Array Processing for Detection and Localization of Narrowband, Wideband and Distributed Sources

by

Shahrokh Valaee

Department of Electrical Engineering

McGill University, Montreal

May, 1994

A Thesis submitted to the Faculty of Graduate Studies and Research

in partial fulfillment of the requirements of the degree of

Doctor of Philosophy

to my parents

Abstract

The detection and estimation techniques that are used in array processing depend on the spatial and temporal characteristics of the signals that arrive at the array. In this dissertation, we consider narrowband as well as wideband signals. For narrowband signals, a detection method based on the predictive stochastic complexity (PSC) is developed. The PSC of data is computed for all the models with order smaller than the number of sensors. The number of signals is selected by choosing the minimum of the PSC over all models. The PSC criterion is on-line and can be used for time varying systems and target tracking.

We also consider wideband signals. One approach to wideband array processing is based on sampling the spectrum of the source signals to generate narrowband signals. Then, using a focusing approach, the information at different frequency bins are combined. Here, an optimal method to select a focusing subspace for the well-known coherent signal-subspace method (CSM) is proposed. It is also shown that with the CSM method unbiased estimation of the directions-of-arrival (DOAs) is not possible. Inspired by the CSM algorithm, a new method for wideband array processing is developed which is based on two-sided transformation of the correlation matrices (TCT). The TCT estimator can generate unbiased estimates of the DOAs and has a lower resolution threshold than the CSM algorithm.

In array processing it is frequently assumed that the signals are generated by point sources. This is an assumption which is not satisfied in reality. In this dissertation, a method is developed for localization of spatially distributed sources. The method is based on generalization of the MUSIC algorithm and is applied to coherent and incoherent distribution of sources.

Résumé

Les techniques de détection et d'estimation utilisées en traitement d'antenne dépendent des caractéristiques spatiales et temporelles des signaux arrivant sur l'antenne. Dans ce mémoire, nous considérons les signaux à bande étroite et à bande large. Pour les signaux à bande étroite, nous developpons une méthode de détection basée sur la compléxité stochastique de prédiction (CSP). La CSP des données est calculée pour tous les modèles d'ordre inférieur au nombre de capteurs. Le nombre de signaux est selectionné en choisissant le minimum de la CSP parmi tous les modèles. Le critère de détection basé sur la CSP est de type récursif et peut être utilisé pour les systèmes variant dans le temps ainsi que pour le suivi de cible mobile.

Nous considérons ensuite les signaux à bande large. Une approche au traitement d'antenne à bande large consiste à échantillonner le spectre des signaux source afin de générer des signaux à bande étroite. En utilisant ensuite des méthodes de focalisation, les éléments d'information correspondants à différentes fréquences sont combinées. Nous proposons ici une technique optimale de sélection d'un sous-espace de focalisation pour la méthode dite de sous-espace signal cohérent (MSC), laquelle est déjà bien établie. Nous montrons aussi que la méthode MSC ne permet pas une estimation non-biaisée des directions d'arrivée. En nous inspirant de l'algorithme MSC, nous développons une nouvelle méthode pour le traitement d'antenne à bande large basée sur une transformation bilatérale des matrices de corrélation. L'estimateur qui en découle génère des estimés non-biaisés des directions d'arrivée et possède un seuil de résolution moindre que celui de l'algorithme MSC.

En traitement d'antenne, on suppose souvent que les signaux sont générés par des sources ponctuelles. En réalité, cette hypothèse est rarement satisfaite. Dans ce mémoire,

nous développons une méthode de localisation pour les sources spatialement distribuées. Cette méthode est basée sur une généralisation de l'algorithme MUSIC et est appliquée à des distributions cohérentes et incohérentes de sources.

Acknowledgement

I would like to extend my sincere gratitude and appreciation to my research supervisor Prof. P. Kabal for his advice, guidance, and encouragement throughout this research. He provided a peace of mind that made the Ph.D. period a pleasant time for me.

I am also grateful to Prof. B. Champagne of INRS-Télécommunications for his constructive comments and invaluable remarks particularly on the material of Chapter 6.

The major financial support during this study was provided by the Ministry of Culture and Higher Education, IRAN. This research was also supported in parts by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) and INRS-Télécommunications, Université du Québec.

This research was conducted at INRS-Télécommunications and I am very grateful for the facilities supplied by them.

I would like to thank all my friends at McGill university and INRS-Télécommunications whose companionship made a pleasant environment for study and friendship. I also would like to thank Abdol-Reza Mansouri and Prof. B. Champagne for translation and review of the french abstract.

The members of my family have always been the strongest support in my life. My biggest debt of gratitude goes to my parents for their generous love and continuous support and encouragement from long distance.

Contents

Abstract i

Résumé ii

Acknowledgement iv

Contents v

List of Figures viii

List of Tables xi

1 Introduction 1

1.1 Array processing 2

1.2 Contributions 7

1.3 Thesis organization 9

2 Array Processing Techniques 11

2.1 Time-space transformation 13

2.2 Deterministic signals 14

2.3 Stochastic signals 16

2.4 Array processing 19

2.4.1 Beamforming 23

2.4.2 Subspace decomposition methods 31

2.4.2.1 Detection . . . 33
2.4.2.2 Localization . . . 35
2.5 Wideband array processing . . . 38
2.6 Distributed sources . . . 42

3 Signal Detection **44**
3.1 The PSC criterion . . . 45
3.2 Detection of noncoherent sources . . . 48
3.3 Subspace Decomposition . . . 49
3.4 Detection of coherent sources . . . 52
3.5 Estimation of the DOAs . . . 54
3.6 Consistency of the PSC algorithm . . . 56
3.7 More discussion on the coherent signal detection . . . 64
3.8 Simulation Results . . . 66

4 Optimum Focusing **71**
4.1 The coherent signal subspace method . . . 72
4.2 Focusing frequency selection . . . 76
4.3 The bias of estimation . . . 79
4.4 Tightness of the upper bound . . . 82
4.5 Simulation Results . . . 83

5 Correlation Transformation **88**
5.1 Matrix approximation . . . 89
5.2 Two-sided Correlation Transformation method . . . 91
5.2.1 The TCT criterion . . . 91
5.2.2 Selecting the focusing subspace . . . 93
5.2.3 The TCT algorithm . . . 96
5.3 The error of transformation . . . 97
5.4 Eigenvalues of the universal focused correlation matrix . . . 99
5.4.1 Analytical study . . . 99

5.4.2 Experimental results . . . 101
5.5 The Bias of estimation . . . 102
5.6 The effect of noise on estimation . . . 105
5.7 Performance comparison . . . 107
5.7.1 Simulation results . . . 108

6 Distributed Source **115**
6.1 Models for spatially distributed sources . . . 117
6.2 Localization . . . 120
6.2.1 A generalization of the MUSIC algorithm . . . 121
6.2.2 The coherently distributed source localizer . . . 126
6.2.3 The incoherently distributed source localizer . . . 127
6.3 The array gain . . . 130
6.3.1 Coherently distributed sources . . . 131
6.3.2 Incoherently distributed sources . . . 133
6.4 The uniqueness problem . . . 135
6.4.1 Coherently distributed sources . . . 136
6.4.2 Incoherently distributed sources . . . 138
6.5 Simulation results and performance comparison . . . 141

7 Summary and Conclusions **147**
7.1 Future work . . . 149

A Uniqueness of the Subspace Decomposition **151**

B Consistency of the Estimator **152**

C Singular Values of a Matrix Product **155**

D Minimization of the Subspace Fitting Error **157**

E Unbiased Condition for a Single Source **159**

List of Figures

1.1 Frequency-wavenumber spectrum 6

2.1 Ambiguity function 16
2.2 Frequency-wavenumber domain 18
2.3 An array processor 19
2.4 Some array configurations 20
2.5 A single planar wavefront 21
2.6 Spatial sampling by a uniform linear array 22
2.7 A delay-and-sum beamformer 24
2.8 The beampattern of a uniform linear array of 8 sensors. 26
2.9 Broad-band beamformer 28
2.10 The generalized sidelobe canceler 30
2.11 An array of matched doublets 37

3.1 The averaged second term of the PSC algorithm for a linear array of 8 sensors exposed to 3 uncorrelated far-field planar wavefronts at 10, 15, and 20 degrees. Each curve represents a different model order (orders, 1, 2, 3, 4). 60
3.2 The probability of resolution for the PSC and MDL algorithms using an array of eight sensors exposed to two uncorrelated far field signals. 67
3.3 The PSC and the MDL criteria for a single source arriving from 10 degrees at a uniform linear array of eight sensors. 69

3.4 Detected number of signals using the PSC and the MDL algorithms versus the number of samples. Three sources are at 10, 15, and 30 degrees and arrive at a uniform linear array of 12 sources. The source at 30 degrees is turned off after 50 samples. 70

4.1 Relationship between frequency and the singular values of the location matrix 79

4.2 The summation of the singular values of $\mathbf{A}_0\mathbf{A}_j^H$ normalized with respect to the norm of the vector $\mu = (\mu_1, \ldots, \mu_J)$ as a function of the number of frequency samples for a configuration with two wideband sources with a 40% relative bandwidth arriving from 10 and 14 degrees at a uniform linear array of 8 sensors. 83

4.3 The subspace fitting error as a function of the focusing frequency for two uncorrelated far-field wideband sources arriving at a linear array of 8 sensors with a 20 dB SNR. 84

4.4 The norm of the bias vector for the DOA estimation of two uncorrelated far-field wideband sources arriving at a linear array of 8 sensors with a 20 dB SNR. 85

4.5 The mean square error of the CSM algorithm for the two focusing frequencies 1 and 1.1 for two uncorrelated far-field wideband sources at 10 and 14 degrees arriving at a linear array of 8 sensors. 86

5.1 The probability of resolution for two closely separated sources using the TCT and CSM algorithms. 110

5.2 Comparison of the resolution of the two algorithms, TCT and CSM. The resolution criterion is defined as the difference between the average of the spatial spectrum at the peak points in the MUSIC algorithm and the spatial spectrum in the valley. 111

5.3 The MUSIC spatial spectrum for the two methods, TCT and CSM. 111

5.4 The MUSIC spatial spectrum for four sources at 8°, 13°, 33°, and 37°. The source at 13° is a delayed version of the source at 8°. 113

5.5 The norm of the averaged bias vector for limited number of observations (40 snapshots of 32 samples) versus SNR for a configuration with four uncorrelated sources. 114
5.6 The norm of the averaged variance vector for limited number of observations (40 snapshots of 32 samples) versus SNR for a configuration with 4 uncorrelated sources. 114

6.1 Geometry of the linear operator and the adjoint operator. 123
6.2 The spatial cross-correlation for a uniform CD source. (The first sensor is placed at the phase reference point.) 132
6.3 The array gain for a uniform CD source for different extension widths Δ in degrees. 134
6.4 The array gain for a uniform ID source for different extension widths Δ in degrees. 135
6.5 The probability of resolution for the conventional MUSIC and the DSPE versus SNR. 142
6.6 The bias of estimation versus SNR for the source at 10 degrees. 143
6.7 The bias of estimation versus SNR for the source at 13 degrees. 144
6.8 The standard deviation versus SNR for the source at 10 degrees. 144
6.9 The standard deviation versus SNR for the source at 13 degrees. 145
6.10 The eigenvalues of a configuration with two SD signals. 145
6.11 The spectrum of the DSPE algorithm. 146

List of Tables

3.1 Detection capability of the MDL and PSC algorithms for a configuration with two uncorrelated sources for 100 independent trials. 68

3.2 Detection capability of the MDL and PSC algorithms for a configuration with two coherent sources for 100 independent trials. 68

4.1 The resolution capability of the CSM algorithm for two closely spaced noncoherent wideband signals at 10 and 14 degrees arriving at a uniform linear array of 8 sensors, versus the focusing frequency. The resolution capability is measured in terms of the percentage of detection. 86

4.2 The norm of the bias vector for two closely spaced noncoherent wideband signals at 10 and 14 degrees arriving at a uniform linear array of 8 sensors, versus the focusing frequency. 87

5.1 The eigenvalues of the correlation matrices $\mathbf{R}_{\mathrm{CSM}}$ and $\mathbf{R}_{\mathrm{TCT}}$ for a configuration of 4 wideband sources. The sources have 40% relative bandwidth and are arriving from the angles 8°, 13°, 33°, and 37° degrees at a linear array of 16 sensors in a noise-free environment. 102

5.2 The estimation results for the first example. 109

5.3 The estimation results for the second example. The blank entries mean that CSM did not resolve the DOA. 112

5.4 The estimation results for the second example after two iterations. 112

Chapter 1

Introduction

We live in a society which places much emphasis on information. Today, information is the fuel of business and also increasingly important for leisure activities. The revolutionary advances in computer and telecommunications technology have spawned this information age.

Information is conveyed through signals. A *signal* is an information carrying vehicle of a physical system. It depends on time and space and assigns a unique value to each time instant and each spatial point. The system that generates a signal is called the *source*. For instance, a radio station is a source of information. The processing of signals is used in radar, sonar, seismology, communications, speech and image transmission, biomedical engineering, data compression, and so on.

In this thesis, we discuss the use of arrays of sensors to acquire signals. In general, the term *array* is denoted to any regularly ordered or arranged set. Here, by an array we mean *a sensory device that collects signals at different spatial coordinates*. This definition is complete in the sense that it includes both discrete and continuous arrays. A discrete array is a combination of two or more distinct sensors located in space. As an example, our eyes form an array of two sensors. Using two eyes, we can see objects in three dimensions and can estimate their distance. Note that in array processing signals are sent or received at different spatial positions. If we close one eye, the image of the objects will become 2-dimensional and the range information will be lost.

In the past, array processing has been restricted to applications that are not cost-sensitive. With DSP processing capabilities increasing, array processing will become more common in a wide variety of applications. The benefit of using array processing techniques is that they exploit the spatial characteristics of signals for information extraction.

1.1. Array processing

From a statistical point of view signals can be categorized as having a *deterministic* or a *random* waveshape. By a deterministic waveshape we mean a known waveform with possibly some unknown parameters such as amplitude, delay, and/or scaling. Deterministic waveshapes are encountered in applications such as radar, active sonar, and data communications where the waveform of the transmitted signal is known to the receiver. In many applications, the transmitted signal is unknown or is affected by a transmission environment whose exact impulse response is either unknown or changes with time and space. In such a case, the waveshape is random. For a random waveshape, all the information that can be obtained is contained in the probability distribution function.

In practice the signals are perturbed by noise or interferences. Noise and interferences are destructive signals that mix with the useful signal and change its properties or obscure it. Interference usually has the same characteristics as the signal and might be generated with a similar source. Electromagnetic coupling between two adjacent wires, diffusion of the energy of one channel into the neighboring channel in a multiplex communication, simultaneous pick up of two or more stations by a single radio receiver, the results of multipath transmission, and smart jamming can be counted as examples of interference. The destructive nature of interference is due to the fact that it can be confused with the useful signal.

Noise is usually generated by phenomena that are independent from the signal. In array processing, the wavefields that exist in space and are independent from the signal can be modeled as a spatial noise-field. If the noise-field is generated by many uncorrelated random waves propagating in all directions, a *spherically isotropic noise* is formed. It is usually assumed that the noise-field is stationary in time and space.

The information may also be conveyed by multi-dimensional signals. For instance, in communication over a fading channel, a multi-dimensional signal is created using time, frequency, or spatial diversity. In array processing, the spatial field is sampled by multiple sensors. The collection of signals at different channels form a multi-dimensional signal. Often, a multi-dimensional signal has a common information along each dimension. The combination of the information at different dimensions can improve the performance of processing. For instance, diversity in communication reduces the reconstruction error, and using a beamformer in array processing increases the signal-to-noise ratio (SNR).

The objectives of an array processor can be categorized as *signal enhancement* or *field characterization*. Signal enhancement involves improving the signal-to-noise ratio at the array output beyond that for a single sensor. This can be performed by steering a beam towards the direction of source and/or inserting nulls in the beampattern in the direction of noise and interference. A conventional method of beam steering is to place delay elements at the output of the sensors and compute a weighted sum of the delayed outputs. With a proper selection of the delays, the signals arriving from a specific direction will appear with the same phase at the output of the delay elements. This is termed *beamforming*. If a source is located in the direction of the beam, the signal power at the array output can be increased by the square magnitude of the number of sensors. For an uncorrelated intersensor noise, the noise power at the beamformer output increases linearly with the number of sensors. Thus, the SNR can be increased by using a conventional beamformer.

Field characterization is used to estimate the spatial properties of the sources such as their range, azimuth, elevation, velocity, and direction of movement. Field characterization is performed in two steps. First, the number of emitting sources is determined. This step is called *detection*. Detection is followed by a *localization* step which is the estimation of the signal position in space. For stationary signals, the spatial parameters of source are range, azimuth, and elevation. If the sources are located in the far field of the sensors, the arriving wavefields are planar and only azimuth and elevation can be estimated. Furthermore, it is frequently assumed that the sources and the array are in the same plane. In such a case, the direction-of-arrival (DOA) is the only spatial parameter of an emitting source. Many

array processing methods have been developed for DOA estimation. In most cases, it is straightforward to generalize these methods to azimuth and elevation estimation.

Many present and potential canonical applications can be found where an array of sensors is used to detect and localize sources or to enhance the signal-to-noise ratio. Consider an array of antennas monitoring the air traffic around an airport. Each airplane sends a narrowband signal that is received by an antenna array. Location of the airplanes can be estimated by using narrowband array processing techniques. Another example of narrowband array processing can be found in wireless communications. A moving transmitter in mobile or indoor communications emits narrowband signals. The receiver consists of an array of antennas. The array receives the original signal from the source and its reflections from the surrounding objects. If the location of the source is estimated, the antenna array can steer a beam towards the direction of the source reducing the effect of the reflected wavefronts. In a transmission mode, the antenna array sends the power only in the direction of the source by forming a steered beam. This way, energy is conserved and since the power is only transmitted in a certain direction it has a smaller interfering effect on other receivers.

Array processing techniques can also be applied to wideband signals. For a wideband signal the frequency bandwidth is relatively large compared to the center frequency. For instance, an array of microphones can be used to localize a speaker in a room. The arriving signal at the array is a wideband speech signal along with its reflections from the walls or any reflecting object. The effect of the reflections, which interferes with the direct signal, can be compensated by steering a beam towards the direction of the speaker. A microphone array can also be used inside a car for hands-free mobile communications. For hands-free communications, the voice of the driver is collected by an array of microphones. The noise of the car environment can also be reduced by forming a beampattern with nulls in the noise direction.

Another example of wideband array processing is passive sonar. A passive sonar system consists of a set of hydrophones that listen to the undersea sound environment. If a ship is located within the detection range of a passive sonar, the sound of the propeller can

be sensed by the sonar and the ship will be detected. The sound from the propeller is a wideband signal.

So far we have considered the sources that are highly localized in space (point sources). In some situations, point source modeling is not appropriate. An example where point source assumption does not hold can be found in the measurements of the seabed by using a multibeam echo sounder. A multibeam echo sounder is an active device that is used to map the seabed. A sound signal is transmitted towards the seabed. The depth of the sea floor is estimated by measuring the reflected echos. Reflection of the signal from the seabed and penetration into the lower layers of the sea floor creates a spatially distributed source. A similar effect can be observed in seismology. In seismology, the layers of the earth are examined by sending sound signals into the earth. The sound is reflected from different layers. Reflection of the wave from different layers in the earth creates a distributed source. The received signal in a communication link that uses the ionosphere or the troposphere as a medium for signal scattering is also observed as a spatially distributed source.

A signal which is a function of time and space is called a *spatiotemporal* signal. For a spatiotemporal signal a multi-dimensional Fourier transform can be defined which operates on time and space variables. The Fourier transform (spectrum) of a spatiotemporal signal is a function of frequency and wavenumber. For random signals, the Fourier transform is applied to the correlation function to give the power spectral density of the signal.

In Fig. 1.1 some particular examples of the spectrum of spatiotemporal signals are depicted. Signal A in this figure corresponds to a narrowband signal which is highly localized in space (point source). The spectrum of Signal A is the product of two impulses in the frequency and wavenumber axes. Many array processing methods have been developed for this model. Narrowband point source modeling is an appropriate assumption in some applications. For instance, an aircraft which is transmitting a narrowband signal and is located far from the observation point can be modeled as a narrowband point source.

In some cases, the arriving wavefronts are generated by wideband sources. Microphone array and passive sonar are two applications of wideband array processing. The frequency-wavenumber spectrum of a point wideband signal takes values on a line paral-

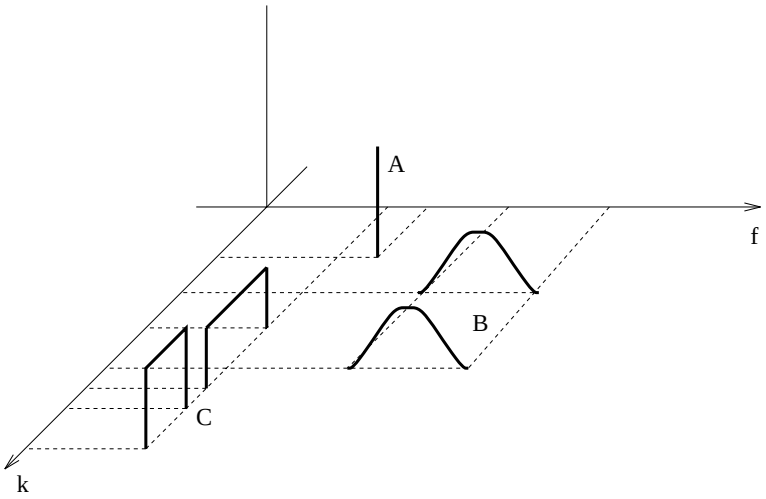

Fig. 1.1 Frequency-wavenumber spectrum

lel to the frequency axis. An example of two wideband signals separated in space, each with a bell-shaped frequency spectrum and the same bandwidth, is depicted as Signal B in Fig. 1.1. These two signals differ in wavenumber. The spatiotemporal processing of wideband signals can be performed by using a wideband array processor.

A third case is when the spectrum of signal takes on values on a line parallel to the wavenumber axis. This case corresponds to a spatially distributed narrowband source. Signal C in Fig. 1.1 consists of two narrowband spatially distributed sources with a flat distribution pattern. The detection and estimation methods that are developed for narrowband point source models usually do not operate well in distributed signal cases.

In general, the spectrum of a spatiotemporal signal extends in both frequency and wavenumber coordinates. However, many applications fall into the three special cases that we cited as Signals A, B, and C in Fig. 1.1. In this thesis, we will only discuss these three types of signal.

A spatiotemporal signal can be localized by spatiotemporal filtering. An array of sensors with delay elements added at the output of the sensors can be used to implement a spatiotemporal filter. The characteristics of a spatiotemporal signal reveal a duality between time and space. All the parameters that are related to temporal behavior of a signal have duals in spatial domain. For instance, the concept of wavelength and wavenumber

in spatial domain are duals of period and frequency in temporal domain. With this duality, the processing of signal using a uniform linear array of sensors is similar to finite impulse response (FIR) filtering for temporal signals. In such an analogy, the time delay for an FIR filter is replaced with the traveling time of the wavefront between two adjacent sensors of the array. A spatiotemporal filter is designed so that it generates a desired frequency-wavenumber spectrum.

1.2. Contributions

This dissertation contributes several new techniques for detection and parameter estimation. In this section, a brief introduction to the methods is presented. Here, the aim is to discuss the objectives of this study and the approaches taken. We will only cite the open problems that were solved in this dissertation. A full discussion on the methods will be presented later.

In Chapter 3, a new detection method for point narrowband signals (Signal A in Fig. 1.1) is developed. Any detection procedure chooses between two or more hypotheses that are possible representative models for the data. The role of the detector is to select the best model based on some criterion. In array processing, detection is the process of determining the number of signals that generate the wavefield. For instance, in an antenna array application, the number of airplanes in the vicinity of an airport is the number of sources. One class of detection methods computes an *information theoretic criterion*. The information theoretic criteria are developed based on minimizing a distance (Kullback-Leibler distance) between the observed data and the hypothesized model. The method that is developed in this dissertation is also an information theoretic technique. Here, the concept of *stochastic complexity* is used to determine the number of signals arriving at an array of sensors. Stochastic complexity is the codelength of data when coded with respect to a given generating class (probability density function). The codelength minimization is an appropriate criterion for model selection since it is directly related to the amount of information that can be obtained for data when it is conditioned on the given model. If a model gives more information about the data, it results in a smaller codelength. In the

new algorithm, the data is coded with respect to each competing model considered as the generating class. The selected model is the one that gives the smallest codelength for the data.

One objective of introducing a new method for signal detection was to develop an iterative technique that can be suitable for on-line use. The information theoretic detection methods that exist in the literature operate on a batch of data. In these methods, the detection process must be delayed until the whole batch of data is observed. In the new method, detection is performed with each new observation and the model selection is updated at each sample time. The dynamic structure of the new technique permits changes in the generating class. For instance, if at some time in the window of observations the signal of a source is turned off, the new method is able to detect a change in the number of sources.

It was mentioned earlier that in some applications the arriving wavefronts at the array are generated by wideband point signals. If the frequency contents of the signals do not overlap, they can be separated by using bandpass filters. Difficulty arises when the signals have the same frequency band. For such signals, wideband array processing methods are used for localization of sources. There are different approaches to wideband array processing.

In some wideband methods, the frequency spectrum of the wideband signals is sampled to generate narrowband signals. These narrowband signals have the same DOA. The frequency of each signal depends on the sampling bin. Since the frequency of processing for these narrowband signals are different, each narrowband signal belongs to a coordinate system that is different from the coordinates of the other signals. In a coherent approach, the data at different frequency bins are transformed (focussed) into a unique coordinate system by using some transformation matrices. The shortcoming of the well-known coherent signal-subspace processing method (CSM) is that it suffers from an asymptotic bias of estimation. In this dissertation, we show that unbiased estimation of the DOA is not possible using a CSM algorithm. However, with a proper selection of the focusing coordinates the bias of DOA estimation can be minimized.

Some alternative coherent methods are evolved from the CSM algorithm that asymptotically can generate unbiased estimates of the DOAs. These methods however require an increase in the computational complexity. In Chapter 5, we develop a new algorithm for coherent wideband signal localization. The new method is evolved from the CSM algorithm and uses a similar focusing approach. The focusing matrices in the new algorithm are determined by using the correlation matrices of the array output. We show that by applying the new algorithm unbiased estimates of the DOA can be asymptotically obtained. Furthermore, the new method shows a modest increase in the computational complexity over that of the CSM algorithm.

Another contribution of this dissertation is a method for distributed source localization. In many applications, point source modeling is not an appropriate assumption. For instance in troposcatter communications the scattered waves from the troposphere appear as a spatially distributed source. Previous work in the literature was based on the approximation of a distributed source by a combination of many point sources located next to each other with small spacing between them. The algorithms that are developed for point source modeling can lead to erroneous results when applied to distributed sources. Point source approximation of a distributed source has problems in determining a unique solution for the localization.

In this dissertation, the output of an array exposed to a distributed source is modeled as an integral of the response of the array to a point source taken over some interval in space. It is further assumed that the distribution pattern of the source belongs to a class of parametric functions. With this assumption, a limited number of sensors can give a unique solution to the localization problem of distributed sources. The model and the results for distributed sources are, to the best of our knowledge, new.

1.3. Thesis organization

The dissertation is organized as follows. In the following chapter some basic array processing techniques are reviewed. In Chapter 3, a new detection method based on the concept of stochastic complexity is developed. In Chapter 4, an optimal method for focusing in

wideband array processing is proposed and it is shown that by using this method the bias of estimation can be reduced. In Chapter 5, a new algorithm for localization of wideband sources is developed. The new method does not suffer from asymptotic bias of estimates. In Chapter 6, a high resolution method for localization of distributed sources is introduced. The techniques developed in this dissertation have been reported in conference proceedings or submitted to journals for publication [38] [39] [40] [41] [42] [43] [44] [45] [46] [47] [48].

Chapter 2

Array Processing Techniques

Array processing has been the subject of active research for more than three decades. Many methods have been developed for processing of signals by using an array of sensors. Mathematically, the localization of narrowband sources in array processing is related to harmonic retrieval and frequency estimation. This relationship has opened the rich literature in spectrum estimation to array processing. In this chapter, we briefly review some important methods in array processing.

Two approaches can be taken for spectrum estimation: parametric methods, and non-parametric methods. In a parametric approach, a model with a fixed number of parameters is assigned to the observation and the parameters of the model are selected so that the observed data fits the model. In a non-parametric method, no model is imposed upon the data. Analysis of a system by using Fourier transformation is a nonparametric approach. Any parametric method that is applied to a system identification problem is directly related to the modeling of that system. If a model of the system is available, parametric methods are preferred. This dissertation considers parametric methods. For these methods, a careful modeling of the arriving signals is essential for good performance.

The signals can be modeled with respect to temporal or spatial characteristics. For instance, from a temporal point of view waveforms can be deterministic or stochastic. The signals can also have wideband or narrowband spectrum, or they might have some other structures such as being the realizations of AR or ARMA processes, and so on.

In this dissertation, we consider stochastic signals with temporally uncorrelated samples. Although this assumption is made for the signals, we will also review an optimal estimation method for deterministic signals. It should be noted that in categorizing the signal as deterministic or stochastic we follow the terminology that is established in the literature. This terminology can be slightly misleading. A signal with deterministic waveform can also be stochastic. An alternative terminology might be as deterministic or non-deterministic [8], or as conditional or unconditional [34].

Spatial signals can be modeled as point or distributed sources. Most techniques in array processing use point source modeling. In some situations point source assumption is not realistic. In a tropospheric scatter link the signal is transmitted between two locations through scattering from the troposphere. The signal arriving at the receiver antenna is observed as a distributed source. In this application the rays of the signal at different angles are the scattered wavefronts and have random phase and amplitude. Thus, the rays at two different angles are uncorrelated. This is an incoherently distributed source. In some applications the rays of the signal at different angles can be completely correlated. Such a signal is coherently distributed. There are not many methods which assume distributed source modeling. One approximation to a distributed signal is to assume that the distributed source is decomposed to many closely placed point sources. This approximation has some drawbacks that we will discuss later.

An array processing is usually performed in two steps: detection and localization. Detection is a terminology used for the procedure that determines the number of signals arriving at the array. Localization (sometimes called estimation) is a process to estimate the spatial parameters of the signals such their DOAs. The methods that are used for detection and localization might be categorized as beamforming or subspace decomposition techniques. The beamforming techniques have an easier implementation and are widely used in practice. The resolution of these methods is lower than for the subspace decomposition methods. We will review delay-and-sum, minimum variance, and linear prediction beamforming methods. For subspace decomposition techniques we will only consider information theoretic criteria for signal detection and the MUSIC and the ESPRIT algorithms

for localization.

2.1. Time-space transformation

For a temporal signal, information is conveyed by variation of the signal in time. For such a signal, the conventional Fourier transformation yields

$$S(\omega) = \int_{-\infty}^{\infty} e^{-j\omega t} s(t) dt \tag{2.1}$$

where $s(t)$ is the signal and ω is the processing radian frequency. The inverse Fourier transform is given by

$$s(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{j\omega t} S(\omega) d\omega. \tag{2.2}$$

The equations (2.1) and (2.2) are called a Fourier transform pair which are widely used in signal processing.

In general, a signal can also be a function of space. A spatiotemporal signal is represented by $s(\mathbf{r}, t)$ where $\mathbf{r}$ is the spatial parameter and t is time. A signal which is a function of space and time can be Fourier transformed by

$$S(\omega, \mathbf{k}) = \int\int\int\int s(t, \mathbf{r})\ e^{-j(\omega t - \mathbf{k}.\mathbf{r})}\ dt\ dx\ dy\ dz \tag{2.3}$$

where the x, y, and z are the components of the spatial vector $\mathbf{r}$, and $\mathbf{k}$ is the wavenumber vector defined as

$$\mathbf{k} = \frac{\omega}{c} \hat{\mathbf{k}} \tag{2.4}$$

where c is the wave velocity and $\hat{\mathbf{k}}$ is a unit vector normal to the wavefront and pointed to the direction of propagation. The inverse Fourier transform is given by

$$s(t, \mathbf{r}) = \frac{1}{(2\pi)^4} \int\int\int\int S(\omega, \mathbf{k})\ e^{j(\omega t - \mathbf{k}.\mathbf{r})}\ d\omega\ dk_x\ dk_y\ dk_z \tag{2.5}$$

where k_x, k_y, and k_z are the components of $\mathbf{k}$ and are called the *spatial frequencies* or *wavenumbers*.

Fourier transform pair (2.3) and (2.5) suggest a duality between time and space and their corresponding parameters. In fact the time and frequency are the duals of space and wavenumber. The length of the wavenumber vector is the dual of the radian frequency; the period of a temporal signal is the dual of the wavelength of a wavefront; and so on. This can be shown as

$$t \longleftrightarrow \omega = \frac{2\pi}{T} \tag{2.6}$$

$$|\mathbf{r}| \longleftrightarrow |\mathbf{k}| = \frac{2\pi}{\lambda} \tag{2.7}$$

where T is the period of the signal, λ is the wavelength and $|.|$ is the length of vector. The filtering of a temporal signal in time domain can also have a dual for spatial signals. The filtering of a spatial signal is performed by using an array of sensors. Array processing for spatial signals has the same effect of FIR filtering for temporal signals. See [51] for a further discussion of the similarities between array processing and FIR filtering.

2.2. Deterministic signals

In some applications the waveform of the received signal is known and only some parameters such as delay and amplitude are to be estimated. The modeling of such signals is based on a deterministic waveform assumption. A precise definition for deterministic signals can be found in [8] and is related to whether the signal can be predicted from its past with zero error. Radar and active sonar are two applications in which the deterministic signal modeling is employed. An optimal processing method for detection and parameter estimation of deterministic signals is by *matched-filtering*. The impulse response of a matched-filter is the complex conjugate of the received signal. A matched-filter maximizes the signal-to-noise ratio [49].

In radar or active sonar a short pulse is emitted. The reflected signal from the target is then received by an antenna or an array of sensors. From the delay between the transmitted pulse and the received waveform the range of the target can be estimated. Assume that the signal $s(t), 0 \leq t \leq T$ is transmitted. The received echo is $x(t) = \alpha(r)s(t-2r/c)$, where

r is the range of the target, c is the wave velocity, and $\alpha(r)$ is an attenuation factor due to propagation and reflection. For a moving target with a radial velocity v, $r = r_0 + vt$ and $x(t) = \alpha(r_0 + vt)s((1 - 2v/c)t - 2r_0/c)$. The variation of the attenuation factor $\alpha(r_0 + vt)$ with distance is usually very slow. Thus it can be approximated by $\alpha(r_0)$. Ignoring the attenuation factor, the output of the matched-filter can be shown as [21]

$$l(\tau, \beta) = \int_0^T \tilde{s}((1 - \beta)t)\tilde{s}^*(t + \tau)dt \tag{2.8}$$

where $\tilde{s}(t)$ is the complex envelope of the bandpass signal $s(t)$, $\beta = 2v/c$, $\tau = 2r_0/c$, and $*$ denotes complex conjugation. Usually the target velocity is much smaller than the wave velocity, which yields $\beta \ll 1$. In such a case, it can be shown that

$$l(\tau, \beta) = \int_0^T \tilde{s}(t)\tilde{s}^*(t + \tau)e^{-j2\pi\beta f_0 t}dt \tag{2.9}$$

where f_0 is the center frequency of the spectrum of the transmitted signal. The square value of the matched-filter output is termed the *ambiguity function* and is denoted by

$$c(\tau, \omega) = \left| \int_0^T \tilde{s}(t)\tilde{s}^*(t + \tau)e^{-j\omega t}dt \right|^2 . \tag{2.10}$$

The ambiguity function can be used for signal detection and range and radial velocity estimation. A signal is detected if for some values of τ and ω the ambiguity function is larger than a threshold. The range and the velocity of the target is estimated from τ and ω. A typical ambiguity function is depicted in Fig. 2.1. The name of the ambiguity function comes from the fact that all the values of τ and ω in the dashed region in Fig. 2.1 can be estimates of the delay and the frequency. It can be shown that the volume under the ambiguity function is constant. Thus, reducing the main lobe will increase the sidelobes. The transmitted signal can be selected to trade off the width of the mainlobe with the amplitude of the sidelobes [21].

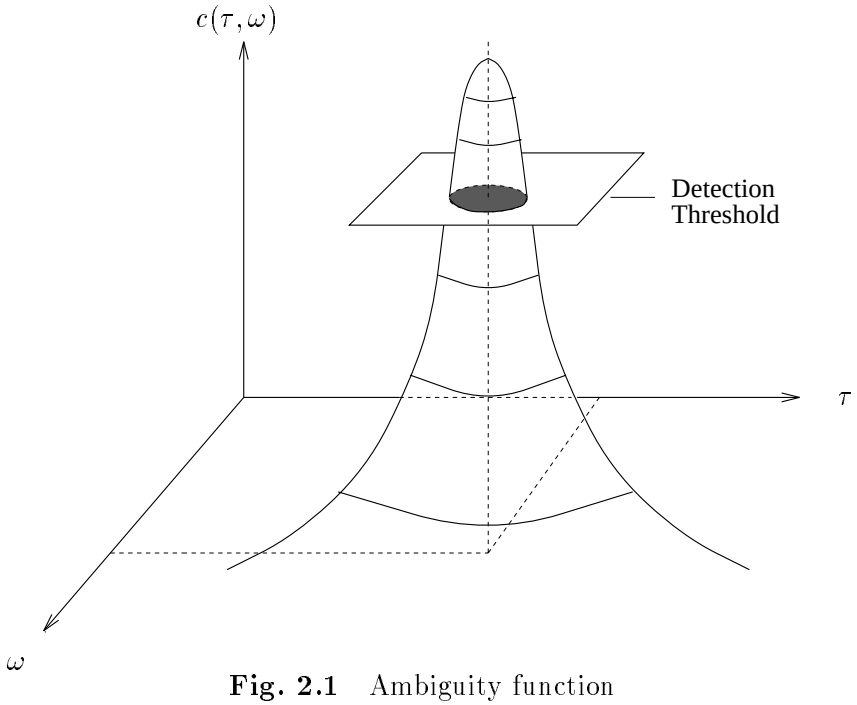

Fig. 2.1 Ambiguity function

2.3. Stochastic signals

In many applications the received signals are random in nature or are perturbed by a random medium. For such signals matched-filtering is not applicable. Stochastic signals are modeled as the realizations of stochastic sources. The modeling of a stochastic signal depends on the application in which the signal is used. Usually it is assumed that the signal samples have a Gaussian probability distribution function. This is an appropriate assumption since in many cases the random signal is generated by combination of a large number of independent sources. The central limit theorem indicates that the distribution of such a signal is Gaussian. For Gaussian signals the second moment is a sufficient statistic and the higher moments are either zero or can be determined by the second moment.

The correlation function of a random signal is defined as

$$R_{ss}(t_1, t_2, \mathbf{r}_1, \mathbf{r}_2) = E\{s^*(t_1, \mathbf{r}_1)s(t_2, \mathbf{r}_2)\}. \tag{2.11}$$

This is a function of eight variables, two time instants and six coordinates. If the signal is homogeneous in time, the correlation is a function of time difference $\tau = t_1 - t_2$. The *cross spectral density* for such a signal is defined as

$$S_{ss}(\omega, \mathbf{r}_1, \mathbf{r}_2) = \int R_{ss}(t_1, t_2, \mathbf{r}_1, \mathbf{r}_2) e^{-j\omega\tau} d\tau. \tag{2.12}$$

Signals can also be homogeneous in space. For spatial homogeneity the correlation function is shown as $R_{ss}(\tau, \mathbf{r})$ where $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$. For these signals, the frequency-wavenumber spectrum is defined as

$$P_{ss}(\omega, \mathbf{k}) = \int\int\int\int R_{ss}(\tau, \mathbf{r}) e^{-j(\omega\tau - \mathbf{k}.\mathbf{r})} d\tau dx dy dz \tag{2.13}$$

which is the spatiotemporal Fourier transform of the correlation function. The frequency-wavenumber spectrum (from now on spectrum) is the transformation of the time-space parameters of signal into the frequency-wavenumber domain. Some special cases of interest are as follows.

For a narrowband point source the spectrum of a signal can be represented by

$$P_{ss}(\omega, \mathbf{k}) = \delta(\omega - \omega_0)\delta(\mathbf{k} - \mathbf{k}_0) \tag{2.14}$$

where $\delta(.)$ is the Dirac delta. In the frequency-wavenumber domain this is depicted by a dot such as point A in Fig. 2.2. Note that Fig. 2.2 is a symbolic representation of the frequency-wavenumber plane. In fact the wavenumber is a three dimensional vector.

In some applications the signals are generated by wideband point sources. The spectrum of these signals which can be represented by

$$P_{ss}(\omega, \mathbf{k}) = P(\omega)\delta(\mathbf{k} - \mathbf{k}_0) \tag{2.15}$$

takes values on lines parallel to the frequency axis. A typical example is shown as signal B in Fig. 2.2 which consists of two wideband sources with distinct wavenumbers and the same frequency bandwidth.

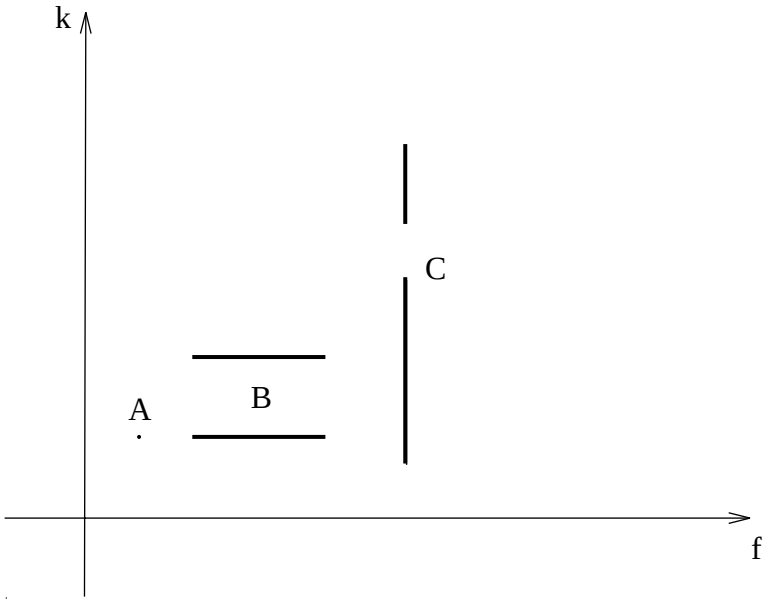

Fig. 2.2 Frequency-wavenumber domain

The signals can also be spatially distributed. For spatially distributed narrowband signals the wavenumber-frequency spectrum consists of lines parallel to the wavenumber axis. The spectrum of these signals can be decomposed as

$$P_{ss}(\omega, \mathbf{k}) = \delta(\omega - \omega_0) P(\mathbf{k}). \tag{2.16}$$

An example of two narrowband distributed signals with the same center frequency is depicted as signal C in Fig. 2.2.

Note that the concepts of narrowband/wideband-point/distributed sources also exist for deterministic signals. For deterministic signals the spatiotemporal Fourier transformation is applied to the signal waveform. A similar discussion can be used to categorize deterministic sources by their spatial characteristics. In this dissertation, we will only consider stochastic sources.

When two temporal signals have non-overlapping frequency contents, a bandpass filter can separate the signals. A similar processing can be carried out for spatial signals. If two signals have different wavenumbers it is possible to separate them by using spatial

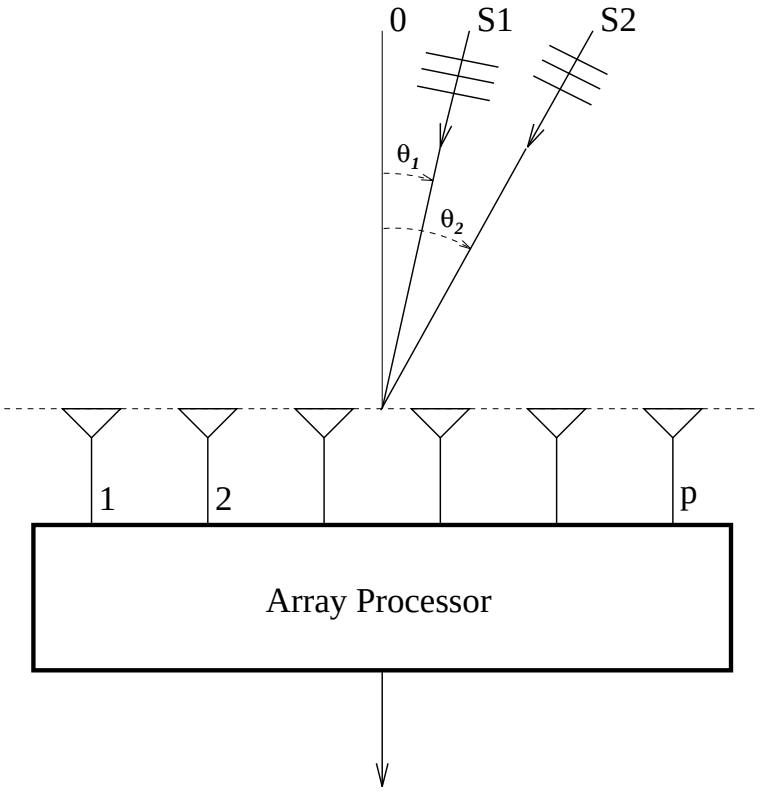

Fig. 2.3 An array processor

filters. The spatial filtering can be performed by an array of sensors. The phase and the amplitude of the received signal at each sensor is a function of the arriving angle of the emitting source and the relative location of that sensor with respect to a phase reference point. The output of the sensors can be processed to find the spatial position of the sources. The following section reviews some array processing techniques.

2.4. Array processing

Fig. 2.3 schematically depicts an array of p sensors exposed to two planar wavefronts. In general, the spatial arrangement of the array sensors is arbitrary. Some typical arrangements of the sensors are shown in Fig. 2.4. The methods that we will propose in this dissertation are applicable to general array geometry. However, for simplicity of implementation we have used uniformly spaced linear arrays in the computer simulations.

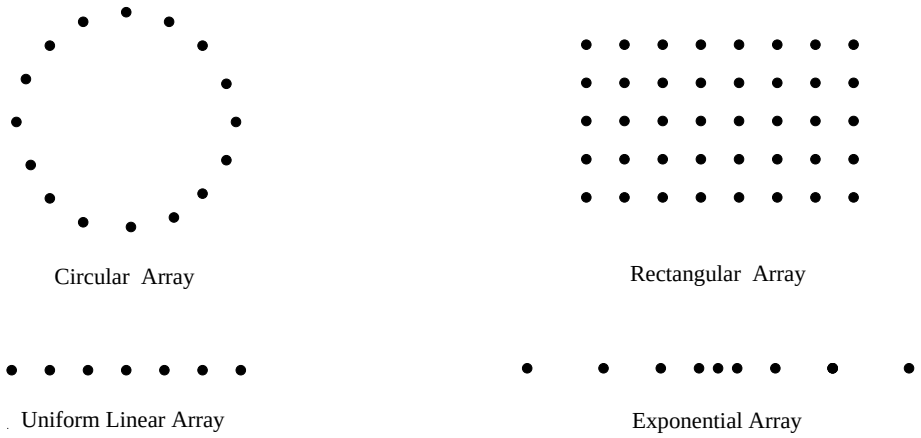

Fig. 2.4 Some array configurations

Throughout this dissertation we will assume that the transmission of the wave is done in the x-y plane. It is further assumed that the array of sensors is in the same plane as the sources. These assumptions are not restrictive and simply are made for ease of formulation. The results that are derived here can easily be extended to general case. With these assumptions, the only information about the position of a source is in the direction of arrival (DOA). The DOA is the angle between the direction of the propagation and the broadside of the array. Fig. 2.5 depicts a planar wave arriving from a DOA θ. The phase reference point is the origin of the coordinates. The received signal at point d is equal to $s(t + d\sin\theta/c)$. If the signal is narrowband, the complex envelope of the signal will approximately be $s(t)e^{j\omega d\sin\theta/c}$ [24].

The *location vector* of an array is defined as the frequency response of the array for a given DOA. For an array of p sensors, the location vector is a column vector with p components and is represented by $\mathbf{a}(\omega,\theta)$ where ω and θ are the frequency of processing and the DOA. The location vector is a function of the array geometry. The location vector of a uniform linear array with the phase reference taken at the first sensor is given by

$$\mathbf{a}(\omega_0,\theta) = [1 \; e^{j\frac{\omega_0 d}{c}\sin\theta} \; \ldots \; e^{j\frac{\omega_0 d}{c}(p-1)\sin\theta}]^T \tag{2.17}$$

where d is the distance between two consecutive sensors, c is the wave velocity, ω_0 is the radian frequency of the source signal, and T indicates the transpose of a vector. The

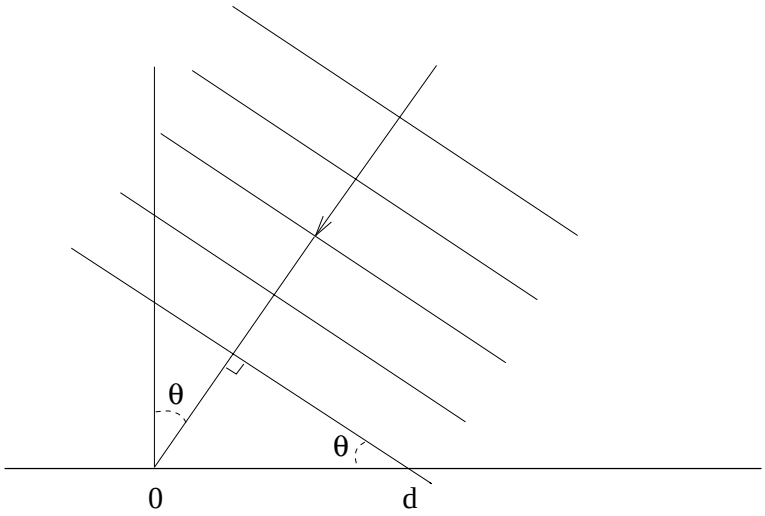

Fig. 2.5 A single planar wavefront

location vector of a uniform linear array can also be shown as

$$\mathbf{a}(\theta) = [1 \; e^{j\pi \frac{2d}{\lambda} \sin\theta} \; \ldots \; e^{j\pi \frac{2d}{\lambda}(p-1)\sin\theta}]^T \tag{2.18}$$

where λ is the wavelength.

In (2.18), if $2d > \lambda$ there exist θ_1 and θ_2 in $[-\frac{\pi}{2}, \frac{\pi}{2}]$ such that

$$e^{j\pi \frac{2d}{\lambda} \sin\theta_1} = e^{j\pi \frac{2d}{\lambda} \sin\theta_2}. \tag{2.19}$$

This is called *spatial aliasing*. In such a case, the signals that are located at θ_1 and θ_2 give the same array output. To clarify the terminology used for this phenomenon, consider a uniform linear array in the wavefield of a single narrowband far-field source arriving from the direction θ. The received wavefield at the line of the array is a sinusoidal wavefront with the wavelength $\lambda / \sin\theta$ (see Fig. 2.6). This wavefield is spatially sampled by the uniform linear array. Aliasing will occur if the sampling interval is larger than half the wavelength. Since the maximum value of $\sin\theta$ is one, to prevent spatial aliasing in a uniform linear array for all θ, the spacing between two consecutive sensors should be smaller than

$$d < \frac{\lambda}{2}. \tag{2.20}$$

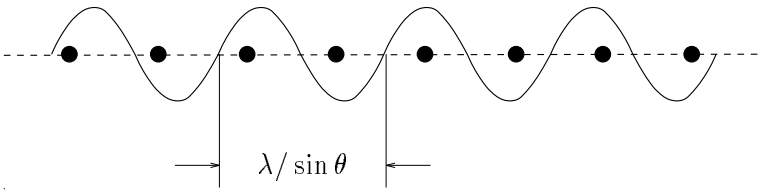

Fig. 2.6 Spatial sampling by a uniform linear array

Usually, it is assumed that d is equal to half the wavelength. In such a case the location vector of a uniform linear array is given by $\mathbf{a}(\theta) = [1 \; e^{j\pi \sin\theta} \; \ldots \; e^{j\pi(p-1)\sin\theta}]^T$.

If a spatiotemporal signal is observed through a continuous aperture (continuous array) along the x-axis, the output can be expressed as

$$r(x,t) = w(x)s(x,t) \tag{2.21}$$

where $w(x)$ is the weighting function of the aperture. The spectrum of the observed signal is given by

$$R(\omega, k_x) = \frac{1}{2\pi} \int W(k_x - k')S(\omega, k')dk' \tag{2.22}$$

where $W(.)$ is the spectrum of the aperture weighting function. For a uniform aperture extended in an interval with the length L,

$$w(x) = \begin{cases} 1 & |x| < \frac{L}{2} \\ 0 & \text{otherwise} \end{cases} \tag{2.23}$$

and the spectrum of the weighting function is

$$W(k_x) = \frac{\sin(k_x L/2)}{k_x/2}. \tag{2.24}$$

The first zero in the spectrum $W(k_x)$ occurs at

$$k_x = \frac{2\pi}{L}. \tag{2.25}$$

Two closely spaced sources can be resolved if their wavenumber is separated at least by $\frac{2\pi}{L}$. This is called *Rayliegh criterion*. Rayleigh criterion can be used to compare the resolution capabilities of two or more apertures. For a wavefront arriving form the DOA θ, $k_x = \frac{2\pi \sin\theta}{\lambda}$. Thus, for a uniform linear array with the half wavelength spacing between sensors, the Rayleigh criterion, based on the DOA, is given by

$$\theta = \sin^{-1}(\frac{2}{p-1}). \tag{2.26}$$

And for large values of p it can be approximated by

$$\theta \approx \frac{2}{p-1}. \tag{2.27}$$

As was depicted in Fig. 2.3 the output of the sensors is imported to an array processor. The output of the array processor depends on the application in which the array is being used. An array processor might be applied to signal enhancement. In such a case, a beamformer is used to create a spatial beam that is directed towards the source. The wavefront which is generated by the source is located in the beamwidth and can be received by the array. The signals arriving from the angles which is not covered by the beamwidth are received with a reduced amplitude. In the following section, some basic beamforming methods are reviewed.

2.4.1. Beamforming

Assume that an array of p sensors is receiving the wavefield of a single narrowband source. The output of the i-th sensor is represented by

$$x_i(t) = s(t - \tau_i) + n_i(t) \qquad \text{for} \quad i = 1, \ldots, p \tag{2.28}$$

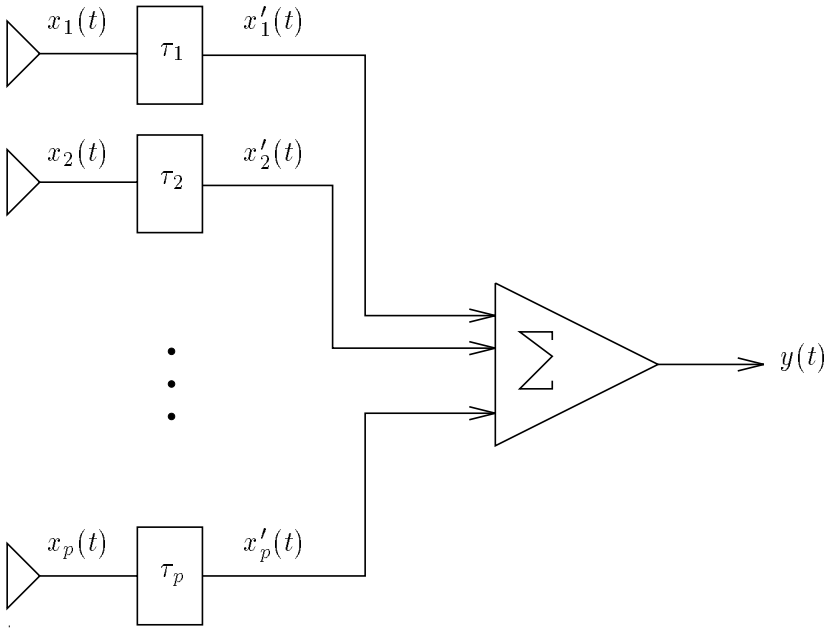

Fig. 2.7 A delay-and-sum beamformer

where τ_i is the delay of the signal at the i-th sensor relative to the phase reference point and $n_i(t)$ is the noise of the i-th sensor. For a uniform linear array with the phase reference at the first sensor and exposed to a source at θ degrees, the delays are given by $\tau_i = (i-1)(d/c)\sin\theta,\ i = 1, \ldots, p$, where d is the spacing between two consecutive sensors, and c is the velocity of wave. It is assumed that the noise components for different sensors are uncorrelated and have the same variance.

The simplest beamformer consists of a delay-and-sum configuration which is depicted in Fig. 2.7. The delay at each sensor depends on the location of that sensor and the signal direction. The output of sensors are delayed so that the signal components from the direct direction at the output of the delay elements have the same phase. If these outputs are shown by $x'_i(t)$, we have

$$x'_i(t) = s(t) + n_i(t + \tau_i) \qquad \text{for} \quad i = 1, \ldots, p. \tag{2.29}$$

The output of the adder is given by

$$y(t) = p\ s(t) + \sum_{i=1}^{p} n_i(t + \tau_i). \tag{2.30}$$

The SNR of the output can be computed as

$$\mathrm{SNR}_p = \frac{p^2 E|s(t)|^2}{E|\sum_{i=1}^{p} n_i(t+\tau_i)|^2}. \tag{2.31}$$

Since the noise components are uncorrelated and have the same power,

$$\mathrm{SNR}_p = \frac{p^2 E|s(t)|^2}{pE|n_i(t)|^2} \tag{2.32}$$

$$= p\ \mathrm{SNR}_1 \tag{2.33}$$

where SNR_1 is the signal-to-noise ratio at the output of a single sensor. It is seen that if the direction of source is known and the intersensor noise components are uncorrelated, a coherent addition of the array output can increase the SNR beyond the one of a single sensor. The *array gain* is defined as the ratio of the SNR at the output of the array to the SNR at the output of a single sensor [20],

$$G = \frac{\mathrm{SNR}_p}{\mathrm{SNR}_1}. \tag{2.34}$$

For the above example the array gain is p which is the same as the number of sensors. Thus, for a single narrowband point source the array gain can be linearly increased with the number of sensors. Later, we will see that linear increment of the array gain is not possible for spatially distributed sources.

The *beampattern* of a delay-and-sum beamformer is defined as

$$F(\theta) = \left| \frac{1}{p} \sum_{i=1}^{p} e^{j\omega\tau_i(\theta)} \right|^2 \tag{2.35}$$

where the summand is the Fourier transform of the delay elements. This is the normalized impulse response of the array to a single far-field source located at the broadside of the

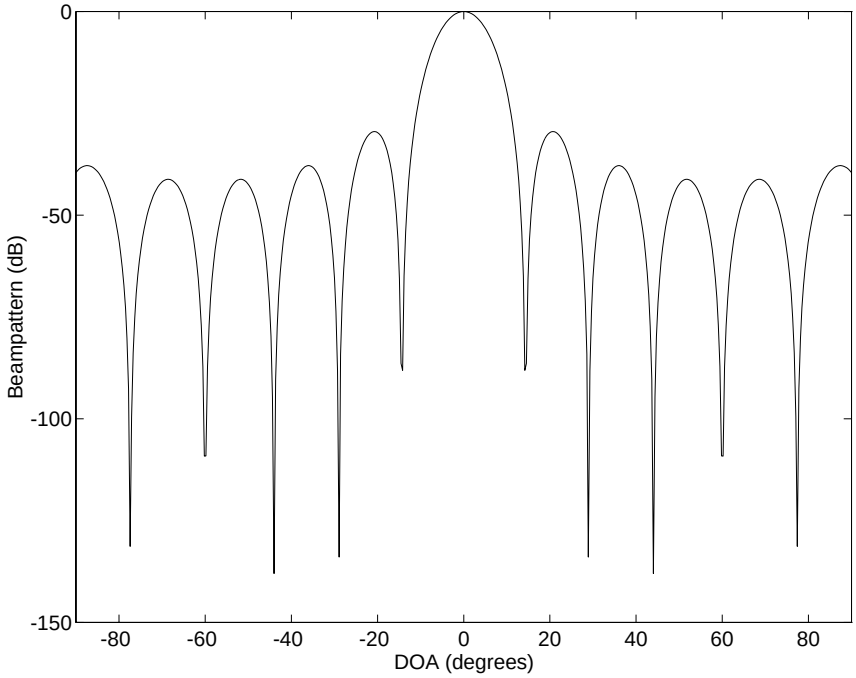

Fig. 2.8 The beampattern of a uniform linear array of 8 sensors.

array. For a uniform linear array with half wavelength spacing

$$F(\theta) = \left| \frac{1}{p} \sum_{i=1}^{p} e^{j\pi(i-1)\sin\theta} \right|^2 \tag{2.36}$$

$$= \left[\frac{\sin(\pi p \sin\theta/2)}{p \sin(\pi \sin\theta/2)} \right]^2 . \tag{2.37}$$

The beampattern of a uniform linear array with 8 sensors is plotted in Fig. 2.8. The first zero crossing point is at $\sin^{-1}(2/p)$. For large p this can be approximated by $2/p$ which is very close to Rayleigh angle (2.27). If some point sources are located at the nulls of the beampattern, they will be blocked and do not have any effect at the array output. This can be used to null out point noise source or interference. The beampattern can be designed so that it places nulls at the direction of these destructive noises.

The beamformer structure for wideband signals is depicted in Fig. 2.9. Since the spectrum of the wideband signals is spread in frequency, temporal filtering of the received data may be required to localize the sources. The output of each sensor is filtered by an

FIR filter with $J-1$ delay elements. The output of these filters are then added to form

$$y(t) = \sum_{i=1}^{p} \sum_{j=1}^{J} w_{ij} x_i(t-(j-1)T). \tag{2.38}$$

where $x_i(t-jT)$ is the output of the i-th sensor at the j-th delay element and w_{ij} is the corresponding weight. Let us define $\mathbf{x}(t) = [\alpha^T(t), \ldots, \alpha^T(t-(J-1)T)]^T$, where $\alpha(t) = [x_1(t), \ldots, x_p(t)]^T$, and $\mathbf{w} = [\beta_1^T, \ldots, \beta_J^T]^T$, where $\beta_j = [w_{1j}, \ldots, w_{pj}]^T$. Then the output of the array processor can be shown as

$$y(t) = \mathbf{w}^H \mathbf{x}(t) \tag{2.39}$$

where H indicates the Hermitian transpose. This is a general formula which also holds for narrowband beamformers. For narrowband sources, $\mathbf{x}(t)$ is a p-dimensional sensor output vector. The average power P at the array output is given by

$$P = E[|y(t)|^2] = \mathbf{w}^H E[\mathbf{x}(t)\mathbf{x}(t)^H]\mathbf{w} = \mathbf{w}^H \mathbf{R} \mathbf{w} \tag{2.40}$$

where $\mathbf{R}$ is the correlation matrix of the observation vector.

In the *minimum variance beamformer* [9] the array weights are selected so that the contribution of the noise and interference to the array output power is minimized. Assume that the array location vector for the desired source is given by $\mathbf{a}$. The weights are selected by

$$\begin{aligned} &\min_{\mathbf{w}} \ \mathbf{w}^H \mathbf{R} \mathbf{w} \\ &\text{s.t. } \mathbf{w}^H \mathbf{a} = 1. \end{aligned} \tag{2.41}$$

The constraint in (2.41) guarantees that the signal of the source will appear at the array output without degradation. The solution to (2.41) is given by

$$\mathbf{w} = \frac{\mathbf{R}^{-1}\mathbf{a}}{\mathbf{a}^H \mathbf{R}^{-1} \mathbf{a}}. \tag{2.42}$$

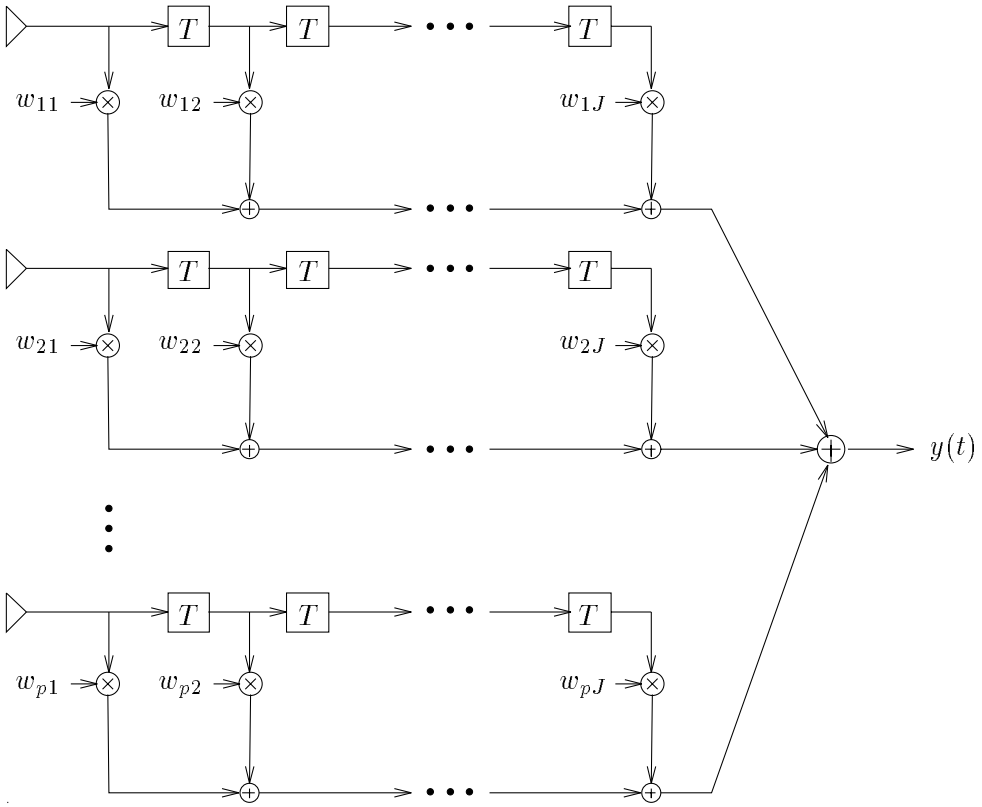

Fig. 2.9 Broad-band beamformer

Applying (2.42) in (2.40) yields

$$P = \frac{1}{\mathbf{a}^H \mathbf{R}^{-1} \mathbf{a}}. \tag{2.43}$$

The *resolution* of an array processer is the capability of localizing two closely spaced signals. The resolution of the minimum variance beamformer is higher than the conventional delay-and-sum beamformer.

We stated earlier that the beampattern can be designed so that it places nulls in the direction of noise or interference. This can be done by using a *linearly constrained minimum variance beamformer*. The approach is based on imposing constraints on the beampattern so that the signal passes through the beamformer and the noise and interference cancel out. The weights of the linearly constrained minimum variance beamformer are the solutions of

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R} \mathbf{w} \tag{2.44}$$

$$\text{s.t. } \mathbf{C}^H\mathbf{w} = \mathbf{f}$$

where $\mathbf{C}$ is the constraint matrix and $\mathbf{f}$ is the response vector. The weight vector that solves (2.44) is given by

$$\mathbf{w} = \mathbf{R}^{-1}\mathbf{C}[\mathbf{C}^H\mathbf{R}^{-1}\mathbf{C}]^{-1}\mathbf{f}. \tag{2.45}$$

Assuming L constraints imposed on weights, $\mathbf{C}$ is a $p \times L$ dimensional matrix. If $\mathbf{w}$ is a real vector, it is selected with $p - L$ degrees of freedom.

An alternative formulation of the linearly constrained minimum variance beamformer is by using a *generalized sidelobe canceler* [13]. Assume that the weight vector can be decomposed into two orthogonal components, $\mathbf{w} = \mathbf{w}_0 - \mathbf{v}_n$, where $\mathbf{w}_0$ and $\mathbf{v}_n$ are in the range of $\mathbf{C}$ and the null space of $\mathbf{C}^H$, respectively. Since $\mathbf{v}_n$ is in the null space of $\mathbf{C}^H$, for $\mathbf{w}$ satisfying the linear constraints of (2.44), we have

$$\mathbf{w}_0 = \mathbf{C}(\mathbf{C}^H\mathbf{C})^{-1}\mathbf{f}. \tag{2.46}$$

Suppose that the columns of the matrix $\mathbf{C}_n$ form a basis for the null space of $\mathbf{C}^H$. It is possible to find $\mathbf{w}_n$ such that $\mathbf{v}_n = \mathbf{C}_n\mathbf{w}_n$. The minimum variance beamformer is now transformed to

$$\min_{\mathbf{w}_n}(\mathbf{w}_0 - \mathbf{C}_n\mathbf{w}_n)^H\mathbf{R}(\mathbf{w}_0 - \mathbf{C}_n\mathbf{w}_n). \tag{2.47}$$

It is seen that the constrained minimization (2.45) has been reduced to an unconstrained minimization (2.47). The solution to (2.47) is given by

$$\mathbf{w}_n = (\mathbf{C}_n^H\mathbf{R}\mathbf{C}_n)^{-1}\mathbf{C}_n^H\mathbf{R}\mathbf{w}_0. \tag{2.48}$$

Fig. 2.10 shows implementation of a generalized sidelobe canceler. This structure is very useful for adaptive array processing since only $p - L$ dimensional vector $\mathbf{w}_n$ should be adapted. The complexity is reduced by performing the adaptation in a space with a smaller dimensionality.

The concept of linear prediction for time series can also be used for beamforming [37].

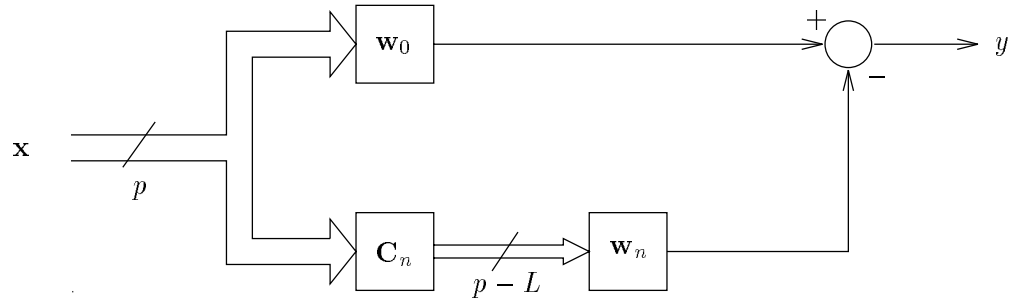

Fig. 2.10 The generalized sidelobe canceler

The output of one sensor is predicted based on a linear combination of the outputs of the other sensors. Let the outputs of an array of p sensors be represented by $x_1(t), \ldots, x_p(t)$. A *linear prediction beamformer* selects the weights $w_1, \ldots, w_{p-1}$ such that

$$\epsilon_p = E|x_p(t) - \sum_{i=1}^{p-1} w_i x_i(t)|^2 \tag{2.49}$$

is minimized. The orthogonality principle yields [23]

$$E[x_p(t)x_j^*(t)] = \sum_{i=1}^{p-1} w_i E[x_i(t)x_j^*(t)] \qquad \text{for } j = 1, \ldots, p-1. \tag{2.50}$$

These equation can be arranged as

$$\mathbf{R}\mathbf{w} = \mathbf{r}_p \tag{2.51}$$

where $\mathbf{R}$ is a $(p-1) \times (p-1)$ matrix of the entries $R_{ij} = E[x_i(t)x_j^*]$ and $\mathbf{r}_p$ is a $(p-1) \times 1$ vector with the i-th component $r_i = E[x_p(t)x_i^*(t)]$. Assuming that $\mathbf{R}$ is full rank, the weights can be found from

$$\mathbf{w} = \mathbf{R}^{-1}\mathbf{r}_p \tag{2.52}$$

It can be shown that the resolution of the linear prediction beamformer is higher than that for the minimum variance and the delay-and-sum techniques [24].

2.4.2. Subspace decomposition methods

We saw that the width of the mainlobe of the beampattern for a uniform linear array with half wavelength spacing is inversely proportional to the number of sensors. If two sources have a spacing smaller than the width of the mainlobe, they cannot be resolved by using a delay-and-sum beamformer. To increase the resolution of a beamformer it is necessary to increase the number of sensors. In practice, physical limitations do not allow a very large array aperture. Furthermore, the spatial correlation of the signal is often a decreasing function of distance. If the spacing between two sensors is larger than the spatial correlation of the signal, their output cannot be coherently added to increase the array gain. For these reasons alternative methods which achieve the same resolution quality with a smaller array aperture should be considered. The array processing techniques which are based on the signal and noise subspace decomposition have higher resolution than beamforming methods.

The objective of introducing high resolution methods is to obtain a better performance in field characterization. Any field characterization problem can be decomposed into two parts: detection and localization. By *detection* we mean a method that can determine the number of signals which are arriving at the array. The *localization* methods are the techniques that are used to estimate the spatial parameters of the sources.

Assume that an array of p sensors receives the wavefield of $q < p$ narrowband sources. For a unique localization, the number of sources should be smaller than the number of sensors [5] [58]. The i-th snapshot of the array output is represented by

$$\mathbf{x}_i = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}_i + \mathbf{n}_i \tag{2.53}$$

where $\mathbf{x}_i$ is the $p \times 1$ observation vector, $\mathbf{s}_i$ is the complex envelope of the source signals arranged in a $q \times 1$ vector, $\mathbf{n}_i$ is the $p \times 1$ noise vector, and $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_q)]$ is the $p \times q$ location matrix. The *array manifold* is defined as $\mathcal{A} = \{\mathbf{a}(\theta) | \theta \in \Theta\}$ where Θ is the set of all values of θ. It is assumed that the array manifold is known and any p location vectors $\mathbf{a}(\theta_i),\ i = 1, \ldots, p$, with distinct θ_i's are linearly independent.

The signal snapshots are modeled as independent identically distributed (i.i.d.) sequence of complex circular Gaussian random vectors with an unknown covariance matrix $\mathbf{S}$. The noise snapshots are i.i.d. sequence of complex circular Gaussian random vectors with unknown covariance matrix $\sigma^2\mathbf{I}$ and are independent of the signal samples. Generalization to nonwhite noise is by pre-whitening [4]. With these assumptions, the observation vectors will be the samples of a complex circular Gaussian process with zero mean and the correlation matrix

$$\mathbf{R} = E[\mathbf{x}_i\mathbf{x}_i^H|\boldsymbol{\theta}] \tag{2.54}$$

$$= \mathbf{A}(\boldsymbol{\theta})\mathbf{S}\mathbf{A}^H(\boldsymbol{\theta}) + \sigma^2\mathbf{I} \tag{2.55}$$

where $\boldsymbol{\theta}$ is the $q \times 1$ vector of DOAs. As it is seen, the correlation matrix $\mathbf{R}$ is a function of $q, \boldsymbol{\theta}, \sigma^2$ and $\mathbf{S}$. The objective in a detection and localization problem is to estimate the number of sources q and their DOAs $\boldsymbol{\theta}$.

Assume that the signals of the sources are noncoherent. The coherent case will be considered later. The eigen-decomposition of the array correlation matrix $\mathbf{R}$ is represented by

$$\mathbf{R} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H \tag{2.56}$$

where $\boldsymbol{\Lambda}$ is the diagonal matrix of the eigenvalues $\lambda_i,\ i = 1, \ldots, p$, arranged in nonincreasing order and $\mathbf{V}$ is the matrix of corresponding eigenvectors. It is possible to show that [30]

$$\lambda_i = \sigma^2 \qquad \text{for} \quad i = q+1, \ldots, p. \tag{2.57}$$

Since the last $p - q$ eigenvalues of $\mathbf{R}$ are equal to the noise variance, they are called the *noise eigenvalues*. The contribution of the signal to the correlation matrix is only along the first q eigenvectors. These eigenvectors correspond to the *signal eigenvalues*. The *signal subspace* is defined as the subspace spanned by the signal component. Assuming noncoherent sources, the signal subspace is the span of the columns of the array location matrix. The dimensionality of the signal subspace is equal to the number of noncoherent sources q. From the structure of the array correlation matrix (2.55) the signal subspace

is identical to the span of the first q columns of $\mathbf{V}$. The *noise subspace* is defined as the orthogonal complement of the signal subspace. The dimensionality of the noise subspace for noncoherent sources is $p-q$.

Many array processing methods are based on the decomposition of the observation space into the signal and noise subspaces. The first step in these techniques is to estimate the signal and noise subspaces by decomposing the array correlation matrix into its eigenstructure form. The subspace spanned by the eigenvectors of $\mathbf{R}$ corresponding to dominant eigenvalues is the signal subspace. The detection methods use the fact that the signal eigenvalues are larger than the noise eigenvalues. These techniques estimate the number of signals by separating the dominant eigenvalues of the correlation matrix or by choosing the eigenvalues that optimize a given information theoretic criterion. These methods will be reviewed in the following section.

2.4.2.1. Detection

The techniques that can be used to determine the number of sources are called detection methods. In any detection problem there are two or more hypotheses which are represented by $\mathcal{H}_i,\ i=0,\ldots,M-1$, where M is the number of hypotheses. There is a probability associated with each hypothesis, say $p(\mathcal{H}_i)$. The detection problem is then to select an appropriate hypothesis by optimizing a cost function. In array processing it is frequently assumed that $\mathcal{H}_i$ is the hypothesis that the wavefield is being generated by i sources. For a unique localization of the sources it is required that the number of sources be smaller than the number of sensors. Thus, the number of hypotheses M is p. It is further assumed that $p(\mathcal{H}_i)$ is uniform which indicates that no prior information about the number of sources is available.

In the previous section, we mentioned that for a scenario of q noncoherent sources the smallest eigenvalue of the array correlation matrix has a multiplicity $p-q$. This fact can be used to detect the number of sources. Let the hypothesis $\mathcal{H}_i$ be that the smallest eigenvalue of the array correlation matrix $\mathbf{R}$ has a multiplicity $p-i$. If we had the correlation matrix of the array, we could simply count the multiplicity of the smallest identical eigenvalues.

In practice, the array correlation matrix is estimated by using sample correlation matrix

$$\bar{\mathbf{R}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i^H. \tag{2.58}$$

Since the observation time is finite, the last $p-q$ eigenvalues of $\bar{\mathbf{R}}$ are not equal. In such a case, the number of sources is estimated by examining the differences between consecutive eigenvalues. We represent the ordered eigenvalues of $\bar{\mathbf{R}}$ by $\bar{\lambda}_i$, $i = 1, \ldots, p$. For a high SNR and q sources, the difference $\bar{\lambda}_q - \bar{\lambda}_{q+1}$ is large.

There are two shortcomings with this method of detection. First, if the SNR is low, $\bar{\lambda}_q$ is close to $\bar{\lambda}_{q+1}$ and the detection method may not be accurate. Second, if some of the sources are completely correlated (coherent), the multiplicity of the smallest eigenvalue of $\mathbf{R}$ is not equal to $p-q$. In other words, the coherent sources cannot be detected by using this method. For coherent signal scenarios rank of the noise-free correlation matrix is smaller than the number of signals. To apply this method, a *smoothing* process is required to increase the rank of the noise-free correlation matrix [31].

An alternative detection method is based on the *information theoretic* approach. Two major information theoretic criteria are Akaike's information criterion (AIC) [1], and Rissannen's minimum description length (MDL) [25]. These methods minimize the Kullback-Leibler distance between the hypothetical model and the observed data. Mathematically, these two criteria are shown as

$$\mathrm{AIC}(N) = -\log f(\mathbf{X}^N; \hat{\boldsymbol{\theta}}_N) + k \tag{2.59}$$

$$\mathrm{MDL}(N) = -\log f(\mathbf{X}^N; \hat{\boldsymbol{\theta}}_N) + \frac{k}{2}\log N \tag{2.60}$$

where $\mathbf{X}^N$ is the $p \times N$ matrix of observations up to time N, $\hat{\boldsymbol{\theta}}_N$ is the maximum likelihood (ML) estimate of the parameter vector based on N snapshots, k is the number of free elements of the parameter vector, and f is the generating model class (probability density function). In these formulae, the first term is the log-likelihood function of the observation vectors and the second term penalizes overparameterization. The detection is performed by computing these criteria for all the models of order $\tilde{q} \in \{0, 1, \ldots, p-1\}$. Choosing the

minimum of the computed criterion over all $\tilde{q}$ gives the estimate of the number of sources. It has been shown that the AIC detector is not consistent and tends to overestimate the true order of the system for a large number of observations [55]. The MDL method on the other hand is a consistent detector. In both methods buffering of data is required.

In Chapter 3, a new detection method is developed using the predictive stochastic complexity (PSC). The PSC principle is based on the concept of predictive coding [27]. The PSC is the length of the coded data which is produced by a predictive encoder. The codelength minimization is an appropriate criterion for model selection due to the fact that the best model which fits to data is the one that gives the most information about it; having more information results in a smaller codelength. The predictive stochastic complexity of the observation vector $\mathbf{x}_i,\ i = 1, \ldots, N$, is defined as

$$\mathrm{PSC}(N) = -\sum_{i=1}^{N} \log f(\mathbf{x}_i|\hat{\boldsymbol{\theta}}_{i-1}) \tag{2.61}$$

where $f(\mathbf{x}|\boldsymbol{\theta})$ is the conditional parametric probability density function of the observation vector, and $\hat{\boldsymbol{\theta}}_{i-1}$ is the ML estimate of the parameter vector with respect to the observations up to time $(i-1)$. Unlike AIC and MDL, the PSC is a recursive detector and can be used on-line. Furthermore, it is a consistent estimator and asymptotically estimates the true order of system.

2.4.2.2. Localization

The objectives of an array processor for field characterization is to determine the number of signals and estimate their spatial parameters. Parameter estimation means localization of the signals by estimating their DOAs. In the literature, there are several approaches to DOA estimation. Here, we only consider multiple signal classification (MUSIC) [30] and estimation of the signal parameters via rotational invariance technique (ESPRIT) [29] which are the two most popular methods.

The MUSIC algorithm uses the fact that the span of the location matrix is the same as the span of the first q eigenvectors of the correlation matrix. Since the eigenvectors

form an orthonormal basis for the observation space, we have

$$\mathbf{a}^H(\theta_i)\,\mathbf{e}_j = 0 \qquad \text{for} \quad i = 1,\ldots,q, \quad j = q+1,\ldots,p \tag{2.62}$$

where the $\mathbf{e}_j$, $j = 1,\ldots,p$, are the eigenvectors of the correlation matrix arranged such that their corresponding eigenvalues are in nonincreasing order. Using this property, the MUSIC frequency estimator is given by

$$P_{\text{MUSIC}} = \max_{\theta} \frac{1}{\mathbf{a}^H(\theta)\mathbf{E}_n\mathbf{E}_n^H\mathbf{a}(\theta)} \tag{2.63}$$

where $\mathbf{E}_n = [\mathbf{e}_{q+1},\ldots,\mathbf{e}_p]$ is the noise eigenvector matrix. To perform the maximization, the array manifold is searched for the maxima of the MUSIC spectrum. This requires complete knowledge of the array manifold. For practical scenarios with imperfect knowledge of the sensors, a calibration step is required. The *calibration* of array for each angle θ is the measurement of the array response for a source located at θ. This process is cumbersome and usually needs a large memory to store the calibrated data. Furthermore, if the characteristics of the sensors change with time, the calibration process should be reperformed from time to time.

ESPRIT is an alternative method that does not need a calibration step. In the ESPRIT algorithm it is assumed that the signals are received by an array of matched doublets. Suppose that q far-field sources arrive at an array of p doublets. Each doublet consists of two identical sensors. The array can be decomposed into two subarrays such that the subarray are completely identical and are displaced from each other with a known directional vector $\vec{\Delta}$. Fig. 2.11 shows a configuration with two uniform linear subarrays. The outputs of the subarrays are represented by p-vectors $\mathbf{x}(t)$ and $\mathbf{y}(t)$ with the i-th components

$$x_i(t) = \sum_{l=1}^{q} s_l(t - \tau_i(\theta_l)) + n_{x_i}(t), \tag{2.64}$$

$$y_i(t) = \sum_{l=1}^{q} s_l(t - \tau_i(\theta_l) - \tau_d(\theta_l)) + n_{y_i}(t), \tag{2.65}$$

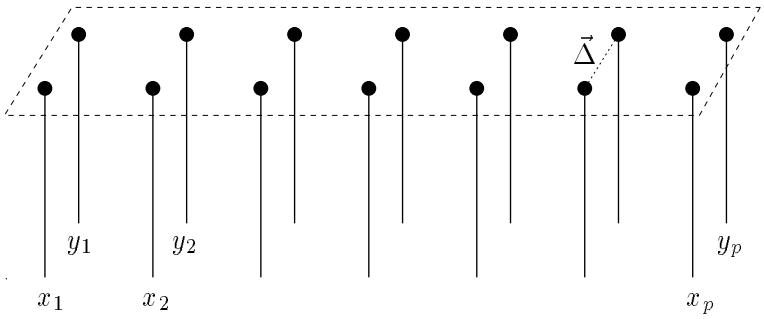

Fig. 2.11 An array of matched doublets

for $i = 1, \ldots, p$, where s_l is the l-th source signal, θ_l is angle between the wavefront of the l-th source and vector $\vec{\Delta}$, $\tau_i(\theta_l)$ is the propagation delay for the l-th source at the i-th sensor of the first subarray with respect to the reference point and $\tau_d(\theta_l) = \Delta \sin\theta_l / c$ is the delay in the propagation between the two subarrays, where c is the propagation velocity and Δ is the distance between two subarrays. For linear array with uniform spacing, $\tau_i(\theta_l) = (i-1)d \sin\theta_l / c$, where d is the spacing between two consecutive sensors and the reference point is at the first sensor of the first subarray. The noise components of the subarrays are represented by $n_{x_i}(t)$ and $n_{y_i}(t)$.

The complex envelope representation of (2.64) and (2.65) in vector form is given by

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n}_x, \tag{2.66}$$

$$\mathbf{y} = \mathbf{A}\Phi\mathbf{s} + \mathbf{n}_y. \tag{2.67}$$

In this equation,

$$\Phi = \text{diag}(e^{-j\omega\tau_d(\theta_1)}, \ldots, e^{-j\omega\tau_d(\theta_1)}) \tag{2.68}$$

is the rotation matrix of the phase delays between the two subarrays. The DOAs are found by estimating the diagonal elements of the rotation matrix Φ as follows [28].

Let us define the vectors $\mathbf{z}$ and $\mathbf{n}_z$ by

$$\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}, \qquad \mathbf{n}_z = \begin{bmatrix} \mathbf{n}_x \\ \mathbf{n}_y \end{bmatrix}. \tag{2.69}$$

The correlation matrices of $\mathbf{z}$ and $\mathbf{n}_z$ are represented by $\mathbf{R}_{zz}$ and $\mathbf{R}_{nn}$, respectively. The generalized eigenvalue decomposition of the matrix pencil $(\mathbf{R}_{zz}, \mathbf{R}_{nn})$ is given by

$$\mathbf{R}_{zz}\tilde{\mathbf{E}} = \mathbf{R}_{nn}\tilde{\mathbf{E}}\boldsymbol{\Lambda} \tag{2.70}$$

where $\boldsymbol{\Lambda}$ is the diagonal matrix of generalized eigenvalues arranged in nonincreasing order. The matrix of generalized eigenvectors can be used to generate

$$\begin{bmatrix} \mathbf{E}_x \\ \mathbf{E}_y \end{bmatrix} = \mathbf{R}_{nn}\tilde{\mathbf{E}}_q \tag{2.71}$$

where $\tilde{\mathbf{E}}_q$ is a submatrix of $\tilde{\mathbf{E}}$ containing the first q columns. We rearrange $\mathbf{E}_x$ and $\mathbf{E}_y$ in the following form and compute the eigen-decomposition

$$\begin{bmatrix} \mathbf{E}_x^H \\ \mathbf{E}_y^H \end{bmatrix} [\mathbf{E}_x \ \mathbf{E}_y] = \mathbf{E}\boldsymbol{\Sigma}\mathbf{E}^H \tag{2.72}$$

Let $\mathbf{E}$ be partitioned into $q \times q$ matrices such as

$$\mathbf{E} = \begin{bmatrix} \mathbf{E}_{11} & \mathbf{E}_{12} \\ \mathbf{E}_{21} & \mathbf{E}_{22} \end{bmatrix}. \tag{2.73}$$

The diagonal elements of Φ are then given by

$$\phi_k = \lambda_k(-\mathbf{E}_{12}\mathbf{E}_{22}^{-1}), \qquad \text{for} \quad k = 1, \ldots, q \tag{2.74}$$

where $\lambda_k(\mathbf{B})$ is the k-th eigenvalue of $\mathbf{B}$.

2.5. Wideband array processing

The MUSIC and ESPRIT algorithms were originally developed for narrowband signal localization. They cannot be directly applied to wideband cases. For wideband signals, the frequency-wavenumber spectrum is extended along the frequency axis. Since the spectrum

is spread in frequency, a temporal filtering is required to separate signals. The role of a temporal filter is to use the content of the frequency spectrum of the signals to extract the useful information. There are different approaches to wideband array processing. In some methods the observation vectors are used to construct a wideband correlation matrix. In [35] the concept of the signal and noise subspaces, which is originally derived for narrowband signal models, is extended to wideband signals. In this method, the elements of the vector space are the signals with rational spectrum. Using these concepts, a MUSIC type algorithm is derived for wideband signal localization.

Alternative methods use sampling of the sensors output in the frequency domain to create narrowband signals. In the incoherent signal-subspace method (ISM) [56], the narrowband signals are processed to estimate the DOAs. Then these results are combined to obtain the final solution. It has been shown that coherent sources cannot be resolved by this approach. Furthermore, the effectiveness of this method deteriorates for closely separated sources and low SNR.

The coherent signal-subspace method (CSM) [52] is an alternative to ISM that improves the estimation by condensing the energy of narrowband signals in a pre-defined subspace. This process is called *focusing*. The universal correlation matrix that has been generated by focusing narrowband components, has the characteristics of a narrowbnad correlation matrix. Any narrowband detection method can be applied to the universal correlation matrix to localize the signals. Wang and Kaveh [52] propose using the MUSIC algorithm for DOA estimation.

Assume that the output of sensors is decomposed into nonoverlapping snapshots of J samples each. An FFT algorithm is applied to each snapshot to sample the signal spectrum in J points. Let us represent these samples by $\mathbf{x}_j,\ j = 1, \ldots, J$. The difficulty in processing these vectors is that the signal subspace is a function of the processing frequency and is different for different frequency bins. The CSM algorithm transforms these samples into the focusing subspace by

$$\mathbf{y}_j = \mathbf{T}_j \mathbf{x}_j \qquad j = 1, \ldots, J, \tag{2.75}$$

where $\mathbf{T}_j$ is the focusing matrix and $\mathbf{y}_j$ is the focused observation vector for the j-th frequency bin. In [17] the focusing matrices are the solution of

$$\min_{\mathbf{T}_j} \|\mathbf{A}_0 - \mathbf{T}_j\mathbf{A}_j\| \tag{2.76}$$

$$\text{s.t.} \quad \mathbf{T}_j^H\mathbf{T}_j = \mathbf{I} \tag{2.77}$$

where $\mathbf{A}_0$ is the focusing location matrix and $\mathbf{A}_j$ is the location matrix at the j-th frequency bin. The signal subspaces of the y_j's are identical or very close to each other. Thus signal subspace processing of these narrowband signals can be performed by combining the correlation matrices of these focused observation vectors. The CSM algorithm averages the correlation matrices at different frequency bins to generate a universal narrowband correlation matrix. Then the MUSIC algorithm is applied to estimate the DOAs.

It has been shown that the CSM improves the resolution threshold and resolves coherent sources [52]. Despite the fact that CSM is very effective in wideband signal detection and estimation, it suffers from an asymptotic bias of the peaks as a result of error in estimated focusing DOAs [36]. The bias increases with the bandwidth of the sources and deviation of the focusing DOAs from the true directions of arrival. In Chapter 4, we will show that with a proper selection of the focusing frequency, the estimation bias can be decreased. However, in general, by using the CSM algorithm an unbiased estimation of the DOA is not possible.

Two alternative methods have been evolved from the CSM algorithm that can asymptotically generate unbiased estimates of the DOAs. In [7] a broadband signal-subspace spatial-spectrum estimation (BASS-ALE) algorithm is proposed. The method forms a broadband covariance matrix with the low rank character of broadband signal observation. The dimensionality of the signal subspace is equal to the source time-bandwidth product [6]. In this method, the estimation bias is reduced by increasing the dimensionality of the location vectors. The trade-off is an increase in the computational complexity.

A second method for bias reduction in wideband array processing was proposed in [22]. The algorithm is based on the steered covariance matrix (STCM). In this technique,

delay elements are introduced at the front-end of the array and the covariance matrix is computed after the delays. With a proper choice of the delays, a steering beam can be formed. It has been shown that when the steering beam coincides with a true DOA, the STCM contains a dc term equal to the power of the corresponding source regardless of its spectral signature. Thus, by steering the space and locating the peaks of the dc component in STCM, the DOA is estimated. The bias is reduced by forming a broadband covariance matrix for which the signal subspace has a rank one representation. An increase in the computational complexity is the price to be paid.

In Chapter 4, an optimal method for focusing subspace selection in the CSM algorithm is introduced. The method is based on minimizing a subspace fitting error. The subspace fitting error is defined as the Euclidean distance between the focusing location matrix and the transformed location matrix at each frequency bin. The subspace fitting error for the j-th frequency is defined as

$$\epsilon_j = \|\mathbf{A}_0 - \mathbf{T}_j\mathbf{A}_j\| \tag{2.78}$$

where $\|.\|$ is the Frobenius matrix norm. In the past, the focusing frequency of $\mathbf{A}_0$ has been chosen to be the center frequency of the spectrum of the signals. If the spectrum of the signal is not symmetric around the center frequency, or the sampling in the frequency domain is not uniform, this choice is not optimal. Here, we choose a focusing frequency which is selected by

$$\begin{aligned} \min_{f_0} \ \min_{\mathbf{T}_j} & \sum_{j=1}^{J} w_j \|\mathbf{A}_0 - \mathbf{T}_j\mathbf{A}_j\|^2 \\ \text{s.t.} \quad & \mathbf{T}_j^H\mathbf{T}_j = \mathbf{I}, \\ & \mathbf{A}_0 \in \boldsymbol{\mathcal{A}}(\boldsymbol{\theta}) \end{aligned} \tag{2.79}$$

where $\boldsymbol{\mathcal{A}}(\boldsymbol{\theta})$ is the set of all location matrices for given DOA $\boldsymbol{\theta}$, and w_j is a weighting factor proportional to the SNR at the j-th frequency bin with $\sum_{j=1}^{J} w_j = 1$. We show that by using the proposed method for focusing subspace selection in the CSM algorithm, the bias of the DOA estimate is minimized and the resolution threshold SNR is reduced.

Although we have found a method to minimize the bias of the estimation in the CSM algorithm, it is not possible to produce unbiased estimates by using this technique. In Chapter 5, we introduce a new method for wideband array processing. The method is termed two-sided correlation transformation (TCT) and is based on computing the focusing transformation matrices by

$$\begin{aligned} &\min_{\mathbf{U}_j} \|\mathbf{P}_0 - \mathbf{U}_j \mathbf{P}_j \mathbf{U}_j^H\| \qquad \text{for} \quad j = 1, \ldots, J \\ &\text{s.t.} \quad \mathbf{U}_j^H \mathbf{U}_j = \mathbf{I}, \end{aligned} \tag{2.80}$$

where $\mathbf{P}_j$ is the correlation matrix of the sensor output at the j-th frequency bin in a noise-free environment, $\mathbf{P}_0$ is the focusing noise-free correlation matrix, and $\mathbf{U}_j$ is the j-th transformation matrix. The TCT method has a lower resolution threshold SNR and a smaller bias than the CSM algorithm. Also it is asymptotically unbiased. We show that the generalized variance of the TCT algorithm is smaller than the CSM method. This results in an estimation of the DOAs with a smaller variance.

2.6. Distributed sources

In array processing it is frequently assumed that the signals are generated by point sources. However, many practical examples can be found where the point source assumption is an unrealistic modeling of the sources. In an undersea echo beam sounder, the reflection of the signal and penetration into the lower levels of the seabed creates a spatial distribution of the receiving waveform [19]. In radar, the received signal is a superposition of the reflections of the pulse from different parts of the target. If the target is spread in range, it appears as a distributed source [50]. In sonar, multipath exhibits an angularly extended interference [14]. This interference can be treated as a spatially distributed signal. A source distribution in space can also be observed in the transmission of radio-waves through ionospheric and tropospheric scatter links, and the propagation of audio signals in a reverberant room.

For narrowband point-source configurations, the dimension of the signal subspace is

equal to the number of noncoherent signals. Thus, each source has a one-dimensional representation in the signal subspace. In previous works, distributed sources have been viewed as a combination of closely spaced point sources [19]. For a good approximation, the number of point sources should be large. If the number of point sources increases beyond the number of sensors, a unique solution for the localization problem may not exist. In such a case, the corresponding location matrix spans the whole space and the noise subspace is empty. This explains why the conventional array processing techniques such as MUSIC [30] and ESPRIT [28], which are based on the signal and noise subspace decomposition for point source scenarios, often lead to erroneous results when applied to such models of distributed sources [19].

In Chapter 6, we introduce a technique for localization of distributed sources. The method is based on generalizing the MUSIC algorithm for extended source applications. We assume that the correlation kernel of the distributed source belongs to a parametric class of functions. The parameter vector of the correlation kernel is estimated by locating the peaks of

$$\hat{\boldsymbol{\psi}} = \arg\max_{\boldsymbol{\psi}} \frac{1}{\int_{\frac{-\pi}{2}}^{\frac{\pi}{2}} \int_{\frac{-\pi}{2}}^{\frac{\pi}{2}} \mathbf{a}^H(\theta) \mathbf{E}_n p(\theta, \theta'; \boldsymbol{\psi}) \mathbf{E}_n^H \mathbf{a}(\theta') d\theta d\theta'} \tag{2.81}$$

where $\mathbf{a}(\theta)$ is the array location vector, $\mathbf{E}_n$ is the noise eigenvector matrix, and $p(\theta, \theta'; \boldsymbol{\psi})$ is the correlation kernel for the signals. The parametric assumption for the correlation kernel is necessary for uniqueness of localization. The method has been formulated for coherently and incoherently distributed sources. We compare the new technique with the conventional MUSIC algorithm. The resolution SNR for the new technique is considerably lower than for the MUSIC algorithm.

Chapter 3

Signal Detection

The first step in processing of spatiotemporal signals using an array of sensors is to ascertain the number of the sources that generate the wavefield. The process of determining the number of sources is often called detection. There are different approaches to signal detection. Recently, much attention has been given to information theoretic criteria [55] [57] [62] [59] [60]. Two important information theoretic criteria are Akaike's information criterion (AIC) [1] and the minimum description length (MDL) [25]. These methods minimize the Kullback-Leibler distance between the hypothetical model and the observed data. The number of sources is detected by computing these criteria for the models of order $\hat{q} \in \{0, 1, \ldots, p-1\}$ and choosing the one which gives the minimum distance. A general class of information theoretic criteria evolving from MDL is also repored in literature [61] [63].

The AIC and MDL criteria consist of the log-likelihood function of the observation and an additive over-parameterization term. The over-parameterization term in the AIC criterion is the number of free parameters. It has been shown that the AIC criterion is not consistent [55]. AIC tends to overestimate the number of parameters as the observation length increases. This can be attributed to the fact that the over-parameterization term is independent of the observation length while the log-likelihood function grows with the observation length.

The MDL is based on minimization of the length of the code that is required to describe

data. Codelength minimization is appropriate for model selection since the model which best fits the data is the one that gives the most information about it — having more information results in a smaller codelength. Coding of data in the MDL criterion is performed in two steps. First, the data is coded using a uniquely decodable prefix code. Then the parameter vector is coded and added as a preample to the codeword of data. The over-parameterization term in the MDL principle represents the number of digits required to encode the parameter vector to an optimal precision [26]. Restriction of the coding method to a two-step scheme in the MDL criterion increases the codelength. It has been shown that the MDL is consistent [55]. For both the AIC and MDL methods, buffering of data is essential.

In this chapter, we develop an algorithm based on Rissanen's predictive stochastic complexity (PSC) [26] [27]. The PSC criterion is the addition of the log-likelihood functions of the observation vectors such that at each time instant the maximum likelihood (ML) estimate of the parameter based on the past data is used in the probability distribution function. It has been shown that the PSC achieves the shortest codelength of the data relative to the generating model class [27]. The algorithm is consistent and its structure makes it suitable for on-line use.

3.1. The PSC criterion

Suppose that an array of p sensors is exposed to $q < p$ far-field sources. The signals from the sources can be partially or fully correlated. The fully correlated case (also called the *coherent* case) arises from multipath propagation or smart jamming and is of practical importance in signal processing. We assume narrowband signals with known center frequency. If the size of the array is much smaller than the range of the sources, it can be assumed that the array of sensors is in the far field of the sources. With this model, the arriving wavefronts at the array are planar and do not carry any information about the range of the signals. It is also frequently assumed that the source and the sensors are in the same plane. In such a case, the only information about the position of a source is its direction-of-arrival (DOA). Assume that the signals arrive at the array from distinct

directions $\theta_1, \ldots, \theta_q$. The objective is to find the number of sources q and their directions of arrival $\boldsymbol{\theta}^q = [\theta_1, \ldots, \theta_q]$.

The i-th snapshot of the array output is represented by

$$\mathbf{x}_i = \mathbf{A}(\boldsymbol{\theta}^q)\mathbf{s}_i + \mathbf{n}_i \tag{3.1}$$

where $\mathbf{x}_i$ is the $p \times 1$ observation vector, $\mathbf{s}_i$ is the $q \times 1$ signal vector, $\mathbf{n}_i$ is the $p \times 1$ noise vector, and $\mathbf{A}(\boldsymbol{\theta}^q) = [\mathbf{a}(\theta_1) \ldots \mathbf{a}(\theta_q)]$ is the $p \times q$ location matrix. The array manifold is defined by $\mathcal{A} = \{\mathbf{a}(\theta) | \theta \in \Theta\}$ where Θ is the region of search. It is assumed that the array manifold is known and any p location vectors $\mathbf{a}(\theta_i),\ i = 1, \ldots, p$, with distinct θ_i's are linearly independent. The methods that we present in this chapter can be applied to a general array manifold. However, for simplicity we have used linear arrays in our simulations. For a uniformly spaced linear array the i-th location vector is represented by $\mathbf{a}(\theta_i) = [1\ \mu_i\ \mu_i^2\ \ldots\ \mu_i^{p-1}]^T$ with $\mu_i = \exp(j\omega_0 d \sin\theta_i / c)$, where d is the distance between two consecutive sensors, c is the wave speed, and ω_0 is the center frequency of the source signal.

The signal snapshots are modeled as an independent identically distributed (i.i.d.) sequence of complex circular Gaussian random vectors with an unknown covariance matrix $\mathbf{S}^q$. The noise snapshots are i.i.d. sequence of complex circular Gaussian random vectors with unknown covariance matrix $\sigma^2\mathbf{I}$, independent of the signal samples. Pre-whitening can be used to accommodate non-white noise. With these assumptions, the observation vectors will be the samples of a complex circular Gaussian process with zero mean and correlation matrix

$$\begin{aligned} \mathbf{R}^q &= E[\mathbf{x}\mathbf{x}^H | q, \boldsymbol{\theta}^q, \sigma^2, \mathbf{S}^q] \\ &= \mathbf{A}(\boldsymbol{\theta}^q)\mathbf{S}^q\mathbf{A}^H(\boldsymbol{\theta}^q) + \sigma^2\mathbf{I}. \end{aligned} \tag{3.2}$$

The conditional probability density function of the observation vector is given by

$$f(\mathbf{x}|\boldsymbol{\theta}^q) = \frac{1}{\pi^p\ |\mathbf{R}^q|} \exp\{-\mathbf{x}^H[\mathbf{R}^q]^{-1}\mathbf{x}\} \tag{3.3}$$

where $|.|$ represents the determinant of a matrix.

The correlation matrix $\mathbf{R}^q$ is a function of the parameter set $\boldsymbol{\phi} = [q, \boldsymbol{\theta}^q, \sigma^2, \mathbf{S}^q]$. The objective in a detection and localization problem is to estimate the number of sources q and their DOAs $\boldsymbol{\theta}^q$. It might be thought that the unknown parameters of $\mathbf{R}^q$ can be estimated by jointly maximizing the likelihood function. However, the maximum allowable value of the likelihood function is an increasing function of the order of the system. In other words, direct ML estimation always gives the maximum value for the number of sources q. For this reason, the AIC and MDL algorithms include a second term which is added to the log-likelihood function to penalize over-parameterization.

Mathematically, the AIC and MDL criteria are represented by

$$\mathrm{AIC}(N) = -\log f(\mathbf{X}^N|\hat{\boldsymbol{\phi}}_N) + k \tag{3.4}$$

$$\mathrm{MDL}(N) = -\log f(\mathbf{X}^N|\hat{\boldsymbol{\phi}}_N) + \frac{k}{2}\log N \tag{3.5}$$

where $\mathbf{X}^N$ is the $p \times N$ matrix of observations up to time N, $\hat{\boldsymbol{\phi}}_N$ is the ML estimate of the parameter vector based on N snapshots, k is the number of free elements of the parameter vector, and f is the generating model class (probability density function). In these formulae, the first term is the log-likelihood function of the observation vectors and the second term compensates for over-parameterization.

The predictive stochastic complexity of the observation vectors $\mathbf{x}_i, i = 1, \ldots, N$, is defined as

$$\mathrm{PSC}(N) = -\sum_{i=1}^{N} \log f(\mathbf{x}_i|\hat{\boldsymbol{\phi}}_{i-1}) \tag{3.6}$$

where $\hat{\boldsymbol{\phi}}_{i-1}$ is the ML estimate of the parameter vector with respect to the observations up to time $(i-1)$. The PSC principle is based on predictive coding of data. At each time instant the parameter vector is selected by using the past observations. The i-th term, $-\log f(\mathbf{x}_i|\hat{\boldsymbol{\phi}}_{i-1})$, is basically the codelength of the prediction error [27].

3.2. Detection of noncoherent sources

It is assumed that $q < p$, which means $q \in \mathcal{P} = \{0, 1, \ldots, p-1\}$. For any $k \in \mathcal{P}$, an appropriate model of order k can be constructed. In order selection methods, p models run simultaneously. The output of array for a given $k \in \mathcal{P}$ is expressed as

$$\mathbf{x}_i = \mathbf{A}(\boldsymbol{\theta}^k)\mathbf{s}_i^k + \mathbf{n}_i \tag{3.7}$$

where $\mathbf{A}(\boldsymbol{\theta}^k)$ is the $p \times k$ parameterized location matrix with respect to the parameter vector $\boldsymbol{\theta}^k$, and $\mathbf{s}_i^k(t)$ represents a $k \times 1$ signal vector. The conditional probability density function of the observation vector for model k is given by

$$f(\mathbf{x}|\mathbf{R}^k) = \frac{1}{\pi^p \, |\mathbf{R}^k|} \exp\{-\mathbf{x}^H [\mathbf{R}^k]^{-1} \mathbf{x}\} \tag{3.8}$$

where $\mathbf{R}^k$ is the correlation matrix of the observation vector for the k-th model.

In this chapter, the predictive stochastic complexity is computed for each model and is minimized over all models to estimate q. From (3.8) the PSC for a model of order k at time instant N is given by

$$\begin{aligned} \mathrm{PSC}_k(N) &= -\sum_{i=1}^{N} \log f(\mathbf{x}_i|\hat{\mathbf{R}}_{i-1}^k) \\ &= \sum_{i=1}^{N} \Big(\log |\hat{\mathbf{R}}_{i-1}^k| + \mathbf{x}_i^H [\hat{\mathbf{R}}_{i-1}^k]^{-1} \mathbf{x}_i \Big) \end{aligned} \tag{3.9}$$

where $\hat{\mathbf{R}}_{i-1}^k$ is the ML estimate of the correlation matrix for the model of order k based on the observations up to time $(i-1)$.

The sample correlation matrix

$$\bar{\mathbf{R}}_{i-1} = \frac{1}{i-1} \sum_{l=1}^{i-1} \mathbf{x}_l \mathbf{x}_l^H \tag{3.10}$$

is used to determine the ML estimate of the true correlation matrix. Assume that $\bar{\lambda}_j, j = 1, \ldots, p$, are the eigenvalues of $\bar{\mathbf{R}}_{i-1}$ arranged in nonincreasing order with the

corresponding eigenvectors $\bar{\mathbf{v}}_j, j = 1, \ldots, p$. It is possible to show that the eigenvalues and eigenvectors of the ML estimator $\hat{\mathbf{R}}_{i-1}^k$ are given by [2]

$$\hat{\lambda}_j = \bar{\lambda}_j \qquad \text{for} \quad j = 1, \ldots, k \tag{3.11}$$

$$\hat{\lambda}_j = \frac{1}{p-k} \sum_{l=k+1}^{p} \bar{\lambda}_l \qquad \text{for} \quad j = k+1, \ldots, p \tag{3.12}$$

$$\hat{\mathbf{v}}_j = \bar{\mathbf{v}}_j \qquad \text{for} \quad j = 1, \ldots, p \tag{3.13}$$

Using these eigenvalues and eigenvectors, $\hat{\mathbf{R}}_{i-1}^k$ is found for each model. Then, the PSC criterion is computed for all orders and the minimum is used to estimate the number of signals.

The method that is proposed here for the ML estimation of $\mathbf{R}^k$ uses the fact that the correlation matrix of the k-th model can be expressed as $\mathbf{R}^k = \mathbf{Q}^k + \sigma^2\mathbf{I}$, where $\mathbf{Q}^k$ is a positive definite matrix with rank $k < p$. The structure of $\mathbf{Q}^k$ is not exploited in this approach. The only information used is the multiplicity of the smallest eigenvalue. The drawback of the method is that when applied to coherent signals it does not estimate the true number of sources. However, it can be used successfully in noncoherent signal detection. We have compared this version of the PSC algorithm with the method of Wax and Kailath [55]. The results are presented in Section 3.8.

3.3. Subspace Decomposition

In order to handle coherent sources, we will make use of a subspace decomposition. Let us assume that the p-dimensional complex observation vector space is represented by $\mathbb{C}^p$. In the subspace decomposition approach, this vector space is decomposed into two orthogonal subspaces called the signal and the noise subspaces. The *signal subspace* is a subspace which is spanned by the column vectors of the location matrix $\mathbf{A}(\boldsymbol{\theta}^q)$. For noncoherent sources the column span of $\mathbf{A}(\boldsymbol{\theta}^q)$ coincides with the span of the eigenvectors of $\mathbf{R}^q$ corresponding to the q largest eigenvalues. Assuming that a one-to-one relationship exists between $\mathbf{A}(\boldsymbol{\theta}^q)$ and $\boldsymbol{\theta}^q$, an estimate of the signal subspace can be obtained through

estimation of the DOAs. The ML estimate of the signal subspace for a model of order k at time instant i is shown by $\mathbb{C}_{s_i}^k$. This is also the column span of the matrix $\mathbf{A}(\hat{\boldsymbol{\theta}}_i^k)$. The dimensionality of $\mathbb{C}_{s_i}^k$ is k. The *noise subspace*, $\mathbb{C}_{n_i}^k$, is the orthogonal complement of the signal subspace. The dimensionality of the noise subspace for a model with k signals is $(p-k)$. The true signal and noise subspaces are represented by $\mathbb{C}_s^q$ and $\mathbb{C}_n^q$, respectively. It should be noted that once the signal subspace is known, the corresponding DOAs can be determined uniquely. In Appendix A, we prove the following theorem on the uniqueness of the signal and noise subspace decomposition.

Theorem 3.1. *For a fixed $\boldsymbol{\theta}^q$ the subspace decomposition is unique and for any subspace decomposition there is a unique $\boldsymbol{\theta}^q$ that can generate that decomposition.* □

The projection matrices onto the signal and noise subspaces are given by

$$\mathbf{P}_s(\boldsymbol{\theta}^q) = \mathbf{A}(\boldsymbol{\theta}^q)\Big(\mathbf{A}^H(\boldsymbol{\theta}^q)\mathbf{A}(\boldsymbol{\theta}^q)\Big)^{-1}\mathbf{A}^H(\boldsymbol{\theta}^q) \tag{3.14}$$

$$\mathbf{P}_n(\boldsymbol{\theta}^q) = \mathbf{I} - \mathbf{P}_s(\boldsymbol{\theta}^q). \tag{3.15}$$

Using these matrices the observation vector can be decomposed into two orthogonal components

$$\mathbf{x} = \mathbf{P}_s(\boldsymbol{\theta}^q)\mathbf{x} + \mathbf{P}_n(\boldsymbol{\theta}^q)\mathbf{x} \tag{3.16}$$

$$= \mathbf{x}_s + \mathbf{x}_n. \tag{3.17}$$

Note that the component of the observation vector in the noise subspace $\mathbf{x}_n$ is due to the additive noise only and is independent of the signal and the component of noise in the signal subspace. Thus the correlation matrix of the array output can be represented as

$$\mathbf{R}^q = \mathbf{R}_s^q + \mathbf{R}_n^q \tag{3.18}$$

where

$$\mathbf{R}_s^q = \mathbf{P}_s(\boldsymbol{\theta}^q)\mathbf{R}^q\mathbf{P}_s(\boldsymbol{\theta}^q) \tag{3.19}$$

$$\mathbf{R}_n^q = \mathbf{P}_n(\boldsymbol{\theta}^q)\mathbf{R}^q\mathbf{P}_n(\boldsymbol{\theta}^q) \tag{3.20}$$

are the projections of the correlation matrix onto the signal and the noise subspaces, respectively. Using this formulation, the ML estimation of the correlation matrix can be formed by adding the ML estimates of $\mathbf{R}_s^q$ and $\mathbf{R}_n^q$.

Lemma 3.1. *Let* $\mathbf{A}$, $\mathbf{B}$ *be* $n \times n$ *Hermitian matrices orthogonal to each other such that* $\mathbf{A}^H\mathbf{B} = \mathbf{B}^H\mathbf{A} = \mathbf{0}$. *If the matrix* $\mathbf{C}$ *is given by*

$$\mathbf{C} = \mathbf{A} + \mathbf{B} \tag{3.21}$$

where $\mathbf{C}$ *is full rank, then*

$$|\mathbf{C}| = \zeta(\mathbf{A})\ \zeta(\mathbf{B}) \tag{3.22}$$

where $|.|$ *is the determinant, and* $\zeta(.)$ *represents the multiplication of the nonzero eigenvalues.*

Proof: Assume that $\mathbf{V}_a$ and $\mathbf{V}_b$ are the eigenvectors of $\mathbf{A}$ and $\mathbf{B}$ corresponding to nonzero eigenvalues. Then $\mathbf{C}$ can be written as

$$\begin{aligned}\mathbf{C} &= \mathbf{V}_a\Lambda_a\mathbf{V}_a^H + \mathbf{V}_b\Lambda_b\mathbf{V}_b^H \\ &= \mathbf{V}\begin{bmatrix}\Lambda_a & 0\\ 0 & \Lambda_b\end{bmatrix}\mathbf{V}^H\end{aligned} \tag{3.23}$$

where $\mathbf{V} = [\mathbf{V}_a\mathbf{V}_b]$. Since $\mathbf{C}$ is a full rank Hermitian matrix it is unitarily diagonalizable and the orthonormal matrix $\mathbf{V}$ is its eigenvalue matrix . Thus the determinant of $\mathbf{C}$ is equal to

$$|\mathbf{C}| = |\Lambda_a|\ |\Lambda_b|. \tag{3.24}$$

Note that $|\Lambda_a|$ and $|\Lambda_b|$ are the multiplication of the nonzero eigenvalues of $\mathbf{A}$ and $\mathbf{B}$. □

Since the matrices $\mathbf{R}_s^q$ and $\mathbf{R}_n^q$ satisfy the conditions of Lemma 3.1, the determinant of the true correlation matrix can be expressed as

$$|\mathbf{R}^q| = \zeta(\mathbf{R}_s^q)\ \zeta(\mathbf{R}_n^q) \tag{3.25}$$

where $\zeta(\mathbf{B})$ is the product of nonzero eigenvalues of $\mathbf{B}$.

3.4. Detection of coherent sources

The correlation matrix $\mathbf{Q}^k$ is of the form $\mathbf{A}(\boldsymbol{\theta}^k)\mathbf{S}^k\mathbf{A}^H(\boldsymbol{\theta}^k)$. For noncoherent signals $\mathbf{S}^k$ is a $k \times k$ Hermitian positive definite matrix of full rank. When some signals are coherent the rank of $\mathbf{S}^k$ is smaller than k. The structure of $\mathbf{Q}^k$ can be exploited to derive a method that can be applied to coherent signal scenarios. Note that for a full rank matrix $\mathbf{S}^k$, the column span of $\mathbf{A}(\boldsymbol{\theta}^k)$ is the same as the span of the eigenvectors of the correlation matrix $\mathbf{R}^k$ corresponding to the k largest eigenvalues. This is an important fact that permits use of the subspace decomposition method. In the sequel, we use this concept to develop a PSC algorithm that can be applied to coherent and noncoherent signal detection.

In the preceding section, we showed that the true correlation matrix of the array output can be represented as a sum of two orthogonal matrices $\mathbf{R}_s^q$ and $\mathbf{R}_n^q$. It was also proved that the determinant of $\mathbf{R}^q$ is equal to multiplication of the nonzero eigenvalues of $\mathbf{R}_s^q$ and $\mathbf{R}_n^q$. In practice the true correlation matrix is unknown and is estimated by the sample correlation matrix. Similarly, we can project the sample correlation matrix $\bar{\mathbf{R}}_{i-1}$ onto the signal and noise subspaces. The projected correlation matrices for the k-th model are shown as

$$\bar{\mathbf{R}}_{s_i}^k = \mathbf{P}_s(\boldsymbol{\theta}^k)\bar{\mathbf{R}}_i\mathbf{P}_s(\boldsymbol{\theta}^k) \tag{3.26}$$

$$\bar{\mathbf{R}}_{n_i}^k = \mathbf{P}_n(\boldsymbol{\theta}^k)\bar{\mathbf{R}}_i\mathbf{P}_n(\boldsymbol{\theta}^k). \tag{3.27}$$

We use these matrices to find the ML estimate of the correlation matrix for the k-th model.

The ML estimate of the correlation matrix for the k-th model and the $(i-1)$-th

snapshot can be represented by

$$\hat{\mathbf{R}}_{i-1}^{k} = \hat{\mathbf{R}}_{s_{i-1}}^{k} + \hat{\mathbf{R}}_{n_{i-1}}^{k} \tag{3.28}$$

where $\hat{\mathbf{R}}_{s_{i-1}}^{k}$ and $\hat{\mathbf{R}}_{n_{i-1}}^{k}$ are the ML estimates of the projection of the correlation matrices onto the signal and the noise subspaces. If $\hat{\boldsymbol{\theta}}_{i-1}^{k}$ is the ML estimator of the DOAs [2]

$$\hat{\mathbf{R}}_{s_{i-1}}^{k} = \bar{\mathbf{R}}_{s_{i-1}}^{k}. \tag{3.29}$$

With a similar method it is possible to show that $\hat{\mathbf{R}}_{n_{i-1}}^{k}$ has the same eigenvectors as $\bar{\mathbf{R}}_{n_{i-1}}^{k}$ and a single eigenvalue with multiplicity $(p-k)$ which is found from

$$\hat{\sigma}^2(\hat{\boldsymbol{\theta}}_{i-1}^{k}) = \frac{1}{p-k}\mathrm{tr}\bar{\mathbf{R}}_{n_{i-1}}^{k}. \tag{3.30}$$

Note that $\hat{\mathbf{R}}_{n_{i-1}}^{k}$ can be obtained by applying a linear transformation $\mathbf{T}_{i}^{k}$ on the matrix $\bar{\mathbf{R}}_{n_{i-1}}^{k}$ such as

$$\hat{\mathbf{R}}_{n_{i-1}}^{k} = \mathbf{T}_{i-1}^{k}\bar{\mathbf{R}}_{n_{i-1}}^{k} \tag{3.31}$$

$$\mathbf{T}_{i-1}^{k} = \bar{\mathbf{V}}_{n,p-k}\mathrm{diag}\left[\frac{\hat{\sigma}^2(\hat{\boldsymbol{\theta}}_{i-1}^{k})}{\lambda_j(\bar{\mathbf{R}}_{n_{i-1}})}\right]\bar{\mathbf{V}}_{n,p-k}^{H} \tag{3.32}$$

where $\lambda_j(\bar{\mathbf{R}}_{n_{i-1}})$, $j = 1,\ldots,p-k$, are the nonzero eigenvalues of $\bar{\mathbf{R}}_{n_{i-1}}$, and $\bar{\mathbf{V}}_{n,p-k}$ is the $p \times (p-k)$ matrix of corresponding eigenvectors. The diag[.] is a representation for a diagonal matrix that has been formed by the elements in the brackets.

The ML estimate of the determinant of the correlation matrix is determined by the multiplication of the nonzero eigenvalues of its projected components

$$|\hat{\mathbf{R}}_{i-1}^{k}| = \zeta(\hat{\mathbf{R}}_{s_{i-1}}^{k})\,\zeta(\hat{\mathbf{R}}_{n_{i-1}}^{k}) \tag{3.33}$$

where from (3.29) and (3.30)

$$\zeta(\hat{\mathbf{R}}_{s_{i-1}}^{k}) = \zeta(\bar{\mathbf{R}}_{s_{i-1}}^{k}) \tag{3.34}$$

$$\zeta(\hat{\mathbf{R}}^k_{n_{i-1}}) = \left(\hat{\sigma}^2(\hat{\boldsymbol{\theta}}^k_{i-1})\right)^{p-k}. \tag{3.35}$$

Using the definition of the PSC and these results, we have

$$\begin{aligned}
\mathrm{PSC}_k(N) &= \sum_{i=1}^{N}\left(\log|\hat{\mathbf{R}}^k_{i-1}| + \mathbf{x}_i^H[\hat{\mathbf{R}}^k_{i-1}]^{-1}\mathbf{x}_i\right) \\
&= \sum_{i=1}^{N}\Big[\log\zeta(\hat{\mathbf{R}}^k_{s_{i-1}})\ \ \log\zeta(\hat{\mathbf{R}}^k_{n_{i-1}}) + \mathbf{x}_i^H(\hat{\mathbf{R}}^k_{s_{i-1}} + \hat{\mathbf{R}}^k_{n_{i-1}})^{-1}\mathbf{x}_i\Big] \\
&= \sum_{i=1}^{N}\Big[\log\zeta(\bar{\mathbf{R}}^k_{s_{i-1}}) + (p-k)\log(\frac{1}{p-k}\mathrm{tr}\bar{\mathbf{R}}^k_{n_{i-1}}) \\
&\qquad\qquad + \mathbf{x}_i^H(\bar{\mathbf{R}}^k_{s_{i-1}} + \mathbf{T}^k_{i-1}\bar{\mathbf{R}}^k_{n_{i-1}})^{-1}\mathbf{x}_i\Big].
\end{aligned} \tag{3.36}$$

Unlike the MDL method which has separate terms to account for coding the data and coding the model, the PSC has no such easily indentifiable terms. It is seen that the computation of PSC depends on the estimation of the angles of arrival $\hat{\boldsymbol{\theta}}^k_{i-1}$. In the original version of the PSC algorithm, the ML estimate of the parameter vector is used. In the following section, we propose a suboptimal method for DOA estimation which is computationally more attractive. In a later section we discuss how PSC works for coherent source detection.

3.5. Estimation of the DOAs

To obtain $\hat{\boldsymbol{\theta}}^k_{i-1}$ an estimation problem has to be solved. The ML estimator of the DOAs for the stochastic signal model is given by

$$\hat{\boldsymbol{\theta}}^k_{i-1} = \arg\min_{\boldsymbol{\psi}^k}\left\{\log\left[\zeta\Big(\mathbf{P}_s(\boldsymbol{\psi}^k)\bar{\mathbf{R}}_{i-1}\mathbf{P}_s(\boldsymbol{\psi}^k)\Big)\ \left(\hat{\sigma}^2(\boldsymbol{\psi}^k)\right)^{p-k}\right]\right\}. \tag{3.37}$$

This is a multivariate nonlinear optimization problem and computationally expensive. To reduce the computational complexity, we use the alternating projection method [64] with a cost function based on ML estimation of a deterministic signal model. In each step of this algorithm an optimization is performed to determine the best value for one parameter

element while keeping the rest of the elements constant. Therefore, the ML estimation is decomposed into several one-variable nonlinear optimization problems. A suitable cost function for the alternating projection algorithm is given by

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{i-1}^{k} &= \arg\min_{\boldsymbol{\psi}^k} \{\log(\hat{\sigma}^2(\boldsymbol{\psi}^k))\} \\ &= \arg\min_{\boldsymbol{\psi}^k} \operatorname{tr}[\mathbf{P}_n(\boldsymbol{\psi}^k)\bar{\mathbf{R}}_{i-1}] \end{aligned} \tag{3.38}$$

which is the ML estimator of the deterministic signal model. We use (3.38) for DOA estimation. Although with this choice the optimality of the PSC algorithm has been impaired, it can still be successfully applied to coherent signal detection.

To give more insight into the algorithm, we present a simple example. Suppose that we seek the parameter vector for a model of order 3. We use the value of $\hat{\boldsymbol{\theta}}_{i-1}^{3}$ to determine $\hat{\boldsymbol{\theta}}_{i}^{3}$. Let us assume $\hat{\boldsymbol{\theta}}_{i-1}^{3} = [a_{i-1}\ b_{i-1}\ c_{i-1}]^T$. To obtain $\hat{\boldsymbol{\theta}}_{i}^{3}$, a minimization problem is solved 3 times to get a_i, b_i, and c_i. At each step we use the latest computed values of the elements of the parameter vector. In other words, the optimization is accomplished in the following manner

$$\begin{aligned} a_i &= \arg\min_{a\in D_a} C(\ a\ , b_{i-1}\ , c_{i-1}\ ) \\ b_i &= \arg\min_{b\in D_b} C(\ a_i\ ,\ \ b\ \ , c_{i-1}\ ) \\ c_i &= \arg\min_{c\in D_c} C(\ a_i\ ,\ \ b_i\ \ ,\ \ c\ \ ) \end{aligned}$$

where C is the cost function, and $D_a = [-\frac{\pi}{2}, b_{i-1}]$, $D_b = [a_i\ ,\ c_{i-1}]$, and $D_c = [b_i\ ,\ \frac{\pi}{2}]$ are the intervals of search. It is seen that for a model of order k, the cost function is optimized k times to get $\hat{\boldsymbol{\theta}}^{k}$. Here, we use one iteration for each dimension at each sample time, but the number of iterations can be increased to obtain more precise estimates of the DOAs. In the following theorem, we study the asymptotic behavior of the estimator (3.38) for true number of signals.

Theorem 3.2. *For $k = q$, the estimator* (3.38) *is consistent.*

Proof: See Appendix B. A proof of a related theorem is given in [60] which is based on partial convergence of the parameter vector $\hat{\theta}^k$ to θ^q for two cases of $q \le k \le p/2$ and

$k < q \leq p/2$.

Since the information theoretic criteria are sensitive to the estimate of the parameter vector, asymptotic bias in the DOA estimation degrades the performance of the detector. The alternating projection algorithm usually converges to a local minimum. If this method is applied on a cost function that is determined by the data collected in a batch of snapshots, such as in the MDL algorithm, a local minimum is found. If this local minimum is far from the global optimum point, the performance of detection will deteriorate. In an iterative method of estimation, since the location of the local minimum changes with every new sample of data, there is a possibility that the estimator moves out of the local minimum with the observation of next sample. For all the trials in our simulations, the recursive estimator converged to the true DOAs within a few sample times.

3.6. Consistency of the PSC algorithm

The PSC criterion for a model of order k up to time N is described by

$$\mathrm{PSC}_k(N) = \sum_{i=1}^{N} \Big(\log|\hat{\mathbf{R}}_{i-1}^k| + \mathbf{x}_i^H [\hat{\mathbf{R}}_{i-1}^k]^{-1} \mathbf{x}_i \Big) \tag{3.39}$$

where $\hat{\mathbf{R}}_{i-1}^k$ is the ML estimate of the correlation matrix. In this section, we prove the consistency of the PSC detector. We show that the PSC algorithm asymptotically estimates the true order of the system. Mathematically,

$$\lim_{N\to\infty} \mathrm{PSC}_q(N) < \lim_{N\to\infty} \mathrm{PSC}_k(N) \qquad \text{for } k \neq q. \tag{3.40}$$

According to the law of large numbers the sum (3.39) converges to the addition of the expected values of its terms. We seek the asymptotic mean value of the two terms in the summation. We consider two cases for the model order k.

- *Case I.* $k < q$

As $i \to \infty$,

$$\hat{\mathbf{R}}_{i-1}^k \to \hat{\mathbf{R}}^k = \mathbf{R}_s^k + \hat{\mathbf{R}}_n^k \tag{3.41}$$

where $\mathbf{R}_s^k$ is the projection of $\mathbf{R}^q$ onto the signal subspace with dimensionality k and $\hat{\mathbf{R}}_n^k$ is the ML estimate of the projection of the correlation matrix onto the noise subspace. If $\mathbf{R}_n^k$ is the projection of $\mathbf{R}^q$ onto the noise subspace with dimensionality $(p-k)$, then $\hat{\mathbf{R}}_n^k$ will be a matrix with the same eigenvectors as $\mathbf{R}_n^k$ and an eigenvalue $\hat{\sigma}_k^2$ with multiplicity $(p-k)$ which is obtained from

$$\hat{\sigma}_k^2 = \frac{1}{p-k}\mathrm{tr}\mathbf{R}_n^k. \tag{3.42}$$

First, we discuss the properties of the second term for a large number of observations. Let the random variable b_i be

$$b_i = \mathbf{x}_i^H[\hat{\mathbf{R}}_{i-1}^k]^{-1}\mathbf{x}_i \tag{3.43}$$

$$= \mathrm{tr}([\hat{\mathbf{R}}_{i-1}^k]^{-1}\mathbf{x}_i\mathbf{x}_i^H). \tag{3.44}$$

Since $\mathbf{x}_i$ and $\hat{\mathbf{R}}_{i-1}^k$ are independent, taking expectation of both sides of (3.44) gives

$$\mathrm{E}b_i = \mathrm{tr}(\mathrm{E}[\hat{\mathbf{R}}_{i-1}^k]^{-1}\mathbf{R}^q). \tag{3.45}$$

Let $i \to \infty$, then

$$\lim_{i\to\infty} \mathrm{E}b_i = \mathrm{tr}(\lim_{i\to\infty} \mathrm{E}[\hat{\mathbf{R}}_{i-1}^k]^{-1}\mathbf{R}^q). \tag{3.46}$$

Assuming $[\hat{\mathbf{R}}_i^k]^{-1}$ converges to a fixed matrix, the order of the expectation and the limit can be exchanged

$$\begin{aligned}\lim_{i\to\infty} \mathrm{E}b_i &= \mathrm{tr}(\mathrm{E}\lim_{i\to\infty}[\hat{\mathbf{R}}_{i-1}^k]^{-1}\mathbf{R}^q) \\ &= \mathrm{tr}([\hat{\mathbf{R}}^k]^{-1}\mathbf{R}^q)\end{aligned} \tag{3.47}$$

Lemma 3.2. *Let* $\mathbf{A}$, $\mathbf{B}$ *be* $n \times n$ *Hermitian matrices orthogonal to each other such that* $\mathbf{A}^H\mathbf{B} = \mathbf{B}^H\mathbf{A} = \mathbf{0}$. *If the matrix* $\mathbf{C}$ *is given by*

$$\mathbf{C} = \mathbf{A} + \mathbf{B} \tag{3.48}$$

where $\mathbf{C}$ *is full rank, then*

$$\mathbf{C}^{-1} = \mathbf{A}^{\dagger} + \mathbf{B}^{\dagger} \tag{3.49}$$

where $\mathbf{A}^{\dagger}$ *and* $\mathbf{B}^{\dagger}$ *are the pseudo-inverse of* $\mathbf{A}$ *and* $\mathbf{B}$.

Proof: Assume that $\mathbf{V}_a$ and $\mathbf{V}_b$ are the eigenvectors of $\mathbf{A}$ and $\mathbf{B}$ corresponding to nonzero eigenvalues. Then $\mathbf{C}$ can be written as

$$\begin{aligned} \mathbf{C} &= \mathbf{V}_a \Lambda_a \mathbf{V}_a^H + \mathbf{V}_b \Lambda_b \mathbf{V}_b^H \\ &= \mathbf{V} \begin{bmatrix} \Lambda_a & 0 \\ 0 & \Lambda_b \end{bmatrix} \mathbf{V}^H \end{aligned} \tag{3.50}$$

where $\mathbf{V} = [\mathbf{V}_a \mathbf{V}_b]$. Since $\mathbf{C}$ is a full rank Hermitian matrix, it is unitarily diagonalizable and the orthonormal matrix $\mathbf{V}$ is its eigenvalue matrix . Thus the inverse of $\mathbf{C}$ is equal to

$$\mathbf{C}^{-1} = \mathbf{V} \begin{bmatrix} \Lambda_a^{-1} & 0 \\ 0 & \Lambda_b^{-1} \end{bmatrix} \mathbf{V}^H. \tag{3.51}$$

Thus

$$\mathbf{C}^{-1} = \mathbf{V}_a \Lambda_a^{-1} \mathbf{V}_a^H + \mathbf{V}_b \Lambda_b^{-1} \mathbf{V}_b^H. \tag{3.52}$$

And the proof is complete. □

Note that $\hat{\mathbf{R}}^k$ is given by

$$\hat{\mathbf{R}}^k = \mathbf{P}_s(\theta^k) \mathbf{R}^q \mathbf{P}_s(\theta^k) + \mathbf{T}^k \mathbf{P}_n(\theta^k) \mathbf{R}^q \mathbf{P}_n(\theta^k). \tag{3.53}$$

Assume

$$\mathbf{P}_s(\theta^k) \mathbf{R}^q \mathbf{P}_s(\theta^k) = \mathbf{V}_s \Lambda_s \mathbf{V}_s^H \tag{3.54}$$

$$\mathbf{P}_n(\theta^k) \mathbf{R}^q \mathbf{P}_n(\theta^k) = \mathbf{V}_n \Lambda_n \mathbf{V}_n^H \tag{3.55}$$

Then

$$\mathbf{T}^k \mathbf{P}_n(\theta^k) \mathbf{R}^q \mathbf{P}_n(\theta^k) = \hat{\sigma}_k^2 \mathbf{V}_n \mathbf{V}_n^H \tag{3.56}$$

where $\hat{\sigma}_k^2$ is given by (3.42). From this (3.53) can be written as

$$\mathbf{P}_s(\theta^k)\mathbf{R}^q\mathbf{P}_s(\theta^k) + \mathbf{T}^k\mathbf{P}_n(\theta^k)\mathbf{R}^q\mathbf{P}_n(\theta^k) = [\mathbf{V}_s\mathbf{V}_n]\begin{bmatrix} \Lambda_s & 0 \\ 0 & \hat{\sigma}_k^2\mathbf{I} \end{bmatrix}\begin{bmatrix} \mathbf{V}_s^H \\ \mathbf{V}_n^H \end{bmatrix}. \tag{3.57}$$

Using these results (3.47) becomes

$$\mathrm{tr}\left\{[\mathbf{V}_s\mathbf{V}_n]\begin{bmatrix} \Lambda_s^{-1} & 0 \\ 0 & \hat{\sigma}_k^{-2}\mathbf{I} \end{bmatrix}\begin{bmatrix} \mathbf{V}_s^H \\ \mathbf{V}_n^H \end{bmatrix}\mathbf{R}^q\right\} \tag{3.58}$$

Note that

$$\mathbf{P}_s(\theta^k) = \mathbf{V}_s\mathbf{V}_s^H \tag{3.59}$$

$$\mathbf{P}_n(\theta^k) = \mathbf{V}_n\mathbf{V}_n^H. \tag{3.60}$$

Thus

$$\mathbf{V}_s^H\mathbf{R}^q\mathbf{V}_s = \Lambda_s \tag{3.61}$$

$$\mathbf{V}_n^H\mathbf{R}^q\mathbf{V}_n = \Lambda_n \tag{3.62}$$

$$\mathbf{V}_n^H\mathbf{R}^q\mathbf{V}_s = \mathbf{V}_s^H\mathbf{R}^q\mathbf{V}_n = 0 \tag{3.63}$$

Using these equalities (3.58) can be written as

$$\mathrm{tr}([\hat{\mathbf{R}}^k]^{-1}\mathbf{R}^q) = \mathrm{tr}\left\{\begin{bmatrix} \Lambda_s^{-1} & 0 \\ 0 & \hat{\sigma}_k^{-2}\mathbf{I} \end{bmatrix}\begin{bmatrix} \Lambda_s & 0 \\ 0 & \Lambda_n \end{bmatrix}\right\} = p. \tag{3.64}$$

This equality shows that the effect of the second term in (3.39) fades as the number of observations increases. Thus consistency of the PSC depends on the asymptotic characteristics of the first term. Note that for a limited number of observations the second term cannot be removed from the PSC algorithm.

For a configuration with 3 uncorrelated sources arriving at a linear array of 8 sensors we found the second term of the PSC algorithm. The results are averaged with respect

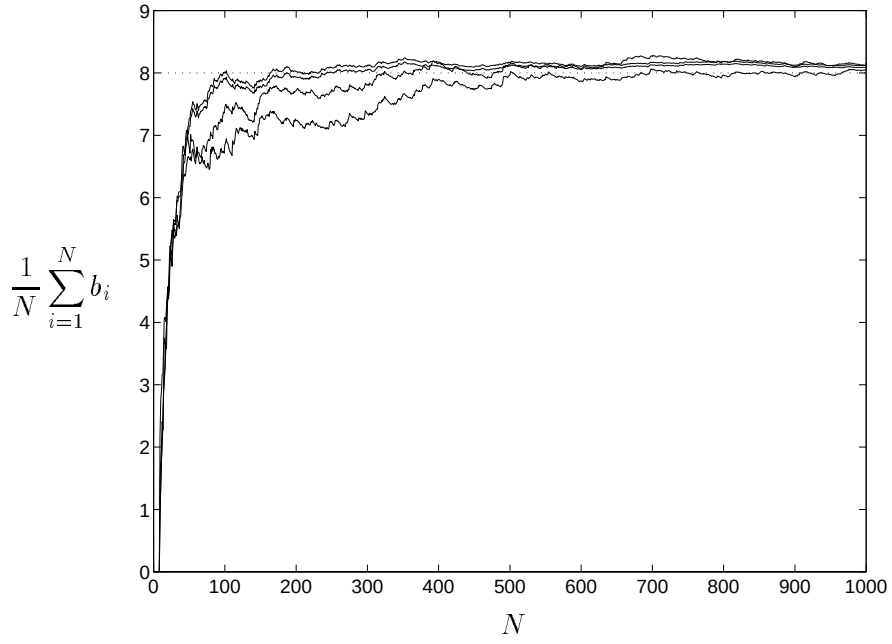

Fig. 3.1 The averaged second term of the PSC algorithm for a linear array of 8 sensors exposed to 3 uncorrelated far-field planar wavefronts at 10, 15, and 20 degrees. Each curve represents a different model order (orders, 1, 2, 3, 4).

to the number of observations and are depicted in Fig. 3.1. For all models it is seen that $\frac{1}{N}\sum_{i=1}^{N} b_i$ approaches 8 which is the number of sensors.

The first term of the summand in (3.39) can be shown as

$$a_i = \log(\zeta(\bar{\mathbf{R}}^k_{s_{i-1}})[\hat{\sigma}^2(\hat{\boldsymbol{\theta}}^k_{i-1})]^{p-k}) \tag{3.65}$$

where $\hat{\sigma}^2(\hat{\boldsymbol{\theta}}^k_{i-1}) = \dfrac{1}{p-k}\mathrm{tr}\bar{\mathbf{R}}^k_{n_{i-1}}$. Since the sample correlation matrix of the array asymptotically approaches the true correlation matrix, as $i \longrightarrow \infty$ we get

$$\bar{\mathbf{R}}^k_{s_{i-1}} \longrightarrow \mathbf{R}^k_s = \mathbf{P}_s(\hat{\boldsymbol{\theta}}^k)\mathbf{R}^q\mathbf{P}_s(\hat{\boldsymbol{\theta}}^k) \tag{3.66}$$

$$\bar{\mathbf{R}}^k_{n_{i-1}} \longrightarrow \mathbf{R}^k_n = \mathbf{P}_n(\hat{\boldsymbol{\theta}}^k)\mathbf{R}^q\mathbf{P}_n(\hat{\boldsymbol{\theta}}^k). \tag{3.67}$$

In Theorem 3.2 we proved that the estimator (3.38) is consistent. Thus the determinant

of the true correlation matrix is

$$\begin{aligned} |\mathbf{R}^q| &= \zeta(\mathbf{R}_s^q)\ \zeta(\mathbf{R}_n^q) \\ &= (\prod_{i=1}^{q} \lambda_i)(\sigma^2)^{p-q} \end{aligned} \tag{3.68}$$

where the λ_i are nonzero eigenvalues of $\mathbf{R}_s^q$ and σ^2 is the noise variance. For a model of order k we have

$$\zeta(\mathbf{R}_s^k)\ \zeta(\mathbf{R}_n^k) = \zeta\Big(\mathbf{P}_s(\hat{\boldsymbol{\theta}}^k)\mathbf{R}^q\mathbf{P}_s(\hat{\boldsymbol{\theta}}^k)\Big)\ \zeta\Big(\mathbf{P}_n(\hat{\boldsymbol{\theta}}^k)\mathbf{R}^q\mathbf{P}_n(\hat{\boldsymbol{\theta}}^k)\Big). \tag{3.69}$$

The parameter vector is found by minimizing the trace of $(\mathbf{P}_n(\hat{\boldsymbol{\theta}}^k)\mathbf{R}^q\mathbf{P}_n(\hat{\boldsymbol{\theta}}^k)\Big)$. In the best situation, the projector $\mathbf{P}_n(\hat{\theta}^k)$ is such that it chooses a parameter vector that projects $\mathbf{R}^q$ onto the subspace spanned by its smallest eigenvalues. In such a case, we have

$$\zeta(\mathbf{R}_s^k) = \prod_{i=1}^{k} \lambda_i \tag{3.70}$$

$$\zeta(\mathbf{R}_n^k) = (\prod_{i=k+1}^{q} \lambda_i)(\sigma^2)^{p-q}. \tag{3.71}$$

In the PSC algorithm, we used the ML estimates of $\mathbf{R}_s^k$ and $\mathbf{R}_n^k$ which are represented by $\hat{\mathbf{R}}_s^k$ and $\hat{\mathbf{R}}_n^k$. We know that

$$\zeta(\hat{\mathbf{R}}_s^k) = \prod_{i=1}^{k} \lambda_i \tag{3.72}$$

$$\zeta(\hat{\mathbf{R}}_n^k) = (\hat{\sigma}^2)^{p-k} \tag{3.73}$$

where from (3.71)

$$\hat{\sigma}^2 = \frac{1}{p-k}\Big(\sum_{i=k+1}^{q} \lambda_i + (p-q)\sigma^2\Big). \tag{3.74}$$

The next step is to show that

$$(\hat{\sigma}^2)^{p-k} > (\prod_{i=k+1}^{q} \lambda_i)(\sigma^2)^{p-q}. \tag{3.75}$$

Lemma 3.3. *For positive scalars* $\alpha_i, i = 1, \ldots, n$, *with* $\sum_{i=1}^{n} \alpha_i = K$, *where* K *is a constant, the maximum of* $\prod_{i=1}^{n} \alpha_i$ *is achieved for* $\alpha_i = \frac{K}{n}$, $i = 1, \ldots, n$.

Using Lemma 3.3, (3.75) is satisfied. Thus we can conclude that

$$\log(\prod_{i=1}^{q} \lambda_i) + (p-q)\log(\sigma^2) < \log(\prod_{i=1}^{k} \lambda_i) + (p-k)\log(\mu^2) \tag{3.76}$$

$$\log \zeta(\mathbf{R}_s^q) + \log \zeta(\hat{\mathbf{R}}_n^q) < \log \zeta(\mathbf{R}_s^k) + \log \zeta(\hat{\mathbf{R}}_n^k). \tag{3.77}$$

The two sides of this inequality are the asymptotic values of the first term of the PSC criterion for the models of order q and k. Thus we conclude that (3.40) is satisfied for $k < q$.

- *Case II.* $k > q$

For this case, we use the following lemma [10].

Lemma 3.4. (Davisson's Formula) *For a probability density function* $f(\mathbf{x}_n|\boldsymbol{\theta}^q)$ *where* $\boldsymbol{\theta}^q$ *is the true parameter vector with dimension* q, *we have*

$$E_{\boldsymbol{\theta}^q}[-\log f(\mathbf{x}_i|\hat{\boldsymbol{\theta}}_{i-1}^q) + \log f(\mathbf{x}_i|\boldsymbol{\theta}^q)] = \frac{q}{2i}(1 + o(1)) \tag{3.78}$$

where $o(1) \to 0$ *as* $i \to \infty$ *and* $\hat{\boldsymbol{\theta}}_{i-1}^q$ *is the ML estimator of* $\boldsymbol{\theta}^q$ *based on the observations up to time* $(i-1)$.

The cumulative effect of the parameter uncertainty is then given by

$$E_{\boldsymbol{\theta}^q} \sum_{i=1}^{N} [-\log f(\mathbf{x}_i|\hat{\boldsymbol{\theta}}_{i-1}^q) + \log f(\mathbf{x}_i|\boldsymbol{\theta}^q)] = \sum_{i=1}^{N} \frac{q}{2i}(1 + o(1)) \tag{3.79}$$

$$= \frac{q}{2}(1 + o(1)) \log N. \tag{3.80}$$

Assume that the PSC of two models of order q and $k > q$ with the parameter vectors $\boldsymbol{\theta}^q$ and $\boldsymbol{\psi}^k$ are compared. The Kullback-Leibler distance between the two models is shown

by

$$I = \int f(\mathbf{x}|\boldsymbol{\theta}^q) \log \frac{f(\mathbf{x}|\boldsymbol{\theta}^q)}{f(\mathbf{x}|\boldsymbol{\psi}^k)} d\mathbf{x} \tag{3.81}$$

which is zero if and only if $f(\mathbf{x}|\boldsymbol{\theta}^q) = f(\mathbf{x}|\boldsymbol{\psi}^k)$. Now we expand $\boldsymbol{\theta}^q$ to a k-dimensional parameter vector $\boldsymbol{\psi}^k$ by introducing $(k-q)$ new sources at arbitrary angles such that the column vectors of $\mathbf{A}(\boldsymbol{\psi})$ are linearly independent. In such a case q components of $\boldsymbol{\theta}^q$ and $\boldsymbol{\psi}^k$ are equal. Suppose that the power of the auxiliary sources is zero. Then the probability distribution functions for the two models are identical,

$$f(\mathbf{x}|\boldsymbol{\theta}^q) = f(\mathbf{x}|\boldsymbol{\psi}^k). \tag{3.82}$$

With this assumption the Kullback-Leibler distance between the true model and the model with the parameter vector $\boldsymbol{\psi}^k$ is zero. The probability distribution function $f(\mathbf{x}|\boldsymbol{\psi}^k)$ can also be considered as the true generating class with a parameter vector which has a higher dimensionality.

The PSC of each model is a metric that represents the distance from the true generating class. By computing the PSC for the two models we get the codelengths

$$\mathrm{PSC}_q(N) = \sum_{i=1}^{N} -\log f(\mathbf{x}_i|\hat{\boldsymbol{\theta}}^q_{i-1}) \tag{3.83}$$

$$\mathrm{PSC}_k(N) = \sum_{i=1}^{N} -\log f(\mathbf{x}_i|\hat{\boldsymbol{\psi}}^k_{i-1}) \tag{3.84}$$

which can also be written as

$$\mathrm{PSC}_q(N) = \sum_{i=1}^{N} [-\log f(\mathbf{x}_i|\hat{\boldsymbol{\theta}}^q_{i-1}) + \log f(\mathbf{x}_i|\boldsymbol{\theta}^q)] - \sum_{i=1}^{N} \log f(\mathbf{x}_i|\boldsymbol{\theta}^q) \tag{3.85}$$

$$\mathrm{PSC}_k(N) = \sum_{i=1}^{N} [-\log f(\mathbf{x}_i|\hat{\boldsymbol{\psi}}^k_{i-1}) + \log f(\mathbf{x}_i|\boldsymbol{\psi}^k)] - \sum_{i=1}^{N} \log f(\mathbf{x}_i|\boldsymbol{\psi}^k) \tag{3.86}$$

Using (3.82) and Lemma 3.4 we have

$$E(\mathrm{PSC}_k(N) - \mathrm{PSC}_q(N)) = \frac{k-q}{2} \log N. \tag{3.87}$$

The empirical procedure is to drop the expectation and compute $\mathrm{PSC}_k(N) - \mathrm{PSC}_q(N)$ as

$$\lim_{N\to\infty} \frac{\mathrm{PSC}_k(N) - \mathrm{PSC}_q(N)}{\log N} = \frac{k-q}{2} \qquad \text{almost surely.} \tag{3.88}$$

Since $k > q$, the proof is complete.

3.7. More discussion on the coherent signal detection

A technique for dealing with coherent signals is to use the spatial smoothing [31]. There are two shortcomings with this method. First, the spatial smoothing is only applicable to uniform linear arrays. Second, the spatial smoothing is based on decomposing the array into subarrays and hence it does not use the whole available array aperture. The PSC algorithm can be applied to a general array geometry and uses the whole array aperture. In this section we discuss how our method detects coherent sources.

A wrong number of signals might be detected if there is an ambiguity in separating the signal and noise subspaces. The methods that choose the order of the system by decomposing the set of the eigenvalues of the correlation matrix into the signal and noise eigenvalues, produce incorrect results when applied to coherent source scenarios. For such methods, there is not a unique way to choose a q-dimensional signal subspace. The method that we have introduced here does not suffer from this shortcoming since it does not encounter ambiguity in the signal and noise subspace decomposition.

It is possible to show that the estimator (3.38) successfully finds the true parameter vector even in the situations where the sources are coherent. Consider any two parameter vectors θ and ψ with arbitrary dimensions k and k'. Assume that $\theta \cap \psi = \phi$, where ϕ is a vector of dimension m. Let $\mathcal{C}_s^k(\theta), \mathcal{C}_s^{k'}(\psi)$ and $\mathcal{C}_s^m(\phi)$ be the subspaces spanned by $\mathbf{A}(\theta), \mathbf{A}(\psi)$ and $\mathbf{A}(\phi)$, respectively.

Lemma 3.5. *For $k + k' - m < p$, the intersection of the two subspaces $\mathcal{C}_s^k(\theta)$ and $\mathcal{C}_s^{k'}(\psi)$ is identical to $\mathcal{C}_s^m(\phi)$.*

Proof: Take any $\mathbf{e} \in \mathcal{C}_s^k(\theta) \cap \mathcal{C}_s^{k'}(\psi)$. Then,

$$\mathbf{A}(\theta)\mathbf{b}^k = \mathbf{e} \quad \Longrightarrow \quad \sum_{i=1}^{k} \mathbf{a}(\theta_i) b_i = \mathbf{e} \tag{3.89}$$

$$\mathbf{A}(\psi)\mathbf{d}^{k'} = \mathbf{e} \quad \Longrightarrow \quad \sum_{i=1}^{k'} \mathbf{a}(\psi_i) d_i = \mathbf{e} \tag{3.90}$$

Without loss of generality suppose that first m components of θ and ψ are equal, i.e. $\theta_i = \psi_i,\ i = 1, \ldots, m,\ m \leq k,\ m \leq k'$. Then,

$$\sum_{i=1}^{m} \mathbf{a}(\theta_i)(b_i - d_i) + \sum_{i=m+1}^{k} \mathbf{a}(\theta_i) b_i - \sum_{i=m+1}^{k'} \mathbf{a}(\psi_i) d_i = 0. \tag{3.91}$$

Independence of the steering vectors for $k + k' - m < p$ implies

$$b_i = d_i \qquad i = 1, \ldots, m \tag{3.92}$$

$$b_i = 0 \qquad i = m+1, \ldots, k \tag{3.93}$$

$$d_i = 0 \qquad i = m+1, \ldots, k'. \tag{3.94}$$

In words, the vector $\mathbf{e}$ is in the space spanned by the steering vectors generated by the parameter vector $\phi = \{\theta_i,\ i = 1, \ldots, m\}$. □

Suppose that none of the sources has zero power. This assumption implies that the signal vector $\mathbf{A}(\theta^q)\mathbf{s}$ is not in a subspace of $\mathcal{C}_s^q$ which is spanned by a subset of steering vectors. This is the key point to the uniquely estimation of the parameter vector θ^q.

Now suppose that the model of order q is investigated. For coherent source case, the signal vector $\mathbf{A}(\theta^q)\mathbf{s}$ spans a one-dimensional subspace of the signal space. And as mentioned earlier $\mathbf{A}(\theta^q)\mathbf{s}$ cannot be spanned by any subset of the steering vectors. In such a case, rank of $\mathbf{Q} = \mathbf{A}(\theta^q)\mathbf{S}\mathbf{A}^H(\theta^q)$ is equal to one and there are $p - 1$ eigenvectors which span a subspace that is orthogonal to $\mathbf{A}(\theta^q)\mathbf{s}$. Estimator (3.38) finds a $\hat{\theta}$ which minimizes the trace of the projection of the correlation matrix into the noise subspace $\mathbf{R}_n^q$. In coherent case, every subspace with dimension $p - q$ and orthogonal to $\mathbf{A}(\theta^q)\mathbf{s}$ can be a candidate for the solution of the optimization problem. It is obvious that the true value of

the parameter vector is a solution to the minimization problem. Now, suppose that there is a $\psi^q \neq \theta^q$ which also satisfies (3.38). If ψ^q and θ^q contain m identical components and if $2q - m < p$, then, by Lemma 3.5, $\mathbf{A}(\theta^q)\mathbf{s}$ has to be in a subspace which is spanned by m columns of the steering matrix. And this is in contradiction to the assumption that the sources do not have zero powers.

It is seen that $2q - m < p$ is a sufficient condition for unique determination of the signals. The worst value for q is obtained when $m = 0$. Hence, the maximum number of sources which can surely be detected by a linear array is equal to half the number of sensors. This is in agreement with the results of [58].

The projection of the correlation matrix into the signal and noise subspaces forces the observation vectors to obtain a special form which is dictated with the location matrix. In other words, the signal part of the correlation matrix should be in the span of a location matrix say $\mathbf{A}(\theta')$. Since the estimator is consistent, for a model of order q we have $\theta' = \theta^q$. This is the unique DOA that can generate the signal component of the observation vector. This is to say that there is no ambiguity in decomposing the observation space into the signal and the noise subspaces. Thus the part of the observation vector which is coded with the PSC criterion and corresponds to the true component of the signal. And hence the method can be used for coherent source localization.

3.8. Simulation Results

Consider a linear array of 8 sensors exposed to 2 planar wavefronts arriving from 10 and 15 degrees. The spacing between two consecutive sensors is half the wave length. We perform 100 Monte-Carlo runs for different signal-to-noise ratios and count the number of times that the MDL and the PSC detect the true number of signals. For each trial 100 snapshots are processed. In the first example, we assume that the sources are uncorrelated. The results of the simulation for 100 independent trials are represented in Table 3.1 and Fig. 5.1. MDL1 is the MDL algorithm of Wax and Kailath [55], PSC1 is the method of Section 3.2, MDL2 is the MDL method of Wax [53], and PSC2 is the stochastic complexity method given by (3.36). Both MDL2 and PSC2 use the signal subspace decomposition and are

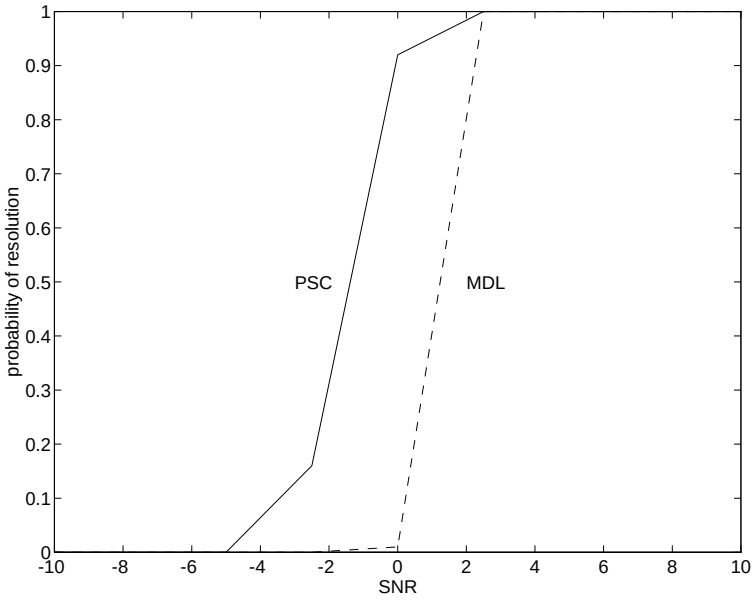

Fig. 3.2 The probability of resolution for the PSC and MDL algorithms using an array of eight sensors exposed to two uncorrelated far field signals.

computationally more expensive than the MDL1 and PSC1. However, since they exploit more information about the structure of the correlation matrix, they perform better.

Table 3.2 represent the results of detection for completely correlated sources. For coherent source scenarios MDL2 shows a smaller SNR threshold than the PSC2. However, the probability of resolution for small SNR in the PSC2 algorithm is larger than the MDL2 method. For this case, PSC1 detects only one signal.

To compare the sensitivity of the two methods for uncertainty in the DOA estimation, we consider a single source scenario. Assume that a planar wavefront is arriving from 10 degrees at the same linear array. The array output is observed for 100 snapshots. The PSC and the MDL criteria are computed for a model of order 1 as a function of the DOA estimates. The results are depicted in Fig. 3.3. The MDL algorithm has a sharper slope in the vicinity of the true DOA which makes it be more sensitive to the uncertainty in the DOA estimate than the PSC.

It was stated earlier that the PSC method can detect changes in the number of signals. To study the performance of the PSC algorithm in a nonstationary environment, we

detection method	k	SNR (dB)									
		−12.5	−10	−7.5	−5	−2.5	0	2.5	5	7.5	10
MDL1	1	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	1	100	100	100	100
	3	100	100	100	100	100	99	0	0	0	0
PSC1	1	100	100	100	100	84	8	0	0	0	0
	2	0	0	0	0	16	92	100	100	100	100
	3	0	0	0	0	0	0	0	0	0	0
MDL2	1	100	98	91	55	6	0	0	0	0	0
	2	0	2	9	45	94	100	100	100	100	100
	3	0	0	0	0	0	0	0	0	0	0
PSC2	1	91	86	71	24	1	0	0	0	0	0
	2	9	14	29	76	99	100	100	100	100	100
	3	0	0	0	0	0	0	0	0	0	0

Table 3.1 Detection capability of the MDL and PSC algorithms for a configuration with two uncorrelated sources for 100 independent trials.

detection method	k	SNR (dB)				
		−20	−10	0	10	20
MDL2	1	100	91	0	0	0
	2	0	9	100	100	100
	3	0	0	0	0	0
PSC2	1	96	70	0	0	0
	2	4	29	93	100	100
	3	0	1	7	0	0

Table 3.2 Detection capability of the MDL and PSC algorithms for a configuration with two coherent sources for 100 independent trials.

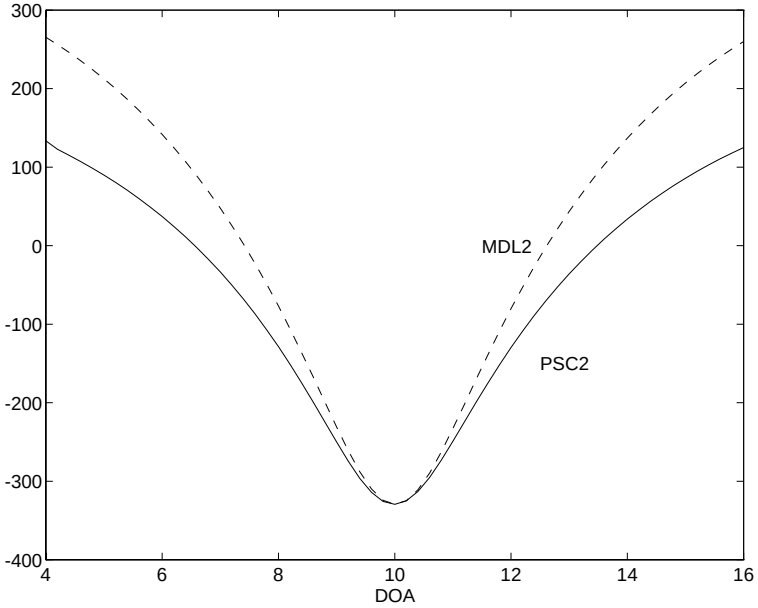

Fig. 3.3 The PSC and the MDL criteria for a single source arriving from 10 degrees at a uniform linear array of eight sensors.

consider a configuration with three sources at 10, 15, and 30 degrees arriving at a uniform linear array of 12 sensors with half wavelength spacing. The SNR is -5 dB. It is assumed that after 50 sample times the source at 30 degrees is turned off. The PSC and MDL methods were used on 2000 data samples to determine the number of signals. The MDL is computed for all window lengths smaller than 2000. The results are reported in Fig. 3.4. The PSC algorithm shows a change in the number of signals after about 700 samples. For window lengths smaller than 1600, the MDL will detect two sources. The time delay for detecting a change in the number of signals for the PSC algorithm is smaller than that for the MDL.

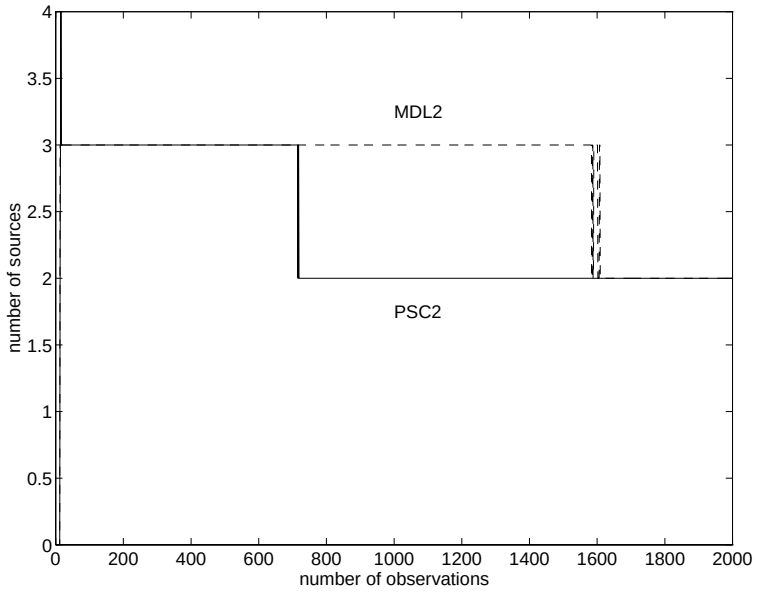

Fig. 3.4 Detected number of signals using the PSC and the MDL algorithms versus the number of samples. Three sources are at 10, 15, and 30 degrees and arrive at a uniform linear array of 12 sources. The source at 30 degrees is turned off after 50 samples.

Chapter 4

Optimum Focusing

Array processing techniques can be used to locate wideband signals. A wideband signal has a relatively large bandwidth with a value comparable to the center frequency. The frequency-wavenumber spectrum of a wideband point signal takes values on a line parallel to the frequency axis (Signal B in Fig. 1.1). Several methods for the processing of wideband signals using an array of sensors have been proposed in the literature. The first step in some of these techniques is to sample the signal in frequency domain [52] [7] [22]. This sampling can be performed by using a discrete Fourier transformation of the time samples or by using filter banks. The samples of the spectrum can be uniformly or nonuniformly distributed in the frequency domain.

Many array processing techniques use the concept of the signal subspace which is the span of the location vectors of the array. Since each location vector is a function of the observation frequency, the signal subspace depends on the frequency of the observation. For wideband signals the signal subspaces at different frequencies are different and do not overlap. This fact prevents the observation vectors at the frequency bins from being directly added to each other. Wang and Kaveh [52] propose *focusing* of the observation vectors. Focusing involves transforming the signal subspaces at different frequency bins into the *focusing subspace*. They choose an arbitrary frequency, say f_0 the center frequency of the spectrum of the signals, and transform all the subspaces at the frequency bins into the subspace created by the span of the location vectors at the frequency f_0. Then, they

use a high resolution algorithm such as MUSIC to estimate the directions-of-arrival of the sources. This method is called the coherent signal subspace method (CSM). Note that the correlation matrix, which is generated by focusing, has the characteristics of a narrowbnad signal and any narrowbnad localization method can be used to estimate the directions-of-arrival (DOAs). Focusing in the CSM reduces the resolution threshold signal-to-noise ratio. If the integral of the signal covariance matrix taken over the frequency spectrum is full rank, the method can also be applied to coherent signal localization. Hung and Kaveh [17] prove that the best performance is obtained if the mapping of the subspaces is done through a unitary transformation. They do not discuss how to choose the best focusing subspace.

In this chapter, we propose a method for focusing subspace selection. The method is based on minimizing the subspace fitting error. The subspace fitting error for each frequency bin is defined as the distance between the focusing matrix and the transformed location matrix. Later, we minimize a tight bound for the error. The focusing frequency selection is performed in two steps. First, the singular values of the location matrix are found. Then, a one variable nonlinear minimization problem is solved to obtain the focusing subspace.

Swingler and Krolik [36] prove that for a single-source scenario it is possible to have an unbiased estimate of the DOA if the centroid of the source spectrum is selected as the focusing frequency. Later we will show that for multiple-source cases the CSM algorithm cannot provide unbiased estimates of the DOAs. However, with the method that we propose here the bias of the DOAs estimate is minimized. The simulation results show that with this method of selecting the focusing subspace, the resolution threshold SNR is also reduced.

4.1. The coherent signal subspace method

Consider an array of p sensors exposed to q far-field wideband sources. To satisfy the constraints for unique solution, it is assumed that the number of sources is smaller than the number of sensors, $q < p$. The signals of the sources can be partially or fully correlated.

The output of the sensors is shown by p-vector $\mathbf{x}(t)$ with the i-th component

$$x_i(t) = \sum_{l=1}^{q} s_l(t - \tau_i(\theta_l)) + n_i(t), \qquad 1 \le i \le p \tag{4.1}$$

where s_l is the l-th source signal, θ_l is the DOA of the l-th source, and $\tau_i(\theta_l)$ is the propagation delay for the l-th source at the sensor i with respect to the reference point of the array. Using the first sensor as a reference point, for a linear array with uniform spacing, $\tau_i(\theta_l) = (i-1)\frac{d}{c}\sin\theta_l$, where d is the spacing between two consecutive sensors, and c is the propagation velocity. It is also assumed that the observation is corrupted by an additive noise which is represented in the model by $n_i(t)$.

In the frequency domain, after arrangement in vector form, the sensors output is represented by

$$\mathbf{x}(\omega) = \mathbf{A}(\omega, \boldsymbol{\theta})\mathbf{s}(\omega) + \mathbf{n}(\omega) \tag{4.2}$$

where $\mathbf{x}(\omega), \mathbf{s}(\omega)$ and $\mathbf{n}(\omega)$ are the Fourier transforms of the observation, the signal and the noise vectors, respectively. The $p \times q$ matrix $\mathbf{A}(\omega, \boldsymbol{\theta}) = [\mathbf{a}(\omega, \theta_1) \ldots \mathbf{a}(\omega, \theta_q)]$ is the location matrix of the array of sensors and is assumed to be full rank. In other words, the steering vectors $\mathbf{a}(\omega, \theta_i),\ i = 1, \ldots, q$, are independent for every ω.

The signal samples are generated independently by a circular complex Gaussian distribution with an unknown covariance matrix $\mathbf{S}(\omega)$. The noise samples are an i.i.d. sequence of complex Gaussian random vectors with unknown covariance matrix $\sigma^2\mathbf{I}$ and are independent of the signal samples. It is assumed that the noise is spatially white. This assumption can be relaxed if the correlation matrix of the noise is known but for a scale factor. In that case, a prewhitening step is required to create uncorrelated intersensor noise [4]. From (4.2) and using the assumptions on the signal and noise samples, the covariance matrix of the observation vector at frequency ω is given by

$$\mathbf{R}(\omega) = \mathbf{A}(\omega, \boldsymbol{\theta})\mathbf{S}(\omega)\mathbf{A}^H(\omega, \boldsymbol{\theta}) + \sigma^2\mathbf{I} \tag{4.3}$$

where the superscript H represents the Hermitian transpose.

In practice, the correlation matrix is unknown and must be estimated. A sufficiently long duration of sensor output is observed and sampled in time. The sampled data are divided into N snapshots, each containing J samples. In each snapshot, an FFT algorithm is applied to transform the data onto the frequency domain. If the correlation time of the signal is sufficiently smaller than the length of a snapshot, the Fourier transformed data at different frequency bins are approximately uncorrelated and their correlation matrix satisfies (4.3). To obtain the sample correlation matrices, an averaging in time is required. Let us represent the frequency samples by $\mathbf{x}_j^{(i)},\ j = 1, \ldots, J,\ i = 1, \ldots, N$. Then the samples correlation matrix at the j-th frequency bin can be found from

$$\mathbf{R}_j = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_j^{(i)} \mathbf{x}_j^{(i)H} \tag{4.4}$$

Since the location matrix is a function of the frequency, the signal subspaces at different frequency bins are different. The CSM algorithm transforms these subspaces and overlaps them in a predefined subspace, the so-called focusing subspace. The focusing matrices, $\mathbf{T}_j$, are the solutions of the equation

$$\mathbf{T}_j \mathbf{A}_j = \mathbf{A}_0, \qquad j = 1, \ldots, J \tag{4.5}$$

where $\mathbf{A}_j$ is the location matrix at the j-th frequency bin and $\mathbf{A}_0$ is the *focusing location matrix*. The matrices $\mathbf{A}_0$ and $\mathbf{A}_j$ are functions of the DOAs, $\boldsymbol{\theta}$. An ordinary beamforming pre-process gives an estimate of the angles-of-arrival which can be used in (4.5). Using the focusing matrices $\mathbf{T}_j$, the observation vectors at different frequency bins are transformed into the focusing subspace. In particular, new observation vectors are formed by

$$\mathbf{y}_j = \mathbf{T}_j \mathbf{x}_j, \qquad j = 1, \ldots, J. \tag{4.6}$$

These transformed observation vectors are used to construct the sample correlation matrices

$$\mathbf{R}_j^{(y)} = \frac{1}{N} \sum_{l=1}^{N} \mathbf{y}_j^{(l)} \mathbf{y}_j^{(l)H} \tag{4.7}$$

where the transformed data vector for frequency bin j and for batch l is represented by $\mathbf{y}_j^{(l)}$. An average of these aligned correlation matrices over the frequency bins gives a universal focused sample correlation matrix that can be used for detection and estimation. If this matrix is represented by $\mathbf{R}$, we will have

$$\mathbf{R} = \frac{1}{J}\sum_{j=1}^{J}\mathbf{R}_j^{(y)} \tag{4.8}$$

$$= \mathbf{A}_0\mathbf{R}_s\mathbf{A}_0^H + \sigma^2\mathbf{R}_n \tag{4.9}$$

where

$$\mathbf{R}_s = \frac{1}{J}\sum_{j=1}^{J}\mathbf{S}_j \tag{4.10}$$

$$\mathbf{R}_n = \frac{1}{J}\sum_{j=1}^{J}\mathbf{T}_j\mathbf{T}_j^H \tag{4.11}$$

and $\mathbf{S}_j = \mathbf{S}(\omega_j)$ is the source correlation matrix at the j-th frequency bin.

This transformation improves the efficiency of the estimation by condensing the energy of sub-bands in the focusing signal subspace. Yet, it creates a problem. It is seen that the focusing removes the whiteness of the noise. This in turn changes the SNR at the output of the processor. The *focusing loss* is defined as the ratio of the array SNR after and before focusing. Using this quantity, Hung and Kaveh [17] show that the focusing is lossless if $\mathbf{T}_j$'s are unitary transformations. Specifically, they propose using the transformation matrices which are obtained by the constrained minimization problem

$$\min_{\mathbf{T}_j}\|\mathbf{A}_0 - \mathbf{T}_j\mathbf{A}_j\| \tag{4.12}$$
$$\text{s.t.}\quad \mathbf{T}_j^H\mathbf{T}_j = \mathbf{I}$$

for $j = 1, \ldots, J$. They use the Frobenius matrix norm which is defined by

$$\|\mathbf{B}\| = \Big[\sum_{i,j} b_{ij}^2\Big]^{\frac{1}{2}} = \Big[\text{tr}(\mathbf{B}^H\mathbf{B})\Big]^{\frac{1}{2}} \tag{4.13}$$

where tr(.) stands for the trace of matrix. In factor analysis, this is known as the problem of finding a *procrustean transformation* of $\mathbf{A}_j$. The solution to (4.12) is given by [15] [17]

$$\mathbf{T}_j = \mathbf{V}_j \mathbf{W}_j^H \tag{4.14}$$

where $\mathbf{V}_j$ and $\mathbf{W}_j$ are the left and right singular vectors of $\mathbf{A}_0 \mathbf{A}_j^H$.

4.2. Focusing frequency selection

We define a criterion based on the error involved in the transformation of the signal subspaces. The minimization of this criterion gives the focusing frequency. Specifically, we seek an f_0 which is the solution to the following minimization,

$$\begin{aligned} \min_{f_0} \ \min_{\mathbf{T}_j} & \sum_{j=1}^{J} w_j \|\mathbf{A}_0 - \mathbf{T}_j \mathbf{A}_j\|^2 \\ \text{s.t.} \ & \mathbf{T}_j^H \mathbf{T}_j = \mathbf{I} \\ & \mathbf{A}_0 \in \boldsymbol{\mathcal{A}}(\boldsymbol{\theta}) \end{aligned} \tag{4.15}$$

where $\boldsymbol{\mathcal{A}}(\boldsymbol{\theta})$ is the set of all location matrices for given DOA $\boldsymbol{\theta}$, and w_j is a weighting factor proportional to the SNR at the j-th frequency bin with $\sum_{j=1}^{J} w_j = 1$.

Using (4.14) the subspace fitting error is given by

$$\begin{aligned} \sum_{j=1}^{J} w_j \|\mathbf{A}_0 - \mathbf{T}_j \mathbf{A}_j\|^2 &= \sum_{j=1}^{J} w_j \Big[\|\mathbf{A}_0\|^2 + \|\mathbf{A}_j\|^2 - 2\Re\Big(\mathrm{tr}(\mathbf{A}_0 \mathbf{A}_j^H \mathbf{T}_j^H)\Big) \Big] \\ &= 2Jpq - 2 \sum_{j=1}^{J} \sum_{i=1}^{q} w_j \sigma_i(\mathbf{A}_0 \mathbf{A}_j^H) \end{aligned} \tag{4.16}$$

where the $\sigma_i(\mathbf{B})$, $i = 1, \ldots, q$, are the singular values of the matrix $\mathbf{B}$ arranged in nonincreasing order, $\Re(.)$ represents the real part of a complex number. Here, we have used the equality

$$\|\mathbf{A}\|^2 = \sum_{i=1}^{q} \|\mathbf{a}_i\|^2 = pq \tag{4.17}$$

which holds for an arbitrary array manifold.

From (4.16) it is seen that the minimization problem (4.15) is identical to

$$\begin{aligned} &\max_{f_0} \sum_{j=1}^{J} \sum_{i=1}^{q} w_j \sigma_i(\mathbf{A}_0 \mathbf{A}_j^H). \\ &\text{s.t.} \quad \mathbf{A}_0 \in \boldsymbol{\mathcal{A}}(\boldsymbol{\theta}) \end{aligned} \tag{4.18}$$

Direct maximization of (4.18) is very complicated and the computational complexity increases with the number of frequency samples. In the sequel, we present a suboptimal method which is based on maximizing an upper bound to (4.18). We show that in the vicinity of the maximum point the bound is tight. The tightness of the bound at the maximum point indicates that the method performs very close to the optimal case. The following lemma establishes a lower bound on the norm of the difference of two matrices.

Lemma 4.1. *If* $\mathbf{A}, \mathbf{B} \in \mathbf{M}_{m,n}$ *(an* $m \times n$ *matrix) are given matrices with ordered singular values* $\sigma_1(\mathbf{A}) \geq \ldots \geq \sigma_q(\mathbf{A}) \geq 0$ *and* $\sigma_1(\mathbf{B}) \geq \ldots \geq \sigma_q(\mathbf{B}) \geq 0$*, where* $q = \min\{m, n\}$*, then*

$$\|\mathbf{A} - \mathbf{B}\|^2 \geq \sum_{i=1}^{q} [\sigma_i(\mathbf{A}) - \sigma_i(\mathbf{B})]^2. \tag{4.19}$$

Proof: See [15].

Application of Lemma 4.1 to (4.16) gives

$$\sum_{j=1}^{J} \sum_{i=1}^{q} w_j \sigma_i(\mathbf{A}_0 \mathbf{A}_j^H) \leq \sum_{j=1}^{J} \sum_{i=1}^{q} w_j \sigma_i(\mathbf{A}_0) \sigma_i(\mathbf{A}_j^H). \tag{4.20}$$

We propose maximizing the right hand side of (4.20),

$$\begin{aligned} &\max_{f_0} \sum_{j=1}^{J} \sum_{i=1}^{q} w_j \sigma_i(\mathbf{A}_0) \sigma_i(\mathbf{A}_j^H). \\ &\text{s.t.} \quad \mathbf{A}_0 \in \boldsymbol{\mathcal{A}}(\boldsymbol{\theta}) \end{aligned} \tag{4.21}$$

This maximization is performed in two steps. First, the singular values of the focusing location matrix $\mathbf{A}_0$ are determined. Then, using the known structure of the location

matrix, the focusing frequency f_0 is found. Let us define

$$\mu_i \stackrel{\triangle}{=} \sum_{j=1}^{J} w_j \sigma_i(\mathbf{A}_j). \tag{4.22}$$

Using (4.22), the maximization problem (4.21) is written as

$$\begin{aligned} \max_{f_0} \ & \sum_{i=1}^{q} \mu_i \sigma_i(\mathbf{A}_0) \\ \text{s.t.} \ & \mathbf{A}_0 \in \boldsymbol{\mathcal{A}}(\boldsymbol{\theta}). \end{aligned} \tag{4.23}$$

This is a one-variable maximization problem that can be solved by searching for the best f_0 in the interval of interest.

In the following, we suggest a method that can reduce the computational complexity of the search (4.23). The method is based on performing the search in a smaller interval. The search interval can be made smaller if we can store the singular values of $\mathbf{A}_0$ for some f_0. To see that, consider the maximization problem

$$\begin{aligned} \max_{\sigma_i} \ & \sum_{i=1}^{q} \mu_i \sigma_i \\ \text{s.t.} \ & \sum_{i=1}^{q} \sigma_i^2 = pq \\ & \sigma_i \geq 0 \end{aligned} \tag{4.24}$$

where we have used (4.17). Using Lagrange multipliers, the solutions of (4.24) can be determined to be

$$\tilde{\sigma}_i = \frac{\mu_i \sqrt{pq}}{\sqrt{\sum_{l=1}^{q} \mu_l^2}}, \qquad i = 1, \ldots, q \tag{4.25}$$

with the maximum value $\sqrt{pq \sum_{l=1}^{q} \mu_l^2}$. Because of the constraints on the structure of the location matrix, usually the $\tilde{\sigma}_i$'s cannot be the singular values of $\mathbf{A}_0$. The singular values of $\mathbf{A}_0$ are continuous functions of the frequency f_0. The continuum of the singular values is a curve located on the surface of a sphere which is separated by the positive quadrants. Fig. 4.1 depicts the relationship between f_0 and the singular values. The optimum singular

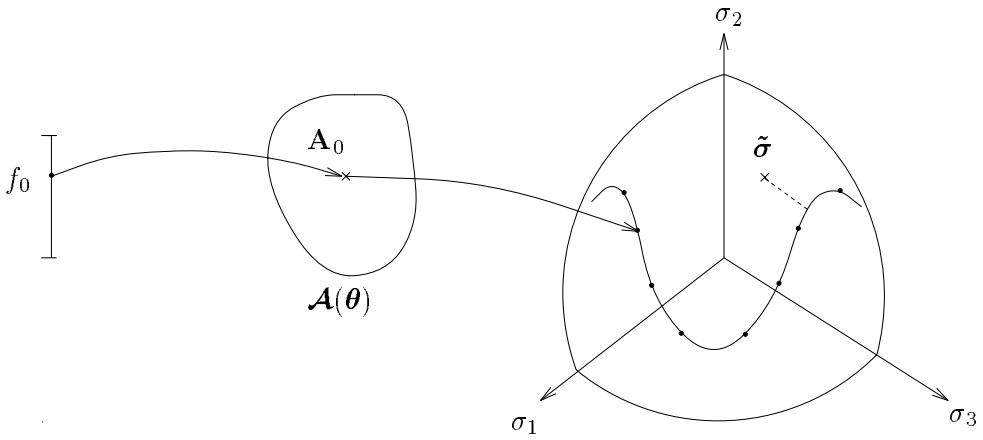

Fig. 4.1 Relationship between frequency and the singular values of the location matrix

value vector of $\mathbf{A}_0$ is the one who is the closest point on the singular value continuum to $\tilde{\boldsymbol{\sigma}} = (\tilde{\sigma}_1, \ldots, \tilde{\sigma}_q)$. This can be found by minimizing the following one-variable nonlinear equation,

$$\begin{aligned} &\min_{f_0} \sum_{i=1}^{q} [\sigma_i(\mathbf{A}_0) - \tilde{\sigma}_i]^2 \\ &\text{s.t.} \quad \mathbf{A}_0 \in \boldsymbol{\mathcal{A}}(\boldsymbol{\theta}). \end{aligned} \tag{4.26}$$

In practice the singular values of the location matrix $\mathbf{A}_0$ can be computed and stored for some coarse digitized values of f_0 and DOA's. Using the pre-estimates of DOA, the $\tilde{\sigma}_i$'s can be determined. The maximization (4.26) is then performed in the vicinity of the stored value of f_0 which has the closest singular values to $\tilde{\sigma}_i$'s. The complexity of this minimization is independent of the number of frequency samples.

4.3. The bias of estimation

Despite the fact that the CSM algorithm is very effective in wideband array processing, it suffers from asymptotic bias of the DOA estimates. The bias increases with the bandwidth of the sources and the deviation of the focusing points from the true DOA. In Chapter 5,

we introduce a new method for unbiased wideband array processing. There, we show that using the CSM algorithm, unbiased estimation is not possible. However, our results here show that there exists a focusing frequency that minimizes the estimation bias in the CSM algorithm.

Note that the MUSIC estimator is asymptotically unbiased. The bias in the peak locations which is created in the CSM algorithm is the consequence of focusing. A proper selection of the focusing frequency minimizes the bias of estimation. We show that the proposed method for focusing frequency selection also minimizes the peak bias. We start by discussing the mechanism that generates the bias.

For a given ω in the source bandwidth, the array manifold is a curve in the p-dimensional complex space that is created from the location vectors $\mathbf{a}(\omega, \theta)$ for all θ. For every pure delay environment, the norm of the location vector is equal to the square root of the number of the sensors. Thus the array manifold lies on the surface of a sphere with the radius $\sqrt{p}$. We represent this sphere by $\mathcal{S}$. It is also important to note that the array manifold is continuous on ω and θ. The MUSIC algorithm finds the intersection of the subspace spanned by the eigenvectors of the correlation matrix corresponding to the q largest eigenvalues with the array manifold. If the true correlation matrix is applied to the MUSIC, the DOAs are estimated without bias. However, deviation from the true signal subspace will cause bias in the estimation.

Now consider the case in which the location matrices $\mathbf{A}_j$'s are transformed by the unitary matrices $\mathbf{T}_j$'s to the vicinity of the focusing location matrix $\mathbf{A}_0$. The transformed location matrices form a cluster around $\mathbf{A}_0$. It is quite clear that the closer the transformed matrices are, the better the performance is. In an ideal case all the transformed location matrices superimpose on $\mathbf{A}_0$. We call this case *perfect focusing.* In perfect focusing, the column vectors of the location matrix $\mathbf{A}_j$ are transformed to the corresponding column vectors of $\mathbf{A}_0$. This is also seen from the characteristics of the Frobenius norm; the square of the Frobenius norm of a matrix is equal to addition of the square of the Euclidean norm of the column vectors. Hung and Kaveh [17] show that for successful application of the unitary transformation method, it is necessary to add two extra focusing DOAs. They

suggest placing the extra DOAs at $\pm 0.25B_W$ (B_W is the beamwidth) from the estimated directions-of-arrival. For instance, if the i-th DOA is found at $\hat{\theta}_i$ by the pre-processing step, the focusing points for the i-th angle are chosen at $(\hat{\theta}_i - 0.25B_W\ ,\ \hat{\theta}_i\ ,\ \hat{\theta}_i + 0.25B_W)$. These angles determine an interval on the array manifold. If this interval is small compared to the curvature of the array manifold at $\hat{\theta}_i$, it is transformed to a corresponding interval at the array manifold with the processing frequency ω_0. This is attributed to the continuity of the array manifold and the unitary transformation. Thus, in perfect focusing the location vectors of each frequency bin that are located at the true DOA are transformed to the corresponding vectors at the focusing manifold; hence the estimation is unbiased.

In practice the conditions for perfect focusing are not satisfied. The transformed matrices are clustered around $\mathbf{A}_0$. The closest distance between these matrices is obtained if and only if $\mathbf{A}_0$ is on the average of the transformed matrices. To see this take any $p \times q$ matrix $\mathbf{C}$ that has column vectors on the sphere $\mathcal{S}$. To have a tight cluster, we should perform the following minimization,

$$\min_{\mathbf{C}} \min_{\mathbf{T}_j} \sum_{j=1}^{J} w_j \|\mathbf{C} - \mathbf{T}_j \mathbf{A}_j\|^2 \tag{4.27}$$

$$\text{s.t.} \quad \mathbf{T}_j^H \mathbf{T}_j = \mathbf{I} \tag{4.28}$$

$$\mathbf{C} \in \mathcal{S} \tag{4.29}$$

The minimum (4.27) is obtained for

$$\mathbf{C} = \left[\sum_{j=1}^{J} w_j \mathbf{T}_j \mathbf{A}_j \right] \mathbf{K} \tag{4.30}$$

where $\mathbf{K}$ is a diagonal normalization matrix. It is seen that $\mathbf{C}$ is formed by the average of the transformed location vectors for each source. The method that we proposed for focusing subspace selection gives the closest $\mathbf{A}_0$ to the matrix $\mathbf{C}$. This suggests that the bias is also near the minimum value for the selected focusing frequency.

4.4. Tightness of the upper bound

It was stated earlier that the upper bound of (4.18) is tight in the vicinity of the optimum point. In this section we discuss this issue. It has been already shown that

$$\sum_{j=1}^{J}\sum_{i=1}^{q} w_j\sigma_i(\mathbf{A}_0\mathbf{A}_j^H) \leq \sum_{j=1}^{J}\sum_{i=1}^{q} w_j\sigma_i(\mathbf{A}_0)\sigma_i(\mathbf{A}_j) \tag{4.31}$$

$$\leq \left(pq\sum_{i=1}^{q}\mu_i^2\right)^{\frac{1}{2}} = \sqrt{pq}\|\mu\| \tag{4.32}$$

where $\mu_i, i = 1,\ldots,q$, are given by (4.22), $\mu = (\mu_1,\ldots,\mu_J)$, and $\|.\|$ is the Euclidean norm of a vector. We make some observations on (4.32).

- The right-hand-side of (4.32) is independent of the focusing frequency.
- (4.24) is a convex maximization problem.
- If $\mathbf{A}_0 = \mathbf{T}_j\mathbf{A}_j,\ j = 1,\ldots,J$, then the bound is achieved.
  This case corresponds to perfect focusing. It is seen that such $\mathbf{A}_0$ is an optimum focusing matrix. In other words, in perfect focusing the left-hand-side of (4.32) attains its maximum value. However, in practice perfect focusing is not fulfilled. The criterion that we defined in (4.15) is the closest case to perfect focusing. Thus the optimum value that is obtained from (4.15) is very close to the bound.

Based on these observations, we see that the proposed method operates very close to the optimal case. This performance is, however, achieved with a considerable reduction in the computational complexity.

The closeness of the left-hand-side of (4.32) to the bound is a function of the number of frequency samples. Fig. 4.2 represents the left-hand-side of (4.32) normalized with respect to $\|\mu\|$, as a function of the number of frequency samples for a typical configuration of 2 sources arriving at 8 sensors. As it is seen, by increasing the number of frequency samples the value approaches the bound $\sqrt{pq} = 4$.

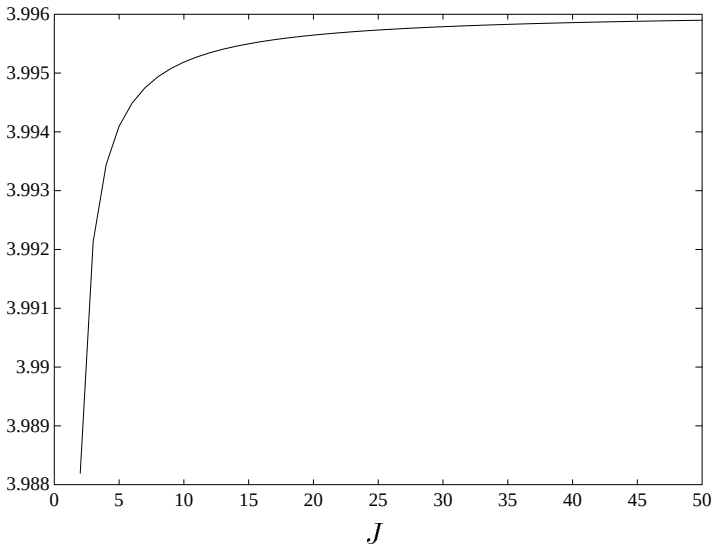

Fig. 4.2 The summation of the singular values of $\mathbf{A}_0\mathbf{A}_j^H$ normalized with respect to the norm of the vector $\mu = (\mu_1, \ldots, \mu_J)$ as a function of the number of frequency samples for a configuration with two wideband sources with a 40% relative bandwidth arriving from 10 and 14 degrees at a uniform linear array of 8 sensors.

4.5. Simulation Results

Assume that a uniform linear array of 8 sensors is exposed to 2 far-field wideband sources arriving from 10 and 14 degrees. The signals of sources are uncorrelated and have a 40 percent bandwidth relative to the center frequency. The spectrum of the signals is given by

$$S(f) = \begin{cases} 5f - 4 & 0.8 \leq f \leq 1.2 \\ 0 & \text{otherwise} \end{cases} \tag{4.33}$$

The spectrum of the signals is sampled using a 16 point FFT algorithm. The data at each frequency bin contain 100 snapshots. Using an ordinary beamformer a single source is detected at 13 degrees. We add two extra DOA's at 9 and 17 degrees as the focusing angles. The subspace fitting error is shown in Fig. 4.3. It is seen that the error is minimized at the frequency 1.1 which is 10 percent higher than the center frequency and 3 percent higher than the centroid frequency. For a fixed SNR at 20 dB we have found the bias of

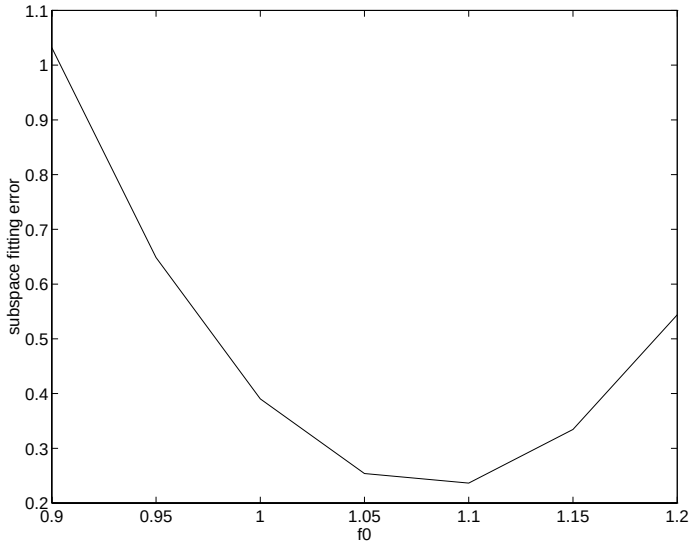

Fig. 4.3 The subspace fitting error as a function of the focusing frequency for two uncorrelated far-field wideband sources arriving at a linear array of 8 sensors with a 20 dB SNR.

the DOA estimates for different focusing frequencies. The results are depicted in Fig. 4.4. The bias of the DOA estimate for the centroid frequency is 0.1246 compared to 0.0487 for the optimum focusing frequency.

To compare the resolution capability of the CSM algorithm for different focusing frequencies, we performed 100 independent trials and counted the number of times that the method resolved the two DOAs by detecting two peaks in the spectrum of the MUSIC algorithm. The results are compared in Table 4.1. It is seen that the resolution SNR for the optimum focusing frequency is lower than that for the center frequency. It should be noted that there are two facts that decrease the resolution SNR. First, optimum focusing frequency has a smaller subspace fitting error and second it is higher than the center frequency. The separation between the location vectors for higher frequencies is larger which tends to give a better resolution performance. The resolution at 1.2 is better than the other frequencies. However, the results of Table 4.2 and Fig. 4.4 show that for SNRs above the resolution threshold the bias is minimized for the optimum focusing frequency. The mean square error of the CSM algorithm for the two focusing frequencies 1 and 1.1

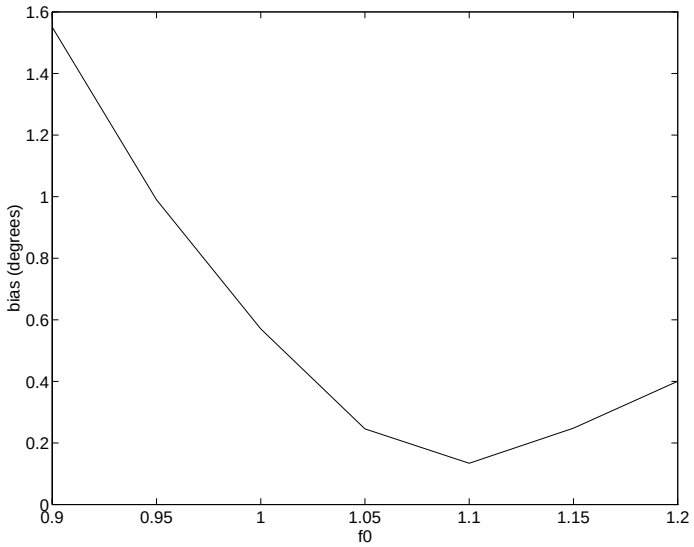

Fig. 4.4 The norm of the bias vector for the DOA estimation of two uncorrelated far-field wideband sources arriving at a linear array of 8 sensors with a 20 dB SNR.

is compared in Fig. 4.5.

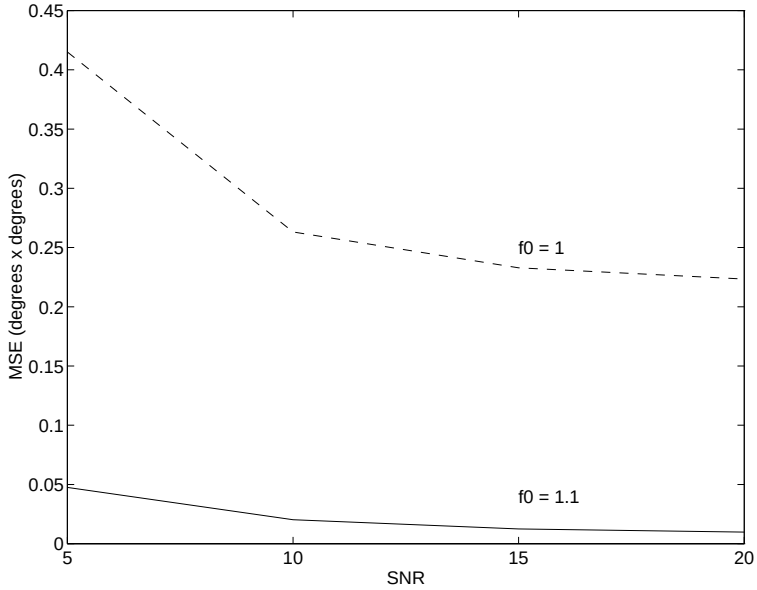

Fig. 4.5 The mean square error of the CSM algorithm for the two focusing frequencies 1 and 1.1 for two uncorrelated far-field wideband sources at 10 and 14 degrees arriving at a linear array of 8 sensors.

	SNR			
f_0	-5	0	5	10
0.90	0	6	36	64
0.95	0	22	86	100
1.00	2	41	99	100
1.05	1	71	100	100
1.10	6	89	100	100
1.15	9	99	100	100
1.20	11	100	100	100

Table 4.1 The resolution capability of the CSM algorithm for two closely spaced non-coherent wideband signals at 10 and 14 degrees arriving at a uniform linear array of 8 sensors, versus the focusing frequency. The resolution capability is measured in terms of the percentage of detection.

	SNR			
f_0	-5	0	5	10
0.90	–	1.22	1.48	1.55
0.95	–	1.13	1.10	0.98
1.00	1.57	0.88	0.68	0.57
1.05	0.59	0.76	0.36	0.24
1.10	1.08	0.54	0.22	0.13
1.15	0.58	0.42	0.25	0.24
1.20	0.79	0.37	0.39	0.40

Table 4.2 The norm of the bias vector for two closely spaced noncoherent wideband signals at 10 and 14 degrees arriving at a uniform linear array of 8 sensors, versus the focusing frequency.

Chapter 5

Correlation Transformation

In this chapter, we introduce a new technique for broadband array processing. Our method is similar to CSM in the sense that transformation of the signal subspaces is done through focusing matrices. A high resolution spectral estimation algorithm, such as MUSIC, is then applied to determine the DOA. In the new method, we apply a two-sided unitary transformation on the correlation matrix. In [17] it has been shown that unitary transformations have good performance in terms of *focusing loss* and *relative information index*. The motivation for using the correlation matrices instead of the location vectors is based on the fact that most of the high resolution spectral estimation algorithms use an eigenstructure decomposition of the correlation matrix.

The new method is termed two-sided correlation transformation (TCT) wideband array processing. The main objective of introducing a new technique for coherent wideband array processing is to eliminate or reduce the bias of DOA estimation in the CSM method. We compare the eigenvalues of the noise-free focused correlation matrix of the CSM and TCT algorithm. It is shown that in the CSM algorithm, the energy of signal is extended into the noise subspace. This signal extension acts as a colored noise with an unknown correlation matrix. Using these results, we show that the CSM cannot asymptotically generate unbiased estimates of the DOA. To the contrary, in the TCT algorithm the energy is confined to the signal subspace. Thus, an unbiased estimation of the DOAs can be achieved by using the TCT method. Furthermore, we show that the subspace

fitting error in the TCT algorithm is smaller than that for the CSM method. A smaller subspace fitting error results in a lower resolution threshold. It is also shown that the TCT algorithm has a smaller generalized variance than CSM. This results in an estimation with smaller variance.

5.1. Matrix approximation

The problem of approximating a given matrix by a matrix in a specified class arises in applications such as multivariate analysis, factor analysis, estimation of residuals in linear models, and the theory of generalized inverse of a matrix. In each case, a minimization problem is solved to obtain the closest distance between the two matrices. The distance between the matrices is measured with respect to an appropriate norm. One class of norms is known as the *unitarily invariant norms.* In this class, the norm of a matrix is invariant under unitary transformation. The spectral norm and the Frobenius norm are unitarily invariant. In this chapter, we use the Frobenius norm.

The error of the one-sided transformation used in the CSM algorithm from the previous chapter is given by

$$\begin{aligned}\mathcal{E} = \sum_{j=1}^{J} \|\mathbf{A}_0 - \mathbf{T}_j\mathbf{A}_j\|^2 &= \sum_{j=1}^{J}\Big[\|\mathbf{A}_0\|^2 + \|\mathbf{A}_j\|^2 - 2\Re\ \mathrm{tr}(\mathbf{A}_0\mathbf{A}_j^H\mathbf{T}_j^H)\Big] \\ &= 2Jpq - 2\sum_{j=1}^{J}\sum_{i=1}^{q}\sigma_i(\mathbf{A}_0\mathbf{A}_j^H). \end{aligned} \tag{5.1}$$

The following lemma produces a lower bound on the error in (5.1).

Lemma 5.1. *Let* $\mathbf{A}, \mathbf{B} \in \mathbf{M}_{m\times n}$ *(an* $m \times n$ *matrix) and* $q = \min\{m, n\}$. *Denote by* $\sigma_i(\mathbf{A}), \sigma_i(\mathbf{B})$ *and* $\sigma_i(\mathbf{A}\mathbf{B}^H)$, $i = 1, \ldots, q$, *the nonzero singular values of the corresponding matrices arranged in nonincreasing order. Then*

$$\sum_{i=1}^{q}\sigma_i(\mathbf{A}\mathbf{B}^H) \le \sum_{i=1}^{q}\sigma_i(\mathbf{A})\sigma_i(\mathbf{B}). \tag{5.2}$$

Proof: See Appendix C.

Lemma 5.1 is presented as a theorem in [16]. We have provided a self-contained proof using a different approach in Appendix C.

From (5.1) and (5.2) it is seen that the error of transformation is lower bounded as

$$2Jpq - 2\sum_{j=1}^{J}\sum_{i=1}^{q}\sigma_i(\mathbf{A}_0)\sigma_i(\mathbf{A}_j) \leq \mathcal{E} \tag{5.3}$$

where the error is defined in (5.1). This lower bound cannot be reached in general using the one-sided transformation of the CSM method.

Now consider the alternative of a two-sided transformation. For such a transformation, we will be able to achieve the lower bound to the error. In a two-sided transformation the objective is to find the unitary matrices $\mathbf{U}$ and $\mathbf{V}$ such that the following criterion is minimized,

$$\begin{aligned} &\min_{\mathbf{U},\mathbf{V}} \|\mathbf{A} - \mathbf{U}\mathbf{B}\mathbf{V}^H\|^2 \\ &\text{s.t.} \quad \mathbf{U}^H\mathbf{U} = \mathbf{I}, \\ &\qquad \mathbf{V}^H\mathbf{V} = \mathbf{I}. \end{aligned} \tag{5.4}$$

Theorem 5.1. *The solutions of* (5.4) *are given by* $\mathbf{U} = \mathbf{E}\mathbf{X}^H$ *and* $\mathbf{V} = \mathbf{F}\mathbf{Y}^H$ *where* $\mathbf{A} = \mathbf{E}\Sigma\mathbf{F}^H$ *and* $\mathbf{B} = \mathbf{X}\Lambda\mathbf{Y}^H$ *are the singular value decompositions of* $\mathbf{A}$ *and* $\mathbf{B}$*, respectively, and the error of transformation is given by*

$$\mathcal{E} = \|\mathbf{A}\|^2 + \|\mathbf{B}\|^2 - 2\sum_{i=1}^{q}\sigma_i(\mathbf{A})\sigma_i(\mathbf{B}). \tag{5.5}$$

Proof: See Appendix D

Corollary 5.1. *If* $\mathbf{A}$ *and* $\mathbf{B}$ *are square Hermitian matrices, the transformation matrices* $\mathbf{U}$ *and* $\mathbf{V}$ *will be identical and equal to* $\mathbf{E}\mathbf{X}^H$ *where* $\mathbf{E}$ *and* $\mathbf{X}$ *contain eigenvectors of* $\mathbf{A}$ *and* $\mathbf{B}$*, respectively.*

From Theorem 5.1 and Lemma 5.1, it is seen that the error of transformation is minimized for the two-sided unitary transformation. Since in practice the location matrix is measured in combination with the source signal vector, the two-sided transformation of

the location matrices is not practical. However, it is well known that for q noncoherent sources, the space spanned by the location matrix is the same as the span of the eigenvectors of the correlation matrix which correspond to q largest eigenvalues (signal subspace). Our method is based on two-sided transformation of a basis for the signal subspace. This will be discussed in the following section.

5.2. Two-sided Correlation Transformation method

In this section, a new method for wideband array processing is introduced. The method is based on transformation of the signal-subspaces into the focusing subspace. Like the CSM algorithm, the signal subspaces are transformed by using focusing matrices. The focusing matrix at each frequency bin is unitary and minimizes the distance between the focusing subspace and the transformed signal-subspace. In the new method, the transformation of the subspaces is performed through two-sided transformation applied on the correlation matrix. The motivation for using the correlation matrix, instead of the location matrix, is twofold. First, a two-sided transformation can be applied which results in a lower error. Second, many of the high resolution methods for DOA estimation are based on the eigenstructure decomposition of the correlation matrix. Thus the closer the transformed correlation matrices are, the better the results will be.

5.2.1. The TCT criterion

The TCT method is based on transformation of the matrices

$$\mathbf{P}_j = \mathbf{A}_j \mathbf{S}_j \mathbf{A}_j^H, \qquad j = 0, 1, \ldots, J \tag{5.6}$$

where $\mathbf{P}_j$ is the correlation matrix of the sensor output at the j-th frequency bin in a noise-free environment. The matrix $\mathbf{P}_0$ is the focusing noise-free correlation matrix. The TCT focusing matrices are found by minimizing

$$\min_{\mathbf{U}_j} \|\mathbf{P}_0 - \mathbf{U}_j \mathbf{P}_j \mathbf{U}_j^H\| \tag{5.7}$$

$$\text{s.t.} \quad \mathbf{U}_j^H \mathbf{U}_j = \mathbf{I}$$

for $j = 1, \ldots, J$. From Corollary 5.1, the solution of (5.7) is obtained as

$$\mathbf{U}_j = \mathbf{X}_0 \mathbf{X}_j^H \tag{5.8}$$

where $\mathbf{X}_0$ and $\mathbf{X}_j$ are the eigenvector matrices of $\mathbf{P}_0$ and $\mathbf{P}_j$, respectively. The matrix $\mathbf{U}_j$ can be used to transfer the observation vector $\mathbf{x}_j$ into $\mathbf{y}_j$ through

$$\mathbf{y}_j = \mathbf{U}_j \mathbf{x}_j. \tag{5.9}$$

The observation vectors $\mathbf{y}_j$, $j = 1, \ldots, J$ are in the focusing subspace. The correlation matrices of $\mathbf{y}_j$'s can be averaged to find the universal focused sample correlation matrix.

In computing $\mathbf{U}_j$ the matrices $\mathbf{A}_j$ and $\mathbf{S}_j$ are assumed to be known. In practice a pre-processing step is required to estimate these matrices. A low resolution beamformer is applied to estimate the number and the directions of arrival of the sources. Closely separated and correlated sources may not be resolved at this stage. Like [17], we add two extra directions of arrival at $\pm 0.25 B_W$ (BeamWidth) of the estimated DOA. For instance if the i-th DOA is found at $\hat{\theta}_i$ by the pre-processing, the focusing points are chosen at $(\hat{\theta}_i - 0.25B_W, \hat{\theta}_i, \hat{\theta}_i + 0.25B_W)$. In TCT, the number of focusing angles should be larger than the true number of sources. Using the results of this pre-processing step an estimate of the location matrix, $\mathbf{A}_j$, is obtained. Then the eigenvalues of the sample correlation matrices, $\hat{\mathbf{R}}_j$, $j = 1, \ldots, J$, are computed and sorted in decreasing order. The noise power at the j-th frequency bin is estimated by

$$\hat{\sigma}_j^2 = \frac{1}{p-q} \sum_{i=q+1}^{p} \lambda_i(\hat{\mathbf{R}}_j) \tag{5.10}$$

where $\lambda_i(\mathbf{B})$ is the i-th eigenvalue of $\mathbf{B}$. The source correlation matrix is then found from

$$\hat{\mathbf{S}}_j = (\mathbf{A}_j^H \mathbf{A}_j)^{-1} \mathbf{A}_j^H [\hat{\mathbf{R}}_j - \hat{\sigma}_j^2 \mathbf{I}] \mathbf{A}_j (\mathbf{A}_j^H \mathbf{A}_j)^{-1}. \tag{5.11}$$

In general, the estimated source correlation matrix (5.11) may have negative eigenvalues. However, our simulation studies have shown that as far as the estimation of the DOAs is concerned the TCT algorithm still can be used. As an alternative to (5.10) and to guarantee the nonnegativeness of the estimated source correlation matrix, the noise power can be estimated from

$$\hat{\sigma}_j^2 = \lambda_p(\hat{\mathbf{R}}_j) \tag{5.12}$$

where $\lambda_p(\hat{\mathbf{R}}_j)$ is the smallest eigenvalue of $\hat{\mathbf{R}}_j$.

We will see later that the matrix $\mathbf{S}_j$ is used to determine $\mathbf{P}_0$, the focusing noise-free correlation matrix. In practice, $\mathbf{P}_j$ is directly computed from

$$\mathbf{P}_j = (\hat{\mathbf{R}}_j - \hat{\sigma}_j^2\mathbf{I}). \tag{5.13}$$

If a good estimate of the noise power is used in (5.13), $\mathbf{P}_j$ can be interpreted as the correlation matrix of the cleaned data. By the cleaned data we mean the output of a pre-processing step that removes or decreases the effect of noise. The computational complexity of (5.13) is relatively low since the Lanczos algorithm [12] can be applied to obtain a few of smallest eigenvalues of $\hat{\mathbf{R}}_j$.

5.2.2. Selecting the focusing subspace

The noise-free focusing correlation matrix $\mathbf{P}_0$ is a function of the DOAs, the frequency of focusing f_0, and the focusing source correlation matrix $\mathbf{S}_0$. The initial DOAs are found using ordinary beamforming. We still have to choose f_0 and $\mathbf{S}_0$. The subspace fitting error is defined as

$$\mathcal{E} = \sum_{j=1}^{J} \|\mathbf{P}_0 - \mathbf{U}_j\mathbf{P}_j\mathbf{U}_j^H\|^2. \tag{5.14}$$

Using this error we can select f_0 and $\mathbf{S}_0$ in two steps. In the first step, we choose the focusing source correlation matrix $\mathbf{S}_0$ as follows.

In an ideal case all the transformed location matrices superimpose on $\mathbf{A}_0$. This case is called perfect focusing. In perfect focusing the column vectors of the location matrix $\mathbf{A}_j$

are transformed to the corresponding columns of $\mathbf{A}_0$, i.e.

$$\mathbf{A}_0 = \mathbf{U}_j \mathbf{A}_j, \qquad j = 1, \ldots, J. \tag{5.15}$$

In such a case the subspace fitting error can be shown to be

$$\mathcal{E} = \sum_{j=1}^{J} \| \mathbf{A}_0 (\mathbf{S}_0 - \mathbf{S}_j) \mathbf{A}_0^H \|^2. \tag{5.16}$$

Minimization of (5.16) for $\mathbf{S}_0$ gives

$$\mathbf{S}_0 = \frac{1}{J} \sum_{j=1}^{J} \mathbf{S}_j. \tag{5.17}$$

The estimate of the $\mathbf{S}_j$, given by (5.11), can be used in (5.17).

In practice, perfect focusing cannot be achieved if the transformation is constrained to be unitary. However, the transformed location matrices are close to the focusing location matrix. Thus, the same focusing source correlation matrix (5.17) can be used in general case. The focusing source correlation matrix (5.17) has yet another important property. For coherent sources the estimated source correlation matrix (5.11) might be singular. However, the average $\mathbf{S}_0$ is of full rank. Hence (5.17) removes the coherence by smoothing the spectrum of the source signals. With this averaging, the TCT algorithm can be applied to coherent cases.

As it is seen from (5.14), the error of focusing is a function of f_0. To minimize the error of focusing a suitable selection of the focusing frequency is needed. We seek a frequency f_0 that minimizes

$$\begin{aligned} \mathcal{E} = \min_{f_0} \; \min_{\mathbf{U}_i} \sum_{j=1}^{J} \| \mathbf{P}_0 - \mathbf{U}_j \mathbf{P}_j \mathbf{U}_j^H \|^2 \\ \text{s.t.} \quad \mathbf{U}_i^H \mathbf{U}_i = \mathbf{I}, \quad i = 1, \ldots, J. \end{aligned} \tag{5.18}$$

For a fixed $\mathbf{P}_0$ the transformation matrices $\mathbf{U}_j$ are obtained from (5.8). By using these

matrices in (5.18), the focusing error is given by

$$\mathcal{E} = \sum_{j=1}^{J} \left[\|\mathbf{P}_0\|^2 + \|\mathbf{P}_j\|^2 - 2\sum_{i=1}^{q} \sigma_i(\mathbf{P}_0)\sigma_i(\mathbf{P}_j) \right]. \tag{5.19}$$

Since $\mathbf{P}_j$'s are independent from the focusing frequency, f_0 can be determined from

$$\min_{f_0} \sum_{j=1}^{J} \left[\sum_{i=1}^{q} \sigma_i^2(\mathbf{P}_0) - 2\sum_{i=1}^{q} \sigma_i(\mathbf{P}_0)\sigma_i(\mathbf{P}_j) \right]. \tag{5.20}$$

To select the best focusing frequency we proceed as follows. First, we find the singular values of the optimum focusing subspace. Then using these values the focusing frequency is selected.

Let us define

$$\mu_i = \sum_{j=1}^{J} \sigma_i(\mathbf{P}_j). \tag{5.21}$$

Using this definition the criterion (5.20) is represented as

$$\min_{f_0} \sum_{i=1}^{q} \left[J\sigma_i^2(\mathbf{P}_0) - 2\mu_i\sigma_i(\mathbf{P}_0) \right]. \tag{5.22}$$

The minimum of (5.22) is achieved when

$$\sigma_i(\mathbf{P}_0) = \frac{\mu_i}{J}, \qquad i = 1, \ldots, q. \tag{5.23}$$

Due to structural constraint on $\mathbf{P}_0$, in general (5.23) is not attainable. Instead, we minimize

$$\min_{f_0} \sum_{i=1}^{q} \left| \sigma_i(\mathbf{P}_0) - \frac{\mu_i}{J} \right|^2. \tag{5.24}$$

This is a one-variable optimization problem and a search procedure can be applied to find the minimum point. In practice, it is sometimes convenient to choose a pre-defined frequency such as the center frequency of the spectrum for focusing. However, to improve the performance, a focusing frequency which produces the smallest error should be selected.

5.2.3. The TCT algorithm

The TCT algorithm is summarized as follows:

1) use an ordinary beamformer to scan the space and find an initial estimate of the DOAs;

2) express the output of the sensors as concatenated snapshots and apply a DFT in each snapshot to sample the spectrum of data;

3) form the location and the source correlation matrices using the results of the pre-processing step (5.11);

4) average the source correlation matrices to obtain the focusing source correlation matrix as in (5.17);

5) find $\mathbf{P}_0 = \mathbf{A}_0\mathbf{S}_0\mathbf{A}_0^H$ and the $\mathbf{P}_j$'s using (5.13);

6) determine the unitary transformation matrices (5.8);

7) multiply these matrices by the sample correlation matrices and average the results;

8) use a detection method (AIC, MDL, or PSC) to find the true number of sources;

9) apply MUSIC or any other high resolution spectral estimation method to find the DOAs;

10) to improve the performance, iterate steps 3 to 9.

Comparison of the new algorithm with the CSM method shows that the second part of Step 3 and Steps 4 and 5 do not have counterparts in CSM. The presence of these steps in TCT increases the complexity of computation. The increase in the computation is due to three parts: (*i*) estimating the noise power in each frequency bin, (*ii*) estimating the source correlation matrix from (5.11), (*iii*) forming the focusing correlation matrix $\mathbf{P}_0 = \mathbf{A}_0\mathbf{S}_0\mathbf{A}_0^H$. To estimate the noise power for each frequency bin we only need to compute a few of smallest eigenvalues of the sample correlation matrix. To find those

eigenvalues the Lanczos algorithm [12] that needs $O(p^2)$ flops for processing can be used. If we use (5.12) for noise power estimation, only one eigenvalue needs to be estimated. Furthermore, if it is known that the noise is white in frequency spectrum, the noise power estimation can be done for one frequency bin.

The source correlation matrix at the j-th frequency bin is found from (5.11) where with defining $\mathbf{B}_j = (\mathbf{A}_j^H \mathbf{A}_j)^{-1} \mathbf{A}_j^H$ can be written as

$$\hat{\mathbf{S}}_j = \mathbf{B}_j [\hat{\mathbf{R}}_j - \hat{\sigma}_j^2 \mathbf{I}] \mathbf{B}_j^H. \tag{5.25}$$

The computation of $\mathbf{B}_j$ is performed by inverting $(\mathbf{A}_j^H \mathbf{A}_j)$, which is a $q \times q$ Toeplitz matrix, and multiplying by $\mathbf{A}_j^H$. Since $(\mathbf{A}_j^H \mathbf{A}_j)$ is Toeplitz, it can be inverted in $O(q^2)$ flops [12]. Using $\mathbf{B}_j$ the source correlation matrix $\hat{\mathbf{S}}_j$ is obtained with two matrix multiplications. The focusing correlation matrix $\mathbf{P}_0$ can also be formed by two matrix multiplications. It should be noted that the increase of the computational complexity is usually small compared to the load of Step 6.

Another difference in the computational complexity of the two methods appears at Step 6 where an eigenvalue decomposition is performed. To find the unitary matrices in the CSM algorithm J singular value decompositions are needed. However, in TCT, $(J+1)$ eigenvalue decompositions of Hermitian matrices need to be performed which should result in a smaller computational load.

5.3. The error of transformation

In this section, we compare the error of transformation for the CSM and TCT algorithms. There are two sources of error for the transformation: the error due to noisy observation, and the error of transformation. The concern of the present section is the later one. In this section, we consider a noise-free environment where the error is only due to the focusing process.

The objective of focusing is to align the subspaces at different frequency bins. It should be noted that alignment of the subspaces alone is not sufficient for a good estimation. The

subspaces might be twisted in the process of focusing which causes augmentation of the noise in some directions. This in turn reduces the focusing SNR and might result in a biased estimation of DOAs. To prevent warping of the subspaces we can use unitary matrices for focusing. However, using unitary transformation matrices for focusing does not necessarily produce an unbiased estimate. Furthermore, there is not a unique solution for the unitary focusing matrices. Here, we define an error of focusing that can be used as a comparison measure between different focusing methods.

Since the objective is to transform the noise-free correlation matrix at each frequency bin to the focusing noise-free correlation matrix, the subspace fitting error is given by

$$\mathcal{E} = \|\mathbf{P}_0 - \mathbf{W}_j \mathbf{P}_j \mathbf{W}_j^H\|^2 \tag{5.26}$$

where $\mathbf{W}_j$ is the focusing matrix. The error of transformation for the TCT algorithm can be obtained by substituting $\mathbf{U}_j$ in (5.26) which simplifies to

$$\begin{aligned} \mathcal{E}_{\text{TCT}} &= \|\mathbf{P}_0\|^2 + \|\mathbf{P}_j\|^2 - 2\Re\,\text{tr}(\mathbf{P}_0 \mathbf{U}_j \mathbf{P}_j \mathbf{U}_j^H) \\ &= \|\mathbf{P}_0\|^2 + \|\mathbf{P}_j\|^2 - 2\sum_{i=1}^{q} \sigma_i(\mathbf{P}_0)\sigma_i(\mathbf{P}_j). \end{aligned} \tag{5.27}$$

Correspondingly, the error of transformation for the CSM algorithm is represented as

$$\mathcal{E}_{\text{CSM}} = \|\mathbf{P}_0\|^2 + \|\mathbf{P}_j\|^2 - 2\Re\,\text{tr}(\mathbf{P}_0 \mathbf{T}_j \mathbf{P}_j \mathbf{T}_j^H) \tag{5.28}$$

where $\mathbf{T}_j$ is the focusing matrix given by (4.14). It is possible to show that the error of transformation for the CSM algorithm is the same as

$$\mathcal{E}_{\text{CSM}} = \|\mathbf{P}_0\|^2 + \|\mathbf{P}_j\|^2 - 2\sum_{i=1}^{q} \sigma_i(\mathbf{P}_0 \mathbf{T}_j \mathbf{P}_j \mathbf{T}_j^H). \tag{5.29}$$

Using Lemma 5.1, this error can be written as

$$\mathcal{E}_{\text{CSM}} \geq \|\mathbf{P}_0\|^2 + \|\mathbf{P}_j\|^2 - 2\sum_{i=1}^{q} \sigma_i(\mathbf{P}_0)\sigma_i(\mathbf{T}_j \mathbf{P}_j \mathbf{T}_j^H) \tag{5.30}$$

$$= \|\mathbf{P}_0\|^2 + \|\mathbf{P}_j\|^2 - 2\sum_{i=1}^{q} \sigma_i(\mathbf{P}_0)\sigma_i(\mathbf{P}_j) \tag{5.31}$$

$$= \mathcal{E}_{\mathrm{TCT}} \tag{5.32}$$

where we have used the property that the matrices related with the similarity transformation have the same eigenvalues [15]. Thus, the error of transformation for the TCT algorithm is always smaller than that for CSM.

5.4. Eigenvalues of the universal focused correlation matrix

One of the major drawbacks of the CSM algorithm is the asymptotic bias of the peak locations. It is known that the CSM algorithm generates an estimate of DOA that is asymptotically biased [17]. The bias increases with the bandwidth of processing and deviation of the initial DOAs from the true DOAs. In this section, we study the eigenvalues of the universal focused correlation matrix of the CSM and TCT methods. Using the eigenvalues of the universal focused correlation matrix, we show that the signal power in the CSM algorithm is randomly extended into the noise subspace. This extension acts as a spatially colored noise with an unknown correlation matrix which produces biased estimates of the DOAs.

5.4.1. Analytical study

The universal focused correlation matrix for the CSM algorithm is found from

$$\mathbf{R}_{\mathrm{CSM}} = \frac{1}{J}\sum_{j=1}^{J} \mathbf{T}_j \hat{\mathbf{R}}_j \mathbf{T}_j^H. \tag{5.33}$$

To study the mechanism that generates the asymptotic bias, a noise-free environment is considered. In such a case, the correlation matrix $\hat{\mathbf{R}}_j$ is equal to the array noise-free correlation matrix $\mathbf{P}_j$. Thus the universal focused correlation matrix will be

$$\mathbf{R}_{\mathrm{CSM}} = \frac{1}{J}\sum_{j=1}^{J} \mathbf{T}_j \mathbf{P}_j \mathbf{T}_j^H$$

$$= \frac{1}{J}\sum_{j=1}^{J} \mathbf{V}_j \mathbf{W}_j^H \mathbf{A}_j \mathbf{S}_j \mathbf{A}_j^H \mathbf{W}_j \mathbf{V}_j^H. \tag{5.34}$$

As it is seen the focused correlation matrix is a function of the pre-estimate of the DOA and the bandwidth of sources. In general, $\mathbf{R}_{\text{CSM}}$ is of full rank and has nonzero eigenvalues in the noise subspace. In other words, the received power is distributed in a p-dimensional space. The components of the signal which diffuse into the noise subspace act as a nonwhite noise with an unknown correlation matrix. The MUSIC algorithm that operates on $\mathbf{R}_{\text{CSM}}$ will provide biased estimates of the DOAs if the spatial noise structure is unknown since for colored noise case, the eigenvectors of the correlation matrix corresponding to q largest eigenvalues are not in the space spaned by the location matrix.

For TCT the universal focused correlation matrix is given by

$$\begin{aligned} \mathbf{R}_{\text{TCT}} &= \frac{1}{J}\sum_{j=1}^{J} \mathbf{U}_j \mathbf{P}_j \mathbf{U}_j^H \\ &= \frac{1}{J}\sum_{j=1}^{J} \mathbf{X}_0 \mathbf{X}_j^H \mathbf{P}_j \mathbf{X}_j \mathbf{X}_0^H \end{aligned} \tag{5.35}$$

where (5.8) has been used to compute $\mathbf{U}_j$. Suppose that the diagonal matrices of the eigenvalues of $\mathbf{P}_j, j = 1, \ldots, J$, are shown by $\mathbf{\Gamma}_j, j = 1, \ldots, J$. Then for any pre-estimate of DOA, (5.35) simplifies to

$$\begin{aligned} \mathbf{R}_{\text{TCT}} &= \frac{1}{J}\sum_{j=1}^{J} \mathbf{X}_0 \mathbf{\Gamma}_j \mathbf{X}_0^H \\ &= \mathbf{X}_0 (\frac{1}{J}\sum_{j=1}^{J} \mathbf{\Gamma}_j) \mathbf{X}_0^H \\ &= \mathbf{X}_0 \mathbf{\Gamma}_0 \mathbf{X}_0^H. \end{aligned} \tag{5.36}$$

Since $\mathbf{P}_j$ is computed directly from the sample correlation matrix using (5.13), the diagonal matrix $\mathbf{\Gamma}_0$ is independent of the pre-estimates of DOA. Note that each diagonal matrix $\mathbf{\Gamma}_j$ has only q nonzero entries which in turn implies that $\mathbf{\Gamma}_0$ has only q nonzero components. The matrix $\mathbf{X}_0$ is orthonormal and hence (5.36) is an eigenvalue decomposition of $\mathbf{R}_{\text{TCT}}$.

It is concluded that in TCT the transformed subspaces for different frequencies are aligned and the eigenvalues at the noise subspace are zero. In other words, the focused correlation matrix $\mathbf{R}_{\mathrm{TCT}}$ has eigenvalues in a q-dimensional subspace. This is an important property of the TCT algorithm that makes it possible to provide unbiased estimates of the DOAs.

It is useful to compare (5.36) with perfect focusing. In perfect focusing the noise-free focused correlation matrix is given by $\mathbf{A}_0\mathbf{R}_s\mathbf{A}_0^H$ where $\mathbf{R}_s$ is defined in (4.10). The perfect focusing can be obtained using the transformation matrices (4.5). These transformation matrices are not unitary. In general it is not possible to establish perfect focusing through unitary transformations. In (5.36), perfect focusing is achieved by applying the transformation on the eigenvectors of the correlation matrices. Since the eigenvectors form an orthonormal basis, it is always possible to use a unitary transformation to transfer them into another orthonormal basis. Note that the true DOA implicitly effects the selection of the focusing subspace through the estimation of $\mathbf{S}_j$. Thus assignment of the orthonormal basis for the focusing subspace in the TCT method is implied by the true DOAs.

5.4.2. Experimental results

We present the results of a computer simulation to study the eigenvalues of the universal correlation matrix of the two methods. As an example, a configuration with 4 equipower wideband signals arriving at a linear array of 16 sensors in a noise-free environment is considered. The true DOAs are 8°, 13°, 33°, and 37° degrees. The spectrum of the signals is flat with 40% relative bandwidth. The initial DOAs are taken at 6.7°, 10.5°, 14.3°, 31°, 35°, and 39° degrees. The output of the sensors is decomposed into 50 snapshots with each snapshot containing 64 samples. An FFT algorithm is used in each snapshot to sample the frequency spectrum of the signals at 33 equispaced points. We applied the CSM and TCT algorithms to obtain the focusing matrices. The eigenvalues of the corresponding matrices are tabulated in Table 5.1. It is seen that $\mathbf{R}_{\mathrm{CSM}}$ has nonzero eigenvalues in the noise subspace due to signal diffusion. Since the focusing matrices, $\mathbf{T}_j$ and $\mathbf{U}_j$, are unitary, the trace of $\mathbf{R}_{\mathrm{CSM}}$ is equal to the trace of $\mathbf{R}_{\mathrm{TCT}}$. This means that the summation of eigenvalues in Table 5.1 is identical for each matrix. This suggests that the energy of the

eigenvalues	$\mathbf{R}_{\text{CSM}}$	$\mathbf{R}_{\text{TCT}}$
λ_1	900.27	918.76
λ_2	806.70	808.09
λ_3	241.68	244.14
λ_4	155.39	140.99
λ_5	4.39	0.00
λ_6	1.35	0.00
λ_7	0.40	0.00
λ_8	0.34	0.00
λ_9	0.31	0.00
λ_{10}	0.27	0.00
λ_{11}	0.22	0.00
λ_{12}	0.19	0.00
λ_{13}	0.15	0.00
λ_{14}	0.12	0.00
λ_{15}	0.09	0.00
λ_{16}	0.07	0.00

Table 5.1 The eigenvalues of the correlation matrices $\mathbf{R}_{\text{CSM}}$ and $\mathbf{R}_{\text{TCT}}$ for a configuration of 4 wideband sources. The sources have 40% relative bandwidth and are arriving from the angles 8°, 13°, 33°, and 37° degrees at a linear array of 16 sensors in a noise-free environment.

signals after transformation in the two methods is identical. However, the TCT method condenses the total received energy in a q-dimensional subspace and hence improves the performance.

5.5. The Bias of estimation

One of the major motivations for introducing the TCT algorithm is to reduce the asymptotic bias of the peaks in CSM. It is important to note that the MUSIC algorithm is intrinsically unbiased. The bias in CSM is introduced by focusing which implies that one can reduce the bias of estimation with a proper selection of the focusing method. In the sequel, we will discuss this issue and show that TCT can asymptotically generate unbiased estimate of the DOAs. We will also generalize the work of Swingler and Krolik [36]. They showed that for a single-source scenario it is possible to have an unbiased estimate provided that the focusing frequency is chosen at the centroid of the source spectrum.

Here we show that *for a multi-source case the bias is eliminated if the focusing correlation matrix for the true values of DOA is at the average of the transformed correlation matrices.* Since the bias of estimation is independent of the noise, in the rest of this section a noise-free environment is considered.

The noise-free universal focused correlation matrix is shown as

$$\mathbf{R} = \frac{1}{J}\sum_{j=1}^{J} \mathbf{W}_j \mathbf{P}_j \mathbf{W}_j^H \tag{5.37}$$

where $\mathbf{W}_j$ is the focusing matrix for the j-th frequency bin. The following lemma gives a necessary and sufficient condition for unbiased estimation of the DOAs.

Lemma 5.2. *For q uncorrelated sources in a noise-free environment the MUSIC estimator of the DOAs is unbiased if and only if*

$$\text{Span}\{\bar{\mathbf{A}}_0\} = \text{Span}\{\mathbf{V}_q\} \tag{5.38}$$

where $\bar{\mathbf{A}}_0$ is the true location matrix at the focusing frequency and $\mathbf{V}_q$ is the matrix of the q largest eigenvectors of $\mathbf{R}$.

Proof: The MUSIC estimator intersects the subspace spanned by the q largest eigenvectors of the correlation matrix with the array manifold. Therefore if (5.38) is satisfied the estimation is unbiased. For the necessary condition recall that the array location matrix is of full rank for any combination of the DOAs. Thus if (5.38) is not satisfied, there are q independent vectors on the array manifold which span the same space as $\mathbf{V}_q$. The matrix of these vectors is the estimated location matrix which is different from $\bar{\mathbf{A}}_0$. □

We showed earlier that the CSM algorithm extends the power of the sources into the noise subspace. Thus the rank of $\mathbf{R}$ is larger than the rank of $\bar{\mathbf{A}}_0$. For an unbiased estimation of the DOAs, the span of the column vectors of $\bar{\mathbf{A}}_0$ should be equal to the subspace spanned by those eigenvectors of $\mathbf{R}$ that correspond to the q largest eigenvalues. We assume that the space spanned by the column vectors of $\bar{\mathbf{A}}_0$ is a subspace of the span of the columns of $\mathbf{R}$. This is an appropriate assumption since otherwise the estimation will

be biased. Note that the sample correlation matrix $\mathbf{R}$ is a Wishart distributed random matrix and with probability one its q largest eigenvectors will not span the same subspace as $\bar{\mathbf{A}}_0$. In other words, with probability one the CSM algorithm provides biased estimates of the DOAs. On the contrary, TCT condenses the power of the sources in a q-dimensional subspace. Therefore it can produce unbiased estimates provided the focusing matrix, $\mathbf{U}_j$, transforms the space spanned by the columns of $\mathbf{P}_j$ into the space spanned by $\bar{\mathbf{A}}_0$. A sufficient condition for unbiased estimation is given by the following lemma.

Lemma 5.3. *Any focusing method that satisfies*

$$\bar{\mathbf{A}}_0\mathbf{S}_0\bar{\mathbf{A}}_0^H = \frac{1}{J}\sum_{j=1}^{J}\mathbf{W}_j\mathbf{P}_j\mathbf{W}_j^H \tag{5.39}$$

where $\mathbf{S}_0$ is a nonsingular, Hermitian, positive-definite matrix, produces an unbiased estimate of the DOAs.

Proof: Note that (5.39) satisfies (5.38). Thus using Lemma 5.2 the proof is complete. □

It is important to notice that (5.39) is a general condition for unbiased estimation regardless of the method which has been applied for focusing. In [36] Swingler and Krolik showed that for a single source scenario, an unbiased estimation of the DOA is possible if the centroid of the frequency spectrum is chosen as the focusing frequency. Lemma 5.3 shows that in a multi-source case, the focusing correlation matrix should be the average of the focused correlation matrices for an unbiased estimation. We show in Appendix E that the same results as [36] can be obtained using (5.39) with diagonal focusing matrices in a single source scenario.

For further discussion of the bias generating mechanism we consider the special case of perfect focusing. In perfect focusing the transformed correlation matrices $\mathbf{U}_j\mathbf{A}_j$ are superimposed on $\mathbf{A}_0$. In such a case the focusing correlation matrix is an average of the correlation matrices at the frequency bins and the following equality is satisfied,

$$\mathbf{A}_0\mathbf{S}_0\mathbf{A}_0^H = \frac{1}{J}\sum_{j=1}^{J}\mathbf{U}_j\mathbf{A}_j\mathbf{S}_j\mathbf{A}_j^H\mathbf{U}_j^H. \tag{5.40}$$

As noted earlier, for any $\hat{\theta}_i$ estimated by the pre-processing, step the focusing points for the i-th angle are chosen at $(\hat{\theta}_i - 0.25B_W \ , \ \hat{\theta}_i \ , \ \hat{\theta}_i + 0.25B_W)$. These angles determine an interval on the array manifold. If this interval is small compared to the curvature of the array manifold at all the points in that interval, it is transferred to a corresponding interval on the array manifold at the processing frequency f_0. This is a consequence of the continuity of the array manifold and the unitary transformation. Thus in perfect focusing the location vectors of each frequency bin that are located at the true DOA are transferred to the corresponding vectors at the focusing manifold. It is seen that in such a case the estimation of DOA can be unbiased.

In practice, perfect focusing is not plausible. The transformed matrices are clustered around $\mathbf{A}_0$. However, as far as the equality (5.39) is satisfied for the true DOA, estimation could be unbiased. It is straightforward to show that the TCT algorithm forms a very good approximation of (5.39). Taking gradient of (5.18) with respect to $\mathbf{P}_0$ and equating to zero proves that the minimum of (5.18) is achieved if and only if $\mathbf{P}_0$ is the average of the matrices $\mathbf{U}_j\mathbf{P}_j\mathbf{U}_j^H$, $j = 1, \ldots, J$. The TCT algorithm is based on the minimization of (5.18) which means that it places the focusing correlation matrix at the centroid of the matrices $\mathbf{U}_j\mathbf{P}_j\mathbf{U}_j^H$, $j = 1, \ldots, J$. However, because of unknown DOAs and structural constraints on the correlation matrix it is not possible to satisfy (5.39) with equality. The TCT algorithm provides a very close approximation to (5.39) which explains its capability to provide asymptotically unbiased estimates.

5.6. The effect of noise on estimation

It is possible to show that the sample correlation matrices of the two methods, CSM and TCT, are Wishart distributed [3] with JN degrees of freedom with the correlation matrices $\mathbf{R}_{\text{CSM}}$ and $\mathbf{R}_{\text{TCT}}$, respectively. Note that simply considering the degrees of freedom is not sufficient for comparison, since any transformation of the signal subspaces results in a correlation matrix which is Wishart distributed with JN degrees of freedom. The important factor is how the transformed observation vectors are distributed in the p-dimensional subspace. In this section, we will show that the variance of the noise is

smaller for the TCT which results in a better estimate of the directions of arrival.

The observation can be considered as a p-dimensional signal vector in the q-dimensional signal subspace perturbed by a p-dimensional noise vector. It is important to note that the noise component in the signal subspace has no effect in the estimation process since the MUSIC algorithm estimates the DOA's by intersecting the signal subspace with the array manifold. If the noise is restricted to the signal subspace, there will be no error in the estimation. However, the power of the noise in the noise subspace is very important in introducing error in the estimation. The effect of the orthogonal noise can be discussed based on the theory of generalized variance [3]. The *generalized variance* of a multivariate random vector is defined as the determinant of the correlation matrix. The generalized variance is a metric for the spread of the observation and is equal to the sum of squares of the volumes of all different parallelotopes formed by using any p observation vectors as the principal edges.

In a noise-free environment the observation vectors are constrained to the signal subspace and all the parallelotopes have zero volume in the p-dimensional observation space. Thus, the generalized variance in such a case is zero. This is the key point to the estimation of the DOA's without error. When the noise is introduced in the system, the observation vectors are expanded into the noise subspace. Extension of the signal vectors into the noise subspace results in a nonzero volume and hence a nonzero generalized variance. Thus, for a noisy environment the estimation of DOA's is usually yielded with error. The smaller the volume of the observation vectors is, the better the DOA's are estimated. Therefore a small generalized variance results in a small variance of estimation. In what follows we show that the generalized variance of the TCT algorithm is smaller than that for CSM.

Consider the following maximization problem

$$\begin{aligned} \max \quad & |\mathbf{R}| \\ \text{s.t.} \quad & \mathrm{tr}\mathbf{R} = \sum_{j=1}^{J} \mathrm{tr}\hat{\mathbf{R}}_j \\ & \mathbf{R} > 0 \end{aligned} \tag{5.41}$$

where $|.|$ stands for the determinant of a matrix and $\mathbf{R} > 0$ means that the matrix is positive-definite. The first constraint in (5.41) assures that the energy is not lost during the focusing process. It is known that the maximum of $|\mathbf{R}|$ is obtained if all the eigenvalues are equal. The maximum corresponds to an equilibrium point where the energy is equally distributed in all dimensions. Assume that the eigenvalues are components of a vector. The vector of the eigenvalues that satisfy the constraints of (5.41) are located in a portion of a hyperplane that is cut by the positive quadrant. The equilibrium point which is the solution of (5.41) is at the centroid of this region. The maximization (5.41) is a convex problem. Thus, the farther we are from the equilibrium point, the smaller $|\mathbf{R}|$ will be.

We have shown in Section 5.4 that in the CSM algorithm, the signal energy is extended into the noise subspace. In other words, the eigenvalues of $\mathbf{R}_{\text{TCT}}$ in the noise subspace are smaller than the eigenvalues of the $\mathbf{R}_{\text{CSM}}$ while the sum of eigenvalues for the two methods are equal. Thus, the vector of eigenvalues is farther from the equilibrium point in the TCT algorithm than in the CSM algorithm. This means that the universal focused correlation matrix of the TCT algorithm has smaller determinant or generalized variance. The smaller generalized variance of the TCT algorithm results in a smaller variance in the estimation of DOA's.

5.7. Performance comparison

Recently, Doron and Weiss [11] introduced a method for wideband array processing using Signal Subspace Transformation (SST). The focusing matrix in their method is found from minimization of $\|\mathbf{A}_0\mathbf{D}_0\mathbf{A}_0^H - \mathbf{T}_j\mathbf{A}_j\mathbf{D}_j\mathbf{A}_j^H\mathbf{T}_j^H\|$, where $\mathbf{D}_0$ and $\mathbf{D}_j$ are any Hermitian positive definite matrices. They used identity matrices for $\mathbf{D}_0$ and $\mathbf{D}_j$ in their simulation. The results of the simulation shows that the method is biased. There are two major differences between the TCT and the SST algorithms. First, in TCT, the average of the estimated source correlation matrices at the frequency bins is chosen as the focusing source correlation matrix in place of $\mathbf{D}_0$. Second, instead of $\mathbf{A}_j\mathbf{D}_j\mathbf{A}_j^H$ the estimated noise-free correlation matrices $\mathbf{P}_j$ are used for focusing matrix determination. Note that in $\mathbf{A}_j\mathbf{D}_j\mathbf{A}_j^H$ the estimated DOAs are utilized, however, for $\mathbf{P}_j$ the true DOAs are implicitly used.

There is also a difference between the model equations of the CSM and the TCT algorithms. In CSM, the focusing model equations are $\mathbf{T}_j\mathbf{A}_j = \mathbf{A}_0$. The solution to this equation is not unique and might be singular. It has been shown [17] that from an estimation point of view, unitary transformations of signal subspaces are the most effective focusing methods for direction finding. In contrast, in TCT, the model equations are $\mathbf{U}_j\mathbf{P}_j\mathbf{U}_j^H = \mathbf{P}_0$, where $\mathbf{P}_j = \hat{\mathbf{R}}_j - \hat{\sigma}_j^2\mathbf{I}$, and $\mathbf{P}_0 = \mathbf{A}_0\mathbf{S}_0\mathbf{A}_0^H$, where $\mathbf{S}_0$ is given by (5.17). Note that the noise-free correlation matrix $\mathbf{P}_j$ is directly estimated from the data and the pre-estimate of the DOAs is not used in its determination.

5.7.1. Simulation results

Here, we present the simulation results for two DOA estimation scenarios. In the first example a configuration with two sources is considered. For this example, we have compared the bias, the resolution threshold, and the spatial spectrum of the MUSIC algorithm of the CSM and TCT methods. The second example is a multi-group DOA estimation problem with the angles taken from [17]. We have found the bias of the estimation of the CSM and TCT algorithms and compared them. We have also shown that the TCT algorithm can locate coherent sources.

Two sources

In the first example, we investigate a configuration with two equipower uncorrelated sources arriving from the angles 11° and 13° degrees off broadside. The signal-to-noise ratio is 10 dB. A linear array of 8 sensors is used. The spacing between adjacent sensors is equal to half the wavelength at the center frequency. A preliminary beamformer output gives a peak at 12 degrees. Two extra focusing points are added at 9° and 15° degrees. Sources are sampled with 33 frequency bins in the frequency domain. We imported the actual correlation matrix to the CSM and TCT algorithms and used the high resolution MUSIC algorithm for DOA estimation. The results of the estimation for 40 and 100 percent bandwidth and for different focusing frequencies are given in Table 5.2. The bias columns in this table are the Euclidean norm of the bias vectors. TCT does not have bias

BW = 0.4	CSM			TCT		
f_0	11	13	bias	11	13	bias
0.8	11.89	12.13	1.24	11.00	13.00	0.00
0.9	11.25	12.75	0.35	11.00	13.00	0.00
1.0	11.01	12.99	0.01	11.00	13.00	0.00
1.1	10.88	13.12	0.17	11.00	13.00	0.00
1.2	10.78	13.22	0.31	11.00	13.00	0.00
BW = 1.0	CSM			TCT		
f_0	11	13	bias	11	13	bias
0.8	–	12.01	–	11.01	12.99	0.01
0.9	11.42	12.58	0.59	11.00	13.00	0.00
1.0	11.12	12.88	0.17	11.00	13.00	0.00
1.1	10.95	13.05	0.07	11.00	13.00	0.00
1.2	10.84	13.16	0.23	11.00	13.00	0.00

Table 5.2 The estimation results for the first example.

regardless of the bandwidth of processing.

For this example, we can investigate the threshold SNR for the two methods. The output of each sensor is separated to 50 snapshots of 16 samples each. Then an FFT algorithm is applied in each snapshot to sample the spectrum of the observation. To find the resolution threshold 100 trials of the same scenario was run for each SNR. The number of times each algorithm resolved the sources was counted to estimate the probability of resolution. The sources were assumed to be resolved when two peaks in the spatial spectrum of the MUSIC algorithm were observed. Fig. 5.1 shows the probability of the resolution for the two methods. Two versions of the CSM algorithm are used here. By UCSM and DCSM we mean unitary and diagonal versions of the CSM algorithm, respectively. For DCSM the focusing angle is chosen at 12° degrees. It is seen that TCT has a lower SNR threshold compared to the UCSM and DCSM algorithms.

We also examine the resolution capability of the two algorithms. We increase the number of sensors to 16 and consider a 40 percent bandwidth. It is assumed that only 20 snapshots of data are available. Again at each snapshot a 64-point DFT is applied to obtain 33 frequency samples in the frequency domain. The resolution criterion is defined

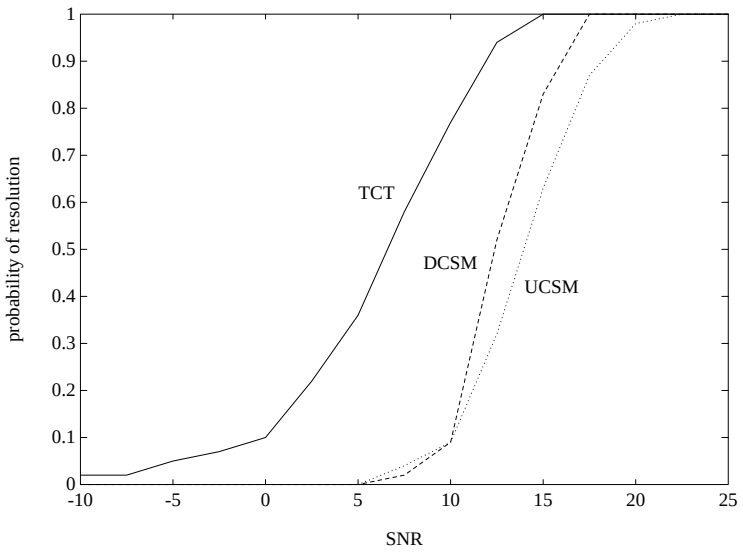

Fig. 5.1 The probability of resolution for two closely separated sources using the TCT and CSM algorithms.

as the difference between the average of the spatial spectrum at the peak points in the MUSIC algorithm and the spatial spectrum in the valley. It is measured on a dB scale for different SNR's. The results are given in Fig. 5.2. As it is seen the performance of TCT is about 6 dB better than CSM. The spatial spectra of the two methods are overlapped in Fig. 5.3 for comparison.

Four sources

For the second configuration we investigate Example 1 in [17]. The same array has been used to estimate the DOA of 4 equipower uncorrelated sources impinging from 8°, 13°, 33° and 37° degrees. The bandwidth of the sources is equal to 40 and 100 percent of the center frequency in two different trials. The pre-estimates of the DOA are given by 6.7°, 10.5°, 14.3°, 31°, 35°, and 39° degrees. The true cross correlation matrices are used for estimation. Table 5.3 presents the results of estimation for this example. Although TCT performs better, it is not an unbiased estimation procedure. To improve the resolution and reduce or eliminate the bias, we can iterate the algorithm. We consider an array of 16 sensors with the same signals as the second example. Application of the TCT method

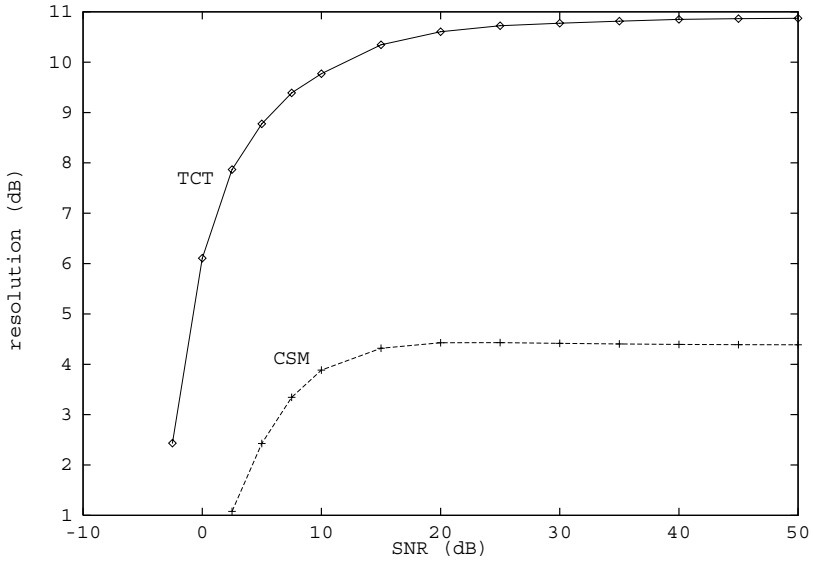

Fig. 5.2 Comparison of the resolution of the two algorithms, TCT and CSM. The resolution criterion is defined as the difference between the average of the spatial spectrum at the peak points in the MUSIC algorithm and the spatial spectrum in the valley.

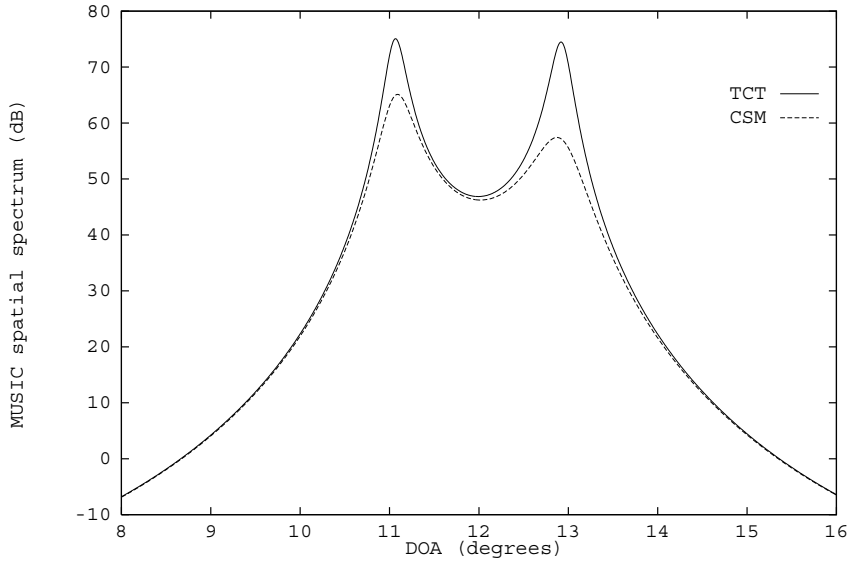

Fig. 5.3 The MUSIC spatial spectrum for the two methods, TCT and CSM.

BW=0.4			CSM					TCT		
f_0	8	13	33	37	bias	8	13	33	37	bias
0.8	–	11.75	34.25	–	–	8.03	13.03	32.98	36.98	0.05
0.9	8.42	12.55	33.90	–	–	8.01	13.01	32.98	36.98	0.03
1.0	8.07	13.01	33.12	36.82	0.23	8.02	13.02	32.97	36.97	0.05
1.1	7.87	13.21	32.53	37.35	0.64	8.07	13.06	32.92	36.91	0.15
1.2	7.72	13.29	32.22	37.66	1.10	8.17	13.13	32.83	36.81	0.33
BW=1.0			CSM					TCT		
f_0	8	13	33	37	bias	8	13	33	37	bias
0.8	–	11.17	34.22	–	–	8.04	13.08	32.86	36.88	0.20
0.9	–	11.85	33.88	–	–	8.07	13.09	32.88	36.89	0.20
1.0	8.25	12.53	33.58	–	–	8.10	13.10	32.87	36.87	0.23
1.1	7.98	12.93	33.21	36.97	0.22	8.15	13.13	32.85	36.85	0.29
1.2	7.85	13.13	32.78	37.44	0.53	8.22	13.19	32.72	36.71	0.50

Table 5.3 The estimation results for the second example. The blank entries mean that CSM did not resolve the DOA.

BW=0.4			CSM					TCT		
f_0	8	13	33	37	bias	8	13	33	37	bias
0.8	8.04	12.99	33.03	36.90	0.11	8.00	13.00	33.00	37.00	0.00
0.9	8.02	12.99	33.03	36.95	0.06	8.00	13.00	33.00	37.00	0.00
1.0	7.99	12.99	33.04	37.03	0.05	8.00	13.00	33.00	37.00	0.00
1.1	7.99	13.01	32.98	37.03	0.04	8.00	13.00	33.00	37.00	0.00
1.2	7.98	13.03	32.94	37.03	0.08	8.00	13.00	33.00	37.00	0.00

Table 5.4 The estimation results for the second example after two iterations.

gives the estimates of the DOAs at 7.94°, 13.03°, 33.09° and 37.08° degrees. We use the following angles for the pre-estimates: 7°, 7.94°, 9°, 12°, 13.03°, 14°, 32°, 33.09°, 34°, 36°, 37.08° and 38° degrees. The results for the two methods are given in Table 5.4. As it is seen the TCT algorithm outperforms CSM and removes the bias of the estimation. In general to eliminate the bias this procedure must be iterated several times.

We can also investigate the capability of the TCT algorithm to resolve coherent sources. In Fig. 5.4, the MUSIC spatial spectrum for the coherent source scenario at 10 dB SNR is depicted. It is assumed that the source at 13° is a delayed version of the source at 8° with a one sampling time delay. As it is seen the TCT algorithm resolves all the sources including the coherent ones.

To study the performance of the TCT algorithm for a limited number of observations,

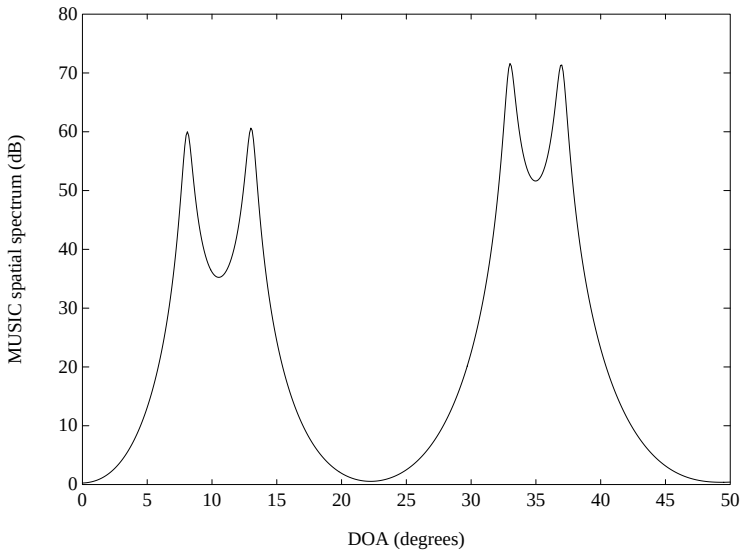

Fig. 5.4 The MUSIC spatial spectrum for four sources at 8°, 13°, 33°, and 37°. The source at 13° is a delayed version of the source at 8°.

the same 4 sources are received by a linear array of 16 sensors. The observation interval is decomposed into 40 snapshots with each containing 32 samples. The focusing angles are taken at 6.7°, 10.5°, 14.3°, 31°, 35°, and 39° degrees. The SNR is varied and the bias and the variance are averaged over 100 independent trials. In Fig. 5.5 and Fig. 5.6 the norm of the bias and the variance vectors for the two methods are compared. These examples show that the TCT algorithm has smaller bias and variance for a limited number of observations.

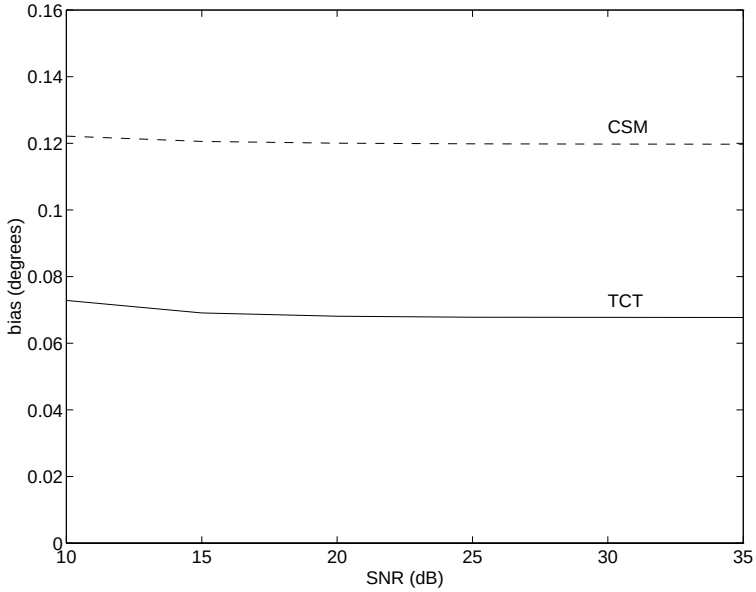

Fig. 5.5 The norm of the averaged bias vector for limited number of observations (40 snapshots of 32 samples) versus SNR for a configuration with four uncorrelated sources.

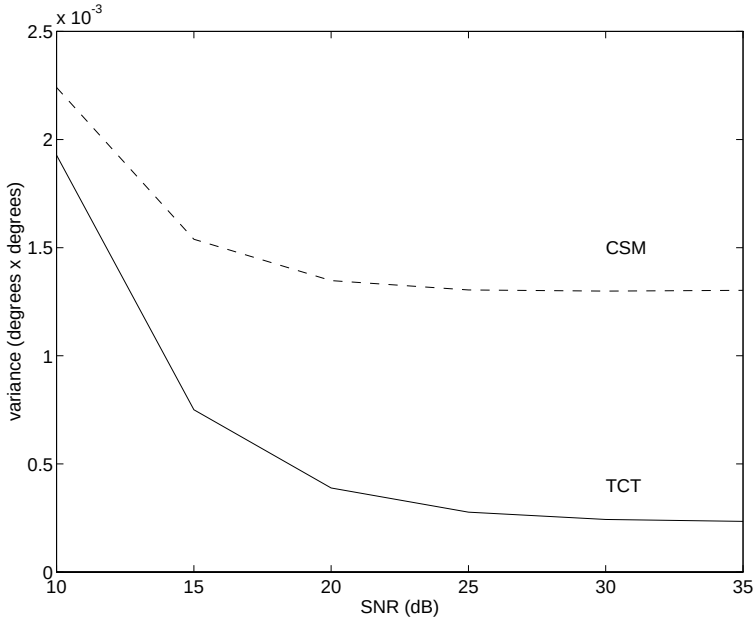

Fig. 5.6 The norm of the averaged variance vector for limited number of observations (40 snapshots of 32 samples) versus SNR for a configuration with 4 uncorrelated sources.

Chapter 6

Distributed Source

In array processing it is frequently assumed that the signals of interest are generated by point sources. This is a modeling constraint that is seldom satisfied in reality. Many practical examples can be found where the point source assumption does not hold. In an undersea echo beam sounder, the reflection of the signal and penetration into the lower levels of the seabed creates a spatial distribution of the receiving waveform [19]. Another example arises in the context of radar using short signal pulses. The received signal is a superposition of the reflections of the pulse from different parts of the target. If the target is spread in range, it appears as a distributed source [50]. A source distribution in space can also be observed in the transmission of radio-waves through ionospheric and tropospheric scatter links and the propagation of audio signals in a reverberant room.

For narrowband point-source configurations, the dimension of the signal subspace (defined as the span of the correlation matrix in a noise-free environment) is equal to the number of noncoherent signals. Thus, each source has a one-dimensional representation in the signal subspace. In previous work, distributed sources have been viewed as a combination of a large number of closely spaced point sources [19]. The corresponding location matrix generally spans the whole space and the noise subspace is empty. This explains why the conventional array processing techniques such as MUSIC [30] and ESPRIT [28], which are based on the signal and noise subspace decomposition for point source scenarios, lead to erroneous results when applied to such models of distributed sources.

In this chapter, we propose a parametric approach to distributed narrowband source localization (Signal C in Fig. 1.1). We assume that the angular cross-correlation kernel is chosen from a parametric class of functions with each function in this class being represented by a parameter vector. With this assumption, the localization problem becomes the one of parameter estimation. The proposed method is based on an extension of the MUSIC algorithm for distributed source localization. In this approach, the noise-free observation vector is modeled as the output of a linear operator. Using the adjoint of this operator, the noise eigenvectors are transformed back into an appropriate source subspace. The parameter vector is estimated by minimizing a norm of the transformed noise eigenvectors. Two cases are considered for the distributed source. If the signal rays of the source for all the angles at the distribution width are completely correlated, the signal is coherently distributed (CD). When the signal rays at different angles are uncorrelated the source is incoherently distributed (ID). The proposed localization method is based on generalization of the MUSIC algorithm and is applied to both CD and ID signals.

The dimension of the signal subspace depends on the assumptions on the signal distribution. For CD signals the dimension of the signal subspace is equal to the number of sources. For this case, the generalization of the MUSIC algorithm involves defining a new array manifold which is based on the integration of the conventional array location vector weighted by the angular signal density. If the signals are incoherently distributed the dimension of the signal subspace is generally equal to the number of sensors and the signal component spans the whole observation space. In such a case, the algorithms which are based on the signal and noise subspace decomposition are not applicable directly.

We study the shortcomings of beamforming techniques by examining the array gain. For distributed signals, the array gain is bounded and cannot be increased linearly with the number of sensors. Specifically, it is shown that for coherently distributed sources the array gain decreases for a very large number of sensors.

To compare our method to the conventional MUSIC algorithm we have performed a numerical study of the two methods. The simulation results show a dramatic improvement in performance for the new algorithm. In our example, the resolution SNR threshold for

the new method is considerably lower than for the MUSIC algorithm. It is also observed that, unlike MUSIC, the new method is asymptotically unbiased.

6.1. Models for spatially distributed sources

Consider an array of p sensors monitoring a wave field of q spatially distributed narrowband sources in additive background noise. For simplicity, it is assumed that the sensors and the sources are in the same plane, however, the method can be easily extended to the 3-dimensional case. The complex envelope representation of the array output observation vector at the k-th snapshot is given by

$$\mathbf{x}_k = \sum_{i=1}^{q} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathbf{a}(\theta) s_{i,k}(\theta; \boldsymbol{\psi}_i) d\theta + \mathbf{n}_k \tag{6.1}$$

where $\mathbf{a}(\theta)$ is the $p \times 1$ location vector of the array, $s_{i,k}(\theta; \boldsymbol{\psi}_i)$ is the angular signal density of the i-th source in the direction θ, and $\mathbf{n}_k$ is the $p \times 1$ additive noise vector. A parametric representation is assumed for $s_{i,k}(\theta; \boldsymbol{\psi}_i)$ with $\boldsymbol{\psi}_i$ being the corresponding parameter vector. Examples of the parameters are the two limits of the direction of arrival (DOA) for a uniform angular signal density or the angle of maximum power and the -3 dB extension width for a bell-shaped density. In the sequel, we drop the snapshot subscript k for simplicity of notation.

For a uniform linear array the location vector is given by $\mathbf{a}(\theta) = [1 \ \mu_\theta \ \ldots \ \mu_\theta^{p-1}]^T$, where $\mu_\theta = \exp(j\omega_0 d \sin(\theta)/v)$, and T indicates the transpose of a vector. In μ_θ, d is the distance between two consecutive sensors, v is the wave speed, and ω_0 is the frequency of the source signal. We have assumed that the phase reference point is at the first sensor. If the distance d is equal to half the wavelength at frequency ω_0, μ_θ will be given by $\mu_\theta = \exp(j\pi \sin\theta)$. The localization method that we present is applicable to an array with arbitrary geometry. However, in this chapter a linear array has been used in the computer simulations.

The signal and noise snapshots are modeled as independent, zero-mean, circular, complex Gaussian random variables. The signals and noise are also assumed to be uncorre-

lated. The correlation matrix of the noise vector $\mathbf{n}$ is known but for a scale factor, σ_n^2. In the sequel, we will consider only spatially white noise. Generalization to the nonwhite case can be done by pre-whitening. With these assumptions the correlation matrix of the observation vector $\mathbf{x}$ is given by

$$\begin{aligned} \mathbf{R}_x &= E(\mathbf{x}\mathbf{x}^H) \\ &= \mathbf{R}_s(\boldsymbol{\psi}) + \sigma_n^2 \mathbf{I} \end{aligned} \tag{6.2}$$

where

$$\mathbf{R}_s(\boldsymbol{\psi}) = \sum_{i=1}^{q} \sum_{j=1}^{q} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathbf{a}(\theta)\, \rho_{ij}(\theta, \theta'; \boldsymbol{\psi}_i, \boldsymbol{\psi}_j)\, \mathbf{a}^H(\theta')\, d\theta\, d\theta' \tag{6.3}$$

is the noise-free correlation matrix, $\mathbf{I}$ is the $p \times p$ identity matrix, and

$$\rho_{ij}(\theta, \theta'; \boldsymbol{\psi}_i, \boldsymbol{\psi}_j) = E[s_i(\theta; \boldsymbol{\psi}_i) s_j^*(\theta'; \boldsymbol{\psi}_j)] \tag{6.4}$$

is the angular cross-correlation kernel which is parameterized with the unknown parameter vectors $\boldsymbol{\psi}_i$ and $\boldsymbol{\psi}_j$. The superscripts H and $*$ represent Hermitian transposition and the complex conjugation, respectively.

If the signals from different sources are uncorrelated, the angular cross-correlation kernel simplifies to

$$\rho_{ij}(\theta, \theta'; \boldsymbol{\psi}_i, \boldsymbol{\psi}_j) = \rho(\theta, \theta'; \boldsymbol{\psi}_i) \delta_{ij} \tag{6.5}$$

where δ_{ij} is the Kronecker delta. Here, it is assumed that the shape of the angular correlation kernel is fixed and just the parameters of the shape change. The noise-free correlation matrix (6.3) is then given by

$$\mathbf{R}_s(\boldsymbol{\psi}) = \sum_{i=1}^{q} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathbf{a}(\theta)\, \rho(\theta, \theta'; \boldsymbol{\psi}_i)\, \mathbf{a}^H(\theta')\, d\theta\, d\theta'. \tag{6.6}$$

In the rest of the chapter, we assume that the sources are uncorrelated. Note that this is not a restrictive assumption. If the sources are correlated, they can be treated as a single source with a new angular correlation kernel which is the addition of the angular

correlation kernels of the sources. We consider two cases.

Case I – Coherently distributed signal

A kernel $\rho(\theta, \theta'; \boldsymbol{\psi})$ is called separable if it can be written as

$$\rho(\theta, \theta'; \boldsymbol{\psi}) = \sum_{k=1}^{K} \eta_k \phi_k(\theta; \boldsymbol{\psi}) \phi_k^*(\theta'; \boldsymbol{\psi}) \tag{6.7}$$

where K is a finite integer, η_k is a scalar and $\phi_k(\theta; \boldsymbol{\psi})$ is a function of θ and the parameter vector $\boldsymbol{\psi}$. This kernel is frequently invoked in multi-target radar problems where K is the number of targets [50]. For a single target the separable kernel is shown as

$$\rho(\theta, \theta'; \boldsymbol{\psi}) = \eta g(\theta; \boldsymbol{\psi}) g^*(\theta'; \boldsymbol{\psi}) \tag{6.8}$$

where $g(\theta; \boldsymbol{\psi})$ is a complex-valued deterministic angular signal density defined in the interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$. In our study, it is convenient to normalize the function $g(\theta; \boldsymbol{\psi})$,

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} g(\theta; \boldsymbol{\psi}) d\theta = 1. \tag{6.9}$$

With this normalization, the scalar η in (6.8) represents the signal power observed at the reference point of the array. A signal with the angular correlation kernel (6.8) is termed a *coherently distributed* (CD) signal.

Case II – Incoherently distributed signal

In some applications, the signal rays arriving from different directions can be assumed uncorrelated. For example in the transmission of the radio-waves through tropospheric scatter links, the signal rays reflected from different layers of the troposphere have uncorrelated phases. A similar effect is observed when the signal rays are the reflections from different parts of a rough surface[1]. The angular correlation kernel for such a case is

[1] According to the Rayleigh criterion, a surface is rough if $h \sin\theta > \lambda/8$, where h is the height of the roughness in the surface, θ is the reflection angle measured from the normal, and λ is the wavelength of the reflected signal.

written as

$$\rho(\theta, \theta'; \boldsymbol{\psi}) = \rho(\theta; \boldsymbol{\psi})\delta(\theta - \theta') \tag{6.10}$$

where $\delta(\theta)$ is the Dirac delta function. A distributed source with the angular correlation kernel (6.10) is called the *incoherently distributed* (ID) signal. The noise-free correlation matrix of these signals is shown as

$$\mathbf{R}_s(\boldsymbol{\psi}) = \sum_{i=1}^{q} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathbf{a}(\theta)\, \rho(\theta; \boldsymbol{\psi}_i)\, \mathbf{a}^H(\theta)\, d\theta. \tag{6.11}$$

In this chapter, the main objective is to locate the distributed signals. The localization is done by estimating the unknown parameter vectors $\boldsymbol{\psi}_i$ of the angular correlation kernel $\rho(\theta, \theta'; \boldsymbol{\psi}_i)$. In the following section, we propose a localization technique for CD and ID signals. In practice an intermediate situation might occur that corresponds to a partially correlated signal where the rays of signal which are arriving from different angles are partially correlated. Partially correlated signals can also be localized using the same method as proposed for the ID signal.

6.2. Localization

For the distributed source model of the preceding section a maximum likelihood (ML) estimator can be derived. However, ML localization is complex. In fact, if the dimension of the parameter vector for each source is given by m, a nonlinear search on a $q \times m$-dimensional space, where q is the number of signals, must be performed.

In this section, we propose a parametric localization technique which is based on *a priori* knowledge of the distribution pattern of the signals. It is assumed that the angular correlation kernel of the signal is chosen from a parametric class of functions. With this assumption, the localization of distributed sources is the same as a parameter estimation problem. The new algorithm trades off optimality and computational complexity. We use the concept of linear operators to generalize the MUSIC algorithm for the distributed source model.

6.2.1. A generalization of the MUSIC algorithm

Let us denote by $L^2[-\frac{\pi}{2}, \frac{\pi}{2}]$ the Hilbert space of all complex valued square integrable functions defined over the interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$. The inner product and the norm in this subspace are defined by

$$\langle s_i, s_j \rangle_c = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} s_i^*(\theta)\, s_j(\theta)\, d\theta \tag{6.12}$$

$$\|s_i\|_c = \sqrt{\langle s_i, s_i \rangle_c} \tag{6.13}$$

where the subscript c denotes a norm of the continuous waveform. According to (6.1), the observation vector $\mathbf{x}$ at the array output can be expressed as

$$\mathbf{x} = \sum_{i=1}^{q} \mathcal{L} s_i(.; \boldsymbol{\psi}_i) + \mathbf{n} \tag{6.14}$$

where $\mathcal{L}$ is a linear operator that maps $L^2[-\frac{\pi}{2}, \frac{\pi}{2}]$ into a p-dimensional complex observation vector space $\mathbb{C}^p$ according to

$$\mathcal{L} : L^2[-\tfrac{\pi}{2}, \tfrac{\pi}{2}] \longrightarrow \mathbb{C}^p \tag{6.15}$$

$$\mathcal{L}s = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathbf{a}(\theta)\, s(\theta)\, d\theta. \tag{6.16}$$

The inner product and the norm in $\mathbb{C}^p$ are defined by

$$\langle \mathbf{x}_i, \mathbf{x}_j \rangle_d = \mathbf{x}_i^H \mathbf{x}_j \tag{6.17}$$

$$\|\mathbf{x}_i\|_d = \sqrt{\langle \mathbf{x}_i, \mathbf{x}_i \rangle_d} \tag{6.18}$$

where the subscript d indicates a norm of discrete functions.

The adjoint operator $\mathcal{L}^+ : \mathbb{C}^p \rightarrow L^2[-\frac{\pi}{2}, \frac{\pi}{2}]$ satisfies

$$\langle \mathcal{L}s, \mathbf{x} \rangle_d = \langle s, \mathcal{L}^+ \mathbf{x} \rangle_c. \tag{6.19}$$

For the linear operator (6.16), we have

$$\begin{aligned}\langle \mathcal{L}s, \mathbf{x}\rangle_d &= [\mathcal{L}s]^H \mathbf{x} \\ &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} s^*(\theta)\, \mathbf{a}^H(\theta)\, d\theta\ \mathbf{x} \\ &= \langle s, \mathbf{a}^H \mathbf{x}\rangle_c. \end{aligned} \tag{6.20}$$

Thus the adjoint is given by

$$\mathcal{L}^+ \mathbf{x} = \mathbf{a}^H(\theta)\mathbf{x}. \tag{6.21}$$

As a starting point, we want to extend the definition of the signal and noise subspaces to distributed sources. Note that the source signal $s_i(\theta;\boldsymbol{\psi}_i)$ in (6.14) is a random signal which is also a function of the DOA θ and the parameter vector $\boldsymbol{\psi}_i$. By the *source subspace* we mean the span of all realizations of the source signals $s_i(\theta;\boldsymbol{\psi}_i),\ i = 1,\ldots,q$ where the $\boldsymbol{\psi}_i$'s are fixed. This subspace is shown by $\mathcal{S}$ and is defined as

$$\mathcal{S} = \text{Span}\{s_i(\theta;\boldsymbol{\psi}_i) : i = 1,\ldots,q, \text{ and all realizations}\}. \tag{6.22}$$

The source subspace $\mathcal{S}$ is a subspace of $L^2[-\frac{\pi}{2},\frac{\pi}{2}]$. The range of the linear operator $\mathcal{L}$ under $\mathcal{S}$ is the *signal subspace* and is represented by

$$\mathcal{R} = \{\mathcal{L}s \ :\ \text{all } s \in \mathcal{S}\}. \tag{6.23}$$

The orthogonal complement of $\mathcal{R}$ is the *noise subspace* and is denoted by $\mathcal{R}^\perp$. It can be shown that the range of the adjoint operator $\mathcal{L}^+$, when the domain is restricted to the noise subspace $\mathcal{R}^\perp$, is the orthogonal complement of $\mathcal{S}$ which is represented by $\mathcal{S}^\perp$. Fig. 6.1 schematically depicts the relationship between the linear operators and the subspaces.

The above concept of the signal and noise subspaces can be reconciled with the conventional definitions for the point source case as follows. The angular signal density of a point source at the DOA θ_i can be shown as

$$s_i(\theta) = \alpha_i \delta(\theta - \theta_i) \tag{6.24}$$

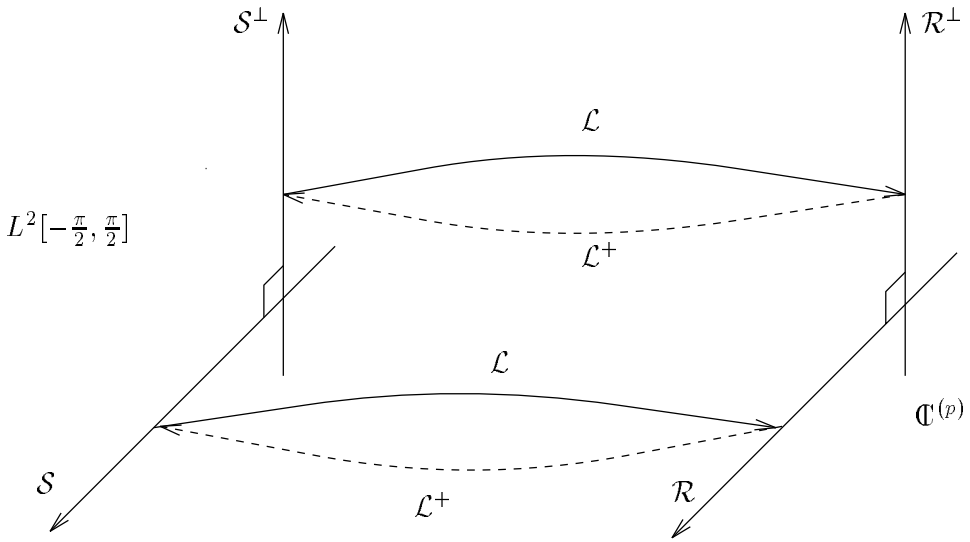

Fig. 6.1 Geometry of the linear operator and the adjoint operator.

where α_i is the complex envelope of the i-th signal. The source subspace for a point source scenario is then given by

$$\mathcal{S} = \text{Span}\{\delta(\theta - \theta_1), \ldots, \delta(\theta - \theta_q)\}. \tag{6.25}$$

Applying the linear operator $\mathcal{L}$ to (6.25) gives the signal subspace

$$\mathcal{R} = \text{Span}\{\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_q)\} \tag{6.26}$$

which corresponds to the conventional definition of the signal subspace for point sources.

We now use these concepts to interpret the MUSIC algorithm for the point sources given in (6.25). Suppose we know a basis for $\mathcal{R}^\perp$, say $\mathbf{e}_i \in \mathcal{R}^\perp,\ i = 1, \ldots, p-q$. Then

$$\mathcal{L}^+ \mathbf{e}_i = \mathbf{a}^H(\theta)\mathbf{e}_i \in \mathcal{S}^\perp, \qquad i = 1, \ldots, p-q. \tag{6.27}$$

Since the back-transformed vector is in the orthogonal complement of the source subspace, it is orthogonal to any vector in the source subspace,

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathbf{a}^H(\theta)\, \mathbf{e}_i s(\theta)\, d\theta = 0, \qquad \text{for any } s(\theta) \in \mathcal{S}, \ \text{ and } i = 1, \ldots, p-q. \tag{6.28}$$

Using (6.25) we have

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathbf{a}^H(\theta)\, \mathbf{e}_i\, \delta(\theta - \theta_j)\, d\theta = \mathbf{a}^H(\theta_j)\mathbf{e}_i = 0, \qquad \text{for } \ i = 1, \ldots, p-q, \quad j = 1, \ldots, q. \tag{6.29}$$

Defining $\mathbf{E}_n = [\mathbf{e}_1, \ldots, \mathbf{e}_{p-q}]$, we have

$$\mathbf{a}^H(\theta_j)\mathbf{E}_n = \mathbf{0}, \qquad \text{for } \ j = 1, \ldots, q. \tag{6.30}$$

The MUSIC algorithm estimates the DOAs of multiple point sources by maximizing the following "frequency detector" with respect to the DOA parameter $\theta \in \Theta$, where Θ is the

parameter set,

$$P_{\text{MUSIC}}(\theta) = \frac{1}{\mathbf{a}^H(\theta)\mathbf{E}_n\mathbf{E}_n^H\mathbf{a}(\theta)} \tag{6.31}$$

$$= \frac{1}{\|\mathbf{a}^H(\theta)\mathbf{E}_n\|^2}. \tag{6.32}$$

We use the same approach to derive a MUSIC type algorithm for distributed source parameter estimation. Suppose that $\mathcal{R}^\perp$ has dimension $p-q$ and we have a basis for $\mathcal{R}^\perp$, say $\mathbf{e}_1, \ldots, \mathbf{e}_{p-q}$, and let $\mathbf{E}_n = [\mathbf{e}_1, \ldots, \mathbf{e}_{p-q}]$. Since $\mathbf{e}_i$'s are in $\mathcal{R}^\perp$, their transformation under $\mathcal{L}^+$ will be in $\mathcal{S}^\perp$, i.e.

$$\mathcal{L}^+\mathbf{e}_i = \mathbf{a}^H(\theta)\mathbf{e}_i \in \mathcal{S}^\perp, \qquad i = 1, \ldots, p-q. \tag{6.33}$$

Thus for all $s(\theta) \in \mathcal{S}$ we have

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathbf{a}^H(\theta)\, \mathbf{E}_n\, s(\theta)\, d\theta = \mathbf{0}. \tag{6.34}$$

In (6.22) the source subspace $\mathcal{S}$ was defined as a span of the functions $s_i(\theta;\boldsymbol{\psi}_i)$. Hence (6.34) can be written as

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathbf{a}^H(\theta)\, \mathbf{E}_n\, s_i(\theta;\boldsymbol{\psi}_i)\, d\theta = \mathbf{0} \tag{6.35}$$

for all realizations of $s_i(\theta;\boldsymbol{\psi}_i)$, and for $i = 1, \ldots, q$. Since $s_i(\theta;\boldsymbol{\psi}_i)$ is a random function, this is equivalent to

$$E[\| \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathbf{a}^H(\theta)\, \mathbf{E}_n\, s_i(\theta;\boldsymbol{\psi}_i)\, d\theta\|^2] = 0, \qquad \text{for} \quad i = 1, \ldots, q. \tag{6.36}$$

Using (6.5) this equation can be expressed as

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathbf{a}^H(\theta)\, \mathbf{E}_n\, \rho(\theta,\theta';\boldsymbol{\psi}_i)\, \mathbf{E}_n^H\mathbf{a}(\theta')\, d\theta\, d\theta' = 0, \quad i = 1 \ldots, q. \tag{6.37}$$

Following the approach of the point source case leading to (6.32), we propose that the

parameter vector be estimated by locating the peaks of

$$\hat{\boldsymbol{\psi}} = \arg\max_{\boldsymbol{\psi}} \frac{1}{\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathbf{a}^H(\theta)\, \mathbf{E}_n\, \rho(\theta, \theta'; \boldsymbol{\psi})\, \mathbf{E}_n^H\, \mathbf{a}(\theta')\, d\theta d\theta'}. \tag{6.38}$$

We call this method the distributed signal parameter estimator (DSPE). The algorithm is implemented in two steps. First, the array is calibrated using signals with known parameter vectors from the same family of kernels and the results are stored. Then using these results, the spectrum of the DSPE algorithm is searched in an m-dimensional space for q largest local maxima. The maximum points correspond to the estimates of the parameter vector. The calibration step is due to imperfect knowledge of the sensors response. If the response of the sensors is known, the calibration step is not required.

6.2.2. The coherently distributed source localizer

The criterion (6.38) can be further simplified for coherently distributed sources. For CD sources, the rays of arriving waves at different angles are delayed and scaled version of the same signal. In such a case, the angular signal density can be represented as

$$s_i(\theta; \boldsymbol{\psi}_i) = \gamma_i g(\theta; \boldsymbol{\psi}_i) \tag{6.39}$$

where γ_i is a random variable and $g(\theta; \boldsymbol{\psi}_i)$ is a deterministic angular signal density. Let us define

$$\mathbf{b}(\boldsymbol{\psi}) = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathbf{a}(\theta) g(\theta; \boldsymbol{\psi}) d\theta, \tag{6.40}$$

and let $\mathbf{B}(\boldsymbol{\psi})$ be the matrix of the column vectors, $\mathbf{b}(\boldsymbol{\psi}_i)$, $i = 1, \ldots, q$. The correlation matrix of the array is then given by

$$\mathbf{R} = \mathbf{B}(\boldsymbol{\psi}) \boldsymbol{\Gamma} \mathbf{B}^H(\boldsymbol{\psi}) + \sigma^2 \mathbf{I} \tag{6.41}$$

where $\boldsymbol{\Gamma}$ is a correlation matrix with the ij-th component defined as $E[\gamma_i \gamma_j^*]$. If the sources are uncorrelated with each other, $\boldsymbol{\Gamma}$ will be diagonal. From (6.41) it is seen that

for CD signals the signal subspace is spanned by the eigenvectors of the correlation matrix corresponding to the q largest eigenvalues. The number of signals can be estimated to be the rank of the noise-free correlation matrix. The localization criterion (6.38) for the CD sources with the angular signal density $g(\theta;\boldsymbol{\psi})$ is given by

$$\hat{\boldsymbol{\psi}} = \arg\max_{\boldsymbol{\psi}} \frac{1}{\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} g^H(\theta;\boldsymbol{\psi})\, \mathbf{a}^H(\theta)\, \mathbf{E}_n\, \mathbf{E}_n^H\, \mathbf{a}(\theta')\, g(\theta';\boldsymbol{\psi})\, d\theta\, d\theta'} \tag{6.42}$$

which is found from using (6.8) in (6.38). With the definition of $\mathbf{b}(\boldsymbol{\psi}_i)$ the criterion (6.42) simplifies to

$$\hat{\boldsymbol{\psi}} = \arg\max_{\boldsymbol{\psi}} \frac{1}{\|\mathbf{b}^H(\boldsymbol{\psi})\mathbf{E}_n\|^2} \tag{6.43}$$

which is similar in form to (6.32). The difference is that the array manifold for the distributed source is an integral of the location vector $\mathbf{a}(\theta)$ weighted by the angular signal density $g(\theta;\boldsymbol{\psi})$. To instrument the estimation, the array is calibrated with the new array manifold which is shown by $\mathbf{b}(\boldsymbol{\psi})$ and the results are saved for later use. For localization, a search step is performed on an m-dimensional space to find the maxima of (6.42). These maxima are the estimates of the signal parameter vectors.

6.2.3. The incoherently distributed source localizer

For ID sources, the noise subspace is generally zero and (6.38) which is based on the signal and noise subspace decomposition cannot be directly applied. However, for a uniform distribution we show that with a proper choice of the signal and noise subspaces, it is possible to use the DSPE algorithm. By a uniform distribution we mean an angular correlation kernel which is flat over its extension width. The study of uniform extension gives insight into the ID source localization problem. Later we will generalize the method for nonuniform kernels.

Although for uniform ID signals the whole observation space is occupied by the signal components, most of the energy of the signals is concentrated in a few eigenvalues of the array correlation matrix. The number of these eigenvalues is denoted as the *effective* dimension of the signal subspace which can be used in the localization algorithm. In what

follows, we derive an analytic expression for the effective dimension of the signal subspace for a single ID source case. For correct localization, the number of the signal eigenvalues in the DSPE algorithm should be chosen equal to or greater than the effective dimension of the signal subspace.

Assume that an ID source with the uniform correlation kernel

$$\rho(\theta;\psi) = \begin{cases} \dfrac{1}{2\Delta} & \text{if } |\theta-\theta_0| \le \Delta \\ 0 & \text{otherwise} \end{cases} \tag{6.44}$$

is observed by a linear continuous array. In a continuous array, the signal is observed at every point z in the interval $[-\frac{L}{2},\frac{L}{2}]$ where L is the array length. Using (6.1), we can show that for a single ID source uniformly distributed in a noise-free environment, the spatial cross-correlation function is given by

$$E[x^*(z)x(z')] = e^{j\frac{2\pi}{\lambda}(z'-z)\sin\theta_0}\mathrm{sinc}[\frac{2}{\lambda}(z'-z)\Delta\cos\theta_0] \tag{6.45}$$

where $x(z)$ is the output vector of the array observed at the point z, and $\mathrm{sinc}(x) = \sin\pi x/(\pi x)$.

To find the effective dimension of the signal subspace we need to perform an eigenvalue analysis of (6.45) by solving

$$\int_{-\frac{L}{2}}^{\frac{L}{2}} e^{j\frac{2\pi}{\lambda}(z-z')\sin\theta_0}\mathrm{sinc}[\frac{2}{\lambda}(z-z')\Delta\cos\theta_0]\,\phi_n(z')\,dz' = \mu_n\phi_n(z). \tag{6.46}$$

The eigenfunctions $\phi_n(z)$ are the *angular prolate spheroidal functions* given by [32]

$$\phi_n(z) = S_{0n}(c,\frac{2}{L}z) \tag{6.47}$$

where c is a parameter defined as

$$c = \pi\Delta\frac{L}{\lambda}\cos\theta_0. \tag{6.48}$$

The eigenvalues μ_n are equal to

$$\mu_n = 2[R_{0n}^{(1)}(c,1)]^2 \tag{6.49}$$

where $R_{0n}^{(1)}(c,1),\ n = 0, 1, \ldots$, are the *radial prolate spheroidal functions* [32].

For a fixed c the radial prolate spheroidal function $R_{0n}^{(1)}(c,1)$ decreases exponentially with n. From the tables of the prolate spheroidal functions [33] it can be seen that more than 95 percent of the energy is concentrated in the first $\lceil c \rceil$ eigenvalues where $\lceil c \rceil$ indicates the smallest integer number larger than c. The effective dimension of the signal subspace is the number of significant eigenvalues in a noise-free environment, which we will take to be $\lceil c \rceil$. Once the signal subspace dimension is determined, the DSPE algorithm is used to localize the sources.

From the above discussion it is seen that the effective dimension of the signal subspace is directly related to the parameter c which is a function of the extension width, array length, signal wavelength, and the signal location in space. For a linear array with a half wavelength spacing, c becomes

$$c = \frac{\pi}{2}\Delta(p-1)\cos\theta_0. \tag{6.50}$$

In practice if c is underestimated, a large error in localization will be occurred. That is because some of the eigenvectors that contribute to the signal subspace are not used in the localization process. When c is overestimated, no great loss of performance is observed. In such a case the resolution capability is slightly reduced due to the importation of some noise components in the signal subspace.

For general nonuniform ID sources an analytic expression for the effective dimension of the signal subspace is not available. An eigenvalue analysis can be used to determine the number of significant signal components. With this approach, the DSPE algorithm can be applied to multiple source cases.

Using (6.10) the DSPE criterion for ID sources simplifies to

$$\hat{\boldsymbol{\psi}} = \arg\max_{\boldsymbol{\psi}} \frac{1}{\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathbf{a}^H(\theta)\, \mathbf{E}_n\, \rho(\theta;\boldsymbol{\psi})\, \mathbf{E}_n^H\, \mathbf{a}(\theta)\, d\theta}. \tag{6.51}$$

This criterion can also be represented as

$$\hat{\boldsymbol{\psi}} = \arg\max_{\boldsymbol{\psi}} \frac{1}{\mathrm{tr}(\mathbf{E}_n^H \mathbf{H}(\boldsymbol{\psi}) \mathbf{E}_n)} \tag{6.52}$$

where tr(.) stands for the trace of a matrix and $\mathbf{H}(\boldsymbol{\psi})$ is defined by

$$\mathbf{H}(\boldsymbol{\psi}) = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathbf{a}(\theta)\, \rho(\theta;\boldsymbol{\psi})\, \mathbf{a}^H(\theta)\, d\theta. \tag{6.53}$$

Note that $\mathbf{H}(\boldsymbol{\psi})$ is independent of the observation and can be computed and stored in the calibration process. For a uniform linear array, $\mathbf{H}(\boldsymbol{\psi})$ has a Hermitian Toeplitz form. In such a case, only p complex numbers need be computed for each parameter vector $\boldsymbol{\psi}$.

For partially correlated distributed signals, the effective signal subspace dimension is between q and $\lceil c \rceil$. Using the number of dominant eigenvalues as the effective dimension of the signal subspace, the DSPE algorithm can be used for partially correlated distribution.

6.3. The array gain

Beamformers improve the array output SNR by steering a beam towards the direction of signal. Because of the ease of implementation, these methods are practically important. However, they have relatively low resolution. In a conventional beamformer, to achieve a higher resolution, a large number of sensors must be used. For point sources the array gain can be improved by increasing the number of sensors. Here, we show that for distributed sources the spatial correlation function of the signal is upper bounded by an exponentially decreasing function. Then, we derive the array gain and show that it is bounded and cannot be linearly increased with the number of sensors. For the specific case of CD sources, we show that the array gain attains a maximum and decreases exponentially for

very large number of sensors.

The gain of an array of sensors is defined as the ratio of the SNR at the array output to the SNR at a single sensor [20]. Assuming that the noise is spatially white and that a conventional beamformer is used, the array gain is given by

$$G_a = \frac{\mathbf{a}^H \mathbf{R}_s \mathbf{a}}{\mathbf{a}^H \mathbf{a}} \tag{6.54}$$

where $\mathbf{a}$ is the location vector of the array steered towards the direction of interest and $\mathbf{R}_s$ is the correlation matrix of the array output in a noise-free environment.

6.3.1. Coherently distributed sources

Assume that the array output can be observed along a continuous linear array. If the observation at point z is shown by $x(z)$, for a CD source in a noise-free environment we have

$$x(z) = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} e^{j\frac{2\pi z}{\lambda}\sin\theta} \gamma g(\theta;\boldsymbol{\psi}) d\theta \tag{6.55}$$

where γ is a zero-mean complex Gaussian random variable and $g(\theta;\boldsymbol{\psi})$ is the normalized deterministic angular signal density. Assuming that the source is uniformly distributed by

$$g(\theta;\boldsymbol{\psi}) = \begin{cases} \dfrac{1}{2\Delta} & \text{if } |\theta - \theta_0| \le \Delta \\ 0 & \text{otherwise} \end{cases} \tag{6.56}$$

the observation vector can be written as

$$x(z) = \gamma \frac{1}{2\Delta} \int_{\theta_0-\Delta}^{\theta_0+\Delta} e^{j\frac{2\pi z}{\lambda}\sin\theta} d\theta. \tag{6.57}$$

For a small Δ, it is straightforward to show that

$$x(z) \approx \gamma e^{j\frac{2\pi z}{\lambda}\sin\theta_0} \mathrm{sinc}(\frac{2z}{\lambda}\Delta\cos\theta_0). \tag{6.58}$$

From (6.58) we arrive at the following result.

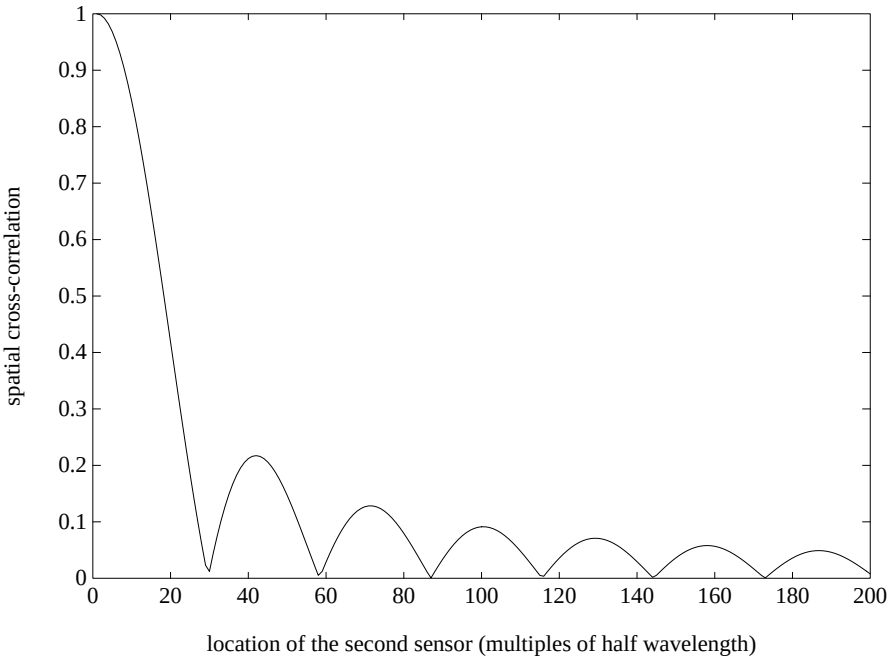

Fig. 6.2 The spatial cross-correlation for a uniform CD source. (The first sensor is placed at the phase reference point.)

Property 6.1. *For a uniform CD source (with small extension) the spatial cross-correlation function at z_1 and z_2 in a noise-free environment is bounded by*

$$|E[x(z_1)x^*(z_2)]| \leq K z_1^{-1} z_2^{-1} \tag{6.59}$$

where K is a positive scalar.

An example of the correlation between two points on a linear array for a uniform CD source is depicted in Fig. 6.2. It is assumed that the first point is the phase reference of the array. The second point varies along the array. The envelope of the correlation function exponentially decreases with distance. Thus, as the aperture length of the array increases, the correlation between far end sensors decreases. In other words, the signals at widely separated sensors cannot be coherently added to increase the SNR. This suggests that the array gain does not increase linearly with the number of sensors.

For a uniform linear array with half the wavelength spacing between sensors, the component of the observation (6.58) at the position of the l-th sensor is

$$x_l = \gamma e^{j\pi l \sin\theta_0} \mathrm{sinc}(l\Delta \cos\theta_0). \tag{6.60}$$

Assuming that the power of the source is unity and $\theta_0 = 0$, the array gain is given by

$$G_a = \frac{1}{p}\Big[\sum_{l=0}^{p-1} \mathrm{sinc}(l\Delta)\Big]^2. \tag{6.61}$$

Note that for $\Delta = 0$ the array gain is equal to p which is the gain of a point source scenario. For $\Delta \neq 0$ and large p, the sum in (6.61) is approximated by $\pi/2$ which reveals that the array gain decreases with a rate of $1/p$. The array gain for a CD source as a function of the number of sensors p is illustrated in Fig. 6.3. The array gain has a maximum which depends on the extension width. Increasing the number of the sensors beyond the maximum point decreases the array gain. We have found that at the maximum point the array length p_{MAX} can be approximated as

$$p_{\mathrm{MAX}} \approx \frac{40}{\Delta^\circ} \tag{6.62}$$

where Δ° is the extension width measured in degrees.

6.3.2. Incoherently distributed sources

From (6.45) we can easily arrive at the following result.

Property 6.2. *For an ID source uniformly distributed with a small extension width, the spatial correlation function decreases exponentially with distance and is upper bounded by*

$$|E[x(z_1)x^*(z_2)]| \leq \frac{K}{z_1 - z_2} \tag{6.63}$$

where K is a positive scalar.

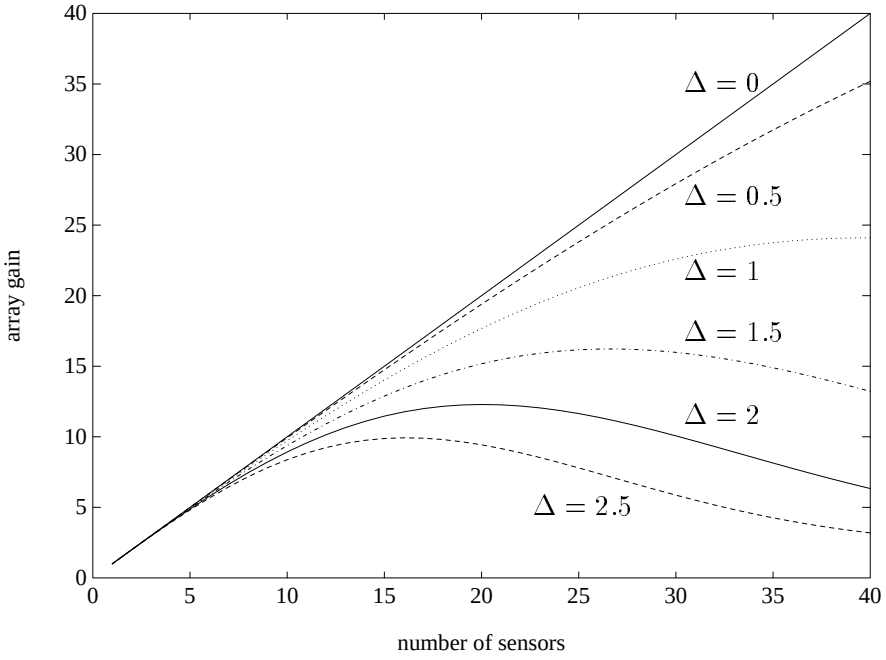

Fig. 6.3 The array gain for a uniform CD source for different extension widths Δ in degrees.

Since the spatial correlation function decreases exponentially with distance, the array gain cannot increase linearly with the number of sensors. For a uniform linear array with half the wavelength spacing between sensors, the spatial cross-correlation function between the l-th and the k-th sensors is given by

$$E[x_l x_k^*] = e^{j\pi(l-k)\sin\theta_0}\mathrm{sinc}[(l-k)\Delta\cos\theta_0]. \tag{6.64}$$

Assuming that $\theta_0 = 0$, the array gain is given by

$$G_a = \frac{1}{p}\Big[\sum_{l=0}^{p-1}\sum_{k=0}^{p-1}\mathrm{sinc}[(l-k)\Delta]\Big]. \tag{6.65}$$

Again it is seen that for $\Delta = 0$ we get the same result as a point source case. With a change of variable the array gain can be represented as

$$G_a = \frac{1}{p}\Big[p + 2\sum_{r=1}^{p-1}(p-r)\mathrm{sinc}(r\Delta)\Big]. \tag{6.66}$$

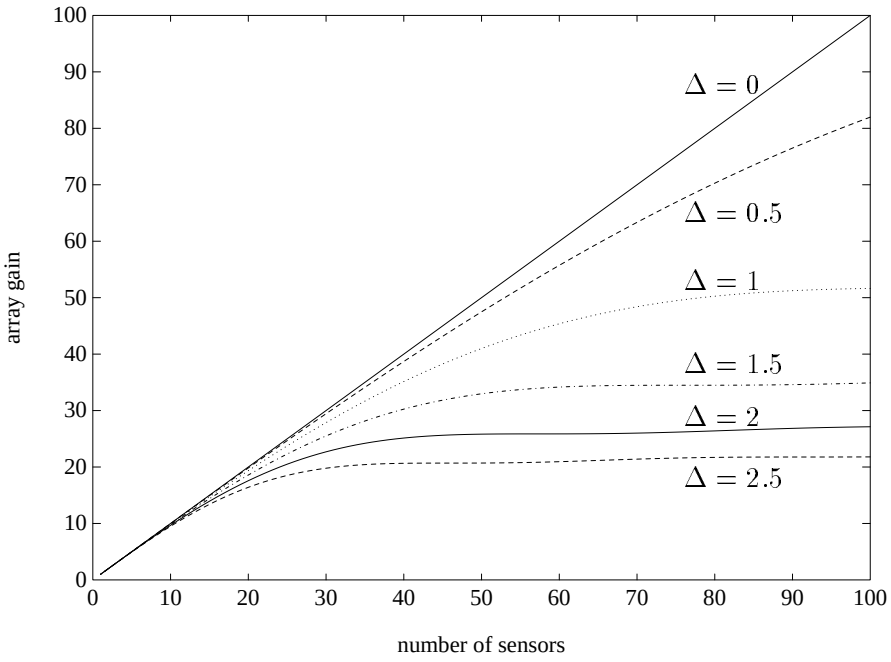

Fig. 6.4 The array gain for a uniform ID source for different extension widths Δ in degrees.

The array gain for an ID source is depicted in Fig. 6.4. For a fixed extension width, the maximum array gain for the ID source is higher than for the CD source.

6.4. The uniqueness problem

A natural question that might be raised is: "Using the observations at the output of an array, are the distributed sources uniquely localizable?" Equation (6.1) shows that the output of an array exposed to a distributed source in a noise-free environment is an integral of the location vector weighted by the angular signal density. Using this interpretation the uniqueness is stated as: For a given array output in a noise-free environment, a unique angular signal density can be found that generates that observation.

The array output for a distributed source can be approximated by the superposition of the array response to a large number of closely spaced point sources. The approximation error decreases by increasing the number of point sources and/or decreasing their spacing. In the limit a distributed source corresponds to an infinite number of point sources. In this approximation the number of free parameters increases with increasing the number

of point sources. If a classical point source localization method, such as ordinary MUSIC, is applied to this scenario, a unique solution may not be attained due to a limited number of sensors. For a unique solution, the number of point sources must be smaller than the number of sensors.

Note that the correlation matrix of these point sources is related to the kernel of the distributed source. If this kernel belongs to a parametric family, the number of free parameters will usually be much smaller than the number of point sources. In this section, we show that if the point sources are related through some parametric constraints, then for any observation at the array output, almost surely there is a unique solution which generates that output. From this, we conclude that everywhere except in a set with probability zero, a unique solution can be found for the localization problem of an extended source. We use the concept of the "topological dimension" of a set. The topological dimension of a set is the number of free (real) parameters that are required to describe all the elements of that set. For a detailed discussion of the topological dimension see [18]. The same concept has been used by Wax [54] to find the number of resolvable sources when the source signals are constrained to some real or complex loci.

We will discuss the uniqueness problems separately for the CD and ID sources. For each case a legitimate set is found which contains all the signals that are chosen from the parametric class of the angular correlation kernel. Every element in the legitimate set can be a candidate for the localization problem. The ambiguity set is a subset of the legitimate set that contains all the signals that can generate nonunique solutions for the localization problem. The objective here is to find the conditions under which the ambiguity set has a smaller dimension than the legitimate set. In such a case, it can be shown that the probability of the ambiguity set is equal to zero.

6.4.1. Coherently distributed sources

Let the interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$ be quantized into a grid of $\tilde{q}$ points. It is assumed that a distributed source with the angular signal density $g(\theta; \boldsymbol{\psi})$, where $\boldsymbol{\psi}$ is an m-dimensional parameter vector, is discretized so that it can take values on the quantized grid. Initially, we assume

a single source scenario.

The output of an array of p sensors in a noise-free environment for N snapshots can be represented as

$$\mathbf{X} = \mathbf{A}\mathbf{S}(\boldsymbol{\psi}) \tag{6.67}$$

where $\mathbf{A}$ is the $p \times \tilde{q}$ location matrix of the array, $\mathbf{S}(\boldsymbol{\psi})$ is the $\tilde{q} \times N$ source signal matrix, and $\mathbf{X}$ is the $p \times N$ observation matrix. Since the source is coherently distributed in space, the rank of the matrix $\mathbf{S}(\boldsymbol{\psi})$ is equal to one. The source signal matrix can be expressed as $\mathbf{S}(\boldsymbol{\psi}) = [\mathbf{s}_1(\boldsymbol{\psi})\ \mathbf{S}_2(\boldsymbol{\psi})]$ where $\mathbf{s}_1(\boldsymbol{\psi})$ is a $\tilde{q} \times 1$ vector and $\mathbf{S}_2(\boldsymbol{\psi})$ is a $\tilde{q} \times (N-1)$ matrix. Similar to [54] we can show that it suffices to solve the uniqueness problem only for

$$\mathbf{x}_1 = \mathbf{A}\mathbf{s}_1(\boldsymbol{\psi}). \tag{6.68}$$

The source signal matrix $\mathbf{s}_1(\boldsymbol{\psi})$ can be shown as

$$\mathbf{s}_1(\boldsymbol{\psi}) = \gamma_1 \mathbf{g}(\boldsymbol{\psi}) \tag{6.69}$$

where $\mathbf{g}(\boldsymbol{\psi})$ is a $\tilde{q} \times 1$ vector with its i-th component equal to the value of $g(\theta; \boldsymbol{\psi})$ computed at the location of the i-th quantized DOA, and γ is a complex level scale. The vectors that satisfy (6.69) for all $\boldsymbol{\psi}$, generate a set which is called the "legitimate set" and is denoted by $\mathcal{G}$. Since it is assumed that there is a one-to-one relationship between $\mathbf{g}(\boldsymbol{\psi})$ and $\boldsymbol{\psi}$, to determine $\mathbf{s}_1(\boldsymbol{\psi})$ we need to determine $2+m$ real parameters. Thus the dimensionality of $\mathcal{G}$ is equal to $2+m$.

A nonunique solution for the localization problem can be found if

$$\mathbf{x}_1 = \mathbf{A}\mathbf{s}_1(\boldsymbol{\psi}) = \mathbf{A}\mathbf{s}_1'(\boldsymbol{\psi}') \tag{6.70}$$

which also be written as

$$\mathbf{A}(\gamma_1 \mathbf{g}(\boldsymbol{\psi}) - \gamma_1' \mathbf{g}(\boldsymbol{\psi}')) = \mathbf{0}. \tag{6.71}$$

The legitimate vectors that satisfy this equality for any $\boldsymbol{\psi}$ and $\boldsymbol{\psi}'$, form the "ambiguity set"

which is represented by $\mathcal{D}$. To represent each vector in the form of $(\gamma_1 \mathbf{g}(\boldsymbol{\psi}) - \gamma_1' \mathbf{g}(\boldsymbol{\psi}'))$ we need to determine $2(m+2)$ real parameters. However, (6.71) shows that for the vectors in the ambiguity set $2p$ constraints should be applied to their parameters. Thus, total number of parameters that can freely be set to satisfy (6.71) is equal to $2(m+2) - 2p$. This is the dimensionality of $\mathcal{D}$. Since $\mathbf{s}_1(\boldsymbol{\psi})$ is a random vector, a unique solution for the localization problem can be found with probability one if

$$\dim \mathcal{D} < \dim \mathcal{G} \tag{6.72}$$

where "dim" is the dimension operator. This criterion is equal to

$$m < 2p - 2. \tag{6.73}$$

As far as (6.73) is satisfied, with probability one a unique solution for the localization problem can be found. Note that (6.73) is independent of $\tilde{q}$ the number of quantized sources. Thus infinite number of point source (an extended source) are localizable if they are related through some parametric constraints.

A multi-source case can be treated similarly with an angular signal density equal to the addition of the angular signal density of the single sources. For a multi-source scenario the dimensionalities of $\mathcal{G}$ and $\mathcal{D}$ are equal to $q(m+2)$ and $q(m+2) - 2p$, respectively, where q is the number of coherent sources. Thus the uniqueness constraint implies that

$$q < \frac{2p}{m+2} \tag{6.74}$$

Note that for point source case $m = 1$ and we have $q < \frac{2p}{3}$, which is the well known sufficient condition for unique localization of coherent point sources [58].

6.4.2. Incoherently distributed sources

The sample correlation matrix is a Hermitian random matrix with jointly Wishart distributed elements [3]. The true correlation matrix is the limit of the sample correlation

matrix when the observation time tends to infinity. In the sequel, we find the dimensionality of the true correlation matrix, keeping in mind that it is the limit of a random matrix. The error between the true and the sample correlation matrices can be arbitrarily reduced by increasing the observation time. We consider a subset of the sample correlation matrices which generate ambiguous solutions for the localization problem. Then we show that this set converges into a set that has a smaller dimension than the set of all possible correlation matrices.

Assume that the interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$ is uniformly sampled into a grid of $\tilde{q}$ points. A distributed source with the angular correlation kernel $\rho(\theta, \theta'; \boldsymbol{\psi})$, where $\boldsymbol{\psi}$ is an m-dimensional parameter vector, takes its values on this grid in a noise-free environment. Again initially assume a single source in a noise-free environment. The correlation matrix of the array output is shown as

$$\mathbf{R}_x = \mathbf{A}(\theta)\mathbf{R}_s\mathbf{A}^H(\theta) \tag{6.75}$$

where $\mathbf{A}(\theta)$ is the $p \times \tilde{q}$ dimensional location matrix of the array and $\mathbf{R}_s$ is the $\tilde{q} \times \tilde{q}$ correlation matrix of the point sources. Since the point sources are the samples of the distributed sources, their cross-correlation matrix satisfies

$$\mathbf{R}_s = \mathbf{P}(\boldsymbol{\psi}), \qquad \text{for some } \boldsymbol{\psi} \in \boldsymbol{\Psi} \tag{6.76}$$

where $\boldsymbol{\Psi}$ is the parameter set and the components of $\mathbf{P}(\boldsymbol{\psi})$ are the values of the angular correlation kernel of the distributed source $\rho(\theta, \theta'; \boldsymbol{\psi})$ computed on the grid. All the correlation matrices $\mathbf{R}_x$ that satisfy (6.75) with the constraint (6.76) form the "legitimate set" $\mathcal{G}$. Since $\mathbf{P}(\boldsymbol{\psi})$ is a function of m free (real) parameters, the topological dimension of the legitimate set is m.

Let us define

$$\mathbf{F} = \mathbf{A}(\theta)\mathbf{P}(\boldsymbol{\psi})\mathbf{A}^H(\theta) - \mathbf{A}(\theta)\mathbf{P}(\boldsymbol{\psi}')\mathbf{A}^H(\theta) \tag{6.77}$$

for some $\boldsymbol{\psi}$ and $\boldsymbol{\psi}'$. The set of all matrices which can be represented by (6.77) has a

dimensionality $2m$. A nonunique solution for the DOA estimation problem is found if

$$\mathbf{F} = \mathbf{0}. \tag{6.78}$$

Note that for ID sources, $\mathbf{P}(\boldsymbol{\psi})$ is a diagonal matrix. Thus (6.78) provides p^2 complex constraints but only p of them are independent. The number of parameters that can be chosen freely to satisfy (6.78) is equal to $2m - p$. Define the ambiguity set as

$$\mathcal{D} = \{\mathbf{R}_x : \mathbf{A}(\theta)\mathbf{P}(\boldsymbol{\psi})\mathbf{A}^H(\theta) = \mathbf{A}(\theta)\mathbf{P}(\boldsymbol{\psi}')\mathbf{A}^H(\theta)\}. \tag{6.79}$$

The elements of $\mathcal{D}$ produce nonunique solutions for the DOA estimator. The topological dimension of $\mathcal{D}$ is equal to $2m - p$. With probability one, a unique solution can be found for the localization problem if

$$2m - p < m, \quad \text{or} \quad m < p. \tag{6.80}$$

This suggests that distributed sources are uniquely resolvable if they are chosen from a parametric class of angular correlation kernels with the dimension of the parameter vector smaller than the number of sensors.

A multi-source case can be treated similarly with an angular correlation kernel equal to the addition of the angular correlation kernels of the single sources. In a multi-source case with q uncorrelated sources the dimensionality of $\mathcal{G}$ and $\mathcal{D}$ are equal to qm and $(2qm - p)$, respectively. The uniqueness criterion is then given by

$$qm < p. \tag{6.81}$$

For point sources $m = 1$, and we have $q < p$, which is the well known sufficient condition for unique localization of uncorrelated sources [58].

6.5. Simulation results and performance comparison

A) CD sources

We investigate a configuration with two equipower uncorrelated narrowband CD sources arriving at a uniform linear array of 20 sensors. The spacing between adjacent sensors is equal to half the wavelength at the operating frequency. The angular signal density of the i-th source is given by

$$g(\theta;\boldsymbol{\psi}_i) = \frac{K_i}{1 + j\frac{\theta-\theta_i}{\Delta_i}} \qquad i = 1,2 \tag{6.82}$$

where K_i is a normalization factor, θ_i is the central angle of arrival, and Δ_i is the -3 dB extension width. The parameter vector in this example is $\boldsymbol{\psi}_i = (\theta_i, \Delta_i)$. The angular correlation kernel for such a signal density is given by

$$\rho(\theta;\boldsymbol{\psi}_i) = \frac{K^2}{1 + (\frac{\theta-\theta_i}{\Delta_i})^2} \tag{6.83}$$

which has a Butterworth spectrum and is proposed in [14, chap 3] as a model for noise sources. In our simulations, θ_1 and θ_2 are taken as 10 and 13 degrees with the extension widths $\Delta_1 = 1$ and $\Delta_2 = 2$ degrees. It is seen that the sources have a significant overlap in space.

A Monte-Carlo simulation of 50 independent runs with 50 snapshots for each trial was performed for different signal-to-noise ratios. The resolution performances of the conventional MUSIC and the DSPE are compared in Fig. 6.5. For the conventional MUSIC algorithm, the two signals are considered to be resolved when two peaks are observed in the MUSIC spectrum. For the DSPE algorithm each source is considered detected if the estimates of θ_i and the distribution widths Δ_i are within 1 degree of the true values. Note that these definitions of detection tends to favor the MUSIC algorithm more than the DSPE estimator. The resolution threshold of the DSPE is about 15 dB lower than the conventional MUSIC algorithm.

For this source configuration we found the bias and the standard deviation of the MUSIC and the DSPE estimators. For both sources the estimated central DOA is biased

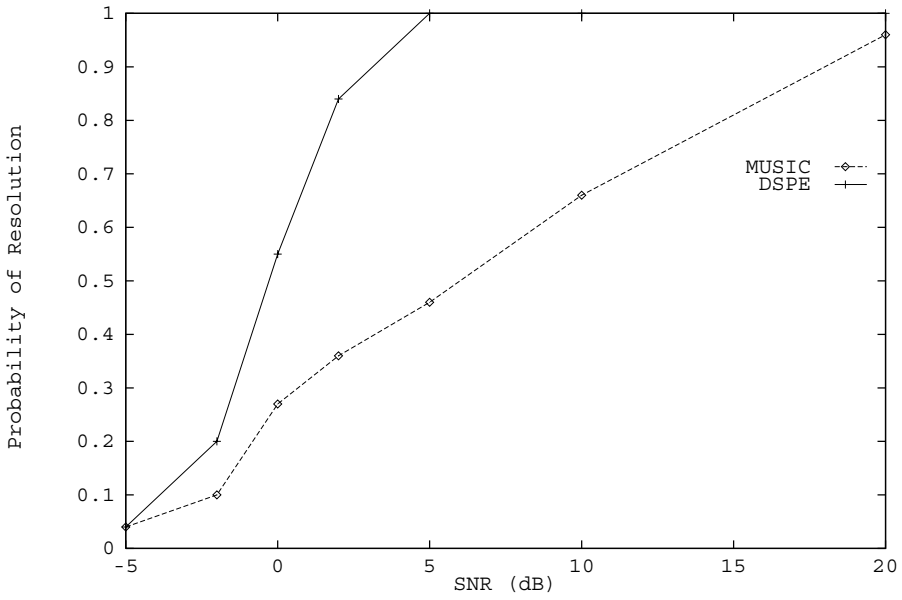

Fig. 6.5 The probability of resolution for the conventional MUSIC and the DSPE versus SNR.

in the conventional MUSIC algorithm and the bias cannot be decreased by increasing the SNR (see Fig. 6.6 and Fig. 6.7.) The DSPE algorithm provides a smaller bias in the DOA estimation. Furthermore, the bias can be reduced by increasing the signal-to-noise ratio. Fig. 6.8 and Fig. 6.9 show that the standard deviation of the DSPE estimators is less than that for the MUSIC algorithm.

B) ID sources

For the ID signal scenario we examine a configuration with two uniformly distributed sources with the angular correlation kernel

$$\rho(\theta;\psi_i) = \begin{cases} \dfrac{1}{2\Delta_i} & \text{if } |\theta-\theta_i| \le \Delta_i, \quad i=1,2 \\ 0 & \text{otherwise} \end{cases} \tag{6.84}$$

arriving at an array of 20 sensors. In our simulation, the central DOAs are selected as $\theta_1 = 8$ and $\theta_2 = 15$ degrees with extension widths $\Delta_1 = 1$ and $\Delta_2 = 1.5$ degrees, respectively. The signal-to-noise ratio is 30 dB and 200 snapshots are observed. For a

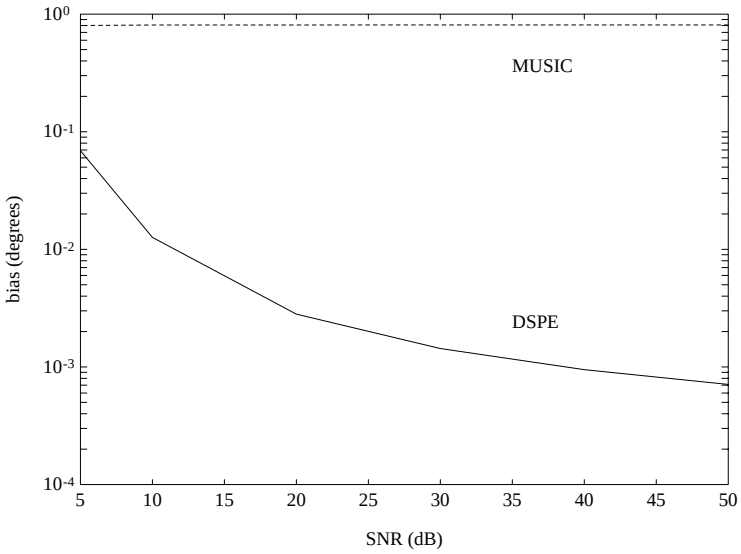

Fig. 6.6 The bias of estimation versus SNR for the source at 10 degrees.

single source with 3 degree extension width, the parameter c is smaller than 1.6. The eigenvalues of the sample correlation matrix for this scenario are shown in Fig. 6.10. It is seen that the first 4 eigenvalues dominate. This number agrees with the value given by the analytical study of the effective dimension of the signal subspace. The DSPE algorithm was run for this example with 16 noise eigenvectors. The DSPE spectrum is illustrated in Fig. 6.11. The two prominent peaks estimate the central DOAs at 7.92 and 15.04 degrees with extension widths 1.86 and 3.10 degrees, respectively. Note that when $\Delta = 0$ the DSPE algorithm coincides with the MUSIC algorithm. The MUSIC spectrum is the $\Delta = 0$ case in Fig. 6.11.

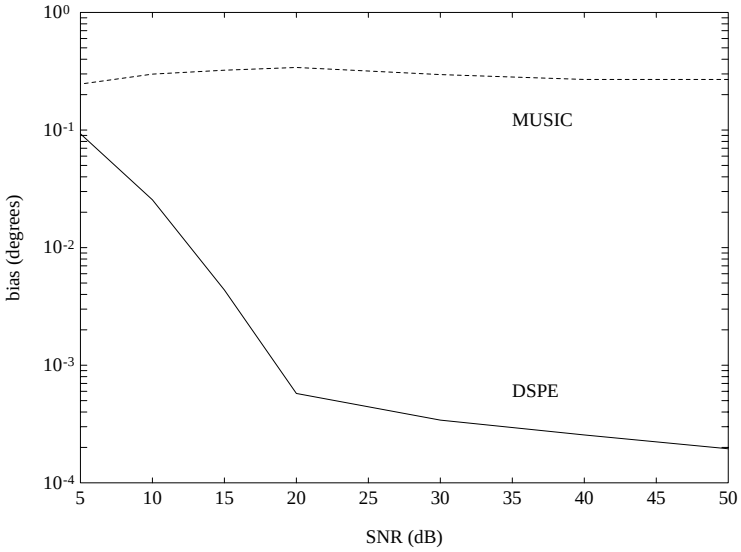

Fig. 6.7 The bias of estimation versus SNR for the source at 13 degrees.

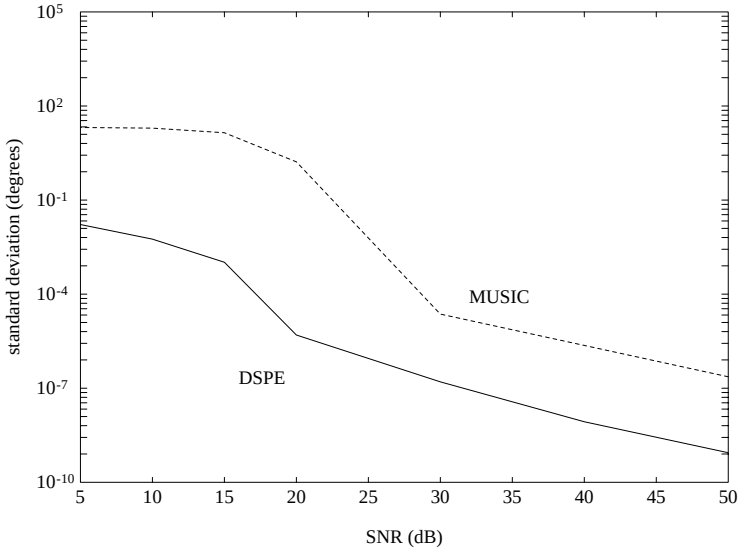

Fig. 6.8 The standard deviation versus SNR for the source at 10 degrees.

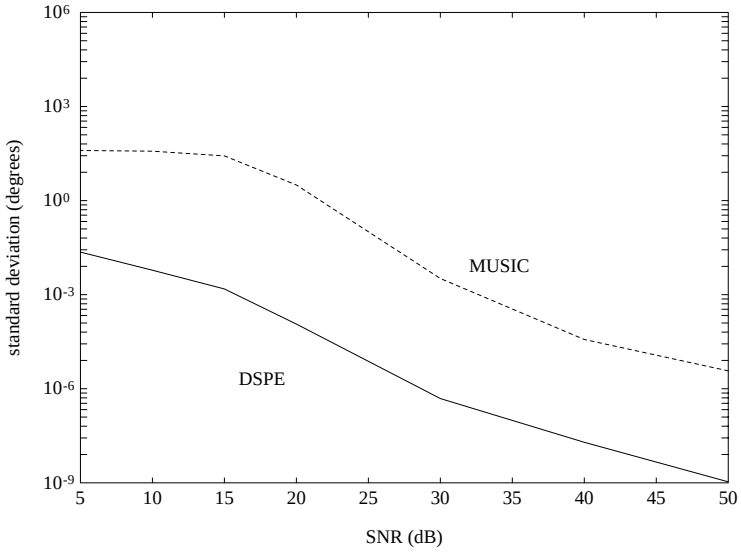

Fig. 6.9 The standard deviation versus SNR for the source at 13 degrees.

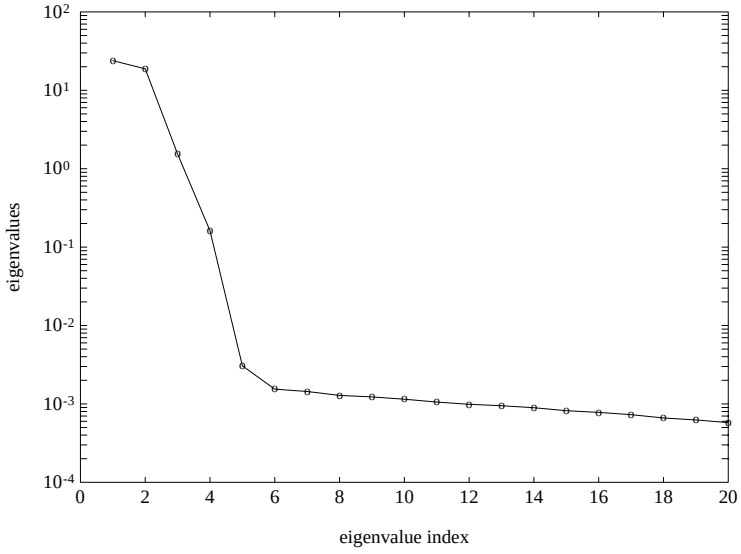

Fig. 6.10 The eigenvalues of a configuration with two SD signals.

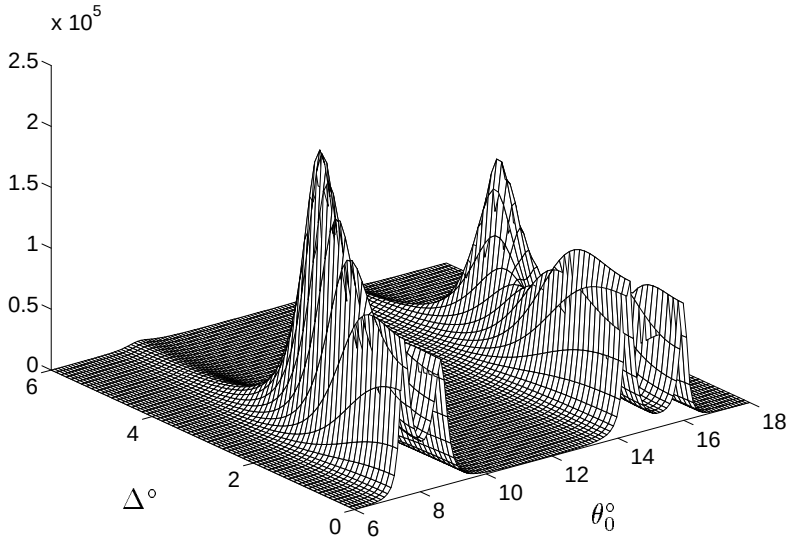

Fig. 6.11 The spectrum of the DSPE algorithm.

Chapter 7

Summary and Conclusions

This dissertation concerns the detection and localization of spatiotemporal signals. A spatiotemporal signal is function of time and space. An spatiotemporal signal can be Fourier transformed into the frequency-wavenumber domain. We have considered three special cases of interest: narrowband, wideband and spatially distributed signals.

In Chapter 3, we have introduced a new information theoretic method for detection of the number of narrowband point signals. The approach is based on the predictive stochastic complexity (PSC). The PSC is the length of a predictive code that encodes the data. We have derived two versions of the PSC algorithm. The first method is based on the maximum likelihood (ML) estimate of the correlation matrix. Since the only information used in this approach is multiplicity of the smallest eigenvalue of the correlation matrix, it can only detect noncoherent sources. In the second approach, we have exploited the structure of the correlation matrix and decomposed it into two parts in the signal and noise subspaces. The ML estimation of the correlation matrix in this method is based on the ML estimation of the direction-of-arrival (DOA). We have shown that this method detects coherent as well as noncoherent sources.

To reduce the computational complexity in the ML estimation of the parameter vector, we used a hill climbing method. The simulation results show that the performance of the PSC algorithm is better or comparable to the MDL method. It has also been shown that the new algorithm has a smaller sensitivity for the uncertainties in the DOA estimates.

Both MDL and PSC have been developed for large data size.

In Chapter 4, a method to determine the optimal focusing frequency for the coherent signal-subspace method (CSM) with unitary transformations is proposed. We have defined a criterion based on the subspace fitting error and optimized a tight upper bound to it. The optimization was done in two steps. First, the singular values of the focusing location matrix are obtained. Then a one-variable nonlinear minimization problem is solved to obtain the focusing frequency. The simulation results show that the method successfully finds the global optimum value and improves the performance of the estimation by reducing the bias and the resolution signal-to-noise ratio (SNR) threshold.

In Chapter 5, we have introduced a new method for localization of broadband signals. The method is based on the two-sided unitary transformation of the correlation matrices. The motivation for this work was to reduce the error of the subspace fitting and to remove the asymptotic bias of estimation that is involved in the CSM algorithm. The bias of estimation in the CSM is a function of the focusing points and the bandwidth of processing. We have shown that the noise-free universal focused sample correlation matrix has nonzero eigenvalues in the noise subspace. This nonuniform expansion of the source energy into the noise subspace acts as a colored noise with an unknown correlation matrix. Since the form of the signal extension is unknown, the estimation will be biased in general.

It has been shown that the TCT algorithm does not suffer from this shortcoming. For a scenario with q noncoherent sources, the noise-free universal focused correlation matrix of the TCT algorithm has q nonzero eigenvalues and its columns span a q-dimensional subspace regardless of the processing bandwidth. With iterative use of the TCT algorithm it is possible to get this q-dimensional subspace to coincide with the true signal subspace; hence resulting in unbiased estimation. We have also shown that the generalized variance of the TCT algorithm is smaller than its counterpart for the CSM. Thus, in a noisy situation, TCT can generate more robust estimates of the DOAs.

In Chapter 6, we have discussed the problem of localizing spatially distributed sources. It has been assumed that the angular cross-correlation kernel of the source signals belongs to a parametric class. We proved that this assumption guarantees a unique solution for the

localization problem. We have proposed a MUSIC type distributed signal parameter estimator (DSPE) which is based on minimizing a norm of the transformed noise eigenvectors in the source subspace.

The method was applied to two cases. First, we used the DSPE algorithm to localize coherently distributed (CD) signals. For the coherent distribution of signals it has been shown that the new method is similar to the MUSIC algorithm with an array manifold which is the integral of the location vector weighted with the angular signal density. We have also considered incoherently distributed (ID) signals. For these signals, it has been shown that the effective dimension of the signal subspace is a function of the product of the extension width, the array aperture, wavelength, and the signal location. The DSPE algorithm is applied to ID sources using the effective dimension of the signal subspace.

We have found the spatial cross-correlation of the coherently and incoherently distributed signals and shown that in both cases the spatial correlation decreases exponentially with distance. For uniformly distributed CD and ID signals, the array gain is bounded and cannot be increased linearly with the number of sensors. In both cases the maximum array gain changes with the extension width; increasing the extension width decreases the array gain.

Computer simulations were conducted to compare the new method and the conventional MUSIC algorithm. It was shown that the resolution threshold for the new method is lower than that for the MUSIC algorithm. The DSPE algorithm has a smaller bias, and unlike the MUSIC estimator, the bias can be reduced by increasing the SNR. It was also shown that the DSPE method has a smaller standard deviation than the MUSIC algorithm.

7.1. Future work

An extension of the work of Chapter 3 can be to compare the sensitivity of the MDl and PSC methods as a function of the number of observations. Although it was proposed that the PSC algorithm can be applied to time varying systems, a computer simulation for time varying environments have not been conducted. We have demonstrated that the PSC can

detect a change in the number of signals faster than the MDL method. For the future, the capability of the PSC method for detecting sources in a nonstationary situation needs to be studied.

We have explained that the DSPE algorithm needs a calibration process. The results of calibration are saved to be used for the search process. Note that the calibration should be implemented for an m-dimensional set where m is the dimension of the parameter vector of the angular cross-correlation kernel. This makes DSPE computationally more complex than the MUSIC algorithm.

Future work for the results of Chapter 6 can be in the direction of avoiding the calibration process by using an ESPRIT-type method. In the ESPRIT algorithm, it is assumed that the signal wave field is sampled with an array of perfectly matched doublets. For a distributed source, we have shown that the spatial correlation function is exponentially decreasing with distance. This fact should be considered in deriving an ESPRIT-type algorithm for the distributed source case.

More works is warranted on the DSPE method for localizing extended sources. We have started in this direction by calculating the Cramer-Rao bound for the estimation of extended sources and in particular coherent sources. The performance of DSPE relative to this bound needs to be evaluated.

Appendix A

Uniqueness of the Subspace Decomposition

In this appendix we prove Theorem 3.1. The objective is to show that for every signal subspace $\mathbb{C}_s^q = \text{span}\{\mathbf{a}(\theta_i)|\ i = 1, \ldots, q\}$ there is a unique set $\Theta = \{\theta_i|\ i = 1, \ldots, q\}$ that generates that space and vice versa. Note that the location vectors $\mathbf{a}(\theta_i) \in \mathcal{A}$ are independent and form a basis for the signal subspace. We claim that this basis is unique. If it is not the case, there will be another set $\Psi = \{\psi_i|\ i = 1, \ldots, q\}$ with $\mathbf{a}(\psi_i) \in \mathcal{A}$ which is also a basis for $\mathbb{C}_s^q$. Then every vector in $\mathbb{C}_s^q$ could be represented with respect to this basis. Take for instance $\mathbf{a}(\theta_1)$. The independence of the location vectors suggests that θ_1 has to be in the set Ψ. The continuation of this argument implies $\Psi = \Theta$.

The inverse is straightforward. If $\Psi = \Theta$, then $\mathbf{A}(\boldsymbol{\theta}^q) = \mathbf{A}(\boldsymbol{\psi}^q)$. Since the signal subspace is the span of the columns of $\mathbf{A}(\boldsymbol{\theta}^q)$, the proof is complete.

Appendix B

Consistency of the Estimator

In this appendix, we show that the estimator (3.38) is consistent. In other words, as the number of observations approaches infinity the true DOAs are estimated. As $i \to \infty$, the sample correlation matrix tends to the true correlation matrix. In the limit the criterion (3.38) is equal to

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\psi}} \mathrm{tr}[\mathbf{P}_n(\boldsymbol{\psi})\mathbf{R}^q]. \tag{B.1}$$

Using $\mathbf{P}_n(\boldsymbol{\psi}) = \mathbf{I} - \mathbf{P}_s(\boldsymbol{\psi})$, the optimization criterion can be shown as

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\psi}} \mathrm{tr}[\mathbf{P}_s(\boldsymbol{\psi})\mathbf{R}^q] \tag{B.2}$$

$$= \arg\max_{\boldsymbol{\psi}} \mathrm{tr}[\mathbf{A}^H(\boldsymbol{\theta}^q)\mathbf{P}_s(\psi)\mathbf{A}(\boldsymbol{\theta}^q)\mathbf{S}] \tag{B.3}$$

where the true angle of arrivals is shown by $\boldsymbol{\theta}^q$. The source correlation matrix $\mathbf{S}$ can be represented by its eigenvalues and eigenvectors as

$$\mathbf{S} = \sum_{i=1}^{q} \lambda_i \mathbf{v}_i \mathbf{v}_i^H. \tag{B.4}$$

Substituting (B.4) in (B.3) yields

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\psi}} \sum_{i=1}^{q} \lambda_i \mathbf{v}_i^H \mathbf{A}^H(\boldsymbol{\theta}^q)\mathbf{P}_s(\boldsymbol{\psi})\mathbf{A}(\boldsymbol{\theta}^q)\mathbf{v}_i \tag{B.5}$$

$$= \arg\max_{\boldsymbol{\psi}} \sum_{i=1}^{q} \lambda_i \mathbf{b}_i^H \mathbf{P}_s(\boldsymbol{\psi}) \mathbf{b}_i \tag{B.6}$$

where $\mathbf{b}_i = \mathbf{A}(\boldsymbol{\theta}^q)\mathbf{v}_i$ is a linear combination of the true location vectors. The projection operator $\mathbf{P}_s(\boldsymbol{\psi})$ is a positive-semidefinite matrix with q eigenvalues equal to 1 and the rest equal to zero. Using Rayleigh-Ritz theorem [15], we have

$$\mathbf{b}_i^H \mathbf{P}_s(\boldsymbol{\psi}) \mathbf{b}_i \leq |\mathbf{b}_i|^2. \tag{B.7}$$

Therefore

$$\sum_{i=1}^{q} \lambda_i \mathbf{b}_i^H \mathbf{P}_s(\boldsymbol{\psi}) \mathbf{b}_i \leq \sum_{i=1}^{q} \lambda_i |\mathbf{b}_i|^2. \tag{B.8}$$

As it is seen the upper bound of the criterion is independent of $\boldsymbol{\psi}$. This bound is achieved if and only if $\mathbf{P}_s(\boldsymbol{\psi})$ is a projection onto the subspace spanned by the columns of $\mathbf{A}(\boldsymbol{\theta}^q)$, i.e.

$$\mathbf{P}_s(\boldsymbol{\psi}) \mathbf{b}_i = \mathbf{b}_i \qquad \forall i. \tag{B.9}$$

This equality also holds for all $\mathbf{b} \in \mathcal{C}_s = \text{span}\{\mathbf{b}_i,\ i = 1, \ldots, q\}$. Using the definition of $\mathbf{b}_i$,

$$\mathbf{P}_s(\boldsymbol{\psi}) \mathbf{A}(\boldsymbol{\theta}^q) \mathbf{v}_i = \mathbf{A}(\boldsymbol{\theta}^q) \mathbf{v}_i \qquad \forall i. \tag{B.10}$$

This equality should hold independent of the source correlation matrix. Thus

$$\mathbf{A}(\boldsymbol{\psi}) [\mathbf{A}^H(\boldsymbol{\psi}) \mathbf{A}(\boldsymbol{\psi})]^{-1} \mathbf{A}^H(\boldsymbol{\psi}) \mathbf{A}(\boldsymbol{\theta}^q) = \mathbf{A}(\boldsymbol{\theta}^q). \tag{B.11}$$

It is obvious that $\boldsymbol{\theta}^q$ satisfies the equation (B.11). Since $\mathbf{P}_s(\boldsymbol{\theta}^q)$ is an orthogonal projection, it is unique. We claim that $\boldsymbol{\theta}^q$ is also unique. Suppose it is not the case. Then there is $\boldsymbol{\psi}' \neq \boldsymbol{\theta}^q$ for which

$$\mathbf{P}_s(\boldsymbol{\psi}') \mathbf{b} = \mathbf{b} \qquad \forall\ \mathbf{b} \in \mathcal{C}_s. \tag{B.12}$$

Uniqueness of the orthogonal projection implies

$$\mathbf{P}_s(\boldsymbol{\psi}') = \mathbf{P}_s(\boldsymbol{\theta}^q). \tag{B.13}$$

The set of independent vectors $\mathcal{A}(\boldsymbol{\psi}') = \{\mathbf{a}(\psi_i'); \ i = 1, \ldots, q\}$ form a basis for a q-dimensional subspace. Since it is assumed that any p location vectors are independent, there is at least one vector in the set $\mathcal{A}(\boldsymbol{\psi}')$ which cannot be represented with respect to the basis $\mathcal{A}(\boldsymbol{\theta}^q) = \{\mathbf{a}(\theta_i); \ i = 1, \ldots, q\}$. The projector $\mathbf{P}_s(\boldsymbol{\psi}')$ is the projection matrix onto the subspace spanned by the vectors of the set $\mathcal{A}(\boldsymbol{\psi}')$ which is different from the true signal subspace spanned by the columns of $\mathbf{A}(\boldsymbol{\theta}^q)$. This is in contradiction to (B.13). Thus, no $\boldsymbol{\psi}' \neq \boldsymbol{\theta}^q$ can satisfy (B.13).

Appendix C

Singular Values of a Matrix Product

In this appendix Lemma 5.1 is proved. The following lemma is adopted from [15].

Lemma C.1 . *Let* $\mathbf{A}, \mathbf{B} \in \mathbf{M}_{m\times n}$ *be* $m \times n$ *matrices with* $q = \min\{m, n\}$. *Denote by* $\sigma_i(\mathbf{A}), \sigma_i(\mathbf{B})$ *and* $\sigma_i(\mathbf{AB}^H)$, $i = 1, \ldots, q$, *the nonzero singular values of the corresponding matrices arranged in nonincreasing order. Then for* $\mathbf{U} \in \mathbf{M}_{m\times m}, \mathbf{V} \in \mathbf{M}_{n\times n}$ *being unitary*

$$\max_{\mathbf{U},\mathbf{V}} \{\Re \text{ tr} \mathbf{A}\mathbf{V}^H\mathbf{B}^H\mathbf{U}^H\} = \sum_{i=1}^{q} \sigma_i(\mathbf{A})\sigma_i(\mathbf{B}). \tag{C.1}$$

We use Lemma C.1 to prove Lemma 5.1. Suppose that the singular value decomposition of the matrices $\mathbf{A}$ and $\mathbf{B}$ are given by

$$\mathbf{A} = \mathbf{E}\mathbf{\Lambda}_a\mathbf{F}^H \tag{C.2}$$

$$\mathbf{B} = \mathbf{X}\mathbf{\Lambda}_b\mathbf{Y}^H. \tag{C.3}$$

Then

$$\begin{aligned} \sum_{i=1}^{q} \sigma_i(\mathbf{AB}^H) &= \Re \text{ tr}(\mathbf{AB}^H) \\ &= \Re \text{ tr}(\mathbf{\Lambda}_a\mathbf{F}^H\mathbf{Y}\mathbf{\Lambda}_b\mathbf{X}^H\mathbf{E}). \end{aligned} \tag{C.4}$$

Define

$$\mathbf{V} = \mathbf{Y}^H\mathbf{F} \tag{C.5}$$

$$\mathbf{U} = \mathbf{E}^H\mathbf{X}. \tag{C.6}$$

Hence

$$\sum_{i=1}^{q} \sigma_i(\mathbf{A}\mathbf{B}^H) = \Re \text{ tr}(\mathbf{\Lambda}_a\mathbf{V}^H\mathbf{\Lambda}_b\mathbf{U}^H). \tag{C.7}$$

Using Lemma C.1 , the maximum of the right-hand side of (C.7) is given by the multiplication of the singular values of the diagonal matrices $\mathbf{\Lambda}_a$ and $\mathbf{\Lambda}_b$. Thus, we have

$$\sum_{i=1}^{q} \sigma_i(\mathbf{A}\mathbf{B}^H) \leq \sum_{i=1}^{q} \sigma_i(\mathbf{A})\sigma_i(\mathbf{B}). \tag{C.8}$$

And the proof is complete.

Appendix D

Minimization of the Subspace Fitting Error

In this appendix we prove Theorem 5.1. The error of the two-sided unitary transformation is given by

$$\begin{aligned}\mathcal{E} &= \|\mathbf{A} - \mathbf{U}\mathbf{B}\mathbf{V}^H\|^2 \\ &= \|\mathbf{A}\|^2 + \|\mathbf{B}\|^2 - 2\Re\,\mathrm{tr}(\mathbf{A}\mathbf{V}\mathbf{B}^H\mathbf{U}^H). \end{aligned} \tag{D.1}$$

Minimization of (D.1) with respect to the unitary matrices $\mathbf{U}$ and $\mathbf{V}$ is identical to maximization of

$$\max_{\mathbf{U},\mathbf{V}} \Re\,\mathrm{tr}(\mathbf{A}\mathbf{V}\mathbf{B}^H\mathbf{U}^H) \tag{D.2}$$

$$\text{s.t.} \quad \mathbf{U}^H\mathbf{U} = \mathbf{I}$$

$$\mathbf{V}^H\mathbf{V} = \mathbf{I} \tag{D.3}$$

subject to $\mathbf{V}$ and $\mathbf{U}$ being unitary transformations. From Lemma C.1 it is seen that the maximum value of (D.2) is given by $\sum_{i=1}^{q}\sigma_i(\mathbf{A})\sigma_i(\mathbf{B})$. Let us represent the singular value

decomposition of the two matrices $\mathbf{A}$ and $\mathbf{B}$ by

$$\mathbf{A} = \mathbf{E}\boldsymbol{\Lambda}_a\mathbf{F}^H \tag{D.4}$$

$$\mathbf{B} = \mathbf{X}\boldsymbol{\Lambda}_b\mathbf{Y}^H \tag{D.5}$$

where the diagonal elements of $\boldsymbol{\Lambda}_a$ and $\boldsymbol{\Lambda}_b$ are the singular values of $\mathbf{A}$ and $\mathbf{B}$, and $\mathbf{E}$, $\mathbf{F}$, $\mathbf{X}$, and $\mathbf{Y}$ are the corresponding left and right matrices of singular vectors. It is straightforward to show that with

$$\mathbf{U} = \mathbf{E}\mathbf{X}^H \tag{D.6}$$

$$\mathbf{V} = \mathbf{F}\mathbf{Y}^H \tag{D.7}$$

the maximum is achieved. This completes the proof.

Appendix E

Unbiased Condition for a Single Source

In this appendix, we show that for a single source scenario and the diagonal unitary transformation matrices

$$\mathbf{T}_j = \text{diag}[1, e^{-j(\omega_0-\omega_j)\tau_0}, \ldots, e^{-j(p-1)(\omega_0-\omega_j)\tau_0}] \tag{E.1}$$

where τ_0 is the propagation delay estimated by the pre-processing, the result of Lemma 5.3 is in agreement with the work of Swingler and Krolik [36]. Using the diagonal matrix (E.1), the condition of lemma (5.39) can be shown as

$$s_0\mathbf{a}_0\mathbf{a}_0^H = \frac{1}{J}\sum_{j=1}^{J} s_j\mathbf{b}_j\mathbf{b}_j^H \tag{E.2}$$

where the transformed column vector $\mathbf{b}_j$ is given by

$$\mathbf{b}_j = [1 \; e^{-j\omega_0\tau_0+j\omega_j(\tau_0-\tau_1)} \; \ldots \; e^{-j(p-1)\omega_0\tau_0+j(p-1)\omega_j(\tau_0-\tau_1)}]^T \tag{E.3}$$

where τ_1 is the true DOA and the superscript T stands for transpose. The direction of arrival is estimated by equating

$$\begin{aligned} s_0\ e^{-j\omega_0\hat{\tau}_1} &= \sum_{j=1}^{J} s_j e^{-j\omega_0\tau_0+j\omega_j(\tau_0-\tau_1)} \\ &= e^{-j\omega_0\tau_0}\sum_{j=1}^{J} s_j e^{j\omega_j(\tau_0-\tau_1)} \end{aligned} \tag{E.4}$$

where $\hat{\tau}_1$ is the estimate of τ_1. Assuming

$$\omega_j(\tau_0-\tau_1) \ll 1, \tag{E.5}$$

we can approximate (E.4) to get

$$s_0\ e^{-j\omega_0\hat{\tau}_1} \approx e^{-j\omega_0\tau_0}\sum_{j=1}^{J} s_j[1+j\omega_j(\tau_0-\tau_1)] \tag{E.6}$$

$$= s_0\ e^{-j\omega_0\tau_0}\ [1+j(\tau_0-\tau_1)\frac{1}{s_0}\sum_{j=1}^{J} s_j\omega_j] \tag{E.7}$$

$$= s_0\ e^{-j\omega_0\tau_0}\ [1+j(\tau_0-\tau_1)\bar{\omega}] \tag{E.8}$$

$$= s_0\ e^{-j\omega_0\tau_0}\ e^{j(\tau_0-\tau_1)\bar{\omega}} \tag{E.9}$$

where $\bar{\omega}$ is the centroid frequency. Equating the exponents gives

$$\omega_0\hat{\tau}_1 = \omega_0\tau_0 - (\tau_0-\tau_1)\bar{\omega} \tag{E.10}$$

which is the same result as Swingler and Krolik [36].

Bibliography

[1] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Auto. Cont.*, vol. AC-19, pp. 716–723, 1974.

[2] T. W. Anderson, "Asymptotic theory for principal component analysis," *Ann. J. Math. Stat.*, vol. 34, pp. 122–148, 1963.

[3] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons Inc., 1984.

[4] G. Bienvenu and L. Kopp, "Optimality of high resolution array processing using the eigensystem approach," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 31, pp. 1235–1248, 1983.

[5] Y. Bresler and A. Macovski, "On the number of signals resolvable by a uniform linear array," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, pp. 1361–1375, 1986.

[6] K. M. Buckley, "Spatial/spectral filtering with linearly constrained minimum variance beamformer," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, pp. 249–266, March 1987.

[7] K. M. Buckley and L. J. Griffiths, "Broad-band signal-subspace spatial-spectrum (BASS-ALE) estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 953–964, July 1988.

[8] P. E. Caines, *Linear Stochastic Systems*. John Wiley & Sons Inc., 1988.

[9] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, pp. 1408–1418, August 1969.

[10] L. D. Davisson, "The prediction error of stationary Gaussian time series of unknown covariance," *IEEE Trans. Information Theory*, vol. IT–11, pp. 527–532, October 1965.

[11] M. A. Doron and A. J. Weiss, "On focusing matrices for wide-band array processing," *IEEE Trans. Signal Processing*, vol. SP–40, pp. 1295–1292, June 1992.

[12] G. H. Golub and C. F. V. Loan, *Matrix Computations*. The Johns Hopkins University Press, 1983.

[13] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. on Antenna and Prop*, vol. 30, pp. 27–82, Jan. 1982.

[14] S. Haykin, ed., *Array Signal Processing*. Prentice Hall, 1985.

[15] R. A. Horn and C. A. Johnson, *Matrix Analysis*. Cambridge, England: Cambridge University Press, 1985.

[16] R. A. Horn and C. A. Johnson, *Topics in Matrix Analysis*. Cambridge, England: Cambridge University Press, 1991.

[17] H. Hung and M. Kaveh, "Focusing matrices for coherent signal–subspace processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1272–1281, August 1988.

[18] W. Hurewicz and H. Wallman, *Dimension Theory*. Princeton University Press, 1948.

[19] T. P. Jäntti, "The influence of extended sources on the theoretical performance of the MUSIC and ESPRIT methods: Narrow-Band sources," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. II–429—II–432, March 1992.

[20] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. Prentice Hall, 1993.

[21] W. C. Knight, R. G. Pridham, and S. M. Kay, "Digital signal processing for sonar," *proc. IEEE*, vol. 69, pp. 1451–1506, Nov. 1981.

[22] J. Krolik and D. N. Swingler, "Multiple broad-band source location using steered covariance matrices," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1481–1494, October 1989.

[23] S. J. Orfanidis, *Optimum signal processing: An introduction*. Macmillan, 1985.

[24] S. U. Pillai, *Array Signal Processing*. Springer-Verlag, 1989.

[25] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1986.

[26] J. Rissanen, "Stochastic complexity and modeling," *Annals of statistics*, vol. 14, pp. 1080–1100, 1986.

[27] J. Rissanen, *Stochastic complexity in statistical inquiry*. World Scientific Publisher, 1989.

[28] R. Roy and T. Kailath, "ESPRIT – Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 984–995, July 1989.

[29] R. Roy, A. Paulraj, and T. Kailath, "Direction-of-arrival estimation by subspace rotation method – ESPRIT," *proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 2495–2498, 1986.

[30] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. on Antenna and Prop*, vol. 34, pp. 276–280, March 1986.

[31] T. J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 33, pp. 806–811, August 1985.

[32] D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty – I," *Bell Syst. Tech. J.*, vol. 40, pp. 43–64, 1961.

[33] D. Slepian and E. Sonnenblick, "Eigenvalues associated with prolate spheroidal wave functions of zero order," *Bell Syst. Tech. J.*, vol. 44, pp. 1745–1759, 1965.

[34] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1783–1795, Oct. 1990.

[35] G. Su and M. Morf, "The signal subspace approach for multiple wideband emitter location," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 31, pp. 1502–1522, Dec. 1983.

[36] D. N. Swingler and J. Krolik, "Source location bias in the coherently focused high-resolution broad-band beamformer," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 143–145, Jan. 1989.

[37] D. W. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids: making linear prediction perform like maximum likelihood," *Proceedings of IEEE*, vol. 70, pp. 975–989, Sept. 1982.

[38] S. Valaee, B. Champagne, and P. Kabal, "Parametric localization of distributed sources," submitted to *IEEE Trans. Signal Processing*, Oct. 1993.

[39] S. Valaee and P. Kabal, "Detection of the number of signals using Predictive Stochastic Complexity," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. V–345–V–348, March 1992.

[40] S. Valaee and P. Kabal, "Selection of the focusing frequency in wideband array processing," *Proc. 26-th Ann. Conf. Inform. Sci. Sys.*, Princeton, USA, 1992.

[41] S. Valaee and P. Kabal, "Selection of the focusing frequency in wideband array processing – MUSIC and ESPRIT," *Proc. 16-th Biennial Symposium on Communication,* Kingston, Ontario, pp. 410–414, May 1992.

[42] S. Valaee and P. Kabal, "A unitary transformation algorithm for wideband array processing," *Sixth SP Workshop on Statistical Signal and Array Processing*, pp. 300–303, October 1992.

[43] S. Valaee and P. Kabal, "Alternate proofs for "On unique localization of constrained-signal sources"," submitted to *IEEE Trans. Signal Processing*, Nov. 1993.

[44] S. Valaee and P. Kabal, "Optimal focusing subspace for coherent signal subspace processing," submitted to *IEEE Trans. Signal Processing*, Nov. 1993.

[45] S. Valaee and P. Kabal, "A stochastic complexity approach to signal detection in array processing," submitted to *IEEE Trans. Signal Processing*, Dec. 1993.

[46] S. Valaee and P. Kabal, "Wideband array processing using a two-sided correlation transformation," accepted for publication in *IEEE Trans. Signal Processing*, 1994.

[47] S. Valaee, P. Kabal, and B. Champagne, "Localization of distributed sources," *Fourteenth GRETSI Symposium on Signal and Image Processing* Juan les Pins, France, pp. 289–292, Sept. 1993.

[48] S. Valaee, P. Kabal, and B. Champagne, "A parametric approach for extended source localization," to be presented at *VII European Signal Processing Conference,* (Edinburgh, Scotland), Sept. 1994.

[49] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I.* New York: Wiley, 1968.

[50] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part III.* New York: Wiley, 1968.

[51] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, pp. 4–24, April 1988.

[52] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, pp. 823–831, August 1985.

[53] M. Wax, "Detection and localization of multiple sources via the stochastic signals model," *IEEE Trans. Signal Processing*, vol. 39, pp. 2450–2456, November 1991.

[54] M. Wax, "On unique localization of constrained-signals sources," *IEEE Trans. Signal Processing*, vol. 40, pp. 1542–1547, June 1992.

[55] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. on Acoust., Speech, Signal Processing*, vol. 33, pp. 387–392, April 1985.

[56] M. Wax, T. Shan, and T. Kailath, "Spatio-temporal spectral analysis by eigenstructure methods," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 32, pp. 817–827, August 1984.

[57] M. Wax and I. Ziskind, "Detection of the number of coherent signals by the MDL principle," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1190–1196, August 1989.

[58] M. Wax and I. Ziskind, "On unique localization of multiple sources by passive sensor arrays," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 996–1000, July 1989.

[59] K. M. Wong, Q. T. Zhang, J. P. Reilly, and P. C. Yip, "On Information theoretic criteria for determining the number of signals in high resolution array processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1959–1971, November 1990.

[60] Q. Wu and D. R. Fuhrmann, "A parametric method for determining the number of signals in narrow-band direction finding," *IEEE Trans. Signal Processing*, vol. 39, pp. 1848–1857, August 1991.

[61] Y. Q. Yin and P. R. Krishnaiah, "On some nonparametric methods for detection of the number of signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, pp. 1533–1538, November 1987.

[62] Q. T. Zhang, K. M. Wong, P. C. Yip, and J. P. Reilly, "Statistical analysis of the performance of information theoretic criteria in the detection of the number of signals in array processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1557–1567, October 1989.

[63] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "On detection of number of signals in presence of white noise," *J. Multivariate Anal.*, vol. 20, pp. 1–25, 1986.

[64] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projection," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1553–1560, October 1988.