PERFORMANCE PREDICTION FOR IMPLICITLY DEFINED ESTIMATORS OF NON-RANDOM PARAMETERS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ERDAL MEHMETCİK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

JANUARY 2023

Approval of the thesis:

PERFORMANCE PREDICTION FOR IMPLICITLY DEFINED ESTIMATORS OF NON-RANDOM PARAMETERS

submitted by **ERDAL MEHMETCİK** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Electrical and Electronics Engineering Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. İlkay Ulusoy
Head of Department, **Electrical and Electronics Engineering** _____

Prof. Dr. Çağatay Candan
Supervisor, **Electrical and Electronics Eng. Dept., METU** _____

Prof. Dr. Umut Orguner
Co-supervisor, **Electrical and Electronics Eng. Dept., METU** _____

Examining Committee Members:

Prof. Dr. Sinan Gezici
Electrical and Electronics Eng. Dept., İ. D. Bilkent University _____

Prof. Dr. Çağatay Candan
Electrical and Electronics Eng. Dept., METU _____

Prof. Dr. Tolga Çiloğlu
Electrical and Electronics Eng. Dept., METU _____

Prof. Dr. Emre Aktaş
Electrical and Electronics Eng. Dept., Hacettepe University _____

Assist. Prof. Dr. Gökhan Muzaffer Güvensen
Electrical and Electronics Eng. Dept., METU _____

Date:11.01.2023

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Erdal Mehmetcik

Signature :

ABSTRACT

PERFORMANCE PREDICTION FOR IMPLICITLY DEFINED ESTIMATORS OF NON-RANDOM PARAMETERS

Mehmetcik, Erdal
Ph.D., Department of Electrical and Electronics Engineering
Supervisor: Prof. Dr. Çağatay Candan
Co-Supervisor: Prof. Dr. Umut Orguner

January 2023, 118 pages

This thesis study is concerned with performance prediction for estimators with non-random parameters. A rather general class of estimators, called implicitly defined estimators (IDEs), is of main interest. An implicitly defined estimator declares the minimizer/maximizer of a selected cost/reward function as the parameter estimate. The maximum likelihood (ML) and the least squares estimators are among the well-known examples of this class. An exact MSE expression for implicitly defined estimators with a symmetric and unimodal objective function is derived. It is also shown that the expression reduces to the Cramer-Rao lower bound (CRLB) and mis-specified CRLB in the large sample size regime for ML and mis-specified ML estimation, respectively. The expression is shown to yield the Ziv-Zakai bound (without the valley filling function) for the maximum a posteriori (MAP) estimator when it is used in a Bayesian setting, that is, when an a-priori distribution is assigned to the unknown parameter. Extension of the suggested expression to the case of nuisance parameters is studied and some approximations are given to ease the computations for this case. Numerical results indicate that the suggested MSE expression not only pre-

dicts the estimator performance in the asymptotic region; but it is also applicable for the threshold region analysis, even for IDEs whose objective functions do not satisfy the symmetry and unimodality assumptions. Advantages of the suggested MSE expression are its conceptual simplicity and its relatively straightforward numerical calculation due to the reduction of the estimation problem to a binary hypothesis testing problem, similar to the usage of Ziv-Zakai bounds in random parameter estimation problems. The proposed approach for MSE approximation is adapted for bias prediction applications, and similar numerical studies are repeated. Several possible applications of the proposed performance prediction method are studied, and example cases are given.

Keywords: Performance prediction, maximum likelihood, parameter estimation, Rician fading, bias prediction, Cramér-Rao lower bound, Ziv-Zakai lower bound

ÖZ

RASTGELE OLMAYAN PARAMETRELİ ÖRTÜLÜ TANIMLANAN KESTİRİMCİLER İÇİN PERFORMANS TAHMİNİ

Mehmetcik, Erdal
Doktora, Elektrik ve Elektronik Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. Çağatay Candan
Ortak Tez Yöneticisi: Prof. Dr. Umut Orguner

Ocak 2023 , 118 sayfa

Bu tez çalışmasında rastgele olmayan parametreler için kullanılan kestirimcilerin performans kestirimi ele alınmaktadır. Oldukça genel bir kestirimci sınıfı olan örtülü olarak tanımlı kestirimciler ana ilgi odağıdır. Örtülü olarak tanımlı kestirimciler seçilen maliyet/ödül fonksiyonunu minimize/maksimize eden parametreyi kestirim değeri olarak sunan kestirimcilerdir. En yüksek olabilirlik kestirimcisi ve en küçük kareler kestirimcisi bu sınıfa ait en iyi bilinen kestirimciler arasındadır. Simetrik ve tek tepeli hedef fonksiyonuna sahip kestirimciler için hatasız şekilde ortalama kare hata değeri (MSE) veren bir ifade türetilmiştir. Büyük örnek sayısı bölgesinde en yüksek olabilirlik (ML) ve yanlış modellenmiş en yüksek olabilirlik kestirimcileri için, türetilen ifadenin sırasıyla Cramer-Rao alt sınırına ve yanlış modellenme durumundaki Cramer-Rao alt sınırına yakınsadığı gösterilmiştir. Rastgele parametreler için (bilinmeyen parametreye bir dağılım atandığı durumda) ve maksimum artçıl kestirimcisi (MAP) için, türetilen ifadenin Ziv-Zakai alt sınırına (vadi doldurma fonksiyonu olmayan) yakınsadığı gösterilmiştir. Türetilen ifadenin, belirsizliği arttıran parametrelerin

bulunduğu durumda da kullanılabilen hali geliştirilmiş ve hesaplamaları kolaylaştırmak için bazı yaklaşık ifadeler önerilmiştir. Nümerik sonuçlarla, hedef fonksiyonları simetrik ve tek tepeli olmayan kestirimciler için dahi, önerilen MSE ifadesinin sadece asimptotik bölgede değil, aynı zamanda eşik değer bölgesindeki analiz için de kullanılabileceği gösterilmiştir. Önerilen ifade, kestirim problemini ikili hipotez testi problemine çevirdiğinden (rastgele parametre kestirimi problemindeki Ziv-Zakai alt sınırındaki kullanıma benzer olarak) kavramsal olarak daha basit ve nümerik hesaplamalar için daha uygun bir yapı sağlamaktadır. MSE tahmini için türetilen ifadeler yanlılık tahmini için de uyarlanmış ve benzer nümerik testler tekrarlanmıştır. Önerilen ifadelerin uygulamaları için çok sayıda örnek sunulmuştur.

Anahtar Kelimeler: Performans kestirimi, en yüksek olabilirlik, parametre kestirimi, Rician sönümlenme, yanlılık tahmini, Cramér-Rao alt sınırı, Ziv-Zakai alt sınırı

In memory of my dear mother

ACKNOWLEDGMENTS

I would like to express my gratitude to Prof. Dr. Çağatay Candan and Prof. Dr. Umut Orguner for their guidance, encouragement, suggestions and patience. It is a real privilege to have the opportunity to work with them.

I would also like to thank the monitoring committee members Prof. Dr. Tolga Çiloğlu, Prof. Dr. Emre Aktaş, and members of thesis defence committee Prof. Dr. Sinan Gezici and Assist. Prof. Dr. Gökhan M. Güvensen for their valuable comments and suggestions.

I would also like to thank my friends Ali Rıza Kaderoğlu, Alper Avcıoğlu, Ferdi Karaduman and Ramazan Cilasın for their continuous support.

Special thanks to my dear nephew Kağan (the red eagle) for the joy, hope and inspiration he brought into our lives. His immense contributions enabled me to finally complete this work. Speaking of joy reminded me of one other very important contributor, my cat Zarife, thank you for being the kindest creature in my life.

This thesis (and my engineering education) would not be possible without the support and sacrifice of my family, my dear mother Zekiye, my father Ömer and my brother Sezgin. I would like to take this opportunity to express how grateful I am for everything.

TABLE OF CONTENTS

ABSTRACT v

ÖZ vii

ACKNOWLEDGMENTS x

TABLE OF CONTENTS xi

LIST OF FIGURES xiv

LIST OF ABBREVIATIONS xvii

CHAPTERS

1 INTRODUCTION 1

1.1 Motivation and Problem Definition 1

1.2 Contributions and Novelties 5

1.3 The Outline of the Thesis 6

2 BACKGROUND 9

2.1 Random and non-random parameter estimation 9

2.2 Summary of approximate performance prediction methods in literature 12

2.2.1 Cramér-Rao Lower Bound (CRLB) 12

2.2.2 Misspecified Cramér-Rao Lower Bound (MCRB) 13

2.2.3 Barankin Bound (BB) 13

2.2.4 Ziv-Zakai Bound (ZZB) 14

2.2.5 Fessler's method, [1] 14

2.2.6 So's method, [2] 15

3 PERFORMANCE PREDICTION OF IMPLICITLY DEFINED ESTIMATORS 17

3.1 Problem Definition 17

3.2 Case of a Scalar Parameter with Symmetric and Unimodal Objective Functions 18

3.3 Relationship to Performance Bounds 24

3.3.1 Relationship to CRLB 25

3.3.2 Relationship to MCRLB 28

3.3.3 Relationship to ZZB 32

3.4 Extension to the Case with Nuisance Parameters 34

3.5 Application to ML Estimation with the Parametric Mean Model with Gaussian Noise 35

3.5.1 Case of a Scalar Parameter with No Nuisance Parameters . . . 36

3.5.2 Case of a Scalar Parameter with Nuisance Parameters 37

3.5.3 Application to ML Estimation under Model Mismatch 42

3.6 Numerical Results 43

3.6.1 DOA Estimation (No Model Mismatch) 44

3.6.2 DOA Estimation (Model Mismatch) 49

3.6.3 Bayesian DOA Estimation 52

3.7 Approximate bias prediction 54

3.8 Numerical Results 56

4 APPLICATIONS IN DOA ESTIMATION 59

4.1 Direction of Arrival Estimation under Rician Fading 59

4.1.1 Cramér-Rao Lower Bound calculation 60

4.1.2 Approximate performance prediction for ML with parameterized mean model under Rician fading 64

4.1.3 On the accuracy of far-field approximation in direction finding 69

4.2 DOA Estimation by an IDE (ESPRIT) 73

4.3 Online array performance prediction in case of sensor failures 77

4.4 Array layout optimization by minimizing the probability of gross errors 81

5 CONCLUSION 87

REFERENCES 91

APPENDICES

A INFINITE SUPPORT BOUNDED M(·) CASE 97

B MOMENT GENERATING FUNCTION OF ESTIMATION ERROR 99

C INCONSISTENCY OF TAYLOR EXPANSION BASED METHODS . . . 103

D MAXIMUM LIKELIHOOD ESTIMATOR DERIVATION UNDER RICIAN FADING 107

E STEIN'S UNIFIED ANALYSIS OF THE ERROR PROBABILITY 111

F IMPLEMENTATION DETAILS OF THE METHODS USED IN SECTION 3.6 AND SECTION 4.2 113

F.1 Implementation Details for Section 3.6.1 113

F.2 Implementation Details for Section 3.6.2 114

F.3 Implementation Details for Section 4.2 115

F.4 Implementation Details for Section 3.6.3 115

CURRICULUM VITAE 117

LIST OF FIGURES

FIGURES

Figure 1.1 Regions of operations for maximum likelihood estimator and the threshold SNR for a direction of arrival problem. 3

Figure 1.2 Histograms of error signals at different regions of operation. . . . 4

Figure 3.1 Illustration of the fact that the events $(\hat{\theta} - \theta) \geq \epsilon$ and $(\hat{\theta} - \theta) \leq -\epsilon$ are equivalent to the events $\mathcal{L}(\mathbf{x}; \theta + 2\epsilon) \geq \mathcal{L}(\mathbf{x}; \theta)$ and $\mathcal{L}(\mathbf{x}; \theta - 2\epsilon) \geq \mathcal{L}(\mathbf{x}; \theta)$, respectively, when the assumptions of Theorem 1 hold. . 19

Figure 3.2 A Matlab code for predicting the MSE of the ML estimator for the frequency estimation problem. 38

Figure 3.3 Array configuration and array beampattern at true DOA. 46

Figure 3.4 Azimuth estimation performance curves for the array configuration in Figure 3.3a, without and with nuisance parameter (azimuth angle). Note that the y-axis limits are the same for both figures. True DOA: $\bar{\phi} = 25°$, $\bar{\theta} = 60°$. 47

Figure 3.5 Elevation estimation performance curves for the array configuration in Figure 3.3a, without and with nuisance parameter (elevation angle). Note that the y-axis limits are the same for both figures. True DOA: $\bar{\phi} = 25°$, $\bar{\theta} = 60°$. 48

Figure 3.6 12 element uniform circular array with a radius of $\frac{5}{3}\lambda$ and target of interest at 5λ range. 49

Figure 3.7 Far-field performance (a) and near-field performance (b) of a 12 element uniform circular array (Figure 3.6), for a target at 5λ distance. Note that the y-axis limits are the same for both figures. 51

Figure 3.8 Bayesian DOA estimation performance of MAP and ML estimators along with the values of BCRLB, ZZB and the proposed MSE prediction expressions (for ML and MAP). 53

Figure 3.9 Azimuth and elevation estimation bias curves for the array configuration in Figure 3.3a. 57

Figure 3.10 Near-field and far-field DOA estimation bias of a 12 element uniform circular array, for a target at 5λ distance. 58

Figure 4.1 ML performance and CRLB curves for different Rician factors, for a 12-element uniform circular array with $5\lambda/3$ radius. Note that CRLB stays overly optimistic in characterizing the ML performance in for low (≤ 10) Rician factors over a wide range of SNR values. 64

Figure 4.2 Near-field (model mismatch) and far-field (no mismatch) performance of a 12 element uniform circular array, for a target at 5λ distance, under Rician fading, (a) for $K = 10$, and (b) for $K = 15$ 68

Figure 4.3 Beampatterns for a 5 element uniform line array with $\lambda/2$ spacing (using spherical spreading) for different target ranges. 69

Figure 4.4 Beampatterns for a 5 element uniform line array with $\lambda/2$ spacing (using spherical spreading) for different target ranges 70

Figure 4.5 Ratio of predicted MSE values for ML estimator to CRLB for various SNR and target range values, for a Rician factor of 5. Note that the color scale is set between 5 and > 20. 72

Figure 4.6 Ratio of predicted MSE values for ML estimator to CRLB for various SNR and target range values, for a Rician factor of 50. Note that the color scale is set between 1 and > 2. 72

Figure 4.7 Non-random DOA estimation performance of ESPRIT along with values of different bounds and MSE prediction expressions. 74

Figure 4.8 Direction of arrival estimation bias for a uniform linear array of 15 elements with $\lambda/2$ element spacing using ESPRIT. 76

Figure 4.9 Sensor array configuration when all sensors are operational. . . . 79

Figure 4.10 Sensor array performance when all sensors are operational. . . . 79

Figure 4.11 Sensor array configuration when 3 sensors are lost. 80

Figure 4.12 Sensor array performance when 3 sensors are lost. 80

Figure 4.13 8 element uniform circular array and layout optimized array configurations. Note that the search for genetic optimization is constrained to the feasible areas (one sensor within each closed region). 83

Figure 4.14 MSE values for each direction of arrival angle in Φ (using ML estimator) for the uniform circular array in Figure 4.13. 84

Figure 4.15 MSE values for each direction of arrival angle in Φ (using ML estimator) for the optimized array configuration in Figure 4.13. 84

Figure 4.16 Average of MSE values (for ML estimator) for direction of arrival angles in $\Phi = \{\frac{2\pi n}{36}\}_{n=0}^{35}$, for each array configuration. 10^4 Monte Carlo runs are carried out for each azimuth angle. 85

Figure F.1 Grid points of the MIE for azimuth estimation. 114

LIST OF ABBREVIATIONS

BB	Barankin Bound
BCRLB	Bayesian Cramér-Rao Lower Bound
CRLB	Cramér-Rao Lower Bound
DOA	Direction Of Arrival
ESPRIT	Estimation of Signal Parameters via Rotational Invariant Techniques
FIM	Fisher Information Matrix
GLRT	Generalized Likelihood Ratio Test
HCRB	Hammersley Chapman Robbins Bound
IDE	Implicitly Defined Estimator
KL	Kullback Leibler
LLR	Log Likelihood Ratio
MAP	Maximum A Posteriori
MCRLB	Misspecified Cramér-Rao Lower Bound
MIE	Method of Interval Errors
MIMO	Multiple Input Multiple Output
ML	Maximum Likelihood
MML	Misspecified Maximum Likelihood
MSE	Mean Square Error
SNR	Signal to Noise Ratio
ULA	Uniform Line Array
WWB	Weiss Weinstein Bound
ZZB	Ziv-Zakai Bound

CHAPTER 1

INTRODUCTION

1.1 Motivation and Problem Definition

Parameter estimation is one of the major subjects in statistical signal processing. Almost every sensor system utilizes a form of parameter estimation method, either to make sense of the sensor readings (i.e., estimating a parameter) or to improve the accuracy they provide. Two fundamental questions are of interest about the system and the estimator performance,

1. What is the utmost accuracy we can achieve using the given sensor/measurement system?
2. What is the accuracy of a given sensor system and the estimation method we decide to use?

The answer to the first question is provided by the lower limit expressions known as "performance bounds" [3]. These bounds have different uses in many applications. For instance, in the system design stage, one can assess whether the performance requirements are feasible or beyond the fundamental limits by checking these bounds, without the need for the actual system.

Note that, the first question is not specific to which estimator is to be used. In practice one needs to choose an estimator and calculate its performance. Performance bounds do not provide any information about a specific estimator's performance, which in turn brings us to the second question. When a specific estimator is chosen, its performance can be assessed either by deriving the error statistics analytically or resorting to Monte Carlo methods (which is a rather sophisticated way of saying computer

simulations). Deriving the error statistics is in general (except for some trivial cases) is a tedious task. And conducting Monte Carlo analysis requires implementing the estimator along with the simulated test environment and data.

A relatively easier way might be to calculate the performance bounds for a specific estimator, as they can be used to characterize the performance of some estimators (for certain signal to noise ratio values). Such estimators (whose performance converge to the bound) are called *efficient* estimators. An important example to this case is the Cramer Rao Lower Bound (CRLB) [4, 5, 6] and the Maximum Likelihood (ML) estimator [7], under additive white Gaussian noise. CRLB is perhaps the most celebrated lower bound for estimator performance, and the calculation of CRLB is rather straightforward for most problems. Consider the Gaussian case where we have the noisy observation vector $\mathbf{x} \in \mathbb{R}^N$ with the likelihood function $f(\mathbf{x};\theta)$ where θ is the (non-random) parameter to be estimated. CRLB is a lower limit on the variance of any *unbiased* estimator $\hat{\theta}$ with the following expression,

$$\mathrm{var}(\hat{\theta}) \geq \frac{1}{-\mathbb{E}\left\{\frac{\partial^2 \ln f(\mathbf{x};\theta)}{\partial \theta^2}\right\}} = \mathrm{CRLB} \tag{1.1}$$

The maximum likelihood estimator for this problem has the following form:

$$\hat{\theta}_{\mathrm{ML}} = \arg\max_{\theta} f(\mathbf{x};\theta). \tag{1.2}$$

For high signal to noise ratio (SNR) values, ML is known to be an efficient estimator, whose performance approaches CRLB, as illustrated by computer simulations in Figure 1.1. This figure also emphasizes the major drawback of using CRLB in ML performance characterization, as the bound is not tight at all SNR values. The regions in Figure 1.1 have the following names and interpretations;

- *Asymptotic region*, where the performance closely follows CRLB,
- *Threshold region*, where CRLB does not provide a reachable bound, but the estimator is still able to extract information from the observations,
- *No information region*, where the estimator performs (asymptotically) as if the observation vector is composed only of the noise component.

Although the ML estimator performance is calculated using the CRLB at high SNR values, the ML performance quickly deteriorates as SNR decreases below a certain

threshold SNR. In most cases, the main concern is determining this threshold SNR value, as the estimator accuracy starts to depreciate with a small SNR drop from this value.

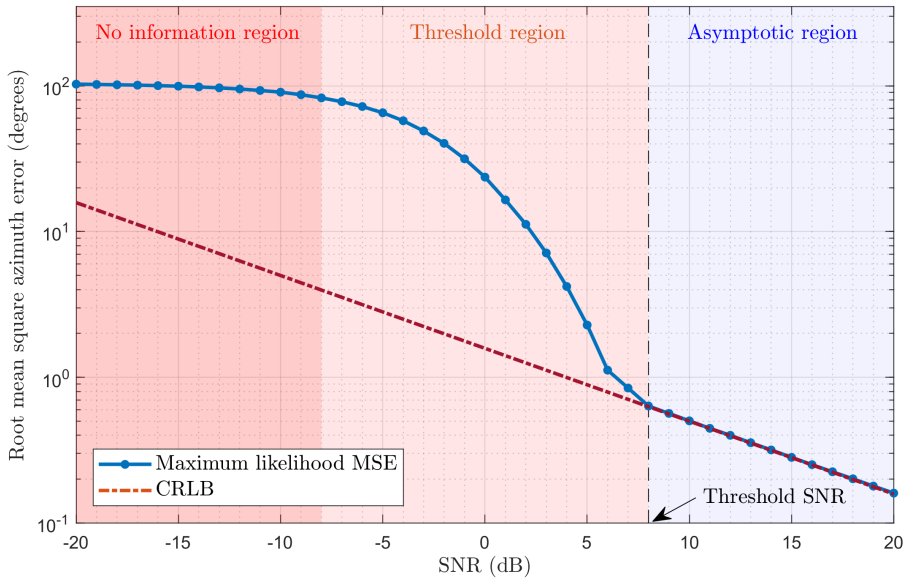


Figure 1.1: Regions of operations for maximum likelihood estimator and the threshold SNR for a direction of arrival problem.

Note that CRLB is still a lower bound in all regions of operation, however it is an *optimistic* bound (meaning that the lower bound is not reachable) outside the asymptotic region. The reason that CRLB diverges from the actual ML performance is the fact that CRLB formulation is concerned with the local errors around the true value of the parameter (note that the definition is only concerned with the derivative at the true parameter value). In other words, as SNR value decreases the estimator output starts to produce results that are far-off from the true parameter. These large deviations are called *gross errors*, and are the main reason of the optimistic behavior of CRLB at low SNR values. Figure 1.2 illustrates the effect of gross errors on ML performance curve. For three different SNR values the histogram of the error signal is provided here, and as seen in the histogram for SNR $= 8$ dB the error distribution has a local peak around zero (i.e. the estimates are within close proximity of the true parameter, no gross errors). On the other hand, in the threshold region the error distribution consists of multiple local peaks. In this specific array processing problem, the

local peaks correspond to the side-lobe positions of the array beam-pattern at the true DOA. Lastly, in the no-information region, the error is distributed almost uniformly over the possible values of the true parameter.

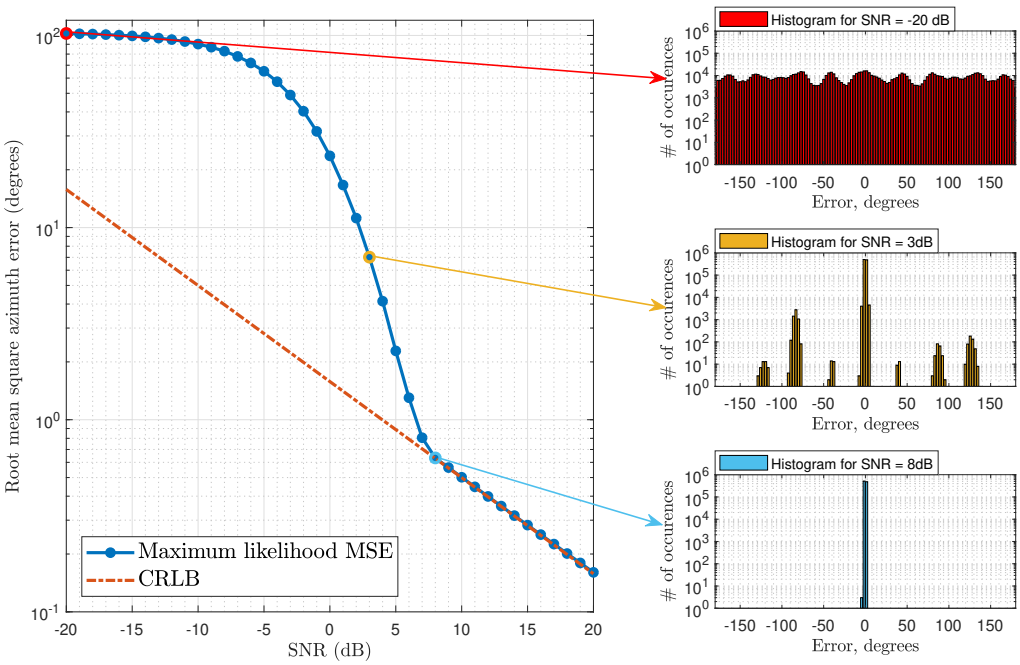


Figure 1.2: Histograms of error signals at different regions of operation.

There are different performance bounds in literature, which can model the threshold effect to a certain degree, as we will explain the issues in detail in Chapter 2.

Another approach to predict the estimator performance is to resort to some approximations. In [1] and [2], to simplify the probability calculations, the Taylor series expansion of the objective function that is minimized by the estimator (in case of ML estimator this is the likelihood function) is used to derive an approximate performance prediction. This approach yields very simple formulas, however as CRLB, the method fails to model the threshold effect, as the Taylor series expansion is only carried out around the true parameter value. Another drawback of this method is that it may yield different performance predictions for the equivalent estimators. For instance, maximizing the likelihood function and the log-likelihood function yields the same estimates, however their Taylor series expansions are not necessarily identical

(An example case is given in Appendix C explaining this phenomenon). Another approximation, the method of interval errors (MIE, [8]), overcomes this problem by approximately calculating the gross error probabilities and using these probability values to generate a weighted sum of asymptotic performance and no-information region performance. The main difficulty here is to approximate these gross error probabilities. After that the calculation of asymptotic performance is also required.

This thesis is mainly concerned about approximately predicting the performance of an estimator (of non-random parameters) for all regions of operations using a rather simple form, as an alternative to the Taylor series expansion based and MIE based methods. The proposed approach is derived for *implicitly defined estimators* (IDEs), which constitute a more general form of estimators maximizing an objective function, with the following form;

$$\hat{\theta} = \arg\max_{\theta} \mathcal{L}(\mathbf{x}; \theta). \tag{1.3}$$

Note that the aforementioned Taylor series expansion based methods and MIE can be applied for this type of estimators as well (with varying degrees of difficulty). The main idea in the proposed approach is to assume (approximate) the objective function $\mathcal{L}(\mathbf{x}; \cdot)$ as a symmetric, uni-modal and strictly decreasing function around the true parameter. This approach simplifies the equations for certain types of problems, as will be explained in detail in the following chapters.

1.2 Contributions and Novelties

In this study, an approximate MSE expression for IDEs of non-random parameters has been proposed, which

- gives the true MSE when the objective function of the IDE is symmetric and unimodal,
- reduces to the CRLB in the large sample size regime for ML estimation,
- reduces to the misspecified CRLB (MCRLB) [9, 10] in the large sample size regime for misspecified ML estimation [10, 11],

- reduces to the ZZB when an a-priori distribution is assigned to the parameter of interest for maximum a posteriori (MAP) estimation.

The proposed approach is extended for bias prediction as well. It is also shown (in Appendix B) that the proposed approach can be utilized to predict any higher order statistics of the error signal through the calculation of the moment generating function. Several applications of the proposed approach are presented. These applications are mentioned briefly in the outline of the thesis in the next section. For example, detailed derivations for the DOA estimation under Rician fading case are given where closed form expressions for the error probability for the proposed approach are calculated. It is also shown that the classical Cramér-Rao Bound fails to represent the Maximum-Likelihood performance even at high SNR values when fading is dominant. Parts of this study have been published in [12] and [13].

1.3 The Outline of the Thesis

The organization of the thesis is as follows;

- Chapter 2 summarizes the background material on random and non-random parameter estimation, performance lower bounds and approximate performance prediction methods for non-random parameter estimation.
- Chapter 3 explains the proposed approach for an approximate performance prediction of implicitly defined estimators (of non-random parameters) in detail. The material in this chapter is published in [13]. The extension of this MSE prediction method to the bias prediction problem is given in this chapter as well.
- Chapter 4 provides different applications using the proposed performance prediction method;

  * Direction of arrival estimation performance prediction under Rician fading channels.

  * Online performance prediction in case of sensor failures (part of this section was published in [12]),

* Array layout optimization by minimizing the probability of gross errors,

* Performance prediction for DOA estimation by ESPRIT method,

- Chapter 5 summarizes the fundamental findings of the overall study and concluding remarks.

- Appendix A describes an exceptional case where the proposed approximation fails to converge.

- Appendix B contains the derivation of the moment generating function of the estimation error using the proposed approximations. Hence higher order statistics of the estimation error can be predicted using the same approach.

- Appendix C provides an example where Taylor series expansion based performance prediction methods [1, 2] yield different performance predictions for equivalent estimators.

- Appendix D contains the derivation of Maximum Likelihood estimator, for parameterized mean model under Rician fading.

- Appendix E contains the summary of results of Stein's unified analysis of the error probability, which was used in Section 4.1.

- Appendix F contains the implementation details of the numerical simulations in Section 3.6.

CHAPTER 2

BACKGROUND

2.1 Random and non-random parameter estimation

The topic of parameter estimation can be divided into two classes, namely the estimation of random and non-random (deterministic) parameters. The random parameter estimation (Bayesian estimation) assumes that the parameter of interest is a random variable with an a-priori distribution and the observations on a realization of the unknown parameter are obtained according to a known probabilistic mapping. Under this setting, the optimal estimator that minimizes the risk, say mean square error (MSE) or mean absolute error, is a functional of the posterior density of the parameter. For instance, the optimal estimator minimizing the MSE is the mean value of the parameter with respect to the posterior density [7]. In general, the posterior density calculation is the key step for the Bayesian formulation. Unfortunately, a closed form expression for the posterior density (and its moments) which does not involve integration, differentiation and limit operations is rarely available. In many problems, one has to resort to the Monte Carlo methods or approximate inference techniques for an inexact realization of the optimal Bayesian estimator. In such problems, the estimator success is typically evaluated by comparisons with the performance bounds. Bayesian performance bounds have a vast literature [3]. Typically, these bounds do not impose any constraints on the estimator. For instance, the Bayesian Cramer Rao lower bound (CRLB) [3, 8], Weiss Weinstein bound (WWB) [14], Bayesian Bhattacharya bound [8] are derived using the covariance inequality principle (hence, sometimes referred as covariance bounds) and applicable in general to any type of estimators. Another main class is the Ziv-Zakai bound (ZZB) [15, 16, 17] type bounds which are derived by converting the estimation problem into a binary detection problem. Ba-

yesian CRLB is one of most fundamental bounds and provides the achievable MSE in the asymptotic region which is the high signal-to-noise ratio (SNR) region. However, it suffers from the threshold effect [15, 16, 17], meaning that it provides unachievable (optimistic) lower bounds at medium or low SNR values. ZZB and WWB are among the tightest Bayesian bounds in all regions of operation [3, 17]. Bayesian bounds continue to be an active research area. Recently, Bayesian bounds for estimating periodic parameters (e.g., phase) have been developed [18, 19, 20].

Non-random parameter estimation involves some challenges unique to this setting. In this setting, an estimator can be improved for a specific value of the parameter at the expense of performance for other parameter values [7, 21]. Since there is no a-priori distribution associated to the parameter of interest, it is not possible to balance the performance gains and losses for different parameter values as in the Bayesian setting. For example, the estimator ignoring the measurements and producing a constant value, say α, as the estimate has no error if the unknown parameter is indeed α; but, suffers from performance losses at all other parameter values. The development of lower bounds for the non-random parameter estimation also suffers from similar inherent admissibility problems. To overcome these problems, the estimators in this setting are typically restricted to the class of unbiased estimators and examined under the title of minimum variance unbiased estimators [7].

The performance bounds for the non-random parameter estimation are also developed for a specific class of estimators. For example, CRLB (for non-random parameters), Hammersley-Chapman-Robbins Bound (HCRB) [22], Barankin Bound [23, 24] require the estimator to be unbiased in an open neighborhood of a point, over a set of two-points and over a set of many-points, respectively (also see [25]). In [26], a general bound form for unbiased estimators is given and it is shown that CRLB, HCRB and BB can be derived by a proper choice of the kernel function of their integral transform.

Note that the unbiasedness condition may not be practical or may be difficult to satisfy, especially for the parameters with a finite support, since the estimation error approaches a one-sided distribution at the edges of the parameter space in such cases [15]. Although it has been shown that for some problems with periodic parameters

[27, 28, 29], the problem with the one-sided error distribution at the edges may vanish; uniformly unbiased estimators do not exist in these cases either [30].

Furthermore, the unbiasedness condition may not even be desirable in some problems. It is known that there exist realizable biased-estimators for some problems whose MSE is lower than the Cramer-Rao bound for unbiased estimators [31]. Perhaps, the most important aspect of unbiasedness condition is in relation with the maximum likelihood estimator. It is well known that the maximum likelihood estimator is unbiased and efficient in the large sample size regime, under fairly general conditions, providing a basis for the theoretical and practical adoption of the unbiasedness condition [32, 33].

The main problem considered in this thesis is prediction of implicitly defined estimators (IDEs) of non-random parameters. IDEs are estimators which produce an estimate by maximizing an objective function of the measurements over the parameter set under consideration. The maximum likelihood (ML) estimator, least squares estimators are some well known examples.

In this study, we present an approximate MSE expression for IDEs of non-random parameters that

- gives the true MSE when the objective function of the IDE is symmetric and unimodal,
- reduces to the CRLB in the large sample size regime for ML estimation,
- reduces to the misspecified CRLB (MCRLB) [9, 10] in the large sample size regime for misspecified ML estimation [10, 11],
- reduces to the ZZB when an a-priori distribution is assigned to the parameter of interest for maximum a posteriori (MAP) estimation.

There are already some approximate MSE expressions available in the literature for IDEs of non-random parameters. For instance, [1] provides formulas for MSE and bias of IDEs using Taylor series expansion of the cost function along with some approximate expressions for certain expectations and derivatives. The study in [2] also uses Taylor expansion approach and derives different approximations in the scalar

parameter case. Both of these approaches are based on the Taylor series expansion around the true parameter value and provide simple MSE expressions; but do not take into account the gross errors which becomes the significant factor as the estimator nonlinearity increases and/or SNR is decreased below the threshold SNR.

The majority of work on the estimator performance prediction focuses on the performance of the ML estimator. The ML estimator is known to be asymptotically efficient (performance approaching CRLB at large sample size) under some regularity conditions [8, 32, 33]. The method of interval errors (MIE) is a celebrated method that was proposed by Van Trees [8] to assess the performance of the ML estimator in the threshold region. This method depends on a careful selection of intervals in the parameter space and the calculation of their probabilities. Different approximations have been proposed to approximate the probabilities [34, 35, 36]. The MSE expression proposed in the present study can be interpreted as a more principled version of the method of interval errors where the need for the interval selection and the gross error probability calculation or approximation is not required.

Notation: Throughout the thesis lower and uppercase letters denote scalars, e.g., a, A. Bold lowercase letters denote vectors, e.g., $\mathbf{a}$. Bold uppercase letters denote matrices, e.g., $\mathbf{A}$. The ith element of the vector $\mathbf{a}$ is denoted by $[\mathbf{a}]_i$. The i, jth element of the matrix $\mathbf{A}$ is denoted by $[\mathbf{A}]_{i,j}$. $\Re\{\cdot\}$ denotes the real part of the complex argument.

2.2 Summary of approximate performance prediction methods in literature

In this section we briefly give the descriptions of some well known performance bounds and approximate performance prediction methods. In all cases, we assume a noisy observation $\mathbf{x} \sim f(\mathbf{x}; \theta)$ of the unknown parameter θ.

2.2.1 Cramér-Rao Lower Bound (CRLB)

The Cramér-Rao Lower Bound is probably the most celebrated lower bound in literature. For any unbiased estimator, the variance of its estimate is bounded below by the

following expression [7],

$$\mathrm{var}(\hat{\theta}) \geq \frac{1}{-\mathbb{E}\left\{\frac{\partial^2 \ln f(\mathbf{x};\theta)}{\partial \theta^2}\right\}} = \mathrm{CRLB} \tag{2.1}$$

where it is assumed that the probability distribution function $f(\mathbf{x};\theta)$ satisfies the following regularity condition

$$\mathbb{E}\left\{\frac{\partial \ln f(\mathbf{x};\theta)}{\partial \theta}\right\} = 0, \forall \theta. \tag{2.2}$$

Equation (2.1) can also be written in the following form,

$$\mathrm{var}(\hat{\theta}) \geq \frac{1}{\mathbb{E}\left\{\left[\frac{\partial \ln f(\mathbf{x};\theta)}{\partial \theta}\right]^2\right\}} = \mathrm{CRLB} \tag{2.3}$$

Note that this form of CRLB is a bound for non-random parameter estimation.

2.2.2 Misspecified Cramér-Rao Lower Bound (MCRB)

When the assumed model $\mathbf{x} \sim f(\mathbf{x};\theta)$ is different than the true model $\mathbf{x} \sim \bar{f}(\mathbf{x})$, it is called a model mismatch. MCRLB is a lower bound for such cases. Following the formulation by [37], the MCRLB is defined as

$$\mathrm{MCRLB} \triangleq \frac{\mathcal{B}(\theta_*)}{\mathcal{A}^2(\theta_*)} \tag{2.4}$$

where,

$$\mathcal{A}(\theta_*) \triangleq \mathbb{E}_{\bar{f}}\left[\frac{\partial^2}{\partial \theta^2} \ln f(x;\theta_*)\right], \quad \mathcal{B}(\theta_*) \triangleq \mathbb{E}_{\bar{f}}\left[\left(\frac{\partial}{\partial \theta} \ln f(x;\theta_*)\right)^2\right]. \tag{2.5}$$

Note that the expectations are with respect to the true density $\bar{f}$, while the assumed density is f. Here θ_* is the parameter value that the mismatched ML estimator converges asymptotically, i.e., as the number of observations increases towards infinity.

2.2.3 Barankin Bound (BB)

Barankin Bound [23, 24, 38] is one of the classical lower bounds on parameter estimation, which makes use of a grid of test points over the parameter space to find a minimum error limit.

$$\mathrm{var}(\hat{\theta}) \geq \mathbf{h}(\mathbf{B} - \mathbf{1}\mathbf{1}^{\mathrm{T}})^{-1}\mathbf{h}^{\mathrm{T}} = \mathrm{BB} \tag{2.6}$$

where

$$\mathbf{h} \triangleq \begin{bmatrix} \theta_1 - \theta & \theta_2 - \theta & \dots & \theta_N - \theta \end{bmatrix}^{\mathrm{T}} \tag{2.7}$$

is the vector composed of the so called test points, $\mathbf{1} = [\ 1 \quad 1 \quad \dots \quad 1\ ]^{\mathrm{T}}$, and the Barankin matrix is $[\mathbf{B}]_{i,j} \triangleq [\mu(\theta_i, \theta_j)]$ with the following definition

$$\mu(\theta_i, \theta_j) \triangleq \int \frac{f(\mathbf{x}|\theta_i) f(\mathbf{x}|\theta_j)}{f(\mathbf{x}|\theta)} \mathrm{d}\mathbf{x} \tag{2.8}$$

Note that BB with a single test point optimized over a grid, is equivalent to the Hammersley-Chapman Robbins Bound (HCRB [22, 25]).

2.2.4 Ziv-Zakai Bound (ZZB)

The Ziv-Zakai Bound (ZZB) is a Bayesian lower bound, different than the non-Bayesian bounds explained in this section. The reason of studying this bound is its relation to the proposed method in Chapter 3. ZZB represents a class of bounds which relate the MSE in an estimation problem to the probability of error in a binary hypothesis testing problem, [3]. ZZB is originally developed in [15] and many improvements over the years have been made [16, 17, 39, 40, 41]. We present the following result without the valley filling function for the scalar parameter case of Ziv-Zakai Bound,

$$\mathrm{ZZB} \triangleq \frac{1}{2} \int_0^{\infty} h \left[\int_{-\infty}^{\infty} \left[p(\theta) + p(\theta + h) \right] P_{\min}(\theta, \theta + h) \mathrm{d}\theta \right] \mathrm{d}h. \tag{2.9}$$

Here $p(\theta)$ is the prior pdf of the random parameter θ, and $P_{\min}(\theta, \theta + h)$ is the minimum probability of error for the following binary hypothesis testing problem: $\mathcal{H}_0 : \mathbf{x} \sim f(\mathbf{x}|\theta)$ and $\mathcal{H}_1 : \mathbf{x} \sim f(\mathbf{x}|\theta + h)$ with the prior hypothesis probabilities of $p(\theta)$ and $p(\theta + h)$, respectively.

2.2.5 Fessler's method, [1]

The method presented by Fessler in [1] is the first approximate performance prediction method we investigate. Note that this is not a lower bound on mean square error (MSE). Fessler's method is concerned with the approximate MSE for estimators with

the following form,

$$\hat{\theta} = \arg\max_{\theta} \mathcal{L}(\mathbf{x}; \theta) \tag{2.10}$$

where $\mathbf{x} = [\ \mathrm{x}_1 \quad \mathrm{x}_2 \quad \ldots \quad \mathrm{x}_N\ ]^{\mathrm{T}}$ is the noisy observation vector and $\mathcal{L}(\cdot;\cdot)$ is the objective function. These types of estimators are called implicitly defined estimators. Fessler's method is based on using the Taylor series expansion of the cost function and implicit function theorem to approximate the MSE of the estimator $\hat{\theta} = h(\mathbf{x})$ as follows,

$$\mathrm{Cov}(\hat{\theta}) \approx \left[\nabla^{20}\mathcal{L}(\bar{\mathbf{x}}; \check{\theta})\right]^{-1} \nabla^{11}\mathcal{L}(\bar{\mathbf{x}}; \check{\theta})\Sigma_{\mathbf{x}} \left[\nabla^{11}\mathcal{L}(\bar{\mathbf{x}}; \check{\theta})\right]^{\mathrm{H}} \left[\nabla^{20}\mathcal{L}(\bar{\mathbf{x}}; \check{\theta})\right]^{-1} \tag{2.11}$$

where $\bar{\mathbf{x}}$ denotes the noiseless measurements, $\check{\theta} = h(\bar{\mathbf{x}})$ is the parameter estimate generated by the estimator function with noiseless measurements, and $\Sigma_{\mathbf{x}}$ is the covariance matrix of the measurement vector. The $(j,k)^{th}$ element of the operators ∇^{20} and ∇^{11} are as follows,

$$\left[\nabla^{20}\right]_{j,k} = \frac{\partial^2}{\partial\theta_j \partial\theta_k}, \quad \left[\nabla^{11}\right]_{j,k} = \frac{\partial^2}{\partial\theta_j \partial\mathrm{x}_k}. \tag{2.12}$$

2.2.6 So's method, [2]

The method by So et al. [2] is another approximate method for MSE prediction. We again consider the estimators of the form given in (2.10). MSE prediction by So's method is as follows,

$$\mathrm{var}\,\hat{\theta} \approx \frac{\mathbb{E}\left\{|\frac{\partial}{\partial\theta}\mathcal{L}(\mathbf{x};\theta)|^2\right\}}{\mathbb{E}\left\{\frac{\partial^2}{\partial\theta^2}\mathcal{L}(\mathbf{x};\theta)\right\}\mathbb{E}\left\{\left[\frac{\partial^2}{\partial\theta^2}\mathcal{L}(\mathbf{x};\theta)\right]^*\right\}} \tag{2.13}$$

CHAPTER 3

PERFORMANCE PREDICTION OF IMPLICITLY DEFINED ESTIMATORS

In this section we explain the main findings of this thesis study. The proposed MSE prediction method for implicitly defined estimators of non-random parameters is explained in detail. As a natural extension of the proposed MSE prediction method, we also present the bias prediction version. The results are tested with different estimation problems using Monte Carlo simulations.

3.1 Problem Definition

We consider the estimation of the non-random real-valued vector $\boldsymbol{\theta} \triangleq [\theta_1 \ \theta_2 \ \cdots \ \theta_J]^{\mathrm{T}}$ from the measurements $\mathbf{x} \triangleq [x_0 \ x_1 \ \cdots \ x_{N-1}]^{\mathrm{T}} \in \mathbb{C}^N$ distributed according to $f(\mathbf{x}; \bar{\boldsymbol{\theta}})$ where $\bar{\boldsymbol{\theta}} \triangleq [\bar{\theta}_1 \ \bar{\theta}_2 \ \cdots \ \bar{\theta}_J]^{\mathrm{T}}$ denotes the true value of $\boldsymbol{\theta}$. An implicitly defined estimator (IDE) generates an estimate $\hat{\boldsymbol{\theta}} \triangleq [\hat{\theta}_1 \ \hat{\theta}_2 \ \cdots \ \hat{\theta}_J]^{\mathrm{T}}$ by maximizing an objective function $\mathcal{L}(\cdot, \cdot)$ of the measurements and the parameters as shown below:

$$\hat{\boldsymbol{\theta}} \triangleq \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\mathbf{x}; \boldsymbol{\theta}). \tag{3.1}$$

The most well-known example of IDEs is the ML estimator where the objective function $\mathcal{L}(\cdot, \cdot)$ is the likelihood function $f(\mathbf{x}; \boldsymbol{\theta})$. Other examples of IDEs are M-estimators and (nonlinear) least square estimators. We see that the estimate $\hat{\boldsymbol{\theta}}$ given by (3.1) is determined by the measurements implicitly, hence the name *implicitly defined estimator*.

In this study we are interested in the performance of IDEs and we give an expression

for the (diagonal elements of the) MSE matrix of the estimate $\hat{\boldsymbol{\theta}}$ which is defined as

$$\mathrm{MSE}(\bar{\boldsymbol{\theta}}) \triangleq \mathbb{E}\big[(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}})^{\mathrm{T}}\big], \tag{3.2}$$

Here it should be mentioned that the methodology presented in the current work can be straightforwardly extended to the other moments of the estimation error $\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}$.

Except for few cases like ML estimation for Gaussian likelihoods with linear models, the estimate $\hat{\boldsymbol{\theta}}$ in (3.1) cannot be analytically expressed in terms of the measurements $\mathbf{x}$, i.e., one cannot find a closed form expression for the function $\mathbf{h}(\cdot)$ such that $\hat{\boldsymbol{\theta}} = \mathbf{h}(\mathbf{x})$. As a consequence, the determination, evaluation and comparison of performance (say, in terms of MSE) of an IDE usually involves extensive Monte Carlo studies and/or problem specific approximations. In this work we first give an MSE expression which is exact for an IDE of a scalar parameter whose objective function is both symmetric (around the estimate) and unimodal in Section 3.2. Since the symmetry and unimodality conditions are typically satisfied by the objective functions of IDEs in the asymptotic or small error region, as further examined in Section 3.3; we suggest using the MSE expression to study the performance of IDEs in the small error and threshold regions. We refrain from calling the suggested MSE expression as a bound due to the lack of performance guarantees in the non-asymptotic regions. The suggested expression can be considered to be in the same league as the MIE [8] which lacks a performance guarantee in all regions including the asymptotic region. Such expressions are also called approximate bounds in some studies [36, 42]. Yet, our main goal in this study is to develop an MSE expression similar to ZZB, which is known to be a tight random parameter estimation bound in the threshold and asymptotic regions, for non-random parameters.

3.2 Case of a Scalar Parameter with Symmetric and Unimodal Objective Functions

In this section we are going to restrict ourselves to a scalar unknown parameter $\theta \in \mathbb{R}$ (i.e., $J = 1$) and provide a predicted MSE expression which is equal to the true MSE for an IDE whose objective function satisfies symmetry and unimodality assumptions. Our main results are given in the following theorem and its corollary.

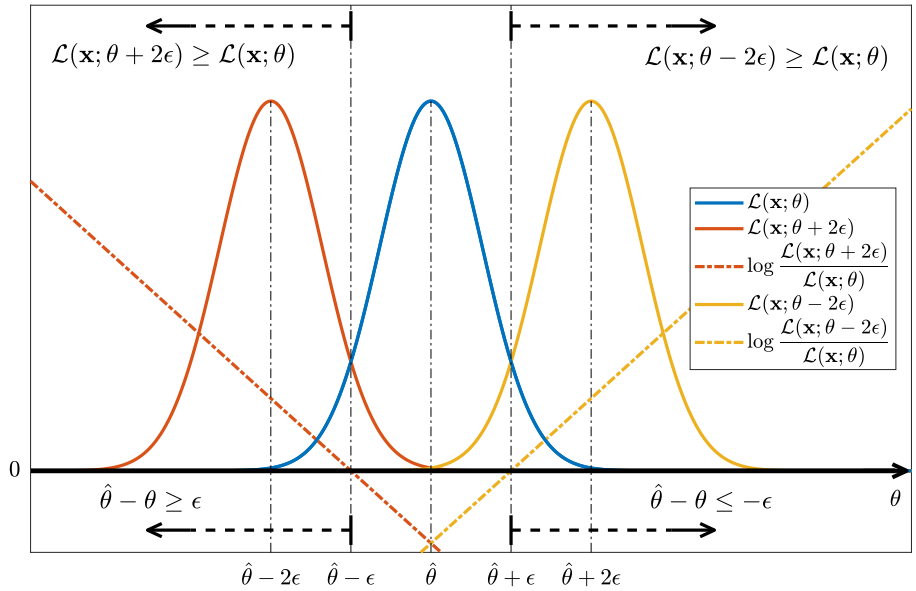


Figure 3.1: Illustration of the fact that the events $(\hat{\theta}-\theta) \geq \epsilon$ and $(\hat{\theta}-\theta) \leq -\epsilon$ are equivalent to the events $\mathcal{L}(\mathbf{x};\theta+2\epsilon) \geq \mathcal{L}(\mathbf{x};\theta)$ and $\mathcal{L}(\mathbf{x};\theta-2\epsilon) \geq \mathcal{L}(\mathbf{x};\theta)$, respectively, when the assumptions of Theorem 1 hold.

Theorem 1 *Consider the IDE given as*

$$\hat{\theta} \triangleq \arg\max_{\theta} \mathcal{L}(\mathbf{x};\theta). \tag{3.3}$$

Let the objective function $\mathcal{L}(\mathbf{x};\cdot)$ satisfy the following conditions for all $\mathbf{x} \in \mathbb{C}^N$.

1. *$\mathcal{L}(\mathbf{x};\hat{\theta}+h) = \mathcal{L}(\mathbf{x};\hat{\theta}-h)$ for all $h \in \mathbb{R}$, i.e., the objective function is symmetric around its peak.*

2. *$\mathcal{L}(\mathbf{x};\theta)$ is strictly-increasing (strictly-decreasing) for $\theta < \hat{\theta}$ ($\theta > \hat{\theta}$).*

Define the true estimator statistic $V_{\hat{\theta}}(\theta)$ as

$$V_{\hat{\theta}}(\theta) \triangleq E[(\hat{\theta}-\theta)^2], \tag{3.4}$$

where θ is an arbitrary fixed parameter value. Then,

$$V_{\hat{\theta}}(\theta) = \widehat{V}_{\hat{\theta}}(\theta), \tag{3.5}$$

where the predicted statistic $\widehat{V}_{\hat{\theta}}(\theta)$ is defined as

$$\boxed{\widehat{V}_{\hat{\theta}}(\theta) \triangleq 2\int_{-\infty}^{\infty} |\epsilon| P\left(\mathcal{L}(\mathbf{x};\theta+2\epsilon) \geq \mathcal{L}(\mathbf{x};\theta)\right) \mathrm{d}\epsilon.} \tag{3.6}$$

Proof: The main idea of the proof is to show that the events $(\hat{\theta}-\theta) \geq \epsilon$ and $(\hat{\theta}-\theta) \leq -\epsilon$ are equivalent to the events $\mathcal{L}(\mathbf{x};\theta+2\epsilon) \geq \mathcal{L}(\mathbf{x};\theta)$ and $\mathcal{L}(\mathbf{x};\theta-2\epsilon) \geq \mathcal{L}(\mathbf{x};\theta)$, respectively, when the assumptions of the theorem are held. An illustration of this equivalence is given in Figure 3.1. Note that the true estimator statistic $V_{\hat{\theta}}(\theta)$ defined in (3.4) can be written as [17, 43]

$$\begin{aligned} V_{\hat{\theta}}(\theta) &= 2\int_0^{\infty} \epsilon P\big(|\hat{\theta}-\theta| \geq \epsilon\big)\, \mathrm{d}\epsilon \\ &= 2\int_0^{\infty} \epsilon\big[P\big(\hat{\theta}-\theta \geq \epsilon\big) + P\big(\hat{\theta}-\theta \leq -\epsilon\big)\big]\, \mathrm{d}\epsilon, \end{aligned} \tag{3.7}$$

and the expression (3.5) follows from (3.7) if the equalities

$$P(\hat{\theta}-\theta \geq \epsilon) = P\left(\mathcal{L}(\mathbf{x};\theta+2\epsilon) \geq \mathcal{L}(\mathbf{x};\theta)\right), \tag{3.8a}$$

$$P(\hat{\theta}-\theta \leq -\epsilon) = P\left(\mathcal{L}(\mathbf{x};\theta-2\epsilon) \geq \mathcal{L}(\mathbf{x};\theta)\right), \tag{3.8b}$$

hold for $\epsilon > 0$. In the following we first show that the equalities in (3.8) indeed hold under the symmetry and unimodality assumptions of Theorem 1. Only the proof of the equality (3.8a) will be made since the proof for (3.8b) is very similar. In order to prove (3.8a), we will show that $\hat{\theta}-\theta \geq \epsilon$ if and only if $\mathcal{L}(\mathbf{x};\theta+2\epsilon) \geq \mathcal{L}(\mathbf{x};\theta)$. The proof has two parts.

- **Proof of the implication $\hat{\theta}-\theta \geq \epsilon \Rightarrow \mathcal{L}(\mathbf{x};\theta+2\epsilon) \geq \mathcal{L}(\mathbf{x};\theta)$:** Suppose that $\hat{\theta} \geq \theta+\epsilon$. Since $\epsilon > 0$, it is clear that $\theta < \hat{\theta}$. If $\theta+2\epsilon < \hat{\theta}$, since $\mathcal{L}(\mathbf{x};\theta)$ is strictly increasing for all $\theta < \hat{\theta}$ and since $\theta < \theta+2\epsilon < \hat{\theta}$, we would have $\mathcal{L}(\mathbf{x};\theta+2\epsilon) > \mathcal{L}(\mathbf{x};\theta)$ and this would make the inequality $\mathcal{L}(\mathbf{x};\theta+2\epsilon) \geq \mathcal{L}(\mathbf{x};\theta)$ hold. Hence, we only need to consider the case $\theta < \hat{\theta} \leq \theta+2\epsilon$. In this case we will show that the inequality $\mathcal{L}(\mathbf{x};\theta+2\epsilon) \geq \mathcal{L}(\mathbf{x};\theta)$ holds by contraposition. Suppose that the reverse inequality, i.e., $\mathcal{L}(\mathbf{x};\theta+2\epsilon) < \mathcal{L}(\mathbf{x};\theta)$, holds. By the symmetry property we have

$$\begin{aligned} \mathcal{L}(\mathbf{x};\theta+2\epsilon) &= \mathcal{L}(\mathbf{x};\hat{\theta}+(\theta+2\epsilon-\hat{\theta})) \\ &= \mathcal{L}(\mathbf{x};\hat{\theta}-(\theta+2\epsilon-\hat{\theta})) \\ &= \mathcal{L}(\mathbf{x};2\hat{\theta}-\theta-2\epsilon), \end{aligned} \tag{3.9}$$

which shows that $\mathcal{L}(\mathbf{x}; 2\hat{\theta} - \theta - 2\epsilon) < \mathcal{L}(\mathbf{x}; \theta)$. Since $2\hat{\theta} - \theta - 2\epsilon \leq \hat{\theta}$ (since $2\hat{\theta} - \theta - 2\epsilon$ is the mirror image of $\theta + 2\epsilon$ (with respect to $\hat{\theta}$), which is greater than or equal to $\hat{\theta}$) and since $\mathcal{L}(\mathbf{x}; \theta)$ is strictly increasing for all $\theta < \hat{\theta}$, the inequality $\mathcal{L}(\mathbf{x}; 2\hat{\theta} - \theta - 2\epsilon) < \mathcal{L}(\mathbf{x}; \theta)$ implies that $2\hat{\theta} - \theta - 2\epsilon < \theta$. This inequality is equivalent to the inequality $\hat{\theta} - \theta < \epsilon$, which completes the proof.

- **Proof of the implication $\mathcal{L}(\mathbf{x}; \theta + 2\epsilon) \geq \mathcal{L}(\mathbf{x}; \theta) \Rightarrow \hat{\theta} - \theta \geq \epsilon$:** Suppose that $\mathcal{L}(\mathbf{x}; \theta + 2\epsilon) \geq \mathcal{L}(\mathbf{x}; \theta)$. Since $\mathcal{L}(\mathbf{x}; \theta)$ is strictly decreasing for all $\theta > \hat{\theta}$, we cannot have $\theta > \hat{\theta}$. Hence, we need to have $\theta \leq \hat{\theta}$. If $\theta + 2\epsilon < \hat{\theta}$, we have $\hat{\theta} - \theta > 2\epsilon > \epsilon$, which makes the inequality $\hat{\theta} - \theta \geq \epsilon$ hold. Hence, we only need to consider the case $\theta \leq \hat{\theta} \leq \theta + 2\epsilon$. By the symmetry property (3.9) we see that $\mathcal{L}(\mathbf{x}; 2\hat{\theta} - \theta - 2\epsilon) = \mathcal{L}(\mathbf{x}; \theta + 2\epsilon) \geq \mathcal{L}(\mathbf{x}; \theta)$. Since we have $2\hat{\theta} - \theta - 2\epsilon \leq \hat{\theta}$ and $\theta \leq \hat{\theta}$ and since $\mathcal{L}(\mathbf{x}; \theta)$ is increasing for all $\theta < \hat{\theta}$, we need to have $2\hat{\theta} - \theta - 2\epsilon \geq \theta$. This inequality is equivalent to the inequality $\hat{\theta} - \theta \geq \epsilon$, which completes the proof.

Hence the equalities in (3.8) hold and we can write (3.7) as

$$V_{\hat{\theta}}(\theta) \triangleq 2 \int_0^{\infty} \epsilon \big[P\left(\mathcal{L}(\mathbf{x}; \theta + 2\epsilon) \geq \mathcal{L}(\mathbf{x}; \theta)\right) + P\left(\mathcal{L}(\mathbf{x}; \theta - 2\epsilon) \geq \mathcal{L}(\mathbf{x}; \theta)\right) \big] \, \mathrm{d}\epsilon, \tag{3.10a}$$

$$= 2 \int_0^{\infty} \epsilon P\left(\mathcal{L}(\mathbf{x}; \theta + 2\epsilon) \geq \mathcal{L}(\mathbf{x}; \theta)\right) \, \mathrm{d}\epsilon + 2 \int_0^{\infty} \epsilon P\left(\mathcal{L}(\mathbf{x}; \theta - 2\epsilon) \geq \mathcal{L}(\mathbf{x}; \theta)\right) \, \mathrm{d}\epsilon, \tag{3.10b}$$

$$= 2 \int_0^{\infty} \epsilon P\left(\mathcal{L}(\mathbf{x}; \theta + 2\epsilon) \geq \mathcal{L}(\mathbf{x}; \theta)\right) \, \mathrm{d}\epsilon + 2 \int_0^{-\infty} \epsilon P\left(\mathcal{L}(\mathbf{x}; \theta + 2\epsilon) \geq \mathcal{L}(\mathbf{x}; \theta)\right) \, \mathrm{d}\epsilon, \tag{3.10c}$$

$$= 2 \int_0^{\infty} \epsilon P\left(\mathcal{L}(\mathbf{x}; \theta + 2\epsilon) \geq \mathcal{L}(\mathbf{x}; \theta)\right) \, \mathrm{d}\epsilon - 2 \int_{-\infty}^{0} \epsilon P\left(\mathcal{L}(\mathbf{x}; \theta + 2\epsilon) \geq \mathcal{L}(\mathbf{x}; \theta)\right) \, \mathrm{d}\epsilon, \tag{3.10d}$$

$$= 2 \int_{-\infty}^{\infty} |\epsilon| P\left(\mathcal{L}(\mathbf{x}; \theta + 2\epsilon) \geq \mathcal{L}(\mathbf{x}; \theta)\right) \, \mathrm{d}\epsilon \triangleq \widehat{V}_{\hat{\theta}}(\theta), \tag{3.10e}$$

which completes the proof. ■

The following remark applies Theorem 1 to find the true MSE of the estimator $\hat{\theta}$.

Remark 1 (MSE of IDE) *The true MSE of the IDE $\hat{\theta}$ in Theorem 1 is given as*

$$\mathrm{MSE}(\bar{\theta}) = \widehat{\mathrm{MSE}}(\bar{\theta}) \tag{3.11}$$

where the predicted MSE, denoted as $\widehat{\mathrm{MSE}}(\bar{\theta})$, is defined as

$$\boxed{\widehat{\mathrm{MSE}}(\bar{\theta}) \triangleq 2\int_{-\infty}^{\infty} |\epsilon| P\left(\mathcal{L}(\mathbf{x};\bar{\theta}+2\epsilon) \geq \mathcal{L}(\mathbf{x};\bar{\theta})\right) \mathrm{d}\epsilon.} \tag{3.12}$$

Proof: The proof is trivial by realizing that $\mathrm{MSE}(\bar{\theta}) = V_{\hat{\theta}}(\bar{\theta})$ and $\widehat{\mathrm{MSE}}(\bar{\theta}) = \widehat{V}_{\hat{\theta}}(\bar{\theta})$. ■

In the special case of a parameter θ with finite support, e.g., $\theta \in [\theta_{\min}, \theta_{\max}]$, the estimation error $\hat{\theta} - \bar{\theta}$ is restricted to the interval $[\theta_{\min} - \bar{\theta},\ \theta_{\max} - \bar{\theta}]$ and the predicted MSE becomes

$$\widehat{\mathrm{MSE}}(\bar{\theta}) = 2\int_{\frac{\theta_{\min}-\bar{\theta}}{2}}^{\frac{\theta_{\max}-\bar{\theta}}{2}} |\epsilon| P\left(\mathcal{L}(\mathbf{x};\bar{\theta}+2\epsilon) \geq \mathcal{L}(\mathbf{x};\bar{\theta})\right)\ \mathrm{d}\epsilon. \tag{3.13}$$

We can interpret the MSE expression (3.12) intuitively as follows. When the probability $P\left(\mathcal{L}(\mathbf{x};\bar{\theta}+2\epsilon) \geq \mathcal{L}(\mathbf{x};\bar{\theta})\right)$ is large for large values of $|\epsilon|$, then it is probable for the IDE $\hat{\theta}$ to make gross errors, resulting in a large MSE. On the other hand, if this probability is small for large values of $|\epsilon|$, the contribution of gross errors in the MSE becomes negligible, resulting in a small MSE. Consequently, IDEs with a small MSE would have the probability $P\left(\mathcal{L}(\mathbf{x};\bar{\theta}+2\epsilon) \geq \mathcal{L}(\mathbf{x};\bar{\theta})\right)$ (thought of as a function of ϵ) highly concentrated in a small neighborhood of $\epsilon = 0$ and *quickly* vanishing elsewhere. More specifically, a sufficient and necessary condition for existence of the integral in the MSE expression (3.12) is $P\left(\mathcal{L}(\mathbf{x};\bar{\theta}+2\epsilon) \geq \mathcal{L}(\mathbf{x};\bar{\theta})\right) = o(1/|\epsilon|^2)$, i.e., the probability $P\left(\mathcal{L}(\mathbf{x};\bar{\theta}+2\epsilon) \geq \mathcal{L}(\mathbf{x};\bar{\theta})\right)$ decaying strictly faster than $1/|\epsilon|^2$ as $|\epsilon| \to \infty$. An exceptional case where the decay is slower than $1/|\epsilon|^2$ with an infinite parameter support is given in Appendix A. The integral in the MSE expression (3.13), on the other hand, always exists.

We can put the expression $\widehat{\mathrm{MSE}}(\bar{\theta})$ to a test by considering the optimal but infeasible estimator $\hat{\theta} = \bar{\theta}$. This estimator can be formulated as an IDE using the objective function $\mathcal{L}(\mathbf{x};\theta) \triangleq -(\theta - \bar{\theta})^2$. Since the objective function $\mathcal{L}(\mathbf{x};\cdot)$ does not depend on the measurements $\mathbf{x}$, we see that the probability of the deterministic event $\mathcal{L}(\mathbf{x};\bar{\theta}+$

$2\epsilon) \geq \mathcal{L}(\mathbf{x};\bar{\theta})$ is given as

$$P\left(\mathcal{L}(\mathbf{x};\bar{\theta}+2\epsilon) \geq \mathcal{L}(\mathbf{x};\bar{\theta})\right) = \begin{cases} 1, & \epsilon = 0 \\ 0, & \text{otherwise} \end{cases}. \tag{3.14}$$

When (3.14) is substituted into (3.12), we have $\widehat{\mathrm{MSE}}(\bar{\theta}) = 0$, which is the true MSE. Similarly, for the feasible (but biased) version of this estimator $\hat{\theta} = \theta_0$, where $\theta_0 \in \mathbb{R}$, the objective function is $\mathcal{L}(\mathbf{x};\theta) \triangleq -(\theta - \theta_0)^2$ and the corresponding probability becomes

$$\begin{aligned} P\left(\mathcal{L}(\mathbf{x};\bar{\theta}+2\epsilon) \geq \mathcal{L}(\mathbf{x};\bar{\theta})\right) &= \begin{cases} 1, & -(\bar{\theta}+2\epsilon-\theta_0)^2 \geq -(\bar{\theta}-\theta_0)^2 \\ 0, & \text{otherwise} \end{cases} \\ &= \begin{cases} 1, & 0 \geq \epsilon \geq -\bar{\theta}+\theta_0 \\ 1, & -\bar{\theta}+\theta_0 \geq \epsilon \geq 0 \\ 0, & \text{otherwise} \end{cases}, \end{aligned} \tag{3.15}$$

which yields $\widehat{\mathrm{MSE}}(\bar{\theta}) = (\bar{\theta} - \theta_0)^2$ when substituted into (3.12) for both $\theta_0 \geq \bar{\theta}$ and $\theta_0 \leq \bar{\theta}$. This also is the true MSE. The following corollary applies the result in Theorem 1 to ML estimation.

Corollary 1 (MSE of ML Estimator) *If the likelihood function $f(\mathbf{x};\cdot)$ satisfies the conditions in Theorem 1, then the true MSE of the ML estimate $\hat{\theta}$ is given as*

$$\mathrm{MSE}_{\mathrm{ML}}(\bar{\theta}) = \widehat{\mathrm{MSE}}_{\mathrm{ML}}(\bar{\theta}) \tag{3.16}$$

where the predicted MSE, denoted as $\widehat{\mathrm{MSE}}_{\mathrm{ML}}(\bar{\theta})$, is defined as

$$\boxed{\widehat{\mathrm{MSE}}_{\mathrm{ML}}(\bar{\theta}) \triangleq 2\int_{-\infty}^{\infty} |\epsilon| P\left(\frac{f(\mathbf{x};\bar{\theta}+2\epsilon)}{f(\mathbf{x};\bar{\theta})} \geq 1\right) \mathrm{d}\epsilon.} \tag{3.17}$$

Proof: Since we have $\mathcal{L}(\mathbf{x};\theta) \triangleq f(\mathbf{x};\theta)$, we can write

$$P\left(\mathcal{L}(\mathbf{x};\bar{\theta}+2\epsilon) \geq \mathcal{L}(\mathbf{x};\bar{\theta})\right) = P\left(f(\mathbf{x};\bar{\theta}+2\epsilon) \geq f(\mathbf{x};\bar{\theta})\right) \tag{3.18a}$$

$$= \int I\left(f(\mathbf{x};\bar{\theta}+2\epsilon) \geq f(\mathbf{x};\bar{\theta})\right) f(\mathbf{x};\bar{\theta})\, \mathrm{d}\mathbf{x} \tag{3.18b}$$

$$= \int_{f(\mathbf{x};\bar{\theta})\neq 0} I\left(f(\mathbf{x};\bar{\theta}+2\epsilon) \geq f(\mathbf{x};\bar{\theta})\right) f(\mathbf{x};\bar{\theta})\, \mathrm{d}\mathbf{x} \tag{3.18c}$$

$$= \int_{f(\mathbf{x};\bar{\theta})\neq 0} I\left(\frac{f(\mathbf{x};\bar{\theta}+2\epsilon)}{f(\mathbf{x};\bar{\theta})} \geq 1\right) f(\mathbf{x};\bar{\theta})\, \mathrm{d}\mathbf{x} \tag{3.18d}$$

$$= P\left(\frac{f(\mathbf{x};\bar{\theta}+2\epsilon)}{f(\mathbf{x};\bar{\theta})} \geq 1\right), \tag{3.18e}$$

where $I(\cdot)$ denotes the indicator function for event arguments. Substituting the last probability into the integral in (3.12) completes the proof. ■

Note that the expression (3.17) connects the MSE of the ML estimate to the error probability of a likelihood ratio test. This connection between estimation and detection theory is further explored in Section 3.3.3 in relation with the ZZB.

Theorem 1 and its corollary provide compact expressions to evaluate the MSE of an implicitly defined estimator exactly even when there is no explicit analytical expression connecting the estimate $\hat{\theta}$ to the measurements $\mathbf{x}$. However, it has some limitations imposed by the assumptions required for its validity. In fact, almost all practical estimation problems violate one of the assumptions of symmetry, unimodality and infinite support (of the parameter θ). For these problems it is certainly possible to have $\mathrm{MSE}(\bar{\theta}) \neq \widehat{\mathrm{MSE}}(\bar{\theta})$. Hence, in general the proposed expressions in (3.12) and (3.17) can only serve as approximate MSE performance prediction tools. Keeping this fact in mind, we show several relations between the suggested MSE expression and well-known bounds in Section 3.3.

3.3 Relationship to Performance Bounds

In this section we present the relationship of the suggested MSE expression (3.17) to some well-known performance bounds.

3.3.1 Relationship to CRLB

In this section, we consider the ML estimation for a scalar parameter $\theta \in \mathbb{R}$. In order to use the large sample asymptotic results for the ML estimate $\hat{\theta}$, we are going to assume that the elements x_n, $n = 0, \ldots, N-1$, of the measurement vector $\mathbf{x}$ are independent and identically distributed as $x_n \sim f(x_n; \bar{\theta})$ [1]. The likelihood for the measurement vector $\mathbf{x}$ is then given as

$$f(\mathbf{x}; \theta) = \prod_{n=0}^{N-1} f(x_n; \theta). \tag{3.19}$$

The relationship of the suggested MSE expression $\widehat{\mathrm{MSE}}_{\mathrm{ML}}(\bar{\theta})$ to CRLB is given in the following proposition.

Proposition 1 *Assume that*

A0 *The parameter θ has finite support, i.e., $\theta \in [\theta_{\min}, \theta_{\max}]$, and the true parameter value $\bar{\theta}$ satisfies $\bar{\theta} \in (\theta_{\min}, \theta_{\max})$.*

A1 *The first three derivatives of $\ln f(x; \theta)$ with respect to θ exist for all θ and are continuous with respect to θ.*

A2 *For every θ, the functions $|\frac{\partial^i}{\partial \theta^i} \ln f(x; \theta)|$, $i = 0, 1, 2, 3$, are dominated by functions $b_i(x)$, $i = 0, 1, 2, 3$, which all have finite variance.*

A3 *The KL divergence $D(f(x; \bar{\theta}) || f(x; \theta))$, where*

$$D(f(x)||g(x)) \triangleq \int f(x) \ln \frac{f(x)}{g(x)} \, \mathrm{d}x, \tag{3.20}$$

has a unique minimum with respect to θ at $\theta = \bar{\theta}$.

A4 *The expectation $\mathbb{E}\left[\frac{\partial^2}{\partial \theta^2} \ln f(x; \bar{\theta})\right]$ is non-zero.*

The assumption A0 is sufficient (but not necessary) for the convergence of the integral in (3.17). *Under the assumptions A1-A3, it can be shown that (See [32, Theorem 2.1])*

[1] In order not to incorporate additional notation, we will keep here the individual measurements $x_n \in \mathbb{C}$ as scalars but the same results can be obtained for the case when x_n is a vector.

the ML estimate $\hat{\theta} \triangleq \arg\max_\theta \ln f(\mathbf{x};\theta)$ *is consistent, i.e.,* $\hat{\theta} \overset{a.s.}{\to} \bar{\theta}$ *as* $N \to \infty$ *. Then, we have*

$$\widehat{\mathrm{MSE}}_{\mathrm{ML}}(\bar{\theta}) \to \mathrm{MSE}_{\mathrm{ML}}(\bar{\theta}) \tag{3.21}$$

as $N \to \infty$*, i.e., the finite support version of the MSE expression* $\widehat{\mathrm{MSE}}_{\mathrm{ML}}(\bar{\theta})$ *in* (3.17) *converges to the true large sample asymptotic MSE of the ML estimate* $\hat{\theta}$ *as* $N \to \infty$*. If the ML estimate is also asymptotically efficient, then we have*

$$\widehat{\mathrm{MSE}}_{\mathrm{ML}}(\bar{\theta}) \to \mathcal{C}(\bar{\theta}), \tag{3.22}$$

as $N \to \infty$ *where* $\mathcal{C}(\bar{\theta}) \triangleq \mathcal{I}^{-1}(\bar{\theta})$ *with* $\mathcal{C}(\bar{\theta})$ *and* $\mathcal{I}(\bar{\theta})$ *denoting the CRLB and the Fisher information matrix, respectively, at the true parameter value* $\bar{\theta}$*.*

Proof: We can write the probability in the integrand of $\widehat{\mathrm{MSE}}_{\mathrm{ML}}(\bar{\theta})$ in (3.17) as

$$P\left(\frac{f(\mathbf{x};\bar{\theta}+2\epsilon)}{f(\mathbf{x};\bar{\theta})} \geq 1\right) = P\left(\ln\frac{f(\mathbf{x};\bar{\theta}+2\epsilon)}{f(\mathbf{x};\bar{\theta})} \geq 0\right) \tag{3.23a}$$

$$= P\left(\frac{1}{N}\ln\frac{f(\mathbf{x};\bar{\theta}+2\epsilon)}{f(\mathbf{x};\bar{\theta})} \geq 0\right) \tag{3.23b}$$

$$= P\left(\frac{1}{N}\ln\frac{f(\mathbf{x};\bar{\theta}+2\epsilon)}{f(\mathbf{x};\bar{\theta})} + D(\bar{\theta}||\bar{\theta}+2\epsilon) \geq D(\bar{\theta}||\bar{\theta}+2\epsilon)\right) \tag{3.23c}$$

$$\leq P\left(\left|\frac{1}{N}\ln\frac{f(\mathbf{x};\bar{\theta}+2\epsilon)}{f(\mathbf{x};\bar{\theta})} + D(\bar{\theta}||\bar{\theta}+2\epsilon)\right| \geq D(\bar{\theta}||\bar{\theta}+2\epsilon)\right) \to 0 \tag{3.23d}$$

for $\epsilon \neq 0$ as $N \to \infty$ where $D(\bar{\theta}||\bar{\theta}+2\epsilon)$ stands for $D(f(x;\bar{\theta})||f(x;\bar{\theta}+2\epsilon))$. This is because we have

$$\frac{1}{N}\ln\frac{f(\mathbf{x}|\bar{\theta}+2\epsilon)}{f(\mathbf{x}|\bar{\theta})} \overset{\mathrm{p}}{\to} -D(\bar{\theta}||\bar{\theta}+2\epsilon) \tag{3.24}$$

as $N \to \infty$ by the law of large numbers and $D(\bar{\theta}||\bar{\theta}+2\epsilon) > 0$ for $\epsilon \neq 0$ due to the assumption A3. As a result, as $N \to \infty$, the integration in (3.17) will be effectively only over an infinitesimal neighborhood of $\epsilon = 0$ and it is only the behavior of the probability $P\big(f(\mathbf{x};\bar{\theta}+2\epsilon)/f(\mathbf{x};\bar{\theta}) \geq 1\big)$ as $\epsilon \to 0$ which determines the MSE expression $\widehat{\mathrm{MSE}}_{\mathrm{ML}}(\bar{\theta})$ in (3.17).

Using the assumption A1, we can now obtain the Taylor expansion of $\ln f(\mathbf{x};\bar{\theta}+2\epsilon)$ around $\epsilon = 0$ given as

$$\begin{aligned}\ln f(\mathbf{x};\bar{\theta}+2\epsilon) = &\ln f(\mathbf{x};\bar{\theta}) + 2\frac{\partial}{\partial\theta}\ln f(\mathbf{x};\bar{\theta})\epsilon \\ &+ 2\frac{\partial^2}{\partial\theta^2}\ln f(\mathbf{x};\bar{\theta})\epsilon^2 + \frac{4}{3}\frac{\partial^3}{\partial\theta^3}\ln f(\mathbf{x};\tilde{\theta})\epsilon^3,\end{aligned} \tag{3.25}$$

where $\tilde{\theta}$ is between $\bar{\theta}$ and $\bar{\theta}+2\epsilon$. Since the $\frac{\partial^3}{\partial\theta^3}\ln f(\mathbf{x};\tilde{\theta})$ is bounded by assumption A2 as $N\to\infty$, the approximation

$$\ln f(\mathbf{x};\bar{\theta}+2\epsilon)\approx\ln f(\mathbf{x};\bar{\theta})+2\frac{\partial}{\partial\theta}\ln f(\mathbf{x};\bar{\theta})\epsilon+2\frac{\partial^2}{\partial\theta^2}\ln f(\mathbf{x};\bar{\theta})\epsilon^2 \tag{3.26}$$

becomes valid as $\epsilon\to 0$. By rearranging, we can write

$$\ln\frac{f(\mathbf{x};\bar{\theta}+2\epsilon)}{f(\mathbf{x};\bar{\theta})}\approx 2\frac{\partial}{\partial\theta}\ln f(\mathbf{x};\bar{\theta})\epsilon+2\frac{\partial^2}{\partial\theta^2}\ln f(\mathbf{x};\bar{\theta})\epsilon^2 \tag{3.27}$$

as $\epsilon\to 0$.

We can also write the Taylor expansion of $\frac{\partial}{\partial\theta}\ln f(\mathbf{x};\hat{\theta})$ around $\theta=\bar{\theta}$ given as

$$\begin{aligned}0=&\frac{\partial}{\partial\theta}\ln f(\mathbf{x};\hat{\theta})\\ =&\frac{\partial}{\partial\theta}\ln f(\mathbf{x};\bar{\theta})+\frac{\partial^2}{\partial\theta^2}\ln f(\mathbf{x};\bar{\theta})(\hat{\theta}-\bar{\theta})+\frac{1}{2}\frac{\partial^3}{\partial\theta^3}\ln f(\mathbf{x};\theta')(\hat{\theta}-\bar{\theta})^2,\end{aligned} \tag{3.28}$$

where θ' is between $\hat{\theta}$ and $\bar{\theta}$. Since the $\frac{\partial^3}{\partial\theta^3}\ln f(\mathbf{x};\theta)$ is bounded by assumption A2 as $N\to\infty$, the approximation

$$0\approx\frac{\partial}{\partial\theta}\ln f(\mathbf{x};\bar{\theta})+\frac{\partial^2}{\partial\theta^2}\ln f(\mathbf{x};\bar{\theta})(\hat{\theta}-\bar{\theta}) \tag{3.29}$$

becomes valid as $\hat{\theta}\overset{\text{a.s.}}{\to}\bar{\theta}$ as $N\to\infty$. Rearranging, we obtain

$$\frac{\partial}{\partial\theta}\ln f(\mathbf{x};\bar{\theta})\approx-\frac{\partial^2}{\partial\theta^2}\ln f(\mathbf{x};\bar{\theta})(\hat{\theta}-\bar{\theta}) \tag{3.30}$$

as $N\to\infty$. Substituting $\frac{\partial}{\partial\theta}\ln f(\mathbf{x};\bar{\theta})$ in (3.30) into (3.27), we get

$$\ln\frac{f(\mathbf{x};\bar{\theta}+2\epsilon)}{f(\mathbf{x};\bar{\theta})}\approx-2\epsilon\frac{\partial^2}{\partial\theta^2}\ln f(\mathbf{x};\bar{\theta})(\hat{\theta}-\bar{\theta}-\epsilon) \tag{3.31}$$

as $\epsilon\to 0$ and $N\to\infty$. We can now substitute the result (3.31) into the the probability in the integrand of $\widehat{\mathrm{MSE}}_{\mathrm{ML}}(\bar{\theta})$ in (3.17) to obtain

$$P\left(\frac{f(\mathbf{x};\bar{\theta}+2\epsilon)}{f(\mathbf{x};\bar{\theta})}\geq 1\right)=P\left(-\epsilon\frac{\partial^2}{\partial\theta^2}\ln f(\mathbf{x};\bar{\theta})(\hat{\theta}-\bar{\theta}-\epsilon)\geq 0\right). \tag{3.32}$$

Using assumptions A1-A4, it can be shown that (See [37, Lemma 2.1 Part-i] or [44, Lemma 4.1 Part-i])

$$-\frac{1}{N}\frac{\partial^2}{\partial\theta^2}\ln f(\mathbf{x};\bar{\theta})\overset{\text{p}}{\to}-\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\ln f(x;\bar{\theta})\right]=\mathcal{I}(\bar{\theta})>0 \tag{3.33}$$

as $N \to \infty$ where we used the law of large numbers. This allows us to write (3.32) as

$$P\left(\frac{f(\mathbf{x};\bar{\theta}+2\epsilon)}{f(\mathbf{x};\bar{\theta})} \geq 1\right) = P\big(\epsilon(\hat{\theta}-\bar{\theta}-\epsilon) \geq 0\big) = \begin{cases} P\big(\hat{\theta}-\bar{\theta} \geq \epsilon\big), & \epsilon > 0 \\ 1, & \epsilon = 0 \\ P\big(\hat{\theta}-\bar{\theta} \leq \epsilon\big), & \epsilon < 0 \end{cases}, \tag{3.34}$$

for $\epsilon \to 0$ and $N \to \infty$. Note that the probabilities $P\big(\hat{\theta}-\bar{\theta} \geq \epsilon\big)$, $\epsilon > 0$ and $P\big(\hat{\theta}-\bar{\theta} \leq \epsilon\big)$, $\epsilon < 0$ would vanish as $N \to \infty$, just as the probability $P\big(f(\mathbf{x};\bar{\theta}+2\epsilon)/f(\mathbf{x};\bar{\theta}) \geq 1\big)$, $\epsilon \neq 0$, itself, thanks to the fact that $\hat{\theta} \xrightarrow{\text{a.s.}} \bar{\theta}$. As a result, we can substitute the right hand side of (3.34) into the finite support version of the integral (3.17) to get

$$\begin{aligned}
\widehat{\mathrm{MSE}}_{\mathrm{ML}}(\bar{\theta}) &\triangleq 2\int_{\frac{\theta_{\min}-\bar{\theta}}{2}}^{\frac{\theta_{\max}-\bar{\theta}}{2}} |\epsilon| P\left(\frac{f(\mathbf{x};\bar{\theta}+2\epsilon)}{f(\mathbf{x};\bar{\theta})} \geq 1\right) \mathrm{d}\epsilon \\
&= 2\int_{\theta_{\min}-\bar{\theta}}^{\theta_{\max}-\bar{\theta}} |\epsilon| P\left(\frac{f(\mathbf{x};\bar{\theta}+2\epsilon)}{f(\mathbf{x};\bar{\theta})} \geq 1\right) \mathrm{d}\epsilon && \text{(3.35a)} \\
&= -2\int_{\theta_{\min}-\bar{\theta}}^{0} \epsilon P\big(\hat{\theta}-\bar{\theta} \leq \epsilon\big)\, \mathrm{d}\epsilon \\
&\quad + 2\int_{0}^{\theta_{\max}-\bar{\theta}} \epsilon P\big(\hat{\theta}-\bar{\theta} \geq \epsilon\big)\, \mathrm{d}\epsilon && \text{(3.35b)} \\
&= \int_{\theta_{\min}-\bar{\theta}}^{0} \epsilon^2 f_{\hat{\theta}-\bar{\theta}}(\epsilon)\, \mathrm{d}\epsilon + \int_{0}^{\theta_{\max}-\bar{\theta}} \epsilon^2 f_{\hat{\theta}-\bar{\theta}}(\epsilon)\, \mathrm{d}\epsilon && \text{(3.35c)} \\
&= \int_{\theta_{\min}-\bar{\theta}}^{\theta_{\max}-\bar{\theta}} \epsilon^2 f_{\hat{\theta}-\bar{\theta}}(\epsilon)\, \mathrm{d}\epsilon = \mathbb{E}[(\hat{\theta}-\bar{\theta})^2] \triangleq \mathrm{MSE}_{\mathrm{ML}}(\bar{\theta}), && \text{(3.35d)}
\end{aligned}$$

as $N \to \infty$, which completes the proof of (3.21). The proof of (3.22) follows trivially if the ML estimate $\hat{\theta}$ is also asymptotically efficient. ■

3.3.2 Relationship to MCRLB

In this section, we consider the misspecified ML (MML) estimation [10, 11] for a scalar parameter $\theta \in \mathbb{R}$. In order to use the asymptotic results for the MML estimate $\hat{\theta}$, we are going to assume that the elements x_n, $n = 0, \ldots, N-1$, of the measurement vector $\mathbf{x}$ are independent and identically distributed as $x_n \sim \bar{f}(x_n)$ where $\bar{f}(\cdot)$ denotes the true measurement distribution. The true distribution for the measurement vector

$\mathbf{x}$ is then given as

$$\bar{f}(\mathbf{x}) = \prod_{n=0}^{N-1} \bar{f}(x_n). \tag{3.36}$$

We assume that MML estimate $\hat{\theta}$ is calculated by maximizing the assumed likelihood $f(\mathbf{x};\theta)$ given in (3.19). The relationship of the suggested variance expression $\widehat{V}_{\hat{\theta}}(\theta)$ to MCRLB [9, 10] is given in the following proposition.

Proposition 2 *Assume that*

- ***A0*** *The parameter θ has finite support, i.e., $\theta \in [\theta_{\min}, \theta_{\max}]$.*
- ***A1*** *The first three derivatives of $\ln f(x;\theta)$ with respect to θ exist for all θ and are continuous with respect to θ.*
- ***A2*** *For every θ, the functions $|\frac{\partial^i}{\partial\theta^i} \ln f(x;\theta)|$, $i = 0, 1, 2, 3$, are dominated by functions $b_i(x)$, $i = 0, 1, 2, 3$, which all have finite variance with respect to the true measurement distribution $\bar{f}(x)$.*
- ***A3*** *The KL divergence $D(\bar{f}(x)||f(x;\theta))$ has a unique minimum with respect to θ at $\theta = \theta_* \in (\theta_{\min}, \theta_{\max})$.*[2]
- ***A4*** *The expectation $\mathbb{E}_{\bar{f}}\left[\frac{\partial^2}{\partial\theta^2} \ln f(x;\theta_*)\right]$ is non-zero.*

The assumption A0 is sufficient (but not necessary) for the convergence of the integral in (3.6). *Under the assumptions A1-A3 it can be shown that (See [32, Theorem 2.1]) the MML estimate $\hat{\theta} \triangleq \arg\max_\theta \ln f(\mathbf{x};\theta)$ is misspecified consistent, i.e., $\hat{\theta} \overset{a.s.}{\to} \theta_*$ as $N \to \infty$. Then, we have*

$$\widehat{V}_{\mathrm{MML}}(\theta_*) \to V_{\mathrm{MML}}(\theta_*) \tag{3.37}$$

as $N \to \infty$, i.e., the finite support version of the expression $\widehat{V}_{\mathrm{MML}}(\theta_*)$ converges to the true large sample asymptotic variance $V_{\mathrm{MML}}(\theta_*)$ of the MML estimate $\hat{\theta}$ as $N \to \infty$. If the MML estimate is also asymptotically misspecified efficient, then we have

$$\widehat{V}_{\mathrm{MML}}(\theta_*) \to \frac{\mathcal{B}(\theta_*)}{\mathcal{A}^2(\theta_*)} \tag{3.38}$$

[2] Note that the existence of the KL divergence $D(\bar{f}(x)||f(x;\theta))$ necessitates additionally the existence of the $E_{\bar{f}}[\ln \bar{f}(x)]$, which we implicitly assume for the sake of conceptual simplicity. We may eliminate the need for the existence of $E_{\bar{f}}[\ln \bar{f}(x)]$ by stating this assumption differently as in [32, 37, 44].

as $N \to \infty$ where the quantity $\frac{\mathcal{B}(\theta_*)}{\mathcal{A}^2(\theta_*)}$ is the MCRLB and

$$\mathcal{A}(\theta_*) \triangleq \mathbb{E}_{\bar{f}}\left[\frac{\partial^2}{\partial\theta^2}\ln f(x;\theta_*)\right], \quad \mathcal{B}(\theta_*) \triangleq \mathbb{E}_{\bar{f}}\left[\left(\frac{\partial}{\partial\theta}\ln f(x;\theta_*)\right)^2\right]. \tag{3.39}$$

Proof: In the case of MML estimation, we set $\mathcal{L}(\mathbf{x};\theta) \triangleq \ln f(\mathbf{x};\theta)$ in the probability in the integrand of $\widehat{V}_{\hat{\theta}}(\theta)$ in (3.6). We can now write the probability in $\widehat{V}_{\hat{\theta}}(\theta)$ as

$$P\big(\ln f(\mathbf{x};\theta_*+2\epsilon) \geq \ln f(\mathbf{x};\theta_*)\big) = P\left(\frac{1}{N}\ln\frac{f(\mathbf{x};\theta_*+2\epsilon)}{f(\mathbf{x};\theta_*)} \geq 0\right) \tag{3.40a}$$

$$= P\left(\frac{1}{N}\ln\frac{\bar{f}(\mathbf{x})}{f(\mathbf{x};\theta_*)} - \frac{1}{N}\ln\frac{\bar{f}(\mathbf{x})}{f(\mathbf{x};\theta_*+2\epsilon)} \geq 0\right) \tag{3.40b}$$

$$= P\Bigg(\frac{1}{N}\ln\frac{\bar{f}(\mathbf{x})}{f(\mathbf{x};\theta_*)} - \frac{1}{N}\ln\frac{\bar{f}(\mathbf{x})}{f(\mathbf{x};\theta_*+2\epsilon)} - \big(D(\theta_*) - D(\theta_*+2\epsilon)\big) \geq D(\theta_*+2\epsilon) - D(\theta_*)\Bigg) \tag{3.40c}$$

$$\leq P\Bigg(\left|\frac{1}{N}\ln\frac{\bar{f}(\mathbf{x})}{f(\mathbf{x};\theta_*)} - \frac{1}{N}\ln\frac{\bar{f}(\mathbf{x})}{f(\mathbf{x};\theta_*+2\epsilon)} - \big(D(\theta_*) - D(\theta_*+2\epsilon)\big)\right| \geq D(\theta_*+2\epsilon) - D(\theta_*)\Bigg) \to 0 \tag{3.40d}$$

for $\epsilon \neq 0$ as $N \to \infty$ where $D(\theta)$ stands for $D(\bar{f}(x)||f(x;\theta))$. This is because we have

$$\frac{1}{N}\ln\frac{\bar{f}(\mathbf{x})}{f(\mathbf{x};\theta_*)} - \frac{1}{N}\ln\frac{\bar{f}(\mathbf{x})}{f(\mathbf{x};\theta_*+2\epsilon)} \xrightarrow{\text{p}} D(\theta_*) - D(\theta_*+2\epsilon) \tag{3.41}$$

as $N \to \infty$ by the law of large numbers and $D(\theta_*+2\epsilon) > D(\theta_*)$ for $\epsilon \neq 0$ due to the assumption A3. As a result, as $N \to \infty$, the integration in (3.12) will be effectively only over an infinitesimal neighborhood of $\epsilon = 0$ and it is only the behavior of the probability $P\big(\ln f(\mathbf{x};\theta_*+2\epsilon) \geq \ln f(\mathbf{x};\theta_*)\big)$ as $\epsilon \to 0$ which determines the expression $\widehat{V}_{\hat{\theta}}(\theta_*)$.

Following a similar approach that is used for obtaining (3.27), we can write

$$\ln\frac{f(\mathbf{x};\theta_*+2\epsilon)}{f(\mathbf{x};\theta_*)} \approx 2\frac{\partial}{\partial\theta}\ln f(\mathbf{x};\theta_*)\epsilon + 2\frac{\partial^2}{\partial\theta^2}\ln f(\mathbf{x};\theta_*)\epsilon^2 \tag{3.42}$$

as $\epsilon \to 0$. Using an approach similar to that used for obtaining (3.30) we can get

$$\frac{\partial}{\partial\theta}\ln f(\mathbf{x};\theta_*) \approx -\frac{\partial^2}{\partial\theta^2}\ln f(\mathbf{x};\theta_*)(\hat{\theta}-\theta_*) \tag{3.43}$$

as $N \to \infty$. Substituting $\frac{\partial}{\partial\theta}\ln f(\mathbf{x};\theta_*)$ in (3.43) into (3.42), we get

$$\ln\frac{f(\mathbf{x};\theta_*+2\epsilon)}{f(\mathbf{x};\theta_*)} \approx -2\epsilon\frac{\partial^2}{\partial\theta^2}\ln f(\mathbf{x};\theta_*)(\hat{\theta}-\theta_*-\epsilon) \tag{3.44}$$

as $\epsilon \to 0$ and $N \to \infty$. We can now substitute the result (3.44) into the probability in the integrand of $\widehat{V}(\theta_*)$

$$P\big(\ln f(\mathbf{x};\theta_*+2\epsilon) \geq \ln f(\mathbf{x};\theta_*)\big) = P\bigg(-\epsilon\frac{\partial^2}{\partial\theta^2}\ln f(\mathbf{x};\theta_*)(\hat{\theta}-\theta_*-\epsilon) \geq 0\bigg). \tag{3.45}$$

Using the assumptions A1-A4, it can be shown that (See [37, Lemma 2.1 Part-i] or [44, Lemma 4.1 Part-i])

$$-\frac{1}{N}\frac{\partial^2}{\partial\theta^2}\ln f(\mathbf{x};\theta_*) \xrightarrow{\text{p}} -\mathbb{E}_{\bar{f}}\left[\frac{\partial^2}{\partial\theta^2}\ln f(x;\theta_*)\right] = -\mathcal{A}(\theta_*) > 0 \tag{3.46}$$

as $N \to \infty$ where we used the law of large numbers. This allows us to write (3.45) as

$$\begin{aligned} P\big(\ln f(\mathbf{x};\theta_*+2\epsilon) \geq \ln f(\mathbf{x};\theta_*)\big) &= P\big(\epsilon(\hat{\theta}-\theta_*-\epsilon) \geq 0\big) \\ &= \begin{cases} P\big(\hat{\theta}-\theta_* \geq \epsilon\big), & \epsilon > 0 \\ 1, & \epsilon = 0 \\ P\big(\hat{\theta}-\theta_* \leq \epsilon\big), & \epsilon < 0 \end{cases}, \end{aligned} \tag{3.47}$$

for $\epsilon \to 0$ and $N \to \infty$. Note that the probabilities $P\big(\hat{\theta}-\theta_* \geq \epsilon\big)$, $\epsilon > 0$ and $P\big(\hat{\theta}-\theta_* \leq \epsilon\big)$, $\epsilon < 0$ would vanish as $N \to \infty$, just as the probability $P\big(\ln f(\mathbf{x};\theta_*+2\epsilon) \geq \ln f(\mathbf{x};\theta_*)\big)$, $\epsilon \neq 0$, itself, thanks to the fact that $\hat{\theta} \xrightarrow{\text{a.s.}} \theta_*$. As a result, we can substitute the right hand side of (3.47) into the finite support version of the integral in $\widehat{V}_{\hat{\theta}}(\theta_*)$ to get

$$\widehat{V}_{\text{MML}}(\theta_*) \triangleq 2\int_{\frac{\theta_{\min}-\theta_*}{2}}^{\frac{\theta_{\max}-\theta_*}{2}} |\epsilon| P\left(\ln f(\mathbf{x};\theta_*+2\epsilon) \geq \ln f(\mathbf{x};\theta_*)\right)\, \mathrm{d}\epsilon \tag{3.48a}$$

$$= 2\int_{\theta_{\min}-\theta_*}^{\theta_{\max}-\theta_*} |\epsilon| P\left(\ln f(\mathbf{x};\theta_*+2\epsilon) \geq \ln f(\mathbf{x};\theta_*)\right)\, \mathrm{d}\epsilon \tag{3.48b}$$

$$\begin{aligned} &= -2\int_{\theta_{\min}-\theta_*}^{0} \epsilon P\big(\hat{\theta}-\theta_* \leq \epsilon\big)\, \mathrm{d}\epsilon \\ &\quad + 2\int_{0}^{\theta_{\max}-\theta_*} \epsilon P\big(\hat{\theta}-\theta_* \geq \epsilon\big)\, \mathrm{d}\epsilon \end{aligned} \tag{3.48c}$$

$$= \int_{\theta_{\min}-\theta_*}^{0} \epsilon^2 f_{\hat{\theta}-\theta_*}(\epsilon)\, \mathrm{d}\epsilon + \int_{0}^{\theta_{\max}-\theta_*} \epsilon^2 f_{\hat{\theta}-\theta_*}(\epsilon)\, \mathrm{d}\epsilon \tag{3.48d}$$

$$= \int_{\theta_{\min}-\theta_*}^{\theta_{\max}-\theta_*} \epsilon^2 f_{\hat{\theta}-\bar{\theta}}(\epsilon)\, \mathrm{d}\epsilon \tag{3.48e}$$

$$= \mathbb{E}[(\hat{\theta}-\theta_*)^2] \triangleq V_{\text{MML}}(\theta_*), \tag{3.48f}$$

as $N \to \infty$, which completes the proof of (3.37). The proof of (3.38) follows trivially if the MML estimate $\hat{\theta}$ is also asymptotically misspecified efficient. ■

Note that according to the proposition, $\widehat{V}(\theta_*)$ converges to the asymptotic variance $V_{\mathrm{MML}}(\theta_*)$ of the MML estimate $\hat{\theta}$. When the true measurement distribution $\bar{f}(\cdot)$ admits the same parameterization as the assumed measurement distribution $f(\cdot;\theta)$ with the true parameter value $\theta = \bar{\theta}$, i.e., $\bar{f}(x) = \bar{f}(x;\bar{\theta})$, then we might predict the MSE performance of the MML estimator $\hat{\theta}$ as

$$\widehat{\mathrm{MSE}}_{\mathrm{MML}}(\bar{\theta}) \triangleq \widehat{V}_{\mathrm{MML}}(\theta_*) + (\theta_* - \bar{\theta})^2, \tag{3.49}$$

which would converge to the true MSE of the MML estimator as $N \to \infty$ if the assumptions of Proposition 2 are satisfied.

3.3.3 Relationship to ZZB

We consider a Bayesian estimation problem where the parameter θ is assigned with the prior distribution $f(\theta)$. The MAP estimate of θ can then be defined as follows.

$$\hat{\theta} \triangleq \arg\max_{\theta} f(\mathbf{x}|\theta) f(\theta) \tag{3.50}$$

where the likelihood $f(\mathbf{x};\theta)$ is shown with the conditioning notation as $f(\mathbf{x}|\theta)$ since θ is now a random variable. Note that the MAP estimator given above corresponds to an IDE with the objective function $\mathcal{L}(\mathbf{x};\theta) \triangleq f(\mathbf{x}|\theta)f(\theta)$. The true MSE of the MAP estimator is given as

$$\begin{aligned}
\mathrm{MSE}_{\mathrm{MAP}} &\triangleq \int\int (\hat{\theta} - \theta)^2 f(\mathbf{x}|\theta)\, \mathrm{d}\mathbf{x} f(\theta)\, \mathrm{d}\theta && (3.51\text{a})\\
&= \mathbb{E}\big[\mathbb{E}\big[(\hat{\theta} - \theta)^2 \big| \theta\big]\big] && (3.51\text{b})\\
&= \mathbb{E}\big[\,\mathrm{MSE}_{\mathrm{MAP}}(\theta)\big], && (3.51\text{c})
\end{aligned}$$

where the outer expectation in (3.51b) is with respect to the random variable θ and $\mathrm{MSE}_{\mathrm{MAP}}(\theta)$ denotes the true MSE of the MAP estimator when θ is given, i.e.,

$$\mathrm{MSE}_{\mathrm{MAP}}(\theta) \triangleq \mathbb{E}\big[(\hat{\theta} - \theta)^2 \big| \theta\big], \tag{3.52}$$

where the expectation is only with respect to the noisy measurements $\mathbf{x}$ given θ. Since the problem becomes a non-random parameter estimation problem when θ is given,

we can predict $\mathrm{MSE}_{\mathrm{MAP}}(\theta)$ of the MAP estimator using (3.12) as follows.

$$\widehat{\mathrm{MSE}}_{\mathrm{MAP}}(\theta) = 2\int_{-\infty}^{\infty} |\epsilon| P\Big(f(\mathbf{x}|\theta+2\epsilon)f(\theta+2\epsilon) \geq f(\mathbf{x}|\theta)f(\theta)\Big|\theta\Big)\,\mathrm{d}\epsilon. \tag{3.53}$$

By substituting the MSE estimate $\widehat{\mathrm{MSE}}_{\mathrm{MAP}}(\theta)$ in (3.53) into the place of $\mathrm{MSE}_{\mathrm{MAP}}(\theta)$ in (3.51c) we can predict the overall MSE of the MAP estimate as follows.

$$\widehat{\mathrm{MSE}}_{\mathrm{MAP}} \triangleq \mathbb{E}[\widehat{\mathrm{MSE}}_{\mathrm{MAP}}(\theta)] \tag{3.54a}$$

$$\triangleq 2\int_{-\infty}^{\infty} f(\theta) \int_{-\infty}^{\infty} |\epsilon| P\Big(f(\mathbf{x}|\theta+2\epsilon)f(\theta+2\epsilon) \geq f(\mathbf{x}|\theta)f(\theta)\Big|\theta\Big)\,\mathrm{d}\epsilon\,\mathrm{d}\theta \tag{3.54b}$$

$$\begin{aligned}=&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\epsilon| f(\theta) P\Big(f(\mathbf{x}|\theta+2\epsilon)f(\theta+2\epsilon) \geq f(\mathbf{x}|\theta)f(\theta)\Big|\theta\Big)\,\mathrm{d}\epsilon\,\mathrm{d}\theta\\ &+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\epsilon| f(\theta-2\epsilon) P\Big(f(\mathbf{x}|\theta-2\epsilon)f(\theta-2\epsilon) \geq f(\mathbf{x}|\theta)f(\theta)\Big|\theta\Big)\,\mathrm{d}\epsilon\,\mathrm{d}\theta\end{aligned} \tag{3.54c}$$

$$\begin{aligned}=&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\epsilon| f(\theta) P\Big(f(\mathbf{x}|\theta+2\epsilon)f(\theta+2\epsilon) \geq f(\mathbf{x}|\theta)f(\theta)\Big|\theta\Big)\,\mathrm{d}\epsilon\,\mathrm{d}\theta\\ &+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\epsilon| f(\theta+2\epsilon) P\Big(f(\mathbf{x}|\theta)f(\theta) \geq f(\mathbf{x}|\theta+2\epsilon)f(\theta+2\epsilon)\Big|\theta+2\epsilon\Big)\,\mathrm{d}\epsilon\,\mathrm{d}\theta\end{aligned} \tag{3.54d}$$

$$\begin{aligned}=&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\epsilon|(f(\theta)+f(\theta+2\epsilon))\\ &\times\Big[\pi_1 P\Big(\pi_2 f(\mathbf{x}|\theta+2\epsilon) \geq \pi_1 f(\mathbf{x}|\theta)\Big|\theta\Big)\\ &\qquad+\pi_2 P\Big(\pi_1 f(\mathbf{x}|\theta) \geq \pi_2 f(\mathbf{x}|\theta+2\epsilon)\Big|\theta+2\epsilon\Big)\Big]\,\mathrm{d}\epsilon\,\mathrm{d}\theta\end{aligned} \tag{3.54e}$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\epsilon|(f(\theta)+f(\theta+2\epsilon))P^{e}_{\min}(\theta,\theta+2\epsilon)\,\mathrm{d}\epsilon\,\mathrm{d}\theta \tag{3.54f}$$

$$\begin{aligned}=&2\int_{0}^{\infty}\int_{-\infty}^{\infty} \epsilon(f(\theta)+f(\theta+2\epsilon))P^{e}_{\min}(\theta,\theta+2\epsilon)\,\mathrm{d}\theta\,\mathrm{d}\epsilon\\ =&\frac{1}{2}\int_{0}^{\infty}\int_{-\infty}^{\infty} \epsilon(f(\theta)+f(\theta+\epsilon))P^{e}_{\min}(\theta,\theta+\epsilon)\,\mathrm{d}\theta\,\mathrm{d}\epsilon,\end{aligned} \tag{3.54g}$$

where $P^{e}_{\min}(\theta_1,\theta_2)$ is the minimum probability of error for the binary hypothesis testing problem given below.

$$\mathcal{H}_1 : \mathbf{x} \sim f(\mathbf{x}|\theta_1), \tag{3.55a}$$

$$\mathcal{H}_2 : \mathbf{x} \sim f(\mathbf{x}|\theta_2), \tag{3.55b}$$

with the prior hypothesis probabilities $P(\mathcal{H}_1) = \pi_1$ and $P(\mathcal{H}_2) = \pi_2 = 1-\pi_1$ where

$$\pi_1 \triangleq \frac{f(\theta_1)}{f(\theta_1)+f(\theta_2)}, \qquad \pi_2 \triangleq \frac{f(\theta_2)}{f(\theta_1)+f(\theta_2)}. \tag{3.56}$$

The expression (3.54g) can be seen to be the ZZB (See [17, Eqn. (14)]) without the so-called valley filling function. As a result $\widehat{\mathrm{MSE}}_{\mathrm{MAP}}$ calculated using (3.12) in a Bayesian framework is equal to the ZZB. Note that this equality is satisfied irrespective of whether the objective function $f(\mathbf{x}|\theta)f(\theta)$ satisfies the assumptions of Theorem 1 or not. If the objective function $f(\mathbf{x}|\theta)f(\theta)$, which is actually the joint density $f(\mathbf{x},\theta)$ of $\mathbf{x}$ and θ, also satisfies the conditions of Theorem 1, then this would mean that $\mathrm{MSE}_{\mathrm{MAP}}(\theta) = \widehat{\mathrm{MSE}}_{\mathrm{MAP}}(\theta)$ for all $\theta \in \mathbb{R}$ and hence $\mathrm{MSE}_{\mathrm{MAP}} = \widehat{\mathrm{MSE}}_{\mathrm{MAP}} = \mathrm{ZZB}$ and hence ZZB would have to be tight, i.e., ZZB would have to be equal to the true average MSE of the MAP estimate $\hat{\theta}$. As a result, the conditions of Theorem 1 are also a set of sufficient conditions for ZZB to be tight.

3.4 Extension to the Case with Nuisance Parameters

Suppose now that we have $J > 1$ unknown scalar parameters, i.e., $\boldsymbol{\theta} \in \mathbb{R}^J$, and we would like to estimate only one of them while keeping the others as unknown nuisance parameters. Without loss of generality we assume that we would like to estimate θ_1 while treating the other parameters $\theta_2, \ldots, \theta_J$ as nuisance parameters. We can express the estimate $\hat{\theta}_1$ for θ_1 as

$$\hat{\theta}_1 \triangleq [\hat{\boldsymbol{\theta}}]_1 = \arg\max_{\theta_1} \underbrace{\left[\max_{\boldsymbol{\theta}_{\setminus 1}} \mathcal{L}(\mathbf{x};\boldsymbol{\theta}) \right]}_{\triangleq \mathcal{L}_1(\mathbf{x},\theta_1)} = \arg\max_{\theta_1} \mathcal{L}_1(\mathbf{x},\theta_1), \tag{3.57}$$

where $\boldsymbol{\theta}_{\setminus 1} \triangleq [\theta_2\ \theta_3\ \cdots\ \theta_J]^{\mathrm{T}}$. If we assume that the function $\mathcal{L}_1(\mathbf{x};\theta_1)$ defined as $\mathcal{L}_1(\mathbf{x};\theta_1) \triangleq \max_{\boldsymbol{\theta}_{\setminus 1}} \mathcal{L}(\mathbf{x};\boldsymbol{\theta})$ satisfies the conditions in Theorem 1, applying the result of Remark 1 to the IDE in (3.57) would give

$$\begin{aligned}\widehat{\mathrm{MSE}}(\bar{\theta}_1) &= 2\int_{-\infty}^{\infty} |\epsilon| P\left(\mathcal{L}_1(\mathbf{x};\bar{\theta}_1+2\epsilon) \geq \mathcal{L}_1(\mathbf{x};\bar{\theta}_1)\right)\, \mathrm{d}\epsilon, \\ &= 2\int_{-\infty}^{\infty} |\epsilon| P\Big(\max_{\boldsymbol{\theta}_{\setminus 1}} \mathcal{L}(\mathbf{x};\bar{\theta}_1+2\epsilon,\boldsymbol{\theta}_{\setminus 1}) \geq \max_{\boldsymbol{\theta}_{\setminus 1}} \mathcal{L}(\mathbf{x};\bar{\theta}_1,\boldsymbol{\theta}_{\setminus 1})\Big)\, \mathrm{d}\epsilon. \end{aligned} \tag{3.58}$$

With the selection $\mathcal{L}(\mathbf{x};\boldsymbol{\theta}) \triangleq f(\mathbf{x};\boldsymbol{\theta}) \geq 0$, we can obtain the MSE of the ML estimate $\hat{\theta}_1$ of θ_1 similarly to Corollary 1 from (3.58) as

$$\widehat{\mathrm{MSE}}_{\mathrm{ML}}(\bar{\theta}_1) = 2\int_{-\infty}^{\infty} |\epsilon| P\left(\frac{\max_{\boldsymbol{\theta}_{\setminus 1}} f(\mathbf{x};\bar{\theta}_1+2\epsilon,\boldsymbol{\theta}_{\setminus 1})}{\max_{\boldsymbol{\theta}_{\setminus 1}} f(\mathbf{x};\bar{\theta}_1,\boldsymbol{\theta}_{\setminus 1})} \geq 1 \right) \mathrm{d}\epsilon, \tag{3.59}$$

connecting the MSE of the ML estimator to the error probability of a generalized likelihood ratio test (GLRT) (instead of a likelihood ratio test) in the presence of nuisance parameters [45].

In Section 3.5.2 below, we are going to investigate the expression (3.59) further on the specific case of the parametric mean model and make approximations to facilitate its calculation, which are later extended to the general case in a remark.

3.5 Application to ML Estimation with the Parametric Mean Model with Gaussian Noise

We consider ML estimator with the measurement model given as

$$\mathbf{x} = \mathbf{m}(\bar{\boldsymbol{\theta}}) + \mathbf{v}, \tag{3.60}$$

where $\mathbf{v} \sim \mathcal{CN}(\mathbf{v}; \mathbf{0}, \sigma^2 \mathbf{I}_N)$ represents the measurement noise and the manifold function $\mathbf{m} : \mathbb{R}^J \to \mathbb{C}^N$ is, in general, a complex-valued function of the unknown parameter vector $\boldsymbol{\theta} \in \mathbb{R}^J$. The measurement model in (3.60) is widely used in signal processing applications. For example, a linear manifold function $\mathbf{m}(\bar{\boldsymbol{\theta}}) = \mathbf{H}\bar{\boldsymbol{\theta}}$ may represent a multi-input multi-output (MIMO) communication system; a non-linear manifold function may represent the array response in the direction of arrival estimation problems [7].

The likelihood function for an arbitrary $\boldsymbol{\theta}$ is given as

$$f(\mathbf{x}; \boldsymbol{\theta}) = \mathcal{CN}(\mathbf{x}; \mathbf{m}(\boldsymbol{\theta}), \sigma^2 \mathbf{I}_N). \tag{3.61}$$

We investigate the cases of a scalar parameter with and without nuisance parameters in different subsections below. In order to calculate the predicted MSE values we will need the following log-likelihood ratio expression.

$$\ln \frac{f(\mathbf{x}; \boldsymbol{\theta})}{f(\mathbf{x}; \bar{\boldsymbol{\theta}})} = \frac{1}{\sigma^2} \big(2\Re\{\tilde{\mathbf{m}}^{\mathrm{H}}(\boldsymbol{\theta}; \bar{\boldsymbol{\theta}})(\mathbf{x} - \mathbf{m}(\bar{\theta}))\} - \|\tilde{\mathbf{m}}(\boldsymbol{\theta}; \bar{\boldsymbol{\theta}})\|^2 \big), \tag{3.62}$$

where $\tilde{\mathbf{m}}(\boldsymbol{\theta}_1; \boldsymbol{\theta}_2) \triangleq \mathbf{m}(\boldsymbol{\theta}_1) - \mathbf{m}(\boldsymbol{\theta}_2)$.

3.5.1 Case of a Scalar Parameter with No Nuisance Parameters

Suppose now that we have a scalar parameter θ with the true value $\bar{\theta}$ ($J = 1$). Note that this case can also be interpreted to be the case when we have multiple parameters $\boldsymbol{\theta} = [\theta_1 \;\; \boldsymbol{\theta}_{\backslash 1}]$ and the true values $\bar{\boldsymbol{\theta}}_{\backslash 1}$ of the nuisance parameters $\boldsymbol{\theta}_{\backslash 1}$ are perfectly known. We can evaluate the probability in (3.17) as

$$P\left(\frac{f(\mathbf{x};\bar{\theta}+2\epsilon)}{f(\mathbf{x};\bar{\theta})} \geq 1\right) = P\left(\ln\frac{f(\mathbf{x};\bar{\theta}+2\epsilon)}{f(\mathbf{x};\bar{\theta})} \geq 0\right) \tag{3.63a}$$

$$= P\left(2\Re\{\tilde{\mathbf{m}}^{\mathrm{H}}(\bar{\theta}+2\epsilon;\bar{\theta})\mathbf{v}\} \geq \|\tilde{\mathbf{m}}(\bar{\theta}+2\epsilon;\bar{\theta})\|^2\right) \tag{3.63b}$$

$$= \mathcal{N}_{\mathrm{ccdf}}\left(\|\tilde{\mathbf{m}}(\bar{\theta}+2\epsilon;\bar{\theta})\|^2; 0, 2\sigma^2\|\tilde{\mathbf{m}}(\bar{\theta}+2\epsilon;\bar{\theta})\|^2\right) \tag{3.63c}$$

$$= \mathcal{N}_{\mathrm{ccdf}}\left(\|\tilde{\mathbf{m}}(\bar{\theta}+2\epsilon;\bar{\theta})\|; 0, 2\sigma^2\right), \tag{3.63d}$$

under the assumption that $\|\tilde{\mathbf{m}}(\bar{\theta}+2\epsilon;\bar{\theta})\| \neq 0$, where $\mathcal{N}_{\mathrm{ccdf}}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma})$ denotes the complementary cumulative distribution function (ccdf) of a real Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ evaluated at $\mathbf{x}$. Assuming that $\|\tilde{\mathbf{m}}(\bar{\theta}+2\epsilon;\bar{\theta})\| \neq 0$ for almost all $\epsilon \in \mathbb{R}$, we can substitute this probability expression into (3.17) to get

$$\widehat{\mathrm{MSE}}_{\mathrm{ML}}(\bar{\theta}) = 2\int_{-\infty}^{\infty} |\epsilon| \mathcal{N}_{\mathrm{ccdf}}\left(\|\tilde{\mathbf{m}}(\bar{\theta}+2\epsilon,\bar{\theta})\|; 0, 2\sigma^2\right) \, \mathrm{d}\epsilon. \tag{3.64}$$

Remark 2 *If both the function* $\mathbf{m}(\cdot)$ *and the measurement noise* $\mathbf{v}$ *are real-valued, i.e., if we have* $\mathbf{m} : \mathbb{R} \to \mathbb{R}^N$ *and* $\mathbf{v} \sim \mathcal{N}(\mathbf{v}; \mathbf{0}, \sigma^2\mathbf{I}_N)$*, then, instead of* (3.64)*, one needs to use*

$$\widehat{\mathrm{MSE}}_{\mathrm{ML}}(\bar{\theta}) = 2\int_{-\infty}^{\infty} |\epsilon| \mathcal{N}_{\mathrm{ccdf}}\left(\|\tilde{\mathbf{m}}(\bar{\theta}+2\epsilon,\bar{\theta})\|; 0, 4\sigma^2\right) \, \mathrm{d}\epsilon. \tag{3.65}$$

Using this expression amounts to replacing the variance σ^2 *in* (3.64) *with* $2\sigma^2$. ■

Note that the likelihood (3.61) does not satisfy the conditions of Theorem 1 and its corollary in general except for some trivial cases, e.g., the case of linear or affine manifold function $\mathbf{m}(\theta)$. As a result, the predicted MSE expressions in (3.64) and (3.65) are expected to be only an approximate estimate of the true MSE of the ML estimator. Furthermore, a closed form solution rarely exists for the integrals in (3.64) and (3.65). Therefore, numerical integration methods have to be used as shown in Example 1 below.

The relations in (3.64) and (3.65) provide some insight on the suggested MSE expression. As $\|\tilde{\mathbf{m}}(\bar{\theta}+2\epsilon,\bar{\theta})\|$, which is the norm of the difference between manifold vectors $\mathbf{m}(\bar{\theta}+2\epsilon)$ and $\mathbf{m}(\bar{\theta})$, gets larger, it should be easier to accurately estimate θ and we get a smaller predicted MSE value (since the function $\mathcal{N}_{\text{ccdf}}(\cdot)$ monotonically decreases as its argument gets larger). Also, it is interesting to see that, for the simplest case $\mathbf{m}(\theta) \triangleq \theta$, the expression for the predicted MSE in (3.64) simplifies to;

$$\widehat{\text{MSE}}_{\text{ML}}(\bar{\theta}) = 2\int_{-\infty}^{\infty} |\epsilon|\mathcal{N}_{\text{ccdf}}\left(|2\epsilon|;0,2\sigma^2\right)\,\mathrm{d}\epsilon = \frac{\sigma^2}{2}, \tag{3.66}$$

which can be obtained by integration by parts. This result is expected, as the corresponding ML estimator is $\hat{\theta} = \Re\{x\}$, hence, the true MSE must be equal to half of the noise variance. We have an exact result since the objective function is a Gaussian likelihood satisfying the conditions of Theorem 1. We finally consider the following example in order to illustrate the practical simplicity of the expression (3.64).

Example 1 (Frequency estimation using ML) *Consider the following signal model.*

$$x_n = Ae^{j\bar{\omega}n} + v_n, \quad n = 0,\ldots,N-1, \tag{3.67}$$

where $A \in \mathbb{C}$ is the known complex amplitude; $\bar{\omega} \in [-\pi,\pi]$ is the unknown true frequency to be estimated using the ML estimator; $v_n \sim \mathcal{CN}(v_n;0,\sigma^2)$, $n = 0,\ldots,N-1$, is the white measurement noise. MSE of the ML estimator based on the measurements x_n, $n = 0,\ldots,N-1$, can be calculated with the Matlab function given in Figure 3.2, which involves only three lines of code. A sample run can be made using the command $MSE_ML_frequency(pi/2,1,1,16)$ for the true frequency value $\bar{\omega} = \pi/2$, amplitude $A = 1$, noise variance $\sigma^2 = 1$ and number of samples $N = 16$ gives $\widehat{\text{MSE}}_{\text{ML}}(\frac{\pi}{2}) =$ `6.417e-4` rad^2. Note that the function `calculateMSEhat(.)` *in Figure 3.2 can be used for predicting the MSE performance of the ML estimator for any measurement model of type* (3.60) *for a scalar parameter $\theta \in [\theta_{\min},\theta_{\max}]$.*

3.5.2 Case of a Scalar Parameter with Nuisance Parameters

When some nuisance parameters exist, we consider the case in Section 3.4 and use the MSE expression in (3.59). Unfortunately it is analytically difficult to calculate the maxima and the probabilities in the integrands on the right hand side of (3.59)

```matlab
function MSEhatML = MSE_ML_frequency(tw, A, sigma2, N)
% Calculate MSEhatML for ML frequency estimate
%     tw: true value of the frequency (scalar) (rad/sec) (-pi < tw < pi)
%      A: known complex amplitude (scalar)
% sigma2: known complex normal noise variance (scalar)
%      N: number of samples (scalar)
m = @(w) A * exp(1i * (0 : N-1)' * w);
MSEhatML = calculateMSEhat(m, tw, -pi, pi, sigma2);

function MSEhatML = calculateMSEhat(m, tt, tMin, tMax, sigma2)
% Calculates the MSEhat for the parameterized mean model
%      m: function handle: m(.) takes a 1xL array of parameter values
%         [theta1 theta2 ... thetaL] and it returns the Nm x L matrix
%         [m(theta1) m(theta2) ... m(thetaL)].
%     tt: True theta value (scalar)
%   tMin: Minimum value of theta (scalar)
%   tMax: Maximum value of theta (scalar)
% sigma2: Complex normal noise variance (scalar)
MSEhatML = 2 * integral(@(e) abs(e) .* ...
    (1 - normcdf(vecnorm(m(tt + 2*e) - m(tt), 2, 1) / sqrt(2*sigma2))),...
    (tMin-tt)/2, (tMax-tt)/2);
```

Figure 3.2: A Matlab code for predicting the MSE of the ML estimator for the frequency estimation problem.

exactly. In the following, we are going to make some approximations to facilitate the calculation. Similar approximations can also be made for the more general case in (3.58) (See Remark 3 below).

$$\begin{aligned}\frac{\max_{\boldsymbol{\theta}_{\backslash 1}} f(\mathbf{x};\bar{\theta}_1+2\epsilon,\boldsymbol{\theta}_{\backslash 1})}{\max_{\boldsymbol{\theta}_{\backslash 1}} f(\mathbf{x};\bar{\theta}_1,\boldsymbol{\theta}_{\backslash 1})} &\approx \frac{\max_{\boldsymbol{\theta}_{\backslash 1}} f(\mathbf{x};\bar{\theta}_1+2\epsilon,\boldsymbol{\theta}_{\backslash 1})}{f(\mathbf{x};\bar{\theta}_1,\bar{\boldsymbol{\theta}}_{\backslash 1})} \\ &= \max_{\boldsymbol{\theta}_{\backslash 1}} \frac{f(\mathbf{x};\bar{\theta}_1+2\epsilon,\boldsymbol{\theta}_{\backslash 1})}{f(\mathbf{x};\bar{\theta}_1,\bar{\boldsymbol{\theta}}_{\backslash 1})} \\ &\approx \max_{\boldsymbol{\theta}_{\backslash 1}\in\boldsymbol{\Theta}_{\backslash 1}} \frac{f(\mathbf{x};\bar{\theta}_1+2\epsilon,\boldsymbol{\theta}_{\backslash 1})}{f(\mathbf{x};\bar{\theta}_1,\bar{\boldsymbol{\theta}}_{\backslash 1})},\end{aligned} \tag{3.68}$$

where the first approximation in (3.68) is made by assuming that the maximum in the denominator of the left hand side is achieved approximately at the true values of the nuisance parameters, i.e., at $\boldsymbol{\theta}_{\backslash 1} = \bar{\boldsymbol{\theta}}_{\backslash 1}$, which is reasonable under asymptotic conditions. The set $\boldsymbol{\Theta}_{\backslash 1} \triangleq \{\boldsymbol{\theta}^1_{\backslash 1}, \boldsymbol{\theta}^2_{\backslash 1}, \ldots, \boldsymbol{\theta}^{N_\theta}_{\backslash 1}\}$ appearing in (3.68) is a set of grid points including the true value $\bar{\boldsymbol{\theta}}_{\backslash 1}$ of $\boldsymbol{\theta}_{\backslash 1}$. Using these approximations, we can approximate

the probability in (3.59) as

$$
\begin{aligned}
&P\left(\frac{\max_{\boldsymbol{\theta}_{\setminus 1}} f(\mathbf{x};\bar{\theta}_1+2\epsilon,\boldsymbol{\theta}_{\setminus 1})}{\max_{\boldsymbol{\theta}_{\setminus 1}} f(\mathbf{x};\bar{\theta}_1,\boldsymbol{\theta}_{\setminus 1})}\geq 1\right)\\
&\quad\approx P\left(\max_{\boldsymbol{\theta}_{\setminus 1}\in\boldsymbol{\Theta}_{\setminus 1}} \frac{f(\mathbf{x};\bar{\theta}_1+2\epsilon,\boldsymbol{\theta}_{\setminus 1})}{f(\mathbf{x};\bar{\theta}_1,\bar{\boldsymbol{\theta}}_{\setminus 1})}\geq 1\right) && (3.69a)\\
&\quad= P\left(\max_{\boldsymbol{\theta}_{\setminus 1}\in\boldsymbol{\Theta}_{\setminus 1}} \ln\frac{f(\mathbf{x};\bar{\theta}_1+2\epsilon,\boldsymbol{\theta}_{\setminus 1})}{f(\mathbf{x};\bar{\theta}_1,\bar{\boldsymbol{\theta}}_{\setminus 1})}\geq 0\right) && (3.69b)\\
&\quad= P\Bigg(\max_{\boldsymbol{\theta}_{\setminus 1}\in\boldsymbol{\Theta}_{\setminus 1}} \Big[2\Re\{\tilde{\mathbf{m}}^{\mathrm{H}}(\bar{\theta}_1+2\epsilon,\boldsymbol{\theta}_{\setminus 1};\bar{\theta}_1,\bar{\boldsymbol{\theta}}_{\setminus 1})\mathbf{v}\}\\
&\qquad\qquad - \|\tilde{\mathbf{m}}(\bar{\theta}_1+2\epsilon,\boldsymbol{\theta}_{\setminus 1};\bar{\theta}_1,\bar{\boldsymbol{\theta}}_{\setminus 1})\|^2\Big]\geq 0\Bigg) && (3.69c)\\
&\quad= 1-P\Bigg(\max_{\boldsymbol{\theta}_{\setminus 1}\in\boldsymbol{\Theta}_{\setminus 1}} \Big[2\Re\{\tilde{\mathbf{m}}^{\mathrm{H}}(\bar{\theta}_1+2\epsilon,\boldsymbol{\theta}_{\setminus 1};\bar{\theta}_1,\bar{\boldsymbol{\theta}}_{\setminus 1})\mathbf{v}\}\\
&\qquad\qquad - \|\tilde{\mathbf{m}}(\bar{\theta}_1+2\epsilon,\boldsymbol{\theta}_{\setminus 1};\bar{\theta}_1,\bar{\boldsymbol{\theta}}_{\setminus 1})\|^2\Big]\leq 0\Bigg) && (3.69d)
\end{aligned}
$$

where $\tilde{\mathbf{m}}(\theta_1^1,\boldsymbol{\theta}_{\setminus 1}^1;\theta_1^2,\boldsymbol{\theta}_{\setminus 1}^2)\triangleq\mathbf{m}(\theta_1^1,\boldsymbol{\theta}_{\setminus 1}^1)-\mathbf{m}(\theta_1^2,\boldsymbol{\theta}_{\setminus 1}^2)$. Let us now define the matrix $\widetilde{\mathbf{M}}_\epsilon\in\mathbb{R}^{N\times N_\theta}$ and the vector $\tilde{\boldsymbol{\mu}}_\epsilon\in\mathbb{R}^{N_\theta}$ as

$$
\widetilde{\mathbf{M}}_\epsilon\triangleq\begin{bmatrix}\tilde{\mathbf{m}}^{\mathrm{H}}(\bar{\theta}_1+2\epsilon,\boldsymbol{\theta}_{\setminus 1}^1;\bar{\theta}_1,\bar{\boldsymbol{\theta}}_{\setminus 1})\\ \vdots\\ \tilde{\mathbf{m}}^{\mathrm{H}}(\bar{\theta}_1+2\epsilon,\boldsymbol{\theta}_{\setminus 1}^{N_\theta};\bar{\theta}_1,\bar{\boldsymbol{\theta}}_{\setminus 1})\end{bmatrix}^{\mathrm{H}} \tag{3.70}
$$

$$
\tilde{\boldsymbol{\mu}}_\epsilon\triangleq\begin{bmatrix}\left\|\tilde{\mathbf{m}}(\bar{\theta}_1+2\epsilon,\boldsymbol{\theta}_{\setminus 1}^1;\bar{\theta}_1,\bar{\boldsymbol{\theta}}_{\setminus 1})\right\|^2\\ \vdots\\ \left\|\tilde{\mathbf{m}}(\bar{\theta}_1+2\epsilon,\boldsymbol{\theta}_{\setminus 1}^{N_\theta};\bar{\theta}_1,\bar{\boldsymbol{\theta}}_{\setminus 1})\right\|^2\end{bmatrix}. \tag{3.71}
$$

We can now write (3.69d) as

$$
\begin{aligned}
P\left(\frac{\max_{\boldsymbol{\theta}_{\setminus 1}} f(\mathbf{x};\bar{\theta}_1+2\epsilon,\boldsymbol{\theta}_{\setminus 1})}{\max_{\boldsymbol{\theta}_{\setminus 1}} f(\mathbf{x};\bar{\theta}_1,\boldsymbol{\theta}_{\setminus 1})}\geq 1\right)&\approx 1-P\Big(2\Re\{\widetilde{\mathbf{M}}_\epsilon^{\mathrm{H}}\mathbf{v}\}\leq\tilde{\boldsymbol{\mu}}_\epsilon\Big|\bar{\boldsymbol{\theta}}\Big)\\
&=\mathcal{N}_{\mathrm{ccdf}}\big(\tilde{\boldsymbol{\mu}}_\epsilon;\mathbf{0},2\sigma^2\Re\{\widetilde{\mathbf{M}}_\epsilon^{\mathrm{H}}\widetilde{\mathbf{M}}_\epsilon\}\big), \qquad (3.72)
\end{aligned}
$$

under the assumption that $\|\tilde{\mathbf{m}}(\bar{\theta}_1+2\epsilon,\boldsymbol{\theta}_{\setminus 1}^i;\bar{\theta}_1,\bar{\boldsymbol{\theta}}_{\setminus 1})\|\neq 0$ for $i=1,\ldots,N_\theta$. Note that the inequalities between vector quantities above should be interpreted in an elementwise manner. The probability given in (3.72) is the generalization of the single parameter probability in (3.63c) to the case of (presence of) nuisance parameters. In fact, when we select the set $\boldsymbol{\Theta}_{\setminus 1}$ as $\boldsymbol{\Theta}_{\setminus 1}=\{\bar{\boldsymbol{\theta}}_{\setminus 1}\}$, i.e., when we have a grid composed of only the true nuisance parameter $\bar{\boldsymbol{\theta}}_{\setminus 1}$, the probability (3.72) reduces to the

probability (3.63c). Moreover, since the set $\boldsymbol{\Theta}_{\setminus 1}$ contains the true value $\bar{\boldsymbol{\theta}}_{\setminus 1}$, the probability (3.72) is always larger than or equal to the probability (3.63c). Substituting the result (3.72) into (3.59) we get

$$\widehat{\mathrm{MSE}}(\bar{\theta}_1) = 2\int_{-\infty}^{\infty} |\epsilon| \mathcal{N}_{\mathrm{ccdf}}\big(\tilde{\boldsymbol{\mu}}_\epsilon; \mathbf{0}, 2\sigma^2 \Re\big\{\widetilde{\mathbf{M}}_\epsilon^{\mathrm{H}} \widetilde{\mathbf{M}}_\epsilon\big\}\big)\, \mathrm{d}\epsilon. \tag{3.73}$$

Note that since the probability (3.72) is always larger than or equal to the probability (3.63c), the predicted MSE in (3.73) is always larger than or equal to the single parameter predicted MSE in (3.64).

Although we ended up with an analytical expression for the predicted MSE in the nuisance parameter case, unfortunately, the calculation of the predicted MSE in (3.73) involves the numerical calculation of the N_θ-variate normal (c)cdf which can be carried out for only small values of the number of grid points N_θ. Furthermore, the covariance matrix $\Re\big\{\widetilde{\mathbf{M}}_\epsilon^{\mathrm{H}} \widetilde{\mathbf{M}}_\epsilon\big\}$ might be ill-conditioned or singular which makes the calculation of the probability even more difficult. As a result, the calculation of the predicted MSE in (3.73) would be computationally infeasible for large grid sizes N_θ. To avoid this calculation, we might follow an alternative approach by approximating the right hand side of (3.69c) as

$$\begin{aligned}
&P\left(\frac{\max_{\boldsymbol{\theta}_{\setminus 1}} f(\mathbf{x}|\bar{\theta}_1 + 2\epsilon, \boldsymbol{\theta}_{\setminus 1})}{\max_{\boldsymbol{\theta}_{\setminus 1}} f(\mathbf{x}|\bar{\theta}_1, \boldsymbol{\theta}_{\setminus 1})} \geq 1\right) \\
&\quad \approx P\Bigg(\max_{\boldsymbol{\theta}_{\setminus 1} \in \boldsymbol{\Theta}_{\setminus 1}} \Big[2\Re\{\tilde{\mathbf{m}}^{\mathrm{H}}(\bar{\theta}_1 + 2\epsilon, \boldsymbol{\theta}_{\setminus 1}; \bar{\theta}_1, \bar{\boldsymbol{\theta}}_{\setminus 1})\mathbf{v}\} \\
&\qquad\qquad - \|\tilde{\mathbf{m}}(\bar{\theta}_1 + 2\epsilon, \boldsymbol{\theta}_{\setminus 1}; \bar{\theta}_1, \bar{\boldsymbol{\theta}}_{\setminus 1})\|^2\Big] \geq 0\Bigg) \qquad (3.74a) \\
&\quad \approx \max_{\boldsymbol{\theta}_{\setminus 1} \in \boldsymbol{\Theta}_{\setminus 1}} P\Bigg(\Big[2\Re\{\tilde{\mathbf{m}}^{\mathrm{H}}(\bar{\theta}_1 + 2\epsilon, \boldsymbol{\theta}_{\setminus 1}; \bar{\theta}_1, \bar{\boldsymbol{\theta}}_{\setminus 1})\mathbf{v}\} \\
&\qquad\qquad - \|\tilde{\mathbf{m}}(\bar{\theta}_1 + 2\epsilon, \boldsymbol{\theta}_{\setminus 1}; \bar{\theta}_1, \bar{\boldsymbol{\theta}}_{\setminus 1})\|^2\Big] \geq 0\Bigg) \qquad (3.74b)
\end{aligned}$$

After these approximations, we can calculate the probability expression for the max-

imization problem as follows:

$$= \max_{\boldsymbol{\theta}_{\setminus 1} \in \boldsymbol{\Theta}_{\setminus 1}} P\Big(2\Re\{\tilde{\mathbf{m}}^{\mathrm{H}}(\bar{\theta}_1 + 2\epsilon, \boldsymbol{\theta}_{\setminus 1}; \bar{\theta}_1, \bar{\boldsymbol{\theta}}_{\setminus 1})\mathbf{v}\} \geq \|\tilde{\mathbf{m}}(\bar{\theta}_1 + 2\epsilon, \boldsymbol{\theta}_{\setminus 1}; \bar{\theta}_1, \bar{\boldsymbol{\theta}}_{\setminus 1})\|^2\Big) \tag{3.75a}$$

$$= \max_{\boldsymbol{\theta}_{\setminus 1} \in \boldsymbol{\Theta}_{\setminus 1}} \mathcal{N}_{\mathrm{ccdf}}\Big(\|\tilde{\mathbf{m}}(\bar{\theta}_1 + 2\epsilon, \boldsymbol{\theta}_{\setminus 1}; \bar{\theta}_1, \bar{\boldsymbol{\theta}}_{\setminus 1})\|^2; 0, 2\sigma^2\|\tilde{\mathbf{m}}(\bar{\theta}_1 + 2\epsilon, \boldsymbol{\theta}_{\setminus 1}; \bar{\theta}_1, \bar{\boldsymbol{\theta}}_{\setminus 1})\|^2\Big) \tag{3.75b}$$

$$= \max_{\boldsymbol{\theta}_{\setminus 1} \in \boldsymbol{\Theta}_{\setminus 1}} \mathcal{N}_{\mathrm{ccdf}}\big(\|\tilde{\mathbf{m}}(\bar{\theta}_1 + 2\epsilon, \boldsymbol{\theta}_{\setminus 1}; \bar{\theta}_1, \bar{\boldsymbol{\theta}}_{\setminus 1})\|; 0, 2\sigma^2\big) \tag{3.75c}$$

$$= \mathcal{N}_{\mathrm{ccdf}}\Big(\min_{\boldsymbol{\theta}_{\setminus 1} \in \boldsymbol{\Theta}_{\setminus 1}} \|\tilde{\mathbf{m}}(\bar{\theta}_1 + 2\epsilon, \boldsymbol{\theta}_{\setminus 1}; \bar{\theta}_1, \bar{\boldsymbol{\theta}}_{\setminus 1})\|; 0, 2\sigma^2\Big), \tag{3.75d}$$

under the assumption that $\min_{\boldsymbol{\theta}_{\setminus 1} \in \boldsymbol{\Theta}_{\setminus 1}} \|\tilde{\mathbf{m}}(\bar{\theta}_1 + 2\epsilon, \boldsymbol{\theta}_{\setminus 1}; \bar{\theta}_1, \bar{\boldsymbol{\theta}}_{\setminus 1})\| \neq 0$. The approximation sign in (3.74a) represents the approximations made until reaching (3.69c). The approximation sign in (3.74b) can be replaced with a greater than equal to sign, i.e., the right hand side of it is a lower bound for the left hand side. Substituting (3.75d) into (3.59) gives the predicted MSE expression shown below:

$$\widehat{\mathrm{MSE}}_{\mathrm{ML}}(\bar{\theta}_1) = 2\int_{-\infty}^{\infty} |\epsilon| \mathcal{N}_{\mathrm{ccdf}}\Big(\min_{\boldsymbol{\theta}_{\setminus 1} \in \boldsymbol{\Theta}_{\setminus 1}} \|\tilde{\mathbf{m}}(\bar{\theta}_1 + 2\epsilon, \boldsymbol{\theta}_{\setminus 1}; \bar{\theta}_1, \bar{\boldsymbol{\theta}}_{\setminus 1})\|; 0, 2\sigma^2\Big)\, \mathrm{d}\epsilon \tag{3.76}$$

The predicted MSE in (3.76) is always smaller than or equal to the computationally prohibitive predicted MSE in (3.73) due to the approximation made in (3.74b), however, it requires the calculation of the ccdf of only a univariate normal random variable. Note that the MSE in (3.76) is still always larger than or equal to the single parameter MSE in (3.64) since the parameter grid $\boldsymbol{\Theta}_{\setminus 1}$ contains the true value $\bar{\boldsymbol{\theta}}_{\setminus 1}$ of $\boldsymbol{\theta}_{\setminus 1}$. This is because of the fact that

$$\min_{\boldsymbol{\theta}_{\setminus 1} \in \boldsymbol{\Theta}_{\setminus 1}} \|\tilde{\mathbf{m}}(\bar{\theta}_1 + 2\epsilon, \boldsymbol{\theta}_{\setminus 1}; \bar{\theta}_1, \bar{\boldsymbol{\theta}}_{\setminus 1})\| \leq \|\tilde{\mathbf{m}}(\bar{\theta}_1 + 2\epsilon, \bar{\boldsymbol{\theta}}_{\setminus 1}; \bar{\theta}_1, \bar{\boldsymbol{\theta}}_{\setminus 1})\| = \|\tilde{\mathbf{m}}(\bar{\theta}_1 + 2\epsilon; \bar{\theta}_1)\| \tag{3.77}$$

and that the function $\mathcal{N}_{\mathrm{ccdf}}(\cdot, 0, 2\sigma^2)$ monotonically increases as its argument gets smaller.

The intuitive meaning of the MSE expression (3.76) can be explained as follows. When the nuisance parameters $\boldsymbol{\theta}_{\setminus 1}$ are known, i.e., we have the case of a single parameter in Section 3.5.1, the MSE is seen to be dependent on the distance between the mean vector $\mathbf{m}(\bar{\theta}_1 + 2\epsilon, \bar{\boldsymbol{\theta}}_{\setminus 1})$ and the true mean vector $\mathbf{m}(\bar{\theta}_1, \bar{\boldsymbol{\theta}}_{\setminus 1})$, which was shown

as (the magnitude of) the vector $\tilde{\mathbf{m}}(\bar{\theta}_1 + 2\epsilon; \bar{\theta}_1) \triangleq \tilde{\mathbf{m}}(\bar{\theta}_1 + 2\epsilon, \bar{\boldsymbol{\theta}}_{\backslash 1}; \bar{\theta}_1, \bar{\boldsymbol{\theta}}_{\backslash 1})$ in (3.64). On the other hand, when the the nuisance parameters $\boldsymbol{\theta}_{\backslash 1}$ are not known, the predicted MSE is dependent on minimum distance between the mean vectors $\mathbf{m}(\bar{\theta}_1 + 2\epsilon, \boldsymbol{\theta}_{\backslash 1})$, where $\boldsymbol{\theta}_{\backslash 1}$ takes values in a grid containing the true nuisance parameter value $\bar{\boldsymbol{\theta}}_{\backslash 1}$, and the fixed true mean vector $\mathbf{m}(\bar{\theta}_1, \bar{\boldsymbol{\theta}}_{\backslash 1})$. Hence if the vector $\mathbf{m}(\bar{\theta}_1 + 2\epsilon, \boldsymbol{\theta}_{\backslash 1})$ is similar to the fixed vector $\mathbf{m}(\bar{\theta}_1, \bar{\boldsymbol{\theta}}_{\backslash 1})$ for some values of the nuisance parameter $\boldsymbol{\theta}_{\backslash 1}$ in the grid, the resulting predicted MSE would get larger.

Remark 3 *The approximations made in this section on* (3.59) *can be applied to the general case* (3.58) *as follows.*

$$
\begin{aligned}
&P\Big(\max_{\boldsymbol{\theta}_{\backslash 1}} \mathcal{L}(\mathbf{x}; \bar{\theta}_1 + 2\epsilon, \boldsymbol{\theta}_{\backslash 1}) \geq \max_{\boldsymbol{\theta}_{\backslash 1}} \mathcal{L}(\mathbf{x}; \bar{\theta}_1, \boldsymbol{\theta}_{\backslash 1})\Big) \\
&\quad \approx P\Big(\max_{\boldsymbol{\theta}_{\backslash 1} \in \boldsymbol{\Theta}_{\backslash 1}} \mathcal{L}(\mathbf{x}; \bar{\theta}_1 + 2\epsilon, \boldsymbol{\theta}_{\backslash 1}) \geq \mathcal{L}(\mathbf{x}; \bar{\theta}_1, \bar{\boldsymbol{\theta}}_{\backslash 1})\Big), \qquad (3.78a) \\
&\quad \approx \max_{\boldsymbol{\theta}_{\backslash 1} \in \boldsymbol{\Theta}_{\backslash 1}} P\big(\mathcal{L}(\mathbf{x}; \bar{\theta}_1 + 2\epsilon, \boldsymbol{\theta}_{\backslash 1}) \geq \mathcal{L}(\mathbf{x}; \bar{\theta}_1, \bar{\boldsymbol{\theta}}_{\backslash 1})\big). \qquad (3.78b)
\end{aligned}
$$

3.5.3 Application to ML Estimation under Model Mismatch

In this section we consider the problem of ML estimation with the parametric mean model under model mismatch, also known as misspecified ML (MML) estimation in the literature [11, 10]. For the sake of simplicity we consider only the scalar parameter case, i.e., $\theta \in \mathbb{R}$. The measurements $\mathbf{x}$ are modeled as $\mathbf{x} = \bar{\mathbf{m}}(\bar{\theta}) + \mathbf{v}$, where $\bar{\mathbf{m}}(\cdot)$ denotes the true mean function and $\mathbf{v} \sim \mathcal{CN}(\mathbf{v}; \mathbf{0}, \bar{\sigma}^2 \mathbf{I}_N)$ represents the measurement noise with the true variance $\bar{\sigma}^2$. This model corresponds to the true likelihood $\bar{f}(\mathbf{x}; \bar{\theta}) \triangleq \mathcal{CN}(\mathbf{x}; \bar{\mathbf{m}}(\bar{\theta}), \bar{\sigma}^2 \mathbf{I}_N)$. We are interested in the MSE of the mismatched ML estimator $\hat{\theta}$ of θ given as

$$
\hat{\theta} \triangleq \arg\max_{\theta} f(\mathbf{x}; \theta), \qquad (3.79)
$$

where the objective function is the assumed likelihood $f(\mathbf{x}; \theta)$ given as

$$
f(\mathbf{x}; \theta) \triangleq \mathcal{CN}(\mathbf{x}; \mathbf{m}(\theta), \sigma^2 \mathbf{I}). \qquad (3.80)
$$

For predicting the performance of the MML estimator given above, we can use the MSE expression of Remark 1 by setting $\mathcal{L}(\mathbf{x}; \theta) \triangleq f(\mathbf{x}; \theta)$ and calculating the probability in the integrand of (3.12) with respect to the true measurement distribution

$\bar{f}(\cdot;\bar{\theta})$. The log-likelihood ratio $\ln \frac{f(\mathbf{x};\bar{\theta}+2\epsilon)}{f(\mathbf{x};\bar{\theta})}$ in this case is given as

$$\ln \frac{f(\mathbf{x};\bar{\theta}+2\epsilon)}{f(\mathbf{x};\bar{\theta})} = \frac{1}{\sigma^2}\Big(2\Re\big\{\tilde{\mathbf{m}}^{\mathrm{H}}(\bar{\theta}+2\epsilon;\bar{\theta})(\mathbf{x}-\bar{\mathbf{m}}(\bar{\theta}))\big\} - \|\tilde{\mathbf{m}}(\bar{\theta}+2\epsilon;\bar{\theta})\|^2 + 2\Re\big\{\tilde{\mathbf{m}}^{\mathrm{H}}(\bar{\theta}+2\epsilon;\bar{\theta})\boldsymbol{\mu}(\bar{\theta})\big\}\Big) \tag{3.81}$$

where the following definitions are used,

$$\tilde{\mathbf{m}}(\boldsymbol{\theta}_1;\boldsymbol{\theta}_2) \triangleq \mathbf{m}(\boldsymbol{\theta}_1) - \mathbf{m}(\boldsymbol{\theta}_2), \tag{3.82}$$

$$\boldsymbol{\mu}(\theta) \triangleq \bar{\mathbf{m}}(\theta) - \mathbf{m}(\theta). \tag{3.83}$$

We can now calculate the probability of the event $\ln(f(\mathbf{x};\bar{\theta}+2\epsilon)/f(\mathbf{x};\bar{\theta})) \geq 0$ with respect to the true measurement distribution $\bar{f}(\cdot;\bar{\theta})$ as

$$P\Big(\ln \frac{f(\mathbf{x};\bar{\theta}+2\epsilon)}{f(\mathbf{x};\bar{\theta})} \geq 0\Big) = \mathcal{N}_{\mathrm{ccdf}}\Big(\|\tilde{\mathbf{m}}(\cdot)\| - 2\Re\Big\{\frac{\tilde{\mathbf{m}}^{\mathrm{H}}(\cdot)}{\|\tilde{\mathbf{m}}(\cdot)\|}\boldsymbol{\mu}(\bar{\theta})\Big\}; 0, 2\bar{\sigma}^2\Big) \tag{3.84}$$

under the assumption that $\|\tilde{\mathbf{m}}(\cdot)\| \neq 0$, where we dropped the arguments of the function $\tilde{\mathbf{m}}(\bar{\theta}+2\epsilon;\bar{\theta})$ for brevity. Substituting this expression into the integrand of (3.12) we get the following predicted MSE for the MML estimate.

$$\widehat{\mathrm{MSE}}_{\mathrm{MML}}(\bar{\theta}) = 2\int_{-\infty}^{\infty} |\epsilon| \mathcal{N}_{\mathrm{ccdf}}\Big(\|\tilde{\mathbf{m}}(\cdot)\| - 2\Re\Big\{\frac{\tilde{\mathbf{m}}^{\mathrm{H}}(\cdot)}{\|\tilde{\mathbf{m}}(\cdot)\|}\boldsymbol{\mu}(\bar{\theta})\Big\}; 0, 2\bar{\sigma}^2\Big)\, \mathrm{d}\epsilon. \tag{3.85}$$

3.6 Numerical Results

In this section, we examine the performance of the proposed MSE expression on four different direction of arrival (DOA) estimation problems. The first two problems study the conventional and misspecified ML estimation respectively. In the third one, we investigate the performance of an IDE whose objective function is not the likelihood function, but a function derived from the manifold characteristics. The fourth problem investigates Bayesian DOA estimation. The implementation details of the numerical experiments are given in Appendix F.

3.6.1 DOA Estimation (No Model Mismatch)

Consider the DOA estimation problem with an N-element sensor array with the following array manifold.

$$\mathbf{a}_{\boldsymbol{\psi}} = [a_1, a_2, \ldots, a_N]^{\mathrm{T}}, \quad a_n = \exp\left(j\frac{2\pi}{\lambda}\mathbf{p}_n^{\mathrm{T}}\mathbf{u}_{\boldsymbol{\psi}}\right), \tag{3.86}$$

$$\mathbf{u}_{\boldsymbol{\psi}} = \begin{bmatrix} \cos(\phi)\sin(\theta) \\ \sin(\phi)\sin(\theta) \\ \cos(\theta) \end{bmatrix}, \quad \mathbf{p}_n = \begin{bmatrix} p_n^{\mathsf{x}} \\ p_n^{\mathsf{y}} \\ p_n^{\mathsf{z}} \end{bmatrix}, \tag{3.87}$$

where, $\boldsymbol{\psi} \triangleq [\phi,\ \theta]^{\mathrm{T}}$ denotes the unknown DOA vector composed of azimuth $\phi \in [0,\ 2\pi)$ rads (measured from the x-axis in counter-clockwise direction) and elevation $\theta \in [0,\ \pi)$ rads (measured from the z-axis). $\mathbf{p}_n$ is the position vector of the nth sensor containing the x, y and z-coordinates; $N = 11$ is the number of sensors; λ denotes the wavelength. The sensor array is illustrated in Figure 3.3a. The sensor positions for this configuration have the following analytical form

$$p_n^x = \frac{1}{4}\cos\left(n\frac{\pi}{4}\right), \quad p_n^y = \frac{1}{4}\sin\left(n\frac{\pi}{4}\right), \quad p_n^z = \frac{1}{4}\sin\left(n\frac{\pi}{4}\right). \tag{3.88}$$

The sensor measurement vector $\mathbf{x} \in \mathbb{C}^N$ under additive noise is modeled as

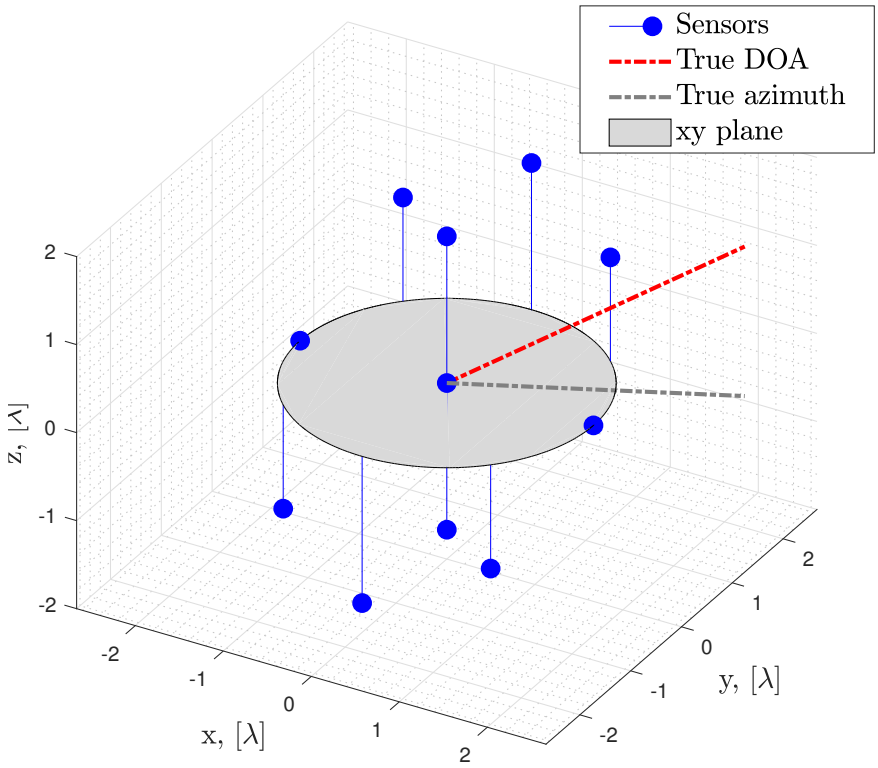
$$\mathbf{x} = \beta\mathbf{a}_{\bar{\boldsymbol{\psi}}} + \mathbf{v}, \tag{3.89}$$

where $\mathbf{v} \sim \mathcal{CN}(\mathbf{v}; \mathbf{0}, \sigma^2\mathbf{I}_N)$; $\beta \in \mathbb{R}$ (β is taken as a real-valued scalar with no loss of generality due to the circular symmetry of the complex Gaussian noise) and $\bar{\boldsymbol{\psi}} \triangleq [\bar{\phi},\ \bar{\theta}]^{\mathrm{T}}$ denotes the true value of the angle vector $\boldsymbol{\psi}$. The true target angular positions are $\bar{\phi} = 25°$ and $\bar{\theta} = 60°$. The beampattern of the array, obtained using the conventional, i.e., Bartlett, beamformer with coefficients steered to the true DOA, for this angular position is shown in Figure 3.3b. The beampattern contains sidelobes as high as -2 dB, with a response normalized to 0 dB at the true DOA. Consequently, the array is prone to gross errors. With these definitions, the ML estimator involves the following optimization problem:

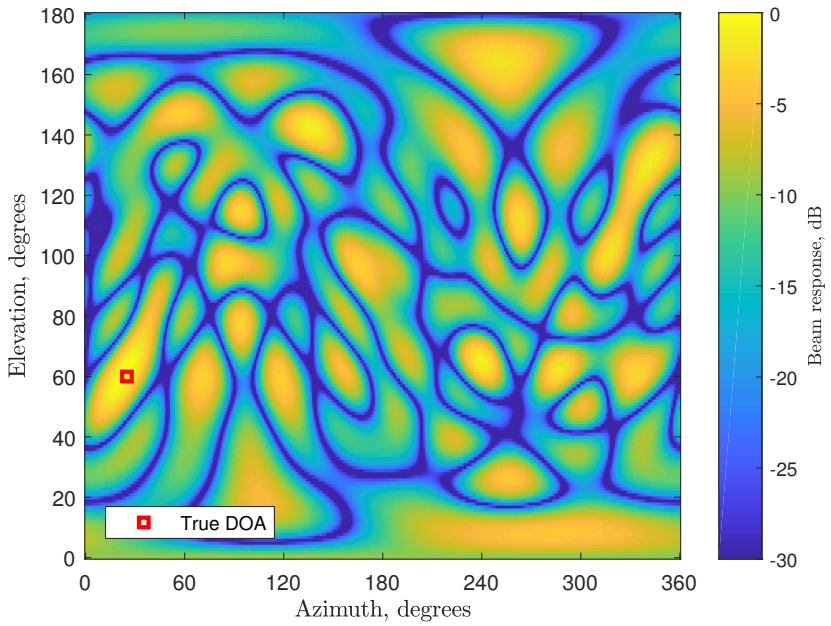
$$\hat{\boldsymbol{\psi}} = \arg\max_{\boldsymbol{\psi}} \Re\{\mathbf{x}^{\mathrm{H}}\mathbf{a}_{\boldsymbol{\psi}}\}. \tag{3.90}$$

Since the signal model is a parametric mean model with the mean function $\mathbf{m}(\boldsymbol{\psi}) = \beta\boldsymbol{a}_{\boldsymbol{\psi}}$, the finite support versions of the expressions (3.64) and (3.76) can be utilized

for MSE prediction. We consider three different cases: (i) Azimuth ϕ is unknown, but elevation $\theta = \bar{\theta}$ is known; (ii) Elevation θ is unknown, but azimuth $\phi = \bar{\phi}$ is known; (iii) Both azimuth ϕ and elevation θ are unknown. The results of 10^5 Monte Carlo simulations are given in Figure 3.4 and Figure 3.5 for azimuth and elevation estimation problems respectively. Without and with nuisance parameter cases are given in both figures. For comparison purposes the corresponding CRLBs (see [46] for the analytical expressions), Barankin bounds (BBs) [24] with single test point optimized over a grid, Fessler's method [1], So et al.'s method [2], and method of interval errors (MIE) [34] are also illustrated. As seen in these figures, the proposed method is able to predict the threshold SNR below which the ML estimator starts following the CRLB and tracks the CRLB in the asymptotic region as expected. BB, on the other hand, converges to CRLB at a much smaller SNR value than the ML estimator. MIE closely follows the ML estimator in the threshold region. This is essentially due to the problem specific selection of the intervals and accurate gross error probability calculation. Note that MIE does not have any assumptions on the objective function, such as symmetry or unimodality, leading to a better tracking of ML estimator performance especially in the threshold region. Taylor expansion based methods of Fessler and So et al. follow the CRLB values in all regions of operation and they are unable to take into account the gross errors the ML estimator makes below the threshold SNR.

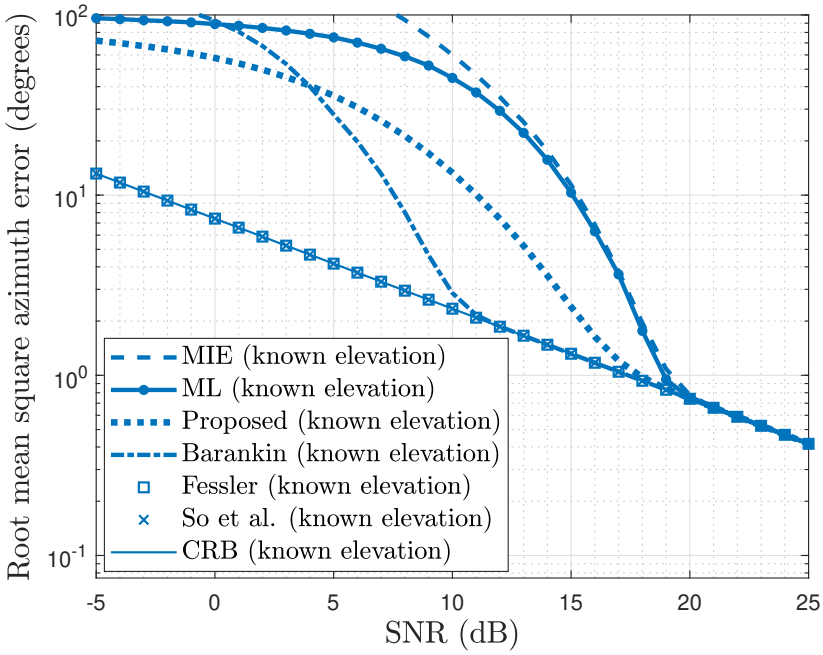


(a) Array configuration.

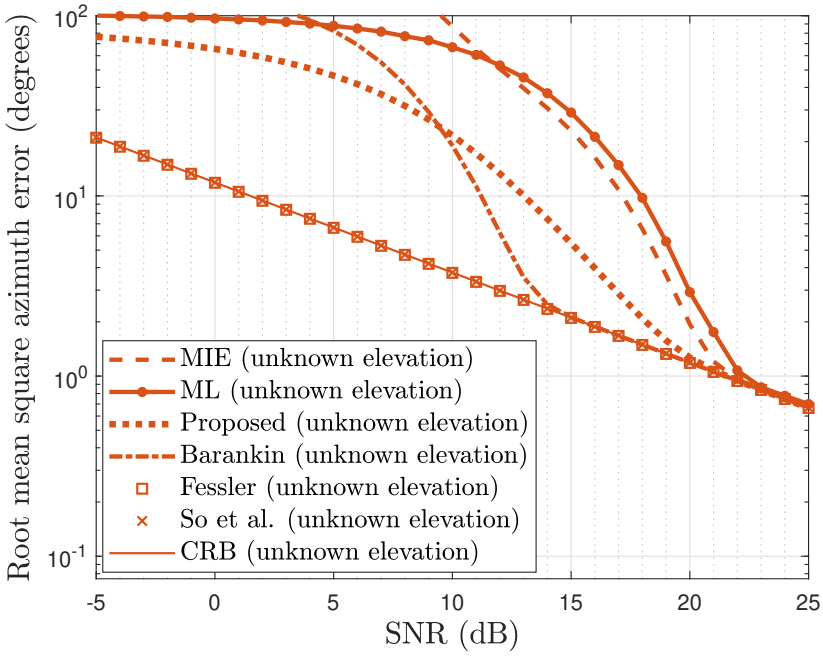


(b) Array beampattern at true DOA, $\phi = 25°$, $\theta = 60°$.

Figure 3.3: Array configuration and array beampattern at true DOA.

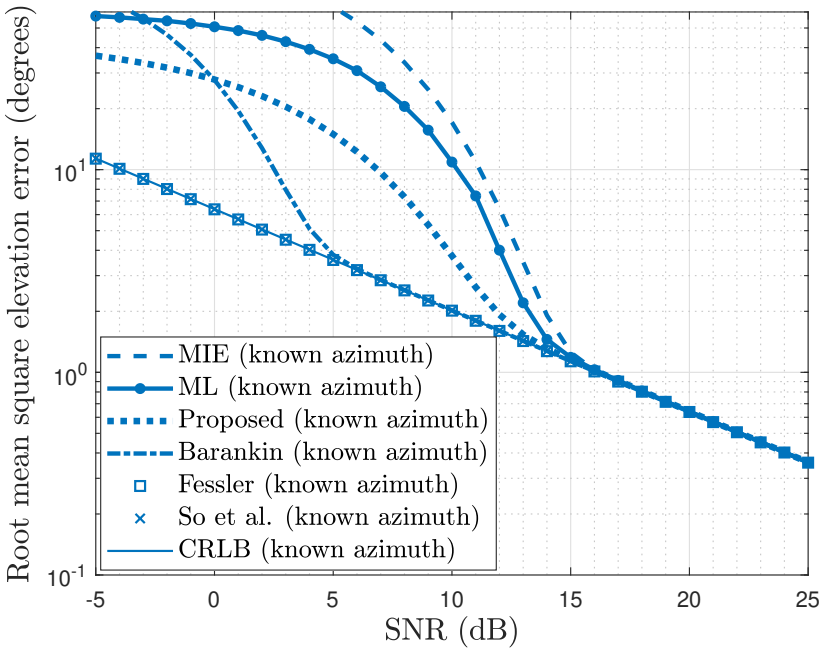


(a) Known elevation

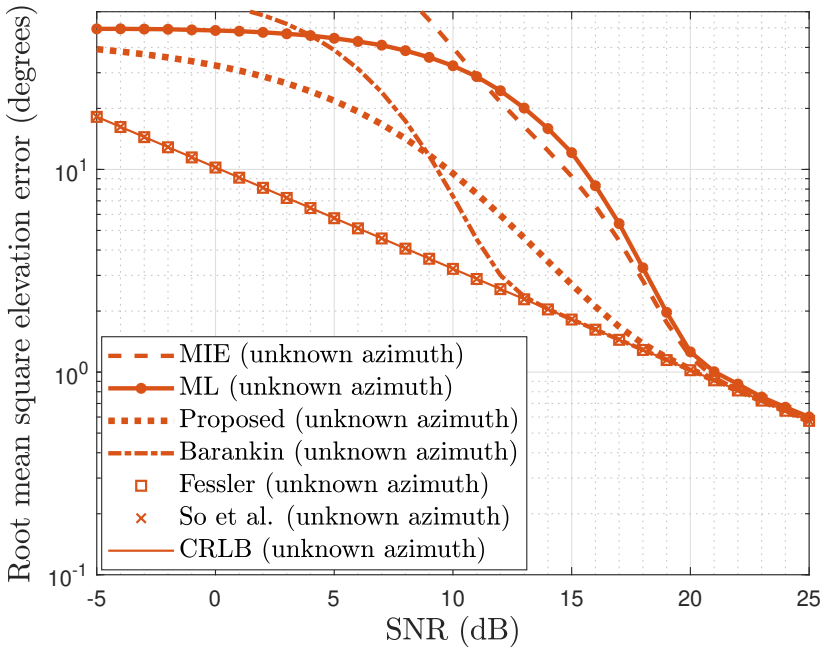


(b) Unknown elevation

Figure 3.4: Azimuth estimation performance curves for the array configuration in Figure 3.3a, without and with nuisance parameter (azimuth angle). Note that the y-axis limits are the same for both figures. True DOA: $\bar{\phi} = 25°$, $\bar{\theta} = 60°$.



(a) Known azimuth



(b) Unknown azimuth

Figure 3.5: Elevation estimation performance curves for the array configuration in Figure 3.3a, without and with nuisance parameter (elevation angle). Note that the y-axis limits are the same for both figures. True DOA: $\bar{\phi} = 25°$, $\bar{\theta} = 60°$.

3.6.2 DOA Estimation (Model Mismatch)

In this section we consider the misspecified ML estimation problem examined in Section 3.5.3 on the parameterized mean model. For this purpose, we consider the near field azimuth estimation problem with known elevation angle, in which the estimator uses the plane wave propagation assumption (far field assumption) rather than the true propagation model which is the spherical spreading.

A uniform circular array of radius $5\lambda/3$ with 12 elements is used. The signal of interest emanates from a target at a range of 5λ, which is closer than the far-field limit $2(10\lambda/3)^2/\lambda = 200\lambda/9$ [47]. The array configuration and the target position are illustrated in Figure 3.6. The true signal model is given as $\bar{\mathbf{m}}(\bar{\phi}) =$

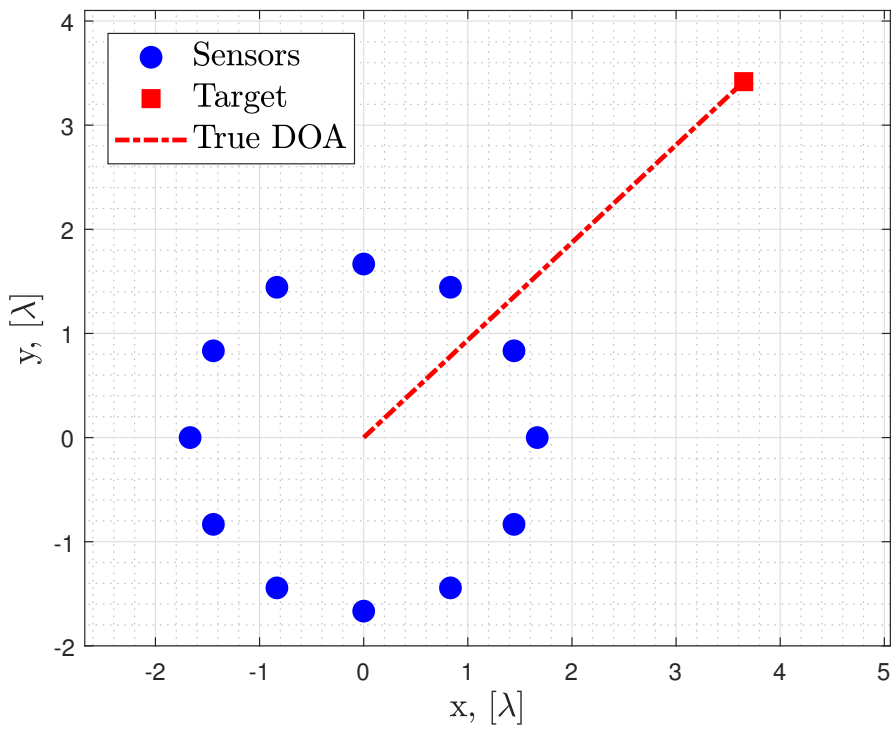


Figure 3.6: 12 element uniform circular array with a radius of $\frac{5}{3}\lambda$ and target of interest at 5λ range.

$[\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_N]^{\mathrm{T}}, \bar{a}_n = \exp\big(-j\frac{2\pi}{\lambda}d_n(\bar{\phi})\big)$, where $d_n(\phi) = \|\mathbf{p}_n - r\mathbf{u}_\phi\|$, $\mathbf{u}_\phi = [\cos(\phi), \sin(\phi)]^{\mathrm{T}}$, $\mathbf{p}_n = [p_n^{\mathrm{x}}, p_n^{\mathrm{y}}]^{\mathrm{T}}$ and r is the range of the target from the array center as illustrated in Figure 3.6. The assumed model by the estimator is the plane wave model, given as $\mathbf{m}(\phi) = [a_1, a_2, \ldots, a_N]^{\mathrm{T}}$, $a_n = \exp\big(j\frac{2\pi}{\lambda}\mathbf{p}_n^{\mathrm{T}}\mathbf{u}_\phi\big)$. There is no

misspecification in the noise variance, i.e., $\sigma^2 = \bar{\sigma}^2$. The MSE values for this experiment with 10,000 Monte Carlo runs are given in Figure 3.7 along with the corresponding MCRLBs [9, 10], BBs [24], and the results for Fessler's [1] and So et al.'s [2] methods. Note that the Matlab codes for these simulations are available at [48]. MCRLB reduces to the following expression for this specific problem.

$$\mathrm{MCRLB}(\bar{\phi}) = \mathbf{C}_D^{-1}(\bar{\phi})\mathcal{I}_D(\bar{\phi})\mathbf{C}_D^{-1}(\bar{\phi}), \tag{3.91}$$

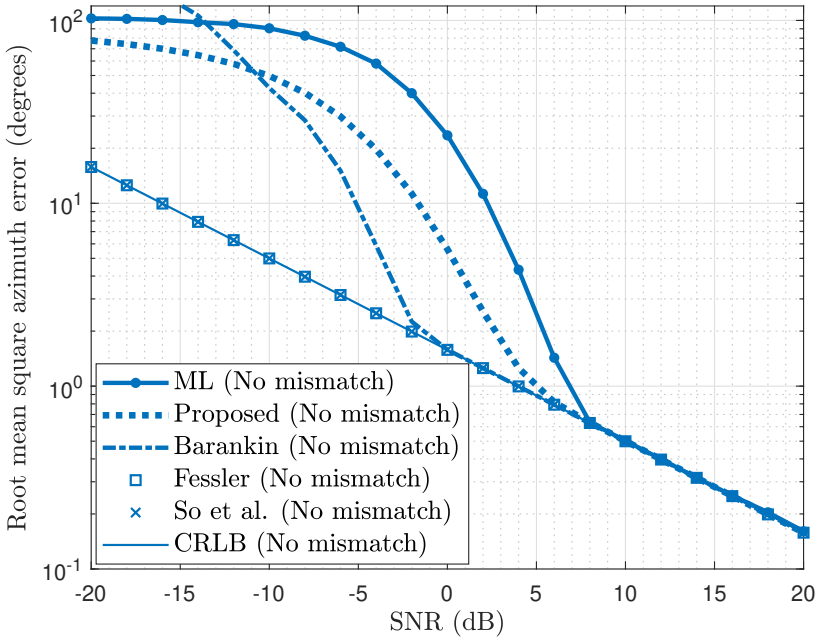
where

$$\mathcal{I}_D(\phi) = \frac{2}{\sigma^2}\left\|\frac{\partial \mathbf{m}(\phi)}{\partial \phi}\right\|^2, \quad \mathbf{C}_D(\theta) = -\mathcal{I}_D(\phi) + 2\Re\left\{\left[\frac{\partial^2 \mathbf{m}(\phi)}{\partial \theta^2}\right]^{\mathrm{H}} \boldsymbol{\mu}(\phi)\right\}, \tag{3.92}$$

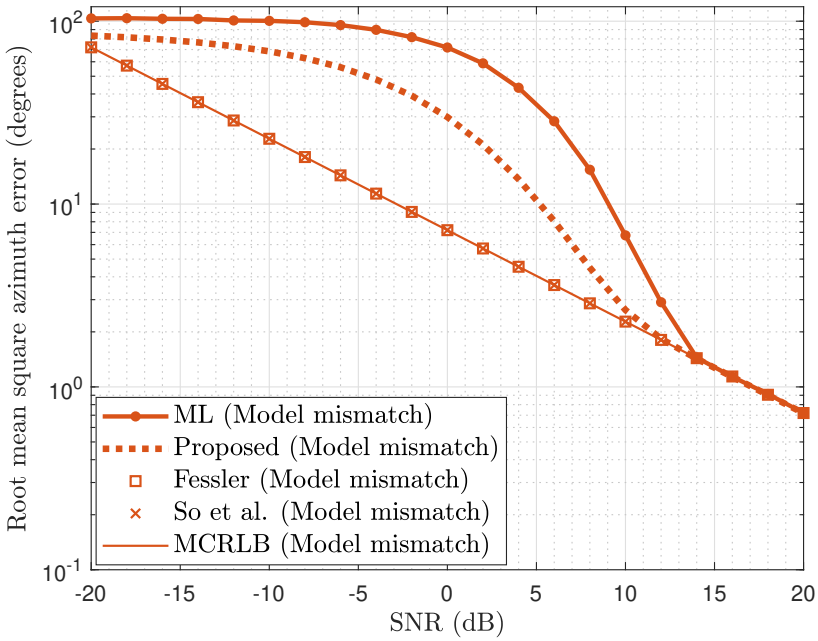
and $\boldsymbol{\mu}(\cdot)$ was defined in (3.83). BB [24] with a single test point optimized over a grid, which is also the HCRB [22, 25] can be expressed as follows.

$$\mathrm{BB}(\bar{\phi}) = \mathrm{HCRB}(\bar{\phi}) = \max_{\phi} \frac{(\phi - \bar{\phi})^2}{e^{\frac{2}{\sigma^2}\|\mathbf{m}(\phi) - \mathbf{m}(\bar{\phi})\|^2} - 1}. \tag{3.93}$$

The results given in Figure 3.7 indicate that the proposed MSE expression again predicts the threshold SNR quite closely and tracks MCRLB in the small error region. On the other hand BB is optimistic about the threshold SNR and both Fessler's and So et al.'s methods yield the same results as CRLB and MCRLB for the no model mismatch and model mismatch cases respectively.



(a) No model mismatch, both models use plane wave propagation



(b) Model mismatch, assumed model uses plane wave propagation, while the actual model is spherical spreading

Figure 3.7: Far-field performance (a) and near-field performance (b) of a 12 element uniform circular array (Figure 3.6), for a target at 5λ distance. Note that the y-axis limits are the same for both figures.

3.6.3 Bayesian DOA Estimation

In this section we study the DOA estimation problem in a Bayesian framework. We consider a uniform linear array composed of $N = 15$ sensors with $\lambda/2$ element spacing. The signal model is as follows

$$x_n = \underbrace{\alpha e^{j\pi\cos(\phi)n}}_{m_n(\phi)} + w_n, \quad n = 0, 1, \ldots, N-1, \tag{3.94}$$

where $\alpha \in \mathbb{C}$ is the unknown complex amplitude, $w_n \sim \mathcal{CN}(w_n, 0, \sigma_w^2)$, and ϕ is the unknown azimuth angle to be estimated. The unknown angle ϕ has now a prior density $f(\phi)$, which is given as the symmetric beta distribution

$$\begin{aligned} \beta(a,b) &\triangleq \int_0^1 \phi^{a-1}(1-\phi)^{b-1}\mathrm{d}\phi, \quad 0 \le \phi \le \pi \\ f(\phi) &= \frac{1}{\pi\beta(a,a)}\left(\frac{\phi}{\pi}\right)^{a-1}\left(\frac{\pi-\phi}{\pi}\right)^{a-1}, \end{aligned} \tag{3.95}$$

with $a = 10$, and the performance of the ML and MAP estimators is examined. The proposed Bayesian MSE expression for the ML estimator can be expressed as $\widehat{\mathrm{MSE}}_{\mathrm{ML}} = \int_0^\pi f(\phi)\widehat{\mathrm{MSE}}_{\mathrm{ML}}(\phi)\mathrm{d}\phi$ where

$$\widehat{\mathrm{MSE}}_{\mathrm{ML}}(\phi) = 2\int_{\frac{-\phi}{2}}^{\frac{\pi-\phi}{2}} |\epsilon|\mathcal{N}_{\mathrm{ccdf}}\left(\|\tilde{\mathbf{m}}(\phi+2\epsilon;\phi)\|; 0, 2\sigma_w^2\right)\mathrm{d}\epsilon, \tag{3.96}$$

where the integration limits are selected as in (3.13) with $\phi_{\min} \triangleq 0$ and $\phi_{\max} \triangleq \pi$. The proposed MSE expression for the MAP estimator can be expressed as $\widehat{\mathrm{MSE}}_{\mathrm{MAP}} = \int_0^\pi f(\phi)\widehat{\mathrm{MSE}}_{\mathrm{MAP}}(\phi)\mathrm{d}\phi$ where

$$\begin{aligned} \mu_I(\phi,\epsilon,\sigma_w^2) &\triangleq \|\tilde{\mathbf{m}}(\phi+2\epsilon;\phi)\| + \frac{\sigma_w^2}{\|\tilde{\mathbf{m}}(\phi+2\epsilon;\phi)\|}\log\left(\frac{f(\phi)}{f(\phi+2\epsilon)}\right) \\ \widehat{\mathrm{MSE}}_{\mathrm{MAP}}(\phi) &= 2\int_{-\pi}^{\pi} |\epsilon|\mathcal{N}_{\mathrm{ccdf}}\left(\mu_I(\phi,\epsilon,\sigma_w^2); 0, 2\sigma_w^2\right)\mathrm{d}\epsilon. \end{aligned} \tag{3.97}$$

Note that $\widehat{\mathrm{MSE}}_{\mathrm{MAP}}(\phi)$ in (3.97) reduces to $\widehat{\mathrm{MSE}}_{\mathrm{ML}}(\phi)$ in (3.96) when the prior is flat. BCRLB for this Bayesian estimation problem is given as [3]

$$\begin{aligned} \mathrm{BCRLB} = &\Bigg(\pi^2\mathrm{SNR}\frac{N(N-1)(2N-1)}{3} \\ &\times \int_0^\pi \sin^2\phi f(\phi)\,\mathrm{d}\phi + \frac{4(a-1)(2a-1)}{\pi^2(a-2)}\Bigg)^{-1}, \end{aligned} \tag{3.98}$$

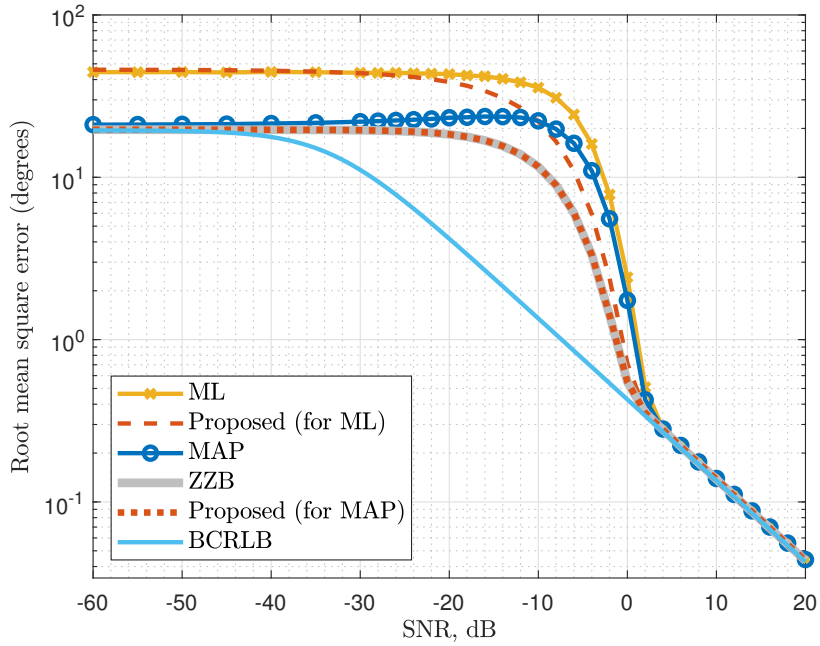


Figure 3.8: Bayesian DOA estimation performance of MAP and ML estimators along with the values of BCRLB, ZZB and the proposed MSE prediction expressions (for ML and MAP).

where $\mathrm{SNR} \triangleq \frac{|\alpha|^2}{\sigma_w^2}$ and $N = 15$. ZZB (without the valley filling function) for the problem can be expressed as

$$\mathrm{ZZB} = \frac{1}{2}\int_0^{\pi}\int_0^{\pi} h(f(\phi) + f(\phi + h))P_{\min}^{e}(\phi, \phi + h)\mathrm{d}\phi\mathrm{d}h, \tag{3.99}$$

where the minimum error probability $P_{\min}^{e}(\phi_1, \phi_2)$ can be calculated as

$$\begin{aligned} P_{\min}^{e}(\phi_1, \phi_2) = \pi_1 \mathcal{N}_{\mathrm{ccdf}}&\left(\|\tilde{\mathbf{m}}(\phi_2; \phi_1)\| + \frac{\sigma_w^2}{\|\tilde{\mathbf{m}}(\phi_2; \phi_1)\|}\log\frac{\pi_1}{\pi_2}; 0, 2\sigma_w^2\right) \\ &+ \pi_2 \mathcal{N}_{\mathrm{ccdf}}\left(\|\tilde{\mathbf{m}}(\phi_1; \phi_2)\| + \frac{\sigma_w^2}{\|\tilde{\mathbf{m}}(\phi_1; \phi_2)\|}\log\frac{\pi_2}{\pi_1}; 0, 2\sigma_w^2\right), \end{aligned} \tag{3.100}$$

with the prior probabilities $\pi_1 \triangleq \frac{f(\phi_1)}{f(\phi_1)+f(\phi_2)}$ and $\pi_2 \triangleq 1 - \pi_1$. Figure 3.8 shows the RMSE performances of the MAP and ML estimators over 10,000 Monte Carlo runs for each SNR value along with the values of BCRLB, ZZB and the proposed MSE prediction expressions $\widehat{\mathrm{MSE}}_{\mathrm{ML}}$ and $\widehat{\mathrm{MSE}}_{\mathrm{MAP}}$. The values of ZZB and the proposed MSE prediction expression $\widehat{\mathrm{MSE}}_{\mathrm{MAP}}$ are identical, as expected from the results of Section 3.3.3.

3.7 Approximate bias prediction

In this section we present the natural extension of the proposed MSE prediction method in Section 3.1 to the bias prediction. We repeat the Monte Carlo experiments to test the bias prediction performance of the proposed method. Bounds on bias performance of estimators received limited attention in literature over the years, [39, 49]. On the other hand, approximate bias prediction methods [1, 2] are Taylor expansion based methods and suffer from the threshold effect problem, explained in detail in the preceding sections. We begin by proposing the following theorem for bias prediction of implicitly defined estimators of the form $\hat{\theta} = \arg\max_{\theta} \mathcal{L}(\mathbf{x};\theta)$ as follows;

Theorem 2 (Bias of IDE) *Let the objective function $\mathcal{L}(\mathbf{x},\cdot)$ satisfy the following conditions.*

1. *$\mathcal{L}(\mathbf{x};\hat{\theta}+h) = \mathcal{L}(\mathbf{x};\hat{\theta}-h)$ for all $h \in \mathbb{R}$, i.e., the objective function is symmetric around its peak.*
2. *$\mathcal{L}(\mathbf{x};\theta)$ is strictly-increasing (strictly-decreasing) for $\theta < \hat{\theta}$ ($\theta > \hat{\theta}$).*

Then the bias of the IDE $\hat{\theta}$ is given as

$$\boxed{\widehat{\mathrm{B}}(\theta) = \int_{-\infty}^{\infty} \mathrm{sgn}(\epsilon) P\left(\mathcal{L}(\mathbf{x};\theta+2\epsilon) \geq \mathcal{L}(\mathbf{x};\theta)\right) \mathrm{d}\epsilon}. \tag{3.101}$$

Proof: The proof follows a simple analytical derivation, up to a point where we use the equivalence of the events $(\hat{\theta}-\theta) \geq \epsilon$ and $(\hat{\theta}-\theta) \leq -\epsilon$ to the events $\mathcal{L}(\mathbf{x};\theta+2\epsilon) \geq \mathcal{L}(\mathbf{x};\theta)$ and $\mathcal{L}(\mathbf{x};\theta-2\epsilon) \leq \mathcal{L}(\mathbf{x};\theta)$, respectively, when the assumptions of the theorem hold. This equivalence is proven in Section 3.2 for Theorem 1, and will not be repeated here. The proof starts from the definition of bias,

$$\begin{aligned}
\widehat{\mathrm{B}}(\theta) = \mathbb{E}\{\hat{\theta}-\theta;\theta\} &\triangleq \mathbb{E}\{e;\bar{\theta}\} \\
&= \int_{-\infty}^{\infty} e f_e(e;\theta)\mathrm{d}e \\
&= \int_{-\infty}^{0} e f_e(e;\theta)\mathrm{d}e + \int_{0}^{\infty} e f_e(e;\theta)\mathrm{d}e.
\end{aligned} \tag{3.102}$$

Using integration by parts ($d\nu = f_e(e;\theta)\mathrm{d}e$, $\nu = \mathrm{F}(e;\theta)$), we have

$$\begin{aligned}\int_{-\infty}^{0} e f_e(e;\theta)\mathrm{d}e &= e\mathrm{F}(e;\theta)|_{-\infty}^{0} - \int_{-\infty}^{0} \mathrm{F}(e;\theta)\mathrm{d}e \\ &= -\int_{-\infty}^{0} P(\hat{\theta} - \theta \leq \epsilon)\mathrm{d}\epsilon. \end{aligned} \tag{3.103}$$

Making a change of variable $\epsilon \leftarrow -\epsilon$ we get

$$\int_{-\infty}^{0} e f_e(e;\theta)\mathrm{d}\epsilon = \int_{\infty}^{0} P(\hat{\theta} - \theta \leq -\epsilon)\mathrm{d}\epsilon. \tag{3.104}$$

Similarly for the second term in (3.102) we have,

$$\begin{aligned}\int_{0}^{\infty} e f_e(e;\theta)\mathrm{d}e &= e\mathrm{F}_{\mathrm{c}}(e;\theta)|_{0}^{\infty} - \int_{0}^{\infty} \mathrm{F}_{\mathrm{c}}(e;\theta)\mathrm{d}e \\ &= \int_{0}^{\infty} P(\hat{\theta} - \theta \leq \epsilon)\mathrm{d}\epsilon. \end{aligned} \tag{3.105}$$

Note that, $f_e(\cdot;\cdot)$ is the probability density function, $\mathrm{F}(\cdot;\cdot)$ is the cumulative distribution function and $\mathrm{F}_{\mathrm{c}}(\cdot;\cdot)$ is the complementary cumulative distribution function. Using (3.104) and (3.105) in (3.102) we have,

$$\begin{aligned}\widehat{\mathrm{B}}(\theta) &= \int_{\infty}^{0} P(\hat{\theta} - \theta \leq -\epsilon)\mathrm{d}\epsilon + \int_{0}^{\infty} P(\hat{\theta} - \theta \leq \epsilon)\mathrm{d}\epsilon \\ &= \int_{\infty}^{0} P\left(\mathcal{L}(\mathbf{x};\theta - 2\epsilon) \geq \mathcal{L}(\mathbf{x};\theta)\right)\mathrm{d}\epsilon \\ &\quad + \int_{0}^{\infty} P\left(\mathcal{L}(\mathbf{x};\theta + 2\epsilon) \geq \mathcal{L}(\mathbf{x};\theta)\right)\mathrm{d}\epsilon \\ &= -\int_{-\infty}^{0} P\left(\mathcal{L}(\mathbf{x};\theta + 2\epsilon) \geq \mathcal{L}(\mathbf{x};\theta)\right)\mathrm{d}\epsilon \\ &\quad + \int_{0}^{\infty} P\left(\mathcal{L}(\mathbf{x};\theta + 2\epsilon) \geq \mathcal{L}(\mathbf{x};\theta)\right)\mathrm{d}\epsilon. \end{aligned} \tag{3.106}$$

Note that the equivalence of the events $(\hat{\theta} - \theta) \geq \epsilon$ and $(\hat{\theta} - \theta) \leq -\epsilon$ to the events $\mathcal{L}(\mathbf{x}, \theta + 2\epsilon) \geq \mathcal{L}(\mathbf{x}, \theta)$ and $\mathcal{L}(\mathbf{x}, \theta - 2\epsilon) \leq \mathcal{L}(\mathbf{x}, \theta)$, respectively was used in the last expressions. We can further simplify the last equation to the following general form.

$$\widehat{\mathrm{B}}(\theta) = \int_{-\infty}^{\infty} \mathrm{sgn}(\epsilon) P\left(\mathcal{L}(\mathbf{x};\theta + 2\epsilon) \geq \mathcal{L}(\mathbf{x};\theta)\right)\mathrm{d}\epsilon. \tag{3.107}$$

■

The following corollary applies the result in Theorem 2 to ML estimation.

Corollary 2 (Bias of ML Estimator) *If the likelihood function $f(\mathbf{x};\cdot)$ satisfies the conditions in Theorem 2, then the bias of the ML estimate $\hat{\theta}$ is given as*

$$\widehat{\mathrm{B}}_{\mathrm{ML}}(\theta) = 2\int_{-\infty}^{\infty} \mathrm{sgn}(\epsilon) P\left(\frac{f(\mathbf{x};\theta+2\epsilon)}{f(\mathbf{x};\theta)} \geq 1\right) \mathrm{d}\epsilon. \tag{3.108}$$

Proof: The proof is trivially provided using the result in Section 3.2. In (3.18) it is shown that $P\left(\mathcal{L}(\mathbf{x};\bar{\theta}+2\epsilon) \geq \mathcal{L}(\mathbf{x};\bar{\theta})\right) = P\left(\frac{f(\mathbf{x};\bar{\theta}+2\epsilon)}{f(\mathbf{x};\bar{\theta})} \geq 1\right)$. Using this equality in (3.107) we can reach (3.108). ■

For the case of nuisance parameters, we can adapt the results in Section 3.4. For this case we have $J > 1$ unknown scalar parameters, i.e., $\boldsymbol{\theta} \in \mathbb{R}^J$, and we would like to estimate only one of them (θ_1) while keeping the others ($\boldsymbol{\theta}_{\backslash 1} \triangleq [\theta_2\ \theta_3\ \cdots\ \theta_J]^{\mathrm{T}}$) as unknown nuisance parameters. Using the same approach in Section 3.4, we can express bias of the estimate $\hat{\theta}_1$ for θ_1 as follows

$$\begin{aligned}\widehat{\mathrm{B}}(\bar{\theta}_1) &= 2\int_{-\infty}^{\infty} \mathrm{sgn}(\epsilon) P\left(\mathcal{L}_1(\mathbf{x};\bar{\theta}_1+2\epsilon) \geq \mathcal{L}_1(\mathbf{x};\bar{\theta}_1)\right) \mathrm{d}\epsilon, \\ &= 2\int_{-\infty}^{\infty} \mathrm{sgn}(\epsilon) P\Big(\max_{\boldsymbol{\theta}_{\backslash 1}} \mathcal{L}(\mathbf{x};\bar{\theta}_1+2\epsilon,\boldsymbol{\theta}_{\backslash 1}) \geq \max_{\boldsymbol{\theta}_{\backslash 1}} \mathcal{L}(\mathbf{x};\bar{\theta}_1,\boldsymbol{\theta}_{\backslash 1})\Big) \mathrm{d}\epsilon.\end{aligned} \tag{3.109}$$

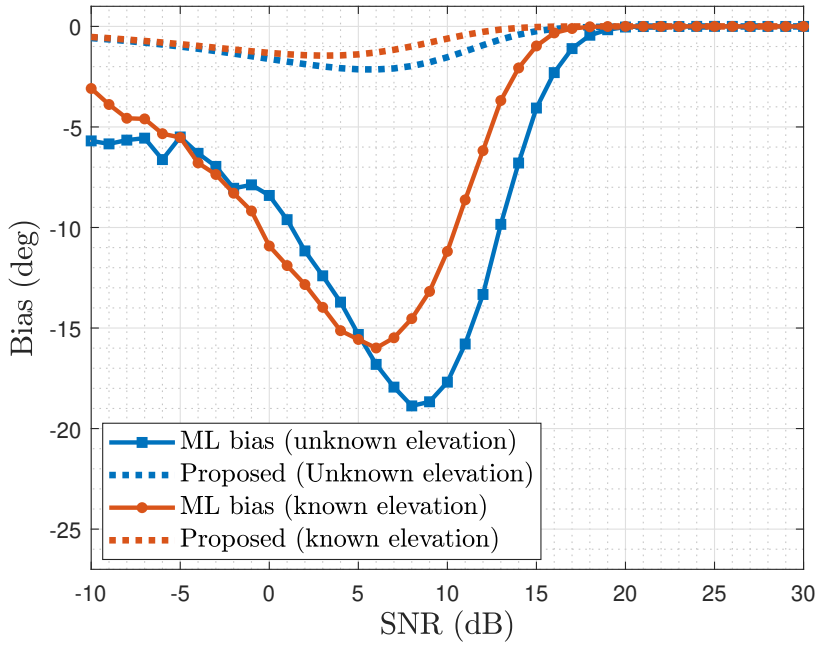
With the selection $\mathcal{L}(\mathbf{x};\boldsymbol{\theta}) \triangleq f(\mathbf{x};\boldsymbol{\theta}) \geq 0$, we can obtain the bias of the ML estimate $\hat{\theta}_1$ of θ_1 similarly to Corollary 2 from (3.109) as

$$\widehat{\mathrm{B}}_{\mathrm{ML}}(\bar{\theta}_1) = 2\int_{-\infty}^{\infty} \mathrm{sgn}(\epsilon) P\left(\frac{\max_{\boldsymbol{\theta}_{\backslash 1}} f(\mathbf{x};\bar{\theta}_1+2\epsilon,\boldsymbol{\theta}_{\backslash 1})}{\max_{\boldsymbol{\theta}_{\backslash 1}} f(\mathbf{x};\bar{\theta}_1,\boldsymbol{\theta}_{\backslash 1})} \geq 1\right) \mathrm{d}\epsilon. \tag{3.110}$$

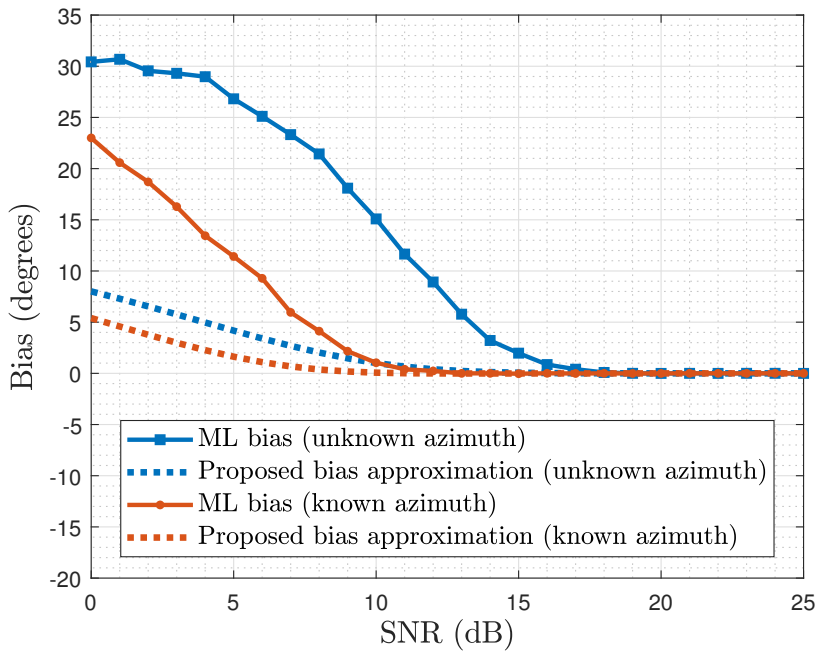
Note that the only difference in bias expressions is instead of $|\epsilon|$ we have $\mathrm{sgn}(\epsilon)$ in the integrand. Consequently, we can apply similar approximations for the calculation of the probability within the integrand in (3.110) as in Section 3.5.2.

3.8 Numerical Results

We consider the same numerical problems given in Section 3.6. Bias prediction results for azimuth and elevation estimation problems for ML estimation are given in Figure 3.9, model mismatch case is given in Figure 3.10. Although the bias prediction results do not seem to closely follow the actual bias values of ML estimator below the threshold SNR, the sign of the bias is predicted with a higher accuracy.



(a) Azimuth estimation at true DOA: $\bar{\phi} = 25°$, $\bar{\theta} = 60°$.



(b) Elevation estimation at true DOA: $\bar{\phi} = 25°$, $\bar{\theta} = 60°$.

Figure 3.9: Azimuth and elevation estimation bias curves for the array configuration in Figure 3.3a.

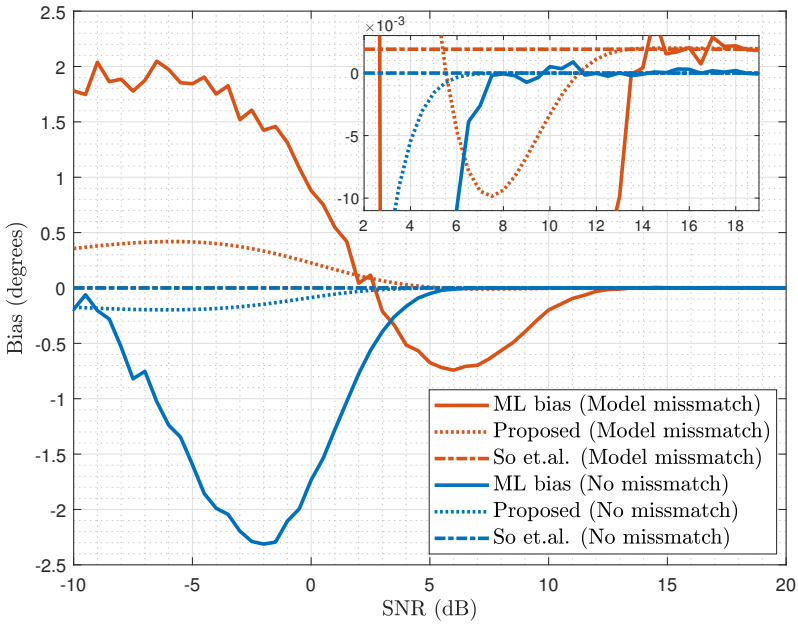


Figure 3.10: Near-field and far-field DOA estimation bias of a 12 element uniform circular array, for a target at 5λ distance.

CHAPTER 4

APPLICATIONS IN DOA ESTIMATION

In this chapter we present the applications of the proposed performance prediction method in direction of arrival (DOA) estimation.

4.1 Direction of Arrival Estimation under Rician Fading

In this section we examine the DOA problem under fading conditions. We start from the signal model and derive the analytical form of the Cramér-Rao Lower Bound first. Using Monte Carlo analysis we show that the CRLB is overly optimistic in characterizing ML performance even at high SNR values under severe fading conditions. We then derive the exact probability expressions for the proposed method in Chapter 3 for ML estimation and show that the proposed method is able to characterize the ML performance even under severe fading, using Monte Carlo analysis.

We consider the problem of estimating the parameter θ using the following measurement vector,

$$\mathbf{x} = \beta \mathbf{a}_\theta + \mathbf{v}, \tag{4.1}$$

where $\beta \sim \mathcal{CN}(\mu_F, \sigma_F^2)$ is the fading coefficient, $\mathbf{a}_\theta \in \mathbb{C}^N$ is the manifold vector and $\mathbf{v} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$ is the additive white Gaussian noise vector. This type of modeling induces randomness to the observed deterministic signal, and generally used to model the multi-path effects in the environment. Since the arrival times of different multi-path signals would be changing in time (hence their phase), the overall sum of the components would induce a variation in the signal of interest, in our case $\mathbf{a}_\theta$. Note that a large σ_F would indicate a large effect due to multipath signals. Also, a zero mean fading coefficient can be used to model the case where the direct path signal is

not observed. To represent the degree of variation in the direct path signal, the following parameter called Rician factor is defined; $\mathrm{K} \triangleq \frac{|\mu_F|^2}{\sigma_F^2}$. A large Rician factor K means that the direct path signal dominates the multi-path signals and a Rician factor of 0 means there is no direct path signal, only multi-path signals. Note that we have the following signal to noise ratio definition with the signal model in (4.1),

$$\mathrm{SNR} = \frac{\mathbb{E}\{|[\beta \mathbf{a}_\theta]_k|^2\}}{\mathbb{E}\{|[\mathbf{v}]_k|^2\}} = \frac{|\mu_F|^2 + \sigma_F^2}{\sigma^2} = \frac{\sigma_F^2}{\sigma^2}(\mathrm{K}+1) \tag{4.2}$$

Note that as $\sigma_F^2 \to 0$ (or $K \to \infty$), the SNR expression reduces to the expression for the non-fading case, $\mathrm{SNR} \to |\mu^2|/\sigma^2$.

4.1.1 Cramér-Rao Lower Bound calculation

Using the derivation in [7, Appendix 15C], Cramér-Rao Lower Bound for complex Gaussian case with a scalar parameter θ has the following form

$$\mathrm{FIM} = \mathrm{tr}\left\{\mathbf{C}_\mathbf{x}^{-1}\frac{\partial \mathbf{C}_\mathbf{x}}{\partial \theta}\mathbf{C}_\mathbf{x}^{-1}\frac{\partial \mathbf{C}_\mathbf{x}}{\partial \theta}\right\} + 2\Re\left\{\frac{\partial \boldsymbol{\mu}_\mathbf{x}^\mathrm{H}}{\partial \theta}\mathbf{C}_\mathbf{x}^{-1}\frac{\partial \boldsymbol{\mu}_\mathbf{x}}{\partial \theta}\right\}. \tag{4.3}$$

Where FIM is the Fisher Information Matrix and is equal to CRLB^{-1}. Note that $\mathbf{C}_\mathbf{x} = \mathbf{C}_\mathbf{x}(\theta)$ and $\boldsymbol{\mu}_\mathbf{x} = \boldsymbol{\mu}_\mathbf{x}(\theta)$, and we dropped the dependence on θ for brevity. The following identity will be used to simplify the first term in (4.3),

$$\begin{aligned}
&\mathbf{C}_\mathbf{x}\mathbf{C}_\mathbf{x}^{-1} = \mathbf{I}, \\
&\frac{\partial \mathbf{C}_\mathbf{x}}{\partial \theta}\mathbf{C}_\mathbf{x}^{-1} + \mathbf{C}_\mathbf{x}\frac{\partial(\mathbf{C}_\mathbf{x}^{-1})}{\partial \theta} = 0, \\
&\frac{\partial}{\partial \theta}\mathbf{C}_\mathbf{x}^{-1} = -\mathbf{C}_\mathbf{x}^{-1}\frac{\partial \mathbf{C}_\mathbf{x}}{\partial \theta}\mathbf{C}_\mathbf{x}^{-1},
\end{aligned} \tag{4.4}$$

which can be used to write (4.3) as

$$\mathrm{FIM} = \mathrm{tr}\left\{-\frac{\partial \mathbf{C}_\mathbf{x}}{\partial \theta}\frac{\partial(\mathbf{C}_\mathbf{x})^{-1}}{\partial \theta}\right\} + 2\Re\left\{\frac{\partial \mu_\mathbf{x}^\mathrm{H}}{\partial \theta}\mathbf{C}_\mathbf{x}^{-1}\frac{\partial \mu_\mathbf{x}}{\partial \theta}\right\}. \tag{4.5}$$

For the signal model in (4.1) we have the following expressions for $\boldsymbol{\mu}_\mathbf{x}$, $\mathbf{C}_\mathbf{x}$ and $\mathbf{C}_\mathbf{x}^{-1}$,

$$\boldsymbol{\mu}_\mathbf{x} = \mu_F \mathbf{a}_\theta, \tag{4.6}$$

$$\mathbf{C}_\mathbf{x} = \sigma^2\mathbf{I} + \sigma_F^2\mathbf{a}_\theta\mathbf{a}_\theta^\mathrm{H}, \tag{4.7}$$

$$\mathbf{C}_\mathbf{x}^{-1} = \frac{1}{\sigma^2}\mathbf{I} - \gamma\mathbf{a}_\theta\mathbf{a}_\theta^\mathrm{H}, \quad \gamma \triangleq \frac{\sigma_F^2}{\sigma^2(\sigma^2 + \sigma_F^2\underbrace{\|\mathbf{a}_\theta\|^2}_{N})}. \tag{4.8}$$

Note that the simple form for $\mathbf{C}_{\mathbf{x}}^{-1}$ is obtained bu using the spectral theorem to invert a matrix ([50, p.45]) for matrices of the form $\mathbf{I}+\alpha\mathbf{u}\mathbf{u}^{\mathbf{H}}$ (where $\alpha \in \mathbb{R}$ and $\mathbf{u} \in \mathbb{C}^N$). To calculate the first derivatives with respect to θ, the special form of the array manifold vector $\mathbf{a}_\theta$ will be of some use;

$$\begin{aligned}
[\mathbf{a}_\theta]_k &= \exp\left(j\frac{2\pi}{\lambda}\mathbf{p}_k^{\mathrm{T}}\mathbf{u}_\theta\right), \\
\left[\frac{\partial \mathbf{a}_\theta}{\partial \theta}\right]_k &= j\frac{2\pi}{\lambda}\mathbf{p}_k^{\mathrm{T}}\dot{\mathbf{u}}_\theta \exp\left(j\frac{2\pi}{\lambda}\mathbf{p}_k^{\mathrm{T}}\mathbf{u}_\theta\right), \\
\frac{\partial \mathbf{a}_\theta}{\partial \theta} &= \mathbf{D}\mathbf{a}_\theta, \quad \frac{\partial \mathbf{a}_\theta^{\mathrm{H}}}{\partial \theta} = \mathbf{a}_\theta^{\mathrm{H}}\mathbf{D}^*, \\
\mathbf{D} &\triangleq \mathrm{diag}\left(\frac{j2\pi\mathbf{p}_1^{\mathrm{T}}\dot{\mathbf{u}}_\theta}{\lambda}, \ldots, \frac{j2\pi\mathbf{p}_N^{\mathrm{T}}\dot{\mathbf{u}}_\theta}{\lambda}\right),
\end{aligned} \tag{4.9}$$

where $\mathbf{p}_k$ is the position vector of the kth sensor, and $\mathbf{u}_\theta = [\ \cos(\theta) \quad \sin(\theta)\ ]^{\mathrm{T}}$. Consequently, the first derivatives of for $\boldsymbol{\mu}_{\mathbf{x}}$, $\mathbf{C}_{\mathbf{x}}$ and $\mathbf{C}_{\mathbf{x}}^{-1}$ are as follows;

$$\frac{\partial \boldsymbol{\mu}_{\mathbf{x}}}{\partial \theta} = \mu_F \mathbf{D}\mathbf{a}_\theta, \tag{4.10}$$

$$\frac{\partial \mathbf{C}_{\mathbf{x}}}{\partial \theta} = \sigma_F^2\left(\mathbf{D}\mathbf{a}_\theta\mathbf{a}_\theta^{\mathrm{H}} + \mathbf{a}_\theta\mathbf{a}_\theta^{\mathrm{H}}\mathbf{D}^*\right), \tag{4.11}$$

$$\frac{\partial (\mathbf{C}_{\mathbf{x}}^{-1})}{\partial \theta} = -\gamma\left(\mathbf{D}\mathbf{a}_\theta\mathbf{a}_\theta^{\mathrm{H}} + \mathbf{a}_\theta\mathbf{a}_\theta^{\mathrm{H}}\mathbf{D}^*\right). \tag{4.12}$$

Using these expressions in (4.5), we have,

$$\begin{aligned}
\mathrm{FIM} =& \mathrm{tr}\{\gamma\sigma_F^2\left(\mathbf{D}\mathbf{a}_\theta\mathbf{a}_\theta^{\mathrm{H}} + \mathbf{a}_\theta\mathbf{a}_\theta^{\mathrm{H}}\mathbf{D}^*\right)^2\} \\
&+ \Re\left\{|\mu_F|^2\mathbf{a}_\theta^{\mathrm{H}}\mathbf{D}^*\left(\frac{1}{\sigma^2}\mathbf{I} - \gamma\mathbf{a}_\theta\mathbf{a}_\theta^{\mathrm{H}}\right)\mathbf{D}\mathbf{a}_\theta\right\} \\
=& \gamma\sigma_F^2\Big[\mathrm{tr}\{\mathbf{D}\mathbf{a}_\theta\mathbf{a}_\theta^{\mathrm{H}}\mathbf{D}\mathbf{a}_\theta\mathbf{a}_\theta^{\mathrm{H}}\} + \mathrm{tr}\{\mathbf{D}\mathbf{a}_\theta\mathbf{a}_\theta^{\mathrm{H}}\mathbf{a}_\theta\mathbf{a}_\theta^{\mathrm{H}}\mathbf{D}^*\} \\
&\qquad + \mathrm{tr}\{\mathbf{a}_\theta\mathbf{a}_\theta^{\mathrm{H}}\mathbf{D}^*\mathbf{D}\mathbf{a}_\theta\mathbf{a}_\theta^{\mathrm{H}}\} + \mathrm{tr}\{\mathbf{a}_\theta\mathbf{a}_\theta^{\mathrm{H}}\mathbf{D}^*\mathbf{a}_\theta\mathbf{a}_\theta^{\mathrm{H}}\mathbf{D}^*\}\Big] \\
&+ \frac{2|\mu_F|^2}{\sigma^2}\Re\{\mathbf{a}_\theta^{\mathrm{H}}\mathbf{D}^*\mathbf{D}\mathbf{a}_\theta\} - 2\gamma|\mu_F|^2\Re\{\mathbf{a}_\theta^{\mathrm{H}}\mathbf{D}^*\mathbf{a}_\theta\mathbf{a}_\theta^{\mathrm{H}}\mathbf{D}\mathbf{a}_\theta\}
\end{aligned} \tag{4.13}$$

Note that;

$$\mathbf{a}_\theta^{\mathrm{H}}\mathbf{D}\mathbf{a}_\theta = \sum_{n=1}^{N} j\frac{2\pi}{\lambda}\mathbf{p}_n^{\mathrm{T}}\dot{\mathbf{u}}_\theta.$$

Using this, and the identity $\mathrm{tr}\{\mathbf{AB}\} = \mathrm{tr}\{\mathbf{BA}\}$, the FIM expression can be put into the following form,

$$\begin{aligned}
\mathrm{CRLB}_{\text{fading}}^{-1} =& \gamma\sigma_F^2 \Bigg[\left(\sum_{n=1}^{N} \frac{j2\pi \mathbf{p}_n^{\mathrm{T}} \dot{\mathbf{u}}_\theta}{\lambda} \right)^2 + \sum_{n=1}^{N} \|\mathbf{a}_\theta\|^2 \left| \frac{2\pi \mathbf{p}_n^{\mathrm{T}} \dot{\mathbf{u}}_\theta}{\lambda} \right|^2 \\
&+ \sum_{n=1}^{N} \|\mathbf{a}_\theta\|^2 \left| \frac{2\pi \mathbf{p}_n^{\mathrm{T}} \dot{\mathbf{u}}_\theta}{\lambda} \right|^2 + \left(\sum_{n=1}^{N} \frac{-j2\pi \mathbf{p}_n^{\mathrm{T}} \dot{\mathbf{u}}_\theta}{\lambda} \right)^2 \Bigg] \\
&+ \frac{2|\mu_F|^2}{\sigma^2} \sum_{n=1}^{N} \|\mathbf{a}_\theta\|^2 \left| \frac{2\pi \mathbf{p}_n^{\mathrm{T}} \dot{\mathbf{u}}_\theta}{\lambda} \right|^2 - 2\gamma|\mu_F|^2 \left| \sum_{n=1}^{N} \frac{j2\pi \mathbf{p}_n^{\mathrm{T}} \dot{\mathbf{u}}_\theta}{\lambda} \right|^2 \\
=& \left(2\gamma\sigma_F^2 + \frac{2|\mu_F|^2}{\sigma^2} \right) N \sum_{n=1}^{N} \left| \frac{2\pi \mathbf{p}_n^{\mathrm{T}} \dot{\mathbf{u}}_\theta}{\lambda} \right|^2 \\
&- \left(2\gamma|\mu_F|^2 + 2\gamma\sigma_F^2 \right) \left| \sum_{n=1}^{N} \frac{2\pi \mathbf{p}_n^{\mathrm{T}} \dot{\mathbf{u}}_\theta}{\lambda} \right|^2 . \qquad (4.14)
\end{aligned}$$

We can write this general form for CRLB under Rician fading in terms of Rician factor K and signal to noise ratio SNR as follows,

$$\begin{aligned}
\mathrm{CRLB}_{\text{fading}}^{-1} =& 2\,\mathrm{SNR} \left(\frac{\mathrm{K}}{\mathrm{K}+1} + \frac{\mathrm{SNR}}{1 + \mathrm{SNR}\,\frac{N}{\mathrm{K}+1}} \right) N \sum_{n=1}^{N} \left| \frac{2\pi \mathbf{p}_n^{\mathrm{T}} \dot{\mathbf{u}}_\theta}{\lambda} \right|^2 \\
&- 2\,\mathrm{SNR}\, \frac{\frac{\mathrm{SNR}}{\mathrm{K}+1}}{1 + \mathrm{SNR}\,\frac{N}{\mathrm{K}+1}} \left| \sum_{n=1}^{N} \frac{2\pi \mathbf{p}_n^{\mathrm{T}} \dot{\mathbf{u}}_\theta}{\lambda} \right|^2 . \qquad (4.15)
\end{aligned}$$

When there is no fading, by simply setting $\sigma_F^2 = 0$ and $\gamma = 0$ in (4.14), we have

$$\mathrm{CRLB}_{\text{no-fading}}^{-1} = 2\frac{|\mu_F|^2}{\sigma^2} N \sum_{n=1}^{N} \left| \frac{2\pi \mathbf{p}_n^{\mathrm{T}} \dot{\mathbf{u}}_\theta}{\lambda} \right|^2 = 2\,\mathrm{SNR} \times N \sum_{n=1}^{N} \left| \frac{2\pi \mathbf{p}_n^{\mathrm{T}} \dot{\mathbf{u}}_\theta}{\lambda} \right|^2 . \quad (4.16)$$

For a uniform line array with $\lambda/2$ element spacing, we have $\mathbf{p}_k^{\mathrm{T}} \dot{\mathbf{u}}_\theta = -\frac{\lambda}{2} k \sin(\theta)$. And the corresponding CRLB value for Rician fading case is as follows;

$$
\begin{aligned}
\mathrm{CRLB}^{-1}_{\mathrm{ULA,fading}} =& 2\,\mathrm{SNR}\left(\frac{\mathrm{K}}{\mathrm{K}+1}+\frac{\mathrm{SNR}}{1+\mathrm{SNR}\,\frac{N}{\mathrm{K}+1}}\right) N\sum_{n=1}^{N}|\pi k\sin(\theta)|^2 \\
&- 2\,\mathrm{SNR}\,\frac{\frac{\mathrm{SNR}}{\mathrm{K}+1}}{1+\mathrm{SNR}\,\frac{N}{\mathrm{K}+1}}\left|\sum_{n=1}^{N}\pi k\sin(\theta)\right|^2 \\
=& 2\,\mathrm{SNR}\left(\frac{\mathrm{K}}{\mathrm{K}+1}+\frac{\mathrm{SNR}}{1+\mathrm{SNR}\,\frac{N}{\mathrm{K}+1}}\right) N\pi^2\sin^2(\theta)\sum_{n=1}^{N}k^2 \\
&- 2\,\mathrm{SNR}\,\frac{\frac{\mathrm{SNR}}{\mathrm{K}+1}}{1+\mathrm{SNR}\,\frac{N}{\mathrm{K}+1}}\pi^2\sin^2(\theta)\left|\sum_{n=1}^{N}k\right|^2 \\
=& 2\,\mathrm{SNR}\left(\frac{\mathrm{K}}{\mathrm{K}+1}+\frac{\mathrm{SNR}}{1+\mathrm{SNR}\,\frac{N}{\mathrm{K}+1}}\right)\pi^2\sin^2(\theta)\frac{N^2(N+1)(2N+1)}{6} \\
&- 2\,\mathrm{SNR}\,\frac{\frac{\mathrm{SNR}}{\mathrm{K}+1}}{1+\mathrm{SNR}\,\frac{N}{\mathrm{K}+1}}\pi^2\sin^2(\theta)\frac{N^2(N+1)^2}{4}. \qquad (4.17)
\end{aligned}
$$

Similarly, when fading is not present ($\sigma_F^2 = 0$), we have the CRB expression from (4.16) as follows:

$$
\mathrm{CRLB}^{-1}_{\mathrm{ULA,no-fading}} = 2\,\mathrm{SNR}\,\pi^2\sin^2(\theta)\frac{N^2(N+1)(2N+1)}{6}. \qquad (4.18)
$$

Although it is not readily apparent, the CRLB for Rician fading does not introduce a big change especially at high SNR values. At high SNR values the CRLB will be dominated by the $2\frac{|\mu_F|^2}{\sigma^2}$ term, which in turn is the CRLB value for non-fading case. It is harder to make the same deduction for the general case (non-ULA arrays). For this reason, we can test the CRLB expression with Monte Carlo simulations for different levels of fading. Figure 4.1 shows the performance curves for ML estimator for a 12-element uniform circular array with a radius of $5\lambda/3$. CRLB curves are also present in the same figure. Clearly the CRLB under Rician fading is overly optimistic. For low Rician factors the ML performance is not characterized by CRLB even at high SNR values. In the following section we derive the closed form expression of probabilities that are used in the proposed performance prediction method (Chapter 3) and show that the proposed method is able to characterize ML performance even for low Rician factor values.

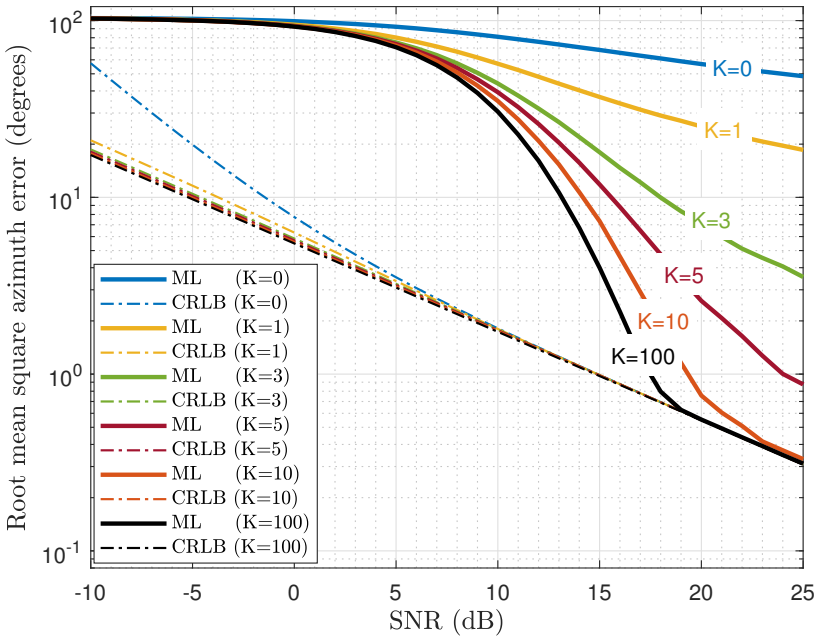


Figure 4.1: ML performance and CRLB curves for different Rician factors, for a 12-element uniform circular array with $5\lambda/3$ radius. Note that CRLB stays overly optimistic in characterizing the ML performance in for low ($\leq$ 10) Rician factors over a wide range of SNR values.

4.1.2 Approximate performance prediction for ML with parameterized mean model under Rician fading

We are interested in the mean square error (MSE) of the maximum likelihood (ML) estimator $\hat{\theta}$ of θ given as

$$\hat{\theta} \triangleq \arg\max_{\theta} f(\mathbf{x};\theta), \tag{4.19}$$

where the objective function is the likelihood function $f(\mathbf{x};\cdot)$ given as

$$\begin{aligned} f(\mathbf{x};\theta) &\triangleq \mathcal{CN}\left(\mathbf{x};\mathbf{m}_\theta,\mathbf{\Sigma}_\theta\right), \\ \mathbf{m}_\theta &= \mu_F \mathbf{a}_\theta, \\ \mathbf{\Sigma} &= \sigma^2\mathbf{I} + \sigma_F^2 \mathbf{a}_\theta \mathbf{a}_\theta^{\mathbf{H}}. \end{aligned} \tag{4.20}$$

Note that the ML estimator has the following form for this problem (see Appendix D),

$$\hat{\theta}_{\mathrm{ML}} = \arg\max_{\theta} 2\mu_F \Re\{\mathbf{a}_\theta^{\mathrm{H}}\mathbf{x}\} + \frac{\sigma_F^2}{\sigma_v^2}|\mathbf{a}_\theta^{\mathrm{H}}\mathbf{x}|^2. \tag{4.21}$$

In order to use the result in (3.17) for ML estimation under Rician fading, we need the likelihood ratio, or equivalently the log-likelihood ratio (LLR),

$$\begin{aligned}\Lambda =& -(\mathbf{x}-\mathbf{m}_{\theta+2\epsilon})^{\mathbf{H}}\mathbf{\Sigma}_{\theta+2\epsilon}^{-1}(\mathbf{x}-\mathbf{m}_{\theta+2\epsilon})+(\mathbf{x}-\mathbf{m}_{\theta})^{\mathbf{H}}\mathbf{\Sigma}_{\theta}^{-1}(\mathbf{x}-\mathbf{m}_{\theta})\\ =& -(\mathbf{x}-\mathbf{m}_{\theta+2\epsilon})^{\mathbf{H}}\left(\frac{1}{\sigma^2}\mathbf{I}-\frac{\sigma_F^2}{\sigma^2(\sigma^2+\sigma_F^2\|\mathbf{a}_{\theta+2\epsilon}\|^2)}\mathbf{a}_{\theta+2\epsilon}\mathbf{a}_{\theta+2\epsilon}^{\mathbf{H}}\right)(\mathbf{x}-\mathbf{m}_{\theta+2\epsilon})\\ &+(\mathbf{x}-\mathbf{m}_{\theta})^{\mathbf{H}}\left(\frac{1}{\sigma^2}\mathbf{I}-\frac{\sigma_F^2}{\sigma^2(\sigma^2+\sigma_F^2\|\mathbf{a}_{\theta}\|^2)}\mathbf{a}_{\theta}\mathbf{a}_{\theta}^{\mathbf{H}}\right)(\mathbf{x}-\mathbf{m}_{\theta}),\end{aligned} \tag{4.22}$$

where we used the identity in (4.8) for matrix inverses. Note that $\|\mathbf{a}_\theta\|^2=\|\mathbf{a}_{\theta+2\epsilon}\|^2=N$, where N is the number of sensors in the array. Let $\gamma\triangleq\frac{\sigma_F^2}{\sigma^2(\sigma^2+\sigma_F^2N)}$, then

$$\begin{aligned}\Lambda =& -\frac{1}{\sigma^2}\|\mathbf{x}-\mathbf{m}_{\theta+2\epsilon}\|^2+\gamma|\mathbf{a}_{\theta+2\epsilon}^{\mathbf{H}}(\mathbf{x}-\mathbf{m}_{\theta+2\epsilon})|^2\\ &+\frac{1}{\sigma^2}\|\mathbf{x}-\mathbf{m}_{\theta}\|^2-\gamma|\mathbf{a}_{\theta}^{\mathbf{H}}(\mathbf{x}-\mathbf{m}_{\theta})|^2\\ =&\frac{1}{\sigma^2}\left(-\|\mathbf{x}\|^2+2\Re\{\mathbf{m}_{\theta+2\epsilon}^{\mathbf{H}}\mathbf{x}\}-\|\mathbf{m}_{\theta+2\epsilon}\|^2+\|\mathbf{x}\|^2-2\Re\{\mathbf{m}_{\theta}^{\mathbf{H}}\mathbf{x}\}+\|\mathbf{m}_{\theta}\|^2\right)\\ &+\gamma|\mathbf{a}_{\theta+2\epsilon}^{\mathbf{H}}\mathbf{x}-\mathbf{a}_{\theta+2\epsilon}^{\mathbf{H}}\mathbf{m}_{\theta+2\epsilon}|^2-\gamma|\mathbf{a}_{\theta}^{\mathbf{H}}\mathbf{x}-\mathbf{a}_{\theta}^{\mathbf{H}}\mathbf{m}_{\theta}|^2\\ =&\frac{1}{\sigma^2}\left(2\Re\{\mathbf{m}_{\theta+2\epsilon}^{\mathbf{H}}\mathbf{x}\}-2\Re\{\mathbf{m}_{\theta}^{\mathbf{H}}\mathbf{x}\}\right)\\ &+\gamma|\mathbf{a}_{\theta+2\epsilon}^{\mathbf{H}}\mathbf{x}-\mathbf{a}_{\theta+2\epsilon}^{\mathbf{H}}\mathbf{m}_{\theta+2\epsilon}|^2-\gamma|\mathbf{a}_{\theta}^{\mathbf{H}}\mathbf{x}-\mathbf{a}_{\theta}^{\mathbf{H}}\mathbf{m}_{\theta}|^2.\end{aligned} \tag{4.23}$$

Note that $\|\mathbf{m}_\theta\|^2=\|\mathbf{m}_{\theta+2\epsilon}\|^2=N|\mu_F|^2$, hence they cancelled each other out in the last step. Also $\mathbf{a}_\theta^{\mathbf{H}}\mathbf{m}_\theta=\mathbf{a}_{\theta+2\epsilon}^{\mathbf{H}}\mathbf{m}_{\theta+2\epsilon}=N\mu_F$. Let $y_0=\mathbf{a}_\theta^{\mathbf{H}}\mathbf{x}$ and $y_+=\mathbf{a}_{\theta+2\epsilon}^{\mathbf{H}}\mathbf{x}$, then we have

$$\begin{aligned}\Lambda =&\frac{1}{\sigma^2}\left(2\Re\{\mu_F^*y_+\}-2\Re\{\mu_F^*y_0\}\right)+\gamma|y_+-N\mu_F|^2-\gamma|y_0-N\mu_F|^2\\ =&\frac{1}{\sigma^2}\left(2\Re\{\mu_F^*y_+\}-2\Re\{\mu_F^*y_0\}\right)+\gamma|y_+|^2-2\gamma\Re\{N\mu_F^*y_+\}+\gamma|N\mu_F|^2\\ &-\gamma|y_0|^2+2\gamma\Re\{N\mu_F^*y_0\}-\gamma|N\mu_F|^2\\ =&2\left(\frac{1}{\sigma^2}-N\gamma\right)\Re\{\mu_F^*y_+\}-2\left(\frac{1}{\sigma^2}-N\gamma\right)\Re\{\mu_F^*y_0\}+\gamma|y_+|^2-\gamma|y_0|^2\\ =&\gamma\Bigg[|y_+|^2+2\left(\frac{1}{\gamma\sigma^2}-N\right)\Re\{\mu_F^*y_+\}+\left(\frac{1}{\gamma\sigma^2}-N\right)^2|\mu_F|^2\\ &-|y_0|^2-2\left(\frac{1}{\gamma\sigma^2}-N\right)\Re\{\mu_F^*y_0\}-\left(\frac{1}{\gamma\sigma^2}-N\right)^2|\mu_F|^2\Bigg]\\ =&\gamma\left|y_++\left(\frac{1}{\gamma\sigma^2}-N\right)\mu_F\right|^2-\gamma\left|y_0+\left(\frac{1}{\gamma\sigma^2}-N\right)\mu_F\right|^2.\end{aligned} \tag{4.24}$$

Expanding the γ term within parentheses, we can make further simplifications as,

$$\begin{aligned}\Lambda =&\gamma \left| y_+ + \left(\frac{1}{\frac{\sigma_F^2}{\sigma^2(\sigma^2+\sigma_F^2 N)}\sigma^2} - N \right) \mu_F \right|^2 - \gamma \left| y_0 + \left(\frac{1}{\frac{\sigma_F^2}{\sigma^2(\sigma^2+\sigma_F^2 N)}\sigma^2} - N \right) \mu_F \right|^2 \\ =&\gamma \left| y_+ + \frac{\sigma^2}{\sigma_F^2}\mu_F \right|^2 - \gamma \left| y_0 + \frac{\sigma^2}{\sigma_F^2}\mu_F \right|^2 . \end{aligned} \tag{4.25}$$

Consequently we have reached the difference of two non-central χ^2 (chi-square) distributed random variables of one degree of freedom. Luckily we are not forced to delve into these distributions as we are interested in $\mathrm{P}(\Lambda \geq 0)$, which has the following form.

$$\begin{aligned}\mathrm{P}(\Lambda \geq 0) =&\mathrm{P}\left(\gamma \left| y_+ + \frac{\sigma^2}{\sigma_F^2}\mu_F \right|^2 - \gamma \left| y_0 + \frac{\sigma^2}{\sigma_F^2}\mu_F \right|^2 \geq 0 \right) \\ =&\mathrm{P}\left(\left| y_+ + \frac{\sigma^2}{\sigma_F^2}\mu_F \right|^2 \geq \left| y_0 + \frac{\sigma^2}{\sigma_F^2}\mu_F \right|^2 \right) . \end{aligned} \tag{4.26}$$

The probability in equation (4.26) has an analytical form (see Appendix E, Stein's unified analysis of the error probability). To use the analytical forms, we need the first two moments of each Gaussian (each term within $|\cdot|^2$) and their cross correlation.

$$\mathrm{z}_0 \triangleq y_0 + \frac{\sigma^2}{\sigma_F^2}\mu_F = \mathbf{a}_\theta^{\mathbf{H}}\mathbf{x} + \frac{\sigma^2}{\sigma_F^2}\mu_F \tag{4.27}$$

$$\mathrm{z}_+ \triangleq y_+ + \frac{\sigma^2}{\sigma_F^2}\mu_F = \mathbf{a}_{\theta+2\epsilon}^{\mathbf{H}}\mathbf{x} + \frac{\sigma^2}{\sigma_F^2}\mu_F \tag{4.28}$$

Since $\mathbf{x} \sim \mathcal{CN}(\mathbf{m}_x, \mathbf{R}_x)$ and y_0, y_1 is defined as $y_0 = \mathbf{a}_\theta^{\mathbf{H}}\mathbf{x}$ and $y_+ = \mathbf{a}_{\theta+2\epsilon}^{\mathbf{H}}\mathbf{x}$; the random variables z_0 and z_+ are complex-valued circularly symmetric Gaussian random variables with the following mean, variance and covariance;

$$\begin{aligned}\mathbb{E}\{\mathrm{z}_0\} &= \mathbf{a}_\theta^{\mathbf{H}}\mathbf{m}_x + \frac{\sigma^2}{\sigma_F^2}\mu_F, \\ \mathbb{E}\{\mathrm{z}_+\} &= \mathbf{a}_{\theta+2\epsilon}^{\mathbf{H}}\mathbf{m}_x + \frac{\sigma^2}{\sigma_F^2}\mu_F, \\ \mathbb{E}\{|\mathrm{z}_0 - \bar{\mathrm{z}}_0|^2\} &= \mathbf{a}_\theta^{\mathbf{H}}\mathbf{R}_x\mathbf{a}_\theta, \\ \mathbb{E}\{|\mathrm{z}_+ - \bar{\mathrm{z}}_+|^2\} &= \mathbf{a}_{\theta+2\epsilon}^{\mathbf{H}}\mathbf{R}_x\mathbf{a}_{\theta+2\epsilon}, \\ \mathrm{cov}(z_0, z_+) &= \mathbf{a}_\theta^{\mathbf{H}}\mathbf{R}_x\mathbf{a}_{\theta+2\epsilon}. \end{aligned} \tag{4.29}$$

When there is no model mismatch, we have

$$\mathbf{m}_x = \mu_F \mathbf{a}_\theta \tag{4.30a}$$

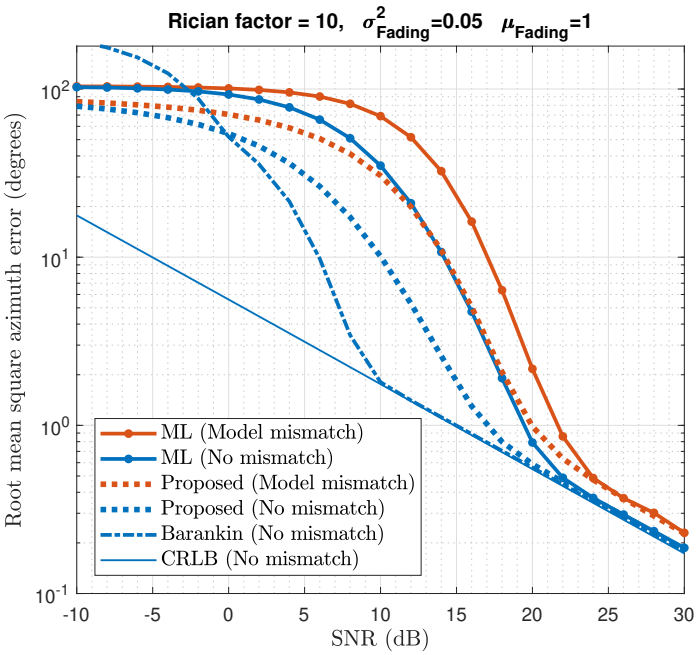
$$\mathbf{R}_x = \sigma^2 \mathbf{I} + \sigma_F^2 \mathbf{a}_\theta \mathbf{a}_\theta^{\mathbf{H}}. \tag{4.30b}$$

When the assumed model differs from the true model as $\mathbf{x} = \beta \mathbf{g}_\theta + \mathbf{v}$, we have

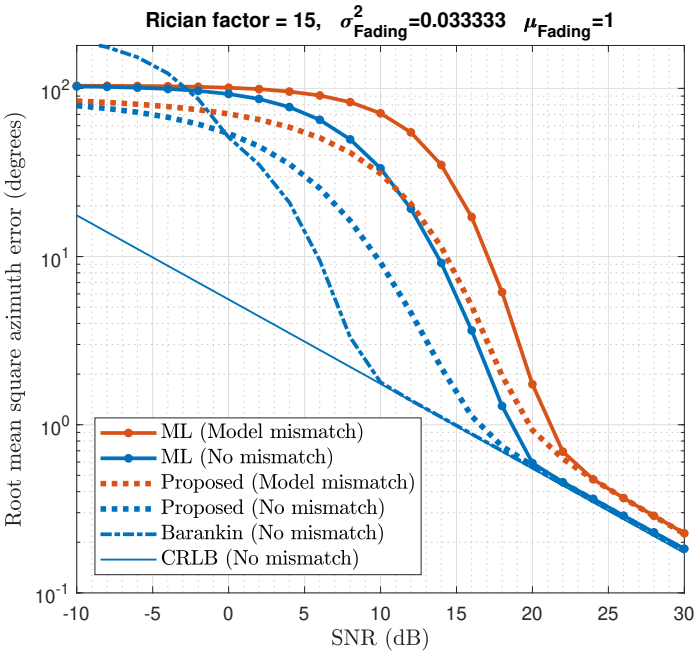
$$\mathbf{m}_x = \mu_F \mathbf{g}_\theta \tag{4.31a}$$

$$\mathbf{R}_x = \sigma^2 \mathbf{I} + \sigma_F^2 \mathbf{g}_\theta \mathbf{g}_\theta^{\mathbf{H}}. \tag{4.31b}$$

With these derivations, we can use the results in Section 3.5 for parameterized mean model and model mismatch. To check the derivations, we reconsider the DOA problem with a 12 element circular array in Section 3.6.2, under Rician fading with the observation model in (4.1). The same model mismatch (near-field and far-field) is applied by changing the array manifold vector, as in Section 3.6.2. The results for different Rician factors ($K = 10$ and $K = 15$) are given in Figure 4.2. Note that the proposed method is able to predict the threshold SNR as well as the asymptotic performance for both model match and model mismatch cases.



(a)



(b)

Figure 4.2: Near-field (model mismatch) and far-field (no mismatch) performance of a 12 element uniform circular array, for a target at 5λ distance, under Rician fading, (a) for $K = 10$, and (b) for $K = 15$.

4.1.3 On the accuracy of far-field approximation in direction finding

In this section, we investigate the far-field region assumption and try to verify the commonly used results with a different approach and discuss its tightness using the threshold behavior of performance curves. In most direction of arrival (DOA) applications, the target is assumed in the far-field region. Hence the signal model assumes a plane wave propagation. The far-field limit of a sensor array with an array aperture of D operating at the center frequency f_c is defined in [51] as,

$$R_{\text{far-field}} = \frac{2D^2}{\lambda}, \tag{4.32}$$

where $\lambda = \frac{c}{f_c}$ is the wavelength, and c is the propagation speed of the wave within the medium.

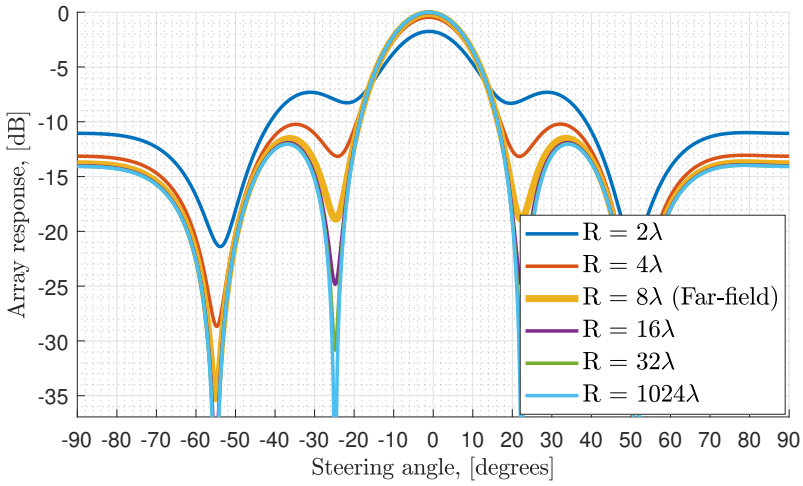


Figure 4.3: Beampatterns for a 5 element uniform line array with $\lambda/2$ spacing (using spherical spreading) for different target ranges.

The effect of far-field region can be illustrated by observing the beampatterns for different ranges of a target (using spherical spreading propagation model). The results for a 5 element uniform line array with $\lambda/2$ element spacing is given in Figure 4.3 and Figure 4.4.

In many cases, identifying the threshold region is crucial, as the performance of the estimator dramatically increases in the threshold region with a small increase in SNR. As explained in Chapter 1, this threshold behavior is attributed to gross error

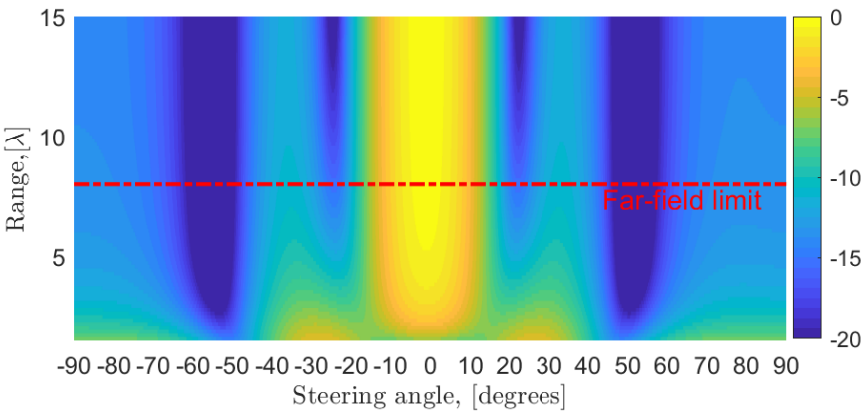


Figure 4.4: Beampatterns for a 5 element uniform line array with $\lambda/2$ spacing (using spherical spreading) for different target ranges

events. The probability of gross error events have a direct correlation with the array beampattern. Additionally, if the DOA estimation system uses the far-field propagation model, a model mismatch will be induced as well. For a complex exponential signal $\exp(j\omega t)$, emitted/reflected from the target with a position vector $\mathbf{p}_T$, the received signal has the following form:

$$r_k(t) = A\exp(j\omega t)\exp\left(-j\frac{2\pi}{\lambda}\|\mathbf{p}_T - \mathbf{p}_k\|\right), \tag{4.33}$$

where A is the complex valued amplitude and λ is the wavelength corresponding to the radial frequency ω. Dropping the carrier term the received signal (for the kth sensor) can be put into the following form for a linear array, using $\mathbf{p}_k = [\ x_k \quad 0\ ]^{\mathrm{T}}$:

$$\begin{aligned}
r_k &= A\exp\left(-j\frac{2\pi}{\lambda}\|\mathbf{p}_T - \mathbf{p}_k\|\right) \\
&= A\exp\left(-j\frac{2\pi}{\lambda}\left\|\begin{bmatrix} R\cos(\theta) & -x_k \\ R\sin(\theta) & \end{bmatrix}\right\|\right) \\
&= A\exp\left(-j\frac{2\pi}{\lambda}\sqrt{(R\cos(\theta) - x_k)^2 + (R\sin(\theta))^2}\right) \\
&= A\exp\left(-j\frac{2\pi}{\lambda}\sqrt{R^2 - 2R\cos(\theta)\,x_k + x_k^2}\right) \\
&= A\exp\left(-j\frac{2\pi}{\lambda}\sqrt{(R - x_k\cos(\theta))^2 + x_k^2\sin^2(\theta)}\right) \\
&= A\exp\left(-j\frac{2\pi}{\lambda}\left\|\begin{bmatrix} R - x_k\cos(\theta) \\ x_k\sin(\theta) \end{bmatrix}\right\|\right).
\end{aligned} \tag{4.34}$$

The benefit of the last equation is that it enables us to use the norm approximation (given in [52]) to be used:

$$\|\mathbf{x}\|_2 = |x_m| \sqrt{1 + \frac{\sum_{k=1,k\neq m}^{N} x_k^2}{x_m^2}}, \tag{4.35}$$

$$m = \underset{k=1,\ 2,\ ...,\ N}{\arg\max} \ |x_k| \tag{4.36}$$

$$\|\mathbf{x}\|_2 \approx |x_m| \left(1 + \frac{1}{2} \sum_{k=1,\ k\neq m}^{N} \frac{x_k^2}{x_m^2}\right) \tag{4.37}$$

where $\mathbf{x} \in \mathbb{R}^N,\ \mathbf{x} \neq \mathbf{0}$. Here the only approximation is $\sqrt{1+x} \approx 1 + \frac{x}{2}$ for small x. For $R > 2x_k$, we can use the following form for r_k:

$$r_k \approx A\exp\left(-j\frac{2\pi}{\lambda}\ \left((R - x_k\cos(\theta)\ ) + \frac{1}{2}\frac{x_k^2\sin^2(\theta)}{R - x_k\cos(\theta)}\right)\right). \tag{4.38}$$

For large values of R (far field), the equation reduces to the following form,

$$\begin{aligned} r_k &\approx A' \exp\left(-j\frac{2\pi}{\lambda}\ (R - x_k\cos(\theta))\right) \\ &= A' \exp\left(-j\frac{2\pi}{\lambda}R\right) \exp\left(j\frac{2\pi}{\lambda}\ x_k\cos(\theta)\right) \\ &= A'' \exp\left(j\frac{2\pi}{\lambda}\ x_k\cos(\theta)\right), \end{aligned} \tag{4.39}$$

which is the common assumption in most direction of arrival (DOA) applications and antenna theory [51].

We study the same example in Chapter 3 for model mismatch case. In that test scenario, the target was placed closer than the far-field limit. We can calculate the proposed MSE prediction expression for a grid of SNR and target range values and compare the results with the Cramer-Rao Lower Bound to observe whether the far-field limit expression is succesful at predicting the performance loss to the range. This ratio will represent the shift of the threshold region, as CRLB will be much smaller than the predicted MSE values below the threshold SNR value. Results for a uniform circular array with a $5\lambda/3$ radius is given in Figure 4.5 and Figure 4.6. Note that the array aperture for this array is the array diameter, $10\lambda/3$. Consequently, the far field region begins at $200\lambda/9$. The results in this figure show that, the far-field assumption is a valid one even at very low Rician factors. This on the other hand can be interpreted that the far-field limit is not a tight limit value for direction finding (DF) applications.

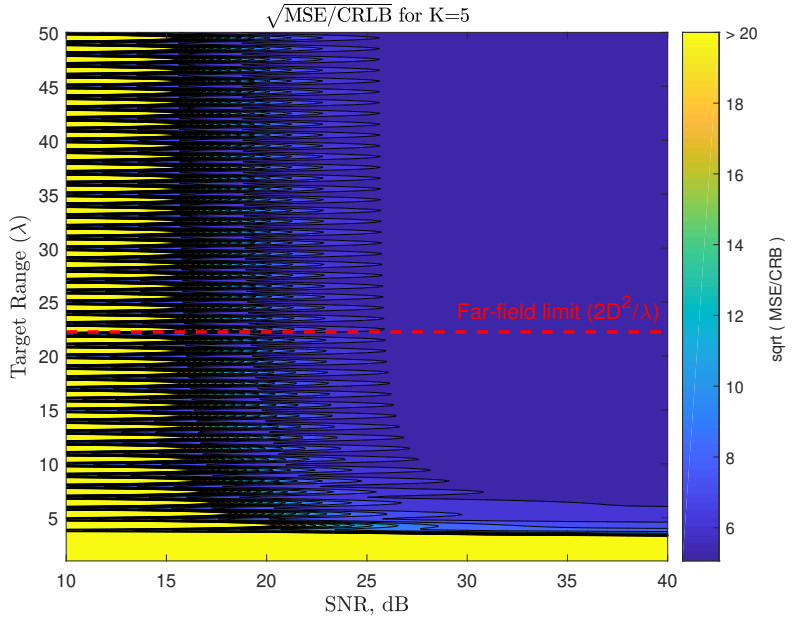


Figure 4.5: Ratio of predicted MSE values for ML estimator to CRLB for various SNR and target range values, for a Rician factor of 5. Note that the color scale is set between 5 and > 20.

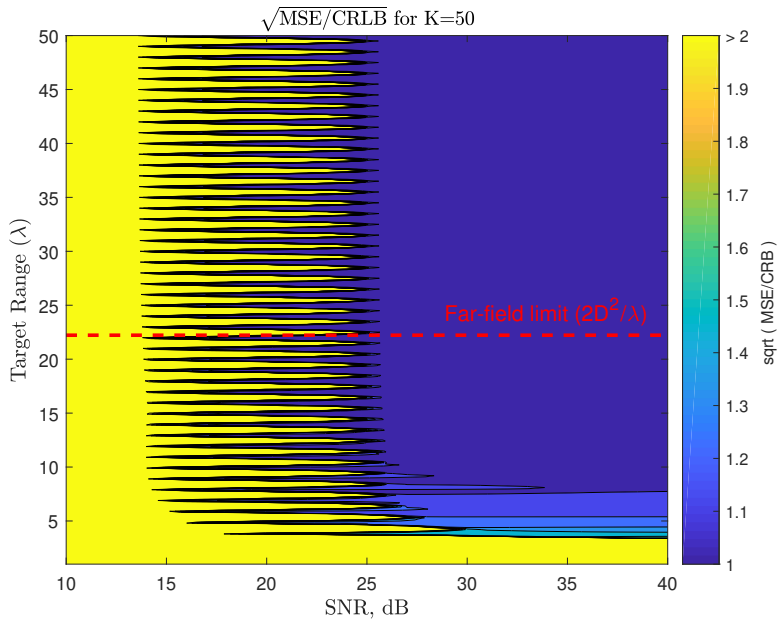


Figure 4.6: Ratio of predicted MSE values for ML estimator to CRLB for various SNR and target range values, for a Rician factor of 50. Note that the color scale is set between 1 and > 2.

4.2 DOA Estimation by an IDE (ESPRIT)

In this section we present the application of the proposed MSE prediction method in Section 3 to an implicitly defined estimator (other than ML estimator), namely ESPRIT. We consider a uniform linear array composed of $N = 15$ sensors with $\lambda/2$ element spacing. The signal model is as follows

$$x_n = \underbrace{\alpha e^{j\pi\cos(\bar{\phi})n}}_{\triangleq m_n(\bar{\phi})} + w_n, \quad n = 0, 1, \ldots, N-1, \tag{4.40}$$

where $\alpha \in \mathbb{C}$ is the unknown complex amplitude, $w_n \sim \mathcal{CN}(w_n, 0, \sigma_w^2)$, and $\bar{\phi} = 35\pi/180$ rad is the unknown true azimuth angle to be estimated. We denote the spatial frequency with $\bar{\omega}$ and define $\bar{\omega} \triangleq \pi\cos(\bar{\phi})$.

Due to the structure of uniform linear arrays, we can write $m_n(\bar{\phi}) = e^{j\bar{\omega}} m_{n-1}(\bar{\phi})$ for the elements of the array manifold vector $m_n(\bar{\phi})$ in (4.40), which is the rotational invariance property exploited in ESPRIT [53]. Using this property, we can define a somewhat adhoc cost function as follows

$$J(\omega) = \sum_{n=1}^{N-1} |x_n - e^{j\omega} x_{n-1}|^2. \tag{4.41}$$

By minimizing (4.41), we can get an estimate for ω as $\hat{\omega} \triangleq \arg\min_\omega J(\omega) = \arg\left(\sum_{n=1}^{N-1} x_{n-1}^* x_n\right)$; from which an estimate for the DOA can be generated as $\hat{\phi} \triangleq \arccos\left(\frac{\hat{\omega}}{\pi}\right)$, which we call the ESPRIT estimate. Note that the cost function $J(\cdot)$ is neither symmetric around the estimate, nor is unimodal. Hence it does not satisfy the conditions for which the proposed method yields the true MSE. The cost function $J(\omega)$ in (4.41) can be written in matrix form as follows.

$$\begin{aligned} J(\omega) &= \left\|\mathbf{A}_1\mathbf{x} - e^{j\omega}\mathbf{A}_0\mathbf{x}\right\|^2 \\ &= \mathbf{x}^{\mathrm{H}} \left(\mathbf{A}_1 - e^{j\omega}\mathbf{A}_0\right)^{\mathrm{H}} \left(\mathbf{A}_1 - e^{j\omega}\mathbf{A}_0\right) \mathbf{x}, \end{aligned} \tag{4.42}$$

where

$$\begin{aligned} \mathbf{x} &= \begin{bmatrix} x_0 & x_1 & \ldots & x_{N-1} \end{bmatrix}^{\mathrm{T}}, \\ \mathbf{A}_0 &= \begin{bmatrix} \mathbf{I}_{(N-1)} & \mathbf{0}_{(N-1)\times 1} \end{bmatrix}, \\ \mathbf{A}_1 &= \begin{bmatrix} \mathbf{0}_{(N-1)\times 1} & \mathbf{I}_{(N-1)} \end{bmatrix}. \end{aligned} \tag{4.43}$$

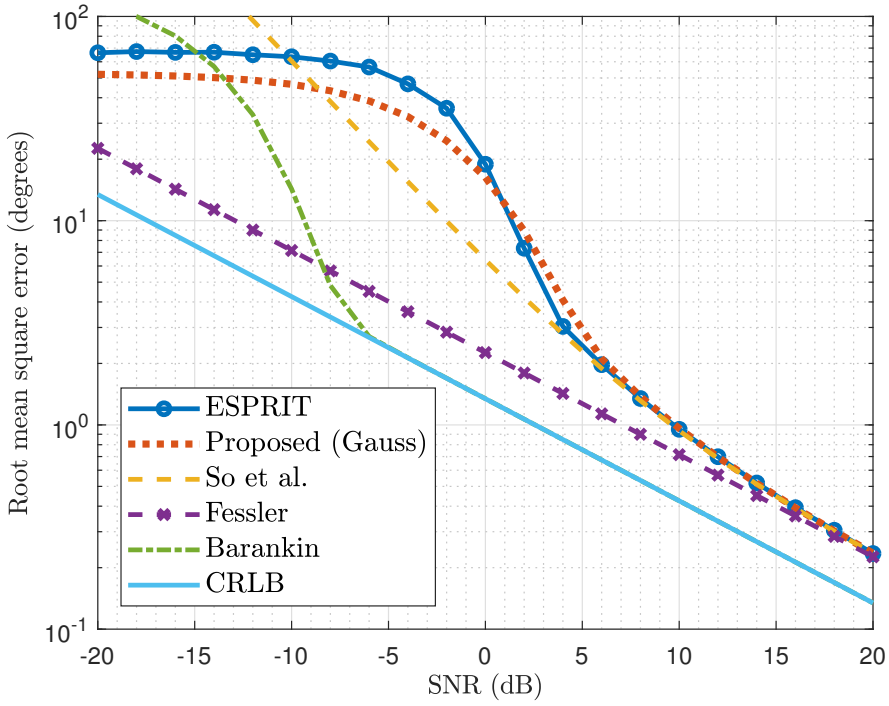


Figure 4.7: Non-random DOA estimation performance of ESPRIT along with values of different bounds and MSE prediction expressions.

Using $\omega \triangleq \pi \cos(\phi)$, we get,

$$J(\phi) = \mathbf{x}^{\mathrm{H}}\big(\mathbf{A}_1 - e^{j\pi\cos(\phi)}\mathbf{A}_0\big)^{\mathrm{H}}\big(\mathbf{A}_1 - e^{j\pi\cos(\phi)}\mathbf{A}_0\big)\mathbf{x}. \tag{4.44}$$

Note that in order to use the approximate MSE expression in (3.12), we need to evaluate the following probability,

$$P\left(J(\bar{\phi}+2\epsilon) \leq J(\bar{\phi})\right) = P\left(J(\bar{\phi}+2\epsilon) - J(\bar{\phi}) \leq 0\right) \triangleq P(\Delta J_{2\epsilon} \leq 0), \tag{4.45}$$

where the inequalities are the reverse of those in Remark 1 since we have a minimization problem instead of a maximization problem in our IDE. Using (4.44) and after some basic algebraic operations we can express $\Delta J_{2\epsilon} \triangleq J(\bar{\phi}+2\epsilon) - J(\bar{\phi})$ as $\Delta J_{2\epsilon} = \mathbf{x}^{\mathrm{H}}\mathbf{Q}\mathbf{x}$ where

$$\mathbf{Q} \triangleq (e^{j\pi\cos(\bar{\phi})} - e^{j\pi\cos(\bar{\phi}+2\epsilon)})\mathbf{A}_1^{\mathrm{H}}\mathbf{A}_0 + (e^{-j\pi\cos(\bar{\phi})} - e^{-j\pi\cos(\bar{\phi}+2\epsilon)})\mathbf{A}_0^{\mathrm{H}}\mathbf{A}_1. \tag{4.46}$$

Even though the density of the quadratic form $\Delta J_{2\epsilon} = \mathbf{x}^{\mathrm{H}}\mathbf{Q}\mathbf{x}$ is known to be the generalized chi-squared distribution and can be evaluated numerically [54, Appendix

A], we pursue a Gaussian fit to the density in order to simplify the probability calculations. To do that, we evaluate the first two moments of $\Delta J_{2\epsilon}$. Using the fact that $\mathbf{x}^{\mathrm{H}}\mathbf{Q}\mathbf{x}$ is always real, we can reach the following expressions (after some algebra)

$$\mu_{\Delta}(\bar{\phi},\epsilon) = \mathbb{E}\{\Delta J_{2\epsilon}\} = \sigma_w^2 \mathrm{tr}\big(\mathbf{Q}\big) + \mathbf{m}^{\mathrm{H}}(\bar{\phi})\mathbf{Q}\mathbf{m}(\bar{\phi}), \tag{4.47a}$$

$$\sigma_{\Delta}^2(\bar{\phi},\epsilon) = \mathrm{Var}\{\Delta J_{2\epsilon}\} = \sigma_w^4 \mathrm{tr}(\mathbf{Q}^2) + 2\sigma_w^2 \mathbf{m}^{\mathrm{H}}(\bar{\phi})\mathbf{Q}^2\mathbf{m}(\bar{\phi}), \tag{4.47b}$$

where $\mathbf{m}(\phi) \triangleq [m_0(\phi),\, m_1(\phi), \cdots, m_{N-1}(\phi)]^{\mathrm{T}}$ and we used the result $\mathbb{E}[(\mathbf{y}^{\mathrm{H}}\mathbf{Q}\mathbf{y})^2] = \mathrm{tr}^2(\mathbf{Q}\boldsymbol{\Sigma}) + \mathrm{tr}((\mathbf{Q}\boldsymbol{\Sigma})^2)$ for any Hermitian matrix $\mathbf{Q}$ and $\mathbf{y} \sim \mathcal{CN}(\mathbf{y}; \mathbf{0}, \boldsymbol{\Sigma})$ [55, Ch. V, Lemma 2.2]. With the Gaussian fit, an approximation to the suggested MSE expression becomes

$$\widehat{\mathrm{MSE}}(\bar{\phi}) \approx 2\int_{-\frac{\bar{\phi}}{2}}^{\frac{\pi-\bar{\phi}}{2}} |\epsilon| \mathcal{N}_{\mathrm{cdf}}\left(0; \mu_{\Delta}(\bar{\phi},\epsilon), \sigma_{\Delta}^2(\bar{\phi},\epsilon)\right) \mathrm{d}\epsilon. \tag{4.48}$$

where we used (3.13) with $\phi_{\min} = 0$; $\phi_{\max} = \pi$ to set the integration limits and the cdf of the normal distribution is used instead of the ccdf due to the reversal of the inequalities in (4.45). Figure 4.7 shows the results of 10,000 Monte Carlo runs for this experiment. The CRLB, BB with single test point optimized over a grid, Fessler's [1] and So et al.'s [2] methods are also illustrated for comparison purposes. Note that the estimator in this experiment is not efficient, hence its performance does not reach CRLB at high SNR. Consequently, the estimator performance is not characterized by the CRLB in any SNR region. Therefore, one needs the asymptotic MSE values as well as the pairwise error probabilities in order to calculate the MSE prediction using MIE. BB provides a very optimistic prediction for this specific problem as in the earlier examples. Although Fessler's and So et al.'s methods predicted the estimator performance well at high SNR region, they have difficulty in representing gross errors of the estimator for low SNR values. The proposed method, on the other hand, closely follows the estimator performance in all SNR regions.

Using the results in Section 3.7 we can also examine the bias performance. The approximate expression in this case reduces to,

$$\widehat{\mathrm{MSE}}(\bar{\phi}) \approx \int_{-\frac{\bar{\phi}}{2}}^{\frac{\pi-\bar{\phi}}{2}} \mathrm{sgn}(\epsilon) \mathcal{N}_{\mathrm{cdf}}\left(0; \mu_{\Delta}(\bar{\phi},\epsilon), \sigma_{\Delta}^2(\bar{\phi},\epsilon)\right) \mathrm{d}\epsilon. \tag{4.49}$$

Bias prediction results of the method are given in Figure 4.8. As seen in this figure, the Taylor expansion based method by So et al. [2] is unable to predict estimator bias

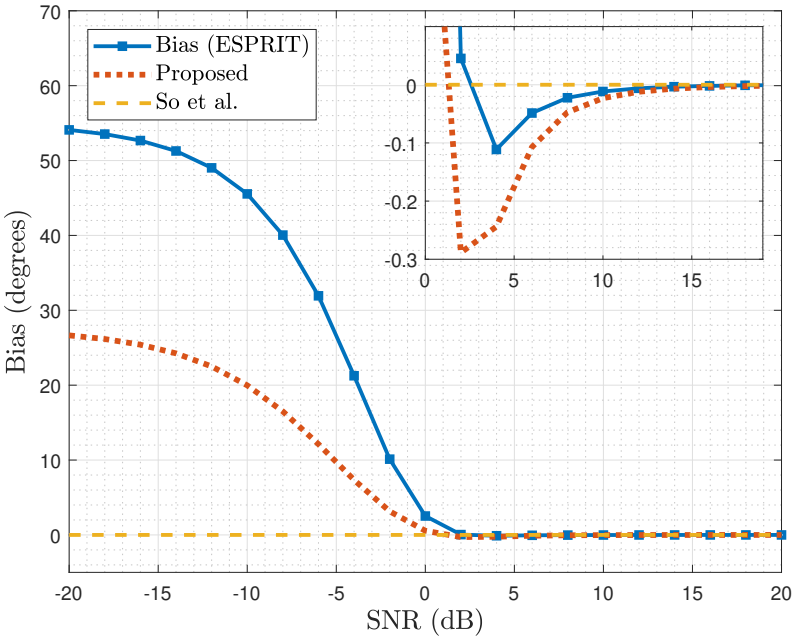


Figure 4.8: Direction of arrival estimation bias for a uniform linear array of 15 elements with $\lambda/2$ element spacing using ESPRIT.

for low SNR values. The proposed method is able to track the behavior of estimator bias along with the sign of the bias as seen in the zoomed subplot.

4.3 Online array performance prediction in case of sensor failures

In this section we present an application of the proposed method in Chapter 3 to predict the sensor array performance (for direction of arrival estimation) in case one or more sensors fail to operate. This application may have some use in military sensor arrays, as they quite often are exposed to very harsh conditions. Although these arrays and their sensors are designed to withstand the expected harsh conditions, there is always a non-zero probability to observe a faulty sensor. Also, it is always possible to be in an unfortunate accident which results in a certain portion of the array to be damaged and some sensors may be unable to function properly. In such cases, the operator or the system may decide to disable certain sensors within the array due to its unreliable readings. The system performance will also change depending on the number of the disabled sensors, and depending on the "importance" of those sensors. Even if most of the array sensors are not in use, the overall system can still be useful for a particular angular sector. In such cases, the operator needs a performance prediction to assess the reliability of array outputs. As the number of combinations of sensors that might be disabled is in general very large (2^N subsets for an N-element array), it is more practical to assess the array performance online. The problem here is to provide the operator with a prediction of the current performance statistics, in a rather simple and abstract manner. Since the performance deviates depending on the signal to noise ratio, bearing and frequency, representation of 4-dimensional data as simple as possible is also a challenge.

Using Bayesian performance bounds is not an appropriate solution, because the Bayesian bound represents the weighted (by apriori density of the parameter) sum of the performance for all possible DOA angles. As a result, the operator cannot understand performance in which angular sector is affected from a rise in the Bayesian bound. Consequently, we need to resort to non-Bayesian methods, bounds or approximate performance metrics. Using approximate metrics is a better way for this problem, since the estimation method is known and actually its current performance is questioned. As explained in detail in Chapter 2 the threshold region performance must also be predicted, since the decrease in the number of healthy sensors is likely to shift the threshold SNR to higher values. For these reasons, the proposed method in

Chapter 3 is a good candidate for this application. The system performance can be calculated online using the parametric mean model presented in Section 3.5 to calculate the system performance for different direction of arrival angles for different SNR values.

To illustrate this application, a 12 element circular array with directional elements is used. The sensor directivity is defined as,

$$d_n(\phi) = |\cos(\phi - \phi_n) + 0.5|^8, \quad \phi_n = \frac{2\pi n}{N} \tag{4.50}$$

where, $N = 12$ is the number of sensors in the array. With directional elements, we have the following parametric mean model for direction of arrival estimation on a plane (azimuth only),

$$\mathbf{a}_\psi = [a_1, a_2, \ldots, a_N]^{\mathrm{T}}, \quad a_n = d_n(\phi) \exp\left(j\frac{2\pi}{\lambda}\mathbf{p}_n^{\mathrm{T}}\mathbf{u}_\phi\right), \tag{4.51}$$

$$\mathbf{u}_\phi = \begin{bmatrix} \cos(\phi) \\ \sin(\phi) \end{bmatrix}, \quad \mathbf{p}_n = \begin{bmatrix} p_n^{\mathsf{x}} \\ p_n^{\mathsf{y}} \end{bmatrix}, \tag{4.52}$$

where, ϕ is the azimuth angle $\in [0,\, 2\pi)$ rads (measured from the x-axis in counter-clockwise direction). $\mathbf{p}_n$ is the position vector of the nth sensor containing the x and y. The sensor measurement vector $\mathbf{x} \in \mathbb{C}^N$ under additive noise is modeled as

$$\mathbf{x} = \mathbf{a}_\phi + \mathbf{v}, \quad \mathbf{x} \sim \mathcal{CN}(\mathbf{a}_\phi, \sigma^2\mathbf{I}_N). \tag{4.53}$$

With these definitions, we can use (3.64) for a grid of azimuth angles and SNR values. The configuration and performance of fully operational sensor array are given in Figure 4.9 and 4.10. Array configuration and performance when three sensors are not operational are given in Figure 4.11 and 4.12. Since the array sensors are directional, and most of the information from an angular sector is collected from the sensors directed to that direction, we can expect the array performance to deteriorate in the angular sectors of the broken sensors for this problem. Figure 4.12 clearly indicates a performance drop in these sectors.

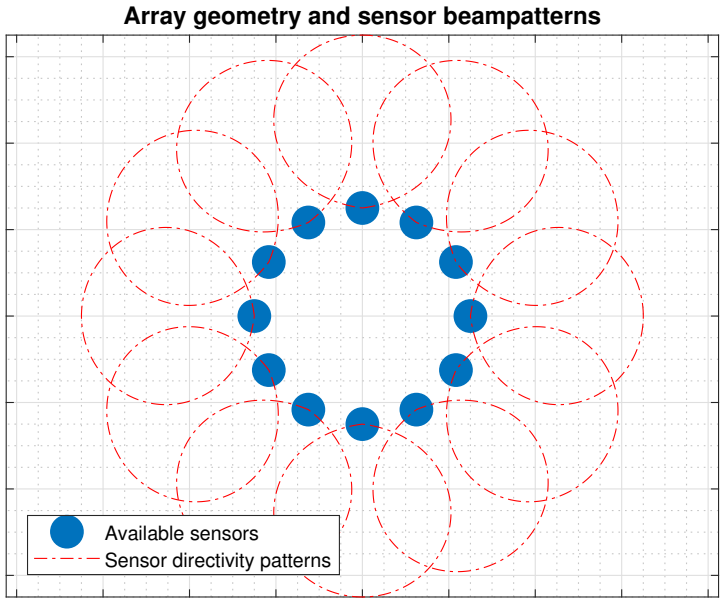


Figure 4.9: Sensor array configuration when all sensors are operational.

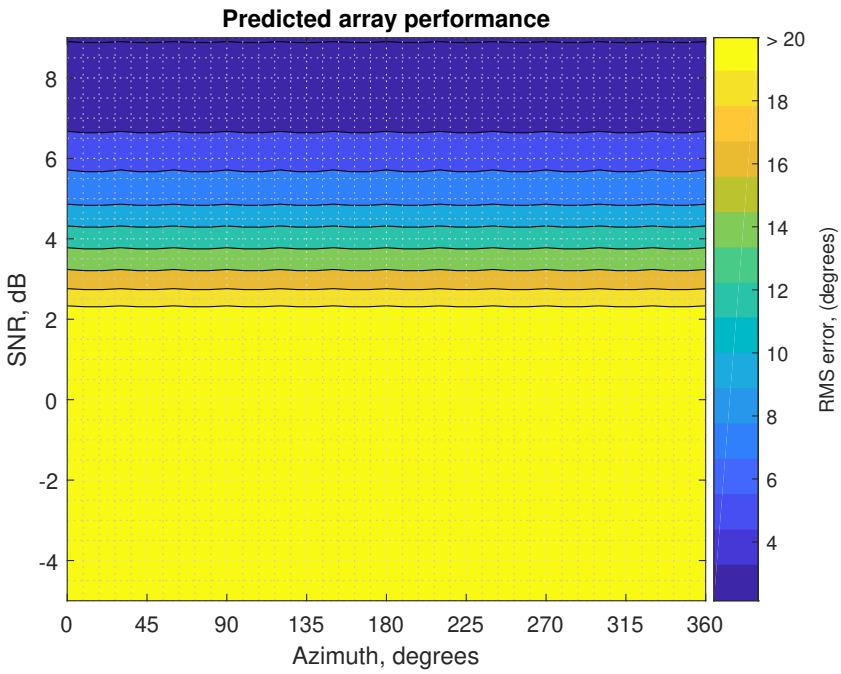


Figure 4.10: Sensor array performance when all sensors are operational.

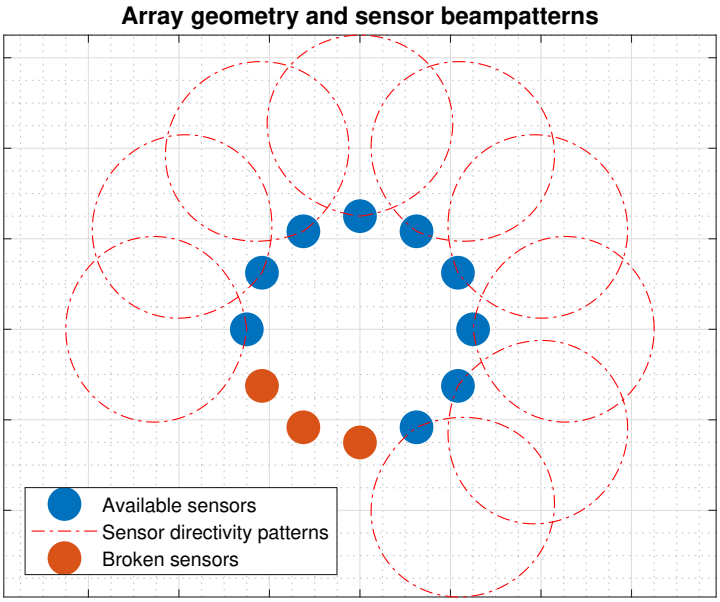


Figure 4.11: Sensor array configuration when 3 sensors are lost.

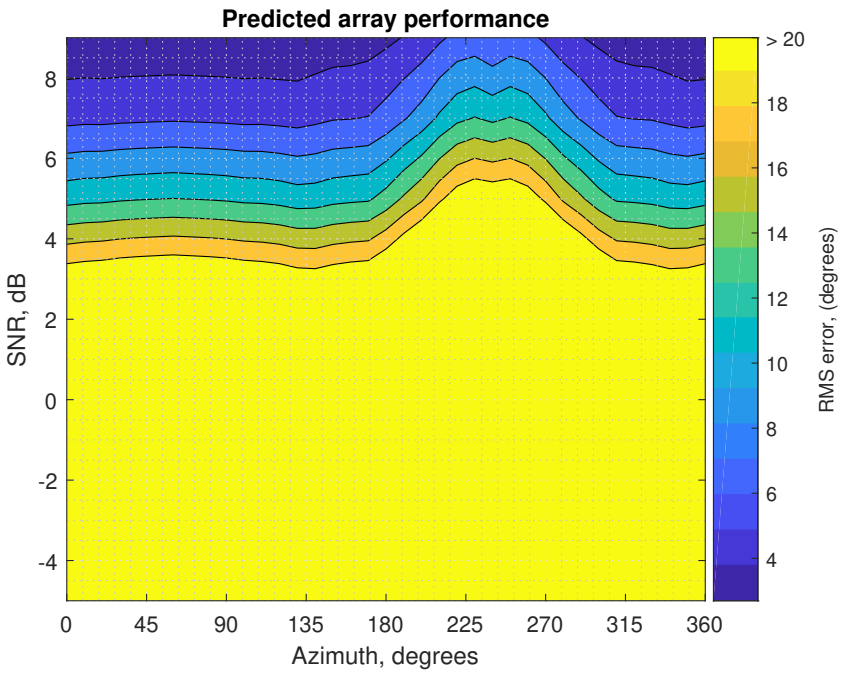


Figure 4.12: Sensor array performance when 3 sensors are lost.

4.4 Array layout optimization by minimizing the probability of gross errors

As explained in detail in previous chapters, the performance of ML estimator is affected drastically below a certain threshold SNR value. And the reason for this abrupt change in performance is gross errors. Consequently, if the threshold SNR value is to be lowered, the probability of gross errors must be decreased. In this section we present a possible application of the proposed performance prediction method in Chapter 3 to sensor array layout optimization by minimizing the gross error probability. The idea to optimize sensor positions to reduce the probability of gross errors is not new (see for instance [56]). We present an extension of this idea, i.e. using the performance prediction method as the cost function of the optimization routine.

Consider the direction of arrival problem, where we have a set of sensors with noisy observations. Array processing literature tells us that to increase the directivity index of a sensor array, the critically sampled array aperture must be increased. This is an analogous observation to frequency estimation, in which we need a longer observation period to better estimate the center frequency. However, increasing the array aperture may not be feasible due to mechanical constraints, or the number of sensors may be too large for the available processing power in the system. In such cases, the sensor positions are still a design parameter and can be re-arranged (within certain constraints). The cost function in this case can be chosen as the average or maximum of predicted MSE values for ML estimator for the desired angle of arrivals, and a specific SNR value. The optimization problem can be constructed as follows;

$$\min_{\mathbf{p}_s} \sum_{\phi \in \Phi} \widehat{\mathrm{MSE}}(\phi), \quad \text{s.t.} \quad \mathbf{p}_s \in S, \quad \mathrm{SNR} = \eta_0 \tag{4.54}$$

where, $\mathbf{p}_s$ is an N by 2 matrix (x and y coordinates, for a 2D array optimization), composed of sensor positions. S is the set of feasible sensor positions, and Φ is the set of angles for which the probability of error is to be minimized at an SNR value of η_0. To simplify the optimization definition, we can use polar coordinates for sensor positions, and define the constraints accordingly as,

$$\min_{\substack{r_n \in [r^n_{\min}, r^n_{\max}] \\ \theta_n \in [\theta^n_{\min}, \theta^n_{\max}]}} \sum_{\phi \in \Phi} \widehat{\mathrm{MSE}}(\phi), \quad \text{s.t.} \quad \mathrm{SNR} = \eta_0 \tag{4.55}$$

Choosing this SNR value is critical at this point, since if we choose an SNR value above the threshold SNR, the optimization is forced to minimize the CRLB essentially. This in turn does not provide any improvement for reducing the gross error probability. The threshold SNR value under consideration here belongs to the baseline array configuration, whose gross error performance is to be improved. As a result, we need to find the threshold SNR before beginning the optimization routine. Note that choosing an SNR value in no-information region is again a futile attempt for this problem, since we have the same number of sensors and cannot move the no-information region a useful amount.

Note that the optimization problem in (4.55) requires a 2N-dimensional search. Heuristic optimization algorithms (such as genetic algorithm, simulated annealing etc. [57]) can be used here to provide "a" solution (not necessarily "the" global optimum solution). In the following example, genetic algorithm is used to optimize the array layout to minimize the probability of gross errors. An 8-element uniform circular array with a $\frac{5\lambda}{3}$ radius is used as a baseline array (Figure 4.13), whose gross error probability is to be reduced. The optimization in (4.55) is carried out for 8 sensor positions $\{r_k, \theta_k\}_{k=1}^{8}$ for a set of azimuth angles covering $[0, 2\pi)$ with 10 degree steps, i.e. $\Phi = \{\frac{2\pi n}{36}\}_{n=0}^{35}$. The cost function in (4.55) is evaluated for an SNR value of 6 dB, which is within the threshold region of the average performance curve for the azimuth angles in Φ (as seen in Figure 4.16). The feasible area for each sensor position, is defined as follows;

$$\begin{aligned} r_{\text{min}}^{n} &= \frac{5\lambda}{6}, \quad r_{\text{max}}^{n} = \frac{5\lambda}{3}, \\ \theta_{\text{min}}^{n} &= \frac{2\pi n}{8} - \frac{\pi}{12}, \quad \theta_{\text{min}}^{n} = \frac{2\pi n}{8} + \frac{\pi}{12}, \quad n = 1, \ldots, 8. \end{aligned} \tag{4.56}$$

The default routine in Matlab for genetic algorithm is used to find a solution with these constraints. The optimized array configuration and feasible search areas is given in Figure 4.13. The performance of the optimized array is tested with 10^4 Monte Carlo runs for each azimuth angle in Φ. The average performance for each sensor configuration is given in Figure 4.16. The performance for each azimuth angle in Φ is given in Figure 4.14

Note that the optimization objective was to reduce the predicted MSE value at 6 dB SNR. As seen in the results in Figure 4.16, the optimization routine succeeded in finding a solution as the performance is significantly improved for this SNR value.

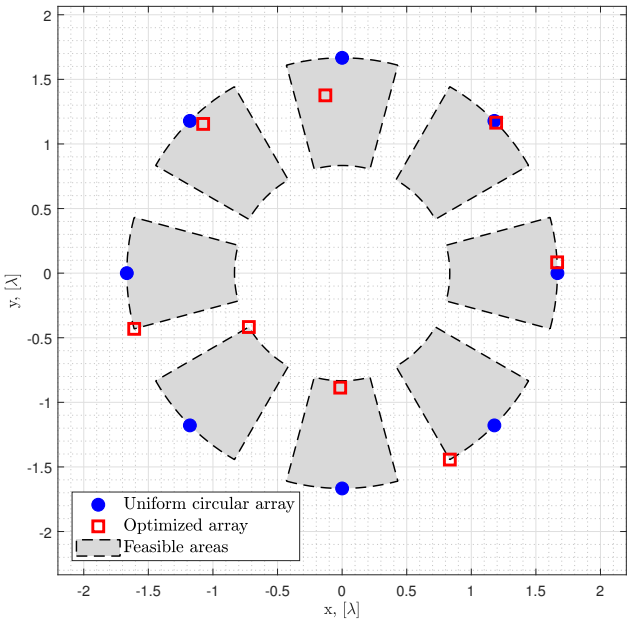


Figure 4.13: 8 element uniform circular array and layout optimized array configurations. Note that the search for genetic optimization is constrained to the feasible areas (one sensor within each closed region).

The actual goal of the optimization was also satisfied, the threshold SNR is shifted to lower levels (i.e. gross error probability is decreased for identical SNR values). Note that this improvement comes with a trade-off, which is the asymptotic performance. The uniform circular array provides a higher accuracy at high SNR values, which was not included in the objective of the optimization.

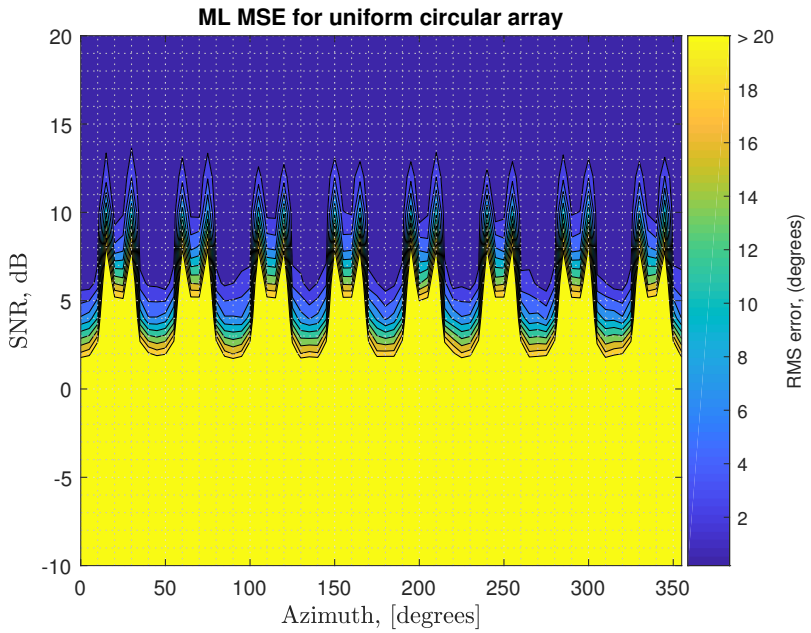


Figure 4.14: MSE values for each direction of arrival angle in Φ (using ML estimator) for the uniform circular array in Figure 4.13.

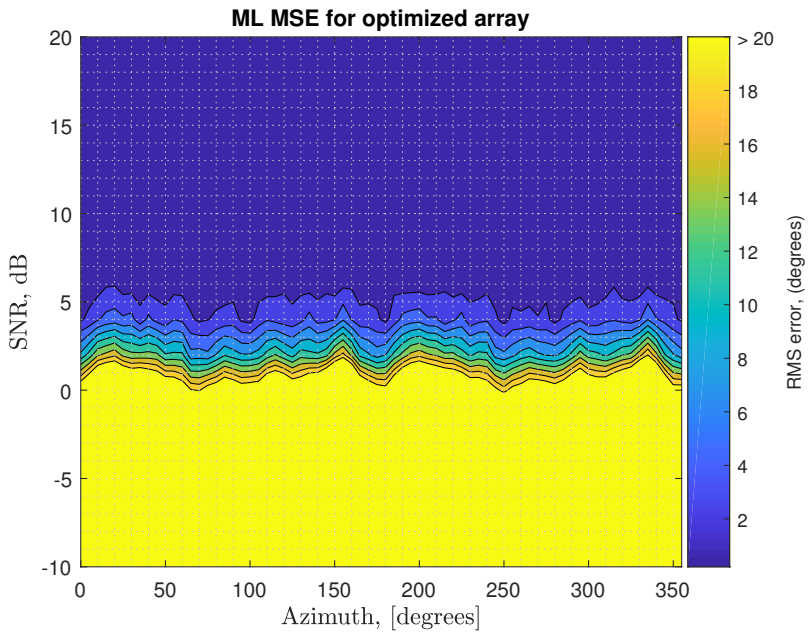


Figure 4.15: MSE values for each direction of arrival angle in Φ (using ML estimator) for the optimized array configuration in Figure 4.13.

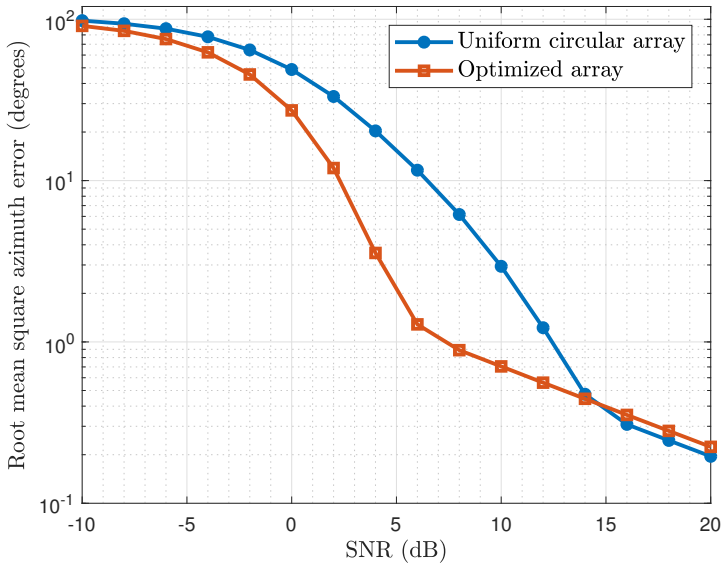


Figure 4.16: Average of MSE values (for ML estimator) for direction of arrival angles in $\Phi = \{\frac{2\pi n}{36}\}_{n=0}^{35}$, for each array configuration. 10^4 Monte Carlo runs are carried out for each azimuth angle.

CHAPTER 5

CONCLUSION

In this study we propose an MSE expression for the performance prediction of IDEs of non-random parameters. The method provides the exact MSE value when the objective function of the IDE is unimodal and symmetric. Even though, this is a rather stringent restriction for the general practice; the symmetric unimodal objective function assumption is in alignment with the operation of consistent estimators in the asymptotic region. The maximum likelihood estimator is the prime example for consistent estimators. Specific to the maximum likelihood estimator, it has been shown that the suggested MSE expression reduces to the CRLB and MCRLB in no-misspecification and misspecification cases, respectively. Furthermore, the suggested expression also yields the ZZB when an a-priori distribution is assigned to the unknown parameter for the MAP estimator.

An extension of the suggested MSE expression to the parameter estimation in the presence of nuisance parameters is given. Numerically friendly, but approximate, versions of the MSE expression are developed and some application examples are given. Numerical results show that the expression not only predicts the performance in the asymptotic region, but also provides valuable information in the threshold region. We consider that the applicability of the expression in other regions is related with the gradual degradation of asymptotic region operation conditions as the operating point moves from asymptotic region to the threshold region, say, with the reduction of SNR.

A possible interpretation for the MSE expression can be given in relation to the method of intervals (MIE). The MIE predicts the MSE by taking into account both small and gross error events via CRLB and the interval error probabilities, respectively. The suggested MSE expression for the ML estimator uses the likelihood ratio

for the same purpose; but it does not have a problem specific interval selection.

Another interpretation for the MSE expression can be given in connection with the ZZB. As in ZZB, the suggested MSE expression is based on the pairwise error probabilities. Furthermore, the average of the expression for the MAP estimator exactly reproduces ZZB for random parameters. Hence, the suggested MSE expression for the ML estimator can also be considered, at least informally, as the non-random parameter version of the ZZB.

The relationship of the proposed MSE expression with hybrid performance bounds is also an interesting subject and is left as a future work.

An interesting observation in the non-random parameter case was that, for medium and low SNR, the proposed MSE expressions usually slightly underestimated the true MSE of the estimators. Hence, a potential future study is to investigate whether the proposed expressions have any lower bounding properties in medium and/or low SNR regions under some conditions.

As a natural extension of the proposed MSE prediction method, bias prediction expression is also derived, and similar numerical studies are repeated.

Several applications are presented where the proposed expressions are utilized.

- Direction of arrival estimation under Rician fading problem is examined in detail, and it is shown that CRLB is not a useful tool for predicting the ML performance under Rician fading even in the high SNR region (or at least feasibly high SNR regime). Exact probability of error expressions are derived for the proposed prediction method, and it is shown (by numerical results) that the proposed method is able to predict the performance in all SNR regions. Note that the fading model under consideration produced the same fading effect on all sensors for a given snapshot. As a possible extension to this study, the case where each sensor is affected with a different realization from the same fading distribution can be studied. The mismatch effect in this case is also an interesting subject to be studied as well.

- A nonlinear IDE example (DOA estimation by ESPRIT) is given, and proba-

bility of error expressions are derived. Certain probability expressions are approximated using a Gaussian distribution by matching the first two moments, and numerical results are presented.

- Commonly used far-field assumption is shown to be an overly cautious factor by using the mismatched model performance of ML estimator.

- Calculation of the proposed MSE estimation method for a grid of values in the parameter space is proposed for online performance prediction of large sensor arrays, which will be especially useful in case of sensor failures. This approach gives the array operator to make a decision to use the sensor array for certain parameter values, even multiple sensors at random positions are defective.

- The proposed expression is used as the cost function to optimize sensor positions, that is to minimize the probability of gross error events. An example is given where a constrained optimization problem is constructed to improve the threshold SNR value. It is shown (using Monte Carlo simulations) that the gross error performance of a uniform circular array is improved, at the cost of asymptotic performance.

REFERENCES

[1] J. A. Fessler, "Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography," *IEEE Trans. Image Proc.*, vol. 5, no. 3, pp. 493–506, 1996.

[2] H. C. So, Y. T. Chan, K. C. Ho, and Y. Chan, "Simple formulas for bias and mean square error computation," *IEEE Sig. Proc. Mag.*, pp. 162–165, July 2013.

[3] H. L. V. Trees and K. L. Bell, *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*. Wiley-IEEE Press, 2007.

[4] R. A. Fisher, "On the mathematical foundations of theoretical statistics," *Philosophical Transactions of the Royal Society of London*, vol. 222, p. 309–368, April 1922.

[5] C. R. Rao, "On the mathematical foundations of theoretical statistics," *Bulletin of Calcutta Mathematical Society*, vol. 37, pp. 81–91, 1945.

[6] H. Cramer, *Mathematical Methods of Statistics*. Princeton, New Jersey: Princeton University Press, 1946.

[7] S. M. Kay, *Fundamentals of Statistical Signal Processing, Vol I: Estimation Theory*. Prentice-Hall, 1993.

[8] H. L. V. Trees, *Detection, Estimation and Modulation Theory*. John Wiley and Sons Inc., 1968.

[9] S. Fortunati, F. Gini, M. S. Greco, and C. D. Richmond, "Performance bounds for parameter estimation under misspecified models: Fundamental findings and applications," *IEEE Sig. Proc. Mag.*, vol. 34, no. 6, pp. 142–157, 2017.

[10] C. D. Richmond and L. L. Horowitz, "Parameter bounds on estimation accuracy under model misspecification," *IEEE Trans. Sig. Proc.*, vol. 63, pp. 2263–2278, May 2015.

[11] H. White, "Maximum likelihood estimation of misspecified models," *Econometrica*, vol. 50, no. 1, pp. 1–25, 1982.

[12] E. Mehmetcik and Ç. Candan, "Prediction and representation of array performance under sensor failure," in Underwater Defence Technology Europe*, Stockholm, Sweden*, May 2019.

[13] E. Mehmetcik, U. Orguner, and Ç. Candan, "An approximate MSE expression for maximum likelihood and other implicitly defined estimators of non-random parameters," *Signal Processing*, vol. 204, p. 108838, 2023.

[14] A. J. Weiss and E. Weinstein, "Fundamental limitations in passive time delay estimation - Part I : Narrow-band systems," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 31, pp. 472–486, April 1983.

[15] J. Ziv and M. Zakai, "Some lower bounds on signal parameter estimation," *IEEE Trans. Inf. Theory*, vol. 15, pp. 386–391, May 1969.

[16] S. Bellini and G. Tartara, "Bounds on error in signal parameter estimation," *IEEE Trans. Comm.*, vol. 22, pp. 340–342, Mar. 1974.

[17] K. L. Bell, Y. Steinberg, Y. Ephraim, and H. L. Van Trees, "Extended Ziv-Zakai lower bound for vector parameter estimation," *IEEE Trans. Inf. Theory*, vol. 43, pp. 624–637, March 1997.

[18] E. Nitzan, T. Routtenberg, and J. Tabrikian, "A new class of Bayesian cyclic bounds for periodic parameter estimation," *IEEE Trans. Sig. Proc.*, vol. 64, no. 1, pp. 229–243, 2016.

[19] A. Xu and T. Coleman, "Minimax lower bounds for circular source localization," in *2020 IEEE Int. Symp. on Inf. Theory (ISIT)*, pp. 1242–1247, 2020.

[20] T. Routtenberg and J. Tabrikian, "Bayesian periodic Cramér-Rao bound," *IEEE Sig. Proc. Lett.*, vol. 29, pp. 1878–1882, 2022.

[21] R. G. Gallager, *Stochastic Processes: Theory for Applications*. New York: Cambridge Uni. Press, 2014.

[22] D. G. Chapman and H. Robbins, "Minimum variance estimation without regularity assumptions," *Ann. Math. Statist.*, vol. 22, pp. 581–586, Dec 1951.

[23] E. W. Barankin, "Locally best unbiased estimates," *Ann. Math. Statist.*, vol. 20, pp. 477–501, 1949.

[24] R. McAulay and E. Hofstetter, "Barankin bounds on parameter estimation," *IEEE Trans. Inf. Theory*, vol. 17, no. 6, pp. 669–676, 1971.

[25] P. Forster and P. Larzabal, "On lower bounds for deterministic parameter estimation," in *IEEE Int. Conf. on Acoustics, Speech, and Sig. Proc.*, vol. 2, pp. II–1137–II–1140, 2002.

[26] K. Todros and J. Tabrikian, "General classes of performance lower bounds for parameter estimation - Part I: Non-Bayesian bounds for unbiased estimators," *IEEE Trans. Inf. Theory*, vol. 56, October 2010.

[27] T. Routtenberg and J. Tabrikian, "Non-Bayesian periodic Cramér-Rao bound," *IEEE Trans. Sig. Proc.*, vol. 61, no. 4, pp. 1019–1032, 2013.

[28] T. Routtenberg and J. Tabrikian, "Cyclic Barankin-type bounds for non-Bayesian periodic parameter estimation," *IEEE Trans. Sig. Proc.*, vol. 62, no. 13, pp. 3321–3336, 2014.

[29] T. Routtenberg and J. Tabrikian, "Cyclic Cramér-Rao-type bounds for periodic parameter estimation," in *2016 19th Int. Conf. on Inf. Fusion (FUSION)*, pp. 1797–1804, 2016.

[30] K. Todros, R. Winik, and J. Tabrikian, "On the limitations of Barankin type bounds for MLE threshold prediction," *Signal Processing*, vol. 108, pp. 622–627, 2015.

[31] P. Stoica and R. L. Moses, "On biased estimators and the unbiased Cramer-Rao lower bound," *Signal Processing*, vol. 21, pp. 349–350, Dec 1990.

[32] H. White, "Consequences and detection of misspecified nonlinear regression models," *Journal of the American Statistical Association*, vol. 76, no. 374, pp. 419–433, 1981.

[33] L. Wasserman, *All of Statistics: A Concise Course in Statistical Inference*. New York: Springer, 2010.

[34] F. Athley, "Threshold region performance of maximum likelihood direction of arrival estimators," *IEEE Trans. Sig. Proc.*, vol. 53, no. 4, pp. 1359–1373, 2005.

[35] C. D. Richmond, "Mean-squared error and threshold SNR prediction of maximum likelihood signal parameter estimation with estimated colored noise covariances," *IEEE Trans. Inf. Theory*, vol. 52, pp. 2146–2164, May 2006.

[36] A. Mallat, S. Gezici, D. Dardari, C. Craeye, and L. Vandendorpe, "Statistics of the MLE and approximate upper and lower bounds - Part I: Application to TOA estimation," *IEEE Trans. Sig. Proc.*, vol. 62, pp. 5663–5676, Nov 2014.

[37] Q. H. Vuong, "Cramer-Rao bounds for misspecified models." `https://resolver.caltech.edu/CaltechAUTHORS:20170823-162930200`, Oct. 1986. Unpublished.

[38] J. Tabrikian and J. Krolik, "Barankin bounds for source localization in an uncertain ocean environment," *IEEE Transactions on Signal Processing*, vol. 47, no. 11, pp. 2917–2927, 1999.

[39] L. Seidman, "Performance limitations and error calculations for parameter estimation," *Proceedings of the IEEE*, vol. 58, no. 5, pp. 644–652, 1970.

[40] D. Chazan, M. Zakai, and J. Ziv, "Improved lower bounds on signal parameter estimation," *IEEE Transactions on Information Theory*, vol. 21, no. 1, pp. 90–93, 1975.

[41] K. K. Bell, Y. Eprahim, and H. L. V. Trees, "Explicit Ziv-Zakai lower bound for bearing estimation," *IEEE Trans. Sig. Proc.*, vol. 44, no. 11, 1996.

[42] A. Mallat, S. Gezici, D. Dardari, C. Craeye, and L. Vandendorpe, "Statistics of the MLE and approximate upper and lower bounds - Part II: Threshold computation and optimal pulse design for TOA estimation," *IEEE Trans. Sig. Proc.*, vol. 62, pp. 5677–5689, Nov 2014.

[43] K. L. Bell, *Performance Bounds in Parameter Estimation with Application to Bearing Estimation*. PhD thesis, George Mason University, 1995.

[44] S. Fortunati, F. Gini, and M. S. Greco, *Parameter bounds under misspecified*

models for adaptive radar detection, vol. 7 of *Academic Press Library in Signal Processing*, ch. 4, pp. 197–252. Academic Press, Nov. 2017.

[45] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*. Prentice-Hall, 1998.

[46] D. Vu, A. Renaux, R. Boyer, and S. Marcos, "A Cramer Rao bounds based analysis of 3D antenna array geometries made from ULA branches," *Multidimensional Systems and Sig. Proc.*, vol. 24, pp. 121–155, 2013.

[47] H. L. V. Trees, *Optimum Array Processing*. Wiley and Sons Inc., 2002.

[48] E. Mehmetcik, U. Orguner, and Ç. Candan, "An approximate MSE expression for maximum likelihood and other implicitly defined estimators of non-random parameters." `https://codeocean.com/capsule/6744831/`, Nov. 2022. Matlab codes.

[49] J. Ziv, "Bounds on the bias of signal parameter estimators," *The Bell System Technical Journal*, vol. 48, no. 6, pp. 2023–2030, 1969.

[50] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. USA: John Wiley and Sons, Inc., 1st ed., 1996.

[51] C. A. Balanis, *Antenna Theory: Analysis and Design*. Wiley-Interscience, 2005.

[52] G. B. Thomas and R. L. Finney, *Calculus and analytic geometry*. Addison-Wesley Publishing Company, 1992.

[53] R. Roy and T. Kailath, "ESPRIT–Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 37, no. 7, pp. 984–995, 1989.

[54] M. S. Paolella, *Linear Models and Time-Series Analysis: Regression, ANOVA, ARMA and GARCH*. John Wiley & Sons, 2018.

[55] K. S. Miller, *Complex Stochastic Processes: An Introduction to Theory and Application*. Addison-Wesley Publishing Company, 1974.

[56] T. Birinci and Y. Tanık, "Optimization of nonuniform array geometry for DOA estimation with the constraint on gross error probability," *Signal Processing*,

vol. 87, no. 10, pp. 2360–2369, 2007. Special Section: Total Least Squares and Errors-in-Variables Modeling.

[57] S. S. Skiena, *The Algorithm Design Manual*. London: Springer, 2008.

[58] M. K. Simon and M. S. Alouni, *Digital Communication over fading channels: A unified approach to performance analysis*. John Wiley and Sons, 2000.

[59] S. Stein, "Unified analysis of certain coherent and noncoherent binary communication systems," *IEEE Trans. Inf. Theory*, vol. IT-10, pp. 43–51, 1964.

APPENDIX A

INFINITE SUPPORT BOUNDED M(·) CASE

In this section we analyze a case when the approximate performance prediction method presented in Chapter 3 does not produce a finite prediction. Consider the estimation problem of $\theta \in \mathbb{R}$ from the following single measurement;

$$\mathrm{x} = \arctan(\bar{\theta}) + w \tag{A.1}$$

where $w \sim \mathcal{N}(0, \sigma_w^2)$. ML estimator for this case is $\hat{\theta}_{\mathrm{ML}} = \tan(\mathrm{x})$, as a result we have,

$$m(\theta) = \arctan(\theta) \tag{A.2}$$

$$\tilde{m}(\bar{\theta} + 2\epsilon; \bar{\theta}) = m(\bar{\theta} + 2\epsilon) - m(\bar{\theta}) \tag{A.3}$$

$$= \arctan(\bar{\theta} + 2\epsilon) - \arctan(\bar{\theta}) \tag{A.4}$$

According to our method;

$$\widehat{\mathrm{MSE}}(\bar{\theta}) = 2 \int_{-\infty}^{\infty} \mathcal{N}_{\mathrm{ccdf}} \left(|\arctan(\bar{\theta} + 2\epsilon) - \arctan(\bar{\theta})|; 0, 4\sigma_w^2 \right) \mathrm{d}\epsilon \tag{A.5}$$

Note that the variance is $4\sigma_w^2$, since the noise term is real (i.e., $w \in \mathbb{R}$). We know that $|\arctan(\theta) \leq \pi/2|$ for all θ. Then,

$$\begin{aligned} |\arctan(\bar{\theta} + 2\epsilon) - \arctan(\bar{\theta})| &\leq |\arctan(\bar{\theta} + 2\epsilon)| + |\arctan(\bar{\theta})| \\ &\leq \frac{\pi}{2} + \frac{\pi}{2} = \pi. \end{aligned} \tag{A.6}$$

Since $\mathcal{N}_{\mathrm{ccdf}}(\cdot)$ is a monotonically decreasing function, we have

$$\mathcal{N}_{\mathrm{ccdf}} \left(\left| \arctan(\bar{\theta} + 2\epsilon) - \arctan(\bar{\theta}) \right|; 0, 2\sigma_w^2 \right) \geq \mathcal{N}_{\mathrm{ccdf}}(\pi; 0, 2\sigma_w^2) > 0. \tag{A.7}$$

Using this in our MSE prediction expression we have,

$$\begin{aligned}\widehat{\mathrm{MSE}}(\bar{\theta}) &\geq 2\int_{-\infty}^{\infty} |\epsilon| \mathcal{N}_{\mathrm{ccdf}}(\pi; 0, 2\sigma_w^2) \mathrm{d}\epsilon \\ &= 2\underbrace{\mathcal{N}_{\mathrm{ccdf}}(\pi; 0, 2\sigma_w^2)}_{\geq 0} \underbrace{\int_{-\infty}^{\infty} |\epsilon| \mathrm{d}\epsilon}_{\to\infty} \to \infty \end{aligned} \tag{A.8}$$

Hence for this problem $\widehat{\mathrm{MSE}} \to \infty$ for all $\bar{\theta}$ and σ_w. This happens because θ has infinite support and the function $|m(\cdot)|$ is bounded. When $m(\cdot)$ is bounded, the probability $P\left(f(\mathrm{x}; \bar{\theta} + 2\epsilon) \geq f(\mathrm{x}; \bar{\theta})\right)$ might be lower bounded by a positive constant for all ϵ and might not tend to zero as $\epsilon \to \infty$, resulting in an infinite MSE prediction.

APPENDIX B

MOMENT GENERATING FUNCTION OF ESTIMATION ERROR

In this section we will derive the moment generating function of the estimation error for the cases where the assumptions of Theorem 1 hold (symmetric and unimodal objective function). Consider the IDE for $\theta \in \mathbb{R}$.

$$\hat{\theta} \triangleq \arg\max_{\theta} \mathcal{L}(\mathbf{x};\theta) \tag{B.1}$$

We have proven that for a symmetric and unimodal $\mathcal{L}(\mathbf{x};\cdot)$, we have

$$\mathrm{P}\left(\hat{\theta}-\bar{\theta} \geq \epsilon\right) = \mathrm{P}\left(\mathcal{L}\left(\mathbf{x};\bar{\theta}+2\epsilon\right) \geq \mathcal{L}\left(\mathbf{x};\bar{\theta}\right)\right) \tag{B.2}$$

$$\mathrm{P}\left(\hat{\theta}-\bar{\theta} \leq -\epsilon\right) = \mathrm{P}\left(\mathcal{L}\left(\mathbf{x};\bar{\theta}-2\epsilon\right) \geq \mathcal{L}\left(\mathbf{x};\bar{\theta}\right)\right) \tag{B.3}$$

Define the estimation error η as

$$\eta \triangleq \hat{\theta}-\bar{\theta}. \tag{B.4}$$

We would like to calculate the moment generating function defined as

$$\mathrm{M}(s) \triangleq \mathbb{E}\left\{\mathrm{e}^{s\eta}\right\}. \tag{B.5}$$

We can see that

$$1-\mathrm{M}(s) = \mathbb{E}\left\{1-\mathrm{e}^{s\eta}\right\} \tag{B.6}$$

$$= \int_{-\infty}^{\infty} \left(1-\mathrm{e}^{s\eta}\right) f_{\eta}(\epsilon) d\epsilon \tag{B.7}$$

$$= \int_{0}^{\infty} \left(1-\mathrm{e}^{s\eta}\right) f_{\eta}(\epsilon) d\epsilon + \int_{-\infty}^{0} \left(1-\mathrm{e}^{s\eta}\right) f_{\eta}(\epsilon) d\epsilon \tag{B.8}$$

$$= \mathrm{I}_{+} + \mathrm{I}_{-} \tag{B.9}$$

We now calculate the integrals on the right hand side above separately. For the first integral I_+, using integration by parts we have

$$u = 1 - \mathrm{e}^{s\eta}, \quad du = -s\mathrm{e}^{s\epsilon} d\epsilon, \tag{B.10}$$

$$dv = f_\eta(\epsilon) d\epsilon, \quad v = -\int_\epsilon^\infty f_\eta(\epsilon') d\epsilon' = -\mathrm{F}_\eta^c(\epsilon). \tag{B.11}$$

Where $\mathrm{F}_\eta(\epsilon)$ is the cumulative distribution function. Then,

$$\begin{aligned}
\mathrm{I}_+ &= \int_0^\infty \left(1 - \mathrm{e}^{s\epsilon}\right) f_\eta(\epsilon) d\epsilon \\
&= -\left(1 - \mathrm{e}^{s\epsilon}\right) \mathrm{F}_\eta^c(\epsilon) \Big|_{\epsilon=0}^\infty - \int_0^\infty \left(-\mathrm{F}_\eta^c(\epsilon)\right) \left(-s\mathrm{e}^{s\epsilon}\right) d\epsilon \\
&= -s \int_0^\infty \mathrm{F}_\eta^c\left(\epsilon\right) \mathrm{e}^{s\epsilon} d\epsilon \\
&= -s \int_0^\infty \mathrm{P}\left(\eta \geq \epsilon\right) \mathrm{e}^{s\epsilon} d\epsilon \\
&= -s \int_0^\infty \mathrm{P}\left(\mathcal{L}\left(\mathbf{x}; \bar{\theta} + 2\epsilon\right) \geq \mathcal{L}\left(\mathbf{x}; \bar{\theta}\right)\right) \mathrm{e}^{s\epsilon} d\epsilon.
\end{aligned} \tag{B.12}$$

Similarly, using integration by parts for the second integral I_- we have,

$$u = 1 - \mathrm{e}^{s\eta}, \quad du = -s\mathrm{e}^{s\epsilon} d\epsilon, \tag{B.13}$$

$$dv = f_\eta(\epsilon) d\epsilon, \quad v = \int_{-\infty}^\epsilon f_\eta(\epsilon') d\epsilon' = \mathrm{F}_\eta(\epsilon). \tag{B.14}$$

Then,

$$\begin{aligned}
\mathrm{I}_- &= \int_{-\infty}^0 \left(1 - \mathrm{e}^{s\epsilon}\right) f_\eta(\epsilon) d\epsilon \\
&= \left(1 - \mathrm{e}^{s\epsilon}\right) \mathrm{F}_\eta(\epsilon) \Big|_{\epsilon \to -\infty}^0 - \int_{-\infty}^0 \mathrm{F}_\eta(\epsilon) \left(-s\mathrm{e}^{s\epsilon}\right) d\epsilon \\
&= s \int_{-\infty}^0 \mathrm{F}_\eta\left(\epsilon\right) \mathrm{e}^{s\epsilon} d\epsilon \\
&= -s \int_\infty^0 \mathrm{F}_\eta\left(-\epsilon\right) \mathrm{e}^{-s\epsilon} d\epsilon \\
&= -s \int_\infty^0 \mathrm{P}\left(\eta \geq -\epsilon\right) \mathrm{e}^{-s\epsilon} d\epsilon \\
&= -s \int_\infty^0 \mathrm{P}\left(\mathcal{L}\left(\mathbf{x}; \bar{\theta} - 2\epsilon\right) \geq \mathcal{L}\left(\mathbf{x}; \bar{\theta}\right)\right) \mathrm{e}^{-s\epsilon} d\epsilon \\
&= s \int_{-\infty}^0 \mathrm{P}\left(\mathcal{L}\left(\mathbf{x}; \bar{\theta} + 2\epsilon\right) \geq \mathcal{L}\left(\mathbf{x}; \bar{\theta}\right)\right) \mathrm{e}^{s\epsilon} d\epsilon.
\end{aligned} \tag{B.15}$$

Combining (B.12) and (B.15) we get

$$\begin{aligned}1-\mathrm{M}(s) =& -s\int_{0}^{\infty}\mathrm{P}\left(\mathcal{L}\left(\mathbf{x};\bar{\theta}+2\epsilon\right)\geq\mathcal{L}\left(\mathbf{x};\bar{\theta}\right)\right)\mathrm{e}^{s\epsilon}d\epsilon\\&+s\int_{-\infty}^{0}\mathrm{P}\left(\mathcal{L}\left(\mathbf{x};\bar{\theta}+2\epsilon\right)\geq\mathcal{L}\left(\mathbf{x};\bar{\theta}\right)\right)\mathrm{e}^{s\epsilon}d\epsilon\\=& -s\int_{-\infty}^{\infty}\mathrm{sign}(\epsilon)\mathrm{P}\left(\mathcal{L}\left(\mathbf{x};\bar{\theta}+2\epsilon\right)\geq\mathcal{L}\left(\mathbf{x};\bar{\theta}\right)\right)\mathrm{e}^{s\epsilon}d\epsilon\end{aligned}\tag{B.16}$$

which gives,

$$\mathrm{M}(s)=1+s\int_{-\infty}^{\infty}\mathrm{sign}(\epsilon)\mathrm{P}\left(\mathcal{L}\left(\mathbf{x};\bar{\theta}+2\epsilon\right)\geq\mathcal{L}\left(\mathbf{x};\bar{\theta}\right)\right)\mathrm{e}^{s\epsilon}d\epsilon.\tag{B.17}$$

For the specific case of ML estimation, we have

$$\mathrm{M}(s)=1+s\int_{-\infty}^{\infty}\mathrm{sign}(\epsilon)\mathrm{P}\left(\frac{f(\mathbf{x};\bar{\theta}+2\epsilon)}{f(\mathbf{x};\bar{\theta})}\geq 1\right)\mathrm{e}^{s\epsilon}d\epsilon.\tag{B.18}$$

We can see that,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\mathrm{M}(s)=&\int_{-\infty}^{\infty}\mathrm{sign}(\epsilon)\mathrm{P}\left(\mathcal{L}\left(\mathbf{x};\bar{\theta}+2\epsilon\right)\geq\mathcal{L}\left(\mathbf{x};\bar{\theta}\right)\right)\mathrm{e}^{s\epsilon}d\epsilon\\&+s\int_{-\infty}^{\infty}|\epsilon|(P)\left(\mathcal{L}(\mathbf{x};\bar{\theta}+2\epsilon)\geq\mathcal{L}(\mathbf{x};\bar{\theta})\right)\mathrm{e}^{s\epsilon}d\epsilon,\end{aligned}\tag{B.19}$$

where we used $\mathrm{sign}(\epsilon)\epsilon=|\epsilon|$. Substituting $s=0$ we get

$$\begin{aligned}\mathbb{E}\{\eta\}=&\frac{\mathrm{d}}{\mathrm{d}s}\mathrm{M}(s)\Big|_{s=0}\\=&\int_{-\infty}^{\infty}\mathrm{sign}(\epsilon)\mathrm{P}\left(\mathcal{L}(\mathbf{x};\bar{\theta}+2\epsilon)\geq\mathcal{L}(\mathbf{x};\bar{\theta})\right)\mathrm{d}\epsilon\end{aligned}\tag{B.20}$$

which is the expression we found for the bias. We can calculate the 2nd order derivative of $\mathrm{M}(s)$ w.r.t. s as follows.

$$\begin{aligned}\frac{\mathrm{d}^2}{\mathrm{d}s^2}\mathrm{M}(s)=&\int_{-\infty}^{\infty}|\epsilon|\mathrm{P}\left(\mathcal{L}\left(\mathbf{x};\bar{\theta}+2\epsilon\right)\geq\mathcal{L}\left(\mathbf{x};\bar{\theta}\right)\right)\mathrm{e}^{s\epsilon}d\epsilon\\&+\int_{-\infty}^{\infty}|\epsilon|\mathrm{P}\left(\mathcal{L}\left(\mathbf{x};\bar{\theta}+2\epsilon\right)\geq\mathcal{L}\left(\mathbf{x};\bar{\theta}\right)\right)\mathrm{e}^{s\epsilon}d\epsilon\\&+s\int_{-\infty}^{\infty}\epsilon|\epsilon|\mathrm{P}\left(\mathcal{L}\left(\mathbf{x};\bar{\theta}+2\epsilon\right)\geq\mathcal{L}\left(\mathbf{x};\bar{\theta}\right)\right)\mathrm{e}^{s\epsilon}d\epsilon.\end{aligned}\tag{B.21}$$

Substituting $s=0$ we get

$$\begin{aligned}\mathbb{E}\{\eta^2\}=&\frac{\mathrm{d}^2}{\mathrm{d}s^2}\mathrm{M}(s)\Big|_{s=0}\\=&2\int_{-\infty}^{\infty}|\epsilon|\mathrm{P}\left(\mathcal{L}(\mathbf{x};\bar{\theta}+2\epsilon)\geq\mathcal{L}(\mathbf{x};\bar{\theta})\right)\mathrm{d}\epsilon,\end{aligned}\tag{B.22}$$

which is the expression we found for the MSE.

APPENDIX C

INCONSISTENCY OF TAYLOR EXPANSION BASED METHODS

In this section, we consider the prediction of the MSE of the maximum likelihood (ML) estimate (MLE) using the method presented in [2], which we will refer to as So's approach.

We consider the ML estimation of a real scalar parameter $\theta \in \mathbb{R}$, with the true value $\bar{\theta}$, using the measurement likelihood denoted as $f(\mathbf{x};\theta)$ where $\mathbf{x} \in \mathbb{C}^N$ is the measurement vector. The ML estimate can be posed in two different ways:

$$\hat{\theta}_{\text{ML-1}} = \arg\max_{\theta} \ln f(\mathbf{x};\theta), \tag{C.1}$$

and

$$\hat{\theta}_{\text{ML-2}} = \arg\max_{\theta} f(\mathbf{x};\theta), \tag{C.2}$$

which are obviously equivalent, i.e., $\hat{\theta}_{\text{ML-1}} = \hat{\theta}_{\text{ML-2}}$, for all $\mathbf{x}$, since the natural logarithm function i.e., $\ln(\cdot)$, is a monotonically increasing function. The approximate MSE prediction expression in So's approach can be summarized (for the case of a real scalar parameter) as follows [2].

$$\widehat{\text{MSE}}_{\text{So}}(\bar{\theta}) \triangleq \frac{\mathbb{E}\left\{\left(\frac{\partial}{\partial\theta}\mathcal{L}(\mathbf{x};\bar{\theta})\right)^2\right\}}{\left(\mathbb{E}\left\{\frac{\partial^2}{\partial\theta^2}\mathcal{L}(\mathbf{x};\bar{\theta})\right\}\right)^2}, \tag{C.3}$$

where $\mathcal{L}(\mathbf{x};\theta)$ is the objective function to be maximized and all expectations are taken over the random vector $\mathbf{x}$. Note that in the estimation problems (C.1) and (C.2), the objective functions are $\mathcal{L}_1(\mathbf{x};\theta) \triangleq \ln f(\mathbf{x};\theta)$ and $\mathcal{L}_2(\mathbf{x};\theta) \triangleq f(\mathbf{x};\theta)$, respectively.

If the objective function $\mathcal{L}_1(\mathbf{x};\theta) \triangleq \ln f(\mathbf{x};\theta)$ is used in ML, So's approach gives the following MSE prediction.

$$\widehat{\text{MSE}}_{\text{So-1}}(\bar{\theta}) = \frac{\mathbb{E}\left\{\left(\frac{\partial}{\partial\theta}\ln f(\mathbf{x};\bar{\theta})\right)^2\right\}}{\left(\mathbb{E}\left\{\frac{\partial^2}{\partial\theta^2}\ln f(\mathbf{x};\bar{\theta})\right\}\right)^2}. \tag{C.4}$$

Note that the CRLB for this problem is given as

$$\mathrm{CRLB}(\bar{\theta}) = \frac{-1}{\mathbb{E}\left\{\frac{\partial^2}{\partial\theta^2}\ln f(\mathbf{x};\bar{\theta})\right\}} = \frac{1}{\mathbb{E}\left\{\left(\frac{\partial}{\partial\theta}\ln f(\mathbf{x};\bar{\theta})\right)^2\right\}}. \tag{C.5}$$

Consequently, the MSE expression in (C.4) can be written as

$$\begin{aligned}\widehat{\mathrm{MSE}}_{\text{So-1}}(\bar{\theta}) &= \frac{\mathbb{E}\left\{\left(\frac{\partial}{\partial\theta}\ln f(\mathbf{x};\bar{\theta})\right)^2\right\}}{\left(\mathbb{E}\left\{\frac{\partial^2}{\partial\theta^2}\ln f(\mathbf{x};\bar{\theta})\right\}\right)^2} \\ &= \frac{\mathbb{E}\left\{\left(\frac{\partial}{\partial\theta}\ln f(\mathbf{x};\bar{\theta})\right)^2\right\}}{\left(\mathbb{E}\left\{\left(\frac{\partial}{\partial\theta}\ln f(\mathbf{x};\bar{\theta})\right)^2\right\}\right)^2} \\ &= \frac{1}{\mathbb{E}\left\{\left(\frac{\partial}{\partial\theta}\ln f(\mathbf{x};\bar{\theta})\right)^2\right\}} = \mathrm{CRLB}(\bar{\theta}).\end{aligned} \tag{C.6}$$

As a result, So's approach yields CRLB as the predicted MSE for (C.1). We are now going to examine So's approach for (C.2).

If the objective function $\mathcal{L}_2(\mathbf{x};\theta) \triangleq f(\mathbf{x};\theta)$ is used in ML, So's approach gives the following MSE prediction.

$$\widehat{\mathrm{MSE}}_{\text{So-2}}(\bar{\theta}) = \frac{\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}f(\mathbf{x};\bar{\theta})\right)^2\right]}{\left(\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}f(\mathbf{x};\bar{\theta})\right]\right)^2}. \tag{C.7}$$

Unfortunately, in general, the MSE expression in (C.7) is not equal to that in (C.4). We are going to prove this considering a very simple estimation problem. Consider the estimation problem for a real scalar $\theta \in \mathbb{R}$ measured under Gaussian noise, i.e., we consider the model

$$x = \theta + \nu, \quad \nu \sim \mathcal{N}(0, \sigma^2), \tag{C.8}$$

which has the likelihood function given as

$$f(x;\theta) = \mathcal{N}(x;\theta,\sigma^2). \tag{C.9}$$

$\widehat{\mathrm{MSE}}_{\text{So-1}}(\bar{\theta})$ for this problem is given as $\widehat{\mathrm{MSE}}_{\text{So-1}}(\bar{\theta}) = \mathrm{CRLB}(\bar{\theta}) = \sigma^2$. In order to calculate $\widehat{\mathrm{MSE}}_{\text{So-2}}$, we can derive the following identities for a univariate Gaussian distribution.

$$\frac{\partial}{\partial\theta}\mathcal{N}(x;\theta,\sigma^2) = \frac{(x-\theta)}{\sigma^2}\mathcal{N}(x;\theta,\sigma^2), \tag{C.10a}$$

$$\frac{\partial^2}{\partial\theta^2}\mathcal{N}(x;\theta,\sigma^2) = \left(\frac{(x-\theta)^2}{\sigma^4} - \frac{1}{\sigma^2}\right)\mathcal{N}(x;\theta,\sigma^2). \tag{C.10b}$$

The numerator in (C.7) is then given as

$$\mathbb{E}\left\{\left(\frac{\partial}{\partial\theta}\mathcal{N}(x;\bar{\theta},\sigma^2)\right)^2\right\} = \int \frac{(x-\bar{\theta})^2}{\sigma^4}\mathcal{N}^2(x;\bar{\theta},\sigma^2)\mathcal{N}(x;\bar{\theta},\sigma^2)\mathrm{d}x. \tag{C.11}$$

Noting that we have

$$\mathcal{N}^2(x;\bar{\theta},\sigma^2) = \mathcal{N}(\bar{\theta};\bar{\theta},2\sigma^2)\mathcal{N}\left(x;\bar{\theta},\frac{\sigma^2}{2}\right) = \frac{1}{\sqrt{4\pi\sigma^2}}\mathcal{N}\left(x;\bar{\theta},\frac{\sigma^2}{2}\right), \tag{C.12}$$

we can write (C.11) as

$$\mathbb{E}\left\{\left(\frac{\partial}{\partial\theta}\mathcal{N}(x;\bar{\theta},\sigma^2)\right)^2\right\} = \frac{1}{\sigma^4\sqrt{4\pi\sigma^2}}\int (x-\bar{\theta})^2\mathcal{N}\left(x;\bar{\theta},\frac{\sigma^2}{2}\right)\mathcal{N}\left(x;\bar{\theta},\sigma^2\right)\mathrm{d}x. \tag{C.13}$$

We can simplify the integrand in (C.13) as

$$\mathcal{N}\left(x;\bar{\theta},\frac{\sigma^2}{2}\right)\mathcal{N}\left(x;\bar{\theta},\sigma^2\right) = \mathcal{N}\left(\bar{\theta};\bar{\theta},\frac{3\sigma^2}{2}\right)\mathcal{N}\left(x;\bar{\theta},\frac{\sigma^2}{3}\right) = \frac{1}{\sqrt{3\pi\sigma^2}}\mathcal{N}\left(x;\bar{\theta},\frac{\sigma^2}{3}\right). \tag{C.14}$$

Hence we can write (C.13) as

$$\mathbb{E}\left\{\left(\frac{\partial}{\partial\theta}\mathcal{N}(x;\bar{\theta},\sigma^2)\right)^2\right\} = \frac{1}{\sigma^4\sqrt{4\pi\sigma^2}}\int (x-\bar{\theta})^2\frac{1}{\sqrt{3\pi\sigma^2}}\mathcal{N}\left(x;\bar{\theta},\frac{\sigma^2}{3}\right)\mathrm{d}x \tag{C.15a}$$

$$= \frac{1}{\sigma^4\sqrt{4\pi\sigma^2}\sqrt{3\pi\sigma^2}}\frac{\sigma^2}{3} \tag{C.15b}$$

$$= \boxed{\frac{1}{3\pi\sigma^4\sqrt{12}}.} \tag{C.15c}$$

For the denominator in (C.7), we can write

$$\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\mathcal{N}(x;\bar{\theta},\sigma^2)\right] = \int\left(\frac{(x-\bar{\theta})^2}{\sigma^4} - \frac{1}{\sigma^2}\right)\mathcal{N}^2(x;\theta,\sigma^2)\mathrm{dx} \tag{C.16a}$$

$$= \frac{1}{\sigma^4}\int (x-\bar{\theta})^2\frac{1}{\sqrt{4\pi\sigma^2}}\mathcal{N}\left(x;\bar{\theta},\frac{\sigma^2}{2}\right)\mathrm{d}x$$

$$- \frac{1}{\sigma^2}\int \frac{1}{\sqrt{4\pi\sigma^2}}\mathcal{N}\left(x;\bar{\theta},\frac{\sigma^2}{2}\right)\mathrm{d}x \tag{C.16b}$$

$$= \frac{\sigma^2}{2\sigma^4\sqrt{4\pi\sigma^2}} - \frac{1}{\sigma^2\sqrt{4\pi\sigma^2}} \tag{C.16c}$$

$$= \boxed{-\frac{1}{2\sigma^2\sqrt{4\pi\sigma^2}}.} \tag{C.16d}$$

Using (C.15c) and (C.16d), we can obtain $\widehat{\mathrm{MSE}}_{\text{So-2}}$ in (C.7) as,

$$\begin{aligned}
\widehat{\mathrm{MSE}}_{\text{So-2}}(\bar{\theta}) &= \frac{\mathbb{E}\left\{x\left(\frac{\partial}{\partial\theta}\mathcal{N}(x;\bar{\theta},\sigma^2)\right)^2\right\}}{\mathbb{E}\left\{\frac{\partial^2}{\partial\theta^2}\mathcal{N}(x;\bar{\theta},\sigma^2)\right\}^2} \\
&= \frac{\frac{1}{3\pi\sigma^4\sqrt{12}}}{\frac{1}{4\sigma^4 4\pi\sigma^2}} \\
&= \frac{8}{3\sqrt{3}}\sigma^2 \\
&\neq \sigma^2 = \widehat{\mathrm{MSE}}_{\text{So-1}}(\bar{\theta}).
\end{aligned} \tag{C.17}$$

As a result, So's approach might not give the same MSE predictions for monotone transformations of the cost function. In other words, different MSE predictions, e.g., $\widehat{\mathrm{MSE}}_{\text{So-1}}(\bar{\theta}) \neq \widehat{\mathrm{MSE}}_{\text{So-2}}(\bar{\theta})$ can be obtained for identical estimators, e.g., $\hat{\theta}_{\text{ML-1}} = \hat{\theta}_{\text{ML-2}}$. Using a similar methodology, we can show that Fessler's approach [1], which is based on Taylor series approximations like So's approach, has a similar problem for some estimation problems. The proposed approach in this thesis study (in Chapter 3) does not have this problem since the event

$$f(\mathrm{x};\bar{\theta}+2\epsilon) \geq f(\mathrm{x};\bar{\theta}) \tag{C.18}$$

is the same as the event

$$\ln f(\mathrm{x};\bar{\theta}+2\epsilon) \geq \ln f(\mathrm{x};\bar{\theta}) \tag{C.19}$$

for all $\epsilon \in \mathbb{R}$ and hence their probabilities are the same. Similarly, any monotone function of the objective function would not change these events, keeping the MSE predictions identical.

APPENDIX D

MAXIMUM LIKELIHOOD ESTIMATOR DERIVATION UNDER RICIAN FADING

Consider the DOA estimation problem under additive white Gaussian noise and Rician fading conditions, with the following signal model,

$$\mathbf{x} = \beta \mathbf{a}_\theta + \mathbf{v} \tag{D.1}$$

where $\beta \sim \mathcal{CN}(\mu_F, \sigma_F^2)$ is the fading coefficient, $\mu_f \in \mathbb{R}$ is the fading mean and σ_F^2 is the fading variance. $\mathbf{a}_\theta \in \mathbb{C}^L, \|\mathbf{a}_\theta\| = 1, \forall\theta$ is the array manifold vector (of an L-element array). Lastly $\mathbf{v} \sim \mathcal{CN}(0, \sigma_v^2\mathbf{I})$ is the additive white Gaussian noise component.

Note that μ_F can be real or complex in general. We assume a real fading mean component without any loss of generality, as the noise component is circularly symmetric complex Gaussian, and normalizing the overall observation vector with $\mu_F^*/|\mu_F|$ would yield the same formulation.

Note that the observation vector is still a complex Gaussian vector,

$$\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}_\mathbf{x}, \mathbf{C}_\mathbf{x}), \quad \boldsymbol{\mu}_\mathbf{x} = \mu_F \mathbf{a}_\theta, \quad \mathbf{C}_\mathbf{x} = \sigma_F^2 \mathbf{a}_\theta \mathbf{a}_\theta^{\mathrm{H}} + \sigma_v^2 \mathbf{I} \tag{D.2}$$

$$f_\mathbf{X}(\mathbf{x}) = \frac{1}{\pi^L |\mathbf{C}_\mathbf{x}|} \exp\left(-(\mathbf{x} - \boldsymbol{\mu}_\mathbf{x})^{\mathrm{H}} \mathbf{C}_\mathbf{x}^{-1} (\mathbf{x} - \boldsymbol{\mu}_\mathbf{x})\right). \tag{D.3}$$

Consequently, the log-likelihood function $\Lambda(\theta)$ can be written as follows,

$$\begin{aligned} \Lambda(\theta) &= \ln f_\mathbf{X}(\mathbf{x}; \theta) \\ &= \ln \pi^L - \ln |\mathbf{C}_\mathbf{x}| - (\mathbf{x} - \boldsymbol{\mu}_\mathbf{x})^{\mathrm{H}} \mathbf{C}_\mathbf{x}^{-1} (\mathbf{x} - \boldsymbol{\mu}_\mathbf{x}) \end{aligned} \tag{D.4}$$

The Maximum Likelihood estimator has the following definition;

$$\hat{\theta}_{\mathrm{ML}} \triangleq \arg\max_\theta \Lambda(\theta) = \mathrm{argmin}_\theta -\Lambda(\theta). \tag{D.5}$$

And for the problem under consideration we have,

$$\hat{\theta}_{\mathrm{ML}} = \mathrm{argmin}_{\theta} \ln |\mathbf{C}_{\mathbf{x}}| + (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})^{\mathrm{H}} \mathbf{C}_{\mathbf{x}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \tag{D.6}$$

Note that we omitted the constant term $\ln \pi^L$, as it does not affect minimization. Note that the determinant covariance matrix of $\mathbf{x}$ can be simplified as,

$$\begin{aligned}
|\mathbf{C}_{\mathbf{x}}| &= \left|\sigma_v^2 \mathbf{I} + \sigma_F \mathbf{a}_\theta \mathbf{a}_\theta^{\mathrm{H}}\right| \\
&= \left|\sigma_v^2 \left(\mathbf{I} + \frac{\sigma_F^2}{\sigma_v^2} \mathbf{a}_\theta \mathbf{a}_\theta^{\mathrm{H}}\right)\right| \\
&= (\sigma_v^2)^L \left|\left(\mathbf{I} + \frac{\sigma_F^2}{\sigma_v^2} \mathbf{a}_\theta \mathbf{a}_\theta^{\mathrm{H}}\right)\right| \\
&= (\sigma_v^2)^L \left|\left(1 + \frac{\sigma_F^2}{\sigma_v^2} \mathbf{a}_\theta^{\mathrm{H}} \mathbf{a}_\theta\right)\right| \\
&= (\sigma_v^2)^L \left(1 + \frac{\sigma_F^2}{\sigma_v^2}\right) \\
&= (\sigma_v^2)^L + \sigma_F^2 (\sigma_v^2)^{L-1}.
\end{aligned} \tag{D.7}$$

Note that in the above derivations we used the identity $|\mathbf{I} + \mathbf{u}\mathbf{v}^{\mathrm{H}}| = 1 + \mathbf{u}^{\mathrm{H}}\mathbf{v}$, see [50]. Note that the determinant term is independent of θ and can be omitted from minimization considerations. Using a similar derivation for $\mathbf{C}_{\mathbf{x}}^{-1}$, we can arrive at the following expression,

$$\mathbf{C}_{\mathbf{x}}^{-1} = \frac{1}{\sigma_v^2}\mathbf{I} - \frac{\sigma_F^2}{\sigma_v^2(\sigma_v^2 + \sigma_F^2)} \mathbf{a}_\theta \mathbf{a}_\theta^{\mathrm{H}} = \frac{1}{\sigma_v^2}\mathbf{I} - \gamma \mathbf{a}_\theta \mathbf{a}_\theta^{\mathrm{H}} \tag{D.8}$$

where we defined $\gamma \triangleq \frac{\sigma_F^2}{\sigma_v^2(\sigma_v^2+\sigma_F^2)}$. As a result the quadratic term can be simplified as follows,

$$\begin{aligned}
\hat{\theta}_{\mathrm{ML}} &= \mathrm{argmin}_{\theta} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})^{\mathrm{H}} \mathbf{C}_{\mathbf{x}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \\
&= \mathrm{argmin}_{\theta} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})^{\mathrm{H}} \frac{1}{\sigma_v^2} \mathbf{I} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) - \gamma |\mathbf{a}_\theta^{\mathrm{H}} (\mathbf{x} - \mu_{\mathbf{x}})|^2 \\
&= \mathrm{argmin}_{\theta} \frac{1}{\sigma_v^2} \|\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})\|^2 - \gamma |\mathbf{a}_\theta^{\mathrm{H}} (\mathbf{x} - \mu_{\mathbf{x}})|^2 \\
&= \mathrm{argmin}_{\theta} \frac{1}{\sigma_v^2} \left(\|\mathbf{x}\|^2 + \|\mu_{\mathbf{x}}\|^2 - 2\Re\{\mu_{\mathbf{x}}^{\mathrm{H}} \mathbf{x}\}\right) - \gamma |\mathbf{a}_\theta^{\mathrm{H}} (\mathbf{x} - \mu_{\mathbf{x}})|^2 \\
&= \mathrm{argmin}_{\theta} \frac{1}{\sigma_v^2} \left(-2\mu_F \Re\{\mathbf{a}_\theta^{\mathrm{H}} \mathbf{x}\}\right) - \gamma |\mathbf{a}_\theta^{\mathrm{H}} \mathbf{x} - \mu_F|^2 \\
&= \mathrm{argmin}_{\theta} -2\mu_F \Re\{\mathbf{a}_\theta^{\mathrm{H}} \mathbf{x}\} \left(\frac{1}{\sigma_v^2} - \gamma\right) - \gamma |\mathbf{a}_\theta^{\mathrm{H}} \mathbf{x}|^2 \\
&= \mathrm{argmin}_{\theta} -\frac{2\mu_F}{\sigma_v^2 + \sigma_F^2} \Re\{\mathbf{a}_\theta^{\mathrm{H}} \mathbf{x}\} - \frac{\sigma_F^2}{\sigma_v^2(\sigma_v^2 + \sigma_F^2)} |\mathbf{a}_\theta^{\mathrm{H}} \mathbf{x}|^2.
\end{aligned}$$

Getting rid of the common multipliers we can arrive at the following result,

$$\hat{\theta}_{\mathrm{ML}} = \arg\max_{\theta} 2\mu_F \Re\{\mathbf{a}_\theta^{\mathrm{H}}\mathbf{x}\} + \frac{\sigma_F^2}{\sigma_v^2}|\mathbf{a}_\theta^{\mathrm{H}}\mathbf{x}|^2. \tag{D.9}$$

It is interesting to see that the resulting estimator is a weighted sum of coherent ($\Re\{\mathbf{a}_\theta^{\mathrm{H}}\mathbf{x}\}$) and non-coherent estimators ($|\mathbf{a}_\theta^{\mathrm{H}}\mathbf{x}|^2$). This is similar to combining two noisy measurements of a parameter with different variances. To obtain an estimate with a lower variance, their weights should be proportional to the information each measurement carries. In this case, the overall estimator tends to lean towards the coherent estimator when the direct path signal is strong, hence its coefficient is proportional to μ_F. And similarly, the non-coherent estimator has a weight that is proportional to the power of multi-path signal component, σ_F^2.

APPENDIX E

STEIN'S UNIFIED ANALYSIS OF THE ERROR PROBABILITY

Consider two correlated complex Gaussian random variables, $z_{1f} \sim \mathcal{CN}(\bar{z}_{1f}, N_{1f})$ and $z_{2f} \sim \mathcal{CN}(\bar{z}_{2f}, N_{2f})$. The probability $P(|z_{1f}| \leq |z_{2f}|)$ can be calculated using Stein's unified analysis of the error probability [58, 59] with the following equation:

$$\begin{aligned}\mathrm{P}(|z_{1f}| < |z_{2f}|) =& \frac{1}{2}\left[1 - Q_1(\sqrt{b}, \sqrt{a}) + Q_1(\sqrt{a}, \sqrt{b})\right] \\ &- \frac{A}{2}\exp\left(-\frac{a+b}{2}\right) I_0(\sqrt{ab})\end{aligned} \tag{E.1}$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind of order zero, and $Q_1(\cdot)$ is the first order Marcum Q-function, and we have used the following definitions,

$$\bar{z}_{if} = m_{if} + j\mu_{if} = |\bar{z}_{if}|e^{j\theta_{if}}, i = 1, 2. \tag{E.2}$$

$$S_{if} = \frac{1}{2}|\bar{z}_{if}|^2, \tag{E.3}$$

$$N_{if} = \frac{1}{2}\mathbb{E}\left\{|z_{if} - \bar{z}_{if}|^2\right\}, \tag{E.4}$$

$$\rho_f = \frac{1}{2\sqrt{N_{1f}N_{2f}}}\mathbb{E}\left\{(z_{1f} - \bar{z}_{1f})^*(z_{2f} - \bar{z}_{2f})\right\}, \tag{E.5}$$

$$\phi = \arg(\rho_{cf} + j\rho_{sf}), \tag{E.6}$$

$$\begin{aligned}\left\{\begin{array}{c} a \\ b \end{array}\right\} =& \frac{1}{2}\frac{S_{1f} + S_{2f} + 2\sqrt{S_{1f}S_{2f}}\cos(\theta_{1f} - \theta_{2f} + \phi)}{N_{1f} + N_{2f} + 2\sqrt{N_{1f}N_{2f}|\rho_f|^2}} \\ &+ \frac{1}{2}\frac{S_{1f} + S_{2f} - 2\sqrt{S_{1f}S_{2f}}\cos(\theta_{1f} - \theta_{2f} + \phi)}{N_{1f} + N_{2f} - 2\sqrt{N_{1f}N_{2f}|\rho_f|^2}} \\ &\mp \frac{1}{2}\frac{2(S_{1f} - S_{2f})}{\sqrt{(N_{1f} + N_{2f})^2 - 4N_{1f}N_{2f}|\rho_f|^2}},\end{aligned} \tag{E.7}$$

$$A = \frac{N_{1f} - N_{2f}}{\sqrt{(N_{1f} + N_{2f})^2 - 4N_{1f}N_{2f}|\rho_f|^2}}. \tag{E.8}$$

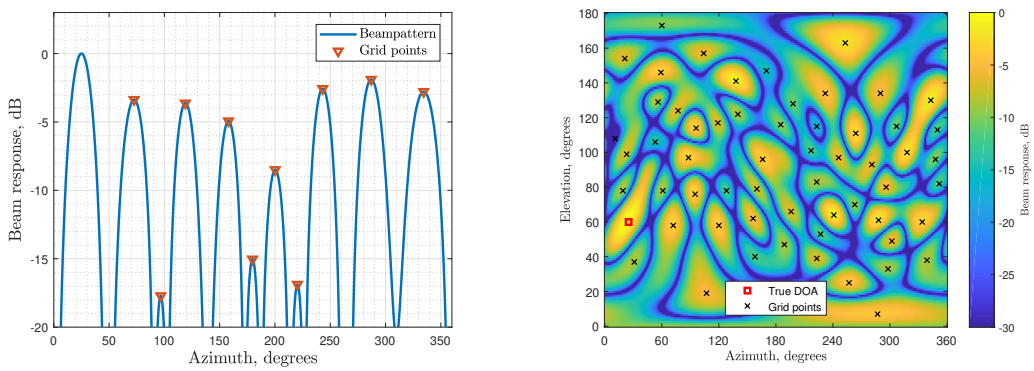
APPENDIX F

IMPLEMENTATION DETAILS OF THE METHODS USED IN SECTION 3.6 AND SECTION 4.2

F.1 Implementation Details for Section 3.6.1

- ML estimate is calculated on a grid. The grid consists of 3600 uniformly spaced points in the interval $[-\pi,\, \pi]$ for (scalar) azimuth estimation, and 3600 uniformly spaced points in the interval $[0,\, \pi]$ for (scalar) elevation estimation. For the case of nuisance parameter, azimuth and elevation angles are selected such that the corresponding unit vectors have an approximately uniform distribution over the unit sphere. To do so, elevation interval $[0,\, \pi]$ is divided into 200 equally spaced values $\{\theta_k\}_{k=1}^{200}$, and each constant elevation circle on the unit sphere is divided into $\lceil (100 \sin(\theta_k) \rceil$ points where $\lceil \cdot \rceil$ denotes the ceiling function. By doing so, a total of 198192 grid points (ϕ-θ pairs) are obtained.

  BBs [24] are calculated with a single test point which is optimized on a grid. The single and multiple parameter BB grids are selected the same as the grids used for ML. MIE grid points are selected as the local maxima outside the mainlobe in the beampattern for the scalar parameter estimation case, which requires a peak finding algorithm over the 1D beampattern function. For the case of a nuisance parameter, the beampattern function is a 2D surface, and grid points are the local maxima on this surface. For both cases the beampattern function is calculated over the ML grid, and local maxima are found using Matlab built-in functions (`findpeaks(·)` for maxima on the 1D curve, `imregionalmax(·)` for maxima on the 2D surface). For both cases, grid points are illustrated in Figure F.1.



(a) Known elevation (scalar parameter)

(b) Unknown elevation (Nuisance parameter)

Figure F.1: Grid points of the MIE for azimuth estimation.

Proposed method does not require grid point selection for the scalar parameter case. For the case of a nuisance parameter, grid points are selected over the entire support of the nuisance parameter with logarithmic spacing around the true value, making the grid denser as the grid points approach the true parameter value. In Matlab notation, the grid is defined as follows

$$\boldsymbol{\delta\theta} = \texttt{logspace}(-7, \log_{10} e_{\text{max}}, 60), \tag{F.1a}$$

$$\boldsymbol{\Theta}_{\setminus 1} = [\bar{\theta},\, \bar{\theta} - \boldsymbol{\delta\theta},\, \bar{\theta} + \boldsymbol{\delta\theta}]. \tag{F.1b}$$

The statement (F.1a) generates 60 logarithmically spaced points within the interval $[10^{-7}, e_{\text{max}}]$. We selected $e_{\text{max}} = \pi/2$ for the elevation angle as the nuisance parameter and $e_{\text{max}} = \pi$ for the azimuth angle as the nuisance parameter.

F.2 Implementation Details for Section 3.6.2

- ML grid consists of 3600 uniformly spaced points in the interval $[-\pi,\, \pi]$ for (scalar) azimuth estimation. BB grid is the same as the ML grid. Other methods do not require grid points.

F.3 Implementation Details for Section 4.2

- BB grid is the same as the ML grid for the problem in Section 3.6.2. Other methods do not require grid points.

The numerical integrals of the proposed method in all subsections up to this point are taken using the Matlab function `integral(·)` with the following tolerance values: `AbsTol=1e-5`, `RelTol=1e-5`. The integration limits for the numerical integrals of the proposed method were set as $\epsilon \in \left[\frac{-\pi-\bar{\phi}}{2}, \frac{\pi-\bar{\phi}}{2}\right]$ for azimuth estimation and $\epsilon \in \left[-\frac{\bar{\theta}}{2}, \frac{\pi-\bar{\theta}}{2}\right]$ for elevation estimation.

F.4 Implementation Details for Section 3.6.3

- ML and MAP use the same grid, which consists of 8192 uniformly spaced points over $[-\pi, \pi]$ for $\omega = \pi\cos(\phi)$. The (non-uniform) grid points for ϕ are calculated using expression $\arccos(\frac{\omega}{\pi})$ from the grid for ω.
- The double integral for ZZB is taken using the numerical integration function of Matlab `integral2(·)` with the default tolerance settings.
- The numerical integrals of $\widehat{\mathrm{MSE}}_{\mathrm{ML}}(\phi)$ and $\widehat{\mathrm{MSE}}_{\mathrm{MAP}}(\phi)$ are taken using the Matlab function `integral(·)` with the tolerance values `AbsTol=1e-18`, `RelTol=1e-12`. The integrals with respect to the prior are calculated over a uniform grid over the inverval $[0, \pi]$ with the grid spacing 0.01.

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name	**:**	Mehmetcik, Erdal
e-mail	**:**	erdalmehmetcik@gmail.com

EDUCATION

Degree	**Institution**	**Graduation**
M.S.	Electrical and Electronics Engineering, METU	2011
B.S.	Electrical and Electronics Engineering, METU	2009
High School	Bursa Anatolian High School	2005

PROFESSIONAL EXPERIENCE

Year	**Place**	**Enrollment**
2009-2010	METU EEE Signal Processing Lab	Research Assistant
2010-2011	METU EEE	Teaching Assistant
2011-...	ASELSAN	Systems Engineer

PUBLICATIONS

Master Thesis

E. Mehmetcik, **"Speech enhancement utilizing phase continuity between consecutive analysis windows"**, Electrical and Electronics Engineering, Middle East Technical University, 2011.

Patents

- E. Mehmetcik, S. Yilmaz, S. Erdogdu, E. Topcu, T. Ciloglu, C. Çandan, U. Orguner, O. Sipahigil, E. Epcacan, **A method for tow-ship noise suppression for triplet arrays**, TPE 2015/12762, 2015.

Journal Publications

- E. Mehmetcik and U. Orguner and C. Çandan, **An Approximate MSE Expression for Maximum Likelihood and Other Implicitly Defined Estimators of Non-Random Parameters**, Signal Processing, vol. 204, p.108838, 2023.

International Conference Publications

- E. Mehmetcik, T. Çiloğlu, **Speech enhancement by maintaining phase continuity**, in Proc. Meetings Acoust. Soc. Amer., vol. 18, no. 055002, Nov. 2012.
- E. Mehmetcik, U. Orguner. **Centralized target tracking using propagation delayed measurements**. In Proceedings of 16th International Conference on Information Fusion, Istanbul, Turkey, July 2013.
- E. Mehmetcik, C. Çandan, **Prediction and representation of array performance under sensor failure**, Underwater Defence Technology Europe, Stockholm, Sweden, May 2019.

National Conference Publications

- E. Mehmetcik and T. Çiloğlu. **Faz sürekliliğini sağlayarak konuşma iyileştirme**. 20. Sinyal İşleme ve Haberleşme Uygulamaları Konferansı (SIU), pp. 1-4, IEEE, 2012.
- E. Mehmetcik, E. Topçu, and M. N. Şenlik. **Akustik transdüserler için denkleştirme filtresi uygulaması**. 22. Sinyal İşleme ve Haberleşme Uygulamaları Konferansı (SIU), p. 1339-1342. IEEE, 2014.