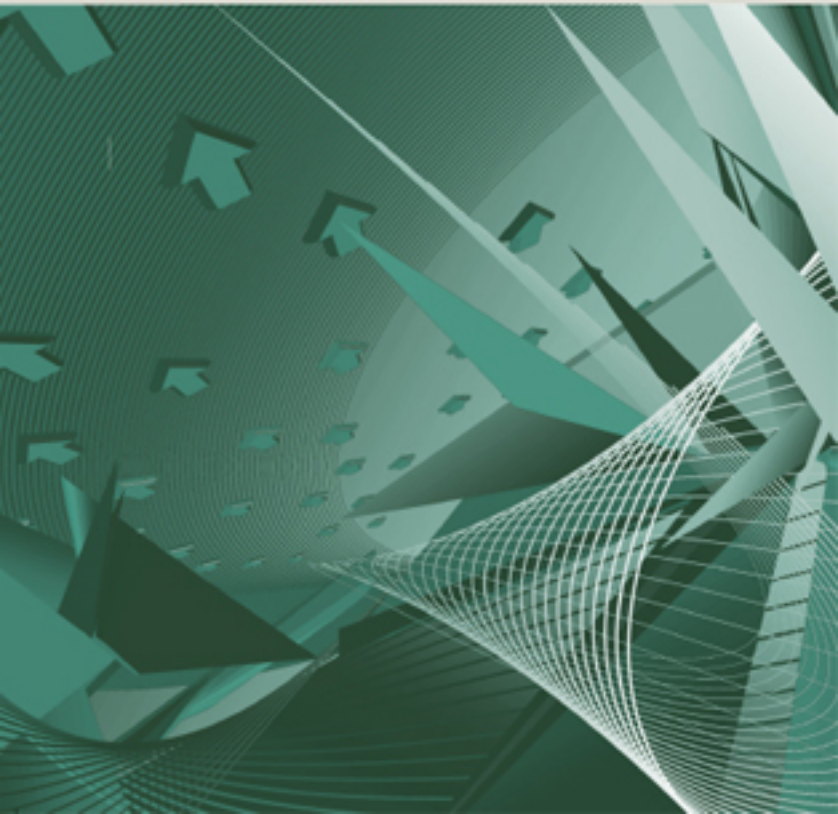
INTRODUCTION TO

Direction-of-Arrival Estimation

Zhizhang Chen
Gopal Gokeda
Yiqiang Yu



Introduction to Direction-of-Arrival Estimation

For a listing of related titles in the
Artech House Signal Processing Library
turn to the back of this book.

Introduction to Direction-of-Arrival Estimation

Zhizhang Chen
Gopal Gokeda
Yiqiang Yu

ARTECH
HOUSE
BOSTON | LONDON
artechhouse.com

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the U.S. Library of Congress.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library.

Cover design by Vicki Kane

ISBN 13: 978-1-59693-089-6

685 Canton Street
Norwood, MA 02062

10 9 8 7 6 5 4 3 2 1

Contents

Preface

Direction-of-arrival (DOA) estimation (or direction finding) essentially concerns the estimation of direction-of-arrival of signals, either in the form of electromagnetic (i.e., radio) or acoustic waves, impinging on a sensor or antenna array. The requirement for DOA estimation arises from the needs of locating and tracking signal sources in both civilian and military applications, such as search and rescue, law enforcement, sonar, seismology, and wireless 911 emergency call locating.

Various theories and techniques have been developed for array signal processing related to DOA estimations. A large body of literature has also existed on the subject. However, during our course of research and development for real-world implementations, we have found that relevant publications have been quite scattered, making it hard for the beginners or students who want to enter the area in a relatively short time. In other words, few review books are available that systematically describe the principles and basic techniques of DOA estimation under one roof.

This book is intended to cover the issue by providing an overview and performance analysis of the basic DOA algorithms and comparisons among themselves. In particular, systematic descriptions, performance analysis, and comparisons of various DOA algorithms are presented with a final focus on the family of ESPRIT (estimation of signal parameters via rotational invariance techniques).

This book is aimed at beginners such as graduate students or engineers or government regulators who need to gain insight into the fundamentals of DOA estimations in a relatively quick manner. It is also

suitable for those who specialize in the area but would like to refresh their knowledge of the basics of DOA estimations. It is our hope that this book can present sufficient information on theoretical foundations of DOA estimation techniques to a reader so that he or she can understand DOA basics and move on to advanced DOA topics if he or she wishes.

1

Introduction

Wireless technology applications have spread into many areas, such as environmental monitoring, sensor networks, public security, and search and rescues. In light of these developments, many technological policies have been established to accommodate the needs of various demands. For instance, a mandatory rule was passed by the FCC [1] that requires 125-m location accuracy on wireless emergency calls. As well, search and rescue always require the location of electromagnetic beacon sources. All these applications perhaps can be counted as the main reason for the recent increased interest in determining the direction of arrival (DOA) of radio signals in wireless systems. In fact, estimating the direction of arrival of several radio signals impinging on an array of sensors is required in a variety of other applications as well, including radar, sonar, and seismology. Another technology that has become equally glamorous is smart antenna technology [2, 3]. In smart antenna technology, a DOA estimation algorithm is usually incorporated to develop systems that provide accurate location information for wireless services [4].

A smart antenna, for this book discussion, is a system that combines multiple antenna elements with a signal processing capability to optimize its radiation and/or reception pattern automatically in response to the system's signal environment. This technology is particularly found useful in mobile communications in lieu of an increasing number of mobile subscribers and limited resources. Smart antennas can be used to enhance the coverage through range extension and to increase system capacity [2, 3]. Smart antennas can also be used to spatially separate signals, allowing

different subscribers to share the same spectral resources, provided that they are spatially separable at the base station. This spatial division multiple access (SDMA) method allows multiple users to operate in the same cell and on the same frequency/time slot provided by utilizing the adaptive beamforming techniques of the smart antennas. Since this approach allows more users to be supported within a limited spectrum allocation, compared with conventional antennas, SDMA can lead to improved capacity.

The smart antenna technology can be divided into three major categories depending on their choice in transmit strategy:

- *Switched lobe (SL):* This is the simplest technique and comprises only a basic switching function between predefined beams of an array. When a signal is received, the setting that gives the best performance, usually in terms of received power, is chosen for the system to operate with.
- *Dynamically with phased array (PA):* This technique allows continuous tracking of signal sources by including a direction-of-arrival (DOA) finding algorithm in the system; as a result, the transmission from the array can be controlled intelligently based on the DOA information of the array. The PA technique can be viewed as a generalization of the switched lobe concept.
- *Adaptive array (AA):* In this case, a DOA algorithm for determining the directions of interference sources (e.g., other users) is also incorporated in addition to finding the DOA of the desired source. The beam pattern can then be adjusted to null out the interferers while maximizing the transmit power at the desired source.

The importance of DOA estimation for smart antenna can be understood by studying the architecture of smart antenna as described in the following section.

1.1 Smart Antenna Architecture

Typical smart antenna architectures for a base station can be divided into following functional blocks, as shown in Figures 1.1 and 1.2:

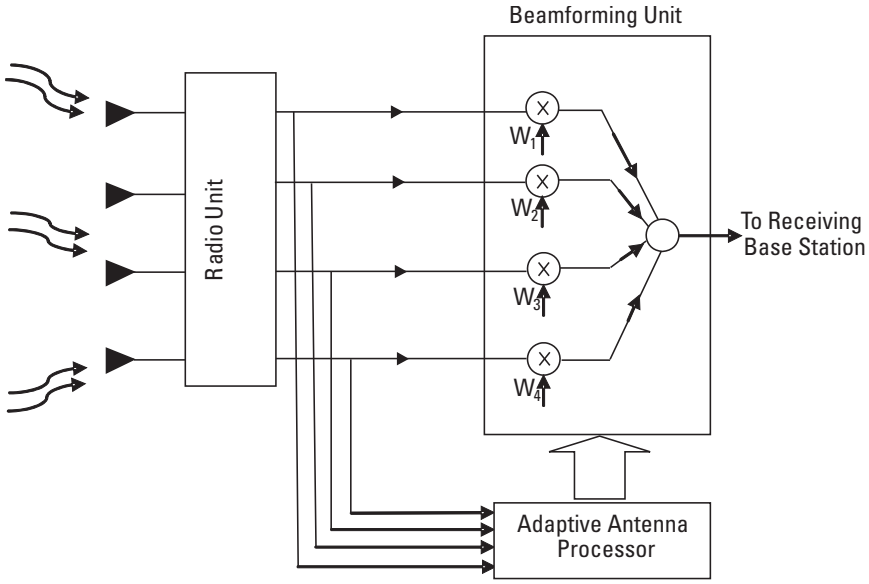


Figure 1.1 Smart antenna receiver.

- *Radio unit:* This unit mainly consists of: (1) antenna arrays that intercept radio frequency (RF) signals from the air, (2) downconversion chains that remove the carrier(s) of the RF signals received by the antenna array, and (3) analog-to-digital converters that convert the no-carrier signals to the corresponding digital signals for further processing. Antenna arrays can be one-, two-, or even three-dimensional, depending on the dimension of the space one wants to access. The radiation pattern of the array depends on the element type, the relative positions, and the excitation (amplitude and phase) to each element [5].

- *Beamforming unit:* The beamforming unit is responsible for forming and steering the beam in the desired direction. In it, the weighting of the received (or transmitted) signals is applied. Basically, the data signals $x_{k,\ k=0,\ \ldots,\ M-1}$ received by an M-element array are directly multiplied by a set of weights to form a beam at a desired angle. In other words, by multiplying the data signals with appropriate sets of weights, it is possible to form a set of beams with pointing angles directed at the desired angles, resulting in a signal peak at the output of a beamformer. Mathematically it can be expressed as:

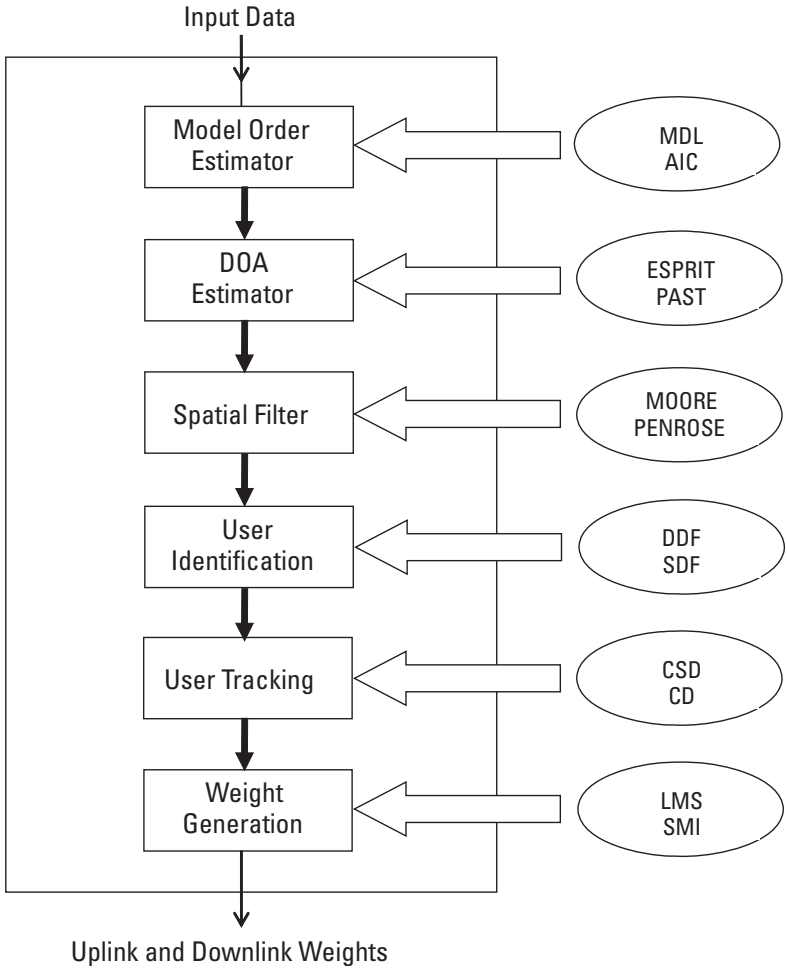


Figure 1.2 Adaptive antenna processor.

$$y(\theta_i) = \sum_{k=0}^{M-1} w_k^i x_k \tag{1.1}$$

where $y(\theta_i)$ is the output of a beamformer, x_k is the data sample from the kth array element and w_k^i is the weight for forming a beam or null at angle θ_i. More descriptions on (1.1) are given in Chapters 2 and 3.

By selecting the appropriate values for the set of the weights $w_{k\ k=0,\ldots,M-1}^i$, one can implement beam steering, adaptive nulling,

and beamshaping. These weights that determine the radiation pattern are generated by the adaptive processor unit.

- *Adaptive antenna processor:* The function of the adaptive processor unit is to determine the complex weights for the beamforming unit. The weights can be optimized from two main types of criteria: maximization of the data signal from the desired source (e.g., switched lobe or phased array) or maximization of the signal-to-interference ratio (SIR) by suppressing the signal from the interference sources (adaptive array). In theory with M antenna elements, one can "null out" $M-1$ interference sources, but due to multipath propagation, this number will be normally lower. The method for calculating the weights will differ depending on the types of optimization criteria. When a switched lobe is used, the receiver will test all the predefined weights and choose the one that gives the strongest received data signal. However, if the phased array or adaptive array approach is used, which consists of directing a maximum gain towards the strongest signal component, the directions of arrival (DOAs) of the signals are first estimated and the weights are then calculated in accordance with the desired steering angle.

In general, as shown in Figure 1.2, the adaptive antenna processor consists of several computation processes:

- *Model order estimator:* From the input data $x_{k,\ k=0,\ \ldots,\ M-1}$ received by the antenna elements, the number of wavefronts impinging on the array is estimated using model order estimation algorithms, such as AIC or MDL, which will be described in Chapter 4. The knowledge of number of the signals impinging on the array is crucial to DOA estimation algorithms; hence, these algorithms are run prior to DOA estimation algorithms.
- *DOA estimator:* This forms the vital stage of the adaptive antenna processor where algorithms like MUSIC or ESPRIT are used for estimating the direction of arrival of all the signals impinging on the array. This stage gives DOAs of all the relevant signals of the user sources and other interference sources. To make the process faster, instead of estimating the signal space every time, subspace tracking algorithms like dPAST and PAST (Projection

Approximation Subspace Tracking) are used to recursively track the signal subspace. Usually, the signal subspace is only slowly time-varying. It is therefore more efficient to track those changes than to perform full subspace estimation. The DOAs can then be estimated faster from these signal subspaces. Detailed descriptions of these DOA algorithms are presented in Chapter 3, 4, and 5.

- *Spatial filter:* After the DOAs of all the signals impinging on the array are obtained, the signals are filtered by reconstructing the signals for each of the DOAs estimated. Estimating the signals from the estimated DOAs is usually called signal reconstruction or signal copy. With the knowledge of DOAs, the corresponding steering vectors **a** and eventually the estimated steering matrix **A** are constructed. The signal is then reconstructed from

$$\mathbf{S} = \mathbf{W}\mathbf{X}$$

where **S** is the matrix of impinging signals extracted from the noise corrupted signals (or data signals) received by an array **X** and the weighting matrix **W** is chosen to be the Moore-Penrose pseudo inverse of the estimated array steering matrix **A**. More details are given in Chapter 3.

- *User identification:* Once the signals are separated with respect to their distinct DOAs, the desired user corresponding to these DOAs needs to be identified. By comparing the received mid-ambles (training sequences) with the desired user mid-amble, the number of bit errors within the training sequence can be calculated. A spatially resolved wavefront and thus the corresponding DOA are attributed to a user, when the number of bit errors is smaller than a threshold. In this way not only a single user path but also all paths that correspond to the intended user can be identified. The DOA of the user path with the strongest instantaneous power is then detected. As a training sequence detector, standard sequence estimators like delayed decision feedback (DDF) and soft decision feedback (SDF) are applied [6].
- *User tracking:* A fast reactive adaptive estimator is needed for tracking changes of signal parameters. The tracker does not only prevent far-off estimates from disturbing the beamforming, but also prevents the DOA estimates from changing too much

between two consecutive signal bursts. This is reasonable, since a mobile user in real time does not move far during one frame of observation. Hence, the variation in DOA is gradual. The approach to user tracking is based on recursive formulation for tracking DOAs where small changes in the difference of correlation matrix estimates result in small changes in difference of steering matrices through which the DOAs are extracted. One of the problems that these algorithms pose is the data association problem, that is, association of DOA estimates established at the previous time instant with that of current time instant. To overcome this problem, algorithms like constrained steepest descent (CSD), steepest descent (SD), and conjugate gradient (CD) are developed and employed [7, 8].

- *Weight generation:* The tracked DOA with the strongest instantaneous power is selected and thus weights for the beamforming unit to adaptively focus the beams at a desired source can be generated. Adaptive antenna algorithms like least mean square (LMS) or sample matrix inversion (SMI) [9] are used to generate these weights. The choice of the adaptive algorithm for deriving the adaptive weights is highly important in that it determines both the speed of convergence and hardware complexity required to implement the algorithm.

1.2 Overview of This Book

DOA estimation algorithms form the heart of smart antenna systems. Across the board, there are a variety of DOA estimation algorithms for use in smart antenna systems targeted at position-location applications [10], and more are emerging [3, 11]. The challenge for a designer, however, is to choose the right DOA estimation algorithm because the most advanced smart antenna implementations involve simultaneously maximizing the useful signal and nulling out the interference sources. Even with the powerful signal processing units available today, it is a challenging task to perform this in real time. Consequently, efficiency and effectiveness of an algorithm become critical in selecting a DOA estimation algorithm. In other words, a thorough understanding of a DOA estimation algorithm is needed for a designer who will use the smart antenna technique.

This book is intended to provide an introduction to the basic concepts of DOA estimations with an overview of basic techniques. In addition, the algorithms that belong to the family of ESPRIT (estimation of signal parameters via rotational invariance techniques) will be elaborated in more detail. This is because of ESPRIT's simplicity and high-resolution capability as a signal subspace-based DOA estimation algorithm; based on it, many state-of-the-art techniques have been developed.

Currently, few books are available commercially that systematically describe the principles and basic techniques of DOA estimations. This book is intended to add to that knowledge base by providing an overview and a performance analysis of the basic DOA algorithms and comparisons among themselves. In particular, systematic study, performance analysis, and comparisons of the DOA algorithms belonging to the family of ESPRIT are provided. In short, this book lays out the theoretical foundations of DOA estimation techniques for a reader to understand DOA basics and to continue on to advanced DOA topics if he or she wishes [10, 11].

The book is aimed at beginners such as graduate students or engineers or government regulators who need to gain rapid insight into the fundamentals of DOA estimations. It is also suitable for those who specialize in the area but would like to refresh their knowledge of the basics of DOA estimations.

1.3 Notations

In this work, vectors and matrices are denoted in bold letters, either lowercase or uppercase. The superscripts $(\cdot)^H$ and $(\cdot)^T$ denote complex conjugate transposition and transposition without complex conjugation, respectively. The overbar $\overline{()}$ denotes complex conjugate. $E\{\cdot\}$ denotes expectation operator. $R^{m\times n}$ refers to a real number matrix of m rows and n columns. $C^{m\times n}$ refers to a complex number matrix of m rows and n columns. $\in$ means belong to. Frequently used exchange matrix Π_p and diagonal matrix are briefly explained in the appendix. Also, a definition of Kronecker products is provided in the appendix.

References

[1] Federal Communications Commission, http://www.fcc.gov/e911/.

[2] Liberti, Jr., J. C., and T. S. Rappaport, *Smart Antennas for Wireless Communications: IS-95 and Third Generation CDMA Applications*, Upper Saddle River, NJ: Prentice-Hall, 1999.

[3] Sarkar, T. K. *Smart Antennas*, New York: IEEE Press/Wiley-Interscience, 2003.

[4] Kuchar, A., et al., "A Robust DOA-Based Smart Antenna Processor for GSM Base Stations," *Proc. IEEE Intl. Conf. on Communications*, Vol. 1, June 6–10, 1999, pp. 11–16.

[5] Balanis, C. A., *Antenna Theory: Analysis and Design*, 3rd ed., New York: Wiley, 2005.

[6] Hallen, A. D., and C. Heegard, "Delayed Decision Feedback Sequence Estimation," *IEEE Trans. on Communications*, Vol. 37, No. 5, May 1989, pp. 428–436.

[7] Griffiths, L. J., and C. W. Jim, "An Alternative Approach to Linearly Constrained Beamforming," *IEEE Trans. on Antennas and Propagation*, Vol. AP-30, No. 1, January 1982, pp. 27–34.

[8] Karttunen, P., and R. Baghaie, "Conjugate Gradient Based Signal Subspace Mobile User Tracking," *Proc. IEEE Vehicular Technology Conference*, Vol. 2, May 16–20, 1999, pp. 1172–1176.

[9] Compton, Jr., R. T., *Adaptive Antennas*, Englewood Cliffs, NJ: Prentice-Hall, 1988.

[10] Krim, H., and M. Viberg, "Two Decades of Array Signal Processing Research," *IEEE Signal Processing Magazine*, Vol. 13, No. 4, July 1996, pp. 67–94.

[11] Van Tree, H. L., *Optimum Array Processing*, New York: John Wiley & Sons.

2

Antennas and Array Receiving System

An antenna is basically a transducer designed to transmit or receive electromagnetic waves or radio signals that propagate in air or in a medium [1]. A transmit antenna takes in currents out of electronic devices and transforms them into the corresponding electromagnetic or radio signals that radiate into air or a medium. A receive antenna does the opposite: it captures or intercepts electromagnetic waves or radio signals in the air or a medium and then transforms them into corresponding currents and sends them to the connected electronic devices. In short, an antenna serves as an interfacing device between air (or a medium) and electronic devices, making possible the realization of a wireless system as shown in Figure 2.1. They are often seen and used in systems such as radio broadcasting, radio communications, cellular phone systems, wireless local area networks (LAN), radar, and space exploration. Without antennas, there would be no modern wireless communications and telemetry industry.

Physically, an antenna is an arrangement of conductors and surrounding materials that will either generate radiating electromagnetic waves in response to currents or voltages applied to the antennas or induce currents and voltages in it due to electromagnetic waves or radio signals impinging on it. Different ways of the arrangements or positioning of the conductors and materials lead to various antennas with different behaviors and properties. Figure 2.2 shows a monopole antenna

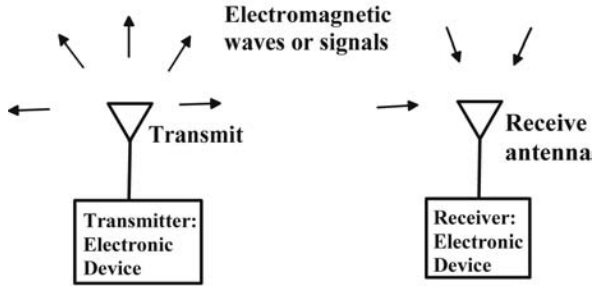


Figure 2.1 Wireless transmission via antennas.

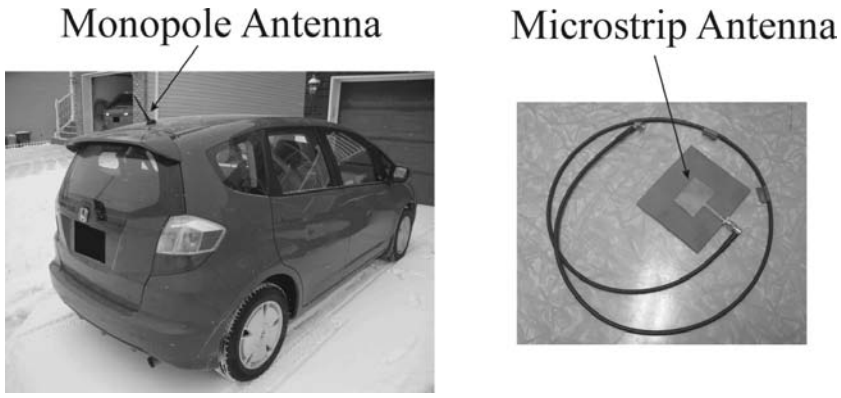


Figure 2.2 A monopole antenna mounted on a car and a microstrip patch antenna.

mounted on a car for receiving radio broadcasting and a microstrip patch antenna seen in some cellular phones or wireless USBs.

In theory, both transmit and receive antennas need to be studied separately because of their different functions. Fortunately, it has been found that transmit and receive functions of an antenna are reciprocal. Therefore, in most literature so far, only transmission properties of an antenna are studied and analyzed; when an antenna is used for receiving, its receiving properties are simply extracted from its transmitting properties through the reciprocity.

Various antenna parameters have been introduced to quantify the performance of an antenna. In this chapter, we will describe the important parameters and concepts that are pertinent to the DOA estimations.

We will first discuss a single transmit antenna, a single receive antenna, and then antenna arrays that consist of many single antennas and are used for DOA estimations.

2.1 Single Transmit Antenna

It has been found and proven that the following properties exist for a transmit antenna:

1. The electromagnetic energy radiated out of an antenna is often not uniformly distributed in space. Radiation intensity is normally stronger in one direction than in other directions.
2. Not all the frequencies of radio signals are radiated effectively out of an antenna. Some frequencies are transmitted into air or a medium very efficiently and others not at all.

To describe these two phenomena, the following parameters have been introduced for an antenna:

1. Directivity and gain that describe the degree of energy concentration in a direction by an antenna;
2. Radiation pattern that shows the relative radiation intensities in all the directions;
3. Equivalent resonant circuit and bandwidth that shows how an antenna behaves in terms of frequency responses.

2.1.1 Directivity and Gain

These two parameters basically quantify how well an antenna can concentrate its radiated energy in a specific direction. To quantify, a reference antenna needs to be taken. In most cases, an omnidirectional antenna that radiates equally in all directions is taken as the reference antenna as shown in Figure 2.3.

The directivity of an antenna is then defined as the ratio of the radiated power intensity of an antenna in one direction to that of the omnidirectional antenna in the same direction when both antennas radiate the same total power. In other words, the directivity of an antenna measures the amount of power intensity by an antenna over that by the

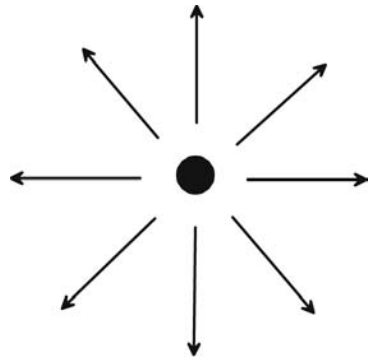


Figure 2.3 The omnidirectional antenna and its radiation distribution.

reference antenna (the omnidirectional isotropic antenna) in a given direction at an arbitrary distance. If the directivity is unity in all the direction, the antenna is an omnidirectional antenna.

Strictly speaking, the directivity of an antenna depends on the directions at which the radiation intensity is measured and compared with that of the reference antenna. However, in a normal circumstance, the term directivity refers to the maximum directivity among all the directions by an antenna.

The gain of an antenna is directly related to its directivity. The gain is equal to the product of the directivity and the antenna efficiency. The efficiency accounts for internal power ohmic loss of the antenna; it is equal to the ratio of the total power radiated into air or a medium by an antenna to the power input to the antenna by the electronic device connected to the antenna (see Figure 2.4). The efficiency is less than or equal to 100%. As a result, the gain is less than or equal to the directivity.

2.1.2 Radiation Pattern

Radiation pattern of an antenna is the geometric pattern of the relative strengths of the field emitted by the antenna in respect to the radiation directions. For the ideal isotropic omnidirectional antenna, this would be a sphere, meaning that the radiation intensity generated is the same in every direction. For a typical dipole, however, this would be a toroid or "donut." Figure 2.5 shows the patterns of two antennas, a dipole and a waveguide horn. In the radiation pattern, the longer radius of the pattern represents the stronger radiation intensity. By examining a radiation

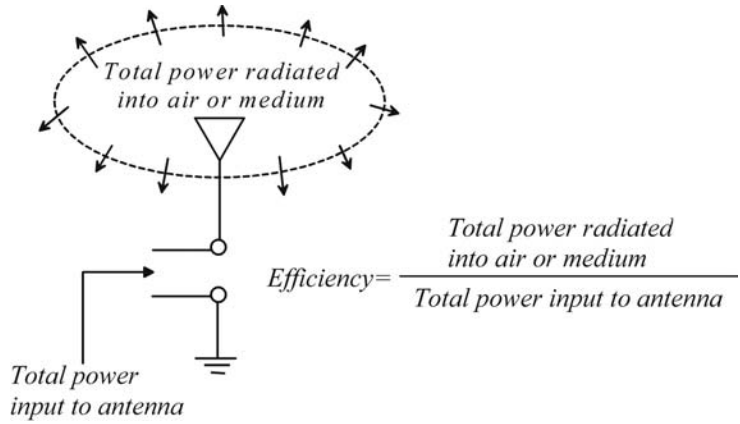


Figure 2.4 Power relationship in a transmit antenna.

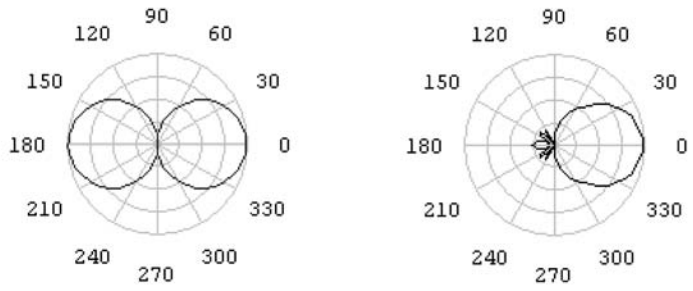


Figure 2.5 Radiation patterns of a dipole and a waveguide horn (open-ended waveguide).

pattern of an antenna, we are able to see in which direction an antenna radiates the most and in which direction the least.

2.1.3 Equivalent Resonant Circuits and Bandwidth

Like many field-based devices, an antenna can be made equivalent to a resonant circuit in terms of its input terminal voltage and current, as shown in Figure 2.6. In the circuit, in addition to a resistor that represents the internal ohmic power loss of an antenna, there is another resistor called a radiation resistor that describes the amount of power radiated into air or a medium. The inductor and capacitor represent the frequency-selective nature of an antenna. The input signals of the frequencies near the resonant frequency of the antenna can easily go in the

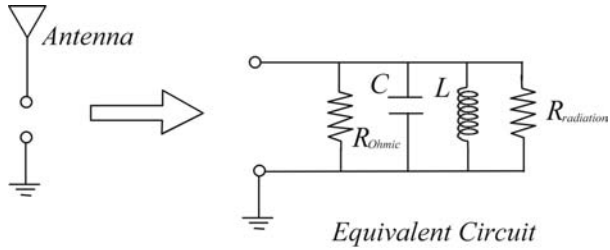


Figure 2.6 Antenna and its equivalent circuit.

radiation resistor and equivalently send their energy into the air or a medium.

A typical relationship between the input power and the frequency is shown in Figure 2.7. Like the bandwidth definition for a resonant circuit, the difference between the two frequency points at which the power drops to the half of its maximum is defined as the bandwidth of an antenna. Within the bandwidth, we consider that the degree of the power radiation is acceptable.

2.2 Single Receive Antenna

As mentioned before, a receive antenna does the opposite of a transmit antenna: it captures and intercepts electromagnetic energy in air or a medium and converts it to currents and voltages. Theoretically, a receive antenna should also have its own parameters that quantify its performance. In particular, it should have its own receive directivity, gain, and

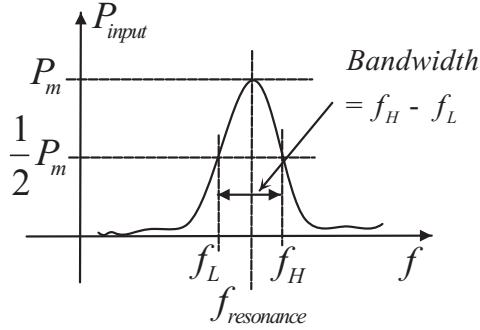


Figure 2.7 A typical relationship between the frequency and the power input to an antenna.

pattern that describe how a receive antenna captures different energy amounts of electromagnetic or radio signals coming from different directions in reference to an omnidirectional receive antenna. However, it has been found and can be proven that these parameters have the same values or shape as the transmit directivity, gain, and pattern when the same antenna is used as a transmit antenna. As a result, in almost all the antenna analysis, directivity, gain, and radiation pattern are referred to those when an antenna is used for transmission. In other words, when an antenna is to be used for receiving, its receive parameters are simply taken from those transmit parameters when the antenna is analyzed for transmission; the exception is, however, the equivalent circuit.

The equivalent circuit of a receive antenna is shown in Figure 2.8. It is exactly the same as the equivalent circuit of the antenna when it is used for transmission, except a voltage or current source is added. The current source represents the equivalent current induced by the presence of the electromagnetic fields near the receive antenna. The resistances, inductance, and capacitance remain the same as those of the transmit antenna.

2.3 Antenna Array

An antenna array is a group of antennas that are used for transmitting or receiving the same signals. Each individual antenna is often called an array element. For a receive array, the signals received by all array elements will be combined and then processed for various applications, including the DOA estimations.

For illustration purposes, let us consider a three-element that is aligned uniformly along a line as shown in Figure 2.9. The rightmost

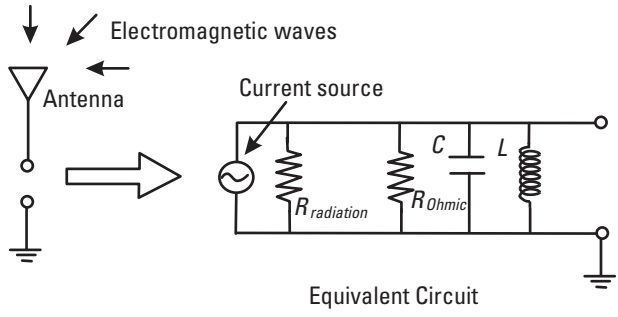


Figure 2.8 Equivalent circuit of a receive antenna.

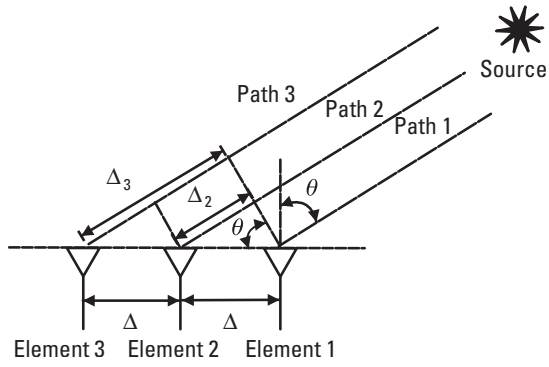


Figure 2.9 Configuration of a three-element antenna array.

element is numbered as element 1 and taken as the reference element. The other two elements are numbered as element 2 and element 3 in a sequence from right to left. The distance between two neighboring elements is Δ. A source emits electromagnetic waves at such a far distance that the three propagation line paths from the source to the three elements can be approximately considered as parallel. The waves impinge on the array at an angle of θ. As a result, the line paths from the source to elements 2 and 3 are longer than the path to element 1 (reference element) by the extra distance of

$$\Delta_m = (m-1)\Delta \cdot \sin\theta \quad {}_{m=1,2,3} \tag{2.1}$$

Now assume that the wave signal received by the reference element (i.e., element 1) is

$$x_1(t) = s(t) \tag{2.2}$$

without taking into account of noises from air. Then the signals received by element 2 and element 3 can be written as:

$$x_2(t) = s(t)e^{-j\beta\Delta_2} = s(t)e^{-j\frac{2\pi\Delta}{\lambda}\sin\theta} \tag{2.3}$$

$$x_3(t) = s(t)e^{-j\beta\Delta_3} = s(t)e^{-j2\frac{2\pi\Delta}{\lambda}\sin\theta} \tag{2.4}$$

Here $\beta = \dfrac{2\pi}{\lambda}$ is the phase shift constant of the wave propagating in air with λ being the wavelength. The phase shift term $e^{-j\beta\Delta_m}$ in (2.3) and (2.4) is the result of the signal propagating over an extra distance Δ_m in comparison with the path to the rightmost element [1].

In a more generalized way, the signals received by the three elements can be written as:

$$\mathbf{x} = \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} = \begin{bmatrix} 1 \\ e^{-j\frac{2\pi\Delta}{\lambda}\sin\theta} \\ e^{-j2\cdot\frac{2\pi\Delta}{\lambda}\sin\theta} \end{bmatrix} s(t) = \begin{bmatrix} 1 \\ e^{j\mu} \\ e^{-j2\mu} \end{bmatrix} = \mathbf{a}(\mu)s(t) \tag{2.5}$$

where $\mu = \dfrac{2\pi\Delta}{\lambda}\sin\theta$ and $\mathbf{a}(\mu) = [1\ e^{-j\mu}\ e^{-j2\mu}]^T$, which is often called the array steering vector.

Equation (2.5) can be extended to an *M*-element array. It can be rewritten as:

$$\mathbf{x} = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \cdots \\ x_3(t) \end{bmatrix} = \begin{bmatrix} 1 \\ e^{-j\frac{2\pi\Delta}{\lambda}\sin\theta} \\ \cdots \\ e^{-j(M-1)\cdot\frac{2\pi\Delta}{\lambda}\sin\theta} \end{bmatrix} s(t) = \begin{bmatrix} 1 \\ e^{-j\mu} \\ \cdots \\ e^{-j(M-1)\mu} \end{bmatrix} s(t) = \mathbf{a}(\mu)s(t) \tag{2.6}$$

and

$$\mathbf{a}(\mu) = \left[1\ e^{-j\mu}\ \ldots\ e^{-j(M-1)\mu}\right]^T$$

Although (2.6) is for a linear array, the expression $\mathbf{x} = \mathbf{a}(\mu)s(t)$ is applicable to any other arrays.

2.4 Conclusion

In this chapter, we have reviewed the basic concepts of antennas and antenna arrays pertinent to the DOA estimation algorithms to be discussed in this book. In short, an antenna is a transducer device that interfaces electronic devices with air (or a medium) for the purpose of

transmitting or receiving electromagnetic waves or radio signals. It can convert voltages/currents to electromagnetic waves and vice versa. Due to the fact that an antenna radiates its energy nonuniformly in space, parameters such as directivity, gain, and radiation pattern are often used to quantify performances of an antenna. Since an antenna is also frequency-selective, an equivalent circuit can be introduced to describe an antenna in terms of its terminal voltage and current. Because of the reciprocity of transmit and receive operations of an antenna, the performances of an antenna are often investigated for its transmit properties; the results can be directly used for an assessment of its receive performance.

There is a huge body of published literature on various topics of antennas. Such a large amount of available information usually gets a nonantenna specialist lost or confused in selecting a good text as a starting point. In order to avoid the similar confusion, we list only a single reference in this chapter. This reference is the excellent textbook by C. A. Balanis [1]; it presents an introduction to various topics of antenna technologies including the basic concepts of antennas and arrays; the contents of the book are very much sufficient for the potential readers of this book.

Reference

[1] Balanis, C. A., *Antenna Theory: Analysis and Design*, 3rd ed., New York: John Wiley & Sons, 2005.

3

Overview of Basic DOA Estimation Algorithms

3.1 Introduction

This chapter presents an overview of some of the popular algorithms for DOA estimation. In general, the direction-of-arrival (DOA) estimation techniques can be broadly classified into conventional beamforming techniques, subspace-based techniques, and maximum likelihood techniques. This chapter is organized as follows. In Section 3.1, the signal and data model used in this book is described. A uniform linear array is used to explain this model. In Section 3.2, the concept of centro-symmetric sensor arrays, which are demanded by many DOA algorithms, is discussed. Two such arrays, uniform linear array (ULA) and uniform rectangular array (URA), are described. Later, the basic principles of a few basic algorithms belonging to each DOA class are briefly discussed.

An excellent doctoral dissertation on array signal processing for DOA estimations by M. Haardt was published by Shaker Verlag [1]. It presents a comprehensive review and study on one of the DOA techniques, the ESPRIT technique. This book uses it as one of the major references. In order to facilitate a reader in checking on this major reference without much confusion, this book employs the same mathematical symbols and technical terms as those in [1]; these symbols and terms have also been used in other related literature.

3.2 Data Model

Most of the modern approaches to signal processing are model-based, in the sense that they rely on certain assumptions on the data observed in the real world. The prevailing data model used in the remainder of this book is discussed in this section.

The following scenario is assumed and maintained throughout our description of DOA estimation algorithms [2]:

- *Isotropic and linear transmission medium:* There are d sources that emit d signals; the d signals travel through a medium and then impinge onto and an M-element antenna or sensor array. The transmission medium between the sources and the array is assumed to be isotropic and linear; in other words, the medium has its physical properties the same in all different directions and the signals or waves at any particular point can be superposed linearly. The isotropic and linear property of the medium ensures: (1) that the propagation property of the waves do not change with the DOAs of signals, and (2) that the signals traveling through the medium and then impinging on or received by any element of the M-element array can be computed as a linear superposition of d signal wavefronts generated by the d sources. In addition, the gain of each antenna or sensor element is assumed to be one.

- *Far-field assumption:* The d signal sources are located far from the array such that the wavefront generated by each source arrives at all the elements at an equal direction of propagation and the wavefront is planar (far-field approximation). Thus, the propagating fields of the d signals arrived at the array are considered as parallel to each other. This assumption can, in general, be realized by making the distance between the signal sources and the array much larger than the dimension of the antenna array. A rule of thumb is that the distance is larger than $2D^2/\lambda$, with D being the dimension of the array and λ being the wavelength of the signals [3].

- *Narrowband assumption:* The d signals from the d sources have the same carrier frequency such that their frequency contents are concentrated in the vicinity of carrier frequency f_c. Then,

mathematically, any one of the instantaneous signals coming out of the sources can be expressed as

$$s_i^r(t) = \alpha_i(t)\cos\left[2\pi f_c t + \beta_i(t)\right], \qquad 1 \le i \le d \tag{3.1}$$

The signals are narrowband as long as their amplitudes $\alpha_i(t)$ and information-bearing phases $\beta_i(t)$ vary slowly with respect to τ, which is the time for the wave signals to propagate from one element to another. In other words,

$$\alpha_i(t-\tau) \approx \alpha_i(t) \text{ and } \beta_i(t-\tau) \approx \beta_i(t) \tag{3.2}$$

The slow-varying $\alpha_i(t)$ and phases $\beta_i(t)$ ensure that Fourier transform of (3.1) have most frequency contents in the neighborhood of the carrier frequency f_c. An expression convenient for mathematical analysis can be obtained by defining the complex envelope of the signal or phasor as

$$s_i^{env}(t) = \alpha_i(t)e^{j\beta_i(t)} \text{ such that } s_i^r(t) = \mathrm{Re}\left\{s_i^{env}(t)e^{j2\pi f_c t}\right\} \tag{3.3}$$

This form of complex (or analytic) signals is supported by most receivers, which decompose the received signals into both in-phase (the real part) and quadrature components (the imaginary part).

- *AWGN channel:* The noise is assumed to be of a complex white Gaussian process. The additive noise is taken from a zero mean, spatially uncorrelated random process, which is uncorrelated with the signals. The noises have a common variance σ_N^2 at all the array elements and are uncorrelated among all elements or sensors.

Figure 3.1 depicts the scenario with the above assumptions.

3.2.1 Uniform Linear Array (ULA)

Now consider a uniform linear array (ULA) consisting of M identical and omnidirectional elements that are aligned and equally spaced on a line. Let the distance between the two adjacent elements be Δ and the distance

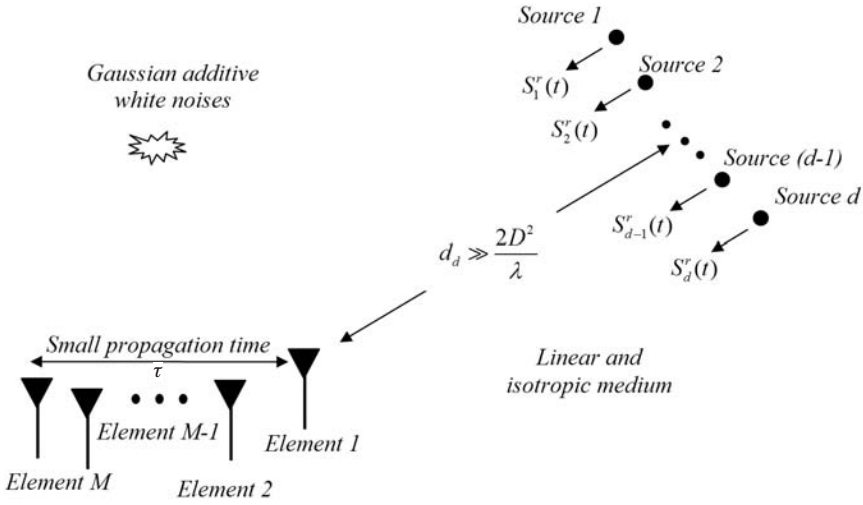


Figure 3.1 The scenario under consideration in this book.

between the source and the first (rightmost) element is d_d. The configuration is shown in Figure 3.2.

Suppose that a plane wave signal generated by source i impinges on the array at an angle θ_i and the signal generated by source i is a

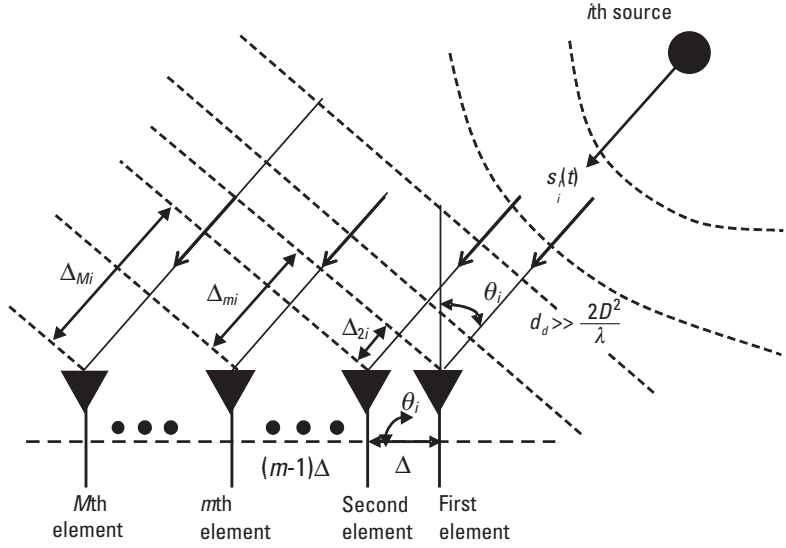


Figure 3.2 Data model for DOA estimation of d sources with a linear array of the M element.

narrowband signal $s_i^r(t)$; it then travels over a distance of d_d at a speed of c [3] and reaches the first (rightmost) element. In other words, the signal received by the rightmost element is the delayed version of the $s_i^r(t)$ with a delay of $\tau_d = \frac{d_d}{c}$. That is,

$$\begin{aligned} s_{i1}(t) &= s_i^r(t-\tau_d) = \alpha_i(t-\tau_d)\cos\left[2\pi f_c(t-\tau_d)+\beta_i(t-t_d)\right] \\ &= \mathrm{Re}\left\{s_i(t)\right\} \end{aligned} \tag{3.4}$$

with the corresponding phasor signal being $s_i(t) = \alpha_i(t-\tau_d)e^{j[2\pi f_c t+\beta_i(t-t_d)+2\pi f_c\tau_d]}$.

Because the ULA positions the array elements on a line, the signal traveling to the mth element will take an extra distance in comparison with the signal arriving at the rightmost element (see Figure 3.2). This extra distance can be found as:

$$\Delta_{mi} = (m-1)\Delta\sin\theta_i \qquad m=1,2,\ldots,M \tag{3.5}$$

Therefore, the signal arriving at the mth element will take the additional delay, which is equal to:

$$\tau_{mi} = \frac{\Delta_{mi}}{c} = (m-1)\frac{\Delta\sin\theta_i}{c} \tag{3.6}$$

The signal received by the mth element is then the delayed version of the signal $s_{i1}(t)$ (which is received by the first element) with the additional delay of τ_{mi}:

$$\begin{aligned} s_{im}(t) &= s_{i1}(t-\tau_{mi}) = s_i^r(t-\tau_d-\tau_{mi}) \\ &= \alpha_i(t-\tau_d-\tau_{mi})\cos\left[2\pi f_c(t-\tau_d-\tau_{mi})+\beta_i(t-\tau_d-\tau_{mi})\right] \\ &\approx \alpha_i(t-\tau_d)\cos\left[2\pi f_c(t-\tau_d)+\beta_i(t-t_d)-(m-1)\mu_i\right] \\ &= \mathrm{Re}\left[s_i(t)e^{j(m-1)\mu_i}\right] \end{aligned} \tag{3.7}$$

where $\mu_i = -\frac{2\pi f_c}{c}\Delta\sin\theta_i = -\frac{2\pi}{\lambda}\Delta\sin\theta_i$; it is called spatial frequency that is associated with the ith source that generates the signal of incident

angle θ_i. $\lambda = c/f_c$ denotes the wavelength corresponding to the carrier frequency f_c. In deriving (3.7), approximation (3.2) is applied.

In the complex phasor form, the above received signal corresponds to:

$$s_{im}(t) \approx \alpha_i(t - \tau_d)e^{j[2\pi f_c(t-\tau_d)+\beta_i(t)]}e^{j(m-1)\mu_i} = s_i(t)e^{j(m-1)\mu_i} \quad (3.8)$$

Equation (3.8) shows that the signal received by the mth element from the ith source is the same as that received by the first (rightmost) element but with an additional phase shift factor of $e^{j(m-1)\mu_i}$; this factor is dependent only on spatial frequency μ_i and the position of the element relative to the first element. For each incident angle θ_i that determines a source, there is a corresponding spatial frequency μ_i. Therefore, the whole objective of estimating a DOA (i.e., θ_i) is to extract this spatial frequency μ_i from the signals received by the array.

In order to be able to determine θ_i uniquely from μ_i, one-to-one correspondence is desired between them. As a result, the spatial frequencies μ_i are limited to $-\pi \le \mu_i \le \pi$ and the range of possible DOAs is restricted to the interval of $-90° \le \theta_i \le 90°$. This, in turn, requires that the element spacing satisfies $\Delta \le \lambda/2$. If the element spacing does not satisfy this relation, there is an ambiguity in DOA determination, that is, there will be two solutions for the angles from a specific value of μ_i; this will result in an array having the grating lobes: lobes other than the main lobe permits the signals from undesired directions. Technically this is called spatial aliasing. It is analogous to the Nyquist sampling rate for a frequency-domain analysis of a signal.

Now consider that the all the signals generated by all the d sources, $s_i(t)$, $1 \le i \le d$, the overall signal and noises received by the mth element at time t can be expressed as:

$$\begin{aligned} x_m(t) &= \sum_{i=1}^{d} s_i(t) + n_m(t) \\ &= \sum_{i=1}^{d} s_i(t)e^{j(m-1)\mu_i} + n_m(t) \\ &= s_i(t)\sum_{i=1}^{d} e^{j(m-1)\mu_i} + n_m(t) \quad {}_{m=1,2,\ldots,M} \end{aligned} \quad (3.9)$$

To differentiate the pure signals generated by the sources and the noise-added or corrupted signals received or detected, the latter is hereafter called the data and denoted with symbol x or $\mathbf{x}$.

In a matrix form, (3.9) can be written as:

$$\mathbf{x}(t) = \left[\mathbf{a}(\mu_1), \mathbf{a}(\mu_2) \ldots \mathbf{a}(\mu_d)\right] \begin{bmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_d(t) \end{bmatrix} + \mathbf{n}(t) = \mathbf{A}s(t) + \mathbf{n}(t) \quad (3.10)$$

where $\mathbf{x}(t) = [x_1(t)\ x_2(t)\ \ldots\ x_M(t)]^T$ is the data column vector received by the array, $\mathbf{s}(t) = [s_1(t)\ s_2(t)\ \ldots\ s_M(t)]^T$ is the signal column vector generated by the sources, $\mathbf{n}(t) = [n_1(t)\ n_2(t)\ \ldots\ n_M(t)]^T$ is a zero-mean spatially uncorrelated additive noises with spatial covariance matrix equal to $\sigma_N^2 \mathbf{I}_M$. The array steering column vector $\mathbf{a}(\mu_i)$ is defined as:

$$\mathbf{a}(\mu_i) = \left[1 \quad e^{j\mu_i} \quad e^{j2\mu_i} \quad \ldots \quad e^{j(M-1)\mu_i}\right]^T \quad (3.11)$$

It is functions of unknown spatial frequencies μ_i; it forms the columns of the $M \times d$ steering matrix $\mathbf{A}$:

$$\begin{aligned} \mathbf{A} &= \left[a(\mu_1) \ \ldots\ \mathbf{a}(\mu_i) \ \ldots\ \mathbf{a}(\mu_d)\right] \\ &= \begin{bmatrix} 1 & 1 & \ldots & 1 \\ e^{j\mu_1} & e^{j\mu_2} & \ldots & e^{j\mu_d} \\ \ldots & \ldots & \ldots & \ldots \\ e^{j(M-1)\mu_1} & e^{j(M-1)\mu_2} & \ldots & e^{j(M-1)\mu_d} \end{bmatrix} \end{aligned} \quad (3.12)$$

Depending upon the different configurations of an array, different array steering matrices can be formed. Many algorithms particularly demand arrays to have centro-symmetric configurations [4]. Fortunately, many array configurations used, such as the uniform linear array and the rectangular array, satisfy this requirement. In the next section, we review two very common centro-symmetric arrays: ULA and URA.

3.3 Centro-Symmetric Sensor Arrays

In this section, the basic definitions of centro-Hermitian and centro-symmetric arrays are defined. These arrays are often used in practice with special properties. They are normally described by centro-symmetric and centro-Hermitian matrices. In this section, we will elaborate them based on the descriptions presented in [1, 5].

First, let us denote $C^{p\times q}$ as the collections or space of the $p \times q$ matrix. Then the following definitions represent what will be a centro-symmetric or centro-Hermitian matrix.

Definition 3.1

A complex $p \times q$ matrix $\mathbf{M} \in C^{p\times q}$ is called centro-symmetric [1, 5] if

$$\Pi_p \mathbf{M} \Pi_q = \mathbf{M} \tag{3.13}$$

Here Π_p is the $p \times p$ exchange matrix as defined in Appendix A.2; it has one on its antidiagonal elements and zeros elsewhere:

$$\Pi_p = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix} \tag{3.14}$$

It is not difficult to demonstrate that premultiplication of a matrix by Π_p will reverse the order of the rows of the matrix (i.e., the first row becomes the last row, the second row becomes the second last row, and the last row will become the first row, and so forth). Postmultiplication of a matrix by Π_p will reverse the order of the columns of the matrix (i.e., the first column becomes the last column, the second column becomes the second last column, and the last column becomes the first column). Similarly, Π_q is the $q \times q$ exchange matrix.

Condition (3.13) basically states that $\mathbf{M} \in C^{p\times q}$ has an inversion center: when $\mathbf{M}$ is rotated by 180°, it will be the same as the original. Spatially, it means that for every point (x, y, z), there is an indistinguishable point $(-x, -y, -z)$.

A complex matrix $\mathbf{M} \in C^{p\times q}$ is called centro-Hermitian-symmetric if [1, 5]

$$\Pi_p \overline{\mathbf{M}} \Pi_q = \mathbf{M} \tag{3.15}$$

where $\overline{\mathbf{M}}$ is the conjugate of $\mathbf{M}$.

Condition (3.15) basically states that $\mathbf{M} \in C^{p\times q}$ has a conjugate inversion center; when $\mathbf{M}$ is conjugated and then rotated by 180°, it will become the same as the original. Spatially, it can be seen that for every point (x, y, z), there is an indistinguishable conjugated point $(-x, -y, -z)$.

Definition 3.2

Matrix $\mathbf{Q} \in C^{p\times q}$ satisfying

$$\Pi_p \overline{\mathbf{Q}} = \mathbf{Q} \Leftrightarrow \Pi_p \mathbf{Q} = \overline{\mathbf{Q}} \tag{3.16}$$

is called the left Π real symmetric matrix.

For instance, unitary matrices

$$Q_{2n}^{(s)} = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{I}_n & j\mathbf{I}_n \\ \Pi_n & -j\Pi_n \end{bmatrix} \text{and } \mathbf{Q}_{2n+1}^{(s)} = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{I}_n & \mathbf{0} & j\mathbf{I}_n \\ \mathbf{0}^T & \sqrt{2} & 0^T \\ \Pi_n & 0 & -j\Pi_n \end{bmatrix} \tag{3.17}$$

are left Π real-matrices of even and odd orders, respectively. Here $\mathbf{I}_n$ is an identity matrix of size $n \times n$. Left Π real-matrices can also be obtained by postmultiplying a left real-Π matrix $\mathbf{Q} \in C^{p\times q}$ with an arbitrary real matrix $\mathbf{R}$. That is, for every matrix left Π matrix $\mathbf{Q} \in C^{p\times q}$,

$$\mathbf{Q}_R = \mathbf{QR} \text{ with } \mathbf{R} \in R^{q\times r} \tag{3.18}$$

is also a left Π real-matrix.

Definition 3.3

A sensor array of M elements is a centro-symmetric array as long as its element locations are symmetric with respect to the centroid and the radiation characteristics of paired elements are the same. Their array steering matrix $\mathbf{A}$, therefore, can be proven to satisfy

$$\Pi_M \overline{\mathbf{A}} = \mathbf{A}\Lambda \tag{3.19}$$

for some unitary diagonal matrix Λ. Note that here the sensor elements are assumed have the identical radiation characteristics.

In next section, two commonly used centro-symmetric arrays, the uniform linear array (ULA) and the uniform rectangular array (URA), are discussed, and the expressions are derived for their data and array steering matrix.

3.3.1 Uniform Linear Array

Now consider the case of a uniform linear array (ULA), which is among the most common arrays used in practice. An example configuration is shown in Figure 3.2. As described in Section 3.2.1, the array steering vector corresponding to the spatial frequency $\mu_i = -\frac{2\pi}{\lambda}\Delta \sin\theta_i$ can be written as

$$\mathbf{a}(\mu_i) = \begin{bmatrix} 1 & e^{j\mu_i} & e^{j2\mu_i} & \dots & e^{j(M-1)\mu_i} \end{bmatrix}^T,\ 1 \le i \le d \tag{3.20}$$

and the steering matrix $\mathbf{A}$ has the following Vandermonde structure:

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} \mathbf{a}(\mu_1)\mathbf{a}(\mu_2) & \dots & \mathbf{a}(\mu_d) \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & \dots & 1 \\ e^{j\mu_1} & e^{j\mu_2} & \dots & e^{j\mu_d} \\ \dots & \dots & \dots & \dots \\ e^{j(M-1)\mu_1} & e^{j(M-1)\mu_2} & \dots & e^{j(M-1)\mu_d} \end{bmatrix} \end{aligned} \tag{3.21}$$

The ULAs are centro-symmetric because the following relationship can be easily proven:

$$\Pi_M \overline{\mathbf{A}} = \mathbf{A}\Lambda_d \text{ with}$$

$$\Lambda_d = \Phi^{-(M-1)} = \begin{bmatrix} e^{-j(M-1)\mu_1} & 0 & \dots & 0 \\ 0 & e^{-j(M-1)\mu_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & e^{j(M-1)\mu_d} \end{bmatrix} \tag{3.22}$$

Again, if the center of the array is chosen as the phase reference, the resulting array steering matrix $\mathbf{A}_c$ can be obtained by scaling the columns of $\mathbf{A}$ in the following fashion,

$$\mathbf{A}_c = \mathbf{A} \cdot \Phi^{-\left(\frac{M-1}{2}\right)} \tag{3.23}$$

It is not difficult to show that $\mathbf{A}_c$ satisfies (3.14); therefore, $\mathbf{A}_c$ is centro-symmetric. This result will be used in Chapter 5. The steering vectors of this array, that is, the columns of $\mathbf{A}_c$ are

$$\mathbf{a}_M(\mu_1) = e^{-j\frac{M-1}{2}\mu_i}\left[1 \quad e^{j\mu_i} \quad e^{j2\mu_i} \quad \ldots \quad e^{j(M-1)\mu_i}\right]^T, 1 \le i \le d \tag{3.24}$$

Here, the subscript M indicates the dimension of the left Π steering vector $\mathbf{a}_M(\mu_i)$.

3.3.2 Uniform Rectangular Array (URA)

Two-dimensional arrays are employed when one wants to estimate the angles in both azimuth and elevation of a source, as shown in Figure 3.3. Such an array configuration is shown in Figure 3.4. It includes three different forms of the array element positioning, some without their center elements [Figure 3.4(b)], while others consisting of orthogonal lines of the array elements with common phase centers [Figure 3.4(a, c)] [6].

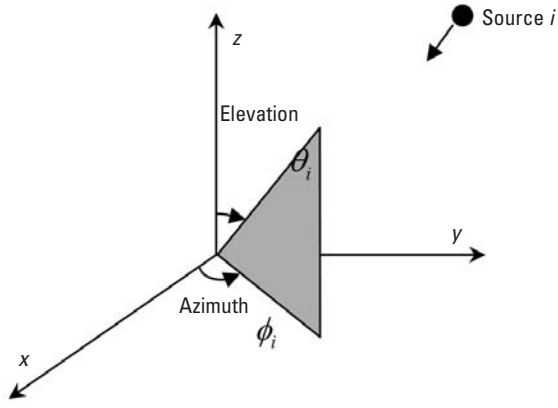


Figure 3.3 Definition of azimuth and elevation angles.

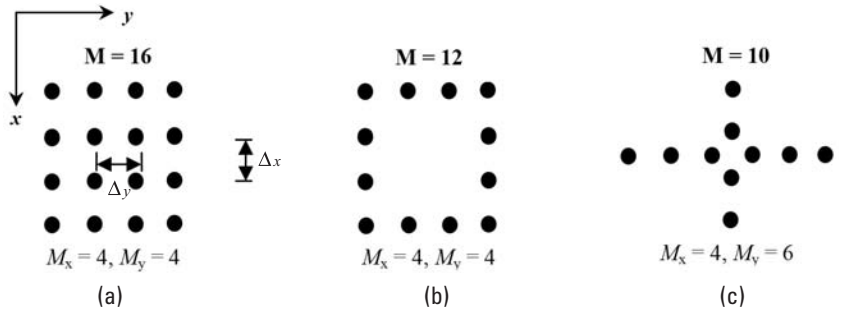


Figure 3.4 (a–c) Configurations of the centro-symmetric two-dimensional array.

We now consider Figure 3.4(a). The dimensions of the two-dimensional array are $M_x \times M_y$. The array element is numbered by (k_x, k_y) with $1 \le k_x \le M_x$ and $1 \le k_y \le M_y$. By following the data model described before, we assume that the d narrowband signals of wavelength λ are emitted by d sources and the planar signal wavefront from the ith source impinges on the array at azimuth angle ϕ_i and elevation angle θ_i, $1 \le i \le d$ (see Figure 3.3). Let

$$u_i = \cos\phi_i \sin\theta_i \text{ and } v_i = \sin\phi_i \sin\theta_i,\ 1 \le i \le d \tag{3.25}$$

Denote

$$\xi_i = u_i + jv_i = \sin\theta_i e^{j\phi_i} \tag{3.26}$$

A simple formula to determine azimuth ϕ_i and elevation θ_i from the corresponding direction cosines u_i and v_i, respectively, is then given by

$$\phi_i = \arg(\xi_i) \text{ and } \theta_i = \arcsin(\|\xi_i\|) \tag{3.27}$$

Equation (3.27) shows that to determine the DOA of a source is to determine corresponding direction cosines u_i and v_i, respectively.

If the data received by array element (k_x, k_y) is given by $x_{k_x,k_y}(t_n)$ at time snapshot $t = t_n$, there are $M = M_x \cdot M_y$ samples at each snapshot. By extending the one-dimensional result (3.9), it is not difficult to find the superposition of the d signals and noises received by the element, which can be written as

$$x_{k_x,k_y}(t_n) = \sum_{i=1}^{d} s_i(t_n) e^{j(k_x-1)\mu_i} e^{j(k_y-1)v_i} + n_{k_x,k_y}(t_n) \tag{3.28}$$

Let the two-dimensional matrix $\chi(t_n) \in C^{M_x \times M_y}$ contain the data received by all the $M_x \times M_y$ elements at a $t = t_n$ snapshot. Then, $\chi(t_n)$ can be written as

$$\chi(t_n) = \sum_{i=1}^{d} \mathbf{a}(\mu_i)\mathbf{a}^T(v_i)s_i(t_n) + N(t_n) \tag{3.29}$$

where $\mathbf{a}(\mu_i) \in C^{M_x}$ and $\mathbf{a}(v_i) \in C^{M_y}$ are defined by (3.20). These two vectors can be interpreted as array steering vectors of ULA of size M_x and M_y, respectively. The matrix $N(t_n) \in C^{M_x \times M_y}$ in (3.29) are the noise samples.

It is desirable that the array data model, regardless for one-dimensional or two-dimensional arrays, is conformal to (3.10) in order to simplify and unify the DOA analysis. However, (3.29) for the two-dimensional array is not conformal to the data model represented by (3.10), since the left-hand side $\chi(t_n) \in C^{M_x \times M_y}$ is an $M_x \times M_y$ matrix rather than a one-column vector. To overcome the problem, (3.29) can be stacked to form a one-column vector by applying the vector mapping operation, vec{·}. In order words, the vec{·} operator maps an $M_x \times M_y$ dimensional matrix to a column vector of size $M_x \cdot M_y$. The resulting data column vector is then:

$$x(t_n) = \text{vec}\{\chi(t_n)\} \tag{3.30}$$

This stacking procedure can be illustrated by taking an example [Figure 3.4(a)]. Let the URA be of size $M = M_x \times M_y = 4 \times 4 = 16$ elements. The data received by the antenna elements are then stacked column-wise. More specifically, the first element of $\mathbf{x}(t_n)$ is the data received by the element in the upper left corner of the array. Then the process sequentially goes downwards, along the positive *x*-axis, such that the fourth element of $\mathbf{x}(t_n)$ is the data received by the antenna element in the bottom left corner of the array. The fifth element of $\mathbf{x}(t_n)$ is the data received by the element at the top of the second column of the array, the eighth element of $\mathbf{x}(t_n)$ is the data received by the element at the bottom

of the second column, and so on. This forms an $M = 16$ dimensional column vector $\mathbf{x}(t_n)$ at each sampling instant t_n.

In correspondence to the stacking on the left-hand size of (3.29), the right-hand side of (3.29) needs to be rearranged in representation of the two-dimensional array manifold as the array steering vector $\mathbf{a}(\mu_i, v_i) \in C^M$, which is also one-column wise. Through mathematical manipulations, we can have the following relationships:

$$\mathbf{a}(\mu_i, v_i) = \text{vec}\{A(\mu_i, v_i)\} \tag{3.31}$$

where $A(\mu_i, v_i) = \mathbf{a}(\mu_i)\mathbf{a}(v_i)^T \in C^{M_x \times M_y}$ with $\mathbf{a}(\mu_i) \in C^{M_x}$ and $\mathbf{a}(v_i) \in C^{M_y}$ being the steering vectors of the ULAs in the x direction and the y direction, respectively. By applying the vec{ } operator to $A(\mu_i, v_i)$, the two-dimensional array steering vectors of the URA can be shown to be the Kronecker products of the array steering vectors of the ULA in the x and y directions (the definition of the Kronecker product can be found in Appendix A.1). Hence, the array steering matrix $\mathbf{A}$ for the two-dimensional rectangular array can be written as

$$\mathbf{A} = \left[\mathbf{a}(\mu_i, v_i)\ \mathbf{a}(\mu_2, v_2)\ \ldots\ \mathbf{a}(\mu_d, v_d)\right] \tag{3.32}$$

where

$$\mathbf{a}(\mu_i, v_i) = \mathbf{a}(v_i) \otimes \mathbf{a}(\mu_i)^T,\ 1 \le i \le d \tag{3.33}$$

Here $\otimes$ represents the Kronecker product. The spatial frequencies are in the x direction. μ_i, and the spatial frequencies in the y direction, v_i, are given by

$$\mu_i = \frac{2\pi}{\lambda}\Delta_x u_i \quad \text{and} \quad v_i = \frac{2\pi}{\lambda}\Delta_x v_i \tag{3.34}$$

With the construction of (3.32) as the steering matrix, date model of a two-dimensional matrix is made conformal to the standard form of (3.10) introduced at the beginning of this chapter. The difference in the steering matrix of a one-dimensional array and that of a two-dimensional array is that there are two spatial frequencies, μ_i, and v_i, in the steering matrix of the two-dimensional array; these two frequencies are related to

elevation and azimuthal angles of an incoming signal generated by a source. The DOA process is then to extract these two frequencies from the data model.

In the previous two sections we discussed the expressions for data vector **x** using the array steering matrices of some commonly used antenna arrays. In the next section, we discuss the covariance matrix of these data vectors, an important quantity that is used in DOA algorithms.

3.3.3 Covariance Matrices

The signals received by an array are noise-corrupted in the real world. These noises are normally uncorrelated while the pure signals received by different elements are correlated as they are originated from the same sources. By using this property, one may be able to extract effectively the DOA information. To do so, the concept of cross-covariance information among the noise-corrupted signals, spatial covariance matrix, is introduced and employed in finding DOAs. The spatial covariance matrix of the data (i.e., signals plus noises) received by an array is defined as

$$\mathbf{R}_{xx} = E\left\{\mathbf{x}(t)\mathbf{x}^H(t)\right\} \tag{3.35}$$

where E{ } denotes the statistical expectation.

Equation (3.35) quantifies the degree of correlation of the data signals received by array elements. The higher values of its elements, the higher degree of correlations among the signals.

By substituting (3.10) into (3.35), we have

$$\mathbf{R}_{xx} = E\left\{\mathbf{x}(t)\mathbf{x}^H(t)\right\} = \mathbf{A}\mathbf{R}_{ss}\mathbf{A}^H + \sigma_N^2\mathbf{I}_M \tag{3.36}$$

where $\mathbf{R}_{ss} = E\{\mathbf{s}(t)\mathbf{s}^H(t)\}$ is the signal covariance matrix and σ_N^2 is the common variance of the noises.

In practice, the exact covariance matrix of $\mathbf{R}_{xx}$ is difficult to find due to the limited number of data sets received and processed by an array. Therefore, an estimation is made. By assuming that all underlying random noise processes are ergodic, the ensemble average (or statistical expectation) can be replaced by a time average. Let **X** be denoted as the noise corrupted signal (or data) matrix composed of N snapshots of $\mathbf{x}(t_n)$, $1 \le n \le N$,

$$\begin{aligned}\mathbf{X} &= \left[\mathbf{x}(t_1)\ \mathbf{x}(t_2)\ \ldots\ \mathbf{x}(t_N)\right]^T \\ &= \mathbf{A}\cdot\left[\mathbf{s}(t_1)\ \mathbf{s}(t_2)\ \ldots\ \mathbf{s}(t_N)\right]^T + \left[\mathbf{n}(t_1)\ \mathbf{n}(t_2)\ldots\ \mathbf{n}(t_N)\right]^T \\ &= \mathbf{A}\cdot\mathbf{S}+\mathbf{N}\end{aligned} \tag{3.37}$$

Note that (3.37) is different from the basic data model (3.10) in that the left-hand side term, **X**, is a stacked noised corrupted signals (or data) received by the array elements at different snapshots, $\mathbf{x}(t_n)$, $n = 1, 2, 3, \ldots, N$. Similarly, the stacking applies to the pure signal vector **s** and the noise vector **n**.

An estimate of the data covariance matrix $\mathbf{R}_{xx}$ as a time average can then be computed with

$$\mathbf{R}_{xx} \approx \hat{\mathbf{R}}_{xx} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(t_n)\mathbf{x}^H(t_n) = \frac{1}{N}\mathbf{X}^H\mathbf{X} \tag{3.38}$$

This fundamental data matrix is used in all our future description of DOA estimation algorithms. Many DOA estimation algorithms basically try to extract the information from this array data covariance matrix.

With this knowledge of data model and fundamental necessary mathematics, we now discuss some popular DOA estimation techniques.

3.4 Beamforming Techniques

The basic idea behind beamforming techniques [2] is to "steer" the array in one direction at a time and measure the output power. When the "steered" direction coincides with a DOA of a signal, the maximum output power will be observed. The development of the DOA estimation schemes is essentially the design of an appropriate form of output power that will be strongly related to the DOA.

Given the knowledge of array steering vector, an array can be steered electronically just as a fixed antenna can be steered mechanically. However, the array pattern can change shape in addition to changing orientation. A weight vector **w** can be designed and then used to linearly combine the data received by the array elements to form a single output signal $y(t)$,

$$y(t) = \mathbf{w}^H\mathbf{x}(t) \tag{3.39}$$

The total averaged output power out of an array over N snapshots can be expressed as

$$P(\mathbf{w}) = \frac{1}{N}\sum_{n=1}^{N}\left|y(t_n)\right|^2 = \frac{1}{N}\sum_{n=1}^{N}\mathbf{w}^H\mathbf{x}(t_n)\mathbf{x}^H(t_n)\mathbf{w} = \mathbf{w}^H\hat{\mathbf{R}}_{xx}\mathbf{w} \tag{3.40}$$

Different beamforming techniques have been developed by measuring the above output power with different choices of the weighting vector $\mathbf{w}$. Two main techniques of this class are discussed here.

3.4.1 Conventional Beamformer

In the conventional beamforming approach [6, 7], $\mathbf{w} = \mathbf{a}(\theta)$ with θ being a scanning angle that is scanned over the angular region of interest for a ULA of M elements. $\mathbf{w} = \mathbf{a}(\theta)$ is defined similarly as the steering vector (3.11), but with an arbitrary scanning angle θ:

$$\mathbf{a}(\theta) = \begin{bmatrix} 1 & e^{j\mu} & e^{j2\mu} & \cdots & e^{j(M-1)\mu} \end{bmatrix}^T \tag{3.41}$$

with $\mu = -\dfrac{2\pi f_c}{c}\Delta\sin\theta = -\dfrac{2\pi}{\lambda}\Delta\sin\theta$.

For each look or scanned direction θ, the average power output $P(\theta)$ of the steered array is then measured or computed with (3.40).

In other words, the output power versus θ is recorded with (3.40). It can be shown that when $\theta = \theta_i$, an impinging angle of the signal from source i, the output power $P(\theta)$ will reach a peak or maximum point. At this moment, $\mathbf{w} = \mathbf{a}(\theta - \theta_i)$ aligns the phases of the signal components received by all the elements of the array, causing them to add constructively and produce a maximum power.

In practical computations, $\mathbf{w} = \mathbf{a}(\theta)$ is normalized as

$$\mathbf{w} = \mathbf{w}_{CON} = \frac{\mathbf{a}(\theta)}{\sqrt{\mathbf{a}^H(\theta)\mathbf{a}(\theta)}} \tag{3.42}$$

By inserting the weight vector equation (3.42) into (3.40), the output power as a function of angle of arrival, or termed as spatial spectrum, is obtained as

$$P(\theta) = P_{CON}(\theta) = \frac{\mathbf{a}^H(\theta)\hat{\mathbf{R}}_{xx}\mathbf{a}(\theta)}{\mathbf{a}^H(\theta)\mathbf{a}(\theta)} \tag{3.43}$$

The weight vector (3.42) can be interpreted as a spatial filter; it is matched to the impinging spatial angles of the incoming signal to produce a peak but attenuate the output power for signals not coming from the angles of the incoming signals. Intuitively, it equalizes the different signal delays experienced by the array elements [i.e., (3.6)] and maximally combine their respective contributions to form a peak in output power at the angles of the incoming signals.

A simulation was conducted by employing a 6-element ULA with its omnidirectional array elements separated by a half wavelength. Two equally powered uncorrelated signals were made to impinge on the array from 10° and 30° (the uncorrelated signals makes $\hat{\mathbf{R}}_{xx}$ nonsingular). In such a case, $d = 2$, $M = 6$, $\theta_1 = 10°$, and $\theta_2 = 30°$. An SNR (signal-to-noise ratio) of 10 dB was assumed. Figure 3.5 shows the result. Two peaks can be seen at 10° and 30°, but the 30° peak is not so obvious and are somewhat "averaged" with the peak at 10°. In other words, the spread of each peak is large, and if two impinging angles are close to each other, the two

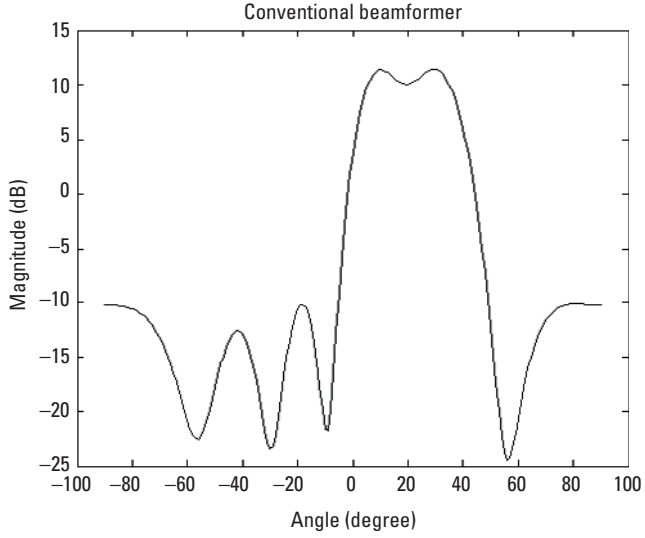


Figure 3.5 DOA estimation with the conventional beamformer; the signal impinges at 10° and 30°.

peaks may be "blurred" into one pair. By further numerical experiments, it is found that the best resolution of a peak is about 19°.

In a more general term, although (3.43) is simple to implement, the width of the beam associated with a peak and the height of the sidelobes, as seen in Figure 3.5, are relatively large; they limit the method's effectiveness when signals arriving from multiple directions and/or sources are present. This technique has poor resolution. Although it is possible to increase the resolution by adding more array elements, it leads to the increase in the numbers of receivers and the amount of storage required for the data.

3.4.2 Capon's Beamformer

The conventional beamformer works described earlier on the premise that pointing the strongest beam in a particular direction yields the peak power arriving in that direction. In other words, all the degrees of freedom available to the array were used in forming a beam in the required look direction. This works well when there is only one incoming signal present. But when there is more than one signal present, the array output power contains signal contributions from the desired angle as well as from the undesired angles.

Capon's method [9] overcomes this problem by using the degrees of freedom to form a beam in the look direction and at the same time the nulls in other directions in order to reject other signals. In terms of the array output power, forming nulls in the directions from which other signals arrive can be accomplished by constraining a beam (or at least maintaining unity gain) in the look direction. Thus, for a particular look direction, Capon's method uses all but one of the degrees of the freedom to minimize the array output power while using the remaining degrees of freedom to constrain the gain in the look direction to be unity:

$$\min P(\mathbf{w}) \text{ subject to } \mathbf{w}^H \mathbf{a}(\theta) = 1 \tag{3.44}$$

The weight vector chosen in this way is often referred to as the minimum variance distortion-less response (MVDR) beamformer, since, for a particular look direction, it minimizes the variance (average power) of the array output signal while passing the signal arriving from the look direction with no distortion. The resulting weight vector is shown to be given by

$$\mathbf{w} = \mathbf{w}_{CAP} = \frac{\hat{\mathbf{R}}_{xx}^{-1}\mathbf{a}(\theta)}{\mathbf{a}^H(\theta)\hat{\mathbf{R}}_{xx}^{-1}\mathbf{a}(\theta)} \tag{3.45}$$

By substituting the above weight vector into (3.40), the following spatial power spectrum is obtained:

$$P(\theta) = P_{CAP} = \frac{1}{\mathbf{a}^H(\theta)\hat{\mathbf{R}}_{xx}^{-1}\mathbf{a}(\theta)}$$

A simulation was conducted by employing a 6-element ULA with its omnidirectional elements separated by a half wavelength. Two equally powered uncorrelated signals were made to impinge on the array from 10° and 30°. In such a case, $d = 2$, $M = 6$, $\theta_1 = 10°$, and $\theta_2 = 30°$. An SNR of 10 dB was assumed. Figure 3.6 shows the result. It can be seen that in comparison with Figure 3.5, the peaks at 10° and 30° are much sharper and better separated compared to that of the conventional beamformer. The side peaks or lobes at other angles are also reduced, making them less likely to confuse the interpretation of the output power. The best resolution achieved was 10°. However, this increased resolution comes at the cost of increased computing time or power, due to the need to invert a large matrix in (3.45).

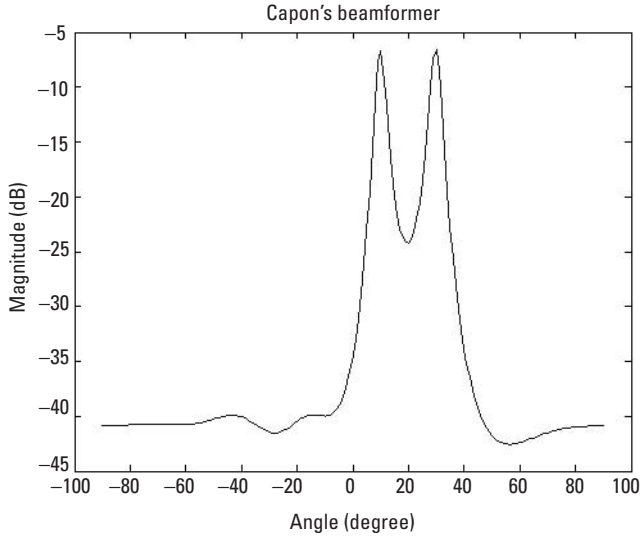


Figure 3.6 DOA estimation with Capon's beamformer; the signal impinges at 10° and 30°.

Though it provides a better resolution in comparison with the conventional beamforming technique, Capon's method suffers from other disadvantages. Capon's method fails if the signals that are correlated are present because it inadvertently uses the correlation matrix $\hat{\mathbf{R}}_{xx}^{-1}$, which becomes singular for the correlated signals. In other words, the correlated components will be combined destructively in the process of minimizing the output power. Also, Capon's method requires the computation of a matrix inverse, which can be expensive for large arrays.

3.4.3 Linear Prediction

This method minimizes the mean output signal power of the array elements subject to the constraint that the weight on a selected element is unity [10]. Expressions for the array weights and the power spectrum are given, respectively, by

$$\mathbf{w} = \frac{\hat{\mathbf{R}}_{xx}^{-1}\mathbf{u}}{\mathbf{u}^H\hat{\mathbf{R}}_{xx}^{-1}\mathbf{u}} \tag{3.46}$$

and

$$P(\theta) = P_{LP}(\theta) = \frac{\mathbf{u}^H\hat{\mathbf{R}}_{xx}^{-1}\mathbf{u}}{\left|\mathbf{u}^H\hat{\mathbf{R}}_{xx}^{-1}\mathbf{a}(\theta)\right|^2} \tag{3.47}$$

where $\mathbf{u}$ is a column vector of all zeros except for the specified element, which is equal to 1. The position of the specified element in the column corresponds to the position of the selected element in the array. There is no hard criterion for the proper choice of this element. The choice of this element, however, affects the resolution capability of the DOA estimates, and this effect is dependent upon the SNR and the separation of the directional sources. The linear prediction method performs well in a moderately low SNR environment and is a good compromise in situations where sources are of approximately equal strength and are nearly coherent.

A simulation was conducted by employing a 6-element ULA with its omnidirectional elements separated by a half wavelength. Two equally powered uncorrelated signals were made to impinge on the array from 10° and 30°. In such a case, $d = 2$, $M = 6$, $\theta_1 = 10°$, and $\theta_2 = 30°$. An SNR

of 10 dB was assumed. Figure 3.7 shows the result. In comparison with Figure 3.6, the linear prediction method performs better than Capon's method, with defined peaks at 10° and 30°. However, the sidelobes are more prominent. In addition, the formula is more complex, requiring slightly more computing time and power. In our tests, the best resolution achieved was found to be 7°.

3.5 Maximum Likelihood Techniques

Maximum likelihood (ML) techniques were some of the first techniques investigated for DOA estimation [11]. Since ML techniques were computationally intensive, they are less popular than other techniques. However, in terms of performance, they are superior to other estimators, especially at low SNR conditions.

Given an received data sequence $\mathbf{x}(t_n)$, it is desired to reconstruct the components of the data due only to the desired signals. The parameter values for which the reconstruction approximates the received data with maximal accuracy are taken to be the DOA and desired signal waveform estimates. The approach taken here is to subtract from $\mathbf{x}(t_n)$ an estimate $\mathbf{A}(\hat{\theta})\hat{\mathbf{s}}(t_n)$ of the signal components $\mathbf{A}(\theta)\mathbf{s}(t_n)$. If the estimates $\hat{\theta}$ and

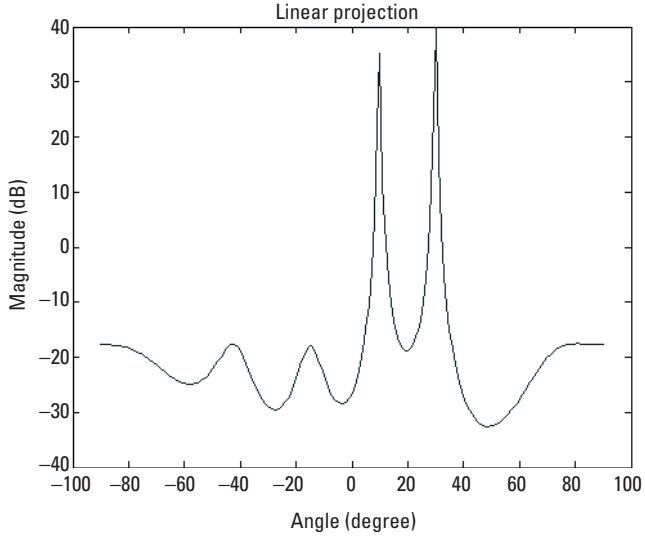


Figure 3.7 DOA estimation with the linear prediction; the signal impinges at 10° and 30°.

$\hat{\mathbf{s}}(t_n)$ are sufficiently good, the residual $\mathbf{x}(t_n) - \mathbf{A}(\hat{\theta})\hat{\mathbf{s}}(t_n)$ will consist primarily of noise and interference with smallest energy. In other words, minimizing the energy in the residual $\mathbf{x}(t_n) - \mathbf{A}(\hat{\theta})\hat{\mathbf{s}}(t_n)$ by proper choice of $\hat{\theta}$ and $\hat{\mathbf{s}}(t_n)$ can result in accurate estimates of $\theta_i \approx \hat{\theta}$ and $\mathbf{s}(t_n) \approx \hat{\mathbf{s}}(t_n)$.

The method can be stated mathematically in a least-squares form as

$$\min_{\theta, \hat{\mathbf{s}}(t_n)} \left\langle \left\| \mathbf{x}(t_n) - \mathbf{A}\left(\hat{\theta}\right)\hat{\mathbf{s}}(t_n) \right\|^2 \right\rangle_N \tag{3.48}$$

for which the best least squares fit between the received signals and a reconstruction of the signal components is sought. After mathematical manipulations, the solution for $\hat{\mathbf{s}}(t_n)$ in terms of any $\hat{\theta}$ can be found as

$$\hat{\mathbf{s}}(t_n) = \left(\mathbf{A}^H\left(\hat{\theta}\right)\mathbf{A}\left(\hat{\theta}\right)\right)^{-1}\mathbf{A}^H\left(\hat{\theta}\right)\mathbf{x}(t_n) = \mathbf{W}^H\mathbf{x}(t_n) \tag{3.49}$$

By substituting this signal estimate back into the residual and minimizing it over the DOA estimates, $\hat{\theta}$ can be shown to be equivalent to maximizing a matrix trace

$$\max_{\hat{\theta}} \text{trace}\left\{\mathbf{P}_A\left(\hat{\theta}\right)\mathbf{R}_{xx}\right\} \tag{3.50}$$

where $\mathbf{P}_A(\hat{\theta})$ is the matrix for the space spanned by the columns of $\mathbf{A}(\hat{\theta})$,

$$\mathbf{P}_A\left(\hat{\theta}\right) = \mathbf{A}\left(\hat{\theta}\right)\left(\mathbf{A}^H\left(\hat{\theta}\right)\mathbf{A}\left(\hat{\theta}\right)\right)^{-1}\mathbf{A}^H\left(\hat{\theta}\right) \tag{3.51}$$

If $\hat{\theta}$ is the exact solution (direction of the arrival), the reconstructed data $\mathbf{A}(\hat{\theta})\hat{\mathbf{s}}(n)$ is equal to the true signal components plus residential noise,

$$\begin{aligned} \mathbf{X}(n) - \mathbf{A}\left(\hat{\theta}\right)\hat{\mathbf{S}}(n) &= \mathbf{X}(n) - \mathbf{P}_A\left(\hat{\theta}\right)\mathbf{X}(n) \\ &= \left(\mathbf{A}\left(\hat{\theta}\right)\hat{\mathbf{S}}(n) + \mathbf{N}\right) - \left(\mathbf{A}\left(\hat{\theta}\right)\hat{\mathbf{S}}(n) + \mathbf{P}_A\left(\hat{\theta}\right)\mathbf{N}\right) \\ &= \left(\mathbf{I} - \mathbf{P}_A\left(\hat{\theta}\right)\right)\mathbf{N} \end{aligned} \tag{3.52}$$

Assuming that the interference and the noise are spatially uncorrelated, the average power in the residual is the minimum value compared with the values with any other choice of $\hat{\theta} \neq \theta_i$.

3.6 Subspace-Based Techniques

These techniques basically rely on the following proven properties of the matrix space defined by $\mathbf{R}_{xx}$:

1. The space, spanned by its eigenvectors, can be partitioned into two orthogonal subspaces, namely, the signal subspace and noise subspace.
2. The steering vectors correspond to the signal subspace.
3. The noise subspace is spanned by the eigenvectors associated with the smaller eigenvalues of the correlation matrix.
4. The signal subspace is spanned by the eigenvectors associated with the larger eigenvalues.

More elaborations are given next.

3.6.1 Concept of Subspaces

With a given matrix $\mathbf{X}$ of size $M \times N$, columns (rows) of this matrix can be parallel to (i.e., dependent of) each other or nonparallel to (i.e., independent of) each other. If there are $d(d \leq M;\ d \leq N)$ independent columns in $\mathbf{X}$, this matrix is said to have a d dimensional range or column space, which is a subspace of the M-dimensional Euclidean space $\mathbf{C}^M$. The rank of the matrix is the dimension of this subspace. If $d = M$, the matrix is of full rank, and if $d < M$, it is rank deficient. $\mathbf{C}^M$ can be spanned by the columns of any unitary matrix in $\mathbf{C}^{M \times M}$, which is defined as the Euclidean space of square and complex valued M dimensional matrices.

The same holds for $\mathbf{C}^N$ of which the row space of $\mathbf{X}$ is a d dimensional subspace: there are $d(d \leq M;\ d \leq N)$ independent rows in $\mathbf{X}$ and $\mathbf{C}^N$ can be spanned by the rows of any unitary matrix in $\mathbf{C}^{N \times N}$.

Suppose $d(d \leq M;\ d \leq N)$. Then a unitary matrix $\mathbf{U}$ can be chosen such that the d dimensional column space of $\mathbf{X}$ is spanned by a subset of d

columns of **U**, say, the first d columns, which together form a matrix $\mathbf{U}_s$. Let the remaining $M_s - d$ columns together form a matrix $\mathbf{U}_o$. Then,

$$\mathbf{U} = [\mathbf{U}_s \mathbf{U}_o] \tag{3.53}$$

Since **U** is a unitary matrix, it can be observed that

1. From $\mathbf{U}^H\mathbf{U} = \mathbf{I}_M$

$$\mathbf{U}_s^H \mathbf{U}_s = \mathbf{I}_d \tag{3.54}$$

$$\mathbf{U}_s^H \mathbf{U}_o = 0 \tag{3.55}$$

$$\mathbf{U}_o^H \mathbf{U}_o = \mathbf{I}_{M-d} \tag{3.56}$$

2. From $\mathbf{U}\mathbf{U}^H = \mathbf{I}_M$

$$\mathbf{U}_s \mathbf{U}_s^H + \mathbf{U}_o \mathbf{U}_o^H = \mathbf{I}_M \tag{3.57}$$

where $\mathbf{I}_d$ is the identity matrix of rank d, and $\mathbf{I}_{M-d}$ is the identity matrix of rank $(M - d)$. Relations (3.54) through (3.57) indicate that any vector $\mathbf{u} \in \mathbf{C}^M$ can be decomposed into two mutually orthogonal vectors $\mathbf{u}_s$ and $\mathbf{u}_o$ in the spaces spanned by $\mathbf{U}_s$ and $\mathbf{U}_o$, respectively. These two spaces are d dimensional and $M - d$ dimensional orthogonal subspaces in $\mathbf{C}^M$, and their direct sum is equal to $\mathbf{C}^M$. As observed, the two subspaces are orthogonal to each other. In our case, the dominant subspace is due to the signals and is referred to as the signal subspace while its complimentary space is referred to as noise subspace [11].

Hence, a given matrix is decomposed in order to obtain these two subspaces. In our case, matrix **X** is decomposed as $\mathbf{X} = \mathbf{U}\Sigma\mathbf{U}$, where **U** and **V** are matrices whose columns span the column and row spaces of **X**, respectively, and Σ is an invertible $d \times d$ matrix. One such way to carry out the decomposition is the singular value decomposition (SVD), which is discussed briefly next.

3.6.1.1 SVD

The tool used to decompose the range space of the data matrix into two complimentary subspaces is the singular value decomposition (SVD).

The SVD is computationally very robust and allows for high resolution discrimination against noise space or contamination.

In light of the earlier discussion of the subspaces, the singular value decomposition of an $M \times N$ matrix $\mathbf{X}$, which is assumed to have rank d, gives rise to following decomposition:

$$\mathbf{X} = \mathbf{U} \sum \mathbf{V}^H = [\mathbf{U}_s \mathbf{U}_o] \begin{bmatrix} \Sigma_s & 0 \\ 0 & \Sigma_o \end{bmatrix} \begin{bmatrix} \mathbf{V}_s^H \\ \mathbf{V}_o^H \end{bmatrix} \tag{3.58}$$

where Σ is an $M \times N$ diagonal matrix containing the singular values σ_{si} of $\mathbf{X}$. These are positive numbers ordered in the following manner

$$\sigma_{s1} \geq \sigma_{s2} \geq \ldots \geq \sigma_{sd} \geq \sigma_{s\,d+1} = \ldots = 0 \tag{3.59}$$

Note that only d singular values are nonzero. The d columns of $\mathbf{U}_s$ corresponding to these nonzero singular values span the column space of $\mathbf{X}$ and are called left singular vectors. Similarly, the d rows of $\mathbf{V}_o$ are called right singular vectors and span the row space of $\mathbf{X}$ (or the column space of $\mathbf{X}^H$). However, only the d largest singular values are of interest to us.

Another way of decomposition is the eigenvalue decomposition on of data covariance matrix $\mathbf{X}\mathbf{X}^H$. The main difference between these two approaches is that SVD-based algorithms operate directly on the data matrix $\mathbf{X}$ instead of "squared" matrix $\mathbf{X}\mathbf{X}^H$, and thus make them more effective in finite precision computations.

In principle, the subspace-based methods search for directions such that the steering vectors associated with these directions are orthogonal to the noise subspace and are contained in the signal subspace. Once the signal subspace has been determined, the data model parameters are extracted from it. This insight gives rise to a number of subspace based approaches. Associated with each of these approaches is a certain algorithm: a computational scheme. In the following section, one such revolutionary technique called Multiple Signal Classification (MUSIC) is discussed. Another pioneering algorithm of this class, Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT), is described in Chapter 5.

3.6.2 MUSIC

MUSIC (Multiple Signal Classification) is one of the earliest proposed and a very popular method for super-resolution direction finding [13]. In terms of the data model described in Section 3.1, the input data covariance matrix $\mathbf{R}_{xx}$ is written as

$$\mathbf{R}_{xx} = \mathbf{A}\mathbf{R}_{ss}\mathbf{A}^H + \sigma_N^2\mathbf{I}_M \tag{3.60}$$

where $\mathbf{R}_{ss}$ is the signal correlation matrix, σ_N^2 is the noise common variance, and $\mathbf{I}_M$ is the identity matrix of rank M.

Suppose that the eigenvalues of $\mathbf{R}_{xx}$ are $\{\lambda_1, \ldots, \lambda_M\}$ so that

$$\left|\mathbf{R}_{xx} - \lambda_i\mathbf{I}_M\right| = 0 \tag{3.61}$$

Then, substitution of (3.60) into (3.61) reads

$$\left|\mathbf{A}\mathbf{R}_{ss}\mathbf{A}^H + \sigma_N^2\mathbf{I}_M - \lambda_i\mathbf{I}_M\right| = 0 \tag{3.62}$$

Assume that $\mathbf{A}\mathbf{R}_{ss}\mathbf{A}^H$ has eigenvalues e_i; then

$$e_i = \lambda_i - \sigma_N^2 \tag{3.63}$$

Since $\mathbf{A}$ is comprised of the steering vectors of an array that are linearly independent, it has full column rank and the signal correlation matrix $\mathbf{R}_{ss}$ is nonsingular as long as the incident signals are not highly correlated [14].

It can be shown that a full column rank $\mathbf{A}$ and a nonsingular $\mathbf{R}_{ss}$ guarantee that when the number of incident signals d is less than number of elements M, the matrix $\mathbf{A}\mathbf{R}_{ss}\mathbf{A}^H$ is positive semidefinite with rank d. This implies that $M - d$ of eigenvalues, e_i of $\mathbf{A}\mathbf{R}_{ss}\mathbf{A}^H$, are zero. From (3.63), this means that $M - d$ of the eigenvalues of $\mathbf{R}_{xx}$ are equal to the noise variance σ_N^2 and they are also the smallest. That is,

$$\lambda_{d+1} = \lambda_{d+1} = \ldots \lambda_M = \lambda_{\min} = \sigma_N^2 \tag{3.64}$$

In other words, once the multiplicity k of the smallest eigenvalue is determined, an estimate of the number of signals, d, can be obtained from the relation

$$d = M - k$$

In practice, however, when the autocorrelation matrix $\mathbf{R}_{xx}$ is estimated from a finite data sample set of the received signals, all the eigenvectors corresponding to the noise space will not be exactly identical. Instead, they will appear as a closely spaced cluster, with the variance of their spread decreasing as the number of samples used to obtain an estimate of $\mathbf{R}_{xx}$ is increased. Therefore, techniques such as AIC and MDL are used. These methods are left for discussion Chapter 4. Here we assume that we get the expected theoretical eigenvalues.

The eigenvector associated with a particular eigenvalue λ_i, denoted as $\mathbf{q}_i$, satisfies

$$(\mathbf{R}_{xx} - \lambda_i \mathbf{I}_M)\mathbf{q}_i = 0, \qquad i = d+1, d+2, \ldots, M \tag{3.65}$$

For these eigenvectors that are associated with the $M - d$ smallest eigenvalues, we can have

$$\begin{aligned}(\mathbf{R}_{xx} - \sigma_i \mathbf{I}_M)\mathbf{q}_i &= \mathbf{A}\mathbf{R}_{ss}\mathbf{A}^H\mathbf{q}_i + \sigma_N^2\mathbf{I}_M\mathbf{q}_i - \sigma_N^2\mathbf{q}_i \\ &= \mathbf{A}\mathbf{R}_{ss}\mathbf{A}^H\mathbf{q}_i = 0\end{aligned} \tag{3.66}$$

Since $\mathbf{A}$ has full rank and $\mathbf{R}_{ss}$ is nonsingular, this implies that

$$\mathbf{A}^H\mathbf{q}_i = 0 \tag{3.67}$$

This means that the eigenvectors associated with the $M - d$ smallest eigenvalues are orthogonal to the d steering vectors that make up $\mathbf{A}$.

$$\{\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_d)\} \perp \{\mathbf{q}_{d+1}, \ldots, \mathbf{q}_M\} \tag{3.68}$$

This remarkable observation forms the cornerstone of almost all of the subspace-based methods. It means that one can estimate the steering vectors associated with the received signals by finding the steering vectors, which are orthogonal to the $M - d$ eigenvectors associated with the eigenvalues of $\mathbf{R}_{xx}$ that are approximately equal to σ_N^2.

This analysis shows that the eigenvectors of covariance matrix $\mathbf{R}_{xx}$ belong to either of the two orthogonal subspaces, one being the principal eigensubspace (signal subspace) and the other being the nonprincipal

eigensubspace (noise subspace). The steering vectors corresponding to the DOA lie in the signal subspace and are hence orthogonal to the noise subspace. By searching through all possible array steering vectors to find those that are perpendicular to the space spanned by the nonprincipal eigenvectors, the DOAs can be determined [7, 8].

To form the noise subspace, a matrix containing the noise eigenvectors needs to be formed:

$$\mathbf{V}_n = [\mathbf{q}_{d+1}, \ldots, \mathbf{q}_M]$$

Since the steering vectors corresponding to signal components are orthogonal to the noise subspace eigenvectors [i.e., (3.43)], $\mathbf{a}^H(\theta)\mathbf{V}_n\mathbf{V}_n^H\mathbf{a}(\theta) = 0$ for $\theta = \theta_i$ corresponding to the DOA of an incoming signal. Then the following MUSIC spectrum is constructed by taking the inverse of $\mathbf{a}^H(\theta)\mathbf{V}_n\mathbf{V}_n^H\mathbf{a}(\theta)$:

$$P(\theta) = P_{MUSIC}(\theta) = \frac{1}{\mathbf{a}^H(\theta)\mathbf{V}_n\mathbf{V}_n^H\mathbf{a}(\theta)} \tag{3.69}$$

The DOAs of the multiple incident signals can be estimated by locating the peaks of the (3.69). The d largest peaks in the MUSIC spectrum above correspond to the DOAs of the signals impinging on the array.

A simulation was conducted by employing a 6-element ULA with its omnidirectional elements separated by a half wavelength. Three equally powered uncorrelated signals were made to impinge on the array from 10°, 20°, and 40°. In such a case, $d = 3$, $M = 6$, $\theta_1 = 10°$, $\theta_2 = 20°$, and $\theta_2 = 40°$. An SNR of 10 dB was assumed. Fifty trials or datasets were taken with each trial (or dataset) averaged over 250 snapshots. Figure 3.8 shows the result.

In comparisons with the methods described so far, MUSIC has the best performance with a resolution of less than 5° in the cases studied.

The MUSIC algorithm can be summarized as follows:

- Step 1: Collect input samples $\mathbf{x}(t_n)$, $n = 1, 2, \ldots, N$ and estimate the input covariance matrix

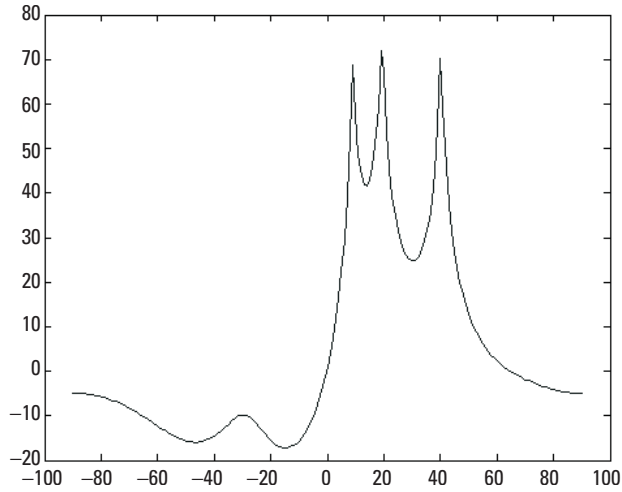


Figure 3.8 DOA estimation with MUSIC; the radio signals impinges at 10°, 20°, and 40°.

$$\mathbf{R}_{xx} \approx \hat{\mathbf{R}}_{xx} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}(t_n)\mathbf{x}^H(t_n) \tag{3.70}$$

- Step 2: Perform eigendecomposition on $\hat{\mathbf{R}}_{xx}$

$$\hat{\mathbf{R}}_{xx}\mathbf{V} = \mathbf{V}\Lambda \tag{3.71}$$

where $\Lambda = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_M\}$, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_M$, are the eigenvalues and $\mathbf{V}$ contains all the corresponding eigenvectors of $\hat{\mathbf{R}}_{xx}$.

- Step 3: Estimate the multiplicity k of the smallest eigenvalue $\lambda_{\min}$ and then the number of signals d from as

$$d = M - k$$

- Step 4: Compute the MUSIC spectrum

$$P(\theta) = P_{MUSIC}(\theta) = \frac{1}{\mathbf{a}^H(\theta)\mathbf{V}_n\mathbf{V}_n^H\mathbf{a}(\theta)} \tag{3.72}$$

where $\mathbf{V}_n = [\mathbf{q}_{d+1}, \ldots, \mathbf{q}_M]$ with $\mathbf{q}_l$, $l = d+1, d+2, \ldots, M$ being the eigenvectors corresponding the smallest eigenvalue $\lambda_{\min}$.

- Step 5: Find the d largest peaks of $P_{MUSIC}(\theta)$ to obtain DOA estimates.

3.6.3 Minimum Norm

The minimum norm method is applicable for linear arrays and can be considered as an improved version of MUSIC in computing DOA estimates. The general expression for the minimum norm method is to search for the locations of the peaks in the power spectrum here:

$$P(\theta) = P_{MIN}(\theta) = \frac{1}{\left|\mathbf{w}^H \mathbf{a}(\theta)\right|} \tag{3.73}$$

with an array weight **w,** which is of minimum norm. The weight vector **w** should have its first element equal to unity and is contained in the noise subspace [15]. The final form of the power spectrum is then:

$$P(\theta) = P_{MIN}(\theta) = \frac{1}{\left|\mathbf{a}^H(\theta)\mathbf{V}_n \mathbf{V}_n^H \mathbf{W}\mathbf{V}_n \mathbf{V}_n^H a(\theta)\right|} \tag{3.74}$$

with the vector $\mathbf{W} = p_1 p_1^T$ where p_1 equals the first column of an $M \times M$ identity matrix.

Equation (3.74) can be seen as a squaring of the denominator of the power equation of (3.72) of MUSIC. As the denominator is squared, near zero values should serve to boost the power output to even higher levels. The $\mathbf{W}$ matrix is necessary to ensure that the matrix dimensions match mathematically

A simulation was conducted by employing a 6-element ULA with its omnidirectional elements separated by a half wavelength. Three equally powered uncorrelated signals were made to impinge on the array from 5°, 25°, and 45°. In such a case, $d = 3$, $M = 6$, $\theta_1 = 5°$, $\theta_2 = 25°$, and $\theta_2 = 45°$. An SNR of 10 dB was assumed. Fifty trials were taken with each trial averaged over 250 snapshots. Figure 3.9 shows the result. As can be seen from the figure, the three peaks are sharp with very good resolutions.

3.6.4 ESPRIT

Due to its simplicity and high resolution capability, ESPRIT has become one of the most popular signal subspace-based DOA estimating schemes. ESPRIT is applicable to array geometries that are composed of two identical subarrays and is restricted to use with array geometries that exhibit invariances. This requirement, however, is not very prohibitive in practical applications since many of the common array geometries used in practice exhibit these invariances [16]. There are three primary steps in any ESPRIT based DOA estimation algorithm:

1. *Signal subspace estimation:* Computation of a basis matrix for the estimated signal subspace.
2. *Solution of the invariance equation*: Solution of an (in general) overdetermined system of equations, the invariance equation, derived from the basis matrix.
3. *DOA estimation:* Computation of the eigenvalues of the solution of the invariance equation formed in step 2.

This algorithm will be explained and described in depth in Chapter 5, as this forms the subject of this book.

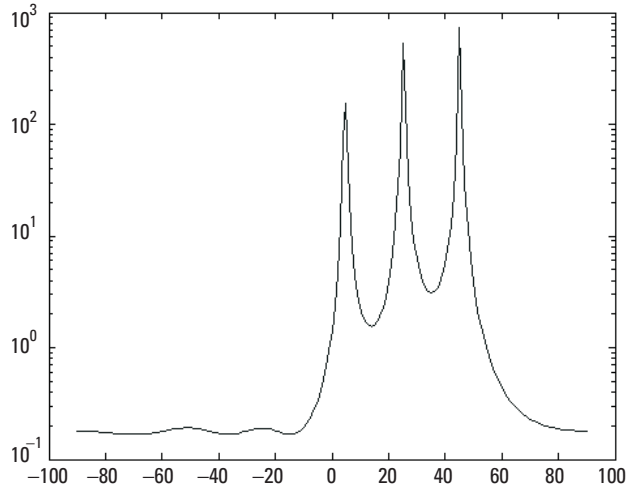


Figure 3.9 DOA estimation with the minimum norm technique.

3.7 Conclusion

This chapter discussed the signal and data model that we are using for this book. Centro-symmetric arrays that are demanded by many DOA estimation algorithms are reviewed. A few well-known DOA estimation techniques of simple to complicated algorithms are described; they are all based on the power spectrum that is inversely proportional to covariance of received signals. Preliminary computations were performed to illustrate the performance of these DOA estimation schemes.

References

[1] Haardt, M., *Efficient One-, Two-, and Multidimensional High-Resolution Array Signal Processing*, New York: Shaker Verlag, 1997.

[2] Krim, H., and M. Viberg, "Two Decades of Array Signal Processing Research," *IEEE Signal Processing Magazine*, Vol. 13, No. 4, July 1996, pp. 67–94.

[3] Balanis, C. A., *Antenna Theory: Analysis and Design*, 3rd ed., New York: Wiley, 2005.

[4] Xu, G., R. H. Roy, and T. Kailath, "Detection of Number of Sources Via Exploitation of Centro-Symmetry Property," *IEEE Trans. on Signal Processing*, Vol. 42, No. 1, January 1994, pp. 102–112.

[5] Lee, A., "Centrohermitian and Skew-Centrohermitian Matrices," *Linear Algebra and Its Applications*, Vol. 29, 1980, pp. 205–210.

[6] Haardt, M., et al., "2D Unitary ESPRIT for Efficient 2D Parameter Estimation," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Vol. 3, May 9–12, 1995, pp. 2096–2099.

[7] Liberti, Jr., J. C., and T. S. Rappaport, *Smart Antennas for Wireless Communications: IS-95 and Third Generation CDMA Applications*, Upper Saddle River, NJ: Prentice-Hall, 1999.

[8] Litva, J., and T. K. Y. Lo, *Digital Beamforming in Wireless Communications*, Norwood, MA: Artech House, 1996.

[9] Capon, J., "High-Resolution Frequency-Wavenumber Spectrum Analysis," *Proc. IEEE*, Vol. 57, No. 8, August 1969, pp. 2408–2418.

[10] Makhoul, J., "Linear Prediction: A Tutorial Review," *Proc. IEEE*, Vol. 63, No. 4, April 1975, pp. 561–580.

[11] Ziskind, I., and M. Wax, "Maximum Likelihood Localization of Multiple Sources by Alternating Projection," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. 36, No. 10, October 1998, pp. 1553–1560.

[12] vander Veen, A. J., E. F. Deprettere, and A. L. Swindlehurst, "Subspace Based Signal Analysis Using Singular Value Decomposition," *Proc. IEEE*, Vol. 81, No. 9, September 1993, pp. 1277–1308.

[13] Schmidt, R. O., "Multiple Emitter Location and Signal Parameter Estimation," *IEEE Trans. on Antennas Propagation*, Vol. 34, No. 3, March 1986, pp. 276–280.

[14] Wax, M., "Detection and estimation of superimposed signals," Ph.D. dissertation, Stanford University, Stanford, CA, 1985.

[15] Ronhovde, A., et al., "High-Resolution Beamforming for Multibeam Echo Sounders Using Raw EM3000 Data," http://www.ifi.uio.no/sverre/papers/Oceans.pdf.

[16] Roy, R., and T. Kailath, "ESPRIT-Estimation of Signal Parameters Via Rotational Invariance Techniques," *IEEE Trans. on Acoust., Speech, Signal Processing*, Vol. 37, No. 7, July 1989, pp. 984–995.

4

Preprocessing Schemes and Model Order Estimation

4.1 Introduction

The DOA estimation algorithms described in the previous chapter assume noncoherent impinging signals. However, if the impinging signals are highly correlated or coherent to each other, most of the algorithms will fail to give reliable DOA estimates because the data covariance matrix $\mathbf{R}_{xx}$ received by an array becomes singular or ill-conditioned. This situation is not uncommon in multipath scenarios. In a rich multipath environment, the signals impinging on the array are often delayed and scaled versions of each other and hence are highly correlated or coherent. To circumvent this problem, the data covariance matrix is processed before "feeding" them to the DOA estimation algorithms. These techniques are called preprocessing techniques and they play a vital, if not a mandatory, role in estimating a DOA. Two well-known and established preprocessing schemes, namely spatial smoothing and forward backward averaging, will be discussed in this chapter.

In addition to the signal coherence issue, the number of signals impinging on the array is assumed to be known so far. In reality, however, this number is not known and has to be estimated from the data received. The DOA estimation algorithms depend completely on the assumption of knowledge of number of impinging signals. Hence, estimation of this number plays another key role in a DOA estimation process and invites

research on a separate class of techniques called model order estimation techniques. Model order or source order estimators are mandatory algorithms that are to be run before the DOA estimation algorithms are executed. This chapter gives a brief introduction to some of these techniques.

4.2 Preprocessing Schemes

The subspace-based DOA estimation algorithms are completely based on the full rank condition of the data covariance matrix $\mathbf{R}_{xx}$ given in (2.20). It is written here again for ready reference.

$$\mathbf{R}_{xx} = E\left\{\mathbf{x}(t)\mathbf{x}^H(t)\right\} = \mathbf{A}\mathbf{R}_{ss}\mathbf{A}^H + \sigma_N^2\mathbf{I}_M \tag{4.1}$$

The condition of $\mathbf{R}_{xx}$ depends on that of signal covariance matrix $\mathbf{R}_{ss}$. When the d impinging signal wavefronts are not correlated, the signal covariance matrix $\mathbf{R}_{ss}$ has full rank d; it is diagonal and nonsingular. $\mathbf{R}_{ss}$ becomes nondiagonal and singular when the signals are partially correlated, or when some signals are fully correlated (coherent). For instance, if one of the impinging signal wavefronts is highly correlated or coherent with another, their cross-correlation coefficient has a magnitude of 1 and the rank of $\mathbf{R}_{ss}$ is now reduced to $d-1$. Therefore, the full rank condition in the previously described DOA algorithms is no longer satisfied (i.e., rank $\{\mathbf{A}\mathbf{R}_{ss}\mathbf{A}^H\} = d-1$); a DOA algorithm will then fail. In general, if there are P coherent wavefronts, $\mathbf{R}_{ss}$ is of rank $M-P$. In such a case, $\mathbf{R}_{xx}$ does not have d large positive eigenvalues, as it otherwise does when the signals are uncorrelated and $\mathbf{R}_{ss}$ is diagonal. This implies that the matrix $\mathbf{R}_{ss}$ is likely to have fewer than d large positive eigenvalues, creating errors in the DOA estimation processes.

For instance, assume that two of the impinging signals are coherent; say the first two are coherent, that is, $s_2(t) = \alpha s_1(t)$, with α being a complex scalar describing the gain and phase relationship between the two coherent signals. Then $s(t)$ is a signal vector given by

$$\mathbf{s}(t) = \left[s_1(t) \;\; \alpha s_1(t) \;\; s_3(t) \ldots s_d(t)\right]^H \tag{4.2}$$

and the array steering matrix is given by

$$\mathbf{A} = \left[\mathbf{a}(\theta_1)\ \mathbf{a}(\theta_2)\ \mathbf{a}(\theta_3)\ldots \mathbf{a}(\theta_d)\right] \tag{4.3}$$

With these modified signals, the signal covariance matrix $\mathbf{R}_{ss} = \mathrm{E}\{\mathbf{s}(t)\mathbf{s}^H(t)\}$ is now a $d \times d$ singular matrix and no longer nonsingular. The following two errors can be observed:

1. The multiplicity of the smallest eigenvalues of $\mathbf{R}_{xx}$ now is $k = M - d + 1$ and not $k = M - d$. This gives an error in estimating the number of signals as $d - 1 (= M - k)$ instead of d.
2. Only $\{\theta_3, \ldots, \theta_d\}$ can be resolved due to the reduced number of resolvable largest eigenvalues of $\mathbf{R}_{xx}$.

To deal with the issue of the coherent signal reception, a preprocessing scheme such as forward-backward averaging or spatial smoothing can be applied; they ensure $\mathbf{R}_{ss}$ (after being smoothed) to be of full rank and nonsingular even when all the received signals are coherent. In other words, the schemes are used basically to decorrelate the signals before estimating their DOAs [1, 2].

4.2.1 Forward-Backward Averaging

Forward-backward (FB) averaging is a popular preprocessing scheme employed in many signal processing applications including radio direction finding. Forward-backward averaging can be implemented if the arrays are centro-symmetric in nature. The basic operation of forward backward averaging relies on the fact that the steering vectors of ULA remains the same, even if their elements are reversed and complex conjugated.

Let Π_M be an $M \times M$ exchange matrix. Then for ULA it can be shown that [1]

$$\Pi_M \bar{\mathbf{a}}(\theta_i) = e^{-j(M-1)\mu_i}\mathbf{a}(\theta_i) \tag{4.4}$$

A backward data covariance matrix $\mathbf{R}_{back}$ can then be constructed from the actual data covariance matrix $\mathbf{R}_{xx}$ as

$$R_{back} = \Pi_M \bar{\mathbf{R}}_{xx} \Pi_M \tag{4.5}$$

Consider a simple case of two coherent sources impinging on a ULA. The forward-backward averaged data covariance matrix is obtained by averaging the actual (forward) covariance matrix and the backward covariance matrix, that is,

$$\begin{aligned}\mathbf{R}_{xx}^{fb} &= \frac{1}{2}\left(\mathbf{R}_{xx} + \mathbf{R}_{back}\right) = \frac{1}{2}\left(\mathbf{R}_{xx} + \Pi_M \overline{\mathbf{R}}_{xx} \Pi_M\right) \\ &= \frac{1}{2}\left\{E\left\{\mathbf{x}(t)\mathbf{x}^H(t)\right\} + \Pi_M E\left\{\overline{\mathbf{x}}(t)\overline{\mathbf{x}}^H(t)\right\}\Pi_M\right\} \\ &= \mathbf{A}\frac{1}{2}\left(\mathbf{R}_{ss} + \Delta\mathbf{R}_{ss}\Delta^H\right)\mathbf{A}^H + \sigma_n^2 \mathbf{I}_M\end{aligned} \tag{4.6}$$

with some unitary diagonal matrix $\Delta \in \mathbf{C}^{d\times d}$ [3]. Comparing (4.6) with (4.1), the new signal covariance matrix is now given by

$$\mathbf{R}_{ss}^{fb} = \frac{1}{2}\left(\mathbf{R}_{ss} + \Delta\overline{\mathbf{R}}_{ss}\Delta^H\right) \tag{4.7}$$

This new forward backward signal covariance matrix has rank d even if the two signals are coherent; it enables the separation of two coherent or highly correlated signals. A forward backward averaged edition of any covariance based algorithm can thus be obtained by replacing $\hat{\mathbf{R}}_{xx}$ with $\hat{\mathbf{R}}_{xx}^{fb}$ in a DOA algorithm. This technique is sometimes used even in uncorrelated signal environments to obtain more reliable estimates.

In practice, to compute $\mathbf{R}_{xx}^{fb}$ with data $\mathbf{X} \in \mathbf{C}^{M\times N}$ received by an array, a new extended data matrix $\mathbf{Z}$ is used; it is defined as

$$\mathbf{Z} = \begin{bmatrix}\mathbf{X} & \Pi_M \mathbf{X} \Pi_N\end{bmatrix} \in \mathbf{C}^{M\times 2N} \tag{4.8}$$

An estimate of $\mathbf{R}_{xx}^{fb}$ can then be calculated as

$$\hat{\mathbf{R}}_{xx}^{fb} = \frac{1}{2N}\mathbf{Z}\mathbf{Z}^H = \frac{1}{2N}\left(\mathbf{X}\mathbf{X}^H + \Pi_M \overline{\mathbf{X}\mathbf{X}}^H \Pi_M\right) \tag{4.9}$$

In Chapter 5, it shall be shown that the forward backward averaging is achieved inherently in a unitary ESPRIT algorithm.

4.2.2 Spatial Smoothing

In a rich multipath area, we may encounter more than two coherent signals. If this is the case, forward-backward averaging alone is not able to decorrelate the signals. Spatial smoothing is another preprocessing technique that can be used to tackle this problem. In spatial smoothing, the antenna array is divided into a number of smaller overlapping subarrays and the data covariance matrices obtained from each subarray are averaged [1–3]. In this section, spatial smoothing techniques applied to ULA and URA are described. These techniques are later applied with ESPRIT when dealing with coherent signals.

4.2.2.1 One-Dimensional Spatial Smoothing

Consider a ULA consisting of M elements as shown in Figure 4.1. Let this ULA be divided into L subarrays, each containing $M_{sub} = M - L + 1$ elements. There are two types of the divisions shown in Figure 4.1; the upper arrays are called the forward subarrays and the lower arrays are called the backward arrays; either of these types can be used.

Consider the forward or upper subarrays first. The number and size of the subarrays are determined from the number of sources under

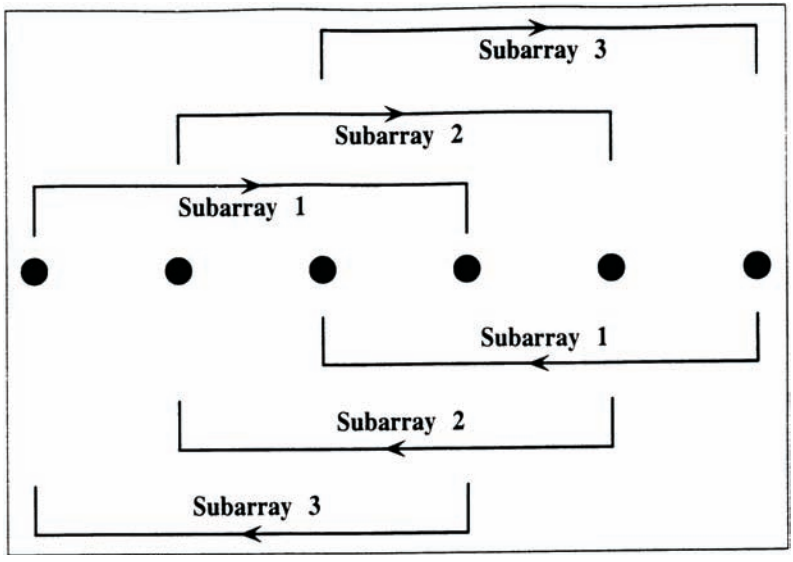


Figure 4.1 An example of two different subarrays for one dimensional smoothing. The upper subarrays are called forward subarrays and the lower subarrays are called the backward subarrays.

consideration. The approach is to form a covariance matrix of each of these subarrays and average them. The formation of the subarrays can be represented mathematically by a selection matrix. For instance, for the lth subarray, the selection matrix is formed here [2, 3]:

$$\mathbf{J}_l^{(M)} = \left[\mathbf{0}\ \mathbf{I}_{M_{sub}}\ \mathbf{0}\right] \in R^{M_{sub} \times M}, \quad 1 \le l \le L \tag{4.10}$$

The data matrix corresponding to the lth subarray can then be given as

$$\mathbf{X}_l = \mathbf{J}_l^{(M)}\mathbf{X} = \mathbf{A}_1 \Phi^{l-1}\mathbf{S} + \mathbf{J}_l^{(M)}\mathbf{N}, \quad 1 \le l \le L \tag{4.11}$$

where $\mathbf{A}_1 = \mathbf{J}_1^{(M)}\mathbf{A}$. The data covariance matrix of the lth subarray is then obtained as

$$\mathbf{R}_l = \mathbf{A}_1 \Phi^{l-1}\mathbf{S} + \mathbf{J}_l^{(M)}\mathbf{N}, \quad 1 \le l \le L \tag{4.12}$$

The "spatially smoothed data" matrix of the complete array is now obtained from the data matrices of the individual subarrays as

$$\mathbf{X}_{ss} = \left[\mathbf{J}_1^{(M)}\mathbf{X}\ \ \mathbf{J}_2^{(M)}\mathbf{X} \ldots\ \mathbf{J}_L^{(M)}\mathbf{X}\right] \in C^{M_{sub} \times NL} \tag{4.13}$$

Finally, the spatially smoothed data covariance matrix is obtained by taking average of data covariance matrices of the individual subarrays as shown next:

$$\begin{aligned}
\mathbf{R}_{xx}^{ss} &= \frac{1}{L}\sum_{l=1}^{L}\mathbf{R}_l \\
&= \frac{1}{L}\sum_{l=1}^{L}\mathbf{J}_l^{(M)}\mathbf{R}_{xx}\mathbf{J}_l^{(M)T} \\
&= \mathbf{A}_1\left(\frac{1}{L}\sum_{l=1}^{L}\Phi^{l-1}\mathbf{R}_{ss}\overline{\Phi}^{l-1}\right)\mathbf{A}_1^H + \sigma^2\mathbf{I}_{M_{sub}} \\
&= \mathbf{A}_1\mathbf{R}_{ss}^{ss}\mathbf{A}_l^H + \sigma^2\mathbf{I}_{M_{sub}}
\end{aligned} \tag{4.14}$$

where $\mathbf{R}_{ss}^{ss} = \frac{1}{L}\sum_{l=1}^{L} \Phi^{l-1}\mathbf{R}_{ss}\overline{\Phi}^{l-1}$ is the newly obtained spatially smoothed signal covariance matrix. Consequently, if there are L coherent wavefronts, the rank of $\mathbf{R}_{ss}$ is $d - L$, but the spatially smoothed signal covariance matrix $\mathbf{R}_{xx}^{ss}$ will have the required rank d. Therefore the direction of arrivals of all d signals can be estimated.

It can be observed from (4.13) that the number of available snapshots is extended from N to NL, whereas the effective aperture of the array is reduced from M to M_{sub}. This is the price for the smoothing. Conventionally, if there are d signals, the minimum number of elements required is $M = d + 1$. Now, when the array is divided into L subarrays, the number of elements in each subarray is $M_{sub} = d + 1$. Thus, for detecting d coherent signals, at least $M_{sub} = d + 1$ elements are needed.

This can be considered as forward spatial smoothing as the antenna array is divided into subarrays in the forward direction (the upper subarrays in Figure 4.1). Similarly, the array can be divided and averaged in the backward direction (the lower subarrays in Figure 4.1). It is then called backward spatial smoothing and the result is the similar to that of forward spatial smoothing.

If spatial smoothing is applied by taking the average in both forward and backward directions (i.e., the forward and backward subarrays), an improvement can be achieved in the array size. By simultaneous use of forward and backward subarray averaging scheme, it was shown [4, 5] that the number of elements required to estimate d coherent signals can be reduced to $(3d/2)$. For example, if there are four correlated signals, the minimum number of elements required with forward or averaged smoothing technique along is $2d = 8$. If forward-backward averaged spatial smoothing is employed, the minimum number of antenna elements required to decorrelate and estimate the DOAs is equal to $(3d/2) = 6$. However, if the four signals are uncorrelated, only five elements are sufficient to estimate the DOAs.

A simulation of MUSIC algorithm with and without forward/backward smoothing is shown in Figure 4.2. Two coherent signals of equal power with SNRs of 20 dB arrive at a six-element array with an interelement spacing of a half-wavelength at 10° and 20°. The linear array is divided into two subarrays with five elements in each. It can be seen from Figure 4.2 that the conventional MUSIC without smoothing fails

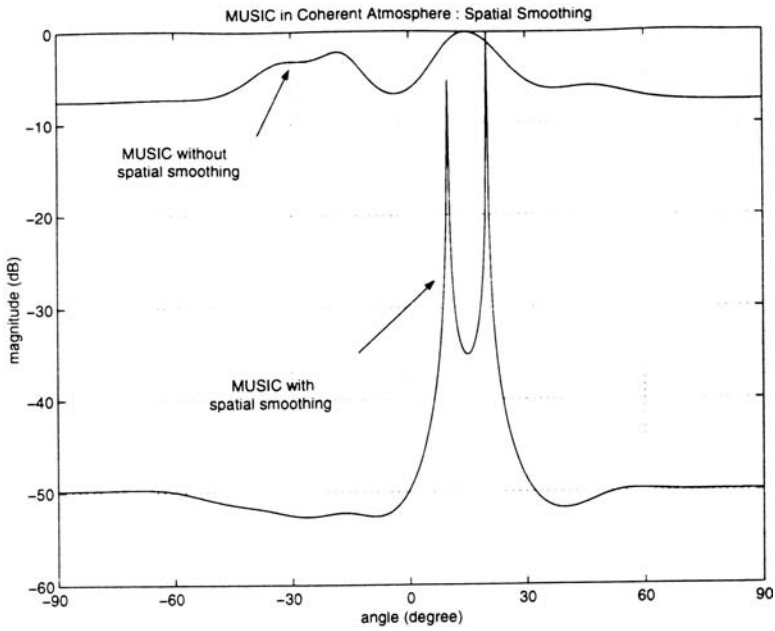


Figure 4.2 Effect of spatial smoothing.

miserably in this coherent environment while the spatial smoothing identified the two angles clearly with two peaks of output.

4.2.2.2 Two-Dimensional Spatial Smoothing

In this section, the method of spatial smoothing is extended to two-dimensional arrays, which are used to find joint azimuthal and elevational angles of a signal source. Two-dimensional spatial smoothing can thus be used as a preprocessing scheme to decorrelate coherent signals and extend the number of available snapshots. The key is to consider the rectangular array as linear arrays in the x and y directions. Then, the smoothing operation can still be applied in the same line as in the one-dimensional case [3, 7, 8].

Consider a URA of size $M = 4 \times 5$ elements as shown in Figure 4.3. The ULA in the y direction contains $M_y = 5$ elements. Now, divide each of these linear arrays into $L_y = 3$ subarrays. Then each subarray will have $M_{sub_y} = M_y - L_y + 1 = 3$ elements in the y direction, as shown in Figure 4.3.

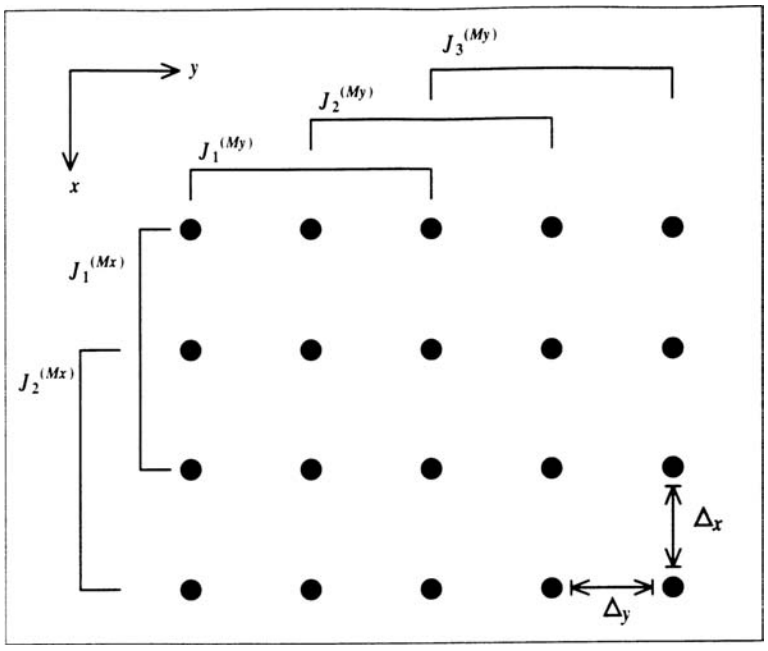


Figure 4.3 An example of two different subarrays for two-dimensional spatial smoothing.

Similarly, ULAs in the x direction contain $M_x = 4$ elements. By dividing each of these linear arrays into $L_x = 2$ subarrays, each subarray contains $M_{sub_y} = M_y - L_y + 1 = 3$ elements in the x direction.

As a result, the whole URA of size $M = M_x \times M_y$ elements is divided into $L = L_x \times L_y$ small rectangular subarrays, each containing $M_{sub} = M_{sub_x} \times M_{sub_y}$ elements. In this example, we divide an $M = 20$ element URA into $L = 2 \times 3 = 6$ rectangular subarrays, each having $M_{sub} = 3 \times 3 = 9$ elements.

Once the linear subarrays in the x and y directions are chosen, the selection matrices corresponding to each of these subarrays can be formed. As described before, the one-dimensional selection matrix corresponding to the l_x^{th} subarray of ULA in the x direction is defined as

$$\mathbf{J}_{l_x}^{(M_x)} = \left[\mathbf{0} \quad \mathbf{I}_{M_{sub_x}} \quad \mathbf{0}\right], \; 1 \le l_x \le L_x \tag{4.15}$$

Similarly, the one-dimensional selection matrices for the subarrays in the y direction are given by

$$\mathbf{J}_{l_y}^{(M_y)} = \begin{bmatrix} \mathbf{0} & \mathbf{I}_{M_{sub\,y}} & \mathbf{0} \end{bmatrix}, \; 1 \le l_y \le L_y \tag{4.16}$$

Suppose that data matrix $\chi(t_n)$ is obtained as derived in Section 3.3.2. One-dimensional spatial smoothing is then applied along the x direction rows of $\chi(t_n)$ and the y direction columns of $\chi(t_n)$ using the one-dimensional selection matrices defined in (4.15) and (4.16), respectively. For the $L = L_x \times L_y$ rectangular subarrays, the two-dimensional selection matrices for each of these subarrays can be obtained from the $(L_x + L_y)$ one-dimensional selection matrices of linear arrays as

$$\mathbf{J}_{l_x,l_y} = \mathbf{J}_{l_y}^{(M_y)} \otimes \mathbf{J}_{l_x}^{(M_x)}, \; 1 \le l_x \le L_x, \; 1 \le l_y \le L_y \tag{4.17}$$

where $\otimes$ denotes the Kronecker product (see Appendix, Section A.1). The spatially smoothed data matrix is then given similar to (4.13) as

$$X_{ss} = \begin{bmatrix} \mathbf{J}_{1,1}\mathbf{X} & \mathbf{J}_{1,2}\mathbf{X} & \cdots & \mathbf{J}_{1,L_y}\mathbf{X} & \mathbf{J}_{2,1}\mathbf{X} \cdots \mathbf{J}_{L_x,L_y}\mathbf{X} \end{bmatrix} \in C^{M_{sub} \times NL} \tag{4.18}$$

This two-dimensional spatial smoothing scheme shall be used as the preprocessing step for two-dimensional ESPRIT algorithms.

4.3 Model Order Estimators

As discussed before, in the real world the number of signals impinging on the array is not known and has to be estimated from the signal data obtained from the array. Since all the DOA estimation algorithms requires a priori knowledge of this number, techniques whose function is to estimate the number of signals need to be run prior to running the DOA estimation algorithms. Such techniques are usually called model order estimation techniques. In the following section, a few of these techniques are described.

4.3.1 Classical Technique

This is the simplest and the most direct way of estimating the number of signals where the estimate can be determined from the data covariance matrix [8]. It is based on the fact that the number of signals impinging on an array is the same as number of "large" eigenvalues of the data covariance matrix. Hence, an estimate of the number of sources can be obtained from an estimate of the number of repeated small eigenvalues other than the large ones. As defined before, the spatial data covariance matrix is given as

$$\mathbf{R}_{xx} = E\left\{\mathbf{x}(t)\mathbf{x}^H(t)\right\} = \mathbf{A}\mathbf{R}_{ss}\mathbf{A}^H + \sigma_N^2\mathbf{I}_M \tag{4.19}$$

Theoretically, $M - d$ small eigenvalues of $\mathbf{R}_{xx}$ have the minimum value equal to the noise variance σ_N^2. In other words, if the multiplicity k of this smallest eigenvalue is found, an estimate of the number of signals, d, can be obtained directly as

$$d = M - k \tag{4.20}$$

In practice, however, when the covariance data matrix $\mathbf{R}_{xx}$ is estimated from a set of finite data samples, the smallest eigenvalues corresponding to noise power will not be identical. Instead they will appear as a closely spaced cluster, with the variance of their spread decreased as the number of data samples is increased. Therefore, various statistical methods have been proposed to test for the equality or closeness of eigenvalues in order to determine the multiplicity k or simply the number d. Two such criteria are discussed in the following sections.

4.3.2 Minimum Descriptive Length Criterion

The minimum descriptive length (MDL) criterion is an old and standard technique for estimating the number of signals [9]. In the MDL-based approach, the number of signals d is obtained as the value of $d \in \{0, 1, \ldots, M-1\}$, which minimizes the following criterion:

$$MDL(d) = -\log\left\{\frac{\prod_{i=d+1}^{M}\lambda_i^{\frac{1}{(M-d)}}}{\frac{1}{M-d}\sum_{i=d+1}^{M}\lambda_i}\right\} + \frac{1}{2}\hat{p}(k)\log N \tag{4.21}$$

where M denotes number of the elements, N is the number of snapshots, λ_i are the eigenvalues of the estimated data covariance matrix, and $\hat{p}(k)$ is a function of number of independent parameters called penalty function. Depending upon the properties of the covariance matrix and preprocessing scheme applied, the penalty function is given by

$$\hat{p}(k) = \begin{Bmatrix} d(2M-d+1) & \text{for real matrices} \\ d(M+d+1) & \text{for real FB}-\text{averaging} \\ d(2M-d) & \text{for complex matricies} \\ 0.5d(2M-d+1) & \text{for complex FB}-\text{averaging} \end{Bmatrix} \quad (4.22)$$

The MDL criterion is comprised of two parts. The first part is the log-likelihood function (4.21) and the second part is a penalty factor to punish the increase of the model order. The model order is obtained by varying the model order d and finding the d that minimizes the MDL criterion. When d is increased to $d+1$ (more than the actual number of signals), the influence of an additional eigenvalue and the corresponding eigenvector is added to the data covariance matrix of the data model; the result will cause the first part of the MDL criterion to decrease but at the same time the second part to increase. Therefore, the model order d is a value that minimizes (4.21), a compromised fit to the data model and the penalty factor.

In case the impinging signals are coherent to each other, spatial smoothing or another preprocessing technique has to be applied to decorrelate the signals before applying (4.21). In this case, the number of elements M refers to the size of the subarrays and the number of snapshots in (4.21) is taken as the number obtained after applying smoothing and/or forward-backward averaging.

For a ULA with M elements and N snapshots, the maximum number of signals that can be estimated by this criterion is given by

$$d_{\max} = \{M, N\} \quad (4.23)$$

For a URA with $M_x \times M_y$ elements and N snapshots, the maximum possible model order obtained by this criterion is given by

$$d_{\max} = \{M_x \cdot M_y, N\} \quad (4.24)$$

The MDL criterion is simulated in two experiments with a 10-element ULA under SNR of 0 dB. Results are averaged over 50 trials (or datasets) with 250 snapshots per trial (or dataset). In the first experiment, two uncorrelated signals were impinged on the array. In the second experiment, four uncorrelated signals were considered to be impinging on the array. As seen in Figure 4.4, the MDL function achieves a minimum at the numbers of the signals, 2 and 4, respectively. In other words, the MDL method successfully estimates the numbers of the signals impinging on the array.

4.3.3 Akaike Information Theoretic Criterion

Similar to MDL, a technique called the Akaike Information Theoretic Criterion (AIC) was proposed for estimation of the number of signals or model order [10, 11]; the number of signals d is determined as the argument that minimizes the following criterion

$$AIC(d) = -\log\left\{\frac{\prod_{i=d+1}^{M}\lambda_i^{\frac{1}{(M-d)}}}{\frac{1}{M-d}\sum_{i=d+1}^{M}\lambda_i}\right\}^{(M-d)N} + \hat{p}(k) \qquad (4.25)$$

Here again, the first term is derived directly from a log-likelihood function, and the second term is the penalty factor added by the AIC.

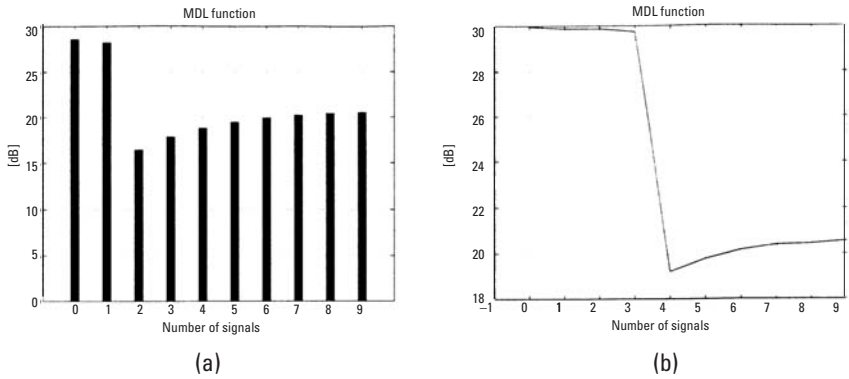


Figure 4.4 Histograms of the MDL for estimating the number of signals. (a) Two signals impinging and (b) four signals impinging.

Similar to the MDL criterion, this criterion is used after the spatial smoothing is applied to the array.

Depending upon the properties of the covariance matrix and the preprocessing scheme applied, the penalty function is given by

$$\hat{p}(k) = \begin{Bmatrix} d(2M-d+1) & \text{for real matrices} \\ d(M+d+1) & \text{for real FB} - \text{averaging} \\ d(2M-d) & \text{for complex matrices} \\ 0.5d(2M-d+1) & \text{for complex FB} - \text{averaging} \end{Bmatrix} \quad (4.26)$$

Also, the maximum number of signals that can be estimated is given in a way similar to MDL. For a ULA with M elements and N snapshots, the maximum number of the signals determined by this criterion is given by

$$d_{\max} = \{M, N\} \quad (4.27)$$

For a URA with $M_x \times M_y$ elements and N snapshots, the maximum number of the signals is given by

$$d_{\max} = \{M_x \cdot M_y, N\} \quad (4.28)$$

The AIC criterion is simulated in two experiments with a 10-element ULA under SNR of 0 dB. Results are averaged over 50 trials with 250 snapshots per trial. In the first experiment, one signal was impinging on the array. In the second experiment, three signals were impinging on the array. Figure 4.5 shows the results. As seen, the AIC function achieves minimum right at the numbers of the signals impinging. In other words, the AIC method successfully estimates the numbers of the signals impinging.

4.4 Conclusion

In this chapter, techniques to deal with correlated signals and to determine the number of signals are described with literature references provided for further detailed presentations of each technique. Simulations were run and results are presented to give a better understanding of the techniques. As observed, these techniques play a vital role in DOA

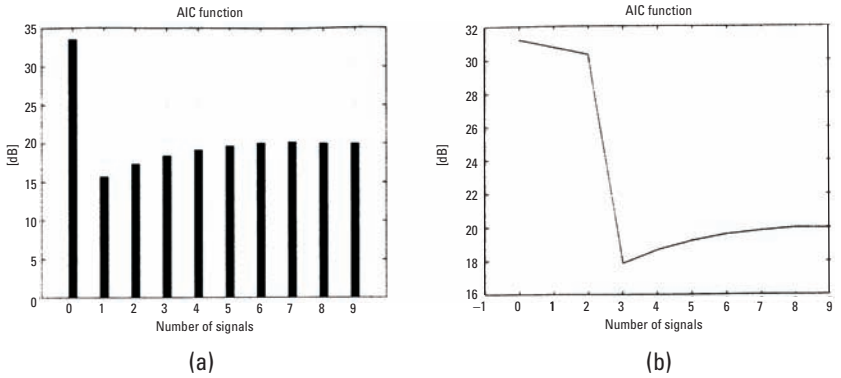


Figure 4.5 Histogram of the AIC for computing the number of signals. (a) Two signals impinging and (b) three signals impinging.

estimation and thus form a mandatory requirement before any DOA algorithms are run. We assume hereafter that while discussing the DOA estimation algorithms, we have had the number of signal sources and have resolved the issue of the correlation among the signals, by applying the techniques described in this chapter. In the following chapters, we will focus on the one of the most popular subspace-based DOA estimation scheme, ESPRIT, with these assumptions.

References

[1] Pillai, S. U., and B. H. Kwon, "Forward/Backward Spatial Smoothing Techniques for Coherent Signal Identification," *IEEE Trans. on Acoustics, Speech and Signal Processing*, Vol. 37, No. 1, January 1989, pp. 8–15.

[2] Evans, J. E., J. R. Johnson, and D. F. Sun, *High Resolution Angular Spectrum Estimation Techniques for Terrain Scattering Analysis and Angle of Arrival Estimation in ATC Navigation and Surveillance System*, M.I.T. Lincoln Lab., Lexington, MA.

[3] M. Haardt, *Efficient One-, Two-, and Multidimensional High-Resolution Array Signal Processing*, New York: Verlag, 1997.

[4] Bachl, R., "The Forward-Backward Averaging Technique Applied to TLS-ESPRIT Processing," *IEEE Trans. on Signal Processing*, Vol. 43, No. 11, November 1995, pp. 2691–2699.

[5] Pillai, S. U., *Array Signal Processing*, New York: Springer-Verlag, 1989.

[6] Wang, H., and K. J. R. Liu, "2-D Spatial Smoothing for Multipath Coherent Signal Separation," *IEEE Trans. on Aerospace and Electronic Systems*, Vol. 34, No. 2, April 1998, pp. 391–405.

[7] Chandna, R., and A. Mahadar, "2D Beamspace ESPRIT with Spatial Smoothing," *IEEE Trans. on Acoust., Speech, Signal Processing*, 1998.

[8] Liberti, Jr., J., and T. S. Rappaport, *Smart Antennas for Wireless Communications: IS-95 and Third Generation CDMA Applications*, Upper Saddle River, NJ: Prentice-Hall, 1999.

[9] Rissanen, J., "Modeling by the Shortest Data Description," *Automatica*, Vol. 14, 1978, pp. 465–471.

[10] Akaike, H., "Information Theory and Extension of the Maximum Likelihood Principle," *Proc. of 2nd Intl. Symp. on Information Theory*, 1973, pp. 267–281.

[11] Wax, M., and T. Kailath, "Detection of Signals by Information Theoretic Criterion," *IEEE Trans. on Acoust., Speech, Signal Processing*, Vol. ASSP-33, No. 2, April 1985, pp. 387–392.

5

DOA Estimations with ESPRIT Algorithms

5.1 Introduction

Most of the algorithms discussed up to now depend on the precise knowledge of the array steering matrix $\mathbf{A}(\theta)$. For every θ_i, the corresponding array response, $\mathbf{a}(\theta_i)$, must be known. This is obtained by either direct calibration in the field, or by analytical means using information about the position and the response of each individual element of the array; this is normally an expensive and time-consuming task. Furthermore, errors in the calibration may seriously degrade the estimation accuracy. Also, the spectral-based DOA algorithms involve an exhaustive search through all possible angles or steering vectors to find the locations of the power spectral peaks and to estimate the DOA, which is computationally intensive.

ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques) overcomes these problems with a dramatic reduction in computational and storage requirements by exploiting a property called the shift invariance of the array [1]. Unlike most DOA estimation methods such as MUSIC, ESPRIT does not require that the array manifold steering vectors be precisely known, so the array calibration requirements are not stringent. This chapter reviews the basics of algorithms belonging to the family of ESPRIT. Section 5.1 explains the fundamental principle of ESPRIT-based algorithms. Section 5.2 applies this principle to uniform linear arrays. Section 5.3 discusses the computation of signal

subspaces. Finally, in Sections 5.4 and 5.5, unitary versions of ESPRIT that employ real-valued computations are discussed for element space and beamspace.

As noted in Chapter 3, an excellent doctoral dissertation on array signal processing for DOA estimations by M. Haardt was published by Shaker Verlag [1]. It presents a comprehensive review and study on the ESPRIT technique. In order not to confuse a reader with different mathematical symbols and technical terms, this chapter uses the same technical language and covers similar topics while taking into account the fact that these symbols and terms have also been used in many other related literature.

5.2 Basic Principle

In this section, the basic principle behind the ESPRIT algorithm is explained. ESPRIT achieves a reduction in computational complexity by imposing a constraint on the structure of an array. The ESPRIT algorithm assumes that an antenna array is composed of two identical subarrays (see Figure 5.1). The subarrays may overlap, that is, an array element may be a member of both subarrays. If there are a total of

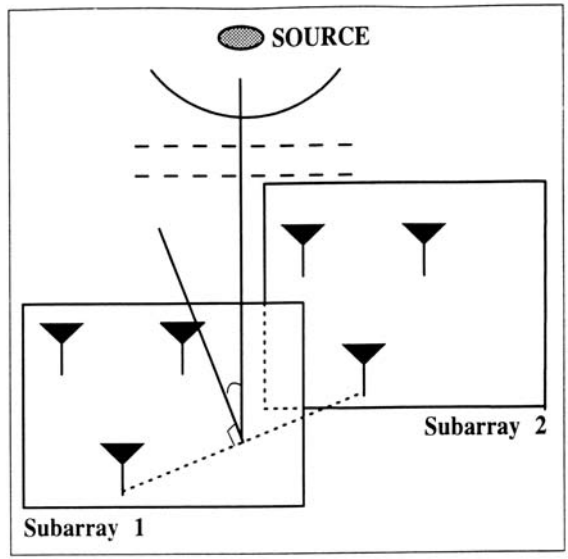


Figure 5.1 Antenna array structure for ESPRIT-based algorithms.

M elements in an array and m elements in each subarray, the overlap implies that $M \leq 2m$. For subarrays that do not overlap, $M = 2m$.

The individual elements of each subarray can have arbitrary polarization, directional gain, and phase response, provided that each has an identical twin in its companion subarray. Elements of each pair of identical sensors, or doublet, are assumed to be separated physically by a fixed displacement (translational) vector. The array thus possesses a displacement (translational) invariance (i.e., array elements occur in matched pairs with identical displacement vectors). This property leads to the rotational invariance of signal subspaces spanned by the data vectors associated with the spatially displaced subarrays; the invariance is then utilized by ESPRIT to find DOAs. We will detail these concepts in the following section.

5.2.1 Signal and Data Model

Assume d signals impinging onto the array. Let $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ represent the signal received by the two subarrays corrupted by additive noise $\mathbf{n}_1(t)$ and $\mathbf{n}_2(t)$. Each of the subarrays has m elements. From Section 2.2.1, the signals received can then be expressed as

$$\begin{aligned}\mathbf{x}_1(t) &= \left[\mathbf{a}(\mu_1), \ldots, \mathbf{a}(\mu_d)\right]s(t) + \mathbf{n}_1(t) \\ &= \mathbf{A}\mathbf{s}(t) + \mathbf{n}_1(t)\end{aligned} \tag{5.1a}$$

$$\begin{aligned}\mathbf{x}_2(t) &= \left[\mathbf{a}(\mu_1)e^{j\mu_1}, \ldots, \mathbf{a}(\mu_d)e^{j\mu_d}\right]\mathbf{s}(t) + \mathbf{n}_2(t) \\ &= \mathbf{A}\Phi\mathbf{s}(t) + \mathbf{n}_2(t)\end{aligned} \tag{5.1b}$$

where $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ are the $m \times 1$ vectors representing the data received by the first and second subarrays, respectively. $\mathbf{n}_1(t)$ and $\mathbf{n}_2(t)$ are the $m \times 1$ vectors representing the noises received by the two subarrays, respectively. $\mathbf{A} = [\mathbf{a}(\mu_1), \ldots, \mathbf{a}(\mu_d)]$ is the $m \times d$ steering matrix of the subarray. $\mathbf{s}(t)$ is the signals received by the first subarray. $\Phi = \text{diag}\left[e^{j\mu_1}, \ldots, e^{j\mu_d}\right]$ is a $d \times d$ diagonal matrix that relates the signals received by the two subarrays and is called the rotation operator. It is caused by the fact that the signals arriving at the second subarray will experience an extra delay

due to the fixed displacement Δ between the two subarrays; here $\mu_i = -\frac{2\pi f_c}{c}\Delta \sin\theta_i = -\frac{2\pi}{\lambda}\Delta \sin\theta_i$ as explained in Section 3.2.1.

Equations (5.1a) and (5.1b) can be combined to form the total array output vector as

$$\mathbf{x}(t) = \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} = \begin{bmatrix} \mathbf{A} \\ \mathbf{A}\Phi \end{bmatrix} \mathbf{s}(t) + \begin{bmatrix} \mathbf{n}_1(t) \\ \mathbf{n}_2(t) \end{bmatrix} = \tilde{\mathbf{A}}\mathbf{s}(t) + \mathbf{n}(t) \tag{5.2}$$

Given N snapshots, $\mathbf{x}(t_1)$, $\mathbf{x}(t_2)$, …, $\mathbf{x}(t_n)$, the objective of the ESPRIT technique is to estimate the DOAs via an estimation of μ_i by determining $\Phi = \text{diag}[e^{j\mu_1}, \ldots, e^{j\mu_d}]$. In doing so, two steps are required based on the data received by the array: estimating the signal subspace and then estimating the subspace rotation operator [1–3]. They are further elaborated in Section 5.2.2.

5.2.2 Signal Subspace Estimation

For a single signal, if $\mathbf{x}(t) = \mathbf{a}(\theta)\mathbf{s}(t)$ (i.e., in the ideal noise-free situation), the data is confined to a one-dimensional subspace of $\mathbf{C}^M$ characterized by the steering vector $\mathbf{a}(\theta)$. For the d signals, the observed data vectors $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$ are constrained to d-dimensional signal subspace of $\mathbf{C}^M$ called the signal subspace (as described in Section 2.6). The objective here is to estimate this d-dimensional signal subspace. Let $\mathbf{E}_1$ and $\mathbf{E}_2$ denote two sets of vectors that span the same signal subspace, which is also ideally spanned by the columns of $\mathbf{A}$. The signal subspace can be obtained from an array output covariance $\mathbf{R}_{xx}$ as explained in Section 3.3.3. The data covariance matrix $\mathbf{R}_{xx}$ has the following form:

$$\mathbf{R}_{xx} = E\left[\mathbf{x}(t)\mathbf{x}^H(t)\right] = \tilde{\mathbf{A}}R_{ss}\tilde{\mathbf{A}}^H \tag{5.3}$$

Both $\mathbf{R}_{ss}$ and the steering matrix $\tilde{\mathbf{A}}$ are assumed to have a full rank d. Suppose that the signal subspace is spanned as $\mathbf{E}_s = [\mathbf{e}_1, \ldots, \mathbf{e}_d]$. Since $\mathbf{R}_{ss}$ has a full rank, $\mathbf{E}_s$ spans the same space as $\tilde{\mathbf{A}}$. As a result, there must exist a unique nonsingular matrix $\mathbf{T}$ such that

$$\mathbf{E}_s = \tilde{\mathbf{A}}\mathbf{T}$$

$\mathbf{E}_s$ can be decomposed into $\mathbf{E}_1$ and $\mathbf{E}_2$ of the two subarrays such that

$$\mathbf{E}_s = \begin{bmatrix} \mathbf{E}_1 \\ \mathbf{E}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{AT} \\ \mathbf{A\Phi T} \end{bmatrix} \tag{5.4}$$

from which it can be observed that

$$\text{Range}\{\mathbf{E}_1\} = \text{Range}\{\mathbf{E}_2\} = \text{Range}\{\mathbf{A}\} \tag{5.5}$$

Equation (5.5) basically indicates that the two subarrays span the same signal subspace and have the same dimension; this is because they are identically configured. As a result, a nonsingular $d \times d$ matrix denoted as Ψ can be found such that

$$\mathbf{E}_1 \Psi = \mathbf{E}_2 \Rightarrow \mathbf{AT}\Psi = \mathbf{A\Phi T} \tag{5.6}$$

and

$$\Psi = \mathbf{T}^{-1}\mathbf{\Phi T} \tag{5.7}$$

As can be seen now, Ψ and Φ are related via an eigenvalue-preserving similarity transformation. The diagonal elements (or eigenvalues) of Φ are equal to eigenvalues of the matrix Ψ that maps (rotates) the m-dimensional signal subspace matrix $\mathbf{E}_1$ associated with the first subarray to the m-dimensional signal subspace matrix $\mathbf{E}_2$ associated with the second subarray. In other words, the shift-invariance property is now expressed in terms of signal eigenvectors that span the signal subspace. Therefore, in ESPRIT, instead of finding the spatial rotational Φ directly, we find the subspace rotating operator Ψ and then its eigenvalues that contain the DOA information.

5.2.3 Estimation of the Subspace Rotating Operator Ψ

In a practical situation, only a finite number of noisy data are received and available. $\mathbf{E}_s$ is then estimated from the data matrix $\mathbf{X}$ or covariance matrix $\mathbf{R}_{xx} = E[\mathbf{x}(t)\mathbf{x}^H(t)] = \tilde{\mathbf{A}}\mathbf{R}_{ss}\tilde{\mathbf{A}}^H + \sigma_N^2\mathbf{I}_{2m}$. Due to these noises, range{$\mathbf{E}_s$} ≠ range{$\tilde{\mathbf{A}}$} and range {$\mathbf{E}_1$} ≠ range {$\mathbf{E}_2$}. Thus, $\mathbf{E}_1\Psi = \mathbf{E}_2$ as in (5.6) cannot be solved exactly. Therefore, an approach needs to be developed to obtain

a suitable estimate for Ψ. Two approaches generally employed for problems of this nature are least squares (LS) [4] and total least squares (TLS) [5], giving rise to two versions of ESPRIT.

The two approaches can be explained by considering (5.6) as a simple matrix model $\mathbf{AX} = \mathbf{B}$. To estimate $\mathbf{X}$, the least squares technique assumes that the matrix $\mathbf{A}$ is known and any errors in the problem are attributed to noise in $\mathbf{B}$. The least squares solution is then given as

$$\hat{\mathbf{X}} = \left[\mathbf{A}\mathbf{A}^H\right]^{-1}\mathbf{A}^H\mathbf{B}$$

where $\hat{\mathbf{X}}$ is the estimate of $\mathbf{X}$.

However, $\mathbf{A}$ is also determined from the received data containing noises or disturbances, and hence it may have errors too. Consequently, the total least squares (TLS) technique is preferred and used to solve the invariance equation that takes into account the noises in both $\mathbf{A}$ and $\mathbf{B}$. The TLS criterion can be stated as finding residual matrices $\mathbf{R}_A$ and $\mathbf{R}_B$ of minimum Frobenius norm such that

$$\left[\mathbf{A} + \mathbf{R}_A\right]\hat{\mathbf{X}} = \mathbf{B} + \mathbf{R}_B$$

The actual computation procedure for TLS can be found in many open literatures including [5].

For comparison purposes, both ESPRIT approaches are considered and elaborated upon in the following sections.

5.3 Standard ESPRIT

With the basic concepts and properties of ESPRIT presented earlier, we now describe the standard ESPRIT algorithm. For clarity, we take the uniform linear array (ULAs) as an example to illustrate the standard ESPRIT procedure. Consider a ULA consisting of M elements. A few different configurations of the array suitable for ESPRIT are shown in Figure 5.2.

Let the two subarrays be chosen with maximum overlap, that is, each subarray contains $m = M - 1$ elements [see Figure 5.2(a)]. The overall array steering matrix $\mathbf{A}$ of a ULA has the following Vandermonde structure:

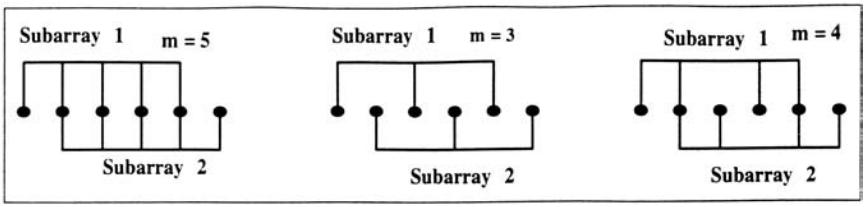


Figure 5.2 Centro-symmetric linear arrays and the corresponding subarrays for ESPRIT.

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{j\mu_1} & e^{j\mu_2} & \cdots & e^{j\mu_d} \\ \vdots & \vdots & \ddots & \vdots \\ e^{j(M-1)\mu_1} & e^{j(M-1)\mu_2} & \cdots & e^{j(M-1)\mu_d} \end{bmatrix} \tag{5.8}$$

and $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t)$. Each row in the steering matrix $\mathbf{A}$ corresponds to each element of the linear array. Choosing a particular subarray configuration can be made mathematically by applying a selection matrix to the overall steering matrix (5.7); it picks m rows of $\mathbf{A}$.

For example, consider a six-element ULA as shown in Figure 5.2. In the case of the maximum overlap, the first five elements form the first subarray and the last five elements form the second subarray. The selection matrices can then be formed by $\mathbf{J}_1$, which picks up the first $m = M - 1 = 5$ rows of $\mathbf{A}$ and $\mathbf{J}_2$, which selects the last $m = M - 1 = 5$ rows of the array steering matrix $\mathbf{A}$. The selection matrices are given by

$$\mathbf{J}_1 = [\mathbf{I}_m \; 0] \in R^{m \times M} \text{ and } \mathbf{J}_2 [0 \; \mathbf{I}_m] \in R^{m \times M} \tag{5.9}$$

where $\mathbf{I}_m$ is an $m \times m$ identity matrix. In general, the subarrays or equivalently the two selection matrices are chosen to be centro-symmetric with respect to one another. That is,

$$\mathbf{J}_2 = \Pi_m \mathbf{J}_1 \Pi_M \tag{5.10}$$

Here Π_p is a $p \times p$ exchange matrix ($p = m$ or M). Because of the fixed displacement between the first and second subarrays, the array steering vector of the second subarray $\mathbf{J}_2\mathbf{a}(\mu_i)$ is just a scaled version of the array steering vector of the first subarray $\mathbf{J}_1\mathbf{a}(\mu_i)$:

$$\mathbf{J}_1\mathbf{a}(\mu_i)e^{j\mu_i} = \mathbf{J}_2\mathbf{a}(\mu_i),\ 1 \le i \le d \tag{5.11}$$

This shift invariance property of all the d steering vectors $\mathbf{a}(\mu_i)$ can be expressed in a compact matrix form as:

$$\mathbf{J}_1\mathbf{A}\Phi = \mathbf{J}_2\mathbf{A} \tag{5.12}$$

where $\Phi = \text{diag}[e^{j\mu_1}, \ldots, e^{j\mu_d}]$ is a unitary diagonal $d \times d$ matrix whose diagonal elements contain the desired DOA information in their phases, μ_i. ESPRIT type algorithms are based on this shift invariance property of the array steering matrix $\mathbf{A}$. However, as discussed in previous sections, instead of finding Φ directly, ESPRIT algorithms estimate the subspace rotating operator Ψ, which is directly related to Φ. In doing so, an estimation of the signal subspace is needed as described next.

5.3.1 Signal Subspace Estimation

5.3.1.1 True Signal Subspace

Due to the assumption that the d impinging signal wavefronts are not correlated (or have been "decorrelated" as described in Section 4.2), the signal covariance matrix $\mathbf{R}_{ss} = E\{\mathbf{s}(t)\mathbf{s}^H(t)\}$ has full rank, namely rank $\{\mathbf{R}_{ss}\} = d$. Thus the covariance matrix of the noise-corrupted signals or data [4] is

$$\mathbf{R}_{xx} = E\left\{\mathbf{x}(t)^H\mathbf{x}(t)\right\} = \mathbf{A}\mathbf{R}_{ss}\mathbf{A}^H + \sigma_N^2\mathbf{I}_M \quad \in C^{M\times M} \tag{5.13}$$

Without additive noise (i.e., $\sigma_N^2 = 0$, $\mathbf{R}_{xx}$), $\mathbf{R}_{xx}$ would be rank deficient since

$$\text{rank}\left\{\mathbf{A}\mathbf{R}_{ss}\mathbf{A}^H\right\} = d < M \tag{5.14}$$

In this case, its $M - d$ smallest eigenvalues are equal to zero, and an eigenvalue decomposition of $\mathbf{R}_{xx}$ can be expressed as

$$\begin{aligned}\mathbf{R}_{xx}\big|_{\sigma_N^2=0} &= \sum_{k=1}^{M}\lambda_k\mathbf{u}_k\mathbf{u}_k^H \\ &= \begin{bmatrix}\mathbf{U}_s & \mathbf{U}_o\end{bmatrix}\begin{bmatrix}\Lambda_d & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}\mathbf{U}_s^H\\ \mathbf{U}_o^H\end{bmatrix} = \mathbf{U}_s\Lambda_d\mathbf{U}_s^H\end{aligned} \tag{5.15}$$

where the diagonal matrix $\Lambda_d = \text{diag}[\lambda_1, \lambda_2, \ldots, \lambda_d]$ contains the nonzero eigenvalues. Here, the eigenvalues λ_k are ordered according to their magnitudes such that

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d \geq \lambda_{d+1} = \lambda_{d+2} = \ldots = \lambda_M = 0 \tag{5.16}$$

The columns of

$$\mathbf{U}_s = [\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_d] \text{ and } \mathbf{U}_o = [\mathbf{u}_{d+1}\ \mathbf{u}_{d+2} \ldots \mathbf{u}_M]$$

span the *d*-dimensional signal subspace *S* and its orthogonal complement is then called noise subspace, respectively, as described in Section 3.6. With additive noise, the eigendecomposition of the covariance matrix $\mathbf{R}_{xx}$ in (5.13) gives

$$\begin{aligned} \mathbf{R}_{xx} &= \sum_{k=1}^{M} \rho_k \mathbf{u}_k \mathbf{u}_k^H \\ &= [\mathbf{U}_s\ \mathbf{U}_o] \left(\begin{bmatrix} \Lambda_d & 0 \\ 0 & 0 \end{bmatrix} + \sigma_N^2 \mathbf{I}_M \right) \begin{bmatrix} \mathbf{U}_s^H \\ \mathbf{U}_o^H \end{bmatrix} \end{aligned} \tag{5.17}$$

Obviously, from (5.15) and (5.17), the eigenvalues of $\mathbf{R}_{xx}$ now become $\rho_k = \lambda_k + \sigma_N^2$, $1 \leq k \leq d$; they are shifted by σ_N^2, but their corresponding eigenvectors remain the same as those of (5.15). From (5.13) and (5.15), we can see that

$$\mathbf{R}_{xx}\big|_{\sigma_N^2 = 0} = \mathbf{A}\mathbf{R}_{ss}\mathbf{A}^H = \mathbf{U}_s \Lambda_d\ \mathbf{U}_s^H$$

From the above relation, it can be observed that the columns of the array steering matrix **A** also span the *d*-dimensional signal subspace, that is,

$$S = \text{Range}\{\mathbf{A}\} = \text{Range}\{\mathbf{U}_s\} \tag{5.18}$$

Therefore, there exists a nonsingular $d \times d$ matrix $\mathbf{T}_A$ such that $\mathbf{A} = \mathbf{U}_s\mathbf{T}_A$. Hence, the shift-invariance property of (5.12) can be expressed in terms of the signal eigenvectors $\mathbf{U}_s$ that span the signal subspace:

$$\mathbf{J}_1\mathbf{U}_S\mathbf{T}_A\Phi = \mathbf{J}_2\mathbf{U}_S\mathbf{T}_A \Leftrightarrow \mathbf{J}_1\mathbf{U}_S\Psi = \mathbf{J}_2\mathbf{U}_S \tag{5.19}$$

where $\Psi = \mathbf{T}_A \Phi \mathbf{T}_A^{-1}$ is a nonsingular $d \times d$ matrix signal subspace rotating operator. As a result, if the signal subspace, as represented by $\mathbf{U}_s$, can be estimated, Ψ and its eigenvalues containing DOAs information can then be found.

5.3.1.2 Estimation of True Signal Subspace

As mentioned before, the ESPRIT algorithm is based on an estimation of signal subspace from the data covariance matrix. Let $\mathbf{X}$ denote the matrix composed of noisy data of N snapshots $\mathbf{x}(t_n)$, $1 \le n \le N$.

$$\begin{aligned}\mathbf{X} &= \left[\mathbf{x}(t_1)\ \mathbf{x}(t_2)\ \ldots\ \mathbf{x}(t_N)\right] \\ &= \mathbf{A}\left[\mathbf{s}(t_1)\ \mathbf{s}(t_1)\ \ldots\ \mathbf{s}(t_N)\right] + \left[\mathbf{n}(t_1)\ \mathbf{n}(t_1)\ \ldots\ \mathbf{n}(t_N)\right] \\ &= \mathbf{A} \cdot \mathbf{S} + \mathbf{N}\end{aligned} \tag{5.20a}$$

Then an estimate of the data covariance matrix is given by

$$\hat{\mathbf{R}}_{xx} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}(t_n)\mathbf{x}^H(t_n) = \frac{1}{N}\mathbf{X}\mathbf{X}^H \tag{5.20b}$$

The signal subspace can be estimated from this data covariance matrix. From (5.20b), as observed earlier, the eigenvectors that correspond to the d largest eigenvalues of $\hat{\mathbf{R}}_{xx}$ span the estimated signal subspace.

As described in Section 2.6.1, two approaches can be used to estimate the signal subspace: the covariance approach by taking eigenvalue decomposition (EVD) of $\mathbf{X}\mathbf{X}^H$ or the direct data approach by taking the singular value decomposition (SVD) of $\mathbf{X}$. The later approach has an advantage over the former in that it does not involve the "squaring" of the data numbers. In addition, when implementing the algorithm on any signal processor in real time, we have to work with finite precision numbers; the covariance approach gives "big" numbers, which may cause numerical problems like round-off error and overflow. As a result, the direct data approach is preferred; the SVD of the data matrix $\mathbf{X}$ is then given by

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^H = \begin{bmatrix}\mathbf{U}_s & \mathbf{U}_o\end{bmatrix}\begin{bmatrix}\Sigma_s & 0 \\ 0 & \Sigma_o\end{bmatrix}\begin{bmatrix}\mathbf{V}_s^H \\ \mathbf{V}_o^H\end{bmatrix} \tag{5.21}$$

where the diagonal matrix $\Sigma \in R^{d\times d}$ contains the d largest singular values and $\mathbf{U}_s$ spans the signal subspace.

5.3.2 Solution of Invariance Equation

After the estimation of the matrix $\mathbf{U}_s$ that spans the estimated signal subspace, based on (5.19), the two known selection matrices $\mathbf{J}_1$ and $\mathbf{J}_2$ are applied to form the following invariance equation.

$$\mathbf{J}_1\mathbf{U}_s\Psi \approx \mathbf{J}_2\mathbf{U}_s \in C^{m\times d} \tag{5.22}$$

where Ψ is the signal subspace rotating operator, defined by (5.17) or (5.6). In contrast to (5.19), this invariance equation might not have an exact solution; this is because the signal subspace is estimated from an estimated data covariance matrix (5.21) that is not the true or exact signal covariance matrix. Also, the size of the subarrays $M-1$ should be at least equal to d in order to compute all the DOAs; otherwise, the invariance system (5.22) would be underdetermined.

Equation (5.22) is solved by using the least squares (LS) or total least squares (TLS) solutions to get an estimate of the subspace rotating operator Ψ as explained in the previous section. In real time, while implementing these solutions, the more efficient algorithm called the QR decomposition may also be used to solve the least squares. This is because the direct matrix inverse calculations are prone to error and consume lot of time in computations [5].

5.3.3 Spatial Frequency and DOA Estimation

Once $\Psi \in C^{d\times d}$ is found, the desired DOA information can be estimated from it. The eigenvalues of the estimated $\Psi \in C^{d\times d}$ can be calculated by its eigendecomposition; this is because

$$\Psi = \mathbf{T}\Phi\mathbf{T}^{-1} \text{ where } \Phi = diag\left[e^{j\mu_1}, \ldots, e^{j\mu_d}\right] \tag{5.23}$$

Therefore, the eigenvalue of Ψ, ϕ_i, represents estimates of the phase factors $e^{j\mu_i}$. Once the estimates of the spatial frequencies μ_i are found, the corresponding DOAs θ_i are obtained via the relationships

$$\mu_i = \arg(\phi_i) \text{ and } \theta_i = \arcsin\left(-\frac{\lambda}{2\pi\Delta}\mu_i\right),\ 1 \le i \le d \qquad (5.24)$$

A brief summary of standard ESPRIT computation steps is shown in Table 5.1 [1].

As seen from Table 5.1, the standard ESPRIT computations involve complex number operations. To avoid this, an improved ESPRIT employing only real-valued computations has been developed. It is described in Section 5.4.

5.4 Real-Valued Transformation

The standard ESPRIT is comprised of complex-valued computations throughout the algorithm. This makes the algorithm computationally expensive. However, it is observed that if centro-symmetric arrays are used, the signal subspace can be estimated by real-valued computations. Exploiting this property for centro-symmetric arrays leads to the unitary ESPRIT with only real-valued computations. Centro-symmetric arrays introduced in Chapter 2 thus are essential for implementing unitary ESPRIT. Unitary ESPRIT is formulated and based on the following theorem [5].

Table 5.1
Summary of Standard ESPRIT Algorithm

1. *Signal Subspace Estimation:* Compute $\mathbf{U}_s$

   as the d dominant left singular vectors of $\mathbf{X}$ (square-root approach) or the d dominant eigenvectors of $\mathbf{XX}^H$ (covariance approach)

2. *Solution of the Invariance Equation:* Solve the following equation for Ψ

   $$\mathbf{J}_1\mathbf{U}_s\Psi \approx \mathbf{J}_2\mathbf{U}_s$$

   by means of *least-squares* or *total least-squares* techniques

3. *DOA Estimation:* Calculate the eigenvalues of the resulting complex-valued solution

   $$\Psi = \mathbf{T}\Phi\mathbf{T}^{-1} \text{ with } \Phi = \text{diag}[\phi_1,\ldots,\phi_d]$$

   and then extract the angular information via

   $$\mu_i = \arg(\phi_i),\quad 1 \le i \le d$$
   $$\theta_i = \arcsin\left(-\frac{\lambda}{2\pi\Delta}\mu_i\right)$$

Theorem 5.1

Let $\mathbf{Q}_p$ and $\mathbf{Q}_q$ denote the left Π-real matrices of $p \times p$ and $q \times q$, respectively, defined in (3.16). Then given any $p \times q$ centro-Hermitian matrix $\mathbf{G}$, $\mathbf{Q}_p^{-1}\mathbf{G}\mathbf{Q}_q$ is a real-valued $p \times q$ matrix.

Consider the extended data matrix defined in (4.8). It is written here again for ready reference.

$$\mathbf{Z} = \left[\mathbf{X} \;\; \Pi_M \overline{\mathbf{X}} \Pi_N\right] \tag{5.25}$$

It can be shown that $\mathbf{Z}$ is centro-Hermitian [6]. Also, similar to what was described in Section 4.2, the estimate $\mathbf{R}_{xx}^{fb}$ of the forward-backward averaged covariance matrix is centro-Hermitian [6].

$$\hat{\mathbf{R}}_{xx}^{fb} = \frac{1}{2}\left(\hat{\mathbf{R}}_{xx} + \Pi_M \overline{\hat{\mathbf{R}}}_{xx} \Pi_M\right) \tag{5.26}$$

Now, the unitary transformation of the complex-valued data matrix to real-valued matrix can be obtained by applying Theorem 5.1 to the extended data matrix.

$$\Gamma(\mathbf{X}) = \varphi(\mathbf{Z}) = \mathbf{Q}_M^H \left[\mathbf{X} \;\; \Pi_M \overline{\mathbf{X}} \Pi_n\right] \mathbf{Q}_{2N} \in R^{M \times 2N} \tag{5.27}$$

The transformation $\Gamma(\mathbf{X})$ thus accomplishes forward-backward averaging by first extending the complex-valued data matrix $\mathbf{X}$ of size $M \times N$ to a complex-valued centro-Hermitian matrix $\mathbf{Z}$ of size $M \times 2N$ and then transforming $\mathbf{Z}$ into a real-valued matrix of the same size. As a result, we can obtain

$$\begin{aligned} \varphi\left(\hat{\mathbf{R}}_{xx}^{fb}\right) &= \mathbf{Q}_M^H \hat{\mathbf{R}}_{xx}^{fb} \mathbf{Q}_M = \frac{1}{2}\left(\mathbf{Q}_M^H \hat{\mathbf{R}}_{xx} \mathbf{Q}_M + \mathbf{Q}_M^H \Pi_M \overline{\hat{\mathbf{R}}}_{xx} \Pi_{\mathbf{M}} \mathbf{Q}_M\right) \\ &= \frac{1}{2}\left(\mathbf{Q}_M^H \hat{\mathbf{R}}_{xx} \mathbf{Q}_M + \overline{\mathbf{Q}}_M^H \overline{\hat{\mathbf{R}}}_{xx} \mathbf{Q}_M\right) \\ &= \mathrm{Re}\left\{\mathbf{Q}_M^H \hat{\mathbf{R}}_{xx} \mathbf{Q}_M\right\} \in R^{M \times M} \end{aligned} \tag{5.28}$$

As seen, the forward-backward averaging can be achieved automatically and implicitly by taking the real part of the transformed covariance

matrix $\mathbf{Q}_M^H \hat{\mathbf{R}}_{xx} \mathbf{Q}_M$. Hence, the real-valued signal subspace $\mathbf{E}_s$ can be found as the d dominant left singular vectors of $\Gamma(\mathbf{X}) \in R^{M\times 2N}$ (direct data approach) or $\Gamma(\mathbf{X})\Gamma(\mathbf{X})^H \in R^{M\times M}$ (covariance approach). This principle is applied to formulate a new version of ESPRIT called the unitary ESPRIT, which is described next.

5.5 Unitary ESPRIT in Element Space

Unitary ESPRIT is applicable to centro-symmetric array configurations and it includes forward-backward averaging [3]. The complex-valued factorizations like SVD or EVD are transformed into real-valued factorizations. The element space implementation of unitary ESPRIT is discussed in detail in this section. Another approach, the unitary ESPRIT in DFT beamspace, which focuses on a particular DOA sector of interest with the reduced computational complexity, will be described later.

5.5.1 One-Dimensional Unitary ESPRIT in Element Space

Unitary ESPRIT performs the computations in real instead of complex numbers from the beginning to the end of the algorithm. The concepts explained with centro-symmetric arrays in Section 3.3 and the fundamental theorem reviewed in the previous section are employed to achieve this real-valued estimation. In this section, each step of the algorithm, that is, estimation of the signal subspace estimation, the solution of the invariance equation, and the computation of DOAs, will be presented with real-valued calculations.

5.5.1.1 Real-Valued Subspace Estimation

Consider a uniform linear array (ULA) with maximum overlap. Assume the signal and data model discussed in Chapter 2. Suppose that the real-valued data matrix from the transformation $\Gamma(\mathbf{X})$ and its real-valued covariance matrix $\mathbf{R}_{xx}$ are obtained using unitary transformations as discussed in the previous section. Also, this unitary transformation includes forward-backward averaging. Let the d dominant left singular vectors of $\Gamma(\mathbf{X}) \in R^{M\times 2N}$ (direct data approach) or the d dominant eigenvectors of $\Gamma(\mathbf{X})\Gamma(\mathbf{X})^H \in R^{M\times M}$ (covariance approach) be $\mathbf{E}_s \in R^{M\times d}$. Then,

$$\mathbf{U}_s = \mathbf{Q}_M \mathbf{E}_s \quad \text{or} \quad \mathbf{E}_s = \mathbf{Q}_M^E \mathbf{U}_s \tag{5.29}$$

contains a basis for the estimated signal subspace. More specifically, the columns of $\mathbf{U}_s$ in (5.29) contain d dominant left singular vectors of the extended data matrix $\mathbf{Z}$ or d dominant eigenvectors of $\mathbf{ZZ}^H$. From now, we refer to the subspace spanned by $\mathbf{E}_s$ as the real-valued signal subspace. The real-valued and complex-valued subspaces are related by (5.29). In the next step, the real-valued invariance equation is formulated.

5.5.1.2 Real-Valued Invariance Equation

In this section, the approach to transforming the complex-valued invariance equation of a uniform linear array with a maximum overlap into a real-valued equation is presented.

Consider the complex-valued invariance equation (5.11) rewritten here for ready reference.

$$\mathbf{J}_1 \mathbf{a}(\mu_i) e^{j\mu_1} = \mathbf{J}_2 \mathbf{a}(\mu_i), \; 1 \le i \le d \tag{5.30}$$

Denote $\mathbf{b}(\mu_i)$ as the real-valued transformed steering vector; then by Theorem 5.1, it can be obtained as:

$$\mathbf{b}(\mu_i) = \mathbf{Q}_M^H \mathbf{a}(\mu_i) \tag{5.31}$$

where $\mathbf{Q}_m$ and $\mathbf{Q}_M$ are unitary and left Π-real matrices. Since $\mathbf{Q}_M$ is unitary, it follows that

$$\mathbf{Q}_M \mathbf{Q}_M^H = \mathbf{I}_M \tag{5.32}$$

By substituting (5.32) into (5.30), we get

$$\mathbf{J}_1 \mathbf{Q}_M \mathbf{Q}_M^H \mathbf{a}(\mu_i) e^{j\mu_i} = \mathbf{J}_2 \mathbf{Q}_M \mathbf{Q}_M^H \mathbf{a}(\mu_i), \; 1 \le i \le d \tag{5.33}$$

By substituting (5.31) into (5.33), we get

$$\mathbf{J}_1 \mathbf{Q}_M \mathbf{b}(\mu_i) e^{j\mu_i} = \mathbf{J}_2 \mathbf{Q}_M \mathbf{b}(\mu_i) \tag{5.34}$$

Premultiplying both sides by $\mathbf{Q}_m^H$ gives the following invariance relationship

$$\mathbf{Q}_m^H \mathbf{J}_1 \mathbf{Q}_M \mathbf{b}(\mu_i) e^{j\mu_i} = \mathbf{Q}_m^H \mathbf{J}_2 \mathbf{Q}_M \mathbf{b}(\mu_i), \quad 1 \le i \le d \tag{5.35}$$

The selection matrices $\mathbf{J}_1$ and $\mathbf{J}_2$ are real-valued and satisfy

$$\mathbf{J}_1 = \Pi_m \mathbf{J}_2 \Pi_M \tag{5.36}$$

Also, $\mathbf{Q}_M$ satisfies the relationship

$$\Pi_M = \overline{\mathbf{Q}}_M \mathbf{Q}_M^H \tag{5.37}$$

and Π_M satisfies the relation

$$\Pi_M^H = \Pi_M \tag{5.38}$$

Combining (5.36), (5.37), and (5.38), we get

$$\begin{aligned} \mathbf{Q}_m^H \mathbf{J}_2 \mathbf{Q}_M &= \mathbf{Q}_m^H \Pi_m \Pi_m \mathbf{J}_2 \Pi_M \Pi_M \mathbf{Q}_M \\ &= \overline{\mathbf{Q}_m^H \mathbf{J}_1 \mathbf{Q}_M} \end{aligned} \tag{5.39}$$

Let $\mathbf{K}_1$ and $\mathbf{K}_2$ be the real and imaginary parts of $\mathbf{Q}_M^H \mathbf{J}_2 \mathbf{Q}_M$. Then (5.35) can be written as

$$(\mathbf{K}_1 - j\mathbf{K}_2)\mathbf{b}(\mu_i) e^{j\mu_i} = (\mathbf{K}_1 + j\mathbf{K}_2)\mathbf{b}(\mu_i), \quad 1 \le i \le d \tag{5.40}$$

Rearranging the terms we have

$$\mathbf{K}_1 \mathbf{b}(\mu_i) \cdot \left(e^{j\mu_i} - 1\right) = \mathbf{K}_2 \mathbf{b}(\mu_i) \cdot j\left(e^{j\mu_i} + 1\right) \tag{5.41}$$

By definition of tangent function, the above invariance relationship satisfied by $\mathbf{b}(\mu_i)$ can be expressed finally in real-valued quantities [6]:

$$\mathbf{K}_1 \mathbf{b}(\mu_i) \cdot \tan\left(\frac{\mu_i}{2}\right) = \mathbf{K}_2 \mathbf{b}(\mu_i), \quad 1 \le i \le d \tag{5.42}$$

For d impinging signals, we define the real-valued transformed steering matrix of size $M \times d$ as

$$\mathbf{B} = \mathbf{Q}_M^H \mathbf{A} = \left[d(\mu_1) d(\mu_2) \ldots d(\mu_d) \right] \tag{5.43}$$

The shift invariance relation in (5.42) can then be written in a matrix form as

$$\mathbf{K}_1 \mathbf{B} \Omega = \mathbf{K}_2 \mathbf{B} \tag{5.44}$$

where $\Omega = \text{diag}\left\{ \tan\left(\frac{\mu_1}{2}\right), \tan\left(\frac{\mu_2}{2}\right), \ldots, \tan\left(\frac{\mu_i}{2}\right), \ldots, \tan\left(\frac{\mu_d}{2}\right) \right\}$.

Let the columns of $\mathbf{U}_s \in C^{M \times d}$ and $\mathbf{E}_s \in C^{M \times d}$ span the estimated complex- valued signal subspace and real-valued signal subspace, respectively. In the noise-free case, $\mathbf{E}_s = \mathbf{Q}_M^H \mathbf{U}_s$ and $\mathbf{B} = \mathbf{Q}_M^H \mathbf{A}$ span the same d-dimensional signal subspace, that is, there is a nonsingular $d \times d$ matrix $\mathbf{T}_A$ such that $\mathbf{B} = \mathbf{E}_s \mathbf{T}_A$. Substitution of this observation into (5.44) gives the real-valued invariance equation

$$\mathbf{K}_1 \mathbf{E}_s \Upsilon \approx \mathbf{K}_2 \mathbf{E}_s \tag{5.45}$$

where $\Upsilon = \mathbf{T}_A \Omega \mathbf{T}_A^{-1}$ contains the desired DOA information and the equality is replaced by an approximation to account for noisy data.

The above observation can be summarized as follows: the complex-valued invariance equation $\mathbf{J}_1 \mathbf{U}_s \Psi = \mathbf{J}_2 \mathbf{U}_s$, given by (5.18), can be replaced by the real-valued invariance equation (5.45) which is of the size $m \times d$, with $\mathbf{K}_1$ and $\mathbf{K}_2$ being transformed selection matrices defined as

$$\begin{aligned} \mathbf{K}_1 &= \mathbf{Q}_m^H (\mathbf{J}_1 + \mathbf{J}_2) \mathbf{Q}_M = 2 \cdot \text{Re}\left\{ \mathbf{Q}_m^H \mathbf{J}_2 \mathbf{Q}_M \right\} \\ \mathbf{K}_2 &= \mathbf{Q}_m^H j(\mathbf{J}_1 - \mathbf{J}_2) \mathbf{Q}_M = 2 \cdot \text{Im}\left\{ \mathbf{Q}_m^H \mathbf{J}_2 \mathbf{Q}_M \right\} \end{aligned} \tag{5.46}$$

5.5.1.3 DOA Estimation

By proceeding as before for the standard ESPRIT, the spatial frequency estimates μ_i, $1 \le i \le d$, can directly be obtained from the eigenvalues of the real-valued matrix $\Upsilon \in R^{d \times d}$ by the means of the LS solution or the TLS solution of the real-valued invariance equation (5.45). Let

$$\Upsilon = \mathbf{T} \Omega \mathbf{T}^{-1} \tag{5.47}$$

with $\Omega = \text{diag}\{\omega_1, \omega_2, \ldots, \omega_i, \ldots, \omega_d\}$ being the eigenvalues of Υ. The eigenvalues ω_i represent estimates of $\tan\left(\frac{\mu_1}{2}\right)$. The spatial frequency is then solved as

$$\mu_i = 2\arctan(\omega_i),\ 1 \le i \le d \tag{5.48}$$

A summary of computational steps of one-dimensional unitary ESPRIT is tabulated in Table 5.2 [1].

In the next section, unitary ESPRIT is extended to two-dimensional arrays for estimating the DOA in both azimuthal and elevational angles.

5.5.2 Two-Dimensional Unitary ESPRIT in Element Space

A fundamental restriction of one dimensional array is that they can only estimate a single direction parameter. For a joint azimuthal and elevation angle estimation, a two-dimensional array is required. Consider a uniform rectangular array (URA) of size $M = M_x \times M_y$ in the *x*-*y* plane where the interelement spacing in the *x* and *y* directions are Δ_x and Δ_y, respectively. In addition, assume that the centro-symmetric arrays also exhibits a dual invariance, that is, two identical subarrays of m_x elements are displaced by Δ_x along the *x*-axis and another pair of identical subarrays, each

Table 5.2
Summary of One-Dimensional Unitary ESPRIT Algorithm

1. *Real-Valued Subspace Estimation:* Compute $\mathbf{E}_S$

   as the *d* dominant left singular vectors of $\Gamma_S(\mathbf{X})$ (square-root approach)
   or as the *d* dominant eigenvectors of $\Gamma_S(\mathbf{X})\Gamma_S(\mathbf{X})^H$ (covariance approach)

2. *Real-Valued Invariance Equation:* Solve the following equation for Υ

   $$\mathbf{K}_1\mathbf{E}_S\Upsilon \approx \mathbf{K}_2\mathbf{E}_S$$

   by means of LS or TLS

3. *DOA Estimation:* Calculate the eigenvalues of the resulting complex-valued solution

   $$\Upsilon = \mathbf{T}\Omega\mathbf{T}^{-1} \text{ with } \Omega = \text{diag}\{\omega_1, \omega_2, \ldots \omega_i, \ldots, \omega_d\}$$

   and then extract the angular information via

   $$\mu_i = 2\arctan(\omega_i),\quad 1 \le i \le d$$
   $$\theta_i = \arcsin\left[-\frac{\lambda}{2\pi\Delta}\right]$$

consisting of m_y elements, is displaced by Δ_y along the y-axis. Consider that the subarrays are maximally overlapped here. An example of an URA is shown in Figure 5.3.

Assume d signals impinging on the array. Let the array outputs at time t be stacked in a column vector $\mathbf{x}(t)$ as explained in Section 2.3.2. As in the one-dimensional case, the d impinging signals are combined to form a column vector $\mathbf{s}(t)$. Also, assume that the additive noise vector $\mathbf{n}(t)$ is taken from a zero mean and spatially uncorrelated random process with spatial covariance matrix $\sigma_N^2 \mathbf{I}_M$. As discussed in Section 3.3.2, the signals received by the array can be expressed as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \tag{5.49}$$

where the array steering matrix is formulated in such a way that

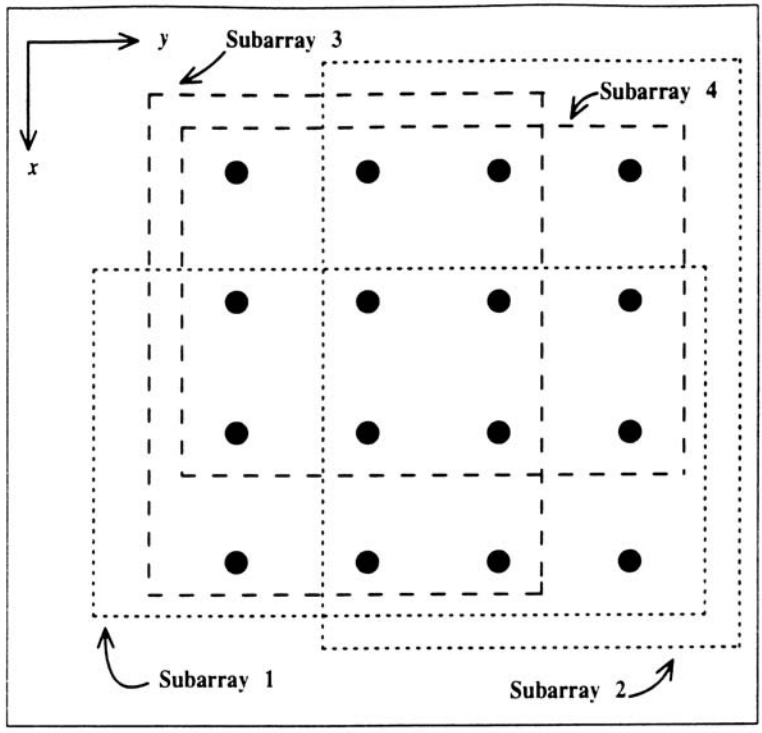


Figure 5.3 Subarray selection for URA of $M = 4 \times 4 = 16$ array elements (maximum overlap in both directions).

$$\mathbf{A} = \left[\mathbf{a}(\mu_1, v_1)\ \mathbf{a}(\mu_2, v_2) \ldots\ \mathbf{a}(\mu_d, v_d) \right] \tag{5.50}$$

which satisfies the centro-symmetric condition. The approach is to decompose the two-dimensional URA problem into two independent one-dimensional problems [1, 8]. Then, we can estimate the spatial frequencies with respect to each array separately. The spatial frequencies μ_i in the x direction and ν_i in the y direction are scaled versions of the corresponding direction cosines, namely,

$$\mu_i = \frac{2\pi}{\lambda} \Delta_x u_i \text{ and } \nu_i = \frac{2\pi}{\lambda} \Delta_y v_i, \quad 1 \le i \le d$$

As described in Section 2.3.2, the DOAs can be found through the following relationship:

$$\phi_i = \arg(u_i + jv_i) \text{ and } \theta_i = \arcsin\left(\left\|u_i + jv_i\right\|\right)$$

Like the one-dimensional unitary ESPRIT, to find these values, the signal subspace and the invariance equation need to be estimated.

5.5.2.1 Real-Valued Signal Subspace Estimation

As the first step, the real-valued signal subspace is obtained from the data matrix. This is obtained identically as the case for the one-dimensional ULA except that the data matrix employed here consists of stacked data from a uniform rectangular array. The real-valued data matrix from the transformation $\Gamma(\mathbf{X})$ and the real-valued covariance matrix $\mathbf{R}_{xx}$ are obtained using the unitary transformations as discussed in Section 5.5.1. It was observed that this unitary transformation also includes forward-backward averaging. Let the d dominant left singular vectors of $\Gamma(\mathbf{X}) \in R^{M \times 2N}$ (direct data approach) or the d dominant eigenvectors of $\Gamma(\mathbf{X})\Gamma(\mathbf{X})^H \in R^{M \times M}$ (covariance approach) be $\mathbf{E}_s \in R^{M \times d}$. Then,

$$\mathbf{U}_s = \mathbf{Q}_M \mathbf{E}_s \tag{5.51}$$

contains a basis for the estimated signal subspace. The columns of $\mathbf{U}_s$ in (5.51) contain d dominant left singular vectors of the extended data matrix $\mathbf{Z}$ or d dominant eigenvectors of $\mathbf{Z}\mathbf{Z}^H$. The real-valued and complex-valued subspaces are related through (5.51). As the next step two

invariance equations are formulated, which are later transformed into real-valued.

5.5.2.2 Real-Valued Invariance Equation

As the URA is subdivided into two pairs of subarrays with maximum overlap, two pairs of selection matrices are defined similar to (5.9): one-dimensional selection matrices for the uniform linear arrays in the x direction are

$$\mathbf{J}_1^{(M_x)} = \left[I_{M_{x-1}} \; 0 \right] \text{ and } \mathbf{J}_2^{(M_x)} = \left[0 \; I_{M_{x-1}} \right] \tag{5.52}$$

Similarly, the one-dimensional selection matrices for the ULAs in the y direction are given by

$$\mathbf{J}_1^{(M_y)} = \left[I_{M_{y-1}} \; 0 \right] \text{ and } \mathbf{J}_2^{(M_y)} = \left[0 \; I_{M_{y-1}} \right] \tag{5.53}$$

The array manifold matrices satisfy the invariance relation and can be expressed

$$\begin{aligned} \mathbf{J}_1^{(M_x)} \mathbf{A}(\mu_i, v_i) e^{j\mu_i} &= \mathbf{J}_2^{(M_x)} \mathbf{A}(\mu_i, v_i) \\ \mathbf{A}(\mu_i, v_i) \mathbf{J}_1^{(M_y)} e^{jv_i} &= \mathbf{A}(\mu_i, v_i) \mathbf{J}_2^{(M_y)} \end{aligned} \tag{5.54}$$

Now applying the vec{•} operator to (5.54) and using the property in the Appendix, Section A.1,

$$\begin{aligned} \mathbf{J}_{\mu 1} \mathbf{a}(\mu_i, v_i) e^{j\mu_i} &= \mathbf{J}_{\mu 2} \mathbf{a}(\mu_i, v_i) \\ \mathbf{J}_{v1} \mathbf{a}(\mu_i, v_i) e^{jv_i} &= \mathbf{J}_{v2} \mathbf{a}(\mu_i, v_i) \end{aligned} \tag{5.55}$$

The two-dimensional selection matrices of the URA (corresponding to maximum overlap) can be obtained as the following Kronecker products [1, 7]:

$$\mathbf{J}_{\mu 1} = \mathbf{I}_{M_y} \otimes \mathbf{J}_1^{(M_x)} \text{ and } \mathbf{J}_{\mu 2} = \mathbf{I}_{M_y} \otimes \mathbf{J}_2^{(M_x)} \tag{5.56a}$$

$$\mathbf{J}_{v1} = \mathbf{J}_1^{(M_y)} \otimes \mathbf{I}_{M_x} \text{ and } \mathbf{J}_{v2} = \mathbf{J}_2^{(M_y)} \otimes \mathbf{I}_{M_x} \tag{5.56b}$$

These two pairs of selection matrices are centro-symmetric with respect to each other, that is,

$$\mathbf{J}_{\mu 1} = \Pi_{m_x} \mathbf{J}_{\mu 2} \Pi_M \text{ and } \mathbf{J}_{v1} = \Pi_{m_y} \mathbf{J}_{v2} \Pi_M \tag{5.57}$$

The array steering matrix $\mathbf{A}$ then satisfies the following two invariance equations for uniform rectangular arrays that are complex in nature.

$$\begin{aligned} \mathbf{J}_{\mu 1}\mathbf{A}\Phi_\mu &= \mathbf{J}_{\mu 2}\mathbf{A} \\ \mathbf{J}_{v1}\mathbf{A}\Phi &= \mathbf{J}_{v2}\mathbf{A} \end{aligned} \tag{5.58}$$

where the complex-valued diagonal matrices

$$\begin{aligned} \phi_\mu &= \text{diag}\left\{e^{j\mu_1}, e^{j\mu_2}, \ldots, e^{j\mu_i}, e^{j\mu_d}\right\} \text{ and} \\ \phi_v &= \text{diag}\left\{e^{jv_1}, e^{jv_2}, \ldots, e^{jv_i}, e^{jv_d}\right\} \end{aligned} \tag{5.59}$$

contain the desired two-dimensional DOA angle information to be estimated.

Next, these complex-valued invariance equations for URA are transformed into real-valued invariance equations similar to the one-dimensional case in Section 5.5.1. Define the transformed two-dimensional array steering matrix as $\mathbf{B} = \mathbf{Q}_M^H \mathbf{A}$. Based on the two invariance properties of the two-dimensional array steering matrix $\mathbf{A}$ in (5.58), the transformed array steering matrix $\mathbf{B}$ satisfies

$$\begin{aligned} \mathbf{K}_{\mu 1}\mathbf{B} \cdot \Omega_\mu &= \mathbf{K}_{\mu 2}\mathbf{B} \\ \mathbf{K}_{v1}\mathbf{B} \cdot \Omega_v &= \mathbf{K}_{v2}\mathbf{B} \end{aligned} \tag{5.60}$$

The two pairs of transformed selection matrices are given by

$$\mathbf{K}_{\mu 1} = 2 \cdot \text{Re}\left\{\mathbf{Q}_{m_x}^H \mathbf{J}_{\mu 2} \mathbf{Q}_M\right\} \quad \mathbf{K}_{\mu 2} = 2 \cdot \text{Im}\left\{\mathbf{Q}_{m_x}^H \mathbf{J}_{\mu 2} \mathbf{Q}_M\right\} \tag{5.61a}$$

$$\mathbf{K}_{v1} = 2 \cdot \text{Re}\left\{\mathbf{Q}_{m_y}^H \mathbf{J}_{v2} \mathbf{Q}_M\right\} \quad \mathbf{K}_{v2} = 2 \cdot \text{Im}\left\{\mathbf{Q}_{m_y}^H \mathbf{J}_{v2} \mathbf{Q}_M\right\} \tag{5.61b}$$

These real-valued selection matrices are thus used to obtain the real-valued invariance equation.

5.5.2.3 DOA Estimation

Let the columns of $\mathbf{E}_s$ span the estimated real-valued signal subspace, which is the dominant subspace of $\Gamma(\mathbf{X})$ as discussed in Section 5.4. As before, $\mathbf{E}_s$ may be computed as the d largest singular vectors of $\Gamma_s(\mathbf{X})$ (direct data approach) or the d dominant eigenvectors of $\Gamma_s(\mathbf{X})\Gamma_s(\mathbf{X})^H$ (covariance approach). $\mathbf{E}_s$ and $\mathbf{B}$ span the same signal subspace. Therefore, there is a nonsingular matrix $\mathbf{T}_A$ of size $d \times d$ such that

$$\mathbf{B} = \mathbf{E}_s \mathbf{T}_A \tag{5.62}$$

Substitution of the above relation into (5.60) gives two real-valued invariance equations:

$$\begin{aligned} \mathbf{K}_{\mu 1}\mathbf{E}_s \Upsilon_\mu &\approx \mathbf{K}_{\mu 2}\mathbf{E}_s \in R^{m_x \times d} \\ \mathbf{K}_{v 1}\mathbf{E}_s \Upsilon_v &\approx \mathbf{K}_{v 2}\mathbf{E}_s \in R^{m_y \times d} \end{aligned} \tag{5.63}$$

where the real-valued matrices can be expressed as

$$\begin{aligned} \Upsilon_\mu &= \mathbf{T}_A \Omega_\mu \mathbf{T}_A^{-1} \in R^{d \times d} \\ \Upsilon_v &= \mathbf{T}_A \Omega_v \mathbf{T}_A^{-1} \in R^{d \times d} \end{aligned} \tag{5.64}$$

The real-valued diagonal matrices

$$\begin{aligned} \Omega_\mu &= \text{diag}\left\{\tan\left(\frac{\mu_1}{2}\right), \tan\left(\frac{\mu_2}{2}\right), \ldots, \tan\left(\frac{\mu_d}{2}\right)\right\} \\ \Omega_v &= \text{diag}\left\{\tan\left(\frac{v_1}{2}\right), \tan\left(\frac{v_2}{2}\right), \ldots, \tan\left(\frac{v_d}{2}\right)\right\} \end{aligned} \tag{5.65}$$

contain the desired spatial frequency (DOA) information.

As in the one-dimensional case, the two real-valued invariance equations in (5.63) can be solved independently by LS or TLS. However, the real-valued eigenvalues of the solution Υ_μ and Υ_v to the above invariance equations are given by $\tan(\mu_i/2)$ and $\tan(v_i/2)$, respectively. If the eigenvalues of these matrices are calculated independently, it is very

difficult to pair the resulting two distinct estimates of the frequency estimates (i.e., the elevation angle pair of azimuthal angle). To overcome this problem of pairing, an automatic pairing scheme is incorporated in the two-dimensional unitary ESPRIT [9].

5.5.2.4 Automatic Pairing of Spatial Frequency Estimates

The real-valued matrices Υ_μ and Υ_ν share the same set of eigenvectors. It is possible to choose one real-valued eigenvector matrix $\mathbf{T}$ such that $\Upsilon_\mu = \mathbf{T}\Omega_\mu\mathbf{T}^{-1}$ and $\Upsilon_\nu = \mathbf{T}\Omega_\nu\mathbf{T}^{-1}$ are real-valued. However, due to additive noise and the finite number of snapshots N, the real-valued matrices Υ_μ and Υ_ν do not exactly have the same set of eigenvectors. As a result, obtaining eigenvectors from only one of the matrices will give wrong estimates as this would be a random choice since the information in the other matrix is not being used. Also, Υ_μ and Υ_ν might have some degenerate (multiple) eigenvalues, though they have common set of eigenvectors.

This problem is solved in two-dimensional unitary ESPRIT by "making a complex matrix" $\Upsilon_\mu + j\Upsilon_\nu$ [9]. An automatic pairing of the spatial frequency estimates μ_i and ν_i can then be obtained by computing the eigenvalues of this matrix as

$$\Upsilon_\mu + j\Upsilon_\nu = \mathbf{T}\left(\Omega_\mu + j\Omega_\nu\right)\mathbf{T}^{-1} \tag{5.66}$$

Hence, the automatically paired estimates of Ω_μ and Ω_ν are obtained by the real and imaginary parts of the complex eigenvalues of $\Omega_\mu + j\,\Omega_\nu$. Mathematically,

$$\Omega_\mu + j\Omega_\nu = \mathbf{T}\Lambda\mathbf{T}^{-1} \text{ with } \Lambda = \text{diag}\left\{\lambda_1, \lambda_2, \ldots, \lambda_i, \ldots, \lambda_d\right\} \tag{5.67}$$

where

$$\mu_i = 2\arctan\left(\text{Re}\left\{\lambda_i\right\}\right), \quad 1 \le i \le d \tag{5.68a}$$

$$\nu_i = 2\arctan\left(\text{Im}\left\{\lambda_i\right\}\right), \quad 1 \le i \le d \tag{5.68b}$$

A summary of computations of two-dimensional unitary ESPRIT is given in Table 5.3 [1].

Table 5.3
Summary of Two-Dimensional Unitary ESPRIT in Element Space

1. *Real-Valued Signal Subspace Estimation:* Compute $\mathbf{E}_S$

   as the d dominant left singular vectors of $\Gamma_S(\mathbf{X})$ (square-root approach)
   or as the d dominant eigenvectors of $\Gamma_S(\mathbf{X})\Gamma_S(\mathbf{X})^H$ (covariance approach)

2. *Real-Valued Invariance Equation:* Solve this equation for Υ_μ and Υ_ν

   $$\mathbf{K}_{\mu 1}\mathbf{E}_\mathbf{s}\Upsilon_\mu \approx \mathbf{K}_{\mu 2}\mathbf{E}_\mathbf{s} \quad \text{and} \quad \mathbf{K}_{\nu 1}\mathbf{E}_\mathbf{s}\Upsilon_\nu \approx \mathbf{K}_{\nu 2}\mathbf{E}_\mathbf{s}$$

   by means of LS or TLS

3. *DOA Estimation:* Calculate the eigenvalues of the complex-valued $d \times d$ matrix

   $$\Omega_\mu + j\Omega_\nu = \mathbf{T}\Lambda\mathbf{T}^{-1} \text{ with } \Lambda = \text{diag}\{\lambda_i\}_{i=1}^{d}$$

   and then extract the DOA angular information via

   $$\mu_i = 2\arctan(\text{Re}\{\lambda_i\}), \quad 1 \le i \le d$$
   $$\nu_i = 2\arctan(\text{Im}\{\lambda_i\}), \quad 1 \le i \le d$$
   $$\phi_i = \arg(u_i - jv_i) \text{ and } \theta_i = \arcsin(\|u_i + jv_i\|)$$

5.6 Beamspace Transformation

The performance improvement of DOA estimation algorithms is generally called for antenna arrays composed of a large number of elements. Since the computational requirements are directly related to the dimensions of the collected data, the burden increases rapidly with the number of elements. The ESPRIT described so far requires that the data received by all elements of the array be available in digital form. In many applications, arrays of 10 or more elements are not uncommon, for example, in radar applications. This requires high number of front-ends and A/D converters that may be prohibitive in light of element space processing [10, 11]. In this section, we describe a technique called beam space processing to reduce the dimension of the observation vector. Algorithms that operate in beamspace first project the received (element space) data into a subspace of lower dimensions, the beamspace. Then the beamspace data are processed in the lower dimensional beamspace, reducing computational complexity.

Consider an M-element ULA with N snapshots of data taken. If $\mathbf{x}(t)$ is an element space snapshot of data matrix that is of size $M \times N$, then a reduced dimensional beamspace snapshot, $\mathbf{y}(t)$ of size $L \times N$ with $L < M$, is obtained by a linear transformation of the form

$$\mathbf{y}(t) = \mathbf{W}^b\mathbf{x}(t)$$

The reduced-dimension observation vector $\mathbf{y}(t)$ is usually referred as the beamspace data and $\mathbf{W}^b$ is referred to as the beamspace transformation of the size $L \times M$. In the case of uniform linear arrays, fast Fourier transform algorithms like DFT are used for beamspace transformation. Naturally, processing the beamspace data, though reducing computational loads, also implies a loss of information. The beamspace transformation can be thought of as a multichannel beamformer. By designing the beamformers (the rows of $\mathbf{W}^b$), a relatively narrow DOA sector is focused on, with the essential information in $\mathbf{x}(t)$ regarding incoming signals retained in $\mathbf{y}(t)$. Hence, beamspace processing is best applied when one has a priori information on the general angular locations of the signal arrivals, as in radar applications. In this technique, some sort of mapping that reduces the dimension of the data set is used before any signal DOA processing algorithm is applied. Another advantage is that if this transformation mapping is implemented in analog hardware, a smaller number of analog-to-digital converters are needed, compared to the number needed in element space implementation. An illustration of the beamspace processing is shown in Figure 5.4(a) where the beamspace processing has the narrowed DOA sector, while Figure 5.4(b) is the general element space DOA estimation.

ESPRIT-based algorithms require a special structure of array that has a shift invariance. In the following section, it is explained how this

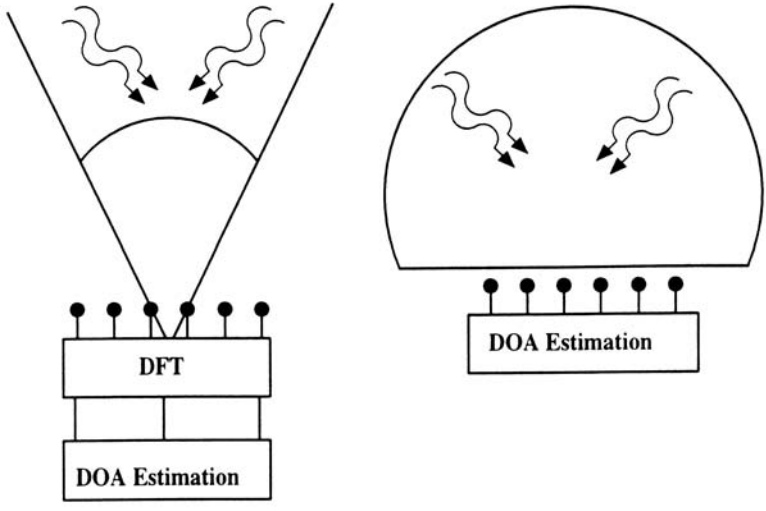


Figure 5.4 (a) Beamspace and (b) element space processing.

shift invariance structure is preserved in the array after the beamspace transformation is applied.

5.6.1 DFT Beamspace Invariance Structure

In this section, invariance structure of beamspace data is shown. To prove this, the element space array steering matrix is considered. In case of a ULA, the transformation from element space to beamspace is accomplished by premultiplying the element space data with those rows of a DFT matrix that forms beams encompassing the DOA sector of interest. This is carried out by exploiting the fact that each row of the DFT matrix forms a beam towards a specific angle.

Consider a ULA with maximum overlap. Assume the signal and data model of Chapter 3 and assume that the center of the ULA is chosen as the phase reference center. The element space array steering matrix $\mathbf{A}_c$ of this array, as discussed in Chapter 3, is given by (3.23). The columns of $\mathbf{A}_c$ are written here again for ready reference.

$$\mathbf{a}_c(\mu_i) = e^{-j\left(\frac{M-1}{2}\right)\mu_i}\left[1 \;\; e^{j\mu_i} \;\; e^{j2\mu_i} \;\; \ldots \;\; e^{j(M-1)\mu^i}\right]^T, \quad 1 \le i \le d \tag{5.69}$$

First, suppose that all M DFT beams are in use. That is, beamspace is made equal to element space. In other words, a scaled M-point DFT matrix, $\mathbf{W}^b \in C^{M\times M}$, is applied. Each row of this DFT matrix $\mathbf{W}^b$ represents a beamforming weight vector needed to steer the array to the direction $(k-1)\cdot 2\pi/M$, $1 \le k \le M$; it can be expressed as:

$$\begin{aligned}\mathbf{w}_k^b &= e^{j\left(\frac{M-1}{2}\right)(k-1)\frac{2\pi}{M}}\left[1 \;\; e^{-j(k-1)\frac{2\pi}{M}} \;\; e^{-j2(k-1)\frac{2\pi}{M}} \;\; \ldots \;\; e^{-j(M-1)(k-1)\frac{2\pi}{M}}\right] \\ &= e^{j\left(\frac{M-1}{2}\right)\gamma_k}\left[1 \;\; e^{-j\gamma_k} \;\; e^{-j2\gamma_k} \;\; \ldots \;\; e^{-j(M-1)\gamma_k}\right], \quad 1 \le k \le M\end{aligned} \tag{5.70}$$

where $\gamma_k = (k-1)\cdot 2\pi/M$, $1 \le k \le M$. The row vector $\mathbf{w}_k^b$ in beamspace represents a DFT beam steered at the spatial frequency γ_k.

Now, a DFT beamspace steering matrix $\mathbf{B}$ is obtained from element space steering matrix $\mathbf{A}_c$ by

$$\mathbf{B} = \mathbf{W}^b\mathbf{A}_c = \left[\mathbf{b}(\mu_1)\;\mathbf{b}(\mu_2)\ldots\mathbf{b}(\mu_d)\right] \in R^{M\times d} \tag{5.71}$$

where $\mathbf{b}(\mu_i)$, $1 \le i \le d$, represents a DFT beamspace steering vector. This DFT beamspace transformed matrix is real-valued since the transformation matrix $\mathbf{W}^b$ is left Π-real as $\overline{\mathbf{W}}^b = \mathbf{W}^b$.

The relation between two adjacent components of the beamspace is derived in the following, which helps to get an insight on how to focus on a particular sector [12]. Consider the ith component of the real-valued DFT beamspace steering vector

$$\mathbf{b}(\mu_i) = \left[b_1(\mu_i)\ b_2(\mu_i) \ldots b_M(\mu_i)\right]^T, \quad 1 \le i \le d \tag{5.72}$$

From (5.71),

$$\begin{aligned} b_k(\mu_i) &= \mathbf{w}_k^b \mathbf{a}_c(\mu_i) \\ &= e^{j\left(\frac{M-1}{2}\right)\gamma_k} e^{-j\left(\frac{M-1}{2}\right)\mu_i} \sum_{l=1}^{M} e^{-j(l-1)\mu_i} \\ &= e^{j\left(\frac{M-1}{2}\right)(\gamma_k-\mu_i)} \sum_{l=0}^{M-1} e^{jl(\mu_i-\gamma_k)} \\ &= \frac{e^{-j\frac{M}{2}(\mu_i-\gamma_k)}}{e^{-j\frac{1}{2}(\mu_i-\gamma_k)}} \cdot \frac{1-e^{jM(\mu_i-\gamma_k)}}{1-e^{j(\mu_i-\gamma_k)}} \\ &= \frac{e^{-j\frac{M}{2}(\mu_i-\gamma_k)} - e^{j\frac{M}{2}(\mu_i-\gamma_k)}}{e^{-j\frac{1}{2}(\mu_i-\gamma_k)} - e^{j\frac{1}{2}(\mu_i-\gamma_k)}} = \frac{\sin\left(\frac{M}{2}(\mu_i-\gamma_k)\right)}{\sin\left(\frac{1}{2}(\mu_i-\gamma_k)\right)} \end{aligned} \tag{5.73}$$

where the sum of the finite geometric series $\sum_{l=0}^{n} a^l = \frac{1-a^{n+1}}{1-a}$ is used.

Consider a component adjacent to $b_k(\mu_i)$. From (5.73), we have

$$b_{k+1}(\mu_i) = \frac{\sin\left(\frac{M}{2}(\mu_i-\gamma_{k+1})\right)}{\sin\left(\frac{1}{2}(\mu_i-\gamma_{k+1})\right)} = \frac{\sin\left(\frac{M}{2}\left(\mu_i-(k+1)\frac{2\pi}{M}\right)\right)}{\sin\left(\frac{1}{2}\left(\mu_i-(k+1)\frac{2\pi}{M}\right)\right)} \tag{5.74}$$

By comparing (5.74) with (5.73), it can be readily observed that the numerator of $b_{i,k+1}(\mu_i)$ in (5.74) is negative of that of $b_k(\mu_i)$, since

$$\sin\left(\frac{M}{2}\left(\mu_i - k\frac{2\pi}{M}\right)\right) = \sin\left(\frac{M}{2}\left(\mu_i - (k-1)\frac{2\pi}{M}\right) - \pi\right)$$
$$= -\sin\left(\frac{M}{2}(\mu_i - \gamma_k)\right)$$

Hence, any two successive components of the beamspace steering vector $\mathbf{b}_i(\mu_i)$ are related as

$$\sin\left(\frac{1}{2}(\mu_i - \gamma_k)\right) b_k(\mu_i) = -\sin\left(\frac{1}{2}(\mu_i - \gamma_{k+1})\right) b_{k+1}(\mu_i)$$

Using that trigonometric identity $\sin(a + b) = \sin(a)\sin(b) - \cos(a)\cos(b)$, we get

$$\begin{aligned}&\left\{\sin\left(\frac{\mu_i}{2}\right)\cos\left(\frac{\gamma_k}{2}\right) - \cos\left(\frac{\mu_i}{2}\right)\sin\left(\frac{\gamma_k}{2}\right)\right\} b_k(\mu_i) \\ &= -\left\{\sin\left(\frac{\mu_i}{2}\right)\cos\left(\frac{\gamma_{k+1}}{2}\right) - \cos\left(\frac{\mu_i}{2}\right)\sin\left(\frac{\gamma_{k+1}}{2}\right)\right\} b_{k+1}(\mu_i)\end{aligned} \tag{5.75}$$

with $\gamma_k = (k-1)\cdot\frac{2\pi}{M}$. Therefore,

$$\begin{aligned}&\tan\left(\frac{\mu_i}{2}\right)\left\{\cos\left((k-1)\frac{\pi}{M}\right) b_k(\mu_i) + \cos\left(k\frac{\pi}{M}\right) b_{k+1}(\mu_i)\right\} \\ &= \sin\left((k-1)\frac{\pi}{M}\right) b_k(\mu_i) + \sin\left(k\frac{\pi}{M}\right) b_{k+1}(\mu_i)\end{aligned} \tag{5.76}$$

with $1 \le k \le M$ and $1 \le i \le d$. This is the relation between any two successive components of the beamspace steering vector.

Moreover, it can be observed that the first and last components, that is, beams with indices $k = M$ and $k = 1$, are physically adjacent to each other, since they are steered at the spatial frequencies $\gamma_M - 2\pi = (M-1)\frac{2\pi}{M} - 2\pi = -\frac{2\pi}{M}$ and $\gamma_0 = 0$, respectively. This is depicted in Figure 5.5. This observation is used to derive an invariance relationship for the DFT beamspace steering vectors $b_k(\mu_i)$ necessary for the implementation of ESPRIT.

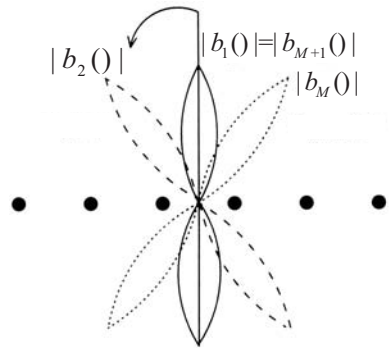


Figure 5.5 Directions of main lobes of $|b_1(\mu)|$, $|b_2(\mu)|$, …, $|b_M(\mu)|$ of a ULA.

Also the following relationship exists:

$$\begin{aligned} b_{M+1}(\mu_i) &= \frac{\sin\left(\frac{M}{2}\left(\mu_i - M\frac{2\pi}{M}\right)\right)}{\sin\left(\frac{1}{2}\left(\mu_i - M\frac{2\pi}{M}\right)\right)} \\ &= \frac{\sin\left(\frac{M}{2}\mu_i - M\pi\right)}{\sin\left(\frac{1}{2}\mu_i - \pi\right)} = \frac{(-1)^M \cdot \sin\left(\frac{M}{2}\mu_i\right)}{-\sin\left(\frac{1}{2}\mu_i\right)} \end{aligned} \tag{5.77}$$

By substituting $k = M$ into (5.76), the following relationship between $b_M(\mu_i)$ and $b_1(\mu_i)$ is then obtained as

$$\begin{aligned} &\tan\left(\frac{\mu_i}{2}\right)\left\{\cos\left((M-1)\frac{\pi}{M}\right)b_M(\mu_i) + \cos(\pi)(-1)^{(M-1)}b_1(\mu_i)\right\} \\ &= \sin\left((M-1)\frac{\pi}{M}\right)b_M(\mu_i) + \sin(\pi)(-1)^{(M-1)}b_1(\mu_i), \quad 1 \le i \le d \end{aligned} \tag{5.78}$$

Compiling all M equations in a vector form for $1 \le k \le M$ gives rise to an invariance relationship for the DFT beamspace steering vectors $\mathbf{b}_M(\mu_i)$; the relationship is similar to the invariance relationship satisfied by $\mathbf{d}(\mu_i)$ in (5.42):

$$\Gamma_1 \mathbf{b}(\mu_i) \cdot \tan\left(\frac{\mu_i}{2}\right) = \Gamma_2 \mathbf{b}(\mu_i), \quad 1 \le i \le d \tag{5.79}$$

The selection matrices Γ_1 and Γ_2 of size $M \times N$ are defined as

$$\Gamma_1 = \begin{bmatrix} 1 & \cos\left(\frac{\pi}{M}\right) & 0 & 0 & \cdots & 0 & 0 \\ 0 & \cos\left(\frac{\pi}{M}\right) & \cos\left(\frac{2\pi}{M}\right) & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cos\left(\frac{2\pi}{M}\right) & \cos\left(\frac{3\pi}{M}\right) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \cos\left((M-2)\frac{\pi}{M}\right) & \cos\left((M-1)\frac{\pi}{M}\right) \\ (-1)^M & 0 & 0 & 0 & \cdots & 0 & \cos\left((M-1)\frac{\pi}{M}\right) \end{bmatrix}$$

$$\Gamma_2 = \begin{bmatrix} 0 & \sin\left(\frac{\pi}{\mathrm{M}}\right) & 0 & 0 & \cdots & 0 & 0 \\ 0 & \sin\left(\frac{\pi}{\mathrm{M}}\right) & \sin\left(\frac{2\pi}{M}\right) & 0 & \cdots & 0 & 0 \\ 0 & 0 & \sin\left(\frac{2\pi}{M}\right) & \sin\left(\frac{3\pi}{M}\right) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \sin\left((M-2)\frac{\pi}{M}\right) & \sin\left((M-1)\frac{\pi}{M}\right) \\ 0 & 0 & 0 & 0 & \cdots & 0 & \sin\left((M-1)\frac{\pi}{M}\right) \end{bmatrix}$$

It can be seen that the last rows of Γ_1 and Γ_2 are linear combinations of the other rows, that is, both $M \times N$ selection matrices are rank-deficient. They are only of rank $(M-1)$. One of the M rows of Γ_1 and Γ_2 should therefore be dropped if all M DFT beams are employed. In reduced dimension processing, however, only a subset of row vectors encompassing the desired DOA sector is applied to the data matrix $\mathbf{X}$. Only those subblocks of the selection matrices Γ_1 and Γ_2 that relate the corresponding components of $\mathbf{b}(\mu_i)$ will be used. With d impinging signals, the real-valued DFT beamspace steering matrix $\mathbf{B}$ satisfies a shift invariance property, given by

$$\Gamma_1 \mathbf{B} \Omega = \Gamma_1 \mathbf{B},$$

where

$$\Omega = \text{diag}\left\{\tan\left(\frac{\mu_1}{2}\right), \tan\left(\frac{\mu_2}{2}\right), \ldots, \tan\left(\frac{\mu_1}{2}\right), \ldots, \tan\left(\frac{\mu_d}{2}\right)\right\} \quad (5.80)$$

Hence, it is observed that beamspace array steering matrix **B** satisfies a shift invariance property similar to the element space array steering matrix given in (5.44), thus enabling the implementation of ESPRIT-based algorithms.

5.6.2 DFT Beamspace in a Reduced Dimension

In the previous section, the beamspace transformation matrix was taken as $\mathbf{W}^b$. Hence, each beamspace steering vector $\mathbf{b}(\mu_i)$ has M components. If one has a priori knowledge of angle(s) of the impinging signal(s), then it is possible to choose only those components that span the desired DOA sector around the angle(s) [8], since each row of the matrix equation (5.80) relates two successive components of the DFT beamspace steering vectors $\mathbf{b}(\mu_i)$ (i.e., they point to successive three angles).

This observation allows us to apply $L \leq M$ successive rows of $\mathbf{W}^b$ beginning at row $k_{\min}$, $1 \leq k_{\min} \leq M$, instead of all M rows to the data matrix **X**. In the reduced number of rows, denote the resulting beamforming matrix that produces L consecutive beams as $\mathbf{W}_L^b \in C^{B\times M}$. The beamformer $\mathbf{W}_L^b$ thus narrows the scope of the search for DOAs to the spatial sector specified by

$$(k_{\min} - 1)\frac{2\pi}{M} \leq \gamma \leq (k_{\min} + L)\frac{2\pi}{M} \quad (5.81)$$

where a steering angle $\gamma > \pi$ is identical to the steering angle $\gamma - 2\pi$.

The number of rows B depends on the angle width of the DOA sector of interest and can be substantially less than the number of elements M. Consequently, the dimension of the SVD of (5.71), and therefore $\mathbf{E}_s$, along with that the real-valued invariance equation, will reduce to $L \times d$. By selecting appropriate subblocks of Γ_1 and Γ_2, the beamspace processing performs the same as element space except for its reduced dimensionality. The resulting selection matrices of size $(L-1) \times L$ shall be called $\Gamma_1^{(L)}$ and $\Gamma_2^{(L)}$ such that the B-dimensional DFT beamspace steering vectors $\mathbf{b}_L(\mu_i) = \mathbf{W}_L^b \mathbf{a}_c(\mu_i)$ still satisfy the invariance relation.

$$\Gamma_1^{(B)}\mathbf{b}_L(\mu_i)\cdot\tan\left(\frac{\mu_i}{2}\right)=\Gamma_2^{(B)}\mathbf{b}_L(\mu_i),\quad 1\le i\le d \tag{5.82}$$

Beam patterns of a ULA with $M = 8$ sensors was computed. A DFT beamspace transformation $\mathbf{W}_8^b$ was employed to form $L = 3$ beams. When rows 4, 5, and 6 were employed to form the beams, the corresponding beam patterns $b_1(\mu_i)$, $b_4(\mu_i)$, and $b_5(\mu_i)$ are obtained as shown in Figure 5.6(a). Another computation was made to verify that first and last components are physically adjacent to each other. Rows 1, 2, and 8 of $\mathbf{W}_8^b$ are employed to form $L = 3$ beams. The corresponding beam patterns are adjacent as shown in Figure 5.6(b). Employing these insights, the following sections explain the working of unitary ESPRIT in DFT beamspace.

5.7 Unitary ESPRIT in DFT Beamspace

In this section, the unitary ESPRIT algorithm is applied to uniform linear arrays in DFT beamspace. A real-valued beamspace array steering matrix is obtained by choosing the reference point at the center of the array and by applying beamspace transformation. This real-valued transformation results in a considerable reduction in computational complexity and is readily extended to uniform rectangular array [9].

5.7.1 One-Dimensional Unitary ESPRIT in DFT Beamspace

Consider a ULA of M elements with maximum overlap. Assume the data and signal model discussed in Chapter 3 and an L-dimensional beamspace. The unitary ESPRIT in DFT beamspace can be formulated in the following steps.

5.7.1.1 Transformation to Beamspace

The element space data vector $\mathbf{X}$ is first transformed into the beamspace data vector $\mathbf{Y}$ as

$$\mathbf{Y} = \mathbf{W}^b\mathbf{X} \tag{5.83}$$

The transformation to DFT beamspace $\mathbf{Y} = \mathbf{W}^b\mathbf{X} \in C^{M\times N}$ can be implemented using an FFT that exploits the Vandermonde form of the

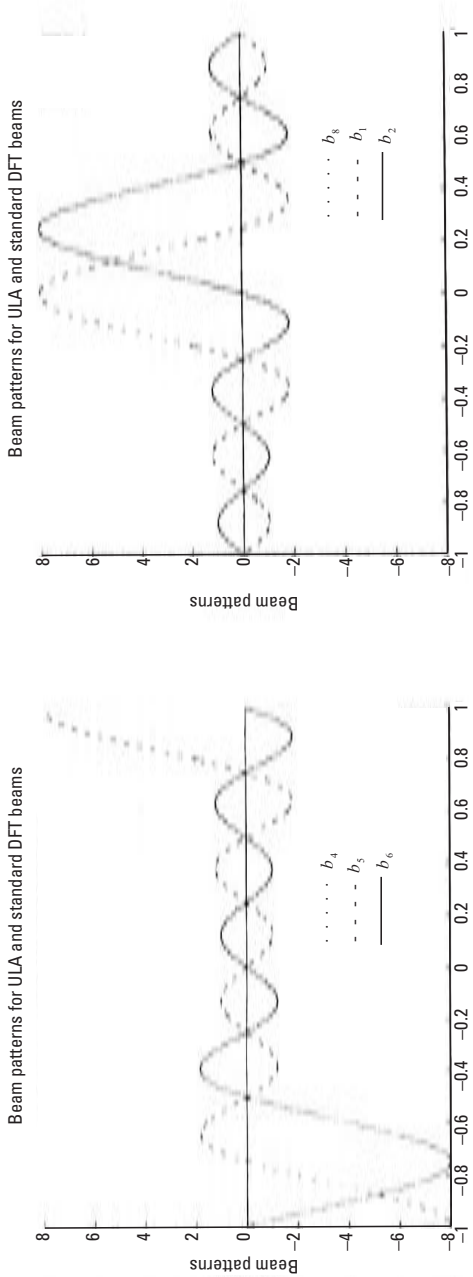


Figure 5.6 (a, b) Beam pattern of an eight-element uniform linear array.

rows of the FFT matrix, followed by an appropriate scaling of the rows of resulting matrix.

5.7.1.2 Real-Valued Signal Subspace Estimation

The unitary transformation is now applied as discussed in Section 5.4. An SVD of the real-valued matrix

$$\left[\text{Re}\,\mathbf{Y} \quad \text{Im}\,\mathbf{Y}\right] \in R^{M\times 2N} \text{ with } \mathbf{Y} = \mathbf{Y}\mathbf{W}^{b}\mathbf{X} \in C^{M\times N} \tag{5.84}$$

is computed to estimate the d left singular vectors that correspond to the d largest singular values of (5.84); otherwise, an EVD of

$$\text{Re}\{\mathbf{X}\}\,\text{Re}\{\mathbf{Y}\}^{H} + \text{Im}\{\mathbf{X}\}\,\text{Im}\{\mathbf{Y}\}^{H} \tag{5.85}$$

can also be computed (i.e., covariance approach). This transformation, as described in Section 5.4, automatically achieves forward-backward averaging.

5.7.1.3 Real-Valued Invariance Equation

Let the d singular vectors corresponding to the d largest singular values of (5.84) be denoted by $\mathbf{E}_s \in R^{M\times d}$. Asymptotically, the real-valued matrices $\mathbf{E}_s$ and $\mathbf{B}$ span the same d-dimensional signal subspace, so there is a nonsingular matrix $\mathbf{T}_A$ such that

$$\mathbf{B} \approx \mathbf{E}_s\,\mathbf{T}_A \tag{5.86}$$

Substituting this relation into (5.45), we have the real-valued invariance equation

$$\Gamma_1\mathbf{E}_s\mathbf{Y} \approx \Gamma_{2s} \in R^{M\times d}$$

where

$$\mathbf{Y} = \mathbf{T}_A\Omega\mathbf{T}_A^{-1} \tag{5.87}$$

which is solved by LS or TLS as explained in previous sections.

5.7.1.3 DOA Estimation

The spatial frequencies or DOAs are then estimated by taking eigenvalues of the solution of the invariance equation given by

$$\Gamma = \mathbf{T}_A \Omega \mathbf{T}_A^{-1}$$

where

$$\Omega = \text{diag}\{\omega_1, \omega_2, \ldots, \omega_i, \ldots, \omega_d\}$$

contains the desired DOA information. The whole algorithm that operates in an L-dimensional DFT beamspace is summarized in Table 5.4 and its performance is analyzed in Chapter 6 [1].

In the next section, the concept of DFT beamspace is extended to two-dimensional arrays.

5.7.2 Two-Dimensional Unitary ESPRIT in DFT Beamspace

In this section, DFT beamspace technique is applied to two-dimensional arrays, more specifically to uniform rectangular arrays. Consider a uniform rectangular array of size $M = M_x \times M_y$. Assume the signal and data

Table 5.4
Summary of One-Dimensional Unitary ESPRIT in DFT Beamspace

1. *Transformation to Beamspace:* $\mathbf{Y} = \mathbf{W}^b\mathbf{X}$
2. *Signal Subspace Estimation:* Compute $\mathbf{E}_S$
   as the d dominant left singular vectors of [Re{**Y**} Im{**Y**}] (square-root approach)
   or as the d dominant eigenvectors of $\text{Re}\{\mathbf{X}\}\text{Re}\{\mathbf{Y}\}^H + \text{Im}\{\mathbf{X}\}\text{Im}\{\mathbf{Y}\}^H$ (covariance approach)
3. *Solution of the Invariance Equation:* Solve the following equation for Υ
   $$\gamma_1^{(B)}\mathbf{E}_s\Upsilon \approx \Gamma_2^{(B)}\mathbf{E}_s$$
   by means of LS or TLS
4. *DOA Estimation:* Calculate the eigenvalues of the real-valued solution
   $$\gamma = \mathbf{T}\Omega\mathbf{T}^{-1} \text{ with } \Omega = \text{diag}\{\omega_i\}_{i=1}^{d}$$
   Then extract the DOA angular information via
   $$\mu_i = 2\arctan(\omega_i), \quad 1 \le i \le d$$
   $$\theta_i = \arcsin\left(-\frac{\lambda}{2\pi\Delta}\mu_i\right)$$

model discussed in Chapter 3. By following the similar line for beamspace processing of one-dimensional ULA, we can compute the two-dimensional DFT beamspace ESPRIT in a straightforward manner.

5.7.2.1 Transformation to Beamspace

Consider the case of reduced dimensionality directly and assume that L_x out of M_x beams in the x direction and L_y out of M_y beams in the y direction are formed, respectively. The total number of beams is $L = L_x \times L_y$. The corresponding scaled DFT matrices $\mathbf{W}_{L_x}^{b} \in C^{B_x \times M_x}$ and $\mathbf{W}_{L_y}^{b} \in C^{B_y \times M_y}$ are formed as discussed in Section 5.6.2.

Let the noise corrupted signals or data received by the array at any given time t_n be given by the matrix $\chi(t_n) \in C^{M_x \times M_y}$. Premultiplying this matrix by $\mathbf{W}_{L_x}^{b}$ and postmultiplying by $\mathbf{W}_{L_y}^{b}$ and applying the vec{•} operator, we can have

$$y(t_n) = \text{vec}\left\{\mathbf{W}_{L_x}^{H} \cdot \mathbf{X}(t_n) \cdot \overline{\mathbf{W}}_{L_y}^{b}\right\} \in C^{B}, \quad L = L_x \times L_y \tag{5.88}$$

We place this column vector as a column of the matrix $\mathbf{Y} \in C^{L \times N}$. The vec{•} operator maps a $B_x \times B_y$ matrix to a $B \times 1$ vector by stacking the columns of the matrix. Again by applying the property of Kronecker products (Appendix, Section A.1) to (5.88), we get

$$\begin{aligned} Y &= \left[\mathbf{y}(t_1)\ \mathbf{y}(t_2) \ldots\ \mathbf{y}(t_n)\right] \\ &= \left(W_{B_y}^{b} \otimes W_{B_x}^{b}\right)\left[\mathbf{x}(t_1)\ \mathbf{x}(t_1) \ldots\ \mathbf{x}(t_n)\right] \\ &= \left(W_{B_y}^{b} \otimes W_{B_x}^{b}\right) \cdot \mathbf{X} \in C^{B \times N} \end{aligned} \tag{5.89}$$

Signal subspace now can be estimated from this data.

5.7.2.2 Real-Valued Subspace Estimation

Let $\mathbf{E}_s$ denote the real-valued signal subspace. The columns of $\mathbf{E}_s \in R^{L \times d}$ contain the d left singular vectors of

$$\left[\text{Re}\{\mathbf{Y}\} \quad \text{Im}\{\mathbf{Y}\}\right] \in R^{L \times 2N} \tag{5.90}$$

corresponding to its d largest singular values. Otherwise, using EVD, the d dominant eigenvectors of

$$\mathrm{Re}\{\mathbf{Y}\}\mathrm{Re}\{\mathbf{Y}\}^H + \mathrm{Im}\{\mathbf{Y}\}\mathrm{Im}\{\mathbf{Y}\}^H \tag{5.91}$$

spans the real-valued signal space.

5.7.2.3 Real-Valued Invariance Equation

The transformed real-valued array manifold matrix is now given by

$$\begin{aligned} B(\mu_i, v_i) &= \mathbf{W}_{L_x}^b \cdot A(\mu_i, v_i) \cdot \overline{\mathbf{W}}_{L_y}^b \\ &= \mathbf{W}_{L_x}^b \cdot \mathbf{a}_{M_x}(\mu_i)\mathbf{a}_{M_y}^T(v_i)\overline{\mathbf{W}}_{L_y}^b \\ &= \mathbf{b}_{L_x}(\mu_i)\mathbf{b}_{L_y}^T(v_i) \in R^{L_x \times L_y} \end{aligned} \tag{5.92}$$

where

$$\mathbf{b}_{L_x}(\mu_i) = \mathbf{W}_{L_x}^b \cdot \mathbf{a}_{M_x}(\mu_i) \in R^{L_x} \text{ and } \mathbf{b}_{L_y}(v_i) = \mathbf{W}_{L_y}^v \cdot \mathbf{a}_{M_y}(v_i) \in R^{L_y}$$

are the one-dimensional DFT beamspace manifold vectors in a reduced dimensional space. The one-dimensional array steering vectors $\mathbf{a}_{M_x}(\mu_i) \in C^{M_x}$ and $\mathbf{a}_{M_y}(v_i) \in C^{My}$ are the same as those defined in (2.24).

Consider the one-dimensional DFT beamspace vectors in the x direction. As $\mathbf{b}_{L_x}(\mu_i)$ satisfies the invariance relationship in (5.82), it follows that $B(\mu_i, v_i)$ satisfies

$$\Gamma_1^{(L_x)} B(\mu_i, v_i) \cdot \tan\left(\frac{\mu_i}{2}\right) = \Gamma_2^{(L_x)} B(\mu_i, v_i) \tag{5.93}$$

where the one-dimensional selection matrices in the x direction, $\Gamma_1^{(L_x)}$ and $\Gamma_2^{(L_x)}$, were defined as before for the one-dimensional DFT beamspace. By using the property of the vec{•} operator in the Appendix, Section A.1, $L = L_x \times L_y$ dimensional DFT beamspace steering vectors $\mathbf{b}(\mu_i, v_i) = \mathrm{vec}\{\boldsymbol{B}(\mu_i, v_i)\}$satisfy

$$\Gamma_{\mu 1}\mathbf{b}(\mu_i, v_i)\cdot\tan\left(\frac{\mu_i}{2}\right)=\Gamma_{\mu 2}\mathbf{b}(\mu_i, v_i),\quad 1\le i\le d \tag{5.94}$$

where the two-dimensional selection matrices in the x direction

$$\Gamma_{\mu 1}=\mathbf{I}_{L_y}\otimes\Gamma_1^{(L_x)}\ \text{and}\ \Gamma_{\mu 2}=\mathbf{I}_{L_y}\otimes\Gamma_2^{(L_x)} \tag{5.95}$$

are of size $b_x\times L$ with $b_x=(L_x-1)\cdot L_y$. Similarly, the one-dimensional DFT beamspace manifold vectors in the y direction, $\mathbf{b}_{L_y}(\nu_i)$, satisfy

$$\Gamma_1^{(L_y)}\mathbf{b}_{L_y}(v_i)\cdot\tan\left(\frac{v_i}{2}\right)=\Gamma_2^{(L_y)}\mathbf{b}_{L_y}(v_i) \tag{5.96}$$

such that

$$B(\mu_i, v_i)\Gamma_1^{(L_y)}\cdot\tan\left(\frac{v_i}{2}\right)=B(\mu_i, v_i)\Gamma_2^{(L_y)} \tag{5.97}$$

Again after the vec{•} operator is applied,

$$\Gamma_{v1}\mathbf{b}(\mu_i, v_i)\cdot\tan\left(\frac{v_i}{2}\right)=\Gamma_{v2}\mathbf{b}(\mu_i, v_i),\quad 1\le i\le d \tag{5.98}$$

where the two-dimensional selection matrices in the y direction

$$\Gamma_{v1}=\Gamma_1^{(L_x)}\otimes\mathbf{I}_{L_x}\ \text{and}\ \Gamma_{v2}=\Gamma_2^{(L_y)}\otimes\mathbf{I}_{L_y} \tag{5.99}$$

are of size $b_y\times L$ with $b_y=L_x\cdot(L_y-1)$.

The real-valued two-dimensional DFT beamspace steering matrix is then given as

$$\begin{aligned}\mathbf{B}&=\left(\mathbf{W}_{L_y}^{H}\otimes\mathbf{W}_{L_x}^{H}\right)\mathbf{A}_c\\&=\left[b(\mu_1, v_1)\ b(\mu_2, v_2)\cdots b(\mu_d, v_d)\right]\in R^{L\times d}\end{aligned} \tag{5.100}$$

Hence, **B** satisfies the following two invariance properties,

$$\Gamma_{\mu 1}\mathbf{B}\cdot\Omega_{\mu}=\Gamma_{\mu 2}\mathbf{B} \tag{5.101}$$

$$\Gamma_{v_1}\mathbf{B}\cdot\Omega_{v}=\Gamma_{v 2}\mathbf{B} \tag{5.102}$$

where the real-valued diagonal matrices

$$\begin{aligned}\Omega_{\mu}&=\operatorname{diag}\left\{\tan\left(\frac{\mu_1}{2}\right),\tan\left(\frac{\mu_2}{2}\right),\ldots\tan\left(\frac{\mu_d}{2}\right)\right\}\text{ and}\\ \Omega_{v}&=\operatorname{diag}\left\{\tan\left(\frac{v_1}{2}\right),\tan\left(\frac{v_2}{2}\right),\ldots\tan\left(\frac{v_d}{2}\right)\right\}\end{aligned} \tag{5.103}$$

contain the desired DOA information to be estimated. Similar to the one-dimensional approach, it is observed that $\mathbf{E}_s$ and $\mathbf{B}$ span the same d-dimensional subspace. Therefore, there is a nonsingular matrix $\mathbf{T}_A$ of size $d\times d$ such that $\mathbf{B}\approx\mathbf{E}_s\mathbf{T}_A$. Substitution of this relationship into (5.94) yields two real-valued invariance equations

$$\Gamma_{\mu 1}\mathbf{E}_s\Upsilon_{\mu}\approx\Gamma_{\mu 2}\mathbf{E}_s\in R^{b_x\times d}\text{ and }\Gamma_{v1}E_s\Upsilon_{v}\approx\Gamma_{v 2_s}\,\mathbf{E}\in R^{b_y\times d} \tag{5.104}$$

5.7.2.4 DOA Estimation

As in the element space version of the two-dimensional unitary ESPRIT, the real-valued matrices $\Upsilon_{\mu}\in R^{d\times d}$ and $\Upsilon_{v}\in R^{d\times d}$, which contain the information about the spatial frequencies, can be calculated as the LS or TLS solution of these two real-valued invariance equations (5.104). Finally, automatically paired spatial frequency estimates μ_i and v_i, $1\le i\le d$, are obtained from the real and imaginary parts of the eigenvalues of the "complexified" matrix $\Upsilon_{\mu}+j\Upsilon_{v}$. Here the maximum number of the signals that can be handled by the array is the smaller of b_x and b_y, under the assumption that at least $d/2$ snapshots are available.

A complete summary of the computational steps is tabulated in Table 5.5 [1].

5.8 Conclusion

This chapter discussed in detail and systematically summarized the algorithms based on ESPRIT. ESPRIT is applicable to array geometries that are composed of two identical subarrays with a fixed displacement

Table 5.5
Summary of Two-Dimensional Unitary ESPRIT in DFT Beamspace

1. *Transformation to Beamspace:* Compute the two-dimensional DFT of array outputs at each snapshot and apply the vec(•) operator and place the result as column

$$\Upsilon = \left(\mathbf{W}_{B_y}^{b} \otimes \mathbf{W}_{B_x}^{b}\right)\mathbf{X}, \quad B = B_x \cdot B_y$$

2. *Signal Subspace Estimation:* Compute $\mathbf{E}_S$

   as the d dominant left singular vectors of [Re{**Y**} Im{**Y**}] (square-root approach)

   or as the d dominant eigenvectors of Re{**Y**} Re{**Y**}^{H} + Im{**Y**}Im{**Y**}^{H} (covariance approach) .

3. *Solution of the Invariance Equation:* Solve the following equations for

$$\Gamma_{\mu 1}\mathbf{E}_s \Upsilon_\mu \approx \Gamma_{\mu 2}\mathbf{E}_s \text{ and } \Gamma_{\nu 1}\mathbf{E}_s \Upsilon_\nu \approx \Gamma_{\nu 2}\mathbf{E}_s$$

   by means of LS or TLS

4. *Spatial Frequency Estimation:* Calculate the eigenvalues of the complex-valued $d \times d$ matrix

$$\Gamma_\mu + j\Upsilon_\nu = \mathbf{T}\Lambda\mathbf{T}^{-1} \text{ with } \Lambda = \text{diag}\{\lambda_i\}_{i=1}^{\delta}$$

   Then extract the angular DOA information via

$$\mu_i = 2\arctan\left(\text{Re}\{\lambda_i\}\right), \quad 1 \le i \le d$$
$$\nu_i = 2\arctan\left(\text{Im}\{\lambda_i\}\right), \quad 1 \le i \le d$$

between them. ESPRIT exploits such a displacement property which translates into an underlying rotational invariance of signal subspaces spanned by two data vectors received by two subarrays.

Unitary ESPRIT retains the simplicity and high resolution capability of a standard ESPRIT scheme and achieves a superior performance with a reduced computational complexity. It is used with centro-symmetric array configurations and includes forward-backward averaging. In contrast to standard ESPRIT, unitary ESPRIT is formulated in terms of real-valued computations. In the two-dimensional case, unitary ESPRIT can obtain automatically paired azimuth and elevation angle estimates.

Finally, DFT beamspace versions of one- and two-dimensional unitary ESPRIT are also described. DFT beamspace unitary ESPRIT enables the DOA processing on a particular directional sector of interest and thereby reduces the computational complexity significantly. However, this is under the assumption that a priori knowledge of where the signal sources of interest is required.

References

[1] Haardt, M., *Efficient One-, Two-, and Multidimensional High-Resolution Array Signal Processing*, New York: Shaker Verlag, 1997.

[2] Roy, R., and T. Kailath, "ESPRIT-Estimation of Signal Parameters Via Rotational Invariance Techniques," *IEEE Trans. on Acoust., Speech, Signal Processing*, Vol. 37, No. 7, July 1989, pp. 984–995.

[3] Rao, B. D., and K. S. Arun, "Model Based Processing of Signals: A State Space Approach," *Proc. IEEE*, Vol. 80, No. 2, February 1992, pp. 283–309.

[4] Kariya, T., and H. Kurata, *Generalized Least Squares*, New York: John Wiley & Sons, 2004.

[5] Golub, G. H., and C. F. van Loan, *Matrix Computations*, 3rd ed., Baltimore, MD: John Hopkins University Press, 1996.

[6] Haardt, M., and J. A. Nossek, "Unitary ESPRIT: How to Obtain Increased Estimation Accuracy with a Reduced Computational Burden," *IEEE Trans. on Signal Processing*, Vol. 43, No. 5, May 1995, pp. 1232–1242.

[7] Lee, A., "Centrohermitian and Skew-Centrohermitian Matrices," *Linear Algebra and Its applications*, Vol. 29, 1980, pp. 205–210.

[8] Haardt, M., " 2D Unitary ESPRIT For Efficient 2D Parameter Estimation," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Vol. 3, May 9–12, 1995, pp. 2096–2099.

[9] Zoltowski, M. D., M. Haardt, and C. P. Mathews, "Closed-Form 2D Angle Estimation with Rectangular Arrays in Element Space or Beamspace Via Unitary ESPRIT," *IEEE Trans. on Signal Processing*, Vol. 44, No. 2, February 1996, pp. 316–328

[10] Xu, G., et al., "Beamspace ESPRIT," *IEEE Trans. on Signal Processing*, Vol. 42, No. 2, February 1994, pp. 349–355.

[11] Gansman, J. A., M. D. Zoltowski, and J. V. Krogmeier, "Multidimensional Multirate DOA Estimation in Beamspace," *IEEE Trans. on Signal Processing*, Vol. 44, No. 11, November 1996, pp. 2780–2792.

6

Analysis of ESPRIT-Based DOA Estimation Algorithms

6.1 Introduction

ESPRIT-based DOA estimation algorithms are computationally efficient and simple compared to other subspace-based algorithms. In particular, they do not need any array calibration. They are also able to combat coherent environments efficiently. Thereby, algorithms belonging to the family of ESPRIT became popular and are opted for direction finding applications. In this chapter, we present various simulations conducted with analyses of the performance of the algorithms both in uncorrelated and coherent environments. The performance of these algorithms is extensively tested and analyzed against the following parameters:

- *Number of snapshots:* This simulation allows the understanding of the impact of the number of samples of data (or snapshots taken) necessary in obtaining a reliable estimate of the DOA of the signal. Obviously, the more the number of data samples, the more accurate is the estimate obtained. However, the cost of having more data is the requirement of more memory and more computational time. Therefore, it is practically important to study the trade-off between the number of snapshots necessary and the DOA estimation accuracy.

- *Array signal-to-noise ratio (SNR):* The SNR observed at the antenna array plays a vital role as that in all communication problems. This experiment demonstrates the variations in performance with a variation of SNR. The higher the SNR is, the better the performance. This simulation studies the behaviors versus SNR.

- *Number of signals:* The number of signals impinging on the array (or the model order) casts a basic limitation on performance of an algorithm because of the fundamental fact that the number of signals should be less than the number of elements minus one for an antenna array. Theoretically, an M element antenna can detect $M - 1$ number of signals at maximum. However, in a realistic environment with noises, the performance deteriorates and an array does not behave as theoretically expected. This simulation attempts to study this by checking the performance against the number of signals.

- *Number of elements:* This simulation studies the DOA estimation accuracy versus the number of antenna elements. Intuitively, the accuracy is high with more numbers of antenna elements and especially when the number of signals is less than the number of elements. However, the computational expenditure and hardware implementation complexity are directly proportional to the number of antenna elements. As a result, it is necessary and desirable to use the number of elements just sufficient to obtain a reliable DOA estimate.

- *Variation of power levels:* This simulation is of importance due to the fact that signal sources are located at various distances from the receiving antenna array. Hence, the signals from different sources arrive at an array at different power levels. This experiment studies how the power level differences can affect the performance of a DOA estimation algorithm, in particular, in the presence of the weak signals and strong signals.

- *Angle of separation:* This simulation directly measures the resolution of a DOA estimation algorithm. The resolution of an algorithm is counted on its capability of resolving two closely spaced signals. This experiment studies variations in error of DOA estimates when the impinging signals are closely placed.

- *Computational complexity:* This simulation attempts to give an idea about the implementation of a DOA finding algorithm in terms of the FLOPS algorithm demands. The algorithms are programmed and the simulations are conducted on the MATLAB platform. Therefore, the MATLAB FLOP counter has been used for this study [1, 2].

In all the simulations, the performance of the algorithms in coherent scenarios is also studied. In these scenarios, spatial smoothing with forward-backward averaging is used to decorrelate the signals. The main objective is to show that the algorithms perform well in coherent environments provided the data is preprocessed. However, only simulations with respect to the key parameters such as snapshots and SNR are presented since the performance with other parameters is similar to that of uncorrelated signals.

In all the simulations to follow, RMS (root mean square) error is used as a metric to measure the accuracy of the DOA estimations. For a single trial of a DOA estimation, the RMS error of the estimated DOAs is defined as

$$RMS = \sqrt{\frac{1}{d}\left\{\hat{\theta}_i - \theta_i\right\}^2} \quad i = 1, 2, \ldots, d \tag{6.1}$$

where $\hat{\theta}_i$ is the DOA estimate obtained by a DOA estimation algorithm and θ_i is the actual angle of arrival.

In our study, the DOA estimations are run with several trials (say, K independent trials) in order to assess the overall errors of a DOA algorithm. As a result, an estimate of the RMS error should involve the results from the K independent trials; the corresponding RMS error is then redefined as

$$RMSE = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\left\{\hat{\theta}_{ik} - \theta_i\right\}^2} \tag{6.2}$$

where $\hat{\theta}_{ik}$ is the DOA estimate obtained by a DOA estimation algorithm at the kth trial based on (6.1).

Finally all the simulations abide to the data model assumptions presented in Chapter 3.

6.2 Performance Analysis

6.2.1 Standard ESPRIT

In this section, we present performance analysis of standard ESPRIT in one dimension. Various simulations in both uncorrelated and coherent signal environments under variety of signal conditions are run. Fifty trials (i.e., $K = 50$) are carried out, with each trial involving 250 snapshots (i.e., $K = 250$). These simulation parameters are maintained throughout the experiments with the standard ESPRIT unless otherwise mentioned. Three uncorrelated signals of equal power are considered to be arriving from $\theta = 5°$, $\theta = 10°$, and $\theta = 20°$. Other parameters are given in the caption of each figure that shows the simulation results.

The performance of the standard ESPRIT with respect to number of snapshots taken to estimate the DOA is presented in Figures 6.1 and 6.2. The five-element ULA is employed. In the first simulation, the SNR is kept at 5 dB and the corresponding RMS errors are shown in Figure 6.1. In the second simulation, SNR is kept at 10 dB and the corresponding RMS errors are shown in Figure 6.2. As seen from Figures 6.1 and 6.2, the

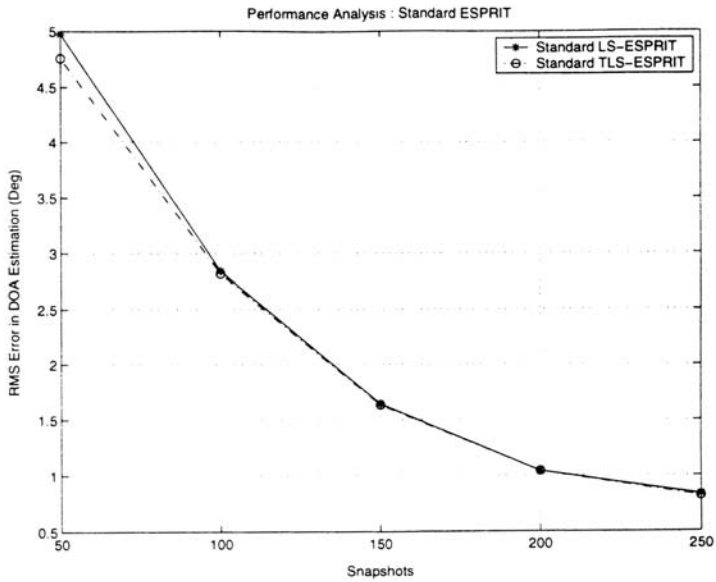


Figure 6.1 RMS error in DOA estimation as a function of the number of snapshots (SNR = 5 dB, five-element ULA, trials = 50, three uncorrelated signals at 5°, 10°, and 20°).

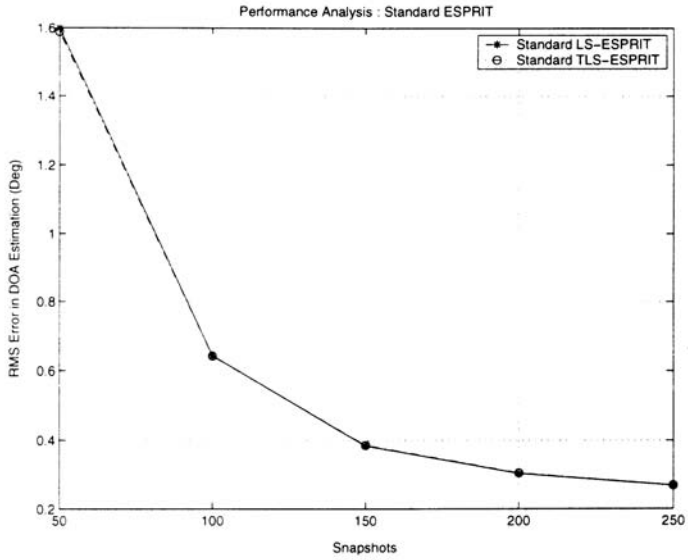


Figure 6.2 RMS error in DOA estimation as a function of the number of snapshots (SNR = 10 dB, five-element ULA, trials = 50, three uncorrelated signals at 5°, 10°, and 20°).

RMS errors are reduced with the increase of the number of the snapshots. The RMS error with 250 snapshots is five times smaller than that with 50 snapshots. In other words, DOA estimation errors are reduced with the increase of the number of snapshots. Also, the total-least-square (TLS) version gives better results than the simple standard least square (LS) version when the number of snapshots is less than 100.

Figures 6.3 shows the performance with respect to SNR varied from −5 dB to 5 dB and with the number of snapshots equal to 250. Figure 6.4 shows the performance with respect to the same SNR variation but with the number of snapshots equal to 500. Again, a five-element ULA is employed. In both cases, the errors reduce with the increase of the SNR. With 250 snapshots, when the SNR is less than −3 dB, both the LS and TLS algorithms have an error of more than 6°. However, the errors reduce to about 0.5° when the SNR is larger than 3 dB. With 500 snapshots, when the SNR is less than −3 dB, both the LS and TLS algorithms have an error of slightly less than 3°. However, the errors reduce to about 0.5° when the SNR is larger than 3 dB. In other words, if SNR increases, the errors reduce.

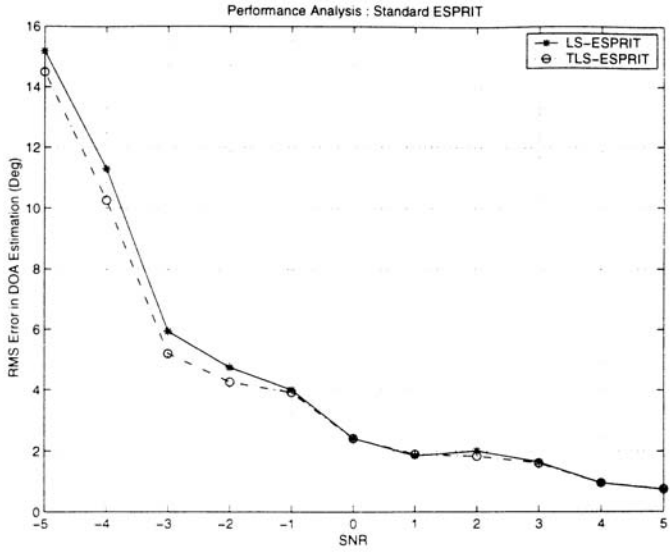


Figure 6.3 RMS error in DOA estimation as a function of the SNR (snapshots = 250, five-element ULA, trials = 50, three uncorrelated signals at 5°, 10°, and 20°).

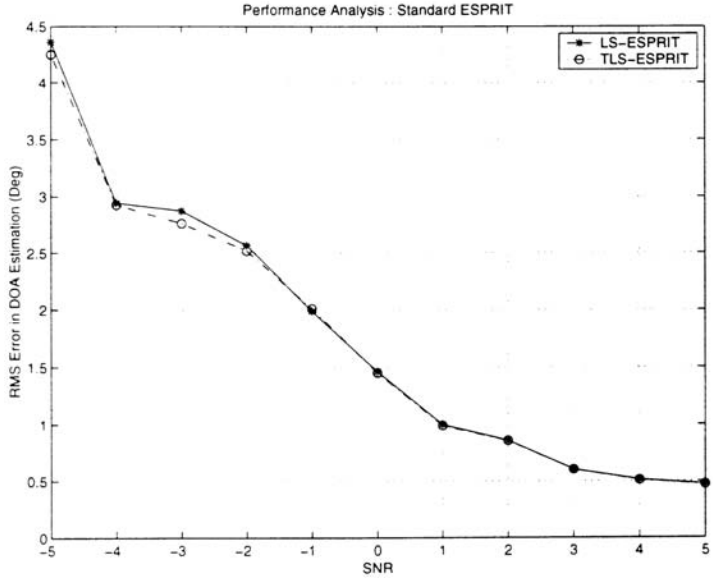


Figure 6.4 RMS error in DOA estimation as a function of the SNR (snapshots = 500, five-element ULA, trials = 50, three uncorrelated signals at 5°, 10°, and 20°).

Figure 6.5 shows the errors of the standard ESPRIT versus number of signals with an SNR of 5 dB, while Figure 6.6 shows the errors of the standard ESPRIT versus number of signals with an SNR of 10 dB. A 10-element antenna array is employed and nine signals are impinging on the array. The two simulations are conducted taking 500 snapshots per trial and an SNR of 5 dB and 10 dB are assumed, respectively. It is observed that the standard ESPRIT can handle about seven signals very well with an error of less than 0.2° in both SNR = 5 dB and SNR = 10 dB cases. For the estimation of 8 signals, the standard ESPRIT present an RMS error of 2° under SNR = 5 dB and an error of 0.5° under SNR = 10 dB. For the estimation of nine signals, the errors increase dramatically: the standard ESPRIT presents an RMS error of 12° under SNR = 5 dB and an error of 4.5° under SNR = 10 dB. This is an indication that in a practical situation, the number of signals that the standard ESPRIT can handle is the number of array elements less than 3 for high accurate DOA estimations, rather than less than 1.

Figure 6.7 shows the estimation error versus the number of elements in an array under SNR = 5 dB, while Figure 6.8 shows the estimation error versus the number of elements in an array under SNR = 10 dB.

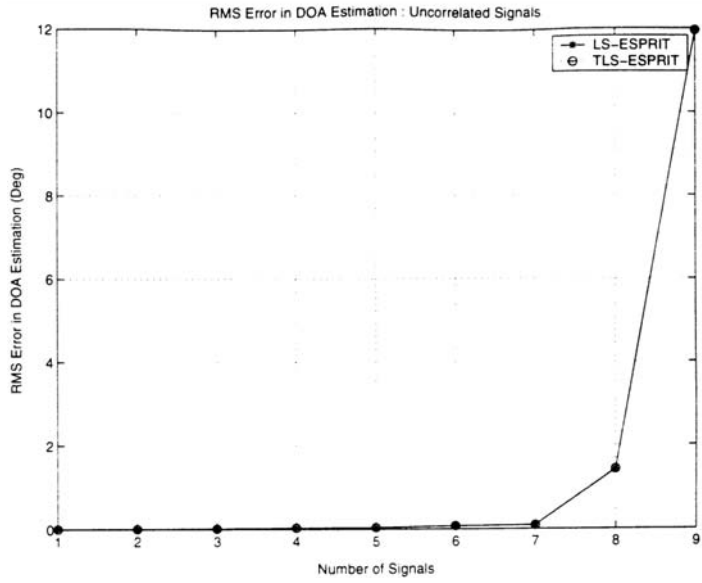


Figure 6.5 RMS error in a DOA estimation as a function of the number of signals (SNR = 5 dB, snapshots = 500, 10-element ULA, trials = 50).

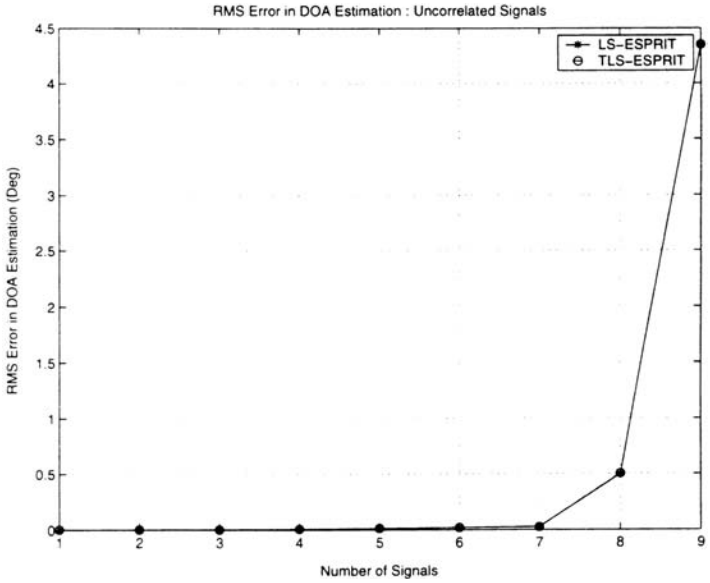


Figure 6.6 RMS error in a DOA estimation as a function of the number of signals (SNR = 10 dB, snapshots = 500, 10-element ULA, trials = 50).

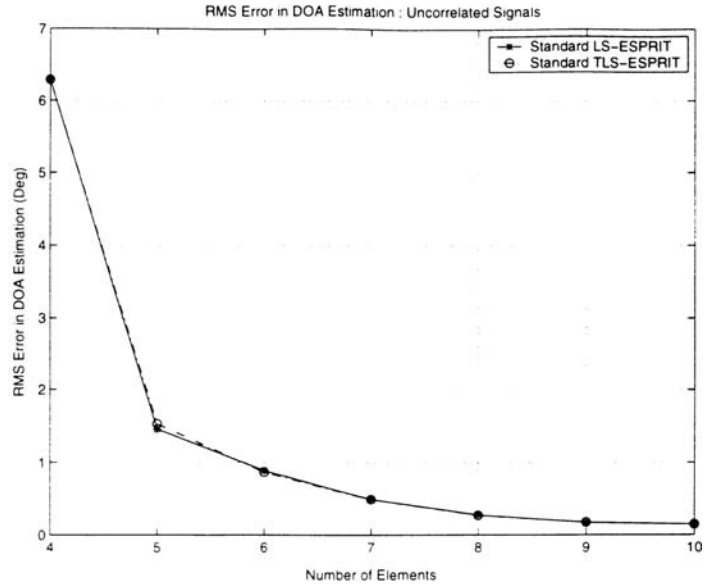


Figure 6.7 RMS error in a DOA estimation as a function of the number of elements (SNR = 5 dB, snapshots = 250, three uncorrelated signals at 5°, 10°, and 20°).

Again, the three signals are assumed to be impinging on the array at 5°, 10°, and 20°. The number of antenna elements is varied from 4 to 10 and two SNR scenarios, 5 dB and 10 dB, are assumed with 250 snapshots. It is observed that under SNR = 5 dB, a minimum of five elements is required to get the errors of less than 1.5° and seven elements to get an error of 0.6°. However, when the SNR is 10 dB, five elements are sufficient to have an error of less than 0.6° and seven elements to get an error of 0.2°. In short, when the number of elements increases, the RMS error is decreased.

Figure 6.9 shows the estimation error versus power levels of the received signals with a 10-element antenna array under SNR = 5 dB, while Figure 6.10 shows the estimation error versus power levels of the received signals with a 10-element antenna array under SNR = 10 dB. Simulations were performed with the strong signal arriving at 5° and the weak signals arriving at 10° and 20° with the respective power levels of −20 dB to −2 dB less than the power of the strongest signal arriving at 5°. Simulations were conducted again in two different SNR scenarios, 5 dB and 10 dB. It can be seen from Figure 6.9 that when the SNR is 5 dB, the

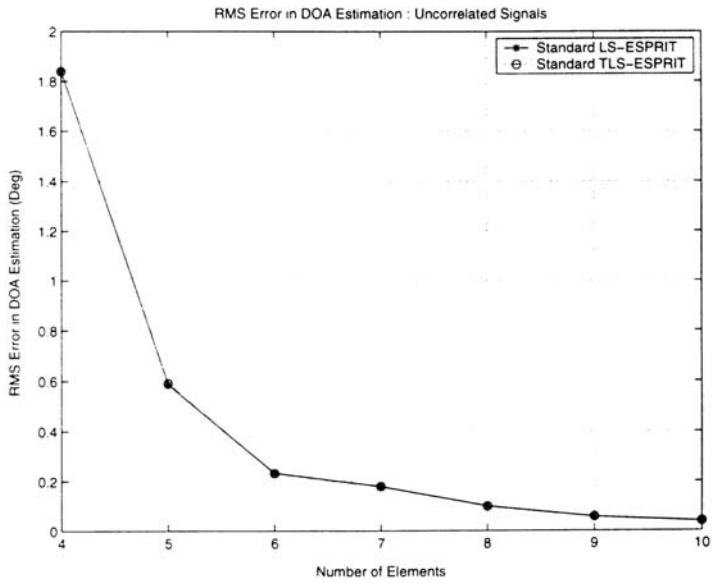


Figure 6.8 RMS error in a DOA estimation as a function of the number of elements (SNR = 10 dB, snapshots = 250, three uncorrelated signals at 5°, 10°, and 20°).

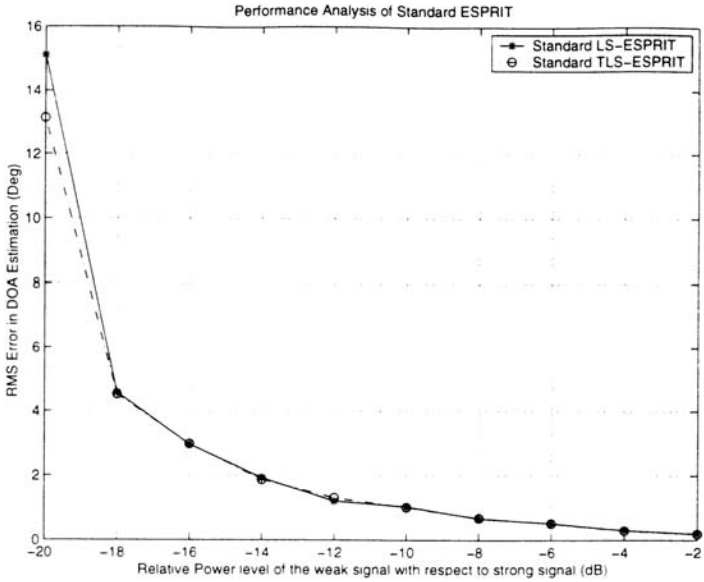


Figure 6.9 RMS error in a DOA estimation as a function of the power level (SNR = 5 dB, 10-element ULA, three uncorrelated signals at 5°, 10°, and 20°).

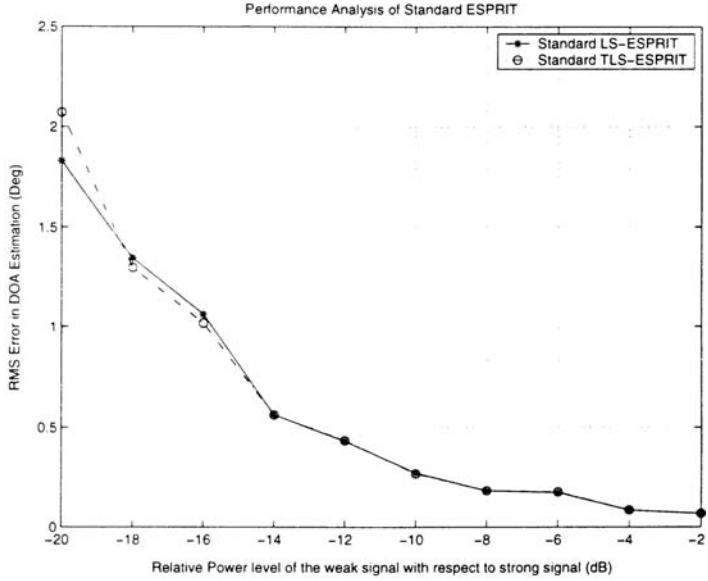


Figure 6.10 RMS error in a DOA estimation as a function of the power level (SNR = 10 dB, 10-element ULA, three uncorrelated signals at 5°, 10°, and 20°).

weak signal has to be higher than −14 dB of the strong signal in order to achieve less than 2° in error. However, when the SNR is 10 dB, the weak signals need only to be higher than −20 dB of the strong signal (see Figure 6.10). In general, the more even the signal power level received by array elements, the less DOA estimation errors there are.

Figures 6.11, 6.12, and 6.13 show the resolution capability of the standard ESPRIT. Here three signals are impinging on a 10-element ULA with the angular separation among them varied from 2° to 20°. One signal is made to come from −40° and other signals are made to arrive from $-40 + \delta$, $-40 + 2\delta$, where δ is varied from 2° to 20°. Two hundred and fifty snapshots are taken in two scenarios of SNR = 5 dB and 10 dB, respectively. Figure 6.11 shows the results with SNR = 5 dB, and Figure 6.12 shows the results with SNR = 10 dB. When SNR = 5 dB, for an error of less than 1.5°, the angle of separation should be larger than 4°; for an error of less than 0.4°, the angle of separation should be larger than 4°. However, when SNR = 10 dB, for the angle of separation larger than 2°, the error is less than 0.75°; for the angle of separation larger than 4°, the

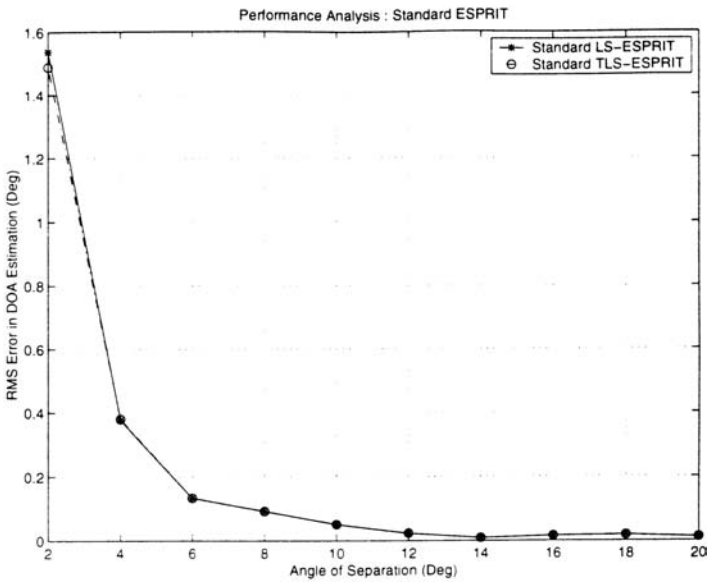


Figure 6.11 RMS error in a DOA estimation as a function of the separation of signals (SNR = 5 dB, snapshots = 250, trials = 50, 10-element ULA).

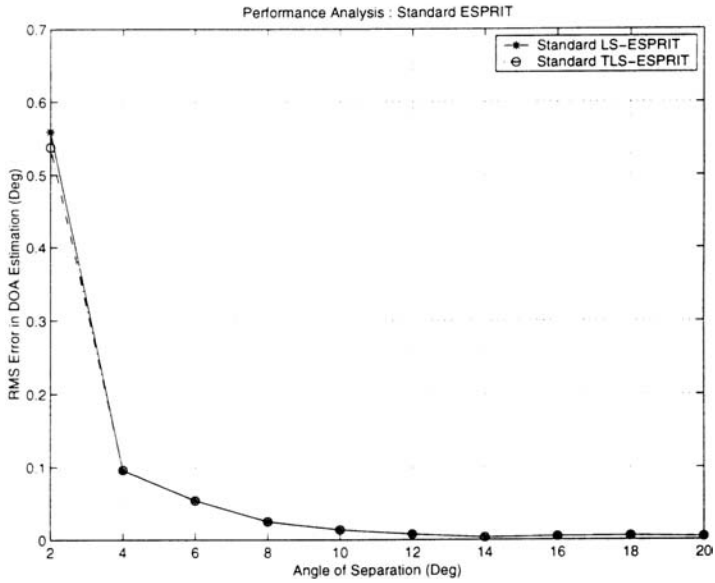


Figure 6.12 RMS error in a DOA estimation as a function of the separation of signals (SNR = 10 dB, snapshots = 250, 10-element ULA, trials = 50).

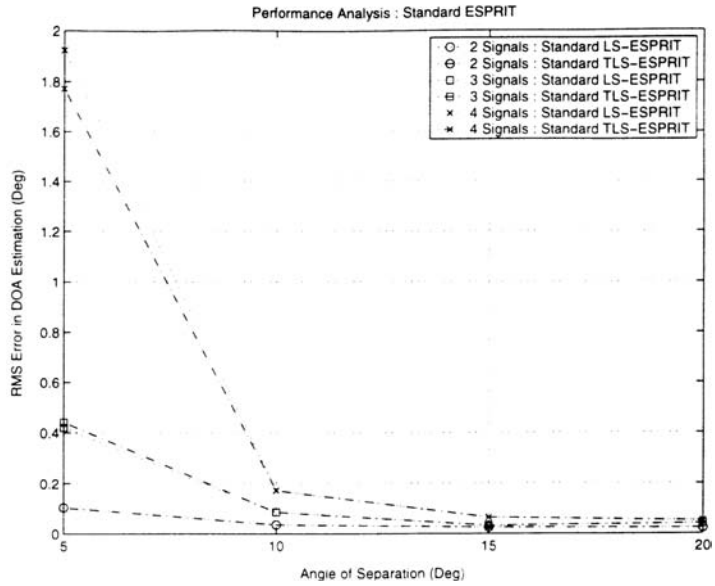


Figure 6.13 RMS error in a DOA estimation as a function of the separation of signals (SNR = 5 dB, snapshots = 250, 10-element ULA).

error is less than 0.1°. In general, when the angle of separation of the signals is larger, the error is smaller.

Another simulation is conducted where the number of signals is increased from 2 to 4, one at a time. Figure 6.13 shows the results versus the angle of separation with the increase of the number of elements. As seen, the larger number of the signals, the larger angle of separation required for a specified error. For instance, to have an error of less than 0.1°, the angle of separation needs to be larger than 5° in the case of two signals, 10° in the case of three signals, and 15° in the case of four signals.

Figure 6.14 shows FLOPS used in the MATLAB platform. It is observed that the TLS version demands more FLOPS as the number of elements increase, while the LS version does slightly. The needed FLOPS with the TLS are almost exponentially proportional to the number of the elements. Therefore, cautions need to be taken when the TLS version is used and the number of the array elements is to increase.

The above simulations are obtained with the uncorrelated signals. In the following, performance of the standard ESPRIT was evaluated in a

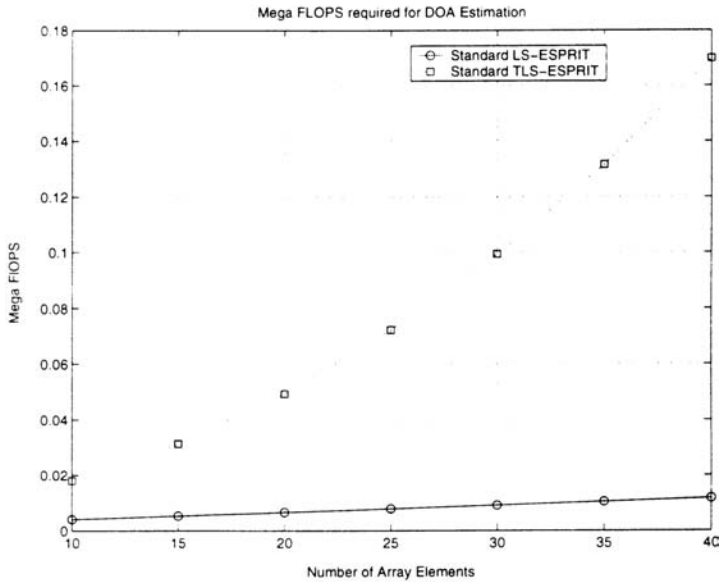


Figure 6.14 FLOPS as a function of the number of antenna array elements (SNR = 5 dB, snapshots = 250, 50 trials).

coherent environment. A six-element ULA is considered and is divided into three subarrays, each of them having four elements. Then two coherent signals are impinged on the array from 5° and 10°.

Figure 6.15 shows the estimation error as a function of snapshots with an SNR of 5 dB, while Figure 6.16 shows the estimation error as a function of snapshots with SNR of 10 dB. As can be seen, the error with the number of snapshots being 250 is more than five times smaller than that when the number of snapshots is 50, an indication of strong benefits with increase of the number of the snapshots.

Figure 6.17 shows the error versus SNR with 250 snapshots, while Figure 6.18 shows the error versus SNR with 500 snapshots. As seen from the figures, the DOA errors reduce as SNR increases as well as the number of snapshots. For instance, when SNR is increased from −5 dB to +5 dB, the errors are reduced from 3.7° to 0.3° with 250 snapshots and 1.4° to 2° with 500 snapshots. In addition, in a low SNR environment, the larger number of snapshots improves the DOA estimation accuracy more than that in a high SNR environment.

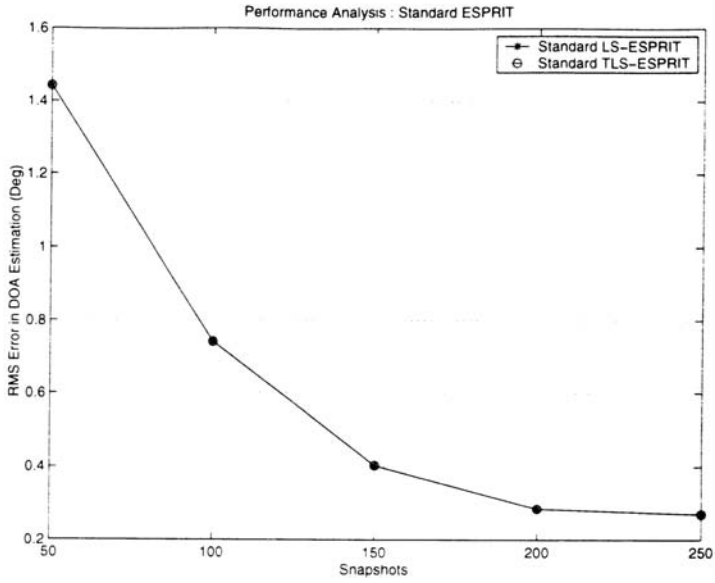


Figure 6.15 RMS error in a DOA estimation as a function of the number of snapshots (SNR = 0 dB, six-element ULA, three subarrays, two coherent signals impinge at 5° and 10°).

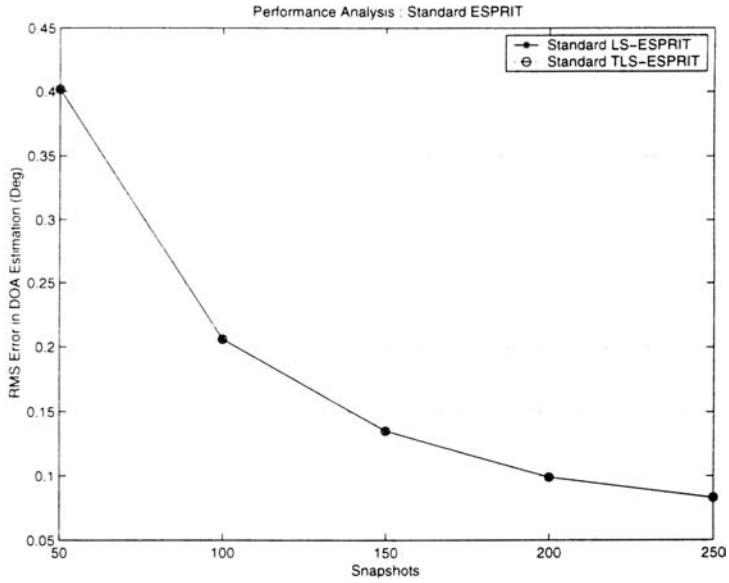


Figure 6.16 RMS error in a DOA estimation as a function of the number of snapshots (SNR = 10 dB, six-element ULA, three subarrays, two coherent signals impinge at 5° and 10°).

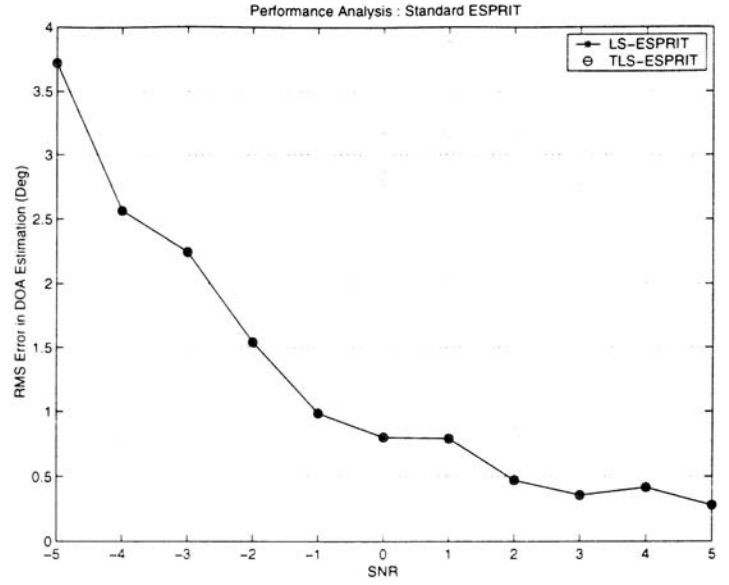


Figure 6.17 RMS error in a DOA estimation as a function of the SNR (snapshots = 250, six-element ULA, three subarrays, two coherent signals impinge at 5° and 10°).

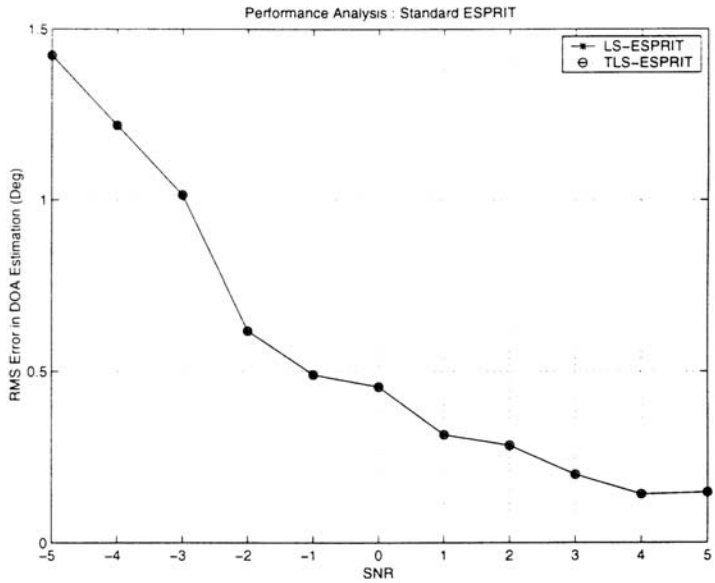


Figure 6.18 RMS error in a DOA estimation as a function of the SNR (snapshots = 500, six-element ULA, three subarrays, two coherent signals impinge at 5° and 10°).

6.2.2 The One-Dimensional Unitary ESPRIT

In this section, the performance of the unitary ESPRIT in one dimension is analyzed. Various simulation results run in both uncorrelated and coherent signal environments under a variety of signal conditions are presented. Simulations were conducted with a 10-element ULA with maximum overlap. Fifty trials were carried out, with each trial involving 250 snapshots. These simulation parameters are maintained throughout the experiments with the unitary one-dimensional ESPRIT unless otherwise mentioned. Four uncorrelated signals of equal power are considered to be arriving at $\theta_1 = -10°$, $\theta_2 = -5°$, $\theta_3 = 5°$, and $\theta_4 = 10°$.

Figure 6.19 shows the results of the unitary one-dimensional ESPRIT simulation with respect to number of snapshots taken to estimate the DOA with SNR = 0 dB, while Figure 6.20 shows the results of the unitary one-dimensional ESPRIT simulation with respect to number of snapshots taken to estimate the DOA with SNR = 10 dB. The results show an error of about 4° or less at SNR of 0 dB and a smaller error of about 0.17° or less when the SNR is increased to 10 dB. The DOA estimation accuracy improves with the increase of the number of the

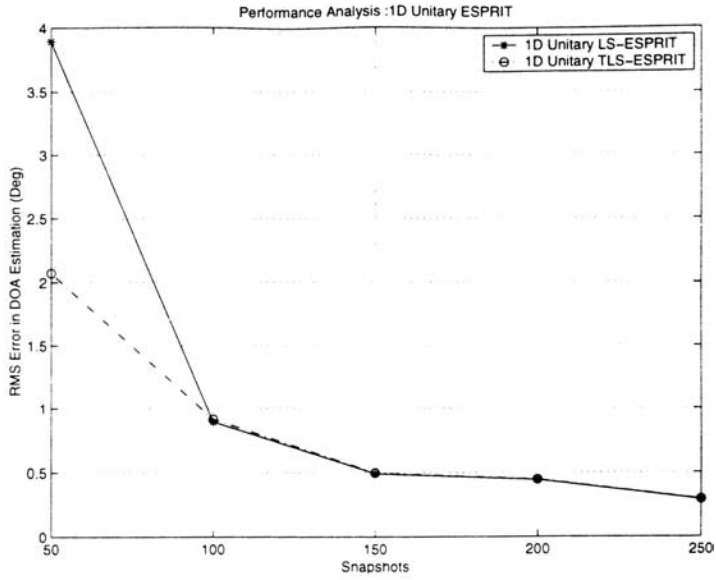


Figure 6.19 RMS error in a DOA estimation as a function of the snapshots (SNR = 0 dB, 10-element ULA, trials = 50, four uncorrelated signals at $\theta_1 = -10°$, $\theta_2 = -5°$, $\theta_3 = 5°$, and $\theta_4 = 10°$).

snapshots taken. The improvement is more than five times in the cases simulated when the number of the snapshots is increased from 50 to 250.

Figures 6.21 and 6.22 show the RMS errors with respect to different SNR values. SNR is varied from −5 dB to 5 dB. Figure 6.21 is obtained with 250 snapshots taken, while Figure 6.22 is obtained with 500 snapshots taken. As can be seen, when SNR is improved from −5 dB to 5 dB, the DOA estimation accuracy is improved more than 10 times in both figures. It is also observed that the unitary TLS-ESPRIT is performing better than the unitary LS-ESPRIT, giving a 1° error at SNR= −5 dB compared to a 4° error given by the LS version. However, when the SNR is above −4 dB or the numbers of snapshots are more than 500, both the TLS- and LS-ESPRIT perform almost the same.

Figure 6.23 shows the RMS error versus number of the signals with SNR = 0 dB, while Figure 6.24 shows the RMS error versus number of the signals with SNR = 10 dB. A 10-element antenna array is employed and nine signals impinge on the array. Five hundred snapshots per trial are taken. It is observed that when the SNR is 0 dB, the LS version can

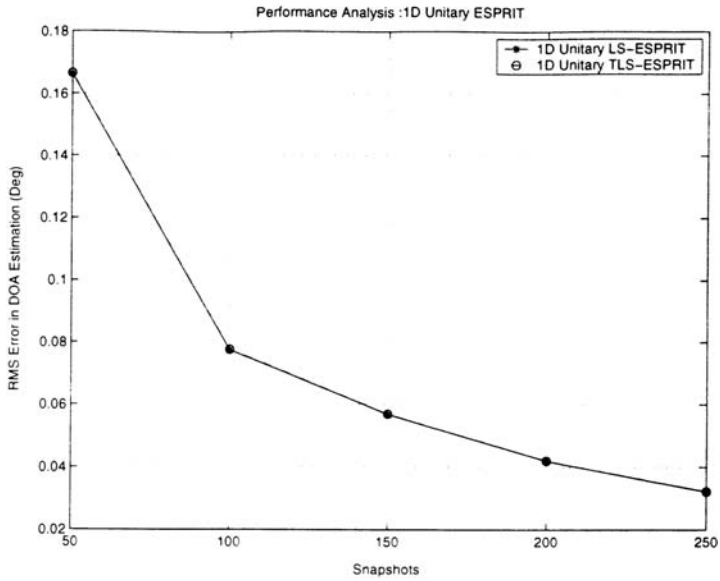


Figure 6.20 RMS error in a DOA estimation as a function of the snapshots (SNR = 10 dB, 10-element ULA, trials = 50, four uncorrelated signals at $\theta_1 = -10°$, $\theta_2 = -5°$, $\theta_3 = 5°$, and $\theta_4 = 10°$).

estimate DOAs up to seven signals under a 10° accuracy, whereas the TLS version is able to estimate up to eight signals. When the SNR is 10 dB, the LS version is estimating eight signals, while the TLS version is estimating DOAs of all nine signals to a 10° accuracy. In general, the number of signals that the unitary ESPRIT algorithms can estimate is lower than the ideal case and the number of the array elements less than one, unless in a high SNR environment.

Figures 6.25 and 6.26 present the performance of the unitary one-dimensional ESPRIT with respect to the number of elements in the array. The four signals are impinging on the array and a minimum of five elements are theoretically needed to resolve them. The total number of antenna elements is varied from 5 to 10 and two SNRs are used, 0 dB and 10 dB. It is observed that with SNR = 0 dB, a minimum of six elements are needed to get an accuracy below 10°. When the SNR is increased to 10 dB, the error reduces to less than 1°. If the number of elements is less than 6, the algorithm is performing badly even at high SNRs. In short, in a practical noisy environment, the minimum number of the array elements required to achieve a good DOA estimation accuracy is normally

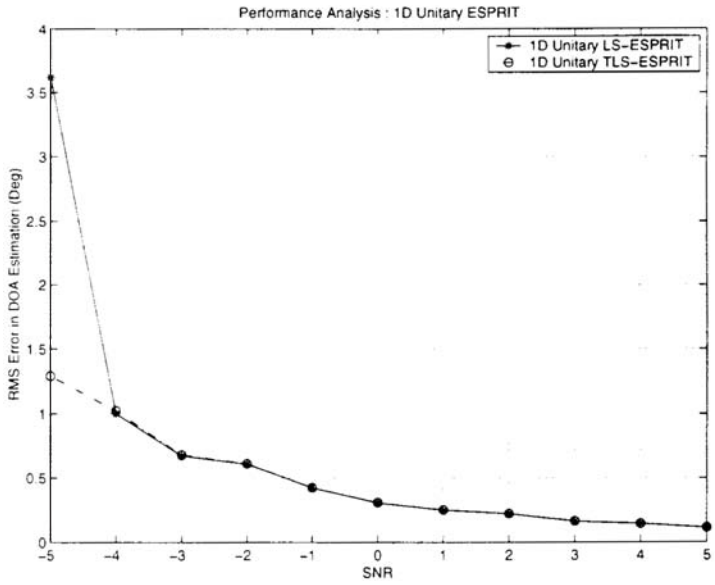


Figure 6.21 RMS error in a DOA estimation as a function of the SNR (snapshots = 250, 10-element ULA, trials = 50, four uncorrelated signals at $\theta_1 = -10°$, $\theta_2 = -5°$, $\theta_3 = 5°$, and $\theta_4 = 10°$).

higher than the theoretically predicted number, the number of the signals plus one.

Figures 6.27 and 6.28 present the performance of the unitary one-dimensional ESPRIT with respect to the power level of the received signal with a 10-element antenna array. The numbers of signals arriving at the array were taken to be 2, 3, and 4, respectively. In all cases, one signal is made to be strong, and the powers of the remaining signals are varied equally from −20 dB to −5 dB below the power of the strong signal. Simulations are conducted in two different SNR scenarios of 0 dB and 10 dB, respectively. The results indicate that the performance is bad when four signals are impinging with unequal power. When SNR = 0 dB, at a −20-dB relative power level, the error is 12° for four signals, 4° for three signals, and 2° for two signals. In general, with a lesser degree of inequality in the power levels received by the array elements, the DOA estimation accuracy improves. It is also observed that TLS version is performing better than the LS version; the performance of the TLS version is far better when SNR is 10 dB: the reduction in error at the −20-dB relative power level is about 90%.

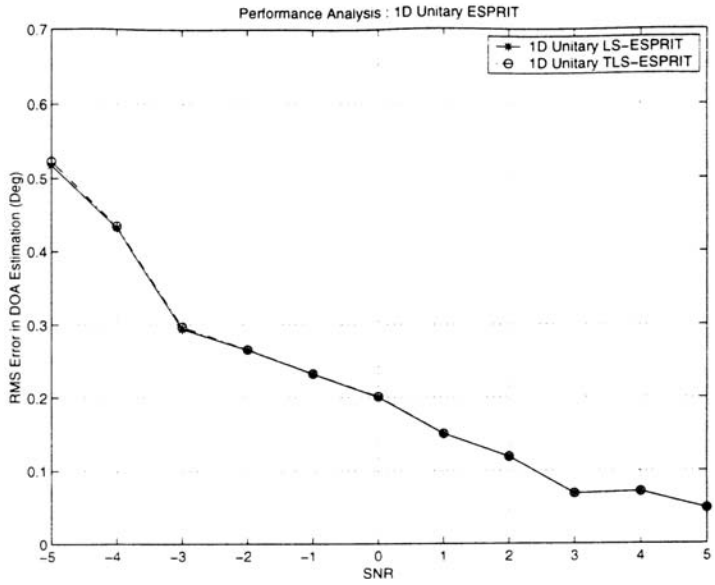


Figure 6.22 RMS error in a DOA estimation as a function of the SNR (snapshots = 500, 10-element ULA, trials = 50, four uncorrelated signals at $\theta_1 = -10°$, $\theta_2 = -5°$, $\theta_3 = 5°$, and $\theta_4 = 10°$).

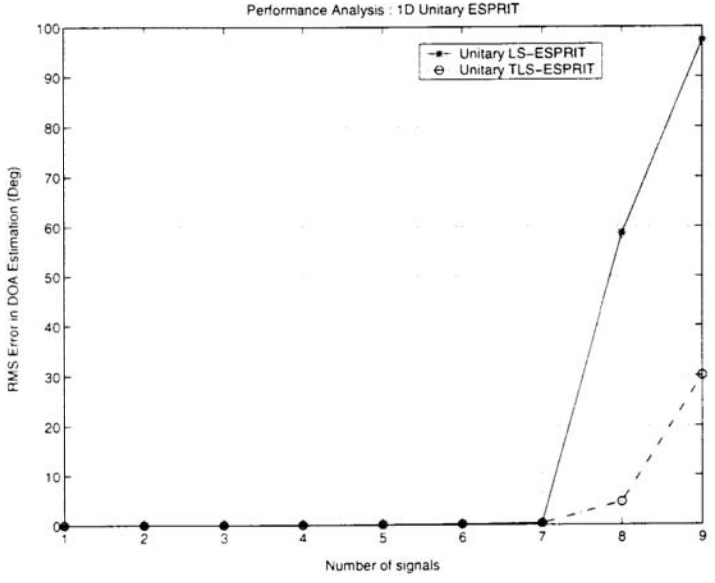


Figure 6.23 RMS error in a DOA estimation as a function of the number of signals (10-element ULA, trials = 50, snapshots = 250, SNR = 0 dB).

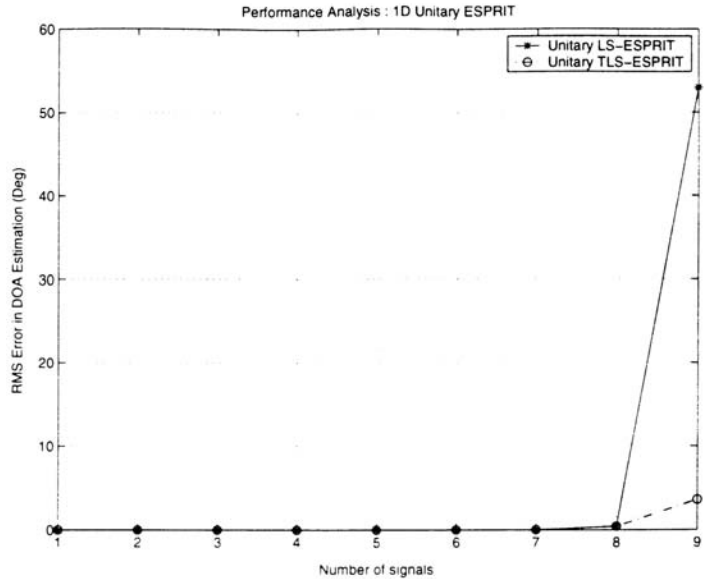


Figure 6.24 RMS error in a DOA estimation as a function of the number of signals (10-element ULA, trials = 50, snapshots = 250, SNR = 10 dB).

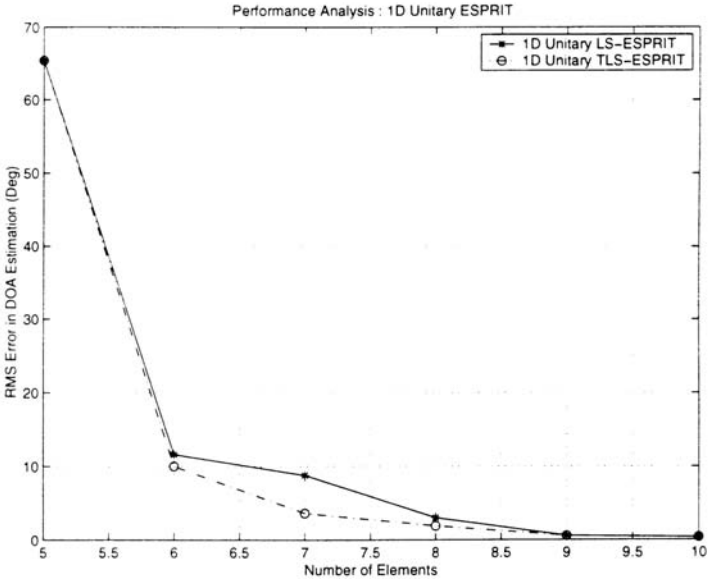


Figure 6.25 RMS error in a DOA estimation as a function of the number of elements (snapshots = 250, SNR = 0 dB, four uncorrelated signals at $\theta_1 = -10°$, $\theta_2 = -5°$, $\theta_3 = 5°$, and $\theta_4 = 10°$).

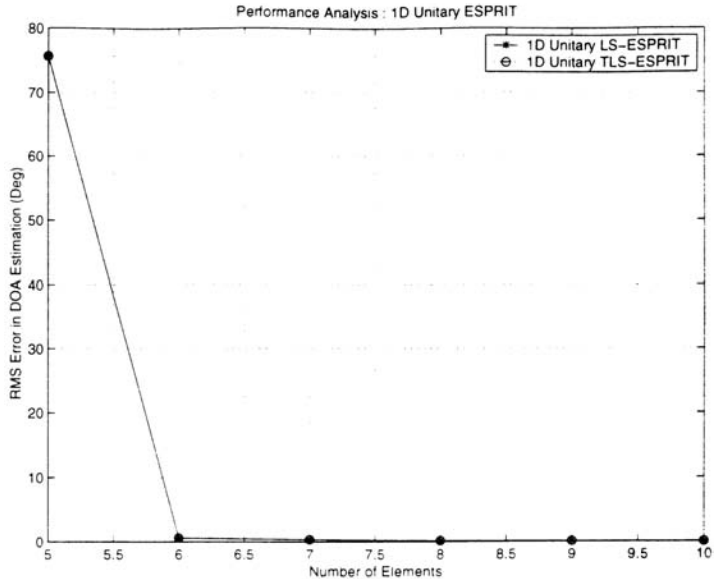


Figure 6.26 RMS error in a DOA estimation as a function of the number of elements (snapshots = 250, SNR = 10 dB, four uncorrelated signals $\theta_1 = -10°$, $\theta_2 = -5°$, $\theta_3 = 5°$, and $\theta_4 = 10°$).

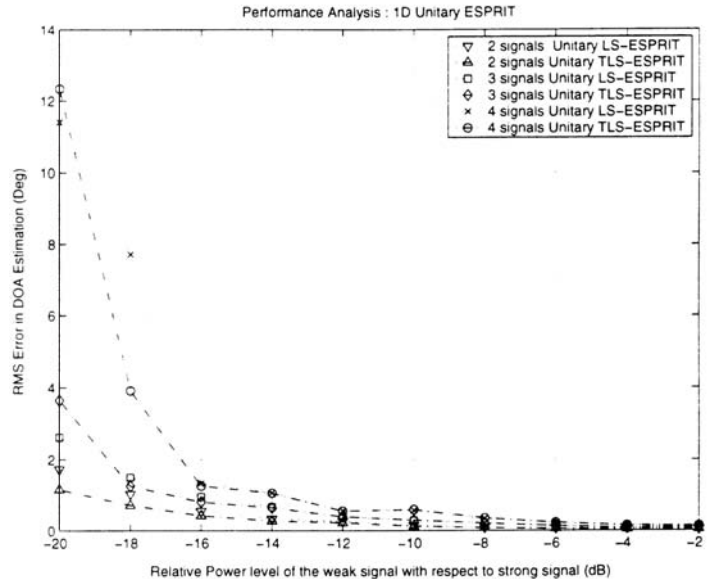


Figure 6.27 RMS error in a DOA estimation as a function of the relative power level (10-element ULA, uncorrelated signals, snapshots = 250, SNR = 0 dB).

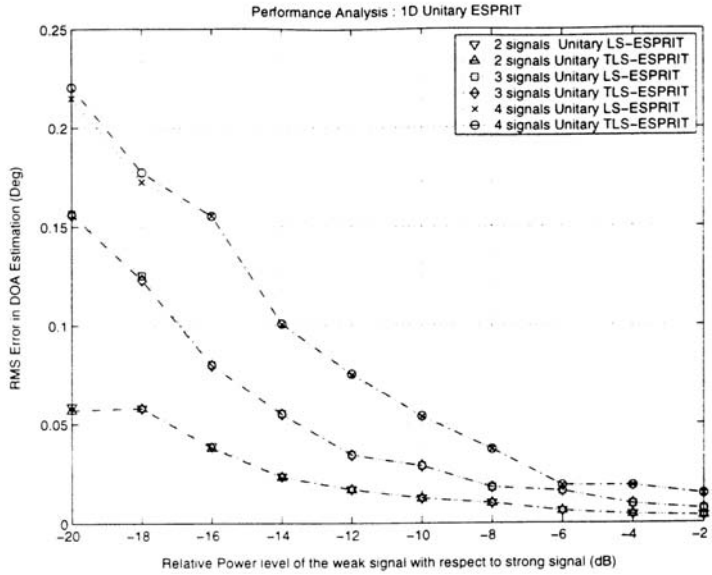


Figure 6.28 RMS error in a DOA estimation as a function of the relative power level (10-element ULA, uncorrelated signals, snapshots = 250, SNR = 10 dB).

Figures 6.29 and 6.30 show the resolution capability of the unitary one-dimensional ESPRIT. Initially, two signals were taken. One signal is arriving at 5° and other signal is made to come from $5° + \delta$, where δ is varied from 2° to 20°. There were 250 snapshots taken in two scenarios of SNR = 0 dB and 10 dB, respectively. As seen from Figures 6.29 and 6.30, the error in both scenarios is less than 0.45°, a small error mainly due to the use of 10 elements. It is observed that error in higher SNR scenario is four to five times less than the error at SNR = 0 dB.

Now the number of signals is increased to 2, 3, and 4 signals, respectively, on the 10-element ULA with the angular separation between them varied from 2° to 20°. One signal is made to come from 10° and other signals are made to arrive to come from $10° + \delta$, $10° + 2\delta$, and $10° + 3\delta$, where δ is varied from 2° to 20°. Figure 6.31 shows the simulation results. When SNR is 0 dB, as can be seen, the unitary ESPRIT is able to estimate DOAs of up to three signals placed as closely as 5° with an error less than 5°. If the separation is 10°, it is able to resolve all of the four signals to 2° accuracy.

The computational expenditure of the unitary ESPRIT is measured in terms of MATLAB FLOPS. Figure 6.32 shows the FLOPS usage. As

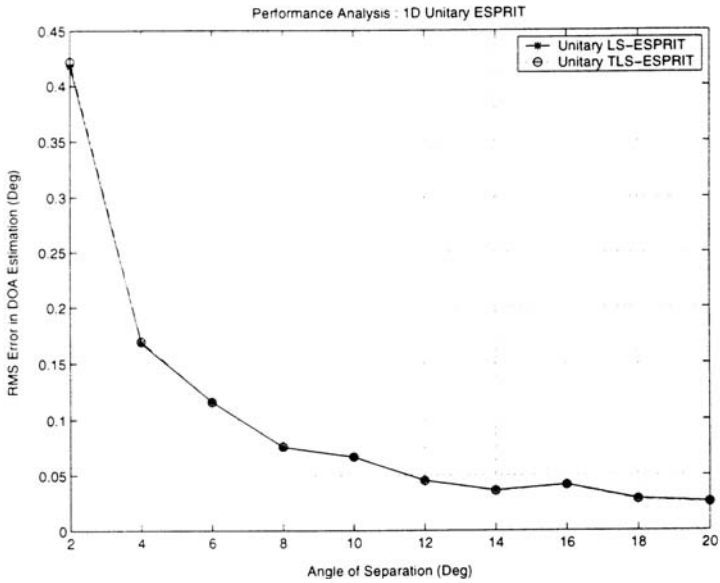


Figure 6.29 RMS error in DOA estimation as a function of separation (10-element ULA, trials = 50, snapshots = 250, SNR = 0 dB, two uncorrelated signals).

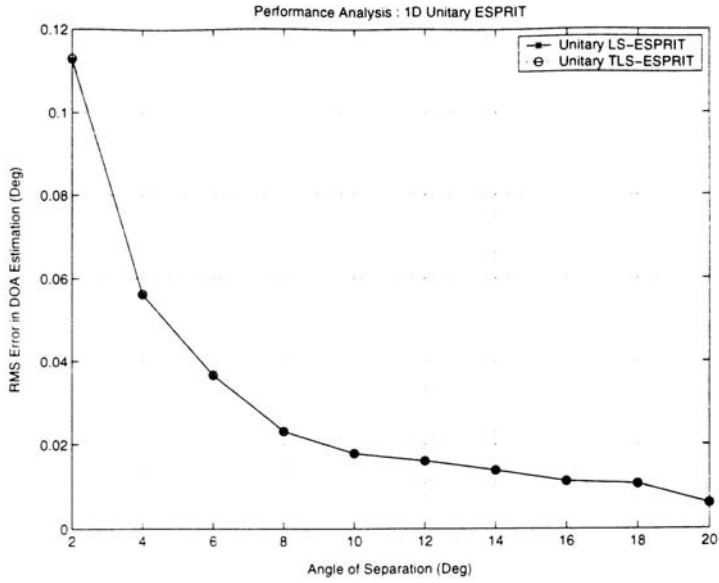


Figure 6.30 RMS error in DOA estimation as a function of separation (10-element ULA, trials = 50, snapshots = 250, SNR = 10 dB, two uncorrelated signals).

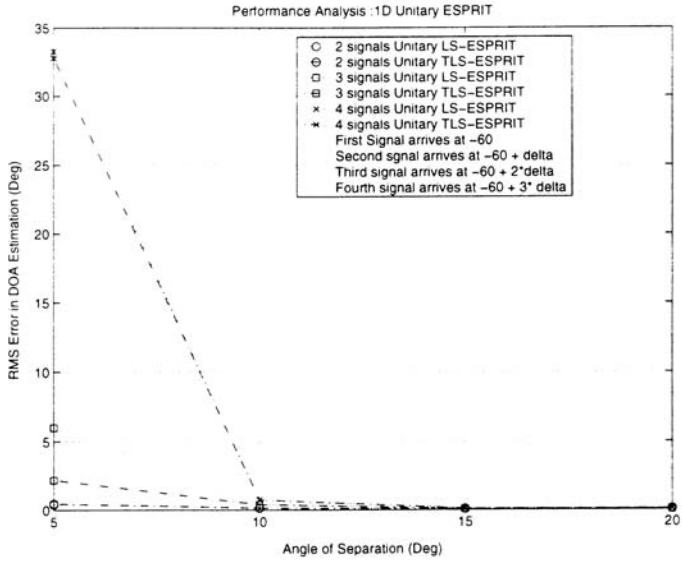


Figure 6.31 RMS error in a DOA estimation as a function of the separation (10-element ULA, trials = 50, snapshots = 250, SNR = 0 dB).

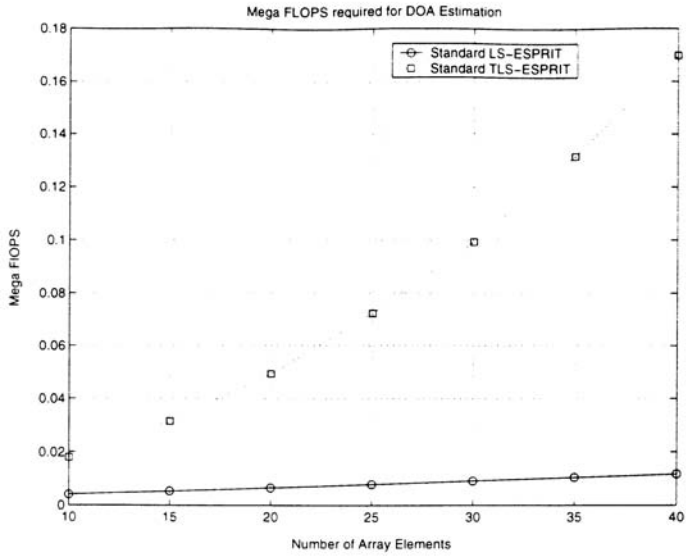


Figure 6.32 FLOPS as a function of number of antenna array elements (SNR = 5 dB, snapshots = 250, 10-element ULA).

expected, the unitary TLS version is demanding more FLOPS compared to the unitary LS.

The behavior of the algorithm in the coherent environment is studied next. A six-element linear array is considered. The array is divided into three subarrays, each of them having four elements. Two coherent signals are impinged on the array at −5° and 10°. The results are shown in Figures 6.33, 6.34, 6.35, and 6.36. In general, the performance in the coherent signal environment is worse than that in the uncorrelated environments; however, similar conclusions can be drawn on the relationship between the performance and various parameters.

6.2.3 The Two-Dimensional Unitary ESPRIT

In this section, the performance of the two-dimensional unitary ESPRIT is evaluated. Simulations are performed in both uncorrelated and coherent signal environments under variety of signal conditions. Simulations are conducted using a $4 \times 4 = 16$ element uniform rectangular array with maximum overlap in both the x and y directions. As usual, the RMS error in the DOA estimation is used as the metric for measuring the

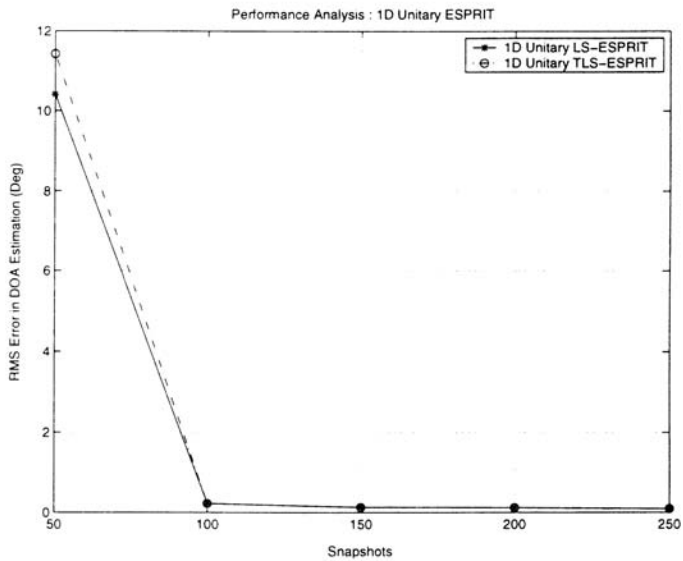


Figure 6.33 RMS error in a DOA estimation as a function of the number of snapshots (six-element ULA, trials = 50, SNR = 0 dB, two coherent signals at –5° and 10°).

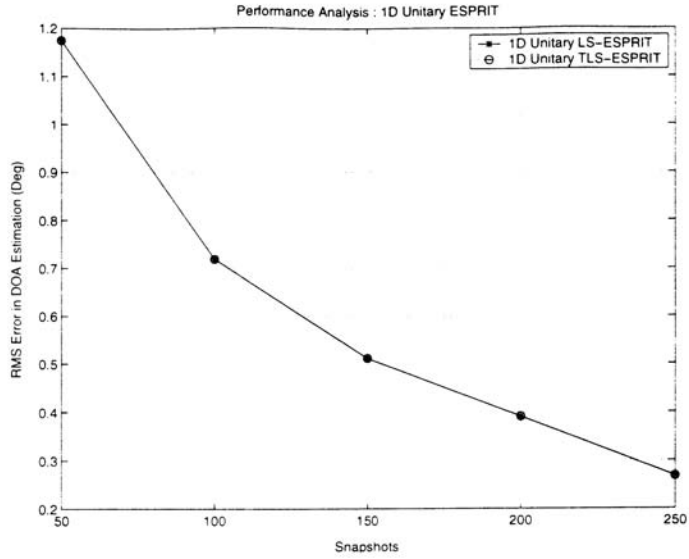


Figure 6.34 RMS error in a DOA estimation as a function of snapshots (6-element ULA, trials = 50, SNR = 10 dB, two coherent signals at –5° and 10°).

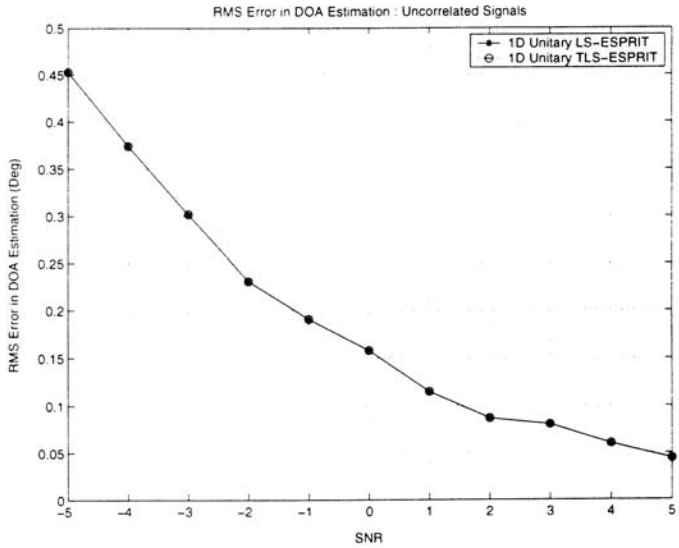


Figure 6.35 RMS error in a DOA estimation as a function of the SNR (six-element ULA, trials = 50, snapshots = 250, two coherent signals at –5° and 10°).

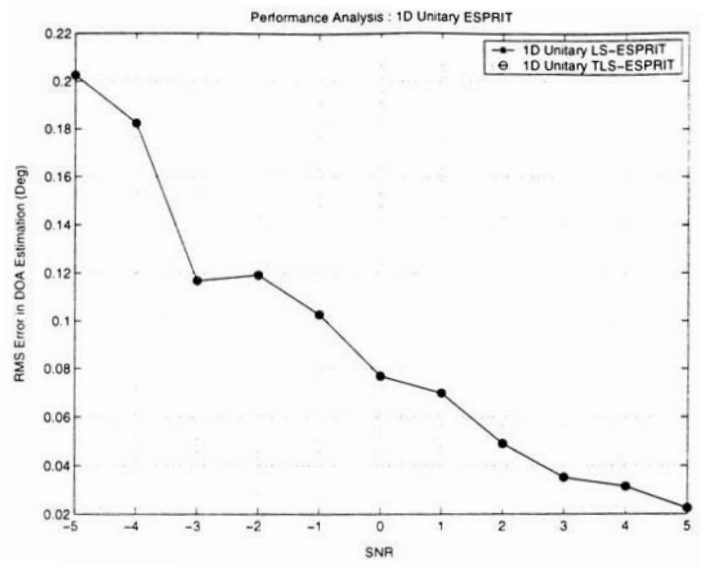


Figure 6.36 RMS error in a DOA estimation as a function of the SNR (six-element ULA, trials = 50, snapshots = 500, two coherent signals at –5° and 10°).

performance. Fifty trials were run. These simulation parameters are maintained throughout the experiments with the two-dimensional unitary ESPRIT unless otherwise mentioned. Two uncorrelated signals of equal power are chosen to be arriving from $(\theta_1, \phi_1) = (-20°, -15°)$ and $(\theta_2, \phi_2) = (20°, 15°)$.

In first simulation, the SNR is kept at 10 dB and 50 to 150 snapshots of data are taken. Figure 6.37 shows the DOA errors. It is observed that, like in the one-dimensional case, the RMS error reduces with the increased number of snapshots.

Next, performance with respect to the SNR is evaluated. The SNR is varied from −5 dB to 5 dB. In the first simulation, 250 snapshots of data are taken, and in the second simulation, 500 snapshots of the data are taken with the same variation in the SNR; the results are shown in Figures 6.38 and 6.39, respectively. As can be seen, the DOA estimation errors reduce with the increase of SNR. With 250 snapshots, the error is reduced from 12° to 0.1° when the SNR is increased from −5 dB to 5 dB; with 500 snapshots, the RMS error decreases from 9° to 0.1° when SNR is increased from −5 dB to 5 dB. It can also be noticed that to achieve an error of less than 2°, the SNR should be at minimum −1 dB.

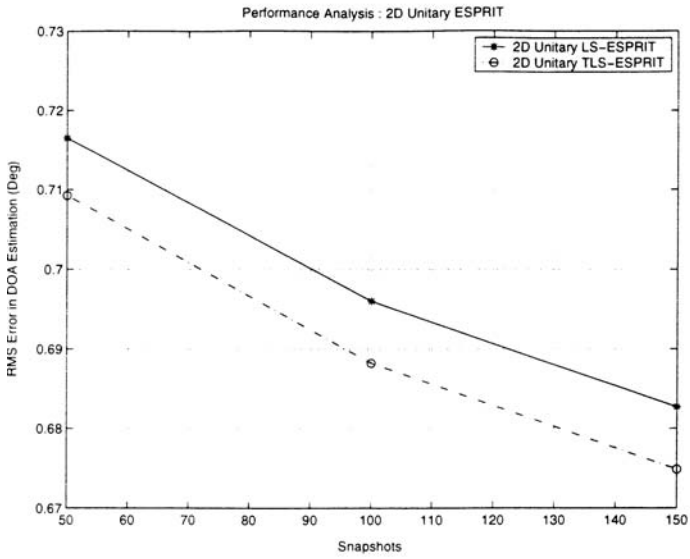


Figure 6.37 RMS error in a DOA estimation as a function of snapshots (SNR = 10 dB, 16-element URA, trials = 50, two uncorrelated signals at $(\theta_1, \phi_1) = (-20°, -15°)$ and $(\theta_2, \phi_2) = (20°, 15°)$).

The performance of the two-dimensional unitary ESPRIT is further studied to check how many signals the algorithm can handle. Eight signals are made to impinge on the array. Two simulations are conducted, taking 250 snapshots per trial with an SNR of 5 dB and 10 dB, respectively. Figures 6.40 and 6.41 are the simulation results. It is observed that, with the given SNR value, the algorithms are able to estimate about seven signals under an error of 5° above which the accuracy degrades rapidly. In other words, the number of the signals that can be estimated is less than the ideally predicted limit, which is the number of the array elements less than one. Also, the LS version shows a better performance than TLS while TLS is computationally more intensive.

The performance of the two-dimensional unitary ESPRIT with respect to the number of elements of the array is shown in Figures 6.42 and 6.43. It can be seen that with the increase of the number of the array elements, the error reduces. For instance, when the number of the elements is increased from 4 to 25, the error is reduced from about 5.4° to 3.7° under SNR = 5 dB and from about 0.125° to 0.045° under SNR = 10 dB. It is also observed that the TLS version is performing slightly better than the LS version, especially for low SNR values.

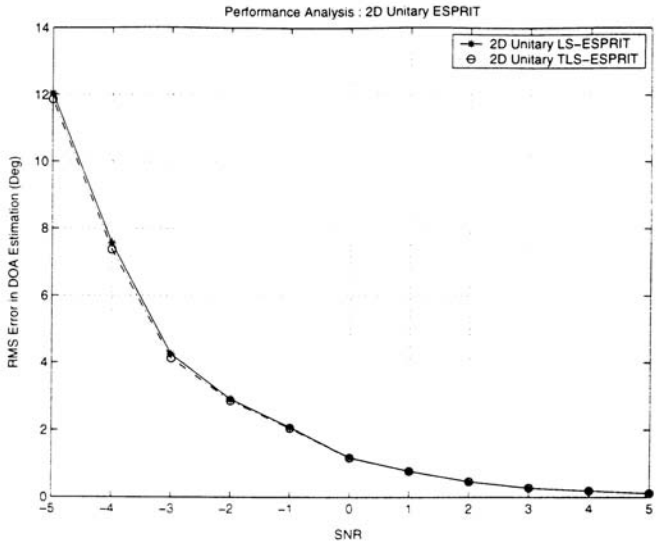


Figure 6.38 RMS error in a DOA estimation as a function of the SNR (snapshots = 250, 16-element ULA, trials = 50, two uncorrelated signals at (θ_1, ϕ_1) = (–20°, –15°) and (θ_2, ϕ_2) = (20°, 15°)).

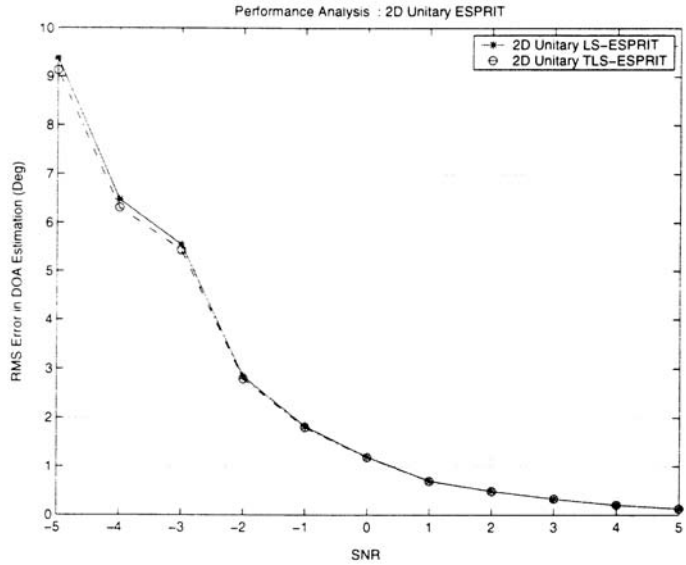


Figure 6.39 RMS error in a DOA estimation as a function of the SNR (snapshots = 500, 16-element URA, trials = 50, two uncorrelated signals at (θ_1, ϕ_1) = (–20°, –15°) and (θ_2, ϕ_2) = (20°, 15°)).

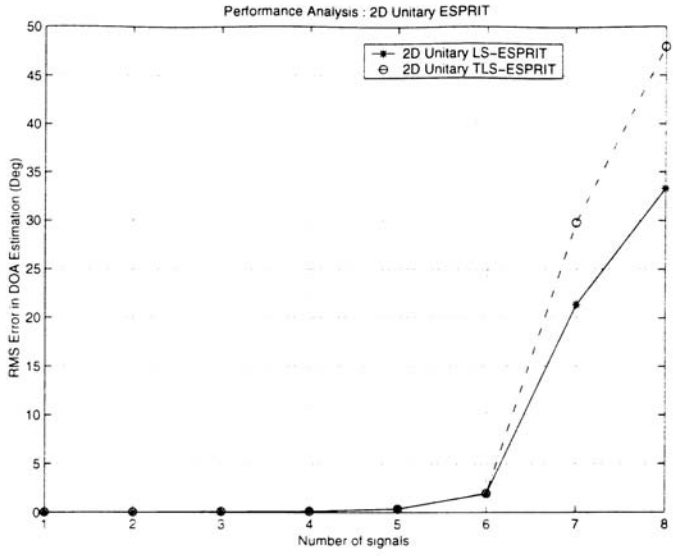


Figure 6.40 RMS error in a DOA estimation as a function of the number of signals (SNR = 5 dB, 16-element URA, trials = 50, snapshots = 250).

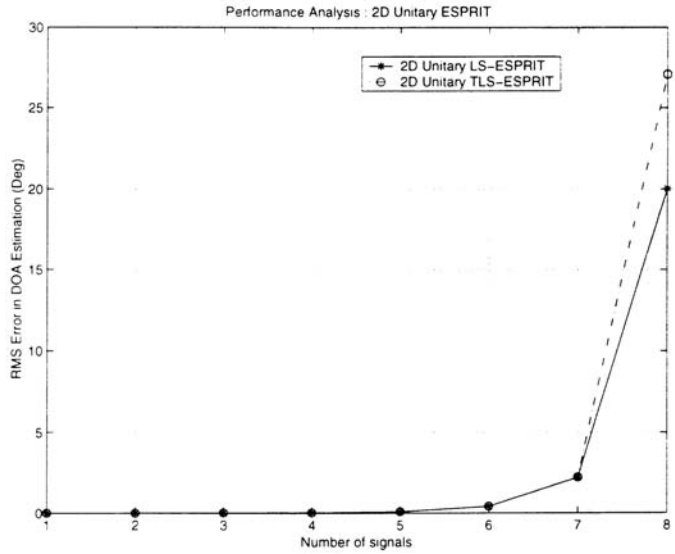


Figure 6.41 RMS error in a DOA estimation as a function of the number of signals (SNR = 10 dB, 16-element URA, trials = 50, snapshots = 250).

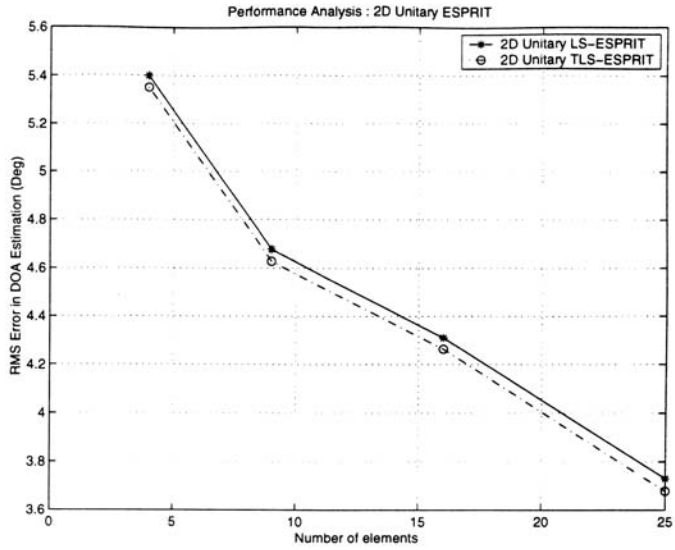


Figure 6.42 RMS error in a DOA estimation as a function of the number of elements (SNR = 5 dB, snapshots = 250, two uncorrelated signals at $(\theta_1, \phi_1) = (-20°, -15°)$ and $(\theta_2, \phi_2) = (20°, 15°)$).

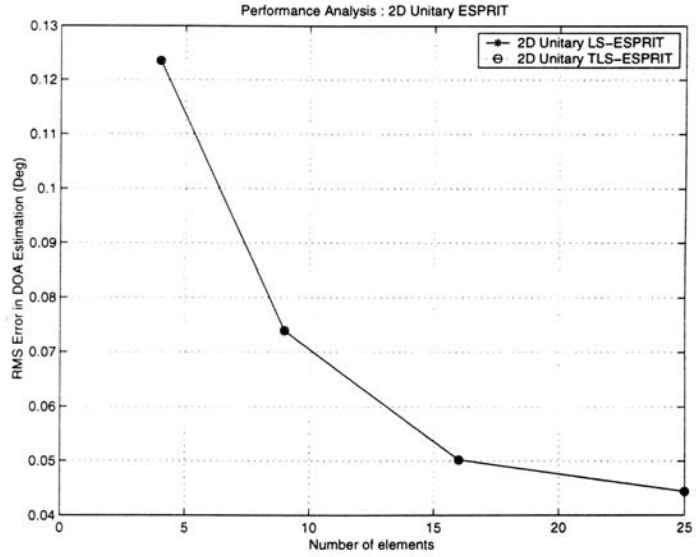


Figure 6.43 RMS error in a DOA estimation as a function of the number of elements (SNR = 10 dB, snapshots = 250, two uncorrelated signals at $(\theta_1, \phi_1) = (-20°, -15°)$ and $(\theta_2, \phi_2) = (20°, 15°)$).

The performance of the two-dimensional unitary ESPRIT with respect to power levels of the received signals is also studied and the results are shown in Figures 6.44 and 6.45. Simulations are performed with one strong signal arriving at $(\theta_1, \phi_1) = (-20°, -15°)$ and another weaker signal arriving at $(\theta_2, \phi_2) = (20°, 15°)$ with a power level of −20 dB to −5 dB less than the power of the strong signal. Simulations are conducted in two different SNR scenarios of 5 dB and 10 dB, respectively. As can be seen, when the difference of the power levels is reduced, the DOA estimation error is reduced. For instance, under SNR = 5 dB, the error is about 20° when the weak signal has a power, which is −20 dB below that of the strong signal; the error reduces to less than 0.5° when the power is −2 dB below.

The resolution capability of the two-dimensional unitary ESPRIT is shown in Figures 6.46 and 6.47. Here two signals are impinged with the angular separation between them varied from 2° to 10°. There are 250 snapshots taken for DOA estimation in two scenarios of SNR = 5 dB and 10 dB, respectively. As observed, the larger the signal angles of separation, the smaller the errors.

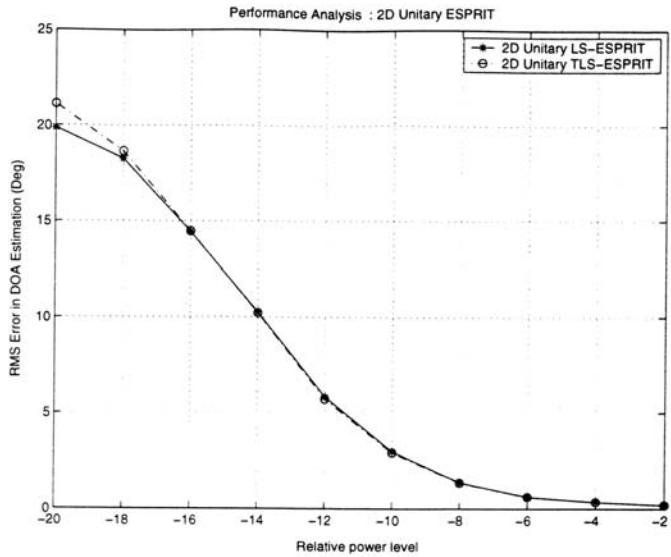


Figure 6.44 RMS error in a DOA estimation as a function of the relative power level (SNR = 5 dB, 16-element URA, two uncorrelated signals at $(\theta_1, \phi_1) = (-20°, -15°)$ and $(\theta_2, \phi_2) = (20°, 15°)$).

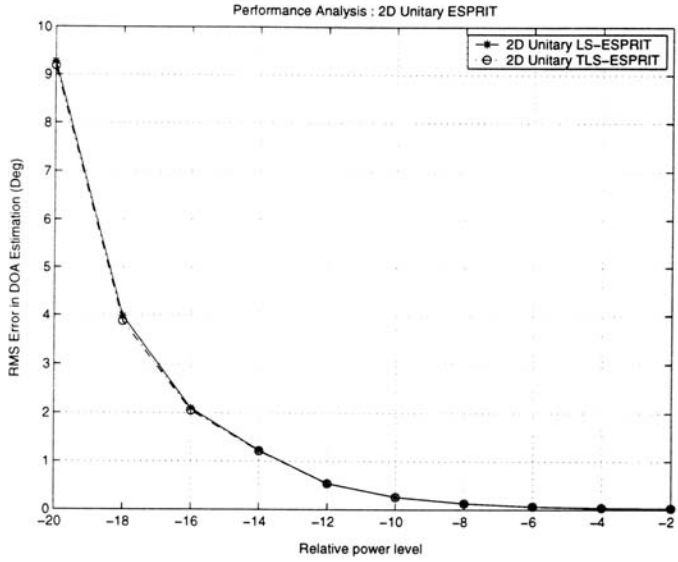


Figure 6.45 RMS error in a DOA estimation as a function of the relative power level (SNR = 10 dB, 16-element URA, two uncorrelated signals at $(\theta_1, \phi_1) = (-20°, -15°)$ and $(\theta_2, \phi_2) = (20°, 15°)$).

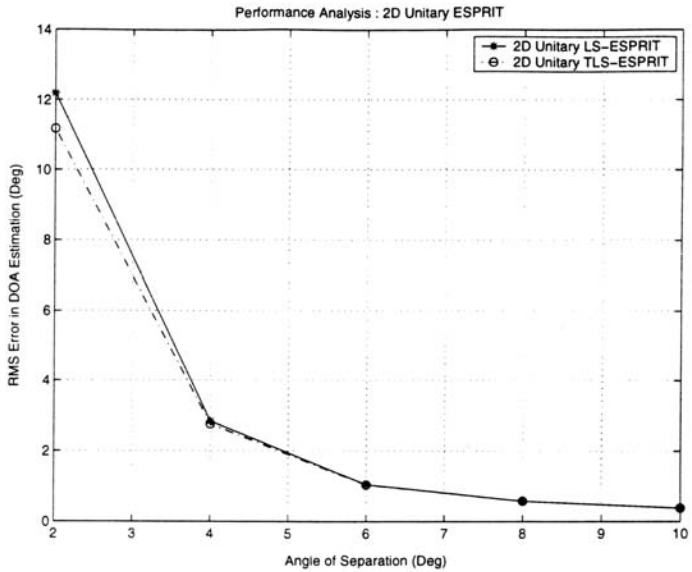


Figure 6.46 RMS error in a DOA estimation as a function of the separation (SNR = 5 dB, 16-element URA, trials = 50, two uncorrelated signals at $(\theta_1, \phi_1) = (-20°, -15°)$ and $(\theta_2, \phi_2) = (20°, 15°)$).

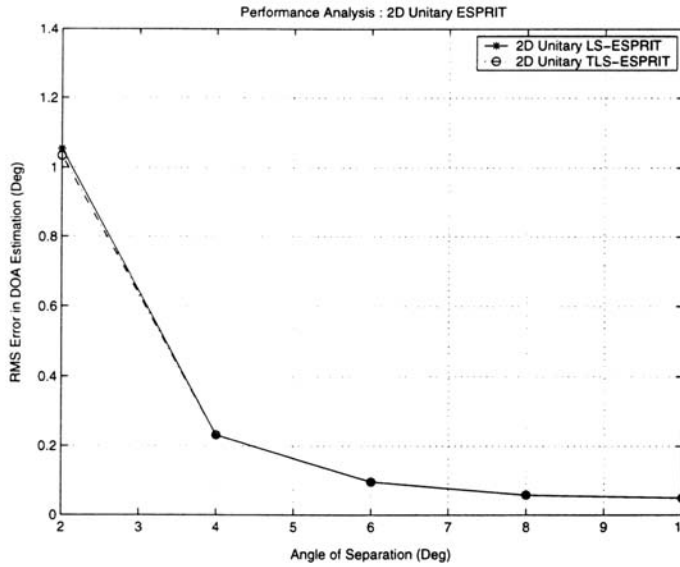


Figure 6.47 RMS error in a DOA estimation as a function of the separation (SNR = 10 dB, 16-element URA, trials = 50, two uncorrelated signals at $(\theta_1, \phi_1) = (-20°, -15°)$ and $(\theta_2, \phi_2) = (20°, 15°)$).

The computational expenditure measured in terms of MATLAB FLOPS is shown in Figure 6.48. The results show that the TLS version asks for more FLOPS than the LS version as the number of elements increase.

Finally, the performance of the algorithm in the coherent environment is presented in Figures 6.49 and 6.50. A $5 \times 5 = 25$ element uniform rectangular array is considered that is divided into nine rectangular subarrays, each of them having nine elements. Then two coherent signals are impinged on the array at $(\theta_1, \phi_1) = (10°, 12°)$ and $(\theta_2, \phi_2) = (12°, 25°)$. The effect of the SNR on the errors is shown by varying the SNR from −5 dB to 5 dB. It is observed from Figures 6.49 and 6.50 that the error is relatively more compared to uncorrelated sources, due to the reduction of the detection capability of the algorithm as explained in Chapter 3. However, the relationships between the DOA estimation accuracy and various parameters are similar to those in uncorrelated signal environments.

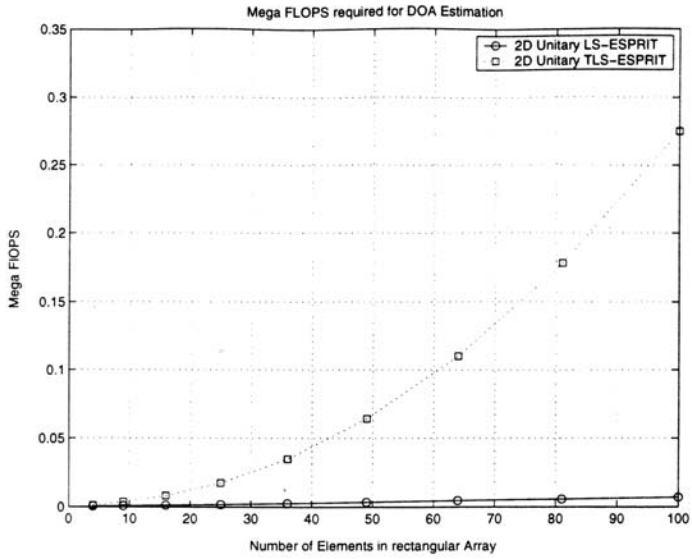


Figure 6.48 Measured FLOPS as a function of the number of elements (SNR = 10 dB, trials = 50, two uncorrelated signals at $(\theta_1, \phi_1) = (-20°, -15°)$ and $(\theta_2, \phi_2) = (20°, 15°)$).

6.3 Comparative Analysis

In the previous sections, analysis of the ESPRIT-based algorithms for DOA estimation is presented. As discussed before, the unitary ESPRIT reduces the computational complexity by doing all mathematical computations in real numbers than in complex numbers as compared to standard ESPRIT. In this section, the performance comparison between the standard ESPRIT and the unitary ESPRIT is presented with a six-element one-dimensional ULA.

Again the RMS error in the DOA estimation is used as the metric of performance. Fifty trials are run, with each trial involving 250 snapshots. These simulation parameters are maintained throughout the simulations unless otherwise mentioned. Two equal power uncorrelated signals at 5° and 10° are impinging on the array.

Figures 6.51 and 6.52 show the performance with respect to the number of snapshots. In Figure 6.51, the SNR is kept at 5 dB. In Figure 6.52, the SNR is kept at 15 dB. It is observed that the unitary ESPRIT achieves a better performance than the standard ESPRIT in both cases. The unitary ESPRIT has an error of about 8% less than standard

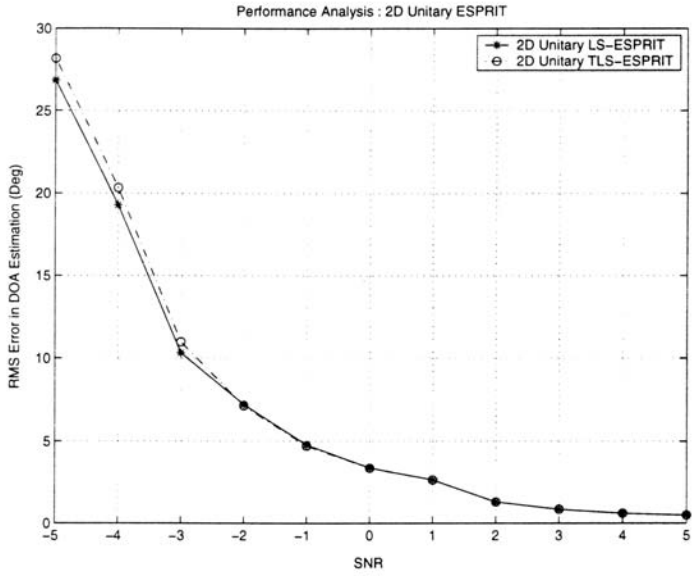


Figure 6.49 RMS error in a DOA estimation as a function of the array SNR (25 element URA, trials = 50, two coherent signals at $(\theta_1, \phi_1) = (10°, 12°)$ and $(\theta_2, \phi_2) = (12°, 25°)$).

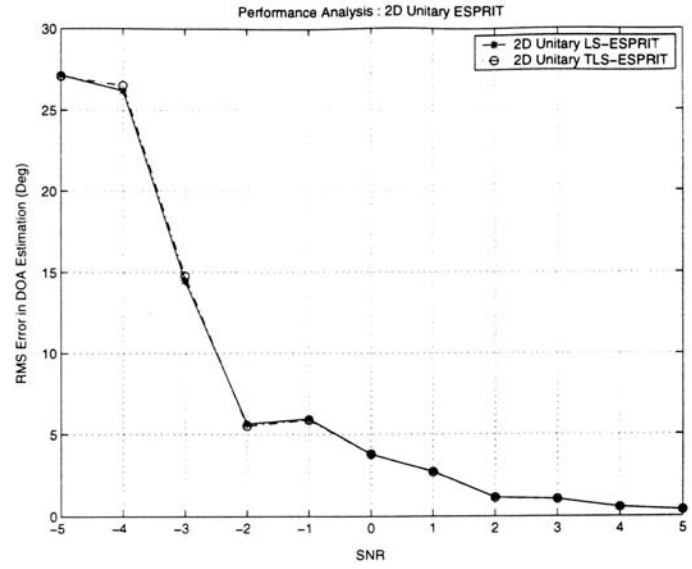


Figure 6.50 RMS error in DOA estimation as a function of SNR (25-element URA, trials = 50, two coherent signals at $(\theta_1, \phi_1) = (10°, 12°)$ and $(\theta_2, \phi_2) = (12°, 25°)$).

ESPRIT at the SNR of 5 dB, and it has about 15% less at the higher SNR of 15 dB.

Figures 6.53 and 6.54 show the performance with respect to the SNR. In Figure 6.53, the SNR is varied from −5 dB to 5 dB, and 250 snapshots of data are taken. In Figure 6.54, 500 snapshots of data are taken with the same variation in the SNR. Again, the unitary ESPRIT outperforms the standard ESPRIT even at low SNR conditions. At −5 dB and with 500 snapshots of data available, it can be seen that the RMS error with the unitary ESPRIT is about 33% less than that with the standard ESPRIT.

The estimation errors of the standard ESPRIT and unitary ESPRIT versus the number of signals are also compared. Figures 6.55 and 6.56 show the results. A 10-element array is taken and nine signals are made impinging on the array. Two simulations are conducted with 500 snapshots per trial at an SNR of 5 dB and 10 dB, respectively. As can be seen from Figures 6.55 and 6.56, both the standard ESPRIT and the unitary ESPRIT perform similarly in picking up the maximum number of the signals with the acceptable errors.

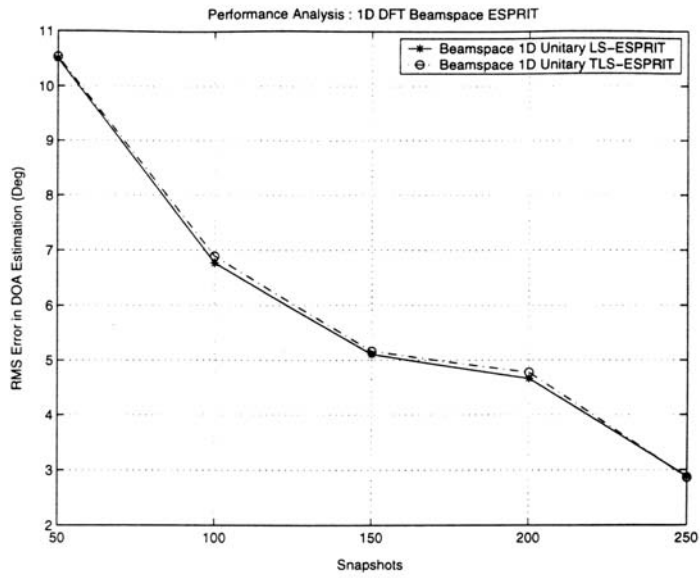


Figure 6.51 RMS error in a DOA estimation as a function of the snapshots (SNR = 5 dB, six-element ULA, trials = 50, two uncorrelated signals at 5° and 10°).

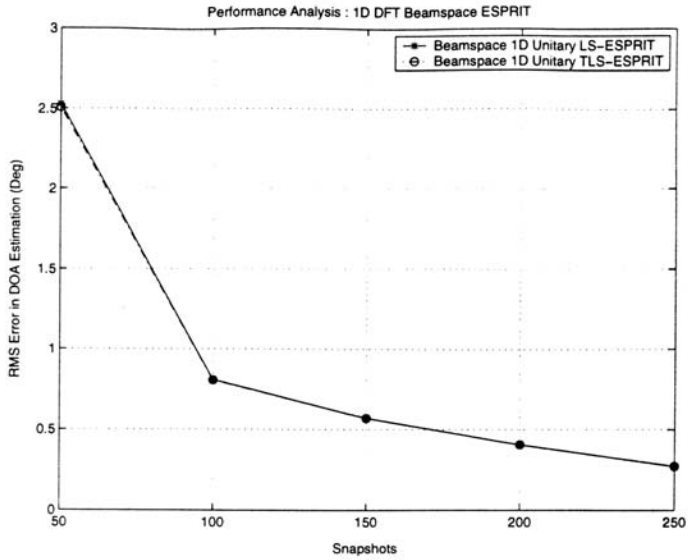


Figure 6.52 RMS error in a DOA estimation as a function of the separation (SNR = 15 dB, six-element ULA, trials = 50, two uncorrelated signals at 5° and 10°).

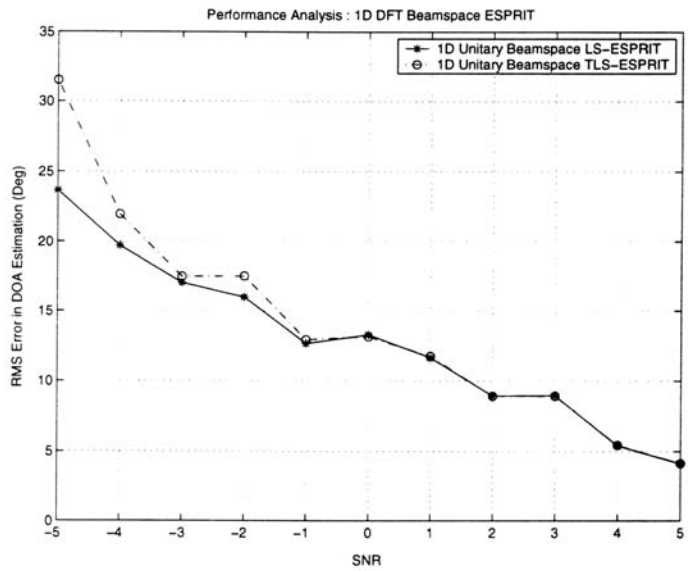


Figure 6.53 RMS error in a DOA estimation as a function of the SNR (snapshots = 250, six-element ULA, trials = 50, two uncorrelated signals at 5° and 10°).

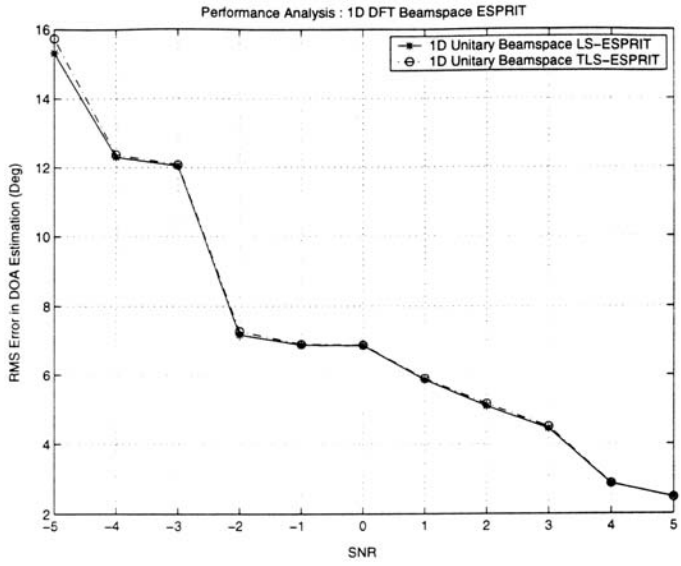


Figure 6.54 RMS error in a DOA estimation as a function of the SNR (snapshots = 500, six-element ULA, trials = 50, two uncorrelated signals at 5° and 10°).

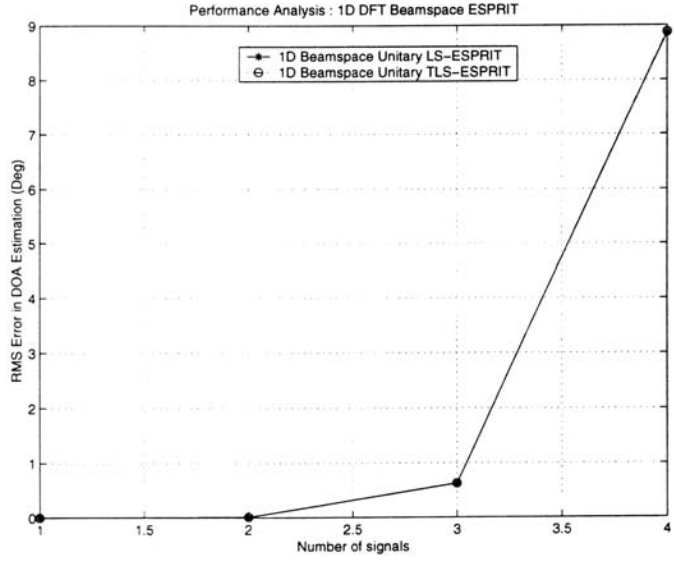


Figure 6.55 RMS error in a DOA estimation as a function of the signals (SNR = 5 dB, six-element ULA, trials = 50, two uncorrelated signals at 5° and 10°).

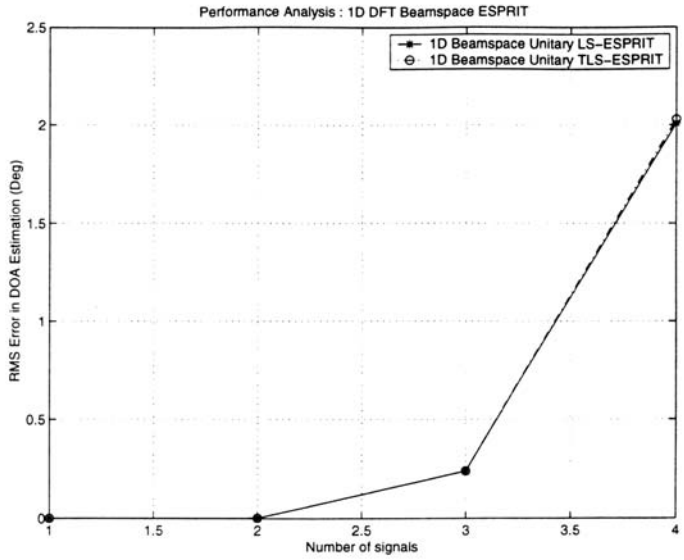


Figure 6.56 RMS error in a DOA estimation as a function of the signals (SNR = 15 dB, 25-element ULA, trials = 50, two uncorrelated signals at 5° and 10°).

The performance of the standard ESPRIT is compared to that of the unitary ESPRIT with respect to the number of elements in the array. The results are shown in Figures 6.57 and 6.58. The number of antenna elements is varied from 4 to 10 and two SNR scenarios of 5 dB and 10 dB are simulated. To estimate the DOA, 250 snapshots are taken. Overall, the unitary ESPRIT shows the errors that are about 10% smaller than those of the standard ESPRIT as the number of elements are increased.

The performance of standard ESPRIT and unitary ESPRIT with respect to power level of the received signal is shown in Figures 6.59 and 6.60 with a 10-element antenna array. Simulations are performed with a signal that is arriving at 5° as the strong signal, and the power of the signals arriving at 10° is varied from −20 dB to −5 dB less than the power of the strong signal. Simulations were conducted in two different SNR scenarios with 5 dB and 10 dB. The results show an error with unitary ESPRIT about 40% smaller than those with the standard ESPRIT at large power differences. However, at low power level differences, both the algorithms perform similarly.

The resolution capabilities of the standard ESPRIT and unitary ESPRIT are compared and presented in Figures 6.61 and 6.62. Here

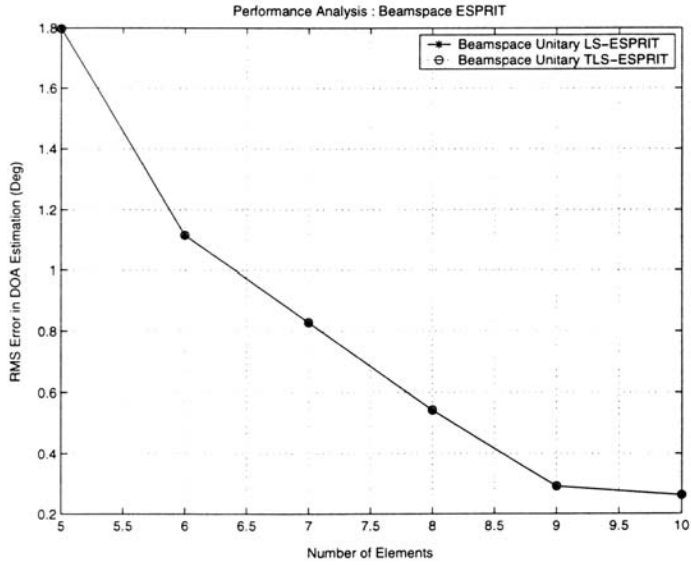


Figure 6.57 RMS error in a DOA estimation as a function of the elements (SNR = 5 dB, six-element ULA, trials = 50, two uncorrelated signals at 5° and 10°).

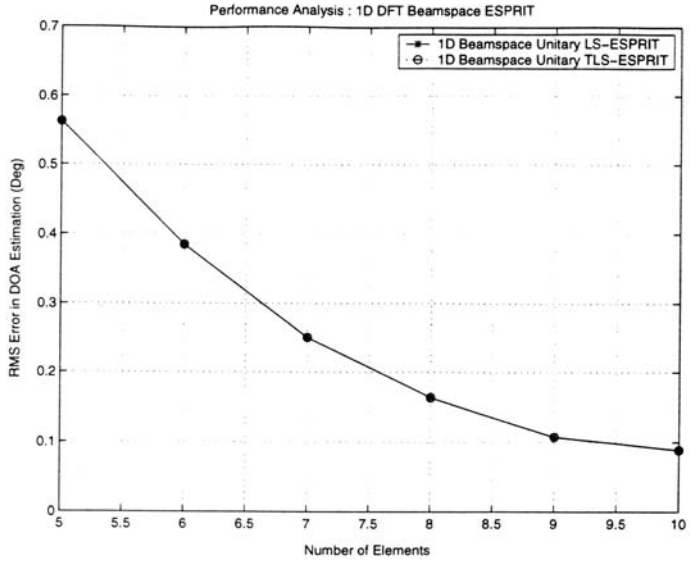


Figure 6.58 RMS error in a DOA estimation as a function of the elements (SNR = 15 dB, 25-element ULA, trials = 50, two uncorrelated signals at 5° and 10°).

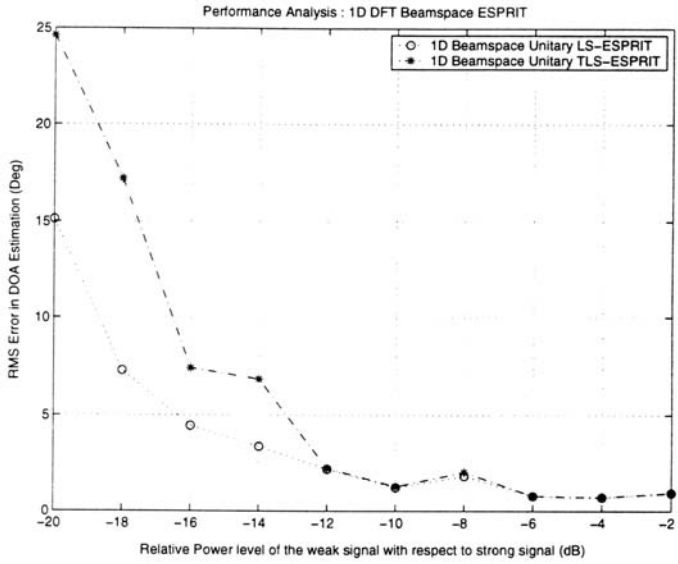


Figure 6.59 RMS error in a DOA estimation as a function of the power level (SNR = 5 dB, six-element ULA, trials = 50, two uncorrelated signals at 5° and 10°).

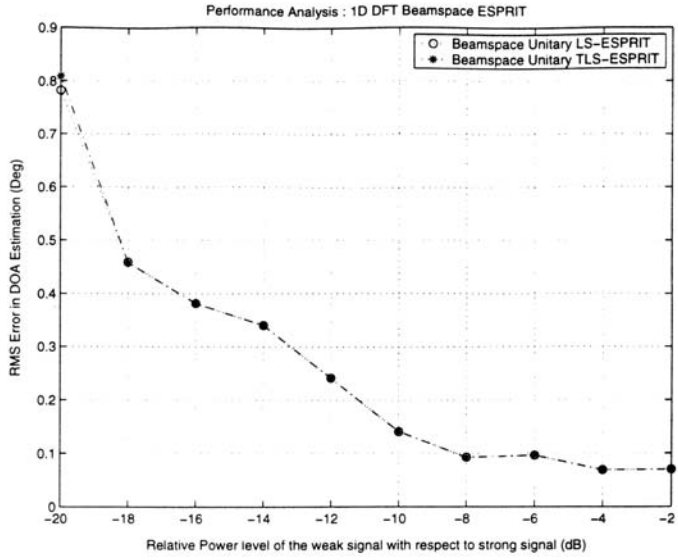


Figure 6.60 RMS error in a DOA estimation as a function of the power level (SNR = 15 dB, six-element ULA, trials = 50, two uncorrelated signals at 5° and 10°).

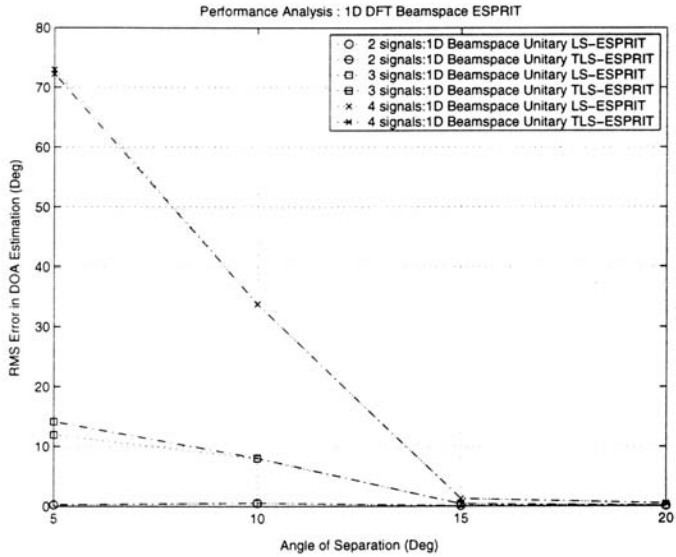


Figure 6.61 RMS error in a DOA estimation as a function of the separation (SNR = 0 dB, six-element ULA, trials = 50).

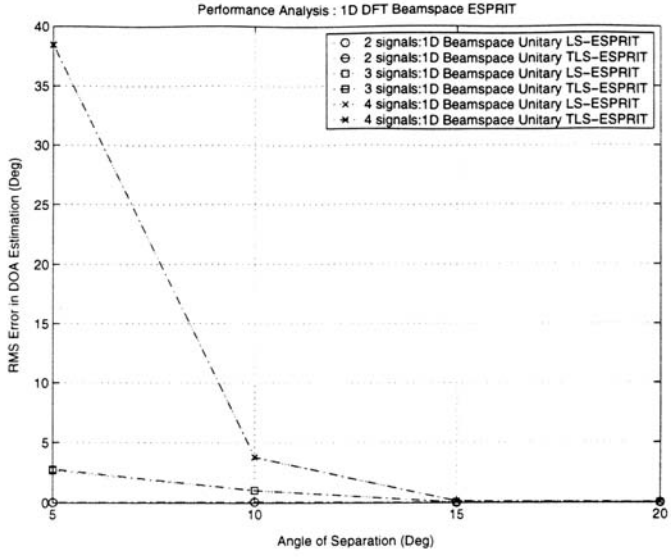


Figure 6.62 RMS error in a DOA estimation as a function of the separation (SNR = 15 dB, 25-element ULA, trials = 50).

three signals impinge on a 10-element ULA with the angular separation between them being varied from 2° to 20°. For DOA estimations, 250 snapshots are taken in two scenarios of SNR = 5 dB and 10 dB. It can be observed that both the algorithms are able to resolve the DOA efficiently. However, the unitary ESPRIT gives better performance than the standard version by the errors of about 1° smaller.

The computational expenditures of the standard and unitary ESPRIT algorithms are compared and shown in Figure 6.63 in terms of MATLAB FLOPS. It is clearly evident that the unitary ESPRIT is demanding less FLOPS that the standard ESPRIT, in particular, as the number of elements increase. Especially for the TLS versions, with a 10-element array, the unitary TLS ESPRIT is consuming about 20,000 FLOPS less than that that of the TLS version of the standard ESPRIT.

6.4 Discussions

In previous sections of this chapter, simulation results obtained by subjecting the ESPRIT-based algorithms to a variety of signal conditions are

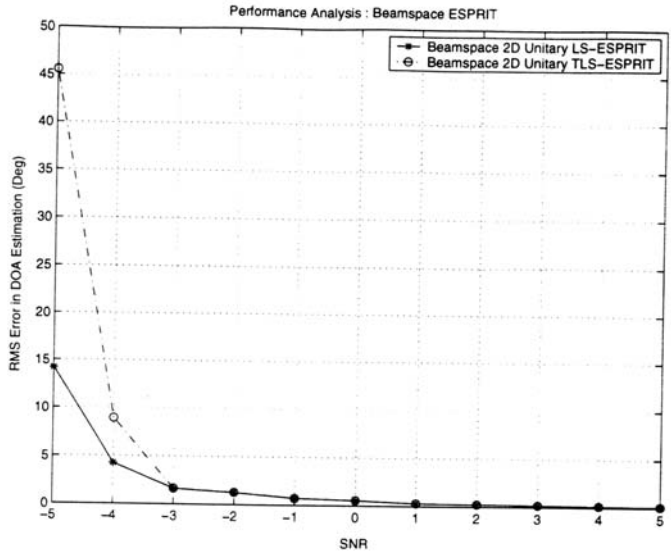


Figure 6.63 Mega FLOPS used as a function of the array elements (SNR = 15 dB, snapshots = 250, two uncorrelated signals at 5° and 10°).

presented. In this section, an attempt has been made to evaluate the obtained results and explain the behavior of the algorithms.

Subspace-based algorithms like ESPRIT basically work on the observation that column space of the observed data matrix can be decomposed into two orthogonal subspaces called signal subspace and noise subspace. In a noise-free environment, the data matrix $\mathbf{X}$ is d-dimensional and contains the spatial frequency information related to DOAs. This d-dimensional subspace is often the signal subspace. In the presence of the noise, the data matrix gets perturbed with noise and the dimension of the matrix becomes greater than d. If the d-dimensional signal subspace can be extracted from the noise perturbed data matrix, the spatial frequencies (in turn, the DOAs) can still be successfully determined. ESPRIT-based algorithms exploit the observation that if the array steering matrix has a shift invariance structure, the estimated signal subspace and the columns of the array steering matrix $\mathbf{A}$ span the same d-dimensional signal subspace. Then the signal subspace can be extracted from the signal space as explained in detail in Chapter 5.

As a result, the performance of all ESPRIT-based algorithms depends upon how closely they can estimate or approximate the signal subspace. As discussed in Chapters 3 and 5, modern methods can estimate the signal subspace or the noise subspace from a noisy data matrix or its covariance matrix, and they do so by using a singular value decomposition (SVD) or an eigenvalue decomposition (EVD).

Suppose that $\mathbf{X}_s$ is an $M \times N$ data matrix without noises and with rank $d < M$, where M is the number of elements and N is the number of snapshots. As described in Chapter 3, denote the SVD of $\mathbf{X}_s$ as $\mathbf{X}_s = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$. Then the SVD is linked to the EVD by

$$\mathbf{X}_s = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H \quad \Rightarrow \quad \mathbf{X}_s\mathbf{X}_s^H = \mathbf{U}\mathbf{\Sigma}^2\mathbf{V}^H \tag{6.3}$$

The singular values of $\mathbf{X}_s$ are the positive square roots of the eigenvalues $\mathbf{X}_s\mathbf{X}_s^H$, while the left singular vectors of $\mathbf{X}_s$ are the eigenvectors of $\mathbf{X}_s\mathbf{X}_s^H$. Let $\mathbf{X}_s$ now be perturbed by some noisy matrix $\Delta\mathbf{X}$ as would be the case in a real world: $\mathbf{X} = \mathbf{X}_s + \Delta\mathbf{X}$.

Assume that the entries of $\Delta\mathbf{X}$ represent uncorrelated and zero mean white noises with variance σ_N^2. Then the expectation $E(\Delta\mathbf{X}\Delta\mathbf{X}^H)$ is

asymptotically (for a number of snapshots $N \to \infty$) given by $E(\Delta\mathbf{X}\Delta\mathbf{X}^H/N) = \sigma_N^2\mathbf{I}_M$. It leads to

$$E\left(\mathbf{X}\mathbf{X}^H/N\right) = E\left(\mathbf{X}_s\mathbf{X}_s^H/N\right) + \sigma_N^2\mathbf{I}_M \tag{6.4}$$

so that, for large N, the SVD of $\mathbf{X}$ is given by

$$\mathbf{X} = \mathbf{U}\sqrt{\Sigma^2 + N\sigma_N^2\mathbf{I}_M}\,\mathbf{U}^H \tag{6.5}$$

The above expression shows that, for large N, that is, for a large number of snapshots N and small σ_N^2, the singular values of $\mathbf{X}$ increases by an amount approximately equal to $\sigma_N\sqrt{N}$, while the left singular vectors of $\mathbf{X}$ remains the same as those of $\mathbf{X}_s$. However, unlike $\mathbf{X}_s$, $\mathbf{X}$ is of full rank and its $M - d$ singular values are no longer zero, but equal to $\sigma_N\sqrt{N}$. Fortunately, the range space of $\mathbf{X}_s$, as estimated from $\mathbf{X}$, is spanned by the left singular vectors corresponding to the d largest singular values of $\mathbf{X}$. Therefore, the signal subspace can still be extracted.

In the case of a small number of snapshots (i.e., smaller N), the singular values of $\mathbf{X}_s$ are raised by an amount on the order of $||\Delta\mathbf{X}||$, the largest singular value of $||\Delta\mathbf{X}||$; this is because $E(\Delta\mathbf{X}\Delta\mathbf{X}^H/N)$ is proportional to $||\Delta\mathbf{X}||$ in this situation. The singular vectors are also perturbed. The amount of perturbation in the estimated signal subspace that they span is in the order of $||\Delta\mathbf{X}||$.

To summarize, the singular values and the subspace spanned by the left singular vectors are relatively insensitive to added noise perturbations on the entries of the matrix if the number of snapshots is reasonably large. This explains the better accuracy obtained with an increasing number of snapshots.

As described in Chapter 5, the performances of ESPRIT depend on the extraction of singular values and the left eigenvectors of the data matrix. In particular, the resolution of the ESPRIT can be found to be dependent on the smallest singular value of the data matrix in relation to the noise level. This singular value is related to the smallest eigenvector, which can be constructed with linear combinations of the columns of the matrix. Therefore, it will be small when the columns are more or less aligned, as illustrated in the example of Figure 6.64. The figure shows the

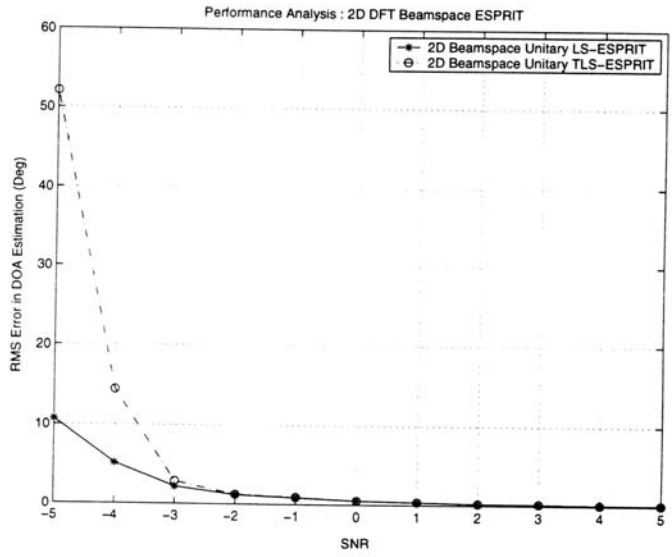


Figure 6.64 Construction of the left singular vectors and values of the matrix $X = [x_1\ x_2]$, where x_1 and x_2 have an equal length.

construction of the left singular vectors of a matrix $\mathbf{X} = [\mathbf{x}_1\ \mathbf{x}_2]$, whose columns $\mathbf{x}_1$ and $\mathbf{x}_2$ are of equal length. The largest possible singular vector, say, $\mathbf{u}_1$, is in the direction of the sum of $\mathbf{x}_1$ and $\mathbf{x}_2$ (i.e., the "common" direction of the two vectors), and the corresponding singular value σ_1 is equal to $\sigma_1 = \|\mathbf{x}_1 + \mathbf{x}_2\|/\sqrt{2}$. On the other hand, the smallest possible singular vector, say, $\mathbf{u}_2$, can be the difference of $\mathbf{x}_1$ and $\mathbf{x}_2$, with its corresponding singular value being $\sigma_2 = \|\mathbf{x}_2 + \mathbf{x}_1\|/\sqrt{2}$. If $\mathbf{x}_1$ and $\mathbf{x}_2$ become more aligned, σ_2 will be smaller and $\mathbf{X}$ will be closer to a singular matrix, thus leading to the failure of the algorithm.

Now, consider the array steering matrix $\mathbf{A}$ as consisting of a linear combination of vectors of $\mathbf{a}(\mu_i)$. If two directions, say, 1 and 2, are close together, $\mu_1 \approx \mu_2$ and $\mathbf{a}(\mu_1)$ points in about the same direction as $\mathbf{a}(\mu_2)$, which will be the direction of $\mathbf{u}_1$. The smallest singular value is dependent on the difference of the directions of $\mathbf{a}(\mu_1)$ and $\mathbf{a}(\mu_2)$. With a noise matrix $\Delta\mathbf{X}$ added, separating the two signals will become difficult if σ_2 is approximately the same or smaller than $||\Delta\mathbf{X}||$. However, if the number of snapshots is high and the EVD approach is employed, all singular values squared will increase by the same amount $||\Delta\mathbf{X}||^2$. In such a case, σ_2 is

automatically greater than ||ΔX|| and locating the second signal is possible. This is why the accuracy of ESPRIT in resolving two closely spaced signals is increasing with the increase in the number of snapshots.

Also, it is observed that smallest singular value is strongly dependent on the value of the M, the length of the vectors (i.e., the number of elements). If M is increased, the difference between $\mathbf{a}(\mu_1)$ and $\mathbf{a}(\mu_2)$ becomes more pronounced, so that σ_2 becomes larger and the resolution increases. This effect is stronger than increasing the number of snapshots N, the number of observation vectors [3].

Standard ESPRIT algorithm can handle, at maximum, $d = M - 1$ uncorrelated signals for a given total number of M array elements. When the correlated signals are to be processed, as explained before, the forward-backward averaging can decorrelate pairs of coherent signals while spatial smoothing can be employed additionally to decorrelate larger sets of coherent signals. In particular, with M array elements, it is possible to estimate the parameters of $d \leq 2/3M$ signals irrespective of the signal correlations when applying the forward backward averaging and spatial smoothing technique to subspace-based methods [4]. This is particularly the case if spatial smoothing is used as a preprocessing step for the unitary ESPRIT, since the unitary ESPRIT inherently includes forward-backward averaging as explained in Chapters 4 and 5. The maximum number of sources that the two-dimensional unitary ESPRIT can handle is the minimum of m_x and m_y. In the case of the two-dimensional ESPRIT in DFT beamspace, the maximum number of sources that can be handled is given by the minimum of b_x and b_y as described in Chapter 5.

Subspace-based algorithms like ESPRIT are computationally intensive. In particular, the SVD or EVD consumes the majority of time, thereby hindering the real-time implementation. If N is the number of snapshots whose data are to be processed and M is the number of elements, forming the data covariance matrix requires the order of NM^2 operations. Eigendecompositions of matrices of $R^{M\times N}$ require the order of $10M^3$ operations, whereas the standard SVD of $\mathbf{X}$ requires approximately $2NM + 4M^3$ if only the singular values and the left singular vectors are computed.

Since most high-resolution DOA estimation algorithms (such as ESPRIT) require an $O(M^3)$ eigendecomposition [5], the increase of M will cause a cubic increase of computational expenditure. Also, the conventional eigendecomposition algorithms are difficult to implement in parallel for a reduction in computational time. Therefore, high

computational complexity may represent a fundamental barrier for implementing the high resolution DOA algorithms in real time.

The unitary ESPRIT reduces the computational burden by performing all the computations in real numbers. Since multiplications of complex numbers generally require four times a product of real numbers, the overall complexity of matrix inversions and eigendecompositions is reduced almost by a factor of 4.

For large arrays (i.e., large M), the beamspace approach can be adopted for a less computational load [6]. In the beamspace approach, the original data vector is decomposed into several lower dimensional subbands or beamspaces via a transformation (e.g., Fourier transformation), and then the DOA estimation is carried out based on the subband data. Since the data processing of each subband is independent, it can be carried out in parallel. If the dimension of each subband is $L < M$, the computational time can be reduced from $O(M^3)$ to $O(L^3)$. Since the processing in each beamspace can be carried out in parallel, the total computation time is also $O(L^3)$. In fact, using the recently proposed fast signal-subspace decomposition (FSD) approach instead of standard decomposition, the computation time can be further reduced to $O(L^2 d)$. If the parallel computation is incorporated (i.e., if L or L^2 simple array processors are used), the computation time can be cut down to $O(Ld)$ or $O(\log Ld)$, respectively. Since the beamspace transformation can be done in $O(ML)$ flops for each data vector, the beamspace sample covariance matrix can be accumulated in real time. In addition, for a ULA, FFT can be employed to further reduce the computational complexity.

The two-dimensional DFT beamspace ESPRIT can be applied in a reduced dimensional beamspace to perform parallel sector-wise searches for DOAs in different regions of the angle space. This is brought about by working with a subset of two-dimensional DFT beams that have main lobes in the spatial sector of interest. The computational complexity can be explained by taking an example described in [6] where an 8×8 URA was employed. The two-dimensional DFT beamspace ESPRIT could then employ 9 beams (a 3×3 subset of the two-dimensional DFT beams), with 9 such overlapping beam sets spanning the entire arrival angle space. Letting three uncorrelated sources be in the sector being spanned, the computations in element space required for a single DOA trial run are:

1. Sixty-four additions for each of 64 snapshots to transform from complex-valued space to real-valued space;
2. Calculation of the three "largest" left singular vectors of a 64×128 real-valued matrix;
3. Calculation of the solution to two systems of equations in the form $\mathbf{AX} = \mathbf{B}$, where $\mathbf{A}$ and $\mathbf{B}$ are both 64×3 and real-valued;
4. Calculation of the eigenvalues of a 3×3 matrix.

In the case of a reduced dimension, the two-dimensional DFT beamspace, the computations required for a single trial run are:

1. Nine sets of 64 multiplications and 63 additions for each of 64 snapshots to transform from element space to beamspace;
2. Calculation of the three "largest" left singular vectors of a 9×128 real-valued matrix;
3. Calculation of the eigenvalues of a 3×3 matrix.

As can be seen now, the ESPRIT in the DFT beamspace requires less computational expenditure than that in the element space.

6.5 Conclusion

In this chapter, in-depth analysis of ESPRIT-based algorithms for a DOA estimation is performed. The performances of the standard ESPRIT and the unitary ESPRIT are presented. In addition, the beamspace approaches for the unitary ESPRIT in one and two dimensions are also analyzed in detail. Finally, the computational complexity of the algorithms is studied in terms of FLOPS necessary for the algorithm.

It is found that the accuracy in a DOA estimation increases with an increasing number of snapshots, the SNR, and the number of array elements. Also, it is observed that there is a limitation in the number of signals that the DOA algorithms can handle and the resolution that the algorithms can achieve. In a practical noisy environment, the number of the signals that an array can resolve is often less than the theoretically predicted one, the number of elements less than one. Finally, the unitary ESPRIT is shown to have the best performance in general with the least the computational expenditures needed.

References

[1] Recktenwald, G., *Numerical Methods with MATLAB Implementation and Application*, Upper Saddle River, NJ: Prentice-Hall, 2000.

[2] Hanselman, D., and B. Littlefield, *Mastering MATLAB 5: A Comprehensive Tutorial and Reference*, Upper Saddle River, NJ: Prentice-Hall, 1998.

[3] vander Veen, A. J., E. F. Deprettere, and A. L. Swindlehurst, "Subspace Based Signal Analysis Using Singular Value Decomposition," *Proc. IEEE*, Vol. 81, No. 9, September 1993, pp. 1277–1308.

[4] Pillai, S. U., *Array Signal Processing*, New York: Springer-Verlag, 1989.

[5] Roy, R., and T. Kailath, "ESPRIT-Estimation of Signal Parameters Via Rotational Invariance Techniques," *IEEE Trans. on Acoust., Speech, Signal Processing*, Vol. 37, No. 7, July 1989, pp. 984–995.

[6] Zoltowski, M. D., M. Haardt, and C. P. Mathews, "Closed-Form 2D Angle Estimation with Rectangular Arrays in Element Space or Beamspace Via Unitary ESPRIT," *IEEE Trans. on Signal Processing*, Vol. 44, No. 2, February 1996, pp. 316–328.

7

Discussions and Conclusion

7.1 Summary

This book has provided an overview of a few basic DOA estimation algorithms and their operational principle; the book is aimed for students, engineers, or government regulators who need to gain an insight into the fundamentals of DOA estimations. Major DOA estimation algorithms including beamforming, maximum likelihood, and subspace-based techniques have been discussed in a systematic way. As a result, this book forms a single source of reference for basic DOA estimation algorithms and places them under one roof. Simulation results are presented in support of the discussions. Also, a broad introduction to preprocessing schemes and model order estimation techniques has been provided with simulation examples.

This book has given an in-depth comprehensive performance analysis of ESPRIT-based algorithms for DOA estimation. As part of this, the performances of standard ESPRIT, unitary ESPRIT, and DFT beamspace ESPRIT have been thoroughly examined by exposing them to variety of signal conditions. The algorithms have been studied for their performance in both one and two dimensions and both in uncorrelated and coherent environments. Therefore, this book contributes to form an instructive guide for the designers employing these algorithms for direction finding applications. The designer shall be able to understand the trade-offs and the requirements of the algorithms by studying the analytical and simulation results presented. In summary, this book helps to gain

experience on DOA estimation algorithms and lays the theoretical foundation for real-time physical implementation.

7.2 Advanced Topics on DOA Estimations

This book aims to provide an introduction to the fundamentals of the basic DOA estimation methods. The radio signals under study have been assumed to be of narrowband and the antenna arrays have been assumed to be stationary in one dimension or two dimensions with their elements uniformly distributed in space. The sources have also been assumed to be stationary and in the far-field regions of the arrays.

More advanced topics on DOA estimations deal mainly with the removal of one or more of the above assumptions in addition to further improvement of DOA estimation accuracies and computational efficiency. These topics include DOA estimations

1. With nonuniform arrays;
2. With three-dimensional arrays;
3. Of wideband signal sources;
4. Of moving sources;
5. With delay or time-of-arrival estimations;
6. With pulsed systems;
7. In the presence of interferences and mutual coupling among elements;
8. With other techniques such as switched arrays, mechanically rotating arrays, and improving algorithms by signal property exploitation.

These advanced topics can be considered as based on or as extensions of the DOA estimation algorithms described in this book. A large body of literature can be found in public domain on these advanced topics, for instance, in IEEE Xplore. References [1–18] are a subset of this literature. In particular, Chandran [1] covers quite a wide range of topics on modern DOA techniques.

Finally, the physical implementation of these algorithms in a DSP-based smart antenna system with DOA capability forms a very exciting and challenging task for the ultimate proof of the utility of the

DOA algorithms in a real-world application. In this respect, not very much documentation has been seen in the public domain.

References

[1] Chandran, S., (ed.), *Advances in Direction-of-Arrival Estimation*, Norwood, MA: Artech House, 2005.

[2] Van Trees, H. L., *Detection, Estimation, and Modulation Theory, Part IV, Optimum Array Processing*, New York: John Wiley & Sons, 2002.

[3] Krim, H., and M. Viberg, "Two Decades of Array Signal Processing Research," *IEEE Signal Processing Magazine*, Vol. 13, No. 4, July 1996, pp. 67–94.

[4] Schoor, M., and B. Yang, "Subspace Based DOA Estimation in the Presence of Correlated Signals and Model Errors," *Proceedings of IEEE International Conf. on Acoustics, Speech and Signal Processing*, April 19–24, 2009, pp. 2161–2164.

[5] Li, W., W. Yao, and P. J. Duffer-Smith, "Comparative Study of Joint TOA/DOA Estimation Techniques for Mobile Positioning Applications," *Proceedings of IEEE Consumer Communications and Networking Conference*, January 10–13, 2009, pp. 1–5.

[6] Liang, J., "Joint Azimuth and Elevation Direction Finding Using Cumulant," *IEEE Sensors Journal*, Vol. 9, No. 4, April 2009, pp. 390–398.

[7] Harabi, F., A. Gharsallah, and S. Marcos, "Three-Dimensional Antennas Array for the Estimation of Direction of Arrival," *IET Microwaves, Antennas & Propagation*, Vol. 3, No. 5, August 2009, pp. 843–849.

[8] Ye, Z., Y. Zhang, and C. Liu, "Direction-of-Arrival Estimation for Uncorrelated and Coherent Signals with Fewer Sensors," *IET Microwaves, Antennas & Propagation*, Vol. 3, No. 3, April 2009, pp. 473–482.

[9] Xu, X., Z. Ye, and Y. Zhang, "DOA Estimation for Mixed Signals in the Presence of Mutual Coupling," *IEEE Trans. on Signal Processing*, 2010.

[10] Gotsis, K. A., K. Siakavara, and J. N. Sahalos, "On the Direction of Arrival (DOA) Estimation for a Switched-Beam Antenna System Using Neural Networks," *IEEE Trans. on Antennas and Propagations*, Vol. 57, No. 5, May 2009, pp. 1399–1411.

[11] Zhao, L., Y. Zhao, and H. Cui, "High Resolution Wideband DOA Estimation Based on Modified MUSIC Algorithm," *Proceedings of 2008 International Conference on Information and Automation*, June 20–23, 2008, pp. 20–22.

[12] Yasar, T. K., and T. E. Tuncer, "Wideband DOA Estimation for Nonuniform Linear Arrays with Wiener Array Interpolation," *Proceedings of 2008 IEEE Sensor Array and Multichannel Signal Processing Workshop*, July 21–23, 2008, pp. 207–211.

[13] Tosa, L., B. Denis and B. Uguen, "Direct Path DOA and DoD Finding Through IR-UWB Communications," *Proceedings of 2008 IEEE International Conference on Ultra-Wideband*, Vol. 2, September 10–12, 2008, pp. 223–227.

[14] M. Navarro and M. Najar, "TOA and DOA Estimation for Positioning and Tracking in IR-UWB," *Proceedings of 2007 IEEE International Conference on Ultra-Wideband*, September 24–26, 2007, pp. 574–579.

[15] Kisliansky, A., R. Shavit, and J. Tabrikian, "Direction of Arrival Estimation in the Presence of Noise Coupling in Antenna Arrays," *IEEE Trans. on Antennas and Propagation*, Vol. 55, No. 7, July 2007, pp. 1940–1947.

[16] Wang, G., Y. Pan, and W. Dong, "A New Anti-Interference Preprocess Method for DOA Estimation Based on Uniform Circular Array," *Proceedings of 2006 International Conference on Radar*, October 16–19, 2006, pp. 1–4.

[17] Feng, J., Y. Yang, and C. Sun, "Beamspace DOA Estimation Algorithms with Robustness Against Fast-Moving Strong Interferers," *Digests of 2006 IEEE Antennas and Propagation Society Internal Symposium*, July 9–14, 2006, pp. 2553–2556.

[18] Zatman, M. A., "Superresolution Direction Finding with Rotating Arrays," *Proceedings of the 34th Asilomar Conference on Signals, Systems and Computers*, Vol. 1, October 29–November 1, 2009, pp. 36–40.

Appendix

A.1 Kronecker Product

Kronecker products, also known as direct products or tensor products, are used frequently in this book, especially when working with 2D arrays. Given two matrices $\mathbf{A} = [a_{ij}] \in C^{m \times n}$ and $\mathbf{B} = [b_{ij}] \in C^{p \times q}$, the Kronecker product of **A** and **B** is defined as the partitioned matrix:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{n1}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{n2}\mathbf{B} \\ \dots & \dots & \dots & \dots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \dots & a_{mn}\mathbf{B} \end{bmatrix} \tag{A.1}$$

Very often, Kronecker products are used with the vec{·} operator. Here, vec{**A**} denotes a vector-valued function that maps an $m \times n$ matrix **A** into an $m \cdot n$-dimensional column vector by stacking the columns of the matrix **A**. Given $\mathbf{Y}_1 = [y_{1ij}] \in C^{y_1 \times y_2}$, $\mathbf{Y}_2 = [y_{2ij}] \in C^{y_2 \times y_3}$, and $\mathbf{Y}_3 = [y_{3ij}] \in C^{y_3 \times y_4}$, the following important identity relates the vec{·} operator with the Kronecker products:

$$vec\{\mathbf{Y}_1\mathbf{Y}_2\mathbf{Y}_3\} = (\mathbf{Y}_3^T \otimes \mathbf{Y}_1)vec\{\mathbf{Y}_2\} \tag{A.2}$$

A.2 Special Vectors and Matrix Notations

This section gives a brief summary of the notations used in this book. If for any positive integer p, $\mathbf{I}_p$ denotes the $p \times p$ identity matrix and Π_p denotes the $p \times p$ exchange matrix with 1 on its antidiagonal and zeros elsewhere:

$$\Pi_p = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 1 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 1 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \end{bmatrix} \in R^{p \times p} \tag{A.3}$$

Π_p is a symmetric matrix and has the property that $\Pi_p^2 = \mathbf{I}_p$. It is not difficult to show that the premultiplication of a matrix by Π_p will reverse the order of its rows, whereas the postmultiplication of a matrix by Π_p will reverse the order of its columns.

A diagonal matrix Φ_d with the diagonal elements $\phi_1, \phi_2, \dots, \phi_d$ is denoted as

$$\Phi_d = \begin{bmatrix} \phi_1 & 0 & \dots & 0 & 0 \\ 0 & \phi_2 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \phi_{d-1} & 0 \\ 0 & 0 & \dots & 0 & \phi_d \end{bmatrix} \in C^{d \times d} \text{ or } \in R^{d \times d} \tag{A.4}$$

A.3 FLOPS

The easiest way to measure computational effort is to count the number of floating point operations (FLOPS) in an algorithm. A FLOP is an addition, subtraction, multiplication, or division. As a rough model of computational expenditure, it is assumed that each FLOP takes the same amount of computational time. Thus, the algorithms that have higher FLOP counts usually take longer to run than algorithms with lower FLOP counts.

For example, consider the inner product of two column vectors u and v with five elements each:

$$\sigma = u^T v = u_1 v_1 + u_2 v_2 + u_3 v_3 + u_4 v_4 + u_5 v_5 \tag{A.5}$$

The computation of $u^T v$ appears to take five multiplications and four additions, or nine FLOPS in total. One more FLOP is hidden when the inner product is written in the preceding mathematical form. Hence, if u and v are of length n, then an inner product of u and v takes $2n$ FLOPS. The product of an $m \times r$ matrix and an $r \times n$ matrix involves mn inner products of length r, for a total of $2mnr$ FLOPS. If two matrices are square (i.e., $m = n = r$), then the matrix-matrix product takes $2n^3$, or $O(n^3)$ FLOPS. In discussing the number of FLOPS required by matrix operations, one is usually only concerned with a single significant digit in the work estimated. Greater precision is not useful, because the actual execution time can be greatly influenced by implementation details in the software and the design of the computer hardware.

List of Abbreviations

1D One-dimensional

2D Two-dimensional

AA Adaptive array

AIC Akaike information criterion

AWGN Additive white Gaussian noise

DFT Discrete Fourier transform

DOA Direction of arrival

ESPRIT Estimation of signal parameters via rotational invariance techniques

EVD Eigenvalue decomposition

FLOPS Floating point operations

LS Least square

LMS Least mean square

MDL Minimum descriptive length

ML Maximum likelihood

MUSIC Multiple Signal Classification

PA Phased array

PAST Project approximation subspace tracking

RMS Root mean square

RMSE Root mean square error

SDMA Space division multiple access

SL Switched lobe

SNR Signal-to-noise ratio

SVD Singular value decomposition

TLS Total least square

ULA Uniform linear array

URA Uniform rectangular array

URFA Uniform rectangular frame array

About the Authors

Zhizhang (David) Chen received his B. Eng degree from Fuzhou University, P. R. China; his M.A. Sc. degree from Southeast University, P. R. China; and his Ph.D. degree from the University of Ottawa in 1992. From January to August of 1993, he was a NSERC postdoctoral fellow with the ECE Department of McGill University, Montreal, Canada. In 1993, he joined Dalhousie University, Halifax, Canada, where he is presently a full professor and Killam Chair in wireless technology.

Dr. Chen has authored and coauthored over 170 journal and conference papers in computational electromagnetics and RF/microwave electronics. He received the 2005 Nova Scotia Engineering Award, a 2006 Dalhousie graduate teaching award, and the 2007 Dalhousie Faculty of Engineering Research Award. His current research interests include numerical modeling and simulation, RF/microwave electronics, smart antennas, and wireless transceiving technology and applications. He is a fellow of the IEEE.

Gopal Gokeda received his M.A. Sc. degree in electrical engineering from Dalhousie University, Halifax, Nova Scotia, Canada, in 2002. He was a technology and design manger with Sensor Technology Ltd., Collingwood, Ontario, Canada, from 2006 to 2009. Between 2003 and 2006, he worked with Orion Electronics Ltd. (now Cobham Tracking and Locating Ltd.), Nova Scotia, Canada, as a project engineer. He was with the Defense Research and Development Organization in India as an electronic engineer from 1998–1999. His areas of expertise include

project management, CPM scheduling and earned value analysis; government contracting and tendering; ISO quality and compliance control; wireless and acoustic system design; and hardware, software, and firmware design.

Yiqiang Yu received his M.A. Sc. degree with distinction in communication systems in 2003 and his Ph.D. in microwave communications engineering in 2007, both from Swansea University, United Kingdom. He is now a postdoctoral research fellow with the Department of Electrical and Computer Engineering, Dalhousie University, Halifax, Canada.

Dr. Yu was a recipient of the Overseas Research Scholarship, awarded from the U.K. Overseas Research Award Scheme during 2004–2007.

Dr. Yu's primary interest is in the applications of computational electromagnetics—in particular, the use of finite-difference methods, method of moments, and fast multipole methods in both the time and frequency domains. His interests also include RF/microwave components, antennas design and measurement, EMI/EMC testing, and iterative solvers and preconditioning techniques for large-scale matrix computation.

Index

Recent Titles in the Artech House Signal Processing Library

Computer Speech Technology, Robert D. Rodman

Digital Signal Processing and Statistical Classification, George J. Miao and Mark A. Clements

Handbook of Neural Networks for Speech Processing, Shigeru Katagiri, editor

Hilbert Transforms in Signal Processing, Stefan L. Hahn

Introduction to Direction-of-Arrival Estimation, Zhizhang Chen, Gopal Gokeda, and Yi-qiang Yu, Editors

Phase and Phase-Difference Modulation in Digital Communications, Yuri Okunev

Signal Processing in Digital Communications, George J. Miao

Signal Processing Fundamentals and Applications for Communications and Sensing Systems, John Minkoff

Signals, Oscillations, and Waves: A Modern Approach, David Vakman

Statistical Signal Characterization, Herbert L. Hirsch

Statistical Signal Characterization Algorithms and Analysis Programs, Herbert L. Hirsch

Voice Recognition, Richard L. Klevans and Robert D. Rodman

For further information on these and other Artech House titles, including previously considered out-of-print books now available through our In-Print-Forever® (IPF®) program, contact:

Artech House
685 Canton Street
Norwood, MA 02062
Phone: 781-769-9750
Fax: 781-769-6334
e-mail: artech@artechhouse.com

Artech House
46 Gillingham Street
London SW1V 1AH UK
Phone: +44 (0)20 7596-8750
Fax: +44 (0)20 7630-0166
e-mail: artech-uk@artechhouse.com

Find us on the World Wide Web at: www.artechhouse.com