$\tilde{\phi} = \arg \min_{\phi \in D_\phi} [\tilde{\gamma} - f(\phi)]^T \mathbf{\Gamma}^{-1} [\tilde{\gamma} - f(\phi)] \quad (2.49)$$

is an asymptotically (in K) valid approximation of the ML estimate $\hat{\phi}$, where $\mathbf{\Gamma}$ is a consistent (in K) estimate of the covariance matrix of $\tilde{\gamma}$.

Let $\tilde{\mathbf{r}} = \text{vec}(\tilde{\mathbf{R}})$ where $\tilde{\mathbf{R}}$ is a sample covariance matrix. Then $\tilde{\gamma} = \mathbf{F}\tilde{\mathbf{r}}$. Since $\mathbf{\Gamma} = \mathbf{F}\mathbf{C}\mathbf{F}^H$ where $\mathbf{C} = \text{cov}(\tilde{\mathbf{r}})$, Eq. (2.49) is equivalent to

$$\tilde{\phi} = \arg \min_{\phi \in D_\phi} [\tilde{\mathbf{r}} - \mathbf{r}(\phi)]^H \mathbf{C}^{-1} [\tilde{\mathbf{r}} - \mathbf{r}(\phi)] \quad (2.50)$$

Finally, they obtain the final formula of AML for estimating a Hermitian Toeplitz covariance matrix.

$$\hat{\phi} = (\mathbf{\Psi}^H \tilde{\mathbf{C}}^{-1} \mathbf{\Psi})^{-1} (\mathbf{\Psi}^H \tilde{\mathbf{C}}^{-1} \tilde{\mathbf{r}}) \quad (2.51)$$

Al-Homidan proposed an algorithm that finds the nearest symmetric positive semi-definite Toeplitz matrix \mathbf{T} to given a matrix \mathbf{F} [12]. The problem is formulated as a nonlinear minimization problem with positive semi-definite Toeplitz matrix as constraints. An algorithm with rapid convergence is obtained by l_1 Sequential Quadratic Programming (SQP) method.

Consider the following problem: Given a data matrix $\mathbf{F} \in \mathbb{R}^{N \times N}$, find the nearest symmetric positive semi-definite Toeplitz matrix \mathbf{T} to \mathbf{F} and $\text{rank}(\mathbf{T}) = r$, that is,

$$\begin{aligned} \min \quad & \phi = \|\mathbf{F} - \mathbf{T}\| \\ \text{subject to} \quad & \mathbf{T} \in T, \quad \text{rank}(\mathbf{T}) = r \end{aligned} \quad (2.52)$$

Using partial LDL^T factorization of \mathbf{T} , the partial factors $\mathbf{T} = \mathbf{L}\mathbf{D}\mathbf{L}^T$ can be calculated

such that

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_{11} & \mathbf{0} \\ \mathbf{L}_{21} & \mathbf{I}_{N-r} \end{bmatrix}, \mathbf{D} = \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{bmatrix}, \mathbf{T} = \begin{bmatrix} \mathbf{T}_{11} & \mathbf{T}_{21}^T \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{bmatrix} \quad (2.53)$$

where \mathbf{L}_{11} , \mathbf{D}_1 and \mathbf{T}_{11} are $r \times r$ matrices; \mathbf{D}_2 and \mathbf{T}_{22} are $(n-r) \times (n-r)$ matrices; \mathbf{L}_{21} and \mathbf{T}_{21} are $(n-r) \times r$ matrices; \mathbf{D}_1 is diagonal and positive definite and $\mathbf{D}_2 = \mathbf{0}$. Since

$$\mathbf{D}_2(\mathbf{T}) = \mathbf{T}_{22} - \mathbf{T}_{21} \mathbf{T}_{11}^{-1} \mathbf{T}_{21}^T \quad (2.54)$$

and if the structure of the matrix \mathbf{T} is in a Toeplitz form, i.e.,

$$\mathbf{T} = \begin{bmatrix} x_1 & x_2 & \cdots & x_N \\ x_2 & x_1 & \cdots & x_{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_N & x_{N-1} & \cdots & x_1 \end{bmatrix} \quad (2.55)$$

the constraint $\mathbf{T} \in T$ to be written in the form

$$\mathbf{D}_2(\mathbf{T}(\mathbf{x})) = \mathbf{0} \quad (2.56)$$

Hence, the optimization problem can be expressed as

$$\begin{aligned} \min \quad & \phi \\ \text{subject to} \quad & \mathbf{D}_2(\mathbf{T}) = \mathbf{0} = \mathbf{Z}^T \mathbf{T} \mathbf{Z} \end{aligned} \quad (2.57)$$

where

$$\mathbf{Z} = \begin{bmatrix} -\mathbf{T}_{11}^{-1} \mathbf{T}_{21}^T \\ \mathbf{I} \end{bmatrix}$$

Now the cost function is

$$\begin{aligned} \phi &= \sum_{i,j=1}^N (f_{ij} - t_{ij})^2 \\ &= \sum_{i,j=1}^N (f_{ij} - x_{|i-j+1|})^2 \end{aligned}$$

and the constraints $\mathbf{D}_2(\mathbf{T}) = \mathbf{0}$ can be written in the following form:

$$d_{ij}(\mathbf{x}) = x_{|i-j+1|} - \sum_{k,l=1}^r x_{i-k+1} [\mathbf{T}_{11}^{-1}]_{kl} x_{j-l+1} = 0 \quad (2.58)$$

where $i, j = r+1, \dots, N$ and $[\mathbf{T}_{11}^{-1}]_{kl}$ denotes the element of \mathbf{T}_{11}^{-1} in kl -position. Thus

$$\begin{aligned} \min_{\phi} \quad & \phi = \sum_{i,j=1}^N (f_{ij} - x_{|i-j+1|})^2 \\ \text{subject to} \quad & d_{ij}(\mathbf{x}) = 0 \end{aligned} \quad (2.59)$$

the SQP method applied to above problem requires the solution of the QP subproblem

$$\begin{aligned} \min_{\boldsymbol{\delta}} \quad & \phi + \nabla \phi^T \boldsymbol{\delta} + \frac{1}{2} \boldsymbol{\delta}^T \mathbf{W} \boldsymbol{\delta} \quad \boldsymbol{\delta} \in \mathbb{R}^r \\ \text{subject to} \quad & d_{ij} + \nabla d_{ij}^T \boldsymbol{\delta} = 0 \quad i, j = r+1, \dots, N \end{aligned} \quad (2.60)$$

giving a correction vector $\boldsymbol{\delta}^{(k)}$, so that $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \boldsymbol{\delta}^{(k)}$.

2.3 Expected Likelihood Approach

Abramovich *et al.* [30] proposed the expected likelihood estimation criterion which inherently justifies the appropriate selection of parameters such as loading factor based on direct likelihood matching. Expected likelihood approach is motivated by invariance properties of the likelihood ratio (LR) value which is given by

$$\text{LR}(\mathbf{R}, \mathbf{Z}) \equiv \left[\frac{f(\mathbf{Z}|\mathbf{R})}{f(\mathbf{Z}|\mathbf{S})} \right]^{1/K} \quad (2.61)$$

$$= \frac{|\mathbf{R}^{-1}\mathbf{S}| \exp N}{\exp[\text{tr}(\mathbf{R}^{-1}\mathbf{S})]} \quad (2.62)$$

under a Gaussian assumption on the observations, \mathbf{z}_i 's. Furthermore, the unconstrained ML solution \mathbf{S} has the LR value of 1. That is,

$$\max_{\mathbf{R}} \text{LR}(\mathbf{R}, \mathbf{Z}) = \text{LR}(\mathbf{S}, \mathbf{Z}) = 1 \quad (2.63)$$

However, as shown in [30] the LR values of the true covariance matrix \mathbf{R}_0 are much lower than that of the ML solution \mathbf{S} . Therefore, it seems natural to replace the ML estimate by one that generates LR values consistent with what is expected for the true

covariance matrix. More importantly, Abramovich *et al.* showed [30] that the pdf of the LR for the true covariance matrix, which is given by

$$\text{LR}(\mathbf{R}_0, \mathbf{Z}) = \frac{|\mathbf{R}_0^{-1}\mathbf{S}| \exp N}{\exp[\text{tr}(\mathbf{R}_0^{-1}\mathbf{S})]} \quad (2.64)$$

$$= \frac{|\mathbf{R}_0^{-1/2}\mathbf{S}\mathbf{R}_0^{-1/2}| \exp N}{\exp[\text{tr}(\mathbf{R}_0^{-1/2}\mathbf{S}\mathbf{R}_0^{-1/2})]} \quad (2.65)$$

does not depend on the true covariance itself since

$$\hat{\mathbf{C}} \equiv N\mathbf{R}_0^{-1/2}\mathbf{S}\mathbf{R}_0^{-1/2} \sim \mathcal{CW}(K, N, \mathbf{I}) \quad (2.66)$$

where \mathcal{CW} represents complex Wishart distribution which is determined entirely by K and N and does not need \mathbf{R}_0 . Therefore, the pdf of LR values for the true covariance matrix can be precalculated for given K and N and indeed the moments of distribution of the LR values were derived by Abramovich *et al.* in their paper [30].

Based on the invariance of the pdf of LR values, the EL approach can be used to determine values of parameters in estimation problems. For instance, the EL estimator for a diagonally loaded SMI technique under homogeneous interference training conditions and fluctuating target with known power is given by [30]

$$\hat{\mathbf{R}}_{\text{LSMI}} = \hat{\beta}\mathbf{I} + \mathbf{S} \quad (2.67)$$

where

$$\hat{\beta} \equiv \arg_{\beta} \left\{ \frac{|(\beta\mathbf{I} + \mathbf{S})^{-1}\mathbf{S}| \exp N}{\exp(\text{tr}[(\beta\mathbf{I} + \mathbf{S})^{-1}\mathbf{S}])} \equiv \text{LR}_0 \right\} \quad (2.68)$$

and LR_0 is the reference median statistic, which can be precalculated from the pdf of the LR values

$$\int_0^{\text{LR}_0} f[\text{LR}(\mathbf{R}_0, \mathbf{Z})] d\text{LR} = 0.5 \quad (2.69)$$

where $f[\text{LR}(\mathbf{R}_0, \mathbf{Z})]$ is the invariant pdf of the LR values.

Rank Constrained Maximum Likelihood Estimation

3.1 Introduction

Space time adaptive processing (STAP), i.e. joint adaptive processing in the spatial and temporal domains [2, 3, 47] is the cornerstone of radar signal processing and creates the ability to suppress interfering signals while simultaneously preserving gain on the desired signal. For STAP to be successful though, interference statistics must be estimated from training samples. Training, therefore plays a pivotal role in adaptive radar systems.

The radar STAP filter rejects clutter and noise by using an adaptive multidimensional finite impulse response (FIR) filter structure which consists of P time taps and J spatial taps. The optimal STAP complex weight vector which maximizes the return signal from a desired target requires formation and inversion of the disturbance covariance matrix. Further, this covariance matrix is central to test statistics involved in target detection, such as the normalized matched filter (NMF) [23] and the generalized likelihood ratio test (GLRT) [41].

The disturbance covariance matrix must in practice be estimated. In the absence of any prior knowledge about the interference environment, a large number of homogeneous (target free) disturbance training samples are required to obtain accurate estimates. The quality and quantity of training data are governed by the scale with which the disturbance statistics change with respect to space and time as well as by systems considerations such as bandwidth. A compelling challenge for radar STAP emerges since the availability of generous homogeneous training data is often unrealistic [4]. This problem is exacerbated

because the estimation process must be repeated for each Doppler and range bin of interest. Therefore, much recent research in radar STAP has focused on circumventing the lack of generous homogeneous training. One approach to this problem includes the use of *a priori* information about the radar environment and is widely referred to in the literature as knowledge-based processing [5, 6, 44, 48–51]. A subset of these techniques concern themselves with intelligent training selection [5]. Other methods try to reduce the spatio-temporal DOF to reduce both the number of required training samples and computational cost [6–8]. Enforcing structure on covariance matrices (such as Toeplitz) and shrinkage estimation techniques have also been considered [11, 21]. A more thorough review of these techniques is provided in Section 3.2.

Our work introduces a new advance in robust covariance estimation for STAP, known as the rank-constrained maximum likelihood (RCML) estimator of structured covariance matrices. The disturbance covariance matrix exhibits a structure that comprises the sum of noise (white) and clutter covariances with the clutter component being positive semi-definite and rank-deficient. What is of particular interest is that in airborne radar scenarios involving land clutter, this rank can in fact be determined using the Brennan rule [24] under nominal conditions. Incorporation of this rank information in our research is a logical evolution of the pioneering work by Steiner and Gerlach [20] wherein they demonstrated that the fast maximum likelihood estimator (FML) which is cognizant of the eigenstructure of the disturbance covariance can in fact outperform competing approaches in the literature.

Contributions: From an analytic viewpoint, the contributions of this work are in formulating and solving the rank-constrained ML estimation problem for structured covariance matrices. We note that the value of the rank has been identified in statistics [52] and in radar signal processing via eigencancellation approaches [19], [2]. Our work overlaps with these techniques in that we use eigenvectors identical to those of the sample covariance matrix for the estimated covariance matrix whose optimal choice can formally be shown to agree with those obtained from an SVD of the data matrix of training data. We demonstrate additionally that despite the presence of the challenging (non-convex) rank-constraint, our estimation problem can in fact be reduced to a convex optimization problem over the eigenvalues and further, closed form expressions are derived for the estimator. Our central analytical result is the RCML estimator, which like FML and eigencancellation approaches above exploits the knowledge of the radar noise floor. The noise power is typically determined by placing the radar in “receive only” mode prior to going active [53]. For mathematical completeness, we also derive another estimator

called RCML_{LB} for the case when the noise floor is assumed unknown and only a lower bound (LB) is available. We note our RCML_{LB} estimator is in fact a generalization of the result reported by Wax and Kailath [54], who quote Anderson [52] for the result.

Our experimental contributions include an extensive performance analysis of the RCML method and performance comparison with a host of competing methods with demonstrated success. Our experimental data was obtained from the DARPA KASSPER data set where ground truth covariance is available which helps in evaluation via well known figures of merit such as the normalized signal to interference and noise ratio (SINR). The merits of RCML are most pronounced in the low training regime, particularly the case of $K < N$ training samples which is a stiff practical challenge for radar STAP. Additionally, the proposed RCML estimator is robust to perturbations against the true knowledge of the rank making it even more appealing from a practical standpoint.

The rest of the chapter is organized as follows. Section 3.2 reviews existing covariance estimation literature and further motivates our contribution. Section 3.3.1 formulates the RCML estimation problem and subsequent derivation is provided in Section 3.3.2. The solutions of the estimation/optimization problem are derived for the two cases of both known and unknown (known lower bound) noise levels. Experimental validation of our methods is provided in Section 3.4 wherein we report the performance of the proposed estimator and compare it against well-known existing methods in terms of SINR. In addition, rank sensitivity of the proposed RCML estimator is also evaluated. Finally, concluding remarks with directions for future work are presented in Section 3.5.

3.2 Motivation and Review

Because the covariance matrix plays a crucial role in the detection statistic (see Eq. (2.8)), it is very important to estimate it reliably. Widrow et al. and Applebaum proposed least-squares method [55] and maximum signal-to-noise-ratio criterion [56], respectively, using feedback loops. However, these methods were slow to converge to the steady-state solution. Reed, Mallet, and Brennan [57] verified that the sample matrix inverse (SMI) method demonstrated considerably better convergence. In the sample matrix inverse method, the disturbance covariance matrix can be estimated using K data ranges for training

$$\hat{\mathbf{R}}_d = \frac{1}{K} \sum_{k=1}^K \mathbf{x}_k \mathbf{x}_k^H = \frac{1}{K} \mathbf{X} \mathbf{X}^H \quad (3.1)$$

where K is the number of training data we received, $\mathbf{x}_k \in \mathbb{C}^N$, $N = JP$ is the k th vector of training data, and $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_K] \in \mathbb{C}^{N \times K}$. It is well known that the sample covariance is the *unconstrained* maximum likelihood estimator when $K \geq N$. Despite this virtue, there remain fundamental problems with the SMI approach. First, typically $K > N$ training samples are needed to guarantee the non-singularity of the estimated covariance matrix. In fact, when $K < N$ the estimate is singular and therefore precludes implementation of the STAP processor. As much past research has shown [8], the estimate also does quite poorly in the vicinity of $K = N$ training samples.

As observed in Section 3.1, large number of homogeneous training samples are generally not available [4]. To overcome the practical issue of limited training, covariance matrix estimation techniques that enforce and exploit particular structure have been pursued. Examples of structure include persymmetry [9], the Toeplitz property [11, 13, 45], circulant structure [15], multichannel autoregressive models [58, 59] and physical constraints [60]. The FML method [20] which enforces special eigenstructure also falls in this category and in fact is shown to be the most competitive technique experimentally [8, 23]. In particular, the disturbance covariance matrix \mathbf{R} represents the exhibits the following structure

$$\mathbf{R} = \sigma^2 \mathbf{I} + \mathbf{R}_c \quad (3.2)$$

where \mathbf{R}_c denotes the clutter matrix which has a low rank and is positive semi-definite and \mathbf{I} is an identity matrix. Steiner and Gerlach's FML technique ensures that the estimated covariance matrix has eigenvalues all greater than σ^2 . Recently, the work by Aubry *et al.* [22] has also improved upon FML by the introduction of a condition number constraint. Other approaches include Bayesian covariance matrix estimators [5, 61–64] and the use of knowledge-based covariance models [44, 51, 65, 66]. Finally, shrinkage estimation methods have been also considered [21, 67–69].

3.3 Rank Constrained ML Estimation of Structured Covariance Matrices

3.3.1 Overview of contribution

The principal contribution of our work is to incorporate the rank of the clutter covariance matrix, \mathbf{R}_c , explicitly into ML estimation of the disturbance covariance matrix¹. Under ideal conditions (no mutual coupling between array elements and no internal clutter

¹Preliminary version of the work has appeared at the 2012 IEEE Radar conference [32].

motion), Brennan rule [24] states that the rank of \mathbf{R}_c in the airborne linear phased array radar problem is given by

$$\text{rank}(\mathbf{R}_c) = J + \gamma(P - 1) \quad (3.3)$$

where $\gamma = 2v_p T/d$ is the slope of the clutter ridge, with v_p denoting the platform velocity, T denoting the pulse repetition interval, and d denoting the inter-element spacing. Even if there is mutual coupling in practice, \mathbf{R}_c has rank r which is much less than the spatio-temporal product $N = JP$ in many practical airborne radar applications. In addition, powerful techniques have been developed [23] to determine the rank fairly accurately.

We first set up the optimization problem to estimate the disturbance covariance matrix with a structural constraint on \mathbf{R} and the rank constraint on \mathbf{R}_c . The estimation problem when seen as an optimization over \mathbf{R} is unfortunately not a convex problem, since neither the cost function nor the constraints (rank) are convex (elaborated upon in Section 3.3.2). We will however show that using a transformation of variables, reduction to a convex form is possible and further by invoking KKT conditions [31] for the resulting convex problems, it is in fact possible to derive a closed form solution. Akin to FML, we initially assume (Section 3.3.2.1) that the noise power σ^2 is known while setting up and solving the problem. Then we extend our results (Section 3.3.2.2) to the case of unknown noise variance. In that case, we assume that only a lower bound on the noise power is available. That is, we know \hat{c} where the covariance matrix is expressed by

$$\mathbf{R} = c\mathbf{I} + \mathbf{R}_c \quad (3.4)$$

where c is the noise variance (to be estimated/optimized) such that $c > \hat{c}$.

3.3.2 ML Estimation

Let $\mathbf{z}_i \in \mathbb{C}^N$ be the i th realization of the target-free (stochastic) disturbance vector and K be the number of training samples. That is, $i = 1, 2, \dots, K$ and $N = JP$. Therefore, under each training sample, \mathbf{z}_i , under assumption of zero mean, obeys

$$f(\mathbf{z}_i) = \frac{1}{\pi^N |\mathbf{R}|} \exp(-\mathbf{z}_i^H \mathbf{R}^{-1} \mathbf{z}_i) \quad (3.5)$$

which comes from a zero-mean complex circular Gaussian distribution and \mathbf{R} is the $N \times N$ disturbance covariance matrix. Further, $|\mathbf{R}|$ denotes the determinant of \mathbf{R} , \mathbf{z}_i^H is the Hermitian (conjugate transpose) of \mathbf{z}_i . Let \mathbf{Z} be the $N \times K$ complex matrix whose i -th column is the observed vector \mathbf{z}_i . Since each observations \mathbf{z}_i are i.i.d, the likelihood

of observing \mathbf{Z} given \mathbf{R} is given by

$$f(\mathbf{R}; \mathbf{Z}) = \frac{1}{\pi^{NK}} |\mathbf{R}|^{-K} \exp(-\text{tr}\{\mathbf{Z}^H \mathbf{R}^{-1} \mathbf{Z}\}) \quad (3.6)$$

$$= \frac{1}{\pi^{NK}} |\mathbf{R}|^{-K} \exp(-\text{tr}\{\mathbf{R}^{-1} \mathbf{Z} \mathbf{Z}^H\}) \quad (3.7)$$

$$= \frac{1}{\pi^{NK}} |\mathbf{R}|^{-K} \exp(-K \cdot \text{tr}\{\mathbf{R}^{-1} \mathbf{S}\}) \quad (3.8)$$

where $\mathbf{S} = \frac{1}{K} \mathbf{Z} \mathbf{Z}^H$ is the well-known sample covariance matrix. Our goal is to find the positive definite matrix \mathbf{R} that maximizes the likelihood function $f(\mathbf{R})(\mathbf{Z})$. The logarithm of the likelihood term is

$$\log f(\mathbf{R}; \mathbf{Z}) = -K \cdot \text{tr}\{\mathbf{R}^{-1} \mathbf{S}\} - K \log(|\mathbf{R}|) - NK \log(\pi). \quad (3.9)$$

Maximizing the log-likelihood as a function of \mathbf{R} is equivalent to minimizing the function given by

$$\text{tr}\{\mathbf{R}^{-1} \mathbf{S}\} + \log(|\mathbf{R}|). \quad (3.10)$$

Therefore, the optimization problem for obtaining the maximum likelihood solution under the rank constraint is given by

$$\begin{cases} \min_{\mathbf{R}} & \text{tr}\{\mathbf{R}^{-1} \mathbf{S}\} + \log(|\mathbf{R}|) \\ s.t. & \mathbf{R} = \sigma^2 \mathbf{I} + \mathbf{R}_c \\ & \text{rank}(\mathbf{R}_c) = r \\ & \mathbf{R}_c \succeq \mathbf{0} \end{cases} \quad (3.11)$$

Since the cost function is not a convex function in \mathbf{R} , we apply a transformation variables i.e., let $\mathbf{X} = \sigma^2 \mathbf{R}^{-1}$ and $\bar{\mathbf{S}} = \frac{1}{\sigma^2} \mathbf{S}$. Then, the revised cost function in the optimization variable \mathbf{X} becomes

$$\begin{aligned} & \text{tr}\{\mathbf{R}^{-1} \mathbf{S}\} + \log(|\mathbf{R}|) \\ &= \text{tr}\{\bar{\mathbf{S}} \mathbf{X}\} - \log\left(\left|\frac{1}{\sigma^2} \mathbf{X}\right|\right) \end{aligned} \quad (3.12)$$

$$= \text{tr}\{\bar{\mathbf{S}} \mathbf{X}\} - \log(|\mathbf{X}|) + \log \sigma^{2N}. \quad (3.13)$$

Since $\log \sigma^{2N}$ in Eq. (3.13) is a constant, the final cost function to be minimized is

$$\text{tr}\{\bar{\mathbf{S}} \mathbf{X}\} - \log(|\mathbf{X}|). \quad (3.14)$$

Note that $\text{tr}\{\bar{\mathbf{S}}\mathbf{X}\} = \sum_{i=1}^N \sum_{j=1}^N \bar{s}_{ji}x_{ij}$ is affine and $\log(|\mathbf{X}|)$ is concave, which implies $-\log(|\mathbf{X}|)$ is convex. Therefore, the final cost function, Eq. (3.14) is convex in the variable \mathbf{X} .

We now express \mathbf{X} and $\bar{\mathbf{S}}$ in terms of their eigenvalue decomposition, i.e., $\mathbf{X} = \Phi \Lambda \Phi^H$ and let $\bar{\mathbf{S}}$ be decomposed as $\bar{\mathbf{S}} = \mathbf{V} \mathbf{D} \mathbf{V}^H$ where Φ and \mathbf{V} are orthonormal eigenvector matrices of \mathbf{X} and $\bar{\mathbf{S}}$, Λ and \mathbf{D} are diagonal matrices with diagonal entries which are eigenvalues of \mathbf{X} arranged in ascending order and $\bar{\mathbf{S}}$ in descending order, respectively. Using the eigendecompositions, the cost function can be simplified as

$$\begin{aligned} \text{tr}\{\bar{\mathbf{S}}\mathbf{X}\} - \log(|\mathbf{X}|) &= \text{tr}\{\mathbf{V} \mathbf{D} \mathbf{V}^H \Phi \Lambda \Phi^H\} - \log(|\Phi \Lambda \Phi^H|) \end{aligned} \quad (3.15)$$

$$= \text{tr}\{\mathbf{D} \mathbf{V}^H \Phi \Lambda \Phi^H \mathbf{V}\} - \log(|\Lambda|) \quad (3.16)$$

Therefore, the optimization problem using the eigenvalues and the eigenvectors is given by

$$\left\{ \begin{array}{ll} \min_{\Lambda, \Phi} & \text{tr}\{\mathbf{D} \mathbf{V}^H \Phi \Lambda \Phi^H \mathbf{V}\} - \log(|\Lambda|) \\ s.t. & \Lambda \mathbf{1} \preceq \sigma^2 \mathbf{1} \\ & \mathbf{U} \Lambda \mathbf{1} \preceq \mathbf{0} \\ & \Lambda \mathbf{e} = \mathbf{e} \\ & \Phi^H \Phi = \mathbf{I} \end{array} \right. \quad (3.17)$$

where $\mathbf{1}$ and $\mathbf{0}$ represent $N \times 1$ vectors with constant elements of 1 and 0, respectively, $\mathbf{e} = [0, \dots, 0_r, 1, \dots, 1]^T$, and a matrix \mathbf{U} is given in Eq. (3.24). Since the first three constraints are only for the eigenvalues of \mathbf{X} , that is, Λ , we can obtain the optimal eigenvector matrix Φ by solving the following optimization problem

$$\left\{ \begin{array}{ll} \min_{\Phi} & \text{tr}\{\mathbf{D} \mathbf{V}^H \Phi \Lambda \Phi^H \mathbf{V}\} - \log(|\Lambda|) \\ s.t. & \Phi^H \Phi = \mathbf{I} \end{array} \right. \quad (3.18)$$

and the solution of the optimization problem (3.18) is in fact fairly well known from standard unconstrained ML estimation. That is, over the space of unitary matrices the optimal Φ that maximizes the likelihood is the one that matches with the eigenvector matrix of the sample covariance \mathbf{V} [42]. The cost function can be more simplified by

substituting Φ with \mathbf{V} ,

$$\text{tr}\{\mathbf{D}\mathbf{V}^H\Phi\Lambda\Phi^H\mathbf{V}\} - \log(|\Lambda|) = \text{tr}\{\mathbf{D}\Lambda\} - \log(|\Lambda|) \quad (3.19)$$

$$= \mathbf{d}^T \boldsymbol{\lambda} - \mathbf{1}^T \log \boldsymbol{\lambda} \quad (3.20)$$

where \mathbf{d} and $\boldsymbol{\lambda}$ are vectors with entries of eigenvalues of $\bar{\mathbf{S}}$ and \mathbf{X} respectively and $\log \boldsymbol{\lambda} = [\log \lambda_1, \log \lambda_2, \dots, \log \lambda_N]^T$. Hence, the optimization may be focused on the vector of eigenvalues $\boldsymbol{\lambda}$. Many previous algorithms such as FML [20], eigncanceler [19], and other eigen-based techniques [2] have utilized the same eigenvectors without eigenvalue optimization.

3.3.2.1 Known Noise Level Case

We first assume the noise power is known. Then the constraints of the optimization problem are

$$\begin{cases} \mathbf{R} = \sigma^2 \mathbf{I} + \mathbf{R}_c \\ \text{rank}(\mathbf{R}_c) = r \\ \mathbf{R}_c \succeq \mathbf{0} \\ \mathbf{R} \succeq \sigma^2 \mathbf{I} \end{cases} \quad (3.21)$$

Since $\text{rank}(\mathbf{R}_c) = r$, \mathbf{R}_c has r non-negative eigenvalues and the rest eigenvalues are all zero. Hence, from Eq. (3.2), \mathbf{R} has r eigenvalues which are greater than or equal to σ^2 and the rest eigenvalues equal to σ^2 . That is, the eigenvalue matrix of \mathbf{R} has structure of the form

$$\begin{pmatrix} \bar{\lambda}_1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \bar{\lambda}_2 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \ddots & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \bar{\lambda}_r & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & \sigma^2 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & \sigma^2 \end{pmatrix} \quad (3.22)$$

where $\bar{\lambda}_1 \geq \bar{\lambda}_2 \geq \dots \geq \bar{\lambda}_r \geq \sigma^2$. Hence, $\mathbf{\Lambda}$, the eigenvalue matrix of $\mathbf{X} = \sigma^2 \mathbf{R}^{-1}$, has the following structure

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \ddots & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \lambda_r & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 1 \end{pmatrix} \quad (3.23)$$

where $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_r \leq 1$. Now the constraints can be expressed in vector and matrix forms. The first constraint is $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$, that is, $\mathbf{U}\mathbf{\lambda} \preceq \mathbf{0}$ where

$$\mathbf{U} = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & \dots & \vdots \\ 0 & 0 & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 1 & -1 \\ 0 & 0 & \dots & 0 & -1 \end{pmatrix} \in \mathbb{R}^{N \times N} \quad (3.24)$$

The second constraint is $0 < \lambda_i \leq 1$ which can be expressed by

$$\boldsymbol{\varepsilon} \preceq \mathbf{\lambda} \preceq \mathbf{1} \quad (3.25)$$

where $\boldsymbol{\varepsilon}$ is a vector with all entries equal to the same constant ϵ such that ϵ is picked close to zero. Note, that this is done to avoid solutions where any of the λ_i exactly equals zero since that would lead to a singular \mathbf{X} . The final constraint is $\lambda_{r+1} = \lambda_{r+2} = \dots = \lambda_N = 1$, and is expressed as

$$\mathbf{E}\mathbf{\lambda} = \mathbf{h} \quad (3.26)$$

where

$$\mathbf{E} = \begin{pmatrix} \mathbf{0}_{r \times r} & \mathbf{0}_{r \times (N-r)} \\ \mathbf{0}_{(N-r) \times r} & \mathbf{I}_{N-r} \end{pmatrix} \in \mathbb{R}^{N \times N} \quad (3.27)$$

and $\mathbf{h} = [0, 0, \dots, 0_r, 1, 1, \dots, 1]^T$.

We therefore have the following optimization problem.

$$\left\{ \begin{array}{ll} \min_{\boldsymbol{\lambda}} & \mathbf{d}^T \boldsymbol{\lambda} - \mathbf{1}^T \log \boldsymbol{\lambda} \\ s.t. & \mathbf{U} \boldsymbol{\lambda} \preceq \mathbf{0} \\ & -\boldsymbol{\lambda} \preceq -\boldsymbol{\varepsilon} \\ & \boldsymbol{\lambda} \preceq \mathbf{1} \\ & \mathbf{E} \boldsymbol{\lambda} = \mathbf{h} \end{array} \right. \quad (3.28)$$

Remark: Note that the constraint $\boldsymbol{\lambda} \preceq \mathbf{1}$ in fact forces the maximum rank of \mathbf{R}_c to be r . That is, we use a relaxation of the strict inequality which forces an exact rank constraint. As will be shown shortly, this relaxation allows us to obtain a closed form solution while the optimal solutions to the two problems (original vs. relaxed) are hardly distinguishable [70].

Combining the inequality constraints into one, we have

$$\left\{ \begin{array}{ll} \min_{\boldsymbol{\lambda}} & \mathbf{d}^T \boldsymbol{\lambda} - \mathbf{1}^T \log \boldsymbol{\lambda} \\ s.t. & \mathbf{F} \boldsymbol{\lambda} \preceq \mathbf{g} \\ & \mathbf{E} \boldsymbol{\lambda} = \mathbf{h} \end{array} \right. \quad (3.29)$$

where $\mathbf{F} = \begin{bmatrix} \mathbf{U} \\ -\mathbf{I} \\ \mathbf{I} \end{bmatrix}$, $\mathbf{g} = \begin{bmatrix} \mathbf{0} \\ -\boldsymbol{\varepsilon} \\ \mathbf{1} \end{bmatrix}$, $\mathbf{E} = \begin{bmatrix} \mathbf{0}_{r \times r} & \mathbf{0}_{r \times (N-r)} \\ \mathbf{0}_{(N-r) \times r} & \mathbf{I}_{N-r} \end{bmatrix}$, and $\mathbf{h} = [0, 0, \dots, 0_r, 1, 1, \dots, 1]^T$.

Here, $\boldsymbol{\lambda}, \mathbf{d}, \mathbf{h} \in \mathbb{R}^N$, $\mathbf{g} \in \mathbb{R}^{3N}$, $\mathbf{U}, \mathbf{E} \in \mathbb{R}^{N \times N}$, and $\mathbf{F} \in \mathbb{R}^{3N \times N}$. The optimization problem (3.29) is obviously a convex optimization problem because the cost function is a convex function and feasible constraint sets are convex as well.

Lemma 1. *A closed form solution of (3.29) can be obtained by KKT conditions and is given by*

$$\lambda_i^* = \begin{cases} \min(1, \frac{1}{d_i}) & \text{for } i = 1, 2, \dots, r \\ 1 & \text{for } i = r+1, r+2, \dots, N \end{cases} \quad (3.30)$$

Proof. Before solving the problem using the KKT conditions, we can simplify the problem further. Note that the constraint (3.26) tells us we already have $\lambda_{r+1} = \lambda_{r+2} = \dots = \lambda_N = 1$. This means we do not need to optimize $\lambda_{r+1}, \lambda_{r+2}, \dots, \lambda_N$. Therefore,

we can arrive at an r variable-optimization problem given by

$$\begin{cases} \min_{\boldsymbol{\lambda}} & \mathbf{d}^T \boldsymbol{\lambda} - \mathbf{1}^T \log \boldsymbol{\lambda} \\ \text{s.t.} & \mathbf{F} \boldsymbol{\lambda} \preceq \mathbf{g} \end{cases} \quad (3.31)$$

where $\mathbf{F} = \begin{bmatrix} \mathbf{U}^T & -\mathbf{I} & \mathbf{I} \end{bmatrix}^T$, $\mathbf{g} = \begin{bmatrix} \mathbf{0}^T & -\boldsymbol{\varepsilon}^T & \mathbf{1}^T \end{bmatrix}^T$. Here, $\boldsymbol{\lambda}, \mathbf{d} \in \mathbb{R}^r$, $\mathbf{g} \in \mathbb{R}^{3r}$, $\mathbf{U} \in \mathbb{R}^{r \times r}$, and $\mathbf{F} \in \mathbb{R}^{3r \times r}$. We can rewrite the Lagrangian and KKT conditions to solve the problem.

The *Lagrangian* $L : \mathbb{R}^r \times \mathbb{R}^{3r} \times \mathbb{R}^r \mapsto \mathbb{R}$ associated with the problem is:

$$L(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{d}^T \boldsymbol{\lambda} - \mathbf{1}^T \log \boldsymbol{\lambda} + \boldsymbol{\mu}^T (\mathbf{F} \boldsymbol{\lambda} - \mathbf{g}) \quad (3.32)$$

where $\boldsymbol{\mu} \in \mathbb{R}^{3r}$. The KKT conditions for any $\boldsymbol{\lambda}$ to be a minimizer are:

Primal inequality constraints:

$$\mathbf{F} \boldsymbol{\lambda} \preceq \mathbf{g} \quad (3.33)$$

Dual inequality constraints:

$$\boldsymbol{\mu} \succeq \mathbf{0}, \boldsymbol{\nu} \succeq \mathbf{0}, \mathbf{v} \succeq \mathbf{0} \quad (3.34)$$

Complementary slackness:

$$\begin{aligned} \mu_1(\lambda_1 - \lambda_2) &= 0 \\ &\vdots \\ \mu_{r-1}(\lambda_{r-1} - \lambda_r) &= 0 \\ \mu_r \lambda_r &= 0 \\ \nu_1(\lambda_1 - \varepsilon) &= 0 \\ &\vdots \\ \nu_r(\lambda_r - \varepsilon) &= 0 \\ v_1(\lambda_1 - 1) &= 0 \\ &\vdots \\ v_r(\lambda_r - 1) &= 0 \end{aligned}$$

Stationarity:

$$\nabla_{\lambda} L(\lambda, \mu) = \mathbf{d} - \frac{1}{\lambda} + \mathbf{F}^T \mu = \mathbf{0} \quad (3.35)$$

where $\frac{1}{\lambda} = [\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \dots, \frac{1}{\lambda_r}]^T$. To make the notation clear, we let $\mu = [\mu_1, \mu_2, \dots, \mu_r]^T$, $\nu = [\mu_{r+1}, \mu_{r+2}, \dots, \mu_{2r}]^T$, and $\mathbf{v} = [\mu_{2r+1}, \mu_{2r+2}, \dots, \mu_{3r}]^T$. Then the primal inequality constraints mean

$$\begin{cases} \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_r \\ \varepsilon \leq \lambda_i \leq 1 \end{cases} \quad \text{for } i = 1, 2, \dots, r \quad (3.36)$$

First, from complementary slackness, we get $\nu = \mathbf{0}$ since we let ε be a very small number which is enough close to zero such that any λ_i cannot equal to ε . In addition, the slackness means

$$\begin{cases} \mu_i = 0 & \text{if } \lambda_i \neq \lambda_{i+1} \\ v_i = 0 & \text{if } \lambda_i \neq 1 \end{cases} \quad (3.37)$$

Next we analyze the stationarity condition

$$\begin{aligned} d_1 - \frac{1}{\lambda_1} + \mu_1 - v_1 &= 0 \\ d_2 - \frac{1}{\lambda_2} - \mu_1 + \mu_2 - v_2 &= 0 \\ &\vdots \\ d_r - \frac{1}{\lambda_r} - \mu_{r-1} - \mu_r - v_r &= 0. \end{aligned}$$

We first assume that $\lambda_1 \neq 1$ which leads to $v_1 = 0$. We now have two cases which are $\lambda_1 \neq \lambda_2$ and $\lambda_1 = \lambda_2$. In the former case, we also know $\mu_1 = 0$. Therefore,

$$d_1 - \frac{1}{\lambda_1} = 0. \quad (3.38)$$

This leads to $\lambda_1 = \frac{1}{d_1}$ and by the primal inequality constraint, $\lambda_1 \leq 1$, finally $\lambda_1 = \min(\frac{1}{d_1}, 1)$. In the latter case, we can also consider two cases here which are $\lambda_1 = \lambda_2 \neq \lambda_3$ and $\lambda_1 = \lambda_2 = \lambda_3$. In the first case, $\mu_2 = v_1 = v_2 = 0$. Therefore, the first two equations

in the stationarity condition become

$$\begin{cases} d_1 - \frac{1}{\lambda_1} + \mu_1 = 0 \\ d_2 - \frac{1}{\lambda_1} - \mu_1 = 0 \end{cases}. \quad (3.39)$$

From these two equations, we get $\mu_1 = (d_2 - d_1)/2$. Because $d_1 \geq d_2$, this implies $\mu_1 \leq 0$. However, $\mu_1 \geq 0$ from the dual inequality constraints. Therefore, $\mu_1 = 0$ and $\lambda_1 = \min(\frac{1}{d_1}, 1)$. Next, the case of $\lambda_1 = \lambda_2 = \lambda_3$ can be split into two which are $\lambda_1 = \lambda_2 = \lambda_3 \neq \lambda_4$ and $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4$. By following an approach similar to solving simultaneous equations, we conclude that $\boldsymbol{\mu} = \mathbf{0}$ and therefore $\lambda_i = \min(\frac{1}{d_i}, 1)$. \square

From Lemma 1, the optimal solution \mathbf{X}^* is

$$\mathbf{X}^* = \mathbf{V}\boldsymbol{\Lambda}^*\mathbf{V}^H \quad (3.40)$$

and the optimal covariance matrix estimator \mathbf{R}^* is

$$\mathbf{R}^* = \sigma^2 \mathbf{X}^{*-1} = \sigma^2 \mathbf{V}\boldsymbol{\Lambda}^{*-1}\mathbf{V}^H \quad (3.41)$$

where \mathbf{V} is the eigenvector matrix of the sample covariance matrix \mathbf{S} and $\boldsymbol{\Lambda}^*$ is a diagonal matrix with diagonal entries λ_i^* . It should be noted that this is a generalization of the FML solution in [20] with the rank-information additionally incorporated.

3.3.2.2 Unknown Noise Level Case

For mathematical completeness, we also derive the case when σ^2 is also estimated as a part of the estimation process. We however assume that a lower bound (LB) is known and call this estimator RCML_{LB}.

We proceed by defining $\mathbf{X} = \mathbf{R}^{-1}$ (instead of $\mathbf{X} = \sigma^2 \mathbf{R}^{-1}$ as in Section 3.3.2.1) because we assume that the exact noise power is unknown. Using simplifications similar to the ones in Section 3.3.2.1, we have the cost function to be minimized as

$$\text{tr}\{\mathbf{R}^{-1}\mathbf{S}\} + \log(|\mathbf{R}|) = \text{tr}\{\mathbf{S}\mathbf{X}\} - \log(|\mathbf{X}|). \quad (3.42)$$

The righthand side of Eq. (3.42) is also a convex function in \mathbf{X} via the same reasoning as in Section 3.3.2.1. Note that we do not use \mathbf{S}' any more in this case. Now the constraints

are written as

$$\left\{ \begin{array}{l} \mathbf{R} = c\mathbf{I} + \mathbf{R}_c \\ \text{rank}(\mathbf{R}_c) \leq r \\ \mathbf{R}_c \succeq \mathbf{0} \\ \mathbf{R} \succeq c\mathbf{I} \\ c \geq \hat{c} \end{array} \right. . \quad (3.43)$$

The eigenvalue matrix of \mathbf{R} has structure similar to that in Eq. (3.22)

$$\begin{pmatrix} \bar{\lambda}_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \bar{\lambda}_2 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \bar{\lambda}_r & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & c & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & c \end{pmatrix} \quad (3.44)$$

where $\bar{\lambda}_1 \geq \bar{\lambda}_2 \geq \cdots \geq \bar{\lambda}_r \geq c$, $\mathbf{\Lambda}$, and hence the eigenvalue matrix of $\mathbf{X} = \mathbf{R}^{-1}$, has the following structure

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \lambda_r & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \frac{1}{c} & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & \frac{1}{c} \end{pmatrix} \quad (3.45)$$

Using reductions similar to the ones in Section 3.3.2.1, we can formulate the following optimization problem

$$\left\{ \begin{array}{l} \min_{\boldsymbol{\lambda}, c} \quad \mathbf{d}^T \boldsymbol{\lambda} - \mathbf{1}^T \log \boldsymbol{\lambda} \\ s.t. \quad \mathbf{U} \boldsymbol{\lambda} \preceq \mathbf{0} \\ \quad \quad -\boldsymbol{\lambda} \preceq -\boldsymbol{\epsilon} \\ \quad \quad \boldsymbol{\lambda} \preceq \frac{1}{c} \mathbf{1} \\ \quad \quad \mathbf{E} \boldsymbol{\lambda} = \mathbf{h} \\ \quad \quad c \geq \hat{c} \end{array} \right. \quad (3.46)$$

Note, \mathbf{U} and \mathbf{E} are same as in Section 3.3.2.1, but $\mathbf{h} = [0, 0, \dots, 0_r, \frac{1}{c}, \frac{1}{c}, \dots, \frac{1}{c}]^T$.

Again, inequality constraints may be combined into one

$$\left\{ \begin{array}{ll} \min_{\boldsymbol{\lambda}, c} & \mathbf{d}^T \boldsymbol{\lambda} - \mathbf{1}^T \log \boldsymbol{\lambda} \\ s.t. & \mathbf{F} \boldsymbol{\lambda} \preceq \mathbf{g} \\ & \mathbf{E} \boldsymbol{\lambda} = \mathbf{h} \\ & c \geq \hat{c} \end{array} \right. \quad (3.47)$$

where $\mathbf{F} = \begin{bmatrix} \mathbf{U} \\ -\mathbf{I} \\ \mathbf{I} \end{bmatrix}$, $\mathbf{g} = \begin{bmatrix} \mathbf{0} \\ -\varepsilon \\ \frac{1}{c} \mathbf{1} \end{bmatrix}$, $\mathbf{E} = \begin{bmatrix} \mathbf{0}_{r \times r} & \mathbf{0}_{r \times (N-r)} \\ \mathbf{0}_{(N-r) \times r} & \mathbf{I}_{N-r} \end{bmatrix}$, and $\mathbf{h} = [0, 0, \dots, 0_r, \frac{1}{c}, \frac{1}{c}, \dots, \frac{1}{c}]^T$.

Here, $\boldsymbol{\lambda}, \mathbf{a}, \mathbf{h} \in \mathbb{R}^N$, $\mathbf{g} \in \mathbb{R}^{3N}$, $\mathbf{U}, \mathbf{E} \in \mathbb{R}^{N \times N}$, and $\mathbf{F} \in \mathbb{R}^{3N \times N}$.

To solve (3.47), we first fix the noise level c to $\bar{c} \geq \hat{c}$ and solve the optimization problem with respect to $\boldsymbol{\lambda}$ to obtain the optimal solution $\boldsymbol{\lambda}^*(\bar{c})$ which is a function of \bar{c} . By substituting the optimal $\boldsymbol{\lambda}^*(\bar{c})$ back into the cost function, we can get an auxiliary optimization over the scalar variable c .

Once we fix $c = \bar{c}$, the optimization problem (3.47) can be reduced to a problem in just $\boldsymbol{\lambda}$ which is nearly the same problem as (3.29)

$$\left\{ \begin{array}{ll} \min_{\boldsymbol{\lambda}} & \mathbf{d}^T \boldsymbol{\lambda} - \mathbf{1}^T \log \boldsymbol{\lambda} \\ s.t. & \mathbf{F} \boldsymbol{\lambda} \preceq \mathbf{g} \\ & \mathbf{E} \boldsymbol{\lambda} = \mathbf{h} \end{array} \right. \quad (3.48)$$

where $\mathbf{F} = \begin{bmatrix} \mathbf{U} \\ -\mathbf{I} \\ \mathbf{I} \end{bmatrix}$, $\mathbf{g} = \begin{bmatrix} \mathbf{0} \\ -\varepsilon \\ \frac{1}{\bar{c}} \mathbf{1} \end{bmatrix}$, $\mathbf{E} = \begin{bmatrix} \mathbf{0}_{r \times r} & \mathbf{0}_{r \times (N-r)} \\ \mathbf{0}_{(N-r) \times r} & \mathbf{I}_{N-r} \end{bmatrix}$, and $\mathbf{h} = [0, 0, \dots, 0_r, \frac{1}{\bar{c}}, \frac{1}{\bar{c}}, \dots, \frac{1}{\bar{c}}]^T$.

Here, $\boldsymbol{\lambda}, \mathbf{a}, \mathbf{h} \in \mathbb{R}^N$, $\mathbf{g} \in \mathbb{R}^{3N}$, $\mathbf{U}, \mathbf{E} \in \mathbb{R}^{N \times N}$, and $\mathbf{F} \in \mathbb{R}^{3N \times N}$. Note that (3.29) and (3.48) are the same except that (3.48) uses $\frac{1}{\bar{c}} \mathbf{1}$ and $\frac{1}{\bar{c}}$ instead of $\mathbf{1}$ and 1 in \mathbf{g} and \mathbf{h} . Using a derivation similar to the one in Lemma 1, the optimal solution $\boldsymbol{\lambda}^*$ of (3.48) is

$$\lambda_i^* = \begin{cases} \min(\frac{1}{\bar{c}}, \frac{1}{d_i}) & \text{for } i = 1, 2, \dots, r \\ \frac{1}{\bar{c}} & \text{for } i = r+1, r+2, \dots, N \end{cases} \quad (3.49)$$

If we substitute the optimal $\boldsymbol{\lambda}^*$ as in Eq. (3.49) back into (3.47), we get the following

problem to solve in the unknown noise level variable c

$$\begin{cases} \min_c & \sum_{i=1}^N G_i(c) \\ \text{s.t.} & c \geq \hat{c} \end{cases} \quad (3.50)$$

where

$$G_i(c) = \begin{cases} 1 + \log d_i & \text{if } c \leq d_i \\ \frac{d_i}{c} + \log c & \text{if } c > d_i \end{cases} \quad (3.51)$$

for $i = 1, 2, \dots, r$ and

$$G_i(c) = \frac{d_i}{c} + \log c \quad (3.52)$$

for $i = r+1, r+2, \dots, N$. This problem too admits a closed form which is derived in Lemma 2.

Lemma 2. *The optimal solution of c in (3.50) is given by*

$$c^* = \begin{cases} \hat{c} & \text{if } \hat{c} > \frac{\sum_{i=r+1}^N d_i}{N-r} \\ \frac{\sum_{i=r+1}^N d_i}{N-r} & \text{if } \hat{c} \leq \frac{\sum_{i=r+1}^N d_i}{N-r} \end{cases} \quad (3.53)$$

Proof. Let c^* be an optimal solution to the following optimization problem

$$\begin{cases} \min_c & \sum_{i=1}^N G_i(c) \\ \text{s.t.} & c \geq \sigma^2 \end{cases} \quad (3.54)$$

where, $G_i(c) = d_i \lambda_i^* - \log \lambda_i$, namely

$$G_i(c) = \begin{cases} 1 + \log d_i & \text{if } c \leq d_i \\ \frac{d_i}{c} + \log c & \text{if } c > d_i \end{cases} \quad (3.55)$$

for $i = 1, 2, \dots, r$ and

$$G_i(c) = \frac{d_i}{c} + \log c \quad (3.56)$$

for $i = r+1, r+2, \dots, N$.

First, note that $G_i(c)$ is a constant for $c < d_i$ and monotonically increasing in c

for $c \geq d_i$ when $i = 1, 2, \dots, r$ because the first derivative of the function $\frac{d_i}{c} + \log c$, $-\frac{d_i}{c^2} + \frac{1}{c} > 0$ for $c > d_i$. Then, for $i = 1, \dots, r$,

$$G_i(c) = \begin{cases} 1 + \log d_i & \text{if } c \leq d_i & : \text{ constant} \\ \frac{d_i}{c} + \log c & \text{if } c > d_i & : \text{ increasing} \end{cases} \quad (3.57)$$

Because we assumed $d_1 \geq d_2 \geq \dots \geq d_r$, we have

$$\begin{aligned} & \sum_{i=1}^r G_i(c) \\ &= \begin{cases} \sum_{i=1}^r (1 + \log d_i) & \text{if } c \leq d_r \\ \sum_{i=1}^{r-1} (1 + \log d_i) + \frac{d_r}{c} + \log c & \text{if } d_r \leq c \leq d_{r-1} \\ \vdots & \vdots \\ \sum_{i=1}^r (\frac{d_i}{c} + \log c) & \text{if } c > d_1 \end{cases}. \end{aligned} \quad (3.58)$$

In Eq. (3.58), $\sum_{i=1}^r G_i(c)$ is a constant when $c \leq d_r$ and an increasing function otherwise. Finally, since $\sum_{i=1}^r G_i(c)$ is continuous at all d_i , we can conclude that

$$\sum_{i=1}^r G_i(c) = \begin{cases} \text{constant} & \text{if } c \leq d_r \\ \text{increasing} & \text{if } c > d_r \end{cases} \quad (3.59)$$

Now, $G_i(c) = \frac{d_i}{c} + \log c$ for $i = r+1, r+2, \dots, N$. Hence,

$$\sum_{i=r+1}^N G_i(c) = \sum_{i=r+1}^N (\frac{d_i}{c} + \log c) \quad (3.60)$$

$$= \frac{1}{c} \sum_{i=r+1}^N d_i + (N-r) \log c \quad (3.61)$$

We can easily see that

$$\sum_{i=r+1}^N G_i(c) = \begin{cases} \text{decreasing} & \text{if } c \leq \hat{d} \\ \text{increasing} & \text{if } c > \hat{d} \end{cases} \quad (3.62)$$

where $\hat{d} = \frac{\sum_{i=r+1}^N d_i}{N-r}$. In addition, it is obvious $\hat{d} < d_r$ because $d_r > d_i$ for all $i = r+1, r+2, \dots, N$ and \hat{d} which is the mean value of $d_{r+1}, d_{r+2}, \dots, d_N$. From Eq. (3.59) and Eq. (3.62),

$$\begin{aligned} G(c) &= \sum_{i=1}^N G_i(c) = \sum_{i=1}^r G_i(c) + \sum_{i=r+1}^N G_i(c) \\ &= \begin{cases} \text{decreasing} & \text{if } c \leq \hat{d} \\ \text{increasing} & \text{if } c > \hat{d} \end{cases} \end{aligned} \quad (3.63)$$

Consequently, the optimal solution c^* which minimizes the cost function $G(c)$ is determined as

$$c^* = \begin{cases} \hat{c} & \text{if } \hat{c} > \hat{d} \\ \hat{d} & \text{if } \hat{c} \leq \hat{d} \end{cases} \quad (3.64)$$

□

Therefore, the optimal solution λ^* is

$$\lambda_i^* = \begin{cases} \min(\frac{1}{c^*}, \frac{1}{d_i}) & \text{for } i = 1, 2, \dots, r \\ \frac{1}{c^*} & \text{for } i = r+1, r+2, \dots, N \end{cases}. \quad (3.65)$$

Combining the optimal estimated eigenvalue set in (3.65) (expressed in terms of the optimal noise level variable c^*) with the eigenvectors of the sample covariance leads to the optimal estimate of the (inverse of) covariance estimate of \mathbf{R} under the rank constraint. This result is generalization of the result in Wax and Kailath [54] by including and exploiting a lower bound on the noise floor when available. We discuss these two algorithms in Section 3.4.3 in more detail.

3.4 Experimental Results

3.4.1 Experimental Setup and Methods Compared

Data from the L-band data set of the Knowledge Aided Sensor Signal Processing and Expert Reasoning (KASSPER) program [71] is used for the performance analysis discussed in this section. The KASSPER data is the result of a significant effort by DARPA to provide a publicly available resource for the evaluation and benchmarking of radar STAP algorithms. As elaborated in [5], the KASSPER data set was carefully captured to

Table 3.1. KASSPER Dataset-1 parameters

Parameter	Value
Carrier Frequency	1240 MHz
Bandwidth (BW)	10 MHz
Number of Antenna Elements	11
Number of Pulses	32
Pulse Repetition Frequency	1984 Hz
1000 Range Bins	35 km to 50 km
91 Azimuth Angles	87°, 89°, ... 267°
128 Doppler Frequencies	-992 Hz, -976.38 Hz, ..., 992 Hz
Clutter Power	40 dB
Number of Targets	226 (200 detectable targets)
Range of Target Dop. Freq.	-99.2 Hz to 372 Hz

represent real-world ground clutter and captures variations in underlying terrain, foliage and urban/manmade structures. Further, the KASSPER data set exhibits two very desirable characteristics from the viewpoint of evaluating covariance estimation techniques: 1.) the low-rank structure of clutter in KASSPER has been verified by researchers before [5, 23], and 2.) the true covariance matrices for each range bin have been made available - this facilitates comparisons via powerful figures of merit where the theoretical upper/lower bounds are known.

The L-band data set consists of a data cube of 1000 range bins corresponding to the returns from a single coherent processing interval from 11(= J) channels and 32(= P) pulses. Therefore, the dimension of observations (or the spatio-temporal product) N is $11 \times 32 = 352$. Other key parameters are detailed in Table 3.1. Finally, a clutter rank² of $r = J + P - 1 = 42$ was used by our RCML estimator in all the results to follow, unless explicitly stated otherwise.

We evaluate and compare four different covariance estimation techniques:

- **Sample Covariance Matrix:** The sample covariance matrix is given in Eq. (3.1). It is well known that the sample covariance is the unconstrained maximum likelihood estimator under Gaussian disturbance statistics. Consistent with radar literature [57], we'll refer to the use of this technique as SMI.
- **Fast Maximum Likelihood:** The fast maximum likelihood (FML) [20] uses the structural constraint of the covariance matrix which is given in Eq. (3.2). The FML method just involves calculating the eigenvalue decomposition of the sample

²We set clutter ridge parameters so that $\gamma = 1$.

covariance and perturbing eigenvalues to conform to the structure in Eq. (3.2). The noise variance σ^2 is assumed known or pre-estimated. FML's success in radar STAP is widely known [6, 8, 23].

- **Leave-one-out shrinkage estimator:** Shrinkage estimators are powerful estimators of covariance for high dimensional data that are known to also perturb the eigenstructure of the sample covariance matrix³ [21] - often to ensure non-singularity of the estimated covariance. While a variety of shrinkage techniques are known [21, 67–69], we choose the leave-one-out covariance matrix estimate (LOOC) shrinkage estimator [16],

$$\mathbf{R} = \beta \text{diag}(\mathbf{S}) + (1 - \beta)\mathbf{S} \quad (3.66)$$

The value of β is determined via a cross-validation technique so that the average likelihood of omitted samples is maximized. We pick this estimator because it is most suited to the problem at hand and has demonstrated success in the $K \leq N$ training regime [16].

- **Eigencanceler:** The eigencanceler (EigC) is based on the eigenanalysis which suggests a small number of eigenvalues contain all the information about interferences (jammers and clutter), and therefore, the span of the eigenvectors associated with these significant eigenvalues includes all the position vectors that comprise the interference signals [19]. Since we assume that the rank is known a priori, the eigencanceler can be compared with our estimator as we use r dominant eigenvectors as interference eigenvectors. The covariance matrix can be expressed by

$$\mathbf{R} = \sum_{i=1}^r p_i \mathbf{v}_i \mathbf{v}_i^H + \sigma^2 \mathbf{I} \quad (3.67)$$

where p_i and \mathbf{v}_i are the clutter power and the eigenvector corresponding to r dominant eigenvalues, respectively. For $p_i \gg \sigma^2$, it follows from [23, 72] that the estimated inverse covariance matrix can be approximated as $\hat{\mathbf{R}}^{-1} \approx \frac{1}{\sigma^2}(\mathbf{I} - \mathbf{P})$

where $\mathbf{P} = \sum_{i=1}^r \mathbf{v}_i \mathbf{v}_i^H$. We apply this inverse covariance matrix in computing the SINR.

- **Rank Constrained Maximum Likelihood:** Our proposed estimator (abbrevi-

³Via this definition, the FML and RCML can also be seen as a special class of shrinkage estimators.

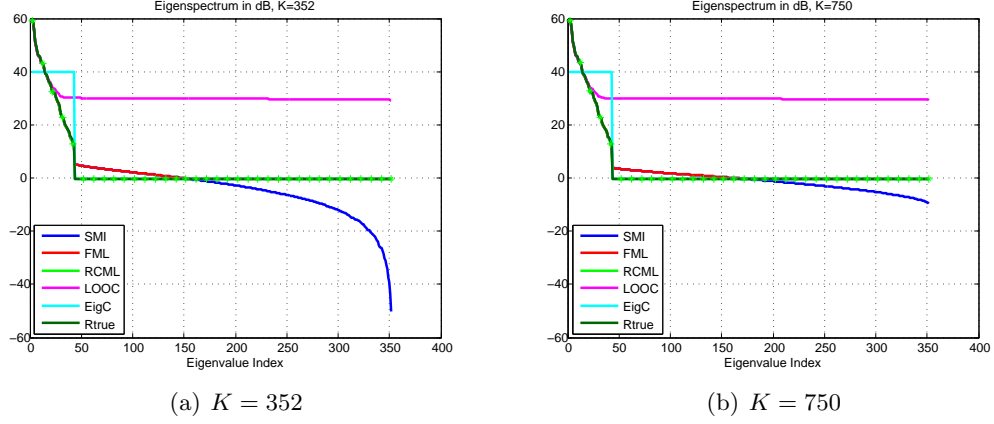


Figure 3.1. Eigenspectra of the estimated covariance matrices.

ated to RCML) incorporates the structural constraint and for the first time the information of the rank of the clutter component.

3.4.2 Experimental Evaluation

The normalized signal to interference and noise ratio (SINR) is used for evaluation the aforementioned covariance estimation techniques. The SINR is desired to be as high as possible. This figure of merit is plotted against azimuthal angle as well as Doppler frequency for distinct training regimes, i.e. low, representative and generous training. We also show the plot of SINR performance versus the number of training samples. Finally, we also evaluate the robustness of our RCML estimator against perturbations in the knowledge of the true rank.

3.4.2.1 Eigenspectrum

First, we provide the eigenspectra plots, i.e. a plot of the eigenvalues of the estimated covariance matrices vs. those of the ground truth covariance. Figure 3.1 shows the eigenspectra of various estimators for $K = N = 352$ and $K = 750$ in dB. We can see the decay of eigenvalues of the true covariance matrix is readily apparent. The first few dominant eigenvalues are well predicted by every method but we notice that after the index 20 or so RCML shows the tightest overlap with the eigenvalues of the true covariance matrix with FML and EigC approaches not far behind. In particular, while FML shows slightly better agreement in capturing the decay, EigC like RCML is more accurate in capturing the rank. The SMI, FML and LOOC approaches are in fact very

significantly off in their estimate of the clutter rank which is determined as an outcome for these methods. In the next Section, we will demonstrate the benefits of incorporating rank and structural information about the disturbance covariance for widely used figures of merit in the radar literature.

3.4.2.2 Normalized SINR vs. angle and Doppler

The normalized SINR measure [47] is commonly used in the radar literature and is given by

$$\eta = \frac{|\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s}|^2}{|\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{R} \hat{\mathbf{R}}^{-1} \mathbf{s}| |\mathbf{s}^H \mathbf{R}^{-1} \mathbf{s}|} \quad (3.68)$$

where \mathbf{s} is the spatio-temporal steering vector, $\hat{\mathbf{R}}$ is an estimated covariance matrix, and \mathbf{R} is the corresponding true covariance matrix. It is easily seen that $0 < \eta \leq 1$ and $\eta = 1$ if and only if $\hat{\mathbf{R}} = \mathbf{R}$. Since the steering vector is a function of both azimuthal angle and Doppler frequency, we evaluate the normalized SINR in both angle and Doppler domain. This would lead to a SINR surface as a function of azimuthal angle and Doppler and comparing surface plots across different covariance estimation techniques is cumbersome. We therefore obtain plots as a function of one variable (i.e. just angle/Doppler) by marginalizing (averaging) over the other variable. The SINR is plotted in dB. in all figures in this dissertation, that is, $\text{SINR}_{\text{dB}} = 10 \log_{10} \eta$. Therefore, $\text{SINR}_{\text{dB}} \leq 0$.

Figures 3.2 and 3.3 plot the variation of normalized SINR as a function of the azimuthal angle and the Doppler frequency for varying number of training samples, K . Specifically, Figures 3.2(a) and 3.2(b) are corresponding to $K = 300 < N = 352$, Figures 3.2(c) and 3.2(d) plot results for $K = 352 = N$, likewise Figures 3.3(a), 3.3(b) and 3.3(c), 3.3(d) are corresponding to $K = 750 \approx 2N$ and $K = 3000 \gg N$ respectively.

Figures 3.2 reports results for the challenging regime of $K \leq N$. When $K < N$ the sample covariance matrix is not invertible, hence for the results in Figures 3.2 and 3.3 we used its pseudo-inverse as a substitute. Unsurprisingly, the sample covariance technique suffers tremendously when $K \leq N$ as is evident from Figures 3.2. LOOC shrinkage does considerably better than SMI because it forces a reasonably good eigenstructure. The informed estimators, i.e. FML, EigC, and RCML perform appreciably well with RCML affording the best overall performance. It is useful to note that RCML in fact offers about 1 dB improvement over FML.

Even for representative training in Figures 3.3(a) and 3.3(b), the vastly superior performance of the FML, EigC, and RCML techniques is apparent. Again, by virtue

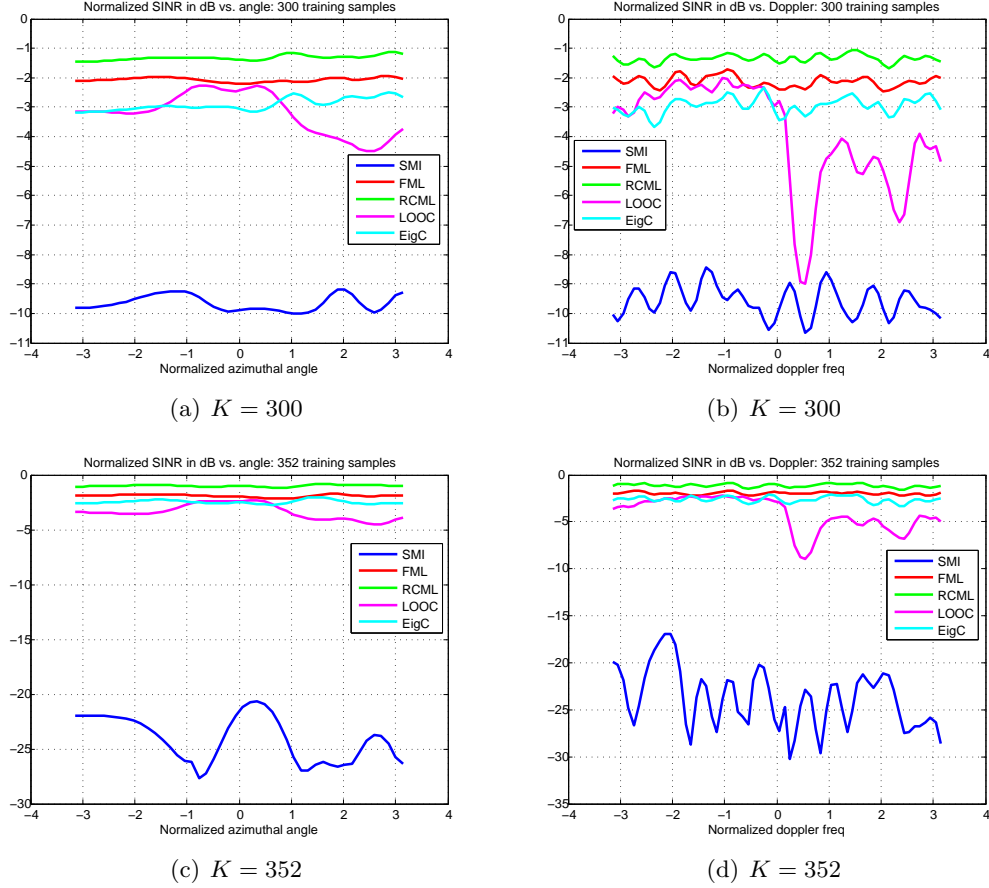


Figure 3.2. Normalized SINR vs. normalized azimuthal angle and doppler frequency.

of incorporating the rank information, the proposed RCML estimator outperforms the competing methods. Finally, Figures 3.3(c) and 3.3(d) confirm the intuition that as training becomes close to asymptotic, the gap between the various methods begins to decrease - of course, such generous training is typically impossible to obtain in practice. This is due to the fact that all the covariance matrix estimates considered converge to the true covariance matrix in the limit of large training data.

3.4.2.3 Performance vs. number of training samples

While the results in Sections 3.4.2.2 do explore performance against training to some extent - here we present bar graphs to explore this issue with a finer granularity. To obtain a single scalar performance measure as a function of training, averaging was carried out over both the angle and Doppler variables.

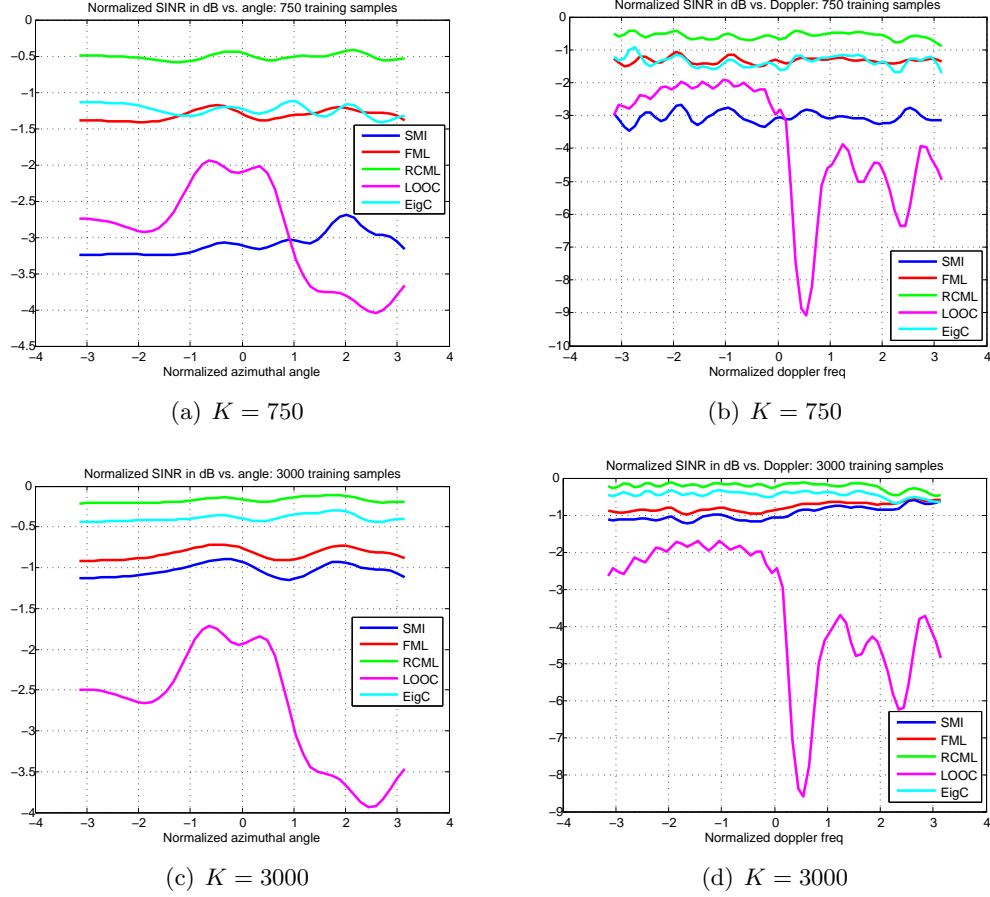


Figure 3.3. Normalized SINR vs. normalized azimuthal angle and doppler frequency.

Figure 3.4 presents bar graphs that quantify the SINR (in dB) as a function of training samples K , where K is varied from as low as 60 to as high as 3000. Two trends are evident from Figure 3.4: 1.) as intuitively expected, the SINR values increases monotonically with an increase in the number of training samples for all methods (except for the sample covariance technique in the $K \leq N$ regime which is a well-known phenomena observed in past work as well [20]) and 2.) the RCML estimator exhibits remarkably good performance in all training regimes.

3.4.2.4 Rank Sensitivity

The KASSPER data, the clutter rank conforms to Eq. (3.3) - the Brennan rule. For the parameters used in our experiments, this would lead to a predicted ideal rank of $r = J + P - 1 = 42$. In a practical situation, departures from the ideal behavior are

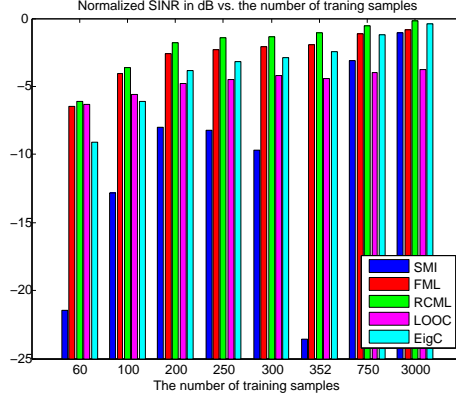


Figure 3.4. Normalized SINR vs. the number of training samples.

expected and hence we explore the performance our proposed RCML estimator even as *incorrect* rank information is used.

The results in Figure 3.5 demonstrate the robustness of RCML to perturbations in the clutter rank. Figure 3.5 presents bar graphs that show averaged SINR results for $K = 352$ and $K = 750$ training samples. We determined numerically that the “true” rank of the clutter covariance for the range bin of choice was in fact 43 which is a mild departure from the 42 predicted by the Brennan rule. Comparisons are made between FML and RCML with the difference that seven variants of RCML are presented - with rank from 34 to 45. As Figure 3.5 reveals, using the true rank of 43 indeed yields the best covariance matrix estimator but the penalty of the small departure, i.e. using a rank of 40 to 45 which are close to the true rank 43 leads to a very small performance loss. On the other hand, Figs. 3.5 also shows variants of the RCML result with a somewhat bigger departure, i.e. a rank of 34. In this case, the performance of RCML with rank 34 is appreciably lower against using rank values around the true rank 43. Remarkably, RCML with rank 34 is still competitive with FML. Overall Figure 3.5 therefore provides two valuable insights: 1.) since rank information is predicted using the Brennan rule - small departures in practice are possible and our estimator exhibits desirable robustness against such small perturbations to rank, and 2.) the value of using the rank information is simultaneously revealed - because RCML with rank 34 is competitive with FML which does not exploit the rank information.

The experiments in Section 3.4.2.1 to Section 3.4.2.4, however, assume that we have access to homogeneous training samples, which is often not available in practice. This section provides more realistic and challenging practical evaluation by means of two new flavors of experimental results: 1.) plots of probability of detection versus SNR for a

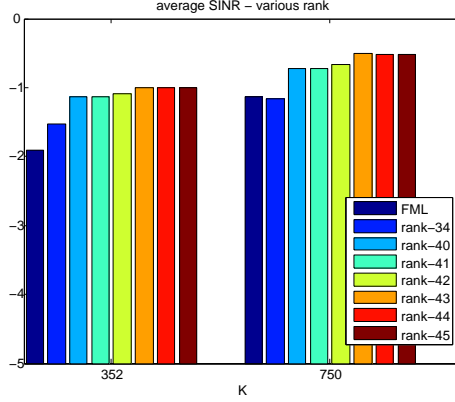


Figure 3.5. Normalized SINR of rank constrained maximum likelihood (RCML) for various rank information

variety of detection statistics, and 2.) normalized SINR performance in the presence of *heterogeneous* training samples which are corrupted by the target information. Here, we consider an experimental environment which reflects real-world scenarios by considering non-homogenous training. We perform two experimental investigations. First, we examine if incorporating rank-information really leads to better target detection. Second, robustness to target contamination in training samples is investigated - while outlier removal techniques have been proposed [73,74], in practice target contamination of training data cannot be entirely ruled out. We evaluate and compare three different covariance estimation techniques, SMI, FML and our proposed RCML. We show that the RCML estimator can still outperform alternatives in that detection probability is second only to the theoretic upper bound when the true covariance is known, and the rank information is invaluable in yielding meaningful estimates even as almost all available training samples are corrupted.

3.4.2.5 Probability of Detection Vs. SNR

We apply three test statistics, the normalized matched filter (NMF), the adaptive matched filter (AMF) [41], and the generalized likelihood ratio test (GLRT) [75]. The test statistics are given by

$$\text{NMF: } \frac{|\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{e}|^2}{(\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s})(\mathbf{e}^H \hat{\mathbf{R}}^{-1} \mathbf{e})} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda_{\text{NMF}} \quad (3.69)$$

$$\text{AMF: } \frac{|\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{e}|^2}{\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s}} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda_{\text{AMF}} \quad (3.70)$$

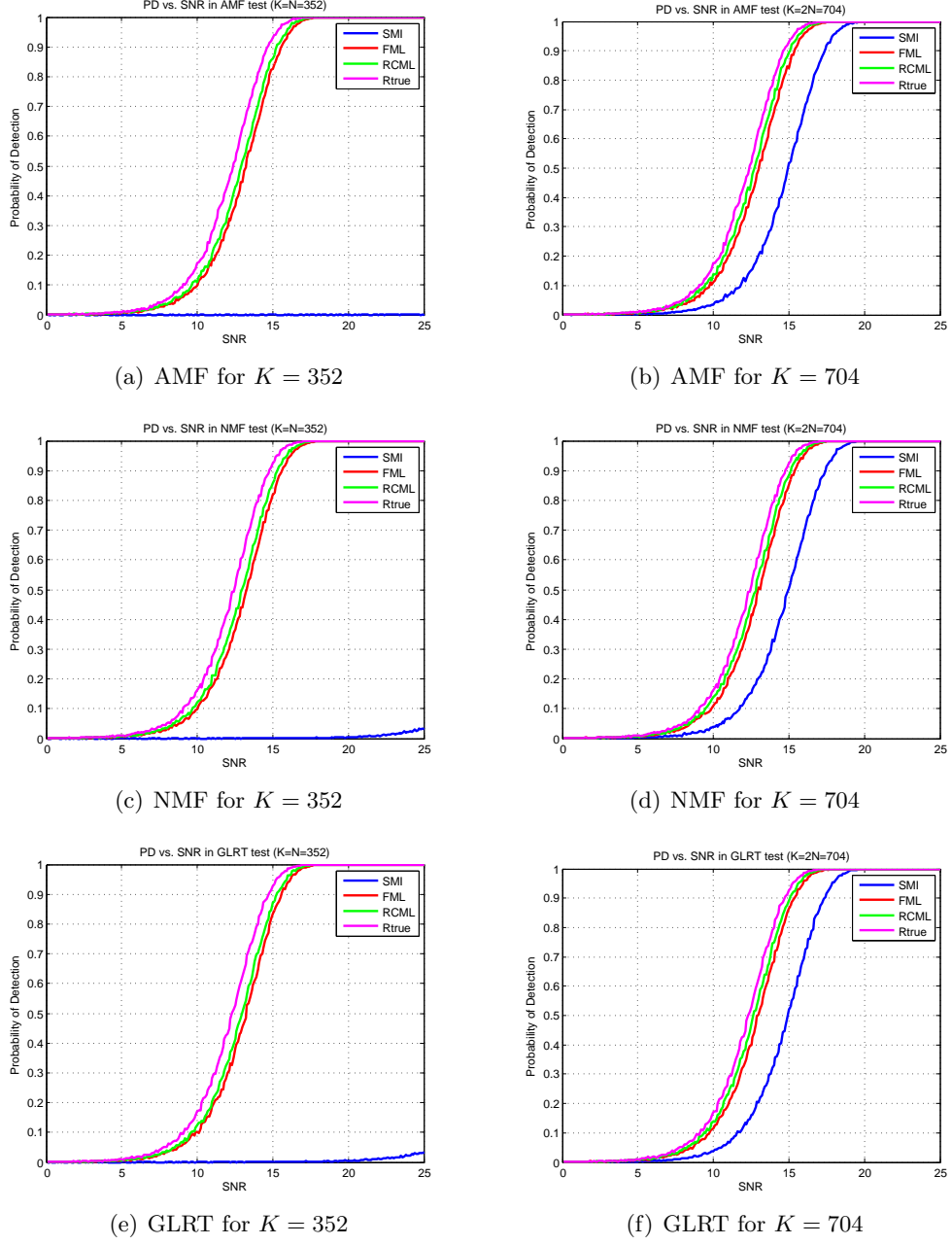


Figure 3.6. Probability of detection vs. SNR.

$$\text{GLRT: } \frac{|s^H \hat{\mathbf{R}}^{-1} \mathbf{e}|^2}{s^H \hat{\mathbf{R}}^{-1} \mathbf{s} \left(1 + \frac{1}{K} \mathbf{e}^H \hat{\mathbf{R}}^{-1} \mathbf{e} \right)} \underset{H_0}{\overset{H_1}{\geq}} K \lambda_{\text{GLRT}} \quad (3.71)$$

Table 3.2. Running time (sec) for various estimators for KASSPER dataset. FML, RCML, and EigC involve eigenvalue decomposition which takes 0.0497 sec.

K	352	528	704
SMI	0.0153	0.0241	0.0294
FML	0.2877	0.3121	0.3292
RCML	0.1054	0.1216	0.1311
EigC	0.2853	0.3178	0.3319
LOOC	12.0666	13.0476	14.5000

where \mathbf{s} , $\hat{\mathbf{R}}$, \mathbf{e} , and K are the steering vector, the estimated covariance matrix, the observation vector, and the number of training samples, respectively. The detection probability P_d is defined as the probability that the value of test statistic is greater than a threshold conditioned on the hypothesis that the received data includes target information. Therefore, it depends on signal to noise ratio (SNR, by virtue of \mathbf{s} ,) and the estimated covariance matrix. Since P_d does not typically admit a closed form, we first generate a number of samples from the L-band data set of KASSPER program to determine λ corresponding to the fixed false alarm rate and then employ Monte Carlo simulations to evaluate P_d corresponding to each estimator for each of the test statistics. We set a constant false alarm rate to 10^{-4} .

Figure 3.6 shows the detection probability P_d plotted as a function of SNR for different estimators and detection statistics. Figures 3.6(a) and 3.6(b) plot P_d for AMF test, Figures 3.6(c) and 3.6(d) are corresponding to the NMF test, and Figures 3.6(c) and 3.6(d) plot results for the GLRT. We use $K = N = 352$ and $K = 2N = 704$ training samples to estimate the covariance matrix for each of the test statistics. Figures 3.6(a), 3.6(c), and 3.6(e) are for $K = 352$ and Figures 3.6(b), 3.6(d), and 3.6(f) are for $K = 704$. It is well-known that $K = 2N$ training samples are needed to keep the performance within 3dB. Indeed, we can see that the sample covariance matrix has about 3dB loss vs. the true covariance matrix in all of test statistics. The proposed RCML estimator is the closest to the P_d achieved by using the true covariance matrix (upper bound) and FML follows RCML. As expected, each estimator shows higher detection probability when $K = 2N$ vs. $K = N$, i.e. an increase in training. Note finally that the RCML estimator performs the best no matter which test statistic is applied and in every regime of training.

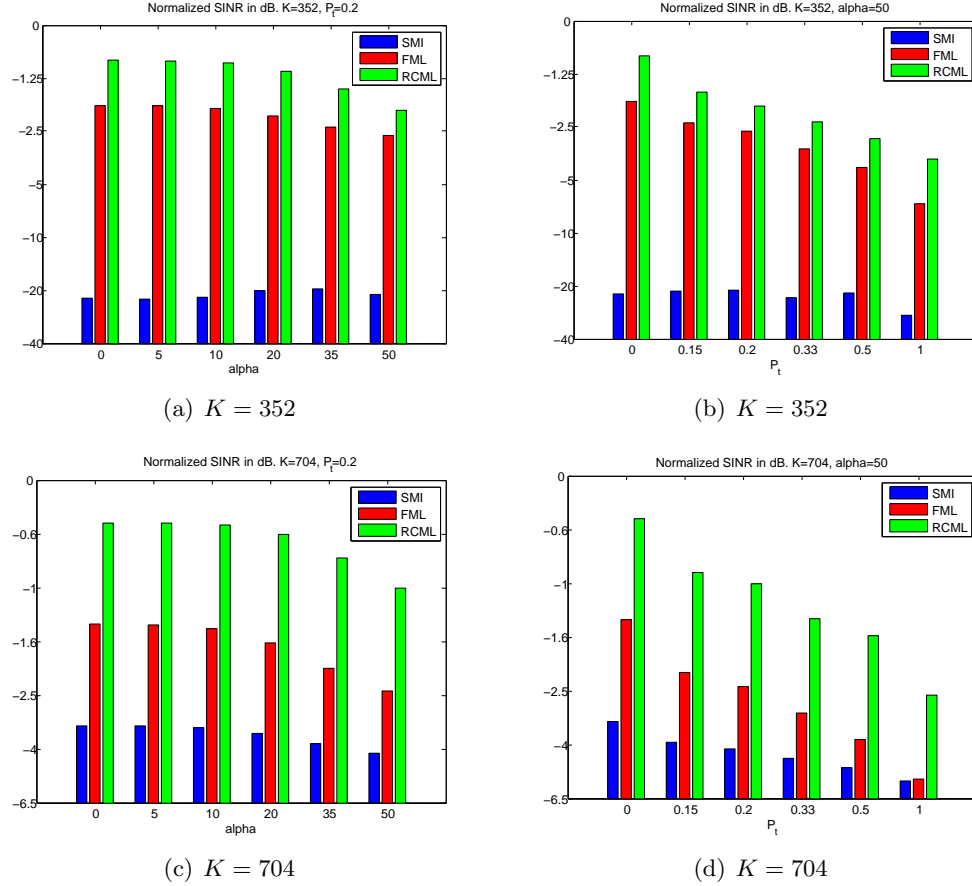


Figure 3.7. Normalized SINR vs. target intensity α and percentage corruption P_t .

3.4.2.6 Complexity Comparison

We compare computational complexity of the compared methods. Table 3.2 shows running times in second for SMI, FML, RCML, eigencanceler, and LOOC. We take average values of results of 100 trials for each estimator and the experiments are performed on the desktop with Intel Core i7-2600 CPU 3.40 GHz and 8.00 GB RAM. As shown in the table, running times increase for all methods as the number of training samples increase. SMI is fastest in all training regimes as expected and RCML is the second best in terms of running time. FML and eigencanceler show running time close to each other. Please note that FML, RCML, and eigencanceler involve eigenvalue decomposition and running time of eigenvalue decomposition in this experiment is 0.0492 second. This result confirms that the RCML estimator not only outperforms the other estimators in the sense of the normalized SINR but also is even computationally cheaper than alternatives.

3.4.2.7 Robustness to Nonhomogeneous Training Samples

We investigate two different scenarios to evaluate robustness to nonhomogeneous training samples. First, we fix the ratio of the number of corrupted samples including target information to target-free samples. This ratio is given by

$$P_t = \frac{\text{the number of corrupted samples by target information}}{K(=\text{the number of total training samples})} \quad (3.72)$$

and the intensity of target signal by α , that is, the received data \mathbf{z} can be expressed by

$$\mathbf{z} = \alpha \mathbf{s}(\theta_t, f_t) + \mathbf{d} \quad (3.73)$$

where $\mathbf{d} = \mathbf{c} + \mathbf{j} + \mathbf{n}$ represents the overall disturbance which is the sum of \mathbf{c} , clutter, \mathbf{j} , jammers, \mathbf{n} , the background white noise, and comes from a zero-mean complex circular Gaussian distribution. \mathbf{s} is a known spatio-temporal steering vector [6] which is drawn from a distribution independent of \mathbf{d} . In particular, we examine performance as the percentage of corrupted samples, i.e., P_t is varied while keeping a fixed intensity of the target signal, α . Our second investigation involves varying α for a fixed P_t .

We use two evaluation measures: the normalized SINR and a trace deviation measure, $\text{TRD}(\hat{\mathbf{R}})$. Figure 3.7 presents bar graphs that show averaged SINR_{dB} results for $K = 352$ and $K = 704$ training samples. Because the steering vector is a function of both azimuthal angle and Doppler frequency, we evaluate the normalized SINR in both angle and Doppler domain and average over both domains to get the normalized SINR value represented by each bar. Figures 3.7(a) and 3.7(b) are corresponding to $K = N = 352$ and Figures 3.7(c) and 3.7(d) plot results for $K = 2N = 704$. In particular, Figures 3.7(a) and 3.7(c) plot the variation of the normalized SINR for varying intensity of the steering vector α , where α is varied from as low as 0 to as high as 50. We fixed $P_t = 0.2$ in these plots. Two trends are evident from Figures 3.7(a) and 3.7(c): 1.) as intuitively expected, the SINR values decreases monotonically with an increase in α for all methods (except for the sample covariance technique in the $K = N$ regime) and 2.) the RCML estimator exhibits appreciably good performance in all training regimes. Figures 3.7(b) and 3.7(d) plot the SINR performance for varying P_t where α remains a constant, $\alpha = 50$. The range of P_t is from 0 (no target corruption) to 1 (all the samples are corrupted by target information). Similar trends are observed as well in Figures 3.7(b) and 3.7(d). Again, the RCML estimator consistently outperforms the other methods. An interesting observation is that SINR_{dB} drops more rapidly as a

function of increasing P_t vs. increasing α , which reveals that P_t is a more critical factor than α in influencing estimation with heterogeneous training.

We define a trace deviation measure, $\text{TRD}(\hat{\mathbf{R}}) = |\text{tr}\{\mathbf{R}\hat{\mathbf{R}}^{-1}\}/N - 1|$ that is an alternate way of evaluating the performance of covariance matrix estimators. Intuitively, we can see $\text{tr}\{\mathbf{R}\hat{\mathbf{R}}^{-1}\}/N = 1$ when $\hat{\mathbf{R}} = \mathbf{R}$. Therefore, we can say the goal of estimation is to keep $\text{TRD}(\hat{\mathbf{R}})$ as small as possible, ideally close to 0. Figure 3.8 shows plots bar graphs in the same training regime as Figure 3.7. We plots values of $\text{TRD}(\hat{\mathbf{R}})$ for varying α and P_t and the number of training samples are $K = 352$ and $K = 704$.

We can observe trends similar to those in Figure 3.7. The TRD values monotonically increase as α and P_t increase for all methods. The proposed RCML estimator consistently outperforms other techniques considered in all experiments. Additionally, the merits of RCML in robust estimation are brought out. The TRD values corresponding to both the sample covariance matrix and the FML estimator increase quite dramatically with an increase in α and, especially, P_t . However, in the case of the RCML estimator this increase is more gradual. The TRD values corresponding to RCML in Figure 3.8(d) are in fact still close to 0 even under severe target corruption, i.e. $P_t = 1$.

3.4.3 RCML vs. RCML_{LB} and Wax and Kailath Estimator

Following Anderson's result [52] in statistics, Wax and Kailath [54] reported an ML estimator of a structured covariance estimator also under the rank-constraint as follows:

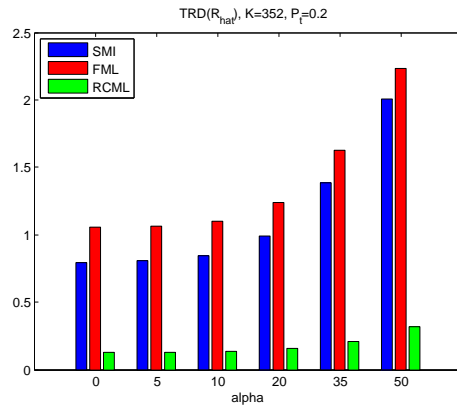
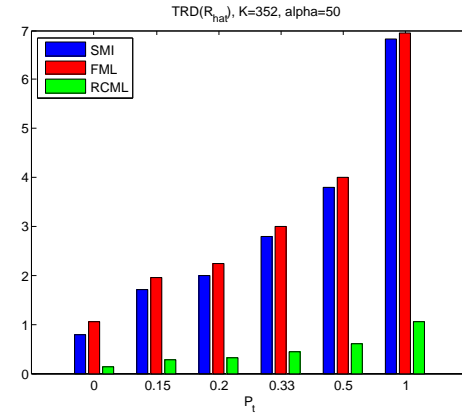
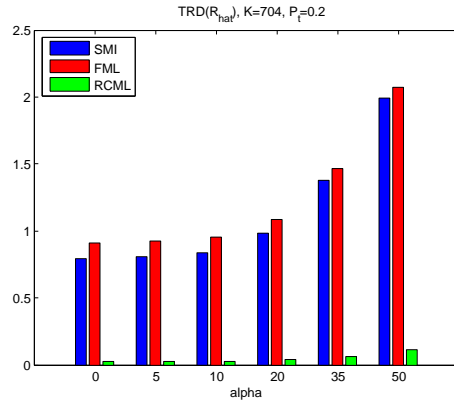
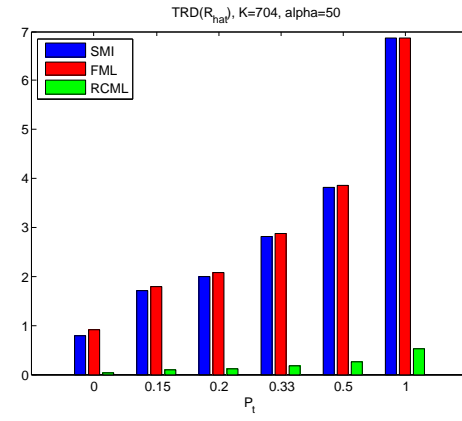
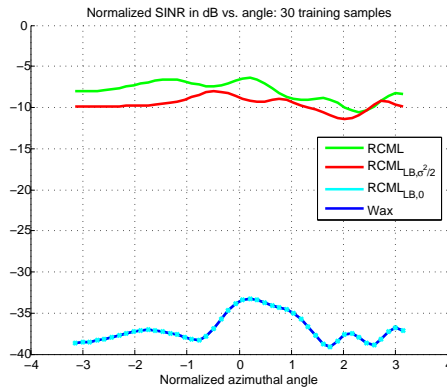
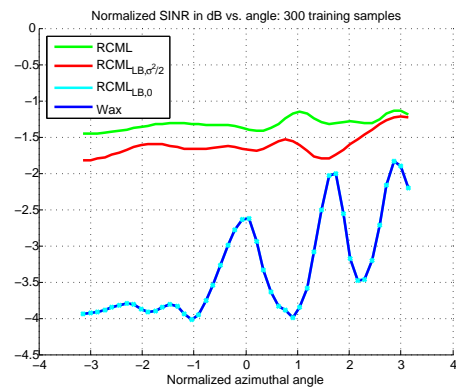
$$\hat{\lambda}_i = \bar{\lambda}_i \quad (3.74)$$

$$\hat{\sigma}^2 = \frac{1}{N-r} \sum_{i=r+1}^N \bar{\lambda}_i \quad (3.75)$$

$$\hat{\Phi}_i = \mathbf{V}_i \quad (3.76)$$

where $\bar{\lambda}_1 \geq \bar{\lambda}_2 \geq \dots \geq \bar{\lambda}_N$ denote the eigenvalues of the sample covariance matrix. It is easy to see that the RCML_{LB} estimator in Eq. (3.65) is a generalization of this result by employing a lower bound on the noise floor.

To further emphasize the value of RCML in Eq. (3.30) is the estimator of choice for practical radar STAP, we now perform an experimental comparison of RCML, RCML_{LB} and Wax and Kailath [54] estimator in the challenging low training regime. Figure 3.9 shows the performance of RCML, RCML_{LB}, and Wax and Kailath estimators for $K = 30$ and $K = 300$ training samples, respectively. Two versions of RCML_{LB} are reported with the lower bound \hat{c} set to $\sigma^2/2$ and 0 and the estimators labeled as RCML_{LB, $\sigma^2/2$} and

(a) $K = 352$ (b) $K = 352$ (c) $K = 704$ (d) $K = 704$ **Figure 3.8.** TRD($\hat{\mathbf{R}}$) vs. target intensity α and percentage corruption P_t .(a) $K = 30$ (b) $K = 300$ **Figure 3.9.** Normalized SINR vs. normalized azimuthal angle.

$\text{RCML}_{\text{LB},0}$ in the plots. Figures 3.9(a) and 3.9(b) clearly reveal that: 1.) RCML is clearly the best estimator, while the Wax and Kailath estimator is about 30 dB worse for $K = 30$ and about 3 dB below for $K = 300$ training samples. 2.) the knowledge of a lower bound helps RCML_{LB} in that $\text{RCML}_{\text{LB},\sigma^2/2}$ is a better estimator than $\text{RCML}_{\text{LB},0}$, which in fact overlaps with the the Wax and Kailath estimator.

3.5 Conclusion

We developed a new estimator of structured covariance matrices (identity plus a positive semi-definite component) which employs rank of the positive semi-definite matrix as an explicit constraint in ML estimation. In radar applications, the rank-deficient component corresponds to the clutter and its rank can be determined using the Brennan rule for airborne radar interacting with land clutter. We demonstrated that despite the presence of the challenging rank-constraint, the estimation problem can in fact be reduced to a convex optimization problem and admits a closed form solution. Experimentally, rank information plays a vital role and rigorous evaluation over the KASSPER data set establishes merits of the proposed estimator when evaluated via widely used figures of merits such as normalized SINR. Future work could consider the incorporation of more constraints on the clutter/disturbance matrix such as Toeplitz structure as well as the use of physically inspired probabilistic priors in a Bayesian setting.

Computationally Efficient Toeplitz Approximation under a Rank Constraint

4.1 Introduction

Radar systems using multiple antenna elements that coherently process multiple pulses offer significant benefits in many applications. The directivity and resolution limits of a single sensor can be overcome by using an adaptive array of spatially distributed sensors makes multiple temporal snapshots processing possible. Specifically, joint adaptive processing in the spatial and temporal domains [2,3,47] called space time adaptive processing (STAP) creates an ability to suppress interference signals while simultaneously preserving gain on the desired signal. For STAP to be successful though, interference statistics, in particular the covariance matrix of the disturbance or interference must be estimated from target free training data, and therefore training plays a pivotal role in adaptive radar systems.

To obtain accurate estimates of the disturbance covariance matrix, a large number of homogeneous (target free) disturbance training samples are required in the absence of any prior knowledge about the interference environment. A compelling challenge for radar STAP emerges since generous homogeneous training is often not available in practice [4]. This problem is exacerbated because the estimation process must be repeated for each range bin of interest. Much recent research in radar STAP has been proposed to overcome the lack of generous homogeneous training. One approach to this problem

uses *a priori* information about the radar environment and is widely referred to in the literature as knowledge-based processing [5, 6, 44, 48–51, 76]. A subset of this technique deals with intelligent training selection for reducing both the number of required training samples and computational cost [6–8]. Another approach to improve the target detection performance is data selection screening among the training data to excise potential outliers [77, 78].

Covariance matrix estimation techniques that enforce and exploit specific structure inherent to the disturbance phenomenon have merit in the regime of extremely limited training data. Examples of structure include persymmetry [9], eigenstructure [19, 20], circulant structure [15], rank constraint [32, 33], multichannel autoregressive models [58, 59], physical constraints [60] and so on. In particular, since the covariance matrix from a stationary stochastic signal is Hermitian and Toeplitz, estimating Toeplitz covariance benefits many applications such as array processing and time series analysis. Such a Hermitian Toeplitz matrix models the covariance of a random vector obtained by sampling a wide sense stationary noise field with a uniform linear array and uncorrelated narrow-band interferers [45]. The seminal work by Burg *et al.* [10] proposed an *iterative* method for estimation of structured covariance matrices using the ML method in its full generality. Li *et al.* developed the asymptotic maximum likelihood (AML) estimation for structured covariance matrices [11] using the extended invariance principle (EXIP) [46]. Approximation of arbitrary matrices by a (Hermitian) Toeplitz matrix using matrix decompositions and outer approximations has separately been pursued in applied mathematics [12, 79–81]. While the techniques in [12, 79–81] were not conceived for signal processing or radar STAP, they can potentially be used in conjunction with classical covariance estimation. Of particular interest is Al-Homidan’s l_1 sequential quadratic programming (SQP) method to find the nearest symmetric positive semi-definite Toeplitz matrix to given a matrix [12].

4.1.1 Motivation and Challenges

Various estimation and approximation techniques of Toeplitz covariance matrices have been proposed [25–28]. It is well known [45] though that there is no closed-form solution for the ML estimation of a Hermitian Toeplitz covariance matrix. Many Toeplitz covariance estimation techniques need the assumption of large sample size (i.e. observed training) for computational tractability [11], [28]. In the regime of realistic training, methods rely on numerical optimization (often non-convex), are computationally involved and hence unsuitable for real-time/practical deployment.

Previous works, notably in statistics [52, 54, 82, 83] (and references therein) have also shown that the rank of the structured interference can be exploited in a tractable manner. Rank is a powerful constraint in covariance estimation and can often be determined via underlying radar physics. Under nominal assumptions, the Brennan rule [24] may be used to determine the rank of the structured interference. Related work also addresses the problem of determining rank in non-ideal scenarios [84]. Recently, Kang *et al.* proposed the rank constrained ML (RCML) estimation of structured covariance matrices [34] which exploits the knowledge of the radar noise floor. Kang *et al.* [34] also report another estimator called RCML_{LB} for the case when the noise floor is assumed unknown and only a lower bound (LB) is available. The RCML_{LB} estimator generalizes the well-known result in statistics [52]. In the radar context though, the noise variance is assumed known since it can be determined by placing the radar in receive only mode [85]. Notable contributions which deal with both the rank information and Toeplitz structure of the covariance matrix jointly include the iterated Toeplitz approximation method (ITAM) [14] proposed by Wilkes and Hayes and the iterative approach by Forster *et al.* [27]. Both approaches are based on a computationally expensive iterative procedure. The ITAM estimator in particular has been shown to be effective under very low training because of its ability to exploit structure but does not yield scalable performance improvements as realistic or generous training is made available.

4.1.2 Our contributions

It may be inferred that for adequate performance under limited training, computationally involved estimators such as ITAM [14] are needed but online covariance estimation is often needed in near real-time. While fast, closed form estimators such as AML [11] can be used, they do not excel under low or realistic training. Our contribution aims to break this classical trade-off. We develop a computationally efficient approximation of structured Toeplitz covariance under a rank (EASTR) constraint¹. Specifically, our key contributions are listed next.

- **Analytically tractable framework for exploiting both Toeplitz structure and the rank of the structured interference.** Our proposed estimator, i.e. EASTR, satisfies both Toeplitz structure property (at least approximately) and the rank information of the structured interference at the same time. Decades of

¹Preliminary version of the work has appeared at IEEE Asilomar Conference on Signals, Systems, and Computers, November 2013 [35] and Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), December 2013 [36].

research has shown that enforcing even each constraint individually can be quite onerous (e.g. rank is a non-convex constraint and no known closed form exists under the Toeplitz constraint for all training). The rank constrained ML estimation problem can be made convex as shown in [34] via a transformation of variables. However, this does not apply when the Toeplitz constraint is added. We propose to decouple the rank and Toeplitz constraints, which lends analytical tractability. Crucially, the EASTR solution does not need iterative steps like ITAM and as will be established in Section 4.3, Furthermore, our results demonstrate that EASTR consistently outperforms ITAM.

- **Computationally efficient and fast estimation and approximation.** Our proposed method, EASTR, essentially involves a cascade of two steps where a closed form solution is available in each step. First a closed form solution using maximum likelihood employing the rank constraint is obtained from the RCML [34, 52] estimator. Next, we propose a new method to perturb the eigenvalues of the RCML estimator in a rank preserving manner so as to impose the Toeplitz structure. We formulate a new quadratic programming (QP) optimization problem that solves for the eigenvalues while incorporating Toeplitz constraints and demonstrate that this problem also admits a closed form solution.
- **Experimental insights and improved performance in low training regimes.** The merits of EASTR are also verified experimentally over both simulated data and realistic data sets such as Knowledge Aided Sensor Signal Processing and Expert Reasoning (KASSPER). ITAM works well particularly in low training regimes but is numerically expensive. The asymptotic ML estimation gives us a fast closed form solution but shows good performances only in high training regimes. EASTR excels across all training regimes while still permitting closed form solutions attractive for practical deployment.

We consider two cases in this chapter: 1.) when the Toeplitz constraint is satisfied exactly, we obtain the exact Toeplitz estimate satisfying the rank constraint and Toeplitz property and 2.) when the Toeplitz constraint is not exactly satisfied, we make a slight modification to the Toeplitz constraint and derive a modified optimization problem to obtain approximately Toeplitz estimate. In practice, the available data dictates which of the two cases is invoked. Experimental investigation shows that EASTR can outperform alternatives in the sense of 1.) normalized SINR and 2.) the probability of detection.

The remainder of this chapter is organized as follows. Section 4.2 develops our proposed estimator, i.e. computationally efficient approximation of structured Toeplitz covariance under a rank constraint (EASTR). Experimental validation of the proposed method is provided in Section 4.3 wherein we report the performance of the proposed method and compare it against widely used existing radar STAP covariance estimators in terms of normalized SINR and the probability of detection. Validation is performed using a popular disturbance covariance simulation model as well as real-world data from the benchmark KASSPER dataset.

4.2 Efficient Approximation of Structured Covariance

The maximum likelihood covariance estimate \mathbf{R} is one which maximizes the likelihood function based on a zero-mean complex circular Gaussian distribution:

$$f(\mathbf{R}; \mathbf{Z}) = \frac{1}{\pi^{NK}} |\mathbf{R}|^{-K} \exp(-\text{tr}\{\mathbf{Z}^H \mathbf{R}^{-1} \mathbf{Z}\}) \quad (4.1)$$

under both Toeplitz and rank constraints. In (4.1), K is the number of training samples, N is the dimension of observations, and \mathbf{Z} is an $N \times K$ matrix whose each column is an i.i.d. observation vector. With some algebraic manipulations, the final optimization problem may be written as

$$\begin{cases} \min_{\mathbf{R}} & \text{tr}\{\mathbf{R}^{-1} \mathbf{S}\} + \log(|\mathbf{R}|) \\ s.t. & \mathbf{R} = \sigma^2 \mathbf{I} + \mathbf{R}_c \\ & \text{rank}(\mathbf{R}_c) = r \\ & \mathbf{R}_c \in T \end{cases} \quad (4.2)$$

where $\mathbf{S} = \frac{1}{K} \mathbf{Z} \mathbf{Z}^H$ is the sample covariance matrix, \mathbf{R}_c denotes the interference covariance matrix, \mathbf{I} is an $N \times N$ identity matrix, and σ^2 is the radar noise floor which can be readily determined using standard techniques [85], and lastly T is the set of all $N \times N$ Hermitian positive semi-definite Toeplitz matrices,

$$T = \{\mathbf{T} : \mathbf{T} \in \mathbb{C}^{N \times N}, \mathbf{T}^H = \mathbf{T}, \mathbf{T} \succeq \mathbf{0} \text{ and } \mathbf{T} \in \mathcal{T}\} \quad (4.3)$$

where \mathcal{T} is the set of all Toeplitz matrices. The optimization problem (4.2) is particularly hard to solve because 1.) the problem is not convex hence a global minimum would be difficult to find, 2.) from a numerical standpoint, solutions are known to be

computationally burdensome under the Toeplitz constraint alone [10], [14]. Adding the rank constraint only exacerbates the problem.

In view of the aforementioned challenges, we focus on covariance matrix estimation that: 1.) is fast and based on analytical closed forms so as to facilitate practical deployment, and 2.) exploits previously known insights in radar STAP so that performance in the sense of high SINR and P_d can be obtained across *all* training regimes.

Our proposed solution decouples the rank and Toeplitz constraints, and develops a cascade of two closed forms as the final estimator. The first closed form is obtained by employing the rank constrained ML (RCML) estimator of structured covariance [34, 52]. The final RCML solution is given by

$$\mathbf{R}^* = \sigma^2 \mathbf{X}^{*-1} = \sigma^2 \mathbf{\Phi} \mathbf{\Lambda}^{*-1} \mathbf{\Phi}^H \quad (4.4)$$

where $\mathbf{\Phi}$ is the eigenvector matrix of the sample covariance matrix \mathbf{S} and $\mathbf{\Lambda}^*$ is a diagonal matrix with optimal diagonal entries λ_i^* which is given by

$$\lambda_i^* = \begin{cases} \min(1, \frac{1}{d_i}) & \text{for } i = 1, 2, \dots, r \\ 1 & \text{for } i = r + 1, r + 2, \dots, N \end{cases} \quad (4.5)$$

where d_i is the i th eigenvalue of the sample covariance matrix normalized by σ^2 , $\mathbf{S}' = \frac{1}{\sigma^2} \mathbf{S}$.

4.2.1 Conditions for Eigenvalues of Toeplitz Covariance

Our approach now involves enforcing the Toeplitz structure on top of the RCML estimator in (4.5). Let the eigenvector matrix of \mathbf{S} be $\mathbf{\Phi}$ and the eigenvalues of \mathbf{R}_c be $\lambda_1, \lambda_2, \dots, \lambda_r, \dots, \lambda_N$. Since we want to preserve the rank constraint of the structured interference $\text{rank}(\mathbf{R}_c) = r$, \mathbf{R}_c should have only r positive eigenvalues and the rest of them should be zero, that is

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > \lambda_{r+1} = \lambda_{r+2} = \dots = \lambda_N = 0 \quad (4.6)$$

Therefore, \mathbf{R}_c can be expressed as

$$\mathbf{R}_c = \mathbf{\Phi} \mathbf{\Lambda} \mathbf{\Phi}^H \quad (4.7)$$

where

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_r & \mathbf{0}_{r \times (N-r)} \\ \mathbf{0}_{(N-r) \times r} & \mathbf{0}_{N-r} \end{bmatrix} \quad (4.8)$$

and $\mathbf{\Lambda}_r$ is an $r \times r$ diagonal matrix with diagonal entries $\lambda_1, \dots, \lambda_r$. Therefore, we know that ij th component of \mathbf{R}_c is given by

$$(\mathbf{R}_c)_{ij} = \sum_{k=1}^r \lambda_k \phi_{ik} \phi_{jk}^* \quad (4.9)$$

where ϕ_{ij} is the (i, j) element of $\mathbf{\Phi}$. Note that \mathbf{R}_c is already Hermitian, that is, $(\mathbf{R}_c)_{ij} = (\mathbf{R}_c)_{ji}^*$. Now in order for \mathbf{R}_c to be Toeplitz matrix, all entries on each diagonal in the lower triangular part in \mathbf{R}_c must have same values, i.e., following equations must hold.

$$\left\{ \begin{array}{llll} (\mathbf{R}_c)_{11} & = & (\mathbf{R}_c)_{22} & = \dots = (\mathbf{R}_c)_{NN} \\ (\mathbf{R}_c)_{21} & = & (\mathbf{R}_c)_{32} & = \dots = (\mathbf{R}_c)_{N,N-1} \\ & & \vdots & \\ & (\mathbf{R}_c)_{N-1,1} & = & (\mathbf{R}_c)_{N2} \end{array} \right. \quad (4.10)$$

Let us examine the first condition in (4.10), $(\mathbf{R}_c)_{11} = (\mathbf{R}_c)_{22}$,

$$\sum_{k=1}^r \lambda_k \phi_{1k} \phi_{1k}^* = \sum_{k=1}^r \lambda_k \phi_{2k} \phi_{2k}^* \quad (4.11)$$

It can be also expressed as

$$\sum_{k=1}^r \lambda_k (\phi_{1k} \phi_{1k}^* - \phi_{2k} \phi_{2k}^*) = 0 \quad (4.12)$$

In vector form, the first equation is given by

$$\begin{bmatrix} \phi_{11} \phi_{11}^* - \phi_{21} \phi_{21}^* & \dots & \phi_{1r} \phi_{1r}^* - \phi_{2r} \phi_{2r}^* \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_r \end{bmatrix} = 0 \quad (4.13)$$

Since the elements ϕ_{ij} of the eigenvalue matrix $\mathbf{\Phi}$ are known ($\mathbf{\Phi}$ is the eigenvector matrix of the sample covariance matrix), we now have the first constraint for Toeplitz covariance matrix as a linear combination of the eigenvalues. Other equations in Eqs. (4.10) also can be expressed in a vector form as in (4.13). Consequentially, we have a total of $N(N-1)/2$ equations and finally get the following equation which is the equality constraint of our

optimization problem.

$$\mathbf{\Psi}\mathbf{\lambda} = \mathbf{0} \quad (4.14)$$

where each row of $\mathbf{\Psi} \in \mathbb{C}^{N(N-1)/2 \times r}$ denotes coefficients of λ_i which come from each of equations in Eqs. (4.10) and $\mathbf{\lambda} = \begin{bmatrix} \lambda_1 & \lambda_2 & \cdots & \lambda_r \end{bmatrix}^T$.

Since $\mathbf{\Psi}$ Eq. (4.14) is a tall matrix, (4.14) in general is a overdetermined linear system, that is, we have more equations than unknowns. The solution set therefore depends on the rank of $\mathbf{\Psi}$. The first case is that we have an infinite set of solutions when the column rank of $\mathbf{\Psi}$ is less than r . On the other hand, when $\mathbf{\Psi}$ has a full column rank, we have the trivial solution, $\mathbf{\lambda} = \mathbf{0}$. That is, the covariance matrix can only be made approximately (and not exactly) Toeplitz in this case - a remedy for this case is discussed in Section 4.2.3.

4.2.2 Exact Toeplitz Solution

When the column rank of $\mathbf{\Psi}$ is less than r , Eq. (4.14) has an infinite number of solutions. In this case, we can obtain the exact Toeplitz solution. First, let $\mathbf{\lambda}_{\text{RCML}}$ be the eigenvalues obtained from the RCML estimation, which is given by Eq. (4.5). We already know the eigenvalues $\mathbf{\lambda}$ from the RCML estimate are the optimal ML estimate of the true structured covariance matrix under only the rank constraint. Therefore, we want the eigenvalues of the interference covariance matrix to satisfy Eq. (4.14) and to be as close to the RCML solution as possible. Since Eq. (4.14) has an infinite number of solutions, we can find the closest vector of the eigenvalues to $\mathbf{\lambda}_{\text{RCML}}$ by solving the following convex optimization problem.

$$\begin{aligned} \min_{\mathbf{\lambda}} \quad & ||\mathbf{\lambda}_{\text{RCML}} - \mathbf{\lambda}||^2 \\ \text{subject to :} \quad & \mathbf{\Psi}\mathbf{\lambda} = \mathbf{0} \end{aligned} \quad (4.15)$$

The optimization problem (4.15) is a well known quadratic programming (QP) optimization problem with an equality constraint and therefore the closed form solution is available using KKT condition [31] and it is given by solving the following equation.

$$\begin{bmatrix} 2\mathbf{I} & \mathbf{\Psi}^T \\ \mathbf{\Psi} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{\lambda}^* \\ \boldsymbol{\nu}^* \end{bmatrix} = \begin{bmatrix} 2\mathbf{\lambda}_{\text{RCML}} \\ \mathbf{0} \end{bmatrix} \quad (4.16)$$

where $\boldsymbol{\nu}^*$ is a vector of Lagrange multipliers.

However, the matrix on the left-hand side of Eq. (4.16) is actually singular because $\mathbf{\Psi}$ has not full column rank. So we introduce a new matrix $\check{\mathbf{\Psi}}$ instead of $\mathbf{\Psi}$ to make the

left matrix invertible when we solve it. That is,

$$\begin{bmatrix} 2\mathbf{I} & \check{\Psi}^T \\ \check{\Psi} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \lambda^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} 2\lambda_{\text{RCML}} \\ \mathbf{0} \end{bmatrix} \quad (4.17)$$

where $\check{\Psi}$ is a matrix consists of $\text{rank}(\Psi)$ linearly independent rows of Ψ . Obviously, Eq. (4.16) and Eq. (4.17) have the same solution because linearly independent $\text{rank}(\Psi)$ rows of Ψ determine the set of solutions of the equation and removing redundant rows does not make any changes to the solution. It follows that the final closed form solution using blockwise inversion property is given by

$$\lambda^* = (\mathbf{I} - \check{\Psi}^T(\check{\Psi}\check{\Psi}^T)^{-1}\check{\Psi})\lambda_{\text{RCML}} \quad (4.18)$$

and the final covariance matrix can be obtained by

$$\mathbf{R}^* = \sigma^2\mathbf{I} + \Phi \text{diag}(\lambda^*)\Phi^H \quad (4.19)$$

4.2.3 Toeplitz Approximation

In the case that Ψ has a full column rank, Eq. (4.14) has the only one solution, $\lambda = \mathbf{0}$, which does not yield a meaningful covariance matrix. In this case, the optimization problem to enforce the Toeplitz structure must be modified. One possibility is to explicitly incorporate the eigenvector matrix into the optimization. This however, will lead to a computationally expensive problem because the optimization must constrain the eigenvector matrix to be unitary. Further, using an eigenvector matrix to agree with Φ , i.e. the one obtained from sample covariance has been known to be very successful in radar STAP [2, 22, 34].

We therefore take the approach of building an *approximately* as opposed to exactly Toeplitz matrix. This can be done by computing the closest rank deficient matrix $\tilde{\Psi}$ to Ψ . Consider the singular value decomposition of Ψ ,

$$\Psi = \mathbf{U}\Sigma\mathbf{V}^H \quad (4.20)$$

The well-known theorem, Eckart-Young theorem [86], says that a matrix $\tilde{\Psi}$ with the column rank less than r that minimizes $\|\Psi - \tilde{\Psi}\|_F$ is given by

$$\tilde{\Psi} = \mathbf{U}\tilde{\Sigma}\mathbf{V}^H \quad (4.21)$$

where $\tilde{\Sigma}$ is the diagonal matrix obtained from Σ by replacing the r -th diagonal element which is the smallest diagonal element by zero. By substituting Ψ with $\tilde{\Psi}$ in Eq. (4.14), we obtain the infinite number of solutions for λ . Now, the optimization problem becomes

$$\begin{aligned} \min_{\lambda} \quad & ||\lambda_{\text{RCML}} - \lambda||^2 \\ \text{subject to :} \quad & \tilde{\Psi}\lambda = \mathbf{0} \end{aligned} \quad (4.22)$$

Finally, a Toeplitz matrix is obtained by solving the above optimization problem in the same way done in the case of exact Toeplitz solution, that is,

$$\lambda^* = (\mathbf{I} - \check{\Psi}^T(\check{\Psi}\check{\Psi}^T)^{-1}\check{\Psi})\lambda_{\text{RCML}} \quad (4.23)$$

where $\check{\Psi}$ is a matrix consists of $r - 1$ linearly independent rows of $\tilde{\Psi}$.

Remark: It should be noted that the actual rank of Ψ which is derived from Φ depends on the training data. If the true covariance is indeed Toeplitz, we expect training samples to reflect that particularly in the regime of $K \gg N$ training samples (asymptotic regime), this is indeed what we observe in practice.

4.3 Experimental Investigation

4.3.1 Experimental Setup and Methods Compared

In this section, we compare the performance of proposed estimator against state of the art Toeplitz STAP estimators. Two data sets are used: 1.) A radar covariance simulation model and 2.) the well known KASSPER [71] data set.

First, we model a radar system with an N -element uniform linear array. The overall disturbance is composed of jammer and white interference. Therefore, the external wideband noise environment via its input covariance matrix can be modeled by

$$\mathbf{R}(n, m) = \sum_{i=1}^J \sigma_i^2 \text{sinc}[0.5\beta_i(n - m)\phi_i] e^{j(n-m)\phi_i} + \sigma_a^2 \delta(n - m) \quad (4.24)$$

where $n, m \in \{1, \dots, N\}$, J is the number of jammers, σ_i^2 is the power associated with the i th jammer, ϕ_i is the jammer phase angle with respect to the antenna phase center, β_i is the fractional bandwidth, σ_a^2 is the actual power level of the white disturbance term, and $\delta(n, m)$ has the value of 1 only when $n = m$ and 0 otherwise. This simulation model has in fact been widely and very successfully used in previous literature [20, 22, 44, 87] for

performance analysis. It is easily seen that \mathbf{R} is Hermitian and Toeplitz since $\mathbf{R}(n, m)$ depends on only $n - m$ and sinc function is an even function. In addition, \mathbf{R} generally has a rank less than N . Therefore, this model can not only be used to simulate radar disturbance samples but also makes ground truth covariance available.

Data from the L-band data set of KASSPER program is the other data set used in our experiments. Note, the KASSPER dataset also makes the true ground truth covariance available and we picked range bins such that their covariance matrices were exactly or approximately Toeplitz. The L-band data set consists of a data cube of 1000 range bins corresponding to the returns from a single coherent processing interval from 11 channels and 32 pulses. Therefore, the dimension of observations (or the spatio temporal product) N is $11 \times 32 = 352$. Other key parameters are detailed in Table 3.1.

We compare the following six different covariance estimation techniques: A host of competing techniques like FML, eigencanceller, and shrinkage estimators have been compared with the RCML method in Chapter 3. The results of Chapter 3 demonstrate that RCML outperforms these techniques under all conditions of training data support and hence they are not reproduced here.

- **Sample Covariance Matrix:** The sample covariance matrix is given by $\mathbf{S} = \frac{1}{K} \mathbf{Z} \mathbf{Z}^H$. It is well known that the sample covariance is the unconstrained maximum likelihood estimator under Gaussian disturbance statistics. We refer to the use of this technique as SML.
- **Iterated Toeplitz Approximation Method:** The iterated Toeplitz approximation method (ITAM) [14] alternatively estimates a rank deficient matrix using the eigenvalue decomposition and then makes the resulting matrix Toeplitz by substituting diagonal entries with the average value of themselves for each diagonal of the estimated matrix. After that, the same process is repeated until the estimated Toeplitz matrix has a desired rank. The estimated covariance satisfies both a desired rank and Toeplitz property and it is closer to the true covariance matrix in the sense of Frobenius norm than the sample covariance matrix.
- **Asymptotic Maximum Likelihood:** The asymptotic maximum likelihood (AML) [11] exploits Toeplitz property of the structured covariance matrix. The authors derived a closed-form formula for Toeplitz covariance matrix estimation and it facilitates computationally efficient implementation. However, they assumed a large number of training samples and their closed-form solution is asymptotically valid. That is, in the low/realistic training regime, estimation performance invariably

suffers.

- **Rank Constrained ML estimator:** The RCML estimator [34] proposed in Chapter 3 exploits the clutter rank information of the structured covariance matrix but not Toeplitz property. It is also the first step of the closed form solution of our proposed method.
- **Sequential Quadratic Programming:** Al-Homidan proposed a sequential quadratic programming (SQP) algorithm to find the nearest symmetric positive semi-definite Toeplitz matrix to given a matrix [12]. There are many other Toeplitz approximation algorithms in applied mathematics [79–81]. We choose the SQP algorithm largely because it guarantees a global minima in approximation error and the l_1 SQP method is considerably faster [12] than alternatives. In practice, the estimator is developed by making a Toeplitz approximation to the RCML estimator. This makes the technique analogous to our proposal of decoupling the rank and Toeplitz constraints in Section 4.2. However, using applied math approximations in a ‘black-box’ manner has two major drawbacks: 1.) the approximation may not necessarily preserve rank and radar STAP specific structure (e.g. eigenvector matrix is perturbed as well), and 2.) the techniques are numerically involved particularly with an increase in data dimension.
- **EASTR:** The proposed Efficient Approximation of Structured covariance under joint Toeplitz and Rank (EASTR) constraints. It incorporates Toeplitz structure, the rank of the clutter component as well as the STAP structural constraint.

In the results to follow, the ITAM, RCML, SQP and EASTR exploit rank information. The clutter rank for the simulation model covariance is of course known and for the KASSPER data set was inferred via the Brennan rule.

4.3.2 Whiteness Test

Before using popular radar STAP measures, we apply a ‘whiteness test’. The trace deviation measure [33] is one way of evaluating covariance matrix estimators since it captures the extent to which the estimated covariance matrix whitens the true covariance matrix. It is given by

$$TRD(\hat{\mathbf{R}}) = |tr\{\mathbf{R}^{-1}\hat{\mathbf{R}}\}/N - 1| \quad (4.25)$$

Intuitively, we can see that its lower bound is zero when $\hat{\mathbf{R}} = \mathbf{R}$ and smaller value of TRD means better performance.

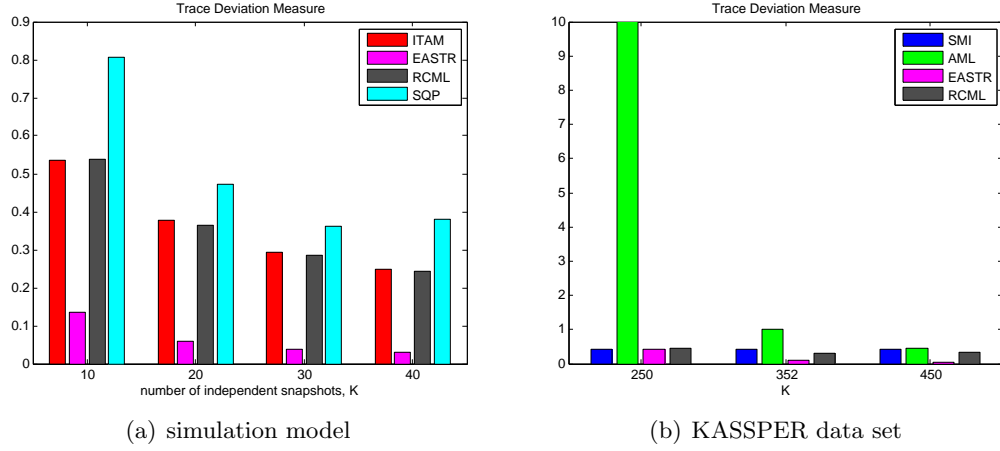


Figure 4.1. Trace deviation measure vs. the number of training samples.

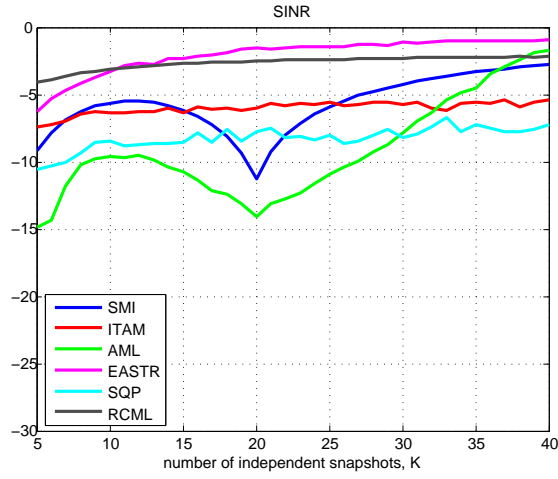


Figure 4.2. Normalized SINR versus the number of training samples for the simulation model

Figure 4.1 shows bargraphs of the performance of compared methods for simulation model and KASSPER data set respectively. Figure 4.1(a) shows bargraphs of the performance in terms of TRD measure versus the number of training samples. Because the SMI and the AML show very high TRD values, we do not plot them in Figure 4.1(a). Figure 4.1(b) similarly shows the result of TRD measure across three training regimes for the KASSPER data set. The TRD measure results in Figures 4.1(a) and 4.1(b) reveal hence that EASTR is in fact “structurally” the closest to the true covariance matrix.

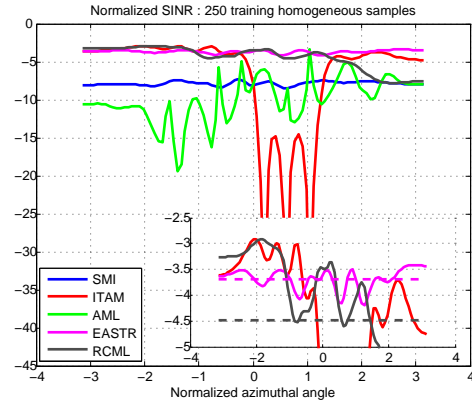
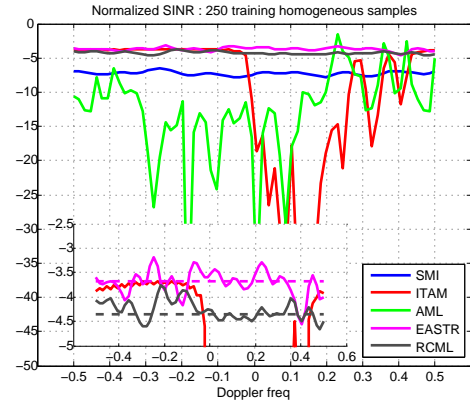
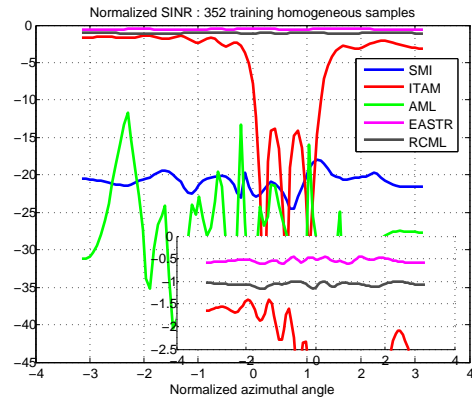
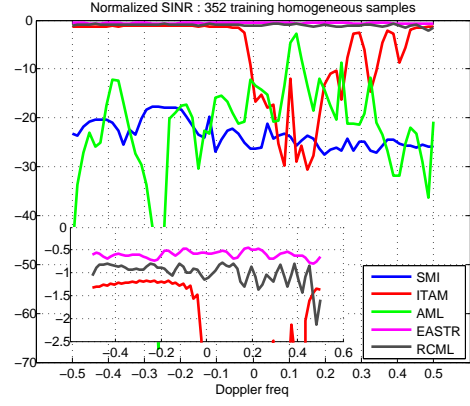
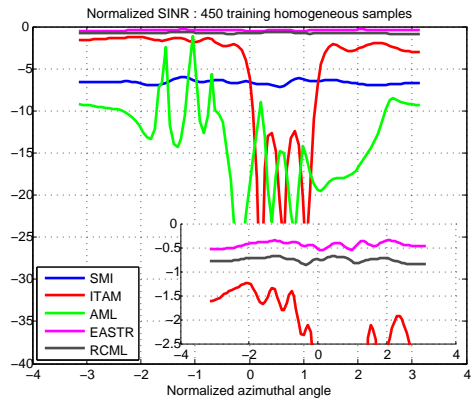
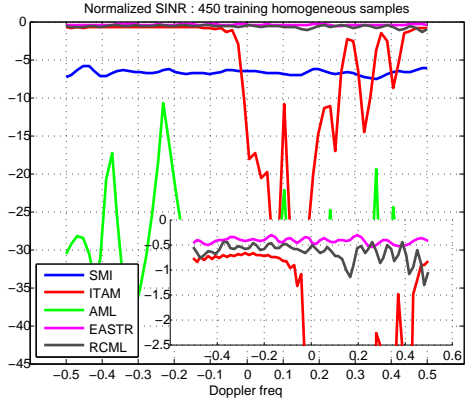
(a) $K(= 250) < N$ (b) $K(= 250) < N$ (c) $K = N = 352$ (d) $K = N = 352$ (e) $K(= 450) > N$ (f) $K(= 450) > N$

Figure 4.3. Normalized SINR versus azimuthal angle Doppler frequency for KASSPER data set.

4.3.3 Normalized SINR

The normalized SINR measure [47] is commonly used in the radar literature. We plot the normalized average SINR versus the number of training samples K for the simulation model in Figure 4.2. In this case, we consider the presence of wideband jamming $J = 3$. In particular, the fractional bandwidth $\beta_i = [0.2, 0, 0.3]$, the powers and phases of jammers are 10 dB, 20 dB, 30 dB and 20 deg, 40 deg, and 60 deg, respectively. $\sigma_a^2 = 1$ and the rank of $r = 7$ are used in the experiments. When $K < N$ the sample covariance is singular, therefore we used its pseudo-inverse instead of inverse itself. Note also that SMI and AML have a dip when $K = 20$ due to numerical instabilities in the $K = N = 20$ training regime. In contrast, ITAM, RCML, EASTR, and SQP guarantee nonsingularity in all training regimes. Interpreting the results in Figure 4.2, it is useful to start with AML which well performs when training is generous $K \gg N$. However, because AML is asymptotically based - its performance is poor when $K < N$ or $K \approx N$ due to numerical instabilities. Even, the SMI and SQP estimators outperform AML when training is low/realistic. ITAM is effective in very low training as expected because it exploits both rank and Toeplitz constraints (though in a largely heuristic way) - ITAM does not exhibit scalable improvements as training support is increased. However, EASTR performs the best overall, even better than RCML (which was recently demonstrated to be the most competitive radar STAP estimator [34]) by virtue of additionally capturing the Toeplitz structure on covariance.

Figure 4.3 plots the normalized SINR results for KASSPER data set as a function of the azimuthal angle and Doppler frequency (averaged over all ranges of Doppler frequency and the azimuthal angle, respectively) for three different training regimes. Specifically, each row of Figure 4.3 is corresponding to $K = 250 (< N)$, $K = N = 352$, and $K > N = 450$ training samples, respectively. Zoomed in versions of ITAM, EASTR, and RCML are shown in each plot to make difference clearly seen. The sample covariance technique and the AML suffer tremendously when $K \leq N$. For low training, ITAM shows comparable performance to the EASTR and the RCML estimators in some ranges of the azimuthal angle but is worse in some other ranges. On an average (over azimuthal angle and Doppler frequency), EASTR is easily the best in Figure 4.3, even providing appreciably gains over the second best RCML estimator. Further, EASTR is stable and effective across all training regimes $K < N$, $K \approx N$ and $K > N$.

4.3.4 Probability of Detection vs. SNR

In order to compute probability of detection, P_d , we apply the normalized matched filter (NMF) [88] as the test statistic given by (3.69). The detection probability P_d is defined as the probability that the value of test statistic is greater than a threshold conditioned on the hypothesis that the received data includes target information. Therefore, it depends on signal to noise ratio (SNR, by virtue of \mathbf{s}_t) and the estimated covariance matrix. Since P_d does not typically admit a closed form, we first generate a number of samples from the true covariance to determine λ corresponding to the fixed false alarm rate and then employ Monte Carlo simulations to evaluate P_d corresponding to each estimator. We set a constant false alarm rate to 10^{-4} .

Figure 4.4 shows the detection probability P_d for simulation model plotted as a function of SNR for different estimators. We use $K = N = 20$ and $K = 2N = 40$ training samples to estimate the covariance matrix in Figure 4.4(a) and Figure 4.4(b), respectively. It is well-known that $K = 2N$ training samples are needed to keep the performance within 3 dB. Indeed, we see that the sample covariance matrix has about 3 dB loss vs. the true covariance matrix in Figure 4.4(b). The proposed EASTR is the closest to the P_d achieved by using the true covariance matrix (upper bound) for both cases. In Figure 4.4(a), we do not plot for ITAM and SQP because they do not guarantee positive semi-definiteness of final estimate in the case of $K = N = 20$, so we cannot calculate the detection probabilities for them.

Figure 4.5 also shows the probability of detection versus SNR plots. We use the same training regimes as used in Section 4.3.3. Figure 4.5(a) and Figure 4.5(b) plot results for $K = 352$ and $K = 2N = 704$, respectively. We can see similar trends in Figure 4.5 hence for KASSPER data to the ones for the simulation model in Figure 4.4. EASTR exhibits the best performance in both plots.

4.3.5 Complexity Comparison

We compare computational complexity of the compared methods for the simulation model. We take average values of results of 100 trials for each estimator and the experiments are performed on the desktop with Intel Core i7-2600 CPU 3.40 GHz and 8.00 GB RAM. Table 4.1 shows running times in second for SMI, ITAM, AML, SQP, and EASTR. Tables 4.1(a), 4.1(b), and 4.1(c) show the results for various data dimensions $N = 20, 40, 80$, respectively. ITAM and EASTR involve eigenvalue decomposition and running times of eigenvalue decomposition for these experiments are shown in caption

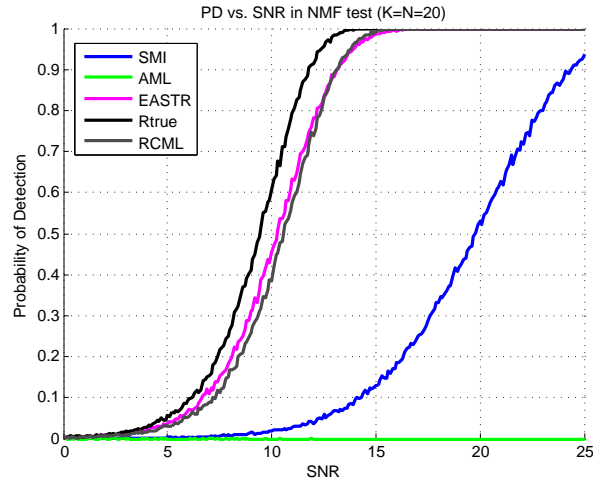
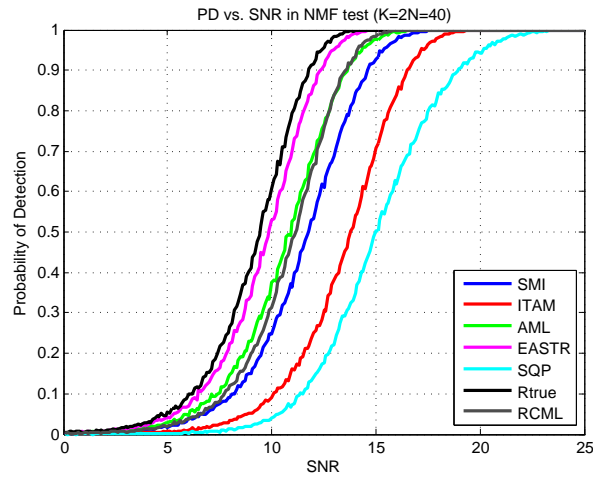
(a) $K = N = 20$ (b) $K = 2N = 40$

Figure 4.4. Probability of detection vs. SNR for simulation model via normalized matched filter (NMF) test.

lines. EASTR is the cheapest method except the sample covariance matrix even though it shows the best performance in the sense of the normalized SINR and the probability of detection. Though a closed form solution is available for AML, the running time of AML increases exponentially as the data dimension increases since AML involves calculations of matrices of which size is N^2 . This results shows that EASTR is the best estimator for the performance as well as for computational complexity.

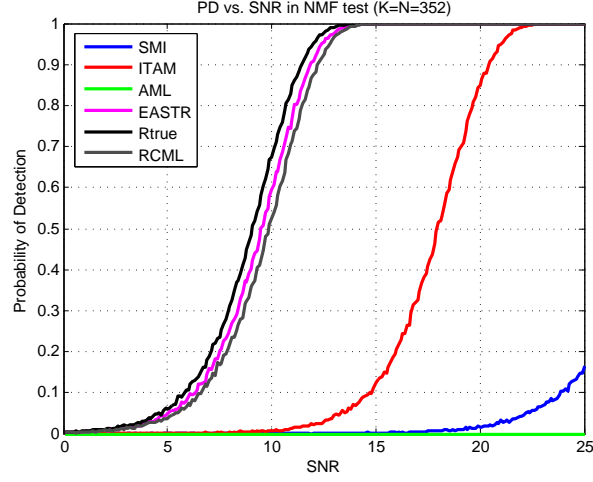
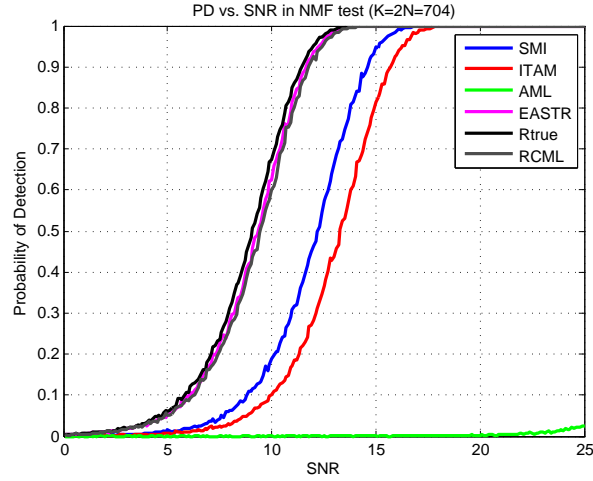
(a) $K = N = 352$ (b) $K = 2N = 704$

Figure 4.5. Probability of detection vs. SNR for KASSPER data set via normalized matched filter (NMF) test.

4.4 Conclusion

Our work focuses on jointly exploiting a Toeplitz structure as well as a rank constraint on the clutter covariance for radar STAP. The problem is inherently hard because it is well known that there is no closed form solution for ML estimation under Toeplitz constraint for all training regimes. While past work has provided iterative often expensive solutions, we develop a new estimator that is based on a cascade of two closed forms. The first closed form is the recently proposed RCML estimator. Our core contribution, the

Table 4.1. Running time (sec) for various estimators for simulation model(a) $N = 20$. Eigenvalue decomposition: 3.4546×10^{-4}

K	20	30	40
SMI	7.8362×10^{-5}	1.0345×10^{-4}	1.3710×10^{-4}
ITAM	0.0939	0.1024	0.1157
AML	0.0615	0.0585	0.0514
SQP	0.2265	0.2128	0.2006
EASTR	0.0212	0.0210	0.0234

(b) $N = 40$. Eigenvalue decomposition: 0.0017

K	40	60	80
SMI	1.4791×10^{-4}	5.5810×10^{-4}	3.7152×10^{-4}
ITAM	0.3915	0.4541	0.4749
AML	0.6593	0.7182	0.7012
SQP	2.7752	2.8421	2.8463
EASTR	0.1115	0.1167	0.1119

(c) $N = 80$. Eigenvalue decomposition: 0.0071

K	80	120	160
SMI	4.7273×10^{-4}	5.1667×10^{-4}	6.7018×10^{-4}
ITAM	2.6020	2.8289	2.5226
AML	15.8855	16.2998	15.9827
SQP	110.3448	109.8137	114.0377
EASTR	0.4939	0.5162	0.5447

second step of Toeplitz approximation performs constrained optimization of eigenvalues to either exactly or approximately satisfy the Toeplitz constraint without compromising the rank. Crucially, this optimization also has a closed form making the overall estimator very friendly from a computational standpoint. Via performance analysis evaluating probability of detection, normalized SINR, and trace deviation measure, our estimator is shown to outperform traditional efforts in Toeplitz and low rank covariance estimation including those based on expensive numerical solutions. Recently, the optimality of the fast maximum likelihood [20] covariance estimator has been proven with respect to cost functions involving the Frobenius or the spectral norm [89]. EASTR can also be investigated for similar notions of optimality. In addition, more analysis of our estimator such as asymptotic convergence can be performed. Finally, practical evaluation may be performed on other radar data sets involving departures from idealized scenarios.

Robust Covariance Estimation under Imperfect Constraints using Expected Likelihood Approach

5.1 Introduction

Radar systems using multiple antenna elements and processing multiple pulses are widely used in modern radar signal processing since it helps overcome the directivity and resolution limits of a single sensor. Joint adaptive processing in the spatial and temporal domains for the radar systems, called space time adaptive processing (STAP) [2, 3, 47], enables to suppress interfering signals as well as to preserve gain on the desired signal. Interference statistics, in particular the covariance matrix of the disturbance, which must be estimated from secondary training samples in practice plays a critical role on success of STAP. To obtain a reliable estimate of the disturbance covariance matrix, a large number of homogeneous training samples are necessary. This gives rise to a compelling challenge for radar STAP because such generous homogeneous (target free) training is generally not available in practice [4].

Much recent research for radar STAP has been developed to overcome this practical limitation of generous homogeneous training. Specifically, the knowledge-based processing which uses *a priori* information about the interference environment is widely referred in the literature [5, 6] and has merit in the regime of limited training data. These techniques include intelligent training selection [5] and the spatio-temporal degrees of freedom reduction [6–8]. In addition, covariance matrix estimation techniques the enforce and

exploit a particular structure have been pursued as one approach of these techniques. Examples of structure include persymmetry [9], Toeplitz structure [11, 13, 45], circulant structure [15], and eigenstructure [20, 22, 34]. In particular, the fast maximum likelihood (FML) method [20] which enforces a special eigenstructure that the disturbance covariance matrix represents a scaled identity matrix plus a rank deficient and positive semidefinite clutter component also falls in this category and is shown to be the most competitive technique experimentally.

Recently, the works by Kang *et al.* [34] and Aubry *et al.* [22] have also improved upon the FML by exploiting practical constraints inspired by physical radar environment, specifically the eigenstructure of the disturbance covariance matrix for radar STAP. They employed a rank of the clutter subspace and a condition number of the interference covariance matrix respectively as a constraint as well as the structural constraint used in the FML into the optimization problem. For both methods, though the initial optimization problems are non-convex, the estimation problems are reduced to a convex optimization problems and admit closed-form solutions. Their methods have also been shown to enable higher normalized SINR over the state-of-the art alternatives for the simulation model and the knowledge-aided sensor signal processing and expert reasoning (KASSPER) data set.

In [34], the authors assume the rank of the clutter is given by Brennan rule [24] under ideal conditions of no coupling. However, in practice (under non-ideal conditions) the clutter rank departs from the Brennan rule prediction due to antenna errors and internal clutter motion. In this case, the rank is not known precisely and needs to be determined before using with the RCML estimator. Determination of the number of signals in a measurement record is a classical eigenvalue problem, which has received considerable attention in the past 60 years. It is important to note that the problem does not have a simple and unique solution. Consequently, a number of techniques have been developed to address this problem [54, 90–93]. In addition, the noise level and the condition number should be estimated as well if they are unknown or non precisely known in practice.

Expected likelihood (EL) approach [30] has been proposed to determine a regularization parameter based on the statistical invariance property of the likelihood ratio (LR) values. More specifically, the probability distribution function (pdf) of LR values for the true covariance matrix depends on only the number of training samples (K) and the dimension of the true covariance matrix (N), not the true covariance itself under a Gaussian assumption on the observations. This statistical independence of LR values on the true covariance itself enables pre-calculation of LR values even though the true co-

variance is unknown. Finally, the regularization parameters are selected so that the LR value of the estimate agrees as closely as possible with the *median* LR value determined via its pre-characterized pdf.

Contributions: In view of the aforementioned observations, we develop covariance estimation methods which automatically and adaptively determines the values of practical constraints via an expected likelihood approach for practical radar STAP. Our main contributions are:

- **A method of choice of constraints using the EL approach:** We propose a method of a choice of practical constraints employed in the optimization problems for covariance estimation in radar STAP using the expected likelihood approach. The proposed method guides the selection of the constraints via the expected likelihood criteria in the case that the knowledge of the constraints is imperfectly known in practice. We consider three different cases of the constraints in this chapter: 1) only the clutter rank constraint, 2) both the clutter rank and the noise power constraints, and 3) the condition number constraint.
- **Analytical results with formal proofs for three different cases of imperfect constraints:** For each case mentioned above, we develop significant analytical results. We first formally prove that the rank selection problem based on the expected likelihood approach has a unique solution. This guarantees there is only one rank which is the best (global optimal) rank in the sense of the EL approach. Second, we derive a closed form solution of the optimal noise power in the sense of the EL approach for a given rank. This means we do not need iterative or numerical method to find the optimal noise power and enables fast implementation. Finally, we also prove there exists the unique condition number for the condition number selection method via the EL approach.
- **Experimental Results through simulated model and the KASSPER data set:** Experimental investigation on a simulation model and on the KASSPER data set shows that the proposed methods for three different cases outperform alternatives such as the FML, leading rank selection methods in radar literature and statistics, and the ML estimation of the condition number constraint in the sense of normalized SINR.

The rest of the chapter is organized as follows. We provide the proposed methods of the constraint selection problems via the EL approach in Section 5.2. Experimental validation of our method is provided in Section 5.3 wherein we report the performance

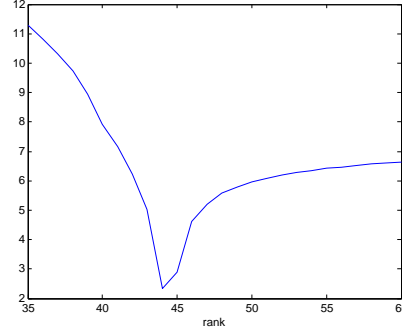


Figure 5.1. $\left(\log \left(\text{LR}(\mathbf{R}_{\text{RCML}}(r), \mathbf{Z}) / \text{LR}_0 \right) \right)^2$ versus r for KASSPER dataset ($K = 2N = 704$)

of the proposed method and compare it against existing methods in terms of normalized SINR on both the simulation model and the KASSPER data set.

5.2 Constraints selection method via Expected Likelihood Approach

5.2.1 Imperfect rank constraint

In Chapter 3, we discuss that the RCML estimator is not only powerful in practice but also computationally cheap and the EL approach is shown to be useful to select parameters so that the estimate is consistent with the true covariance matrix in the sense of the LR value in Section 2.3. From Eq. (4.5), we see the RCML solution is a function of the rank r and d_i 's which are given in the problem. We propose to use the EL approach to refine and find the *optimal* rank when the rank determined by underlying physics is not necessarily accurate.

Now we set up the optimization criterion to find the rank via the EL approach. Since the rank is an integer, there may not exist the rank which exactly satisfies Eq. (2.68). Therefore, we instead find a rank which makes its corresponding LR value closer to LR_0 than any other ranks. That is,

$$\hat{\mathbf{R}}_{\text{RCML}_{\text{EL}}} = \sigma^2 \mathbf{V} \mathbf{\Lambda}^{\star -1}(\hat{r}) \mathbf{V}^H \quad (5.1)$$

where

$$\hat{r} \equiv \arg \min_{r \in \mathbb{Z}} \left| \text{LR}(\mathbf{R}_{\text{RCML}}(r), \mathbf{Z}) - \text{LR}_0 \right|^2 \quad (5.2)$$

and $\text{LR}(\mathbf{R}_{\text{RCML}}(r), \mathbf{Z})$ is given by Eq. (5.3).

Now we investigate the optimization problem (5.2) for the rank selection. Since the eigenvectors of \mathbf{R}_{RCML} are identical to those of the sample covariance matrix \mathbf{S} as shown in Eq. (3.41), the LR value of \mathbf{R}_{RCML} in Eq. (5.2) can be reduced to the function of the eigenvalues of \mathbf{R}_{RCML} and \mathbf{S} . Let the eigenvalues of \mathbf{R}_{RCML} and \mathbf{S} be λ_i and d_i (arranged in descending order). Then the LR value of \mathbf{R}_{RCML} can be simplified to a function of ratio of d_i to λ_i , $\frac{d_i}{\lambda_i}$. That is,

$$\text{LR}(\mathbf{R}_{\text{RCML}}(r), \mathbf{Z}) = \frac{|\hat{\mathbf{R}}_{\text{RCML}}^{-1}(r)\mathbf{S}| \exp N}{\exp \left(\text{tr} [\hat{\mathbf{R}}_{\text{RCML}}^{-1}(r)\mathbf{S}] \right)} \quad (5.3)$$

$$= \frac{\prod_{i=1}^N \frac{d_i}{\lambda_i} \cdot \exp N}{\exp \left[\sum_{i=1}^N \frac{d_i}{\lambda_i} \right]} \quad (5.4)$$

Lemma 3. *The LR value of the RCML estimator, $\text{LR}(\mathbf{R}_{\text{RCML}}(r), \mathbf{Z})$, is a monotonically increasing function with respect to the rank r and there is only one unique \hat{r} in the optimization problem (5.2).*

Proof. First, let r be the largest i such that $d_{i+1} \geq \sigma^2$. Then, from the closed form solution of the RCML estimator, the eigenvalues of the RCML estimator with rank i and $i+1$ for given $i < r$ will be

- $\hat{\mathbf{R}}_{\text{RCML}}(i) : d_1, d_2, \dots, d_i, \sigma^2, \dots, \sigma^2$
- $\hat{\mathbf{R}}_{\text{RCML}}(i+1) : d_1, d_2, \dots, d_i, d_{i+1}, \sigma^2, \dots, \sigma^2$

since $d_{i+1} \geq \sigma^2$. Then $\frac{d_i}{\lambda_i}$ should be

- $\hat{\mathbf{R}}_{\text{RCML}}(i) : 1, 1, \dots, 1_i, \frac{d_{i+1}}{\sigma^2}, \dots, \frac{d_N}{\sigma^2}$
- $\hat{\mathbf{R}}_{\text{RCML}}(i+1) : 1, 1, \dots, 1_i, 1_{i+1}, \frac{d_{i+2}}{\sigma^2}, \dots, \frac{d_N}{\sigma^2}$

From Eq. (5.4), the LR values of the RCML estimators with the ranks i and $i+1$ are

$$\text{LR}(i) = \frac{\frac{\exp N}{\sigma^{2(N-i)}} \prod_{k=i+1}^N d_k}{\exp(i + \frac{1}{\sigma^2} \sum_{k=i+1}^N d_k)} \quad (5.5)$$

$$\text{LR}(i+1) = \frac{\frac{\exp N}{\sigma^{2(N-i-1)}} \prod_{k=i+2}^N d_k}{\exp(i+1 + \frac{1}{\sigma^2} \sum_{k=i+2}^N d_k)} \quad (5.6)$$

From Eq. (5.5) and Eq. (5.6), we obtain

$$\text{LR}(i+1) = \frac{\frac{\exp N}{\sigma^{2(N-i-1)}} \prod_{k=i+2}^N d_k}{\exp(i+1 + \frac{1}{\sigma^2} \sum_{k=i+2}^N d_k)} \quad (5.7)$$

$$= \frac{\frac{\exp N}{\sigma^{2(N-i)}} \prod_{k=i+1}^N d_k \cdot \frac{\sigma^2}{d_{i+1}}}{\exp(i + \frac{1}{\sigma^2} \sum_{k=i+1}^N d_k) \exp(1 - \frac{d_{i+1}}{\sigma^2})} \quad (5.8)$$

$$= \text{LR}(i) \cdot \frac{\sigma^2}{d_{i+1}} \cdot \exp(\frac{d_{i+1}}{\sigma^2} - 1) \quad (5.9)$$

Eq. (5.9) tells us $\text{LR}(i+1)$ can be calculated by multiplying $\text{LR}(i)$ by the coefficient $\frac{\sigma^2}{d_{i+1}} \cdot \exp(\frac{d_{i+1}}{\sigma^2} - 1)$. Fig. 5.2 shows that

$$\frac{\sigma^2}{d_{i+1}} \cdot \exp(\frac{d_{i+1}}{\sigma^2} - 1) \geq 1 \quad (5.10)$$

for all values of $\frac{\sigma^2}{d_{i+1}}$. Therefore, it is obvious that

$$\text{LR}(i+1) \geq \text{LR}(i), \quad (5.11)$$

which means the LR value monotonically increases with respect to i .

Now, let's consider the other case, $i \geq r$. In this case, since $d_{i+1} < \sigma^2$, it is easily shown that

$$\mathbf{R}_{\text{RCML}}(i) = \mathbf{R}_{\text{RCML}}(i+1) \quad (5.12)$$

Therefore,

$$\text{LR}(i+1) = \text{LR}(i) \quad (5.13)$$

This proves that $\text{LR}(i)$ monotonically increases for all $1 \leq i \leq N$.

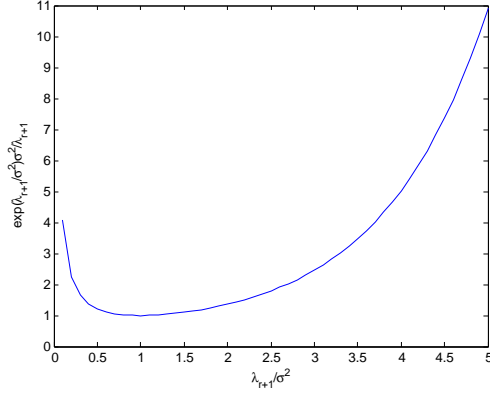


Figure 5.2. The value of the coefficient $\frac{\sigma^2}{\lambda_{r+1}} \cdot \exp(\frac{\lambda_{r+1}}{\sigma^2} - 1)$

□

Lemma 3 gives us a significant analytical result that is the EL approach leads to a unique value of the rank, i.e., when searching over the various values of the rank it is impossible to come up with multiple choices. That also means that it is guaranteed that we can always find the global optimum of r not local optima (minima) for the optimization problem (5.2) regardless of an initial value of r . We plot the values of $\left(\log \left(\text{LR}(\mathbf{R}_{\text{RCML}}(r), \mathbf{Z}) / \text{LR}_0 \right) \right)^2$ versus the rank r for one realization for the KASSPER dataset ($K = 2N = 704$) in Fig. 5.1. Since the LR values are too small in this case, we use a log scale and the ratio between two instead of the distance to see the variation clearly. Note that monotonic increase of the value of $\text{LR}(\mathbf{R}_{\text{RCML}}(r), \mathbf{Z})$ w.r.t r guarantees a unique optimal rank even if the optimization function as defined in (5.2) is not necessarily convex in r .

The algorithm to find the optimal rank is simple and not computationally expensive due to the analytical results above. For a given initial rank such as Brennan rule for the KASSPER data set and the number of jammers for a simulation model, we first determine a direction of searching and then find the optimal rank. The procedure of finding the optimal rank is shown in Algorithm 1 in detail.

5.2.2 Imperfect rank and noise power constraints

In this section, we investigate the second case that both the rank r and the noise power σ^2 are not perfectly known. We propose the estimation of both the rank and the noise level based on the EL approach. The estimator with both the rank and the noise power

Algorithm 1 The proposed algorithm to select the rank via EL

- 1: Initialize the rank r by physical environment such as Brennan rule.
 - 2: Evaluate $\text{LR}(r-1)$, $\text{LR}(r)$, $\text{LR}(r+1)$, the LR values of RCML estimators for the ranks $r-1$, r , $r+1$, respectively.
 - if $|\text{LR}(r+1) - \text{LR}_0| < |\text{LR}(r) - \text{LR}_0|$
 \rightarrow increase r by 1 until $|\text{LR}(r) - \text{LR}_0|$ is minimized to find \hat{r} .
 - elseif $|\text{LR}(r-1) - \text{LR}_0| < |\text{LR}(r) - \text{LR}_0|$
 \rightarrow decrease r by 1 until $|\text{LR}(r) - \text{LR}_0|$ is minimized to find \hat{r} .
 - else $\hat{r} = r$, the initial rank.
-

obtained by the EL approach is given by

$$\hat{\mathbf{R}}_{\text{RCML}_{\text{EL}}} = \hat{\sigma}^2 \mathbf{V} \mathbf{\Lambda}^{\star-1}(\hat{r}) \mathbf{V}^H \quad (5.14)$$

where

$$(\hat{r}, \hat{\sigma}^2) \equiv \arg \min_{r \in \mathbb{Z}, \sigma^2 > 0} \left| \text{LR}(\mathbf{R}_{\text{RCML}}(r, \sigma^2), \mathbf{Z}) - \text{LR}_0 \right|^2 \quad (5.15)$$

In section 5.2.1, we have shown that the optimal rank via the EL approach is uniquely obtained for a fixed σ^2 . Now we analyze the LR values of the RCML estimator for various σ^2 and a fixed rank.

Lemma 4. *For a fixed rank, the LR value of the RCML estimator, which is a function of σ^2 , has a maximum value at $\sigma^2 = \sigma_{ML}^2$. It monotonically increases for $\sigma^2 < \sigma_{ML}^2$ and monotonically decreases for $\sigma^2 > \sigma_{ML}^2$.*

Proof. In this section, I investigate the LR values for varying noise level σ^2 and a given rank r . From Eq. (5.5) we obtain the LR value when the rank is r ,

$$\text{LR}(\sigma^2) = \frac{\frac{\exp N}{\sigma^{2(N-r)}} \prod_{k=r+1}^N d_k}{\exp\left(r + \frac{1}{\sigma^2} \sum_{k=r+1}^N d_k\right)} \quad (5.16)$$

For simplicity, let $\sigma^2 = t$ then Eq. (5.16) can be simplified as

$$\text{LR}(t) = \frac{e^{N-r} \prod_{k=r+1}^N d_k}{t^{N-r} e^{\frac{\sum_{k=r+1}^N d_k}{t}}} \quad (5.17)$$

Now let $\sum_{k=r+1}^N d_k = d_s$ and $\prod_{k=r+1}^N d_k = d_p$, then

$$\text{LR}(t) = \frac{e^{N-r} d_p}{t^{N-r} e^{\frac{d_s}{t}}} \quad (5.18)$$

$$= d_p e^{N-r} t^{r-N} e^{-\frac{d_s}{t}} \quad (5.19)$$

To analyze increasing or decreasing property Eq. (5.19), I calculate its first derivative. Since $d_p e^{N-r}$ is a positive constant, it does not affect increasing or decreasing of the function. Therefore,

$$\begin{aligned} & (t^{r-N} e^{-\frac{d_s}{t}})' \\ &= (r-N)t^{r-N-1} e^{-d_s/t} + t^{r-N} e^{-d_s/t} \frac{d_s}{t^2} \end{aligned} \quad (5.20)$$

$$= (r-N)t^{r-N-1} e^{-d_s/t} + t^{r-N-2} e^{-d_s/t} d_s \quad (5.21)$$

$$= t^{r-N-2} ((r-N)t + d_s) e^{-d_s/t} \quad (5.22)$$

Since t^{r-N-2} and $e^{-d_s/t}$ are always positive, the first derivative $(t^{r-N} e^{-\frac{d_s}{t}})' = 0$ if and only if

$$t = \frac{d_s}{N-r} = \frac{\sum_{k=r+1}^N d_k}{N-r} \quad (5.23)$$

and it is positive when $t < \frac{\sum_{k=r+1}^N d_k}{N-r}$ and negative otherwise. This means that $\text{LR}(\sigma^2)$ increases for $\sigma^2 < \frac{\sum_{k=r+1}^N d_k}{N-r}$ and decreases for $\sigma^2 > \frac{\sum_{k=r+1}^N d_k}{N-r}$. The LR value is maximized when $\sigma^2 = \frac{\sum_{k=r+1}^N d_k}{N-r}$. Note that $\frac{\sum_{k=r+1}^N d_k}{N-r}$ is the average value of $N-r$ smallest eigenvalues of the sample covariance matrix and in fact a maximum likelihood solution of σ^2 as shown in the RCML estimator [34]. \square

Fig. 5.3 shows an example of the LR values as a function of the noise level σ^2 . As shown in Lemma 4, we see that the LR value is maximized for the ML solution of σ^2

and monotonically increases and decreases for each direction. It is obvious that we have three cases of the solution of the optimal noise power from Lemma 4: 1) no solution, 2) only one solution, and 3) two optimal solution. Now we discuss how to obtain the optimal noise power for a fixed rank.

Lemma 5. *The noise power obtained by the expected likelihood approach, $\hat{\sigma}_{EL}^2$, is given by*

$$\hat{\sigma}_{EL}^2 = \exp \left(W_k \left(\frac{b}{a} e^{-\frac{c}{a}} \right) + \frac{c}{a} \right) \quad (5.24)$$

where $W_k(z)$ is the k -th branch of Lambert W function and

$$\begin{cases} a = r - N \\ b = \sum_{k=r+1}^N d_k \\ c = \log \text{LR}_0 - \log \left(\prod_{k=r+1}^N d_k \right) + a \end{cases} \quad (5.25)$$

Proof. For a given rank r , the optimal solution of the noise power via the EL approach, $\hat{t}(= \hat{\sigma}_{EL}^2)$, is the solution of $\text{LR}(t) = \text{LR}_0$. From Eq. (5.19), that is, \hat{t} is the solution of the equation given by

$$d_p e^{N-r} t^{r-N} e^{-\frac{d_s}{t}} = \text{LR}_0 \quad (5.26)$$

Taking log on both side leads

$$\log d_p + N - r + (r - N) \log t - \frac{d_s}{t} = \log \text{LR}_0 \quad (5.27)$$

For simplification, we take substitutions of variables,

$$\begin{cases} a = r - N \\ b = \sum_{k=r+1}^N d_k \\ c = \log \text{LR}_0 - \log \left(\prod_{k=r+1}^N d_k \right) + a \end{cases} \quad (5.28)$$

Then, Eq. (5.27) is simplified to an equation of t ,

$$a \log t - \frac{b}{t} = c \quad (5.29)$$

Again, let $u = \log t$. Then, since $t = e^u$, we obtain

$$au - be^{-u} = c \quad (5.30)$$

$$e^{-u} = \frac{a}{b}u - \frac{c}{b} \quad (5.31)$$

Now let $s = u - \frac{c}{a}$. Then, the equation is

$$e^{-s-\frac{c}{a}} = \frac{a}{b}s \quad (5.32)$$

$$se^s = \frac{b}{a}e^{-\frac{c}{a}} \quad (5.33)$$

The solution of Eq. (5.33) is known to be obtained using Lambert W function [94]. That is,

$$s = W\left(\frac{b}{a}e^{-\frac{c}{a}}\right) \quad (5.34)$$

where $W(\cdot)$ is a Lambert W function which is defined to be the function satisfying

$$W(z)e^{W(z)} = z \quad (5.35)$$

Finally, we obtain

$$u = W\left(\frac{b}{a}e^{-\frac{c}{a}}\right) + \frac{c}{a} \quad (5.36)$$

and

$$\hat{\sigma}_{\text{EL}}^2 = \hat{t} = \exp\left(W\left(\frac{b}{a}e^{-\frac{c}{a}}\right) + \frac{c}{a}\right) \quad (5.37)$$

□

Lemma 5 shows that there is a closed form solution of the optimal noise power for a fixed rank. Therefore we do not need any iterative and numerical algorithms to obtain both the optimal rank and noise power.

Now we propose the method to alternately find the optimal solution of both the rank and the noise power. For a fixed σ^2 , we can obtain the optimal rank via Algorithm 1. For a fixed rank, we should consider three cases described above. The first case implies that the LR value corresponding σ_{ML}^2 is less than LR_0 and therefore, we increase the rank until the solution of σ^2 exists. In the second case, we can easily determine $\hat{\sigma}^2 = \sigma_{\text{ML}}^2$. For the third case that there are two solutions of σ^2 , we have to choose one among two EL solutions and the ML solution. We experimentally observe that the threshold in the test statistics such as the normalized matched filter is typically smaller for the better estimator in the sense of the normalized SINR and the probability of detection from our experiments. Therefore, we choose one of σ_{ML}^2 , σ_{EL1}^2 , σ_{EL2}^2 , which generates the smallest value of the test statistics. The detail procedure of the algorithm is described in Algorithm 2.

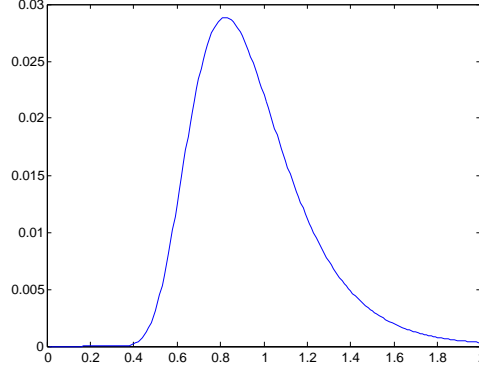


Figure 5.3. The LR value versus σ^2 for the simulation model, $N = 20$, $K = 40$, $r = 5$

Algorithm 2 The proposed algorithm to select the rank and the noise level via EL

- 1: Initialize the rank r by physical environment such as Brennan rule or the number of jammers.
 - 2: If there is no solution of σ^2 for given r , increase r until the solution of σ^2 exists.
 - 3: Obtain $\sigma_{\text{ML}}^2 = \frac{1}{N-r} \sum_{i=r+1}^N d_i$.
 - 4: For given σ_{ML}^2 , find a new r using Algorithm 1.
 - 5: Repeat Step 3 and Step 4 until the rank r converges.
 - 6: After r is determined, choose $\hat{\sigma}^2$ among σ_{ML}^2 , σ_{EL1}^2 , σ_{EL2}^2 .
-

5.2.3 Imperfect condition number constraint

Now we discuss the proposed method to determine the condition number constraint through the EL approach in this section. As shown in Eq. (2.24) through Eq. (2.29), the condition number constrained ML estimator is a function of u which is a function of the condition number K_{max} . Therefore, the final estimate is also a function of K_{max} . Similar to what we have done in previous sections, we find an optimal condition number so that the LR value of the estimated covariance matrix should be same as a statistical median value of the LR value of the true covariance matrix, that is

$$\hat{\mathbf{R}}_{\text{CNCML}_{\text{EL}}} = \hat{\sigma}^2 \mathbf{V} \mathbf{\Lambda}^{\star-1} (\hat{K}_{\text{max}}) \mathbf{V}^H \quad (5.38)$$

where

$$\hat{K}_{\text{max}} \equiv \arg \min_{K_{\text{max}} \geq 1} \left| \text{LR}(\mathbf{R}_{\text{CNCML}}(K_{\text{max}}), \mathbf{Z}) - \text{LR}_0 \right|^2 \quad (5.39)$$

Before we discuss the algorithm to find the optimal condition number, we analyze the closed form solution for the condition number constrained ML estimation which is proposed in [22]. We derive a more explicit closed form solution.

Lemma 6. *The more simplified closed form solution of the condition number constrained*

ML estimator is given by

$$1. d_1 \leq \sigma^2,$$

$$\hat{\mathbf{R}}_{CN} = \sigma^2 \mathbf{I} \quad (5.40)$$

$$2. \sigma^2 \leq d_1 \leq \sigma^2 K_{\max},$$

$$\hat{\mathbf{R}}_{CN} = \hat{\mathbf{R}}_{FML} \quad (5.41)$$

$$3. d_1 > \sigma^2 K_{\max} \text{ and } K_{\max} \geq \frac{\sum_{i=1}^c d_i}{c - \sum_{N+1}^{\bar{N}} (d_i - 1)},$$

$$\hat{\mathbf{R}}_{CN} = \Phi \text{diag}(\boldsymbol{\lambda}^*) \Phi^H \quad (5.42)$$

where

$$\boldsymbol{\lambda}^* = [\sigma^2 K_{\max}, \dots, \sigma^2 K_{\max}, d_{c+1}, \dots, d_{\bar{N}}, \sigma^2, \dots, \sigma^2], \quad (5.43)$$

c and \bar{N} are the vector of the eigenvalues of the estimate, the largest indices so that $d_c > \sigma^2 K_{\max}$, and $d_{\bar{N}} \geq \sigma^2$

$$4. d_1 > \sigma^2 K_{\max} \text{ and } K_{\max} < \frac{\sum_{i=1}^c d_i}{c - \sum_{N+1}^{\bar{N}} (d_i - 1)},$$

$$\boldsymbol{\lambda}^* = \left[\frac{\sigma^2}{u}, \dots, \frac{\sigma^2}{u}, d_{p+1}, \dots, d_q, \frac{\sigma^2}{u K_{\max}}, \dots, \frac{\sigma^2}{u K_{\max}} \right] \quad (5.44)$$

And the condition numbers of the estimates are 1, $\frac{d_1}{\sigma^2}$, K_{\max} , and K_{\max} , respectively.

Proof. We consider 5 cases provided in [22].

$$1. d_1 \leq \sigma^2 \leq \sigma^2 K_{\max}$$

$$\text{Since } u^* = \frac{1}{K_{\max}},$$

$$\lambda_i^* = \min(\min(K_{\max} u^*, 1), \max(u^*, \frac{1}{d_i})) \quad (5.45)$$

$$= \min(\min(1, 1), \max(\frac{1}{K_{\max}}, \frac{1}{d_i})) \quad (5.46)$$

$$= \min(1, \frac{1}{d_i}) = 1 \quad (5.47)$$

Therefore,

$$\hat{\mathbf{R}}_{CN} = \sigma^2 \mathbf{I} \quad (5.48)$$

and the condition number is 1.

2. $\sigma^2 < d_1 \leq K_{\max}$
 Since $u^* = \frac{1}{d_1}$,

$$\lambda_i^* = \min(\min(K_{\max} u^*, 1), \max(u^*, \frac{1}{\bar{d}_i})) \quad (5.49)$$

$$= \min(\min(\frac{K_{\max}}{\bar{d}_1}, 1), \max(\frac{1}{\bar{d}_1}, \frac{1}{\bar{d}_i})) \quad (5.50)$$

$$= \min(1, \frac{1}{\bar{d}_i}) \quad (5.51)$$

$$= \begin{cases} \frac{1}{\bar{d}_i} & \bar{d}_i \geq 1 \\ 1 & \bar{d}_i < 1 \end{cases} \quad (5.52)$$

Therefore,

$$\hat{\mathbf{R}}_{\text{CN}} = \hat{\mathbf{R}}_{\text{FML}} \quad (5.53)$$

and the condition number is $\frac{d_1}{\sigma^2}$.

3. $d_1 > \sigma^2 K_{\max}$ and $u^* = \frac{1}{d_1}$

Since u^* is the optimal solution of the optimization problem (2.27), $\frac{dG(u)}{du}|_{u=\frac{1}{d_1}}$ must be zero if $u^* = \frac{1}{d_1}$. From, Eq. (2.28) and Eq. (2.29), the first derivative of $G_i(u)$ is given by

$$G'_i(u) = \begin{cases} -\frac{1}{u} + K_{\max} \bar{d}_i & \text{if } 0 < u \leq \frac{1}{K_{\max}} \\ 0 & \text{if } \frac{1}{K_{\max}} \leq u \leq 1 \end{cases} \quad (5.54)$$

for $\bar{d}_i \leq 1$, and

$$G'_i(u) = \begin{cases} -\frac{1}{u} + K_{\max} \bar{d}_i & \text{if } 0 < u \leq \frac{1}{K_{\max} \bar{d}_i} \\ 0 & \text{if } \frac{1}{K_{\max} \bar{d}_i} < u \leq \frac{1}{\bar{d}_i} \\ -\frac{1}{u} + \bar{d}_i & \text{if } \frac{1}{\bar{d}_i} \leq u \leq 1 \end{cases} \quad (5.55)$$

for $\bar{d}_i > 1$. Therefore,

$$\frac{dG(u)}{du}|_{u=\frac{1}{d_1}} = \sum_{i=\bar{N}+1}^N (K_{\max} \bar{d}_i - \bar{d}_1) + \sum_{i=p}^{\bar{N}} (K_{\max} \bar{d}_i - \bar{d}_1) \quad (5.56)$$

where p is the greatest index such that $\frac{1}{d_1} < \frac{1}{K_{\max} \bar{d}_p}$. For $i = \bar{N}, \dots, N$, since $\bar{d}_i \leq 1$,

$$K_{\max} \bar{d}_i - \bar{d}_1 < K_{\max} - \bar{d}_1 < 0 \quad (5.57)$$

and for $i = p, \dots, \bar{N} - 1$, since $\bar{d}_1 > K_{\max} \bar{d}_i$, $K_{\max} \bar{d}_i - \bar{d}_1 < 0$. Therefore, in this case, it is obvious that

$$\frac{dG(u)}{du} \Big|_{u=\frac{1}{\bar{d}_1}} < 0 \quad (5.58)$$

which implies $u = \frac{1}{\bar{d}_1}$ can not be the optimal solution of (2.27).

4. $d_1 > \sigma^2 K_{\max}$ and $u^* = \frac{1}{K_{\max}}$

Aubry *et al.* [22] showed that $u^* = \frac{1}{K_{\max}}$ if $\frac{dG(u)}{du} \Big|_{u=\frac{1}{K_{\max}}} \leq 0$. From Eq. (5.54) and Eq. (5.55),

$$\frac{dG(u)}{du} \Big|_{u=\frac{1}{K_{\max}}} = \sum_{i=\bar{N}+1}^N K_{\max}(\bar{d}_i - 1) + \sum_{i=1}^p (\bar{d}_i - K_{\max}) \quad (5.59)$$

where p is the greatest index such that $\bar{d}_p > K_{\max}$. Therefore,

$$\frac{dG(u)}{du} \Big|_{u=\frac{1}{K_{\max}}} \leq 0 \quad (5.60)$$

$$\Leftrightarrow \sum_{i=\bar{N}+1}^N K_{\max}(\bar{d}_i - 1) + \sum_{i=1}^p (\bar{d}_i - K_{\max}) \leq 0 \quad (5.61)$$

$$\Leftrightarrow K_{\max}(\sum_{i=\bar{N}+1}^N (\bar{d}_i - 1) - p) + \sum_{i=1}^p \bar{d}_i \leq 0 \quad (5.62)$$

$$\Leftrightarrow K_{\max}(\sum_{i=\bar{N}+1}^N (\bar{d}_i - 1) - p) \leq -\sum_{i=1}^p \bar{d}_i \quad (5.63)$$

$$\Leftrightarrow K_{\max} \geq \frac{\sum_{i=1}^p \bar{d}_i}{p - \sum_{i=\bar{N}+1}^N (\bar{d}_i - 1)} \quad (5.64)$$

In this case,

$$\lambda_i^* = \min(\min(K_{\max} u^*, 1), \max(u^*, \frac{1}{\bar{d}_i})) \quad (5.65)$$

$$= \min(\min(1, 1), \max(\frac{1}{K_{\max}}, \frac{1}{\bar{d}_i})) \quad (5.66)$$

$$= \min(1, \max(\frac{1}{K_{\max}}, \frac{1}{\bar{d}_i})) \quad (5.67)$$

$$= \begin{cases} \min(1, \frac{1}{K_{\max}}) & \bar{d}_i \geq K_{\max} \\ \min(1, \frac{1}{\bar{d}_i}) & \bar{d}_i < K_{\max} \end{cases} \quad (5.68)$$

$$= \begin{cases} \frac{1}{K_{\max}} & \bar{d}_i \geq K_{\max} \\ \frac{1}{\bar{d}_i} & \bar{d}_i < K_{\max} \\ 1 & \bar{d}_i < 1 \end{cases} \quad (5.69)$$

Finally we obtain

$$\lambda^* = [\sigma^2 K_{\max}, \dots, \sigma^2 K_{\max}, d_{p+1}, \dots, d_{\bar{N}}, \sigma^2, \dots, \sigma^2], \quad (5.70)$$

where p and \bar{N} are the largest indices so that $d_p > \sigma^2 K_{\max}$ and $d_{\bar{N}} \geq \sigma^2$, respectively.

$$5. \ d_1 > \sigma^2 K_{\max} \text{ and } K_{\max} < \frac{\sum_{i=1}^p \bar{d}_i}{p - \sum_{i=\bar{N}+1}^N (\bar{d}_i - 1)}$$

In this case, since $\frac{1}{d_1} < u^* < \frac{1}{K_{\max}}$,

$$\lambda_i^* = \min(\min(K_{\max} u^*, 1), \max(u^*, \frac{1}{\bar{d}_i})) \quad (5.71)$$

$$= \min(K_{\max} u^*, \max(u^*, \frac{1}{\bar{d}_i})) \quad (5.72)$$

$$= \begin{cases} \min(K_{\max} u^*, u^*) & \bar{d}_i \geq \frac{1}{u^*} \\ \min(K_{\max} u^*, \frac{1}{\bar{d}_i}) & \bar{d}_i < \frac{1}{u^*} \end{cases} \quad (5.73)$$

$$= \begin{cases} u^* & \bar{d}_i \geq \frac{1}{u^*} \\ \frac{1}{\bar{d}_i} & \frac{1}{K_{\max} u^*} \leq \bar{d}_i \leq \frac{1}{u^*} \\ K_{\max} u^* & \bar{d}_i < \frac{1}{K_{\max} u^*} \end{cases} \quad (5.74)$$

Therefore, we obtain

$$\lambda^* = [\frac{\sigma^2}{u^*}, \dots, \frac{\sigma^2}{u^*}, d_{p+1}, \dots, d_q, \frac{\sigma^2}{u^* K_{\max}}, \dots, \frac{\sigma^2}{u^* K_{\max}}] \quad (5.75)$$

where p and q are the largest indices so that $d_p > \frac{\sigma^2}{u}$ and $d_q > \frac{\sigma^2}{u K_{\max}}$, respectively. \square

From Lemma 6, for the first two cases that is $d_1 \leq \sigma^2 K_{\max}$, the estimator is either a scaled identity matrix or the FML. Therefore, there is no need to find an optimal condition number in these cases since the estimator is not a function of the condition number.

Now we investigate uniqueness of the optimal condition number as we have done in the case of only rank constraint for the last two cases where the optimal eigenvalues are functions of the condition number.

Lemma 7. *The LR value of the condition number ML estimator is a monotonically increasing function with respect to the condition number K_{\max} and there is only one unique $K_{\max_{EL}}$.*

Proof. 1. $d_1 \leq \sigma^2$

$$\hat{\mathbf{R}}_{\text{CN}} = \sigma^2 \mathbf{I} \quad (5.76)$$

In this case, $\hat{\mathbf{R}}_{\text{CN}}$ does not change, so $\text{LR}(\mathbf{K}_{\text{max}})$ is a constant.

2. $\sigma^2 \leq d_1 \leq \sigma^2 K_{\text{max}}$

$$\hat{\mathbf{R}}_{\text{CN}} = \hat{\mathbf{R}}_{\text{FML}} \quad (5.77)$$

In this case, $\hat{\mathbf{R}}_{\text{CN}}$ does not change, so $\text{LR}(\mathbf{K}_{\text{max}})$ is a constant.

3. $d_1 > \sigma^2 K_{\text{max}}$ and $K_{\text{max}} \geq \frac{\sum_{i=1}^p d_i}{c - \sum_{i=\bar{N}+1}^{\bar{N}} (d_i - 1)}$

$$\hat{\mathbf{R}}_{\text{CN}} = \mathbf{\Phi} \text{diag}(\boldsymbol{\lambda}^*) \mathbf{\Phi}^H \quad (5.78)$$

where

$$\boldsymbol{\lambda}^* = [\sigma^2 K_{\text{max}}, \dots, \sigma^2 K_{\text{max}}, d_{p+1}, \dots, d_{\bar{N}}, \sigma^2, \dots, \sigma^2], \quad (5.79)$$

p and \bar{N} are the largest indices so that $d_p > \sigma^2 K_{\text{max}}$ and $d_{\bar{N}} \geq \sigma^2$, respectively.

$$\begin{aligned} & \text{LR}(\mathbf{K}_{\text{max}}) \\ &= \frac{\prod_{i=1}^N \frac{d_i}{\lambda_i} e^N}{\exp(\sum_{i=1}^N \frac{d_i}{\lambda_i})} \\ &= \frac{\prod_{i=1}^p \frac{d_i}{\sigma^2 K_{\text{max}}} \cdot \prod_{i=p+1}^{\bar{N}} 1 \cdot \prod_{i=\bar{N}+1}^N \frac{d_i}{\sigma^2} \cdot e^N}{\exp(\sum_{i=1}^p \frac{d_i}{\sigma^2 K_{\text{max}}} + \sum_{i=p+1}^{\bar{N}} 1 + \sum_{i=\bar{N}+1}^N \frac{d_i}{\sigma^2})} \\ &= \frac{\prod_{i=1}^p \frac{d_i}{\sigma^2 K_{\text{max}}} \cdot \prod_{i=\bar{N}+1}^N \frac{d_i}{\sigma^2} \cdot e^N}{\exp(\sum_{i=1}^p \frac{d_i}{\sigma^2 K_{\text{max}}}) \cdot e^{\bar{N}-p} \cdot \exp(\sum_{i=\bar{N}+1}^N \frac{d_i}{\sigma^2})} \end{aligned}$$

(a) within the range where p remains same

$$\begin{aligned} & \text{LR}(\mathbf{K}_{\text{max}}) \\ &= \frac{\prod_{i=1}^p \frac{d_i}{\sigma^2 K_{\text{max}}} \cdot \prod_{i=\bar{N}+1}^N \frac{d_i}{\sigma^2} \cdot e^N}{\exp(\sum_{i=1}^p \frac{d_i}{\sigma^2 K_{\text{max}}}) \cdot e^{\bar{N}-p} \cdot \exp(\sum_{i=\bar{N}+1}^N \frac{d_i}{\sigma^2})} \end{aligned}$$

$$\begin{aligned}
&= c_1 \frac{\prod_{i=1}^p \frac{d_i}{\sigma^2 K_{\max}}}{\exp(\sum_{i=1}^p \frac{d_i}{\sigma^2 K_{\max}})} \\
&= c_1 \frac{\frac{1}{(\sigma^2 K_{\max})^p} \prod_{i=1}^p d_i}{\exp(\frac{1}{\sigma^2 K_{\max}} \sum_{i=1}^p d_i)} \\
&= c_1 \frac{\frac{1}{(\sigma^2 K_{\max})^p} \prod_{i=1}^p d_i}{(\exp(\sum_{i=1}^p d_i))^{\frac{1}{\sigma^2 K_{\max}}}} \\
&= c_2 \frac{(\frac{1}{K_{\max}})^p}{c_3^{\frac{1}{K_{\max}}}} \\
&= c_2 \frac{1}{(K_{\max})^p \cdot c_3^{\frac{1}{K_{\max}}}}
\end{aligned}$$

where $c_1 = \frac{\prod_{i=\bar{N}+1}^N \frac{d_i}{\sigma^2} \cdot e^N}{\exp(\bar{N}-p) \cdot \exp(\sum_{i=\bar{N}+1}^N \frac{d_i}{\sigma^2})}$, $c_2 = c_1 \frac{\prod_{i=1}^p d_i}{\sigma^{2p}}$, and $c_3 = \exp(\frac{1}{\sigma^2} \sum_{i=1}^p d_i)$. Now let's evaluate the first derivative of the denominator of Eq. (5.80).

$$\begin{aligned}
&((K_{\max})^p \cdot c_3^{\frac{1}{K_{\max}}})' \\
&= p(K_{\max})^{p-1} c_3^{\frac{1}{K_{\max}}} + (K_{\max})^p \frac{c_3^{\frac{1}{K_{\max}}} \log c_3}{-(K_{\max})^2} \\
&= p(K_{\max})^{p-1} c_3^{\frac{1}{K_{\max}}} - (K_{\max})^{p-2} c_3^{\frac{1}{K_{\max}}} \log c_3 \\
&= (K_{\max})^{p-2} c_3^{\frac{1}{K_{\max}}} (p K_{\max} - \log c_3) \\
&= (K_{\max})^{p-2} c_3^{\frac{1}{K_{\max}}} (p K_{\max} - \frac{1}{\sigma^2} \sum_{i=1}^p d_i)
\end{aligned}$$

Since $d_1 > d_2 > \dots > d_p > \sigma^2 K_{\max}$,

$$p K_{\max} - \frac{1}{\sigma^2} \sum_{i=1}^p d_i < 0 \quad (5.80)$$

This implies the denominator of Eq. (5.80) is a decreasing function, and therefore, $LR(K_{\max})$ is an increasing function with respect to K_{\max} .

(b) $p \rightarrow p+1$ as K_{\max} decreases

The $LR(K_{\max})$ is a continuous function since $\lambda_{p+1} = d_{p+1}$ at the moment that $\sigma^2 K_{\max} = d_{p+1}$ and there is no discontinuity of λ_i . Therefore, $LR(K_{\max})$ is an increasing function in this case.

$$4. d_1 > \sigma^2 K_{\max} \text{ and } K_{\max} < \frac{\sum_{i=1}^c d_i}{c - \sum_{i=\bar{N}+1}^N (d_i - 1)}$$

$$\hat{\mathbf{R}}_{\text{CN}} = \mathbf{\Phi} \text{diag}(\boldsymbol{\lambda}^*) \mathbf{\Phi}^H \quad (5.81)$$

where

$$\boldsymbol{\lambda}^* = \left[\frac{\sigma^2}{u}, \dots, \frac{\sigma^2}{u}, d_{p+1}, \dots, d_q, \frac{\sigma^2}{uK_{\max}}, \dots, \frac{\sigma^2}{uK_{\max}} \right] \quad (5.82)$$

p , q , and \bar{N} are the vector of the eigenvalues of the estimate, the largest indices so that $d_p > \frac{\sigma^2}{u}$, $d_q > \frac{\sigma^2}{uK_{\max}}$, and $d_{\bar{N}} \geq \sigma^2$, respectively.

Before we prove the increasing property of $\text{LR}(K_{\max})$, we show u decreases as K_{\max} increases. u is the optimal solution of the optimization problem. In this case, u^* , the optimal solution of the optimization problem (2.27) is obtained by making the first derivative of the cost function 0. Let u_1 and u_2 be the optimal solutions for $K_{\max 1}$ and $K_{\max 2}$, respectively. Then, $\sum_{i=1}^N G'_i(u_1) = 0$ for $K_{\max 1}$. Since $\frac{1}{d_i} \leq u_1 \leq \frac{1}{K_{\max 1}}$ in this case, for $K_{\max 2} < K_{\max 1}$, the value of $G'_i(u_1)$ decreases for $d_i \leq 1$. $G'_i(u)$ also decreases for $d_i > 1$ and $u \leq \frac{1}{K_{\max} d_i}$ and remain same for $d_i > 1$ and $\frac{1}{K_{\max} d_i} < u$. Therefore, $\sum_{i=1}^N G'_i(u_1) < 0$ for $K_{\max 2}$. Finally, since $\sum_{i=1}^N G'_i(u_2)$ must be zero for $K_{\max 2}$, it is obvious that $u_1 < u_2$. This shows that u decreases as K_{\max} increases.

Now we show the increasing property of $\text{LR}(K_{\max})$.

(a) within the range where p and q remain same

In this case, We show $\text{LR}(u)$ is a decreasing function of u and an increasing function of K_{\max} for each of u and K_{\max} .

i. Proof of $\text{LR}(u)$ is a decreasing function.

$$\begin{aligned} \text{LR}(u) &= \frac{\prod_{i=1}^p \frac{ud_i}{\sigma^2} \cdot \prod_{i=q+1}^{\bar{N}} \frac{K_{\max} ud_i}{\sigma^2} \cdot e^N}{\exp(\sum_{i=1}^p \frac{ud_i}{\sigma^2} + \sum_{i=p+1}^q 1 + \sum_{i=q+1}^N \frac{K_{\max} ud_i}{\sigma^2})} \\ &= \frac{u^p \prod_{i=1}^p \frac{d_i}{\sigma^2} \cdot u^{N-q} \prod_{i=q+1}^{\bar{N}} \frac{K_{\max} d_i}{\sigma^2} \cdot e^N}{\exp(u(\sum_{i=1}^p \frac{d_i}{\sigma^2} + \sum_{i=q+1}^N \frac{K_{\max} d_i}{\sigma^2}) + q - p)} \\ &= \frac{c_1 u^{N-q+p}}{\exp(c_2 u + c_3)} \\ &= c_4 \frac{u^{N-q+p}}{c_5^u} \end{aligned}$$

where $c_1 = \prod_{i=1}^p \frac{d_i}{\sigma^2} \cdot \prod_{i=q+1}^{\bar{N}} \frac{K_{\max} d_i}{\sigma^2} \cdot e^N$, $c_2 = \sum_{i=1}^p \frac{d_i}{\sigma^2} + \sum_{i=q+1}^N \frac{K_{\max} d_i}{\sigma^2}$, $c_3 = q - p$, $c_4 = \frac{c_1}{e^{c_3}}$, and $c_5 = e^{c_2}$. The first derivative of Eq. (5.83) is obtained by

$$\begin{aligned}
\text{LR}'(u) &= (N - q + p)u^{N-q+p-1}c_5^{-u} \\
&\quad - u^{N-q+p} \log c_5 \cdot c_5^{-u} \\
&= u^{N-q+p-1}c_5^{-u}(N - q + p - u \log c_5) \\
&= u^{N-q+p-1}c_5^{-u}(N - q + p - c_2 u) \\
&= u^{N-q+p-1}c_5^{-u}(N - q + p \\
&\quad - u(\sum_{i=1}^p \frac{d_i}{\sigma^2} + \sum_{i=q+1}^N \frac{K_{\max} d_i}{\sigma^2}))
\end{aligned}$$

Since $\frac{\sigma^2}{u} \leq d_p$,

$$\begin{aligned}
&N - q + p - u(\sum_{i=1}^p \frac{d_i}{\sigma^2} + \sum_{i=q+1}^N \frac{K_{\max} d_i}{\sigma^2}) \\
&\leq N - q + p - u(\frac{p}{u} + \frac{N - q}{u} \cdot K_{\max}) \\
&= N - q - K_{\max}(N - q)
\end{aligned}$$

Since $K_{\max} > 1$, $\text{LR}'(u) < 0$ which implies $\text{LR}(u)$ is a decreasing function with respect to u .

ii. Proof of $\text{LR}(K_{\max})$ is an increasing function.

$$\begin{aligned}
&\text{LR}(K_{\max}) \\
&= \frac{\prod_{i=1}^p \frac{ud_i}{\sigma^2} \cdot \prod_{i=q+1}^{\bar{N}} \frac{K_{\max} ud_i}{\sigma^2} \cdot e^N}{\exp(\sum_{i=1}^p \frac{ud_i}{\sigma^2} + \sum_{i=p+1}^q 1 \\
&\quad + \sum_{i=q+1}^N \frac{K_{\max} ud_i}{\sigma^2})} \\
&= \frac{c_1 K_{\max}^{N-q}}{\exp(c_2 K_{\max} + c_3)} \\
&= c_4 \frac{K_{\max}^{N-q}}{c_5^{K_{\max}}}
\end{aligned}$$

where $c_1 = \prod_{i=1}^p \frac{ud_i}{\sigma^2} \cdot \prod_{i=q+1}^{\bar{N}} \frac{ud_i}{\sigma^2} \cdot e^N$, $c_2 = \sum_{i=q+1}^N \frac{ud_i}{\sigma^2}$, $c_3 = \sum_{i=1}^p \frac{ud_i}{\sigma^2} +$

$q - p$, $c_4 = \frac{c_1}{e^{c_3}}$, and $c_5 = e^{c_2}$. The first derivative is

$$\begin{aligned} \text{LR}'(K_{\max}) &= (N - q) K_{\max}^{N-q-1} c_5^{-K_{\max}} \\ &\quad - K_{\max}^{N-q} \log c_5 \cdot c_5^{-K_{\max}} \end{aligned} \quad (5.83)$$

$$\begin{aligned} &= K_{\max}^{N-q-1} \\ &\quad \times c_5^{-K_{\max}} (N - q - K_{\max} \log c_5) \end{aligned} \quad (5.84)$$

$$\begin{aligned} &= K_{\max}^{N-q+p-1} \\ &\quad \times c_5^{-u} (N - q - c_2 K_{\max}) \end{aligned} \quad (5.85)$$

$$\begin{aligned} &= K_{\max}^{N-q+p-1} \\ &\quad \times c_5^{-u} (N - q - K_{\max} \sum_{i=q+1}^N \frac{ud_i}{\sigma^2}) \end{aligned} \quad (5.86)$$

Since $\frac{\sigma^2}{u K_{\max}} \leq d_{q+1}$,

$$\begin{aligned} N - q - K_{\max} \sum_{i=q+1}^N \frac{ud_i}{\sigma^2} \\ \geq N - q - K_{\max} \left(\frac{N - q}{K_{\max}} \right) = 0 \end{aligned} \quad (5.87)$$

Therefore, $\text{LR}'(K_{\max}) \geq 0$ and $\text{LR}(K_{\max})$ is an increasing function with respect to K_{\max} .

These two proofs show that $\text{LR}(u, K_{\max})$ is an increasing function with respect to K_{\max} .

(b) p and q changes as K_{\max} decreases

The $\text{LR}(u, K_{\max})$ is a continuous function, and therefore, $\text{LR}(u, K_{\max})$ is an increasing function in this case.

□

Lemma 7 formally proves that there exist only one optimal condition number and therefore we can find the optimal condition number numerically. The algorithm of finding the optimal condition number is shown in Algorithm 3. We first set the initial condition number as the ML condition number obtained by [22]. Then we increase or decrease the condition number to the direction where the LR value decreases. Reducing the stepsize as the direction is reversed, we find the optimal condition number as precisely

Algorithm 3 The proposed algorithm to select condition number via EL

- 1: Obtain the ML solution of the condition number $K_{\max_{\text{ML}}}$ by the method in [22] and set the initial value of $K_{\max} = K_{\max_{\text{ML}}}$
 - 2: Set the initial step, $\Delta = K_{\max}/100$
 - 3: Evaluate $\text{LR}(K_{\max} - \Delta)$, $\text{LR}(K_{\max})$, $\text{LR}(K_{\max} + \Delta)$
 - if $|\text{LR}(K_{\max_{\text{ML}}} + \Delta) - \text{LR}_0| < |\text{LR}(K_{\max_{\text{ML}}}) - \text{LR}_0|$
 \rightarrow increase K_{\max} by Δ until it does not hold.
 \rightarrow then $\Delta = -\Delta/10$
 - elseif $|\text{LR}(K_{\max_{\text{ML}}} + \Delta) - \text{LR}_0| > |\text{LR}(K_{\max_{\text{ML}}}) - \text{LR}_0|$
 \rightarrow decrease K_{\max} by Δ until it does not hold.
 \rightarrow then $\Delta = -\Delta/10$
 - 4: Repeat Step 3 until $\Delta < 0.0001$.
-

as we want.

5.3 Experimental Validation

5.3.1 Experimental setup

In this section, we compare the proposed methods with alternative covariance estimation algorithms and parameter estimation algorithms. Two data sets are used in the experiments: 1) a radar covariance simulation model and 2) the KASSPER dataset [71].

First, we consider a radar system with an N -element uniform linear array for the simulation model. The overall covariance which is composed of jammer and additive white noise can be modeled by

$$\mathbf{R}(n, m) = \sum_{i=1}^J \sigma_i^2 \text{sinc}[0.5\beta_i(n-m)\phi_i] e^{j(n-m)\phi_i} + \sigma_a^2 \delta(n, m) \quad (5.88)$$

where $n, m \in \{1, \dots, N\}$, J is the number of jammers, σ_i^2 is the power associated with the i th jammer, ϕ_i is the jammer phase angle with respect to the antenna phase center, β_i is the fractional bandwidth, σ_a^2 is the actual power level of the white disturbance term, and $\delta(n, m)$ has the value of 1 only when $n = m$ and 0 otherwise. This simulation model has been widely and very successfully used in previous literature [20, 22, 44, 87] for performance analysis.

Data from the L-band data set of KASSPER program is the other data set used in our experiments. Note that the KASSPER data set exhibits two desirable characteris-

tics: 1) the low-rank structure of clutter and 2) the true covariance matrices for each range bin have been made available. These two characteristics facilitate comparisons via powerful figures of merit. The L-band data set consists of a data cube of 1000 range bins corresponding to the returns from a single coherent processing interval from 11 channels and 32 pulses. Therefore, the dimension of observations (or the spatio-temporal product) N is $11 \times 32 = 352$. Other key parameters are detailed in Table 3.1.

We measure the normalized signal to interference and noise ratio (SINR). The normalized SINR measure is commonly used in the radar literature and given by

$$\eta = \frac{|\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s}|^2}{|\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{R} \hat{\mathbf{R}}^{-1} \mathbf{s}| |\mathbf{s}^H \mathbf{R}^{-1} \mathbf{s}|} \quad (5.89)$$

where \mathbf{s} is the spatio-temporal steering vector, $\hat{\mathbf{R}}$ is the data-dependent estimate of \mathbf{R} , and \mathbf{R} is the true covariance matrix. It is easily seen that $0 < \eta < 1$ and $\eta = 1$ if and only if $\hat{\mathbf{R}} = \mathbf{R}$. The SINR is plotted in decibels in all our experiments, that is, $\text{SINR}(\text{dB}) = 10 \log_{10} \eta$. Therefore, $\text{SINR}(\text{dB}) \leq 0$. For the KASSPER data set, since the steering vector is a function of both azimuthal angle and Doppler frequency, we obtain plots as a function of one variable (azimuthal angle or Doppler) by marginalizing over the other variable. We evaluate and compare different covariance estimation techniques and parameter selection methods in the following three subsections:

- **Sample Covariance Matrix:** The sample covariance matrix is given by $\mathbf{S} = \frac{1}{K} \mathbf{Z} \mathbf{Z}^H$. It is well known that \mathbf{S} is the unconstrained ML estimator under Gaussian disturbance statistics. We refer to this as SML.
- **Fast Maximum Likelihood:** The fast maximum likelihood (FML) [20] uses the structural constraint of the covariance matrix. The FML method just involves the eigenvalue decomposition of the sample covariance and perturbing eigenvalues to conform to the structure. The FML also can be considered as the RCML estimator with the rank which is the greatest index i satisfying $\lambda_i > \sigma^2$ where λ_i 's are the eigenvalues of the sample covariance in descending order. Therefore, a rank can be considered as an output of the FML. The FML's success in radar STAP is widely known [23].
- **Rank Constrained ML Estimators:** The RCML estimator with the rank or the rank and the noise level obtained by the proposed methods using the expected likelihood approach. The rank is obtained by the EL approach in the case of the imperfect rank constraint and both of the rank and the noise level are obtained by

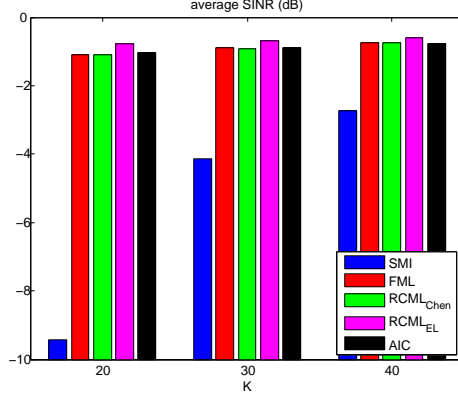


Figure 5.4. Normalized SINR in dB versus number of training samples K ($N = 20$) for the simulation model.

the EL approach in the case of imperfect rank and noise power constraints. We refer to these as RCML_{EL} .

- **Chen *et al.* Rank Selection Method:** Chen *et al.* [95] proposed a statistical procedure for detecting the multiplicity of the smallest eigenvalue of the structured covariance matrix using statistical selection theory. The rank can be estimated from their methods using pre-calculated parameters. We refer to this method as $\text{RCML}_{\text{Chen}}$.
- **AIC:** Akaike [90] proposed the information theoretic criteria for model selection. The AIC selects the model that best fits the data for given a set of observations and a family of models, that is, a parameterized family of probability densities. Wax and Kailath [54] proposed the method to determine the number of signals from the observed data based on the AIC. We compare Wax and Kailath's method.
- **Condition number constrained ML estimators:** The maximum likelihood estimation method of the covariance matrix with a condition number [22] proposed by Aubry *et al.* is considered for evaluating the performance with three different condition numbers. 1) CNCML : the condition number obtained by the proposed method in [22], 2) CNCML_{EL} : the condition number obtained by the expected likelihood approach, and 3) $\text{CNCML}_{\text{true}}$: the true condition number.

5.3.2 Imperfect rank constraint

First, we compare the rank estimation method proposed in Section 5.2.1 with alternative algorithms including SMI, FML, AIC, and Chen's algorithm. We plot the normalized

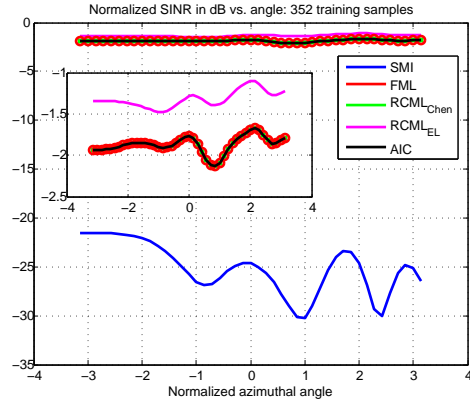
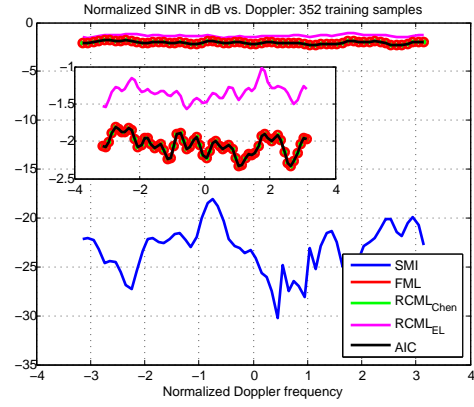
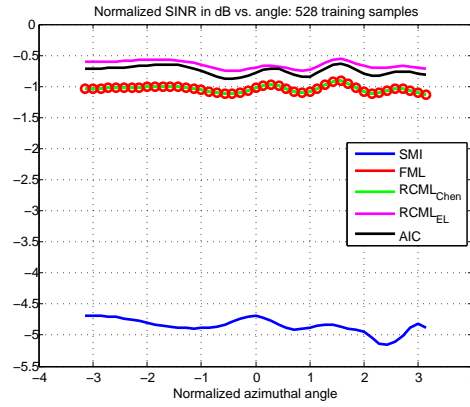
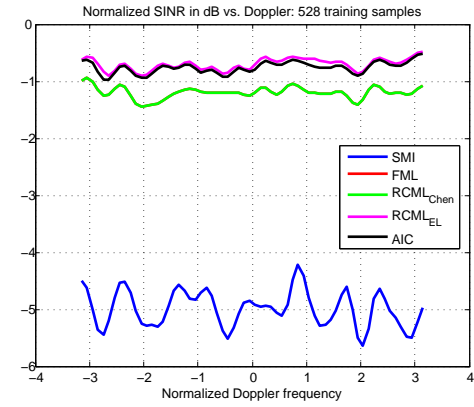
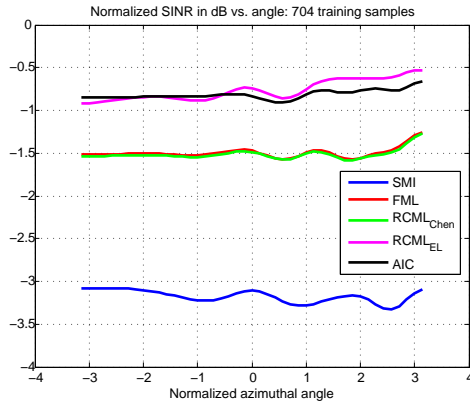
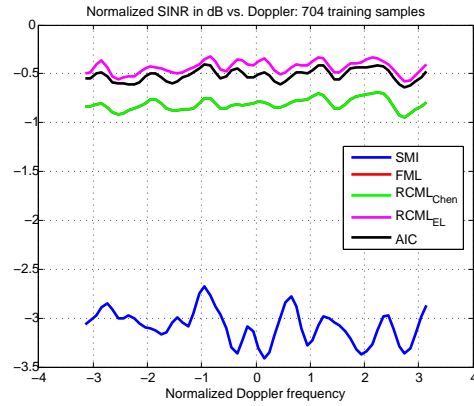
(a) $K = N = 352$ (b) $K = N = 352$ (c) $K = 1.5N = 528$ (d) $K = 1.5N = 528$ (e) $K = 2N = 704$ (f) $K = 2N = 704$

Figure 5.5. Normalized SINR versus azimuthal angle and Doppler frequency for the KASSPER data set.

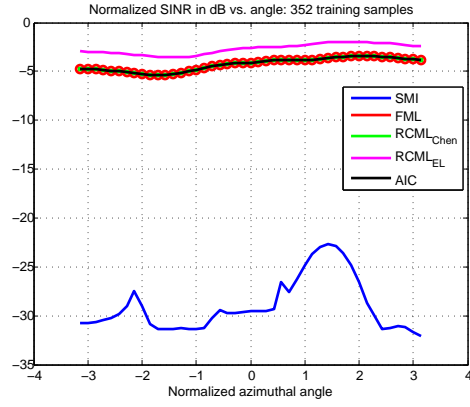
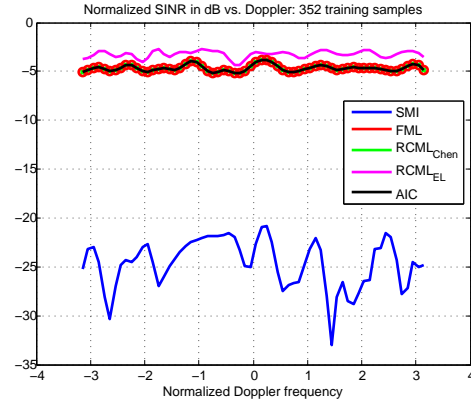
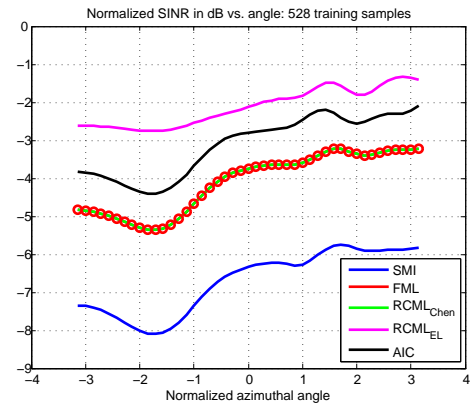
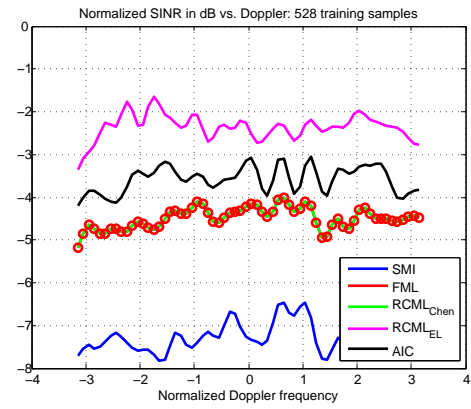
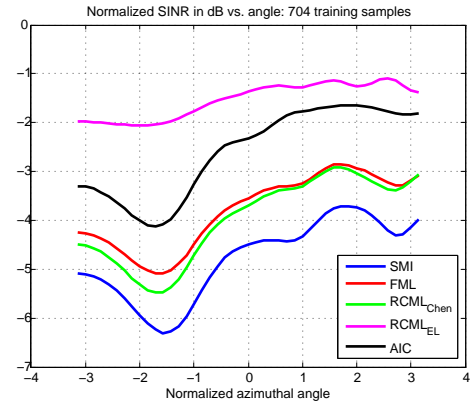
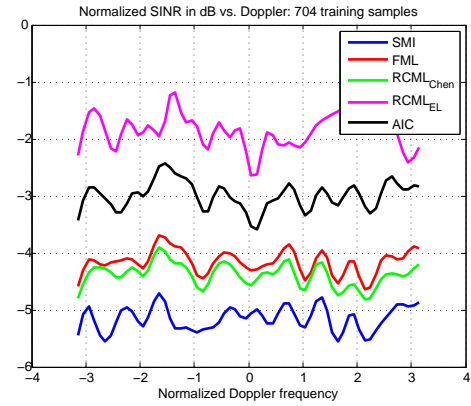
(a) $K = N = 352$ (b) $K = N = 352$ (c) $K = 1.5N = 528$ (d) $K = 1.5N = 528$ (e) $K = 2N = 704$ (f) $K = 2N = 704$

Figure 5.6. Normalized SINR versus azimuthal angle and Doppler frequency for the KASSPER data set.

SINR (in dB) versus the number of training samples, 20, 30, and 40 in Fig. 5.4 for the simulation model. The SINR values are obtained by averaging SINR values from 500 Monte Carlo trials. It is shown that the SINR values increases monotonically as K increases. The RCML_{EL} exhibits the best performance in all training regimes. Particularly, the difference between RCML_{EL} and other methods increases when training samples are limited.

Figure 5.5 shows the normalized SINR values for various number of training samples for the KASSPER data set. We plot the averaged SINR values in decibel over either azimuth angle or Doppler frequency domain. The left and right column show the results for angle and Doppler, respectively. Similarly to the results for the simulation model, RCML_{EL} outperforms all the other compared methods in all training regime. This implies that the rank obtained by the EL approach is more accurate and closer to the rank predicted by Brennan rule ($M + P - 1 = 42$) than any other methods.

Realistic case of contaminated observations: In practice, homogeneous training samples are hard to obtain and the received signals are often corrupted by target information. Therefore, it is meaningful to compare the performance for nonhomogeneous observation to investigate which algorithm indeed works well and is robust in practice. In this case, the received signal is given by

$$\mathbf{z} = \alpha \mathbf{s} + \mathbf{d} \quad (5.90)$$

where \mathbf{s} and \mathbf{d} are the deterministic steering vector and stochastic disturbance vector, respectively. Figure 5.6 shows the normalized SINR values when a half of the training samples contain \mathbf{s} with $\alpha = 50$. The gap between RCML_{EL} and the others is bigger than that in Figure 5.5. In particular, AIC shows compatible performance in homogeneous cases though, Figure 5.6(e) and Figure 5.6(f) show that the difference between AIC and RCML_{EL} is larger for non-homogeneous case. Therefore the results show remarkably that RCML_{EL} still excels under target contamination or heterogeneous training where other techniques face severe degradation in performance.

5.3.3 Imperfect rank and noise power constraints

In this section, we show experimental results for estimation of both a rank and a noise power via the expected likelihood approach, which is proposed in Section 5.2.2. In this case, we assume that both the rank and the noise power are unknown and to be estimated for both the simulation model and the KASSPER data set. Since the previous works such

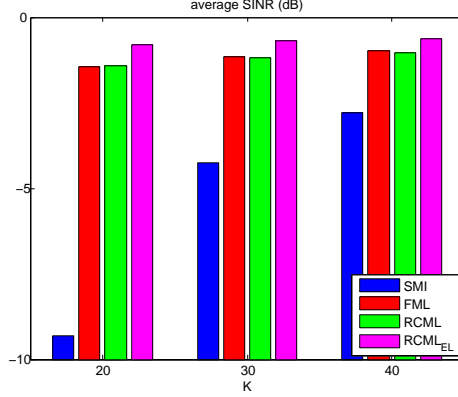


Figure 5.7. Normalized SINR in dB versus number of training samples K ($N = 20$) for the simulation model.

as AIC and Chen’s algorithm are for only estimating the rank and can not be extended to estimate both the rank and the noise power, we compare the proposed EL method with the sample covariance, FML, and the RCML estimator with a prior knowledge of the rank. For the RCML estimator, we employ the number of jammers ($r = 5$) and the Brennan rule ($r = 42$) as the clutter rank for the simulation model and the KASSPER data set, respectively. In addition, since the FML method requires a prior knowledge of the noise power, we calculate and use the maximum likelihood estimate of the noise power for a rank given by a prior knowledge for the FML.

Figure 5.7 shows the performance of various estimators in the sense of the normalized SINR values for the simulation model. Similarly to the case of only rank estimation, the proposed method show the best performance in all training regimes.

Figure 5.8 shows the performance of the methods in the normalized SINR for the KASSPER data set. The proposed method is comparable with or slightly better than the RCML estimator using the rank by Brennan rule. This means that the proposed method estimates both the rank and the noise power adaptively from training samples whereas the rank by Brennan rule is fixed regardless of the training samples.

5.3.4 Imperfect condition number constraint

Now we show experimental results for the condition number estimation method proposed in Section 5.2.3. We compare the proposed method, denoted by CNCML_{EL} , with four different covariance estimation methods, the sample covariance matrix (SMI), FML, CNCML , and $\text{CNCML}_{\text{true}}$.

Tables 5.1 and 5.2 show the normalized SINR values for the simulation model. We

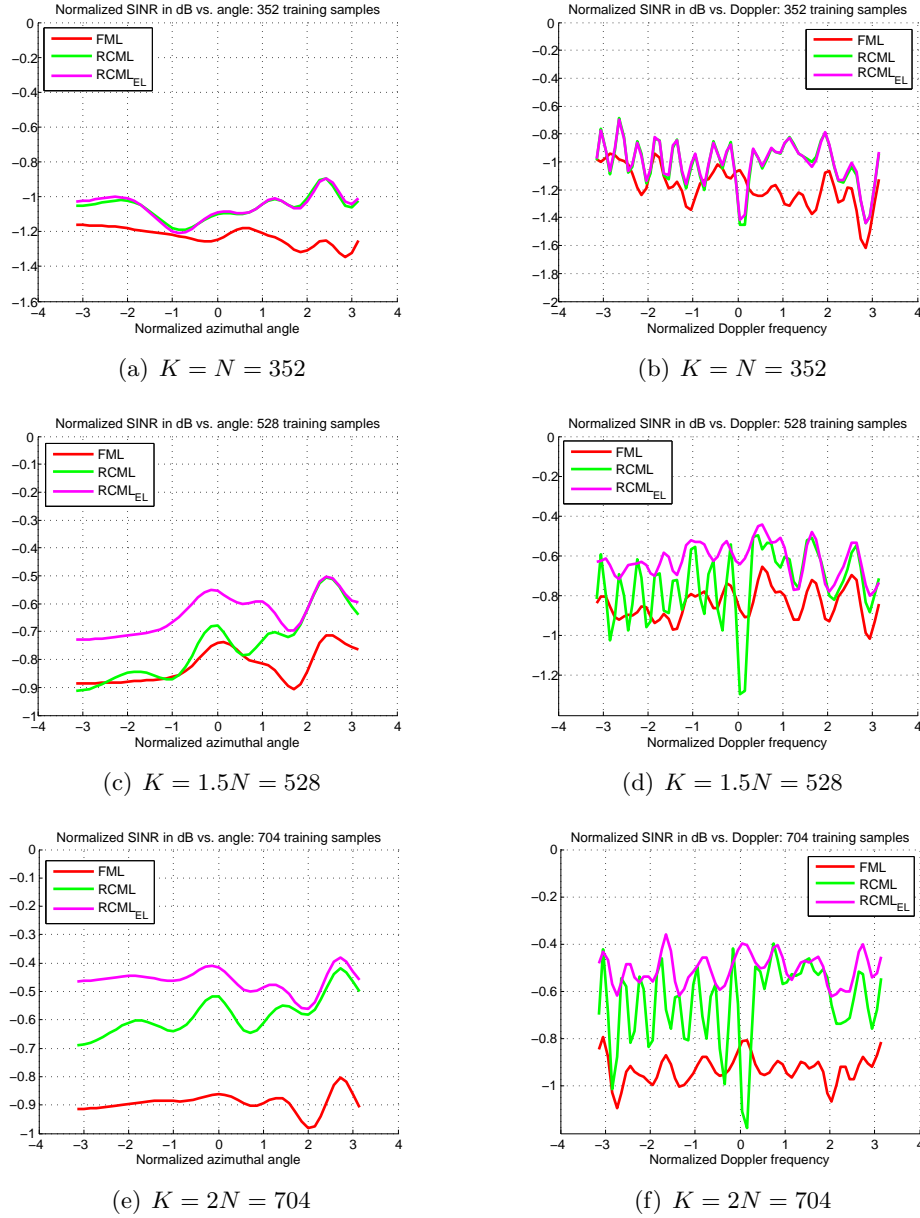


Figure 5.8. Normalized SINR versus azimuthal angle and Doppler frequency for the KASSPER data set.

analyze five different scenarios with different parameters of the simulated covariance model given by Eq. (5.88). We use the same parameters as those used in [22] to evaluate the performances and they are shown in Table 5.2(c).

For the narrowband scenarios ($B_f = 0$) in Table 5.1(a) and Table 5.1(c), CNCML_{EL} outperforms the alternatives for the limited training regime and FML is the best in other

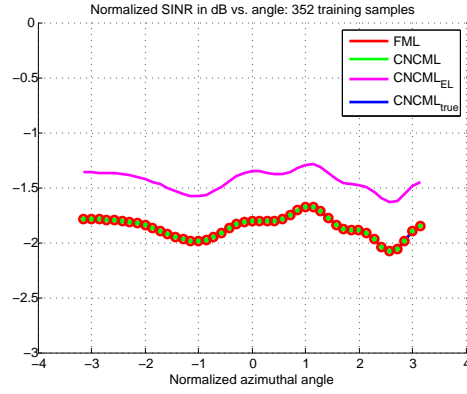
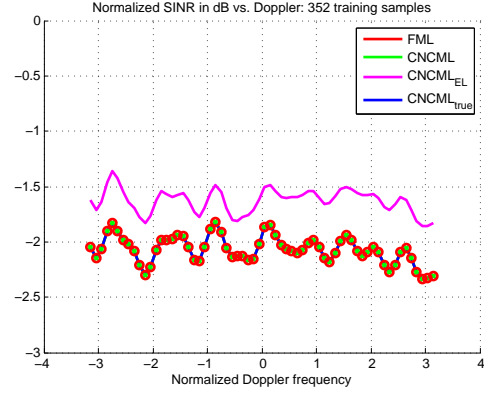
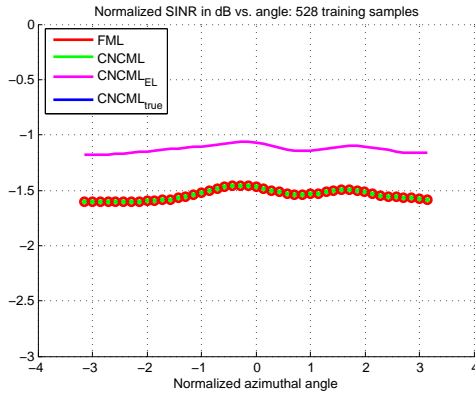
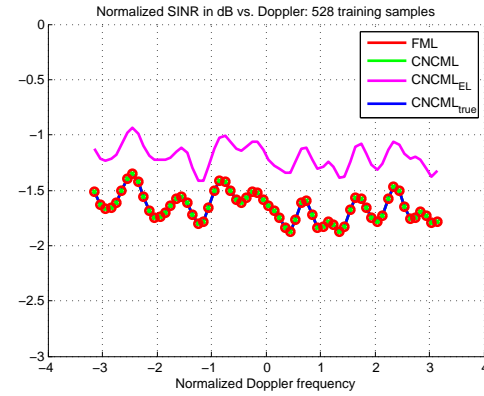
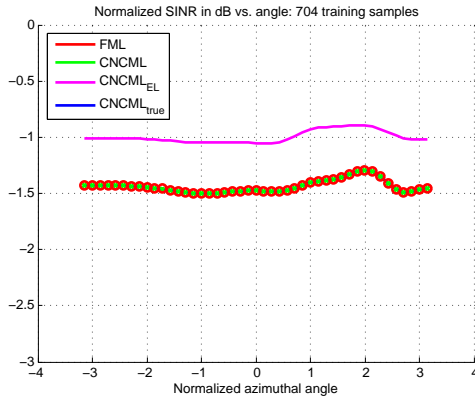
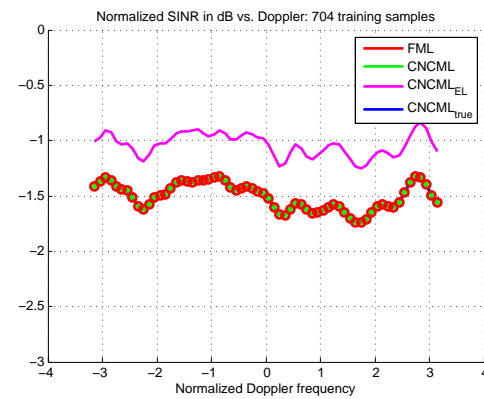
(a) $K = N = 352$ (b) $K = N = 352$ (c) $K = 1.5N = 528$ (d) $K = 1.5N = 528$ (e) $K = 2N = 704$ (f) $K = 2N = 704$

Figure 5.9. Normalized SINR versus azimuthal angle and Doppler frequency for the KASSPER data set. (a) and (b) for $K = N = 352$, (c) and (d) for $K = 1.5N = 528$, and (e) and (f) for $K = 2N = 704$

training regimes. Note that the gap between CNCML_{EL} and FML (at most 0.002) is much smaller than that of the limited training regime (at least 0.3).

On the other hand, for the wideband scenarios in Tables 5.1(b), 5.2(a), and 5.2(b), CNCML_{EL} shows the best performance in most cases including $\text{CNCML}_{\text{true}}$ using the true condition number.

The experimental results for the KASSPER data set are shown in Figure 5.9. We do not plot the sample covariance matrix to clarify the difference among the estimators. In every case, FML, CNCML, and $\text{CNCML}_{\text{true}}$ are very close to one another and CNCML_{EL} is the best estimator. Note that CNCML_{EL} is based on the same algorithm as CNCML and $\text{CNCML}_{\text{true}}$ and differs from them in the point that CNCML_{EL} uses a different condition number which is estimated by the expected likelihood approach. Again, this shows that the expected likelihood criterion is really useful and powerful to estimate parameters which is imperfectly known and leads to an adaptive and robust covariance estimator.

5.4 Conclusion

We propose robust covariance estimation algorithms which automatically determine the optimal values of practical constraints via the expected likelihood criterion for radar STAP in this chapter. Three different cases of practical constraints which is exploited in recent works including the rank constrained ML estimation and the condition number constrained ML estimation are investigated. Significant analytic results and proofs are derived for each case. Uniqueness of the optimal values of the rank constraint and the condition number constraint is formally proved and a closed form solution of the noise level is derived. Experimental results show that the estimators with the constraints obtained by the expected likelihood outperform previous works which uses constraints obtained by other parameter estimation methods including the maximum likelihood solution of the constraints.

(d)

σ^2	K	SMI	FML	CNCML	CNCML _{EL}	CNCML _{true}
-5	20	-9.3785	-0.5195	-0.5212	-0.4822	-0.5200
	30	-4.2579	-0.4242	-0.4257	-0.4256	-0.4250
	40	-2.7424	-0.3460	-0.3476	-0.3476	-0.3468
0	20	-9.3196	-0.5511	-0.5521	-0.5141	-0.5515
	30	-4.2276	-0.4202	-0.4221	-0.4220	-0.4210
	40	-2.7649	-0.3513	-0.3530	-0.3528	-0.3521
5	20	-9.0922	-0.5269	-0.5279	-0.4875	-0.5272
	30	-4.2172	-0.4348	-0.4364	-0.4362	-0.4357
	40	-2.7300	-0.3484	-0.3503	-0.3505	-0.3493
10	20	-9.3511	-0.5355	-0.5305	-0.4998	-0.5360
	30	-4.1955	-0.4164	-0.4180	-0.4175	-0.4175
	40	-2.7491	-0.3501	-0.3515	-0.3518	-0.3509

(e)

σ^2	K	SMI	FML	CNCML	CNCML _{EL}	CNCML _{true}
-5	20	-9.3069	-1.7371	-1.7322	-1.7358	-1.7350
	30	-4.1795	-1.2399	-1.2388	-1.2347	-1.2397
	40	-2.7535	-0.9496	-0.9492	-0.9456	-0.9493
0	20	-9.1354	-1.6944	-1.6928	-1.7027	-1.6940
	30	-4.2345	-1.2986	-1.2987	-1.2955	-1.2990
	40	-2.7545	-1.0041	-1.0043	-1.0023	-1.0046
5	20	-9.2524	-1.3976	-1.4016	-1.3244	-1.4000
	30	-4.2309	-1.0737	-1.0784	-1.0666	-1.0762
	40	-2.7523	-0.8848	-0.8876	-0.8818	-0.8866
10	20	-9.3660	-1.2567	-1.2569	-1.2115	-1.2570
	30	-4.3013	-0.9526	-0.9545	-0.9450	-0.9537
	40	-2.7350	-0.7171	-0.7197	-0.7139	-0.7186

(f)

σ^2	K	SMI	FML	CNCML	CNCML _{EL}	CNCML _{true}
-5	20	-9.3702	-0.5340	-0.5349	-0.4925	-0.5340
	30	-4.2791	-0.4302	-0.4316	-0.4315	-0.4308
	40	-2.7856	-0.3493	-0.3510	-0.3509	-0.3501
0	20	-9.2898	-0.5485	-0.5501	-0.5104	-0.5491
	30	-4.2648	-0.4202	-0.4219	-0.4220	-0.4209
	40	-2.7274	-0.3604	-0.3621	-0.3621	-0.3611
5	20	-9.0582	-0.5318	-0.5328	-0.4899	-0.5322
	30	-4.1548	-0.4142	-0.4155	-0.4152	-0.4149
	40	-2.7655	-0.3515	-0.3531	-0.3533	-0.3521
10	20	-9.3632	-0.5352	-0.5363	-0.4974	-0.5360
	30	-4.2728	-0.4328	-0.4348	-0.4349	-0.4337
	40	-2.7577	-0.3538	-0.3554	-0.3547	-0.3547

Table 5.1. Normalized SINR for various values of parameters for the simulation model.

(a)

σ^2	K	SMI	FML	CNCML	CNCML _{EL}	CNCML _{true}
-5	20	-9.0316	-1.7161	-1.7131	-1.7634	-1.7150
	30	-4.1465	-1.1704	-1.1691	-1.1659	-1.1693
	40	-2.7727	-0.9390	-0.9384	-0.9351	-0.9387
0	20	-9.2091	-1.6706	-1.6701	-1.6674	-1.6706
	30	-4.2004	-1.2681	-1.2682	-1.2633	-1.2682
	40	-2.7423	-1.0102	-1.0117	-1.1009	-1.0106
5	20	-9.3538	-1.3980	-1.4028	-1.3216	-1.4004
	30	-4.2203	-1.0869	-1.0910	-1.0785	-1.0889
	40	-2.7079	-0.8694	-0.8721	-0.8666	-0.8713
10	20	-9.221	-1.2446	-1.2455	-1.1982	-1.2452
	30	-4.2116	-0.9428	-0.9460	-0.9382	-0.9444
	40	-2.7563	-0.7235	-0.7264	-0.7226	-0.7253

(b)

σ^2	K	SMI	FML	CNCML	CNCML _{EL}	CNCML _{true}
-5	20	-9.2679	-1.1593	-1.1616	-1.1150	-1.1610
	30	-4.2234	-0.9262	-0.9286	-0.9242	-0.9278
	40	-2.8271	-0.7705	-0.7729	-0.7712	-0.7723
0	20	-9.2934	-0.9052	-0.9051	-0.8422	-0.9059
	30	-4.1617	-0.6909	-0.6920	-0.6862	-0.6924
	40	-2.7387	-0.5711	-0.5724	-0.5676	-0.5725
5	20	-9.4154	-0.8398	-0.8334	-0.7909	-0.8399
	30	-4.2284	-0.6273	-0.6231	-0.6070	-0.6278
	40	-2.7208	-0.5034	-0.5022	-0.4945	-0.5046
10	20	-9.1447	-0.7388	-0.7225	-0.6815	-0.7392
	30	-4.2046	-0.5931	-0.5803	-0.5535	-0.5931
	40	-2.7241	-0.4821	-0.4738	-0.4576	-0.4827

(c)

	J	σ_J^2	ϕ	B_f
(a)	1	30	20°	0
(b)	1	30	20°	0.3
(c)	3	[30 30 30]	[20° 40° 60°]	[0 0 0]
(d)	3	[30 30 30]	[20° 40° 60°]	[0.3 0.3 0.3]
(e)	3	[10 20 30]	[20° 40° 60°]	[0.2 0 0.3]

Table 5.2. Normalized SINR for various values of parameters for the simulation model.

Conclusion

6.1 Summary of Contributions

This dissertation developed robust covariance estimation methods for radar STAP. Here, I briefly summarize the contributions of this dissertation.

Covariance matrix estimation under practical constraints: In Chapter 3 and 4, we focused on maximum likelihood estimation methods of structured covariance estimation exploiting practical constraints for radar STAP. It is shown that introducing practical constraints into covariance estimation problems leads to powerful estimators in the radar literatures. In particular, the rank of the clutter subspace is really powerful constraint because the rank can be either calculated by the well-known Brennan rule under ideal condition or estimated by robust rank estimation methods. We develop the RCML estimator which exploits the structural constraint and the rank of clutter subspace. Though the rank is a challenging constraint, we reduce the estimation problem to a convex optimization problem and derive a closed form solution. We also derive an estimator for when the noise level is unknown but only a lower bound is available.

Chapter 4 focuses on exploiting a Toeplitz structure as well as the rank constraint on the structured interference for radar STAP. The problem is inherently hard because there is no closed form solution for ML estimation of Toeplitz covariance matrices for all training regimes. We develop a new estimator that is based on a cascade of two closed forms. The first closed form is the RCML estimator and the second step of the Toeplitz approximation performs constrained optimization of eigenvalues to either exactly or approximately satisfy the Toeplitz constraint with preserving the rank.

Robust covariance estimation under imperfect constraints: In Chapter 5,

we develop robust covariance estimation algorithms under imperfect constraints. The proposed algorithms automatically determine the optimal values of practical constraints which are imperfectly or approximately known via the expected likelihood approach. We consider three different cases of practical constraints, 1) the rank constraint, 2) the rank constraint and the noise level, and 3) the condition number constraint, which are exploited in recent works including the RCML estimator and the condition number ML estimator. We develop significant analytical results and derive formal proof for each case of the constraints.

6.2 Future Research

6.2.1 Newer Constraint: Sparsity or Block Sparsity

The principal merit of recent radar STAP estimator including the methods proposed in this dissertation is that they employ constrained optimization theory to yield tractable solutions. We can take this theme to fruition by investigating newer increasingly relevant constraints/scenarios that represent covariance estimation in modern radar STAP. The recently proposed [96] graphical lasso technique is an L_1 regularized maximum likelihood estimate, such that

$$\mathbf{R} = \arg \max_{\mathbf{R} \in \psi} \left\{ \log (p(\mathbf{Z}|\mathbf{R})) - \rho \|\mathbf{R}^{-1}\|_1 \right\} \quad (6.1)$$

Here, ψ denotes the set of covariance matrices of interest, i.e., for our problem the set of positive definite matrices which can be expressed as a sum of an identity and a rank-deficient positive semi-definite (the clutter) matrix. Note the cost function consists of a standard log likelihood term plus a regularization term. The regularization term $\rho \|\mathbf{R}^{-1}\|$ encourages the inverse of the covariance matrix to be sparse. That is, if the ij th component of \mathbf{R}^{-1} is zero, then clutter variables i and j are conditionally independent given the other variables. The sparsity constraint may often capture real-world clutter scenarios. In fact, physical effects such as locally pronounced clutter correlation over the spatio-temporal range may in fact give rise to “block sparsity.”

6.2.2 Non-Homogeneity Detection

The presence of outliers in the training data for radar STAP causes target cancellation resulting in degraded output signal to interference ratio and perforce degraded detection performance. A common signal processing method in this context is to excise

outliers from the training data and use the resulting outlier free training data for covariance matrix estimation. Several algorithms for outlier removal have been proposed recently [23, 73, 74, 97, 98] in a variety of dense target environments. We take the perspective of observing interference or clutter specific training data as normal and outlier contamination as an anomaly. In machine learning, anomaly detection is an extensively studied area [99]. Anomaly detection refers to the problem of finding patterns in data that do not conform to expected behavior. By treating outlier (target) contamination as an anomaly we can benefit from machine learning techniques devoted to identifying which data samples are anomalous and subsequently use the remaining normal data for estimation purposes. We can particularly focus on methods that can identify anomalies even as a small number of observations is presented because training samples are not abundant in practice. We can investigate and apply two categories of anomaly detection techniques for the radar non-homogeneity detection (NHD) problem, 1.) supervised learning based anomaly detection [100–102] and 2.) unsupervised learning or clustering based anomaly detection [103, 104].

6.2.3 Expansion of the EL Approach

As shown in this dissertation, the expected likelihood approach is powerful to estimate regularization parameters employed as constraints in covariance estimation problems. Unsurprisingly, other constraints which have been commonly exploited in previous covariance estimation methods in radar literature can be also selected or estimated using the expected likelihood approach. For example, the condition number constraint and the noise level can be jointly estimated and the intensity parameter in the shrinkage estimator also can be estimated by the expected likelihood approach. Furthermore, we investigate experimental results for only $K \geq N$, i.e., the number of training samples are greater than or equal to the data dimension since the complex Wishart distribution is defined only for $K \geq N$. Analyzing theoretical properties of the likelihood ratio values for true covariance and focusing on how to apply the expected likelihood approach to the case of $K < N$ can be another future work. Finally, using the expected likelihood approach not just for tuning parameters but for direct estimation of disturbance covariance can be investigated in the future. For example, the EL approach can be applicable to estimate a whole set of eigenvalues and eigenvectors of the covariance matrix.

Bibliography

- [1] R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays*. New York: Wiley, 1980.
- [2] J. R. Guerci, *Space-Time Adaptive Processing for Radar*. Artech House Publishers, July 2003.
- [3] R. Klemm, *Principles of space-time adaptive processing*. IEE Publishing, April 2002.
- [4] B. Himed and W. L. Melvin, "Analyzing Space-Time Adaptive Processors Using Measured Data," in *Conference Record of The Thirty-First Asilomar Conference on Signals, Systems and Computers*, vol. 1, November 1997, pp. 930–935.
- [5] J. R. Guerci and E. J. Baranoski, "Knowledge-Aided Adaptive Radar at DARPA: An overview," *IEEE Signal Processing Mag.*, vol. 23, no. 1, pp. 41–50, January 2006.
- [6] M. C. Wicks, M. Rangaswamy, R. Adve, and T. B. Hale, "Space-Time Adaptive Processing: A knowledge-based perspective for airborne radar," *IEEE Signal Processing Mag.*, vol. 23, no. 1, pp. 51–65, January 2006.
- [7] H. Wang and L. Cai, "A Localized Adaptive MTD Processor," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 27, no. 3, pp. 532–539, May 1991.
- [8] F. Gini and M. Rangaswamy, *Knowledge Based Radar Detection, Tracking and Classification*. Wiley-Interscience publication, 2008.
- [9] R. Nitzberg, "Application of Maximum Likelihood Estimation of Persymmetric Covariance Matrices to Adaptive Processing," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-16, no. 1, pp. 124–127, January 1980.
- [10] J. P. Burg, D. G. Luenberger, and D. L. Wenger, "Estimation of Structured Covariance Matrices," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 963–974, September 1982.

- [11] H. Li, P. Stoica, and J. Li, "Computationally Efficient Maximum Likelihood Estimation of Structured Covariance Matrices," *IEEE Trans. Signal Processing*, vol. 47, no. 5, pp. 1314–1323, May 1999.
- [12] S. S. Al-Homidan, "SQP Algorithms for Solving Toeplitz Matrix Approximation Problem," *Numerical Linear Algebra with Applications*, vol. 9, no. 8, pp. 619–627, 2002.
- [13] Y. I. Abramovich, D. A. Gray, A. Y. Gorokhov, and N. K. Spencer, "Positive-Definite Toeplitz Completion in DOA Estimation for Nonuniform Linear Antenna Arrays-Part I: Fully Augmentable Arrays," *IEEE Trans. Signal Processing*, vol. 46, no. 9, pp. 2458–2471, September 1998.
- [14] D. M. Wilkes and M. H. Hayes, "Iterated Toeplitz Approximation of Covariance Matrices," in *Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on*, vol. 3, April 1988, pp. 1663–1666.
- [15] E. Conte, M. Lops, and G. Ricci, "Adaptive Detection Schemes in Compound-Gaussian Clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 4, pp. 1058–1069, October 1998.
- [16] J. P. Hoffbeck and D. A. Landgrebe, "Covariance Matrix Estimation and Classification with Limited Training Data," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, no. 7, pp. 763–767, July 1996.
- [17] Y. I. Abramovich, "A controlled method for adaptive optimization of filters using the criterion of maximum SNR," *Radio Engineering and Electronic Physics*, vol. 3, pp. 87–95, 1981.
- [18] B. D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 397–401, July 1988.
- [19] A. Haimovich, "The Eigencanceler: Adaptive Radar by Eigenanalysis Methods," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 2, pp. 532–542, April 1996.
- [20] M. Steiner and K. Gerlach, "Fast Converging Adaptive Processors for a Structured Covariance Matrix," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 4, pp. 1115–1126, October 2000.
- [21] Y. Chen, A. Wiesel, Y. C. Eldar, and A. O. Hero, "Shrinkage Algorithms for MMSE Covariance Estimation," *IEEE Trans. Signal Processing*, vol. 58, no. 10, pp. 5016–5029, October 2010.
- [22] A. Aubry, A. D. Maio, L. Pallotta, and A. Farina, "Maximum Likelihood Estimation of a Structured Covariance Matrix with a Condition Number Constraint," *IEEE Trans. Signal Processing*, vol. 60, no. 6, pp. 3004–3021, June 2012.

- [23] M. Rangaswamy, F. C. Lin, and K. R. Gerlach, "Robust adaptive signal processing methods for heterogeneous radar clutter scenarios," *Signal Processing*, vol. 84, no. 9, pp. 1653–1665, September 2004.
- [24] J. Ward, "Space-Time Adaptive Processing for Airborne Radar," MIT Lincoln Laboratory, Tech. Rep., December 1994.
- [25] M. I. Miller and D. L. Snyder, "The Role of Likelihood and Entropy in Incomplete-Data Problems: Application to Estimating Point-Process Intensities and Toeplitz Constrained Covariances," *Proceedings of the IEEE*, vol. 75, no. 7, pp. 892–907, July 1987.
- [26] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, 2nd ed. A John Wiley and Sons, 2002.
- [27] P. Forster and G. Vezzosi, "Optimal Toeplitzification with a Given Rank Constraint," in *Acoustics, Speech, and Signal Processing, 1989. ICASSP-89., 1989 International Conference on*, vol. 4, May 1989, pp. 2783–2786.
- [28] M. Jansson and B. Ottersten, "Structured Covariance Matrix Estimation: A Parametric Approach," in *Acoustics, Speech, and Signal Processing, 2000. ICASSP '00., 2000 IEEE International Conference on*, vol. 5, June 2000, pp. 3172–3175.
- [29] B. Kang, V. Monga, and M. Rangaswamy, "Computationally Efficient Toeplitz Approximation of Structured Covariance under a Rank Constraint," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 1, pp. 775–785, 2015.
- [30] Y. I. Abramovich, N. K. Spencer, and A. Y. Gorokhov, "Modified GLRT and AMF Framework for Adaptive Detectors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 3, pp. 1017–1051, July 2007.
- [31] S. Boyd and L. Vandenberghe, *Convex Optimization*, 2nd ed. Cambridge University Press, 2004.
- [32] V. Monga and M. Rangaswamy, "Rank Constrained ML Estimation of Structured Covariance Matrices with Applications in Radar Target Detection," in *IEEE Radar Conference*, May 2012, pp. 475–480.
- [33] B. Kang, V. Monga, and M. Rangaswamy, "On the Practical Merits of Rank constrained ML Estimator of Structured Covariance Matrices," in *IEEE Radar Conference*, May 2013, pp. 1–6.
- [34] —, "Rank-Constrained Maximum Likelihood Estimation of Structured Covariance Matrices," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 1, pp. 501–516, January 2014.
- [35] —, "Efficient Approximation of Structured Covariance under Joint Toeplitz and Rank Constraints," in *Signals, Systems and Computers, 2013 Conference Record of the Forty Seventh Asilomar Conference on*, November 2013, pp. 692–696.

- [36] —, “Constrained ML estimation of structured covariance matrices with applications in radar STAP,” in *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2013 IEEE 5th International Workshop on*, December 2013, pp. 101–104.
- [37] —, “Estimation of Structured Covariance Matrices for Radar STAP under Practical Constraints,” in *IEEE Radar Conference*, May 2014, pp. 585–590.
- [38] B. Kang, “Student Research Highlight: Estimation of Structured Covariance Matrices for Radar STAP,” *IEEE Aerosp. Electron. Syst. Mag.*, vol. 30, no. 2, pp. 24–25, 2015.
- [39] B. Kang, V. Monga, M. Rangaswamy, and Y. I. Abramovich, “Automatic Rank Estimation for Practical STAP Covariance Estimation via an Expected Likelihood Approach,” in *IEEE Radar Conference*, 2015.
- [40] —, “Robust Covariance Estimation under Imperfect Constraints using Expected Likelihood Approach,” *IEEE Trans. Aerosp. Electron. Syst.*, 2015.
- [41] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nizberg, “A CFAR Adaptive Matched Filter Detector,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208–216, January 1992.
- [42] G. Cao and C. Bouman, “Covariance Estimation for High Dimensional Data Vectors Using the Sparse Matrix Transform,” Purdue University, Tech. Rep., April 2008.
- [43] L. Vandenberghe, S. Boyd, and S. P. Wu, “Determinant Maximization with Linear Matrix Inequality Constraints,” *SIAM. J. Matrix Anal. and Appl.*, vol. 19, no. 2, pp. 499–533, 1998.
- [44] A. D. Maio, S. D. Nicola, L. Landi, and A. Farina, “Knowledge-Aided Covariance Matrix Estimation: a MAXDET Approach,” *IET Radar, Sonar and Navigation*, vol. 3, no. 4, pp. 341–356, August 2009.
- [45] D. R. Fuhrmann, “Application of Toeplitz Covariance Estimation to Adaptive Beamforming and Detection,” *IEEE Trans. Signal Processing*, vol. 39, no. 10, pp. 2194–2198, October 1991.
- [46] P. Stoica and T. Söderström, “On reparameterization of loss functions used in estimation and the invariance principle,” *Signal Processing*, vol. 17, no. 4, pp. 383–387, August 1989.
- [47] R. A. Monzingo and T. Miller, *Introduction to Adaptive Arrays*, 1st ed. SciTech Publishing, 2004.
- [48] G. T. Capraro, A. Farina, H. Griffiths, and M. C. Wicks, “Knowledge-Based Radar Signal and Data Processing: A tutorial overview,” *IEEE Signal Processing Mag.*, vol. 23, no. 1, pp. 18–29, January 2006.

- [49] S. Haykin, "Cognitive Radar: A way of the future," *IEEE Signal Processing Mag.*, vol. 23, no. 1, pp. 30–40, January 2006.
- [50] S. Miranda, C. Baker, K. Woodbridge, and H. Griffiths, "Knowledge-Based Resource Management for Multifunction Radar: A look at scheduling and task prioritization," *IEEE Signal Processing Mag.*, vol. 23, no. 1, pp. 66–76, January 2006.
- [51] A. D. Maio, G. Foglia, A. Farina, and M. Piezzo, "Estimation of the Covariance Matrix Based on Multiple A-Priori Models," in *IEEE Radar Conference*, May 2010, pp. 1025–1029.
- [52] T. W. Anderson, "Asymptotic Theory for Principal Component Analysis," *The Annals of Mathematical Statistics*, vol. 34, no. 1, pp. 122–148, 1963.
- [53] M. Skolnik, *Radar Handbook*. McGraw-Hill, New York, 1990.
- [54] M. Wax and T. Kailath, "Detection of Signals by Information Theoretic Criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, no. 2, pp. 387–392, April 1985.
- [55] B. Widrow, P. E. Mantey, L. J. Griffiths, and B. B. Goode, "Adaptive Antenna Systems," in *Proceedings of the IEEE*, vol. 55, December 1967, pp. 2143–2159.
- [56] S. P. Applebaum, "Adaptive arrays," Syracuse University Research Corporation, Tech. Rep., August 1966.
- [57] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid Convergence Rate in Adaptive Arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-10, no. 6, pp. 853–863, 1974.
- [58] J. R. Roman, M. Rangaswamy, D. W. Davis, Q. Zhang, B. Himed, and J. H. Michels, "Parametric Adaptive Matched Filter for Airborne Radar Applications," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 2, pp. 677–692, April 2000.
- [59] P. Wang, H. Li, and B. Himed, "A Simplified Parametric GLRT for STAP Detection," in *IEEE Radar Conference*, May 2009, pp. 1–5.
- [60] A. L. Kraay and A. B. Baggeroer, "A Physically Constrained Maximum-Likelihood Method for Snapshot-Deficient Adaptive Array Processing," *IEEE Trans. Signal Processing*, vol. 55, no. 8, pp. 4048–4063, August 2007.
- [61] A. D. Maio and A. Farina, "Adaptive Radar Detection: A Bayesian Approach," in *International Radar Symposium*, May 2006, pp. 1–4.
- [62] A. D. Maio, A. Farina, and G. Foglia, "Adaptive Radar Detection: A Bayesian Approach," in *IEEE Radar Conference*, April 2007, pp. 624–629.
- [63] O. Besson, J. Y. Tournet, and S. Bidon, "Knowledge-Aided Bayesian Detection in Heterogeneous Environments," *IEEE Signal Processing Lett.*, vol. 14, no. 5, pp. 355–358, May 2007.

- [64] P. Wang, H. Li, and B. Himed, "A Bayesian Parametric Test for Multichannel Adaptive Signal Detection in Nonhomogeneous Environments," *IEEE Signal Processing Lett.*, vol. 17, no. 4, pp. 351–354, April 2010.
- [65] P. R. Gurram and N. A. Goodman, "Spectral-Domain Covariance Estimation with A Priori Knowledge," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 3, pp. 1010–1020, July 2006.
- [66] W. L. Melvin and G. A. Showman, "An Approach to Knowledge-Aided Covariance Estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 3, pp. 1021–1042, July 2006.
- [67] O. Ledoit and M. Wolf, "Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection," *Journal of Empirical Finance*, vol. 10, no. 5, pp. 603–621, December 2003.
- [68] P. Stoica, J. Li, X. Zhu, and J. R. Guerci, "On Using A-Priori Knowledge in Space-Time Adaptive Processing," *IEEE Trans. Signal Processing*, vol. 56, no. 6, pp. 2598–2602, June 2008.
- [69] J. H. Won, J. Lim, S. J. Kim, and B. Rajaratnam, "Maximum Likelihood Covariance Estimation with a Condition Number Constraint," Department of Statistics, Stanford University, Tech. Rep., August 2009.
- [70] B. Kang and V. Monga, "Strict and relaxed rank constrained ML estimator of structured covariance matrices," The Pennsylvania State University, Tech. Rep., November 2012, http://signal.ee.psu.edu/TechReport_StrictRCML.pdf.
- [71] J. Bergin and P. Techau, "High-fidelity site specific radar simulation: Kassper'02 workshop datacube," DAPRA Technical Report, Tech. Rep. ISL-SCRD-TR-02-105, May 2002.
- [72] I. Kirsteins and D. Tufts, "Adaptive Detection Using Low Rank Approximation to a Data Matrix," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, no. 1, pp. 55–67, January 1994.
- [73] M. Rangaswamy, "Statistical Analysis of the Nonhomogeneity Detector for Non-Gaussian Interference Backgrounds," *IEEE Trans. Signal Processing*, vol. 53, no. 6, pp. 2101–2111, June 2005.
- [74] S. D. Blunt, K. R. Gerlach, and M. Rangaswamy, "The Enhanced FRACTA Algorithm With Knowledge-Aided Covariance Estimation," in *Proc. Third IEEE Workshop on Sensor Array and Multichannel Processing*, April 2004, pp. 638–642.
- [75] E. J. Kelly, "An Adaptive Detection Algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 1, pp. 115–127, March 1986.
- [76] A. Aubry, A. D. Maio, L. Pallota, and A. Farina, "Radar Detection of Distributed Targets in Homogeneous Interference Whose Inverse Covariance Structure is Defined via Unitary Invariant Functions," *IEEE Trans. Signal Processing*, vol. 61, no. 20, pp. 4949–4961, October 2013.

- [77] P. Chen, W. L. Melvin, and M. C. Wicks, "Screening among Multivariate Normal Data," *Journal of Multivariate Analysis*, vol. 69, no. 1, pp. 10–29, April 1999.
- [78] A. Aubry, A. D. Maio, L. Pallotta, and A. Farina, "Covariance Matrix Estimation via Geometric Barycenters and its Application to Radar Training Data Selection," *IET Radar, Sonar and Navigation*, vol. 7, no. 6, pp. 600–614, July 2013.
- [79] L. Davis, R. Evans, and E. Polak, "Maximum Likelihood Estimation of Positive Definite Hermitian Toeplitz Matrices Using Outer Approximations," in *Statistical Signal and Array Processing, 1998. Proceedings., Ninth IEEE SP Workshop on*, September 1998, pp. 49–52.
- [80] A. K. Shaw, S. Pokala, and R. Kumaresan, "Toeplitz and Hankel Matrix Approximation using Structured Approach," in *Acoustics, Speech, and Signal Processing, 1998. ICASSP-98., 1998 International Conference on*, vol. 4, May 1998, pp. 2349–2352.
- [81] T. J. Suffridge and T. L. Hayden, "Approximation by a Hermitian Positive Semidefinite Toeplitz Matrix," *SIAM Journal on Matrix Analysis and Applications*, vol. 14, no. 3, pp. 721–734, July 1993.
- [82] M. E. Tipping and C. M. Bishop, "Probabilistic Principal Component Analysis," *Journal of the Royal Statistical Society*, vol. 61, no. 3, pp. 611–622, October 1999.
- [83] P. Stoica and M. Jansson, "On Maximum Likelihood Estimation in Factor Analysis - An Algebraic Derivation," *Signal Processing*, vol. 89, no. 6, pp. 1260–1262, June 2009.
- [84] N. A. Goodman and J. M. Stiles, "On Clutter Rank Observed by Arbitrary Arrays," *IEEE Trans. Signal Processing*, vol. 55, no. 1, pp. 178–186, January 2007.
- [85] M. Skolnik, *Radar Handbook*, 3rd ed. The McGraw-Hill Professional, 2008.
- [86] C. Eckart and G. Young, "The Approximation of One Matrix by Another of Lower Rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, September 1936.
- [87] L. Pallotta, A. Aubry, A. D. Maio, and A. Farina, "Estimation of a Structured Covariance Matrix with a Condition Number Constraint for Radar Applications," in *IEEE Radar Conference*, May 2012, pp. 778–783.
- [88] E. Conte, M. Lops, and G. Ricci, "Asymptotically Optimum Radar Detection in Compound-Gaussian Clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 2, pp. 617–625, April 1995.
- [89] A. Aubry, A. D. Maio, and V. Carotenuto, "Optimality Claims for the FML Covariance Estimator with respect to Two Matrix Norms," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 49, no. 3, pp. 2055–2057, July 2013.
- [90] H. Akaike, "A New Look at the Statistical Model Identification," *IEEE Trans. Auto. Contr.*, vol. AC-19, no. 6, pp. 716–723, December 1974.

- [91] J. Rissanen, "Modeling by the Shortest Data Description," *Automatica*, vol. 14, no. 5, pp. 465–471, September 1978.
- [92] Y. Q. Yin and P. R. Krishnaiah, "On Some Nonparametric Methods for Detection of the Number of signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, no. 11, pp. 1533–1538, November 1987.
- [93] D. W. Tufts and A. A. Shah, "Rank Determination in Time-Series Analysis," in *Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., 1994 International Conference on*, vol. 4, April 1994, pp. IV21–IV24.
- [94] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lamber W function," *Advances in Computational Mathematics*, vol. 5, no. 1, pp. 329–359, 1996.
- [95] P. Chen, M. C. Wicks, and R. S. Adve, "Development of a statistical procedure for detecting the number of signals in a radar measurement," *IEE Proceedings – Radar, Sonar and Navigation*, vol. 148, no. 4, pp. 219–226, August 2001.
- [96] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse Invariance Covariance Estimation with Graphical Lasso," *Biostatistics*, vol. 9, pp. 432–441, July 2008.
- [97] J. H. Michels, B. Himed, and M. Rangaswamy, "Robust STAP detection in a dense signal airborne radar environment," *Signal Processing*, vol. 84, no. 9, pp. 1625–1636, September 2004.
- [98] M. Rangaswamy, J. H. Michels, and B. Himed, "Statistical analysis of the non-homogeneity detector for STAP applications," *Digital Signal Processing*, vol. 14, no. 3, pp. 253–267, May 2004.
- [99] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A Survey," *ACM Computing Surveys*, vol. 41, no. 3, July 2009.
- [100] N. V. Chawla, N. Japkowicz, and A. Kotcz, "Editorial: Special Issue on Learning from Imbalanced Data Sets," *SIGKDD Explorations*, vol. 6, no. 1, pp. 1–6, 2004.
- [101] C. Phua, D. Alahakoon, and V. Lee, "Minority Report in Fraud Detection: Classification of Skewed Data," *SIGKDD Explorer Newsletter*, vol. 6, no. 1, pp. 50–59, 2004.
- [102] G. Ratsch, S. Mika, B. Scholkopf, and K. R. Muller, "Constructing Boosting Algorithms from SVMs: An Application to One-Class Classification," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 9, pp. 1184–1199, 2002.
- [103] P. N. Tan, M. Steinbach, and V. Kumar, *Algorithms for Clustering Data*. Prentice-Hall Inc., 1988.
- [104] B. C. Levy, *Principles of Signal Detection and Parameter Estimation*. Springer, 2008.

Vita

Bosung Kang

Mr. Kang received the B. S. and M. S. degrees from Yonsei University, Seoul, Korea, in 2005 and 2007, respectively. He worked at LG Electronics as a research engineer, Seoul, Korea, from 2007 to 2011. He developed image and video signal processing algorithms for mobile cameras and monitors in LG Electronics. He was an intern in Air Force Research Lab, WPAFB, OH in 2013. His research interest include statistical signal processing, detection and estimation, and convex optimization. Mr. Kang received the Best Student Paper Award at IEEE International Radar Conference in 2014

Journal Publications:

1. Bosung Kang, Vishal Monga, and Muralidhar Rangaswamy, "Rank Constrained Maximum Likelihood Estimation of Structured Covariance Matrices," IEEE Transactions on Aerospace and Electronic Systems, vol. 50, no. 1, pp.501-515, Jan. 2014.
2. Bosung Kang, Vishal Monga, and Muralidhar Rangaswamy, "Computationally Efficient Toeplitz Approximation of Structured Covariance Under a Rank Constraint," IEEE Transactions on Aerospace and Electronic Systems, vol. 51, no. 1, pp. 775-785, Jan. 2015.
3. Bosung Kang, Vishal Monga, Muralidhar Rangaswamy, and Yuri Abramovich, "Robust Covariance Estimation under Imperfect Constraints using Expected Likelihood Approach," in preparation.