University of Alberta

Evaluation of the Performance Characteristics of Parameter Estimation Techniques Based on Cramer-Rao Lower Bounds

by

Zhili Chen ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **Master of Science**.

Department of Electrical and Computer Engineering

Edmonton, Alberta
Spring 2006

Your file Votre référence
ISBN: 0-494-13803-3
Our file Notre référence
ISBN: 0-494-13803-3

To my wife and my parents.

Abstract

This thesis is concerned with the evaluation of the performance characteristics of parameter estimation techniques based on the classical Cramer-Rao lower bound (CRLB) for unbiased estimation, and uniform CRLB (UCRLB) for biased estimation. These evaluations are conducted in the context of both single-tone and multiharmonic sinusoidal signals contaminated by AWGN.

Three unbiased frequency estimators are studied in the case of contaminated single-tone signals. The linear regression estimation (LRE) algorithm is derived in detail for the full range of SNR, and its performance is compared against CRLB together with the other estimators. The parameter estimation for harmonic/subharmonic sinusoidal signal is studied in detail under the assumption that noise variance is unknown a priori. The CRLBs are derived and verified by simulation. For biased estimation, three SNR estimators are evaluated by applying the UCRLBs, and a modified algorithm for estimating the bias gradient is presented.

Acknowledgements

I would like to thank my supervisor, Dr. B. Nowrouzian, who has given me countless advice, guidance, continuous encouragement and financial support. He has played an important role in my final work. Without him, I would not be here.

I would also like to thank my former supervisor, Dr. C. J. Zarowski, for his academic guidance and financial support during the initial stages of my graduate study.

I wish to thank Dr. Bin Han and Dr. Venkata Dinavahi for serving on my thesis committee.

Finally, my deepest gratitude goes to my wife, Jing, for her love, encouragement as well as support at home.

Contents

List of Figures

Acronyms

AWGN Additive White Gaussian Noise

BLUE Best Liner Unbiased Estimator

BUE Best Unbiased Estimator

CRLB Cramer-Rao Lower Bound

DTFT Discrete Time Fourier Transform

FFT Fast Fourier Transform

FIM Fisher Information Matrix

i.i.d. Independent and Identically Distributed

LLF Log Likelihood Function

LMMSE Linear Minimum Mean Square Error

LPE Linear Prediction Estimator

LSE Least Squares Estimator

MAP Maximum A Posteriori

MICA Modified Iterative Cosinor Algorithm

ML Maximum Likelihood

MLE Maximum Likelihood Estimator

MMSE Minimum Mean-Squared Error

MSE Mean-Squared Error

MUSIC Multiple Signal Classification

MVUE Minimum Variance Unbiased Estimator

PDF Probability Density Function

PML Penalized Maximum Likelihood

RHS Right-Hand-Side

SAGE Space Alternating Generalized Expectation-Maximization

SNR Signal-to-Noise Ratio

SPECT Single Photon Emission Computed Tomography

UCRLB Uniform CRLB

WLSE Weighted Least-Squares Estimation

Notations

- X, Y : Uppercase letters are used to represent random variables, random vectors, and random processes.
- x, y : Lowercase letters represent the realizations of the random variables, random vectors, and random processes.
- $\mathbf{R}$: the set of real numbers.
- $\mathbf{\Theta}$: the N-dimensional parameter space.
- $\mathbf{R}^{N \times P}$: the $N \times P$ real-valued matrix space.
- $\delta(m)$: the Kronecker delta sequence.
- $H\{\cdot\}$: the discrete-time Hilbert transform operator.
- $\underline{\theta}$: a parameter vector $\underline{\theta} = [\theta_1, \theta_2, \cdots, \theta_p]^T$.
- $\hat{\underline{\theta}} = \hat{\underline{\theta}}(X)$: an estimator of $\underline{\theta}$.
- $\|\underline{d}\|$: the Euclidean norm of vector $\underline{d}$.
- $f(\underline{X}; \underline{\theta})$: the probability density function of $\underline{X}$ parameterized by the parameter vector $\underline{\theta}$.
- $E[\cdot]$: the expectation operation.
- $\mathcal{N}(0, \sigma^2)$: the Gaussian probability density function with a mean of zero and a variance of σ^2.
- $\underline{m}(\underline{\theta})$: the mean vector $E\left(\hat{\underline{\theta}}\right)$ of $\hat{\underline{\theta}}$, where $\underline{\theta}$ is the true underlying parameter.

- $\underline{b}(\underline{\theta})$: the bias of $\hat{\underline{\theta}}$, i.e. $\underline{m}(\underline{\theta}) - \hat{\underline{\theta}}$.

- $erf(x)$: the error function.

- $(\cdot)^T$: matrix or vector transpose operation.

- $(\cdot)^{-1}$: matrix inverse operation.

- $cov_{\underline{\theta}}\left(\hat{\underline{\theta}}\right)$ or $\mathbf{C}_{\hat{\theta}}$: the $N \times N$ covariance matrix of $\hat{\underline{\theta}}$, $E\left[\left(\hat{\underline{\theta}} - \underline{m}(\underline{\theta})\right)\left(\hat{\underline{\theta}} - \underline{m}(\underline{\theta})\right)^T\right]$.

- $var_{\underline{\theta}}\left(\hat{\underline{\theta}}\right)$: the variance of $\hat{\underline{\theta}}$.

- $F(\underline{\theta})$: the $N \times N$ Fisher information matrix for vector $\underline{\theta}$.

- δ : a user-specified upper bound on the length of the bias-gradient vector.

- ∇ : the gradient operator $\nabla = \left[\frac{\partial}{\partial\theta_1}, \cdots, \frac{\partial}{\partial\theta_p}\right]^T$.

- $B(\underline{\theta}, \delta)$: the uniform Cramer-Rao lower bound on $var_{\underline{\theta}}\left(\hat{\underline{\theta}}\right)$ for estimators with bias $b(\underline{\theta})$ such that $\|\nabla b\| < \delta$.

- $\underline{d}^T_{\min}$: the minimizing bias-gradient vector which characterizers $B(\underline{\theta}, \delta)$.

- λ : a scaling constant determined by the solution to the constraint equation on the bias gradient (Eqn. (4.13)).

- $I_k(x)$: the kth order modified Bessel function of the first kind.

- ${}_1F_1(a, b, x)$: the confluent hypergeometric function of the first kind.

- $\Gamma(\cdot)$: the Gamma Function.

- I_N : the $N \times N$ identity matrix.

- P_H : an orthogonal projection matrix $P_H = H\left(H^T H\right)^{-1} H^T \in \mathbf{R}^{N \times N}$.

Chapter 1

Introduction

1.1 Introduction to Parameter Estimation

In signal processing systems, an important task is to extract information from the observed data contaminated by noise. Usually, one needs to estimate the value of one parameter or the values of several parameters from the noisy data. For example, in the field of communications, one needs to estimate the carrier frequency of a modulated signal so that the signal can be demodulated to baseband [1]. In radar processing systems, one is interested in determining the position of an aircraft, as in the case of airport surveillance radar [2]. In speech recognition, it is needed to extract the spectral envelop of the voice by using a model of speech called linear prediction coding which is characterized by the model parameters [3]. In all these systems, one is faced with the problem of extracting values of parameters based on the observed data set. This is the problem of parameter estimation.

Let us take a look at a simple case in which the DC value is to be determined in a noisy data set:

$$x[n] = A + w[n], \ n = 0, 1, \cdots, N-1 \tag{1.1}$$

where the parameter A represents the DC value and w represents additive white Gaussian noise (AWGN) with a Gaussian probability density function (PDF) of zero mean and of variance σ^2, denoted by $\mathcal{N}(0, \sigma^2)$. Therefore, based on the observed data set $\underline{X} = [x[0], x[1], \cdots, x[N-1]]^T \subseteq \mathbf{R}^N$, one would like to estimate A. Intuitively, it would be reasonable to estimate A as

$$\hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x[n], \tag{1.2}$$

i.e. as the sample mean of the data.

1.1.1 Mathematical Model for Estimation

The choice of an estimator that will perform well for a particular application depends upon many considerations. Of primary concern is the selection of a suitable data model. The data model should be comprehensive enough to embody the principal features of the data, but at the same time simple enough to allow for an estimator that is optimal and easily implemented.

Mathematically, one could conduct an experiment and obtain N independent observations or a N-point observed data set $\underline{X}$ that depends on an unknown but deterministic parameter vector $\underline{\theta} = [\theta_1, \theta_2, \cdots, \theta_p]^T$ taking on values in a parameter space $\Theta \subseteq \mathbf{R}^p$. Based on the data, $\underline{\theta}$ can be determined or an estimator can be defined as [4] :

$$\hat{\underline{\theta}} = g(x[0], x[1], \cdots, x[N-1]), \tag{1.3}$$

where g is some function. As the data set consists of random variables and the data size is definite, all the theory and techniques of statistical estimation can be used to address the problem.

1.1.2 Bias, Variance and MSE

As an estimator itself is a random variable, its performance can only be described completely statistically or by its PDF. Suppose now that there is an unknown real parameter $\theta \in \mathbf{R}$. An estimator of θ is actually a real-valued statistic $\hat{\theta}$, characterized by a distribution, a mean, a variance, and so on. The mean value of an estimator is defined by

$$m_\theta = E\left[\hat{\theta}\right], \tag{1.4}$$

where $E[\cdot]$ denotes the expectation operation.

Bias is defined as the expected value of the error, the difference due to mismatch between the estimation algorithm and the actual value of θ

$$b_\theta(\hat{\theta}) = E\left[\hat{\theta} - \theta\right] = m_\theta - \theta. \tag{1.5}$$

An estimator is said to be unbiased if the bias is 0 for all values of θ, or equivalently, if the expected value of the estimator is the parameter being estimated.

The variance σ_θ^2 of the estimation arises from the statistical fluctuations due to statistical uncertainty in the observed data $\underline{X}$ and is defined as

$$\sigma_\theta^2 = E\left[\left(\hat{\theta} - m_\theta\right)^2\right]. \tag{1.6}$$

To measure the overall quality of an estimator, the mean square error (MSE) is often computed. It is well known that the MSE is a function of both the bias and the variance in accordance with

$$\begin{aligned} MSE(\hat{\theta}) &= E\left[\left(\hat{\theta} - \theta\right)^2\right] \\ &= E\left[\left(\hat{\theta} - m_\theta + m_\theta - \theta\right)^2\right] \\ &= E\left[\left(\hat{\theta} - m_\theta\right)^2\right] + (m_\theta - \theta)^2 \\ &= \sigma_\theta^2 + b_\theta^2(\hat{\theta}). \end{aligned} \tag{1.7}$$

In particular, if the estimator is unbiased, then the mean square error is simply the variance of the estimator.

1.1.3 Cramer-Rao Lower Bound

An important question in parameter estimation is whether an estimator has certain desired properties, in particular, if it converges to the actual value of the unknown parameter it is estimating. One typical property of an estimator is unbiasedness, meaning that on the average, the estimator hits its target. From (1.7), the overall quality of an estimator is equal to the estimation variance if the bias is zero. Therefore, for unbiased estimation, the natural question is whether the estimator shares some optimality properties in terms of its sampling variance. The best estimator should be the one with the smallest possible variance which is bounded by some theoretical lower bound.

There is such a bound, well known as the Cramer-Rao lower bound (CRLB) [4], giving the minimal achievable variance for any unbiased estimator. Suppose the goal is to estimate a vector parameter $\underline{\theta}$ which is deterministic but unknown, and assume that the estimator $\hat{\underline{\theta}}$ is unbiased. The CRLB bound will allow us to place a bound on the variance of each element. Let $f\left(\underline{X};\underline{\theta}\right)$ be the PDF of measured data which is parameterized by the parameter vector $\underline{\theta}$, where the semicolon is used to denote this dependence. The following theorem is a definition of CRLB bound.

Theorem 1.1 (Cramer-Rao Lower Bound) Let us assume that the PDF satisfies the "regularity" conditions

$$E\left[\frac{\partial \ln f\left(\underline{X};\underline{\theta}\right)}{\partial \underline{\theta}}\right] = 0, \quad \text{for all } \underline{\theta} \tag{1.8}$$

where the expectation is taken with respect to $f\left(\underline{X};\underline{\theta}\right)$. *Then, the covariance matrix of any unbiased estimator* $\hat{\theta}$ *satisfies*

$$\mathbf{C}_{\hat{\theta}} - F^{-1}\left(\underline{\theta}\right) \geqslant 0 \tag{1.9}$$

where the $\geqslant 0$ *sign is interpreted as meaning that the matrix is positive semidefinite. The Fisher information matrix* $F\left(\underline{\theta}\right)$ *is given as*

$$\left[F(\underline{\theta})\right]_{i,j} = -E\left[\frac{\partial^2 \ln f\left(\underline{X};\underline{\theta}\right)}{\partial \theta_i \partial \theta_j}\right] \tag{1.10}$$

for $i = 1, 2, \ldots, p$; $j = 1, 2, \ldots, p$, *where the derivatives are evaluated at the true value* *of $\underline{\theta}$ and the expectation is taken with respect to $f(\underline{X};\underline{\theta})$.*

In the above theorem, $\dfrac{\partial \ln f(\underline{X};\underline{\theta})}{\partial \underline{\theta}}$ is called the score function. It is defined to be the gradient of log-likelihood function $\ln f(\underline{X};\underline{\theta})$, given a statistical model with a PDF of $f(\underline{X};\underline{\theta})$.

Although many variance bounds exist [5–8], the CRLB bound is by far the easiest to derive. It sets a lower bound on the variance of any unbiased estimator. This can be extremely useful in several ways:

1. It might allow one to assert that an estimator is the minimum variance unbiased (MVU) estimator. If an estimator is found to achieve the CRLB bound, then it is an MVU estimator.

2. It provides a benchmark against which one can compare the performance of any unbiased estimator. If the variance is close to the CRLB bound, this means the unbiased estimator is "good".

3. It alerts one to the physically impossible task of finding an unbiased estimator whose variance is less than the bound. That is, it is physically impossible to find an unbiased estimator whose variance outperforms the CRLB bound. This is very useful in feasibility studies.

4. The theory behind the CRLB bound can tell us if an estimator exists that achieves the bound.

In parameter estimation problems, the goal is to obtain an optimal estimator, if no such estimator exists, then one can resort to an approximately optimal estimator. If this is still not available, a suboptimal estimator will need to be found.

1.1.4 Overview to Existing Estimation Approaches

Basically, there are two different classes of estimation approaches available for the parameter estimation.

1. Classical approaches to estimation. In these approaches, the unknown $p \times 1$ parameter vector $\underline{\theta}$ is assumed to be a deterministic constant vector, and the data

information is characterized by the PDF $f(\underline{X};\underline{\theta})$, where the PDF is functionally dependent on $\underline{\theta}$.

2. Bayesian approaches. In contrast to classical approaches, here the parameter vector $\underline{\theta}$ is assumed to be a random vector. This approach augments the data information with a prior PDF $f(\underline{\theta})$ which describes the knowledge about $\underline{\theta}$ (before any data are observed). This is summarized by the joint PDF $f(\underline{X},\underline{\theta})$ or, equivalently, by the conditional PDF of $f(\underline{X}|\underline{\theta})$ (data information) and the prior PDF $f(\underline{\theta})$ (prior information).

This thesis is mainly concerned with the classical approaches of parameter estimation. Before proceeding further, some concepts regarding the estimation accuracy need to be introduced.

- **Efficiency**: An estimator is said to be an efficient estimator of a parameter if:

  a) it is unbiased,

  b) it attains CRLB bound.

- **Asymptotically unbiased**: an estimator is asymptotically unbiased if $E\left[\hat{\theta}\right] \rightarrow \theta$, as data size $N \rightarrow \infty$.

- **Asymptotically efficient**: an estimator asymptotically efficient if $var\left[\hat{\theta}\right] \rightarrow CRLB$, as data size $N \rightarrow \infty$.

- **Consistency**: an estimator $\hat{\theta}$ is consistent if, given any $\epsilon > 0$,

$$\lim_{N\rightarrow\infty} \Pr\left\{|\hat{\theta}-\theta| > \epsilon\right\} = 0$$

  where N is the sample data size.

- **Threshold effect**: the MSE rises very rapidly as SNR decreases within a low range of SNR. The SNR at which this effect is first apparent is called the threshold.

Classical Estimation Approaches

In the classical estimation, the parameters of interest are assumed to be deterministic but unknown. For this category, the following approaches are available [4, 9]: minimum variance unbiased estimator (MVUE), best linear unbiased estimator (BLUE), maximum likelihood estimator (MLE), least squares estimator (LSE) and the method of moments, etc. In the following, the estimation methods are introduced respectively.

Minimum Variance Unbiased Estimation The MVUE estimator is an unbiased estimator with minimum variance for all θ. That is, the variance of MVUE for each component is minimum among all unbiased estimators. In practice, it is desirable to find the estimator whose bias is to be zero and the estimator minimizes the variance. Sometime it is called the best unbiased estimator (BUE), because it is the "best" estimator one can get using only unbiased estimators. For example, when trying to estimate the mean of a random variable with Gaussian distribution, given a measurement data set $\underline{X} = [\ x(0), x(1), \cdots, x(N-1)]^T \subseteq \mathbf{R}^N$, two estimators can be chosen. One is the sample mean $\widehat{\overline{x}} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$ and the other is the sample median $\widehat{\tilde{x}}$, which is the middle of a distribution. They are both natural estimators of the mean of a normal population. It can be proven that both are unbiased, but $\widehat{\overline{x}}$ has a smaller variance than the estimator of median $\widehat{\tilde{x}}$. In fact, $\widehat{\overline{x}}$ is the MVUE for the random variable mean.

In MVU estimation, $\hat{\theta}$ achieves the CRLB bound, and is therefore said to be efficient. However, It is not always possible to find MVUE estimators. Sometimes the efficient estimator may not exist, and hence this approach may fail.

Best Linear Unbiased Estimator An estimator is called best linear unbiased estimator (BLUE) if it is

a) Linear (linear function of a random variable);

b) Unbiased.

c) Efficient.

A BLUE estimator can be defined as

$$\hat{\theta} = \sum_{n=0}^{N-1} a_n x[n] \tag{1.11}$$

where $x[n]$ is the data set whose PDF $f(\underline{X}; \underline{\theta})$ depends on an unknown parameter θ. The estimator has the minimum variance of all unbiased estimators that are linear in $\underline{X}$. If the data are Gaussian, then the BLUE is also the MVU estimator.

Maximum Likelihood Estimator This estimator is defined as the value of $\underline{\theta}$ that maximizes the likelihood function, which is functionally the same in form as the PDF of the sample data. Actually, the likelihood function L is the function obtained by reversing the roles of $\underline{X}$ and $\underline{\theta}$ in the PDF; that is, $\underline{\theta}$ is viewed as the variable and $\underline{X}$ as the given information (which is precisely the point of view in estimation):

$$L(\underline{\theta}|\underline{X}) = f(\underline{X}|\underline{\theta}) \tag{1.12}$$

for $\underline{\theta}$ in parameter space $\boldsymbol{\Theta}$ and $\underline{X}$ in data set $\mathbf{S}$.

The method is intuitively appealing, as the values of the parameters to be estimated are the ones that would have most likely produced the observed data. Since the natural logarithm function ln is monotonically increasing function of its argument, the maximum value of $L(\underline{\theta}|\underline{X})$, if it exists, will occur at the same points as the maximum value of $ln[L(\underline{\theta}|\underline{X})]$. This latter function is called the log likelihood function and in many cases is easier to work with than the likelihood function (usually because the density $f(\underline{X}|\underline{\theta})$ has a product structure).

MLE is not optimal in general. Under certain conditions on the PDF, however, the MLE is efficient for large data sets as $N \to \infty$ (asymptotically). Hence, asymptotically it is the MVU estimator.

From a statistical point of view, this method is considered to be more robust (with some exceptions) and yields estimators with good statistical properties. In other words, MLE methods are versatile and apply to most models and to different types of data. In addition, they provide efficient methods for quantifying uncertainty through confidence bounds. The MLE estimators become minimum variance unbiased estimators as the sample size increases.

There are only two drawbacks to MLE's, but they are important ones. Firstly, maximum likelihood estimates can be heavily biased for small number of samples. The optimality properties may not apply for small number of samples. Secondly, although the idea behind the maximum likelihood estimation is simple, the implementation is mathematically intense. Calculating MLE's often requires specialized software for solving complex non-linear equations. Using today's computer power, however, mathematical complexity is not a big obstacle.

Least Squares Estimator The least squares approach to parameter estimation chooses $\underline{\theta}$ to minimize the sum of the squared deviations between the observed responses and the functional portion of the model

$$J\left(\underline{\theta}\right) = \sum_{n=0}^{N-1} \left(x\left[n\right] - s\left[n\right]\right)^2, \tag{1.13}$$

where the signal s depends on $\underline{\theta}$, and x is the signal plus noise w: $x\left[n\right] = s\left(n; \underline{\theta}\right) + w\left[n\right]$. The minimum value of $J\left(\underline{\theta}\right)$ is called the least square error. For example, let us consider the straight-line model,

$$x\left[n\right] = \theta_0 + \theta_1 y\left[n\right] + w\left[n\right], \tag{1.14}$$

where $y[n]$ is a known sequence, θ_0 and θ_1 are the two parameters to be estimated, which represent the intercept and slope of the line respectively. For this model the least squares estimates of the parameters would be computed by minimizing

$$J\left(\underline{\theta}\right) = \sum_{n=0}^{N-1} \left[x\left[n\right] - \left(\hat{\theta}_0 + \hat{\theta}_1 y[n]\right)\right]^2. \tag{1.15}$$

By taking partial derivatives of $J(\underline{\theta})$ with respect to $\hat{\theta}_0$ and $\hat{\theta}_1$, by setting each partial derivative to zero, and by solving the resulting system of two equations with two unknowns, one gets the following estimators for the parameters:

$$\hat{\theta}_1 = \frac{\sum_{n=0}^{N-1}(x[n]-\bar{x})(y[n]-\bar{y})}{\sum_{n=0}^{N-1}(y[n]-\bar{y})^2}, \tag{1.16}$$

$$\hat{\theta}_0 = \bar{x}-\hat{\theta}_1\bar{y}, \tag{1.17}$$

where $\bar{x}=(1/N)\sum_{n=0}^{N-1}x[n]$, $\bar{y}=(1/N)\sum_{n=0}^{N-1}y[n]$.

Minimizing a LSE error criterion does not in general translate into minimizing the estimation error. Also, if w is a Gaussian random vector, then the LSE is equivalent to the MLE.

Method of Moments This method is based on the solution of a theoretical equation involving the moments of a PDF. It is a way of generating estimators: set the distribution moments equal to the sample moments, and solve the resulting equations for the parameters of the distribution. For example, let us consider the kth moment of a random variable with zero mean given data set $\underline{X}$. The kth statistical moment is

$$\mu_k = E_{\underline{\theta}}[x^k[n]] \tag{1.18}$$

and the kth sample moment is

$$m_k = (1/N)\sum_{n=0}^{N-1}x^k[n]. \tag{1.19}$$

Then, one can choose as estimates those values of the parameters that are solutions of the equations

$$\mu_k = m_k, \tag{1.20}$$

where $k=1,2,\cdots,p$.

When moment methods are available, they have the advantage of simplicity, that is, easy to determine and simple to implement. The disadvantage is that they are often not available and they do not have the desirable optimality properties of MLE and LSE estimators. The primary use of moment estimates is as starting values for the more precise maximum likelihood and least squares estimates.

Although for the moment method the estimator has no optimality properties, it can be useful if the data record is long enough.

Bayesian Estimation Approach [4]

In the Bayesian estimation approach, the parameter $\underline{\theta}$ of interest is a random vector whose particular realization must be estimated. The motivation for doing so is twofold. First, if some prior knowledge about $\underline{\theta}$ is available, it can be incorporated into the estimator. The mechanism for doing this requires us to assume that $\underline{\theta}$ is a random vector with a given prior PDF. Classical estimation, on the other hand, finds it difficult to make use of any prior knowledge. The Bayesian approach, when applicable, can therefore improve the estimation accuracy.

Second, Bayesian estimation is useful in situations where an MVU estimator cannot be found, as for example, when the variance of an unbiased estimator may not be uniformly less than that of all other estimators. In this instance, it may be true that for most values of the parameter an estimator can be found whose mean square error may be less than that of all other estimators. By assigning a PDF to $\underline{\theta}$, strategies can be devised to find that estimator. The resulting estimator can then be said to be optimal "on average," or with respect to the assumed prior PDF of $\underline{\theta}$.

There are some general estimation approaches based on Bayesian philosophy. Here a brief introduction is given just to the following methods.

Minimum Mean Square Error (MMSE) Estimator The data model is that the joint PDF of $\mathbf{X}$,$\underline{\theta}$ or $f\left(\underline{X},\underline{\theta}\right)$ is known, where $\underline{\theta}$ is now considered to be a random vector $\underline{\theta}=[\theta_1,\theta_2,\cdots,\theta_p]^T$. Usually, $f(\underline{X}|\underline{\theta})$ is specified as the data model and $f\left(\underline{\theta}\right)$ as the prior PDF for $\underline{\theta}$, so that $f\left(\underline{X},\underline{\theta}\right)=f(\underline{X}|\underline{\theta})f\left(\underline{\theta}\right)$. The estimator θ_i minimizes the Bayesian MSE:

$$Bmse\left(\theta_i\right) = E\left[\left(\theta_i - \widehat{\theta}_i\right)^2\right], \qquad i = 1, 2, \ldots, p \tag{1.21}$$

where the expectation $E[\cdot]$ is with respect to $f\left(\underline{X}, \theta_i\right)$.

In the non-Gaussian case, this will be difficult to implement.

Maximum A Posteriori (MAP) Estimator In MAP estimation, the data model is the same as for the MMSE estimator. The estimator is the value of $\underline{\theta}$ that maximizes or, equivalently, the value that maximizes $f(\underline{X}|\underline{\theta})f\left(\underline{\theta}\right)$. The performance of the estimator depends on the joint PDF $f\left(\underline{X}, \underline{\theta}\right)$, therefore, no general formula of the estimator is available. If $\underline{X}, \underline{\theta}$ are jointly Gaussian, then the performance of the estimator is identical to that of the MMSE estimator.

Linear Minimum Mean Square Error (LMMSE) Estimator This data model is that the first two moments of the joint PDF $f(\underline{X}|\underline{\theta})$ are known. The estimator is

$$\widehat{\underline{\theta}} = E\left(\underline{\theta}\right) + \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\left(\underline{X} - E\left(\underline{X}\right)\right) \tag{1.22}$$

and the error $\varepsilon_i = \theta_i - \widehat{\theta}_i$ has zero mean and variance. If $\underline{X}, \underline{\theta}$ are jointly Gaussian, then this is identical to the MMSE and MAP estimators.

Actually, if $\underline{X}, \underline{\theta}$ are jointly Gaussian, all the above estimation techniques are essentially the same.

Choosing an Estimator

It has been seen that at times it is not possible to assure the existence of an optimal estimator, an example being the search for the MVU estimator in classical estimation. In other instances, even though the optimal estimator could easily be found, it could not be implemented, an example being the MMSE estimator in Bayesian estimation [4]. For a particular problem, one is neither assured of finding an optimal estimator, or even if the optimal estimator can be found, it is not possible to implement it.

Therefore, it becomes critical to have at one's disposal knowledge of the estimators that are optimal and easily implemented, and furthermore, to understand under what conditions one can justify their use.

In general, the search for an appropriate estimator for a signal-processing problem should begin with the search for an optimal estimator that is computationally feasible. If the search fails, then suboptimal estimators should be investigated.

Depending on the assumptions on the signal and noise, the data may have the form of the classical or Bayesian linear model, so one could find the optimal estimator easily. However, even if the estimator is optimal for the assumed data model, its performance may not be adequate. Thus, the data model may need to be modified. The flowchart in Fig 1.1 describes the considerations in the selection of an estimator.

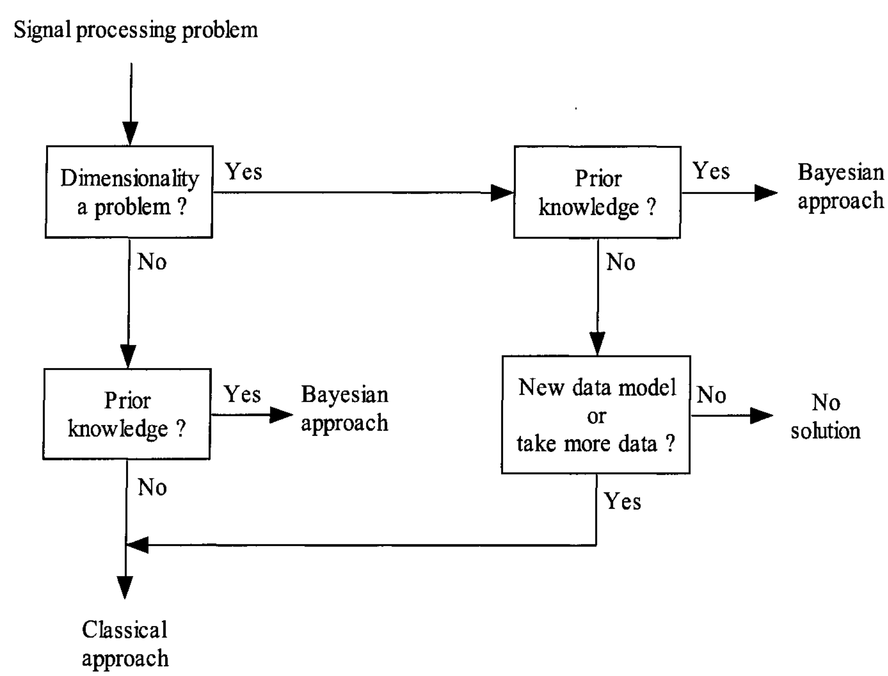

Figure 1.1: Classical versus Bayesian Estimation

With prior knowledge such as the PDF of estimator, a Bayesian approach could be used. Even if dimensionality is not a problem, the use of prior knowledge as embodied by the prior PDF will improve the estimation accuracy in the Bayesian sense. That is to say, the Bayesian MSE will be reduced.

If no prior knowledge is available, the data model will be forced to be re-evaluated (to reduce dimensionality of the problem) or else more data points are needed. This may result in bias errors due to modeling inaccuracies, but at least the variability of any resultant estimator would be reduced. Then, one could resort to a classical approach.

1.2 Concluding Remarks

This chapter has introduced the basic parameter estimation problems and the commonly used estimation methods. There are two kinds of estimation methods, the Classical estimation method and the Bayesian estimation method. The former method assumes that the parameter vector $\underline{\theta}$ is a deterministic constant vector, where the latter the parameter vector $\underline{\theta}$ is assumed to be the observation of a random vector. The performance of an estimator can be evaluated in terms of its bias and variance. Preferably, one wants to find an unbiased estimator with the minimum variance. But in reality this may not be always possible. Therefore, it is needed to compare different estimators with the same criteria in order to locate the "best" estimator. In this regard, the classical CRLB bound may be used to compare the variance of the unbiased estimators. The following chapters will introduce different applications of CRLB bound in evaluating the estimators performance.

Chapter 2

Parameter Estimation for Single Tone Sinusoidal Signals

The preceding chapter introduced estimation theory and Cramer-Rao Lower Bound (CRLB) on the estimation variance. This chapter is mainly concerned with the estimation problems for single-tone sinusoidal signal contaminated by additive white Gaussian noise (AWGN). Three existing estimation techniques are reviewed, namely linear regression estimator (LRE), maximum likelihood estimator (MLE), and linear prediction estimation (LPE). Extensions to linear regression are also presented for both low and high SNR ranges. The performance of these techniques is investigated by comparing with the CRLB bound.

2.1 Introduction

Estimating the frequency of sinusoidal signals contaminated by noise has been the focus of research for quite some time. It has important applications in such areas as communications, radar, sonar and geophysical seismology. For example, in location systems, e.g. radar or sonar, the measurement of the Doppler frequency is important because it would permit the estimation of the radial velocity of the target.

Generally speaking, frequency estimation problems can be categorized into three different cases.

Case 1: single-tone frequency estimation. In this case, the signal consists of a single, constant-frequency, sinusoid. This is the oldest and simplest frequency estimation problem. A classical paper by Rife and Boorstyn [10] examined this case using the maximum likelihood estimation (MLE) method. Also see the papers in [11–13].

Case 2: multi-harmonic frequency estimation. In this case, the signal is composed of sum of harmonically related sinusoids. For example, suppose frequency information is to be from acoustic sources such as rotating machinery. Then, non-linear effects within the generating system often give rise to harmonics and subharmonics of the fundamental frequency of interest. In these situations, case 1 does not model the physical situation adequately, so a signal model which accounts for the added harmonics should be used. See the papers in [14–18].

Case 3: multi-tone frequency estimation. This problem occurs in certain environments where several tonal sources of differing frequencies may be present in one signal. In some applications, it may be possible to apply single-tone techniques in this situation, but it is more desirable to account for the extra problem complexity by altering the signal model. Refer to the papers in [19–21] for further information.

Each of the above cases assumes a different signal model which is to be fitted to measured data, requiring different estimation algorithms for different models. There are numerous papers on the estimation problems for sinusoidal signal contaminated in noise. In this thesis, discussions are confined to the first two of the above-mentioned cases, i.e. single-tone frequency estimation and multi-harmonic frequency estimation. This chapter will discuss case 1 and the next chapter will deal with case 2.

Let us consider a single-tone sinusoidal signal which has been corrupted by noise in the receiver. The received samples $x(n)$ can be expressed as

$$x(n) = A_0 \cos\left[\omega_0 T n + \theta_0\right] + w(n), \quad n = 0, 1, \cdots, N-1, \tag{2.1}$$

where N is the number of samples received, and $w(n)$ represents the additive white Gaussian noise (AWGN) with a mean of zero and a variance of σ^2. Therefore,

$$E\left[w\left(n\right)\right] = 0,$$

and

$$E\left[w\left(n\right)w\left(n+m\right)\right]=\sigma^2\delta(m),$$

where $\delta(m)$ is the Kronecker delta sequence, and $E\left[\cdot\right]$ represents the statistical expectation operator. The goal will be to estimate the frequency ω_0 from N-point noisy data $\{x(n),\ n=0,1,\cdots,N-1\}$. Three estimation techniques based on this data model are presented in the following sections.

2.2 Linear Regression Estimator

This section is firstly concerned with a review of single frequency estimation technique proposed by Tretter [22]. This is followed by a generalization of Tretter's work by lifting the restriction on high SNR to general values of SNR.

2.2.1 Analytic Signal Model and Hilbert Transform

Let us rewrite Eqn. (2.1) in the form

$$x(n)=A_0\cos\left[\phi(n)\right]+w(n),\quad n=0,1,\cdots,N-1, \tag{2.2}$$

where $\phi(n)=\ \omega_0Tn+\theta_0$. In this estimation method, the signal $x(n)$ is replaced by the corresponding "analytic signal" [1]

$$\begin{aligned} r(n) &= x(n)+jH\left\{x(n)\right\} \\ &= A_0\cos[\phi(n)]+jA_0H\left\{\cos\left[\phi(n)\right]\right\}+w(n)+jH\left\{w(n)\right\} \\ &= A_0e^{j\phi(n)}+z(n),\quad n=0,1,\cdots,N-1, \end{aligned} \tag{2.3}$$

where $H\left\{\cdot\right\}$ represents the discrete-time Hilbert transform operator, and

$$z(n)=w(n)+jH\left\{w(n)\right\}. \tag{2.4}$$

[1] In an "analytic signal", the spectrum of the signal vanishes either for positive or for negative frequencies.

The Hilbert transform operator can be visualized as a linear time-invariant noncasual digital filter operator with an impulse response sequence [23]

$$h(n) = \begin{cases} \frac{2}{\pi}\frac{\sin^2(n\pi/2)}{n}, & n \neq 0 \\ 0 & , n = 0 \end{cases}, \tag{2.5}$$

or the frequency response

$$H(e^{j\omega}) = \begin{cases} j\,, & -\pi < \omega < 0 \\ -j, & 0 < \omega < \pi \end{cases}. \tag{2.6}$$

Let us define

$$\eta(n) = H\{w(n)\}. \tag{2.7}$$

Then, $\eta(n)$ can be written in terms of a convolution sum in accordance with

$$\eta(n) = \sum_{k=-\infty}^{\infty} w(k)h(n-k), \; n = 0, 1, \cdots, N-1. \tag{2.8}$$

It can be shown (see Appendix A) that $w(n)$ and $\eta(n)$ are zero-mean Gaussian random variables, each with a variance of σ^2, and they are uncorrelated (and so are statistically independent as they are Gaussian).

As in [22], one can express Eqn. (2.3) as

$$r(n) = \left[1 + \frac{1}{A_0}z(n)e^{-j(\omega_0 Tn+\theta_0)}\right] A_0 e^{j(\omega_0 Tn+\theta_0)}. \tag{2.9}$$

Let us define

$$\begin{aligned} v(n) &\triangleq \frac{1}{A_0}z(n)e^{-j(\omega_0 Tn+\theta_0)} \\ &= v_I(n) + jv_Q(n), \end{aligned} \tag{2.10}$$

where subscript "I" denotes the "in-phase" component, and subscript "Q" denotes the "quadrature" component of $v(n)$. The sequence defined by Eqn. (2.10) is a complex Gaussian process. This is due to the fact that the Hilbert transform is a linear operator and any linear operation on a Gaussian random variable yields a Gaussian random variable.

Some details of the statistical characteristics of $v(n)$ is summarized as follows. From the definition of $v(n)$, it can be shown that

$$E[v(n)] = 0, \tag{2.11a}$$

$$E\left[|v(n)|^2\right] = \frac{2\sigma^2}{A_0^2}. \tag{2.11b}$$

From Eqn. (2.10), one has

$$\begin{aligned} v_I(n) &= \mathrm{Re}\left[\frac{1}{A_0} z(n) e^{-j(\omega_0 Tn + \theta_0)}\right] \\ &= \frac{1}{A_0}\left[w(n)\cos(\omega_0 Tn + \theta) + \eta(n)\sin(\omega_0 Tn + \theta)\right], \end{aligned} \tag{2.12}$$

and

$$\begin{aligned} v_Q(n) &= \mathrm{Im}\left[\frac{1}{A_0} z(n) e^{-j(\omega_0 Tn + \theta_0)}\right] \\ &= \frac{1}{A_0}\left[\eta(n)\cos(\omega_0 Tn + \theta) - w(n)\sin(\omega_0 Tn + \theta)\right]. \end{aligned} \tag{2.13}$$

Since $w(n)$ and $\eta(n)$ are both zero-mean and independent Gaussian distributed random variables, it is easy to verify that

$$E[v_I(n)] = E[v_Q(n)] = 0, \tag{2.14}$$

$$E[v_I(n) v_Q(n)] = 0, \tag{2.15}$$

and that

$$E\left[v_I^2(n)\right] = E\left[v_Q^2(n)\right] = \frac{\sigma^2}{A_0^2}. \tag{2.16}$$

From Eqns. (2.14) - (2.16), one can easily arrive at Eqn. (2.11). Therefore $v_I(n)$ and $v_Q(n)$ are uncorrelated Gaussian random variables of zero mean and variance of σ^2/A_0^2. Consequently, they are also statistically independent due to the fact that they have Gaussian statistics. These characteristics will be used for the derivation of the probability density function (PDF) of phase noise.

2.2.2 Derivation of the PDF of Phase Noise for Full SNR Range

Tretter [22] proposed a linear regression estimator for the amplitude A_0, frequency ω_0 and phase θ_0 of the signal model $r(n)$. The parameters A_0, ω_0 and θ_0 of the tone were assumed to be constant over the observation interval. Tretter considered the high SNR case which leads to a model of the phase noise that is Gaussian. This chapter extends his results by considering the PDF of the phase noise over the full range of SNR.

Let us rewrite Eqn. (2.9) as

$$r(n) = [(1 + v_I(n)) + jv_Q(n)]\, A_0 e^{j(\omega_0 T n + \theta_0)}. \tag{2.17}$$

Moreover, let

$$a(n)e^{j\phi(n)} \triangleq (1 + v_I(n)) + jv_Q(n), \tag{2.18}$$

where

$$a(n) = \left[[1 + v_I(n)]^2 + v_Q^2(n)\right]^{1/2}, \tag{2.19a}$$

$$\phi(n) = \tan^{-1}\left[\frac{v_Q(n)}{1 + v_I(n)}\right]. \tag{2.19b}$$

Then, the analytic signal in Eqn. (2.17) can be simplified as

$$r(n) = A_0 a(n) e^{j[\omega_0 T n + \theta_0 + \phi(n)]}. \tag{2.20}$$

From Eqn. (2.20), $\phi(n)$ acts as an additive noise on the phase θ_0 and is called as "phase noise". It is desirable to know $p_\Phi(\phi)$, the PDF of random variable $\phi(n)$.

For convenience, omit the explicit dependence of $a(n)$ and $\phi(n)$ on time n, and relabel $v_I(n)$ as i and $v_Q(n)$ as q. Then, the following relationships exist between random variables i and q, and random variables a and ϕ :

$$a \;=\; \left[(1+i)^2 + q^2\right]^{1/2}, \tag{2.21a}$$

$$\phi \;=\; \tan^{-1}\left(\frac{q}{1+i}\right). \tag{2.21b}$$

In order to determine $p_\Phi(\phi)$, one must find the joint PDF of a and ϕ first. This is given by [24]

$$p_{A,\Phi}(a,\phi) = p_{I,Q}(i,q)\left|J\left(\frac{i,q}{a,\phi}\right)\right|, \tag{2.22}$$

where $p_{I,Q}(i,q)$ is the joint PDF of i and q, and $J\left(\frac{i,q}{a,\phi}\right)$ is the Jacobian of the transformation as given by

$$J\left(\frac{i,q}{a,\phi}\right) = \begin{vmatrix} \frac{\partial i}{\partial a} & \frac{\partial i}{\partial \phi} \\ \frac{\partial q}{\partial a} & \frac{\partial q}{\partial \phi} \end{vmatrix}. \tag{2.23}$$

Recall that i and q are statistically independent Gaussian random variables with mean of zero and variance of σ^2/A_0^2 (c.f. Eqns. (2.14)-(2.16)). It is also convenient to set

$$s^2 = \frac{\sigma^2}{A_0^2}, \tag{2.24}$$

reducing the joint PDF of i and q to

$$p_{I,Q}(i,q) = p_I(i)p_Q(q) = \frac{1}{2\pi s^2}\exp\left(-\frac{i^2+q^2}{2s^2}\right). \tag{2.25}$$

It can be observed from Eqn. (2.24) that s^2 has the property of being the reciprocal of the signal to noise ratio (SNR). In this way, $SNR = 1/s^2$. From Eqn. (2.21) one has

$$i = a\cos\phi - 1, \tag{2.26a}$$

$$q = a\sin\phi, \tag{2.26b}$$

so that

$$\begin{aligned} J\left(\frac{i,q}{a,\phi}\right) &= \begin{vmatrix} \cos\phi & -a\sin\phi \\ \sin\phi & a\cos\phi \end{vmatrix} \\ &= a. \end{aligned} \tag{2.27}$$

From Eqn. (2.26)

$$i^2 + q^2 = a^2 - 2a\cos\phi + 1. \tag{2.28}$$

Substitute Eqn. (2.28) into Eqn. (2.25), then invoke the result together with Eqn. (2.27) in Eqn. (2.22) to obtain

$$p_{A,\Phi}(a,\phi) = \frac{1}{2\pi s^2} a \exp\left(-\frac{a^2 - 2a\cos\phi + 1}{2s^2}\right), \tag{2.29}$$

where $a \geqslant 0$, and $-\pi \leqslant \phi < \pi$. Thus, Eqn. (2.29) is the desired joint PDF of random variables a and ϕ.

The PDF of the phase noise is given by

$$\begin{aligned} p_{\Phi}(\phi) &= \int_0^\infty p_{A,\Phi}(a,\phi)\, da & (2.30) \\ &= \frac{1}{2\pi s^2} \exp\left(-\frac{1}{2s^2}\right) \int_0^\infty a \exp\left(-\frac{a^2}{2s^2} + \frac{a\cos\phi}{s^2}\right) da. & (2.31) \end{aligned}$$

By using the identity [25]

$$\begin{aligned} &\int_0^\infty x \exp\left(-\mu x^2 - 2vx\right) dx \\ &= \frac{1}{2\mu} - \frac{v}{2\mu}\sqrt{\frac{\pi}{\mu}} \exp\left(\frac{v^2}{\mu}\right)\left[1 - \operatorname{erf}\left(\frac{v}{\sqrt{\mu}}\right)\right] \end{aligned} \tag{2.32}$$

in Eqn. (2.31) and by identifying $\mu = \frac{1}{2s^2}$ and $v = -\frac{\cos\phi}{2s^2}$, one can obtain

$$p_{\Phi}(\phi) = \frac{1}{2\pi s^2} \exp\left(-\frac{1}{2s^2}\right)\left\{s^2 + \sqrt{\frac{\pi s^2}{2}} \cos\phi \exp\left(\frac{\cos^2\phi}{2s^2}\right)\left[1 - \operatorname{erf}\left(-\frac{\cos\phi}{\sqrt{2s^2}}\right)\right]\right\}, \tag{2.33}$$

where

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp\left(-t^2\right) dt$$

is the error function.

Eqn. (2.33) is the exact expression for the phase noise PDF over all SNR value, which is an extension of the result in [22] for the PDF of the phase noise. It is clear that for high SNR (i.e., for small s^2), this expression tends to the limiting Gaussian PDF

$$g(\phi) = \frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{\phi^2}{2s^2}\right). \tag{2.34}$$

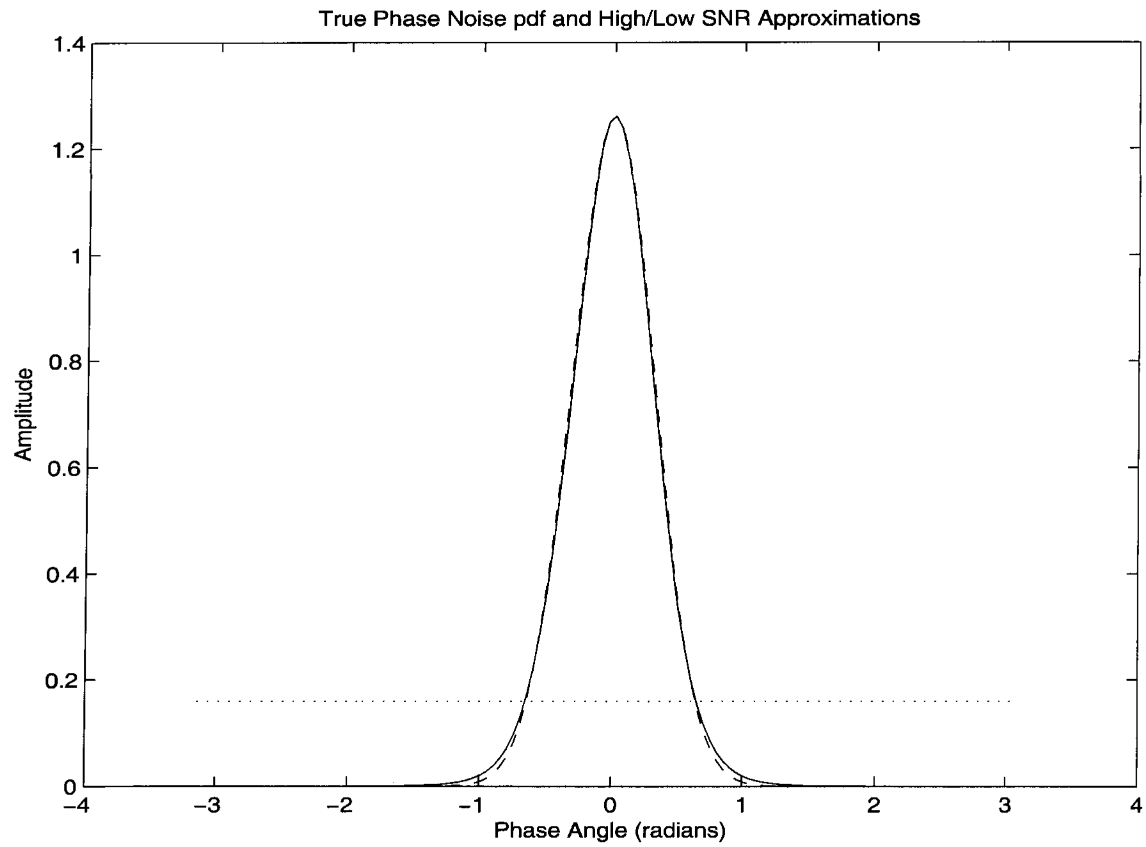

Figure 2.1: Plots of Eqns. (2.33) (solid line), (2.34) (dashed line), and (2.35) (dotted line) for $s^2 = 0.1$.

For low SNR (i.e., for large s^2), the limiting PDF is the uniform distribution over interval $[-\pi, \pi)$ as

$$l(\phi) = \begin{cases} \frac{1}{2\pi}, & -\pi \leqslant \phi < \pi \\ 0, & \text{elsewhere} \end{cases} . \qquad (2.35)$$

Fig. 2.1 shows the plot of the PDFs as in Eqns. (2.33), (2.34) and (2.35) for $s^2 = 0.1$, and it is noted that the phase noise PDF in all cases is symmetric about a mean of $\overline{\phi} = E[\phi(n)] = 0$. This is also obvious from Eqn. (2.33) since $\cos\phi = \cos(-\phi)$. Furthermore, one can see from 2.1 that for high SNR, the limiting PDF of Eqn. (2.34) is a good approximation to Eqn. (2.33).

2.2.3 Detailed Derivation of Tretter's Linear Regression Estimator

The Tretter [22] linear regression estimator should be derived here with the new phase PDF in Eqn. (2.33) in the full SNR range. However, the approach adopted here is somewhat different from that in [22].

Recall from Eqn. (2.20) that

$$r(n) = A_0 a(n) e^{j[\omega_0 Tn + \theta_0 + \phi(n)]}. \tag{2.36}$$

Let us set $\theta(n) = \omega_0 Tn + \theta_0 + \phi(n)$ which is the instantaneous phase of $r(n)$ obtained by applying a phase unwrapping algorithm to the argument of $r(n)$. Let us assume as in [22] that $\theta(n)$ is available for $n = n_0, \cdots, n_0 + N - 1$, where n_0 is some initial or starting time instance.

Also as in [22], define the error sequence energy

$$E(\omega_0, \theta_0) = \sum_{n=n_0}^{n_0+N-1} \left[\theta(n) - \omega_0 Tn - \theta_0\right]^2. \tag{2.37}$$

This implies that it is desirable to attempt to fit the best straight line segment $\omega_0 Tn + \theta_0$ to the data $\{\theta(n)\}$, where "best" is defined in the least squares sense. This is a classical "linear regression" problem. The notation in Eqn. (2.37) assumes that n_0, N and T are fixed constants predetermined by the experimental setup used for the measurement sequence $\{r(n)\}$. It is desired to minimize E with respect to ω_0 and θ_0. Thus, one can write

$$\left(\hat{\omega}_0, \hat{\theta}_0\right) = \arg\min_{(\omega_0, \theta)} E(\omega_0, \theta_0), \tag{2.38}$$

i.e. $\hat{\omega}_0$ and $\hat{\theta}_0$ are the choices of ω_0 and θ_0 that minimize E, and so are the best estimates of frequency and phase in the least squares sense. Let us define the vectors

$\underline{\alpha} = [\omega_0 \ \theta_0]^T$, and $\underline{\beta} = [nT \ 1]^T$. Then, Eqn. (2.37) can be rewritten as

$$\begin{aligned} E(\underline{\alpha}) &= \sum_{n=n_0}^{n_0+N-1} \left[\theta(n) - \underline{\beta}^T \underline{\alpha}\right]^2 \\ &= \sum_{n=n_0}^{n_0+N-1} \left[\theta^2(n) - 2\theta(n)\underline{\beta}^T\underline{\alpha} + \underline{\alpha}^T \underline{\beta}\underline{\beta}^T \underline{\alpha}\right] \\ &= \rho - 2\underline{q}^T\underline{\alpha} + \underline{\alpha}^T P \underline{\alpha}, \end{aligned} \tag{2.39}$$

where $\rho = \sum_{n=n_0}^{n_0+N-1} \theta^2(n)$, $\underline{q} = \sum_{n=n_0}^{n_0+N-1} \underline{\beta}\theta(n)$, and $P = \sum_{n=n_0}^{n_0+N-1} \underline{\beta}\underline{\beta}^T$. Let $\hat{\underline{\alpha}} = \left[\hat{\omega}_0 \ \hat{\theta}_0\right]^T$ be the estimator of $\underline{\alpha}$ that minimizes $E(\underline{\alpha})$ so that

$$\frac{\partial}{\partial \underline{\alpha}} E(\underline{\alpha})\Big|_{\underline{\alpha}=\hat{\underline{\alpha}}} = 0. \tag{2.40}$$

Then, by substituting Eqn. (2.39) into Eqn. (2.40), one can obtain linear system of equations

$$P\hat{\underline{\alpha}} = \underline{q}. \tag{2.41}$$

Therefore, the estimator is given by

$$\hat{\underline{\alpha}} = P^{-1}\underline{q}, \tag{2.42}$$

where $\underline{q} = \begin{bmatrix} \sum_{n=n_0}^{n_0+N-1} nT\theta(n) \\ \sum_{n=n_0}^{n_0+N-1} \theta(n) \end{bmatrix}$, $P = \begin{bmatrix} T^2 \sum_{n=n_0}^{n_0+N-1} n^2 & T\sum_{n=n_0}^{n_0+N-1} n \\ \sum_{n=n_0}^{n_0+N-1} n & \sum_{n=n_0}^{n_0+N-1} 1 \end{bmatrix}$.

Through some straightforward manipulations, one can finally obtain the following estimator

$$\begin{bmatrix} \hat{\omega}_0 \\ \hat{\theta}_0 \end{bmatrix} = \frac{12}{T^2N^2(N^2-1)} \begin{bmatrix} N & -T(Nn_0 + A) \\ -T(Nn_0 + A) & T^2(Nn_0^2 + 2An_0 + B) \end{bmatrix} \begin{bmatrix} \sum_{n=n_0}^{n_0+N-1} nT\theta(n) \\ \sum_{n=n_0}^{n_0+N-1} \theta(n) \end{bmatrix}, \tag{2.43}$$

where

$$A = \frac{N(N-1)}{2}, \tag{2.44}$$

$$B = \frac{(N-1)N(2N-1)}{6}. \tag{2.45}$$

Eqn. (2.43) is the same as Eqn. (12) in [22].

As suggested in [22], let us now assume that n_0 is chosen to satisfy

$$Nn_0 + A = 0, \tag{2.46}$$

so that $n_0 = -\frac{1}{2}(N-1)$.[2] Then, Eqn. (2.43) can be simplified to

$$\begin{bmatrix} \hat{\omega}_0 \\ \hat{\theta}_0 \end{bmatrix} = \frac{12}{T^2N^2(N^2-1)} \begin{bmatrix} N & 0 \\ 0 & \frac{1}{12}T^2(N-1)N(N+1) \end{bmatrix} \begin{bmatrix} \sum_{n=n_0}^{n_0+N-1} nT\theta(n) \\ \sum_{n=n_0}^{n_0+N-1} \theta(n) \end{bmatrix}. \tag{2.47}$$

From Eqn. (2.47), the simplified estimator can be written as :

$$\hat{\omega}_0 = \frac{12}{TN(N^2-1)} \sum_{n=-(N-1)/2}^{(N-1)/2} n\theta(n), \tag{2.48a}$$

$$\hat{\theta}_0 = \frac{1}{N} \sum_{n=-(N-1)/2}^{(N-1)/2} \theta(n). \tag{2.48b}$$

From Eqn. (2.48), the mean of $\hat{\omega}_0$ and that of $\hat{\theta}_0$ can be obtained as

$$\begin{aligned} E\left[\hat{\omega}_0\right] &= \frac{12}{TN(N^2-1)} E\left[\sum_{n=-(N-1)/2}^{(N-1)/2} n\theta(n)\right] \\ &= \frac{12}{TN(N^2-1)} \sum_{n=-(N-1)/2}^{(N-1)/2} nE\left[\omega_0 Tn + \theta_0 + \phi(n)\right] \\ &= \frac{12}{TN(N^2-1)} \sum_{n=-(N-1)/2}^{(N-1)/2} n\left(\omega_0 Tn + \theta_0\right) \\ &= \frac{12\omega_0 T}{TN(N^2-1)} \sum_{n=-(N-1)/2}^{(N-1)/2} n^2 \end{aligned}$$

[2]This implies that N is odd.

$$= \frac{12\omega_0}{N(N^2-1)} \frac{N(N^2-1)}{12} = \omega_0, \quad (2.49)$$

and

$$\begin{aligned} E\left[\hat{\theta}_0\right] &= \frac{1}{N} \sum_{n=-(N-1)/2}^{(N-1)/2} E\left[\theta\right] \\ &= \frac{1}{N} \sum_{n=-(N-1)/2}^{(N-1)/2} (\omega_0 T n + \theta_0) \\ &= \frac{1}{N} N\theta_0 = \theta_0. \end{aligned} \quad (2.50)$$

From Eqn. (2.49), it can be seen that the frequency estimator is unbiased for all SNRs (and the estimator for initial phase is also unbiased from Eqn. (2.50)). This conclusion is an extension of the result by Tretter which was based on the assumption of high SNRs. So the unbiased CRLB bound can be applied to evaluate the performance of the estimator in the full range of SNR.

2.3 Maximum Likelihood Estimator

The basic theory can be found at Scharf [9] (see also Rife [10]). The MLE estimation $\hat{\omega}_0$ is the value of ω_0 that maximizes the PDF of the observation data.

The data model is the same as Eqn. (2.1) in the above. The data has a Gaussian distribution with a PDF of

$$\begin{aligned} f(\underline{x}; \underline{\theta}) &= \prod_{n=0}^{N-1} f\left(x(n); \underline{\theta}\right) \\ &= \frac{1}{\left(2\pi\sigma^2\right)^{N/2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} \left[x(n) - \mu_n\right]^2\right], \end{aligned} \quad (2.51)$$

where $\underline{\theta} = [\omega_0\ A_0\ \theta_0\ \sigma^2]^T$ is the parameter vector to be estimated, and $\mu_n = E\left[x(n)\right] = A_0 \cos\left(\omega_0 T n + \theta_0\right)$ is the mean of data.

The maximization of $f(\underline{x}; \underline{\theta})$ in Eqn. (2.51) is not an easy task, as it consists of the multiplication of exponential functions. One can resort to using the *ln* operation to the equation to get the so called log likelihood function (LLF). Because the *ln* function is a monotonically increasing function, the maximum of $f(\underline{x}; \underline{\theta})$ will occur at the maximum of its LLF

$$\ln f(\underline{x}; \underline{\theta}) = -\frac{N}{2} \ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} \left[x(n) - \mu_n\right]^2. \tag{2.52}$$

2.3.1 Algorithm

This algorithm includes 4 steps to estimate $\underline{\theta}$.

1) Estimation of noise variance σ^2: Let the derivative of LLF with respect to the variance be set zero

$$\left.\frac{\partial}{\partial\sigma^2} \ln f(\underline{x}; \underline{\theta})\right|_{\sigma^2=\hat{\sigma}^2} = \left. -\frac{N}{2\sigma^2} + \frac{1}{2\left(\sigma^2\right)^2} \sum_{n=0}^{N-1} \left[x(n) - \mu_n\right]^2 \right|_{\sigma^2=\hat{\sigma}^2} = 0. \tag{2.53}$$

Then, one can obtain $\hat{\sigma}^2$, the estimator of σ^2, as

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{n=0}^{N-1} \left[x(n) - \mu_n\right]^2. \tag{2.54}$$

2) Estimation of initial phase θ_0: Let

$$V^2 = \sum_{n=0}^{N-1} \left[x(n) - \mu_n\right]^2 \tag{2.55}$$

and substitute Eqn. (2.54) into Eqn. (2.51) with $\sigma^2 = \hat{\sigma}^2 = \frac{V^2}{N}$. Then, Eqn. (2.51) becomes

$$\ln f(\underline{x}; \underline{\theta}) = -\frac{N}{2} \ln\left(\frac{2\pi}{N} V^2\right) - \frac{N}{2}. \tag{2.56}$$

From Eqn. (2.56), θ_0 can be estimated as $\hat{\theta}_0$ that maximizes $\ln f(\underline{x}; \underline{\theta})$, or equally, minimizes V^2. From Eqn. (2.55), V^2 can be obtained as

$$
\begin{aligned}
V^2 &= \sum_{n=0}^{N-1} [x(n) - \mu_n]^2 = \sum_{n=0}^{N-1} \left[x^2(n) - 2\mu_n x(n) + \mu_n^2\right] \\
&= \sum_{n=0}^{N-1} x^2(n) - 2A_0 \sum_{n=0}^{N-1} x(n) \cos(\omega_0 Tn + \theta_0) \\
&\quad + A_0^2 \sum_{n=0}^{N-1} \cos^2(\omega_0 Tn + \theta_0) \\
&\approx \sum_{n=0}^{N-1} x^2(n) - 2A_0 \sum_{n=0}^{N-1} x(n) \cos(\omega_0 Tn + \theta_0) + A_0^2 \frac{N}{2} \\
&= \sum_{n=0}^{N-1} x^2(n) - 2A_0 \,\mathrm{Re}\left[e^{-j\theta_0} \sum_{n=0}^{N-1} x(n) e^{-j\omega_0 Tn}\right] + A_0^2 \frac{N}{2}. \qquad (2.57)
\end{aligned}
$$

From Eqn. (2.57), θ_0 can be estimated as $\hat{\theta}_0$ that maximizes PDF by letting

$$\frac{\partial}{\partial \theta_0} V^2 = -2A_0 \frac{\partial}{\partial \theta_0} \mathrm{Re}\left[e^{-j\theta_0} \sum_{n=0}^{N-1} x(n) e^{-j\omega_0 Tn}\right] = 0. \qquad (2.58)$$

Then,

$$\mathrm{Re}\left[-je^{-j\theta_0} \sum_{n=0}^{N-1} x(n) e^{-j\omega_0 Tn}\right] = 0. \qquad (2.59)$$

By setting

$$\sum_{n=0}^{N-1} x(n) e^{-j\omega_0 Tn} = X(\omega_0) = |X(\omega_0)| e^{j\phi_X(\omega_0)} \qquad (2.60)$$

where $X(\omega)$ is the discrete-time Fourier transform (DTFT) of $x(n)$, $n = 0, 1, \cdots, N-1$,

$$X(\omega) = \sum_{n=0}^{N-1} x(n) e^{-j\omega Tn}.$$

Eqn. (2.59) can be rewritten as

$$\mathrm{Re}\left\{-j\,|X(\omega_0)|\, e^{j[\phi_X(\omega_0) - \theta_0]}\right\} = 0. \qquad (2.61)$$

Therefore, from Eqn. (2.61), the estimator of initial phase can be obtained as

$$\hat{\theta}_0 = \phi_X(\omega_0) = \arg\left[X(\omega_0)\right]. \qquad (2.62)$$

3) Estimation of amplitude A_0: Substitute Eqns. (2.60) and (2.62) into Eqn. (2.57), one has

$$V^2 = \sum_{n=0}^{N-1} x^2(n) - 2A_0 \left|X(\omega_0)\right| + \frac{A_0^2 N}{2}. \tag{2.63}$$

By setting the derivative of V^2 with respect to A_0 to zero

$$\frac{\partial}{\partial A_0} V^2 = -2\left|X(\omega_0)\right| + A_0 N = 0, \tag{2.64}$$

the estimate $\hat{A}_0$ of A_0 is obtained as

$$\hat{A}_0 = \frac{2}{N}\left|X(\omega_0)\right|. \tag{2.65}$$

4) Estimation of frequency ω_0: Substituting Eqn. (2.65) into Eqn. (2.63) will lead to

$$\begin{aligned} V^2 &= \sum_{n=0}^{N-1} x^2(n) - 2\frac{2}{N}\left|X(\omega_0)\right|^2 + \frac{4}{2N^2}\left|X(\omega_0)\right|^2 N && (2.66)\\ &= \sum_{n=0}^{N-1} x^2(n) - \frac{2}{N}\left|X(\omega_0)\right|^2. && (2.67)\end{aligned}$$

So, the estimate $\hat{\omega}_0$ of ω_0 is given by

$$\hat{\omega}_0 = \arg\min_{\omega_0} V^2 = \arg\max_{\omega_0} \left|X(\omega_0)\right|. \tag{2.68}$$

To summarize, the ML estimator of $\underline{\theta} = [\omega_0\ A_0\ \theta_0\ \sigma^2]^T$ is obtained as

$$\begin{aligned} \hat{\omega}_0 &= \arg\max_{\omega_0} \left|X(\omega_0)\right|, && (2.69)\\ \hat{A}_0 &= \frac{2}{N}\left|X(\hat{\omega}_0)\right|, && (2.70)\\ \hat{\theta}_0 &= \arg\left[X(\hat{\omega}_0)\right], && (2.71)\\ \hat{\sigma}^2 &= \frac{1}{N}\sum_{n=1}^{N}[x(n) - \hat{A}_0\cos(\hat{\omega}_0 T n + \hat{\theta}_0)]^2. && (2.72)\end{aligned}$$

The MLE can reach the CRLB bound when the data size is large enough. One of the disadvantages of this is that it is computationally intense. However, the relationship of ML estimation to the discrete Fourier transform can be exploited to increase the computational speed with the help of fast Fourier transform (FFT) algorithm.

2.4 Linear Prediction Estimator

A computationally simpler algorithm called linear prediction estimation (LPE) for frequency was first proposed by Kay in [11]. This technique is based on the complex-valued sinusoidal signal contaminated in AWGN noise given by

$$x(n) = A_0 e^{j(\omega_0 n + \theta_0)} + w(n), \quad n = 0, ..., N-1 \tag{2.73}$$

where A_0 represents the amplitude, $\omega_0 \in [-\pi, \pi)$ is the angular frequency, θ_0 represents the initial phase, and $w(n)$ represents the zero mean complex-valued AWGN noise variable with variance σ^2 and are i.i.d.. The parameters (A_0, ω_0, θ_0) are deterministic but unknown constants and the frequency ω_0 is to be estimated. The other two parameters A_0 and ω_0 are considered to be nuisance parameters.

There are two kinds of LPE estimators, namely, unweighted LPE estimator and weighted linear prediction estimator (WLPE). The algorithms are as follows:

1) The unweighted LPE frequency estimator is simply given by

$$\hat{\omega}_0 = \angle\left[\hat{R}\right] = \mathrm{Im}\left[\log \hat{R}\right], \tag{2.74}$$

where

$$\hat{R} = \sum_{n=1}^{N-1} \frac{1}{N-1} x(n) x^*(n-1). \tag{2.75}$$

In Eqn. (2.74), $\angle[\cdot]$ denotes the phase of a complex number, and *log* represents the principal value of the log function. In Eqn. (2.75), $*$ denotes complex conjugate operation.

2) The WLPE frequency estimator is given by

$$\hat{\omega}_0 = \angle\left[\hat{R}\right] = \mathrm{Im}\left[\log \hat{R}\right], \tag{2.76}$$

where

$$\hat{R} = \sum_{n=1}^{N-1} v(n)\, x(n) x^*(n-1), \tag{2.77}$$

and where

$$v(n) = \frac{6n(N-n)}{N(N^2-1)}. \tag{2.78}$$

The further discussion regarding this technique can be readily found in [26–28].

2.5 Unbiased CRLB bound and Performance Comparison

In an estimation system, it is important to have benchmarks that identify the best estimation that can be made with the available data. The unbiased CRLB bound introduced in Chapter 1 is a good candidate for this purpose. Although there are many other bounds exist, the CRLB is the most used criterion for its simplicity and easy-to-calculate property.

In this section, computer simulations are carried out to evaluate the single-tone frequency estimation performances of the above reviewed methods (i.e., the LRE, MLE and two LPE estimators) by comparing their estimation variances against the CRLB bounds for various SNR values. [3]

In [10], the unbiased CRLB bound on the estimation variance for the data model similar as in this chapter was derived as

$$CRLB = \frac{12\sigma^2}{A_0^2 T^2 N\left(N^2-1\right)} = \frac{6}{SNR \cdot T^2 N\left(N^2-1\right)}. \tag{2.79}$$

The performances comparison is carried out by the Monte Carlo simulations of the above frequency estimators with the single-tone data sizes of $N = 31$, 61 and 101, respectively. Figs. 2.2 to 2.4 are three typical results of the simulations. All numerical results are obtained by averaging over 500 Monte Carlo simulations. It can be observed from Figs. 2.2 to 2.4 that:

1) The variances of the three estimators are higher than the CRLB bounds for all SNRs. This is as expected, because the unbiased CRLB bound determines the best achieved accuracy for all unbiased estimators.

---

[3]The normalized standard deviation is used as the the square root of variance divided by the corresponding SNR.

2) The Tretter estimator has the smallest variance among the four estimators considered at very low SNRs. This is very desired at conditions where the received signal is very noisy. At very high SNRs, the Tretter estimator is slightly better than the other two estimators in terms of the estimator variance. Therefore, it is the "best" estimator for the above frequency estimation problem at these SNR ranges. Only at the middle range of SNRs (i.e. -5 dB to 8 dB), the Tretter estimator is outperformed by the MLE estimator, but is still better than the LPE estimators.

3) The estimator variances are in reversely proportional to the SNRs for all four estimators. Especially, with increasing SNRs, the variances of MLE and LRE estimators tend to approach the CRLB bound. Hence, they are efficient frequency estimators.

4) By comparing Figs. 2.2 to 2.4, it can be found that with lager N, i.e. more data points, the estimator variances and CRLB bounds decrease their values.

5) It needs to be noticed that the variance of MLE estimator decreased dramatically when SNR is smaller than 3 dB for $N = 31$ (c.f. Fig. 2.2). This phenomenon is called "threshold effect". The Tretter estimator has the similar threshold effect but with a different start point of SNR $= 8$ dB for $N = 31$. It is well known that nonlinear estimation is generally plagued by threshold effect. At low SNRs, there is usually a range of SNRs in which the estimation error rises very rapidly as SNR decreases. The SNR at which this effect first becomes apparent is called the threshold.

6) The threshold effect is undesirable in parameter estimation and should be avoided as much as possible. It can be observed that for different data size N, the threshold value is different for the same estimator. For example, the threshold of the MLE estimator is 0 dB for $N = 61$ (c.f. Fig. 2.3), and is -3 dB for $N = 101$ (c.f. Fig. 2.4). The starting point of threshold effect in Fig. 2.2 ($N = 31$) occurs sooner than in Fig. 2.3 ($N = 61$) and much sooner than Fig. 2.4 ($N = 101$), implying that larger data size will lead to a later occurring of the threshold effect. This is because that larger data size means more information regarding the parameter available, leading

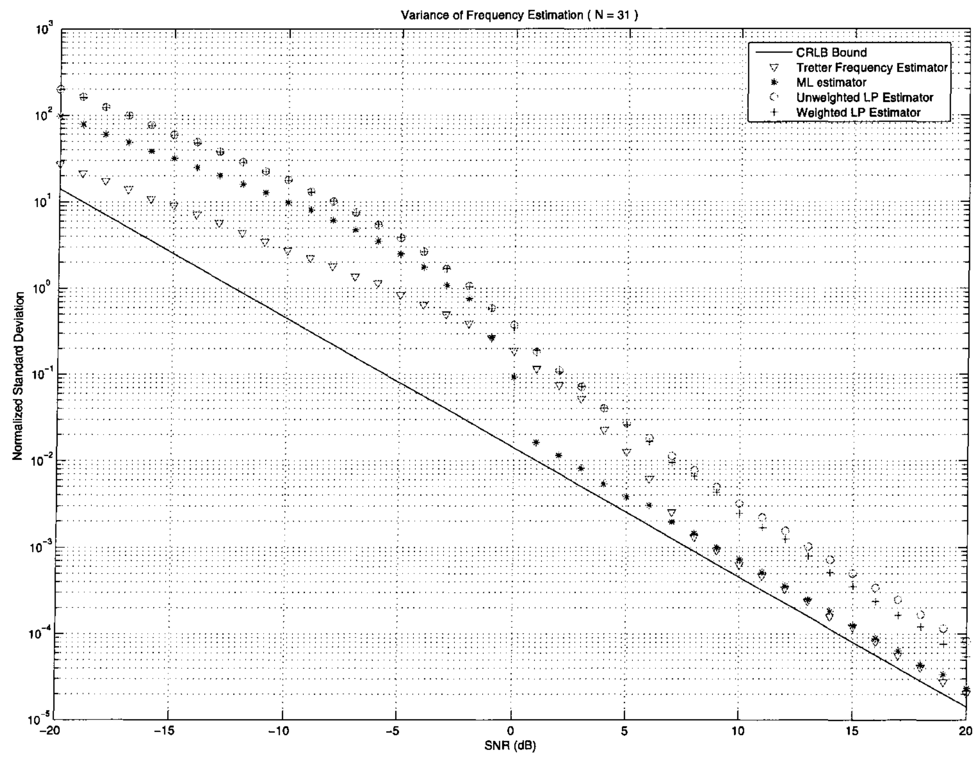

Figure 2.2: Comparisons of the sample variance of the estimators against the CRLB bound for $N = 31$.

to more accurate parameter estimation, or less estimator variance at the same SNR. Therefore, it is preferable to get as much data points as possible in order to obtain a better estimation accuracy when the SNR is low.

7) The last point is that the MLE estimator has the lowest threshold value among all estimators at the same data size N. In situation where data size is a constraint condition, the MLE estimator should be considered as the most preferable one.

2.6 Conclusions

In this chapter, the LRE estimator by Tretter [22] has been developed in a more general manner. The Tretter estimator has the disadvantage that it requires as input

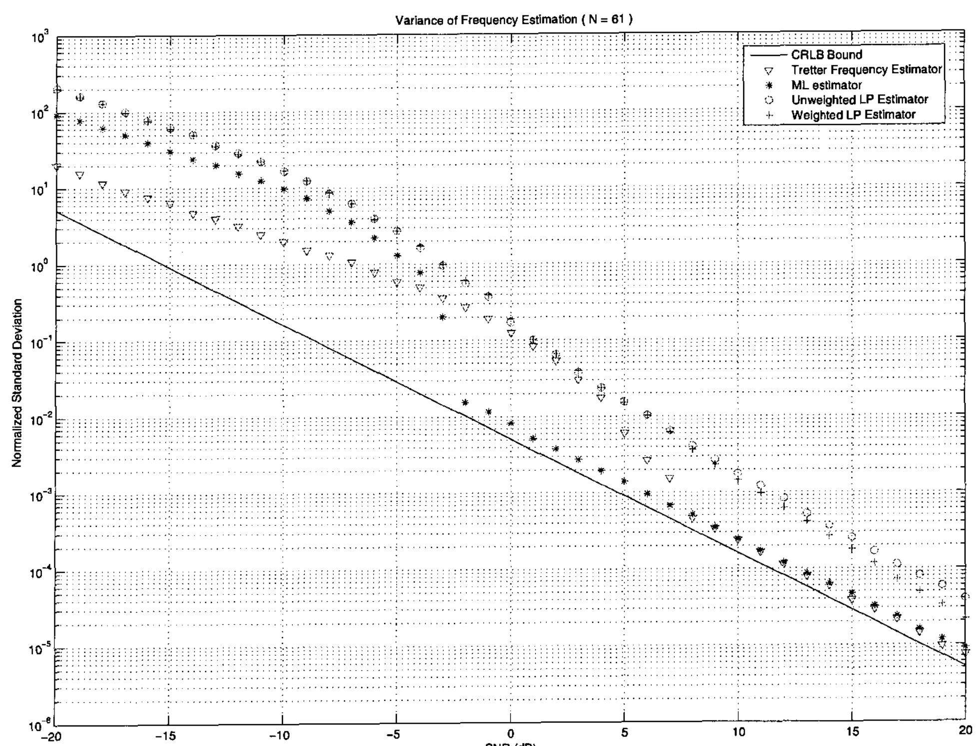

Figure 2.3: Comparisons of the sample variance of the estimators against the CRLB bound for $N = 61$.

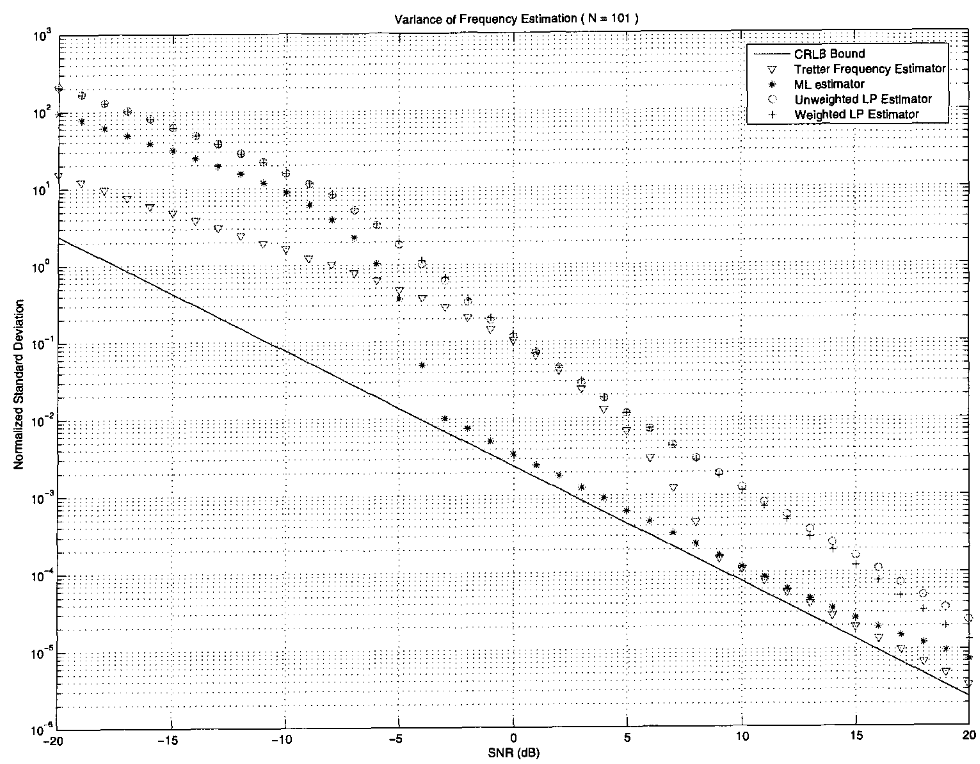

Figure 2.4: Comparisons of the sample variance of the estimators against the CRLB bound for $N = 101$.

an unwrapped estimate of the phase. If the SNR is high, this is not a great problem. However, if the SNR is low, the noise causes $\pm 2\pi$ radian phase ambiguities which are an additional source of error in the frequency estimates, and which was not accounted for by the models considered either in this thesis or in [22]. This is perhaps the best justification for only considering the high SNR case in [22]. But, in general, Tretter's LRE estimator has the smallest variance among the four estimators concerned at both very low SNRs and very high SNRs, except that at the middle range of SNRs (e.g. -5dB to 8dB), the Tretter estimator is outperformed by the MLE estimator. In all SNR range, LRE estimator is better than the two LPE estimators in the sense that it has less estimator variance.

The MLE estimator has smaller threshold SNR values than the Tretter estimator which is a very desired characteristic in application where SNR is a constrained condition. One of the disadvantages of MLE estimator is that it is computationally intense. However, the relationship of ML estimation to the discrete Fourier transform can be exploited to increase the computational speed with the help of fast Fourier transform (FFT) algorithm.

The LPE estimators seem to be worse than the other two methods in the sense that it has higher estimation variance for all SNRs and larger starting SNR values of threshold effect under the same data size. But it has the advantage that it does not require unwrapped phase estimates as input, therefore, it is simple to compute and implement.

Chapter 3

Parameter Estimation for Multiharmonic Sinusoids

The preceding chapter was concerned with parameter estimation problems for single frequency sinusoidal signal contaminated by AWGN noise. This chapter deals with the corresponding parameter estimation problems for multiharmonic sinusoids signal contaminated by AWGN noise.

3.1 Introduction

Parameter estimation for multitone sinusoidal signals in noise has been well reported in the hitherto literature. Typical techniques include the maximum likelihood estimation (MLE) [20], multiple signal classification (MUSIC) estimation [29], and the subspace method [14]. However, these techniques have paid little attention to sinusoidal signals with harmonics and subharmonics. In [20], the MLE approach was applied to the signals consisting of multiple sinusoids in noise under the assumption that there is no special harmonic relationship between the sinusoidal components. Harmonic and subharmonic signal components exist in many signal processing applications. For example, in speech signal processing problems, many acoustic sources such as rotating machinery have non-linear effects within the generating system, and often give rise to harmonics and subharmonics besides the fundamental component. To properly characterize such signals, the harmonics and subharmonics should be taken into account. The work of [30] contained in part a derivation of the Cramer-Rao Lower Bound (CRLB) for the case of a real multiharmonic signal, measurements of which are assumed to commence at time $t = 0$.

Recently, Zarowski and Kmpyvnytskyy [31] developed a modified iterative cosinor algorithm (MICA) for the temperature data containing circadian rhythms for head-injured patients. In the MICA model, the period of the circadian rhythm for the patient, called the patient's "tau" and denoted by T, is to be estimated together with the harmonic and subharmonic components of the data. In [31], the CRLB bound for the estimation of tau was derived but it was assumed that the noise variance is known a priori. This paper presents the derivation of CRLB bounds for the case that the variance is unknown. The derivations also include the CRLB bounds for other parameters such as amplitudes of harmonics and subharmonics, noise variance, and the SNR. Computer simulation results are also given to verify and interpret the CRLB bounds.

The remainder of this chapter is organized as follows. Section 2 is the statement of the problem. The CRLB bounds are derived for the case that the noise variance is not known a priori. in section 3. Section 4 presents the simulation results and performance comparison. The conclusions are given in Section 5.

3.2 Statement of the Problem

Let us consider the MICA algorithm for a circadian rhythm data set $\{s(n)\}$ given by

$$s(n) = \sum_{k=1}^{N_H} [A_k \cos(\omega k n) + B_k \sin(\omega k n)] + \sum_{k=2}^{N_S} [C_k \cos(\omega n/k) + D_i \sin(\omega n/k)] + w(n), \tag{3.1}$$

where $n = 0, 1, \cdots, N-1$, with N being the length of the data set. Moreover, $\omega = 2\pi T_s/T \in (0, \pi)$, with T_s being the sampling period, $A_k, B_k, C_k, D_k \in \mathbf{R}$, and $\{w(n)\}$ is a sequence of independent, identically distributed (i.i.d.) AWGN noises with zero mean and an unknown variance of σ^2. Then, one can define the amplitude vector

$$\begin{aligned} \underline{\alpha} &= [A_1, \cdots, A_{N_H}, B_1, \cdots, B_{N_H}, \\ &\quad C_2, \cdots, C_{N_S}, D_2, \cdots, D_{N_S}]^T \in \mathbf{R}^P, \end{aligned} \tag{3.2}$$

and the sinusoidal signal vector

$$\begin{aligned} \underline{x}[n;T] &= [\cos(\omega n), \cdots, \cos(\omega N_H n), \sin(\omega n), \cdots, \\ &\quad \sin(\omega N_H n), \cos(\omega n/2), \cdots, \cos(\omega n/N_S), \\ &\quad \sin(\omega n/2), \cdots, \sin(\omega n/N_S)]^T \in \mathbf{R}^P, \end{aligned} \tag{3.3}$$

where $P = 2(N_H + N_S - 1)$. Compactly, Eqn. (3.1) can be rewritten as

$$s(n) = \underline{x}^T[n;T]\,\underline{\alpha} + w(n). \tag{3.4}$$

Let us define the parameter vector

$$\underline{\theta} = \left[\underline{\alpha}^T \; T \; \sigma^2\right]^T = [\theta_1, \theta_2, \cdots, \theta_P, \; \theta_{P+1}, \theta_{P+2}]^T, \tag{3.5}$$

where $[\theta_1, \theta_2, \cdots, \theta_P] = \underline{\alpha}^T$, $\theta_{P+1} = T$, and $\theta_{P+2} = \sigma^2$. Then, the problem under consideration is to estimate $\underline{\theta}$ from N points of noisy data $\underline{s} = [s(0), s(1), \cdots, s(N-1)]^T \in \mathbf{R}^N$. The PDF of $\underline{s}$ is given by

$$p(\underline{s};\underline{\theta}) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left[s(n) - \underline{x}^T[n;T]\,\underline{\alpha}\right]^2\right\}. \tag{3.6}$$

It is also convenient to define the matrix

$$X(T) = \begin{bmatrix} \underline{x}^T[0;T] \\ \underline{x}^T[1;T] \\ \vdots \\ \underline{x}^T[N-1;T] \end{bmatrix} \in \mathbf{R}^{N \times P}, \tag{3.7}$$

and

$$A(T) = X^T(T)\,X(T) \in \mathbf{R}^{P \times P}, \tag{3.8}$$

together with $\rho = \underline{s}^T \underline{s} \in \mathbf{R}$, and $\underline{g}(T) = X^T(T)\,\underline{s} \in \mathbf{R}^P$. [1] Then, Eqn. (3.6) may be rewritten as a log likelihood function (LLF) as[2]

$$\begin{aligned}\ln p(\underline{s};\underline{\theta}) &= \ln \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} - \frac{1}{2\sigma^2}\left(\underline{s} - X\underline{\alpha}\right)^T\left(\underline{s} - X\underline{\alpha}\right) \\ &= -\frac{N}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\left(\underline{\alpha}^T A\underline{\alpha} - 2\underline{g}^T\underline{\alpha} + \rho\right).\end{aligned} \tag{3.9}$$

The data model developed above will be used in the next section for the derivation of the desired CRLB bounds.

3.3 The Derivation of CRLBs

For the CRLB bounds to exist, the regularity condition [4]

$$E\left[\frac{\partial \ln p(\underline{s};\underline{\theta})}{\partial \underline{\theta}}\right] = \mathbf{0} \tag{3.10}$$

must be satisfied. From Theorem 3.2 of [4] it must be the case that for all $\underline{\theta}$ the regularity condition of Eqn. (3.10) holds. To begin, let $\underline{s} = [s(0), s(1), \cdots, s(N-1)]^T = X\underline{\alpha}$ and $\underline{w} = [w(0), w(1), \cdots, w(N-1)]^T$, so that

$$\underline{s} = X\underline{\alpha} + \underline{w}, \tag{3.11}$$

$$E[\underline{s}] = X\underline{\alpha}, \tag{3.12}$$

and

$$\begin{aligned}E[\underline{g}] &= E\left[X^T\underline{s}\right] \\ &= E\left[X^T(X\underline{\alpha} + \underline{w})\right] = X^T X\underline{\alpha} \\ &= A\underline{\alpha}.\end{aligned} \tag{3.13}$$

[1] Here, T is used in two different contexts, namely, as a subscript to denote matrix transposition, and as the signal period.

[2] In the following, signal dependence on T is omitted from the notation for the sake of simplicity. So $\underline{g}$ stands for $\underline{g}(T)$, X for $X(T)$, and A for $A(T)$, etc.

From Eqn. (3.9), the three components of the gradient of LLF are obtained as

$$\frac{\partial \ln p(\underline{s};\theta)}{\partial \underline{\alpha}} = \frac{1}{\sigma^2}\left(\underline{g} - A\underline{\alpha}\right), \tag{3.14}$$

$$\frac{\partial \ln p(\underline{s};\theta)}{\partial T} = \frac{1}{\sigma^2}\frac{\partial \underline{g}^T}{\partial T}\underline{\alpha} - \frac{1}{2\sigma^2}\underline{\alpha}^T\frac{\partial A}{\partial T}\underline{\alpha}, \tag{3.15}$$

$$\frac{\partial \ln p(\underline{s};\theta)}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\left(\sigma^2\right)^2}\left(\underline{\alpha}^T A\underline{\alpha} - 2\underline{g}^T\underline{\alpha} + \rho\right). \tag{3.16}$$

From Eqn. (3.14), one has

$$\begin{aligned} E\left[\frac{\partial \ln p(\underline{s};\theta)}{\partial \underline{\alpha}}\right] &= \frac{1}{\sigma^2}E\left[\underline{g} - A\underline{\alpha}\right] \\ &= \frac{1}{\sigma^2}\left[A\underline{\alpha} - A\underline{\alpha}\right] \\ &= 0, \end{aligned} \tag{3.17}$$

and the second regularity component

$$\begin{aligned} E\left[\frac{\partial \ln p(\underline{s};\theta)}{\partial T}\right] &= E\left[\frac{1}{\sigma^2}\frac{\partial \underline{g}^T}{\partial T}\underline{\alpha} - \frac{1}{2\sigma^2}\underline{\alpha}^T\frac{\partial A}{\partial T}\underline{\alpha}\right] \\ &= \frac{1}{2\sigma^2}\left[2E\left[\underline{s}^T\right]\frac{\partial X}{\partial T}\underline{\alpha} - \underline{\alpha}^T\frac{\partial A}{\partial T}\underline{\alpha}\right] \\ &= \frac{1}{2\sigma^2}\left[2\underline{\alpha}^T X^T\frac{\partial X}{\partial T} - \underline{\alpha}^T\left(\frac{\partial X^T}{\partial T}X + X^T\frac{\partial X}{\partial T}\right)\underline{\alpha}\right] \\ &= \frac{1}{2\sigma^2}\left[\underline{\alpha}^T X^T\frac{\partial X}{\partial T}\underline{\alpha} - \underline{\alpha}^T\frac{\partial X^T}{\partial T}X\underline{\alpha}\right] \\ &= \frac{1}{2\sigma^2}\left[\underline{\alpha}^T X^T\frac{\partial X}{\partial T}\underline{\alpha} - \underline{\alpha}^T X^T\frac{\partial X}{\partial T}\underline{\alpha}\right] \\ &= 0. \end{aligned} \tag{3.18}$$

Let us take a look at the third component of regularity condition:

$$\begin{aligned} E\left[\frac{\partial \ln p(\underline{s};\theta)}{\partial \sigma^2}\right] &= -\frac{N}{2\sigma^2} + \frac{1}{2\left(\sigma^2\right)^2}E\left[\underline{\alpha}^T A\underline{\alpha} - 2\underline{g}^T\underline{\alpha} + \rho\right] \\ &= -\frac{N}{2\sigma^2} + \frac{1}{2\left(\sigma^2\right)^2}\left[\underline{\alpha}^T A\underline{\alpha} - 2\underline{\alpha}^T A\left(T\right)\underline{\alpha} + E\left[\rho\right]\right]. \end{aligned} \tag{3.19}$$

By taking into account the fact that

$$\begin{aligned} E[\rho] &= E\left[\underline{s}^T \underline{s}\right] \\ &= E\left[[X\underline{\alpha} + \underline{w}]^T [X\underline{\alpha} + \underline{w}]\right] \\ &= E\left[\underline{\alpha}^T X^T X \underline{\alpha} + \underline{\alpha}^T X^T \underline{w} + \underline{w}^T X \underline{\alpha} + \underline{w}^T \underline{w}\right] \\ &= \underline{\alpha}^T A \underline{\alpha} + N\sigma^2, \end{aligned} \tag{3.20}$$

Eqn. (3.19) can be simplified as

$$\begin{aligned} E\left[\frac{\partial \ln p(\underline{s};\underline{\theta})}{\partial \sigma^2}\right] &= -\frac{N}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\left[\underline{\alpha}^T A \underline{\alpha} - 2\underline{\alpha}^T A \underline{\alpha} + \underline{\alpha}^T A \underline{\alpha} + N\sigma^2\right] \\ &= -\frac{N}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} N\sigma^2 \\ &= 0. \end{aligned} \tag{3.21}$$

In accordance with Eqns. (3.17), (3.18) and (3.21), the regularity condition in Eqn. (3.10) is satisfied.

The next step is to determine the elements of the CRLB matrix. Let us define

$$\Lambda = diag\left\{\sigma^2, \sigma^2, \cdots, \sigma^2\right\} \in \mathbf{R}^{N\times N} \tag{3.22}$$

which is the covariance matrix for a AWGN process. In this case, the last term of Eqn. (3.9) can be rewritten as

$$\ln p(\underline{s};\underline{\theta}) = -\frac{N}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2}\left[s - X\underline{\alpha}\right]^T \Lambda^{-1}\left[s - X\underline{\alpha}\right]. \tag{3.23}$$

In the derivations to follow, it is useful to note that

$$\begin{aligned} \Lambda &= E\left[(s - X\underline{\alpha})(s - X\underline{\alpha})^T\right] \\ &= E\left[\underline{ss}^T - \underline{s\alpha}^T X^T - X\underline{\alpha s}^T + X\underline{\alpha\alpha}^T X^T\right] \\ &= E\left[\underline{ss}^T\right] - X\underline{\alpha\alpha}^T X^T - X\underline{\alpha\alpha}^T X^T + X\underline{\alpha\alpha}^T X^T \\ &= E\left[\underline{ss}^T\right] - X\underline{\alpha\alpha}^T X^T, \end{aligned} \tag{3.24}$$

so that

$$E\left[\underline{ss}^T\right] = \Lambda + X\underline{\alpha\alpha}^T X^T. \tag{3.25}$$

From [4], the covariance matrix $C_{\hat{\theta}}$ of an unbiased estimate $\hat{\theta}$ of θ satisfies the CRLB bond

$$C_{\hat{\theta}} \geqslant F^{-1}(\underline{\theta}), \tag{3.26}$$

where $F(\underline{\theta})$ is the Fisher information matrix (FIM). For the problem under consideration, the FIM matrix is defined as

$$\begin{aligned} F(\underline{\theta}) &= \{F_{ij}(\underline{\theta}),\ i,j=1,2,3\} \\ &= E\left\{ \begin{bmatrix} \dfrac{\partial \ln p(\underline{s};\underline{\theta})}{\partial \underline{\alpha}} \\ \dfrac{\partial \ln p(\underline{s};\underline{\theta})}{\partial T} \\ \dfrac{\partial \ln p(\underline{s};\underline{\theta})}{\partial \sigma^2} \end{bmatrix} \begin{bmatrix} \left(\dfrac{\partial \ln p(\underline{s};\underline{\theta})}{\partial \underline{\alpha}}\right)^T & \dfrac{\partial \ln p(\underline{s};\underline{\theta})}{\partial T} & \dfrac{\partial \ln p(\underline{s};\underline{\theta})}{\partial \sigma^2} \end{bmatrix} \right\} && (3.27) \\ &\triangleq E\left\{G(\underline{\theta})\right\}, && (3.28) \end{aligned}$$

where

$$G(\underline{\theta}) = \begin{bmatrix} G_{11}(\underline{\theta}) & G_{12}(\underline{\theta}) & G_{13}(\underline{\theta}) \\ G_{21}(\underline{\theta}) & G_{22}(\underline{\theta}) & G_{23}(\underline{\theta}) \\ G_{31}(\underline{\theta}) & G_{32}(\underline{\theta}) & G_{33}(\underline{\theta}) \end{bmatrix}. \tag{3.29}$$

From Eqn. (3.27), it can be observed that $G(\underline{\theta}) = G^T(\underline{\theta})$. Then, the elements of $G(\underline{\theta})$ are in turn defined as

$$\begin{aligned} G_{11}(\underline{\theta}) &= \frac{\partial \ln p(\underline{s};\underline{\theta})}{\partial \underline{\alpha}} \left[\frac{\partial \ln p(\underline{s};\underline{\theta})}{\partial \underline{\alpha}}\right]^T, && (3.30) \\ G_{12}(\underline{\theta}) &= \frac{\partial \ln p(\underline{s};\underline{\theta})}{\partial \underline{\alpha}} \frac{\partial \ln p(\underline{s};\underline{\theta})}{\partial T}, && (3.31) \\ G_{13}(\underline{\theta}) &= \frac{\partial \ln p(\underline{s};\underline{\theta})}{\partial \underline{\alpha}} \frac{\partial \ln p(\underline{s};\underline{\theta})}{\partial \sigma^2}, && (3.32) \\ G_{21}(\underline{\theta}) &= G_{12}^T(\underline{\theta}), && (3.33) \\ G_{22}(\underline{\theta}) &= \frac{\partial \ln p(\underline{s};\underline{\theta})}{\partial T} \frac{\partial \ln p(\underline{s};\underline{\theta})}{\partial T}, && (3.34) \\ G_{23}(\underline{\theta}) &= \frac{\partial \ln p(\underline{s};\underline{\theta})}{\partial T} \frac{\partial \ln p(\underline{s};\underline{\theta})}{\partial \sigma^2}, && (3.35) \\ G_{31}(\underline{\theta}) &= G_{13}^T(\underline{\theta}), && (3.36) \\ G_{32}(\underline{\theta}) &= G_{23}^T(\underline{\theta}), && (3.37) \\ G_{33}(\underline{\theta}) &= \left[\frac{\partial \ln p(\underline{s};\underline{\theta})}{\partial \sigma^2}\right]^2. && (3.38) \end{aligned}$$

In this way, one should compute the matrix elements $G_{11}(\underline{\theta})$, $G_{12}(\underline{\theta})$, $G_{13}(\underline{\theta})$, $G_{22}(\underline{\theta})$, $G_{23}(\underline{\theta})$ and $G_{33}(\underline{\theta})$. It can be shown that

$$\begin{aligned} G_{11}(\underline{\theta}) &= \frac{1}{(\sigma^2)^2}\left[\underline{g} - A\underline{\alpha}\right]\left[\underline{g} - A\underline{\alpha}\right]^T \\ &= \frac{1}{(\sigma^2)^2}\left[X^T\underline{s} - X^T X\underline{\alpha}\right]\left[\underline{s}^T X - \underline{\alpha}^T X^T X\right] \\ &= \frac{1}{(\sigma^2)^2} X^T \underline{s}\underline{s}^T X - X^T \underline{s}\underline{\alpha}^T X^T X - X^T X\underline{\alpha}\underline{s}^T X + X^T X\underline{\alpha}\underline{\alpha}^T X X^T. \end{aligned} \quad (3.39)$$

As $F_{11}(\underline{\theta}) = E\left[G_{11}(\underline{\theta})\right]$, one can obtain

$$\begin{aligned} &(\sigma^2)^2 F_{11}(\underline{\theta}) \\ &= X^T E\left[\underline{s}\underline{s}^T\right] X - X^T E\left[\underline{s}\right]\underline{\alpha}^T X^T X - X^T X\underline{\alpha} E\left[\underline{s}^T\right] X + X^T X\underline{\alpha}\underline{\alpha}^T X X^T \\ &= X^T\left(\Lambda + X\underline{\alpha}\underline{\alpha}^T X^T\right) X - X^T X\underline{\alpha}\underline{\alpha}^T X^T X - X^T X\underline{\alpha}\underline{\alpha}^T X^T X \\ &\quad + X^T X\underline{\alpha}\underline{\alpha}^T X X^T \\ &= X^T \Lambda X, \end{aligned} \quad (3.40)$$

leading to

$$F_{11}(\underline{\theta}) = \frac{1}{\sigma^2} X^T X. \quad (3.41)$$

Eqns. (3.12) and (3.20) have been used in the derivation of Eqn. (3.40).

Now, substituting Eqns. (3.14) and (3.15) into Eqn. (3.31) leads to

$$\begin{aligned} &G_{12}(\underline{\theta}) \\ &= \frac{1}{(\sigma^2)^2}\left[\underline{\alpha}^T \frac{\partial \underline{g}}{\partial T} - \frac{1}{2}\underline{\alpha}^T \frac{\partial A}{\partial T}\underline{\alpha}\right]\left[\underline{g} - A\underline{\alpha}\right] \\ &= \frac{1}{(\sigma^2)^2}\left[\underline{\alpha}^T \left(\frac{\partial X}{\partial T}\right)^T \underline{s} - \frac{1}{2}\underline{\alpha}^T \frac{\partial A}{\partial T}\underline{\alpha}\right]\left[X^T\underline{s} - A\underline{\alpha}\right] \\ &= \frac{1}{(\sigma^2)^2}\left[X^T \underline{s}\underline{s}^T \frac{\partial X}{\partial T}\underline{\alpha} - A\underline{\alpha}\underline{s}^T \frac{\partial X}{\partial T}\underline{\alpha} - \frac{1}{2} X^T \underline{s}\underline{\alpha}^T \frac{\partial A}{\partial T}\underline{\alpha} + \frac{1}{2} A\underline{\alpha}\underline{\alpha}^T \frac{\partial A}{\partial T}\underline{\alpha}\right]. \end{aligned} \quad (3.42)$$

As $F_{12}(\underline{\theta}) = E[G_{12}(\underline{\theta})]$, one has

$$
\begin{aligned}
(\sigma^2)^2 F_{12}(\underline{\theta}) &= (\sigma^2)^2 E[G_{12}(\underline{\theta})] \\
&= X^T \left[\Lambda + X\underline{\alpha}\underline{\alpha}^T X^T\right] \frac{\partial X}{\partial T}\underline{\alpha} - \Lambda\underline{\alpha}\underline{\alpha}^T X^T \frac{\partial X}{\partial T}\underline{\alpha}(\sigma^2)^2 \\
&\quad - \frac{1}{2} X^T X \underline{\alpha}\underline{\alpha}^T \frac{\partial \Lambda}{\partial T}\underline{\alpha} + \frac{1}{2}\Lambda\underline{\alpha}\underline{\alpha}^T \frac{\partial \Lambda}{\partial T}\underline{\alpha} \\
&= X^T \Lambda \frac{\partial X}{\partial T}\underline{\alpha},
\end{aligned} \tag{3.43}
$$

where Eqns. (3.12) and (3.20) have been used in the derivation of Eqn. (3.43). Therefore, one can obtain

$$
F_{12} = \frac{1}{\sigma^2} X^T \frac{\partial X}{\partial T}\underline{\alpha}. \tag{3.44}
$$

Next, from Eqn. (3.34) and $F_{22}(\underline{\theta}) = E[G_{22}(\underline{\theta})]$, it can be shown that

$$
\begin{aligned}
(\sigma^2)^2 F_{22}(\underline{\theta}) &= (\sigma^2)^2 E[G_{22}(\underline{\theta})] \\
&= E\left[\left(\underline{\alpha}^T \frac{\partial \underline{g}}{\partial T} - \frac{1}{2}\underline{\alpha}^T \frac{\partial \Lambda}{\partial T}\underline{\alpha}\right)^2\right] \\
&= E\left\{\left[\underline{\alpha}^T \left(\frac{\partial X}{\partial T}\right)^T \underline{s} - \frac{1}{2}\underline{\alpha}^T \frac{\partial \Lambda}{\partial T}\underline{\alpha}\right]^2\right\} \\
&= \underline{\alpha}^T \left(\frac{\partial X}{\partial T}\right)^T E\left[\underline{s}\underline{s}^T\right] \frac{\partial X}{\partial T}\underline{\alpha} - \frac{1}{2}\underline{\alpha}^T \frac{\partial \Lambda}{\partial T}\underline{\alpha}\underline{\alpha}^T \left(\frac{\partial X}{\partial T}\right)^T X\underline{\alpha} \\
&\quad - \frac{1}{2}\underline{\alpha}^T \frac{\partial \Lambda}{\partial T}\underline{\alpha}\underline{\alpha}^T \left(\frac{\partial X}{\partial T}\right)^T X\underline{\alpha} + \frac{1}{4}\underline{\alpha}^T \frac{\partial \Lambda}{\partial T}\underline{\alpha}\underline{\alpha}^T \frac{\partial \Lambda}{\partial T}\underline{\alpha} \\
&= \underline{\alpha}^T \left(\frac{\partial X}{\partial T}\right)^T \left[\Lambda + X\underline{\alpha}\underline{\alpha}^T X^T\right] \frac{\partial X}{\partial T}\underline{\alpha} - \frac{1}{2}\underline{\alpha}^T \frac{\partial \Lambda}{\partial T}\underline{\alpha}\underline{\alpha}^T \left(\frac{\partial X}{\partial T}\right)^T E[\underline{s}] \\
&\quad - \frac{1}{2}\underline{\alpha}^T \frac{\partial \Lambda}{\partial T}\underline{\alpha}\underline{\alpha}^T \left(\frac{\partial X}{\partial T}\right)^T E[\underline{s}] + \frac{1}{4}\underline{\alpha}^T \frac{\partial \Lambda}{\partial T}\underline{\alpha}\underline{\alpha}^T \frac{\partial \Lambda}{\partial T}\underline{\alpha} \\
&= \underline{\alpha}^T \left(\frac{\partial X}{\partial T}\right)^T \left[\Lambda + X\underline{\alpha}\underline{\alpha}^T X^T\right] \frac{\partial X}{\partial T}\underline{\alpha} - \underline{\alpha}^T \frac{\partial \left(X^T X\right)}{\partial T}\underline{\alpha}\underline{\alpha}^T \left(\frac{\partial X}{\partial T}\right)^T X\underline{\alpha} \\
&\quad + \frac{1}{4}\underline{\alpha}^T \frac{\partial \Lambda}{\partial T}\underline{\alpha}\underline{\alpha}^T \frac{\partial \Lambda}{\partial T}\underline{\alpha} \\
&= \underline{\alpha}^T \left(\frac{\partial X}{\partial T}\right)^T \Lambda \frac{\partial X}{\partial T}\underline{\alpha} - \underline{\alpha}^T X^T \frac{\partial X}{\partial T}\underline{\alpha}\underline{\alpha}^T X \frac{\partial X}{\partial T}\underline{\alpha} + \frac{1}{4}\underline{\alpha}^T \frac{\partial \Lambda}{\partial T}\underline{\alpha}\underline{\alpha}^T \frac{\partial \Lambda}{\partial T}\underline{\alpha},
\end{aligned} \tag{3.45}
$$

where Eqns. (3.12) and (3.20) have been used in the derivation of Eqn. (3.45). The last term of Eqn. (3.45) can be written as

$$
\begin{aligned}
\frac{1}{4}\underline{\alpha}^T\frac{\partial A}{\partial T}\underline{\alpha\alpha}^T\frac{\partial A}{\partial T}\underline{\alpha} &= \frac{1}{4}\underline{\alpha}^T\left[\frac{\partial X^T}{\partial T}X + X^T\frac{\partial X}{\partial T}\right]\underline{\alpha\alpha}^T\left[\frac{\partial X^T}{\partial T}X + X^T\frac{\partial X}{\partial T}\right]\underline{\alpha} \\
&= \frac{1}{4}\left[\underline{\alpha}^T\frac{\partial X^T}{\partial T}X\underline{\alpha\alpha}^T\frac{\partial X^T}{\partial T}X\underline{\alpha} + \underline{\alpha}^T\frac{\partial X^T}{\partial T}X\underline{\alpha\alpha}^T X^T\frac{\partial X}{\partial T}\underline{\alpha}\right. \\
&\quad \left. + \underline{\alpha}^T X^T\frac{\partial X}{\partial T}\underline{\alpha\alpha}^T\frac{\partial X^T}{\partial T}X\underline{\alpha} + \underline{\alpha}^T X^T\frac{\partial X}{\partial T}\underline{\alpha\alpha}^T X^T\frac{\partial X}{\partial T}\underline{\alpha}\right] \\
&= \frac{1}{4}\left[\underline{\alpha}^T\frac{\partial X^T}{\partial T}X\underline{\alpha\alpha}^T X^T\frac{\partial X}{\partial T}\underline{\alpha} + \underline{\alpha}^T\frac{\partial X^T}{\partial T}X\underline{\alpha\alpha}^T X^T\frac{\partial X}{\partial T}\underline{\alpha}\right. \\
&\quad \left. + \underline{\alpha}^T X^T\frac{\partial X}{\partial T}\underline{\alpha\alpha}^T X^T\frac{\partial X}{\partial T}\underline{\alpha} + \underline{\alpha}^T X^T\frac{\partial X}{\partial T}\underline{\alpha\alpha}^T X^T\frac{\partial X}{\partial T}\underline{\alpha}\right] \\
&= \frac{1}{2}\left[\underline{\alpha}^T X^T\frac{\partial X}{\partial T}\underline{\alpha\alpha}^T X^T\frac{\partial X}{\partial T}\underline{\alpha} + \underline{\alpha}^T X^T\frac{\partial X}{\partial T}\underline{\alpha\alpha}^T X^T\frac{\partial X}{\partial T}\underline{\alpha}\right] \\
&= \underline{\alpha}^T X^T\frac{\partial X}{\partial T}\underline{\alpha\alpha}^T X^T\frac{\partial X}{\partial T}\underline{\alpha}. \qquad (3.46)
\end{aligned}
$$

Substituting Eqn. (3.46) into Eqn. (3.45) gives

$$
\begin{aligned}
F_{22}(\underline{\theta}) &= E\left[G_{22}(\underline{\theta})\right] \qquad (3.47) \\
&= \frac{1}{(\sigma^2)^2}\underline{\alpha}^T\left(\frac{\partial X}{\partial T}\right)^T A\frac{\partial X}{\partial T}\underline{\alpha} \qquad (3.48) \\
&= \frac{1}{\sigma^2}\underline{\alpha}^T\left(\frac{\partial X}{\partial T}\right)^T\frac{\partial X}{\partial T}\underline{\alpha}. \qquad (3.49)
\end{aligned}
$$

It should be pointed out that the identity $\frac{1}{\sigma^2}A = I \in \mathbf{R}^{N\times N}$, where I is a $N \times N$ identity matrix., has been used in the derivation from Eqn. (3.47) to Eqn. (3.49).

By following derivations similar to those leading to Eqns. (3.49), one can obtain

$$
\begin{aligned}
E\left[G_{13}(\underline{\theta})\right] &= E\left\{\frac{1}{\sigma^2}\left[\underline{g} - A\underline{\alpha}\right]\left[-\frac{N}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\left[\underline{\alpha}^T A\underline{\alpha} - 2\underline{g}^T\underline{\alpha} + \rho\right]\right]\right\} \\
&= \frac{1}{2(\sigma^2)^3}E\left[\left(\underline{g} - A\underline{\alpha}\right)\left(\underline{\alpha}^T A\underline{\alpha} - 2\underline{g}^T\underline{\alpha} + \rho\right)\right] \\
&= \frac{1}{2(\sigma^2)^3}E\left[\underline{g\alpha}^T A\underline{\alpha} - 2\underline{gg}^T\underline{\alpha} + \rho\underline{g} - A\underline{\alpha\alpha}^T A\underline{\alpha} + 2A\underline{\alpha g}^T\underline{\alpha} - A\underline{\alpha}\rho\right] \\
&= \frac{1}{2(\sigma^2)^3}\left[A\underline{\alpha\alpha}^T A\underline{\alpha} - 2\left(A\underline{\alpha\alpha}^T A + \sigma^2 A\right)\underline{\alpha} + A\underline{\alpha\alpha}^T A\underline{\alpha} + (N+2)\sigma^2 A\underline{\alpha}\right. \\
&\quad \left. - A\underline{\alpha\alpha}^T A\underline{\alpha} + 2A\underline{\alpha\alpha}^T A\underline{\alpha} - A\underline{\alpha}\left(\underline{\alpha}^T A\underline{\alpha} + N\sigma^2\right)\right] \\
&= \frac{1}{2(\sigma^2)^3}\left[A\underline{\alpha\alpha}^T A\underline{\alpha} - 2\sigma^2 A\underline{\alpha} + (N+2)\sigma^2 A\underline{\alpha} - A\underline{\alpha\alpha}^T A\underline{\alpha} - N\sigma^2 A\underline{\alpha}\right] \\
&= 0. \qquad (3.50)
\end{aligned}
$$

It should also be pointed out that the following identities (see Appendix C for the derivations)

$$E\left[\underline{g}\underline{g}^T\right] = A\underline{\alpha}\underline{\alpha}^T A\alpha + \sigma^2 A, \tag{3.51}$$

$$E\left[\rho\underline{g}\right] = A\underline{\alpha}\underline{\alpha}^T A\underline{\alpha} + (N+2)\sigma^2 A\underline{\alpha}, \tag{3.52}$$

and

$$E\left[\rho\right] = \underline{\alpha}^T A\underline{\alpha} + N\sigma^2, \tag{3.53}$$

have been used in the derivation of Eqn. (3.50).

Proceeding further, one can obtain

$$F_{13}\left(\underline{\theta}\right) = E\left[G_{13}\left(\underline{\theta}\right)\right] = 0. \tag{3.54}$$

Similarly, one has

$$G_{23}\left(\underline{\theta}\right) = \left(\frac{1}{\sigma^2}\frac{\partial \underline{g}^T}{\partial T}\underline{\alpha} - \frac{1}{2\sigma^2}\underline{\alpha}^T\frac{\partial A}{\partial T}\underline{\alpha}\right)\left[-\frac{N}{2\sigma^2} + \frac{1}{2\left(\sigma^2\right)^2}\left(\underline{\alpha}^T A\underline{\alpha} - 2\underline{g}^T\underline{\alpha} + \rho\right)\right], \tag{3.55}$$

and it can be shown that

$$\begin{aligned}
F_{23}\left(\underline{\theta}\right) &= E\left[G_{23}\left(\underline{\theta}\right)\right] \\
&= -\frac{N}{2\sigma^2}E\left[\frac{\partial \ln p(\underline{s};\underline{\theta})}{\partial T}\right] \\
&\quad +\frac{1}{2\sigma^4}E\left[\left(\underline{\alpha}^T A\underline{\alpha} - 2\underline{g}^T\underline{\alpha} + \rho\right)\left(\frac{1}{\sigma^2}\frac{\partial \underline{g}^T}{\partial T}\underline{\alpha} - \frac{1}{2\sigma^2}\underline{\alpha}^T\frac{\partial A}{\partial T}\underline{\alpha}\right)\right] \\
&= \frac{1}{4\sigma^6}E\left[\left(\underline{\alpha}^T A\underline{\alpha} - 2\underline{g}^T\underline{\alpha} + \rho\right)\left(2\underline{s}^T\frac{\partial X}{\partial T}\underline{\alpha} - \underline{\alpha}^T\frac{\partial A}{\partial T}\underline{\alpha}\right)\right] \\
&= \frac{1}{4\sigma^6}E\left[2\underline{\alpha}^T A\underline{\alpha}\underline{s}^T\frac{\partial X}{\partial T}\underline{\alpha} - 4\underline{\alpha}^T\underline{g}\underline{s}^T\frac{\partial X}{\partial T}\underline{\alpha} + 2\rho\underline{s}^T\frac{\partial X}{\partial T}\underline{\alpha}\right. \\
&\quad \left. - \left[\underline{\alpha}^T A\underline{\alpha} - 2\underline{g}^T\underline{\alpha} + \rho\right]\underline{\alpha}^T\frac{\partial A}{\partial T}\underline{\alpha}\right] \\
&= \frac{1}{4\sigma^6}\left[2\underline{\alpha}^T A\underline{\alpha}\underline{\alpha}^T X^T\frac{\partial X}{\partial T}\underline{\alpha} - 4\underline{\alpha}^T\left(A\underline{\alpha}\underline{\alpha}^T X^T + \sigma^2 X^T\right)\frac{\partial X}{\partial T}\underline{\alpha} + \right. \\
&\quad +2\left(\underline{\alpha}^T A\underline{\alpha}\underline{\alpha}^T X^T + (N+2)\sigma^2\underline{\alpha}^T X^T\right)\frac{\partial X}{\partial T}\underline{\alpha} - \\
&\quad \left. - \left(\underline{\alpha}^T A\underline{\alpha} - 2\underline{g}^T\underline{\alpha} + \underline{\alpha}^T A\underline{\alpha} + N\sigma^2\right)\underline{\alpha}^T\frac{\partial A}{\partial T}\underline{\alpha}\right]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{4\sigma^6}\left[2N\sigma^2\underline{\alpha}^T X^T \frac{\partial X}{\partial T}\underline{\alpha} - \left(\underline{\alpha}^T A\underline{\alpha} - 2\underline{\alpha}^T A\underline{\alpha} + \underline{\alpha}^T A\underline{\alpha} + N\sigma^2\right)\underline{\alpha}^T\frac{\partial A}{\partial T}\underline{\alpha}\right] \\
&= \frac{1}{4\sigma^6}\left(2N\sigma^2\underline{\alpha}^T X^T \frac{\partial X}{\partial T}\underline{\alpha} - N\sigma^2\underline{\alpha}^T\frac{\partial A}{\partial T}\underline{\alpha}\right) \\
&= \frac{N}{4\sigma^4}\left[2\underline{\alpha}^T X^T \frac{\partial X}{\partial T}\underline{\alpha} - \underline{\alpha}^T\left(\frac{\partial X^T}{\partial T}X + X^T\frac{\partial X}{\partial T}\right)\underline{\alpha}\right] \\
&= \frac{N}{4\sigma^4}\left[\underline{\alpha}^T X^T \frac{\partial X}{\partial T}\underline{\alpha} - \underline{\alpha}^T\frac{\partial X^T}{\partial T}X\underline{\alpha}\right] \\
&= 0,
\end{aligned} \tag{3.56}
$$

where the following identity has been used (see Appendix C for the derivations):

$$
E\left[\underline{g}\underline{s}^T\right] = A\underline{\alpha}\underline{\alpha}^T X^T + \sigma^2 X^T, \tag{3.57}
$$

and

$$
E\left[\rho\underline{s}^T\right] = \underline{\alpha}^T A\underline{\alpha}\underline{\alpha}^T X^T + (N+2)\sigma^2\underline{\alpha}^T X^T. \tag{3.58}
$$

Next, let us derive $F_{33}(\underline{\theta})$ by firstly looking at

$$
\begin{aligned}
G_{33}(\underline{\theta}) &= \left\{-\frac{N}{2\sigma^2} + \frac{1}{2\left(\sigma^2\right)^2}\left[\underline{\alpha}^T A\underline{\alpha} - 2\underline{g}^T\underline{\alpha} + \rho\right]\right\}^2 \\
&= \frac{1}{4\left(\sigma^2\right)^4}\left(\underline{\alpha}^T A\underline{\alpha} - 2\underline{g}^T\underline{\alpha} + \rho - N\sigma^2\right)^2.
\end{aligned} \tag{3.59}
$$

After some mathematical manipulations on Eqn. (3.59), one can obtain [3]

$$
\begin{aligned}
4\left(\sigma^2\right)^4 E\left[G_{33}(\underline{\theta})\right] &= E\left[\left(\underline{\alpha}^T A\underline{\alpha} - 2\underline{g}^T\underline{\alpha} + \rho - N\sigma^2\right)^2\right] \\
&= 2N\left(\sigma^2\right)^2.
\end{aligned} \tag{3.60}
$$

Then,

$$
F_{33}(\underline{\theta}) = E\left[G_{33}(\underline{\theta})\right] = \frac{N}{2\left(\sigma^2\right)^2}. \tag{3.61}
$$

[3] See Appendix C for detailed derivation.

Finally, one can arrive at the FIM matrix as

$$F(\underline{\theta}) = \frac{1}{\sigma^2}\begin{bmatrix} X^TX & X^T\frac{\partial X}{\partial T}\underline{\alpha} & 0 \\ \underline{\alpha}^T\left(\frac{\partial X}{\partial T}\right)^T X & \underline{\alpha}^T\left(\frac{\partial X}{\partial T}\right)^T\frac{\partial X}{\partial T}\underline{\alpha} & 0 \\ 0 & 0 & \frac{N}{2\sigma^2} \end{bmatrix} \quad (3.62)$$

$$= \begin{bmatrix} B & 0 \\ 0 & \frac{N}{2\left(\sigma^2\right)^2} \end{bmatrix}, \quad (3.63)$$

where

$$B = \frac{1}{\sigma^2}\begin{bmatrix} X^TX & X^T\frac{\partial X}{\partial T}\underline{\alpha} \\ \underline{\alpha}^T\left(\frac{\partial X}{\partial T}\right)^T X & \underline{\alpha}^T\left(\frac{\partial X}{\partial T}\right)^T\frac{\partial X}{\partial T}\underline{\alpha} \end{bmatrix}. \quad (3.64)$$

The inverse of the FIM matrix in Eqn. (3.62) is given by

$$F^{-1}(\underline{\theta}) = \begin{bmatrix} B^{-1} & 0 \\ 0 & \frac{2\left(\sigma^2\right)^2}{N} \end{bmatrix}. \quad (3.65)$$

In order to calculate the inverse of B, let us represent B in a partitioned form as

$$B = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}. \quad (3.66)$$

Then, the inverse is given by

$$B^{-1} = \begin{bmatrix} \left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right)^{-1} & -\left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right)^{-1}A_{12}A_{22}^{-1} \\ -\left(A_{22} - A_{21}A_{11}^{-1}A_{12}\right)^{-1}A_{21}A_{11}^{-1} & \left(A_{22} - A_{21}A_{11}^{-1}A_{12}\right)^{-1} \end{bmatrix}, \quad (3.67)$$

provided that A_{11} and A_{22} are invertible. By comparing Eqn. (3.66) to Eqn. (3.64), one can make the following identifications:

$$A_{11} = \frac{1}{\sigma^2}X^TX, \quad (3.68)$$

$$A_{12} = \frac{1}{\sigma^2}X^T\frac{\partial X}{\partial T}\underline{\alpha}, \quad (3.69)$$

$$A_{21} = \frac{1}{\sigma^2}\underline{\alpha}^T\left(\frac{\partial X}{\partial T}\right)^T X, \quad (3.70)$$

$$A_{22} = \frac{1}{\sigma^2}\underline{\alpha}^T\left(\frac{\partial X}{\partial T}\right)^T\frac{\partial X}{\partial T}\underline{\alpha}. \quad (3.71)$$

If $\widehat{T}$ is an unbiased estimate of T, then it has a variance $var(\widehat{T})$, around its mean value. Then, $var(\widehat{T}) \geq V$, where V is the CRLB bound given by

$$\begin{aligned} V &= \left(A_{22} - A_{21}A_{11}^{-1}A_{12}\right)^{-1} \\ &= \left[\frac{1}{\sigma^2}\underline{\alpha}^T\left(\frac{\partial X}{\partial T}\right)^T\frac{\partial X}{\partial T}\underline{\alpha} - \frac{1}{\sigma^2}\underline{\alpha}^T\left(\frac{\partial X}{\partial T}\right)^T X\left(\frac{1}{\sigma^2}X^TX\right)^{-1}\frac{1}{\sigma^2}X^T\frac{\partial X}{\partial T}\underline{\alpha}\right]^{-1} \\ &= \sigma^2\left[\underline{\alpha}^T\left(\frac{\partial X}{\partial T}\right)^T\frac{\partial X}{\partial T}\underline{\alpha} - \underline{\alpha}^T\left(\frac{\partial X}{\partial T}\right)^T X\left(X^TX\right)^{-1}X^T\frac{\partial X}{\partial T}\underline{\alpha}\right]^{-1} \\ &= \sigma^2\left[\underline{\alpha}^T\left(\frac{\partial X}{\partial T}\right)^T\left[I - X\left[X^TX\right]^{-1}X^T\right]\frac{\partial X}{\partial T}\underline{\alpha}\right]. \end{aligned} \tag{3.72}$$

Similarly, if $\hat{\underline{\alpha}}$ is an unbiased estimate of $\underline{\alpha}$, then it has a variance $var(\hat{\underline{\alpha}})$, around its mean value. So, $var(\hat{\underline{\alpha}}) \geq U$, where U is the CRLB bound for estimation of $\underline{\alpha}$ given by

$$\begin{aligned} U &= \left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right)^{-1} \\ &= \left[\frac{1}{\sigma^2}X^TX - \frac{1}{\sigma^2}X^T\frac{\partial X}{\partial T}\underline{\alpha}\left(\frac{1}{\sigma^2}\underline{\alpha}^T\left(\frac{\partial X}{\partial T}\right)^T\frac{\partial X}{\partial T}\underline{\alpha}\right)^{-1}\frac{1}{\sigma^2}\underline{\alpha}^T\left(\frac{\partial X}{\partial T}\right)^T X\right]^{-1} \\ &= \sigma^2\left[X^TX - X^T\frac{\partial X}{\partial T}\underline{\alpha}\left(\underline{\alpha}^T\left(\frac{\partial X}{\partial T}\right)^T\frac{\partial X}{\partial T}\underline{\alpha}\right)^{-1}\underline{\alpha}^T\left(\frac{\partial X}{\partial T}\right)^T X\right]^{-1} \\ &= \sigma^2\left[X^T\left[I - \frac{\partial X}{\partial T}\underline{\alpha}\left(\left(\frac{\partial X}{\partial T}\underline{\alpha}\right)^T\frac{\partial X}{\partial T}\underline{\alpha}\right)^{-1}\left(\frac{\partial X}{\partial T}\underline{\alpha}\right)^T\right]X\right]^{-1}. \end{aligned} \tag{3.73}$$

The structure of Eqn. (3.73) has a geometric interpretation. From [32], let us consider a real-valued matrix $H \in \mathbf{R}^{N\times p}$ such that $p < N$, where p is the rank of H. Then, the matrix operator

$$P_H = H\left(H^TH\right)^{-1}H^T \in \mathbf{R}^{N\times N} \tag{3.74}$$

is an orthogonal projection matrix (i.e., it is idempotent[4] and symmetric). This implies that the real-valued vector $x \in \mathbf{R}^N$ is projected into the subspace of vector

[4] A matrix operator P is idempotent if $P^2 = P$

space $\mathbf{R}^N$ spanned by the columns of H under the operation $P_H x$. This theorem will be used in the problem under consideration.

Let

$$\underline{y} = \frac{\partial X}{\partial T}\underline{\alpha} \in \mathbf{R}^N \tag{3.75}$$

This implies that

$$U = \sigma^2 \left[\underline{y}^T \left[I - \underline{y}\left(\underline{y}^T \underline{y}\right)^{-1} \underline{y}^T\right] X\right]^{-1} \tag{3.76}$$

(due to the idempotency and symmetry of orthogonal projections). The matrix operator

$$P_Y = \underline{y}\left(\underline{y}^T \underline{y}\right)^{-1} \underline{y}^T \in \mathbf{R}^{N \times N} \tag{3.77}$$

is an orthogonal projection matrix. Therefore,

$$\begin{aligned} U &= \sigma^2 \left[X^T P_Y^{\perp} X\right]^{-1} && (3.78) \\ &= \sigma^2 \left[\left(P_Y^{\perp} X\right)^T P_Y^{\perp} X\right]^{-1} && (3.79) \\ &\in \mathbf{R}^{P \times P}. && (3.80) \end{aligned}$$

In Eqn. (3.78), $P_Y^{\perp} X$ is a length N column vector, hence $\left(P_Y^{\perp} X\right)^T P_Y^{\perp} X$ is its energy. This means that U is a ratio of noise power σ^2 to signal power, having the connotation of the inverse of SNR. This is to be expected.

3.3.1 CRLB bound for SNR estimator

According to [4], the CRLB bound for estimator $\alpha = g\left(\underline{\theta}\right)$ is defined as

$$\frac{\partial g\left(\underline{\theta}\right)}{\partial \underline{\theta}} F\left(\underline{\theta}\right) \frac{\partial g\left(\underline{\theta}\right)^T}{\partial \underline{\theta}}. \tag{3.81}$$

Here, the SNR ratio for the MICA signal model is defined as the ratio of the sum of fundamental frequency component power to the noise power in accordance with

$$\beta = h\left(\underline{\theta}\right) = \frac{A_1^2 + B_1^2}{\sigma^2}.$$

Therefore, the CRLB bound for the SNR estimation is obtained as

$$S = \frac{\partial h}{\partial \underline{\theta}} F^{-1}(\underline{\theta}) \left(\frac{\partial h}{\partial \underline{\theta}} \right)^T, \tag{3.82}$$

where

$$\frac{\partial h}{\partial \underline{\theta}} = \left[\underbrace{\frac{2A_1}{\sigma^2} 0 \cdots 0}_{N_H} \underbrace{\frac{2B_1}{\sigma^2} 0 \cdots 0}_{N_H} \underbrace{\underline{00 \cdots 0}}_{N_S - 1} \underbrace{\underline{00 \cdots 0}}_{N_S - 1} \, 0 \, \frac{-A_1^2 - B_1^2}{(\sigma^2)^2} \right].$$

The above derived CRLB bounds for parameter estimation are useful in estimating how much data needs to be collected in order to achieve the desired accuracy for the MICA model. Much simulation and analytical work will be needed to verify and interpret the bounds in Eqns. (3.72), (3.73) and (3.82). This will be done by computer simulation in the next section.

3.4 The Simulation Results

This section is concerned with a computational confirmation of the CRLB bounds in Eqns. (3.72), (3.73) and (3.82). This is achieved by the Monte Carlo simulations for the above parameters using MLE algorithm. The simulations are based on $\underline{\alpha} = [1, 0.2, 0.05, 0.01, 1, 0.2, 0.05, 0.01, 4, 6, 4, 6]^T$ in Eqn. (3.2). The data size is chosen as $N = 500$, as during the simulations, if the data size is too small (e.g. $N = 100$), the condition number of the matrix A becomes very large, rendering A close to singular. This makes the estimation less accurate or even unreliable. Therefore, the data size of 500 is used to obtain the desired estimation accuracy. For each SNR point, 500 independent Monte Carlo simulation results are carried out and the results are averaged to arrive at an estimation.

The estimation variances for parameters $A1$, $B1$ are plotted versus SNR (in dB) in Fig. 3.1, the estimation variance of σ^2 and SNR are plotted in Fig. 3.2, and the estimation variance of T is plotted in Fig. 3.3. It is observed from Figs. 3.1 - 3.2 that the simulated variance approaches the CRLB bound as SNR increases (as expected).

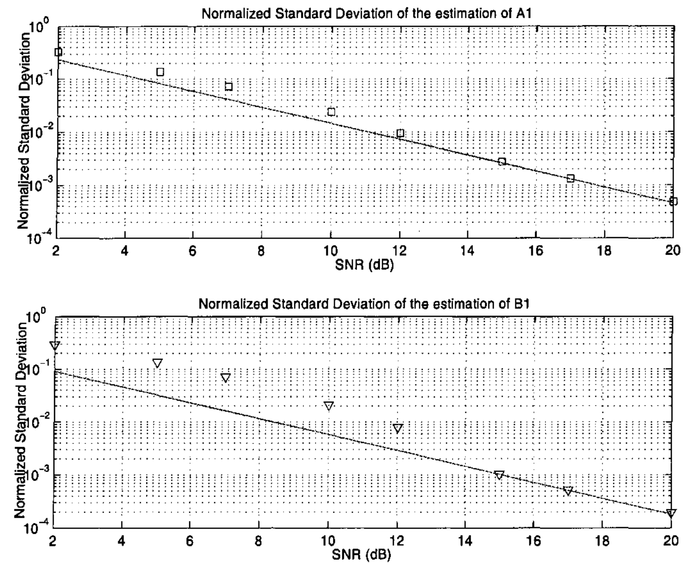

Figure 3.1: The normalized standard deviation of the estimation of $A1$ (top) and $B1$ (bottom) versus SNR together with the CRLB bound (solid curves).

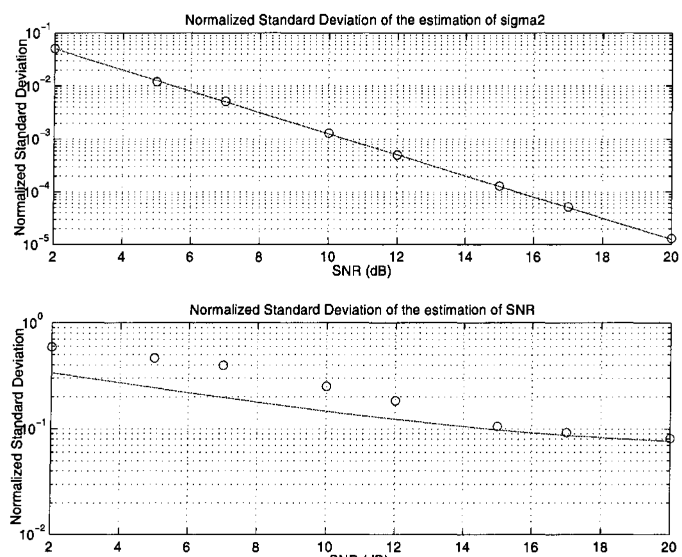

Figure 3.2: The normalized standard deviation of the estimation of σ^2 (top) and SNR (bottom) versus SNR together with the CRLB bound (solid curves).

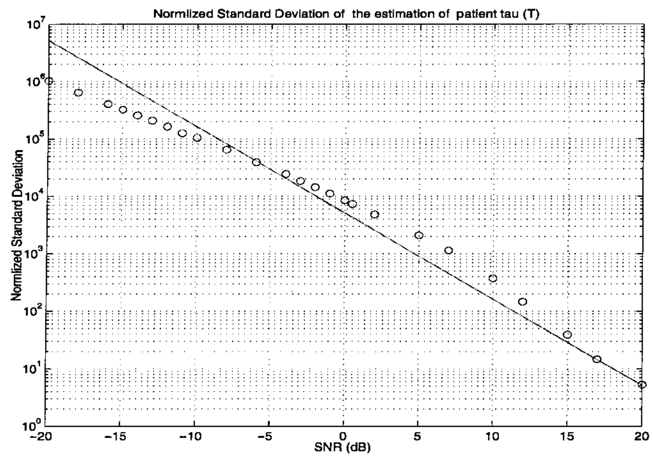

Figure 3.3: The normlized standard deviation of the estimation of T together with CRLB bound (solid curve).

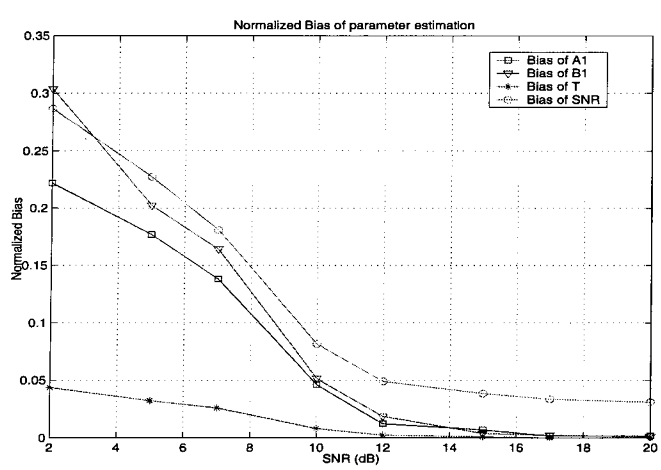

Figure 3.4: The estimation bias versus SNR.

Since CRLB bound is the theoretical limit on the best possible unbiased estimation performance, no other estimation method can be expected to perform better than MLE at high SNRs. With decreasing SNR, the estimation variance becomes worse than the bound (again, as expected). The simulation results verify the validity of the derived CRLB bounds. However, at SNRs smaller than -5 dB, the simulation estimation variance for T does not conform to the bound (c.f. Fig. 3.3), simply because the CRLB bound is valid for unbiased estimation only. At very low SNRs, the estimator tends to be biased as clear from Fig. 3.4, and the unbiased CRLB bound is not the valid bound anymore. In this case, the biased CRLB bound should be considered instead. For the simulation results, the unbiased CRLB bound is a very good predictor of the true variance for SNRs greater than about 14 or 15 dB.

3.5 Conclusions

This chapter has presented CRLB bounds for the unbiased estimation of parameters in circadian rhythm data model MICA under the assumption that noise variance is not known a priori. Much attention has been given to the detailed derivation of the Fisher information matrix in order to arrive at the unbiased CRLB bounds for the model under consideration. The resulting CRLB bounds have been confirmed through Monte Carlo simulations using the MLE estimation method. It has been observed that the estimators achieve the CRLB bound at high SNRs, but that at very low SNRs, the estimators tend to be biased and therefore the unbiased CRLB bound is no longer the valid bound. In this way, future work involves the study of the biased CRLB bound in MICA model.

Chapter 4

Uniform Cramer-Rao Lower Bound

The performance of unbiased estimators can be evaluated by using the classical Cramer-Rao lower bound (CRLB) as discussed in the proceeding chapters. For biased estimators, the unbiased classical CRLB bound is no longer applicable because it does not take into account the estimation bias. Therefore, a new CRLB bound is needed when dealing with biased estimation problems. This chapter presents an overview of a recent uniform CRLB bound (UCRLB) proposed by Hero et al. [33]. This bound is applicable to any estimator whose bias gradient satisfies a user specified length constraint. Normally, the bias gradient needs to be estimated in order to apply the UCRLB bound. An algorithm for estimating the bias gradient was derived in the paper by Hero et al., but there is an oversight in their derivation of the estimation variance as shown in Eqn. (24) of [33]. In this chapter, the variance will be re-derived to correct this error and the resulting formula has been acknowledged by the author [34, 35]. It is then shown how to use the UCRLB bound to trace a curve separating the regions of achievable and unachievable performance in the bias-variance trade-off plane. The results will be used for the remainder of this thesis.

This chapter is organized as follows. Section 1 gives an introduction to the performance evaluation for biased estimation problems. Section 2 discusses the conventional biased CRLB bound and its limitations. Section 3 presents an overview of the UCRLB bound. Section 4 introduces a modified version of the bias gradient estimation algorithm to estimate the bias gradient length. This is followed by the modified formula for the variance of this estimation in Section 5. The concluding remarks of this chapter are given in Section 6.

4.1 Introduction

As in the previous chapters, let us consider the problem of the estimation of an p-dimensional parameter vector

$$\underline{\theta} = [\theta_1, \theta_2, \cdots, \theta_p]^T \in \mathbf{\Theta} \tag{4.1}$$

given an observation vector of random variables, say $\underline{Y} \in \mathbf{R}^N$, having a probability density function (PDF) $f_{\underline{Y}}(\underline{y}; \underline{\theta})$. The parameter space $\mathbf{\Theta}$ is assumed to be an open subset of p-dimensional Euclidean space $\mathbf{R}^p$. When an estimator $\hat{\underline{\theta}}$ is biased, the mean-square error (MSE) of the estimator is an important measure of the estimation performance. As pointed out in Chapter 1, the MSE error is a function of both the bias $b_{\underline{\theta}}(\hat{\underline{\theta}})$ and the variance $\sigma^2_{\underline{\theta}}$ of the estimator in according with

$$MSE_\theta(\hat{\underline{\theta}}) = \sigma^2_{\underline{\theta}}(\hat{\underline{\theta}}) + b^2_{\underline{\theta}}(\hat{\underline{\theta}}). \tag{4.2}$$

Recall that the MSE error for unbiased estimator is only dependent on the variance of estimation. However, if the estimation bias is not equal to zero, i.e. in biased estimation, one can observe from Eqn. (4.2) that any increase in MSE error can be due to an increase in the bias or variance of $\hat{\underline{\theta}}$. While bias is due to "mismatch" between the average value of the estimation and its true value, variance is due to statistical fluctuations in the estimation. Usually, there exists a tradeoff between the bias and variance of the estimated parameter. One can reduce the variance only at the expense of increasing the bias or vice versa, reduce the bias only at the expense of increasing the variance. In the case of biased estimation algorithms, biased CRLB [36] is available to place a bound on the total variance of any estimator with a given bias, but it is strongly estimator dependent. That is, it will usually change when the estimator changes due to changes in the bias.

Hero et al. [33] developed a lower bound, called uniform CRLB (UCRLB), on the estimation variance for any estimator whose bias gradient satisfies a user-specified length constraint. In this way, the resulting UCRLB bound is independent of the estimator under consideration. By means of UCRLB bound, one can explore the fundamental trade-offs between the bias and variance in biased estimation problems. Different estimators can effectively be compared by tracing out their performance on a bias-variance trade-off plane. This new UCRLB bound also provides a means for specifying achievable and unachievable regions in the bias-variance trade-off plane. In order to apply UCRLB bound in performance comparisons, the bias gradient has to be determined. In practice, the bias gradient is often not available analytically and must also be estimated. An algorithm for estimating bias gradient using computer simulations is presented in [33]. This algorithm only considers the problem of using a single data for estimating the parameter under consideration.

In this chapter, a modified relationship is developed for estimating the bias gradient, mainly because the algorithm in the original paper by Hero et al. cannot be directly applied to many biased estimation problems, e.g. the SNR estimation problems to be discussed in next chapter. For this modified method, multiple data points are used to estimate a parameter instead of using a single data point. In conjunction with the resulting method, a new relationship formula is also developed for the determination of the variance of this bias gradient estimation.

4.2 Biased CRLB bound

Let us consider a non-random but unknown parameter vector $\underline{\theta}$ and assume that the observable random variable vector $\underline{Y}$ has a PDF of $f_{\underline{Y}}\left(\underline{y};\underline{\theta}\right)$ conditioned upon $\underline{\theta}$. Let $\hat{t} = \hat{t}\left(\underline{Y}\right)$ be an estimation of the scalar $t_{\underline{\theta}}$, where $t : \mathbf{\Theta} \rightarrow \mathbf{R}$ is a specified function. Let this estimator have mean of $m_{\underline{\theta}} = E_{\underline{\theta}}\left[\hat{t}\right]$. Then, the estimation bias is defined as

$$b_{\underline{\theta}} = m_{\underline{\theta}} - t_{\underline{\theta}}, \tag{4.3}$$

and the estimation variance is defined as

$$\sigma_{\underline{\theta}}^2 = E_{\underline{\theta}}\left[\left(\hat{t} - m_{\underline{\theta}}\right)^2\right]. \tag{4.4}$$

The so called biased CRLB bound is given by [9]

$$\begin{aligned} \sigma_{\underline{\theta}}^2 &\geq \nabla m_{\theta}^T F_{\underline{Y}}^+ \left(\underline{\theta}\right) \nabla m_{\theta} && (4.5) \\ &= (\nabla t_{\underline{\theta}} + \nabla b_{\underline{\theta}})^T F_{\underline{Y}}^+ \left(\underline{\theta}\right) (\nabla t_{\underline{\theta}} + \nabla b_{\underline{\theta}}), && (4.6) \end{aligned}$$

where $F_{\underline{Y}}^+ \left(\underline{\theta}\right)$ is the Moore-Penrose pseudoinverse of the $N \times N$ Fisher information matrix (FIM) defined as

$$F_{\underline{Y}}\left(\underline{\theta}\right) = E_{\theta}\left\{\left[\nabla \ln f_{\underline{Y}}\left(\underline{y};\underline{\theta}\right)\right]\left[\nabla \ln f_{\underline{Y}}\left(\underline{y};\underline{\theta}\right)\right]^T\right\}, \tag{4.7}$$

and where ∇ represents the gradient operator $\nabla = \left[\frac{\partial}{\partial \theta_1}, \cdots, \frac{\partial}{\partial \theta_p}\right]^T$.

Eqn. (4.6) is the classical biased CRLB bound which can be seen to be estimator-dependent due to the fixed bias gradient ∇b_{θ} of the estimator. This means that it can only be applied to the class of estimators which have a particular bias gradient function of ∇b_{θ}. For example, it cannot be used to simultaneously place a bound for two estimators which have different but perhaps acceptable biases. Therefore, this biased CRLB bound only applies to a limited class of practical estimators, since seldom do two different estimators have the same bias gradient. A more general class includes the set of estimators whose bias gradient length is smaller than a specified threshold.

For unbiased estimators, $b_{\theta} = 0$, and if the FIM is full-rank, the unbiased CRLB bound satisfies the relationship

$$\sigma_{\underline{\theta}}^2 \geq \nabla^T t_{\underline{\theta}} F_{\underline{Y}}^{-1}\left(\underline{\theta}\right) \nabla t_{\underline{\theta}}. \tag{4.8}$$

The right hand side (RHS) of Eqn. (4.8) is only dependent on the FIM and is hence estimator-independent. The classical unbiased CRLB bound has been previously applied to compare different unbiased estimators.

4.3 Uniform CRLB Bound

In biased estimation, the general form of the CRLB bound in Eqn. (4.6) is not estimator independent as pointed out before, but is a function of the gradient of the estimator bias $\nabla b_{\underline{\theta}} = [\partial b_{\underline{\theta}}/\partial\theta_1, \cdots, \partial b_{\underline{\theta}}/\partial\theta_p]^T$. This means that the classical biased CRLB bound in Eqn. (4.5) is only applicable to estimators with fixed bias gradient. The bias gradient vector is simply a measure of the sensitivity or coupling of a particular estimator's bias function with respect to variations in the parameter $\underline{\theta}$. The kth component of this bias gradient vector specifies the influence of the kth parameter θ_k. A large value indicates that the estimator is sensitive to changes in θ_k. The length $\|\nabla b_{\underline{\theta}}\|$ of the bias gradient is a measure of the overall non-removable estimator bias. Since many different bias gradient vectors can correspond to the same overall bias (as measured by the bias gradient length), the restriction of a lower bound to a fixed bias gradient is not justified. Therefore, the classical biased CRLB bound is not capable of providing a meaningful comparison of different biased estimators that have acceptable bias but different bias gradients.

Hero et al. [33] presented a uniform CRLB bound on the variance of biased estimators as shown in Theorem 4.1 for a simplified version.

Theorem 4.1 *Let the PDF* $f_{\underline{Y}}(\underline{y};\underline{\theta})$ *have an associated FIM matrix* $F_{\underline{Y}} = F_{\underline{Y}}(\underline{\theta})$ *as in Eqn. (4.7), and let* $\hat{t}$ *be an estimator of the scalar differentiable function* $t_{\underline{\theta}}$ *of the parameter* $\underline{\theta} = [\theta_1, \theta_2, \cdots, \theta_p]^T$. *For a fixed* $\delta > 0$, *let the bias gradient of* $\hat{t}$ *satisfy the norm constraint* $\|\nabla b_{\underline{\theta}}(\hat{t})\| \le \delta$, *where* $\|\underline{z}\|^2 = \sum_{i=1}^{N} z_i^2$ *for an* N*-element real vector* $\underline{z}$. *Then the variance of* $\hat{t}$ *satisfies*

$$var_{\underline{\theta}}(\hat{t}) \ge B(\underline{\theta}, \delta) \tag{4.9}$$

$$= [\nabla t_{\underline{\theta}} + \underline{d}_{\min}]^T F_{\underline{Y}}^{-1} [\nabla t_{\underline{\theta}} + \underline{d}_{\min}] \tag{4.10}$$

$$= \lambda^2 \nabla t_{\underline{\theta}} [I + \lambda F_{\underline{Y}}]^{-1} F_{\underline{Y}} [I + \lambda F_{\underline{Y}}]^{-1} \nabla t_{\underline{\theta}}, \tag{4.11}$$

where in Eqn. (4.10)

$$\underline{d}_{\min} = -\left[I + \lambda F_{\underline{Y}}\right]^{-1} \nabla t_{\underline{\theta}}, \tag{4.12}$$

and where λ is given by the unique non-negative solution of the following equation involving the monotone decreasing, strictly convex function $g(\lambda) \in \left[0, \left\|\nabla \beta_{\underline{\theta}}\right\|^2\right]$ given by

$$g(\lambda) = \underline{d}_{\min}^T \underline{d}_{\min} = \nabla t_{\underline{\theta}} \left[I + \lambda F_{\underline{Y}}\right]^{-2} \nabla t_{\underline{\theta}} = \delta^2. \tag{4.13}$$

For each δ, $B(\underline{\theta}, \delta)$ provides a uniform lower bound on the variance $var_{\underline{\theta}}(\hat{t})$ for a large class of estimators. A more general version of the UCRLB bound is given in [33] and applicable to a singular $F_{\underline{Y}}$.

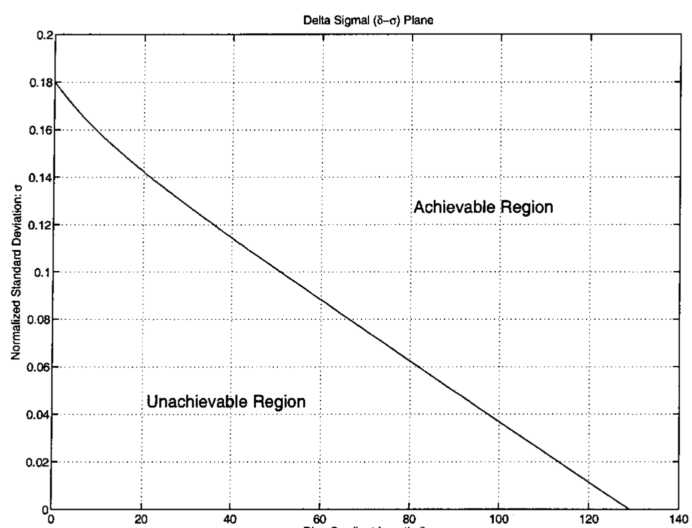

Figure 4.1: The Normalized UCRLB bound on the $\delta - \sigma$ trade-off plane.

The plot of the lower bound versus δ gives rise to a curve $(\delta, B(\underline{\theta}, \delta))$ which indicates the best estimation performance of the estimators over a wide range of bias gradient vector length. The curve divides the bias gradient vector length and variance trade-off plane (or $\delta - \sigma$ plane) into achievable and unachievable regions. Different estimators can be placed in the achievable region of the $\delta - \sigma$ plane and their performance can be effectively compared.

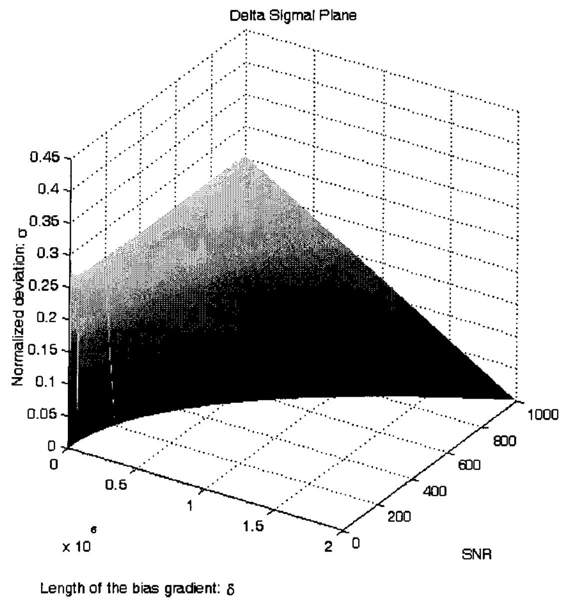

Figure 4.2: The Normalized UCRLB bound on the $\delta - \sigma$ tradeoff plane together with the corresponding SNR values.

The idea behind the UCRLB bound is that the length or norm of the bias gradient vector can be viewed as a measure of the total bias error of an estimator. Among all possible estimators with a given bias gradient vector length, there exists an ideal minimum variance estimator whose variance is the lower bound of all possible estimators with the given gradient vector length.

4.4 Estimation of the Bias Gradient

In order to compare the performance of an estimator against the UCRLB bound, one needs to determine both the estimator variance and the bias gradient vector length. From Eqn. (4.3), the bias gradient can be obtained as

$$\nabla b_\theta = \nabla m_\theta - \nabla t_{\underline{\theta}}, \tag{4.14}$$

where $b_\theta \triangleq b_\theta(\hat{\theta})$ for simplified notation.

Unfortunately, in most cases it is not possible to get an analytical relationship for

∇m_θ. Consequently, one must resort to estimating the bias b_θ and bias gradient ∇b_θ also.

A statistically consistent relationship for estimating ∇b_θ by computer simulations was presented in Eqn. (23) of [33] as given by:

$$\widehat{\nabla b_{\underline{\theta}}} = \frac{1}{L-1}\sum_{i=1}^{L}\left[\hat{t}\left(Y_i\right) - \frac{1}{L}\sum_{j=1}^{L}\hat{t}\left(Y_j\right)\right]\nabla \ln f_Y(Y_i;\underline{\theta}) - \nabla t_{\underline{\theta}}, \tag{4.15}$$

where $\hat{t}\left(Y_i\right)$ represents the estimation $\hat{t}(Y)$ of $t(Y)$ based on the single measurement Y_i taken from the set of L measurements $\{Y_i \mid i = 1, 2, \cdots, L\}$.

Unfortunately, Eqn. (4.15) cannot be applied to some estimators. For example, let us consider the SNR estimator defined as

$$\hat{\beta} = \hat{t}\left(Y\right) = \frac{\hat{A}^2}{2\hat{\sigma}^2}, \tag{4.16}$$

where for the moment-based estimation algorithm, then one has

$$\hat{A}^2 = \left[2\left(\overline{Y^2}\right)^2 - \overline{Y^4}\right]^{1/2}, \tag{4.17a}$$

$$\hat{\sigma}^2 = \frac{1}{2}\left[\overline{Y^2} - \hat{A}^2\right]. \tag{4.17b}$$

The above estimator will be discussed in detail in the next chapter. However, for the time being, one can observe that Eqn. (4.15) (which assumes estimation based on one data point only) cannot be applied in this situation. This is because for $N = 1$, one has $\overline{Y^2} = Y_1^2$, and $\overline{Y^4} = Y_1^4$, so $\hat{A}^2 = Y_1^2$ but $\hat{\sigma}^2 = 0$. Thus, $\hat{\beta}$ does not exist for $N = 1$. In this case, we propose a modified estimation for the bias gradient as

$$\widehat{\nabla b_{\underline{\theta}}} = \frac{1}{L-1}\sum_{i=1}^{L}\left[\hat{t}\left(Y^{(i)}\right) - \frac{1}{L}\sum_{j=1}^{L}\hat{t}\left(Y^{(j)}\right)\right]\nabla \ln f_Y(Y^{(i)};\underline{\theta}) - \nabla t_{\underline{\theta}}, \tag{4.18}$$

where $Y^{(i)} = [Y_1^{(i)} \cdots Y_N^{(i)}]^T$ is the ith set of N statistically independent measured random variables, and $f_Y(Y^{(i)};\underline{\theta})$ is their joint PDF.

Eqn. (4.18) is a more general formula for estimating the bias gradient. When $N = 1$, it is the same as the Eqn. (24) in [33]. This modified estimation algorithm will be used in the next chapter for the discussion of SNR estimators.

4.5 Derivation of Variance of Bias Gradient Estimation

In [33], the variance of the estimation of bias gradient is given by

$$\begin{aligned} S\left(\widehat{\nabla b_{\underline{\theta}}}\right) &= \frac{1}{L} cov_{\underline{\theta}} \left[\left(\hat{t}\left(Y_i\right) - m_{\underline{\theta}}\right) \nabla \ln f_Y \left(Y_i; \underline{\theta}\right) \right] \\ &+ \frac{1}{L\left(L-1\right)} \left[var_{\underline{\theta}} \left(\hat{t}\left(Y_i\right)\right) \mathbf{F}_{Y_i} + \nabla b_{\underline{\theta}} \nabla^T b_{\underline{\theta}} \right], \end{aligned} \tag{4.19}$$

where $\mathbf{F}_{Y_i} = E_{\underline{\theta}} \left\{ \left[\nabla \ln f_Y \left(Y_i; \underline{\theta}\right)\right] \left[\nabla \ln f_Y \left(Y_i; \underline{\theta}\right)\right]^T \right\}$. Unfortunately, the last term in Eqn. (4.19) is incorrect. A correct variance relationship will be derived as follows.

From Eqn. (4.15), one can write the covariance matrix as

$$\begin{aligned} S\left(\widehat{\nabla b_{\underline{\theta}}}\right) &= \frac{1}{L^2} cov_{\underline{\theta}} \left[\sum_{i=1}^{L} \left(\hat{t}\left(Y_i\right) - \frac{1}{L-1} \sum_{\substack{j=1 \\ j \neq i}}^{L} \hat{t}\left(Y_j\right) \right) \nabla \ln f_Y \left(Y_i; \underline{\theta}\right) \right] \\ &= \frac{1}{L^2} cov_{\underline{\theta}} \left[\sum_{i=1}^{L} \left(\hat{t}\left(Y_i\right) - m_{\underline{\theta}} + m_{\underline{\theta}} - \frac{1}{L-1} \sum_{\substack{j=1 \\ j \neq i}}^{L} \hat{t}\left(Y_j\right) \right) \nabla \ln f_Y \left(Y_i; \underline{\theta}\right) \right] \\ &= \frac{1}{L^2} cov_{\underline{\theta}} \left[\sum_{i=1}^{L} \left(\hat{t}\left(Y_i\right) - m_{\underline{\theta}}\right) \nabla \ln f_Y \left(Y_i; \underline{\theta}\right) \right. \\ &\quad \left. + \sum_{i=1}^{L} \left(m_{\underline{\theta}} - \frac{1}{L-1} \sum_{\substack{j=1 \\ j \neq i}}^{L} \hat{t}\left(Y_j\right) \right) \nabla \ln f_Y \left(Y_i; \underline{\theta}\right) \right] \\ &= \frac{1}{L^2} cov_{\underline{\theta}} \left[\underbrace{\sum_{i=1}^{L} \left(\hat{t}\left(Y_i\right) - m_{\underline{\theta}}\right) \nabla \ln f_Y \left(Y_i; \underline{\theta}\right)}_{U} \right. \\ &\quad \left. \underbrace{+ \frac{1}{L-1} \sum_{i=1}^{L} \sum_{\substack{j=1 \\ j \neq i}}^{L} \left(m_{\underline{\theta}} - \hat{t}\left(Y_j\right)\right) \nabla \ln f_Y \left(Y_i; \underline{\theta}\right)}_{V} \right]. \end{aligned} \tag{4.20}$$

In order to simplify Eqn. (4.20), the following two theorems will be used.

Theorem 4.2 Suppose $U, V \in \mathbf{R}^m$ are random vectors conditional on $\underline{\theta} \in \mathbf{R}^m$ with

$m_U = E_{\underline{\theta}}[U]$, *and* $m_V = E_{\underline{\theta}}[V]$, *so that*

$$cov_{\underline{\theta}}(U+V) = E_{\underline{\theta}}\left[(U+V)(U+V)^T - (m_U+m_V)(m_U+m_V)^T\right]. \tag{4.21}$$

Suppose $m_V = 0$, *but* $m_U \neq 0$. *Then,*

$$\begin{aligned} cov_{\underline{\theta}}(U+V) &= E_{\underline{\theta}}\left[UU^T + VV^T + UV^T + VU^T - m_U m_U^T\right] &(4.22)\\ &= E_{\underline{\theta}}\left[UU^T - m_U m_U^T\right] + E_{\underline{\theta}}\left[VV^T\right] + \underbrace{E_{\underline{\theta}}[UV^T] + E_{\underline{\theta}}\left[VU^T\right]}_{\triangleq cov_{\underline{\theta}}(U,V)} &(4.23)\\ &= cov_{\underline{\theta}}(U) + cov_{\underline{\theta}}(V) + cov_{\underline{\theta}}(U,V). &(4.24) \end{aligned}$$

Theorem 4.3 For any random variable Z *with PDF of* $f_z(Z;\underline{\theta})$, *there exists the following identity*

$$E_{\underline{\theta}}\left[Z\frac{\partial \ln f_z(Z;\underline{\theta})}{\partial \theta_i}\right] = \frac{\partial}{\partial \theta_i}E_{\underline{\theta}}[Z]. \tag{4.25}$$

Proof:

$$\begin{aligned} E_{\underline{\theta}}\left[Z\frac{\partial \ln f_z(Z;\underline{\theta})}{\partial \theta_i}\right] &= \int_Z z\frac{\partial \ln f_z(z;\underline{\theta})}{\partial \theta_i} f_z(z;\underline{\theta})\,dz \\ &= \int_Z z\frac{\partial f_z(z;\underline{\theta})}{\partial \theta_i}dz \\ &= \frac{\partial}{\partial \theta_i}\int_Z z f_z(z;\underline{\theta})\,dz \\ &= \frac{\partial}{\partial \theta_i}E_{\underline{\theta}}[Z]. \end{aligned} \tag{4.26}$$

From the above Theorem 4.2, Eqn. (4.20) can be rewritten as:

$$S\left(\widehat{\nabla b_{\underline{\theta}}}\right) = \frac{1}{L^2}cov_{\underline{\theta}}[U+V]. \tag{4.27}$$

It can be shown that

$$\begin{aligned} m_U &= E_{\underline{\theta}}\left[\sum_{i=1}^{L}\left(\hat{t}(Y_i) - m_{\underline{\theta}}\right)\nabla \ln f_Y(Y_i;\underline{\theta})\right] &(4.28)\\ &= \sum_{i=1}^{L} E_{\underline{\theta}}\left[\hat{t}(Y_i)\nabla \ln f_Y(Y_i;\underline{\theta})\right] \neq 0, &(4.29) \end{aligned}$$

and

$$m_V = E_{\underline{\theta}}\left[\frac{1}{L-1}\sum_{i=1}^{L}\sum_{\substack{j=1\\ j\neq i}}^{L}\left(m_{\underline{\theta}} - \hat{t}\left(Y_j\right)\right)\nabla \ln f_Y\left(Y_i;\underline{\theta}\right)\right] \tag{4.30}$$

$$= \frac{1}{L-1}\sum_{i=1}^{L}\sum_{\substack{j=1\\ j\neq i}}^{L} E_{\underline{\theta}}\left[\left(m_{\underline{\theta}} - \hat{t}\left(Y_j\right)\right) E_{\underline{\theta}}\left[\nabla \ln f_Y\left(Y_i;\underline{\theta}\right)\right]\right] = 0. \tag{4.31}$$

From Theorem 4.1, one has

$$S\left(\widehat{\nabla b_{\underline{\theta}}}\right) = \frac{1}{L^2}\left[cov_{\underline{\theta}}\left(U\right) + cov_{\underline{\theta}}\left(V\right) + E_{\underline{\theta}}\left(UV^T\right) + E_{\underline{\theta}}\left(VU^T\right)\right]. \tag{4.32}$$

The first term in Eqn. (4.32) is given by

$$cov_{\underline{\theta}}\left(U\right) = cov_{\underline{\theta}}\left[\sum_{i=1}^{L}\left(\hat{t}\left(Y_i\right) - m_{\underline{\theta}}\right)\nabla \ln f_Y\left(Y_i;\underline{\theta}\right)\right] \tag{4.33}$$

$$= L cov_{\underline{\theta}}\left[\left(\hat{t}\left(Y_i\right) - m_{\underline{\theta}}\right)\nabla \ln f_Y\left(Y_i;\underline{\theta}\right)\right], \tag{4.34}$$

and the second term is given by

$$cov_{\underline{\theta}}\left(V\right) = cov_{\underline{\theta}}\left[\frac{1}{L-1}\sum_{i=1}^{L}\sum_{\substack{j=1\\ j\neq i}}^{L}\left(m_{\underline{\theta}} - \hat{t}\left(Y_j\right)\right)\nabla \ln f_Y\left(Y_i;\underline{\theta}\right)\right]$$

$$= \frac{1}{\left(L-1\right)^2} cov_{\underline{\theta}}\left[\sum_{i=1}^{L}\sum_{\substack{j=1\\ j\neq i}}^{L}\left(m_{\underline{\theta}} - \hat{t}\left(Y_j\right)\right)\nabla \ln f_Y\left(Y_i;\underline{\theta}\right)\right]$$

$$= \frac{1}{\left(L-1\right)^2} E_{\underline{\theta}}\left[\sum_{i=1}^{L}\sum_{\substack{j=1\\ j\neq i}}^{L}\sum_{k=1}^{L}\sum_{\substack{l=1\\ l\neq k}}^{L}\left(m_{\underline{\theta}} - \hat{t}\left(Y_j\right)\right)\nabla \ln f_Y\left(Y_i;\underline{\theta}\right)\cdot\right.$$

$$\left.\cdot\left[\nabla \ln f_Y\left(Y_k;\underline{\theta}\right)\right]^T\left(m_{\underline{\theta}} - \hat{t}\left(Y_l\right)\right)\right]. \tag{4.35}$$

One can further simplify Eqn. (4.35) as follows:

$$
\begin{aligned}
&cov_{\underline{\theta}}(V) \\
&= \frac{1}{(L-1)^2}\sum_{i=1}^{L}\sum_{\substack{j=1\\ j\neq i}}^{L}\sum_{\substack{l=1\\ l\neq k}}^{L} E_{\underline{\theta}}\left\{\left(m_{\underline{\theta}}-\hat{t}(Y_j)\right)\left(m_{\underline{\theta}}-\hat{t}(Y_l)\right)\left[\nabla \ln f_Y(Y_i;\underline{\theta})\right]\cdot\right. \\
&\quad \left.\cdot\left[\nabla \ln f_Y(Y_i;\underline{\theta})\right]^T\right\} + \sum_{i=1}^{L}\sum_{\substack{j=1\\ j\neq i}}^{L}\sum_{\substack{k=1\\ k\neq i}}^{L}\sum_{\substack{l=1\\ l\neq k}}^{L} E_{\underline{\theta}}\left\{\left[m_{\underline{\theta}}-\hat{t}(Y_j)\right]\left[m_{\underline{\theta}}-\hat{t}(Y_l)\right]\cdot\right. \\
&\quad \left.\cdot\left[\nabla \ln f_Y(Y_i;\underline{\theta})\right]\left[\nabla \ln f_Y(Y_k;\underline{\theta})\right]^T\right\} \\
&= \frac{1}{(L-1)^2}\left\{\sum_{i=1}^{L}\sum_{\substack{j=1\\ j\neq i}}^{L} E_{\underline{\theta}}\left[\left(m_{\underline{\theta}}-\hat{t}(Y_j)\right)^2\left[\nabla \ln f_Y(Y_i;\underline{\theta})\right]\left[\nabla \ln f_Y(Y_i;\underline{\theta})\right]^T\right]+\right. \\
&\quad \left.+\sum_{i=1}^{L}\sum_{\substack{j=1\\ j\neq i}}^{L}\sum_{\substack{l=1\\ l\neq j}}^{L} E_{\underline{\theta}}\left[\left(m_{\underline{\theta}}-\hat{t}(Y_j)\right)\left(m_{\underline{\theta}}-\hat{t}(Y_l)\right)\left[\nabla \ln f_Y(Y_i;\underline{\theta})\right]\left[\nabla \ln f_Y(Y_j;\underline{\theta})\right]^T\right]\right\} \\
&= \frac{1}{(L-1)^2}\sum_{i=1}^{L}\sum_{\substack{j=1\\ j\neq i}}^{L} E_{\underline{\theta}}\left\{\left[\left(m_{\underline{\theta}}-\hat{t}(Y_j)\right)^2\right] E_{\underline{\theta}}\left[\left[\nabla \ln f_Y(Y_i;\underline{\theta})\right]\left[\nabla \ln f_Y(Y_i;\underline{\theta})\right]^T\right]+\right. \\
&\quad \left.\sum_{i=1}^{L}\sum_{\substack{j=1\\ j\neq i}}^{L} E_{\underline{\theta}}\left[\left(m_{\underline{\theta}}-\hat{t}(Y_i)\right)\left[\nabla \ln f_Y(Y_i;\underline{\theta})\right]\left(m_{\underline{\theta}}-\hat{t}(Y_j)\right)\left[\nabla \ln f_Y(Y_j;\underline{\theta})\right]^T\right]\right\} \\
&= \frac{L(L-1)}{(L-1)^2}\left\{var_{\underline{\theta}}\left(\hat{t}(Y_i)\right)\mathbf{F}_{Y_i} + E_{\underline{\theta}}\left[\left(m_{\underline{\theta}}-\hat{t}(Y_i)\right)\nabla \ln f_Y(Y_i;\underline{\theta})\right]\cdot\right. \\
&\quad \left.\cdot E_{\underline{\theta}}\left[\left(m_{\underline{\theta}}-\hat{t}(Y_j)\right)\left[\nabla \ln f_Y(Y_j;\underline{\theta})\right]^T\right]\right\} \\
&= \frac{L}{L-1}\left\{var_{\underline{\theta}}\left(\hat{t}(Y_i)\right)\mathbf{F}_{Y_i} + E_{\underline{\theta}}\left[\hat{t}(Y_i)\nabla \ln f_Y(Y_i;\underline{\theta})\right] E_{\underline{\theta}}\left[\left[\hat{t}(Y_j)\nabla \ln f_Y(Y_j;\underline{\theta})\right]^T\right]\right\}
\end{aligned}
\tag{4.36}
$$

The above derivation is based on the assumption that the Y_i are independently identical distributed (i.i.d.) random variables, and $E_{\underline{\theta}}\left[\nabla \ln f_Y(Y_i;\underline{\theta})\right] = 0$.[1]

From Eqn. (4.25), one can obtain

$$\nabla m_{\underline{\theta}} = E_{\underline{\theta}}\left[\hat{t}(Y_i)\nabla \ln f_Y(Y_i;\underline{\theta})\right], \tag{4.37}$$

[1] This is due to Theorem 4.3 and the fact that $\int f_Y(Y_i;\underline{\theta})\,dY_i = 1$.

therefore,

$$cov_{\underline{\theta}}(V) = \frac{L}{L-1}\left[var_{\underline{\theta}}\left(\hat{t}(Y_i)\right)\mathbf{F}_{Y_i} + \nabla m_{\underline{\theta}} \nabla^T m_{\underline{\theta}}\right]. \tag{4.38}$$

Also the third term in Eqn. (4.32) can be obtained as

$$\begin{aligned}
E_{\underline{\theta}}\left(UV^T\right) &= E\left\{_{\underline{\theta}}\left[\sum_{i=1}^{L}\left(\hat{t}(Y_i) - m_{\underline{\theta}}\right)\nabla \ln f_Y(Y_i;\underline{\theta})\right]\cdot\right.\\
&\quad \left.\cdot\left[\frac{1}{L-1}\sum_{k=1}^{L}\sum_{\substack{l=1\\l\neq k}}^{L}\left(m_{\underline{\theta}} - \hat{t}(Y_l)\right)\nabla \ln f_Y(Y_k;\underline{\theta})\right]^T\right\}\\
&= \frac{1}{L-1}\left\{\sum_{i=1}^{L}\sum_{k=1}^{L}\sum_{\substack{l=1\\l\neq k}}^{L} E_{\underline{\theta}}\left[\hat{t}(Y_i) - m_{\underline{\theta}}\right]\cdot\right.\\
&\quad \left.\cdot\nabla \ln f_Y(Y_i;\underline{\theta})\left[\nabla \ln f_Y(Y_k;\underline{\theta})\right]^T\left[m_{\underline{\theta}} - \hat{t}(Y_l)\right]\right\}\\
&= \frac{1}{L-1}\left\{\sum_{i=1}^{L}\sum_{\substack{l=1\\l\neq i}}^{L} E_{\underline{\theta}}\left[m_{\underline{\theta}} - \hat{t}(Y_l)\right]\left[\hat{t}(Y_i) - m_{\underline{\theta}}\right]\cdot\right.\\
&\quad \left.\cdot\nabla \ln f_Y(Y_i;\underline{\theta})\left[\nabla \ln f_Y(Y_i;\underline{\theta})\right]^T\right\}\\
&= 0.
\end{aligned} \tag{4.39}$$

Similarly, the fourth term of Eqn. (4.32) is given by

$$\begin{aligned}
E_{\underline{\theta}}\left(VU^T\right) &= E_{\underline{\theta}}\left\{\left[\frac{1}{L-1}\sum_{i=1}^{L}\sum_{\substack{j=1\\j\neq i}}^{L}\left(m_{\underline{\theta}} - \hat{t}(Y_j)\right)\nabla \ln f_Y(Y_i;\underline{\theta})\right]\cdot\right.\\
&\quad \left.\cdot\left[\sum_{k=1}^{L}\left(\hat{t}(Y_k) - m_{\underline{\theta}}\right)\nabla \ln f_Y(Y_k;\underline{\theta})\right]^T\right\}\\
&= \frac{1}{L-1}\sum_{i=1}^{L}\sum_{\substack{j=1\\j\neq i}}^{L}\sum_{k=1}^{L} E_{\underline{\theta}}\left\{\left[m_{\underline{\theta}} - \hat{t}(Y_j)\right]\cdot\right.\\
&\quad \left.\cdot\nabla \ln f_Y(Y_i;\underline{\theta})\left[\nabla \ln f_Y(Y_k;\underline{\theta})\right]^T\left[\hat{t}(Y_k) - m_{\underline{\theta}}\right]\right\}\\
&= \frac{1}{L-1}\sum_{i=1}^{L}\sum_{\substack{j=1\\j\neq i}}^{L} E_{\underline{\theta}}\left\{\left[m_{\underline{\theta}} - \hat{t}(Y_j)\right]\left[\hat{t}(Y_i) - m_{\underline{\theta}}\right]\cdot\right.
\end{aligned}$$

$$\cdot \nabla \ln f_Y (Y_i; \underline{\theta}) [\nabla \ln f_Y (Y_i; \underline{\theta})]^T \Big\}$$
$$= \quad 0 \tag{4.40}$$

Finally, one can get

$$\begin{aligned} & S\left(\widehat{\nabla b_{\underline{\theta}}}\right) \\ &= \frac{1}{L^2}\left\{L cov_{\underline{\theta}}\left[\left(\hat{t}\left(Y_i\right) - m_{\underline{\theta}}\right)\nabla \ln f_Y\left(Y_i;\underline{\theta}\right)\right] + \frac{L}{L-1}\left[var_{\underline{\theta}}\left(\hat{t}\left(Y_i\right)\right)\mathbf{F}_{Y_i} + \nabla m_{\underline{\theta}}\nabla^T m_{\underline{\theta}}\right]\right\} \\ &= \frac{1}{L} cov_{\underline{\theta}}\left[\left(\hat{t}\left(Y_i\right) - m_{\underline{\theta}}\right)\nabla \ln f_Y\left(Y_i;\underline{\theta}\right)\right] + \frac{1}{L\left(L-1\right)}\left[var_{\underline{\theta}}\left(\hat{t}\left(Y_i\right)\right)\mathbf{F}_{Y_i} + \nabla m_{\underline{\theta}}\nabla^T m_{\underline{\theta}}\right]. \end{aligned} \tag{4.41}$$

Eqn. (4.41) is the desired covariance of the biased estimation, correcting the error in Eqn. (24) in [33].

The variance of the proposed general bias gradient estimation in Eqn. (4.18) can be obtained in a similar manner as

$$\begin{aligned} S\left(\widehat{\nabla b_{\underline{\theta}}}\right) = & \frac{1}{L} cov_{\underline{\theta}}\left[\left(\hat{t}\left(Y^{(i)}\right) - m_{\underline{\theta}}\right)\nabla \ln f_Y\left(Y^{(i)};\underline{\theta}\right)\right] \\ & + \frac{1}{L\left(L-1\right)}\left[var_{\underline{\theta}}\left(\hat{t}\left(Y^{(i)}\right)\right)\mathbf{F}_{Y^{(i)}} + \nabla m_{\underline{\theta}}\nabla^T m_{\underline{\theta}}\right], \end{aligned} \tag{4.42}$$

where

$$\mathbf{F}_{Y^{(i)}} = E_{\underline{\theta}}\left\{\left[\nabla \ln f_Y\left(Y^{(i)};\underline{\theta}\right)\right]\left[\nabla \ln f_Y\left(Y^{(i)};\underline{\theta}\right)\right]^T\right\}. \tag{4.43}$$

The first term in the RHS of Eqn. (4.42) asymptotically decreases to zero with L as $1/L$ and is independent of the mean $m_{\underline{\theta}}$. The second term in Eqn. (4.41) depends on $m_{\underline{\theta}}$ only through its gradient and asymptotically decreases to zero with L as $1/L^2$. Therefore, this bias gradient estimation is consistent.

The bias gradient estimation in Eqn. (4.18) is also unbiased which can be established as follows:

$$\begin{aligned} E\left[\widehat{\nabla b_{\underline{\theta}}}\right] &= \frac{1}{L-1} E\left\{\sum_{i=1}^{L}\left[\hat{t}\left(Y^{(i)}\right) - \frac{1}{M}\sum_{j=1}^{L}\hat{t}\left(Y^{(j)}\right)\right]\nabla \ln f_Y(Y^{(i)};\underline{\theta}) - \nabla t_{\underline{\theta}}\right\} \\ &= \frac{1}{L-1}\sum_{i=1}^{L} E\left\{\left[\hat{t}\left(Y^{(i)}\right) - \frac{1}{L}\sum_{j=1}^{L}\hat{t}\left(Y^{(j)}\right)\right]\nabla \ln f_Y(Y^{(i)};\underline{\theta})\right\} - \nabla t_{\underline{\theta}} \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{L-1}\sum_{i=1}^{L} E\left[\frac{L-1}{L}\hat{t}\left(Y^{(i)}\right)\nabla \ln f_Y(Y^{(i)};\underline{\theta})\right] - \nabla t_{\underline{\theta}} \\
&= E\left[\hat{t}\left(Y^{(i)}\right)\nabla \ln f_Y(Y^{(i)};\underline{\theta})\right] - \nabla t_{\underline{\theta}} \\
&= \nabla m_{\theta} - \nabla t_{\underline{\theta}} = \nabla b_{\underline{\theta}}.
\end{aligned} \tag{4.44}
$$

The above proof has used Eqn. (4.37). From Eqns. (4.44) and (4.42), the proposed bias gradient estimation in Eqn. (4.18) is an unbiased and consistent estimate for bias gradient $\nabla b_{\underline{\theta}}$.

4.6 Computation of UCRLB and Its Applications

The computation of UCRLB bound from Theorem 4.1 can be a difficult task when the number of unknown parameters is large due to the required large matrix inversions. Often, one needs to compute the bound for several different values of δ, and, equivalently, for several different values of λ. Therefore, it is important to have a computationally efficient algorithm to compute Eqn. (4.12). The delta-sigma tradeoff curve can be computed by sweeping out λ in the following step-by-step procedure.

1) Select $\lambda \in (0, \infty)$.

2) Compute $\underline{d}_{\text{min}} = -\left[I + \lambda F_{\underline{Y}}\right]^{-1} \nabla t_{\underline{\theta}}$.

3) Compute the point $(\delta,\, B(\delta))$ via

$$
B(\delta) = \lambda^2 \underline{d}_{\text{min}}^T F_{\underline{Y}} \underline{d}_{\text{min}}, \tag{4.45}
$$

and

$$
\delta = \sqrt{\underline{d}_{\text{min}}^T \underline{d}_{\text{min}}}. \tag{4.46}
$$

Applying this method for several values of λ allows one to trace the curve $B(\delta)$ in the $\delta - \sigma$ plane, with or without SNR as an axis as shown in Figs. (4.1) and (4.2), respectively. A more general procedure may be found in [33] for the case of a large number of unknown parameters of θ_k, $k = 1, 2, \cdots, p$.

The UCRLB bound provides a useful means for the determinations of lower bound for any biased estimator whose bias gradient length is smaller than a specified value.

Its applications can be readily found in the hitherto literature mainly in image estimation [33, 37, 38]. In [33], the UCRLB is used in for linear Gaussian and non-linear Poisson inverse problems. In [37], two applications were presented to study the bias-variance trade-offs. One is a particular class of roughness penalized maximum-likelihood (PML) in single photon emission computed tomography (SPECT) image reconstruction, and the other one concerned one-dimensional edge localization estimation problem. In [38], two other applications in image processing problems are presented. The first one is the space alternating generalized expectation-maximization (SAGE), and the second one is for the penalized weighted least-squares estimation (WLSE). All these applications show pertinent results in studying the bias-variance trade-offs and performance comparison with the help of UCRLB bounds.

4.7 Concluding Remarks

This chapter has been concerned with a study of UCRLB bound for biased estimation. The UCRLB bound provides an effective way to compare the performance of different biased estimators. Different biased estimation algorithms can be placed in the same variance-bias gradient plane to identify the best estimator. In most cases, computation of the UCRLB bound requires numerical methods. It should be noticed that in general, the bias gradient also needs to be estimated, and a computer simulation method is available for this purpose as discussed in section 4. For some estimation problems, the method cannot be used directly such as in SNR estimation, therefore we have proposed a modified estimation method, also included in Section 4. Section 5 derived the formula for determination of variance of bias gradient estimation, and correct an oversight of the original formula in the paper by Hero et al. [33]. Computational procedure and some applications of UCRLB bound were reviewed in Section 6.

The next chapter presents a new application of UCRLB bound, namely the SNR estimation for the single-tone signals contaminated by noise. The reason for applying

UCRLB bound to the SNR estimation problem is twofold. Firstly, SNR is a very important parameter in signal processing applications and most parameter estimation problems are SNR dependent. Secondly, the SNR estimation is a nonlinear estimation, as it is defined as the ratio of the squared amplitude to the noise variance. Normally, the amplitude and variance must be estimated in order to obtain the SNR value by the nonlinear calculation. Since nonlinear estimation is inherently biased, the classical CRLB bound is not applicable anymore. A detailed discussion will be seen in the next chapter.

Chapter 5

Application of UCRLB to Estimation Techniques

This chapter presents a novel investigation of estimator performance based on the UCRLB bound. This investigation consists of two different applications, namely SNR estimation for a pure tone contaminated by additive white Gaussian noise (AWGN), and a comparison of the frequency and SNR estimators for multiharmonic signal in AWGN noise.

Before the formal introduction of UCRLB bound, the various estimators in the existing literature were compared by applying the classical CRLB bound on the estimator variance. This led to the fundamental problem that for biased estimators, the classical CRLB bound becomes grossly inaccurate at low SNR values, rendering itself an invalid lower bound[1]. In practical situations, many estimators, such as SNR estimators, are inherently biased. In this chapter, a performance comparison for biased estimators is conducted by tracing out the estimator performance on the bias gradient-variance space and comparing it to the surface specified by the UCRLB bound. The best SNR estimator can be identified as the one with the lowest estimator variance. At the mean time, it is the one closest to the UCRLB bound surface, because the UCRLB bound is the lower bound on the variance of any estimators whose bias gradient satisfies a user specified length constraint for the same data model.

[1] As discussed in Chapter 4, the classical CRLB bound is only valid for unbiased estimators.

Section 1 provides an introduction to the signal model for SNR estimation problems. Sections 2 to 4 are concerned with the investigation of three commonly used SNR estimation techniques available in the hitherto literature. These techniques include maximum likelihood estimation (MLE), and methods of lower order moments and higher order moments. The investigations are carried out in terms of estimating the SNR for a pure tone contaminated in AWGN. Section 5 presents the performance comparison of SNR estimators via Monte Carlo simulation and identifies the best SNR estimator out of the three estimators considered by employing the UCRLB bound. Section 6 presents the performance comparison using UCRLB bound for multiharmonic signal estimation problem discussed in Chapter 3. Section 7 will conclude this chapter.

5.1 Signal Model for SNR estimation

The problems associated with SNR estimation are of great importance in many areas of digital signal processing and digital communication systems [39–41]. This is mainly due to the fact that SNR is an important characteristic of these systems. Moreover, the knowledge of SNR is critical in other parameter estimation techniques. As is known, SNR estimators are inherently biased. Therefore, to compare the performance of SNR estimators, the UCRLB bound renders itself as a valid lower bound.

In the following sections, three SNR estimation techniques are discussed and their performance characteristics are compared with the UCRLB bound. The data model used is a single-tone signal contaminated in AWGN noise as given by

$$x(n) = A\cos(\omega_0 Tn + \phi) + w(n), \tag{5.1}$$

where $w(n)$ represents an AWGN noise with zero mean and a variance of σ^2. Moreover, ω_0 represents the frequency, ϕ represents the initial phase, and T represents the system sampling period (assumed to be a known constant).

Let us define the deterministic but unknown parameter vector $\underline{\theta} = [\theta_1,\ \theta_2,\ \theta_3,\ \theta_4]^T$

$= [A,\ \omega_0,\ \phi,\ \sigma^2]^T$ which is to be estimated. The SNR is defined as

$$\beta = \frac{A^2}{2\sigma^2} = \frac{\theta_1^2}{2\theta_4}. \tag{5.2}$$

The CRLB bound for the data model in Eqn. (5.1) is determined in terms of the Fisher Information Matrix (FIM) as given by [4]

$$[F(\underline{\theta})]_{i,j} = -E\left[\frac{\partial^2 \ln p_X(\underline{x};\underline{\theta})}{\partial\theta_i \partial\theta_j}\right] \qquad (i,j = 1,2,3,4), \tag{5.3}$$

where $\underline{x} = [x(1),\ x(2)\ ,\cdots,\ x(N)]^T$ is the vector of N independent and identically distributed (i.i.d.) random variables with a PDF given by

$$p_X(\underline{x};\underline{\theta}) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}\left[x(n) - A\cos(\omega_0 Tn + \phi)\right]^2\right\}. \tag{5.4}$$

The data vector $\underline{x}$ is used by the estimator for the purpose of estimating SNR of β via the estimates of $\underline{\theta}$ in accordance with Eqn. (5.2).

The three SNR estimators discussed in this chapter will use the envelope $\tilde{x}(n)$ of an "analytic signal" $x(n) + jH(x(n))$, given by

$$\tilde{x}(n) = \sqrt{x^2(n) + H^2\left[x(n)\right]}, \tag{5.5}$$

where $H[\cdot]$ represents the Hilbert transform operation (see Chapter 2).

It is known that the PDF of the envelop of a Gaussian signal is Ricean [1]. Suppose that there are N i.i.d. envelope random variables $\tilde{x}(n)$, each having a Ricean PDF. Then, the joint PDF of the N elements is

$$p(\underline{\tilde{x}};\underline{\theta}) = \prod_{n=1}^{N}\left\{\frac{\tilde{x}(n)}{\sigma^2} I_0\left(\frac{A\tilde{x}(n)}{\sigma^2}\right)\exp\left[-\frac{\tilde{x}^2(n) + A^2}{2\sigma^2}\right]\right\}, \tag{5.6}$$

where $\underline{\tilde{x}} = [\tilde{x}(1),\ \tilde{x}(2)\ ,\cdots,\ \tilde{x}(N)]^T$ is the vector of the envelopes of the random variables. The above PDF of Eqn. (5.6) will be used for the following sections in the derivation of the SNR estimators.

5.2 Maximum Likelihood SNR Estimator

The maximum likelihood (ML) method can be used to estimate the squared amplitude $\hat{A}^2$, the frequency $\hat{\omega}_0$, the phase $\hat{\phi}$, and the noise variance $\hat{\sigma}^2$. Subsequently, the SNR can be estimated in accordance with

$$\hat{\beta} = \frac{\hat{A}^2}{2\hat{\sigma}^2}. \tag{5.7}$$

A similar estimation method can be found in [10] where the original data set $\underline{x}$ was used instead of the envelope signal $\underline{\tilde{x}}$.

It is convenient to define the log-likelihood function (LLF) as

$$\begin{aligned} L &= \ln p\left(\underline{\tilde{x}}; \underline{\theta}\right) \\ &= \sum_{n=1}^{N} \left\{ \ln \left[\tilde{x}(n) I_0 \left(\frac{A\tilde{x}(n)}{\sigma^2} \right) \right] - \ln \sigma^2 - \frac{\tilde{x}^2(n) + A^2}{2\sigma^2} \right\}, \end{aligned} \tag{5.8}$$

where $I_k(x)$ denotes the kth order modified Bessel function of the first kind.

In ML method, one needs to solve the following equations for $\hat{A}^2$ and $\hat{\sigma}^2$:

$$\left. \frac{\partial L}{\partial A^2} \right|_{A^2 = \hat{A}^2} = 0, \tag{5.9a}$$

$$\left. \frac{\partial L}{\partial \sigma^2} \right|_{\sigma^2 = \hat{\sigma}^2} = 0. \tag{5.9b}$$

From Eqn. (5.9a), one has

$$\left. \sum_{n=1}^{N} \left\{ \frac{\partial}{\partial A^2} \ln \left[I_0 \left(\frac{A\tilde{x}(n)}{\sigma^2} \right) \right] - \frac{1}{2\sigma^2} \right\} \right|_{A^2 = \hat{A}^2} = 0. \tag{5.10}$$

Equivalently,

$$\sum_{n=1}^{N} \frac{I_1\left(\dfrac{\hat{A}\tilde{x}(n)}{\sigma^2} \right)}{I_0\left(\dfrac{\hat{A}\tilde{x}(n)}{\sigma^2} \right)} \frac{\tilde{x}(n)}{2\sigma^2\sqrt{\hat{A}^2}} - \frac{N}{2\sigma^2} = 0, \tag{5.11}$$

or

$$\frac{1}{N\sqrt{\hat{A}^2}} \sum_{n=1}^{N} x(n) \frac{I_1\left(\dfrac{\hat{A}\tilde{x}(n)}{\sigma^2} \right)}{I_0\left(\dfrac{\hat{A}\tilde{x}(n)}{\sigma^2} \right)} - 1 = 0. \tag{5.12}$$

From Eqn. (5.9b) one has

$$\sum_{n=1}^{N}\left\{\frac{\partial}{\partial\sigma^2}\ln\left[\tilde{x}(n)I_0\left(\frac{\hat{A}\tilde{x}(n)}{\sigma^2}\right)\right]-\frac{1}{\sigma^2}+\frac{\tilde{x}^2(n)+\hat{A}^2}{2\left(\sigma^2\right)^2}\right\}\Bigg|_{\sigma^2=\hat{\sigma}^2}=0, \tag{5.13}$$

which can be simplified as:

$$\sum_{n=1}^{N}\left\{-\tilde{x}(n)\frac{I_1\left(\frac{\hat{A}\tilde{x}(n)}{\hat{\sigma}^2}\right)}{I_0\left(\frac{\hat{A}\tilde{x}(n)}{\hat{\sigma}^2}\right)}\frac{\hat{A}}{\left(\hat{\sigma}^2\right)^2}+\frac{\tilde{x}^2(n)}{2\left(\hat{\sigma}^2\right)^2}\right\}-\frac{N}{\hat{\sigma}^2}+\frac{N\hat{A}^2}{2\left(\hat{\sigma}^2\right)^2}=0. \tag{5.14}$$

From Eqn. (5.12), it can be shown that

$$\sum_{n=1}^{N}\tilde{x}(n)\frac{I_1\left(\frac{\hat{A}\tilde{x}(n)}{\sigma^2}\right)}{I_0\left(\frac{\hat{A}\tilde{x}(n)}{\sigma^2}\right)}=N\sqrt{\hat{A}^2}. \tag{5.15}$$

Substituting Eqn. (5.15) into Eqn. (5.14), one can get

$$-N\sqrt{\hat{A}^2}\frac{\hat{A}}{\left(\hat{\sigma}^2\right)^2}+\sum_{n=1}^{N}\frac{\tilde{x}^2(n)}{2\left(\hat{\sigma}^2\right)^2}-\frac{N}{\hat{\sigma}^2}+\frac{N\hat{A}^2}{2\left(\hat{\sigma}^2\right)^2}=0, \tag{5.16}$$

which can be simplified as

$$-2\hat{A}^2+\frac{1}{N}\sum_{n=1}^{N}\tilde{x}^2(n)+\hat{A}^2-2\hat{\sigma}^2=0. \tag{5.17}$$

Finally, one can obtain

$$\hat{\sigma}^2=\frac{1}{2}\left[\overline{\tilde{x}^2}-\hat{A}^2\right], \tag{5.18}$$

where $\overline{\tilde{x}^2}\triangleq\frac{1}{N}\sum_{n=1}^{N}\tilde{x}^2(n)$ is the average of N squared-sample values of $\tilde{x}(n)$. Eqn. (5.18) is basically the same as Eqn. (15) in [42].

From Eqn. (5.18), in order to get $\hat{\sigma}^2$, one must first find $\hat{A}^2$. The method to estimate A^2 is discussed next. Let

$$W(n)=\frac{\hat{A}\tilde{x}(n)}{\sigma^2}\Bigg|_{\sigma^2=\hat{\sigma}^2}=\frac{2\tilde{x}(n)\sqrt{\hat{A}^2}}{\overline{\tilde{x}^2}-\hat{A}^2}. \tag{5.19}$$

Then, Eqn. (5.12) can be rewritten as

$$\frac{1}{N\sqrt{\hat{A}^2}}\sum_{n=1}^{N}\frac{\tilde{x}(n)I_1\left[W(n)\right]}{I_0\left[W(n)\right]}-1=0. \tag{5.20}$$

In general, one must use a numerical method in order to solve Eqn. (5.20) for $\hat{A}^2$. In [43], a method was presented that converted the problem to solving a differential equation by means of the classical Runge-Kutta integration method [44] as follows.

Let $\hat{A}_1^2$ be the first estimate of the solution of Eqn. (5.20),

$$f\left(\hat{A}^2\right)=\frac{1}{N\sqrt{\hat{A}^2}}\sum_{n=1}^{N}\frac{\tilde{x}(n)I_1\left[W(n)\right]}{I_0\left[W(n)\right]}-1. \tag{5.21}$$

One can get $\hat{A}^2$ by finding the root of Eqn. (5.21) as

$$\hat{A}^2=\hat{A}_1^2+\int_{f(\hat{A}_1^2)}^{\infty}\frac{d\hat{A}^2}{df}df. \tag{5.22}$$

The classical Runge-Kutta integration method can then be directly applied to solving Eqn. (5.22). Let

$$h_1=-f\left(\hat{A}_1^2\right) \tag{5.23}$$

and

$$F\left(\hat{A}^2\right)=\left[\frac{df}{d\hat{A}^2}\right]^{-1}. \tag{5.24}$$

If a fourth-order integration is used, the second approximation $\hat{A}^2$ is given by

$$\hat{A}_2^2=\hat{A}_1^2+\frac{1}{6}\left(k_1+2k_2+2k_3+k_4\right), \tag{5.25}$$

where

$$k_1 = h_1F\left(\hat{A}_1^2\right), \tag{5.26}$$

$$k_2 = h_1F\left(\hat{A}_1^2+\frac{1}{2}k_1\right), \tag{5.27}$$

$$k_3 = h_1F\left(\hat{A}_1^2+\frac{1}{2}k_2\right), \tag{5.28}$$

$$k_4 = h_1F\left(\hat{A}_1^2+k_3\right). \tag{5.29}$$

From Eqn. (5.21) one can calculate the derivative

$$\frac{df\left(\hat{A}^2\right)}{d\hat{A}^2} = \frac{-\left(\hat{A}^2\right)^{-3/2}}{2N}\sum_{n=1}^{N}\frac{\tilde{x}(n)I_1\left[W(n)\right]}{I_0\left[W(n)\right]} + \frac{1}{N\sqrt{\hat{A}^2}}\sum_{n=1}^{N}\tilde{x}(n)\frac{I_1'\left[W(n)\right]I_0\left[W(n)\right]-I_1^2\left[W(n)\right]}{I_0^2\left[W(n)\right]}\frac{d\left[W(n)\right]}{d\hat{A}^2}. \quad (5.30)$$

But, from [45]

$$I_k'(x) = I_{k-1}(x) - \frac{k}{x}I_k(x). \quad (5.31)$$

Therefore, Eqn. (5.30) can be rewritten as

$$\frac{df\left(\hat{A}^2\right)}{d\hat{A}^2} = \frac{-\left(\hat{A}^2\right)^{-3/2}}{2N}\sum_{n=1}^{N}\frac{\tilde{x}(n)I_1\left[W(n)\right]}{I_0\left[W(n)\right]} + \frac{1}{N\sqrt{\hat{A}^2}}\sum_{n=1}^{N}\tilde{x}(n)\left[1-\frac{I_1\left[W(n)\right]}{W(n)I_0\left[W(n)\right]}-\frac{I_1^2\left[W(n)\right]}{I_0^2\left[W(n)\right]}\right]\frac{d\left[W(n)\right]}{d\hat{A}^2}, \quad (5.32)$$

where

$$\frac{d\left[W(n)\right]}{d\hat{A}^2} = \frac{d}{d\hat{A}^2}\left[\frac{2\tilde{x}(n)\sqrt{\hat{A}^2}}{\overline{\tilde{x}^2}-\hat{A}^2}\right] \quad (5.33)$$

$$= \tilde{x}(n)\frac{\overline{\tilde{x}^2}+\hat{A}^2}{\sqrt{\hat{A}^2}\left(\overline{\tilde{x}^2}-\hat{A}^2\right)^2}. \quad (5.34)$$

Care should be exercised when choosing the initial guess for $\hat{A}_1^2$. This is due to the fact that Eqn. (5.20) has two zeros. One of these zeros is at $\hat{A}=0$ and the other zero is the desired one. One should choose $\hat{A}_1^2$ so that the converged iterative solution for A_1^2 could be achieved. For this purpose, an approximate-likelihood function for choosing $\hat{A}_1^2$ was presented in [42] . From [46], one can have the following approximation equation

$$\ln I_0\left[W(n)\right] = \frac{aW^2(n)}{1+bW(n)}, \quad (5.35)$$

where $a = 0.36699205, b = 0.36482285$. By using Eqns. (5.35) and (5.19), one can rewrite Eqn. (5.10) as

$$\sum_{n=1}^{N}\left\{\frac{\partial}{\partial A^2}\left[\frac{a\left(\frac{A\tilde{x}(n)}{\sigma^2}\right)^2}{1+b\frac{A\tilde{x}(n)}{\sigma^2}}\right]\right\}\Bigg|_{A^2=\hat{A}^2} - \frac{N}{2\sigma^2} = 0,$$

which can be simplified as

$$\sum_{n=1}^{N}\left\{\frac{\partial}{\partial A^2}\left[\frac{A^2\tilde{x}^2(n)}{\sigma^2+bA\tilde{x}(n)}\right]\right\}\Bigg|_{A^2=\hat{A}^2}-\frac{N}{2a}=0,$$

or finally as

$$\sum_{n=1}^{N}\frac{2\tilde{x}^2(n)\left[2+bW(n)\right]}{\left(\overline{\tilde{x}^2}-\hat{A}^2\right)\left[1+bW(n)\right]^2}-\frac{N}{a}=0. \tag{5.36}$$

Eqn. (5.36) is called approximate-likelihood equation which has no root at $\hat{A}=0$, while preserving the second root that gives an approximate MLE estimate of A^2.

The ML estimation procedure is summarized as follows:

1) use a numerical technique to determine the solution of the approximate-likelihood equation of Eqn. (5.36) as $\hat{A}_1^2$;

2) use $\hat{A}_1^2$ as the first estimate for iterations into Eqn. (5.20) to get a solution of $\hat{A}^2$ with the desired accuracy.

3) obtain $\hat{\sigma}^2$ simply by substituting $\hat{A}^2$ into Eqn. (5.18).

4) finally, obtain the SNR estimator as $\hat{\beta}=\dfrac{\hat{A}^2}{2\hat{\sigma}^2}$.

5.3 SNR Estimation Based on Lower-Order Moments

The data model is the same as in Eqn. (5.6). This estimation method uses the averages of the first and second powers of envelope data points [42]. From [1] (see Eqn. (1.1.140) on p. 31 with $n=2$), a random variable $\tilde{x}$ with PDF given by Eqn. (5.6) has its k^{th} moments as given by

$$E\left[\tilde{x}^k\right]=\left(2\sigma^2\right)^{k/2}\exp\left(-\frac{A^2}{2\sigma^2}\right)\,{}_1F_1\left(\frac{k+1}{2};1;\frac{A^2}{2\sigma^2}\right), \tag{5.37}$$

where ${}_1F_1\left(a,b,\tilde{x}\right)$ is the confluent hypergeometric function of the first kind given by [47,48]

$${}_1F_1\left(a,b,\tilde{x}\right)=\sum_{i=0}^{\infty}\frac{(a)_i\,\tilde{x}^i}{(b)_i\;i!}, \tag{5.38}$$

where

$$(a)_i = \prod_{k=0}^{i-1}(a+k) = \frac{\Gamma(a+i)}{\Gamma(a)}, \tag{5.39}$$

and where $\Gamma(\cdot)$ represents the Gamma Function. In this way, it can be shown that

$$E[\tilde{x}] = \sqrt{\frac{\pi}{2}}\sigma \exp\left(-\frac{A^2}{2\sigma^2}\right) {}_1F_1\left(\frac{3}{2}; 1; \frac{A^2}{2\sigma^2}\right), \tag{5.40}$$

$$E\left[\tilde{x}^2\right] = A^2 + 2\sigma^2. \tag{5.41}$$

By using the following mathematical average to substitute the statistical average

$$\overline{\tilde{x}}^2 = \left\{\frac{1}{N}\sum_{n=1}^{N}[\tilde{x}(n)]\right\}^2 \tag{5.42}$$

$$\overline{\tilde{x}^2} = \frac{1}{N}\sum_{n=1}^{N}[\tilde{x}(n)]^2, \tag{5.43}$$

one can obtain $\hat{A}^2$, the estimate of A^2, by using

$$\overline{\tilde{x}} = \sqrt{\frac{\pi}{2}}\sigma \exp\left(-\frac{A^2}{2\sigma^2}\right) {}_1F_1\left(\frac{3}{2}; 1; \frac{A^2}{2\sigma^2}\right), \tag{5.44}$$

or

$$\overline{\tilde{x}}^2 = \frac{\pi}{2}\sigma^2\left[\exp\left(-\frac{A^2}{2\sigma^2}\right) {}_1F_1\left(\frac{3}{2}; 1; \frac{A^2}{2\sigma^2}\right)\right]^2. \tag{5.45}$$

From Eqn. (5.41), by replacing σ^2 by $\hat{\sigma}^2$ and A^2 by $\hat{A}^2$, one gets

$$\hat{\sigma}^2 = \frac{1}{2}\left[\overline{\tilde{x}^2} - \hat{A}^2\right]. \tag{5.46}$$

By substituting Eqn. (5.46) into Eqn. (5.45), one can obtain

$$\overline{\tilde{x}}^2 = \frac{\pi}{4}\left(\overline{\tilde{x}^2} - \hat{A}^2\right)\left[\exp\left(-\frac{A^2}{\overline{\tilde{x}^2} - \hat{A}^2}\right) {}_1F_1\left(\frac{3}{2}; 1; \frac{A^2}{\overline{\tilde{x}^2} - \hat{A}^2}\right)\right]^2$$

or

$$\frac{\overline{\tilde{x}}^2}{\overline{\tilde{x}^2}} = \frac{\pi}{4}\left(1 - \frac{\hat{A}^2}{\overline{\tilde{x}^2}}\right)\left[\exp\left(-\frac{A^2}{\overline{\tilde{x}^2} - \hat{A}^2}\right) {}_1F_1\left(\frac{3}{2}; 1; \frac{A^2}{\overline{\tilde{x}^2} - \hat{A}^2}\right)\right]^2. \tag{5.47}$$

From the following property of confluent hypergeometric function [49]

$${}_1F_1(a, b, \tilde{x}) = \exp(\tilde{x})\,{}_1F_1(b-a, b, -\tilde{x}), \tag{5.48}$$

one can simplify the third term of Eqn. (5.47) as

$$\exp\left(-\frac{A^2}{\overline{\tilde{x}^2}-\hat{A}^2}\right)\ {}_1F_1\left(1-\left(-\frac{1}{2}\right);1;\frac{A^2}{\overline{\tilde{x}^2}-\hat{A}^2}\right) \tag{5.49}$$

$$= {}_1F_1\left(-\frac{1}{2};1;-\frac{A^2}{\overline{\tilde{x}^2}-\hat{A}^2}\right). \tag{5.50}$$

Finally, the estimations for $\hat{A}^2$ and $\hat{\sigma}^2$ can be obtained by using the relationships

[2]

$$\frac{\overline{\tilde{x}}^2}{\overline{\tilde{x}^2}} = \frac{\pi}{4}\left(1-\frac{\hat{A}^2}{\overline{\tilde{x}^2}}\right)\left[{}_1F_1\left(-\frac{1}{2};1;-\frac{\hat{A}^2/\overline{\tilde{x}^2}}{1-\hat{A}^2/\overline{\tilde{x}^2}}\right)\right]^2, \tag{5.51}$$

$$\hat{\sigma}^2 = \frac{1}{2}\left[\overline{\tilde{x}^2}-\hat{A}^2\right]. \tag{5.52}$$

These relationships permit a simple procedure for the determination of $\hat{A}^2$. Since the quantity $\overline{\tilde{x}}^2/\overline{\tilde{x}^2}$ can be calculated directly form Eqn. (5.51) for a given value of $\hat{A}^2/\overline{\tilde{x}^2}$, values of $\overline{\tilde{x}}^2/\overline{\tilde{x}^2}$ over the range of admissible values of $\hat{A}^2/\overline{\tilde{x}^2}$ are first tabulated. Moreover, the solution of $\hat{A}^2$ can be obtained by simple table look-up and interpolation. Then, $\hat{\sigma}^2$ can be obtained from Eqn. (5.52). Finally, the SNR estimate is obtained as $\hat{\beta} = \frac{\hat{A}^2}{2\hat{\sigma}^2}$.

5.4 SNR Estimation Based on Higher-Order Moments

This moment based method was originally introduced in [42], discussing how to estimate SNR from the envelope of the analytic signal associated with a single-tone signal contaminated in AWGN noise.

For the moment based SNR estimation, one just needs the second (M_2) and fourth moments (M_4) of $\tilde{x}$. The second moment is the same as that in Eqn. (5.41), while the fourth order moment is given from Eqn. (5.37) as

$$E\left[\tilde{x}^4\right] = A^4 + 8\sigma^2 A^2 + 8\sigma^4. \tag{5.53}$$

[2] There is an error in Eqn. (25) of [42] (c.f. Eqn. (5.51)).

Using numerical averages to replace statistical means, and solving Eqns. (5.41) and (5.53), one can obtain the estimate for the squared-amplitude and signal variance explicitly as

$$\hat{A}^2 = \left[2\left(\overline{\tilde{x}^2}\right)^2 - \overline{\tilde{x}^4}\right]^{1/2} \tag{5.54}$$

and

$$\hat{\sigma}^2 = \frac{1}{2}\left[\overline{\tilde{x}^2} - \hat{A}^2\right],$$

where

$$\overline{\tilde{x}^2} = \frac{1}{N}\sum_{n=1}^{N}\left[\tilde{x}(n)\right]^2, \tag{5.55}$$

$$\overline{\tilde{x}^4} = \frac{1}{N}\sum_{n=1}^{N}\left[\tilde{x}(n)\right]^4. \tag{5.56}$$

Therefore, the SNR estimator in terms of second and fourth order moments can be expressed as:

$$\hat{\beta} = \frac{\sqrt{2\left(\overline{\tilde{x}^2}\right)^2 - \overline{\tilde{x}^4}}}{\overline{\tilde{x}^2} - \hat{A}^2}. \tag{5.57}$$

5.5 Investigations and Comparisons

This section will present Monte Carlo simulation results for the comparison of the performance characteristics of the three SNR estimation methods for the case of a single-tone signal contaminated in AWGN noise, as discussed in sections 2-4. The data size for the each run of the simulation is denoted by N, and the number of runs for each estimate is denoted by M. In the resulting plots, the standard deviation is used which is defined as the square root of estimator variance. The biases or variances are then divided by the corresponding SNR values to obtain the normalized values. The Monte Carlo simulations are repeated for various data size N to observe its effect on the estimation performance. For each data set, $M = 500$ simulations were made to obtain the average value of the estimate.

5.5.1 Application of the Classical CRLB Bound

The classical CRLB bound is given by the right-hand side of the following relationship [42]

$$\frac{\sigma_{\hat{\beta}}}{\beta} \geqslant \frac{1}{\sqrt{N}} \frac{1}{\beta} \sqrt{\frac{1 + (\beta)^2 [F(\beta) - 1]}{-1 + (1 + 2\beta)[F(\beta) - 1]}}, \tag{5.58}$$

where $\sigma_{\hat{\beta}}$ is the standard deviation of the SNR estimator, and $F(\beta)$ represents a certain function (c.f. Eqn. (5) of [42]) as

$$F(\beta) = \int_0^\infty \frac{v^3}{2\beta} \left[\frac{I_1^2\left(\sqrt{2\beta} v\right)}{I_0\left(\sqrt{2\beta} v\right)} \right] \exp\left(-\frac{v^2 + 2\beta}{2} \right) dv. \tag{5.59}$$

To simplify the formula, Benedict and Soong [42] gave an approximation to $F(\beta)$ as

$$F(\beta) \approx \frac{2 + 2\beta}{1 + 2\beta}. \tag{5.60}$$

Simulations are conducted for three values of N, namely $N = 51, 101, 201$. Figs. 5.1 to 5.3 are the plots of simulation results for the three estimators showing their normalized standard deviations versus SNRs together with the classical CRLB bound. Fig. 5.1 is for $N = 51$, Fig. 5.2 is for $N = 101$ and Fig. 5.3 is for $N = 201$. From the three figures, one can make the following observations. Firstly, when SNRs are small (e.g. smaller than 1), the simulated normalized standard deviations are less than what the CRLB bound predicts. Therefore, the classical CRLB bound is no longer the lower bound at very low SNRs. Secondly, increasing N will decrease both the CRLB bound and the variance of the SNR estimators. For the same SNR, Fig. 5.1 shows the biggest variance, whereas Fig. 5.3 shows the lowest variance among the three cases. This is because larger N means that more data points are available to make an estimate. Therefore, the estimates are more accurate resulting less estimator variance.

Figs. 5.4 to 5.6 are the plots of simulation results for normalized estimator biases versus SNRs for $N = 51$, 101 and 201, respectively. It can be ovserved that the three

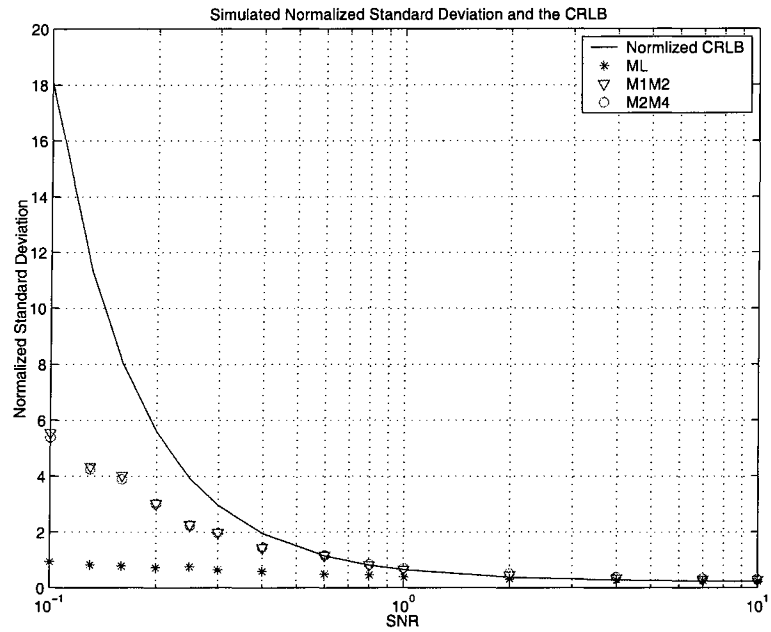

Figure 5.1: Normalized standard deviation versus SNR together with the classical CRLB for $N = 51$.

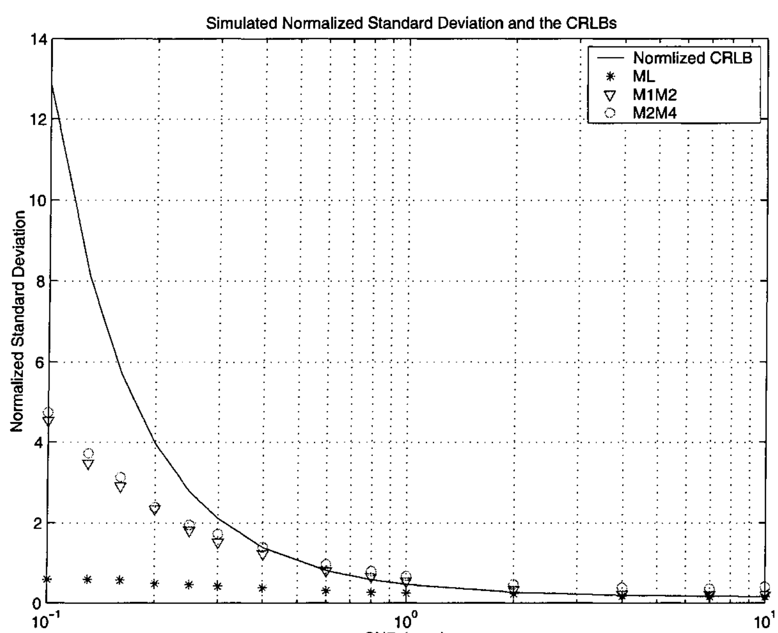

Figure 5.2: Normalized standard deviation versus SNR together with the classical CRLB for $N = 101$.

estimators are biased when SNRs are smaller than 2. Therefore, it is not appropriate to use the classical CRLB bound as a performance metric to compare the estimator performance. Among the three estimators, the ML method has the smallest bias and errors in all cases.

5.5.2 Application of the UCRLB Bound to SNR Estimation

From the discussion in section 5.5.1, the classical CRLB bound is not a valid lower bound on the SNR estimator variance. Therefore, the UCRLB bound should be used instead. As discussed in Chapter 4, the bias gradient must be known at the outset in order to use the UCRLB bound. Unfortunately, the bias gradient cannot be expressed in a closed-form analytical relationship. Consequently, one has to resort to an estimation technique to obtain its numerical value. In [33], an algorithm was developed for the numerical calculation of the bias gradient.

Unfortunately, the algorithm in [33] is not directly applicable to the above three estimators, mainly due to the fact that it uses only one data point to obtain the estimate. As mentioned in Chapter 4, it is not possible to use only one data point to estimate the SNR. For example, in the moments-based estimation algorithm, if one data point $x(1)$ is used for the estimation, then from Eqns. (5.55) and (5.56), $\overline{\tilde{x}^2} = x^2(1)$ and $\overline{\tilde{x}^4} = x^4(1)$. But from Eqn. (5.54), one gets $\hat{A}^2 = 0$ which is not a valid estimate. Therefore, the algorithm fails.

To circumvent the above problem, we propose an algorithm involving M simulation runs of estimates based on N data points, so as to get one average estimate. From Eqn. (4.18) in Chapter 4, one can estimate the bias gradient for the SNR estimator in accordance with

$$\nabla\widehat{b}(\beta) = \frac{1}{M-1}\sum_{i=1}^{M}\left[\hat{\beta}^2 - \frac{1}{M}\sum_{j=1}^{M}\hat{\beta}^2\right]\nabla \ln p\left(\underline{\tilde{x}}^{(i)};\underline{\theta}\right) - \nabla g(\theta), \tag{5.61}$$

where $x^{(i)} = [x_1^{(i)} \cdots x_L^{(i)}]^T$ is the ith set of L statistically independent measured

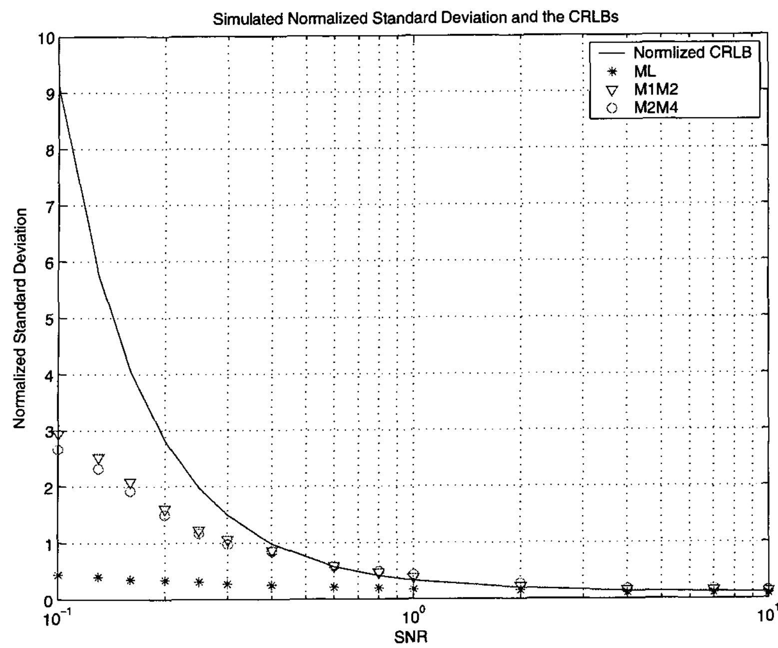

Figure 5.3: Normalized standard deviation versus SNR together with the classical CRLB for $N = 201$.

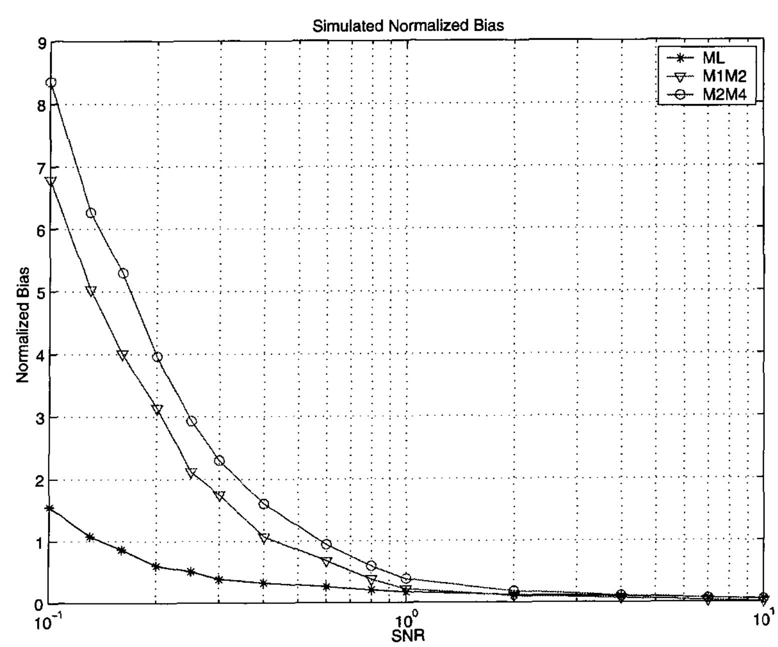

Figure 5.4: Normalized Bias versus SNR for $N = 51$.

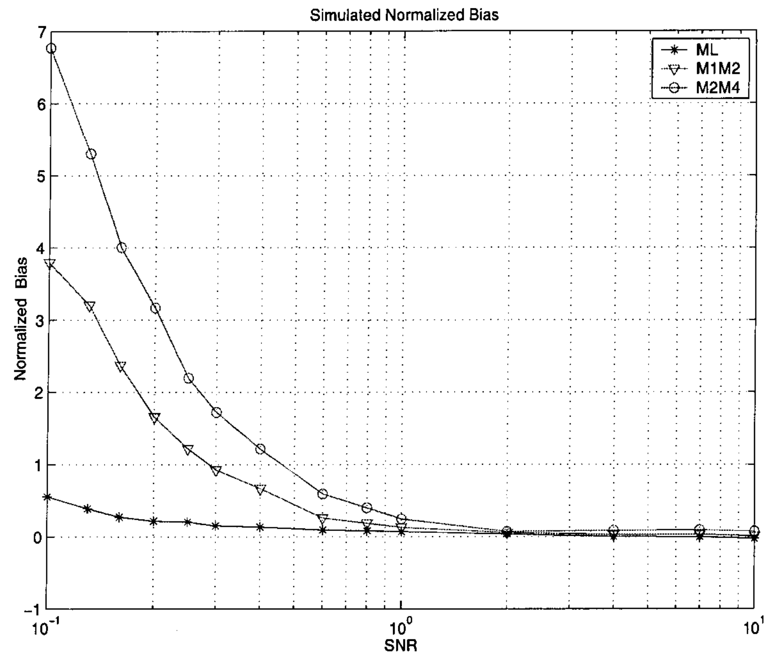

Figure 5.5: Normalized Bias versus SNR for $N = 101$.

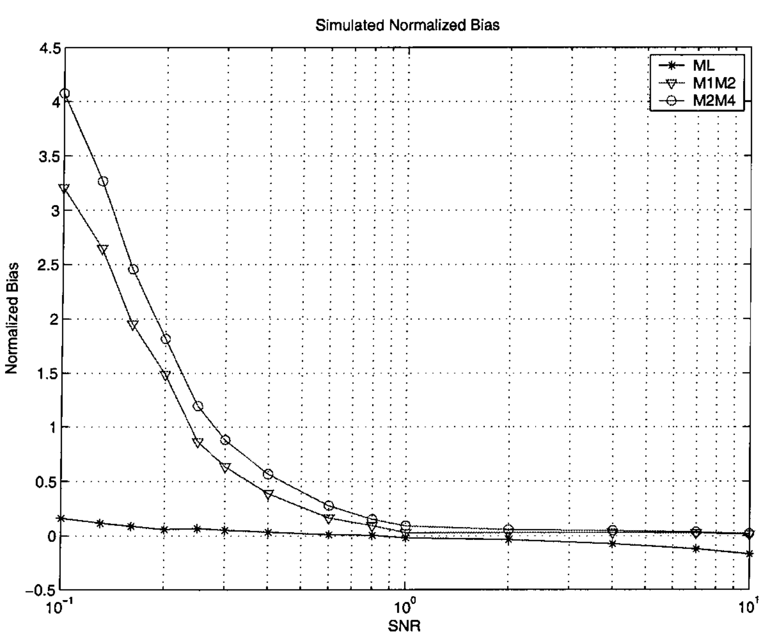

Figure 5.6: Normalized Bias versus SNR for $N = 201$.

random variables and $p\left(\underline{\tilde{x}}^{(i)};\underline{\theta}\right)$ is their joint PDF in accordance with

$$p\left(\underline{\tilde{x}}^{(i)};\underline{\theta}\right)=\prod_{k=1}^{L}\frac{\tilde{x}_k^{(i)}}{\sigma^2}I_0\left(\frac{A\tilde{x}_k^{(i)}}{\sigma^2}\right)\exp\left[-\frac{\left(\tilde{x}_k^{(i)}\right)^2+A^2}{2\sigma^2}\right]. \tag{5.62}$$

The elements of $\nabla\ln p(\underline{\tilde{x}}^{(i)};\underline{\theta})$ in Eqn. (5.61) are given by

$$\frac{\partial\ln p\left(\underline{\tilde{x}}^{(i)};\underline{\theta}\right)}{\partial\theta_1} = \sum_{n=1}^{N}\left\{-\frac{1}{2\theta_2}+\frac{1}{2\sqrt{\theta_1}\theta_2}\frac{I_1\left(\sqrt{\theta_1}\tilde{x}_k^{(i)}/\theta_2\right)}{I_0\left(\sqrt{\theta_1}\tilde{x}_k^{(i)}/\theta_2\right)}\right\}, \tag{5.63}$$

$$\frac{\partial\ln p\left(\underline{\tilde{x}}^{(i)};\underline{\theta}\right)}{\partial\theta_2} = \sum_{n=1}^{N}\left\{-\frac{1}{\theta_2}+\frac{\left(\tilde{x}_k^{(i)}\right)^2+\theta_1}{2\theta_2^2}-\frac{\sqrt{\theta_1}}{\theta_2^2}\frac{I_1\left(\sqrt{\theta_1}\tilde{x}_k^{(i)}/\theta_2\right)}{I_0\left(\sqrt{\theta_1}\tilde{x}_k^{(i)}/\theta_2\right)}\right\}. \tag{5.64}$$

Figs. 5.7 to 5.9 are the plots of simulation results showing the estimator variances for the above three SNR estimators relative to the UCRLB bound for $N=51,101$ and 201, respectively. The surface plot in each figure shows the bias-variance trade-off representing the UCRLB bound in the $\sigma-\delta-SNR$ space. The space above the bias-variance trade-off surface is the achievable region and all the SNR estimators should be placed here. The space below the surface is unachievable region which means no SNR estimator can reach it. If an estimator is lied on the surface, one can say that the estimator has reached the UCRLB bound.

It can be observed from Figs. 5.7 to 5.9 that all three estimators are within the achievable region and are bounded by the UCRLB bound. This implies that the UCRLB bound is a valid lower bound for the SNR estimators. The ML estimator has a better estimation accuracy than the other two moments-based methods. The higher order moments method has less computational complexity, as it only involves calculating the average of the squared data values as well as the average of the fourth power of data values. However, its performance is the worst among the three estimators concerned according to the simulation results. In this way, there is generally a tradeoff between the implementation complexity and estimation accuracy. By comparison, the ML estimator has the best performance at only moderate complexity. Finally, it can be observed that increasing the data size N will decrease the UCRLB

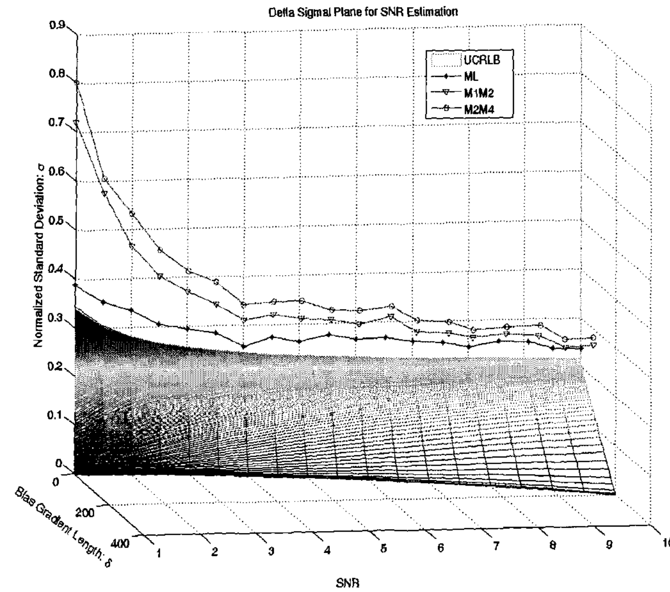

Figure 5.7: Estimation variance of SNR estimators whthin $\sigma - \delta - SNR$ space for $N = 51$.

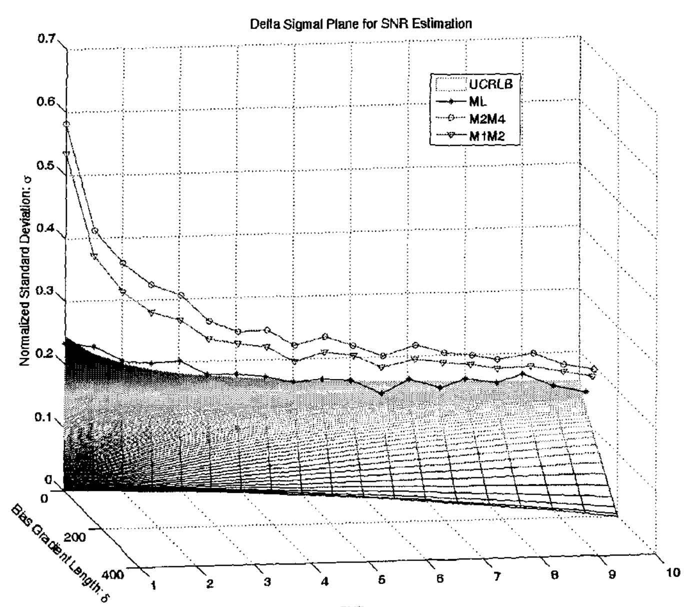

Figure 5.8: Estimation variance of SNR estimators whthin $\sigma - \delta - SNR$ space for $N = 101$.

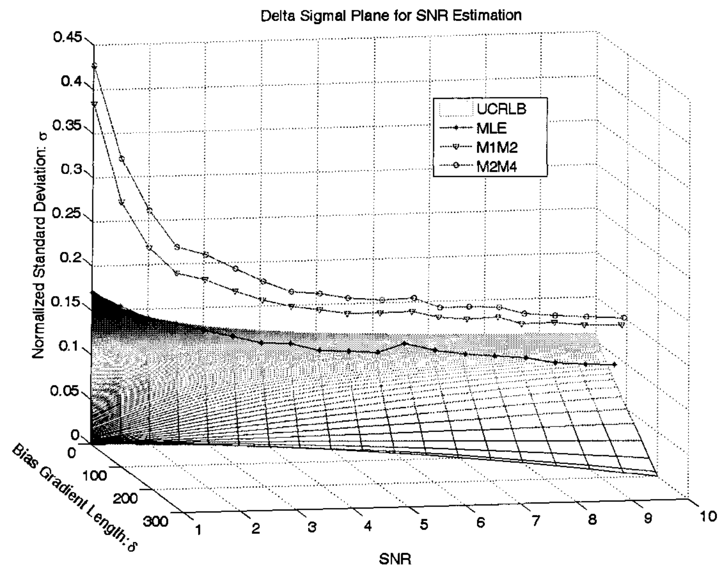

Figure 5.9: Estimation variance of SNR estimators whthin $\sigma - \delta - SNR$ space for $N = 201$.

bound as well as the estimator variances. Similarly, as in Figs. 5.1 to 5.3, increasing the data size N will decrease the UCRLB bound and the estimator variances. For the same SNR, Fig. 5.7 shows the largest variance and the highest UCRLB bound surface where $N = 51$, whereas Fig. 5.9 shows the smallest variance and the lowest UCRLB bound surface where $N = 201$.

5.5.3 Application of the UCRLB Bound to Multiharmonic Estimation

In Chapter 3, the estimator performance for multiharmonic signal was compared to the classical CRLB bound. As noted there, at SNRs smaller than -5 dB, the simulation estimation variance for T, the signal period, does not conform to the bound (c.f. Fig. 3.3). This is because the estimator exhibits noticeable bias at very low SNRs (c.f. Fig. 3.4) and the classical CRLB bound is valid for unbiased estimation only. It was also mentioned there that some biased CRLB bound should be considered instead. Here, the UCRLB bound will be employed to resolve the above problem. Figs. 5.12 - 5.13 are the simulation results for $N = 500$ and $N = 1000$ respectively. It is clearly shown that the UCRLB bound is the valid lower bound on estimator variance at very low SNRs.

Also one can employ UCRLB bound to other estimators, i.e. the SNR estimator in harmonics signal contaminated in AWGN noise as discussed in Chapter 3. Remember that in Chapter 3, the classical CRLB bound is not the valid lower bound on estimator variance at very low SNRs. The estimator is based on $\underline{\alpha} = [1, 0.2, 0.05, 0.01, 1, 0.2, 0.05, 0.01, 4, 6, 4, 6]^T$ as in Eqn. (3.2) in Chapter 3. To calculate the UCRLB bound, one needs to estimate the bias gradient which can be obtained via the numerical method as discussed before. For the estimation of T, its gradient to the parameter $\underline{\theta}$ is obtained as

$$\frac{\partial T}{\partial \underline{\theta}} = [0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0]^T.$$

For the estimation of SNR, its gradient to the parameter $\underline{\theta}$ is

$$\frac{\partial \beta}{\partial \underline{\theta}} = \left[\frac{2A_1}{\sigma^2}\ 0\ 0\ 0\ \frac{2B_1}{\sigma^2}\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ \frac{-A_1^2 - B_1^2}{(\sigma^2)^2}\right]^T,$$

where β represents the SNR.

The simulation results for SNR estimation are shown Figs. 5.10 ($N = 500$) and 5.11 ($N = 1000$). It can be seen that the UCRLB bound is the valid bound on the estimator variance with regard to the estimation of SNR for both figures. This is because the UCRLB has taken into consideration the estimation bias where the classic CRLB assumes the estimator is unbiased which is not true in cases being considered. One can still find the basic tendency that when N increases, i.e. the data points increase, the estimator variance becomes close to the UCRLB bound and the estimation variance also decreases. This is because of the fact that the more data points available for estimation, the more accurately one can estimate the parameters. Similar conclusions can be drawn from the simulation results for T estimation as shown in Figs. 5.12 ($N = 500$) and 5.13 ($N = 1000$).

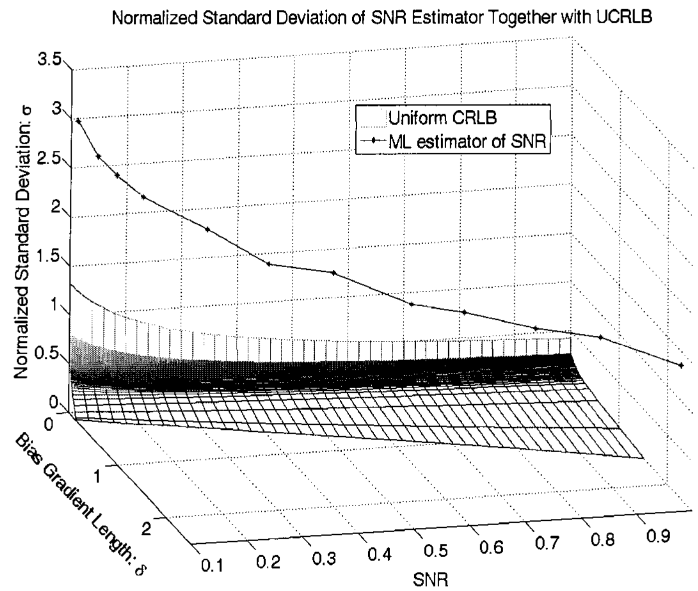

Figure 5.10: Estimation variance of SNR estimator for harmonics whthin $\sigma - \delta - SNR$ space for $N = 500$.

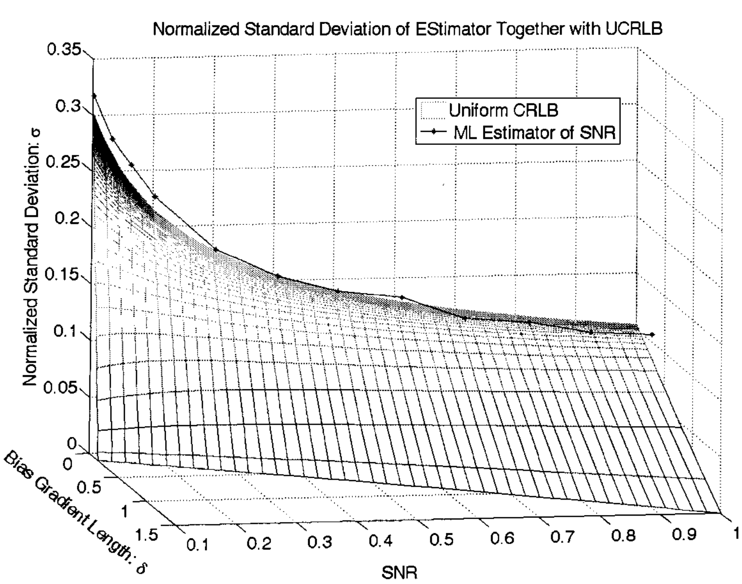

Figure 5.11: Estimation variance of SNR estimator for harmonics whthin $\sigma - \delta - SNR$ space for $N = 1000$.

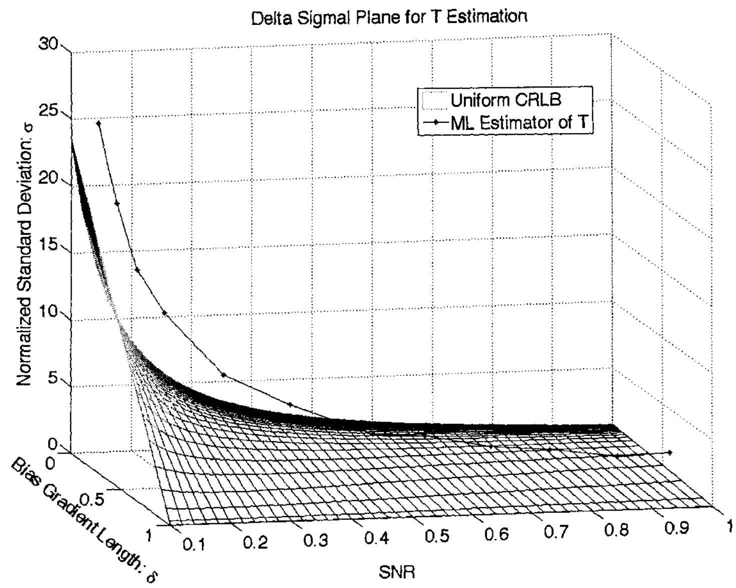

Figure 5.12: Estimation variance of T estimator whthin $\sigma - \delta - SNR$ space for $N = 500$.

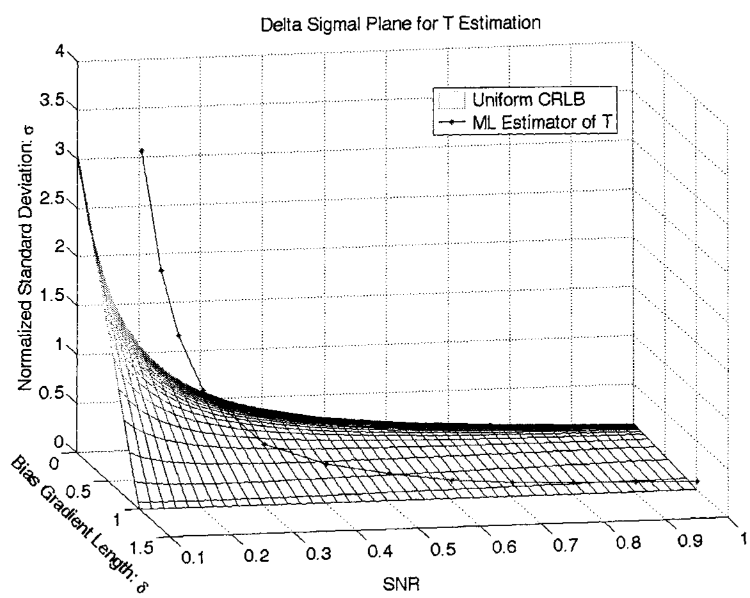

Figure 5.13: Estimation variance of T estimator whthin $\sigma - \delta - SNR$ space for $N = 1000$.

5.6 Conclusion

This chapter investigated three commonly-used SNR estimation techniques available in the literature and compared their performance to the newly developed UCRLB bounds. The investigation firstly shows that the classical CRLB bound is not a valid lower bound when SNR is small. Then the UCRLB bound is employed and shown to be the valid bound for SNR estimators. By placing the SNR estimators on the same variance-bias gradient plane for performance comparison, one can identify the "best" SNR estimator in terms of the estimation variance. Among the three SNR estimators, the ML method is better than the other two moments based estimation in the sense that it has less variance at all SNR values. The higher order moments based method is less accurate compared to lower order moments method while the former has less computation complexity. The other application of UCRLB also discussed in section 5.5 for the estimation of harmonic signal in AWGN noise. The simulation results show that the UCRLB is a valid bound for the estimator variance even at small SNRs.

Chapter 6

Summary and Future Work

6.1 Contribution and Main Results

This thesis has been concerned with the estimation of frequency, amplitude, phase, noise variance and SNR for two different data models, namely, single-tone and multi-harmonic sinusoidal signals contaminated by AWGN noise. Investigations have been undertaken to compare the performance characteristic of different estimation techniques to identify the best estimator in terms of the estimation variance relative to CRLB bound for unbiased estimation, and UCRLB bound for biased estimation especially. The contributions and main results of this thesis are summarized in the following [50, 51].

1. The linear regression estimator (LRE) for the frequency of single-tone sinusoidal signals contaminated by AWGN noise [22] has been derived in a general manner in Chapter 2: [22] only considered the high SNR case, but in this thesis, the full SNR range has been taken into account and the detailed PDF of phase noise has been derived. The performance of three estimators, namely, MLE estimator, linear prediction estimator (LPE), and LRE estimator, has been compared against the CRLB bound by Monte Carlo simulations. The results have shown that the LRE estimator is slightly more accurate than the other two methods in terms of estimation variance, except it has larger threshold than the MLE estimator.

2. In Chapter 3, the unbiased estimation of parameters in circadian rhythm

data model MICA [31] for harmonic/subharmonic sinusoidal signal contaminated in AWGN noise has been studied under the assumption that noise variance is not known a priori.: [31] assumed that the noise variance is known a priori. A new derivation of the Fisher information matrix has been given in order to arrive at the unbiased CRLB bounds for the model under consideration. The resulting CRLB bounds have been confirmed through Monte Carlo simulations using the MLE estimation method. It has been shown that the estimators achieve the CRLB bound at high SNRs, but that at very low SNRs, the estimators tend to be biased and therefore the unbiased CRLB bound is no longer the valid bound.

3. To compare the performance of biased estimators, a new bound called UCRLB bound [33] for biased estimation has been introduced in Chapter 4. A modified algorithm for estimating the bias gradient has been proposed in the general case when the estimation is based on multiple data points instead of a single data point: the latter was only considered in [33]. The original algorithm in [33] is not applicable to many estimation problems that use multiple data points, as in the case of SNR estimators discussed in Chapter 5. The proposed modified algorithm is therefore very useful when applying the UCRLB bound to the performance evaluation for those estimators.

4. A new equation for determining the estimation variance of the proposed bias gradient estimator has been presented in Chapter 4. The result has shown that the proposed bias gradient estimator is an unbiased estimator and its variance reaches zero when the data size approaches infinity. Based on this new derivation, an error has been located in Eqn. 24 of [33] which has a similar structure as the proposed one. A correct equation with detailed derivation has been e-mailed to Dr. Hero, the first author of [33], who later announced this error on his web page, c.f. [35].

5. Two new applications of UCRLB bound have been presented in Chapter 5. The first application concerns the performance comparison of SNR estimators for the single-tone signal contaminated by AWGN noise. The reason for applying UCRLB

bound to the SNR estimation problem is twofold. Firstly, SNR is a very important parameter in signal processing applications and most parameter estimation problems are SNR dependent. Secondly, the SNR estimation is inherently biased, and the classical CRLB bound is not applicable to performance evaluations anymore. Chapter 5 investigated three commonly-used SNR estimation techniques available in the literature and compared their performance to the newly developed UCRLB bound. By placing the SNR estimators on the same variance-bias gradient plane for performance comparison, one can identify the "best" SNR estimator in terms of the estimation variance. Among the above mentioned three SNR estimators, the simulation results show that the MLE method is better than the other two moments based estimation in the sense that it has less variance at all SNR values. The higher order moments based method is less accurate compared to lower order moments method while the former has less computation complexity.

The second application of UCRLB is in the estimation of harmonically related multitone signals contaminated by AWGN noise. In Chapter 3, it was argued that the classical CRLB bound is not a valid lower bound at low SNRs in that the estimator variance is smaller that the bound. In Chapter 5, this problem has been resolved by applying the new UCRLB bound as the metric for performance comparison. Also, the original UCRLB bound has been extended for multiple estimators, such as the estimation of multiple amplitude components in MICA model. The simulation results show that the UCRLB is a valid bound for the estimator variance even at small SNRs.

6.2 Future Work

1. The SNR estimation problems considered in Chapter 5 could be extended to other data models, such as the 8-PSK in complex AWGN noise as studied in [52] for the wireless communication systems. The UCRLB bounds should be derived and applied on those situations as well.

2. One may compare different SNR estimators, other than methods discussed in

this thesis, with the UCRLB bounds to identify the best estimator. In this thesis, only the most commonly used methods has been considered. There are more SNR estimation techniques available in the literature, and also one can develop new algorithms to try to obtain the best biased estimator by comparing the estimation performances against UCRLB bound.

3. There may be a better way to estimate the norm of the bias gradient for the calculation of UCRLB bound. In this thesis, a numerical method has been employed to estimate the bias gradient, and the length of bias gradient has been obtained by a simple square root operation of the average value to its squared components. This may cause some inaccuracy problems when plotting the points in delta-sigma plane. The boot strap method mentioned in [33] should be a good starting point to solve this problem.

4. Some other lower bounds may be studied. In this regard, Bellini-Tartara bounds [53, 54] may prove to be choices as these bounds do not make any recourse if the estimator is biased or not. Presently, the UCRLB bound seems to be the best suitable bound for biased SNR estimators.

Appendix A

This appendix provides a proof for the validity of the following theorem.

Let $w(n)$ represent an additive white Gaussian noise (AWGN) with a mean of zero and a variance of σ^2, and let

$$\eta(n) = \sum_{k=-\infty}^{\infty} w(k)h(n-k), n = 0, 1, \cdots, N-1, \tag{A.1}$$

where $h(n)$ is Hilbert transform sequence defined as

$$h(n) = \begin{cases} \frac{2}{\pi}\frac{\sin^2(n\pi/2)}{n}, & n \neq 0 \\ 0 & , n = 0 \end{cases} .$$

Then,

1) $\eta(n)$ is also a zero-mean Gaussian variable with a variance of σ^2,

2) $\eta(n)$ and $w(n)$ are uncorrelated.

Proof: From Eqn. (A.1), one has

$$E[\eta(n)] = \sum_{k=-\infty}^{\infty} E[w(n)]\, h(n-k) = 0. \tag{A.2}$$

But,

$$\phi_{ww}(m) = E[w(n)\, w(n+m)] = \sigma^2 \delta(m),$$

where $\delta(m)$ is the Kronecker delta sequence. Therefore,

$$\begin{aligned}
\phi_{\eta\eta}(m) &= E\left[\eta(n)\,\eta(n+m)\right] \\
&= E\left[\sum_{k=-\infty}^{\infty} w(k)h(n-k) \sum_{l=-\infty}^{\infty} w(l)h(n+m-l)\right] \\
&= \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} E\left[w(k)w(l)\right] h(n-k)h(n+m-l) \\
&= \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} \sigma^2\delta(l-k)\, h(n-k)h(n+m-l) \\
&= \sum_{k=-\infty}^{\infty} \sigma^2 h(n-k)h(n+m-k) \\
&= \sigma^2 \sum_{l=-\infty}^{\infty} h(l)h(l+m) \\
&= \sigma^2 c(m),
\end{aligned} \tag{A.3}$$

where $c(m) = \sum_{l=-\infty}^{\infty} h(l)h(l+m)$. Then,

$$\begin{aligned}
E\left[\eta^2(n)\right] &= \phi_{\eta\eta}(0) \\
&= \sigma^2 c(0) \\
&= \sigma^2 \sum_{l=-\infty}^{\infty} h^2(l).
\end{aligned} \tag{A.4}$$

From Parseval's theorem

$$\sum_{l=-\infty}^{\infty} h^2(l) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|H(e^{j\omega})\right|^2 d\omega, \tag{A.5}$$

Substituting Eqn. (2.6) into Eqn. (A.5), on can obtain

$$\sum_{l=-\infty}^{\infty} h^2(l) = 1.$$

Consequently, Eqn. (A.4) can be simplified to

$$\phi_{\eta\eta}(m) = \sigma^2.$$

Therefore, 1) is satisfied.

Furthermore, it can be shown that

$$\begin{aligned} E\left[w\left(n\right)\eta\left(m\right)\right] &= \sum_{k=-\infty}^{\infty} E\left[w\left(n\right)w(k)\right]h(m-k) \\ &= \sigma^2 \sum_{k=-\infty}^{\infty} \delta(n-k)h(m-k) \\ &= \sigma^2 h(m-n), \end{aligned}$$

and in particular

$$E\left[w\left(n\right)\eta\left(n\right)\right] = \sigma^2 h(0) = 0,$$

where Eqn. (2.3) has been used. Therefore, random variables $w\left(n\right)$ and $\eta\left(n\right)$ are uncorrelated. Therefore, 2) is satisfied.

Appendix B

This is to prove the validity of Eqns. (2.14) - (2.16) in Chapter 2, i.e.,

$$E\left[v_I(n)\right] = E\left[v_Q(n)\right] = 0, \tag{2.14}$$

$$E\left[v_I(n)v_Q(n)\right] = 0, \tag{2.15}$$

$$E\left[v_I^2(n)\right] = E\left[v_Q^2(n)\right] = \frac{\sigma^2}{A_0^2}. \tag{2.16}$$

Proof: From Eqns. (2.12) and (2.13),

$$\begin{aligned} v_I(n) &= \mathrm{Re}\left[\frac{1}{A_0} z(n) e^{-j(\omega_0 Tn + \theta_0)}\right] \\ &= \frac{1}{A_0}\left[w(n)\cos(\omega_0 Tn + \theta) + \eta(n)\sin(\omega_0 Tn + \theta)\right], \end{aligned} \tag{B.1}$$

and

$$\begin{aligned} v_Q(n) &= \mathrm{Im}\left[\frac{1}{A_0} z(n) e^{-j(\omega_0 Tn + \theta_0)}\right] \\ &= \frac{1}{A_0}\left[\eta(n)\cos(\omega_0 Tn + \theta) - w(n)\sin(\omega_0 Tn + \theta)\right]. \end{aligned} \tag{B.2}$$

Since $w(n)$ and $\eta(n)$ are both zero-mean Gaussian distributed random variables, it is easy to verify that

$$\begin{aligned} E\left[v_I(n)\right] &= \frac{1}{A_0}\left\{E\left[w(n)\right]\cos(\omega_0 Tn + \theta) + E\left[\eta(n)\right]\sin(\omega_0 Tn + \theta)\right\} \\ &= 0. \end{aligned} \tag{B.3}$$

Similarly,

$$E\left[v_Q(n)\right] = 0.$$

Therefore, the validity of Eqn. (2.14) is established.

Since $w(n)$ and $\eta(n)$ are also independent Gaussian random variables, each with a variance of σ^2, one can obtain

$$\begin{aligned}
& E\left[v_I(n)v_Q(n)\right] \\
= & \frac{1}{A_0^2}E\left\{\left[w(n)\cos(\omega_0 Tn+\theta)+\eta(n)\sin(\omega_0 Tn+\theta)\right]\cdot\right. \\
& \left.\left[\eta(n)\cos(\omega_0 Tn+\theta)-w(n)\sin(\omega_0 Tn+\theta)\right]\right\} \\
= & \frac{1}{A_0^2}\left\{E\left[w(n)\eta(n)\right]\cos^2(\omega_0 Tn+\theta)-E\left[w^2(n)\right]\cos(\omega_0 Tn+\theta)\sin(\omega_0 Tn+\theta)\cdot\right. \\
& \left.+E\left[\eta^2(n)\right]\cos(\omega_0 Tn+\theta)\sin(\omega_0 Tn+\theta)-E\left[w(n)\eta(n)\right]\sin^2(\omega_0 Tn+\theta)\right\} \\
= & \frac{1}{A_0^2}\left[-\sigma^2\cos(\omega_0 Tn+\theta)\sin(\omega_0 Tn+\theta)+\sigma^2\cos(\omega_0 Tn+\theta)\sin(\omega_0 Tn+\theta)\right] \\
= & 0. \qquad\qquad \text{(B.4)}
\end{aligned}$$

Therefore, the validity of Eqn. (2.15) is established. Moreover,

$$\begin{aligned}
& E\left[v_I^2(n)\right] \\
= & \frac{1}{A_0^2}E\left[w(n)\cos(\omega_0 Tn+\theta)+\eta(n)\sin(\omega_0 Tn+\theta)\right]^2 \\
= & \frac{1}{A_0^2}\left\{E\left[w^2(n)\right]\cos^2(\omega_0 Tn+\theta)+2E\left[\eta(n)w(n)\right]\cos(\omega_0 Tn+\theta)\sin(\omega_0 Tn+\theta)+\right. \\
& \left.+E\left[\eta^2(n)\right]\sin^2(\omega_0 Tn+\theta)\right\} \\
= & \frac{1}{A_0^2}\left[\sigma^2\cos^2(\omega_0 Tn+\theta)+\sigma^2\sin^2(\omega_0 Tn+\theta)\right] \\
= & \frac{\sigma^2}{A_0^2}. \qquad\qquad \text{(B.5)}
\end{aligned}$$

Similarly,

$$E\left[v_Q^2(n)\right]=\frac{\sigma^2}{A_0^2}.$$

Therefore, the validity of Eqn. (2.16) is established.

Appendix C

This appendix provides detailed derivations for some of the relationships originally appearing in Chapter 3. These include,

$$E\left[\underline{g}\underline{g}^T\right] = A\underline{\alpha}\underline{\alpha}^T A + \sigma^2 A, \tag{3.51}$$

$$E\left[\rho\underline{g}\right] = A\underline{\alpha}\underline{\alpha}^T A\underline{\alpha} + (N+2)\sigma^2 A\underline{\alpha}, \tag{3.52}$$

$$E\left[\underline{g}\underline{s}^T\right] = A\underline{\alpha}\underline{\alpha}^T X^T + \sigma^2 X^T, \tag{3.57}$$

$$E\left[\rho\underline{s}^T\right] = \underline{\alpha}^T A\underline{\alpha}\underline{\alpha}^T X^T + (N+2)\sigma^2\underline{\alpha}^T X^T, \tag{3.58}$$

$$4\left(\sigma^2\right)^4 E\left[G_{33}\left(\underline{\theta}\right)\right] = 2N\left(\sigma^2\right)^2. \tag{3.60}$$

Recall from Chapter 3 that $\underline{s} = X\underline{\alpha} + \underline{w}$, where $\underline{w} = \left[w_1, w_2, \ldots, w_N\right]^T \in \mathbf{R}^N$ is a sequence of independent, identically distributed (i.i.d.) AWGN variables with zero mean and an unknown variance of σ^2. Recall also that $\underline{s} \in \mathbf{R}^N$, $X \in \mathbf{R}^{N\times P}$, $\underline{\alpha} \in \mathbf{R}^P$, $A = X^T X \in \mathbf{R}^{P\times P}$, $\rho = \underline{s}^T\underline{s} \in \mathbf{R}$, and $\underline{g} = X^T\underline{s} \in \mathbf{R}^P$. Note that the following relationships of Eqns. (C.1) - (C.6) will be used in the derivations:

$$E\left[\underline{w}\right] = 0, \tag{C.1}$$

$$E\left[\underline{w}^T\underline{w}\right] = N\sigma^2, \tag{C.2}$$

$$E\left[\underline{w}\underline{w}^T\right] = \sigma^2 I_N, \tag{C.3}$$

$$E\left[\underline{w}\underline{w}^T\underline{w}\right] = 0, \tag{C.4}$$

$$E\left[\underline{w}^T\underline{w}\underline{w}^T\right] = 0, \tag{C.5}$$

$$E\left[\underline{w}^T\underline{w}\underline{w}^T\underline{w}\right] = (N^2 + 2N)\sigma^4. \qquad \text{(C.6)}$$

1. Derivation of Eqn. (3.51) :

$$E\left[\underline{g}\underline{g}^T\right] = E\left[X^T\underline{s}\underline{s}^T X\right] = X^T E\left[\underline{s}\underline{s}^T\right] X. \qquad \text{(C.7)}$$

But, from Eqn. (3.25),

$$E\left[\underline{s}\underline{s}^T\right] = X\underline{\alpha}\underline{\alpha}^T X^T + \sigma^2 I_N. \qquad \text{(C.8)}$$

Substituting Eqn. (C.8) into Eqn. (C.7) gives

$$\begin{aligned} E\left[\underline{g}\underline{g}^T\right] &= X^T\left(X\underline{\alpha}\underline{\alpha}^T X^T + \sigma^2 I_N\right) X \\ &= A\underline{\alpha}\underline{\alpha}^T A + \sigma^2 A, \end{aligned} \qquad \text{(C.9)}$$

which is the result in Eqn. (3.51).

2. Derivation of Eqn. (3.52) :

$$\begin{aligned} E\left[\rho\underline{g}\right] &= E\left[\underline{g}\rho\right] \\ &= E\left[X^T\underline{s}\underline{s}^T\underline{s}\right] \\ &= X^T E\left[\underline{s}\underline{s}^T\underline{s}\right]. \end{aligned} \qquad \text{(C.10)}$$

But,

$$\begin{aligned} E\left[\underline{s}\underline{s}^T\underline{s}\right] &= E\left[(X\underline{\alpha} + \underline{w})\left(\underline{\alpha}^T X^T + \underline{w}^T\right)(X\underline{\alpha} + \underline{w})\right] \\ &= E\left[X\underline{\alpha}\underline{\alpha}^T X^T X\underline{\alpha} + X\underline{\alpha}\underline{w}^T\underline{w} + \underline{w}\underline{\alpha}^T X^T\underline{w} + \underline{w}\underline{w}^T X\underline{\alpha}\right] \\ &= X\underline{\alpha}\underline{\alpha}^T X^T X\underline{\alpha} + X\underline{\alpha}E\left[\underline{w}^T\underline{w}\right] + E\left[\underline{w}\underline{w}^T\right] X\underline{\alpha} + E\left[\underline{w}\underline{w}^T\right] X\underline{\alpha} \\ &= X\underline{\alpha}\underline{\alpha}^T A\underline{\alpha} + N\sigma^2 X\underline{\alpha} + 2\sigma^2 X\underline{\alpha} \\ &= X\underline{\alpha}\underline{\alpha}^T A\underline{\alpha} + (N+2)\sigma^2 X\underline{\alpha}. \end{aligned} \qquad \text{(C.11)}$$

Substituting Eqn. (C.11) into Eqn. (C.10) leads to

$$\begin{aligned} E\left[\rho\underline{g}\right] &= X^T\left[X\underline{\alpha}\underline{\alpha}^T A\underline{\alpha} + (N+2)\sigma^2 X\underline{\alpha}\right] \\ &= X^T X\underline{\alpha}\underline{\alpha}^T A\underline{\alpha} + (N+2)\sigma^2 X^T X\underline{\alpha} \\ &= A\underline{\alpha}\underline{\alpha}^T A\underline{\alpha} + (N+2)\sigma^2 A\underline{\alpha}, \end{aligned} \qquad \text{(C.12)}$$

which is the result in Eqn. (3.52).

3. Derivation of Eqn. (3.57) :

$$\begin{aligned} E\left[\underline{g}\underline{s}^T\right] &= E\left[X^T \underline{s}\underline{s}^T\right] \\ &= X^T E\left[\underline{s}\underline{s}^T\right] \\ &= X^T\left[X\underline{\alpha}\underline{\alpha}^T X^T + \sigma^2 I_N\right] \\ &= A\underline{\alpha}\underline{\alpha}^T X^T + \sigma^2 X^T, \end{aligned} \tag{C.13}$$

which is Eqn. (3.57).

4. Derivation of Eqn. (3.58) :

$$\begin{aligned} E\left[\rho \underline{s}^T\right] &= E\left[\underline{s}^T \underline{s}\underline{s}^T\right] \\ &= E\left[\left(\underline{\alpha}^T X^T + \underline{w}^T\right)\left(X\underline{\alpha} + \underline{w}\right)\left(\underline{\alpha}^T X^T + \underline{w}^T\right)\right] \\ &= E\left[\left(\underline{\alpha}^T X^T X\underline{\alpha} + \underline{\alpha}^T X^T \underline{w} + \underline{w}^T X\underline{\alpha} + \underline{w}^T \underline{w}\right)\left(\underline{\alpha}^T X^T + \underline{w}^T\right)\right] \\ &= E\left[\underline{\alpha}^T X^T X\underline{\alpha}\underline{\alpha}^T X^T + \underline{w}^T \underline{w}\underline{\alpha}^T X^T + \underline{\alpha}^T X^T \underline{w}\underline{w}^T + \underline{w}^T X\underline{\alpha}\underline{w}^T\right] \\ &= \underline{\alpha}^T X^T X\underline{\alpha}\underline{\alpha}^T X^T + E\left[\underline{w}^T \underline{w}\right]\underline{\alpha}^T X^T + \underline{\alpha}^T X^T E\left[\underline{w}\underline{w}^T\right] + \underline{\alpha}^T X^T E\left[\underline{w}\underline{w}^T\right] \\ &= \underline{\alpha}^T X^T X\underline{\alpha}\underline{\alpha}^T X^T + N\sigma^2 \underline{\alpha}^T X^T + \sigma^2 \underline{\alpha}^T X^T + \sigma^2 \underline{\alpha}^T X^T \\ & \underline{\alpha}^T A\underline{\alpha}\underline{\alpha}^T X^T + (N+2)\sigma^2 \underline{\alpha}^T X^T, \end{aligned} \tag{C.14}$$

which is the result in Eqn. (3.58).

5. Derivation of Eqn. (3.60) :

From Eqn. (3.59),

$$\begin{aligned} G_{33}(\underline{\theta}) &= \left\{-\frac{N}{2\sigma^2} + \frac{1}{2\left(\sigma^2\right)^2}\left[\underline{\alpha}^T A\underline{\alpha} - 2\underline{g}^T \underline{\alpha} + \rho\right]\right\}^2 \\ &= \frac{1}{4\left(\sigma^2\right)^4}\left(\underline{\alpha}^T A\underline{\alpha} - 2\underline{g}^T \underline{\alpha} + \rho - N\sigma^2\right)^2. \end{aligned} \tag{C.15}$$

Then,

$$\begin{aligned} & 4\left(\sigma^2\right)^4 G_{33}(\underline{\theta}) \\ = \; & \underline{\alpha}^T A\underline{\alpha}\underline{\alpha}^T A\underline{\alpha} - 4\underline{\alpha}^T A\underline{\alpha}\underline{\alpha}^T A\underline{\alpha} + 4\underline{\alpha}^T \underline{g}\underline{g}^T \underline{\alpha} + 2\underline{\alpha}^T A\underline{\alpha}\left(\underline{\alpha}^T A\underline{\alpha} + N\sigma^2\right) - 2N\sigma^2 \underline{\alpha}^T A\underline{\alpha} \end{aligned}$$

$$-4\underline{\alpha}^T\underline{g}\rho + 4N\sigma^2\underline{\alpha}^T A\underline{\alpha} + \rho^2 - 2N\sigma^2\left(\underline{\alpha}^T A\underline{\alpha} + N\sigma^2\right) + N^2\sigma^4$$

$$= -\underline{\alpha}^T A\underline{\alpha}\underline{\alpha}^T A\underline{\alpha} + 4\underline{\alpha}^T\underline{g}\underline{g}^T\underline{\alpha} - 4\underline{\alpha}^T\underline{g}\rho + 2N\sigma^2\underline{\alpha}^T A\underline{\alpha} + \rho^2 - N^2\sigma^4,$$

leading to

$$4\left(\sigma^2\right)^4 E\left[G_{33}\left(\underline{\theta}\right)\right]$$
$$= -\underline{\alpha}^T A\underline{\alpha}\underline{\alpha}^T A\underline{\alpha} + 4\underline{\alpha}^T E\left[\underline{g}\underline{g}^T\right]\underline{\alpha} - 4\underline{\alpha}^T E\left[\underline{g}\rho\right] + 2N\sigma^2\underline{\alpha}^T A\underline{\alpha} + E\left[\rho^2\right] - N^2\sigma^4. \quad \text{(C.16)}$$

But,

$$E\left[\rho^2\right] = E\left[\underline{s}^T\underline{s}\underline{s}^T\underline{s}\right]$$
$$= E\left[\left(\underline{\alpha}^T X^T + \underline{w}^T\right)\left(X\underline{\alpha} + \underline{w}\right)\left(\underline{\alpha}^T X^T + \underline{w}^T\right)\left(X\underline{\alpha} + \underline{w}\right)\right] \quad \text{(C.17)}$$
$$= E\left[\underline{\alpha}^T A\underline{\alpha}\underline{\alpha}^T A\underline{\alpha} + 4\underline{w}^T X\underline{\alpha}\underline{w}^T X\underline{\alpha} + \underline{w}^T\underline{w}\underline{w}^T\underline{w} + 4\underline{\alpha}^T A\underline{\alpha}\underline{w}^T X\underline{\alpha}\right.$$
$$\left. + 2\underline{\alpha}^T A\underline{\alpha}\underline{w}^T\underline{w} + 4\underline{w}^T X\underline{\alpha}\underline{w}^T\underline{w}\right]$$
$$= \underline{\alpha}^T A\underline{\alpha}\underline{\alpha}^T A\underline{\alpha} + 4\underline{\alpha}^T X^T E\left[\underline{w}\underline{w}^T\right]X\underline{\alpha} + E\left[\underline{w}^T\underline{w}\underline{w}^T\underline{w}\right]$$
$$+ 2N\sigma^2\underline{\alpha}^T A\underline{\alpha} + 4\underline{\alpha}^T X^T E\left[\underline{w}\underline{w}^T\underline{w}\right]$$
$$= \underline{\alpha}^T A\underline{\alpha}\underline{\alpha}^T A\underline{\alpha} + 4\sigma^2\underline{\alpha}^T X^T X\underline{\alpha} + (N^2 + 2N)\sigma^4 + 2N\sigma^2\underline{\alpha}^T A\underline{\alpha}$$
$$= \underline{\alpha}^T A\underline{\alpha}\underline{\alpha}^T A\underline{\alpha} + (4 + 2N)\sigma^2\underline{\alpha}^T A\underline{\alpha} + (N^2 + 2N)\sigma^4. \quad \text{(C.18)}$$

Substituting Eqn. (C.18) together with Eqns. (C.9) and (C.12) into Eqn. (C.16) leads to

$$4\left(\sigma^2\right)^4 E\left[G_{33}\left(\underline{\theta}\right)\right]$$
$$= -\underline{\alpha}^T A\underline{\alpha}\underline{\alpha}^T A\underline{\alpha} + 4\underline{\alpha}^T\left[A\underline{\alpha}\underline{\alpha}^T A + \sigma^2 A\right]\underline{\alpha} - 4\underline{\alpha}^T\left[A\underline{\alpha}\underline{\alpha}^T A\underline{\alpha} + (N+2)\sigma^2 A\underline{\alpha}\right]$$
$$+ 2N\sigma^2\underline{\alpha}^T A\underline{\alpha} + \left[\underline{\alpha}^T A\underline{\alpha}\underline{\alpha}^T A\underline{\alpha} + (4 + 2N)\sigma^2\underline{\alpha}^T A\underline{\alpha} + (N^2 + 2N)\sigma^4\right] - N^2\sigma^4$$
$$= 4\sigma^2\underline{\alpha}^T A\underline{\alpha} - 4(N+2)\sigma^2\underline{\alpha}^T A\underline{\alpha} + 2N\sigma^2\underline{\alpha}^T A\underline{\alpha} + (4 + 2N)\sigma^2\underline{\alpha}^T A\underline{\alpha} + 2N\sigma^4$$
$$= 2N\left(\sigma^2\right)^2, \quad \text{(C.19)}$$

which is the result in Eqn. (3.60).

Bibliography

[1] J. G. Proakis, *Digital Communications.* New York: McGraw-Hill, 1983.

[2] M. I. Skolnik, *Introduction to Radar Systems.* New York: MaGraw-Hill, 1986.

[3] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals.* Englewood Cliffs, N. J.: Prentice-Hall, 1978.

[4] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory,* vol. 1. Upper Sadle River, New Jersey: Prentice Hall, 1993.

[5] R. J. McAulay and E. M. Hofstetter, "Barakin bounds on parameter estimation.," *IEEE Transactions on Information Theory,* vol. 17, pp. 669–679, November 1971.

[6] S. M. Kendall and A. Stuart, *The Advanced Theory of Statistics.* Macmillan, New York, 1979.

[7] L. P. Seidman, "Performance limitations and error calculations for parameter estimation," *Proceedings of the IEEE,* vol. 58, p. 644, May 1970.

[8] J. Ziv and M. Zakai, "Some lower bounds on signal parameter estimation," *IEEE Trans. on Info. Theo.*, vol. IT-15, no. 3, pp. 386–391, May 1969.

[9] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation and Time Series Analysis.* Massachusetts: Addison-Wesley, 1991.

[10] D. C. Rife and R. R. Boorstyn, "Single-tone parameter estimation from discrete-time observations," *IEEE Trans. on Information Theo.*, vol. IT-20, pp. 591–598, Sept. 1974.

[11] S. M. Kay, "A fast and accurate single frequency estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1987–1990, Dec. 1989.

[12] D. Kim, M. J. Narasimha, and D. C. Cox, "An improved single frequency estimator," *IEEE Signal Process. Letter*, vol. 3, pp. 212–214, 1996.

[13] T. Brown and M. Wang, "An iterative algorithm for single-frequency estimation," *IEEE Trans. Signal Process.*, vol. 50, pp. 2671–2682, 2002.

[14] V. F. Pisarenko, "The retrieval of harmonics by linear prediction," *Geophys. J. R. Astron. Soc.*, vol. 33, pp. 347–366, 1973.

[15] R. F. Barrett and D. R. A. McMahon, "ML estimation of the fundamental frequency of a harmonic series," *the International Symposium of Signal Processing and its Applications*, 1987.

[16] B. James, B. D. O. Anderson, and R. C. Williamson, "Conditional mean and maximum likelihood approaches to multiharmonic frequency estimation," *IEEE Transactions on Signal Processing*, vol. 42, pp. 1366 – 1375, June 1994.

[17] H. Li, P. Stoica, and J. Li, "Computationally efficient parameter estimation for harmonic sinusoidal signals," *Signal Process.*, vol. 80, pp. 1937–1944, 2000.

[18] K. W. Chan and I. H. C. So, Member, "Accurate frequency estimation for real harmonic sinusoids," *IEEE SIGNAL PROCESSING LETTERS*, vol. 11, July 2004.

[19] E. J. Hannan, "The estimation of frequency," *Journal of Applied Probability*, vol. 10(3), pp. 510–519, 1973.

[20] D. C. Rife and R. R. Boorstyn, "Multiple tone parameter estimation from discrete time observations," *Bell Syst. Tech. J.*, vol. 55, pp. 1389–1410, Nov. 1976.

[21] D. W. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood," *Proc. IEEE*, vol. 70, pp. 975–989, Sept 1982.

[22] S. A. Tretter, "Estimating the frequency of a noisy sinusoid by linear regression," *IEEE Trans. on Information Theory*, vol. IT-31, pp. 832–835, Nov. 1985.

[23] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing.* Englewood Cliffs, New Jersey: Prentice Hall, 1989.

[24] J. W. B. Davenport and W. L. Root, *An Introduction to the Theory of Random Signals and Noise.* McGraw-Hill Book Company, 1958.

[25] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products.* Boston, Massachusetts: Academic Press, fifth ed., 1994. Equation ETI 146(31)a on p. 382.

[26] B. G. Quinn, I. Vaughan, L. Clarkson, and P. J. Kootsookos, "Comments on 'On the performance of the weighted linear predictor frequency estimator'," *IEEE Trans. Signal Processing*, vol. 46, no. 2, pp. 526–527, 1998.

[27] P. Handel, "On the performance of the weighted linear predictor frequency estimator," *IEEE Trans. Signal Processing*, vol. 43, pp. 3070–3071, Dec. 1995.

[28] V. Clarkson, P. J. Kootsookos, and B. G. Quinn, "Analysis of the variance threshold of Kay's weighted linear predictor frequency estimator," *IEEE Trans. Signal Processing*, vol. 42, pp. 2370–2379, Sept. 1994.

[29] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, pp. 276–280, March 1985.

[30] A. Nehorai and B. Porat, "Adaptive comb filtering for harmonic signal enhancement," *IEEE Trans. Acoust. Speech Signal Processing*, vol. ASSP-34, pp. 1124–1 138, Oct. 1986.

[31] C. Zarowski and I. Kropyvnytskyy, "Cramer-Rao bounds for circadian rhythm parameter estimation," *IEEE CCECE 2002*, vol. 2, pp. 1083–1086, May 2002.

[32] R. T. Behrens and L. L. Scharf, "Signal processing applications of oblique projection operators," *IEEE Trans. on Signal Proc.*, vol. 42, pp. 1413–1424, 1994.

[33] A. O. Hero, J. A. Fessler, and M. Usman, "Exploring estimator bias-variance tradeoffs using the uniform CR bound," *IEEE Trans. on Signal Proc.*, vol. 44, pp. 2026–2041, Aug. 1996.

[34] Z. Chen and A. O. Hero *Personal Communication (Email records)*, Feb. 2005.

[35] A. O. Hero *http://www.eecs.umich.edu/~hero.*

[36] H. L. V. Trees, *Detection, Estimation, and Modulation Theory: Part I.* New York: John Wiley, 1968.

[37] M. Usman, A. Hero, and J. A. Fessler, "Bias-variance tradeoffs analysis using uniform CR bound for images," *Proceedings of the 1994 IEEE Conference on Image Processing , Austin*, pp. 835–839, Nov. 1994.

[38] M. Usman, A. Hero, and J. A. Fessler, "Uniform CR bound: implementation issues and applications to image reconstruction," *Proceedings of the IEEE 1994 Nuclear Science Symposium and Medical Imaging Conference, Norfolk*, Oct. 1994.

[39] M. K. Simon and A. Mileant, "SNR estimation for the baseband assembly," *Jet Propulsion Lab., Pasadena, CA, Telecommunications and Data Acquisition Prog. Rep. 42-85*, May 15, 1986.

[40] B. Shah and S. Hinedi, "The split symbol moments SNR estimator in narrow-band channels," *IEEE Trans. Aerospace and Electronic Systems*, vol. 25, no. 5, pp. 737–747, 1990.

[41] R. M. Gagliardi and C. M. Thomas, "PCM data reliability monitoring through estimation of signal-to-noise ratio," *IEEE Trans. Commun.*, vol. COM-16, pp. 479–486, June 1968.

[42] T. R. Benedict and T. T. Soong, "The joint estimation of signal and noise from the sum envelope," *IEEE Trans. Inform. Theory*, vol. IT-13, no. 3, pp. 447–454, 1967.

[43] W. Kizner, "A numerical method for finding solutions of nonlinear equations," *J. SIAM*, vol. 12, pp. 424–428, June 1964.

[44] F. B. Hildebrand, *Introduction to Numerical Analysis.* New York: McGraw-Hill, 1956.

[45] eFund Engineering Fundamentals, "The modified Bessel's differential equation," *http://www.efunda.com/math/bessel/modifiedbessel.cfm.*

[46] E. H. Inselmann, "Hypothesis testing of Gaussian processes with composite alternatives," *J. SIAM*, vol. 12, pp. 370–385, June 1964.

[47] E. W. Weisstein, "Confluent hypergeometric function of the first kind," *MathWorld – A Wolfram Web Resource.* http://mathworld.wolfram.com/Confluent HypergeometricFunctionoftheFirstKind.html.

[48] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* New York: Dover, 9 ed., 1972.

[49] K. E. Muller, "Computing the confluent hypergeometric function, M(a,b,x)," *Published online May 4, 2001, Springer-Verlag 2001.* http://www.bios.unc.edu /~muller/2001.pdf.

[50] Z. Chen, B. Nowrouzian, and C. Zarowski, "An investigation of SNR estimation techniques based on uniform Cramer-Rao lower bound," *IEEE International Midwest Symposium on Circuits and Systems (MWSCAS 2005) USA*, 2005.

[51] Z. Chen, B. Nowrouzian, and C. Zarowski, "Cramer-Rao lower bound for harmonic and subharmonic estimation," *The 30th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2006), Toulouse, France*, 2006.

[52] D. R. Pauluzzi and N. C. Beaulieu, "A comparison of SNR estimation techniques in the AWGN channel," *IEEE Symposium, Victoria, Canada*, May 1995.

[53] S. Bellini and G. Tartara, "Bounds on error in signal parameter estimation," *IEEE Trans. on Communications*, vol. COM-22, pp. 340–342, March 1974.

[54] S. Bellini and G. Tartara, "Correction to 'Bounds on error in signal parameter estimation'," *IEEE Trans. on Communications*, vol. COM-23, p. 486, April 1975.