TEL AVIV UNIVERSITY אוניברסיטת תל-אביב
FACULTY OF ENGINEERING הפאקולטה להנדסה

THE IBY AND ALADAR FLEISCHMAN FACULTY OF ENGINEERING

Practical performance of the Maximum Likelihood Estimator for problems characterized by zero information points

A thesis submitted toward the degree of
Master of Science in Electrical and Electronics Engineering

by
Bashan Eran

December, 2003

TEL AVIV UNIVERSITY אוניברסיטת תל-אביב
FACULTY OF ENGINEERING הפאקולטה להנדסה

THE IBY AND ALADAR FLEISCHMAN FACULTY OF ENGINEERING

Practical performance of the Maximum Likelihood Estimator for problems characterized by zero information points

A thesis submitted toward the degree of
Master of Science in Electrical and Electronics Engineering

by

Bashan Eran

This research was carried out in the
Department of Electrical Engineering - Systems,
under the supervision of Prof. Anthony J. Weiss

December, 2003

Acknowledgments

I wish to thank professor Anthony J. Weiss for the professional guidance, for many helpful discussions and comments, and for his availability all along this research.
I would also like to thank professor Isaac Meilijson for his contribution to this work in several sessions, and to professor Alexander Apelblat for his help in deriving some of the mathematical results presented in the appendix.

To Iris, for her patience, support and understanding.

Abstract

In this thesis we study issues related to estimators behavior. We examine the general scalar measurement equation $y_n = h(\theta) + v_n$, from which we wish to estimate the parameter θ. Specifically, we concentrate on problems where θ is a continuous parameter and the Fisher Information Measure (FIM) equal zero at isolated points θ_i. The well-known *Cramér-Rao Lower Bound* (CRLB) on the variance of any unbiased estimator is the inverse of the FIM, hence we define the points where the FIM equal zero as singular points of the CRLB. These conditions are common in many practical applications. One such known case is the problem of Direction Of Arrival (DOA) estimation using a linear array of sensors. In the DOA problem the singularity occurs when the transmitter is located at the end-fire of the array (aligned with the array direction). Our target is to derive the statistical description of the Maximum Likelihood Estimator for the general case of singular points.

Maximum Likelihood Estimators (MLE) are very popular since in many cases they are relatively easy to derive, and under regularity conditions (see section 2.2.1) known to be asymptotically unbiased and efficient. These properties led to the common belief that the MLE is useless in the vicinity of singular points of the CRLB. In this thesis we show that the MLE, for this problem, is useful over the entire parameter space. Moreover, we derive the estimator Probability Distribution Function (PDF) from which we derive expressions for its bias and Mean Square Error (MSE), specifically showing that the latter is finite. To solve the apparent contradiction with the bound specified by the CRLB we emphasize a property shared by all (asymptotically) efficient estimators, then show that the MLE does not posses this property. Furthermore, we define the vicinity where the performance of the practical MLE cannot be predicted by the CRLB but is given precisely by our statistical description. In addition, we specify the conditions under which the MLE is locally unbiased at the singular point.

We demonstrate our results using three cases: Phase estimation, DOA estimation and a special polynomial case. In each example we define the singular point, derive the estimator PDF and present simulation results. Specifically, we analyze the DOA problem and show that the MSE is reasonable at the array end-fire (this allows a designer to accurately define the array search range). Moreover we provide an upper bound for the MSE at the singular point. For the special polynomial case, we present close-form expression for both the bias and MSE of the estimator. These expressions are used to emphasize the points raised in this thesis.

Contents

List of Abbreviations

$\boldsymbol{MLE}$	Maximum likelihood estimator
$\boldsymbol{MSE}$	Mean square error
$\boldsymbol{VAR}$	Variance
$\boldsymbol{STD}$	Standard deviation
$\boldsymbol{ULA}$	Uniform linear array
$\boldsymbol{CRLB}$	*Cramér-Rao Lower Bound*
$\boldsymbol{BB}$	Barankin Bound
$\boldsymbol{WW}$	Weiss-Weinstein bound
$\boldsymbol{ZZ}$	Ziv-Zakai bound
$\boldsymbol{PDF}$	Probability distribution function
$\boldsymbol{FIM}$	Fisher information measure
$\boldsymbol{i.i.d.}$	Independently identical distributed
$\boldsymbol{DOA}$	Direction of arrival
$\boldsymbol{TDOA}$	Time difference of arrival
$\boldsymbol{J(\theta)}$	FIM of θ
$\boldsymbol{\Gamma(\cdot)}$	Gamma function
$\boldsymbol{{}_1F_1(a;b;z)}$	Confluent hypergeometric function

List of Figures

Chapter 1

Preface

1.1 Background

1.1.1 The general estimation problem

Estimation problems are widely encountered both in practical applications and in theoretical analysis. A common measure for the efficiency of an estimator is the gap between the estimator variance and the *Cramér-Rao Lower Bound* (CRLB). An estimator that achieves the CRLB is termed efficient. Among the different estimation schemes, Maximum Likelihood Estimators are often chosen due to the fact that they can be derived easily from the distribution function and under regularity conditions are asymptotically unbiased and efficient. However, in the case that the measurements are (locally) loosely dependent on the parameter, the CRLB approaches infinity. Estimator performance depends on the noise characteristics of the problem, mainly Signal to Noise Ratio (SNR), and on the measurements sensitivity to changes in the parameters. This two aspects are represented by the CRLB, which was widely investigated in the past. However, most studies had concentrated on the SNR dependency either through the noise characteristics or through the amount or type of the available data. We emphasize the dependency on the measurement sensitivity to the parameter values. The latter is also the cause for singular points in the CRLB. One might conclude that the variance of the asymptotically efficient MLE diverges in the vicinity of a singular point, thus the estimator is useless.

In this work we analyze the performance of MLE in the vicinity of singular points of the CRLB. We present a thorough investigation of the MLE, show analytically its PDF, and develop analytical expression for its bias and MSE. We prove a general property possessed by asymptotically efficient estimators, then show that, for any given N^1, the MLE is not efficient in the vicinity of a singular point. Furthermore, we show that the estimator MSE is bounded, and identify the area where the MLE performance diverge from the CRLB. Moreover, we specify the case where the MLE is locally unbiased at the singular point. Finally, we apply all of the above for three examples: phase estimation, DOA estimation and a special polynomial case. In each case we analyze the estimator performance over the entire parameter range, including the singular points, and present simulation results. for the last case we derive close form expressions for the expectation and MSE.

Performance of estimators is a problem that has been long studied by many researchers. In this field a great deal of research was devoted to the problem of bounds on the estimation performance. The majority of the research on performance bounds has focused on the CRLB (established in [1], [2] and [3]) and the Barankin bound (BB) [4], which have been derived for a variety of scenarios. The CRLB and BB belong to a family of deterministic "covariance inequality" bounds that treat the parameter θ as an unknown deterministic quantity and provide bounds on the variance in estimating any selected value of the parameter, say θ_0. The *Cramér-Rao* inequality state that the variance of any unbiased estimator of θ_0 must be greater or equal to the inverse of the FIM at θ_0. The CRLB has been popular because it is relatively easy to evaluate for many scenarios. The main drawback of the CRLB and of the BB is that the estimator is required to be unbiased. Other type of bounds have developed from the Bayesian approach. Bounds such as the Ziv-Zakai bound [5] or the Weiss-Weinstein bound [6] assume that the parameter θ is

[1]Where N is the sample size, or the amount of the available data.

a random variable with a known *a priori* distribution. Bayesian type bounds incorporate knowledge of the *a priori* parameter space via the prior distribution of θ. Since the Bayesian approach assumes that the parameter is random the bound includes integration over the entire parameter space. Therefore, the Bayesian bound do not provide information on the accuracy as a function of the parameters value. Given that Bayesian bounds do not bound the performance for a specific θ_0 they are most useful when the estimator performance are nearly the same over all the parameter space.

A different approach is taken in [7] the bias-variance tradeoff curve defined on the $\delta\sigma$ plane[2] specify an unachievable region in that space. Hero et. al. treat the case where θ is deterministic but the estimator is biased, therefore the CRLB cannot be used. However, by construction the $\delta\sigma$ plane does not contain information about the parameter space. Their approach ties the maximal changes in the estimator bias to a relevant estimation error variance. Furthermore, it does not supply a method to bound the estimator behavior over (specific points) the parameter space. We are interested in a parameter space characterized by isolated singular points in the CRLB. In that space we seek a description of the estimator performance at each point for both the bias and the variance. This description cannot be obtained by the method described in [7].

In the case of isolated zero information points the estimator performance is not equal over the parameter space. Bounds originated from the Bayesian approach yield a uniform bound for the entire parameter space, therefore it is not advisable to use them for such a case. In addition the CRLB suggests that the variance of an unbiased estimator is infinite at the singular points hence if an estimator is (asymptotically) unbiased it must be ineffective around these points. We provide a statistical description of the MLE, from which it is pos-

[2]Hero et. al. introduce a plane, which they call the delta-sigma plane, that is indexed by the norm of the estimator bias gradient and the variance of the estimator. The norm of the bias gradient is related to the maximum variation in the estimator bias function over a neighborhood of parameter space.

sible to calculate its MSE for any specific point within the parameter space. None of the methods previously described provides such a description for this case.

1.1.2 Prior research

Maximum Likelihood Estimators have been previously investigated by many researchers ([8], [9] to name some). However, we are not aware of a prior research designated to the mentioned topic. Let $\theta \in \Theta$ be an unknown non-random parameter that parameterize the density $f_Y(y;\theta)$ of an observed random variable Y. The FIM defined as $J(\theta) = E\{(\frac{\partial \log(f_Y(y;\theta))}{\partial \theta})^2\}$ specify the dependency of the measurements on the parameter θ. The *Cramér-Rao Lower Bound* of any unbiased estimator $\hat{\theta}$ of θ, is given by $\mathrm{VAR}(\hat{\theta}) \geq J^{-1}(\theta)$. If $J(\theta_0) = 0$ (which we had defined as a singular point) it is commonly said that there is a fundamental fault in the problem formulation in the vicinity of θ_0. This means that the measurement are very loosely dependent on θ in that neighborhood. Consequently, the CRLB assumes infinite values. Assume $\theta \in [a, b] \subset R$, the random variable Y relates to θ through some smooth and finite function $h(\theta)$ and $J(\theta) = 0$ only at isolated points θ_i. Note that since θ is continuous and $h(\theta)$ is smooth (i.e. its second derivative with respect to θ is continuous) this setup create a vicinity $\{\theta : \ |\theta - \theta_i| \leq \delta, \ \delta \geq 0\}$ in which the performance of an unbiased estimator, bounded by the CRLB, degrades in proportion to $(\frac{\partial h(\theta)}{\partial \theta})^{-2}$. We have defined θ_i to be isolated, but are interested in the effects that this setup has on the MLE in the vicinity of θ_i. Under the regularity conditions the MLE is asymptotically efficient (i.e. the variance of $\hat{\theta}_{ML}$ converges to the CRLB). Often it is assumed that this means that the MLE remains so in the vicinity of θ_0 and therefore it is useless in that area.

In the DOA problem (also known as Angle Of Arrival (AOA) or Bearing estimation) often a linear array of receivers is used to estimate a single side target location. This means that θ, the target direction, is in $[-\frac{\pi}{2}, \frac{\pi}{2}]$ and the singular points are at $\theta = \pm\frac{\pi}{2}$ which is known as the array end-fire. To avoid the problem of estimating the DOA at the array end-fire many applications uses reduced searching range. We analyze the performance of the MLE at the end-fire, thus allowing the developer to determined whether the performance is appropriate all over the search range. A great deal of research has been devoted to this problem and to bounds on estimation performance. Some of the papers on the subject are collected in [10], which deals with a planner array of receivers. However, for the discussed case bounds of the Bayseian type such as the extended Ziv-Zakai bound will not provide the desired answer. Another interesting analysis of this problem concentrating on the MUSIC estimator and the MLE is presented in [11],[12]. A specific expression for the performance of the MUSIC estimator is derived and a comparison to the MLE performance is performed. However in both papers there is no treatment for singularity in the parameter space. The derived results depends on the array size and the amount of available data (number of "snapshots").

Chapter 2

The general scalar estimation problem

2.1 Problem formulation

We introduce the basic scalar estimation problem. Let $\theta \in \Theta$ be an unknown non-random parameter that parameterize the density $f_Y(y;\theta)$ of an observed random variable Y. Assume that the measurement equation is as follows

$$\begin{aligned} & y_n = h(\theta) + v_n, && n = 1, 2, \dots N \\ & v_n \sim N(0, \sigma_v^2), && i.i.d.\ sequence \\ & Y = [y_1, y_2, \dots y_N]^T \end{aligned} \tag{2.1}$$

and we are given the set of measurements $\{y_n\}_{n=1}^N$. The problem on hand is to estimate θ out of this set of measurements. We define $\hat{\theta}$ as the estimator of θ and wish to characterize the estimator performance. It is common to ask questions about the estimator expectation (bias) variance and MSE. The mentioned joint distribution function $f_Y(y;\theta)$ for the discussed case is

$$f_Y(y;\theta) = \prod_{n=1}^{N} f_{Y_n}(y_n;\theta) = \prod_{n=1}^{N} \left(\frac{1}{\sqrt{2\pi\sigma_v^2}} \cdot e^{-\frac{1}{2\sigma_v^2}(y_n - h(\theta))^2}\right) = \tag{2.2}$$

$$\frac{1}{(2\pi\sigma_v^2)^{N/2}} \cdot e^{-\frac{1}{2\sigma_v^2}\sum_{n=1}^{N}(y_n - h(\theta))^2}$$

The *Cramér-Rao* inequality provide some useful information about this problem.

2.2 *Cramér-Rao* Inequality

H. Cramér and C. R. Rao had independently introduced the following inequality that set a lower bound on the MSE of any estimator

$$E(\theta - \hat{\theta})^2 \geq \frac{(\frac{\partial}{\partial\theta} E\hat{\theta})^2}{E(\frac{\partial \log f_Y(y;\theta)}{\partial\theta})^2} \tag{2.3}$$

An estimator is termed unbiased if $\forall\theta,\ E\hat{\theta} = \theta$. For an unbiased estimator we have $MSE(\hat{\theta}) = VAR(\hat{\theta})$ and in addition $\frac{\partial}{\partial\theta}E\hat{\theta} = \frac{\partial}{\partial\theta}\theta = 1$. The result is the well known *Cramér-Rao Lower Bound* (CRLB) on the variance of any unbiased estimator which is given by

$$VAR\{\hat{\theta}\} \geq \frac{1}{E(\frac{\partial \log f_Y(y;\theta)}{\partial\theta})^2} \tag{2.4}$$

The most obvious difference between (2.3) and (2.4) is that the right hand side of (2.3) strictly depends on the specific estimator performance. On the contrary (2.4) sets a general lower bound for the variance of any unbiased estimator $\hat{\theta}$ that depends only on the problem formulation, the lower bound in (2.3) is valid only for estimators with the same bias. Therefore (2.3) is not a bound in the general meaning. However, it still applies for all estimators.

2.2.1 Efficient estimator

An efficient estimator is an estimator that achieves the CRLB, meaning the inequality from (2.4) is replaced with equality. Note that if an estimator is efficient it must also be unbiased. Unfortunately, it is well known that for a general non-linear function $h(\theta)$ no efficient estimator of θ exists. In [3] Cramér had introduce the idea of an asymptotically efficient estimators. If $\hat{\theta}$ is an asymptotically efficient estimator of θ then $f_{\hat{\theta}}(y) \overset{N\to\infty}{\sim} N(\theta, J^{-1}(\theta))$ with probability. Cramér had proved that if the following regularity conditions hold the MLE is an asymptotically efficient estimator. For brevity denote $f = f_Y(y;\theta)$, the regularity conditions are

1. For allmost all y, the derivatives $\frac{\partial \log f}{\partial \theta}$, $\frac{\partial^2 \log f}{\partial \theta^2}$ and $\frac{\partial^3 \log f}{\partial \theta^3}$ exist for every θ belonging to a non-degenerate interval Θ.

2. For every θ in Θ, we have $|\frac{\partial f}{\partial \theta}| < F_1(y)$, $|\frac{\partial^2 f}{\partial \theta^2}| < F_2(y)$ and $|\frac{\partial^3 f}{\partial \theta^3}| < H(y)$, the functions F_1 and F_2 being integrable over $(-\infty, \infty)$, while $\int_{-\infty}^{\infty} H(y) f(y;\theta) dy < M$, where M is independent of θ.

3. For every θ in Θ, the integral $\int_{-\infty}^{\infty} (\frac{\partial \log f}{\partial \theta})^2 f dy$ is finite and positive.

The third condition is the demand that $\forall\theta \in \Theta,\ J(\theta) > 0$. Points of zero information, which we define as singular points of the CRLB, contradicts this condition. However, since θ is a continuous parameter it is interesting to examine the behavior of an estimator in the vicinity of the singular point θ_0. We introduce a property shared by all efficient estimators. This property will help to light on the practical behavior of Maximum Likelihood Estimators in the vicinity of a singular point.

A property of efficient estimators

Theorem 1. *If an estimator is asymptotically efficient then*

$$\forall\theta \in \Theta,\ \lim_{N\to\infty} \{\frac{\partial}{\partial \theta} E\hat{\theta}\} = 1$$

Proof. Assume that $\hat{\theta}$ is an asymptotically efficient estimator of θ. Denote $E\hat{\theta} = \theta + \beta(\theta, \hat{\theta})$ where $\beta(\theta, \hat{\theta})$ is the estimator bias. Note that $\beta(\theta, \hat{\theta})$ depends on both the true parameter value and the specific estimator realization. Since $\hat{\theta}$ is asymptotically efficient it is also asymptotically unbiased, therefore

$$\lim_{N\to\infty} \{\beta(\theta, \hat{\theta})\} = 0 \Rightarrow \lim_{N\to\infty} \{\beta^2(\theta, \hat{\theta})\} = 0 \tag{2.5}$$

In addition, the fact that $\hat{\theta}$ is asymptotically efficient also means that

$$\lim_{N\to\infty}\{VAR(\hat{\theta})\} = \frac{1}{E(\frac{\partial \log f_Y(y;\theta)}{\partial\theta})^2} \tag{2.6}$$

Since (2.3) holds for any estimator we can write

$$MSE(\hat{\theta}) = VAR(\hat{\theta}) + \beta^2(\theta,\hat{\theta}) \geq \frac{(\frac{\partial}{\partial\theta}E\hat{\theta})^2}{E(\frac{\partial \log f_Y(y;\theta)}{\partial\theta})^2} \tag{2.7}$$

But the fact that $\hat{\theta}$ is asymptotically efficient means that if we take the limit $(N \to \infty)$ in the latter we may replace inequality with equality. Moreover, inserting (2.5) and (2.6) we get

$$\frac{1}{E(\frac{\partial \log f_Y(y;\theta)}{\partial\theta})^2} = \frac{\lim_{N\to\infty}\{(\frac{\partial}{\partial\theta}E\hat{\theta})^2\}}{E(\frac{\partial \log f_Y(y;\theta)}{\partial\theta})^2} \tag{2.8}$$

which actually means

$$\lim_{N\to\infty}\{(\frac{\partial}{\partial\theta}E\hat{\theta})^2\} = 1 \Rightarrow \lim_{N\to\infty}\{\frac{\partial}{\partial\theta}E\hat{\theta}\} = \pm 1 \tag{2.9}$$

But due to the asymptotically unbiased property we are left with

$$\lim_{N\to\infty}\{\frac{\partial}{\partial\theta}E\hat{\theta}\} = 1 \tag{2.10}$$

Thus the proof is complete.

□

Note that:

i. The theorem is general and holds for any estimator, and is not limited to the MLE.

ii. We will show several examples where a MLE, that satisfy the regularity conditions, is actually not efficient (for any given N).

iii. Surprisingly, the fact that the MLE is not asymptotically efficient is beneficial since it means that it does not obey the CRLB.

iv. The latter suggests that we should understand the term "Asymptotically unbiased" not only as $\lim_{N\to\infty}\{E\hat{\theta}\} = \theta$, but also as $\lim_{N\to\infty}\{\frac{\partial}{\partial\theta}E\hat{\theta}\} = 1$.

Example for *theorem* 1

The following simple example may help to understand where *theorem* 1 becomes relevant. Assume that the random variable y is an estimator of some unknown variable $x \in R$, and the expectation of y is given by

$$Ey = x - xe^{-Nx^2} \tag{2.11}$$

clearly

$$\forall x, \lim_{N\to\infty} Ey = x \tag{2.12}$$

however

$$\frac{\partial}{\partial x}Ey = 1 - e^{-Nx^2}(1 - 2Nx^2) \tag{2.13}$$

so

$$\lim_{N\to\infty} \frac{\partial}{\partial x}Ey = \begin{cases} 0, & x = 0 \\ 1, & \text{Otherwise} \end{cases} \tag{2.14}$$

The result may seem a little bit strange, (2.14) actually means that although the estimator is asymptotically unbiased everywhere, it is not asymptotically unbiased uniformly over R. Note that in the limit ($N \to \infty$) we may interpret (2.14) as a delta function, meaning $x = 0$ becomes a singular point. However, note that (2.11) is a continuous function. Therefore for any given N (which is the practical case) there exists a vicinity, that is a function of N, for which the expectation derivative will be almost zero. According to *theorem* 1, in that vicinity the estimator is <u>not</u> asymptotically efficient. Later on we will show that this is exactly what happens for the MLE at the vicinity of a singular points of the CRLB.

Chapter 3

Maximum Likelihood Estimator (MLE)

3.1 The standard MLE

Maximum Likelihood Estimators are relatively easy to derive, they only require the knowledge of $f_Y(y;\theta)$. For the basic model given in (2.1) with the distribution function described in (2.2) the MLE is given as the solution of the following likelihood equation

$$\frac{\partial}{\partial\theta}\log(f_Y(y;\theta)) = 0 \tag{3.1}$$

Assume that $h(\theta)$ is a smooth and bounded function and the first derivative of $h(\theta)$ equals zero only at isolated points θ_i. The CRLB for this setup, according to (2.4), is given by

$$VAR\{\hat{\theta}\} \geq \frac{1}{\frac{N}{\sigma_v^2}(\frac{\partial h(\theta)}{\partial\theta})^2} \tag{3.2}$$

This suggests that no useful unbiased estimation scheme can be found in the vicinity of θ_i where $\frac{\partial h(\theta_i)}{\partial\theta} = 0$. It is well known that for a general nonlinear function $h(\theta)$ no efficient estimator of θ can be found. However, the invariance property of the maximum likelihood estimator leads to the following standard solution. Set $\rho \equiv h(\theta)$ and rewrite (2.1) in terms of the auxiliary parameter ρ

$$y_n = \rho + v_n, \quad n = 1, 2, \ldots N \tag{3.3}$$

The result is a linear measurement equation. We know that an estimator of ρ is given by the sample mean and has the following properties.

$$\begin{aligned} &\hat{\rho} = \frac{1}{N}\sum_{n=1}^{N} y_n \\ &VAR\{\hat{\rho}\} = \frac{\sigma_v^2}{N} = \sigma_\rho^2 \\ &\hat{\rho} \sim N(h(\theta), \sigma_\rho^2) \end{aligned} \tag{3.4}$$

since $h(\theta)$ is assumed to be smooth we know that it is at least piecewise monotonous. Without loss of generality we restrict the discussion to a monotonous section of $h(\theta)$. In such a section the inverse function $h^{-1}(\theta)$ exists and may be used to estimate θ. Define

$$\hat{\theta} = h^{-1}(\hat{\rho}) = g(\hat{\rho}) \tag{3.5}$$

The invariance property of the MLE suggests that (3.5) is the MLE of θ. For $\theta \neq \theta_i$ the regularity conditions hold, therefore $\hat{\theta}$ is asymptotically unbiased and efficient almost everywhere. Denote

$$\hat{\rho} = h(\theta) + \sigma_\rho z \tag{3.6}$$

$$z \sim N(0, 1)$$

Then

$$\hat{\theta} = g(\hat{\rho}) = g(h(\theta) + \sigma_\rho z) \tag{3.7}$$

Note that $\sigma_\rho^2 \propto N^{-1}$ and thus

$$\hat{\rho} \xrightarrow{N \to \infty} h(\theta)$$

where the convergence is either in L^2 or in probability. Hence, we may use Taylor expansion of $g(\hat{\rho})$ at the point $\hat{\rho} = h(\theta)$ to express $g(\hat{\rho})$

$$\begin{aligned} &\hat{\theta} = g(h(\theta) + \sigma_\rho z) = g(h(\theta)) + \frac{\partial g(h(\theta))}{\partial \rho}\sigma_\rho z + \frac{\partial^2 g(h(\theta))}{2!\partial \rho^2}(\sigma_\rho z)^2 + O(\sigma_\rho^2 z^2) \\ &\cong \theta + \frac{\partial g(h(\theta))}{\partial \rho}\sigma_\rho z + \frac{\partial^2 g(h(\theta))}{2\partial \rho^2}\sigma_\rho^2 z^2 \end{aligned} \tag{3.8}$$

Hence

$$E\hat{\theta} \cong \theta + \frac{\partial^2 g(\rho)}{2\partial\rho^2}\sigma_\rho^2 = \theta + \frac{\partial^2 g(\rho)}{2\partial\rho^2}\frac{\sigma_v^2}{N} \tag{3.9}$$

In addition

$$E(\theta - \hat{\theta})^2 = E(\frac{\partial g(\rho)}{\partial\rho}\sigma_\rho z)^2 + O(\frac{z^3}{N^{3/2}}) \cong (\frac{\partial g(\rho)}{\partial\rho})^2\frac{\sigma_v^2}{N} \tag{3.10}$$

Note that

$$\frac{\partial g(\rho)}{\partial\rho} = \frac{\partial\theta}{\partial h(\theta)} = \frac{1}{\partial h(\theta)/\partial\theta} \tag{3.11}$$

Therefore we may rewrite (3.10) in the form of (3.2) to show that the suggested MLE of θ is indeed asymptotically efficient as its estimation error variance converges to the CRLB.

$$E(\theta - \hat{\theta})^2 \cong \frac{\sigma_v^2}{N(\frac{\partial h(\theta)}{\partial\theta})^2} \tag{3.12}$$

Unfortunately this feature of the estimator suggests that it is useless whenever $\frac{\partial h(\theta)}{\partial\theta} = 0$. Moreover since $h(\theta)$ is smooth the estimator performance degrades in the vicinity of the isolated points where the first derivative equals zero. Assume $\frac{\partial h(\theta_0)}{\partial\theta} = 0$ From (3.2) one may derive:

$$\lim_{\theta\to\theta_0} E(\theta - \hat{\theta})^2 \geq \lim_{\theta\to\theta_0} \frac{\sigma_v^2}{N(\frac{\partial h(\theta)}{\partial\theta})^2} = \infty \tag{3.13}$$

It is commonly assumed that the MLE of θ will not be useful in the vicinity of θ_0. We show that this is not the case. We begin by deriving the distribution function of the MLE for this case.

3.1.1 A finite parameter space

Assume that the unknown parameter θ belongs to a finite interval Θ (i.e. $\theta \in [a, b]$ where $a < b$). In addition we assume that $f_Y(y;\theta)$ is a unimodal distribution. We derive the MLE for this case. Recall that the MLE is defined as

$$\hat{\theta}_{ML} = \max_{\theta \in \Theta}\{f_Y(y;\theta)\} \tag{3.14}$$

Previously, we had defined $\hat{\theta}_{ML}$ as the solution of the likelihood equation

$$\frac{\partial}{\partial \theta} \log(f_Y(y;\theta)) = 0 \tag{3.15}$$

Denote the solution of the latter as $\tilde{\theta}$, and note that $\tilde{\theta}$ does not necessarily belongs to Θ. From basic calculus, we know that for a continuous and finite function $f(x)$, defined over a finite interval $x \in [a,b]$ we have

$$\max_{x \in [a,b]}\{f(x)\} = \begin{cases} f(x_0), & \text{If } x_0 \text{ is a local maxima} \\ f(a) \text{ or } f(b), & \text{No local maximas in the interval} \end{cases} \tag{3.16}$$

The result is that $\tilde{\theta}$, the solution of (3.15), may not be in Θ, and therefore cannot be considered as $\hat{\theta}_{ML}$. In this case we define the MLE solution as

$$\hat{\theta}_{ML} = \begin{cases} a, & \tilde{\theta} < a \\ b, & b < \tilde{\theta} \end{cases} \tag{3.17}$$

This setup is extremely important when we expect the parameter to be close to the interval limits (as in the DOA problem when the transmitter is located at the end-fire of the array). On top of that this setup also allows another interpretation of (3.14), introduce the following modification of the latter

$$\tilde{\theta} = \max_{\theta \in R}\{f_Y(y;\theta)\}$$

$$\hat{\theta}_{ML} = \begin{cases} a, & \tilde{\theta} < a \\ \tilde{\theta}, & a \le \tilde{\theta} \le b \\ b, & b < \tilde{\theta} \end{cases} \tag{3.18}$$

This way we derive $\tilde{\theta}$ using a linear operation and $\hat{\theta}_{ML}$ through a non-linear truncation of $\tilde{\theta}$. The benefit of this solution will become clear in the next section when we derive the PDF of $\hat{\theta}_{ML}$.

3.2 Performance evaluation of the MLE

3.2.1 PDF of the MLE

Since the (asymptotical) distribution function of $\hat{\rho}$ is well defined let us examine the distribution function of $\hat{\theta}$ in terms of $\hat{\rho}$. Without loss of generality we assume that $h(\theta)$ is non-decreasing over the examined interval $\theta \in [a, b]$. If singularity occurs within this interval it would either be an arbitrary located saddle point or an extremum point at the interval limits. Denote

$$Pr(\hat{\theta} \leq s) = Pr(h^{-1}(\hat{\rho}) \leq s) = Pr(\hat{\rho} \leq h(s)) = F_{\hat{\rho}}(h(s)) \tag{3.19}$$

Assuming that (3.19) is valid, it allows to derive the PDF of $\hat{\theta}$ by taking its derivative with respect to θ.

$$f_{\hat{\theta}}(s) = \frac{\partial}{\partial s} F_{\hat{\rho}}(h(s)) = \frac{\partial h(s)}{\partial s} f_{\hat{\rho}}(h(s)) \tag{3.20}$$

Note that (3.20) is well defined and has no singular values when $\frac{\partial h(\theta_0)}{\partial \theta} = 0$. From the PDF of $\hat{\theta}$ we may calculate all the probabilistic characterization of $\hat{\theta}$. In particular we are interested in the MSE that is given by:

$$E\{(\hat{\theta} - \theta)^2\} = \int_{\Theta} (s - \theta)^2 f_{\hat{\theta}}(s) ds = \int_{\Theta} (s - \theta)^2 \frac{\partial h(s)}{\partial s} f_{\hat{\rho}}(h(s)) ds \tag{3.21}$$

In addition we may use (3.20) to compute the estimation bias

$$\beta(\theta, \hat{\theta}) = \theta - E\hat{\theta} = \theta - \int_{\Theta} s f_{\hat{\Theta}}(s) ds \tag{3.22}$$

We distinguish the case of a finite Θ from the infinite case. If Θ is infinite the problem is simplified as the only type of singularity that we consider is saddle points in $h(\theta)$. For this case $h(\theta)$ is strictly monotonous in R, and the integral limits in (3.21) are $\pm\infty$.

PDF of $\hat{\theta}_{ML}$ in case of a finite parameter space

In the case that the parameter space is finite we use the interpretation of the MLE that was introduced in section 3.1.1. Assume $\theta \in [a, b]$, we use $\tilde{\theta}$ as an auxiliary variable in the following manner. Define $x \in R$ and $g(x)$ a smooth and monotonous function in R obeying

$$g(x) = \begin{cases} g_1(x), & x < a \\ h(x), & a \leq x \leq b \\ g_2(x), & b < x \end{cases} \tag{3.23}$$

where $g_1(x)$ and $g_2(x)$ are some convenient monotonous extension of $h(x)$. Replace the original measurements equation in (2.1) with the following one

$$y_n = g(x) + v_n, \;\; n = 1, 2, \ldots N \tag{3.24}$$

Define $\tilde{\theta}$ as the MLE of x, then its PDF can be immediately derived in terms of (3.19) and (3.20). Next we define $\hat{\theta}_{ML}$ in terms of (3.18) as

$$\hat{\theta}_{ML} = \begin{cases} a, & \tilde{\theta} < a \\ \tilde{\theta}, & a \leq \tilde{\theta} \leq b \\ b, & b < \tilde{\theta} \end{cases} \tag{3.25}$$

The probability distribution function of $\hat{\theta}_{ML}$ is given by

$$P_r(\hat{\theta}_{ML} \leq s) = \begin{cases} 0, & s < a \\ P_r(\tilde{\theta} \leq s), & a \leq s \leq b \\ 1, & b < s \end{cases} \tag{3.26}$$

Therefore, its PDF is

$$f_{\hat{\theta}_{ML}}(s) = \alpha \cdot \delta(s-a) + \beta \cdot \delta(s-b) + f_{\tilde{\theta}}(s) \cdot I(a \leq s \leq b) \tag{3.27}$$

and

$$\alpha = F_{\tilde{\theta}}(a);\ \beta = 1 - F_{\tilde{\theta}}(b)$$

while $I(x)$ is an indicator function that equals one as long as its condition is 'true'. Replacing the latter into (3.21) allows the direct computation of the MSE of $\hat{\theta}_{ML}$.

Assume now that Θ is finite and $a < \theta < b$, meaning the true parameter value belongs to the open interval (a, b). Depending on $f_{\tilde{\theta}}(s)$ it is possible to show that

$$\lim_{N\to\infty} \alpha = 0;\ \lim_{N\to\infty} \beta = 0$$

which suggest that the PDF of $\hat{\theta}_{ML}$ almost equals $f_{\tilde{\theta}}(s)$, or in other words

$$\forall s \text{ not in } [a, b],\ \lim_{N\to\infty} f_{\tilde{\theta}}(s) = 0$$

Therefore, if $\theta \neq a$ or $\theta \neq b$ we can assume that for large enough N we may use $f_{\tilde{\theta}}(s)$ to predict the performance of $\hat{\theta}_{ML}$. However, if for example $\theta = a$ then it is very important to choose $g_1(x)$ properly. It will be later on showed that if the extended function is antisymmetric around the true parameter value then the integrand in (3.21) is a symmetric function.

Conditions for a finite MSE

Common knowledge suggest that the MSE of any estimator is bounded as long as the parameter range Θ is finite. This of course also applies to the MLE. The following claim indicates when the MSE, of the MLE, is finite even if Θ is infinite.

Claim 1. *For the problem formulation as defined in (2.1) where* $\frac{\partial h(\theta)}{\partial \theta} = 0$ *at isolated points and* $h(\theta)$ *is smooth the MSE of the MLE is bounded if either one of the following conditions holds: a. The interval* Θ *is finite. b. The distribution function of* $\hat{\theta}$ *decays at least like* $\frac{1}{s^4}$*.*

Proof. If 'a' holds the proof is trivial: Since $f_{\hat{\theta}}(s)$ is a distribution function, by definition we have

$$\int_{\Theta} f_{\hat{\theta}}(s)ds = 1 \tag{3.28}$$

In addition $(s-\theta)^2 \leq C^2$, where C is the size of Θ, therefore

$$\int_{\Theta} (s-\theta)^2 f_{\hat{\theta}}(s)ds \leq \int_{\Theta} C^2 f_{\hat{\theta}}(s)ds = C^2 < \infty \tag{3.29}$$

So the first part of the claim is proved. To prove the second part we must show that the integral tail is bounded. Without loss of generality, assume $\theta = 0$, and $f_{\hat{\theta}}(s) = f_{\hat{\theta}}(-s)$, hence

$$MSE(\theta) = \int_{R} s^2 f_{\hat{\theta}}(s)ds = \int_{-C}^{C} s^2 f_{\hat{\theta}}(s)ds + 2\int_{C}^{\infty} s^2 f_{\hat{\theta}}(s)ds \tag{3.30}$$

Note that

$$\int_{-C}^{C} f_{\hat{\theta}}(s)ds < 1 \tag{3.31}$$

and therefore,

$$\int_{-C}^{C} s^2 f_{\hat{\theta}}(s)ds \leq C^2 \int_{-C}^{C} f_{\hat{\theta}}(s)ds < C^2 \tag{3.32}$$

In addition, due to the constrain that $f_{\hat{\theta}}(s)$ decays at least like $\frac{1}{s^4}$ we have

$$2\int_{C}^{\infty} s^2 f_{\hat{\theta}}(s)ds \leq 2\int_{C}^{\infty} s^2 \frac{1}{s^4} ds = 2\int_{C}^{\infty} s^{-2} ds < \frac{2}{C} \tag{3.33}$$

Therefore, for some arbitrary C, we have

$$MSE(\theta) < C^2 + \frac{2}{C} < \infty \tag{3.34}$$

Thus the proof is complete.

□

Since 'b' only depends on the noise characteristics of the problem, it holds in many practical problems[1]. For example if the noise is Gaussian then condition 'b' immediately holds. Furthermore, if the regularity conditions hold the asymptotical distribution of $\hat{\theta}$ is normal, hence the actual distribution must decay exponentially and 'b' holds.

Hence we may conclude that:

i. The mean square error of the MLE of θ is bounded (regardless of the points where $\frac{\partial h(\theta)}{\partial \theta} = 0$).

ii. It is possible to calculate the MSE directly from (3.21) and therefore predict the performance of the MLE.

This may seem like a contradiction to the fact that the MLE is asymptotically unbiased and efficient and at the same time the CRLB is infinite where $\frac{\partial h(\theta)}{\partial \theta} = 0$. To explain the apparent contradiction we must first distinguish unbiased estimators from asymptotically unbiased estimators. It is easy to show that, for any given N the MLE is a biased estimator of θ. This can be either viewed from (3.9) or understood by means of the Jenssen inequality. The Jenssen inequality states that for any convex function $f(x)$: $Ef(x) \geq f(Ex)$, while for any concave function $f(x)$: $Ef(x) \leq f(Ex)$. Hence, the equality $Ef(x) = f(Ex)$ holds only for the case that $f(x)$ is both concave and convex at the same time, i.e. $f(x)$ is a linear functions[2]. Note that $E\hat{\theta} = Eh^{-1}(\hat{\rho})$. An estimator is defined as unbiased if $E\hat{\theta} = \theta$, but for a general nonlinear function $h^{-1}(\theta)$ by the Jennsen inequality we have $Eh^{-1}(\hat{\rho}) \neq h^{-1}(E\hat{\rho}) = \theta$. Hence $E\hat{\theta} \neq \theta$ or equivalently the MLE of θ is biased.

[1]This suggests that condition 'b' is not very strict, and that for most common application the MSE of the MLE is indeed bounded.

[2]In the case that $f(x)$ is neither strictly concave nor strictly convex we may utilize the Jenssen inequality for either parts separately. Still the equality will hold only if $f(x)$ is strictly linear.

3.2.2 Finite lower bound on the MSE

Since we saw that the MLE is a biased estimator we cannot use (2.4) to bound its variance and are forced to use (2.3). Recall that according to (2.3) we have

$$E(\theta - \hat{\theta})^2 \geq \frac{(\frac{\partial}{\partial\theta} E\hat{\theta})^2}{E(\frac{\partial \log f_Y(y;\theta)}{\partial\theta})^2}$$

We will now show that the right hand side of the latter is bounded for the discussed case (MLE). Note that in accordance with (3.2) the denominator in (2.3) equals $\frac{N}{\sigma_v^2}(\frac{\partial h(\theta)}{\partial\theta})^2$ and the singularity is due to the fact that $\frac{\partial h(\theta_i)}{\partial\theta} = 0$. To evaluate the numerator of (2.3) we will use (3.4) and (3.20) to conclude that

$$f_{\hat{\theta}}(s) = \frac{\partial h(s)}{\partial s} \frac{1}{\sqrt{2\pi\sigma_\rho^2}} e^{-\frac{1}{2\sigma_\rho^2}(h(s)-h(\theta))^2} \tag{3.35}$$

Therefore

$$\frac{\partial}{\partial\theta} E\hat{\theta} = \frac{\partial}{\partial\theta} \int_\Theta s \frac{\partial h(s)}{\partial s} \frac{1}{\sqrt{2\pi\sigma_\rho^2}} e^{-\frac{1}{2\sigma_\rho^2}(h(s)-h(\theta))^2} ds \tag{3.36}$$

By the same arguments as in *claim* 1 we conclude that the expectation of $\hat{\theta}$ is finite, therefore it is possible to exchange the integration and derivation order. The result is

$$\frac{\partial}{\partial\theta} E\hat{\theta} = \int_\Theta s \frac{\partial h(s)}{\partial s} \frac{1}{\sqrt{2\pi\sigma_\rho^2}} \frac{\partial}{\partial\theta} e^{-\frac{1}{2\sigma_\rho^2}(h(s)-h(\theta))^2} ds \tag{3.37}$$

Taking the derivative with respect to θ yield

$$\frac{\partial}{\partial\theta} E\hat{\theta} = \int_\Theta s \frac{\partial h(s)}{\partial s} \frac{1}{\sqrt{2\pi\sigma_\rho^2}} \frac{h(s)-h(\theta)}{\sigma_\rho^2} \frac{\partial h(\theta)}{\partial\theta} e^{-\frac{1}{2\sigma_\rho^2}(h(s)-h(\theta))^2} ds \tag{3.38}$$

Which can be also written as

$$\frac{\partial}{\partial\theta} E\hat{\theta} = \frac{1}{\sigma_\rho^2} \frac{\partial h(\theta)}{\partial\theta} \int_\Theta s(h(s)-h(\theta)) f_{\hat{\theta}}(s) ds \tag{3.39}$$

Or finally

$$\frac{\partial}{\partial\theta}E\hat{\theta} = \frac{1}{\sigma_\rho^2}\frac{\partial h(\theta)}{\partial\theta}(E\{\hat{\theta}h(\hat{\theta})\} - h(\theta)E\hat{\theta}) \tag{3.40}$$

From (3.39) and (3.40) it immediately follows that the MLE is asymptotically biased in the vicinity of the singular point[3] θ_i. Replacing the numerator of (2.3) with the right hand side that of (3.40) and the denominator of (2.3) with the denominator from (3.2) yields

$$E(\theta - \hat{\theta})^2 \geq \frac{[\frac{1}{\sigma_\rho^2}\frac{\partial h(\theta)}{\partial\theta}(E\{\hat{\theta}h(\hat{\theta})\} - h(\theta)E\hat{\theta})]^2}{\frac{N}{\sigma_v^2}(\frac{\partial h(\theta)}{\partial\theta})^2} = \tag{3.41}$$

$$\frac{(\frac{N}{\sigma_v^2})^2(\frac{\partial h(\theta)}{\partial\theta})^2(E\{\hat{\theta}h(\hat{\theta})\} - h(\theta)E\hat{\theta}))^2}{\frac{N}{\sigma_v^2}(\frac{\partial h(\theta)}{\partial\theta})^2}$$

Canceling out equivalent terms and rearranging yields

$$E(\theta - \hat{\theta})^2 \geq \frac{N}{\sigma_v^2}[E\{(h(\hat{\theta}) - h(\theta))\hat{\theta}\}]^2 \tag{3.42}$$

Although the last expression seems to be proportional to N this is not the case. Remember that we had defined $\rho = h(\theta)$ and $\hat{\theta} = h^{-1}(\rho)$, therefore we have $h(\hat{\theta}) = h(h^{-1}(\hat{\rho})) = \hat{\rho}$. Replacing the latter into (3.42) we get

$$E(\theta - \hat{\theta})^2 \geq \frac{N}{\sigma_v^2}[E\{(\hat{\rho} - \rho)\hat{\theta}\}]^2 \tag{3.43}$$

but

$$\hat{\rho} - \rho = \frac{1}{N}\sum_{n=1}^{N} y_n - \rho = \frac{1}{N}\sum_{n=1}^{N}(\rho + v_n) - \rho = \frac{1}{N}\sum_{n=1}^{N} v_n$$

and

$$\frac{N}{\sigma_v^2}[E\{\hat{\theta}\frac{1}{N}\sum_{n=1}^{N} v_n\}]^2 = \frac{1}{N\sigma_v^2}[E\hat{\theta}\sum_{n=1}^{N} v_n]^2 = \frac{1}{N\sigma_v^2}[\sum_{n=1}^{N} E\{\hat{\theta}v_n\}]^2$$

[3]Remember that we had showed that if an estimator is asymptotically unbiased then $\forall\theta,\ \lim_{N\to\infty}\frac{\partial}{\partial\theta}E\hat{\theta} = 1$. However, from (3.40) we can see that at $\theta = \theta_i$, $\frac{\partial}{\partial\theta}E\hat{\theta} \equiv 0$.

Finally we may insert the latter into (3.43) which yields the desired result

$$E(\theta-\hat{\theta})^2 \geq \frac{1}{N}\cdot\frac{[\sum_{n=1}^{N}E\{\hat{\theta}v_n\}]^2}{\sigma_v^2} \tag{3.44}$$

Before we continue to discuss the last result we show that the right hand side of (3.44) is finite. Applying the Cauchy-Schwarz inequality to the term on the right side yields

$$\frac{1}{N\sigma_v^2}[E\hat{\theta}\sum_{n=1}^{N}v_n]^2 \leq \frac{1}{N\sigma_v^2}E(\hat{\theta}^2)E(\sum_{n=1}^{N}v_n)^2 = E(\hat{\theta}^2) \tag{3.45}$$

so we end up with

$$\frac{1}{N\sigma_v^2}[E\hat{\theta}\sum_{n=1}^{N}v_n]^2 \leq E(\hat{\theta}^2) = \int_{\Theta} s^2\frac{\partial h(s)}{\partial s}f_{\hat{P}}(h(s))ds \tag{3.46}$$

So we showed that the right hand side of (3.44) is finite. To continue the discussion on (3.44) we replace $\hat{\theta}$ with its definition, this yields

$$E(\theta-\hat{\theta})^2 \geq \frac{1}{N}\cdot\frac{[\sum_{n=1}^{N}E\{h^{-1}(\hat{\rho})v_n\}]^2}{\sigma_v^2} = \tag{3.47}$$

$$\frac{1}{N}\cdot\frac{[\sum_{n=1}^{N}E\{h^{-1}(\frac{1}{N}\sum_{j=1}^{N}(h(\theta)+v_j))v_n\}]^2}{\sigma_v^2}$$

or finally

$$E(\theta-\hat{\theta})^2 \geq \frac{1}{N}\cdot\frac{[\sum_{n=1}^{N}E\{h^{-1}(h(\theta)+\frac{1}{N}\sum_{j=1}^{N}v_j)v_n\}]^2}{\sigma_v^2} \tag{3.48}$$

Unfortunately, there is no simple interpretation of the latter. Although it is tempting to use Taylor expansion on the denominator of the right hand side, it does not help to simplify it (this can be also understood from (3.8)). However, there are several things to note about either (3.44) or (3.48)

i. First we point out that the lower bound on the MSE is finite and bounded by the second moment of $\hat{\theta}$. The latter is finite due to the same reasoning argued in the proof of *claim* 1.

ii. Although (3.44) seems like a general bound on the MSE, of any estimator, it is not the case. Note that the right hand side is specifically dependent on the properties of the selected MLE.

iii. Having said that, it is still a lower bound on the MSE of the proposed MLE.

iv. The actual bound depends on the ratio between the square sum of the correlation between the residual noise transformed into $\hat{\theta}$ with each noise sample and the average noise power.

v. Equation (3.43) can also be understood in terms of the orthogonality principle, meaning that it attains its minimum when the estimation error $(\rho - \hat{\rho})$ is orthogonal to $\hat{\theta} = g(\hat{\rho})$.

Example: MSE bound for the linear case

Since (3.44) does not seem very intuitive, we will show that at least for the linear case it coincides with common knowledge. Assume $h(\theta) = a\theta$ and the measurements are given by $y_n = a\theta + v_n$. For this case we know that the MLE is an efficient estimator and is given by

$$\hat{\theta} = \frac{1}{aN} \sum_{n=1}^{N} y_n \tag{3.49}$$

It can be easily verified that $\hat{\theta}$ is unbiased (i.e. $E\hat{\theta} = \theta$) and that its variance is given by $VAR(\hat{\theta}) = \frac{\sigma_v^2}{Na^2}$. Let us verify that (3.44) yield the expected results

$$\textstyle\sum_{n=1}^{N} E\{\hat{\theta} v_n\} = \sum_{n=1}^{N} E\{\frac{1}{aN} \sum_{j=1}^{N} (a\theta + v_j) v_n\} =$$

$$\textstyle\sum_{n=1}^{N} E\{(\theta + \frac{1}{aN} \sum_{j=1}^{N} v_j) v_n\} = \frac{\sigma_v^2}{a}$$

Setting the results into (3.44) yields

$$E(\theta-\hat{\theta})^2 = VAR(\hat{\theta}) \geq \frac{1}{N}\cdot\frac{[\sum_{n=1}^{N}E\{\hat{\theta}v_n\}]^2}{\sigma_v^2} = \frac{1}{N}\cdot\frac{(\frac{\sigma_v^2}{a})^2}{\sigma_v^2} = \frac{\sigma_v^2}{Na^2} \qquad (3.50)$$

3.3 Regions where the MLE does not obey the CRLB

Our next objective is to point out the region where (2.3) deviates from (3.2). To do so we introduce another interpretation of (2.3). Assume $E\hat{\theta} = \theta + \beta(\theta,\hat{\theta})$, where $\beta(\theta,\hat{\theta})$ is the estimator bias, in that case $\frac{\partial}{\partial\theta}E\hat{\theta} = 1 + \frac{\partial}{\partial\theta}\beta(\theta,\hat{\theta})$. Previously we had mentioned that the term asymptotically unbiased should be also understood in the sense that $\lim_{N\to\infty}(\frac{\partial}{\partial\theta}E\hat{\theta}) = 1$, so we may now check the limit of the bias derivative. It is well known that "far enough" from the singularity points, the performance of the MLE is described well by the CRLB. This means that in this region $\lim_{N\to\infty}(\frac{\partial}{\partial\theta}\beta(\theta,\hat{\theta})) = 0$. Moreover, since the MLE in that region is asymptotically unbiased we may assume $\beta(\theta,\hat{\theta}) \cong 0$. However, from (3.40) we have $\frac{\partial}{\partial\theta}E\hat{\theta}|_{\theta=\theta_i} \equiv 0$ which force $\lim_{N\to\infty}(\frac{\partial}{\partial\theta}\beta(\theta,\hat{\theta})) = -1$ at $\theta=\theta_i$.

Before we continue we examine the behavior of $E\hat{\theta}$ in the vicinity of a singular point θ_0. This depends on the properties and nature of $h(\theta)$ over the interval Θ.

Theorem 2. *If $h(\theta)$ is antisymmetric around $(\theta_0, h(\theta_0))$ and the domain Θ is symmetric around θ_0 then $\beta(\theta_0) = 0$ (equivalently $E\hat{\theta}|_{\theta=\theta_0} = \theta_0$). If the domain Θ is not symmetric around θ_0 then $\lim_{N\to\infty}\{\beta(\theta_0)\} = 0$.*

Proof. From (3.35) we have

$$E\hat{\theta} = \int_{\Theta} s\frac{\partial h(s)}{\partial s}\frac{1}{\sqrt{2\pi\sigma_\rho^2}}e^{-\frac{1}{2\sigma_\rho^2}(h(s)-h(\theta))^2}ds \qquad (3.51)$$

It is well known that if $h(\theta)$ is antisymmetric around $(\theta_0, h(\theta_0))$, then $\frac{\partial}{\partial\theta}h(\theta)$ is symmetric around θ_0. Let us calculate $E(\hat{\theta} - \theta_0)$, from (3.51) we get

$$E(\hat{\theta} - \theta_0) = \int_{\Theta} (s - \theta_0) \frac{\partial h(s)}{\partial s} \frac{1}{\sqrt{2\pi\sigma_\rho^2}} e^{-\frac{1}{2\sigma_\rho^2}(h(s)-h(\theta_0))^2} ds \tag{3.52}$$

Note that we have one antisymmetric function around θ_0, that is $(s - \theta_0)$, and two symmetric functions, which are $\frac{\partial h(s)}{\partial s}$ and $\exp\{-\frac{1}{2\sigma_\rho^2}(h(s) - h(\theta_0))^2\}$, so the integrand in (3.52) is antisymmetric around θ_0. If Θ is symmetric around θ_0 it immediately follows that (3.52) equal zero and the first claim is proved. If this is not the case we define

$$\delta = \min(|a - \theta_0|, |b - \theta_0|)$$

as the distance from θ_0 to the closest edge of Θ. Without loss of generality we will assume that θ_0 is closer to the lower interval limit a. We now rewrite (3.52) as

$$E(\hat{\theta} - \theta_0) = \int_a^{a+2\delta} (s - \theta_0) \frac{\partial h(s)}{\partial s} \frac{1}{\sqrt{2\pi\sigma_\rho^2}} e^{-\frac{1}{2\sigma_\rho^2}(h(s)-h(\theta_0))^2} ds + \tag{3.53}$$

$$\int_{a+2\delta}^{b} (s - \theta_0) \frac{\partial h(s)}{\partial s} \frac{1}{\sqrt{2\pi\sigma_\rho^2}} e^{-\frac{1}{2\sigma_\rho^2}(h(s)-h(\theta_0))^2} ds$$

The first integral equal zero since it is an integral over a symmetric interval on an antisymmetric function. The limit of the second integral when $N \to \infty$ is zero because the integrand limit is zero. Note that the limit in N is

$$\lim_{N\to\infty} \{\sqrt{N} e^{-\frac{1}{2\sigma_\rho^2}(h(s)-h(\theta_0))^2}\} \stackrel{\dagger}{=} \lim_{N\to\infty} \{C \frac{1}{\sqrt{N}} e^{-\frac{N}{2\sigma_v^2}(h(s)-h(\theta_0))^2}\} = 0$$

Where equality † is due to L'Hospital rule. In the case that θ_0 is at one of the interval edges we may always assure that the necessary extension of $h(\theta)$ is antisymmetric as required, therefore it will not affect the completeness of the proof. Thus the proof is complete.

□

Assume now that $h(\theta)$ is antisymmetric around $(\theta_0, h(\theta_0))$. In this case we have

$$\lim_{N\to\infty}\{\beta(\theta,\hat{\theta})\} = \begin{cases} 0, & \theta \text{ “far from” } \theta_0 \\ 0, & \theta = \theta_0 \end{cases}$$

$$\lim_{N\to\infty}\{\tfrac{\partial}{\partial\theta}\beta(\theta,\hat{\theta})\} = \begin{cases} 0, & \theta \text{ “far from” } \theta_0 \\ -1, & \theta = \theta_0 \end{cases}$$

This allows the characterization of $\beta(\theta, \hat{\theta})$, one function that exhibit such a behavior is $f(x) = (x_0 - x)\exp(-\eta(x - x_0)^2)$. In figure 3.1 we illustrate such a function for $\eta = 10$ and $x_0 = \frac{\pi}{2}$. Later on we will show that this would be the case in the DOA problem. If $h(\theta)$ is not antisymmetric the only change would be $\lim_{N\to\infty}\{\beta(\theta_0)\} \neq 0$. This means that it is harder to exemplify the behavior of $\beta(\theta, \hat{\theta})$, still it would have to vary in some way since the derivative $\frac{\partial}{\partial\theta}\beta(\theta_0) = -1$.

Corollary 1. *All of the above suggests that for any given N: a) The MLE is inefficient in the vicinity of a singular point. b) The MLE is biased in that vicinity.*

We are left with the task of identifying that vicinity. Until now we had discussed the asymptotical behavior of the MLE but in practical cases we cannot assume $(N \to \infty)$. Since we always face a finite N, we are looking for an environment (that is N dependent) at which the performance of the MLE cannot be predicted by the CRLB. We suggest two options to identify this area.

i. Find the points where the ratio $\frac{|\beta(\theta,\hat{\theta})|}{STD(\hat{\theta})} \geq 0.1$, those points mark the vicinity where the MLE of θ is biased.

ii. Find the area where, for the given N, $|\frac{\partial}{\partial\theta}\beta(\theta,\hat{\theta})| \geq 0.1$ this is the area where the MLE of θ is biased.

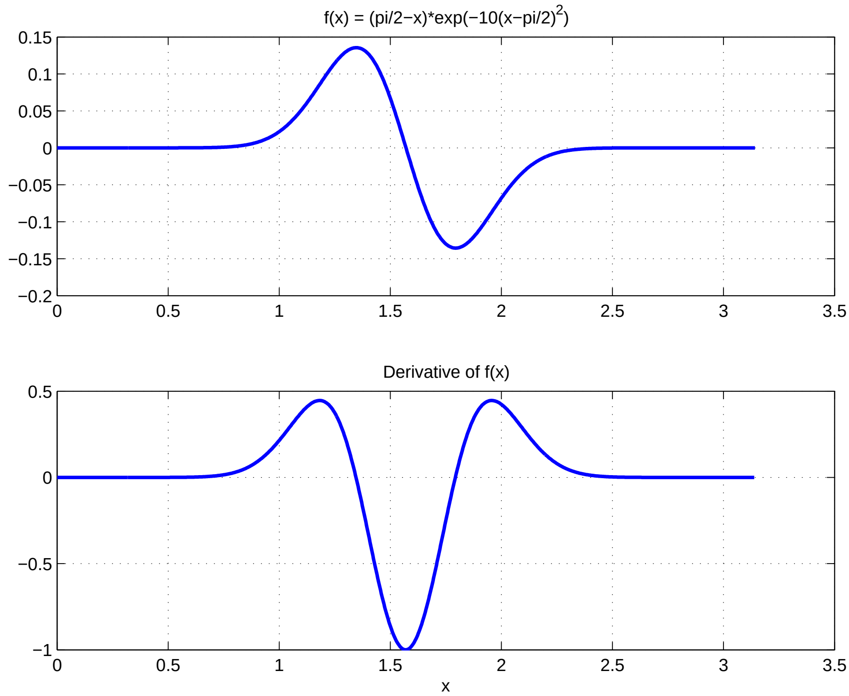

Figure 3.1: A possible behavior of $\beta(\theta,\hat{\theta})$ for the case that $h(\theta)$ is antisymmetric.

Where $\beta(\theta,\hat{\theta})$ is defined in (3.22) and $STD(\hat{\theta}) = \sqrt{VAR(\hat{\theta})}$, and the variance is defined as

$$VAR(\hat{\theta}) = \int_{\Theta} \beta^2(\theta,s) f_{\hat{\theta}}(s) ds \tag{3.54}$$

Nevertheless, we recall that we may always use (3.21) to predict the MLE performance.

Corollary 2. *The fact that the MLE is not governed by the CRLB also means that its variance, in the vicinity of the a singular point θ_i, can be finite (as opposed to the CRLB that forces $VAR(\hat{\theta})|_{\theta=\theta_i} \geq \infty$). In addition, if the conditions in* claim *1 hold then the MSE is indeed finite, and therefore the MLE is still useful in the vicinity of θ_i. Moreover, in some cases it is also locally unbiased at the singular point. Such a case is the DOA problem.*

Chapter 4

Examples

Surprisingly, the simple model in (2.1) is enough to fully characterize the more complex problem of DOA estimation by a Uniform Linear Arrays (ULA). In order to apply our analysis to that problem, we shall first consider the problem of phase estimation. The parameter space Θ in the first two examples is finite so we use the MLE interpretation presented in section 3.1.1.

4.1 Phase estimation

Assume that we wish to estimate the phase of a known signal from measurements corrupted with additive noise. We use the following model, based on (2.1), to analyze this problem

$$\begin{aligned} y_n &= h(\theta) + v_n \\ h(\theta) &= \sin(\theta) \\ \theta &\in [-\tfrac{\pi}{2}, \tfrac{\pi}{2}] \end{aligned} \tag{4.1}$$

From (3.2) we find that the singularity points are at $\theta_0 = \pm\frac{\pi}{2}$. We may use (3.4) to get $\hat{\rho} \sim N(\sin(\theta), \sigma_\rho^2)$. Since $\hat{\rho} \in R$ (rather than limited to $[-1, 1]$) we cannot use the inverse sine function to extract $\hat{\theta}$. Nonetheless, for large enough N we may assume that $P_r(|\hat{\rho} - 1| > \epsilon) \approx 0$, thus we need a function that gently extends the domain of the inverse sine beyond [-1,1]. Moreover, coming to use (3.20) we must pay close attention to the assumption that $h(\theta)$

must be non-decreasing. In this case, the singular points that are the edge points of the parameter space Θ are an extremum points of $h(\theta)$. Hence it is not monotonous in the vicinity of those points. Before we may apply (3.20) we must extend the domain of θ beyond $[-\frac{\pi}{2}, \frac{\pi}{2}]$, and $h(\theta)$ in a monotonous manner over the extended domain. Since we a priori know that $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ we may extend $h(\theta)$ in any convenient way. Introduce the following function

$$h(x) = \begin{cases} -2 - \sin(x), & -\frac{3\pi}{2} \leq x < -\frac{\pi}{2} \\ \sin(x), & -\frac{\pi}{2} \leq x \leq \frac{\pi}{2} \\ 2 - \sin(x), & \frac{\pi}{2} < x \leq \frac{3\pi}{2} \end{cases} \tag{4.2}$$

This function is monotonous if $x \in [-\frac{3\pi}{2}, \frac{3\pi}{2}]$ (it is also antisymmetric around $\pm\frac{\pi}{2}$), and it allows us to define $h^{-1}(x)$ in a monotonous manner over the interval $[-3, 3]$. Since for large enough N we have $P_r(|\hat{\rho}| \geq 3) \approx 0$ this extension is enough. Define $h^{-1}(x)$ in the following form

$$h^{-1}(x) = g(x) = \begin{cases} -\pi + \sin^{-1}(x+2), & -3 \leq x < -1 \\ \sin^{-1}(x), & -1 \leq x \leq 1 \\ \pi + \sin^{-1}(x-2), & 1 < x \leq 3 \end{cases} \tag{4.3}$$

Next we may apply (3.20) directly to get

$$f_{\hat{\theta}}(s) = \frac{\partial h(s)}{\partial s} f_{\hat{\rho}}(h(s)) = \tag{4.4}$$

$$= \begin{cases} \frac{-\cos(s)}{\sqrt{2\pi\sigma_\rho^2}} exp\{-\frac{1}{2\sigma_\rho^2}(2+\sin(s)+\sin(\theta))^2\}, & -\frac{3\pi}{2} \le s < -\frac{\pi}{2} \\ \frac{\cos(s)}{\sqrt{2\pi\sigma_\rho^2}} exp\{-\frac{1}{2\sigma_\rho^2}(\sin(s)-\sin(\theta))^2\}, & -\frac{\pi}{2} \le s \le \frac{\pi}{2} \\ \frac{-\cos(s)}{\sqrt{2\pi\sigma_\rho^2}} exp\{-\frac{1}{2\sigma_\rho^2}(2-\sin(s)-\sin(\theta))^2\}, & \frac{\pi}{2} < s \le \frac{3\pi}{2} \end{cases}$$

Simulation results of this estimator shows that it is indeed biased when $\theta \to \pm\frac{\pi}{2}$ and that its MSE is bounded. Moreover (4.4) allows to compute, by means of (3.20) and (3.22), (numerically) both the bias and MSE of the estimator. Simulation results of the phase estimation problem are presented in figure 4.1. It is evident that the estimator is characterized by (4.4). In addition, as claimed in *theorem* 2, we may note that the estimation bias decreases when $\theta \to \pm\frac{\pi}{2}$, this is due to the antisymmetric nature of the extended function defined in (4.2). The vertical lines mark the area where the performance of the MLE diverge from those predicted by the CRLB. In this particular example the two options (bias/STD ratio and the absolute bias value) point to the same θ. We are now ready to continue to the problem of DOA estimation.

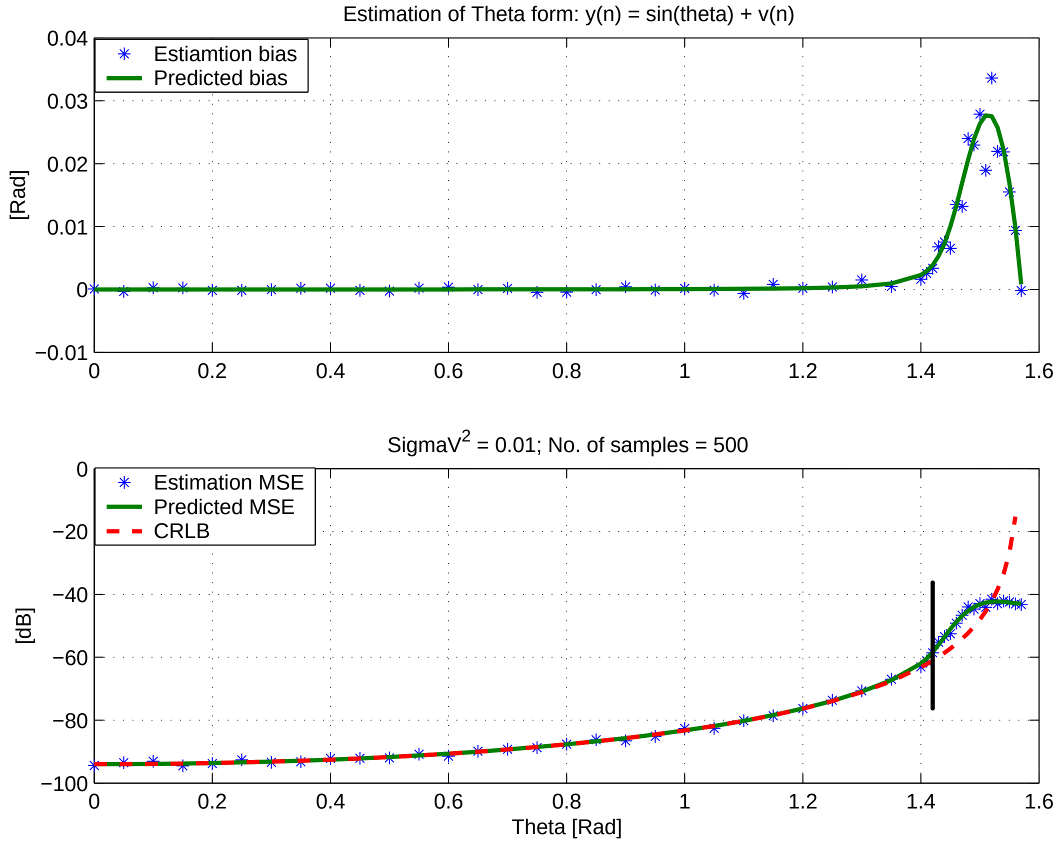

Figure 4.1: Simulation results of phase estimation using MLE.

4.2 DOA Estimation

The DOA estimation problem was widely discussed, in many papers and books, in the past. We consider the ULA model presented in [13]. For this case the singularity occur when the received signal arrives from the array end-fire. We present a MLE which is a slight modification of the MLE solution presented in [11]. Assume $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}] \equiv \Theta$, and consider the following model (standard baseband model) describing a linear array of uniformly spaced M identical sensors with one target signal.

$$y(k;\theta) = A(\theta)s(k) + v(k), \;\; k = 1, 2, \ldots, N \tag{4.5}$$

Where $y(k;\theta) \in C^{M\times 1}$ is the noisy data vector, $s(k) \in C^{1\times 1}$ is the signal complex envelop and $v(k) \in C^{M\times 1}$ is an additive white noise (i.i.d.). Due to the fact that we deal with only one target, $A(\theta) \in C^{M\times 1}$ is a vector of phase delays. We assume

$$s(k) = e^{j\eta k},\ \ \eta \sim U[0, 2\pi]$$

$$A(\theta_0) = [1, \exp\{-j\tfrac{2\pi d}{\lambda}\sin(\theta_0)\}, \ldots, \exp\{-j\tfrac{2\pi d}{\lambda}(M-1)\sin(\theta_0)\}]^T$$

where θ_0 is the true angle of arrival, d is the distance between two adjacent receivers, and λ denoting the wavelength of the incident signal. The angle θ_0 is measured relative to the array normal, and $(\cdot)^T$ stands for the transpose operation. The standard MLE [11] for this case is

$$\hat{\theta}_{ML} = \max_{\theta}\{A^H(\theta)RA(\theta)\} \tag{4.6}$$

$$R = \frac{1}{N}\sum_{k=1}^{N} y(k)y^H(k)$$

Where $(\cdot)^H$ is the conjugate transpose. This solution performs a scan over the parameter space Θ to find the maximum likelihood solution. The CRLB for this problem as given in [9] (it can be shown to be equal to the one in [11]) is

$$VAR\{\hat{\theta}\} \geq \frac{12\sigma_v^2}{NM(M^2-1)(\frac{2\pi d}{\lambda})^2(\cos^2(\theta))} \tag{4.7}$$

Obviously, (4.7) may lead to the conclusion that it is impossible to efficiently estimate the DOA when $\theta \to \pm\frac{\pi}{2}$. In addition, in (4.6) we perform a search over a finite domain Θ this guarantees biased estimation if $\theta = \pm\frac{\pi}{2}$.

Consider the following modification: replace $\sin(\theta)$ in $A(\theta)$ with ρ which represents the time shift between the received signals. Although the true setup force $\rho \in [-1, 1]$, we allow $\rho \in [-3, 3]$ and assume that $\rho = h(\theta)$ as defined in (4.2). This formulation will solve the previously mentioned problem of bias when $\theta = \pm\frac{\pi}{2}$. What we actually do is extend the search range to a level where the search range edge point is beyond the actual parameter value (meaning not in Θ). The suggested extension is a standard antisymmetric extension which is made possible by the *a priori* knowledge of the parameter space Θ. So our problem has the following model

$$y(k;\rho) = A(\rho)s(k) + v(k), \;\; k = 1, 2, \ldots, N \tag{4.8}$$

$$\rho \in [-3, 3]$$

$$\rho = h(\theta) = \begin{cases} -2 - \sin(\theta), & -\frac{3\pi}{2} \le \theta < -\frac{\pi}{2} \\ \sin(\theta), & -\frac{\pi}{2} \le \theta \le \frac{\pi}{2} \\ 2 - \sin(\theta), & \frac{\pi}{2} < \theta \le \frac{3\pi}{2} \end{cases}$$

The advantage of this model is the fact that the measured time shift is a linear transformation of ρ. Hence, in the TDOA problem there are no singular points in the CRLB (for $\hat{\rho}$). The CRLB for this case can be easily derived from (4.7) and is

$$VAR(\hat{\rho}) \ge \frac{12\sigma_v^2}{NM(M^2-1)(\frac{2\pi d}{\lambda})^2} = \sigma_\rho^2 \tag{4.9}$$

At the same time we may use (4.6) to estimate $\hat{\rho}_{ML}$ as

$$\hat{\rho}_{ML} = \max_{\rho}\{A^H(\rho)RA(\rho)\} \tag{4.10}$$

In addition, for the TDOA problem all the regularity conditions holds. This means that the distribution function of the random variable $\hat{\rho}$ is asymptotically Normal with $h(\theta)$ as its mean and σ_ρ^2 (defined by the FIM and given in (4.9)) as its variance. The result is

$$\hat{\rho} \overset{a}{\sim} N(h(\theta), \sigma_\rho^2)$$

which is very close to the setup we had in the previously discussed phase estimation problem. The difference is that previously we had $\hat{\rho}$ normally distributed and now we have $\hat{\rho}$ asymptotically normal distribution. We continue by applying the same technics as in the phase estimation example to the asymptotically normal random variable $\hat{\rho}$. The result would be the asymptotical distribution of $\hat{\theta}$. Define $\hat{\theta}$ in terms of (4.3), this yields

$$\hat{\theta} = h^{-1}(\hat{\rho}) = g(\hat{\rho}) = \begin{cases} -\pi + \sin^{-1}(\hat{\rho}+2), & -3 \leq \hat{\rho} < -1 \\ \sin^{-1}(\hat{\rho}), & -1 \leq \hat{\rho} \leq 1 \\ \pi + \sin^{-1}(\hat{\rho}-2), & 1 < \hat{\rho} \leq 3 \end{cases} \tag{4.11}$$

The invariance property of the MLE assure that this will yield $\hat{\theta}_{ML}$. We may apply the distribution transformation technique, previously introduced, to the asymptotical distribution of $\hat{\rho}$ to derive the asymptotical distribution of $\hat{\theta}_{ML}$. This PDF is given by

$$f_{\hat{\theta}}(s) = \frac{\partial h(s)}{\partial s} f_{\hat{\rho}}(h(s)) =$$

$$= \begin{cases} \frac{-\cos(s)}{\sqrt{2\pi\sigma_\rho^2}} exp\{-\frac{1}{2\sigma_\rho^2}(2+\sin(s)+\sin(\theta))^2\}, & -\frac{3\pi}{2} \leq s < -\frac{\pi}{2} \\ \frac{\cos(s)}{\sqrt{2\pi\sigma_\rho^2}} exp\{-\frac{1}{2\sigma_\rho^2}(\sin(s)-\sin(\theta))^2\}, & -\frac{\pi}{2} \leq s \leq \frac{\pi}{2} \\ \frac{-\cos(s)}{\sqrt{2\pi\sigma_\rho^2}} exp\{-\frac{1}{2\sigma_\rho^2}(2-\sin(s)-\sin(\theta))^2\}, & \frac{\pi}{2} < s \leq \frac{3\pi}{2} \end{cases} \tag{4.12}$$

and

$$\sigma_\rho^2 = \frac{12\sigma_v^2}{NM(M^2-1)(\frac{2\pi d}{\lambda})^2} \tag{4.13}$$

From hereon we use (4.12) to fully characterize the proposed MLE.

Note that this way we may end up with $|\hat{\theta}| \geq \frac{\pi}{2}$, we apply the truncation proposed in section 3.1.1 to the results to assure $\hat{\theta} \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. From the asymptotical distribution of $\hat{\theta}$ we are able to compute both the bias and the MSE of the proposed MLE. Simulations of this problem for the case where $M = 2$ were conducted to demonstrate the results. The results are presented in figure 4.2. It is evident that the estimator is characterized by (4.12). In addition we may note again that the estimation bias decreases when $\theta \to \pm\frac{\pi}{2}$. The vertical lines mark the area where the performance of the MLE diverge from those predicted by the CRLB. The solid line is the bias/STD ratio and the dash-dot line is the absolute bias value indicator. In this case we see that the measure based on the estimator bias absolute value points out the requested area more accurately.

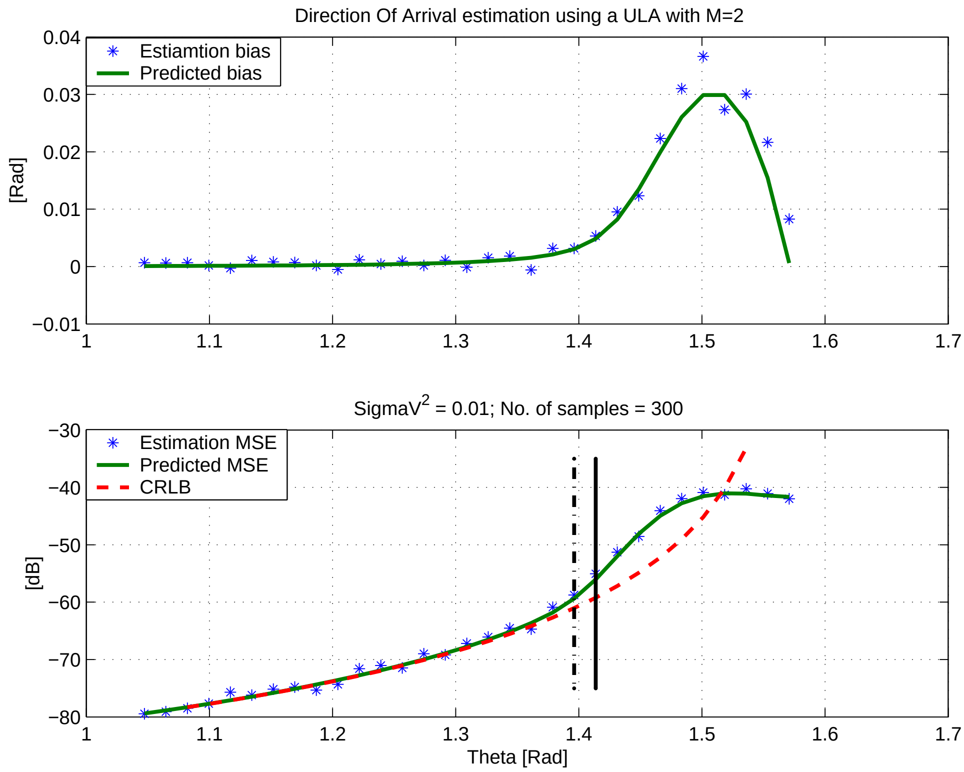

Figure 4.2: Simulation results of DOA estimation using MLE.

4.2.1 Upper bound on the MSE at the singular point

Our next objective is to find a simple upper bound on the MSE at $\theta = \pm\frac{\pi}{2}$, which is the singular points. It is clear that the symmetric nature of this problem assures that it is enough to bound one point, say $\theta = \frac{\pi}{2}$. Note that from (3.21) and (4.12) it is clear that the integrand is symmetric, hence the MSE at the singular point equals

$$MSE(\theta) = 2\int_{-\infty}^{\frac{\pi}{2}} (s - \frac{\pi}{2})^2 f_{\hat{\theta}}(s)ds \qquad (4.14)$$

Since Θ is finite we can change the lower limit in the integral to $-\frac{\pi}{2}$. In addition, it is easy to see that in the examined interval

$$\lim_{N\to\infty}(s-\frac{\pi}{2})^2\frac{\cos(s)}{\sqrt{2\pi\sigma_\rho^2}}e^{\frac{-1}{2\sigma_\rho^2}(\sin(s)-1)^2} \leq \frac{\pi^2}{\sqrt{2\pi\sigma_\rho^2}}e^{\frac{-1}{2\sigma_\rho^2}(\sin(s)-1)^2} = g(s,\sigma_\rho^2) \quad (4.15)$$

The right hand side of the latter is a Gaussian function with a maximum (with respect to s) at $s=\frac{\pi}{2}$. Since we are only interested in the left side of the function, over the interval $[-\frac{\pi}{2},\frac{\pi}{2}]$, we have

$$s_1 > s_2 \Rightarrow g(s_1,\sigma_\rho^2) > g(s_2,\sigma_\rho^2)$$

so we may conclude that $\max_{s\in[-\frac{\pi}{2},0]}\{g(s,\sigma_\rho^2)\} = g(0,\sigma_\rho^2)$ and therefore

$$\lim_{N\to\infty}(s-\frac{\pi}{2})^2\frac{\cos(s)}{\sqrt{2\pi\sigma_\rho^2}}e^{\frac{-1}{2\sigma_\rho^2}(\sin(s)-1)^2} \leq \frac{\pi^2}{\sqrt{2\pi\sigma_\rho^2}}e^{\frac{-1}{2\sigma_\rho^2}} \quad (4.16)$$

but

$$\sigma_\rho^2 = \frac{12\sigma_v^2}{NM(M^2-1)(\frac{2\pi d}{\lambda})^2}$$

so we can rewrite (4.16) as

$$\lim_{\sigma_\rho^2\to 0}(s-\frac{\pi}{2})^2\frac{\cos(s)}{\sqrt{2\pi\sigma_\rho^2}}e^{\frac{-1}{2\sigma_\rho^2}(\sin(s)-1)^2} \leq \frac{\pi^2}{\sqrt{2\pi\sigma_\rho^2}}e^{\frac{-1}{2\sigma_\rho^2}} = 0 \quad (4.17)$$

Where the latter can be easily verified through standard limit techniques. All of the above suggest the the MSE of $\hat{\theta}$ is very well approximated by the following integral

$$MSE(\theta) \cong 2\int_0^{\frac{\pi}{2}}(s-\frac{\pi}{2})^2\frac{\cos(s)}{\sqrt{2\pi\sigma_\rho^2}}e^{\frac{-1}{2\sigma_\rho^2}(\sin(s)-1)^2}ds \quad (4.18)$$

Where for large enough N the integral tail $[-\frac{\pi}{2},0]$ can be safely omitted.

Lemma 1. *The MSE of the MLE, in the DOA problem, when $\theta = \frac{\pi}{2}$, is finite and bounded by*

$$MSE(\theta) \leq \frac{\sigma_\rho \pi^4}{16\sqrt{2\pi}}(1 - e^{\frac{-1}{2\sigma_\rho^2}})$$

where σ_ρ^2, defined in (4.13), is a function of the noise level σ_v^2 the number of receivers M and the number of snapshots N and given by

$$\sigma_\rho^2 = \frac{12\sigma_v^2}{NM(M^2-1)(\frac{2\pi d}{\lambda})^2} \propto \frac{1}{SNR}$$

Proof. Assume $x \in [0, \frac{\pi}{2}]$, use the tangent of the cosine function at $x = \frac{\pi}{2}$ to get

$$\cos(x) \leq \frac{\pi}{2} - x \tag{4.19}$$

According to *claim* 2 (which follows)

$$\forall x \in [0, \frac{\pi}{2}],\ (\sin(x) - 1)^2 \geq (\frac{2}{\pi}(x - \frac{\pi}{2}))^4 \tag{4.20}$$

Next we use (4.19) and (4.20) to bound (4.18), the result is

$$2\int_0^{\frac{\pi}{2}} (s - \frac{\pi}{2})^2 \frac{\cos(s)}{\sqrt{2\pi\sigma_\rho^2}} e^{\frac{-1}{2\sigma_\rho^2}(\sin(s)-1)^2} ds \leq \frac{-2}{\sqrt{2\pi\sigma_\rho^2}} \int_0^{\frac{\pi}{2}} (s - \frac{\pi}{2})^3 e^{\frac{-8}{\sigma_\rho^2 \pi^4}(s-\frac{\pi}{2})^4} ds \tag{4.21}$$

Fortunately, the right hand side can be analytically solved. The solution is

$$\frac{-2}{\sqrt{2\pi\sigma_\rho^2}} \int_0^{\frac{\pi}{2}} (s - \frac{\pi}{2})^3 e^{\frac{-8}{\sigma_\rho^2 \pi^4}(s-\frac{\pi}{2})^4} ds = \frac{\sigma_\rho \pi^4}{16\sqrt{2\pi}}(1 - e^{\frac{-1}{2\sigma_\rho^2}}) \tag{4.22}$$

Thus the proof is complete.

□

Claim 2. *For all $x \in [0, \frac{\pi}{2}]$ the following inequality hold*

$$(\sin(x) - 1)^2 \geq (\frac{2}{\pi}(x - \frac{\pi}{2}))^4$$

Proof. To prove the claim we divide the interval $[0, \frac{\pi}{2}]$ to two adjacent sections, then prove the inequality in each section. First we take positive roots of each side of the suggested inequality, this yields

$$1 - \sin(x) \geq \frac{4}{\pi^2}(x - \frac{\pi}{2})^2 = 1 + \frac{4x^2}{\pi^2} - \frac{4x}{\pi} \tag{4.23}$$

Or equivalently we have to show

$$\frac{4x}{\pi} - \frac{4x^2}{\pi^2} \geq \sin(x) \tag{4.24}$$

Define $z1 \in [0, 0.6]$, $z2 \in [0.6, \frac{\pi}{2}]$, we show that for $z1$ the latter is true. By the tangent of the sine function at $z1 = 0$ we have $z1 > \sin(z1)$, hence

$$\frac{4z1}{\pi} - \frac{4z1^2}{\pi^2} \geq z1 \geq \sin(z1) \tag{4.25}$$

Solving the inequality on the left we get

$$z1 \leq \frac{\pi^2}{4}(\frac{4}{\pi} - 1) \cong 0.674 \tag{4.26}$$

So the claim holds for $z1$. We now show that it also holds for $z2$. We may use Taylor expansion of $\sin(z2)$ around $z2 = \frac{\pi}{2}$ to derive

$$\sin(z2) \leq 1 - \frac{1}{2}(z2 - \frac{\pi}{2})^2 + \frac{1}{24}(z2 - \frac{\pi}{2})^4 \tag{4.27}$$

Assume that the latter holds (the prove follows), we now solve the following inequality

$$1 - \sin(z2) \geq 1 - (1 - \frac{1}{2}(z2 - \frac{\pi}{2})^2 + \frac{1}{24}(z2 - \frac{\pi}{2})^4) \geq \frac{4}{\pi^2}(z2 - \frac{\pi}{2})^2 \tag{4.28}$$

or

$$\frac{1}{2}(z2-\frac{\pi}{2})^2-\frac{1}{24}(z2-\frac{\pi}{2})^4)\geq\frac{4}{\pi^2}(z2-\frac{\pi}{2})^2$$

Solving the latter yields

$$0.063\cong\frac{\pi}{2}-\sqrt{12-\frac{96}{\pi^2}}\leq z2\leq\frac{\pi}{2} \tag{4.29}$$

Combining (4.29) with (4.26) proves that the claim hold $\forall x\in[0,\frac{\pi}{2}]$. To complete the proof we need to show that (4.27) is indeed true. We use Taylor expansion of the sine function around $x=\frac{\pi}{2}$ to get

$$\begin{aligned} &\sin(x)=1-\tfrac{1}{2!}\delta^2+\tfrac{1}{4!}\delta^4-\tfrac{1}{6!}\delta^6+...\\ &\delta=x-\tfrac{\pi}{2}\end{aligned} \tag{4.30}$$

Denote

$$\begin{aligned} &\sin(x)=\textstyle\sum_{k=0}^{\infty}a_k\\ &a_k=(-1)^k\tfrac{\delta^{2k}}{(2k)!}\end{aligned} \tag{4.31}$$

we claim:

$$\sin(x)\leq\sum_{k=0}^{2N}a_k \tag{4.32}$$

First we note that this is true for the case of $N=0$

$$\sin(x)\leq a_0=1 \tag{4.33}$$

Next we prove by contradiction, assume that for some N we have

$$\sin(x)>\sum_{k=0}^{2N}a_k=C \tag{4.34}$$

The next element in the series has a negative sign so

$$\sum_{k=0}^{2N+1}a_k=C-a_{2N+1}\leq C<\sin(x) \tag{4.35}$$

Since $\delta < 1$ we have $\forall k,\ |a_k| > |a_{k+1}|$ so the next element in the series is smaller than the last one so

$$\sum_{k=0}^{2N+2} a_k = C - a_{2N+1} + a_{2N+2} \leq C < \sin(x) \tag{4.36}$$

The result is that C becomes the supremum of the series summation

$$\sum_{k=0}^{\infty} a_k \leq C < \sin(x) \tag{4.37}$$

So the series will not converge to its limit, which contradicts the basic assumption. Hence (4.27) holds, thus the proof is complete.

□

In figure 4.3 we present the upper bound from *lemma* 1 versus the actual calculated MSE for SNR ranging from (5 ∼ 90 dB), in the upper subplot. On the lower subplot we show the distance from the bound to the actual MSE. This difference is less then 4dB, and appear to converge. In *lemma* 1 we proved that

$$MSE(\theta) \leq \frac{\sigma_\rho \pi^4}{16\sqrt{2\pi}}(1 - e^{\frac{-1}{2\sigma_\rho^2}}) < \frac{\sigma_\rho \pi^4}{16\sqrt{2\pi}}$$

But for all practical purpose we note that $e^{\frac{-1}{2\sigma_\rho^2}} \cong 0$ so we do not lose much by using the simpler form to bound the MSE.

$$\boxed{MSE(\theta) \leq \frac{\sigma_\rho \pi^4}{16\sqrt{2\pi}} = \frac{\pi^4}{16\sqrt{2\pi}} \cdot \sqrt{\frac{12\sigma_v^2}{NM(M^2-1)(\frac{2\pi d}{\lambda})^2}}} \tag{4.38}$$

Corollary 3. *Since the MSE is proportional to σ_ρ it is also proportional to $N^{-\frac{1}{2}}$ and to $M^{-\frac{3}{2}}$. This suggests that*

$$\lim_{N\to\infty} MSE(\hat{\theta}) = 0,\ \lim_{M\to\infty} MSE(\hat{\theta}) = 0$$

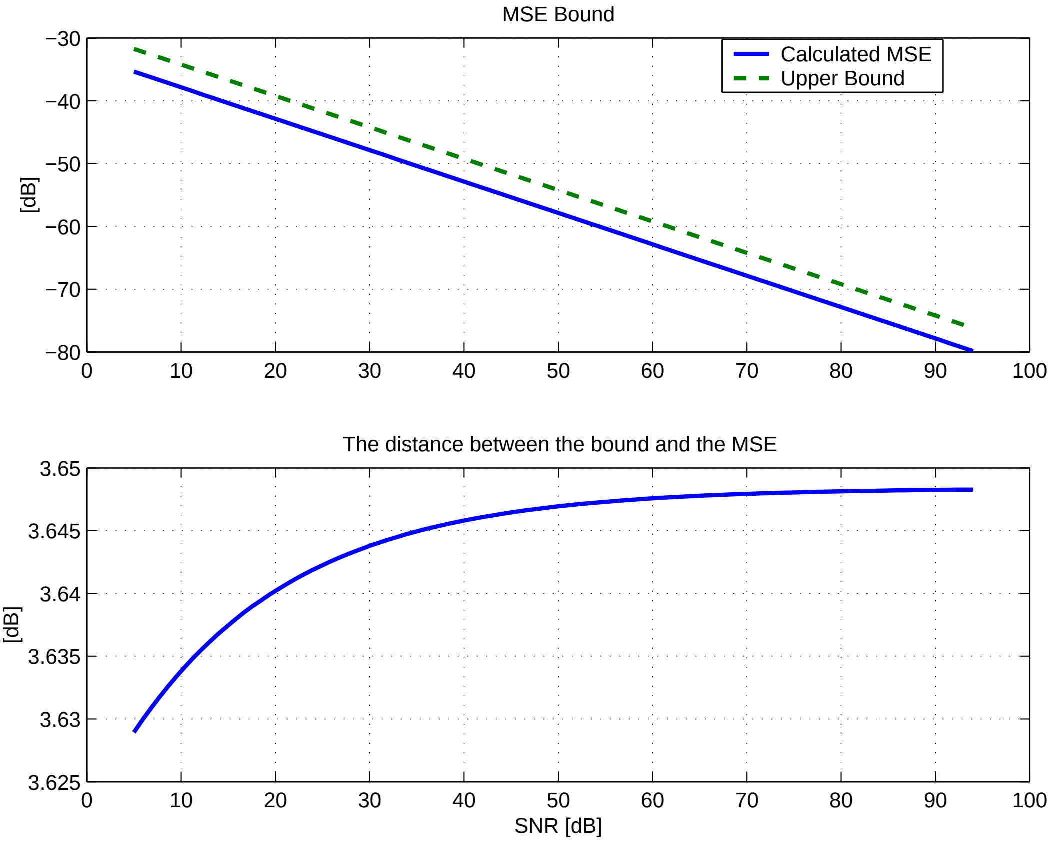

Figure 4.3: Calculated MSE versus the proposed MSE bound.

4.3 A special polynomial case

Another interesting case would be to examine a strictly monotonous and anti-symmetric function $h(\theta)$. Consider the case where $h(\theta) = \theta^{2k+1}, \quad k = 1, 2, \ldots,$ a simple polynomial antisymmetric around zero. The problem formulation will be:

$$\begin{aligned} & y_n = h(\theta) + v_n, \quad n = 1, 2, \ldots N \\ & v_n \sim N(0, \sigma_v^2), \quad \textit{i.i.d. sequence} \\ & \theta \in [-\infty, \infty] \end{aligned} \tag{4.39}$$

From (3.2) we can immediately derive the CRLB for the variance of $\hat{\theta}$

$$VAR(\hat{\theta}) \geq \frac{\sigma_v^2}{N(2k+1)^2 \theta^{4k}} \tag{4.40}$$

Note that the only singularity is at $\theta_0 = 0$, and $h(\theta)$ is antisymmetric around θ_0. From (3.20) and (3.4) we get

$$f_{\hat{\theta}}(s) = \frac{(2k+1)s^{2k}}{\sqrt{2\pi\sigma_\rho^2}} \exp\{-\frac{1}{2\sigma_\rho^2}(s^{2k+1} - \theta^{2k+1})^2\} \tag{4.41}$$

The latter enables the direct computation of the expectation and MSE of $\hat{\theta}$ (see Appendix A for the direct derivation of those expression). The result is

$$E\hat{\theta} = \lambda_E(k)\frac{\theta^{2k+1}}{\sigma_\rho^{\frac{2k}{2k+1}}} \ {}_1F_1(\frac{k}{2k+1}; \frac{3}{2}; \frac{-\theta^{2(2k+1)}}{2\sigma_\rho^2}) \tag{4.42}$$

$$MSE(\hat{\theta}) = \lambda_M(k)\sigma_\rho^{\frac{2}{2k+1}} \ {}_1F_1(\frac{-1}{2k+1}; \frac{1}{2}; \frac{-\theta^{2(2k+1)}}{2\sigma_\rho^2}) + \theta^2 - 2\theta E\hat{\theta} \tag{4.43}$$

Where

$$\lambda_E(k) = \frac{\Gamma(\frac{2k+2}{2k+1})}{\Gamma(\frac{k+1}{2k+1})} \cdot 2^{\frac{k}{2k+1}}; \quad \lambda_M(k) = \frac{\Gamma(\frac{2k+3}{2k+1})}{\Gamma(\frac{2k+2}{2k+1})2^{\frac{1}{2k+1}}}$$

and ${}_1F_1(a;b;z)$ is the confluent hypergeometric function, also known as the Kummer function. The expression in (4.42) can give some interesting insights about this problem. Consider the following expression ${}_1F_1(\frac{k}{2k+1}; \frac{3}{2}; -z)$ where $z = \frac{N\theta^{2(2k+1)}}{2\sigma_v^2} \geq 0$. Note that $z = z(\theta, N) \propto N\theta^\alpha$ and $\alpha \geq 6$. Since we know that for some $\theta >> 0$ the MLE is asymptotically efficient we may conclude that

$$\lim_{z\to+\infty} {}_1F_1(\frac{k}{2k+1}; \frac{3}{2}; -z) = \frac{\Gamma(\frac{k+1}{2k+1})}{\Gamma(\frac{2k+2}{2k+1})2^{\frac{2k}{2k+1}}} \frac{1}{z^{\frac{k}{2k+1}}} \tag{4.44}$$

Replacing the latter into (4.42) yields the following: $\forall\theta$ "far enough" from zero

$$\lim_{N\to\infty} E\hat{\theta} = \theta \tag{4.45}$$

However, by definition where $ab \neq 0$ we have ${}_1F_1(a;b;0) = 1$, therefore

$$\lim_{N\to\infty} E\hat{\theta}|_{\theta\cong 0} = \frac{\Gamma(\frac{2k+2}{2k+1})}{\Gamma(\frac{k+1}{2k+1})} \frac{\theta^{2k+1}}{\sigma_\rho^{\frac{2k}{2k+1}}} \cdot 2^{\frac{k}{2k+1}} \neq \theta \tag{4.46}$$

Since z is dominated both by N and θ it is interesting to examine the joint limit where N approaches infinity and θ approaches zero. This limit is of the form

$$\lim_{x\to 0, y\to\infty}\{x^{\alpha}y\}$$

unfortunately this limit does not exist (the latter can be easily verified by setting either $y = x^{-\alpha}$ or $y = x^{-(\alpha+1)}$ then taking the one parameter limit $x \to 0$). Nevertheless, since $x^{\alpha}y$ is a smooth function there must be a smooth transition area between the two sections. Hence there exists an area where the MLE of θ is biased as previously suggested.

4.3.1 The case of $k = 1$

It is interesting to examine the resulting PDF for the case where $k = 1$. In this case the problem formulation and the CRLB in (2.1) and (3.2) becomes

$$y_n = \theta^3 + v_n, \quad n = 1, 2, \ldots N$$

$$VAR(\hat{\theta}) \geq \frac{\sigma_v^2}{9N\theta^4}$$

The distribution function in (4.41) is given by

$$f_{\hat{\theta}}(s) = \frac{3s^2}{\sqrt{2\pi\sigma_\rho^2}} exp\{-\frac{1}{2\sigma_\rho^2}(s^3 - \theta^3)^2\} \tag{4.47}$$

and the bias and MSE are

$$E\hat{\theta} = 0.8309\frac{\theta^3}{\sigma_\rho^{\frac{2}{3}}}\ {}_1F_1(\frac{1}{3}; \frac{3}{2}; \frac{-\theta^6}{2\sigma_\rho^2}) \tag{4.48}$$

$$MSE(\hat{\theta}) = 0.8024\sigma_\rho^{\frac{2}{3}}\ {}_1F_1(\frac{-1}{3}; \frac{1}{2}; \frac{-\theta^6}{2\sigma_\rho^2}) + \theta^2 - 2\theta E\hat{\theta} \tag{4.49}$$

Note that when $\theta = 0$ from (4.48) we have $E\hat{\theta} = 0$ (as claimed in *theorem* 2), therefore $MSE(\hat{\theta}) = VAR(\hat{\theta}) = 0.8024\sigma_\rho^{\frac{2}{3}}$. This means that the suggested

MLE is locally unbiased while its variance (or MSE) is bounded. Moreover, replacing $\sigma_\rho^2 = \frac{\sigma_v^2}{N}$ we get $VAR(\hat{\theta}) = 0.8024\frac{\sigma_v^{2/3}}{N^{1/3}}$ which means that we have asymptotical convergence in L^2 of the estimation to the real value at the singular point! The PDF from (4.47) is presented in figure 4.4 for different values of θ. It can be seen that for $\theta = 0$ the PDF has two identical maximum points. Due to the antisymmetric nature of $h(\theta)$, the estimator bias reduces around $\theta = 0$. It is also obvious that the MSE of the estimation is bounded.

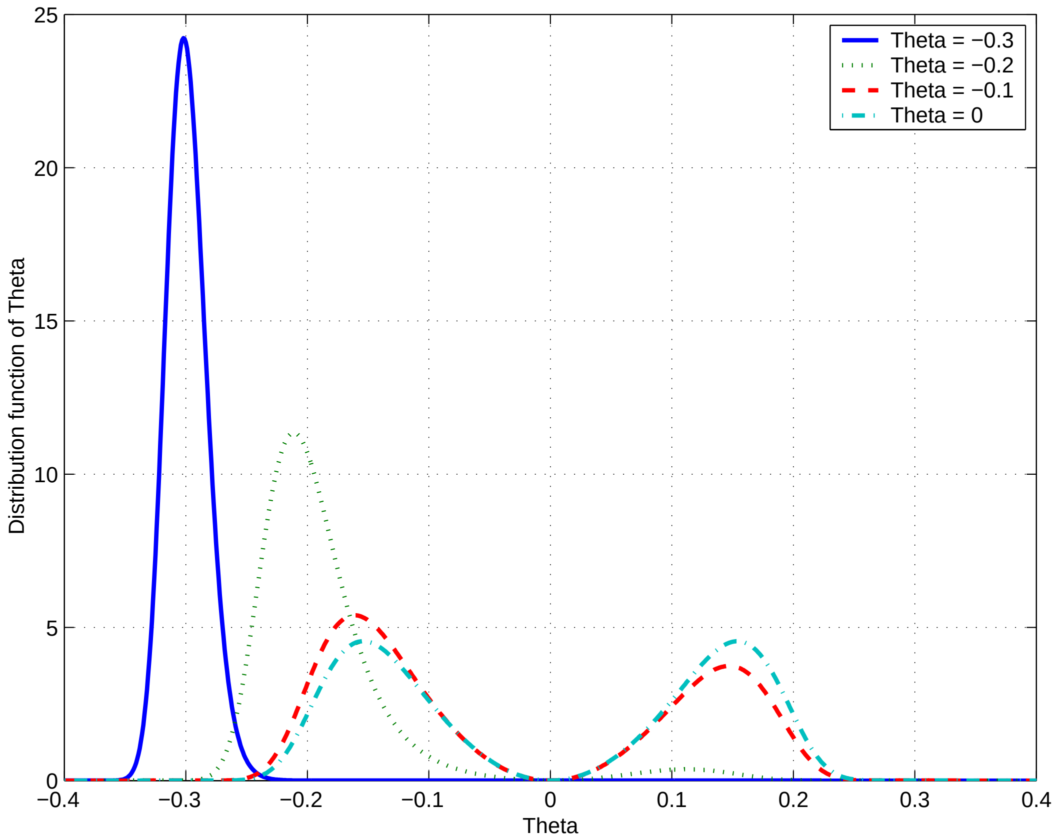

Figure 4.4: PDF of $\hat{\theta}$ for different values of θ for the special polynomial case.

Simulation results of this case with $\sigma_v^2 = 10^{-2}$ and $N = 500$ can be seen in figure 4.5. It is evident that the experiment results match the predicted MSE from (4.49). Note that where the estimator is asymptotically unbiased the MSE converge to the CRLB. As mentioned in *theorem* 2 the bias of $\hat{\theta}$ in the vicinity of $\theta = 0$ is very small, moreover it is exactly zero at the singular point while the MSE of $\hat{\theta}$ remains bounded. The vertical lines mark the area where the performance of the MLE diverge from those predicted by the CRLB. In this particular example the dash-dot vertical line, which point the area where the absolute bias derivative is greater than 0.1, marks more accurately the region where the CRLB does not predict the MLE performance.

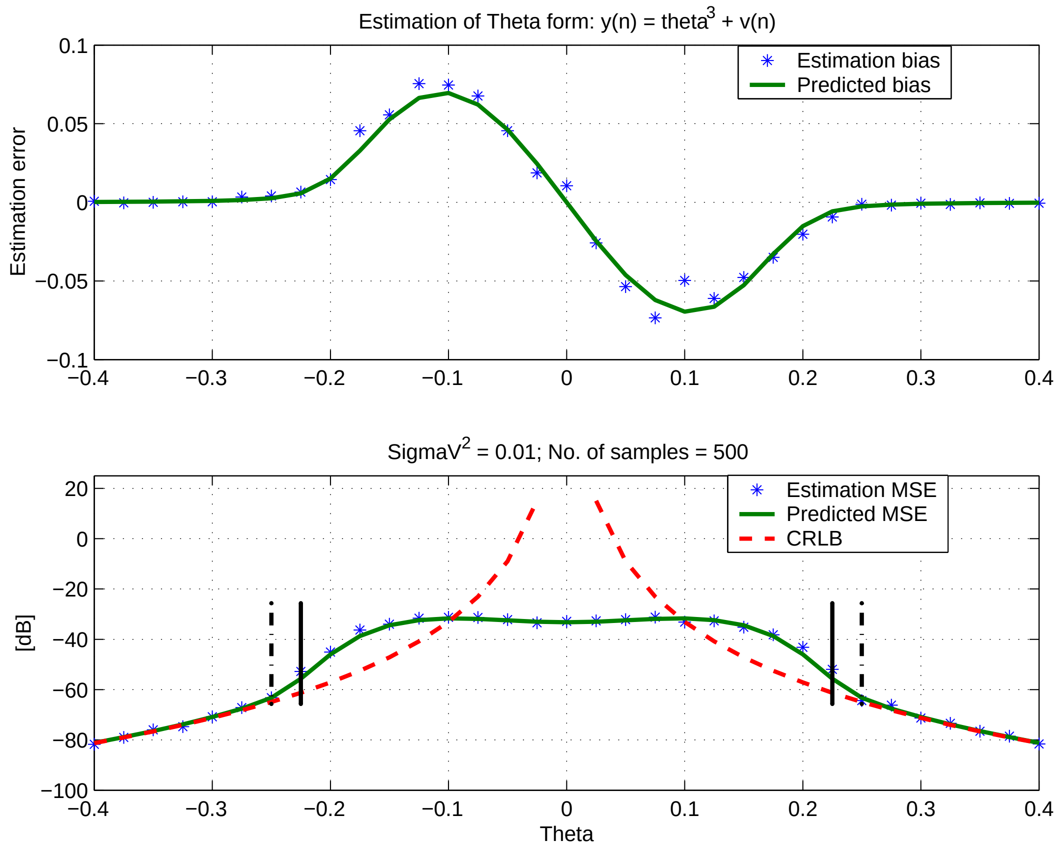

Figure 4.5: Simulation results of the special polynomial case using MLE.

Chapter 5

Summary

The primary object of this research was to obtain a statistical description for the MLE of θ for a parameter space characterized by singular points. Specifically, we were interested in the estimator performance for the DOA problem. Assuming the measurement equation is $y_n = h(\theta) + v_n, \ n = 1, 2, \ldots N$ we proposed the following scheme to find this description:

1. Define $g(x)$ as a monotonous extension of $h(\theta)$.
2. Set $\rho \equiv g(x)$ then use the equation $y_n = \rho + v_n$ to derive the MLE $\hat{\rho}$.
3. Define the MLE: $\hat{x} = g^{-1}(\hat{\rho})$, then apply the PDF transformation to derive $f_{\hat{x}}(s)$, the PDF of $\hat{x}$, from $\hat{\rho} \sim N(g(x), \sigma_\rho^2)$.
4. Define $\hat{\theta} = \min_{\theta \in \Theta} |\theta - \hat{x}|$ as the MLE of θ.
5. Use $f_{\hat{x}}(s)$ to derive $f_{\hat{\theta}}(s)$ the PDF of $\hat{\theta}_{ML}$.
6. The latter is the desired statistical model and it is used to fully characterize $\hat{\theta}_{ML}$.

The proposed scheme was used to show that for any given N there is a vicinity, around the singular point of the CRLB, where the MLE is biased and not efficient. Apparently, this turned out to be beneficial since it means that the estimator MSE is not bounded by the CRLB that approaches infinity in this vicinity. We showed that the MSE of the MLE is finite over the entire parameter space, including singular points. Moreover, we showed that for the MLE the lower bound on its MSE, as given by the *Cramér-Rao* inequality for biased estimators, is bounded and finite. In addition we showed that the MLE is locally unbiased at the singular point for the DOA problem, as well as for any other case with antisymmetric nature. Some close form expression for the MLE bias and MSE were presented for the special polynomial case. These expressions were then used to emphasize the properties of the MLE in the vicinity of a singular point. Finally we had suggested two options to point out the neighborhood of a singular point, where the MLE performance cannot be predicted by the CRLB.

5.1 Future work

It would be interesting the examine the implications of this work for the vector case, when $\underline{\theta} \in R^{(M \times 1)}$ is an unknown parameters vector. The methods suggested in this work can be applied when it is possible to replace some non-linear function $h(\underline{\theta})$ in the measurement equation with a linear matrix, thus generating an auxiliary parameter vector $\underline{\rho}$. This new vector will have asymptotical normal distribution, from which it may be possible to derive the PDF of the desired parameter vector. Problems where the FIM is an ill-conditioned matrix are common in image processing applications.

Bibliography

[1] R. A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society*, 222:309–368, 1922.

[2] C. R. Rao. Information and the accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37:81–91, 1945.

[3] Harald Cramér. *Mathematical Methods of Statistics*. Princeton University Press, Princeton, 1946.

[4] Robert J. McAulay and Edward M. Hofstetter. Barankin bounds on parameter estimation. *IEEE Transactions Information Theory*, IT-17(6):669–676, November 1971.

[5] J. Ziv and M. Zakai. Some lower bounds on signal parameter estimation. *IEEE Transactions Information Theory*, IT-15(3):386–391, May 1969.

[6] Ehud Weinstein and Anthony J. Weiss. A general class of lower bounds in parameter estimation. *IEEE Transactions on Information Theory*, 35:338–342, March 1988.

[7] A. Hero, J. Fessler, and Mohammad Usman. Exploring estimator bias-variance tradeoffs using the uniform cr bound. *IEEE Transactions on Signal Processing*, 44(8):2026–2041, August 1996.

[8] Harry L. Van Trees. *Detection, Estimation, and Linear Modulation Theory*. John Wilet and Sons, New York, 1968.

[9] Steven M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, Inc., Upper Saddle River, New Jersey, 1993.

[10] Y. Ephraim K. L. Bell and H. L. Van Trees. Explicit ziv-zakai lower bound for bearing estimation. *IEEE Transaction Signal Processing*, 44(11):2810–2824, November 1996.

[11] P. Stoica and A. Nehorai. Music, maximum likelihood, and cramer-rao bound. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(5):720–741, May 1989.

[12] P. Stoica and A. Nehorai. Music, maximum likelihood, and cramer-rao bound: Further results and comparisons. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(12):2140–2150, December 1990.

[13] P. Stoica J. Li, B. Halder and M. Vilberg. Computationally efficient angle estimation for signals with known waveforms. *IEEE Transactions on Signal Processing*, 43(9):2154–2163, September 1995.

[14] I.S. Gradshteyn and I.M. Ryzhik. *Table of Integrals, Series, and Products*. Academic Press, New York, N.Y., 1965.

[15] Wilhelm Magnus and Fritz Oberhettinger. *Formulas and Theorems for the Functions of Mathematical Physics*. Chelsea, New York, N.Y., 1954.

[16] Milton Abramowitz and Irene A. Stegun. *Table of Integrals, Series, and Products*. Dover Publicatoins, New York, 1965.

Appendix A

Derivation of the expectation and MSE for the special polynomial case

From (4.41) we have the distribution of $\hat{\theta}$ so we can compute $E\hat{\theta}$ directly by utilizing

$$E\hat{\theta} = \int_{-\infty}^{\infty} s \frac{(2k+1)s^{2k}}{\sqrt{2\pi\sigma_\rho^2}} \exp\{-\frac{1}{2\sigma_\rho^2}(s^{2k+1} - \theta^{2k+1})^2\} ds = f(\theta) \tag{A.1}$$

Dividing the integral of (A.1) to two separate integrals yield:

$$f(\theta) = \frac{(2k+1)}{\sqrt{2\pi\sigma_\rho^2}} [\int_{-\infty}^{0} (\cdot) ds + \int_{0}^{\infty} (\cdot) ds] = \alpha[I_1(\theta) + I_2(\theta)] \tag{A.2}$$

We change variables on $I_1(\theta)$ to replace integration range to $(0, \infty)$. By setting $y = -s$ we get

$$I_1(\theta) = -\int_{0}^{\infty} y^{2k+1} \exp\{-\frac{1}{2\sigma_\rho^2}(y^{2k+1} + \theta^{2k+1})^2\} dy \tag{A.3}$$

Now we may collect all items under the same integral to get

$$f(\theta) = \frac{(2k+1)}{\sqrt{2\pi\sigma_\rho^2}} \int_{0}^{\infty} s^{2k+1} [\exp\{-\frac{1}{2\sigma_\rho^2}(s^{2k+1} - \theta^{2k+1})^2\} - \tag{A.4}$$

$$\exp\{-\frac{1}{2\sigma_\rho^2}(s^{2k+1} + \theta^{2k+1})^2\}] ds = \alpha[\widetilde{I}_1 - \widetilde{I}_2]$$

Finally we change variables $x = s^{2k+1}$ to get

$$f(\theta) = \frac{1}{\sqrt{2\pi}\sigma_\rho} \int_0^\infty x^{\frac{1}{2k+1}} [\exp\{-\frac{1}{2\sigma_\rho^2}(x - \theta^{2k+1})^2\} - \exp\{-\frac{1}{2\sigma_\rho^2}(x + \theta^{2k+1})^2\}] dx \tag{A.5}$$

Some additional algebraic manipulation yields

$$f(\theta) = \frac{\exp(\frac{-\theta^{4k+2}}{2\sigma_\rho^2})}{\sqrt{2\pi}\sigma_\rho} \int_0^\infty x^{\frac{1}{2k+1}} [\exp\{-\frac{1}{2\sigma_\rho^2}(x^2 - 2x\theta^{2k+1})\} - \tag{A.6}$$

$$\exp\{-\frac{1}{2\sigma_\rho^2}(x^2 + 2x\theta^{2k+1})\}] dx = \widetilde{\alpha}[\widetilde{I_1} - \widetilde{I_2}]$$

This integral has a solution [14] and is given by

$$\begin{aligned} \widetilde{I_1} &= \Gamma(v)\sigma_\rho^v \exp(\tfrac{\theta^{4k+2}}{4\sigma_\rho^2}) D_{-v}(\tfrac{\theta^{2k+1}}{\sigma_\rho}) \\ \widetilde{I_2} &= \Gamma(v)\sigma_\rho^v \exp(\tfrac{\theta^{4k+2}}{4\sigma_\rho^2}) D_{-v}(\tfrac{-\theta^{2k+1}}{\sigma_\rho}) \\ v &= \tfrac{2k+2}{2k+1} \end{aligned} \tag{A.7}$$

Where $D_p(z)$ is the parabolic cylindrical function of the order p. So the result would be

$$f(\theta) = \frac{\Gamma(v)}{\sqrt{2\pi}} \sigma_\rho^{v-1} \exp(\frac{-\theta^{4k+2}}{4\sigma_\rho^2})[D_{-v}(\frac{-\theta^{2k+1}}{\sigma_\rho}) - D_{-v}(\frac{\theta^{2k+1}}{\sigma_\rho})] \tag{A.8}$$

Note that we arrived at the structure $D_p(-z) - D_p(z)$ from [15] we know that

$$D_p(z) = 2^{\frac{p}{2}} e^{\frac{-z^2}{4}} [\frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{1-p}{2})}\ {}_1F_1(\frac{-p}{2}; \frac{1}{2}; \frac{z^2}{2}) + \frac{z}{\sqrt{2}} \frac{\Gamma(\frac{-1}{2})}{\Gamma(\frac{-p}{2})}\ {}_1F_1(\frac{1-p}{2}; \frac{3}{2}; \frac{z^2}{2})] \tag{A.9}$$

$$= A + B$$

Therefore $D_p(-z) - D_p(z) = -2B$ so we get

$$f(\theta) = -\frac{\Gamma(v)\Gamma(\frac{-1}{2})}{\Gamma(\frac{v}{2}) 2^{\frac{v}{2}} \sqrt{\pi}} \sigma_\rho^{v-2} \theta^{2k+1} \exp(\frac{-\theta^{4k+2}}{2\sigma_\rho^2})\ {}_1F_1(\frac{1+v}{2}; \frac{3}{2}; \frac{\theta^{4k+2}}{\sigma_\rho^2}) \tag{A.10}$$

But ${}_1F_1(a;b;z) = e^z\ {}_1F_1(b-a;b;-z)$, replacing that into the latter yields

$$f(\theta) = -\frac{\Gamma(\frac{2k+2}{2k+1})\Gamma(\frac{-1}{2})}{\Gamma(\frac{k+1}{2k+1})2^{\frac{k+1}{2k+2}}\sqrt{\pi}}\ \frac{\theta^{2k+1}}{\sigma_\rho^{\frac{2k}{2k+1}}}\ {}_1F_1(\frac{k}{2k+1};\frac{3}{2};\frac{-\theta^{4k+2}}{\sigma_\rho^2}) \tag{A.11}$$

Using the reflection identity $-\Gamma(\frac{-1}{2}) = 2\Gamma(\frac{1}{2})$, but $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ so we get the desired result

$$E\hat{\theta} = \lambda(k)\frac{\theta^{2k+1}}{\sigma_\rho^{\frac{2k}{2k+1}}}\ {}_1F_1(\frac{k}{2k+1};\frac{3}{2};\frac{-\theta^{2(2k+1)}}{2\sigma_\rho^2}) \tag{A.12}$$

Where

$$\lambda(k) = \frac{\Gamma(\frac{2k+2}{2k+1})}{\Gamma(\frac{k+1}{2k+1})} \cdot 2^{\frac{k}{2k+1}}$$

Properties of the Kummer function can also be found in [16]. By a very similar manner we will compute (4.43).

$$E\{(\hat{\theta}-\theta)^2\} = E\{\hat{\theta}^2 - 2\theta\hat{\theta} + \theta^2\} = E\{\hat{\theta}^2\} + \theta^2 - 2\theta E\hat{\theta} \tag{A.13}$$

Since $\hat{\theta}$ is the only random variable and its expectation is given in (A.12) we only have to calculate the second moment of $\hat{\theta}$. Define

$$m_2(\hat{\theta}) = \int_{-\infty}^{\infty} s^2 \frac{\partial h(s)}{\partial s} f_{\hat{\rho}}(h(s))ds \tag{A.14}$$

Then $E\{(\hat{\theta}-\theta)^2\} = m_2(\hat{\theta}) + \theta^2 - 2\theta E\hat{\theta}$. We compute $m_2(\hat{\theta})$ directly from (A.14) by the same methods we used for the computation of $E\hat{\theta}$. From (4.41) we have

$$m_2(\hat{\theta}) = \int_{-\infty}^{\infty} s^2 \frac{(2k+1)s^{2k}}{\sqrt{2\pi\sigma_\rho^2}} \exp\{-\frac{1}{2\sigma_\rho^2}(s^{2k+1} - \theta^{2k+1})^2\}ds \tag{A.15}$$

Using the manipulations presented in (A.2)-(A.4) we get

$$m_2(\hat{\theta}) = \frac{(2k+1)}{\sqrt{2\pi\sigma_\rho^2}} \int_0^{\infty} s^{2(k+1)}[\exp\{-\frac{1}{2\sigma_\rho^2}(s^{2k+1} - \theta^{2k+1})^2\} + \tag{A.16}$$

$$\exp\{-\frac{1}{2\sigma_\rho^2}(s^{2k+1}+\theta^{2k+1})^2\}]ds$$

Substituting $x = s^{2k+1}$ we get the following integral

$$m_2(\hat{\theta}) = \frac{(2k+1)}{\sqrt{2\pi\sigma_\rho^2}} \int_0^\infty x^{\frac{2(k+1)}{2k+1}} [\exp\{-\frac{1}{2\sigma_\rho^2}(x-\theta^{2k+1})^2\}+ \tag{A.17}$$

$$\exp\{-\frac{1}{2\sigma_\rho^2}(x+\theta^{2k+1})^2\}]\frac{dx}{(2k+1)x^{\frac{2k}{2k+1}}}$$

Collecting some terms and rearranging the latter we get

$$m_2(\hat{\theta}) = \frac{exp\{\frac{-\theta^{2(2k+1)}}{2\sigma_\rho^2}\}}{\sqrt{2\pi\sigma_\rho^2}} \int_0^\infty x^{\frac{2}{2k+1}} [\exp\{-\frac{1}{2\sigma_\rho^2}(x^2-2\theta^{2k+1}x)\}+ \tag{A.18}$$

$$\exp\{-\frac{1}{2\sigma_\rho^2}(x^2+2\theta^{2k+1}x)\}]dx$$

But from the usage of (A.7)-(A.9) we get for $m_2(\hat{\theta})$ an expression of the form $D_p(-z) + D_p(z) = 2A$ or

$$m_2(\hat{\theta}) = \frac{\Gamma(v)\Gamma(\frac{1}{2})}{\Gamma(\frac{1+v}{2})2^{\frac{v-1}{2}}\sqrt{\pi}}\sigma_\rho^{v-1}\exp(\frac{-\theta^{4k+2}}{2\sigma_\rho^2})\ {}_1F_1(\frac{v}{2};\frac{1}{2};\frac{\theta^{4k+2}}{\sigma_\rho^2}) \tag{A.19}$$

Or after the final manipulation (as in (A.10)-(A.12)) we get the desired result

$$m_2(\hat{\theta}) = \lambda_M(k)\sigma_\rho^{\frac{2}{2k+1}}\ {}_1F_1(\frac{-1}{2k+1};\frac{1}{2};\frac{-\theta^{2(2k+1)}}{2\sigma_\rho^2}) \tag{A.20}$$

Where

$$\lambda_M(k) = \frac{\Gamma(\frac{2k+3}{2k+1})}{\Gamma(\frac{2k+2}{2k+1})2^{\frac{k}{2k+1}}}$$

Now we can use (A.13) to get (4.43).