Spring 2000

# Cramer-Rao bound and optimal amplitude estimator of superimposed sinusoidal signals with unknown frequencies

Shaohui Jia

# CRAMER-RAO BOUND AND OPTIMAL AMPLITUDE ESTIMATOR OF SUPERIMPOSED SINUSOIDAL SIGNALS WITH UNKNOWN FREQUENCIES

by

Shaohui Jia, M.S.

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

COLLEGE OF ENGINEERING AND SCIENCE
LOUISIANA TECH UNIVERSITY

May 2000

UMI Number: 9963600



Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

**LOUISIANA TECH UNIVERSITY**

**THE GRADUATE SCHOOL**

May 10, 2000
Date

We hereby recommend that the dissertation prepared under our supervision by Shaohui Jia

entitled Cramer-Rao Bound and Optimal Amplitude Estimator of Superimposed Sinusoidal Signals with Unknown Frequencies

be accepted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

Supervisor of Dissertation Research

Head of Department

Applied Computational Analysis and Modeling
Department

Recommendation concurred in:

Advisory Committee

Approved: Director of Graduate Studies

Dean of the College

Approved: Director of the Graduate School

GS Form 13
2/97

# ABSTRACT

This dissertation addresses optimally estimating the amplitudes of superimposed sinusoidal signals with unknown frequencies. The Cramer-Rao Bound of estimating the amplitudes in white Gaussian noise is given, and the maximum likelihood estimator of the amplitudes in this case is shown to be asymptotically efficient at high signal to noise ratio but finite sample size. Applying the theoretical results to signal resolutions, it is shown that the optimal resolution of multiple signals using a finite sample is given by the maximum likelihood estimator of the amplitudes of signals.

## APPROVAL FOR SCHOLARLY DISSEMINATION

The author grants to the Prescott Memorial Library of Louisiana Tech University the right to reproduce, by appropriate methods, upon request, any or all portions of this Dissertation. It is understood that "proper request" consists of the agreement, on the part of the requesting party, that said reproduction is for his personal use and that subsequent reproduction will not occur without written approval of the author of this Dissertation. Further, any portions of the Dissertation used in books, papers, and other works must be appropriately referenced to this Dissertation.

Finally, the author of this Dissertation reserves the right to publish freely, in the literature, at any time, any or all portions of this Dissertation.

Author Jia, Shaohui

Date May 12, 2000

GS Form 14
2/97

# TABLE OF CONTENTS

# LIST OF FIGURES

# NOTATION

| | |
|---|---|
| $\mathrm{Re}(\bullet)$ | real part of a complex number, vector, or matrix |
| $\mathrm{Im}(\bullet)$ | imaginary part of a complex number, vector, or matrix |
| $(\bullet)^T$ | transposition of a vector or matrix |
| $(\bullet)^*$ | conjugate of a vector or matrix |
| $(\bullet)^H$ | conjugated transposition of a vector or matrix |
| $E(\bullet)$ | expectation of a random variable |
| $\lvert\bullet\rvert$ | absolute value of a complex number |
| $\lVert\bullet\rVert$ | norm of a vector |
| $\mathbf{I}$ | identity matrix |

# ACKNOWLEDGMENT

I wish to express my gratitude and sincere appreciation to my advisor in ACAM, Dr. Huaijin Gu, for directing this dissertation. Without his time, patience, and encouragement, this dissertation could not have been possible.

I would like to thank Dr. Richard Greechie for his comments on this dissertation, especially for his kindness and financial support that provided me the opportunity to study at Louisiana Tech University. Special thanks are due to my advisor in Computer Science, Dr. Ben Choi, for the time he invested in discussing and suggesting improvements to this dissertation. Thanks must also go to the other committee members, Dr. Louis Roemer for his advice, and Dr. Raja Nassar for his service on the advisory committee.

I also appreciate the help from Margaret Dunn for formatting this dissertation.

Finally and most importantly, the encouragement and support of my wife and family through the graduate program is acknowledged.

# CHAPTER 1

# INTRODUCTION

We deal with the problem of optimally estimating the amplitudes of superimposed sinusoidal signals in this dissertation, which occurs in a variety of fields ranging from radar, sonar, oceanography, and seismology to medical imaging and radio-astronomy. The study of this problem has been the subject of numerous books and papers, and its history dates back at least 200 years.

In this chapter, we model and formulate the problem, review some of its classical methods, and present a summary of the content and the contributions of this dissertation.

This chapter is organized as follows. Section 1.1 presents several examples. Section 1.2 then presents a detailed model and formulation of the problem. Section 1.3 reviews the previous work on the problem. Finally, Section 1.4 presents an outline of this dissertation.

## 1.1 Background

Our formulation of the superimposed sinusoidal signal problem was motivated by several specific problems that we now briefly describe.

### 1.1.1 The Passive Sensor Array Problem

Consider a passive sensor array composed of $N$ sensors with arbitrary locations and arbitrary directional characteristics. Assume that $K$ radiating sources are located in the far-field of the array. The far-field assumption implies that the wavefronts received by the array can be well modeled as planewaves. Assuming for simplicity that the array and the sources are confined to a plane, it follows that the position of the $k$-th source is characterized by a single parameter--its direction-of-arrival $\theta_k$.

With this parameterization, the signal received by the $i$-th sensor can be expressed as

$$x_i(t) = \sum_{k=1}^{K} a_i(\theta_k) s_k(t - \tau_i(\theta_k)) + w_i(t) \tag{1.1}$$

where $s_k(\bullet)$ is the signal of the $k$-th wavefront as observed at a reference point in the array, $a_i(\theta_k)$ is the amplitude response of the $i$-th sensor to a wavefront impinging from direction $\theta_k$, $\tau_i(\theta_k)$ is the propagation delay between the reference point and the $i$-th sensor for a wavefront impinging from direction $\theta_k$, and $w_i(\bullet)$ is the white noise at the $i$-th sensor.

Certain simplifications of (1.1) arise if the signals are narrowband and have the same unknown center frequency, say $\omega_0$. In this case, the $k$-th signal can be expressed as

$$s_k(t) = u_k(t)\cos(\omega_0 t + v_k(t)) \tag{1.2}$$

where $u_k(\bullet)$ and $v_k(\bullet)$ represent slowly varying signals that modulate the amplitude and phase of $s_k(\bullet)$. In the context of passive sensor arrays, there is a more specific restriction; it is assumed that the modulating signals do not change significantly during the time it takes for the wavefront to propagate across the array, implying that the following approximation is valid:

$$s_k(t-\tau_i(\theta_k)) \approx u_k(t)\cos(\omega_0(t-\tau_i(\theta_k))+v_k(t)) \quad (1.3)$$

The form of (1.3) is still not very convenient for our purposes. A simpler form results when the complex representation is used. For any real signal *x(t),* the Fourier transform of *x(t)*, $X(\omega)$, is given by

$$X(\omega)=\int_{-\infty}^{\infty} x(t)e^{-j\omega t}dt \quad (1.4)$$

which has the following property:

$$X(\omega)=(X(-\omega))^{*} \quad (1.5)$$

As a result of (1.5), we restrict our attention only to the half line and define

$$X_{+}(\varpi)=\begin{cases} 2X(\omega) & \omega>0 \\ 0 & otherwise \end{cases} \quad (1.6)$$

The inverse function of $X_{+}(\varpi)$ is given by

$$x_{+}(t)=\int_{0}^{\infty} X_{+}(\omega)e^{j\omega t}d\omega \quad (1.7)$$

With these definitions, the complex representation of the signal *x(t),* $\tilde{x}(t)$, is defined as

$$\tilde{x}(t)=e^{-j\omega_0 t}x_{+}(t) \quad (1.8)$$

From (1.4) to (1.8), it follows that $\tilde{s}_k(t)=u_k(t)e^{jv_k(t)}$ which implies that (1.3) can be rewritten as

$$\tilde{s}_k(t-\tau_i(\theta_k)) \approx \tilde{s}_k(t)e^{-j\varpi_0\tau_i(\theta_k)} \quad (1.9)$$

Notice the simplicity resulting from the complex representation; the time-delay has been transformed to multiplication by an exponential.

Using (1.9), we can now rewrite (1.1) as

$$\tilde{x}_i(t) = \sum_{k=1}^{K} a_i(\theta_k) e^{-j\varpi_0 \tau_i(\theta_k)} \tilde{s}_k(t) + \tilde{w}_i(t)$$

$$= \sum_{k=1}^{K} \tilde{a}_i(\theta_k, t) e^{-j\varpi_0 \tau_i(\theta_k)} + \tilde{w}_i(t) \qquad (1.10)$$

where

$$E(\tilde{w}_i(t)\tilde{w}_j^*(t)) = E(w_i(t)w_j^*(t)) = \sigma^2 \delta(i-j) \qquad (1.11)$$

$$E(\tilde{w}_i(t)\tilde{w}_j(t)) = E(w_i(t)w_j(t)) = 0 \qquad (1.12)$$

where $\delta(\bullet)$ is a delta function, which is given below

$$\delta(\theta) = \begin{cases} 1 & \theta = 0 \\ 0 & \theta \neq 0 \end{cases}$$

Given a snapshot of $N$ complex-valued samples, the passive sensor array problem is to estimate the directions-of-arrival $\theta_1, \cdots \theta_K$ and the amplitudes $\tilde{a}_1(\bullet), \cdots, \tilde{a}_K(\bullet)$ of the impinging wavefronts.

### 1.1.2 The Harmonic Retrieval Problem

Consider a signal $x(\bullet)$ composed of $K$ sinusoids with unknown frequencies $\varpi_1, \cdots \varpi_K$ embedded in additive noise,

$$x(t) = \sum_{k=1}^{K} a_k \cos(\varpi_k t + \phi_k) + w(t) \qquad (1.13)$$

where $a_k$ and $\phi_k$ are the amplitude and phase of the $k$-th sinusoid.

Let a tapped-delay-line with $N$ equally spaced taps $D$ delay units apart be used to sample the signal. The signal at the $(i+1)$-th tap is given by

$$x(t-iD)=\sum_{k=1}^{K}a_k\cos(\varpi_k(t-iD)+\phi_k)+w_i(t) \tag{1.14}$$

As in the sensor array problem, it turns out that the complex representation is more convenient. Using the complex representation, we can express (1.14) as

$$\tilde{x}(t-iD)=\sum_{k=1}^{K}\tilde{a}_k e^{-j\varpi_k iD}+\tilde{w}_i(t) \tag{1.15}$$

where $\tilde{a}_k = a_k e^{j(\varpi_k t+\phi_k)}$, (1.11) and (1.12) still hold for $\tilde{w}_i(\bullet)$.

Given a snapshot of $N$ complex-valued samples, the harmonic retrieval problem is to estimate the frequencies $\varpi_1,\cdots\varpi_K$ and the amplitudes $\tilde{a}_1,\cdots,\tilde{a}_K$ of the sinusoids.

### 1.1.3 The Pole Retrieval Problem

Consider a linear system that is excited by an impulse. Assume that the system has $K$ unknown poles at locations $s_1,\cdots,s_K$ in the complex plane, $s_k=\alpha_k+j\varpi_k$. Assuming that all the poles are distinct, the response of the system can be expressed as

$$x(t)=\sum_{k=1}^{K}a_k e^{\alpha_k t}\cos(\varpi_k t+\phi_k)+w(t) \tag{1.16}$$

where $a_k$ is the residue at the $k$-th pole, $\phi_k$ is the phase at the $k$-th pole, and $w(t)$ is the white noise. Assuming that the response is sampled by a tapped-delay-line with $N$ equally spaced taps $D$ delay units apart, the output at the $(i+1)$-th tap is given by

$$x(t-iD)=\sum_{k=1}^{K}a_k e^{\alpha_k(t-iD)}\cos(\varpi_k(t-iD)+\phi_k)+w_i(t) \tag{1.17}$$

Transforming again to the complex representation, (1.17) can be expressed as

$$\tilde{x}(t-iD)=\sum_{k=1}^{K}\tilde{a}_k e^{-j\varpi_k iD}+\tilde{w}_i(t) \tag{1.18}$$

where $\tilde{a}_k = a_k e^{j(s_k t+\phi_k)}$, (1.11) and (1.12) still hold for $\tilde{w}_i(\bullet)$.

Given a snapshot of $N$ complex-valued samples, the pole retrieval problem is to estimate the locations $s_1,\cdots,s_K$ of the poles and the corresponding residues $\tilde{a}_1,\cdots,\tilde{a}_K$.

#### 1.1.4 The Echo Retrieval Problem

Consider a radar or sonar system that transmits a known pulse $s(\bullet)$ and receives a backscattered signal. The backscattered signal can be modeled as a superposition of $K$ scaled and delayed echoes embedded in additive noise,

$$x(t)=\sum_{k=1}^{K}m_k s(t-\tau_k)+w(t) \tag{1.19}$$

where $m_k$ is the amplitude of the $k$-th echo and $w(t)$ is the additive noise. Assume that the received signal is sampled by a tapped-delay-line with $N$ equally spaced taps $D$ delay units apart; then the complex representation at the ($i$+1)-th tap is expressed as

$$\tilde{x}(t-iD)=\sum_{k=1}^{K}\tilde{m}_k\tilde{s}(t-iD-\tau_k)+\tilde{w}(t-iD) \tag{1.20}$$

Given a snapshot of $N$ complex-valued samples, the echo retrieval problem is to estimate the delays $\tau_1,\cdots,\tau_K$ and the amplitudes $\tilde{m}_1,\cdots,\tilde{m}_K$ of the echoes.

## 1.2 Modeling and Formulating

Motivated by the examples presented in the previous section, in particular the passive sensor array problem, we now formulate the superimposed sinusoidal signal problem.

Let $x(0),\ldots, x(N-1)$ denote $N$ complex-valued samples or observations of a complex-valued process $x(\bullet)$,

$$x(n) = s(n) + w(n) = \sum_{k=1}^{K} a_k e^{j\varpi_k T_s n} + w(n), \quad n = 0,1,\ldots,N-1 \tag{1.21}$$

where $a_k$ is the complex amplitude of the $k$th sinusoid having frequency $\omega_k$, $\{w(n)\}$ are the observation noises, assumed to be statistically independent and white Gaussian noises with mean zeroes, and $T_s$ is the sampling interval.

Stacking the observation data into vectors, we can rewrite the model (1.21) as

$$\mathbf{x} = [\mathbf{s}(\varpi_1),\mathbf{s}(\varpi_2),\ldots,\mathbf{s}(\varpi_K)]\mathbf{a} + \mathbf{w} = \mathbf{S}(\varpi)\mathbf{a} + \mathbf{w} \tag{1.22}$$

where $\mathbf{x} = \begin{pmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{pmatrix}$, $\mathbf{w} = \begin{pmatrix} w(0) \\ w(1) \\ \vdots \\ w(N-1) \end{pmatrix}$, $\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_K \end{pmatrix}$, $\mathbf{s}(\varpi) = \begin{pmatrix} 1 \\ e^{j\varpi T_s} \\ \vdots \\ e^{j\varpi T_s (N-1)} \end{pmatrix}$,

$\mathbf{S}(\varpi) = [\mathbf{s}(\varpi_1),\cdots,\mathbf{s}(\varpi_K)]$ is a $N \times K$ matrix with $K < N$, and where the white noise satisfies

$$E(\mathbf{w}) = 0, \; E(\mathbf{w}\mathbf{w}^H) = \sigma^2 \mathbf{I} \text{ and } E(\mathbf{w}\mathbf{w}^T) = 0 \tag{1.23}$$

The sample space associated with model (1.22) is $(X, \mathfrak{I}, P)$, where $X = \mathbf{C}^N$, $\mathfrak{I}$ is the Borel set of $\mathbf{C}^N$, and for any event $B$ in $\mathfrak{I}$, the probability $P(\bullet)$ is given by

$$P(B \in \mathfrak{I}) = \int_{B \in \mathfrak{I}} \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{|x(i) - \sum_{k=1}^{K} a_k e^{j\varpi_k T_s i}|^2}{2\sigma^2}\right) d\mathbf{x}$$

$$= \int_{B \in \mathfrak{I}} \frac{1}{(\sqrt{2\pi}\sigma)^N} \exp\left(-\frac{\|\mathbf{x} - \mathbf{S}(\varpi)\mathbf{a}\|^2}{2\sigma^2}\right) d\mathbf{x}$$

The superimposed sinusoidal signal problem can be stated as follows.

Given a snapshot of observation data $\mathbf{x}$ described by equation (1.21) or (1.22), estimate the following three sets of unknown parameters,

1) The number of sinusoids $K$.

2) The frequencies $\varpi_1, \ldots, \varpi_K$.

3) The amplitudes $a_1, \ldots, a_K$.

## 1.3 Previous Method Review

In this section, we review some of the existing methods to the superimposed sinusoidal signal problem.

As the classical method, Fourier transform,

$$X(\varpi) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-j\varpi T_s n} \tag{1.24}$$

can work on the problem. The number of sinusoids $K$ can be estimated by the number of local maxima of $|X(\varpi)|$ which are over a threshold; the frequencies can be estimated by picking the $K$ values of $\omega$ which locally maximizes $|X(\varpi)|$, we denote the estimated frequencies as $\hat{\varpi}_k$, $k = 1,2,\cdots,K$; and, the amplitudes can be estimated by $\hat{a}_k = X(\hat{\varpi}_k)$, $k = 1,2,\cdots,K$.

The classical technique is still popular even today but is challenged by the following problems:

1) The window length $T = NT_s$ is finite and imposes a limitation upon the resolution capability of two sinusoidal signals. As is well known, two sinusoidal signals cannot be resolved by the classical method, if their frequency gap is less than $1/T$.
2) In estimating the amplitude for one sinusoid, say $\omega_1$, all other sinusoids interfere with the estimation as clutters, possibly severely damaging the estimation.

These problems have motivated the modern developments in signal, array, and imaging processing such as maximum entropy spectral estimation, adaptive filtering, or adaptive antenna.

One method to improve upon the Fourier transform method was presented by Burg (1967, 1975). He attributed the poor resolution of the Fourier transform method to the limited window length which yields only a finite number of the spatial covariance lags while the other covariance lags are inadvertently assumed to be zero. To improve the resolution, he proposed to extrapolate the covariance function beyond the given segment. In principle, there is an infinite number of such possible extrapolations; the "all zeros" extrapolation assumed by the Fourier transform method is one of them, a very arbitrary one as a matter of fact. Burg proposed selecting that extrapolation for which the entropy of the signal is maximized and showed that the maximum entropy is achieved by fitting an AR model to the data. If the order of the AR model is $p$, the AR$(p)$ model

$$x_t - \sum_{k=1}^{p} \phi_k x_{t-k} = \theta_0 \varepsilon_t \tag{1.25}$$

amounts to solving for the AR coefficients $\{\phi_i\}$ that minimize the expected prediction error

$$\min_{\phi_1 \_ \phi_p} E|x_t - \sum_{k=1}^{p} \phi_k x_{t-k}|^2 \tag{1.26}$$

All the parameters $(\phi_1, \cdots, \phi_p, \theta_0^2)$ can be obtained equivalently by solving the following Yule-Walker equation:

$$\mathbf{R}\begin{pmatrix} 1 \\ \phi_1 \\ \vdots \\ \phi_p \end{pmatrix} = \begin{pmatrix} \theta_0^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{1.27}$$

where $\mathbf{R} = \begin{pmatrix} R(0) & R(1) & \cdots & R(p) \\ R(1) & R(0) & \cdots & R(p-1) \\ \vdots & \vdots & \ddots & \vdots \\ R(p) & R(p-1) & \cdots & R(0) \end{pmatrix}$ is the true $(p+1)\times(p+1)$ covariance matrix of the complex-valued process $x(\bullet)$ and can be estimated using the sample-covariance matrix $\hat{R}$. Thus, the equation (1.2-4) can be rewritten as

$$\hat{\mathbf{R}}\begin{pmatrix} 1 \\ \hat{\phi}_1 \\ \vdots \\ \hat{\phi}_p \end{pmatrix} = \begin{pmatrix} \hat{\theta}_0^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{1.28}$$

The spatial spectral density function is thus given by

$$f(\varpi) = \frac{\hat{\theta}_0^2}{2\pi|1 + \sum_{k=1}^{p} \hat{\phi}_k e^{-jk\varpi}|^2} \tag{1.29}$$

and the frequencies are determined as the peaks of this spectrum.

Noticing that the poor resolution is caused by undesirable contributions from the other sources, Capon (1969) proposed a different approach adaptively filtering the interference so as to increase the resolution. His approach was as follows. Suppose we want to estimate the $k$-th amplitude $a_k$ by a linear filter

$$\hat{a}_k = \mathbf{h}_k^T \mathbf{x} \tag{1.30}$$

where $\mathbf{h}_k$ is a $N \times 1$ complex vector. Using (1.22) we can rewrite this as

$$\begin{aligned} \hat{a}_k &= a_1 \mathbf{h}^T \mathbf{s}(\varpi_1) + a_2 \mathbf{h}^T \mathbf{s}(\varpi_2) + \cdots + a_K \mathbf{h}^T \mathbf{s}(\varpi_K) + \mathbf{h}^T \mathbf{w} \\ &= a_k \mathbf{h}^T \mathbf{s}(\varpi_k) + \sum_{i \neq k} a_i \mathbf{h}^T \mathbf{s}(\varpi_i) + \mathbf{h}^T \mathbf{w} \end{aligned} \tag{1.31}$$

Note that the first term is a scaled version of the desired signal, while the other terms are contributions from the other signals and the noise. Regarding all but the first term as undesirable interference, Capon proposed selecting $\mathbf{h}_k$ so as to minimize the power of the right-hand side, subject to the constraint that the desired signal be undistorted. More formally, he proposed selecting $\mathbf{h}_k$ as the solution of the following minimization problem:

$$\min_{\mathbf{h}} E(|\mathbf{h}^T \mathbf{x}|^2) = \min_{\mathbf{h}} (\mathbf{h}^T \mathbf{R} \mathbf{h}^*) \tag{1.32}$$

subject to the constraint

$$\mathbf{h}^T \mathbf{s}(\varpi_k) = 1 \tag{1.33}$$

The solution to this minimization problem is given by

$$\hat{h}_k = c_k^2 \mathbf{R}^{-1} \mathbf{s}(\varpi_k) \tag{1.34}$$

where $c_k^2$ is positive scalar given by

$$c_k^2 = \frac{1}{\mathbf{s}^H(\varpi_k) \mathbf{R}^{-1} \mathbf{s}(\varpi_k)} \tag{1.35}$$

The directions-of-arrival can be estimated as the values, which locally maximize

$$C(\varpi)=\frac{1}{\mathbf{s}^H(\varpi)\hat{\mathbf{R}}^{-1}\mathbf{s}(\varpi)} \tag{1.36}$$

Because the maximum entropy method was claimed to have the super capability of frequency resolution over the Fourier transform, significant attention (Ligget, 1973; Pisarenko, 1973; Berni, 1975; Schmidt, 1979, 1981; Bienvenu and Kopp, 1979, 1980, 1981) has been paid to frequency estimation. The eigenstructure of the covariance matrix $\mathbf{R}$ of the observation vector $\mathbf{x}$ has thus been extensively studied, and a class of approaches has been developed to estimating the frequencies. A representative member of this class is the MUSIC (Multiple Signal Characterization) algorithm. Before we describe MUSIC, we point out certain properties of the covariance matrix $\mathbf{R}$. From (1.22), assuming that the signals are uncorrelated zero-mean random processes that are independent of the noise, the covariance matrix is given by

$$\mathbf{R}=\mathbf{S}(\varpi)\mathbf{A}\mathbf{S}^H(\varpi)+\sigma^2\mathbf{I} \tag{1.37}$$

where $\mathbf{A}$ is the $K\times K$ covariance matrix of the signals, $\mathbf{A}=diag\{a_1^2,\cdots,a_K^2\}$

Let $\lambda_1\geq\lambda_2\cdots\geq\lambda_N$ and $\upsilon_1\geq\upsilon_2\cdots\geq\upsilon_N$ denote the eigenvalues of $\mathbf{R}$ and $\mathbf{S}(\varpi)\mathbf{A}\mathbf{S}^H(\varpi)$, respectively. From the structure of $\mathbf{R}$ given by (1.37), these two sets of eigenvalues are related by

$$\lambda_i=\upsilon_i+\sigma^2 \tag{1.38}$$

Assuming that the matrix $\mathbf{S}(\varpi)$ is of full column rank, it follows that the rank of matrix $\mathbf{S}(\varpi)\mathbf{A}\mathbf{S}^H(\varpi)$ is $K$, implying that the $N$-$K$ smallest eigenvalues of $\mathbf{S}(\varpi)\mathbf{A}\mathbf{S}^H(\varpi)$ are equal to zero

$$\upsilon_{K+1} = \cdots = \upsilon_N = 0 \tag{1.39}$$

Thus by (1.38), we have

$$\lambda_{K+1} = \cdots = \lambda_N = \sigma^2 \tag{1.40}$$

That is, the smallest eigenvalues of $\mathbf{R}$ are equal to $\sigma^2$ with multiplicity $N$-$K$.

Denote the unit-norm eigenvectors associated with $\lambda_1, \cdots, \lambda_K$ by $\mu_1, \cdots, \mu_K$, and those corresponding to $\lambda_{K+1}, \cdots, \lambda_N$ by $\mathbf{g}_1, \cdots, \mathbf{g}_{N-K}$. Also define

$$\mu = (\mu_1 \cdots \mu_K)_{N \times K}$$

$$\mathbf{G} = (g_1 \cdots g_{N-K})_{N \times (N-K)}$$

Next, observe that

$$\mathbf{R}\mathbf{G} = \mathbf{S}(\varpi)\mathbf{A}\mathbf{S}^H(\varpi)\mathbf{G} + \sigma^2\mathbf{G} = \sigma^2\mathbf{G} \tag{1.41}$$

which readily implies $\mathbf{S}^H(\varpi)\mathbf{G} = 0$, or equivalently

$$\mathbf{s}^H(\varpi_k)\mathbf{G}\mathbf{G}^H\mathbf{s}(\varpi_k) = 0 \qquad k = 1,2,\cdots,K \tag{1.42}$$

Since the normalized eigenvectors $\{\mu_i, \mathbf{g}_j\}$ are orthonormal,

$$\mu\mu^H + \mathbf{G}\mathbf{G}^H = \mathbf{I} \tag{1.43}$$

It follows that (1.42) can also be written as

$$\mathbf{s}^H(\varpi_k)(\mathbf{I} - \mu\mu^H)\mathbf{s}(\varpi_k) = 0 \qquad k = 1,2,\cdots,K \tag{1.44}$$

It is not difficult to see that the true parameter values $\varpi_1, \ldots, \varpi_K$ are the only solutions of (1.42) or (1.44).

The basic idea of the MUSIC algorithm is the exploitation of the property (1.42) or (1.44) of the true covariance matrix **R** that can be estimated by sample covariance matrix $\hat{\mathbf{R}}$.

Similar to the eigendecomposition of **R**, let $\{\hat{\mu}_{1_1},\cdots,\hat{\mu}_K,\hat{g}_1,\cdots,\hat{g}_{N-K}\}$ denote the unit-norm eigenvectors of $\hat{\mathbf{R}}$, arranged in the descending order of the associated eigenvalues, and let $\hat{\mu}$ and $\hat{\mathbf{G}}$ denote the matrices $\mu$ and **G** made of $\{\hat{\mu}_i\}$ and, respectively, $\{\hat{g}_j\}$. Define

$$f(\varpi)=\mathbf{s}^H(\varpi)\hat{\mathbf{G}}\hat{\mathbf{G}}^H\mathbf{s}(\varpi)=\mathbf{s}^H(\varpi)(\mathbf{I}-\hat{\mu}\hat{\mu}^H)\mathbf{s}(\varpi) \tag{1.45}$$

The MUSIC estimates of $\{\varpi_k\}$ are obtained by picking the K values of ω for which $f(\varpi)$ is minimized.

There have been many other frequency estimation approaches proposed. Stoica et al. (1989) found the Cramer-Rao bound (CRB) of estimating the frequencies. In principle, there is a large number of frequency estimators; thus a dispute arises since everyone claims his or her estimator takes some advantages over the others. The importance of the CRB is that it provides a criterion for deciding the optimal frequency estimator, and puts the dispute to an end by comparing the performance of a given estimator to the ultimate performance corresponding to the CRB.

However, the challenging problems still remain unsolved. Most of the previous approaches paid significant attention to estimating the frequencies while assuming the number of signals is known. What happens if the number of signals is unknown?

Wax and Kailath (1985) indicated that the estimation of the number of signals is a prerequisite for signal detection and estimation. They applied the results of the eigendecomposition of **R** to the information theoretic criteria (ITC) for model selection introduced by Akaike (AIC) and by Schwartz and Rissanen (MDL), the number of signals is determined as the value for which the AIC or the MDL criteria are minimized. According to the ITC, the number of unknown signals in a likelihood function $f(x|\varpi)$ is selected to minimize the ITC sequence as

$$\text{ITC}(m) = -2\log f(x|\hat{\varpi}^{(m)}) + P(N,m),\ m \in [0, M] \quad (1.46)$$

where $\hat{\varpi}^{(m)}$ is the maximum likelihood estimate (MLE) when the number of signals is $m$, i.e., $m+1 \le k \le M$, and $\varpi = [\varpi_1,\ldots,\varpi_m]^T$ is a $m$-dimensional vector with $\varpi_m \ne 0$; and $P(N,m)$ is a penalty function in the ITC.

A different approach to estimating the number of signals was given by Gu (1998). He proposed that the estimation of the number of signals was based on the resolution of multiple signals, and the latter is given by the amplitude estimation. In his approach, the nested maximum likelihood ratios that are given by

$$\hat{d}_m^2 = 2\log f(\mathbf{x}|\hat{\varpi}^{(m)}) - 2\log f(\mathbf{x}|\hat{\varpi}^{(m-1)}) \quad (1.47)$$

were shown to be asymptotically independent and sufficient for estimating the number of signals. It is also shown that the statistic $\hat{d}_m$ is asymptotically normally distributed with a standard deviation of 1 and a mean value $d_m$. Thus, $d_m = 0$, $m = K+1,\cdots,M$, and $d_K \ne 0$ if and only if the number of signals is $K$. As a result, estimating the number of signals can be reduced to a sequence of independent tests of hypothesis $d_m = 0$ against its alternative $d_m \ne 0$, which has the reject region

$$|\hat{d}_m| > D_m \qquad m = 1,\cdots,M \quad (1.48)$$

Applying the sequence $\{\hat{d}_m\}$ to estimating the number of signals in (1.22), it is noted that

$$-2\log f(\mathbf{x}|\hat{\varpi}^{(m)}) = \min_{\mathbf{s}(\varpi)\in\mathbf{L}_m} \|\mathbf{x}-\mathbf{s}(\varpi)\|^2 \big/ \sigma^2 \qquad m = 1,\cdots,M \qquad (1.49)$$

where $\mathbf{L}_m$ is the subspace spanned by $s(\varpi_1),\cdots,s(\varpi_m)$. Thus, from the Pythagorean theorem, it follows that

$$\hat{d}_m^2 = 2\log f(\mathbf{x}|\hat{\varpi}^{(m)}) - 2\log f(\mathbf{x}|\hat{\varpi}^{(m-1)}) = |\mathbf{x}^T\mathbf{u}_m^*|^2 \big/ \sigma^2 \qquad (1.50)$$

where $\mathbf{u}_m$ is a unit vector in the subspace $\mathbf{L}_m$ and perpendicular to $\mathbf{L}_{m-1}$. The linear processing $\mathbf{x}^T\mathbf{u}_m^*$ is the maximum likelihood estimator of the amplitude of signal $s(\varpi_m)$. In other words, $\{\hat{d}_m\}$ are given by the estimation of the amplitudes.

## 1.4 Dissertation Outline

The Cramer-Rao Bound (CRB) of estimating the frequencies was established by Stoica et al. (1989) while treating the amplitudes as nuisances; no CRB of estimating the amplitudes with unknown frequencies has been reported yet. However, the estimation of the amplitudes plays a significant role in signal resolution and detection. A question of interest is what the performance bound and the optimal estimator are in the estimation of the amplitudes of signals using a finite sample. This question is the major concern of this dissertation.

The organization of this dissertation is as follows. In Chapter 2, we derive the formula of the CRB of estimating the amplitudes of superimposed sinusoidal signals with unknown frequencies in white noise after we present a brief review of the least square estimator (LSE) of the amplitudes of superimposed sinusoidal signals with known frequencies in white noise. In Chapter 3, the maximum likelihood estimator (MLE) of the amplitudes with unknown frequencies in white noise is given and shown to be asymptotically efficient at high signal to noise ratio (SNR) but finite sample. In Chapter 4, we present the simulation results that illustrate the performance of the MLE. In Chapter 5, we present some concluding remarks.

# CHAPTER 2

# CRAMER-RAO BOUND

As mentioned in Chapter 1, the amplitude estimation of superimposed sinusoidal signals plays a significant role in signal resolution. In principle, there is a large number of estimators of the amplitudes such as the least square estimator (LSE), the weighted least square estimator (WLSE), and the matched-filterbank estimator (MAFI) which has been introduced recently (Stoica et al., 2000). Accordingly a natural question is which one is optimal. The Cramer-Rao Bound (CRB) of estimating the amplitudes derived in this chapter can be employed to answer the question. The importance of the CRB is that it provides a criterion for deciding optimal estimators of the amplitudes by comparing the performance of a given estimator to the ultimate performance corresponding to the CRB.

In this chapter, we introduce the concept of the Constrained Matched Filter (CMF), then review some properties of the least square estimator (LSE) of the amplitudes in model (1.22) when the frequencies are known and finally derive the formula of the CRB of estimating the amplitudes when the frequencies are unknown.

## 2.1 Constrained Matched Filter

As an extension of the classical matched filters, the Constrained Matched Filter introduced in this section maximizes the signal to noise ratio (SNR) for a desired signal and cancels out unwanted signals in the input data.

In model (1.22), a weight vector $\mathbf{h}$ of the CMF $\mathbf{s}(\varpi_m)$ against $\{\mathbf{s}(\varpi_i), i = 1,\cdots,K, i \neq m\}$ must satisfy

$$\mathbf{h}^T \mathbf{s}(\varpi_m) = 1 \tag{2.1}$$

$$\mathbf{h}^T \mathbf{s}(\varpi_i) = 0\,, \quad i = 1,\cdots,K, i \neq m \tag{2.2}$$

and maximize the SNR for the desired signal $\mathbf{s}(\varpi_m)$

$$\mathrm{SNR} = \frac{|a_m \mathbf{h}^T \mathbf{s}(\varpi_m)|^2}{\sigma^2 \mathbf{h}^H \mathbf{h}} \tag{2.3}$$

To maximize SNR subject to ( 2.1), break $\mathbf{s}(\varpi_m)$ to $p_m + q_m$, where $p_m$ is the projection of $\mathbf{s}(\varpi_m)$ onto the subspace $\mathbf{L}_m$ spanned by $\{\mathbf{s}(\varpi_i), i = 1,\cdots,K, i \neq m\}$, and $q_m$ is orthogonal to $\mathbf{L}_m$. Thus, $\mathbf{h}^T p_m = 0$ because of (2.1), $p_m \in \mathbf{L}_m$, and

$$\mathrm{SNR} = \frac{|a_m \mathbf{h}^T q_m|^2}{\sigma^2 \mathbf{h}^H \mathbf{h}} = \frac{|a_m|^2 \, |\mathbf{h}^T q_m|^2}{\sigma^2 \mathbf{h}^H \mathbf{h}} \leq \frac{|a_m|^2 \, |q_m|^2}{\sigma^2} = SNR_{\max} \tag{2.4}$$

as a result of Schwarz inequality. The equality holds if and only if $\mathbf{h} = c(q_m)^*$, where $c$ is a constant complex number. Therefore, $\mathbf{h}$ is a weight vector of the CMF of $\mathbf{s}(\varpi_m)$ against $\{\mathbf{s}(\varpi_i), i = 1,\cdots,K, i \neq m\}$ if and only if $\mathbf{h} = c(q_m)^*$.

## 2.2 LSE of the Amplitudes with Known Frequencies

In this section, we review some properties of the LSE of the amplitudes in model (1.22) when the frequencies are known.

The least squares estimator (LSE) of the amplitudes $\mathbf{a}$ is given by

$$\hat{\mathbf{a}} = [\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_k]^T = \arg\min_{\mathbf{a}} \|\mathbf{x} - \mathbf{S}(\varpi)\mathbf{a}\| \tag{2.5}$$

$$= [\mathbf{S}^H(\varpi)\mathbf{S}(\varpi)]^{-1} \mathbf{S}^H(\varpi)\mathbf{x} \tag{2.6}$$

$\hat{\mathbf{a}}$ is an unbiased estimator of the amplitudes $\mathbf{a}$, since $E(\hat{\mathbf{a}}) = \mathbf{a}$.

$\hat{\mathbf{a}}$ is consistent, because $\lim_{\sigma \to 0} \hat{\mathbf{a}} = \lim_{\sigma \to 0}(\mathbf{a} + [\mathbf{S}^H(\varpi)\mathbf{S}(\varpi)]^{-1}\mathbf{S}^H(\varpi)\mathbf{w}) = \mathbf{a}$, almost surely.

The covariance matrix of $\hat{\mathbf{a}}$ is given by

$$\Sigma = E([\hat{\mathbf{a}} - \mathbf{a}][\hat{\mathbf{a}} - \mathbf{a}]^H) = \sigma^2[\mathbf{S}^H(\varpi)\mathbf{S}(\varpi)]^{-1} \tag{2.7}$$

The matrix in (2.7) is the Cramer-Rao Bound (CRB) of estimating the complex amplitudes of the superimposed sinusoidal signals with known frequencies in white noise (Stoica et al., 2000 ).

Separating the amplitudes $\mathbf{a}$ into two parts, $\mathbf{a} = \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{pmatrix}$, then the model (1.22) can be rewritten as

$$\mathbf{x} = \mathbf{S}_1\mathbf{a}_1 + \mathbf{S}_2\mathbf{a}_2 + \mathbf{w} \tag{2.8}$$

where $\mathbf{a}_1 = [a_1, a_2, \cdots, a_M]^T$, $\mathbf{a}_2 = [a_{M+1}, a_{M+2}, \cdots, a_K]^T$, $\mathbf{S}_1(\varpi) = [\mathbf{s}(\varpi_1), \cdots, \mathbf{s}(\varpi_M)]$, and $\mathbf{S}_2(\varpi) = [\mathbf{s}(\varpi_{M+1}), \cdots, \mathbf{s}(\varpi_K)]$.

If $\mathbf{S}_1(\varpi)$ and $\mathbf{S}_2(\varpi)$ are mutually orthogonal, i.e., $\mathbf{S}_1^H(\varpi)\mathbf{S}_2(\varpi) = 0$, then we have the following theorem (Gu, 2000):

**Theorem 1:** The LSE of $\mathbf{a}_1$ in model (2.8) can be obtained from the following model:

$$\mathbf{x} = \mathbf{S}_1\mathbf{a}_1 + \mathbf{w} \tag{2.9}$$

while the LSE of $\mathbf{a}_2$ in model (2.8) can be obtained from the following model:

$$\mathbf{x} = \mathbf{S}_2\mathbf{a}_2 + \mathbf{w} \tag{2.10}$$

The LSE of $\mathbf{a}_1$ and the LSE of $\mathbf{a}_2$ are uncorrelated.

Besides, the LSE $\hat{\mathbf{a}}$ is a bank of filters, $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_K$. For example, $\hat{a}_1 = \mathbf{h}^T\mathbf{x}$ is a linear filter of $\mathbf{x}$ with a weight vector $\mathbf{h}^T$. Since $\hat{a}_1$ is unbiased, thus

$$E(\hat{a}_1) = E(\mathbf{h}^T\mathbf{x}) = a_1\mathbf{h}^T\mathbf{s}(\varpi_1) + a_2\mathbf{h}^T\mathbf{s}(\varpi_2) + \cdots + a_K\mathbf{h}^T\mathbf{s}(\varpi_K) = a_1$$

and hence

$$\mathbf{h}^T\mathbf{s}(\varpi_1) = 1, \quad \mathbf{h}^T\mathbf{s}(\varpi_2) = \cdots = \mathbf{h}^T\mathbf{s}(\varpi_k) = 0 \tag{2.11}$$

(2.11) means that the filter $\hat{a}_1$ has nulls at $\omega_2$, $\omega_3$, $\ldots, \omega_K$. Furthermore, it is shown by Gu (1997) that the filter $\hat{a}_1$ is a constrained matched filter which maximizes the SNR for the signal at $\omega_1$ subject to the nulls in (2.7). The LSE perfectly solves the challenging problems with the Fourier transform if the frequencies are known.

## 2.3 CRB of Estimating theAmplitudes with Unknown Frequencies

In this section, we are considering the model (1.21) or the vector version (1.22) with unknown frequencies $\boldsymbol{\varpi}$ and amplitudes $\mathbf{a}$ in the case of white Gaussian noise. The CRB of estimating $\boldsymbol{\varpi}$ in this case has been established by Stoica et al. (1989); no CRB of estimating the amplitudes with unknown frequencies has been reported yet. Before we derive the CRB of estimating the amplitudes $\mathbf{a}$ in this case, let us introduce the linearization method proposed by Gu (2000).

We first generalize the model (1.22) as follows:

$$\mathbf{x} = \mathbf{f}(\theta) + \mathbf{w} \tag{2.12}$$

where $\mathbf{f}(\theta)$ is an $N$-dimensional complex-valued vector with a $K$-dimensional real-valued unknown parameters $\theta$ $(K < N)$, and the assumption for $\mathbf{w}$ in (1.23) still holds.

Expanding $\mathbf{f}(\theta)$ to Taylor series at the true value $\theta^0$ of $\theta$, the nonlinear model (2.12) can be transformed into the following linear model:

$$\mathbf{x} = \mathbf{f}(\theta) + \mathbf{w} = \mathbf{f}(\theta^0) + \frac{\partial \mathbf{f}(\theta^0)}{\partial \theta^T}(\theta - \theta^0) + \mathbf{w} \tag{2.13}$$

where

$$\frac{\partial \mathbf{f}(\theta)}{\partial \theta^T} = [\frac{\partial \mathbf{f}(\theta)}{\partial \theta_1}, \frac{\partial \mathbf{f}(\theta)}{\partial \theta_2}, \ldots, \frac{\partial \mathbf{f}(\theta)}{\partial \theta_K}]$$

and

$$\frac{\partial \mathbf{f}^H(\theta)}{\partial \theta} = [\frac{\partial \mathbf{f}(\theta)}{\partial \theta_1}, \frac{\partial \mathbf{f}(\theta)}{\partial \theta_2}, \ldots, \frac{\partial \mathbf{f}(\theta)}{\partial \theta_K}]^H$$

Letting $\mathbf{y} = \mathbf{x} - \mathbf{f}(\theta^0)$, (2.13) can be rewritten as

$$\mathbf{y} = \frac{\partial \mathbf{f}(\theta^0)}{\partial \theta^T}(\theta - \theta^0) + \mathbf{w} \tag{2.14}$$

Gu (2000) proved the following theorem:

**Theorem 2:** The CRB of estimating $\theta$ in the nonlinear model (2.12) is given by computing the covariance matrix of the LSE of $(\theta - \theta^0)$ in the linear model (2.14) when $\theta$ is real-valued.

The theorem holds for real-valued parameters, But, how can we apply the theorem to the complex-valued amplitude parameters?

Extending $\theta$ to the complex value, we obtain the following theorem:

**Theorem 3:** Theorem 2 holds when $\theta$ is complex-valued

Proof: For any unbiased estimator of $\theta^0$, we denote it by $\hat{\phi} = \hat{\phi}_1 + j\hat{\phi}_2$, we have,

$$E(\hat{\phi}) = \theta^0, \ E(\hat{\phi}_1) = \theta_1^0 \text{ and } E(\hat{\phi}_2) = \theta_2^0$$

where $\theta_1^0 = \mathrm{Re}(\theta^0)$, $\theta_2^0 = \mathrm{Im}(\theta^0)$, and $\theta^0 = \theta_1^0 + j\theta_2^0$.

We denote the LSE of $\theta^0$ by $\hat{\psi} = \hat{\psi}_1 + j\hat{\psi}_2$, so that

$$E(\hat{\psi}) = \theta^0,\ E(\hat{\psi}_1) = \theta_1^0 \text{ and } E(\hat{\psi}_2) = \theta_2^0$$

Obviously, $\hat{\phi}_1$ and $\hat{\psi}_1$ are unbiased estimators of $\theta_1^0$ while $\hat{\phi}_2$ and $\hat{\psi}_2$ are unbiased estimators of $\theta_2^0$.

We also denote the covariance matrices of $\hat{\phi}$ and $\hat{\psi}$ by $\Sigma$ and $\Omega$, respectively, where $\Sigma = E([\hat{\phi} - \theta^0][\hat{\phi} - \theta^0]^H)$ and $\Omega = E([\hat{\psi} - \theta^0][\hat{\psi} - \theta^0]^H)$ are positive definite Hermitian matrices.

To prove the theorem, for any $K$-dimensional complex-valued vector $\mathbf{C} = \mathbf{C}_1 + j\mathbf{C}_2$, where $\mathbf{C}_1$ and $\mathbf{C}_2$ are K-dimensional real valued vectors, we need to show that the variance of $\mathbf{C}^T\hat{\phi}$, which is a linear combination of $\hat{\phi}$, must be equal to or greater than the variance of the corresponding linear combination of $\hat{\psi}$, $\mathbf{C}^T\hat{\psi}$, that is, we need to show that the following inequality holds,

$$E([\mathbf{C}^T(\hat{\phi} - \theta^0)][\mathbf{C}^T(\hat{\phi} - \theta^0)]^*) \geq E([\mathbf{C}^T(\hat{\psi} - \theta^0)][\mathbf{C}^T(\hat{\psi} - \theta^0)]^*) \qquad (2.15)$$

or, equivalently,

$$\mathbf{C}^T\Sigma(\mathbf{C}^T)^H \geq \mathbf{C}^T\Omega(\mathbf{C}^T)^H \qquad (2.16)$$

Noticing the fact that every complex number consists of two parts, real and imaginary, the model (2.12) can be considered as estimating $2K$-dimensional real-valued unknown parameters $\begin{pmatrix}\theta_1^0 \\ \theta_2^0\end{pmatrix}$.

Applying Theorem 2, the CRB of estimating the $2K$-dimensional real-valued unknown parameters can be easily obtained by computing the covariance matrix of the LSE of $\begin{pmatrix}\theta_1^0\\ \theta_2^0\end{pmatrix}$ in linear model (2.14).

The LSE of $\begin{pmatrix}\theta_1^0\\ \theta_2^0\end{pmatrix}$ in linear model (2.14) is exactly $\begin{pmatrix}\hat{\psi}_1\\ \hat{\psi}_2\end{pmatrix}$, and we denote the variance matrix of $\begin{pmatrix}\hat{\psi}_1\\ \hat{\psi}_2\end{pmatrix}$ by $\Psi$, which is given by

$$\Psi = E([\hat{\psi}_1 - \theta_1^0]^T,[\hat{\psi}_2 - \theta_2^0]^T)\begin{pmatrix}\hat{\psi}_1 - \theta_1^0\\ \hat{\psi}_2 - \theta_2^0\end{pmatrix} = \begin{pmatrix}\Psi_{11} & \Psi_{12}\\ \Psi_{21} & \Psi_{22}\end{pmatrix}$$

where

$$\Psi = \Psi^T, \Psi_{11} = \Psi_{11}^T,\ \Psi_{22} = \Psi_{22}^T,\ \Psi_{12} = \Psi_{21}^T$$

$$\Psi_{11} = E([\hat{\psi}_1 - \theta_1^0]^T[\hat{\psi}_1 - \theta_1^0])$$

$$\Psi_{22} = E([\hat{\psi}_2 - \theta_2^0]^T[\hat{\psi}_2 - \theta_2^0])$$

$$\Psi_{12} = E([\hat{\psi}_1 - \theta_1^0]^T[\hat{\psi}_2 - \theta_2^0])$$

$$\Psi_{21} = E([\hat{\psi}_2 - \theta_2^0]^T[\hat{\psi}_1 - \theta_1^0])$$

The vector $\begin{pmatrix}\hat{\phi}_1\\ \hat{\phi}_2\end{pmatrix}$ is an unbiased estimator of $\begin{pmatrix}\theta_1^0\\ \theta_2^0\end{pmatrix}$ in linear model (2.14), and we denote the variance matrix of $\begin{pmatrix}\hat{\phi}_1\\ \hat{\phi}_2\end{pmatrix}$ by $\Phi$, which is given by

$$\Phi = E([\hat{\phi}_1 - \theta_1^0]^T,[\hat{\phi}_2 - \theta_2^0]^T)\begin{pmatrix}\hat{\phi}_1 - \theta_1^0\\ \hat{\phi}_2 - \theta_2^0\end{pmatrix} = \begin{pmatrix}\Phi_{11} & \Phi_{12}\\ \Phi_{21} & \Phi_{22}\end{pmatrix}$$

where

$$\Phi = \Phi^T,\ \Phi_{11} = \Phi_{11}^T,\ \Phi_{22} = \Phi_{22}^T,\ \Phi_{12} = \Phi_{21}^T$$

$$\Phi_{11} = E([\hat{\phi}_1 - \theta_1^0]^T[\hat{\phi}_2 - \theta_2^0])$$

$$\Phi_{12} = E([\hat{\phi}_1 - \theta_1^0]^T[\hat{\phi}_2 - \theta_2^0])$$

$$\Phi_{22} = E([\hat{\phi}_2 - \theta_2^0]^T[\hat{\phi}_2 - \theta_2^0])$$

$$\Phi_{21} = E([\hat{\phi}_2 - \theta_2^0]^T[\hat{\phi}_1 - \theta_1^0])$$

By Theorem 2, the variance matrix of the LSE $\begin{pmatrix} \hat{\psi}_1 \\ \hat{\psi}_2 \end{pmatrix}$ is exactly the CRB of estimating the unknown parameter $\begin{pmatrix} \theta_1^0 \\ \theta_2^0 \end{pmatrix}$ in linear model (2.14).

Hence, the variance of any linear combination of an unbiased estimator of $\begin{pmatrix} \theta_1^0 \\ \theta_2^0 \end{pmatrix}$, say, $\begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{pmatrix}$, is equal to or greater than the variance of the corresponding linear combination of $\begin{pmatrix} \hat{\psi}_1 \\ \hat{\psi}_2 \end{pmatrix}$, or,

$$E|[\mathbf{C}_1^T, \mathbf{C}_2^T]\begin{pmatrix} \hat{\phi}_1 - \theta_1^0 \\ \hat{\phi}_2 - \theta_2^0 \end{pmatrix}|^2 \geq E|[\mathbf{C}_1^T, \mathbf{C}_2^T]\begin{pmatrix} \hat{\psi}_1 - \theta_1^0 \\ \hat{\psi}_2 - \theta_2^0 \end{pmatrix}|^2 \qquad (2.17)$$

where

$$E|[\mathbf{C}_1^T, \mathbf{C}_2^T]\begin{pmatrix} \hat{\phi}_1 - \theta_1^0 \\ \hat{\phi}_2 - \theta_2^0 \end{pmatrix}|^2 = E|\mathbf{C}_1^T[\hat{\phi}_1 - \theta_1^0] + \mathbf{C}_2^T[\hat{\phi}_2 - \theta_2^0]|^2$$

$$= E(\mathbf{C}_1^T[\hat{\phi}_1 - \theta_1^0] + \mathbf{C}_2^T[\hat{\phi}_2 - \theta_2^0])([\hat{\phi}_1 - \theta_1^0]^T\mathbf{C}_1 + [\hat{\phi}_2 - \theta_2^0]^T\mathbf{C}_2)$$

$$= E(\mathbf{C}_1^T[\hat{\phi}_1 - \theta_1^0][\hat{\phi}_1 - \theta_1^0]^T\mathbf{C}_1 + \mathbf{C}_1^T[\hat{\phi}_1 - \theta_1^0][\hat{\phi}_2 - \theta_2^0]^T\mathbf{C}_2$$
$$+ \mathbf{C}_2^T[\hat{\phi}_2 - \theta_2^0][\hat{\phi}_1 - \theta_1^0]^T\mathbf{C}_1 + \mathbf{C}_2^T[\hat{\phi}_2 - \theta_2^0][\hat{\phi}_2 - \theta_2^0]^T\mathbf{C}_2)$$

$$= \mathbf{C}_1^T E[\hat{\phi}_1 - \theta_1^0][\hat{\phi}_1 - \theta_1^0]^T \mathbf{C}_1 + \mathbf{C}_1^T E[\hat{\phi}_1 - \theta_1^0][\hat{\phi}_2 - \theta_2^0]^T \mathbf{C}_2 + \mathbf{C}_2^T E[\hat{\phi}_2 - \theta_2^0][\hat{\phi}_1 - \theta_1^0]^T \mathbf{C}_1 + \mathbf{C}_2^T E[\hat{\phi}_2 - \theta_2^0][\hat{\phi}_2 - \theta_2^0]^T \mathbf{C}_2)$$

$$= \mathbf{C}_1^T \Phi_{11} \mathbf{C}_1 + \mathbf{C}_1^T \Phi_{12} \mathbf{C}_2 + \mathbf{C}_2^T \Phi_{21} \mathbf{C}_1 + \mathbf{C}_2^T \Phi_{22} \mathbf{C}_2 \tag{2.18}$$

Likewise,

$$E\|[\mathbf{C}_1^T, \mathbf{C}_2^T]\begin{pmatrix} \hat{\psi}_1 - \theta_1^0 \\ \hat{\psi}_2 - \theta_2^0 \end{pmatrix}\|^2 = \mathbf{C}_1^T \Psi_{11} \mathbf{C}_1 + \mathbf{C}_1^T \Psi_{12} \mathbf{C}_2 + \mathbf{C}_2^T \Psi_{21} \mathbf{C}_1 + \mathbf{C}_2^T \Psi_{22} \mathbf{C}_2 \tag{2.19}$$

so that

$$\mathbf{C}_1^T \Phi_{11} \mathbf{C}_1 + \mathbf{C}_1^T \Phi_{12} \mathbf{C}_2 + \mathbf{C}_2^T \Phi_{21} \mathbf{C}_1 + \mathbf{C}_2^T \Phi_{22} \mathbf{C}_2 \geq \mathbf{C}_1^T \Psi_{11} \mathbf{C}_1 + \mathbf{C}_1^T \Psi_{12} \mathbf{C}_2 + \mathbf{C}_2^T \Psi_{21} \mathbf{C}_1 + \mathbf{C}_2^T \Psi_{22} \mathbf{C}_2 \tag{2.20}$$

We now calculate both sides of (2.15). The left-hand side of the inequality is given by the following:

$$E([\mathbf{C}^T(\hat{\phi} - \theta^0)][\mathbf{C}^T(\hat{\phi} - \theta^0)]^*) = E(|[\mathbf{C}_1^T + j\mathbf{C}_2^T][\hat{\phi}_1 - \theta_1^0 + j[\hat{\phi}_2 - \theta_2^0]|^2)$$

$$= E(|\mathbf{C}_1^T[\hat{\phi}_1 - \theta_1^0] - \mathbf{C}_2^T[\hat{\phi}_2 - \theta_2^0] + j(\mathbf{C}_1^T[\hat{\phi}_2 - \theta_2^0] + \mathbf{C}_2^T[\hat{\phi}_1 - \theta_1^0])|^2)$$

$$= E(|\mathbf{C}_1^T[\hat{\phi}_1 - \theta_1^0] - \mathbf{C}_2^T[\hat{\phi}_2 - \theta_2^0]|^2 + |\mathbf{C}_1^T[\hat{\phi}_2 - \theta_2^0] + \mathbf{C}_2^T[\hat{\phi}_1 - \theta_1^0]|^2)$$

$$= E((\mathbf{C}_1^T[\hat{\phi}_1 - \theta_1^0] - \mathbf{C}_2^T[\hat{\phi}_2 - \theta_2^0])([\hat{\phi}_1 - \theta_1^0]^T \mathbf{C}_1 - [\hat{\phi}_2 - \theta_2^0]^T \mathbf{C}_2) + (\mathbf{C}_1^T[\hat{\phi}_2 - \theta_2^0] + \mathbf{C}_2^T[\hat{\phi}_1 - \theta_1^0])([\hat{\phi}_2 - \theta_2^0]^T \mathbf{C}_1 + [\hat{\phi}_1 - \theta_1^0]^T \mathbf{C}_2))$$

$$= E(\mathbf{C}_1^T[\hat{\phi}_1 - \theta_1^0][\hat{\phi}_1 - \theta_1^0]^T \mathbf{C}_1 - \mathbf{C}_1^T[\hat{\phi}_1 - \theta_1^0][\hat{\phi}_2 - \theta_2^0]^T \mathbf{C}_2 - \mathbf{C}_2^T[\hat{\phi}_2 - \theta_2^0][\hat{\phi}_1 - \theta_1^0]^T \mathbf{C}_1 + \mathbf{C}_2^T[\hat{\phi}_2 - \theta_2^0][\hat{\phi}_2 - \theta_2^0]^T \mathbf{C}_2 + \mathbf{C}_1^T[\hat{\phi}_2 - \theta_2^0][\hat{\phi}_2 - \theta_2^0]^T \mathbf{C}_1 + \mathbf{C}_2^T[\hat{\phi}_1 - \theta_1^0][\hat{\phi}_2 - \theta_2^0]^T \mathbf{C}_1 + \mathbf{C}_1^T[\hat{\phi}_2 - \theta_2^0][\hat{\phi}_1 - \theta_1^0]^T \mathbf{C}_2 + \mathbf{C}_2^T[\hat{\phi}_1 - \theta_1^0][\hat{\phi}_1 - \theta_1^0]^T \mathbf{C}_2)$$

$$= \mathbf{C}_1^T E[\hat{\phi}_1 - \theta_1^0][\hat{\phi}_1 - \theta_1^0]^T \mathbf{C}_1 - \mathbf{C}_1^T E[\hat{\phi}_1 - \theta_1^0][\hat{\phi}_2 - \theta_2^0]^T \mathbf{C}_2 - \mathbf{C}_2^T E[\hat{\phi}_2 - \theta_2^0][\hat{\phi}_1 - \theta_1^0]^T \mathbf{C}_1 + \mathbf{C}_2^T E[\hat{\phi}_2 - \theta_2^0][\hat{\phi}_2 - \theta_2^0]^T \mathbf{C}_2 + \mathbf{C}_1^T E[\hat{\phi}_2 - \theta_2^0][\hat{\phi}_2 - \theta_2^0]^T \mathbf{C}_1 + \mathbf{C}_2^T E[\hat{\phi}_1 - \theta_1^0][\hat{\phi}_2 - \theta_2^0]^T \mathbf{C}_1 + \mathbf{C}_1^T E[\hat{\phi}_2 - \theta_2^0][\hat{\phi}_1 - \theta_1^0]^T \mathbf{C}_2 + \mathbf{C}_2^T E[\hat{\phi}_1 - \theta_1^0][\hat{\phi}_1 - \theta_1^0]^T \mathbf{C}_2)$$

$$= \mathbf{C}_1^T\Phi_{11}\mathbf{C}_1 - \mathbf{C}_1^T\Phi_{12}\mathbf{C}_2 - \mathbf{C}_2^T\Phi_{21}\mathbf{C}_1 + \mathbf{C}_2^T\Phi_{22}\mathbf{C}_2$$
$$+ \mathbf{C}_1^T\Phi_{22}\mathbf{C}_1 + \mathbf{C}_1^T\Phi_{21}\mathbf{C}_2 + \mathbf{C}_2^T\Phi_{12}\mathbf{C}_1 + \mathbf{C}_2^T\Phi_{11}\mathbf{C}_2 \quad (2.21)$$

Likewise, the right-hand side of (2.15) is given by the following,

$$E([\mathbf{C}^T(\hat{\psi} - \theta^0)][\mathbf{C}^T(\hat{\psi} - \theta^0)]^*) = \mathbf{C}_1^T\Psi_{11}\mathbf{C}_1 - \mathbf{C}_1^T\Psi_{12}\mathbf{C}_2 - \mathbf{C}_2^T\Psi_{21}\mathbf{C}_1 + \mathbf{C}_2^T\Psi_{22}\mathbf{C}_2$$
$$+ \mathbf{C}_1^T\Psi_{22}\mathbf{C}_1 + \mathbf{C}_1^T\Psi_{21}\mathbf{C}_2 + \mathbf{C}_2^T\Psi_{12}\mathbf{C}_1 + \mathbf{C}_2^T\Psi_{11}\mathbf{C}_2 \quad (2.22)$$

Noticing that (2.20) holds for any $K$-dimensional real-valued vector, we exchange $\mathbf{C}_1$ and $\mathbf{C}_2$ in (2.20), and obtain the following expression

$$\mathbf{C}_1^T\Phi_{22}\mathbf{C}_1 + \mathbf{C}_1^T\Phi_{21}\mathbf{C}_2 + \mathbf{C}_2^T\Phi_{12}\mathbf{C}_1 + \mathbf{C}_2^T\Phi_{11}\mathbf{C}_2 \geq$$
$$\mathbf{C}_1^T\Psi_{22}\mathbf{C}_1 + \mathbf{C}_1^T\Psi_{21}\mathbf{C}_2 + \mathbf{C}_2^T\Psi_{12}\mathbf{C}_1 + \mathbf{C}_2^T\Psi_{11}\mathbf{C}_2 \quad (2.23)$$

Replacing $\mathbf{C}_1$ by $-\mathbf{C}_1$ in (2.20), we have

$$\mathbf{C}_1^T\Phi_{11}\mathbf{C}_1 - \mathbf{C}_1^T\Phi_{12}\mathbf{C}_2 - \mathbf{C}_2^T\Phi_{21}\mathbf{C}_1 + \mathbf{C}_2^T\Phi_{22}\mathbf{C}_2 \geq$$
$$\mathbf{C}_1^T\Psi_{11}\mathbf{C}_1 - \mathbf{C}_1^T\Psi_{12}\mathbf{C}_2 - \mathbf{C}_2^T\Psi_{21}\mathbf{C}_1 + \mathbf{C}_2^T\Psi_{22}\mathbf{C}_2 \quad (2.24)$$

Combining (2.23) and (2.24), we obtain

$$\mathbf{C}_1^T\Phi_{11}\mathbf{C}_1 - \mathbf{C}_1^T\Phi_{12}\mathbf{C}_2 - \mathbf{C}_2^T\Phi_{21}\mathbf{C}_1 + \mathbf{C}_2^T\Phi_{22}\mathbf{C}_2$$
$$+\mathbf{C}_1^T\Phi_{22}\mathbf{C}_1 + \mathbf{C}_1^T\Phi_{21}\mathbf{C}_2 + \mathbf{C}_2^T\Phi_{12}\mathbf{C}_1 + \mathbf{C}_2^T\Phi_{11}\mathbf{C}_2$$

$$\geq \mathbf{C}_1^T\Psi_{11}\mathbf{C}_1 - \mathbf{C}_1^T\Psi_{12}\mathbf{C}_2 - \mathbf{C}_2^T\Psi_{21}\mathbf{C}_1 + \mathbf{C}_2^T\Psi_{22}\mathbf{C}_2$$
$$+\mathbf{C}_1^T\Psi_{22}\mathbf{C}_1 + \mathbf{C}_1^T\Psi_{21}\mathbf{C}_2 + \mathbf{C}_2^T\Psi_{12}\mathbf{C}_1 + \mathbf{C}_2^T\Psi_{11}\mathbf{C}_2$$

that is,

$$E([\mathbf{C}^T(\hat{\phi} - \theta^0)][\mathbf{C}^T(\hat{\phi} - \theta^0)]^*) \geq E([\mathbf{C}^T(\hat{\psi} - \theta^0)][\mathbf{C}^T(\hat{\psi} - \theta^0)]^*)$$

which completes the proof.

With Theorem 1, Theorem 2, and Theorem 3, we can obtain the formula of the CRB of estimating the amplitudes in model (1.22) that is given by the following theorem:

**Theorem 4:** The CRB of estimating the amplitudes in model (1.22) is given by

$$\sigma^2[\mathbf{S}^H(\varpi_0)\mathbf{S}(\varpi_0)]^{-1}+(\sigma^2/2)\ \mathbf{D}(\varpi_0)\mathbf{R}^{-1}\ \mathbf{D}^H(\varpi_0) \qquad (2.25)$$

Proof: Applying the Taylor expansion, the model (1.22) can be linearized as follows:

$$\begin{aligned}\mathbf{x} &= \mathbf{S}(\varpi_0)\mathbf{a}_0+\mathbf{S}(\varpi_0)[\mathbf{a}-\mathbf{a}_0]+\mathbf{S}'(\varpi_0)\mathbf{A}_0[\varpi-\varpi_0]+\mathbf{w}\\ &=\mathbf{S}(\varpi_0)\mathbf{a}+\mathbf{S}'(\varpi_0)\mathbf{A}_0[\varpi-\varpi_0]+\mathbf{w}\end{aligned} \qquad (2.26)$$

where $\varpi_0 = [\omega_1^0,\omega_2^0,\ldots\omega_K^0]^T$ and $\mathbf{a}_0 = [a_1^0,a_2^0,\ldots,a_K^0]^T$ are the true values of $\varpi$ and $\mathbf{a}$, respectively; $\mathbf{S}'(\varpi_0) = [\mathbf{s}'(\omega_1^0),\mathbf{s}'(\omega_2^0),\ldots,\ \mathbf{s}'(\omega_K^0)]$, and $\mathbf{s}'(\omega)$ is the derivative of $\mathbf{s}(\omega)$ with respect to $\omega$ ; $\mathbf{A}_0$ is the diagonal matrix $\{a_1^0,a_2^0,\ldots,a_K^0\}$.

Let $$\mathbf{P}(\varpi_0)=\mathbf{S}(\varpi_0)[\mathbf{S}^H(\varpi_0)\mathbf{S}(\varpi_0)]^{-1}\mathbf{S}^H(\varpi_0) \qquad (2.27)$$

and $$\mathbf{Q}(\varpi_0)=\mathbf{I}-\mathbf{P}(\varpi_0) \qquad (2.28)$$

From (2.26)-(2.28), it follows that

$$\begin{aligned}\mathbf{x} &= \mathbf{S}(\varpi_0)\mathbf{a}+\mathbf{P}(\varpi_0)\mathbf{S}'(\varpi_0)\mathbf{A}_0[\varpi-\varpi_0]+\mathbf{Q}(\varpi_0)\mathbf{S}'(\varpi_0)\mathbf{A}_0[\varpi-\varpi_0]+\mathbf{w}\\ &= \mathbf{S}(\varpi_0)\{\mathbf{a}+[\mathbf{S}^H(\varpi_0)\mathbf{S}(\varpi_0)]^{-1}\mathbf{S}^H(\varpi_0)\mathbf{S}'(\varpi_0)\mathbf{A}_0[\varpi-\varpi_0]\}\\ &\quad+\mathbf{Q}(\varpi_0)\mathbf{S}'(\varpi_0)\mathbf{A}_0[\varpi-\varpi_0]+\mathbf{w}\\ &= \mathbf{S}(\varpi_0)\mathbf{b}+\mathbf{Q}(\varpi_0)\mathbf{S}'(\varpi_0)\mathbf{A}_0[\varpi-\varpi_0]+\mathbf{w}\end{aligned} \qquad (2.29)$$

In (2.29), $$\mathbf{b} = \mathbf{a}+\mathbf{D}(\varpi_0)[\varpi-\varpi_0] \qquad (2.30)$$

and $$\mathbf{D}(\varpi_0)=[\mathbf{S}^H(\varpi_0)\mathbf{S}(\varpi_0)]^{-1}\mathbf{S}^H(\varpi_0)\mathbf{S}'(\varpi_0)\mathbf{A}_0 \qquad (2.31)$$

Because $\mathbf{S}^H(\varpi_0)\mathbf{Q}(\varpi_0)=\mathbf{S}^H(\varpi_0)-\mathbf{S}^H(\varpi_0)\mathbf{P}(\varpi_0)=\mathbf{S}^H(\varpi_0)-\mathbf{S}^H(\varpi_0)=0$, so $\mathbf{S}^H(\varpi_0)\mathbf{Q}(\varpi_0)\mathbf{S}'(\varpi_0)\mathbf{A}_0=0$, i.e., $\mathbf{S}(\varpi_0)$ and $\mathbf{Q}(\varpi_0)\mathbf{S}'(\varpi_0)\mathbf{A}_0$ are orthogonal; by Theorem 1, the LSE $\tilde{\mathbf{b}}$ and $(\tilde{\varpi}-\varpi_0)$ of the linear expansion (2.25) are uncorrelated and can be separately obtained from the following models:

$$\mathbf{x} = \mathbf{S}(\varpi_0)\mathbf{b} + \mathbf{w} \tag{2.32}$$

and

$$\mathbf{x} = \mathbf{Q}(\varpi_0)\mathbf{S}^{'}(\varpi_0)\mathbf{A}_0[\varpi - \varpi_0] + \mathbf{w} \tag{2.33}$$

From (2.32), we can easily obtain

$$\tilde{\mathbf{b}} = [\mathbf{S}^H(\varpi_0)\mathbf{S}(\varpi_0)]^{-1}\mathbf{S}^H(\varpi_0)\mathbf{x} \tag{2.34}$$

As the real-valued LSE in the linear model (2.33), $\tilde{\varpi} - \varpi_0$ can be explicitly given in a closed form. As a matter of fact, let

$$\mathbf{x} = \mathbf{x}_r + j\mathbf{x}_i, \qquad \mathbf{w} = \mathbf{w}_r + j\mathbf{w}_i$$

$$\mathbf{Q}(\varpi_0)\mathbf{S}^{'}(\varpi_0)\mathbf{A}_0 = \mathbf{U} + j\mathbf{V}$$

where $\mathbf{x}_r$, $\mathbf{w}_r$, and $\mathbf{U}$ are the real part, and $\mathbf{x}_i$, $\mathbf{w}_i$, and $\mathbf{V}$ are the imaginary part. Thus, we have

$$\mathbf{x}_r = \mathbf{U}(\varpi - \varpi_0) + \mathbf{w}_r, \qquad \mathbf{x}_i = \mathbf{V}(\varpi - \varpi_0) + \mathbf{w}_i$$

According to (1.23), $\mathbf{w}_r$ and $\mathbf{w}_i$ are independent Gaussian random vectors with

$$E(\mathbf{w}_r\mathbf{w}_r^T) = E(\mathbf{w}_i\mathbf{w}_i^T) = (\sigma^2/2)\mathbf{I}$$

It is straightforward to have

$$(\tilde{\varpi} - \varpi_0) = (\mathbf{U}^T\mathbf{U} + \mathbf{V}^T\mathbf{V})^{-1}(\mathbf{U}^T\mathbf{x}_r + \mathbf{V}^T\mathbf{x}_i)$$

where

$$\mathbf{U}^T\mathbf{U} + \mathbf{V}^T\mathbf{V} = \mathrm{Re}\{\mathbf{A}_0^H[\mathbf{S}^{'}(\varpi_0)]^H\mathbf{Q}(\varpi_0)\mathbf{S}^{'}(\varpi_0)\mathbf{A}_0\}$$

$$\mathbf{U}^T\mathbf{x}_r + \mathbf{V}^T\mathbf{x}_i = \mathrm{Re}\{\mathbf{A}_0^H[\mathbf{S}^{'}(\varpi_0)]^H\mathbf{Q}(\varpi_0)\mathbf{x}\}$$

Thus,

$$(\tilde{\varpi} - \varpi_0) = \mathbf{R}^{-1}\mathrm{Re}\{\mathbf{A}_0^H[\mathbf{S}^{'}(\varpi_0)]^H\mathbf{Q}(\varpi_0)\mathbf{x}\} \tag{2.35}$$

where

$$\mathbf{R} = \mathrm{Re}\{ \mathbf{A}_0^H [\mathbf{S}'(\varpi_0)]^H \mathbf{Q}(\varpi_0)\mathbf{S}'(\varpi_0)\mathbf{A}_0 \}, \tag{2.36}$$

From (2.30) it follows that the LSE $\tilde{\mathbf{a}}$ is given by

$$\tilde{\mathbf{a}} = \tilde{\mathbf{b}} - \mathbf{D}(\varpi_0)[\tilde{\varpi} - \varpi_0] \tag{2.37}$$

and

$$\tilde{\mathbf{a}} - \mathbf{a}_0 = \tilde{\mathbf{b}} - \mathbf{b}_0 - \mathbf{D}(\varpi_0)[\tilde{\varpi} - \varpi_0] \equiv \mathbf{N} + \tilde{\mathbf{C}} \tag{2.38}$$

here

$$\mathbf{N} = \tilde{\mathbf{b}} - \mathbf{b}_0 = [\mathbf{S}^H(\varpi_0)\mathbf{S}(\varpi_0)]^{-1} \mathbf{S}^H(\varpi_0)\mathbf{w} \tag{2.39}$$

and

$$\sum(\mathbf{N}) = E(\mathbf{N}\mathbf{N}^H) = \sigma^2 [\mathbf{S}^H(\varpi_0)\mathbf{S}(\varpi_0)]^{-1} \tag{2.40}$$

$\tilde{\mathbf{C}}$ in (2.38) is caused by the error between $\tilde{\varpi}$ and $\varpi_0$. It is known (Stoica, 1989) that

$$\mathrm{E}\,[\tilde{\varpi} - \varpi_0][\tilde{\varpi} - \varpi_0]^T = (\sigma^2/2)\ \mathbf{R}^{-1} \tag{2.41}$$

where $\mathbf{R}$ is given in (2.36), and where (2.41) is the CRB of estimating $\varpi$.

Since $\mathbf{N}$ and $\tilde{\mathbf{C}}$ are uncorrelated, thus the covariance matrix of $\hat{\mathbf{a}}$ is given by

$$\mathrm{E}\,[\tilde{\mathbf{a}} - \mathbf{a}_0][\tilde{\mathbf{a}} - \mathbf{a}_0]^H = \mathrm{E}\,\mathbf{N}\mathbf{N}^H + \mathrm{E}\,\tilde{\mathbf{C}}\tilde{\mathbf{C}}^H =$$

$$= \sigma^2 [\mathbf{S}^H(\varpi_0)\mathbf{S}(\varpi_0)]^{-1} + (\sigma^2/2)\ \mathbf{D}(\varpi_0)\mathbf{R}^{-1}\mathbf{D}^H(\varpi_0) \tag{2.42}$$

By Theorem 3, the right-hand side in (2.42) is the CRB of estimating the amplitudes with unknown frequencies in the case of white noise, which completes the proof.

# CHAPTER 3

# OPTIMAL AMPLITUDE ESTIMATOR

In Chapter 2, we derived the formula of the CRB of estimating the amplitudes, which should be very useful in practical applications and theoretical studies to compare the performance of a given estimator to the ultimate performance corresponding to the CRB. We say an estimator of a parameter is optimal if it is both consistent and efficient. The LSE of the amplitudes in Section 2.1 is optimal when the frequencies are known. It is not such a case that we can find an efficient estimator of a parameter all the time; however, we can always find an asymptotically efficient estimator of the parameter; that is, the covariance of the estimator asymtotically approaches the CRB of estimating the parameter when the signal noise ratio (SNR) goes to infinity. The maximum likelihood estimator (MLE) of the amplitudes, as shown in this chapter, is consistent and asymptotically efficient.

We have organized this chapter as follows. In Section 1.1, we derive the MLE of the amplitudes in the superimposed sinusoidal signal model (1.21) or (1.22). Then in Section 1.2, we show that the MLE of the amplitudes is asymptotically efficient. Finally, in Section 1.3, we show that the optimal resolution of signals using a finite sample is given by the MLE of the amplitudes of superimposed sinusoidal signals.

## 3.1 MLE of the Amplitudes

To evaluate the statistical model of the observed data, recall that the noise $w(\bullet)$ was assumed to be a complex Gaussian random process with zero mean and covariance matrix of the form $\sigma^2\mathbf{I}$. Thus, it follows that the joint density function of the observed data $\mathbf{x}$ is given by

$$f(\mathbf{x}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{|x(i) - \sum_{k=1}^{K} a_k e^{j\varpi_k T_i}|^2}{2\sigma^2}\right)$$

$$= \frac{1}{(\sqrt{2\pi}\sigma)^N} \exp\left(-\frac{\|\mathbf{x} - \mathbf{S}(\varpi)\mathbf{a}\|^2}{2\sigma^2}\right) \tag{3.1}$$

The log-likelihood function of $(\mathbf{a}, \varpi)$ is given by

$$\mathbf{L}(\mathbf{a}, \varpi) = -\frac{N}{2}\log(2\pi\sigma^2) - \frac{\|\mathbf{x} - \mathbf{S}(\varpi)\mathbf{a}\|^2}{2\sigma^2} \tag{3.2}$$

The maximum likelihood estimates (MLE) of the unknown amplitudes $\mathbf{a}$ and unknown frequencies $\varpi$ are the values that maximize the log-likelihood function in (3.2) or, equivalently, the values that minimize the following expression:

$$\mathbf{U}(\mathbf{a}, \varpi) = \|\mathbf{x} - \mathbf{S}(\varpi)\mathbf{a}\|^2 \tag{3.3}$$

The minimization of $\mathbf{U}(\mathbf{a}, \varpi)$, for fixed $\varpi$, can be achieved when the vector $\mathbf{x} - \mathbf{S}(\varpi)\mathbf{a}$ is orthogonal to the subspace $S(\varpi)$ spanned by $\mathbf{s}(\varpi_1), \cdots, \mathbf{s}(\varpi_K)$. The orthogonality requires that the complex inner product of $\mathbf{x} - \mathbf{S}(\varpi)\mathbf{a}$ with each of $\mathbf{s}(\varpi_1), \cdots, \mathbf{s}(\varpi_K)$ be nullified, i.e.,

$$\mathbf{s}^H(\varpi_i)(\mathbf{x} - \mathbf{S}(\varpi)\mathbf{a}) = 0, \qquad i = 1, 2, \cdots, K \tag{3.4}$$

that is,

$$\mathbf{S}^H(\varpi)(\mathbf{x} - \mathbf{S}(\varpi)\mathbf{a}) = 0 \tag{3.5}$$

The solution for $\mathbf{a}$ in (3.5), for fixed $\varpi$, is then given by

$$\mathbf{a}(\varpi) = [\mathbf{S}^H(\varpi)\mathbf{S}(\varpi)]^{-1}\mathbf{S}^H(\varpi)\mathbf{x} \tag{3.6}$$

Applying the Pythagorean theorem to (3.3) leads to

$$\begin{aligned} U(\mathbf{a}(\varpi), \varpi) &= \|\mathbf{x} - \mathbf{S}(\varpi)\mathbf{a}(\varpi)\|^2 \\ &= \mathbf{x}^H\mathbf{x} - \|\mathbf{S}(\varpi)\mathbf{a}(\varpi)\|^2 \\ &= \mathbf{x}^H\mathbf{x} - \mathbf{a}^H(\varpi)\mathbf{S}^H(\varpi)\mathbf{S}(\varpi)\mathbf{a}(\varpi) \\ &= \mathbf{x}^H\mathbf{x} - \mathbf{x}^H\mathbf{S}(\varpi)[\mathbf{S}^H(\varpi)\mathbf{S}(\varpi)]^{-1}\mathbf{S}^H(\varpi)\mathbf{x} \\ &= \mathbf{x}^H\mathbf{x} - \mathbf{x}^H\mathbf{P}(\varpi)\mathbf{x} \\ &= \mathbf{x}^H\mathbf{Q}(\varpi)\mathbf{x} \end{aligned} \tag{3.7}$$

where $\mathbf{P}(\varpi) = \mathbf{S}(\varpi)[\mathbf{S}^H(\varpi)\mathbf{S}(\varpi)]^{-1}\mathbf{S}^H(\varpi)$ is the projection operator over the vector space $\mathbf{C}^N$, $\mathbf{Q}(\varpi) = \mathbf{I} - \mathbf{P}(\varpi)$ and $\mathbf{Q}(\varpi)\mathbf{S}(\varpi) = 0$.

To minimize $U(\mathbf{a}(\varpi), \varpi)$ over $\varpi$, the MLE $\hat{\varpi}$ needs to maximize $\mathbf{x}^H\mathbf{P}(\varpi)\mathbf{x}$, i.e.,

$$\hat{\varpi} = \arg \max_{\varpi} \mathbf{x}^H \mathbf{P}(\varpi)\mathbf{x} \tag{3.8}$$

and

$$\hat{\mathbf{a}} = [\mathbf{S}^H(\hat{\varpi})\mathbf{S}(\hat{\varpi})]^{-1}\mathbf{S}^H(\hat{\varpi})\mathbf{x} \tag{3.9}$$

## 3.2 Asymptotic Efficiency of the MLE

In this section, we are going to show that the MLE $\hat{\varpi}$ in (3.8) and the MLE $\hat{\mathbf{a}}$ are asymptotically efficient as the noise power $\sigma^2 \to 0$, or, equivalently, as the SNR is large. The white Gaussian noise $\mathbf{w} = \sigma\,\mathbf{z}$ and $\mathbf{z}$ satisfies

$$E(\mathbf{z}) = \mathbf{0}, \quad E(\mathbf{z}\mathbf{z}^H) = \mathbf{I} \quad \text{and} \quad E(\mathbf{z}\mathbf{z}^T) = \mathbf{0}$$

We define

$$C(\varpi) \equiv \mathbf{x}^H\mathbf{P}(\varpi)\mathbf{x} = \|\mathbf{P}(\varpi)\mathbf{x}\|^2$$

$$F(\varpi,\varpi_0) \equiv \mathbf{a}_0^H \mathbf{S}^H(\varpi_0)\mathbf{P}(\varpi)\mathbf{S}(\varpi_0)\mathbf{a}_0 = \|\mathbf{P}(\varpi)\mathbf{S}(\varpi_0)\mathbf{a}_0\|^2$$

Over the domain $\Omega$ that is a compact region in the $\varpi$-space containing $\varpi_0$ as an interior point.

The geometric interpretation of $C(\varpi)$ is the length of the projection of $\mathbf{x}$ onto the subspace $S(\varpi)$. Obviously, $F(\varpi,\varpi_0)$ has a unique maximum at $\varpi_0$ in $\Omega$.

**Theorem 5**: $\hat{\varpi}$ is statistically consistent as $\sigma \to 0$.

Proof: When $\mathbf{x} = \mathbf{S}(\varpi_0)\mathbf{a}_0 + \mathbf{w} = \mathbf{S}(\varpi_0)\mathbf{a}_0 + \sigma\,\mathbf{z}$, it follows from (3.8) that

$$C(\varpi) = [\mathbf{S}(\varpi_0)\mathbf{a}_0 + \sigma\,\mathbf{z}]^H\,\mathbf{P}(\varpi)[\mathbf{S}(\varpi_0)\mathbf{a}_0 + \sigma\,\mathbf{z}] \to F(\varpi,\varpi_0) \tag{3.10}$$

uniformly in $\varpi \in \Omega$, as $\sigma \to 0$, almost surely.

Thus, for any $\varepsilon > 0$, let

$$\Omega_\varepsilon = \{\varpi,\ \varpi \in \Omega \text{ and } \|\varpi - \varpi_0\| \geq \varepsilon\}$$

and

$$d = F(\varpi_0,\varpi_0) - \max_{\varpi \in \Omega_\varepsilon} F(\varpi,\varpi_0) > 0, \tag{3.11}$$

Thus, almost surely, $\exists\,\delta > 0$ such that whenever $\sigma < \delta$, we have , for all $\varpi \in \Omega$

$$|C(\varpi) - F(\varpi,\varpi_0)| < d/2 \tag{3.12}$$

In particular, for $\varpi = \varpi_0$, we have

$$F(\varpi_0,\varpi_0) - d/2 < C(\varpi_0) < F(\varpi_0,\varpi_0) + d/2 \tag{3.13}$$

On the other hand, for $\varpi$ in $\Omega_\varepsilon$, we have

$$C(\varpi) < F(\varpi,\varpi_0) - d/2 < \max_{\varpi \in \Omega_\varepsilon} F(\varpi,\varpi_0) + d/2 = F(\varpi_0,\varpi_0) - d/2 \tag{3.14}$$

(3.13.) and (3.14) imply that $\hat{\varpi} \in \Omega$, maximizing $C(\varpi)$, cannot be in $\Omega_\varepsilon$. In other words, $\|\hat{\varpi} - \varpi_0\| < \varepsilon$, when $\sigma < \delta$. This completes the proof of consistency.

**Theorem 6:** $\hat{\varpi}$ is asymptotically efficient as $\sigma \to 0$.

Proof: Expanding $C'(\varpi)$ in $(\varpi - \varpi_0)$ and noting $C'(\hat{\varpi}) = 0$ lead to

$$C'(\hat{\varpi}) - C'(\varpi_0) = C''(\bar{\varpi})(\hat{\varpi} - \varpi_0) = -C'(\varpi_0) \tag{3.15}$$

where $\bar{\varpi} \to \varpi_0$ as $\hat{\varpi} \to \varpi_0$ and where $C'(\varpi)$ is the vector of derivatives of $C(\varpi)$ and $C''(\varpi)$ is the matrix of 2nd derivatives of $C(\varpi)$.

From Theorem 5 and (3.10), it follows that, as $\sigma \to 0$,

$$C''(\bar{\varpi}) \to F''(\varpi_0, \varpi_0), \text{ almost surely,} \tag{3.16}$$

Differentiate $F(\varpi, \varpi_0)$ with respect to $\varpi_i$ in $\varpi = [\varpi_1, \cdots, \varpi_K]$

$$\frac{\partial}{\partial \varpi_i} \mathbf{F}(\varpi, \varpi_0) = \mathbf{a}_0^H \mathbf{S}^H(\varpi_0) \frac{\partial}{\partial \varpi_i} \mathbf{P}(\varpi) \mathbf{S}(\varpi_0) \mathbf{a}_0 \tag{3.17}$$

Making use of general rules for differentiation, and letting

$$\frac{\partial}{\partial \varpi_i} \mathbf{S}(\varpi) = [0, \cdots, 0, \mathbf{s}'(\varpi_i), 0, \cdots, 0] = D_i \mathbf{S}(\varpi)$$

Noting that

$$\frac{\partial}{\partial \varpi_i} [\mathbf{S}^H(\varpi)\mathbf{S}(\varpi)]^{-1} = [\mathbf{S}^H(\varpi)\mathbf{S}(\varpi)]^{-1} \{D_i \mathbf{S}^H(\varpi)\mathbf{S}(\varpi) + \mathbf{S}^H(\varpi) D_i \mathbf{S}(\varpi)\} [\mathbf{S}^H(\varpi)\mathbf{S}(\varpi)]^{-1}$$

we have

$$\begin{aligned}
\frac{\partial}{\partial \varpi_i} \mathbf{P}(\varpi) &= \frac{\partial}{\partial \varpi_i} [\mathbf{S}(\varpi)[\mathbf{S}^H(\varpi)\mathbf{S}(\varpi)]^{-1} \mathbf{S}^H(\varpi)] \\
&= D_i \mathbf{S}(\varpi)[\mathbf{S}^H(\varpi)\mathbf{S}(\varpi)]^{-1}\mathbf{S}^H(\varpi) + \mathbf{S}(\varpi)[\mathbf{S}^H(\varpi)\mathbf{S}(\varpi)]^{-1} D_i \mathbf{S}^H(\varpi) \\
&\quad - \mathbf{S}(\varpi)[\mathbf{S}^H(\varpi)\mathbf{S}(\varpi)]^{-1}\{D_i \mathbf{S}^H(\varpi)\mathbf{S}(\varpi) + \mathbf{S}^H(\varpi) D_i \mathbf{S}(\varpi)\}[\mathbf{S}^H(\varpi)\mathbf{S}(\varpi)]^{-1}\mathbf{S}^H(\varpi) \\
&= \{\mathbf{I} - \mathbf{S}(\varpi)[\mathbf{S}^H(\varpi)\mathbf{S}(\varpi)]^{-1}\mathbf{S}^H(\varpi)\} D_i \mathbf{S}(\varpi)[\mathbf{S}^H(\varpi)\mathbf{S}(\varpi)]^{-1}\mathbf{S}^H(\varpi) \\
&\quad + \mathbf{S}(\varpi)[\mathbf{S}^H(\varpi)\mathbf{S}(\varpi)]^{-1} D_i \mathbf{S}^H(\varpi)\{\mathbf{I} - \mathbf{S}(\varpi)[\mathbf{S}^H(\varpi)\mathbf{S}(\varpi)]^{-1}\mathbf{S}^H(\varpi)\}
\end{aligned}$$

$$= \mathbf{Q}(\varpi)D_i\mathbf{S}(\varpi)[\mathbf{S}^H(\varpi)\mathbf{S}(\varpi)]^{-1}\mathbf{S}^H(\varpi) + \mathbf{S}(\varpi)[\mathbf{S}^H(\varpi)\mathbf{S}(\varpi)]^{-1}D_i\mathbf{S}^H(\varpi)\mathbf{Q}(\varpi) \quad (3.18)$$

Combining (3.17) and (3.18) leads to

$$\frac{\partial}{\partial\varpi_i}\mathbf{F}(\varpi,\varpi_0) = 2\,\mathrm{Re}\{\mathbf{a}_0^H\mathbf{S}^H(\varpi_0)\mathbf{S}(\varpi)[\mathbf{S}^H(\varpi)\mathbf{S}(\varpi)]^{-1}D_i\mathbf{S}^H(\varpi)\mathbf{Q}(\varpi)\mathbf{S}(\varpi_0)\mathbf{a}_0\} \quad (3.19)$$

Differentiating with respect to $\varpi_j$ and noting $\mathbf{Q}(\varpi)\mathbf{S}(\varpi) = 0$, we have

$$\frac{\partial^2}{\partial\varpi_i\partial\varpi_j}\mathbf{F}(\varpi_0,\varpi_0) = -2\,\mathrm{Re}\{\mathbf{a}_0^H D_i\mathbf{S}^H(\varpi_0)\mathbf{Q}(\varpi_0)D_j\mathbf{S}(\varpi_0)\mathbf{a}_0\}$$

and therefore

$$\mathbf{F}''(\varpi_0,\varpi_0) = -2\,\mathrm{Re}\{\mathbf{A}_0^H[\mathbf{S}'(\varpi_0)]^H\,\mathbf{Q}(\varpi_0)\mathbf{S}'(\varpi_0)\mathbf{A}_0\} = -2\,\mathbf{R} \quad (3.20)$$

Differentiating $\mathbf{C}(\varpi)$ with respect to $\varpi_i$ in $\varpi = [\varpi_1,\cdots,\varpi_K]$ and using (3.18), we have

$$\frac{\partial}{\partial\varpi_i}C(\varpi) = \mathbf{x}^H\frac{\partial}{\partial\varpi_i}\mathbf{P}(\varpi)\mathbf{x}$$

$$= \mathbf{x}^H\mathbf{Q}(\varpi)D_i\mathbf{S}(\varpi)[\mathbf{S}^H(\varpi)\mathbf{S}(\varpi)]^{-1}\mathbf{S}^H(\varpi)\mathbf{x} + \mathbf{x}^H\mathbf{S}(\varpi)[\mathbf{S}^H(\varpi)\mathbf{S}(\varpi)]^{-1}D_i\mathbf{S}^H(\varpi)\mathbf{Q}(\varpi)\mathbf{x}$$

$$= 2\,\mathrm{Re}\{\mathbf{x}^H\mathbf{Q}(\varpi)D_i\mathbf{S}(\varpi)[\mathbf{S}^H(\varpi)\mathbf{S}(\varpi)]^{-1}\mathbf{S}^H(\varpi)\mathbf{x}\}$$

Hence,

$$\frac{\partial}{\partial\varpi_i}C(\varpi_0) = 2\,\mathrm{Re}\{\mathbf{x}^H\mathbf{Q}(\varpi_0)D_i\mathbf{S}(\varpi_0)[\mathbf{S}^H(\varpi_0)\mathbf{S}(\varpi_0)]^{-1}\mathbf{S}^H(\varpi_0)\mathbf{x}\}$$

$$= 2\,\mathrm{Re}\{(\mathbf{a}_0^H\mathbf{S}^H(\varpi_0) + \mathbf{w}^H)\mathbf{Q}(\varpi_0)D_i\mathbf{S}(\varpi_0)[\mathbf{S}^H(\varpi_0)\mathbf{S}(\varpi_0)]^{-1}\mathbf{S}^H(\varpi_0)(\mathbf{S}(\varpi_0)\mathbf{a}_0 + \mathbf{w})\}$$

$$= 2\,\mathrm{Re}\{\mathbf{w}^H\mathbf{Q}(\varpi_0)D_i\mathbf{S}(\varpi_0)(\mathbf{a}_0 + [\mathbf{S}^H(\varpi_0)\mathbf{S}(\varpi_0)]^{-1}\mathbf{S}^H(\varpi_0)\mathbf{w})\}$$

Ignoring high order,

$$\frac{\partial}{\partial\varpi_i}C(\varpi_0) \approx 2\,\mathrm{Re}\{\mathbf{w}^H\mathbf{Q}(\varpi_0)D_i\mathbf{S}(\varpi_0)\mathbf{a}_0\} = 2\,\mathrm{Re}\{\mathbf{a}_0^H D_i\mathbf{S}^H(\varpi_0)\mathbf{Q}(\varpi_0)\mathbf{w}\}$$

Hence,

$$C'(\varpi_0) \approx 2\mathrm{Re}\{\mathbf{A}_0^H[\mathbf{S}'(\varpi_0)]^H\mathbf{Q}(\varpi_0)\mathbf{w}\}=2\mathrm{Re}\{\mathbf{A}_0^H[\mathbf{S}'(\varpi_0)]^H\mathbf{Q}(\varpi_0)\mathbf{x}\} \quad (3.21)$$

and (3.15) - (3.21) lead to, as $\sigma \rightarrow 0$,

$$\hat{\varpi} - \varpi_0 = \mathbf{R}^{-1}\mathrm{Re}\{\mathbf{A}_0^H[\mathbf{S}'(\varpi_0)]^H\mathbf{Q}(\varpi_0)\mathbf{x}\} \quad (3.22)$$

and thus $\hat{\varpi}$ is asymptotically normally distributed with mean $\varpi_0$ and covariance matrix

$$\mathrm{E}[\hat{\varpi}-\varpi_0][\hat{\varpi}-\varpi_0]^T = (\sigma^2/2)\,\mathbf{R}^{-1} \quad (3.23)$$

which completes the proof of Theorem 6.

**Theorem 7:** $\hat{\mathbf{a}}$ is consistent as $\sigma \rightarrow 0$.

Proof: From (3.9), it follows immediately that

$$\hat{\mathbf{a}} = [\mathbf{S}^H(\hat{\varpi})\mathbf{S}(\hat{\varpi})]^{-1}\mathbf{S}^H(\hat{\varpi})[\mathbf{S}(\varpi_0)\mathbf{a}_0 + \mathbf{w}] \quad (3.24)$$

The Taylor expansion of $\mathbf{S}(\varpi)$ in $(\varpi - \varpi_0)$ leads to

$$\mathbf{S}(\hat{\varpi})\mathbf{a}_0 = \mathbf{S}(\varpi_0)\mathbf{a}_0 + \mathbf{S}'(\bar{\varpi})\mathbf{A}_0[\hat{\varpi}-\varpi_0] \quad (3.25)$$

and

$$\hat{\mathbf{a}} = [\mathbf{S}^H(\hat{\varpi})\mathbf{S}(\hat{\varpi})]^{-1}\mathbf{S}^H(\hat{\varpi})\{\mathbf{S}(\hat{\varpi})\mathbf{a}_0 - \mathbf{S}'(\bar{\varpi})\mathbf{A}_0[\hat{\varpi}-\varpi_0] + \mathbf{w}\}$$

$$= \mathbf{a}_0 + [\mathbf{S}^H(\hat{\varpi})\mathbf{S}(\hat{\varpi})]^{-1}\mathbf{S}^H(\hat{\varpi})\mathbf{w}$$

$$- \hat{\mathbf{a}}\,[\mathbf{S}^H(\hat{\varpi})\mathbf{S}(\hat{\varpi})]^{-1}\mathbf{S}^H(\hat{\varpi})\mathbf{S}'(\bar{\varpi})\mathbf{A}_0[\hat{\varpi}-\varpi_0] \quad (3.26)$$

$$= \mathbf{a}_0 + \sigma[\mathbf{S}^H(\hat{\varpi})\mathbf{S}(\hat{\varpi})]^{-1}\mathbf{S}^H(\hat{\varpi})\mathbf{z} - [\mathbf{S}^H(\hat{\varpi})\mathbf{S}(\hat{\varpi})]^{-1}\mathbf{S}^H(\hat{\varpi})\mathbf{S}'(\bar{\varpi})\mathbf{A}_0[\hat{\varpi}-\varpi_0]$$

where the interpolation $\bar{\varpi} \rightarrow \varpi_0$ as $\hat{\varpi} \rightarrow \varpi_0$.

By Theorem 5, $\hat{\varpi} \rightarrow \varpi_0$, almost surely at high SNR. Hence, $\hat{\mathbf{a}} \rightarrow \mathbf{a}_0$, almost surely, as $\sigma \rightarrow 0$, which completes the proof of the Theorem.

**Theorem 8:** $\hat{\mathbf{a}}$ is asymptotically efficient as $\sigma \rightarrow 0$.

Proof: Neglecting higher order errors, (3.23) leads to, as SNR is large,

$$\hat{\mathbf{a}} = \mathbf{a}_0 + [\mathbf{S}^H(\varpi_0)\mathbf{S}(\varpi_0)]^{-1}\mathbf{S}^H(\varpi_0)\mathbf{w} -$$

$$[\mathbf{S}^H(\varpi_0)\mathbf{S}(\varpi_0)]^{-1}\mathbf{S}^H(\varpi_0)\mathbf{S}'(\varpi_0)\mathbf{A}_0[\hat{\varpi} - \varpi_0]$$

$$= \mathbf{a}_0 + \mathbf{N} - \mathbf{D}(\varpi_0)[\hat{\varpi} - \varpi_0] = \mathbf{a}_0 + \mathbf{N} + \hat{\mathbf{C}} \tag{3.27}$$

Equation (3.27) is exactly the expansion (2.38) except $\hat{\mathbf{C}}$ is caused by the error between the MLE $\hat{\varpi}$ and the true value $\varpi_0$. However, it follows from (3.19) that ($\hat{\varpi}$ - $\varpi_0$) has exactly the same expansion as ($\tilde{\varpi}$ - $\varpi_0$) when SNR are large. And it proves the asymptotic efficiency of the MLE $\hat{\mathbf{a}}$ at high SNR.

## 3.3 Optimal Resolution of Signals in a Finite Sample

The potential capability of signal resolutions using a finite sample and the optimal resolution has been a controversial issue (Kay, 1988; Marple, 1987). However, all agreed that the signal resolution, by its nature, is the detection of each of the multiple signals instead of the visual impressions of a simulation (Kay, 1988; Marple, 1987; Gu, 1998). In the superimposed sinusoidal signal model (2.21), the signal at $\omega_1$, for example, is the signal to be detected; then all other signals play a role of clutter or interference. The optimal filter of detecting $a_1 \mathbf{s}(\omega_1)$ against interference and noise is thus required to be the optimal unbiased estimator of the amplitude $a_1$. The requirement of unbiasness is for rejecting or nulling the interference, and the requirement of optimality is to minimize the noise variance or, equivalently, to maximize the SNR. As an immediate result of the asymptotic efficiency of the MLE of amplitudes, the optimal unbiased estimator of the amplitude $a_1$ is given by its MLE $\hat{a}_1$. Hence, interestingly, the optimal resolution for each of the $K$ signals is given by the MLE of the amplitudes, $\hat{\mathbf{a}}$.

From (3.8), it can be seen that $\hat{a}_1$ is a constrained matched filter (CMF) which nulls the signals at the estimated frequencies: $\hat{\omega}_2,...,\hat{\omega}_K$ and maximizes the SNR for the signal at $\hat{\omega}_1$. In general, the mismatch between $\hat{\varpi}$ and $\varpi_0$ causes a leakage of the interference at the output of the filter bank $\hat{\mathbf{a}}$. The leakage is given in (3.24) by $\hat{\mathbf{C}}$, while N in (3.24) is the noise output. The output interference power and noise power are given in the CRB in (2.42). The CRB can be used to set up an optimal CFAR (constant false alarm rate) detection for each of the $K$ signals.

# CHAPTER 4

# SIMULATIONS

In Chapter 2 we derived the formula of the CRB of estimating the amplitudes, and in Chapter 3 we showed that the MLE of the amplitudes is asymptotically efficient and the optimal resolution of signals using a finite sample is given by the MLE. In this chapter, we present the simulation results that illustrate the performances of the MLE and the Fourier transform.

In simulations, we are considering the following model:

$$x(n) = A\, e^{j2\pi\phi_1 T_s n} + B\, e^{j2\pi\phi_2 T_s n} + a\, e^{j2\pi\phi T_s n} + w(n), \quad n = 0,1,...31, \tag{4.1}$$

or

$$\mathbf{x} = A\ \mathbf{s}(\phi_1) + B\ \mathbf{s}(\phi_2) + a\,\mathbf{s}(\phi) + \mathbf{w} \tag{4.2}$$

where $A\,\mathbf{s}(\phi_1)$ and $B\,\mathbf{s}(\phi_2)$ are two strong clutters with unknown frequencies, $\phi_1$, $\phi_2$, and unknown amplitudes, A, B; $a\,\mathbf{s}(\phi)$ is a weak return echo with unknown Doppler frequency $\phi$ and amplitude a; $T_s$ is the radar repetition interval and $f_s = (1/T_s) = 320$ Hz, and $\mathbf{w}$ is vector of white noise with mean 0 and covariance $\sigma^2\mathbf{I}$.

We now present some examples in which $\sigma^2 = \frac{1}{12}$ and the classical frequency resolution limit is $\frac{1}{NT_s} = 10$ Hz. Note that this model is a special case of the general model presented in (1.21) or (1.22).

This chapter is organized as follows. In Section 4.1, we illustrate the performance of Fourier transform in signal resolution. Then, In Section 4.2, we illustrate the performance of the MLE.

## 4.1 Fourier Transform

In this section, we present three examples to illustrate the performance of the Fourier transform. It holds for all three examples that the input SNR of the strong clutters, $|A|^2/\sigma^2$, is 25.4 dB and the input SNR of the weak echo, $|a|^2/\sigma^2$, is 5.4 dB.

In the first example, three frequencies are far apart beyond the classical frequency resolution limit 10Hz:

Amplitude of Clutter One A=$100e^{j\pi/4}$, Frequency of Clutter One $\phi_1$=32Hz;

Amplitude of Clutter Two B=$100e^{j\pi/3}$, Frequency of Clutter Two $\phi_2$=64Hz;

Amplitude of Weak Echo a=$e^{j\pi/6}$, Doppler of Weak Echo $\phi$=96Hz.

The observation data x(n) are shown in Figure 4.1 while the Fourier transform F($\phi$) =$|X(\omega)|$ is shown in Figure 4.2. There are two peaks that are far apart and correspond to strong clutters, but the echo is too weak to be observed in F($\phi$) in Figure 4.2.

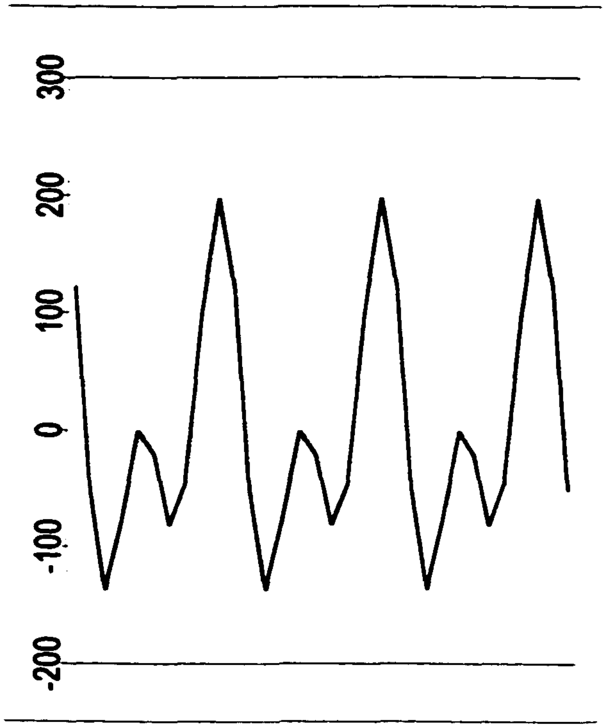


Figure 4.1 Example one ( Observation vs Time)

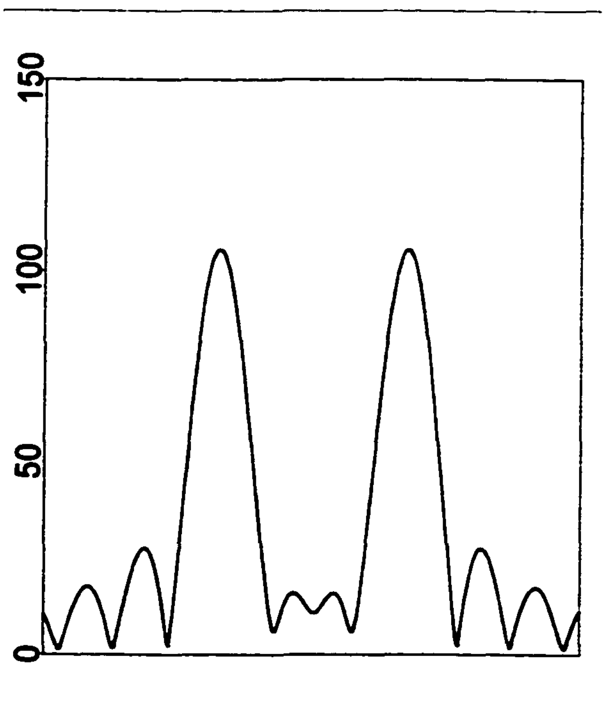


Figure 4.2 Example one ( Fourier Transform vs Frequency)

In the second example, the three frequencies are closer to each other than they are in the first example, but they are still beyond the classical frequency resolution limit:

Amplitude of Clutter One A=$100e^{j\pi/4}$, Frequency of Clutter One $\phi_1$=32Hz;

Amplitude of Clutter Two B=$100e^{j\pi/3}$, Frequency of Clutter Two $\phi_2$=48Hz;

Amplitude of Weak Echo a=$e^{j\pi/6}$, Doppler of Weak Echo $\phi$=64Hz.

The observation data x(n) are shown in Figure 4.3 while the Fourier transform $F(\phi) = |X(\omega)|$ is shown in Figure 4.4. There are still two peaks that corresponds to the two strong clutters even though they are partly overlapped. The echo cannot be seen here just as in the first example in $F(\phi)$ in Figure 4.4.

In the third example, two strong clutters are much closer to each other than they are in the first two examples such that they fall into the classical frequency resolution limit 10Hz.

Amplitude of Clutter One A=$100e^{j\pi/4}$, Frequency of Clutter One $\phi_1$=32Hz;

Amplitude of Clutter Two B=$100e^{j\pi/3}$, Frequency of Clutter Two $\phi_2$=35.2Hz;

Amplitude of Weak Echo a=$e^{j\pi/6}$, Doppler of Weak Echo $\phi$=64Hz.

The observation data x(n) are shown in Figure 4.5 while the Fourier transform $F(\phi) = |X(\omega)|$ is shown in Figure 4.6. At this time, we can merely see one peak in $F(\phi)$ in Figure 4.6.

Thus, the Fourier transform works only on signal resolution if the signals are strong and the frequencies of the signals are far apart beyond the traditional frequency

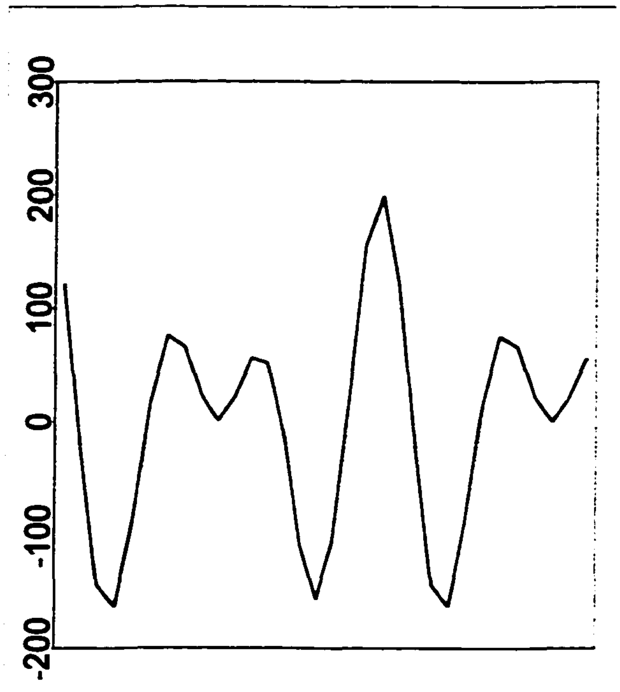


Figure 4.3 Example two ( Observation vs Time)

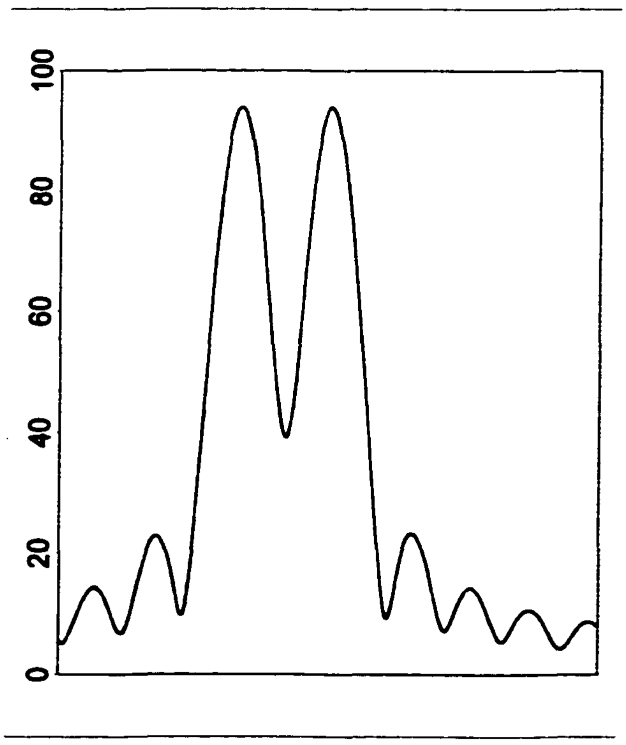


Figure 4.4 Example two ( Fourier Transform vs Frequency)

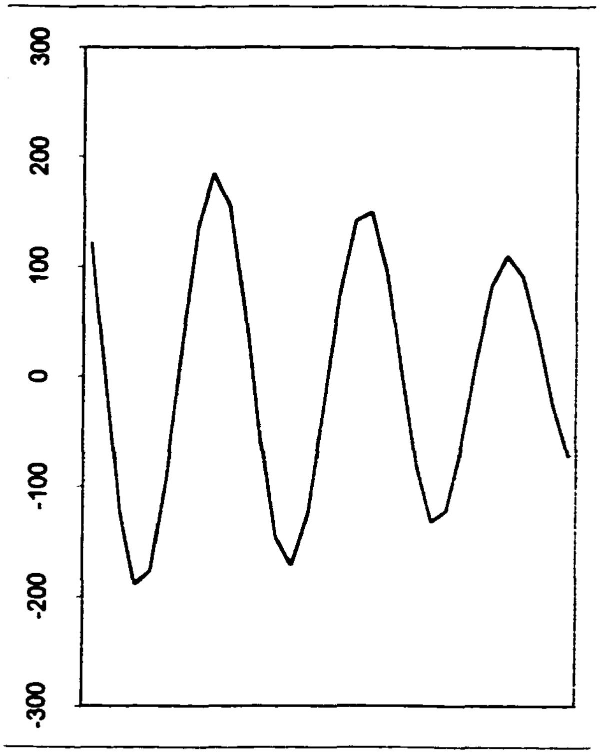


Figure 4.5 Example three ( Observation vs Time)

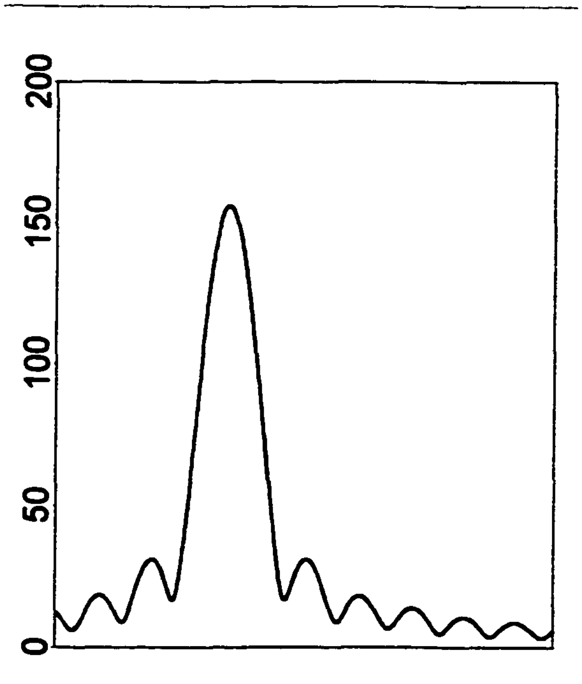


Figure 4.6 Example three ( Fourier Transform vs Frequency)

resolution limit. Any weak signals cannot be detected since they are concealed by even the side lobes of strong signals.

## 4.2 Maximum Likelihood Estimator

However, the MLE can be employed to detect adaptively all signals if the SNRs are reasonably given. The MLE of $\phi_1$ and $\phi_2$ can be determined easily. The MLE of $\phi$ and a is given by the following model

$$\mathbf{x} = A\ \mathbf{s}(\hat{\phi}_1) + B\ \mathbf{s}(\hat{\phi}_2) + a\,\mathbf{s}(\phi) + \mathbf{w} \tag{4.3}$$

Break $\mathbf{s}(\phi)$ into two parts:

$$\mathbf{s}(\phi) = \mathbf{s}_p(\phi) + \mathbf{s}_q(\phi) \tag{4.4}$$

where $\mathbf{s}_q(\phi)$ is orthogonal to the space spanned by $\mathbf{s}(\hat{\phi}_1)$ and $\mathbf{s}(\hat{\phi}_2)$ while $\mathbf{s}_p(\phi)$ belongs to the space. Thus, the equation (4.3) can be rewritten as

$$\mathbf{x} = \tilde{A}\ \mathbf{s}(\hat{\phi}_1) + \tilde{B}\ \mathbf{s}(\hat{\phi}_2) + a\,\mathbf{s}_q(\phi) + \mathbf{w} \tag{4.5}$$

The MLE of $\phi$ and a can be determined as

$$\hat{\phi} = \arg\max_{\phi} \frac{|\mathbf{s}_q^H(\phi)\mathbf{x}|}{|\mathbf{s}_q(\phi)|} \equiv \arg\max_{\phi} c(\phi) \tag{4.6}$$

and

$$\hat{a} = \frac{\mathbf{s}_q^H(\hat{\phi})\mathbf{x}}{|\mathbf{s}_q(\hat{\phi})|^2} \tag{4.7}$$

The optimal CFAR detection of the weak echo is given by

$$|\hat{a}|/\sigma_a \quad > \text{Threshold} \tag{4.8}$$

where $\sigma_a$ is the standard deviation of $\hat{a}$ given in the CRB, and $|\hat{a}| = c(\hat{\phi})/|s_q(\hat{\phi})|$.

In this section, we present three examples to illustrate the performance of the MLE of the amplitudes.

The fourth example is exactly the same as the third example in Section 4.1. There is a dominant maximum in Figure 4.6, so we assume here that there was only one signal in the observation data and that the frequency of this assumed signal should be $\hat{\phi}$=33.6. The white noise should be left only after we cancel the interference caused by the assumed signal with frequency $\hat{\phi}$=33.6Hz if there was really one signal. We cancel the interference caused by the assumed signal, then obtain Figure 4.7. There are two dominant maxima that are symmetrically located on $\hat{\phi}$=33.6Hz in Figure 4.7 and significantly over the threshold in (4.8), which means the above assumption that there was only one signal in the data is not correct; that is, there are at least two signals contained in the observation data. We again assume here that there were two signals in the observation data, and that the frequencies of these two assumed signals should be $\hat{\phi}_1$=33.216Hz and $\hat{\phi}_2$=33.984. Starting with these two initial points, we try to find the stable points of the maximum likelihood function $c(\phi)$. Our algorithm is to cancel the interference caused by either of the initial points, say, $\hat{\phi}_2$=33.984, then find the point which maximizes $c(\phi)$ as a new initial point. We cancel the interference caused by $\hat{\phi}_2$=33.984Hz, then obtain Figure 4.8. There is one dominant maximum in Figure 4.8; thus the new initial point is chosen as $\hat{\phi}_1$=33.544Hz. Again, we cancel the interference caused by this new initial point $\hat{\phi}_1$=33.544Hz, then obtain Figure 4.9. In Figure 4.9

there is one dominant maximum, so another initial point is found as $\hat{\phi}_2$=35.232Hz. We continue this way in Figure 4.10 and Figure 4.11 until we find the stable points, $\hat{\phi}_1$=32.064Hz and $\hat{\phi}_2$=35.232Hz. Again, there should be only the white noise left after we cancelled the interference caused by the assumed signals with frequencies, $\hat{\phi}_1$=32.064Hz and $\hat{\phi}_2$=35.232Hz, if they were the only signals. We cancel the interference caused by $\hat{\phi}_1$=32.064Hz and $\hat{\phi}_2$=35.232Hz, then obtain Figure 4.12. There is one dominant maximum coming up in Figure 4.12 even though it is very weak, but it is over the threshold in (4.8), so at least there were three signals in the observation data. We assume the frequency of the third signal as $\phi$ =63.264Hz. We treat these three detected signals again as initial points, do the same convergent procedure described above, then obtain the stable points $\hat{\phi}_1$=31.968Hz, $\hat{\phi}_2$=35.232Hz, and $\phi$ =64.32Hz. We cancel the interference caused by these three signals, then obtain Figure 4.14, which is tested as a white noise. Finally, we detect all three signals although the two strong signals are very close and the echo signal is very weak. The amplitudes are estimated as follows:

$$A = 99.4412e^{j0.256078\pi},$$

$$B = 99.4954e^{j0.327229\pi},$$

$$a = 0.90636e^{j0.153587\pi}.$$

In the fourth example, as is seen, that the weak echo signal can be detected if it is far away from the other signals beyond the classical frequency limit 10Hz. We will

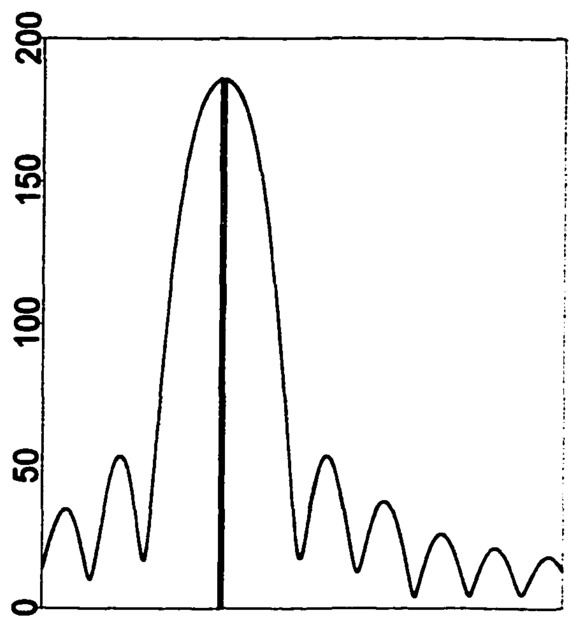


Figure 4.7 Example four (MLE vs Frequency)

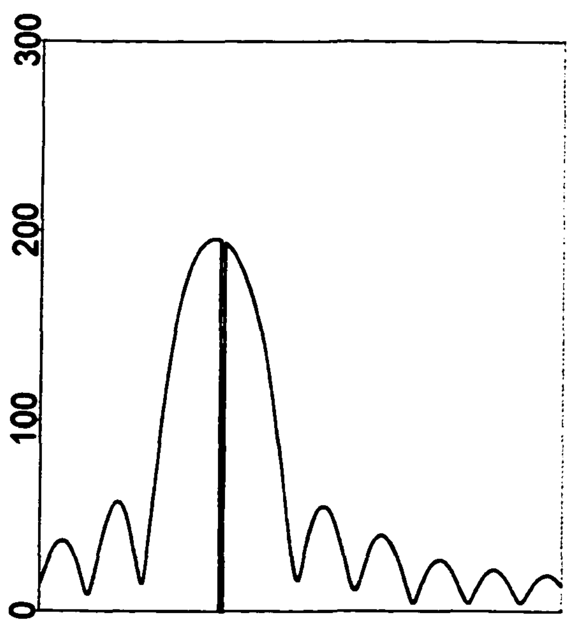


Figure 4.8 Example four

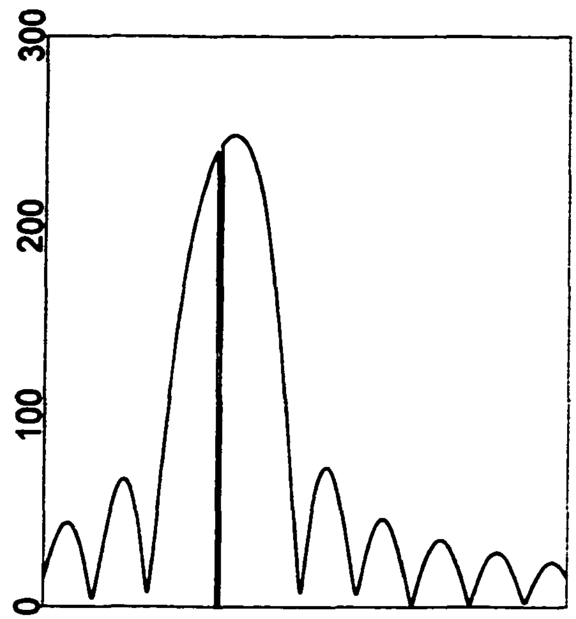


Figure 4.9 Example four

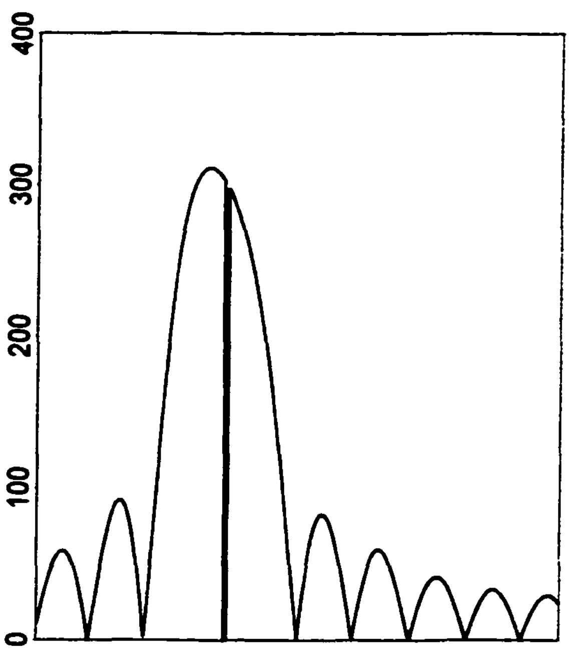


Figure 4.10 Example four

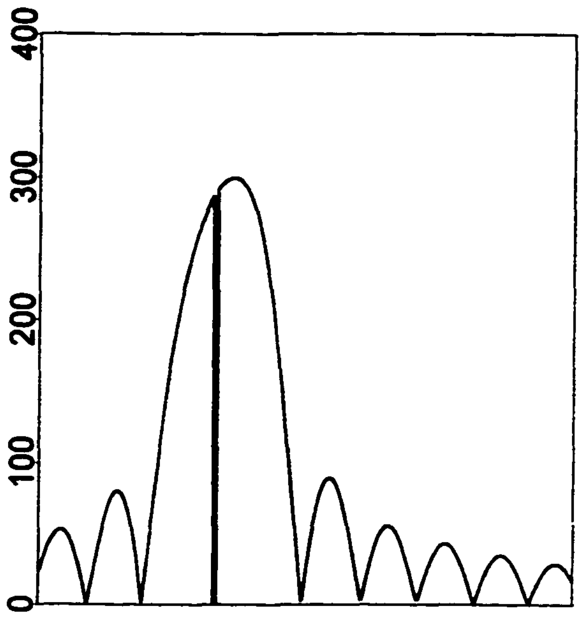


Figure 4.11 Example four

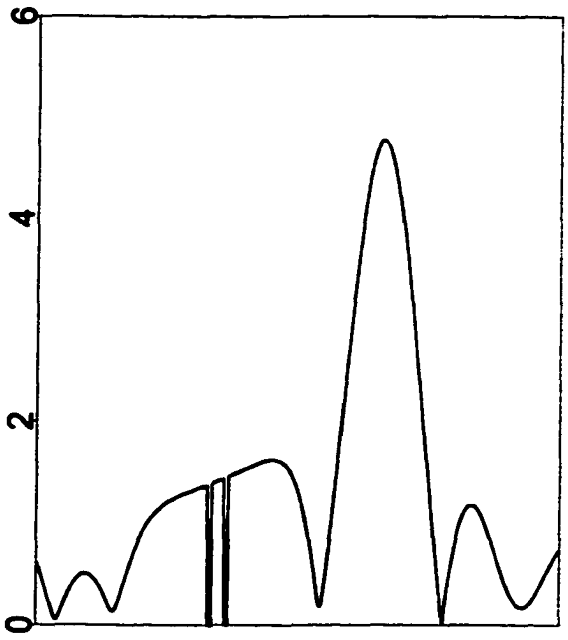


Figure 4.12 Example four

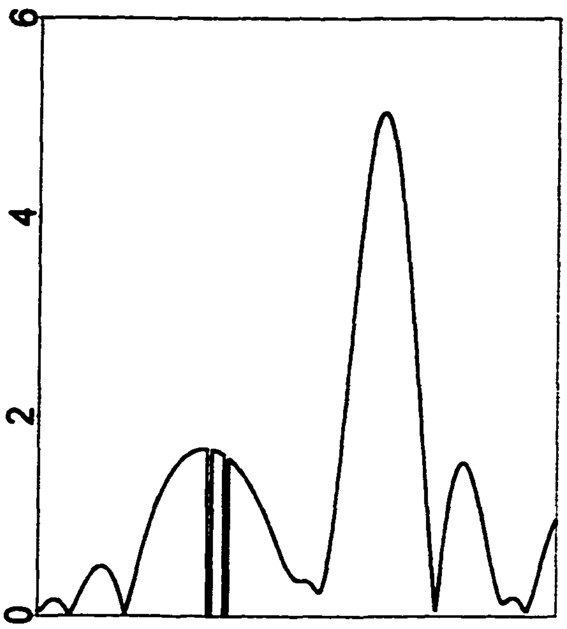


Figure 4.13 Example four

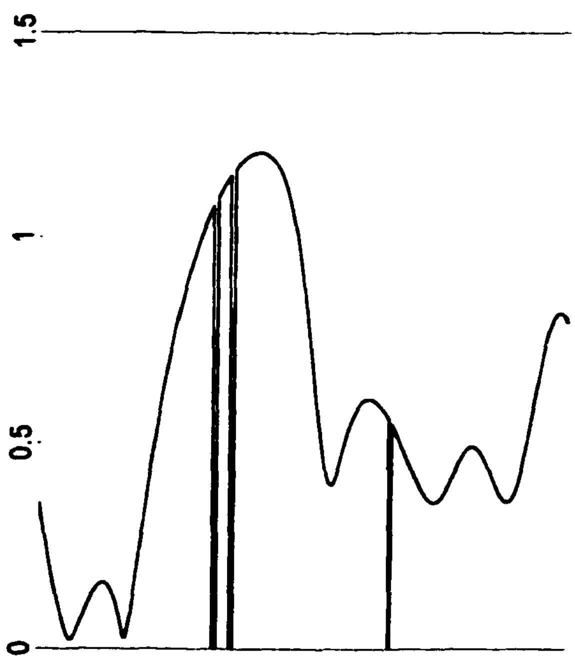


Figure 4.14 Example four

show in the fifth example that the echo signal still can be detected even though it is located in the classical frequency limit.

Amplitude of Clutter One A=$100e^{j\pi/4}$, Frequency of Clutter One $\phi_1$=32Hz;

Amplitude of Clutter Two B=$100e^{j\pi/3}$, Frequency of Clutter Two $\phi_2$=35.2Hz;

Amplitude of Weak Echo a=$e^{j\pi/6}$, Doppler of Weak Echo $\phi$=41.6Hz.

The observation data x(n) and the Fourier transform $F(\phi)=|X(\omega)|$ look almost the same as in Figure 4.5 and Figure 4.6. We do the similar procedures as we did in Example 4 After we cancel the interference caused by the first two signals, we obtain Figure 4.15. There is one dominant maximum that is over the threshold in Figure 4.15, so we find the third signal even though it is concealed by the strong signals. After we cancel the interference caused by the three signals, we obtain Figure 4.16 that is a white noise.

Our sixth example is as follows:

Amplitude of Clutter One A=$100e^{j\pi/4}$, Frequency of Clutter One $\phi_1$=12Hz;

Amplitude of Clutter Two B=$60e^{j\pi/3}$, Frequency of Clutter Two $\phi_2$=16Hz;

Amplitude of Weak Echo a=$e^{j\pi/6}$, Doppler of Weak Echo $\phi$=20Hz.

The observation data x(n) are shown in Figure 4.17 while the Fourier transform $F(\phi)=|X(\omega)|$ is shown in Figure 4.18. What we can say about $F(\phi)$ in Figure 4.18 is that there is merely one signal. We do the similar procedure as we did before. We first assume there was only signal in the observation data; the frequency of the assumed signal should be $\hat{\phi}$=12.96. We cancel the interference caused by the assumed signal,

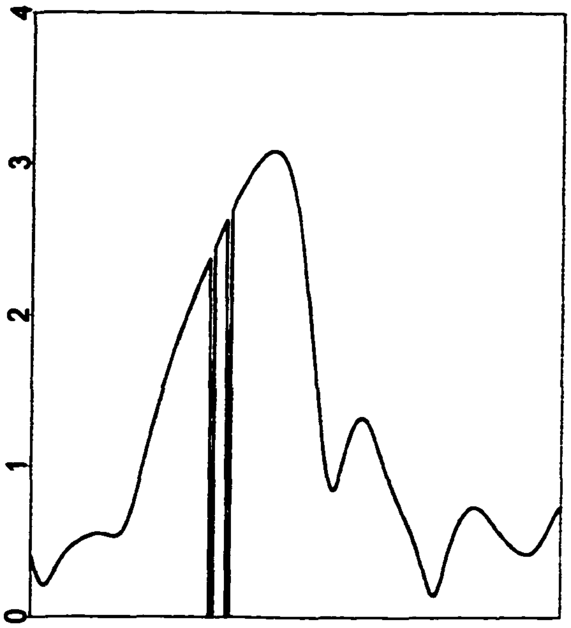


Figure 4.15 Example five

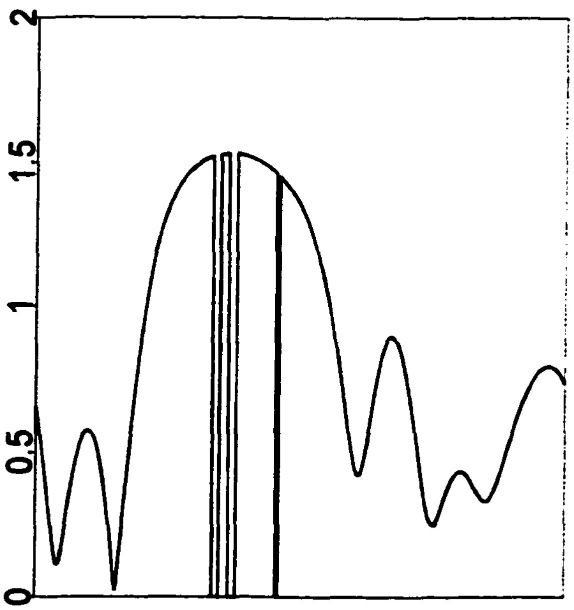


Figure 4.16 Example five

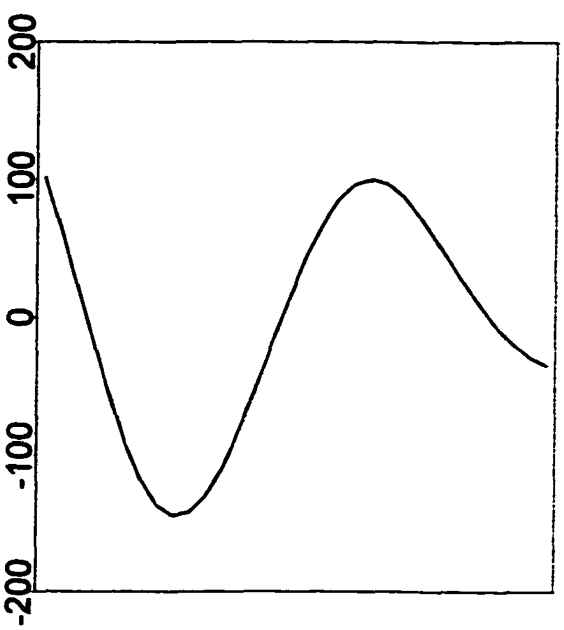


Figure 4.17 Example six (Observation vs Time)

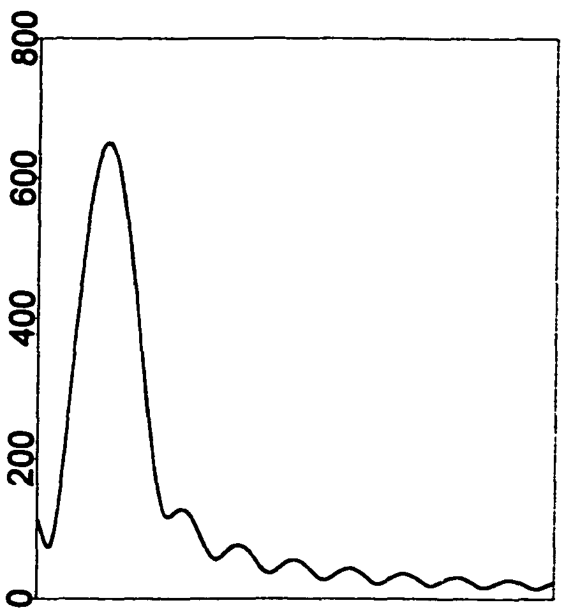


Figure 4.18 Example six (Fourier Transform vs Frequency)

then obtain Figure 4.19. There is one dominant maximum in Figure 4.19 that is significantly over the threshold, which implies that the above assumption is not correct. We again assume here that there were two signals in the observation data and that the frequencies of these two assumed signals can be obtained by canceling the interference caused by either of these two assumed signals. Figure 4.20, Figure 4.21, and Figure 4.22 describe the above-mentioned process, and, finally, $\hat{\phi}_1$=11.904Hz and $\hat{\phi}_2$=16.032Hz. We again cancel the interference caused by the two assumed signals, then obtain Figure 4.23. There is one dominant maximum in Figure 4.23 that is over the threshold, so we detect the third signal and the frequency of this signal is $\phi$=21.504Hz. We start with these three frequencies as initial points; at each time we fix two points and cancel the interference caused by them, then obtain a new point. We proceed this way until we find the stable points, which are $\hat{\phi}_1$=12.000Hz, $\hat{\phi}_2$=16.032Hz, and $\hat{\phi}$=22.752Hz. Figure 4.23, Figure 4.24, Figure 4.25, Figure 4.26, and Figure 4.27 describe the above-mentioned process. After we cancel the interference caused by the three signals, we obtain Figure 4.28 that is a white noise. The amplitudes are estimated as follows:

$$A = 99.9905e^{j0.250886\pi},$$

$$B = 59.9451e^{j0.330967\pi},$$

$$a = 0.798866e^{j0.115677\pi}.$$

Thus, the MLE of the amplitudes detects not only the strong signals that are far apart beyond the traditional frequency resolution limit, but also those that fall into the limit. Furthermore, the MLE can also detect the weak signals that are covered up by the side lobes of the strong signals.

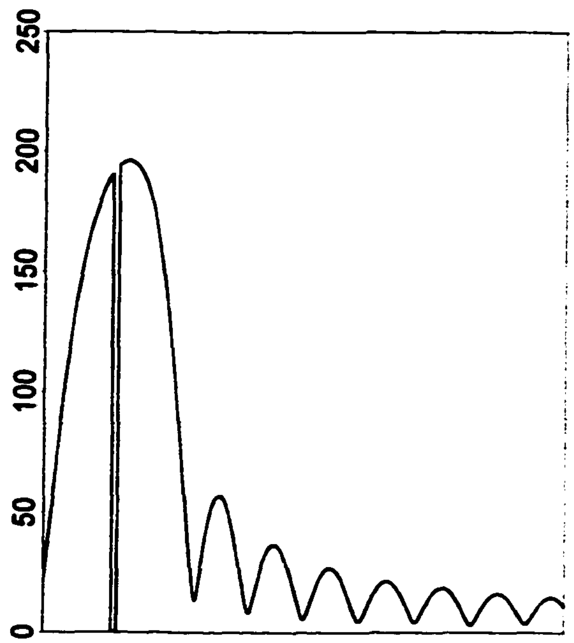


Figure 4.19 Example six

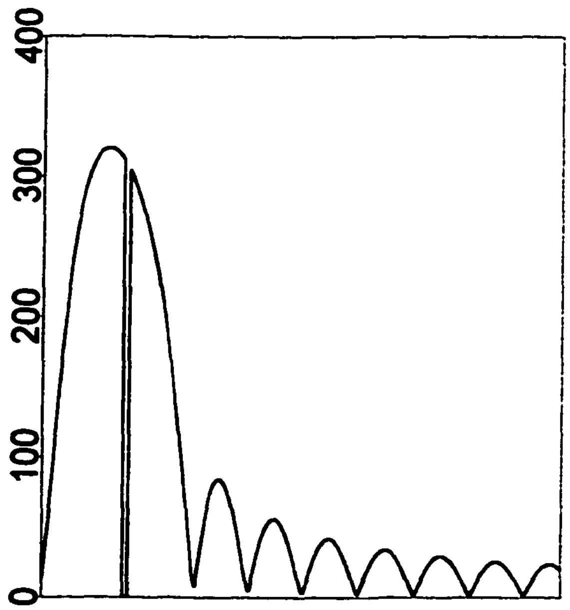


Figure 4.20 Example six

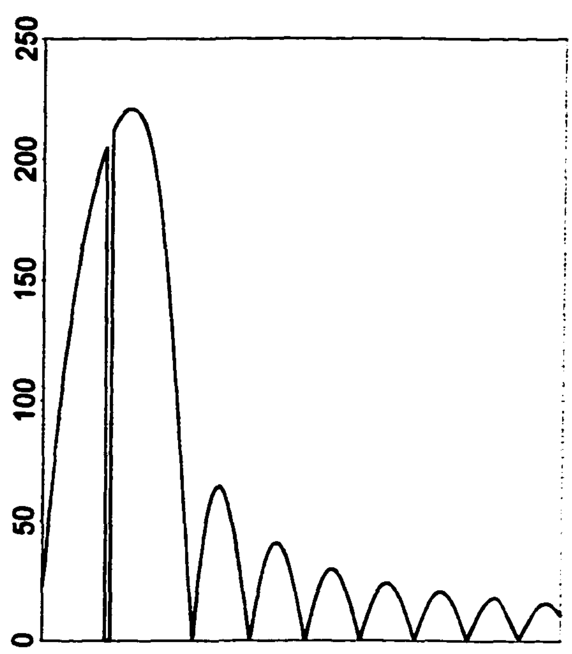


Figure 4.21 Example six

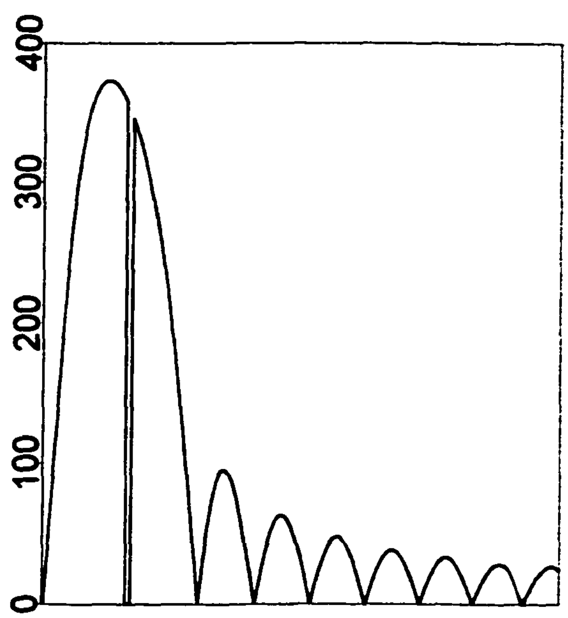


Figure 4.22 Example six

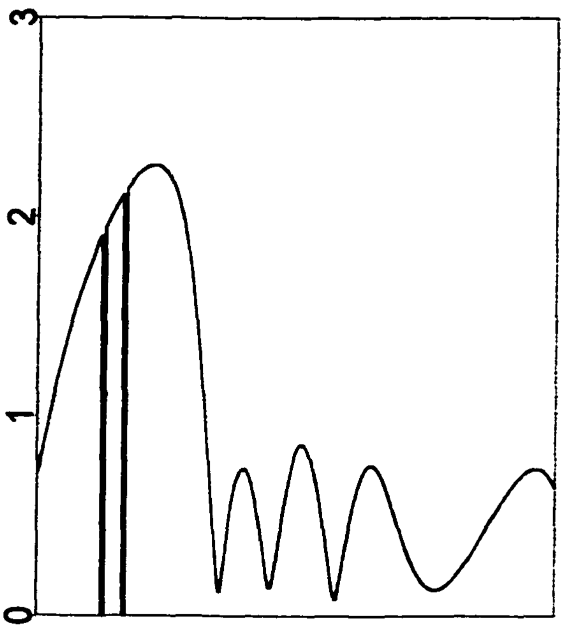


Figure 4.23 Example six

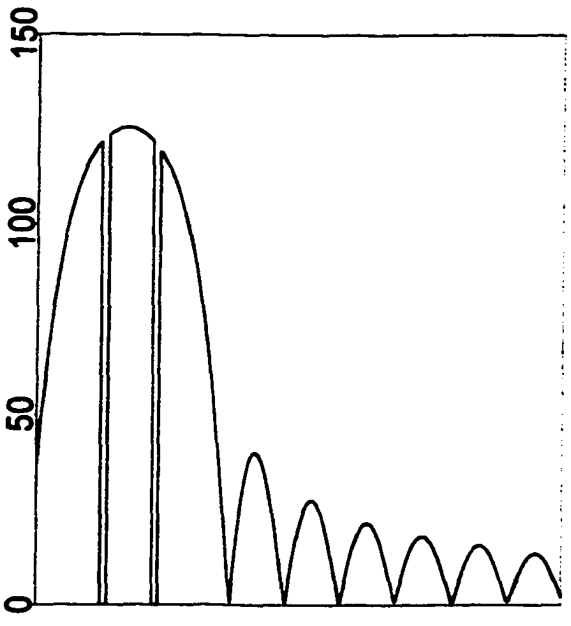


Figure 4.24 Example six

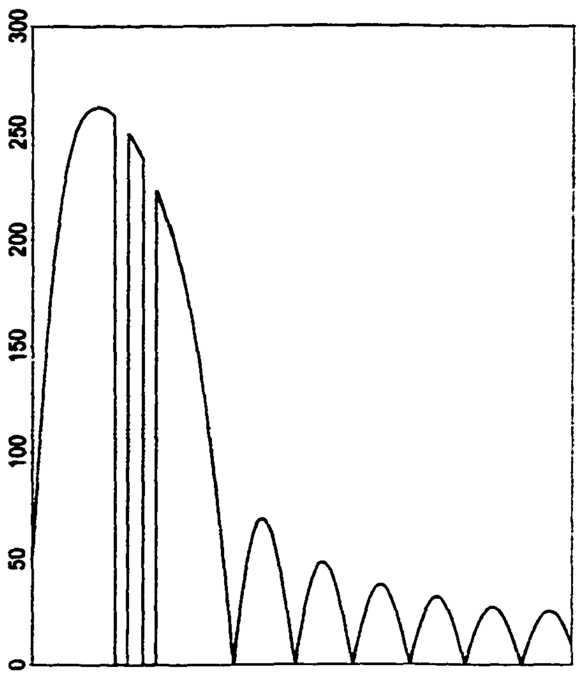


Figure 4.25 Example six

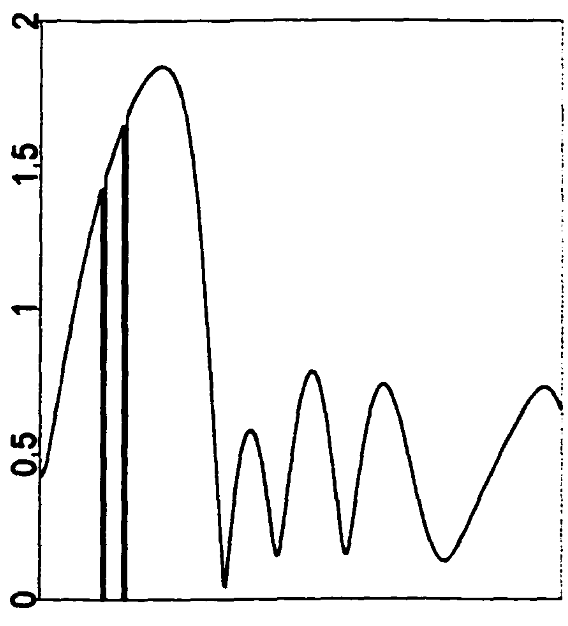


Figure 4.26 Exmaple six

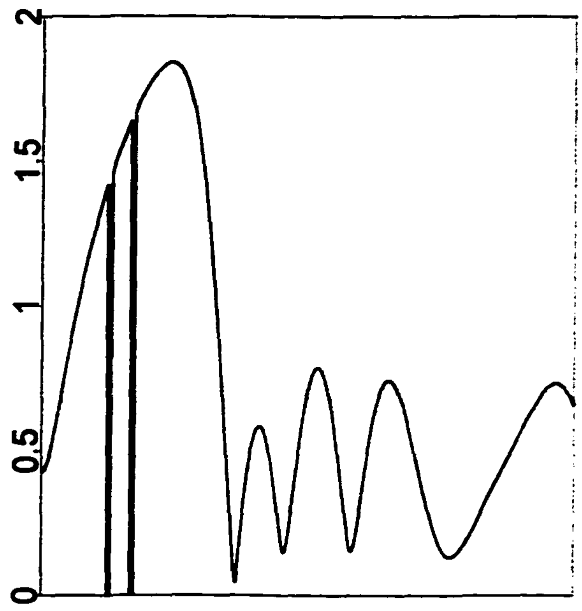


Figure 4.27 Example six

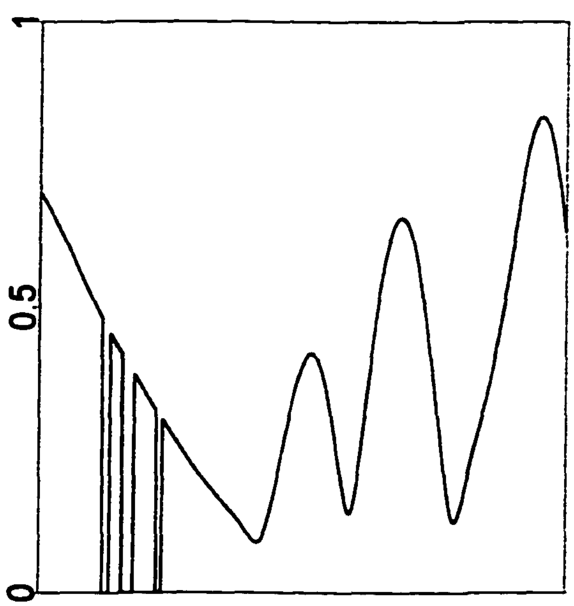


Figure 4.28 Example six

# CHAPTER 5

# CONCLUDING REMARKS

The CRB of estimating the amplitudes of superimposed sinusoidal signals with unknown frequencies in the case of white Gaussian noise and the asymptotic efficiency of the MLE in this case at high SNR but finite sample have been established. Consequently, the optimal resolution of $K$ sinusoidal signals with unknown frequencies in a finite window is given by the MLE of their amplitudes.

It is well known that the detection of a signal from colored interference and noise can also be given by a filter with the optimal weights (Brennan, L. and Reed, I., 1973) as follows:

$$\mathbf{h} = [\Sigma^{-1}\mathbf{s}]^{*}$$

where $\mathbf{s}$ is the signal to be detected and $\Sigma$ is the covariance matrix of the interference and noise. In practice, $\Sigma$ is unknown and needs to be estimated using extra data in another dimension which is supposed to be statistically stationary or homogeneous. Unfortunately, the extra data may not be available because of a lack of homogeneity. Furthermore, the inversion of $\Sigma$ is also an obstinate problem if the order of $\Sigma$ is large. However, the colored interference can be approximated by a superposition of sinusoidal signals with unknown frequencies and amplitudes because of the spectrum theorem of a stationary random process (Yaglom, 1972). Practically, the superimposed model (2.21) can be applied to the detection and estimation of signals in colored observation noise. In this approach, the unknown frequencies and amplitudes in the colored noise can be estimated without resort to a covariance matrix or extra data in another dimension.

# REFERENCES


Akaike, H. (1974) "A New Look at the Statistical Model Identification," IEEE Trans. Automat. Contr., Vol. AC-19, pp. 716-723.

Akhiezer, N. I. And Glazman, I. M. (1961) "Theory of Linear Operators in Hilbert Space," Frederick Ungar Publishing Co.

Applebaum, S. P. (1976) "Adaptive Arrays," IEEE Trans. AP-24, pp. 585-598.

Box, G. E. P. and Jenkins, G. M. (1970) "Time Series Analysis, Forcasting and Control," Holden –Day, San Francisco.

Brennan, L. and Reed, I. (1973) "Theory of Adaptive Radar," IEEE Trans. on AES, Vol. 9, pp. 237-251.

Bretthorst, G. (1988) "Bayesian Spectrum Analysis and Paramter Estimation," Spring -Verlag, Berlin.

Brillinger, D. R. (1981) "Time Series: Data Analysis and Theory," Holden-Day, Inc.

Burg, J. (1978) "Maximum Entropy Power Spectral Analysis," Modern Spectrum Analysis, Ed. By Childers, D., IEEE PRESS, New York, pp. 34-41.

Capon, J. (1969) "High Resolution Frequency Wave Number Spectrum Analysis," Proc. IEEE, Vol. 57, pp. 1408-1418.

Davenport, W. B. and Root, W. L. (1958) "Random Signals and Noise," McGraw-Hill, New York.

Doob, J. L. (1953) "Stochastic Processes," Wiley, New York.

Fante, R. (1999) "Adaptive Nulling of SAR Sidelobe Discretes," IEEE Trans. on AES, Vol. 35, No. 4, pp 1212-1218.

Frost, O. L. (1972) "An Algorithm for Linearly Constrained Adaptive Array Processing," Proc. IEEE, Vol. 60, pp. 926-935.

Gihman, I. I. And Skorohod, A.V. (1974) "The Theory of Stochastic Processes," Spring-Verlag, Berlin.

Gu, H. (1996) "Ambiguity Function and Cramer-Rao Bound in the Multi-signal Case," IEE Proceedings on Radar, Sonar and Navigation, Vol. 143, pp. 227-231.

Gu, H. and Gao, R. (1997) "Resolution of Overlapping Echoes and Constrained Matched Filters," IEEE Trans. SP, Vol.45, No. 7, pp. 1854-1857.

Gu, H. (1998) "Estimating the Number of Signals and Signal Resolution," IEEE Trans. on Signal Processing, Vol. 46, No. 8, pp. 2267-2270.

Gu, H. (2000) "Linearization Method for Finding Cramer-Rao Bounds in Signal Processing," IEEE Trans. on SP, Vol. 48, No. 2, pp. 543-545.

Kaveh, M. and Barabell, A. J. (1986) "The Statistical Performance of the MUSIC and the Minimum-Norm Algorithm in Resolving Plane Waves in Noise," IEEE Trans. ASSP, Vol., ASSP-34, No. 2, pp. 331-341.

Kay, S. (1988) "Modern Spectral Estimation," Prentice-Hall, Englewood Cliffs, N. J.

Kelly, E. J., Reed, I. S., and Root, W. L. (1960) "The Detection of Radar Echoes in Noise, I and II," J. SIAM, 8, pp. 309-341, 481-507.

Lehman, E.(1959) "Testing Statistical Hypotheses," Wiley, New York.

Lecam, L. M. (1973) "On Some Asymptotic Properties of Maximum Likelihood Estimates and Related Bayes Estimates," University of California publication in Statistics, 1, pp. 277-328.

Ligget, W. S. (1973) "Passive Sonar: Fitting Models to Multiple Time Series," in NATO ASI on signal processing, (J. W. R. Griffiths et al. Eds.), New York: Academic Press, pp. 327-345.

Lindgren, B.W. (1976) "Statistical Theory," Macmillan Publishing Co., Inc.

Marple, L. (1987) "Digital Spectral Analysis, Prentice-Hall," Englewood Cliffs, N. J.

Murdoch, D. (1970) "Linear Algebra," Wiley, New York.

Pisarenko, V.F. (1973) "The Retrieval of Harmonics from Covariance Functions," Geophys. J. Royal Astronomical Soc., Vol. pp. 347-366.

Rao, C. R., Zhang, L., and Zhao, L. C. (1993) "Multiple Target Angle Tracking Using Sensor Array Outputs," IEEE Trans. AES-29, pp. 268-271.

Scharf, L. and Friedlander, B (1994) "Matched Subspace Detector," IEEE Trans. On SP Vol 42, pp. 2146-2157.

Schmidt, R. O. (1979) "Multiple Emitter Location and Signal Parameter Estimation," Proc. RADC spectrum estimation workshop, (Griffis AFB, N.Y), pp. 243-258.

Schmidt, R. O. (1981) "A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation," Ph.D. dissertation, Stanford University, CA.

Stoica, P. and Nehorai, A. (1989) "MUSIC, Maximum Likelihood, and Cramer-Rao Bound," IEEE Trans., ASSP-37, No. 5, pp. 720-743.

Stoica, P., Jakobsson, A. and Li, J. (1997) "Cisoid Parameter Estimation in the Colored Noise Case: Asymptotic Cramer-Rao Bound, Maximum Likelihood, and Nonlinear Least Squares," IEEE Trans. SP-45, pp. 2048-2059.

Stoica, P., Li, H. and Li, J. (2000) "Amplitude Estimation of Sinusoidal Signals: Survey, New Results, and an Application," IEEE Trans. on SP Vol. 48, No. 2, pp.338-352.

Stoica, P., Moses. R., Friedlander, B., and Soderstrom, T. (1989) "Maximum Likelihood Estimation of the Parameters of Multiple Sinusoids from Noisy Measurements," IEEE Trans., ASSP-37, pp. 378-391.

Tufts, W. and Kumaresan, R. (1982) "Frequency Estimation of Multiple Sinusoids: Making Linear Prediction Perform Like Maximum Likelihood," Proc IEEE, 70, pp. 975-989.

Wald, A. (1943) "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations Is Large," Trans. Amer. Math. Soc., Vol. 54, pp. 426-482.

Wax, T. and Kailath, T. (1985) "Detection of Signals by Information Theoretic Criteria," IEEE Trans on ASSP, Vol. ASSP-33, No. 2.

Wax, M. (1985) "Detection and Estimation of Superimposed Signals," PhD. Dissertation, Stanford University, CA.

Widrow, B. "Adaptive Filters," in "Aspects of Network and System Theory" Ed. By R. Kalmam, Holt, Rinehart & Winston, New York, NY, 1070, pp. 563-587.

Widrow, P. E., Duvall, K. M., Gooch, R. P., and Newman, W. C. (1982) "Signal Cancellation Phenomena in Adaptive Antenna: Causes and Cures," IEEE Trans. Antennas Propagation, Vol. AP-30, pp. 469-478.

Woodward, P. M. (1953) "Probability and Information Theory, with Applications to Radar," London, U.K.: Pergamon.

Xie, Z. (1986) "Time Series Analysis," Peking University Press, Beijing.

Yaglom, A. (1972) "An Introduction to the Theory of Stationary Random Functions," New York.

Ziskind, I., and Wax, M. (1988) "Maximum Likelihood Localization of Multiple Sources by Alternating Projection," IEEE Trans., ASSP-36, 10, pp. 1553-1560.

## VITA

Shaohui Jia, the son of Juhua Zhang and Guoming Jia, was born on June 3, 1966, in Shanxi, China. He is married to Yanting Zhou. He received his Bachelor of Science in Mathematics and his Master of Science in Statistics from Peking University, Beijing, China, in July 1987 and July 1994, respectively. He studied at the Department of Mathematics and Statistics, University of Saskatchewan, Saskatoon, Canada, from September 1996 to August 1997. Then he transferred to Louisiana Tech University and obtained his Master of Science in Mathematics in August 1998. At present, he is a candidate for the Doctor of Philosophy degree in the Applied Computational Analysis and Modeling Program (ACAM) at Louisiana Tech University, and his advisor is Dr. Huaijin Gu.