

Effects of Model Misspecification and Uncertainty on the Performance of Estimators

A Dissertation Presented

by

Gerald LaMountain

to

The Department of Electrical and Computer Engineering

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in

Electrical Engineering

Northeastern University

Boston, Massachusetts

December 2024

“All models are wrong; some models are useful.”

—George E. P. Box

“The imagination of nature is far, far greater than the imagination of man.”

—Richard Feynman

*In loving memory of Zorida Shalaby,
and to my family and rabbit.*

Contents

List of Figures	v
List of Acronyms	vii
Acknowledgments	ix
Abstract of the Dissertation	xi
1 Introduction	1
1.1 Background	1
1.2 Motivation	2
1.3 Objectives	2
1.4 Summary of Contributions	3
1.5 Dissertation Outline	6
2 Optimizing Estimator Design using Performance Bounds	8
2.1 Synthetic Aperture DOA Estimation: Applications and Challenges	8
2.2 Spatial filtering of RF signals with Synthetic Aperture Processing	10
2.2.1 Signal Model for Synthetic Aperture Processing	10
2.2.2 Cramér-Rao Bound Derivation	12
2.3 Maximum-Likelihood Validation	14
2.4 Maneuver Optimization via CRB Minimization	15
2.5 Maneuver Parametrization	16
2.6 Optimization Methods	16
2.6.1 Generalized Pattern Search	17
2.6.2 Bayesian Optimization	18
2.7 Experimental Validation for Optimizing SA-DOA Trajectories	18
2.8 Conclusions	19
3 Misspecified MUSIC and MLE for Multi-Sensor Arrays	22
3.1 Array Estimation: Applications and Challenges	22
3.2 Signal Models	24
3.3 Maximum-Likelihood Estimation and Model Misspecification	25
3.3.1 Optimal Estimators	26

3.3.2	Subspace Estimation Methods for Multi-Sensor Estimation	27
3.3.3	Estimation under Model Misspecification	28
3.4	Relationship between MUSIC and MLE under Model Misspecification	30
3.5	Applications in Array Processing	32
3.5.1	Sensor Position Error	32
3.5.2	Mutual Coupling Error	33
3.6	Simulation results	34
3.6.1	Single signal experiments	36
3.6.2	Multiple signals experiments	38
3.7	Conclusions	39
Appendices		41
3.A	Proof of Theorem 1	41
3.B	Proof of Corollary 1.1	44
3.C	Proof of Theorem 2	45
4	Bayesian Statistics Estimation in Gaussian Filtering	46
4.1	Dynamic State Estimation: Applications and Challenges	46
4.2	Review of Recursive Gaussian Filtering	48
4.3	Bayesian Estimation of Distribution Parameters	49
4.3.1	Method of Conjugate Priors	49
4.3.2	Initial Prior	51
4.4	Leveraging Bayesian Estimation for Recursive Gaussian Filtering	51
4.4.1	Initializing the Bayesian Statistics Estimator	52
4.4.2	Accounting for Changes in the Noise Characteristics	53
4.5	Implementation Using Linear and Cubature Kalman Filtering Strategies	57
4.5.1	Linear Filtering with Standard Kalman Filter	57
4.5.2	Nonlinear Filtering with Cubature Kalman Filter	58
4.6	Simulation-Based Experimental Validation	60
4.6.1	Linear/Gaussian Multivariate System	61
4.6.2	Nonlinear/Gaussian Multivariate System	67
4.7	Conclusions	70
5	A Bayesian Pseudotrue State for Dynamic State Estimation	74
5.1	Dynamic Model Misspecification: Applications and Challenges	74
5.2	On Parametric Misspecification for Bayesian Estimation	75
5.3	The Posterior Pseudotrue State for Dynamic State Estimation	76
5.3.1	Computing the Posterior Pseudotrue State for Linear Gaussian Systems	77
5.4	Simulation-Based Experimental Validation	78
5.5	Conclusions	85
Appendices		87
5.A	Proof of Equality (5.24)	87

6 Summary of Key Results and Conclusions	89
6.1 Dissertation Contributions	89
6.2 Directions for Future Research	90
6.3 Final Remarks	91
Bibliography	92

List of Figures

2.1	Possible System Outline	10
2.2	RMSE for 50° Azimuth, various Elevation angles (1000 Monte Carlo averages) . .	15
2.3	direction of arrival (DOA)-Trace Objective function minimum using generalized pattern search (GPS) optimization. Ground truth of 50° Azimuth, various Elevation angles.	20
2.4	DOA-Trace Objective function minimum using Bayesian optimization. Ground truth of 50° Azimuth, various Elevation angles.	21
3.1	Array configurations for various spacing errors and various directions of arrival. . .	36
3.2	Estimator performance with respect to the pseudotrue parameter for position error misspecification.	37
3.3	Misspecified estimator performance with respect to the pseudotrue parameter for mutual coupling model misspecification.	38
3.4	Difference between misspecified maximum-likelihood (MML) and misspecified MULTIPLE SIGNAL CLASSIFICATION (MMUSIC) estimator output with respect to the pseudotrue parameter for random additive misspecification.	39
4.4.1	Block diagram of the augmented Gaussian filter (GF) with Bayesian estimation of innovations' noise statistics.	52
4.4.2	Bayesian Re-initialization scheme for the augmented GF.	56
4.6.1	Measurement noise covariance estimates. Experiments shown represent a linear/Gaussian multivariate system with step increase in measurement noise (top) and step decrease in measurement noise (bottom), respectively. Comparison of the true model covariance with Myers approach ($L = 8$, $L = 100$), and the proposed Bayesian methodology with both informative priors and time-averaged NIS re-initialization. .	63
4.6.2	Measurement noise covariance RMSE. Experiments shown represent a linear/Gaussian multivariate system with step increase in measurement noise (top) and step decrease in measurement noise (bottom), respectively.	64
4.6.3	Measurement noise covariance RMSE cumulative distribution (CDF). Experiments shown represent a linear/Gaussian multivariate system with step increase in measurement noise (top) and step decrease in measurement noise (bottom), respectively. Plots shown for both the state (left) and measurement noise covariance (right) estimation errors.	65

4.6.4 Time-averaged Normalized Innovation Squared (NIS). Experiments shown represent a linear/Gaussian multivariate system with step increase in measurement noise (top) and step decrease in measurement noise (bottom), respectively.	66
4.6.5 Measurement noise covariance estimates. Experiments shown represent a nonlinear/Gaussian radar system system with step increase in measurement noise (top) and step decrease in measurement noise (bottom), respectively. Comparison of the true model covariance with Myers approach ($L = 8$, $L = 100$), and the proposed Bayesian methodology with both informative priors and time-averaged NIS reinitialization.	68
4.6.6 Measurement noise covariance root mean squared error (RMSE). Experiments shown represent a nonlinear/Gaussian radar system system with step increase in measurement noise (top) and step decrease in measurement noise (bottom), respectively.	69
4.6.7 Measurement noise covariance RMSE cumulative distribution (CDF). Experiments shown represent a nonlinear/Gaussian radar system with step increase in measurement noise (top) and step decrease in measurement noise (bottom), respectively. Plots shown for both the State (left) and measurement noise covariance (right) estimation errors.	71
4.6.8 Time-averaged Normalized Innovation Squared (NIS). Experiments shown represent a nonlinear/Gaussian radar system with step increase in measurement noise (top) and step decrease in measurement noise (bottom), respectively.	72
5.4.1 Comparison of pseudotrue state to average linear KF mismatched estimate, true state to average KF matched estimate. Misspecified state transition matrix.	80
5.4.2 Comparison of pseudotrue state to average linear KF mismatched estimate, true state to average KF matched estimate. Misspecified measurement matrix.	81
5.4.3 Comparison of pseudotrue state to average linear KF mismatched estimate, true state to average KF matched estimate. Misspecified measurement covariance.	82
5.4.4 Comparison of pseudotrue state to average linear KF mismatched estimate, true state to average KF matched estimate. Misspecified state transition covariance.	83
5.4.5 Comparison of pseudotrue state to average linear KF mismatched estimate, true state to average KF matched estimate. Total model misspecification.	84
5.4.6 Comparison of pseudotrue state to average linear KF mismatched estimate, true state to average KF matched estimate. Randomized assumed model.	86

List of Acronyms

BSE	Bayesian statistics estimation
CKF	cubature Kalman filter
CRB	Cramér-Rao lower bound
DML	deterministic maximum-likelihood
DOA	direction of arrival
FA-DOA	fixed aperture direction of arrival
FIM	Fisher Information Matrix
GF	Gaussian filter
GNSS	global navigation satellite systems
GPS	generalized pattern search
KF	Kalman filter
KLD	Kullback–Leibler divergence
MAP	maximum a posteriori
MBCRB	misspecified Bayesian Cramér-Rao lower bound
MCM	mutual coupling matrix
MCRB	misspecified Cramér-Rao lower bound
MLE	maximum-likelihood estimation
ML	maximum-likelihood
MMLE	misspecified maximum-likelihood estimation
MML	misspecified maximum-likelihood
MMSE	minimum mean-squared error

MMUSIC misspecified Multiple Signal Classification
MPCRB misspecified posterior Cramér-Rao lower bound
MS-unbiased misspecified unbiased
MUSIC Multiple Signal Classification
NIS normalized innovation squared
PF particle filter
PNT Positioning, Navigation and Timing
RDF radio direction finding
RMSE root mean-squared error
SA-DOA synthetic aperture direction of arrival
SCKF square-root cubature Kalman filter
SML stochastic maximum-likelihood
TDoA time difference of arrival
UAV unmanned aerial vehicle

Acknowledgments

They say that it takes a village to raise a child. Over the past several years, I've learned that the same goes for raising a Ph.D. The work that you will read about in this dissertation would not have been possible without the guidance, mentorship, inspiration, encouragement, and support of the many people in my life.

The first person to whom I wish to express my gratitude is my advisor, Professor Pau Closas, who has taught me almost all I know about being a researcher. His patient teaching and guidance led me to a bevy of opportunities to grow and expand my horizons, to identify opportunities to contribute to the knowledge of this world, and to overcome the many challenges that stood between me and publishing this dissertation. I first met Pau when he had just come to Northeastern, and looking back it is almost impossible to believe how far we have come together. His ability to bring people together, identify innovative solutions, and continuously advance the field's boundaries is an example that every researcher should aspire to.

I would also like to express my thanks to the other members of my dissertation committee, Professor Deniz Erdoğan and Professor Aanjhan Ranganathan. As far back as the first year of my undergraduate studies, Deniz has been helping me to find opportunities to develop into the researcher I am today. His rigorous approach to research is something that I have always admired and striven for, and will continue to strive for in my future endeavors. Although I have known Aanjhan for a shorter period, the energy he brings to every discussion, his intellectual curiosity, and his capacity to engage deeply with a broad range of material have been no less impactful on determining the kind of academic I want to be.

I am grateful as well to my supporters from outside the university, Dr. Jordi Vilà-Valls, now professor at ISAE-SUPAERO, and Dr. Tales Imbiriba, now professor at UMASS Boston. I have met few people who tackle both research and life (not to mention the mountains) with the determination and strength that Jordi brings. Tales has been a great mentor to me since he first came to Northeastern, and has continued to help me as a researcher even while pursuing his new opportunities. I consider both of them to be great friends, and I hope to have opportunities to continue to collaborate with both of them in the future.

I am also indebted to the past and present members of the Information Processing Laboratory at Northeastern University—Haoqing Li, Parisa Borhani, Peng Wu, Shuo Tang, Guillermo Hernandez, Yorgos Stratis, Helena Calatrava, and Mariona Jaramillo Civill—whose camaraderie, insights, and support have made IPL feel like home. I look forward to seeing all of the incredible things that you all accomplish in the future. In addition, I want to thank my friends and colleagues from outside of the lab, Dr. Richard Obermeier, now at Systems Technology Research, and Dr. Lorenzo Ortega who is now professor at ISAE-SUPAERO. They are both some of the smartest people I have ever met, and both have been inspirational to me in pursuit of my Ph.D.

Thank you to Professor Masoud Salehi, for his guidance all the way back during my undergraduate capstone project, and for helping me navigate the process of becoming a Ph.D student here at Northeastern University.

Finally, and of course not least of all, thank you to my family. To my parents Jason and Annie LaMountain, who have pushed me and given me the encouragement and strength to be the best version of myself that I could be. To my siblings, Liam and Fariel LaMountain, who have each been inspirational to me in so many ways. I am so proud of both of you, and I can't wait to see where you two go from here. To the smallest members of the family, Jaden and Pepper LaMountain. Jaden, my childhood Jack Russell, whose presence at home has always brought comfort and warmth, and Pepper, my sweet baby bun, who provided companionship through countless late nights of writing. It's no exaggeration to say that this dissertation might not have been possible without her gentle presence and calming influence.

Abstract of the Dissertation

Effects of Model Misspecification and Uncertainty on the Performance of
Estimators

by

Gerald LaMountain

Doctor of Philosophy in Electrical Engineering

Northeastern University, December 2024

Prof. Pau Closas, Advisor

Prof. Deniz Erdogmus, Committee Member

Prof. Aanjhan Ranganathan, Committee Member

System designers across all disciplines of technology face the need to develop machines capable of independently processing and analyzing data and predicting future data. This is the fundamental problem of interest in “estimation theory,” wherein probabilistic analyses are used to isolate relationships between variables, and in “statistical inference,” wherein those variables are used to make inferences about real-world quantities. In practice, all estimators are designed based on limited statistical generalizations about the behavior of the observed and latent variables of interest; however, these models are rarely fully representative of reality. In such cases, there exists a “model misspecification,” and the resulting estimators will produce results that differ from those of the properly specified estimators. Evaluating the performance of a given estimator may sometimes be done by direct comparison of estimator outputs to known ground truth. However, in many cases, there is no guarantee that future latent variables will follow the same rules as the limited ground truth available for analysis. When this is the case, it is useful to employ a second evaluation methodology: statistical generalization about estimator behavior based on candidate models of observed reality.

The overall objective of this dissertation is to evaluate and expand upon state-of-the-art approaches to estimation and estimator analysis under various types of modeling errors. Four main contributions are provided in this area. First, we contribute a use-case demonstration of applying estimator performance bounds analysis in a practical application: improving estimator design by optimization of system parameters, applied to optimizing sensor trajectories in synthetic aperture direction of arrival (SA-DOA) based on the Cramér-Rao lower bound (CRB). Next, we contribute

an analysis of the efficacy of computationally efficient signal subspace algorithms (e.g. Multiple Signal Classification (MUSIC)) for performing multi-antenna radio direction finding in the presence of various modeling errors (such as antennae mutual coupling or sensor placement errors). To that aim, we leverage the framework of misspecified Cramér-Rao lower bound (MCRB) to characterize the impact of modeling errors. This thesis provides an analysis that extends the known results on the equivalence of maximum-likelihood estimation (MLE) and MUSIC methods to cases of model misspecification. Next, we contribute an online method of Bayesian covariance estimation that leverages conjugate prior analysis, which, when embedded within Kalman-type filtering architectures, may be used to adapt to real-time changes in sensor performance. Finally, we focus on the analysis of state estimation problems under modeling. In those situations, we show that the expected value of a state estimator converges to the so-called pseudotrue state, which is formalized in this thesis for the first time. This result has important implications in defining theoretical estimation bounds under model misspecification in filtering contexts, for instance the misspecified posterior Cramér-Rao lower bound (MPCRB) which is applicable to practical dynamical estimation problems, and solutions, including the popular Kalman filter and other similar estimators. Particularization of the proposed pseudotrue expression to linear/Gaussian systems results in a form which may be cheaply computed for an arbitrary set of hypothetical ground truth states, allowing for rapid analysis of dynamic estimator performance against any number of proposed ground truth models.

Chapter 1

Introduction

1.1 Background

Throughout the history of modern technological development, one of the most often recurring questions to come to the mind of engineers has surely been “how can we use this data to learn more about the underlying processes at work?” Across the diverse fields of engineering, science, and technology, there has long been a fundamental need to devise principled and accurate processes and algorithms for analyzing data and making meaningful inferences about quantities that cannot be directly observed. Over these many years, numerous approaches have been applied to this task, including those that fall under the dual umbrellas of “estimation theory” and “statistical inference.” Methodologies within these fields rely on measurements or observations, referred to as the “data,” which are combined with “statistical models” of the relationships between observable and hidden quantities to infer information about those hidden quantities.

As the field of estimation theory continued to develop throughout the 20th century, two distinct approaches to statistical inference began to take shape. The first to emerge, commonly referred to as “classical statistics” or “frequentist statistics,” develops an understanding of probability based on the frequency at which observed data exhibits expected qualities. This approach has been shown to be a powerful tool in a number of estimation and inference contexts, and includes popular methods such as MLE and statistical hypothesis testing. The second was the re-emergence of “Bayesian statistics,” which develops a competing understanding of probability based around a concept of “belief” in a given event. The Bayesian approach has several advantages in comparison to frequentist approaches that make it a powerful tool in its own right, not the least of which is the inclusion of statistical representations of prior beliefs. Bayesian methods include Gaussian filters

and maximum a posteriori (MAP) estimation.

1.2 Motivation

Whether one is selecting a hammer for driving nails or choosing an estimation algorithm for a radar array, part of the process of identifying the best tool for the job involves analyzing the performance of the available options, and, where possible, refining them to improve performance with respect to relevant metrics. This principle applies equally to estimation algorithms as well as statistical models: models based on human understanding of existing relationships are rarely fully representative of the reality of the observed and latent variables of interest, and will work better or worse depending on the quality of representation and the particulars of the problem being addressed. Misrepresentation, or misspecification, in statistical models can occur for a number of reasons, including unforeseen confounding environmental factors and design factors, such as simplifying assumptions for performance or cost.

There exists an extensive body of research investigating the behavior of parameter estimators under various cases of model misspecification, leading to the development of powerful analytical frameworks, such as the “pseudotrue parameter” and the misspecified Cramér-Rao lower bound (MCRB) [1], that quantify how estimators perform when the assumed model differs from the true one. At the same time, there has also been extensive research showing the optimality of estimators, including frequentist and Bayesian estimators for both static and dynamic systems, under optimal conditions with well-known and precise dynamic models [2]. However, there exists a gap in the state-of-the-art when it comes to bounding the performance of dynamic Bayesian estimators and their derivatives (e.g. Kalman filter) under model misspecification. This represents a significant limitation, not only because of the ubiquity of dynamic systems and associated estimators in many different fields of technology from radar tracking to navigation, but because misspecification analysis is perhaps even more important in dynamic state estimation than in static state estimation, where gaps in understanding compound upon themselves as the parameters of interest propagate through multiple iterations of the system.

1.3 Objectives

The overarching objective of this thesis is to expand upon established approaches to estimation and estimator analysis under various types of misspecification, including those caused

CHAPTER 1. INTRODUCTION

by common environmental and design factors, such as unknown or variable observation noise. Building upon these estimation and analysis methodologies, this work seeks to narrow the gap in our understanding of how to effectively evaluate and improve dynamic state estimator performance when the assumed model differs from the true underlying process. Approaching these objectives from multiple angles, we have identified and pursued several key research components:

1. *Rectifying Classes of Model Uncertainty*: Identifying and implementing principled approaches to handling model uncertainty, for example, estimating Gaussian estimator tuning parameters.
2. *Applying and Extending Statistical Bounds*: Leveraging analytical frameworks, including the pseudotrue parameter and the misspecified Cramér-Rao lower bound (MCRB), to quantify the impact of model misspecification on estimator performance.
3. *Enhancing Estimator Performance Through Bounds Analysis*: Translating bounding analyses into actionable strategies that improve estimator performance, for example, optimizing sensor placement based on minimizing performance losses based on the Cramér-Rao lower bound (CRB).
4. *Extending Misspecification Analysis Tools to Dynamic State Estimation Problems*: Extending the principles and tools developed to a broader range of estimation problems, including dynamic state estimation, for which current research into misspecification analysis is limited.

To validate these approaches, the thesis employs a simulation-based methodology which involves generating test data from a hidden “true model,” and applying a separate “assumed model” estimator to this data. By comparing the misspecified estimator outputs against the known true states and the derived analysis tools (*i.e.* pseudotrue parameters and associated bounds) using metrics like the root mean-squared error (RMSE) across multiple Monte Carlo simulations, we aim to rigorously assess the effectiveness and applicability of the proposed solutions.

1.4 Summary of Contributions

This dissertation summarizes several contributions to the application and expansion of modern tools for misspecified estimator and model analysis. These contributions span a fairly wide gamut of research topics from establishing the convergence properties of computationally efficient estimation algorithms to establishing new misspecified estimator performance analysis tools. As such, there are several ways to break down this body of work; however, the intended breakdown

CHAPTER 1. INTRODUCTION

matches the progression from applying existing tools which exist for static parameter estimation to developing new tools for dynamic systems. With regard to the former, the key contributions for parameter estimation are

- A framework for using bounds analysis for improving and tuning estimation strategies, applied to matched model conditions based on optimization of the CRB for flight pattern planning in drone-based SA-DOA [3].
- Conditions for the asymptotic equivalence of computationally efficient eigenspace methods (MUSIC) and maximum-likelihood for DOA estimation using multi-antenna arrays [4].

With regard to the latter, the key contributions for dynamic state estimation are

- A framework for using Bayesian statistics estimation (BSE) to rectify incomplete measurement model information within the Kalman filter (KF) framework [3, 5].
- A definition of the pseudotrue state that extends the concept of a pseudotrue parameter from static to dynamic state estimation, including a recursive form for the linear KF.

Part of the materials presented in this thesis have been the subject of various publications as cited, with others awaiting review and publication at the time of writing:

- G. LaMountain and P. Closas, “Performance of MUSIC and Maximum-Likelihood DOA Estimation under Model Misspecification” awaiting review in IEEE Transactions on Aerospace and Electronic Systems.
- G. LaMountain, J. Vilà-Valls, and P. Closas, “Bayesian Statistics Estimation in Gaussian Filtering” awaiting review in EURASIP Journal on Advances in Signal Processing.
- G. LaMountain, S. Tang, T. Imbiriba and P. Closas, “A Bayesian Pseudotrue State for Gaussian Filters” awaiting review in IEEE Signal Processing Letters.

Other Research Contributions

As is the case in many engineering disciplines, much of the work in this dissertation is based on mathematical principles that may readily be applied to various applications and technologies. During the course of developing this body of research, there have been a number of other projects outside the primary scope of this dissertation that have contributed to both the broader research strategy and specific mathematical understandings that motivated the works described here.

CHAPTER 1. INTRODUCTION

The first set of projects that merit special mention are those that were carried out in coordination with GNSS-SDR [6–8], including those performed under the Google Summer of Code initiative. GNSS-SDR is an open source multiplatform system that accelerates research by providing a modifiable implementation of a global navigation satellite systems (GNSS) software-defined receiver. Many of the fundamental challenges in GNSS positioning may be expressed as estimation problems and, as such, although not explicitly stated in later chapters, several projects in this dissertation were at some point implemented in GNSS-SDR for testing using real-world data for, for example, carrier phase synchronization. This was the case with “Proposal to implement Bayesian Covariance Estimation for Kalman Filter based Digital Carrier Synchronization in GNSS-SDR,” for Google Summer of Code 2018. Other projects that utilized GNSS-SDR that are beyond the scope of the main contributions of this dissertation include:

- “Nonlinear State Estimation for Carrier Synchronization in GNSS-SDR,” Summer 2019. Participant: Gerald LaMountain. Mentors: Jordi Vilà-Valls. Google Summer of Code 2019
- “Robust Interference Mitigation for anti-jamming,” Summer 2020. Participant: Haoqing Li. Mentors: Gerald LaMountain, Pau Closas. Google Summer of Code 2020
- “Anti-spoofing Techniques for GNSS Receivers,” Summer 2021. Participant: Harshad Sathaye. Mentors: Gerald LaMountain. Google Summer of Code 2021

Furthermore, there are other projects that contributed to an understanding of various applications in Positioning, Navigation and Timing (PNT) and other fields that, although not directly included in the scope of this dissertation, were informational regarding general estimation problems and the needs and limitations of current analysis tools in this field:

- [9] T. Imbiriba, P. Wu, G. LaMountain, D. Erdoğmuş, and P. Closas, “Recursive Gaussian processes and fingerprinting for indoor navigation,” in 2020 IEEE/ION Position, Location and Navigation Symposium (PLANS), pp. 933–940, 2020
- [10] T. Imbiriba, G. LaMountain, P. Wu, D. Erdoğmuş, and P. Closas, “Change detection and Gaussian process inference in piecewise stationary environments under noisy inputs,” in 2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), pp. 530–534, IEEE, 2019
- [11] P. Wu, T. Imbiriba, G. LaMountain, J. Vilà-Valls, and P. Closas, “Wifi fingerprinting and tracking using neural networks,” in Proceedings of the 32nd International Technical Meeting

CHAPTER 1. INTRODUCTION

of the Satellite Division of The Institute of Navigation (ION GNSS+ 2019), pp. 2314–2324, 2019

- [12] G. Hernandez, G. LaMountain, and P. Closas, “Privacy-preserving cooperative positioning,” in Proceedings of the 33rd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2020), pp. 2667–2675, 2020
- [13] G. Hernandez, G. LaMountain, and P. Closas, “Proximity-based positioning scheme with multi-layer privacy,” in 2023 IEEE/ION Position, Location and Navigation Symposium (PLANS), 2023, pp. 235–242.
- [14] G. Hernandez, G. LaMountain, and P. Closas, “Privacy-preserving cooperative GNSS positioning,” *NAVIGATION: Journal of the Institute of Navigation*, vol. 70, no. 4, 2023. [Online]. Available: <https://navi.ion.org/content/70/4/navi.625>
- [15] H. Sathaye, G. LaMountain, P. Closas, and A. Ranganathan, “Semperfi: A spoofer eliminating GPS receiver for UAVs,” in Proceedings of Network and Distributed System Security Symposium (NDSS 2022), 2022

These and related projects contributed to the development and submission of two patents (Pending):

- [16] M. T. Kling, D. Lau, K. L. Witham, and G. M. LaMountain, “System for closed-loop GNSS simulation,” U.S. Patent Application 0 365 223, Dec. 09, 2024.
- [17] P. Closas, G. Hernandez, and G. LaMountain, “Method and apparatus for determining a geospatial location of an unlocated device,” U.S. Patent Application 0 276 226, Dec. 09, 2024.

Finally, there is one article that was part of the same project as, and instrumental in laying the foundation to, one of the key contributions of this dissertation highlighted in Chapter 5, that is

- [18] S. Tang, G. LaMountain, T. Imbiriba, and P. Closas, “On parametric misspecified Bayesian Cramér-Rao bound: An application to linear/Gaussian systems,” in ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1–5, 2023

1.5 Dissertation Outline

This dissertation is structured in five chapters, each of which are based on several key publications and submissions contributing to the objectives stated in 1.3. The intended audience is

CHAPTER 1. INTRODUCTION

those who have an interest in dynamic state and parameter estimation and wish to identify methods of approaching model limitations in associated problems thereof, and those who have an interest in model misspecification and understanding the available tools for performing misspecification analysis. This section is intended to help navigate the various contributions and briefly summarize each chapter. At the beginning of each chapter, a short abstract provides further details.

This initial **Chapter 1** has introduced the background and motivation which inspired this dissertation, described the breakdown of objectives upon which it is built, and summarized the primary contributions that form its foundation. In addition, this section has highlighted several ancillary contributions that have been informational regarding general engineering knowledge and understanding of motivating applications in estimation theory.

Each of the four core chapters following the introduction represent one of the contributions described in 1.4, which in turn correspond to the key research components identified in 1.3.

Chapter 2 includes contributions on the use of bounds analysis tools for optimizing state parameters, with applications in the context of SA-DOA in the *matched case*. This chapter serves to demonstrate the potential of using bounds analysis for improving and tuning estimation strategies.

Chapter 3 describes contributions centered on the comparison of frequentist and eigenspace methods for estimation in array-based DOA estimation, under conditions of model misspecification. Conditions are provided for the asymptotic equivalence of MMUSIC and MML estimators.

Chapter 4 focuses on contributions related to Gaussian filters and model uncertainty rather than model misspecification. It introduces a novel Bayesian statistics estimation (BSE) methodology for online estimation of the innovations covariance matrix within linear and nonlinear Gaussian filtering strategies. The chapter also provides a framework for incorporating BSE within existing Gaussian filter architectures, with particular attention given to the class of sigma-point filters (e.g. cubature Kalman filter (CKF)) for nonlinear Gaussian filtering.

Chapter 5 describes the final main contribution. This contribution centers around the extension of the “pseudotrue parameter,” established in the literature for misspecified parameter estimation, to the class of Gaussian dynamic state estimators. Derivations and a framework for calculating the “pseudotrue state” in such systems are provided, as well as experiments validating this misspecified analysis tool in linear systems.

Finally, **Chapter 6** concludes this dissertation with a summary of the main findings, conclusions, and future avenues for research and practical applications extending from this work.

Chapter 2

Optimizing Estimator Design using Estimator Performance Bounds: Synthetic Aperture DOA

GNSS denial via jamming is a low skilled attack which can be performed by nearly anyone using tools which are readily available through the online marketplace. Methods of jammer mitigation such as beamforming or other active methodologies require an estimation of the location of the jamming signal source. There is interest in developing systems which can be used to identify and locate the sources of broadcast signals, either for the purposes of augmenting mitigation or for the purposes of taking direct action to eliminate the sources of jamming signals. One way that such localization systems might be deployed is on rotor-wing aircraft which may be utilized to perform synthetic aperture direction of arrival (SA-DOA). In this contribution we propose a methodology for identifying circular maneuver trajectories which minimize the estimation error of performing SA-DOA using such a platform. This methodology demonstrates the utility of using estimator performance bounds in a practical way: improving estimator performance by parameter optimization.

2.1 Synthetic Aperture DOA Estimation: Applications and Challenges

As GNSS usage in PVT applications has become increasingly ubiquitous over the past several decades, denial or manipulation of GNSS signals has become an increasingly critical vulnerability in many civilian and military systems [19–21]. In particular, GNSS denial via jamming is a low skilled attack which can be performed by nearly anyone using tools which are readily available through the online marketplace. In order to combat these types of attacks on critical

systems, substantial effort has been made into developing robust GNSS signal processing techniques. Some of these techniques have shown impressive results in mitigating the effect of jamming attacks, however there are no techniques capable of completely countering the loss of performance produced by jamming on GNSS based PVT [22, 23]. For this reason there is interest in developing systems which can be used to identify and locate the sources of broadcast signals, either for the purposes of augmenting mitigation techniques such as those which utilize beamforming or for the purposes of taking direct action to eliminate the sources of jamming signals. Passive jammer localization consists of two stages: first, the direction of arrival (DOA) of the jamming signal is estimated from several known locations, and second, the multiple direction of arrival estimates are used to compute an estimated location of the signal source. The final accuracy and precision of the location estimate depends on that of the direction of arrival estimates, and as such it is critical that the DOA estimates be as accurate and precise as possible.

For some applications, it is neither feasible nor desirable to deploy a fixed network of DOA estimators in established locations around potential jamming sources, due to constraints such as cost and mobility. Deploying a bulky and complex GNSS antenna array [24] on a battlefield, for instance, may not be practical even using some of the most compact solutions that can be found in the literature [25]. In these scenarios, DOA estimation may need to rely on mobile platforms, such as unmanned aerial vehicles (UAVs), which can be relocated in the field to perform localization synthetically. The literature provides examples of UAV systems equipped with small fixed sensor arrays of two or three antennae or, in larger systems, rotating arrays used in airborne early warning and control (AEW&C) systems. Since jammer localization problems often occur on a scale that is not conducive to using large aircraft, many of these systems are restricted to small arrays, limiting the precision that is achievable in comparison to larger arrays. In this contribution, we propose a broadcast source localization system leveraging the maneuvering capabilities of multi-rotor UAVs, such as commercially available quadcopters. This system performs SA-DOA estimation, which can achieve more reliable estimates than comparable fixed aperture direction of arrival (FA-DOA) architectures on similarly sized aircraft. Additionally, we propose a methodology to optimize aircraft maneuvers for SA-DOA estimation. By iteratively adjusting maneuvers based on present DOA estimates and the information inequality, this approach aims to improve estimation accuracy.

To realize the potential of this proposed system, we outline an architecture that achieves these goals through several intermediate steps, as detailed in Fig. 2.1. This chapter derives a maximum-likelihood (ML) estimator for DOA estimates using a synthetic aperture antenna and provides its estimation bound as determined by the CRB. In the proposed architecture, the UAV

follows an arbitrary trajectory, with its locations determined using, for example, an on-board inertial measurement unit (IMU). From this trajectory, an initial DOA estimate is obtained. This initial estimate is required to start the iterative process of minimizing the CRB. To improve estimation accuracy, we consider two optimization approaches, GPS and Bayesian optimization, to iteratively refine the trajectory of the UAV and minimize the CRB. These approaches ensure that the synthetic aperture structure is optimized to achieve the best possible precision within the bounds of the CRB.

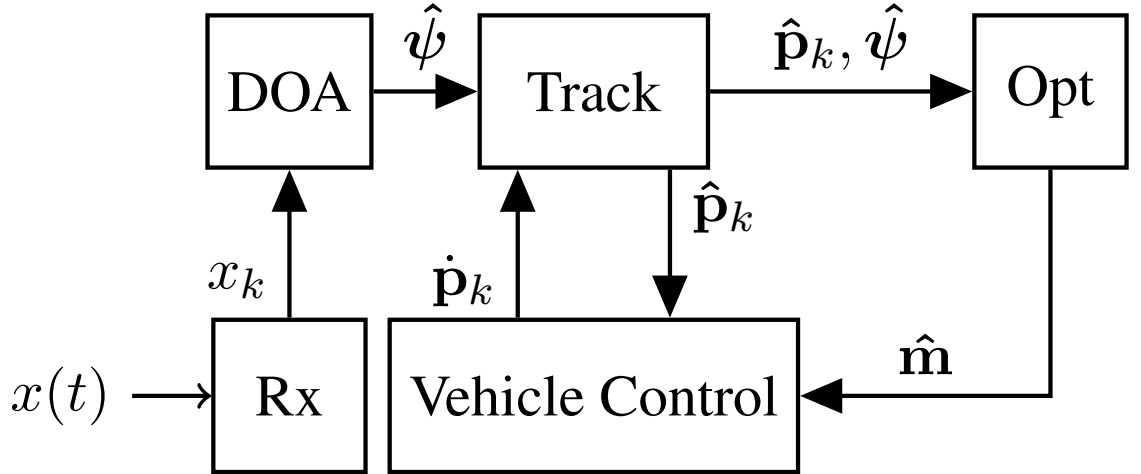


Figure 2.1: Possible System Outline

2.2 Spatial filtering of RF signals with Synthetic Aperture Processing

2.2.1 Signal Model for Synthetic Aperture Processing

The complex baseband representation of a line-of-sight (LOS) RF signal measured by a sensor is expressed as

$$x(t) = a(t)s(t - \tau(t)) \exp(-j2\pi f_c \tau(t)) + n(t) \quad (2.1)$$

Here, $s(t)$ is the baseband received signal content and $a(t)$ is the complex amplitude of the *received* signal produced when a signal of complex amplitude u impinges on the antenna with directional gain and relative polarization $G(t) = F(t)\gamma(t)$. $\tau(t)$ is the delay caused by the propagation of the signal through space. f_c is the nominal carrier frequency of the transmitted signal and $n(t)$ represents any additive noise, except from multipath reflections.

CHAPTER 2. OPTIMIZING ESTIMATOR DESIGN USING PERFORMANCE BOUNDS

The delays observed in the received signal are separated into two components: the *reference* delay representing the propagation time for the transmitted signal to the origin of the local frame of reference of the receiver, and the *relative* delay representing the small differences in propagation time for different local antenna locations within that local frame of reference. That is,

$$\tau(t) = \tau_{\text{ref}} + \tau_{\text{rel}} . \quad (2.2)$$

For small observation times and low relative velocities, the reference delay can be approximated to be of a constant value $\tau_{\text{ref}}(t) = \tau_0$, while the relative delay is expressed in terms of the position of the receiver in the local frame and a unit vector $\hat{\mathbf{r}}_\psi$ pointing in the direction of the transmitter as in (3.32). The relative delay, then, is consequently a function of the direction of arrival ψ , comprised of the relative azimuth ϕ and elevation θ of the transmitter from the perspective of the receiver. When we collect a set of K measurements of this signal at the receiver, we observe a measurement vector $\mathbf{x} \in \mathbb{C}^{K \times 1}$

$$\mathbf{x} = (\mathbf{d} \odot \mathbf{g})u + \mathbf{n} , \quad (2.3)$$

where $\mathbf{d} \in \mathbb{C}^{K \times 1}$ is the basis-function vector with elements expressed in terms of the sample time t_k as

$$[\mathbf{d}]_k = s(t - \tau(t_k)) \exp(-j2\pi f_c \tau(t_k)) . \quad (2.4)$$

$u \in \mathbb{C}^1$ is the complex constant amplitude of the signal impinging on the receiving antenna and $\mathbf{g} \in \mathbb{C}^{K \times 1}$ is the complex gain induced by the directionality and relative polarization of the antenna for each of the collected samples. For complex, real world radiation patterns this vector should be determined experimentally, but for some hypothetical cases however this may be computed analytically. In this chapter we consider the case where each antenna is isotropic and right-hand circularly polarized (RHCP). Under these constraints, the value of \mathbf{g} will depend only on the polarization mismatch coefficient $\gamma(t)$, which can be obtained as

$$\gamma(\mathbf{r}_\psi, \mathbf{\Upsilon}_{Rx}, \mathbf{\Upsilon}_{Tx}, \mathbf{\Gamma}_{Tx}, \mathbf{\Gamma}_{Rx}) = \phi_{g,Tx}^H \phi_{g,Rx} , \quad (2.5)$$

where the vectors $\phi_{g,Tx}$ and $\phi_{g,Rx}$ represent the normalized Jones polarization vectors of the transmitting and receiving antennas transformed respectively into a global frame of reference. This transformation requires the application of two pairs of rotations. The first, $\mathbf{\Upsilon}_{Rx}, \mathbf{\Upsilon}_{Tx} \in \mathbb{R}^{3 \times 3}$, are defined such that they each represent a transformation from the antenna's directional frame of

reference (*i.e.* that in which the boresight of the antenna is aligned with the z axis) into the global frame of reference. The second, $\mathbf{\Gamma}_{Rx}, \mathbf{\Gamma}_{Tx} \in \mathbb{R}^{3 \times 3}$, are defined such that they each represent a transformation from the antenna's relative frame of reference (*i.e.* that in which the z axis is aligned with \mathbf{r}_ψ and $-\mathbf{r}_\psi$ respectively) into the global frame of reference. Then, given the normalized Jones vector for RHCP, $\phi_{RHCP} = 1/\sqrt{2} [1, -j, 0]^\top$, we can compute

$$\phi_{g,Rx} = \mathbf{\Gamma}_{Rx} \mathbf{\Upsilon}_{Rx} \phi_{RHCP} \quad (2.6)$$

$$\phi_{g,Tx} = \mathbf{\Gamma}_{Tx} \mathbf{\Upsilon}_{Tx} \phi_{RHCP} \quad (2.7)$$

2.2.2 Cramér-Rao Bound Derivation

The Cramér-Rao lower bound (CRB) provides a lower bound on the covariance for joint, unbiased estimation of the deterministic parameters of a given model. The CRB is formed by taking the inverse of the Fisher Information Matrix (FIM) defined as

$$[\mathbf{J}(\boldsymbol{\xi})]_{ij} = -\mathbb{E} \left\{ \frac{\partial^2 \ln p(\mathbf{x}; \boldsymbol{\xi})}{\partial \xi_i \partial \xi_j} \right\} \quad (2.8)$$

where \mathbf{x} is the vector of measurements and $\boldsymbol{\xi}$ is the vector of parameters to be estimated. Under the assumption that the measurement vector is distributed as $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}(\boldsymbol{\xi}), \boldsymbol{\Sigma}_x)$, where $\boldsymbol{\Sigma}_x$ does not depend on the parameters $\boldsymbol{\xi}$, application of the Slepian-Bang's formula yields

$$[\mathbf{J}(\boldsymbol{\xi})]_{ij} = 2\Re \left\{ \frac{\partial \boldsymbol{\mu}^H(\boldsymbol{\xi})}{\partial \xi_i} \boldsymbol{\Sigma}_x^{-1} \frac{\partial \boldsymbol{\mu}(\boldsymbol{\xi})}{\partial \xi_j} \right\} \quad (2.9)$$

For the model described in (2.3), we have

$$\boldsymbol{\mu}(\boldsymbol{\xi}) = (\mathbf{d}(\mathbf{v}, \boldsymbol{\psi}) \odot \mathbf{g}(\boldsymbol{\psi}))u \quad (2.10)$$

with parameters

$$\boldsymbol{\xi} = \begin{bmatrix} \mathbf{u} \\ \boldsymbol{\psi} \\ \tau_0 \end{bmatrix}; \quad \mathbf{u} = \begin{bmatrix} \Re\{u\} \\ \Im\{u\} \end{bmatrix}; \quad \boldsymbol{\psi} = \begin{bmatrix} \theta \\ \phi \end{bmatrix}. \quad (2.11)$$

Under the assumption of a known channel spatio-temporal covariance matrix $\Sigma_{\mathbf{x}}$, the FIM submatrices are given by [26, 27]

$$\mathbf{J}_{u_i u_j} = 2\Re \left\{ \frac{\partial \mathbf{u}^H}{\partial u_i} (\mathbf{d} \odot \mathbf{g})^H \Sigma_{\mathbf{x}}^{-1} (\mathbf{d} \odot \mathbf{g}) \frac{\partial \mathbf{u}}{\partial u_j} \right\} \quad (2.12)$$

$$\mathbf{J}_{u_i \psi_j} = 2\Re \left\{ \frac{\partial \mathbf{u}^H}{\partial u_i} (\mathbf{d} \odot \mathbf{g})^H \Sigma_{\mathbf{x}}^{-1} \frac{\partial (\mathbf{d} \odot \mathbf{g})}{\partial \psi_j} \mathbf{u} \right\} \quad (2.13)$$

$$\mathbf{J}_{u_i \tau_0} = 2\Re \left\{ \frac{\partial \mathbf{u}^H}{\partial u_i} (\mathbf{d} \odot \mathbf{g})^H \Sigma_{\mathbf{x}}^{-1} \left(\frac{\partial \mathbf{d}}{\partial \tau_0} \odot \mathbf{g} \right) \mathbf{u} \right\} \quad (2.14)$$

$$\mathbf{J}_{\psi_i \psi_j} = 2\Re \left\{ \mathbf{u} \frac{\partial (\mathbf{d} \odot \mathbf{g})^H}{\partial \psi_i} \Sigma_{\mathbf{x}}^{-1} \frac{\partial (\mathbf{d} \odot \mathbf{g})}{\partial \psi_j} \mathbf{u} \right\} \quad (2.15)$$

$$\mathbf{J}_{\psi_i \tau_0} = 2\Re \left\{ \mathbf{u} \frac{\partial (\mathbf{d} \odot \mathbf{g})^H}{\partial \psi_i} \Sigma_{\mathbf{x}}^{-1} \left(\frac{\partial \mathbf{d}}{\partial \tau_0} \odot \mathbf{g} \right) \mathbf{u} \right\} \quad (2.16)$$

$$\mathbf{J}_{\tau_0 \tau_0} = 2\Re \left\{ \mathbf{u} \left(\frac{\partial \mathbf{d}}{\partial \tau_0} \odot \mathbf{g} \right)^H \Sigma_{\mathbf{x}}^{-1} \left(\frac{\partial \mathbf{d}}{\partial \tau_0} \odot \mathbf{g} \right) \mathbf{u} \right\} \quad (2.17)$$

The partial derivatives of \mathbf{d} are separated as

$$\frac{\partial \mathbf{d}}{\partial \xi_i} = \frac{\partial \mathbf{d}}{\partial \tau(t_k)} \frac{\partial \tau(t_k)}{\partial \xi_i} \quad (2.18)$$

with

$$\begin{aligned} \left[\frac{\partial \mathbf{d}}{\partial \tau(t_k)} \right]_k &= -\dot{s}(t_k - \tau(t_k)) \times \exp(j2\pi f_c \tau(t_k)) \\ &\quad - j2\pi f_c s(t_k - \tau(t_k)) \times \exp(j2\pi f_c \tau(t_k)) \end{aligned} \quad (2.19)$$

and

$$\left[\frac{\partial \tau(t_k)}{\partial \xi_i} \right]_i = \begin{cases} 1, & \text{if } \xi_i \in \tau_0 \\ c^{-1} \frac{\partial \Delta \rho(t_k)}{\partial \xi_i}, & \text{if } \xi_i \in \psi \end{cases} \quad (2.20)$$

The values of $\partial \mathbf{u} / \partial \xi_i$ are $[1, 0]^\top$ for $\partial \mathbf{u} / \partial \Re\{u\}$ and $[0, j]^\top$ for $\partial \mathbf{u} / \partial \Im\{u\}$ respectively, with j the unit imaginary operator. Finally, for \mathbf{g} we need to compute the partial derivatives of the gain matrix with respect to the DOA parameters ψ . These derivatives depend on the direction of arrival ψ , as well as the radiation pattern and relative polarization of both the transmitting and receiving antennas. For the isotropic right-hand circularly polarized (RHCP) case, these derivatives can be computed analytically by applying the chain rule to (2.5) using the partial derivatives of the DOA vector

$$\frac{\partial \mathbf{r}_\psi}{\partial \theta} = \begin{bmatrix} -\sin \theta \cos \phi \\ -\sin \theta \sin \phi \\ \cos \theta \end{bmatrix}; \quad \frac{\partial \mathbf{r}_\psi}{\partial \phi} = \begin{bmatrix} -\cos \theta \sin \phi \\ \cos \theta \cos \phi \\ 0 \end{bmatrix}; \quad (2.21)$$

2.3 Maximum-Likelihood Validation

Under the assumption that the vector of received signal measurements is distributed as $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}(\boldsymbol{\xi}), \boldsymbol{\Sigma}_{\mathbf{x}})$ with known spatio-temporal covariance matrix $\boldsymbol{\Sigma}_{\mathbf{x}}$, the probability of observing a given measurement vector \mathbf{x} is given by the likelihood function

$$p(\mathbf{x}; \boldsymbol{\xi}) = \frac{1}{\pi^n \det(\boldsymbol{\Sigma}_{\mathbf{x}}(\boldsymbol{\xi}))} \exp \left\{ -(\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\xi}))^H \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} (\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\xi})) \right\} \quad (2.22)$$

from which we define the maximum-likelihood solution for parameter vector $\boldsymbol{\xi}$ as

$$\hat{\boldsymbol{\xi}}_{\text{ML}} = \arg \max_{\boldsymbol{\xi}} p(\mathbf{x}; \boldsymbol{\xi}) = \arg \max_{\boldsymbol{\xi}} \ln p(\mathbf{x}; \boldsymbol{\xi}) \quad (2.23)$$

$$= \arg \min_{\boldsymbol{\xi}} (\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\xi}))^H \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} (\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\xi})). \quad (2.24)$$

Here, $\ln p(\mathbf{x}; \boldsymbol{\xi})$ is the log-likelihood function and the mean of the measurements is defined as in (2.3).

$$\boldsymbol{\mu}(\boldsymbol{\xi}) = \mathbb{E}\{\mathbf{x}\} = (\mathbf{d} \odot \mathbf{g}) \mathbf{u} \quad (2.25)$$

Finding the maximum-likelihood estimate for the parameters $\hat{\boldsymbol{\xi}}$ is performed by solving an optimization problem with cost given by

$$\Lambda(\boldsymbol{\xi}) \triangleq (\mathbf{x} - \mathbf{h}\mathbf{u})^H \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} (\mathbf{x} - \mathbf{h}\mathbf{u}) \quad (2.26)$$

where $\mathbf{h} \triangleq (\mathbf{d} \odot \mathbf{g})$. Since the mean function $\boldsymbol{\mu}(\boldsymbol{\xi})$ is linear with respect to the measurement amplitudes \mathbf{u} , we can formulate the maximum-likelihood solution for those parameters in closed-form as

$$\hat{\mathbf{u}}_{\text{ML}} = (\mathbf{h}^H \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} \mathbf{h})^{-1} \mathbf{h}^H \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} \mathbf{x} \quad (2.27)$$

Substituting this back into (2.26) yields a new cost function in terms of the other parameters. Eliminating terms which are independent of the parameters $\boldsymbol{\xi}$ and simplifying yields the reduced cost function

$$[\Lambda(\boldsymbol{\xi})]_{\mathbf{u}=\hat{\mathbf{u}}_{\text{ML}}} = \mathbf{x}^H \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} \mathbf{h} (\mathbf{h}^H \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} \mathbf{h})^{-1} \mathbf{h}^H \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} \mathbf{x}, \quad (2.28)$$

to be maximized in this case. Recalling the definition of the basis function \mathbf{d} given in (2.4), we see that, with respect to the parameters in $\boldsymbol{\psi}$ and $\boldsymbol{\tau}_0$, the cost function given in (2.28) is both nonlinear and nonconvex. Maximizing this function must be done numerically rather than analytically using

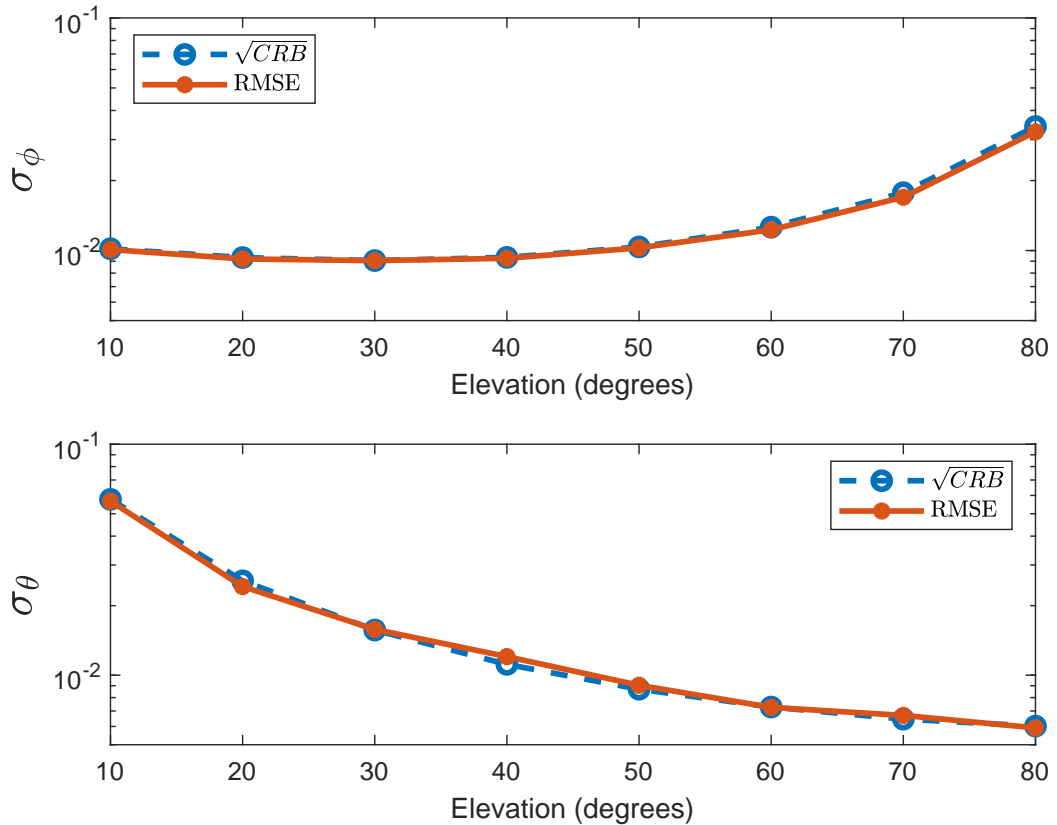


Figure 2.2: RMSE for 50° Azimuth, various Elevation angles (1000 Monte Carlo averages)

a nonlinear optimization algorithm. Figure 2.2 shows the results of applying one such algorithm, the Nelder-Mead simplex method [28], to the cost function found in (2.4) for a known, uniform spatio-temporal covariance matrix of the form $\Sigma_{\mathbf{x}} = \sigma^2 \mathbf{I}$. One-thousand trials were performed for different ground-truth directions of arrival and a noise variance of $\sigma = 10^{-1}$. The root mean-squared error (RMSE) of these trials was computed and is shown in comparison to the results of the CRB computation described in 2.2.2 for a ground-truth azimuth of 50° and varying elevation.

2.4 Maneuver Optimization via CRB Minimization

We are interested in finding the set of sensor positions $\mathbf{p}(t)$ for a given observation interval T_{SA} which will yield the best possible estimate of the parameters of interest in ξ . To do this, we need to find the set of k positions which minimize the Cramér-Rao lower bound defined in 2.2.2. In particular, we are interested in minimizing the mean-squared error of our DOA estimate which can

be done by minimizing the trace of the CRB submatrix corresponding to ψ :

$$\hat{\mathbf{p}} = \arg \min_{\mathbf{p}(t_k), \forall t_k \in T_{SA}} \sum_i [\mathbf{J}(\boldsymbol{\xi})]_{ii}^{-1}, \forall \xi_i \in \psi \quad (2.29)$$

With $\hat{\mathbf{p}} \in \mathbb{R}^{3 \times k}$, estimating this set of positions is infeasible for even modest sample rates and observation intervals, and would yield a solution which would not be useful in controlling the sensor platform. Instead, we need to apply a more compact and tractable characterization of the position of the aircraft during the collection period.

2.5 Maneuver Parametrization

To formulate our problem in a way that is independent of the sampling rate and observation interval, we need to identify a parametrization which yields a continuous function of position in terms of a handful of tunable design variables. This function defines a trajectory along which the sensor will travel during the observation interval. This can be done in a number of different ways, but in the case of synthetic aperture mounted on an aerial vehicle, it is appropriate to restrict ourselves to the set of circular trajectories incident to the current location of the vehicle. This yields the following characterization

$$\mathbf{p}(t) = f(t; \rho, \alpha, \beta, \gamma, \mathbf{p}_0, \mathbf{v}_0) \quad (2.30)$$

where ρ is the radius of the maneuver, \mathbf{p}_0 and \mathbf{v}_0 are the initial position and velocity of the vehicle at the start of the maneuver, and α, β , and γ are the yaw, pitch and roll of the maneuver with respect to the initial position of the receiver, in the global frame of reference. By adjusting these parameters it is possible to form any complete or fractional circular trajectory incident to the current location of the receiver. Utilizing this parametrization for the position of the sensor into the formulation for the CRB given in 2.2.2 yields the following parametrization for the relative delay in (3.32) as a function of time

$$\tau_{\text{rel}}(t) = \frac{\hat{\mathbf{r}}_{\psi}^{\top}}{c} f(t; \rho, \alpha, \beta, \gamma, \mathbf{p}_0, \mathbf{v}_0) \quad (2.31)$$

Finally, this delay can be used to compute the value of \mathbf{d} that is used throughout the CRB computation.

2.6 Optimization Methods

The objective function to be minimized is the CRB described in 2.2.2 parameterized by the sample positions computed using (2.30). This nested complexity and lack of smoothness excludes

the possibility of minimization using gradient-based optimization methods, so we must instead rely on *direct search methods* which require evaluating only the objective function at selected points. In our case, evaluating the objective function means computing the Cramér-Rao lower bound for a given set of trajectory parameters. To do so we must simulate the action of the sensor and signal for the entire duration of T_{SA} . Due to this computational cost, it is important that we minimize the number of evaluations required to reach a suitable global minimum. We consider two such methods for performing this direct search optimization: an algorithm based on generalized pattern search (GPS) using an adaptive mesh [29], and a Bayesian optimization method utilizing an expected improvement acquisition function [30].

2.6.1 Generalized Pattern Search

Generalized pattern search (GPS) describes a family of iterative algorithms for derivative-free unconstrained or linearly constrained optimization. These methods work by evaluating the objective function at a fixed set of points, called a mesh, which are located at a given distance in different directions away from the currently estimated minimum. This process is repeated for several iterations, with the mesh distance being adjusted depending on whether or not the method was successful in identifying an improved minimum in the previous iteration. When an iteration identifies a new estimated minimum the mesh expands by a predetermined factor, otherwise it contracts. Application of linear constraints in the GPS algorithm is done by application of a barrier function of the form [31]

$$f_{\Omega}(\mathbf{x}) = \begin{cases} f(\mathbf{x}), & \text{if } \mathbf{x} \in \Omega \\ \infty, & \text{otherwise} \end{cases} \quad (2.32)$$

where Ω is the feasible set for \mathbf{x} . In this way, values which fall outside of the constraints are functionally discarded and the method proceeds with the remaining points in the mesh. The process of polling the objective function and updating the mesh repeats until a predetermined number of iterations have occurred, or the size of the mesh has reached a predetermined minimum. Although these methods can be applied to problems for which the objective function is non-differentiable or even non-continuous, they are not guaranteed to converge to a *global* minimum.

2.6.2 Bayesian Optimization

Bayesian optimization is another black box approach to globally optimizing objective functions, particularly those which are difficult or computationally expensive to evaluate or for which the objective function is not available in a closed form. It is similar to the pattern search algorithm in that it seeks to identify a global minimum in as few objective evaluations as possible, however the way in which it approaches this problem is fundamentally different. Rather than immediately searching the objective function for increasingly smaller values of the objective function, Bayesian optimization operates by constructing a probabilistic model over the feasible domain of the objective function. As observations are made by evaluating the objective function at different points, the posterior is updated and the model iteratively becomes a more refined representation of the true objective function [30].

Determining where to sample the objective function is done by invoking an acquisition function which quantifies the “utility” of a candidate point, taking into account the uncertainty in the probabilistic model and trading off between *exploration* and *exploitation*. Here, “exploration” refers to sampling the objective at points where the model uncertainty is high while, in the context of objective *minimization*, while “exploitation” refers to sampling the objective at points where the model expectation is low. By balancing these properties Bayesian optimization avoids becoming trapped in local minima, and in applications where it’s useful can be exploited to find multiple local or global minima. For our application we selected the *expected improvement* acquisition function, which has the form

$$EI(\mathbf{x}, Q) = \mathbb{E}[\max(0, \mu_Q(\mathbf{x}_{\text{best}}) - f(\mathbf{x}))]. \quad (2.33)$$

where $f(x)$ is the objective function, Q is the Gaussian process posterior, and \mathbf{x}_{best} and $\mu_Q(\mathbf{x}_{\text{best}})$ are the location and value of its minimum. At each iteration the posterior distribution of the model is updated using samples of the objective function which correspond to the maximum of this function as computed in the previous step, and a new maximum is computed for the next iteration. This process is repeated for a fixed number of iterations corresponding to the total number of allotted objective function evaluations.

2.7 Experimental Validation for Optimizing SA-DOA Trajectories

To compare the two methods for CRB optimization under the maneuver model described in 2.5, we performed simulated experiments assuming known signal waveform $s(t)$ of a form consistent

with a narrowband continuous-wave (CW) jammer modulated to a frequency appropriate for jamming GPS L1 (1575.42 MHz). Since the CRB for a uniform circular array is agnostic to relative azimuth, received signals were generated for an azimuth of 50° above the xz -plane and elevations varied from $10^\circ - 90^\circ$ above the xy -plane in increments of 10° .

For each experiment, the initial direction-of-arrival was first computed using maximum likelihood estimation and an arbitrary trajectory with a radius of 2 meters in the xy -plane, executed at a vehicle speed of 2π m/s (π rads/sec). This parameters given by this maximum likelihood result were then used to form the model for performing CRB optimization using both the pattern search and Bayesian optimization methods. Each method was initialized with the flat, arbitrary trajectory used to perform initial estimation and executed for a total of one-hundred function evaluations. After each iteration, the minimum objective and total number of function calls for each method was recorded. In the case of Bayesian optimization, additional random initialization points are selected each time the method is executed. In order to determine mean performance the method was called fifty times for each experiment and the results averaged.

Figures 2.3 and 2.4 show the progression of estimated minima for the pattern search and Bayes algorithms respectively as a function the number of objective evaluations. Both methods were successful in identifying circular trajectories which perform better than the initial, flat trajectory. In comparison to each other the two methods performed similarly on average, with the GPS methods performing slightly better on average for evaluation counts lower than ten. This is consistent with the approach employed by each method: while the GPS methods are designed to rapidly converge on a local minimum, the Bayesian methods aim to strike a balance between searching for the minimum and mapping the function over the feasible domain.

2.8 Conclusions

In this chapter, we focused on applying bounds analysis to improving estimator performance in the matched case. Specifically, we proposed a method to minimize estimation error in unbiased estimators by identifying tunable parameters that minimize the corresponding Cramér-Rao lower bound (CRB). We applied this method to the problem of synthetic aperture direction of arrival (SA-DOA) estimation in the context of GNSS jammer localization, and showed that this approach can be used to design trajectories for performing DOA estimation using an antenna mounted on rotor wing aircraft. Two different optimization strategies were applied to solving this minimization problem: one using generalized pattern search (GPS) and another using Bayesian optimization with

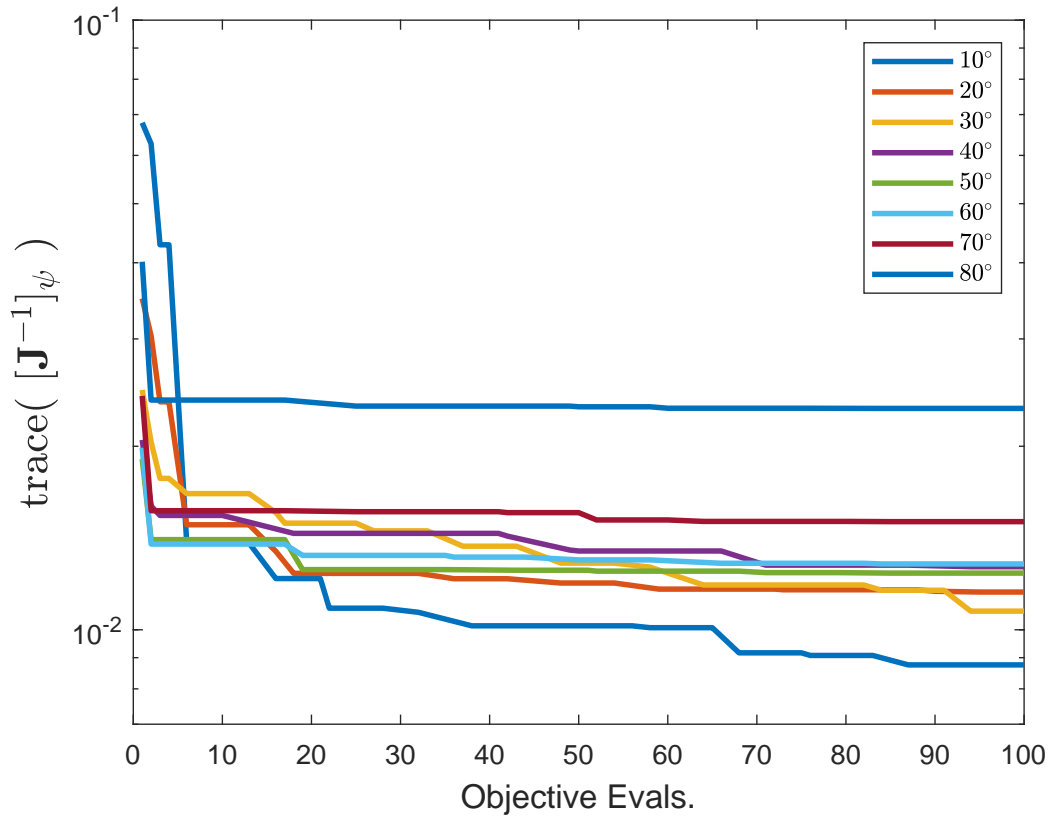


Figure 2.3: DOA-Trace Objective function minimum using GPS optimization. Ground truth of 50° Azimuth, various Elevation angles.

an expected improvement acquisition function. By applying these methods in simulations of the GNSS jammer localization scenario, we determined that both methods perform similarly for the given scenarios. We conclude that for the particular problem of optimizing DOA estimation in GNSS synthetic aperture, there is not a substantial benefit to applying more computationally complex Bayesian optimization methods where more efficient pattern search methods will suffice.

These results are significant as they demonstrate the utility of applying estimation bounds analysis to refining estimators and enhancing estimator performance. This shows that the theory described in this dissertation has the potential to be adapted into an actionable strategy for improving estimator performance, as demonstrated here in the matched case. Future work in this area should focus on identifying places where similar analyses can be applied to *misspecified* bounds analysis to achieve improved performance with respect to other key metrics.

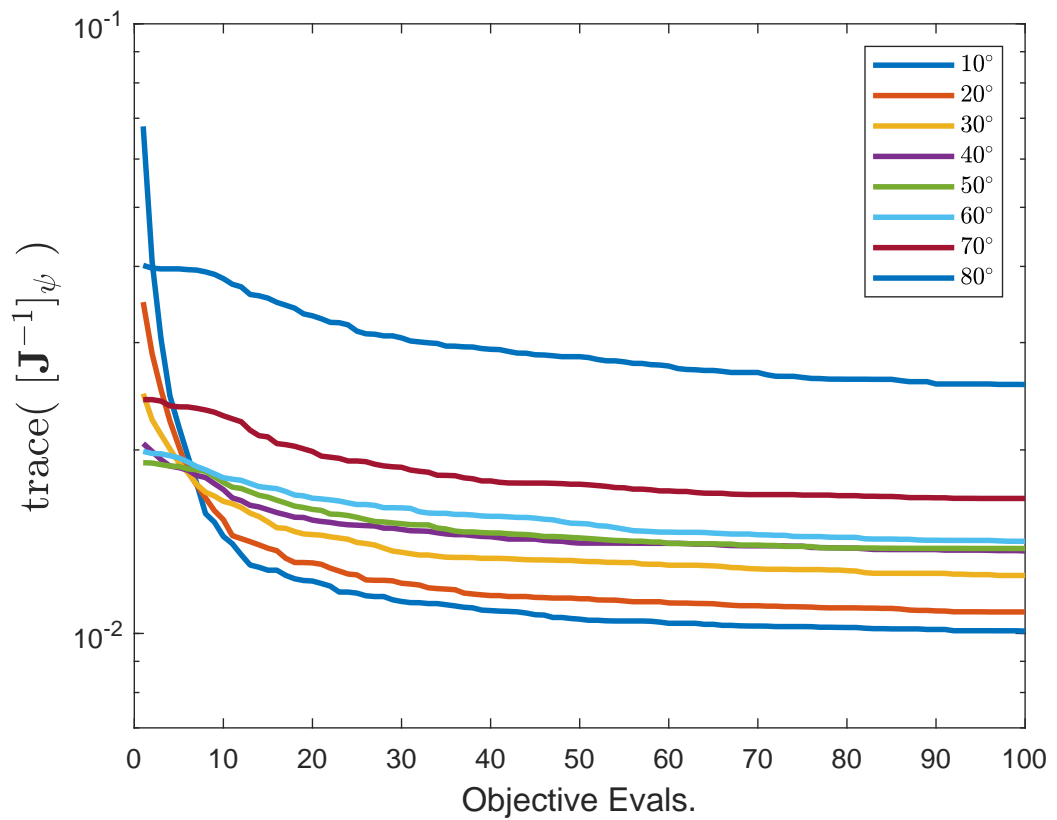


Figure 2.4: DOA-Trace Objective function minimum using Bayesian optimization. Ground truth of 50° Azimuth, various Elevation angles.

Chapter 3

Misspecified MUSIC and Maximum-Likelihood Estimation for Multi-Sensor Array Estimation

Radio direction finding (RDF) is vital in communication and defense systems for estimating the direction of arrival (DOA) of radio sources. This process utilizes the spatiotemporal characteristics of a receiver to gather multiple observations. Various receiver configurations, including dynamic single-antenna and multi-antenna arrays, have been employed for radio direction finding. Multiple Signal Classification (MUSIC) has gained popularity as it can estimate DOA or frequency for diverse, uncorrelated signals using eigenspace methods. In this chapter, we utilize Gaussian stochastic models to explore scenarios of model misspecification, wherein one or more assumptions in estimator design are violated. We demonstrate that the MUSIC algorithm performs comparably to classical parametric methods like maximum-likelihood estimation (MLE) when facing small deviations from the assumed model. We present comparisons of both methods with the misspecified Cramér-Rao lower bound (MCRB), indicating that both estimation approaches are asymptotically optimal and unbiased within the discussed class of misspecification.

3.1 Array Estimation: Applications and Challenges

The ubiquity of radio transmissions in both civilian and military communications has led to decades of research into understanding the impact that the spatial characteristics of the real world have on the propagation of radio signals. In various civil and defense applications, such as tracking unmanned aerial vehicles or monitoring radio frequency emissions, it is necessary to

obtain or estimate the DOA of radio signals. Under adversarial or uncooperative conditions, or in environments in which active radar localization is infeasible, approaches based on passive observation at one or more known locations may be preferable. RDF is a field dedicated to this purpose, and throughout its lifetime numerous receiver configurations have been demonstrated to be effective for this purpose, including various dynamic single-antenna systems [27, 32, 33] and multi-antenna arrayed systems [24, 34]. Using the local spatiotemporal characteristics of a given receiver, it is possible to provide multiple simultaneous “looks” at one or more nearby radio sources and use time difference of arrival (TDoA) methodology to estimate direction of arrival [35, 36].

One of the key challenges in radio direction finding is to efficiently process simultaneous or overlapping observations originating from multiple sources or radio reflectors. This is particularly challenging when faced with incomplete or incorrect information about the structure or nature of the observed signals. To address this challenge, several estimation techniques have been proposed, among which Multiple Signal Classification (MUSIC) has been widely employed. As an eigenspace method, MUSIC is non-parametric with respect to the distribution of the observed signal, but does require that the observed signals are mutually uncorrelated. This approach alleviates some of the issues associated with handling multiple simultaneously observed signals, including the so-called curse of dimensionality associated with parametric models such as MLE.

In this chapter, we investigate the performance of DOA estimation algorithms in the presence of model misspecification, where one or more of the assumptions in estimator design are violated [37]. In particular, the assumed model corresponds to a widely used class of Gaussian models, while we discuss how a more general class of Gaussian models characterize a variety of practical antenna array mismatches. The equivalence between MLE and MUSIC under certain modeling conditions is known when models are Gaussian, and the parameters of the output distributions are correctly known. We additionally show that under the general class of Gaussian model mismatches of interest, the performance of mismatched MUSIC is comparable to that of classical statistical inference methods such as mismatched MLE (sometimes refer to as quasi-MLE [38]). To further evaluate the efficacy of both estimation approaches, we compare them to the misspecified Cramér-Rao lower bound (MCRB), which is an established tool for understanding the performance characteristics of misspecified systems [39, 40]. This analysis reveals that mismatched MUSIC and MLE are both asymptotically optimal and unbiased within certain classes of misspecification [41]. This approach aims to provide valuable insights into the robustness of the MUSIC algorithm under various conditions and its potential applicability in real-world radio direction finding scenarios. We investigate two relevant mismatch scenarios, namely the effect of cross-channel mutual coupling

and estimation distortions due to antenna elements misspecified locations. Such an analysis can be leveraged in understanding and improving tuning of DOA and RDF systems with the aim of improving robustness under misspecified conditions [3].

3.2 Signal Models

We are interested in the classical problem of estimating the bearing from a multi-channel, multi-antenna radio receiver to a number of signal sources broadcasting from multiple locations [36]. In particular, we consider the realistic case where the assumed signal model does not entirely capture the actual signal propagation interaction. This section presents an assumed signal model (\mathcal{M}), which is generally accepted in the literature. This assumed signal model is mismatched to the *true* signal model (\mathcal{M}_*). We discuss that \mathcal{M}_* can be seen as a generalization of \mathcal{M} that can account for a variety of antenna array impairments such as mutual coupling or sensor misplacement, as discussed in more detail in Section 3.5. This model is taken from the canonical stochastic maximum-likelihood (SML) model, and is proposed as the baseline “assumed model” for multiple reasons. First, in contrast to the deterministic maximum-likelihood (DML) model, the explicit structure of the observed signal is not required, only the statistical characterization of the observations. In this way, the statistical model can be applied to adversarial scenarios where the structure of the incoming signals is variable or cannot be known. Second, estimators based on the DML model have been shown *not* to attend the Cramér-Rao lower bound as the number of observations goes to infinity. This is because the number of unknowns scales with the number of observations leading to a situation where the Cramér-Rao lower bound can no longer be used to predict estimator performance.

Consider the impingement of signals from N stochastic narrowband signal sources located at angles $\boldsymbol{\theta} = [\boldsymbol{\theta}_1 \dots \boldsymbol{\theta}_N] \in \mathbb{R}^{2 \times N}$ with respect to a reference position associated with an antenna array of M sensors. Each $\boldsymbol{\theta}_n \in \mathbb{R}^{2 \times 1}$ is further composed of azimuth ψ and elevation θ as $\boldsymbol{\theta}_n = [\psi_n, \theta_n]^\top$. Under the assumed model, the observation vector for the k^{th} snapshot is modeled as

$$\mathcal{M} : \mathbf{y}_k = \mathbf{A}(\boldsymbol{\theta})\mathbf{x}_k + \mathbf{w}_k, \quad k = 1, 2, \dots, K \quad (3.1)$$

such that $\mathbf{y}_{1:K} = [\mathbf{y}_1 \dots \mathbf{y}_K] \in \mathbb{C}^{M \times K}$. The vector of signal amplitudes \mathbf{x}_k is assumed to be a complex circularly stationary Gaussian random vector with zero mean and covariance $\boldsymbol{\Sigma}_x$. The additive noise random term \mathbf{w}_k is also assumed to be a circularly stationary complex Gaussian random variable with zero mean and covariance given by $\boldsymbol{\Sigma}_w = \sigma_w^2 \mathbf{I}$. Signal values are assumed to be observed through the basis-function matrix $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\boldsymbol{\theta}_1) \dots \mathbf{a}(\boldsymbol{\theta}_N)] \in \mathbb{C}^{M \times N}$, which is

formed by evaluating the array manifold function at the corresponding DOA for the n^{th} signal, $\mathbf{a}(\boldsymbol{\theta}_n)$. As a consequence, the observation vector \mathbf{y}_k also comes from a circularly symmetric complex normal distribution $\mathbf{y}_k|\mathbf{x}_k \sim \mathcal{CN}(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y)$ [42]. The moments of this distribution are then related to the parameter vector $\boldsymbol{\theta}$ by means of the basis-function matrix such that

$$\boldsymbol{\mu}_y = \mathbb{E}\{\mathbf{y}_k\} = \mathbf{0} \quad (3.2)$$

$$\boldsymbol{\Sigma}_y = \mathbb{E}\{\mathbf{y}_k\mathbf{y}_k^H\} = \sigma_x^2 \mathbf{A}(\boldsymbol{\theta})\boldsymbol{\Sigma}_x\mathbf{A}^H(\boldsymbol{\theta}) + \sigma_w^2 \mathbf{I}. \quad (3.3)$$

This assumed model \mathcal{M} is an incomplete or incorrect characterization of the true distribution of observations \mathbf{y}_k . In general, it is not possible or necessary to explicitly identify the entirety of the *true* model, \mathcal{M}_* . We will, however, make some assumptions about its structure which are broadly applicable to the problem of multi-sensor DOA estimation. This can be represented using a ground-truth model of the form

$$\mathcal{M}_* : \mathbf{y}_k = \mathbf{H}(\boldsymbol{\theta})\mathbf{x}_k + \mathbf{n}_k, \quad k = 1, 2, \dots, K \quad (3.4)$$

where the true basis-function matrix $\mathbf{H}(\boldsymbol{\theta}) = [\mathbf{h}(\boldsymbol{\theta}_1) \dots \mathbf{h}(\boldsymbol{\theta}_N)] \in \mathbb{C}^{M \times N}$ accounts for the nominal steering of the signals, as well as any effects not modeled in the assumed model $\mathbf{A}(\boldsymbol{\theta})$. Section 3.5 provides a discussion of the composition of $\mathbf{H}(\boldsymbol{\theta})$ and its relation to $\mathbf{A}(\boldsymbol{\theta})$ for a set of relevant applications in array signal processing. Furthermore, the signal covariance matrix for $\mathbf{x}_k \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_x)$ can be potentially different from the assumed model, thus accounting for covariance mismatches. This results in an observation vector that also follows a circularly symmetric complex normal distribution $\mathbf{y}_k|\mathbf{x}_k \sim \mathcal{CN}(\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_*)$ with moments

$$\boldsymbol{\mu}_* = \mathbb{E}\{\mathbf{y}_k\} = \mathbf{0} \quad (3.5)$$

$$\boldsymbol{\Sigma}_* = \mathbb{E}\{\mathbf{y}_k\mathbf{y}_k^H\} = \mathbf{H}(\boldsymbol{\theta})\boldsymbol{\Sigma}_x\mathbf{H}^H(\boldsymbol{\theta}) + \sigma_w^2 \mathbf{I}. \quad (3.6)$$

The structure of these two models is similar, but allows for several important differences stemming from assumptions made about the composition of the underlying signal \mathbf{x}_k and its relationship to the signal direction of arrival given by $\mathbf{H}(\boldsymbol{\theta})$ and $\mathbf{A}(\boldsymbol{\theta})$ respectively. In the following sections, these differences and the effect of incorrectly made assumptions on estimator performance will be explored in this contribution by way of the MCRB.

3.3 Maximum-Likelihood Estimation and Model Misspecification

This section provides expressions for the MLE estimators under the (true and assumed) models in Section 3.2. Then, a discussion on the pseudotrue parameter is given for the type of model

mismatch considered in this chapter. Finally, the section derives the MCRB, the lowest achievable error bound under model misspecification that the mismatched ML estimator is known to attain.

3.3.1 Optimal Estimators

In this contribution, we consider an estimator to be “optimal” in the sense that it minimizes the minimum mean-squared error (MMSE) for a given set of observations and associated model. This class of estimator minimizes the average of the squared residual between the estimate provided by the estimator and the actual value or, as we will discuss later in this section, between the estimate and the “pseudotrue” value in cases of model misspecification. As we are performing statistical inference with Gaussian models, the optimal estimator that follows from the Lehmann-Scheffé theorem is the maximum likelihood estimator (MLE). For the class of models under consideration, and as described in Section 3.2, the stochastic likelihood models \mathcal{M} and \mathcal{M}_\star follow a Gaussian form with distributions

$$\mathcal{M} : f(\mathbf{y}|\boldsymbol{\theta}) = \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_y) \quad (3.7)$$

$$\mathcal{M}_\star : p(\mathbf{y}) = \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_\star) . \quad (3.8)$$

Note that in these and following definitions we refer to the parameter $\boldsymbol{\theta}$ to denote the entire set of unknown model parameters, which include but are not necessarily limited to the directions of arrival $\boldsymbol{\theta}$. Applying maximum likelihood to these models results in the so-called stochastic maximum likelihood method, the canonical solution to which is given by optimizing with respect to the respective likelihood functions for the stochastic model [42]. Given in logarithmic form these are

$$\ln \mathcal{L}_{\mathcal{M}}(\mathbf{y}_{1:K}) = -K \ln |\boldsymbol{\Sigma}_y| - \sum_{k=1}^K \mathbf{y}_k^H \boldsymbol{\Sigma}_y^{-1} \mathbf{y}_k + \kappa \quad (3.9)$$

$$\ln \mathcal{L}_{\mathcal{M}_\star}(\mathbf{y}_{1:K}) = -K \ln |\boldsymbol{\Sigma}_\star| - \sum_{k=1}^K \mathbf{y}_k^H \boldsymbol{\Sigma}_\star^{-1} \mathbf{y}_k + \kappa \quad (3.10)$$

where K is the number of independent observations of \mathbf{y} and $\kappa = KM \ln(\pi)$ is a constant term not influencing in the optimization. The optimal maximum-likelihood estimation for both of these models is a function of the observed data and the variances thereof $\boldsymbol{\Sigma}_y$ and $\boldsymbol{\Sigma}_\star$ under each respective model [43]. That is,

$$\hat{\boldsymbol{\theta}}(\mathbf{y}_{1:K}; \boldsymbol{\Sigma}) = \arg \max_{\boldsymbol{\theta}} \left\{ -K \ln |\boldsymbol{\Sigma}| - \sum_{k=1}^K \mathbf{y}_k^H \boldsymbol{\Sigma}^{-1} \mathbf{y}_k \right\} \quad (3.11)$$

In particular, the expression can be further resolved for each of the two models as

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \left\{ \ln |\boldsymbol{\Sigma}_y| + \text{Tr}\{\boldsymbol{\Sigma}_y^{-1} \hat{\mathbf{R}}_K\} \right\} \quad (3.12)$$

$$\hat{\boldsymbol{\theta}}_0 = \arg \min_{\boldsymbol{\theta}} \left\{ \ln |\boldsymbol{\Sigma}_*| + \text{Tr}\{\boldsymbol{\Sigma}_*^{-1} \hat{\mathbf{R}}_K\} \right\} \quad (3.13)$$

where $\hat{\mathbf{R}}_K$ is the sample covariance given by

$$\hat{\mathbf{R}}_K = \frac{1}{K} \sum_{k=1}^K \mathbf{y}_k \mathbf{y}_k^H. \quad (3.14)$$

3.3.2 Subspace Estimation Methods for Multi-Sensor Estimation

For many applications, the computational cost and complexity of employing a full MLE strategy becomes prohibitive when considering other design constraints, such as power consumption or hardware weight, and alternative, more efficient approaches need to be explored. One class of estimators that has been shown to be quite powerful in this regard, particularly for array-based frequency and DOA estimation, is that of “eigenspace methods,” which are based on the eigen-decomposition of the sample covariance matrix as expressed in (3.14). Using the characteristics of the associated eigenvalues and eigenvectors it is possible to make an inference about the array manifold $\mathbf{a}(\boldsymbol{\theta})$.

Among the class of eigenspace methods, one algorithm that has seen widespread adoption and attention is Multiple Signal Classification (MUSIC), which will be used as the representative algorithm for this analysis. MUSIC leverages the orthogonality of the computed eigenvectors to identify a signal subspace \mathcal{U}_s and a noise subspace \mathcal{U}_n based on the magnitude of the associated eigenvalues. This method requires some knowledge about the number of signals present to determine the number of respective signal and noise eigenvalues, with signal eigenvalues taken to be the N greatest eigenvalues, and the remainder to be the noise eigenvalues. The associated eigenvectors span the signal and noise subspaces respectively. Since the true array manifold $\mathbf{a}(\boldsymbol{\theta}_*)$ must be orthogonal to the noise subspace, we can scan through candidate array manifold vectors to find an estimate of the correct parameter $\boldsymbol{\theta}$. Given the stochastic models of the forms given in (3.1) and (3.4) the MUSIC estimate may be found by maximizing the pseudo-spectrum

$$\hat{\boldsymbol{\theta}} = \arg \max \text{Tr}\{\mathbf{A}^H(\boldsymbol{\theta}) \mathbf{G} \mathbf{G}^H \mathbf{A}(\boldsymbol{\theta})\}^{-1} \quad (3.15)$$

where $\mathbf{G} \in \mathbb{C}^{M \times (M-N)}$ is the matrix of noise eigenvectors $\mathbf{G} \perp \mathcal{U}_s$ associated with the $M - N$ weakest eigenvalues.

3.3.3 Estimation under Model Misspecification

The performance of the DOA stochastic maximum-likelihood estimator under model misspecification is characterized by the misspecified mean and variance, which can be calculated using the so-called “pseudotrue” parameter and MCRB respectively [44]. The MCRB is a generalization of the classical Cramér-Rao lower bound widely used in evaluating “best case” estimator variance in numerous applications [45]. It can be shown that an optimal unbiased estimator will asymptotically adhere to the CRB as the signal-to-noise ratio goes to infinity. Similarly, it can be expected that an optimal misspecified unbiased (MS-unbiased) estimator will be bounded in performance by the MCRB. In this section we will provide definitions for both the pseudotrue parameter and misspecified Cramér-Rao lower bound in the context of the proposed DOA estimators, and discuss approaches to finding them for a given model misspecification.

3.3.3.1 Pseudotrue Parameter

We consider estimators for the assumed model MS-unbiased with respect to the “pseudotrue parameter” for the assumed and true models. In contrast to the standard Cramér-Rao lower bound (CRB) which provides a lower bound for the error between the true parameter θ_0 and the best possible estimate $\hat{\theta}_0$, the MCRB provides a lower bound for the error between this “pseudotrue” parameter $\tilde{\theta}_0$ and the best estimate $\hat{\theta}(\mathbf{y})$ that can be found using the misspecified estimator. More precisely, the MCRB bounds the error covariance $\mathbb{E}_p\{(\hat{\theta}(\mathbf{y}) - \tilde{\theta}_0)(\hat{\theta}(\mathbf{y}) - \tilde{\theta}_0)^\top\}$ of a MS-unbiased estimator such that $\mathbb{E}_p\{\hat{\theta}(\mathbf{y})\} = \tilde{\theta}_0$. In this way the pseudotrue parameter represents a type of optimum which is particular to both the true and assumed models together. This parameter is found by computing the unique interior point $\tilde{\theta}_0 \in \Theta$ that minimizes the KL divergence between the output distributions of the true and assumed models,

$$\tilde{\theta}_0 = \arg \min_{\theta \in \Theta} \mathcal{D}(p(\mathbf{y}) || f(\mathbf{y}|\theta)) , \quad (3.16)$$

where $p(\mathbf{y})$ and $f(\mathbf{y}|\theta)$ are the likelihood models as specified in (3.7) and (3.8). The parameter $\tilde{\theta}_0$ is a function of the true parameter θ , and represents the optimum to which our MS-unbiased estimator is shown to converge. The stochastic maximum likelihood estimator described in the DOA literature is known to adhere to the MS-unbiasedness property [46] and as such, finding the pseudotrue parameter for these models is then a matter of numerically solving the optimization given by (3.16). Applying the analytic form of the distributions, we find that the pseudotrue parameter for

complex-valued zero-mean Gaussian distributions is given by

$$\tilde{\boldsymbol{\theta}}_0 = \arg \min_{\boldsymbol{\theta}} \left\{ \ln \frac{|\boldsymbol{\Sigma}_y|}{|\boldsymbol{\Sigma}_0|} + \text{Tr} \{ \boldsymbol{\Sigma}_y^{-1} \boldsymbol{\Sigma}_0 \} \right\} \quad (3.17)$$

where $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_* |_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ is the true covariance of the observations. This term is no longer parameterized by $\boldsymbol{\theta}$, but rather a function of the true quantity $\boldsymbol{\theta}_0$.

3.3.3.2 Misspecified Cramér-Rao Lower Bound

Just as the regular Fisher Information Matrix is used to derive the Cramér-Rao lower bound for correctly specified models, the misspecified Cramér-Rao lower bound can be derived by the pair of matrices $\mathbf{A}_{\tilde{\boldsymbol{\theta}}_0}$ and $\mathbf{B}_{\tilde{\boldsymbol{\theta}}_0}$ which constitute the irregular Fisher Information matrices. The elements of these matrices are found from the partial derivatives of the log-likelihood in a similar fashion to the standard Fisher Information Matrix (FIM), however unlike the regular FIM wherein the elements can be computed from either the second partial derivatives or the product of the first partial derivatives, this equivalency no longer holds and the $\{i, j\}$ -elements are strictly defined as

$$[\mathbf{A}_{\tilde{\boldsymbol{\theta}}_0}]_{ij} \equiv [\mathbb{E}_p \{ \nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}}^{\top} \ln f(\mathbf{y}|\boldsymbol{\theta}) \}]_{ij} \quad (3.18)$$

$$[\mathbf{B}_{\tilde{\boldsymbol{\theta}}_0}]_{ij} \equiv [\mathbb{E}_p \{ \nabla_{\boldsymbol{\theta}} \ln f(\mathbf{y}|\boldsymbol{\theta}) \cdot \nabla_{\boldsymbol{\theta}}^{\top} \ln f(\mathbf{y}|\boldsymbol{\theta}) \}]_{ij} \quad (3.19)$$

The required partial derivatives follow from the expanded forms of the Gaussian likelihood distributions in (3.7)–(3.8) [47], and more specifically from the logarithmic forms given in (3.9)–(3.10). In order to compute $\mathbf{B}_{\tilde{\boldsymbol{\theta}}_0}$ we consider the expression $\mathbb{E}_p \left\{ \frac{\partial}{\partial \theta_i} \ln f(\mathbf{y}|\boldsymbol{\theta}) \frac{\partial}{\partial \theta_j} \ln f(\mathbf{y}|\boldsymbol{\theta}) \right\}$. The first partial derivative of (3.9) with respect to the i^{th} element of $\boldsymbol{\theta}$ is

$$\frac{\partial}{\partial \theta_i} \ln |\boldsymbol{\Sigma}_y| = \text{Tr} \{ \boldsymbol{\Sigma}_y^{-1} \frac{\partial \boldsymbol{\Sigma}_y}{\partial \theta_i} \} \quad (3.20)$$

$$\begin{aligned} \frac{\partial}{\partial \theta_i} \ln \text{Tr} \{ \boldsymbol{\Sigma}_y^{-1} \mathbf{Y} \} &= \text{Tr} \left\{ \frac{\partial \ln \text{Tr} \{ \boldsymbol{\Sigma}_y^{-1} \mathbf{Y} \}}{\partial \boldsymbol{\Sigma}_y} \frac{\partial \boldsymbol{\Sigma}_y}{\partial \theta_i} \right\} \\ &= \text{Tr} \left\{ \boldsymbol{\Sigma}_y^{-1} \mathbf{Y} \boldsymbol{\Sigma}_y^{-1} \frac{\partial \boldsymbol{\Sigma}_y}{\partial \theta_i} \right\} \\ &= \text{Tr} \left\{ \boldsymbol{\Sigma}_y^{-1} \frac{\partial \boldsymbol{\Sigma}_y}{\partial \theta_i} \boldsymbol{\Sigma}_y^{-1} \mathbf{Y} \right\} \end{aligned} \quad (3.21)$$

where $\mathbf{Y} = \mathbf{y}_k \mathbf{y}_k^H$. Combining these terms yields the entire first partial derivative,

$$\frac{\partial}{\partial \theta_i} \ln f(\mathbf{x}|\boldsymbol{\theta}) = \text{Tr} \left\{ \boldsymbol{\Sigma}_y^{-1} \frac{\partial \boldsymbol{\Sigma}_y}{\partial \theta_i} (\mathbf{I} - \boldsymbol{\Sigma}_y^{-1} \mathbf{Y}) \right\} \quad (3.22)$$

For $\mathbf{A}_{\tilde{\theta}_0}$ we consider the expression $\mathbb{E}_p \left\{ \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln f(\mathbf{x}|\boldsymbol{\theta}) \right\}$. The second partial derivative of (3.9) with respect to the i^{th} element of $\boldsymbol{\theta}$ follows from (3.22)

$$\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln f(\mathbf{x}|\boldsymbol{\theta}) = \text{Tr} \left\{ \frac{\partial \boldsymbol{\Sigma}_y}{\partial \theta_i} \boldsymbol{\Sigma}_y^{-1} \frac{\partial \boldsymbol{\Sigma}_y}{\partial \theta_i} + \left(\frac{\partial \boldsymbol{\Sigma}_y}{\partial \theta_j} \boldsymbol{\Sigma}_y^{-1} \frac{\partial \boldsymbol{\Sigma}_y}{\partial \theta_i} + \frac{\partial^2}{\partial \theta_i \partial \theta_j} \boldsymbol{\Sigma}_y \right) (\mathbf{I} - \boldsymbol{\Sigma}_y^{-1} \mathbf{Y}) \right\} \quad (3.23)$$

Applying the expectation with respect to the true distribution, the resulting expression for $[\mathbf{A}_{\tilde{\theta}_0}]$ is given by

$$\begin{aligned} \mathbb{E}_p \left\{ \frac{\partial^2 \ln f(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right\} &= \text{Tr} \left\{ -\boldsymbol{\Sigma}_y^{-1} \frac{\partial^2 \boldsymbol{\Sigma}_y}{\partial \theta_i \partial \theta_j} (\mathbf{I} - \boldsymbol{\Sigma}_y^{-1} \boldsymbol{\Sigma}_0) \right\} \\ &+ \text{Tr} \left\{ \boldsymbol{\Sigma}_y^{-1} \frac{\partial \boldsymbol{\Sigma}_y}{\partial \theta_i} \boldsymbol{\Sigma}_y^{-1} \frac{\partial \boldsymbol{\Sigma}_y}{\partial \theta_j} \boldsymbol{\Sigma}_y^{-1} \boldsymbol{\Sigma}_0 \right\} + \text{Tr} \left\{ \boldsymbol{\Sigma}_y^{-1} \frac{\partial \boldsymbol{\Sigma}_y}{\partial \theta_j} \boldsymbol{\Sigma}_y^{-1} \frac{\partial \boldsymbol{\Sigma}_y}{\partial \theta_i} (\mathbf{I} - \boldsymbol{\Sigma}_y^{-1} \boldsymbol{\Sigma}_0) \right\} \end{aligned} \quad (3.24)$$

while the expression for $[\mathbf{B}_{\tilde{\theta}_0}]$ is given by [48]

$$\begin{aligned} \mathbb{E}_p \left\{ \frac{\partial \ln f(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_i} \cdot \frac{\partial \ln f(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_j} \right\} &= \text{Tr} \left\{ \boldsymbol{\Sigma}_y^{-1} \frac{\partial \boldsymbol{\Sigma}_y}{\partial \theta_i} \right\} \text{Tr} \left\{ \boldsymbol{\Sigma}_y^{-1} \frac{\partial \boldsymbol{\Sigma}_y}{\partial \theta_j} \right\} \\ &+ \text{Tr} \left\{ \boldsymbol{\Sigma}_y^{-1} \frac{\partial \boldsymbol{\Sigma}_y}{\partial \theta_i} \right\} \text{Tr} \left\{ \mathbf{Q}_j \boldsymbol{\Sigma}_0 \right\} + \text{Tr} \left\{ \boldsymbol{\Sigma}_y^{-1} \frac{\partial \boldsymbol{\Sigma}_y}{\partial \theta_j} \right\} \text{Tr} \left\{ \mathbf{Q}_i \boldsymbol{\Sigma}_0 \right\} \\ &- \text{Tr} \left\{ \mathbf{Q}_i \boldsymbol{\Sigma}_0 \right\} \text{Tr} \left\{ \mathbf{Q}_j \boldsymbol{\Sigma}_0 \right\} + \text{Tr} \left\{ \mathbf{Q}_i \boldsymbol{\Sigma}_0 \mathbf{Q}_j \boldsymbol{\Sigma}_0 \right\} \end{aligned} \quad (3.25)$$

with

$$\mathbf{Q}_i = \boldsymbol{\Sigma}_y^{-1} \frac{\partial \boldsymbol{\Sigma}_y}{\partial \theta_i} \boldsymbol{\Sigma}_y^{-1}, \quad \mathbf{Q}_j = \boldsymbol{\Sigma}_y^{-1} \frac{\partial \boldsymbol{\Sigma}_y}{\partial \theta_j} \boldsymbol{\Sigma}_y^{-1} \quad (3.26)$$

Finally, the error bound can be computed for K independent snapshots as

$$\mathbb{E}_p \{ (\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}_0)(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}_0)^\top \} \geq \frac{1}{K} [\mathbf{A}_{\tilde{\theta}_0}^{-1} \mathbf{B}_{\tilde{\theta}_0} \mathbf{A}_{\tilde{\theta}_0}^{-1}], \quad (3.27)$$

where the inequality indicates semi-positive definiteness. Note that in the matched case, this simplifies to the canonical FIM given by the Slepian-Bangs formula in the absence of model misspecification

$$[\mathbf{J}_{\boldsymbol{\theta}}]_{ij} = K \text{Tr} \left\{ \frac{\partial \boldsymbol{\Sigma}_y}{\partial \theta_i} \boldsymbol{\Sigma}_y^{-1} \frac{\partial \boldsymbol{\Sigma}_y}{\partial \theta_j} \boldsymbol{\Sigma}_y \right\}. \quad (3.28)$$

3.4 Relationship between MUSIC and MLE under Model Misspecification

Multiple Signal Classification (MUSIC) is a vector signal processing technique based on eigenspace methodology. Used frequently in array signal processing for frequency estimation

and radio direction finding, MUSIC behaves as a large sample realization of stochastic maximum-likelihood [49]. It follows then, that as the MLE is asymptotically efficient, that is that under ideal conditions it attains the CRB, for large samples the MUSIC estimator is asymptotically efficient as well. This section provides analogous results to characterize the performance of MUSIC and MLE in the class of mismatched models described in Section 3.2. We refer to those as MMUSIC and MML, respectively. The results are organized in the form of a theorem, which states the conditions under which both estimators are asymptotically equivalent for the general case where $N \geq 1$ signals. The result is accompanied by a corollary, showing that for the single signal case, $N = 1$, the equivalence is unconditional on the relation between $\mathbf{h}(\boldsymbol{\theta})$ and $\mathbf{a}(\boldsymbol{\theta})$, the true and assumed array manifold vectors respectively.

Theorem 1. *Under the conditions and models described in Section 3.2, for cases of small errors between the true and assumed models, the MMUSIC estimator behaves as a large sample ($K \gg 0$) realization of the MML estimator, conditional on the following: If the relationship between the true and assumed models can be expressed or approximated by an additive error of the form $\mathbf{H} = \mathbf{A} + \mathbf{E}$, where \mathbf{E} may or may not be given as a function of $\boldsymbol{\theta}$, then the error must be such that the contribution to the ML cost function of the term $\text{Tr} \{ \mathbf{A}^H \mathbf{G} \mathbf{G}^H \mathbf{A} \}$ associated with the MMUSIC solution is much greater than the contribution to the MML solution due to the error. That is,*

$$\text{Tr} \{ \mathbf{A}^H \mathbf{G} \mathbf{G}^H \mathbf{A} \} \gg \text{Tr} \{ \Re \{ \mathbf{E}^H \boldsymbol{\Pi}_x \mathbf{G} \mathbf{G}^H \mathbf{A} \} \} + \text{Tr} \{ [\boldsymbol{\Pi}_x \mathbf{E}]^H \mathbf{G} \mathbf{G}^H [\boldsymbol{\Pi}_x \mathbf{E}] \} \quad (3.29)$$

where \mathbf{G} is the noise eigenvector matrix and $\boldsymbol{\Pi}_x = \mathbf{A} \mathbf{A}^\dagger$.

Proof. See Appendix 3.A. □

Corollary 1.1. *If there is exactly one signal of interest ($N = 1$) impinging upon the detecting array, that is, when the basis-function matrices $\mathbf{H}(\boldsymbol{\theta})$ and $\mathbf{A}(\boldsymbol{\theta})$ are both vector functions of $\boldsymbol{\theta}$, the additional model constraints described in Theorem 1 no longer apply, and the two estimators can be shown to converge in general terms.*

Proof. See Appendix 3.B. □

Theorem 2. *If the additive error term \mathbf{E} described in Theorem 1 may be described stochastically, with a structure resulting from a first-order perturbation of the array manifold function $\mathbf{a}(\boldsymbol{\theta})$ given by*

$$\mathbf{H}(\boldsymbol{\theta}) = \mathbf{A}(\boldsymbol{\theta}) + \mathbf{E} \quad (3.30)$$

where \mathbf{E} represents an additive circularly symmetric complex error term with zero mean and $\mathbb{E}\{\mathbf{E}\mathbf{E}^H\} = \sigma_e^2 \mathbf{I}$, and under the standard assumptions for MUSIC used in Theorem 1, the condition described in (3.29) may be reduced to a simple condition on the variance of the elements of the error terms

$$\sigma_e^2 \ll \frac{1}{N \text{Tr}\{(\mathbf{A}^H \mathbf{A})^{-1}\}} \quad (3.31)$$

Proof. See Appendix 3.C. □

3.5 Applications in Array Processing

Due to the versatility of statistical representations, the models described in Section 3.2 can be adapted to represent several different phenomena in the field of array signal processing. In this section, we provide two examples of applications where the results in this chapter are applicable. Namely, we discuss the mismatch cases where the sensor positions contain errors and the case where mutual coupling among array elements is not properly accounted for.

3.5.1 Sensor Position Error

One way that the models described in Section 3.2 may be applied is to account for errors in the specified spatial characteristics of the sensor array. The relative delay at a given array element is expressed in terms of the additional propagation path $\Delta\rho(t)$ of the impinging signal, defined as the projection of the position of the receiver in the local frame onto a unit vector $\hat{\mathbf{r}}_\theta$ pointing in the direction of the transmitter as [26]

$$\tau_{\text{rel}}(t) = \frac{\Delta\rho(t)}{c} = \frac{\hat{\mathbf{r}}_\theta^\top \mathbf{p}(t)}{c}. \quad (3.32)$$

Therefore, the relative delay is a function of the direction of arrival $\boldsymbol{\theta}$, which consists of the relative azimuth ψ and elevation θ of the transmitter from the receiver's perspective, and the position $\mathbf{p}(t)$ of the element in question. Error in the specified positions of the sensors is reflected in the signal model as an additive phase shift which is different for each channel of the array.

We consider a mismatch on the array manifold resulting from an additive error in the assumed positions of each array element. The standard model for the array manifold vector is [26]

$$\mathbf{a}(\boldsymbol{\theta}) = \exp\{j\mathbf{k}^\top(\boldsymbol{\theta})\mathbf{P}\} \quad (3.33)$$

which is a function of the matrix of antenna positions $\mathbf{P} \in \mathbb{R}^{3 \times M}$ the array steering vector $\mathbf{k}(\boldsymbol{\theta}) \in \mathbb{R}^{M \times 1}$ that is

$$\mathbf{k}(\boldsymbol{\theta}) = \frac{2\pi}{\lambda} \mathbf{u}(\boldsymbol{\theta}) \quad (3.34)$$

with $\mathbf{u}(\boldsymbol{\theta}) \in \mathbb{R}^{3 \times 1}$ the unit vector pointing in the direction of the signal source at angle $\boldsymbol{\theta}$. The concatenation of N instances of the array manifold vector form the basis-function matrix

$$\mathbf{A}(\boldsymbol{\theta}) = \exp\{j\mathbf{K}^\top(\boldsymbol{\theta})\mathbf{P}\} \quad (3.35)$$

where $\mathbf{K}(\boldsymbol{\theta}) \in \mathbb{R}^{3 \times N}$ is formed by concatenating the steering vectors $\mathbf{k}(\boldsymbol{\theta}_n)$. An additive error on either $\mathbf{K}(\boldsymbol{\theta})$ or \mathbf{P} can be represented as a complex, multiplicative error on $\mathbf{A}(\boldsymbol{\theta})$. Here we consider errors with regard to sensor position where the true position $\mathbf{P}_0 = \mathbf{P} + \boldsymbol{\Delta}_p$ is assumed to be \mathbf{P} , with the incremental term $\boldsymbol{\Delta}_p$ not accounted in \mathcal{M} , such that the true basis-function matrix may be represented as

$$\begin{aligned} \mathbf{H}(\boldsymbol{\theta}) &= \exp\{j\mathbf{K}^\top(\boldsymbol{\theta})(\mathbf{P} + \boldsymbol{\Delta}_p)\} \\ &= \exp\{j\mathbf{K}^\top(\boldsymbol{\theta})\mathbf{P}\} \odot \exp\{j\mathbf{K}^\top(\boldsymbol{\theta})\boldsymbol{\Delta}_p\} \end{aligned} \quad (3.36)$$

Applying a first-order Taylor approximation¹ to this product yields the small-error approximation

$$\begin{aligned} \mathbf{H}(\boldsymbol{\theta}) &\approx \mathbf{1} + j\mathbf{K}^\top(\boldsymbol{\theta})(\mathbf{P} + \boldsymbol{\Delta}_p) \\ &= \mathbf{1} + j\mathbf{K}^\top(\boldsymbol{\theta})\mathbf{P} + j\mathbf{K}^\top(\boldsymbol{\theta})\boldsymbol{\Delta}_p. \end{aligned} \quad (3.37)$$

Using a further first-order Taylor approximation of

$$\mathbf{A}(\boldsymbol{\theta}) = \exp\{j\mathbf{K}^\top(\boldsymbol{\theta})\mathbf{P}\} \approx \mathbf{1} + j\mathbf{K}^\top(\boldsymbol{\theta})\mathbf{P} \quad (3.38)$$

we obtain an approximation of $\mathbf{H}(\boldsymbol{\theta})$ for small values of $\boldsymbol{\Delta}_p$ given by

$$\mathbf{H}(\boldsymbol{\theta}) \approx \mathbf{A}(\boldsymbol{\theta}) + j\mathbf{K}^\top(\boldsymbol{\theta})\boldsymbol{\Delta}_p \quad (3.39)$$

yielding an additive small-error model of the form used in Section 3.4, with $\mathbf{E} = j\mathbf{K}^\top(\boldsymbol{\theta})\boldsymbol{\Delta}_p$.

3.5.2 Mutual Coupling Error

The class of models described in Section 3.2 can incorporate a statistical representation of the mutual coupling phenomenon present in many real-world array processing applications. Typical representations of mutual coupling in the literature utilize an approximation called the mutual

¹The Taylor expansion of e^x is given by $\sum_{k=1}^{\infty} 1 + x + \frac{x^2}{2} + \dots + \frac{x^k}{k!}$, with a first-order approximation given by truncating this series at the first-order term x .

coupling matrix (MCM) as a modification of the standard array manifold representation. This matrix quantifies the interactions between the channels of the array caused by various properties of the array, including certain characteristics of the antennas, inductive coupling of internal circuitry, etc.

The mutual coupling matrix is defined as a matrix of $M \times M$ coupling coefficients.

$$\mathbf{M} = \begin{bmatrix} m_{1,1} & m_{1,2} & \dots & m_{1,M} \\ \vdots & \vdots & \ddots & \vdots \\ m_{M,1} & m_{M,2} & \dots & m_{M,M} \end{bmatrix} \quad (3.40)$$

The coupled manifold is produced by modification of the uncoupled array manifold equations using the MCM as [50]

$$\mathbf{H} = \mathbf{M}\mathbf{A}(\boldsymbol{\theta}) = \begin{bmatrix} m_{1,1} & \dots & m_{1,M} \\ \vdots & \ddots & \vdots \\ m_{M,1} & \dots & m_{M,M} \end{bmatrix} \begin{bmatrix} a_1(\boldsymbol{\theta}) \\ \vdots \\ a_M(\boldsymbol{\theta}) \end{bmatrix} \quad (3.41)$$

This coupling representation can be easily applied to the observation model in (3.4)

$$\boldsymbol{\mu}_\star = \mathbb{E}\{\mathbf{y}_k\} = \mathbb{E}\{\mathbf{M}\mathbf{A}(\boldsymbol{\theta})\mathbf{x}_k\} = \mathbf{0} \quad (3.42)$$

$$\boldsymbol{\Sigma}_\star = \mathbb{E}\{\mathbf{y}_k\mathbf{y}_k^H\} = \sigma_x^2\mathbf{M}\mathbf{A}(\boldsymbol{\theta})\mathbf{A}^H(\boldsymbol{\theta})\mathbf{M}^H + \sigma_w^2\mathbf{I} \quad (3.43)$$

The effect of mutual coupling may be expressed as an additive error in a similar fashion as (3.37). Without loss of generality,

$$\mathbf{H} = \mathbf{M}\mathbf{A} = \mathbf{A} + (\mathbf{M} - \mathbf{I})\mathbf{A} \quad (3.44)$$

yielding an additive error model of the form used in Section 3.4, with $\mathbf{E} = (\mathbf{M} - \mathbf{I})\mathbf{A}(\boldsymbol{\theta})$.

3.6 Simulation results

The validation process for the misspecified Cramér-Rao lower bound was conducted via simulation of observations under two different conditions for validating the results of Corollary 1.1 for single-signal cases, and for validating the convergence results of Theorem 2 respectively. The basic approach is thus: data is generated according to a candidate ground-truth model, and the outputs of that model are used as the input for estimators derived according to the assumed model. This approach of simulating both “ground-truth” and misspecified models demonstrates one potential real-world application of mismatched modeling, specifically as a means of limit testing to evaluate the performance of a given system under different error conditions. In this case, those known failure

points are represented as cross-channel interference and imprecise array measurements, though they are by no means limited to these conditions. Through this perspective of limit testing, these results offer key insights into the system’s limitations, which are particularly relevant to the design of multi-antenna array systems for deployment in the field. The findings contribute to establishing hardware design tolerances, offering a more informed understanding of how to manage distortions.

For experiments validating the results of Corollary 1.1 with a single signal, experiments were performed using observations simulated according to an eight-antenna circular sensor array for 2-D (*i.e.* azimuth, elevation) directions-of-arrival. The observed signal was simulated according to (3.4) as an appropriately demodulated pseudorandom baseband signal, with a carrier frequency of $f_c = 150$ MHz and a signal-to-noise ratio of 9 dB, situated at a constant source DOA of $\boldsymbol{\theta} = [\psi, \theta]^\top = [50^\circ, 40^\circ]^\top$. Variations were applied to the model according to the application-specific modifications described in Section 3.5. Over the course of 250 Monte Carlo trials, each maximum likelihood and MUSIC estimator was tested using $K = 50000$ samples, which corresponds approximately to a sample interval of $T_s = 50$ ms. The assumed or “nominal” standard model, as defined in (3.1), was used to derive the MML and MMUSIC estimators, and the performance of each approach was evaluated based on the accumulated mean-squared error of the estimations with respect to the pseudotrue direction-of-arrival. Finally, comparisons were made between the estimation performance of model \mathcal{M} using both MMUSIC and MML estimators to the MCRB.

The simulations were designed to represent two error conditions: *i*) errors due to the mutual coupling model as per Friedlander et al. [50] with a range of coupling coefficients; and *ii*) sensor positioning errors corresponding to changes in the radius of the circular array, as shown in Fig. 3.1.

For experiments validating the results of Theorem 2, observations were simulated according to an eight-antenna linear sensor array for 1-D (*i.e.* azimuth only) directions-of-arrival. Two observed signals were simulated according to the same signal properties as for previous experiments, and with constant ground-truth source directions-of-arrival of $\boldsymbol{\theta} = [\psi_1, \psi_2] = [140^\circ, 20^\circ]$. Variations were applied to the model according to the additive error model described in Theorem 2 for various values of σ_e^2 and the error with respect to the misspecified maximum-likelihood estimation (MMLE) result averaged to obtain the *average* performance of the estimator under the additive noise conditions, with respect to the best unbiased misspecified estimator.

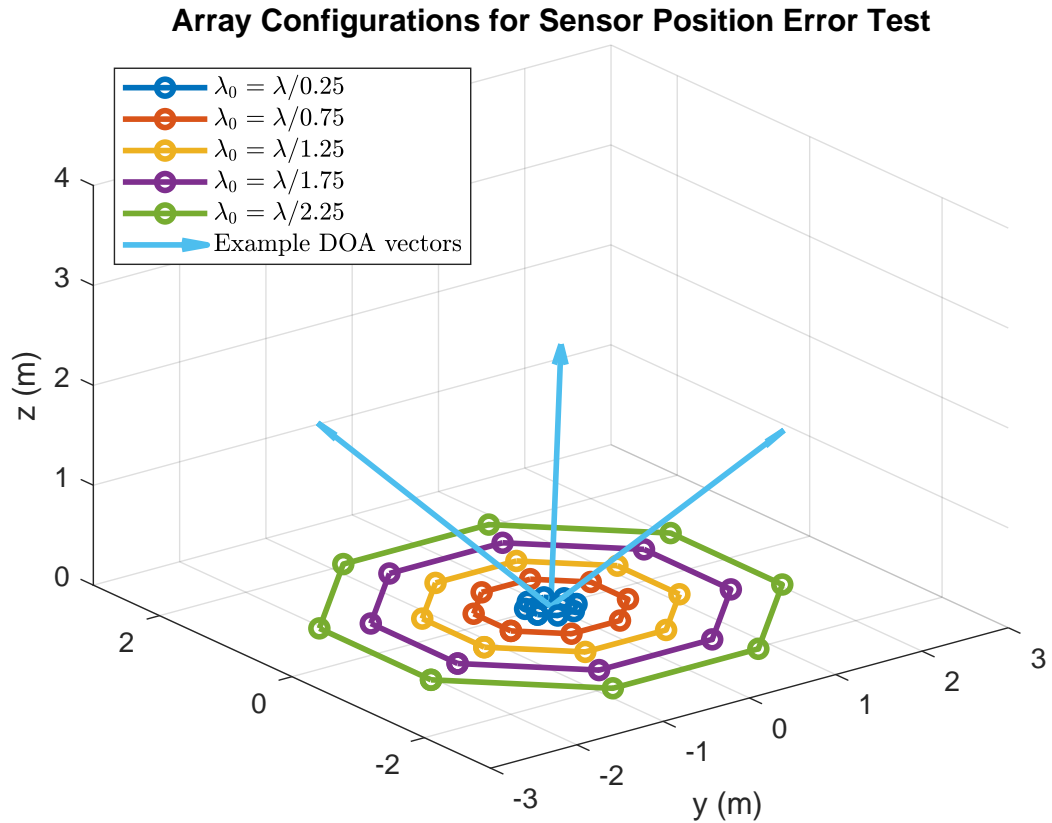


Figure 3.1: Array configurations for various spacing errors and various directions of arrival.

3.6.1 Single signal experiments

In regards to sensor displacement errors as described in 3.5.1 of Section 3.5, Fig. 3.2 depicts the results of the single-signal case using an eight-antenna circular sensor array. The root mean-squared error (RMSE) for the mismatched estimates is plotted with respect to the pseudotrue parameter, against the nearest-neighbor antenna spacing as a fraction of the signal wavelength. We note that, in the matched case, optimal performance is expected when the spacing is half of the signal wavelength, a result that is unsurprisingly not consistent in the mismatched case. Interestingly, we observed that the variance of the elevation pseudotrue error produced by the mismatched estimators in this case differed in behavior from that of the error in terms of azimuth, with pseudotrue error increasing with respect to elevation and decreasing with respect to azimuth as separation was increased. MML and MMUSIC estimates in these scenarios performed as expected, demonstrating consistency with the MCRB for their respective cases.

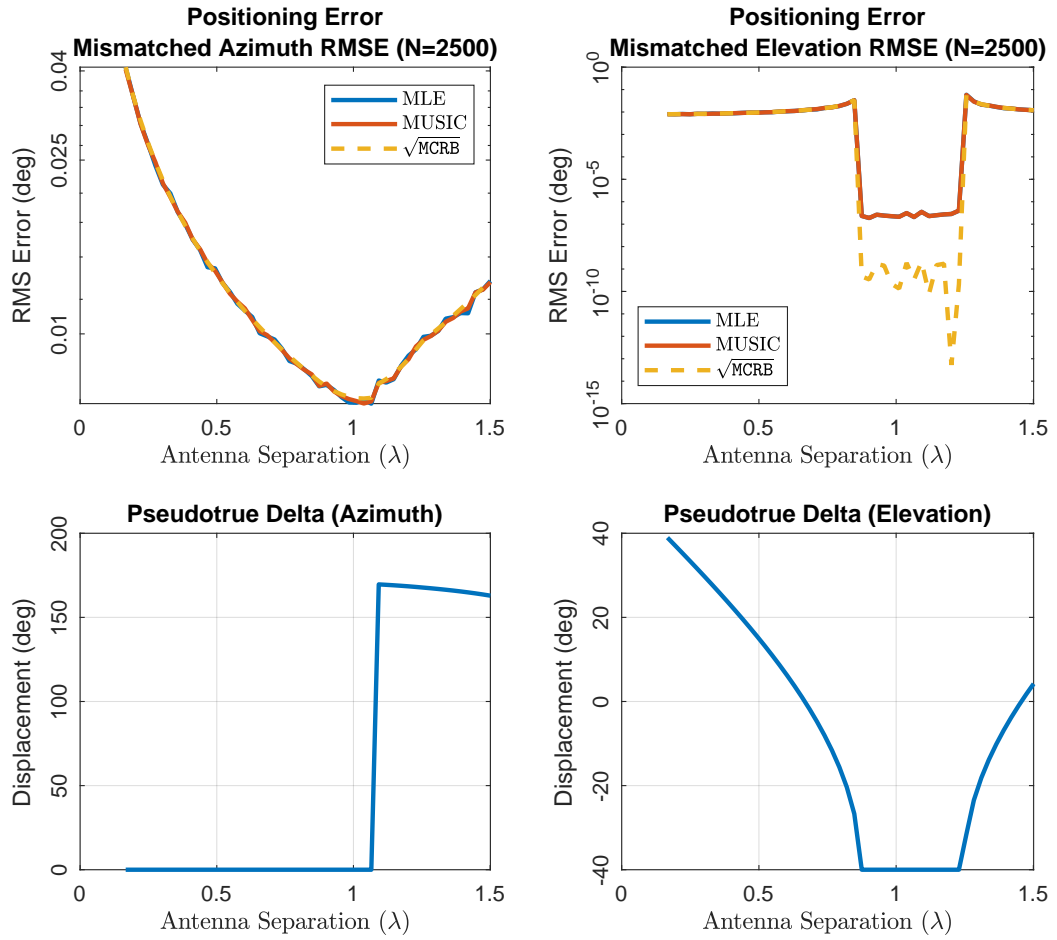


Figure 3.2: Estimator performance with respect to the pseudotrue parameter for position error misspecification.

Fig. 3.3 presents the estimation error for the mutual coupling model with respect to the pseudotrue parameter. The coupling coefficient is plotted against root mean-squared error (RMSE) and, as shown, both the maximum-likelihood and MUSIC methodologies reveal a consistent impact on model performance with increasing coupling coefficients. The decrease in performance is more notable in azimuth, while elevation performance appears to be somewhat more resilient. In both scenarios, the ML and MUSIC estimates were found to be strongly consistent with each other and with the MCRB. These findings support the findings in Corollary 1.1 regarding the influence of model misspecification on system performance with regard to both ML and MUSIC estimation in the single-signal case. While the nominal standard model produced results that were inconsistent with the real DOA in most scenarios, this inconsistency was accurately captured and represented

using the MCRB and its associated pseudotrue parameter.

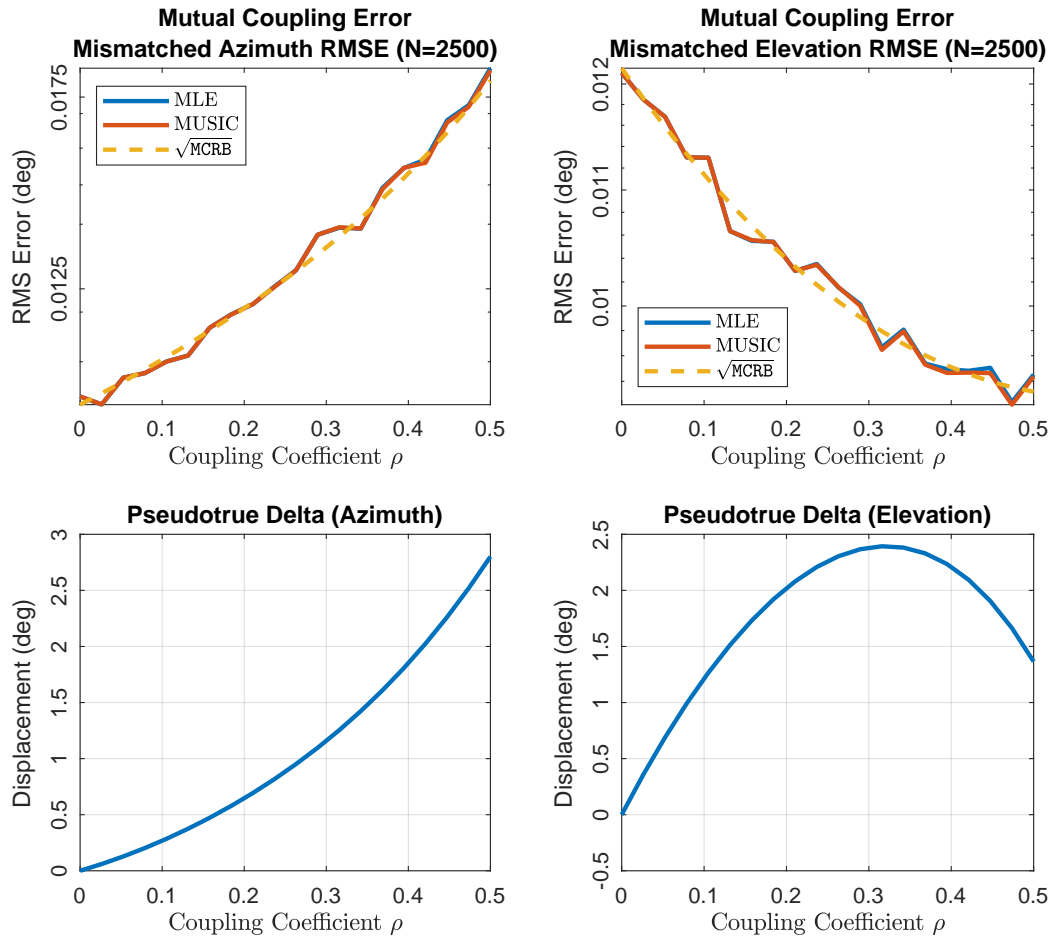


Figure 3.3: Misspecified estimator performance with respect to the pseudotrue parameter for mutual coupling model misspecification.

3.6.2 Multiple signals experiments

In regards to multiple signal experiments, Fig. 3.4 shows the offset between the MML and MMUSIC estimators as a function of the variance of the additive misspecification term \mathbf{E} . Note that the so-called “Objective Similarity Threshold” represents the point at which the term of the objective corresponding to the MMUSIC objective function begins to dominate the terms corresponding to the error term, which is unaccounted for in the pseudospectrum optimization of MMUSIC, as described in (3.31). As expected, for values of σ_e^2 that are at least an order of magnitude smaller than the objective similarity threshold, the MMUSIC and MML results converge, and subsequently diverge

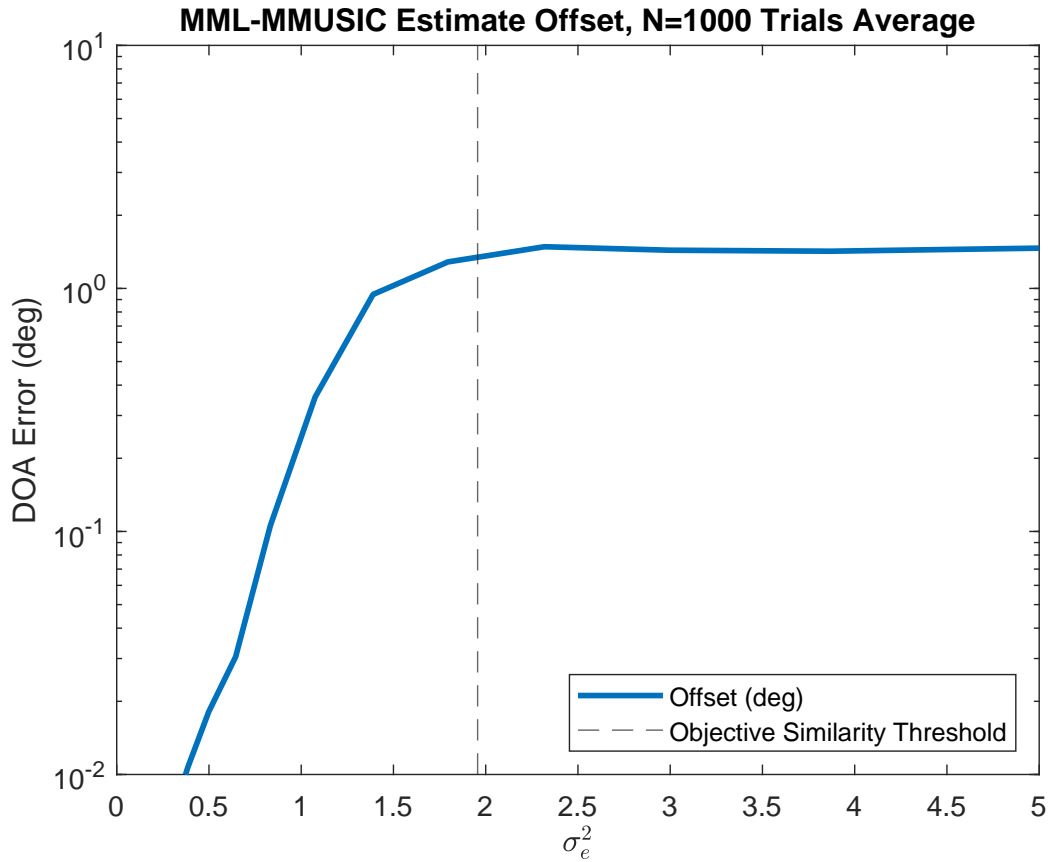


Figure 3.4: Difference between MML and MMUSIC estimator output with respect to the pseudotrue parameter for random additive misspecification.

as the power of \mathbf{E} continues to increase. This is consistent with the proposition given by Theorem 1 and Theorem 2 with respect to MMUSIC behaving as a large sample realization of MMLE in this small-error regime.

3.7 Conclusions

In this chapter, we investigated a class of misspecifications in the array manifold model for array-based direction finding applications. We demonstrated that under these classes of misspecification, and under several conditions described in 3.4, including those that are required for the application of the MUSIC algorithm, the behavior of the MMUSIC estimator is as a small-error realization of the MML estimator. These conditions are established in Theorem 1 and Theorem 2 for cases where multiple signals are being observed concurrently, while Corollary 1.1 establishes the

CHAPTER 3. MISSPECIFIED MUSIC AND MLE FOR MULTI-SENSOR ARRAYS

generality of the equivalence for cases where only one signal is being observed. These conditions were validated through simulation of uniform circular arrays for the single-signal case, and through simulation of uniform linear arrays in the multi-signal, stochastic error case. These results demonstrate the utility of deterministic and stochastic modeling of misspecifications in understanding the behavior of Gaussian systems. Although further research is required to generalize these findings to other classes of misspecification, the results of this chapter indicate a promising avenue for future exploration.

Appendix

3.A Proof of Theorem 1

The MML estimate $\hat{\boldsymbol{\theta}}(\mathbf{y}_{1:K})$ is defined as the value of $\boldsymbol{\theta}$ that results in maximization of the objective function

$$g(\boldsymbol{\theta}) = \text{Tr} \{ \mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \hat{\mathbf{R}}_K \} \quad (3.45)$$

Here $\hat{\mathbf{R}}_K$ is the sample covariance matrix defined in (3.14), the eigendecomposition of which is expressed as

$$\hat{\mathbf{R}}_K = \mathbf{S} \boldsymbol{\Lambda} \mathbf{S}^H + \mathbf{G} \boldsymbol{\Sigma} \mathbf{G}^H \quad (3.46)$$

where \mathbf{S} denotes the matrix containing the eigenvectors associated with the signal eigenvalues $\lambda_1, \dots, \lambda_N$ of the sample covariance matrix $\hat{\mathbf{R}}_K$. These eigenvalues form the diagonal matrix $\boldsymbol{\Lambda}$, which has dimensions $N \times N$. For the right side of the expression, \mathbf{G} represents the eigenvectors corresponding to the eigenvalues $\lambda_{N+1}, \dots, \lambda_M$, with $\boldsymbol{\Sigma}$ being a diagonal matrix with dimensions $M - N \times M - N$ and non-zero elements $\lambda_{N+1}, \dots, \lambda_M$. Applying (3.46) to the estimator function in (3.45) produces the following result.

$$\begin{aligned} g(\boldsymbol{\theta}) &= \text{Tr} \{ \mathbf{S}^H \mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{S} \boldsymbol{\Lambda} \} \\ &\quad + \text{Tr} \{ (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{G} \boldsymbol{\Sigma} \mathbf{G}^H \mathbf{A} \} \end{aligned} \quad (3.47)$$

Leveraging the orthogonality and independence properties of the noise components, this optimization simplifies to

$$g(\boldsymbol{\theta}) = \text{Tr} \{ [\mathbf{S} \mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{S} - \mathbf{I}] [\boldsymbol{\Lambda} - \sigma_w^2 \mathbf{I}] \} \quad (3.48)$$

Defining as in [49] $\mathbf{\Delta} = \mathbf{A} - \mathbf{S}\mathbf{S}^H\mathbf{A}$ provides useful simplifications of the estimator $g(\hat{\boldsymbol{\theta}})$. In particular,

$$\begin{aligned}\mathbf{\Delta}^H\mathbf{\Delta} &= \mathbf{A}^H\mathbf{A} - \mathbf{A}^H\mathbf{S}\mathbf{S}^H\mathbf{A} \\ &= \mathbf{A}^H\mathbf{G}\mathbf{G}^H\mathbf{A}\end{aligned}\quad (3.49)$$

With this notation we can write the *minimization* objective as

$$\text{Tr}\{\mathbf{\Delta}^H\mathbf{\Delta}(\mathbf{S}^H\mathbf{A})^{-1}[\mathbf{\Lambda} - \sigma_w^2\mathbf{I}](\mathbf{A}^H\mathbf{S})^{-1}\}\quad (3.50)$$

It is at this point that our derivation deviates somewhat from that for the standard maximum-likelihood and MUSIC solutions using correctly specified models. We noted earlier that the the optimal estimator for the measurement covariance is $\hat{\mathbf{R}}_K$. The assumptions associated with this estimator are only that the observations are zero mean, and as such it is unaffected by any other modelling errors. It follows that the sample covariance asymptotically approaches the true measurement covariance,

$$\hat{\mathbf{R}}_K \simeq \boldsymbol{\Sigma}_* = \mathbf{H}(\boldsymbol{\theta})\boldsymbol{\Sigma}_x\mathbf{H}(\boldsymbol{\theta})^H + \sigma_w^2\mathbf{I}\quad (3.51)$$

Since

$$\begin{aligned}\boldsymbol{\Sigma}_* &= \mathbf{H}(\boldsymbol{\theta})\boldsymbol{\Sigma}_x\mathbf{H}(\boldsymbol{\theta})^H + \sigma_w^2\mathbf{I} \\ &= \mathbf{S}\mathbf{\Lambda}\mathbf{S}^H + \mathbf{G}\mathbf{G}^H\sigma_w^2 + \mathcal{O}(1/\sqrt{N})\end{aligned}\quad (3.52)$$

From here, we leverage several properties of the eigenvector matrices, namely that *i*) $\mathbf{S}^H\mathbf{S} = \mathbf{I}$; and *ii*) $\mathbf{S}^H\mathbf{G}\mathbf{G}^H\mathbf{S} = \mathbf{0}$ where $\mathbf{0}$ is the $n_s \times n_s$ matrix of zeros. Such that

$$\mathbf{S}^H\mathbf{H}\boldsymbol{\Sigma}_x\mathbf{H}^H\mathbf{S} + \sigma_w^2\mathbf{I} = \mathbf{\Lambda} + \mathcal{O}(1/\sqrt{N})\quad (3.53)$$

Employing the Moore-Penrose pseudoinverse for non-square matrices, we obtain the important identity

$$[\mathbf{S}^H\mathbf{A}]^\dagger[\mathbf{\Lambda} - \sigma_w^2\mathbf{I}][\mathbf{A}^H\mathbf{S}]^\dagger = [\mathbf{S}^H\mathbf{A}]^\dagger\mathbf{S}^H\mathbf{H}\boldsymbol{\Sigma}_x\mathbf{H}^H\mathbf{S}[\mathbf{A}^H\mathbf{S}]^\dagger\quad (3.54)$$

Under the condition that the signal is independent and identically distributed, and consequently that the signal covariance $\boldsymbol{\Sigma}_x$ is diagonal, substituting this back into (3.50) and applying (3.49) results in the simplified MMLE solution

$$\tilde{g}(\boldsymbol{\theta}) = \text{Tr}\{[\mathbf{A}^\dagger\mathbf{H}]^H\mathbf{A}^H\mathbf{G}\mathbf{G}^H\mathbf{A}[\mathbf{A}^\dagger\mathbf{H}]\}\quad (3.55)$$

This solution can be expressed in terms of the signal subspace projection matrix $\boldsymbol{\Pi}_x = \mathbf{A}\mathbf{A}^\dagger$

$$\text{Tr}\{[\mathbf{A}^\dagger\mathbf{H}]^H\mathbf{A}^H\mathbf{G}\mathbf{G}^H\mathbf{A}[\mathbf{A}^\dagger\mathbf{H}]\} = \text{Tr}\{[\boldsymbol{\Pi}_x\mathbf{H}]^H\mathbf{G}\mathbf{G}^H[\boldsymbol{\Pi}_x\mathbf{H}]\}\quad (3.56)$$

Applying an additive error of the general form $\mathbf{H}(\boldsymbol{\theta}) = \mathbf{A}(\boldsymbol{\theta}) + \mathbf{E}$ and expanding produces an objective function comprised of three terms

$$\tilde{g}(\boldsymbol{\theta}) = \text{Tr}\{\mathbf{A}^H \mathbf{G} \mathbf{G}^H \mathbf{A}\} + 2 \text{Tr}\{\Re\{\mathbf{E}^H \mathbf{A} \mathbf{A}^\dagger \mathbf{G} \mathbf{G}^H \mathbf{A}\}\} + \text{Tr}\{[\mathbf{A}^\dagger \mathbf{E}]^H \mathbf{A} \mathbf{G} \mathbf{G}^H \mathbf{A} [\mathbf{A}^\dagger \mathbf{E}]\} \quad (3.57)$$

where the first term is notably identical to the objective function used in the MUSIC algorithm. It follows, then, that optimization over this objective function approaches the MUSIC solution if this is the dominant term of the objective, that is

$$\text{Tr}\{\mathbf{A}^H \mathbf{G} \mathbf{G}^H \mathbf{A}\} \gg \text{Tr}\{\Re\{\mathbf{E}^H \boldsymbol{\Pi}_x \mathbf{G} \mathbf{G}^H \mathbf{A}\}\} + \text{Tr}\{[\boldsymbol{\Pi}_x \mathbf{E}]^H \mathbf{G} \mathbf{G}^H [\boldsymbol{\Pi}_x \mathbf{E}]\} \quad (3.58)$$

This is the so-called similarity condition for convergence of the MMUSIC and MML solutions. Notably, this condition is guaranteed when the error term $\mathbf{E} \rightarrow \mathbf{0}$ such that $\mathbf{H} \rightarrow \mathbf{A}$.

3.B Proof of Corollary 1.1

The proof of general applicability in the single signal case is derived from the misspecified maximum-likelihood objective given in (3.45). In the particular case of the $N = 1$ signal, the array manifold vectors \mathbf{H} and \mathbf{A} are substituted with \mathbf{h} and \mathbf{a} , respectively. Equation (3.55) can subsequently be expressed for the single-signal case as

$$\text{Tr}\{[(\mathbf{a}^H \mathbf{a})^{-1} \mathbf{a}^H \mathbf{h}]^H \mathbf{a}^H \mathbf{G} \mathbf{G}^H \mathbf{a} [(\mathbf{a}^H \mathbf{a})^{-1} \mathbf{a}^H \mathbf{h}]\} = M^{-2} \text{Tr}\{[\mathbf{a}^H \mathbf{h}]^H \mathbf{a}^H \mathbf{G} \mathbf{G}^H \mathbf{a} [\mathbf{a}^H \mathbf{h}]\} \quad (3.59)$$

Applying the properties of the trace for scalar products, this is rewritten as

$$\min_{\theta} M^{-2} \text{Tr}\{\mathbf{a}^H \mathbf{G} \mathbf{G}^H \mathbf{a} [\mathbf{a}^H \mathbf{h} \mathbf{h}^H \mathbf{a}]\} = \min_{\theta} M^{-2} \text{Tr}\{\mathbf{a}^H \mathbf{G} \mathbf{G}^H \mathbf{a}\} \text{Tr}\{\mathbf{a}^H \mathbf{h} \mathbf{h}^H \mathbf{a}\} \quad (3.60)$$

This results in a simpler necessary condition for equivalence between the MMUSIC and MMLE cost functions: that $\text{Tr}\{\mathbf{a}^H \mathbf{h} \mathbf{h}^H \mathbf{a}\}$ is independent of θ . To show that this is the case, we refer to the diagonalization of the matrix $\mathbf{h} \mathbf{h}^H$

$$\mathbf{a}^H \mathbf{h} \mathbf{h}^H \mathbf{a} = \mathbf{a}^H \mathbf{U} \mathbf{\Lambda}_h \mathbf{U}^H \mathbf{a} \quad (3.61)$$

As the matrix $\mathbf{h} \mathbf{h}^H$ is, by definition, Hermitian, it follows that the diagonal matrix $\mathbf{\Lambda}_h$ is both real and strictly positive. Using the notation $\mathbf{U}^H \mathbf{a} = \mathbf{z}$, the minimization of this term can be expressed as

$$\begin{aligned} \min_{\theta} \text{Tr}\{\mathbf{z} \mathbf{\Lambda}_h \mathbf{z}^H\} &= \min_{\theta} \text{Tr}\{\mathbf{z} \mathbf{z}^H\} \\ &= \min_{\theta} \text{Tr}\{\mathbf{z}^H \mathbf{z}\} \end{aligned} \quad (3.62)$$

Reversing the notational change we find that

$$\min_{\theta} \text{Tr}\{\mathbf{a}^H \mathbf{h} \mathbf{h}^H \mathbf{a}\} = \min_{\theta} \text{Tr}\{\mathbf{a}^H \mathbf{U} \mathbf{U}^H \mathbf{a}\} \quad (3.63)$$

From the properties of the decomposition we once again are able to leverage the identity $\mathbf{U} \mathbf{U}^H = \mathbf{I}$ to show that

$$\begin{aligned} \min_{\theta} \text{Tr}\{\mathbf{a}^H \mathbf{h} \mathbf{h}^H \mathbf{a}\} &= \min_{\theta} \text{Tr}\{\mathbf{a}^H \mathbf{U} \mathbf{U}^H \mathbf{a}\} \\ &= \min_{\theta} \text{Tr}\{\mathbf{a}^H \mathbf{a}\} \end{aligned} \quad (3.64)$$

And finally, since \mathbf{a} is defined as a unit magnitude complex vector function of θ , $\mathbf{a}^H \mathbf{a} = M$ and

$$\min_{\theta} \text{Tr}\{\mathbf{a}^H \mathbf{h} \mathbf{h}^H \mathbf{a}\} = M \quad (3.65)$$

which is unconditionally independent of θ and thus the property is shown to hold regardless of the relationship between \mathbf{h} and \mathbf{a} .

3.C Proof of Theorem 2

Considering stochastic characterizations of the additive error term \mathbf{E} allows us to consider the *average* performance of the misspecified maximum-likelihood (MML) and misspecified Multiple Signal Classification (MMUSIC) solutions for small errors. Taking the expectation of the MML solution given in (3.58) results in a stochastic similarity condition

$$\text{Tr}\{\mathbf{A}^H \mathbf{G} \mathbf{G}^H \mathbf{A}\} \gg \text{Tr}\{\Re\{\mathbb{E}[\mathbf{E}^H] \mathbf{\Pi}_x^H \mathbf{G} \mathbf{G}^H \mathbf{A}\}\} + \text{Tr}\{\mathbf{\Pi}_x^H \mathbf{G} \mathbf{G}^H \mathbf{\Pi}_x \mathbb{E}[\mathbf{E} \mathbf{E}^H]\} \quad (3.66)$$

As an interesting particular case, we consider a first-order perturbation of the array manifold function $\mathbf{a}(\theta)$ given by

$$\mathbf{h}(\theta) = \mathbf{a}(\theta) + \mathbf{e} \quad (3.67)$$

where \mathbf{e} represents an additive circularly symmetric complex error term with zero mean and known covariance $\mathbf{\Sigma}_e$. Considering the multiple-signal problem we have

$$\mathbf{H}(\theta) = \mathbf{A}(\theta) + \mathbf{E} \quad (3.68)$$

where \mathbf{E} is a matrix comprised of the horizontal concatenation of N i.i.d. random vectors $\mathbf{E} = [\mathbf{e}_1 \dots \mathbf{e}_N]$, such that $\mathbb{E}\{\mathbf{E}\} = \mathbf{0}$ and $\mathbb{E}\{\mathbf{E} \mathbf{E}^H\} = N \mathbf{\Sigma}_e$. Under these conditions (3.66) collapses to

$$\text{Tr}\{\mathbf{A}^H \mathbf{G} \mathbf{G}^H \mathbf{A}\} \gg \text{Tr}\{\mathbf{\Pi}_x^H \mathbf{G} \mathbf{G}^H \mathbf{\Pi}_x N \mathbf{\Sigma}_e\} \quad (3.69)$$

As the quadratic terms are guaranteed to be positive semi-definite, we are able to consider the trace as an inner-product and leverage the Cauchy-Schwarz inequality to show that

$$1 \gg N \text{Tr}\{[(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H] \mathbf{\Sigma}_e [\mathbf{A} (\mathbf{A}^H \mathbf{A})^{-1}]\} \quad (3.70)$$

If the elements of the random perturbation are independent and identically distributed for each sensor such that $\mathbf{\Sigma}_e = \sigma_e^2 \mathbf{I}$, this expression simplifies further to provide a condition on the variance of the perturbation as

$$\sigma_e^2 \ll \frac{1}{N \text{Tr}\{(\mathbf{A}^H \mathbf{A})^{-1}\}} \quad (3.71)$$

Chapter 4

Bayesian Statistics Estimation in Gaussian Filtering

Gaussian filtering provides a Bayesian approach to dynamic state estimation, but requires precise statistical information about observation noise. When this information is unavailable, it is necessary to estimate the measurement noise covariance based on the observation and/or innovation sequences. Common approaches are based on frequentist paradigms, such as maximum-likelihood or least squares criteria, which are dependent on having a large number of observations and do not guarantee the structural validity (*i.e.* positive semi-definiteness) of the output matrix. We propose an alternative Bayesian approach to recursively estimate the innovation covariance concurrently with state estimation, guaranteeing the requisite properties of the covariance and eliminating the need for an explicit measurement noise statistics model. The proposed Bayesian statistics estimation (BSE) method is embedded within a sigma-point Kalman filter, and adapted to non-stationary noise processes using Gaussian filter consistency tests. The adaptive filter is validated through simulation, and compared to an established online approach based on frequentist covariance matching.

4.1 Dynamic State Estimation: Applications and Challenges

The problem under study concerns the derivation of new robust methods to solve the Bayesian filtering problem, where robustness refers to the ability of a method to continue to function in spite of possible system uncertainties or model mismatch. General Gaussian state-space models of interest are expressed as [51, 52],

$$\mathbf{x}_k = \mathbf{f}_{k-1}(\mathbf{x}_{k-1}) + \boldsymbol{\nu}_{k-1}, \quad \boldsymbol{\nu}_{k-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{k-1}), \quad (4.1)$$

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{n}_k, \quad \mathbf{n}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k), \quad (4.2)$$

where $\mathbf{x}_k \in \mathbb{R}^{n_x}$ and $\mathbf{y}_k \in \mathbb{R}^{n_y}$ are the hidden state of the system and measurements at discrete time k ; $\mathbf{f}_{k-1}(\cdot)$ and $\mathbf{h}_k(\cdot)$ are known (possibly nonlinear) functions of the state; ν_{k-1} and \mathbf{n}_k are assumed to be independent. The solution to the Bayesian filtering problem is given by the posterior distribution $p(\mathbf{x}_k | \mathbf{y}_{1:k})$, which gathers all the information about the states at time k given by the available measurements up to time k , $\mathbf{y}_{1:k} = (\mathbf{y}_1, \dots, \mathbf{y}_k)$. Different alternatives are available in the literature, including particle filters (PFs) [53], the most popular method for general nonlinear/non-Gaussian systems, and the class of sigma-point GFs [54–60] under the Gaussian assumption of interest here. The latter is a family of derivative-free algorithms that are based on a weighted sum of function values at specified (*i.e.*, deterministic) points within the domain of integration, in contrast to the stochastic sampling performed by PFs. One of the major challenges is that all these methods assume fully known noise statistics (*i.e.*, Gaussian noise covariances \mathbf{Q}_{k-1} and \mathbf{R}_k , in the Gaussian context), an assumption that does not hold in real-life applications. Therefore, robust solutions must be developed to apply standard filtering techniques.

A plethora of methods face the Gaussian noise statistics estimation problem within the KF framework [61]. *Correlation methods* [62], which are based on the correlation function of the innovation sequence and its direct relation to the unknown parameters, are the most popular for the estimation of Gaussian noise covariance matrices. Several extensions and modifications of these standard techniques have appeared during the last 30 years, but the recently introduced Autocovariance Least Square (ALS) method [63, 64] is probably the most promising solution [61]. Another typical solution is the *covariance matching method* [65, 66], which is equivalent to the maximum-likelihood (ML) estimator [67] if the system noise is zero mean [68]. This method uses the state estimation residuals to compute the process and measurement noise covariances. In general, ML solutions may not accept an analytic expression and expectation-maximization simulation-based solutions are needed [69–71]. The main drawback of these methods is that the positive definiteness of the covariance matrix is not guaranteed. Recently, some related contributions appeared based on Monte Carlo sampling [72], variational Bayes [73, 74] and marginalized PF [75].

A key point on the noise statistics estimation problem within the KF framework is the joint state/covariance identifiability. It has been shown in [76] that noise covariance matrices cannot be jointly estimated along the states. The time update step is computed using only the first two moments, but the third-order cross-conditional moment is nonzero, thus making these covariances not identifiable. Therefore, the standard Bayesian recursive characterization of the joint posterior distribution [77, 78] is not a valid approach and a divide-and-conquer strategy must be adopted. In this contribution, we provide a fully Bayesian solution to this problem for unknown noise

measurement covariance matrices \mathbf{R}_k together with the corresponding adaptive GF formulation, and its implementation using both linear KF and sigma point strategies for both linear and nonlinear dynamic systems.

4.2 Review of Recursive Gaussian Filtering

The filtering solution is given by the *a posteriori* distribution, $p(\mathbf{x}_k|\mathbf{y}_{1:k})$, which in the case of linear/Gaussian systems turns to be Gaussian whose parameters can be computed analytically. If the system is nonlinear and Gaussian, both the transition and likelihood densities are Gaussian distributed, and the predictive and posterior densities are typically approximated as Gaussian yielding to the family of so-called GFs such that [51, 54],

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, \Sigma_{k|k-1}), \quad (4.3)$$

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}, \Sigma_{k|k}), \quad (4.4)$$

are recursively computed in two steps: *prediction* and *update*. In the prediction step, the filter is interested in calculating the mean and corresponding prediction error covariance of the predictive density ¹

$$\hat{\mathbf{x}}_{k|k-1} = \int \mathbf{f}(\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})d\mathbf{x}_{k-1}, \quad (4.5)$$

$$\Sigma_{k|k-1} = \int \mathbf{f}^2(\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})d\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k|k-1}^2 + \mathbf{Q}_{k-1}. \quad (4.6)$$

After the prediction step, the filter receives a new measurement, \mathbf{y}_k , and is able to construct the innovation, $\mathbf{z}_k = \mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1}$, where the predicted measurement is

$$\hat{\mathbf{y}}_{k|k-1} = \int \mathbf{h}(\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k-1})d\mathbf{x}_k. \quad (4.7)$$

Both innovation and cross-covariance matrices are computed by the following integrals,

$$\Sigma_{y,k|k-1} = \int \mathbf{h}^2(\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k-1})d\mathbf{x}_k - \hat{\mathbf{y}}_{k|k-1}^2 + \mathbf{R}_k, \quad (4.8)$$

$$\Sigma_{xy,k|k-1} = \int \mathbf{x}_k \mathbf{h}^\top(\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k-1})d\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1} \hat{\mathbf{y}}_{k|k-1}^\top, \quad (4.9)$$

and the optimal Kalman gain is given by,

$$\mathbf{W}_k = \Sigma_{xy,k|k-1} \Sigma_{y,k|k-1}^{-1}. \quad (4.10)$$

¹We have omitted the dependence with time of $\mathbf{f}_{k-1}(\cdot)$ and $\mathbf{h}_k(\cdot)$ for the sake of notational clarity.

Notice that $\mathbf{z}_k \sim \mathcal{N}(\mathbf{z}_k; \mathbf{0}, \Sigma_{y,k|k-1})$ and the innovation covariance depend on the unknown measurement covariance, which typically has to be estimated $\mathbf{R}_k \approx \hat{\mathbf{R}}_k$. Finally, in the subsequent update step, the filter approximates the posterior by the corrected filter update equations,

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{W}_k (\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1} - \boldsymbol{\mu}_k), \quad (4.11)$$

$$\Sigma_{k|k} = \Sigma_{k|k-1} - \mathbf{W}_k \hat{\Sigma}_{y,k|k-1} \mathbf{W}_k^\top. \quad (4.12)$$

4.3 Bayesian Estimation of Distribution Parameters

We are interested in solving a problem in which we are given n samples drawn independently from the same normal distribution $\mathbf{z}_{1:n} = (\mathbf{z}_1, \dots, \mathbf{z}_n) \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$, and the statistics of that distribution must be inferred. This is the fundamental problem discussed in this chapter. More precisely, we are interested in the posterior distribution of these statistics given the observed data. This section provides a general discussion of the Bayesian inference solution, leaving the application of the result to the context of GFs for Section 4.4, and its implementation for Section 4.5.

Application of Bayes' rule to the distribution of the observed sequence yields the expression,

$$p(\boldsymbol{\mu}, \mathbf{C} | \mathbf{z}_{1:n}) \propto p(\mathbf{z}_{1:n} | \boldsymbol{\mu}, \mathbf{C}) p(\boldsymbol{\mu}, \mathbf{C}), \quad (4.13)$$

where the term $p(\mathbf{z}_{1:n} | \boldsymbol{\mu}, \mathbf{C})$ corresponds to the likelihood of observations given the unknown mean and covariance $\boldsymbol{\mu}$ and \mathbf{C} , and $p(\boldsymbol{\mu}, \mathbf{C})$ term represents the a priori knowledge about those parameters. We assume that the likelihood distribution of the observed sequence is Gaussian,

$$p(\mathbf{z}_{1:n} | \boldsymbol{\mu}, \mathbf{C}) = \mathcal{N}(\boldsymbol{\mu}, \mathbf{C}), \quad (4.14)$$

where both mean and covariance are unknown. Given this statistical characterization, we have to identify a statistically principled method of estimating the unknown parameters $\hat{\boldsymbol{\mu}}$ and $\hat{\mathbf{C}}$, given the a priori information $p(\boldsymbol{\mu}, \mathbf{C})$ and new observations of the observed sequence $\mathbf{z}_{1:n}$. One such method which fits well into the Bayesian recursive estimation framework is the method of conjugate priors.

4.3.1 Method of Conjugate Priors

In this section, we will give a generalized overview of the method of recursively estimating distribution statistics using conjugate priors. Selecting a prior distribution which is conjugate to the likelihood distribution results in a closed-form expression for the posterior distribution that is of the same type as the prior. For the purposes of GF, we are interested in scenarios where the

observed sequence is a multivariate normal likelihood with unknown mean and covariance matrix, in which case the conjugate prior and posterior take the form of a normal-inverse-Wishart distribution ($\mathcal{NW}^{-1}(\cdot)$) [79],

$$p(\boldsymbol{\mu}, \mathbf{C}) = \mathcal{NW}^{-1}(\boldsymbol{\mu}, \mathbf{C}; \boldsymbol{\mu}_0, \kappa_0, \nu_0, \boldsymbol{\Psi}) \quad (4.15)$$

$$= \mathcal{N}(\boldsymbol{\mu} | \boldsymbol{\mu}_0, \frac{1}{\kappa_0} \mathbf{C}) \mathcal{W}^{-1}(\mathbf{C} | \boldsymbol{\Psi}, \nu_0). \quad (4.16)$$

With this selection of prior, the estimated parameters of the likelihood distribution are linked to the hyperparameters of the conjugate distribution, with estimated mean $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu}_0$ and covariance $\hat{\mathbf{C}} = \boldsymbol{\Psi}/\nu_0$, based on κ_0 and ν_0 pseudo-observations of the likelihood, respectively, where the quantity of pseudo-observations is based on the implied number of observations κ_0 and ν_0 represented within the prior. The posterior is obtained by updating the parameters of the \mathcal{NW}^{-1} prior using observations \mathbf{z}_n of the likelihood distribution as,

$$p(\boldsymbol{\mu}, \mathbf{C} | \mathbf{z}_{1:n}) = \mathcal{NW}^{-1}(\boldsymbol{\mu}, \mathbf{C}; \tilde{\boldsymbol{\mu}}_0, \tilde{\kappa}_0, \tilde{\nu}_0, \tilde{\boldsymbol{\Psi}}), \quad (4.17)$$

$$\tilde{\boldsymbol{\mu}}_0 = \frac{\kappa_0 \boldsymbol{\mu}_0 + n \bar{\mathbf{z}}}{\kappa_0 + n}, \quad (4.18)$$

$$\tilde{\kappa}_0 = \kappa_0 + n, \quad (4.19)$$

$$\tilde{\nu}_0 = \nu_0 + n, \quad (4.20)$$

$$\tilde{\boldsymbol{\Psi}} = \boldsymbol{\Psi} + \boldsymbol{\Psi}_n + \frac{\kappa_0 n}{\kappa_0 + n} (\bar{\mathbf{z}} - \boldsymbol{\mu}_0)(\bar{\mathbf{z}} - \boldsymbol{\mu}_0)^\top, \quad (4.21)$$

with $\boldsymbol{\Psi}_n = \sum_{i=1}^n (\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{z}_i - \bar{\mathbf{z}})^\top$ and $\bar{\mathbf{z}}$ the sample mean of $\mathbf{z}_{1:n}$. Note that here the reliance on the prior when determining the posterior mean $\boldsymbol{\mu}_0$ and sum of pairwise deviations $\boldsymbol{\Psi}$, depends on the number of pseudo-observations κ_0 and ν_0 , while the posterior hyperparameters combine the pseudo-observations with quantity κ_0 and ν_0 with the new, real observations with quantity n .

We are interested in obtaining estimates from the marginal posterior $p(\mathbf{C} | \mathbf{z}_{1:n})$, which can be obtained integrating out $\boldsymbol{\mu}$ in $p(\boldsymbol{\mu}, \mathbf{C} | \mathbf{z}_{1:n})$. This yields,

$$p(\mathbf{C} | \mathbf{z}_{1:n}) = \mathcal{W}^{-1}(\mathbf{C}; \tilde{\nu}_0, \tilde{\boldsymbol{\Psi}}). \quad (4.22)$$

We can compute the mean of this marginal as

$$\hat{\mathbf{C}}_{\text{mean}} = \frac{\tilde{\boldsymbol{\Psi}}}{\tilde{\nu}_0 - n_y - 1}, \quad (4.23)$$

with $\tilde{\nu}_0 > n_y + 1$. Additionally, the mode is

$$\hat{\mathbf{C}}_{\text{mode}} = \frac{\tilde{\boldsymbol{\Psi}}}{\tilde{\nu}_0 + n_y + 1}. \quad (4.24)$$

4.3.2 Initial Prior

Although it is possible to begin the Bayesian recursion loop with an objective or uninformative prior with $\kappa_0 \rightarrow 0$, $\nu_0 \rightarrow 0$, this will generally result in a poor initialization and a longer than desirable convergence time. In the proposed method, the informative prior is given by interpretation of the hyperparameters of the prior distribution given in (4.15). Namely, we can take the initial N_0 samples of the sequence and obtain a sample mean and covariance of the observed sequence, and transform them into the prior hyperparameters,

$$\boldsymbol{\mu}_0 = \hat{\boldsymbol{\mu}}_{\mathbf{z}} = \frac{1}{N_0} \sum_{k=1}^{N_0} \mathbf{z}_k, \quad (4.25)$$

$$\kappa_0 = N_0, \quad (4.26)$$

$$\nu_0 = N_0, \quad (4.27)$$

$$\boldsymbol{\Psi}_0 = \nu_0 \hat{\boldsymbol{\Sigma}}_0 = \sum_{k=1}^{N_0} \mathbf{z}_k \mathbf{z}_k^\top - \hat{\boldsymbol{\mu}}_0 \hat{\boldsymbol{\mu}}_0^\top. \quad (4.28)$$

From here the hyperparameters may be updated recursively for future observations as described in (4.18)-(4.21).

4.4 Leveraging Bayesian Estimation for Recursive Gaussian Filtering

Remarkably, it is possible to use the BSE approach described in Section 4.3 to eliminate one of the necessary assumptions of the standard Gaussian filter: that the measurement noise covariance \mathbf{R}_k is precisely known for each filter update at each time k . Recall from (4.8) that the measurement noise covariance appears only when computing the innovation covariance $\boldsymbol{\Sigma}_{y,k|k-1}$, and thus if we are able to estimate the innovation covariance by another method, we can completely bypass the need for \mathbf{R}_k . Instead of relying on the model, we can instead rely on the observed innovations to estimate the covariance directly. Concatenation of the innovations at successive instants yields the so-called innovation sequence, $\mathbf{z}_{1:k} = (\mathbf{z}_1, \dots, \mathbf{z}_k)$, the distribution of which we wish to estimate for use within the GF. Application of the BSE method requires that the observations are independent at each update, which is guaranteed for optimal filtering by the *innovations property* [80–82] of the GF. Moreover, it is possible to use a fully sequential estimator where $n = 1$ and the estimate of the innovation covariance $\hat{\boldsymbol{\Sigma}}_{y,k|k-1}$ is updated at each filter update at time k . A batch estimator considering more samples is straightforward to obtain, but we are interested in the sequential form which may be embedded within the sequential, recursive structure of the GF.

Once the innovations are computed, and selecting some values for the a priori parameters $(\boldsymbol{\mu}_0^S, \kappa_0^S, \nu_0^S, \boldsymbol{\Psi}^S)$, the posterior parameters are updated as

$$\tilde{\boldsymbol{\mu}}_0^S = \frac{\kappa_0^S \boldsymbol{\mu}_0^S + \mathbf{z}_k}{\kappa_0^S + 1}, \quad (4.29)$$

$$\tilde{\kappa}_0^S = \kappa_0^S + 1, \quad (4.30)$$

$$\tilde{\nu}_0^S = \nu_0^S + 1, \quad (4.31)$$

$$\tilde{\boldsymbol{\Psi}}^S = \boldsymbol{\Psi}^S + \frac{\kappa_0^S}{\kappa_0^S + 1} (\mathbf{z}_k - \boldsymbol{\mu}_0^S)(\mathbf{z}_k - \boldsymbol{\mu}_0^S)^\top, \quad (4.32)$$

which would become the parameters of the new a priori at $k + 1$. In the subsequent update step, the GF approximates the mean and covariance of the state posterior using the update equations [83] as described in (4.10)-(4.12). An overview of the augmented filter architecture is shown in Figure 4.4.1.

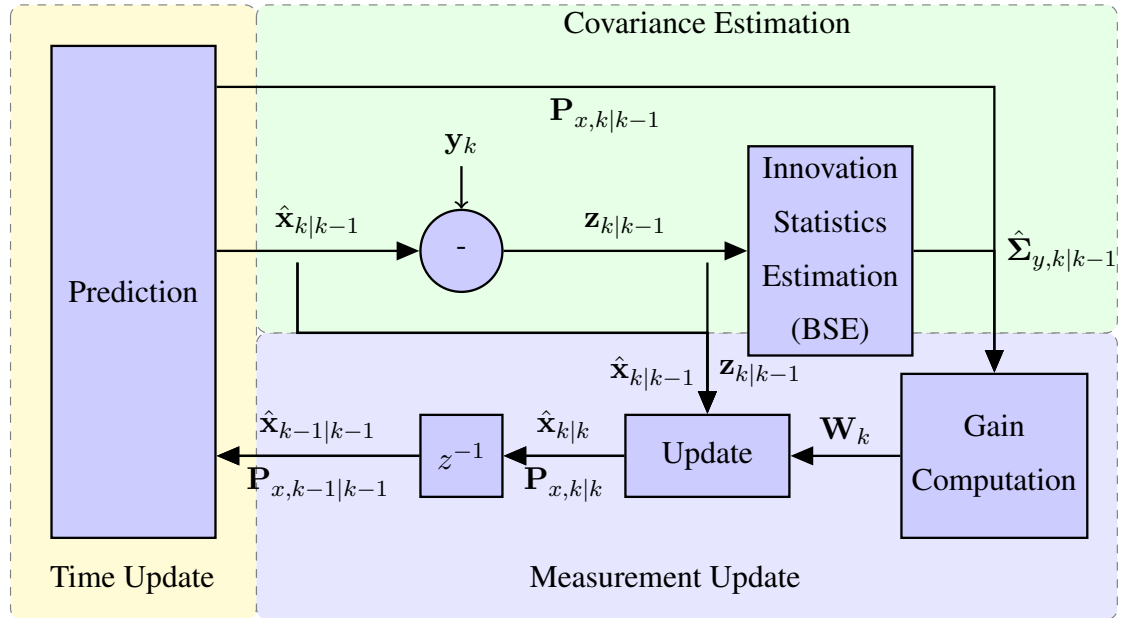


Figure 4.4.1: Block diagram of the augmented GF with Bayesian estimation of innovations' noise statistics.

4.4.1 Initializing the Bayesian Statistics Estimator

Considering the recursive nature of the Bayesian estimator, its performance with respect to the unknown true measurement noise covariance, and the corresponding innovation covariance, is

highly dependent on the initial values of the prior hyperparameters. One method is to select a non-informative with zero mean $\boldsymbol{\mu}_0$ and pseudo-observation parameters κ_0 , and ν_0 , and a conservative initial estimate of the covariance and associated $\boldsymbol{\Psi}_0$. This gives the Bayesian filter the highest level of adaptability, in exchange for greater initial instability and a longer convergence time. It is important to note that using this highly variable covariance estimate directly in the gain computation is likely to result in a longer convergence time in the state estimate. This in turn will result in a longer period in which the filter is performing sub-optimally, thus violating the innovations property and consequently affecting the quality of the initial covariance estimates.

To avoid this outcome, the proposed method leverages the adaptive bandwidth of the optimal GF by providing a conservative estimate of the measurement noise covariance $\hat{\mathbf{R}}_k$, allowing the filter to process for an initial “transient” period of N_0 filter iterations. If the assumed measurement noise covariance is greater than the optimal true value, this results in a bias towards the model and against the measurements for the purposes of computing the Kalman gain. Assuming that the process model and initial states of the filter are both accurately specified, it is possible for the filter function, albeit suboptimally, for a number of initial iterations. Over the course of this initial transient period, the Bayesian filter remains inactive, while the sample statistics are computed for the first set of innovations, from which the prior hyperparameters may be initialized *informatively* as in (4.25)-(4.28),

$$\boldsymbol{\mu}_0^S = \sum_{k=0}^{N_0-1} \mathbf{z}_k, \quad (4.33)$$

$$\kappa_0^S = N_0, \quad (4.34)$$

$$\nu_0^S = N_0, \quad (4.35)$$

$$\boldsymbol{\Psi}^S = \sum_{k=0}^{N_0-1} \mathbf{z}_k \mathbf{z}_k^\top - \boldsymbol{\mu}_0^S \boldsymbol{\mu}_0^{S\top}. \quad (4.36)$$

Using this informative prior, the Bayesian covariance estimate converges quickly towards the true noise characteristics of the filter, and from here the conservative estimate of $\hat{\mathbf{R}}_k$ used during the transient period to compute the innovation covariance in (4.8) may be replaced directly with the Bayesian covariance estimate given in (4.23) as $\boldsymbol{\Sigma}_{y,k|k-1} = \hat{\mathbf{C}}_{\text{mean}}$ for subsequent iterations.

4.4.2 Accounting for Changes in the Noise Characteristics

Readers who are familiar with Bayesian estimation and, in particular, the use of conjugate priors for estimating distribution parameters, will notice that there is a limitation which has thus far

not been accounted for in the proposed methodology. That is, the measurement noise \mathbf{R}_k that we are interested in eliminating from the Gaussian filtering process, and which is additively related to the innovation covariance that we are trying to estimate (4.8) is here specified as being parameterized by, and possibly varying according to, the time index k while, in the Bayesian framework, we have thus far established for estimation of the noise characteristics, the distribution parameters of interest $\boldsymbol{\mu}$ and \mathbf{C} are not. In order to obtain a fully Bayesian approach to estimate time varying parameters, it is necessary to have a model for the behavior of the parameters over time. Although this is straightforward to do with the mean $\boldsymbol{\mu}$, which may be assumed to remain static by way of the innovations property of the filter, modelling the relationships between the n_y^2 parameters of the covariance matrix as they evolve through time is generally infeasible. In this section, we propose an alternate methodology for dealing with time-varying noise characteristics based on identifying deviations in the estimated and observed innovations, and reinitializing the Bayesian estimator accordingly.

4.4.2.1 Identifying changes in the noise characteristics

As the accuracy of the BSE is dependent on the amount of pseudo-observations used in determining the parameters of interest, it is important to preserve as much information as possible while throwing out data which no longer corresponds to the distribution of interest. To this end, handling time-varying statistics requires precise and reliable detection of changes in the underlying distribution. Although there are many possible approaches to make this determination, the relationship between the BSE and Gaussian filtering allows us to identify these changes based on the performance of the GF. That is, when our estimate of the innovation covariance is no longer accurately representative of the measurement noise distribution, the performance of the filter will decrease. This inconsistency is detectable, and there are several metrics in the literature for online determination of the consistency of the recursive GF based on the statistics of the innovations sequence. In our testing the innovations' metric which gave the most reliable and responsive indicator of a change in the underlying noise characteristics is the so-called time-averaged normalized innovation squared (NIS) [84],

$$\bar{\epsilon}_k = \frac{1}{K_D} \sum_{k_0=0}^{K_D-1} \mathbf{z}_{k-k_0}^\top \hat{\mathbf{S}}_{k-k_0}^{-1} \mathbf{z}_{k-k_0}. \quad (4.37)$$

This metric tests that the innovations are white, zero-mean and distributed with covariance \mathbf{S}_{k-k_0} for a given window size in samples K_D . During optimal filtering, wherein the innovations adhere to these conditions, the time-averaged NIS metric can be modelled as a chi-square random variable with

$K_D n_z$ degrees of freedom, where n_z is the dimension of the innovations. Evaluating the consistency of the filter, and subsequently of the BSE, may be performed by determining how likely the computed time-averaged NIS values are to have come from that chi-squared distribution. The literature also provides a whiteness test for the innovations based on the autocorrelation of the innovations sequence,

$$\bar{\rho}_k = \sum_{k_0=0}^{K_D-1} \mathbf{z}_k^\top \left[\sum_{k_0=0}^{K_D-1} \mathbf{z}_k \mathbf{z}_k^\top \right]^{-\frac{1}{2}} \left[\sum_{k_0=0}^{K_D-1} \mathbf{z}_{k+1} \mathbf{z}_{k+1}^\top \right]^{-\frac{1}{2}} \mathbf{z}_k, \quad (4.38)$$

which, again assuming that the innovations are white and for large enough window size K_D , can be modelled as a normal distribution with variance $1/K_D$. In the case of both the time-averaged NIS metric and the whiteness metric, determining whether the corresponding metric does belong to the modeled distribution can be done either by hypothesis testing with a given confidence region or by other methods to determine the fitness of a sequence to a given distribution, for example, the Kolmogorov–Smirnov test [85]. Among these possible approaches, the application of hypothesis testing to the time-averaged NIS metric (4.37) stands out for being easily represented in a low-memory, recursive form using recursive averaging,

$$\begin{aligned} \bar{\epsilon}_k &= \frac{1}{K_D} \epsilon_k + \frac{1}{K_D} \sum_{k_0=1}^{K_D-1} \epsilon_{k-k_0}, \\ &= \frac{1}{K_D} \epsilon_k + \frac{k_0 - 1}{k_0} \bar{\epsilon}_{k-1} \end{aligned} \quad (4.39)$$

where ϵ_k is the instantaneous NIS at time k given by $\epsilon_k = \mathbf{z}_k^\top \hat{\mathbf{S}}_k^{-1} \mathbf{z}_k$. This form is easily incorporated into the already recursive, low-memory structure of the Gaussian and Bayesian filtering already employed within the proposed methodology.

4.4.2.2 Responding to changes in the noise characteristics

Upon detecting a change in the noise characteristics as described in 4.4.2.1, the parameters of the BSE must be adjusted to account for this change. This process is similar to the initialization procedure described in 4.4.1. As the filter has been running for some time, however, it is not necessary in this instance to disconnect the BSE from the GF and we can instead use the sample statistics from

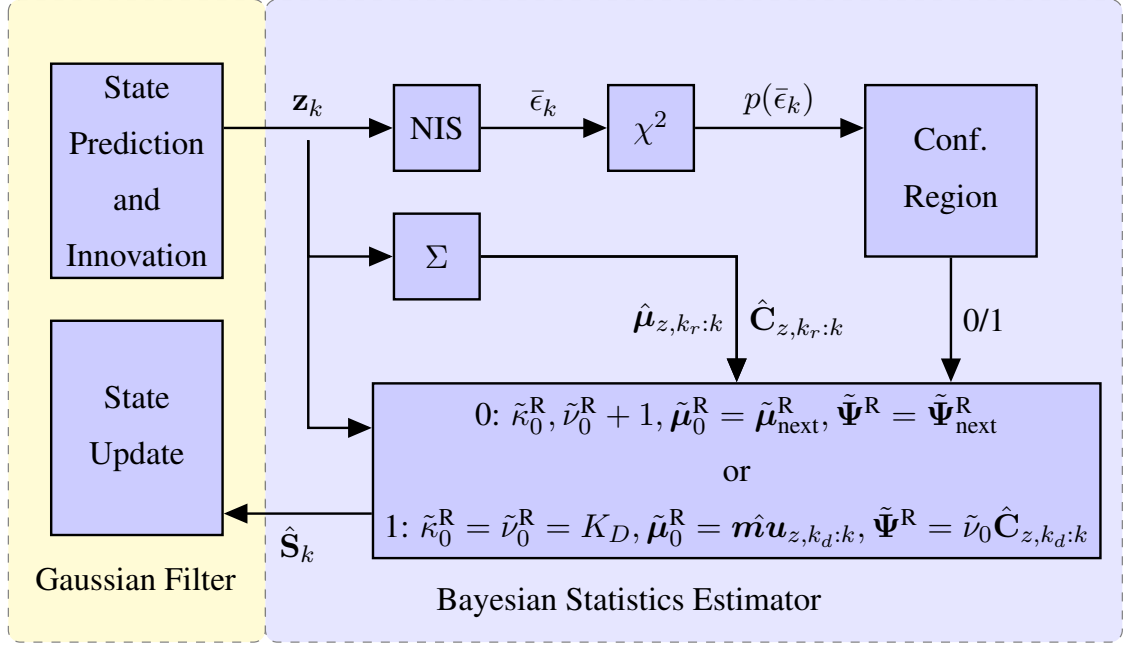


Figure 4.4.2: Bayesian Re-initialization scheme for the augmented GF.

a period of given length K_R , leading up to the detected change,

$$\boldsymbol{\mu}_0^S = \hat{\boldsymbol{\mu}}_{z, k_r: k} = \sum_{k_0=0}^{K_R-1} \mathbf{z}_{k-k_0}, \quad (4.40)$$

$$\kappa_0^S = K_R, \quad (4.41)$$

$$\nu_0^S = K_R, \quad (4.42)$$

$$\boldsymbol{\Psi}^S = \sum_{k_0=0}^{K_R-1} \mathbf{z}_{k-k_0} \mathbf{z}_{k-k_0}^\top - \boldsymbol{\mu}_0^S \boldsymbol{\mu}_0^{S\top}. \quad (4.43)$$

Notably, since these reinitialized prior hyperparameters are based on the sample averages, it is also possible to construct a recursive form in the same manner as in (4.39). From this point the filter updates may proceed as usual and the filter will converge to reflect the new measurement noise statistics until another change is detected.

4.5 Implementation Using Linear and Cubature Kalman Filtering Strategies

When it comes to implementing the recursive GF, the multidimensional integrals in the prediction and update steps are analytically intractable in the general case. If the system is linear/Gaussian, these equations turn out to be the standard KF [83]. In the nonlinear and Gaussian case we can resort to sigma-point GFs [54, 56, 86], which approximate the integrals by propagating a fixed set of points through the system non-linearity. In this section, we provide an overview of both the standard, linear KF, as well as the lauded and highly accurate CKF [87], for linear and nonlinear models, respectively, and highlight some of the challenges and limitations of using these methods with the Bayesian statistics estimator as proposed and detailed in the previous sections.

4.5.1 Linear Filtering with Standard Kalman Filter

In the linear Gaussian case the state-space model given in (4.1) and (4.2) degenerates to the linear state transition and measurement equations,

$$\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \boldsymbol{\nu}_{k-1}, \quad \boldsymbol{\nu}_{k-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{k-1}), \quad (4.44)$$

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{n}_k, \quad \mathbf{n}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k). \quad (4.45)$$

From this state-space representation we obtain the familiar set of KF prediction equations [88],

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_k \hat{\mathbf{x}}_{k|k-1}, \quad (4.46)$$

$$\boldsymbol{\Sigma}_{k|k-1} = \mathbf{F}_k \boldsymbol{\Sigma}_{k-1|k-1} \mathbf{F}_k + \mathbf{Q}_{k-1}, \quad (4.47)$$

$$\boldsymbol{\Sigma}_{xy,k|k-1} = \boldsymbol{\Sigma}_{k|k-1} \mathbf{H}_k^\top. \quad (4.48)$$

It is at this point that we make a deviation from the standard KF equations, wherein the innovation covariance would be computed from the predicted error covariance $\boldsymbol{\Sigma}_{k|k-1}$, and the terms of the measurement equation (4.45). Instead, following the prediction step of the KF, we compute the innovation $\mathbf{z}_k = \mathbf{y}_k - \mathbf{F}_k \hat{\mathbf{x}}_{k|k-1}$ and provide this input to the BSE. The output of the BSE is exactly the value of the innovation covariance obtained as in (4.23),

$$\boldsymbol{\Sigma}_{y,k|k-1} \approx \hat{\mathbf{S}}_{k,\text{mean}} = \frac{\tilde{\boldsymbol{\Psi}}}{\tilde{\nu}_0 - n_y - 1}. \quad (4.49)$$

The rest of the KF recursion proceeds as in the standard formulation by computing the optimal gain as in (4.10), which may then be used to perform the filter update and obtain the updated state estimate and covariance. Notice that this solution avoids the need to estimate \mathbf{R}_k as it is not used.

4.5.2 Nonlinear Filtering with Cubature Kalman Filter

Some of the most popular approximations of the recursive GF in nonlinear systems are referred to as sigma-point filters. These approaches define a set of so-called sigma-points and their corresponding weights [56, 87, 89], $\{\xi_i, \omega_i\}_{i=1, \dots, L_s}$, where the number of sigma points L_s depends on the specific rules associated with the type of filter and the state dimension n_x . The integrals in (4.5), (4.6), (4.8) and (4.9) can be approximated as,

$$\hat{\mathbf{x}}_{k|k-1} = \sum_{i=1}^{L_s} \omega_i \mathbf{f}(\mathbf{x}_{i,k-1|k-1}), \quad (4.50)$$

$$\hat{\mathbf{y}}_{k|k-1} = \sum_{i=1}^{L_s} \omega_i \mathbf{h}(\mathbf{x}_{i,k|k-1}),$$

$$\Sigma_{k|k-1} = \sum_{i=1}^{L_s} \omega_i \mathbf{f}^2(\mathbf{x}_{i,k-1|k-1}) - (\hat{\mathbf{x}}_{k|k-1})^2 + \mathbf{Q}_{k-1}, \quad (4.51)$$

$$\Sigma_{y,k|k-1} = \sum_{i=1}^{L_s} \omega_i \mathbf{h}^2(\mathbf{x}_{i,k|k-1}) - (\hat{\mathbf{y}}_{k|k-1})^2 + \mathbf{R}_k, \quad (4.52)$$

$$\Sigma_{xy,k|k-1} = \sum_{i=1}^{L_s} \omega_i \mathbf{x}_{i,k|k-1} \mathbf{h}(\mathbf{x}_{i,k|k-1})^\top - \hat{\mathbf{x}}_{k|k-1} (\hat{\mathbf{y}}_{k|k-1})^\top, \quad (4.53)$$

with the transformed sets of sigma-points²

$$\mathbf{x}_{i,k-1|k-1} = \mathbf{S}_{k-1|k-1} \xi_i + \hat{\mathbf{x}}_{k-1|k-1}, \quad (4.54)$$

$$\mathbf{x}_{i,k|k-1} = \mathbf{S}_{k|k-1} \xi_i + \hat{\mathbf{x}}_{k|k-1}. \quad (4.55)$$

We use these approximations to build filters which belong to the family of adaptive sigma-point GFs (ASPGFs). One sigma-point GF approach which has gained significant traction over the past decade is the CKF [87], so called for its use of efficient third-degree spherical-radial cubature rules in defining its set of sigma-points for propagation through the system model, resulting in the $m = 2n_x$ set of cubature points

$$\begin{aligned} \xi_i &= \sqrt{\frac{m}{2}} [1]_i, \\ \omega_i &= \frac{1}{m}, \quad i = 1, 2, \dots, m, \end{aligned} \quad (4.56)$$

where $[1]_i$ is a row vector of length $2n$ of zeros except for the i^{th} and $n + i^{\text{th}}$ position where it is 1 and -1 , respectively. The steps of the CKF may be found in the literature, however use of the CKF

² $\Sigma_{k-1|k-1} = \mathbf{S}_{k-1|k-1} \mathbf{S}_{k-1|k-1}^\top$ and $\Sigma_{k|k-1} = \mathbf{S}_{k|k-1} \mathbf{S}_{k|k-1}^\top$.

with BSE deviates in a way which parallels the deviations with respect to the standard KF in the linear case. Namely, during the measurement update, the innovation covariance as estimated using the cubature rules,

$$\mathbf{P}_{zz,k|k-1} = \frac{1}{m} \sum_{i=1}^m \mathbf{Z}_{i,k|k-1} \mathbf{Z}_{i,k|k-1}^\top - \hat{\mathbf{z}}_{k|k-1} \hat{\mathbf{z}}_{k|k-1}^\top + \mathbf{R}_k \quad (4.57)$$

is substituted by

$$\mathbf{P}_{zz,k|k-1} \approx \Sigma_{y,k|k-1} \approx \hat{\mathbf{S}}_{k,\text{mean}}, \quad (4.58)$$

with the Bayesian estimations of the innovation covariance as in (4.49) and the subsequent Kalman gain computed as in (4.10), thus avoiding the explicit use of \mathbf{R}_k in (4.57).

4.5.2.1 Square-Root Cubature Kalman Filtering (SCKF)

One of the key challenges with implementation of the CKF is the need to preserve the numerical stability of the estimated covariance matrices. In particular, it is imperative that throughout each stage of operation the error covariance matrices stored in the CKF maintain the requisite properties of covariance matrices: symmetry and positive semi-definite. As identified in the literature there are a number of operations in the standard CKF update cycle which are not guaranteed to maintain these properties, including subtracting of positive definite matrices. To address these issues in practical applications, we leverage the so-called square-root cubature Kalman filter (SCKF) [87], with a square-root factorization to avoid such numerical issues. However, we will see that the use of the square-root form requires a different integration of the proposed Bayesian scheme within the filtering equations. In particular, we are interested in the square-root of the predicted error covariance

$$\mathbf{S}_{k|k-1} = \mathbf{Tri}([\mathcal{X}_{k|k-1}^* \mathbf{S}_{Q,k-1}]), \quad (4.59)$$

and the square-root of the innovation covariance

$$\mathbf{S}_{zz,k|k-1} = \mathbf{Tri}([\mathcal{Z}_{k|k-1} \mathbf{S}_{R,k}]), \quad (4.60)$$

as defined in [87]. Note that in the typical SCKF algorithm this factorization of the innovation covariance comes from the measurement propagated, centered cubature points $\mathcal{Z}_{k|k-1}$, as well as the square-root factor $\mathbf{S}_{R,k}$ defined such that $\mathbf{R}_k = \mathbf{S}_{R,k} \mathbf{S}_{R,k}^\top$. We instead wish to use the estimate of the innovation covariance derived from the BSE equations in Section 4.3, the square-root factor of which may be computed as,

$$\mathbf{S}_{zz,k|k-1} = \mathbf{Tri}([\Sigma_{y,k|k-1}]). \quad (4.61)$$

The challenge comes in the measurement update stage of the SCKF, wherein we need to compute the posterior mean and covariance of the estimate, respectively. In the traditional SCKF procedure we compute this as

$$\mathbf{S}_{k|k} = \mathbf{Tri}([\mathcal{X}_{k|k-1} - \mathbf{W}_k \mathcal{Z}_{k|k-1} \quad \mathbf{W}_k \mathbf{S}_{R,k}]) \quad (4.62)$$

where \mathbf{W}_k is the Kalman gain. Note that once again $\mathbf{S}_{R,k}$ appears in the equations, and this poses a particular problem since, for the reasons mentioned above, we wish to avoid using \mathbf{R}_k at all to avoid numerical issues involved with subtracting covariance matrices. In the standard CKF the innovation covariance matrix is computed as in (4.57) which is a particularization of the general expression in (4.52).

One possible approach to solving the problem of maintaining positive definiteness in our estimations of the relevant covariance matrices is through numerical optimization. Namely, we wish to find $\hat{\mathbf{S}}_{R,k}$ which most closely corresponds to the value of $\mathbf{S}_{zz,k|k-1}$ obtained from the BSE³, which may be found by constrained optimization and represented as,

$$\hat{\mathbf{S}}_{R,k} = \arg \min_{\mathbf{S}_{R,k}} \|\Sigma_{y,k|k-1} - \sum_{i=1}^{L_s} \omega_i \mathbf{h}^2(\mathbf{x}_{i,k|k-1}) + (\hat{\mathbf{y}}_{k|k-1})^2 - \mathbf{S}_{R,k} \mathbf{S}_{R,k}^\top\|_F^2 \quad (4.63)$$

$$= \arg \min_{\mathbf{S}_{R,k}} \|\Sigma_{y,k|k-1} - \frac{1}{m} \sum_{i=1}^m \mathbf{z}_{i,k|k-1} \mathbf{z}_{i,k|k-1}^\top + \hat{\mathbf{z}}_{k|k-1} \hat{\mathbf{z}}_{k|k-1}^\top - \mathbf{S}_{R,k} \mathbf{S}_{R,k}^\top\|_F^2. \quad (4.64)$$

4.6 Simulation-Based Experimental Validation

We tested the proposed Bayesian statistics augmentation approach through simulation-based testing using both linear and nonlinear models, and for both increasing and decreasing changes in the measurement noise covariance. To this end, we identified and developed simulated representations of both a multivariate linear system and a multivariate system with nonlinear observations. The method was evaluated using several key metrics related to both the performance of the encapsulating GF and that of the BSE, the most important of which was the average, or RMSE between the true innovation covariance \mathbf{S}_k and the innovation covariance $\hat{\mathbf{S}}_k$ as estimated by the Bayesian covariance estimator. Recalling from (4.8) that the innovation covariance of interest is related to the predictive state error covariance and measurement noise covariance by way of the measurement model, we find that for simplicity of comparison it is useful in this case to eliminate the component which is related to the state error and compare the difference to the value of \mathbf{R}_k used in generating the noisy observations.

³ $\mathbf{P}_{zz,k|k-1} = \mathbf{S}_{zz,k|k-1} \mathbf{S}_{zz,k|k-1}^\top$

For the linear experiments, a standard KF was implemented and used as the baseline filter to generate the innovation sequence that BSE is using to estimate the covariance of the noise term. When the method was assessed under the nonlinear system, a CKF approach was used for this same purpose. Performance of the proposed method was compared against one of the most popular online covariance estimation methodologies found in the literature [65], hereafter referred to as the Myers' method, which operates similarly to a windowed moving sample average. For completeness, this control methodology was implemented twice in each experiment: once with a short, more adaptive window length of $L = 8$ samples, and once with a longer, more stable window length of $L = 100$ samples. The proposed methodology was implemented with the window for the time-averaged NIS threshold check set to a length of $K_d = 8$ samples. In each case, both control filters and the proposed methodology were evaluated only during their "active" iterations (e.g., performance of the $L = 100$ Myers approach was only evaluated from time $t = 100$: the point at which the method has access to a full window length worth of samples from which to produce an estimate). Note that each model under consideration uses a sampling rate of $T_s = 1$ seconds resulting in the time index $t = kT_s$.

Half way through each experiment the system is exposed to a sudden increase or decrease in the signal-to-noise ratio of the observations, leading to an instantaneous change in the measurement noise covariance. The respective methodologies then are required to adapt to the change in the observation quality throughout continued operation of the encapsulating GF. This process was repeated with a given ground-truth trajectory for the state, and with independently redrawn measurements and state-estimation filter initialization for $N = 2000$ Monte Carlo trials, and the results then averaged. In this section, we will discuss the average performance of the proposed methodology in comparison to the Myers' methodologies for each of these experiments.

4.6.1 Linear/Gaussian Multivariate System

For our evaluation of the BSE methodology under a linear multivariate system, we consider a simple two-dimensional constant velocity motion problem, with position (ξ and η) and velocity ($\dot{\xi}$ and $\dot{\eta}$) composing the state vector $\mathbf{x}_k = (\xi_k, \dot{\xi}_k, \eta_k, \dot{\eta}_k)^\top$, at instant k . The state evolves according to the state transition equation,

$$\mathbf{x}_k = \begin{pmatrix} 1 & 0 & T_s & 0 \\ 0 & 1 & 0 & T_s \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \mathbf{x}_{k-1} + \boldsymbol{\nu}_{k-1}, \quad (4.65)$$

where $\boldsymbol{\nu}_{k-1} \sim \mathcal{N}(\mathbf{0}, 10^{-6}\mathbf{I})$ is the time-invariant process noise, which is independently and identically distributed for each k . The initial condition is $\mathbf{x}_0 = (100m, 5m/s, -100m, 2m/s)^\top$. The position of the system is observed directly, which is represented as,

$$\mathbf{y}_k = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mathbf{x}_k + \mathbf{n}_k, \quad (4.66)$$

where $\mathbf{n}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$ and $\mathbf{R}_0 = \text{diag}(1, 0.5)$ is the initial value of the measurement covariance matrix we aim at indirectly estimating through the use of the proposed Bayesian and Myers methodologies.

Estimation of the state of this system was performed with an instance of a linear KF corresponding to the proposed Bayesian and control Myers methodologies, each with access to the true state transition and measurement equations. Over the course of an initial transient period of $T_n = 10s$ for the Bayesian estimator or $T_n = 10 + L/T_s$ for the Myers methods, it is assumed that the filter has access to the true initial estimate of the noise covariance. After this transient period, each filter has access only to the estimates provided by the embedded Bayesian or Myers method respectively. Each filter state was initialized randomly for each Monte Carlo trial, with mean drawn from a normal distribution with mean given by the true initial state \mathbf{x}_0 and state covariance $\mathbf{P}_0 = \text{diag}(20, 0.1, 20, 0.1)$.

Figure 4.6.1 shows the average output of the Myers and BSE implementations over the $N = 2000$ Monte Carlo trials across the $T = 200s$ duration of the experiment. Figure 4.6.2 shows the corresponding RMSE with respect to the measurement noise covariance diagonal true value. The Myers method with $L = 8$ samples demonstrates a high degree of variability throughout the course of the experiment, in contrast to the $L = 100$ Myers method which is slow to transition, but precisely attains and remains at the true value of the measurement covariance towards the end of the experiment duration. The proposed Bayesian methodology, which experiences a reset according to the time-averaged NIS threshold, shown averaged in Figure 4.6.4 before quickly and stably converging towards the new ground-truth measurement covariance. Figure 4.6.3 shows the error distributions, both with respect to the state estimation and the covariance estimation, averaged over the entire experiment, which reflects the ‘‘best of both worlds’’ benefit to the proposed approach resulting in a lower overall error distribution. It is worth noting, however, that for the linear model the precision of the innovation covariance provided to KF had a relatively low impact on the state estimation performance of that filter, with the encapsulating KF performing similarly for all three control and proposed approaches.

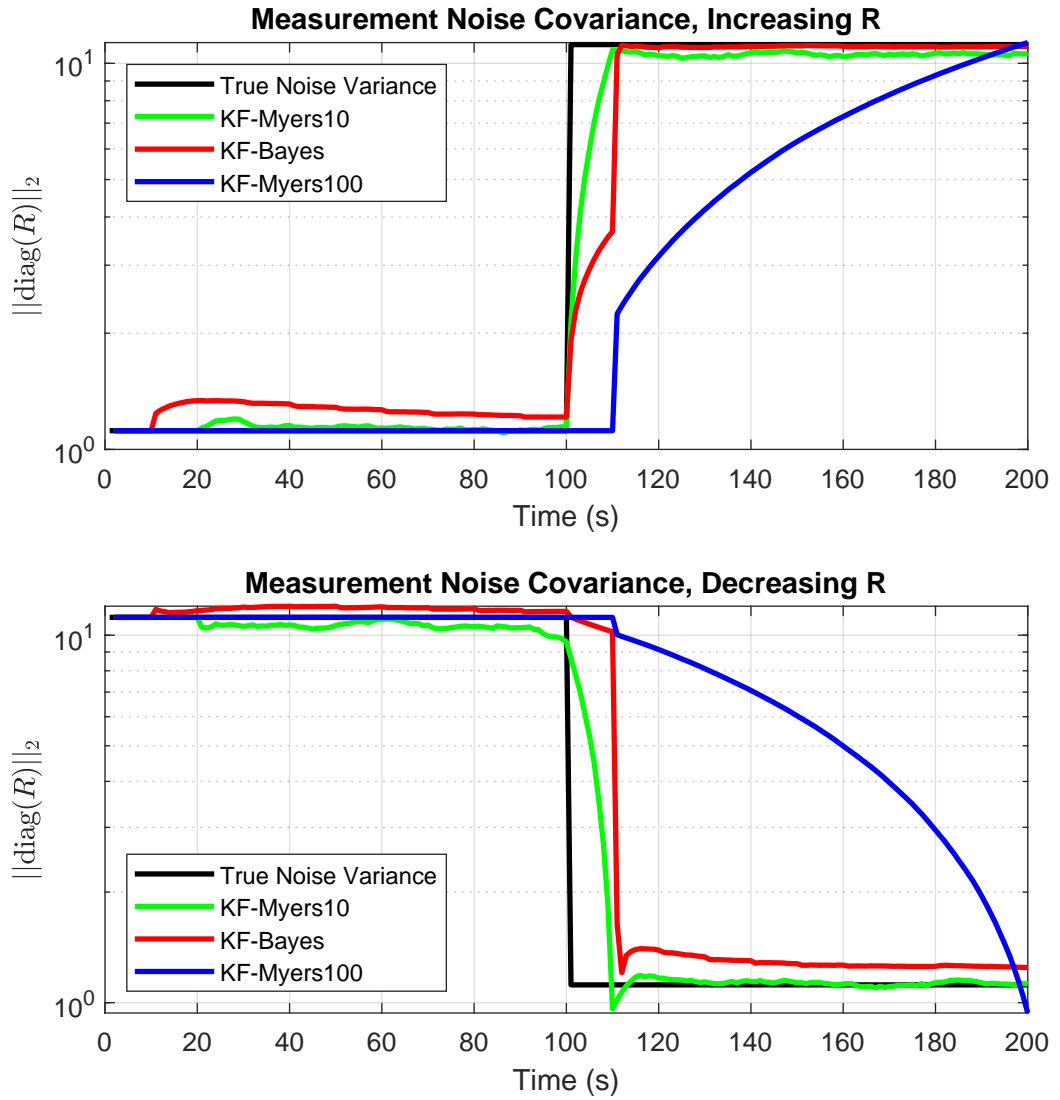


Figure 4.6.1: Measurement noise covariance estimates. Experiments shown represent a linear/-Gaussian multivariate system with step increase in measurement noise (top) and step decrease in measurement noise (bottom), respectively. Comparison of the true model covariance with Myers approach ($L = 8$, $L = 100$), and the proposed Bayesian methodology with both informative priors and time-averaged NIS re-initialization.

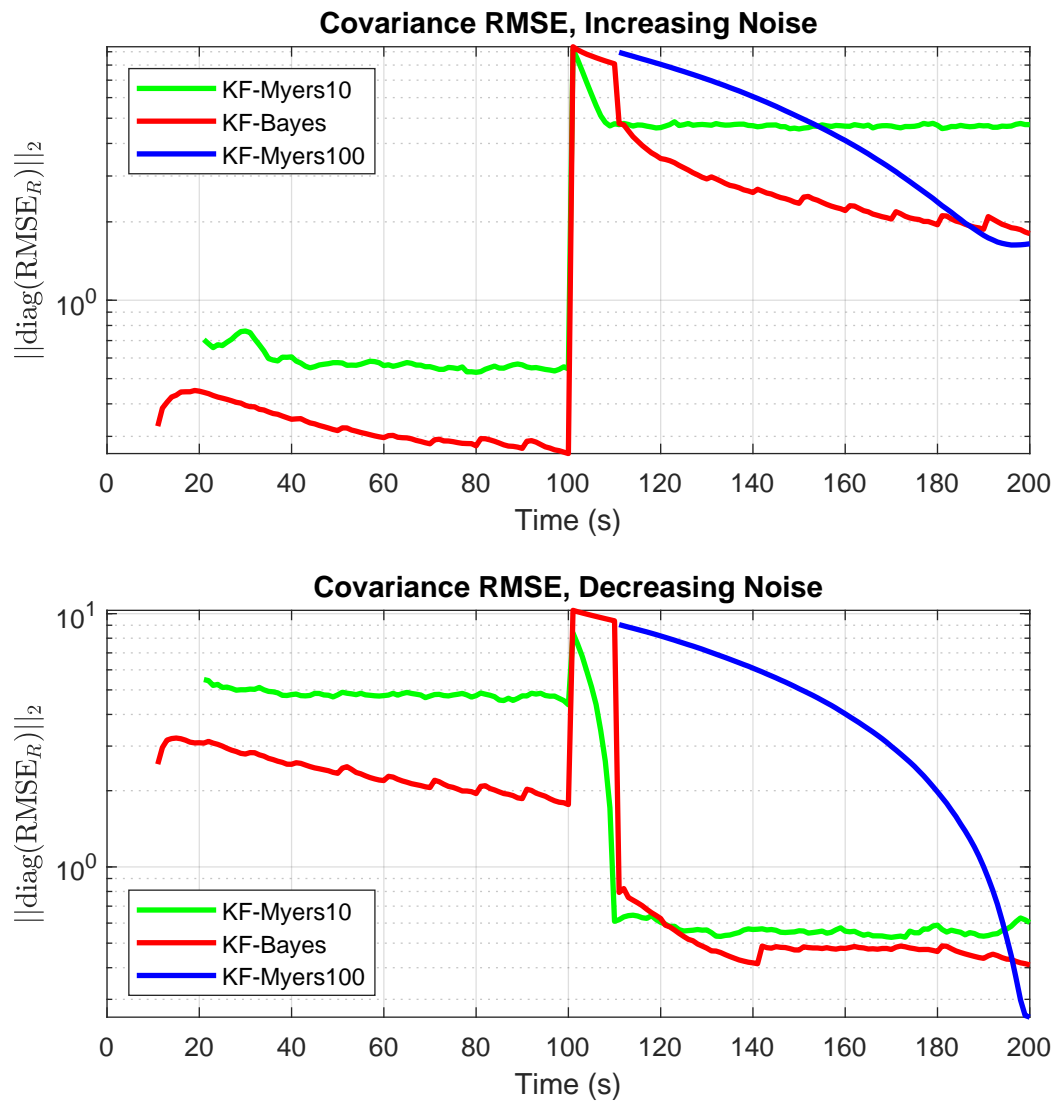


Figure 4.6.2: Measurement noise covariance RMSE. Experiments shown represent a linear/Gaussian multivariate system with step increase in measurement noise (top) and step decrease in measurement noise (bottom), respectively.

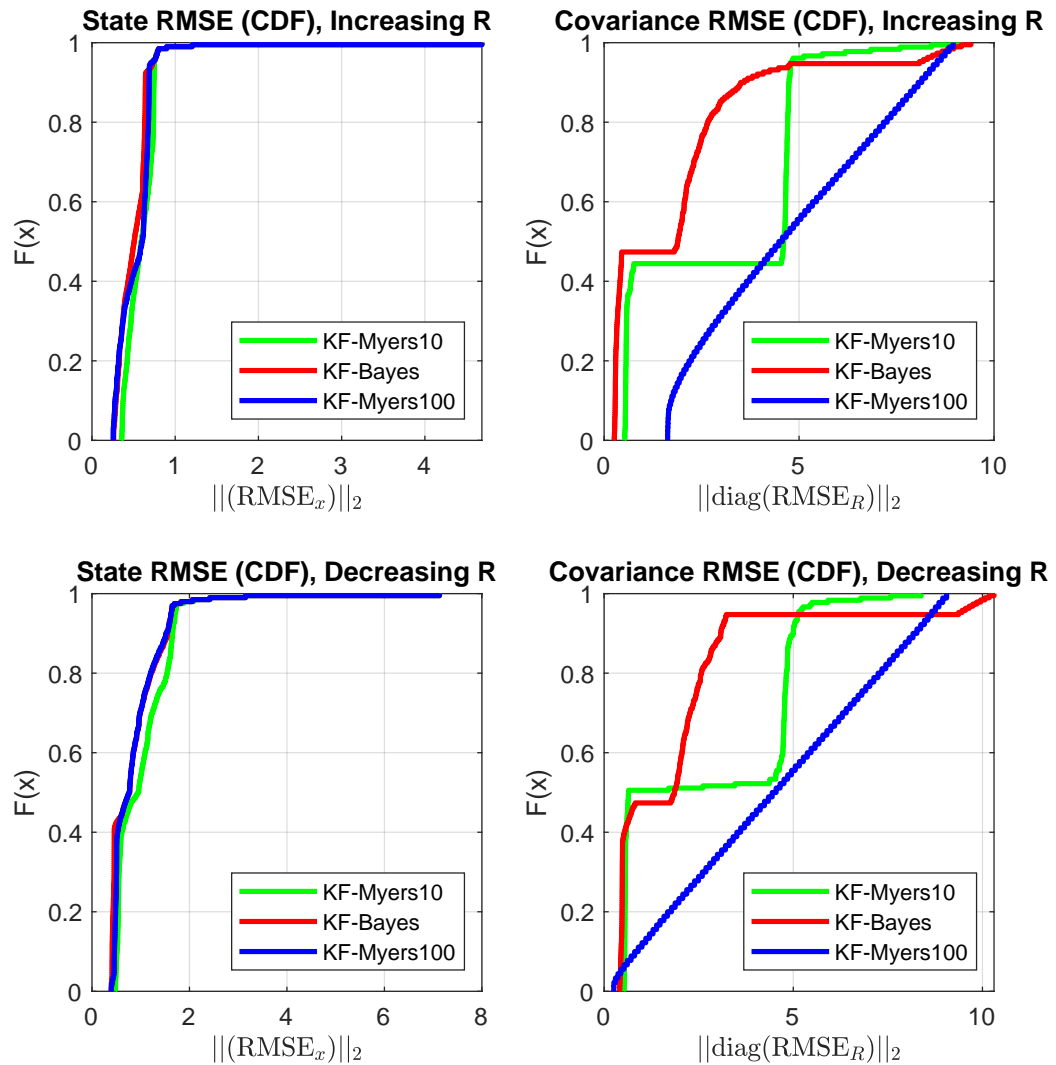


Figure 4.6.3: Measurement noise covariance RMSE cumulative distribution (CDF). Experiments shown represent a linear/Gaussian multivariate system with step increase in measurement noise (top) and step decrease in measurement noise (bottom), respectively. Plots shown for both the state (left) and measurement noise covariance (right) estimation errors.

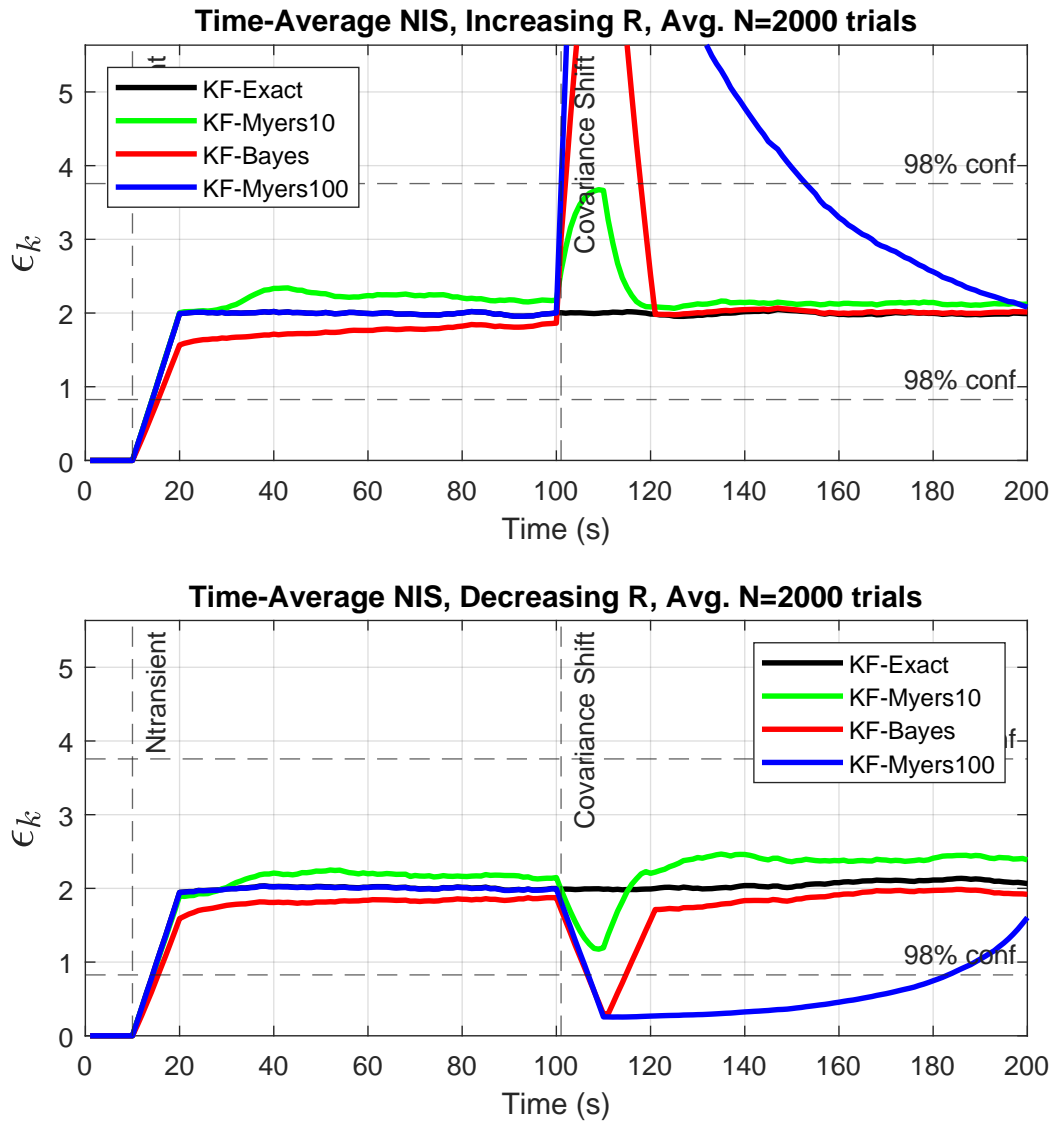


Figure 4.6.4: Time-averaged Normalized Innovation Squared (NIS). Experiments shown represent a linear/Gaussian multivariate system with step increase in measurement noise (top) and step decrease in measurement noise (bottom), respectively.

4.6.2 Nonlinear/Gaussian Multivariate System

For evaluation under a nonlinear multivariate system, we consider an aircraft tracking problem using radar observables identical to that used to demonstrate the efficacy of the CKF in [87]. The behavior of the aircraft under observation is characterized by the elements of the state vector $\mathbf{x}_k = (\xi_k, \dot{\xi}_k, \eta_k, \dot{\eta}_k, \Omega)^\top$, with ξ_k , $\dot{\xi}_k$, η_k , and $\dot{\eta}_k$ representing the same 2-D position and velocity as in the linear example, and Ω representing the turn rate in radians per second. The state evolves according to the state transition equation

$$\mathbf{x}_k = \begin{pmatrix} 1 & \frac{\sin \Omega T_s}{\Omega} & 0 & -\frac{1-\cos \Omega T_s}{\Omega} & 0 \\ 0 & \cos \Omega T_s & 0 & -\sin \Omega T_s & 0 \\ 0 & \frac{1-\cos \Omega T_s}{\Omega} & 1 & \frac{\sin \Omega T_s}{\Omega} & 0 \\ 0 & \sin \Omega T_s & 0 & \cos \Omega T_s & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \mathbf{x}_{k-1} + \boldsymbol{\nu}_{k-1} \quad (4.67)$$

where $\boldsymbol{\nu}_{k-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ is once again the independently and identically distributed process noise for each update⁴. The initial condition is $\mathbf{x}_0 = (1000m, 300m/s, 1000m, 0m/s, -3 \text{ deg } s^{-1})^\top$. The position of the aircraft in this system is observed indirectly through the radar range r and bearing θ , which is represented in the measurement equation as,

$$\mathbf{y}_k = \begin{pmatrix} r_k \\ \theta_k \end{pmatrix} = \begin{pmatrix} \sqrt{\xi_k^2 + \eta_k^2} \\ \tan^{-1}(\frac{\eta_k}{\xi_k}) \end{pmatrix} + \mathbf{n}_k, \quad (4.68)$$

where $n_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$ and $\mathbf{R}_0 = \text{diag}(100, 10)$.

An instance of the SCKF was created for each of the covariance estimation methods considered and instantiated with the same mean drawn from $\hat{\mathbf{x}}_0 \sim \mathcal{N}(\mathbf{x}_0, \mathbf{P}_0)$ and covariance $\mathbf{P}_0 = \text{diag}(100m^2, 10m^2s^{-2}, 100m^2, 10m^2s^{-2}, 100\text{mrad}^2s^{-2})$, and with access to the true initial measurement covariance \mathbf{R}_0 for the transient period before switching to the Myers or Bayesian estimate. After this initial transient period, which was configured in the same way as in the linear case, each filter was made to rely only on the internal estimate given by the respective statistics estimation method.

Figure 4.6.5 shows the responses of the implementations of the proposed and control methodologies to the change in measurement noise covariance, while Figure 4.6.6 shows the RMSE with respect to the measurement noise covariance diagonal true value. As predicted, the Myers

⁴ \mathbf{Q} is the nonsingular, block diagonal matrix $\mathbf{Q} = \text{diag}[q_1\mathbf{M}, q_1\mathbf{M}, q_2T_s]$ with $\mathbf{M} = \begin{pmatrix} T_s^3/3 & T_s^2/2 \\ T_s^2/2 & T_s \end{pmatrix}$ and $q_1 = 0.1m^2s^{-3}$, $q_2 = 1.75 \times 10^{-4}s^{-3}$.

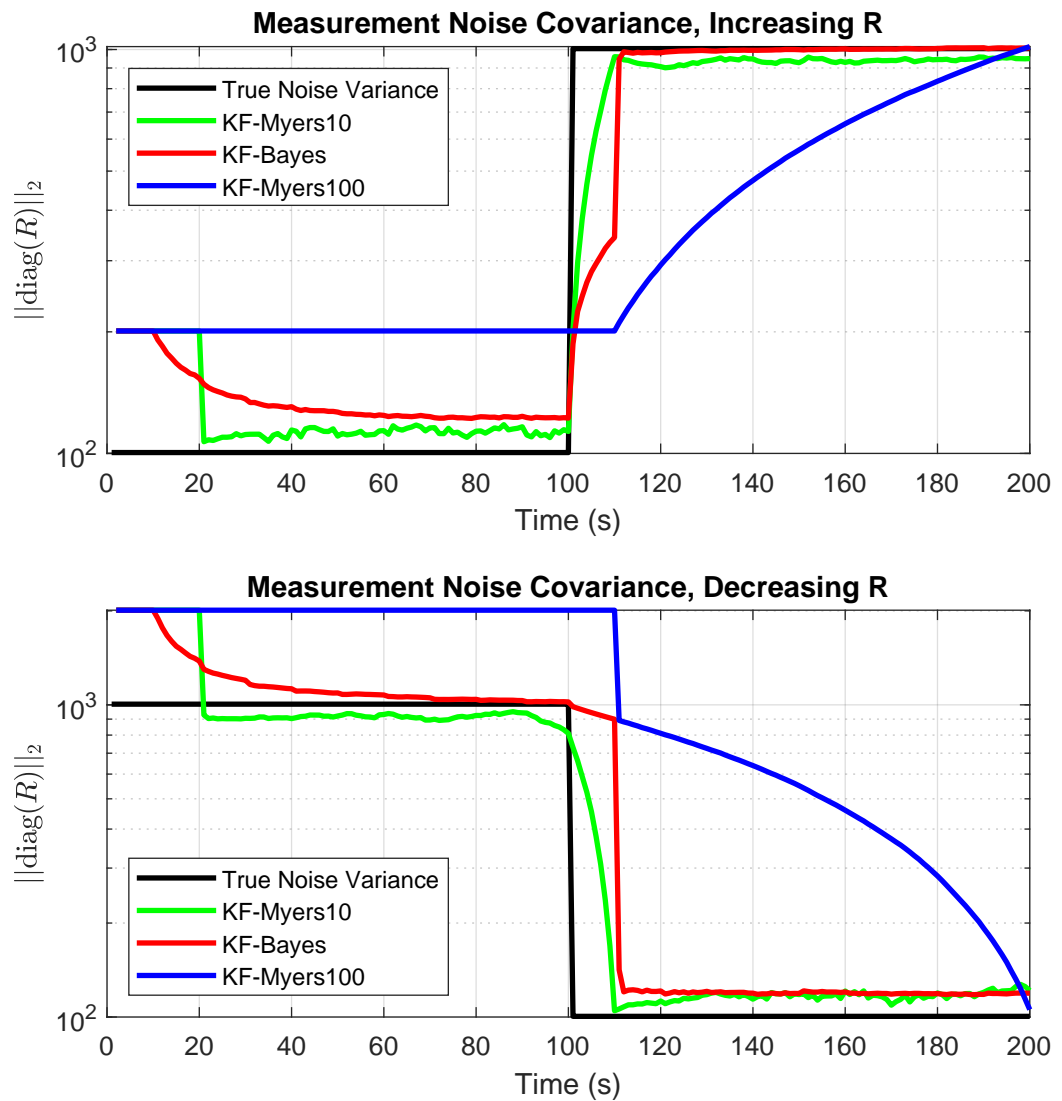


Figure 4.6.5: Measurement noise covariance estimates. Experiments shown represent a nonlinear/Gaussian radar system system with step increase in measurement noise (top) and step decrease in measurement noise (bottom), respectively. Comparison of the true model covariance with Myers approach ($L = 8$, $L = 100$), and the proposed Bayesian methodology with both informative priors and time-averaged NIS reinitialization.

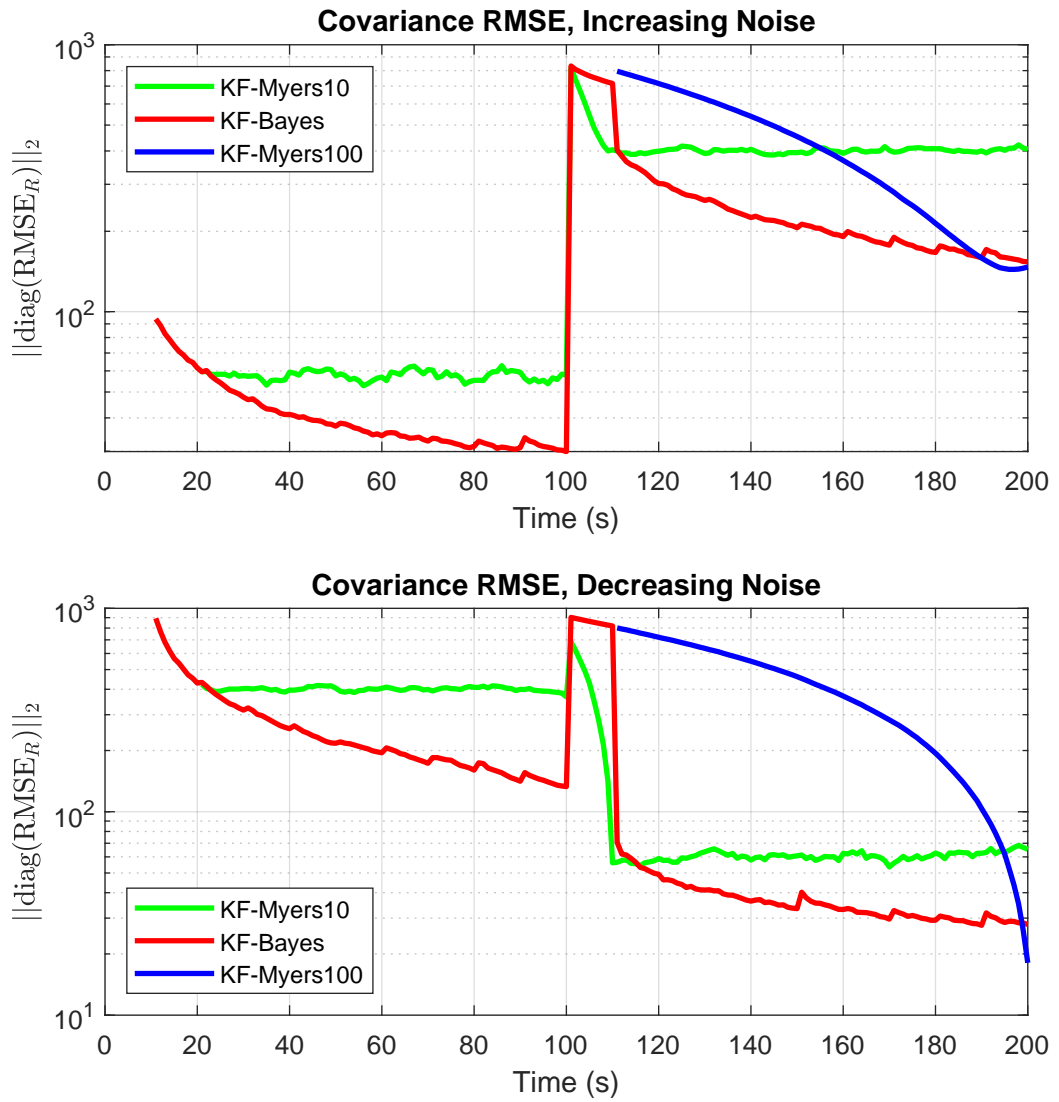


Figure 4.6.6: Measurement noise covariance root mean squared error (RMSE). Experiments shown represent a nonlinear/Gaussian radar system system with step increase in measurement noise (top) and step decrease in measurement noise (bottom), respectively.

method with $L = 8$ samples shows the most rapid response to the model change, with the Bayesian methodology being reset to a comparable value at $t = 110$ using the time-averaged NIS methodology, shown again averaged over the Monte Carlo trials in Figure 4.6.8. Finally, the Myers method with $L = 100$ has the slowest, but smoothest response as its window gradually shifts into the new regime. Figure 4.6.7 shows the error distributions, both with respect to the state estimation and the covariance estimation, averaged over the entire experiment. The proposed methodology performed well in terms of the covariance estimation error both before and after the noise distribution change, leading to an overall better performance over the $T = 200s$ duration of the experiment. This resulted in an expected and marked gain in state estimator performance over the $L = 100$ Myers approach and a smaller gain over the $L = 8$ Myers approach.

4.7 Conclusions

Our experiments with the proposed methodology in a simulated testing environment demonstrated a number of strengths and a few key weaknesses of the proposed Bayesian statistics estimation methodology. Many of these advantages and disadvantages stem from the synergy between the Bayesian covariance estimate and the time-averaged NIS based re-initialization. This synergistic operation allows the filter to respond quickly to changes in the noise environment of the filter, while converging to more precise estimates during periods where the noise environment remains stationary. This approach closely ties the behavior of the Bayesian statistics estimation to the performance of the Gaussian state estimation filter, and thus the performance of the system overall is dependent on the quality and stability of the system being tracked. Under optimal filtering conditions this is ideal for obtaining a high degree of precision both in noise and state estimation, however this also reveals a weakness to this approach. Namely, that when the filter is performing sub-optimally due to a poor characterization on the part of the system model or because of highly nonlinear dynamics, the properties of the innovations sequence that the Bayesian statistics estimate relies on are no longer guaranteed which can lead to instability of the estimator system. Additionally, we note that the Bayesian covariance methodology is principally designed for estimating the statistics of stationary distributions, and the addition of the time-averaged NIS based resetting approach to overcome this limitation adds complexity to the system as well as new “tuning parameters” that must be adjusted to get the best possible performance.

In spite of this, we have demonstrated through these experiments that the methodology of applying Bayesian statistics estimation to account for unknown or uncertain measurement noise

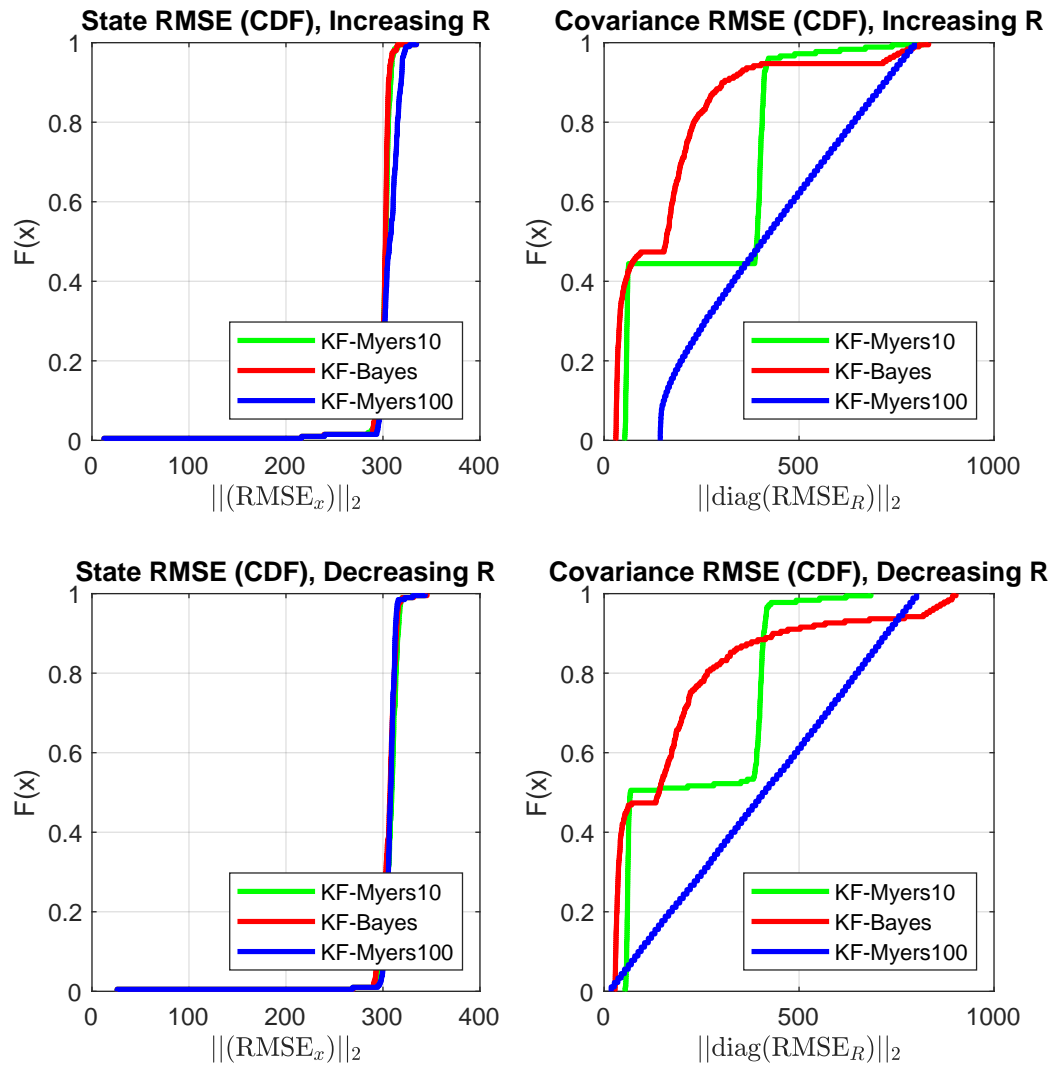


Figure 4.6.7: Measurement noise covariance RMSE cumulative distribution (CDF). Experiments shown represent a nonlinear/Gaussian radar system with step increase in measurement noise (top) and step decrease in measurement noise (bottom), respectively. Plots shown for both the State (left) and measurement noise covariance (right) estimation errors.

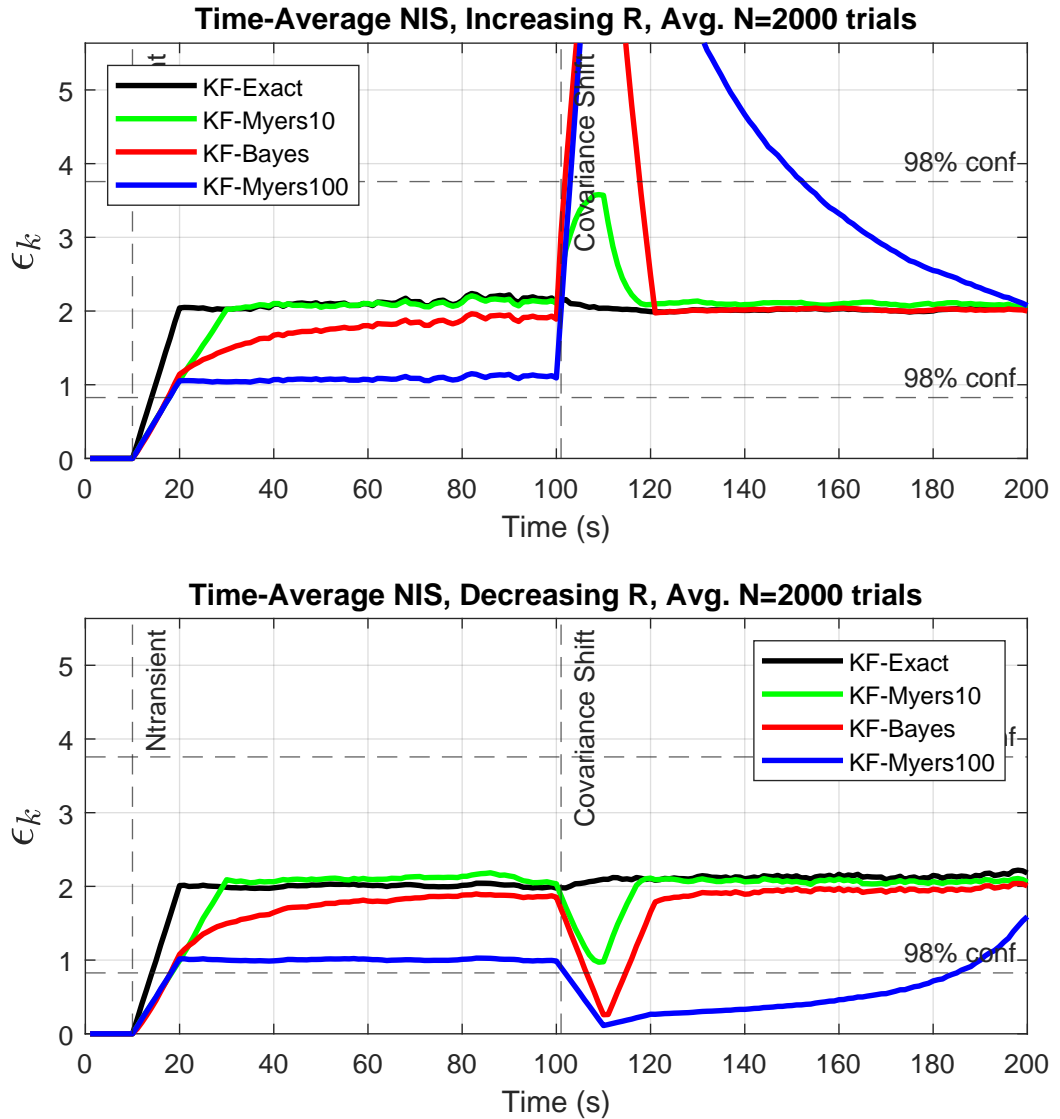


Figure 4.6.8: Time-averaged Normalized Innovation Squared (NIS). Experiments shown represent a nonlinear/Gaussian radar system with step increase in measurement noise (top) and step decrease in measurement noise (bottom), respectively.

CHAPTER 4. BAYESIAN STATISTICS ESTIMATION IN GAUSSIAN FILTERING

characteristics may be applied to both linear and nonlinear KF algorithms, including the standard linear KF and SCKF. Throughout our experiments with both linear dynamic and nonlinear radar based models, our approach successfully adapted to changes in the noise characteristics of the observed data and adjusted accordingly to provide better filtering performance. We conclude that this method of using Bayesian statistics estimation within dynamic Gaussian filtering techniques has significant potential, and future work will be dedicated to refining this method for use in a wider range of nonlinear dynamic state estimation scenarios.

Chapter 5

A Bayesian Pseudotrue State for Dynamic State Estimation

Dynamic state estimation plays a pivotal role in numerous civilian and military applications, where estimator failures can be costly. It is thus of critical importance to be able to analyze the performance of dynamic state estimators under model misspecification. Although the concept of a pseudotrue parameter provides a valuable benchmark for analyzing model misspecification in static estimation scenarios, a corresponding framework for dynamic state estimation has yet to be established. In this work, we introduce the notion of a “pseudotrue state” to bridge this gap. By extending the principles of static misspecification analysis to the dynamic context, we derive a solution for the pseudotrue state based on the true and assumed models. Leveraging the Kullback–Leibler divergence (KLD) that underpins the pseudotrue parameter, we obtain a recursive, analytic expression for the pseudotrue state in linear Gaussian state space models.

5.1 Dynamic Model Misspecification: Applications and Challenges

Dynamic state estimation problems occur in diverse fields of technology, from sophisticated radar defense systems to simple GPS trackers. Estimators designed to address dynamic problems require an accurate model of the system dynamics and noise characteristics. In the case that this requirement is not met, the dynamic nature of the problem results in compounding errors and ultimate failure of the estimator. Evaluating the behavior of the estimator in response to various types of model misspecification is crucial to the security of these systems; however, existing misspecification analysis methods are applicable mainly to static parameter estimation, with similar concepts in dynamic state estimation remaining an open challenge. In static parameter estimation literature, the pseudotrue parameter provides a benchmark against which one can measure estimator performance

under model misspecification. In this contribution, we extend this idea to the dynamic context, introducing the notion of a “pseudotrue state” applicable to a wide range of dynamic state estimation problems, and demonstrating its realizations in Gaussian filters. This conceptual framework provides a direct benchmark for understanding and evaluating estimator performance when the assumed dynamic model deviates from reality. We obtain a solution for the pseudotrue state based on the true and assumed models that is a function of the true state. We identify a recursive, analytic expression for this pseudotrue state for linear Gaussian state space models. In the following sections, we detail the derivation of the pseudotrue state from the Kullback–Leibler divergence (KLD) between dynamic models, provide analytic expressions for linear Gaussian state space models, and demonstrate its utility in understanding estimator performance under model misspecification.

5.2 On Parametric Misspecification for Bayesian Estimation

Given dynamic state-transition and measurement models of the form

$$\mathcal{M}_* : \mathbf{x}_n = h^*(\boldsymbol{\psi}_n) + \mathbf{v}_n, \quad \mathbf{v}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_n^*) \quad (5.1)$$

$$\boldsymbol{\psi}_n = f^*(\boldsymbol{\psi}_{n-1}) + \mathbf{w}_{n-1}^*, \quad \mathbf{w}_{n-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{n-1}^*). \quad (5.2)$$

where $\boldsymbol{\psi}$ represents a hidden state that evolves dynamically over the discrete time index n according to the deterministic process equation f^* and additive process noise $\mathbf{w}_{n-1}^* \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{n-1}^*)$, and is observable only through the measurement \mathbf{x} according to the deterministic measurement equation h^* and additive measurement noise $\mathbf{v}_n^* \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_n^*)$. We consider a situation where a system designer has erroneously come to understand the measurements \mathbf{x} as having a relationship to a hypothesized dynamic hidden state $\boldsymbol{\theta}$, and wishes to produce online estimates of the quantity $\boldsymbol{\theta}$ at each n . To do so, this designer produces an “assumed model” around this quantity of the form

$$\mathcal{M} : \mathbf{x}_n = h(\boldsymbol{\theta}_n) + \mathbf{v}_n, \quad \mathbf{v}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_n) \quad (5.3)$$

$$\boldsymbol{\theta}_n = f(\boldsymbol{\theta}_{n-1}) + \mathbf{w}_{n-1}, \quad \mathbf{w}_{n-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{n-1}) \quad (5.4)$$

where $\boldsymbol{\theta}$ represents the a hypothesized state that evolves dynamically according to model parameters that are analogous to that of the true model \mathcal{M}_* . We note that although the assumed hidden state evolves over the same time index n as the true hidden state $\boldsymbol{\psi}$, it does not have any assumed relationship to $\boldsymbol{\psi}$, apart from the relationship that each has with respect to the observable quantity \mathbf{x} . Our objective with this contribution is to identify a method of characterizing the behavior of

recursive Bayesian estimators developed under these conditions, across various classes of model misspecification between the true and assumed models.

5.3 The Posterior Pseudotrue State for Dynamic State Estimation

The so-called ‘‘pseudotrue parameter’’ for misspecified models is defined as a unique interior point of the assumed parameter space Θ that minimizes the KLD between the probability density functions of the measurements given by the true and assumed models [1, 44, 46, 90, 91]. Typically, this is defined in terms of the likelihood of the measurements given the assumed parameter, $f(\mathbf{x}|\boldsymbol{\theta})$, and the unconditional true distribution of the measurements, $p(\mathbf{x})$.

$$\boldsymbol{\theta}^0(\boldsymbol{\psi}) = \arg \min_{\boldsymbol{\theta}} \mathcal{D}\left(p(\mathbf{x}|\boldsymbol{\psi})||f(\mathbf{x})\right). \quad (5.5)$$

However, this approach is less suitable in a Bayesian context, where the parameters are influenced by both *a priori* information and the likelihood of the measurements. More suitable is the Bayesian pseudotrue parameter associated with the misspecified Bayesian Cramér-Rao lower bound (MBCRB) defined as [18]

$$\boldsymbol{\theta}_1(\boldsymbol{\psi}) = \arg \min_{\boldsymbol{\theta}} \mathcal{D}\left(p(\mathbf{x}|\boldsymbol{\psi})||f(\mathbf{x}, \boldsymbol{\theta})\right). \quad (5.6)$$

To adapt this definition of the Bayesian pseudotrue parameter for recursive Bayesian estimators and dynamic state estimation problem, we need to expand this definition to incorporate all measurements and all states up to the current time n , rather than just a single one. Rather than attempting to define a posterior pseudotrue parameter that also spans all measurements and all states, we propose an iterative approach to computing the pseudotrue parameter of the current state in terms of all previous information.

$$\boldsymbol{\theta}_n^0(\boldsymbol{\psi}_{0:n}) = \arg \min_{\boldsymbol{\theta}_n} \mathcal{D}\left(p(\mathbf{x}_{0:n}|\boldsymbol{\psi}_{0:n})||f(\mathbf{x}_{0:n}, \boldsymbol{\theta}_n)\right) \quad (5.7)$$

$$= \arg \min_{\boldsymbol{\theta}_n} \mathbb{E}_{\mathbf{x}_{0:n}|\boldsymbol{\psi}_{0:n}} \left\{ -\log f(\mathbf{x}_{0:n}, \boldsymbol{\theta}_n) \right\} \quad (5.8)$$

Although this is not the only possible definition of the pseudotrue state, the inclusion of a particular realization of the hidden state across all previous samples $\boldsymbol{\psi}_{0:n}$ makes it particularly well suited for characterizing the performance of recursive Bayesian estimators in practical applications.

It is possible to obtain an analytic solution to this minimization problem by breaking down the joint density function $f(\mathbf{x}_{0:n}, \boldsymbol{\theta}_n)$ into composite densities. From the properties of the joint

density function,

$$f(\mathbf{x}_{0:n}, \boldsymbol{\theta}_n) = p(\mathbf{x}_n | \mathbf{x}_{0:n-1}, \boldsymbol{\theta}_n) p(\mathbf{x}_{0:n-1}, \boldsymbol{\theta}_n) \quad (5.9)$$

$$= p(\mathbf{x}_n | \mathbf{x}_{0:n-1}, \boldsymbol{\theta}_n) p(\boldsymbol{\theta}_n | \mathbf{x}_{0:n-1}) p(\mathbf{x}_{0:n-1}). \quad (5.10)$$

By Markovianity, \mathbf{x}_n is independent of previous measurements $\mathbf{x}_{0:n-1}$ conditioned on $\boldsymbol{\theta}_n$, therefore

$$p(\mathbf{x}_{0:n}, \boldsymbol{\theta}_n) = p(\mathbf{x}_n | \mathbf{x}_{0:n-1}, \boldsymbol{\theta}_n) p(\boldsymbol{\theta}_n | \mathbf{x}_{0:n-1}) p(\mathbf{x}_{0:n-1}) \quad (5.11)$$

$$= p(\mathbf{x}_n | \boldsymbol{\theta}_n) p(\boldsymbol{\theta}_n | \mathbf{x}_{0:n-1}) p(\mathbf{x}_{0:n-1}). \quad (5.12)$$

The middle term $p(\boldsymbol{\theta}_n | \mathbf{x}_{0:n-1})$ may be expressed in the form of a Chapman-Kolmogorov integral

$$p(\boldsymbol{\theta}_n | \mathbf{x}_{0:n-1}) = \int p(\boldsymbol{\theta}_n | \boldsymbol{\theta}_{n-1}) p(\boldsymbol{\theta}_{n-1} | \mathbf{x}_{0:n-1}) d\boldsymbol{\theta}_{n-1}. \quad (5.13)$$

Thus, the joint density function of the measurements and assumed state $\boldsymbol{\theta}_n$ can be written as

$$f(\mathbf{x}_{0:n}, \boldsymbol{\theta}_n) = p(\mathbf{x}_n | \boldsymbol{\theta}_n) \int p(\boldsymbol{\theta}_n | \boldsymbol{\theta}_{n-1}) p(\boldsymbol{\theta}_{n-1} | \mathbf{x}_{0:n-1}) d\boldsymbol{\theta}_{n-1} p(\mathbf{x}_{0:n-1}) \quad (5.14)$$

applying the properties of the logarithm in (5.7) and leveraging the fact that the trailing term $p(\mathbf{x}_{0:n-1})$ has no dependency on the optimization parameter $\boldsymbol{\theta}_n$, the expression for the pseudotrue state may be rewritten as

$$\begin{aligned} \boldsymbol{\theta}_n^0(\boldsymbol{\psi}_{0:n}) = \arg \min_{\boldsymbol{\theta}_n} \mathbb{E}_{\mathbf{x}_{0:n} | \boldsymbol{\psi}_{0:n}} \left\{ -\log p(\mathbf{x}_n | \boldsymbol{\theta}_n) \right\} \\ - \mathbb{E}_{\mathbf{x}_{0:n} | \boldsymbol{\psi}_{0:n}} \left\{ \log \int p(\boldsymbol{\theta}_n | \boldsymbol{\theta}_{n-1}) p(\boldsymbol{\theta}_{n-1} | \mathbf{x}_{0:n-1}) d\boldsymbol{\theta}_{n-1} \right\}. \end{aligned} \quad (5.15)$$

5.3.1 Computing the Posterior Pseudotrue State for Linear Gaussian Systems

Although there is no general closed-form solution to this integral and subsequent optimization, in the case that the assumed state transition and posterior distributions are linear and Gaussian with

$$f(\boldsymbol{\theta}_n | \boldsymbol{\theta}_{n-1}) = \mathcal{N}(\boldsymbol{\theta}_{n-1}, \mathbf{Q}_{n-1}) \quad (5.16)$$

$$f(\boldsymbol{\theta}_{n-1} | \mathbf{x}_{0:n-1}) = \mathcal{N}(\tilde{\boldsymbol{\theta}}_{n-1}, \tilde{\boldsymbol{\Sigma}}_{n-1}), \quad (5.17)$$

the integral has a closed-form solution

$$\int f(\boldsymbol{\theta}_n | \boldsymbol{\theta}_{n-1}) f(\boldsymbol{\theta}_{n-1} | \mathbf{x}_{0:n-1}) d\boldsymbol{\theta}_{n-1} = \mathcal{N}(\mathbf{F}\tilde{\boldsymbol{\theta}}_{n-1}, \mathbf{F}\tilde{\boldsymbol{\Sigma}}_{n-1}\mathbf{F}^\top + \mathbf{Q}_{n-1}) = \mathcal{N}(\tilde{\boldsymbol{\theta}}_n, \tilde{\boldsymbol{\Sigma}}_n). \quad (5.18)$$

Using these same linearity and Gaussianity assumptions, the measurement likelihood is Gaussian with

$$f(\mathbf{x}_n|\boldsymbol{\theta}_n) = \mathcal{N}(\mathbf{H}\boldsymbol{\theta}_n, \mathbf{R}) \quad (5.19)$$

The log distribution of the assumed model then is given as the sum of the respective terms

$$\begin{aligned} \log f(\mathbf{x}_{0:n}, \boldsymbol{\theta}_n) &= \log p(\mathbf{x}_{0:n-1}) + \log \mathcal{N}(\mathbf{H}\boldsymbol{\theta}_n, \mathbf{R}) \\ &+ \log \mathcal{N}(\mathbf{F}\tilde{\boldsymbol{\theta}}_{n-1}, \mathbf{F}\tilde{\boldsymbol{\Sigma}}_{n-1}\mathbf{F}^\top + \mathbf{Q}_{n-1}) \end{aligned} \quad (5.20)$$

where, as in (5.15), $\log p(\mathbf{x}_{0:n-1})$ is constant with respect to the hypothesized state $\boldsymbol{\theta}_n$ and can therefore be ignored. It follows that, under the aforementioned linearity and Gaussianity assumptions, the pseudotrue stat at time instance n can be determined by solving the optimization:

$$\begin{aligned} \boldsymbol{\theta}_n^0(\boldsymbol{\psi}_{0:n}) &= \arg \min_{\boldsymbol{\theta}} \mathbb{E}_{\mathbf{x}_{0:n}|\boldsymbol{\psi}_{0:n}} \left\{ \frac{1}{2}(\mathbf{x}_n - \mathbf{H}\boldsymbol{\theta}_n)^\top \mathbf{R}^{-1}(\mathbf{x}_n - \mathbf{H}\boldsymbol{\theta}_n) \right. \\ &\quad \left. + \frac{1}{2}(\boldsymbol{\theta}_n - \mathbf{F}\tilde{\boldsymbol{\theta}}_{n-1})^\top (\mathbf{F}\tilde{\boldsymbol{\Sigma}}_{n-1}\mathbf{F}^\top + \mathbf{Q}_{n-1})^{-1}(\boldsymbol{\theta}_n - \mathbf{F}\tilde{\boldsymbol{\theta}}_{n-1}) \right\} \end{aligned} \quad (5.21)$$

This optimization also has a closed form solution under these conditions, given by the product of matrices

$$\boldsymbol{\theta}_n^0(\boldsymbol{\psi}_{0:n}) = \left[\mathbf{H}^\top \mathbf{R}^{-1} \mathbf{H} + \tilde{\boldsymbol{\Sigma}}_n^{-1} \right]^{-1} \left[\mathbf{H}^\top \mathbf{R}^{-1} \mathbf{H}^* \boldsymbol{\psi}_n + \tilde{\boldsymbol{\Sigma}}_n^{-1} \mathbf{F} \mathbb{E}_{\mathbf{x}_{0:n}|\boldsymbol{\psi}_{0:n}} \{ \tilde{\boldsymbol{\theta}}_{n-1} \} \right] \quad (5.22)$$

with

$$\tilde{\boldsymbol{\Sigma}}_n = \mathbf{F}\tilde{\boldsymbol{\Sigma}}_{n-1}\mathbf{F}^\top + \mathbf{Q}_{n-1}. \quad (5.23)$$

Furthermore, it can be shown that the posterior mean and pseudotrue state are equivalent (See Appendix 5.A)

$$\mathbb{E}_{\mathbf{x}_{0:n}|\boldsymbol{\psi}_{0:n}} \{ \tilde{\boldsymbol{\theta}}_{n-1} \} = \boldsymbol{\theta}_{n-1}^0(\boldsymbol{\psi}_{0:n-1}) \quad (5.24)$$

leading to the satisfyingly recursive result that

$$\boldsymbol{\theta}_n^0(\boldsymbol{\psi}_{0:n}) = \left[\mathbf{H}^\top \mathbf{R}^{-1} \mathbf{H} + \tilde{\boldsymbol{\Sigma}}_n^{-1} \right]^{-1} \left[\mathbf{H}^\top \mathbf{R}^{-1} \mathbf{H}^* \boldsymbol{\psi}_n + \tilde{\boldsymbol{\Sigma}}_n^{-1} \mathbf{F} \boldsymbol{\theta}_{n-1}^0(\boldsymbol{\psi}_{0:n-1}) \right]. \quad (5.25)$$

5.4 Simulation-Based Experimental Validation

The proposed pseudotrue state was validated through simulation-based testing using multivariate linear models based on the forms given in (5.1) and (5.3), and for misspecifications on

all relevant parameters: the process equation $f(\boldsymbol{\theta})$, the measurement equation $h(\boldsymbol{\theta})$, as well as the covariances of the process and measurement noises, \mathbf{Q}_{n-1} and \mathbf{R}_n , respectively. The true states, $\boldsymbol{\psi}_{0:M}$, were generated for $M = 100$ seconds according to a ground-truth model based on a 1-D constant velocity tracking problem with

$$\mathbf{F}_n^* = \begin{pmatrix} 1 & T_s \\ 0 & 1 \end{pmatrix}, \mathbf{H}_n^* = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \mathbf{R}_n^* = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.5 \end{pmatrix}, \mathbf{Q}_{n-1}^* = 10^{-3}\mathbf{I} \quad (5.26)$$

The parameters of the assumed model were given based on these parameters, with the following variations based on the class of misspecification under consideration:

$$\tilde{\mathbf{F}}_n = \begin{pmatrix} 1 & 0 \\ T_s & 1 \end{pmatrix}, \tilde{\mathbf{H}}_n = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \tilde{\mathbf{R}}_n = 0.1\mathbf{R}_n^*, \tilde{\mathbf{Q}}_{n-1} = 10\mathbf{Q}_{n-1}^*. \quad (5.27)$$

The initial set of experiments tested the mismatch on only one of the parameters, with the rest of the parameters being identical to that of the true model. For example, the first test case tested a misspecification on the state transition matrix, with $\mathbf{F}_n = \tilde{\mathbf{F}}_n$ and $\mathbf{H}_n = \mathbf{H}_n^*$, $\mathbf{R}_n = \mathbf{R}_n^*$, and $\mathbf{Q}_n = \mathbf{Q}_n^*$. $N = 10000$ instances of a linear Kalman filter were created based on this assumed model and randomly initialized with $\tilde{\boldsymbol{\theta}}_{0|0} \sim \mathcal{N}(\boldsymbol{\psi}_0, \boldsymbol{\Sigma}_{0:0})$ and $\boldsymbol{\Sigma}_{0:0} = \text{diag}([20, 0.1])$. For the purposes of establishing a baseline, or control case, another N instances of a linear Kalman filter were created based on this true model. The inputs to these filters were generated according to the true model and true state, and regenerated for each of the N trials. The upper plots of Figure 5.4.1 show the average two-state Kalman filter output for both the control and test cases in comparison to the true states $\boldsymbol{\psi}_{1:M}$ and the associated pseudotrue states $\boldsymbol{\theta}_{1:M}^0$. As expected, the pseudotrue state is closely followed by the estimates produced by the Kalman filters that were based on the assumed model. This result validates this interpretation of the pseudotrue state for linear Kalman filtering. This process was then repeated for individual model misspecifications on \mathbf{H}_n , \mathbf{R}_n and \mathbf{Q}_n in Figures 5.4.1, 5.4.1 and 5.4.1 respectively with similar results. For completeness, two more experiments were performed based on a ‘‘Total Mismatch’’ scenario, each with the same procedure as with the previous single-parameter mismatch cases. The first of these experiments was a total mismatch scenario with $\mathbf{F}_n = \tilde{\mathbf{F}}_n$ and $\mathbf{H}_n = \tilde{\mathbf{H}}_n$, $\mathbf{R}_n = \tilde{\mathbf{R}}_n$, and $\mathbf{Q}_n = \tilde{\mathbf{Q}}_n$, with similar results shown in Figure 5.4.5. Finally, the experiment was repeated with a total mismatch scenario where the assumed model parameters were all randomly generated, with with all elements of \mathbf{F}_n and \mathbf{H}_n generated according to $F_{i,j} \sim \mathcal{N}(0, 1)$ and $H_{i,j} \sim \mathcal{N}(0, 1)$. \mathbf{R}_n and \mathbf{Q}_{n-1} were similarly generated and then squared to guarantee positive semidefiniteness of the noise covariance matrices. This approach adds

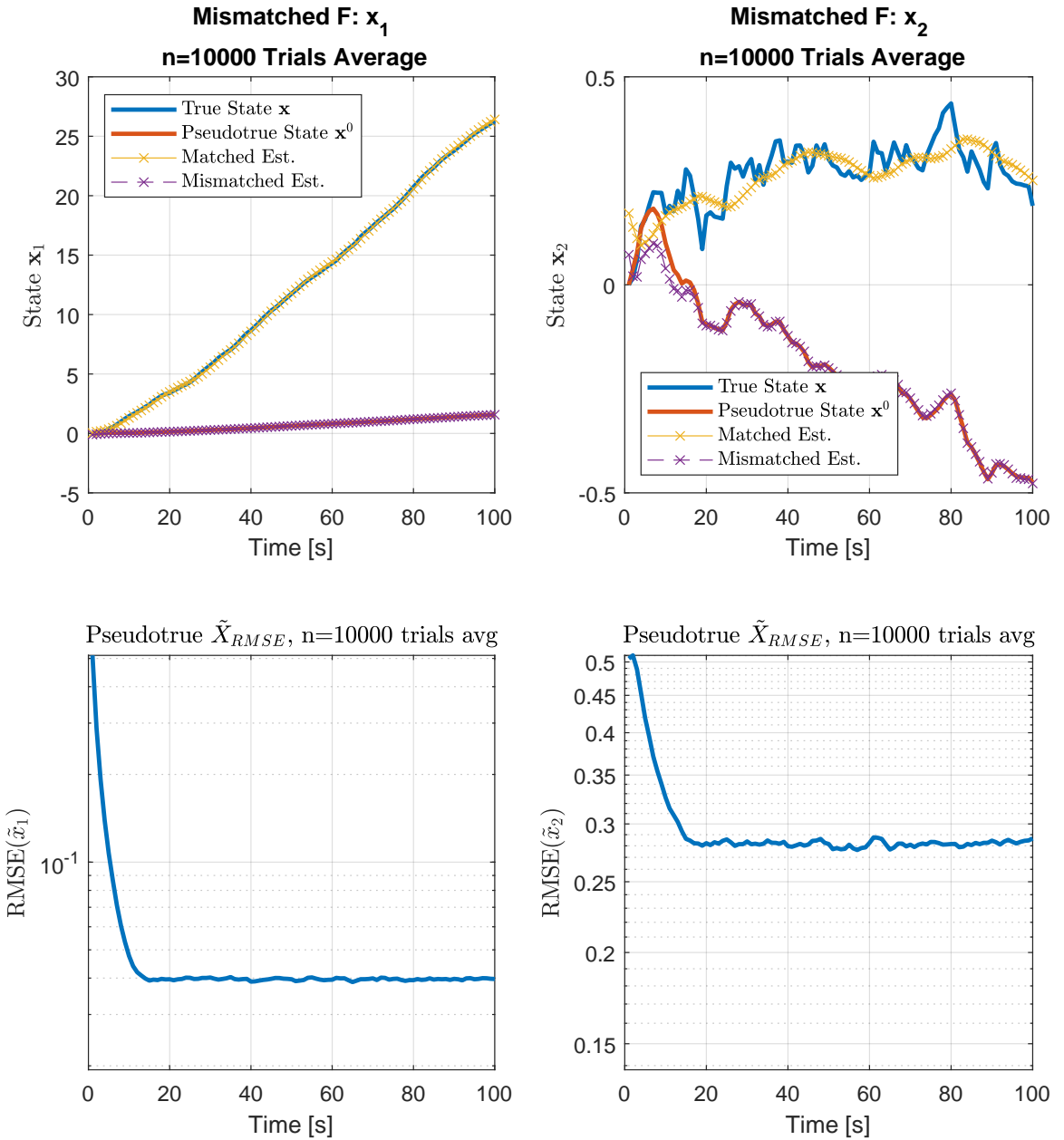


Figure 5.4.1: Comparison of pseudotrue state to average linear KF mismatched estimate, true state to average KF matched estimate. Misspecified state transition matrix.

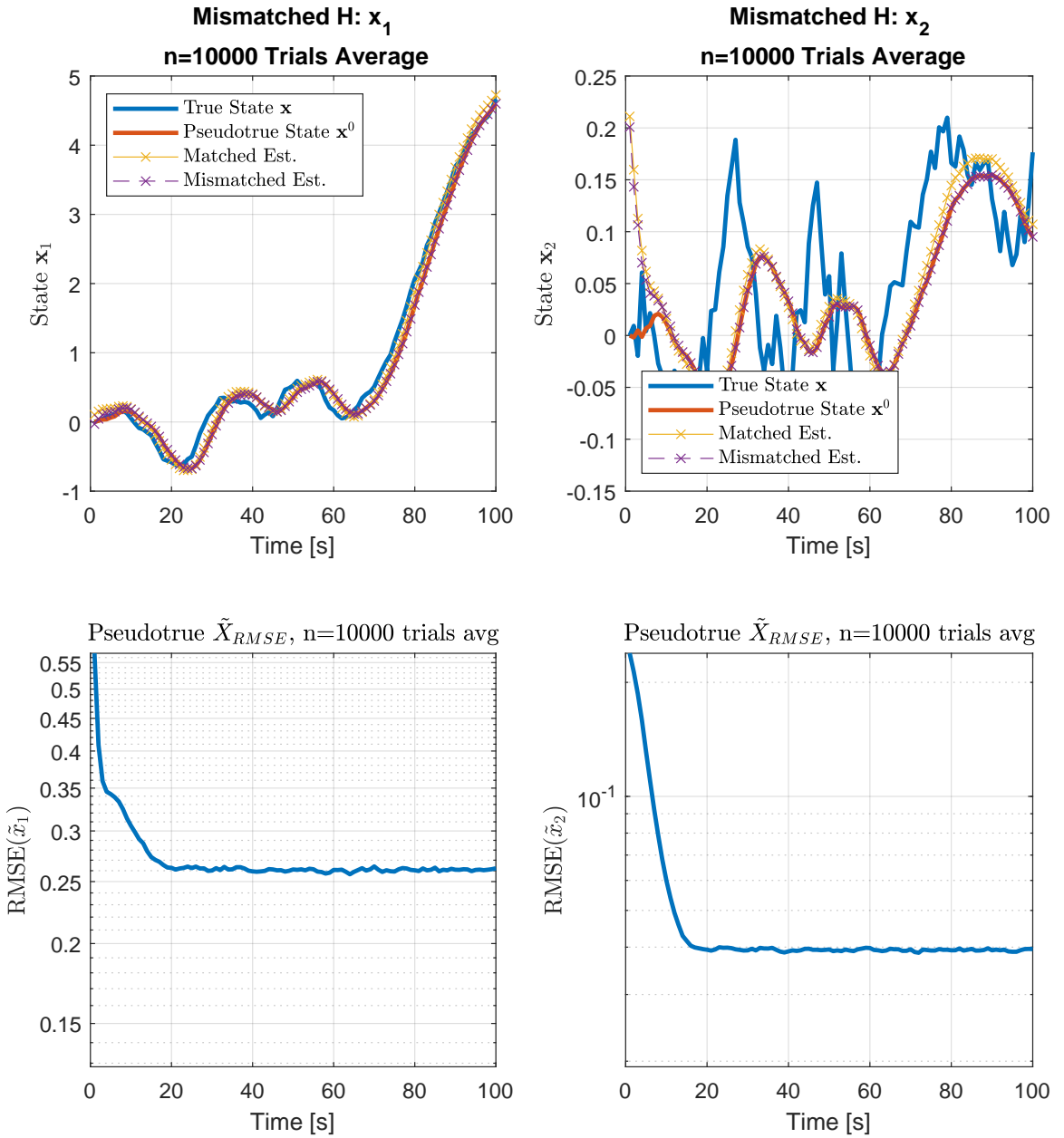


Figure 5.4.2: Comparison of pseudotrue state to average linear KF mismatched estimate, true state to average KF matched estimate. Misspecified measurement matrix.

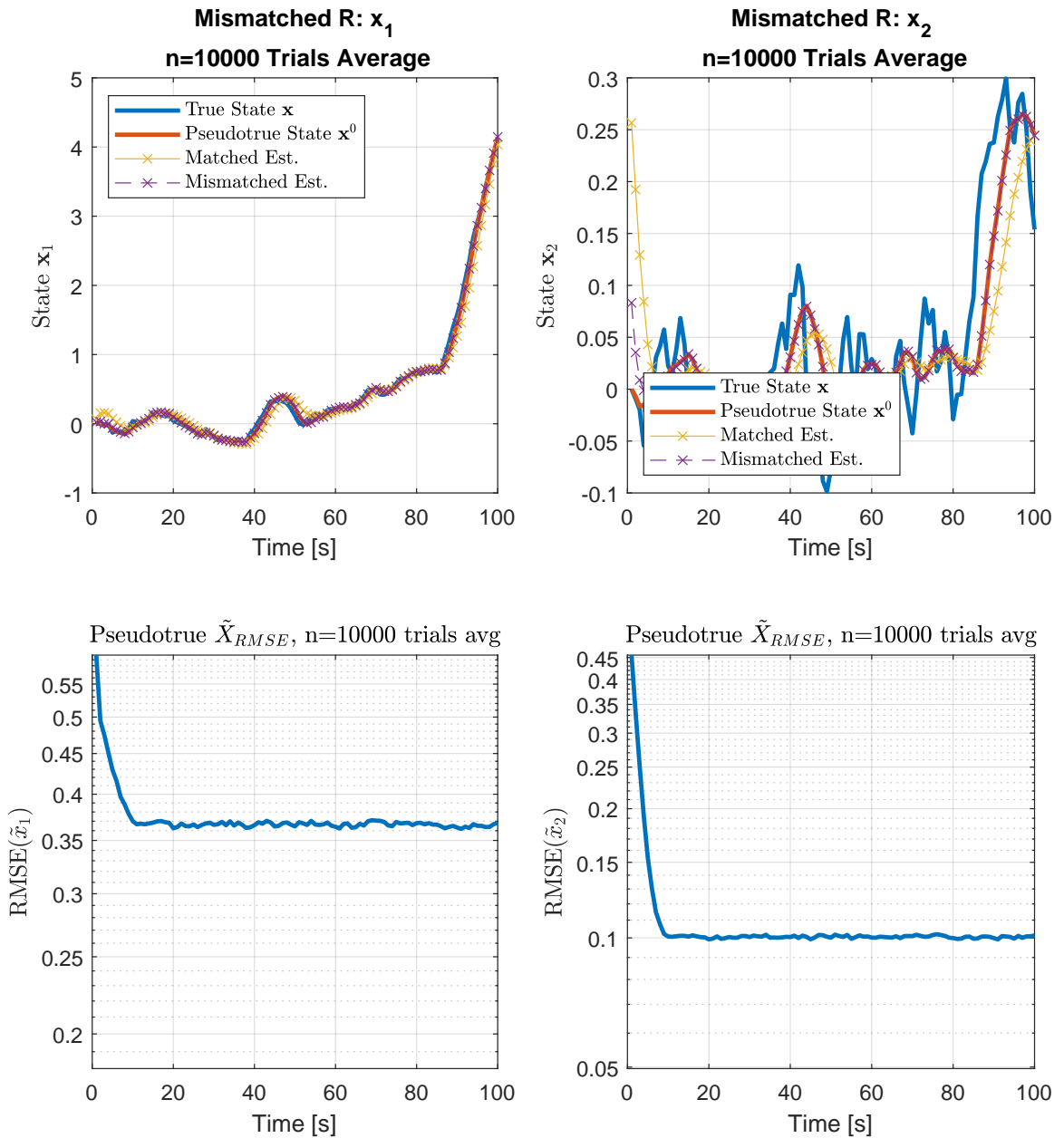


Figure 5.4.3: Comparison of pseudotrue state to average linear KF mismatched estimate, true state to average KF matched estimate. Misspecified measurement covariance.

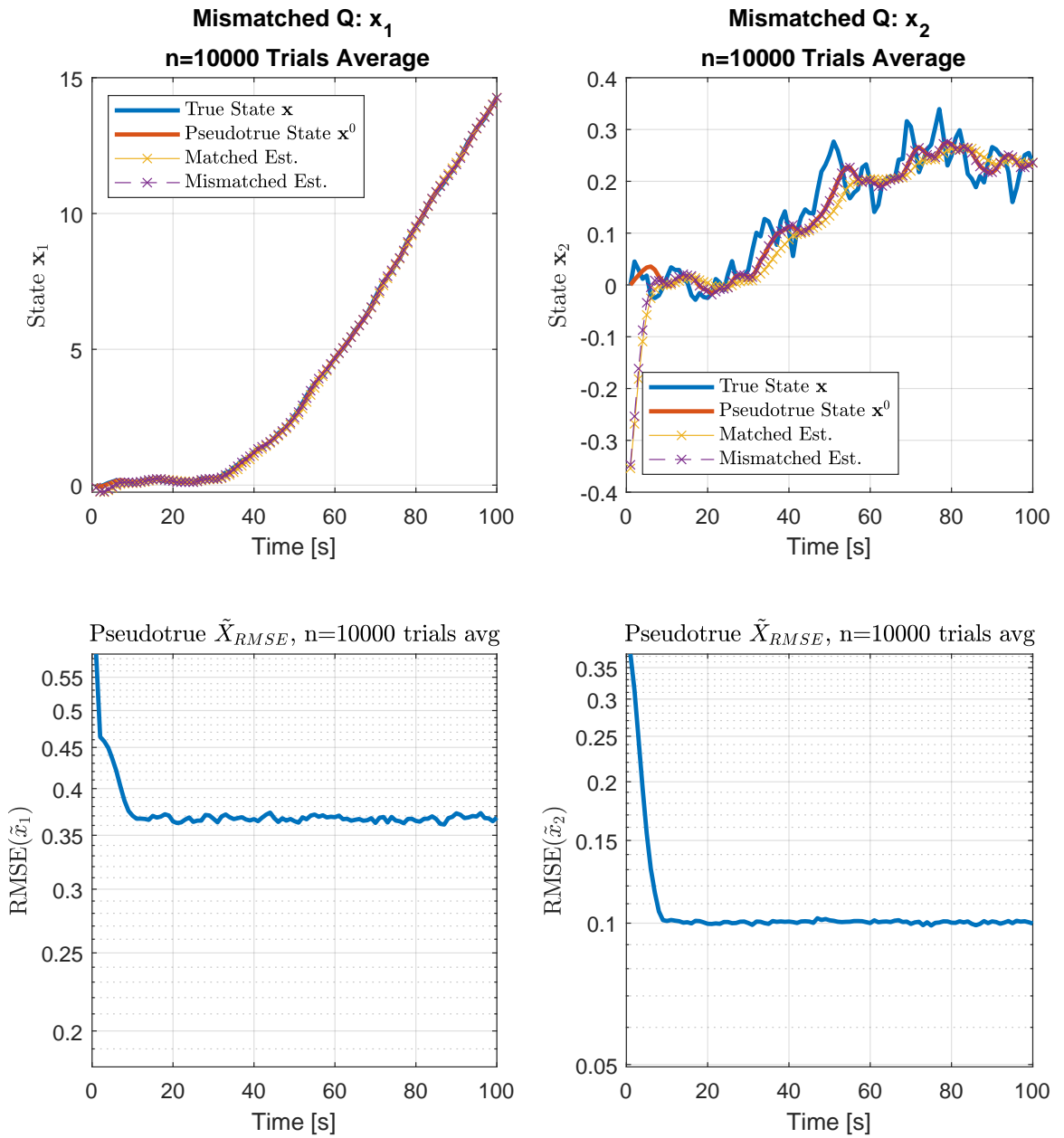


Figure 5.4.4: Comparison of pseudotrue state to average linear KF mismatched estimate, true state to average KF matched estimate. Misspecified state transition covariance.

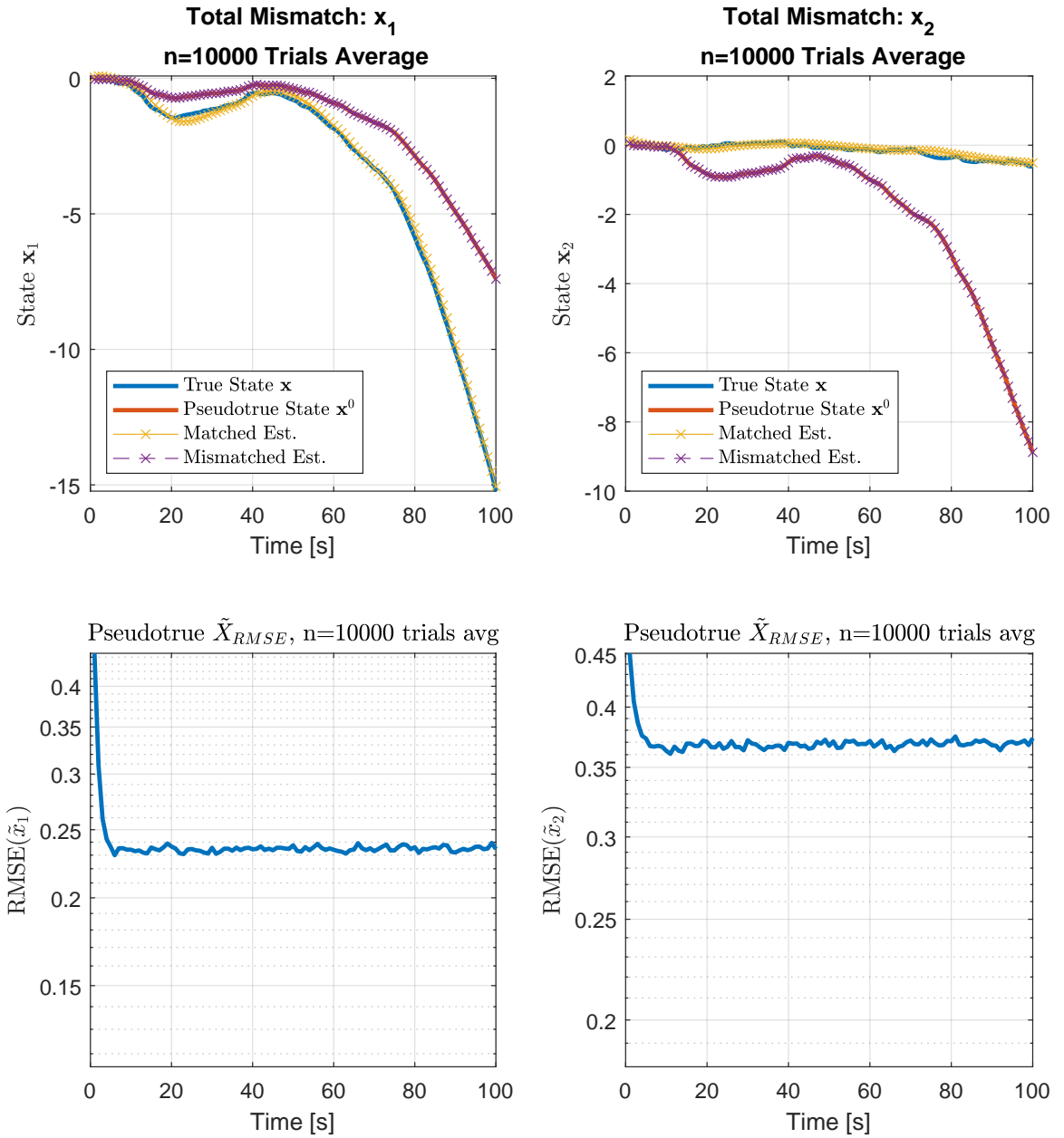


Figure 5.4.5: Comparison of pseudotrue state to average linear KF mismatched estimate, true state to average KF matched estimate. Total model misspecification.

some complexity in comparison to the assumed system dynamics, including assumed measurement and process noises which are not independent for each observation and state respectively. The results shown in Figure 5.4.6 tell a similar story as the other experiments, validating our pseudotrue parameter for an arbitrary set of assumed system parameters.

5.5 Conclusions

Our experiments with the proposed recursion demonstrate the ability for our definition of the pseudotrue state to accurately predict the average response of a misspecified linear Kalman filter to data generated for a given set of hidden parameters ψ . As with other applications of the pseudotrue parameter in parameter estimation problems, the pseudotrue state provides a powerful tool for understanding the behavior of a designed estimator in various conditions of model misspecification. For example, based on the scenarios that we tested in Section 5.4, it can be said that misspecification of the process and measurement noise covariances on the order of magnitude that we tested have relatively little effect on the output performance of the estimators in comparison to the true parameter, in comparison to misspecification on the state transition and measurement matrices. Such experiments could be repeated for different proposed true and assumed model compositions to construct models which are resilient to the uncertainties inherent in constructing estimators.

Of note in each of the experiments we performed was that the root mean-squared error of the misspecified estimators with respect to the pseudotrue state, shown in the bottom half of each of the figures in this chapter, demonstrates a clear asymptotic behavior towards a non-zero offset from the pseudotrue state. Future developments based on this work will focus on characterizing this behavior using the MPCRB, which will further assist in the kind of model misspecification analyses discussed here.

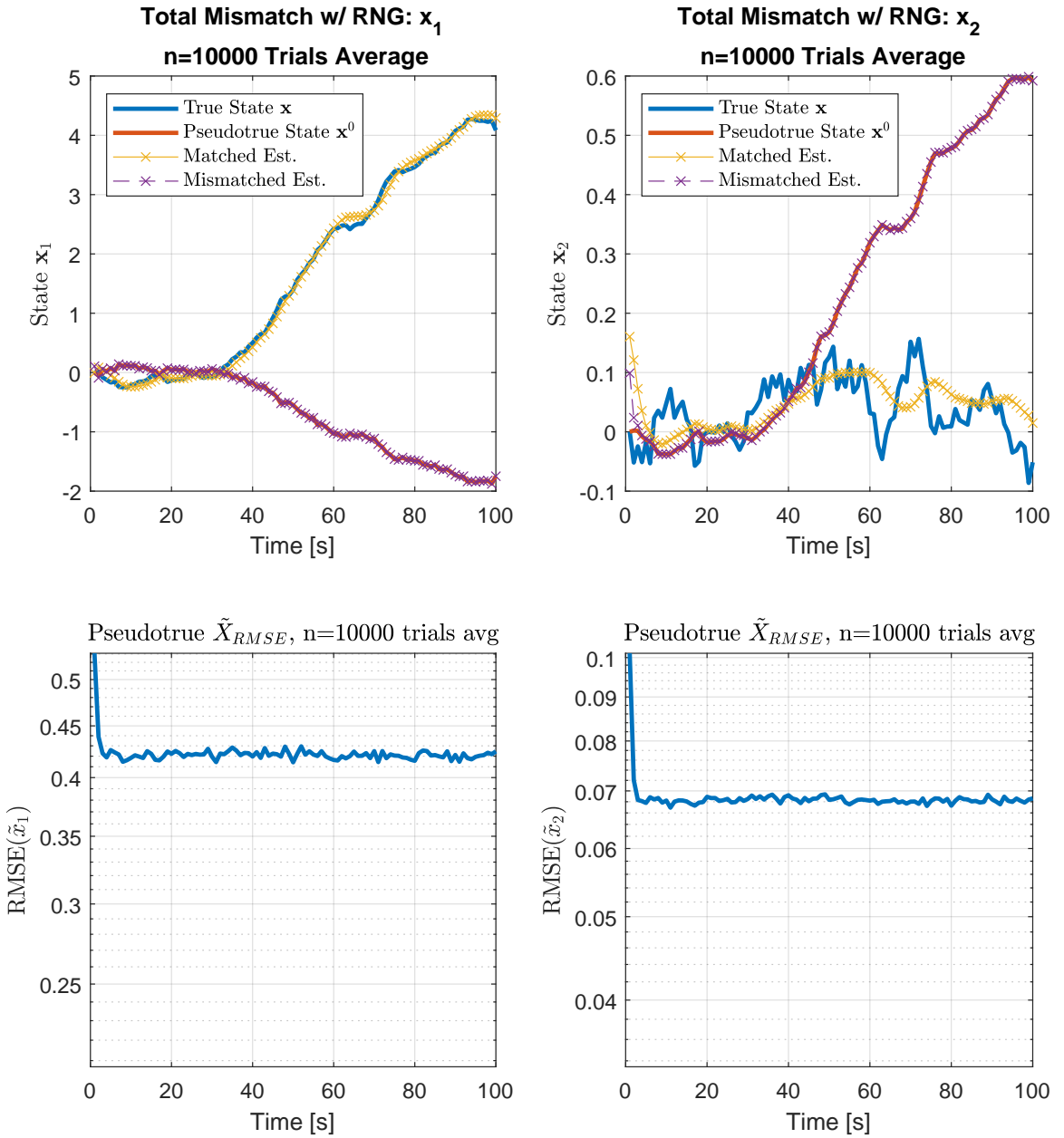


Figure 5.4.6: Comparison of pseudotrue state to average linear KF mismatched estimate, true state to average KF matched estimate. Randomized assumed model.

Appendix

5.A Proof of Equality (5.24)

Let $\tilde{\boldsymbol{\theta}}_{n|n}$ be the mean of the posterior density $p(\boldsymbol{\theta}_n|\mathbf{x}_{0:n})$ as in the linear Kalman filter,

$$\tilde{\boldsymbol{\theta}}_{n|n} = \mathbf{F}\tilde{\boldsymbol{\theta}}_{n-1|n-1} + \mathbf{K}_n(\mathbf{x}_n - \mathbf{H}\mathbf{F}\tilde{\boldsymbol{\theta}}_{n-1|n-1}) \quad (5.28)$$

$$= \mathbf{K}_n\mathbf{x}_n + (\mathbf{F} + \mathbf{K}_n\mathbf{H}\mathbf{F})\tilde{\boldsymbol{\theta}}_{n-1|n-1} \quad (5.29)$$

where \mathbf{K}_n is the Kalman gain $\mathbf{K}_n = \tilde{\boldsymbol{\Sigma}}_n^{-1}\mathbf{H}\mathbf{S}_n^{-1}$ defined in terms of predictive and innovations covariances given by

$$\tilde{\boldsymbol{\Sigma}}_n^{-1} = \mathbf{F}\boldsymbol{\Sigma}_{n-1|n-1}\mathbf{F}^\top + \mathbf{Q}_{n-1}, \quad (5.30)$$

$$\mathbf{S}_n = \mathbf{H}\tilde{\boldsymbol{\Sigma}}_n^{-1}\mathbf{H}^\top + \mathbf{R}_k. \quad (5.31)$$

Taking an expectation over the ground truth likelihood yields the expression

$$\mathbb{E}_{\mathbf{x}_{0:n}|\boldsymbol{\psi}_{0:n}}\{\tilde{\boldsymbol{\theta}}_{n|n}\} = \mathbf{K}_n\mathbf{H}_n^*\boldsymbol{\psi}_n + (\mathbf{F} + \mathbf{K}_n\mathbf{H}\mathbf{F})\mathbb{E}_{\mathbf{x}_{0:n}|\boldsymbol{\psi}_{0:n}}\{\tilde{\boldsymbol{\theta}}_{n-1|n-1}\}. \quad (5.32)$$

Expanding this expression using the Kalman gain gives us

$$\begin{aligned} \mathbb{E}_{\mathbf{x}_{0:n}|\boldsymbol{\psi}_{0:n}}\{\tilde{\boldsymbol{\theta}}_{n|n}\} &= \tilde{\boldsymbol{\Sigma}}_n^{-1}\mathbf{H}^\top(\mathbf{H}\tilde{\boldsymbol{\Sigma}}_n^{-1}\mathbf{H}^\top + \mathbf{R}_n)^{-1}\mathbf{H}^*\boldsymbol{\psi}_n \\ &+ [\mathbf{F} - \tilde{\boldsymbol{\Sigma}}_n^{-1}\mathbf{H}^\top(\mathbf{H}\tilde{\boldsymbol{\Sigma}}_n^{-1}\mathbf{H}^\top + \mathbf{R}_n)^{-1}\mathbf{H}\mathbf{F}]\mathbb{E}_{\mathbf{x}_{0:n}|\boldsymbol{\psi}_{0:n}}\{\tilde{\boldsymbol{\theta}}_{n-1|n-1}\}. \end{aligned} \quad (5.33)$$

Recalling the expression for the pseudotrue state given in (5.22), and applying the matrix inversion lemma [92]

$$\left[\mathbf{H}^\top\mathbf{R}^{-1}\mathbf{H} + \tilde{\boldsymbol{\Sigma}}_n^{-1}\right]^{-1} = \tilde{\boldsymbol{\Sigma}}_n - \tilde{\boldsymbol{\Sigma}}_n\mathbf{H}^\top(\mathbf{R}_n + \mathbf{H}\tilde{\boldsymbol{\Sigma}}_n\mathbf{H}^\top)^{-1}\mathbf{H}\tilde{\boldsymbol{\Sigma}}_n, \quad (5.34)$$

we have

$$\begin{aligned} \boldsymbol{\theta}_n^0(\boldsymbol{\psi}_{0:n}) &= \left[\tilde{\boldsymbol{\Sigma}}_n - \tilde{\boldsymbol{\Sigma}}_n\mathbf{H}(\mathbf{H}\tilde{\boldsymbol{\Sigma}}_n^{-1}\mathbf{H}^\top + \mathbf{R}_n)^{-1}\mathbf{H}\tilde{\boldsymbol{\Sigma}}_n\right]\mathbf{H}^\top\mathbf{R}_n^{-1}\mathbf{H}^*\boldsymbol{\psi}_n \\ &+ \left[\tilde{\boldsymbol{\Sigma}}_n - \tilde{\boldsymbol{\Sigma}}_n\mathbf{H}^\top(\mathbf{H}\tilde{\boldsymbol{\Sigma}}_n^{-1}\mathbf{H}^\top + \mathbf{R}_n)^{-1}\mathbf{H}\tilde{\boldsymbol{\Sigma}}_n\right]\tilde{\boldsymbol{\Sigma}}_n^{-1}\mathbf{F}\mathbb{E}_{\mathbf{x}_{0:n}|\boldsymbol{\psi}_{0:n}}\{\tilde{\boldsymbol{\theta}}_{n-1|n-1}\} \end{aligned} \quad (5.35)$$

Applying (5.34) to the second term of (5.33) and (5.35) and simplifying reveals that

$$\mathbf{F} - \tilde{\Sigma}_n^{-1} \mathbf{H}^\top (\mathbf{H} \tilde{\Sigma}_n^{-1} \mathbf{H}^\top + \mathbf{R}_n)^{-1} \mathbf{H} \mathbf{F} = [\mathbf{I} - \tilde{\Sigma}_n \mathbf{H} (\mathbf{H} \tilde{\Sigma}_n^{-1} \mathbf{H}^\top + \mathbf{R}_n)^{-1} \mathbf{H}] \mathbf{F} \quad (5.36)$$

$$= (\tilde{\Sigma}_n - \tilde{\Sigma}_n \mathbf{H}^\top (\mathbf{H} \tilde{\Sigma}_n^{-1} \mathbf{H}^\top + \mathbf{R}_n)^{-1} \mathbf{H} \tilde{\Sigma}_n) \tilde{\Sigma}_n^{-1} \mathbf{F}. \quad (5.37)$$

Furthermore, using another important result of the matrix inversion lemma, namely that

$$(\mathbf{H} \tilde{\Sigma}_n^{-1} \mathbf{H}^\top + \mathbf{R}_n)^{-1} \mathbf{H} \tilde{\Sigma}_n = \mathbf{R}_n^{-1} \mathbf{H} (\mathbf{H}^\top \mathbf{R}_n^{-1} \mathbf{H} + \tilde{\Sigma}_n^{-1})^{-1}, \quad (5.38)$$

it is shown that

$$\tilde{\Sigma}_n^{-1} \mathbf{H}^\top (\mathbf{H} \tilde{\Sigma}_n^{-1} \mathbf{H}^\top + \mathbf{R}_n)^{-1} = (\tilde{\Sigma}_n - \tilde{\Sigma}_n \mathbf{H} (\mathbf{H} \tilde{\Sigma}_n^{-1} \mathbf{H}^\top + \mathbf{R}_n)^{-1} \mathbf{H} \tilde{\Sigma}_n) \mathbf{H}^\top \mathbf{R}_n^{-1} \quad (5.39)$$

$$= \tilde{\Sigma}_n^{-1} \mathbf{H}^\top (\mathbf{R}_n^{-1} - \mathbf{R}_n^{-1} \mathbf{H} (\mathbf{H}^\top \mathbf{R}_n^{-1} \mathbf{H} + \tilde{\Sigma}_n^{-1})^{-1} \mathbf{H}^\top \mathbf{R}_n^{-1}). \quad (5.40)$$

Applying the equalities shown in (5.36-5.37) and (5.39-5.40) to the expressions for the posterior mean given in (5.33) and the pseudotrue state given in (5.35) it is shown that

$$\mathbb{E}_{\mathbf{x}_{0:n} | \psi_{0:n}} \{\tilde{\boldsymbol{\theta}}_{n|n}\} = \boldsymbol{\theta}_{n-1}^0(\psi_{0:n-1}). \quad (5.41)$$

and furthermore that, as the recursion is shown to be identical between the expectation recursion in (5.33) and the pseudotrue state recursion, that

$$\mathbb{E}_{\mathbf{x}_{0:n} | \psi_{0:n}} \{\tilde{\boldsymbol{\theta}}_{n-1}\} = \boldsymbol{\theta}_{n-1}^0(\psi_{0:n-1}). \quad (5.42)$$

Chapter 6

Summary of Key Results and Conclusions

This dissertation explored new methods to improve estimator performance through bounds analysis and misspecified model analysis. The key objectives, as identified in **Chapter 1**, included addressing model uncertainty, applying statistical bounds, enhancing estimator performance through these bounds, and extending misspecification analysis tools to dynamic state estimation problems. Each chapter of this dissertation has served to present contributions that we have made to furthering knowledge with respect to each of these components.

6.1 Dissertation Contributions

In **Chapter 2** we sought to address the objective of enhancing estimator performance through these bounds by demonstrating the utility of employing the CRB as an objective function against which estimator design parameters, in this case the parameters of a UAV flight pattern, could be optimized. Although this section focused on the utility with respect to the *matched case* where no model misspecification has occurred, similar approaches may be applied to misspecified bounds and pseudotrue parameters to achieve optimal results with respect to other metrics.

The contributions described in **Chapter 3** provide powerful results in regard to applying statistical bounds. In it we investigated the applicability of the MCRB to eigenspace methods for DOA estimation by establishing conditions under which the MML estimate, shown in the literature to converge to the MCRB, and the MMUSIC estimate converge. We established a set of conditions on an additive error component of the basis function matrix, that are applicable to either single-signal

CHAPTER 6. SUMMARY OF KEY RESULTS AND CONCLUSIONS

or multi-signal conditions, and tested these conditions using simulations of uniform circular and uniform linear arrays. Our results in this show that under a particular framework the MCRB and MMLE represent useful tools for analyzing the performance of algorithms like MUSIC under model misspecification.

In **Chapter 4**, we addressed an example of model uncertainty by developing a Bayesian framework for rectifying uncertain measurement model information in recursive Gaussian estimators. Our proposed method leverages the conjugate prior distribution to the Gaussian distribution to produce a recursive estimator for the innovations covariance matrix. Included in this framework is a method of using consistency checks on the Gaussian filter to identify changes in the statistics of the innovations, and to reset the parameters of the conjugate prior accordingly. This approach was validated through simulation and yielded promising results.

Finally, in **Chapter 5** we extended existing definitions of the “pseudotrue parameter” associated with the MCRB and the “Bayesian pseudotrue parameter” associated with the MBCRB to the class of dynamic state space models. Our definition of the “pseudotrue state” incorporates all measurements and all states up to the current time, rather than just a single one, and results in a recursive case when applied to linear Gaussian state space models. This result for linear Gaussian models was tested for various cases of dynamic model misspecification, and was shown to be consistent with the output of the misspecified KF averaged over a large number of Monte Carlo trials. This validates this definition of the pseudotrue state for particular realizations of the true underlying parameter.

6.2 Directions for Future Research

The findings and contributions discussed above naturally lead to several directions for future research, which aim to extend and refine the methodologies introduced in this thesis. The most likely next steps from this work is:

- Exploring the utility of applying parameter optimization to the pseudotrue parameter and the MCRB, and identifying applications where this sort of optimization could prove useful.
- Extending the theorems regarding MMUSIC and MMLE performance in the small error and stochastic error domains to broader classes of misspecification.

- Establishing convergence criterion and a rigorous analytic evaluation of the performance of the BSE methodology within KF architectures based on misspecified model analysis.
- Development of a tractable, recursive form of the misspecified posterior Cramér-Rao lower bound (MPCRB) constraining the performance of misspecified dynamic estimators around the pseudotrue state.

6.3 Final Remarks

Our efforts in this thesis have focused on different aspects of evaluating and improving statistical estimation methods, both for static and dynamic models, with the thrust of our effort being with regard to model misspecification. Model misspecification is an often overlooked condition that appears in all forms of estimation. The literature contains a number of tools that may be applied to evaluating and bounding the performance of models under matched conditions and conditions of model misspecification, and while these tools are powerful in the contexts that they have been applied thus far, the challenge of extending these tools to broader classes of estimator remains an open one. The contributions of this thesis demonstrate both the potential of bounds analysis, and the potential of extending these analysis tools to dynamic state estimators.

The methodologies developed here have significant implications for both academic research and practical applications. Misspecification comes in all forms, and it is critical to understand the failure conditions of a given system. Design considerations should be made to minimize the effect of these failure conditions. This work lays the ground work for the further development of analysis tools, and for designing robust estimators capable of continued performance under challenging conditions. This thesis concludes that misspecification analysis are not just useful as theoretical concepts, but represent an opportunity to develop design processes and estimators that meet the challenges of the complicated world in which we live.

Bibliography

- [1] Q. H. Vuong, “Cramér-Rao bounds for misspecified models,” 1986.
- [2] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation: theory algorithms and software*.
- [3] G. LaMountain and P. Closas, “Maneuver optimization for synthetic aperture based doa estimation of GNSS jammers,” in *2020 IEEE/ION Position, Location and Navigation Symposium (PLANS)*. IEEE, 2020, pp. 44–49.
- [4] G. LaMountain, J. Vilà-Valls, and P. Closas, “Bayesian covariance estimation for Kalman filter based digital carrier synchronization,” in *Proceedings of the 31st International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2018)*, 2018, pp. 3575–3586.
- [5] G. LaMountain and P. Closas, “A mismatched bound for stochastic DOA estimation,” in *2020 54th Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2020, pp. 778–783.
- [6] C. Fernández-Prades, J. Arribas, M. Majoral, G. Araujo, A. Arnold, C. Avilés, M. Branzanti, A. Cebrián-Juan, A. Cecilia-Luque, L. Esteve, F. Fabra, D. Fehr, P. Gupta, G. LaMountain, M. Lenhart, J. Melton, D. Miralles, M. Molina, R. Muñoz, C. O’Driscoll, I. Pääkkönen, I. Pérez Riega, D. Pubill, A. Ramos, E. Shin, J. Schindehette, W. Silberman, L. Tonetto, and S. van der Linden, “GNSS-SDR,” Jan. 2024. [Online]. Available: <https://github.com/gnss-sdr/gnss-sdr>
- [7] T. Pany, D. Akos, J. Arribas, M. Z. H. Bhuiyan, P. Closas, F. Dovis, I. Fernandez-Hernandez, C. Fernández-Prades, S. Gunawardena, T. Humphreys *et al.*, “GNSS Software-Defined Radio: History, Current Developments, and Standardization Efforts,” *NAVIGATION: Journal of the Institute of Navigation*, vol. 71, no. 1, 2024.

BIBLIOGRAPHY

- [8] C. Fernández–Prades, J. Arribas, P. Closas, C. Avilés, and L. Esteve, “GNSS-SDR: An open source tool for researchers and developers,” in *Proc. of ION GNSS 2011 Conference*, Portland, Oregon, Sept. 2011.
- [9] T. Imbiriba, P. Wu, G. LaMountain, D. Erdoğmuş, and P. Closas, “Recursive Gaussian processes and fingerprinting for indoor navigation,” in *2020 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, 2020, pp. 933–940.
- [10] T. Imbiriba, G. LaMountain, P. Wu, D. Erdoğmuş, and P. Closas, “Change detection and Gaussian process inference in piecewise stationary environments under noisy inputs,” in *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. IEEE, 2019, pp. 530–534.
- [11] P. Wu, T. Imbiriba, G. LaMountain, J. Vilà-Valls, and P. Closas, “Wifi fingerprinting and tracking using neural networks,” in *Proceedings of the 32nd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2019)*, 2019, pp. 2314–2324.
- [12] G. Hernandez, G. LaMountain, and P. Closas, “Privacy-preserving cooperative positioning,” in *Proceedings of the 33rd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2020)*, 2020, pp. 2667–2675.
- [13] —, “Proximity-based positioning scheme with multi-layer privacy,” in *2023 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, 2023, pp. 235–242.
- [14] —, “Privacy-preserving cooperative GNSS positioning,” *NAVIGATION: Journal of the Institute of Navigation*, vol. 70, no. 4, 2023. [Online]. Available: <https://navi.ion.org/content/70/4/navi.625>
- [15] H. Sathaye, G. LaMountain, P. Closas, and A. Ranganathan, “Semperfi: A spoofer eliminating GPS receiver for UAVs,” in *Proceedings of Network and Distributed System Security Symposium (NDSS 2022)*, 2022.
- [16] M. T. Kling, D. Lau, K. L. Witham, and G. M. LaMountain, “System for closed-loop GNSS simulation,” U.S. Patent Application 0 365 223, Dec. 09, 2024.
- [17] P. Closas, G. Hernandez, and G. LaMountain, “Method and apparatus for determining a geospatial location of an unlocated device,” U.S. Patent Application 0 276 226, Dec. 09, 2024.

BIBLIOGRAPHY

- [18] S. Tang, G. LaMountain, T. Imbiriba, and P. Closas, "On parametric misspecified Bayesian cramer-rao bound: An application to linear/gaussian systems," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.
- [19] D. Borio, C. O'Driscoll, and J. Fortuny, "GNSS jammers: Effects and countermeasures," in *2012 6th ESA Workshop on Satellite Navigation Technologies (Navitec 2012) & European Workshop on GNSS Signals and Signal Processing*. IEEE, 2012, pp. 1–7.
- [20] D. Borio, F. Dovis, H. Kuusniemi, and L. L. Presti, "Impact and detection of GNSS jammers on consumer grade satellite navigation receivers," *Proceedings of the IEEE*, vol. 104, no. 6, pp. 1233–1245, 2016.
- [21] M. G. Amin, P. Closas, A. Broumandan, and J. L. Volakis, "Vulnerabilities, threats, and authentication in satellite-based navigation systems [scanning the issue]," *Proceedings of the IEEE*, vol. 104, no. 6, pp. 1169–1173, 2016.
- [22] D. Medina, C. Lass, E. Pérez-Marcos, R. Ziebold, P. Closas, and J. García, "On GNSS Jamming Threat from the Maritime Navigation Perspective," in *Proceedings of the 22st International Conference on Information Fusion (FUSION), Ottawa, ON, Canada*, 2019, pp. 2–5.
- [23] R. Morales-Ferre, P. Richter, E. Falletti, A. de la Fuente, and E. S. Lohan, "A survey on coping with intentional interference in satellite navigation for manned and unmanned aircraft," *IEEE Communications Surveys & Tutorials*, 2019.
- [24] C. Fernández-Prades, J. Arribas, and P. Closas, "Robust GNSS receivers by array signal processing: Theory and implementation," *Proceedings of the IEEE*, vol. 104, no. 6, pp. 1207–1220, June 2016.
- [25] J. L. Volakis, A. J. O'Brien, and C.-C. Chen, "Small and adaptive antennas and arrays for GNSS applications," *Proceedings of the IEEE*, vol. 104, no. 6, pp. 1221–1232, 2016.
- [26] M. A. Ribot Sanfelix, "Parameter Estimation with GNSS-Reflectometry and GNSS Synthetic Aperture Techniques," Ph.D. dissertation, EPFL, EPFL, Lausanne, 2018. [Online]. Available: <http://infoscience.epfl.ch/record/253111>
- [27] M. A. Ribot, J. Cabeza, P. Closas, C. Botteron, and P.-A. Farine, "Estimation bounds for GNSS synthetic aperture techniques," in *2017 IEEE 7th International Workshop on Computational*

BIBLIOGRAPHY

- Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. IEEE, Dec. 2017. [Online]. Available: <https://doi.org/10.1109/camsap.2017.8313168>
- [28] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The computer journal*, vol. 7, no. 4, pp. 308–313, 1965.
- [29] C. Bogani, M. Gasparo, and A. Papini, "Generalized pattern search methods for a class of nonsmooth optimization problems with structure," *Journal of Computational and Applied Mathematics*, vol. 229, no. 1, pp. 283 – 293, 2009. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0377042708005773>
- [30] B. Shahriari, K. Swersky, Z. Wang, R. Adams, and N. D. Freitas, "Taking the human out of the loop: A review of Bayesian optimization," *Proceedings of the IEEE*, vol. 104, no. 1, pp. 148–175, 2015.
- [31] C. Audet and J. E. Dennis, "Analysis of generalized pattern searches," *SIAM Journal on Optimization*, vol. 13, no. 3, pp. 889–903, Jan. 2002. [Online]. Available: <https://doi.org/10.1137/s1052623400378742>
- [32] Y. H. Chen, F. Rothmaier, D. Akos, S. Lo, and P. Enge, "Towards a practical single element null steering antenna." Monterey, CA: Proc. of the Institute of Navigation International Technical Meeting, 2017.
- [33] Y.-H. Chen, S. Lo, A. Perkins, F. Rothmaier, D. Akos, and P. Enge, "Demonstrating single element null steering antenna direction finding for interference detection," in *Proceedings of the 2018 International Technical Meeting of The Institute of Navigation*, 2018, pp. 240–259.
- [34] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE signal processing magazine*, vol. 13, no. 4, pp. 67–94, 1996.
- [35] R. A. Monzingo and T. W. Miller, *Introduction to adaptive arrays*. Scitech publishing, 2004.
- [36] H. L. Van Trees, *Optimum array processing: Part IV of detection, estimation, and modulation theory*. John Wiley & Sons, 2002.
- [37] M. Greco, S. Fortunati, and F. Gini, "Maximum likelihood covariance matrix estimation for complex elliptically symmetric distributions under mismatched conditions," *Signal Processing*, vol. 104, pp. 381–386, 2014. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0165168414001583>

BIBLIOGRAPHY

- [38] P. J. Huber, "The behavior of maximum likelihood estimates under nonstandard conditions," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 1, 1967, pp. 221–233.
- [39] C. Ren, M. N. El Korso, J. Galy, E. Chaumette, P. Larzabal, and A. Renaux, "Performance bounds under misspecification model for mimo radar application," in *2015 23rd European Signal Processing Conference (EUSIPCO)*. IEEE, 2015, pp. 514–518.
- [40] G. LaMountain and P. Closas, "A mismatched bound for stochastic DOA estimation," in *2020 54th Asilomar Conference on Signals, Systems, and Computers*, 2020, pp. 778–783.
- [41] S. Fortunati, F. Gini, and M. S. Greco, "The misspecified Cramér-Rao bound and its application to scatter matrix estimation in complex elliptically symmetric distributions," *IEEE Transactions on Signal Processing*, vol. 64, no. 9, pp. 2387–2399, 2016.
- [42] A. Jaffer, "Maximum likelihood direction finding of stochastic sources: a separable solution," in *ICASSP-88., International Conference on Acoustics, Speech, and Signal Processing*. IEEE, 1988, pp. 2893–2896 vol.5.
- [43] P. Stoica and K. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 7, pp. 1132–1143, 1990.
- [44] S. Fortunati, F. Gini, M. S. Greco, and C. D. Richmond, "Performance bounds for parameter estimation under misspecified models: Fundamental findings and applications," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 142–157, 2017.
- [45] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, New Jersey, USA: Prentice-Hall, 1993.
- [46] H. White, "Maximum likelihood estimation of misspecified models," *Econometrica*, vol. 50, no. 1, pp. 1–25, 1982.
- [47] E. B. Daher, "Analysis and design of nonuniform arrays for direction finding," 2018.
- [48] C. D. Richmond and L. L. Horowitz, "Parameter bounds on estimation accuracy under model misspecification," *IEEE transactions on signal processing*, vol. 63, no. 9, pp. 2263–2278, 2015.

BIBLIOGRAPHY

- [49] P. Stoica and A. Nehorai, “MUSIC, maximum likelihood, and cramer-rao bound,” *IEEE Transactions on Acoustics, speech, and signal processing*, vol. 37, no. 5, pp. 720–741, 1989.
- [50] B. Friedlander, “On the mutual coupling matrix in array signal processing,” in *2020 54th Asilomar Conference on Signals, Systems, and Computers*, 2020, pp. 1245–1249.
- [51] S. Särkkä, *Bayesian Filtering and Smoothing*. Cambridge University Press, 2013.
- [52] J. Duník, S. K. Biswas, A. G. Dempster, T. Pany, and P. Closas, “State estimation methods in navigation: Overview and application,” *IEEE Aerospace and Electronic Systems Magazine*, vol. 35, no. 12, pp. 16–31, 2020.
- [53] P. M. Djurić, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Míguez, “Particle filtering,” *IEEE Signal Processing Magazine*, vol. 20, no. 5, pp. 19–38, September 2003.
- [54] K. Ito and K. Xiong, “Gaussian filters for nonlinear filtering problems,” *IEEE Trans. on Automatic Control*, vol. 45, no. 5, pp. 910–927, May 2000.
- [55] S. J. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte, “A new method for nonlinear transformation of means and covariances in filters and estimators,” *IEEE Trans. Automatic Control*, vol. 45, no. 3, pp. 472–482, March 2000.
- [56] I. Arasaratnam, S. Haykin, and R. J. Elliot, “Discrete-time nonlinear filtering algorithms using Gauss-Hermite quadrature,” *Proc. of the IEEE*, vol. 95, no. 5, pp. 953–977, 2007.
- [57] I. Arasaratnam and S. Haykin, “Square-root quadrature Kalman filtering,” *IEEE Trans. Signal Processing*, vol. 56, no. 6, pp. 2589–2593, June 2008.
- [58] ———, “Cubature Kalman filters,” *IEEE Trans. Automatic Control*, vol. 54, no. 6, pp. 1254–1269, June 2009.
- [59] R. V. der Merwe, “Sigma-point Kalman filters for probabilistic inference in dynamic state-space models,” Ph.D. dissertation, OHSU, Portland, Oregon, USA, April 2004.
- [60] H. Afshari, S. Gadsden, and S. Habibi, “Gaussian filters for parameter and state estimation: A general review of theory and recent trends,” *Signal Processing*, vol. 135, pp. 218–238, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0165168417300014>

BIBLIOGRAPHY

- [61] J. Duník, O. Straka, O. Kost, and J. Havlík, “Noise covariance matrices in state-space models: A survey and comparison of estimation methods—Part I,” *Int. J. Adapt. Control Signal Process.*, July 2017.
- [62] R. Mehra, “Approaches to adaptive filtering,” *IEEE Trans. on Automatic Control*, vol. 17, no. 10, pp. 693–698, 1972.
- [63] B. J. Odelson, M. R. Rajamani, and J. B. Rawlings, “A new autocovariance least-square method for estimating noise covariances,” *Automatica*, vol. 42, no. 2, pp. 303–308, 2006.
- [64] J. Duník, O. Straka, and M. Simandl, “On Autocovariance Least-Squares Method for Noise Covariance Matrices Estimation,” *IEEE Transactions on Automatic Control*, vol. 62, no. 2, pp. 967–972, Feb. 2017.
- [65] K. A. Myers and B. D. Tapley, “Adaptive sequential estimation with unknown noise statistics,” *IEEE Trans. on Automatic Control*, vol. 21, no. 8, pp. 520–523, 1976.
- [66] J. Leathrum, “On the sequential estimation of state noise variances,” *IEEE Trans. on Automatic Control*, vol. 26, no. 3, pp. 745–746, 1981.
- [67] P. S. Maybeck, *Stochastic models, estimation, and control. Vol. 2.* Academic Press, 1982.
- [68] I. Blanchet and C. Frankignoul, “A comparison of adaptive Kalman filters for a tropical pacific ocean model,” *Monthly Weather Review*, vol. 125, pp. 40–58, 1997.
- [69] G. C. Goodwin and J. C. Aguero, “Approximate EM algorithms for parameter and state estimation in nonlinear stochastic models,” in *Proc. of the Conf. on Decision and Control*, Newcastle, Australia, 2005, pp. 368–373.
- [70] M. A. Gandhi and L. Mili, “Robust Kalman filter based on a generalized Maximum-Likelihood-type estimator,” *IEEE Trans. on Signal Processing*, vol. 58, no. 5, pp. 2509–2520, 2010.
- [71] V. A. Bavdekar, A. P. Deshpande, and S. C. Patwardhan, “Identification of process and measurement noise covariance for state and parameter estimation using extended Kalman filter,” *Journal of Process Control*, vol. 21, no. 4, pp. 585–601, 2011.
- [72] P. Matisko and V. Havlena, “Noise covariance estimation for Kalman filter tuning using Bayesian approach and Monte Carlo,” *Int. J. Adapt. Control Signal Process.*, vol. 27, p. 957–973, 2013.

BIBLIOGRAPHY

- [73] S. Särkkä and A. Nummenmaa, “Recursive Noise Adaptive Kalman Filtering by Variational Bayesian Approximations,” *IEEE Transactions on Automatic Control*, vol. 54, no. 3, pp. 596–600, 2009.
- [74] S. Särkkä and J. Hartikainen, “Non-Linear Noise Adaptive Kalman Filtering via Variational Bayes,” in *Proc. of the MLSP*, 2013.
- [75] E. Özkan, V. Šmídl, S. Saha, C. Lundquist, and F. Gustafsson, “Marginalized adaptive particle filtering for nonlinear models with unknown time-varying noise parameters,” *Automatica*, vol. 49, no. 6, p. 1566–1575, June 2013.
- [76] J. Duník, M. Šimandl, and O. Straka, “Methods for estimating state and measurement noise covariance matrices: aspects and comparison,” in *Proc. IFAC SYSID*, Saint-Malo, France, July 2009, pp. 372–377.
- [77] D. Alspach and A. Abiri, “A Bayesian solution to the problem of the state estimation in an unknown noise environment,” *Int. Journal of Control*, vol. 19, 1974.
- [78] D. Alspach, “A parallel filtering algorithm for linear systems with unknown time varying noise statistics,” *IEEE Trans. on Automatic Control*, vol. 19, no. 5, pp. 552–556, 1974.
- [79] J. M. Bernardo and A. F. Smith, *Bayesian theory*. John Wiley & Sons, 2009, vol. 405.
- [80] T. Kailath, “An innovations approach to least-squares estimation—Part I: Linear filtering in additive white noise,” *IEEE Transactions on Automatic Control*, vol. 13, no. 6, pp. 646–655, December 1968.
- [81] R. Mehra, “On the identification of variances and adaptive filtering,” *IEEE Trans. on Automatic Control*, vol. 15, no. 2, pp. 175–184, 1970.
- [82] P. A. Frost and T. Kailath, “An innovations approach to least-squares estimation—Part III: Nonlinear Estimation in White Gaussian Noise,” *IEEE Trans. on Automatic Control*, vol. 16, no. 3, pp. 217–226, 1971.
- [83] B. Anderson and J. B. Moore, *Optimal filtering*. Englewood Cliffs, New Jersey, USA: Prentice-Hall, 1979.

BIBLIOGRAPHY

- [84] Y. Bar-Shalom, X. Rong Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation*, 1st ed., ser. A Wiley-Interscience publication. Nashville, TN: John Wiley & Sons, June 2001.
- [85] G. Fasano and A. Franceschini, “A multidimensional version of the Kolmogorov–Smirnov test,” *Monthly Notices of the Royal Astronomical Society*, vol. 225, no. 1, pp. 155–170, 03 1987. [Online]. Available: <https://doi.org/10.1093/mnras/225.1.155>
- [86] P. Closas, J. Vilà-Valls, and C. Fernández-Prades, “Computational complexity reduction techniques for quadrature Kalman filters,” in *Proc. of the CAMSAP’15*, Cancun, Mexico, Dec. 2015.
- [87] I. Arasaratnam, “Cubature kalman filtering: Theory & applications,” Ph.D. dissertation, McMaster University, Hamilton, Ontario, Canada, 2009.
- [88] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *J. Basic Eng., Trans. ASME*, vol. 82, no. 1, pp. 35–45, 1960.
- [89] B. Jia, M. Xin, and Y. Cheng, “Sparse-grid quadrature nonlinear filtering,” *Automatica*, vol. 48, no. 2, pp. 327–341, 2012.
- [90] C. D. Richmond and L. L. Horowitz, “Parameter bounds on estimation accuracy under model misspecification,” *IEEE Transactions on Signal Processing*, vol. 63, no. 9, pp. 2263–2278, 2015.
- [91] C. D. Richmond, “On constraints in parameter estimation and model misspecification,” in *2018 21st International Conference on Information Fusion (FUSION)*. IEEE, 2018, pp. 1080–1085.
- [92] K. B. Petersen and M. S. Pedersen, “The matrix cookbook,” *Technical University of Denmark*, vol. 7, no. 15, p. 510, 2012.