First order bias and second order variance of the Maximum Likelihood Estimator with application to multivariate Gaussian data and time delay and Doppler shift estimation

by

Eran Naftali

B.Sc., Electrical Engineering (1993)
Technion – Israel Institute of Technology

Submitted to the Department of Ocean Engineering
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Ocean Engineering
at the
Massachusetts Institute of Technology

February 2000

© 2000 Eran Naftali
All rights reserved

The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of the thesis document in the whole or in part

Signature of author..............................................

Department of Ocean Engineering
January 25, 2000

Certified by..........................................................

Professor Nicholas C. Makris
Associate Professor of Ocean Engineering
Thesis Supervisor

Accepted by.....................................................

Professor Nicholas Patrikalakis
Kawasaki Professor of Engineering
Chairman, Departmental Committee on Graduate Studies
First order bias and second order variance of the Maximum Likelihood Estimator with application to multivariate Gaussian data and time delay and Doppler shift estimation

by

Eran Naftali

Submitted to the Department of Ocean Engineering
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Ocean Engineering

Abstract

In many practical problems the relationship between the data and the parameters to be estimated is nonlinear. This implies that the estimate may be biased and may not attain the minimum variance possible. The maximum likelihood estimator (MLE) is one of the most widely used estimators because if an asymptotically unbiased and minimum variance estimator exists for large sample sizes it is guaranteed to be the maximum likelihood estimator. Since exact expressions for the maximum likelihood estimator’s bias and covariance are often difficult or impractical to find, it has become popular in recent years to compute limiting bounds, since these bounds are usually much easier to obtain. A well known bound is the Cramer - Rao Bound, which can be interpreted as the first order term of the covariance of the maximum likelihood estimator, when the covariance is expended in inverse orders of sample size or Signal-to-Noise ratio (SNR).

When dealing with a small number of samples, or similarly, a small integration time or a low SNR, higher order terms can significantly contribute to the covariance, thus making limiting bounds such as the Cramer - Rao Bound a poor approximation to the true covariance. By applying higher order asymptotics and using tensor analysis, tighter bounds for the covariance and bias of the MLE can be obtained.

Expressions for both the first order bias and the second order covariance term of a general maximum likelihood estimator are presented. The first order bias is then evaluated for multivariate Gaussian data while the second order covariance is evaluated for the two special cases of multivariate Gaussian data that are of great practical significance in signal processing. The first is where the data covariance matrix is independent of the parameters to be estimated, as in standard signal in additive noise scenarios, and the second is where the data mean is zero, as in scenarios where the signal is fully randomized. These expressions are then applied to estimate the variance in time-delay and Doppler-shift estimation, where it is rigorously shown that the Cramer - Rao bound is an unrealistically optimistic estimate of the true covariance when the SNR is below 20 dB, for Gaussian, LFM and HFM signals.

Thesis Supervisor: Professor Nicholas C. Makris
Title: Associate Professor of Ocean Engineering
Acknowledgements

Throughout the last two years I have received enormous support from several individuals and organizations that I would like to specifically thank.

I would like to express my sincere appreciation to my advisor, Associate Professor Nicholas C. Makris, for his excellent guidance throughout my studies. He was always ready to dedicate time to me, and I am grateful for the opportunity he offered me to work on my research topic. His understanding and generosity toward his students are greatly appreciated.

In addition, I would like to thank the Israeli Navy for giving me the opportunity to study at MIT and awarding me a scholarship, especially Rear Admiral Shlomo Markel and Commander Arie Cohen. In addition, I would like to thank Professor Shlomo Maital, Professor Diana McCammon and Professor Israel Bar-David for the recommendations they gave me to study at MIT. Without their intensive efforts and help I would not have been studying here.

I would like to express my deepest regards to the Ocean Engineering department, and especially Professor Arthur B. Baggeroer who was willing to share his knowledge and experience with me. In addition, I would like to thank Michele Zanolin and Aaron Thode for the fruitful discussions we had over many issues in the thesis. The Teaching Assistants Kelli Hendrickson, Yanwu Zhang and Ben Reeder who guided me through my classes, and fellow graduate students Sam Geiger and David Prosper, were always helpful when I needed their advice. Furthermore, I would like to thank Jean Sucharewicz, Beth Tuths and Sabina Rataj for all the administrative work they have done for me.

Last and certainly not least, I would like to thank my parents and my lovely family. Without their help and support during the past one and a half years, I would not have succeeded. Thank you from all my heart!
Dedicated to my wife Michal
and to my sons Nadav and Amit
# Table of Contents

1. **Introduction** ................................................................................................................. 7
   1.1. Background .................................................................................................................. 7
   1.2. Thesis Outline .............................................................................................................. 9
2. **Estimation** ..................................................................................................................... 10
   2.1. Bias, Covariance Matrix and Error Correlation Matrix of an Estimator ............... 11
      2.1.1. Bias ..................................................................................................................... 11
      2.1.2. Covariance Matrix .............................................................................................. 11
      2.1.3. Error Correlation Matrix ................................................................................... 12
   2.2. Minimum Variance Unbiased Estimator ................................................................... 12
   2.3. Maximum Likelihood Estimator ............................................................................... 12
      2.3.1. The Asymptotic Property of the Maximum Likelihood Estimator ..................... 14
      2.3.2. The Invariance Property of the Maximum Likelihood Estimator ...................... 15
3. **The Asymptotic Expansion** ......................................................................................... 16
   3.1. Review .......................................................................................................................... 16
   3.2. Definitions and Common Equalities ......................................................................... 17
   3.3. The Asymptotic Expansion Formula ......................................................................... 18
      3.3.1. The First Order Asymptotic Bias ...................................................................... 19
      3.3.2. The Second Order Asymptotic Error .................................................................. 20
4. **The Asymptotic Expansion for the Gaussian Case** ...................................................... 24
   4.1. The Likelihood Quantities for the Gaussian Case ......................................................... 24
      4.1.1. The General Multivariate Gaussian Case ............................................................ 25
      4.1.2. The Case where the Covariance is Parameter-Independent ................................. 28
      4.1.3. The Case where the Mean is Zero and the Covariance is Parameter Dependent ....... 34
   4.2. A Summary of Results ............................................................................................... 40
5. **Applications of the Asymptotic Expansion – Time Delay and Doppler Shift Estimation** 41
   5.1. Real and Complex Waveform Representation ......................................................... 41
      5.1.1. Real Signal .......................................................................................................... 41
      5.1.2. Complex Representation of Real Signals ............................................................. 42
   5.2. The Matched Filter ..................................................................................................... 43
   5.3. The Likelihood Quantities for a Continuous Gaussian Data where the Covariance is Parameter-Independent .......................................................... 47
      5.3.1. The Joint Moments of the Log – Likelihood Derivatives ..................................... 48
      5.3.2. Time Delay Estimation ....................................................................................... 54
      5.3.3. Doppler Shift Estimation .................................................................................... 59
   5.4. First and Second Order Terms of the Variance for Three Specific Waveforms .......... 65
6. **Summary and Conclusions** ......................................................................................... 90

*Appendix A – Matlab Scripts* ............................................................................................... 92
1. Introduction

1.1. Background

In recent years, a wide variety of acoustic techniques have been developed to probe the marine environment. These techniques typically require the nonlinear inversion of acoustic field data measured on a hydrophone array. The data, however, are often randomized by the addition of naturally occurring ambient noise or by fluctuations in the waveguide's refractive index and boundaries. The nonlinear inversion of random data often leads to estimates with biases and variances that are difficult to quantify analytically. It has become popular in recent years to (1) simply assume that biases are negligible and to (2) compute limiting bounds on the variance of these nonlinear estimators, since these bounds are usually much easier to obtain than the actual variance. A major problem, however, is that the estimators may be strongly biased and that the bounds are only guaranteed to converge to the true variance of the estimator under sufficiently high signal-to-noise ratio (SNR).

If an estimator for a particular parameter, for large sample sizes or high signal to noise, is asymptotically optimal, which means that it becomes unbiased and attains the minimum variance possible, which is known as the Cramer-Rao Bound, it is guaranteed to be the maximum likelihood estimator. This fact is useful because the maximum likelihood estimate can be obtained simply by maximizing the likelihood function with respect to the parameter to be estimated. The likelihood function is simply the probability distribution function for the data, evaluated at the measured values of the data, given the parameters to be estimated which are unknown variables to be determined in the maximization. The linear least squares estimator, which for example is used extensively in ocean-acoustic inversions, is the same as the maximum likelihood estimator only when the data and parameter vectors are linearly related and the data are uncorrelated, have the same variance, follow a multivariate Gaussian distribution, and the parameters to be estimated depend only on the expected value of the data and not the covariance of the data. These assumptions are typically not satisfied in ocean-acoustic inverse problems
even in large sample sizes and high signal-to-noise ratio, so the maximum likelihood approach is preferable since it will at least be asymptotically optimal. In ocean acoustics, the data and parameters to be inverted from the data typically share a complicated nonlinear relationship that often can only be described by numerical modeling using sophisticated wave propagation software. It then becomes impractical to exactly determine the variance and bias of ocean-acoustic parameter estimates analytically or even numerically since extensive computations must be made for every particular case necessary to describe parameter and data.

By applying higher order asymptotic theory, general expressions for the first-order-bias and second-order-covariance of a general maximum likelihood estimate can be derived. These two results can then be applied in a variety of inverse problems, some specific problems that come to mind include

1) ocean acoustic matched field processing, matched field inversion for oceanographic and geophysical parameters, active sonar range and Doppler estimates and target classification,
2) radar range and Doppler estimation, radar imaging,
3) digital image processing,
4) pattern recognition,
5) astrophysical gravity wave detection,
6) genetic protein unfolding,
7) financial analysis of stock rate behavior,

and all natural sciences, wherever the Maximum Likelihood Estimator is being used.

In this thesis expressions for the first-order-bias and second-order-covariance of a general maximum likelihood estimate are presented. The bias expression is then evaluated for general multivariate Gaussian data with repetitive measurements, where both the covariance matrix and the mean depend on the parameters to be estimated. The covariance of the parameter estimate is evaluated for two specific cases. In the first case, the data covariance matrix is independent of the parameters, and in the second the data mean vector is zero. Applying these results to continuous Gaussian data that depend on a single scalar parameter that modifies the mean only, expressions for the bias and variance
of the estimate are obtained. These are then evaluated for both time delay and Doppler shift estimation.

1.2. Thesis Outline

Chapter 2 introduces the estimation problem, provides definitions for the bias, covariance matrix and error correlation matrix of a vector of estimators, and presents the minimum variance unbiased estimator (MVU) concept and the maximum likelihood estimator (MLE).

In Chapter 3 the asymptotic expansion of the log - likelihood function is presented, and expressions for the bias to first order and the covariance matrix and error correlation matrix to second order are derived.

In Chapter 4 the bias to first order is evaluated for general multivariate Gaussian data, where both the data mean vector and the data covariance matrix are parameter-dependent. Both the parameter-vector estimator error correlation matrix and covariance matrix are then evaluated for two special cases. In the first case the Gaussian data has a covariance matrix that is parameter-independent, and in the second the data is zero-mean.

In Chapter 5 applications of the results obtained in Chapter 4 are applied to the MLE of the time delay (known as the matched filter) and to the MLE of the Doppler shift, followed by implementation for three different signals.

Chapter 6 contains the summary and conclusion of the thesis.
2. Estimation

Estimation is the heart of all the experimental sciences and engineering, as it is used to extract information from observed data.

Let us assume that we have a set of \( n \) data observations \([x_1, x_2, \ldots, x_n]\), comprising a vector \( \mathbf{X} \) (where bold notation is used to denote a vector here and elsewhere). In addition, let us assume the data depend on a set of \( m \) unknown parameters \([\theta_1, \theta_2, \ldots, \theta_m]\) comprising a vector \( \mathbf{\theta} \). We define \( \hat{\mathbf{\theta}}(\mathbf{X}) \) as the parameter vector estimator of \( \mathbf{\theta} \).

The first step in determining the performance of the estimator is modeling the data. Since the data is inherently random, we describe the model by its probability density function (PDF). The PDF is parameterized by the \( m \) unknown parameters \([\theta_1, \theta_2, \ldots, \theta_m]\), i.e., we have a class of PDF’s where each one is different due to a different value of \( \mathbf{\theta} \). In an actual problem we are not given a PDF but must choose one that is not only consistent with the problem constraints and any prior knowledge, but one that is also mathematically tractable. Once the PDF has been specified, the problem becomes one of determining an optimal estimator, as a function of the data. It is clear that the performance of the estimator will be critically dependent upon the assumption we make about the PDF.

There are two approaches used to solve the estimation problem:

a) The *Bayesian* approach

b) The *classical* approach

The *Bayesian* approach assumes the parameter of interest is random, therefore knowledge of its prior density is required for the estimation. The *classical* approach assumes the parameter of interest is deterministic (nonrandom), therefore the only thing known about the parameter is the range of values it can have.

Often, we don’t have sufficient information about the parameter’s a priori density, and we tend to assume that it is uniformly distributed. However, this assumption is misleading and might cause major errors. Here is an example. Let us assume that the parameter to be estimated is the mean of a given PDF. Since we have no prior information on the PDF of the mean, we tend to believe it is uniformly distributed.
between two values. Now, let us assume we want to estimate the square of the mean. Assuming again a uniform distribution, now on the squared mean, leads to an extremely different distribution of the mean than the uniform one we assumed before, hence causing a conflict. This example shows why we are often enforced to go to the classical approach and assume no prior knowledge on the parameter’s density is available.

Next, we define tools to evaluate the performance of an estimator.

2.1. **Bias, Covariance Matrix and Error Correlation Matrix of an Estimator**

The bias, variance and the mean-square error are useful tools to evaluate the performance of an estimator. When dealing with more than one estimator, or an estimator vector, these tools become bias, covariance matrix and error correlation matrix, respectively. We should note that both matrices are symmetric where the diagonal represents each estimated parameter's variance and mean-square error, respectively, while the off diagonal terms represent the cross terms covariance and correlation, respectively.

2.1.1. **Bias**

The bias of an estimator vector is defined as

\[
b(\hat{\theta}) = E[\hat{\theta} - \theta] = E[\hat{\theta}] - \theta.
\]

(2.1)

Here and elsewhere \(E[\bullet]\) denotes the expectation operator.

When the bias of an estimator is identically zero the estimator is *unbiased*, otherwise, the estimator is *biased*.

2.1.2. **Covariance Matrix**

The covariance matrix that gives us a measure of the spread of the error around the expected value of the estimator, is defined as

\[
cov(\hat{\theta}) = E[(\hat{\theta} - E[\hat{\theta}])(\hat{\theta} - E[\hat{\theta}])^T] = E[\hat{\theta}\hat{\theta}^T] - E[\hat{\theta}]E[\hat{\theta}]^T.
\]

(2.2)

When \(\theta\) is a scalar (i.e, we estimate only one parameter) the covariance matrix is reduced to a scalar called variance.
2.1.3. Error Correlation Matrix

Often, we are interested in the spread of the error about the true value of the parameter, rather than mean of the estimator. This spread about $\theta$ is referred to as the error correlation matrix (ECM) and is defined as

$$\text{ECM}(\hat{\theta}) = E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T].$$  

(2.3)

When $\theta$ is a scalar (i.e., we estimate only one parameter) the error correlation matrix is reduced to a scalar called mean-square error (MSE).

For an unbiased estimator, the covariance matrix and the error correlation matrix converge to the same value.

2.2. Minimum Variance Unbiased Estimator

For an estimator to be unbiased, we mean that on the average the estimator will yield the true value of the unknown parameter, or

$$E[\hat{\theta}] = \theta.$$  

(2.4)

In searching for optimal estimators we need to adopt some optimally criterion. A practical and commonly used approach is to restrict our search to unbiased estimators, and find the one, which minimizes the variance. Such an estimator is termed the minimum variance unbiased (MVU) estimator. The question arises as to whether a MVU estimator exists, i.e., an unbiased estimator with minimum variance for all $\theta$. In general, the MVU estimator does not always exists, and even if it does, we might not be able to find it, since there is no known procedure to produce it in general.\(^1\)

2.3. Maximum Likelihood Estimator

When the MVU estimator does not exist, or cannot be found even if it does exist, there is a powerful alternative estimator. This estimator, based on the maximum likelihood principle, is overwhelmingly the most popular approach to obtaining practical estimates. It has the distinct advantage of being derived by a well-defined procedure, allowing it to be implemented for complicated estimation problems. Additionally, for

\(^1\) Kay, Pp. 15-21.
most cases of practical interest its performance is optimal, in terms of attaining the
minimum variance, for large enough data records. More specifically, if an asymptotically
minimum variance unbiased estimator exists it is guaranteed to be the maximum
likelihood one.

We will now define the procedure in which we derive the maximum likelihood
estimator and then two important properties of this estimator will be presented.

If \( p(X;\theta) \) is the PDF of a vector \( X \) given a parameter vector \( \theta \), the MLE for a
vector parameter, \( \theta \), is defined as the one that maximizes the PDF over the allowable
range of \( \theta \)

\[
\frac{dp(X;\theta)}{d\theta} \bigg|_{\theta=\hat{\theta}} = 0. \quad (2.5)
\]

It is important to note that maximizing the PDF and maximizing the logarithm of the PDF
(known as the log – likelihood function) are equivalent since the logarithmic function is
monotonic. Therefore, the MLE can be equivalently defined as

\[
\frac{d \ln(p(X;\theta))}{d\theta} \bigg|_{\theta=\hat{\theta}} = 0. \quad (2.6)
\]

**Example 2.1**

Let us assume \( X \) is a vector of \( n \) independent identically distributed Gaussian
random variables, \( x_1, \ldots, x_n \), each with mean \( \mu \) and variance \( \sigma^2 \). Assuming \( \theta \) is a scalar
parameter that modifies the mean, the PDF of the vector \( X \) is hence

\[
p(X;\theta) = \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu(\theta))^2}.
\quad (2.7)
\]

If the parameter of interest is the mean, maximizing the log – likelihood function, we
obtain the MLE for \( \theta \),

\[
\hat{\theta} = \hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i
\quad (2.8)
\]

which is just average of the data.

\footnote{Kay, p. 157.}
Substituting Eq. (2.8) into the bias and the variance expressions, Eqs. (2.1)- (2.3), respectively, we obtain a zero bias and a \( \frac{\sigma^2}{n} \) variance and MSE for this estimator, therefore the estimator is unbiased and its variance goes to zero as \( n \) increases.

Next, we will present two important properties of the MLE.

2.3.1. The Asymptotic Property of the Maximum Likelihood Estimator

It can be shown that the MLE of the unknown parameter \( \theta \) is asymptotically distributed (for large data records) according to

\[
\hat{\theta} \sim N\left(\theta, I^{-1}(\theta)\right)
\]

where \( I(\theta) \) is the Fisher information, defined as

\[
[I(\theta)]_{ij} = -E\left[ \frac{\partial^2 \ln(P(X;\theta))}{\partial \theta_i \partial \theta_j} \right]
\]

and its inverse, known as the Cramer Rao Lower Bound (CRLB). For a scalar parameter case, the Cramer Rao Lower Bound provides the minimum possible variance. The only conditions that have to be satisfied require the existence of log-likelihood derivatives, a condition on the first log-likelihood derivative, \( E\left[ \frac{\partial \ln(P(X;\theta))}{\partial \theta_i} \right] = 0 \), as well as the Fisher information being nonzero\(^3\).

This result is quite general and forms the basis for claiming optimality of the MLE, since it is seen to be asymptotically unbiased and asymptotically attains the CRLB. However, in practice, the key question is always how large the number of data samples should be in for the asymptotic properties to apply, a question we will answer in the following sections by using the asymptotic expansion of the likelihood function.

\(^3\) Kay, p. 167.
The Invariance Property of the Maximum Likelihood Estimator

The MLE of the parameter $a = g(\theta)$, where the PDF $p(X; \theta)$ is parameterized by $\theta$, is given by $\hat{a} = g(\hat{\theta})$, where $\hat{\theta}$ is the MLE of $\theta$, the parameter that parameterizes the PDF.4

Example 2.2

Continuing example 2.1, we now want to find the MLE for $a=\theta^2$. Using Eq. (2.7) and the invariance property of the MLE, we can easily determine the MLE for $a$ by

$$\hat{a} = (\hat{\mu})^2 = \left(\frac{1}{n} \sum_{i=1}^{n} x_i\right)^2 = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j. \quad (2.11)$$

4 Kay, p. 176.
3. The Asymptotic Expansion

3.1. Review

The use in statistical theory of approximate arguments based on such methods as local linearization and approximate normality has a long history. These ideas play at least a couple of roles.

a. They may give simple approximate answers to distributional problems where an exact solution is known in principle but difficult to implement.

b. They yield higher order expansions from which the accuracy of simple approximations may be assessed and, where necessary, improved.

The approximate arguments are developed by supposing that some quantity, often, the sample size, but generally the amount of observed data, becomes large. It must be stressed this is a technical device for generating approximations whose adequacy always needs to be assessed.

As mentioned before, some of the distributional problems in statistics have exact solutions that can be derived analytically. Many problems of practical interest, however, either have no exact solution or have ones, which are too complicated for direct use. Such situations are usually handled by the so-called asymptotic or large-sample theory. That is, we derive approximations supposing some quantity \( n \), usually the sample size, but more generally the amount of the observed data, is large.

For example, the limit laws of probability theory may show that as \( n \rightarrow \infty \) a distribution of interest approaches some simple standard distribution, particularly the normal distribution associated with the central limit theorem.\(^5\)

3.2. Definitions and Common Equalities

The definitions below follow closely the Barndorff–Nielsen definitions\(^6\).

Given a probability density function (PDF), \( p(X; \theta) \), the log-likelihood function, \( l(\theta) \), is defined as \( l(\theta) = \ln(p(X; \theta)) \). We assume that \( \theta \) is an \( m \) -dimensional parameter vector, and its generic coordinates are denoted by \( \theta^r \), \( \theta^s \), etc. We define the log-likelihood derivative as \( l_r = \frac{\partial l(\theta)}{\partial \theta^r} \). We use the letter \( \nu \) to indicate joint moments of the log-likelihood derivatives. More specifically, if \( R_1 = r_{n_1} \ldots r_{n_l} \), \( R_m = r_{m_1} \ldots r_{m_m} \) are sets of coordinate indices we define

\[
\nu_{R_1, \ldots, R_m} = E[l_{R_1} \ldots l_{R_m}] . \tag{3.1}
\]

**Example 3.1**

Let us assume \( R_1 = r_s \) and \( R_2 = t \). Using the definition of \( \nu \) we can see that

\[
\nu_{R_1} = \nu_{rs} = E[l_{rs}] \quad \text{and} \quad \nu_{R_1, R_2} = \nu_{rs, t} = E[l_{rs}l_{st}] . \tag{3.2}
\]

The expected information, known as the Fisher information, is denoted by the notation \( i_{rs} \), where the indices \( r, s \) are arbitrary, and is defined by \( i_{rs} = E[l_{rs}] \).

Lifting the indices produce quantities which are denoted as follows

\[
i^{rs} = [i]^{-1}_{rs} \tag{3.2}
\]

i.e., the inverse of the expected information, and

\[
\nu_{R_1, \ldots, R_m} = i^{r_{n_1} \ldots r_{n_l}} i^{s_{n_m} \ldots s_{n_m}} \nu_{s_{n_1} \ldots s_{m_1} \ldots s_{m_m}} . \tag{3.3}
\]

**Example 3.2**

Let us assume that \( R_1 = rst \) and \( R_2 = u \). From the definition above we can see that

\[
\nu_{R_1} = \nu_{rst} = i^{r_{s_{11}} \ldots r_{s_{13}}} i^{s_{13} s_{11}} \nu_{s_{11} \ldots s_{13}} = i^{ra_i} i^{s_{b_i s_{c_i}}} \nu_{abc} \tag{3.4}
\]

\[
\nu_{R_1, R_2} = \nu_{rst, u} = i^{r_{s_{11}} \ldots r_{s_{13}}} i^{s_{13} s_{11}} i^{s_{21} s_{21}} \nu_{s_{11} \ldots s_{13} s_{21}} = i^{ra_i} i^{s_{b_i s_{c_i}}} i^{u d} \nu_{abc, d} \tag{3.5}
\]
Here, as elsewhere, the Einstein summation convention is used. If an index occurs twice in a term, once in the subscript and once in the superscript, summation over the index is implied.

Following are some equalities that can be derived from the above definitions and from earlier known results:

\[ \nu_r = E[l_r] = 0 \quad (3.6) \]
\[ i_{rs} = E[l_r l_s] = \nu_{r,s} \quad (3.7) \]
\[ \nu_{r,s} + \nu_{rs} = 0 \quad (3.8) \]
\[ i_{rs} i_{tu} = i_{ru}. \quad (3.9) \]

### 3.3. The Asymptotic Expansion Formula

Using the notation defined in the previous section we now derive an asymptotic expansion of \( \hat{\theta} \), the maximum likelihood estimator of \( \theta \), around \( \theta \), for increasing orders of inverse size, \( n^{-1} \). The following derivation of the likelihood function follows Barndorff-Nielsen.

Beginning by expanding the estimated value of the generic component \( l_r \) around \( \theta \), we obtain

\[
\hat{l}_r = l_r + l_{rs} (\hat{\theta} - \theta)^s + \frac{1}{2} l_{rst} (\hat{\theta} - \theta)^s (\hat{\theta} - \theta)^t + \frac{1}{6} l_{rstu} (\hat{\theta} - \theta)^s (\hat{\theta} - \theta)^t (\hat{\theta} - \theta)^u + ... \]

(3.10)

where \( (\hat{\theta} - \theta)^r = \hat{\theta}^r - \theta^r \), etc.

Inverting Eq. (3.10) to an asymptotic expansion for \( (\hat{\theta} - \theta)^r \), and collecting terms of the same asymptotic order, we finally obtain

---

\[(\hat{\theta} - \theta)^r = \frac{i^{rs} l_s}{O_p(n^{-1/2})} \nabla + \frac{1}{2} \nu^{rst} l_s l_t + \frac{i^{rs} i^{tu} H_{st} l_u}{O_p(n^{-1})} \nabla + \frac{1}{6} (\nu^{rst} + 3 \nu^{sv} i^{vw} \nu^{wu}) l_s l_t l_u \]
\[+ \frac{i^{rs} i^{tu} H_{uv} l_s l_t}{O_p(n^{-3/2})} + \frac{1}{2} i^{rs} i^{tv} H_{st} l_u l_v + \frac{1}{2} i^{rs} i^{tu} i^{vw} H_{stv} l_u l_w \]
\[+ \frac{i^{rs} i^{tu} i^{vw} H_{st} H_{uv} l_w}{O_p(n^{-3/2})} \nabla + \ldots \] (3.11)

where \( H_R = l_R - \nu_R \) and the symbol "\( \nabla \)" indicates a drop of in asymptotic magnitude of order \( n^{-1/2} \) under ordinary repeated sampling and the symbol \( O_p(n^{-1}) \) meaning a polynomial of order \( n^{-1} \). For example, \( i^{rs} l_s \) is of order \( n^{-1/2} \), while the following two terms are of order \( n^{-1} \).

The orders of \( n \) are determined as follows:

- \( i^{r_1 r_2} \) is in the order of \( n^{-1} \)
- \( \nu_{r_1 r_2 \ldots r_m} \) is in the order of \( n^1 \)
- \( \nu^{r_1 v_2 \ldots v_m} \) is in the order of \( n^{m+1} \)
- \( l_{r_1} \) is in the order of \( n^{1/2} \)
- \( H_{r_1 r_2 \ldots r_m} \) is in the order of \( n^{1/2} \)

For example, the second term of Eq. (3.11), \( \frac{1}{2} \nu^{rst} l_s l_t \), is in the order of \( O_p(\nu^{rst}) + O_p(l_s) + O_p(l_t) = n^{-2+1/2+1/2} = n^{-1} \) as stated in Eq. (3.11).

3.3.1. The First Order Asymptotic Bias

As mentioned in Eq. (2.1) the bias of an estimator is defined as \( \mathbb{E}[\hat{\theta} - \theta] \). Taking the expectation of the asymptotic expansion, Eq. (3.11), leads to the following result, derived by Barndorff - Nielsen\(^9\),

\[ b(\hat{\theta}) = \mathbb{E}[\left(\hat{\theta} - \theta\right)^r] = \mathbb{E}\left[i^{rs} l_s \nabla + \frac{1}{2} \nu^{rst} l_s l_t + i^{rs} i^{tu} H_{st} l_u \nabla\right] + O_p(n^{-3/2}) = \]

\[
\frac{1}{2} \nu_{rst} \nu_{st,u} + O_p(n^{-3/2}) = \frac{1}{2} \nu_{ra,ib,ic} \nu_{abc,ul} + O_p(n^{-3/2})
\]

From Eq. (3.12) we see that the joint moments of the likelihood-derivatives needed to be evaluated to determine the first order bias are \(i_{r_1 r_2}, \nu_{r_1 r_2 r_3}\) and \(\nu_{r_1 r_2 r_3}\) where \(r_1, r_2\) and \(r_3\) are arbitrary indices.

### 3.3.2. The Second Order Asymptotic Error

In order to determine the error correlation or the covariance of the maximum likelihood estimator, we should square the maximum likelihood expansion obtained in Eq. (3.11) and take its expectation, i.e.,

\[
E\left[ (\hat{\theta} - \theta)^2 \right].
\]

The result I obtain corresponds to the error correlation, since the likelihood expansion is evaluated around the true value of \(\theta\). Substituting Eq. (3.11) into the error correlation expression, Eq. (2.3), I obtain the following result

\[
E\left[ (\hat{\theta} - \theta)^2 \right] = 6i_{r_1} + 2i_{mb} \nu_{lmn} \nu_{s,t,u} \left[ i_{rs} i_{la} + i_{as} i_{lr} \right] + \frac{1}{2} \nu_{rs,ib,ic} \nu_{abc,ul} + O_p(n^{-3/2})
\]

It is important to note that it is not straightforward to determine the orders on \(n\) of each term in Eq. (3.13), since joint moments such as \(\nu_{bce,d,f,s}\) might involve more than one order of \(n\) as we will see later in Chapter 4. Therefore, separating Eq. (3.13) to orders of \(n\) has to be done carefully, by examining the behavior of each term in special cases, such as the Gaussian case, and inducting to the general case. Separating the orders leads to the following expression.
\[
\text{corr}(\hat{\theta'}, \hat{\theta^a}) = E\left[\left(\hat{\theta} - \theta\right)^r \left(\hat{\theta} - \theta\right)^a\right] = \left\{ \begin{array}{l}
O_p(n^{-1}) \\
\text{\text{V\text{V} +}}
\end{array} \right.
\]

\[
+ \left[ \frac{2i^m_{rb} r_{ra} v_{lmn} (r_{rs} i_{ra} + i_{as} i_{rb}) s_{b,c} (n^1)}{O_p(n^{-1})} + \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})} \right]
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \left[ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})} \right]
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
\frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

\[
+ \frac{1}{2} \frac{2i^m_{rb} i_{rb} s_{b,c} (n^2)}{O_p(n^{-1})}
\]

(3.14)

where notations such as \(v_{bee,d,f,s}(n^2)\) meaning only the \(n^2\) order terms of the joint moment \(v_{bee,d,f,s}\).

From Eq. (3.14) we see that in order to evaluate the second order error correlation we have to evaluate a few more joint-moments in addition to the ones we evaluated for the bias. The likelihood-derivatives joint-moments that are needed to be evaluated are therefore \(i_{r,r_2}, v_{r,r_2,r_3}, v_{r,r_2,r_4}, \) the \(n^1\) order of \(v_{r,r_2,r_5}\), the \(n^1\) order of \(v_{r,r_2,r_3,r_4}\), the \(n^2\) order of \(v_{r,r_2,r_3,r_5}\), the \(n^2\) order of \(v_{r,r_2,r_3,r_4,r_5}\), the \(n^2\) order term of \(v_{r,r_2,r_3,r_4,r_5}\), the \(n^1\) order term of \(v_{r,r_2,r_3,r_4,r_5}\), and the \(n^2\) order term of \(v_{r,r_2,r_3,r_4,r_5,r_6}\) where \(r_1, r_2, r_3, r_4, r_5, r_6\) are arbitrary indices.

The covariance expression equals the bias-corrected error correlation,
\[
\text{cov} \left( \hat{\theta}^r, \hat{\theta}^a \right) = E \left[ (\hat{\theta}^r - E\theta^r)(\hat{\theta}^a - E\theta^a) \right] = \\
= E \left[ (\hat{\theta}^r - E\theta^r)(\hat{\theta}^a - E\theta^a + b(\theta^a)) \right] = \\
= E \left[ (\hat{\theta}^r - \theta^r)(\hat{\theta}^a - \theta^a) - b(\theta^r)\theta^a \right] = \\
= \text{corr} \left( \hat{\theta}^r, \hat{\theta}^a \right) - b(\theta^r)\theta^a. \quad (3.15)
\]

Observing the result, we see that since the lowest order of the bias is \( n^{-1} \), the bias can affect only the covariance to second order and higher. Therefore, the covariance has the same first order term as the error correlation does, while its second order term is the second order term of the error correlation, corrected by the first order of the bias. The result we obtain for the variance is therefore

\[
\text{cov} \left( \hat{\theta}^r, \hat{\theta}^a \right) = E \left[ (\hat{\theta}^r - E\theta^r)(\hat{\theta}^a - E\theta^a) \right] = \frac{i^{ra}}{O_p(n^{-1})} \nabla \nabla + \\
\left[ 2i^{mb}i^{nc}v_{lmm} \left( i^{rs}i^{la} + i^{as}i^{lr} \right) \right]_{s,b,c} \left( n^{-1} \right) + \frac{1}{2} i^{cd}i^{ef} \left( i^{rs}i^{ab} + i^{as}i^{rb} \right)_{bce,d,f,s} \left( n^{-2} \right) \\
+ i^{tu} \left( i^{rs}i^{ab}i^{cd} + i^{rd}i^{ab}i^{cs} + i^{ad}i^{rb}i^{cs} \right) \left( n^{-2} \right) + i^{bm}i^{cq}i^{dp}v_{lmm} \left( \frac{1}{4} i^{rl}i^{ao}i^{sm} + \frac{1}{2} i^{rs}i^{al}i^{on} + \frac{1}{2} i^{as}i^{rl}i^{on} \right) \left( n^{-2} \right) \\
+ \frac{1}{2} i^{sm}v_{lmm} \left( i^{tn}i^{cd} \left( i^{rl}i^{ab} + i^{al}i^{rb} \right) + 2i^{bn}i^{cd} \left( i^{rl}i^{at} + i^{al}i^{rt} \right) + i^{cl}i^{tn} \left( i^{ad}i^{rb} + i^{rd}i^{ab} \right) \right) \left( n^{-2} \right) \\
+ \frac{1}{6} i^{mb}i^{nc}i^{od}v_{lmm} \left( i^{rs}i^{la} + i^{as}i^{rl} \right) \left( n^{-2} \right) + 4i^{bm}i^{rs}i^{al} + i^{as}i^{rl} \left( n^{-2} \right) \\
- \frac{1}{4} i^{rs}i^{tu}i^{aw}i^{yz}v_{stu}v_{wyz} + 2v_{stu}v_{wyz} \left( n^{-2} \right) + 4v_{stu} \left( n^{-1} \right) v_{wyz} \left( n^{-1} \right) \nabla \nabla . \quad (3.16)
\]

It is important to note that the first order term of the covariance, \( i^{ra} \), is the inverse of the Fisher information, which corresponds to the Cramer – Rao Bound.
A bound on the lowest possible mean square error of an unbiased estimate $\hat{\theta}$ that involves orders $n$ higher than the first and log-likelihood derivatives was introduced by Bhattacharyya\textsuperscript{10}. Defining $J$ as the matrix that contains the joint-moments of the likelihood derivatives, $[J_{jk}] = E \left[ \frac{\partial l^j}{\partial \theta^j} \frac{\partial l^k}{\partial \theta^k} \right]$, the Bhattacharyya bound is defined as $J^{11}$, the upper-left term of the inverse of $J$, so that

$$\text{MSE}(\hat{\theta}) \geq J^{11} \quad (3.17)$$

The bound is valid only for an unbiased scalar estimator and though it involves derivatives of the likelihood function, it is quite different from the multivariate covariance derived in Eq. (3.16) that is valid for biased estimates. An example of the Bhattacharyya bound is presented in Chapter 4, following example 4.2.

\textsuperscript{10} Van Trees, part I, Pp. 148-149.
4. The Asymptotic Expansion for the Gaussian Case

4.1. The Likelihood Quantities for the Gaussian Case

In this Chapter we derive the joint–moments for the case where the data samples are Gaussian, and then evaluate the corresponding bias, error correlation and covariance using Eqs. (3.12), (3.14) and (3.16), respectively. We begin by deriving the joint-moments required for evaluating the bias, for the case where the data covariance, $C$, and the data mean, $\mu$, depend upon the parameter vector $\theta$, a case we denote as the general multivariate Gaussian case. Then, we evaluate the bias for this case using Eq. (3.12). Since the joint-moments required to evaluate both the error correlation and covariance for this general case are relatively complicated, they are not derived in this thesis, but have been derived by Zanolin, Naftali and Makris in a separate work. Therefore, we define two special cases that have great practical value, since they describe a deterministic signal in additive noise and a fully randomized signal in noise, respectively. Simplified expressions for the joint-moments required for evaluating the error correlation and covariance can be obtained in these cases. In the first case the data covariance, $C$, is independent of the parameter vector, $\theta$, and in the second, the data mean, $\mu$, is zero. After deriving the joint-moments required for evaluating the error correlation and covariance, we simply substitute the results into Eqs. (3.14) and (3.16) to evaluate the error correlation and covariance for these two cases, respectively.

Let us assume we have $n$ independent samples of $N$ joint Gaussian random variables stacked into a vector $X$. In addition, let us assume the data depend on a set of $m$ unknown parameters $[\theta_1, \theta_2, \ldots, \theta_m]$ truncated into a vector $\theta$. The PDF for this case, I denote as the general multivariate Gaussian case, where both the mean, $\mu$, and the covariance matrix, $C$, depend on the parameter vector, $\theta$, is represented by

$$p(X; \theta) = \frac{1}{(2\pi)^{nN/2}} \frac{1}{|C(\theta)|^{n/2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^{n} (X_i - \mu(\theta))^T C^{-1}(\theta)(X_i - \mu(\theta)) \right\}$$

The likelihood function, $l(X; \theta)$, which is defined as the natural logarithm of the PDF is
4.1.1. The General Multivariate Gaussian Case

Below are the three likelihood-derivatives joint-moments I derived that are needed to evaluate the first order bias, Eq. (3.12), for the general multivariate Gaussian case.

\[
i_{ab} = \frac{n}{2} * \text{tr} \left( C^{-1} C_a C^{-1} C_b \right) + n * \mu_a^T C^{-1} \mu_b
\]
\[
v_{abc} = \frac{n}{3} \sum_{a,b,c} \text{tr} \left( C^{-1} C_a C^{-1} C_b C^{-1} C_c \right) - \frac{n}{4} \sum_{a,b,c} \text{tr} \left( C^{-1} C_{ab} C^{-1} C_c \right) - \frac{n}{2} \sum_{a,b,c} \mu_{ab}^T C^{-1} \mu_c
\]
\[
v_{ab,c} = -\frac{n}{2} \sum_{a,b} \text{tr} \left( C^{-1} C_a C^{-1} C_b C^{-1} C_c \right) + n \mu_{ab}^T C^{-1} \mu_c + \frac{n}{2} \text{tr} \left( C^{-1} C_{ab} C^{-1} C_c \right)
\]
\[
- n \sum_{a,b} \mu_a^T C^{-1} C_b C^{-1} \mu_c
\]

where the operator \( \text{tr}(\bullet) \) denotes trace of a matrix, and summations such as \( \sum_{a,b,c} \) meaning a sum over all possible permutations of \( a, b \) and \( c \) (a total of six), and terms such as \( C_{ab} \) and \( \mu_{ab} \) meaning the derivatives of the covariance matrix, \( C \), and the mean vector, \( \mu \), with respect to \( \theta^a \) and \( \theta^b \), respectively.

The first term, \( i_{st} \), is the Fisher -- Information matrix, and its inverse is the first order covariance matrix term, which corresponds to the Cramer -- Rao bound.

It is important to note at this stage that joint-moments for the general multivariate case that are required to evaluate the error correlation and the covariance, such as \( v_{s,t,b,c} \) and \( v_{s,t,b,c,d} \) that appear in Eqs. (3.14) and (3.16) are not derived in this thesis due to their relatively high complexity\(^{11} \). However, joint moments for two special cases, the case where the data covariance is independent of the parameter vector and the case where the mean vector is zero, are derived and will be presented in the following sections.

\(^{11} \) these joint-moments are derived by Zanolin, Naftali and Makris in a separate work.
4.1.1.1. The General First Order Bias of the MLE for multivariate Gaussian Data

Substituting the expressions I derived for the general multivariate case (Eqs. (4.3), (4.4), (4.5)) into the bias Eq. (3.12),

\[ b(\hat{\theta}^r) = \frac{1}{2} \sum_{i}^{n} \sum_{j}^{n} \sum_{k}^{n} v_{i,j,k} (v_{abc} + 2\nu_{ab,c}) \]  

(3.12)

yields the following expression for the first order bias of the maximum likelihood estimator with multivariate Gaussian data.

It is important to see that for the first order bias we need only second order derivatives of both the covariance matrix and mean with respect to the parameters that are being estimated.

Example 4.1

Suppose we would like to evaluate the bias for the vector \( \hat{\theta} = [\hat{\mu}, \hat{C}] \) where \( \hat{\mu} \) and \( \hat{C} \) are the estimators of the mean and variance, respectively, of a scalar Gaussian random variable \( x \). The MLE estimators are \( \hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_{i} \) and \( \hat{C} = \frac{1}{n} \sum_{j=1}^{n} (x_{i} - \frac{1}{n} \sum_{i=1}^{n} x_{i})^2 \) respectively\(^{12}\). The values we expect to obtain for the bias are zero for the mean and \( -\frac{C}{n} \) for the variance\(^{13}\).

We define \( \theta^1 \) as the mean and \( \theta^2 \) as the variance. Evaluating the derivatives required for the joint moments, we obtain \( \frac{\partial \mu}{\partial \theta^1} = \frac{\partial \mu}{\partial \mu} = 1, \frac{\partial C}{\partial \theta^2} = \frac{\partial C}{\partial C} = 1 \), 

\( \frac{\partial \mu}{\partial \theta^2} = \frac{\partial \mu}{\partial C} = \frac{\partial C}{\partial \theta^1} = \frac{\partial C}{\partial \mu} = 0 \) and all second derivatives are identically zero.

Substituting the above results into Eqs. (4.3)-(4.5) we obtain the Fisher Information matrix and the two joint moments required to evaluate the bias. Evaluating

\(^{12}\) Papoulis, p. 188  
\(^{13}\) Papoulis, p. 189
the bias for the mean, we set $r=1$. For any combination of $s,t$ and $u$ one of the terms of the bias expression vanishes. Therefore, the summation totals to zero. Evaluating the bias for the variance, we set $r=2$. The only $s,t$ and $u$ combination which is not zero is the one where $s=2$ and $t=u=1$. In this case, we obtain

$$b(\hat{c}) = -\frac{C}{n}$$

(4.8)

as expected by taking expectation values directly.

4.1.2. The Case where the Covariance is Parameter-Independent

We now define two special cases that have great practical value, where simplified expressions for the joint-moments can be obtained. The first case is where the covariance is parameter independent and so corresponds to a general deterministic signal in additive noise; the next is where the mean is zero and the covariance is parameter dependent and so corresponds to a fully randomized signal in noise, the former is examined in this section, the later in the next.

Restricting ourselves to a specific case where the covariance matrix is independent of the parameters to be estimated, we have the following PDF

$$p(X;\theta) = \frac{1}{(2\pi)^{nN/2}} \frac{1}{|C|^{n/2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^{n} (X_i - \mu(\theta))^T C^{-1} (X_i - \mu(\theta)) \right\}$$

(4.9)
4.1.2.1. The Joint Moments of the Log – Likelihood Derivatives

To evaluate the bias, error correlation and covariance the following likelihood-derivatives joint-moments are needed: $i_{r_1r_2}$, $\nu_{r_1r_2r_3}$, $\nu_{r_1r_2r_3r_4}$, the $n^1$ order of $\nu_{r_1r_2r_3r_4}$, the $n^1$ order of $\nu_{r_1r_2r_3r_4r_5}$, the $n^2$ order of $\nu_{r_1r_2r_3r_4r_5r_6}$, the $n^2$ order of $\nu_{r_1r_2r_3r_4r_5r_6r_7}$, the $n^3$ order of $\nu_{r_1r_2r_3r_4r_5r_6r_7r_8}$, the $n^4$ order of $\nu_{r_1r_2r_3r_4r_5r_6r_7r_8r_9}$, the $n^4$ order of $\nu_{r_1r_2r_3r_4r_5r_6r_7r_8r_9r_{10}}$, the $n^5$ order of $\nu_{r_1r_2r_3r_4r_5r_6r_7r_8r_9r_{10}r_{11}}$, the $n^6$ order of $\nu_{r_1r_2r_3r_4r_5r_6r_7r_8r_9r_{10}r_{11}r_{12}}$, the $n^7$ order of $\nu_{r_1r_2r_3r_4r_5r_6r_7r_8r_9r_{10}r_{11}r_{12}r_{13}}$, the $n^8$ order of $\nu_{r_1r_2r_3r_4r_5r_6r_7r_8r_9r_{10}r_{11}r_{12}r_{13}r_{14}}$, where $r_1, r_2, r_3, r_4, r_5$ and $r_6$ are arbitrary indices.

Below are the joint moments for the parameter-independent covariance matrix $I$ evaluated. It is important to note that some terms involve multiple orders of $n$, such as $\nu_{u,d,st,bc}$ that involves $n^2$ and $n^3$. In order to evaluate the first and the second orders of the error correlation and the covariance we have to pick the correct orders of the joint-moment defined in Eqs. (3.14) and (3.16), as is done in Eqs. (4.10)-(4.19).

\[
i_{ab} = n \mu_a^T C^{-1} \mu_b \tag{4.10}
\]

\[
\nu_{abc}(n^1) = -\frac{n}{2} \sum_{a,b,c} \mu_{ab}^T C^{-1} \mu_c \tag{4.11}
\]

\[
\nu_{ab,c}(n^1) = n \mu_{ab}^T C^{-1} \mu_c \tag{4.12}
\]

\[
\nu_{abcd}(n^1) = -\frac{n}{8} \sum_{a,b,c,d} \mu_{ab}^T C^{-1} \mu_{cd} - \frac{n}{6} \sum_{a,b,c,d} \mu_{abc}^T C^{-1} \mu_d \tag{4.13}
\]

\[
\nu_{a,b,c}(n^1) = 0 \tag{4.14}
\]

\[
\nu_{a,b,c,d}(n^1) = 0 \tag{4.15}
\]

\[
\nu_{a,b,c,d}(n^2) = \frac{n^2}{8} \sum_{a,b,c,d} \mu_a^T C^{-1} \mu_b \mu_c^T C^{-1} \mu_d \tag{4.16}
\]

\[
\nu_{a,b,c,d,e}(n^2) = \frac{n^2}{2} \sum_{a,b,c} \mu_a^T C^{-1} \mu_b \mu_c^T C^{-1} \mu_{de} \tag{4.17}
\]

\[
\nu_{a,b,c,d,e,f}(n^2) = \frac{n^2}{2} \sum_{(a,b) \in (c,d,e,f)} \mu_a^T C^{-1} \mu_{cd} \mu_b^T C^{-1} \mu_{ef} + n^2 \mu_{cd}^T C^{-1} \mu_{ef} \mu_a^T C^{-1} \mu_b \tag{4.18}
\]
\[ v_{a,b,c,def}(n^2) = \frac{n^2}{2} \sum_{a,b,c} \mu_a^T C^{-1} \mu_b \mu_c^T C^{-1} \mu_{def} \]  

(4.19)

where the notation \( \sum \) meaning sum over all possible permutations of \( a \) and \( b \), combined with permutations of \( cd \) and \( ef \) (a total of four).

For the case where the parameter, \( \theta \), is a scalar the joint moments derived in Eqs. (4.10)-(4.19) reduce to

\[ i_{ab} = \frac{n(\mu')^2}{C} \]  

(4.20)

\[ v_{abc}(n^1) = -\frac{3n(\mu'\mu'')}{C} \]  

(4.21)

\[ v_{ab,c}(n^1) = \frac{n(\mu'\mu'')}{C} \]  

(4.22)

\[ v_{abcd}(n^1) = -\frac{4n(\mu''\mu''')}{C} - \frac{3n(\mu')^2}{C} \]  

(4.23)

\[ v_{a,b,c}(n^1) = 0 \]  

(4.24)

\[ v_{a,b,cd}(n^1) = 0 \]  

(4.25)

\[ v_{a,b,c,d}(n^2) = \frac{3n^2(\mu')^4}{C^2} \]  

(4.26)

\[ v_{a,b,c,de}(n^2) = \frac{3n^2(\mu')^3(\mu'')}{C^2} \]  

(4.27)

\[ v_{a,b,c,def}(n^2) = \frac{3n^2(\mu')^2(\mu'')^2}{C^2} \]  

(4.28)

\[ v_{a,b,c,def}(n^2) = \frac{3n^2(\mu')^3(\mu''')}{C^2} \]  

(4.29)

4.1.2.2. The Bias Term to First Order

Substituting Eqs. (4.10), (4.11) and (4.12) into the bias expression, Eq. (3.12), yields the bias for the independent covariance case. The result obtained is simpler than
the one obtained for the general case, Eq. (4.6), since the derivatives of the covariance matrix with respect to the parameter vanish.

\[
b(\hat{r}) = \frac{1}{2} i^{rs} i^{pu} (v_{stu} + 2v_{st,u}) \nabla + O_p(n^{-3/2}) = \\
= \sum_{s=1}^{m} \sum_{t=1}^{m} \sum_{\alpha=1}^{m} \frac{1}{2n} \left( \frac{\partial \mu}{\partial \theta^r} \right)^T C^{-1} \left( \frac{\partial \mu}{\partial \theta^s} \right)^{-1} \frac{\partial^2 \mu}{\partial \theta^t \partial \theta^u} C^{-1} \left( \frac{\partial \mu}{\partial \theta^u} \right)^{-1} \\
= \left( \frac{\partial^2 \mu}{\partial \theta^s \partial \theta^t} \right)^T C^{-1} \left( \frac{\partial \mu}{\partial \theta^t} \right) - \frac{\partial^2 \mu}{\partial \theta^s \partial \theta^u} C^{-1} \left( \frac{\partial \mu}{\partial \theta^u} \right) \\
(4.30)
\]

For the scalar case, the bias is evaluated by substituting Eqs. (4.20), (4.21) and (4.22) into Eq. (3.12). The bias result obtained is therefore

\[
b(\hat{r}) = \frac{1}{2} \left( \frac{C}{n(\mu')^2} \right)^2 \left( -\frac{3n(\mu')(\mu'')}{C} + 2 \left( \frac{n(\mu')(\mu''')}{C} \right) \right) = \\
- \frac{1}{2} \left( \frac{C}{n(\mu')^2} \right)^2 \left( \frac{n}{C} (\mu')(\mu'') \right) = - \frac{C (\mu'')}{2n (\mu')^3} \quad (4.31)
\]

4.1.2.3. The Covariance and the Error Correlation to Second Order

Substituting Eqs. (4.10)-(4.19) into the error correlation and covariance expressions, Eqs. (3.14) and (3.16) respectively, yields the error correlation and covariance for the independent covariance case.

For the scalar case, substituting Eqs. (4.20)-(4.29) into the error correlation and covariance expressions, Eqs. (3.14) and (3.16) respectively, yields the following mean-square error and variance expressions for the independent covariance case

\[
\text{MSE}(\hat{\theta}) = \frac{C}{n(\mu')^2} + \frac{15C^2 (\mu'')^2}{4n^2 (\mu')^6} - \frac{C^2 (\mu''')}{n^2 (\mu')^5} \quad (4.32)
\]
Example 4.2

Assume we are interested in evaluating the bias, MSE and variance of the maximum likelihood estimator for \( \theta = \mu^2 \) where \( x \) is a Gaussian scalar random variable with mean \( \mu \) and variance \( \sigma \), and the \( \sigma \) is independent of \( \mu^2 \).

Using the invariance property of the MLE we can say that

\[
\hat{\theta} = (\hat{\mu})^2 = \frac{1}{n} \sum_{i=1}^{n} x_i x_j
\]

The corresponding bias, mean-square error and variance are

\[
b(\hat{\mu}^2) = E[(\hat{\mu}^2 - \mu^2)] = E\left[ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j - \mu^2 \right] = \frac{C}{n} \tag{4.34}
\]

\[
\text{MSE}(\hat{\mu}^2) = E[(\hat{\mu}^2 - \mu^2)^2] = E\left[ \left( \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j - \mu^2 \right)^2 \right] = \frac{4C\mu^2}{n^2} + \frac{3C^2}{n^2} \tag{4.35}
\]

\[
\text{Var}(\hat{\mu}^2) = E\left[ (\hat{\mu}^2 - E[\hat{\mu}^2])^2 \right] = E\left[ \left( \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j - E\left[ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j \right] \right)^2 \right] = \frac{4C\mu^2}{n^2} + \frac{2C^2}{n^2} \tag{4.36}
\]

Next we evaluate the derivatives of the mean with respect to the scalar parameter. In this example the parameter is \( \mu^2 \), therefore the derivatives are

\[
\frac{\partial \mu}{\partial \theta} = \frac{1}{2\mu} \tag{4.37}
\]

\[
\frac{\partial^2 \mu}{\partial \theta^2} = -\frac{1}{4\mu^3} \tag{4.38}
\]

\[
\frac{\partial^3 \mu}{\partial \theta^3} = \frac{3}{8\mu^5} \tag{4.39}
\]

Substituting these three derivatives into Eqs. (4.31), (4.32) and (4.33) we obtain the bias, the mean-square error and the variance for the scalar case, respectively. The results we
obtain are identical to those obtained by taking direct expectations, which are given in Eqs. (4.34), (4.35) and (4.36), respectively. 

Next, let us examine the Bhattacharyya bound for this case. Since the estimator for $\theta=\mu^2$ is biased, we violate one of the bound restrictions, therefore we do not expect the inequality in Eq. (3.17), $\text{Variance}(\hat{\theta}) \geq J^{11}$, to hold. As mentioned before, $J^{11}$ is the upper-left term of the inverse of the matrix $J$, which its elements are defined by 

$$
\begin{bmatrix}
J_{jk}
\end{bmatrix} = E \left[ \frac{\partial l^j}{\partial \theta^k} \right], \text{ when } l \text{ is the likelihood function.}
$$

Let us assume we are setting the maximum order of the derivatives to 2, hence the $J$ matrix is a 2-by-2 matrix, and its inverse matrix upper-left term, $J^{11}$, can be represented by

$$
J^{11} = \frac{1}{J_{11}} + \frac{J_{12}^2}{J_{11}(J_{11}J_{22} - J_{12}^2)}
$$

where

$$
J_{11} = E \left[ \left( \frac{\partial l}{\partial \theta} \right)^2 \right], \quad J_{22} = E \left[ \left( \frac{\partial^2 l}{\partial \theta^2} \right)^2 \right], \quad J_{12} = E \left[ \frac{\partial l}{\partial \theta} \frac{\partial^2 l}{\partial \theta^2} \right].
$$

Evaluating the three elements, $J_{11}$, $J_{22}$ and $J_{12}$, for $\theta=\mu^2$ using Eqs. (4.37)-(4.39) and substituting the results in Eq. (4.40), we obtain the following lower limit for the variance of the estimator

$$
J^{11} = \frac{4C\mu^2}{n} + \frac{4C^2}{n^2}.
$$

Comparing the result obtained in Eq. (4.42) and the actual variance in Eq. (4.36) we see that the bound for the variance obtained is greater than the actual variance, hence the Bhattacharyya bound cannot be implemented for this estimator because it is biased. Moreover, going to third order derivatives, we find that the 3-by-3 $J$ matrix becomes singular and a bound cannot be evaluated using third order derivatives.
4.1.3. The Case where the Mean is Zero and the Covariance is Parameter Dependent

Next we restrict ourselves to the case where the mean is zero and the covariance matrix is dependent upon the parameters to be estimated, which has the PDF

\[ p(X; \theta) = \frac{1}{nN \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^{n} (X_i)^T C^{-1}(\theta) X_i \right\} \]  

(4.43)

4.1.3.1. The Joint Moments of the Log–Likelihood Derivatives

To evaluate the bias, error correlation and covariance the following likelihood-derivatives joint-moments are needed: \( i_{r_1 r_2} \), \( v_{r_1 r_2 r_3} \), \( v_{r_1 r_2 r_4} \), the \( n^1 \) order of \( v_{r_1 r_2 r_3} \), the \( n^1 \) order of \( v_{r_1 r_2 r_3 r_4} \), the \( n^2 \) order of \( v_{r_1 r_2 r_3 r_4} \), the \( n^2 \) order of \( v_{r_1 r_2 r_3 r_4 r_5} \), the \( n^3 \) order term of \( v_{r_1 r_2 r_3 r_4 r_5} \), the \( n^1 \) order term of \( v_{r_1 r_2 r_3 r_4} \), and the \( n^2 \) order term of \( v_{r_1 r_2 r_3 r_4 r_5} \), where \( r_1, r_2, r_3, r_4, r_5 \) and \( r_6 \) are arbitrary indices.

Below are the joint moments for the zero-mean parameter-dependent covariance matrix I evaluated. It is important to note that some terms involve multiple orders of \( n \), such as \( v_{u,d,st,bc} \) that involves \( n^2 \) and \( n^3 \). In order to evaluate the first and the second orders of the error correlation and the covariance we have to pick the correct orders of the joint-moment defined in Eqs. (3.14) and (3.16), as is done in Eqs. (4.44)-(4.53).

\[
i_{ab} = \frac{n}{2} * \text{tr} \left( C^{-1} C_a C^{-1} C_b \right) \] 

(4.44)

\[
v_{abc} \left( n^1 \right) = \frac{n}{3} \sum_{a,b,c} \text{tr} \left( C^{-1} C_a C^{-1} C_b C^{-1} C_c \right) - \frac{n}{4} \sum_{a,b,c} \text{tr} \left( C^{-1} C_{ab} C^{-1} C_c \right) \] 

(4.45)

\[
v_{ab,c} \left( n^1 \right) = -\frac{n}{2} \sum_{a,b} \text{tr} \left( C^{-1} C_a C^{-1} C_b C^{-1} C_c \right) + \frac{n}{2} \text{tr} \left( C^{-1} C_{ab} C^{-1} C_c \right) \] 

(4.46)

\[
v_{abcd} \left( n^1 \right) = -\frac{3n}{8} \sum_{a,b,c,d} \text{tr} \left( C^{-1} C_a C^{-1} C_b C^{-1} C_c C^{-1} C_d \right) + \frac{n}{2} \sum_{a,b,c,d} \text{tr} \left( C^{-1} C_{ab} C^{-1} C_c C^{-1} C_d \right) \] 

\[
- \frac{n}{16} \sum_{a,b,c,d} \text{tr} \left( C^{-1} C_{ab} C^{-1} C_{cd} \right) - \frac{n}{12} \sum_{a,b,c,d} \text{tr} \left( C^{-1} C_{abc} C^{-1} C_d \right) \] 

(4.47)
\[ \nu_{a,b,c} (n^1) = \frac{n}{6} \sum_{a,b,c} \text{tr} \left( C^{-1} C_a C^{-1} C_b C^{-1} C_c \right) \] (4.48)

\[ \nu_{a,b,c,d} (n^1) = -\frac{n}{2} \sum_{(a,b)\times(c,d)} \text{tr} \left( C^{-1} C_a C^{-1} C_b C^{-1} C_c C^{-1} C_d + \frac{n}{2} \sum_{a,b} \text{tr} \left( C^{-1} C_a C^{-1} C_b C^{-1} C_c C^{-1} C_d \right) \] (4.49)

\[ \nu_{a,b,c,d} (n^2) = \frac{n^2}{32} \sum_{a,b,c,d} \left\{ \text{tr} \left( C^{-1} C_a C^{-1} C_b \right) \text{tr} \left( C^{-1} C_c C^{-1} C_d \right) \right\} \] (4.50)

\[ \nu_{a,b,c,d} (n^2) = -\frac{n^2}{24} \sum_{(a,b,c)\times(d,e)} \left\{ \text{tr} \left( C^{-1} C_d C^{-1} C_e \right) \text{tr} \left( C^{-1} C_a C^{-1} C_b C^{-1} C_c \right) \right\} \] (4.51)

\[ \nu_{a,b,c,d,e} (n^2) = \frac{n^2}{8} \sum_{(c,d)\times(a,b)\times(c,d)\times(e,f)} \left\{ \text{tr} \left( C^{-1} C_a C^{-1} C_b \right) \text{tr} \left( C^{-1} C_c C^{-1} C_d \right) \right\} \] (4.54)
\[-\frac{n^2}{8} \sum_{(cd,ef)\times(a,b)\times(c,d)\times(e,f)} \left\{ tr\left( C^{-1} C_a C^{-1} C_{cd} \right) \times tr\left( C^{-1} C_b C^{-1} C_e C^{-1} C_f \right) \right\} \]

(4.52)

\[v_{a,b,c,def} \left( n^2 \right) = \frac{n^2}{18} \sum_{(a,b,c)\times(d,e,f)} \left\{ tr\left( C^{-1} C_a C^{-1} C_b C^{-1} C_c \right) \times tr\left( C^{-1} C_d C^{-1} C_e C^{-1} C_f \right) \right\} \]

(4.53)

For the case where the parameter, \( \theta \), is a scalar the joint moments derived in Eqs. (4.44)-(4.53) reduce to

\[i_{ab} = \frac{n}{2C^2} \left( C' \right)^2 \]

(4.54)

\[v_{abc} \left( n^1 \right) = \frac{2n}{C^3} \left( C' \right)^3 - \frac{3n}{2C^2} \left( C' \right) \left( C'' \right) \]

(4.55)

\[v_{ab,c} \left( n^1 \right) = -\frac{n}{C^3} \left( C' \right)^3 + \frac{n}{2C^2} \left( C' \right) \left( C'' \right) \]

(4.56)

\[v_{abcd} \left( n^1 \right) = -\frac{9n}{C^4} \left( C' \right)^4 + \frac{12n}{C^3} \left( C' \right)^2 \left( C'' \right) - \frac{3n}{2C^2} \left( C'' \right)^2 - \frac{2n}{C^2} \left( C' \right) \left( C''' \right) \]

(4.57)

\[v_{a,b,c} \left( n^1 \right) = \frac{n}{C^3} \left( C' \right)^3 \]

(4.58)

\[v_{a,b,cd} \left( n^1 \right) = -\frac{2n}{C^4} \left( C' \right)^4 + \frac{n}{C^3} \left( C' \right)^2 \left( C'' \right) \]

(4.59)
\[
\nu_{a,b,c,d}(n^2) = \frac{3n^2}{4C^4}(C')^4
\]  
(4.60)

\[
\nu_{a,b,c,d,e}(n^2) = -\frac{2n^2}{C^5}(C')^5 + \frac{3n^2}{4C^4}(C')^3(C''')
\]  
(4.61)

\[
\nu_{a,b,c,d,e,f}(n^2) = \frac{5n^2}{C^6}(C')^6 - \frac{4n^2}{C^5}(C')^4(C''') + \frac{3n^2}{4C^4}(C')^2(C''')^2
\]  
(4.62)

\[
\nu_{a,b,c,d,e,f}(n^2) = \frac{13n^2}{2C^6}(C')^6 - \frac{6n^2}{C^5}(C')^4(C''') + \frac{3n^2}{4C^4}(C')^3(C''')
\]  
(4.63)

4.1.3.2. The Bias Term to First Order

Substituting Eqs. (4.44), (4.45) and (4.46) into the bias expression, Eq. (3.12), yields the bias for the zero-mean parameter dependent covariance case. The result obtained is simpler than the one obtained for the general case, Eq.(4.6), since the derivatives of the covariance matrix with respect to the parameter vanish.

\[
b(\hat{\theta}) = \frac{1}{2} \sum_i \sum_j \sum_k \left( \text{tr} \left( \begin{bmatrix} \frac{\partial \mathbf{C}}{\partial \theta^r} & \frac{\partial \mathbf{C}}{\partial \theta^s} \end{bmatrix} \right) \right)^{-1} 
\]

\[
\left( \text{tr} \left( \begin{bmatrix} \frac{\partial^2 \mathbf{C}}{\partial \theta^r \partial \theta^t} & \frac{\partial \mathbf{C}}{\partial \theta^t} \end{bmatrix} \right) \right)^{-1} 
\]

\[
\left( \text{tr} \left( \begin{bmatrix} \frac{\partial \mathbf{C}}{\partial \theta^t} \end{bmatrix} \right) \right) 
\]

\[
\left( \text{tr} \left( \begin{bmatrix} \frac{\partial \mathbf{C}}{\partial \theta^u} \end{bmatrix} \right) \right) 
\]

\[
\left( \text{tr} \left( \begin{bmatrix} \frac{\partial \mathbf{C}}{\partial \theta^s} \end{bmatrix} \right) \right) 
\]

(4.64)

For the scalar case, the bias is evaluated by substituting Eqs. (4.54), (4.55) and (4.56) into Eq. (3.12). The bias result obtained is therefore

\[
b(\hat{\theta}) = \frac{1}{2} \left( \frac{2C^2}{n(C''')^2} \right)^2 \left( \frac{2n}{C^3} \left( C'''' \right)^3 - \frac{3n}{2C^2} \left( C'''' \right)^3 \right) + \left( \frac{n}{C^3} \left( C'''' \right)^3 + \frac{n}{2C^2} \left( C'''' \right)^3 \right) = 
\]

\[
\frac{1}{2} \left( \frac{2C^2}{n(C''')^2} \right)^2 \left( - \frac{n}{2C^2} \left( C'''' \right)^3 \right) = - \frac{C^2}{n} \left( C'''' \right)^3
\]  
(4.65)
4.1.3.3. The Covariance and the Error Correlation to Second Order

Substituting Eqs. (4.44)-(4.53) into the error correlation and covariance expressions, Eqs. (3.14) and (3.16) respectively, yields the error correlation and covariance for the zero mean and parameter – dependent covariance case.

For the scalar case, substituting Eqs. (4.54)-(4.63) into the error correlation and covariance expressions, Eqs. (3.14) and (3.16) respectively, yields the following mean-square error and variance expressions for the zero-mean case

\[
\text{MSE}(\hat{\theta}) = \frac{2C^2}{n(C')^2} - \frac{8C^3(C'')}{n^2(C')^4} + \frac{15C^4(C''')^2}{n^2(C')^6} - \frac{4C^4(C''')^2}{n^2(C')^5}
\]

(4.66)

\[
\text{var}(\hat{\theta}) = \frac{2C^2}{n(C')^2} - \frac{8C^3(C'')}{n^2(C')^4} + \frac{14C^4(C''')^2}{n^2(C')^6} - \frac{4C^4(C''')^2}{n^2(C')^5}
\]

(4.67)

**Example 4.3**

Assume we are interested in evaluating the bias, MSE and variance of the maximum likelihood estimator for \(\theta = C^2\) where \(x\) is a zero-mean Gaussian scalar random variable with variance \(C\) which is dependent upon \(\theta\).

Using the invariance property of the MLE we can say that

\[
\hat{\theta} = \left(\hat{C}\right)^2 = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} x_i^2 x_j^2 \]

The corresponding bias, mean-square error and variance are

\[
b\left(\hat{C}^2\right) = E\left[\left(\hat{C}^2 - C^2\right)\right] = E\left[\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} x_i^2 x_j^2 - C^2\right] = \frac{2C^2}{n}
\]

(4.68)

\[
\text{MSE}\left(\hat{C}^2\right) = E\left[\left(\hat{C}^2 - C^2\right)^2\right] = E\left[\left(\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} x_i^2 x_j^2 - C^2\right)^2\right] = \frac{8C^4}{n} + \frac{44C^4}{n^2}
\]

(4.69)

\[
\text{var}\left(\hat{C}^2\right) = E\left[\left(\hat{C}^2 - E[\hat{C}^2]\right)^2\right] =
\]
\[ E \left( \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} x_i^2 x_j^2 - E \left( \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} x_i^2 x_j^2 \right) \right)^2 \] = \frac{8C^4}{n} + \frac{40C^4}{n^2} \tag{4.70}

Next we evaluate the derivatives of the variance with respect to the scalar parameter. In this example the parameter is \( C^2 \), therefore the derivatives are

\[ \frac{\partial C}{\partial \theta} = \frac{1}{2C} \] \hspace{2cm} (4.71)

\[ \frac{\partial^2 C}{\partial \theta^2} = -\frac{1}{4C^3} \] \hspace{2cm} (4.72)

\[ \frac{\partial^3 C}{\partial \theta^3} = \frac{3}{8C^5} \] \hspace{2cm} (4.73)

Substituting these three derivatives into Eqs. (4.65), (4.66) and (4.67) we obtain the bias, the mean-square error and the variance for the scalar case, respectively. The results we obtain are identical to those obtained by taking direct expectations, which are given in Eqs. (4.55), (4.56) and (4.57), respectively.
### 4.2. A Summary of Results

Table 4.1 provides a list of expressions derived in this chapter.

<table>
<thead>
<tr>
<th>Case</th>
<th>Data Covariance</th>
<th>Data Mean</th>
<th>1&lt;sup&gt;st&lt;/sup&gt; Order Bias Expression</th>
<th>Variance to 2&lt;sup&gt;nd&lt;/sup&gt; Order Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Multivariate</td>
<td>√</td>
<td>√</td>
<td>Multivariate Data</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>/ Multiple Parameters – Eq. (4.6)</td>
<td></td>
</tr>
<tr>
<td>Parameter Independent Covariance</td>
<td>√</td>
<td>√</td>
<td>Multiple Data Sensors / Multiple Parameters – Eq. (4.30)</td>
<td>Multiple Data Sensors / Multiple Parameters – Sec. 4.1.2.3</td>
</tr>
<tr>
<td>Zero Mean</td>
<td>√</td>
<td>√</td>
<td>Multiple Data Sensors / Multiple Parameters – Eq. (4.64)</td>
<td>Multiple Data Sensors / Multiple Parameters – Sec. 4.1.3.3</td>
</tr>
</tbody>
</table>

| Table 4.1 – A List of Expressions Derived in Chapter 4 |
5. Applications of the Asymptotic Expansion – Time Delay and Doppler Shift Estimation

5.1. Real and Complex Waveform Representation

5.1.1. Real Signal

A signal with finite energy can be represented either in the time domain, \( \tilde{s}(t) \), or in the frequency domain, \( \tilde{S}(f) \), where a tilde denotes a complex quantity here and elsewhere. Both representations are related through the Fourier transform pairs

\[
\tilde{S}(f) = \int_{-\infty}^{\infty} \tilde{s}(t) e^{-j2\pi ft} \, dt \tag{5.1}
\]

\[
\tilde{s}(t) = \int_{-\infty}^{\infty} \tilde{S}(f) e^{j2\pi ft} \, df \tag{5.2}
\]

The function \( \tilde{s}(t) \) is conventionally called the signal waveform and \( \tilde{S}(f) \) the frequency spectrum. The complex signals we are interested in, both in radar and in sonar applications, have a zero phase in the time domain where they are purely real.

The fact that all the signals of interest are real in the time domain has consequences on the formalism of the waveform analysis. For the real time-domain case we can relate the positive and the negative frequencies by the following equality,

\[
\tilde{s}(-f) = \tilde{S}^*(f). \tag{5.3}
\]

From Eq. (5.3) we can see that for this special case, working solely with the positive frequencies suffices, since the negative frequencies mirror the positive ones in complex conjugate form\textsuperscript{14}.

\textsuperscript{14} Rihaczek, pp.10-11.
5.1.2. Complex Representation of Real Signals

The complex representation, described by Gabor\(^{15}\), consists of solely working with positive frequencies, on the principle that the negative frequencies simply mirror the positive ones in complex conjugate form. Separating the integration in (5.1) over the positive and negative parts of the spectrum, we can write

\[
\tilde{s}(t) = \int_{-\infty}^{\infty} \tilde{S}(f) e^{j2\pi ft} df = \int_{-\infty}^{0} \tilde{S}(f) e^{j2\pi ft} df + \int_{0}^{\infty} \tilde{S}(f) e^{j2\pi ft} df .
\]  

(5.4)

Assuming the time domain signal is real, so that (5.3) holds, and performing a change of variable, we obtain the following expression for \(s(t)\)

\[
s(t) = \int_{0}^{\infty} \tilde{S}^*(f) e^{-j2\pi ft} df + \int_{0}^{\infty} \tilde{S}(f) e^{j2\pi ft} df = \text{Re} \left\{ \int_{0}^{\infty} 2\tilde{S}(f) e^{j2\pi ft} df \right\} = \text{Re} \left\{ \tilde{S}(f) e^{j2\pi ft} df \right\} .
\]

(5.5)

Inspection of Eq. (5.5) shows that the complex signal can be derived from the real signal either by omitting the negative frequencies and doubling the amplitude of the positive ones or, alternatively, by omitting the positive frequencies and doubling the amplitude of the negative ones. We define the analytic time signal as

\[
\tilde{\psi}(t) = \int_{0}^{\infty} 2\tilde{S}(f) e^{j2\pi ft} df
\]

(5.6)

where \(s(t) = \text{Re}[\tilde{\psi}(t)]\). The frequency spectrum of the analytic signal is therefore defined as

\[
\tilde{\phi}(f) = \begin{cases} 
2\tilde{S}(f) & f > 0 \\
0 & f < 0
\end{cases}
\]

(5.7)

where the analytic time signal and its spectrum form a Fourier pair the same way \(\tilde{s}(t)\) and \(\tilde{S}(f)\) did.

The analytic signal, \(\tilde{\psi}(t)\), can be written as \(\tilde{\psi}(t) = s(t) + \tilde{s}(t)\) where \(s(t)\) is the real time-domain signal, and \(\tilde{s}(t)\) is a unique function that is formed to cancel the spectrum of

\(^{15}\) Woodward, pp. 40-41.
\( s(t) \) in the negative frequencies. The two functions are known to be related to each other by the Hilbert transform\(^\text{16}\)

\[
\hat{s}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{s(\xi)}{t - \xi} d\xi
\]

\( s(t) = -\frac{1}{\pi} \int_{-\infty}^{\infty} \hat{s}(\xi) \frac{\xi}{t - \xi} d\xi. \tag{5.9}
\]

It is important to note that if the real waveform contains no frequencies outside the \((-W, W)\) band, the analytic signal will then contain spectral components only in the \((0, W)\) band. Therefore, as opposed to the real waveform, where the signal can be completely described by samples at intervals of \(1/(2W)\), \(1/NyquistRate\), the complex signal can be completely described by complex samples at intervals \(1/W\).

When using the complex notation, the energy of the real physical waveform \(s(t)\) is given by

\[
E = \int s^2(t) dt = \frac{1}{2} \int |\hat{s}(\xi)|^2 d\xi,
\]

hence the complex signal doubles the energy of the real signal.

### 5.2. The Matched Filter

Since the received signal is always immersed in noise and interference from other sources of scatterers, the receiver must be optimized in some manner. The question of what constitutes an optimal receiver can seldom be answered exactly if such problems as prior information about the environment, consequences of errors in the interpretation of the measurement, and practical system implementation are taken into consideration. Among the entire class of receivers, the matched filter attains a unique compromise between theoretical performance and ease of implementation.

For a simple problem of detecting a single signal at unknown time delay in a stationary white Gaussian noise environment, an engineering approach to optimize the results is to maximize the signal-to-noise ratio (SNR) at the receiver output. This approach leads to the matched filter receiver. It was also shown that such a receiver

\(^{16}\) Rihaczek, pp.15-18
preserves all the information the received signal contains\textsuperscript{17}. Although it may be argued that preserving the information is not an adequate criterion for receiver optimally, practical aspects tend to favor the matched filter receiver.

To arrive at the matched filter concept, we consider maximizing the SNR at the output of a passive linear filter. Let us assume the filter has a real impulse response \( h(t) \) with an analytic complex form \( \tilde{\phi}(t) \), and the corresponding spectrums \( H(f) \) and \( \tilde{\phi}(f) \), respectively. The filter’s input is composed of a complex signal, \( \psi_i(t) \) and additive double-sided (-\( W \), \( W \)) white Gaussian noise, \( n_i(t) \), with a power density of \( N_0/2 \).

The average noise output power is

\[
N = \frac{N_0}{2} \int_{-\infty}^{\infty} |H(f)|^2 df = \frac{N_0}{4} \int_{0}^{\infty} |\tilde{\phi}(f)|^2 df , \tag{5.10}
\]

whereas the output signal peak power is

\[
\frac{1}{4} \left| \int_{0}^{\infty} \tilde{\Psi}(f)\tilde{\phi}(f)e^{j2\pi f \tau} df \right|^2 \tag{5.11}
\]

where \( \tilde{\Psi}(f) \) is the analytic time signal spectrum \( \tau \) represents the signal delay in the filter. The SNR to be maximized is defined as the ratio of peak signal power to average noise power, or

\[
SNR = \frac{\left| \int_{0}^{\infty} \tilde{\Psi}(f)\tilde{\phi}(f)e^{j2\pi f \tau} df \right|^2}{N_0 \int_{0}^{\infty} |\tilde{\phi}(f)|^2 df} \tag{5.12}
\]

\textsuperscript{17} Woodward, pp. 40-41.
To find the maximum of $d$, we use the Schwarz inequality for complex functions,

$$\left| \int_{a}^{b} u(x)v(x)dx \right|^2 \leq \int_{a}^{b} |u(x)|^2 dx \int_{a}^{b} |v(x)|^2 dx.$$  

(5.13)

The equality holds only if $u(x) = kv(x)^*$ where $k$ is an arbitrary constant. Identifying $u(x)$ with $\tilde{\Phi}(f)$ and $v(x)$ with $\tilde{\varphi}(f)e^{j2\pi f\tau_d}$ and applying the Schwarz inequality on the SNR term in Eq. (5.12), we obtain

$$d \equiv d_{\text{max}} = \frac{2E}{N_0}$$

(5.14)

and the maximum SNR at the receiver’s output is seen to depend only on the signal energy and the noise power density. Most significantly, it doesn’t depend on the signal waveform.

The filter that maximizes the SNR is defined by $\tilde{\Phi}(f) = ke^{-j2\pi f\tau_d} \tilde{\varphi}^*(f)$, where $\tau_d$ represents the signal delay in the filter. The impulse response of the filter is then $\tilde{\phi}(t) = k\tilde{\varphi}_i^*(\tau_d - t)$, or $h(t) = ks_i(\tau_d - t)$ in the real notation.

Observing the impulse response of the filter, we can see that it is equal to the reversed time input waveform, except for an arbitrary scale factor and a translation in time. For this reason the filter is called a matched filter. As for the frequency domain, the matched filter frequency response is the complex conjugate of the signal spectrum, modified by an arbitrary scale factor and phase. This means that maximization of the SNR is achieved by first removing any nonlinear phase function of the spectrum and then weighting the received spectrum in accordance with the strength of the spectral components of the transmitted signal.

Substituting the transfer function obtained for the matched filter into Eq. (5.11) and transforming it to the time domain, we can define the matched filter as the filter that maximizes the quantity $\int_0^T [(\tilde{\varphi}_i(\xi) + n_i(\xi))\tilde{\varphi}_i^*(\xi - \tau_d) \xi].$ It can then be represented as shown in Fig. 5.2, where time delay is changed until a maximum output is obtained at $\tau = \tau_d$. 

45
We now examine how the matched filter is related to the maximum likelihood estimator.

Since we assume the signal $\tilde{\psi}_i(t)$ is embedded in white Gaussian noise with power density of $N_0$, the filter’s input, $\tilde{\phi}_i(t)$, is Gaussian with mean $\tilde{\psi}_i(t, \theta)$, where $\theta$ is a parameter that modifies the mean, and its autocorrelation function is,

$$E\{\tilde{\phi}_i(t, \theta)\tilde{\phi}_i^*(t + \tau, \theta)\} = N_0 \delta(\tau). \tag{5.15}$$

Its PDF of the signal is,

$$p(\tilde{\phi}_i(t); \theta) = k * \exp\left\{-\frac{1}{2N_0} \int_0^T [(\tilde{\phi}_i(t) - \tilde{\psi}_i(t, \theta))^2] dt\right\} \tag{5.16}$$

for the complex representation, and

$$p(x_i(t); \theta) = k * \exp\left\{-\frac{1}{N_0} \int_0^T [x_i(t) - s_i(t, \theta))^2] dt\right\} \tag{5.17}$$

for the real representation, where $k$ is a normalization factor. Moreover, we assume that $\theta$ is a scalar parameter that modifies the mean only, so that the variance is parameter-independent. The maximum likelihood estimator for $\theta$, is the one that maximizes the likelihood function

$$-\frac{1}{2N_0} \int_0^T [(\tilde{\phi}_i(t) - \tilde{\psi}_i(t, \theta))^2] dt. \tag{5.18}$$

Expanding the integral in Eq. (5.18) we see that it can be separated into three parts,
\[
-\frac{T}{2N_0} \int_0^T [(\varphi_i(t) - \overline{\varphi}_i(t, \theta))^2 dt = -\frac{T}{2N_0} \left\{ \int_0^T \left| \overline{\varphi}_i(t) \right|^2 dt + \int_0^T \left| \overline{\varphi}_i(t, \theta) \right|^2 dt \right\} - 2 \Re \left\{ \int_0^T \overline{\varphi}_i(t) \overline{\varphi}_i^* (t, \theta) dt \right\}. \tag{5.19}
\]

The first two terms of Eq. (5.19) represent the energy of the filter's input signal and the energy of the parameterized original signal, respectively. Assuming that the signal energy is not affected by the parameter, as is the case in the time-delay and Doppler shift estimations, these terms do not play a role in the maximization. Therefore, maximizing the likelihood function when \( \theta = \tau \), reduces to maximizing the quantity \( \int_0^T \overline{\varphi}_i(t) \overline{\varphi}_i^* (t - \tau) dt \). By inspection of Fig. 5.2 we see that choosing the peak time-delay of the matched filter output, is equivalent to maximizing the quantity \( \int_0^T \overline{\varphi}_i(t) \overline{\varphi}_i^* (t - \tau) dt \).

We should note that for the real time-domain signal representation, we maximize the exponent of Eq. (5.17), obtaining the real representation of the matched filter, \( \int_0^T x_i(t) s_i(t - \tau_d) dt \).

### 5.3. The Likelihood Quantities for a Continuous Gaussian Data where the Covariance is Parameter-Independent

Given the Gaussian PDF defined in Eqs. (5.16) and (5.17), we can now evaluate the bias, the mean square error and the variance of an estimator for \( \theta \), the same way we did in section 4.1.2 for the parameter-independent covariance. There are mainly two differences between the derivation of 4.1.2 and the one to be performed. First, we are now dealing with continuous time as opposed to discrete samples. Second, both the parameter to be estimated and the observed data are now scalars rather than vectors. This simplifies the evaluation of the joint moments.
5.3.1. The Joint Moments of the Log – Likelihood Derivatives

Eqs. (5.20)-(5.29) provide the likelihood function joint moments, for the PDF’s defined in Eqs. (5.16) and (5.17) that are required to evaluate the bias, mean square error and the variance of a scalar estimator of \( \theta \). The index notations that were used in the parameter vector case are dropped since the parameter is scalar. The joint moments are first evaluated for the analytic complex signal and then for the real time-domain signal. In order to be consistent with the notation introduced in Chapter four we use \( \tilde{\mu}(t) \) as the Gaussian PDF mean for the analytic case and \( \mu(t) \) as the Gaussian PDF mean for the real case. The moments are, for the analytic case,

\[
\tilde{v}_{11} = \frac{1}{N_0} \int \left( \frac{\partial \tilde{\mu}(t)}{\partial \theta} \right)^* \left( \frac{\partial \tilde{\mu}(t)}{\partial \theta} \right) dt \tag{5.20}
\]

\[
\tilde{v}_{111} = -\frac{3}{2N_0} \left[ \int \left( \frac{\partial \tilde{\mu}(t)}{\partial \theta} \right)^* \left( \frac{\partial^2 \tilde{\mu}(t)}{\partial \theta^2} \right) dt + \int \left( \frac{\partial^2 \tilde{\mu}(t)}{\partial \theta^2} \right)^* \left( \frac{\partial \tilde{\mu}(t)}{\partial \theta} \right) dt \right] \tag{5.21}
\]

\[
\tilde{v}_{1111} = -\frac{2}{N_0} \left[ \int \left( \frac{\partial \tilde{\mu}(t)}{\partial \theta} \right)^* \left( \frac{\partial^3 \tilde{\mu}(t)}{\partial \theta^3} \right) dt + \int \left( \frac{\partial^3 \tilde{\mu}(t)}{\partial \theta^3} \right)^* \left( \frac{\partial \tilde{\mu}(t)}{\partial \theta} \right) dt \right] - \frac{3}{N_0} \left[ \int \left( \frac{\partial^2 \tilde{\mu}(t)}{\partial \theta^2} \right)^* \left( \frac{\partial^2 \tilde{\mu}(t)}{\partial \theta^2} \right) dt \right] \tag{5.22}
\]

\[
\tilde{v}_{1,1,1} = \frac{1}{N_0^2} \left( \int \left( \frac{\partial \tilde{\mu}(t)}{\partial \theta} \right)^* \left( \frac{\partial \tilde{\mu}(t)}{\partial \theta} \right) dt \right)^2 \tag{5.23}
\]

\[
\tilde{v}_{1,1,1} = 0 \tag{5.24}
\]

\[
\tilde{v}_{1,1} = \frac{1}{2N_0} \left[ \int \left( \frac{\partial \tilde{\mu}(t)}{\partial \theta} \right)^* \left( \frac{\partial^2 \tilde{\mu}(t)}{\partial \theta^2} \right) dt + \int \left( \frac{\partial^2 \tilde{\mu}(t)}{\partial \theta^2} \right)^* \left( \frac{\partial \tilde{\mu}(t)}{\partial \theta} \right) dt \right] \tag{5.25}
\]

\[
\tilde{v}_{1,1,1} = \frac{3}{N_0^2} \left( \int \left( \frac{\partial \tilde{\mu}(t)}{\partial \theta} \right)^* \left( \frac{\partial \tilde{\mu}(t)}{\partial \theta} \right) dt \right)^2 \tag{5.26}
\]
\[ \tilde{v}_{1,1,11} = \frac{3}{2N_0^2} \left[ \left( \frac{\partial^2 \tilde{\mu}(t)}{\partial \theta} \right) \frac{\partial^2 \tilde{\mu}(t)}{\partial \theta^2} \right] dt + \int \left( \frac{\partial^2 \tilde{\mu}(t)}{\partial \theta} \right) \frac{\partial \tilde{\mu}(t)}{\partial \theta} dt \left( \frac{\partial \tilde{\mu}(t)}{\partial \theta} \right) \frac{\partial \tilde{\mu}(t)}{\partial \theta} dt \] (5.27)

\[ \tilde{v}_{1,1,111} = \frac{1}{N_0^3} \left\{ \int \left( \frac{\partial\tilde{\mu}(t)}{\partial \theta} \right) \frac{\partial^2 \tilde{\mu}(t)}{\partial \theta^2} dt \right\}^3 \left[ \left( \frac{\partial^2 \tilde{\mu}(t)}{\partial \theta^2} \right)^2 \right] dt + \frac{1}{N_0^3} \left\{ \left( \frac{\partial \tilde{\mu}(t)}{\partial \theta} \right) \frac{\partial \tilde{\mu}(t)}{\partial \theta} dt \right\} \] (5.28)

\[ \tilde{v}_{1,1,111} = \frac{3}{2N_0^2} \left[ \left( \frac{\partial \tilde{\mu}(t)}{\partial \theta} \right) \frac{\partial \tilde{\mu}(t)}{\partial \theta^3} \right] dt + \int \left( \frac{\partial^3 \tilde{\mu}(t)}{\partial \theta^2} \right) \frac{\partial \tilde{\mu}(t)}{\partial \theta} dt \left( \frac{\partial \tilde{\mu}(t)}{\partial \theta} \right) \frac{\partial \tilde{\mu}(t)}{\partial \theta} dt \] (5.29)

and for the real case we obtain

\[ i_{11} = \frac{2}{N_0} \left\{ \int \left( \frac{\partial \mu(t)}{\partial \theta} \right)^2 dt \right\} \] (5.30)

\[ v_{111} = -\frac{6}{N_0} \left\{ \int \left( \frac{\partial \mu(t)}{\partial \theta} \right)^2 \left( \frac{\partial^2 \mu(t)}{\partial \theta^2} \right) dt \right\} \] (5.31)

\[ v_{1111} = -\frac{8}{N_0} \left\{ \int \left( \frac{\partial \mu(t)}{\partial \theta} \right)^2 \left( \frac{\partial^3 \mu(t)}{\partial \theta^3} \right) dt \right\} - \frac{6}{N_0} \left\{ \int \left( \frac{\partial^2 \mu(t)}{\partial \theta^2} \right)^2 dt \right\} \] (5.32)

\[ v_{1,1,11} = \frac{1}{N_0^2} \left\{ \int \left( \frac{\partial \mu(t)}{\partial \theta} \right)^2 dt \right\}^2 \] (5.33)

\[ v_{1,1,11} = 0 \] (5.34)

\[ v_{1,11} = \frac{2}{N_0} \left\{ \int \left( \frac{\partial \mu(t)}{\partial \theta} \right)^2 \left( \frac{\partial^2 \mu(t)}{\partial \theta^2} \right) dt \right\} \] (5.35)
\[ v_{1,1,1,1} = \frac{12}{N_0^2} \left( \int \left( \frac{\partial \mu(t)}{\partial \theta} \right)^2 \, dt \right)^2 \]  
(5.36)

\[ v_{1,1,1,1} = \frac{12}{N_0^2} \left( \int \left( \frac{\partial \mu(t)}{\partial \theta} \right) \left( \frac{\partial^2 \mu(t)}{\partial \theta^2} \right) \, dt \right) \left( \int \left( \frac{\partial \mu(t)}{\partial \theta} \right)^2 \, dt \right) \]  
(5.37)

\[ v_{1,1,1,1} = \frac{4}{N_0^2} \left( \int \left( \frac{\partial \mu(t)}{\partial \theta} \right)^2 \, dt \right) \left( \int \left( \frac{\partial^2 \mu(t)}{\partial \theta^2} \right)^2 \, dt \right) + 2 \left( \int \left( \frac{\partial \mu(t)}{\partial \theta} \right) \left( \frac{\partial^2 \mu(t)}{\partial \theta^2} \right) \, dt \right)^2 \]  
(5.38)

\[ v_{1,1,1,11} = \frac{12}{N_0^2} \left( \int \left( \frac{\partial \mu(t)}{\partial \theta} \right) \left( \frac{\partial^3 \mu(t)}{\partial \theta^3} \right) \, dt \right) \left( \int \left( \frac{\partial \mu(t)}{\partial \theta} \right)^2 \, dt \right) \]  
(5.39)

For simplification we now define the four integrals involved in the moments as

\[ \bar{I}_1 = \int \left( \frac{\partial \bar{\mu}(t)}{\partial \theta} \right) \left( \frac{\partial \bar{\mu}(t)}{\partial \theta} \right) \, dt \]  
(5.40)

\[ \bar{I}_2 = \int \left( \frac{\partial \bar{\mu}(t)}{\partial \theta} \right) \left( \frac{\partial^2 \bar{\mu}(t)}{\partial \theta^2} \right) \, dt \]  
(5.41)

\[ \bar{I}_3 = \int \left( \frac{\partial \bar{\mu}(t)}{\partial \theta} \right) \left( \frac{\partial^3 \bar{\mu}(t)}{\partial \theta^3} \right) \, dt \]  
(5.42)

\[ \bar{I}_4 = \int \left( \frac{\partial^2 \bar{\mu}(t)}{\partial \theta^2} \right) \left( \frac{\partial^2 \bar{\mu}(t)}{\partial \theta^2} \right) \, dt \]  
(5.43)

\[ I_1 = \int \left( \frac{\partial \mu(t)}{\partial \theta} \right)^2 \, dt \]  
(5.44)

\[ I_2 = \int \left( \frac{\partial \mu(t)}{\partial \theta} \right) \left( \frac{\partial^2 \mu(t)}{\partial \theta^2} \right) \, dt \]  
(5.45)

\[ I_3 = \int \left( \frac{\partial \mu(t)}{\partial \theta} \right) \left( \frac{\partial^3 \mu(t)}{\partial \theta^3} \right) \, dt \]  
(5.46)
\[ I_4 = \int \left( \frac{\partial^2 \mu(t)}{\partial \theta^2} \right)^2 dt. \]  

(5.47)

Next we substitute the integrals defined into Eqs. (5.40)-(5.43) and (5.44)-(5.47) into the joint moments derived in Eqs. (5.20)-(5.29) and (5.30)-(5.39) respectively. Eqs. (3.20)-(5.29) and (5.30)-(5.39) then become

\[ \bar{q}_{11} = \frac{1}{N_0} \bar{I}_1 \]  

(5.48)

\[ \bar{v}_{111} = -\frac{3}{N_0} \text{Re}\{\bar{I}_2\} \]  

(5.49)

\[ \bar{v}_{1111} = -\frac{4}{N_0} \text{Re}\{\bar{I}_3\} - \frac{3}{N_0} \bar{I}_4 \]  

(5.50)

\[ \bar{v}_{1,1,11} = \frac{1}{N_0^2} \bar{I}_1^2 = \bar{q}_{11}^2 \]  

(5.51)

\[ \bar{v}_{1,1,1} = 0 \]  

(5.52)

\[ \bar{v}_{1,11} = \frac{1}{N_0} \text{Re}\{\bar{I}_2\} \]  

(5.53)

\[ \bar{v}_{1,1,1,1} = \frac{3}{N_0^2} \bar{I}_1^2 = 3\bar{q}_{11}^2 \]  

(5.54)

\[ \bar{v}_{1,1,1,11} = \frac{3}{N_0} \text{Re}\{\bar{I}_2\} \right) * \frac{1}{N_0} \bar{I}_1 = -\bar{v}_{1111}\bar{q}_{11} \]  

(5.55)
\[
\tilde{v}_{1,1,1,11} = \frac{1}{N_0^2} \left( \tilde{I}_1 \tilde{I}_4 + 2 \text{Re} \left( \tilde{I}_2 \right)^2 \right) + \frac{1}{N_0^3} \tilde{I}_1^3
\]  \hspace{1cm} (5.56)

\[
\tilde{v}_{1,1,1,111} = \frac{3}{N_0^2} \text{Re} \left( \tilde{I}_3 \right) \tilde{I}_1
\]  \hspace{1cm} (5.57)

\[
i_{11} = \frac{2}{N_0} I_1
\]  \hspace{1cm} (5.58)

\[
v_{111} = -\frac{6}{N_0} I_2
\]  \hspace{1cm} (5.59)

\[
v_{1111} = -\frac{8}{N_0} I_3 - \frac{6}{N_0} I_4
\]  \hspace{1cm} (5.60)

\[
v_{1,1,11} = \frac{4}{N_0^2} I_1^2 = i_{11}^2
\]  \hspace{1cm} (5.61)

\[
v_{1,1,1} = 0
\]  \hspace{1cm} (5.62)

\[
v_{1,11} = \frac{2}{N_0} I_2
\]  \hspace{1cm} (5.63)

\[
v_{1,1,1,1} = \frac{12}{N_0^2} I_1^2 = 3i_{11}^2
\]  \hspace{1cm} (5.64)

\[
v_{1,1,1,11} = -v_{1111} i_{11}
\]  \hspace{1cm} (5.65)

\[
v_{1,1,1,111} = \frac{4}{N_0^2} \left( I_1 I_4 + 2I_2^2 \right) + \frac{8}{N_0^3} I_1^3
\]  \hspace{1cm} (5.66)

\[
v_{1,1,1,111} = \frac{12}{N_0^2} I_3 I_1
\]  \hspace{1cm} (5.67)

**5.3.1.1. The Bias, Mean – Square Error and Variance for the Scalar Case**

The bias for the single parameter case is the scalar version of the bias term derived in Chapter three, Eq. (3.12),

\[
b(\hat{\theta}) = E[\hat{\theta} - \theta] = \frac{1}{2} \left( i_{11} \right)^{-2} \left( v_{111} + 2v_{11,1} \right) \sqrt{\nu} + O_p \left( \nu^{-3/2} \right)
\]  \hspace{1cm} (5.68)

The corresponding mean-square error and variance are the scalar versions of the expressions derived in Chapter three, Eqs. (3.14) and (3.16) respectively,
\[
\text{MSE}(\hat{\theta}) = i_{11}^{-1} \nabla + \left[ 4i_{11}^{-4}v_{111}v_{1111} \left( n^1 \right) + \frac{5}{4} i_{11}^{-6}v_{1111}^2 v_{11111} \left( n^2 \right) + 3i_{11}^{-4}v_{111111} \left( n^3 \right) + 4i_{11}^{-5}v_{111111} \left( n^4 \right) + \frac{1}{3} i_{11}^{-5}v_{111111} \left( n^5 \right) + 8i_{11}^{-3}v_{111111} \left( n^6 \right) \right] \nabla \nabla + \sum_{i=1}^{4} v_{i111111} \left( n^2 \right)
\]

\[
\text{var}(\hat{\theta}) = \text{MSE}(\hat{\theta}) - b^2(\hat{\theta})
\]

where the orders of \( n \) in each term of the moment are determined by the number of integrals in its product. For example, the first two terms of \( v_{11111111} \),

\[
\frac{4}{N_0^2} \left( I_1 I_4 + 6I_2^2 \right),
\]

involve a two-integral product \( (I_1 I_4, I_2^2) \) therefore their order is \( n' n'' = n^2 \), whereas the third term \( \frac{8}{N_0^3} I_3^3 \) involves a three-integral product \( (I_3^3) \) therefore its order is \( n' n'' n''' = n^3 \).

Now we are in a stage where we have all the joint moments for the matched-filter scalar parameter estimation problem and the corresponding expressions for the bias, mean-square error and variance. Substituting the joint moments, Eqs. (5.48)-(5.57) and (5.58)-(5.67), into Eqs. (5.68)-(5.70), we obtain the following expressions for the bias, mean-square error and the variance for a scalar parameter estimation,

\[
b(\hat{\theta}) = -\frac{N_0}{2} \left( \frac{\text{Re}(\tilde{T}_2)}{I_1^2} \right) = -\frac{N_0}{4} \left( \frac{I_2}{I_1^2} \right) \quad (5.71)
\]

\[
\text{MSE}(\hat{\theta}) = \frac{N_0}{I_1} \left( \frac{\text{Re}(\tilde{T}_2)}{O_p(n^{-1})} \right) + \frac{15N_0^2}{4I_1^4} \left( \frac{\text{Re}(\tilde{T}_2)}{O_p(n^{-1})} \right)^2 - \frac{N_0^2}{I_1^3} \left( \frac{\text{Re}(\tilde{T}_3)}{O_p(n^{-1})} \right) = \frac{N_0}{2I_1} \left( \frac{I_2}{O_p(n^{-1})} \right) + \frac{15N_0^2 I_2^2}{16I_1^4} - \frac{N_0^2 I_3^3}{4I_1^3} \quad (5.72)
\]

\[
\text{var}(\hat{\theta}) = \frac{N_0}{I_1} \left( \frac{\text{Re}(\tilde{T}_2)}{O_p(n^{-1})} \right) + \frac{14N_0^2}{4I_1^4} \left( \frac{\text{Re}(\tilde{T}_2)}{O_p(n^{-1})} \right)^2 - \frac{N_0^2}{I_1^3} \left( \frac{\text{Re}(\tilde{T}_3)}{O_p(n^{-1})} \right) = \frac{N_0}{2I_1} \left( \frac{I_2}{O_p(n^{-1})} \right) + \frac{14N_0^2 I_2^2}{16I_1^4} - \frac{N_0^2 I_3^3}{4I_1^3} \quad (5.73)
\]

Next, we apply the above expressions to obtain the bias, mean-square error and variance for time delay and Doppler shift estimation problems by first replacing \( \theta \) with the time delay \( \tau \) and then by the Doppler shift \( f_D \).
5.3.2. Time Delay Estimation

In this problem the scalar time delay \( \tau \) is to be estimated. As shown before, the maximum likelihood estimate of the time delay \( \tau \) the time delay corresponds to the peak output of a matched filter for a signal in additive Gaussian noise. To derive the joint moments for \( \tau \) we simply evaluate the first three derivatives of the mean with respect to \( \tau \), substitute the results into the integrals \( \tilde{I}_1, \tilde{I}_2, \tilde{I}_3, \tilde{I}_4 \) for the complex representation, Eqs. (5.40)-(5.43), or \( I_1, I_2, I_3, I_4 \) for the real representation, Eqs. (5.44)-(5.47), and then evaluate the bias, means square error and variance by substituting the results into Eqs. (5.71)-(5.73).

We begin by evaluating the first three derivatives of the mean, \( \tilde{\mu}(t - \tau) \) for the complex case and \( \mu(t - \tau) \) for the real, with respect to \( \tau \). The derivatives are

\[
\frac{\partial \tilde{\mu}(t - \tau)}{\partial \tau} = -\tilde{\mu}'(t - \tau) \tag{5.74}
\]

\[
\frac{\partial^2 \tilde{\mu}(t - \tau)}{\partial \tau^2} = \tilde{\mu}''(t - \tau) \tag{5.75}
\]

\[
\frac{\partial^3 \tilde{\mu}(t - \tau)}{\partial \tau^3} = -\tilde{\mu}'''(t - \tau) \tag{5.76}
\]

\[
\frac{\partial \mu(t - \tau)}{\partial \tau} = -\mu'(t - \tau) \tag{5.77}
\]

\[
\frac{\partial^2 \mu(t - \tau)}{\partial \tau^2} = \mu''(t - \tau) \tag{5.78}
\]

\[
\frac{\partial^3 \mu(t - \tau)}{\partial \tau^3} = -\mu'''(t - \tau) \tag{5.79}
\]

Next we substitute Eqs. (5.74)-(5.76) into the integrals \( \tilde{I}_1, \tilde{I}_2, \tilde{I}_3, \tilde{I}_4 \) for the complex representation, Eqs. (5.40)-(5.43), and Eqs. (5.77)-(5.79) in \( I_1, I_2, I_3, I_4 \) for the real representation, Eqs. (5.44)-(5.47), and apply the Parseval’s theorem to the time domain integrals. We define \( \tilde{\mu} \leftrightarrow \Psi \) and \( \mu \leftrightarrow M \) as Fourier transform pairs for the complex time signals and the real time domain signal, respectively, and assume the signal’s
spectrum is symmetric about the carrier frequency. The following expressions are obtained

\[
\bar{T}_1 = \int \left( \frac{\partial \tilde{\mu}(t)}{\partial t} \right)^* \left( \frac{\partial \tilde{\mu}(t)}{\partial t} \right) dt = (2\pi)^2 \int f^2 \| \tilde{\mu}(f) \|^2 df = (2\pi)^2 \left\{ \int f^2 |\tilde{\psi}(f_c + v)|^2 df + f_c^2 \int |\tilde{\psi}(f_c + v)|^2 df \right\} \tag{5.80}
\]

\[
\bar{T}_2 = \int \left( \frac{\partial \tilde{\mu}(t)}{\partial t} \right)^* \left( \frac{\partial^2 \tilde{\mu}(t)}{\partial t^2} \right) dt = j(2\pi)^3 \int f^2 |\tilde{\psi}(f)|^2 df = j(2\pi)^3 f_c \left\{ \int f^2 |\tilde{\psi}(f_c + v)|^2 df + f_c^2 \int |\tilde{\psi}(f_c + v)|^2 df \right\} \tag{5.81}
\]

\[
\bar{T}_3 = \int \left( \frac{\partial \tilde{\mu}(t)}{\partial t} \right)^* \left( \frac{\partial^3 \tilde{\mu}(t)}{\partial t^3} \right) dt = -j(2\pi)^4 \int f^4 |\tilde{\psi}(f)|^2 df = -(2\pi)^4 f_c^4 \left\{ \int f^4 |\tilde{\psi}(f_c + v)|^2 df + f_c^2 \int |\tilde{\psi}(f_c + v)|^2 df \right\} \tag{5.82}
\]

\[
\bar{T}_4 = \int \left( \frac{\partial^2 \tilde{\mu}(t)}{\partial t^2} \right)^* \left( \frac{\partial^2 \tilde{\mu}(t)}{\partial t^2} \right) dt = (2\pi)^4 \int f^4 |\tilde{\psi}(f)|^2 df = -\bar{T}_3 \tag{5.83}
\]

For real time-domain signals the integrals become

\[
I_1 = \int \left( \frac{\partial \mu(t)}{\partial t} \right)^2 dt = (2\pi)^2 \int \| \mu(f) \|^2 df = 2(2\pi)^2 \left\{ \int f^2 |\psi(f_c + v)|^2 df \right\} \tag{5.84}
\]

\[
I_2 = \int \left( \frac{\partial \mu(t)}{\partial t} \right)^* \left( \frac{\partial \mu(t)}{\partial t} \right) dt = j(2\pi)^3 \int f^3 |\psi(f)|^2 df = 0 \tag{5.85}
\]
\[ I_3 = \int \left( \frac{\partial \mu(t)}{\partial t} \right)^2 \left( \frac{\partial^2 \mu(t)}{\partial t^2} \right) \, dt = -(2\pi)^4 \int_{-\infty}^{\infty} f^4 \left| M(f) \right|^2 \, df = -2(2\pi)^4 \int_{-f_c}^{f_c} f^4 \left| M(f_c + \nu) \right|^2 \, d\nu + \int_{-f_c}^{f_c} (2\pi)^4 \left| M(f_c + \nu) \right|^2 \, d\nu \]

\[ I_4 = \int \left( \frac{\partial^2 \mu(t)}{\partial t^2} \right)^2 \, dt = (2\pi)^4 \int_{-\infty}^{\infty} f^4 \left| M(f) \right|^2 \, df = -I_3. \]  

(5.86)

(5.87)

5.3.2.1. The Bias Term to First Order

To evaluate the bias we need to substitute the values for the integrals \( \bar{I}_1, I_1, \text{Re} \{ \bar{I}_2 \}, I_2 \) into Eq. (5.71). Noting that \( \text{Re} \{ \bar{I}_2 \} = I_2 = 0 \), we obtain an identically zero first order bias for the time delay problem.

5.3.2.2. The MSE and the Variance to Second Order

To evaluate the mean-square error and the variance we need all four integrals. Noting that \( \text{Re} \{ \bar{I}_2 \} = I_2 = 0 \) and \( \text{Re} \{ \bar{I}_3 \} = \bar{I}_3 = -\bar{I}_4 \) the following results are obtained

\[ \text{MSE}(\hat{\tau}) = \frac{N_0}{\bar{I}_1} \nabla \nabla - \left( \frac{N_0}{\bar{I}_1} \right)^2 \frac{\text{Re} \{ \bar{I}_3 \} \nabla \nabla + O_p(n^{-3})}{\bar{I}_1} = \frac{N_0}{2I_1} \nabla \nabla + \left( \frac{2N_0}{I_1} \right)^2 \frac{I_4 \nabla \nabla + O_p(n^{-3})}{I_1} \]

\[ \text{var}(\hat{\tau}) = \text{MSE}(\hat{\tau}) - b^2(\hat{\tau}) = \text{MSE}(\hat{\tau}) \]

(5.88)

(5.89)

where \( \bar{I}_1, \bar{I}_3, I_1, I_4 \) are evaluated in Eqs. (5.80), (5.82), (5.84) and (5.87) respectively.
Example 5.1

Let us assume that we have a Gaussian base-banded signal with a constant energy. Its waveform can be represented as

$$g(t) = \frac{1}{\sqrt{s}} \exp\left(-\pi \frac{t^2}{s^2}\right) \quad |t| \leq T/2$$  \hspace{1cm} (5.90)

Evaluating the two integrals required for the variance, $I_1$ and $I_4$, under the assumption that $s \leq T$ (required for approximating the integrals' span $[-T/2, T/2]$ to $[-\infty, \infty]$) we obtain

$$I_1 \approx \frac{\pi}{\sqrt{2s^2}}$$  \hspace{1cm} (5.91)

$$I_4 \approx \frac{6\pi^2}{2\sqrt{2s^4}}$$  \hspace{1cm} (5.92)

Substituting Eqs. (5.91) and (5.92) into Eq. (5.88) we obtain the first two orders of variance for the time delay estimator

$$\text{var}(\hat{\tau}) = \frac{N_0 \sqrt{2}}{2} \frac{s^2}{\pi} + \left(\frac{N_0}{2}\right)^2 \frac{6}{\pi} s^2 \cdot O_p(n^{-1}) + O_p(n^{-2})$$  \hspace{1cm} (5.93)

Evaluating the signal's energy we find that it equals $\frac{1}{\sqrt{2}}$. Observing Eq. (5.93) we can see that the first and second order terms equalize when the SNR, $\frac{2E}{N_0}$, is 3, or 5 dB. Therefore, for SNR's less than 3, the second order term is higher than the first and hence is definitely not negligible. Moreover, since $\frac{1}{s}$ can be determined as a measure of the signals bandwidth, decreasing $s$, or increasing the signal's bandwidth, will decrease both first and second order variance terms, i.e., improve the time-delay estimate. ◊
Let us examine the expressions obtained for the mean-square error and the variance in a different way. We make the following definitions,

\[ 2E = \int_{-f_c}^{f_c} |\psi(v + f_c)|^2 dv \]  
(5.94)

\[ \tilde{\beta} = \sqrt{\frac{(2\pi)^2 \int_{-f_c}^{f_c} v^2 |\psi(v + f_c)|^2 dv}{2E}} \]  
(5.95)

\[ \tilde{\gamma} = \sqrt{\frac{(2\pi)^4 \int_{-f_c}^{f_c} v^4 |\psi(v + f_c)|^2 dv}{2E}} \]  
(5.96)

\[ E = 2 \int_{-f_c}^{f_c} \left( M(v + f_c) \right)^2 dv \]  
(5.97)

\[ \beta = \sqrt{\frac{2 \cdot (2\pi)^2 \int_{-f_c}^{f_c} v^2 M(v + f_c)^2 dv}{E}} \]  
(5.98)

\[ \gamma = \sqrt{\frac{2 \cdot (2\pi)^4 \int_{-f_c}^{f_c} v^4 M(v + f_c)^2 dv}{E}} \]  
(5.99)

\[ 2\pi f_c = \omega_c \]  
(5.100)

where \( E \) is the total energy of the real signal and \( \beta \) is commonly defined as the signal's root-mean square bandwidth. It is important to note that the values of \( \beta \) and \( \gamma \) as defined for the real and complex cases have the same value, thus, \( \beta = \tilde{\beta} \) and \( \gamma = \tilde{\gamma} \).

We can now define the integrals \( \bar{I}_1, \bar{I}_2, \bar{I}_3, \bar{I}_4 \) and \( I_1, I_2, I_3, I_4 \) by substituting Eqs. (5.94) - (5.96) and (5.100) into Eqs. (5.80)-(5.83) and Eqs. (5.97) - (5.100) into Eqs. (5.84)-(5.86),

\[ \bar{I}_1 = 2E\left(\tilde{\beta}^2 + \omega_c^2\right) \]  
(5.101)

\[ \bar{I}_2 = j2E\omega_c\left(\tilde{\beta}^2 + \omega_c^2\right) \]  
(5.102)
\[ \tilde{I}_3 = -2E \left( \gamma^4 + 6\omega_c^2 \beta^2 + \omega_c^4 \right) \]  
(5.103)

\[ \tilde{I}_4 = 2E \left( \gamma^4 + 6\omega_c^2 \beta^2 + \omega_c^4 \right) \]  
(5.104)

\[ I_1 = E \left( \beta^2 + \omega_c^2 \right) \]  
(5.105)

\[ I_2 = 0 \]  
(5.106)

\[ I_3 = -E \left( \gamma^4 + 6\omega_c^2 \beta^2 + \omega_c^4 \right) \]  
(5.107)

\[ I_4 = E \left( \gamma^4 + 6\omega_c^2 \beta^2 + \omega_c^4 \right). \]  
(5.108)

Substituting Eqs. (5.101), (5.103), (5.105), (5.108) into Eqs. (5.88) and (5.89) we obtain the following result,

\[
\text{var}(\hat{t}) = \text{MSE}(\hat{t}) = \frac{1}{\left(2E/N_0\right)\left(\beta^2 + \omega_c^2\right)} \nabla \nabla + \frac{1}{\left(2E/N_0\right)^2\left(\beta^2 + \omega_c^2\right)^3} \nabla \nabla + O_p\left(n^{-3}\right).
\]

(5.109)

Observing Eq. (5.109) we can see that for a base-banded signal (\(\omega_c=0\)) the first order term is proportional to the inverse of \(\beta^2\), while the second term is proportional to the ratio \(\gamma^4/\left(\beta^2\right)^3\). Increasing the root - mean square bandwidth, while maintaining the SNR, \(\left(2E/N_0\right)\), will then result in decreasing the first order term of the variance. As for the second order variance term, since it involves both \(\gamma\) and \(\beta\), its behavior cannot be determined generally.

### 5.3.3. Doppler Shift Estimation

In this problem the scalar narrowband signal's Doppler shift \(f_D\) is to be estimated. As opposed to the time delay estimation, the maximum likelihood estimation of the Doppler shift \(f_D\) is not equivalent to matched filter processing. The main reason is that a Doppler shifted signal is dilated, hence changing its bandwidth, hence the mean of the PDF is no longer the time delayed transmitted signal, but a time delayed dilated signal.
that we do not necessarily know in advance. To derive the joint moments for $f_D$ we simply evaluate the first three derivatives of the mean with respect to $f_D$, the same way we did for time delay estimation, substitute the results into the integrals $\tilde{I}_1, \tilde{I}_2, \tilde{I}_3, \tilde{I}_4$ for the complex representation, Eqs. (5.40)-(5.43), or $I_1, I_2, I_3, I_4$ for the real representation, Eqs. (5.44)-(5.47), and then evaluate the bias, mean square error and variance by substituting the results into Eqs. (5.71)-(5.73).

We assume the waveform is narrowband waveform so that it can be represented as $\bar{\mu}(t) = \bar{g}(t)e^{j2\pi(f_c + f_D)}$, or $\mu(t) = g(t)\cos(2\pi(f_c + f_D))$ for the real case.

We begin by evaluating the first three derivatives of the mean, $\bar{\mu}(t)$ for the complex case and $\mu(t)$ for the real case with respect to $f_D$. The derivatives are

\[
\frac{\partial \bar{\mu}(t)}{\partial f_D} = j2\pi \bar{\mu}(t) \tag{5.110}
\]

\[
\frac{\partial^2 \bar{\mu}(t)}{\partial f_D^2} = -(2\pi)^2 \bar{\mu}(t) \tag{5.111}
\]

\[
\frac{\partial^3 \bar{\mu}(t)}{\partial f_D^3} = -j(2\pi)^3 \bar{\mu}(t) \tag{5.112}
\]

\[
\frac{\partial \mu(t)}{\partial f_D} = -2\pi g(t)\sin(2\pi(f_c + f_D)) \tag{5.113}
\]

\[
\frac{\partial^2 \mu(t)}{\partial f_D^2} = -(2\pi)^2 g(t)\cos(2\pi(f_c + f_D)) = -(2\pi)^2 \mu(t) \tag{5.114}
\]

\[
\frac{\partial^3 \mu(t)}{\partial f_D^3} = (2\pi)^3 g(t)\sin(2\pi(f_c + f_D)) \tag{5.115}
\]

Next we substitute Eqs. (5.110)-(5.112) in the integrals $\tilde{I}_1, \tilde{I}_2, \tilde{I}_3, \tilde{I}_4$ for the complex representation, Eqs. (5.40)-(5.43), and Eqs. (5.113)-(5.115) in $I_1, I_2, I_3, I_4$ for the real representation, Eqs. (5.44)-(5.47), and apply the Parseval's theorem on the frequency domain integrals, the same way we did before for the time delay problem. The following expressions for the integrals are obtained
\[ \tilde{I}_1 = \int \left( \frac{\partial \mu(t)}{\partial f_D} \right) \left( \frac{\partial \mu(t)}{\partial f_D} \right) dt = \frac{T}{2} \int \frac{t^2 |\bar{g}(t)|^2 dt}{-T/2} \quad (5.116) \]

\[ \tilde{I}_2 = \int \left( \frac{\partial \mu(t)}{\partial f_D} \right)^* \left( \frac{\partial \mu^2(t)}{\partial f_D^2} \right) dt = j(2\pi)^3 \int \frac{t^4 |\bar{g}(t)|^2 dt}{-T/2} \quad (5.117) \]

\[ \tilde{I}_3 = \int \left( \frac{\partial \mu(t)}{\partial f_D} \right)^* \left( \frac{\partial \mu^3(t)}{\partial f_D^3} \right) dt = -(2\pi)^4 \int \frac{t^4 |\bar{g}(t)|^2 dt}{-T/2} \quad (5.118) \]

\[ \tilde{I}_4 = \int \left( \frac{\partial^2 \mu(t)}{\partial f_D^2} \right)^* \left( \frac{\partial \mu^2(t)}{\partial f_D^2} \right) dt = (2\pi)^4 \int \frac{t^4 |\bar{g}(t)|^2 dt}{-T/2} = -\tilde{I}_3. \quad (5.119) \]

For the real time-domain signals the integrals become

\[ I_1 = \int \left( \frac{\partial \mu(t)}{\partial f_D} \right)^2 dt = 0.5*(2\pi)^2 \int \frac{t^2 g^2(t) dt}{-T/2} \quad (5.120) \]

\[ I_2 = \int \left( \frac{\partial \mu(t)}{\partial f_D} \right) \left( \frac{\partial \mu^2(t)}{\partial f_D^2} \right) dt = 0 \quad (5.121) \]

\[ I_3 = \int \left( \frac{\partial \mu(t)}{\partial f_D} \right) \left( \frac{\partial^3 \mu(t)}{\partial f_D^3} \right) dt = -0.5*(2\pi)^4 \int \frac{t^4 g^2(t) dt}{-T/2} \quad (5.122) \]

\[ I_4 = \int \left( \frac{\partial^2 \mu(t)}{\partial f_D^2} \right)^2 dt = 0.5*(2\pi)^4 \int \frac{t^4 g^2(t) dt}{-T/2} = -I_3 \quad (5.123) \]

5.3.3.1. The Bias Term to First Order

To evaluate the bias we need to substitute the values for the integrals \( \tilde{I}_1, I_1, \text{Re}\{\tilde{I}_2\}, I_2 \) into Eq. (5.71). Noting that similar to the time delay problem \( \text{Re}\{\tilde{I}_2\} = I_2 = 0 \), we obtain an identically zero first order bias for the Doppler shift problem.
5.3.3.2. The MSE and the Variance to Second Order

To evaluate the mean-square error and the variance we need all four integrals. Noting that since we obtain the following relations \( \text{Re}\{I_2\} = I_2 = 0 \) and \( \text{Re}\{I_3\} = \tilde{I}_3 = -\tilde{I}_4 \) that are similar to those of the time delay problem, the expressions we obtain for the mean-square error and variance are similar to the ones obtained for the time delay problem

\[
\text{MSE}(\hat{f}_D) = \frac{N_0}{I_1} \nabla + \left( \frac{N_0}{I_1} \right)^2 \frac{\text{Re}\{I_3\}}{I_1} \nabla + O_p(n^{-3}) = \frac{N_0}{2I_1} \nabla + \left( \frac{N_0}{2I_1} \right)^2 \frac{I_4}{I_1} \nabla + O_p(n^{-3})
\]

(5.88)

\[
\text{var}(\hat{f}_D) = \text{MSE}(\hat{f}_D) - b^2(\hat{\tau}) = \text{MSE}(\hat{f}_D)
\]

(5.89)

but with different values for the integrals: \( \tilde{I}_1, \tilde{I}_3, I_1, I_4 \) are evaluated in Eqs. (5.116), (5.118), (5.120) and (5.123) respectively.

**Example 5.2**

Let us assume we have the same Gaussian base-band signal with constant energy that we had in example 5.1. Its waveform can be represented as

\[
g(t) = \frac{1}{\sqrt{s}} \exp\left(-\pi \left(\frac{t^2}{s^2}\right)\right) \quad |t| \leq \frac{T}{2}
\]

(5.90)

Evaluating the two integrals required for the variance, \( I_1 \) and \( I_4 \), under the assumption that \( s \leq T \), required for approximating the integrals' span \([-T/2, T/2]\) to \([-\infty, \infty]\), we obtain

\[
I_1 \approx \frac{\pi s^2}{2\sqrt{2}} \quad \text{(5.124)}
\]

\[
I_4 \approx \frac{6\pi^2 s^4}{4\sqrt{2}} \quad \text{(5.125)}
\]

Substituting Eqs. (5.124) and (5.125) into Eq. (5.88) we obtain the first two orders of variance for the Doppler shift estimator.
\[ \text{var}(\hat{f}_D) = \frac{N_0}{2} \frac{2\sqrt{2}}{\pi} \frac{1}{s^2} \left[ \frac{24}{\pi} \frac{1}{s^2} \right]. \] (5.126)

The signal's energy equals \( 1/\sqrt{2} \) as mentioned in Ex. 5.1. Observing Eq. (5.126) we can see that the first and second order equalize terms when the SNR, \( \frac{2E}{N_0} \), is 6, or 7.8 dB. Therefore, for SNR's less than 6, the second order term is higher than the first and hence is definitely not negligible. Moreover, since \( \sqrt{s} \) can be determined as a measure of the signals bandwidth, increasing \( s \), or decreasing the signal's bandwidth, will decrease both first and second order variance terms, i.e., improve the Doppler shift estimate.

Let us examine the expressions obtained for the mean-square error and the variance the same way we did for the time delay problem. We make the following definitions,

\[ 2E = \int_{-T/2}^{T/2} |\bar{g}(t)|^2 dt \] (5.127)

\[ \bar{\alpha} = \sqrt{\frac{\int_{-T/2}^{T/2} (2\pi)^2 t^2 |\bar{g}(t)|^2 dt}{2E}} \] (5.128)

\[ \bar{\delta} = \sqrt{\frac{\int_{-T/2}^{T/2} (2\pi)^4 t^4 |\bar{g}(t)|^2 dt}{2E}} \] (5.129)

\[ 2E = \int_{-T/2}^{T/2} g(t)^2 dt \] (5.130)
\[ \alpha = \sqrt{\frac{(2\pi)^2}{4E} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int g(t)^2 \, dt} \]  
(5.131)

\[ \delta = \sqrt{\frac{(2\pi)^4}{4E} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int f(t)^2 \, dt} \]  
(5.132)

It is important to point out that the values of \( \alpha \) and \( \delta \), as defined above for the real and complex cases, are equal. We can now define the integrals \( \bar{I}_1, \bar{I}_2, \bar{I}_3, \bar{I}_4 \) and \( I_1, I_2, I_3, I_4 \) by substituting Eqs. (5.127) – (5.129) into Eqs. (5.116)-(5.119) and Eqs. (5.130) – (5.132) into Eqs. (5.120)-(5.123),

\[ \bar{I}_1 = 2E\alpha^2 \]  
(5.133)

\[ \bar{I}_2 = j2E2\pi\alpha^2 \]  
(5.134)

\[ \bar{I}_3 = -2E\delta^4 \]  
(5.135)

\[ \bar{I}_4 = 2E\delta^4 \]  
(5.136)

\[ I_1 = E\alpha^2 \]  
(5.137)

\[ I_2 = 0 \]  
(5.138)

\[ I_3 = -E\delta^4 \]  
(5.139)

\[ I_4 = E\delta^4 \]  
(5.140)

Substituting (5.133), (5.135), (5.137), (5.140) into (5.88) we obtain the following result,

\[ \text{var}(\hat{f}_d) = \text{MSE}(\hat{f}_d) = \frac{1}{(2E/N_0)\alpha^2} \nabla \nabla + \frac{1}{\alpha^6} \frac{\delta^4}{O_p(n^{-1})} \nabla \nabla + O_p\left(n^{-3}\right). \]  
(5.141)
5.4. First and Second Order Terms of the Variance for Three Specific Waveforms

Having derived expressions for the variance to second order for both the time delay and the Doppler shift estimators, Eq. (5.88), we can now implement these results for specific waveforms. I chose to implement the variance expression on three specific waveforms. The waveforms picked are the Gaussian waveform, the Linear Frequency Modulation (LFM) waveform and the Hyperbolic Frequency Modulation (HFM) waveform. All waveforms are demodulated, since the phase of the signals is seldom known.

Since both the first and second order variance terms depend on many parameters such as SNR, the bandwidth of the signal, the time duration of the signal, I decided to pick two parameters, namely the bandwidth of the signal and the SNR, and observe how they affect both first and second order terms. The analysis is therefore separated in two parts: first, I set a specific SNR (0 dB) and observe how changing the bandwidth of the signal affects the first and second order variance terms of its time delay and Doppler shift estimators; then, I set a specific bandwidth (100 Hz) and observe how changing the SNR affects the first and second order variance terms of its time delay and Doppler shift estimators.

We begin by defining the waveforms.

The Gaussian signal is the signal introduced in Exs. 5.1 and 5.2, is defined as

\[ g(t) = \frac{1}{\sqrt{s}} \exp\left(-\pi \frac{t^2}{s^2}\right) \quad |t| \leq T/2 \]  

(5.142)

where \( T = 1 \) sec. The bandwidth of this signal can be roughly represented by \( 1/s \) as indicated in Exs. 5.1 and 5.2. This signal is selected because the first and second order terms of its variance can be derived analytically with relative ease, as already done in Exs. 5.1 and 5.2.

The second signal is the LFM signal\textsuperscript{18}, which is defined by

\[ \mu(t) = \cos \left( \omega_0 t + \frac{1}{2} bt^2 \right) \quad |t| \leq T/2 \] 

(5.143)
where $T=1$ sec and $\omega_0$ is the carrier frequency. The bandwidth of the signal is determined by $\frac{bT}{2\pi}$. Demodulating the signal is done by multiplying it by $\cos(\omega_0 t)$ and passing it through a low pass filter. This signal is extensively used in radar and sonar applications, since its gain over an incoherent linear bandpass filter with the same bandwidth equals the time-bandwidth product. Moreover, since its spectrum is very close to rectangular shape it is easy to be generated.

The third signal is the HFM signal, which is defined by

$$\mu(t) = \sin(\alpha * \log(1 - k(t + T/2))) \quad |t| \leq T/2$$

(5.144)

where $T=1$ sec, $k = \frac{f_2 - f_1}{f_2 T}$, $\alpha = \frac{2\pi f_1}{k}$ and $f_1$ and $f_2$ are the frequencies of the signal's spectrum is within. Demodulating the signal is done by multiplying it by $\cos(2\pi f_0 t)$, where $f_0 = \sqrt{f_1 f_2}$, and passing it through a low pass filter. Moreover, in order to control frequency domain sidelobes, hence reduce wiggles in the spectrum the signal is often multiplied by a window function, or taper. The taper I used is the modified Tukey window that has the form

$$w(t) = \begin{cases} 
  p + (1-p)\sin^2\left(\pi \frac{t + T/2}{2T_w}\right) & 0 \leq t \leq T_w \\
  1 & T_w \leq t \leq T - T_w \\
  p + (1-p)\sin^2\left(\pi \frac{(t + T/2) - (T - 2T_w)}{2T_w}\right) & T - T_w \leq t \leq T 
\end{cases}$$

(5.145)

where $T_w$ is the window duration, 0.125T in my taper, and p is the pedestal for sources, 0.1 in my taper. This signal is extensively used in radar and sonar applications due to being Doppler tolerant, which means that the matched filter response to an echo from a moving point target will have a large peak value, even if the return Doppler shift dilated signal is processed under the assumption that the target is stationary. Alternatively, it is

---

18 DiFranco, p. 153.
19 For further information about the LFM signal see Rihaczek, chaps. 6,7.
20 Orcutt, Pp. 97-98.
the frequency modulated signal that has the optimal correlation between the transmitted signal and the Doppler shifted signal\textsuperscript{21}.

The following graphs show the waveform and spectrum of each of the three signals.

![Gaussian Basebanded Signal Waveform - 100Hz Bandwidth](image)

![Gaussian Basebanded Signal Spectrum - 100Hz Bandwidth](image)

\textbf{Figure 5.3 – Gaussian Signal Waveform and Spectrum}

\textsuperscript{21} For further information about the HFM signal see Drumheller and Kroszczyński.
Figure 5.3 – LFM Signal Waveform and Spectrum
Figure 5.5 – HFM Signal Waveform and Spectrum
Next, we observe how the bandwidth of the signal affects the first and second orders of its time delay and Doppler shift estimators, for the three different signals. The bandwidth dependence is presented in Figs. 5.6-5.8 for the time delay estimator for the three signals, and in Figs. 5.9-5.11 for the Doppler shift estimator. The SNR is 0 dB for all Figures. Each figure presents the first order variance, labeled as '*', and the second order variance, labeled as 'o', for a range of bandwidths starting from 1 Hz to 1000 Hz. The results were evaluated using MATLAB and the script is provided in appendix A.

Following are key features observed for the time delay problem:

- As bandwidth increases both first and second order terms of the time delay variance tend to decrease, so that the resolution increases. While the two terms monotonically decrease for the Gaussian signal, as derived analytically in Ex. 5.1, their behavior for the LFM and HFM signals is less monotonic. They both oscillate, especially the LFM signal. The HFM signal's first and second order variance terms are similar to the Gaussian ones for bandwidths exceeding 100 Hz, while the LFM signal's first and second order variance terms match the Gaussian ones for bandwidths exceeding 400 Hz. The reason for the LFM and HFM behavior is that as bandwidth increases, both approach a Gaussian shape, so that their first and second order variance terms eventually meet the first and second order variance terms of the Gaussian signal.

- For bandwidths smaller than 200 Hz the first order variance is the smallest for the LFM signal and the greatest for the Gaussian signal, while the second order variance is the smallest for the Gaussian signal and greatest for the HFM signal.

- In the entire range of bandwidths at the 0 dB SNR, the second order term is greater than the first order one.

For the Doppler shift estimator, we observe the following key features:

- For the Gaussian signal the both terms monotonically increase as bandwidth increases, as derived analytically in Ex. 5.2. For the LFM and the HFM signals both terms follow oscillations that decay to constant values of approximately 0.6 Hz$^2$ for the first order term and 2 Hz$^2$ for the second order term. The reason that both the LFM and HFM signals variance terms converge to a constant value is that the first and second order variance terms are solely determined by the integrals
\[ \int_{-\infty}^{\infty} \left( \frac{\partial M(f)}{\partial f} \right)^2 df \quad \text{and} \quad \int_{-\infty}^{\infty} \left( \frac{\partial^2 M(f)}{\partial f^2} \right)^2 df \], respectively, as indicated by Eqs. (5.120) and (5.121). As the bandwidth increases, both signals obtain the same rectangular spectrum, therefore both integrals are dominated by the rectangular edges derivatives of the signal's spectrum, which become constant from a specific bandwidth.

- In the entire range of bandwidths for the specific 0 dB SNR, the second order term is greater than the first order one.
Figure 5.6 – Gaussian Signal – Time-Delay Variance Terms as a Function of Bandwidth for 0 dB SNR
Figure 5.7 – LFM Signal – Time-Delay Variance Terms as a Function of Bandwidth for 0 dB SNR
First Order and Second Order Variance vs. Bandwidth for an HFM Signal - Time Delay

Figure 5.8 – HFM Signal – Time-Delay Variance Terms as a Function of Bandwidth for 0 dB SNR
First Order and Second Order Variance vs. Bandwidth for a Gaussian Signal - Doppler Shift

Figure 5.9 – Gaussian Signal – Doppler Shift Variance Terms as a Function of Bandwidth for 0 dB SNR
Figure 5.10 – LFM Signal – Doppler Shift Variance Terms as a Function of Bandwidth for 0 dB SNR
First Order and Second Order Variance vs. Bandwidth for an HFM Signal - Doppler Shift

Figure 5.11 – HFM Signal – Doppler Shift Variance Terms as a Function of Bandwidth for 0 dB SNR
We now show how the SNR of the signal affects the first and second order variance terms of its time delay and Doppler shift estimators, for the three different signals. The SNR dependence is presented in Figs. 5.12-5.14, 5.18-5.19 for the time delay estimator for the three signals, and in Figs. 5.15-5.17 5.20-5.21 for the Doppler shift estimator. The bandwidth is 100 Hz for Figs. 5.12-5.17, representing sonar signals, whereas 10 MHz (and 1msec duration) for Figs. 5.18-5.21, representing a radar signal. Each figure presents the first order variance, labeled as '*', and the second order variance, labeled as 'o', for a range of SNR starting from -20 dB to 40 dB. The results were evaluated using MATLAB and the script is provided in appendix A.

Following are key features observed for both the time delay and the Doppler shift first and second order variance terms:

- As the SNR increases, both first order and second order variance terms decrease.
- The log-log slope of the log of the second order variance term is double that of the first order variance term. The reason is that observing Eq. (5.88) we can see that the second order term is proportional to \( \left( \frac{N_0}{2} \right)^2 \), which is \(-20\log_{10}\left( \frac{N_0}{2} \right)\) on a logarithmic scale, while the first order term is proportional to \( \frac{N_0}{2} \) which is \(-10\log_{10}\left( \frac{N_0}{2} \right)\) on a logarithmic scale. Therefore, by knowing the values of both terms at a specific bandwidth and SNR, we are able to determine their values at the same bandwidth for all SNR's. It is important to note that this type of relation in the log-log scale between the first order variance term and the second order variance term is expected, since the first order term reflects the \( n^{-1} \) dependency terms in the asymptotic expansion, while the second order term reflects the \( n^{-2} \) dependency terms in this expansion, so that the second order variance term slope is expected to double that of the first order variance term.

Table 5.1 provides the SNR's values where,

a) both first order variance and second order variance terms equalize,

b) the first order variance term is greater than the second order variance term by a factor of ten,
for the three signals and for both time delay and Doppler shift problems. It is important to note that the values SNR's where the second order term is less by a factor of ten than the first order term, are interpreted as the lowest SNR's required to regard the estimate as linear.

<table>
<thead>
<tr>
<th></th>
<th>Gaussian Signal</th>
<th>LFM Signal</th>
<th>HFM Signal</th>
<th>LFM Signal</th>
<th>HFM Signal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bandwidth</strong></td>
<td>100 Hz</td>
<td>100 Hz</td>
<td>100 Hz</td>
<td>10 MHz</td>
<td>10 MHz</td>
</tr>
<tr>
<td><strong>Time Delay</strong></td>
<td>5 dB</td>
<td>22 dB</td>
<td>8 dB</td>
<td>4 dB</td>
<td>3 dB</td>
</tr>
<tr>
<td><strong>Doppler Shift</strong></td>
<td>8 dB</td>
<td>6 dB</td>
<td>5 dB</td>
<td>5 dB</td>
<td>5 dB</td>
</tr>
</tbody>
</table>

(a) SNR's required for both terms to equalize

<table>
<thead>
<tr>
<th></th>
<th>Gaussian Signal</th>
<th>LFM Signal</th>
<th>HFM Signal</th>
<th>LFM Signal</th>
<th>HFM Signal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bandwidth</strong></td>
<td>100 Hz</td>
<td>100 Hz</td>
<td>100 Hz</td>
<td>10 MHz</td>
<td>10 MHz</td>
</tr>
<tr>
<td><strong>Time Delay</strong></td>
<td>15 dB</td>
<td>32 dB</td>
<td>18 dB</td>
<td>14 dB</td>
<td>13 dB</td>
</tr>
<tr>
<td><strong>Doppler Shift</strong></td>
<td>18 dB</td>
<td>16 dB</td>
<td>15 dB</td>
<td>15 dB</td>
<td>15 dB</td>
</tr>
</tbody>
</table>

(b) SNR's required for linearity

**Table 5.1**

The important result that we obtain is that for a both small bandwidth signal, 100 Hz, and large bandwidth one, 10 MHz, both time-delay and Doppler shift estimates cannot be regarded as linear until we are in SNR's of approximately 18 dB, with an exceptional requirement of 32 dB for the 100 Hz bandwidth LFM time-delay estimate.
First Order and Second Order Variance vs. SNR for a Gaussian Signal - Time Delay

Figure 5.12 – Gaussian Signal – Time-Delay Variance Terms as a Function of SNR for 100 Hz Bandwidth
First Order and Second Order Variance vs. SNR for a LFM Signal - Time Delay

Figure 5.13 – LFM Signal – Time-Delay Variance Terms as a Function of SNR for 100 Hz Bandwidth
First Order and Second Order Variance vs. SNR for a HFM Signal - Time Delay

![Graph showing the relationship between Variance Terms, s^2, and SNR, dB. The graph includes data points for First Order Term and Second Order Term.](image)

**Figure 5.14** – HFM Signal – Time-Delay Variance Terms as a Function of SNR for 100 Hz Bandwidth
Figure 5.15 — Gaussian Signal – Doppler Shift Variance Terms as a Function of SNR for 100 Hz Bandwidth
Figure 5.16 – LFM Signal – Doppler Shift Variance Terms as a Function of SNR for 100 Hz Bandwidth
Figure 5.17 – HFM Signal – Doppler Shift Variance Terms as a Function of SNR for 100 Hz Bandwidth
Figure 5.18 – LFM Radar Signal – Time-Delay Variance Terms as a Function of SNR for 10 MHz Bandwidth
Figure 5.19 – HFM Radar Signal – Time-Delay Variance Terms as a Function of SNR for 10 MHz Bandwidth
Figure 5.20 – LFM Radar Signal – Doppler Shift Variance Terms as a Function of SNR for 10 MHz Bandwidth
Figure 5.21 – HFM Radar Signal – Doppler Shift Variance Terms as a Function of SNR for 10 MHz Bandwidth
6. Summary and Conclusions

In this thesis we applied higher order asymptotics to derive general expressions for the first order bias and second order error correlation and covariance of a general multivariate maximum likelihood estimate. The second order covariance derived is a tighter bound than the Cramer-Rao lower bound, the most widely used bound in statistics, but requires more effort to implement. The expressions for the first order bias and second order covariance of the maximum likelihood estimate are of a fundamental and completely general nature so that they can be used in any statistical estimation problem.

We started by deriving expressions for the first order bias and second order error correlation and covariance of a general multivariate maximum likelihood estimate. Then, we evaluated these expressions for a multivariate maximum likelihood estimate from Gaussian data. First, the first order bias was evaluated for general multivariate Gaussian data, where both the data covariance matrix and the data mean depend on the estimated parameters. Second, both the second order correlation error and covariance were evaluated for two less general but very practical cases. In the first case, the data covariance matrix is independent of the parameter we estimate, and in the second the data mean is zero. The first case encompasses deterministic signal in additive noise while the second describes fully randomized signal in noise.

Applying the results obtained on the maximum likelihood time delay estimate, known as the matched filter processor, and the maximum likelihood Doppler shift estimate we obtained analytic expressions for the bias, mean-square error and variance of the time delay and Doppler shift estimators. Then, these expressions were evaluated for three specific signals, namely the Gaussian signal, LFM (Linear Frequency Modulation) signal and HFM (Hyperbolic Frequency Modulation) signal.

The following conclusions, regarding the time delay and Doppler shift estimates, were obtained:

- Both maximum likelihood time delay estimate and Doppler shift estimate are unbiased to first order.
• Both first order and second order time delay resolution increase with increasing signal bandwidth.

• For the Gaussian signal both orders of the Doppler shift resolution monotonically decrease as bandwidth increases because the slope of its spectrum decreases. For the LFM and the HFM signals both orders approach constant values as bandwidth increases.

• As the SNR increases, both first order and second order variance terms decrease.

• The log-log slope of the log of the second order variance term is double that of the first order variance term. This relation is expected, since the first order term reflects the $n^1$ dependency terms in the asymptotic expansion, while the second order term reflects the $n^2$ dependency terms in this expansion.

• For the three signals tested the SNR must be in excess of roughly 20-30 dB before the variance of the time delay or Doppler shift estimate can attain the Cramer-Rao bound.
Appendix A – Matlab Scripts


% This procedure evaluates the first and second order variance terms
% for the time-delay and Doppler shift estimation problems. Each term is evaluated
% for Gaussian HFM and LFM signals for different bandwidths. The SNR is 0dB for all cases.

clear;
close all;

delta_tau=1; % signal's duration
dt=delta_tau/16000; % sample time
t=-delta_tau/2:dt:delta_tau/2; % signal's energy
Energy_in=.5;
NO=1; % noise energy
noise_var=NO/2;
Tw=delta_tau/8; % HFM carrier frequency
p=.1; % HFM taper parameter
[B,A] = butter(10,.15); % HFM lowpass filter
w0=0; % Gaussian and LFM carrier frequency

w=ones(1,length(t)); % HFM taper parameter
w(1:Tw/dt)=p+(1-p).*exp(-pi/2/Tw.*t(1:Tw/dt));
w(Tw/dt+1:(deltatau-Tw)/dt)=p;
w((deltatau-Tw)/dt+1:deltatau/dt)=p+(1-p).*exp(-pi/2/Tw.*t((deltatau-Tw)/dt+1:deltatau/dt));

for s=0:.025:3; % bandwidth counter
    tau=(10^(-s)); % Gaussian bandwidth
    signal_innorm1=cos(w0.*t)./tau.*exp(-pi.*t.^2./tau.^2); % Gaussian waveform
    Energy_in_norm1=signal_innorm1*signal_innorm1';
    Al=(Energy_in_norm1).5;
    signal1=Al.*signal_innorm1;
    diff1sr=diff(signal1)/dt;
    diff1sc=diff1sr';
    diff2s2r=diff(signal1,2)/(dt^2);
    diff2s2c=diff2s2r';
    I1_1=diff1sr*diff1sc*dt;
    I4_1=diff2s2r*diff2s2c*dt;
    first_order1=noise_var*I1_1^(-1);
    second_order1=(first_order1^2)*I4_1/I1_1;
    I1D_1=5*(2*pi)^2*sum((signal1.*t^2).*sum((signal1.*t^2)^2).*sum((signal1.*t^2)^2).*sum((signal1.*t^2)^2))*dt;
    I4D_1=5*(2*pi)^4*sum((signal1.*t^2)^2).*sum((signal1.*t^2)^2).*sum((signal1.*t^2)^2).*sum((signal1.*t^2)^2))*dt;
    first_order_dl=noise_var*I1D_1^(-1);
    second_order_dl=(first_order_d1^2)*I4D_1/I1D_1;

    f1=f0-(10^s)*2/2; % lower frequency
    f2=f0+(10^s)*2/2;
    k=(f2-f1)/(delta_tau); % upper frequency
    % HFM parameter
    alpha=-2*pi*f0/k; % HFM parameter
    signal_innorm2=sin(alpha.*log(1-k.*(t+delta_tau/2))).*w.*cos(2*pi*f0.*t); % HFM waveform
signal_innorm2=filtfilt(B,A,signal_innorm2);
Energy_in_norm2=signal_innorm2*signal_innorm2';
A2=(Energy_in_norm2).5;
signal2=signal_innorm2.*A2;
diff2s2r=diff(signal2)/dt;
diff2s2c=diff2s2r';
I1_2=diff2s2r*diff2s2c*dt;
I4_2=diff2s2r*diff2s2c*dt;
first_order2=noise_var*I1_2^(-1);
second_order2=(first_order2^2)*I4_2/I1_2;  % second order variance - time delay
I1D_2=.5*(2*pi)^2*sum((signal2.*t).*(signal2.*t'))*dt;  % I1 evaluation - Doppler shift
I4D_2=.5*(2*pi)^4*sum((signal2.*t.^2).*(signal2.*t.^2'))*dt;  % I4 evaluation - Doppler shift
first_order_d2=noise_var*I1D_2^(-1);  % first order variance - Doppler shift
second_order_d2=(first_order_d2^2)*I4D_2/I1D_2;  % second order variance - Doppler shift

% LFM signal
b=(10^8)/delta_tau;  % LFM bandwidth
signal_in_norm3=cos(2*pi*f0.*t)+(b.*t.^2).*cos(2*pi*f0.*t);  % LFM waveform
Energy_in_norm3=signal_in_norm3*signal_in_norm3'*dt;  % energy normalization
A3=Energy_in_norm3/Energy_in_norm3'*.5;
signal3=signal_in_norm3.*A3;  % energy normalization
diff3sr=diff(signal3)/dt;  % first derivative
diff3sc=diff3sr';  % second derivative
I1_3=diff3sr*diff3sc*dt;  % I1 evaluation - time delay
I4_3=diff3sc*diff3sc*dt;  % I4 evaluation - time delay
first_order_d3=noise_var*I1_3^(-1);  % first order variance - time delay
second_order_d3=(first_order_d3^2)*I4_3/I1D_3;  % second order variance - time delay

% plots
figure(1);
grid on;
title('First Order and Second Order Variance vs. Bandwidth for a Gaussian Signal - Time Delay');
loglog(1/tau,first_order1,'*'),1/tau,second_order1,'o');
xlabel('Bandwidth,Hz');
ylabel('Variance Terms,s^2');
legend('First Order Term','Second Order Term',1);
axis([1 1000 1e-8 1e4]);
hold on;
figure(2);
grid on;
title('First Order and Second Order Variance vs. Bandwidth for a Gaussian Signal - Doppler Shift');
loglog(1/tau,first_order_d1,'*'),1/tau,second_order_d1,'o');
xlabel('Bandwidth,Hz');
ylabel('Variance Terms,Hz^2');
legend('First Order Term','Second Order Term',2);
hold on;
figure(3);
grid on;
title('First Order and Second Order Variance vs. Bandwidth for an HFM Signal - Time Delay');
loglog((f2-f1),first_order2,'*'),(f2-f1),second_order2,'o');
xlabel('Bandwidth,Hz');
ylabel('Variance Terms,s^2');
legend('First Order Term','Second Order Term',1);
axis([5 1000 1e-8 1e4]);
hold on;
figure(4);
grid on;
title('First Order and Second Order Variance vs. Bandwidth for an HFM Signal - Doppler Shift');
loglog((f2-f1),first_order_d2,'*'),(f2-f1),second_order_d2,'o');
xlabel('Bandwidth,Hz');
ylabel('Variance Terms,Hz^2');
legend('First Order Term','Second Order Term');
hold on;
figure(5);
grid on;
title('First Order and Second Order Variance vs. Bandwidth for an LFM Signal - Time Delay');
loglog(b/2/pi*delta_tau,first_order3,'*'),b/2/pi*delta_tau,second_order3,'o');
xlabel('Bandwidth,Hz');
ylabel('Variance Terms,s^2');
axis([1 1000 1e-8 1e4]);
legend('First Order Term','Second Order Term',1);
hold on;
figure(6);
% waveforms and spectrum plots
y=15999;
f=-1/dt/2:1/dt/(y-1):1/dt/2;
S1=fft(signal1,y)*dt;
Spos1=S1(1:(y+1)/2);
Sneg1=S1((y+1)/2+1:y);
Sfixed1=[Sneg1 Spos1];
S2=fft(signal2,y)*dt;
Spos2=S2(1:(y+1)/2);
Sneg2=S2((y+1)/2+1:y);
Sfixed2=[Sneg2 Spos2];
S3=fft(signal3,y)*dt;
Spos3=S3(1:(y+1)/2);
Sneg3=S3((y+1)/2+1:y);
Sfixed3=[Sneg3 Spos3];

figure(7);
subplot(2,1,1);
plot(t,signal1);
grid on;
title('Gaussian Basebanded Signal Waveform - 100Hz Bandwidth');
ylabel('Amplitude');
xlabel('Time [sec]');
axis([-15e-3 15e-3 0 9]);
subplot(2,1,2);
plot(f,10.*log10(abs(Sfixed1)));
title('Gaussian Basebanded Signal Spectrum - 100Hz Bandwidth');
ylabel('Amplitude [dB]');
xlabel('Frequency [Hz]');
axis([-100 100 -50 0]);
grid on;

figure(8);
subplot(2,1,1);
plot(t,signal2);
grid on;
title('HFM Basebanded Signal Waveform - 100Hz Bandwidth');
ylabel('Amplitude');
xlabel('Time[sec]');
axis([- .5 .5 -1.1 1.1]);
subplot(2,1,2);
plot(f,10.*log10(abs(Sfixed2)));
title('HFM Basebanded Signal Spectrum - 100Hz Bandwidth');
ylabel('Amplitude [dB]');
xlabel('Frequency [Hz]');
axis([-100 100 -50 0]);
grid on;

figure(9);
subplot(2,1,1);
plot(t,signal3);
grid on;
title('LFM Basebanded Signal Waveform - 100Hz Bandwidth');
ylabel('Amplitude');
xlabel('Time[sec]');
axis([- .5 .5 -1 1]);
subplot(2,1,2);
plot(f,10.*log10(abs(Sfixed3)));
title('LFM Basebanded Signal Spectrum - 100Hz Bandwidth');
A2. SNR Dependence Code – 100 Hz Bandwidth

% This procedure evaluates the first and second order variance terms
% for the time-delay and Doppler shift estimation problems. Each term is evaluated
% for Gaussian HFM and LFM signals for different SNR's. The bandwidth is 100Hz for all cases.

clear;
close all;

delta_tau=1;
dc=delta_tau/16000;
t=-delta_tau/2:dc:delta_tau/2;
NO=1;
noise_var=NO/2;
f0=7000;
Tw=delta_tau/8;
p=.1;
[B,A] = butter(10,.15);
w0=0;
s=2;
tau=(10^(-s));
f1=f0-(10^s)/2;
f2=f0+(10^s)/2;
k=(f2-f1)/f2/deltatau;
alpha=-2*pi*f1/k;

for q=-2:.1:4;
    Energy_in=.5*10^q;
SNR_in=10*log10(Energy_in/noise_var);

    % Gaussian signal
    signal_in_norm=cos(w0.*t)./tau.*exp((-pi.*t.^2)/(tau^2));  % Gaussian waveform
    Energy_in_norm=signal_in_norm.*signal_in_norm'*dt;
    A1=(Energy_in/Energy_in_norm).^5;
    signal1=A1.*signal_in_norm;
    diff1sr=diff(signal1)/dt;
    diff1sc=diff1sr';
    diff1s2r=diff(signal1,2)/(dt^2);
    diff1s2c=diff1s2r';
    I1_1=diff1sr.*diff1s2r*dt;
    I2_1=diff1sr.*diff1s2c*dt;
    first_order1=noise_var*I1_1^(-1);
    second_order1=(first_order1^2);I4_1/I1_1;    % first order variance - time delay
    I1D_1=.5*(2*pi)^2*sum((signal1.*t).*(signal1.*t).^2).*dt;
    I4D_1=.5*(2*pi)^4*sum((signal1.*t.^2).*(signal1.*t.^2).*dt);  % second order variance - time delay
    first_order_d1=noise_var*I1D_1^(-1);
    second_order_d1=(first_order_d1^2)*I4D_1/I1D_1;   % second order variance - Doppler shift

    % HFM Signal
    signal_in_norm2=sin(alpha.*log1-k.*t) + delta_tau/2))').*w.*cos(2*pi*f0.*t);  % HFM waveform
    signal_in_norm2=filter(B,A,signal_in_norm2);
    Energy_in_norm2=signal_in_norm2.*signal_in_norm2'*dt;
    A2=(Energy_in/Energy_in_norm2).^5;
    signal2=signal_in_norm2.*A2;
    diff2sr=diff(signal2)/dt;
    diff2sc=diff2sr';
    diff2s2r=diff(signal2,2)/(dt^2);
    diff2s2c=diff2s2r';

95
\[ I_{1_2} = \text{diff2sr} \times \text{diff2sc} \times dt; \]
\[ I_{4_2} = \text{diff2s2r} \times \text{diff2s2c} \times dt; \]
\[ \text{first_order2} = \text{noise_var} \times I_{1_2} \times \text{dt}; \]
\[ \text{second_order2} = (\text{first_order2}^2) \times I_{4_2} \times I_{1_2}; \]
\[ \text{I1D}_2 = 0.5 \times (2 \pi)^2 \times \text{sum}((\text{signal1} \times t) \times (\text{signal1} \times t)) \times dt; \]
\[ \text{I4D}_2 = 0.5 \times (2 \pi)^4 \times \text{sum}((\text{signal1} \times t^2) \times (\text{signal1} \times t^2)) \times dt; \]
\[ \text{first_order_d2} = \text{noise_var} \times \text{I1D}_2 \times \text{dt}; \]
\[ \text{second_order_d2} = (\text{first_order_d2}^2) \times I_{4D_2} \times I_{1D_2}; \]

% LFM signal

\[ \text{signal} \_\text{in}\_\text{norm3} = \cos(2 \times \pi \times f_0 \times t + 0.5 \times b \times t^2) \times \cos(2 \pi \times f_0 \times t); \]
\[ \text{energy} \_\text{in}\_\text{norm3} = \text{filtfilt(B,A,signal} \_\text{in}\_\text{norm3}); \]
\[ A_3 = \left(\frac{\text{energy} \_\text{in} \_\text{norm3}}{\text{energy} \_\text{in} \_\text{norm3}}\right)^{0.5}; \]
\[ \text{signal3} = \text{signal} \_\text{in}\_\text{norm3} \times A_3; \]
\[ \text{diff3sr} = \text{diff}((\text{signal3}) \times dt); \]
\[ \text{diff3sc} = \text{diff3sr}; \]
\[ \text{diff3s2r} = \text{diff}((\text{signal3}) \times dt); \]
\[ \text{diff3s2c} = \text{diff3s2r}; \]
\[ \text{I1}_3 = \text{diff3sr} \times \text{diff3sc} \times dt; \]
\[ \text{I4}_3 = \text{diff3s2r} \times \text{diff3s2c} \times dt; \]
\[ \text{first_order3} = \text{noise_var} \times \text{I1}_3 \times \text{dt}; \]
\[ \text{second_order3} = (\text{first_order3}^2) \times I_{4D_3} \times I_{1D_3}; \]

% plots

figure(1);
grid on;
title('First Order and Second Order Variance vs. SNR for a Gaussian Signal - Time Delay');
semilogy(SNR_in,first_order1,'*',SNR_in,second_order1,'o');
xlabel('SNR, dB');
ylabel('Variance Term, s^2');
legend('First Order Term', 'Second Order Term',1);
axis([-20 40 1e-12 1e4]);
hold on;
figure(2);
grid on;
title('First Order and Second Order Variance vs. SNR for a Gaussian Signal - Doppler Shift');
semilogy(SNR_in,first_order1,'*',SNR_in,second_order1,'o');
xlabel('SNR, dB');
ylabel('Variance Term, Hz^2');
legend('First Order Term', 'Second Order Term',1);
axis([-20 40 1e-8 1e10]);
hold on;
figure(3);
grid on;
title('First Order and Second Order Variance vs. SNR for a HFM Signal - Time Delay');
semilogy(SNR_in,first_order2,'*',SNR_in,second_order2,'o');
xlabel('SNR, dB');
ylabel('Variance Term, s^2');
legend('First Order Term', 'Second Order Term',1);
axis([-20 40 1e-12 1e4]);
hold on;
figure(4);
grid on;
title('First Order and Second Order Variance vs. SNR for a HFM Signal - Doppler Shift');
semilogy(SNR_in,first_order2,'*',SNR_in,second_order2,'o');
xlabel('SNR, dB');
ylabel('Variance Term, Hz^2');
legend('First Order Term', 'Second Order Term',1);
axis([-20 40 1e-8 1e10]);
hold on;
figure(5);
grid on;
title('First Order and Second Order Variance vs. SNR for a LFM Signal - Time Delay');
semilogy(SNR_in,first_order3,'*',SNR_in,second_order3,'o');
xlabel('SNR, dB');
ylabel('Variance Term, s^2');
legend('First Order Term', 'Second Order Term',1);
A3. SNR Dependence Code – 10 MHz Bandwidth

% This procedure evaluates the first and second order variance terms 
% for the time-delay and Doppler shift estimation problems for HFM & LFM Radar signals 
% of 10MHz bandwidth for different SNR's.

clear;
close all;

deltatau=1e-3; % signal duration
dt=1/16000/10^4; % sample time
t=-deltatau/2:dt:deltatau/2;
N0=1; % noise energy
s=7; % general bandwidth parameter
b=2*pi*(10^s)/deltatau; % LFM bandwidth
f0=7000*10^4; % HFM carrier frequency
Tw=deltatau/8; % HFM taper parameter
P=.1; % HFM taper parameter
[B,A] = butter(10,.15); % HFM lowpass filter
sdf=(f2-f1)/f2/deltatau; % HFM parameter
alpha=-2*pi*fl/k; % HFM parameter
w(t)=[(2*pi)^2*sum((signal2.*t.2).*dt);].% HFM Taper
w(Tw/dt1)=p+(l-p).*(sin(pi/2/Tw.*t(Tw/Tw/dt1))).^2; % HFM Signal
w(Tw/dt1+dtTw/dt1)=p+(l-p).*(sin(pi/2/Tw...% energy normalization
*1*(t(Tw+dtTw)/dt1+dtTw/dt1+dtTw/dt1))."2;

for q=-2:.1:4; % energy counter
    Energy_in=.5*10^q;
    SNR_in=10*log10(Energy_in/noise_var);
    % HFM Signal
    signal_in_norm2=alpha.*log(l-k.*(t+deltatau/2))).*w.*cos(2*pi*f0.*t); % HFM waveform
    Energy_in_norm2=signal_in_norm2.*filtfilt(B,A,signal_in_norm2); % energy normalization
    signal2=signal_in_norm2.*A2; % first derivative
    first_order2=noise_var*I1_2^-1; % second derivative
    second_order2=(first_order2^2)*I4_2/I1_2; % first order variance - time delay
    first_order_d2=(first_order_d2^2)*I4D_2/I1D_2; % second order variance - time delay
% LFM Signal
signal_in_norm3=cos(2.*pi.*f0.*t+5*b.*t.^2).*cos(2*pi*f0.*t); % LFM waveform
signal_in_norma3=filtfilt(B,A,signal_in_norm3);
Energy_in_norma3=signal_in_norma3*signal_in_norma3'*dt;
A3=(Energy_in/Energy_in_norma3).^0.5;
signal3=signal_in_norma3.*A3; % energy normalization

diff3sr=diff(signal3)/dt; % first derivative
diff3sc=diff3sr';
diff3s2r2=diff3sr*(dt^2);
diff3s2c=diff3s2r2';
I1_3=diff3sr*diff3sc*dt; % I1 evaluation - time delay
I4_3=diff3s2r2*diff3s2c*dt;
first_order3=Noise_Var*I1_3^-1; % first order variance - time delay

second_order3=(first_order3^2)*I4_3/I1_3; % second order variance - time delay

I1D_3=.5*(2*pi)^2*sum((signal3.*t).*signal3.*t')*dt; % I1 evaluation - Doppler shift
I4D_3=.5*(2*pi)^4*sum((signal3.*t.^2).*signal3.*t.^2')*dt; % I4 evaluation - Doppler shift

first_orderd3=Noise_Var*I1D_3^-1; % first order variance - Doppler shift

second_orderd3=(first_orderd3^2)*I4D_3/I1D_3; % second order variance - Doppler shift

% plots
figure(1);
grid on;
title('First Order and Second Order Variance vs. SNR for a LFM Signal - Time Delay');

hold on;
figure(2);
grid on;
title('First Order and Second Order Variance vs. SNR for a LFM Signal - Doppler Shift');

hold on;
figure(3);
grid on;
title('First Order and Second Order Variance vs. SNR for a HFM Signal - Time Delay');

hold on;
figure(4);
grid on;
title('First Order and Second Order Variance vs. SNR for a HFM Signal - Doppler Shift');

end
Bibliography


