Model Order Determination and Characterisation of Direction of Arrival (DOA) Estimators in the Acoustic Context.

A dissertation submitted to the University of Dublin
for the degree of Doctor of Philosophy

Angela Elizabeth Quinlan
Trinity College Dublin, October 2006

Department of Electronic and Electrical Engineering
Trinity College Dublin

To my parents.

Declaration

I hereby declare that this thesis has not been submitted as an exercise for a degree at this or any other University and that it is entirely my own work.

I agree that the Library may lend or copy this thesis upon request.

Signed,

Angela E. Quinlan

October 27, 2006.

Summary

This thesis is concerned with the detection and characterization of acoustical sources using array signal processing techniques. In particular this work is concerned with determining the number of sources present, and estimation of the Direction of Arrival (DOA) of the signals received by the array.

Firstly the problem of DOA estimation using array processing techniques is introduced and the possible applications of DOA estimation are discussed. The model of the wavefront propagating from the source to the array is recalled.

Chapter 2 provides the mathematical basis for the rest of the thesis. The mathematical estimation problem is introduced and the difficulties in finding an optimal estimator for DOA estimation are discussed.

A unified explanation and review of classical array processing DOA estimation techniques is then given. In particular the application of such techniques to the problem of estimating the location of wideband sources is discussed. Recent developments in acoustical source localisation are then reviewed.

The initial step in any estimation scheme is to determine the number of sources present. This process is called model order determination and classical model order determination methods are unsuited to situations where the number of snapshots available is small, however such situations frequently arise, particularly when dealing with non-stationary sources. Based on the profile of the noise eigenvalues of the observed data correlation matrix as introduced by Grouffaud et. al [1], a novel method of determining the number of acoustical sources present is presented. The performance of the proposed method is compared to classical model order determination techniques using both computer simulations and experimental recordings. In particular the effect of reverberation is considered. The proposed method is shown to outperform the classical methods while maintaining low computational complexity.

As it is not possible to find an optimal estimator for the DOA estimation problem, a sub-optimal estimator must instead be used. The choice of estimator depends on the characteristics of both the source and environment of the given estimation problem. Three of the most well known approaches to the DOA estimation problem are subspace-based estimation, in this case we consider the Multiple Signal Classification(MUSIC) algorithm, Maximum Likelihood (ML) estimation and Time-Delay Estimation (TDE) using the cross-correlation method.

The performance of these three approaches are then compared using both computer

simulations and experimental recordings. Analysis of the Mean Square Error (MSE) of each of the estimators for decreasing SNR shows that all three estimators display a threshold SNR below which the MSE increases rapidly. This threshold determines the operational range of the estimators for the given estimation problem and is explained by examination of the behaviour of the estimators as the SNR is decreased. The effect of reverberation is considered by evaluating the performance of the methods for varying Useful-to-Detrimental U_{25} values. Comparison of the results found from each estimator allows some conclusions to be drawn on the suitability of these estimation approaches for different localisation problems.

For any given localisation problem it is therefore of great practical importance to know the best possible performance that can achieved. Such knowledge allows for a decision to be made on whether or not starting from a given algorithm an improvement can be made, or whether or not the system performance requirements can ever be met. It is of particular importance to establish the threshold SNR value, as this value determines the operational range.

In this thesis the best performance that can be achieved by an estimator is evaluated using lower bounds on the Mean Square Error (MSE), allowing for prediction of the operating range of an estimator in a given situation. The best known of these bounds is the Cramér-Rao Bound (CRB). Its popularity can be attributed to both its simplicity of calculation and the fact that asymptotically (high SNR and/or large number of samples) it is reached by the ML estimator. However, the CRB only provides an accurate indication of achievable performance in the asymptotic (or small-error) region, as it does not predict the threshold effect. The resulting performance characterization may therefore be misleadingly optimistic.

For this reason the application of the Barankin Bound (BB), which is the greatest lower bound on the MSE of an unbiased estimator, is instead considered. Unfortunately however, it is generally not computable, and must therefore be approximated. In this thesis a practical means of classification of BB approximations is proposed. This classification scheme includes all previously existing approximation bounds, and highlights their underlying assumptions.

Finally this formalism leads to the derivation of a new practical bound that is tighter than existing bounds, particularly in the threshold region, while maintaining a comparable computational complexity to that of the CRB.

Acknowledgments

Firstly, I would like to thank my supervisor, Prof. Frank Boland for placing his trust in me to begin work in his new research group, and for his help throughout the course of my PhD. It is thanks to your powers of persuasion that I decided to do a PhD, and for this I am very grateful.

I would also like to express my gratitude to the examiners for the time and effort they put into exmaining this thesis.

I am also very grateful to the Irish Research Council for Science Engineering and Technology (IRCSET) for their generous funding of this work. In particular I would like to thank Dawn Carroll and Vicky Garnett of IRCSET, for their rapid help on any matters that arose throughout the course of my PhD.

I want to say a big thank you to the other post-graduate students and members of staff in the Electronics department in Trinity College for their help and friendship throughout the past few years. In particular I would like to thank Gavin, Deepti, Dennis and Damian for all their help and Francis for answering all those very important questions at the end. I also especially want to thank Claire for always helping me to see a funny side during the darker days!

It is hard to know where to begin to express my gratitude to the people in the ENS de Cachan, France. From the moment I arrived at the ENS you did all you could to make me feel at home. Benoit, Christoph, Stephanie, Francois, Mark, Leila, Ana, Eric Vourch, Ann, Dominique and the many many others. I can never express all that your friendship, teaching, humour and patience means to me.

Alex, I especially want to thank you for your incredible generosity and willingness to share your knowledge, your supervisor, and your very unique sense of humour - Thank you!

Also a very special word to Eric Chaumette. I just can't possibly thank you enough for having the patience and energy to introduce me to the wonderful world of minimal performance bounds, and for always being so generous with your time and support. Thank you so very much.

Finally, Jean-Pierre and Pascal, it is you who made it possible for me to be at the ENS, and you who made sure that the work got done (mais toujours comme je voulais!). You have taught me so much about research, signal processing and of course French wine. In your office you achieve a truly remarkable combination of team spirit, hard work and fun, and I know how privileged I am to have been part of this.

I would like to thank F. Asano and his team at AIST, Tsukuba, Japan, who made me so welcome during my time in Tsukuba. Thank you so much for your hospitality and generosity.

I also want to say a big thanks to Zsolt, for the continuous advice and support, and most of all for refusing to let me make excuses!

Finally I thank the people who have always been there for me, and without whom none of this would be possible - my family. Your love, help, encouragement and humour are a constant source of strength, Thank you.

Contents

List of Figures

List of Tables

List of Acronyms

DOA	**D**irection **O**f **A**rrival
pdf	**P**robability **D**ensity **F**unction
MSE	**M**ean **S**quare **E**rror
ML	**M**aximum **L**ikelihood
MUSIC	**MU**ltiple **SI**gnal **C**lassification
TDE	**T**ime **D**elay **E**stimation
SNR	**S**ignal to **N**oise **R**atio
MSE	**M**ean **S**quare **E**rror
CRB	**C**ramér **R**ao **B**ound
MSB	**M**cAulay **S**eidman **B**ound
BB	**B**arankin **B**ound
HCRB	**H**ammersley **C**hapman **R**obbins **B**ound

1 Introduction

1.1 Introduction

Array processing has been an active area of research for many years now, and originally array processing techniques were developed for military applications. However, the dramatic increases in computing power which have taken place over the last number of years have led to the widespread use of Digital Signal Processing (DSP) devices in consumer electronics, for both business and entertainment purposes. The phenomenal growth of this industry has provided many new and challenging problems for signal processing researchers, as there is a constant demand for increased speed, accuracy and robustness, while reducing price and size.

In particular the area of acoustical array processing has become an active research topic. This interest can be attributed to the host of potential applications, for example: sonar applications; medical applications such as lithotripsy; non-destructive testing and human-computer interfacing; as well as a wide range of entertainment applications e.g. tracking of acoustical sources during theatrical performances and acoustical ambiance re-creation.

At its most basic, signal processing is concerned with transmission of a signal that contains some desired information, and manipulation of this signal in order to extract this

information. In array signal processing the signal or signals are emitted and/or received by an array. By array, we mean a set of receivers, that are spatially distributed. This separation of the receivers means that as well as the signal being temporally sampled, i.e. a value is received at each time instant, it is also spatially sampled as a value is received at each of the array elements. The advantage of using an array of receivers is this ability to exploit both the spatial and temporal characteristics of the signal.

Acoustical array processing is concerned with detection and manipulation of signals emitted by acoustical sources and has numerous applications in sonar, medicine, audio, active noise control etc. When dealing with acoustic sources the signals are received by an array of microphones.

1.2 Array Processing

For array signal processing purposes signals can be divided into two groups: those that have a fixed behaviour, called deterministic signals, and those that change randomly, called stochastic or random signals. Deterministic signals can be completely specified as a function of time and therefore the signal can be predicted from a number of previous time samples.

On the other hand stochastic or random signals cannot be easily characterised by a mathematical expression, and instead use must be made of prior knowledge and probabilities in order to analyse the signal behaviour, for example the use of a prior probability of a parameter to estimated the current value.

The signals of interest can be further classified as narrow- or wideband. A narrowband signal is one whose amplitude and phase vary slowly relative to the time taken for it to propagate across the array. A broadband signal is then a signal that is not narrowband, or a signal which has a relatively large frequency bandwidth compared to its centre frequency.

The source emitting the signal can be classified as near-field or far-field. Near-field sources are located close enough to the array for the wavefront arriving at the array to be spherical. On the other hand, the wavefronts arriving from a far-field source have been propagating for long enough for the wavefronts to be planar, i.e in straight horizontal lines as they arrive at the array.

As the signal travels through the propagating medium it will be perturbed by additive noise and interfering signals. The presence of these signals change the properties of the received signal and make it difficult to extract the desired information.

Additive noise is generated by sources in the same environment of the signal, and is usually independent of the signals themselves. In signal processing, an assumption of

white noise, that is noise with equal quantities of all frequencies, is often made. One reason for making such an assumption is that it greatly simplifies the mathematics involved in subsequent signal models. Fortunately, it is also a good approximation of the true noise present in many cases.

In some situations, however, the noise present may not contain all frequencies, and this noise is instead called coloured noise. In these situations, if the spectrum of the coloured noise is known, a whitening filter is usually applied to whiten the noise. Once this has been done, the characteristics of white noise can once again be exploited for any mathematical modelling. However, this requires access to the noise signal separate from the information signal and this is often not possible in practical situations.

Interference, on the other hand, will usually have similar characteristics to the desired source signal, and may be generated by a similar source, e.g. people speaking in the background when trying to extract a specific speech signal. Another type of interference is due to reverberation or multi-path, when the desired signal is reflected off surfaces within the propagation environment causing multiple delayed arrivals of the desired signal. The degradation of the desired signal by interference and reverberation is usually very difficult to deal with, as it becomes confused with the desired signal and therefore cannot be easily identified or removed.

Once the signal has been received by the array, the objectives of any subsequent processing steps can be categorised as either signal enhancement or field characterisation. Signal enhancement occurs when the spatial characteristics of the array are used to improve the SNR of the signal received. This can be done by steering the array so that it receives signals from a certain direction only, thereby ignoring signals arriving from other directions. This technique is called beamforming and in its simplest form is performed by delaying the signals received at each of the array elements, and then adding these delayed signals.

The delay applied corresponds to the time delay experienced by a signal originating from the desired location, as it propagates across the array. Signals originating from other locations will therefore be filtered out. This spatial filtering is very useful in situations where the interfering and noise signals overlap the desired signal spatially or temporally, making other types of filtering difficult.

Field characterisation is concerned with estimation of the spatial properties of the source or sources. Before the location of the sources can be found the first step must be to identify the number of sources present. This process can be called model order determination.

The next step then is localisation of the sources present. For near-field sources emitting

stationary signals the localisation step involves estimation of the range, asimuth and elevation of the source. The number of parameters to be estimated is reduced for far-field sources, as only the asimuth and elevation can be estimated. These parameters are commonly further reduced by assuming the source and array are in the same plane, i.e. at the same height, which reduces the localisation problem to that of estimating the asimuth Direction of Arrival (DOA) of the signal only. In this thesis we are concerned with the estimation of the DOA of signals arriving from a far-field source. Extension to the situation where the source and array are not on the same plane, and the elevation must also be estimated is straight-forward.

Estimation of the DOA of an arriving signal has been an active area of research for many years now, and a vast number of algorithms and approaches exist. While many of these methods were developed originally for narrowband signals, they may also be applied to situations of broadband sources such as those encountered in acoustical array processing. In order to apply narrowband localisation techniques the signals are first broken down into narrowband bins, and localisation is then performed individually on each bin. These individual results are then combined to give an overall DOA estimate.

Despite the many localisation techniques that have been developed, the problem of DOA estimation continues to be a challenging problem. One of the difficulties lies in the fact that there is no guaranteed way of finding the best or optimal method of estimation for all situations. Instead sub-optimal methods must be used, and the most suitable approach is selected by taking into account characteristics of the source and environment.

In order to determine the suitability of a proposed estimation method (i.e. the estimator) its performance can be compared to the best possible performance. This allows a decision to be reached on whether or not the estimator performance is satisfactory, or if it can be improved upon. It can also be determined whether or not it will ever be possible to achieve the required performance. One method of evaluating the best possible performance is by the use of lower bounds on the Mean Square Error (MSE) bounds, which provide a bound on the minimal MSE that can be achieved.

1.3 Thesis Organisation

In this thesis, the problems of detection and estimation of the DOA of far-field acoustical sources using an array of microphones are considered. In chapter 2 array signal processing techniques are introduced and a model for the data received by the array is proposed. Estimation of the unknown parameters of the received signals is then reviewed, and the individual steps in this process are explored, including specification of the Probability

Density Function (pdf) and commonly applied assumptions. The selection of a suitable estimator and the criteria used to evaluate estimator performance are then considered. These criteria can then be applied in order to find the optimal estimator for a given estimation problem. However, it is shown that it is not always possible to find an expression for the optimal estimator, and that DOA estimation is an example of such a situation. We can therefore conclude that for this problem it is instead necessary to select a sub-optimal estimator that has desirable properties for the given situation. The commonly applied sub-optimal DOA estimators are then considered in chapter 3 and the main contributions to the area of DOA estimation are reviewed.

In chapter 4 the initial step in characterisation of the sources, that of model order determination is discussed. Acoustical model order determination presents a challenging problem due to both the wideband nature of the source and the fact that the amount of data available for a given source location may be limited. Classical methods operate well for determining the number of narrowband sources when large amounts of data are available. However, they are unreliable for the difficult situation considered here.

To this end, a method for determining the number of acoustical sources is proposed here. This method is based on the profile of the noise eigenvalues of the observed data correlation matrix as introduced by Grouffaud et al. [1]. This method is suited to the difficult operating scenarios encountered in acoustical source localisation. In particular this method is shown to out-perform classical detection methods for the situations where a small amount of data is available for a given source location.

This then leads to a study of three of the most well-known sub-optimal estimators, in chapter 5. The performance of these estimators is compared using both simulations and experimental recordings for a variety of source numbers and locations. The effects of interference and additive noise on the estimators is considered, allowing us to draw some conclusions on estimator behaviour, and the suitability of the estimators for different localisation problems.

In chapter 6 the best possible performance that can be achieved by an estimator for a given localisation problem is evaluated. This is done using lower bounds on the Mean Square Error (MSE), which provide a bound on the minimum MSE that can possibly be achieved.

Firstly, a new formalism that encompasses all previously derived lower MSE bounds is proposed. This formalism provides a meaningful way of classifying these bounds based on their underlying assumptions. Secondly, with the help of this formalism, a new practical bound is derived, which, while maintaining low computational complexity is closer to the true performance than existing bounds.

Conclusions are drawn in chapter 7 by commenting on the results from the previous chapters and finally, some proposals for future work are presented.

1.4 Contributions

The work done in the course of this thesis has led to the following papers:

1.4.1 Journal Papers:

1. A. Quinlan, J. P. Barbot, and P. Larzabal, "Automatic determination of the number of targets present when using the time reversal operator (TRO)," Journal Acoustical Society of America (JASA), vol.119, n.4, pp. 2220-2225, 2006.

2. A. Quinlan, J. P. Barbot, P. Larzabal, and M. Haardt, "Model order selection for short data: An exponential fitting test (EFT)," EURASIP JASP (European Journal Applied Signal Processing), Accepted for publication.

1.4.2 Conference Papers:

1. A.Quinlan and F. Boland, "Using the singer's formant to reduce inaccuracies in the location of a singer on stage," in Proc. Irish Signals and Systems Conference (ISSC), Belfast, Ireland, 2004.

2. A. Quinlan and F. Boland, "The effect of vibrato on singer localisation," in Proc. GSPx Conference, Santa Clara, CA, 2004.

3. A. Quinlan, E. Chaumette, and P. Larzabal, "A direct method to generate approximations of the Barankin bound," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP), Toulouse, France, 2006.

4. A. Quinlan, F. Boland, J. Barbot, and P. Larzabal, "Determination of the Number of Wideband Acoustical Sources in a Reverberant Environment" In Proc. 14th European Signal Processing Conference (EUSIPCO), Florence, Italy, 2006.

2

Estimation Using Array Processing Techniques

2.1 Introduction

The use of array signal processing techniques for source localization has been an active area of research for many years, however it remains a challenging problem. The work presented in this thesis is concerned with localization of an acoustic source using an array of microphones. In this chapter the mathematical background of estimation is recalled. The terms defined here are then used throughout the thesis.

There are two possible approaches to the analysis of the data received by the microphone array: parametric and non-parametric. Parametric techniques assign a mathematical model with a fixed number of parameters to the observed data, and the parameters of the model are selected so that the observations fit this model. These techniques use information that is known or assumed to be known about the observed data in order to determine a suitable model. The accuracy of the method is therefore reliant on the accuracy of the underlying model. In non-parametric techniques, on the other hand, no model is imposed on the data.

The work presented here focuses on parametric methods, in which the required infor-

mation is expressed as a parameter of the received signal [2]. This parameter is then be estimated from the observations.

2.2 Modelling the Received Data

Firstly we consider the source, which can be classified as near-field or far-field depending on its distance from the array. If the source is in the near-field then the wavefronts reaching the array are spherical. For this type of source there is a vector of spatial parameters to be estimated containing both bearing and range information.

As the source is moved farther from the array the wavefronts become planar and parallel, as shown in figure 2.1. A source is considered to be in the far-field if:

$$R > \frac{2L^2}{\lambda}, \tag{2.1}$$

where R is the distance from the source to the array, L is the length of the array, and λ is the wavelength of the arriving wave. When the source is in the far-field only the bearing information, or Direction of Arrival (DOA), can be used to characterise the source spatially [3]. This is the situation being investigated in this thesis. A common assumption is that the source and array are in the same plane, thereby further reducing the spatial parameters to be estimated, and the wavefronts arriving at consecutive microphones differ only by an amplitude coefficient and a time-delay.

The scenario where there are P far-field sources present, and the emitted signals are received by an arbitrary array of M microphones is now considered. The impulse response of the mth microphone to a signal $s_p(t)$, emitted from the pth source is denoted $h_{pm}(t)$. This impulse response depends on the locations of both the source and the microphone, the propagating medium, and any gain or attenuation introduced by the microphone itself. The time delay of the pth signal arriving at the mth microphone is denoted τ_{pm}. The output at the mth microphone can then be written as:

$$x_m(t) = \sum_{p=1}^{P} h_{pm}(t) * s_p(t - \tau_{pm}) + n_m(t), \tag{2.2}$$

where $(*)$ denotes convolution and $n_m(t)$ is the additive noise signal received by the mth microphone, it is assumed that $n_m(t)$ is independent of $s_p(t)$, and that $P < M$. The array output can then be expressed as:

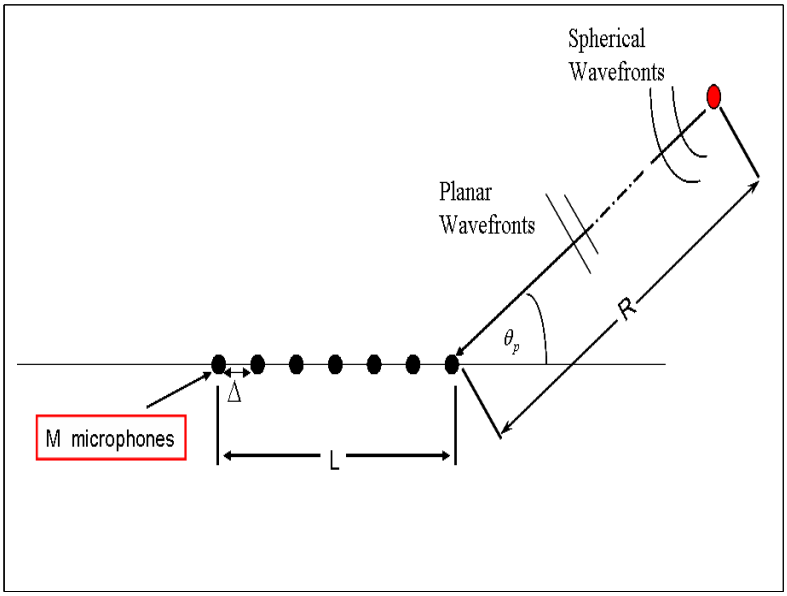


Figure 2.1: *Source in the far field emitting a signal received by the array of microphones.*

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_M(t) \end{bmatrix} = \begin{bmatrix} \sum_{p=1}^{P} h_{p1}(t) * s_p(t-\tau_{p1}) \\ \sum_{p=1}^{P} h_{p2}(t) * s_p(t-\tau_{p2}) \\ \vdots \\ \sum_{p=1}^{P} h_{pM}(t) * s_p(t-\tau_{pM}) \end{bmatrix} + \begin{bmatrix} n_1(t) \\ n_2(t) \\ \vdots \\ n_M(t) \end{bmatrix}. \tag{2.3}$$

2.3 Complex Signal Representation

In array signal processing the complex representation of a signal is often used. This representation is based on the assumption that the signal is narrowband. In many cases this narrowband assumption does not hold, and in these situations the bandwidth of the signal must be divided into smaller frequency bins so that the narrowband assumption will hold across each frequency bin, and the following analysis will then hold for each bin.

The emitted signal $s_p(t)$ can be expressed in terms of a modulated centre frequency ω:

$$s_p(t) = \alpha_p(t)\cos\left[\omega t + \phi_p(t)\right]. \tag{2.4}$$

Assuming that the signal is narrowband we can therefore say:

$$\alpha_p(t-\tau_p) \approx \alpha_p(t), \tag{2.5}$$
$$\phi_p(t-\tau_p) \approx \phi_p(t), \tag{2.6}$$

which implies that:

$$s_p(t-\tau_p) = \alpha_p(t-\tau_p)\cos\left[\omega(t-\tau_p)+\phi_p(t-\tau_p)\right] \tag{2.7}$$
$$\approx \alpha_p(t)\cos\left[\omega t-\omega\tau_p+\phi_p(t)\right]. \tag{2.8}$$

The time-delay can now be modelled as a phase-shift of the carrier frequency. Following from the assumption that the signal is narrowband, the impulse response can be taken to be constant across the frequency band of interest. The stationary response of the *mth* microphone to the signal $s_p(t)$ can therefore be expressed as:

$$r_{pm}(t) = h_{pm}(t) * s_p(t-\tau_{pm}) \tag{2.9}$$
$$= h_{pm}(t) * \alpha_p(t)\cos\left[\omega t-\omega\tau_{pm}+\phi_p(t)\right] \tag{2.10}$$
$$= H_{pm}(\omega)\,\alpha_p(t)\cos\left[\omega t-\omega\tau_{pm}+\phi_p(t)\right], \tag{2.11}$$

where $H_{pm}(\omega)$ is the Fourier transform of the impulse response $h_{pm}(t)$. Expressing $H_{pm}(\omega)$ in terms of its phase, $\arg H_{pm}(\omega)$, and magnitude, $|H_{pm}(\omega)|$, and letting $\beta_{pm} = |H_{pm}(\omega)|$, (2.11) can be re-written as:

$$r_{pm}(t) \backsimeq \beta_{pm}(\omega)\,\alpha_p(t)\cos\left[\omega t-\omega\tau_{pm}+\phi_p(t)+\arg H_{pm}(\omega)\right], \tag{2.12}$$

Using a complex signal representation of the received signal, the time-delay can then be expressed as a multiplication by a complex number. Considering the case where there is no noise present, the complex representation of the signal received by the *mth* microphone is given by:

$$r_{pm}(t) = x_m^i(t) + jx_m^q(t) \tag{2.13}$$
$$= \beta_{pm}(\omega)\,e^{-j\omega\tau_{pm}}\alpha_p(t)\,e^{j\phi_p(t)} \tag{2.14}$$
$$= \beta_{pm}(\omega)\,e^{-j\omega\tau_{pm}}\widetilde{s_p}(t), \tag{2.15}$$

where $\widetilde{s_p}(t)$ is the complex envelope of the signal $s_p(t)$, and $\widetilde{s_p}(t) = \alpha_p(t)\,e^{j\phi_p(t)}$. Then, letting $a_m(\theta_p) = \beta_{pm}(\omega)\,e^{-j\omega\tau_{pm}}$ the array response vector can be defined:

$$\mathbf{a}(\theta_p) = \begin{bmatrix} \beta_{p1}(\omega) e^{-j\omega\tau_{p1}} \\ \beta_{p2}(\omega) e^{-j\omega\tau_{p2}} \\ \vdots \\ \beta_{pM}(\omega) e^{-j\omega\tau_{pM}} \end{bmatrix}. \tag{2.16}$$

This complex vector describes the response of each microphone, including geometric path differences, to a signal arriving at an angle of θ_p. Throughout the thesis this vector is called the array response vector, however it can also be referred to as the steering vector, the location vector or the transfer function vector.
The in-phase and quadrature components of $x_m(t)$, $x_m^i(t)$ and $x_m^q(t)$ respectively, are given by:

$$x_m^i(t) = |H_{pm}(\omega)| \alpha_p(t) \cos\left[\phi_p(t) + \arg H_{pm}(\omega) - \omega\tau_{pm}\right] \tag{2.17}$$
$$x_m^q(t) = |H_{pm}(\omega)| \alpha_p(t) \sin\left[\phi_p(t) + \arg H_{pm}(\omega) - \omega\tau_{pm}\right]. \tag{2.18}$$

2.4 Parametric Signal Model

The required parameter for each source is the Direction of Arrival (DOA), θ_p, and the DOA associated with the each of the P sources present must be estimated from the data received by the array. In the case considered here the angle of elevation is assumed to be zero, as the microphones and source are assumed to be the same height. However, the results derived here can easily be extended to the general case, where both bearing and elevation are unknown.

From equations (2.3) and (2.15), the parameterised data model for the *pth* source is given by:

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_M(t) \end{bmatrix} = \sum_{p=1}^{P} \begin{bmatrix} a_1(\theta_p) \\ a_2(\theta_p) \\ \vdots \\ a_M(\theta_p) \end{bmatrix} \tilde{s_p}(t) + \begin{bmatrix} n_1(t) \\ n_2(t) \\ \vdots \\ n_M(t) \end{bmatrix} \tag{2.19}$$

$$= \sum_{p=1}^{P} \mathbf{a}(\theta_p) \tilde{s}_p(t) + \mathbf{n}(t). \tag{2.20}$$

The received signal model for the case of P parameters can now be expressed as:

$$\mathbf{x}(t) = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \ldots, \mathbf{a}(\theta_P)] [\widetilde{s}_1(t), \ldots, \widetilde{s}_P(t)]^T + \mathbf{n}(t). \tag{2.21}$$

Then defining the $M \times P$ matrix $\mathbf{A}(\boldsymbol{\theta})$ the columns of which are given by the array response vectors associated with each of the P sources, equation (2.21) can be re-expressed as:

$$\mathbf{x}(t) = \mathbf{A}(\boldsymbol{\theta}) \widetilde{\mathbf{s}}(t) + \mathbf{n}(t), \tag{2.22}$$

with:

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_P \end{bmatrix}. \tag{2.23}$$

This model is applicable to an array of arbitrarily located microphones assuming that the sources are located in the far-field; that $P < M$; and that the sources are non-coherent, where two signals are coherent if one is a scaled and delayed version of the other [2]. A Uniform Linear Array (ULA) is commonly used as it leads to simplification of the data model. This is the array configuration considered here. However, the results found are easily extended to other array configurations.

For a ULA with an inter-microphone spacing Δ, τ_{pm} can be related to the DOA by the following expression, as seen in figure (2.2):

$$\cos(\theta_p) = \frac{v\tau_{pm}}{\Delta}, \tag{2.24}$$

where v is the speed of sound as it travels through air. Due to the far-field conditions it can be assumed that β_{pm} is constant for $m = 1, \ldots, M$, and $\forall p = 1, \ldots, P$.. Then, substituting for τ_{pm} in equation (2.16), the array response vector can now be expressed as:

$$\mathbf{a}(\theta_p) = a_1(\theta_p) \begin{bmatrix} 1 \\ e^{-j\omega 2\Delta \frac{\cos(\theta_p)}{v}} \\ \vdots \\ e^{-j(M)\omega\Delta \frac{\cos(\theta_p)}{v}} \end{bmatrix}. \tag{2.25}$$

Now, assuming unit response of the first microphone and substituting for $v = f\lambda$, the array response vector can be re-expressed as:

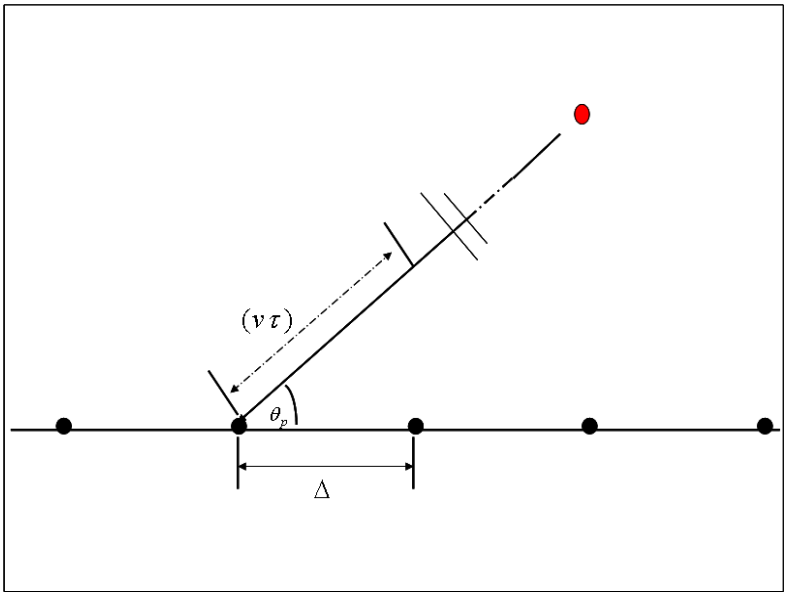


Figure 2.2: *The wavefronts arriving at a Uniform Linear Array (ULA) with inter-microphone spacing* Δ.

$$\mathbf{a}(\theta_p) = \begin{bmatrix} 1 \\ e^{-j(2)\frac{2\pi\Delta}{\lambda}\cos(\theta_p)} \\ \vdots \\ e^{-j(M)\frac{2\pi\Delta}{\lambda}\cos(\theta_p)} \end{bmatrix}. \tag{2.26}$$

If $\Delta > \frac{\lambda}{2}$ then spatial aliasing can occur, i.e. there exists θ_1, $\theta_2 \in [0, \pi]$ such that for $\theta_1 \neq \theta_2$:

$$e^{-j\frac{2\pi\Delta}{\lambda}\cos(\theta_1)} = e^{-j\frac{2\pi\Delta}{\lambda}\cos(\theta_2)}. \tag{2.27}$$

Therefore, in order to avoid this problem, the array must be designed so that:

$$\Delta \leq \frac{\lambda_{\min}}{2}, \tag{2.28}$$

where $\lambda_{\min}$ is the minimum wavelength of the received signal.

2.5 Probability Density Function (pdf) of the Observed Data

Due to the random nature of the observation vector the probability of correctly determining the required parameters will never be equal to 1. Instead the observed data can be characterised by the Probability Density Function (pdf) denoted $f_\theta(\mathbf{x})$. As the pdf is dependent on θ there is a different pdf associated with each different realisation of θ. The accuracy of the estimator is directly linked to that of the pdf, the correct choice of pdf will therefore greatly influence the performance of the estimator.

Estimation based solely on the pdf of the observed data is called classical estimation [4]. This is the estimation problem considered here and it is assumed throughout that no prior information about the parameters is available. Alternatively Bayesian estimation can be used if prior knowledge of the parameter values exists [4]. In this situation the parameter to be estimated is viewed as a realization of the random variable θ.

In this work deterministic estimation is considered, and for the sake of simplicity the case of estimating a single, i.e. $P = 1$, real, deterministic parameter is considered initially. These results are then extended to include estimation of multiple deterministic parameters.

In certain situations it may be more convenient to estimate a function of the required parameter rather than the parameter itself. Therefore in the following the general case of estimating a function $g(\theta)$ is considered, where $g(\theta)$ may or may not be equal to θ. Correspondingly the estimator used to estimate the parameter value based on the observations is denoted $\widehat{g(\theta)}(\mathbf{x})$.

For many practical estimation problems the pdf will not be known and must instead be approximated. A random variable $\mathbf{x}$ is said to be Gaussian, with distribution $N(m_\mathbf{x}, \sigma_\mathbf{x})$ if its pdf has a Normal distribution. Assuming a Gaussian distribution, and letting $P = 1$ the pdf of the received data, as given in equation (2.21), is expressed as:

$$f_\theta(\mathbf{x}) = \frac{1}{\sigma_\mathbf{x}\sqrt{2\pi}} e^{-\frac{1}{2}\left\| \frac{x - m_\mathbf{x}}{\sigma_\mathbf{x}} \right\|^2}, \tag{2.29}$$

where $m_\mathbf{x} = E[\mathbf{x}]$, and $\sigma_\mathbf{x}^2$ is the variance of $[\mathbf{x}]$.

A Gaussian distribution is commonly assumed due to the many convenient mathematical properties of the Gaussian distribution, and also because this assumption is often justifiable due to the Central Limit Theorem [5]. The central limit theorem states that the distribution of the mean tends to be Normal, regardless of the distribution from which the mean is computed.

If $x_1, x_2, \ldots, x_N$ are a set of independent random variables, with mean m and finite variance σ^2, then:

$$\frac{x_1 + \ldots + x_n + \ldots + x_N}{N} \xrightarrow[N\to\infty]{\text{converges to distribution}} \mathcal{N}\left(E\left[\mathbf{x}\right], \sqrt{\frac{\sigma_{\mathbf{x}}^2}{N}}\right), \tag{2.30}$$

Therefore if we have N independent observations of the same random experiment, $x_1, x_2, \ldots, x_N$, for sufficiently large N the estimator $\frac{1}{N}\sum \widehat{g\left(\theta\right)}\left(\mathbf{x}\right)$ will have a Gaussian distribution:

$$\mathcal{N}\left(E\left[\widehat{g\left(\theta\right)}\left(\mathbf{x}\right)\right], \sqrt{\frac{\sigma^2_{\widehat{g(\theta)}}}{N}}\right), \tag{2.31}$$

where $\sigma^2_{\widehat{g(\theta)}}$ is the variance of the estimator $\widehat{g\left(\theta\right)}$. We now need to find an estimator that will assign an estimate to each set of observations $\mathbf{x} = [x\left[1\right], x\left[2\right], \ldots x\left[N\right]]$. The quality of such an estimator must be quantified in order to see if it meets requirements, and ideally we wish to find the optimal estimator for a given situation. This naturally leads to the question of what measurements are needed to evaluate the estimator performance, and what criteria should be met in order for an estimator to be considered "optimal".

2.6 Estimator Performance Evaluation

When evaluating the performance of an estimator the basic question that must be addressed is "How accurate are the results provided by this estimator?", or put another way, "How close will the resulting estimates be to the true values?". The aim then is to select the best possible estimator, but in order to do this the term "best" must be quantified.

The first factor to be considered is whether or not the mean or expected value of the estimator is equal to the true parameter value, i.e. is:

$$E\left[\widehat{g\left(\theta\right)}\left(\mathbf{x}\right)\right] = g\left(\theta\right). \tag{2.32}$$

If this is true, then the estimator is said to be unbiased. If this is not the case, then the estimator is biased, with bias given by:

$$Bias\left(\widehat{g\left(\theta\right)}\left(\mathbf{x}\right)\right) = E\left[\widehat{g\left(\theta\right)}\left(\mathbf{x}\right)\right] - g\left(\theta\right), \tag{2.33}$$

A biased estimator is one that systematically introduces an error, resulting in the mean or expected value of the estimator no longer being equal to the true parameter value.

The effects of such an error therefore cannot be removed by averaging. While an unbiased estimator is not necessarily a good estimator, a biased estimator is generally highly undesirable [4].

The second measurement used to characterize the estimator performance is the variance, which measures the dispersion of the estimates around their expected value:

$$Var\left(\widehat{g(\theta)}(\mathbf{x})\right) = E\left\{\left(\widehat{g(\theta)}(\mathbf{x}) - E\left[\widehat{g(\theta)}(\mathbf{x})\right]\right)^2\right\}. \tag{2.34}$$

Clearly the more accurate the estimator, the smaller the variance will be [4].

The Mean Square Error (MSE) is commonly used as a measurement of the quality or precision of an estimator. Defining Ω as the observation space, the MSE can be expressed as:

$$\begin{aligned} MSE\left[\widehat{g(\theta)}(\mathbf{x})\right] &= E\left[\left(\widehat{g(\theta)}(\mathbf{x}) - g(\theta)\right)^2\right] && (2.35)\\ &= \int_{\Omega}\left(\widehat{g(\theta)}(\mathbf{x}) - g(\theta)\right)^2 f_\theta(\mathbf{x})\, d\mathbf{x} && (2.36)\\ &= \left\|\widehat{g(\theta)} - g(\theta)\right\|_\theta^2. && (2.37) \end{aligned}$$

From equation (2.35) we can see that the MSE can also be expressed as:

$$MSE\left[\widehat{g(\theta)}(\mathbf{x})\right] = Var\left[\widehat{g(\theta)}(\mathbf{x})\right] + Bias^2\left[\widehat{g(\theta)}(\mathbf{x})\right]. \tag{2.38}$$

Then for an unbiased estimator:

$$\begin{aligned} Bias\left(\widehat{g(\theta)}(\mathbf{x})\right) &= E\left[\widehat{g(\theta)}(\mathbf{x})\right] - g(\theta) = 0, && (2.39)\\ \Longrightarrow\ & MSE\left(\widehat{g(\theta)}(\mathbf{x})\right) = Var\left(\widehat{g(\theta)}(\mathbf{x})\right). && (2.40) \end{aligned}$$

A final consideration when evaluating the performance of an estimator is the complexity of the computations involved. While an estimator may provide highly accurate results, it will be of little practical use if it cannot be quickly and easily implemented. In such situations, an estimator with lower accuracy may in fact be preferable, particularly if real-time estimation of the parameters is necessary.

2.7 Finding the Optimal Estimator

Unfortunately there is no straightforward minimisation scheme that is guaranteed always to produce an expression for an optimal and realisable estimator, where we consider an

estimator to be realisable if its definition does not rely on unknown or unobservable quantities. There are however certain approaches that will, in some cases, produce a closed form expression for such an estimator. The most well known of these approaches are discussed here.

Using the MSE as a measurement of the estimator quality, we aim to find a realizable estimator with the minimum MSE. The following questions therefore arise "does such an estimator exist?" and if so, "how can such an estimator be found?".

In the following the general case of estimating a vector of unknown parameters is considered.

2.7.1 Minimum Mean Square Error (MMSE) Estimator

One approach [6] is to minimize the expression for the overall weighted MSE and find the estimator that achieves this, i.e the aim is to find the optimal estimator for all values of **theta**, called the optimal global estimator $\widehat{g(\theta)}(\mathbf{x})_{opt}^{glob}$, with the minimum MSE.

The expression for the weighted MSE is given by [6]:

$$MSE\left[\widehat{g(\theta)}(\mathbf{x}),\partial\right] = \int_{\Theta} MSE\left[\widehat{g(\theta)}(\mathbf{x})\right]\partial(\boldsymbol{\theta})\,d\boldsymbol{\theta} \tag{2.41}$$

$$= \int_{\Theta}\int_{\Omega}\left(\widehat{g(\theta)}(\mathbf{x}) - g(\boldsymbol{\theta})\right)^2 f_{\boldsymbol{\theta}}(\mathbf{x})\,\partial(\boldsymbol{\theta})\,d\mathbf{x}d\boldsymbol{\theta}, \tag{2.42}$$

where $\partial(\boldsymbol{\theta})$ is a strictly positive weighting function defined over Θ, such that $\int_{\Theta}\partial(\boldsymbol{\theta})\,d\boldsymbol{\theta}$=1. The weighted MSE is introduced here in order to express the minimization in a form similar to that of risk minimization in Bayesian estimation, allowing exploitation of results previously established in this area [7]. From equation (2.41) it can be seen that in this case $\partial(\boldsymbol{\theta})$ is equivalent to the prior in Bayesian estimation theory. The $MSE\left[\widehat{g(\theta)}(\mathbf{x}),\partial\right]$ can be re-expressed in certain cases as:

$$MSE\left[\widehat{g(\theta)}(\mathbf{x}),\partial\right] = \int_{\Omega}\int_{\Theta}\left(\widehat{g(\theta)}(\mathbf{x}) - g(\boldsymbol{\theta})\right)^2\partial(\boldsymbol{\theta})\,f_{\boldsymbol{\theta}}(\mathbf{x})\,d\boldsymbol{\theta}d\mathbf{x}. \tag{2.43}$$

Minimization of the MSE is therefore equivalent to minimization of the inner integral:

$$\int_{\Theta}\widehat{g(\theta)}(\mathbf{x})\,\partial(\boldsymbol{\theta})\,f_{\boldsymbol{\theta}}(\mathbf{x})\,d\boldsymbol{\theta} - \int_{\Theta} g(\boldsymbol{\theta})\,\partial(\boldsymbol{\theta})\,f_{\boldsymbol{\theta}}(\mathbf{x})\,d\boldsymbol{\theta} = 0, \tag{2.44}$$

Then setting:

$$\widehat{g(\theta)}(\mathbf{x}) = \widehat{g(\theta)}(\mathbf{x})_{opt}^{glob}, \tag{2.45}$$

this gives:

$$\widehat{g(\theta)}(\mathbf{x})_{opt}^{glob} = \frac{\int_{\Theta} g(\boldsymbol{\theta})\,\partial(\boldsymbol{\theta})\,f_{\boldsymbol{\theta}}(\mathbf{x})\,d\boldsymbol{\theta}}{\int_{\Theta}\partial(\boldsymbol{\theta})\,f_{\boldsymbol{\theta}}(\mathbf{x})\,d\boldsymbol{\theta}}, \tag{2.46}$$

where $\widehat{g(\theta)}(\mathbf{x})$ is a realisable estimator as it is independent of **theta**. Unfortunately however, calculation of this estimator is often impossible and there is no known method for determining a suitable $\partial(\boldsymbol{\theta})$, that will lead to the minimum MSE for all $\boldsymbol{\theta} \in \Theta$ [4].

As the strategy of minimizing the overall or global MSE does not necessarily produce a computable expression for the minimum MSE estimator, the MSE is minimized instead for a given parameter value $\boldsymbol{\theta}$, in order to find a realizable $\widehat{g(\theta)}(\mathbf{x})_{opt}^{loc}$, that is non-trivial $\left(\widehat{g(\theta)}(\mathbf{x})_{opt}^{loc} \neq g(\theta)\right)$. While this approach is simpler than the global approach, as $\widehat{g(\theta)}(\mathbf{x})_{opt}^{loc}$ is based on local optimization it is unlikely to produce a realizable estimator that is not "*clairvoyant*" i.e. that does not depend on the unknown parameter value [6]. Minimization of the local MSE under non-bias constraints is discussed in detail in chapter 6, when it is used for the calculation of the best possible performance of an estimator using minimal performance bounds. It can therefore be concluded that the procedure of minimizing the MSE is unlikely to produce a realizable estimator, and an alternative strategy must be adopted.

2.7.2 Minimum Variance Unbiased (MVU) Estimator

An alternative approach is to search for the unbiased estimator with the minimum variance for all possible values of $\boldsymbol{\theta}$. This Minimum Variance Unbiased (MVU) estimator does not always exist, as it is common for different estimators to have minimum variance depending on the value of $\boldsymbol{\theta}$.

The minimum possible variance that can be attained by an unbiased estimator is characterised by the Cramer-Rao Bound (CRB) [8]. Therefore, if an estimator exists whose variance equals the CRB for each value of $\boldsymbol{\theta}$, this must be the MVU, and taking variance as a measure of optimality, the CRB can then be used to find an expression for the optimal estimator. Any estimator that has variance equal to the CRB is called an efficient estimator.

In order to derive the CRB, the following regularity condition is assumed to be met:

$$E\left[\frac{\partial \ln f_{\boldsymbol{\theta}}(\mathbf{x})}{\partial\theta}\right] = 0 \qquad \forall\boldsymbol{\theta}, \tag{2.47}$$

and an efficient estimator exists if (and only if) the following factorization can be made:

$$\frac{\partial \ln f_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta} = \mathbf{F}(\boldsymbol{\theta})\left(h(\mathbf{x}) - g(\boldsymbol{\theta})\right), \tag{2.48}$$

where $\widehat{g(\boldsymbol{\theta})}^{MVU}(\mathbf{x}) = h(\mathbf{x})$.

The variance of an efficient estimator is equal to the Cramer Rao Bound (CRB) and can be expressed as follows (see chapter 6 for a full discussion on the derivation of the minimal variance of an estimator and estimator performance bounds.)

$$\mathbf{C}_{\widehat{g(\boldsymbol{\theta})}^{MVU}(\mathbf{x})} \geq -E\left[\frac{\partial^2 \ln f_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta \partial \theta^T}\right]^{-1}. \tag{2.49}$$

When a linear model can be used to describe the data this approach yields a closed form expression for the MVU estimator, and its variance is equal to the CRB. Using the classical general linear model [4], the observed data can be described as:

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{n} \tag{2.50}$$

where $\mathbf{x}$ is the $(N \times 1)$ observation vector and $\mathbf{H}$ is the known $(N \times P)$ transfer function matrix. In this case, $\mathbf{H}$ is a linear function of the $(P \times 1)$ vector of unknown parameters $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_P]^T$, and $\mathbf{n}$ is the $(N \times 1)$ noise vector with pdf $\mathcal{N}(0, \mathbf{C})$. The pdf of $\mathbf{x}$ is:

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{N}{2}}\sqrt{\det \mathbf{C}}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T \mathbf{C}^{-1}(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})\right], \tag{2.51}$$

and so:

$$\frac{\partial \ln f_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta} = \left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)\left(\widehat{g(\boldsymbol{\theta})}(\mathbf{x}) - g(\boldsymbol{\theta})\right), \tag{2.52}$$

resulting in the efficient and MVU estimator, which is the weighted Least Squares (LS) estimate:

$$\widehat{g(\boldsymbol{\theta})}^{MVU}(\mathbf{x}) = \left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x}, \tag{2.53}$$

with variance:

$$CRB\left(\widehat{g(\boldsymbol{\theta})}\right) = \left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}. \tag{2.54}$$

It can therefore be seen that in the case of a linear data model it is always possible to find a closed form expression for an efficient and therefore MVU estimator.

However, in many situations it will not be appropriate to use a linear model to describe the data, as arises when estimating the Direction of Arrival (DOA) of a source signal. This situation is now considered, using the data model defined in equation (2.22) where in this case $\mathbf{n}$ is White Gaussian Noise (WGN) with variance σ^2. The pdf of the observed data is given by:

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{x} - \mathbf{A}(\boldsymbol{\theta})\mathbf{s})^T(\mathbf{x} - \mathbf{A}(\boldsymbol{\theta})\mathbf{s})\right], \tag{2.55}$$

leading to:

$$\frac{\partial \ln f_{\boldsymbol{\theta}}(\mathbf{x})}{\partial\theta} = \frac{\mathbf{s}}{2\sigma^2}\left\{\left(\frac{\partial \mathbf{A}(\boldsymbol{\theta})}{\partial\theta}\right)^T(\mathbf{x} - \mathbf{A}(\boldsymbol{\theta})\mathbf{s}) + (\mathbf{x} - \mathbf{A}(\boldsymbol{\theta})\mathbf{s})^T\left(\frac{\partial \mathbf{A}(\boldsymbol{\theta})}{\partial\theta}\right)\right\}, \tag{2.56}$$

where:

$$\frac{\partial \mathbf{A}(\boldsymbol{\theta})}{\partial\theta} = \begin{bmatrix} \frac{\partial \mathbf{A}(\theta_1)}{\partial\theta} \\ \frac{\partial \mathbf{A}(\theta_2)}{\partial\theta} \\ \vdots \\ \frac{\partial \mathbf{A}(\theta_P)}{\partial\theta} \end{bmatrix} \tag{2.57}$$

In this situation, factorisation as shown in equation (2.48) in order to find an expression for the MVU estimator is not possible. In fact, for non-linear data models efficient estimators will only be found under asymptotic conditions (high SNR [9] and/or large number of snapshots [8]). This is due to the fact that asymptotically the pdf becomes more concentrated around the true parameter value $\boldsymbol{\theta}$, causing the estimates to lie in a smaller interval about $\boldsymbol{\theta}$. In this case the estimator is said to be consistent. The relationship is approximately linear in this region and observations rarely occur in the non-linear region, resulting in asymptotic efficiency. (For a detailed discussion on the performance of non-linear estimators, see chapter 6). Therefore unless operating under asymptotic conditions it will not be possible to find an efficient estimator for DOA estimation.

However, in situations where there is no efficient estimator, it is still possible that an MVU estimator exists. The MVU estimator, if one exists, can be determined by finding a minimal sufficient statistic to describe the observed data. A statistic is said to be sufficient if the conditional pdf of the observations after the sufficient statistic has been observed is independent of the parameter to be estimated [4]. A necessary statistic is one which can be computed from any sufficient statistic, without reference to the original data, and a minimal sufficient statistic is one which is both necessary and sufficient. This means

that a minimal sufficient statistic is just as informative as the original data, but it can be computed from any other sufficient statistic; no further compression of the data is possible, without losing some information. If a sufficient statistic is found, then the MVU estimator must be a function of this statistic.

Using the Neyman-Fisher factorization a sufficient statistic can be found directly from the pdf, which is assumed to be known [4]:

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = f(\mathbf{T}(\mathbf{x}), \boldsymbol{\theta})\, h(\mathbf{x}), \tag{2.58}$$

where if $\boldsymbol{\theta}$ is a $p-$dimensional vector, then the statistic $\mathbf{T}(\mathbf{x})$ is a $p-$dimensional function, f is a function depending only on $\mathbf{T}$ and $\boldsymbol{\theta}$, and h is a function depending only on the observations $\mathbf{x}$. The difficulty however, arises in situations where this factorization is not obvious, and in these cases it is possible that a sufficient statistic, other than the observed data itself, does not exist.

If a sufficient statistic does exist, the Rao-Blackwell-Lehmann-Scheffe (RBLS) theorem can be applied in order to determine the corresponding estimator. Firstly, a $p-$dimensional function f must be found such that:

$$E[f(\mathbf{T})] = \boldsymbol{\theta}. \tag{2.59}$$

Then if $f(\mathbf{T})$ is the only unbiased function of the sufficient statistic, the statistic is said to be complete, and the MVU estimator is given by:

$$\widehat{g(\boldsymbol{\theta})}(\mathbf{x}) = f(\mathbf{T}). \tag{2.60}$$

This approach can be seen to provide a means of finding an expression for the MVU estimator provided a minimal sufficient statistic exists. However, even in situations where such a statistic does exist, verification of its completeness can be very difficult.

It can therefore be seen that there is no guaranteed way to find an expression for an optimal estimator for the non-linear data models being considered here, and even if such an estimator exists it is not guaranteed to be realisable, as it may require knowledge of unknown parameters. Sub-optimal estimators must instead be considered in order to select a suitable estimator for the given estimation problem.

2.8 Conclusion

The use of array processing to estimate one or more real deterministic parameters was introduced in this chapter, and the steps involved in modelling the received data were

discussed. The use of the pdf of the data to estimate the desired parameters was then considered, and the criteria that can be used for estimator evaluation were developed and used to define an "optimal estimator".

The optimality criterion initially selected was the minimal MSE, however it was shown that minimization of the MSE does not provide an expression for an optimal estimator. The next optimality criterion considered was the variance, and it was seen that for some estimation problems the CRB can be used to find a direct expression for an efficient estimator, which is also the MVU estimator. However, for non-linear estimation problems such as DOA estimation, it is not possible to find an efficient estimator. The fact that no efficient estimator exists does not exclude the possibility that an MVU estimator exists, and use can be made of sufficient statistics in order to determine the MVU estimator in such a situation. However, determining the sufficient statistic involves factorization of the pdf, which may not be obvious, and in many cases a minimal sufficient statistic does not exist. Even if a minimal sufficient statistic is found, it must be checked for completeness in order to result in an expression for the MVU estimator, and verification of the completeness can be extremely difficult.

It is therefore clear that for the case of non-linear data models, as in the case of DOA estimation, an optimal estimator cannot be easily determined, and instead use must be made of sub-optimal estimators that have desirable characteristics for the problem under consideration.

3

Direction of Arrival (DOA) Estimation

3.1 Acoustic Source Localization

Using a common problem specification, this chapter provides a unified explanation of classical Direction of Arrival (DOA) estimation techniques. Moreover, the classical signal processing techniques, which in many cases were developed for narrowband signal processing, are treated in the context of localization of broadband sources. Recent developments in the application of these classical techniques to acoustical array localisation are also discussed.

3.2 Beamforming

The earliest development of spatial filtering or beamforming dates back to the Second World War, when the conventional beamformer was developed. The aim of this beamformer was to enhance the received signal by "steering" the array in the direction of the desired source. This beamformer is simply an application of Fourier-based spectral analysis to spatio-temporally sampled data [2, 10, 11].

The ability of the beamformer to enhance signals from a desired direction can also be applied to the problem of DOA estimation. The output of the antenna is steered

in each possible direction of interest, and the power of the array output is measured for each direction. The values of $\boldsymbol{\theta}$, resulting in the maximum array output power are then chosen as the DOA estimates. Narrowband beamformers assume that the incident signal has a narrow bandwidth, centred at a particular frequency. If the incident signal is broadband then it can be divided into narrow frequency bands, and a weighted average of the DOA estimates found for each of these bands is then found (see section 3.3.1 for further discussion on these techniques).

Using the data model introduced in the previous chapter the signal received by the array at time t is given by:

$$\mathbf{x}(t) = \mathbf{A}(\boldsymbol{\theta})\,\widetilde{\mathbf{s}}(t) + \mathbf{n}(t), \tag{3.1}$$

where $\mathbf{n}(t)$ is assumed to Gaussian noise. The steered output of the array is found by linearly combining the spatially sampled data received at each sensor, and can be expressed as:

$$y(t) = \mathbf{w}^H \mathbf{x}(t), \tag{3.2}$$

where $\mathbf{w}$ is the complex weighting vector, and acts as a spatial filter applied to the signal which results in one particular direction being emphasized. For the case where N snapshots of the signal are available the output power of the array is given by:

$$\begin{aligned} P(\boldsymbol{\theta}) &= \frac{1}{N}\sum_{t=1}^{N} |\mathbf{y}(t)|^2 && (3.3)\\ &= \frac{1}{N}\sum_{t=1}^{N} \mathbf{w}^H \mathbf{x}(t)\,\mathbf{x}^H(t)\,\mathbf{w} && (3.4)\\ &= \mathbf{w}^H \widehat{\mathbf{R}} \mathbf{w}, && (3.5) \end{aligned}$$

where $\widehat{\mathbf{R}}$ is the estimate of the spatial correlation matrix of the signal $\mathbf{x}(t)$:

$$\widehat{\mathbf{R}} = \frac{1}{N}\sum_{t=1}^{N} \mathbf{x}(t)\,\mathbf{x}^H(t) \tag{3.6}$$

The estimate of the DOA, $\widehat{\boldsymbol{\theta}}$, is then given by:

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \{P(\boldsymbol{\theta})\} \tag{3.7}$$

The type of beamformer used depends on the choice of weight vector $\mathbf{w}$, and there are two general categories of beamformers: data independent and statistically optimum [10].

As suggested by the name, the weights in a data independent beamformer are independent of the observations, and are chosen to produce a specified response regardless of the data received. On the other hand, the weights in a statistically optimum beamformer are chosen to optimize the array response, based on the statistics of the array data.

3.2.1 Delay and Sum Beamformer

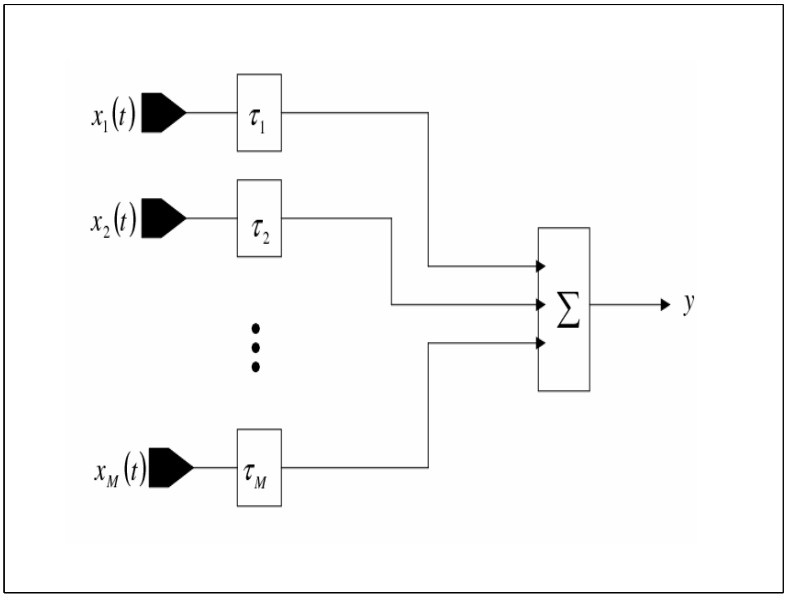


Figure 3.1: *A Delay and Sum beamformer.*

The Delay and Sum Beamformer (DSB) [10] is data independent, as it depends on the array geometry not on the received signal, and it is the simplest type of beamformer. Firstly the delay corresponding to a signal arriving from the direction θ_p is calculated for each of the microphones in the array, and the signal received by each of the microphones is then weighted by the appropriate delay. This results in constructive re-enforcement of the signal arriving from the direction θ_p while signals arriving from other directions are incoherently combined. The power of the array output corresponding to a signal arriving from direction θ_p is given by:

$$y(t) = \sum_{m=1}^{M} x_m(t + \tau_{pm}(\theta)), \tag{3.8}$$

$$P(\boldsymbol{\theta}) = \frac{1}{N}\sum_{t=1}^{N} |y(t)|^2 \tag{3.9}$$

$$= \frac{1}{N}\sum_{t=1}^{N} \left|\sum_{m=1}^{M} x_m(t + \tau_{pm})\right|^2, \tag{3.10}$$

where $\tau_{pm}(\theta)$ is defined in chapter 2 - section 2.2. The DOA estimates are then found by searching for the P maxima of $P(\boldsymbol{\theta})$.

An advantage of this beamforming approach is that it can be applied directly to broadband signals. However, the degree of resolution that can be achieved is strictly limited by the temporal sampling frequency of the data as delay differences less than the sampling rate cannot be resolved. Therefore in order to achieve higher resolution the sampling frequency must be increased resulting in an increased need for storage space and processing power.

3.2.2 Frequency Domain Beamforming

In order to increase the resolution that can be achieved without a corresponding increase in the sampling frequency required, beamforming can instead be performed in the frequency domain. Frequency domain beamforming is inherently narrowband, and therefore broadband signals are divided into narrowband frequency bins centered on f_c:

$$\mathbf{X}(f_c) = \mathbf{A}(f_c, \boldsymbol{\theta})\,\mathbf{S}(f_c) + \mathbf{N}(f_c). \tag{3.11}$$

Assuming once again that the array is a ULA, the array response function $\mathbf{A}(f_c, \boldsymbol{\theta})$ is a function of the incident angle only, and represents the response of the array to P complex exponentials at frequency f_c, which arrive at the array with angles of $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_p]$. The frequency domain representation of the steered output of the array is found by linear combination of these frequency components after applying the appropriate complex weights:

$$Y(f_c) = \mathbf{W}^H \mathbf{X}(f_c). \tag{3.12}$$

where $\mathbf{W}$ is the weighting vector and is chosen so that the signals from the direction under consideration, which is often called the look direction, are added coherently. Signals from

all other directions are attenuated. The output Power Spectral Density (PSD) is given by:

$$\Phi_{YY}(f_c) = Y(f_c) Y^*(f_c) \tag{3.13}$$
$$= \mathbf{W}^H(f_c) \widehat{\mathbf{R}}_{\mathbf{XX}} \mathbf{W}(f_c), \tag{3.14}$$

where $\widehat{\mathbf{R}}_{\mathbf{XX}}$ is defined in 3.6. The DOA estimates are then found for each frequency band $\widehat{\boldsymbol{\theta}}(f_c)$ by selecting the P maxima of the output PSD:

$$\widehat{\boldsymbol{\theta}}(f_c) = \arg\max_{\boldsymbol{\theta}} \{\Phi_{YY}(f_c)\} \tag{3.15}$$

In order to produce a non-trivial solution of equation (3.7) the weight vector is chosen so that $|\mathbf{W}| = 1$, resulting in the following weight vector for a given direction θ at a given frequency:

$$\mathbf{W}(f_c, \theta) = \frac{\mathbf{A}(f_c, \theta)}{\sqrt{\mathbf{A}^H(f_c, \theta) \mathbf{A}(f_c, \theta)}}. \tag{3.16}$$

Substitution of equation (3.16) into equation (3.14) produces the classical spatial spectrum:

$$\Phi_{BF}(f_c) = \frac{\mathbf{A}^H(f_c, \theta) \widehat{\mathbf{R}_{\mathbf{XX}}} \mathbf{A}(f_c, \theta)}{\mathbf{A}^H(f_c, \theta) \mathbf{A}(f_c, \theta)}. \tag{3.17}$$

The conventional beamformer is an extension of the classical Fourier based spectral analysis, and if the array in question is a ULA, then the resulting spatial spectrum in (3.17) can be viewed as a spatial domain version of the classical time-series domain periodogram. This similarity between the conventional beamformer and the time-series periodogram also extends to the resolution threshold experienced in the periodogram, and the maximum resolution that can be achieved for a ULA of M elements is:

$$\Delta = \left(\frac{2\pi}{M}\right) rads. \tag{3.18}$$

3.2.3 Beamforming and Acoustical Source Localization

While beamforming provides an optimal estimate of the source location, the need to calculate the beamformer output for every possible value of θ makes the method computationally prohibitive, particularly if high resolution localization is desired. Another difficulty arises due to the fact that originally the main motivations in the development of many beamforming applications were RADAR [12], and consequently such techniques

may perform poorly when applied to the reverberant situations encountered in acoustical source localization.

Several researchers have shown that the degradation of performance due to such effects can be reduced by the use of any *a priori* information that may be available. In [13] the nature (e.g. statistical non-stationarity, method of production, pitch, voicing, formant structure, and source radiator model) of the speech signal being localized is modelled by the "Dual Excitation Model", providing a specific parameterization model which improves upon the general spatial filtering approach. Information on the nature of the speech signal is also used in [14] to distinguish between real sources and virtual sources arising due to reverberation. *A priori* knowledge has also been used to reduce the computational load of beamformer localization. In [15] the fact that the characteristic wavelengths of speech are comparable to the dimensions of the space being searched is exploited, allowing for the implementation of a coarse-to-fine search criterion in both the spatial and frequency domain.

In [16] a beamformer-based source localization technique within a particle filtering framework is proposed. The use of particle filters avoids the need for a comprehensive search of the source location space, and therefore allows for a computationally efficient beamforming scheme.

The similarity between the DSB and a Bayesian formulation was recently demonstrated in [17]. It was shown that when considered from the point of view of maximizing the likelihood the Bayesian formulation and Beamforming have been shown to be equal except for an energy term weighting which will not effect the likelihood of localizing a stationary signal, making the two methods identical in this case [17].

3.3 Subspace-Based Techniques

Subspace based DOA estimation methods exploit the geometrical properties of the correlation matrix of the received signals. A narrowband signal model is assumed as the signal subspace will differ for the different frequency bands in a broadband signal [18, 19]. Broadband incident signals are therefore transformed into the frequency domain, and divided into narrowband frequency bins as described in the previous section. Operating in the frequency domain, and assuming spatially white, zero-mean Gaussian noise, the correlation matrix of the observed signal is given by:

$$\begin{aligned} \mathbf{R}(f_c) &= E\left\{\mathbf{X}(f_c)\,\mathbf{X}^H(f_c)\right\} && (3.19)\\ &= \mathbf{A}(f_c,\theta)\,\mathbf{R}_s(f_c)\,\mathbf{A}^H(f_c,\theta) + \sigma^2\mathbf{I}, && (3.20) \end{aligned}$$

where:

$$\mathbf{R}_s\left(f_c\right) = E\left\{\mathbf{S}\left(f_c\right)\mathbf{S}^H\left(f_c\right)\right\}, \tag{3.21}$$

and $\mathbf{R}_s$ is assumed to be full rank. Once the matrix $\mathbf{A}\left(f_c,\boldsymbol{\theta}\right)$ has full rank and assuming $P < M$, the matrix $\mathbf{A}\left(f_c,\boldsymbol{\theta}\right)\mathbf{R}_s\left(f_c\right)\mathbf{A}^H\left(f_c,\boldsymbol{\theta}\right)$ has rank P, where P is the number of sources present. Furthermore, every vector in the range space of $\mathbf{A}\left(f_c,\boldsymbol{\theta}\right)\mathbf{R}_s\left(f_c\right)\mathbf{A}^H\left(f_c,\boldsymbol{\theta}\right)$ is an eigenvector of $\mathbf{R}$, associated with eigenvalue λ. Consequently, using the eigendecomposition of the matrix, $\mathbf{R}$ can be re-expressed as:

$$\mathbf{R} = \sum_{m=1}^{M} \lambda_m \mathbf{e}_m \mathbf{e}_m^H. \tag{3.22}$$

Arranging the eigenvectors in order of the decreasing size of their associated eigenvalues, the signal and noise eigenvectors can then be separated:

$$\begin{aligned} \mathbf{R} = & \sum_{m=1}^{P} \lambda_m \mathbf{e}_m \mathbf{e}_m^H + \sum_{m=P+1}^{M} \lambda_m \mathbf{e}_m \mathbf{e}_m^H & (3.23)\\ & = \mathbf{E}_s \boldsymbol{\Lambda}_s \mathbf{E}_s^H + \sigma^2 \mathbf{E}_n \mathbf{E}_n^H, & (3.24) \end{aligned}$$

where $\mathbf{E}_s$ and $\mathbf{E}_n$ are matrices containing respectively the signal and noise eigenvectors:

$$\begin{aligned} \mathbf{E}_s &= \left[\mathbf{e}_1, \ldots, \mathbf{e}_P\right] & (3.25)\\ \mathbf{E}_n &= \left[\mathbf{e}_{P+1}, \ldots, \mathbf{e}_M\right], & (3.26) \end{aligned}$$

and $\boldsymbol{\Lambda}_s = diag\left[\lambda_1, \ldots \lambda_P\right]$ are the eigenvalues associated with the signal eigenvectors.

Any vector orthogonal to $\mathbf{A}\left(f_c,\boldsymbol{\theta}\right)$ is an eigenvector of $\mathbf{R}$ associated with an eigenvalue σ^2 [2]. Therefore $\mathbf{E}_n$ is orthogonal to $\mathbf{A}\left(f_c,\boldsymbol{\theta}\right)\mathbf{R}_s\left(f_c\right)\mathbf{A}^H\left(f_c,\boldsymbol{\theta}\right)$,
and as $\mathbf{A}\left(f_c,\boldsymbol{\theta}\right)\mathbf{R}_s\left(f_c\right)\mathbf{A}^H\left(f_c,\boldsymbol{\theta}\right)$ is full rank, it follows that [2]:

$$\begin{aligned} \mathbb{R}\left\{\mathbf{E}_s\right\} &= \mathbb{R}\left\{\mathbf{A}\left(f_c,\boldsymbol{\theta}\right)\right\} & (3.27)\\ \mathbb{R}\left\{\mathbf{E}_n\right\} &= \mathbb{R}\left\{\mathbf{E}_s\right\}^{\perp} = \mathbb{R}\left\{\mathbf{A}\left(f_c,\boldsymbol{\theta}\right)\right\}^{\perp} = \mathbb{N}\left\{\mathbf{A}^H\left(f_c,\boldsymbol{\theta}\right)\right\} & (3.28) \end{aligned}$$

where $\mathbb{R}\left\{\mathbf{E}_s\right\}$ is the subspace spanned by the range of $\mathbf{E}_s$, and $\mathbb{N}\left\{\mathbf{A}^H\left(f_c,\boldsymbol{\theta}\right)\right\}$ is the nullspace of $\mathbf{A}^H\left(f_c,\boldsymbol{\theta}\right)$. Therefore, if the signal subspace is the subspace spanned by $\mathbf{E}_s$, and the noise subspace is the subspace spanned by $\mathbf{E}_n$, then we can see that the signal and noise subspaces are orthogonal to each other. This relation is the basis for all subspace based estimation techniques.

In practice the matrix $\mathbf{R}\left(f_c\right)$ is unknown, and must be estimated from the observations:

$$\widehat{\mathbf{R}}\left(f_c\right) = \frac{1}{N}\sum_{t=1}^{N} \mathbf{X}\left(f_c\right)\mathbf{X}^H\left(f_c\right), \tag{3.29}$$

where the number of observations N must be large enough to ensure that $\widehat{\mathbf{R}}(f_c)$ is full rank (i.e. rank M). As $N \to \infty$, $\widehat{\mathbf{R}}(f_c) \to \mathbf{R}(f_c)$, however the fact that $\widehat{\mathbf{R}}(f_c)$ is estimated using a finite number of samples can lead to errors in detecting the number of sources present and estimating the associated angles of arrival.

3.3.1 Multiple Signal Classification (MUSIC) Algorithm

The best known of the subspace techniques is the Multiple Signal Classification (MUSIC) algorithm [20, 21]. As with other subspace methods, the basis of the MUSIC algorithm is that the noise eigenvectors are orthogonal to the steering vector or signal subspace.

From equation (3.28) the estimates of the DOAs are the values of $\boldsymbol{\theta}$ that minimize the projection of the steering vector, $\mathbf{A}(f_c, \boldsymbol{\theta})$, into the noise subspace.

The orthogonal projector onto the noise subspace is estimated as:

$$\widehat{\Pi}^{\perp} = \mathbf{E}_n \mathbf{E}_n^H, \tag{3.30}$$

and the MUSIC "spatial spectrum" is then defined as:

$$P_{MUSIC}(\boldsymbol{\theta}) = \frac{\mathbf{A}^H(f_c, \boldsymbol{\theta})\,\mathbf{A}(f_c, \boldsymbol{\theta})}{\mathbf{A}^H(f_c, \boldsymbol{\theta})\,\widehat{\Pi}^{\perp}\mathbf{A}(f_c, \boldsymbol{\theta})}. \tag{3.31}$$

In order for the DOA estimates to be unique, the array is assumed to be unambiguous, i.e. $\mathbf{A}(f_c, \boldsymbol{\theta}_1) \neq \mathbf{A}(f_c, \boldsymbol{\theta}_2)$ unless $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$, over the range of possible values of $\boldsymbol{\theta}$.

MUSIC offers a very large improvement in performance over traditional delay-and-sum beamforming techniques as the number of calculations necessary is reduced.

It has also been shown that assuming asymptotic conditions (high SNR [9] and/or large number of snapshots [8]) and uncorrelated signals, the MUSIC estimator is efficient, i.e. it reaches the Cramer Rao Bound (CRB), as long as M is larger than P [22, 23]. Therefore, unlike beamforming techniques, the MUSIC algorithm provides statistically consistent estimates, i.e. the estimates converge to the true value as the number of snapshots goes to infinity [2] [24].

However, like other spectral-based methods, MUSIC can show large bias in the case of a finite number of snapshots and low SNR, leading to an inability to resolve closely spaced sources. This is a serious difficulty when tracking a moving source, as the number of snapshots will be small in order to continuously update the source position. These resolution problems are even more serious when the received signals are highly correlated, with the algorithm failing to yield consistent estimates in the presence of coherent sources [22, 23]. Two sources are said to be coherent (or fully correlated) sources, if the signals emitted by the sources are identical, except possibly for a multiplicative constant factor [25].

In reverberant environments, highly correlated and even coherent signals may arise due to several replicas of the signal arriving by different paths. In the situation of coherent signals, the signal covariance matrix, $\mathbf{R}_s$, will be rank deficient. This rank deficiency causes the divergence of a signal eigenvector into the noise subspace, with the result that (3.28) no longer holds [2].

In the case of Uniform Linear Arrays (ULAs), a spatial smoothing method may be applied in order to deal with the problem of coherent (or highly correlated) signals. This method, introduced by [26] and later extended in [27–29], consists of splitting the original array into overlapping sub-arrays. Assuming that the steering vector of each sub-array is identical (up to a scaling), the sub-array covariance matrices can now be averaged. This spatial smoothing results in a random phase modulation, which in turn leads to decorrelation of the signals that were the cause of the rank deficiency.

While this spatial smoothing has been shown to allow for localization of coherent sources using subspace methods [29], these techniques perform particularly poorly when dealing with closely spaced sources [28]. Also, the fact that the use of spatial smoothing is limited to a very restricted class of array geometries [2, 28], limits its practical application.

There are two approaches to combining the results found from the individual narrowband frequency bins: Incoherent Wideband Processing [30]; and Coherent Wideband Processing [31], with both approaches offering higher accuracy than the original MUSIC algorithm when dealing with broadband sources.

In the incoherent case, the frequency range of the signal is divided into non-overlapping frequency bins. The narrowband MUSIC algorithm is then applied to each bin, and an incoherent average is taken over all the bins. This approach is very straightforward, and provided that all frequency bins have sufficiently high SNR, performs very well, particularly for sources with highly peaked spectra, where it outperforms the coherent methods [30, 32]. However, the selection of the frequency bins to be included is an important consideration, as outliers due to bins with low SNR reduce the source peaks and can also lead to spurious peaks, resulting in incorrect estimates.

In [33], Asano et al. proposed an eigenvalue weighting of the frequency bins. This method weights each frequency bin by the sum of the P largest eigenvalues of the covariance matrix, which correspond to the energy of the signal in the signal subspace. The average of the weighted MUSIC spectrum across all the frequency bins is then used to find the DOA estimate [33, 34].

While the implementation of coherent broadband processing is more complex than incoherent methods, the resulting accuracy is higher when the sources have relatively flat spectra [30, 32]. The aim of coherent processing is to translate the signal spaces for all

frequency bands into a common signal subspace.

One coherent approach, introduced in [18], is the use of focusing matrices. In this approach an initial DOA estimate is found by applying narrowband techniques at a selected frequency. This initial estimation step will be unable to resolve closely spaced sources, leading to the focusing step, where the information from all the frequency bands is used to focus in on the initial estimate. The frequency averaging of the covariance matrices increases the effective number of snapshots, and consequently the original narrowband estimate can be correctly resolved [18, 19]. The focusing matrix used can greatly influence the statistical characteristics of the DOA estimates [31], and different variations of the original coherent focusing matrix method have been proposed [19, 31, 35].

3.3.2 Root-MUSIC

The Root-MUSIC algorithm is a variation of the original MUSIC algorithm and can be applied in situations where a Uniform Linear Array (ULA) is used.

While the classical MUSIC algorithm searches for the points at which the steering vector is orthogonal to the noise subspace, in Root-MUSIC a polynomial is instead formed using the noise eigenvectors [36]. The DOA estimates are then determined from the roots of this polynomial.

The M-P polynomials corresponding to the noise eigenvectors can be defined as [37]:

$$D_k(z) = \sum_{m=1}^{M} e_{mk} z^{-(m-1)}, k = [P+1, \ldots, M,]. \tag{3.32}$$

where $\mathbf{e_{mk}}$ are the elements of $\mathbf{E_n}$.

The roots of each of these polynomials, $z_p = e^{-j(2)\frac{\pi\Delta}{\lambda}\cos(\theta_p)}$, are the signal zeros. As $z_p = |z_p| e^{j \arg(z_p)}$, it can be seen that:

$$arg(z_p) = (2)\frac{\pi\Delta}{\lambda}\cos(\theta_p), \quad \theta_p = cos^{-1}\left(\frac{\lambda}{2pi\Delta} arg(z_P)\right). \tag{3.33}$$

Define the polynomial:

$$Q(z) = \sum_{k=P+1}^{M} D_k(z) D_k^*\left(\frac{1}{z^*}\right). \tag{3.34}$$

The roots of $Q(z)$ are the same as those of $D_k(z)$. The null spectrum is obtained by evaluating $D_k(z)$ on the unit circle and so there are M double roots lying on the unit circle. These roots correspond to the actual incident signals, and other roots can be ignored.

Under asymptotic conditions this algorithm performs comparably to the original MUSIC algorithm. However, with Root-MUSIC higher resolution is also possible in situations of low SNR or where a limited number of snapshots is available. One of the main advantages of this approach is that it eliminates the need for a numerical search in order to find the maxima of the MUSIC spectrum.

3.3.3 Estimation of Signal Parameters via Rotational Invariance Technique (ESPRIT)

The Estimation of Signal Parameters via Rotational Invariance Technique (ESPRIT) algorithm is a subspace based technique that can be applied when the array configuration meets specific requirements, namely that it can be decomposed into two sub-arrays, such that each microphone in the first sub-array has a corresponding microphone in the second sub-array [38]. These microphone pairs are identical in every way, but are separated from each other physically by a known displacement vector $\mathbf{\Delta}$, of magnitude $\triangle$.

The received signal vectors for the sub-arrays can be expressed as:

$$\mathbf{x}_1(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}_1(t) \tag{3.35}$$

$$\mathbf{x}_2(t) = \mathbf{A}\Phi\mathbf{s}(t) + \mathbf{n}_2(t). \tag{3.36}$$

where Φ is the rotation operator relating the measurements of the sub-arrays, and is a diagonal matrix with elements $\Phi_{pp} = e^{-j\omega 2\Delta \frac{\mathrm{COS}(\theta_p)}{v}}$, for $p = 1, \ldots P$ and as before θ_p is the angle associated with the pth source.

The response matrix of the overall array, $\overline{\mathbf{A}}$ can be expressed as [2]:

$$\overline{\mathbf{A}} = \begin{bmatrix} \mathbf{A} \\ \mathbf{A}\Phi \end{bmatrix}, \tag{3.37}$$

and this can be exploited to obtain the estimates of the diagonal elements of Φ. This allows us to find the DOA estimates without knowing $\mathbf{A}$ itself. We can now define a unique, non-singular matrix $\mathbf{T}$, such that:

$$\mathbf{E}_s = \overline{\mathbf{A}}\mathbf{T}, \tag{3.38}$$

where $\mathbf{E}_s$ is defined in (3.25).

The invariance structure of the array allows for decomposition of $\mathbf{E}_s$ into $\mathbf{E}_{X1} \in \mathbb{C}^{M\times P}$, and $\mathbf{E}_{X2} \in \mathbb{C}^{M\times P}$:

$$\mathbf{E}_s = \begin{bmatrix} \mathbf{E}_{X1} \\ \mathbf{E}_{X2} \end{bmatrix} = \begin{bmatrix} \mathbf{AT} \\ \mathbf{A\Phi T} \end{bmatrix}. \tag{3.39}$$

Defining the unique matrix $\mathbf{F} \in \mathbb{C}^{2P \times P}$ with rank P, such that $\mathbf{F}$ spans the null-space of the matrix given by $[\mathbf{E}_{X1}|\mathbf{E}_{X2}]$, we can say that [38]:

$$\begin{aligned} \mathbf{0} &= [\mathbf{E}_{X1}|\mathbf{E}_{X2}]\,\mathbf{F} \\ &= \mathbf{E}_{X1}\mathbf{F}_1 + \mathbf{E}_{X2}\mathbf{F}_2 \qquad (3.40) \\ &= \mathbf{AT}\mathbf{F}_1 + \mathbf{A\Phi T}\mathbf{F}_2. \qquad (3.41) \end{aligned}$$

We now introduce the matrix $\mathbf{\Psi} \equiv -\mathbf{F}_1\mathbf{F}_2^{-1}$, which combined with (3.40) gives:

$$\mathbf{AT\Psi} = \mathbf{A\Phi T} \Rightarrow \mathbf{AT\Psi T}^{-1} = \mathbf{A\Phi}, \tag{3.42}$$

and then, as long as $\mathbf{A}$ is full rank, we can say:

$$\mathbf{T\Psi T}^{-1} = \Phi. \tag{3.43}$$

The eigenvalues of $\mathbf{\Psi}$, $[\lambda_1, \ldots \lambda_P]$ are equal to the diagonal elements of Φ, i.e. $\lambda_p = \varphi_p$. The signal parameters can then be obtained as nonlinear functions of these eigenvalues, and an estimate of the DOA is therefore given by:

$$\widehat{\theta}_p = \cos^{-1}\left(\frac{c}{2\pi f_0 \Delta}\arg\left\{\varphi_p\right\}\right). \tag{3.44}$$

In practical situations, the effects of additive noise and finite samples, means that it will not be possible to find a matrix $\mathbf{\Psi}$ that exactly meets the criteria $\mathbf{E}_{X1}\mathbf{\Psi} = \mathbf{E}_{X2}$. Instead an approximation is found, using either Least-Squares (LS-ESPRIT) or Total Least-Squares (TLS-ESPRIT) methods [2].

One of the main advantages of the ESPRIT algorithm is that it dramatically reduces the computational and storage costs when compared to the MUSIC algorithm, as the steering vector, $\mathbf{A}$, need not be known and the search for the maxima of the spectrum is eliminated. ESPRIT is also more robust to array imperfections than MUSIC, and produces unbiased estimates, even in situations where the estimates from MUSIC show some bias [38]. However, the variance of the ESPRIT estimates can be much larger than the MUSIC variance due to the fact that less information about the array geometry is used.

3.3.4 Time Reversal

In recent years, the use of acoustic time reversal mirrors and the DORT method (French acronym for decomposition of the time reversal operator) have been shown to be highly accurate methods of focusing a sound field on either a single, or else multiple targets (or scatterers) [39, 40]. The use of such techniques has been demonstrated in a wide variety of applications, for example, underwater acoustics [41], focusing a sound field in a reverberant room [42] and lithotripsy [43].

Iteration of the time reversal process has been shown to result in convergence on the most reflective of the targets (or the source with the highest amplitude) [44]. However, in situations where it is desirable to focus on less reflective targets, or on numerous targets simultaneously, the DORT technique, which is closely related to passive localization techniques such as MUSIC, can be used [45–47]. The DORT method is based on eigendecomposition of the time reversal operator (TRO), and the first step in deriving the TRO is to find the inter-element response matrix $\mathbf{K}(t)$. This matrix is measured by emitting a short pulse from each array element in turn and measuring the resulting response across the array. Then for transmit signal vector $\mathbf{e}(t)$, the output signal vector is $\mathbf{r}(t)$:

$$\mathbf{r}(t) = \mathbf{K}(t) \otimes \mathbf{e}_m(t) \tag{3.45}$$

where $\otimes$ is the Kronecker operator.

Transforming equation (3.45) into the frequency domain, we get:

$$\mathbf{r}(f) = \mathbf{K}(f)\,\mathbf{e}(f) \tag{3.46}$$

$\mathbf{r}(f)$, $\mathbf{K}(f)$ and $\mathbf{e}(f)$ are now evaluated at a single frequency, usually the center frequency of the transducers. For simplicity of notation, the $\mathbf{K}(f)$ will be denoted simply as $\mathbf{K}$.

This leads to the definition of the time reversal operator (TRO) [48]:

$$TRO = \mathbf{K}^H\mathbf{K}. \tag{3.47}$$

The TRO can therefore be considered as an estimate, $\widehat{\mathbf{R}}_K$, of the covariance matrix used in classical localisation techniques [48]. However, while these classical techniques such as the MUSIC technique assume statistically uncorrelated sources and require averaging of the measured data, the DORT method is active, (i.e. the elements of the array are used to both transmit and receive), and deterministic [49, 50].

If the scatterers are assumed to be point-like, then the number of significant eigenvalues of the TRO is equal to the number of scatterers present. The eigenvalue associated with a target is proportional to the reflectivity of the scatterer and its position relative to

the array [47]. The eigenvector $\mathbf{v}_i$, corresponding to the eigenvalue λ_i, gives the phase information necessary to focus on the ith target. One of the main advantages of the DORT method, is that, like all time-reversal processing techniques, it takes advantage of environmental multi-path in order to increase the effective SNR [49], as the presence of reverberation causes the effective array aperture to be increased.

The DORT method can be successfully applied in situations where the speed of sound changes gradually as the signal travels through the propagation medium. However, one of the basic assumptions underlying the time-reversal process is that the medium is lossless, and the presence of absorption will lead to degradation of the resolution that can be achieved. This has led to the introduction of amplitude compensation techniques [49], which increase the effective SNR and consequently the resolution that can be achieved.

In [48] it was shown that the TRO can be regarded as a pseudo-covariance matrix, and by using the TRO instead of the covariance matrix found in passive localization techniques, methods such as MUSIC and ML can be applied to localization using an active array. It was also shown when these techniques were used with the TRO, the resolution achieved was higher than that achieved using the DORT method.

3.4 Maximum Likelihood (ML) Estimation

Parametric methods exploit the underlying data model of the received signal and offer an alternative approach to the source localization problem. The Maximum Likelihood (ML) technique is the best known and most frequently used of these approaches.

The principle of the ML approach is very simple, a search is performed across all the possible DOAs to find the parameter, $\boldsymbol{\theta}$, that maximizes the likelihood function $f_{\mathbf{x}}(\boldsymbol{\theta})$ where $\mathbf{x}$ is the observation vector. This effectively means we are searching to find the parameter values that make the probability of the observations as large as possible. Since $f_{\mathbf{x}}(\boldsymbol{\theta})$ depends on the observations, the ML is therefore a function of the observations. The value of $f_{\mathbf{x}=\mathbf{x}_0}(\boldsymbol{\theta})d\mathbf{x}$ gives the probability of observing $\mathbf{x}$, in the region of volume $d\mathbf{x}$ centered around $\mathbf{x}_0$, as a function of $\boldsymbol{\theta}$.

Using the signal model given in equation (3.11) and assuming the signal is received in the presence of additive white complex Gaussian noise, the observation PDF at a given frequency f_c can be written as:

$$f_{\mathbf{x}}(\boldsymbol{\eta}) = \frac{1}{\left(\pi\sigma^2\right)^M} e^{-\frac{\|\mathbf{X}(f_c)-\mathbf{A}(f_c,\boldsymbol{\theta})\mathbf{S}(f_c)\|^2}{\sigma^2}}, \tag{3.48}$$

where $\boldsymbol{\eta}$ is the vector of unknown signal parameters, and $\boldsymbol{\eta} = \begin{bmatrix}\boldsymbol{\theta} & \mathbf{S} & \sigma^2\end{bmatrix}^T$.

As it is more convenient to find the ML estimate as a minimizing argument of the negative log-likelihood function, the following expression is used:

$$\Lambda\left(\mathbf{Y};\boldsymbol{\eta}\right) \triangleq -\ln\left\{f_{\mathbf{x}}\left(\mathbf{Y};\boldsymbol{\eta}\right)\right\}, \tag{3.49}$$

$$\Lambda\left(\mathbf{Y};\boldsymbol{\eta}\right) = M\ln\pi + M\ln\sigma^2 + \frac{1}{\sigma^2}\left\|\mathbf{X}\left(f_c\right) - \mathbf{A}\left(f_c,\boldsymbol{\theta}\right)\mathbf{S}\left(f_c\right)\right\|^2. \tag{3.50}$$

The ML estimate of the parameters is then given by:

$$\widehat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\theta},\mathbf{S},\sigma^2}\left\{M\ln\sigma^2 + \frac{1}{\sigma^2}\left\|\mathbf{X}\left(f_c\right) - \mathbf{A}\left(f_c,\boldsymbol{\theta}\right)\mathbf{S}\left(f_c\right)\right\|^2\right\}. \tag{3.51}$$

An estimate of σ^2 can be found using:

$$\widehat{\sigma}^2 = \frac{1}{M}\left(\left\|\mathbf{X}\left(f_c\right) - \mathbf{A}\left(f_c,\widehat{\boldsymbol{\theta}}\right)\widehat{\mathbf{S}}\left(f_c\right)\right\|^2\right), \tag{3.52}$$

which involves the non-linear least squares problem of minimising the following term in $\boldsymbol{\theta}$ and $\mathbf{S}$:

$$\left\|\mathbf{X}\left(f_c\right) - \mathbf{A}\left(f_c,\boldsymbol{\theta}\right)\mathbf{S}\left(f_c\right)\right\|^2, \tag{3.53}$$

This criterion is quadratic in relation to $\mathbf{S}$, which can be estimated by:

$$\mathbf{S} = \mathbf{A}^{\dagger}\left(f_c,\boldsymbol{\theta}\right)\mathbf{X}\left(f_c\right). \tag{3.54}$$

where: $\mathbf{A}^{\dagger}\left(\boldsymbol{\theta}\right)$ is the pseudo-inverse of $\mathbf{A}$.

$$\mathbf{A}^{\dagger}\left(f_c,\boldsymbol{\theta}\right) = \left(\mathbf{A}^H\left(f_c,\boldsymbol{\theta}\right)\mathbf{A}\left(f_c,\boldsymbol{\theta}\right)\right)^{-1}\mathbf{A}\left(f_c,\boldsymbol{\theta}\right)^H. \tag{3.55}$$

Then, defining:

$$\mathbf{P}_{\mathbf{A}} = \mathbf{A}\left(f_c,\boldsymbol{\theta}\right)\mathbf{A}^{\dagger}\left(f_c,\boldsymbol{\theta}\right), \tag{3.56}$$

$$\mathbf{P}_{\mathbf{A}}^{\perp} = \mathbf{I} - \mathbf{P}_{\mathbf{A}}, \tag{3.57}$$

the DOA estimate is given by:

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} Tr\left\{\mathbf{P}_{\mathbf{A}}^{\perp}\widehat{\mathbf{R}}\right\} \tag{3.58}$$

where:

$$\widehat{\mathbf{R}} = \mathbf{X}\left(f_c\right)\mathbf{X}^H\left(f_c\right). \tag{3.59}$$

One of the main advantages of the ML estimator is the fact that it is asymptotically efficient, where asymptotic can mean high SNR [9] and/or large number of snapshots [8, p. 500]. However, for practical localization schemes the number of snapshots available is finite, and therefore in order to determine the conditions necessary for the ML estimator to become unbiased and attain minimum variance the effect of the SNR is instead considered [51, 52]. For the case of using the MLE to obtain a time delay estimation, it is shown that the SNR necessary in order to attain the CRB is much larger than the kurtosis of the expected signal's energy spectrum [51]. In [52], it is also shown that for localization of a source in range and depth in a shallow water waveguide, once the Signal to Additive Noise Ratio (SANR) drops below 0dB the variance of the localization estimates obtained by the ML estimator will be greatly underestimated by the CRB.

In order to calculate the deterministic ML estimates the non-linear P-dimensional optimization problem in (3.58) must be solved. In order for the problem to converge rapidly to the true solution the choice of initial value used is very important. If this value is reasonably accurate, it is usually possible to find the true minimum quite rapidly using, for example, a Gauss-Newton technique [53]. On the other hand though, if the initial guess is not sufficiently accurate the optimization step may converge on a local minimum and the desired global minimum may never be found.

In order to obtain a suitable initial DOA estimate, one of the computationally efficient spectral-based methods discussed previously may be used. Choice of the method used, will however, determine the type of situation that the overall estimation method can deal with, e.g. if MUSIC is used to provide an initial estimate, then coherent signals will once again cause problems. In source tracking, the estimate of the previous frame may be used as the initial estimate of the next frame [54].

An alternative to using a multi-dimensional search is use of an alternating projection method [11, 55]. This approach, which is discussed in detail in chapter 5, breaks down the multidimensional parameter search into a sequence of signal source parameter searches, resulting in fast convergence rates. The resulting localization method has also been shown to demonstrate an improvement on previously proposed methods when applied to the case of unknown or inaccurate sensor locations [11, 56].

While these optimization techniques increase the speed of the ML algorithm, the computational time needed to search through each possible angle of arrival makes ML estimation unsuited to most practical localization schemes. This is especially true when operating in the frequency domain as the ML estimate must be found individually for each frequency.

3.4.1 Expectation Maximization (EM) Algorithm

The Expectation Maximization (EM) algorithm is a general method for solving ML estimation problems. It was proposed as a source localisation method in [57] and [58], with the objective of simplifying the computations involved in ML localization. This reduction is achieved by decoupling the multi-dimensional search associated with the direct ML approach into searches in smaller-dimensional parameter subspaces, resulting in considerable simplification of the computations involved [57–59].

The basis of the algorithm is to decompose the observed data into their signal components and then individually estimate the parameters of each of these signal components. These estimates are then used to improve the decomposition of the observations, and the process is iterated until the algorithm converges to a stationary point of the likelihood function, with each iteration improving the likelihood of the estimated parameters.

While the EM algorithm was originally developed for narrowband signals, extension to broadband is performed by transforming the signals into the frequency domain as described for the other narrowband algorithms [60, 61]. The received signal vector at time t and frequency f_c, $\mathbf{x}(f_c,t) = [\mathbf{X}_1(f_c,t),\ldots,\mathbf{X}_M(f_c,t)]^T$, is then modelled as before:

$$\mathbf{x}(f_c,t) = \mathbf{A}(f_c,\boldsymbol{\theta})\,\mathbf{s}(f_c,t) + \mathbf{n}(f_c,t), \tag{3.60}$$

where:

$$\mathbf{s}(f_c,t) = [S_1(f_c,t),\ldots,S_P(f_c,t)]^T \tag{3.61}$$

$$\mathbf{n}(f_c,t) = [\mathbf{N}_1(f_c,t),\ldots,\mathbf{N}_M(f_c,t)]^T. \tag{3.62}$$

Firstly, the received signal vector is decomposed into vectors corresponding to each sound source $\mathbf{y}_p(f_c,t)$ as:

$$\mathbf{x}(f_c,t) = \sum_{p=1}^{P}\mathbf{y}_p(f_c,t) = \mathbf{H}\mathbf{y}(f_c,t), \tag{3.63}$$

where $\mathbf{y}_p(f_c,t) = \mathbf{a}(\theta_p)\,S(f_c,t) + \mathbf{n}_p(f_c,t)$, and $\mathbf{y}(f_c,t) = \left[\mathbf{y}_1(f_c,t)^T,\ldots,\mathbf{y}_P(f_c,t)^T\right]^T$, and $\mathbf{H} = [\mathbf{I},\ldots,\mathbf{I}]$. The decomposition of the noise vector $\mathbf{n}_p(f,t)$ is an arbitrary decomposition, satisfying $\sum_{p=1}^{P}\mathbf{n}_p(f_c,t) = \mathbf{n}(f_c,t)$, and $E\left[\mathbf{n}_p(f_c,t)\,\mathbf{n}_p^H(f_c,t)\right] = \frac{\sigma^2}{P}\mathbf{I_M}$. The complete data set is then defined as the set of decomposed received signal vectors:

$$\mathbf{Y}(f_c) = [\mathbf{y}(f_c,1),\ldots,\mathbf{y}(f_c,N)]. \tag{3.64}$$

The term "'incomplete data"' is used to describe the actual snapshots $\mathbf{x}(f,t)$.

The log-likelihood of the complete data can now be expressed as:

$$L_c\left(\boldsymbol{\theta}, \mathbf{K}_s; \mathbf{Y}\right) = \sum_{p=1}^{P}\left[-N \log\det \mathbf{K}_{yp} - \sum_{t=1}^{N} \mathbf{y}_p^H\left(f_c, t\right) \mathbf{K}_{yp}^{-1} \mathbf{y}_p\left(f, t\right)\right], \tag{3.65}$$

where $\mathbf{K}_{yp}$ denotes the true covariance of the observations:

$$\mathbf{K}_{yp} = \gamma_p \mathbf{a}\left(\theta_p\right) \mathbf{a}^H\left(\theta_p\right) + \frac{\sigma^2}{P}\mathbf{I}_{\mathbf{M}}, \tag{3.66}$$

and the sample covariance of the observations is defined as:

$$\mathbf{C}_{yp} = \frac{1}{N}\sum_{t=1}^{N} \mathbf{y}_p\left(f_c, t\right) \mathbf{y}_p^H\left(f_c, t\right) \tag{3.67}$$

The EM algorithm finds the conditional expectation of $\mathbf{C}_{yp}$ in the E-step, and then uses this to maximize the log likelihood in the M-step. The algorithm is implemented as follows:

E-step:

$$\begin{aligned} \mathbf{C}_{yp}^q &\equiv E\left[\mathbf{C}_{yp} | \mathbf{C}_x; \widehat{\mathbf{K}}_y^q\right] && (3.68)\\ &= \widehat{\mathbf{K}}_{yp}^q - \widehat{\mathbf{K}}_{yp}^q \left(\widehat{\mathbf{K}}_x^q\right)^{-1} \widehat{\mathbf{K}}_{yp}^q + \widehat{\mathbf{K}}_{yp}^q \left(\widehat{\mathbf{K}}_x^q\right)^{-1} \mathbf{C}_x \left(\widehat{\mathbf{K}}_x^q\right)^{-1} \widehat{\mathbf{K}}_{yp}^q, && (3.69)\end{aligned}$$

where the superscript q denotes the iteration number, and $\widehat{\mathbf{K}}_x^q$ and $\widehat{\mathbf{K}}_{yp}^q$ are the model covariance estimations at iteration q, defined as:

$$\begin{aligned} \widehat{\mathbf{K}}_x^q &= \sum_{p=1}^{P} \widehat{\mathbf{K}}_{yp}^q && (3.70)\\ \widehat{\mathbf{K}}_{yp}^q &= \widehat{\gamma}_p^q \mathbf{a}\left(\widehat{\theta}_p^q\right) \mathbf{a}^H\left(\widehat{\theta}_p^q\right) + \frac{\sigma^2}{P}\mathbf{I}_{\mathbf{M}}. && (3.71)\end{aligned}$$

M-step:

$$\begin{aligned} \widehat{\theta}_p^{q+1} &= \arg\max_{\theta p} \frac{\mathbf{a}^H\left(\theta_p\right) \mathbf{C}_{\mathbf{y}p}^q \mathbf{a}\left(\theta_p\right)}{\left|\mathbf{a}\left(\theta_p\right)\right|^2} && (3.72)\\ & && (3.73)\\ \widehat{\gamma}_p^{q+1} &= \frac{\mathbf{a}^H\left(\widehat{\theta}_p^{q+1}\right) \mathbf{C}_{\mathbf{y}p}^q \mathbf{a}\left(\widehat{\theta}_p^{q+1}\right)}{\left|\mathbf{a}\left(\widehat{\theta}_p^{q+1}\right)\right|^4}. && (3.74)\end{aligned}$$

As the EM algorithm is based on the ML estimator, it also has the attractive properties of consistency, asymptotic unbiasedness and asymptotic minimum variance. It has been

shown to provide accurate estimates in situations of small sample size, and/or low SNR where other localization methods fail [58]. It has also been shown to be an accurate method for estimation of the azimuth, elevation and range, of near-field unknown but deterministic sources [59].

While the EM algorithm offers a clear improvement over the ML estimator in computational speed, the simultaneous update of all parameters can lead to slow convergence [62]. Difficulties may also arise with the maximization step of the EM algorithm when smoothness penalties, which may be necessary with a very large parameter space, are used.

In [62], the Space-Alternating Generalized Expectation-Maximization (SAGE) Algorithm is proposed, and an extension to the case of broadband signals is given in [63]. SAGE is a variant of the EM algorithm that involves sequential update of the parameters thereby increasing the rate of asymptotic convergence. In order to update the parameters sequentially, the algorithm alternates between several hidden-data spaces, updating a subset of the parameter vector each time. The use of a separate hidden-data space for each parameter, in this manner, automatically decouples the parameter updates, resulting in a simpler maximization step, and eliminating problems when using smoothness penalties.

The SAGE algorithm has been shown numerically to result in improved convergence rates when compared to the original EM algorithm [64, 65]. However, it has also been demonstrated that SAGE may become unstable for low SNR and small number of snapshots. The sensitivity of the SAGE algorithm to initial estimates is also greater than the sensitivity of the EM algorithm when estimating the direction of arrival of closely located sources [64, 65].

3.5 Subspace Fitting Techniques

In recent years subspace fitting techniques have been developed, with the aim of combining the low computational complexity of subspace based techniques with the statistical properties of the parametric approach [53, 66–68]. Once again these techniques can be implemented in the frequency domain in the case of broadband signals, and as with all subspace techniques they are based on the structure of the covariance matrix $\mathbf{R}$ [2]:

$$\mathbf{R} = \mathbf{A}\left(f_c, \boldsymbol{\theta}\right)\mathbf{R}_s\mathbf{A}^H\left(f_c, \boldsymbol{\theta}\right) + \sigma^2\mathbf{I}. \tag{3.75}$$

Then, as discussed in section 3.3, if there are P sources present, $\mathbf{R}_s$ has rank P, and $\mathbf{E}_s$, defined in equation (3.25), spans the P-dimensional signal subspace. Therefore, the $(P \times P)$ matrix T must exist such that:

$$\mathbf{E}_s = \mathbf{A}\left(f_c, \boldsymbol{\theta}\right)\mathbf{T}. \tag{3.76}$$

This equation forms the basis of the subspace fitting techniques. In practice $\mathbf{R}$ is unknown and must be estimated, and the eigendecomposition of the estimated covariance matrix $\widehat{\mathbf{R}}$ is given by [69, 70]:

$$\widehat{\mathbf{R}} = \widehat{\mathbf{E}}_s \boldsymbol{\Lambda}_s \widehat{\mathbf{E}}_s^H + \sigma_n^2 \widehat{\mathbf{E}}_n \widehat{\mathbf{E}}_n^H. \tag{3.77}$$

Because of this estimation, there is no value of θ which will result in an exact solution to $\widehat{\mathbf{E}}_s = \mathbf{A}\left(f_c, \theta\right)\mathbf{T}$. Instead, the estimates of θ and $\mathbf{T}$ are given by finding the least squares solution of the following:

$$\begin{aligned} \widehat{\theta}, \widehat{\mathbf{T}} &= \arg\min_{\theta, \mathbf{T}} \left\| \widehat{\mathbf{E}}_s - \mathbf{A}\left(f_c, \boldsymbol{\theta}\right)\mathbf{T} \right\|_{\mathbf{W}}^2 && (3.78) \\ &= \arg\min_{\theta, \mathbf{T}} \left\{ \left[\widehat{\mathbf{E}}_s - \mathbf{A}\left(f_c, \boldsymbol{\theta}\right)\mathbf{T} \right] \mathbf{W} \left[\widehat{\mathbf{E}}_s - \mathbf{A}\left(f_c, \boldsymbol{\theta}\right)\mathbf{T} \right]^H \right\}, && (3.79) \end{aligned}$$

where $\mathbf{W}$ is a positive, semi-definite weighting matrix, with optimal value [67, 68]:

$$\mathbf{W}_{opt} = \left(\widehat{\boldsymbol{\Lambda}}_s - \widehat{\sigma}_n^2 \mathbf{I} \right)^2 \widehat{\boldsymbol{\Lambda}}_s^{-1}, \tag{3.80}$$

and $\widehat{\sigma}_n^2$ is a consistent estimate of the noise variance.

From (3.78), it can be seen that the solution of the least squares problem results in the following estimated values of $\mathbf{T}$ and $\boldsymbol{\theta}$:

$$\begin{aligned} \widehat{\mathbf{T}} &= \mathbf{A}^{\dagger}\left(f_c, \boldsymbol{\theta}\right)\widehat{\mathbf{E}}_s && (3.81) \\ \widehat{\boldsymbol{\theta}}^{SSF} &= \arg\min_{\boldsymbol{\theta}} \mathbf{V}_{SSF}\left(\boldsymbol{\theta}\right), && (3.82) \end{aligned}$$

where:

$$\begin{aligned} \mathbf{V}_{SSF}\left(\boldsymbol{\theta}\right) &= \left\| \widehat{\mathbf{E}}_s - \mathbf{A}\left(f_c, \boldsymbol{\theta}\right)\mathbf{A}^{\dagger}\left(f_c, \boldsymbol{\theta}\right)\widehat{\mathbf{E}}_s \right\|_{\mathbf{W}}^2 && (3.83) \\ &= \left\| \mathbf{P}_{\mathbf{A}}^{\perp}\left(f_c, \boldsymbol{\theta}\right)\widehat{\mathbf{E}}_s \right\| && (3.84) \\ &= Tr\left\{ \mathbf{P}_{\mathbf{A}}^{\perp}\left(f_c, \boldsymbol{\theta}\right)\widehat{\mathbf{E}}_s \mathbf{W} \widehat{\mathbf{E}}_s^H \right\}. && (3.85) \end{aligned}$$

The SSF (or Weighted Subspace Fitting (WSF)) estimate of $\boldsymbol{\theta}$ is therefore the vector of DOA values that results in the subspaces of $\widehat{\mathbf{E}}_s$ and $\mathbf{A}\left(f_c, \boldsymbol{\theta}\right)$ being as close as possible.

The second of the subspace fitting techniques is called Noise Subspace Fitting (NSF), and is based on the MUSIC relation given in equation (3.28). The estimate of $\boldsymbol{\theta}$ is then given by:

$$\widehat{\boldsymbol{\theta}}^{NSF} = \arg\min_{\boldsymbol{\theta}} \mathbf{V}_{NSF}(\boldsymbol{\theta}), \tag{3.86}$$

where:

$$\begin{aligned}
\mathbf{V}_{NSF}(\boldsymbol{\theta}) &= \left\| \widehat{\mathbf{E}}_n^H \mathbf{A}(f_c, \boldsymbol{\theta}) \right\|_{\mathbf{U}}^2 && (3.87)\\
&= Tr\left\{ \widehat{\mathbf{E}}_n^H \mathbf{A}(f_c, \boldsymbol{\theta}) \mathbf{U} \mathbf{A}^H(f_c, \boldsymbol{\theta}) \widehat{\mathbf{E}}_n \right\} && (3.88)\\
&= Tr\left\{ \mathbf{U} \mathbf{A}^H(f_c, \boldsymbol{\theta}) \widehat{\mathbf{E}}_n \widehat{\mathbf{E}}_n^H \mathbf{A}(f_c, \boldsymbol{\theta}) \right\}, && (3.89)
\end{aligned}$$

and in this case $\mathbf{U}$ is the weighting matrix. As in the case of the MUSIC technique, this method is not accurate in the presence of coherent signals.

In the special case of a ULA, the structure of the array can once again be used to reduce the necessary computations when using SSF techniques. This approach is similar to the Root-MUSIC algorithm; instead of parameterizing $\mathbf{A}$, the null-space of $\mathbf{A}^H$ is parameterized, thereby once again eliminating the need for a numerical search.

3.6 Time Delay Estimation (TDE)

Time Delay Estimation (TDE) is an indirect approach to finding the location of a source, based on a two stage algorithm. Firstly, the time delay associated with a signal arriving at a pair of microphones is estimated. The delay estimates are converted to distances and a Least Squares (LS) fit is then usually applied to find the source location. In this approach, the time delays for different sensor pairs are estimated directly, making a search over the entire source location space unnecessary.

TDE techniques are computationally far less demanding than beamforming or parametric techniques, a big advantage when designing practical localization systems. However, the resolution that can be achieved using TDE techniques is severely limited by the temporal sampling frequency and the inter-microphone spacing. The advantages and disadvantages of these techniques are discussed in full in chapter 5.

The most common method for estimating the Time Difference of Arrival (TDOA) between a signal arriving arriving at two microphones, is by use of correlation-based techniques, in particular the General Cross Correlation (GCC) method [71]. If $x_1(t)$ and $x_2(t)$ are the signals received by microphones 1 and 2 of a microphone pair due to a source p emitting a signal $s_p(t)$ then the received signals can be expressed as:

$$x_1(t) = s_p(t) + n_1(t) \tag{3.90}$$
$$x_2(t) = \alpha s_p(t + \delta\tau_p) + n_2(t) \tag{3.91}$$

where: $\delta\tau_p$ is the TDOA, $n_1(t), n_2(t)$ are the noise signals received by sensors 1 and 2 respectively, and are assumed to be jointly stationary, uncorrelated and also uncorrelated with $s(t)$.

Defining $X_1(\omega)$ and $X_2(\omega)$ as the frequency transformations of $x_1(t)$ and $x_2(t)$ respectively, the cross-correlation of the received signals $R_{x_1,x_2}(\delta\tau)$ is calculated as the inverse Fourier transform of the received signal cross spectrum $X_1(\omega)$ and $X_2^*(\omega)$, scaled by a weighting function, $W(\omega)$. The weighting function is selected to compensate for signal degradation arising from background noise or reverberation. The time delay estimate, $\widehat{\delta\tau_p}$ is given by the lag value that maximises the cross-correlation between the $x_1(t)$ and $x_2(t)$

$$R_{x_1,x_2}(\delta\tau_p) = F^{-1}\left\{W(\omega) X_1(\omega) X_2^*(\omega)\right\}, \tag{3.92}$$

where the weighting function $W(\omega)$ is selected in order to minimize performance degradation due to background noise or reverberation depending on the environment in question. The time-delay estimate is then the value that maximises the correlation:

$$\widehat{\delta\tau_p} = \arg max_{\delta\tau_p}\left[R_{x_1,x_2}(\delta\tau_p)\right], \tag{3.93}$$

and the corresponding DOA estimate is given by:

$$\widehat{\theta}_{TDE} = \cos^{-1}\left\{\frac{v\widehat{\delta\tau_p}}{f_s\Delta}\right\}. \tag{3.94}$$

From equation (3.94) it can be seen that interpretation of the time delay as a DOA estimate is reliant on knowledge of the propagation speed of the signal. While the propagation speed is usually assumed to be known [72] this is not always the case, and in recent years, methods have been proposed for estimating the Time Delay of Arrival (TDOA) when the propagation speed is unknown [73].

As discussed in section 3.2.1 one of the main disadvantages of this approach is the severe limitation on the resolution achievable due to the fact that delay differences less than the temporal sampling rate cannot be resolved. This quantization effect is especially serious for arrays with small inter-microphone spacing as the true delays may be less than a sample unless operating at high sampling frequencies. A possible approach for broadband

signals is to increase the inter-microphone spacing, the maximum allowable distance before phase-wrapping occurs depends on the fundamental frequency, the characteristics of the incident signal and the levels of reverberation. In [74] it was shown that for certain singing voices the modulating effects of vibrato result in the inter-microphone spacing being determined by the low vibrato or modulating frequency as opposed to the fundamental frequency of the note being sung.

In the presence of background noise a Maximum Likelihood (ML) weighting as given in equation (3.95) is applied that uses estimates of the background noise present in order to emphasize frequency components with higher SNR [71]. This method is called the Generalized Cross-Correlation (GCC) method and is the most widely used TDE technique.

$$W_{ML} = \frac{|X_1(\omega)|\,|X_2(\omega)|}{|N_1(\omega)|^2\,|X_2(\omega)|^2 + |N_2(\omega)|^2\,|X_1(\omega)|^2} \tag{3.95}$$

The GCC performs well in moderately noisy environments, however in the case of low SNR the accuracy of the delay estimates will be significantly reduced [71].The performance also rapidly degrades in the presence of reverberation as the mathematical model is based on the assumption of free space propagation, and it relies on the spectral characteristics of the received signal, which are modified by multipath propagation. One approach to improving performance of the GCC filter, which has recently been investigated is the use of any prior knowledge about the signal structure, such as the harmonic structure of the speech signal [12] or the characteristics of a singing signal [75].

The phase correlation method (PHAT) reduces the effect of multi-path propagation by flattening the magnitude spectrum and thereby de-emphasizing the frequency-dependent weightings:

$$W_{PHAT} = \frac{1}{|X_1(\omega)\,X_2^*(\omega)|}. \tag{3.96}$$

This normalization weighting function places equal emphasis on each component of the cross-spectrum phase [71] and results in improved accuracy under mildly reverberant conditions [76], while also having the benefit of requiring no background noise estimate. However, in highly reverberant environments, serious inaccuracies in the TDOA estimates will still arise, as the reverberation produces spurious peaks that have greater amplitude than the peak due to the true source. These effects are investigated in [77] and [78]. The performance of the PHAT method is also poor in the presence of background noise due to accentuation of components with poor SNR.

Previous attempts to improve the overall confidence level of TDOA estimates have in some situations been successful. However, in highly reverberant or noisy environments large errors in the delay estimates can be expected. By using three separate reliability

criteria the confidence of every TDOA estimate is found in [79]. These reliability criteria are based on both psycho-acoustic knowledge and properties of the GCC function itself, and provide a practical analysis of the quality of the TDOA estimates achieved.

A technique called "'hemisphere sampling"' which improves the robustness of TDE techniques to reverberation and low SNR was recently proposed in [17, 80, 81]. The initial step in this approach consists of computing the cross-correlation, $R_{x_1,x_2}(\delta\tau_p)$ (equation 3.92). For each microphone pair a hemisphere surrounding the microphones is then mapped out from the source location corresponding to each value of the correlation. The likelihood of a given source location is equal to the value of the correlation corresponding to this location. The results from each microphone pair are mapped onto the hemisphere in this manner and the estimate of the source location is given by the peak of the entire hemisphere. As the decision on the source location is delayed and all the information available is used, the resulting estimation technique is more robust than traditional delay estimation techniques.

4

Model Order Determination

4.1 Introduction

The initial step in array processing is the determination of the number of sources present, called the model order determination step. Once the number of sources, or the model order, is known, estimation of the required parameters can be performed. While successful determination of the number of sources is important for all DOA estimation schemes, it is especially important in order to obtain good performance for high-resolution direction finding estimators.

A lot of work has been published concerning the problem of model order selection. Estimating the number of sources is traditionally thought of as being equivalent to the determination of the number of eigenvalues of the covariance matrix that differ in value from the smallest eigenvalue [82]. Such an approach leads to a rank reduction principle in order to separate the noise from the signal eigenvalues [83]. Anderson [84] gave a hypothesis testing procedure based on the confidence interval of the noise eigenvalue, in which a threshold value must be assigned subjectively. He showed that the ratio of the log-likelihood to the number of snapshots is asymptotic to a χ^2 distribution. For a small number of snapshots, James introduced the idea of "modified statistics" [85]. In [86], Chen et al. proposed a method based on prior knowledge of the observation probability

density function that detects the number of sources present by setting an upper bound on the value of the eigenvalues.

For thirty years, Information Theoretic Criteria (ITC) approaches have been widely suggested for detection of multiple sources [87]. The best known of this test family are the Akaike Information Criterion (AIC) [88] and the Minimum Description Length (MDL) [69, 89, 90].

These tests are based on a decomposition of the correlation matrix $\mathbf{R}_x$ into two orthogonal components, the signal and noise subspaces, and are used here as benchmarks.

The aim of the AIC method is to determine the order of a model using information theory. Using the expression given in [69] for the AIC, the number of sources is the integer $\widehat{P}$ which, for $m \in \{0, 1, \ldots, M-1\}$, minimizes the following quantity:

$$AIC(m) = -N(M-m)\log\left(\frac{g(m)}{a(m)}\right) + m(2M-m), \tag{4.1}$$

where $g(m)$ and $a(m)$ are respectively the geometric and arithmetic means of the $(M-m)$ smallest eigenvalues of the covariance matrix of the observation. The first term is the log-likelihood residual error, while the second is a penalty for overfitting. This criterion does not correctly estimate the number of sources with a probability of one, even with an infinite number of samples.

The MDL approach is also based on information theoretic arguments, and the selected model order is the one which minimizes the following expression:

$$MDL(m) = -N(M-m)\log\left(\frac{g(m)}{a(m)}\right) + \frac{1}{2}m(2M-m)\log N. \tag{4.2}$$

The form of the MDL used here is that given in [69]. The difference between the AIC and the MDL methods is in the penalty term which has been modified in the MDL to yield asymptotic consistency.

The AIC is not consistent and tends to over-estimate the number of sources present, even at high Signal to Noise Ratio (SNR) values. While the MDL method is asymptotically consistent, it tends to under-estimate the number of sources at low and moderate SNR. In [91] a theoretical evaluation is given of the probability of over- and under-estimation of source detection methods such as the AIC and MDL, under the assumption of asymptotic conditions.

In an effort to moderate the behaviour of the AIC and MDL methods, Wong et al. proposed a modified ITC approach in [92], in which an alternative to the likelihood function used in the AIC and MDL approaches is proposed. This approach is based on the observation that the eigenvectors are irrelevant in determining the number of sources

present and instead the necessary information is contained in the eigenvalues. This leads to the use of the marginal pdf of the observed eigenvalues as the log-likelihood function. In [82] a general ITC is proposed in which the first term of the criterion can be selected from a set of suitable functions. Based on this method Wu and Fuhrmann then proposed a parametric technique as an alternative method of defining the first term of this criterion.

Using Bayesian methodology Djurić then proposed an alternative to the AIC and MDL methods [93, 94] in which the penalty against over-parameterization was no longer independent of the data. Some authors have also investigated the possible use of eigenvectors for model order selection [95, 96], but they generally suffer from the need to introduce prior knowledge. More recently, Wu et al. [97] proposed two ways of estimating the number of sources by drawing Gerschgorin radii.

These algorithms work correctly when the noise eigenvalues are similar or closely clustered. However for a small sample size, where a sample is defined as small when the number of snapshots is of the same order as the number of microphones, this condition is no longer valid and the profile of noise eigenvalues arranged in decreasing order can instead be seen to have an approximately exponential profile.

Recently this problem of detecting multiple sources was re-addressed by looking directly for a gap between the noise and signal eigenvalues [98]. In this way - and as an alternative to the traditional approaches - a method of obtaining an estimate of the model order based on the profile of the ordered noise eigenvalues was proposed in [1]. Based on this test a method of determining the number of significant targets (or reflective sources) in time reversal imaging was proposed in [99]. Assuming that the smallest eigenvalue is a noise eigenvalue, the exponential profile of the ordered noise eigenvalues can be used to find the theoretical profile of the noise-only eigenvalues. Starting with the smallest eigenvalue a recursive algorithm is then applied in order to detect a mismatch greater than a threshold value between each observed eigenvalue and the corresponding theoretical eigenvalue. The index where such a mismatch first occurs is taken to equal the number of sources present.

The test proposed in [99] uses thresholds obtained from the empirical dispersion of ordered noise eigenvalues as discussed in section 4.5.1. An alternative method of determining the corresponding thresholds for a pre-defined false alarm probability as discussed in section 4.5.2 was then proposed in [70], and the improvement of the proposed method compared to some of the classical tests was demonstrated. In [100] the proposed method was seen to perform significantly better than the classical tests when determining the number of wideband acoustical sources using experimental results taken in a reverberant environment.

4.2 Problem Formulation

4.2.1 Principle of statistical tests based on eigenvalue profile

Using the signal model given in equation (2.22), the noiseless observations $\mathbf{r}(t)$ are defined as a linear combination of $\mathbf{a}(\theta_1), ..., \mathbf{a}(\theta_P)$. Assuming independent source amplitudes $\mathbf{s}(t)$, the random vector $\mathbf{r}(t)$ spans the whole subspace generated by the steering vectors. As discussed in chapter 3 this is the "signal subspace". Assuming $P < M$ and no antenna ambiguity, the signal subspace dimension is P. Consequently the number of non-zero eigenvalues of $\mathbf{R}_r$ is equal to the number of sources P, with $(M - P)$ eigenvalues being zero.

Now, in the presence of white noise, according to (3.20), $\mathbf{R}_x$ has the same eigenvectors as $\mathbf{R}_r$, with eigenvalues $\lambda_x = \lambda_r + \sigma^2$; and σ^2 is a degenerate order $(M - P)$ eigenvalue. Then, from the spectrum of $\mathbf{R}_x$ with eigenvalues in decreasing order, it becomes easy to discriminate between signal or noise eigenvalues and determination of the model order would be an easy task.

In practice, $\mathbf{R}_x$ is unknown and must instead be estimated using:

$$\widehat{\mathbf{R}}_x = \frac{1}{N} \sum_{t=1}^{N} \mathbf{x}(t)\mathbf{x}(t)^H, \tag{4.3}$$

where N is the number of snapshots available. $\widehat{\mathbf{R}}_x$ involves averaging over the number of snapshots available and therefore $\widehat{\mathbf{R}}_x \rightarrow \mathbf{R}_x$, as $N \rightarrow \infty$, resulting in all the noise eigenvalues being equal to σ^2. However, when taken over a finite number of snapshots, the sample matrix $\widehat{\mathbf{R}}_x \neq \mathbf{R}_x$. In the spectrum of ordered eigenvalues, the "signal eigenvalues" are still identified as the P largest ones. But, the noise eigenvalues are no longer equal to each other, and the separation between the signal and noise eigenvalues is not clear (except in the case of high SNR, when a large separation can be observed between signal and noise eigenvalues), making discrimination between signal and noise eigenvalues a difficult task.

4.2.2 Qualification of order estimation performance

Letting P equal the number of sources and $\widehat{P}$ the estimated model order, three mutually exclusive propositions and their corresponding probabilities are considered:

$$\begin{aligned} &\widehat{P} = P : \text{ correct detection}, && P_d = Prob\left[\widehat{P} = P\right] \\ &\widehat{P} > P : \text{ over-estimation}, && P_{fa} = Prob\left[\widehat{P} > P\right] \\ &\widehat{P} < P : \text{ under-estimation}, && P_{nd} = Prob\left[\widehat{P} < P\right] \end{aligned} \tag{4.4}$$

Various methods can now be compared on the basis of P_d and P_{fa} values for various numbers of sources, locations, and numbers of snapshots.

Usually, a detection threshold may be adjusted to provide the best compromise between detection and false alarm. In such situations, a common practice is to set the threshold for a given value of P_{fa} (1% for instance) and to compare the corresponding values of P_d for different methods. The probabilities P_d and P_{fa} will be estimated from statistical occurrence rates by Monte Carlo simulations.

4.3 Eigenvalue Profile Of The Correlation Matrix Under The Noise-Only Assumption

As the noise eigenvalues are no longer equal for a small sample size it is necessary to identify the mean profile of the decreasing noise eigenvalues. The eigenvalue profile of the sample covariance matrix for the noise-only situation $\widehat{\mathbf{R}}_n = \frac{1}{N}\sum_{t=1}^{N}\mathbf{n}(t).\mathbf{n}(t)^H$ is therefore considered. The distribution of the matrix $\widehat{\mathbf{R}}_n$ is a Wishart distribution [101] with N degrees of freedom. This distribution can be seen as a multivariate generalization of the χ^2 distribution. It depends on N, M and σ^2 and is sometimes denoted by $W_M(N, \sigma^2\mathbf{I})$. In order to establish the mean profile of the ordered eigenvalues (denoted as $\lambda_1, ..., \lambda_M$) the joint probability of an ordered M-tuplet has to be known. The joint distribution of the ordered eigenvalues is then [101]

$$p(\lambda_1, ..., \lambda_M) = \alpha \left(-\frac{1}{2\sigma^2}\sum_{i=1}^{M}\lambda_i\right)\left(\prod_{i=1}^{M}\lambda_i\right)^{\frac{1}{2}(N-M-1)}\prod_{i>j}^{M}(\lambda_j - \lambda_i), \tag{4.5}$$

where α is a normalization coefficient. The distribution of each eigenvalue can be found [102], but this requires zonal polynomials and appears to produce unusable results.

Instead an alternative approach is used which consists of finding an approximation of this profile by conserving the first two moments of the trace of the error covariance matrix defined by:

$$\mathbf{\Psi} = \hat{\mathbf{R}}_n - \mathbf{R}_n = \hat{\mathbf{R}}_n - E\left\{\hat{\mathbf{R}}_n\right\} = \hat{\mathbf{R}}_n - \sigma^2\mathbf{I}. \tag{4.6}$$

It follows from $E\{tr[\mathbf{\Psi}]\} = \mathbf{0}$, that

$$M\sigma^2 = \sum_{i=1}^{M}\lambda_i \tag{4.7}$$

Note that $\Psi_{ij} = \frac{1}{N}\sum_{t=1}^{N} n_i(t).n_j^*(t) - \sigma^2\delta_{ij.}$ It is easy to show that $E\left\{\|\Psi_{ij}\|^2\right\} = \frac{\sigma^4}{N}$ in the case of white Gaussian complex circular noise. Since the trace of a matrix remains

unchanged when the base changes, it follows that

$$E\left\{tr\left(\hat{\mathbf{R}}_n - \mathbf{R}_n\right)^2\right\} = \sum_{i,j} E\left\{\|\Psi_{ij}\|^2\right\} = M^2\frac{\sigma^4}{N} = \sum_{i=1}^{M}\left(\lambda_i - \sigma^2\right)^2 \tag{4.8}$$

The decreasing model retained for the approximation is:

$$\lambda_i = \lambda_1 r_{M,N}^{i-1}, \tag{4.9}$$

with $0 < r_{M,N} < 1$. Of course, $r_{M,N}$ depends on M and N, but is denoted by r for simplicity. From equation (4.7) it can be seen that:

$$\lambda_1 = M\frac{1-r}{1-r^M}\sigma^2 = MJ_M\sigma^2, \tag{4.10}$$

where

$$J_M = \frac{1-r}{1-r^M}. \tag{4.11}$$

Considering that $(\lambda_i - \sigma^2) = (MJ_M r^{i-1} - 1)\,\sigma^2$, the relation (4.8) gives

$$\frac{M+N}{MN} = \frac{(1-r)\left(1+r^M\right)}{\left(1-r^M\right)(1+r)}. \tag{4.12}$$

From both simulation results shown in figure 4.1, and experimental results reported in literature (for example see [103]) the decreasing model of the noise-only eigenvalues can be seen to be approximately exponential. Setting $r = e^{-2a}\,(a > 0)$ (4.10) can be re-expressed as:

$$\frac{M.\tanh(a) - \tanh(Ma)}{M.\tanh(Ma)} = \frac{1}{N}, \tag{4.13}$$

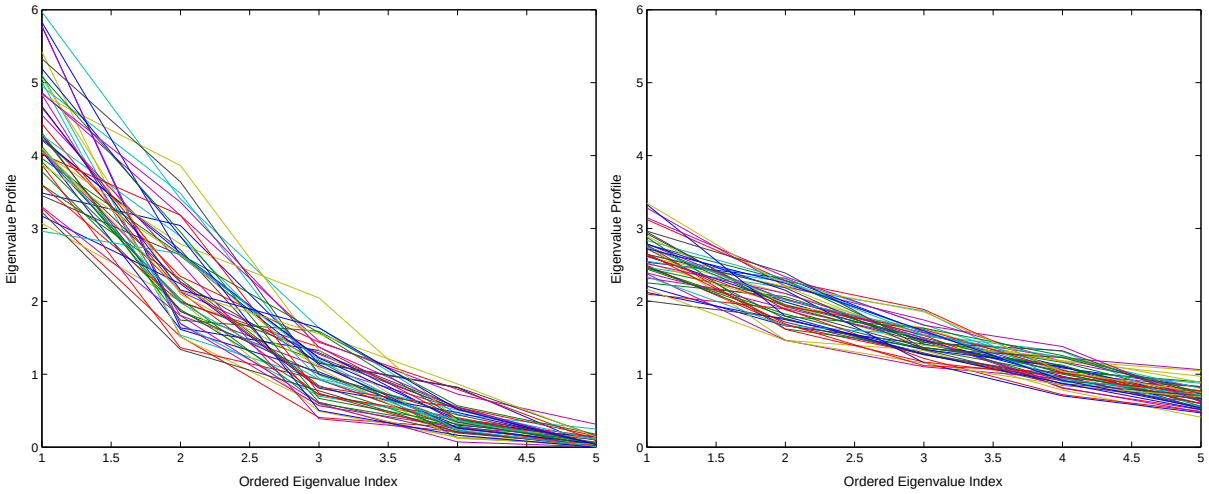
where tanh is the hyperbolic tangent function. An order-4 expansion gives the following bi-quadratic equation in a:

$$a^4 - \frac{15}{M^2+2}a^2 + \frac{45M}{N\left(M^2+1\right)\left(M^2+2\right)} = 0 \tag{4.14}$$

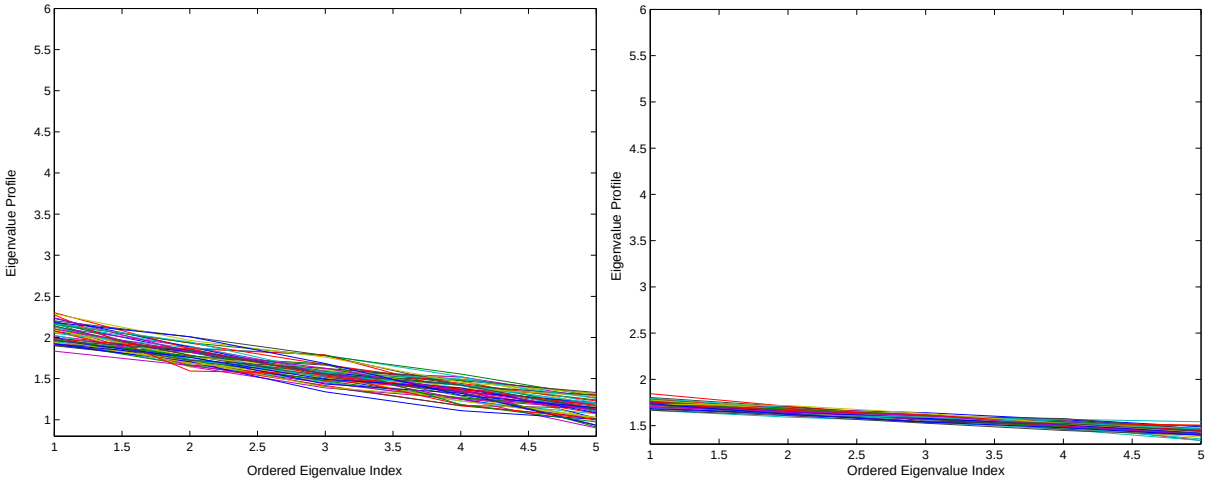
for which the positive solution is given by

$$a(M,N) = \sqrt{\frac{1}{2}\left\{\frac{15}{M^2+2} - \sqrt{\frac{225}{\left(M^2+2\right)^2} - \frac{180M}{N\left(M^2-1\right)\left(M^2+2\right)}}\right\}}. \tag{4.15}$$

As the calculation of the noise-only eigenvalue profile takes into account the number of snapshots N, this profile is valid for all sample sizes, with the exponential tending to a horizontal profile as $N \to \infty$ and the noise eigenvalues become equal.



(a) $M = 5, N = 5$ (b) $M = 5, N = 20$



(c) $M = 5, N = 100$ (d) $M = 5, N = 1000$

Figure 4.1: *Profile of the ordered eigenvalues under the noise-only assumption for 50 independent trials, with M=5 and various values of N.*

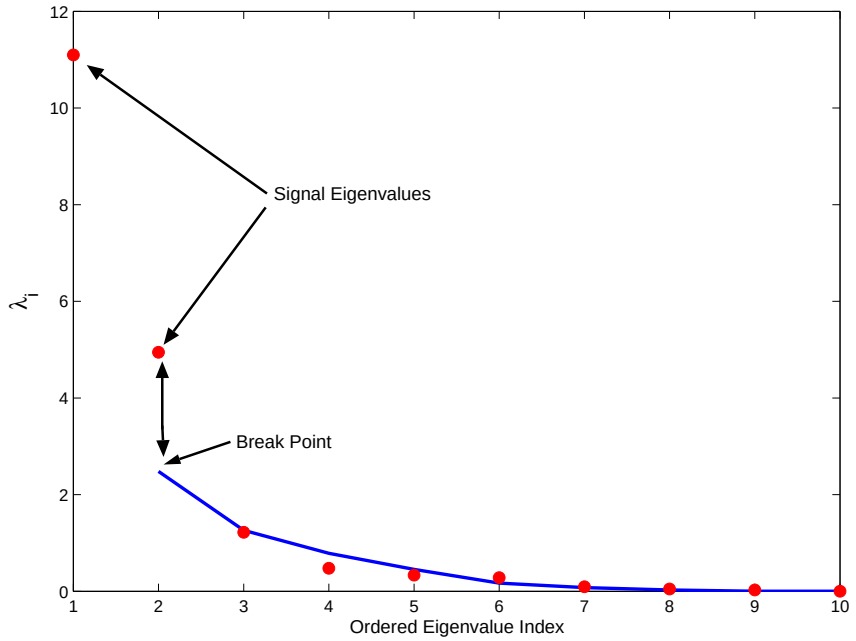


Figure 4.2: *Profile of ordered noise eigenvalues in the presence of 2 sources, and 10 microphones. The ordered profile of the observed eigenvalue is seen to break from the noise eigenvalue distribution, when there are sources present.*

4.4 A Recursive Exponential Fitting Test (EFT)

4.4.1 Test principle

The expressions for the noise-only eigenvalue profile can now be extended to the case where the observations consist of P non-coherent sources corrupted by additive noise. Under these conditions the covariance matrix can be broken down into two complementary subspaces: the source subspace $\mathbf{E}_s$ of dimension P, and the noise subspace $\mathbf{E}_n$ of dimension $Q = M - P$. Consequently, the profile established in the previous section still holds for the Q noise eigenvalues, and the theoretical noise eigenvalues can be found by replacing M with Q in the previous expressions for the noise-only eigenvalue profile.

The proposed test then finds the highest dimension d_n of the candidate noise subspace, such that the profile of these d_n candidate noise eigenvalues is compatible with the theoretical noise eigenvalue profile. The main idea of the test is to detect the eigenvalue index at which a break occurs between the profile of the observed eigenvalues and the theoretical noise eigenvalue profile provided by the exponential model. Figure 4.2 shows how a break point appears between the signal eigenvalues and the theoretical noise eigenvalue profile, while the observed noise eigenvalues are seen to fit the theoretical profile.

Firstly, an eigen-decomposition of the sample covariance matrix is performed and the resulting eigenvalues $\lambda_1, \ldots, \lambda_M$, called the observed eigenvalues in the following discussion, are arranged in order of decreasing size. Beginning with the smallest observed eigenvalue λ_M, this is assumed to be a noise eigenvalue, giving the initial candidate noise subspace dimension $d_n = 1$. Then using λ_M, $d_n = 1$ and the prediction equations (4.10) the next eigenvalue of the theoretical noise eigenvalue profile $\widehat{\lambda}_{M-1}$ can be found [1]:

$$\hat{\lambda}_{M-d_n} = (d_n + 1) J_{d_n+1} \hat{\sigma}^2, \tag{4.16}$$

$$\text{with } J_{d_n+1} = \frac{1 - r_{d_n+1,N}}{1 - (r_{d_n+1,N})^{d_n+1}}, \tag{4.17}$$

$$\text{and } \hat{\sigma}^2 = \frac{1}{d_n + 1} \sum_{i=0}^{d_n} \lambda_{M-i}. \tag{4.18}$$

Now taking both λ_M and $\widehat{\lambda}_{M-1}$ to be noise eigenvalues, corresponding to a candidate noise subspace dimension $d_n = 2$, equations (4.16)-(4.18) are applied again to predict $\widehat{\lambda}_{M-2}$.

These steps are then repeated, and for each step the candidate noise subspace dimension d_n is increased by one. Then taking all the previously estimated noise eigenvalues, the next noise eigenvalue in the theoretical profile $\hat{\lambda}_{M-d_n}$ is found. this process is continued until $d_n = M - 1$, and the M eigenvalues of the theoretical noise-only profile, $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_M$ have been calculated, (where $\widehat{\lambda}_M = \lambda_M$).

The following two hypotheses are defined:

$$H_{d_n+1} \quad : \quad \lambda_{M-d_n} \text{ is a noise eigenvalue.} \tag{4.19}$$

$$\overline{H}_{d_n+1} \quad : \quad \lambda_{M-d_n} \text{ is a signal eigenvalue.} \tag{4.20}$$

Then, starting with the smallest eigenvalue pair (that are not equal) $\widehat{\lambda}_{M-1}$ and λ_{M-1}, the relative distance between each of the theoretical noise eigenvalues and the corresponding observed eigenvalue is found, and compared to the threshold, η_{d_n} found for that eigenvalue index, equations (4.21) and (4.22). Selection of a suitable threshold is discussed in sections 4.5.

$$H_{d_n+1} : \left| \frac{\lambda_{M-d_n} - \widehat{\lambda}_{M-d_n}}{\widehat{\lambda}_{M-d_n}} \right| \leq \eta_{d_n} \tag{4.21}$$

$$\overline{H}_{d_n+1} : \left| \frac{\lambda_{M-d_n} - \widehat{\lambda}_{M-d_n}}{\widehat{\lambda}_{M-d_n}} \right| > \eta_{d_n} \tag{4.22}$$

If the relative difference between the theoretical noise eigenvalue and the observed eigenvalue is less than (or equal to) the corresponding threshold the observed eigenvalue matches the theoretical noise-only eigenvalue profile, and so it is deemed to be a noise eigenvalue, which is the case shown by equation (4.21).

The next eigenvalues $\widehat{\lambda}_{M-2}$ and λ_{M-2} are then compared in the same manner. This process continues until a pair of eigenvalues $\widehat{\lambda}_{M-d_n}$ and λ_{M-d_n}, whose relative difference is greater than the corresponding threshold as shown in equation (4.22), is found.

4.4.1.1 Note on the complexity

The proposed EFT method requires calculation of the sample correlation matrix for each set of observations. An eigenvalue decomposition of this matrix must then be performed and the smallest of the observed eigenvalues is used to predict the theoretical noise-only eigenvalue profile. The computational cost of the EFT method is therefore of the same order as that of the AIC and MDL tests.

4.5 Computation of Thresholds

4.5.1 Using the empirical distribution of the noise-only eigenvalue profile

In many situations a practical approach to selecting a suitable threshold η_{d_n} for each step, is to consider the empirical distribution of the noise-only eigenvalue profile. This is emphasized in figure 4.3, where the large dots denote the mean value of each ordered eigenvalue empirically obtained over 10000 realizations. Assuming a normal distribution for each ordered eigenvalue, this figure reports the width corresponding to a six standard deviation truncation. For clarity of presentation, the figure only shows a few realizations, linking the corresponding eigenvalues. The threshold for each step, d_n, is chosen as half the width of the corresponding eigenvalue in the noise-only distribution. It should be noted that these relative η_{d_n} are independent of the noise power.

4.5.2 Selecting a threshold to ensure a pre-determined false alarm probability is observed

Alternatively it may be desirable to determine the thresholds to ensure that a preset probability of false alarm P_{fa} is observed. From the definition of P_{fa} given in equation (4.4) the expression for the P_{fa} in the noise-only case can be decomposed as follows:

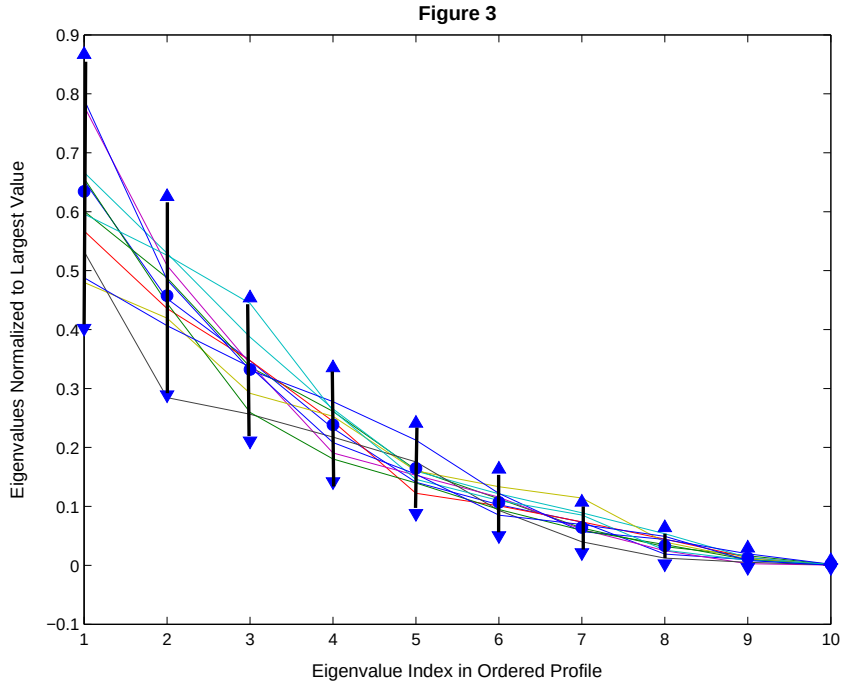


Figure 4.3: *Profile of ordered noise eigenvalues for several realizations. The circles through the centre show the mean value for each eigenvalue. The distance between the upper and lower triangles is the spread of the eigenvalue and the chosen threshold is equal to half this distance.*

$$P_{fa} = \Pr\left[\widehat{P} > P_0 | P = P_0\right] \text{ for } P_0 = 0, 1, 2, ..., M-1. \tag{4.23}$$

For the noise-only case $P_0 = 0$, and the expression for P_{fa} can be decomposed as follows:

$$\begin{aligned} P_{fa} &= \Pr\left[\widehat{P} > P_0 | P = P_0\right] && (4.24)\\ &= \sum_{i=1}^{M-1} \Pr\left[\widehat{P} = i | P = 0\right] && (4.25)\\ &= \sum_{d_n=1}^{M-1} P_{fa}^{(d_n)}, && (4.26) \end{aligned}$$

where $P_{fa}^{(d_n)} = \Pr\left[\widehat{P} = M - d_n | P = 0\right]$ is the contribution of the d_nth step to the total false alarm.

Re-expressing equations (4.21) and (4.22) results in:

$$H_{d_n+1} : Q(d_n) = \left| \frac{\lambda_{M-d_n}}{\sum_{i=M-d_n}^{M} \lambda_i} \right| \leq \left(\eta_{d_n} + 1\right) J_{d_n+1} \tag{4.27}$$

$$\bar{H}_{d_n+1} : Q(d_n) = \left| \frac{\lambda_{M-d_n}}{\sum_{i=M-d_n}^{M} \lambda_i} \right| > \left(\eta_{d_n} + 1\right) J_{d_n+1}, \tag{4.28}$$

resulting in the following expression for $P_{fa}^{(d_n)}$ in the noise-only situation:

$$P_{fa}^{(d_n)} = \Pr\left[Q\left(d_n\right) > \left(\eta_{d_n} + 1\right) J_{d_n+1} | P = 0\right]. \tag{4.29}$$

Then, denoting the distribution of $Q\left(d_n\right)$ as $f_{d_n}\left(q\right)$ and the threshold η_{d_n} is defined by the following integral equation:

$$P_{fa}^{(M-d_n)} = \int_{J_{d_n+1}\left(\eta_{d_n}+1\right)}^{+\infty} f_{d_n}\left(q\right) dq. \tag{4.30}$$

Solution of this equation in order to find η_{d_n} is reliant on knowledge of the distribution $f_{d_n}\left(q\right)$. For $d_n = M$ and $d_n = M-1$ the distribution is known and is given in [69], but is unusable in this application. It appears that this statistical distribution is not known for other values of d_n. Hence, numerical methods must instead be used in order to solve for η_{d_n}.

4.5.2.1 Threshold determination by Monte Carlo Methods

Using $I = P_{fa}^{(M-d_n)}$ for the sake of notational simplicity, equation (4.29) can be rewritten as:

$$I = \int_D p\left(\lambda_1, ..., \lambda_M\right) \prod_{i=1}^{M} d\lambda_i = E\left[1_{\mathcal{D}}\right] \tag{4.31}$$

where $\mathcal{D}$ is the domain of integration defined as follows

$$\mathcal{D} = \left\{0 < \lambda_M < ... < \lambda_1 < \infty | Q\left(d_n\right) > J_{d_n+1}\left(\eta_{d_n} + 1\right)\right\}, \tag{4.32}$$

and $1_{\mathcal{D}}\left(\lambda_1, ..., \lambda_M\right)$ is the indicator function over the domain $\mathcal{D}$. Its value is unity if the eigenvalues belong to $\mathcal{D}$ and zero otherwise. $E\left[1_{\mathcal{D}}\right]$ is the expected value of the indicator function. Equation (4.31) can then be estimated by Monte Carlo simulation, in which the steps are:

- Generation of q noise-only sample correlation matrices, where q is the number of the Monte Carlo trials to be run,
- Computation of the ordered eigenvalues for each of these q matrices: $(\lambda_{1,j}, ..., \lambda_{M,j})$ $1 \leq j \leq q$,
- Estimation of I by $\widehat{I} = \frac{1}{q}\sum_{j=1}^{q} 1_{\mathcal{D}}\left(\lambda_{1,j}, ..., \lambda_{M,j}\right)$.

As the P_{fa} is usually very small, q must be statistically determined in order to obtain a pre-defined precision for the estimation of I. Because of the Central Limit Theorem $\widehat{I}$ follows a Gaussian law. Consequently, denoting the standard deviation of $\widehat{I}$ as σ, the following is true: $\Pr\left[\frac{\sqrt{q}}{\sigma}\left|I - \hat{I}\right| < 1.96\right] = 0.95$, where $\Pr[x < y]$ is the probability that $x < y$. $\sigma^2 = E\left[\left(1_{\mathcal{D}}\left(.\right)\right)^2\right] - I^2 = I - I^2$, and as $I - I^2 \approx I$, we can say that $\sigma = \sqrt{I}$.

4.5.2.2 Application

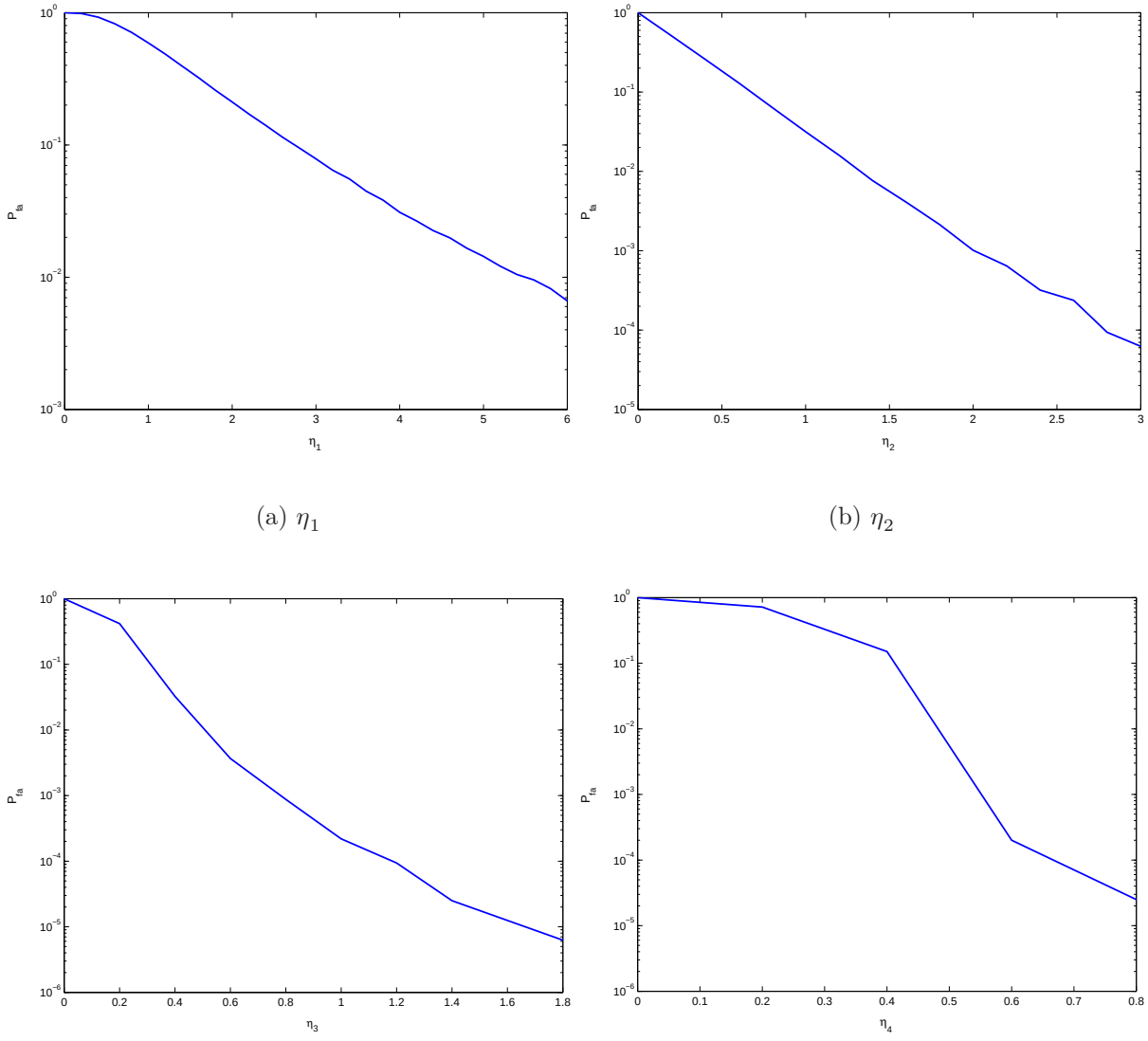
For $M = 5$ microphones and a false alarm probability of 1%, identically distributed over the $M - 1$ steps of the test, $I = P_{fa}^{(M-d_n)1 \leq d_n \leq 4} = \frac{0.01}{4} = 0.0025$.

Then, with a probability of 95%, we want to estimate $\widehat{I_{pfa}}$ with an accuracy of 10%, i.e. we want to find $\widehat{I_{pfa}}$ such that $I_{pfa} - \widehat{I_{pfa}} = 0.1\left(I_{pfa}\right)$, where $0.10.1\left(I_{pfa}\right) = \frac{\sigma^2}{q}$. Solving for q in this situation, we see that $q = 160000$ Monte Carlo trials are needed. In figure 4.4 the $P_{fa}^{(M-d_n)}$ versus η_{d_n} is plotted. From this η_{d_n} is selected for each d_n and for a given P_{fa}.

4.6 Comparison with Classical Tests

4.6.1 Simulation Examples

Initially, in order to evaluate the test performance in white Gaussian complex noise, computed simulations have been performed for a uniform linear array of five omnidirectional



(a) η_1 (b) η_2

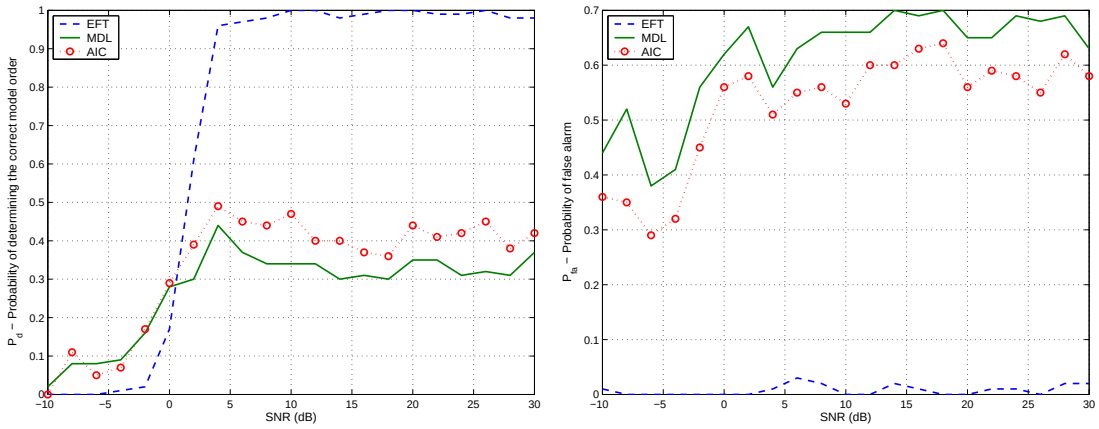
(c) η_3 (d) η_4

Figure 4.4: *Threshold computation for each step in the case where* $M = 5$ *and* $N = 10$.

microphones. Six snapshots are used for the covariance estimation. Two random sources, of the same power, impinge on the array at 0° and +10°. The SNR is defined as:

$$SNR = 10.\log_{10}\left(\frac{\sigma_s^2}{\sigma^2}\right), \tag{4.33}$$

where σ_s^2 is the power of one of the sources, and σ^2 is the noise power.



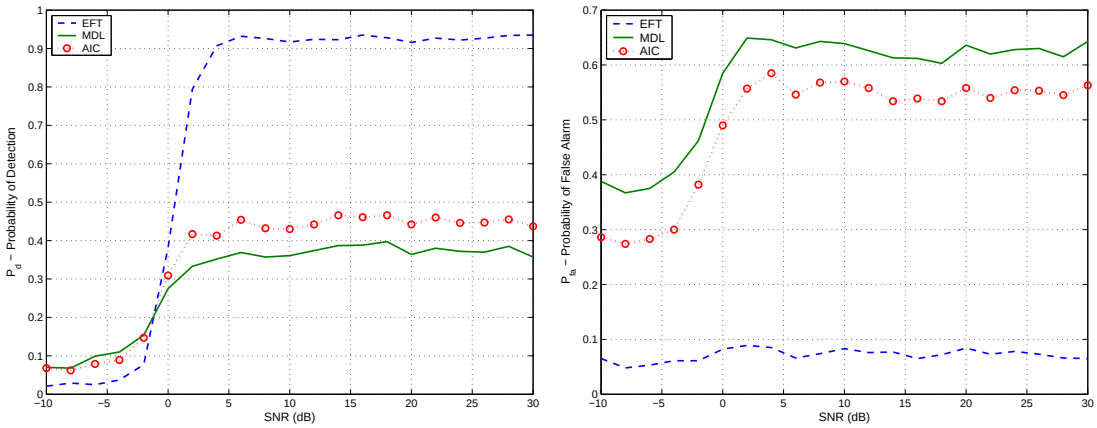
(a) P_d - Probability of detection.

(b) P_{fa} - Probability of false alarm.

Figure 4.5: *Results for the EFT, the MDL and the AIC. In this case the correct model order is 2, the number of snapshots $N = 6$, and the number of microphones $M = 5$. The EFT thresholds have been determined to result in $P_{fa} = 1\%$*

Unlike AIC and MDL, the proposed test provides a way for controlling overestimation if necessary, through the appropriate choice of thresholds. Direct comparison is difficult because AIC and MDL have a higher uncontrollable false alarm rate than the proposed test, which is initially set at 1 %. From figures 4.5 it can also be seen that the proposed algorithm has a far higher probability of correctly detecting the number of sources present for SNR values greater than around $0dB$, even though the P_{fa} is constrained to be 1% and as such is much lower than those of the other tests.

The threshold for the EFT test is then recalculated to allow for a false alarm probability of 10%, as shown in 4.6. Even though the allowable false alarm has been increased the probability of false alarm for the EFT remains significantly lower than that of the other tests as can be seen in figure 4.6. Determining the threshold to observe a very low false alarm probability may lead to non-detection of sources. For many acoustical array processing problems non-detection of sources is undesirable and for this reason the threshold is determined to observe a P_{fa} of 10% for all the following tests.



(a) P_d - Probability of detection.

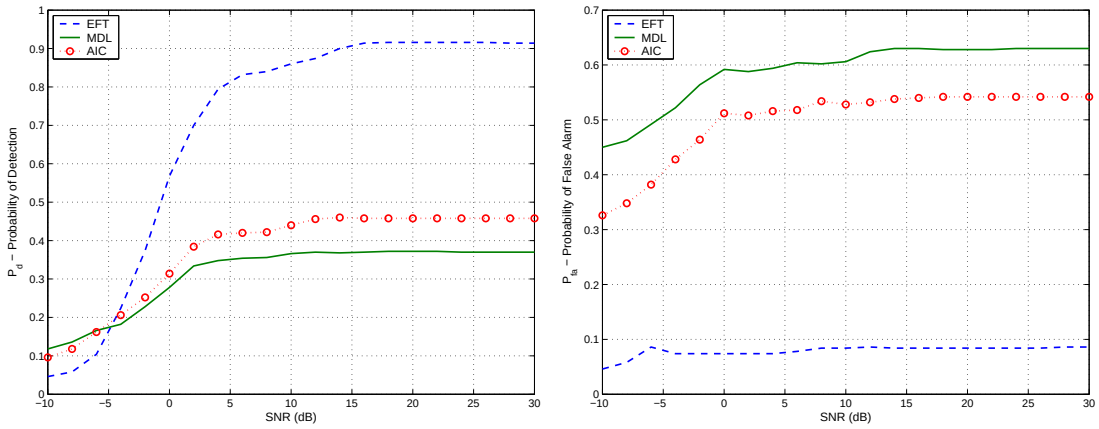
(b) P_{fa} - Probability of false alarm.

Figure 4.6: *Results for the EFT, the MDL and the AIC. In this case the correct model order is 2, the number of samples $N = 6$, and the number of microphones $M = 5$. The thresholds for the EFT have been determined to result in $P_{fa} = 10\%$*

Next the performance of the three tests for simulations of the case of two male speakers is evaluated. The speech signals are received by an antenna of 5 microphones in the presence of Complex White Gaussian Noise and the frame length of 6 samples ($0.375ms$) is used for the covariance estimation. In order to establish the threshold for the EFT, the relative difference between the noise-only eigenvalue distribution and the predicted distribution is considered for 16000 Monte Carlo trials, where the number of trials necessary was found as described in section (4.5.2.1). In figure 4.7(a) it can be seen that for this situation that the EFT outperforms the AIC and MDL methods, and as the SNR increases detects the correct number of sources with a probability of 1. From figure 4.7(b) it can be seen that as expected the P_{fa} for the EFT is around 10%, while that of the other tests is significantly greater.

4.6.2 Room Response Simulations

The previous simulations do not take into the consideration the presence of reverberation, and therefore do not give an accurate indication of the performance that can be expected in a practical situation. Simulations of the following experimental situation using room Enhanced Acoustic Simulator for Engineers (EASE^{TM}) simulation software are therefore considered. This software models the transmission of the signals from the specified sources to the microphone array, taking into account the dimensions and other acoustic characteristics of the room.



(a) Probability of detection.

(b) Probability of false alarm.

Figure 4.7: *Results for the EFT, the MDL and the AIC for simulated case of 2 speech sources received in presence of complex Gaussian White Noise and no reverberation, for the case where $N = 10$ and $M = 5$. The thresholds for the EFT have been determined to result in $P_{fa} = 10\%$*

A microphone array of 5 omni-directional microphones arranged in a Uniform Linear Array (ULA) was simulated with inter-microphone spacing of $3.4cm$. In order to ensure that $\Delta \leq \lambda_{\min}$, where Δ is the distance between the microphones, the signals were filtered to ensure a maximum frequency of $5kHz$. A model of a medium sized classroom or meeting room with dimensions $5.38m \times 6.9m \times 2.44m$, containing desks, chairs and a white board at the top wall was used.

The responses were then found for 2 sources with DOAs of 70^o and 110^o, as shown in figure 4.8. The horizontal distance between the array and the sources was $1.5m$, allowing for the far-field assumption to be made. The height of the microphone array and the sources were equal. The environmental noise came from a room fan located in the ceiling, traffic noise from outside the window and reverberation (reverberation time $T_{60} \approx 0.5s$). The experiment was repeated for two (different) male speakers, two (different) female speakers, and finally one male and one female speaker.

The sampling frequency used was $f_s = 16kHz$, and the number of samples used to determine the model order was $N = 100$, $(6.25ms)$, with a 50% overlap between frames. In order to establish the thresholds recordings were taken of the background noise in order to find the distribution of the noise only eigenvalues. The threshold is then selected from this distribution, in order to observe a false alarm probability over the entire test, as described in section 4.5. In this case the allowable false alarm was set at 10%.

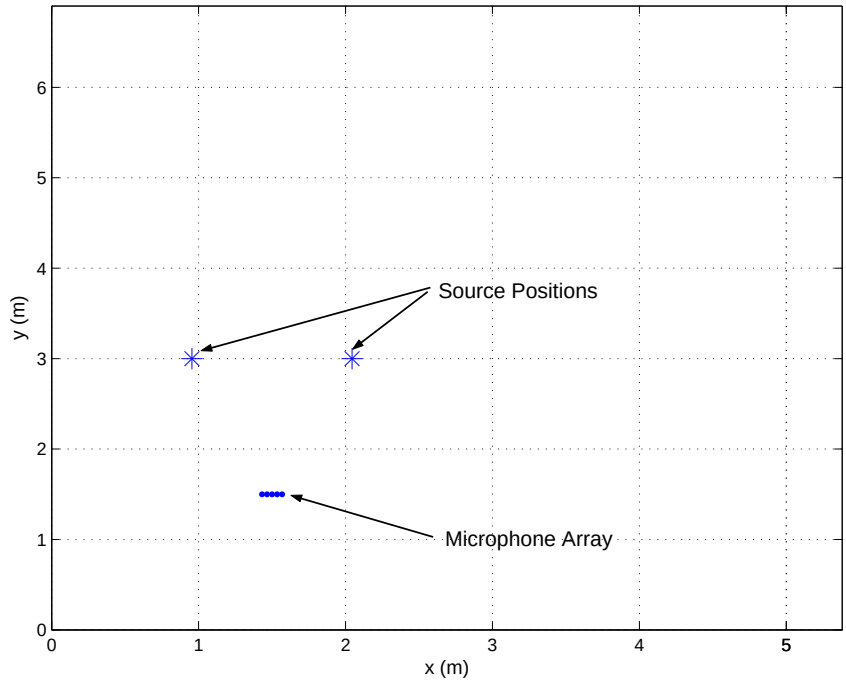
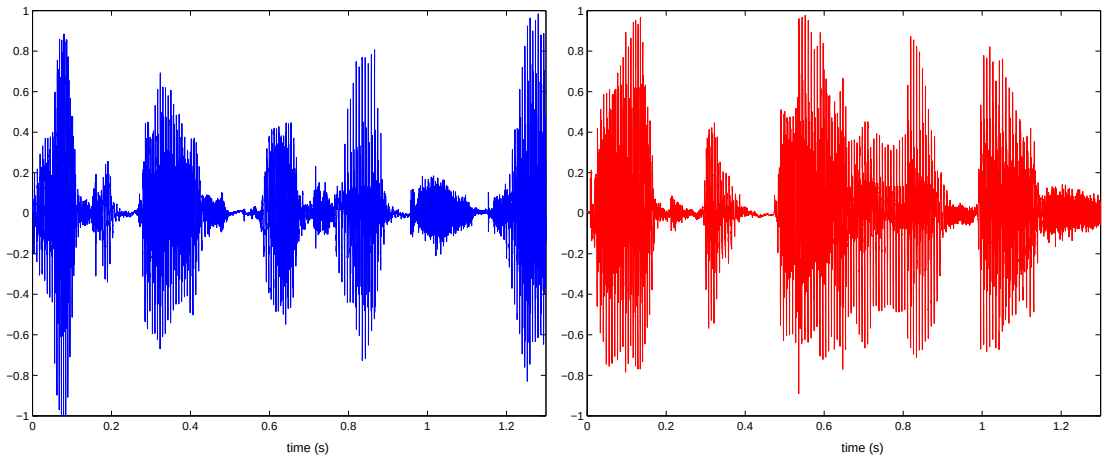


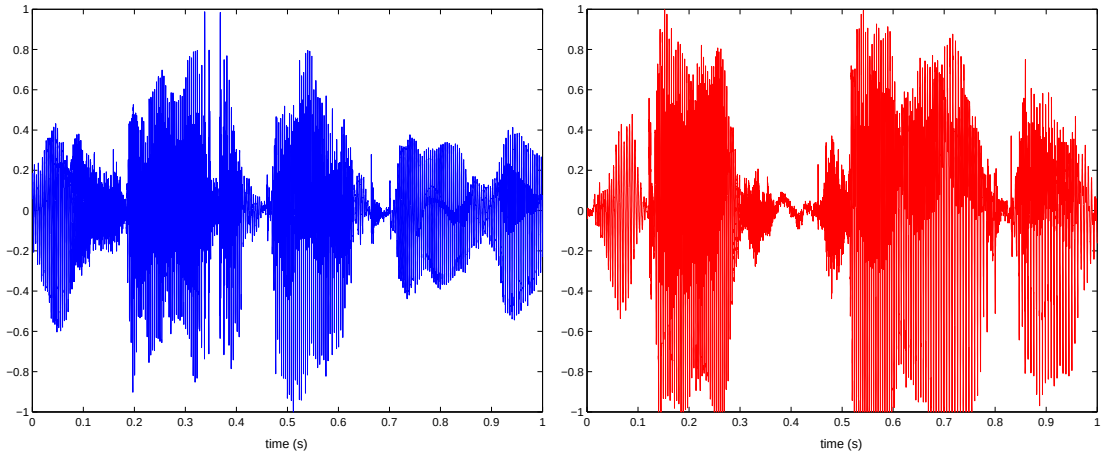
Figure 4.8: *Microphone and sound source positions.*



(a) Male speech 1.

(b) Male speech 2.

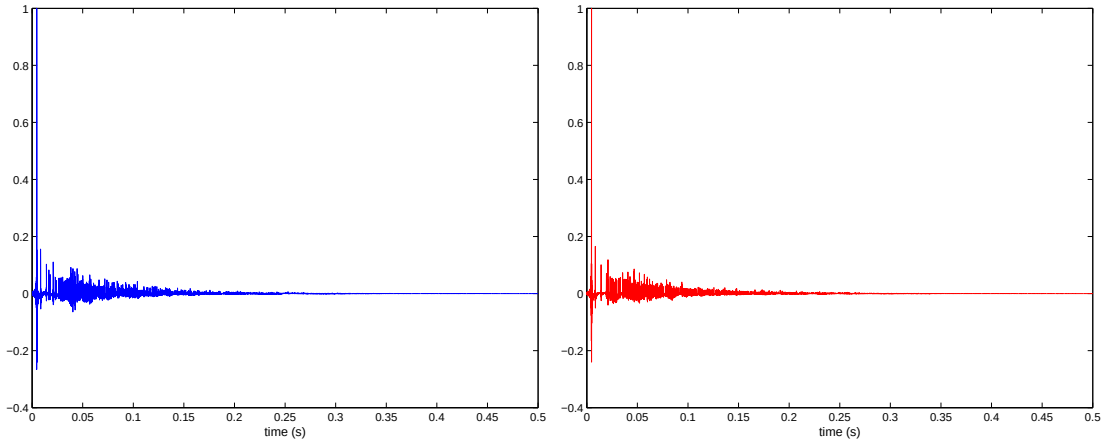
Figure 4.9: *Male source signals*



(a) Female speech 1.

(b) Female speech 2.

Figure 4.10: *Female source signals*



(a) Impulse response from 1st source to the centre microphone in the array.

(b) Impulse response from 2nd source to the centre microphone in the array.

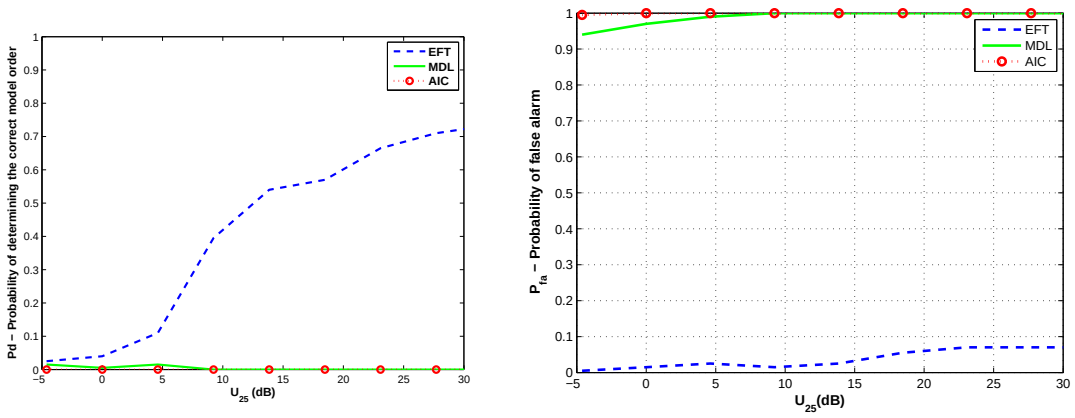
Figure 4.11: *Impulse responses from the sources to the centre microphone in the array. These responses are found using $EASE^{TM}$ acoustic simulation software. The impulse responses are simulated for a given setup based on the acoustical properties of the venue and the geometrical configuration of the source and microphone array.*

The impulse responses produced by this software give a good indication of the effect of reverberation on the test, although they usually simulate higher reverberation levels than those encountered experimentally. The source signals used are two male speakers, and the additive background noise is simulated complex White Gaussian Noise, as in the previous case.

The performance of the tests is evaluated as the Useful-to-detrimental ratio is increased. The Useful-to-detrimental ratio is a measure of the strength of the beneficial early-arriving reflections, to the later arriving sounds and the background noise [104]. The cut-off time between the beneficial and detrimental reflections depends on the room impulse response, and for rooms such as offices or meeting rooms has been shown to lie between $25 - 30ms$ [105]. Here, the energy of the first $25ms$ of the impulse response is used to compute the energy of the early arrivals E, and the energy of the impulse response from $25ms$ to $500ms$ (T_{60}) is the energy of the late arrivals L. The power of the background noise present is N, and the Useful-to-detrimental ratio, U_{25} is defined as:

$$U_{25} = 10 \log \left\{ \frac{E}{L+N} \right\} dB. \tag{4.34}$$

For this case the thresholds are established as described in section 4.5.2, however this time instead of using simulated white noise, recordings were taken in the room when there were no sources present. These recordings are then used to find the eigenvalue distribution in the noise-only case.



(a) P_d - Probability of detection.

(b) P_{fa} - Probability of false alarm.

Figure 4.12: *Probability of false alarm for the EFT, the MDL and the AIC using room simulator EASE, as the Useful-to-Detrimental Ratio, U_{25}, is increased.*

From the results, shown in figure 4.12 it can be seen that the introduction of the

	P_d (%)	P_{fa} (%)	P_{nd} (%)
EFT	87.32	0	12.68
AIC	0.88	99.12	0
MDL	2.36	97.64	0

Table 4.1: *Results found by EFT, AIC and MDL tests using experimental recordings of two different male speakers.*

	P_d (%)	P_{fa} (%)	P_{nd} (%)
EFT	65.96	0	34.04
AIC	0	100	0
MDL	2.46	97.54	0

Table 4.2: *Results found by EFT, AIC and MDL tests using experimental recordings of two different female speakers.*

reverberation causes a serious deterioration in the accuracy of the AIC and MDL tests. While the probability of detection of the EFT is reduced slightly it is much higher than that of the other tests in this case, while the probability of false alarm remains much lower. The results of the AIC and MDL vary only slightly with increasing U_{25}, and it is clear that both these tests are highly unsuitable in the presence of reverberation.

4.6.3 Experimental Results

In this section the results of the EFT, the MDL and the AIC are compared for the physical environment modelled in the previous section. Firstly the case of two male speakers is considered, then the case of the two female speakers and finally the case of one male and one female speaker. Once again the thresholds are found using recordings taken in the absence of a source, and observe a pre-defined false alarm probability of 10%. Tables 4.1-4.3 respectively, show the results for the cases of: two male speakers, two female speakers, and one male and one female speaker.

The results from the experimental recordings confirm those of the simulated room response. Once again the AIC and MDL methods perform very poorly in the presence

	$P_d\,(\%)$	$P_{fa}\,(\%)$	$P_{nd}\,(\%)$
EFT	74.48	0.59	24.93
AIC	0.59	99.41	0
MDL	3.56	96.44	0

Table 4.3: *Results found by EFT, AIC and MDL tests using experimental recordings of one male and one female speakers.*

of reverberation, and they over-estimate the true number of sources. The EFT greatly outperforms the other two methods, showing its suitability for determining the number of speakers in such an environment.

The robustness of the EFT to the effects of reverberation can be attributed to the fact that the threshold is selected using recordings taken in the room when there is no source present. The selection of the threshold then acts as a training step for the test taking the room impulse response into account. This initial training step makes the EFT less likely to mistakenly classify a reflection as a source. However, on the other hand the AIC and MDL methods have no such training step and therefore mis-classify reflections as sources.

4.7 Conclusion

In any source localization scheme, accurate knowledge of the number of sources is essential. As well as being necessary for accurate localization, such information may also be used for noise cancellation. In this chapter, a new test for model order selection based on the geometrical profile of noise-only eigenvalues, has been proposed. The noise eigenvalues for white Gaussian noise have been shown to fit an exponential law whose parameters have been predicted. Contrary to traditional algorithms, this test performs well when there is a small number of snapshots used for the estimation of the correlation matrix. Another important advantage over classical tests is that the false alarm probability observed can be controlled through selection of the threshold. Moreover, the computational cost of the proposed method is of the same order as those of the AIC and MDL.

The performance of the proposed test was then compared to that of the AIC and MDL tests for the problem of determining the number of speakers in a moderately reverberant environment. It has been shown that in the absence of reverberation, all three tests perform well in the presence of white Gaussian noise for SNR levels of higher than $5dB$.

Using EASE^{TM} (Enhanced Acoustic Simulator for Engineers) room simulation software, it has then been shown that once reverberation is introduced the performance of all three tests is reduced, and the EFT greatly outperforms the AIC and MDL methods, both of which overestimate the number of sources present. These results were then confirmed using experimental recordings.

The novel contributions made in this chapter include the application of the test originally proposed in [1] to model order determination in the acoustic context. The performance of this model order determination method was then compared to that of the AIC and MDL tests using both computer simulations and experimental recordings, which take into account the effect of both SNR and the presence of reverberation on the performance of the model order determination methods.

5

Comparison of Direction of Arrival Estimators

5.1 Introduction

In chapter 2 it was shown that there is no standard technique guaranteed to produce an optimal estimator for DOA estimation, and instead a suitable sub-optimal estimator must be selected. In order to select a suitable estimator from the large number of sub-optimal estimators that exist, the properties of both the signal and the surrounding environment must be considered. For difficult estimation problems, e.g. low SNR and/or small number of samples, it may be necessary to use any available *a priori* knowledge of the signal or environment in order to improve the estimator performance [75].

In this chapter, we consider three of the commonly used DOA estimation schemes: Maximum Likelihood (ML) Estimation (section 3.4), the Multiple Signal Classification (MUSIC) technique (section 3.3.1) and Time-Delay Estimation (TDE) (section 3.6). The characteristics of these methods are examined and the advantages and disadvantages pertaining to each are established. We then evaluate the performance of the algorithms for various localization situations, and comparisons are made using results from both simulations and experimental recordings. In all cases, we assume the number of sources

N_s is known. A full discussion of estimating N_s is given in chapter 4.

5.1.1 Maximum Likelihood (ML) Estimator

As discussed in chapter 3 the Maximum Likelihood (ML) estimates are calculated as the values of the unknown parameters that maximize the likelihood function, or put another way ML estimators select the set of parameters that make the observed data most likely.

Using the expression given in equation (3.58) an estimate $\widehat{\theta}$ is found for each individual frequency component within the range of interest. The estimate for each frequency component is then weighted by the magnitude of the singular values at this frequency [33], and the overall ML estimate $\widehat{\theta}_{ML}$ is then given by:

$$\widehat{\theta}_{ML}(f) = \arg\min_{\theta} Tr\left\{\psi(f)\,\mathbf{P}_{\mathbf{A}}^{\perp}(f,\theta)\,\widehat{\mathbf{R}}_{XX}(f)\right\}, \tag{5.1}$$

where, $\psi(f)$ is the weighting found from the trace of the estimated covariance matrix:

$$\psi(f) = \sum_{i=1}^{P} \lambda_i(f). \tag{5.2}$$

The resolution that can be achieved using ML techniques depends on the grid, or range of values of θ, over which equations (5.1) are calculated. This makes the ML approach computationally very expensive, in particular for the case of $P > 1$ sources, which results in a $P-$dimensional search as discussed in chapter (3). In the following the Alternating Projection (AP) algorithm is used in order to break the multi-dimensional parameter search into a sequence of single-parameter searches, resulting in faster convergence rates than iterative methods. Defining the ML ambiguity function:

$$J(\theta) = Tr\left\{\psi(f)\,\mathbf{P}_{\mathbf{A}}^{\perp}(f,\theta)\,\widehat{\mathbf{R}}_{XX}(f)\right\}, \tag{5.3}$$

the steps of the algorithm are as follows for the case of two sources [106]:

1. Estimate the DOA of the stronger source using a single source grid, (i.e. in the same manner as when a single source is present)
$$\widehat{\theta}_1^{(0)} = \arg\min_{\theta_1}\{J(\theta_1)\}. \tag{5.4}$$

2. Estimate the DOA of the weaker source on a single-source grid under the assumption of a two-source model while keeping the first source location estimate from Step 1 constant
$$\widehat{\theta}_2^{(0)} = \arg\min_{\theta_2}\{J([\theta_1,\theta_2])\}. \tag{5.5}$$

3. A grid search is then performed to find the DOA of the first source while keeping the estimate of the second source DOA from Step 2 constant

$$\widehat{\theta}_1^{(i)} = \arg\min_{\theta_1} \left\{ J\left([\theta_1, \theta_2]\right) \right\}. \tag{5.6}$$

4. Holding the estimate of the first source DOA found in Step 3 constant, a grid search is then performed to find the next estimate of the second source DOA.

$$\widehat{\theta}_2^{(i)} = \arg\min_{\theta_2} \left\{ J\left([\theta_1, \theta_2]\right) \right\}. \tag{5.7}$$

These steps are then repeated until convergence.

5.1.2 Multiple Signal Classification (MUSIC)

Multiple Signal Classification or MUSIC [21] is the best known of the subspace estimation algorithms, as discussed in section 3.3.1. The DOA estimate is found by maximizing the weighted MUSIC ambiguity function often referred to as the MUSIC "spatial spectrum":

$$\widehat{\theta}_{MUSIC}\left(f\right) = \arg\max_{\theta} \left\{ \psi\left(f\right) \mathbf{P}_{MUSIC}\left(f, \theta\right) \right\}, \tag{5.8}$$

where $\psi\left(f\right)$ is given by (5.2) $\mathbf{P}_{MUSIC}$ is defined in (3.31).

This method is easily extended to the case of P sources, by selecting the P values of θ that maximize equation (5.8), and therefore the presence of multiple sources does not significantly increase the computational complexity of the algorithm.

The performance of the MUSIC algorithm approaches that of the ML method as the number of snapshots increases to infinity, and for uncorrelated signals the MUSIC estimator is efficient as long as M is also large [22, 23]. The advantage of the MUSIC technique is that it offers significant computational savings when compared to the ML estimator, and a large increase in resolution when compared with TDE techniques. Unfortunately however, MUSIC can show large bias in the case of a finite number of samples and low SNR, leading to an inability to resolve closely spaced sources. The performance of the method degrades when the signals are correlated, and fails when the signals are coherent. For a more detailed discussion on the characteristics of MUSIC see chapter 3.

5.1.3 Time Delay Estimation (TDE)

As discussed in section 3.6, Time Delay Estimation (TDE) techniques provide a computationally inexpensive method of estimating the location of a source, based on estimation

of the time delay between the signal arriving at a pair of microphones, and this approach is directly applicable to broadband signals.

This chapter considers the performance of the Phase-Correlation or PHAT technique defined in equation (3.96).

The estimate of the Time Difference of Arrival (TDOA) for each microphone pair is given by equation (3.93), and for a Uniform Linear Array (ULA) the TDOA will be the same for each microphone pair. The overall TDOA estimate is then found, by adding the weighted correlation for each microphone pair and selecting the delay value that maximises the sum of these individual correlations:

$$\widehat{\delta\tau} = \arg max_{\delta\tau} \left[\sum_{m=1}^{M-1} R_{X_m, X_{m+1}} (\delta\tau) \right] \tag{5.9}$$

This delay estimate is then substituted into equation (3.94) to find the corresponding DOA estimate.

Multiple sources will result in separate peaks in the cross-correlation function, and therefore the technique is easily extended to the multiple source case. However for the case of multiple sources cross-correlation based techniques do not allow for distinction between multiple sources with similar delay values, as the presence of several close impulses leads to a spread in the peak of the correlation function.

The biggest disadvantage of TDE techniques is that the resolution is strictly limited by the temporal sampling frequency and the microphone spacing. The delays found can only be equal to an integer number of samples, therefore the maximum resolution that can be obtained is equivalent to the number of degrees corresponding to one sample:

$$\widehat{\theta}_{RES} = \cos^{-1} \left\{ \frac{v}{f_s \Delta} \right\}. \tag{5.10}$$

This quantization effect results in a severe limitation on the resolution achievable at lower sampling frequencies, in particular when the inter-microphone spacing Δ is small. This effect becomes more pronounced as the DOA increases, as will be seen in section 5.2.

5.2 Simulation Results

Initially, the performance of the three localization methods are compared using simulations of the scenario of a uniform linear array of five omnidirectional microphones. The distance between the microphones is $3.4cm$, which is equal to half the minimum wavelength (taking the maximum frequency $f_{\max} = 5kHz$). The sampling frequency used is $f_s = 16kHz$,

the speed of sound is taken to be $v = 340m/s$ and the number of samples is $N = 100$, $(6.25ms)$, with a 50% overlap between frames.

5.2.1 Single Source

Initially the situation of single source emitting a speech signal is considered. The source is located at *i*) 20^o, *ii*) 70^o and *iii*) 90^o, for SNR increasing from $-10dB$ to $30dB$. For each SNR, 100 Monte Carlo trials are run and the average DOA estimate and the MSE are found. These results are shown in figure 5.1.

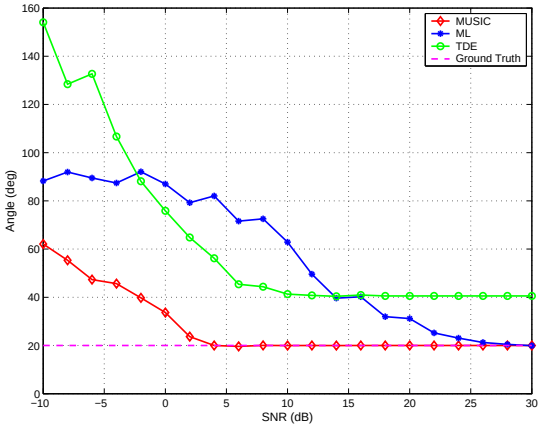
From the results it can be seen that for SNR values of above $0dB$ the results from the estimators do not change, and the MSE is constant for increasing SNR. However, below the threshold point of $0dB$ the MSE increases rapidly with decreasing SNR.

The MSE is greatest for $DOA = 20^o$ and smallest for $DOA = 90^o$. This is due to the fact that sources perpendicular to the array are easiest to locate and as the angle decreases localisation becomes more difficult. The true angle of arrival can therefore be seen to have an effect on the estimator performance. This is a characteristic of a ULA and is not the case for a circular array where the minimum variance that can be achieved is independent of the DOA [106].

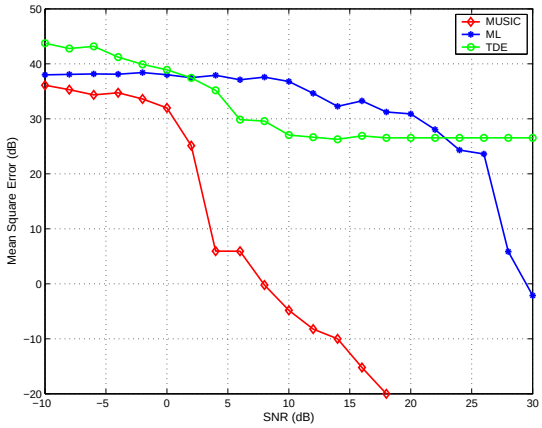
For all SNR values above the threshold SNR, the MUSIC and ML estimators both estimate the true DOA. As these methods are not based on an initial time delay estimate, an arbitrary accuracy can be achieved and is dependent only on the number of samples being sufficiently large, and the signal model sufficiently accurate. The performance of these two estimators is similar, with the ML method having slightly lower MSE, which is expected as MUSIC is known to have the same large sample accuracy as the ML estimator [22, 107].

The TDE results can be seen to be limited by the temporal sampling frequency f_s, and the inter-microphone spacing. This is because the time delay can only be estimated to within an accuracy of one sample (equation (5.10)) corresponding in this case to an angle of 51.32^o. The introduction of this quantization error means that even for the cases when the time-delay is correctly estimated, it is only possible to convert this to a precise DOA estimate if the DOA happens to fall on one of the quantized delay values. However, the low computational cost of TDE make it an attractive approach for situations where a rough location estimate is sufficient.

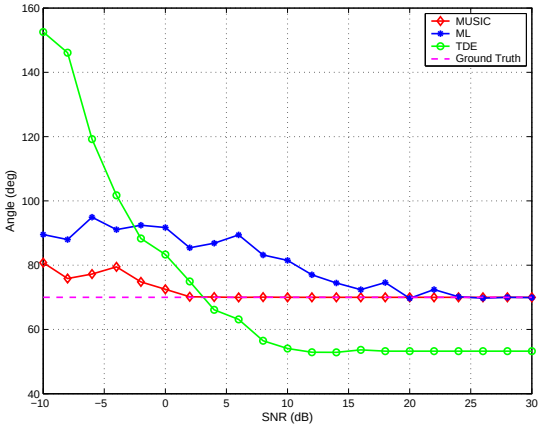
In figures 5.2 - 5.4 the spatial spectra (cf. chapter 3) of the three estimators at different SNR levels is examined. In each case the arrows indicate the true DOA estimate. It can be seen that for $SNR = 30dB$ the spatial spectra show clear global maxima (or a global minimum in the case of the MLE). All the estimators will find the true DOA value at this



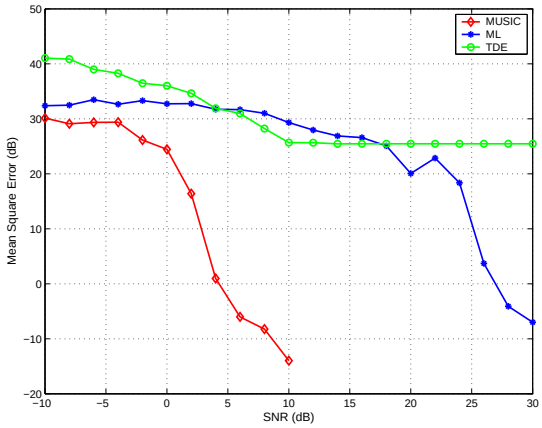
(a) DOA estimates, $DOA = 20^o$.



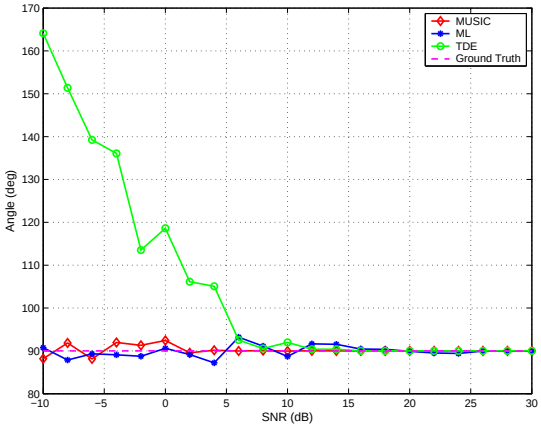
(b) Mean Square Error, $DOA = 20^o$..



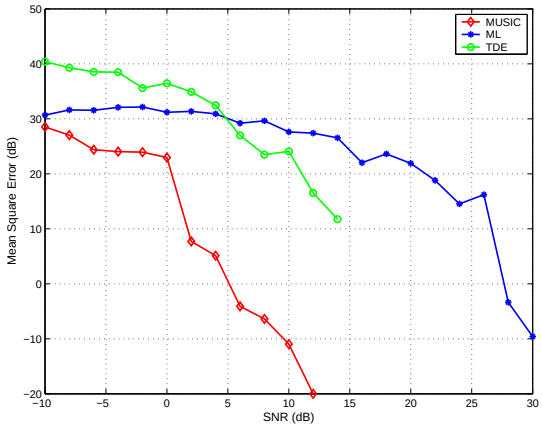
(c) DOA estimates, $DOA = 70^o$.



(d) Mean Square Error, $DOA = 70^o$.

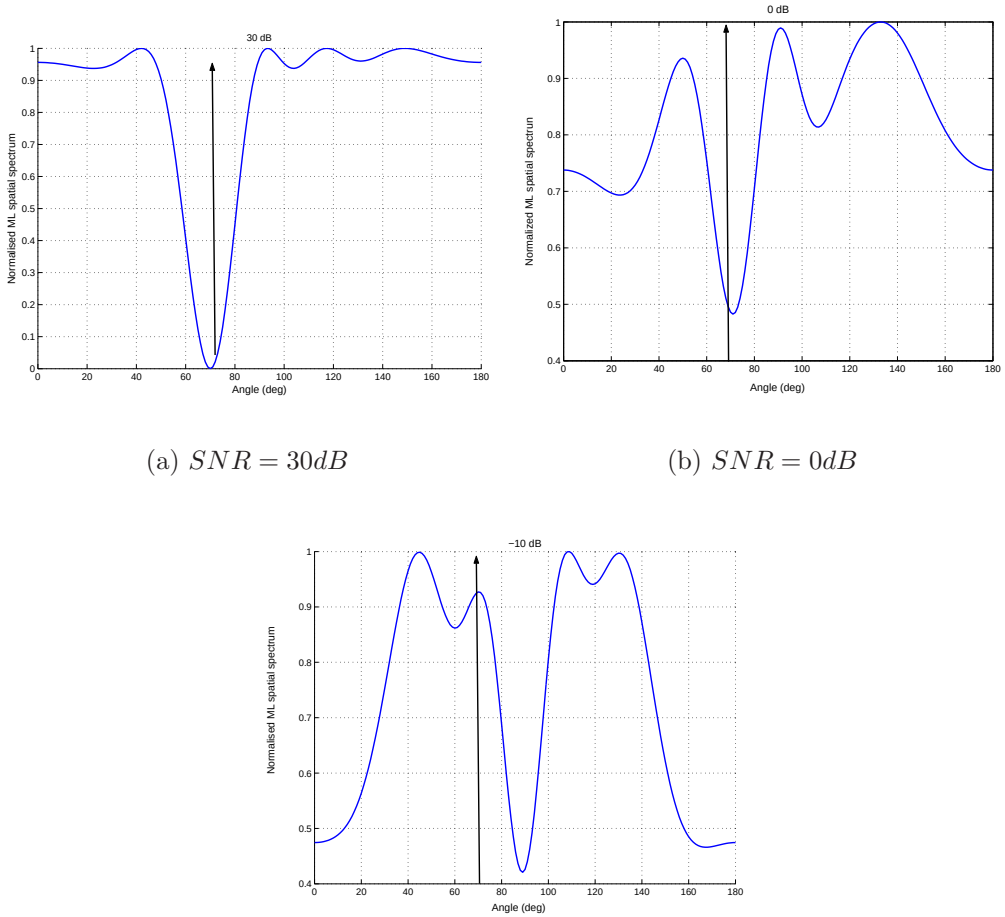


(e) DOA estimates, $DOA = 90^o$.



(f) Mean Square Error, $DOA = 90^o$.

Figure 5.1: *Performance of MUSIC, MLE and TDE techniques, for a speech signal arriving at an array of $M = 5$ microphones and window length $N = 50$ samples. Results are taken over 100 Monte Carlo trials for increasing SNR.*



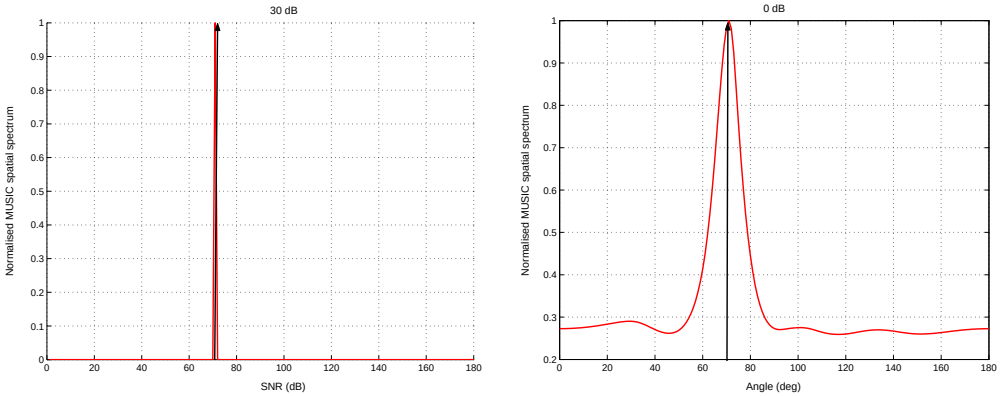
(a) $SNR = 30dB$

(b) $SNR = 0dB$

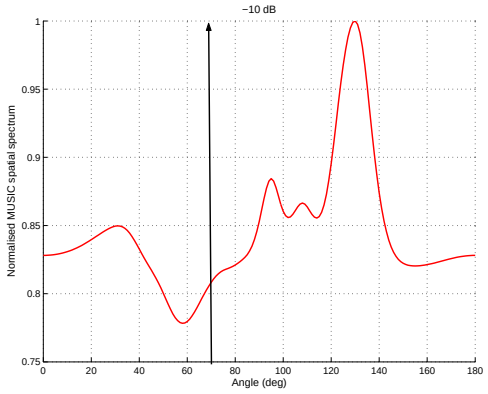
(c) $SNR = -10dB$

Figure 5.2: *The normalised ML spatial spectrum for decreasing SNR (equation (5.1)). The true DOA value is indicated by the arrow.*

SNR. When the SNR is decreased to $0dB$, the width of the mainlobe increases, and the magnitude of the sidelobes can be seen to begin to increase. However, the mainlobe still lies close to the true DOA value and the estimates will accordingly be close to the true value, and operation can be said to be in the small-error region as the MSE will be small (see chapter 6 for a full discussion on regions of operation). However, when the SNR is reduced to $-10dB$ the magnitude of the sidelobes increase while the mainlobe decreases. The DOA estimates from all three estimators will no longer be close to the true DOA value and we can see that this corresponds to the area of larger MSE seen for SNR values below the threshold point of $0dB$ in figure 5.1.

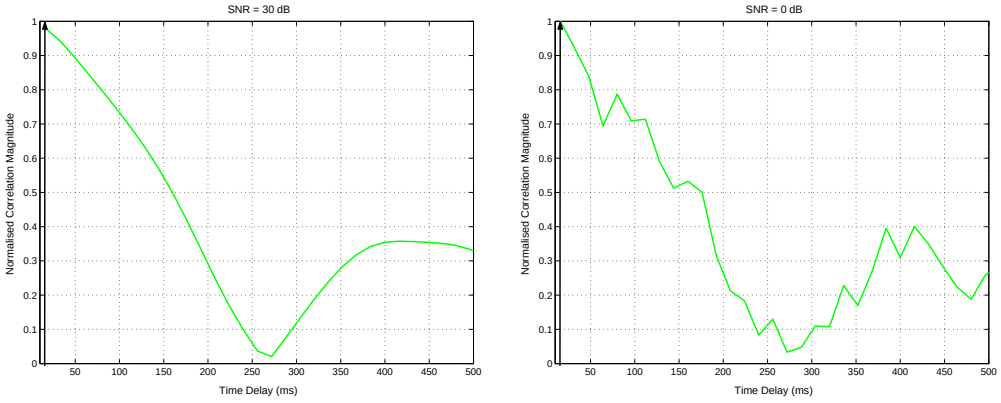


(a) $SNR = 30dB$ (b) $SNR = 0dB$

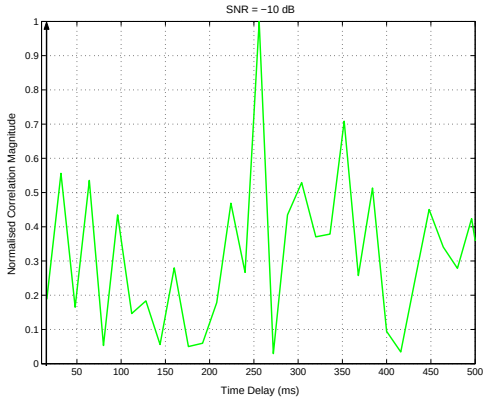


(c) $SNR = -10dB$

Figure 5.3: *The normalised MUSIC spatial spectrum for decreasing SNR (equation (5.8)). The true DOA value is indicated by the arrow.*



(a) $SNR = 30dB$ (b) $SNR = 0dB$



(c) $SNR = -10dB$

Figure 5.4: *The normalised TDE cross-correlation for decreasing SNR (equation (5.9)). The true DOA value is indicated by the arrow.*

5.2.2 Multiple Sources

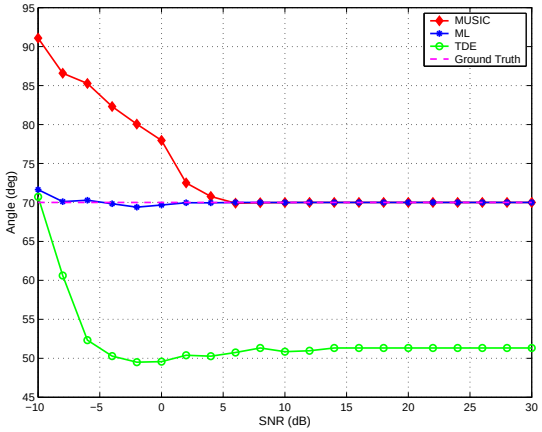
The situation of a two stationary sources emitting speech signals is now considered. The sources are located at 70^o and 110^o and the SNR is increased from $-10dB$ to $30dB$. For each SNR 100 Monte Carlo trials are run and the average DOA estimate and the MSE are found. These results are shown in figure 5.5. Once again we can see from the results that, for high SNR, the MUSIC and ML techniques accurately estimate the DOA, while the accuracy of the TDE results is reduced due to resolution problems. In this case it can be seen that the threshold SNR is higher than for the case of a single source. This is due to the fact that as the SNR decreases the peaks become less distinct from each other as the lobes around the true values become wider.

The ability of the estimators to correctly estimate the DOA as the sources approach each other is now investigated. In this case the SNR is held constant at $10dB$. To begin with, the sources are located at 5^o and 180^0, and then they are moved towards each other simultaneously in steps of 10^o (i.e. the sources are 20^o closer after each step), until they are a minimum distance of 15^o from each other. The results of the three methods are shown in figures (5.6). As the sources move closer together, the advantage of the MUSIC and ML methods becomes obvious, as at higher SNR levels and for sources with large angle DOAs, these methods continue to accurately estimate the true DOA even when the sources are separated by only 15^o. However, when the SNR is reduced the ability of all the techniques to accurately resolve the sources is reduced as the spectral peaks become broader and merge together. Once again it can be seen that sources located at small angles to the array are more difficult to estimate.

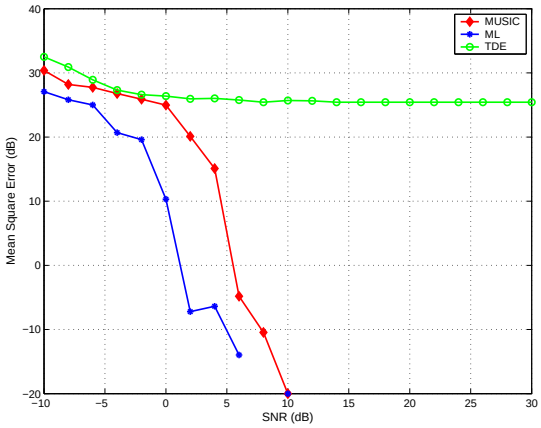
5.2.3 Room Response Simulations

The performance of the estimators in the presence of reverberation is considered using simulated impulse responses found from Enhanced Acoustic Simulator for Engineers^{TM} (EASE) simulation software. The response is found for a room with dimensions $5.38m \times 6.9m \times 2.44m$, containing desks, chairs and a white board at the top wall. We compare the results of the estimators over 100 Monte Carlo trials as the Useful-to-detrimental ratio, U_{25} (as defined in equation (4.34)), is increased.

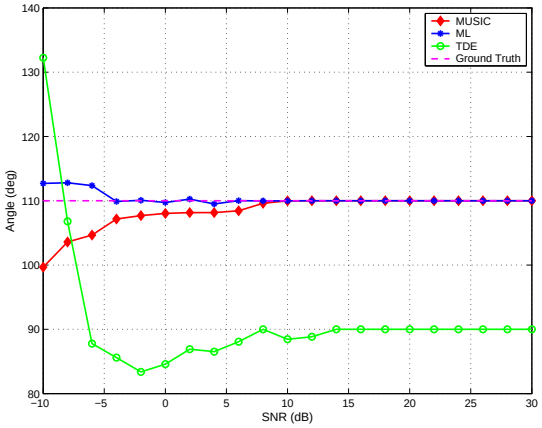
The results are shown in figure 5.7 for the case of a single source emitting a white noise signal with $DOA = 70^o$. The presence of reverberation can be seen to lead to a slight reduction in accuracy of the MUSIC and ML estimators compared to the earlier cases. In particular, the MUSIC algorithm can be seen to suffer from the presence of reverberation and does not find the exact source location. The TDE results, above the threshold point



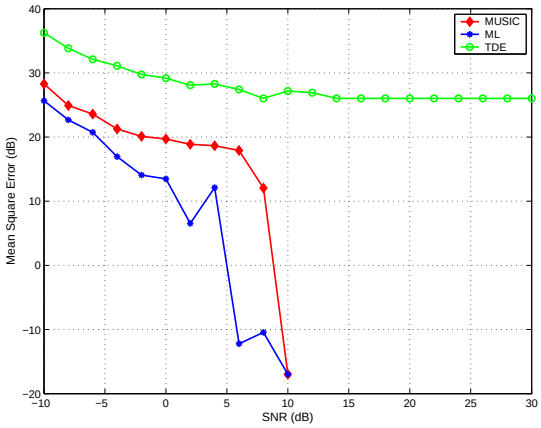
(a) DOA estimates, $DOA = 70^o$.



(b) Mean Square Error, $DOA = 70^o$.

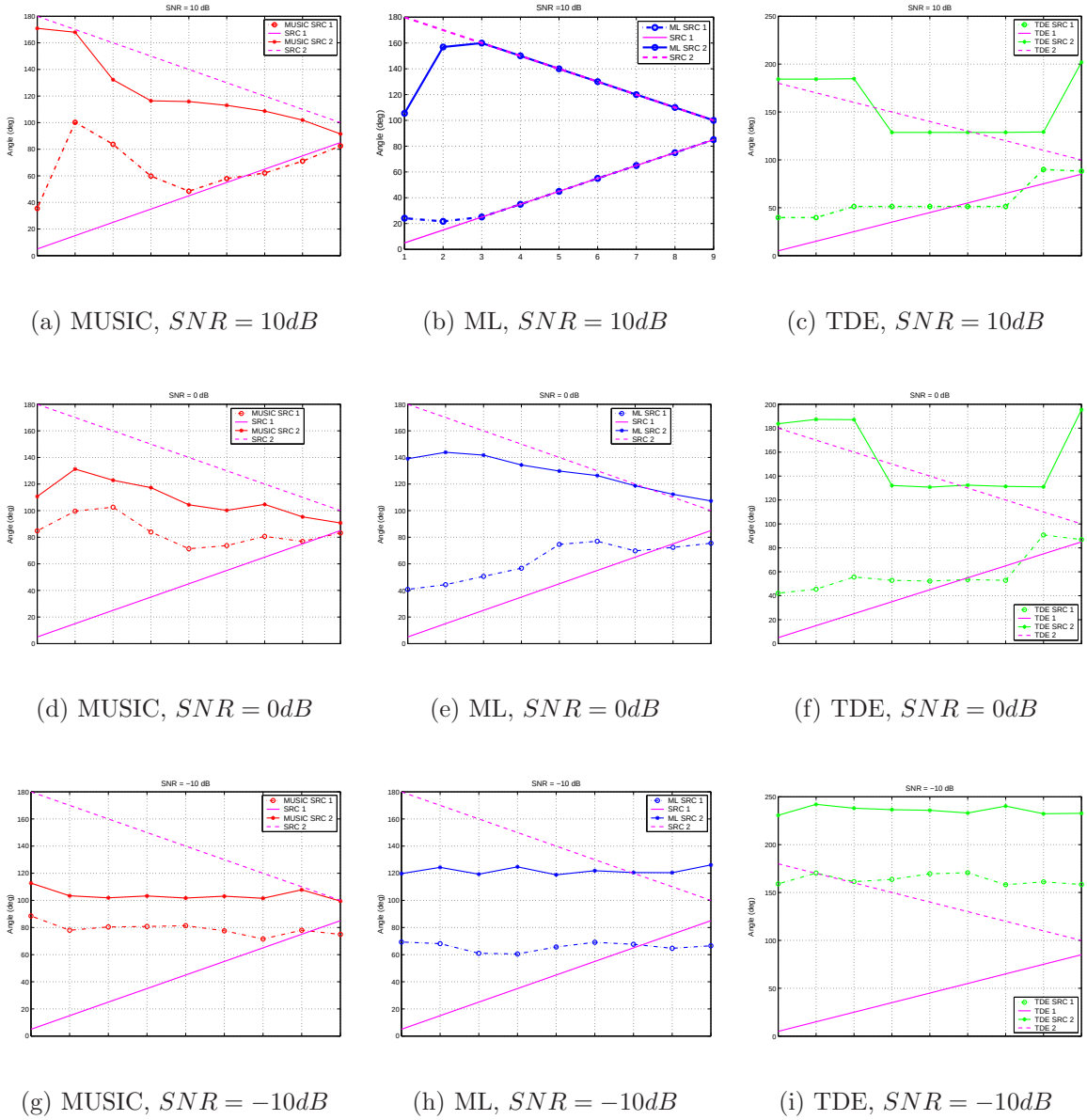


(c) DOA estimates, $DOA = 110^o$



(d) Mean Square Error, $DOA = 110^o$

Figure 5.5: *Performance of MUSIC, MLE and TDE techniques, when 2 speech signals arrive at an array of $M = 5$ microphones, and window length $N = 50$ samples. The results are averaged over 100 Monte Carlo trials for increasing SNR.*

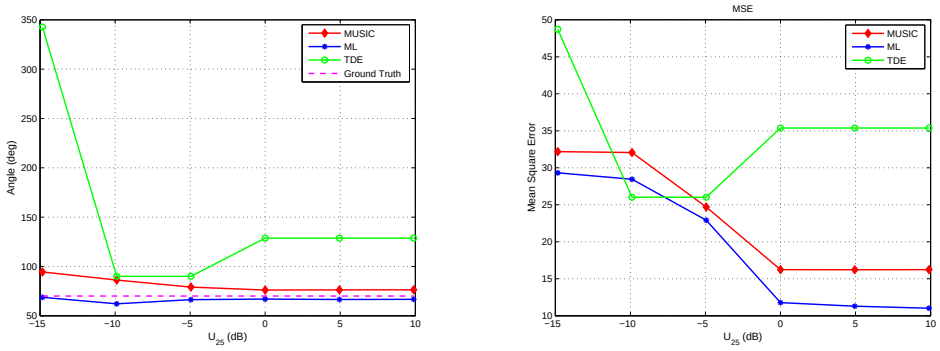


(a) MUSIC, $SNR = 10dB$ (b) ML, $SNR = 10dB$ (c) TDE, $SNR = 10dB$

(d) MUSIC, $SNR = 0dB$ (e) ML, $SNR = 0dB$ (f) TDE, $SNR = 0dB$

(g) MUSIC, $SNR = -10dB$ (h) ML, $SNR = -10dB$ (i) TDE, $SNR = -10dB$

Figure 5.6: *Results of MUSIC, MLE and TDE techniques, for estimating the DOA of 2 white noise sources arriving at an array of $M = 5$ microphones, and window length $N = 100$ samples. The sources move toward each other in steps of 10^o, and the results are averaged over 100 Monte Carlo trials at each position.*



(a) DOA estimates.

(b) Mean Square Error.

Figure 5.7: *Performance of MUSIC, MLE and TDE techniques, for the case of a single speech signal arriving at an array of $M = 5$ microphones, with window length $N = 50$ samples. The DOA is equal to 70^o and the results are shown for increasing U_{25}, with averages taken over 100 Monte Carlo trials.*

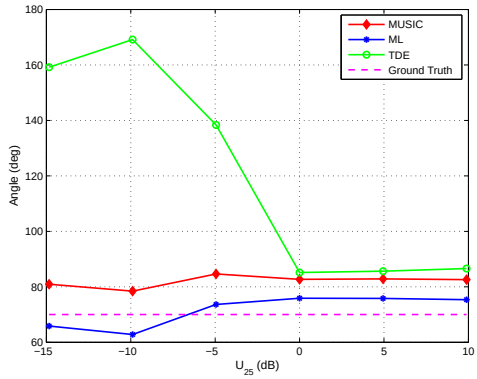
of $U_{25} = 0dB$, are similar to those in the non-reverberant case. Once again, the main cause of the reduction in accuracy of the TDE approach is due to the limited resolution as a result of the low sampling frequency and the short distance between the microphones.

Next the performance of the estimators is examined when there are two sources present. The results are shown in figures (5.8). It can be seen that the presence of reverberation leads to a large degradation in the performance of the estimators when compared to the case of additive noise only, shown in figure 5.5. From comparison with the results for the case of a single source, shown in (5.7), it can be seen that the presence of a second source only results in a small reduction in accuracy, with the MUSIC algorithm showing the greatest deterioration in performance. This can be attributed to the early arriving reflections (in this case reflections arriving in the first $25ms$) causing spreading of the mainlobe.

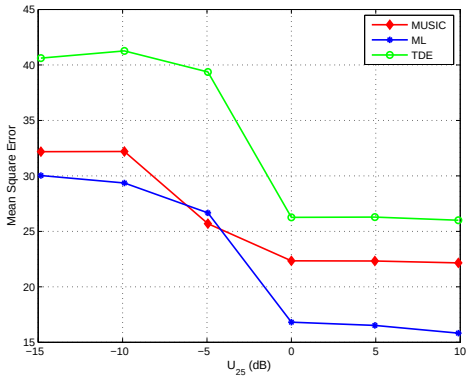
5.3 Experimental Results

5.3.1 Experimental Setup

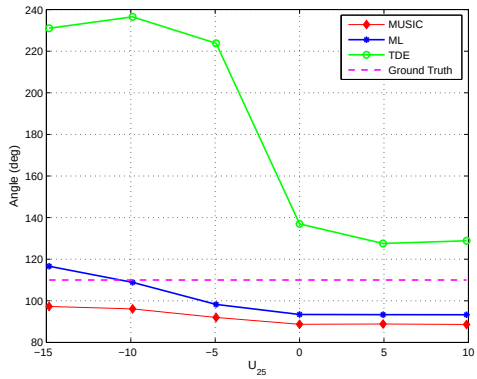
A microphone array of 5 omni-directional Rhode microphones were arranged in a Uniform Linear Array (ULA) with inter-microphone spacing of $3.4cm$. The experimental setup is the same as was modelled by the room simulation software in the previous section, with environmental noise coming from a room fan located in the ceiling and traffic noise from



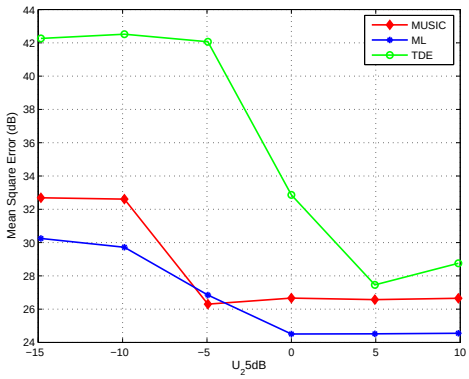
(a) DOA estimates, DOA=70^o



(b) Mean Square Error, DOA=70^o



(c) DOA estimates, DOA=110^o



(d) Mean Square Error, DOA=110^o

Figure 5.8: *Performance of MUSIC, MLE and TDE techniques, for two speech signals arriving at an array of $M = 5$ microphones, with window length $N = 100$ samples. The DOAs are equal to 70^o and 110^o, and the results shown here are the average estimates taken over 100 Monte Carlo trials for increasing U_{25}.*

outside the window. Once again the reverberation time $T_{60} \approx 0.5s$. For all cases the height of the microphone array and the sources were equal, and the horizontal distance between the sources and the array is sufficient to allow for the sources to be considered far-field, resulting in planar wavefronts arriving at the array as discussed in chapter (2).

5.3.2 Single Source

We consider the case of a single source emitting white noise, located at a distance of $2.5m$ from the microphone array, with and DOAs 90^o and 120^o. The DOA estimation results, using a window length of $N = 100$ samples, are shown in tables (5.1(a)) and (5.1(b)). It can be seen that all three methods perform well for a DOA of 90^o, with the TDE approach performing the best, as it correctly estimates the true DOA value for each window. This corresponds to the earlier simulations where it was seen that the accuracy achieved by the estimators was dependent on the true DOA value, and the highest accuracy was achieved for $DOA = 90^o$.

The performance of all three estimators deteriorates for $DOA = 120^o$, as seen in table 5.1(b). In particular the average result of the MUSIC algorithm is seen to differ from the true DOA by approximately 20^o. The ML estimator is seen to have minimum bias. While the TDE average result is farther from the true DOA than that of the ML, the low MSE of the TDE results indicate that the difference between the true DOA value and the results is mainly due to the limited resolution capabilities of the TDE approach.

From these results it can be seen that for situations where a robust estimate is required from each data window, the TDE approach is the most appropriate, provided high resolution is not required. For situations where high resolution results are crucial, the ML method is more suitable. However, the computational time required by the ML method, and the need to average over a number of estimates makes it undesirable for most practical applications.

5.3.3 Multiple Sources

We now consider the performance of the estimators using experimental recordings for the situation of two stationary sources. Firstly, using the same experimental set-up as for the single source case we consider the case of two uncorrelated white noise sources located at 90^o and 120^o. The results are shown in table (5.2).

As expected from the results in the case of a single source (tables 5.1(a) and 5.1(b)) we see that all three estimators accurately estimate the source with $DOA = 90^o$, with the MUSIC algorithm showing just a very slight deterioration. However, the source with

Estimator	**Mean DOA Estimate** $DOA = 90^o$	**Root Mean Square Error (RMSE)**
MUSIC	89.81^o	0.5^o
ML	91.29^o	3.7^o
TDE	90^o	-

(a) $DOA = 90^o$

Estimator	**Mean DOA Estimate** $DOA = 120^o$	**Root Mean Square Error (RMSE)**
MUSIC	101.64^o	11.64^o
ML	116.24^o	35.46^o
TDE	128.68^o	38.68^o

(b) $DOA = 120^o$

Table 5.1: *Results found by MUSIC, ML and TDE for experimental recordings of a white noise source with* $DOA = 120^o$*. The results are averaged over* 100 *Monte Carlo trials using a window length of* $N = 100$ *samples.*

Estimator	**Mean Estimate** $DOA = 90^o$	**Mean Estimate** $DOA = 120^o$	**Root Mean Square Error (RMSE)** $DOA = 90^o$	**Root Mean Square Error (RMSE)** $DOA = 120^o$
MUSIC	89.20^o	92.62^o	1.45^o	27.35^o
ML	88.28^o	113.92^o	21.03^o	23.33^o
TDE	90^o	128.68^o	0^o	8.68^o

Table 5.2: *Results found by MUSIC, ML and TDE for experimental recordings of 2 white noise sources with DOAs of* 90^o *and* 120^o*. The results are averaged over* 100 *Monte Carlo trials using a window length of* $N = 100$ *samples.*

$DOA = 120^o$ proves more problematic for the MUSIC algorithm, and the accuracy of the DOA estimates for this source are seen to be further reduced by the presence of the second source. Compared to the results for the single source, the ML method shows the greatest deterioration due to the presence of the second source. The TDE estimates are again very close to the true DOA values, and the source with $DOA = 90^o$ is correctly estimated for each window. The error in the estimates of the source with $DOA = 120^o$ is due to the restricted resolution that is characteristic of the TDE approach.

Then 2 speech sources (two different male speakers) were placed at angles of 70^o and 110^o from the array. The results in table 5.3 show that the accuracy of all three of the estimators is reduced for this case.

The estimates of the MUSIC algorithm indicate that it continuously fails to discriminate between the sources, which can be attributed to the spreading of the peaks corresponding to each of the sources in the MUSIC spatial spectrum.

While the ML method shows better discrimination between the sources, the accuracy of the mean estimate is significantly lower than that of the previous case. However, the total MSE is approximately the same for both cases. The TDE method is clearly the most suited to this situation as the estimates remain close to the true DOA values for both sources.

Estimator	**Mean Estimate** $DOA = 70^o$	**Mean Estimate** $DOA = 110^o$	**Root Mean Square Error (RMSE)** $DOA = 70^o$	**Root Mean Square Error (RMSE)** $DOA = 110^o$
MUSIC	85.47^o	92.1^o	16.06^o	17.01^o
ML	73.51^o	109.85^o	9.17^o	7.81^o
TDE	74.24^o	114.75^o	19.46^o	19.58^o

Table 5.3: *Results found by MUSIC, ML and TDE for experimental recordings of 2 male speakers with DOAs of* 70^o *and* 110^o*. The results are averaged over* 100 *Monte Carlo trials using a window length of* $N = 100$ *samples.*

5.4 Conclusions

We have considered three of the most well-known approaches, MUSIC, ML and TDE, for estimation of the DOA of one or more sources. The performance of these methods was compared using both computer simulations and experimental recordings, allowing us to draw several conclusions on the comparative behaviour of these estimators. Firstly it was seen that once the SNR decreases below a threshold point the MSE increases dramatically. The SNR value of this threshold point was shown to depend on the number of sources, and the true value of the DOA. Examination of the spatial spectra of the three estimators showed that once the SNR decreases below the threshold level, the sidelobes can be greater than the mainlobe, and the width of all lobes increases.

From the results of the different simulations it could be seen that the resolution achievable by the ML and MUSIC methods is much greater than that of the TDE approach. However, the TDE approach was shown to be more robust to reverberation and the number of samples available, while also being computationally far less demanding. This makes it suitable for situations where the estimator must be robust to moderate levels of reverberation and low SNR, and a rough estimate of the DOA is sufficient, where rough means that the DOA can only be estimated to within a certain range determined by the temporal sampling frequency. The ML algorithm has the advantage of much higher resolution capabilities than the TDE method, and is seen to achieve higher accuracy than

the MUSIC algorithm in the presence of reverberation. However, the disadvantage here is the large number of calculations needed, particularly for the case of multiple sources. The MUSIC algorithm also has the advantage of high resolution capabilities without the computational expense of the ML method. Unfortunately however, its performance has been seen to deteriorate significantly in the presence of reverberation, particularly for smaller window lengths when there are multiple sources present.

The novel contributions made in this chapter include the comparison of the performance of three of the classical DOA estimation methods for a number of different situations using both computer simulations and experimental recordings. These comparisons considered the effects of both varying SNR and Useful-to-Detrimental ratio on the performance of the estimators.

6

Lower Bounds on the Mean Square Error (MSE) of an Estimator

6.1 Introduction

For any acoustic source loclisation scheme, knowledge of the best possible performance that can be achieved by the estimator is very important. This knowledge can be used to determine the operating range of the estimator and therefore to establish whether or not it is possible to meet operational requirements, thus making it a crucial step in localisation feasibility studies. Comparison of the best possible performance with the performance of a proposed estimator can also be used to see if, starting from a particular algorithm, an improvement in performance is possible. Lower bounds on the Mean Square Error (MSE) allow for calculation of the best performance that can be achieved in the MSE sense, when estimating a parameter of a signal corrupted by noise.

Historically the first MSE lower bound for deterministic parameters to be derived was the Cramér Rao Bound (CRB) [8, 108] and since then it has remained by far the most widely used for evaluation of acoustical estimator performance [106, 109]. Its popularity is largely due to its computational simplicity (in comparison to other lower bounds); the fact that in many cases it can be achieved asymptotically (high SNR [110] and/or large

number of snapshots [8]) by the ML estimator [7]; and last but not least, its noticeable property of being the lowest bound on the MSE of unbiased estimators. This last property is due to the fact that the CRB derives from the *weakest* formulation of unbiasedness, that is, non-bias local to (i.e. in the vicinity of) the true value of the parameter to be estimated [111–113].

However, at low Signal to Noise Ratio (SNR), DOA estimators, such as those discussed in chapter 5, suffer from severe ambiguity problems. This is because the ambiguous peaks in the estimator ambiguity function (or estimator spectrum - equations (3.31,5.3), which are located far from the true location, become larger as the SNR decreases (as seen in figures 5.2-5.4). When the estimator can no longer distinguish between these ambiguous peaks and the peak at the true source location, the MSE deviates rapidly from the CRB [114], and therefore the CRB no longer provides a meaningful performance bound. The SNR value at which this deviation occurs is called the SNR threshold, and this threshold is clearly visible in the results presented in chapter 5.

The initial characterization of locally unbiased estimators given by the CRB was firstly improved by Bhattacharyya's works [7, 111, 115] which refined the characterization of local unbiasedness, and significantly generalized by Barankin's works [111], who established the general form of the greatest lower bound of any absolute moment of an unbiased estimator. In the particular case of MSE, Barankin's work allows the derivation of the greatest lower bound on MSE since it takes into account the *strongest* formulation of unbiasedness, that is to say unbiasedness over a continuous interval of parameter values including the true parameter value. Unfortunately, however the Barankin Bound (BB) is generally not computable [116].

Therefore, since then, numerous works [112, 113, 117, 118] have been devoted to deriving computable approximations of the BB. These works have shown that in non-linear estimation problems, such as DOA estimation, three distinct regions of operation can be observed as seen in figure 6.1. In the asymptotic region, the MSE is small and, in many cases, close to the Small-Error bounds (CRB). In the *a priori,* or maximum MSE performance region where the number of independent snapshots and/or the SNR are very low, the observations provide little information and the MSE is close to that obtained from the prior knowledge of the problem. Between these two extremes, there is an additional ambiguity region, called the transition region. In this region, the MSE of ML estimation results usually deteriorates rapidly with respect to Small-Error bounds and exhibits a threshold behavior corresponding to a "performance breakdown" highlighted by Large-Error bounds (BB) [113, 118, 119], and seen in experimental results (chapter 5).

However, when compared to estimators such as the ML estimator, it can be seen that

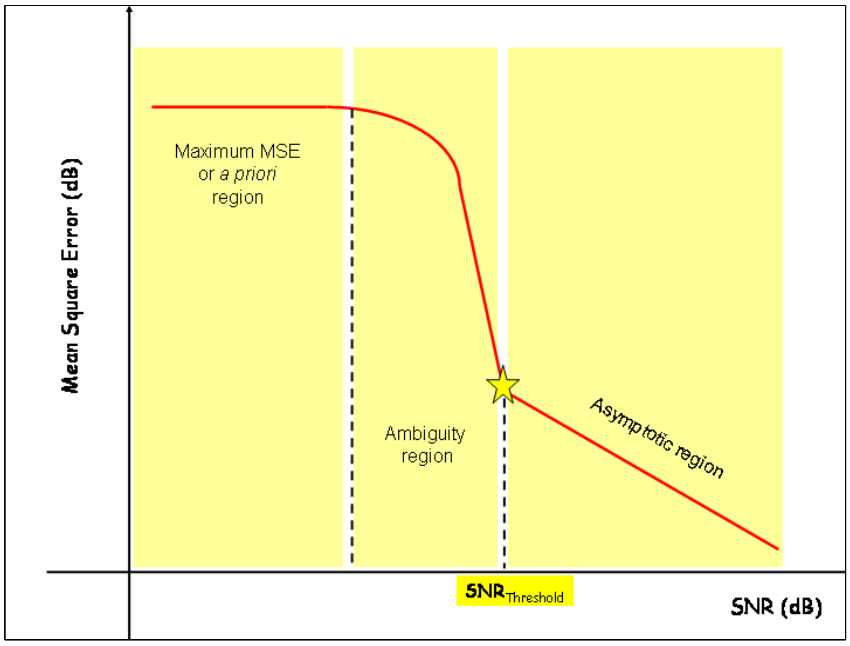


Figure 6.1: *Three regions of operation observed for a non-linear estimator.*

the performance characterisation provided by existing approximations of the BB differs considerably from the true estimator performance, particularly in prediction of the SNR threshold. Therefore the search for an easily computable approximation of the BB, that is closer to the true performance achievable, is still a subject worth investigation. Indeed, from a theoretical standpoint, the accurate knowledge of the BB should allow a better prediction of the SNR value at the start of the transition region and avoid misleadingly optimistic conclusions on achievable performance being drawn from the computation of the CRB at low SNR. Moreover, from a practical standpoint, it is also the SNR value at which estimator performance will "break down".

In this chapter, a new formalism is presented that allows not only the derivation of a general class of Barankin Bound approximations, but also gives a clear and straightforward interpretation of these approximations. Firstly, it will be shown that this formalism includes all previously derived bounds, and provides a meaningful way to classify these bounds. Secondly, with the help of this formalism a new practical approximation of the BB is derived, whose computational complexity is comparable to that of the CRB.

The new BB approximation is then applied to the problem of estimating the DOA of a single source emitting a signal that is received by a Uniform Linear Array (ULA) of microphones. From the results it can be seen that the bound produced by this new BB approximation provides a much closer prediction of the threshold SNR than existing approximations [112, 113, 115, 118]. However, the results also show that in the Maximum

MSE Region, the performance indicated by the new BB approximation is much worse than that achieved by the estimator, suggesting that deterministic Large Error bounds are not suitable for characterisation of the estimator performance below the SNR threshold. Instead, their use is limited to prediction of the SNR threshold.

6.2 Mean Square Error of an Estimator: an algebraic quantity

In the following, unless otherwise stated, $\mathbf{x}$ denotes the random observations vector and Ω is the observation space, where $\mathbf{x}$ consists of a signal source function of a single unknown deterministic real parameter θ and embedded in a nuisance signal (generally Gaussian noise). $f_\theta(\mathbf{x})$ denotes the Probability Density Function (pdf) of observations depending on $\theta \in \Theta$, where Θ denotes the parameter space.

First let us recall that, if $\widehat{g(\theta_0)}(\mathbf{x})$ is an unbiased estimator of $g(\theta_0)$ where $g(\ )$ is a real function and θ_0 is a selected value of θ, then $MSE_{\theta_0}\left[\widehat{g(\theta_0)}\right] = Var\left[\widehat{g(\theta_0)}\right]$, as discussed previously in chapter (2), and therefore finding the lower bound on the MSE is equivalent to finding the lower bound on the estimator variance. Secondly, it is worth noting that the MSE is an algebraic quantity, in the sense that it can always be interpreted as a norm associated with a scalar product, where the MSE can be expressed as:

$$\begin{aligned} MSE_{\theta_0}\left[\widehat{g(\theta_0)}\right] &= E_{\theta_0}\left[\left(\widehat{g(\theta_0)}(\mathbf{x}) - g(\theta_0)\right)^2\right] && (6.1)\\ &= \int_\Omega \left(\widehat{g(\theta_0)}(\mathbf{x}) - g(\theta_0)\right)^2 f_{\theta_0}(\mathbf{x})\, d\mathbf{x} && (6.2)\\ &= \left\|\widehat{g(\theta_0)}(\mathbf{x}) - g(\theta_0)\right\|_{\theta_0}^2, && (6.3) \end{aligned}$$

and the scalar product is given by:

$$\langle g(\mathbf{x}) | h(\mathbf{x})\rangle_{\theta_0} = E_{\theta_0}\left[g(\mathbf{x})\ \ h(\mathbf{x})\right] = \int_\Omega \left[g(\mathbf{x})\ \ h(\mathbf{x})\right] f_{\theta_0}(\mathbf{x})\, d\mathbf{x}. \quad (6.4)$$

Among the many results established on vector spaces with a scalar product, two fundamental results are especially important in the search for a lower bound on the MSE: the generalization of the Cauchy-Schwartz inequality to Gram matrices [8, 108, 118, 120] and the minimization of a norm under linear constraints [116, 117].

Generalized Cauchy-Schwartz inequality

Let $\mathbb{U}$ be a Euclidean vector space of any dimension (finite or infinite) on the field of real numbers $\mathbb{R}$ which has a scalar product $\langle \mid \rangle$.

The Gram matrix $\mathbf{G}\left(\mathbf{u}_{[1,N]}, \mathbf{c}_{[1,K]}\right) \in \mathbb{R}^{K\times N}$, associated with the two vector families $\mathbf{u}_{[1,N]} = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N)$ and $\mathbf{c}_{[1,K]} = (\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_K)$ of $\mathbb{U}$ is then defined by [121]:

$$\mathbf{G}\left(\mathbf{u}_{[1,N]}, \mathbf{c}_{[1,K]}\right)_{n,k} = \langle \mathbf{u}_k \mid \mathbf{c}_n \rangle = \begin{bmatrix} \langle \mathbf{u}_1 \mid \mathbf{c}_1 \rangle & \langle \mathbf{u}_2 \mid \mathbf{c}_1 \rangle & \ldots & \langle \mathbf{u}_N \mid \mathbf{c}_1 \rangle \\ \langle \mathbf{u}_1 \mid \mathbf{c}_2 \rangle & \langle \mathbf{u}_2 \mid \mathbf{c}_2 \rangle & \ldots & \langle \mathbf{u}_N \mid \mathbf{c}_K \rangle \\ \vdots & \vdots & \ldots & \vdots \\ \langle \mathbf{u}_1 \mid \mathbf{c}_K \rangle & \langle \mathbf{u}_2 \mid \mathbf{c}_K \rangle & \ldots & \langle \mathbf{u}_N \mid \mathbf{c}_K \rangle \end{bmatrix} \tag{6.5}$$

and if $\mathbf{G}\left(\mathbf{c}_{[1,K]}, \mathbf{c}_{[1,K]}\right)$ is invertible, i.e. if $\mathbf{c}_{[1,K]}$ is a free family, the following inequality holds [6]:

$$\mathbf{G}\left(\mathbf{u}_{[1,N]}, \mathbf{u}_{[1,N]}\right) \geq \mathbf{G}\left(\mathbf{u}_{[1,N]}, \mathbf{c}_{[1,K]}\right)^T \mathbf{G}\left(\mathbf{c}_{[1,K]}, \mathbf{c}_{[1,K]}\right)^{-1} \mathbf{G}\left(\mathbf{u}_{[1,N]}, \mathbf{c}_{[1,K]}\right) \tag{6.6}$$

Moreover, if $(\mathbf{u}_1, \ldots, \mathbf{u}_N, \mathbf{c}_1, \ldots, \mathbf{c}_K)$ forms a free family of $\mathbb{U}$, then the inequality (6.6) is a strict inequality. The Cauchy-Schwartz inequality is a special case of (6.6) in the form:

$$\mathbf{G}(\mathbf{u}_1, \mathbf{u}_1) \geq \mathbf{G}(\mathbf{u}_1, \mathbf{c}_1)\, \mathbf{G}(\mathbf{c}_1, \mathbf{c}_1)^{-1}\, \mathbf{G}(\mathbf{u}_1, \mathbf{c}_1)^T \quad \Leftrightarrow \quad \|\mathbf{u}_1\|^2 \geq \frac{\langle \mathbf{u}_1 \mid \mathbf{c}_1 \rangle^2}{\|\mathbf{c}_1\|^2} \tag{6.7}$$

When the scalar product $\langle \mid \rangle$ corresponds to an expectation, as is the case for the MSE (6.3), then the inequality (6.6) is generally referred to as the "covariance inequality" [118, 120].

Norm minimization under linear constraints

Let $\mathbf{c}_{[1,K]}$ be a free family of K vectors of $\mathbb{U}$ and $\mathbf{v} = (v_1, \ldots, v_K)^T$ a vector of $\mathbb{R}^K$. The problem of the minimization of $\|\mathbf{u}\|^2 \, \forall \mathbf{u} \in \mathbb{U}$ under the K linear constraints $\langle \mathbf{u} \mid \mathbf{c}_k \rangle = v_k$, $k \in [1, K]$ then has the solution [116, 117]:

$$\min_{\mathbf{u}} \left\{ \|\mathbf{u}\|^2 \right\} = \mathbf{v}^T \mathbf{G}\left(\mathbf{c}_{[1,K]}, \mathbf{c}_{[1,K]}\right)^{-1} \mathbf{v}, \quad \text{with the unique minimizer} \quad \mathbf{u}_{opt} = \sum_{k=1}^{K} \alpha_k \mathbf{c}_k \tag{6.8}$$

$$(\alpha_1, \ldots, \alpha_K)^T = \boldsymbol{\alpha} = \mathbf{G}\left(\mathbf{c}_{[1,K]}, \mathbf{c}_{[1,K]}\right)^{-1} \mathbf{v} \tag{6.9}$$

Equivalence

Although discussed independently in the open literature, the results given by equations (6.6) and (6.8) may be considered equivalent. Consider the following minimization problem for any $\boldsymbol{\lambda} \in \mathbb{R}^N, \boldsymbol{\lambda} \neq \mathbf{0}$:

$$\min_{\mathbf{u}_{[1,N]}} \left\{ \left\| \sum_{z=1}^{Z} \lambda_n \mathbf{u}_n \right\|^2 \right\} \text{ under } \mathbf{G}\left(\mathbf{u}_{[1,N]}, \mathbf{c}_{[1,K]}\right) = \mathbf{V} \tag{6.10}$$

This is equivalent to:

$$\min \left\{ \left\| \sum_{z=1}^{Z} \lambda_n \mathbf{u}_n \right\|^2 \right\} \text{ under } \mathbf{G}\left(\mathbf{u}_{[1,N]}, \mathbf{c}_{[1,K]}\right) \boldsymbol{\lambda} = \mathbf{G}\left(\sum_{z=1}^{Z} \lambda_n \mathbf{u}_n, \mathbf{c}_{[1,K]}\right) = \mathbf{V}\boldsymbol{\lambda} \tag{6.11}$$

and thus satisfies according to (6.8):

$$\left\| \sum_{z=1}^{Z} \lambda_n \mathbf{u}_n \right\|^2 = \boldsymbol{\lambda}^T \mathbf{G}\left(\mathbf{u}_{[1,N]}, \mathbf{u}_{[1,N]}\right) \boldsymbol{\lambda} \geq (\mathbf{V}\boldsymbol{\lambda})^T \mathbf{G}\left(\mathbf{c}_{[1,K]}, \mathbf{c}_{[1,K]}\right)^{-1} (\mathbf{V}\boldsymbol{\lambda}) \tag{6.12}$$

or equivalently:

$$\boldsymbol{\lambda}^T \mathbf{G}\left(\mathbf{u}_{[1,N]}, \mathbf{u}_{[1,N]}\right) \boldsymbol{\lambda} \geq \boldsymbol{\lambda}^T \left(\mathbf{V}^{\mathbf{T}} \mathbf{G}\left(\mathbf{c}_{[1,K]}, \mathbf{c}_{[1,K]}\right)^{-1} \mathbf{V} \right) \boldsymbol{\lambda} \tag{6.13}$$

which is precisely the inequality (6.6).

In the following the form (6.8) is preferred, as its use provides a better understanding of the hypotheses associated with the different lower bounds of the MSE. Indeed, formulation of the minimal MSE performance bounds can be seen as a result of minimizing a norm (MSE) under the non-bias constraints associated with the given bound. The main advantage of this formulation is that it explicitly raises the problem of formulating the relevant non-bias constraints, which then determine the value of the lower bound of the norm.

6.3 Lower Bounds on the Mean Square Error (MSE) of an Estimator

As stated earlier, the most commonly used of the lower bounds on MSE is the CRB [8, 108]. The fact that the CRB is the lowest bound on the MSE of unbiased estimators is a result of the non-bias constraint that it imposes on the minimization of the MSE, which is the

weakest possible formulation of non-bias, i.e. that the estimator be non-biased at the true parameter value θ_0, and in its vicinity $\theta_0 + d\theta$ [122]:

$$\lim_{d\theta \to 0} \begin{cases} E_{\theta_0}\left[\widehat{g(\theta_0)}(\mathbf{x})\right] = g(\theta_0) \\ E_{\theta_0+d\theta}\left[\widehat{g(\theta_0)}(\mathbf{x})\right] = g(\theta_0 + d\theta) \end{cases} \tag{6.14}$$

which defines the constraint of "local non-bias":

$$E_{\theta_0+d\theta}\left[\widehat{g(\theta_0)}(\mathbf{x})\right] = g(\theta_0 + d\theta) + O(d\theta) = g(\theta_0) + \frac{\partial g(\theta_0)}{\partial \theta} d\theta + O(d\theta), \tag{6.15}$$

where $O(d\theta)$ is used to describe the error term in the approximation of the non-bias constraint.

Equivalently, by uniqueness of series expansion of $g(\theta_0 + d\theta)$ and $f_{\theta_0+d\theta}(\mathbf{x})$:

$$\begin{cases} E_{\theta_0}\left[\widehat{g(\theta_0)}(\mathbf{x})\right] = g(\theta_0) \\ E_{\theta_0}\left[\widehat{g(\theta_0)}(\mathbf{x}) \frac{\partial f_{\theta_0}(\mathbf{x})}{\partial \theta}\right] = \frac{\partial g(\theta_0)}{\partial \theta} \end{cases} \Leftrightarrow \begin{cases} E_{\theta_0}\left[\widehat{g(\theta_0)}(\mathbf{x}) - g(\theta_0)\right] = 0 \\ E_{\theta_0}\left[\left(\widehat{g(\theta_0)}(\mathbf{x}) - g(\theta_0)\right) \frac{\partial f_{\theta_0}(\mathbf{x})}{\partial \theta}\right] = \frac{\partial g(\theta_0)}{\partial \theta} \end{cases} \tag{6.16}$$

Then a direct application of (6.8) under (6.16) leads to the CRB.

Bhattacharyya extended the 1^{rst} order "locally non-bias" constraint of the CRB (6.15) to the L^{th} order in the vicinity of the true parameter value by looking for estimators satisfying [7, 111, 115, 117, 122]:

$$E_{\theta_0+d\theta}\left[\widehat{g(\theta_0)}(\mathbf{x})\right] = g(\theta_0 + d\theta) + O\left(d\theta^L\right) \tag{6.17}$$

The higher the order of differentiation, L, the closer the approximation of the non-bias constraint is to the BB non-bias constraint, and accordingly the smaller $O\left(d\theta^L\right)$. The resulting bound is therefore greater than the CRB, but practical results have shown that values of L for which there is a noticeable difference requires a huge increase in computational complexity.

Barankin then introduced arguably one of the most significant characterizations of estimator performance, the Barankin Bound (BB) [111], which is the greatest possible lower bound on any absolute moment of order k for an unbiased estimator. The lower bound on the MSE is then a particular case of the general Barankin Bound, where $k = 2$, and the resulting bound can be seen to include the initial characterizations of MSE estimator performance provided by Cramér-Rao and Bhattacharyya. The Barankin Bound is the

result of imposing the strongest possible formulation of non-bias, that is non-bias across the entire parameter interval, on the norm minimization (6.3):

$$E_{\theta_0+d\theta}\left[\widehat{g\left(\theta_0\right)}\left(\mathbf{x}\right)\right] = g\left(\theta_0 + d\theta\right), \quad \forall \theta_0 + d\theta \in \Theta. \tag{6.18}$$

Unfortunately the Barankin Bound is the solution of an integral equation [116, 123] , which generally does not have a computable analytic solution. Therefore, since its introduction, numerous works [112–114, 117, 118, 124] have been devoted to deriving computable approximations of the Barankin Bound and have shown that the CRB and Barankin bound (BB) can be regarded as key representatives of two general classes of bounds, respectively the Small-Error bounds and the Large-Error bounds.

6.3.1 Derivation of the Barankin Bound

In the previous section we saw that the Barankin Bound corresponds to the minimization problem:

$$\min\left\{MSE_{\theta_0}\left[\widehat{g\left(\theta_0\right)}\right]\right\} \text{ under } E_\theta\left[\widehat{g\left(\theta_0\right)}\left(\mathbf{x}\right)\right] = g\left(\theta\right), \quad \forall \theta \in \Theta \tag{6.19}$$

where the expectation is taken over the entire parameter range.

Therefore, any unbiased estimator satisfying this Barankin non-bias constraint will satisfy:

$$E_{\theta_z}\left[\widehat{g\left(\theta_0\right)}\left(\mathbf{x}\right)\right] = g\left(\theta_z\right) = \int_\Omega \widehat{g\left(\theta_0\right)}\left(\mathbf{x}\right) f_{\theta_z}\left(\mathbf{x}\right) d\mathbf{x}, \quad \forall \theta_z \in \Theta. \tag{6.20}$$

Consequently, for $1 \leq z \leq Z$ and $\forall \mathbf{w} = \left[w_1, \ldots, w_Z\right]^T \in \mathbb{R}^Z$:

$$\sum_{z=1}^{Z} w_z E_{\theta_z}\left[\widehat{g\left(\theta_0\right)}\left(\mathbf{x}\right) - g\left(\theta_0\right)\right] = \sum_{z=1}^{Z} w_z \left(g\left(\theta_z\right) - g\left(\theta_0\right)\right) \tag{6.21}$$

or alternatively:

$$E_{\theta_0}\left[\left(\widehat{g\left(\theta_0\right)}\left(\mathbf{x}\right) - g\left(\theta_0\right)\right)\left(\sum_{z=1}^{Z} w_z \frac{f_{\theta_z}\left(\mathbf{x}\right)}{f_{\theta_0}\left(\mathbf{x}\right)}\right)\right] = \sum_{z=1}^{Z} w_z \left(g\left(\theta_z\right) - g\left(\theta_0\right)\right) \tag{6.22}$$

Then, according to (6.8), the minimization of $MSE_{\theta_0}\left[\widehat{g\left(\theta_0\right)}\right]$ under the constraint (6.22) leads to [111]:

$$MSE_{\theta_0}\left[\widehat{g\left(\theta_0\right)}\right] \geq \frac{\left[\sum_{z=1}^{Z} w_z \left(g\left(\theta_z\right) - g\left(\theta_0\right)\right)\right]^2}{E_{\theta_0}\left[\left(\sum_{z=1}^{Z} w_z \frac{f_{\theta_z}(\mathbf{x})}{f_{\theta_0}(\mathbf{x})}\right)^2\right]} \tag{6.23}$$

The inequality (6.23) is valid for any subset of test points $\{\theta_z\}_{[1,Z]}$ of Θ considered and $\forall \mathbf{w} \in \mathbb{R}^Z$. From this, Barankin deduced that, if $\widehat{g(\theta_0)}(\mathbf{x})$ is an unbiased estimator of $g(\theta_0)$, then its MSE is bounded by:

$$MSE_{\theta_0}\left[\widehat{g(\theta_0)}\right] \geq \lim_{Z\longrightarrow\infty} \sup_{\mathbf{w},\{\theta_z\}_{[1,Z]}} \left\{ \frac{\left[\sum_{z=1}^{Z} w_z \left(g(\theta_z) - g(\theta_0)\right)\right]^2}{E_{\theta_0}\left[\left(\sum_{z=1}^{Z} w_z \frac{f_{\theta_z}(\mathbf{x})}{f_{\theta_0}(\mathbf{x})}\right)^2\right]} \right\} \tag{6.24}$$

A more concise form can nevertheless be found by noting that [113]:

$$\frac{\left[\sum_{z=1}^{Z} w_z \left(g(\theta_z) - g(\theta_0)\right)\right]^2}{E_{\theta_0}\left[\left(\sum_{z=1}^{Z} w_z \frac{f_{\theta_z}(\mathbf{x})}{f_{\theta_0}(\mathbf{x})}\right)^2\right]} = \frac{\left(\mathbf{w}^T\boldsymbol{\Delta}\mathbf{g}\right)^2}{\mathbf{w}^T\mathbf{R}\mathbf{w}}, \tag{6.25}$$

$$R_{z,y} = \int_{\Omega} \frac{f_{\theta_z}(\mathbf{x})\, f_{\theta_y}(\mathbf{x})}{f_{\theta_0}(\mathbf{x})} d\mathbf{x}, \quad \Delta g_z = g(\theta_z) - g(\theta_0) \tag{6.26}$$

In fact, $\frac{\left(\mathbf{w}^T\boldsymbol{\Delta}\mathbf{g}\right)^2}{\mathbf{w}^T\mathbf{R}\mathbf{w}}$ is maximal for $\mathbf{w} = \lambda\mathbf{R}^{-1}\boldsymbol{\Delta}\mathbf{g}$ and its value is then $\boldsymbol{\Delta}\mathbf{g}^T\mathbf{R}^{-1}\boldsymbol{\Delta}\mathbf{g}$. Consequently another expression of (6.23) is:

$$MSE_{\theta_0}\left[\widehat{g(\theta_0)}\right] \geq \boldsymbol{\Delta}\mathbf{g}^T\mathbf{R}^{-1}\boldsymbol{\Delta}\mathbf{g} \tag{6.27}$$

leading to the "reduced" form of the Barankin bound:

$$MSE_{\theta_0}\left[\widehat{g(\theta_0)}\right] \geq \lim_{Z\longrightarrow\infty} \sup_{\{\theta_z\}_{[1,Z]}} \left\{\boldsymbol{\Delta}\mathbf{g}^T\mathbf{R}^{-1}\boldsymbol{\Delta}\mathbf{g}\right\} \tag{6.28}$$

Note that (6.27) is also the solution of the minimization problem:

$$\min\left\{MSE_{\theta_0}\left[\widehat{g(\theta_0)}\right]\right\} \text{ under } E_{\theta_z}\left[\widehat{g(\theta_0)}(\mathbf{x})\right] = g(\theta_z), \ \{\theta_z\}_{[1,Z]} \in \Theta \tag{6.29}$$

which proves that the greatest lower bound of the MSE for a finite number of test points $\{\theta_z\}_{[1,Z]}$ is obtained by simply expressing the "unbiased" constraint at the test points [113].

Finally, the unbiased and locally optimal estimator in the MSE sense $\widehat{g(\theta_0)}_{opt}^{loc}(\mathbf{x})$, i.e. the estimator that has the lowest MSE for estimation of the given parameter θ_0, satisfies (6.27):

$$\lim_{Z\longrightarrow\infty}\begin{cases}\mathbf{R}\mathbf{w}\lambda^{-1}=\boldsymbol{\Delta}\mathbf{g}\\ \widehat{g\left(\theta_0\right)}_{opt}^{loc}\left(\mathbf{x}\right)-g\left(\theta_0\right)=\sum\limits_{z=1}^{Z}\frac{w_z}{\lambda}\frac{f_{\theta_z}\left(\mathbf{x}\right)}{f_{\theta_0}\left(\mathbf{x}\right)}\\ MSE_{\theta_0}\left[\widehat{g\left(\theta_0\right)}\right]\geq\boldsymbol{\Delta}\mathbf{g}^T\mathbf{R}^{-1}\boldsymbol{\Delta}\mathbf{g}=\boldsymbol{\lambda}^{-1}\boldsymbol{\Delta}\mathbf{g}^T\mathbf{w}\end{cases}\tag{6.30}$$

that leads to, defining $\frac{1}{\lambda}=d\theta=\theta_{z+1}-\theta_z$, [116, 123]:

$$\int_{\Theta}K\left(\theta,\theta'\right)w\left(\theta'\right)d\theta' = g\left(\theta\right)-g\left(\theta_0\right),\tag{6.31a}$$

$$K\left(\theta,\theta'\right) = \int_{\Omega}\frac{f_{\theta}\left(\mathbf{x}\right)f_{\theta'}\left(\mathbf{x}\right)}{f_{\theta_0}\left(\mathbf{x}\right)}d\mathbf{x}\tag{6.31b}$$

$$\widehat{g\left(\theta_0\right)}_{opt}^{loc}\left(\mathbf{x}\right)-g\left(\theta_0\right) = \int_{\Theta}\frac{f_{\theta}\left(\mathbf{x}\right)}{f_{\theta_0}\left(\mathbf{x}\right)}w\left(\theta\right)d\theta\tag{6.31c}$$

$$MSE_{\theta_0}\left[\widehat{g\left(\theta_0\right)}\right] \geq \int_{\Theta}\left(g\left(\theta\right)-g\left(\theta_0\right)\right)w\left(\theta\right)d\theta\tag{6.31d}$$

Unfortunately, as discussed previously in section 2.7.1, in most practical cases, it is impossible to find an analytical solution of (6.31a) and to obtain an explicit form of $\widehat{g\left(\theta_0\right)}_{opt}$ and of the lower bound of the MSE, which somewhat limits its interest. However, while this discretization of the continuous constraint does not provide an analytical solution of the Barankin bound, its interest lies elsewhere, in the fact that it provides a general approach for approximating the Barankin Bound.

6.4 Toward a Piecewise Approximation of the Barankin Bound

So far, all previous works dedicated to assessing the true behavior of the BB at low SNR (Large-Error bounds) [112, 113, 116–118] can be reduced to the exploitation of the norm minimization lemma (6.8) associated with a basic discretization (6.29) of Barankin unbiasedness definition (6.18) and possibly including the search for a supremum (6.28). Such a basic discretization is sub-optimal in the scope of BB approximation tightness. Indeed, there is a set $\mathcal{H}$ of functions $h\left(\theta\right)$ of various behaviours that take the same values for a given set of test points $\left(h\left(\theta_z\right)=g\left(\theta_z\right),\left\{\theta_z\right\}_{[1,z]}\right)$, but may differ greatly from $g\left(\theta\right)$ at other points on the parameter interval. Therefore, the lower bound provided by such a

discretization (6.29) may not be a tight BB approximation since it is not a lower bound for the whole set of functions $\mathcal{H}$, except when the number of test points θ_z tends to infinity as $\mathcal{H}$ tends to reduce to $g(\theta)$ only. As discussed previously, the rate at which $\mathcal{H}$ reduces to $g(\theta)$ is not known, and the computational complexity is greatly increased as the number of test points increases. Consequently, in order to reduce the set $\mathcal{H}$ and thereby increase the tightness, it seems intuitively more efficient to resort to more discriminating constraints, such as L^{th} order derivative constraints. This leads to a straightforward, but novel method of approximating the BB [125].

The development of this method requires that both $f_\theta(\mathbf{x})$ and $g(\theta)$ can be approximated by piecewise series expansions of order L_z, that is to say the parameter space Θ can be partitioned in Z real sub-intervals I_z over which, for $\forall \theta_z + d\theta \in I_z$:

$$g(\theta_z + d\theta) = g(\theta_z) + \sum_{l=1}^{L_z} \frac{\partial^l g(\theta_z)}{\partial^l \theta} \frac{d\theta^l}{l!} + O\left(d\theta^{L_z}\right) \tag{6.32}$$

$$f_{\theta_z + d\theta}(\mathbf{x}) = f_{\theta_z}(\mathbf{x}) + \sum_{l=1}^{L_z} \frac{\partial^l f_{\theta_z}(\mathbf{x})}{\partial^l \theta} \frac{d\theta^l}{l!} + o_{\mathbf{x}}\left(d\theta^{L_z}\right) \tag{6.33}$$

We implicitly assume that the integrals $\int \left(\frac{\partial^l f_\theta(\mathbf{x})}{\partial^l \theta}\right)^2 \frac{1}{f_\theta(\mathbf{x})} d\mathbf{x}$ converge and define piecewise continuous functions of θ on Θ, for all $\theta \in \Theta$, to allow order of integration and differentiation interchange [118]. The terms $o\left(d\theta^{L_z}\right)$ and $o_{\mathbf{x}}\left(d\theta^{L_z}\right)$ are the differences between the true function values on each sub-interval and the polynomial approximations. It is clear from this term that the higher the order of the expansion, the closer the approximation is to the true function.

Then, on every sub-interval I_z, a possible general discretization of Barankin unbiasedness definition (6.18) is:

$$E_{\theta_z + d\theta}\left[\widehat{g(\theta_0)}(\mathbf{x})\right] = g(\theta_z + d\theta) + O\left(d\theta^{L_z}\right) \tag{6.34}$$

which, by uniqueness of any series expansion, can be expressed in terms of the $L_z + 1$ following linear constraints:

$$\int \widehat{g(\theta_0)}(\mathbf{x}) \frac{\partial^l f_{\theta_z}(\mathbf{x})}{\partial^l \theta} d\mathbf{x} = \frac{\partial^l g(\theta_z)}{\partial^l \theta}, \quad l \in [0, L_z] \tag{6.35}$$

or equivalently:

$$E_{\theta_0}\left[\left(\widehat{g(\theta_0)}(\mathbf{x}) - g(\theta_0)\right) \frac{\frac{\partial^l f_{\theta_z}(\mathbf{x})}{\partial^l \theta}}{f_{\theta_0}(\mathbf{x})}\right] = \left[\frac{\partial^l \left(g(\theta) - g(\theta_0)\right)}{\partial^l \theta}\right]_{\theta_z} \tag{6.36}$$

Thus, the set of $\sum_{z=1}^{Z}(L_z+1)$ constraints (6.36) deriving from the Z piecewise discretization of (6.18) defines a given approximation of the BB denoted by $\widehat{\mathrm{BB}}_{L_1,...,L_Z}^{I_1,...,I_Z}$ (6.8):

$$\widehat{\mathrm{BB}}_{L_1,...,L_Z}^{I_1,...,I_Z} = \mathbf{v}^T\mathbf{G}^{-1}\mathbf{v}, \quad \mathbf{v}=\begin{bmatrix}\mathbf{v}_1\\ \vdots \\ \mathbf{v}_Z\end{bmatrix}, \quad \mathbf{G}=E_{\theta_0}\left[\mathbf{c}\mathbf{c}^T\right], \quad \mathbf{c}=\begin{bmatrix}\mathbf{c}_1\\ \vdots \\ \mathbf{c}_Z\end{bmatrix} \tag{6.37}$$

(6.38)

$$\mathbf{v}_z = \left[g(\theta_z)-g(\theta_0), \frac{\partial g(\theta_z)}{\partial\theta}, ..., \frac{\partial^{L_z} g(\theta_z)}{\partial^{L_z}\theta}\right]^T \tag{6.39}$$

$$\mathbf{c}_z = \left[f_{\theta_z}(\mathbf{x}), \frac{\partial f_{\theta_z}(\mathbf{x})}{\partial\theta}, ..., \frac{\partial^{L_z} f_{\theta_z}(\mathbf{x})}{\partial^{L_z}\theta}\right]^T \tag{6.40}$$

Moreover, if $\min\{L_1,...,L_Z\}$ tends to infinity, a straightforward exercise in mean square convergence establishes that $\widehat{\mathrm{BB}}_{L_1,...,L_Z}^{I_1,...,I_Z}$ converges in mean-square to the BB. An immediate generalization of expression (6.37), similar to that used in the HCRB [112] and the hybrid bound proposed by Abel [118], consists of taking its supremum over existing degrees of freedom (sub-interval definitions and series expansion orders).

6.4.1 An Alternative Look at Existing BB Approximations

The formalism proposed in the previous section allows exploration of the unbiasedness assumption from its *weakest* to its *strongest* formulation, and therefore encompasses all previously derived BB approximations, which we designate:

• Z-piecewise BB approximation of homogeneous order L, if on all sub-intervals I_z the series expansions are of the same order L,

• Z-piecewise BB approximation of heterogeneous orders $\{L_1,...,L_Z\}$, if otherwise.

The proposed formalism can now be used to provide a new look at the previously derived MSE lower bounds:

• the CRB [8] is a 1-piecewise BB approximation of homogeneous order 1, since the constraints are:

$$E_{\theta_0}\left[\widehat{g(\theta_0)}(\mathbf{x})\right] = g(\theta_0), \quad E_{\theta_0}\left[\widehat{g(\theta_0)}(\mathbf{x})\frac{\frac{\partial f_{\theta_0}(\mathbf{x})}{\partial\theta}}{f_{\theta_0}(\mathbf{x})}\right] = \frac{\partial g(\theta_0)}{\partial\theta} \tag{6.41}$$

• the Bhattacharyya bound [7] of order L is a 1-piecewise BB approximation of homoge-

neous order L, since the constraints, for $l = 1, \ldots, L$, are:

$$E_{\theta_0}\left[\widehat{g(\theta_0)}(\mathbf{x})\right] = g(\theta_0), \quad E_{\theta_0}\left[\widehat{g(\theta_0)}(\mathbf{x})\frac{\frac{\partial^l f_{\theta_0}(\mathbf{x})}{\partial^l \theta}}{f_{\theta_0}(\mathbf{x})}\right] = \frac{\partial^l g(\theta_0)}{\partial^l \theta}. \tag{6.42}$$

• the Hammersley-Chapman-Robbins bound (HCRB) [112] is the supremum of a 2-piecewise BB approximation of homogeneous order 0, over a set of constraints of type:

$$E_{\theta_0}\left[\widehat{g(\theta_0)}(\mathbf{x})\right] = g(\theta_0), \quad E_{\theta_1}\left[\widehat{g(\theta_0)}(\mathbf{x})\right] = g(\theta_1) \tag{6.43}$$

• the McAulay-Seidman bound (MSB_Z) [113] with Z test points is an $Z+1$-piecewise BB approximation of homogeneous order 0, since the constraints, for $z = 1, \ldots, Z$, are:

$$E_{\theta_0}\left[\widehat{g(\theta_0)}(\mathbf{x})\right] = g(\theta_0), \quad E_{\theta_z}\left[\widehat{g(\theta_0)}(\mathbf{x})\right] = g(\theta_z), \tag{6.44}$$

• the McAulay-Hofstetter bound (MHB_Z) [126] is an $Z+1$-piecewise BB approximation of heterogeneous order $\{1, 0, ..., 0\}$, since the constraints, for $z = 1, \ldots, Z$, are:

$$\begin{cases} E_{\theta_0}\left[\widehat{g(\theta_0)}(\mathbf{x})\right] = g(\theta_0), \quad E_{\theta_0}\left[\widehat{g(\theta_0)}(\mathbf{x})\frac{\frac{\partial f_{\theta_0}(\mathbf{x})}{\partial \theta}}{f_{\theta_0}(\mathbf{x})}\right] = \frac{\partial g(\theta_0)}{\partial \theta} \\ E_{\theta_z}\left[\widehat{g(\theta_0)}(\mathbf{x})\right] = g(\theta_z), \end{cases}. \tag{6.45}$$

• the Hybrid Barankin-Bhattacharyya bound ($\text{HBB}_{L,Z}$) [118] is an $Z+1$-piecewise BB approximation of heterogeneous order $\{L, 0, ..., 0\}$, since the constraints, for $z = 1, \ldots, Z$ and $l = 1, \ldots, L$, are:

$$\begin{cases} E_{\theta_0}\left[\widehat{g(\theta_0)}(\mathbf{x})\right] = g(\theta_0), \quad E_{\theta_0}\left[\widehat{g(\theta_0)}(\mathbf{x})\frac{\frac{\partial^l f_{\theta_0}(\mathbf{x})}{\partial^l \theta}}{f_{\theta_0}(\mathbf{x})}\right] = \frac{\partial^l g(\theta_0)}{\partial^l \theta} \\ E_{\theta_z}\left[\widehat{g(\theta_0)}(\mathbf{x})\right] = g(\theta_z) \end{cases} \tag{6.46}$$

Note that $\text{MHB}_Z = \text{HBB}_{1,Z}$, i.e. the Hybrid Barankin-Bhattacharyya bound is a generalization of the McAulay-Hofstetter bound.

6.4.2 A New Practical BB Approximation

The proposed formalism not only provides a new look at previous BB approximations, it also suggests a very straightforward and practical new BB approximation, which we

denote $\widehat{\mathrm{BB}}_1^Z$ in the following: the $Z+1$-piecewise BB approximation of homogeneous order 1 characterized by the set of constraints, for $z = 1, \ldots, Z$:

$$\left\{ \begin{array}{l} E_{\theta_0}\left[\widehat{g\left(\theta_0\right)}\left(\mathbf{x}\right)\right] = g\left(\theta_0\right), \; E_{\theta_0}\left[\widehat{g\left(\theta_0\right)}\left(\mathbf{x}\right)\frac{\frac{\partial f_{\theta_0}(\mathbf{x})}{\partial\theta}}{f_{\boldsymbol{\theta}_0}(\mathbf{x})}\right] = \frac{\partial g(\theta_0)}{\partial\theta} \\ E_{\theta_z}\left[\widehat{g\left(\theta_0\right)}\left(\mathbf{x}\right)\right] = g\left(\theta_z\right), \; E_{\theta_z}\left[\widehat{g\left(\theta_0\right)}\left(\mathbf{x}\right)\frac{\frac{\partial f_{\theta_z}(\mathbf{x})}{\partial\theta}}{f_{\boldsymbol{\theta}_z}(\mathbf{x})}\right] = \frac{\partial g(\theta_z)}{\partial\theta} \end{array} \right. \tag{6.47}$$

Indeed it appears to be the generalization of the CRB when the parameter space is partitioned in more than one sub-interval, as well as the generalization of the usual BB approximation used in the open literature, i.e. the McAulay-Seidman form of the BB. Therefore its computational complexity does not exceed that of these two bounds.

6.4.3 General lower bounds expressions

For any set of $Z+1$ test points $\{\theta_z\}_{[1,Z+1]}$, among the existing lower bounds, only the MSB_Z and the $\mathrm{HBB}_{1,Z}$ (MHB_Z) are of a complexity comparable with $\widehat{\mathrm{BB}}_1^Z$. Nevertheless, the CRB and the HCRB are included in the comparison as they are respectively the simplest representative of Small and Large Error bounds. All mentioned lower bounds can be computed from the components of $\widehat{\mathrm{BB}}_1^Z$ - (6.37) (6.47) with rearrangement -:

$$\widehat{\mathrm{BB}}_1^Z = \mathbf{v}^T\mathbf{G}^{-1}\mathbf{v}, \quad \mathbf{v} = \left[\boldsymbol{\Delta}\mathbf{g}^T, \left(\ldots, \frac{\partial g(\theta_z)}{\partial\theta}, \ldots\right)^T\right]^T, \quad \mathbf{G} = \left[\begin{array}{cc} \mathbf{MSM} & \mathbf{CM} \\ \mathbf{CM}^T & \mathbf{PFIM} \end{array}\right]. \tag{6.48}$$

where **MSM** stands for the McAulay-Seidman Matrix, **PFIM** stands for the Pseudo Fisher Information Matrix and **CM** stands for the Cross Matrix:

$$\mathbf{MSM}_{z,y} = E_{\theta_0}\left[\left(\frac{f_{\theta_z}(\mathbf{x})}{f_{\theta_0}(\mathbf{x})}\right)\left(\frac{f_{\theta_y}(\mathbf{x})}{f_{\theta_0}(\mathbf{x})}\right)\right] \tag{6.49}$$

$$\tag{6.50}$$

$$\mathbf{PFIM}_{z,y} = E_{\theta_0}\left[\left(\frac{\partial f_{\theta_z}(\mathbf{x})}{\partial\theta}\frac{1}{f_{\theta_0}(\mathbf{x})}\right)\left(\frac{\partial f_{\theta_y}(\mathbf{x})}{\partial\theta}\frac{1}{f_{\theta_0}(\mathbf{x})}\right)\right] \tag{6.51}$$

$$\tag{6.52}$$

$$\mathbf{CM}_{z,y} = E_{\theta_0}\left[\left(\frac{f_{\theta_z}(\mathbf{x})}{f_{\theta_0}(\mathbf{x})}\right)\left(\frac{\partial f_{\theta_y}(\mathbf{x})}{\partial\theta}\frac{1}{f_{\theta_0}(\mathbf{x})}\right)\right] \tag{6.53}$$

Noting that:

$$\frac{f_{\theta_z}(\mathbf{x})f_{\theta_y}(\mathbf{x})}{f_{\theta_0}(\mathbf{x})} = e^{2s^2\,\mathrm{Re}\{[\mathbf{a}(\theta_z)-\mathbf{a}(\theta_0)]^H[\mathbf{a}(\theta_y)-\mathbf{a}(\theta_0)]\}}f(\mathbf{x}), \quad f(\mathbf{x}) = \frac{e^{-\|\mathbf{x}-s[\mathbf{a}(\theta_z)+\mathbf{a}(\boldsymbol{\theta}_\mathbf{y})-\mathbf{a}(\theta_0)]\|^2}}{\pi^M} \tag{6.54}$$

then **MSM**, **PFIM**, **CM** matrices are given by:

$$\mathbf{MSM}_{z,y} = e^{2s^2\,\mathrm{Re}\{[\mathbf{a}(\theta_z)-\mathbf{a}(\theta_0)]^H[\mathbf{a}(\theta_y)-\mathbf{a}(\theta_0)]\}}$$

$$\mathbf{CM}_{z,y} = 2s^2\left(\mathbf{MSM}_{z,y}\right)\mathrm{Re}\left\{\left[\mathbf{a}\left(\theta_z\right)-\mathbf{a}\left(\theta_0\right)\right]^H\frac{\partial\mathbf{a}\left(\theta_y\right)}{\partial\theta}\right\}$$

$$\mathbf{PFIM}_{z,y} = 2s^2\left(\mathbf{MSM}_{z,y}\right)\mathrm{Re}\left[\begin{array}{l}\left\{\frac{\partial\mathbf{a}(\theta_y)^H}{\partial\theta}E\left[\left(\mathbf{x}-s\mathbf{a}\left(\theta_y\right)\right)\left(\mathbf{x}-s\mathbf{a}\left(\theta_z\right)\right)^T\right]\frac{\partial\mathbf{a}(\theta_z)^*}{\partial\theta}\right\}+\\ \left\{\frac{\partial\mathbf{a}(\theta_y)^H}{\partial\theta}E\left[\left(\mathbf{x}-s\mathbf{a}\left(\theta_y\right)\right)\left(\mathbf{x}-s\mathbf{a}\left(\theta_z\right)\right)^H\right]\frac{\partial\mathbf{a}(\theta_z)}{\partial\theta}\right\}\end{array}\right]$$

where, (see Appendix A):

$$E\left[\mathbf{x}\right] = s\left[\mathbf{a}\left(\theta_z\right)+\mathbf{a}\left(\theta_y\right)-\mathbf{a}\left(\theta_0\right)\right] \quad (6.55)$$

$$E\left[\mathbf{x}\mathbf{x}^T\right] = E\left[\mathbf{x}\right]E\left[\mathbf{x}\right]^T \quad (6.56)$$

$$E\left[\mathbf{x}\mathbf{x}^H\right] = \mathbf{I}+E\left[\mathbf{x}\right]E\left[\mathbf{x}\right]^H \quad (6.57)$$

$$E\left[\left(\mathbf{x}-s\mathbf{a}\left(\theta_y\right)\right)\left(\mathbf{x}-s\mathbf{a}\left(\theta_z\right)\right)^T\right] = E\left[\mathbf{x}\mathbf{x}^T\right]-sE\left[\mathbf{x}\right]\mathbf{a}\left(\theta_z\right)^T \quad (6.58)$$

$$-s\mathbf{a}\left(\theta_y\right)E\left[\mathbf{x}\right]^T+s^2\mathbf{a}\left(\theta_y\right)\mathbf{a}\left(\theta_z\right)^T \quad (6.59)$$

$$E\left[\left(\mathbf{x}-s\mathbf{a}\left(\theta_y\right)\right)\left(\mathbf{x}-s\mathbf{a}\left(\theta_z\right)\right)^H\right] = E\left[\mathbf{x}\mathbf{x}^H\right]-sE\left[\mathbf{x}\right]\mathbf{a}\left(\theta_z\right)^H \quad (6.60)$$

$$-s\mathbf{a}\left(\theta_y\right)E\left[\mathbf{x}\right]^H+s^2\mathbf{a}\left(\theta_y\right)\mathbf{a}\left(\theta_z\right)^H \quad (6.61)$$

6.5 DOA Estimation Analysis

6.5.1 General observation model

The work presented here is primarily concerned with estimation of the DOA of signals emitted by an acoustic source and received by an array of microphones. In this framework two different signal models can be considered: the deterministic and the stochastic signal models. The discussed signal models are Gaussian and the unknown parameters are connected with the expectation value in the deterministic case and with the covariance matrix in the stochastic one [95]. In the following we focus on the deterministic signal model, although the proposed bound could also be applied to the stochastic signal model.

For the sake of simplicity the case of estimating the DOA of a single source using a single snapshot is considered. This situation corresponds to the signal model given in equation (2.22) for the case where $P = 1$, and $N = 1$. It also corresponds to the signal model in equation (3.11), for the case where estimation is performed using a single frequency bin only. In most situations it is necessary to use a larger number of snapshots or frequency bins in order to achieve reliable estimates, however some algorithms estimate

the DOA using a single snapshot only [127, 128]. The complex observation vector model $\mathbf{x}$ is modelled by:

$$\mathbf{x} = s\mathbf{a}(\theta_0) + \mathbf{n} \tag{6.62}$$

where θ_0 is the unknown DOA to be estimated, s^2 is the SNR ($s > 0$) and $\mathbf{n}$ is complex circular Gaussian noise, with zero mean and a known covariance matrix $\mathbf{C_n}$.

Without any loss of generality, and in order to simplify the resulting expressions it is assumed that the observation vector has been whitened. This is possible as the noise covariance matrix, $\mathbf{C_n}$, is known. Then $\mathbf{C_n} = \mathbf{L_n}\mathbf{L_n}^H$ and $\mathbf{x} \rightarrow \mathbf{L_n}^{-1}\mathbf{x}$ is the whitened observation model**,** which is equivalent to the original observation model but now $\mathbf{C_n} = \mathbf{I}$. As $\sigma^2 = 1$, the SNR is now equal to s^2 ($s > 0$), and the expression for the pdf given in equation (2.29) can be re-written as:

$$f_\theta(\mathbf{x}) = \frac{e^{-\|\mathbf{x} - s\mathbf{a}(\boldsymbol{\theta})\|^2}}{\pi^M} \tag{6.63}$$

From equation (2.26), the response vector is expressed as:

$$\mathbf{a}(\theta) = \left[\ 1, e^{-j(2)\frac{2\pi\Delta}{\lambda}\cos(\theta)}, \ldots, e^{-j(M)\frac{2\pi\Delta}{\lambda}\cos(\theta)}\ \right]^T. \tag{6.64}$$

and

$$\frac{\partial \mathbf{a}(\theta)}{\partial \theta} = \left(j(M)\frac{2\pi\Delta}{\lambda}\sin(\theta)\right)\mathbf{a}(\theta). \tag{6.65}$$

Then, setting $\frac{\Delta}{\lambda} = 1$, the possible values of the possible values of θ are limited to $]60^0, 120^0[$, in order to avoid the problem of ambiguous steering vectors as discussed in chapter (2).
For the case under consideration, the vector $\mathbf{v}$ is given by:

$$\mathbf{v} = [\theta_0 - \theta_0, \ldots, \theta_{Z-1} - \theta_0, 1, \ldots, 1]^T \tag{6.66}$$

In order to provide a fair comparison with the HCRB, the MSB_Z, $\mathrm{HBB}_{1,Z}$, $\widehat{\mathrm{BB}}_1^Z$ are computed as suprema over the possible values of $\{\theta_z\}_{[1,Z+1]}$. For the sake of simplicity $\{\theta_z\}_{[1,Z+1]} = \{0, d\theta, -d\theta\}$. In figure 6.2, the evolution of the various bounds as a function of SNR in the case of $M = 8$, $P = 1$ and $N = 1$.

The variance of the ML estimator is also shown in order to compare the threshold behaviour of the bounds, and it can be seen that the proposed approximation bound allows for a closer prediction of the threshold SNR.

For practical localisation schemes it is desirable to have the estimator operate at all times in the asymptotic region. This makes knowledge of the threshold SNR very

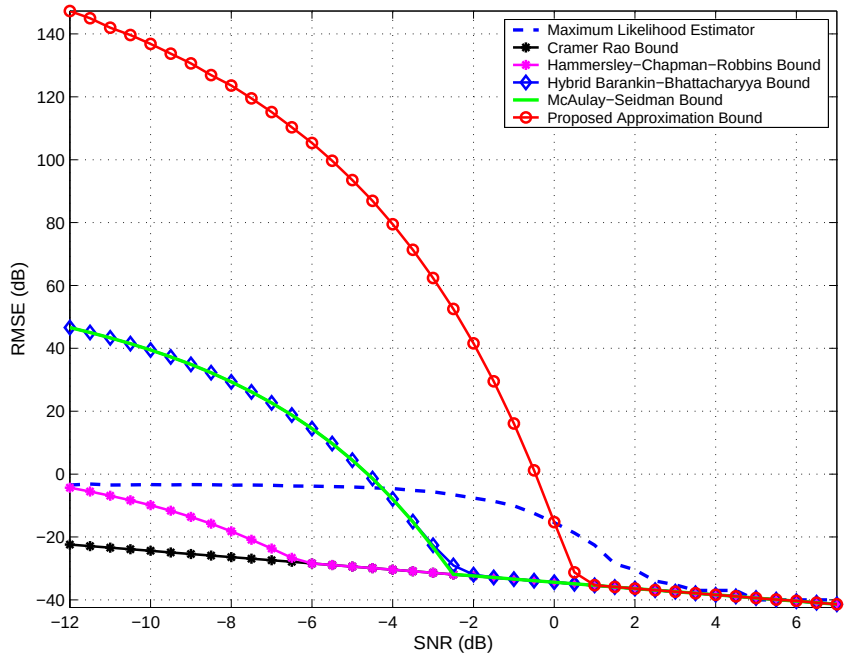


Figure 6.2: *Comparison of MSE lower bounds versus SNR when estimating the DOA of a known signal. In this case a single snapshot is available, $N = 1$, and the number of sensors $M = 8$.*

important as it allows the operating range of the estimator to be determined for a given situation. The proposed approximation bound is therefore suitable for source localisation feasibility studies as it can be used to predict the possible operating range that can be achieved. Knowledge of the possible operating range can also be used for comparison purposes with any proposed localisation algorithms, and in this sense the proposed bound can be used as a tool for evaluation of proposed estimators.

A large difference can be seen between the proposed approximation and the performance of the ML estimator below the threshold SNR, and it is clear that the Large-Error deterministic bounds are not suitable for evaluation of the estimator performance below the threshold SNR. This may be attributed to the fact that as the SNR decreases, the ML estimator may become heavily biased [117, 126], and in fact, under non-asymptotic conditions the ML estimator is not the most accurate estimator [23].

While it is common to evaluate estimator performance with respect to decreasing SNR [114, 129], in many situations of localising an acoustic source, the SNR will not be known. Therefore, future work will focus on evaluation of the proposed bound as a function of varying M, N and θ.

6.6 Conclusions

For design and evaluation of localisation schemes, knowledge of the best possible performance that can be achieved is very useful. This information can be found using lower bounds on the Mean Square Error (MSE). The greatest of these lower bounds is the Barankin Bound (BB), however it is generally not computable and must instead be approximated.

The difference seen between the ML estimator variance and the proposed approximation in the large error region shows that while the ML estimator is optimal under asymptotic conditions this is not the case as the SNR and/or number of samples decreases below the threshold values. Additionally, the present results suggest that the true value of the BB may be significantly underestimated by existing approximations, questioning previously drawn conclusions on prediction of the ML estimator variance by Deterministic Large Error Bounds.

The novel contributions made to this area of research include a new formalism for classification of Lower MSE bounds, which provides a clear explanation of the constraints imposed by these bounds, and their relationship to the BB. The proposed formalism leads to the derivation of a new approximation bound, $\widehat{\mathrm{BB}}_1^Z$, which provides a closer approximation of the BB than existing bounds, allowing for a better prediction of the SNR threshold.

7 Conclusions

7.1 Summary and Conclusions

The work presented here concerns model order determination, and characterisation of DOA estimators in the acoustic context. Firstly the problem of DOA estimation using array processing techniques is introduced and the possible applications of DOA estimation are discussed. Then, chapter 2 provides the mathematical basis for the rest of the thesis. The model of the wavefront propagating from the source to the array is recalled and the mathematical estimation problem is introduced. The possible approaches to finding an optimal estimator and the difficulties in finding such an estimator for DOA estimation are then discussed.

Chapter 3 provides a unified explanation and discussion of the classical array processing DOA estimation techniques. Most of these techniques are based on an initial assumption that the signal of interest is narrowband. This assumption is not valid for acoustic source localisation and the application of these classical techniques to the localisation of wideband signals is discussed, and recent developments in this area are reviewed.

Based on the profile of the noise eigenvalues of the correlation matrix as introduced by Grouffaud et. al [1], a novel method for determining the number of acoustical sources present is presented in chapter 4. This approach is based on the profile of the correlation

matrix eigenvalues in the absence of a source, which is seen to be exponential. The presence of a source is then detected by looking for a break in this profile. This method does not make the assumption that all noise eigenvalues are equal making it more robust to the effects of small data size than classical methods.

The novel work done in this area includes a discussion on the selection of suitable threshold values, and also the application of the method, called the Exponential Fitting Test (EFT), to the area of determining the number of wideband acoustic sources present. The performance of the test was then analysed for this situation using both computer simulations and experimental recordings. In particular the effects of reverberation were considered, and it was shown that the EFT is far more robust to the effects of reverberation. This is due to the fact that the threshold is selected using recordings taken in the room when there is no source present. The selection of the threshold then acts as a training step allowing the room impulse response to be taken into consideration. In the presence of reverberation the classical tests mistakenly classify reflections as sources, however the initial training step for the EFT makes it more robust to the presence of such phantom sources. A further advantage of the EFT is that this increase in accuracy does not come at the price of increased computational complexity.

In chapter 5 the performance of three of the most well known localisation approaches are compared: Multiple Signal Classification (MUSIC), Maximum Likelihood (ML) estimation and Time Delay Estimation (TDE). Using the criteria established in chapter 2 the performance of these three estimators was then analysed and the results for the cases of single and multiple sources were compared for both computer simulations and experimental recordings. The novel work done in this area includes the investigation of the effects of reverberation on the tests using the Useful-to-Detrimental ratio, and also the experimental results which confirmed these simulation results.

The results found show that for SNR values above a certain "threshold" level the results found by the estimators are close to the true DOA and the MSE is low, with the MSE of the MUSIC and ML methods being lower than that of the TDE method, due to the severely limited resolution capabilities of the TDE method. The resolution achievable when using the MUSIC and ML were also shown to be dependent on both the SNR and number of samples available as well as the true DOA value.

However, once the SNR decreases below this threshold value, all the estimators experienced a sharp increase in the MSE. Each of these estimators involves a maximisation of the estimator ambiguity function, or the"spatial spectra" of the estimators, although it should be noted that they are not spectra in any real sense. Analysis of these spectra was then used to explain the performance of the estimators. It was seen that at high SNR the

spectra have sharp mainlobes centred on the true DOA value, resulting in all estimates being very close or equal to the true value. As the SNR decreased to the threshold level it was shown that the sidelobes begin to increase and the mainlobe becomes broader. For SNR levels below the threshold value both the sidelobes and mainlobe are now very broad and the sidelobes are of equal or greater magnitude than the mainlobe, causing the sharp increase in MSE experienced in the results.

The effects of reverberation on the estimator performance were also considered in chapter 5, using both room response simulations and experimental recordings. From the results it was seen that the TDE approach is more robust to the effects of reverberation than the MUSIC and ML methods. For larger sample sizes, the accuracy of both the ML and MUSIC methods increased, with MUSIC being the most sensitive to sample size. However, the ML approach is computationally expensive. This is especially true for the case of $P > 1$, where a P-dimensional search must be implemented. In this work, the need for a multi-dimensional search was avoided by implementing an Alternating Projection (AP) algorithm which instead de-couples the estimates into P 1-dimensional searches. However, the method remains computationally burdensome, particularly when implemented for broadband sources, as the calculations must be repeated across each of the frequency bands of interest.

The novel contributions made in chapter 5 include the evaluation of the estimators using both computer simulations and experimental recordings, for a range of estimation situations. The effect of reverberation on the estimators' performances was also considered by evaluation of the results for increasing values of the Useful-to-Detrimental ratio.

In chapter 6 lower bounds on the MSE are used to predict the best possible performance of an estimator for a given estimation problem. The most commonly used minimal performance bound is the Cramer-Rao Bound (CRB). However, for non-linear estimators the possible performance indicated by the CRB is far too optimistic, and it gives no indication of the threshold value witnessed in chapter 5. Knowledge of this threshold value is of great importance as operation below this point is highly undesirable, as operation below this threshold point leads to very high MSE. Lower bounds on the MSE can therefore be used to predict the SNR threshold for a given estimation problem and thereby define the estimator operating range.

We therefore consider the Barankin Bound (BB), which is the greatest lower bound on the MSE. However, it is generally not computable and therefore must instead be approximated. In chapter 6 we propose a new formalism that allows for classification of BB approximations which also gives a clear interpretation of these approximations. It is shown that existing minimal bounds can be classified using this formalism. This

analysis then leads to the introduction of a new practical method of approximating the BB, which while appearing closer to the BB than existing approximations, does not exceed the computational complexity of the CRB.

The proposed bound is then applied to the DOA estimation problem, for the case of $P = 1$, $N = 1$ and $M = 8$. Comparison of the performance characterisation given by the proposed bound and previous approximation bounds show that the proposed bound allows for closer approximation of the SNR threshold for the ML estimator. This makes it a useful tool for prediction of the SNR threshold for DOA estimation problems.

7.2 Future Work

The analysis of DOA estimator performance can be extended to include the situation where the sources lie in the near-field and the assumption of planar wavefronts no longer holds. Prior knowledge of the signal characteristics can also be applied to improve the estimator performance in situations of low SNR and/or high reverberation.

The work in chapter 4 can be extended to include further analysis of the effects of reverberation on the proposed model order determination method. The presence of reverberation means that the assumption of white noise can no longer be applied. For moderate levels of reverberation, it was shown that the approximation of the noise as white is possible and the proposed method continues to work under such circumstances. However, for higher levels of reverberation this approximation is no longer possible and future work will focus on revising the model for prediction of the noise eigenvalue profile, in order to take this into account.

An extension of the work in chapter (6) is the generalisation of the proposed bound to simultaneous estimation of several functions of multiple parameters. Also the case for $N > 1$ will also be considered.

Another extension is the selection of suitable test points. By removing the supremum from the expression for the proposed bound, the calculational complexity is reduced and this is particularly important for the multiple parameter case in order to avoid the need for a multi-dimensional optimization. This approach can then be used to evaluate the effects of experimental modelling errors. Similarly the proposed practical bound can then be used as a design tool for choosing the optimal array geometry for a given situation.

A

Appendix

Derivation of equation (6.55:6.61)

As we assume a Gaussian distribution the pdf can be expressed as follows:

$$f_\theta(\mathbf{x}) = \frac{e^{-\|\mathbf{x}-\mathbf{m_x}\|^2}}{\pi^M}. \tag{A.1}$$

The expected value of $\mathbf{x}$ is then given by:

$$E[\mathbf{x}] = \mathbf{m_x}. \tag{A.2}$$

Then:

$$E\left[(\mathbf{x}-\mathbf{m_x})(\mathbf{x}-\mathbf{m_x})^T\right] = 0 = E\left[\mathbf{x}\mathbf{x}^T\right] - E[\mathbf{x}]E[\mathbf{x}]^T, \tag{A.3}$$

$$\Rightarrow E\left[\mathbf{x}\mathbf{x}^T\right] = E[\mathbf{x}]E[\mathbf{x}]^T, \tag{A.4}$$

and:

$$E\left[(\mathbf{x}-\mathbf{m_x})(\mathbf{x}-\mathbf{m_x})^H\right] = C_\mathbf{x} = \mathbf{I}_d, \tag{A.5}$$

$$E\left[(\mathbf{x}-\mathbf{m_x})(\mathbf{x}-\mathbf{m_x})^H\right] = E\left[\mathbf{x}\mathbf{x}^H\right] - E[\mathbf{x}]E[\mathbf{x}]^H. \tag{A.6}$$

We can therefore say that:

$$\mathbf{I}_d = E\left[\mathbf{x}\mathbf{x}^H\right] - E[\mathbf{x}]E[\mathbf{x}]^H, \tag{A.7}$$

resulting in:

$$E\left[\mathbf{x}\mathbf{x}^H\right] = \mathbf{I}_d + E[\mathbf{x}]E[\mathbf{x}]^H. \tag{A.8}$$

Bibliography

[1] J. Grouffaud, P. Larzabal, and H. Clergeot, "Some Properties of Ordered Eigenvalues of a Wishart Matrix: Application in Detection Test and Model Order Selection," in *Proc. ICASSP*, 1996, pp. 2463–2466.

[2] H. Krim and M. Viberg, "Two Decades of Array Signal Processing - The Parametric Approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, 1996.

[3] Y. Rockah and P. Schultheiss, "Array Shape Calibration Using Sources in Unknown Locations–Part I: Far-Field Sources," *IEEE Trans. Signal Processing*, vol. 35, no. 3, pp. 286–299, 1987.

[4] S. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory.* London: Prentice Hall, 1993.

[5] W. Feller, *An Introduction to Probability Theory and Its Applications, Volume 1,.* New York: Wiley, 1968.

[6] E. Chaumette, "Contribution a la Caracterisation des Performances des Problémes Conjoints de Détection et D'Estimation," Ph.D. dissertation, Ecole Normale Superieure de Cachan, Paris, France, 2004.

[7] H. L. V. Trees, *Detection, Estimation and Modulation Theory.* New York: Wiley, 1968, vol. 1.

[8] H.Cramér, "Mathematical Methods of Statistics," *Princeton University Press*, 1946.

[9] N. Bertaux, P. Larzabal, C. Adnet, and E. Chaumette, "A Parameterized Maximum Likelihood Method for Multipaths Channels Estimation," in *Proc. Signal Processing Advances in Wireless Communications (SPAWC)*, 1999, pp. 391–394.

[10] B. V. Veen and K. Buckley, "Beamforming: A Versatile Approach to Spatial Filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, 1988.

[11] J. Chen, R. Hudson, and K. Yao, "Maximum likelihood source localization and unknown sensor location estimation for wideband signals in the near-field," *IEEE Trans. Signal Processing*, vol. 50, no. 8, pp. 1843–1854, 2002.

[12] M. Brandstein, "Time-delay Estimation of Reverberated Speech Exploiting Harmonic Structure," *J. Acoust. Soc. Am.*, vol. 105, no. 5, pp. 2914–2919, 1999.

[13] ——, "On the Use of Explicit Speech Modeling in Microphone Array Applications," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Seattle, WA, 1998.

[14] M. Brandstein and D. Ward, "Cell-Based Beamforming (CE-BABE)for Speech Acquisition with Microphone Arrays," *IEEE Trans. Speech Audio Processing*, vol. 8, no. 6, pp. 738–743, 2000.

[15] R. Duraiswami, D. Zotkin, and L. Davis, "Active Speech Source Localization by a Dual Coarse-to-Fine Search," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Salt Lake city, Ut, 2001.

[16] D. Ward and R. Williamson, "Particle Filter Beamforming for Acoustic Source Localization in a Reverberant Environment," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Orlando, Fl, 2002.

[17] S. Birchfield, "A Unifying Framework for Acoustic Localization," in *Proc. 12th European Signal Processing Conference (EUSIPCO)*, Vienna, Austria, 2004.

[18] H. Wang and M. Kaveh, "Estimation of Angles-of-Arrival for Wideband Sources," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, San Diego, CA, 1984.

[19] ——, "Coherent Signal-Subspace Processing for the Detection and Estimation of Angles of Arrival of Multiple Wideband Sources," *IEEE Trans. Signal Processing*, vol. 33, no. 4, pp. 823–831, 1985.

[20] R. O. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," in *Proc RADC Spectrum, Estimation Workshops*, 1979, pp. 243–258.

[21] ——, "A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation," Ph.D. dissertation, Stanford University, Stanford, CA, nov 1981.

[22] P. Stoica and A. Nehorai, "MUSIC, Maximum Likelihood and Cramer-Rao Bound," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, New York, NY, 1988.

[23] ——, "MUSIC, Maximum Likelihood and Cramer-Rao Bound," *IEEE Trans. Signal Processing*, vol. 37, no. 5, pp. 720–741, 1989.

[24] "Exact and Large Sample Maximum Likelihood Techniques for Parameter Estimation and Detection in Array Processing," in *Radar Array Processing*, S. Haykin, J. Litva, and T. J. Shepherd, Eds. Berlin: Springer-Verlag, 1993, ch. 3, pp. 47–98.

[25] S. Valaee and P. Kabal, "An Information Theoretic Approach to Source Enumeration in Array Signal Processing," *IEEE Trans. Signal Processing*, vol. 52, no. 5, pp. 1171–1178, 2004.

[26] N. Wiener, "Extrapolation, Interpolation and Smoothing of Stationary Time Series," *MIT Press, Cambridge MA*, 1949.

[27] J. Evans, J. Johnson, and D. Sun, "Application of Advanced Signal Processing Techniques to Angle of Arrival Estimation in ATC Navigation and Surveillance Systems," *Technical report, MIT Lincoln Laboratory*, June 1982.

[28] B. Friedlander and A. Weiss, "Direction Finding Using Spatial Smoothing with Interpolated Arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 2, pp. 574–587, 1992.

[29] T. Shan, M. Wax, and T. Kailath, "On Spatial Smoothing for Directions of Arrival Estimation of Coherent Signals," *IEEE Trans. Signal Processing*, vol. 33, no. 4, pp. 806–811, 1985.

[30] T. Pham and B. Sadler, "Incoherent and Coherent Aeroacoustic Wideband Direction Finding Algorithms for Ground Vehicles," in *Proc. Acoustical Society America (ASA)*, Honolulu, HI, 1996.

[31] H. Hung and M. Kaveh, "Focussing Matrices for Coherent Signal-Subspace Processing," *IEEE Trans. Signal Processing*, vol. 36, no. 8, pp. 1272–1281, 1988.

[32] T. Pham and B. Sadler, "Aeroacoustic Wideband Array Processing for Detection and Tracking of Ground Vehicles," *J. Acoust. Soc.Am.*, vol. 98, no. 5, p. 2969, 1995.

[33] F. Asano, Y. Motomura, H. Asoh, T. Yoshimura, N. Ichimura, and S. Nakamura, "Detection and Separaton of Speech Segment Using Audio and Video Information Fusion," in *Proc. International Conference on Information and Fusion (IF)*, 2003.

[34] F. Asano, K. Yamamoto, I. Hara, J. Ogata, T. Yoshiumura, Y. Motomura, N. Ichimura, and H. Asoh, "Detection and Separation of Speech Event Using Audio and Video Information Fusion and its Application to Robust Speech Interface," *EURASIP Journal on Applied Signal Processing*, vol. 11, pp. 1727–1738, 2004.

[35] S. Valaee and P. Kabal, "Wideband Array Processing Using a Two-Sided Correlation Transformation," *IEEE Trans. Signal Processing*, vol. 43, no. 1, pp. 160–172, 1995.

[36] A. Barabell, "Improving the Resolution Performance of Eigenstructure Based Direction-Finding Algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Boston, MA, 1983.

[37] B. Rao and K. Hari, "Performance Analysis of ESPRIT and TAM in Determining the Direction of Arrival of Plane Waves in Noise," *IEEE Trans. Signal Processing*, vol. 37, no. 12, pp. 1990–1995, 1989.

[38] R. Roy and T. Kailath, "ESPRIT-Estimation of Signal Parameters via Rotational Invariant Techniques,," *IEEE Trans. Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.

[39] M. Fink, "Time Reversal of Ultrasonic Fields - Part I: Basic Theory," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 39, no. 5, pp. 555–566, 1992.

[40] C. Dorme and M. Fink, "Ultrasonic beam steering through inhomogeneous layers with a time reversal mirror," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 43, no. 1, pp. 167–175, 1996.

[41] G. F. Edelmann, T. Akal, W. S. Hodgkiss, S. Kim, W. A. Kupermann, and H. C. Song, "An initial demonstration of underwater acoustic communication using time reversal," *IEEE J. Oceanic Eng.*, vol. 27, no. 3, pp. 602–609, 2002.

[42] S. Yon, M. Tanter, and M. Fink, "Sound Focusing in Rooms: The Time-Reversal Approach," *J. Acoust. Soc. Am.*, vol. 113, no. 3, pp. 1533–1543, 2003.

[43] F. Wu, J. L. Thomas, and M. Fink, "Time Reversal of Ultrasonic Fields - Part II: Experimental Results," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 39, no. 5, pp. 567–578, 1992.

[44] N. Chakroun, M. Fink, and F. Wu, "Time Reversal Processing in Ultrasonic Nondestructive Testing," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 42, no. 6, pp. 1087–1098, 1995.

[45] N. Mordant, C. Prada, and M. Fink, "Highly Resolved Detection and Selective Focusing in a Waveguide using the D.O.R.T. Method," *J. Acoust. Soc. Am.*, vol. 105, no. 5, pp. 2634–2642, 1999.

[46] E. Kerbrat, C. Prada, D. Cassereau, and M. Fink, "Ultrasonic nondestructive testing of scattering media using the decomposition of the time reversal operator," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 49, no. 8, pp. 1103–1113, 2002.

[47] ——, "Imaging in the presence of grain noise using the decomposition of the time reversal operator," *J. Acoust. Soc. Am.*, vol. 113, no. 3, pp. 1230–1240, 2003.

[48] C. Prada and J. L. Thomas, "Experimental Sub-Wavelength Localization of Scatterers by Decomposition of the Time Reversal Operator Interpreted as a Covariance Matrix." *J. Acoust. Soc. Am.*, vol. 114, no. 1, pp. 235–243, 2003.

[49] T. Folegot, C. Prada, and M. Fink, "Resolution enhancement and separation of reverberation from target echo with the time reversal operator decomposition," *J. Acoust. Soc. Am.*, vol. 113, no. 6, pp. 3155–3160, 2003.

[50] C. Prada and M. Fink, "Separation of interfering acoustic scattered signals using the invariants of the tim-reversal operator. application to lamb waves characterization," *J. Acoust. Soc. Am.*, vol. 104, no. 2, pp. 801–807, 1998.

[51] E. Naftali and N. Makris, "Necessary Conditions for a Maximum Likelihood Estimate to Become Asymptotically Unbiased and Attain the Cramer-Rao Lower Bound. Part I. General Approach with an Application to Time-Delay and Doppler Shift Estimation," *Journal Acoustical Society of America (JASA)*, vol. 110, no. 4, pp. 1917–1930, 2001.

[52] A. Thode, M. Zanolin, E. Naftali, I. Ingram, P. Ratilal, and N. Makris, "Necessary Conditions for a Maximum Likelihood Estimate to Become Asymptotically Unbiased and Attain the Cramer-Rao Lower Bound. Part II. Range and Depth Localization of a Sound Source in an Ocean Waveguide," *Journal Acoustical Society of America (JASA)*, vol. 112, no. 5, pp. 1890–1910, 2002.

[53] M. Viberg and A. Swindlehurst, "A Bayesian Approach to Auto-Calibration for Parametric Array Signal Processing," *IEEE Trans. Signal Processing*, vol. 42, no. 12, pp. 3495–3507, 1994.

[54] J. Chen, R. Hudson, and K. Yao, "A maximum-likelihood parametric approach to source localizations," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Salt Lake City, UT, 2001.

[55] K. Yao, J. Chen, and R. Hudson, "Maximum-Likelihood Acoustic Source Localization: Experimental Results," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Orlando, FL, 2002.

[56] J. Chen, R. Hudson, and K. Yao, "Joint maximum-likelihood source localization and unknown sensor location estimation for near-field wideband signals," in *Proc. SPIE*, 2001.

[57] M. Feder and E. Weinstein, "Parameter Estimation of Superimposed Signals Using the EM Algorithm," *IEEE Trans. Signal Processing*, vol. 36, no. 4, pp. 477–489, 1988.

[58] M. Miller and D. Fuhrmann, "Maximum-Likelihood Narrow-Band Direction Finding and the EM Algorithm," *IEEE Trans. Signal Processing*, vol. 38, no. 9, pp. 1560–1577, 1990.

[59] N. Kabaoglu, H. Cirpan, E. Cekli, and S. Paker, "Maximum Likelihood 3-D Near-Field Source Localization Using the EM Algorithm," in *Proc. Eighth IEEE International Symposium on Computers and Communications (ISCC)*, 2003.

[60] D. Kraus, A. Dhaouadi, and J. Bohme, "EM Dual Maximum Likelihood Estimation for Wideband Source Location," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Minneapolis, MN, 1993.

[61] F. Asano and H. Asoh, "Sound Source Localization and Separation Based on the EM Algorithm," in *In Proc. SAPA*, Jeju, Korea, 2004.

[62] J. Fessler and A. Hero, "Space-Alternating Generalized Expectation-Maximization Algorithm," *IEEE Trans. Signal Processing*, vol. 42, no. 10, pp. 2664–2677, 1994.

[63] N. Cadalli and O. Arikan, "Wideband Maximum Likelihood Direction Finding and Signal Parameter Estimation by Using the Tree-Structured EM Algorithm," *IEEE Trans. Signal Processing*, vol. 47, no. 1, pp. 201–206, 1999.

[64] P. Chung and J. Bohme, "Comparative Convergence Analysis of EM and SAGE Algorithms in DOA estimation," *IEEE Trans. Signal Processing*, vol. 49, no. 12, pp. 2940–2949, 2001.

[65] ——, "Comparative Convergence Analysis of EM and SAGE Algorithms in DOA Estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Salt Lake City, UT, 2001.

[66] P. Stoica and K. Sharman, "Novel Eigenanalysis Method for Direction Estimation," *IEE Proceedings F, Radar and Signal Processing*, vol. 137, no. 1, pp. 19–26, 1990.

[67] ——, "Maximum Likelihood Methods for Direction-of-Arrival Estimation," *IEEE Trans. Signal Processing*, vol. 38, no. 7, pp. 1132–1143, 1990.

[68] M. Viberg, B. Ottersten, and T. Kailath, "Detection and Estimation in Sensor Arrays Using Weighted Subspace Fitting," *IEEE Trans. Signal Processing*, vol. 39, no. 11, pp. 2436–2449, 1991.

[69] M. Wax and T. Kailath, "Detection of Signals by Information Theoretic Criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, no. 2, pp. 387–392, 1985.

[70] A. Quinlan, J. P. Barbot, P. Larzabal, and M. Haardt, "Model Order Selection for Short Data: An Exponential Fitting Test (EFT)," *EURASIP JASP (European Journal Applied Signal Processing)*, Accepted for publication.

[71] C. Knapp and G. Carter, "The Generalized Correlation Method for Estimation of Time Delay," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 24, no. 4, pp. 320–327, 1976.

[72] J. Chen, K. Yao, and R. Hudson, "Source Localization and Beamforming," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 30–39, 2002.

[73] C. Reed, R. Hudson, and K. Yao, "Direct Joint Source Localization and Propagation Speed Estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Phoenix, AZ, 1999.

[74] A. Quinlan and F. Boland, "The Effect of Vibrato on Singer Localisation," in *Proc. GSPX conference, Santa Clara CA*, 2004.

[75] A.Quinlan and F. Boland, "Using the Singer's Formant to Reduce Inaccuracies in the Location of a Singer on Stage," in *Proc. ISSC conference, Belfast Ireland*, 2004.

[76] H. Wang and P. Chu, "Voice Source Localization for Automatic Camera Pointing System in Videoconferencing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Munich, Germany, 1997.

[77] M. Jian, A. Kot, and M. Er, "Performance Study of Time Delay Estimation in a Room Environment [Microphone Arrays]," in *Proc. IEEE Trans. Circuits Syst.*, Monterey, CA, 1998.

[78] T. Gustafsson, B. Rao, and M. Trivedi, "Analysis of Time-Delay Estimation in Reverberant Environments," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Orlando, FL, 2002.

[79] D. Bechler and K. Kroschel, "Three Different Reliability Criteria for Time Delay Estimates," in *Proc. 12th European Signal Processing Conference EUSIPCO*, Vienna, Austria, 2004.

[80] S. Birchfield and D. Gilmor, "Acoustic Source Direction by Hemisphere Sampling," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Salt Lake City, UT, 2001.

[81] S. Birchfield and D. Gillmor, "Fast Bayesian Acoustic Location," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Orlando, Fl, 2002.

[82] 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, n.11, pp. 1533–1538, 1987.

[83] L. L. Scharf and D. W. Tufts, "Rank Reduction for Modeling Stationary Signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, no. 3, pp. 350–355, 1987.

[84] T. W. Anderson, "Asymptotic Theory for Principal Component Analysis," *Ann. Math. Statist.*, vol. 34, pp. 122–148, 1963.

[85] A. T. James, "Test of Equality of Latent Roots of the Covariance Matrix," *Journal of Multivariate Analysis*, pp. 205–218, 1969.

[86] W. Chen, K. M. Wong, and J. Reilly, "Detection of the Number of Signals: A Predicted Eigen-Threshold Approach," *IEEE Trans. Signal Processing*, vol. 39, no. 5, pp. 1088–1098, 1991.

[87] P. Stoica and Y. Selen, "Model-Order Selection: A Review of Information Criterion Rules," *IEEE Signal Processing Magazine*, vol. 21, no. 4, pp. 36–47, 2004.

[88] H. Akaike, "A New Look at the Statistical Model Identification," *IEEE Trans. Automat. Contr.*, vol. 19, no. 6, pp. 716–723, 1974.

[89] J. Rissanen, "Modelling by Shortest Data Description Length," *Automatica*, vol. 14, pp. 465–471, 1978.

[90] M. Wax and I. Ziskind, "Detection of the Number of Coherent Signals by the MDL Principle," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 8, pp. 1190–1196, 1989.

[91] M. Kaveh, H. Wang, and H. Hung, "On the Theoretical Performance of a Class of Estimators of the Number of Narrow-Band Sources," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, n.9, pp. 1350–1352, 1987.

[92] K. Wong, Q. Zhang, J. Reilly, and P. Yip, "On Information Theoretic Criteria for Determining the Number of Signals in High Resolution Array Processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, n.11, pp. 1959–1971, 1990.

[93] P. M. Djurić, "Model Selection Based on Asymptotic Bayes Theory," in *Proc. of the Seventh SP Workshop on Statistical Signal and Array Processing*, 1994, pp. 7 – 10.

[94] W. B. Bishop and P. Djurić, "Model Order Selection of Damped Sinusoids in Noise by Predictive Densities," *IEEE Trans. Signal Processing*, vol. 44, no. 3, pp. 611–619, March 1996.

[95] H. L. V. Trees, *Optimum Array Processing*, ser. Detection, Estimation and Modulation Theory. New York: Wiley-Interscience, 2002, vol. 4.

[96] A. Di, "Multiple Source Location – A Matrix Decomposition Approach," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, no. 4, pp. 1086–1091, 1985.

[97] H. T. Wu, J. F. Yang, and F. K. Chen, "Source Number Estimators Using Transformed Gershgorin Radii," *IEEE Transactions on Signal Processing*, vol. 43, no. 6, pp. 1325–1333, 1995.

[98] A. P. Liavas and P. A. Regalia, "On the Behavior of Information Theoretic Criteria for Model Order Selection," *IEEE Trans. Signal Processing*, vol. 49, no. 8, pp. 1689–1695, 2001.

[99] A. Quinlan, J. P. Barbot, and P. Larzabal, "Automatic Determination of the Number of Targets Present when Using the Time Reversal Operator (TRO)," *Journal Acoustical Society of America (JASA)*, vol. 119, n.4, pp. 2220–2225, 2006.

[100] A. Quinlan, F. Boland, J. Barbot, and P. Larzabal, "Determining the Number of Speakers with a Limited Number of Samples," in *European Signal Processing Conference EUSIPCO*, Florence, 2006.

[101] N. L. Johnson and S. Kotz, in *Distributions in Statistics: Continuous Multivariate Distributions*. New York: John Wiley, 1972, ch. 38-39.

[102] P. R. Krishnaiah and F. J. Schurmann, "On the Evaluation of Some Distribution that Arise in Simultaneous Tests of the Equality of the Latents Roots of the Covariance Matrix," *Journal of Multivariate Analysis*, vol. 4, pp. 265–282, 1974.

[103] M. Tanter, J. L. Thomas, and M. Fink, "Time Reversal and the Inverse Filter," *J. Acoust. Soc. Am.*, vol. 108, no. 1, pp. 223–234, 2000.

[104] J. S. Bradley, R. D. Reich, and S. G. Norcross, "On the Combined Effects of Signal-to-Noise Ratio and Room Acoustics on Speech Intelligibility," *Journal Acoustical Society of America (JASA)*, vol. 108, no. 2, pp. 651–661, 2000.

[105] H. Golzer and M. Kleinschmidt, "Importance of Early and Late Reflections for Automatic Speech Recognition in Reverberant Environments," in *Proc. Elektronische Sprachsignalverarbeitung (ESSV)*, 2003.

[106] J. Chen, K. Yao, and R. Hudson, "Acoustic Source Localization and Beamforming Theory and Practice," *EURASIP Journal on Applied Signal Processing*, vol. 4, pp. 359–370, 2003.

[107] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai, "Exact and Large Sample Maximum Likelihood Techniques for Parameter Estimation and Detection in Array Processing," in *Radar Array Processing*, S. Haykin, J. Litva, and T. J. Shepherd, Eds. Berlin: Springer-Verlag, 1993, ch. 4, pp. 99–151.

[108] M. Fréchet, "Sur l'extension de certaines évaluations statistiques au cas de petits échantillons," *Rev. Int. Stat*, vol. 11, pp. 182–205, 1943.

[109] S. Narasimhan and J. L. Krolik, "Fundamental Limits on Acoustic Source Range Estimation Performance in Uncertain Ocean Channels," *Journal Acoustical Society of America (JASA)*, vol. 97, no. 1, pp. 215–226, 1999.

[110] A. Renaux, P. Forster, E. Chaumette, and P. Larzabal, "On the High SNR Conditional Maximum Likelihood Estimator Full Statistical Characterization," *IEEE Trans. Signal Processing*, vol. Accepted for Publication.

[111] E. W. Barankin, "Locally Best Unbiased Estimates," *Ann. Math. Stat.*, vol. 20, pp. 477–501, 1949.

[112] D. G. Chapman and H. Robbins, "Minimum Variance Estimation Without Regularity Assumptions," *Ann. Math. Stat.*, vol. 22, pp. 581–586, 1951.

[113] R. J. McAulay and L. P. Seidman, "A Useful Form of the Barankin Lower Bound and its Application to PPM Threshold Analysis," *IEEE Trans on Information Theory*, vol. 15, no. 2, pp. 273–279, 1969.

[114] J. Tabrikian and J. Krolik, "Barankin Bounds for Source Localization in an Uncertain Ocean Environment," *IEEE Trans. Signal Processing*, vol. 47, no. 11, pp. 2917–2927, 1999.

[115] A. Bhattacharrya, "On Some Analogues of the Amount of Information and their use in Statistical Estimation," *Sankya Indian J. of Stat.*, vol. 8, pp. 1–14, 201–218, 315–328, 1946.

[116] P. Forster and P. Larzabal, "On Lower Bounds for Deterministic Parameter Estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Orlando, Fl, 2002.

[117] F. E. Glave, "A New Look at the Barankin Lower Bound," *IEEE Trans. Inform. Theory*, vol. 18, no. 3, pp. 349–356, 1972.

[118] J. S. Abel, "A Bound on Mean-Square-Estimate Error," *IEEE Trans. Inform. Theory*, vol. 39, no. 5, pp. 1675–1680, Sept 1993.

[119] D. Rife and R. Boorstyn, "Single Tone Parameter Estimation from Discrete-time Observations," *IEEE Trans. Inform. Theory*, vol. 20, no. 9, pp. 591–598, 1974.

[120] E. Lehmann, *Theory of Point Estimation, 2nd Ed.* New York: Springer, 1998.

[121] R. Horn and C. Johnson, "Matrix Analysis," *Cambridge University Press*, 1999.

[122] E. Chaumette, P. Larzabal, and P. Forster, "On the Influence of a Detection Step on Lower Bounds for Deterministic Parameters estimation," *IEEE Trans. Signal Processing*, vol. 53, no. 11, pp. 4080–4090, Nov. 2005.

[123] T. Marzetta, "Computing the Barankin Bound, by Solving an Unconstrained Quadratic Optimization Problem," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Munich, 1997.

[124] L. Knockaert, "The Barankin Bound and Threshold Behaviour in Frequency Estimation," *IEEE Trans. Signal Processing*, vol. 45, no. 9, pp. 2398–2401, 1997.

[125] A. Quinlan, E. Chaumette, and P. Larzabal, "A Direct Method to Generate Approximations of the Barankin Bound," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Toulouse, 2006.

[126] R. McAulay and E. Hofstetter, "Barankin Bounds on Parameter Estimation," *IEEE Trans. Inform. Theory*, vol. 17, no. 6, pp. 669–676, 1971.

[127] I. Ziskind and M. Wax, "Maximum Likelihood Localization of Multiple Sources by Alternating Projection," *IEEE Trans. Signal Processing*, vol. 36, no. 10, pp. 1553–1560, 1988.

[128] R. T. OBrien and K. Kiriakidis, "Single-Snapshot Robust Direction Finding," *IEEE Trans. Signal Processing*, vol. 53, no. 6, pp. 1964–1978, 2005.

[129] K. Harmanci, J. Tabrikian, and J. Krolik, "Relationships between adaptive minimum variance beamforming and optimal source localization," *IEEE Trans. Signal Processing*, vol. 48, no. 1, pp. 1–12, 2000.

"Using the singer's formant to reduce inaccuracies in the location of a singer on stage," in Proc. Irish Signals and Systems Conference (ISSC), Belfast, Ireland, 2004.

"The effect of vibrato on singer localisation," in Proc. GSPx Conference, Santa Clara, CA, 2004.

"A direct method to generate approximations of the Barankin bound," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP), Toulouse, France, 2006.

"Determination of the Number of Wideband Acoustical Sources in a Reverberant Environment," In Proc. 14th European Signal Processing Conference (EUSIPCO) Florence, 2006.

"Automatic determination of the number of targets present when using the time reversal operator (TRO)," Journal Acoustical Society of America (JASA), vol.119, n.4, pp. 2220-2225, 2006.

"Model order selection for short data: An exponential fitting test (EFT)," EURASIP JASP (European Journal Applied Signal Processing), Accepted for publication.