TRACKING OF MULTIPLE MOVING TARGETS BY PASSIVE ARRAYS AND ASYMPTOTIC PERFORMANCE ANALYSIS

By

YIFENG ZHOU, B. Eng. (Southeast University)
M. Sc. (Shanghai Acoustics Laboratory, Academia Sinica)

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree
Ph. D

McMaster University
March 1995

TRACKING OF MULTIPLE MOVING TARGETS BY PASSIVE ARRAYS AND ASYMPTOTIC PERFORMANCE ANALYSIS

PH. D (1995) MCMASTER UNIVERSITY

(Electrical and Computer Engineering) Hamilton, Ontario

TITLE: **Tracking of Multiple Moving Targets by Passive Arrays and Asymptotic Performance Analysis**

AUTHOR: Yifeng Zhou

B. Eng. (Southeast University)

M. Sc. (Shanghai Acoustics Laboratory, Academia Sinica)

SUPERVISOR(S): Dr. Patrick C. Yip

Professor, Department of Mathematics and Statistics

and Communications Research Laboratory

NUMBER OF PAGES: xvi, 146

TO MY PARENTS

ABSTRACT

This thesis is focused on the topic of tracking multiple moving targets by combining the spatial and temporal information obtained by a passive array. For stationary sources a unified constrained subspace fitting approach for estimating the DOA's of spatially close source signals is presented. The algorithm is based on the Karhunen-Loève expansion of the covariance matrix of the array manifold in a sector of interest and searches for an optimal signal subspace over the array manifold space, which has minimum principal angles with the data signal subspace generated from the array data. This method is shown to be asymptotically consistent. Although this algorithm involves only one-dimensional researches, its performance is comparable to those in which multi-dimensional optimization is used.

We propose a maximum likelihood approach for tracking moving targets by passive arrays. A locally linear model is used for the source target motion dynamics, and the target state is shown to be strongly observable. An MTS (multiple target state) vector is defined to describe the source target state. The maximum likelihood estimator is based on a batch of array data. The initial MTS is estimated as the maximizing point of the likelihood function of a batch of array data, and the subsequent MTS vectors are predicted by the target dynamics. Since the association problem is embedded in the estimation problem, the natural ordering of the target state is kept as long as the initial target DOA's can be successfully resolved by the array. To cope with difficulties involved in the nonlinear optimization process a modified Gauss-Newton algorithm is proposed in which the Hessian is approximated by a positive semi-definite matrix to guarantee that the algorithm is descent. The asymptotic performance analysis has also been fully investigated. We show that the ML estimates of the MTS variables are asymptotically consistent. We also derive explicit formulas

for the asymptotic covariance and the Cramér-Rao bounds for the MTS estimates, and it is found that the ML estimator is relatively efficient. To show the effectiveness of the ML tracking technique we compare the asymptotic performance of the ML estimator with that of the extended Kalman filter (EKF). Its performance is superior to that of the EKF. Numerical results are provided to demonstrate the performance of the ML tracking technique via computer simulations.

Acknowledgement

I wish to express my deep gratitude to Dr. P. Yip, for his constant encouragement, continued assistance and expert guidance and supervision throughout the course of this work; he is not only my academic supervisor but also the mentor of mine. Special thanks are also due to the Drs. J. Reilly, K. M. Wong, and T. Luo for being supportive as supervisory committee members.

I would like to extend my sincere thanks to Dr. Q. T. Zhang, Ryerson Technical University and Dr. H. Leung, Defense Research Establishment, Ottawa, for their continued interests and helpful suggestions. The author is also grateful to Dr. Q. Jin and Ms. J. Li for inspiring discussions with them.

I also like to acknowledge the financial support provided by Telecommunications Research Institute of Ontario, National Science and Engineering Research Council and the Department of Electrical and Computer Engineering, McMaster University.

Finally, I am indebted to my loving wife, Zhaoyang, for her constant understanding, encouragement and support.

Contents

List of Tables

List of Figures

Glossary

Throughout this thesis, we will use the following notation :

$\underline{x}(i)$	array data vector.
$\underline{s}(i)$	signal data vector.
$\underline{w}(i)$	sensor noise vector.
$\Delta\omega_{sc}$	half the source signal frequency bandwidth.
ω_0	central frequency of the source signal.
$\underline{z}_m$	the mth sensor position vector.
$\underline{\zeta}_k$	the kth near-field source position vector.
$\underline{\kappa}_k$	the kth far-field source directional vector.
τ_{km}	time delay on the mth sensor induced by the mth far-field source target.
τ_{max}	maximum time delay among τ_{km}, for $k = 1, 2, \ldots, K$ and $m = 1, 2, \ldots, M$.
τ^s_{km}	time delay on the mth sensor induced by the mth near-field source target.
θ_k	azimuth of the kth source target.
$\theta_k(i)$	azimuth of the kth moving source target at i.
Θ	source DOA parameter set.
$\Theta(i)$	source DOA parameter set of the moving source targets at i.
$\underline{a}(\theta_k)$	array steering vector associated with the kth source signal.
$A(\Theta)$	array composite steering matrix.
$\underline{a}[\theta_k(i)]$	array steering vector associated with the kth moving source signal at i.

$A[\Theta(i)]$	Also denoted as $A(i)$; array composite steering matrix at i for moving source targets.
ν_k	constant speed of the kth source moving target.
$\tau_k(i)$	range of the kth moving target at i.
$\phi_k(i)$	heading direction of the kth moving target.
$\underline{\alpha}_k(i)$	$\alpha_k(i) = [\theta_k(i), q_k(i), \phi_k]^T$, where $q_k(i) = \nu_k/\tau_k(i)$; denote the kth target state vector at i.
$\underline{\alpha}(i)$	$\underline{\alpha}(i) = [\underline{\alpha}_1^T(i), \underline{\alpha}_2^T(i), \ldots, \underline{\alpha}_K^T(i)]^T$ is the multiple target state (MTS) vector at i.
$\underline{F}$	denote the locally linear source motion dynamics.
$\underline{v}_k(i)$	gradient of $\theta_k(i)$ with respect to $\underline{\alpha}(1)$.
$\underline{v}_k(i,j)$	gradient of $\theta_k(i)$ with respect to $\underline{\alpha}(j)$.
$V(i)$	$V(i) = [\underline{v}_1(i), \underline{v}_2(i), \ldots, \underline{v}_K(i)]$.
$V(i,j)$	$V(i,j) = [\underline{v}_1(i,j), \underline{v}_2(i,j), \ldots, \underline{v}_K(i,j)]$.
$P_{A(\Theta)}$	orthogonal projection matrix onto the column space of $A(\Theta)$.
$P^{\perp}_{A(\Theta)}$	$P^{\perp}_{A(\Theta)} = I - P_{A(\Theta)}$; orthogonal projection matrix onto the null space of $A(\Theta)$.
$P_{A(i)}$	Also denoted as P_i; orthogonal projection matrix onto the column space of $A(i)$.
$P^{\perp}_{A(i)}$	Also denoted as $P^{\perp}_i$; orthogonal projection matrix onto the null space of $A(i)$.
R_x	array data covariance matrix.
R_s	source signal covariance matrix.
R_w	array sensor noise covariance matrix.
λ_k	the kth eigen-value of R_x.
$\underline{u}_k$	the kth eigen-vector of R_x.
U_S	$U_S = [\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_r]$, where r is the number of signal eigen-values.
U_N	$U_N = [\underline{u}_{r+1}, \underline{u}_{r+2}, \ldots, \underline{u}_M]$.
E_S	signal subspace or the range space of U_S.
P_{E_S}	orthogonal projection onto E_S.
E_N	noise subspace or the range space of U_N.
P_{E_N}	orthogonal projection matrix onto E_N.

A_M	array manifold in a sector of interest.
S_M	array manifold space.
P_{S_M}	orthogonal projection matrix onto the array manifold space S_M.
S_X	the optimum signal subspace within the array manifold space.
P_{S_X}	orthogonal projection matrix onto the subspace S_X.
σ_s^2	source signal variance.
σ_w^2	array sensor noise variance.
$\hat{A}$	denote an estimate of A.
$E[\cdot]$	expectation.
A^H	conjugate transpose of matrix A.
A^T	transpose of matrix A.
$\oplus$	direct sum notation.
$A \cdot B$	inner product of matrices A and B.
$A \odot B$	Hadamard product of matrices A and B.
$A \otimes B$	Kronecker product of matrices A and B.
$\| A \|_F$	Frobenius norm of A.
$tr\{A\}$	trace of matrix A.
$A^\dagger$	pseudo-inverse of matrix A.
$vecA$	vector function, concatenation of the columns of matrix A.
$diag(A)$	diagonal function. if A is a matrix, $diag(A)$ is a column vector formed from the diagonal elements of A. If A is a vector, $diag(A)$ puts A on the main diagonal.
$det(A)$	determinant of matrix A.
I	identity matrix.
$\mathbf{O}$	all one matrix of size 3×3.
∇	gradient operator.
∇^2	Hessain operator.
$Var[\cdot]$	variance of a random variable.
$Cov[\cdot]$	covariance matrix of a random vector.

$\mathbf{CRB}[\cdot]$	Cramér-Rao bound of an estimate.
$\mathbf{I}[\cdot]$	Fisher information matrix corresponding to a variable.
C_{MU}	asymptotic covariance matrix of the MUSIC estimates.
C_{DML}	asymptotic covariance matrix of the DML estimates.
C_{opt}	asymptotic covariance matrix of the WSF estimates.
CRB_{DET}	deterministic Cramér-Rao bound.
CRB_{STO}	stochastic Cramér-Rao bound.
l_{DML}	DML criterion function.
l_{SML}	SML criterion function.
l_{MD}	criterion function of the MD-MUSIC algorithm.
l_{SF}	criterion function of the subspace fitting technique.
l_{NSF}	criterion function of the noise subspace fitting technique.
J	criterion function of the ML tracking algorithm.
$\overline{J}$	limit function of J.
J_s	criterion function of the ML tracking algorithm under the assumption that the source waveforms are known.
$\overline{J}_s$	limit function of J_s.
G	gradient of J with respect to the MTS vector.
H	Hessain matrix of J with respect to the MTS vector.
G'	gradient of J_s with respect to the MTS vector when the source waveforms are assumed to be known.
H'	Hessain matrix of J_s with respect to the MTS vector when the source waveforms are assumed to be known.
μ_k	step-length used in the Gauss-Newton algorithm at the kth iteration.
W_{opt}	optimal weighting matrix for the subspace fitting technique.
U_{opt}	optimal weighting matrix for the noise subspace fitting technique.

Chapter 1

Introductions

This thesis is devoted to the direction-of-arrival (DOA) estimation of source targets by passive arrays, with emphasis given to the DOA tracking of moving source targets. In this chapter, we will give a brief introduction about array signal processing and provide an outline of the thesis.

1.1 Array signal processing

Sensor array systems have been in use for several decades in many practical applications, such as radar, sonar, seismic exploration and communications antenna systems. In an array system, the sensors are spatially distributed in known locations and collect data transmitted by the source targets in the array's field of view. There are two kinds of array sensor systems : active and passive. In the active sensing situation, a known waveform of finite duration is transmitted by the array. It propagates through the medium and is reflected back by the source targets to the array. The transmitted signals are usually modified by the target characteristics, and the information about the source targets can thus be extracted from the reflected signals. In the passive context, the sensor array only receives signal from the source targets. Propeller or engine noise from submarines is an example in the case of sonar. The received signal at the array is usually random in nature. In addition to the direct signals from the source targets, the signals received at the array may also contain clutter, spurious returns from the medium, medium ambient noise, undesired interference

and intentional jamming signals. Moreover, the multipath effects can create delayed, amplitude-weighted replicas of the direct signals at the array and generate coherent interference.

The signals can be classified as narrow-band or wide-band. A narrow-band signal is adequately characterized by a single frequency, while a wide-band signal occupies a significant frequency band. For reasons of simplicity, wideband signals can usually be transformed into narrow-band signals at the sensor for processing by applying narrow-band filters. In array processing, the definition of a narrow-band signal is given relative to physical size of the array. The source target can also be defined as far-field or near-field by virtue of the distance between the target and the array. For a far-field source signal, its wavefronts are well approximated by plane waves, while the propagation wavefronts of a near-field source signals are better represented by spherical waves.

The practical interest in array signal processing is to extract the source target parameters. The processing is done for detection and/or estimation. Detection deals with the problem of determining the number of source targets present, and estimation is more concerned with the direction-of-arrivals (DOA), signal waveforms, power levels or correlations between source signals. In many passive array applications, the receivers are subject to a variety of interference and noise, including natural or manmade signals other than the desired source signal. These signals which may occupy the same frequency band as the desired signals, can severely degrade the performance of the array systems. In radar and communications systems, it is necessary to suppress the undesired signals and interference to enhance the desired source signals. This task can be accomplished by using digital beamforming techniques. Most often, the desired source signals, undesired source signals and the ambient interference have time varying characteristics, and the digital beamforming systems must be devised with quick learning capabilities of the changing scene to maintain optimal system performance. Adaptation of the array system to the changing environment can be "blind" or "non-blind". Blind adaptation techniques do not require prior knowledge of the desired signal or interference characteristics, which may be difficult, costly, or simply impossible to obtain in some applications. Blind adaptation usually results in suboptimal solutions. In contrast, non-blind techniques usually involve using the desired signal and interference parameters, e.g. the desired source DOA and the ambient noise covariance matrix. But, they are more robust and perform

better than blind techniques provided the signal/interference parameters are sufficiently accurately estimated. In this thesis, we will discuss the problems of estimating the source DOA's, especially those of the moving source targets by passive arrays.

1.2 Outline of thesis

This thesis is focused on exploiting the spatial and temporal information contained in the received array data to estimate the source signal DOA's by passive arrays. It contains two parts. In the first part, we propose a unified constrained subspace fitting approach for estimating the DOA's of stationary source targets. This technique is based on the Karhunen-Loève expansion of the covariance matrix of the array manifold in a sector of interest. When the sources are spatially close, it is found that the signal subspace is contained in the array manifold subspace of small dimension than the number of sensors, and the optimal signal subspace is obtained by searching over the constrained array manifold subspace.

In the second part of the thesis, we deal with the extraction of parameters of moving source targets. We propose the maximum likelihood (ML) tracking technique based on the locally linear model of the source target motion dynamics. We define the multiple target state (MTS) vector to describe the target motion dynamics and show the strong observability of the locally linear motion model of the target in the context of passive array. We also present a modified Gauss-Newton type algorithm for optimizing the nonlinear criterion function. The asymptotic performance of the ML tracking technique has been fully investigated and is compared to that of the extended Kalman filter (EKF).

Chapter 2 is devoted to the study of passive array structure. In this chapter, we present the basic narrow-band array signal model and discuss the effects of far- and near-field signals on their induced time delays on the array. We also study the uniqueness conditions imposed on the array under which the source parameters can be identified.

Chapter 3 is a review of some of the high resolution DOA estimation techniques, including the subspace-based methods, deterministic maximum likelihood estimator (DMLE), stochastic maximum likelihood estimator (SMLE), multi-dimensional MUSIC (MD-MUSIC) and the subspace

fitting techniques. Different techniques are presented and discussed under a unifying subspace fitting framework.

Chapter 4 is devoted to the unified constrained subspace fitting approach for estimating the DOA's of stationary sources in a limited sector of interest. The geometric explanations for MUSIC, MD-MUSIC and DMLE are given, and new technique is developed. We also carry out the asymptotic performance analysis and provide numerical simulation results to show the effectiveness of the technique.

In Chapter 5, we present the ML tracking technique for moving source targets. We discuss the locally linear model for source target motion dynamics and its observability. The multiple target state (MTS) is defined to describe the source target motion. Lastly, we present the ML criterion function based on a batch of array data.

Chapter 6 is devoted to the optimization of the ML criterion. A modified Newton-type algorithm is provided in which the Hessian is approximated by a positive semi-definite matrix to ensure that the Newton-type algorithm is descent. We also present a method for estimating the initial MTS from the consecutive observations of the source target DOA's. Finally, numerical studies are conducted for the ML tracking technique via computer simulations.

In Chapter 7, we investigate the asymptotic performance of the maximum likelihood tracking estimator. We derive analytic formulas for the asymptotic covariance matrix of the MTS estimates and show the strong consistency of the estimates. This chapter ends with a discussion on the Cramér-Rao bounds of the MTS estimates.

Chapter 8 is devoted to the asymptotic performance comparisons between the extended Kalman filter (EKF) and the ML tracking technique. The source waveforms are assumed to be known. The asymptotic performances are analysed for both the EKF and the ML tracking techniques and conlusions are drawn.

Finally, in Chapter 9, we summarize the major contributions and describe future research topics.

Chapter 2

The Passive Array Structure

A passive array consists of a number of sensors located at different points in the field of interest. The passive array sensors collect the signals transmitted by the source targets in the space and estimates both the temporal and the spatial structure of the signal field. The array takes on a variety of different geometries depending on the application of interests, although the most commonly used configuration is the standard linear array in which the sensors are uniformly spaced with half the source signal wavelength.

2.1 Array sensor geometry

Assume that the sensors and the source targets are co-planar on the xy plane. Consider an array of M omni-directional sensors located at $\{\underline{z}_1, \underline{z}_2, \ldots, \underline{z}_M\}$, each vector $\underline{z}_m$ specified by

$$\underline{z}_m = [z_{mx}, z_{my}]^T, \quad m = 1, 2, \ldots, M, \tag{2.1}$$

where superscript T denotes the transpose operator.

Assume that K point source signals are incident in the array. It is also assumed that the medium is nondispersive and that the source wavefronts propagate at a constant velocity. Then, the received data $\bar{x}_m$, at the mth sensor, can be expressed as a sum of the time shifted versions of the original source signals

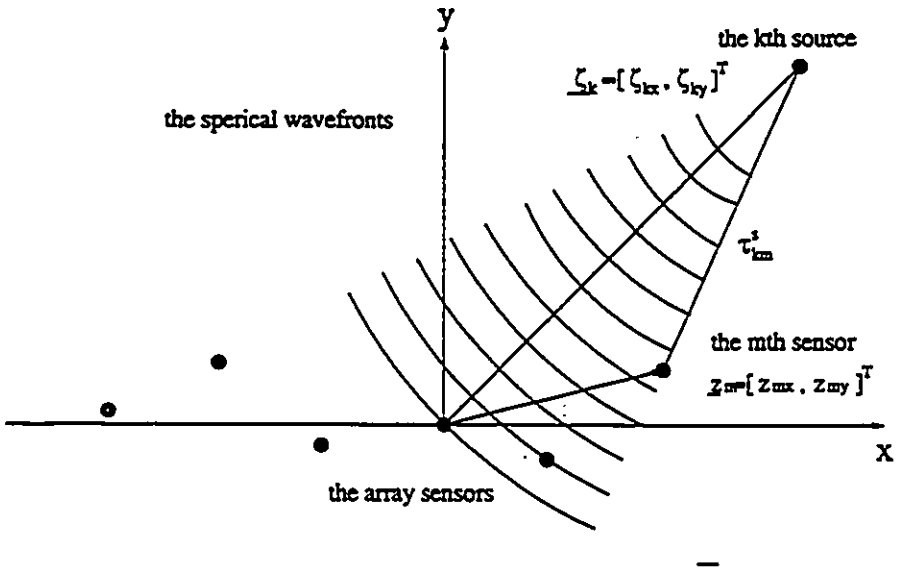

Figure 2.1: The array geometry for near-field source wavefronts.

$$\tilde{x}_m(t) = \sum_{k=1}^{K} \tilde{s}_k(t + \tau_{km}), \tag{2.2}$$

where $\tilde{s}_k(t)$ denotes the waveform of the kth source signal and τ_{km} is the relative time delay induced by the kth source signal on the mth sensor with respect to a specified origin of the co-ordinate system. The time delay τ_{km} depends on the array geometry and the source parameters. It also depends on the propagation pattern of the source wavefronts.

2.1.1 Near-field source wavefronts

A point source is known to propagate with spherical wavefronts in general. In array processing, when a source signal is located relatively close to the array, its spherical wavefronts reaching the array has a high degree of curvature. We define such source signals as near-field source signals. For near-field source signals, the time delay τ_{km} is calculated as the time required for the spherical wavefront to travel from the kth source to the mth sensor. The near-field source wavefronts and the array sensor geometry are shown in Fig. 2.1. Denote the kth source location as $\underline{\zeta}_k = [\zeta_{kx}, \zeta_{ky}]^T$, and the time delay τ^s_{km} can be calculated as

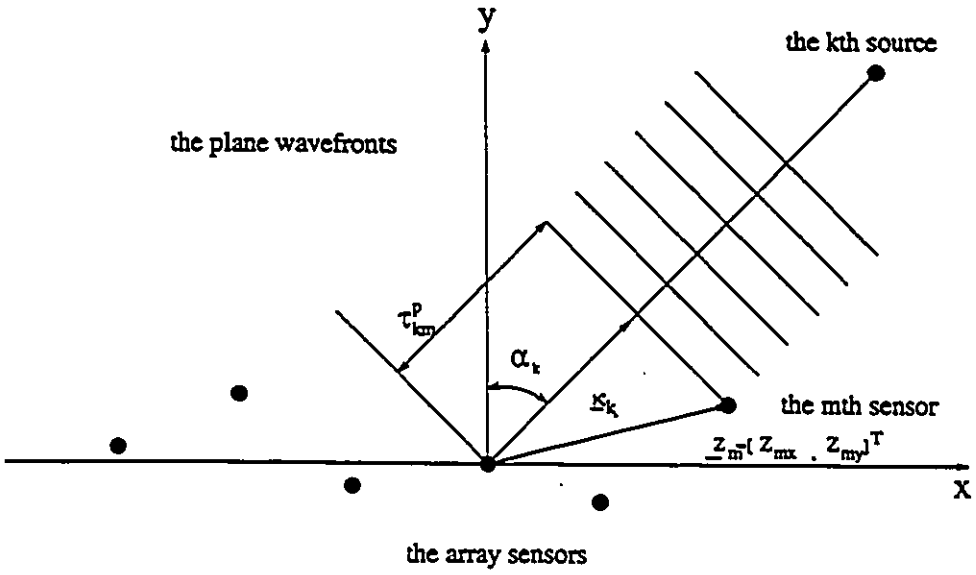

Figure 2.2: The array geometry for far-field source wavefronts.

$$\tau^s_{km} = \frac{1}{c}\sqrt{[z_{mx} - \zeta_{kx}]^2 + [z_{my} - \zeta_{ky}]^2}, \quad 1 \leq m \leq M, \tag{2.3}$$

where c is the propagation velocity in the medium.

2.1.2 Far-field source wavefronts

When the sources are located at a great distance from the array, the curvature of the spherical wavefront at the array is very small and can be neglected. The propagation of the source wavefronts can be well approximated by plane waves. This kind of source signals si usually called far-field. Define the kth source directional parameter vector $\underline{\kappa}_k$ as

$$\underline{\kappa}_k = [\cos(\alpha_k), \sin(\alpha_k)]^T,$$

where α_k is the azimuth angle of the kth source signal. From the array geometry shown in Fig. 2.2, the relative time delay τ_{km} can be obtained as

$$\tau_{km} = \frac{1}{c}\underline{z}_m \cdot \underline{\kappa}_k \quad m = 1, 2, \ldots, M \text{ and } k = 1, 2, \ldots, K. \tag{2.4}$$

Note that for a far-field signal, the directional parameter vector is sufficient to describe it. But, for a near-field source signal, its actual location coo-rdinates are required, i.e., the distance information of the sources must also be included in addition to the directional parameter vectors.

2.2 Assumptions about the source signals

The source signals $\{\tilde{s}_k(t);\ k = 1, 2, \ldots, K\}$ are assumed to be narrow-band and wide-sense stationary with zero mean. Denote the source signal bandwidth as $2\Delta\omega_{sc}$. A process is said to be a narrow-band process if its bandwidth is much less than its central frequency ω_0

$$\frac{2\Delta\omega_{sc}}{\omega_0} \ll 1. \tag{2.5}$$

In the narrow-band array signal model, the bandwidth of the source is assumed to satisfy

$$2\Delta\omega_{sc} \ll 2\pi/\tau_{max}, \tag{2.6}$$

where τ_{max} is the maximum time delay induced by the sources

$$\tau_{max} = \max_{m,n}\{|\ \tau_{km} - \tau_{kn}\ |, k = 1, 2, \ldots, K; m = 1, 2, \ldots, M\}.$$

Let $x'_m(t)$ denote the analytic representation of $\tilde{x}_m(t)$. It can be written as

$$x'_m(t) = \sum_{k=1}^{K} s_k(t + \tau_{km}) \exp\{j\omega_0\tau_{km}\}, \tag{2.7}$$

where $s_k(t)$ denotes the base-band representation of $\tilde{s}_k(t)$. From the assumption (2.5), it follows that $s_k(t)$ will remain essentially constant over the time duration needed to travel across the array,

$$s_k(t + \tau_{km}) \cong s_k(t), \quad k = 1, 2, \ldots, K;\ \ m = 1, 2, \ldots, M, \tag{2.8}$$

and the signal model (2.7) becomes

$$x'_m(t) = \sum_{k=1}^{K} s_k(t) \exp\{j\omega_0\tau_{km}\}. \tag{2.9}$$

2.3 Array signal model

The received data are always corrupted by noise. In the presence of additive noise, the mth sensor data can be written as

$$x_m(t) = x'_m(t) + w_m(t) = \sum_{k=1}^{K} s_k(t + \tau_{km}) \exp\{j\omega_0 \tau_{km}\} + w_m(t), \tag{2.10}$$

where $w_m(t)$ denotes the additive noise of the mth array sensor. It includes the sensor thermal noise and the ambient noise of the medium. Examples of ambient noise are distant ship traffic, ocean turbulence and thermal noise. The additive sensor noise component is assumed to be wide-sense stationary with zero mean and uncorrelated with the signals. Assume that the array sensor output has been sampled with a sampling interval of T. We use index n to denote the nth sampling time nT. In matrix form, the array signal model (2.10) can be written as

$$\underline{x}(n) = A(\Theta)\underline{s}(n) + \underline{w}(n), \tag{2.11}$$

where

$$\underline{x}(n) = [x_1(n), x_2(n), \ldots, x_M(n)]^T$$

$$\underline{s}(n) = [s_1(n), s_2(n), \ldots, s_K(n)]^T$$

$$\underline{w}(n) = [w_1(n), w_2(n), \ldots, w_M(n)]^T$$

are defined respectively as the array data vector, the signal vector and the noise vector. $A(\Theta)$ is the array composite steering matrix given by

$$A(\Theta) = [\underline{a}(\theta_1), \underline{a}(\theta_2), \ldots, \underline{a}(\theta_K)]. \tag{2.12}$$

Here Θ is used to represent the source DOA parameters and θ_k denotes the kth source DOA. The kth column of $A(\Theta)$, $\underline{a}(\theta_k)$, is defined as the array steering vector associated with the kth signal given by

$$\underline{a}(\theta_k) = [\exp\{j\omega_0 \tau_{k1}\}, \exp\{j\omega_0 \tau_{k2}\}, \ldots, \exp\{j\omega_0 \tau_{kM}\}]^T. \tag{2.13}$$

The array steering vector depends on the array structure and the source parameters. For later references, we introduce the following definitions. We define the array manifold as the set of steering vectors, $\{\underline{a}(\theta)\}$, over all possible source DOA parameters and the signal subspace as the K-dimensional subspace spanned by the columns of $A(\Theta)$. Also, a uniqueness condition is applied to $A(\Theta)$. The array is assumed to be unambiguous, i.e. the parameterized $A(\Theta)$ has full rank. The *uniqueness condition* is imposed on the array to enable the unique identification of the DOA parameters, i.e.,

$$A(\Theta_1)T_1 = A(\Theta_2)T_2 \Longrightarrow \Theta_1 = \Theta_2, \tag{2.14}$$

where T_1 and T_2 are any full rank matrices of size $K \times K$ [1].

Chapter 3

High Resolution DOA Estimation Techniques

In this chapter, we present a review of some of the most commonly employed high resolution techniques in array processing. These techniques include the subspace-based techniques[2]-[9], the multi-dimensional MUSIC (MD-MUSIC) algorithm [10], the deterministic and stochastic maximum likelihood estimators [15][30], and the subspace fitting techniques[1][53]. These techniques are discussed within the subspace fitting framework.

3.1 The array data covariance matrix

The second-order statistics of the array data are particularly insightful and useful for the detection and estimation problems in high resolution array signal processing. From the array signal model (2.11), the array data covariance matrix can be written as

$$R_x = E[\underline{x}(n)\underline{x}^H(n)] = A(\Theta)R_s A^H(\Theta) + \sigma_w^2 R_w, \tag{3.1}$$

where E denotes the expectation operator and the superscript H denotes the conjugate transpose operator. The $M \times M$ matrix $R_w = E[\underline{w}(n)\underline{w}^H(n)]$ and the $K \times K$ matrix $R_s = E[\underline{s}(n)\underline{s}^H(n)]$ denote the covariance matrices of the source and the noise processes respectively. Parameter σ_w^2 is defined

as the unknown sensor noise variance and it is assumed that the trace of R_w has been normalized to M. When the source signals are mutually uncorrelated, the source signal covariance matrix R_s is a non-singular diagonal matrix with each element in the diagonal representing the variance of source. When the source signals are partially coherent or correlated, R_s is non-diagonal, but still nonsingular. If a subset of the source signals are coherent or fully correlated, the covariance matrix becomes both nondiagonal and singular. Coherency happens when there is a subset of linearly dependent source signals. Coherent source signals can arise from multipath effects, or they can be induced by evasive signalling techniques. In practice, the exact array data covariance matrix is not available and is usually estimated from the array data by averaging over the array data vector samples

$$\hat{R}_x = \frac{1}{N}\sum_{n=1}^{N} \underline{x}(n)\underline{x}^H(n), \tag{3.2}$$

where N denotes the number of array data samples.

3.2 Subspace-based estimation algorithms

From eigen-decomposition theory, it is known that the generalized eigen-values of the matrix pair (R_x, R_w) are distributed as [10]

$$\lambda_1 \geq \ldots \geq \lambda_r > \lambda_{r+1} = \ldots = \lambda_K = \ldots = \lambda_M = \sigma_w^2, \tag{3.3}$$

where $r \leq K$, with equality holding for incoherent sources. Define the eigen-values $\{\lambda_k,\ k = 1,2,\ldots,r\}$ as the signal eigen-values and their associated eigen-vectors $\{\underline{u}_k,\ k = 1,2,\ldots,r\}$ as the signal eigen-vectors. Define $\{\lambda_k, k = r+1, r+2, \ldots, M\}$ as the noise eigen-values and their associated eigen-vectors $\{\underline{u}_k, k = r+1, r+2, \ldots, M\}$ as the noise eigen-vectors. For incoherent source signals, the signal eigen-vectors span the K-dimensional subspace identical to the signal subspace [10]

$$span\{\underline{u}_1, \ldots, \underline{u}_K\} = span\{\underline{a}(\theta_1), \ldots, \underline{a}(\theta_K)\},$$

and the noise eigen-vectors span the $(M-K)$-dimensional subspace identical to noise subspace. Since the noise subspace is the orthogonal complement of the signal subspce, we have the following

$$\underline{a}(\theta_k)^H \underline{u}_i = 0 \quad i = K+1, \ldots, M \text{ and } k = 1, \ldots, K, \tag{3.4}$$

which is fundamental to the subspace based techniques. The MUSIC algorithm [3][10] fully exploits this orthogonality. It determines the source DOA parameters by locating the peaks of a spatial spectrum formed from the inverse of the norm of the projection of a continuum of steering vectors onto the noise subspace

$$P_{MUSIC}(\theta) = [\underline{a}^H(\theta)\hat{U}_N\hat{U}_N^H\underline{a}(\theta)]^{-1}, \tag{3.5}$$

where $\hat{U}_N = [\hat{\underline{u}}_{K+1}, \ldots, \hat{\underline{u}}_M]$ denotes the estimate of $\underline{U}_N$. There are also the Mini-Norm (MN) estimator [6] and the eigen-vector (EV) method proposed by Johonson and DeGraff [5]. The spatial spectrum of the MN algorithm is formed by projecting the continuum of steering vectors onto a one dimension subspace

$$P_{MN}(\theta) = [\underline{a}^H(\theta)\hat{\underline{u}}_{MN}\hat{\underline{u}}_{MN}^H\underline{a}(\theta)]^{-1}, \tag{3.6}$$

where $\hat{\underline{u}}_{MN}$ is from the noise subspace with its first element set to unity and has minimum norm among the vectors in the noise subspace. The MN estimator usually has a higher estimation variance and resolution capability than the MUSIC algorithm [11]. In the EV method, the projections of the array steering vector to the estimated noise eigen-vectors are each weighted according to the amplitudes of their corresponding eigen-values

$$P_{EV}(\theta) = [\underline{a}^H(\theta)P_{EV}\underline{a}(\theta)]^{-1}, \tag{3.7}$$

where $P_{EV} = \hat{U}_N\hat{D}_\lambda\hat{U}_N^H$ and $\hat{D}_\lambda = diag[\hat{\lambda}_{K+1}, \ldots, \hat{\lambda}_M]$. The EV method has higher resolution than the MUSIC algorithm and is less sensitive to the selection of the number of source signals. The MUSIC estimator is asymptotically efficient when there is only one source signal present. For multiple source signals, the algorithm is not asymptotically efficient in general. However, it

approaches asymptotic efficiency when the SNR's of all sources tend to infinity [12]. The asymptotic covariance matrix of the MUSIC estimates is given by [13]

$$C_{MU} = \frac{\sigma_w^2}{2N}[H_{MU} \odot I]^{-1} Re\{H_{MU} \odot [A^H(\Theta)UA(\Theta)]^T\}[H_{MU} \odot I]^{-1}, \tag{3.8}$$

where Re represents taking the real part and $\odot$ denotes the Hadamard product which is the element-wise multiplcation of two matrices. In (3.8), H_{MU} is defined as

$$H_{MU} = D^H P^{\perp}_{A(\Theta)} D,$$

where D is given by

$$D = \left[\frac{\partial \underline{a}(\theta)}{\partial \theta}\,|_{\theta=\theta_1}, \frac{\partial \underline{a}(\theta)}{\partial \theta}\,|_{\theta=\theta_2}, \ldots, \frac{\partial \underline{a}(\theta)}{\partial \theta}\,|_{\theta=\theta_K}, \right], \tag{3.9}$$

and U is implicitly defined by

$$A^H(\Theta)UA(\Theta) = R_s^{-1} + \sigma_w^2 R_s^{-1}[A^H(\Theta)A(\Theta)]^{-1}R_s^{-1},$$

where $P^{\perp}_{A(\Theta)} = I - A(\Theta)\{A^H(\Theta)A(\Theta)\}^{-1}A^H(\Theta)$ denotes the orthogonal projector onto the noise subspace. The diagonal elements of C_{MU} give the variances of the MUSIC estimates.

3.3 The maximum likelihood estimators

The maximum likelihood method is a popular estimation procedure in statistical studies. When the ML technique is applied to array processing, two classes of estimators are generated depending on the assumptions about the source waveforms. When the sources are modeled as unknown but deterministic, the resulting estimator is referred to as the deterministic ML (DML) estimator. When the sources are modeled as Gaussian random processes, the stochastic ML (SML) estimator is obtained.

3.3.1 The deterministic maximum likelihood method

In many applications, for example, in radar and radio communication systems, the transmitted signal waveforms are far from being random, and it is natural to model the source waveforms

as unknown but deterministic. In (2.11), we assume that sensor noise $\underline{w}(n)$ is a stationary and ergodic complex Gaussian process with zero mean and covariance matrix $\sigma_w^2 I$, where σ_w^2 is an unknown and I denotes the identity matrix. Assume that $\{\underline{w}(n)\}$ are statistically independent. The conditional probability density function of the array is formed by conditioning on the unknown source waveforms, noise variance and the source DOA parameters

$$p(\underline{x}(1), \underline{x}(2), \ldots, \underline{x}(N)) = \prod_{i=1}^{N} \frac{1}{\pi^M \sigma_w^{2M}} \exp\left\{-\frac{1}{\sigma_w^2}\{\underline{x}(i) - A(\Theta)\underline{s}(i)\}^H\{\underline{x}(i) - A(\Theta)\underline{s}(i)\}\right\}, \quad (3.10)$$

and the negative log likelihood function, ignoring the constant term, has the following form

$$L = -\log p = MN \log(\pi\sigma_w^2) + \frac{1}{\sigma_w^2} \sum_{i=1}^{N} \| \underline{x}(i) - A(\Theta)\underline{s}(i) \|_F^2, \quad (3.11)$$

where $\| \cdot \|_F$ is the Frobenius norm. The principle of the ML estimation is to maximize the likelihood function with respect to all the unknown parameters [14] and obtain the maximizing arguments as the corresponding ML estimates. First, by fixing Θ and $\underline{s}(i)$ and minimizing L with respect to σ_w^2, we obtain the noise variance estimate as

$$\hat{\sigma}_w^2 = \frac{1}{MN} \sum_{i=1}^{N} \| \underline{x}(i) - A(\Theta)\underline{s}(i) \|_F^2 . \quad (3.12)$$

Substituting $\hat{\sigma}_w^2$ back into (3.11) shows that the ML estimates of $\underline{s}(i)$ and Θ can be solved from the following nonlinear least-squares problem

$$[\hat{\underline{s}}(i), \hat{\Theta}] = \arg \min_{\underline{s}(i), \Theta} \sum_{i=1}^{N} \| \underline{x}(i) - A(\Theta)\underline{s}(i) \|_F^2 . \quad (3.13)$$

Since criterion in (3.13) is quadratic in $\underline{s}(i)$, fixing Θ and minimizing the criterion with respect to $\underline{s}(i)$ yields

$$\hat{\underline{s}}(i) = \{A^H(\Theta)A(\Theta)\}^{-1} A^H(\Theta)\underline{x}(i). \quad (3.14)$$

By substituting (3.14) back into (3.13), we obtain the objective function for ML estimator of the source DOA parameters

$$
\begin{aligned}
l'_{DML} &= \sum_{i=1}^{N} \| \underline{x}(i) - P_{A(\Theta)}\underline{x}(i) \|_F^2 \\
&= \sum_{i=1}^{N} \| P^{\perp}_{A(Theta)}\underline{x}(i) \|_F^2 \\
&= \sum_{i=1}^{N} tr\{P^{\perp}_{A(\Theta)}\underline{x}(i)\underline{x}^H(i)\}, \qquad (3.15)
\end{aligned}
$$

where $P_{A(\Theta)} = A(\Theta)\{A^H(\Theta)A(\Theta)\}^{-1}A^H(\Theta)$ is the orthonormal projection onto the signal subspace. Since l'_{DML} may not be bounded and is not well defined when N approaches infinity, we divide l'_{DML} by N to obtain an equivalent ML criterion l_{DML} so that the DML estimate of Θ is given by

$$
\hat{\Theta} = \arg\min_{\Theta} l_{DML} = \arg\min_{\Theta} \frac{1}{N}\sum_{i=1}^{N} tr\{P^{\perp}_{A(\Theta)}\hat{R}_x\}. \qquad (3.16)
$$

The DML estimate $\hat{\Theta}$ has been shown to be consistent [1]. However, since the dimension of the unknown source waveform parameter set, $S = [\underline{s}(1), \ldots, \underline{s}(N)]$, grows in proportion to the number of the array samples, the source waveform estimates $\hat{\underline{s}}(i)$ are not consistent. This inconsistency can be seen from (3.14),

$$
\hat{S} = A^{\dagger}(\Theta)X \rightarrow S + A^{\dagger}(\Theta)W_n, \qquad (3.17)
$$

where $W_n = [\underline{w}(1), \ldots, \underline{w}(N)]$, and $A^{\dagger}(\Theta)$ denotes the pseudo-inverse of $A(\Theta)$. Asymptotically, the distribution of the DML estimate error $\sqrt{N}(\hat{\Theta} - \Theta)$ converges in distribution to $N(0, C_{DML})$ with C_{DML} given by [13]

$$
\frac{1}{N}C_{DML} = CRB_{DET} + 2N\ CRB_{DET}Re\{[D^H P^{\perp}_{A(\Theta)}D] \odot [A^H(\Theta)A(\Theta)]^{-T}\}. \qquad (3.18)
$$

where CRB_{DET} is the deterministic Cramér-Rao bound (CRB) given by [13]

$$
CRB_{DET} = \frac{\sigma_w^2}{2}[Re\{[D^H P^{\perp}_{A(\Theta)}D] \odot R_s^T\}]^{-1}. \qquad (3.19)
$$

In (3.18), since the second term can be shown to be positive semi-definite, we have

$$\frac{1}{N}C_{DML} > CRB_{DET},$$

i.e., the asymptotic covariance of the DML estimates does not attain the CRB. Thus, the DML estimates are not asymptotically efficient. Note that this conclusion is not applicable to those cases in which the number of parameters does not grow as the number of the measurements increases. In this latter case, the resulting ML estimates are asymptotically efficient, i.e., their asymptotic covariance coincides with the CRB under Gaussian assumption.

When R_s is diagonal, it can be shown [13] that the covariance of the MUSIC estimates, C_{MUSIC} in (3.7) approaches C_{DML} in (3.18) asymptotically. This implies that the MUSIC estimator is a large sample realization of the DMLE when the source signals are uncorrelated. The ML estimator (3.16) has an appealing geometric interpretation. From array signal model (2.11), it is observed that $\{\underline{x}(n)\}$ stays in the K-dimensional column space of $A(\Theta)$ in the absence of noise and is perturbed out of the signal subspace by the noise $\underline{w}(n)$. It follows that the ML estimates are obtained by searching for the K steering vectors over the array manifold that form the K-dimensional signal subspace closest to the vectors $\{\underline{x}(n)\}$ in the least squares error sense. The closeness is measured by the sum of moduli of the projections of $\{\underline{x}(n)\}$ onto the signal subspace.

3.3.2 The stochastic maximum likelihood estimator

Although in some communications systems the source waveforms turn out be more deterministic than random, in most passive radar and sonar systems, random modeling is adequate due to the physical sources responsible for the generation of these processes. Gaussian assumption is usually used for the distributions of the source waveforms. The Gaussian distribution has the advantage of mathematical convenience and is often motivated by the central limit theorem. By assuming independent array samples, the conditional probability density function of the array data vectors can be written as

$$p(\underline{x}(1),\underline{x}(2),\ldots,\underline{x}(N) \mid \Theta, R_s, \sigma_w^2) = \prod_{i=1}^{N} \frac{1}{\pi^M \mid R_x \mid} \exp\left\{-\underline{x}(i)R_x^{-1}\underline{x}(i)\right\}, \qquad (3.20)$$

where $|\cdot|$ denotes a determinant. The negative log likelihood function can be written as (ignoring the constant terms)

$$l_{SML} = N \log | R_x | + \sum_{i=1}^{N} \underline{x}^H(i) R_x^{-1} \underline{x}(i), \tag{3.21}$$

and the SML estimates are obtained as

$$[\hat{\Theta}, \hat{R}_s, \hat{\sigma}_w^2] = \arg \min_{\Theta, R_s, \sigma_w^2} l_{SML}. \tag{3.22}$$

Criterion l_{SML} can be further written as

$$l_{SML} = \log | R_x | + tr\{R_x^{-1} \hat{R}_x\}, \tag{3.23}$$

in which we have used the matrix property $tr\{AB\} = tr\{BA\}$ for A and B of appropriate dimensions. Minimizing (3.23) with respect to R_s and σ_w^2 yields [15][16],

$$\hat{R}_s = A^{\dagger}(\Theta)(\hat{R}_x - \hat{\sigma}_w^2) A^{\dagger H}(\Theta) \tag{3.24}$$

$$\hat{\sigma}_w^2 = \frac{1}{M-K} tr\{P_{A(\Theta)}^{\perp} \hat{R}_x\}. \tag{3.25}$$

By substituting (3.24) and (3.25) back into (3.22), we obtain the SML estimate for Θ as

$$\hat{\Theta} = \arg \min_{\Theta} l_{SML}, \tag{3.26}$$

where

$$l_{SML} = \log | A(\Theta) \hat{R}_s A^H(\Theta) + \hat{\sigma}_w^2 | . \tag{3.27}$$

Note that in minimizing l_{SML}, we can also include other a-priori information on R_s. For example, the positive semi-definite assumption on R_s leads to a potentially different ML estimator [17]. Somtimes, the rank deficiency condition of R_s due to the source coherency can also be taken into consideration. However, it is found that all these efforts cannot improve the asymptotic statistical properties of the SML estimator but result in a significantly more complicated optimization

problem. The SML estimator is asymptotically efficient. The asymptotic covariance of the SML estimates attains the stochastic CRB which is given by [18]

$$CRB_{STO} = \frac{\sigma_w^2}{2}[Re\{[D^H P^{\perp}_{A(\Theta)} D] \odot [R_s A^H(\Theta) R_x^{-1} A(\Theta) R_s\}]^{-1}. \tag{3.28}$$

3.4 MD-MUSIC and the subspace fitting techniques

The MD-MUSIC algorithm [1][10][20] is a multi-dimensional extension of the one-dimensional MUSIC algorithm. The subspace fitting techniques [1][19] search for the optimal source parameter estimates that forms a subspace closest (or farthest) in distance from the estimated signal (or noise subspace). The subspace fitting framework can be usd to obtain a unified derivation of the asymptotic properties of the MD-MUSIC algorithm, the DMLE and SMLE, and the weighted subspace fitting (WSF) method.

3.4.1 Description of the MD-MUSIC algorithm

The MD-MUSIC algorithm was first reported by Schmidt [10] and formulated by Cadzow [20] and Viberg and Ottersten [1]. Let $U_s = [\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_r]$ and let $\hat{U}_s$ be the estimate from the eigen-decomposition of the array covariance matrix. It follows that the source parameters can be estimated by minimizing the following criterion in the least squares sense

$$L = \| \hat{U}_s - A(\Theta)Q \|_F, \tag{3.29}$$

where Q is a full rank matrix of size $K \times K$. Minimizing L with respect to Q yields

$$Q = A^{\dagger}(\Theta)\hat{U}_s, \tag{3.30}$$

By substituting (3.30) back into (3.29), we get

$$\hat{\Theta} = \arg\max_{\Theta} l_{MD}, \tag{3.31}$$

where

$$l_{MD} = tr\{P_{A(\Theta)}\hat{U}_S\hat{U}_S^H\}. \tag{3.32}$$

The MD-MUSIC algorithm can successfully overcome the deficiency due to source signal coherency and provide better performance than the one-dimensional MUSIC algorithm. In the presence of coherent source signals, it is superior to the DML estimator. But, when the sources are incoherent it is outperformed by DML estimator. Similar to most nonlinear multi-dimensional optimization problems, the MD-MUSIC algorithm requires a significant amount of computation and suffers from the lack of assured global convergence.

The MD-MUSIC estimator is closely related to the DMLE. To see this, we write the spectral representation $\hat{R}_x$ as

$$\hat{R}_x = \sum_{i=1}^{M} \hat{\lambda}_i \underline{\hat{u}}_i \underline{\hat{u}}_i^H. \tag{3.33}$$

and the DMLE criterion in (3.16) becomes

$$l_{DML} = \sum_{i=1}^{M} \hat{\lambda}_i \parallel P_{A(\Theta)}\underline{\hat{u}}_i \parallel^2 . \tag{3.34}$$

where l_{MD} can be written as

$$l_{MD} = \sum_{i=1}^{r} \parallel P_{A(\Theta)}\underline{\hat{u}}_i \parallel^2 . \tag{3.35}$$

It shows that MD-MUSIC has included only the estimated signal eigen-vectors while DMLE has utilized both the estimated signal and the estimated noise eigen-vectors in the algorithm. Also in DMLE, each projection of the eigen-vector onto the signal subspace is weighted by its corresponding eigen-value.

3.4.2 The subspace fitting techniques

The subspace fitting technique [1] is a general framework in which the DMLE, SMLE, MD-MUSIC and the ESPIRIT algorithm [22]-[25] can be formulated. Asymptotically, the subspace fitting method maximizes the following criterion

$$l_{SF} = tr\{P_{A(\Theta)}\hat{U}_s W \hat{U}_s\}, \tag{3.36}$$

where W is a weighting matrix of dimension $r \times r$. Different choices of W will yield potentially different estimators. For example, when $W = I$, l_{SF} is identical to the MD-MUSIC criterion (3.31), and the subspace fitting technique produces the MD-MUSIC estimator. If we choose $W = \Lambda_s - \sigma_w^2 I$, where $\Lambda_s = diag[\lambda_1, \ldots, \lambda_r]$, then the estimator will have the same asymptotic distribution as the DML estimator [1]. This can be shown as follows. Let $\Lambda_n = diag[\lambda_{r+1}, \ldots, \lambda_M]$ and denote $\hat{\Lambda}_n$ as its estimate. Then, the DMLE criterion in (3.16) can be written as

$$l_{DML} = tr\{P_{A(\Theta)}\hat{R}_z\} = tr\{P_{A(\Theta)}(\hat{U}_s\hat{\Lambda}_s\hat{U}_s^H + \hat{U}_n\hat{\Lambda}_n\hat{U}_n^H)\}. \tag{3.37}$$

In [1], it is shown that $\hat{\Lambda}_n$ can be replaced by $\sigma_w^2 I$ without affecting the asymptotic properties of the estimator. Thus, l_{DML} becomes asymptotically

$$l_{DML} = tr\{P_{A(\Theta)}\hat{U}_s(\hat{\Lambda}_s - \sigma_w^2 I)\hat{U}_s^H\}, \tag{3.38}$$

Again, since the replacement of $(\hat{\Lambda}_s - \sigma_w^2 I)$ by $(\Lambda_s - \sigma_w^2 I)$ does not change the asymptotic distribution of the estimates [1], we get the conclusion that the DMLE is asymptotically identical to the subspace fitting technique with $W = \Lambda_s - \sigma_w^2 I$.

The estimates obtained from maximizing the criterion (3.36) are consistent estimates [1]. The weighted subspace fitting (WSF) method [1] [53] is defined as the optimal estimator which provides the minimum variance estimates in the subspace fitting framework. The optimum weighting matrix for WSF is identified as

$$W_{opt} = \tilde{\Lambda}^2\Lambda_s^{-1} = \tilde{\Lambda} + \sigma_w^4\Lambda_s^{-1} - \sigma_w^2 I, \tag{3.39}$$

where $\tilde{\Lambda} = \Lambda_s - \sigma_w^2 I$. The WSF method has been shown [1] to be a large sample realization of the SML estimator, i.e., their asymptotic distributions of the normalized estimation errors coincide.

The subspace fitting technique can also be formulated in the noise subspace when the sources are incoherent. The criterion is described by the weighted least squares measure of the orthogonality between the signal and the noise subspaces

$$
\begin{aligned}
l_{NSF} &= \| \hat{U}_n^H A(\Theta) \|_U^2 \\
&= tr\{U A(\Theta) A^H(\Theta) \hat{U}_n \hat{U}_n^H\},
\end{aligned} \tag{3.40}
$$

where U is a positive semi-definite $K \times K$ weighting matrix. Similar to the signal subspace fitting technique, different choices of U will result in estimates with different asymptotic performances. When U is a diagonal matrix, criterion (3.40) decouples, and the estimator becomes the MUSIC estimator. The signal and the noise subspace fitting techniques are closely related. When W and U are related by

$$
U = A^{\dagger}(\Theta) \hat{U}_s W \hat{U}_s^H A^{\dagger H}(\Theta), \tag{3.41}
$$

the resultant respective signal/noise subspace fitting mthods will be asymptotically equivalent. For example, when $U = A^{\dagger}(\Theta) \hat{U}_s \hat{\Lambda} \hat{U}_s^H A^{\dagger}(\Theta)$, the noise subspace fitting technique produces asymptotically the DML estimator. Consequently, the optimal weighting matrix U_{opt} which minimizes the estimation error covariance is

$$
U_{opt} = A^{\dagger}(\Theta) \hat{U}_s W_{opt} \hat{U}_s^H A^{\dagger H}(\Theta), \tag{3.42}
$$

It is pointed out that when the source signals are incoherent, both of the weighting matrices W and U can be replaced by a consistent estimate, and an aymptotically equivalent estimator is still obtained. However, for coherent source signals, i.e., $r < K$, the replacement of Θ in U_{opt} with a consistent estimate does affect the asymptotic performance of the estimate. Also, for coherent source signals, the optimally weighted noise subspace method is not asymptotically equivalent to the SML estimator.

3.5 Asymptotic performance analysis

Most of the high resolution techniques are based on the eigen-decomposition of the estimated array data covariance matrix $\hat{R}_x$. Under the independent Gaussian assumption for the array data vectors

$\{\underline{x}(n)\}$, the covariance estimate $\hat{R}_x$ in (3.1) can be shown to be consistent, i.e., the elements of $\hat{R}_x$ converge with probability one (w.p.1) to those of R_x. Assume that the signal eigen-values $\{\hat{\lambda}_i,\ i = 1,2,\ldots,r\}$ are distinct. The following limits hold w.p.1 as N approaches infinity

$$\begin{aligned} \hat{\lambda}_k &\rightarrow \lambda_k \quad k = 1,2,\ldots,r \\ \underline{\hat{u}}_k &\rightarrow \underline{u}_k \quad k = 1,2,\ldots,r \end{aligned} \tag{3.43}$$

In what follows, we summarize some of the asymptotic performance comparisons between various high resolution methods.

A. MUSIC $\neq$ SMLE It is known that the maximum likelihood estimator satisfies invariance property [47]. In [50], Sharman claimed that since $\hat{R}_x$ is the ML estimate of R_x, the MUSIC estimates are also the ML estimates by the invariance principle. Actually, this is not true for two reasons. First, $\hat{R}_x$ is not the ML estimate of R_x even under Gaussian assumptions since the structure of R_x has been incorporated in maximizing the likelihood function. $\hat{R}_x$ is usually referred to as the unstructured ML estimate of R_x and is an approximate of the ML estimate. Secondly, the invariance principle is not completely applicable in this case. By the invariance princople, the ML estimate, $\hat{\theta}_k$ for $k = 1,2,\ldots,K$, should satisfy the following relationship

$$[P_{MUSIC}(\hat{\theta}_i)]^{-1} = 0. \tag{3.44}$$

In the MUSIC estimator, the source parameters are obtained as the minimizing arguments of $[P_{MUSIC}(\theta)]^{-1}$. Although the minimum values could be very small under high SNR and large sample size, they are never zero as required by the invariant principle (3.44).

B. THE ASYMPTOTIC ERROR DISTRIBUTION The asymptotic distribution of the estimation error of the general subspace fitting technique is derived as [1]

$$\sqrt{N}(\hat{\Theta} - \Theta) \in N(0, C), \tag{3.45}$$

where

$$C = \overline{V}^{-1} Q \overline{V}^{-1}, \tag{3.46}$$

and matrices $\overline{V}$ and Q are given by

$$\begin{aligned} \overline{V} &= -2Re\{[D^H P^{\perp}_{A(\Theta)} D] \odot [A^{\dagger}(\Theta) U_s W U_s^H A^{\dagger H}(\Theta)]^T\} \\ Q &= 2\sigma_w^2 Re\{[D^H P^{\perp}_{A(\Theta)} D] \odot [A^{\dagger}(\Theta) U_s W \Lambda_s \bar{\Lambda}^{-2} W U_s^H A^{\dagger H}(\Theta)]^T\}, \end{aligned} \tag{3.47}$$

The noise subspace based techniques have the same asymptotic distribution result (3.47) provided that U is replaced conformally with (3.41). Since the MD-MUSIC estimator is asymptotically identical to the result of subspace fitting when W is replaced by I in (3.45), the asymptotic covariance of the MD-MUSIC estimates is given by (3.46) with

$$\begin{aligned} \overline{V} &= -2Re\{[D^H P^{\perp}_{A(\Theta)} D] \odot [A^H(\Theta) A(\Theta)]^{-T}\} \\ Q &= 2\sigma_w^2 Re\{[D^H P^{\perp}_{A(\Theta)} D] \odot \{[A^H(\Theta) A(\Theta) R_s A^H(\Theta) A(\Theta)]^{-1} \\ &\quad + \sigma_w^2 [A^H(\Theta) A(\Theta) R_s A^H(\Theta) A(\Theta) R_s A^H(\Theta) A(\Theta)]\}\}. \end{aligned} \tag{3.48}$$

Similarly, since the replacement of W by $\bar{\Lambda}$ in (3.45) yields asymptotically the DMLE, the asymptotic covariance of the DML estimates can be obtained from (3.46) as

$$C = CRB_{DET} + \overline{V}^{-1} \bar{Q} \overline{V}^{-1}, \tag{3.49}$$

where

$$\begin{aligned} \overline{V} &= -2Re\{[D^H P^{\perp}_{A(\Theta)} D] \odot R_s^T\} \\ \bar{Q} &= 2\sigma_w^4 Re\{[D^H P^{\perp}_{A(\Theta)} D] \odot [A^H(\Theta) A(\Theta)]^{-T}\}, \end{aligned} \tag{3.50}$$

which coincides with the directly obtained covariance (3.18). This reconfirms the conclusion that the subspace fitting estimator with $W = \bar{\Lambda}$ and DML estimator are asymptotically equivalent.

C. THE MUSIC AND DML ESTIMATORS In [13][12], it has been shown that the MUSIC estimates are large sample realizations of the DML estimates when the source signals are uncorrelated. This can be further confirmed by observing the DML weighting in the noise subspace fitting technique

$$U = A^{\dagger}(\Theta)\hat{U}_s\tilde{\Lambda}\hat{U}_s^H A^{\dagger H}(\Theta) - R_s.$$

which implies that the DML weighting matrix U approaches the source covariance matrix with sufficiently large number of array samples. Since R_s is diagonal for uncorrelated source signals, the resulting noise subspace fitting technique asymptotically approaches the MUSIC method. Thus, it is rediscovered that MUSIC is a large sample realization of the DMLE when the sources are uncorrelated.

D. THE ASYMPTOTIC EQUIVALENCE BETWEEN WSF AND SMLE When the source signals are incoherent, the asymptotic distribution of the signal and noise WSF techniques (3.36) and (3.40) can be obtained from (3.46) as

$$C_{opt} = \frac{\sigma_w^2}{2}[Re\{[D^H P^{\perp}_{A(\Theta)} D] \odot [R_s A^H(\Theta) R_x^{-1} A(\Theta) R_s\}]^{-1}, \tag{3.51}$$

which coincides with the covariance of the SML estimates (3.28), or CRB_{STO}. This indicates the fact that the weightings W_{opt} and U_{opt} in (3.36) and (3.40) will result in estimators which are large sample realizations of the SML method. However, as mentioned before, when R_s is rank deficient (coherent sources), the optimal weighting U_{opt} does not produce asymptotically equivalent estimates to those of the SML estimator.

Chapter 4

DOA Estimation on the Array Manifold Space

In this chapter, we present a high resolution technique for estimating DOA's of spatially close source signals. This technique is based on the fact that the array manifold subspace over a sector of interest can be well approximated by a subspace with a dimension smaller than the number of sensors in the array, and the optimal signal subspace is then sought within this array manifold space. Asymptotic performance analysis is carried out, and computer simulations are also provided.

4.1 Introduction and preliminaries

High resolution DOA estimation techniques have been exploited in many sensor systems such as radar, sonar and seismic exploration systems. Subspace-based techniques [2]-[9] constitute some of the more effective procedures for obtaining quality DOA estimates. The subspace-based techniques are known to yield, with high resolution, asymptotically unbiased estimates. The MD-MUSIC algorithm is first proposed by Schmidt [10] and formulated by Cadzow [20] and Viberg and Ottersten [1]. It is statistically robust and effective for coherent source signals. However, the MD-MUSIC algorithm involves nonlinear optimization problems and requires intensive computation. The DML

estimator is another popular technique in array processing. It provides statistically robust estimation results and outperforms the subspace-based algorithms for uncorrelated source signals. But, due to the highly nonlinear nature of the criterion, the DMLE requires that nonlinear programming algorithms be employed and thus, carries a heavy computational burden.

Much effort has been spent to enhance the resolution capability of the subspace-based DOA estimation techniques. For example, in [21], the authors formulated a constrained least squares problem and solved for a weight vector which will maintain a flat response over a spatial region of interest and, at the same time, will remain in the noise subspace. In this chapter, we present a subspace-based technique for estimating DOA's of spatially close source signals. The sources are assumed to lie within a known small sector. We define the array manifold space as the subspace spanned by all the vectors in the array manifold within the known sector. By using the Karhunen-Loève expansion, we show that the array manifold subspace over a limited sector of interest can be well approximated by a subspace with dimension smaller than the number of sensors in the array. The technique has a sound geometric interpretation within the framework of constrained subspace fitting in the least squares sense. This technique searches for the optimal signal subspace over the array manifold subspace which has a minimum distance to the array data signal subspace in the sense of a Frobenius norm. The source DOA parameters are then determined from the spatial spectrum formed from the estimated signal subspace. We examine the quality of the estimates and demonstrate their consistency. Both theoretical analysis and numerical simulation results show that the proposed technique is always superior to the MUSIC algorithm in the sense that the estimated signal subspace is closer in distance to the true signal subspace. Numerical studies also show that the performance of the proposed technique is comparable to that of the DMLE and the MD-MUSIC algorithm. However, since it only involves one dimensional optimization, the computation involved is drastically reduced.

4.2 Array signal model

Consider an array of M omni-directional sensors. Assume that K incoherent narrow-band far-field source signals are incident on the array. The analytical array signal model is described by (2.11) as

$$\underline{x}(n) = A(\Theta)\underline{s}(n) + \underline{w}(n). \tag{4.1}$$

where the array noise vector $\underline{w}(n)$ is uncorrelated with the source signals. The covariance of $\underline{w}(n)$ is denoted by $\sigma_w^2 I$, where σ_w^2 is the unknown variance and I denotes the identity matrix. Denote the K-dimensional signal subspace as E_S and the $(M-K)$-dimensional noise subspace as E_N. E_S and E_N are orthogonal. The eigen-values of R_x, the covariance matrix of $\underline{x}(t)$, are known to be distributed as

$$\lambda_1 \leq \ldots \lambda_K \leq \lambda_{K+1} = \ldots = \lambda_M = \sigma_w^2.$$

Let $\underline{u}_i$ denote the eigen-vector associated with λ_i. Let $U_S = [\underline{u}_1, \ldots, \underline{u}_K]$ and $U_N = [\underline{u}_{K+1}, \ldots, \underline{u}_M]$. Then, the range spaces of U_S and U_I, are identical respectively to the signal subspace E_S and the noise subspace E_N. Since the signal subspace and the noise subspace are orthogonal, the steering vectors $\{\underline{a}(\theta_k);\ k = 1, \ldots, K\}$ are orthogonal to the noise subspace. Let $\hat{\underline{u}}_i$ denote the estimate of $\underline{u}_i$ from the array data. Let $\hat{U}_S$ and $\hat{U}_N$ be the estimates of U_S and U_N, respectively. Define

$$\hat{E}_S = \text{range space of } \hat{U}_S$$

as the array data signal subspace and

$$\hat{E}_N = \text{range space of } \hat{U}_N$$

as the array data noise subspace. Subspaces $\hat{E}_S$ and $\hat{E}_N$ are orthogonal complements and are estimates of E_s and E_n, respectively.

4.3 Geometric interpretations of different algorithms

In this section, we formulate the MUSIC, the DML and the MD-MUSIC algorithms in a unified framework of constrained subspace fitting in the least squares sense. This framework is useful for designing numerical algorithms and obtaining new techniques.

4.3.1 Constrained subspace fitting in the least squares sense

Consider a constrained subspace fitting problem in which we search a for a K-dimensional subspace S_X in the constraint subspace S_C which is closest to a pre-specified subspace C_X in the least squares error sense

$$\hat{S}_X = \arg \min_{S_C, T_X} \| Q_C - Q_S T_X \|_F^2, \quad S_X \subset S_C. \tag{4.2}$$

where the columns of Q_S and Q_C are each a set of independent vectors from subspace S_X and C_X and both matrices are of size $M \times K$. T_X is an arbitrary $K \times K$ matrix. Fixing Q_S and minimizing the Frobenius norm in (4.2) with respect to R yields the pseudo-inverse solution [29]

$$\hat{T}_X = Q_S^{\dagger} Q_C = (Q_S^H Q_S)^{-1} Q_S^H Q_C. \tag{4.3}$$

where $Q_S^{\dagger}$ denotes the pseudo-inverse of Q_S. Substituting (4.3) back into (4.2), we get

$$\hat{S}_X = \arg \max_{S_X} tr\{P_{S_X} Q_C Q_C^H\}, \quad S_X \subset S_C. \tag{4.4}$$

where $P_{S_C} = Q_S (Q_S^H Q_S)^{-1} Q_S^H$ is the projection operator onto C_S. In (4.4), it is seen that different constriants S_C will lead to different solution of $\hat{S}_X$. In the context of array processing, we replace C_S with the array data signal subspace $\hat{E}_S$ and write (4.4) as

$$\hat{S}_X = \arg \max_{S_X} tr\{P_{C_S} \hat{U}_S \hat{U}_S^H\}, \quad S_X \subset S_C. \tag{4.5}$$

4.3.2 The DML and the MUSIC algorithm

Let $X = [\underline{x}(1), \underline{x}(2), \ldots, \underline{x}(N)]$. The DML method [30][27] determines the source DOA parameters by (for details, see Chapter 2)

$$\hat{\Theta} = \arg \min_{\Theta} \| X - A(\Theta) S \|_F^2 . \tag{4.6}$$

which is equivalent to (4.5) by assigning $Q_{C_X} = X$, $Q_{S_X} = A(\Theta)$ and $T_S = S$ in (4.2). Equation (4.6) can also be written as

$$\hat{\Theta} = \arg\max_{\Theta} \sum_{i=1}^{N} \mid P_{A(\Theta)} \underline{x}(i) \mid^2 = \arg\max_{\Theta} tr\{P_{A(\Theta)} \hat{R}_x\}. \tag{4.7}$$

The DMLE has a clear geometric interpretation. From the array signal model (2.11), we see that the array data vector $\underline{x}(i)$ abides within the K-dimensional signal subspace spanned by the columns of $A(\Theta)$ in the absence of noise and is perturbed out of the signal subspace by the noise $\underline{w}(i)$. It follows that the DML estimates are obtained by searching for the K vectors over the array manifold which will form the K-dimensional subspace closest to $\{\underline{x}(i)\}$ in the least squares error sense. This closeness is measured by the sum of moduli of the projections of all the vectors $\{\underline{x}(i)\}$ onto this signal subspace as in (4.7).

The geometric interpretation of the MUSIC algorithm is as follows. In the absence of array sensor noise, the array data vectors are all from the signal subspace. If sensor noise is present, the array data vectors are perturbed out of the signal subspace. Thus, it is natural to estimate the K-dimensional subspace M_S closest to the vectors $\{\underline{x}(i)\}$ by minimizing their projections onto M_S. Thus,

$$\hat{M}_S = \arg\max_{M_S} \sum_{i=1}^{N} \mid P_M \underline{x}(i) \mid^2 = \arg\max_{M_S} tr\{P_M \hat{R}_x\}, \tag{4.8}$$

where P_M denotes the orthogonal projection onto the M_S. Let the columns of U_S denote the orthonormal base vectors of M. Then, U_S is of dimension $M \times K$ and satisfies the following orthonormality condition

$$U_S^H U_S = I. \tag{4.9}$$

Since the orthogonal projection P_M can be written as $P_M = U_S U_s^H$, we rewrite (4.8) as

$$\hat{U}_S = \arg \max_{U_S^H U_S = I} tr\{U_S^H \hat{R}_x U_S\}, \tag{4.10}$$

in which the problem of finding an optimal subspace has been transformed into the problem of searching for a set of orthonormal base vectors. By the Rayleigh theorem [29], the orthogonal base vectors of the estimated signal subspace, $\hat{U}_S$, are identical to the estimated signal eigen-vectors of

$\hat{R}_x$. In other words, they span a subspace identical to that generated by the MUSIC algorithm. In examining the difference between the DMLE and the MUSIC algorithms, we find that in DMLE, the optimization of the signal subspace is limited by the structure of the array manifold, while in the MUSIC algorithm, the search for the signal subspace in the M-dimensional space is unconstrained.

4.3.3 MD-MUSIC : subspace fitting in the eigen-vector domain

Usually, the array data covariance matrix has to be estimated from the array data. Thus, the array data subspace is an estimate of the of the true signal subspace. The MD-MUSIC algorithm searches for the K vectors over the array manifold spanning a subspace closest to the array data signal subspace in the least squares sense. The MD-MUSIC estimate is given by

$$\hat{\Theta} = \arg\max_{\Theta} \sum_{i=1}^{K} \mid P_{A(\Theta)}\hat{u}_i \mid^2 = \arg\max_{\Theta} tr\{P_{A(\Theta)}\hat{U}_S\hat{U}_S^H\}, \tag{4.11}$$

in which the criterion function measures the distance between the range space of $A(\Theta)$ and the array data signal subspace. The MD-MUSIC algorithm provides better performance than the MUSIC algorithm and is capable of resolving coherent soure signals. Its estimate is superior to the DML estimate for coherent signals [20]. The MD-MUSIC algorithm and the DML method are closely related. To see this, we express the DML estimate using

$$\hat{\Theta} = \arg\max_{\Theta} \sum_{i=1}^{M} \hat{\lambda}_i \mid P_{A(\Theta)}\hat{u}_i \mid^2, \tag{4.12}$$

where $\hat{\lambda}_i$ denotes the ith eigen-value of $\hat{R}_x$. It shows that the DML estimate has included both the estimated signal and the estimated noise eigen-vectors under consideration and that each projection of the eigen-vector onto the signal subspace is weighted by its associated eigen-value. The eigen-vector associated with a larger eigen-value will contribute more to the criterion.

4.4 DOA estimation in the array manifold subspace

In the previous section, we discuss the DMLE, the MUSIC and the MD-MUSIC algorithms from the viewpoint of constrained subspace fitting. In this section, assuming that the source signals are

are spatially close, we propose an efficient estimation procedure which is naturally born out of the constrained subspace fitting framework.

4.4.1 Array manifold space in a sector

Define $\Omega_0 = \{\theta_{01} \leq \theta \leq \theta_{02}\}$, where θ_{01} and θ_{02} are two angle parameters. The array manifold over the sector Ω_0, $A_M = \{\underline{a}(\theta);\ \theta \in \Omega_0\}$, is defined as the set of all array steering vectors over sector Ω_0. The array manifold space S_M is defined as the space spanned by all the vectors in the array manifold A_M. Obviously, since the array steering vectors associated with the sources are from the array manifold A_M, the signal subspace E_S is a subspace within the array manifold space S_M. To observe the limited rank character of the array manifold in sector Ω_0, or the limited dimension characteristics of the array manifold space, consider the array manifold covariance matrix

$$R_\Theta = \int_{\Omega_0} \underline{a}(\theta)\underline{a}^H(\theta)d\theta. \tag{4.13}$$

Denote γ_i and $\underline{\psi}_i$, $i = 1, 2, \ldots, M$ as the eigen-values and eigen-vectors of R_Θ, respectively. Assume that $\{\gamma_i;\ \ i = 1, 2, \ldots, M\}$ are ordered in a monotonically nonincreasing fashion. From the Karhunen-Loève (KL) expansion theory [26], we know that the eigen-value γ_i of R_Θ indicates the distribution of the variance of projection of $\underline{a}(\theta)$ onto the corresponding eigenvector $\underline{\psi}_i$. It is known that the KL expansion is the most efficient expansion in the mean-square-error sense, i.e., when each vector in the array manifold is represented as a linear combination of the first d ordered eigen-vectors,

$$\underline{a}(\theta) = \sum_{i=1}^{d} \alpha_i(\theta)\psi_i,$$

where $\alpha_i(\theta)$ is the expansion coefficient with $\underline{\psi}_i$, the following truncation error is minimized

$$\epsilon(d) = \sum_{i=d+1}^{M} E\{|\ \alpha_i(\theta)\ |^2\} = \sum_{i=d+1}^{M} \gamma_i. \tag{4.14}$$

In the extreme case, when R_Θ has $(M-d)$ zero eigen-values, the d eigen-vectors associated with the nonzero eigen-values are sufficient to represent the array manifold without error and the subspace spanned by these eigen-vectors is identical to the array manifold space. If R_Θ has d principal eigen-values, then the array manifold space can be approximated by the the subspace spanned by the

Table 4.1: **The distribution of eigen-values of R_Θ and the effective dimension of the array manifold space, $\Omega_0 = \{-\theta_0 < \theta < \theta_0\}$**

θ_0	1	2	3	4	5	6	Effe. D	Truc. E
5°	0.9755	0.0708	0.0008	0.0000	0.0000	0.0000	3	0.0000
12°	5.5162	2.1319	0.1855	0.0043	0.0000	0.0000	3	0.0043
15°	1.9010	1.0673	0.1666	0.0066	0.0001	0.0000	4	0.0001
25°	2.0186	1.9180	1.0960	0.1940	0.0092	0.0001	5	0.0002
35°	2.0935	2.0509	1.9594	1.0594	0.1618	0.0054	6	0.0000
55°	2.4919	2.3654	2.0997	2.0583	1.9871	0.5166	6	0.0000

principal eigen-vectors with an error depending on the distribution of the remaining eigen-values. When all the sources are spatially close, i.e., the sector Ω_0 is relatively small, we find that the array manifold space can be well approximated by a subspace sapnned by the principal eigen-vectors of R_Θ with a dimension smaller than the number of sensors in the array. The following is an example to show how the array manifold space over a limited sector can be approximated by a subspace of smaller dimension. Consider a linear equispaced array with eight omni-directional sensors. Table. 4.1 shows the distribution of the eigen-values of R_Θ. The effective dimension of the array manifold is defined as the number of non-zero principal eigen-values.

In the table, we see that the eigen-value distribution varies with the size of sector selected. In general, the wider the sector we choose, the larger are the principal eigen-values. When Ω_0 is sufficiently wide, all the eigen-values of R_Θ become significant and the array manifold space spans the entire M-dimensional space. When the sector is limited, we see that some of the eigen-values are very small and insignificant compared with the principal eigen-values. It is appropriate to disregard these trivial eigen-values and use the remaining principal eigen-vectors to represent the array manifold space. The eigen-values to be eliminated are based on a preset threshold, which represents a percentage of the trace of R_Θ. The significant eigen-values are those above the threshhold, while those below the threshold are neglected.

4.4.2 Optimal signal subspace in the array manifold space

Assume that the source signals are spatially close and that the array manifold space can be approximated by a subspace of limited dimension. Since the signal subspace is a subspace in the array manifold space, it is natural that we let the constraint be the array manifold space in (4.5) and obtain the estimate as

$$\hat{S}_X = \arg\max_{S_X} tr\{P_{S_X}\hat{U}_S\hat{U}_S^H\}, \quad S_X \subset S_M \tag{4.15}$$

in which the estimated optimal signal subspace is obtained by searching the K-dimensional subspace over the array manifold space, closest to the array data signal subspace. The optimal signal subspace $\hat{S}_X$ can be interpreted as a K-dimensional subspace in S_M which has minimum principal angles with the array data signal subspace. The estimation procedure is summarized as follows.

- 1) Eigen-decompose the matrix R_Θ and determine the effective dimension $d(d > K)$. Let $\Psi_d = [\underline{\psi}_1, \underline{\psi}_2, \ldots, \underline{\psi}_d]$ in which the columns constitute a orthonormal basis of S_M. $\hat{U}_S$ is obtained from the eigen-decomposition of the array data covariance matrix $\hat{R}_x$, the columns of which are orthonormal basis of the array data signal subspace M.

- 2) Perform singular value decomposition (SVD) of $\hat{U}_S^H\Psi_d$. It follows that if

$$Y^H[\hat{U}_S^H\Psi_d]Z = diag(\sigma_1, \sigma_2, \ldots, \sigma_K), \tag{4.16}$$

  is the SVD of $\hat{U}_S^H\Psi_d$, the principal angle η_k and principal vector $\underline{v}_k$ and $\underline{z}_k$ are defined by

$$[\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_K] = \hat{U}_S Y$$

$$[\underline{z}_1, \underline{z}_2, \ldots, \underline{z}_d] = \Psi_d Z$$

$$\cos(\eta_k) = \sigma_k \quad k = 1, 2, \ldots, K$$

  and the followings hold

$$\underline{v}_i^H\underline{z}_j = \delta_{ij}\sigma_i, \text{ for } i \text{ and } j \leq K$$

$$\underline{v}_i^H\underline{z}_j = 0, \text{ for } i \text{ and } j > K,$$

- 3) The columns of $\hat{Q}_S = [\underline{z}_1, \ldots, \underline{z}_K]$ will be an orthonormal basis for the estimated signal subspace.

In the algorithm, the d-dimensional array manifold space is seen to be partitioned into two orthogonal complementary subspaces S_{M1} and S_{M2}, with S_{M1} having minimum principal angles with $\hat{E}_S$. Let S_M be represented as the direct sum of two arbitrary complementary subspaces S'_{M1} and S'_{M2}, $S_M = S'_{M1} \oplus S'_{M2}$. We have $P_{S_M} = P_{S'_{M1}} + P_{S'_{M2}}$, where $P_{S'_{M1}}$ and $P_{S'_{M2}}$ are orthogonal projections onto S'_{M1} and S'_{M2}, respectively, and P_{S_M} denotes the orthogonal projection onto S_M. We rewrite the criterion in (4.15) as

$$tr\{P_{S'_{M1}} \hat{U}_S \hat{U}_S^H\} = tr\{P_{S_M} \hat{U}_S \hat{U}_S^H\} - tr\{P_{S'_{M2}} \hat{U}_S \hat{U}_S^H\}. \tag{4.17}$$

Since the first term on the right side of (4.17), $tr\{P_{S_M} \hat{U}_S \hat{U}_S^H\}$, is a constant and $tr\{P_{S'_{M2}} \hat{U}_S \hat{U}_S^H\} \geq 0$, the criterion is maximized when $tr\{P_{S'_{M2}} \hat{U}_S \hat{U}_S^H\} = 0$ holds. This is possible only when $S'_{M2} = S_{M2}$, i.e., the estimated signal subspace is the maximizing argument of (4.15).

The DOA parameters can now be estimated by locating the peak positions of spectrum $P(\theta)$ constructed from the estimated signal subspace

$$P(\theta) = \underline{a}^H(\theta) \hat{Q}_S \hat{Q}_S^H \underline{a}(\theta), \quad \theta \in \Omega_0. \tag{4.18}$$

In the absence of noise, this measure will provide the exact parameter estimations. When the array data are perturbed by noise, the parameters obtained will be associated with the array steering vectors which are most nearly orthogonal to the estimated noise subspace.

4.4.3 Comments on the proposed algorithm

When Ω_0 is widened, the eigen-values of R_Θ tend to become non-trivial, and the effective dimension of the array manifold space approaches M. In this case, the optimal signal subspace will be identical to the array data signal subspace. Since the array data signal subspace is the same as the estimated signal subspace from the MUSIC algorithm, the proposed technique reduces to the conventional MUSIC algorithm. The algorithm is expected to perform better than the MUSIC algorithm for

spatially close sources. However, since the MUSIC algorithm does not require the source signals to be limited to a specific sector, it is fair to say that the improved performance over the MUSIC algorithm is extracted from this prior information about the source locations.

4.5 Performance analysis

In this section, we carry out the asymptotic performance analysis of the proposed technique by measuring the distance between the estimated signal subspace and the actual signal subspace, when the number of array samples is sufficiently large. We show that the estimated signal subspace is consistent and is closer to the signal subspace than is the array data signal subspace.

4.5.1 The consistency of the estimated signal subspace

It is known that the eigen-values and eigen-vectors of $\hat{R}_x$ are consistent [1] [27] and the following limits hold with probability one (w. p. 1) as $N \rightarrow \infty$,

$$\hat{\lambda}_i \rightarrow \lambda_i, \quad i = 1, 2, \ldots, M$$

$$\underline{\hat{u}}_i \rightarrow \underline{u}_i, \quad i = 1, 2, \ldots, K$$

To show the consistency of the estimated signal subspace, consider an equivalent algorithm which computes the normalized basis vectors of the estimated signal subspace successively. Let $V_i(\underline{q}) = \underline{\hat{u}}_i^H \underline{q}$. The basis vectors can also be obtained by

$$\underline{q}_i = \arg \max_{\underline{q}} V_i(\underline{q}), \quad i = 1, 2 \ldots, K$$

$$subject\ to:$$

$$\underline{q} \in S_M, \ \| \underline{q} \|_F = 1 \text{ and } \underline{q} \perp \underline{q}_k, \ k = 1, 2 \ldots, i-1$$

Since $\underline{\hat{u}}_i$ converges w.p.1 to $\underline{u}_i$, $V_i(\underline{q})$ converges w.p.1, uniformly in $\underline{q}$ to the limit function $\overline{V}_i(\underline{q}) = \underline{u}_i^H \underline{q}$ subject to the constraint $\| \underline{q} \|_F = 1$. Obviously, $\underline{q}_i$ converges w.p.1 to the maximizing argument of $\overline{V}_i(\underline{q})$. Since the true signal eigen-vectors $\underline{u}_i$ is from the the array manifold space, $\overline{V}_i(\underline{q})$ obtain its maximum if and only if $\underline{q} = \underline{u}_i$. Hence, $\underline{q}_i$ converges w.p.1. to the true signal eigen-vector $\underline{u}_i$.

4.5.2 The consistency of the DOA estimates

Let $\hat{\theta}$ be the DOA estimates determined from the spectral function $P(\theta)$. Define the limit function $\overline{P}(\theta)$ as

$$\overline{P}(\theta) = \underline{a}^H(\theta) U_S U_S^H \underline{a}(\theta). \tag{4.19}$$

Consider the following difference

$$\begin{aligned} \Delta &= \sup_{\theta} \mid \underline{a}^H(\theta) \hat{Q}_S \hat{Q}_S^H \underline{a}(\theta) - \underline{a}^H(\theta) U_S U_S^H \underline{a}(\theta) \mid \\ &\leq \| \underline{a}(\theta) \|_F^2 \times \| \hat{Q}_S \hat{Q}_S^H - U_S U_S^H \|_F \\ &= A_m \| \hat{Q}_S \hat{Q}_S^H - U_S U_S^H \|_F, \end{aligned} \tag{4.20}$$

where

$$A_m = \max_{\theta} \| \underline{a}(\theta) \|_F^2 . \tag{4.21}$$

Since $\underline{q}_i$ converges w.p.1 to $\underline{u}_i$, the difference Δ will tend to zero w.p.1. This indicates that the spectral function $P(\theta)$ converges w.p.1, uniformly in θ to its limit function $\overline{P}(\theta)$. Consequently, the DOA estimates $\hat{\theta}$ converges w.p.1 to the maximizing arguments of $\overline{P}(\theta)$. Since U_S denotes the true signal eigen-vectors, the limit function will be maximized when $U_S = A(\theta)T$ holds, where T is an arbitrary full rank matrix of dimension $K \times K$. However, in view of *the uniqueness condition* (2.14), this is possible if and only if θ represents the true source DOA's. This discussion states that the estimated DOA's are strongly consistent and asymptotically unbiased.

4.5.3 The quality of the estimated signal subspace

Due to finite observation time of the array data, the estimated signal subspaces $\hat{S}_X$ and $\hat{E}_S$, will not be identical to the true signal subspace E_S. We use the closeness between the estimated and true signal subspaces to evaluate the quality of the estimates. This closeness is measured by the distance between the subspaces. An advantage of using the subspace distance as a performance

criterion in the subspace based algorithms is that it includes all the source signals, array geometry and other array parameters under consideration and is independent of the search methods used. The distance between subspaces $\hat{S}_X$ and E_S is defined as [29]

$$dist(\hat{S}_X, E_S) = \| P_{\hat{S}_X} - P_{E_S} \|_F, \tag{4.22}$$

where $P_{\hat{S}_X}$ and P_{E_S} are the orthogonal projection operator onto subspace $\hat{S}_X$ and E_S, repectively. The distance between subspace $\hat{E}_S$ and E_S is simply $dist(\hat{E}_S, E_S) = \| P_{\hat{E}_S} - P_{E_S} \|_F$. Consider the following difference

$$\Delta d = \| P_{\hat{S}_X} - P_{E_S} \|_F^2 - \| P_{\hat{E}_S} - P_{E_S} \|_F^2 . \tag{4.23}$$

From the definition of Frobenius norm, Δd can be written as

$$\begin{aligned} \Delta d &= \| P_{\hat{S}_X} - P_{E_S} \|_F^2 - \| P_{\hat{E}_S} - P_{E_S} \|_F^2 \\ &= tr\{P_{E_S}(P_{\hat{E}_S} - P_{\hat{S}_X})\}. \end{aligned} \tag{4.24}$$

Since $\hat{S}_X$ is a subspace in S_M, we can always partition S_M as the orthogonal direct sum $S_M = \hat{S}_X \oplus S_W$. Denote P_{S_W} as the orthogonal projection onto subspace S_W. We write (4.24) as

$$\begin{aligned} \Delta d &= tr\{P_{E_S}(P_{\hat{E}_S} + P_{S_W} - P_{\hat{S}_X} - P_{S_W})\} \\ &= tr\{P_{E_S}(P_{\hat{E}_S} + P_{S_W} - P_{S_M})\}, \end{aligned} \tag{4.25}$$

where $P_{S_M} = P_{S_X} + P_{S_W}$ is the orthogonal projection onto the array manifold subspace. Since E_S is a subspace of S_M, we have $tr\{P_{E_S} P_{S_M}\} = K$ and (4.25) becomes

$$\Delta d = -K + tr\{P_{E_S}(P_{\hat{E}_S} + P_{S_W})\} = -K + tr\{P_{E_S} P_{C_W})\}. \tag{4.26}$$

where $C_W = \hat{E}_S \ominus S_W$ is a subspace of dimension d and P_{C_W} is the projection operator onto subspace C_W. Since $d > K$, we can partitioned C_W into the direct sum of C_{W1} and C_{W2}, where

C_{W1} is a K-dimensional subspace and C_{W2} is a $(d-K)$-dimensional subspace orthogonal to E_S. It follows that

$$\Delta d = -K + tr\{P_{E_S} P_{C_{W1}}\}. \tag{4.27}$$

Since the second term in (4.27) reaches its maximum value K only when C_{W1} is identical to E, we have the property of $\Delta d \leq 0$, or equivalently,

$$\| P_{\hat{S}_X} - P_{E_S} \|_F \leq \| P_{\hat{E}_S} - P_{E_S} \|_F . \tag{4.28}$$

This inequality indicates that the estimated signal subspace $\hat{S}_X$ is always closer to the true signal subspace E_S than is the array data signal subspace $\hat{E}_S$.

4.6 Numerical simulation results

Computer simulation results are provided to demonstrate the effectiveness of the proposed technique relative to the MUSIC, the DML and the MD-MUSIC estimators. An equi-spaced linear array of six sensors is simulated, with half the source wavelength spacings, and all sensors are assumed to be omni-directional with unit gain. Two narrow-band source signals of equal power impinge from $-6°$ and $6°$ to the normal of the array. Additive sensor noise is assumed to be a spatially white Gaussian process with zero-mean. The selected sector of interest is $\Omega_0 = \{-12^0 \leq \theta \leq 12^o\}$. The distribution of eigen-values of R_Θ is shown in Table.4.1. The effective dimension of the array manifold subspace in Ω_0 is 3 and the mean squares error (MSE) due to truncation is 0.0043 which is 0.055 of the sum of variances. The MSE's of the first DOA estimate are calculated both for uncorrelated and correlated signals. To obtain averaged results, we run 200 independent trials for each point. In each trial, 100 independent snapshots are used. In the simulation, the electrical angle $\phi = \pi \sin\theta$ is used. Denote the DOA parameter vector $\underline{\Phi} = [\phi_1, \ldots, \phi_K]$, and the Cramér-Rao Bound is computed according to (3.19). In all the results, the value of signal-to-noise ratio (SNR) is taken from −6dB to 14dB with a 4dB step.

UNCORRELATED SOURCES Fig.4.1 shows the MSE's of the first DOA estimate versus SNR

for the MUSIC algorithm and the proposed method compared to the square root of the Cramér-Rao bound. The proposed method coincides with the Cramér-Rao bound at high SNR and shows much better performance than the conventional MUSIC method under lower SNR. Fig.4.2 shows the MSE's of the first DOA estimate versus SNR for the proposed method compared to the MD-MUSIC and the MLE methods. The proposed technique is seen to perform almost as well as the MD-MUSIC and the DML estimators.

CORRELATED SOURCES Fig.4.3 and Fig.4.4 compare the MSE's of the first DOA estimate of the proposed method to the MUSIC, the MD-MUSIC, the DMLE and the Cramér-Rao bound when the source signals are correlated. The source signal used has a covariance matrix R_S of the form

$$R_s = E\{\underline{s}(n)\underline{s}^H(n)\} = \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix}, \tag{4.29}$$

where α is the correlation coefficient between the two source signals. The correlation coefficient α is chosen is 0.7 and 0.8, respectively, in Fig.4.3 and Fig.4.4. From the simulation results, the proposed method provides better performance than the MUSIC algorithm at low SNRs. It has the same performance as the MD-MUSIC algorithm and is just slightly outperformed by the DMLE.

4.7 Conclusions

A subspace-based algorithm for estimating DOA's of spatially close sources has been presented. It has been shown both theoretically and numerically that this proposed estimator is consistent and is always superior to the MUSIC estimator. It is slightly outperformed by the DMLE and the MD-MUSIC algorithm. However, since it involves only one-dimensional optimization, the computational burden involved is much less than that of the DMLE and the MD-MUSIC algorithm. Although this technique relies on prior knowledge regarding the extent of the sector in which the source signals are to be found, this is not an unreasonable assumption in many practical applications.

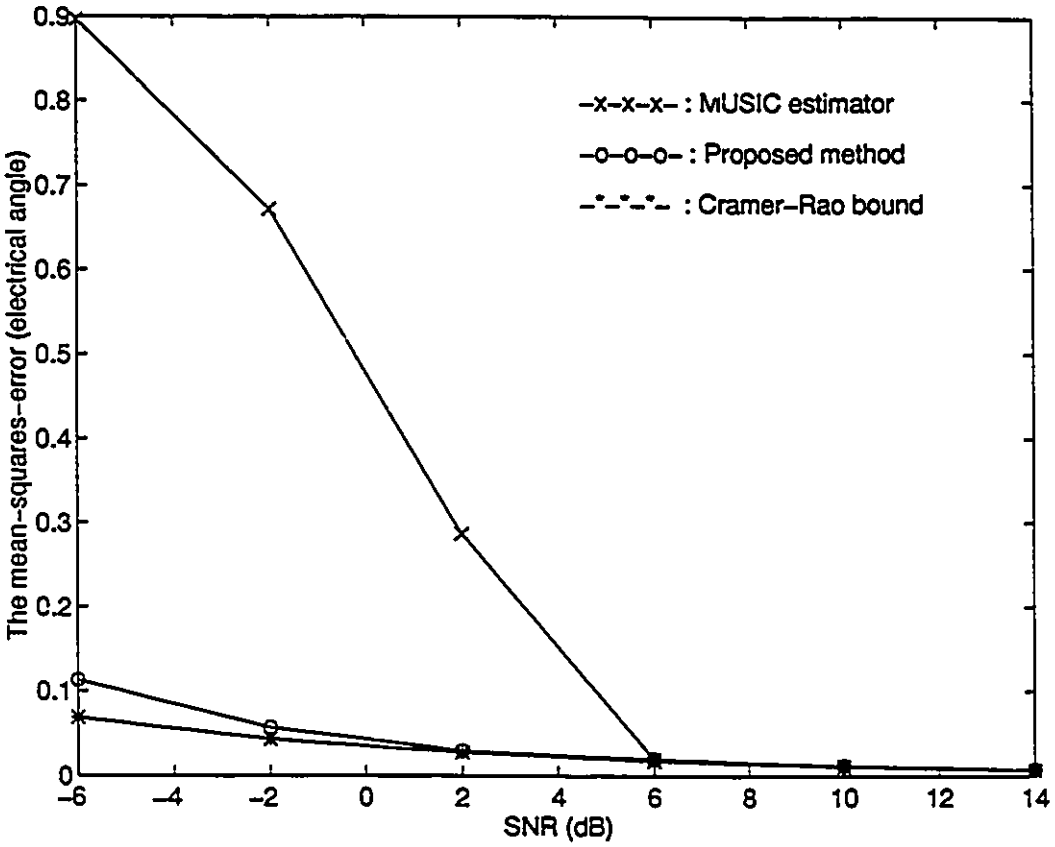

Figure 4.1: The variation of the MSE of the first DOA estimate versus SNR : uncorrelated case

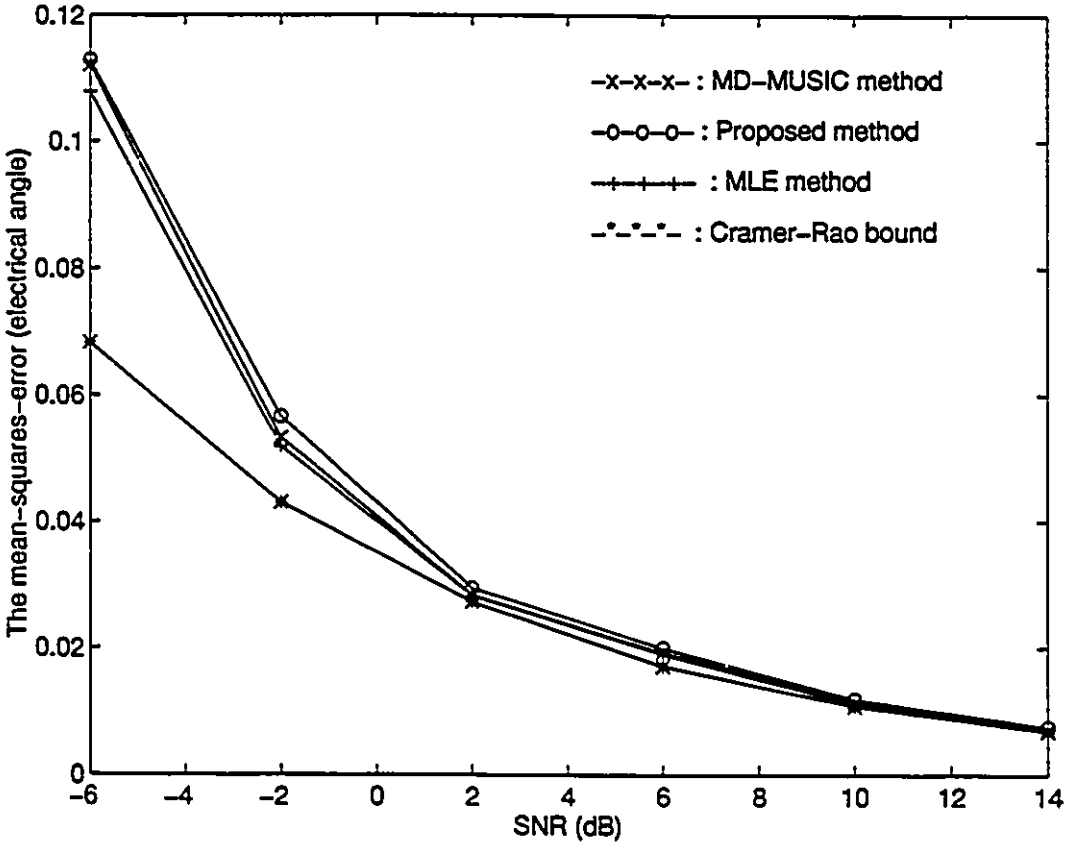

Figure 4.2: The variation of the MSE of the first DOA estimate versus SNR : uncorrelated case

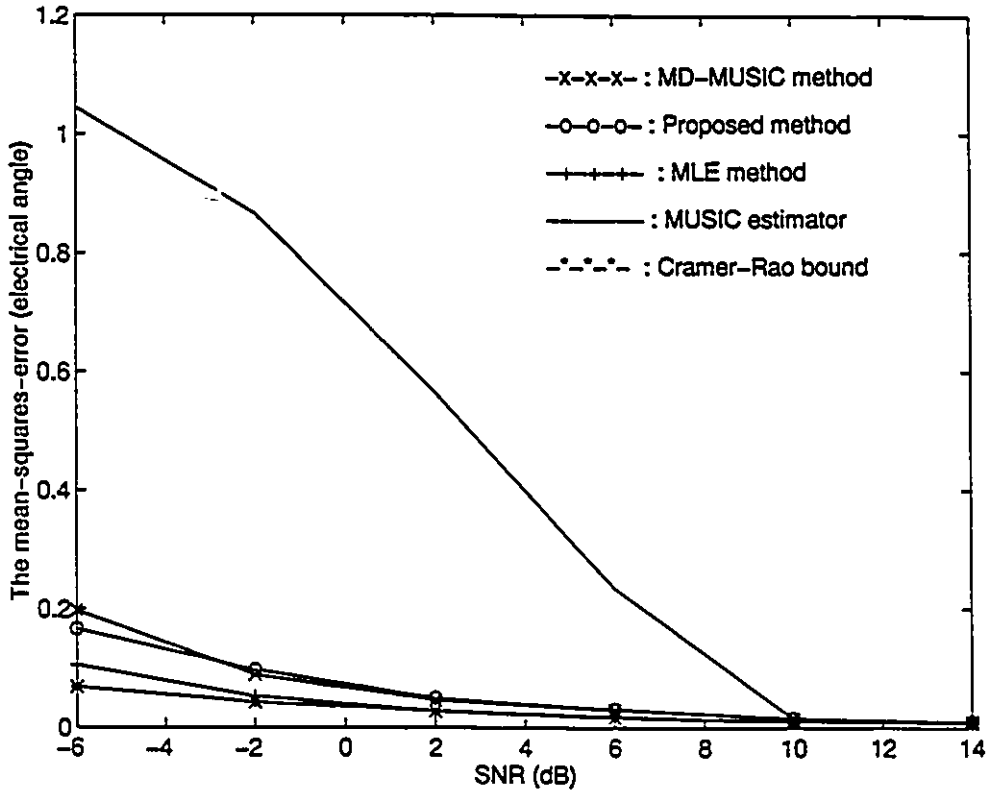

Figure 4.3: The variation of the MSE of the first DOA estimate versus SNR : correlated case with $\alpha = 0.70$

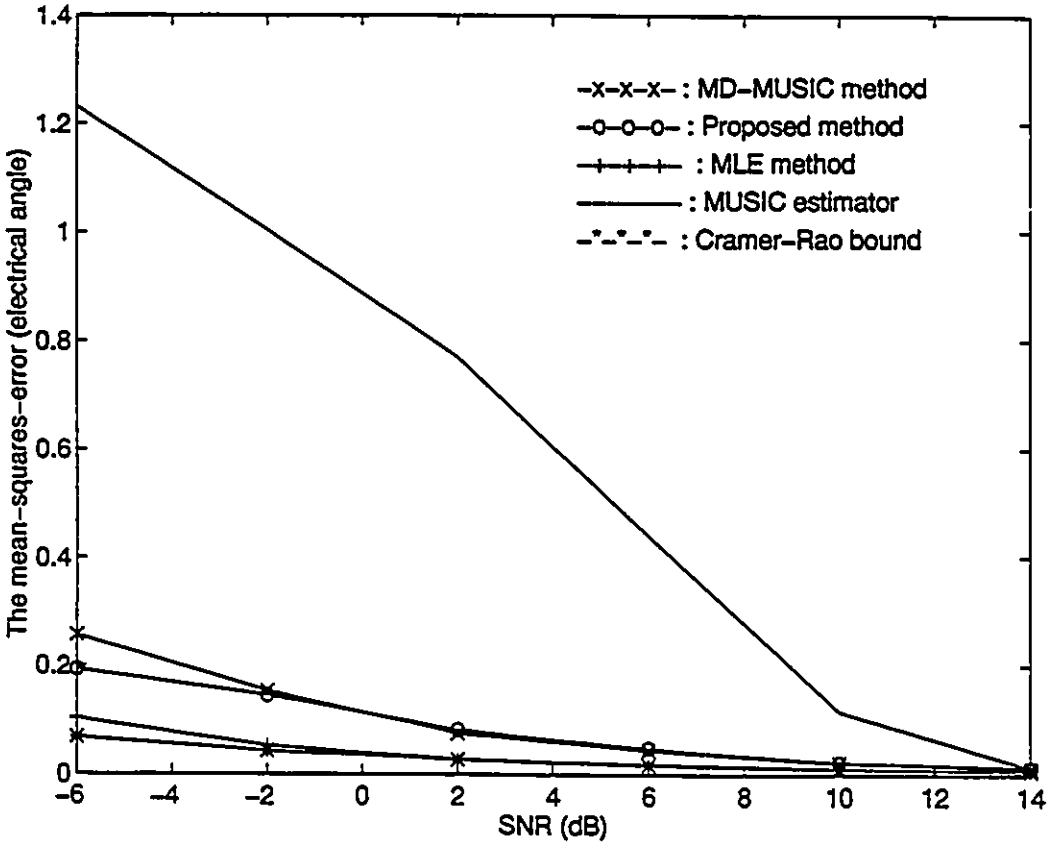

Figure 4.4: The variation of the MSE of the first DOA estimate versus SNR : correlated case with $\alpha = 0.85$

Chapter 5

Maximum Likelihood Estimator for Tracking the DOA's of Multiple Moving Targets

Most high resolution DOA estimation techniques rely on the assumption that the sources are stationary. However, in many communication system applications, estimation and tracking of DOA's of moving targets are of more interest. For example, in personal communication systems, the signal sources are usually satellites moving in orbit or people talking in a moving car. Therefore, the conventional high resloution techniques are not directly applicable, and it is necessary to develop effective estimation techniques to track the source target DOA's accurately. In this chapter, by assuming a locally linear motion model for the moving targets, we present a highly efficient maximum likelihood (ML) tracking algorithm for multiple moving targets by a passive array.

5.1 Introduction

There have been developed many DOA estimation techniques, such as the beamforming method, subspace based techniques and the nonlinear criterion based approaches including the DMLE and SMLE [1][6][10][31][30]. However, these techniques usually fail or have degraded performance in

the presence of moving targets. The tracking of the DOA's of moving targets is of great interest for communications, air traffic control, tactical and strategic defense operations. Take the application of the Applebaum adaptive array [64] in personal communication systems as an example. The Applebaum array is an adaptive array which effectively rejects the undesired signals from other directions and/or spatially spreads the interfering signal with optimum SNR, by forming the optimum beam pattern according to the environment. The Applebaum array requires the desired source target DOA's, which can be generated by a tracking algorithm.

Conventional high resolution techniques exploit the underlying array signal model by temporal integration. When applied to a moving source targets, their performance deteriorates. Generally speaking, target motion will spread the array spatial spectrum and result in degraded performance and poor resolution [33]. The performance worsens as the number of sensors increases. Recently, several techniques have been developed for estimating tracking slowing varying parameters [33]-[38]. A general procedure in these newly developed techniques is to assume that the source parameters are stationary during a limited integration time interval. Kumaresan *et al.* [34] proposed an approach which can instantaneously track the frequency and amplitude variation of multiple components signal by using certain nonlinear operators on the signal samples. In [36], [37] and [38], several tracking algorithms using the sensor outputs have been introduced. These algorithms are focused on the data association problems of the estimated parameters of moving targets, and the estimates are still obtained based on a stationary assumption over a limited subinterval of observation. This class of tracking algorithms does not require knowledge of the target motion dynamics, but, its performance is far from satisfactory due to the inaccurate parameter estimates from each observation sub-interval, especially when there are fast-moving source targets. An alternative is to make the subinterval sufficiently small to provide good stationary approximation. But this will result in a tremendous computational burden in the very short time subinterval and make real-time implemention difficult. Champaigne [35] exploits the nonstationary nature of the array data for slowly varying sources and presents a systematic approach to the general problem of optimum space-time array processing. However, his method can be used only when the array data can be approximated

as stationary over time intervals on the order of the correlation time of the array processor. Another newly developed algorithm is based on the so-called subspace tracking techniques [39] [40] [41]. These algorithms perform adaptively, and the signal/noise subspace estimates are updated each time when a new array data sample comes in. However, due to the lack of knowledge of the moving source parameters, the spread spatial spectrum problem still exists in the updating process, and the success of these techniques is limited.

There are also the standard approaches to multiple target tracking, such as the extended Kalman filter (EKF), the batch filter approach for deterministic systems, the maximum likelihood estimator and the Baysian estimator as in [42]. These methods assume that the dynamical model for the moving targets is known and estimate the target trajectories parameters from noisy measurements provided by the observations. However, these approaches are not appropriate for array processing. In most array applications, the source waveforms are unknown, and the array model cannot be directly used in the standard tracking techniques such as the Kalman filtering process. If we extract the source parameter estimates from small segments of the array data and then apply the tracking techniques, the spread spatial spectrum problem arises again. Also, Kalman filter requires the statistical information about the estimates which is usually not available in practice.

In the following chapters, a maximum likelihood estimator is provided for tracking the DOA's of multiple moving targets by a passive array. The source targets are assumed to be moving smoothly, and the target dynamics is modeled as locally linear with unknown local constant speed. The MTS (multiple target state) variable, including the target DOA state, is defined to decribe the moving source target. By applying the Sontag Theorem [43], we show that the locally linear dynamics is strongly locally observable except under some extreme conditions. The source MTS estimates are obtained by optimizing the likelihood criterion function based on a batch of array data. Since the source MTS's are related through the target motion dynamics, the optimization is performed with respect to the initial MTS, and the subsequent MTS estimates are predicted from the target dynamics equations. A modified Gauss-Newton algorithm is proposed for optimizing the ML criterion function, in which the Hessian is approximated by a positive semi-definite matrix to guarantee that algorithm is descent. Asymptotic performance analysis is carried out, and it is shown that the MTS

estimates are consistent. We also derive the asymptotic mean-squared errors and the Cramér-Rao bounds analytically for the maximum likelihood MTS estimates. The ML estimates are shown to be relatively asymptotically efficient. In the numerical study part, numerical simulation results are presented to show the effectiveness of the algorithm, and comparisons with theoretical conclusions are also provided. In the proposed ML tracking algorithm, since target motion dynamics has been incorperated, the spread spatial spectrum effects are eliminated completely. Also the subsequent MTS estimates are predicted from the initial MTS estimate. So they are automatically associated through their underlying dynamics and there is no further requirement for data association. Once new array data arrive, we form a new batch of data. The estimates from the previous set of data can be used to start the ML tracking algorithm, and the DOA's of the moving sources are expected to be tracked accurately.

5.2 Target motion dynamics and the array model

Assume that the sensors and source targets are coplanar. For a moving source target, its trajectory in the x-y plane is modeled by

$$\begin{bmatrix} \dot{x}(t) \\ \dot{y}(t) \end{bmatrix} = \begin{bmatrix} f_x[x(t)] \\ f_y[y(t)] \end{bmatrix}, \tag{5.1}$$

in the Cartesian co-ordinate system, where $\{x(t), y(t)\}$ denotes the source target position in the $x-y$ plane, and $\dot{x}(t)$ and $\dot{y}(t)$ are the derivatives of $x(t)$ and $y(t)$ with respect to t. Functions $f_x(t)$ and $f_y(t)$ are nonlinear in general. In (5.1), complete decoupling between the x and y components is assumed. This decoupling has been shown to be a sufficiently accurate model in target tracking techniques [42]. We assume that, in the neighbourhood of time t_0, $t \in [t_0, t_0 + \Delta T]$, the target trajectory (5.1) is locally linear so that

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} \nu_x[x(t_0)](t-t_0) \\ \nu_y[y(t_0)](t-t_0) \end{bmatrix}, \tag{5.2}$$

where

$$\nu_x(t_0) = f_x[x(t_0)] \quad \text{and} \quad \nu_y(t_0) = f_y[y(t_0)] \tag{5.3}$$

denote the x and the y components of the velocity of the target at time t_0. The approximation errors can be controlled by selecting an appropriately small finite observation time interval. Note that (5.3) is a first-order approximation of the target trajectory, or a piece-wise linear approximation of the trajectory. By contrast, we can verify that the slowly moving target model which has been used by most tracking techniques is a zero-order approximation of the target trajectory.

5.2.1 Trajectory equations of the source targets

Consider K linearly moving targets in the far-field of the array. For far-field source targets, the array data are directly associated with the source target DOA's. The kth target location at time index n is described in the coordinate system $\{r_k(n), \theta_k(n), \phi_k(n)\}$ rather than in the Cartesian coordinate system. We use $r_k(n)$ to denote the range, $\theta_k(n)$ to denote the target DOA and $\phi_k(n)$ to denote the target's direction of heading. All target parameters are measured with respect to a common reference. The array and source target geometry is shown in Fig. 5.1. Let T be the sampling period and let index n represent the sampling time nT. The first array sensor position is used as the system origin. The source DOA's are measured to the y-axis, the range is measured relative to the system origin and the direction of heading is measured against the x-axis. By the sine law, we can write for the kth source target

$$\begin{cases} q_k(n+1) = \dfrac{\nu_k}{r_k(n+1)} = \dfrac{\sin[\theta_k(n+1) - \theta_k(n)]}{T\cos[\theta_k(n) + \phi_k]} \\ q_k(n) = \dfrac{\nu_k}{r_k(n)} = \dfrac{\sin[\theta_k(n+1) - \theta_k(n)]}{T\cos[\theta_k(n+1) + \phi_k]} \end{cases} \tag{5.4}$$

and

$$\begin{cases} q_k(n+2) = \dfrac{\nu_k}{r_k(n+2)} = \dfrac{\sin[\theta_k(n+2) - \theta_k(n)]}{2T\cos[\theta_k(n) + \phi_k]} \\ q_k(n) = \dfrac{\nu_k}{r_k(n)} = \dfrac{\sin[\theta_k(n+2) - \theta_k(n)]}{2T\cos[\theta_k(n+2) + \phi_k]} \end{cases} . \tag{5.5}$$

where ν_k denotes the speed of the kth source target and

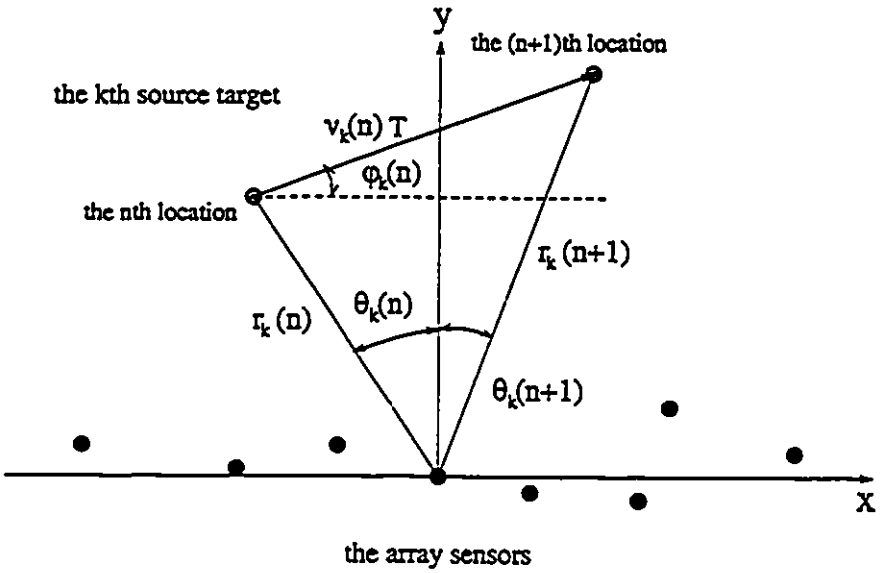

Figure 5.1: The array geometry for the linear moving source targets.

$$\phi_k(i) = \arctan\left\{\frac{v_x}{v_y}\right\}, \tag{5.6}$$

denotes the target's heading direction. In (5.4) and (5.5), we note that $r_k(n)$ and ν_k cannot be solved separately. Thus, for linearly moving target, its speed ν_k and range $r_k(n)$ cannot be resolved from passive array measurements. However, this will not affect the formulation as far as the source DOA tracking is concerned. Define the kth target state vector as

$$\underline{\alpha}_k(n) = [\theta_k(n), q_k(n), \phi_k(n)]^T. \tag{5.7}$$

By the target trajectory equations (5.4) and (5.5), we have

$$\underline{\alpha}_k(n \pm 1) = \underline{F}^{\pm}[\underline{\alpha}_k(n)], \tag{5.8}$$

where $\underline{F}^{\pm}[\underline{\alpha}_k(n)]$ is a column vector given by

$$\underline{F}^{\pm}[\underline{\alpha}_k(n)] = \begin{bmatrix} \arctan\left\{\dfrac{\sin[\theta_k(n)] \pm Tq_k(n)\cos[\phi_k(n)]}{\cos[\theta_k(n)] \pm q_k(n)T\sin[\phi_k(n)]}\right\} \\ \dfrac{q_k(n)}{\sqrt{1 \pm 2Tq_k(n)\sin[\theta_k(n)+\phi_k(n)] + q_k^2(n)T^2}} \\ \phi_k(n) \end{bmatrix}. \tag{5.9}$$

Equation(5.8) can be applied recursively m times and the result is the kth target states $\underline{\alpha}_k(n+m)$ or $\underline{\alpha}_k(n-m)$

$$\underline{\alpha}_k(n \pm m) = (\overbrace{\underline{F}^{\pm} \cdot \underline{F}^{\pm} \cdots \underline{F}^{\pm}}^{m})[\underline{\alpha}_k(n)] = \underline{F}^{\pm m}[\underline{\alpha}_k(n)], \tag{5.10}$$

where

$$\underline{F}^{\pm m}[\underline{\alpha}_k(n)] = \begin{bmatrix} \arctan\left\{\dfrac{\sin[\theta_k(n)] \pm mTq_k(n)\cos[\phi_k(n)]}{\cos[\theta_k(n)] \pm mTq_k(n)\sin[\phi_k(n)]}\right\} \\ \dfrac{q_k(n)}{\sqrt{1 \pm 2Tmq_k(n)\sin[\theta_k(n)+\phi_k(n)] + m^2q_k^2(n)T^2}} \\ \phi_k(n) \end{bmatrix}. \tag{5.11}$$

provided n, m and $n \pm m$ are bounded by $[0, N-1]$.

5.2.2 Array signal model

Consider an array consisting of $M(K < M)$ omni-directional sensors. Assume that the kth array sensor is located at $\underline{Z}_k = [z_{xk}, z_{yk}]^T$ in the xy plane. The medium is assumed to be isotropic and non-dispersive. The source taregts are assumed to be in the far-field of the array. For narrow-band source signals, each sensor output is the sum of the shifted versions of the original source signals,

$$x_m(n) = \sum_{k=1}^{K} s_k[nT + \tau_{km}(n)] + w_m(n), \tag{5.12}$$

where $w_m(n)$ is the mth array sensor additive noise and s_k is the kth signal waveform. Parameter $\tau_{km}(n)$ is the relative time delay induced by the kth signal in the mth sensor at time nT with respect to the selected origin of the co-ordinate system. This time delay is determined by the array geometry and the source signal DOA's and is given by

$$\tau_{km}(n) = \frac{1}{c}\underline{Z}_m \cdot \underline{\kappa}_k(n) \qquad m = 1, 2, \ldots, M, \tag{5.13}$$

where $\underline{\kappa}_k(n) = [\cos\theta_k(n), \sin\theta_k(n)]^T$ is defined as the kth direction vector. Using analytic representation, the array signal model can be described in matrix form

$$\underline{x}(n) = A[\Theta(n)]\underline{s}(n) + \underline{w}(n), \tag{5.14}$$

where $\Theta(n)$ denotes the source target DOA parameters at nT, and $A[\Theta(n)]$ is defined as the array composite steering matrix at time instant nT. For simplicity, we will use $A(n)$ to denote $A[\Theta(n)]$. The kth column of $A(n)$, $\underline{a}[\theta_k(n)]$, is the array steering vector associated with the kth signal at time nT and is given by

$$\underline{a}[\theta_k(n)] = [1, \exp\{j\omega_0\tau_{k2}(n)\}, \ldots, \exp\{j\omega_0\tau_{kM}(n)\}]^T, \tag{5.15}$$

where ω_0 is the source signal frequency. Note that, since the time delays are referred to the first sensor, $\tau_{k1} = 0$. In previous chapters, we introduced the signal subspace. Similarly, we define the signal subspace at time instant nT as the column space of $A(n)$. The signal subspace at time nT depends on the array geometry, time instant and the moving target trajectories. The unambiguity condition can also be extended to the array tracking model. The array is assumed to be unambiguous at n, i.e., $A(n)$ has full rank. The *uniqueness condition* can be stated as

$$A[\Theta_1(n)]T_1 = A[\Theta_2(n)]T_2 \Longrightarrow \Theta_1(n) = \Theta_2(n), \tag{5.16}$$

where T_1 and T_2 are any full rank matrices of size $K \times K$. The details of the uniqueness condition are in [32].

The source signals are assumed to be unknown but deterministic. In communication applications, the signal waveforms are often far from being the Gaussian random processes which have been assumed in some literature [10] [15]. The signal waveform $s_k(n)$ is regarded as an unknown parameter to be estimated. The additive sensor noise $\underline{w}(n)$ is assumed to be jointly Gaussian, zero mean with covariance matrix $\sigma_w^2 I$, where σ_w^2 is the unknown noise covariance and I denotes

identity matrix. $\{\underline{w}(n);\ n = 0, 1, \ldots, N-1\}$, where N is the number of samples, are assumed to be statistically independent.

5.2.3 Observability of locally linear target dynamics

The locally linear dynamics of the kth target are described by (5.9) and (5.11). The question arises as to whether the states can be determined from the observed array data uniquely. This is related to both the observability of the target dynamics and the ambiguity of the array signal model. The ambiguity problem has been solved by imposing the uniqueness condition on the array model. Let's discuss the observability of the locally linear model. For far-field source targets, only the source target DOA states are directly associated with the array data, and the target DOA states can be written as a mapping from the MTS vector space to the DOA parameter domain as

$$\theta_k(n) = \mathcal{F}[\underline{\alpha}_k(n)] = [1, 0, 0]\underline{\alpha}_k(n), \tag{5.17}$$

where $\theta_k(n) \in R^1$, $\underline{\alpha}(n) \in R^3$ and $\mathcal{F}$ denotes the mapping operator, $R^3 \rightarrow R^1$, from the MTS state space to the DOA parameter domain. Since we are primarily concerned with the theoretical aspects of observability, noise-free measurements are considered. Also, because each source target dynamic forms an independent system, it is sufficient to consider a single target. Henceforth, we shall omit the subscript k in this section.

The system observability refers to the ability to reconstruct the state space from the measurement space. Here, in the context of source DOA tracking, the measurement space means specifically the source DOA parameter domain. In a geometric sense, system observability is a functional relationship between the measurement space and the state space. When applied to the moving target dynamics here, the definition of observability can be stated as follows [44].

Definition 1 A state $\underline{\alpha}(n) \in R^3$ is said to be observable if knowledge about measurement sequence $\{\theta(n+m);\ m = 0, 1, \ldots, M_\alpha - 1\}$ is sufficient to determine the system state $\underline{\alpha}(n)$ for a finite integer M_α. If state $\underline{\alpha}(n)$ is observable everywhere on its defined state space, then, the system is said to be completely observable.

Generally speaking, if a system is observable, the corresponding mapping from the state space to the measurement space should be one-to-one, and the state space can be reconstructed from the measurement space. If the system is not observable, sufficient measurement data may be obtained to ascertain the state of interest over one specific domain but, not over another domain. In [45], several notions of observability have been introduced for discrete dynamical systems. A system is *strongly observable* or *finite time observable* at $\underline{\alpha}(n)$ if M_α equals the dimension of the system state space. The system is *strongly observable* if it is strongly observable at every point of the system state space. A system is *strongly locally observable* at $\underline{\alpha}(n)$ if there exists a neighbourhood of $\underline{\alpha}(n)$, $U_{\underline{\alpha}(m)}$, in which $\underline{\alpha}(n)$ is strongly observable. The system is *strongly locally observable* if every point in the state space is strongly locally observable.

The strong local observability is an appropriate criterion for testing the observability of the locally linear model. Besides, strong local observability has the advantage of lending itself to a simple algebraic test. Define a map T_α given by

$$T_\alpha[\underline{\alpha}(m)] = \{\mathcal{F}[\underline{\alpha}(m)], \mathcal{F}[\underline{\alpha}(m+1)], \mathcal{F}[\underline{\alpha}(m+2)]\} = [\theta(m), \theta(m+1), \theta(m+2)]. \tag{5.18}$$

The system is said to satisfy the *observability of rank condition* at $\underline{\alpha}(m)$ if the rank of T_α is 3, and the system is said to satisfy the *observability rank distribution* if this is true at every $\underline{\alpha}(n)$ in the system state space. Let's consider an autonomous dynamical system. An autonomous dynamical system is defined by the following state equations

$$\dot{x} = f(x); \;\; x(t_0) = x_0, \tag{5.19}$$

in which mapping f, also referred to as the vector fieid of the system, is independent of time t. Obviously, the locally linear target dynamics described by (5.8) represent an autonomous system. For testing the observability of an autonomous discrete nonlinear system, we introduce the following theorem by Sontag [43].

Theorem 1 An autonomous discrete nonlinear system is strongly locally observable at $\underline{\alpha}(n)$ if it satisfies the observability rank of condition at $\underline{\alpha}(n)$. If the system satisfies the observability rank

distribution, then the system is strongly locally observable.

It has been shown [46] that the rank of map T_α at $\underline{\alpha}(n)$ is the rank of the Jacobian matrix of the local representation of T_α using local co-ordinates. From (5.18), we obtain the Jacobian matrix $T_\alpha[\underline{\alpha}(m)]$ at $\underline{\alpha}(m)$ as

$$\nabla T_\alpha = \begin{bmatrix} 1 & \frac{\partial\theta(m+1)}{\partial\theta(m)} & \frac{\partial\theta(m+2)}{\partial\theta(m)} \\ 0 & \frac{\partial\theta(m+1)}{\partial q(m)} & \frac{\partial\theta(m+2)}{\partial q(m)} \\ 0 & \frac{\partial\theta(m+1)}{\partial\phi(m)} & \frac{\partial\theta(m+2)}{\partial\phi(m)} \end{bmatrix}. \tag{5.20}$$

In Appendix A, we show that the Jacobian ∇T_α has full rank, and the mapping T_α satisfies the observability rank condition almost everywhere except for an extreme case in which

$$\theta(m) = \pi/2 - \phi(m) \text{ or } \theta(m) = -\pi/2 - \phi(m). \tag{5.21}$$

This condition happens when the source target DOA and its direction of heading are parallel or anti-parallel, i.e., the target moves along the line connecting the target and origin of the co-ordinate system. In this case, consecutive DOA measurements will be identical, and the locally linear target dynamics have a rank deficient Jacobian. The MTS vectors cannot be completely reconstructed from the source target DOA parameter measeurments due to the non-observability of the system. For such a scenario, the locally linear target dynamics (5.8) reduce to

$$\underline{\alpha}(n \pm 1) = \begin{bmatrix} \theta(n \pm 1) \\ q(n \pm 1) \\ \phi(n \pm 1) \end{bmatrix} = \begin{bmatrix} \theta(n) \\ q(n) \pm 1/T \\ \phi(n) \end{bmatrix}. \tag{5.22}$$

in which the source target DOA compoent $\theta(n)$ is no longer related to the other MTS components. Since the source DOA parameters are the only MTS components which are directly observable from the array data, we can see that the information on the q and the ϕ components are totally lost when condition (5.21) prevails. This non-observability reduces the tracking algorithm to the conventional

estimation algorithm for stationary source targets since targets moving with their DOA equal to their heading directionwill appear as stationary targets to the passive array. A possible way to overcome this rank deficiency of the system Jacobian is to move the co-ordinate origin. If the rank deficient Jacobian is detected, we can relocate the co-ordinate origin so that the target DOA and direction of heading in the new co-ordinate system, $\theta'(m)$ and $\phi'(m)$, will satisfy

$$\theta'(m) \neq \pi/2 - \phi'(m) \text{ and } \theta'(m) \neq -\pi/2 - \phi'(m), \tag{5.23}$$

and the transformed locally linear target dynamics will become strongly locally observable in the new co-ordinate system.

The strong local observability of the locally linear target dynamics (5.8) indicates that the source target MTS vectors can be uniquely determined from the source DOA measurements. Since the identification of the source target DOA parameters from the array data is guaranteed by the array uniqueness condition, it is concluded that the locally linear target states can be uniquely determined by the passive array data.

5.3 The maximum likelihood estimator

The maximum likelihood method is a popular estimation procedure in statistical studies. It works with the likelihood function of the array data and computes the unknown parameters as the maximizing arguments of the likelihood function. The ML estimator are consistent and asymptotically efficient [47]. In this section, we will investigate the possibility of applying the ML approach to passive array tracking.

5.3.1 Array data likelihood function

Under the Gaussian assumption of the array sensor noise, the conditional density function of a batch of N sampled array data is given by

$$p(\underline{x}(1), \underline{x}(2), \ldots, \underline{x}(N)) = \prod_{i=1}^{N} \frac{1}{\pi^M \sigma_w^{2M}} \exp\left\{-\frac{1}{\sigma_w^2}\{\underline{x}(i) - A(i)\underline{s}(i)\}^H\{\underline{x}(i) - A(i)\underline{s}(i)\}\right\}, \tag{5.24}$$

and the negative log likelihood function, ignoring the constant term, has the following form

$$L = -\log p = MN \log(\pi\sigma_w^2) + \frac{1}{\sigma_w^2} \sum_{i=1}^{N} \| \underline{x}(i) - A(i)\underline{s}(i) \|_F^2 . \tag{5.25}$$

5.3.2 Optimizing the likelihood function

The principle of the maximum likelihood estimator is to maximize the likelihood function with respect to the unknown parameters [14], and the maximum likelihood estimates are obtained as the minimizing arguments of (5.25). Fixing $\Theta(i)$ and $\underline{s}(i)$ and minimizing the negative log likelihood function with respect to σ_w^2 yields the noise variance estimate

$$\hat{\sigma}_w^2 = \frac{1}{MN} \sum_{i=1}^{N} \| \underline{x}(i) - A(i)\underline{s}(i) \|_F^2 . \tag{5.26}$$

Substitute $\hat{\sigma}_w^2$ back into (5.25) and obtain the array signal vector estimate $\hat{\underline{s}}(i)$ and the DOA estimate $\hat{\Theta}(i)$ as

$$[\hat{\underline{s}}(i), \hat{\Theta}(i)] = \arg \min_{\underline{s}(i), \Theta(i)} \sum_{i=1}^{N} \| \underline{x}(i) - A(i)\underline{s}(i) \|_F^2, \; i = 1, 2, \ldots, N. \tag{5.27}$$

Since the above criterion function in (5.27) is quadratic in the array signal vector $\underline{s}(i)$, fixing $\Theta(i)$ and minimizing with respect to $\underline{s}(i)$ yields the signal waveform estimate

$$\hat{\underline{s}}(i) = \{A^H(i)A(i)\}^{-1} A^H(i)\underline{x}(i). \tag{5.28}$$

Substituting (5.28) back into (5.27) we obtain the criterion function for the DOA estimate

$$\begin{aligned} J' &= \sum_{i=1}^{N} \| \underline{x}(i) - P_{A(i)}\underline{x}(i) \|_F^2 \\ &= \sum_{i=1}^{N} \| P_{A(i)}^{\perp}\underline{x}(i) \|_F^2 \\ &= \sum_{i=1}^{N} tr\{P_{A(i)}^{\perp}\underline{x}(i)\underline{x}^H(i)\} \\ &= \sum_{i=1}^{N} tr\{P_{A(i)}^{\perp}\hat{R}_x(i)\} = \sum_{i=1}^{N} J_i. \end{aligned} \tag{5.29}$$

where

$$P_{A(i)} = A(i)\{A^H(i)A(i)\}^{-1}A^H(i), \tag{5.30}$$

is the projection operator onto the column space of $A(i)$ and

$$P^{\perp}_{A(i)} = I - P_{A(i)} \tag{5.31}$$

denotes the projection operator onto the null space of $A(i)$. The ith component J_i is given by

$$J_i = tr\{P^{\perp}_{A(i)}\hat{R}_x(i)\}, \tag{5.32}$$

and notation $\hat{R}_x(i)$ represents the one-sample estimate of the array data covariance matrix at time iT and is computed as

$$\hat{R}_x(i) = \underline{x}(i)\underline{x}^H(i), \tag{5.33}$$

which is a rank one matrix. Since $\hat{R}_x(i)$ is positive semi-definite Hermitian, J_i is nonnegative. We see that J' is not well defined since when N approaches infinity, J' is not bounded. We divide J' by N and obtain an equivalent likelihood function J given by

$$J = \frac{1}{N}J' = \frac{1}{N}\sum_{i=1}^{N} tr\{P^{\perp}_{A(i)}\hat{R}_x(i)\}. \tag{5.34}$$

which keeps all the properties of the the likelihood function J' [47]. Then, the maximum likelihood estimate of the DOA state parameter $\Theta(i)$ is obtained by

$$\hat{\Theta}(i) = \arg\min_{\Theta(i)} J = \arg\min_{\Theta(i)} \frac{1}{N}\sum_{i=1}^{N} tr\{P^{\perp}_{A(i)}\hat{R}_x(i)\},\ i = 1, 2, \ldots, N. \tag{5.35}$$

Chapter 6

An Iterative Algorithm for Optimizing the Likelihood Function

The source target MTS vectors, including the source DOA parameters can be estimated by optimizing the criterion function J. This optimization process is non-linear in general. The Gauss-Newton type method is one of the most efficient optimization methods and is known for its superior convergence rate. In this chapter, a modified Gauss-Newton algorithm is provided. We derive a close form expression for the gradient of J and approximate the Hessian by a positive semi-definite matrix to guarantee that the algorithm is descent.

6.1 An iterative optimization algorithm

The criterion function J is a function of the source DOA states $\{\Theta(i);\ i = 1, 2, \ldots, N\}$ which are in turn related to the state variables $\{q_k(i), \phi_k(i);\ i = 1, 2, \ldots, N,\ k = 1, 2, \ldots, K\}$ by the source target dynamics. Define the $3K \times 1$ multiple target state (MTS) vector at time index i as

$$\begin{aligned} \underline{\alpha}(i) &= [\underline{\alpha}_1^T(i), \underline{\alpha}_2^T(i), \ldots, \underline{\alpha}_K^T(i)]^T \\ &= [\theta_1(i), q_1(i), \phi_1(i), \theta_2(i), q_2(i), \phi_2(i), \ldots, \theta_K(i), q_K(i), \phi_K(i)]^T. \end{aligned} \tag{6.1}$$

Since J is an implicit function of the MTS vectors and one MTS vector can be determined

completely by another through the locally linear target dynamics, we need only to estimate the initial MTS vector $\underline{\alpha}(1)$. The subsequent MTS vectors, $\{\underline{\alpha}(i);\ i = 2,3,\ldots,N\}$ can be predicted from the assumed target dynamics.

6.1.1 The iterative minimization algorithm

The initial ML estimate of $\underline{\alpha}(1)$ can be obtained as the minimizing argument of J. Analytical solutions are, in general, not available, and we have to resort to numerical optimization techniques. Recently, several optimization methods have been applied to array signal processing, including the alternating projection (AP) method [30], expected maximization (EM) algorithm [48] [49], global search techniques [50]-[52] and different Newton-type algorithms [53]-[55]. It is well-known that the Newton-type method is one of the most efficient optimization methods. It gives, locally, a quadratic convergence to the optimizing arguments [56].

When a Newton-type algorithm is applied to the optimization problem (5.35), the $(k+1)$th iteration for locating the initial MTS vector $\underline{\alpha}(1)$ is computed as

$$\hat{\underline{\alpha}}^{(k+1)}(1) = \hat{\underline{\alpha}}^{(k)}(1) - \mu_k H^{-1} G, \tag{6.2}$$

where μ_k is the kth iteration step-length, H represents the Hessian matrix of the criterion function with respect to the initial state $\underline{\alpha}(1)$ and G is the gradient. Both the Hessian and the gradient are evaluated at $\hat{\underline{\alpha}}(1)$. The essential steps in deriving a Newton-type algorithm involve finding the gradient and the Hessian.

6.1.2 Computation of gradient G

The criterion J is an explicit function of the DOA state $\Theta(i)$ and an implicit function of the initial MTS vector $\underline{\alpha}(1)$. The gradient of the criterion J with respect to $\underline{\alpha}(1)$ can be written as

$$G = \frac{\partial J}{\partial \underline{\alpha}(1)} = \begin{bmatrix} \frac{1}{N}\sum_{i=1}^{N} \frac{\partial J_i}{\partial \theta_1(1)} \\ \frac{1}{N}\sum_{i=1}^{N} \frac{\partial J_i}{\partial q_1(1)} \\ \frac{1}{N}\sum_{i=1}^{N} \frac{\partial J_i}{\partial \phi_1(1)} \\ \vdots \\ \frac{1}{N}\sum_{i=1}^{N} \frac{\partial J_i}{\partial \theta_K(1)} \\ \frac{1}{N}\sum_{i=1}^{N} \frac{\partial J_i}{\partial q_K(1)} \\ \frac{1}{N}\sum_{i=1}^{N} \frac{\partial J_i}{\partial \phi_K(1)} \end{bmatrix} = \begin{bmatrix} \frac{1}{N}\sum_{i=1}^{N} \frac{\partial J_i}{\partial \theta_1(i)}\frac{\partial \theta_1(i)}{\partial \theta_1(1)} \\ \frac{1}{N}\sum_{i=1}^{N} \frac{\partial J_i}{\partial \theta_1(i)}\frac{\partial \theta_1(i)}{\partial q_1(1)} \\ \frac{1}{N}\sum_{i=1}^{N} \frac{\partial J_i}{\partial \theta_1(i)}\frac{\partial \theta_1(i)}{\partial \phi_1(1)} \\ \vdots \\ \frac{1}{N}\sum_{i=1}^{N} \frac{\partial J_i}{\partial \theta_K(i)}\frac{\partial \theta_K(i)}{\partial \theta_K(1)} \\ \frac{1}{N}\sum_{i=1}^{N} \frac{\partial J_i}{\partial \theta_K(i)}\frac{\partial \theta_K(i)}{\partial q_K(1)} \\ \frac{1}{N}\sum_{i=1}^{N} \frac{\partial J_i}{\partial \theta_K(i)}\frac{\partial \theta_K(i)}{\partial \phi_K(1)} \end{bmatrix}, \tag{6.3}$$

Let ∇J_i be the gradient of J_i with respect to the DOA state $\Theta(i)$ given by

$$\nabla J_i = [\frac{\partial J_i}{\partial \theta_1(i)}, \frac{\partial J_i}{\partial \theta_2(i)}, \ldots, \frac{\partial J_i}{\partial \theta_K(i)}]^T. \tag{6.4}$$

Let $r(i) = P^{\perp}_{A(i)}\underline{x}(i)$. Write J_i as $J_i = \| r(i) \|^2_F$ and the mth element of ∇J_i can be written as

$$\frac{\partial J_i}{\partial \theta_m(i)} = 2Re\left\{\left(\frac{\partial r(i)}{\partial \theta_m(i)}\right)^H r(i)\right\} = 2Re\left\{r^H_m(i) r(i)\right\}, \tag{6.5}$$

where $r_m(i) = \partial r(i)/\partial \theta_m(i)$. The first derivative of the projection matrix has been derived as [57][53]

$$\frac{\partial P^{\perp}_{A(i)}}{\partial \theta_m(i)} = -\frac{\partial P_{A(i)}}{\partial \theta_m(i)} = -P^{\perp}_{A(i)} A_m(i) A^{\dagger}(i) - [P^{\perp}_{A(i)} A_m(i) A^{\dagger}(i)]^H, \tag{6.6}$$

where $A_m(i) = \partial A(i)/\partial \theta_m(i)$ and for the parameterized array composite steering matrix $A(i)$,

$$A_m(i) = \left[0, \ldots, \frac{\partial \underline{a}[\theta_m(i)]}{\partial \theta_m(i)}, \ldots, 0\right]. \tag{6.7}$$

Then, (6.5) can be written as

$$\frac{\partial J_i}{\partial \theta_m(i)} = 2Re\left\{\left(\frac{\partial r(i)}{\partial \theta_m(i)}\right)^H r(i)\right\}$$

$$
\begin{aligned}
&= 2Re\left\{\underline{x}^H(i)\left[\frac{\partial P^{\perp}_{A(i)}}{\partial\theta_m(i)}\right]^H P^{\perp}_{A(i)}\underline{x}(i)\right\} \\
&= -2Re\left\{\underline{x}^H(i)P^{\perp}_{A(i)}A_m(i)A^{\dagger}(i)\underline{x}(i)\right\} \\
&= -2Re\left\{tr[A^{\dagger}(i)\hat{R}_x(i)P^{\perp}_{A(i)}A_m(i)]\right\}, \qquad (6.8)
\end{aligned}
$$

in which we have used $A^{\dagger}(i)P^{\perp}_{A(i)} = 0$. The gradient ∇J_i is then

$$
\nabla J_i = -2Re\{diag[A^{\dagger}(i)\hat{R}_x(i)P^{\perp}_{A(i)}D(i)]\}, \qquad (6.9)
$$

where $diag(X)$ denotes a vector formed from the diagonal elements of matrix X, and $D(i)$ is defined as

$$
D(i) = \left[\frac{\partial\underline{a}[\theta_1(i)]}{\partial\theta_1(i)}, \frac{\partial\underline{a}[\theta_2(i)]}{\partial\theta_2(i)}, \ldots, \frac{\partial\underline{a}[\theta_K(i)]}{\partial\theta_K(i)}\right] = \sum_{m=1}^{K} A_m(i). \qquad (6.10)
$$

By (5.11), the kth target DOA state $\theta_k(i)$ is a non-linear function of the MTS vector $\underline{\alpha}_k(i)$. Denote $\underline{v}_k(i)$ as the gradient of $\theta_k(i)$ with respect to $\underline{\alpha}_k(1)$. It can be computed from (5.11) using elementary mathematics as

$$
\underline{v}_k(i) = \begin{bmatrix} \frac{\partial\theta_k(i)}{\partial\theta_k(1)} \\ \frac{\partial\theta_k(i)}{\partial q_k(1)} \\ \frac{\partial\theta_k(i)}{\partial\phi_k(1)} \end{bmatrix} = \begin{bmatrix} \frac{1+(i-1)Tq_k(1)\sin[\theta_k(1)+\phi_k(1)]}{1+(i-1)^2T^2q_k^2(1)+2(i-1)Tq_k(1)\sin[\theta_k(1)+\phi_k(1)]} \\ \frac{(i-1)T\cos[\theta_k(1)+\phi_k(1)]}{1+(i-1)^2T^2q_k^2(1)+2(i-1)Tq_k(1)\sin[\theta_k(1)+\phi_k(1)]} \\ \frac{-(i-1)^2q_k^2(1)T^2-(i-1)Tq_k(1)\sin[\theta_k(1)+\phi_k(1)]}{1+(i-1)^2T^2q_k^2(1)+2(i-1)Tq_k(1)\sin[\theta_k(1)+\phi_k(1)]} \end{bmatrix}. \qquad (6.11)
$$

Define the matrix $V(i)$ as

$$
V(i) = [\underline{v}_1(i), \underline{v}_2(i), \ldots, \underline{v}_K(i)]. \qquad (6.12)
$$

and the gradient G can finally be obtained as

$$
G = vec\{\frac{1}{N}\sum_{i=1}^{N} V(i)diag(\nabla J_i)\} = -2vec\left\{\frac{1}{N}\sum_{i=1}^{N} V(i)Re\{I\odot[A^{\dagger}(i)\hat{R}_x(i)P^{\perp}_{A(i)}D(i)]\}\right\}, \qquad (6.13)
$$

where *vec* stands for the concatenation of the columns of a matrix.

6.1.3 Computing the Hessian matrix of the criterion J

The Hessian matrix of the criterion function J with respect to $\underline{\alpha}(1)$ can be partitioned as a $K \times K$ matrix of 3×3 blocks

$$H = [\frac{\partial^2 J}{\partial \alpha_i(1) \partial \alpha_j(1)}]_{3K \times 3K} = \begin{bmatrix} H_{11} & H_{12} & \dots & H_{1K} \\ H_{21} & H_{22} & \dots & \dots \\ \vdots & & & \\ H_{K1} & \dots & \dots & H_{KK} \end{bmatrix}, \tag{6.14}$$

where the 3×3 matrix H_{mn} is defined as

$$H_{mn} = \frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} \frac{\partial^2 J_i}{\partial \theta_m(1) \partial \theta_n(1)} & \frac{\partial^2 J_i}{\partial \theta_m(1) \partial q_n(1)} & \frac{\partial^2 J_i}{\partial \theta_m(1) \partial \phi_n(1)} \\ \frac{\partial^2 J_i}{\partial q_m(1) \partial \theta_n(1)} & \frac{\partial^2 J_i}{\partial q_m(1) \partial q_n(1)} & \frac{\partial^2 J_i}{\partial q_m(1) \partial \phi_n(1)} \\ \frac{\partial^2 J_i}{\partial \phi_m(1) \partial \theta_n(1)} & \frac{\partial^2 J_i}{\partial \phi_m(1) \partial q_n(1)} & \frac{\partial^2 J_i}{\partial \phi_m(1) \partial \phi_n(1)} \end{bmatrix}. \tag{6.15}$$

The diagonal elements of the submatrix H_{mn} are computed as

$$\begin{aligned} \frac{\partial^2 J_i}{\partial \theta_m(1) \partial \theta_n(1)} &= \frac{\partial}{\partial \theta_m(1)} \cdot \left[\frac{\partial J_i}{\partial \theta_n(i)} \cdot \frac{\partial \theta_n(i)}{\partial \theta_n(1)} \right] \\ &= \frac{\partial^2 J_i}{\partial \theta_m(i) \partial \theta_n(i)} \left[\frac{\partial \theta_m(i)}{\partial \theta_m(1)} \cdot \frac{\partial \theta_n(i)}{\partial \theta_n(1)} \right] + \frac{\partial J_i}{\partial \theta_n(i)} \cdot \frac{\partial^2 \theta_n(i)}{\partial \theta_m(1) \partial \theta_n(1)} \delta_{mn} \\ \frac{\partial^2 J_i}{\partial q_m(1) \partial q_n(1)} &= \frac{\partial}{\partial q_m(1)} \cdot \left[\frac{\partial J_i}{\partial \theta_n(i)} \cdot \frac{\partial \theta_n(i)}{\partial q_n(1)} \right] \\ &= \frac{\partial^2 J_i}{\partial \theta_m(i) \partial \theta_n(i)} \left[\frac{\partial \theta_m(i)}{\partial q_m(1)} \cdot \frac{\partial \theta_n(i)}{\partial q_n(1)} \right] + \frac{\partial J_i}{\partial \theta_n(i)} \cdot \frac{\partial^2 \theta_n(i)}{\partial q_m(1) \partial q_n(1)} \delta_{mn} \\ \frac{\partial^2 J_i}{\partial \phi_m(1) \partial \phi_n(1)} &= \frac{\partial}{\partial \phi_m(1)} \cdot \left[\frac{\partial J_i}{\partial \theta_n(i)} \cdot \frac{\partial \theta_n(i)}{\partial \phi_n(1)} \right] \\ &= \frac{\partial^2 J_i}{\partial \theta_m(i) \partial \theta_n(i)} \left[\frac{\partial \theta_m(i)}{\partial \phi_m(1)} \frac{\partial \theta_n(i)}{\partial \phi_n(1)} \right] + \frac{\partial J_i}{\partial \theta_n(i)} \cdot \frac{\partial^2 \theta_n(i)}{\partial \phi_m(1) \partial \phi_n(1)} \delta_{mn} \end{aligned} \tag{6.16}$$

and the off-diagonal elements are computed as

$$\frac{\partial^2 J_i}{\partial \theta_m(1) \partial q_n(1)} = \frac{\partial}{\partial q_n(1)} \cdot \left[\frac{\partial J_i}{\partial \theta_m(i)} \cdot \frac{\partial \theta_m(i)}{\partial \theta_m(1)} \right]$$

$$
\begin{aligned}
&= \frac{\partial^2 J_i}{\partial\theta_m(i)\partial\theta_n(i)}\left[\frac{\partial\theta_n(i)}{\partial q_n(1)}\cdot\frac{\partial\theta_m(i)}{\partial\theta_m(1)}\right]+\frac{\partial J_i}{\partial\theta_m(i)}\cdot\frac{\partial^2\theta_m(i)}{\partial\theta_m(1)\partial q_n(1)}\delta_{mn}\\
\frac{\partial^2 J_i}{\partial\theta_m(1)\partial\phi_n(1)} &= \frac{\partial}{\partial\phi_n(1)}\cdot\left[\frac{\partial J_i}{\partial\theta_m(i)}\cdot\frac{\partial\theta_m(i)}{\partial\theta_m(1)}\right]\\
&= \frac{\partial^2 J_i}{\partial\theta_m(i)\partial\theta_n(i)}\left[\frac{\partial\theta_n(i)}{\partial\phi_n(1)}\cdot\frac{\partial\theta_m(i)}{\partial\theta_m(1)}\right]+\frac{\partial J_i}{\partial\theta_m(i)}\cdot\frac{\partial^2\theta_m(i)}{\partial\theta_m(1)\partial\phi_n(1)}\delta_{mn}\\
\frac{\partial^2 J_i}{\partial q_m(1)\partial\phi_n(1)} &= \frac{\partial}{\partial\phi_m(1)}\cdot\left[\frac{\partial J_i}{\partial\theta_m(i)}\cdot\frac{\partial\theta_m(i)}{\partial q_m(1)}\right]\\
&= \frac{\partial^2 J_i}{\partial\theta_m(i)\partial\theta_n(i)}\left[\frac{\partial\theta_n(i)}{\partial\phi_n(1)}\cdot\frac{\partial\theta_m(i)}{\partial q_m(1)}\right]+\frac{\partial J_i}{\partial\theta_m(i)}\cdot\frac{\partial^2\theta_m(i)}{\partial q_m(1)\partial\phi_n(1)}\delta_{mn}, \quad (6.17)
\end{aligned}
$$

where δ_{mn} is the Kronecker delta defined as

$$
\delta_{mn}=\begin{cases}1 & \text{if } m=n\\ 0 & \text{if } m\neq n\end{cases}. \quad (6.18)
$$

Written in a compact form, the submatrix H_{mn} is

$$
H_{mn}=\frac{1}{N}\sum_{i=1}^{N}\frac{\partial^2 J_i}{\partial\theta_m(i)\partial\theta_n(i)}\underline{v}_m(i)\underline{v}_n^H(i)+\frac{1}{N}\sum_{i=1}^{N}\frac{\partial J_i}{\partial\theta_m(i)}\Lambda_m(i)\delta_{mn},\quad m=1,2,\ldots,K. \quad (6.19)
$$

where $\underline{v}_m(i)$ is defined in (6.11), and $\Lambda_m(i)$ is the Hessian matrix of the mth target DOA state $\theta_m(i)$ with respect to the mth initial target state $\underline{\alpha}_m(1)$ given by

$$
\Lambda_m(i)=\begin{bmatrix}
\frac{\partial^2\theta_m(i)}{\partial\theta_m(1)\partial\theta_m(1)} & \frac{\partial^2\theta_m(i)}{\partial\theta_m(1)\partial q_m(1)} & \frac{\partial^2\theta_m(i)}{\partial\theta_m(1)\partial\phi_m(1)}\\
\frac{\partial^2\theta_m(i)}{\partial\theta_m(1)\partial q_m(1)} & \frac{\partial^2\theta_m(i)}{\partial q_m(1)\partial q_m(1)} & \frac{\partial^2\theta_m(i)}{\partial q_m(1)\partial\phi_m(1)}\\
\frac{\partial^2\theta_m(i)}{\partial\phi_m(1)\partial\theta_m(1)} & \frac{\partial^2\theta_m(i)}{\partial\phi_m(1)\partial q_m(1)} & \frac{\partial^2\theta_m(i)}{\partial\phi_m(1)\partial\phi_m(1)}
\end{bmatrix}. \quad (6.20)
$$

Denote the Hessian matrix of J_i with respect to the ith DOA state $\Theta(i)$ by $\nabla^2 J_i$ where

$$
\nabla^2 J_i=\left[\frac{\partial^2 J_i}{\partial\theta_m(i)\partial\theta_n(i)}\right]_{K\times K}. \quad (6.21)
$$

The Hessian matrix H can then be expressed as

$$H = \frac{1}{N}\sum_{i=1}^{N}\left\{[\nabla^2 J_i \otimes \mathbf{O}] \odot [vecV(i)vec^H V(i)] + [diag(\nabla J_i) \otimes I] \odot \Lambda(i)\right\}, \tag{6.22}$$

where $\otimes$ denotes the Kronecker product, and $\Lambda(i) = diag[\Lambda_1(i), \Lambda_2(i), \ldots, \Lambda_K(i)]$. Matrix $\mathbf{O}$ is an all one matrix of dimension 3×3

$$\mathbf{O} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}. \tag{6.23}$$

In (6.22), the Hessian $\nabla^2 J_i$ is calculated by differentiating (6.5) with respect to $\theta_n(i)$

$$\frac{\partial^2 J_i}{\partial\theta_m(i)\theta_n(i)} = 2Re\left\{r_m^H(i)r_n(i) + r_{mn}^H(i)r(i)\right\}. \tag{6.24}$$

6.2 Modifying the algorithm by approximating the Hessian

The general idea behind the Newton method is to approximate the criterion function locally in the neighbourhood of the stationary point by a quadratic function. The role of the Hessian can be seen as a modification of the gradient. A negative gradient direction is a descent direction, and the Newton method is a descent method only when the Hessian is positive definite at each iteration [56]. In practice, this positivity cannot be guaranteed, and we usually need to modify the Hessian to ensure that it remains positive definite at every iteration. One of the most elegant ways was first proposed by Levenberg [58] and applied to the Newton method by Goldfeld [59]. A constant ϵ is added to all diagonal elements of the Hessian so that the estimate is iteratively calculated as

$$\underline{\hat{\alpha}}^{(k+1)}(1) = \underline{\hat{\alpha}}^{(k)}(1) - \mu_k(H + \epsilon I)^{-1}G, \tag{6.25}$$

where ϵ is chosen to make $[H_i(i) + \epsilon I]$ positive definite. This idea is very intuitive, but, the search process for parameter ϵ is computationally very costly. Another often used technique for overcoming the non-positive Hessian problem is to use a less complex positive semidefinite matrix to

approximate the Hessian. Below, the derivation of such a simplified positive semi-definite Hessian is presented.

Since the ML estimates $\{\hat{\Theta}(i);\ i = 1, 2, \ldots, N\}$ are each a consistent estimate, the kth iterate $\underline{\alpha}^{(k)}(1)$ is expected to be sufficiently close to the optimum value, and the first-order derivative of J_i with respect to $\Theta(i)$ would be very close to zero,

$$\frac{\partial J_i}{\partial \Theta(i)} \simeq 0. \tag{6.26}$$

Thus, the second term in (6.22) can be neglected and the Hessian is approximated by

$$H = \frac{1}{N}\sum_{i=1}^{N}[\nabla^2 J_i \otimes \mathbf{O}] \odot [vecV(i)vec^H V(i)]. \tag{6.27}$$

In [57], Golub and Pereyra proposed a variable projection method for separable non-linear least squares problems. In the variable projection method, the second derivative term is discarded under the assumption that the residual at the kth iteration is small. For sufficiently large number of array samples, with $\underline{\alpha}^{(k)}(1)$ close to its true value, $\| r \|_F$ is certainly small, and the term in (6.24) containing the residual r can be ignored. Thus, the Hessian (6.24) can be approximated by

$$\frac{\partial^2 J_i}{\partial \theta_m(i)\theta_n(i)} \simeq 2Re\left\{r_m^H(i)r_n(i)\right\}. \tag{6.28}$$

The approximated Hessian (6.28) can be verified to be positive semi-definite and is written as

$$\begin{aligned}\frac{\partial^2 J_i}{\partial \theta_m(i)\theta_n(i)} \simeq\ & 2Re\Big\{tr\{[A^{\dagger H}(i)A_m(i)P^{\perp}_{A(i)}A_n(i)A^{\dagger}(i) \\ & + P^{\perp}_{A(i)}A_m(i)A^{\dagger}(i)A^{\dagger H}(i)A_m^H(i)P^{\perp}_{A(i)}]\hat{R}_x(i)\}\Big\},\end{aligned} \tag{6.29}$$

in which we have used $A^{\dagger}(i)P^{\perp}_{A(i)} = P^{\perp}_{A(i)}A^{\dagger H}(i) = 0$. Kauffman [60] modified the Newton method by further dropping the second term in (6.29). A natural motivation for dropping this term is given in [53], and this is believed to be a better approximation of the Hessian. By the same motivation, we neglect the second term in (6.29) and approximate $\nabla^2 J_i$ as

$$\frac{\partial^2 J_i}{\partial \theta_m(i)\theta_n(i)} \simeq 2Re\{tr[A^{\dagger H}(i)A_m^H(i)P^{\perp}_{A(i)}A_n(i)A^{\dagger}(i)]\}. \tag{6.30}$$

Then in compact matrix form, we have

$$\nabla^2 J_i = 2Re\{[D^H(i)P^{\perp}_{A(i)}D(i)] \odot [A^{\dagger}(i)\hat{R}_x(i)A^{\dagger H}(i)]^T\}. \tag{6.31}$$

Since $D^H(i)P^{\perp}_{A(i)}D(i)$ and $A^{\dagger}(i)\hat{R}_x(i)A^{\dagger H}(i)$ are both positive semi-definite, $\nabla^2 J_i$ is positive semi-definite by the property that the Hardamard product of two positive semidefinite matrices is also positive semi-definite [61] [13]. The following lemma is given for the approximate Hessian H.

Lemma 1 The approximate Hessian matrix H in (6.27) of the criterion J with respect to $\underline{\alpha}(1)$ is positive semi-definite.

Proof The approximate Hessian matrix H has the form

$$H = \frac{1}{N}\sum_{i=1}^{N}[\nabla^2 J_i \otimes \mathbf{O}] \odot [vecV(i)vec^H V(i)], \tag{6.32}$$

where the approximation of $\nabla^2 J_i$ is given in (6.31). Since both $\mathbf{O}$ and the approximate $\nabla^2 J_i$ are positive semi-definite, $\nabla^2 J_i \otimes \mathbf{O}$ is positive semi-definite by the properties of Kronecker product.

The outer product $[vecV(i)vec^H V(i)]$ is rank one and is obviously positive semi-definite. Thus, each term in (6.32)

$$[\nabla^2 J_i \otimes \mathbf{O}] \odot [vecV(i)vec^H V(i)]$$

is positive semi-definite, and it follows that the sum of positive semi-definite matrices is also positive semi-definite. □

In addtion, the step length parameter μ_k can be incorporated to ensure that the Newton algorithm is descent. The parameter μ_k should be selected such that the convergence of the criterion to the local minimum will be guaranteed [56]. An often used scheme for the Newton method is to choose $\mu < 1$ and to take $\mu_k = \mu^i$ for the smallest integer $i \geq 0$ that causes sufficient decrease [56].

6.3 The initial estimate computation

The global convergence of the Gauss-Newton algorithm depends on the initial MTS estimate $\underline{\alpha}^{(1)}(1)$. This initial estimate can be calculated as follows. Assume that a batch of array data of length N

has been collected. We take three overlapping segments of array data of length N_0 ($N_0 < N$), $\{X_i;\ i = m_0, m_1, m_2,\}$, in which

$$X_i = \{\underline{x}(n + m_i);\ i = 0, 1, 2\ \ n = 1, 2, \ldots, N_0\}, \tag{6.33}$$

where m_i are ordered intergers with $m_0 = 1$. For each segment of array data, we assume that the source targets are approximately stationary and estimate the source target DOA states $\Theta(m_i)$ using the classical high resolution DOA techniques, such the MUSIC and DMLE algorithms. Note that, if the subspace based techniques are to be used, the length of each segment should satisfy $N_0 > K$ [13].

With the three target DOA estimates $\{\hat{\Theta}(m_i), i = 0, 1, 2\}$, the initial estimates for $q_k(1)$ and $\phi_k(1)$ can be computed from the target state dynamics (5.4) and (5.5) as

$$\begin{aligned} \phi_k(1) \;\; = \;\; & -\theta_k(m_2) + arctan\{\cot[\theta_k(m_2) - \theta_k(m_1)] \\ & -\zeta \frac{\sin[\theta_k(m_3) - \theta_k(m_1)]}{\sin[\theta_k(m_2) - \theta_k(m_1)]\sin[\theta_k(m_3) - \theta_k(m_2)]}\}. \end{aligned} \tag{6.34}$$

and

$$q_k(m_1) = \frac{\sin[\theta_k(m_3) - \theta_k(m_1)]}{(m_3 - m_1)T\cos[\theta_k(m_3) + \phi_k(1)]}, \tag{6.35}$$

respectively, where $\zeta = (m_2 - m_1)/(m_3 - m_1)$. For a given batch of data, the converged initial MTS estimate $\underline{\hat{\alpha}}(1)$ is independent of $\underline{\alpha}^{(0)}(1)$ as long as $\underline{\alpha}^{(0)}(1)$ is in the region of convergence. Although this region of convergence is not rigorously defined, $\underline{\alpha}^{(0)}(1)$ should be close enough to the optimum $\underline{\hat{\alpha}}(1)$. We emphasize that the methods for obtaining the initial MTS eitimate is useful to initialize the optimization algorithm. Once an optimum estimate for the first batch of N sampled array data is obtained, it can be used to generate the initial estimate for the next batch of N sampled array data, and the initial estimation procedure is no longer repeated.

6.4 The modified Newton algorithm

When the first batch of N array data has been collected, we can calculate the initial MTS vector estimate $\underline{\alpha}^{(0)}(1)$ and start the modified Newton algorithm iteratively as

$$\hat{\underline{\alpha}}^{(k+1)}(1) = \hat{\underline{\alpha}}^{(k)}(1) - \mu_k H^{-1} G, \tag{6.36}$$

where G and H are the gradient and the approximate Hessian of J with respect to $\underline{\alpha}(1)$, given in (6.13) and (6.32) respectively. Both the gradient and the Hessian are evaluated at $\underline{\alpha}^{(k)}(1)$. In the evaluation, the estimated states $\{\underline{\alpha}^{(k)}(i);\ i = 2, 3, \ldots, N\}$ can be predicted from $\underline{\alpha}^{(k)}(1)$ by the target state equations. In the algorithm, each iteration involves the computation of the pseudo-inverse $A^{\dagger}(i)$ and the projection $P_i^{\perp}$. This can be done by QR-decomposition instead of the direct calculation to reduce the computational burden. Thus,

$$A(i) = Q_{A(i)} R_{A(i)} = [Q_1, Q_2] \begin{bmatrix} R_{A(i)} \\ 0 \end{bmatrix}$$

and

$$A^{\dagger}(i) = R_{A(i)}^{-1} Q_1^H, \text{ and } P_i^{\perp} = Q_2 Q_2^H,$$

where matrix Q_1 is of size $M \times K$, Q_2 is $M \times (M - K)$ and R_A is $K \times K$.

At each iteration, the steplength μ_k should be properly selected to make sure that the modified algorithm is a descent algorithm. The iteration continues until some prescribed ending criterion is satisfied. We give some examples of the criterion

- $\mid H^{-1}G \mid$ is within the specified tolerance and $H > 0$.
- no improvement can be found between two successive iterations. This can be detected by either of the following phenonmena

  1. the distance between consecutively computed estimates $\mid \underline{\alpha}^{(k+1)}(1) - \underline{\alpha}^{(k)}(1) \mid_F$ is less than a preset tolerance;
  2. at the kth iteration, the step-length μ_k is less than a prescribed tolerance.

- the number of iterations reaches a specified maximum limit.

Once the initial MTS estimate $\hat{\underline{\alpha}}(1)$ is obtained, the MTS vector at any arbitrary time instant within the observation time can be computed simply from the target motion dynamics (5.10). It is well known that the ML estimators satisfies the invariance property [47]. Since one MTS vector can be reconstructed from another MTS vector uniquely by the target motion dynamics, the estimates $\{\hat{\underline{\alpha}}(j);\ j = 2, 3, \ldots, N\}$ thus computed from the ML estimate $\hat{\underline{\alpha}}(1)$ are also the ML estimates.

An alternative to estimate the MTS vector $\underline{\alpha}(j)$ for $j = 2, 3, \ldots, N$ is to construct the maximum likelihood estimator based directly on $\underline{\alpha}(j)$ and obtain the estimate $\hat{\underline{\alpha}}(j)$ by optimizing the criterion J in (5.34) with respect to $\underline{\alpha}(j)$. When the Newton-type algorithm is applied to the optimization procedure, the gradient and the Hessian of the criterion function are both derived with respect to MTS vector $\underline{\alpha}(j)$. To evaluate the gradient and the Hessian at each iteration, we need to compute the gradient of the kth DOA state $\theta_k(j)$ with respect to the target state $\underline{\alpha}(j)$ where $i \geq j$ or $i < j$. Relation (6.11) is valid only for the gradient computation when $i \geq j$. When $i < j$, we derive the general expression for the gradient as

$$\underline{v}_k(i,j) = \begin{bmatrix} \dfrac{\partial\theta_k(i)}{\partial\theta_k(j)} \\ \dfrac{\partial\theta_k(i)}{\partial q_k(j)} \\ \dfrac{\partial\theta_k(i)}{\partial\phi_k(j)} \end{bmatrix} = \begin{bmatrix} \dfrac{q_k^2(j) + (i-j)Tq_k(j)\sin[\theta_k(j)+\phi_k(j)]}{q_k^2(j) + (i-j)^2T^2 + 2(i-j)Tq_k(j)\sin[\theta_k(j)+\phi_k(j)]} \\ \dfrac{-(i-j)T\cos[\theta_k(j)+\phi_k(j)]}{q_k^2(j) + (i-j)^2T^2 + 2(i-j)Tq_k(j)\sin[\theta_k(j)+\phi_k(j)]} \\ \dfrac{-(i-j)^2T^2 - (i-j)Tq_k(j)\sin[\theta_k(j)+\phi_k(j)]}{q_k^2(j) + (i-j)^2T^2 + 2(i-j)Tq_k(j)\sin[\theta_k(j)+\phi_k(j)]} \end{bmatrix}. \tag{6.37}$$

Accordingly, the gradient and the approximate Hessian are derived as

$$\begin{aligned} G &= -2vec\left\{\frac{1}{N}\sum_{i=1}^{N} V(i,j)Re\{I \odot [A^{\dagger}(i)\hat{R}_x(i)P^{\perp}_{A(i)}D(i)]\}\right\} \\ H &= \frac{1}{N}\sum_{i=1}^{N}[\nabla^2 J_i \otimes \mathbf{O}] \odot [vecV(i,j)vec^H V(i,j)], \end{aligned} \tag{6.38}$$

where matrix $V(i,j)$ is defined as

$$V(i,j) = [\underline{v}_1(i,j), \underline{v}_2(i,j), \ldots, \underline{v}_K(i,j)]. \tag{6.39}$$

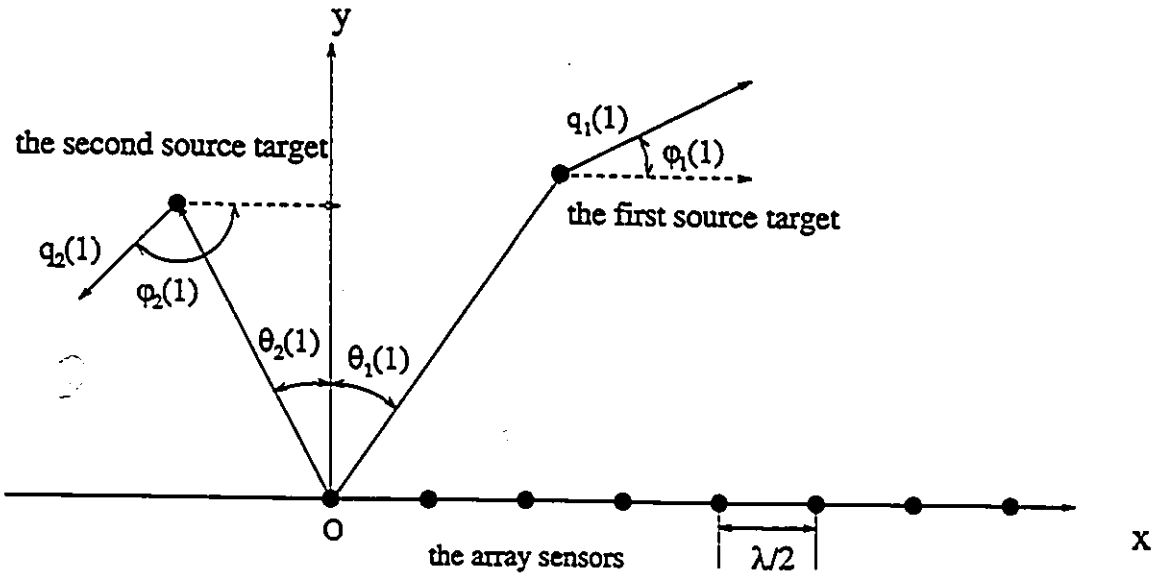

Figure 6.1: The array geometry of an equi-spaced linear array.

and both the gradient and the Hessian are evaluated at $\underline{\hat{\alpha}}(j)$ in each iteration.

When the $(N+1)$th array data measurement is available, we form the new data batch $\{\underline{x}(i);\ i = 2, 3, \ldots, N+1\}$ and apply the same estimation procedure to obtain the ML estimate of $\underline{\alpha}(2)$. This time, the initial guess procedure is not necessary, and the initial estimate can be assigned as $\underline{\alpha}^{(0)}(2) = \underline{\hat{\alpha}}(2)$ which is predicted from $\underline{\hat{\alpha}}(1)$ by (5.10). Since $\underline{\alpha}^{(0)}(2)$ is the ML estimate $\underline{\alpha}(2)$ obtained from the preceeding array data batch, the algorithm is expected to converge to the optimum estimate very quickly when it is applied to the new array data batch. Also, since the estimates are automatically associated with the previous ones, the data association problem does not arise. As new array data continue to arrive, the algorithm is repeated, and the target states including target DOA state can be successively estimated and tracked.

6.5 Numerical simulation studies

Simulation studies are carried out for an $M = 8$ sensors equi-spaced linear array with half source wavelength sensor spacing. The array geometry is shown in Fig. 6.1. In the first part, we study the effects of the source motion on the array and in the second part, we investigate the performance of the ML tracking algorithm via computer simulations.

6.5.1 Effects of source motion on array processing

Generally speaking, the source target motion tends to spread the array spatial spectrum in the spatial domain and result in lower performance in source target dectection and estimation. In the temporal domain, the target motion makes the received array data non-stationary even though the source signal themselves may be stationary. Since most high resolution array processing techniques rely on the stationarity assumption, their applications and performances are limited by the source motion effects.

Consider a single linearly moving source target. To show how source motion will affect array processing, we use the following simplistic model [33] for a small observation interval

$$\begin{aligned} \sin\theta(i+m) &= \sin\theta(i) + \{cos\phi(i) - \sin\theta(i)\sin(\theta(i)-\phi(i))]\}\frac{vmT}{r(i)} \\ &= \sin\theta(i) + k_i m. \end{aligned} \tag{6.40}$$

where

$$k_i = \{cos\phi(i) - \sin\theta(i)\sin(\theta(i)-\phi(i))]\}\frac{vT}{r(i)}. \tag{6.41}$$

In (6.40), the locally linear motion model is approximated by the linear model in $\sin\theta(i+m)$. Let $i = 0$ and (6.40) becomes

$$\sin\theta(m) = \sin\theta(0) + k_0 m. \tag{6.42}$$

The spread spatial spectrum effects due to target motion can be observed from the asymptotic array spatial density which is defined by

$$P_m(\theta) = E\{\underline{a}^H(\theta)\hat{R}_x\underline{a}(\theta)\} = \underline{a}^H(\theta)R_x\underline{a}(\theta). \tag{6.43}$$

By substituting R_x and $\underline{a}(\theta)$ into (6.43) and ignoring the constant scaling factor, we get

$$P_m(\theta) = \sum_{i=0}^{N-1} \underline{a}^H(\theta)\underline{a}[\theta(i)]\underline{a}^H[\theta(i)]\underline{a}(\theta)$$

$$= \sum_{m=1}^{M}\sum_{l=1}^{M} \exp\{-j(m-l)\pi[\sin\theta - \sin\theta(0) - \frac{N-1}{2}k_0]\}\frac{\sin[(m-l)\pi N k_0/2)]}{\sin[(m-l)\pi k_0/2)]}$$

$$= \sum_{m=-M+1}^{M-1} (M-|m|)\exp\{-jm\pi[\sin\theta - \sin\theta(0) - \frac{N-1}{2}k_0]\}\frac{\sin[m\pi N k_0/2)]}{\sin[m\pi k_0/2)]}. \quad (6.44)$$

For a stationary source signal at $\theta(0)$, the asymptotic spatial spectrum is [68]

$$P_s(\theta) = \left\{\frac{\sin[M\pi\sin(\theta-\theta(0))/2]}{\sin[\pi\sin(\theta-\theta(0))/2]}\right\}^2. \quad (6.45)$$

Compared to (6.45), the spatial density (6.44) in the presence of moving target exhibits two effects: biased peak location and enlarged main lobe. The peak location θ_p of $P_m(\theta)$ is found to satisfy the following relationship

$$\sin\theta_p = \sin\theta(0) + \frac{N-1}{2}k_0, \quad (6.46)$$

and if we use $\sin\theta$ to denote the source DOA parameter, $\sin\theta_p$ is seen to have a biase of $(N-1)k_0/2$. Fig. 6.2 shows the array spatial densities in the presence of stationary and moving source targets. They are plotted for different numbers of array data samples. The solid line represents the case of a stationary source target at $\theta(0) = 10°$. The dashed lines are the spatial densities for a linearly moving target with $\theta(0) = 10°$ and $k_0 = 0.005$. The numbers of the array samples used are $N = 50, 100$ and 150. From the plot, the bias is observed to increase with the number of array data. We also observe that the main lobe widens as the array sample number is increased.

Furthermore, these effects increase as the array sensor number becomes greater. Fig. 6.3 shows the variation of the array spatial spectrum with the number of sensors. In the example, the number of array sample is $N = 100$. The spatial spectra have the same peak biases. But, the spatial spectra become flatter and their main lobe wider as the number of sensors is increased. This is different from the case of stationary source where the spectrum becomes narrower and sharper as the number of sensors increases. The widening and flattening of spatial spectrum in the presence of moving targets will result in degraded performance in terms of source DOA resolution and the estimation variance.

Another important effect of the moving targets on array processing is the increased array

data covariance rank. Since most high resolution array processing techniques rely on the eigen-decomposition of the estimated covariance matrix of the array data, the consequences are source splitting and false detection of the number of source targets present. For K stationary source signals, it is known that the array data covariance matrix has K signal eigenvalues. The source motion changes the distribution of the eigenvalues and tends to blur the boundary between the signal and noise eigenvalues, making them difficult to separate. Usually, it will result in a false detection of higher number of the source incident on the array, or a higher dimension for the estimated signal subspace. False detection due to source motion in turn will lead to poor estimation performance.

6.5.2 Tracking the DOA's of the moving source targets

In the simulation study, two narrow-band source targets of equal power are used. The array sensor noise is assumed to be a white Gaussian process with unit variance and be independent between sensors. The sampling period is normalized to $T = 1$ and the source powers are normalized to unity. The array SNR is defined as

$$\text{SNR} = 20 \log \frac{\sigma_s}{\sigma_w} = 20 \log \sigma_s, \tag{6.47}$$

where σ_s^2 denotes the source signal power. At each iteration in the Gauss-Newton type algorithm, we choose an initial $\mu_0 < 1$ and take $\mu_k = (\mu_0)^i$ for the smallest integer i that causes the most decrease in the criterion function.

LINEAR UNIFORMLY MOVING TARGETS Assume that two strictly linear uniformly moving source targets with heading directions $\phi_1(1) = -135°$ and $\phi_2(1) = 30°$. The two source targets start moving at DOA locations $\theta_1(1) = 0°$ and $\theta_2(1) = -20°$. The initial q components of the source MTS vector are set to $q_1(1) = 7.4 \times 10^{-3}$ and $q_2(1) = 5.3 \times 10^{-3}$. We set SNR=5dB. Fig. 6.4, Fig. 6.5 and Fig. 6.6 show the tracking of the source targets DOA's when the two sources are uncorrelated, correlated and coherent. In the correlated case, the correlation coefficient between the two sources is set to 0.75. When simulating the coherent sources, we copy one source waveforms from the other by a complex attenuation factor $0.9e^{j0.31}$. In all three plots, the source DOA trajectories are seen

to be successfully tracked.

In Fig. 6.8, we compare the ML tracking results with those obtained from the traditional ML estimator in which we treat the source targets as stationary in a relatively short data period of 50. Two uncorrelated linearly moving targets are assumed. The total number of array samples is $N = 400$, whereas $SNR = 5$dB. The initial source target parameters are chosen as $\theta_1(1) = 0°$, $\theta_2(1) = -20°$, $q_1(1) = 7.4\times10^{-3}$ and $q_2(1) = 5.3\times10^{-3}$ and the heading directions $\phi_1(1) = 135°$ and $\phi_2(1) = 30°$. From Fig. 6.8, it is observed that the ML tracking algorithm consistently outperforms the traditional ML estimator while the latter exihibits a bias from the actual source target DOA trajectories. This bias is caused by the spread spectrum effects due to target motion. Furthermore, since the linear motion target model has been assumed, the DOA estimates are automatically associated with the previous estimates through their underlying assumed target dynamical models, and no data association problems arise.

Fig. 6.7 shows the tracking results of multiple linear uniformly moving targets. The number of source targets present is $K = 4$ and the initial source parameters are chosen as follows

$$\theta_1(1) = 0° \quad q_1(1) = 0.0148 \quad \phi_1(1) = -135°$$

$$\theta_2(1) = -20° \quad q_2(1) = 0.0071 \quad \phi_2(1) = 30°$$

$$\theta_3(1) = 20° \quad q_3(1) = 0.0154 \quad \phi_3(1) = 135°$$

$$\theta_4(1) = 40° \quad q_4(1) = 0.0114 \quad \phi_4(1) = 0°$$

In Fig. 6.7, the source DOA trajectories are estimated accurately at each point and tracked correctly.

NON-LINEARLY MOVING SOURCE TARGETS We consider non-linearly moving source targets and examine how the ML tracking algorithm performs when local linear motion is used to approximate the actual non-linear target trajectories. Two uncorrelated source targets are simulated with their trajectories described by parabolas

$$
\begin{aligned}
\tilde{x}_i(n) &= \tilde{x}_i(n-1) + \frac{s_i}{\sqrt{1+4\tilde{x}_i^2(n-1)/d_i}} \\
\tilde{y}_i(n) &= \tilde{x}_i^2(n)/d_i; \quad i=1,2 \\
x_i(n) &= x_{i0} + \tilde{x}_i(n)\cos\varphi_i - \tilde{y}_i(n)\sin\varphi_i \\
y_i(n) &= y_{i0} + \tilde{y}_i(n)\sin\varphi_i + \tilde{y}_i(n)\cos\varphi_i,
\end{aligned} \tag{6.48}
$$

where $[x_i(n), y_i(n)]$ denotes the ith source target location at the observation time index n, $[x_{i0}, y_{i0}]$ is the location of the origin of the ith parabola and φ_i denotes the rotation angle of the trajectory. In (6.48), parameter d_i controls the curvature, and s_i represents the line speed of the source target along its trajectory. In the simulation, the source target parameters are chosen as follows : $[x_{10}, y_{10}] = [-150, 600]$ and $[x_{20}, y_{20}] = [-60, 700]$, $\varphi_1 = \varphi_2 = 0$, $d_1 = 4000$ and $d_2 = -3400$, and $s_1 = 4.5$ and $s_2 = -4.0$. The starting points for the two targets are at $x_1(1) = -50$ and $x_2(1) = 890$. Fig. 6.9 plots the source target trajectories in the x-y plane. The aray signal-to-noise ratio is set to SNR=10dB. The length of the data batch is $N = 50$. A total of 300 array samples is simulated. The non-linear source target trajectory is approximated by the piece-wise linear uniform motion, and the maximum likelihood algorithm is then applied in each data batch to track the source DOA trajectories. Fig. 6.10 shows the tracking results of the ML algorithm, and it is seen that there is considerable agreement between the actual and the computed target DOA trajectories. It should be noted that when the ML algorithm is applied to source targets with non-linear trajectories, the selection of the data batch length N is twofold. On one hand, N should be sufficiently large to retain the superior performance of the ML estimates. On the other hand, it should be small enough to provide accurate approximation when locally linear motion model is used. The final tracking performance usually depends on the underlying non-linear structure of the target trajectories.

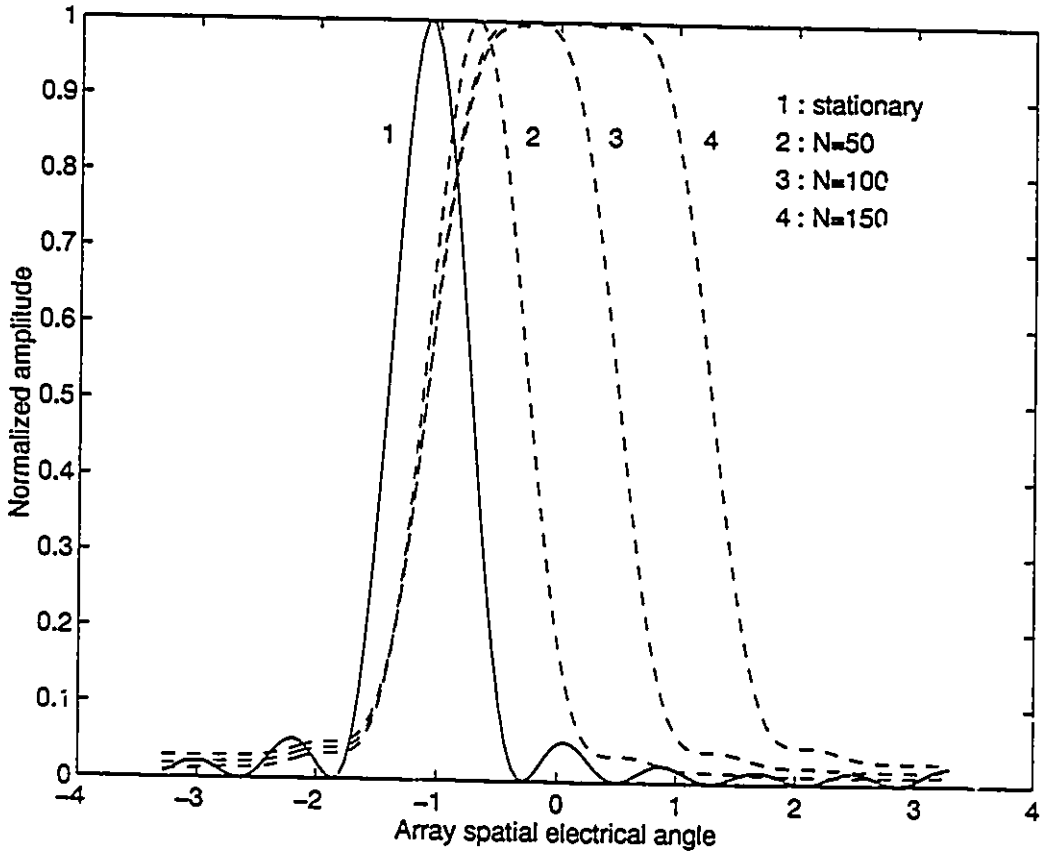

Figure 6.2: Variation of the array spatial density with stationary and moving targets with different array sample size

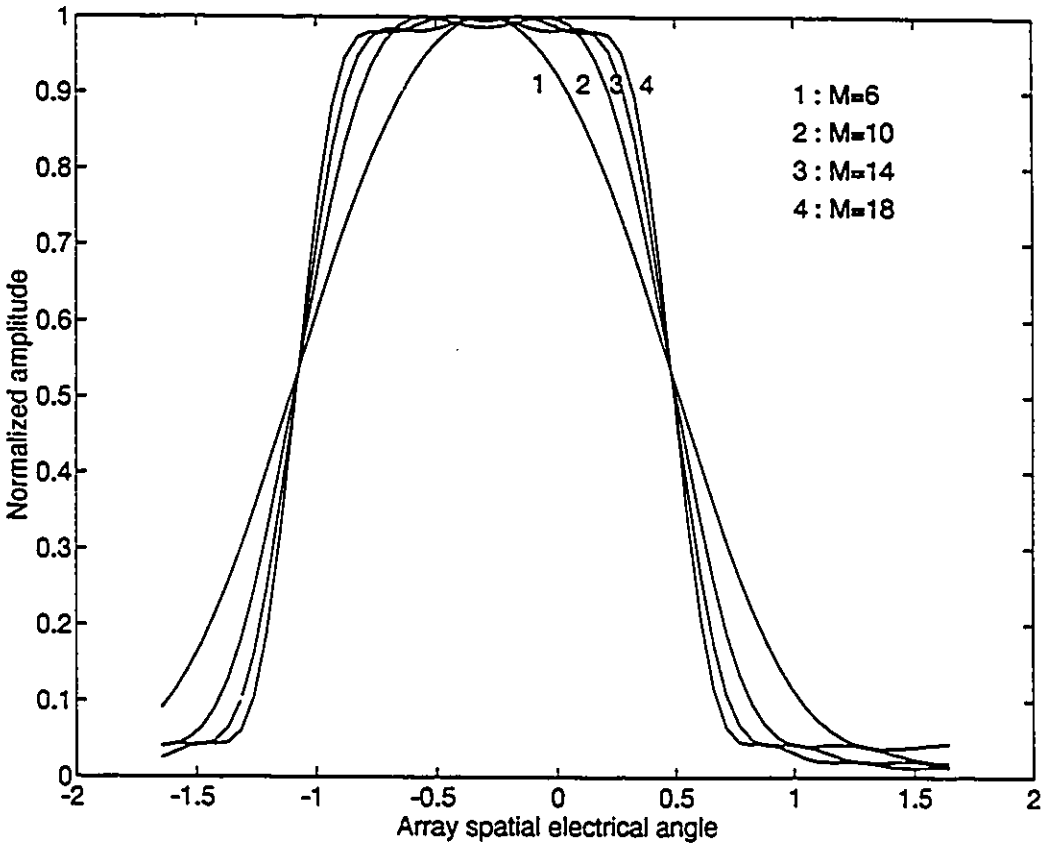

Figure 6.3: Variation of the spatial density with the number of sensors in the array

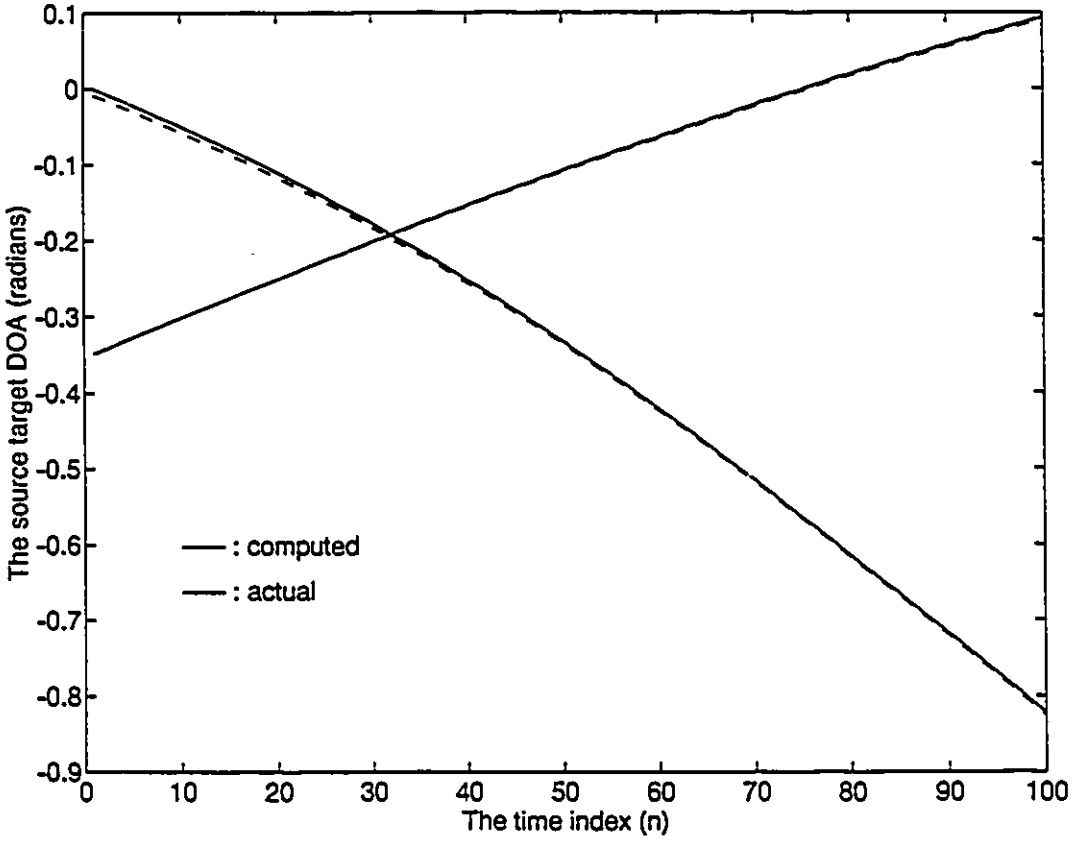

Figure 6.4: The DOA trackings of two linear uniformly moving source targets : uncorrelated case.

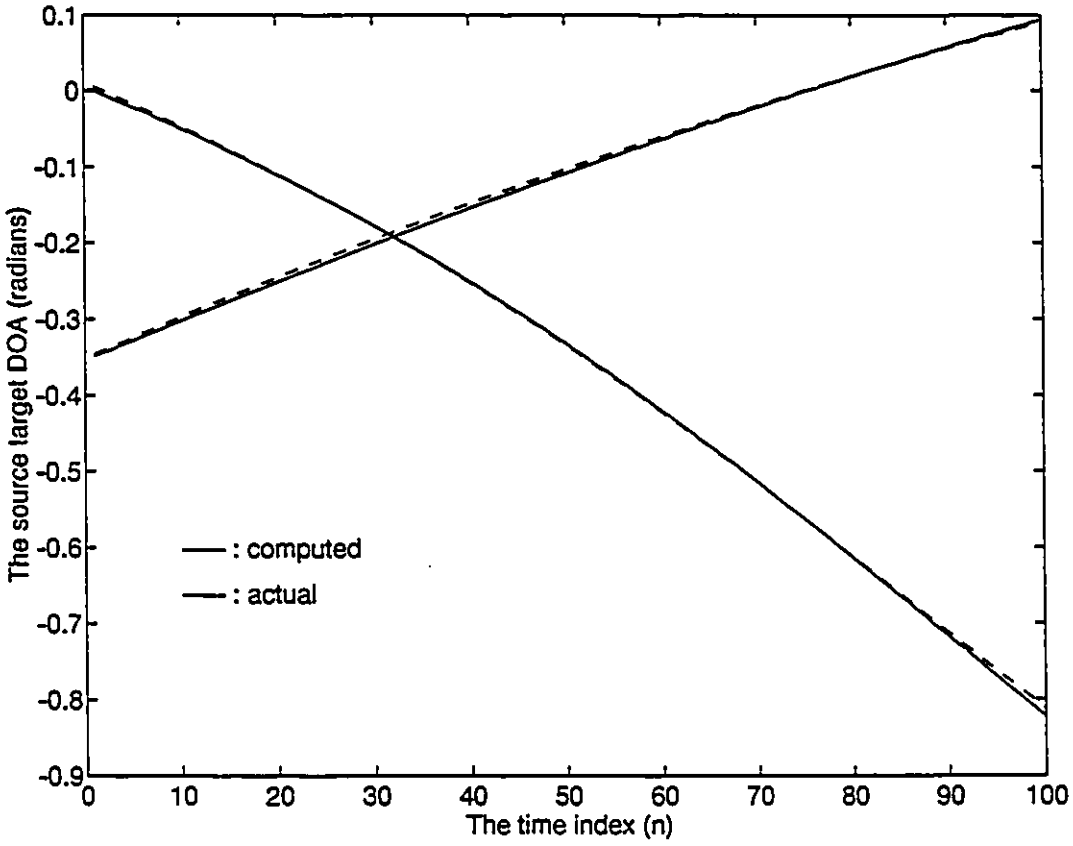

Figure 6.5: The DOA trackings of two linear uniformly moving source targets : correlated case.

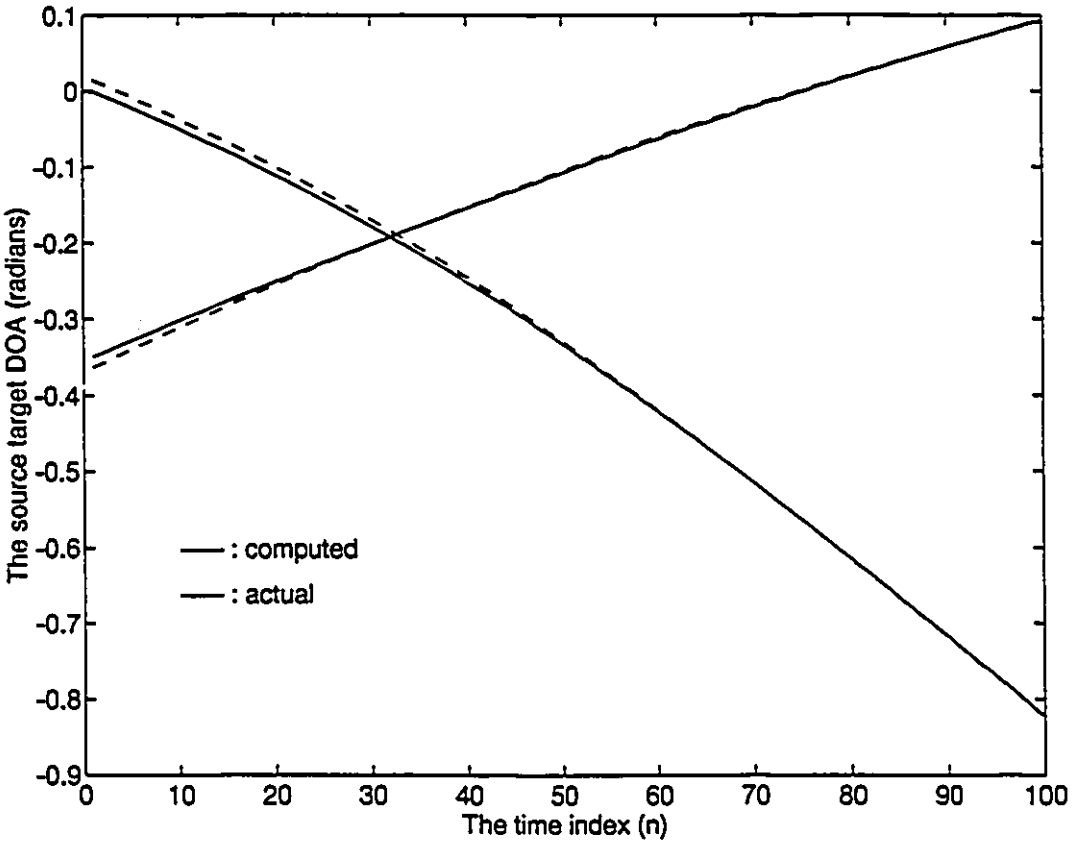

Figure 6.6: The DOA trackings of two linear uniformly moving source targets : coherent case.

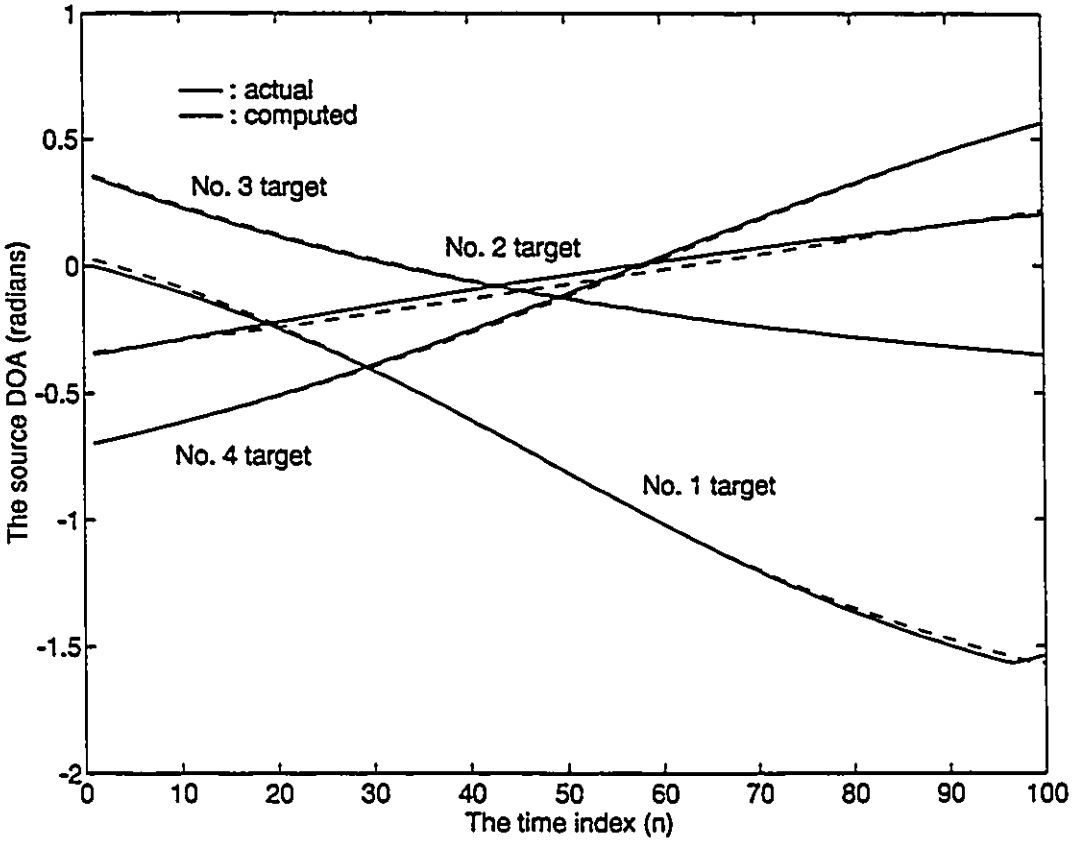

Figure 6.7: The DOA trackings of multiple linear uniformly moving source targets

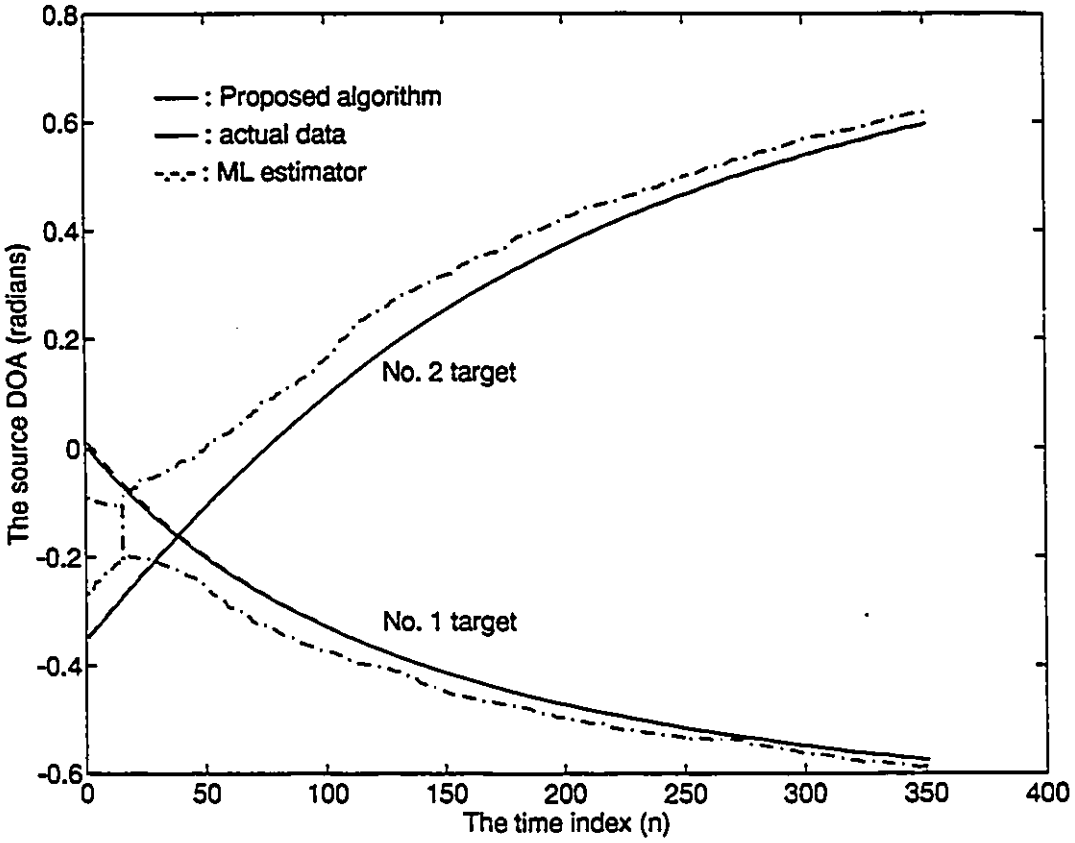

Figure 6.8: Comparison of tracking results between the ML tracking algorithm and the conventional ML estimator.

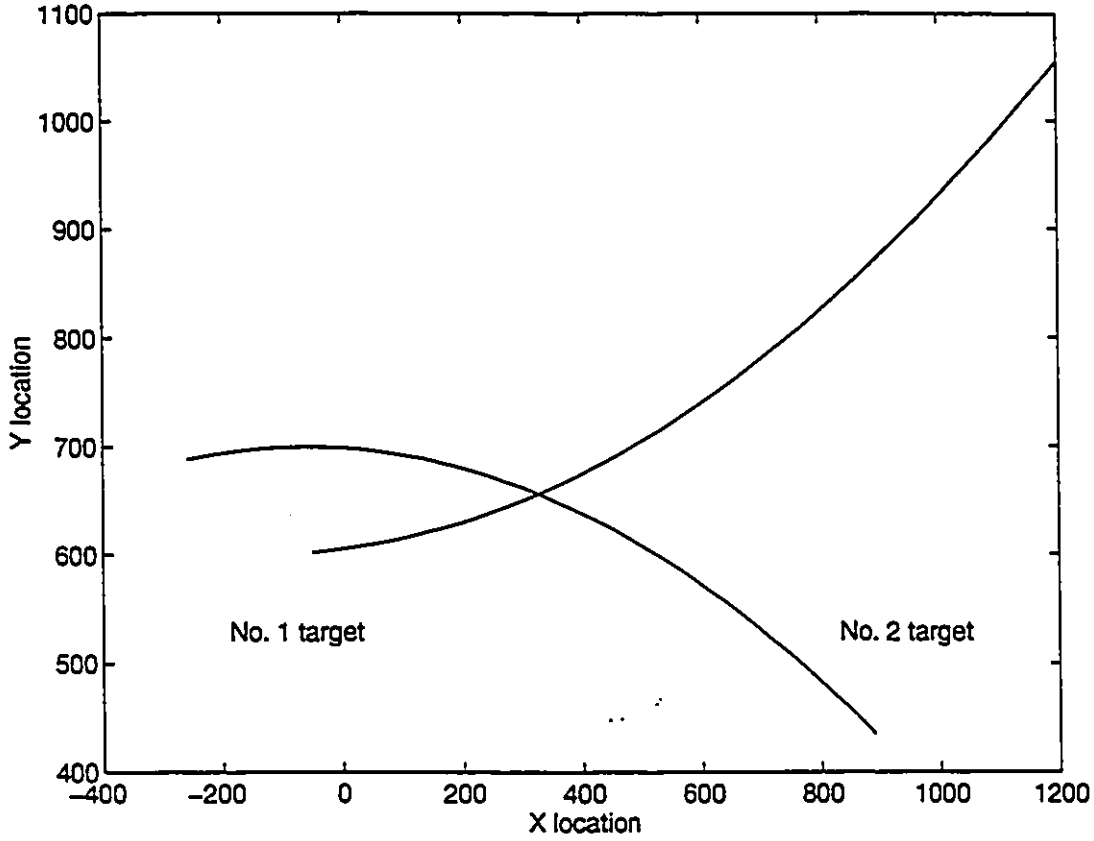

Figure 6.9: The trajectories of two non-linearly moving targets

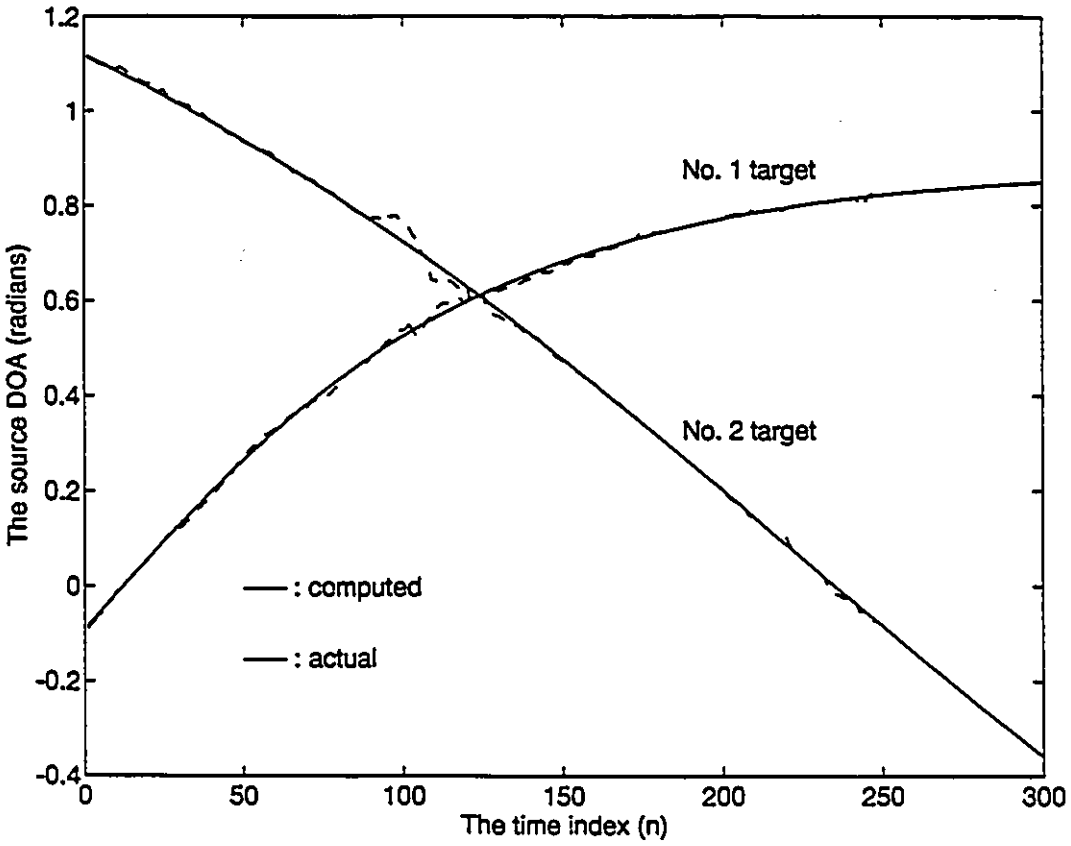

Figure 6.10: The DOA trackings of two non-linear moving source targets

Chapter 7

Asymptotic Performance Analysis and the Cramér-Rao Bound

In this chapter, an asymptotic performance analysis of the ML tracking algorithm is carried out. By asymptotic we mean that the number of samples is sufficiently large. An analytic form for the asymptotic mean-squared error (MSE) of the maximum likelihood MTS estimates is derived, and the strong asymptotic consistency of the estimates is shown. The Cramér-Rao bound (CRB) on the estimator is discussed and is shown to be relatively asymptotically efficient.

7.1 Asymptotic consistency of the estimates

One of the most important measures for an estimator is its consistency. A consistent estimate will converge to the true parameter when the sam[le size tends to infinity. In the following lemma, we show that the ML estimates, $\{\hat{\Theta}(i)\}$; $i = 1, 2, \ldots, N$, are consistent estimates.

Lemma 2 The ML DOA estimate, $\{\hat{\Theta}(i);\ i = 1, 2, \ldots, N\}$, obtained from (5.35) converges w.p.1 each to $\{\Theta(i);\ i = 1, 2, \ldots, N\}$ as the number of the array data approaches infinity.

Proof First, we show that the objective function J converges w.p.1, uniformly in parameter Θ to its limit function $\overline{J}$ given by

$$\overline{J} = \frac{1}{N}\sum_{i=1}^{N} tr\{P^{\perp}_{A(i)} R_x(i)\}, \tag{7.1}$$

where

$$R_x(i) = A(i)R_s(i)A^H(i) + \sigma_w^2 I. \tag{7.2}$$

To simplify notation, we use $P_i^{\perp}$ instead of $P^{\perp}_{A(i)}$. Consider the difference

$$\begin{aligned} | J - \overline{J} | &= | \frac{1}{N}\sum_{i=1}^{N} tr\{P_i^{\perp}[\underline{x}(i)\underline{x}^H(i)]\} - \frac{1}{N}\sum_{i=1}^{N} tr\{P_i^{\perp} R_x(i)\} | \\ &= | \frac{1}{N}\sum_{i=1}^{N} tr\{P_i^{\perp}[A(i)\underline{s}(i)\underline{w}^H(i) + \underline{w}(i)\underline{s}^H(i)A^H(i) + \underline{w}(i)\underline{w}^H(i) - \sigma_w^2 I]\} | \end{aligned} \tag{7.3}$$

and using triangular inequality, we obtain

$$\begin{aligned} | J - \overline{J} | \leq & | \frac{1}{N}\sum_{i=1}^{N} tr\{P_i^{\perp}[A(i)\underline{s}(i)\underline{w}^H(i)]\} | + | \frac{1}{N}\sum_{i=1}^{N} tr\{P_i^{\perp}[\underline{w}(i)\underline{s}^H(i)A^H(i)]\} | \\ & + | \frac{1}{N}\sum_{i=1}^{N} tr\{P_i^{\perp}[\underline{w}(i)\underline{w}^H(i) - \sigma_w^2 I]\} | . \end{aligned} \tag{7.4}$$

For an array noise vector $\underline{w}(i)$ with Gaussion distribution, we show that [See Appendix C] the following limits hold w.p.1

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} tr\{P_i^{\perp}[A(i)\underline{s}(i)\underline{w}^H(i)]\} &= 0 \\ \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} tr\{P_i^{\perp}[\underline{w}(i)\underline{s}^H(i)A^H(i)]\} &= 0 \\ \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} tr\{P_i^{\perp}[\underline{w}(i)\underline{w}^H(i) - \sigma_w^2 I]\} &= 0. \end{aligned} \tag{7.5}$$

Consequently, since the difference (7.4) tends to 0 w.p.1, their minimizing arguments of J and $\overline{J}$ will certainly coincide w.p.1, i.e., the ML estimate $\hat{\Theta}(i)$ will converges w.p.1 to the minimizing argument of $\overline{J}$, $\Theta(i)$. Let $\Theta'(i)$ denote the true parameters. The limit criterion function $\overline{J}$ can be written as

$$\begin{aligned}\overline{J} &= \frac{1}{N}\sum_{i=1}^{N} tr\{P_i^{\perp}[A(i)R_s(i)A^H(i)+\sigma_w^2 I]\} \\ &= \frac{1}{N}\sum_{i=1}^{N} tr\{P_i^{\perp}[A(i)R_s(i)A^H(i)]\}+\frac{1}{N}\sum_{i=1}^{N} tr\{P_i^{\perp}[\sigma_w^2 I]\} \\ &\geq \frac{1}{N}\sum_{i=1}^{N} tr\{P_i^{\perp}[\sigma_w^2 I]\}. \end{aligned} \tag{7.6}$$

Since $A(i)$ is of full rank and $P_i^{\perp}$ is an orthogonal projection matrix, we have therefore,

$$\overline{J} \geq \sigma_w^2(M-K). \tag{7.7}$$

Equality is valid only when $P_i^{\perp}A(i)=0$ which in turn implies $A[\Theta(i)]=A[\Theta'(i)]T_a$, where T_a is a nonsingular matrix of size $K\times K$. By the uniqueness condition of the array model (5.16), this is possible if and only if $\hat{\Theta}(i)=\Theta'(i)$ for $i=1,2,\ldots,N$. This lemma is thus readily established. □

However, the estimated signal waveform $\{\underline{\hat{s}}(i)\}$ and noise variance $\hat{\sigma}_w^2$ are not consistent estimates. Since $\hat{\Theta}(i)$ tends to $\Theta'(i)$ when N approaches infinity, from (5.26) and (5.28), we have

$$\begin{aligned}\lim_{N\to\infty}\underline{\hat{s}}(i) &= \{A^H(i)A(i)\}^{-1}A^H(i)\underline{x}(i) \\ \lim_{N\to\infty}\hat{\sigma}_w^2 &= \frac{1}{MN}\sum_{i=1}^{N} tr\{P_i^{\perp}\underline{w}(i)\underline{w}^H(i)\}. \end{aligned} \tag{7.8}$$

Taking the expectations of the limits of $\underline{\hat{s}}(i)$ and $\hat{\sigma}_w^2$ yields

$$\begin{aligned}\lim_{N\to\infty} E[\underline{\hat{s}}(i)] &= \underline{s}(i) \\ \lim_{N\to\infty} E[\hat{\sigma}_w^2] &= \frac{M-K}{M}\sigma_w^2. \end{aligned} \tag{7.9}$$

Thus, the estimated waveform $\underline{\hat{s}}(i)$ is asymptotically unbiased while the noise variance estimate $\hat{\sigma}_w^2$ is asymptotically biased. Using the array signal model (5.14), we obtain the the following limit

$$\lim_{N\to\infty}\underline{\hat{s}}(i)=\underline{s}(i)+\{A^H(i)A(i)\}^{-1}A^H(i)\underline{w}(i), \tag{7.10}$$

which differs from the true signal waveforms even when N tends to infinity. This, in turn, implies the inconsistency of the ML estimate $\underline{\hat{s}}(i)$. Although $\hat{\sigma}_w^2$ is asymptotically biased, we can still construct an unbiased estimate $\bar{\sigma}_w^2$

$$\bar{\sigma}_w^2 = \frac{M}{M-K}\hat{\sigma}_w^2 = \frac{1}{N(M-K)}\sum_{i=1}^{N} tr\{P_i^{\perp}\underline{w}(i)\underline{w}^H(i)\}, \tag{7.11}$$

which can be shown to be asymptotically unbiased. The aysmptotic convergence of $\bar{\sigma}_w^2$ to the true σ_w^2 can be verified by using the fact that the following limit holds w.p.1 (see Appendix C)

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} tr\{P_i^{\perp}\underline{w}(i)\underline{w}^H(i)\} = (M-K)\sigma_w^2. \tag{7.12}$$

Considering the asymptotic consistency of the MTS estimates $\underline{\hat{\alpha}}(i)$, we have the following corollary.

Corollary 1 The maximum likelihood MTS estimate $\underline{\hat{\alpha}}(1)$ converges w.p.1 to its true value $\underline{\alpha}(1)$ as the number of array samples tends to infinity.

This can be directly inferred from **Lemma 1**. For the kth source target, since the corresponding state vector components $\{q_k(i), \phi_k(i)\}$ in (6.27) and (6.28) are both continuous in $\{\theta_k(i+m);\ m = 0,1,2\}$, they will converge to their true values w.p.1 when the DOA estimates $\hat{\Theta}(i)$ converges to $\Theta(i)$ w.p.1. Consequently, the maximum likelihood MTS estimate $\underline{\hat{\alpha}}(i)$ is asymptotically consistent.

7.2 Asymptotic MSE analysis of the estimates

The consistency of the ML estimates has been shown in the previous section. In this section, we provide an analysis of MSE of the ML estimates by applying first-order perturbation theory. The following analysis is based on the assumption that the number of the array samples is sufficiently large. Define a perturbed $R_x(i)$ as

$$\hat{R}_x(i) = R_x(i) + \Delta R_x(i). \tag{7.13}$$

Taking expectation of ΔR_x gives

$$E[\Delta R_x(i)] = E[\hat{R}_x(i)] - R_x(i) = 0, \tag{7.14}$$

which indicates that $\hat{R}_x(i)$ is an unbiased estimate of $R_x(i)$. Define ΔJ as the perturbation on the limiting criterion function $\overline{J}$ and the criterion J can be written as

$$J = \overline{J} + \Delta J. \tag{7.15}$$

In (7.15), by invoking the perturbation form of $\hat{R}_x(i)$, we obtain the perturbation on $\overline{J}$ as

$$\Delta J = \frac{1}{N}\sum_{i=1}^{N} tr\{P^{\perp}_{A(i)}\Delta R_x(i)\}. \tag{7.16}$$

Since $\underline{\hat{\alpha}}(1)$ is the minimizing argument of the criterion J, it follows that

$$\frac{\partial J}{\partial \underline{\alpha}(1)} \mid_{\underline{\hat{\alpha}}(1)} = 0, \tag{7.17}$$

where $\partial J/\partial \underline{\alpha}(1)$ is the gradient of J with respect to $\underline{\alpha}(1)$. Denote the true value of the initial MTS vector as $\underline{\alpha}'(1)$. For a consistent estimate $\underline{\hat{\alpha}}(1)$, we can write a first-order Taylor series expansion of $\partial J/\partial \underline{\alpha}(1)$ around $\underline{\alpha}'(1)$ for sufficiently large N as

$$0 \simeq \frac{\partial J}{\partial \underline{\alpha}(1)}|_{\underline{\alpha}'(1)} + \frac{\partial^2 J}{\partial \underline{\alpha}^2(1)}|_{\underline{\alpha}'(1)}[\underline{\hat{\alpha}}(1) - \underline{\alpha}'(1)], \tag{7.18}$$

where higher order terms in $[\underline{\hat{\alpha}}(1) - \underline{\alpha}'(1)]$ are neglected. The error term $[\underline{\hat{\alpha}}(1) - \underline{\alpha}'(1)]$ is written as

$$\Delta\underline{\alpha}(1) = \underline{\hat{\alpha}}(1) - \underline{\alpha}'(1) = Q^{-1}[\hat{R}_x, \underline{\alpha}'(1)]P[\hat{R}_x, \underline{\alpha}'(1)], \tag{7.19}$$

where matrix functions P and Q are defined as

$$\begin{aligned} P[\hat{R}_x, \underline{\alpha}'(1)] &= -\frac{\partial J}{\partial \underline{\alpha}(1)}|_{\underline{\alpha}'(1)} \\ Q[\hat{R}_x, \underline{\alpha}'(1)] &= \frac{\partial^2 J}{\partial \underline{\alpha}^2(1)}|_{\underline{\alpha}'(1)}. \end{aligned} \tag{7.20}$$

For convenience, we use the notation $\hat{R}_x$ to represent the matrix set $\{\hat{R}_x(i);\ i = 1, 2, \ldots, N\}$. The first-order perturbation of $P[\hat{R}_x, \underline{\alpha}'(1)]$ and $Q[\hat{R}_x, \underline{\alpha}'(1)]$ can be written

$$
\begin{aligned}
P[\hat{R}_x, \underline{\alpha}'(1)] &= P[R_x, \underline{\alpha}'(1)] + P[\Delta R_x, \underline{\alpha}'(1)] = P_0 + \Delta P \\
Q[\hat{R}_x, \underline{\alpha}'(1)] &= Q[R_x, \underline{\alpha}'(1)] + Q[\Delta R_x, \underline{\alpha}'(1)] = Q_0 + \Delta Q,
\end{aligned} \tag{7.21}
$$

where R_x is used to denote the matrix set $\{R_x(i);\ i = 1, 2, \ldots, N\}$. By substituting the criterion function J and the limiting function $\overline{J}$, we obtain

$$
\begin{aligned}
P_0 &= -\frac{\partial \overline{J}}{\partial \underline{\alpha}(1)}|_{\underline{\alpha}'(1)} \\
Q_0 &= \frac{\partial^2 \overline{J}}{\partial \underline{\alpha}^2(1)}|_{\underline{\alpha}'(1)},
\end{aligned} \tag{7.22}
$$

and

$$
\begin{aligned}
\Delta P &= -\frac{\partial \Delta J}{\partial \underline{\alpha}(1)}|_{\underline{\alpha}'(1)} \\
\Delta Q &= \frac{\partial^2 \Delta J}{\partial \underline{\alpha}^2(1)}|_{\underline{\alpha}'(1)}.
\end{aligned} \tag{7.23}
$$

Since $\underline{\alpha}'(1)$ is the minimizing argument of $\overline{J}$, we have $P_0 = 0$ and (7.19) reduces to

$$
\Delta \underline{\alpha}(1) = (Q_0 + \Delta Q)^{-1} \Delta P. \tag{7.24}
$$

Furthermore, if we expand the inverse matrix in power series as

$$
\Delta \underline{\alpha}(1) = [Q_0^{-1} + Q_0^{-1} \Delta Q Q_0^{-1} + o(\Delta R_x)] \Delta P, \tag{7.25}
$$

we can write $\Delta \underline{\alpha}(1)$ (ignoring the higher orders terms in ΔR_x as

$$
\Delta \underline{\alpha}(1) = Q_0^{-1} \Delta P + o(\Delta R_x). \tag{7.26}
$$

Parameter ΔP is the perturbation on P_0 evaluated at the true parameter values and can be written as

$$\Delta P = -\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial \underline{\alpha}(1)} tr\{P_i^{\perp}\Delta R_x(i)\}|_{\underline{\alpha}'(1)}. \tag{7.27}$$

Denote p_{km} as the $\{3(k-1)+m\}$th element of ΔP, where $m = 1,2,3$ and $k = 1,2,\ldots,K$. It is easy to see that p_{km} is associated with the mth state variable of the kth target state vector. Then, we have

$$p_{km} = -\frac{1}{N}\sum_{i=1}^{N} tr\left\{\frac{\partial P_i^{\perp}}{\partial \theta_k(i)}\frac{\partial \theta_k(i)}{\partial \alpha_{km}(i)}\Delta R_x(i)\right\}|_{\alpha'_{km}(1)}, \tag{7.28}$$

where $\alpha'_{km}(1)$ denotes the true value of the mth state variable of the kth target state vector at the initial time. In the following, for notational simplicity, the quantities are all evaluated at the true parameters by default. Using the array signal model (5.14), we can write the perturbation term $\Delta R_x(i)$ in the following form

$$\Delta R_x(i) = A(i)\underline{s}(i)\underline{w}^H(i) + \underline{w}(i)\underline{s}^H(i)A^H(i) + \underline{w}(i)\underline{w}^H(i) - \sigma_w^2 I. \tag{7.29}$$

Define the matrix $B_{km}(i)$ as

$$B_{km}(i) = \frac{\partial P_i^{\perp}}{\partial \alpha_{km}(1)} = \frac{\partial P_i^{\perp}}{\partial \theta_k(i)}\frac{\partial \theta_k(i)}{\partial \alpha_{km}(1)}. \tag{7.30}$$

Using (7.29) and (7.30), we reduce (7.28) to

$$\begin{aligned} p_{km} &= -\frac{1}{N}\sum_{i=1}^{N} tr\{B_{km}(i)A(i)\underline{s}(i)\underline{w}^H(i)\} \\ &\quad -\frac{1}{N}\sum_{i=1}^{N} tr\{B_{km}(i)\underline{w}(i)\underline{s}^H(i)A^H(i)\} \\ &\quad -\frac{1}{N}\sum_{i=1}^{N} tr\{B_{km}(i)[\underline{w}(i)\underline{w}^H(i) - \sigma_w^2 I]\} \\ &= p_{km}(1) + p_{km}(2) + p_{km}(3), \end{aligned} \tag{7.31}$$

where we have defined the three terms in p_{km} as $p_{km}(1)$, $p_{km}(2)$ and $p_{km}(3)$. The covariance between p_{km} and p_{qn} is derived as (see Appendix D)

$$\begin{aligned}\mathrm{Cov}[p_{km}, p_{qn}] &= E[p_{km}(1)p_{qn}^*(1)] + E[p_{km}(2)p_{qn}^*(2)] + E[p_{km}(3)p_{qn}^*(3)] \\ &= \frac{2\sigma_w^2}{N^2}\sum_{i=1}^{N} Re\left[tr\{B_{km}(i)A(i)R_s(i)A^H(i)B_{qn}^H(i)\}\right] \\ &\quad + \frac{\sigma_w^4}{N^2}\sum_{i=1}^{N} tr\{B_{km}(i)B_{qn}^H(i)\}. \end{aligned} \tag{7.32}$$

Substituting (6.6) into (7.32), and using $P_i^{\perp}A(i) = 0$, we get

$$\begin{aligned}\mathrm{Cov}[p_{km}, p_{qn}] &= \frac{2\sigma_w^2}{N^2}\sum_{i=1}^{N} Re\left[v_{mk}(i)v_{nq}(i)tr\{P_i^{\perp}A_k(i)R_s(i)A_q^H(i)\}\right] \\ &\quad + \frac{\sigma_w^4}{N^2}\sum_{i=1}^{N} tr\{B_{km}(i)B_{qn}^H(i)\}, \end{aligned} \tag{7.33}$$

in which we have used the fact $A^{\dagger}(i)A(i) = I$. Define $P_d = E\{\Delta P \Delta P^H\}$. The $\xi\eta$th element of P_d is given by (7.33), where the integer indices k, m, q and n with $k, q = 1, 2, \ldots, K$ and $m, n = 1, 2, 3$ are related through the following relationships

$$\begin{aligned}\xi &= (k-1)3 + m \\ \eta &= (q-1)3 + n. \end{aligned} \tag{7.34}$$

Denote $Q_{\xi\eta}$ as the $\xi\eta$th element of matrix Q_0. Then according to (7.22), it can be written

$$\begin{aligned}Q_{\xi\eta} &= \frac{1}{N}\sum_{i=1}^{N} \frac{\partial^2}{\partial\alpha_{km}(1)\partial\alpha_{qn}(1)} tr\{P_i^{\perp}R_x(i)\} \\ &= \frac{1}{N}\sum_{i=1}^{N} \frac{\partial^2}{\partial\alpha_{km}(1)\partial\alpha_{qn}(1)} \left[tr\{P_i^{\perp}A(i)R_s(i)A^H(i)\} + (M-K)\sigma_w^2\right] \\ &= \frac{1}{N}\sum_{i=1}^{N} tr\{\frac{\partial^2 P_i^{\perp}}{\partial\alpha_{km}(1)\partial\alpha_{qn}(1)} A(i)R_s(i)A^H(i)\}. \end{aligned} \tag{7.35}$$

The second-order derivative of the projection matrix with respect to $\alpha_{km}(1)$ and $\alpha_{qn}(1)$ can be obtained by the chain rule as

$$\frac{\partial^2 P_i^{\perp}}{\partial \alpha_{km}(1)\partial \alpha_{qn}(1)} = \frac{\partial}{\partial \alpha_{qn}(1)}\left(\frac{\partial P_i^{\perp}}{\partial \theta_k(i)}\frac{\partial \theta_k(i)}{\partial \alpha_{km}(1)}\right)$$
$$= \frac{\partial \theta_k(i)}{\partial \alpha_{km}(1)}\frac{\partial \theta_q(i)}{\partial \alpha_{qn}(1)}\frac{\partial^2 P_i^{\perp}}{\partial \theta_k(i)\partial \theta_q(i)} + \frac{\partial^2 \theta_k(i)}{\partial \alpha_{km}(1)\partial \alpha_{qn}(1)}\frac{\partial P_i^{\perp}}{\partial \theta_k(i)}, \quad (7.36)$$

and then we have

$$Q_{\xi\eta} = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial \theta_k(i)}{\partial \alpha_{km}(1)}\frac{\partial \theta_q(i)}{\partial \alpha_{qn}(1)} tr\{\frac{\partial^2 P_i^{\perp}}{\partial \theta_k(i)\partial \theta_q(i)} A(i)R_s(i)A^H(i)\}$$
$$+\frac{1}{N}\sum_{i=1}^{N}\frac{\partial^2 \theta_k(i)}{\partial \alpha_{km}(1)\partial \alpha_{qn}(1)} tr\{\frac{\partial P_i^{\perp}}{\partial \theta_k(i)} A(i)R_s(i)A^H(i)\}. \quad (7.37)$$

The first-order derivatives of the projection matrix $P_i^{\perp}$ with respect to $\theta_k(i)$ and $\theta_q(i)$ are shown in (6.6), and the second-order derivative has been derived in [57] [53] as

$$\frac{\partial^2 P_i^{\perp}}{\partial \theta_k \partial \theta_q} = P_i^{\perp}A_q(i)A^{\dagger}(i)A_k(i)A^{\dagger}(i) + A^{\dagger H}(i)A_q^H(i)P_i^{\perp}A_k(i)A^{\dagger}(i)$$
$$-P_i^{\perp}A_{kq}(i)A^{\dagger}(i) - P_i^{\perp}A_k(i)(A^H(i)A(i))^{-1}A_q^H(i)P_i^{\perp}$$
$$+P_i^{\perp}A_k(i)A^{\dagger}(i)A_q(i)A^{\dagger}(i) + (\ldots)^H, \quad (7.38)$$

where $(\ldots)^H$ denotes complex conjugate transposes of all proceding terms. $A_{kq}(i)$ is defined as

$$A_{kq}(i) = \frac{\partial^2 A(i)}{\partial \theta_k \partial \theta_q}. \quad (7.39)$$

From (6.6) and (7.38), we obtain

$$Q_{\xi\eta} = \frac{2}{N}\sum_{i=1}^{N}\frac{\partial \theta_k(i)}{\partial \alpha_{km}(1)}\frac{\partial \theta_q(i)}{\partial \alpha_{qn}(1)} Re[tr\{A_q^H(i)P_i^{\perp}A_k(i)R_s(i)\}]$$
$$= \frac{2}{N}\sum_{i=1}^{N} v_{mk}(i)v_{nq}(i) Re[tr\{A_q^H(i)P_i^{\perp}A_k(i)R_s(i)\}]. \quad (7.40)$$

in which we have used $P_i^{\perp}A(i) = 0$ and $A^{\dagger}(i)A(i) = \Gamma$. In matrix form, Q_0 can be written as

$$Q_0 = \frac{2}{N}\sum_{i=1}^{N}[vecV(i)vecV^H(i)] \odot Re\left\{\{[D^H(i)P_i^{\perp}D(i)] \odot R_s^T(i)\} \otimes \mathbf{O}\right\}. \tag{7.41}$$

In the following, we summarize the above results by introducing **Lemma 3.**

Lemma 3 The maximum likelihood MTS estimate $\underline{\hat{\alpha}}(1)$ is asymptotically unbiased, and the asymptotic covariance matrix of $\underline{\hat{\alpha}}(1)$ is given by

$$\text{Cov}[\underline{\hat{\alpha}}(1)] = \frac{\sigma_w^2}{2}C^{-1} + \frac{\sigma_w^4}{4}C^{-1}BC^{-1}, \tag{7.42}$$

where the $\xi\eta$th element of matrix B is given as

$$B_{\xi\eta} = \sum_{i=1}^{N} tr\{B_{km}(i)B_{qn}^H(i)\}. \tag{7.43}$$

and the matrix C is defined as

$$C = \sum_{i=1}^{N}[vecV(i)vecV^H(i)] \odot Re\left\{\{[D^H(i)P_i^{\perp}D(i)] \odot R_s^T(i)\} \otimes \mathbf{O}\right\}. \tag{7.44}$$

Note that all the matrix functions are evaluated at the true initial target state value $\underline{\alpha}'(1)$.

Proof The asymptotic unbiasedness can be shown by taking expectation of (7.26) and neglecting high order terms in $\Delta R_x(i)$

$$E\{\Delta\underline{\alpha}(1)\} = E\{Q_0^{-1}\Delta P\} = 0. \tag{7.45}$$

Since matrices P_d and Q_0 can be expressed in term of C as

$$\begin{aligned} P_d &= \frac{2\sigma_w^2}{N^2}C + \frac{\sigma_w^4}{N^2} \\ Q_0 &= \frac{2}{N}C, \end{aligned} \tag{7.46}$$

the covariance matrix of $\underline{\hat{\alpha}}(1)$ can then be written as

$$\begin{aligned} \mathrm{Cov}[\underline{\hat{\alpha}}(1)] &= E[\underline{\hat{\alpha}}(1)\underline{\hat{\alpha}}^H(1)] \\ &= Q_0^{-1}E\{\Delta P \Delta P^H\}[Q_0^{-1}]^H \\ &= Q_0^{-1}P_d[Q_0^{-1}]^H \\ &= \frac{\sigma_w^2}{2}(C)^{-1} + \frac{\sigma_w^4}{4}C^{-1}BC^{-1}, \end{aligned} \tag{7.47}$$

□

In (7.47), it is observed that, when σ_w is sufficiently small, the higher order term in σ_w^4 is negligible, and the asymptotic covariance matrix of $\underline{\hat{\alpha}}(1)$ has a form identical (up to a scaling factor) to the inverse of the Hessian matrix of the limiting criterion function $\overline{J}$, except that the covariance matrix is evaluated at the true initial MTS vector value. Define matrix $C(i)$ as,

$$C(i) = [vecV(i)vecV^H(i)] \odot Re\left\{\{[D^H(i)P_i^{\perp}D(i)] \odot R_s^T(i)\} \otimes \mathbf{O}\right\}. \tag{7.48}$$

Since $\underline{\alpha}(1)$ is the minimizing argument of the limiting criterion function $\overline{J}$ and $C(i)$ is the Hessian of $\overline{J}$ evaluated at $\underline{\alpha}(1)$ up to a scale factor, then, the positivity of $C(i)$ is readily established. Define a scalar function φ as

$$\varphi(W) = tr\{W^{-1}\}, \quad W \in P^{3K\times 3K}, \tag{7.49}$$

where $P^{3K\times 3K}$ denotes the closed convex set of $3K \times 3K$ positive definite matrices. Then, it has been shown [62] that $\varphi(W)$ is a convex function in the convex region $P^{3K\times 3K}$. Considering the trace of the covariance matrix $Cov[\underline{\hat{\alpha}}(1)]$, we have

$$tr\{\mathrm{Cov}[\underline{\hat{\alpha}}(1)]\} = \frac{\sigma_w^2}{2N}tr\{[\frac{1}{N}\sum_{i=1}^{N}C(i)]^{-1}\} = \frac{\sigma_w^2}{2N}\varphi[\frac{1}{N}\sum_{i=1}^{N}C(i)]. \tag{7.50}$$

For a convex function φ and positive $C(i)$, the following inequalty holds

$$\frac{\sigma_w^2}{2N}\varphi[\frac{1}{N}\sum_{i=1}^{N}C(i)] \le \frac{\sigma_w^2}{2N^2}\sum_{i=1}^{N}\varphi[C(i)]. \tag{7.51}$$

Assume that the sequence $\{\varphi[C(i)];\ i = 1, 2, \ldots, N\}$ is bounded by a constant c_u. Then we have

$$tr\{\text{Cov}[\underline{\hat{\alpha}}(1)]\} \leq \frac{\sigma_w^2 c_u}{2N}. \tag{7.52}$$

As N approaches infinity, we have the following limit

$$\lim_{N\to\infty} tr\{\text{Cov}[\underline{\hat{\alpha}}(1)]\} \leq \lim_{N\to\infty} \frac{\sigma_w^2 c_u}{2N} = 0. \tag{7.53}$$

Since $tr\{\text{Cov}[\underline{\hat{\alpha}}(1)]\}$ is the sum of all the variances of the estimated initial MTS vector elements

$$tr\{\text{Cov}[\underline{\hat{\alpha}}(1)]\} = \sum_{k=1}^{3K} \text{var}[\hat{\alpha}_k(1)] \geq 0, \tag{7.54}$$

and $\text{var}[\hat{\alpha}_k(1)] \geq 0$ for $k = 1, 2, \ldots, 3K$, then,

$$\lim_{N\to\infty} tr\{\text{Cov}[\underline{\hat{\alpha}}(1)]\} = 0 \implies \lim_{N\to\infty} \text{var}[\hat{\alpha}_k(1)] = 0. \tag{7.55}$$

Relationship (7.55) indicates that the variance of each component of the MTS estimate tends to zero when the array data sample number approaches infinity. This also provides an alternative derivation for the asymptotic consistency of the estimator stated earlier.

The estimates $\{\underline{\hat{\alpha}}(j);\ j = 2, 3, \ldots, N\}$ can be obtained either by the direct maximum likelihood algorithm or by prediction using the source target dynamics from the ML estimate of $\underline{\hat{\alpha}}(1)$. Under the conditions of large N and sufficiently high SNR, we can write the asymptotic MSE for the estimated $\underline{\hat{\alpha}}(j)$ by similar derivations as in the case for $\underline{\hat{\alpha}}(1)$

$$\text{Cov}[\underline{\hat{\alpha}}(j)] = \frac{\sigma_w^2}{2} D^{-1}, \tag{7.56}$$

where matrix D is given by

$$D = \sum_{i=1}^{N} [vecV(i,j)vecV^H(i,j)] \odot Re\left\{\{[D^H(i)P_i^{\perp}D(i)] \odot R_s^T(i)\} \otimes \mathbf{O}\right\}. \tag{7.57}$$

and where matrix $V(i,j)$ is defined in (6.39). Obviously, since $\underline{\hat{\alpha}}(j)$ is the ML estimate of the corresponding target state $\underline{\alpha}(j)$, the asymptotic consistency statement is also true.

7.3 The Cramér-Rao lower bound (CRB)

It is known that the variance of unbiased estimates provided by any estimator is limited by the Cramér-Rao lower bound [47]. The comparison between the CRB and the variance of the estimates will give a measure of statistical efficiency of the estimator, which is another important index of performance.

7.3.1 Derivation of the CRB

The Cramér-Rao lower bound can be derived from the Fisher information matrix [47] defined as the second derivative of the log likelihood function with respect to the parameters. Let $\Theta'(i)$, $\underline{s}'(i)$ and $\sigma_w'^2$ denote the true values of the source parameters. From (5.24), the log likelihood function can be written as

$$\Lambda(\Theta(i), \underline{s}(i), \sigma_w^2;\ \ i = 1, 2, \ldots, N) =$$
$$\log p = -MN\log(\pi\sigma_w^2) - \frac{1}{\sigma_w^2}\sum_{i=1}^{N} \| \underline{x}(i) - A(i)\underline{s}(i) \|_F^2 . \tag{7.58}$$

For fixed values of $\{\Theta(i)\}$, the maximization of Λ with respect to $\{\underline{s}(i)\}$ can be achieved by choosing

$$\hat{\underline{s}}(i) = \{A^H(i)A(i)\}^{-1}A^H(i)\underline{x}(i), \tag{7.59}$$

Substituting (7.59) back into (7.58) gives a reduced expression of the log likelihood function of σ_w^2 and $\{\Theta(i)\}$ only

$$\begin{aligned} \Lambda &= -MN\log(\pi\sigma_w'^2) - \frac{1}{\sigma_w^2}\sum_{i=1}^{N} | \underline{x}(i) - P_{A(i)}\underline{x}(i) |_F^2 \\ &= -MN\log(\pi\sigma_w^2) - \frac{1}{\sigma_w^2}\sum_{i=1}^{N} tr\{P_{A(i)}^{\perp}\hat{R}_x(i)\}. \end{aligned} \tag{7.60}$$

The CRB for the covariance matrix of the estimate $\underline{\alpha}(1)$ and σ_w^2 is given by the inverse of the Fisher informatrix matrix

$$\text{CRB}([\underline{\alpha}(i), \sigma_w^2]) = \{I([\underline{\alpha}(i), \sigma_w^2])\}^{-1}, \tag{7.61}$$

and the Fisher information matrix $I([\underline{\alpha}(1), \sigma_w^2])$ can be partitioned as

$$I([\underline{\alpha}(1), \sigma_w^2])_{(3K+1)\times(3K+1)} = \begin{bmatrix} I(\underline{\alpha}(1))_{3K\times 3K} & I(\underline{\alpha}(1), \sigma_w^2)_{3K\times 1} \\ I(\sigma_w^2, \underline{\alpha}(1))_{1\times 3K} & I(\sigma_w^2)_{1\times 1} \end{bmatrix}, \tag{7.62}$$

in which the elements of $I(\underline{\alpha}(1), \sigma_w^2)$, $I(\sigma_w^2, \underline{\alpha}(1))$, $I(\underline{\alpha}(1))$ and $I(\sigma_w^2)$ are defined as

$$\begin{aligned} I(\underline{\alpha}(1), \sigma_w^2)_{\xi 1} &= -\frac{\partial^2}{\partial \alpha_{km}(1) \partial \sigma_w^2} E[\Lambda] \mid_{\underline{\alpha}'(1), \sigma_w'^2} \qquad \xi = 1, 2, \ldots, 3K \\ I(\sigma_w^2, \underline{\alpha}(1))_{1\xi} &= -\frac{\partial^2}{\partial \sigma_w^2 \partial \alpha_{km}(1)} E[\Lambda] \mid_{\underline{\alpha}'(1), \sigma_w'^2} \qquad \xi = 1, 2, \ldots, 3K \\ I(\underline{\alpha}(1))_{\xi\eta} &= -\frac{\partial^2}{\partial \alpha_{km}(1) \partial \alpha_{qn}(1)} E[\Lambda] \mid_{\underline{\alpha}'(1), \sigma_w'^2} \quad \xi, \eta = 1, 2, \ldots, 3K \\ I(\sigma_w^2) &= -\frac{\partial^2}{\partial \sigma_w^2 \partial \sigma_w^2} E[\Lambda] \mid_{\underline{\alpha}'(1), \sigma_w'^2}, \end{aligned} \tag{7.63}$$

respectively, where indices ξ, η, m and n are related by

$$\xi = (k-1)3 + m \text{ and } \eta = (q-1)3 + n. \tag{7.64}$$

Taking the expectation of the reduced log likelihood function (7.60), we obtain

$$E[\Lambda] = -MN \log(\pi \sigma_w^2) - \frac{1}{\sigma_w^2} \sum_{i=1}^{N} tr\{P_{A(i)}^{\perp} R_x(i)\}, \tag{7.65}$$

and substituting (7.2) into (7.65) yields

$$\begin{aligned} E[\Lambda] &= -MN \log(\pi \sigma_w^2) - \frac{\sigma_w'^2}{\sigma_w^2}(M-K) - \frac{1}{\sigma_w^2} \sum_{i=1}^{N} tr\{P_{A(i)}^{\perp} A'(i) R_s(i) A'^H(i)\} \\ &= -MN \log(\pi \sigma_w^2) - \frac{\sigma_w'^2}{\sigma_w^2}(M-K) - \frac{1}{\sigma_w^2} \sum_{i=1}^{N} tr\{A'^H(i) P_{A(i)}^{\perp} A'(i) R_s(i)\}, \end{aligned} \tag{7.66}$$

where $A'(i)$ denotes the composite steering matrix $A(i)$ evaluated at $\Theta'(i)$. In (7.66), we have used the property of the trace $tr\{AB\} = tr\{BA\}$. With the expectation $E[\Lambda]$, we can demonstrate that

that the submatrices, $I(\underline{\alpha}(i), \sigma_w^2)$ and $I(\sigma_w^2, \underline{\alpha}(i))$, in the Fisher information matrix corresponding to cross terms $\underline{\alpha}(1)$ and σ_w^2 are zeros, i.e.,

$$I(\underline{\alpha}(i), \sigma_w^2) = 0 \ \ and \ \ I(\sigma_w^2, \underline{\alpha}(i)) = 0. \tag{7.67}$$

We have therefore

$$I([\underline{\alpha}(1), \sigma_w^2])_{(3K+1)\times(3K+1)} = \begin{bmatrix} I(\underline{\alpha}(1))_{3K\times 3K} & \mathbf{0}_{3K\times 1} \\ \mathbf{0}_{1\times 3K} & I(\sigma_w^2)_{1\times 1} \end{bmatrix}. \tag{7.68}$$

Since the entries of the Fisher information matrix corresponding to $\underline{\alpha}(1)$ are decoupled from the entries corresponding to σ_w^2, we get the CRB for the variance of the estimate $\underline{\alpha}(1)$ as

$$\mathrm{CRB}[\underline{\alpha}(1)] = [I(\underline{\alpha}(1))]^{-1}, \tag{7.69}$$

where the $\xi\eta$th element of $I[\underline{\alpha}(1)]$ is computed as

$$\left\{\mathrm{CRB}^{-1}[\underline{\alpha}(1)]\right\}_{\xi\eta} = \frac{1}{\sigma_w'^2}\sum_{i=1}^{N} \frac{\partial^2}{\partial\alpha_{km}(1)\partial\alpha_{qn}(1)} tr\{A'^H(i)P_{A(i)}^{\perp}A(i)R_s(i)\}\mid_{\underline{\alpha}'(1)}. \tag{7.70}$$

Lemma 4 The Cramér-Rao bound for the estimated initial MTS vector $\underline{\alpha}(1)$ is given by

$$\mathrm{CRB}[\underline{\alpha}(1)] = \frac{\sigma_w^2}{2}C^{-1}, \tag{7.71}$$

where matrix C is defined in (7.44).

Proof Since the matrix $\mathrm{CRB}[\underline{\alpha}(1)]$ can be written as

$$\{\mathrm{CRB}[\underline{\alpha}(1)]\}^{-1} = \frac{N}{\sigma_w^2}Q_0, \tag{7.72}$$

the result follows immediately from (7.46). □

7.3.2 Relative efficiency discussions about the ML estimator

In Lemma 3, we have shown that $\underline{\hat{\alpha}}(1)$ is an asymptotically unbiased estimate. The Cramér-Rao inequality for an unbiased estimate $\underline{\hat{\alpha}}(1)$ can be stated as

$$E\{[\underline{\hat{\alpha}}(1) - \underline{\alpha}(1)][\underline{\hat{\alpha}}(1) - \underline{\alpha}(1)]^H\} \geq \text{CRB}[\underline{\alpha}(1)]. \tag{7.73}$$

An estimate is said to be asymptotically efficient if its covariance achieves the Cramér-Rao bound when the sample number approaches infinity. Specifically, $\underline{\hat{\alpha}}(1)$ is an asymptotically efficient estimate if

$$\lim_{N\to\infty} \text{Cov}[\underline{\hat{\alpha}}(1)]\text{CRB}^{-1}[\underline{\hat{\alpha}}(1)] \longrightarrow I. \tag{7.74}$$

To study the efficiency of the maximum likelihood estimator, we introduce the concept of relative efficiency. An estimate $\underline{\hat{\alpha}}(1)$ is said to be asymptotically relatively efficient if

$$\lim_{\sigma_w'^2\to 0}\lim_{N\to\infty} \text{Cov}[\underline{\hat{\alpha}}(1)]\text{CRB}^{-1}[\underline{\hat{\alpha}}(1)] \longrightarrow I. \tag{7.75}$$

For the maximum likelihood estimator, we are ready to state the following results.

Lemma 5 The ML estimate of $\underline{\hat{\alpha}}(1)$ obtained from (5.35) is not asymptotically efficient, i.e., its asymptotic variance does not attain the Cramér-Rao bound. The asymptotic covariance matrix and the Cramér-Rao bound have the following relationship

$$\text{Cov}[\underline{\hat{\alpha}}(1)] = \text{CRB}[\underline{\alpha}(1)] + \frac{\sigma_w'^4}{4} Q^{-1} B Q^{-1}, \tag{7.76}$$

However, the estimate $\underline{\hat{\alpha}}(1)$ is relatively asymptotically efficient, i.e., when the array sample number N approaches infinity and the array noise variance σ_w^2 approaches zero, the following limit holds

$$\lim_{N\to\infty}\lim_{\sigma_w'^2\to 0} \text{CRB}^{-1}[\underline{\alpha}(1)]\text{Cov}[\underline{\hat{\alpha}}(1)] = I, \tag{7.77}$$

which means that the covariance matrix of $\underline{\hat{\alpha}}(1)$ asymptotically attains the Cramér-Rao bound when SNR is sufficiently high.

Proof From (7.42) and (7.71), the asymptotic covariance matrix of $\underline{\hat{\alpha}}(1)$ can be expressed as

$$\mathrm{Cov}[\underline{\hat{\alpha}}(1)] = \frac{\sigma_w'^2}{2}C^{-1} + \frac{\sigma_w'^4}{4}C^{-1}BC^{-1} = CRB[\underline{\hat{\alpha}}(1)] + \frac{\sigma_w'^4}{4}C^{-1}BC^{-1}. \tag{7.78}$$

When N approaches infinity, we have

$$\begin{aligned}\lim_{N\to\infty} \mathrm{CRB}^{-1}[\underline{\alpha}(1)]\mathrm{Cov}[\underline{\hat{\alpha}}(1)] &= \frac{N^2}{\sigma_w'^2}C[\frac{\sigma_w'^2}{2}C^{-1} + \frac{\sigma_w'^4}{4}C^{-1}BC^{-1}] \\ &= I + \frac{\sigma_w'^2}{2}BC^{-1}, \end{aligned} \tag{7.79}$$

which, from the definition of the asymptotic efficiency (7.75), shows that the estimate $\underline{\hat{\alpha}}(1)$ is not asymptotically efficient.

Since the first and second terms in $\mathrm{Cov}[\underline{\hat{\alpha}}(1)]$ are of order two and four in σ_w', respectively, for sufficiently high SNR, the quantity of the second term is much less significant than the first term. Assuming that $\sigma_w'^2$ approaches zero, consider the following limit

$$\begin{aligned}\lim_{N\to\infty}\lim_{\sigma_w'^2\to 0} \mathrm{CRB}^{-1}[\underline{\alpha}(1)]\mathrm{Cov}[\underline{\hat{\alpha}}(1)] &= I + \lim_{\sigma_w'^2\to 0}\frac{\sigma_w^2}{2}BC^{-1} \\ &= I, \end{aligned} \tag{7.80}$$

which indicates the relative asymptotic efficiency of the maximum likelihood estimator. □

The Cramér-Rao bound for the estimates $\underline{\hat{\alpha}}(j)$ for $j = 2, 3, \ldots, N$ can be derived as

$$\left\{\mathrm{CRB}^{-1}[\underline{\alpha}(j)]\right\}_{\xi\eta} = \frac{1}{\sigma_w'^2}\sum_{i=1}^{N}\frac{\partial^2}{\partial\alpha_{km}(j)\partial\alpha_{qn}(j)}tr\{A'^H(i)P^{\perp}_{A(i)}A'(i)R_s(i)\}\mid_{\underline{\alpha}(j)=}\frac{\sigma_w'^2}{2}D^{-1}, \tag{7.81}$$

where matrix D is given in (7.57). By comparing expressions (7.56) and (7.81), we reach the conclusion that, the maximum likelihood estimate $\underline{\hat{\alpha}}(j)$ is also relatively asymptotically efficient,

i,e., when both N and SNR are sufficiently large, the asymptotic covariance of the ML estimate $\hat{\underline{\alpha}}(j)$ attains the Cramér-Rao bound.

7.4 Numerical studies

Computer simulations are carried out for evaluating the performance of the maximum likelihood tracking algorithm. The array is simulated as an equi-spaced linear array of $M = 8$ sensors with half source wavelength spacing, as shown in Fig. 6.1. Two narrow-band source targets of equal power are assumed, with their initial DOA locations $\theta_1(1) = 0°$ and $\theta_2(1) = -20°$. The initial q components of the source MTS vector are set to $q_1(1) = 7.4 \times 10^{-3}$ and $q_2(1) = 5.3 \times 10^{-3}$, respectively. The results are obtained through Monte-Carlo simulations. Fig. 7.1, Fig. 7.2 and Fig. 7.3 show the variation of the MSE's of the MTS estimates with the SNR. The comparison is between the theoretically evaluated MSE's and that obtained by computer simulations. The SNR ranges from -4dB to 26dB. For each simulation, 100 array samples were used and each test is repeated 100 times. In the figure, it is seen that the simulation results agree with the theoretical results. As the SNR increases, the MSE of the estimates approach the CRB on the estimates, which confirms the theoretical assertion from **Lemma 5** that the estimates are asymptotically relatively efficient.

Fig. 7.4, Fig. 7.5 and Fig. 7.6 show the variation of MSE's of the source MTS estimates with the number of array samples. The SNR is chosen at 5dB. The initial source MTS vector is selected as : $\theta_1(1) = 0°$ and $\theta_2(1) = -20°$, $q_1(1) = 7.4 \times 10^{-3}$ and $q_2(1) = 5.3 \times 10^{-3}$, and the heading directions of motion $\phi_1(1) = 135°$ and $\phi_2(1) = 30°$. Each test is repeated 100 times to obtain the Monte-Carlo results. Again, the theoretical performance curves are in close agreement with those obtained by simulations. The consistency of the estimates can be observed as the array sample number increases, the MSE's of the estimates both evaluated theoretically and by simulations approach zero consistently.

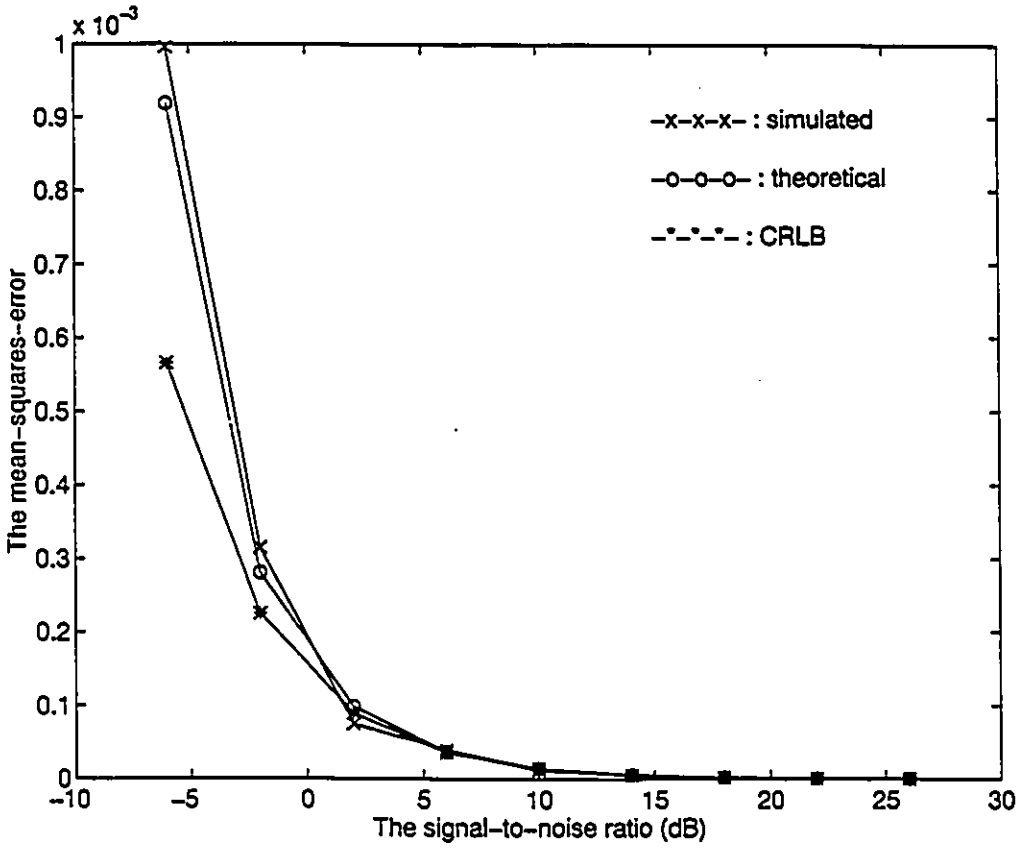

Figure 7.1: The variation of the MSE of the DOA estimate with SNR.

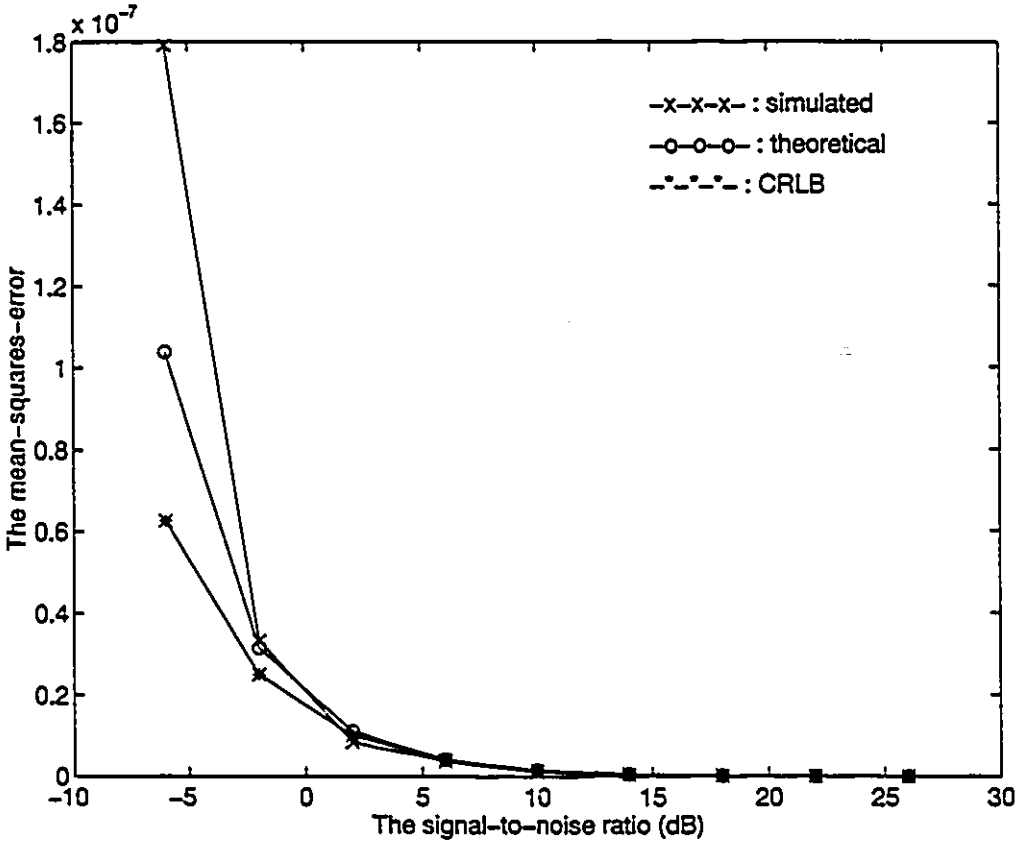

Figure 7.2: The variation of the MSE of q component estimates with SNR

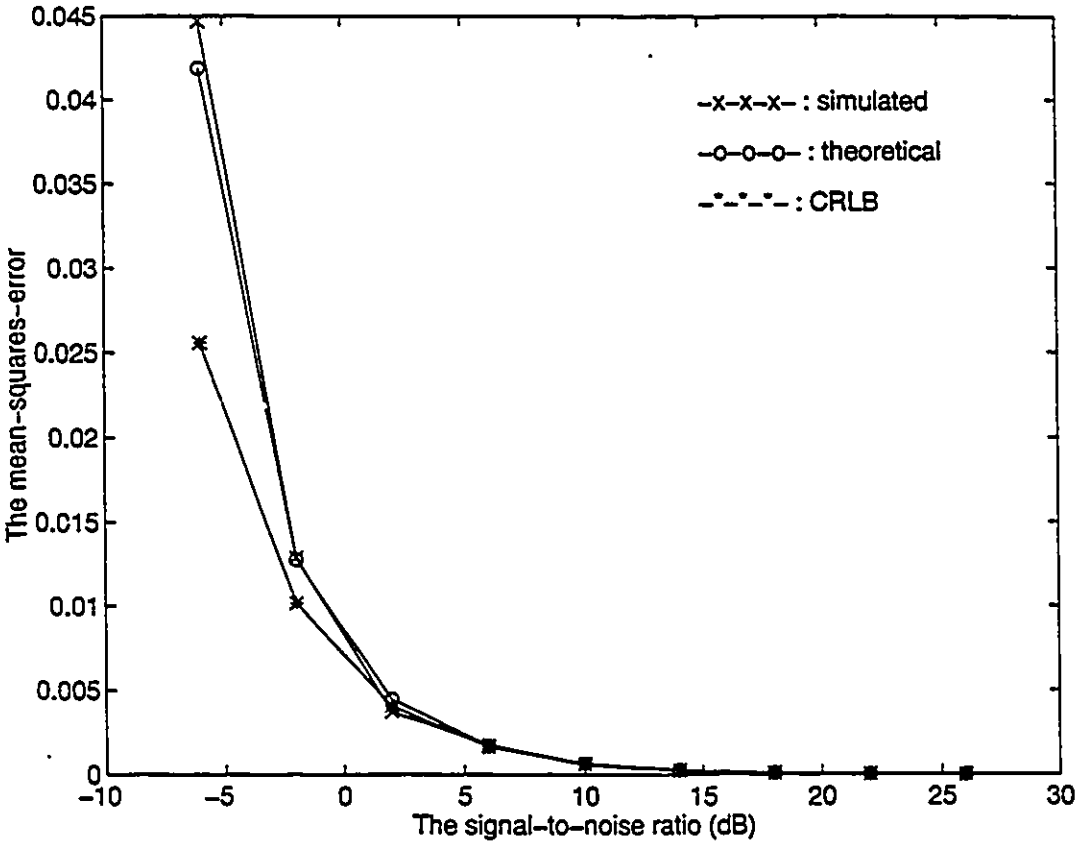

Figure 7.3: The variation of the MSE of the ϕ component estimate with SNR.

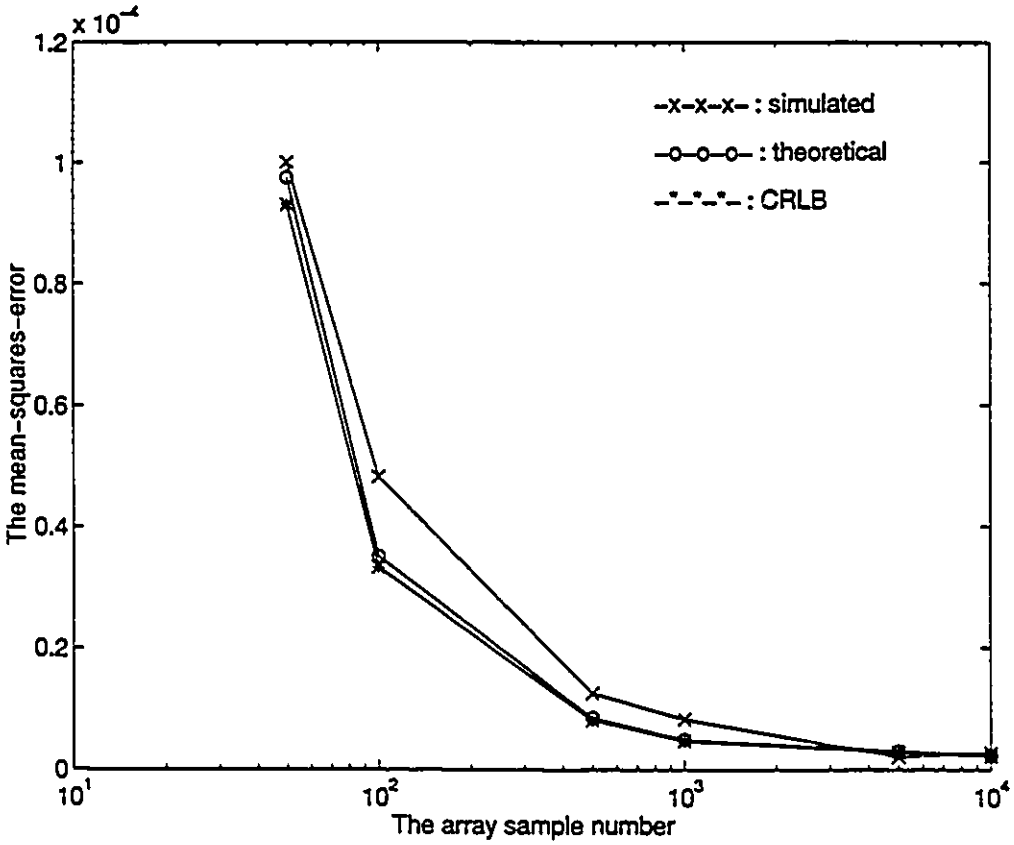

Figure 7.4: The variation of the MSE of the DOA estimate with number of array samples.

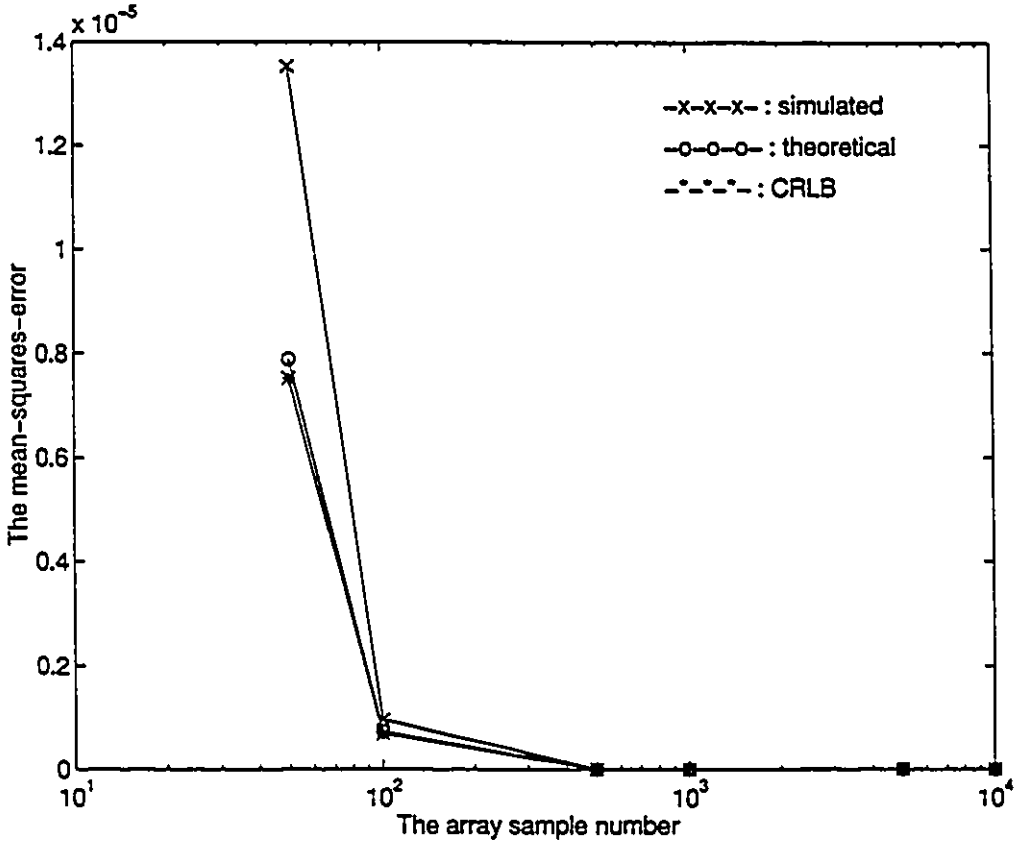

Figure 7.5: The variation of the MSE of q component estimates with the number of array samples.

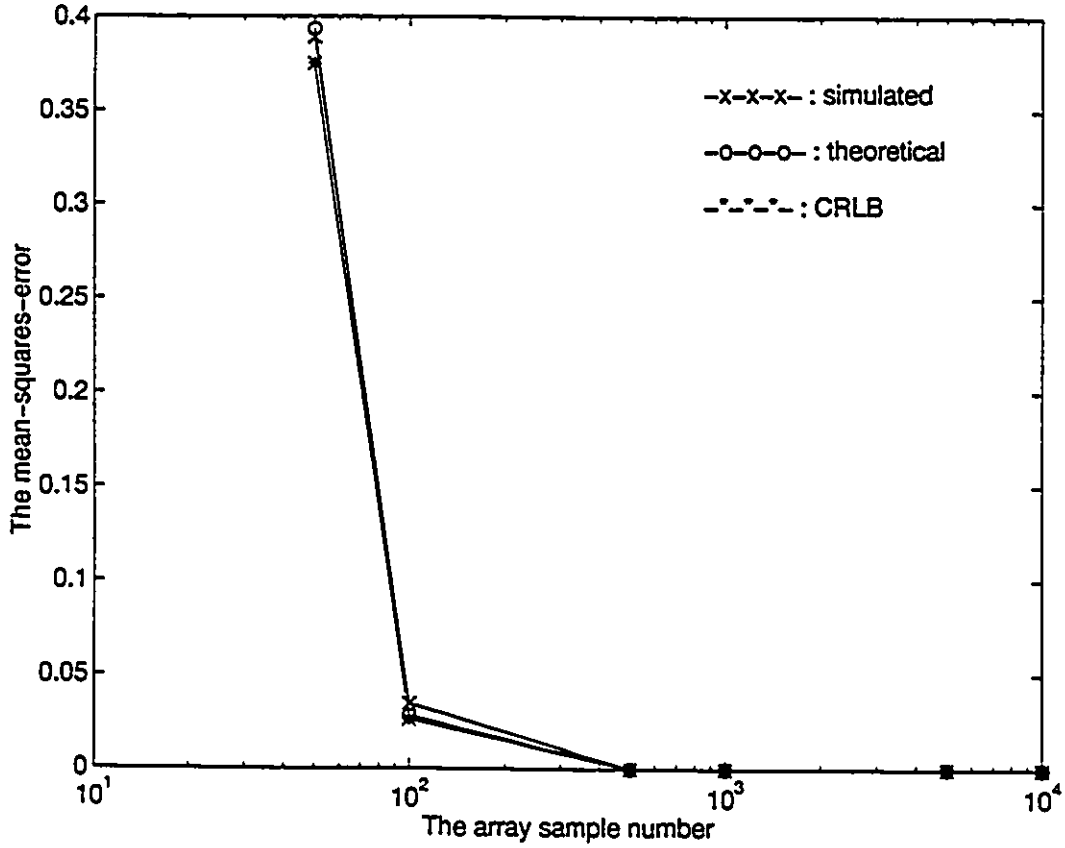

Figure 7.6: The variation of the MSE of the ϕ component estimate with the number of array samples.

Chapter 8

Asymptotic Performance Comparisons with the Extended Kalman Filter

The maximum likelihood estimator is based on a batch of received array data. It estimates the source parameters by iterating over the same batch of data in attempting to minimize the criterion function. This algorithm is fundamentally different from some recursive algorithms such as the extended Kalman filter [63]. As mentioned before, since the array signal model (5.14) contains the unknown deterministic target signal waveforms, it cannot be used as a measurement model directly for the Kalman filter process. In this chapter, our analysis is based on the assumption that the source signal waveforms are known. On one hand, such an assumption is made to provide a fair comparison between the ML and EKF. On the other hand, although such an application does not usually occur in passive array processing, we do find some applications in communication systems, such as the packet radio systems currently under study [64]. In packet radio systems, each packet contains a known pseudo-noise (PN) code acquisition code preamble. An antenna array is used to estimate the packet DOA's, and then to spatially separate colliding packets arriving from different directions.

8.1 The maximum likelihood estimator with known source target waveforms

When the source signal vector $\underline{s}(i)$ is assumed to be known, the array data likelihood function is described by (5.24). Fixing $\Theta(i)$ and minimizing the negative log likelihood function with respect to σ_w^2 yields the estimate

$$\hat{\sigma}_w^2 = \frac{1}{MN}\sum_{i=1}^{N} \| \underline{x}(i) - A(i)\underline{s}(i) \|_F^2 . \tag{8.1}$$

Since $\underline{s}(i)$ is known, there is no need to estimate the source target waveform. By substituting $\hat{\sigma}_w^2$ back into (5.25) and ignoring the constant term, we obtain the likelihood function J_s as

$$\begin{aligned} J_s &= \frac{1}{N}\sum_{i=1}^{N} \| \underline{x}(i) - A(i)\underline{s}(i) \|_F^2 \\ &= \frac{1}{N}\sum_{i=1}^{N} tr\{[\underline{x}(i) - A(i)\underline{s}(i)][\underline{x}(i) - A(i)\underline{s}(i)]^H\} \\ &= \frac{1}{N}\sum_{i=1}^{N} J_{si}. \end{aligned} \tag{8.2}$$

J_s is an implicit function of the MTS variables, and the maximum likelihood estimate of the MTS vector $\underline{\alpha}(j)$ can be obtained by minimizing the criterion J_s with respect to $\underline{\alpha}(j)$

$$\hat{\underline{\alpha}}(j) = \arg\min_{\underline{\alpha}(j)} J_s. \tag{8.3}$$

8.1.1 Optimizing the criterion function J_s

When a Newton-type algorithm is applied to the optimization process (8.3), the iteration process is formulated as

$$\hat{\underline{\alpha}}^{(k+1)}(j) = \hat{\underline{\alpha}}^{(k)}(j) - \mu_k H'^{-1} G', \tag{8.4}$$

where G' and H' are the gradient and Hessian of J_s with respect to the MTS vector $\underline{\alpha}(j)$, respectively. The gradient and Hessian, G' and H', can be derived as

$$G' = vec\{\frac{1}{N}\sum_{i=1}^{N} V(i,j)diag(\nabla J_{si})\}$$

$$H' = \frac{1}{N}\sum_{i=1}^{N}\Big\{[\nabla^2 J_{si}\otimes \mathbf{O}]\odot[vecV(i,j)vec^H V(i,j)] + [diag(\nabla J_{si})\otimes I]\odot\Lambda(i)\Big\}, \quad (8.5)$$

where ∇J_{si} and $\nabla^2 J_{si}$ are the gradient and Hessian of J_{si} with respect to the DOA state $\Theta(i)$. The mth element of ∇J_{si} is computed as

$$\begin{aligned}\frac{\partial J_{si}}{\partial\theta_m(i)} &= 2Re\{tr[A(i)\underline{s}(i)\underline{s}^H(i)A_m^H(i)] - tr[\underline{x}(i)\underline{s}^H(i)A_m^H(i)]\}\\ &= 2Re\{tr\{[A(i)\underline{s}(i)\underline{s}^H(i) - \underline{x}(i)\underline{s}^H(i)]A_m^H(i)\}\}, \end{aligned} \quad (8.6)$$

and in matrix form, ∇J_{si} can be written as

$$\nabla J_{si} = \frac{\partial J_{si}}{\partial\Theta(i)} = 2Re\{diag\{D^H(i)[A(i)R_s(i) - \underline{x}(i)\underline{s}^H(i)]\}\}, \quad (8.7)$$

where $R_s(i) = \underline{s}(i)\underline{s}^H(i)$. Substituting (8.7) into (8.5) yields the expression for the gradient

$$G' = 2vec\{\frac{1}{N}\sum_{i=1}^{N} V(i,j)Re\{I\odot\{D^H(i)[A(i)R_s(i) - \underline{x}(i)\underline{s}^H(i)]\}\}\}. \quad (8.8)$$

When the Newton-type optimzation algorithm is applied, we need to modify the Hessian matrix to guarantee that the algorithm is descent. In the following, we approximate the Hessian by a positive semi-definite matrix. First, we assume that in each iteration, the MTS estimate $\underline{\alpha}^{(k)}(1)$ is close enough to the optimal $\hat{\underline{\alpha}}(1)$. Hence, the derivative term is sufficiently small, i.e.,

$$\nabla J_{si} \simeq 0, \quad (8.9)$$

and the second term in (8.5) can be neglected. The mnth element of $\nabla^2 J_{si}$ is given by

$$\begin{aligned}\frac{\partial^2 J_{si}}{\partial\theta_m(i)\partial\theta_n(i)} &= \frac{\partial}{\partial\theta_m(i)}\{\frac{\partial J_{si}}{\partial\theta_n(i)}\}\\ &= \frac{\partial}{\partial\theta_m(i)}\Big\{2Re\{tr\{[A(i)R_s(i) - \underline{x}(i)\underline{s}^H(i)]A_n^H(i)\}\}\Big\}\end{aligned}$$

$$= 2Re\{tr[A_m(i)R_s(i)A_n^H(i)]\}$$
$$+2Re\{tr[A(i)R_s(i)A_{mn}^H(i) - \underline{x}(i)\underline{s}^H(i)A_{mn}^H(i)]\}. \quad (8.10)$$

Equation (8.10) can be further simplified as

$$\frac{\partial^2 J_{si}}{\partial\theta_m(i)\partial\theta_n(i)} = 2Re\{tr[A_m(i)R_s(i)A_n^H(i)]\} - 2Re\{tr[\underline{w}(i)\underline{s}^H(i)A_{mn}^H(i)]\}. \quad (8.11)$$

By taking (8.11) back into (8.5) and ignoring the second term, we get

$$\begin{aligned} H' &= H_1 + H_2 \\ &= \frac{1}{N}\sum_{i=1}^{N}[\nabla^2 J_{si}^{(1)} \otimes \mathbf{O}] \odot [vecV(i,j)vec^H V(i,j)] \\ &\quad + \frac{1}{N}\sum_{i=1}^{N}[\nabla^2 J_{si}^{(2)} \otimes \mathbf{O}] \odot [vecV(i,j)vec^H V(i,j)], \end{aligned} \quad (8.12)$$

where $\nabla^2 J_{si}^{(1)}$ and $\nabla^2 J_{si}^{(2)}$ are matrices with their mnth elements defined by

$$\begin{aligned} \{\nabla^2 J_{si}^{(1)}\}_{mn} &= 2Re\{tr[A_m(i)R_s(i)A_n^H(i)]\} \\ \{\nabla^2 J_{si}^{(2)}\}_{mn} &= -2Re\{tr[\underline{w}(i)\underline{s}^H(i)A_{mn}^H(i)]\}, \end{aligned} \quad (8.13)$$

respectively. In the following, we introduce a lemma to show the asymptotic zero covergence of the elements of H_2.

Lemma 6 The elements of matrix H_2 asymptotically converge to zero, $O(1/\sqrt{N})$, with probability one.

Proof : Denote the pth element of $\underline{s}^H(i)A_{mn}^H(i)$ as $\rho_p(mn,i)$. Then the mnth element of $\nabla^2 J_{si}^{(2)}$ can be expressed as

$$\begin{aligned} [\nabla^2 J_{si}^{(2)}]_{mn} &= -2Re\{tr[\underline{w}(i)\underline{s}^H(i)A_{mn}^H(i)]\} \\ &= -2\sum_{p=1}^{M} Re[w_p(i)\rho_p(mn,i)] \end{aligned}$$

$$= -\sum_{p=1}^{M}[w_p(i)\rho_p(mn,i) + w_p^*(i)\rho_p^*(mn,i)]. \tag{8.14}$$

Let $v_{\xi\eta}(i,j)$ denote the $\xi\eta$th element of $vecV(i,j)vec^HV(i,j)$, where ξ, η, m and n are related through relationship (7.34). The $\xi\eta$th element of H_2 can be written as

$$H_{\xi\eta}^{(2)} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{p=1}^{M}[w_p(i)\rho_p(mn,i) + w_p^*(i)\rho_p^*(mn,i)]v_{\xi\eta}(i,j). \tag{8.15}$$

The variance of $H_{\xi\eta}^{(2)}$ can be derived as

$$\begin{aligned}
Var\{H_{\xi\eta}^{(2)}\} &= E\{H_{\xi\eta}^{(2)}[H_{\xi\eta}^{(2)}]^*\} \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{p=1}^{M}\sum_{p'}^{M}E\{[w_p(i)\rho_p(mn,i) + w_p^*(i)\rho_p^*(mn,i)] \\
&\quad \cdot[w_{p'}(i')\rho_{p'}(mn,i') + w_{p'}^*(i')\rho_{p'}^*(mn,i')]v_{\xi\eta}^2(i',j)\} \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{p=1}^{M}\sum_{p'}^{M}E\{[w_p(i)\rho_p(mn,i) + w_p^*(i)\rho_p^*(mn,i)] \\
&\quad \cdot[w_{p'}(i')\rho_{p'}(mn,i') + w_p^*(i')\rho_{p'}^*(mn,i')]\}v_{\xi\eta}(i,j)v_{\xi\eta}(i',j) \\
&= \frac{\sigma_w^2}{N^2}\sum_{i=1}^{N}\sum_{p=1}^{M}v_{\xi\eta}^2(i,j)Re[\rho_p^2(mn,i)],
\end{aligned} \tag{8.16}$$

in which we have applied the results of Appendix B. Assume that each term inside the summation is bounded by a finite constant c_k and we get

$$Var\{H_{\xi\eta}^{(2)}\} \leq \frac{M\sigma_w^2 c_k}{N}, \tag{8.17}$$

which implies that $H_{\xi\eta}^{(2)}$ is of order $O(1/\sqrt{N})$ and converges to zero, $O(1/\sqrt{N})$, with probability one, when the array sample number N approaches infinity. □.

Since the elements of $H^{(2)}$ are of order $O(1/\sqrt{N})$, when the array sample number is sufficiently large, $H^{(2)}$ can be neglected, and the approximation is of order $O(1/\sqrt{N})$. The approximated Hessian becomes

$$H' = \frac{1}{N}\sum_{i=1}^{N}[\nabla^2 J_{si}^{(1)} \otimes \mathbf{O}] \odot [vecV(i,j)vec^HV(i,j)], \tag{8.18}$$

where $\nabla^2 J_{si}^{(1)}$, according to (8.13), can be written in matrix form as

$$\nabla^2 J_{si}^{(2)} = 2Re\{[D^H(i)D(i)] \odot R_s(i)^T\}. \tag{8.19}$$

The approximated Hessian (8.18) can be verified to be positive semi-definite. When the approximate Hessian is used, the Gauss-Newton algorithm is guaranteed to be descent.

8.1.2 Covariance of the ML estimates and the Cramér-Rao bound

Under the assumption of known source target waveform, the ML estimator is asmptotically efficient, i.e, it attains the Cramér-Rao bound when the number of samples is sufficiently large. Thus, the asymptotic performance of the ML estimator can be examined by the CRB. From (5.24), the log likelihood function is given by

$$\Lambda_s = \log p = -MN\log(\pi\sigma_w^2) - \frac{1}{\sigma_w^2}\sum_{i=1}^{N} \| \underline{x}(i) - A(i)\underline{s}(i) \|_F^2 . \tag{8.20}$$

We can show that the Fisher information sub-matrix $I[\underline{\alpha}(j)]$ corresponding to the MTS vector $\underline{\alpha}(j)$ can be decoupled from that that of the noise covariance σ_w^2, the CRB on the MTS estimates is obtained as

$$\mathrm{CRB}[\underline{\alpha}(j)] = I^{-1}[\underline{\alpha}(j)]. \tag{8.21}$$

The $\xi\eta$th element of the Fisher information matrix $I[\underline{\alpha}(j)]$ is defined

$$I[\underline{\alpha}(j)] = -\frac{\partial^2}{\partial\alpha_{km}\partial\alpha_{qn}} E[\Lambda_s], \tag{8.22}$$

where the expectation of the likelihood function is

$$\begin{aligned} E[\Lambda_s] &= -MN\log(\pi\sigma_w^2) - \frac{1}{\sigma_w^2}\sum_{i=1}^{N} E\{\| \underline{x}(i) - A(i)\underline{s}(i) \|_F^2\} \\ &= -MN\log(\pi\sigma_w^2) - \frac{1}{\sigma_w^2}\sum_{i=1}^{N}\Big\{tr[A'(i)R_s(i)A'^H(i) + \sigma_w^2 I] \\ &\quad -2Retr[A'(i)R_s(i)A^H(i)] + tr[A(i)R_s(i)A^H(i)]\Big\}. \end{aligned} \tag{8.23}$$

The $\xi\eta$th element of the Fisher information matrix $I[\underline{\alpha}(j)]$ is evaluated at the true $\underline{\alpha}(j)$ as

$$\begin{aligned}\{I[\underline{\alpha}(j)]\}_{\xi\eta} &= \frac{1}{\sigma_w^2}\sum_{i=1}^{N}\frac{\partial^2}{\partial\alpha_{km}(j)\partial\alpha_{qn}(j)}\{tr[A(i)R_s(i)A^H(i)] - 2Retr[A'(i)R_s(i)A^H(i)]\} \\ &= 2\frac{1}{\sigma_w^2}\sum_{i=1}^{N}\frac{\partial\theta_k(i)}{\partial\alpha_{km}(j)}\frac{\partial\theta_q(i)}{\partial\alpha_{qn}(j)}Retr\{A_q(i)R_s(i)A_k^H(i)\}, \end{aligned} \tag{8.24}$$

and in matrix form, the Fisher information matrix can be written as

$$I[\underline{\alpha}(j)] = \frac{2}{\sigma_w^2}\sum_{i=1}^{N}[C_K(i)\otimes \mathbf{O}]\odot[vecV(i,j)vec^HV(i,j)], \tag{8.25}$$

in which $C_K(i) = Re\{[D^H(i)D(i)]\odot R_s^T(i)\}$.

8.2 The extended Kalman filter

The extended Kalman filter (EKF) is an optimal extension of the Kalman filter to the case of nonlinear filtering. When the source target waveforms are assumed to be known, the EKF can be used to estimate and track the MTS vectors. The information model is the locally linear motion dynamics

$$\underline{\alpha}(n+1) = F_I[\underline{\alpha}(n)] = \left\{F^T[\underline{\alpha}_1(n),1], F^T[\underline{\alpha}_2(n),1],\ldots,F^T[\underline{\alpha}_K(n),1]\right\}^T. \tag{8.26}$$

and the array signal model can be used as the measurement model directly

$$\underline{x}(n) = F_M[\underline{\alpha}(n)] + \underline{w}(n) = A[\Theta(n)]\underline{s}(n) + \underline{w}(n). \tag{8.27}$$

The EKF performs a recursive operation. Assume that at k, an optimal MTS estimate $\hat{\underline{\alpha}}(k|k)$ has been obtained based on the array observations $\{\underline{x}(i);\ i = 1,2,\ldots,k\}$. When the new measurement $\underline{x}(k+1)$ arrives, the MTS vector estimate $\hat{\underline{\alpha}}(k+1|k+1)$ is updated by [65]

$$\hat{\underline{\alpha}}(k+1|k+1) = \hat{\underline{\alpha}}(k+1|k) + \frac{1}{\sigma_w^2}P(k|k)Re\{\mathcal{H}(k)^H\{\underline{x}(k+1) - A[\Theta(k+1|k)]\underline{s}(k+1)\}, \tag{8.28}$$

where $\hat{\underline{\alpha}}(k+1|k)$ is predicted from the information model (8.26) from $\hat{\underline{\alpha}}(k|k)$ by

$$\hat{\underline{\alpha}}(k+1|k) = F_I[\hat{\underline{\alpha}}(k|k)], \tag{8.29}$$

and $\mathcal{H}(k)$ is the measurement Jacobian matrix with respect to the MTS vector defined as

$$\mathcal{H}(k) = \frac{\partial F_M[\underline{\alpha}(k)]}{\partial \underline{\alpha}(k)}. \tag{8.30}$$

The measurement Jacobian matrix $\mathcal{H}(k)$ is computed as

$$\mathcal{H}(k) = [s_1(k)\underline{d}[\theta_1(k)], \underline{0}, \underline{0}, s_2(k)\underline{d}[\theta_2(k)], \underline{0}, \underline{0}, \ldots, s_K(k)\underline{d}[\theta_K(k)], \underline{0}, \underline{0}], \tag{8.31}$$

where $\underline{0}$ denotes a zero vector of $M \times 1$ and

$$\underline{d}[\theta_k(i)] = \frac{\partial \underline{a}[\theta_k(i)]}{\partial \theta_k(i)} \mid_{\theta_k'(i)} . \tag{8.32}$$

Matrix $P(k|k) = E\{[\hat{\underline{\alpha}} - \underline{\alpha}(k)][\hat{\underline{\alpha}} - \underline{\alpha}(k)]^H\}$ is the covariance of the MTS estimate $\hat{\underline{\alpha}}(k|k)$. It is known [65] that $P(k|k)$ satisfies the EKF inverse covariance matrix propagation equation

$$P(k|k) = \{[\Phi(k-1)P(k-1|k-1)\Phi^H(k-1)]^{-1} + \frac{1}{\sigma_w^2} Re[\mathcal{H}^H(k)\mathcal{H}(k)]\}^{-1}, \tag{8.33}$$

where the shorter standard notation $\Phi(k) = \Phi(k, k-1)$ denotes the incremental state transition matrix between $\underline{\alpha}(k)$ and $\underline{\alpha}(k-1)$ defined as

$$\Phi(k, k-1) = \frac{\partial \underline{\alpha}(k)}{\partial \underline{\alpha}(k-1)}, \tag{8.34}$$

for the discrete information model. For locally linear target dynamics described by (8.26), the incremental transition matrix $\Phi(k) = \Phi(k, k-1)$ is

$$\Phi(k) = \Phi(k, k-1) = diag[\Phi_{11}(k), \Phi_{22}(k), \ldots, \Phi_{KK}(k)], \tag{8.35}$$

where each submatrix $\Phi_{mm}(k)$ is

$$\Phi_{mm}(k) = \begin{bmatrix} \frac{\partial\theta_m(k)}{\partial\theta_m(k-1)} & \frac{\partial\theta_m(k)}{\partial q_m(k-1)} & \frac{\partial\theta_m(k)}{\partial\phi_m(k-1)} \\ \frac{\partial q_m(k)}{\partial\theta_m(k-1)} & \frac{\partial q_m(k)}{\partial q_m(k-1)} & \frac{\partial q_m(k)}{\partial\phi_m(k-1)} \\ 0 & 0 & 1 \end{bmatrix}. \tag{8.36}$$

The first row of the matrix $\Phi_{mm}(k)$ can be obtained as $\underline{v}(i,j)$ in (6.37) by assigning $j = k$ and $i = j + 1$. The second row is computed from the target state equation (5.8) as

$$\begin{bmatrix} \frac{\partial q_m(k)}{\partial\theta_m(k-1)} \\ \frac{\partial q_m(k)}{\partial q_m(k-1)} \\ \frac{\partial q_m(k)}{\partial\phi_k(k-1)} \end{bmatrix} = \begin{bmatrix} \frac{-q_m^2(k-1)\cos[\theta_m(k-1)+\phi_m(k-1)]}{\{q_m^2(k-1)+2Tq_m(k-1)\sin[\theta_m(k-1)+\dot{\phi}_m(k-1)]+T^2\}^{2/3}} \\ \frac{1+q_m(k-1)\sin[\theta_m(k-1)+\phi_m(k-1)]}{\{q_m^2(k-1)+2Tq_m(k-1)\sin[\theta_m(k-1)+\dot{\phi}_m(k-1)]+T^2\}^{2/3}} \\ \frac{-q_m^2(k-1)\cos[\theta_m(k-1)+\phi_m(k-1)]}{\{q_m^2(k-1)+2Tq_m(k-1)\sin[\theta_m(k-1)+\dot{\phi}_m(k-1)]+T^2\}^{2/3}} \end{bmatrix}. \tag{8.37}$$

For deterministic nonlinear systems, it has been shown in [66] that the Fisher Information matrix can be represented in a recursive form as

$$I[\underline{\alpha}(k)] = [\Phi^{-1}(k-1)]^H I[\underline{\alpha}(k-1)][\Phi^{-1}(k-1)] + \frac{1}{\sigma_w^2} Re[\mathcal{H}^H(k)\mathcal{H}(k)], \tag{8.38}$$

where $I[\underline{\alpha}(k)]$ is the Fisher Information matrix corresponding to $\underline{\alpha}(k)$ based on the first k measurements, and in terms of the Cramér-Rao bound, the recursive form becomes

$$\text{CRB}[\underline{\alpha}(k)]^{-1} = \{\Phi(k-1)]\text{CRB}[\underline{\alpha}(k-1)]\Phi^H(k-1)\}^{-1} + \frac{1}{\sigma_w^2} Re[\mathcal{H}^H(k)\mathcal{H}(k)], \tag{8.39}$$

which is identical to the propagation equation of the inverse covariance matrix (8.33), except that all quantities in (8.39) are evaluated along the true target state trajectory rather than along the estimated target state trajectory $\hat{\underline{\alpha}}(k)$.

8.3 Numerical simulations

Consider an equi-spaced linear array of $M = 8$ sensors with half the source wavelength spacing, as shown in Fig. 6.1. Two narrow-band source targets of equal power are assumed to have known

waveforms, and their initial MTS components are set to $\theta_1(1) = 0^o$ and $\theta_2(1) = -20^o$, $q_1(1) = 7.4 \times 10^{-3}$ and $q_2(1) = 5.3 \times 10^{-3}$, and, $\phi_1(1) = 135^o$ and $\phi_2(1) = 30^o$. The performances are evaluated through Monte-Carlo simulations, and each test is repeated 100 times. Fig. 8.1, Fig. 8.2 and Fig. 8.3 demonstrate the variation of the MSE's of the MTS estimates with SNR, when the MLE and EKF are applied. The SNR ranges from −4dB to 26dB and the comparison is between the MLE, EKF and the CRB. In this simulation, the array number of samples is $N = 100$. In the figures, it is seen that the simulated and theoretically obtained MSE's coincide and attain the Cramér-Rao bound, while the simulated MSE's of the EKF estimate do not. Fig. 8.4 to Fig. 8.6 show the variation of MSE's of the EKF estimates with the array number of samples. As the array sample number increases, the MSE of the EKF estimates approaches the CRB. This agrees with the theoretical derivation that the EKF is asymptotically efficient, since the MSE and the CRB of the EKF estimates have identical propagation equations.

8.4 Conclusions

We have presented the ML tracking algorithm when the source waveforms are assumed to be known and compared it to the EKF. The ML approach has been shown to approach the Cramér-Rao bound asymptotically. For EKF applications, since the Cramér-Rao bound and the covariance matrix of the estimates have identical propagation equation, the covariance on the estimates is an approximation of the Cramér-Rao bound, and at its best, attains the Cramér-Rao bound asymptotically. The EKF is a computationally efficient algorithm and requires a smaller computational load than the ML approach. However, the EKF requires *a priori* known statistical information about the estimates which, in practice, is usually not available. Although, some of the statistical information can be obtained throught measurement, this may lead to increased compuational load, slow convergence or even failure of the EKF.

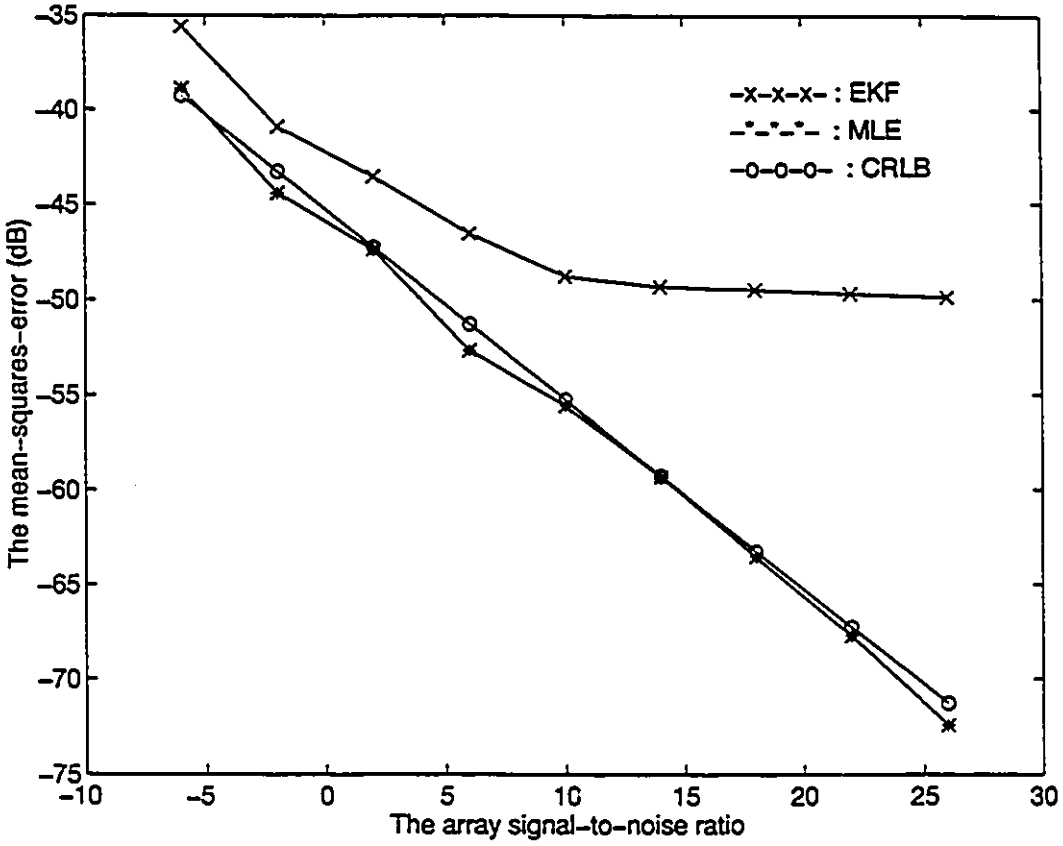

Figure 8.1: The MSE comparison between the MLE and EKF for the DOA estimates

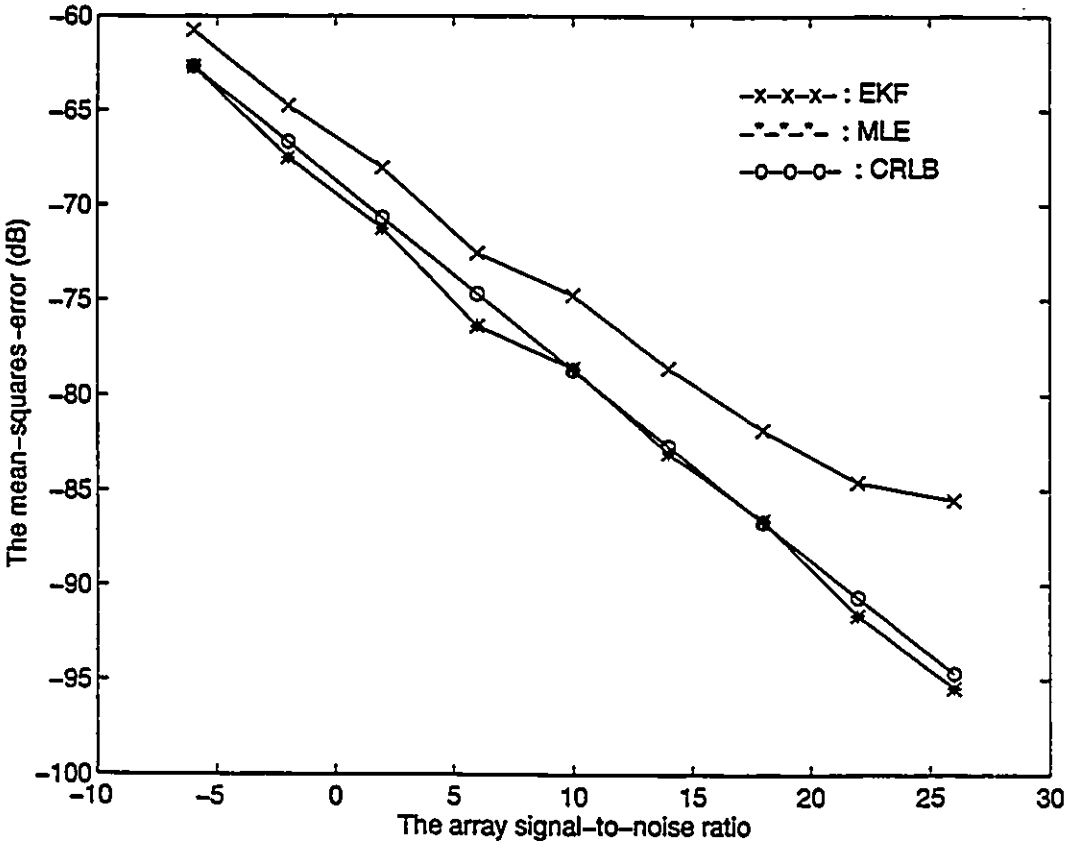

Figure 8.2: The MSE comparison between the MLE and EKF for the q component estimates

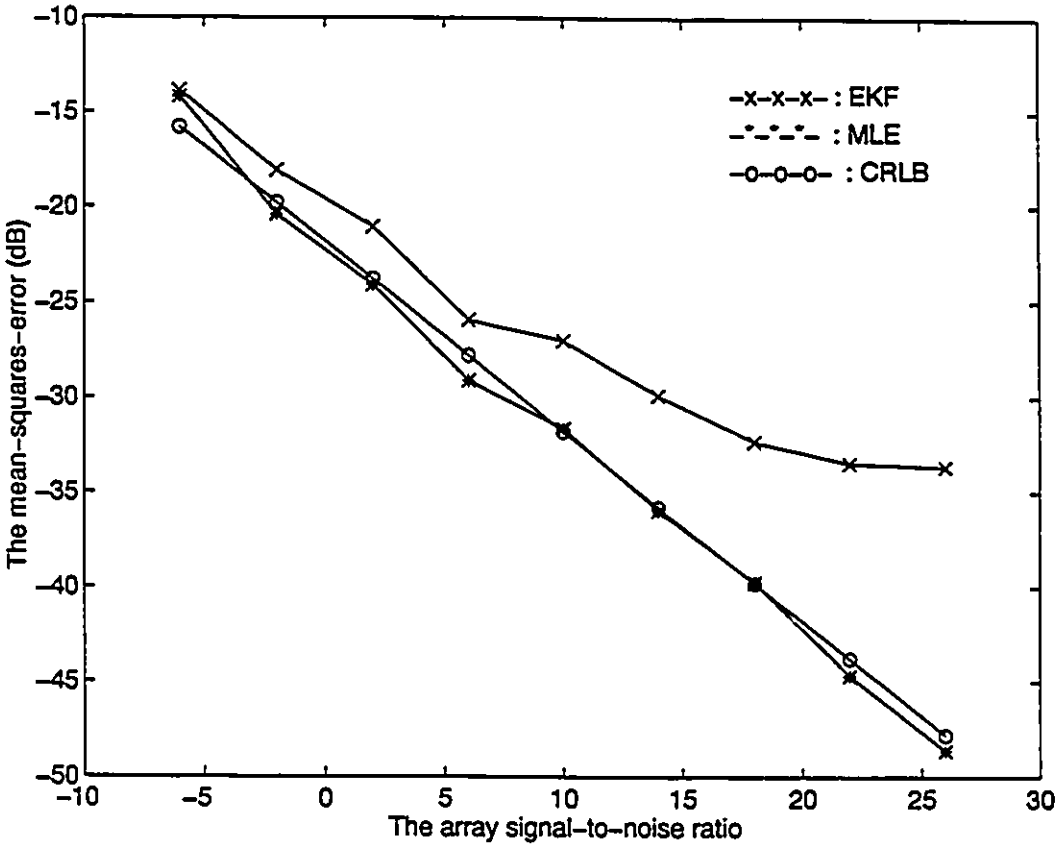

Figure 8.3: The MSE comparison between the MLE and EKF for the ϕ component estimates

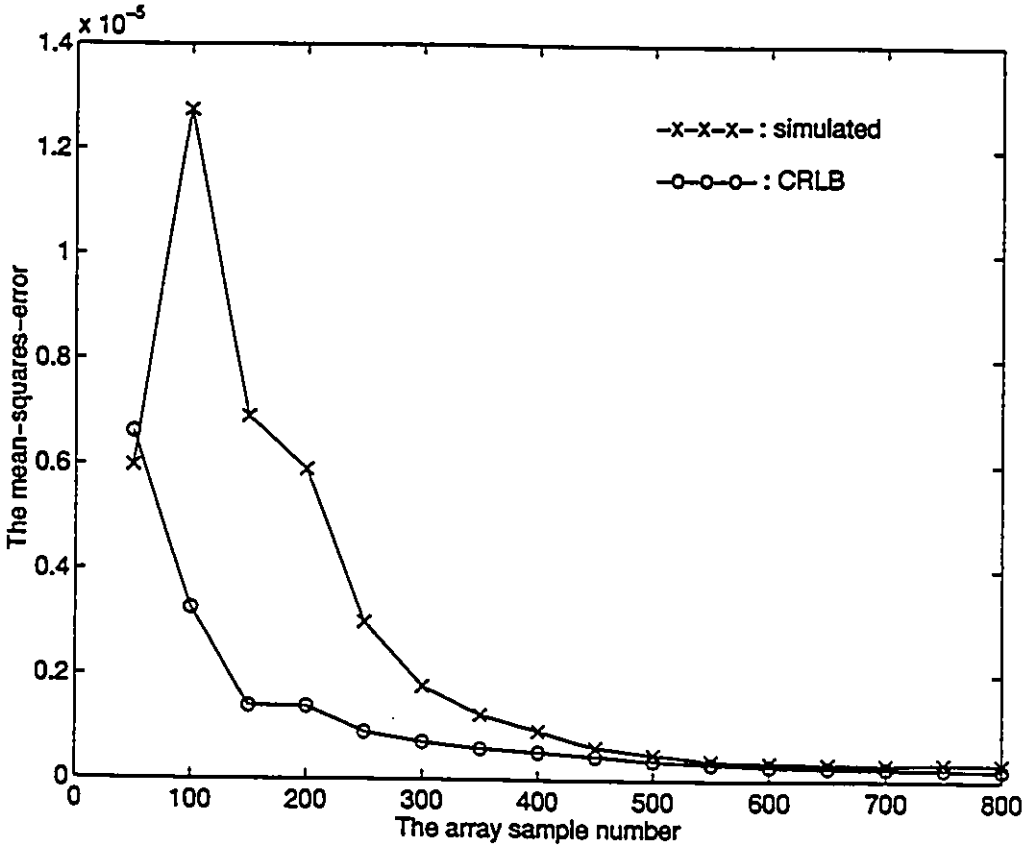

Figure 8.4: The variation of the MSE of the DOA estimate with the number of array samples.

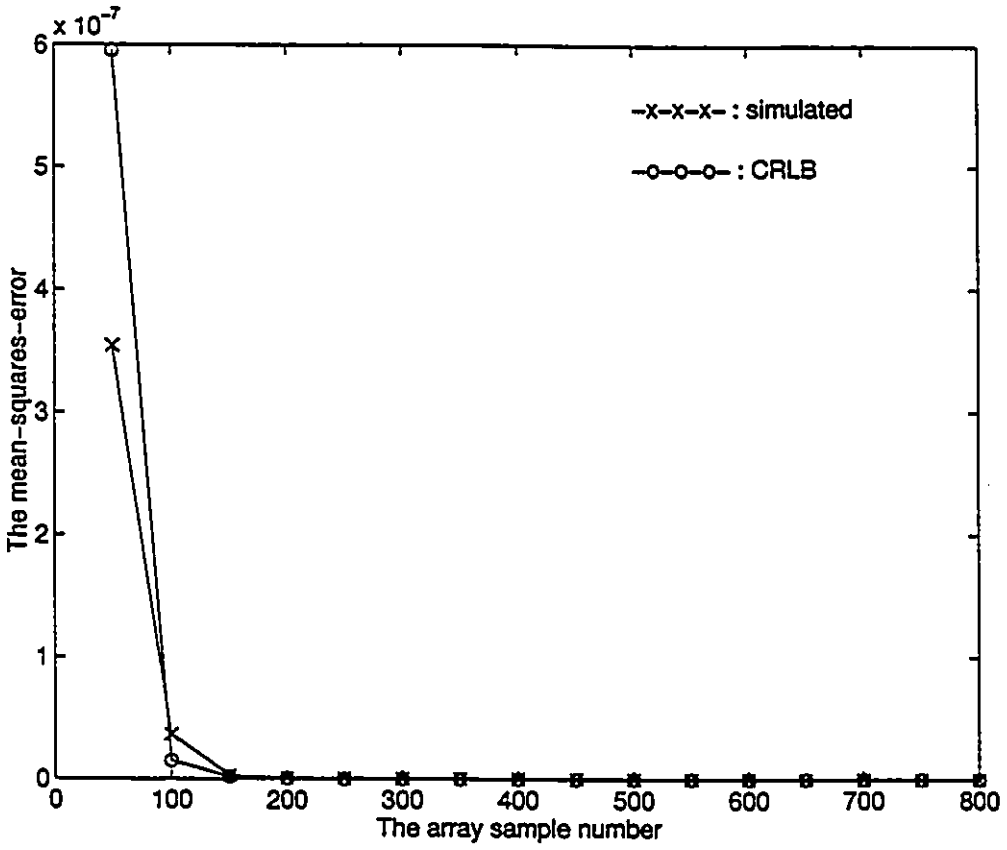

Figure 8.5: The variation of the MSE of the q component estimate with the number of array samples.

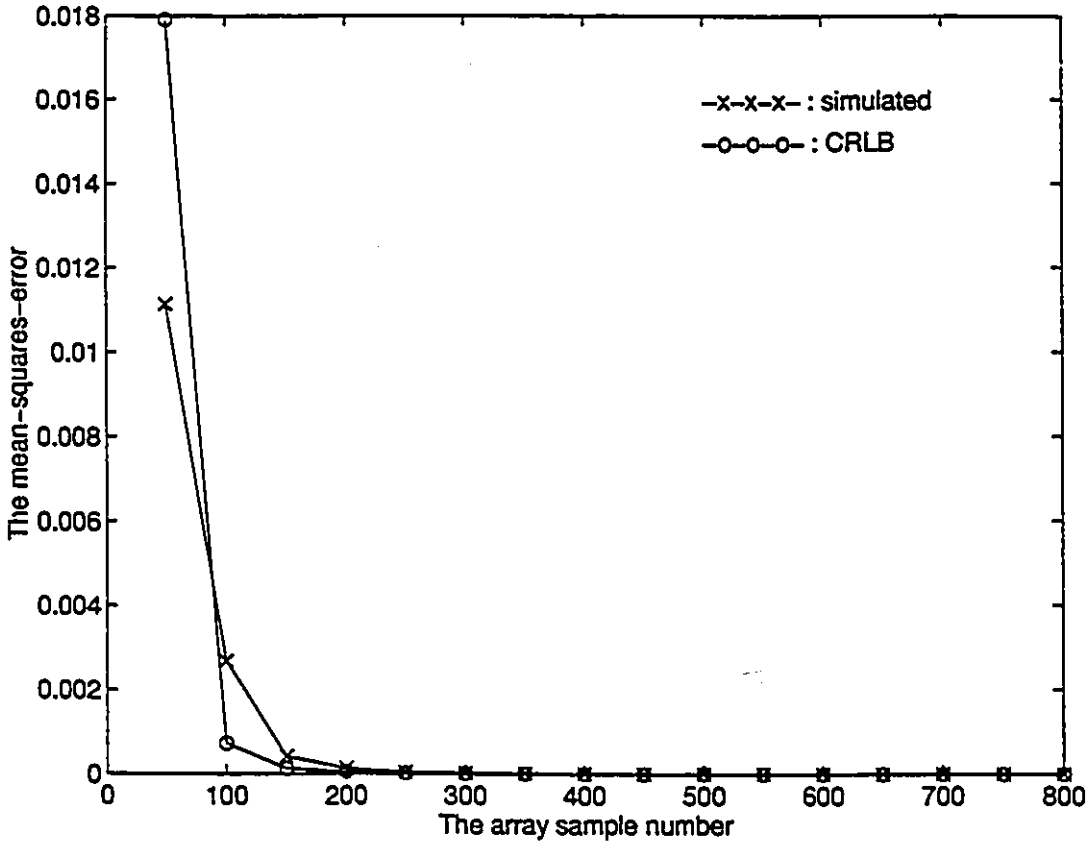

Figure 8.6: The variation of the MSE of the ϕ component estimate with the number of array samples.

Chapter 9

Summary and Future Research Directions

This thesis is mainly focused on the tracking of moving source targets by a passive array and the asymptotic performance analysis. In the following, we summarize our major contributions throughout the course of this work.

- A high resolution DOA estimation algorithm for spatially close source signals has been proposed. The algorithm is based on the Karhunen-Loève expansion of the covariance matrix of the array manifold in a limited sector of interest.
- The asymptotic performance has been analysed. It shows that the proposed technique is consistent and superior to conventional-subspace based techniques. Theoretical results are verified via computer simulations.
- A maximum likelihood estimator for tracking the DOA's of multiple moving targets by a passive array has been presented. The MTS has been defined to describe the state of the moving source targets. The target motion is modeled as locally linear. We show that the locally linear motion dynamics are strongly locally observable, and this observability ensures that the source target parameters can be resolved uniquely.

- A modified Gauss-Newton algorithm has been provided for the optimization of the ML criterion. We have derived a close form expression for the gradient and Hessian of the criterion function. The Hessian has been successfully approximated by a positive semi-definite matrix to ensure that the algorithm is descent.

- The asymptotic performance of the ML tracking estimator has been provided. We have derived the analytic forms for the asymptotic covariance and the Cramér-Rao lower bound for the MTS estimate. We show that the ML estimates of MTS vectors are not asymptotically efficient, but relatively asymptotically efficient.

- The ML tracking algorithm has been compared with the EKF, based on the assumption that the source target waveforms are known. Their asymptotic performance behaviours have also been studied.

It should be noted that, although in this thesis, the emphasis has been on co-planar array configurations, the ML tracking technique is directly applicable to general three-dimensional array geometries. A useful research topic in the future would be the application of adaptive algorithms to the ML tracking technique. When there are many batches of data, the ML algorithm based on a batch of data may be ineffectve because the size of the data makes each iteration costly. An adaptive algorithm updates the estimates each time when a new data point comes in. One advantage of an adaptive algorithm is that the optimal estimates become available as data are accumulated, making it suitable for real-time operation.

Appendix A

Observabiliy of the Locally Linear Motion Model

Consider the target motion dynamics (5.8) and (5.17). The mapping T_α is defined as

$$T_\alpha[\underline{\alpha}(m)] = [\theta(m), \theta(m+1), \theta(m+2)]. \tag{A.1}$$

The observability of rank condition is related to the conditioning of the mapping $T_\alpha[\underline{\alpha}(m)]$ or its Jacobian matrix at $\underline{\alpha}(m)$. The Jacobian matrix ∇T_α has the form given by

$$\nabla T_\alpha = \begin{bmatrix} 1 & \dfrac{\partial\theta(m+1)}{\partial\theta(m)} & \dfrac{\partial\theta(m+2)}{\partial\theta(m)} \\ 0 & \dfrac{\partial\theta(m+1)}{\partial q(m)} & \dfrac{\partial\theta(m+2)}{\partial q(m)} \\ 0 & \dfrac{\partial\theta(m+1)}{\partial\phi(m)} & \dfrac{\partial\theta(m+2)}{\partial\phi(m)} \end{bmatrix}. \tag{A.2}$$

The full rank coditioning of $T_\alpha[\underline{\alpha}(m)]$ can be determined by computing the determinant of the Jacobian matrix. From (5.11), we obtain the the gradient of $\theta(m+i)$ with respect to $\underline{\alpha}(m)$ as

$$\begin{bmatrix} \dfrac{\partial\theta(m+i)}{\partial\theta(m)} \\ \dfrac{\partial\theta(m+i)}{\partial q(m)} \\ \dfrac{\partial\theta(m+i)}{\partial\phi(m)} \end{bmatrix} = \begin{bmatrix} \dfrac{q^2(m) + iTq(m)\sin[\theta(m)+\phi(m)]}{q^2(m) + i^2T^2 + 2iTq(m)\sin[\theta(m)+\phi(m)]} \\ \dfrac{-iT\cos[\theta(m)+\phi(m)]}{q^2(m) + i^2T^2 + 2iTq(m)\sin[\theta(m)+\phi(m)]} \\ \dfrac{-i^2T^2 - iTq(m)\sin[\theta(m)+\phi(m)]}{q^2(m) + i^2T^2 + 2iTq(m)\sin[\theta(m)+\phi(m)]} \end{bmatrix}. \tag{A.3}$$

where $i = 1, 2$, and the determinant of ∇T_α can be computed as

$$det[\nabla T_\alpha] = det \begin{bmatrix} 1 & \frac{\partial \theta(m+1)}{\partial \theta(m)} & \frac{\partial \theta(m+2)}{\partial \theta(m)} \\ 0 & \frac{\partial \theta(m+1)}{\partial q(m)} & \frac{\partial \theta(m+2)}{\partial q(m)} \\ 0 & \frac{\partial \theta(m+1)}{\partial \phi(m)} & \frac{\partial \theta(m+2)}{\partial \phi(m)} \end{bmatrix} \tag{A.4}$$

which by substituting (A.2) can be further simplified as

$$det[\nabla T_\alpha] = \frac{\partial \theta(m+1)}{\partial q(m)} \frac{\partial \theta(m+2)}{\partial \phi(m)} - \frac{\partial \theta(m+2)}{\partial q(m)} \frac{\partial \theta(m+1)}{\partial \phi(m)}. \tag{A.5}$$

Let variables $\xi_1(m)$ and $\xi_2(m)$ be

$$\begin{aligned} \xi_1(m) &= q^2(m) + T^2 + 2Tq(m)\sin[\theta(m) + \phi(m)] \\ \xi_2(m) &= q^2(m) + 4T^2 + 4Tq(m)\sin[\theta(m) + \phi(m)], \end{aligned} \tag{A.6}$$

respectively. Equation (A.5) can be written as

$$\begin{aligned} det[\nabla T_\alpha] = & \frac{-T\cos[\theta(m) + \phi(m)]}{\xi_1(m)} \cdot \frac{-4T^2 - 2Tq(m)\sin[\theta(m) + \phi(m)]}{\xi_2(m)} \\ & + \frac{2T\cos[\theta(m) + \phi(m)]}{\xi_2(m)} \cdot \frac{-T^2 - Tq(m)\sin[\theta(m) + \phi(m)]}{\xi_1(m)}. \end{aligned} \tag{A.7}$$

Consider the situations in which the determinant $det[\nabla T_\alpha]$ will be zero. Setting $det[\nabla T_\alpha] = 0$ is equivalent to

$$\cos[\theta(m) + \phi(m)] = 0, \tag{A.8}$$

where we have use the fact that $\xi_1(m)$ and $\xi_2(m)$ are positive. Then, it is clear that the determinant $det[\nabla T_\alpha]$ equals zero if and only if $\theta(m) = \pi/2 - \phi(m)$ or $\theta(m) = -\pi/2 - \phi(m)$. In other words, the locally linear motion dynamics satisfy the observability of rank condition almost everywhere except when $\theta(m) = \pi/2 - \phi(m)$ or $\theta(m) = -\pi/2 - \phi(m)$.

Appendix B

Calculating the Correlation between the Analytic Signals

Consider the analytic representations of two stationary signals $x(t)$ and $y(t)$ [13]

$$\begin{aligned} x(t) &= x_0(t) + j\hat{x}_0(t) \\ y(t) &= y_0(t) + j\hat{y}_0(t), \end{aligned} \tag{B.1}$$

where $\hat{x}_0(t)$ and $\hat{y}_0(t)$ are the Hilbert transforms of $x_0(t)$ and $y_0(t)$, respectively and $j = \sqrt{-1}$ here. The cross-covariance between $x(t)$ and $y(t)$ is written as

$$\begin{aligned} E[x(t)y(t+\tau)] &= E[(x_0(t) + j\hat{x}_0(t))(y_0(t+\tau) + j\hat{y}_0(t+\tau))] \\ &= E[(x_0(t)y_0(t+\tau)] - E[(\hat{x}_0(t)\hat{y}_0(t+\tau)] \\ &\quad + j\left\{E[(x_0(t)\hat{y}_0(t+\tau)] + E[(\hat{x}_0(t)y_0(t+\tau)]\right\}. \end{aligned} \tag{B.2}$$

From the definition of the Hilbert transform, we have

$$E[(\hat{x}_0(t)\hat{y}_0(t+\tau)] = E[\frac{1}{\pi^2}\int_{-\infty}^{\infty}\frac{x_0(u)}{t-u}du\int_{-\infty}^{\infty}\frac{y_0(v+\tau)}{t-v}dv]$$

$$= \frac{1}{\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{r_{xy}(\tau+v-u)}{t-u}dudv, \tag{B.3}$$

where $r_{xy}(\tau) = E[x_0(t)y_0(t+\tau)]$. Since

$$\frac{1}{\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{r_{xy}(\tau+v-u)}{t-u}dudv = r_{xy}(t+\tau), \tag{B.4}$$

we have

$$E[x_0(t)y_0(t+\tau)] = E[\hat{x}_0(t)\hat{y}_0(t+\tau)]. \tag{B.5}$$

and in a similar manner,

$$E[x_0(t)\hat{y}_0(t+\tau)] = -E[\hat{x}_0(t)y_0(t+\tau)]. \tag{B.6}$$

By combining (B.5) and (B.6), we prove the assertion that

$$E[x(t)y(t+\tau)] = 0. \tag{B.7}$$

Appendix C

Variance of the Likelihood Function

Taking expectation of the criterion function J of (7.1) yields

$$E\left\{\frac{1}{N}\sum_{i=1}^{N} tr\{P^{\perp}_{A(i)}\hat{R}_x(i)\}\right\} = \frac{1}{N}\sum_{i=1}^{N} tr\{P^{\perp}_{A(i)}R_x(i)\} = \overline{J}, \tag{C.1}$$

where R_x is given in (7.2). This means that J is an unbiased estimate of $\overline{J}$. Consider the purterbation of $\overline{J}$

$$\Delta J = J - \overline{J}. \tag{C.2}$$

From (C.2), we know that ΔJ is zero mean. Using (5.14) and (7.2), we can write ΔJ as

$$\begin{aligned}\Delta J &= \frac{1}{N}\sum_{i=1}^{N} tr\{P_i^{\perp}[A(i)\underline{s}(i)\underline{w}^H(i)]\} + \frac{1}{N}\sum_{i=1}^{N} tr\{P_i^{\perp}[\underline{w}(i)\underline{s}^H(i)A^H(i)]\} \\ &\quad + \frac{1}{N}\sum_{i=1}^{N} tr\{P_i^{\perp}[\underline{w}(i)\underline{w}^H(i) - \sigma_w^2 I]\} \\ &= \delta_1 + \delta_2 + \delta_3. \end{aligned} \tag{C.3}$$

It is straightforward that δ_1, δ_2 and δ_3 are zero mean with

$$E[\delta_i] = 0, \quad i = 1, 2, 3. \tag{C.4}$$

Let $\underline{\xi}(i) = P_i^{\perp}A(i)\underline{s}(i)$. Then, δ_1 and δ_2 can be written alternatively as

$$\begin{aligned}\delta_1 &= \frac{1}{N}\sum_{i=1}^{N}\sum_{m=1}^{M}\xi_m(i)w_m^*(i)\\ \delta_2 &= \frac{1}{N}\sum_{i=1}^{N}\sum_{m=1}^{M}\xi_m^*(i)w_m(i),\end{aligned} \tag{C.5}$$

respectively, where $\xi_m(i)$ denotes the mth element of $\underline{\xi}(i)$. Under Gaussian assumption, the covariance of δ_1 can be written as

$$\begin{aligned}E[\delta_1\delta_1^*] &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M}\xi_m(i)\xi_n^*(j)E[w_m^*(i)w_n(j)]\\ &= \frac{\sigma_w^2}{N^2}\sum_{i=1}^{N}\sum_{m=1}^{M}\xi_m(i)\xi_m^*(i)\\ &= \frac{\sigma_w^2}{N^2}\sum_{i=1}^{N}tr\{P_i^{\perp}A(i)R_s(i)A^H(i)\}.\end{aligned} \tag{C.6}$$

Similarly, the covariance of δ_2 is derived as

$$E[\delta_2\delta_2^*] = \frac{\sigma_w^2}{N^2}\sum_{i=1}^{N}tr\{P_i^{\perp}A(i)R_s(i)A^H(i)\} = E[\delta_1\delta_1^*]. \tag{C.7}$$

Assume that $tr\{P_i^{\perp}A(i)R_s(i)A^H(i)\}$ is bounded by a limited constant c_v. Then, the following holds

$$E[\delta_i\delta_i^*] \le \frac{c_v\sigma_w^2}{N}, \quad i = 1, 2. \tag{C.8}$$

As N approaches infinity, the covariances of δ_1 and δ_2 tend to zero, $O(1/N)$, which implies that δ_1 and δ_2 asymptotically converge to zero, $O(1/\sqrt{N})$, with probability one.

The covariance of δ_3 can be expressed as

$$E[\delta_3\delta_3^*] = E\left[\left\{\frac{1}{N}\sum_{i=1}^{N}tr\{P_i^{\perp}[\underline{w}(i)\underline{w}^H(i)\}\right\}\left\{\frac{1}{N}\sum_{i=1}^{N}tr\{P_i^{\perp}[\underline{w}(i)\underline{w}^H(i)\}\right\}^*\right] - \sigma_w^4(M-K)^2. \tag{C.9}$$

Let

$$\mu = \frac{1}{N}\sum_{i=1}^{N}tr\{P_i^{\perp}[\underline{w}(i)\underline{w}^H(i)\}. \tag{C.10}$$

Denote $\underline{\rho}(i) = P_i^{\perp}\underline{w}(i)$ and we have

$$\mu = \frac{1}{N}\sum_{i=1}^{N}\sum_{m=1}^{M}\rho_m(i)\rho_m^*(i), \tag{C.11}$$

where $\rho_m(i)$ is the mth element of $\underline{\rho}(i)$. The covariance of μ can be written as

$$E[\mu\mu^*] = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M}E[\rho_m(i)\rho_m^*(i)\rho_n(j)\rho_n^*(j)]. \tag{C.12}$$

Since $\underline{\rho}(i)$ is a linear combination of Gaussian variables of zero mean, it is also Gaussian of zero mean with covarance $E[\underline{\rho}(i)\underline{\rho}^H(i)] = \sigma_w^2 P_i^{\perp}$. Then, we can rewrite (C.12) as [67]

$$\begin{aligned} E[\mu\mu^*] &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M}E[\rho_m(i)\rho_m^*(i)]E[\rho_n(j)\rho_n^*(j)] \\ &\quad + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M}E[\rho_m(i)\rho_n^*(j)]E[\rho_m^*(i)\rho_n(j)] \\ &\quad + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M}E[\rho_m(i)\rho_n(j)]E[\rho_m^*(i)\rho_n^*(j)] \\ &= \Pi_1 + \Pi_2 + \Pi_3. \end{aligned} \tag{C.13}$$

we can verify that

$$\begin{aligned} \Pi_1 &= \left[\frac{1}{N}\sum_{i=1}^{N}\sum_{m=1}^{M}E[\rho_m(i)\rho_m^*(i)]\right]\left[\frac{1}{N}\sum_{j=1}^{N}\sum_{n=1}^{M}E[\rho_n(j)\rho_n^*(j)]\right] \\ &= \left[\frac{1}{N}\sum_{i=1}^{N}tr\{E[\underline{\rho}(i)\underline{\rho}^H(i)]\}\right]\left[\frac{1}{N}\sum_{j=1}^{N}tr\{E[\underline{\rho}(j)\underline{\rho}^H(j)]\}\right] \\ &= \left[\frac{1}{N}\sum_{i=1}^{N}tr\{\sigma_w^2 P_i^{\perp}\}\right]\left[\frac{1}{N}\sum_{j=1}^{N}tr\{\sigma_w^2 P_i^{\perp}\}\right] \\ &= \sigma_w^4(M-K)^2. \end{aligned} \tag{C.14}$$

and

$$\Pi_2 = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M}E[\rho_m(i)\rho_n^*(i)]E[\rho_m^*(i)\rho_n(i)]$$

$$= \frac{1}{N}\sum_{i=1}^{N} \| E[\underline{\varrho}(i)\underline{\varrho}^H(i)] \|_F^2 = \frac{1}{N^2}\sum_{i=1}^{N} \| \sigma_w^2 P_i^{\perp} \|_F^2$$

$$= \frac{1}{N^2}\sum_{i=1}^{N} tr\{\sigma_w^4 P_i^{\perp}\} = \frac{\sigma_w^4(M-K)}{N}. \qquad \text{(C.15)}$$

For analytic process $\underline{w}(i)$ with zero mean, it is shown in Appendix B that $\Pi_3 = 0$. Then, the covariance of δ_3 becomes

$$E[\delta_3 \delta_3^*] = \frac{\sigma_w^4(M-K)}{N}, \qquad \text{(C.16)}$$

and it follows that the covariance of δ_3 tends to zero $O(1/N)$ as $N \to 0$. This implies that δ_3 asymptotically converges to zero, $O(1/\sqrt{N})$, with probability one when the sample number N approaches infinity.

Appendix D

Covariance between p_{km} and p_{qn}

Consider the covariance between p_{km} and p_{qn}

$$\begin{aligned} cov[p_{km}, p_{qn}] &= E[p_{km}p_{qn}^*] \\ &= E\{[p_{km}(1) + p_{km}(2) + p_{km}(3)][p_{qn}^*(1) + p_{qn}^*(2) + p_{qn}^*(3)]\} \\ &= E[p_{km}(1)p_{qn}^*(1)] + E[p_{km}(1)p_{qn}^*(2)] + E[p_{km}(1)p_{qn}^*(3)] \\ &\quad + E[p_{km}(2)p_{qn}^*(1)] + E[p_{km}(2)p_{qn}^*(2)] + E[p_{km}(2)p_{qn}^*(3)] \\ &\quad + E[p_{km}(3)p_{qn}^*(1)] + E[p_{km}(3)p_{qn}^*(2)] + E[p_{km}(3)p_{qn}^*(3)], \end{aligned} \tag{D.1}$$

where $p_{km}(1)$, $p_{km}(2)$ and $p_{km}(3)$ are defined as

$$\begin{aligned} p_{km}(1) &= -\frac{1}{N}\sum_{i=1}^{N} tr\{B_{km}(i)A(i)\underline{s}(i)\underline{w}^H(i)\} \\ p_{km}(2) &= -\frac{1}{N}\sum_{i=1}^{N} tr\{B_{km}(i)\underline{w}(i)\underline{s}^H(i)A^H(i)\} \\ p_{km}(3) &= -\frac{1}{N}\sum_{i=1}^{N} tr\{B_{km}(i)[\underline{w}(i)\underline{w}^H(i) - \sigma_w^2 I]\}, \end{aligned} \tag{D.2}$$

respectively, and

$$B_{km}(i) = \frac{\partial P_i^{\perp}}{\partial \theta_k(i)} \frac{\partial \theta_k(i)}{\partial \alpha_{km}(i)}. \tag{D.3}$$

It can be shown that p_{km} and $p_{km}(i)$ are zero mean. Denote $\underline{\xi}(i) = B_{km}(i)A(i)\underline{s}(i)$ and $\underline{\eta}(i) = B_{qn}(i)A(i)\underline{s}(i)$. We have

$$\begin{aligned} E[p_{km}(1)p_{qn}^*(2)] &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M} E[\xi_m(i)w_m^*(i)\eta_n(j)w_n^*(i)] \\ &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M} \xi_m(i)\eta_n(i)E[w_m^*(i)w_n^*(i)]. \end{aligned} \tag{D.4}$$

Since for the analytic array noise process, $E[w_m^*(i)w_n^*(i)] = 0$ (see Appendix A), we have

$$E[p_{km}(1)p_{qn}^*(2)] = 0. \tag{D.5}$$

By similar procedures, we can show that

$$E[p_{km}(2)p_{qn}^*(1)] = 0. \tag{D.6}$$

Let $\beta_{mn}(i)$ be the mnth element $B_{qn}(i)$. Then, we can write $p_{qn}(3)$ as

$$p_{qn}(3) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{m=1}^{M}\sum_{r=1}^{M}\beta_{mr}w_r(i)w_m^*(i) + \frac{1}{N}\sum_{i=1}^{N}B_{km}(i)\sigma_w^2, \tag{D.7}$$

and covariance $E[p_{km}(1)p_{qn}^*(3)]$ can be written as

$$\begin{aligned} E[p_{km}(1)p_{qn}^*(3)] &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M}\sum_{r=1}^{M} E[\beta_{mr}^*\xi_n(j)w_r^*(i)w_m(i)w_n^*(j)] \\ &\quad -\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{n=1}^{M} E[\sigma_w^2 B_{km}(i)\xi_n(j)w_n^*(j)] \\ &= \Pi_1 + \Pi_2 \end{aligned} \tag{D.8}$$

With Gaussian zero mean assumption about $\underline{w}(i)$, it is obvious that $E[\Pi_2] = 0$. For analytic process $\underline{w}(i)$, by a moment theorem for Gaussian processes in [67], we have

$$E[w_r^*(i)w_m(i)w_n^*(j)] = 0; \;\; m,n = 1,2,\ldots,M,\; i,j = 1,2,\ldots,N, \tag{D.9}$$

and finally

$$E[p_{km}(1)p_{qn}^*(3)] = 0; \tag{D.10}$$

Similarly, we can show that

$$E[p_{km}(2)p_{qn}^*(3)] = 0 \quad E[p_{km}(3)p_{qn}^*(2)] = 0$$
$$E[p_{km}(3)p_{qn}^*(1)] = 0. \tag{D.11}$$

For the covariance between terms with the same index, we obtain

$$\begin{aligned} E[p_{km}(1)p_{qn}^*(1)] &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M} E[\xi_m(i)w_m^*(i)\eta_n^*(j)w_n(i)] \\ &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M} \xi_m(i)\eta_n^*(i)E[w_m^*(i)w_n(i)] \\ &= \frac{\sigma_w^2}{N^2}\sum_{i=1}^{N}\sum_{m=1}^{M} \xi_m(i)\eta_m^*(i) \\ &= \frac{\sigma_w^2}{N^2}\sum_{i=1}^{N} tr\{B_{km}(i)A(i)R_s(i)A^H(i)B_{qn}^H(i)\}, \end{aligned} \tag{D.12}$$

and

$$E[p_{km}(2)p_{qn}^*(2)] = \frac{\sigma_w^2}{N^2}\sum_{i=1}^{N} tr\{B_{qn}(i)A(i)R_s(i)A^H(i)B_{km}^H(i)\}. \tag{D.13}$$

For the covariance between $p_{km}(3)$ and $p_{qn}(3)$, we have

$$\begin{aligned} E[p_{km}(3)p_{qn}^*(3)] &= E\left[\frac{1}{N}\sum_{i=1}^{N} tr\{B_{km}(i)[\underline{w}(i)\underline{w}^H(i)\}\left\{\frac{1}{N}\sum_{i=1}^{N} tr\{B_{km}(i)[\underline{w}(i)\underline{w}^H(i)\}\right\}^*\right] \\ &\quad -\frac{\sigma_w^4}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} tr\{B_{km}(i)\}tr\{B_{km}(j)\} \\ &= \Lambda_1 + \Lambda_2. \end{aligned} \tag{D.14}$$

Denote γ_{mn} as the mnth element of $B_{km}(i)$. We expand Λ_1 as

$$\begin{aligned}
\Lambda_1 &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M}\sum_{r=1}^{M}\sum_{s=1}^{M} E[\gamma_{mr}\beta^*_{ns}w_r(i)w^*_m(i)w^*_s(j)w_n(j)] \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M}\sum_{r=1}^{M}\sum_{s=1}^{M} E[\gamma_{mr}w_r(i)w^*_m(i)]E[\beta^*_{ns}w^*_s(j)w_n(j)] \\
&\quad + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M}\sum_{r=1}^{M}\sum_{s=1}^{M} \gamma_{mr}\beta^*_{ns}E[w_r(i)w^*_s(j)]E[w^*_m(i)w_n(j)] \\
&\quad + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M}\sum_{r=1}^{M}\sum_{s=1}^{M} \gamma_{mr}\beta^*_{ns}E[n_r(i)n_n(j)]E[n^*_m(i)n^*_s(j)] \\
&= \Omega_1 + \Omega_2 + \Omega_3. \qquad \text{(D.15)}
\end{aligned}$$

The first term Ω_1 can be written directly as

$$\Omega_1 = \frac{\sigma_w^4}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} tr\{B_{km}(i)\}tr\{B_{qn}(j)\}, \qquad \text{(D.16)}$$

which is identical to Λ_2. With the analytic assumption about the array sensor noise, we have

$$\Omega_3 = 0. \qquad \text{(D.17)}$$

We write Ω_2 as

$$\begin{aligned}
\Omega_2 &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{m=1}^{M}\sum_{r=1}^{M} \gamma_{mr}\beta^*_{mr}E[w_r(i)w^*_s(i)]E[w^*_m(i)w_n(i)] \\
&= \frac{\sigma_w^4}{N^2}\sum_{i=1}^{N} tr\{B_{km}(i)B^H_{qn}(i)\}. \qquad \text{(D.18)}
\end{aligned}$$

Then, the covariance between $p_{km}(3)$ and $p_{qn}(3)$ can be written as

$$E[p_{km}(3)p^*_{qn}(3)] = \frac{\sigma_w^4}{N^2}\sum_{i=1}^{N} tr\{B_{km}(i)B^H_{qn}(i)\}, \qquad \text{(D.19)}$$

and the covariance between p_{km} and p_{qn} is calculated as

$$cov[p_{km}, p_{qn}] = E[p_{km}(1)p^*_{qn}(1)] + E[p_{km}(2)p^*_{qn}(2)] + E[p_{km}(3)p^*_{qn}(3)]$$

$$= \frac{2\sigma_w^2}{N^2}\sum_{i=1}^{N} Re\left[tr\{B_{km}(i)A(i)R_s(i)A^H(i)B_{qn}^H(i)\}\right.$$
$$\left.+\frac{\sigma_w^4}{N^2}\sum_{i=1}^{N} tr\{B_{km}(i)B_{qn}^H(i)\}\right]. \qquad \text{(D.20)}$$

Bibliography

[1] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting", IEEE Trans. ASSP, Vol. ASSP-39, pp 1110-1121, 1991

[2] G. Bienvenu and L. Kopp, "Source power estimation method associated with high resolution bearing estimation", ICASSP-1981, pp 302-305, 1981

[3] R. O. Schmidt, "Multiple emitter location and signal parameter estimation", IEEE Tran. AP, Vol. AP-34, pp. 276-280, 1986

[4] V. F. Pisarenko, "The retrieval of harmonics from a covariance function", Geophys. J. Roy. , Vol. 33, pp 247-266, 1973

[5] D. H. Johnson, S. R, DeGraff, "Improving the resolution of bearing arrays by eigenvalue analysis", IEEE Trans. ASSP, Vol. ASSP-30, pp 638-647, 1982

[6] R. Kumaresan and D. W. Tufts, "Estimating the angle of multiple plane waves", IEEE Tran. AES, Vol. AES-19, pp 134-139, 1983

[7] K. M. Buckley and X. Xu, "Spatial-spectrum estimation in a location sector", IEEE Tran. ASSP, Vol. ASSP-38, pp 1842-52, 1990

[8] S. S. Reddi, "Multiple source location - A digital approach", IEEE Tran. AES, Vol. AES-15, pp 95-105, 1979

[9] J. A. Cadzow, "General direction-of-arrival estimation : A signal subspace approach", IEEE Trans. AES, Vol. AES-25, pp 31-47, 1989

[10] R. Schmidt, "A signal subspace approach to multiple emitter location and spectral estimation", Ph. D dissertation, Standford Univ., 1981

[11] M. Kaveh and A. J. Barabell, "The statistical performance of the MUSIC and the mini-norm algorithms in resolving plane waves in noise", IEEE Tran. ASSP, Vol. ASSP-34, pp 331-341, 1986

[12] B. Porat and B. Friedlander, "Analysis of the asymptotic relative efficiency of the MUSIC algorithm", IEEE Trans. ASSP, Vol. ASSP-36, pp. 532-544, 1988

[13] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood and Cramer-Rao Bound : further results and comparisons", IEEE Trans. ASSP, Vol. ASSP-38, pp 2140-2150, 1990

[14] T. W. Anderson, "Introduction to multivariate statistical analysis", New York : John Wiley & Sons, Inc., 1984

[15] A. G. Jaffer, "Maximum likelihood direction finding of stochastic sources : a separable solution", Proc. ICASSP-1988, Vol. 5, pp 2893-2896

[16] J. F. Bohme, "Estimation of spectral parameters of correlated signals in wavefield", Signal Processing, Vol. 10, pp. 329-337, 1986

[17] Y. Bresler, "Maximum likelihood estimation of linear structured covariance with application to antenna array processing", Proc. 4th ASSP Workshop on Spectrum Estimation and Modeling, pp. 172-175, Minneapolis, MN, 1988

[18] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation", IEEE Trans. ASSP, Vol. ASSP-38, pp. 1783-1795, 1990

[19] W. Ligget, "Passive sonar : Fitting models to multiple time series", in NATO ASI on Signal Processing, J. W. R. Griffith et al., Eds. Academic Press, New York, 1973

[20] J A. Cadzow, "A high resolution direction-of-arrival algorithm of narrow-band coherent and incoherent sources", IEEE Trans. ASSP, Vol. ASSP-36, pp 965-979, 1988

[21] J M. H. Er and S. K. Hui, "An improved MUSIC algorithm for superresolution array processing", Proc. ICASSP-1990, Vo. 4, pp. 2843-2846

[22] R. H. Roy, "ESPIRIT, estimation of signal parameters via rotational invariance techniques", Ph. D Dissertation, Standford Univ., Standford, CA, 1987

[23] R. H. Roy, A. Paulraj and T. Kailath, "ESPIRIT - A subspace rotation approach to estimation of parameters of cisoids in noise", IEEE Trans. ASSP, Vol. ASSP-34, pp. 1340-1342, 1986

[24] R. H. Roy and T. Kailath, "ESPIRIT - estimation of signal parameters via rotational invariance techniques", IEEE Trans. ASSP, Vol. ASSP-37, pp. 984-995, 1989

[25] B. D. Rao and K. V. S. Hari, "Performance analysis of ESPIRIT and TAM in determining the direction of arrival of plane waves in noise", IEEE Trans. ASSP, Vol. ASSP-37, pp. 1990-1995, 1989

[26] N. Ahmed and K. R. Rao, "Orthogonal transformations for digital signal processing", Springer-Verlag, New York, 1975

[27] P. Stoica and A. Nehorai, "MUSIC, Maximum Likelihood and Cramer-Rao Bound", IEEE Tran. ASSP, Vol. ASSP-37, pp 720-741, 1989

[28] J. A. Cadzow, "Direction finding : The signal subspace approach", Report, Dept. of Elec. Engin., Vanderbilt Univ., Nashville, Tennessee, 1990

[29] G. H. Golub and C. F. Van Loan, "Matrix computation", Baltimore, MD, John Hopkins University Press, 1984

[30] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projection", IEEE Trans. ASSP, Vol. ASSP-37, pp 1190-1196, 1989

[31] J A. Cadzow, "General direction-of-arrival estimation : A signal subspace approach", IEEE Trans. AES, Vol. AES-25, pp 31-47, 1989

[32] M. Wax and I. Zinskind, "On unique localization of multiple sources by passive sensor arrays", IEEE Trans. ASSP, Vol. ASSP-37, pp 996-1000, 1989

[33] J. P. LeCadre and O. Zugmeyer, "Temporal integration for array processing", Proc. ICASSP-91, Vol. 4, pp 1441-1444, May 1991

[34] R. Kumaresan, A. G. Sadasiv, C. S. Ramalingam and J. F. Kaiser, "Instantaneous non-linear operators for tracking multicomponent signal parameters", Proc. ICASSP-92, Vol. 1, pp 404-407

[35] B. Champagne, "Optimum space-time processing in non-stationary environments", Ph. D Thesis, U. of Toronto, 1990

[36] C. K. Sword, M. Simaan and E. W. Kamen, "Multiple target angle tracking using sensor array output", IEEE Trans. AES, Vol. AES-26, pp 367-372, 1990

[37] C. R. Sastry, E. W. Kamen and M. Simaan, "An efficient algorithm for tracking the angles of arrival of moving targets", IEEE Trans. SP, Vol. SP-39, pp 242-246, 1991

[38] C. R. Rao, L. Zhang and L. C. Zhao, "Multiple target angle tracking using sensor array output", IEEE Trans. AES, Vol. AES-29, pp 268-271, 1993

[39] J. F. Yang and M. Kaveh, "Adaptive eigensubspace algorithms for direction or frequency estimation and tracking", IEEE Trans. SP, Vol. SP-36, pp 241-251, 1988

[40] P. Comon and G. H. Golub, "Tracking a few extreme singular values and vectors in signal processing", Proc. IEEE, Vol. 78, pp 983-996, 1990

[41] R. D. DeGroat, "Noniterative subspace tracking", IEEE Trans. SP, Vol. SP-40, pp 571-577, 1992

[42] C. B. Chang and J. A. Tabaczynski, "Applications of state estimation to target tracking", IEEE Trans. AC, Vol. AC-29, pp 98-109, 1984

[43] E. D. Sontag, "On the observability of polynomial systems, I : finite-time problems", SIAM J. Control, Vol. 17, pp 139-151, 1979

[44] R. R. Mohler and C. S. Hwang, "Nonlinear data observability and information", J. Franklin Inst., Vol. 325, No. 4, pp 443-464, 1988

[45] H. Nijeijer, "Observability of autonomous discrete time nonlinear system ", Int. J. Control, Vol. 36, pp 867-874, 1982

[46] S. P. Banks, "Mathematical theories of nonlinear systems", Prentice Hall, 1988

[47] H. L. Van Tree, "Detection, estimation, and modulation theory ", Part I, John Wiley & Sons, 1986

[48] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm", IEEE Trans. ASSP, Vol. ASSP-36, pp 477-489, 1988

[49] M. I. Miller and D. R. Fuhrmann, "Maximum-likelihood narrow-band direction finding and the EM algorithm", IEEE Trans. ASSP, Vol. ASSP-38, pp 1560-1577, 1990

[50] K. Sharman, "Maximum likelihood estimation by simulated annealing", Proc. ICASSP-1988, Vol. 4, pp 2741-2744

[51] D. Goryn and M. Kaveh, "Neural networks for narrowband and wideband direction finding", Proc. ICASSP-1988, Vol. 4, pp 2716-2719

[52] K. Sharman and G. D. McClurkin, "Genetic algorithms for maximum likelihood parameter estimation", Proc. ICASSP-1989, Vol. 5, pp 2716-2719

[53] M. Viberg, B. Ottersten and T. Kailath, "Detection and estimation in sensor arrays using weighted subspace fitting", IEEE Trans. ASSP, Vol. ASSP-39, pp 2436-2449, 1991

[54] M. Wax and T. Kailath, "Optimal localization of multiple sources by passive arrays", IEEE Trans. ASSP, Vol. ASSP-31, pp 1210-1218, 1983

[55] D. Starter and A. Nehorai, "Newton algorithms for conditional and unconditional maximum likelihood estimation of the parameters of exponential signals in noise", IEEE Trans. ASSP, Vol. ASSP-40, pp 1528-1534, 1992

[56] J. E. Dennisand R. B. Schnabel, "Numerical methods for unconstrained optimization and nonlinear equations", Englewood Cliffs, NJ : Prrntice-Hall, 1983

[57] G. H. Golub and V. Pereyra, "The differentiation of pseudoinverse and nonlinear least squares problem whose variables separate", SIAM J. Numer. Anal., Vol. 10, pp 413-432, 1973

[58] K. Levenberg, "A method for the solution of certain nonlinear problems in least squares", Qu. App. Maths, Vol. 2, pp 164, 1944

[59] S. M. Goldfeld, R. E. Quandt and H. F. Trotter, "Maximization by quardratic hill-climbing", Econometrica, Vol. 34, pp541-551, 1966

[60] L. Kauffman, "A vaiable projection method for solving separable nonlinear least squares problems", BIT, Vol. 15, pp 49-57, 1975

[61] R. Bellman, "Introduction to matrix analysis", 2nd ed. New York : McGraw-Hill, 1972

[62] G. X. Ni, "Matrix theory and methods", Shanghai Science and Technology Publishing House : Shanghai, PRC, 1984

[63] A. H. Jazwinski, "Stochastic processes and filtering theory", New York : Academic, 1970

[64] J. Ward and R. T. Compton, "Improving the performance of a slotted ALOHA packet radio network with an adaptive array", IEEE Trans. Commun, Vol. COMM-40, pp. 292-300, 1992

[65] R. P. Wishner, J. A. tabaczynski and M. Athans, "A comparison of three non-linear filters", Automatica, Vol. 5, pp. 487-496, 1969

[66] J. H. Taylor, "The Cramér-Rao estimation lower bound computation for deterministic nonlinear systems", IEEE Trans. AC, Vol. AC-24, pp. 343-344, 1979

[67] I. S. Reed, "On a moment theorem for complex Gaussian processes", IRE Trans. IT, Vol. IT-37, pp 194-195, 1962

[68] D H. Johnson, "The application of spectral estimation to bearing estimation problems", IEEE Proc., Vol. 70, pp 1018-1028, 1982