Perturbation Analysis and Compensating Algorithm for Subspace Fitting Array Signal Processing Methods

Thesis
Submitted in partial fulfillment of the requirements for the Degree of Master of Philosophy in Electrical and Electronic Engineering of the Hong Kong University of Science and Technology

by
Kong-man CHEUNG
B.Eng., University of Hong Kong, Hong Kong, 1992

May, 1994
Perturbation Analysis and Compensating Algorithm for Subspace Fitting Array Signal Processing Methods

by

Kong-man CHEUNG

Approved by:

Advisor: Dr. Sau-fong YAU
(Supervisor)

Head Of Dept: Prof. Peter Cheung

Research Committee member: Dr. Xiren CAO

Research Committee member: Dr. Bing ZENG

Research Committee member: Dr. Danny H. K. TSANG

Department of Electrical and Electronic Engineering
The Hong Kong University of Science and Technology
May 1994
“The secret of human life doesn’t lie in the fact that one lives, but in what he lives for.”  

_Fyodor Dostoevski._
Acknowledgments

I wish to sincerely thank my supervisor, Dr. Mark S. F. Yau, for his enthusiasm, inspiration, persistence, guidance, encouragement and support during my time at UST. His valuable suggestions on my research and thesis writing are absolutely appreciated. I am especially grateful to have been given the opportunity to work with him. Special thanks must be given to the financial help from him and the HKUST Research Grant R191/92.EG04. Without this financial support, I cannot survive till now.

I would like to thank my family members, especially my mother for their indebted support and encouragement during my two years studies for the degree.

Finally, I would like to express gratitude to my long-suffering colleagues here, especially Joanna Chiu for her advice on LaTeX. Special thanks should be given to Alex Lee for his opinions, help and encouragement through the period.
Abstract

Although the eigen-based subspace algorithms such as MUSIC and ESPRIT has been proven to be superior over other conventional methods, they are prone to model errors and system uncertainties which are ubiquitous in practical situations. Therefore recently much attention has been put on analyzing the behaviour and evaluating the performance of these subspace fitting method under the presence of random perturbations. Though much work has been done by researchers on these aspects, most of them have employed the additive Gaussian error model which may not be physically justifiable. Moreover, the algorithm they proposed to compensate for the errors under this perturbation model is computationally extensive and therefore cannot be put into practice. In this thesis, a novel and more realistic error model is introduced. Despite the fact that the novel idea is generally applicable to a wide variety of array signal processing problems and subspace methods, for the sake of presentation, focus has been made on problem of direction-of-arrivals estimation by MUSIC and ESPRIT. By using the new error model, together with certain assumptions, a compact expression for the perturbed data covariance matrix is derived. Based on this expression, a simple and robust compensating algorithm is proposed which proves to have good performance when the perturbation parameter is known. The case of unknown parameter is also considered and a new algorithm is devised for estimating this unknown parameter. To study the performance of the proposed compensating and estimating algorithms, the Cramér-Rao Bound for the parameters being estimated are derived. Simulated and theoretical results are then compared. Finally, to show the generality of our idea, the compensating algorithm is applied with slight modifications to the field of blind deconvolution. Promising simulation results are obtained.
# TABLE OF CONTENTS

1 INTRODUCTION
1.1 Brief History of Array Signal Processing ......................................... 1
1.2 Recent Research Topics and Thesis Objective ...................................... 3
1.3 Conventions and Notations ................................................................. 4

2 PROBLEM FORMULATION
2.1 Direction-of-Arrival (DOA) Estimation ................................................ 6
   2.1.1 General formulation ........................................................................ 6
   2.1.2 Nature of the radiating sources .................................................... 8
   2.1.3 The statistical behaviour of the signals and noises ....................... 11
2.2 Other Mathematically Equivalent Problems ........................................... 13
   2.2.1 Cisoids retrieval ........................................................................... 14
   2.2.2 Pole retrieval .............................................................................. 14
   2.2.3 Known echo retrieval .................................................................... 15
2.3 An Example: Uniform Linear Array (ULA) ............................................ 16

3 EIGEN-STRUCTURE BASED ALGORITHMS ........................................... 19
3.1 Fundamental Principle of Eigen-Structure
   Based Algorithm .................................................................................... 19
   3.1.1 Array manifold ............................................................................. 19
   3.1.2 Eigen-decomposition .................................................................... 22
   3.1.3 Source of Errors .......................................................................... 24
3.2 MULTiple SIgnal Classification (MUSIC) ............................................. 24
   3.2.1 System requirements .................................................................... 25
   3.2.2 Underlying principle ..................................................................... 25
   3.2.3 Sources of error and limitations .................................................. 26
3.3 Estimation of Signal Parameters via
Rotational Invariance Techniques (ESPRIT) ........................................ 27
3.3.1 System requirements ................................................................. 27
3.3.2 Underlying principle ................................................................. 28
3.3.3 Advantages and sources of errors ............................................. 30
3.4 Other Subspace Fitting Methods .................................................. 31

4 NOVEL APPROACH IN ERROR ANALYSIS ........................................ 32
4.1 Background .................................................................................... 32
4.2 DOA Dependent Errors ................................................................... 34
  4.2.1 Sources of errors ....................................................................... 34
  4.2.2 Assumptions ............................................................................. 36
  4.2.3 Mathematical formulation ......................................................... 39
  4.2.4 Effect of the perturbation matrix ............................................. 42
  4.2.5 Compensating algorithm .......................................................... 44
  4.2.6 Estimating unknown perturbation parameter .......................... 49
  4.2.7 ESPRIT formulation ................................................................. 55
4.3 DOA Independent Errors ............................................................... 58
  4.3.1 Sources of errors and mathematical model ............................... 58
4.4 Simulation Results ......................................................................... 62
  4.4.1 Significance of the compensating algorithm ............................. 62
  4.4.2 Problem of division by zero ...................................................... 67
  4.4.3 Unknown perturbation parameter ........................................... 70

5 CRAMÉR-RAO BOUND ANALYSIS .................................................. 73
5.1 Theoretical Background ............................................................... 73
5.2 Derivation of CRB ......................................................................... 75
5.3 Simulation Results ......................................................................... 77
  5.3.1 Effect of the number of sensors ............................................. 77
  5.3.2 Finite sample effect ................................................................. 80
5.4 Conclusions ................................................................................... 80
6 AN APPLICATION : BLIND DECONVOLUTION

6.1 Blind Deconvolution by ESPRIT ........................................ 82
   6.1.1 Mathematical formulation ......................................... 82
   6.1.2 Simulation results .................................................. 86
6.2 Modified Blind Deconvolution by ESPRIT ............................ 90
   6.2.1 Problems and remedies of the algorithm ....................... 90
   6.2.2 Simulation results .................................................. 92

7 CONCLUDING REMARKS ..................................................... 99

A Proof of Theorem T4.1 .................................................... 101

B Derivation of Perturbation Matrix for Uniform Distributed Errors 103

C Derivation of Perturbation Matrix for Gaussian Distributed Errors104

D Derivation of the Error Vector .......................................... 105

E Proof of Theorem T4.3 ..................................................... 106

F Proof of Theorem T4.4 ..................................................... 108

G Derivation of Covariance Matrix for Blind Deconvolution ........ 109

   Bibliography .............................................................. 112
CHAPTER 1
INTRODUCTION

A brief introduction to the field of array signal processing is first given through its history and importance in every day lives. Follows by that, the major research directions in the field is described together with a simple statement of the objective and the organization of this thesis. Finally, the conventions and notations that will be used throughout the thesis are listed for readers’ reference.

1.1 Brief History of Array Signal Processing

The ultimate goal of signal processing is to extract as much useful and accurate information as possible from the measurement data. Usually the available data are obtained from the environment by different kind of sensors. In the case of multiple sensors (single sensor is inadequate in many situations), special techniques are needed to process and utilize all the output data effectively. This gives rise to the development of array signal processing. In fact, besides the basic problems such as target tracking by radar, underwater sonar applications, oceanography, exploration seismology, obtaining high resolution maps of galaxies and quasars by radio telescope astronomy, array signal processing techniques can be applied to other different areas for information extraction or parameters estimation. One of the example is the so called "cocktail party problem" (recognition and separation of speeches as produced by different sources in a mixed noisy environment). Another example is the use of array processing algorithms in wireless communications like cellular phone design technology in estimating channel response and resolving individual signals in order to increase the channel capacity. Computer tomography by diffracting and non-diffracting sources in medical imaging is also one application where array signal processing methods are
used.

From the above examples, it is obvious that sensor array is becoming more and more common and important. Because of this, much research has been done in this field in the past few decades and great breakthroughs are achieved. Sensor arrays are characterized by their spatial geometry and directional patterns and can be classified into two main categories, namely structured and unstructured sensor arrays. Arrays with very regular geometry and directional patterns are known as structured arrays, otherwise they are classified as unstructured. Due to the regular nature of the structured arrays, most of the array processing algorithms deal with this kind of sensor array. One of these methods is called conventional (classical) beamforming which may be considered as the earliest array signal processing technique. This method is based on the delay-and-sum principle in an attempt to steer a beam in a particular direction. Thus it is useful in Direction-Of-Arrival (DOA) determination of certain propagating electrical signal (which is a fundamental problem of array signal processing). However, this method is applicable when there is only one signal source. Hence based on the classical beamforming, many different methods are devised to handle situations with multiple sources. Some examples are Maximum Entropy Method proposed by Burg [5], Maximum Likelihood Method (Minimum Variance Estimator) suggested by Capon [6] and the AR prediction Techniques based on Prony’s method [21]. Although these beamforming methods may resolve signals of multiple sources, they are restricted by the Rayleigh resolution limit so that DOAs with angle separation smaller than this limit is unresolvable. The limitations can be explained by the fact that these techniques are all non-model based and do not exploit the detailed nature of the mathematical model of the observed processes nor the statistical models of the signals and noises.

By observing the constraints of the above algorithms, another approach known as model-based parametric estimation method was developed by several researchers such as Pisarenko [20] and Liggett [12] in the mid-70's. These methods, which are also known as eigen-structure based techniques, borrowed ideas from the statistical factor analysis literature of the early 1900's. Since these techniques, when under certain assumptions, can give superior performance and higher angular resolution as compared to the beamforming approach, they have arouse much attention in the recent
years. The two representative eigen-based methods are probably MUSIC (Multiple Signal Classification) suggested by Schmidt in 1977 [25] and ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques) developed by Roy in 1986 [24] while all the other eigen-based methods follows similar principle as the above two. These two methods are still very popular till now as they fully exploit the geometrical aspects of the array and hence are very effective in resolving closely spaced radiating sources.

1.2 Recent Research Topics and Thesis Objective

Needless to say, the inventions of MUSIC and ESPRIT may be considered as the two greatest breakthroughs in the field of array signal processing in the last twenty years as they casted a new direction in solving the complicated problem. Despite its effectiveness, however, these methods have their own shortcomings. Firstly, they may fail due to certain system constraints such as source coherency. Moreover, the algorithms depend heavily on the correctness of the mathematical model of the array and hence are susceptible to model uncertainties or errors. These errors are ubiquitous in practical situations and arise from different phenomena such as non-ideal sensor characteristics, imprecise sensors locations and orientations, channel inhomogeneities. In view of these, recently most of the researchers have put their efforts in these two aspects. In fact, success has been achieved in relaxing some system constraints like spatial smoothing [27] [11] in case of coherent sources. For the second problem, work has been done on analyzing the effects of model error performance evaluation of the algorithms in the past four years with many published papers such as [31] [18] [37]. Unfortunately, most of the literatures have focused only on performing theoretical analysis or investigating asymptotic behaviour of the algorithms in the presence of model errors, but very few of them has suggested complete or computationally efficient methods to compensate the errors. On account of this, the main theme of this thesis is to establish a fast and robust errors compensating algorithm in addition to studying the performance of eigen-structure based array signal processing methods under random model perturbations and uncertainties. Focus will be put on both MUSIC and ESPRIT from a novel point of view. Several aspects of the algorithms
are discussed and illustrated with simulation results. The compensating algorithm is shown to be able to incorporate into other problems such as blind deconvolution. In Chapter 2, the mathematical model of the problems will be derived. Brief descriptions concerning the principle and deficiencies of both MUSIC and ESPRIT will be given in Chapter 3 as preliminaries. Based on the materials presented in Chapters 2 and 3, the novel methods for error analysis and compensation are given in Chapter 4 which may be considered as the "soul" of the thesis. To study the performance of the proposed method, Cramér-Rao Bound (CRB) for the parameters being estimated are derived in Chapter 5. Chapter 6 discusses the application of the compensating algorithm to the field of blind deconvolution. To conclude the thesis, brief remarks and future research directions are given in Chapter 7.

1.3 Conventions and Notations

Throughout this thesis, scalar quantities are denoted by regular lower or upper case letters while lower and upper case bold type faces represent vectors and matrices respectively. Thus a (or A), $\mathbf{a}$, $\mathbf{A}$ stand for scalar, vector and matrix in that order. Moreover, the following is a list of symbols and operations that are used frequently in the sequel:

- $\mathbf{A}^T$ - Transpose of $\mathbf{A}$,
- $\mathbf{A}^*$ - Complex conjugate transpose of $\mathbf{A}$,
- $\text{tr}(\mathbf{A})$ - Trace of the matrix $\mathbf{A}$,
- $\text{det}(\mathbf{A})$ - Determinant of the matrix $\mathbf{A}$,
- $\mathbf{A}^{-1}$ - Inverse of $\mathbf{A}$,
- $\mathbf{A}^\dagger$ - Pseudo-inverse of $\mathbf{A}$,
- $\text{diag}(\mathbf{A})$ - a vector containing the diagonal elements of the matrix $\mathbf{A}$.
- $\text{Diag}(\mathbf{a})$ - a diagonal matrix whose diagonal elements are that of the vector $\mathbf{a}$.
- $\text{Re}\{\cdot\}$ - Real part of the complex variable or function $\{\cdot\}$,
• $E\{x(\cdot)\}$ - Expectation of the stochastic process $x(\cdot)$.

• $\mathcal{F}\{x(t)\}$ - Fourier transform of the time function $x(t)$,

• $\rho(A)$ - Rank of the matrix $A$,

• $\mathcal{R}(A)$ - Range space of the matrix $A$,

• $A \odot B$ - Schur-Hadamard product between $A$ and $B$,

• $A \div B$ - Element-by-element division of matrix $A$ by $B$,

• $\hat{\cdot}$ - Estimate of the actual parameter $\cdot$, for example, $\hat{\sigma}$ represents the estimated value of $\sigma$,

• $T$ - A matrix with all its elements being 1's,

• $d$ - Number of sources to be estimated,

• $m$ - Number of sensors within the sensor array,

• $\theta$ - Direction of Arrival (DOA) of a certain source.
CHAPTER 2

PROBLEM FORMULATION

Mathematical model of array signal processing problems is discussed in this chapter. The typical case of Direction-of-arrival (DOA) estimation by an array of sensors is first presented and a basic formulation is derived. We will then show that most of the array signal processing problems are mathematically equivalent to DOA determination model. It allows us to concentrate on studying the DOA estimation problem. To conclude the chapter, we will consider an important kind of array sensor known as Uniform Linear Array (ULA) which will be used in subsequent chapters as example in describing and analyzing the existing array processing techniques.

2.1 Direction-of-Arrival (DOA) Estimation

2.1.1 General formulation

Probably the most fundamental problem in array signal processing is the location of energy-radiating sources in a transmission medium (such as the ocean or sky) by a group of electrical signal sensitive sensors. Consider the situation depicted in Figure 2.1. To simplify the situation, we assume the sources are coplanar with the sensors. Moreover, it is assumed that the transmission medium is isotropic and non-dispersive so that radiations propagate in straight lines (this assumption will be released in the latter part of the thesis).
Figure 2.1: Locating three radiating sources by a passive array consisting of three sensors.

Suppose there are totally $d$ emitting sources and the array consists of $m$ sensors. When the signals impinge on each sensor, electrical response will be produced and the response depends not only on the nature of the sensor, but also on the signal strength and its direction of arrival. Hence the output response $x_i(t)$ of the $i$th sensor due to all the $d$ sources is given as

$$x_i(t) = \sum_{k=1}^{d} a_i(\theta_{ik}) s_k(t - \tau_{ik}) + n_i(t),$$  \hspace{1cm} (2.1)$$

where $a_i(\theta_{ik})$ is the response of the $i$th sensor to the $k$th source arriving at an angle $\theta_{ik}$ and $s_k(t)$ is the signal strength of the $k$th source at time $t$. $n_i(t)$ is the noises at the $i$th sensor and $\tau_{ik}$ is the time delay at the $i$th sensor of the $k$th source. Note that besides the sensor response, $a_i(\theta_{ik})$ also includes the channel (transmission medium) characteristics such as amplitude modulation and phase change. Hence $a_i(\theta_{ik})$ is complex in general. The time delays $\tau_{ik}$ represents information of the spatial distribution of the sensors and the source signals. They are functions of sensor locations, source locations, distance between sensors. Before obtaining the
final mathematical model of the problem, we have to consider various aspects of the emitting sources and formulate the problem in accordance.

2.1.2 Nature of the radiating sources

Far-field and near-field sources

From Figure 2.1, we can see that when the emitting source is very far away from the sensors (e.g. Source 2), the signal's wavefronts may be considered as parallel and will arrive at each sensor with the same angle. This kind of source is called far-field emitters. On the other hand, if the radiating source locates near to the sensors (e.g. Source 3), the wavefronts are no longer parallel with respect to different sensors. Therefore the signal will impinge on the sensors with different angles of arrival. This kind of source is known as near-field source. The presence of near-field sources is a nuisance to most of the array signal processing algorithms as they complicate the problem. Fortunately in many practical situations, all the sources can be considered as far-field and hereafter, we will assume this is the case. With this assumption, the angles of arrival are independent of the sensors and hence we can write $\theta_{ik} = \theta_k$ for all $i$.

Wideband and narrowband signals

In general the signal emitted by the $k$th source will have a complex quadrature representation as follows:

$$s_k(t) = \hat{s}_k(t) \cos(\omega_0 t + \psi_k(t)) - j \hat{s}_k(t) \sin(\omega_0 t + \psi_k(t))$$
$$= \hat{s}_k(t) e^{-j(\omega_0 t + \psi_k(t))},$$

(2.2)

with $\hat{s}_k(t)$ and $\psi_k(t)$ being the signal envelope and phase of $s_k(t)$ respectively and $\omega_0$ being the center frequency of the signal. According to the nature of $\hat{s}_k(t)$ and $\psi_k(t)$, we can classify the problem into two situations, namely wideband and narrowband cases.

A signal is said to be narrowband if the highest frequency component of its complex envelope $\hat{s}_k(t) e^{-j\psi_k(t)}$ is much smaller than the center (carrier) frequency $\omega_0$, otherwise,
it is a wideband signal. In other words, with \( s_k(t) \) in the form of (2.2), if \( \tilde{s}_k(t) \) and \( \psi_k(t) \) are slowly varying functions of time, then \( s_k(t) \) is a narrowband signal. By slowly varying, we mean that the approximations

\[
\tilde{s}_k(t - \tau) \approx \tilde{s}_k(t), \quad (2.3)
\]

\[
\psi_k(t - \tau) \approx \psi_k(t), \quad (2.4)
\]

are valid for all \( \tau \leq \zeta \), where \( \zeta \) is the maximum difference in delay times between any two sensors in the array. That is to say, the variations of the signal envelope over a period of less than \( \zeta \) is negligible. Assume that all the sources are narrowband and let \( \tau_{ik} \) be the time taken for the signal \( s_k(t) \) to reach the \( i \)th sensor. Without loss of generality, let

\[
\tau_{1k} \leq \tau_{2k} \leq \cdots \cdots \leq \tau_{mk}, \quad (2.5)
\]

and

\[
\tau_{mk} - \tau_{1k} \leq \zeta, \quad (2.6)
\]

then we can recast (2.1) as

\[
x_i(t) = \sum_{k=1}^{d} a_i(\theta_k) s_k(t - \tau_{ik}) + n_i(t)
\]

\[
= \sum_{k=1}^{d} a_i(\theta_k) \tilde{s}_k(t - \tau_{ik}) e^{-j\omega_0(t - \tau_{ik})} e^{-j\psi_k(t - \tau_{ik})} + n_i(t)
\]

\[
= \sum_{k=1}^{d} a_i(\theta_k) \tilde{s}_k(t - \tau_{1k} - (\tau_{ik} - \tau_{1k})) e^{-j\omega_0(t - \tau_{1k} - (\tau_{ik} - \tau_{1k}))} e^{-j\psi_k(t - \tau_{1k} - (\tau_{ik} - \tau_{1k}))} + n_i(t)
\]

\[
\approx \sum_{k=1}^{d} a_i(\theta_k) e^{j\omega_0(\tau_{ik} - \tau_{1k})} \tilde{s}_k(t - \tau_{1k}) e^{-j(\omega_0(t - \tau_{1k}) + \psi_k(t - \tau_{1k}))} + n_i(t)
\]

\[
= \sum_{k=1}^{d} a_i(\theta_k) e^{j\omega_0(\tau_{ik} - \tau_{1k})} \tilde{s}_k(t - \tau_{1k}) + n_i(t).
\]

The transition from (2.7) to (2.8) follows from (2.3) and (2.4). By grouping the responses of the sensors into a single vector, and without loss of generality letting \( \tau_{1k} = 0 \), we get a compact form
\[ x(t) = A(\theta)s(t) + n(t), \quad (2.9) \]

where

\[ x(t) = [x_1(t) \ x_2(t) \ \cdots \ x_m(t)]^T, \quad (2.10) \]

\[ \theta = [\theta_1 \ \theta_2 \ \cdots \ \theta_d]^T, \quad (2.11) \]

\[ A = [a(\theta_1) \ a(\theta_2) \ \cdots \ a(\theta_d)], \quad (2.12) \]

\[ a(\theta_k) = [a_1(\theta_k)e^{-j\omega_0\tau_{1k}} \ a_2(\theta_k)e^{-j\omega_0\tau_{2k}} \ \cdots \ a_m(\theta_k)e^{-j\omega_0\tau_{mk}}]^T, \quad (2.13) \]

\[ s(t) = [s_1(t) \ s_2(t) \ \cdots \ s_d(t)]^T, \quad (2.14) \]

\[ n(t) = [n_1(t) \ n_2(t) \ \cdots \ n_m(t)]^T. \quad (2.15) \]

The quantities \( a(\theta) \in \mathbb{C}^m \) and \( A(\theta) \in \mathbb{C}^{m \times d} \) are called the response vector for the emitter at DOA \( \theta \) and steering matrix of the sensor array respectively.

For the wideband case, the change in the signal amplitude and phase is comparable to the time delay \( \tau \) and the approximations in (2.3) and 2.4 are no longer valid. In such case, the array response cannot be written in the compact matrix form as (2.9) in time domain. However, we can reformulate the problem in the frequency domain and get similar results. It can be shown that by taking Fourier Transform, the following equation is obtained for wideband case:

\[ x(\Omega) = A(\Omega, \theta)s(\Omega) + n(\Omega), \quad (2.16) \]

with

\[ A(\Omega, \theta) = [a(\theta_1, \Omega) \ a(\theta_2, \Omega) \ \cdots \ a(\theta_d, \Omega)], \quad (2.17) \]
\[ a(\theta_k, \Omega) = [a_1(\theta_k)e^{-j\Omega \gamma_k} \ a_2(\theta_k)e^{-j\Omega \gamma_2k} \ \ldots \ \ldots \ \ldots \ a_m(\theta_k)e^{-j\Omega m_k}]^T, \]  

(2.18)

\[ s(\Omega) = \mathcal{F}\{s(t)\}, \quad n(\Omega) = \mathcal{F}\{n(t)\}, \]  

(2.19)

and \( \Omega \) being the frequency variable. Comparing (2.9) and (2.16), it can be seen that the major difference between the two cases is that for narrowband signal, \( A(\theta) \) is independent of time and frequency (constant matrix) while \( A(\Omega, \theta) \) is a function of frequency \( \Omega \) (hence changes with frequency) in the wideband case. Various techniques have been developed for the processing of wideband signals where some of these techniques utilized methods for narrowband signals. In this thesis, to simplify the discussions, all signals are assumed to be narrowband.

### 2.1.3 The statistical behaviour of the signals and noises

In the above analysis, the formulation is derived in the continuous time domain. However in reality, due to the use of digital computers, we have to process the data in discrete time domain. Thus sampling of the continuous time signal is required. After sampling, we will get a number of “snapshots” which consist of data vectors of the form

\[ x(t_s) = A(\theta)s(t_s) + n(t_s), \quad s = 1, 2, \ldots, N, \]  

(2.20)

where \( t_s \) is the sampling time. Note that the sampling instances may be arbitrary and may not be equally spaced.

As clear from (2.20), the statistical behaviour of snapshots \( x(t_s) \) depend on that of the signal vectors \( s(t_s) \) and noise vectors \( n(t_s) \). Hence, we have to consider the statistical models of both the signals and noises [7].

**Signal Vectors**

To model the statistics of the signal vectors, two approaches are commonly used, namely the Deterministic Signals (DS) model and the Stochastic Signals (SS) model which are described below:
DS: The signals $s(t_s)$ are regarded as unknown deterministic constants.

SS: The signals $s(t_s)$ are independent and identical Gaussian random vectors with zero mean and covariance matrix $S$.

In both models, the shape of the signals and the correlation amongst them are arbitrary. In particular, the signals can be fully correlated as in the case of multipath propagation in communications.

**Noise Vectors**

Similar to the case of signals, there are also two frequently used statistical models for the noise. The first one is referred to as White Noise (WN) model and the other as Colored Noise (CN) model.

**WN**: The noise samples $n(t_s)$ are statistically independent Gaussian random vectors with zero mean and covariance matrix $\sigma^2 I$.

**CN**: The noise samples $n(t_s)$ are statistically independent and Gaussian with zero mean and covariance matrix $\sigma^2 \Sigma_N$. In fact this is just the generalized case of WN.

**Assumptions adopted in this thesis**

Concerning the statistics of the signals and noises, the following assumptions are made for the remaining of this thesis:

**Assumption A2.1**: The signals are assumed to be Stochastic signals (SS) with a positive definite covariance matrix $S$ which is unknown. The positive definite assumption is equivalent to assuming that the $d$ signal sources are not coherent.

**Assumption A2.2**: The noises are Colored Noises with known noise covariance matrix $\Sigma_N$ but unknown noise factor $\sigma$.

**Assumption A2.3**: The signals and noises are statistically independent of each other. That is to say, $E\{s(t_\alpha)n(t_\beta)^*\} = 0$ for all $\alpha$ and $\beta$. 
As will be explained in the coming chapters, these assumptions are indeed very important. To make the problem to be solvable, two more assumptions about the system are also made:

**Assumption A2.4** : The number of signals $d$ is assumed to be known.

**Assumption A2.5** : The steering matrix $A(\theta)$ (2.12) is assumed to be of full rank.

These two assumptions are very reasonable. Firstly, though in practical situations, $d$ may not be known in advance, it can be estimated accurately by using information theoretic criteria techniques like Akaike’s Information Criterion (AIC) [1] or Rissanen’s Minimum Description Length (MDL) Criterion [22]. Hence for simplicity, we assume it is known in our case. Moreover, since $A(\theta)$ depends not only on $\theta$ but also on the array configuration, by careful design, assumption A2.5 can be achieved. An example is given in the last part of this chapter when Uniform Linear Array (ULA) is considered.

To summarize, we consider the following problem: suppose there are totally $d$ far-field, narrowband emitting sources in an non-dispersive medium with $m$ sensors. By sampling the output of the sensors, we get a batch of snapshots of sensor array response which can be expressed in the following form:

$$x(l) = A(\theta)s(l) + n(l),$$  \hspace{1cm} (2.21)

where $x(l), n(l) \in \mathbb{C}^m$, $s(l) \in \mathbb{C}^d$, $A(\theta) \in \mathbb{C}^{d \times m}$ and $l = 1, 2, \ldots, N$ with $N$ being the total number of snapshots. The problem is to estimate the matrix $A(\theta)$ which carries information of the DOAs. Moreover, certain statistics (e.g. the emitting power or noise factor) of the signals and the noises may also be determined from the available data.

### 2.2 Other Mathematically Equivalent Problems

In this section, several examples of array signal processing problems are given and shown to have similar mathematical formulation as the DOA estimation problem derived in Section 2.1.
2.2.1 Cisoids retrieval

The problem of separating a mixture of cisoids in the presence of noise arises in many situations such as spectral estimation and analysis. Consider a signal $x(t)$ consists of $d$ sinusoids with unknown parameters embedded in additive noise $n(t)$ as

$$x(t) = \sum_{k=1}^{d} e^{-j(\omega_k t + \phi_k)} s_k + n(t),$$

where $s_k$, $\omega_k$ and $\phi_k$ represent the amplitude, frequency and phase of the $k$th sinusoid respectively. Suppose we sample the output signal $x(t)$ by using tapped-delay-line with $m$ equally spaced tapping line of interval time of $t_s$, then the $l$th discrete output from the delay-line can be written in the form

$$x(l) = \sum_{k=1}^{d} e^{-j(\omega_k (l-1) t_s + \phi_k)} s_k + n(l).$$

Hence by stacking outputs of the delay-line into a vector, we get the compact form:

$$x = As + n,$$  \hspace{1cm} (2.22)

with $s = [s_1 e^{-j\phi_1} \ s_2 e^{-j\phi_2} \ \ldots \ \ldots \ s_d e^{-j\phi_d}]^T$, $n = [n(0) \ n(t_s) \ \ldots \ \ldots \ n(mt_s)]^T$ and

$$A = \begin{bmatrix}
e^{-j\omega_1 t_s \cdot 0} & e^{-j\omega_2 t_s \cdot 0} & \ldots & \ldots & e^{-j\omega_d t_s \cdot 0} \\
e^{-j\omega_1 t_s \cdot 1} & e^{-j\omega_2 t_s \cdot 1} & \ldots & \ldots & e^{-j\omega_d t_s \cdot 1} \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
e^{-j\omega_1 t_s (m-1)} & e^{-j\omega_2 t_s (m-1)} & \ldots & \ldots & e^{-j\omega_d t_s (m-1)}
\end{bmatrix}$$

By comparing (2.21) and (2.22), the problem of cisoids separation is similar to that of DOA estimation.

2.2.2 Pole retrieval

Consider a linear system that is excited by an impulse response. Suppose the system has $d$ unknown and distinct poles in the complex plane. By simple linear system theory, the response of the system can be expressed as
\[ x(t) = \sum_{k=1}^{d} e^{\gamma_k + j(\beta_k t + \phi_k)} s_k + n(t), \]

where \( \gamma_k = \alpha_k + j\beta_k \) is the location of the \( k \)th pole in the complex plane with \( s_k \) and \( \phi_k \) being the residue and phase respectively. By applying a similar tapped-delay-line sampling procedure as in Section 2.2.1 and collecting the \( m \) consecutive outputs into a vector, we will obtain an equation similar to (2.22) with the same form of \( A \) (with \( \omega_k \) replaced by \( \beta_k \)) and \( n \) but a slightly different \( s \) given as follows:

\[ s = [s_1 e^{\delta_1 - j\delta_1} \quad s_2 e^{\delta_2 - j\delta_2} \quad \ldots \quad s_d e^{\delta_d - j\delta_d}]^T, \]

Hence once again, the problem is reduced to finding the matrix \( A(\theta) \) as in the case of DOAs estimation.

2.2.3 Known echo retrieval

Consider the situation of a radar or sonar system that transmits a known shape of pulse \( p(t) \) to probe the scatters (assume there are \( d \) such scatters) in a medium. Suppose the backscattered signals are of the same shape of \( p(t) \), then we can write the output of the receiving sensor due to all the \( d \) echoes as

\[ x(t) = \sum_{k=1}^{d} s_k p(t - \tau_k) + n(t), \]

where \( s_k, \tau_k \) represent the received signal strength and delay of the \( k \)th echo. Assume the pulse \( p(t) \) is of duration \( t_p \) and the maximum delay of the echoes are \( \tau_{max} \). By sampling the output at a time interval of \( t_s = \frac{t_p + \tau_{max}}{m} \), and putting the sampled data into a vector, we will get the following familiar equation

\[ \mathbf{x} = A \mathbf{s} + \mathbf{n}, \]  

(2.23)

with \( \mathbf{s} = [s_1 \quad s_2 \quad \cdots \quad s_d]^T \), \( \mathbf{n} = [n(0) \quad n(t_s) \quad \cdots \quad n((m-1)t_s)]^T \) and \( A \) is defined as

\[ A = [p_1 \quad p_2 \quad \cdots \quad p_d], \]  

(2.24)

\[ p_k = [p(0 - \tau_k) \quad p(t_s - \tau_k) \quad \cdots \quad p((m-1)t_s - \tau_k)]^T. \]
Note that if we perform the probing $N$ times, we will get $N$ snapshots and by adding the index $l$ to indicate the $l$th set of data, (2.23) will looks exactly the same as (2.21). Thus the model in this case fits the narrowband DOA estimation problem though with different steering matrix $A$.

In this application, we assume that the pulse shape $p(t)$ is known. In many situations, it may be unknown. In fact in Chapter 6 we will consider the problem of resolving unknown pulse shape (blind deconvolution) which utilize some techniques described in Chapter 4.

![Diagram of an Uniform Linear Array (ULA)](image)

**Figure 2.2:** Configuration of an Uniform Linear Array (ULA)

### 2.3 An Example: Uniform Linear Array (ULA)

In this section, the most common type of array sensor, namely Uniform Linear Array (ULA) is discussed. The basic configuration of an ULA is shown in Figure 2.2. It consists of $m$ identical sensors arranged in a line with equal distance $d_a$ between adjacent sensors. With this structured configuration, the steering matrix $A(\theta)$ has a very special form. To express $A(\theta)$ in terms of the DOAs of the signals, consider the single source case in Figure 2.3. Before proceeding with the derivations, we have the
following definitions which hold for the whole thesis:

Figure 2.3: Wavefronts from an emitting source with DOA $\theta_k$ impinge on a ULA with $m$ sensors separated at a distance of $d_a$. The $i$th and the $j$th sensors are shown to illustrate the relation between the path difference and the DOA $\theta_k$.

**Definition D2.1**: The DOA of a signal source is defined as the angle between the line of sensors and the signal ray impinging on the sensor array ($\theta_k$ as shown in Figure 2.3).

**Definition D2.2**: The leftmost sensor is referred as the reference sensor and the sensors are named in order from left to right as in Figure 2.3. Moreover, the time delay for the reference sensor is defined to be zero, i.e.

$$\tau_{rk} = 0 \quad \forall \ k = 1, \ldots, d.$$ 

With the above definitions, let us concentrate on the $i$th and the $j$th sensors. From Figure 2.3, the path difference between these two sensors is given by $(j - i)d_a \cos \theta_k$, where $\theta_k$ is the DOA of the $k$th emitting source. Similarly, the path difference between the reference and the $i$th sensors is $(i - 1)d_a \cos \theta_k$. Since the time delay of the reference sensor is zero, the time delay at the $i$th sensor due to the $k$th source, $\tau_{ik}$, is given as
\[
\tau_{ik} = \frac{(i-1)d_a \cos \theta_k}{c}, \tag{2.25}
\]

where \(c\) is the speed of signal in the transmission medium. Substituting (2.25) into (2.13), we get the following form of \(a(\theta_k)\)

\[
a(\theta_k) = [a_1(\theta_k)e^{-j\omega_0 \frac{d_a \cos \theta_k}{c}} \quad a_2(\theta_k)e^{-j\omega_0 \frac{d_a \cos \theta_k}{c}} \quad \cdots \quad a_m(\theta_k)e^{-j\omega_0 \frac{(m-1)d_a \cos \theta_k}{c}}]^T, \tag{2.26}
\]

Observing that \(\frac{c}{\omega_0}\) is just the wavelength \(\lambda_s\) of the signal in the medium. Moreover since the sensors are identical, they should have the same response and hence

\[
a_1(\theta_k) = a_2(\theta_k) = \cdots = a_m(\theta_k) = a(\theta_k). \tag{2.27}
\]

By letting \(\frac{d_a}{\lambda_s} = \frac{\tilde{d}_a}{\lambda_s}\), the steering matrix of the ULA problem is written as:

\[
A(\theta) = [a(\theta_1) \quad a(\theta_2) \quad \cdots \quad a(\theta_d)],
\]

\[
\theta = [\theta_1 \quad \theta_2 \quad \cdots \quad \theta_d]^T,
\]

\[
a(\theta_k) = a(\theta_k)[1 \quad e^{-j\tilde{d}_a \cos \theta_k} \quad e^{-j2\tilde{d}_a \cos \theta_k} \quad \cdots \quad e^{-j(m-1)d_a \cos \theta_k}]^T. \tag{2.28}
\]

From (2.28), we can see that \(A(\theta)\) is a Vandermonde matrix and hence if all the \(\theta_k\)'s are different, \(A(\theta)\) will be of full rank. This is a very important feature which is utilized in Chapters 3 and 4 to solve the ULA DOAs estimation problem.
CHAPTER 3

EIGEN-STRUCTURE BASED ALGORITHMS

Since the main theme of this thesis is to investigate the effects of random errors in eigen-structure based algorithms, it is essential to have a study in this kind of array signal processing algorithm. In this chapter, focus are made on the two representative methods, namely MUSIC and ESPRIT. Both system requirements and underlying principle of the techniques are described. Moreover, the major limitations of the methods are also discussed which will serve as a preliminary introduction to the novel analyzing techniques and compensating algorithms of the next chapter. Finally, other subspace fitting algorithms which are based on the above two methods are summarized.

3.1 Fundamental Principle of Eigen-Structure Based Algorithm

In this section, the concept of array manifold and the notion of eigen-decomposition are discussed. These mathematical preliminaries will help us to understand the rationales behind MUSIC and ESPRIT.

3.1.1 Array manifold

To introduce the concept of array manifold, consider (2.9) and (2.12) under the case where $n = 0$. In this case,

$$x(l) = A(\theta)s(l),$$

(3.1)
\[ A(\theta) = [a(\theta_1) \ a(\theta_2) \ \cdots \ a(\theta_d)]. \]

Recall that the columns of \( A(\theta) \), \( a(\theta) \) (the dependence of \( \theta \) on \( k \) is dropped for general case) is an element of \( \mathcal{C}^m \), and when \( \theta \) varies, \( a(\theta) \) varies accordingly, though it remains in \( \mathcal{C}^m \). Based on this feature, the array manifold \( \mathcal{A} \) is defined as follows:

**Definition D3.1** : **Array manifold** \( \mathcal{A} \) is defined as the set of all steering vectors \( a(\theta) \) obtained by varying \( \theta \) over its entire domain. If the value of \( \theta \) is restricted in the range of \([−\pi, \pi]\), then \( \mathcal{A} \) can be expressed mathematically as

\[ \mathcal{A} = \{a(\theta) \in \mathcal{C}^m \mid \theta \in [−\pi, \pi]\}. \quad (3.2) \]

For simplicity, we have assumed \( \theta \) to be the only parameter of \( a(\theta) \) which may not be true in some cases. \( \mathcal{A} \) is completely determined by the sensor geometry and sensitivity patterns. In some sensor array with regular structure and special directivity, the manifold can be determined analytically. As an example, consider the case of ULA in the last section of Chapter 2, by which the array manifold is given as the set of vectors in the form of (2.28). For cases where analytical form is not available, the array manifold can be obtained by calibration.

By (3.1), it can be observed that in the absence of noise, the sensor response \( x(l) \in \mathcal{C}^m \) is a linear combination of the columns of \( A(\theta) \) in proportions according to \( s(l) \). This induces the concept of signal subspace as follows:

**Definition D3.2** : The signal subspace \( \mathcal{S} \) is a vector subspace of \( \mathcal{C}^m \) with bases being the columns of the steering matrix \( A(\theta) \), which is fixed by a particular set of signal parameters \( \{\theta_k, k = 1, \ldots, d\} \).
Figure 3.1: The intersection of the Signal subspace and Array Manifold

By assumption A2.5 in Chapter 2, it can be seen that the signal subspace $S$ is of dimension $d$ as the columns of $A(\theta) \in \mathbb{C}^{m \times d}$ are assumed to be linearly independent. It should be noted that the concept of array manifold is completely different from that of signal subspace. In fact, in our one parameter case, the array manifold can be viewed as a rope weaving through $\mathbb{C}^m$ while the signal subspace $S$ is a $d$-dimensional hyperplane in $\mathbb{C}^m$. However, these two concepts are closely related by the fact that the intersecting points of the array manifold $\mathcal{A}$ and signal subspace $S$ in $\mathbb{C}^m$ will give the steering vectors corresponding to the DOAs. An example of two sources ($d = 2$) and three sensors ($m = 3$) is illustrated in Figure 3.1. From the figure, the signal subspace is determined by finding the two-dimensional plane that contains the response vectors $x(l)$ obtained at different instances. Note that when there are no noises, all the response vectors $x(l)$ must lie on the signal subspace. Hence if the array manifold is available, the steering matrix can be found easily under noiseless case. Unfortunately, noise is unavoidable in practical situations and therefore the output vectors $x(l)$ will not lie exactly on the signal subspace due to non-zero $n(l)$
and the above geometrical method of finding the DOAs fails.

### 3.1.2 Eigen-decomposition

To tackle the problem under noisy case, we need the techniques called eigen-decomposition. To simplify the notation, we will write \( A(\theta) \) as \( A \) hereafter when there are no ambiguities. When there are noises, the response is given by (2.9) as follows:

\[
x(l) = As(l) + n(l).
\]

For simplicity, we will first consider the case when \( n(l) \) is white noise and then generalize our results to the case of colored noise.

**White noise**

The covariance matrix \( R_{xx} \) of \( x \) is given by

\[
R_{xx} = \text{E}\{x(l)x^*(l)\}
= \text{E}\{(As(l) + n(l))(As(l) + n(l))^*\},
\]

By the statistical assumptions A2.1 - A2.3 in Chapter 2, we get the following compact form

\[
R_{xx} = ASA^* + \sigma^2 I,
\]

where \( S = \text{E}\{s(l)s^*(l)\} \) and \( \sigma^2 I = \text{E}\{n(l)n^*(l)\} \). By our assumptions, \( S \in \mathbb{C}^{d \times d} \) is non-singular (i.e. the sources are not coherent) and \( A \) is of full rank, hence \( ASA^* \) (which is positive semi-definite) must be of rank \( d \). Thus \( ASA^* \in \mathbb{C}^{m \times m} \) must have \( d \) positive eigenvalues and \( (m - d) \) zero eigenvalues. As a consequence, from the theory of linear algebra, the smallest \( (m - d) \) eigenvalues of \( R_{xx} \) are all equal to \( \sigma^2 \) and the eigen-decomposition of \( R_{xx} \) is given as

\[
R_{xx} = EAE^*,
\]

where \( E, A \) are matrices comprising of the eigenvectors and eigenvalues of \( R_{xx} \) respectively. By partitioning \( E \) into two submatrices \( E_S \) and \( E_N \) of dimensions
\( m \times d \) and \( m \times (m - d) \) containing the eigenvectors corresponding to the largest and smallest eigenvalues of \( R_{xx} \) respectively, (3.4) can be rewritten as

\[
R_{xx} = \begin{bmatrix} A_S & 0 \\ 0 & A_N \end{bmatrix} \begin{bmatrix} E_S^* \\ E_N^* \end{bmatrix} = E_S A_S E_S^* + E_N A_N E_N^* .
\] (3.5)

It is a well known result [14] that the eigenvectors of a matrix span the same range space as the columns of the matrix. Hence the signal subspace may still be obtained in noisy case as \( \mathcal{S} = \mathcal{R}(E_S) \).

From the above derivations, eigenvectors corresponding to the \( d \) largest eigenvalues of \( R_{xx} \) are related to the signal space. With a similar reasoning, the remaining \( (m-d) \) eigenvectors can be thought of constituting another vector space and it leads to the idea of noise space.

**Definition D3.3**: The noise subspace \( \mathcal{N} \) is a vector subspace in \( \mathcal{C}^m \) which is orthogonal to the signal subspace \( \mathcal{S} \). Note that if the dimension of \( \mathcal{S} \) is \( d \), then the dimension of \( \mathcal{N} \) must be \( (m-d) \) and we may express \( \mathcal{C}^m \) as the direct sum of the two spaces

\[
\mathcal{C}^m = \mathcal{S} \oplus \mathcal{N} .
\]

The space is called noise space as it may be thought as the space spanned by the "noise vectors" \( n(l) \). Hence in our case, the noise space can be determined by finding the range space of the matrix \( E_N \), i.e. \( \mathcal{N} = \mathcal{R}(E_N) \). (Note that \( E_N \) is orthogonal to \( E_S \) as they both contain eigenvectors of \( R_{xx} \)). The concept of noise space is of crucial importance to algorithms such as MUSIC which will be described shortly.

**Colored noise**

When the noise is CN, (3.3) must be rewritten as:

\[
R_{xx} = ASA^* + \sigma^2 \Sigma_N ,
\] (3.6)

Under such case, we have to apply Mahalanobius transformation to pre-whiten the noise. As we have assumed that \( \Sigma_N \) is positive definite and known, we may apply
Cholesky decomposition and decompose $\Sigma_N$ into a product of non-singular matrices as $\Sigma_N = \Sigma_N^{1/2} \Sigma_N^{*1/2}$. Substituting this into (3.6) and multiplying both sides on the left by $\Sigma_N^{-1/2}$ and on the right by $\Sigma_N^{-*1/2}$, we get

$$
\Sigma_N^{-1/2} R_{xx} \Sigma_N^{-*1/2} = (\Sigma_N^{-1/2} A) S (\Sigma_N^{-1/2} A)^* + \sigma^2 I. \quad (3.7)
$$

By letting $\hat{R}_{xx} = \Sigma_N^{-1/2} R_{xx} \Sigma_N^{-*1/2}$ and $\hat{A} = (\Sigma_N^{-1/2} A)$, (3.7) is reduced back to the same form as (3.3). Thus by applying the Mahalanobius transformation, the colored Noise case is reduced back to the white Noise case and the derivation discussed in the WN case can be applied with slight modifications.

To determine the steering matrix (and hence the DOA parameter), we have to find the intersecting point of the array manifold and the signal subspace. There are various methods of doing so, depending on the system configuration and the information available. In the coming sections, MUSIC and ESPRIT are discussed.

### 3.1.3 Source of Errors

In reality, the exact covariance matrix $R_{xx}$ is unknown and an estimate $\hat{R}_{xx}$ has to be found from the available data. Usually, we approximate $R_{xx}$ by the following method

$$
\hat{R}_{xx} = \frac{1}{N} \sum_{l=1}^{N} x(l)x^*(l), \quad (3.8)
$$

where $N$ is the total number of snapshots available. As the covariance matrix used in the eigen-decomposition is an estimate, the signal and noise subspaces obtained are disturbed and this gives rise to errors. This is called finite sample effect and may be a serious problem in eigen-based methods. Asymptotically when $N$ tends to infinity, (3.8) will give the exact covariance matrix. Thus when $N$ is very large, the finite sample effect may be ignored.

### 3.2 Multiple Signal Classification (MUSIC)

MUSIC was first developed in 1979 by Schmidt and since then, modified versions are proposed by several researchers [36] [26]. However, for the sake of simplicity, the
most common one is discussed. Although the case is somewhat simplified, it does reveal the most important aspects of MUSIC.

3.2.1 System requirements

The development of MUSIC had arouse great interests as it offered remarkable breakthroughs in array signal processing at that time. The greatest contributions of MUSIC is that the array used can be of arbitrary geometry or configuration. This will ease the task of designing an array sensor with specified array sensitivity pattern for use in particular situations.

The most fundamental system requirements of MUSIC is that the array geometry is known in advance. That is to say, the complete information of the array manifold $\mathscr{A}$ must be obtained either analytically or by experimental calibration. For structured array, such as ULA, the array manifold can be expressed in mathematical forms (2.28). For unstructured one, experiments must be done to obtain the manifold and stored as data. Note that in this case, continuous manifold information cannot be obtained as the DOAs used for calibration must be in discrete domain. In view of this, structured array is preferred.

3.2.2 Underlying principle

Provided with the complete knowledge of the array manifold and signal subspace (the noise subspace is also available) obtained from the eigen-decomposition method described in the previous section, the major problem is how to search for their intersecting points to find the DOAs. The task may be easy for one dimensional case (only one emitting source) with not more than three sensors because we can search extensively in one dimension. However, for higher dimension (more than two sources with an array consisting of six to ten sensors), exhaustive search is almost impossible. To solve this problem, Schmidt has utilized the unitary property of the signal and noise subspaces. The key idea can be explained by the fact that for a particular DOA (viz. $\theta$), the vector $\mathbf{a}(\theta)$ must lie on the signal subspace and therefore must be unitary to the noise subspace. i.e.
\[ a^*(\theta)E_N = 0 \quad \forall \theta \in \Theta , \quad (3.9) \]

where \( \Theta \) is the set containing the actual DOAs of the emitting sources.

Based on this argument, instead of using (3.9), Schmidt proposed a much better objective function as follows:

\[ P_M(\theta) = \frac{a^*(\theta)a(\theta)}{a^*(\theta)E_N E_N^*a(\theta)} \quad (3.10) \]

This objective function \( P_M(\theta) \), termed as the MUSIC spectrum, is actually the inverse of the norm of the projection function of \( a(\theta) \) on the space spanned by the noise space \( E_N \). The factor \( a^*(\theta)a(\theta) \) serves as a normalization factor and may be omitted. In the ideal situation, where the array manifold is exact (i.e. we know the form or data of \( a(\theta) \) for every \( \theta \)) and the signal subspace is free of noise, the denominator of (3.10) should be zero and \( P_M(\theta) \) is infinite for those \( \theta \) corresponding to the actual DOAs. In the practical application, however, only peaks (not infinity) will occur at points nearest to the true DOAs as the signal subspace (and hence noise subspace) will unavoidably be contaminated with noise due to the finite sample effect. Therefore to determine the DOAs, we can perform an one dimensional search for peaks of the function \( P_M(\theta) \) over reasonable range of \( \theta \) (e.g. from \([-\pi, \pi]\)) and the DOAs are located at points where \( P_M(\theta) \) shows peaks. We can see that Schmidt has successfully reduced the \( d \)-dimensional searching problem to an one-dimensional one.

### 3.2.3 Sources of error and limitations

The greatest limitation of MUSIC is that the array manifold of the system must be known in advance. For certain configurations such as ULA, this is not a problem. But when the array response is not regular, difficulties arise in calibrating the manifold. Calibration experiments may be done in certain range of DOAs only and hence restrict the use of this method in cases with extreme DOAs.

The second drawback of MUSIC is that the performance of the algorithm degrades when the signal to noise ratio (SNR) is low. In this case, peaks of the projection function \( P_M(\theta) \) corresponding to closely spaced emitting sources (with very close DOAs) will cluster together and may be lumped into one single peak. This is due to
the fact that when the DOAs are closed to each other, the condition number of $A$ is very large\(^1\) and the estimated covariance $\hat{R}_{xx}$ is prone to noise perturbations, giving rise to heavily perturbed signal subspace.

Finally, as common to all the subspace fitting methods, MUSIC is very sensitive to model or manifold errors. Moreover random shifts in the array sensor response and changes in sensing environment due to unknown reasons cause undesirable degradation of the algorithm. However, the problem is left till Chapter 4 where a comprehensive discussion of various kind of errors will be presented.

3.3 Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT)

3.3.1 System requirements

As revealed in the previous Section 3.2.3, although MUSIC may considered as the breakthrough method in fully utilizing the knowledge of the array data model, the requirement of knowing the complete information of the array manifold restricts its use. In view of this, Roy proposed another subspace fitting method called ESPRIT. The idea of this algorithm was first published in 1986 and a more complete descriptions was published in 1987 in Roy’s Ph.D dissertation [23]. ESPRIT shares the same eigen-decomposition algorithms and features of MUSIC and retains the flexibility of allowing arbitrary array geometry. The major advantage of ESPRIT over MUSIC is that it does not require complete knowledge of the array manifold. Instead, it imposes a constraint on the structure of the sensor array. The constraint may be considered as doubling the sensor array in a special way and it is illustrated in Figure 3.2.

The sensor array consists of a number of sensor doublets (assume $m$ doublets) which can be located at any position. Each sensor doublet consists of two identical sensors which are separated at a fixed but known displacement $\Delta$. The responses and the sensitivities of the sensor doublets may be arbitrary provided that they are the same within the doublet. There will be obvious advantages if the sensor doublets are the same although it is not a necessary requirements. In fact, the ULA described in

\(^1\)The condition number of a matrix is defined as the ratio of its largest eigenvalue to its smallest one.
Chapter 2 satisfies these constraints and is eligible for ESPRIT. Note that it is the property of fixed translational displacement within the doublets that allows us to find the DOAs without knowing the whole array manifold. For the sake of presentation in the next section, we will adopt a numbering system for the sensors. The first sensors in each doublet is named as odd sensors (Sensor 1, 3, \ldots, (m - 1)) and the second one is termed even sensors (Sensor 2, 4, \ldots, m).

3.3.2 Underlying principle

The basic idea of ESPRIT is to exploit the geometrical translation within the sensor doublets and transform it to a rotational invariance property of the signal subspace. Consider the outputs of all the odd sensors as

\[ \mathbf{x}_o(l) = \mathbf{A}s(l) + \mathbf{n}_o(l), \]  

where \( \mathbf{x}_o(l) \) and \( \mathbf{n}_o(l) \) represent outputs and the noise of the odd sensors of the \( l \)th snapshots. \( \mathbf{A} \) is the steering matrix of the array consisting of odd sensors only and is
of the same form as (2.12) and (2.13). Since the even sensors are all displaced at a distance $\Delta$ from their odd counterparts, it can be shown that their response vector is given by the equation

$$x_e(l) = A\Phi s(l) + n_e(l),$$

with

$$\Phi = \text{Diag}(e^{-j\frac{\pi\Delta}{\lambda} \cos \theta_1}, e^{-j\frac{\pi\Delta}{\lambda} \cos \theta_2}, \ldots, e^{-j\frac{\pi\Delta}{\lambda} \cos \theta_m})^T).$$

Note that $\Phi$ not only contains information about the DOAs of the emitting sources, it also relates the outputs of the odd and even sensors. By stacking the odd and even response vectors, we get

$$z(l) = \begin{bmatrix} x_o(l) \\ x_e(l) \end{bmatrix} = \begin{bmatrix} A \Phi s(l) \\ A \Phi s(l) \end{bmatrix} = \tilde{A}s(l) + \tilde{n}(l),$$

with $\tilde{A} = \begin{bmatrix} A \\ A\Phi \end{bmatrix}$ and $\tilde{n}(l) = \begin{bmatrix} n_o(l) \\ n_e(l) \end{bmatrix}$.

Now we may treat $z(l)$ as a single vector and calculate its covariance matrix as $R_{zz} = E \{ z(l)z^*(l) \}$. By performing eigen-decomposition on $R_{zz}$, let $\tilde{E}_S \in \mathbb{C}^{2m \times d}$ contains the eigenvectors corresponding to the $d$ largest eigenvalues. Since $\tilde{E}_S$ has the same range space of $\tilde{A}$, there exists a non-singular matrix $H$ such that

$$\tilde{E}_S = \begin{bmatrix} E_{So} \\ E_{Se} \end{bmatrix} = \tilde{A}H,$$

and $\mathcal{R}\{E_{So}\}$ and $\mathcal{R}\{E_{Se}\}$ may be considered as the signal subspaces due to $A$ and $A\Phi$ respectively. Since $\Phi$ is a non-singular diagonal matrix, $A$ and $A\Phi$ should share the same signal subspace, i.e.

$$\mathcal{R}\{E_{So}\} = \mathcal{R}\{E_{Se}\} = \mathcal{R}\{A\}.$$

By (3.15), we may write

$$E_{So} = AH,$$
\[ E_{se} = A\Phi H. \]  \hspace{1cm} (3.17)

Multiplying (3.16) on the left by \( H^{-1} \) (\( H \) is non-singular) and substituting into (3.17), it becomes

\[ E_{se} = E_{so} H^{-1} \Phi H = E_{so} \Psi, \]  \hspace{1cm} (3.18)

Note that \( \Phi \) (and \( \theta \)) can be found if \( \Psi \) is known. Hence the problem is reduced to a least square problem of finding a non-singular matrix \( \Psi \) such that (3.18) holds. Since both \( E_{so} \) and \( E_{se} \) are obtained from the estimated covariance matrix, they contain errors and ordinary least square methods involving matrix pseudo-inverse \( (E_{so}^T) \) cannot be used. On account of this, Roy has proposed using Total Least Square (TLS) to find an optimal solution for \( \Psi \). The method requires the use of Singular Value Decomposition (SVD) technique as described in [8] and is omitted here. After finding \( \Psi \), we can perform an eigen-decomposition and the eigenvalues will then be the elements of the diagonal matrix \( \Phi \) as they are related by the equation

\[ \Phi = H^{-1} \Psi H. \]

The values of \( \theta_k \) for \( k = 1, 2, \ldots, d \) can then be deduced by (3.13).

3.3.3 Advantages and sources of errors

The advantage of ESPRIT is that the DOAs are calculated directly without searching. This reduces the computational complexity of the problem and increases the DOA resolving power as compared with MUSIC. Moreover, it relaxes the need to obtain the complete information of the array manifold, hence no calibration is required. However, all of these advantages must be paid for by increasing system resources such as doubling the number of sensors used. Besides, errors are serious if the sensors within the doublets are not exactly the same or if the displacement between them are varying. Thus though ESPRIT seems superior over MUSIC, it has its own drawbacks.
3.4 Other Subspace Fitting Methods

To conclude the chapter, the other subspace fitting methods are mentioned though they are not discussed in details. In fact, other than ESPRIT and MUSIC, Maximum Likelihood (ML) is the other commonly used method. However, this method involves exhaustive search over the space concerned and is not practical in the computational point of view. ML based on the principle of maximizing a likelihood function derived from the statistics of the sources and noises. ML may be divided into Stochastic ML or Deterministic ML depends on whether the sources are treated as deterministic or stochastic processes.

Recently, researchers have proposed improved versions of MUSIC and ESPRIT. MD-MUSIC (Multi-dimensional MUSIC) is a generalization of MUSIC to higher dimensional parameter space. Multiple Invariance ESPRIT improves ESPRIT by allowing the sensor array to have multiple invariance instead of single invariance as described previously. For example, we may treat the ULA as a overlapping ESPRIT array geometry and by overlapping the output data, different degree of invariance can be introduced. Weighted subspace fitting (WSF) improves the algorithms by introducing a weighting matrix to the objective function during optimization [2].
CHAPTER 4

NOVEL APPROACH IN ERROR ANALYSIS

Following the arguments in Chapter 3, the performance of both MUSIC and ESPRIT degrades greatly when there are system errors and model uncertainties. In this chapter, a comprehensive study of the effects of these errors on the two methods are studied. A brief background on the analyzing techniques and error models commonly used by researchers is first given as an introduction. Then an innovative error model is proposed. By using this model, a novel approach of analyzing the behaviour of the algorithms under erroneous situation is devised. By observing the results of the analysis, a compensating algorithm is proposed and shown to be effective in compensating the errors under certain reasonable assumptions. The discussion is divided into two parts. Focus will first be made on DOA dependent perturbations. Then the idea will be carried forward to DOA independent errors. Simulation results are included to demonstrate the effectiveness of the compensating algorithm.

4.1 Background

From Section 3.1 of Chapter 3, it is clear that the performance of subspace fitting methods depends heavily on the estimated covariance matrix which is given by \( \hat{R}_{xx} = \frac{1}{N} \sum_{l=1}^{N} x(l) x^*(l) \). It is because the signal subspaces (and the noise subspaces) are obtained from the eigen-decomposition of \( \hat{R}_{xx} \). As a result, by considering the difference between the ideal covariance matrix \( R_{xx} \) and the estimated one, \( \hat{R}_{xx} \), the effect of various kinds of errors on the algorithms can be deduced. For
this reason, the error analysis of eigen-structure based subspace fitting method are divided into two main directions. The first one concerns about the finite sample effect of the data (i.e. the effect of the number of snapshots $N$ used in estimating $R_{xx}$, (2.14)). The literature published in this aspect mainly study the asymptotic behaviour of the algorithms when $N$ tends to infinity. For example, Ottersten, Viberg and Kailath have made an analysis on the TLS ESPRIT [17]. The Cramér-Rao bound (CRB) for ESPRIT problem formulation is derived (the definition of CRB will be given in Chapter 5). Comparisons by simulations between ESPRIT and other subspace fitting methods are also made with respect to their asymptotic performance. In [18], more generalized results are obtained.

The second line of research mainly deals with the effect of sensor and environment uncertainties which is also the main direction we will follow hereafter. In fact, in most cases, the errors in array response modeling dominates over the effect of finite sample size and when $N$ is reasonably large, the latter can be ignored. To investigate how the deviations in array model affect the estimates of the DOAs, an error model has to be proposed. The most commonly used model is the “additive Gaussian model” by which the steering matrix is written as $A + \delta A$ instead of $A$ with $\delta A \in \mathbb{C}^{m \times d}$ being a Gaussian distributed random matrix. The statistics of $\delta A$ are assumed to be

$$\delta A = [\delta a(\theta_1) \; \delta a(\theta_2) \; \cdots \; \delta a(\theta_d)] ,$$

$$E\{\delta a(\theta_k)\} = 0 ,$$

$$E\{\delta a(\theta_k)\delta a^\ast(\theta_l)\} = \sigma^2_{kl} I \; \delta_{kl} ,$$

where $\delta_{kl}$ is the Kronecker delta. Based on this model, Swindlehurst has derived a compact expression for the covariance matrix of the estimation error [31]. However, as pointed out by Soon and Huang in [28], this model may not be physically justifiable as most of the errors in practical situation cannot be fitted into such formulation. Moreover, although Swindlehurst et al. have made a in-depth analysis of the situation, the algorithm they proposed to compensate for the errors under their error model is computationally extensive and therefore cannot be put into practice. However, by following the arguments in the coming section, we will reinforce Huang’s idea and
a new error model which is more realistic and appropriate for analysis is proposed. Besides, a simple, robust and effective compensating method is devised. Since there are quite a lot of error sources, we will divide the analysis into two parts, namely DOA dependent and DOA independent errors. The former is discussed in details and the principle derived is applied to the latter case in a similar fashion.

4.2 DOA Dependent Errors

4.2.1 Sources of errors

In deriving (3.3), it is assumed that except the Gaussian distributed noise, there are no other sources of errors. However, in reality such ideal case does not exist as there are random perturbations arising from system uncertainties or from artificial interference. For example, in the beginning of Chapter 1, we have assumed that the propagating and the sensing medium is non-dispersive. This assumption is not valid in many situations such as underwater sonar application or radar detection. The transmission media in these cases are dispersive and the DOAs will be disturbed by the dispersion. Moreover, the presence of tiny particles in the media will act as refracting or diffracting sources and alter the directions of waves emitted from the sources, thus affecting the DOAs estimated by the array sensor.

Besides the above uncertainties in sensing environment, there are random errors arising from the sources. The most common one is due to the inconsistency in the positions of the emitting sources. As an example, in target tracking, the concerned objects may be moving around within a small area or even oscillating in all directions. Thus the DOAs will change slightly and randomly from time to time during the period of data acquisition. Another possible cause of randomness is the phenomenon called "jamming". In the military point of view, in order to escape from radar tracking, the signals emitted by the aircraft or small spy satellites may be jammed to interfere opponent's sensors. There are a number of jamming techniques and one of them exploits randomness in the emitting directions of the signals.
Figure 4.1: Figure showing the changes of DOAs of the sources

To illustrate the various kinds of random perturbations mentioned above, some examples are depicted in Figure 4.1. Although all these perturbations are of different nature, they affect the system in the same way: the DOAs of the sources are not consistent at all instances. This means that the snapshots we obtained are in fact corresponding to different DOAs. Hence our derivations for (2.9), (2.12) and (2.13) in Chapter 2 and the form of covariance matrix in (3.3) will not be correct. The subspace fitting techniques described in Chapter 3 cannot be used directly under such
circumstances. Before rederiving the system equations and presenting the modified estimation algorithm under the perturbed situation, some important assumptions about the random perturbations are made. Under these assumptions, we will arrive at another compact expression for the covariance matrix.

4.2.2 Assumptions

Although the idea presented below is generally applicable to a wide variety of subspace methods with arbitrary array geometry, for illustrative purpose we will focus on the case of ULA sensors in MUSIC and ESPRIT. For the sake of explanation, the equations concerning the ULA problem discussed in Chapter 2 are restated as follows:

\[ \mathbf{x}(l) = \mathbf{A}s(l) + \mathbf{n}(l), \]  

\[ \mathbf{A} = [a(\theta_1) \ a(\theta_2) \ \cdots \ a(\theta_d)], \]  

\[ a(\theta_k) = a(\theta_k)[1 \ e^{-j\hat{d}_a \cos \theta_k} \ e^{-j2\hat{d}_a \cos \theta_k} \ \cdots \ e^{-j(m-1)\hat{d}_a \cos \theta_k}]^T, \]  

The assumptions are first listed and then followed by detailed explanations.

**Assumption A4.1**: All the DOA dependent uncertainties can be represented by random shifts in the DOA \( \theta \). Thus for the \( l \)th snapshots, the DOA of the \( k \)th source is \( \theta_k + \delta \theta_k(l) \) rather than \( \theta_k \).

**Assumption A4.2**: The sensor outputs are not affected by the above randomness, i.e. \( a(\theta_k + \delta \theta_k(l)) \approx a(\theta_k) \ \forall \ k = 1, 2, \ldots, d \).

**Assumption A4.3**: The random perturbations are statistically independent with the noises and the signals.

**Assumption A4.4**: Instead of \( \delta \theta_k(l) \), it is \( \delta \theta_k(l) \cdot \sin \theta_k \) which follows a particular statistical distribution (such as Gaussian, uniform, etc.).

**Assumption A4.5**: \( \delta \theta_k(l) \cdot \sin \theta_k \) is independent, identically distributed (i.i.d.) for each \( l \) and \( k \) and may be considered as the outcomes of a random variable, viz. \( \delta \theta \cdot \sin \theta \).
Assumption A4.6: The random perturbations must be of zero means, i.e.

\[ E\{\delta \theta \cdot \sin \theta\} = 0. \]  

(4.4)

Since the perturbations arisen from different reasons are of different nature, it is difficult to introduce different error models for them and perform analysis one by one. By A4.1, however, we considered all the errors as a whole and represented them together as random shifts in the DOAs of the sources. By A4.2, it is assumed that the errors are DOA dependent. The effects of uncertainties in sensor locations, gains or phases will be discussed in the section 4.3 as DOA independent errors. By this assumption, we means that the perturbations only affect the phase shifts caused by the geometry of the array (in ULA case, it refers to the constant displacement between the sensors). The statistical independence of the perturbations with the noises and signals (A4.3) is a reasonable assumption since they are arisen from different origins.

The rationale behind assumption A4.4 can be explained by the help of Figure 4.2 which shows the situation of two signals from two sources with different DOAs \( \theta_i \) and \( \theta_j \) impinging on the sensor array. The key parameter to be noted is the path difference \( \zeta \) as it affects the time delay (hence the phase response and the steering matrix) between the sensors in the array. From the figure, it can be seen that when the incident angle (\( \theta_j \)) is large (and so is \( \sin \theta_j \)), the path difference \( \zeta_j \) between any two sensors will be small. Thus even a small angle perturbation \( \delta \theta_j \) will give rise to large distortion in path difference. On the other hand, when the angle (\( \theta_i \)) is small (so is \( \sin \theta_i \)), \( \zeta_i \) will be large and hence larger angle deviations can be allowed. On account of this, in making assumptions about the statistics of \( \delta \theta_k \), we should take into account the factor \( \theta_k \). Here we propose to add the factor \( \sin \theta_k \) (which is a monotonic increasing function in the range [0, \( \pi / 2 \)]) to \( \delta \theta_k \) so that \( \delta \theta_k \cdot \sin \theta_k \) will follow certain distributions. In fact, we can explain this assumption from another point of view. If the emitting sources are close to each other, \( \sin \theta_k \) is an indication of the distance traveled by the plane waves before reaching the sensors. When \( \theta_k \) is small, the waves will travel a long distance before arriving at the array and thus the random angle deviations \( \delta \theta_k \) will be larger. It is because deviations caused by dispersions and refractive or diffractive tiny particles will increase with the distance traveled by the waves. When \( \theta_k \) is large, the plane waves will undergo smaller angle
perturbations as the distance traveled by the waves is relatively shorter. This explains the appearance of the factor \(\sin \theta_k\) in the statistical distributions of the random angle deviations.

![Diagram showing wave paths and angle changes](image)

Figure 4.2: The relation between the change in path difference and the change in impinging angle

Finally, as we have assumed that the sources are independent of each order (see assumption A1.2) and the snapshots are taken at independent instances, we can treat all the \(\delta \theta_k(l) \cdot \sin \theta_k\) of different sources at different snapshots be i.i.d. as in A4.5. The zero means assumption in A4.6 ensures that the DOAs are perturbed around a nominal value. Equipped with the above assumptions, we are now ready to reformulate the equations of the ULA problem.
4.2.3 Mathematical formulation

By incorporating the random DOAs shifts into the model described in Section 2.3 and employing first order approximation to the perturbations

\[ \cos(\theta_k + \delta \theta_k(l)) \approx \cos \theta_k - \delta \theta_k(l) \cdot \sin \theta_k , \]

(4.5)

(4.1) is rewritten as

\[ \bar{\mathbf{x}}(l) = (\mathbf{A} \circ \delta \mathbf{A}(l)) \mathbf{s}(l) + \mathbf{n}(l) , \]

(4.6)

where \( \circ \) denotes Schur-Hadamard product between two matrices and \( \delta \mathbf{A}(l) \) is given by

\[ \delta \mathbf{A}(l) = [\delta a(\theta_1, \delta \theta_1(l)) \; \delta a(\theta_2, \delta \theta_2(l)) \; \cdots \; \delta a(\theta_d, \delta \theta_d(l))] , \]

(4.7)

\[ \delta a(\theta_k, \delta \theta_k(l)) = [1 \; e^{j \pi \tilde{d}_a(\delta \theta_k(l) - \sin \theta_k) - \sin \theta_k} \; \cdots \; e^{j \pi \tilde{d}_a(m-1)\delta \theta_k(l) - \sin \theta_k}]^T . \]

(4.8)

Note that \( \mathbf{A} \) is still of the same form as in (4.2) and (4.3).

Based on A4.3, A4.4, A4.5 and A4.6, we have the following theorem:

**Theorem T4.1:** The covariance matrix of the perturbed data vector \( \bar{\mathbf{x}} \) is given by

\[ \mathbf{R}_{\bar{x} \bar{x}} = (\mathbf{A} \mathbf{S \mathbf{A}}^*) \circ \mathbf{V} + \sigma^2 \mathbf{\Sigma}_N , \]

(4.9)

where \( \mathbf{V} \) is a perturbation matrix with the \( kl \)th element given by

\[ V(k, l) = E\{\exp(j \pi \tilde{d}_a(k - l) \delta \theta \cdot \sin \theta)\} . \]

(4.10)

**Proof:** see Appendix A.
For example, if $\delta \theta \cdot \sin \theta$ is uniformly distributed with zero mean and maximum value of $L$, i.e. $\delta \theta \cdot \sin \theta \sim \mathcal{U}[-L, L]$, then $V$ is given as (see Appendix B):

$$
V = \begin{bmatrix}
1 & \text{sinc}[\pi \tilde{d}_a L] & \cdots & \cdots & \text{sinc}[(m - 1)\pi \tilde{d}_a L] \\
\text{sinc}[\pi \tilde{d}_a L] & 1 & \ddots & \cdots & \text{sinc}[(m - 2)\pi \tilde{d}_a L] \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
\text{sinc}[(m - 1)\pi \tilde{d}_a L] & \text{sinc}[(m - 2)\pi \tilde{d}_a L] & \cdots & \cdots & 1
\end{bmatrix}
$$

(4.11)

where $\text{sinc}[x]$ represents the sinc function. For zero-mean Gaussian distributed angle perturbations with variance $\sigma_\phi^2$, i.e. $\delta \theta \cdot \sin \theta \sim \mathcal{N}[0, \sigma_\phi]$, $V$ is of the form (see Appendix C)

$$
V = \begin{bmatrix}
1 & e^{-\frac{(\pi \tilde{d}_a \sigma_\phi)^2}{2}} & \cdots & e^{-\frac{((m - 1)\pi \tilde{d}_a \sigma_\phi)^2}{2}} \\
e^{-\frac{(\pi \tilde{d}_a \sigma_\phi)^2}{2}} & 1 & \cdots & e^{-\frac{((m - 2)\pi \tilde{d}_a \sigma_\phi)^2}{2}} \\
\vdots & \vdots & \ddots & \vdots \\
e^{-\frac{((m - 1)\pi \tilde{d}_a \sigma_\phi)^2}{2}} & \cdots & \cdots & 1
\end{bmatrix}
$$

(4.12)

From theorem T4.1, under the assumptions we have made, a compact expression for the perturbed covariance matrix is derived. In fact, these results coincide with that obtained by K. M. Wong [13] under a different approach of formulation. In [13], a signal spatial model is used to investigate the problem of spread source. For single emitting source, Wong has proved that for spread source, the $kl$ element of the covariance matrix (under noiseless condition) is given by

$$
R_{kl} = \frac{1}{\pi} \int_{-\pi}^{\pi} |S(\phi)|^2 e^{j(k-l)\phi} d\phi ,
$$

(4.13)

where $|S(\phi)|^2$ is the average power density of the arriving spread signal and is distributed as a function of the electrical angle $\phi$ in space. By careful inspection, (4.13) is closely related to (4.9) and (4.10) that the former one combined both the latter two equations. For example, in [13], Wong has shown that if the signal spread is Gaussian distributed as
\[ |S(\phi)|^2 = \frac{\sigma_s^2}{\sqrt{2\pi \sigma_\phi^2}} e^{-\frac{(\phi - \bar{\phi})^2}{2\sigma_\phi^2}}, \quad (4.14) \]

then the covariance matrix is given as

\[
R = \frac{1}{2\pi^2 \sigma_s^2} \begin{bmatrix}
1 & e^{-\frac{\sigma_s^2}{2} \delta \delta_1} & \cdots & e^{-\frac{\sigma_s^2}{2} \delta \delta_{m-1}} \\
e^{-\frac{\sigma_s^2}{2} \delta \delta_1} & 1 & \cdots & e^{-\frac{\sigma_s^2}{2} \delta \delta_{m-2}} \\
\vdots & \vdots & \ddots & \vdots \\
e^{-\frac{\sigma_s^2}{2} \delta \delta_{m-1}} & e^{-\frac{\sigma_s^2}{2} \delta \delta_{m-2}} & \cdots & 1
\end{bmatrix} \quad (4.15)
\]

By substituting (4.10) into (4.9) and letting \( d = 1, \theta_1 = \phi, \sigma = 0 \), it has exactly the same form as (4.15). This coincidence of results from two different approaches and formulations of analysis indicates that our assumptions and error model stated previously are reasonable and realistic. Before further analyzing the effect of the random DOA dependent errors on the eigen-structure based algorithms, the properties of the perturbation matrix \( V \) is revealed below.

**Properties of the perturbation matrix**

The properties of \( V \) can be deduced from (4.11) as follows:

**P1.** When \( k = l \), \( V(k, k) = \mathbb{E}\{1\} = 1 \), this means that all the diagonal elements of \( V \) are 1's irrespective of the probability distributions of \( \delta \theta \cdot \sin \theta \).

**P2.** For most of the zero-mean probability distributions (such as uniform, Gaussian etc.), the probability density function (p.d.f.) is symmetrical and we can write

\[
V(k, l) = \mathbb{E}\{\exp(j\pi \tilde{d}_a(k-l)\delta \theta \cdot \sin \theta)\} = \mathbb{E}\{\exp(j\pi \tilde{d}_a(l-k)\delta \theta \cdot \sin \theta)\} = V(l, k).
\]

This means that \( V \) is a symmetric matrix.

**P3.** Since \( V \) is given by the equation *(see Appendix A)*

\[
V = \mathbb{E}\{\delta a(\theta_k, \delta \theta_k(l)) \delta a^*(\theta_k, \delta \theta_k(l))\}, \quad (4.16)
\]
where \( \delta a(\theta_k, \delta \theta_k(l)) \) is a random vector containing \( m \) random variables (4.8), by simple statistical theories, \( V \) is positive definite [32].

\textbf{P4.} Since the \( kl \)th element of \( V \) depends on the difference of \( k \) and \( l \) only, it is obvious that \( V \) is a Toeplitz matrix \(^1\).

All the above properties are of crucial importance in the coming analysis. For example, the compensating algorithm that will be described shortly depend on the first three properties \textbf{P1.} - \textbf{P3.}. Also, \textbf{P4.} will be used to alleviate the problem of signal coherence in Chapter 6.

\subsection{Effect of the perturbation matrix}

In this section, an analysis of the effect of \( V \) on the estimated DOAs will be presented. The discussion is restricted to MUSIC with ULA because the case with ESPRIT is similar. In Chapter 1, we have defined \(( \cdot )\) to be the estimate of \(( \cdot )\) and this notation will be used hereafter. By following the approach adopted by Stoica and Nehorai in [29] and Swindlehurst in [31], an expression for the error of estimation \(( \hat{\theta}_k - \theta_k )\) is derived. Based on this expression, it is shown that if ordinary root-MUSIC is applied to the perturbed covariance matrix \( R_{\hat{x} \hat{x}} \), the DOAs estimated will be biased. Recall that in root-MUSIC, to obtain the DOAs under \textit{ideal situation}, we have to locate the infinite points of the objective function \( P_M(\theta) \) (3.10). It is equivalent to finding the zeros of the function \( p(\theta) \) defined as

\[
p(\theta) = \frac{1}{P_M(\theta)} = \frac{\alpha^*(\theta)E_N E_N^* \alpha(\theta)}{\alpha^*(\theta) \alpha(\theta)} = \text{tr}(P_{\theta} E_N E_N^*). \tag{4.17}
\]

where \( P_{\theta} \) is defined as the projection matrix with \( P_{\theta} = \alpha(\theta)[\alpha^*(\theta) \alpha(\theta)]^{-1} \alpha^*(\theta) \). Due to the presence of the perturbation matrix, the exact \( p(\theta) \) is not available and an approximated one \( \hat{p}(\theta) \) is used.

\(^1\)A matrix \( C \in \mathbb{C}^{m \times m} \) is called Toeplitz if it is of the form

\[
C = \begin{bmatrix}
c_0 & c_{-1} & \cdots & \cdots & c_{-m+2} & c_{-m+1} \\
c_1 & c_0 & \cdots & \cdots & c_{-m+3} & c_{-m+2} \\
\vdots & \vdots & \ddots & \cdots & \vdots & \vdots \\
\vdots & \vdots & \cdots & \ddots & \vdots & \vdots \\
c_{m-2} & c_{m-3} & \cdots & \cdots & c_0 & c_{-1} \\
c_{m-1} & c_{m-2} & \cdots & \cdots & c_1 & c_0
\end{bmatrix}.
\]
\[
\hat{p}(\theta) = \text{tr}(P_{\delta} \hat{E}_N \hat{E}_N^*).
\]

By expanding \( \hat{p}' \) about the estimate \( \hat{\theta}_k \) and rearrange terms, we get

\[
\hat{\theta}_k - \theta_k \approx \frac{\hat{p}'(\theta_k)}{\hat{p}''(\theta_k)}.
\]  

(4.18)

To complete (4.18), we need the expressions of \( \hat{p}'(\theta_k) \) and \( \hat{p}''(\theta_k) \) which have been derived in [31]. By substituting these expressions into (4.18), the first-order estimation error for the DOA is given as

\[
e_k = \hat{\theta}_k - \theta_k \approx \frac{\text{Re}\{d^*(\theta_k) \Sigma_N^{-1/2} E_N E_N^* \Sigma_N^{-1/2} \xi_k\}}{d^*(\theta_k) \Sigma_N^{-1/2} E_N E_N^* \Sigma_N^{-1/2} d(\theta_k)},
\]  

(4.19)

with \( E_N, \Sigma_N^{-1/2} \) defined as in Chapter 3, \( d(\theta_k) \) is the change of the steering vector \( a(\theta_k) \) with respect to the angle \( \theta_k \) as

\[
d(\theta_k) = \frac{\partial a(\theta_k)}{\partial \theta_k},
\]

and \( \xi_k \) is the error vector. In [31], the expression for \( \xi_k \) is very complicated due to the presence of errors in the noise covariance matrix. However, in our case, \( \xi_k \) is rather simple and is given by (see Appendix D)

\[
\xi_k = (a(\theta_k) \odot \delta a(\theta_k, \delta \theta_k)) - a(\theta_k).
\]  

(4.20)

From the approximated estimation error expression in (4.19), we can calculate the means of the estimation error as

\[
e_k = \mathbb{E}\{e_k\}
\]

\[
= \mathbb{E}\left\{ \text{Re}\left\{d^*(\theta_k) \Sigma_N^{-1/2} E_N E_N^* \Sigma_N^{-1/2} \xi_k\right\}\right\} \\
= \frac{\text{Re}\left\{d^*(\theta_k) \Sigma_N^{-1/2} E_N E_N^* \Sigma_N^{-1/2}\right\}}{d^*(\theta_k) \Sigma_N^{-1/2} E_N E_N^* \Sigma_N^{-1/2} d(\theta_k)} \mathbb{E}\{\xi_k\}
\]

\[
= K \cdot (\mathbb{E}\{\delta a(\theta_k, \delta \theta_k)\} - t).
\]  

(4.21)
where $t$ is a vector with all its elements being 1's.

It is easy to see that $E\{\delta a(\theta_k, \delta \theta_k)\}$ will not be a vector with all its elements being 1's. For example, if the angles are uniformly distributed ($\sim \mathcal{U}[-L, L]$) or Gaussian distributed ($\sim \mathcal{N}[0, \sigma_\theta]$), then $\tilde{e}_k$ is given by (4.22) and (4.23) (the derivations can be found in Appendix A and Appendix B) respectively

$$\tilde{e}_k = K \cdot \begin{bmatrix} 0 & \text{sinc}[\pi \tilde{d}_a L] - 1 & \cdots & \text{sinc}[\pi \tilde{d}_a(m - 1)L] - 1 \end{bmatrix}^T, \quad (4.22)$$

$$\tilde{e}_k = K \cdot \begin{bmatrix} 0 & e^{-\frac{(\pi \tilde{d}_a \sigma_\theta)^2}{2}} - 1 & \cdots & e^{-\frac{(\pi \tilde{d}_a(m - 1) \sigma_\theta)^2}{2}} - 1 \end{bmatrix}^T. \quad (4.23)$$

In general, unless $L = 0$ and $\sigma_\theta = 0$, $\tilde{e}_k$ will not be zero vector. This means that if ordinary root MUSIC is applied to $R_{\mathbf{x}\mathbf{x}}$ directly without modification, the DOAs obtained will be biased. We will show this phenomenon graphically when simulation results are presented. Since the estimates are biased, the variances of the estimates errors will not be considered.

### 4.2.5 Compensating algorithm

Though Kailath and Swindlehurst has presented a very detailed and thorough analysis about different kinds of errors concerning the MUSIC algorithm in [31] and [30], they have not proposed methods to compensate the errors. The same situation apply to the other researchers which put all their efforts in analysis ([17], [18], [3]). As revealed by the previous section, the DOA estimates are biased if we apply MUSIC to the perturbed covariance matrix without any modification. In order to get unbiased estimates, the ordinary MUSIC have to be modified. As deduced from (4.9), the angle perturbations modify the covariance matrix by Schur-Hadamard multiplication. By utilizing this property, the following simple and robust method is proposed to compensate the errors.

To generalize the procedures, the noise is assumed to be colored. Define $\otimes$ to be a binary operator between two matrices such that if $C = A \otimes B$, then

$$C(k, l) = \frac{A(k, l)}{B(k, l)}, \quad (4.24)$$

44
i.e. $\odot$ is the point-by-point division between two matrices and may be considered in some sense as the inverse operator of the Schur-Hadamard product operation. Performing an $\odot$ operation on both sides of (4.9) by $V$ gives the following equation

\begin{equation}
R_{\hat{x}\hat{x}} \odot V = ASA^* + \sigma^2 (\Sigma_N \odot V).
\end{equation}

Note that the first term on the RHS is due to the identity $(A \odot B) \odot B = A$. (4.25) is exactly of the same form as (4.9) but with $\Sigma_N$ replaced by $(\Sigma_N \odot V)$. Hence we can apply the techniques of Mahalanobius transform as described in Chapter 3 and the resulting equation is given as

\begin{align*}
Z_{\hat{x}\hat{x}} &= (\Sigma_N \odot V)^{-1/2}(R_{\hat{x}\hat{x}} \odot V)(\Sigma_N \odot V)^{-*1/2} \\
&= (\Sigma_N \odot V)^{-1/2}ASA^*(\Sigma_N \odot V)^{-*1/2} + \sigma^2 I.
\end{align*}

(4.26)

After these procedures, a simple eigen-decomposition on the resulting matrix $Z_{\hat{x}\hat{x}}$ will then give the corrected signal eigenvectors $E_S$ and noise eigenvectors $E_N$ as the case without random perturbations, i.e.

\begin{equation}
Z_{\hat{x}\hat{x}} = \left[(\Sigma_N \odot V)^{-1/2}E_S \right] A_S \left[ (\Sigma_N \odot V)^{-1/2}E_S \right]^* + E_N A_N E_N^*,
\end{equation}

(4.27)

and the DOAs can be estimated by searching the peak points of the inverse of the projection function defined in (3.10).

Although the algorithm seems straightforward, there are two very important points to be noted and they are described below.

**Situation unresolvable by using Mahalanobius transformation**

In the above derivation, both the matrix $\Sigma_N$ and $V$ are positive definite. However, unlike the Schur-Hadamard operation by which positive definiteness is preserved [9], the $\odot$ operation does not preserve positive definiteness and therefore $(\Sigma_N \odot V)$ may not be positive definite. For example, consider the two-dimensional case with

\[
\Sigma_N = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}, \quad V = \begin{bmatrix} 1.0 & 0.1 \\ 0.1 & 1.0 \end{bmatrix},
\]

45
which are both positive definite. But after the \( \odot \) operation, we will get

\[
\Sigma_N \odot V = \begin{bmatrix} 3 & 10 \\ 10 & 3 \end{bmatrix},
\]

which is obviously not positive definite. By observing this, we have the following theorem:

**Theorem T4.2**: If the matrix \( \Sigma \odot V \) is not positive definite, then DOAs of the emitting sources cannot be estimated by using Mahalanobius transformation followed by eigen-based methods.

**Proof**: It is obvious from the above arguments that when \( \Sigma \odot V \) is not positive definite, neither generalized eigen-decomposition nor Mahalanobius transformation can be performed (because Mahalanobius transformation requires the use of Cholesky decomposition which is restricted to positive definite matrix). Therefore we cannot separate the signal subspace from the noise space and hence fail to estimate the DOAs by simple eigen-decomposition.\( \blacksquare \)

Theorem **T4.2** is important as it imposes a bound on the random perturbations and noises for the problem to be solvable by Mahalanobius transformation. By combining (4.11) (4.12) with the arguments of the theorem and studying the effect of the random deviations on the perturbation matrix \( V \) carefully, it can be seen that when the random perturbations are very large, or when the noise covariance matrix is nearly singular, the situation described in **T4.2** will happen. Another point to be noted is that this situation only applies to cases with colored noise. For white noise, this problem does not exist because identity matrix is always obtained after the \( \odot \) operation (due to the property **P1** stated in the previous section) and hence no Mahalanobius transformation is required. This shows a great difference between colored noise and white noise in our error model and compensating algorithm.
Problem of noise amplification

Another precaution which have to be taken care of the modifying procedures concerns with the finite sample effect. Though we have stated that this effect can be ignored in the ideal situation (i.e. the random DOA perturbations are zero), it cause problem when the perturbation matrix is "behaving bad". By this we mean that $V$ is not well conditioned for our compensating algorithm. Since $V$ depends on the statistics of the random angle deviations, the elements of $V$ may be zero or nearly zero. For example, consider the ULA problem with uniformly distributed perturbations. Suppose there are totally twelve sensors and $\tilde{d}_{a} = 1$. By substituting these parameters into (4.11), the $kth$ element of $V$ is given as

$$V(k, l) = \text{sinc} \left[ (k - l)\pi L \right].$$

From this formula, $V(12, 1)$ is negative if $L \geq \frac{1}{11}$ and the elements $V(11, 1)$, $V(10, 1)$ etc. is nearly zero. These nearly zero elements will cause undesirable noise amplification during the point-by-point ($\odot$) operation and the effect is great if the number of samples used to estimate the covariance matrix (3.6) is not large enough. Even worse is that if so happens one of the elements of the perturbation matrix is zero, the $\odot$ operation will cause severe overflow problem. This phenomenon can be explained by the fact that Schur product operation cannot be completely reversible by $\odot$ operation if $V$ contains zero elements as some information is lost during the multiplication process by zeros.

To alleviate the above problem, three methods are proposed. The first method exploit functional dependence of the perturbation matrix on the system parameters. From (4.11) and (4.12), it can be observed that the elements of $V$ are not only dependent on the perturbation parameters ($L$ or $\sigma_{\theta}$) but also on the number of sensors used $m$ and the array configuration factor $\tilde{d}_{a}$. Thus by limiting $m$ or by decreasing the number $\tilde{d}_{a}$, the zero elements problem can be avoided. This method is the simplest and most straightforward one though it put more constraints on the array configuration design.

The second method is based on the fact that some elements of $ASA^{*}$ was set to zero by the zero (or nearly zero) elements of the perturbation matrix. Since the original information is lost, we may assign arbitrary values to these elements instead.
of performing division by zero. A threshold can be set so that if \( V(k,l) \) is smaller than the threshold, assignment is performed instead of division. In the previous example, suppose \( L = 0.1 \) and we set the threshold to be 0.02. By simple calculation, we found that \( V(12,2) = 1.2246 \times 10^{-16} \) which is much smaller than the threshold. So instead of dividing \( R_{xx}(12,2) \) by this small number, a value is assigned to it. Although arbitrary values can be used in the assignment, reasonable ones should be used. One of the suggestions is the neighboring values. Thus we can set \( (R_{xx} \odot V)(12,2) \) to its neighbor value such as \( (R_{xx} \odot V)(11,2) \). This means that the lost information is estimated by the neighboring values. The drawbacks of the method is that this method is non-linear and the positive definiteness of the resulting \( R_{xx} \odot V \) will be lost. This leads to negative eigenvalues when performing eigen-decomposition and complicates the whole algorithm.

The third method uses the principle of regularization. Instead of performing \( \odot \) operation by \( V \), a modified one \( \tilde{V} \) is used where

\[
\tilde{V} = V + \eta T,
\]

where \( \eta \) being a regularization factor lies between zero and one. \( T \) is the regularization matrix with all the elements being 1's. The purpose of adding this regularization matrix is to remove the problem of division by zero. By adjusting the value of \( \eta \), the noise-amplification effect can be reduced to certain extent. However, note that with \( T \), \( \tilde{V} \) will introduce bias to both the noise and signal subspaces as \( T \) destroys the correct proportions of the elements of \( V \). In fact, asymptotically when \( \eta \) is very small, \( \tilde{V} \) will reduce back to \( V \) and no regularization is performed. When \( \eta \) is very large, \( T \) dominates and the modifying procedures degenerate to standard MUSIC. To compare the three methods, simulations are carried out and the results will be presented in the latter part of this chapter.
4.2.6 Estimating unknown perturbation parameter

In the above compensating algorithm, we have assumed a prior knowledge about the whole perturbing matrix. In most practical situations, it is not the case. For example, if the angle deviations are caused by the presence of tiny particles and the dispersive nature of the medium, we may only be able to assume the statistical model of the perturbations but not the model parameters (such as the variance in case of Gaussian distribution) as the maximum degree of deviations is unpredictable. Since it is clear from the last section that the entire compensating algorithm hinges on the perturbation matrix, an accurate estimation of its parameters is required. The major concern of this chapter is to devise a method to estimate the perturbation parameters. Since we have restricted the perturbations to have zero means, the estimation space is reduced to one-dimension (i.e. single parameter is estimated) for uniform or Gaussian distribution. In the sequel, an innovative algorithm exploiting the special properties of \( V \) (P1. - P4.) is proposed. We will show that under ideal situation, this method will give exact estimate of the desired parameter. To have smooth and precise presentation, Gaussian distributed random deviations is assumed with unknown parameter \( \sigma_\theta \). For other kind of statistical distribution (e.g. uniform), the principle of the algorithm is similar and is not discussed in detail. Moreover, \( \Sigma_N \) is set to the identity matrix \( I \) (white noise) for the sake of simplicity. To start with, two theorems are first stated and proved. The rationale of the algorithm which is based on the theorems is described then.

**Theorem T4.3**: Let \( P(\hat{\sigma}_\theta, \sigma_\theta) \) be a \( m \times m \) p.d. \(^2\) matrix of the form

\[
P(\hat{\sigma}_\theta, \sigma_\theta) = \begin{bmatrix}
1 & & & e^{-(m-1)k_\theta (\sigma_\theta^2 - \hat{\sigma}_\theta^2)} \\
& \ddots & & \vdots \\
e^{-k_\theta (\sigma_\theta^2 - \hat{\sigma}_\theta^2)} & \ddots & 1 & e^{-(m-2)k_\theta (\sigma_\theta^2 - \hat{\sigma}_\theta^2)} \\
\vdots & \ddots & \ddots & \ddots \\
e^{-(m-1)k_\theta (\sigma_\theta^2 - \hat{\sigma}_\theta^2)} & \ldots & \ldots & 1
\end{bmatrix}.
\]

(4.29)

where \( k_\theta \) is a constant.

\(^2\)Throughout the whole thesis, p.d., p.s.d., i.d. and n.d. stand for positive definite, positive semi-definite, indefinite and negative definite respectively.
Suppose $R \in \mathcal{G}^m$ is a fixed p.s.d. matrix which can be written as

$$R = ASA^*,$$  

(4.30)

where $A \in \mathcal{G}^{m \times d}$ is a steering matrix as in (4.2) and $S \in \mathcal{G}^{d \times d}$ is a p.d. matrix. The rank of $R$ is given by $\rho(R) = d$. Assume $2 \leq d \leq m - 3$. If

$$R \odot P(\hat{\sigma}_\theta, \sigma_\theta) = Q(\hat{\sigma}_\theta, \sigma_\theta) + \mu I,$$  

(4.31)

where $\mu$ is real positive and is strictly less than the major diagonal elements of $R$, then

$$\rho(Q(\hat{\sigma}_\theta, \sigma_\theta)) \geq d + 1,$$  

(4.32)

with probability 1, $\forall (\sigma_\theta - \hat{\sigma}_\theta) > 0$.

**Proof:** see Appendix E.  

---

**Theorem T4.4:** Let $P(\hat{\sigma}_\theta, \sigma_\theta) \in \mathcal{G}^{m \times m}$ be a matrix of the form as (4.29). Define $P^\odot(\hat{\sigma}_\theta, \sigma_\theta)$ as follows:

$$P^\odot(\hat{\sigma}_\theta, \sigma_\theta) = T \odot P(\hat{\sigma}_\theta, \sigma_\theta),$$  

(4.33)

where $T$ is defined in the previous section, represents the matrix with all its elements being 1's.

If $P(\hat{\sigma}_\theta, \sigma_\theta)$ is indefinite (i.d.), then $P^\odot(\hat{\sigma}_\theta, \sigma_\theta)$ is positive definite and vice versa.

**Proof:** see Appendix F.
As will be seen shortly, theorem T4.3 is of key importance to the estimation algorithm which is described below.

Suppose we have an estimate \( \hat{\sigma}_\theta \) of the standard deviation \( \sigma_\theta \). Using this estimate, the estimated perturbation matrix \( \hat{V}(\hat{\sigma}_\theta) \) is constructed according to (4.12). By setting \( \Sigma_N \) to \( I \) in (4.9), the covariance matrix of the perturbed data is given as

\[
R_{\check{x}\check{x}} = (ASA^*) \odot V(\sigma_\theta) + \sigma^2 I,
\]

(4.34)

Consider performing \( \odot \) operation on both sides of (4.34), which can then be rewritten as

\[
W(\hat{\sigma}_\theta, \sigma_\theta) = (ASA^*) \odot P(\hat{\sigma}_\theta, \sigma_\theta) + \sigma^2 I,
\]

(4.35)

where \( P(\hat{\sigma}_\theta, \sigma_\theta) = V(\sigma_\theta) \odot \hat{V}(\hat{\sigma}_\theta) \) and \( W(\hat{\sigma}_\theta, \sigma_\theta) = R_{\check{x}\check{x}} \odot \hat{V}(\hat{\sigma}_\theta) \). Note that according to its definition, it is easy to show that \( P(\hat{\sigma}_\theta, \sigma_\theta) \) can be written as the same form of (4.29).

Now an eigen-decomposition is performed on the resulting matrix \( W(\hat{\sigma}_\theta, \sigma_\theta) \) and let \( \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_m \) be the corresponding eigenvalues arranged in ascending order. Since \( W(\hat{\sigma}_\theta, \sigma_\theta) \) is Hermitian, all the \( \lambda_i \)'s are real (although not necessary positive). Denote the standard deviation of the smallest \( m-d \) eigenvalues of \( W(\hat{\sigma}_\theta, \sigma_\theta) \) by \( \sigma_\lambda \), i.e.

\[
\sigma_\lambda = \sqrt{\frac{1}{m-d} \sum_{i=1}^{m-d} (\lambda_i - \mu_\lambda)^2},
\]

(4.36)

where \( \mu_\lambda = \sum_{i=1}^{m-d} \lambda_i \) is the mean of the \( m-d \) smallest eigenvalues. Now from (4.35), it can be observed that as \( \sigma_\theta, A, \sigma_\lambda \) are constant, the eigenvalues of \( W(\hat{\sigma}_\theta, \sigma_\theta) \) depend solely on the values of \( \hat{\sigma}_\theta \). This implies that as \( \hat{\sigma}_\theta \) varies, \( \sigma_\lambda \) changes accordingly which may be classified into four cases according to the positive definiteness of both \( P(\hat{\sigma}_\theta, \sigma_\theta) \) and \( W(\hat{\sigma}_\theta, \sigma_\theta) \). The four cases are listed below and the behaviour of \( \sigma_\lambda \) is described in each case.
Case 1: $\hat{\sigma}_\theta = \sigma_\theta$.

Under this situation, $\hat{V}(\hat{\sigma}_\theta) = V(\sigma_\theta)$. Since $P(\hat{\sigma}_\theta, \sigma_\theta) = \hat{V}(\hat{\sigma}_\theta) \otimes V(\sigma_\theta)$, all the elements of $P(\sigma_\theta, \sigma_\theta)$ will equal to one by which the case is reduced back to zero random perturbations and $W(\sigma_\theta, \sigma_\theta) = ASA^* + \sigma^2 I$. Hence by the arguments in Chapter 3, we have $\lambda_1 = \lambda_2 = \cdots = \lambda_{m-d} = \sigma^2$ and thus $\sigma_\lambda = 0$.

Case 2: $\hat{\sigma}_\theta \neq \sigma_\theta$ and $P(\hat{\sigma}_\theta, \sigma_\theta)$ is p.d.

In this case, $\lambda_1 > \sigma^2$. Since $\lambda_1$ is the smallest eigenvalues of $W(\hat{\sigma}_\theta, \sigma_\theta)$, we may express $W(\hat{\sigma}_\theta, \sigma_\theta)$ in the form

$$W(\hat{\sigma}_\theta, \sigma_\theta) = W'(\hat{\sigma}_\theta, \sigma_\theta) + \lambda_1 I,$$  \hspace{1cm} (4.37)

with $W'(\hat{\sigma}_\theta, \sigma_\theta)$ being p.s.d. Substitute (4.37) into (4.35) and rearrange terms, it becomes

$$W'(\hat{\sigma}_\theta, \sigma_\theta) + (\lambda_1 - \sigma^2) I = (ASA^*) \otimes P(\hat{\sigma}_\theta, \sigma_\theta).$$  \hspace{1cm} (4.38)

Note that this equation has the same form as (4.31) with $R = ASA^*$, $Q(\hat{\sigma}_\theta, \sigma_\theta) = W'(\hat{\sigma}_\theta, \sigma_\theta)$ and $\mu = \lambda_1 - \sigma^2$. Hence $\rho(W'(\hat{\sigma}_\theta, \sigma_\theta)) \geq d + 1$ (by theorem T4.3). This means that the $(m - d)$th eigenvalue (which are also arranged in ascending order) of $W'(\hat{\sigma}_\theta, \sigma_\theta)$ must not be zero (and therefore larger than zero as $W'(\hat{\sigma}_\theta, \sigma_\theta)$ is p.s.d.) which implies that $\lambda_{m-d}$ must be larger than $\lambda_1$, i.e. $\lambda_1 \neq \lambda_{m-d}$ and thus $\sigma_\lambda > 0$.

Case 3: $\hat{\sigma}_\theta \neq \sigma_\theta$, $P(\hat{\sigma}_\theta, \sigma_\theta)$ is i.d. but $W(\hat{\sigma}_\theta, \sigma_\theta)$ is p.s.d.

In this case, $0 \leq \lambda_1 \leq \sigma^2$. We will prove by contradiction that $\sigma_\lambda > 0$. Now suppose $\sigma_\lambda = 0$, this implies that $\lambda_1 = \lambda_2 = \cdots = \lambda_{m-d}$ and $W(\hat{\sigma}_\theta, \sigma_\theta)$ can then be decomposed as

$$W(\hat{\sigma}_\theta, \sigma_\theta) = W''(\hat{\sigma}_\theta, \sigma_\theta) + \lambda_1 I,$$  \hspace{1cm} (4.39)

where $W''(\hat{\sigma}_\theta, \sigma_\theta)$ is a p.s.d. matrix and is of rank $d$. By substituting (4.39) into (4.35) and performing an $\otimes$ operation by $P(\hat{\sigma}_\theta, \sigma_\theta)$ on both sides, the resulting equation is given as

52
\[ \textbf{ASA}^* + \sigma^2 I = \left( W''(\hat{\sigma}_\theta, \sigma_\theta) + \lambda_1 I \right) \odot P(\hat{\sigma}_\theta, \sigma_\theta) \]
\[ = \left( W''(\hat{\sigma}_\theta, \sigma_\theta) + \lambda_1 I \right) \odot P^\odot(\hat{\sigma}_\theta, \sigma_\theta). \quad (4.40) \]

The second equality comes from (4.33). After rearranging terms we get

\[ W''(\hat{\sigma}_\theta, \sigma_\theta) \odot P^\odot(\hat{\sigma}_\theta, \sigma_\theta) = \textbf{ASA}^* + (\sigma^2 - \lambda_1) I. \quad (4.41) \]

By theorem T4.4, as \( P(\hat{\sigma}_\theta, \sigma_\theta) \) is indefinite, \( P^\odot(\hat{\sigma}_\theta, \sigma_\theta) \) is positive definite. Thus it can be seen that (4.41) has the same structure as (4.31) with \( R \) replaced by \( W''(\hat{\sigma}_\theta, \sigma_\theta) \), \( Q(\hat{\sigma}_\theta, \sigma_\theta) \) replaced by \( \textbf{ASA}^* \), \( \mu \) replaced by \( \sigma^2 - \lambda_1 \) which is larger than zero and \( P(\hat{\sigma}_\theta, \sigma_\theta) \) replaced by \( P^\odot(\hat{\sigma}_\theta, \sigma_\theta) \). Note that the matrices and constants satisfy the conditions of theorem T4.3, and consequently \( \rho(\textbf{ASA}^*) \geq d + 1 \). However, this leads to contradiction that \( \rho(\textbf{ASA}^*) = d \). Hence our assumption \( \sigma_\lambda = 0 \) must be wrong. Since \( \sigma_\lambda \) is non-negative, therefore \( \sigma_\lambda > 0 \).

Case 4: \( \hat{\sigma}_\theta \neq \sigma_\theta \) and \( W(\hat{\sigma}_\theta) \) is i.d.

For this situation, \( \lambda_1 < 0 \) and by checking the sign of \( \lambda_1 \), this case can be ignored.

From the above arguments, \( \sigma_\lambda \) will be zero if and only if \( \hat{\sigma}_\theta = \sigma_\theta \). Inspiring by this special property, procedures for estimating the unknown parameters are proposed as follows:

1. For a fixed \( \hat{\sigma}_\theta \), construct the corresponding perturbation matrix \( \hat{V}(\hat{\sigma}_\theta) \) according to (4.34).
2. Perform the \( \odot \) operation and obtain the matrix \( W(\hat{\sigma}_\theta, \sigma_\theta) = R_{\hat{x}\hat{x}} \odot \hat{V}(\hat{\sigma}_\theta) \).
3. Perform eigen-decomposition on \( W(\hat{\sigma}_\theta, \sigma_\theta) \) and obtain its \( (m - d) \) smallest eigenvalues.
4. Calculate the standard deviation \( \sigma_\lambda \) of the \( (m - d) \) smallest eigenvalues obtained in 3. by using the formula in (4.36).
By increasing $\hat{\sigma}_\theta$ stepwisely from zero to $\hat{\sigma}_{\theta,\text{max}}$ (which is the maximum $\hat{\sigma}_\theta$ for which $\lambda_1 > 0$), corresponding values of $\sigma_\lambda$ can be obtained. The value of $\hat{\sigma}_\theta$ at which $\sigma_\lambda$ is minimum will then be the best estimate of the actual perturbation parameter $\sigma_\theta$.

A few words must be said about the estimation algorithm. It is clear that the algorithm relies heavily on theorem T4.3. Under ideal situation, (i.e. we have the exact covariance matrix $R_{\hat{x}\hat{x}}$ rather than the estimated one $\hat{R}_{\hat{x}\hat{x}}$ by (3.16)), the algorithm will give exact estimate of $\sigma_\theta$. However, in practical situation, errors and hence inaccurate estimates arise due to the imperfections of $\hat{R}_{\hat{x}\hat{x}}$. To increase the accuracy of the estimates under such situation, the condition restricting the difference between $d$ and $m$ (i.e. $d \leq m - 1$) listed in T4.3 plays its role. In fact, the mentioned condition is of crucial importance as it expresses the minimal difference between $d$ and $m$ for the algorithm to be applicable. It can be shown that in non-ideal case, the larger the value of $(m - d)$, the more accurate the estimated parameters. Since $d$ is fixed and cannot be altered, in order to get better results, more sensors should be used. However, increasing the number of sensors not only increase system requirements but also increase the time needed to find the estimates as eigen-decomposition of matrix with higher dimension has to be done. Thus it is indeed a good example of compromise between computation time and accuracy. This dilemma will be shown in our simulation results in Chapter 5.
4.2.7 ESPRIT formulation

For ESPRIT, the perturbation matrix is of different form from that of MUSIC as the sensor array used in ESPRIT contains sensor doublets. Due to this reason, the effects of deviations in DOAs are reflected by the difference in phase response within the doublets. Consider the situation as shown in Figure 4.3 where we have partitioned the array into two subarrays, namely the odd array and the even array. Suppose the nominal responses of the odd sensors with respect to an emitting source at an angle $\theta_k$ are given as $\rho_1(\theta_k), \rho_2(\theta_k), \ldots, \rho_m(\theta_k)$. Then by following the procedures as in section 3.3.2 of Chapter 3, in the absence of random perturbations, the response vector can be written as

$$x(l) = A_{\rho}(\theta, \Delta)s(l) + n(l), \quad (4.42)$$

where
\[ \theta = [\theta_1 \quad \theta_2 \quad \ldots \ldots \quad \theta_d]^T, \]  

(4.43)

\[ A \rho(\theta, \Delta) = [\rho(\theta_1, \Delta) \quad \rho(\theta_2, \Delta) \quad \ldots \ldots \quad \rho(\theta_d, \Delta)], \]  

(4.44)

\[ \rho(\theta_k, \Delta) = \left[ \rho^T_\circ(\theta_k, \Delta) \quad \rho^T_\circ(\theta_k, \Delta)e^{-j \frac{\pi a_x}{\lambda} \Delta \cos(\theta_k)} \right]^T, \]  

(4.45)

\[ \rho_\circ(\theta_k, \Delta) = [\rho_1(\theta_k) \quad \rho_2(\theta_k) \quad \ldots \ldots \quad \rho_m(\theta_k)]^T. \]  

(4.46)

In practical situations, (4.42) is not correct due to the presence of deviations \( \delta \theta_k(l) \) in the DOA \( \theta_k \). To study the effect of random perturbations on (4.42), the sensor doublets are assumed to be independent of each other. Moreover, it is assumed that \( \Delta \) is much larger than \( d_\alpha \) so that the following approximations are valid

\[ \rho_i(\theta_k + \delta \theta_k(l)) \approx \rho_i(\theta_k) \quad i = 1, 2, \ldots \ldots, m, \quad k = 1, 2, \ldots \ldots, d, \quad l = 1, 2, \ldots \ldots, N. \]  

(4.47)

This assumption is equivalent to saying that the random perturbations \( \delta \theta(l) \) only affect the phase shifts caused by the invariant displacement \( (\Delta) \) within the sensor doublets but they have no effect on the nominal responses of the sensors. Without this assumption, the situation will be very complicated and hence is not considered here. Moreover, similar to the approximation made in (4.5), we assume

\[ \cos(\theta_k + \delta \theta_k(l)) \approx \cos \theta_k - \delta \theta_k(l) \cdot \sin \theta_k. \]  

(4.48)

By using (4.48) and (4.47), the perturbed response vector \( \bar{\rho}(\theta_k, \Delta, \delta \theta_k(l)) \) is given by

\[ \bar{\rho}(\theta_k, \Delta, \delta \theta_k(l)) = \left[ \rho^T_\circ(\theta_k, \Delta) \quad \rho^T_\circ(\theta_k, \Delta)e^{-j \frac{\pi a_x}{\lambda} \Delta \cos(\theta_k) - \delta \theta_k(l) \sin \theta_k} \right]. \]  

(4.49)

Hence (4.42) can be rewritten as

\[ x(l) = \bar{A} \rho(\theta, \Delta, \delta \theta)(l)s(l) + n(l), \]  

(4.50)

with
\[ \tilde{A}_\rho(\theta, \Delta, \delta\theta)(l) = A_\rho(\theta, \Delta) \odot \delta A_\rho(\theta, \Delta, \delta\theta)(l), \] (4.51)

\[ \delta A_\rho(\theta, \Delta, \delta\theta)(l) = [\delta \rho(\theta_1, \Delta, \delta\theta_1(l)) \ \delta \rho(\theta_2, \Delta, \delta\theta_2(l)) \ \cdots \ \delta \rho(\theta_d, \Delta, \delta\theta_d(l))] , \] (4.52)

\[ \delta \rho(\theta_k, \Delta, \delta\theta_k(l)) = [1 \ \cdots \ 1 \ e^{j\pi \Delta \delta\theta_k(l) \cdot \sin \theta_k} \ \cdots \ e^{j\pi \Delta \delta\theta_k(l) \cdot \sin \theta_k}]^T. \] (4.53)

Hence the perturbed covariance matrix is given by

\[ R_{xx} = (A_\rho(\theta, \Delta)SA_\rho(\theta, \Delta)) \odot V + \sigma^2 I , \] (4.54)

where the perturbation matrix \( V \) with uniform distributed perturbations is given as

\[
V = E \begin{bmatrix}
1 \\
\vdots \\
1 \\
e^{j\pi \Delta \delta\theta_k(l) \cdot \sin \theta_k} \\
\vdots \\
e^{j\pi \Delta \delta\theta_k(l) \cdot \sin \theta_k}
\end{bmatrix}
[1 \ \cdots \ 1 \ e^{-j\pi \Delta \delta\theta_k(l) \cdot \sin \theta_k} \ \cdots \ e^{-j\pi \Delta \delta\theta_k(l) \cdot \sin \theta_k}]
\]

\[
= \begin{bmatrix}
T \\
\text{sinc}(\pi \Delta L) \cdot T
\end{bmatrix}.
\] (4.55)

Note that all the elements of \( T \) are 1's as defined in Chapter 1. Although the perturbation matrix for ESPRIT seems much simpler than that of MUSIC, \( V \) is only positive semi-definite and is not Toeplitz though it is still symmetric, the estimated DOAs are still affected by this matrix. This is clear from the arguments below. By (3.18), the perturbed covariance matrix for ESPRIT formulation can be written as
\[ R_{zz} = \left( \begin{bmatrix} E_{S_0} & kE_{S_0}(H^{-1} \Phi H)^*E_{S_0}^* \\ E_{S_0}(H^{-1} \Phi H) & E_{S_0}(H^{-1} \Phi H)^*E_{S_0}^* \end{bmatrix} \Lambda_S \right) \odot V + \sigma^2 I \]

\[ = \left[ \begin{bmatrix} E_{S_0} \Lambda_S E_{S_0}^* & kE_{S_0}(H^{-1} \Phi H)^*E_{S_0}^* \\ kE_{S_0}(H^{-1} \Phi H) \Lambda_S E_{S_0}^* & E_{S_0}(H^{-1} \Phi H) \Lambda_S (H^{-1} \Phi H)^*E_{S_0}^* \end{bmatrix} + \sigma^2 I \right] + \sigma^2 I, \]

\[ = \left[ \begin{bmatrix} \tilde{E}_{S_0} \\ \tilde{E}_{S_e} \end{bmatrix} \right] \tilde{\Lambda}_S \left[ \tilde{E}_{S_0}^* \tilde{E}_{S_e}^* \right] + \sigma^2 I \quad (4.56) \]

where \( k = \text{sinc}(\pi \delta_n L) \) is a constant. From (4.56), due to the presence of \( k \), after performing eigen-decomposition on the perturbed covariance matrix, the subspaces \( \tilde{E}_{S_0} \) and \( \tilde{E}_{S_e} \) will not be related as shown in (3.18) and the principle of TLS cannot be used. In fact, the rank of the resulting covariance matrix will be higher than the number of sources present because \( V \) is p.s.d. and hence the signal subspaces obtained are disturbed non-linearly. To compensate for the random deviations, we can just divide the upper right and the lower left \( (m/2 \times m/2) \) submatrices of the perturbed covariance matrix by the constant \( k \). Note that in this case, the problem of zero elements division does not exist as the constant \( k \) is larger than zero under most cases. Although it is not proven, the algorithm used for estimating the distribution parameter \( (L \) in this case) is also applicable for ESPRIT.

### 4.3 DOA Independent Errors

#### 4.3.1 Sources of errors and mathematical model

The major source of DOA independent errors arises from the random shifts in array response due to drastic changes of weather, unmatched sensors (in the case of ESPRIT), incorrect calibration (in the case of MUSIC), imprecise sensor locations and working out of the system equipment, etc. It is assumed that the errors are independent of the DOAs. In the mathematical formulation, an error model which is similar to
that for DOA dependent errors is introduced. This model is somewhat different from
the additive model proposed by other researchers (e.g. [18], [30]). As usual, the case
of ULA under MUSIC is used as example to illustrate the idea.

In deriving equations for the ULA steering matrix in Chapter 2, it is assumed that
the directivity patterns of the sensors are the same (2.27). However, it is not true
when there is random sensor errors and the responses of the sensor are perturbating
about a nominal value. Let this nominal value be defined as

$$ a(\theta_k) = \rho(\theta_k)e^{-j\phi(\theta_k)}, $$

(4.57)

where $\rho(\theta_k)$ and $\phi(\theta_k)$ represent the amplitude and phase of the response. Since there
are random shifts in responses for each sensor, instead of (2.27), we have

$$ a_i(\theta_k) = (\rho(\theta_k) + \Delta \rho_i)e^{-j(\phi(\theta_k) + \Delta \phi_i)}, $$

(4.58)

with $\Delta \rho_i$ and $\Delta \phi_i$ being the errors in amplitude and phase respectively. Intuitively,$\Delta \rho_i$ and $\Delta \phi_i$ should be zero means random variables following certain statistical
distributions. This model error is more realistic than those proposed in [18], [36]
because the amplitude and phase errors are treated separately. With this model, by
defining $\Delta \rho'_i = (1 + \frac{\Delta \rho_i}{\rho})$ and follow similar procedures as the case of DOA dependent
errors, the output response vector of the $l$th snapshots is given as

$$ y(l) = (A \odot \Delta A(l))s(l) + n(l), $$

(4.59)

where $A$ is defined as usual and

$$ \Delta A(l) = [\Delta a(\theta_1, l) \Delta a(\theta_2, l) \cdots \Delta a(\theta_d, l)] , $$

(4.60)

$$ \Delta a(\theta_k, l) = \begin{bmatrix} 
\Delta \rho'_1(l) \\
\Delta \rho'_2(l) \\
\vdots \\
\Delta \rho'_m(l) 
\end{bmatrix} \odot \begin{bmatrix} 
e^{-j\pi \Delta \phi_1(l)} \\
e^{-j\pi \Delta \phi_2(l)} \\
\vdots \\
e^{-j\pi \Delta \phi_m(l)} 
\end{bmatrix} . $$

(4.61)
It can be observed that (4.59) is exactly the same as (4.6) except $\Delta a$ is given by (4.61) while $\delta a$ is given by (4.8). This coincidence enables us to make the following assumptions which are similar to A4.1 to A4.6:

**Assumption A4.7:** The sensor errors $\Delta \rho_i(l)$ and $\Delta \phi_i(l)$ are independent of each other and are independent of the signals and noises.

**Assumption A4.8:** Following the same reason as A4.4, we assume that instead of $\Delta \rho_i(l)$, it is $\frac{\Delta \rho_i(l)}{\rho}$ that follows certain probability distribution. That is to say, the errors are independent of the DOAs.

**Assumption A4.9:** $\Delta \rho_i(l)$ and $\frac{\Delta \phi_i(l)}{\rho}$ for different $i$ and $l$ are assumed to be the outcomes of two random variables $\Delta \rho$ and $\Delta \phi$ respectively (correspondingly $\Delta \rho_i'(l)$ are outcomes of the random variable $\Delta \rho'$). This assumption bases on the fact that the sensors are the same and the snapshots are taken at independent instances.

Based on assumptions above and using the same formulating procedures as the case of DOA dependent errors, the perturbed covariance matrix of data vector $y$ under white noise case is given as

$$R_{yy} = (ASA^*) \mathcal{T} U + \sigma^2 I,$$  \hspace{1cm} (4.62)

where $U \in \mathbb{R}^{m \times m}$ is the perturbing matrix with $V(k,k) = \sigma^2 \rho$, $V(k,l) = \mu \phi$ for $k \neq l$ and

$$\sigma^2 \rho = E\{(\Delta \rho')^2\} = E\{(1 + \Delta \rho)^2\},$$ \hspace{1cm} (4.63)

$$\mu \phi = E\{e^{-j\pi \phi}\}.$$ \hspace{1cm} (4.64)

As an example, if $\Delta \rho \sim \mathcal{U}[-K,K]$ and $\Delta \phi \sim \mathcal{U}[-L,L]$, then $U$ is of the form
\[
U = \begin{bmatrix}
1 + \frac{k^2}{3} & \text{sinc}^2\pi L & \cdots & \text{sinc}^2\pi L \\
\text{sinc}^2\pi L & 1 + \frac{k^2}{3} & \cdots & \text{sinc}^2\pi L \\
\vdots & \vdots & \ddots & \vdots \\
\text{sinc}^2\pi L & \text{sinc}^2\pi L & \cdots & 1 + \frac{k^2}{3}
\end{bmatrix}
\]  
(4.65)

Note that \( U \) has the same properties as \( V \) in (4.11), i.e. it is p.d., symmetric and Toeplitz. Our algorithm for compensating the perturbations (when the statistical parameters are known) can be used here without modifications. The only point to be noted is that the diagonal elements of \( U \) is greater than one and hence scaling of the noise variance \( \sigma^2 \) must be taken into account when applying the algorithm. Theoretically, in case of unknown perturbing parameters, the estimating algorithm described in Section 4.2.6 is still applicable though it may take longer time as the there are two unknown parameters.

It can be shown very easily that the perturbation matrix is exactly the same for ULA used in ESPRIT if the sensors are assumed to have identical random errors distribution (which is very reasonable for an ULA). Hence the above discussions concerning the behaviour of \( U \), the compensating and estimating algorithms are also applicable to the case for ESPRIT. On account of this, no further discussions are addressed for DOA independent errors occurred in the method of ESPRIT.
4.4 Simulation Results

Simulation results are presented to demonstrate the various aspects of the analysis and the compensating algorithm. Focus will be made on ULA configuration with DOA dependent errors in the case of root-MUSIC. In all the simulation results shown below, there are totally five emitting sources \((d = 5)\) with DOAs at 0.94, 1.04, 1.21, 1.28, 1.4 (all are in radians) and \(\bar{d}_a\) is set to one. The signal covariance matrix \(S\) is given by

\[
S = \begin{bmatrix}
1.00 & 0 & 0 & 0 & 0 \\
0 & 0.75 & 0 & 0 & 0 \\
0 & 0 & 0.94 & 0 & 0 \\
0 & 0 & 0 & 0.90 & 0 \\
0 & 0 & 0 & 0 & 0.80 \\
\end{bmatrix}.
\]  

(4.66)

4.4.1 Significance of the compensating algorithm

In these sets of simulations, we will show that in the presence of perturbations, the estimates obtained from standard MUSIC are biased while unbiased results are obtained with our compensating algorithms. Twelve sensors \((m = 12)\) are used and without loss of generality, they are assumed to have unity response i.e. \(a(\theta_k) = 1\) for all \(k\)’s. Moreover, \(\Sigma_N\) is assumed to be the identity matrix \(I\) and \(\sigma = 0.1\). To reduce the finite samples effect, a large number of snapshots \((2000)\) are used for forming the covariance matrix. Both uniform \((U[-L, L], L = 0.06)\) and Gaussian \((N[0, \sigma_\theta], \sigma_\theta = 0.04)\) distributed perturbations are considered. By (4.11) and (4.12), the elements of the perturbation matrix \(V\) in both cases are not nearly zero. Thus the problem of division by zero is not considered here (this problem will be treated in the next set of simulations). For each kind of distributions, a total of 100 Monte Carlo simulations are carried out. Figure 4.4 and Figure 4.5 respectively show one of the results obtained from uniform and Gaussian distributed perturbations. Note that the values of \(P_M(\theta)\) has been normalized to one for the sake of plotting. The means of the estimated DOAs for the 100 simulations are shown in Table.4.1. Results obtained from three different situations are shown:
(a). Standard MUSIC applied to measurements without random perturbations,

(b). Standard MUSIC applied to data with random perturbations,

(c). MUSIC modified by our approach applied to data with the same perturbations as in (b). The perturbation parameter is assumed to be known.

From Figure 4.4 and Figure 4.5, we can see that if only standard MUSIC is applied to estimate the DOAs of the sources with random angle deviations (dash line (b)), the DOAs estimated are deviated from the actual values. Moreover, the peaks indicating the position of DOAs spread out a lot as compared to the case without angle shifts (solid line (a)). On the contrary, the results improve a lot when our procedures are applied with sharper peaks and much more accurate estimated DOAs (dashed-dotted line (c)). Furthermore, we can deduce from Table.4.1 and Table.4.2 that the DOAs estimated by using our modifying MUSIC is better (at least have similar performance as the case of standard MUSIC without random perturbations) while biased results are obtained when only standard MUSIC is applied to the snapshots (the arguments are clear by looking at the absolute errors of the estimates).

Two more figures are included to illustrate the biasing effect of standard MUSIC under the presence of perturbating errors. Figure 4.6 shows the first fifty set of results obtained from standard MUSIC for the uniformly distributed perturbations. The biasedness can be seen easily (the vertical dashed line indicate the position of the actual DOAs). Figure 4.7 shows the same fifty set of results as in Figure 4.6 but modified by using our compensating method. Though the positions of the peaks varied a little, they tend to clustered around the actual DOAs, revealing the unbiased nature of the algorithm. Since the maximum errors ($L = 0.06$) is comparable to the DOA difference of the third and the forth source ($= 0.07$), it is reasonable to expect poorer results for these two sources. This phenomenon is also reflected by the results in Table.4.1 and Table.4.2 by which the absolute errors of the estimated DOAs of the third and forth emitting sources are relatively larger than the others.
Figure 4.4: Plot of $P_M(\theta)$ against DOA $\theta$ (4.4) for uniform distributed deviations. (a) standard MUSIC without random perturbations in DOAs, (b) standard MUSIC with DOA shifts and (c) modified MUSIC with the same random DOAs shifts as in (b) with known perturbation parameter. The stems indicate the actual DOAs used in the simulations.

| TRUE DOAs | Means of Estimates ($\hat{\theta}_k$) | Errors of Estimates ($|\hat{\theta}_k - \theta_k|$) |
|-----------|--------------------------------------|---------------------------------------------------|
| 0.9400    | 0.9400 0.9290 0.9398                  | 0.0000 0.0110 0.0002                               |
| 1.0401    | 1.0401 1.0290 1.0411                  | 0.0001 0.0110 0.0011                               |
| 1.2105    | 1.2105 1.1995 1.2115                  | 0.0005 0.0105 0.0015                               |
| 1.2800    | 1.2795 1.2890 1.2750                  | 0.0005 0.0090 0.0050                               |
| 1.4000    | 1.3999 1.4090 1.3998                  | 0.0001 0.0090 0.0002                               |

Table 4.1: Table showing the means and errors of the estimated DOAs for uniform distributed perturbations. (a) standard MUSIC without random perturbations in DOAs, (b) standard MUSIC with random DOA shifts and (c) MUSIC modified by our procedures with same shifts in DOAs as in (b) with known perturbation parameter.
Figure 4.5: Plot of $P_{bf}(\theta)$ against DOA $\theta$ (4.4) for Gaussian distributed deviations. Descriptions same as Figure 4.4.

| TRUE DOAs | Means of Estimates ($\hat{\theta}_k$) | Errors of Estimates ($|\hat{\theta}_k - \theta_k|$) |
|-----------|--------------------------------------|-----------------------------------------------|
|           | (a) | (b) | (c) | (a) | (b) | (c) |
| 0.9400    | 0.9401 | 0.9290 | 0.9407 | 0.0001 | 0.0110 | 0.0007 |
| 1.0400    | 1.0400 | 1.0292 | 1.0407 | 0.0000 | 0.0108 | 0.0017 |
| 1.2100    | 1.2106 | 1.2000 | 1.2147 | 0.0006 | 0.0100 | 0.0047 |
| 1.2800    | 1.2794 | 1.2890 | 1.2743 | 0.0006 | 0.0090 | 0.0057 |
| 1.4000    | 1.4000 | 1.4090 | 1.4000 | 0.0000 | 0.0090 | 0.0000 |

Table 4.2: Table showing the means and errors of the estimated DOAs for Gaussian distributed perturbations. (a) standard MUSIC without random perturbations in DOAs, (b) standard MUSIC with random DOA shifts and (c) MUSIC modified by our procedures with same shifts in DOAs as in (b) with known perturbation parameter.
Figure 4.6: Plot of $P_M(\theta)$ against DOA $\theta$ (4.4). The results correspond to unmodified MUSIC with uniform distributed perturbations. The vertical dashed lines indicate the actual DOAs. Biased estimates are obtained.

Figure 4.7: Plot of $P_M(\theta)$ against DOA $\theta$ (4.4). The results correspond to modified MUSIC with uniform distributed perturbations. The vertical dashed lines indicate the actual DOAs. Unbiased estimates are obtained.
4.4.2 Problem of division by zero

In this section, the problem of division by zero is investigated. To show its effect on the compensating algorithm, the number of sensors is increased to eighteen \((m = 18)\) and the maximum range \(L\) for uniformly distributed perturbation is set to \(L = 0.0625\) while other parameters are keeping unchanged. It can be checked easily that \(V(i,j)\) is almost zero when \(|i-j| = 16\). These zero elements will cause catastrophic effect of noise amplification during the \(\odot\) operation and hence our compensating algorithm fails to locate the correct DOAs. The fifty realizations in Figure 4.8 illustrate this effect clearly. To alleviate this problem, the three methods proposed previously have used and their results are shown in the next three figures. Figure 4.9 shows results by reducing the number of sensors used \((m = 12)\). Though the problem is solved, the resolution power is reduced too (the peaks are not as sharp as the case with eighteen sensors). Figure 4.10 illustrate the thresholding method. From the simulation results, it is the best method as it preserves the resolution power yet it does not introduce bias as the case of regularization with results shown in Figure 4.11. Clearly bias are introduced to the estimates by using regularization. As a conclusion, the best method for alleviating the problem of division by zero is thresholding. The means and errors of the 100 estimates are included in Table 4.3 below for reference.

| TRUE DOAs | Means of Estimates \((\hat{\theta}_k)\) | Errors of Estimates \(|\hat{\theta}_k - \theta_k|\) |
|-----------|--------------------------------|--------------------------------|
|           | (a)       | (b)       | (c)       | (a)       | (b)       | (c)       |
| 0.9400    | 0.9399    | 0.9391    | 0.9364    | 0.0001    | 0.0009    | 0.0036    |
| 1.0400    | 1.0398    | 1.0349    | 1.0335    | 0.0002    | 0.0051    | 0.0065    |
| 1.2100    | 1.2121    | 1.2125    | 1.2145    | 0.0021    | 0.0025    | 0.0045    |
| 1.2800    | 1.2751    | 1.2813    | 1.2818    | 0.0049    | 0.0013    | 0.0018    |
| 1.4000    | 1.4001    | 1.4019    | 1.4011    | 0.0001    | 0.0019    | 0.0011    |

Table 4.3: Table showing the means and errors of the estimated DOAs for uniform distributed perturbations by different methods to overcome the zero element problem. (a) reduce number of sensors (b) thresholding (c) regularization.

67
Figure 4.8: Plot of $P_M(\theta)$ against DOA $\theta$ (4.4). The results correspond to eighteen ($m = 18$) sensors with zero elements in the perturbation matrix but no remedies are performed to remove the problem of division by zero. Modified MUSIC fails in this case. The dashed lines indicate positions of the actual DOAs.

Figure 4.9: Plot of $P_M(\theta)$ against DOA $\theta$ (4.4). The results correspond to reduced number of sensors ($m = 12$) to remove zero elements. The peaks spread out more as compared to that of using 18 sensors.
Figure 4.10: Plot of $P_M(\theta)$ against DOA $\theta$ (4.4). The results correspond to eighteen ($m = 18$) sensors with zero elements in the perturbation matrix. Method of thresholding is used to remove the effect of division by zero. Good and unbiased results are obtained.

Figure 4.11: Plot of $P_M(\theta)$ against DOA $\theta$ (4.4). Same data set as used to produce Figure 4.10 but method of regularization is used to remove the effect of division by zero. Biased results are obtained.
4.4.3 Unknown perturbation parameter

In the previous results, statistical parameter of the perturbation is assumed to be known in advance. To demonstrate the usefulness of the parameter estimation algorithm presented in section 4.2.6 the parameter is assumed to be unknown in the following set of simulations. Both uniform ($\sim \mathcal{U}[-L, L], L = 0.045$) and Gaussian ($\sim \mathcal{N}[0, \sigma_{\theta}], \sigma_{\theta} = 0.04$) distributed errors are considered with sixteen ($m = 16$) sensors. Results of fifty Monte Carlo simulations for each distribution are shown. Figure 4.12 and Figure 4.13 compare results obtained from known and estimated $\hat{L}$ respectively for uniform distributed errors. Figure 4.14 and Figure 4.15 shows results from known and estimated $\sigma_{\theta}$ respectively for Gaussian distribution. It is clear that the performance of the compensating algorithm of using known or estimated perturbation parameters are nearly the same indicating that the parameter estimating method is quite efficient. In fact, the means of the estimated parameters (totally 100 simulations) are very close to the actual ones ($\hat{L} = 0.0448$ and $\hat{\sigma}_{\theta} = 0.0396$). These observations can also be deduced from Table 4.4 below which summarizes results for both kind of perturbation distributions. Results are almost the same for known and estimated perturbation parameters.

<table>
<thead>
<tr>
<th>TRUE DOAs</th>
<th>Uniform Distributed Errors</th>
<th>Gaussian Distributed Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Known $L$</td>
<td>Estimated $\hat{L}$</td>
</tr>
<tr>
<td></td>
<td>(a)</td>
<td>(b)</td>
</tr>
<tr>
<td>0.9400</td>
<td>0.9398</td>
<td>0.0002</td>
</tr>
<tr>
<td>1.0400</td>
<td>1.0398</td>
<td>0.0002</td>
</tr>
<tr>
<td>1.2100</td>
<td>1.2100</td>
<td>0.0000</td>
</tr>
<tr>
<td>1.2800</td>
<td>1.2805</td>
<td>0.0005</td>
</tr>
<tr>
<td>1.4000</td>
<td>1.4002</td>
<td>0.0002</td>
</tr>
</tbody>
</table>

Table 4.4: Table showing the means and errors of the estimated DOAs obtained from known and unknown perturbation parameters ($L$ and $\sigma_{\theta}$) for uniform and Gaussian distributions. (a) Means of Estimated DOAs ($\hat{\theta}_k$) (b) Errors of Estimation ($|\hat{\theta}_k - \theta_k|$). The estimates for $L$ and $\sigma_{\theta}$ are $\hat{L} = 0.0448$ and $\hat{\sigma}_{\theta} = 0.0396$. 

70
Figure 4.12: Plot of $P_M(\theta)$ against DOA $\theta$ (4.4). The results correspond to sixteen ($m = 16$) sensors with uniform DOAs perturbations. The parameter $L$ is known to be 0.045.

Figure 4.13: Plot of $P_M(\theta)$ against DOA $\theta$ (4.4). The results correspond to sixteen ($m = 16$) sensors with uniform DOAs deviations. The parameter $L(= 0.045)$ is assumed to be unknown but estimated to have mean value $\hat{L} = 0.0448$. The results are comparable to that in Figure 4.12.
Figure 4.14: Plot of $P_M(\theta)$ against DOA $\theta$ (4.4). The results correspond to sixteen ($m = 16$) sensors with Gaussian DOAs perturbations. The parameter $\sigma_\theta$ is known to be 0.04.

Figure 4.15: Plot of $P_M(\theta)$ against DOA $\theta$ (4.4). The results correspond to sixteen ($m = 16$) sensors with Gaussian DOAs deviations. The parameter $\sigma_\theta (= 0.04)$ is assumed to be unknown but estimated to have mean value $\hat{\sigma}_\theta = 0.0396$. The results are comparable to that in Figure 4.14.
CHAPTER 5
CRAMÉR-RAO BOUND ANALYSIS

Being able to place a lower bound on the variance of any unbiased estimator proves extremely useful in practice. At best, it allows us to assert that an estimator is the minimum variance unbiased (MVU) estimator. This will be the case if the estimator attains the bound for all values of the unknown parameter. At worst, it provides a benchmark against which we can compare the performance of any unbiased estimator. Furthermore, it alerts us to the physical impossibility of finding an unbiased estimator whose variance is less than the bound. The latter is often useful in signal processing feasibility studies. Although many such variance bounds exist, the Cramér-Rao Bound (CRB) is by far the easiest to determine. In this chapter, we will examine the CRB of the estimators we got by using the methods described in Chapter 4.

5.1 Theoretical Background

It is well known from statistics that the Cramér-Rao Bound (CRB) set a limit on the performance of an unbiased estimator. An estimator is said to be efficient if its variance attains the CRB. More specifically, given a data vector \( \mathbf{v} \), which depends on a vector of \( r \) parameters, \( \varphi = [\varphi_1, \ldots, \varphi_r]^T \in \mathbb{C}^r \), the covariance of an unbiased estimate of \( \varphi \), \( \hat{\varphi} \) satisfies

\[
\text{cov}(\hat{\varphi}) \geq J^{-1}(\varphi),
\]

or in other words, the matrix \( \text{cov}(\hat{\varphi}) - J^{-1}(\varphi) \) is non-negative definite where \( J(\varphi) \)
is the Fisher Information Matrix whose elements are defined as

$$ \mathbf{J}_{i,j}(\varphi) = \mathbb{E} \left\{ \frac{\partial \ln p(\mathbf{v} | \varphi)}{\partial \varphi_i} \frac{\partial \ln p(\mathbf{v} | \varphi)}{\partial \varphi_j} \right\}, \quad (5.2) $$

with $p(\mathbf{v} | \varphi)$ being the probability density function of the vector $\mathbf{v}$, given the parameters $\varphi$. Though (5.2) seems simple, the RHS is difficult to evaluate in most situations where $p(\mathbf{v} | \varphi)$ is of complicated form. Fortunately in the case of Gaussian observations, i.e.

$$ \mathbf{v} \sim \mathcal{N}(\mu(\varphi), \mathbf{R}(\varphi)), \quad (5.3) $$

a general expression for Fisher information matrix $\mathbf{J}_{i,j}(\varphi)$ is available (see Appendix 3C of [10]) as

$$ \mathbf{J}_{i,j}(\varphi) = \left[ \frac{\partial \mu(\varphi)}{\partial \varphi_i} \right]^T \mathbf{R}^{-1}(\varphi) \left[ \frac{\partial \mu(\varphi)}{\partial \varphi_j} \right] + \frac{1}{2} \text{tr} \left\{ \mathbf{R}^{-1}(\varphi) \frac{\partial \mathbf{R}(\varphi)}{\partial \varphi_i} \mathbf{R}^{-1}(\varphi) \frac{\partial \mathbf{R}(\varphi)}{\partial \varphi_j} \right\}, \quad (5.4) $$

where

$$ \frac{\partial \mu(\varphi)}{\partial \varphi_i} = \left[ \frac{\partial [\mu(\varphi)]_1}{\partial \varphi_i}, \frac{\partial [\mu(\varphi)]_2}{\partial \varphi_i}, \ldots, \frac{\partial [\mu(\varphi)]_r}{\partial \varphi_i} \right]^T, \quad (5.5) $$

and

$$ \frac{\partial \mathbf{R}(\varphi)}{\partial \varphi_i} = \begin{bmatrix}
\frac{\partial [\mathbf{R}(\varphi)]_{11}}{\partial \varphi_i} & \frac{\partial [\mathbf{R}(\varphi)]_{12}}{\partial \varphi_i} & \ldots & \frac{\partial [\mathbf{R}(\varphi)]_{1r}}{\partial \varphi_i} \\
\frac{\partial [\mathbf{R}(\varphi)]_{21}}{\partial \varphi_i} & \frac{\partial [\mathbf{R}(\varphi)]_{22}}{\partial \varphi_i} & \ldots & \frac{\partial [\mathbf{R}(\varphi)]_{2r}}{\partial \varphi_i} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial [\mathbf{R}(\varphi)]_{r1}}{\partial \varphi_i} & \frac{\partial [\mathbf{R}(\varphi)]_{r2}}{\partial \varphi_i} & \ldots & \frac{\partial [\mathbf{R}(\varphi)]_{rr}}{\partial \varphi_i}
\end{bmatrix}. \quad (5.6) $$

For the special case in which the vector $\mathbf{v}$ is complex, zero mean (i.e. $\mu(\varphi) = 0$), circularly Gaussian (i.e. uncorrelated real and imaginary parts with equal variance) with covariance matrix $\mathbf{R}_{\mathbf{v}\mathbf{v}}(\varphi)$, the probability density function for a set of $N$ independent sample vectors $\mathbf{Z} = [\mathbf{v}_1 \ldots \mathbf{v}_N]$, is

74
\[
p(Z) = \frac{1}{(\pi \det(\hat{R}_{\nu\nu}(\varphi)))^N} e^{-\text{tr}\left\{R^{-1}_{\nu\nu}(\varphi)\hat{R}_{\nu\nu}\right\}}, \tag{5.7}
\]

with \( \hat{R}_{\nu\nu} = \frac{1}{N} ZZ^* \) is the sample correlation matrix. Under this case, the Fisher Information Matrix is given by

\[
J_{i,j}(\varphi) = N \text{ tr} \left\{ R^{-1}_{\nu\nu}(\varphi) \frac{\partial R_{\nu\nu}(\varphi)}{\partial \varphi_i} R^{-1}_{\nu\nu}(\varphi) \frac{\partial R_{\nu\nu}(\varphi)}{\partial \varphi_j} \right\}. \tag{5.8}
\]

With this much simpler formula, we are able to apply the notion of CRB to study the efficiencies of our compensating and estimating algorithms. Note that in the sequel, only Gaussian distributed errors is considered as the case of uniform perturbations is similar and is omitted. Moreover for simplicity, the signal covariance matrix is taken to be diagonal and White Noise is assumed.

### 5.2 Derivation of CRB

Before deriving expressions of the CRB for our problem, we have to identify the unknown parameter vector. By following the arguments of [16], for fixed \( \theta_k, \sigma \) and \( \sigma_\theta \), we are able to find the estimate of the signal covariance matrix by using the following formula

\[
\hat{S} = A^d((\hat{R}_{\vec{x}\vec{x}} \otimes V) - \sigma^2 I)(A^d)^*.
\tag{5.9}
\]

This means that the unknown parameters in our problem are the DOAs \( \theta_k \), the noise variance \( \sigma^2 \) and the perturbation variance \( \sigma_\theta^2 \). Thus

\[
\varphi = \begin{bmatrix}
\theta_1 & \cdots & \theta_d & \sigma^2 & \sigma_\theta^2
\end{bmatrix}^T \in \mathbb{R}^{d+2}.
\tag{5.10}
\]

From (4.6), the perturbed data vector of the array sensor \( \vec{x} \in \mathcal{C}^m \) is given as

\[
\vec{x}(l) = (A \otimes \delta A(l)) s(l) + n,
\tag{5.11}
\]

and hence the mean \( \mu \) is
\[ \mu(\varphi) = \mathbb{E}\{\bar{x}\} \]
\[ = \mathbb{E}\{\mathbf{A} \odot \delta\mathbf{A}(l)s(l) + \mathbf{n}\} \]
\[ = (\mathbf{A} \odot \mathbb{E}\{\delta\mathbf{A}(l)\})\mathbb{E}\{s(l)\} + \mathbb{E}\{\mathbf{n}\} \]
\[ = 0. \quad (5.12) \]

The third equality follows from the fact that the perturbations and the signals are statistically independent and the last equality is due to the zero-mean nature of the signals and noises. The covariance matrix \( R_{\bar{x}\bar{x}} \) is already derived in Theorem T4.1 of Chapter 4 and is given as

\[ R_{\bar{x}\bar{x}}(\varphi) = (\mathbf{A}\mathbf{S}\mathbf{A}^*) \odot \mathbf{V} + \sigma^2 \mathbf{I}. \quad (5.13) \]

The most important assumption made is that the perturbed data is circularly Gaussian distributed with zero mean ((5.12)) and covariance matrix given by (5.13). The rationale is based on the Central Limit Theorem [19] which states that the sum of \( r \) random variables of different probability distribution will be approximately Gaussian if \( r \) is sufficiently large. This assumption allows us to use (5.8) to find the Fisher Information Matrix. The various partial differentials are derived as follows:

\[ \frac{\partial R_{\bar{x}\bar{x}}}{\partial \sigma^2} = \mathbf{I}, \quad (5.14) \]

\[ \frac{\partial R_{\bar{x}\bar{x}}}{\partial \theta_i} = s_i [d(\theta_i)^*a^*(\theta_i) + a(\theta_i)d^*(\theta_i)] \odot \mathbf{V}, \quad (5.15) \]

where \( s_i \) is the \( i \)th diagonal element of \( \mathbf{S} \), \( a(\theta_k) \) is of the same form as (4.3) and \( d(\theta_i) \) is defined to be

\[ d(\theta_i) = \frac{da(\theta)}{d\theta} \bigg|_{\theta=\theta_i}, \quad (5.16) \]

\[ \frac{\partial R_{\bar{x}\bar{x}}}{\partial (\sigma^2)} = (\mathbf{A}\mathbf{S}\mathbf{A}^*) \odot \frac{d\mathbf{V}}{d(\sigma^2)}, \quad (5.17) \]

with
\[
\frac{dV}{d(\sigma^2_\theta)} = -\left[
\begin{array}{cccc}
0 & \frac{\pi^2}{2} & \ldots & \frac{(m-1)\pi^2}{2} \\
\frac{\pi^2}{2} & 0 & \ldots & \frac{(m-2)\pi^2}{2} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{(m-1)\pi^2}{2} & \frac{(m-2)\pi^2}{2} & \ldots & 0
\end{array}\right] \odot V.
\] (5.18)

Because of the complexity introduced by the Schur product operator \( \odot \) in the expression, analytic expressions for the Cramér-Rao Bounds are very difficult to obtain. Typically, the bounds are calculated numerically, as were done in our case. Since \( J(\varphi) \) is of dimension \((d+2) \times (d+2)\) only, extensive computational time is not required and this suits our purpose. Results for various situations are shown in the next section.

### 5.3 Simulation Results

In this section, several factors affecting the performance of the compensating and estimating algorithms are investigated, including the effect of the number of sensors on the results, and the finite sample effects.

#### 5.3.1 Effect of the number of sensors

As revealed in Chapter 4, the number of sensors used have a determining influence on the estimates obtained. Simulations are carried out to show this effect. Five emitting sources are used with DOAs 0.94, 1.04, 1.21, 1.28, 1.4 (all are in radians) and signal covariance matrix \( S \) set as \((4.66)\). Gaussian distributed errors are assumed with parameter \( \sigma_\theta = 0.04 \) and noise variance \( \sigma^2 = 0.02 \). The number of sensors are increased from eight to twenty-three. The theoretical CRBs are found by using the formulae (5.8) and (5.14) through (5.18). The variances of the estimates from 100 Monte Carlo simulations are calculated. For comparisons, the two standard deviations are plotted in the same figures which are shown below.
Figure 5.1: Plots of the theoretical and simulated variances of the estimated parameters against $m$. Boundary for 99% confidence level for the MS estimation error is also plotted. (a). CRB of $\theta_1 = 0.94$, (b). CRB of $\theta_2 = 1.04$.

Figure 5.2: Plots of the theoretical and simulated variances of the estimated parameters against $m$. Boundary for 99% confidence level for the MS estimation error is also plotted. (a). CRB of $\theta_3 = 1.21$, (b). CRB of $\theta_4 = 1.28$. 

Figure 5.3: Plots of the theoretical and simulated variances of the estimated parameters against $m$. Boundary for 99% confidence level for the MS estimation error is also plotted. (a). CRB of $\theta_5 = 1.40$, (b). CRB of $\sigma_\theta = 0.04$.

From Figure 5.1 to Figure 5.3, we can see that the standard deviations obtained from simulations by using our proposed idea is quite close to the theoretical CRB. This shows that our algorithms are very efficient. Moreover, though in principle the results will be better if more sensors are used, it is not the case as deduced from the plots. From the concave shape of the simulated results, it seems that there is an optimal point for the number of sensors in the array. This strange phenomenon can be explained by the "division by zero" effect described in Chapter 4 by which undesirable noise amplification degrades the performance of the algorithms. This problem prevents us from increasing the number of sensors indefinitely. Moreover, the boundary of the 99% confidence level is also plotted in the diagram. This confidence level is obtained by calculating the standard deviation of the squared errors so that the range is within three times of this standard deviation.
5.3.2 Finite sample effect

To demonstrate the effect of the number of snapshots used, simulations are also done on this aspect. To simplify the case, two emitting sources at DOAs 0.94 and 1.04 (both in radians) are used with 10 sensors and signal covariance matrix $S = \text{Diag}([1 \ 0.75]^T)$. Gaussian distributed perturbations with $\sigma_\theta = 0.04$ is used and noise variance is $\sigma = 0.02$. The number of snapshots used are $N = 200, 400, 800, \ldots, 2000$ and the results are plotted in Figure 5.4 and Figure 5.5. In the figures, the 99% of confidence levels are also shown as reference so that they imposes upper bound for the simulated results. The closeness between the theoretical and the simulated curves further reinforce the idea that the algorithm is efficient. Moreover, when the number of snapshots increase, the errors variances decrease steady indicating that when $N \to \infty$, the simulated variances will approach the theoretical CRB.

5.4 Conclusions

Although a compact form of the CRB matrix for our problem is not available, numerical simulations are possible by substituting actual parameters in the equations derived. By performing this numerical CRB analysis, we are able to illustrate that estimates obtained by using the compensating and estimating algorithm are indeed efficient and unbiased. This is a very significant point as achieving unbiasedness and efficiency is of crucial importance in the field of parameter estimation.
Figure 5.4: Plots of the theoretical and simulated variances of the estimated parameters against $N$. Boundary for 99% confidence level for the MS estimation error is also plotted. (a). CRB of $\theta_1 = 0.94$, (b). CRB of $\theta_2 = 1.04$.

Figure 5.5: Plots of the theoretical and simulated variances of the estimated parameters against $N$. Boundary for 99% confidence level for the MS estimation error is also plotted. (a). CRB of $\sigma^2 = 0.02$, (b). CRB of $\sigma_\theta = 0.04$. 
CHAPTER 6
AN APPLICATION : BLIND DECONVOLUTION

Though we have restricted our discussions of the compensating algorithm to problem of DOAs estimation, the method can also be applied to other fields. As an example, an application concerning blind deconvolution is considered in this chapter. We will first generalize Bresler’s idea of resolving overlapping echoes [4] to blind deconvolution. The major disadvantages of the generalized method are then revealed and remedies which exploited the principle of our compensating algorithm is proposed. Simulation results are included to show the significance of the whole deconvolution process.

6.1 Blind Deconvolution by ESPRIT

Though there are a number of literatures such as [33] [34] [35] discussing the problem of blind deconvolution, the problem has not been solved completely. In [4], Bresler has utilized the invariant nature of ESPRIT to resolve overlapping echoes. In this section, a generalization is made so that the method can be applied to blind deconvolution.

6.1.1 Mathematical formulation

Consider the situation as depicted in Figure 6.1. In the diagram, a system with an unknown system response $g(t)$ is excited by a signal $s(t)$ which consists of periodic random impulses is shown. By periodic, we mean that the impulses occurred with a fixed period as shown in Figure 6.1. The output of the system is given by
\[ y(t) = g(t) * s(t). \]

Figure 6.1: Figure showing configuration for blind deconvolution and an example of \( s(t) \).

Since the impulses are periodic, we can consider a period of time \( T \) which is an integral multiple of the impulse period and reformulate that portion of the output as follows:

\[
y(t) = \sum_{k=1}^{d} s_k g(t - \tau_k) + n(t) \quad 0 \leq t \leq T, \tag{6.1}
\]

where \( d \) is the number of effective impulses in the period \( T \), \( s_k \) and \( \tau_k \) are the amplitudes and the delay factors of the impulses respectively, \( n(t) \) is random noises of the system. By considering a total length of time \( NT \), \( N \) snapshots are obtained. Now taking Fourier transform on (6.1) and adding the index \( i \) to represent the \( i \)th snapshots, we get
\[ Y_l(\Omega) = \sum_{k=1}^{d} s_{ik} G(\Omega) \exp(-j\Omega \tau_{ik}) + N_l(\Omega). \]  

(6.2)

On the other hand, by sampling \( y(t) \) at a rate of \( t_o \), we obtained \( m \) samples of the signal as follows

\[ y_l(t_o) = \sum_{k=1}^{d} s_{ik} g(t_o - \tau_{ik}) + n_l(t_o) \quad l = 0, 1, \ldots, m - 1. \]  

(6.3)

Here, it is assumed that the sampling rate is large enough so that aliasing is avoided. By taking a \( 4m \) point DFT of the digital samples, it is equivalent to [15] sampling the continuous spectrum \( Y_l(\Omega) \) at a rate of \( \frac{2\pi}{4mt_o} \), giving

\[ Y_{il} = \sum_{k=1}^{d} s_{ik} G(\frac{2\pi l}{4mt_o}) \exp(-j\frac{2\pi l}{4mt_o} \tau_{ik}) + N_l(\frac{2\pi l}{4mt_o}). \]  

(6.4)

Under ideal case, the snapshots are consistent which means that the delay factors of the impulses within each period will not change with time. The non-ideal case by which the delay factors change with time will be considered shortly. Thus in the mean time, we will assume \( \tau_{ik} = \tau_k \) for all \( i \). Moreover, to simplify the notation, let \( G_l = G(\frac{2\pi l}{4mt_o}) \), \( \phi_k = \exp(-j\frac{2\pi l}{4mt_o} \tau_k) \) and \( n_{il} = N_l(\frac{2\pi l}{4mt_o}) \). The equation can be written in a compact matrix form as

\[ y_i = G M s_i + n_i, \]  

(6.5)

where

\[ M = \begin{bmatrix}
\phi_1^0 & \phi_2^0 & \cdots & \cdots & \phi_d^0 \\
\phi_1^1 & \phi_2^1 & \cdots & \cdots & \phi_d^1 \\
\vdots & \vdots & \ddots & \cdots & \vdots \\
\vdots & \vdots & \ddots & \cdots & \vdots \\
\phi_1^{4m-1} & \phi_2^{4m-1} & \cdots & \cdots & \phi_d^{4m-1}
\end{bmatrix}, \]  

(6.6)

\[ G = \text{diag}[G_0 \quad G_1 \quad \cdots \quad G_{4m-1}] \]  

(6.7)

and

\[ n_i = [n_{i1} \quad n_{i2} \quad \cdots \quad n_{i4m-1}]. \]  

(6.8)
Note that $M$ is a Vandermonde matrix and has similar form as the steering matrix described in Chapter 2. If the diagonal matrix $G$ is known, the delay factors can be found easily by using either MUSIC or ESPRIT. However, in blind deconvolution, $g(t)$ is unknown and under such situation, we may utilize the special rotational invariant properties of ESPRIT to find both the delay factors and $g(t)$ as follows. Assuming all the signals are real, then the DFT points will be symmetric (in amplitude and anti-symmetric in phase). Hence we only need to consider the first $2m$ portions of the DFT sequence. Let's divide the sequence into two parts corresponding to odd and even indices as

$$
y_{oi} = G_o M_o s_i + n_{oi} \in \mathcal{C}^m,
$$

$$
y_{ei} = G_e M_e s_i + n_{ei} \in \mathcal{C}^m,
$$

where

$$
G_o = diag[G_1, G_3, \cdots, G_{2m-1}],
$$

$$
G_e = diag[G_0, G_2, \cdots, G_{2m-2}],
$$

$$
M_o = \begin{bmatrix} m_1 & m_3 & \cdots & m_{2m-1} \end{bmatrix}^T,
$$

$$
M_e = \begin{bmatrix} m_2 & m_4 & \cdots & m_{2m-2} \end{bmatrix}^T,
$$

with

$$
m_i = [\phi_1^i \phi_2^i \cdots \phi_d^i]^T,
$$

By stacking the vectors $y_{oi}$ and $y_{ei}$, we have the following compact form

$$
x_i = \begin{bmatrix} y_{oi} \\ y_{ei} \end{bmatrix} = \begin{bmatrix} G_o M_o s_i \\ G_e M_e s_i \end{bmatrix} + \begin{bmatrix} n_{oi} \\ n_{ei} \end{bmatrix}.
$$

Observe that $M_e = M_o \Phi$, with $\Phi = diag[\phi_1 \phi_2 \cdots \phi_d]$ and if the continuous spectrum $y(\Omega)$ is smooth enough (this can be achieved by increasing the sampling rate), we may assume $G_o \approx G_e$. With these arguments, (6.15) may be rewritten as
\[ x_i \approx \begin{bmatrix} G_o M_o s_i \\ G_o M_o \Phi s_i \end{bmatrix} + \begin{bmatrix} n_{oi} \\ n_{ei} \end{bmatrix} = A_o s_i + \begin{bmatrix} n_{oi} \\ n_{ei} \end{bmatrix} \]  \hspace{1cm} (6.16)

The covariance matrix is given as

\[ R_{xx} = A_o S A_o^* + \sigma^2 I. \]  \hspace{1cm} (6.17)

by which we have assumed the noises to be not only temporally white but also white in the frequency domain and \( S \) records the correlation between the impulses within one period. With this procedure, the deconvolution problem is reformulated into a framework suitable for ESPRIT with \( G_o M_o \) being the steering matrix and \( \Phi \) being the rotational matrix. By following the procedures as described in Chapter 3, \( \Phi \) and hence the delay factors \( \tau_k \) can be estimated. Moreover, the steering matrix may also be determined to within a diagonal, complex scaling factor \( D \). The procedures for estimating the steering matrix are described in [4] and is omitted here. Once the steering matrix is determined, \( G \) and hence \( g(t) \) can be found easily by performing an inverse Fourier transform. Hereafter, this method is referred as UNMODIFIED deconvolution method.

### 6.1.2 Simulation results

To demonstrate the ability of the UNMODIFIED algorithm, simulations are carried out and the results are presented below. The impulse response function \( g(t) \) used is a wavelet waveform. The input impulses \( s(t) \) repeat their statistical properties for a period of four, which means that any four consecutive impulses have the same statistical relation as the next four impulses. The signal covariance matrix of the four impulses within one period is given by
\[
S = \begin{bmatrix}
1.0 & 0 & 0 & 0 \\
0 & 0.75 & 0 & 0 \\
0 & 0 & 0.9 & 0 \\
0 & 0 & 0 & 0.8 \\
\end{bmatrix},
\]

(6.18)

and the delay factors are given as \( \tau_1 = 0, \tau_2 = 0.25, \tau_3 = 0.390 \) and \( \tau_4 = 0.625 \) (which are expressed as fractions of the total period). The SNR ratio is 20dB and totally 1025 snapshots are used to form the covariance matrix. 100 of them are shown in Figure 6.2 as reference. The results are shown in Figure 6.3 to Figure 6.6. Since the estimated pulses are shifted and scaled, for the sake of comparison, all the pulses are normalized to have maximum amplitude of one and shifted to the center accordingly before plotting in the figures. The estimated \( \tau \) are given as \( \hat{\tau}_1 = 0, \hat{\tau}_2 = 0.289, \hat{\tau}_3 = 0.411 \) and \( \hat{\tau}_4 = 0.643 \). Although the estimated delay factors deviate quite a lot from the actual ones, from the figures, it is clear that the deconvolving results are very good as little errors are observed between the original and the estimated impulse responses. The deviations of the estimated delay factors are mainly due to the assumption \( G_o \approx G_e \) which is an unavoidable assumption of the algorithm.

Figure 6.2: Samples of snapshots for performing blind deconvolution. Note that these snapshots are obtained from data with no perturbations in the delay factors.
Figure 6.3: Comparison between actual and estimated pulses (First Pulse) for the ideal case (i.e. there are no deviations in the delay factors).

Figure 6.4: Comparison between actual and estimated pulses (Second Pulse) for the ideal case (i.e. there are no deviations in the delay factors).
Figure 6.5: Comparison between actual and estimated pulses (Third Pulse) for the ideal case (i.e. there are no deviations in the delay factors).

Figure 6.6: Comparison between actual and estimated pulses (Fourth Pulse) for the ideal case (i.e. there are no deviations in the delay factors).
6.2 Modified Blind Deconvolution by ESPRIT

6.2.1 Problems and remedies of the algorithm

Though the results presented in Section 6.1.2 are very promising, they are a little bit ideal. In practical situations, the algorithm has several problems and disadvantages. The most important one is the deviations in the delay factors. In reality, the impulses will not be exactly periodic in the way described previously and delay factors deviate from time to time around a mean value. The situation is common in communication systems where multipath propagation is unavoidable. The second problem is that the period of the signals may be unknown or inexact due to imprecise estimation, equipment error or system properties. Large errors and discrepancies will occur as a result, especially when the snapshots are obtained sequentially from a long but single time series. These random perturbations degrade the performance of the algorithm greatly as will be shown in latter simulations.

To study the effect of the delay factor and period deviations, the random model and correcting techniques as described in Chapter 4 for studying random errors in MUSIC and ESPRIT is intuitively introduced here. First of all, notice that the deviations in the delay factors are random in nature. Moreover, though the perturbations introduced by inexact period is not really random, if this error is small and there are a large number of snapshots which are obtained sequentially from single time series, we may assume the perturbations to be uniform. Hence the delay factors may be expressed in the form

$$\tau_{ik} = \tau_k + \theta_{ik}, \quad (6.19)$$

here \(\tau_k\) is the mean delay factor for the \(k\) impulses and \(\theta_{ik}\) represents the random perturbations from the actual delay. With this, (6.5) may be rewritten as

$$y_i = GM_i s_i + n_i, \quad (6.20)$$

where \(M_i\) is given by
\[ M_i = \begin{bmatrix}
\phi_0^0 & \cdots & \phi_d^0 \\
\phi_1^0 & \cdots & \phi_d^1 \\
\vdots & \cdots & \vdots \\
\phi_1^{4m-1} & \cdots & \phi_d^{4m-1}
\end{bmatrix} \odot \begin{bmatrix}
\Theta_0^{0} & \cdots & \Theta_d^{0} \\
\Theta_1^{0} & \cdots & \Theta_d^{1} \\
\vdots & \cdots & \vdots \\
\Theta_1^{4m-1} & \cdots & \Theta_d^{4m-1}
\end{bmatrix}
\]

\[ = M \odot P_i, \quad (6.21) \]

with \( \Theta_{ik} = \exp^{-j2\pi k \theta_{ik}} \) and \( \odot \) represents Schur product.

Observe that (6.20) has the same form as (4.7) in Chapter 4. In fact, it can be shown easily that under suitable assumptions (such as \( \theta_{ik} \) are i.i.d. so that \( \Theta_{ik} \) may be treated as outcomes of a random variable \( \Theta \)) as stated in Chapter 4, (6.17) can be recasted as (see Appendix G)

\[ R_{xx} = (A_s S A_s^*) \odot V + \sigma^2 I, \quad (6.22) \]

where

\[ V = \begin{bmatrix}
V_1 & V_3 \\
V_2 & V_1
\end{bmatrix}, \quad (6.23) \]

with

\[ V_1(i,j) = \mathbb{E}\{\Theta^{2(i-j)}\} \quad (6.24) \]

\[ V_2(i,j) = \mathbb{E}\{\Theta^{2(i-j)-1}\} \quad (6.25) \]

\[ V_3(i,j) = \mathbb{E}\{\Theta^{2(i-j)+1}\}. \quad (6.26) \]

Although ESPRIT is used for the formulation of the problem and the perturbation model is similar, the results are different from that in Chapter 4 (4.55). It is because the random errors in the present case has been modified by the Fourier transform while in Chapter 4, the perturbations enter the equation directly as phase errors. This cause a major difference between the two cases. Despite of this, as the perturbation matrix affect the covariance matrix through Schur-Hadamard product, the compensating algorithm described previously can still be applied though the estimation algorithm cannot be used due to the following reason. It is because the perturbation matrix \( V \)
will have a lot of zero or nearly zero elements. This means that we have to avoid the problem of division by zero by using one of the three techniques as described in Section 4.2.5 of Chapter 4. Note that the first method (i.e. limiting the value of \(m\)) is not feasible here as the algorithm hinges on the smoothness of the transformed spectrum \((G_0 \approx G_e)\) which depends on the value of \(m\). However, if the other two methods are used, the positive definiteness of the perturbation matrix will be destroyed and hence conditions for using the parameter estimation algorithm are violated and cannot be applied here. Thus in case of unknown parameter, a preliminary guess may be used and an approximate \(g(t)\) is found by using this initial guess. The method of alternating projection [38] can then be used to locate the approximate positions of the pulses. Once this is done, the algorithm may be applied recursively to get better results. The disadvantage of this approach is that performing alternating projection is quite time-consuming and therefore not suitable for real-time application. Hereafter, this method is referred as MODIFIED deconvolution algorithm.

6.2.2 Simulation results

To illustrate the two problems described in Section 6.2.1, the delay factors are set to be uniformly distributed with unknown amplitude about their mean values. The mean delay factors are given as \(\tau_1 = 0, \tau_2 = 0.25, \tau_3 = 0.390\) and \(\tau_4 = 0.625\) with the same signal covariance matrix as (6.18). Moreover, the period of the impulses are inexact and deviated from the actual period by 7%. Though this deviation is not large, it will cause great problems as the snapshots are obtained sequentially from the convolved signals. The deviations will be carried onwards and positions of the pulses will vary from snapshots to snapshots. This situation is depicted in Figure 6.7 where 100 sample snapshots are shown. It can be seen that they look messy. Although the perturbation parameter is unknown, by intuitive reasoning, we may assume its maximum amplitude to be half of the estimated period. In fact by observing Figure 6.7, the deviation caused by wrongly estimated period determines the maximum amplitude of the perturbations to be approximately half of the actual period. Results obtained by UNMODIFIED deconvolving method are shown from Figure 6.8 to Figure 6.11. Results for MODIFIED deconvolving algorithm are plotted from Figure 6.12 to Figure 6.15.
Figure 6.7: Samples of snapshots for performing blind deconvolution. Note that these snapshots are obtained from data with wrongly estimated period and random perturbations in the delay factors.

To obtain the results in Figure 6.12 etc. by MODIFIED deconvolution algorithm, method of "thresholding" is used to eliminate the problem of division by zero. The threshold value used is of utmost importance to the smoothness of the resultant pulses because it determines the portion of high frequency to be retained. By extensive simulations, it is found to be around 0.06. By comparing the figures, it is very clear that the MODIFIED method outperforms that of the UNMODIFIED one when there are errors in estimating the actual period or when the delay factors change with time. This fact is further reinforced by Table 6.1 which summarizes the results of the estimated delay factors and the errors in the estimated pulse shape in all the three cases (including that from Section 6.1.2). From the data in Table 6.1, the MSE between the estimated and actual pulses for MODIFIED deconvolution algorithm is much smaller than that of UNMODIFIED method and the estimated delay factors are more accurate. As a conclusion, our MODIFIED algorithm is quite useful and practical for real situation.
Figure 6.8: Comparison between actual and estimated pulses (First Pulse) obtained from data of wrongly estimated period with perturbations in the delay factors. UNMODIFIED deconvolution method is used and bad results are obtained.

Figure 6.9: Comparison between actual and estimated pulses (Second Pulse) obtained from data of wrongly estimated period with perturbations in delay factors. UNMODIFIED deconvolution method is used and bad results are obtained.
Figure 6.10: Comparison between actual and estimated pulses (Third Pulse) obtained from data of wrongly estimated period with perturbations in delay factors. **UNMODIFIED** deconvolution method is used and *bad* results are obtained.

Figure 6.11: Comparison between actual and estimated pulses (Fourth Pulse) obtained from data of wrongly estimated period with perturbations in delay factors. **UNMODIFIED** deconvolution method is used and *bad* results are obtained.
<table>
<thead>
<tr>
<th>Delay Factors</th>
<th>MSE in Estimated Pulse</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\tau_1)</td>
<td>(\tau_2)</td>
</tr>
<tr>
<td>(a) 0.00</td>
<td>0.250</td>
</tr>
<tr>
<td>(b) 0.00</td>
<td>0.289</td>
</tr>
<tr>
<td>(c) 0.00</td>
<td>0.340</td>
</tr>
<tr>
<td>(d) 0.00</td>
<td>0.318</td>
</tr>
</tbody>
</table>

Table 6.1: Table showing the estimated delay factors and mean squared error between the estimated and actual pulses for different situations. (a) refers to actual values. (b) refers to UNMODIFIED deconvolution method without any perturbations. (c) corresponds to UNMODIFIED deconvolution method with wrongly estimated period and deviations in delay factors. (d) shows results for MODIFIED deconvolution algorithm with the same wrongly estimated period and delay factor deviations as in (c).

Figure 6.12: Comparison between actual and estimated pulses (First Pulse) obtained from data of wrongly estimated period with perturbations in delay factors. MODIFIED deconvolution method is used and good results are obtained.
Figure 6.13: Comparison between actual and estimated pulses (Second Pulse) obtained from data of wrongly estimated period with perturbations in delay factors. MODIFIED deconvolution method is used and good results are obtained.

Figure 6.14: Comparison between actual and estimated pulses (Third Pulse) obtained from data of wrongly estimated period with perturbations in delay factors. MODIFIED deconvolution method is used and good results are obtained.
Figure 6.15: Comparison between actual and estimated pulses (Fourth Pulse) obtained from data of wrongly estimated period with perturbations in delay factors. MODIFIED deconvolution method is used and good results are obtained.
CHAPTER 7

CONCLUDING REMARKS

In this thesis, a novel approach in analyzing random perturbations in subspace fitting array signal processing methods is introduced. To illustrate the whole idea, it is applied to the classical problem of DOAs estimation. As different from the approaches commonly adopted by other researchers, the various kinds of errors are divided into two groups, namely DOA dependent and DOA independent errors which are assumed to be independent. Based on certain reasonable assumptions, the perturbed covariance matrix is shown to be the Schur-Hadamard product between a perturbation matrix and the unperturbed covariance matrix. By observing the special properties of the perturbation matrix, a simple compensating algorithm together with a perturbation parameter estimation method are proposed. Promising results are obtained from simulations by which the algorithms successfully compensate the random errors and give unbiased estimates. Moreover, by comparing the theoretical and simulated CRBs, we have shown that our methods are quite efficient. Finally, to further generalize the idea, we have applied the method to blind deconvolution with slight modifications and excellent results are obtained. This proves that our algorithms are indeed very useful.

The major contribution of the novel idea is that instead of using the impractical additive Gaussian error model, a more realistic one is used for analyzing the random perturbations. Moreover, computationally effective algorithms are devised to find and correct the errors by which high dimensional parameter search (used in the method like Maximum Likelihood, etc.) is not required. This saves a lot of time, especially when there are more than four emitting sources. Although not shown in the thesis, the compensating algorithm is still valid when the emitting sources are completely coherent. This special characteristics together with the robustness of the algorithms
allow us to apply the idea to many other fields such as blind deconvolution and channel identification, etc.

There is certainly more research to be done into the theoretical as well as practical aspects of our algorithms. Some of the key issues that are unresolved include the case of unequal perturbation parameter for each source. Under such situation, the perturbed covariance matrix will have a more complicated form and the compensating algorithms need modifications. Furthermore, as the novel idea is quite general, we may apply it to other subspace fitting methods such as MD-MUSIC, WSF ESPRIT, etc. although the mathematical formulation is quite involved. The effect of division by zero is yet another area for further investigation because we have shown that there exists an optimal number of sensors (see Chapter 5). Though we have utilized theorem T4.3 to prove our estimation algorithm, it applies to p.d. signal covariance matrix $S$ and Vandermonde steering matrix $A$ only. Intuitively, the theorem would also be true for general $S$ and $A$. Due to lack of time, general cases are not considered in this thesis. We also note that this lemma may not only be useful for our purpose, but also of importance to other signal processing problems. Thus besides using probabilistic approach, a more vigorous mathematical proof may be devised though it is a very very difficult task.

Finally, though the blind deconvolution results obtained in this thesis are quite good, several problems concerning the algorithm do exist. For example, the threshold value used to avoid division by zero is found by trial and error and this wastes time. A much better method may be devised to find this threshold. In addition, we have assumed that the responses are the same for each impulse. This may not be the case in some situations such as ECG analysis. Theoretically, a generalization can be done in these cases. In fact, the field of blind deconvolution is itself a very big research topic and thus it is worthwhile to investigate the application of array signal processing techniques to this field.

100
APPENDIX A

Proof of Theorem T4.1

By (4.6)

\[ \bar{x}(l) = (A \odot \delta A(l))s(l) + n(l), \quad (A.1) \]

and

\[
R_{\bar{x}\bar{x}} = E\{\bar{x}\bar{x}^*\} \\
= E\{[(A \odot \delta A)s + n][(A \odot \delta A)s + n]^*\} \\
= E\{(A \odot \delta A)ss^*(A \odot \delta A)^*\} + \sigma^2 I. \quad (A.2)
\]

The last equality follows from assumptions listed in Chapter 2. By using (2.28) and (4.7), (A.2) is written as

\[
R_{\bar{x}\bar{x}} = E\{\sum_{i=1}^d s_i [a(\theta_i) \odot \delta a(\theta_i)] \sum_{j=1}^d s_j^*[a(\theta_j) \odot \delta a(\theta_j)]^*\} \\
= \sum_{i=1}^d \sum_{j=1}^d s_{ij} E\{[a(\theta_i) \odot \delta a(\theta_i)] [a(\theta_j) \odot \delta a(\theta_j)]^*\}, \quad (A.3)
\]

where \(s_{ij} = E\{s_i s_j^*\}\). Note that

\[
E\{[a(\theta_i) \odot \delta a(\theta_i)] [a(\theta_j) \odot \delta a(\theta_j)]^*\} \\
= E\{\text{diag}(a(\theta_i)) [\delta a(\theta_i) \delta a^*(\theta_j)] \text{diag}(a^*(\theta_j))\} \\
= \text{diag}(a(\theta_i)) E\{[\delta a(\theta_i) \delta a^*(\theta_j)] \text{diag}(a^*(\theta_j))\} \\
= \text{diag}(a(\theta_i)) V \text{diag}(a^*(\theta_j)), \quad (A.4)
\]
where $V = \{\mathbf{\delta a(\theta_i)\delta a^*(\theta_j)}\}$. Now by (4.6) and the assumption that $\mathbf{\delta \theta} \cdot \mathbf{\sin \theta}$ are i.i.d. for all $i$, the $k/l$th element of $V$ is given by

$$V(k,l) = \mathbb{E}\{e^{j\pi \mathbf{\delta \theta} \cdot \mathbf{\sin \theta}}\}.$$  \hspace{1cm} (A.5)

Hence the resultant covariance matrix can be expressed as

$$R_{\mathbf{x}x} = \sum_{i=1}^{d} \sum_{j=1}^{d} s_{ij} [\mathbf{a(\theta_i)\mathbf{a}^*(\theta_j)}] \odot V + \sigma^2 \mathbf{I}$$

$$= \left[\sum_{i=1}^{d} \sum_{j=1}^{d} s_{ij} \mathbf{a(\theta_i)\mathbf{a}^*(\theta_j)}\right] \odot V + \sigma^2 \mathbf{I}$$

$$= (\mathbf{A} \mathbf{S} \mathbf{A}^*) \odot V + \sigma^2 \mathbf{I},$$ \hspace{1cm} (A.6)

which is the result of Theorem T4.1.
APPENDIX B

Derivation of Perturbation Matrix for Uniform Distributed Errors

If $\delta\theta \cdot \sin\theta \sim \mathcal{U}[-L, L]$, then for any constant $k$

$$E\{e^{jk\delta\theta\sin\theta}\} = \frac{1}{2L} \int_{-L}^{L} e^{jk\delta\theta\sin\theta} d(\delta\theta \cdot \sin\theta)$$

$$= \frac{1}{2L} \cdot \frac{1}{j k} \left[ e^{jkL} - e^{-jkL} \right]$$

$$= \frac{\sin kL}{kL}.$$  \hspace{1cm} (B.1)

and hence

$$V(k, l) = E\{e^{j\pi d_a (k-l) \delta\theta\sin\theta}\}$$

$$= \frac{\sin[(k-l)\pi d_a L]}{(k-l)\pi d_a L},$$  \hspace{1cm} (B.2)

which is the result as shown in (4.11).
APPENDIX C

Derivation of Perturbation Matrix for Gaussian Distributed Errors

If $\delta \theta \cdot \sin \theta \sim \mathcal{N}[0, \sigma_{\theta}]$, then for any constant $k$

$$
E \{e^{jk \delta \theta \sin \theta} \} = \frac{1}{\sqrt{2\pi \sigma_{\theta}}} \int_{-\infty}^{\infty} e^{jk \delta \theta \sin \theta} e^{-\frac{\delta \theta \sin \theta}{2\sigma_{\theta}^2}} d(\delta \theta \cdot \sin \theta)
$$

$$
= \frac{1}{\sqrt{2\pi \sigma_{\theta}}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma_{\theta}^2} (\delta \theta \sin \theta - jk \sigma_{\theta}^2)^2 - \frac{1}{2\sigma_{\theta}^2} k^2 \sigma_{\theta}^4} d(\delta \theta \cdot \sin \theta)
$$

$$
= e^{-\frac{k^2 \sigma_{\theta}^2}{2}}, \quad (C.1)
$$

where the last equality follows from standard formula [19]. Hence

$$
V(k, l) = E \{e^{j\pi \delta_u(k-l) \delta \theta \sin \theta} \}
$$

$$
= e^{-\frac{(k-l)^2 \tau_{\theta} \sigma_{\theta}^2}{2}}. \quad (C.2)
$$

which is the result as shown in (4.12).
APPENDIX D

Derivation of the Error Vector

In [31], Swindlehurst has derived an expression for the error vector as (equation (13) of [31])

\[ \xi_k = a(\theta_k) \odot \delta a(\theta_k, \delta \theta_k) - a(\theta_k) + f(\Delta R, \Delta \Sigma), \]  

(D.1)

where \( f(\Delta R, \Delta \Sigma) \) is a vector function of the matrix \( \Delta R \) and \( \Delta \Sigma \) which represent the errors due to finite sample size effect and perturbed noise covariance matrix respectively. In our analysis, it is assumed that the data covariance matrix is error-free and the noise covariance matrix is exact. Therefore these two quantities are zero and so is \( f(\Delta R, \Delta \Sigma) \) and hence the error vector \( \xi \) is given by (D.1).
APPENDIX E

Proof of Theorem T4.3

We will prove the theorem from a probabilistic point of view. Suppose with the notation adopted in theorem T4.3, for a fixed value of \( \hat{\theta} \), there exists a real positive \( \mu \) such that

\[
R \odot P(\hat{\theta}, \sigma) = Q(\hat{\theta}, \sigma) + \mu I,
\]

and

\[
\rho(Q(\hat{\theta}, \sigma)) \leq d.
\]

By considering the smallest eigenvalues of the matrix \( R \odot P(\hat{\theta}, \sigma) \), it can be shown that for a fixed value of \( \hat{\theta} \), the corresponding \( \mu \) must be unique. Assume \( m = d + 3 \) (the proof for \( m > d + 3 \) is trivial if the theorem holds for \( m = d + 3 \)). Rewrite (E.1) as

\[
Q(\hat{\theta}, \sigma) = R \odot P(\hat{\theta}, \sigma) - \mu I,
\]

and let \( Q_k \) be the \( k \)th principal matrix of \( Q(\hat{\theta}, \sigma) \). Since \( Q(\hat{\theta}, \sigma) \) is of rank less than or equal to \( d \), then

\[
det(Q_{d+1}) = 0,
\]

\[
det(Q_{d+2}) = 0,
\]

\[
det(Q_{d+3}) = 0,
\]

Thus for fixed values of \( \theta \)'s, \( S \), \( \sigma \), the values of \( \mu \) and \( \hat{\theta} \) which satisfy (E.1) and (E.2) must satisfy (E.4), (E.5) and (E.6) simultaneously. Since this is a system of
three equations with only two variables, the probability that they are consistent is very low. Moreover, since the five equations (E.1), (E.2), (E.4), (E.5) and (E.6) are nonlinear, the function relating \( \mu \) to \( \hat{\sigma}_\theta \) may not be continuous in the \( \mu \times \hat{\sigma}_\theta \) domain. That is to say, there is great probability that the points \((\mu, \hat{\sigma}_\theta)\) which satisfy all the five equations are discrete in the \( \mu \times \hat{\sigma}_\theta \) domain. Since the probability of occurrence for an event corresponds to a discrete point in a two-dimensional probability sample space is zero [19], it can be concluded that these five equations are satisfied with probability zero for any fixed values of \( \theta \)'s, \( S \), \( \sigma_\theta \). This is equivalent to saying that if (E.1) is true with a positive real \( \mu \), then

\[
\rho(Q(\hat{\sigma}_\theta, \sigma_\theta)) \geq d + 1,
\]

is true with a probability 1. Hence by probabilistic arguments, the theorem is proved.
APPENDIX F

Proof of Theorem T4.4

If \( P(\hat{\sigma}_\theta, \sigma_\theta) \) is positive definite, then obviously \( P^\Omega(\hat{\sigma}_\theta, \sigma_\theta) \) is indefinite and hence we need to concentrate on the other direction only. Suppose \( P(\hat{\sigma}_\theta, \sigma_\theta) \) is indefinite, as it is in the form as shown in (4.29), we can assert that \( (\sigma^2_\theta - \hat{\sigma}^2_\theta) \) is not positive. Suppose it is positive, then we may write \( (\sigma^2_\theta - \hat{\sigma}^2_\theta) \) as \( k^2 \) and express \( P(\hat{\sigma}_\theta, \sigma_\theta) \) in a form as (4.12) which means that \( P(\hat{\sigma}_\theta, \sigma_\theta) \) is a perturbation matrix with gaussian distributed errors of zero mean and standard deviation proportional to \( k \). This implies that \( P(\hat{\sigma}_\theta, \sigma_\theta) \) is positive definite (note that all the perturbation matrices are positive definite), contradict with the fact that it is indefinite. Hence \( (\sigma^2_\theta - \hat{\sigma}^2_\theta) \leq 0 \). Since it is not positive, we can define \( K^2 = -(\sigma^2_\theta - \hat{\sigma}^2_\theta) \), where \( K \) is a positive real constant. By this substitution, \( P(\hat{\sigma}_\theta, \sigma_\theta) \) can be rewritten as

\[
P^\Omega(\hat{\sigma}_\theta, \sigma_\theta) = \begin{bmatrix}
1 & e^{k_s(\sigma^2_\theta - \hat{\sigma}^2_\theta)} & \cdots & e^{(m-1)k_s(\sigma^2_\theta - \hat{\sigma}^2_\theta)} \\
e^{k_s(\sigma^2_\theta - \hat{\sigma}^2_\theta)} & 1 & \cdots & e^{(m-2)k_s(\sigma^2_\theta - \hat{\sigma}^2_\theta)} \\
\vdots & \vdots & \ddots & \vdots \\
e^{(m-1)k_s(\sigma^2_\theta - \hat{\sigma}^2_\theta)} & \cdots & \cdots & 1
\end{bmatrix}
\]

\[
= \begin{bmatrix}
1 & e^{-k_sK^2} & \cdots & e^{-(m-1)k_sK^2} \\
e^{-k_sK^2} & 1 & \cdots & e^{-(m-2)k_sK^2} \\
\vdots & \vdots & \ddots & \vdots \\
e^{-(m-1)k_sK^2} & \cdots & \cdots & 1
\end{bmatrix} \quad (F.1)
\]

From (F.1), \( P^\Omega(\hat{\sigma}_\theta, \sigma_\theta) \) is expressed in the form of a perturbation matrix with gaussian distributed perturbations of zero mean and covariance proportional to \( K^2 \) ((4.12). Thus \( P^\Omega(\hat{\sigma}_\theta, \sigma_\theta) \) must be positive definite and this proves the theorem.
APPENDIX G

Derivation of Covariance Matrix for Blind Deconvolution

For the sake of explanation, (6.20) and (6.21) are restated below:

\[ y_i = GM_i s_i + n_i, \]  \hspace{1cm} (G.1)

where \( M_i \) is given by

\[
M_i = \begin{bmatrix}
\phi_1^0 & \cdots & \phi_d^0 \\
\phi_1^4 & \cdots & \phi_d^4 \\
\vdots & \ddots & \vdots \\
\phi_1^{4m-1} & \cdots & \phi_d^{4m-1}
\end{bmatrix} \odot \begin{bmatrix}
\Theta_{i1}^0 & \cdots & \Theta_{id}^0 \\
\Theta_{i1}^1 & \cdots & \Theta_{id}^1 \\
\vdots & \ddots & \vdots \\
\Theta_{i1}^{4m-1} & \cdots & \Theta_{id}^{4m-1}
\end{bmatrix}
\]

\[ = M \odot P_i, \]  \hspace{1cm} (G.2)

with \( \Theta_{ik} = \exp^{-j \frac{2\pi}{4m} \delta_{ik}} \)

After permuting the odd and even portions of the data vector \( y_i \), we get

\[
x_i = \begin{bmatrix}
y_{oi} \\
y_{ei}
\end{bmatrix} = \begin{bmatrix}
G_0 M_{oi} s_i \\
G_e M_{ei} s_i
\end{bmatrix} + \begin{bmatrix}
n_{oi} \\
n_{ei}
\end{bmatrix}
\]

\[ = \begin{bmatrix}
G_0 M_0 \odot P_{oi} s_i \\
G_e M_e \odot P_{ei} s_i
\end{bmatrix} + \begin{bmatrix}
n_{oi} \\
n_{ei}
\end{bmatrix} \]  \hspace{1cm} (G.3)
\[ \begin{bmatrix} G_o M_o \odot P_{oi} s_i \\ G_o M_o \odot P_{ei} s_i \end{bmatrix} + \begin{bmatrix} n_{oi} \\ n_{ei} \end{bmatrix}, \quad \text{(G.4)} \]

where

\[ P_{oi} = \begin{bmatrix} \Theta^1_{i1} & \cdots & \Theta^1_{id} \\ \Theta^3_{i1} & \cdots & \Theta^3_{id} \\ \vdots & \ddots & \vdots \\ \Theta^{4m-1}_{i1} & \cdots & \Theta^{4m-1}_{id} \end{bmatrix}, \quad \text{(G.5)} \]

\[ P_{ei} = \begin{bmatrix} \Theta^0_{i1} & \cdots & \Theta^0_{id} \\ \Theta^2_{i1} & \cdots & \Theta^2_{id} \\ \vdots & \ddots & \vdots \\ \Theta^{4m-2}_{i1} & \cdots & \Theta^{4m-2}_{id} \end{bmatrix}, \quad \text{(G.6)} \]

Thus if we assume \( \theta_{ik} \) to be i.i.d. and treat \( \Theta_{ik} \) be the outcomes of the random variable \( \Theta \), then by following similar procedures in Appendix A, the covariance matrix \( R_{xx} \) is given by

\[ R_{xx} = (A_o S A_o^\ast) \odot V + \sigma^2 I, \quad \text{(G.7)} \]

where

\[ V = E \left\{ \begin{bmatrix} \Theta^1 \\ \vdots \\ \Theta^{4m-1} \\ \Theta^0 \\ \vdots \\ \Theta^{4m-2} \end{bmatrix} \begin{bmatrix} \Theta^{-1} & \cdots & \Theta^{-(4m-1)} & \Theta^0 & \cdots & \Theta^{-(4m-2)} \end{bmatrix} \right\} \]

\[ = \begin{bmatrix} V_1 & V_3 \\ V_2 & V_1 \end{bmatrix}, \quad \text{(G.8)} \]
with

\[ V_1(i, j) = E\{ \Theta^{2(i-j)} \} , \]  

\[ V_2(i, j) = E\{ \Theta^{2(i-1)-(2j-1)} \} \]
\[ = E\{ \Theta^{2i-2-2j+1} \} \]
\[ = E\{ \Theta^{2(i-j)-1} \} , \]  

(G.9)

(G.10)

\[ V_3(i, j) = E\{ \Theta^{2(i-1)-2(j-1)} \} \]
\[ = E\{ \Theta^{2i-1-2j+2} \} \]
\[ = E\{ \Theta^{2(i-j)+1} \} . \]  

(G.11)
Biobliography


113


