DISSERTATION

PARAMETER ESTIMATION FROM COMPRESSED AND SPARSE MEASUREMENTS

Submitted by

Pooria Pakrooh

Department of Electrical and Computer Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2015

Doctoral Committee:

Advisor: Ali Pezeshki
Co-Advisor: Louis L. Scharf

Edwin K. P. Chong
J. Rockey Luo
Chris Peterson

ABSTRACT

PARAMETER ESTIMATION FROM COMPRESSED AND SPARSE MEASUREMENTS

In this dissertation, the problem of parameter estimation from compressed and sparse noisy measurements is studied. First, fundamental estimation limits of the problem are analyzed. For that purpose, the effect of compressed sensing with random matrices on Fisher information, the Cramér-Rao Bound (CRB) and the Kullback-Leibler divergence are considered. The unknown parameters for the measurements are in the mean value function of a multivariate normal distribution. The class of random compression matrices considered in this work are those whose distribution is right-unitary invariant. The compression matrix whose elements are i.i.d. standard normal random variables is one such matrix. We show that for all such compression matrices, the Fisher information matrix has a complex matrix beta distribution. We also derive the distribution of CRB. These distributions can be used to quantify the loss in CRB as a function of the Fisher information of the non-compressed data. In our numerical examples, we consider a direction of arrival estimation problem and discuss the use of these distributions as guidelines for deciding whether compression should be considered, based on the resulting loss in performance.

Then, the effect of compression on performance breakdown regions of parameter estimation methods is studied. Performance breakdown may happen when either the sample size or signal-to-noise ratio (SNR) falls below a certain threshold. The main reason for this threshold effect is that in low SNR or sample size regimes, many high resolution parameter estimation methods, including subspace methods as well as maximum likelihood estimation lose their capability to resolve signal and noise subspaces. This leads to a large error in parameter estimation. This phenomenon is called a subspace swap. The probability of a subspace swap for parameter estimation from compressed data is studied. A lower bound

has been derived on the probability of a subspace swap in parameter estimation from compressed noisy data. This lower bound can be used as a tool to predict breakdown for different compression schemes at different SNRs.

In the last part of this work, we look at the problem of parameter estimation for p damped complex exponentials, from the observation of their weighted and damped sum. This problem arises in spectrum estimation, vibration analysis, speech processing, system identification, and direction of arrival estimation. Our results differ from standard results of modal analysis to the extent that we consider sparse and co-prime samplings in space, or equivalently sparse and co-prime samplings in time. Our main result is a characterization of the orthogonal subspace. This is the subspace that is orthogonal to the signal subspace spanned by the columns of the generalized Vandermonde matrix of modes in sparse or co-prime arrays. This characterization is derived in a form that allows us to adapt modern methods of linear prediction and approximate least squares for estimating mode parameters. Several numerical examples are presented to demonstrate the performance of the proposed modal estimation methods. Our calculations of Fisher information allow us to analyze the loss in performance sustained by sparse and co-prime arrays that are compressions of uniform linear arrays.

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest gratitude to my advisor, Professor Ali Pezeshki, and my co-advisor Professor Louis L. Scharf, for their consistent guidance, encouragement, advice and support throughout my time as their PhD student. They have taught me innumerable lessons and insights, which will be useful not only for my academic research work, but also throughout my life. I have been extremely lucky to have the opportunity of working with them during the past four years. I have always tried to learn as much as I can from their positive attitude and valuable insights.

Special thanks go to my PhD committee members Professor Edwin K. P. Chong, Professor J. Rockey Luo, and Professor Chris Peterson, for their important suggestions and remarks in this research work and for their time and effort in service on my PhD committee despite their already heavy loads of responsibility.

During the past four years, I have had the chance of benefiting fruitful discussions with several great researchers in the field of signal processing. I would especially like to thank Professor Yuejie Chi, Professor Douglas Cochran, Dr. Stephen D. Howard and Professor William Moran for their insightful comments on my research.

I am thankful to my colleagues, Ramin Zahedi, Wenbing Dang, Yajing Luo, Somayeh Hosseini, Zhenliang Zhang and Mehrdad Khatami for all their help and useful discussions.

My heartfelt appreciations go to my girlfriend, Mahsa Sedighi, for her consistent patience and love during all these hard years of living apart. I am very thankful for her presence and support during my good and bad times.

Last but not least, I would like to thank my parents, Nastaran Ayoubi and Abbas Pakrooh, and my sisters, Ronak Pakrooh and Tarkan Pakrooh. I have been blessed with very loving and supportive parents who have sacrificed their lives for their children and provided unconditional love and care. My sisters have been my best friends in my life and I owe them huge thanks for all their support and kindness.

DEDICATION

To my beloved parents

TABLE OF CONTENTS

ABSTRACT ii

ACKNOWLEDGEMENTS iv

DEDICATION v

LIST OF FIGURES viii

1 INTRODUCTION 1

2 ANALYSIS OF FISHER INFORMATION AND THE CRAMÉR-RAO BOUND FOR NONLINEAR PARAMETER ESTIMATION AFTER COMPRESSED SENSING 6

2.1 Introduction 6

2.2 Problem Statement 7

2.3 Distribution of the Fisher Information Matrix after Compression 9

2.4 Numerical Results 13

3 THRESHOLD EFFECTS IN PARAMETER ESTIMATION FROM COMPRESSED DATA 17

3.1 Introduction 17

3.2 Measurement Model 18

3.2.1 Parameterized Mean Case 18

3.2.2 Parameterized Covariance Case 19

3.3 Bound on the Probability of a Subspace Swap after Compression 20

3.3.1 Parameterized Mean Case 22

3.3.2 Parameterized Covariance Case 24

3.4 Simulation Results 26

3.4.1 Parameterized Mean Case 28

3.4.2 Parameterized Covariance Case 30

4 MODAL ANALYSIS USING SPARSE AND CO-PRIME ARRAYS 32

4.1 Motivation 32

4.2 Problem Statement 34

4.3 Characterization of the Orthogonal Subspace for Uniform Line Arrays 37

4.4 Characterization of the Orthogonal Subspaces for Sparse and Co-prime Arrays . . . 39
4.4.1 Sparse Array . . . 39
4.4.2 Co-prime Array . . . 42
4.5 Numerical Results . . . 45
4.6 Acknowledgment . . . 47
5 SUMMARY . . . 51
5.1 Conclusions . . . 51
5.2 Future Work . . . 52
REFERENCES . . . 54
APPENDIX A . . . 58
APPENDIX B . . . 59
APPENDIX C . . . 60

LIST OF FIGURES

2.1 Histogram data and analytical distributions for $\frac{(\mathbf{J}^{-1})_{11}}{(\hat{\mathbf{J}}^{-1})_{11}}$ using 10^5 realizations of i.i.d. Gaussian compression matrices with $n = 128$ and $m = 64$. 14

2.2 Concentration ellipses for the Fisher information matrices before and after compression. 15

2.3 Compression ratios needed so that $(\hat{\mathbf{J}}^{-1})_{11} < \kappa(\mathbf{J}^{-1})_{11}$ for different confidence levels. . . 16

3.1 Signal and noise subspaces. 21

3.2 Geometry of the dense array (a), and co-prime subarrays (b), (c). At $m_1 = 11$ and $m_2 = 9$, $(2m_2 - 1)m_1\lambda/2 = 187\lambda/2$. 27

3.3 Parameterized mean case. Dense 188 element array and 28 element co-prime array. MSE bounds and MSE for ML estimation of $\theta_1 = 0$ in the presence of an interfering source at $\theta_2 = \pi/188$; 200 trials. 29

3.4 Parameterized mean case. Analytical lower bounds (event F) for the probability of subspace swap for estimation of the angle of a source at $\theta_1 = 0$ in the presence of an interfering source at $\theta_2 = \pi/188$ using 188 element dense array and 28 element co-prime array. 29

3.5 Parameterized covariance case. Dense 36 element array and 12 element co-prime array. MSE bounds and MSE for ML estimation of $\theta_1 = 0$ in the presence of an interfering source at $\theta_2 = \pi/36$; 200 snapshots and 200 trials. 30

3.6 Parameterized covariance case. Analytical lower bounds (event G) for the probability of a subspace swap using co-prime compression for the estimation of the angle of a source at $\theta_1 = 0$ in the presence of an interfering source at $\theta_2 = \pi/36$ using 36 element dense array and 12 element co-prime array. 31

4.1 The signal subspace $\langle \mathbf{V}(\mathbf{z}, \mathbb{I}) \rangle$ and the orthogonal subspace $\langle \mathbf{A}(\mathbf{z}, \mathbb{I}) \rangle = \langle \mathbf{V}(\mathbf{z}, \mathbb{I}) \rangle^{\perp}$. In the figure, we have dropped $(\mathbf{z}, , \mathbb{I})$ and have simply used $\mathbf{A}$, $\mathbf{V}$, $\langle \mathbf{A} \rangle$, and $\langle \mathbf{V} \rangle$. 36

4.2 Beampatterns for ULAs with 14 and 50 elements, a sparse array with 14 elements, $d = 4$, and $M = 3$, and a co-prime array with 14 elements, $m_1 = 7$, and $m_2 = 4$. 46

4.3 The CRB in dB for estimating $z_1 = 1$ in the presence of an interfering mode z_2: (a) ULA with 50 elements. (b) sparse array with 14 elements $d = 4$ and $M = 3$. (c) co-prime array with 23 elements $m_1 = 7$ and $m_2 = 4$. For all arrays per sensor SNR is 10 dB. 48

4.4 Estimating two closely spaced modes $z_1 = e^{j0.52}$ and $z_2 = 0.95e^{j0.69}$ using a ULA with 50 elements: (a) Per sensor SNR = 0 dB. (b) Per sensor SNR = -5 dB. 49

4.5 Estimating two closely spaced modes $z_1 = e^{j0.52}$ and $z_2 = 0.95e^{j0.69}$ using a sparse array with 14 elements, $d = 4$ and $M = 3$: (a) Per sensor SNR = 5 dB (b) Per sensor SNR = 0 dB. 49

4.6 Estimating two closely spaced modes $z_1 = e^{j0.52}$ and $z_2 = 0.95e^{j0.69}$ using a co-prime array with 14 elements, $m_1 = 7$ and $m_2 = 4$: (a) Per sensor SNR = 5 dB (b) Per sensor SNR = 0 dB. 50

CHAPTER 1

INTRODUCTION

There are many engineering applications in which the goal is to *invert an image* for the underlying parameters. Some examples are the identification of multipath components in wireless communications, radiating sources in radar and sonar, and light sources in optical imaging. Classical methods for inversion are mainly based on two principles. The first principle is matched filtering, where a sequence of test images is matched to the measured image. The test images are obtained from scanning prototype images (e.g., steering vectors) through frequency, wavenumber, doppler, or delay. The second principle for image inversion is parameter estimation in a separable linear model, wherein a sparse *modal* representation is considered for the image. Estimates of underlying linear parameters (complex amplitudes of modes) and nonlinear mode parameters (frequency, wavenumber, delay, and/or doppler) are extracted, usually based on maximum likelihood, or some variation on linear prediction, using ℓ_2 minimization (see, e.g., [1]). However, the ultimate limitation for image inversion is that any subsampling of the measured image has consequences for resolution (or bias) and for variability (or variance) [2].

Compressed sensing [3]–[5] is a relatively new theory which exploits sparse representations and sparse recovery for inversion. In compressed sensing, typically random compression matrices are used to compress the image. The elements of the compression matrices are normally drawn from specific distributions such as Gaussian or Bernoulli. Using concentration inequalities, the so-called restricted isometry constants are derived for different distributions of the random compression matrices [6]. Based on these restricted isometry constants, there are guarantees for the reconstruction of sparse images after compressed sensing, using greedy or ℓ_1-norm-based reconstruction methods (See e.g. [7]–[11]). Compressed sensing

demonstrates that for images that have sparse representations in an *apriori known* basis, subsampling has manageable consequences [3]–[5].

In [2]–[14], the sensitivity of sparse inversion algorithms to basis mismatch and frame mismatch are studied. It is shown in [2] that mismatch between the actual basis in which a signal has a sparse representation and the basis (or frame) which is used for sparsity in a sparse reconstruction algorithm, such as basis pursuit, has performance consequences on the estimated parameter vector. [13, 14] contain numerical results that characterize the increase in CRB after random compression for the case where the parameters nonlinearly modulate the mean in a multivariate normal measurement model.

As a deterministic compression scheme, co-prime sensor array processing was introduced recently by Vaidyanathan and Pal [15]- [16] as a sparse alternative to uniform line arrays. The concept was extended to sampling in multiple dimensions in subsequent papers by the same authors. In one dimension, the idea is to employ two uniform line arrays with spacings of M and N in units of half-wavelength, where M and N are co-prime. We will also look into the effect of co-prime subsampling on parameter estimation in Chapters 3 and 4.

Chapter 2 of this dissertation addresses a fundamental question: *How much information for nonlinear parameter estimation is retained (or lost) when compressing noisy measurements?* To answer this question, we analyze the effect of compressed sensing on the Fisher information matrix and the Cramér-Rao bound (CRB). The inverse of the Fisher information matrix gives a lower bound on the mean squared error (MSE) of any unbiased estimator of the parameters. Therefore, it quantifies the amount of information we lose due to compression, and the increase we expect in the MSE of the estimation of parameters after compressed sensing. The class of random compression matrices we consider are those whose distributions are invariant under right-unitary transformations. These include i.i.d draws of spherically invariant matrix rows, including, for example, i.i.d. draws of standard normal matrix elements. We consider a measurement model in which the unknown parameters for the measurements

are in the mean value function of a multivariate normal distribution. Then, we consider compression using random matrices whose distribution is right-unitary invariant. As a simple case, we analyze the distribution of the Fisher information matrix and the CRB for the case that the elements of the compression matrix are i.i.d. standard normal random variables. We show that the normalized Fisher information matrix after compressed sensing has a multivariate beta distribution. Then, we show that the same distribution result applies to all compressed sensing matrices of the aforementioned class. We also derive the distribution of the CRB in estimation of the individual parameters. Importantly, the distribution of the ratio of CRBs before and after compression depends only on the number of parameters and the number of measurements. The distribution is invariant to the underlying signal-plus-noise model, in the sense that it is invariant to the underlying Fisher information matrix, before compression. The analytical distributions obtained in Chapter 2 can be used to quantify the amount of loss due to compression. Also, they can be used as guidelines for choosing a suitable compression ratio based on a tolerable loss in the CRB.

In Chapter 3, we study threshold effects associated with the estimation of parameters from compressed noisy measurements. The term threshold effect refers to a catastrophic increase in mean-squared error when the signal-to-noise ratio falls below a threshold SNR. In many cases, the threshold effect is caused by a subspace swap event, when the measured data (or its sample covariance) is better approximated by a subset of components of an orthogonal subspace than by the components of a signal subspace. We consider measurement models in which the parameters to be estimated are either in the mean or in the covariance of a complex multivariate normal set of measurements. For these models, we derive analytical lower bounds on the probability of a subspace swap in compressively measured noisy data. These bounds guide our understanding of threshold effects and performance breakdown for parameter estimation using compression. As a case study, we investigate threshold effects in the maximum likelihood estimation of direction of arrivals of two closely-spaced sources, using co-prime subsampling of a uniform line array. Our results show the impact of compression

on threshold SNR, and can be used as a tool to predict the threshold SNR for different compression regimes. Other authors have addressed the performance breakdown regions of high resolution parameter estimation methods. In [17], approximation of the probability of a subspace swap in the Singular Value Decomposition (SVD) is investigated. In [18] lower bounds on the probability of a subspace swap are derived for the problem of modal analysis. In [19] the performance breakdown regions have been studied in the DOA estimation problem using asymptotic assumptions on the number of antennas and number of samples. It is shown that while a subspace swap is the main reason for the performance breakdown of maximum likelihood, earlier breakdown of MUSIC is due to the loss of resolution in separating closely spaced sources.

In chapter 4, we consider the problem of parameter estimation for p damped complex exponentials, from the observation of their weighted and damped sum. This problem has many applications such as spectrum estimation, vibration analysis, speech processing, system identification, and direction of arrival estimation. There is a vast literature on different modal estimation methods from uniformly sampled time or space series data, starting with the work of Prony [20]. Other methods include approximate least squares or maximum likelihood estimation [21], [22], reduced rank linear prediction [23], [24], MUSIC [25], and ESPRIT [26]. While there are extensions of MUSIC and ESPRIT for direction of arrival estimation from non-uniformly sampled data [27]–[31], Prony-like methods have mainly been developed for uniformly sampled data, and extending such methods to non-uniformly sampled data has not received much attention (exceptions being [32] and [33]).

We consider sparse and co-prime samplings in space, or equivalently sparse and co-prime samplings in time. Our main result is a characterization of the orthogonal subspace. This is the subspace that is orthogonal to the signal subspace spanned by the columns of the generalized Vandermonde matrix of modes in sparse or co-prime arrays. This characterization is derived in a form that allows us to adapt modern methods of linear prediction and approximate least squares for estimating mode parameters. Several numerical examples

are presented to demonstrate the performance of the proposed modal estimation methods. Although we present our numerical results in the context of sensor array processing, all of our results apply to the estimation of complex exponential modes from time series data. Our calculations of Fisher information allow us to analyze the loss in performance sustained by sparse and co-prime arrays that are compressions of uniform linear arrays.

The results of this thesis have been reported in [34]– [36].

CHAPTER 2

ANALYSIS OF FISHER INFORMATION AND THE CRAMÉR-RAO BOUND FOR NONLINEAR PARAMETER ESTIMATION AFTER COMPRESSED SENSING

2.1 Introduction

As mentioned in chapter 1, compressed sensing [3]–[5] is a relatively new theory which exploits sparse representations and sparse recovery for inversion. In this chapter, we analyze the impact of compressed sensing on Fisher information and the Cramér-Rao Bound for nonlinear parameter estimation from noisy measurements. We consider a case where the parameters nonlinearly modulate the mean value of a complex multivariate normal distribution. We derive the distribution of the Fisher information matrix and the CRB for the class of random matrices whose distributions are invariant under right-unitary transformations. These include i.i.d draws of spherically invariant matrix rows, including, for example, i.i.d. draws of standard normal matrix elements. The results in this chapter quantify the amount of information we lose due to compression, and the increase we expect in the MSE of parameter estimators after compressed sensing.

Other studies on the effect of compressed sensing on the CRB and the Fisher information matrix include [37]– [39]. Babadi *et al.* [37] proposed a "Joint Typicality Estimator" to show the existence of an estimator that asymptotically achieves the CRB of sparse parameter estimation for random Gaussian compression matrices. Niazadeh *el al.* [38] generalize the results of [37] to a class of random compression matrices which satisfy the concentration of

measures inequality. Nielsen *et al.* [40] derive the mean of the Fisher information matrix for the same class of random compression matrices that we are considering. Ramasamy *et al.* [39] derive bounds on the Fisher information matrix, but not for the model we are considering. We will clarify the important distinction between our work and [39] after establishing our notation in Section 2.2.

2.2 Problem Statement

Let $\mathbf{y} \in \mathbb{C}^n$ be a complex random vector whose probability density function $f(\mathbf{y};\boldsymbol{\theta})$ is parameterized by an unknown but deterministic parameter vector $\boldsymbol{\theta} \in \mathbb{R}^p$. The derivative of the log-likelihood function with respect to $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_p]$ is called the Fisher score, and the covariance matrix of the Fisher score is the Fisher information matrix which we denote by $\mathbf{J}(\boldsymbol{\theta})$:

$$\mathbf{J}(\boldsymbol{\theta}) = E[(\frac{\partial \log f(\mathbf{y};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})(\frac{\partial \log f(\mathbf{y};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})^H] \in \mathbb{C}^{p\times p}. \tag{2.1}$$

The inverse $\mathbf{J}^{-1}(\boldsymbol{\theta})$ of the Fisher information matrix lower bounds the error covariance matrix for any unbiased estimator $\hat{\boldsymbol{\theta}}(\mathbf{y})$ of $\boldsymbol{\theta}$, that is

$$E[(\hat{\boldsymbol{\theta}}(\mathbf{y}) - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}(\mathbf{y}) - \boldsymbol{\theta})^H] \succeq \mathbf{J}^{-1}(\boldsymbol{\theta}), \tag{2.2}$$

where $\mathbf{A} \succeq \mathbf{B}$ for matrices $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{n\times n}$ means $\mathbf{a}^H\mathbf{A}\mathbf{a} \geq \mathbf{a}^H\mathbf{B}\mathbf{a}$ for all $\mathbf{a} \in \mathbb{C}^n$. The i^{th} diagonal element of $\mathbf{J}^{-1}(\boldsymbol{\theta})$ is the CRB for estimating θ_i and it gives a lower bound on the MSE of any unbiased estimator of θ_i from $\mathbf{y}$ (see, e.g., [41]).

For $\mathbf{y} \in \mathbb{C}^n$ a proper complex normal random vector distributed as $\mathcal{CN}(\mathbf{x}(\boldsymbol{\theta}), \mathbf{C})$ with unknown mean vector $\mathbf{x}(\boldsymbol{\theta})$ parameterized by $\boldsymbol{\theta}$, and known covariance $\mathbf{C} = \sigma^2\mathbf{I}$, the Fisher information matrix is the Grammian

$$\mathbf{J}(\boldsymbol{\theta}) = \mathbf{G}^H\mathbf{C}^{-1}\mathbf{G} = \frac{1}{\sigma^2}\mathbf{G}^H\mathbf{G}. \tag{2.3}$$

The i^{th} column $\mathbf{g}_i$ of $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_p]$ is the partial derivative $\mathbf{g}_i = \frac{\partial}{\partial\theta_i}\mathbf{x}(\boldsymbol{\theta})$, which characterizes the sensitivity of the mean vector $\mathbf{x}(\boldsymbol{\theta})$ to variation of the i^{th} parameter θ_i.

The CRB for estimating θ_i is given by

$$(\mathbf{J}^{-1}(\boldsymbol{\theta}))_{ii} = \sigma^2(\mathbf{g}_i^H(\mathbf{I} - \mathbf{P}_{\mathbf{G_i}})\mathbf{g}_i)^{-1}, \tag{2.4}$$

where $\mathbf{G}_i$ consists of all columns of $\mathbf{G}$ except $\mathbf{g}_i$, and $\mathbf{P}_{\mathbf{G}_i}$ is the orthogonal projection onto the column space of $\mathbf{G}_i$ [42]. This CRB can also be written as

$$(\mathbf{J}^{-1}(\boldsymbol{\theta}))_{ii} = \frac{\sigma^2}{\|\mathbf{g}_i\|_2^2 \sin^2(\psi_i)}, \tag{2.5}$$

where ψ_i is the principal angle between subspaces $\langle \mathbf{g}_i \rangle$ and $\langle \mathbf{G}_i \rangle$. These representations illuminate the geometry of the CRB, which is discussed in detail in [42].

If $\mathbf{y}$ is compressed by a compression matrix $\boldsymbol{\Phi} \in \mathbb{C}^{m \times n}$ to produce $\hat{\mathbf{y}} = \boldsymbol{\Phi}\mathbf{y}$, then the probability density function of the compressed data $\hat{\mathbf{y}}$ is $\mathcal{CN}[\boldsymbol{\Phi}\mathbf{x}(\theta), \sigma^2\boldsymbol{\Phi}\boldsymbol{\Phi}^H]$. The Fisher information matrix $\hat{\mathbf{J}}(\boldsymbol{\theta})$ is given by

$$\hat{\mathbf{J}}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\hat{\mathbf{G}}^H\hat{\mathbf{G}}, \tag{2.6}$$

where $\hat{\mathbf{G}} = \mathbf{P}_{\boldsymbol{\Phi}^H}\mathbf{G}$. The CRB for estimating the i^{th} parameter is

$$(\hat{\mathbf{J}}^{-1}(\boldsymbol{\theta}))_{ii} = \sigma^2(\hat{\mathbf{g}}_i^H(\mathbf{I} - \mathbf{P}_{\hat{\mathbf{G}}_\mathbf{i}})\hat{\mathbf{g}}_i)^{-1}, \tag{2.7}$$

where $\hat{\mathbf{G}}_i = \mathbf{P}_{\boldsymbol{\Phi}^H}\mathbf{G}_i$, and $\mathbf{P}_{\boldsymbol{\Phi}^H} = \boldsymbol{\Phi}^H(\boldsymbol{\Phi}\boldsymbol{\Phi}^H)^{-1}\boldsymbol{\Phi}$ is the orthogonal projection onto the row span of $\boldsymbol{\Phi}$.

Our aim is to study the effect of random compression on the Fisher information matrix and the CRB. In Section 2.3 we investigate this problem by deriving the distributions of the Fisher information matrix and the CRB for the case in which the elements of the compression matrix $\boldsymbol{\Phi}$ are distributed as i.i.d. standard normal random variables. Then we demonstrate that the same analysis holds for a wider range of random compression matrices.

Remark 1: In parallel to our work, Ramasamy et al. [39] have also looked at the impact of compression on Fisher information. However, they have considered a different parameter model. Specifically, their compressed data has density $\mathcal{CN}[\boldsymbol{\Phi}\mathbf{x}(\boldsymbol{\theta}), \sigma^2\mathbf{I}]$, in contrast to ours which is distributed as $\mathcal{CN}[\boldsymbol{\Phi}\mathbf{x}(\boldsymbol{\theta}), \sigma^2\boldsymbol{\Phi}\boldsymbol{\Phi}^H]$. Our model is a signal-plus-noise model, wherein

the noisy signal $\mathbf{x}(\boldsymbol{\theta})+\mathbf{n}$, $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$, is compressed to produce $\boldsymbol{\Phi}\mathbf{x}(\boldsymbol{\theta})+\boldsymbol{\Phi}\mathbf{n}$. In contrast, their model corresponds to compressing a noiseless signal $\mathbf{x}(\boldsymbol{\theta})$ to produce $\boldsymbol{\Phi}\mathbf{x}(\theta)+\mathbf{w}$, where $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$ represents post-compression noise. Note that the Fisher information, CRB and corresponding bounds of these two models are different, as in our model noise enters at the input of the compressor, whereas in [39] noise enters at the output of the compressor. This is an important distinction.

2.3 Distribution of the Fisher Information Matrix after Compression

Let $\mathbf{W}$ be the normalized Fisher Information Matrix after compression, defined as

$$\mathbf{W} = \mathbf{J}^{-1/2}\hat{\mathbf{J}}\mathbf{J}^{-H/2} \in \mathbb{C}^{p\times p}, \tag{2.8}$$

where $\mathbf{J}$ and $\hat{\mathbf{J}}$ are the Fisher information matrix before compression and the Fisher information matrix after compression, defined in (2.3) and (2.6), respectively. Our aim is to derive the distribution of the random matrix $\mathbf{W}$ for the case that the elements of the compression matrix $\boldsymbol{\Phi}_{ij}$ are i.i.d. random variables distributed as $\mathcal{CN}(0,1)$. We assume $n-p \geq m$, which is typical in almost all compression scenarios of interest.

Using (2.3) and (2.6), $\mathbf{W}$ may be written as

$$\mathbf{W} = \mathbf{H}^H\mathbf{P}_{\boldsymbol{\Phi}^H}\mathbf{H}, \tag{2.9}$$

where $\mathbf{H} = \mathbf{G}(\mathbf{G}^H\mathbf{G})^{-H/2}$ is a left-unitary matrix, i.e., $\mathbf{H}^H\mathbf{H} = \mathbf{I}_p$. Define $\tilde{\mathbf{H}} \in \mathbb{C}^{n\times(n-p)}$ such that $\boldsymbol{\Lambda} = [\mathbf{H}|\tilde{\mathbf{H}}]$ is an orthonormal basis for $\mathbb{C}^n$, i.e., $\boldsymbol{\Lambda}\boldsymbol{\Lambda}^H = \boldsymbol{\Lambda}^H\boldsymbol{\Lambda} = \mathbf{I}_n$. Then we have:

$$\mathbf{W} = \left[\begin{array}{c|c} \mathbf{I}_p & \mathbf{0} \end{array}\right] \boldsymbol{\Lambda}^H\mathbf{P}_{\boldsymbol{\Phi}^H}\boldsymbol{\Lambda} \left[\begin{array}{c} \mathbf{I}_p \\ \hline \mathbf{0} \end{array}\right], \tag{2.10}$$

where

$$\begin{aligned} \boldsymbol{\Lambda}^H\mathbf{P}_{\boldsymbol{\Phi}^H}\boldsymbol{\Lambda} &= \boldsymbol{\Lambda}^H\boldsymbol{\Phi}^H(\boldsymbol{\Phi}\boldsymbol{\Phi}^H)^{-1}\boldsymbol{\Phi}\boldsymbol{\Lambda} \\ &= \boldsymbol{\Lambda}^H\boldsymbol{\Phi}^H(\boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^H\boldsymbol{\Phi}^H)^{-1}\boldsymbol{\Phi}\boldsymbol{\Lambda} \\ &= \mathbf{P}_{\boldsymbol{\Lambda}^H\boldsymbol{\Phi}^H}. \end{aligned} \tag{2.11}$$

Because the distribution of $\mathbf{\Phi}$ is right-unitary invariant, the distribution of $\mathbf{\Lambda}^H \mathbf{P}_{\mathbf{\Phi}^H} \mathbf{\Lambda}$ is the same as the distribution of $\mathbf{P}_{\mathbf{\Phi}^H}$. Therefore, the distribution of $\mathbf{W}$ is the same as the distribution of $\mathbf{V} = \mathbf{\Phi}_1^H (\mathbf{\Phi}\mathbf{\Phi}^H)^{-1} \mathbf{\Phi}_1$, where $\mathbf{\Phi} = [\mathbf{\Phi}_1 | \mathbf{\Phi}_2]$, $\mathbf{\Phi}_1 \in \mathbb{C}^{m \times p}$ and $\mathbf{\Phi}_2 \in \mathbb{C}^{m \times (n-p)}$. Now, write $\mathbf{V}$ as $\mathbf{V} = \mathbf{Y}\mathbf{Y}^H$, with $\mathbf{Y} = \mathbf{\Phi}_1^H \mathbf{Z}^{-1/2}$ and $\mathbf{Z} = \mathbf{\Phi}\mathbf{\Phi}^H$. Since $\mathbf{Z} = \mathbf{\Phi}_1 \mathbf{\Phi}_1^H + \mathbf{\Phi}_2 \mathbf{\Phi}_2^H$, and $\mathbf{\Phi}_2 \mathbf{\Phi}_2^H$ is distributed as complex Wishart $\mathcal{W}_c(\mathbf{I}_m, m, n-p)$ for $n - p \geq m$, given $\mathbf{\Phi}_1$ the pdf of $\mathbf{Z}$ is

$$f(\mathbf{Z}|\mathbf{\Phi}_1) = c_1 e^{-tr(\mathbf{Z} - \mathbf{\Phi}_1 \mathbf{\Phi}_1^H)} |\mathbf{Z} - \mathbf{\Phi}_1 \mathbf{\Phi}_1^H|^{n-p-m}. \tag{2.12}$$

The pdf of $\mathbf{\Phi}_1$ can be written as $c_2 e^{-tr(\mathbf{\Phi_1}\mathbf{\Phi_1^H})}$. Therefore, the joint pdf of $\mathbf{Z}$ and $\mathbf{\Phi}_1$ is

$$f(\mathbf{Z}, \mathbf{\Phi}_1) = c_3 e^{-tr(\mathbf{Z})} |\mathbf{Z} - \mathbf{\Phi}_1 \mathbf{\Phi}_1^H|^{n-p-m} \tag{2.13}$$

where c_1, c_2, and $c_3 = c_1 c_2$ are normalization factors. Since $\mathbf{Y} = \mathbf{\Phi}_1^H \mathbf{Z}^{-1/2}$, from (2.13) the joint pdf of $\mathbf{Y}$ and $\mathbf{Z}$ is

$$\begin{aligned} f(\mathbf{Y}, \mathbf{Z}) &= c_3 e^{-tr(\mathbf{Z})} |\mathbf{Z} - \mathbf{Z}^{1/2} \mathbf{Y}^H \mathbf{Y} \mathbf{Z}^{H/2}|^{n-p-m} |\mathbf{Z}|^p \\ &= c_3 e^{-tr(\mathbf{Z})} |\mathbf{I}_m - \mathbf{Y}^H \mathbf{Y}|^{n-p-m} |\mathbf{Z}|^{n-m}. \end{aligned} \tag{2.14}$$

This shows that $\mathbf{Y}$ and $\mathbf{Z}$ are independent and the pdf of $\mathbf{Y}$ is

$$\begin{aligned} f(\mathbf{Y}) &= c_4 |\mathbf{I}_m - \mathbf{Y}^H \mathbf{Y}|^{n-p-m} \\ &= c_4 |\mathbf{I}_p - \mathbf{Y}\mathbf{Y}^H|^{n-p-m}, \end{aligned} \tag{2.15}$$

where c_4 is a normalization factor.

Let $g(\mathbf{Y}\mathbf{Y}^H) = f(\mathbf{Y})$. To derive the distribution of $\mathbf{V} = \mathbf{Y}\mathbf{Y}^H$ we use the following theorem.

Theorem 1: [43] If the density of $\mathbf{Y} \in \mathbb{C}^{p \times m}$ is $g(\mathbf{Y}\mathbf{Y}^H)$, then the density of $\mathbf{V} = \mathbf{Y}\mathbf{Y}^H$ is

$$\frac{|\mathbf{V}|^{m-p} g(\mathbf{V}) \pi^{mp}}{\tilde{\Gamma}_p(m)}, \tag{2.16}$$

where $\tilde{\Gamma}_m(.)$ is the complex multivariate Gamma function. The tilde notation is standard in the literature.

Using Theorem 1 and (2.15), the pdf of $\mathbf{V}$ is

$$c_5|\mathbf{V}|^{m-p}|\mathbf{I}_p - \mathbf{V}|^{n-p-m} \quad \text{for} \quad \mathbf{0} \preceq \mathbf{V} \preceq \mathbf{I}_p, \tag{2.17}$$

which is the Type I complex multivariate beta distribution $\mathbb{C}B_p^I(m, n-m)$ for $c_5 = \frac{\tilde{\Gamma}_p(n)}{\tilde{\Gamma}_p(m)\tilde{\Gamma}_p(n-m)}$. Recall that the distribution of the normalized Fisher information matrix after compression $\mathbf{W} = \mathbf{J}^{-1/2}\hat{\mathbf{J}}\mathbf{J}^{-H/2}$ is identical to that of $\mathbf{V}$. Therefore, $\mathbf{W}$ is also distributed as $\mathbb{C}B_p^I(m, n-m)$, with the pdf of (2.17).

Remark 2: It is important to note that the distribution of $\mathbf{W} = \mathbf{J}^{-1/2}\hat{\mathbf{J}}\mathbf{J}^{-H/2}$ is *invariant* to $\mathbf{J}$, and it depends on only on the parameters $(n-p)-m$ and $m-p$. In this sense, this result for the distribution of $\mathbf{J}^{-1/2}\hat{\mathbf{J}}\mathbf{J}^{-H/2}$ is universal, and reminiscent of the classical result of Reed, Mallat, and Brennan [44] for normalized SNR in adaptive filtering.

Lemma 1: [45] Assume $\mathbf{A} \in \mathbb{C}^{p\times p}$ is a positive definite Hermitian random matrix with a pdf $h(\mathbf{A})$. Then, the joint pdf of eigenvalues $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$ of $\mathbf{A}$ is

$$\frac{\pi^{p(p-1)}}{\tilde{\Gamma}_p(p)} \prod_{i<j}^{p} (\lambda_i - \lambda_j)^2 \int_{\mathcal{U}(p)} h(\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H) d\mathbf{U}, \tag{2.18}$$

where $d\mathbf{U}$ is the invariant Haar measure on the unitary group $\mathcal{U}(p)$.

Using Lemma 1, we can derive the joint distribution of the eigenvalues of $\mathbf{W}$. Replacing h in (2.18) by the pdf of $\mathbf{W} \sim \mathbb{C}B_p^I(m, n-m)$, the joint pdf of the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$ of $\mathbf{W} = \mathbf{J}^{-1/2}\hat{\mathbf{J}}\mathbf{J}^{-H/2}$ is given by

$$\frac{\pi^{p(p-1)}\tilde{\Gamma}_p(n)}{\tilde{\Gamma}_p(p)\tilde{\Gamma}_p(m)\tilde{\Gamma}_p(n-m)} \prod_{i<j}^{p} (\lambda_i - \lambda_j)^2 \prod_{i=1}^{p} \lambda_i^{m-p}(1-\lambda_i)^{n-m-p}. \tag{2.19}$$

Now, from (2.17) and using the transformation $\hat{\mathbf{J}} = \mathbf{J}^{1/2}\mathbf{W}\mathbf{J}^{H/2}$, the Fisher information matrix after compression $\hat{\mathbf{J}}$ is distributed as

$$c_5|\mathbf{J}|^{p-n}|\hat{\mathbf{J}}|^{m-p}|\mathbf{J} - \hat{\mathbf{J}}|^{n-m-p} \quad \text{for} \quad \mathbf{0} \preceq \hat{\mathbf{J}} \preceq \mathbf{J} \tag{2.20}$$

and the inverse of the Fisher information matrix after compression $\hat{\mathbf{K}} = \hat{\mathbf{J}}^{-1}$ is distributed as

$$c_5|\mathbf{J}|^{p-n}|\hat{\mathbf{K}}|^{-n}|\mathbf{J}\hat{\mathbf{K}} - \mathbf{I}_p|^{n-m-p} \quad \text{for} \quad \hat{\mathbf{K}} \succeq \mathbf{J}^{-1}. \tag{2.21}$$

Remark 3: For the class of random compression matrices that have density functions of the form $g(\mathbf{\Phi}\mathbf{\Phi}^H)$, that is, the distribution of $\mathbf{\Phi}$ is right-unitary invariant, $\mathbf{\Phi}^H(\mathbf{\Phi}\mathbf{\Phi}^H)^{-1/2}$ is uniformly distributed on the Stiefel manifold $\mathcal{V}_m(\mathbb{C}^n)$ [46]. Therefore, the distribution of the normalized Fisher information matrix for this class of compression matrices is the same as the one given in (2.17).

Remark 4: Using the properties of a complex multivariate beta distribution [47], we have

$$E[\hat{\mathbf{J}}] = \frac{m}{n}\mathbf{J}, \tag{2.22}$$

and

$$E[\hat{\mathbf{J}}^{-1}] = \frac{n-p}{m-p}\mathbf{J}^{-1}. \tag{2.23}$$

This shows that, on average, compression results in a factor $\frac{n-m}{n}$ loss in the Fisher information and a factor $\frac{n-p}{m-p}$ increase in the CRB $\mathbf{J}^{-1}$.

The distribution of the CRB after compression can be derived using the following Lemma.

Lemma 2: [47] Assume $\mathbf{X} \sim \mathbb{C}B_p^I(a_1, a_2)$. Let $\mathbf{z}$ be a complex vector independent of $\mathbf{X}$. Then, $x = \frac{\mathbf{z}^H\mathbf{z}}{\mathbf{z}^H\mathbf{X}^{-1}\mathbf{z}}$ is distributed as $B_p^I(a_1 - p + 1, a_2)$, which is a Type I univariate beta distribution with the pdf

$$\frac{\Gamma_p(a_1 + a_2)}{\Gamma_p(a_1)\Gamma_p(a_2)} x^{a_1 - 1}(1-x)^{a_2 - 1} \quad \text{for} \quad 0 < x < 1. \tag{2.24}$$

Now consider the CRB on an unbiased estimator of parameter θ_i, after compression, normalized by the CRB before compression:

$$\frac{(\hat{\mathbf{J}}^{-1})_{ii}}{(\mathbf{J}^{-1})_{ii}} = \frac{\mathbf{e}_i^H\hat{\mathbf{J}}^{-1}\mathbf{e}_i}{\mathbf{e}_i^H\mathbf{J}^{-1}\mathbf{e}_i} = \frac{\mathbf{z}^H\mathbf{W}^{-1}\mathbf{z}}{\mathbf{z}^H\mathbf{z}}, \tag{2.25}$$

where $\mathbf{z} = \mathbf{J}^{-1/2}\mathbf{e}_i$, and $\mathbf{e}_i \in \mathbb{C}^p$ is a standard unit vector with 1 as its ith element and zeros as its other elements. By Lemma 1, the above ratio is distributed as the inverse of a univariate beta random variable $B^I(m - p + 1, n - m)$.

Remark 5: From the distribution of the CRB after compression, we have

$$E[(\hat{\mathbf{J}}^{-1})_{ii}] = (\frac{n-p}{m-p})(\mathbf{J}^{-1})_{ii}, \tag{2.26}$$

$$\text{var}[(\hat{\mathbf{J}}^{-1})_{ii}] = \frac{(n-m)(n-p)}{(m-p-1)(m-p)^2}((\mathbf{J}^{-1})_{ii})^2. \tag{2.27}$$

Remark 6: We can also look at the effect of compressed sensing on the Kullback Leibler (KL) divergence of two normal probability distributions, for the class of random compressors already discussed in Remark 3. The KL divergence between $\mathcal{CN}(\mathbf{x}(\boldsymbol{\theta}), \mathbf{C})$ and $\mathcal{CN}(\mathbf{x}(\boldsymbol{\theta}'), \mathbf{C})$ is:

$$D(\boldsymbol{\theta}, \boldsymbol{\theta}') = (\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}'))^H \mathbf{C}^{-1} (\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}')). \tag{2.28}$$

After compression with $\boldsymbol{\Phi}$ we have

$$\begin{aligned}\hat{D}(\boldsymbol{\theta}, \boldsymbol{\theta}') =& (\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}'))^H \boldsymbol{\Phi}^H (\boldsymbol{\Phi}\mathbf{C}\boldsymbol{\Phi}^H)^{-1} \boldsymbol{\Phi} \\ & (\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}')).\end{aligned} \tag{2.29}$$

For the case $\mathbf{C} = \sigma^2 \mathbf{I}$, the normalized KL divergence is

$$\frac{\hat{D}(\boldsymbol{\theta}, \boldsymbol{\theta}')}{D(\boldsymbol{\theta}, \boldsymbol{\theta}')} = \frac{(\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}'))^H \mathbf{P}_{\boldsymbol{\Phi}^H} (\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}'))}{(\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}'))^H (\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}'))}. \tag{2.30}$$

Therefore, the normalized KL divergence, for random compression matrices $\boldsymbol{\Phi}$ whose distributions are invariant to right-unitary transformations, is distributed as $B^I(m, n-m)$.

2.4 Numerical Results

As a special example, we consider the effect of compression on DOA estimation using a uniform line array with n elements. In our simulations, we consider two sources whose electrical angles θ_1 and θ_2 are unknown. The mean vector $\mathbf{x}(\boldsymbol{\theta})$ is $\mathbf{x}(\boldsymbol{\theta}) = \mathbf{x}(\boldsymbol{\theta}_1) + \mathbf{x}(\boldsymbol{\theta}_2)$, where

$$\mathbf{x}(\theta_i) = A_i e^{j\phi_i} [1 \;\; e^{j\theta_i} \;\; e^{j2\theta_i} \;\; \cdots \;\; e^{j(n-1)\theta_i}]^T. \tag{2.31}$$

Here A_i and ϕ_i are the amplitude and phase of the i^{th} source, which we assume known. We set $\phi_1 = \phi_2 = 0$ and $A_1 = A_2 = 1$. We wish to estimate θ_1, whose true value in this

example is zero, in the presence of the interfering source at electrical angle $\theta_2 = \pi/n$ (half the Rayleigh limit of the array). For our simulations, we use Gaussian compression matrices $\mathbf{\Phi}_{m\times n}$ whose elements are i.i.d. $\mathcal{CN}(0, 1/m)$. The Fisher information matrix and the CRB on the estimation of θ_1 are calculated for different realizations of $\mathbf{\Phi}$. Fig. 2.1 shows the CRB on the estimation of θ_1 before compression divided by its corresponding value after compression, i.e. $\frac{(\mathbf{J}^{-1})_{11}}{(\hat{\mathbf{J}}^{-1})_{11}}$ for $m = 64$, $n = 128$. A histogram of actual values of $\frac{(\mathbf{J}^{-1})_{11}}{(\hat{\mathbf{J}}^{-1})_{11}}$ for 10^5 independent realizations of random $\mathbf{\Phi}$ is shown in blue. The red curve represents the pdf of a $B^I(m-p+1, n-m)$ distributed random variable for $p = 2$. This figure simply provides an illustration of the result (2.25).

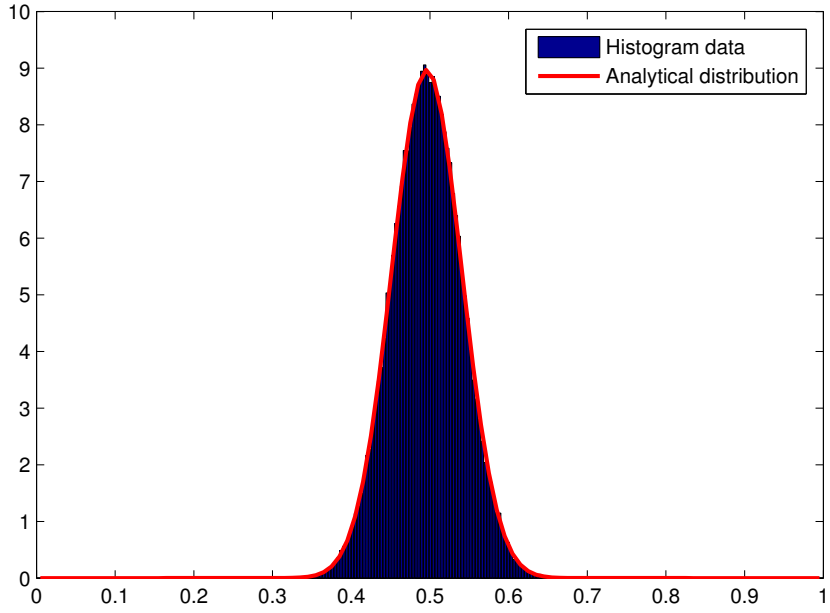


Figure 2.1: Histogram data and analytical distributions for $\frac{(\mathbf{J}^{-1})_{11}}{(\hat{\mathbf{J}}^{-1})_{11}}$ using 10^5 realizations of i.i.d. Gaussian compression matrices with $n = 128$ and $m = 64$.

Recall that the inverse Fisher Information Matrix $\mathbf{J}^{-1}$ lower bounds the error covariance matrix $\mathbf{\Sigma} = E[\mathbf{e}\mathbf{e}^H]$ for unbiased errors $\mathbf{e} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$. So the concentration ellipse $\mathbf{e}^H\mathbf{\Sigma}^{-1}\mathbf{e} \leq \mathbf{e}^H\mathbf{J}\mathbf{e}$ for all $\mathbf{e} \in \mathbb{C}^p$. The ellipses $\mathbf{e}^H\mathbf{J}\mathbf{e} = r^2$ and $\mathbf{e}^H\hat{\mathbf{J}}\mathbf{e}$, with $r^2 = \mathbf{J}_{11}$, are illustrated in Fig. 2.2, demonstrating the effect that compression inflates the concentration ellipse. The

blue curve is the locus of all points $\mathbf{e} \in \mathbb{C}^p$, for which $\mathbf{e}^H \mathbf{J} \mathbf{e} = r^2$. The red curves are the loci of all points $\mathbf{e} \in \mathbb{C}^p$, for which $\mathbf{e}^H \hat{\mathbf{J}} \mathbf{e} = r^2$ for 100 realizations of the Fisher information matrix after compression. As can be seen, the concentration ellipse for the Fisher information matrix before compression has the smallest volume in comparison with all the realizations of the concentration ellipses after compression. Also, for each realization of the Gaussian compression, the orientation of the concentration ellipse is nearly aligned with that of the uncompressed ellipse.

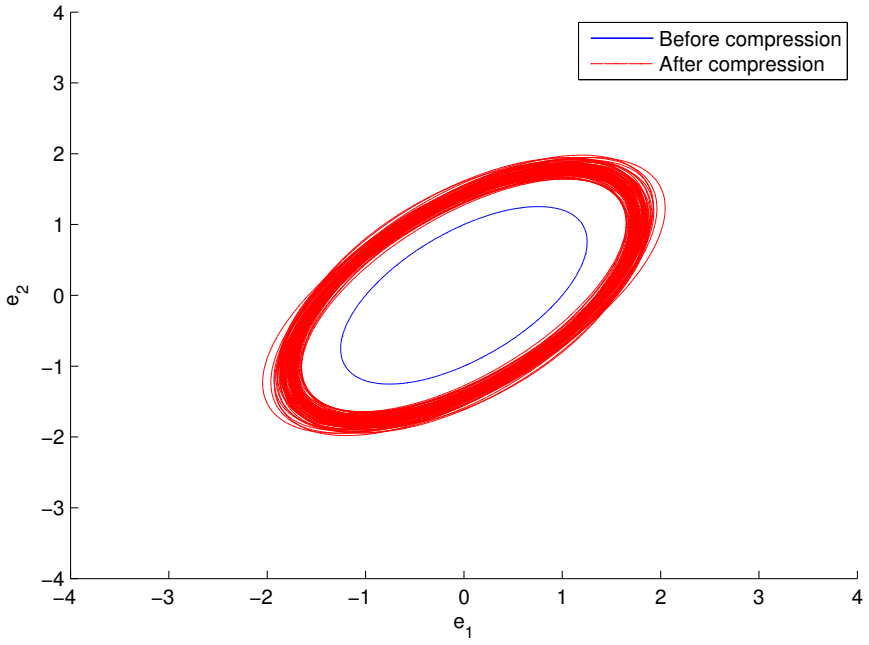


Figure 2.2: Concentration ellipses for the Fisher information matrices before and after compression.

Figure 2.3 shows the compression ratio m/n needed so that the CRB after compression $(\hat{\mathbf{J}}^{-1})_{11}$ does not exceed κ times the CRB before compression $(\mathbf{J}^{-1})_{11}$, at two levels of confidence and for $n = 128$. These curves are plotted using the tail probabilities of a univariate beta random variables. They can be used as guidelines for deriving a satisfactory compression ratio based on a tolerable level of loss in the CRB. Alternatively, we can plot the confidence level curves versus m for fixed values of κ . In that case, the plots may be useful to find a number of measurements that would guarantee that after compression CRB

does not go above a desired bound (corresponding to a particular κ) with a certain level of confidence.

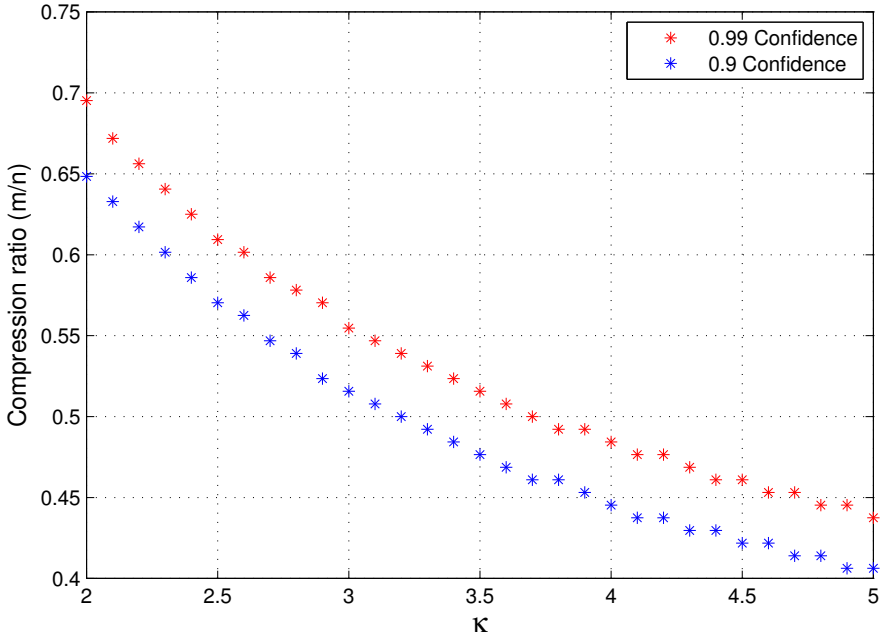


Figure 2.3: Compression ratios needed so that $(\hat{\mathbf{J}}^{-1})_{11} < \kappa(\mathbf{J}^{-1})_{11}$ for different confidence levels.

CHAPTER 3

THRESHOLD EFFECTS IN PARAMETER ESTIMATION FROM COMPRESSED DATA

3.1 Introduction

Many high resolution parameter estimation methods suffer from performance breakdown, where the mean squared error (MSE) increases dramatically at low SNR. Performance breakdown may happen when either the sample size or signal-to-noise ratio (SNR) falls below a certain threshold [17]. Typically, a subspace swap is known to be the main source of this performance breakdown, where one or more components in the orthogonal (noise) subspace better approximate the data than at least one component of the signal subspace, which in turn leads to a large error in parameter estimation [17]- [19].

In this chapter, we study the effect of compression on the probability of a subspace swap. Specifically, we want to see what effects compression has on the threshold SNR at which performance breaks down. To answer this question, we derive lower bounds on the probability of a subspace swap in parameter estimation from compressed noisy data in complex multivariate normal measurement models. These lower bound can be used as a tool to predict breakdown for different compression schemes at different SNRs, and therefore to predict whether specific compression or subsampling schemes are viable in a specific application. For our numerical results, we consider DOA estimation of two closely spaced sources and investigate the effect of compression with co-prime arrays [15,16] on the probability of a subspace swap. Our simulation results indicate that compression brings a cost of about $10 \log_{10} C \; dB$ in threshold SNR, where C is the compression ratio.

3.2 Measurement Model

In the following subsections, we consider two models for the random measurement vector $\mathbf{y} \in \mathbb{C}^n$. In the first-order model, the parameters to be estimated nonlinearly modulate the mean of a complex multivariate normal vector, and in the second-order model the parameters nonlinearly modulate the covariance of a multivariate normal vector.

3.2.1 Parameterized Mean Case

Let $\mathbf{y} \in \mathbb{C}^n$ be a complex measurement vector in a signal plus noise model $\mathbf{y} = \mathbf{x}(\boldsymbol{\theta}) + \mathbf{n}$. Here, we assume that $\mathbf{n}$ is a proper complex white Gaussian noise with covariance $\sigma^2\mathbf{I}$ and $\mathbf{x}(\boldsymbol{\theta})$ is parameterized by $\boldsymbol{\theta} \in \mathbb{C}^p$, $p \leq n$. We assume that the parameters are nonlinearly embedded in $\mathbf{x}(\boldsymbol{\theta})$ as $\mathbf{x}(\boldsymbol{\theta}) = \mathbf{K}(\boldsymbol{\theta})\boldsymbol{\alpha}$, where the columns of $\mathbf{K}(\boldsymbol{\theta}) = [\mathbf{k}(\boldsymbol{\theta}_1)\ \mathbf{k}(\boldsymbol{\theta}_2)\ \cdots\ \mathbf{k}(\boldsymbol{\theta}_p)]$ define the signal subspace, and $\boldsymbol{\alpha} \in \mathbb{C}^p$ is a deterministic vector associated with the mode weights. Therefore, $\mathbf{y}$ is distributed as $\mathcal{CN}_n(\mathbf{K}(\boldsymbol{\theta})\boldsymbol{\alpha}, \sigma^2\mathbf{I})$, and the parameters $\boldsymbol{\theta} \in \mathbb{C}^p$ to be estimated nonlinearly modulate the mean of a complex multivariate normal vector. Assume we compress the measurement vector $\mathbf{y}$ by a unitary compression matrix $\boldsymbol{\Psi} = (\boldsymbol{\Phi}\boldsymbol{\Phi}^H)^{-1/2}\boldsymbol{\Phi}$, where $\boldsymbol{\Phi} \in \mathbb{C}^{m\times n}$, $p \leq m < n$. Then, we obtain $\mathbf{w} = \boldsymbol{\Psi}\mathbf{y}$ which is distributed as $\mathcal{CN}_m(\mathbf{z}(\boldsymbol{\theta}), \sigma^2\mathbf{I})$, where $\mathbf{z}(\boldsymbol{\theta}) = \boldsymbol{\Psi}\mathbf{x}(\boldsymbol{\theta})$. We form the data matrix $\mathbf{W} = [\mathbf{w}_1\ \mathbf{w}_2\ \cdots\ \mathbf{w}_M]$, where $\mathbf{w}_i$'s are independent realizations of $\mathbf{w}$. To specify a basis for the signal subspace and the orthogonal subspace in our problem, we define $\mathbf{H}(\boldsymbol{\theta}) = \boldsymbol{\Psi}\mathbf{K}(\boldsymbol{\theta}) = [\mathbf{h}(\boldsymbol{\theta}_1)\ \mathbf{h}(\boldsymbol{\theta}_2)\ \cdots\ \mathbf{h}(\boldsymbol{\theta}_p)]$, with $\mathbf{h}(\boldsymbol{\theta}_i) = \boldsymbol{\Psi}\mathbf{k}(\boldsymbol{\theta}_i)$. The singular value decomposition of $\mathbf{H}_{m\times p}$ $(p \leq m)$ is

$$\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H \tag{3.1}$$

where

$$\mathbf{U} \in \mathbb{C}^{m\times m} : \mathbf{U}\mathbf{U}^H = \mathbf{U}^H\mathbf{U} = \mathbf{I}$$

$$\mathbf{V} \in \mathbb{C}^{p\times p} : \mathbf{V}\mathbf{V}^H = \mathbf{V}^H\mathbf{V} = \mathbf{I}$$

$$\mathbf{\Sigma} \in \mathbb{C}^{m\times p} : \mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_p \\ \mathbf{0} \end{bmatrix}$$

$$\mathbf{\Sigma}_p = \mathrm{diag}(\sigma_1, \sigma_2, ..., \sigma_p),\ \sigma_1 \geq \sigma_2 \geq ... \geq \sigma_p. \tag{3.2}$$

Now we can define the basis vectors from $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_p | \mathbf{u}_{p+1}, ..., \mathbf{u}_m] = [\mathbf{U}_p | \mathbf{U}_0]$, where $\langle \mathbf{U}_p \rangle$ and $\langle \mathbf{U}_0 \rangle$ represent signal and orthogonal subspaces, respectively. The columns of $\mathbf{U}_p$ and $\mathbf{U}_0$ can be considered as basis vectors for the signal and orthogonal subspaces, respectively.

3.2.2 Parameterized Covariance Case

Assume in the signal plus noise model $\mathbf{y} = \mathbf{x} + \mathbf{n}$, the signal component $\mathbf{x}$ is of the form $\mathbf{x} = \mathbf{K}(\boldsymbol{\theta})\boldsymbol{\alpha}$, where the columns of $\mathbf{K}(\boldsymbol{\theta}) = [\mathbf{k}(\boldsymbol{\theta}_1)\ \mathbf{k}(\boldsymbol{\theta}_2)\ \cdots\ \mathbf{k}(\boldsymbol{\theta}_p)]$ are the modes and $\boldsymbol{\alpha} \in \mathbb{C}^p$ is the vector associated with the random mode weights. We assume $\boldsymbol{\alpha}$ is distributed as $\mathcal{CN}_p(0, \mathbf{R}_{\boldsymbol{\alpha}\boldsymbol{\alpha}})$. Therefore, $\mathbf{R}_{xx}(\boldsymbol{\theta}) = \mathbf{K}(\boldsymbol{\theta})\mathbf{R}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}\mathbf{K}^H(\boldsymbol{\theta})$ is parameterized by $\boldsymbol{\theta} \in \mathbb{C}^p$. We assume $\mathbf{n}$ is a proper complex white Gaussian noise with covariance $\sigma^2\mathbf{I}$, and $\mathbf{x}$ and $\mathbf{n}$ are independent. Therefore, $\mathbf{y}$ is distributed as $\mathcal{CN}_n(\mathbf{0}, \mathbf{R}_{\mathbf{yy}}(\boldsymbol{\theta}))$, where $\mathbf{R}_{\mathbf{yy}}(\boldsymbol{\theta}) = \mathbf{K}(\boldsymbol{\theta})\mathbf{R}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}\mathbf{K}^H(\boldsymbol{\theta}) + \sigma^2\mathbf{I}$. Such a data model arises in many applications such as direction of arrival and spectrum estimation.

Assume we compress the measurement vector $\mathbf{y}$ by a unitary compression matrix $\mathbf{\Psi} = (\mathbf{\Phi}\mathbf{\Phi}^H)^{-1/2}\mathbf{\Phi}$, where $\mathbf{\Phi} \in \mathbb{C}^{m\times n}(m < n)$. Then, we obtain $\mathbf{w} = \mathbf{\Psi}\mathbf{y}$ which is distributed as

$$\mathbf{w} \sim \mathcal{CN}_m(\mathbf{0}, \mathbf{R}_{\mathbf{ww}}) \tag{3.3}$$

where $\mathbf{R}_{\mathbf{ww}} = \mathbf{\Psi}\mathbf{K}(\boldsymbol{\theta})\mathbf{R}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}\mathbf{K}^H(\boldsymbol{\theta})\mathbf{\Psi}^H + \sigma^2\mathbf{I}$. We form the data matrix $\mathbf{W} = [\mathbf{w}_1\ \mathbf{w}_2\ \cdots\ \mathbf{w}_M]$, where $\mathbf{w}_i$'s are independent realizations of $\mathbf{w}$. Each of these i.i.d. realizations consists of an i.i.d. realization of $\mathbf{y}_i$, compressed by a common compressor $\mathbf{\Psi}$ for

all $i = 1, 2, \ldots, M$. We may define the signal covariance matrix after compression as

$$\begin{aligned}\mathbf{R_{zz}} &= \mathbf{\Psi K}(\boldsymbol{\theta})\mathbf{R}_{\boldsymbol{\alpha\alpha}}\mathbf{K}^H(\boldsymbol{\theta})\mathbf{\Psi}^H \\ &= \mathbf{H}(\boldsymbol{\theta})\mathbf{R}_{\boldsymbol{\alpha\alpha}}\mathbf{H}^H(\boldsymbol{\theta}),\end{aligned} \tag{3.4}$$

where $\mathbf{H}(\boldsymbol{\theta}) = [\mathbf{h}(\boldsymbol{\theta}_1)\ \mathbf{h}(\boldsymbol{\theta}_2)\ \cdots\ \mathbf{h}(\boldsymbol{\theta}_p)]$, and $\mathbf{h}(\boldsymbol{\theta}_i) = \mathbf{\Psi k}(\boldsymbol{\theta}_i)$. Now, we can write the singular value decomposition of $\mathbf{R_{zz}}$ and $\mathbf{R_{ww}}$ as

$$\begin{aligned}\mathbf{R_{zz}} &= \mathbf{U\Lambda U}^H \\ \mathbf{R_{ww}} &= \mathbf{U}(\mathbf{\Lambda} + \sigma^2\mathbf{I})\mathbf{U}^H\end{aligned} \tag{3.5}$$

where $\mathbf{U}$ and $\mathbf{\Lambda}$ are defined as

$$\begin{aligned}&\mathbf{U} \in \mathbb{C}^{m\times m} : \mathbf{UU}^H = \mathbf{U}^H\mathbf{U} = \mathbf{I} \\ &\mathbf{\Lambda} \in \mathbb{C}^{m\times m} : \mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \\ &\mathbf{\Lambda}_p = \mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_p),\ \lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p.\end{aligned} \tag{3.6}$$

Assuming $\mathbf{R_{zz}}$ has rank p, the unitary matrix $\mathbf{U}$ can be written as $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_p|\mathbf{u}_{p+1}, ..., \mathbf{u}_m] = [\mathbf{U}_p|\mathbf{U}_0]$. Here $\langle\mathbf{U}_p\rangle$ represents the signal subspace and $\langle\mathbf{U}_0\rangle$ represents the orthogonal subspace which completes $\mathbb{C}^{m\times m}$, assuming $p \leq m < n$. Figure 3.1 gives a geometrical representation of (3.6).

3.3 Bound on the Probability of a Subspace Swap after Compression

To bound the probability of a subspace swap for the compressed measurements $\mathbf{W}$, we define the following events:

- E is the event that one or more modes of the orthogonal subspace resolve more energy in $\mathbf{W}$ than one or more modes of the noise-free signal subspace. Therefore, E may be written as

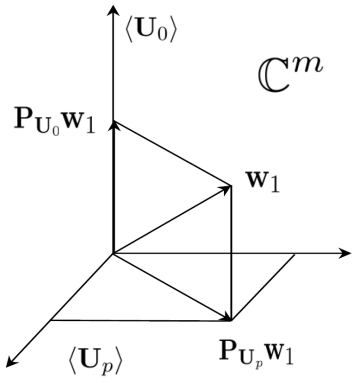


Figure 3.1: Signal and noise subspaces.

$$E = \cup_{q=1}^{p} E(q), \tag{3.7}$$

where $E(q)$ is the following subset of the subspace swap event E,

$$\min_{\mathbf{A} \in \mathcal{I}_{p,q}} \mathrm{tr}(\mathbf{W}^H \mathbf{P}_{\mathbf{HA}} \mathbf{W}) < \max_{\mathbf{B} \in \mathbb{C}^{(n-p)\times q}} \mathrm{tr}(\mathbf{W}^H \mathbf{P}_{\mathbf{U}_0\mathbf{B}} \mathbf{W}), \tag{3.8}$$

and $\mathcal{I}_{p,q}$ is the set of all $p \times q$ slices of the identity matrix $\mathbf{I}_p$. Here, the columns of $\mathbf{H}$ are the modes defined in Section 3.2, and $\mathbf{A}$ selects q of the columns of $\mathbf{H}$.

- F is the event that the average energy resolved in the orthogonal subspace $\langle \mathbf{U}_0 \rangle$ is greater than the average energy resolved in the noise-free signal subspace $\langle \mathbf{U}_p \rangle$ (or equivalently $\langle \mathbf{H} \rangle$). Then, the following bounds establish that F is a subset of $E(1)$, which is in turn a subset of the swap event E:

$$\begin{aligned}
\min_{1 \le i \le p} \mathrm{tr}(\mathbf{W}^H \mathbf{P}_{\mathbf{h}_i} \mathbf{W}) &\le \frac{1}{p} \mathrm{tr}(\mathbf{W}^H \mathbf{P}_{\mathbf{U}_p} \mathbf{W}) \\
&< \frac{1}{m-p} \mathrm{tr}(\mathbf{W}^H \mathbf{P}_{\mathbf{U}_0} \mathbf{w}) \\
&\le \max_{p+1 \le i \le m} \mathrm{tr}(\mathbf{W}^H \mathbf{P}_{\mathbf{u}_i} \mathbf{W}) \\
&\le \max_{\mathbf{b} \in \mathbb{C}^{(n-p)\times 1}} \mathrm{tr}(\mathbf{W}^H \mathbf{P}_{\mathbf{U}_0\mathbf{b}} \mathbf{W}).
\end{aligned} \tag{3.9}$$

- G is the event that the energy resolved in the apriori minimum mode $\mathbf{h}_{min}$ of the noise-free signal subspace $\langle \mathbf{H} \rangle$ (or equivalently $\langle \mathbf{U}_p \rangle$) is smaller than the average energy resolved in the orthogonal subspace $\langle \mathbf{U}_0 \rangle$. For the parameterized mean measurement model, we define $\mathbf{h}_{min}$ as

$$\mathbf{h}_{min} = \underset{\mathbf{h} \in \{\mathbf{h}(\boldsymbol{\theta}_1), \mathbf{h}(\boldsymbol{\theta}_2), \ldots, \mathbf{h}(\boldsymbol{\theta}_p)\}}{\arg\min} |\mathbf{h}^H \mathbf{z}(\boldsymbol{\theta})|^2, \tag{3.10}$$

and for the parameterized covariance measurement model as

$$\mathbf{h}_{min} = \underset{\mathbf{h} \in \{\mathbf{h}(\boldsymbol{\theta}_1), \mathbf{h}(\boldsymbol{\theta}_2), \ldots, \mathbf{h}(\boldsymbol{\theta}_p)\}}{\arg\min} |\mathbf{h}^H \mathbf{R}_{\mathbf{zz}}(\boldsymbol{\theta}) \mathbf{h}|^2. \tag{3.11}$$

Then, the following bounds establish that G is a subset of $E(1)$, which is in turn a subset of the swap event E:

$$\begin{aligned} \min_{1 \le i \le p} \mathrm{tr}(\mathbf{W}^H \mathbf{P}_{\mathbf{h}_i} \mathbf{W}) &\le \mathrm{tr}(\mathbf{W}^H \mathbf{P}_{\mathbf{h}_{min}} \mathbf{W}) \\ &< \frac{1}{m-p} \mathrm{tr}(\mathbf{W}^H \mathbf{P}_{\mathbf{U}_0} \mathbf{w}) \\ &\le \max_{p+1 \le i \le m} \mathrm{tr}(\mathbf{W}^H \mathbf{P}_{\mathbf{u}_i} \mathbf{W}) \\ &\le \max_{\mathbf{b} \in \mathbb{C}^{(n-p) \times 1}} \mathrm{tr}(\mathbf{W}^H \mathbf{P}_{\mathbf{U}_0 \mathbf{b}} \mathbf{W}). \end{aligned} \tag{3.12}$$

Since events F and G are subsets of event E, their probabilities of occurrence give lower bounds on the probability of a subspace swap, $P_{ss} \triangleq P(E)$. We use these events to derive lower bounds on the probability of a subspace swap for the two data models given in Section 3.2.

3.3.1 Parameterized Mean Case

For the parameterized mean measurement model discussed in Section 3.2.1, we start with event F and define

$$\mathbf{T}_F = \frac{1}{m-p} \mathbf{P}_{\mathbf{U}_0} - \frac{1}{p} \mathbf{P}_{\mathbf{U}_p} \tag{3.13}$$

where $\mathbf{P}_{\mathbf{U}_p} = \mathbf{U}_p\mathbf{U}_p^H$ is the orthogonal projection onto the signal subspace and $\mathbf{P}_{\mathbf{U}_0} = \mathbf{U}_0\mathbf{U}_0^H$ is the orthogonal projection onto the orthogonal (noise) subspace. According to the definition of event F we can lower bound the probability of a subspace swap P_{ss} as

$$P_{ss} \geq P(\mathrm{tr}\big[\mathbf{W}^H\mathbf{T}_F\mathbf{W}\big] > 0) \tag{3.14}$$

Therefore, we have

$$\begin{aligned} P_{ss} &\geq P(\mathrm{tr}\big[\mathbf{W}^H\mathbf{T}_F\mathbf{W}\big] > 0) \\ &= P(\frac{\mathrm{tr}\big[\mathbf{W}^H\mathbf{U}_p\mathbf{U}_p^H\mathbf{W}\big]/2p}{\mathrm{tr}\big[\mathbf{W}^H\mathbf{U}_0\mathbf{U}_0^H\mathbf{W}\big]/2(m-p)} < 1) \\ &= P(\frac{\sum_{i=1}^{M}\|\mathbf{U}_p^H\mathbf{w}_i\|_2^2/2p}{\sum_{i=1}^{M}\|\mathbf{U}_0^H\mathbf{w}_i\|_2^2/2(m-p)} < 1). \end{aligned} \tag{3.15}$$

Here, the $\mathbf{U}_p^H\mathbf{w}_i$ are independent and identically distributed as

$$\mathbf{U}_p^H\mathbf{w}_i \sim \mathcal{CN}_p(\mathbf{U}_p^H\mathbf{z}(\boldsymbol{\theta}), \sigma^2\mathbf{I}) \;\; \forall 1 \leq i \leq M. \tag{3.16}$$

Therefore, $\|\mathbf{U}_p^H\mathbf{w}_i\|_2^2/\sigma^2 \sim \chi^2_{2p}(\|\mathbf{z}(\boldsymbol{\theta})\|_2^2/\sigma^2)$, which is the distribution of a complex noncentral chi-squared random variable with $2p$ degrees of freedom and noncentrality parameter $\|\mathbf{z}(\boldsymbol{\theta})\|_2^2/\sigma^2$. Also, since $\langle\mathbf{U}_p\rangle$ and $\langle\mathbf{U}_0\rangle$ are orthogonal, we can conclude that in (3.15), each $\|\mathbf{U}_0^H\mathbf{w}_i\|_2^2/\sigma^2$ is independent of $\|\mathbf{U}_p^H\mathbf{w}_i\|_2^2/\sigma^2$ and is distributed as $\chi^2_{2(m-p)}$. Hence, the term $\frac{\|\mathbf{U}_p^H\mathbf{w}\|_2^2/2p}{\|\mathbf{U}_0^H\mathbf{w}\|_2^2/2(m-p)}$ is the ratio of two independent normalized chi-squared random variables and is distributed as $F_{2pM,2(m-p)M}(\|\mathbf{z}(\boldsymbol{\theta})\|_2^2/\sigma^2)$, which is a noncentral F distribution with $2pM$ and $2(m-p)M$ degrees of freedom and noncentrality parameter $\|\mathbf{z}(\boldsymbol{\theta})\|_2^2/\sigma^2$. Thus, the probability of a subspace swap after compression is lower bounded by the probability that a $F_{2pM,2(m-p)M}(\|\mathbf{z}(\boldsymbol{\theta})\|_2^2/\sigma^2)$ distributed random variable is less than 1. When there is no compression, this lower bound turns into the probability that a $F_{2pM,2(n-p)M}(\|\mathbf{x}(\boldsymbol{\theta})\|_2^2/\sigma^2)$ random variable is less than 1.

For event G, we define

$$\mathbf{T}_G = \frac{1}{m-p}\mathbf{P}_{\mathbf{U}_0} - \mathbf{P}_{\mathbf{h}_{min}}. \tag{3.17}$$

Here, we define $\boldsymbol{\rho}_{min} = \frac{\mathbf{h}_{min}}{\|\mathbf{h}_{min}\|_2}$. Therefore $\mathbf{P}_{\mathbf{h}_{min}} = \boldsymbol{\rho}_{min}\boldsymbol{\rho}_{min}^H$, and we have

$$\begin{aligned} P_{ss} &\geq P(\mathrm{tr}\big[\mathbf{W}^H\mathbf{T}_G\mathbf{W}\big] > 0) \\ &= P(\frac{\mathrm{tr}\big[\mathbf{W}^H\boldsymbol{\rho}_{min}\boldsymbol{\rho}_{min}^H\mathbf{W}\big]/2}{\mathrm{tr}\big[\mathbf{W}^H\mathbf{U}_0\mathbf{U}_0^H\mathbf{W}\big]/2(m-p)} < 1) \\ &= P(\frac{\sum_{i=1}^M \|\boldsymbol{\rho}_{min}^H\mathbf{w}_i\|_2^2/2}{\sum_{i=1}^M \|\mathbf{U}_0^H\mathbf{w}_i\|_2^2/2(m-p)} < 1). \end{aligned} \tag{3.18}$$

Here, we have

$$\boldsymbol{\rho}_{min}^H\mathbf{w}_i \sim \mathcal{CN}(\boldsymbol{\rho}_{min}^H\mathbf{z}(\boldsymbol{\theta}), \sigma^2\mathbf{I}) \;\; \forall 1 \leq i \leq M. \tag{3.19}$$

Therefore, $\|\boldsymbol{\rho}_{min}^H\mathbf{w}_i\|_2^2/\sigma^2 \sim \chi_2^2(|\boldsymbol{\rho}_{min}^H\mathbf{z}(\boldsymbol{\theta})|^2/\sigma^2)$ which is the distribution of a complex noncentral chi-squared random variable with 2 degrees of freedom and noncentrality parameter $|\boldsymbol{\rho}_{min}^H\mathbf{z}(\boldsymbol{\theta})|^2/\sigma^2$. Thus, with the same type of arguments as for event F, we can conclude that the term $\frac{\sum_{i=1}^M \|\boldsymbol{\rho}_{min}^H\mathbf{w}_i\|_2^2/2}{\sum_{i=1}^M \|\mathbf{U}_0^H\mathbf{w}_i\|_2^2/2(m-p)}$ is distributed as $F_{2M,2(m-p)M}(|\boldsymbol{\rho}_{min}^H\mathbf{z}(\boldsymbol{\theta})|^2/\sigma^2)$, which is a noncentral F distribution with $2M$ and $2(m-p)M$ degrees of freedom and noncentrality parameter $|\boldsymbol{\rho}_{min}^H\mathbf{z}(\boldsymbol{\theta})|^2/\sigma^2$. When there is no compression, this turns into the probability that a $F_{2M,2(n-p)M}(|\boldsymbol{\kappa}_{min}^H\mathbf{x}(\boldsymbol{\theta})|^2/\sigma^2)$ random variable is less than 1. Here, $\boldsymbol{\kappa}_{min} = \frac{\mathbf{k}_{min}}{\|\mathbf{k}_{min}\|_2}$, and $\mathbf{k}_{min}$ is the apriori minimum mode of the signal subspace before compression.

3.3.2 Parameterized Covariance Case

For the parameterized covariance measurement model discussed in Section 3.2.2, we start with event F. In this case, the columns of the measurement matrix $\mathbf{W}$ are i.i.d. random vectors distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{R}_{\mathbf{ww}})$, and similar to the mean case we have

$$\begin{aligned} P_{ss} &\geq P(\mathrm{tr}\big[\mathbf{W}^H\mathbf{T}_F\mathbf{W}\big] > 0) \\ &= P(\frac{\mathrm{tr}\big[\mathbf{W}^H\mathbf{U}_p\mathbf{U}_p^H\mathbf{W}\big]/2p}{\mathrm{tr}\big[\mathbf{W}^H\mathbf{U}_0\mathbf{U}_0^H\mathbf{W}\big]/2(m-p)} < 1) \\ &= P(\frac{\sum_{i=1}^M \|\mathbf{U}_p^H\mathbf{w}_i\|_2^2/2p}{\sum_{i=1}^M \|\mathbf{U}_0^H\mathbf{w}_i\|_2^2/2(m-p)} < 1). \end{aligned} \tag{3.20}$$

Here, the $\mathbf{U}_p^H\mathbf{w}_i$ are i.i.d. and distributed as

$$\mathbf{U}_p^H\mathbf{w}_i \sim \mathcal{CN}_p(\mathbf{0}, \boldsymbol{\Lambda}_p + \sigma^2\mathbf{I}_p) \;\; \forall 1 \leq i \leq M. \tag{3.21}$$

Therefore we can write

$$\|\mathbf{U}_p^H \mathbf{w}_i\|_2^2 = \sum_{i=1}^{p} (\lambda_i + \sigma^2)\rho_i, \tag{3.22}$$

where ρ_i's are i.i.d. random variables, each distributed as χ_2^2. Therefore,

$$\sum_{i=1}^{M} \|\mathbf{U}_p^H \mathbf{w}_i\|_2^2 = \sum_{i=1}^{p} (\lambda_i + \sigma^2)\xi_i, \tag{3.23}$$

where ξ_i's are i.i.d. random variables, each distributed as χ_{2M}^2. Also, we can write $\sum_{i=1}^{M} \|\mathbf{U}_0^H \mathbf{w}_i\|_2^2 = \sigma^2 \nu$, where ν is distributed as $\chi_{2M(m-p)}^2$ and is independent of the ξ_i's. Therefore, we have

$$\begin{aligned} P_{ss} &\geq P\left(\frac{\sum_{i=1}^{M} \|\mathbf{U}_p^H \mathbf{w}_i\|_2^2 / 2p}{\sum_{i=1}^{M} \|\mathbf{U}_0^H \mathbf{w}_i\|_2^2 / 2(m-p)} < 1\right) \\ &= P\left(\frac{\sum_{i=1}^{p} (1 + \lambda_i/\sigma^2)\xi_i / 2Mp}{\nu / 2M(m-p)} < 1\right). \end{aligned} \tag{3.24}$$

Here, the term $\frac{\sum_{i=1}^{p}(1+\lambda_i/\sigma^2)\xi_i/2Mp}{\nu/2M(m-p)}$ is distributed as $GF\big[(1+\frac{\lambda_1}{\sigma^2}), \ldots, (1+\frac{\lambda_p}{\sigma^2}); 2Mp; 2M(m-p)\big]$, which is the distribution of a generalized F random variable [48]. Thus, the probability of a subspace swap in this case is lower bounded by the probability that a $GF\big[(1+\frac{\lambda_1}{\sigma^2}), \ldots, (1+\frac{\lambda_p}{\sigma^2}); 2M; 2M(m-p)\big]$ random variable is less than 1. Without compression, this turns into the probability that a $GF\big[(1+\frac{\tilde{\lambda}_1}{\sigma^2}), \ldots, (1+\frac{\tilde{\lambda_p}}{\sigma^2}); 2Mp; 2M(m-p)\big]$ random variable is less than 1. Here $\tilde{\lambda}_i$'s are the eigenvalues of the signal covariance matrix $\mathbf{R_{xx}}$ before compression.

We can also derive the probability of the event G for the parameterized covariance measurement model. In this case we have

$$\begin{aligned} P_{ss} &\geq P(\mathrm{tr}\big[\mathbf{W}^H \mathbf{T}_G \mathbf{W}\big] > 0) \\ &= P\left(\frac{\sum_{i=1}^{M} \|\boldsymbol{\rho}_{min}^H \mathbf{w}_i\|_2^2 / 2}{\sum_{i=1}^{M} \|\mathbf{U}_0^H \mathbf{w}_i\|_2^2 / 2(m-p)} < 1\right), \end{aligned} \tag{3.25}$$

where $\boldsymbol{\rho}_{min} = \frac{\mathbf{h}_{min}}{\|\mathbf{h}_{min}\|_2}$, and $\mathbf{h}_{min}$ is the apriori minimum mode of the signal subspace given by (3.11). Here, $\boldsymbol{\rho}_{min}^H \mathbf{w}_i$ is distributed as

$$\boldsymbol{\rho}_{min}^H \mathbf{w}_i \sim \mathcal{CN}(0, \tau) \quad \forall 1 \leq i \leq M, \tag{3.26}$$

where $\tau = \boldsymbol{\rho}_{min}^H \mathbf{R}_{\mathbf{ww}} \boldsymbol{\rho}_{min}$. Therefore,

$$\sum_{i=1}^{M} \|\boldsymbol{\rho}_{min}^H \mathbf{w}_i\|_2^2 / \tau \sim \chi_{2M}^2, \tag{3.27}$$

and we have

$$\begin{aligned} P_{ss} &\geq P\left(\frac{\sum_{i=1}^{M} \|\boldsymbol{\rho}_{min}^H \mathbf{w}_i\|_2^2/2}{\sum_{i=1}^{M} \|\mathbf{U}_0^H \mathbf{w}_i\|_2^2/2(m-p)} < 1\right) \\ &= P(\vartheta < \frac{\sigma^2}{\tau}), \end{aligned} \tag{3.28}$$

where ϑ is distributed as $F_{2M,2M(m-p)}$, which is a central F random variable with $2M$ and $2M(m-p)$ degrees of freedom. Without compression, this turns into the probability that a $F_{2M,2M(n-p)}$ random variable is less than $\frac{\sigma^2}{\tilde{\tau}}$, where $\tilde{\tau} = \boldsymbol{\kappa}_{min}^H \mathbf{R}_{\mathbf{yy}} \boldsymbol{\kappa}_{min}$, $\boldsymbol{\kappa}_{min} = \frac{\mathbf{k}_{min}}{\|\mathbf{k}_{min}\|_2}$, and $\mathbf{k}_{min}$ is the apriori minimum mode of the signal subspace before compression.

Remark 1: In Sections 3.3.1 and 3.3.2, we have derived lower bounds on the probability of a subspace swap for the case that $\boldsymbol{\Psi} = (\boldsymbol{\Phi}\boldsymbol{\Phi}^H)^{-1/2}\boldsymbol{\Phi}$ is deterministic, as in standard or co-prime subsamplings. In the case that $\boldsymbol{\Psi}$ is random, these probability bounds would have to be integrated over the distribution of $\boldsymbol{\Psi}$ to give lower bounds on marginal probabilities of a subspace swap. For example, for random $\boldsymbol{\Psi}$ and for the subevent F we have

$$P_{ss} = \int P(E|\boldsymbol{\Psi}) P(\boldsymbol{\Psi}) d\boldsymbol{\Psi} \geq \int P(F|\boldsymbol{\Psi}) P(\boldsymbol{\Psi}) d\boldsymbol{\Psi} \tag{3.29}$$

where $P(F|\boldsymbol{\Psi})$ is given in Sections 3.3.1 and 3.3.2 for the parameterized mean and parameterized covariance measurement models, respectively. For the class of random compression matrices that have density functions of the form $g(\boldsymbol{\Phi}\boldsymbol{\Phi}^H)$, that is, the distribution of $\boldsymbol{\Phi}$ is right orthogonally invariant, $\boldsymbol{\Psi}$ is uniformly distributed on the Stiefel manifold $\mathcal{V}_m(\mathbb{C}^n)$ [46]. The compression matrix $\boldsymbol{\Phi}$ whose elements are i.i.d. standard normal random variables is one such matrix.

3.4 Simulation Results

In this Section, we present numerical examples to show the impact of compression on threshold effects for estimating directions of arrival using a sensor array. We consider a

dense uniform line array with n elements at half-wavelength inter-element spacings. We compress this array to m dimensions using co-prime subsampling. In co-prime compression, we uniformly subsample the dense array once by a factor m_1 and once by a factor m_2, where m_1 and m_2 are co-prime. We then interleave these two subarrays to form the co-prime array of $m_1 + 2m_2 - 1$ elements. We note that although we are compressing the array by a factor n/m for the co-prime array, the dense and the compressed arrays still have the same total aperture. The geometry of the dense and co-prime arrays are shown in Figure 3.2. We consider two point sources at far field at electrical angles $\theta_1 = 0$ and $\theta_2 = \pi/n$. We set the amplitudes of these sources $\alpha_1 = \alpha_2 = 1$. The Rayleigh limit of the dense array in electrical angle is $2\pi/n$. Therefore, in our examples the two sources are separated by half the Rayleigh limit of the dense array. We present the results for the parameterized mean and parameterized covariance cases.

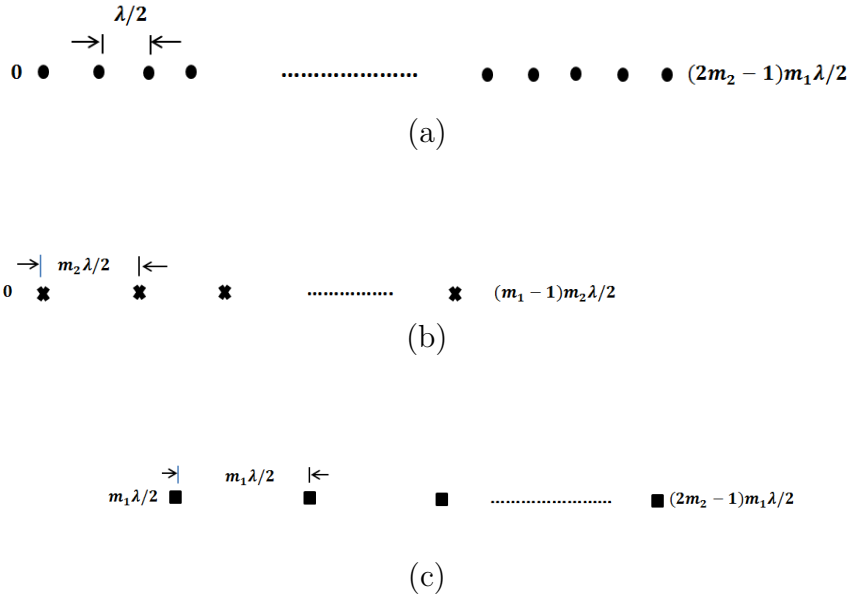


Figure 3.2: Geometry of the dense array (a), and co-prime subarrays (b), (c). At $m_1 = 11$ and $m_2 = 9$, $(2m_2 - 1)m_1\lambda/2 = 187\lambda/2$.

3.4.1 Parameterized Mean Case

In this case, the Monte Carlo simulation consists of 200 independent realizations of the measurement vector $\mathbf{y}$ for a dense array of 188 elements, each for a single snapshot ($M = 1$). Then we compress these measurements to simulate the data for the co-prime compressed array of 28 elements with $m_1 = 11$ and $m_2 = 9$. The compression ratio is $\frac{n}{m} \approx 6.7$. Figure 3.3 shows the MSE for the maximum likelihood estimator of the source at θ_1 in the presence of the interfering source at θ_2. The CRB corresponding to the 188-element dense array is also shown in this figure as a reference for performance analysis. Figure 3.3 also shows approximations to the MSE (in starred solid lines) obtained using the *method of intervals* (introduced in [49] and used in [18]). At each SNR, the approximate MSE σ_T^2 is computed as

$$\sigma_T^2 = P_{ss}\sigma_0^2 + (1 - P_{ss})\sigma_{CR}^2. \tag{3.30}$$

Here, P_{ss} is the probability of the subspace swap as a function of SNR, which we approximate using the lower bound in (3.14); σ_{CR}^2 is the value of the CRB as a function of SNR, and σ_0^2 is the variance of the error given the occurrence of a subspace swap. The justification for using this formula is that when a subspace swap does not occur, MSE almost follows the CRB . However, given the occurrence of the subspace swap (and in the absence of any prior knowledge) the error in estimating the electrical angle θ_1 may be taken to be uniformly distributed between $(-\pi/2, \pi/2)$ and the error variance is $\sigma_0^2 = \pi^2/12$.

Figure 3.3 shows that performance loss, measured by onset of threshold effect is approximately $10\log_{10} n/m$. Our approximations on MSE also predict the same SNR difference in the onset of the performance breakdown. Figure 3.4 shows our bounds on the probability of a subspace swap for the dense and co-prime arrays which are obtained using event F in Section 3.3. The ML curves of Figure 3.3 would approach the CRB at high SNR were it not for the quantization of our ML simulation code.

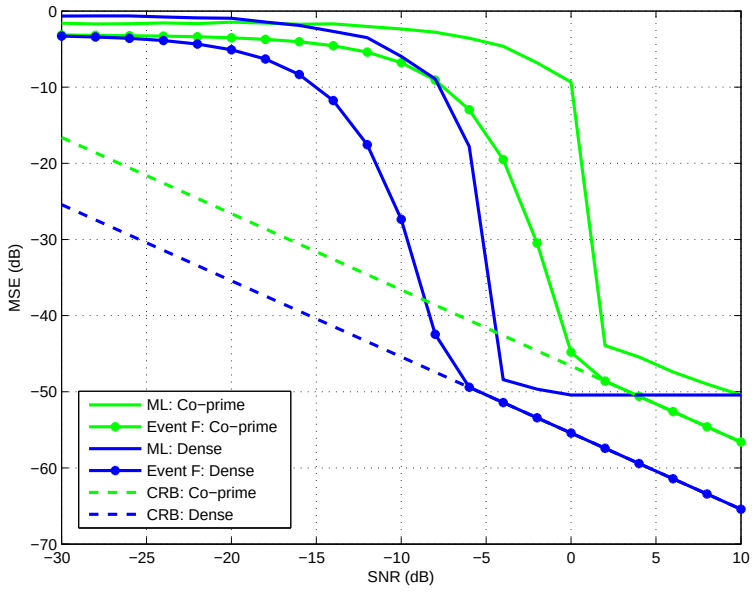


Figure 3.3: Parameterized mean case. Dense 188 element array and 28 element co-prime array. MSE bounds and MSE for ML estimation of $\theta_1 = 0$ in the presence of an interfering source at $\theta_2 = \pi/188$; 200 trials.

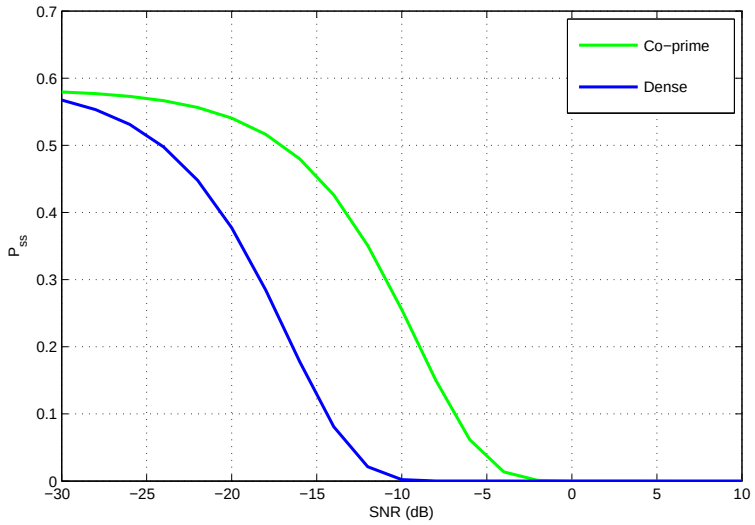


Figure 3.4: Parameterized mean case. Analytical lower bounds (event F) for the probability of subspace swap for estimation of the angle of a source at $\theta_1 = 0$ in the presence of an interfering source at $\theta_2 = \pi/188$ using 188 element dense array and 28 element co-prime array.

3.4.2 Parameterized Covariance Case

We conduct the same set of Monte Carlo simulations for the stochastic data model. Here we draw $M = 200$ independent snapshots for a dense array of 36 elements over 200 independent realizations, and compress them to simulate the data for the co-prime array of 12 elements with $m_1 = 5$ and $m_2 = 4$. The compression ratio is $\frac{n}{m} = 3$. Figure 3.5 shows the results for the MSE of the maximum likelihood estimator of the source at $\theta_1 = 0$ in the presence of the interfering source at $\theta_2 = \pi/36$. Our approximations for the MSE using the *method of intervals* in (3.30) and the Cramér-Rao bound are also shown for each array. Figure 3.5 shows that performance loss, measured by onset of threshold effect is approximately $10\log_{10} n/m$. Our bounds on the probability of a subspace swap using event G in Section 3.3 are shown in Figure 3.6 for the dense and co-prime arrays.

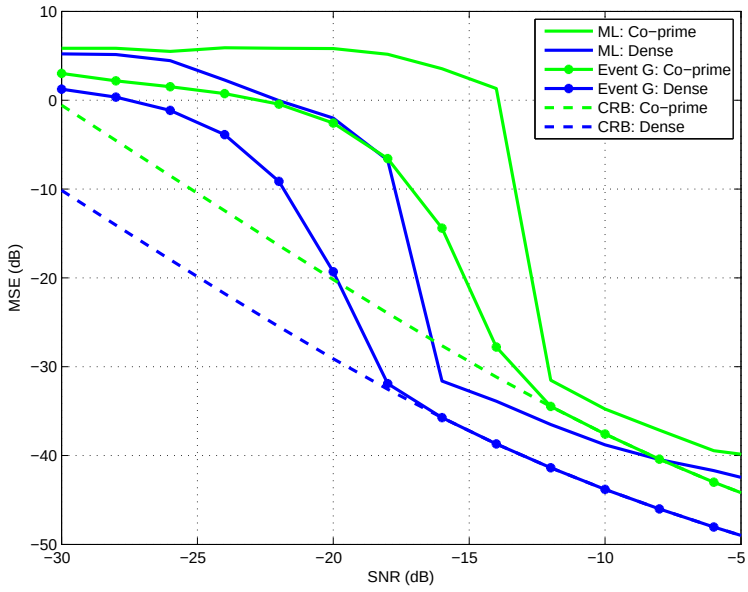


Figure 3.5: Parameterized covariance case. Dense 36 element array and 12 element co-prime array. MSE bounds and MSE for ML estimation of $\theta_1 = 0$ in the presence of an interfering source at $\theta_2 = \pi/36$; 200 snapshots and 200 trials.

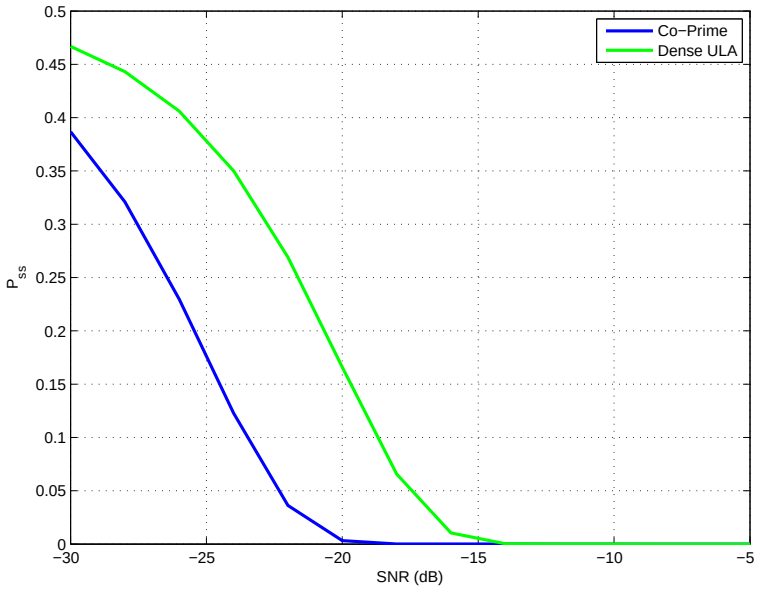


Figure 3.6: Parameterized covariance case. Analytical lower bounds (event G) for the probability of a subspace swap using co-prime compression for the estimation of the angle of a source at $\theta_1 = 0$ in the presence of an interfering source at $\theta_2 = \pi/36$ using 36 element dense array and 12 element co-prime array.

CHAPTER 4

MODAL ANALYSIS USING SPARSE AND CO-PRIME ARRAYS

4.1 Motivation

Non-uniform sensor array geometries, without aliasing ambiguities, have a long history in sensor array processing, dating back to minimum-redundancy arrays [50]. The introduction of co-prime arrays in [15,16] has created renewed interest in such geometries. In this chapter, we consider two specific cases of non-uniform sensor arrays. These are sparse arrays and co-prime arrays. Both of these geometries can be viewed as subsampled (or compressed) versions of a dense uniform line array, whose consecutive elements are separated by a half wavelength in space.[1] Specifically, the sparse array can be thought of as a subsampled version of a dense uniform line array, plus an extra sensor that is positioned at a location on the array that allows us to resolve aliasing ambiguities. The co-prime array consists of two uniform subarrays, each obtained by uniformly subsampling a dense uniform line array with co-prime subsampling factors. The co-prime property allows for resolving aliasing ambiguities.

Naturally, any subsampling in space results in a reduction in signal-to-noise ratio (SNR), by the compression factor, and leads to a loss in estimation performance. Our studies in [51] and [52] address the effect of compression on Fisher information, the Cramer-Rao lower bound, and the probability of a swap between signal and noise subspaces. Assuming that the loss in SNR due to compression has tolerable effect on estimation or detection, or can be compensated by collecting more temporal snapshots (requiring a scene to remain stationary

[1] If we were sampling in time, then the dense sequence of uniform samples would have had spacings equal to the Nyquist interval.

for a longer period), the question is how can methods of linear prediction and approximate least squares be adapted to the estimation of mode parameters in sparse and co-prime arrays? In this chapter, we address this question.

We determine a parameterization of the orthogonal subspace. This is the subspace that is orthogonal to the signal subspace spanned by the columns of a generalized Vandermonde matrix of the modes in sparse and co-prime arrays. This parameterization is of a form that is particularly suitable for utilizing approximate least squares, such as iterative quadratic maximum likelihood (IQML) (see [21] and [22]), for estimating the modes. Although we present our numerical results in the context of sensor array processing, all of our results apply to the estimation of complex exponential modes from time series data. Our numerical results here, and in [51], [52], show that there is a loss in performance sustained by sparse and co-prime arrays that are compressions of uniform linear arrays. A rough rule of thumb is that effective SNR is reduced by 10 $log_{10}C$, where C is the compression ratio. For example, in our experiments a 50-element array is subsampled to a 14-element co-prime array, for a compression ratio of 50/14. The loss in SNR is roughly 5.5 dB.

Remark 1: A small number of other authors have also considered estimating the parameters of complex exponentials from non-uniformly sampled data. In [32], the authors approach the modal estimation problem by fitting a polynomial to the non-uniform samples and estimating the parameters of the exponentials using linear regression. For the case that the modes are on the unit circle, in [33] a truncated window function is fitted to the non-uniform measurements in the least squares sense, and then an approximate Prony method is proposed to estimate the frequencies of the exponentials. These approaches are different from ours and do not involve characterization of orthogonal subspaces for utilizing modern methods of linear prediction.

4.2 Problem Statement

Consider a non-uniform line array of m sensors at locations $\mathbb{I} = \{i_0, i_1, \ldots, i_{m-1}\}$ in units of half wavelength in space. We assume, without loss of generality, that $i_0 = 0$. Suppose the array is observing a weighted superposition of p damped complex exponentials (modes). These modes are determined by the mode parameters $z_k = \rho_k e^{j\theta_k}$, $k = 1, 2, \ldots, p$, where the kth mode has a damping factor ρ_k and an electrical angle $\theta_k \in (-\pi, \pi]$. Suppose the array collects N temporal snapshots. Then, the measurement equation for the lth sensor (located at i_l) can be written as

$$y_l[n] = \sum_{k=1}^{p} x_k[n] z_k^{i_l} + e_l[n], \quad n = 0, 1, \ldots, N-1, \tag{4.1}$$

where n is the snapshot index, $x_k[n]$ denotes the amplitude (or weight) of the kth mode at index n, and $e_l[n]$ is the measurement noise at sensor l. In vector form, we have $\mathbf{y}[n] \in \mathbb{C}^m$,

$$\mathbf{y}[n] = \mathbf{V}(\mathbf{z}, \mathbb{I})\mathbf{x}[n] + \mathbf{e}[n], \quad n = 0, 1, \ldots, N-1, \tag{4.2}$$

where $\mathbf{y}[n] = [y_0[n], y_1[n], \ldots, y_{m-1}[n]]^T$ is the array measurement vector, $\mathbf{x}[n] = [x_1[n], x_2[n], \ldots, x_p[n]]^T$ is the vector of mode amplitudes at index n, $\mathbf{e}[n] = [e_0[n], e_1[n], \ldots, e_{m-1}[n]]^T$ is the noise vector at index n, and $\mathbf{V}(\mathbf{z}, \mathbb{I}) \in \mathbb{C}^{m \times p}$ is a generalized Vandermonde matrix of the modes $\mathbf{z} = [z_1, z_2, \ldots, z_p]^T$, given by

$$\mathbf{V}(\mathbf{z}, \mathbb{I}) = \begin{bmatrix} z_1^{i_0} & z_2^{i_0} & \cdots & z_p^{i_0} \\ z_1^{i_1} & z_2^{i_1} & \cdots & z_p^{i_1} \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{i_{m-1}} & z_2^{i_{m-1}} & \cdots & z_p^{i_{m-1}} \end{bmatrix}. \tag{4.3}$$

We consider the case where $\mathbf{x}[n]$ is free to change with n, and assume that the $\mathbf{e}_l[n]$'s, are i.i.d. complex normal with mean zero and variance σ^2. This means that the measurement vectors $\mathbf{y}[n]$, $n = 0, 1, \ldots, N-1$ are i.i.d proper complex normal with mean $\mathbf{V}(\mathbf{z}, \mathbb{I})\mathbf{x}[n]$ and covariance $\sigma^2\mathbf{I}$. Under this measurement model, the least squares estimation and the maximum likelihood estimation of the modes $\{z_k\}_{k=1}^p$ and mode weights $\{\mathbf{x}[n]\}_{n=0}^{N-1}$ are equivalent

and can be posed as

$$\min_{\mathbf{z},\mathbf{x}[0],\ldots,\mathbf{x}[N-1]} \sum_{n=0}^{N-1} \|\mathbf{y}[n] - \mathbf{V}(\mathbf{z},\mathbb{I})\mathbf{x}[n]\|_2^2. \tag{4.4}$$

The least squares estimate of $\mathbf{x}[n]$ is

$$\hat{\mathbf{x}}[n] = \mathbf{V}^+(\mathbf{z},\mathbb{I})\mathbf{y}[n], \tag{4.5}$$

where $\mathbf{V}^+(\mathbf{z},\mathbb{I}) = (\mathbf{V}^H(\mathbf{z},\mathbb{I})\mathbf{V}(\mathbf{z},\mathbb{I}))^{-1}\mathbf{V}^H(\mathbf{z},\mathbb{I})$ is the Moore-Penrose pseudoinverse of $\mathbf{V}(\mathbf{z},\mathbb{I})$. The least squares estimate of the modes is obtained as

$$\begin{aligned}\hat{\mathbf{z}} &= \arg\min_{\mathbf{z}} \sum_{n=0}^{N-1} \mathbf{y}^H[n](\mathbf{I} - \mathbf{P}_{\mathbf{V}(\mathbf{z},\mathbb{I})})\mathbf{y}[n] \\ &= \arg\min_{\mathbf{z}} \sum_{n=0}^{N-1} \mathbf{y}^H[n]\mathbf{P}_{\mathbf{A}(\mathbf{z},\mathbb{I})}\mathbf{y}[n],\end{aligned} \tag{4.6}$$

where $\mathbf{A}(\mathbf{z},\mathbb{I})$ is a full column rank matrix that satisfies

$$\mathbf{A}^H(\mathbf{z},\mathbb{I})\mathbf{V}(\mathbf{z},\mathbb{I}) = \mathbf{0}_{(m-p)\times p}, \tag{4.7}$$

and $\mathbf{P}_{\mathbf{V}(\mathbf{z},\mathbb{I})}$ and $\mathbf{P}_{\mathbf{A}(\mathbf{z},\mathbb{I})} = \mathbf{I} - \mathbf{P}_{\mathbf{V}(\mathbf{z},\mathbb{I})}$ are the orthogonal projections onto the column spans of $\mathbf{V}(\mathbf{z},\mathbb{I})$ and $\mathbf{A}(\mathbf{z},\mathbb{I})$, respectively. We denote these column spans by the subspaces $\langle\mathbf{V}(\mathbf{z},\mathbb{I})\rangle$ and $\langle\mathbf{A}(\mathbf{z},\mathbb{I})\rangle$. We call $\langle\mathbf{V}(\mathbf{z},\mathbb{I})\rangle$ the *signal subspace* and $\langle\mathbf{A}(\mathbf{z},\mathbb{I})\rangle$ the *orthogonal subspace*. Note that $\langle\mathbf{A}(\mathbf{z},\mathbb{I})\rangle = \langle\mathbf{V}(\mathbf{z},\mathbb{I})\rangle^\perp$. See Figure 4.1.

For a given array geometry, the *basis matrix* $\mathbf{V}(\mathbf{z},\mathbb{I})$ given in (4.3), and the subspace $\langle\mathbf{V}(\mathbf{z},\mathbb{I})\rangle$, are fully characterized by the p modes $\mathbf{z} = [z_1, z_2, \ldots, z_p]^T$. This subspace, parameterized by $\mathbf{z}$, is an element of a Grassmanian manifold of dimension p. Now, let us rewrite $\mathbf{V}(\mathbf{z},\mathbb{I})$, using elementary operations, and with some abuse of notation, as

$$\mathbf{V}(\mathbf{z},\mathbb{I}) = \begin{bmatrix}\mathbf{V}_1(\mathbf{z},\mathbb{I}) \\ \mathbf{V}_2(\mathbf{z},\mathbb{I})\end{bmatrix}, \tag{4.8}$$

where $\mathbf{V}_1(\mathbf{z},\mathbb{I}) \in \mathbb{C}^{p\times p}$ is invertible and $\mathbf{V}_2(\mathbf{z},\mathbb{I}) \in \mathbb{C}^{(m-p)\times p}$. Then the basis matrix $\mathbf{A}(\mathbf{z},\mathbb{I})$ for the orthogonal subspace is the Hermitian transpose of

$$\mathbf{A}^H(\mathbf{z},\mathbb{I}) = [-\mathbf{V}_2(\mathbf{z},\mathbb{I})\mathbf{V}_1^{-1}(\mathbf{z},\mathbb{I}) \mid \mathbf{I}_{m-p}]. \tag{4.9}$$

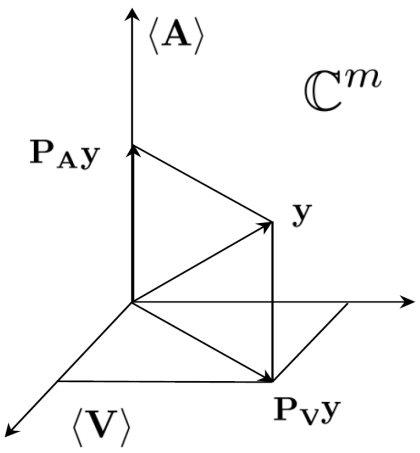


Figure 4.1: The signal subspace $\langle \mathbf{V}(\mathbf{z}, \mathbb{I}) \rangle$ and the orthogonal subspace $\langle \mathbf{A}(\mathbf{z}, \mathbb{I}) \rangle = \langle \mathbf{V}(\mathbf{z}, \mathbb{I}) \rangle^{\perp}$. In the figure, we have dropped $(\mathbf{z}, , \mathbb{I})$ and have simply used $\mathbf{A}$, $\mathbf{V}$, $\langle \mathbf{A} \rangle$, and $\langle \mathbf{V} \rangle$.

Although these p-dimensional characterizations of the signal and orthogonal subspaces have minimum parameterization $\mathbf{z} \in \mathbb{C}^p$, it is not easy to solve the least squares problem (4.6) using these characterizations.

For an m-element uniform line array, a particular p-parameter characterization of $\mathbf{A}(\mathbf{z}, \mathbb{I})$ exists that makes solving (4.6) relatively simple [20]. We will review this characterization in Section 4.3. Then, we derive such suitable parameterizations of $\mathbf{A}(\mathbf{z}, \mathbb{I})$ for two specific non-uniform arrays: sparse and co-prime.

- *Sparse array*: In this case, the location set $\mathbb{I}$ is given by $\mathbb{I}_s = \{0, d, 2d, \ldots, (m-2)d, M\}$, where M and d are co-prime integers, that is, $(M, d) = 1$, and $d > 1$. This array may be thought of as two subarrays. The first is a downsampled version, by a factor d, of an $(m-1)d$-element uniform line array (ULA) with half wavelength interelement spacings. The second is a single sensor at location M in the line array such that M and d are co-prime. We call this the sparse array because of the single element that sits apart from the origin of the first subarray. We note that M need not be greater than $(m-2)d$.

- *Co-prime array:* In this case, $\mathbb{I} = \mathbb{I}_1 \cup \mathbb{I}_2$, where $\mathbb{I}_1 = \{0, m_2, 2m_2, \ldots, (m_1 - 1)m_2\}$,

$\mathbb{I}_2 = \{m_1, 2m_1, \ldots, (2m_2 - 1)m_1\}$, and $(m_1, m_2) = 1$. Again the array is composed of two subarrays. The first is an m_1-element ULA with interelement spacings of m_2 and sensor locations $\mathbb{I}_1$. The second is a $(2m_2 - 1)$-element ULA with interelement spacings of m_1 and sensor locations $\mathbb{I}_2$. This co-prime geometry was recently introduced in [15] and [16].

Remark 2: In both cases, the co-prime constraint guarantees that aliasing ambiguities due to undersampling can be resolved. Although a sparse array can be viewed as a special case of a co-prime array, we consider them separately, because it is easier to first derive a suitable characterization of the orthogonal subspace $\langle \mathbf{A} \rangle$ for a sparse array, and then generalize it to a co-prime array. Our parameterizations are not minimal. They involve $2p$ parameters, instead of p, but as we will show in Section 4.4, they are specifically designed to utilize modern methods of linear prediction and approximate least squares, such as IQML.

Remark 3: By now it should be clear that $\mathbf{A}(\mathbf{z}, \mathbb{I})$, and therefore its parameterization, depend on both the mode vector $\mathbf{z}$ and the array geometry $\mathbb{I}$. Therefore, from here on, we may drop $(\mathbf{z}, \mathbb{I})$ and simply use $\mathbf{A}$, $\mathbf{V}$, $\langle \mathbf{A} \rangle$, and $\langle \mathbf{V} \rangle$.

4.3 Characterization of the Orthogonal Subspace for Uniform Line Arrays

Consider a uniform line array of m equidistant sensors located at $\mathbb{I}_u = \{0, 1, 2, \ldots, m-1\}$, taking measurements from the superposition of p modes as in (4.2). The signal subspace in this case is characterized by the Vandermonde matrix $\mathbf{V}$ in (4.3) with $\mathbb{I} = \mathbb{I}_u$. To characterize the orthogonal subspace $\langle \mathbf{A} \rangle$, consider the polynomial $A(z)$:

$$\begin{aligned} A(z) &= \prod_{k=1}^{p} (1 - z_k z^{-1}) \\ &= \sum_{i=0}^{p} a_i z^{-i}; \quad a_0 = 1 \end{aligned} \tag{4.10}$$

which has $(z_1, z_2, \ldots, z_p)$ as its p complex roots. The $(m-p)$ dimensional orthogonal subspace $\langle \mathbf{A} \rangle$ is spanned by the $m-p$ linearly independent columns of $\mathbf{A}$:

$$\mathbf{A} = \begin{bmatrix} a_p & a_{p-1} & \cdots & a_1 & 1 & 0 & \cdots & 0 \\ 0 & a_p & & & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & & & \ddots & \vdots \\ 0 & \cdots & 0 & a_p & \cdots & \cdots & a_1 & 1 \end{bmatrix}^H . \tag{4.11}$$

Since $\mathbf{A}^H \mathbf{V} = \mathbf{0}$, and the columns of $\mathbf{A}$ are linearly independent, $\mathbf{V}$ and $\mathbf{A}$ span orthogonal subspaces $\langle \mathbf{V} \rangle$ and $\langle \mathbf{A} \rangle$ in $\mathbb{C}^m$. The above parameterization is at the heart of methods of linear prediction, approximate least squares, and IQML (see, e.g., [21]– [24]).

Using the p-parameter representation for $\langle \mathbf{A} \rangle$ in (4.11) we may re-write the least squares problem of (4.6) as

$$\hat{\mathbf{a}} = \underset{\mathbf{a}=[a_1,\ldots,a_p]^T \in \mathbb{C}^p}{\operatorname{argmin}} \sum_{n=0}^{N-1} \mathbf{y}^H[n] \mathbf{P}_{\mathbf{A}} \mathbf{y}[n]. \tag{4.12}$$

There are many algorithms to approximately solve the nonlinear least squares problem in (4.12). One approach is to ignore the $(\mathbf{A}^H \mathbf{A})^{-1}$ term in the projection matrix $\mathbf{P}_{\mathbf{A}} = \mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H$ and solve the following modified least squares or linear prediction problem:

$$\hat{\mathbf{a}} = \underset{\mathbf{a} \in \mathbb{C}^p}{\operatorname{argmin}} \sum_{n=0}^{N-1} \mathbf{y}^H[n] \mathbf{A} \mathbf{A}^H \mathbf{y}[n]. \tag{4.13}$$

The iterative quadratic maximum likelihood (IQML) algorithm (see [21], [22], and [53]) is another method to approximately solve (4.12). In the lth iteration of IQML, the parameters $\mathbf{a}_l$ are estimated by iteratively minimizing the quadratic form

$$\hat{\mathbf{a}}_l = \underset{\mathbf{a}_l \in \mathbb{C}^p}{\operatorname{argmin}} \ \mathbf{a}_l^H \left[\sum_{n=0}^{N-1} \mathbf{Y}^H[n] (\mathbf{A}_{l-1}^H \mathbf{A}_{l-1})^{-1} \mathbf{Y}[n] \right] \mathbf{a}_l, \tag{4.14}$$

where $\mathbf{A}_{l-1}$ is formed as in (4.11) using the estimated parameters $\hat{\mathbf{a}}_{l-1}$ from iteration $(l-1)$ and $\mathbf{y}[n]$ is the following $(m-p) \times p$ Hankel data matrix for snapshot n:

$$\mathbf{Y}[n] = \begin{bmatrix} y_0[n] & \cdots & y_{p-1}[n] & y_p[n] \\ y_1[n] & \cdots & y_p[n] & y_{p+1}[n] \\ \vdots & & \vdots & \vdots \\ y_{m-1-p}[n] & \cdots & y_{m-2}[n] & y_{m-1}[n] \end{bmatrix} . \tag{4.15}$$

After a number of these iterations the sequence $\{\hat{\mathbf{a}}_l\}$ converges to an estimate $\hat{\mathbf{a}} = [\hat{a}_1, \ldots, \hat{a}_p]^T$. The polynomial $\hat{A}(z) = \sum_{i=0}^{p} \hat{a}_i z^{-i}$ is then formed from this estimate and its roots are taken as the mode estimates $(\hat{z}_1, \hat{z}_2, \ldots, \hat{z}_p)$.

4.4 Characterization of the Orthogonal Subspaces for Sparse and Co-prime Arrays

In this Section, we present simple characterizations of the orthogonal subspace $\langle \mathbf{A} \rangle$ for the sparse and co-prime arrays discussed in Section 4.2. Based on our characterizations, we adopt IQML for approximate least squares estimation of complex exponential modes in such arrays.

4.4.1 Sparse Array

Consider the sparse array described in Section 4.2. The set of sensor locations for this array is $\mathbb{I}_s = \{0, d, 2d, \ldots, (m-2)d, M\}$. The generalized Vandermonde matrix $\mathbf{V}$ in this case is

$$\mathbf{V}(\mathbf{z}, \mathbb{I}_s) = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1^d & z_2^d & \cdots & z_p^d \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{(m-2)d} & z_2^{(m-2)d} & \cdots & z_p^{(m-2)d} \\ z_1^M & z_2^M & \cdots & z_p^M \end{bmatrix}. \tag{4.16}$$

For $d > 1$ it is clear that without the use of the last sensor at location M, we cannot unambiguously estimate the modes, because any two modes z_k and $z_k e^{j2\pi q/d}$, $q = 1, 2, \cdots, d-1$ produce the same measurement. This is the aliasing problem for subsampled arrays.

To characterize the $(m-p)$-dimensional orthogonal subspace $\langle \mathbf{A} \rangle$, determined by the modes $\{z_k\}_{k=1}^{p}$, we first form the polynomial $A(z)$ from the dth powers of z_k, namely the $w_k = z_k^d$, $k = 1, 2, \ldots, p$:

$$A(z) = \prod_{k=1}^{p} (1 - w_k z^{-1}) = \sum_{i=0}^{p} a_i z^{-i}; \quad a_0 = 1. \tag{4.17}$$

Since $\{w_k\}_{k=1}^p$ are the roots of $A(z)$, the first $m-p-1$ columns $\mathbf{A}_0$ of $\mathbf{A} \in \mathbb{C}^{m\times(m-p)}$, which is to satisfy $\mathbf{A}^H\mathbf{V} = \mathbf{0}$, can be written as

$$\mathbf{A}_0^H = \begin{bmatrix} a_p & a_{p-1} & \cdots & a_1 & 1 & 0 & \cdots & 0 \\ 0 & a_p & & & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & & & \ddots & \vdots \\ 0 & \cdots & 0 & a_p & \cdots & \cdots & a_1 & 1 & 0 \end{bmatrix}. \tag{4.18}$$

But of course any mode of the form $z_k e^{j2\pi q/d}$, $q = 1, \ldots, d-1$, would produce the same w_k and therefore the same $\mathbf{A}_0$. This is the ambiguity caused by aliasing.

Now, consider the polynomial

$$B(z) = z^M + \sum_{i=1}^{p} b_i z^{(p-i)d}. \tag{4.19}$$

Suppose the coefficient vector $\mathbf{b} = [b_1, b_2, \cdots, b_p]^T$ is such that the actual modes $\{z_k\}_{k=1}^p$ are the roots of $B(z)$. That is, $B(z_k) = 0$ for $k = 1, 2, \ldots, p$. Then, since M and d are co-prime, for $1 \le q \le d-1$ and $1 \le k \le p$ we have

$$\begin{aligned} B(z_k e^{j2\pi q/d}) &= z_k^M e^{j2\pi Mq/d} + \sum_{i=1}^{p} b_i z_k^{(p-i)d} \\ &= z_k^M (e^{j2\pi Mq/d} - 1) \\ &\neq 0 \quad \text{for } q = 1, 2, \ldots, d-1. \end{aligned} \tag{4.20}$$

Therefore, the only common roots of $B(z)$, and the dth roots of $\{w_k\}_{k=1}^p$, are $\{z_k\}_{k=1}^p$, which are the actual modes to be estimated. In this way, $B(z)$ resolves the ambiguities.

Now suppose $\{w_k\}_{k=1}^p$ are known (or estimated). Then from (4.19), $\mathbf{b}$ can be found by solving the linear system of equations

$$\begin{pmatrix} b_p & b_{p-1} & \ldots & b_1 \end{pmatrix} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ z_1^d & z_2^d & \cdots & z_p^d \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{(p-1)d} & z_2^{(p-1)d} & \cdots & z_p^{(p-1)d} \end{pmatrix} = -\begin{pmatrix} z_1^M & z_2^M & \ldots & z_p^M \end{pmatrix}, \tag{4.21}$$

which if $z_i^d \neq z_j^d$ for $i \neq j$ (as we assume) has a unique solution. Using the $2p$ coefficients $\{a_i\}_{i=1}^p$ and $\{b_i\}_{i=1}^p$, we can characterize $\langle \mathbf{A} \rangle$ by writing $\mathbf{A} \in \mathbb{C}^{m \times (m-p)}$ as

$$\mathbf{A} = \begin{bmatrix} a_p & a_{p-1} & \cdots & a_1 & 1 & 0 & \cdots & 0 \\ 0 & a_p & & & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & & & \ddots & \vdots \\ 0 & \cdots & 0 & a_p & \cdots & a_1 & 1 & 0 \\ b_p & b_{p-1} & \cdots & b_1 & 0 & \cdots & 0 & 1 \end{bmatrix}^H . \tag{4.22}$$

To estimate $\mathbf{a} = [a_1, \cdots, a_p]^T$ and $\mathbf{b} = [b_1, \cdots, b_p]^T$, we need to solve the following problem:

$$\min_{\mathbf{a},\mathbf{b}} \sum_{n=0}^{N-1} \mathbf{y}^H[n] \mathbf{P}_{\mathbf{A}} \mathbf{y}[n], \tag{4.23}$$

We approximate the solution to this problem in two steps. First, we ignore the last column of $\mathbf{A}$ and estimate $\mathbf{a}$ as

$$\hat{\mathbf{a}} = \operatorname*{argmin}_{\mathbf{a}} \sum_{n=0}^{N-1} \mathbf{y}^H[n] \mathbf{P}_{\mathbf{A}_0} \mathbf{y}[n]. \tag{4.24}$$

In the noiseless case, where the $\mathbf{y}[n]$, $n = 0, 1, \ldots, N-1$, lie in $\langle \mathbf{V} \rangle$, it can be shown that if $m \geq 2p+1$ then the solution to (4.24) is unique and yields the coefficients of the polynomial $A(z)$ with roots $(w_1, w_2, \ldots, w_p)$. See Appendix A.

The minimization problem in (4.24) can be solved using IQML. Now, given $\hat{\mathbf{a}}$, we form the polynomial

$$\begin{aligned} \hat{A}(z) &= 1 + \sum_{i=1}^{p} \hat{a}_i z^{-i} \\ &= \prod_{k=1}^{p} (1 - \hat{w}_k z^{-1}) \end{aligned} \tag{4.25}$$

and derive its roots as $\{\hat{w}_k\}_{k=1}^p$. We know from the structure of the problem that $\hat{w}_k = \hat{z}_k^d$, and any of the d-th roots of $\hat{z}_k^d$ is a candidate solution. Therefore, we construct the candidate set $\mathcal{R}$, which contains all modes and their aliased versions, as

$$\mathcal{R} = \{(\hat{z}_1 e^{j2\pi q_1/d}, \hat{z}_2 e^{j2\pi q_2/d}, \cdots, \hat{z}_p e^{j2\pi q_p/d}) | 0 \leq q_1, q_2, \ldots, q_p \leq d-1\}. \tag{4.26}$$

In the second step, to find the p actual modes and resolve aliasing ambiguities, we solve the following constrained linear prediction problem:

$$\hat{\mathbf{b}} = \underset{\boldsymbol{\zeta}}{\operatorname{argmin}} \sum_{n=0}^{N-1} |y_{m-1}[n] + \boldsymbol{\zeta}^T \mathbf{u}[n]|^2$$
$$\text{s.t.} \quad B_{\boldsymbol{\zeta}}(\hat{\mathbf{z}}) = 0, \quad \hat{\mathbf{z}} \in \mathcal{R}, \tag{4.27}$$

where $\mathbf{u}[n] = [y_0[n], y_1[n], \ldots, y_{p-1}[n]]^T$, and the polynomial $B_{\boldsymbol{\zeta}}(z)$ is obtained from replacing $\mathbf{b}$ by $\boldsymbol{\zeta}$ in (4.19).

In the noiseless case, where $\mathbf{y}[n]$, $n = 0, 1, \ldots, N-1$ lie in $\langle \mathbf{V} \rangle$, the solution $\hat{\mathbf{b}}$ to (4.27) satisfies (4.21) and yields the actual modes. See Appendix B.

Our algorithm for estimating modes in a sparse array may be summarized in the following steps:

1. Estimate $\hat{\mathbf{a}} = [\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_p]^T$ from (4.24) using IQML;
2. Root $\hat{A}(z)$ to return roots $\{\hat{w}_k\}_{k=1}^p$. Then, recognizing that the dth roots of $\hat{w}_k$ are $\hat{z}_k e^{j2\pi q/d}$ for some $q \in \{0, 1, 2, \ldots, d-1\}$, form the set of candidate modes $\mathcal{R}$ as in (4.26);
3. Solve (4.27) for $\hat{\mathbf{b}}$;
4. Intersect the roots of $\hat{B}(z)$ with $\mathcal{R}$.

4.4.2 Co-prime Array

Consider an $m = m_1 + 2m_2 - 1$ element co-prime array, consisting of two uniform sub-arrays: one with m_1 elements at locations $\mathbb{I}_1 = \{0, m_2, 2m_2, \ldots, (m_1-1)m_2\}$ and the other with $2m_2 - 1$ elements at locations $\mathbb{I}_2 = \{m_1, 2m_1, \ldots, (2m_2-1)m_1\}$, where $(m_1, m_2) = 1$ and $m_1 > m_2$. In this case, the generalized Vandermonde matrix $\mathbf{V} \in \mathbb{C}^{m \times p}$ of modes may be partitioned as

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}(\mathbf{z}, \mathbb{I}_1) \\ \mathbf{V}(\mathbf{z}, \mathbb{I}_2) \end{bmatrix}, \tag{4.28}$$

where

$$\mathbf{V}(\mathbf{z},\mathbb{I}_1)=\begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1^{m_2} & z_2^{m_2} & \cdots & z_p^{m_2} \\ z_1^{2m_2} & z_2^{2m_2} & \cdots & z_p^{2m_2} \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{(m_1-1)m_2} & z_2^{(m_1-1)m_2} & \cdots & z_p^{(m_1-1)m_2} \end{bmatrix} \tag{4.29}$$

and

$$\mathbf{V}(\mathbf{z},\mathbb{I}_2)=\begin{bmatrix} z_1^{m_1} & z_2^{m_1} & \cdots & z_p^{m_1} \\ z_1^{2m_1} & z_2^{2m_1} & \cdots & z_p^{2m_1} \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{(2m_2-1)m_1} & z_2^{(2m_2-1)m_1} & \cdots & z_p^{(2m_2-1)m_1} \end{bmatrix} \tag{4.30}$$

are the Vandermonde matrices for the two individual subarrays of the co-prime array.

Let $\mathbf{A}_1 \in \mathbb{C}^{m_1\times(m_1-p)}$ and $\mathbf{B}_1 \in \mathbb{C}^{(2m_2-1)\times(2m_2-1-p)}$ be matrices that are orthogonal to $\mathbf{V}(\mathbf{z},\mathbb{I}_1)$ and $\mathbf{V}(\mathbf{z},\mathbb{I}_2)$, respectively. That is, $\mathbf{A}_1^H\mathbf{V}(\mathbf{z},\mathbb{I}_1)=\mathbf{0}$ and $\mathbf{B}_1^H\mathbf{V}(\mathbf{z},\mathbb{I}_2)=\mathbf{0}$. Following our results in the sparse case, we may parameterize $\mathbf{A}_1 \in \mathbb{C}^{m_1\times(m_1-p)}$ as

$$\mathbf{A}_1^H=\begin{bmatrix} a_p & a_{p-1} & \cdots & a_1 & 1 & \cdots & 0 \\ 0 & a_p & \cdots & & & \cdots & 0 \\ \vdots & \ddots & \ddots & & \cdots & & \vdots \\ 0 & \cdots & 0 & a_p & \cdots & a_1 & 1 \end{bmatrix}, \tag{4.31}$$

where $\{a_i\}_{i=1}^p$ are the coefficients of a polynomial $A(z)$, whose roots are $w_k = z_k^{m_2}$, $k = 1,2,\ldots,p$. That is,

$$\begin{aligned} A(z) &= \prod_{k=1}^{p}(1-w_k z^{-1}) \\ &= \sum_{i=0}^{p} a_i z^{-i}, \quad a_0 = 1. \end{aligned} \tag{4.32}$$

Similarly, we parameterize $\mathbf{B}_1 \in \mathbb{C}^{(2m_2-1)\times(2m_2-1-p)}$ as

$$\mathbf{B}_1^H = \begin{bmatrix} b_p & b_{p-1} & \cdots & b_1 & 1 & \cdots & 0 \\ 0 & b_p & \cdots & & & \cdots & 0 \\ \vdots & \ddots & \ddots & & \cdots & & \vdots \\ 0 & \cdots & 0 & b_p & \cdots & b_1 & 1 \end{bmatrix}, \tag{4.33}$$

where $\{b_i\}_{i=1}^p$ are the coefficients of a polynomial $B(z)$, whose roots are $s_k = z_k^{m_1}$, $k = 1, 2, \ldots, p$. That is,

$$\begin{aligned} B(z) &= \prod_{k=1}^{p}(1 - s_k z^{-1}) \\ &= \sum_{i=0}^{p} b_i z^{-i}, \quad b_0 = 1. \end{aligned} \tag{4.34}$$

Note that we still need p more independent columns to fully characterize the basis matrix $\mathbf{A}$ for the orthogonal subspace $\langle \mathbf{A} \rangle$. However, using our partial characterization, we can estimate the modes (with no aliasing ambiguities) in the following steps:

1. Separate the measurements of the two subarrays as $\mathbf{u}[n] = \{\mathbf{y}_i[n] | i \in \mathbb{I}_1\}$ and $\mathbf{v}[n] = \{\mathbf{y}_i[n] | i \in \mathbb{I}_2\}$;

2. Estimate $\hat{\mathbf{a}} = [\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_p]^T$ using IQML on $\mathbf{u}[n]$;

3. Root $\hat{A}(z)$ to return the roots $\{\hat{w}_k\}_{k=1}^p$. Then, recognizing that the m_2th roots of $\hat{w}_k$ are $\hat{z}_k e^{j2\pi q/m_2}$ for some $q \in \{0, 1, \ldots, m_2 - 1\}$, form the set of candidate modes $\mathcal{R}_1$ as

$$\mathcal{R}_1 = \{(\hat{z}_1 e^{j2\pi k_1/m_2}, \hat{z}_2 e^{j2\pi k_2/m_2}, \cdots, \hat{z}_p e^{j2\pi k_p/m_2}) \mid 0 \le k_1, k_2, \ldots, k_p \le m_2 - 1\}; \tag{4.35}$$

4. Estimate $\hat{\mathbf{b}} = [\hat{b}_1, \hat{b}_2, \ldots, \hat{b}_p]^T$ using IQML on $\mathbf{v}[n]$;

5. Root $\hat{B}(z)$ to return the roots $\{\hat{s}_k\}_{k=1}^p$. Then, recognizing that the m_1th roots of $\hat{s}_k$ are $\hat{z}_k e^{j2\pi q/m_1}$ for some $q \in \{0, 1, \ldots, m_1 - 1\}$, form the set of candidate modes $\mathcal{R}_2$ as

$$\mathcal{R}_2 = \{(\hat{z}_1 e^{j2\pi k_1/m_1}, \hat{z}_2 e^{j2\pi k_2/m_1}, \cdots, \hat{z}_p e^{j2\pi k_p/m_1}) \mid 0 \le k_1, k_2, \ldots, k_p \le m_1 - 1\}; \tag{4.36}$$

6. Intersect $\mathcal{R}_1$ and $\mathcal{R}_2$, in other words look for the closest (based on the Euclidean metric) p members of the set $\mathcal{R}_1$ to the set $\mathcal{R}_2$.

Remark 4: To complete the $2p-$parameterization of the basis matrix $\mathbf{A}$ for the orthogonal subspace $\langle \mathbf{A} \rangle$, consider the standard representation of $\mathbf{A}$ given in (4.9). Define $\mathbf{V}_1 = \mathbf{V}(\mathbf{z}, \mathbb{I}_{p_1})$ and $\mathbf{V}_2 = \mathbf{V}(\mathbf{z}, \mathbb{I}_{p_2})$ where $\mathbb{I}_{p_1} = \{0, m_2, \ldots, (p-1)m_2\}$ and $\mathbb{I}_{p_2} = \{m_1, 2m_1, \ldots, pm_1\}$. Then, from (4.9) the p remaining columns of $\mathbf{A}$ may be represented in $\mathbf{C}_1 \in \mathbb{C}^{n \times p}$ as:

$$\mathbf{C}_1^H = [\mathbf{C}_0^H |\ \mathbf{0}_{p \times (m_1 - p)} \mid \mathbf{I}_p \mid \mathbf{0}_{p \times (2m_2 - 1 - p)}] \tag{4.37}$$

where $\mathbf{0}_{k \times l}$ denotes a $k \times l$ matrix with zero entries, $\mathbf{I}_p$ is the $p \times p$ identity matrix, and

$$\mathbf{C}_0^H = -\mathbf{V}(\mathbf{z}, \mathbb{I}_{p_2}) \mathbf{V}^{-1}(\mathbf{z}, \mathbb{I}_{p_1}) \in \mathbb{C}^{p \times p}. \tag{4.38}$$

From (4.37) and (4.38) we can see that $\mathbf{C}_1$ only depends on $\{z_k^{m_1}\}_{k=1}^p$ and $\{z_k^{m_2}\}_{k=1}^p$ which are obtained from $\mathbf{a}$ and $\mathbf{b}$ by rooting $A_{\mathbf{a}}(z)$ and $B_{\mathbf{b}}(z)$ in (4.32) and (4.34), respectively. Therefore, the full, and minimally parameterized, characterization of the orthogonal subspace for the co-prime array may be written as

$$\mathbf{A} = \left[\begin{array}{cc|c} \mathbf{A}_1 & 0 & \\ & & \mathbf{C}_1 \\ 0 & \mathbf{B}_1 & \end{array} \right]. \tag{4.39}$$

We note that we do not need this full characterization for estimating the modes. The partial characterization using $\mathbf{A}_1$ and $\mathbf{B}_1$ suffices, at the expense of p fitting equations.

4.5 Numerical Results

In this Section we present numerical results for the estimation of damped complex exponential modes in co-prime, sparse and uniform line arrays. We consider a ULA of 50 elements. We form our co-prime and sparse arrays with 14 elements by subsampling this ULA. For the sparse array, we subsample the measurements of the ULA by a factor of $d = 4$ and place a sensor at $M = 3$. For the co-prime array, the first subarray includes $m_1 = 7$

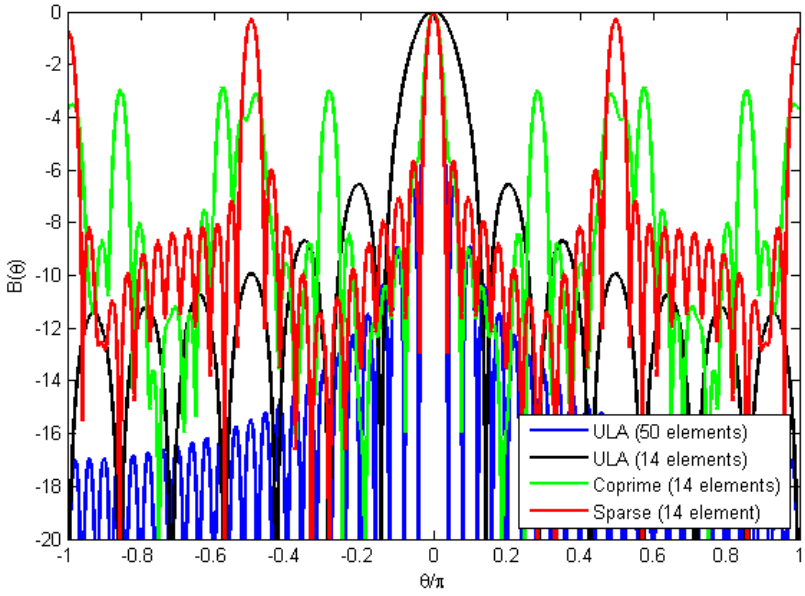


Figure 4.2: Beampatterns for ULAs with 14 and 50 elements, a sparse array with 14 elements, $d = 4$, and $M = 3$, and a co-prime array with 14 elements, $m_1 = 7$, and $m_2 = 4$.

elements with interelement spacing of $m_2 = 4$, and the second subarray includes $2m_2 - 1 = 7$ elements with interelement spacing of $m_1 = 7$.

It is insightful to first look at the beampatterns of sparse, co-prime, and uniform line arrays for the problem of estimating undamped modes. In this case, the beam pattern $B(\theta)$ is

$$B(\theta) = \sum_{l=0}^{m-1} e^{ji_l\theta}. \tag{4.40}$$

Figure 4.2 shows the beam patterns for different array geometries. Although the co-prime and sparse arrays of 14 elements have the same aperture and the same main lobe width as the ULA with 50 elements, we see that they suffer from higher sidelobes, suggesting that there will be performance losses in resolving closely spaced modes using these arrays, relative to the ULA.

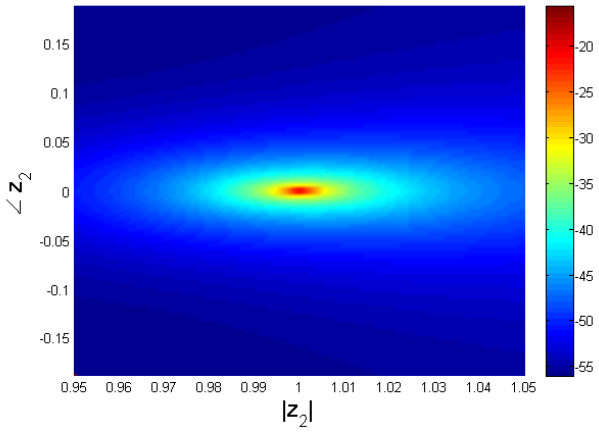
Let us also look at numerical results for the Cramér-Rao bound (CRB) associated with the co-prime, sparse and uniform line arrays (See Appendix C). Figure 4.3 shows the CRB in the estimation of the mode $z_1 = 1$ in the presence of an interfering mode z_2. The per

sensor SNR is 10 dB. As the interfering mode z_2 gets closer to z_1, the CRB in the estimation of the z_1 increases. The CRB for the sparse and co-prime arrays are similar, but they are higher than the CRB for the ULA. Since the aperture of the three arrays are equal, the fewer number of sensors in the sparse and co-prime arrays can be considered the only reason for this difference in the CRB.

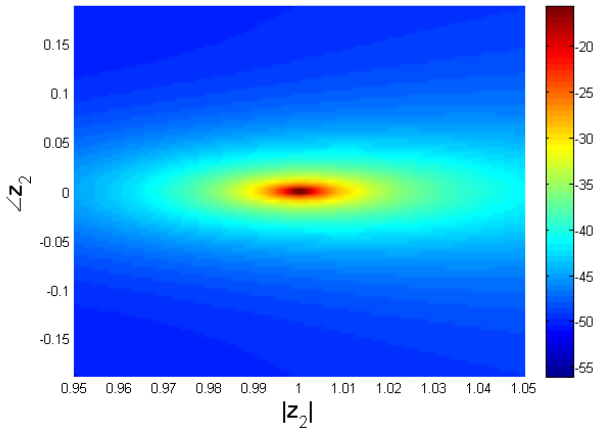
We now consider the performance of the approximate least squares estimation methods, shown in Figs. 4.4-4.6. The two modes to be estimated here are $z_1 = e^{j0.52}$ and $z_2 = 0.95e^{j0.69}$. We choose the per sensor SNR values for the sparse and co-prime arrays to be 5 dB higher than the SNR for the ULA, based on our insight in [52] about the threshold SNR for ULA and co-prime arrays. When SNR is decreased, there comes a point where some of the components of the orthogonal subspace better approximate the measurements than some of the components of the signal subspace. This leads to a performance breakdown in the estimation of the modes. The SNR at which this catastrophic breakdown occurs is called the threshold SNR (see [18] and [52]). For the compression ratio of 50/14, the threshold SNR for the co-prime and sparse arrays is almost 5 dB more than its value for the ULA, which is a consequence of the subsamplings by these arrays. We emphasize that any compression increases the SNR threshold. The use of a co-prime or sparse array instead of a dense uniform line array is only justified in applications where SNR is high enough for the desired estimation resolution, or when SNR can be built up from temporal snapshots using long observation periods. The latter of course requires the scene to remain stationary over the longer estimation period.

4.6 Acknowledgment

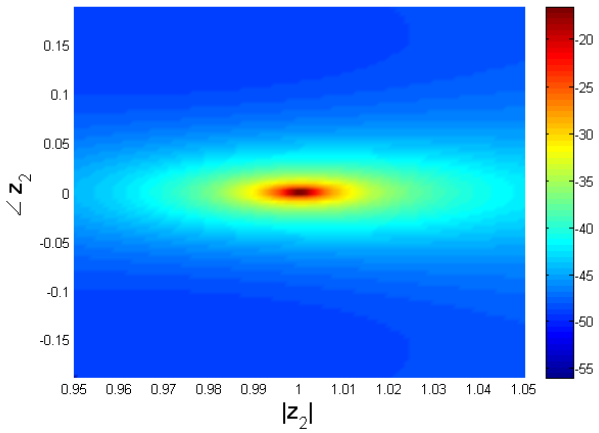
I would like to thank Prof. Chris Peterson for pointing out the minimal, p-dimensional, characterization of the orthogonal subspace in (4.9).



(a) ULA



(b) Sparse array



(c) Co-prime array

Figure 4.3: The CRB in dB for estimating $z_1 = 1$ in the presence of an interfering mode z_2: (a) ULA with 50 elements. (b) sparse array with 14 elements $d = 4$ and $M = 3$. (c) co-prime array with 23 elements $m_1 = 7$ and $m_2 = 4$. For all arrays per sensor SNR is 10 dB.

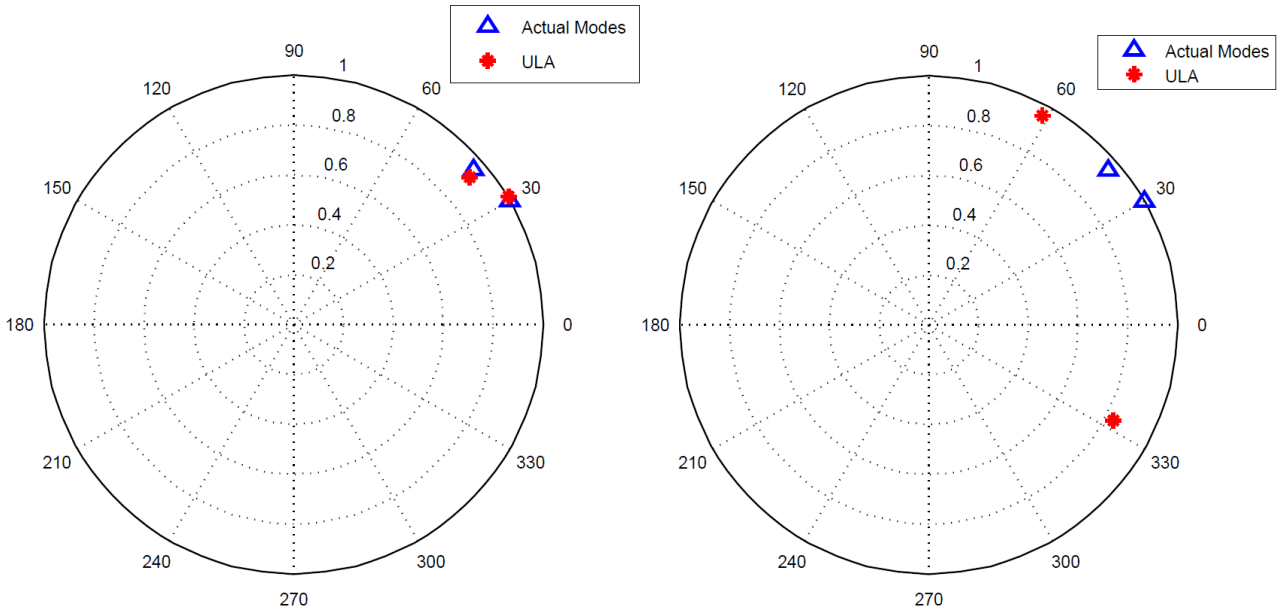


Figure 4.4: Estimating two closely spaced modes $z_1 = e^{j0.52}$ and $z_2 = 0.95e^{j0.69}$ using a ULA with 50 elements: (a) Per sensor SNR $= 0$ dB. (b) Per sensor SNR $= -5$ dB.

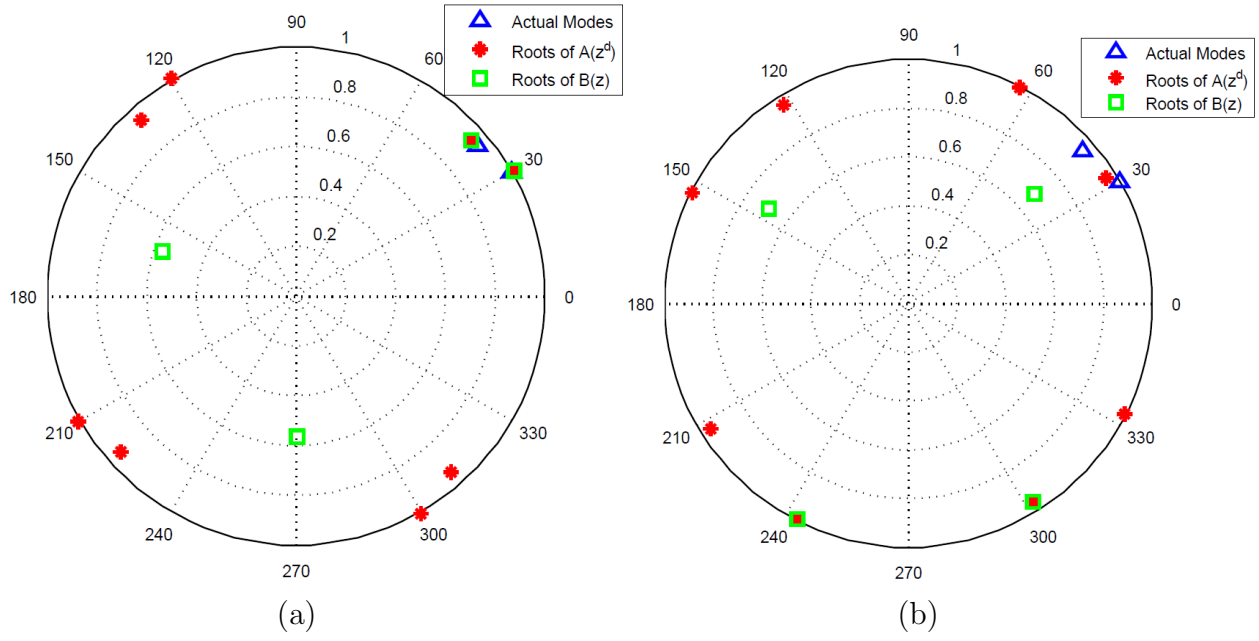


Figure 4.5: Estimating two closely spaced modes $z_1 = e^{j0.52}$ and $z_2 = 0.95e^{j0.69}$ using a sparse array with 14 elements, $d = 4$ and $M = 3$: (a) Per sensor SNR $= 5$ dB (b) Per sensor SNR $= 0$ dB.

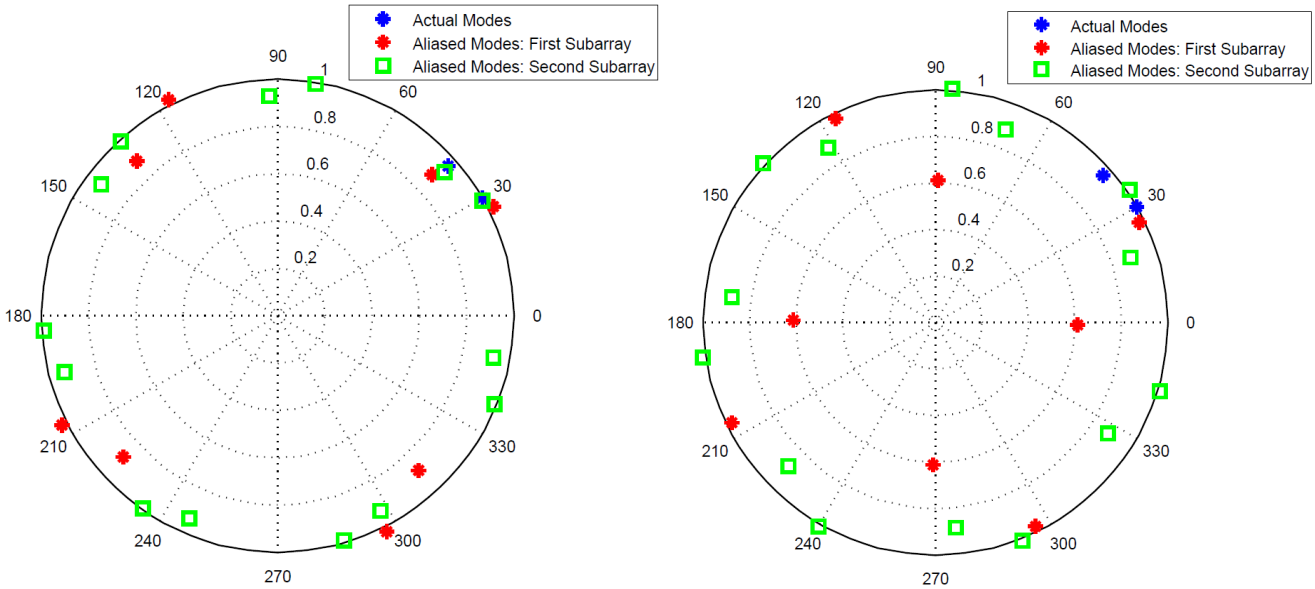


Figure 4.6: Estimating two closely spaced modes $z_1 = e^{j0.52}$ and $z_2 = 0.95e^{j0.69}$ using a co-prime array with 14 elements, $m_1 = 7$ and $m_2 = 4$: (a) Per sensor SNR = 5 dB (b) Per sensor SNR = 0 dB.

CHAPTER 5

SUMMARY

5.1 Conclusions

In this work, we have studied the problem of parameter estimation form sparse and compressed measurements. In Chapter 2, we have studied the effect of random compression of noisy measurements on the CRB for estimating parameters in a nonlinear model. We have considered the class of random compression matrices whose distributions are right-orthogonally invariant. A random compression matrix with i.i.d. standard normal elements is one such compression matrix. The analytical distribution for the normalized Fisher Information Matrix obtained in this chapter can be used to quantify the information loss due to compression. Also, they can be used as guidelines for choosing a suitable compression ratio based on a tolerable loss in the CRB. Importantly, the distribution of the ratios of CRBs before and after compression depends only on the number of parameters and the number of measurements. The distribution is invariant to the underlying signal-plus-noise model, in the sense that it is invariant to the uncompressed Fisher Information Matrix.

In Chapter 3, we have addressed the effect of compression on the probability of a subspace swap. A subspace swap is known to be the main source of performance breakdown in parameter estimation, wherein one or more modes of a noise subspace better approximate a measurement than one or more modes of a signal subspace. We have derived an analytical bound on this probability for two measurement models. In the first-order model, the parameters modulate the mean of a set of complex i.i.d. multivariate normal measurements. In the second-order model, the parameters to be estimated modulate a rank-deficient covariance matrix. Our lower bounds can be used to predict the threshold SNR. At a compression ratio of 7 to 1, our numerical experiments show that the threshold SNR increases by about 8 dB

when estimating a broadside source DOA in interference located at twice the Rayleigh limit of the pre-compressed array.

In summary, our results in Chapter 2 and Chapter 3 show that compression, whether by linear maps (e.g., Gaussian or Bernoulli) or by subsampling (e.g., co-prime) has performance consequences. The CRB in estimation and the onset of threshold SNR increase. Using our analysis in these two chapters, one can quantify the increases to determine if compressive sampling is viable, and if so, at what cost in performance.

In Chapter 4, we have considered the problem of estimating the parameters of p damped complex exponentials, from sparse or co-prime samples (in time or space) of their weighted sum. We have derived a $2p$-parameter characterization of the subspace that is orthogonal to the generalized Vandermonde subspace of the complex exponential modes. We then used this characterization to extend methods of linear prediction and approximate least squares for estimating mode parameters for sparse and co-prime arrays. We have also presented numerical examples demonstrating the performance of the proposed modal estimation approach. Our methods stand in contrast to MUSIC-type algorithms, which would return angles of mode parameters.

5.2 Future Work

An extension to our work in Chapter 2 is to study the distribution of the Fisher information matrix and the CRB after random compression, for the case that the parameters modulate the covariance of a complex multivariate normal model. The Fisher information matrix in this case has a more complicated structure, and as a special simplifying case, one may look at it in the asymptotic regime for large dimensions and fixed compression ratio. Also, the effect of compressed sensing on Bayesian, Bhattacharyya and Weiss-Weinstein bounds may be studied.

In Chapter 4, we derived a $2p$-parameter characterization of the orthogonal subspaces for the sparse and co-prime arrays. We later used these characterizations to apply modern

methods of linear prediction for modal analysis in these arrays. The sensitivity of our parameterization of the orthogonal subspace to sensor location errors is an open problem to be studied. Another path for future work is to consider orthogonal subspace characterizations for other nonuniform arrays such as nested arrays [16]. In the special case of DOA estimation, one may look for the best nonuniform array geometries that minimize the CRB in a given SNR region.

REFERENCES

[1] L. L. Scharf, *Statistical signal processing*, vol. 98. Addison-Wesley Reading, MA, 1991.

[2] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Transactions on Signal Processing*, vol. 59, pp. 2182–2195, May 2011.

[3] E. J. Candès, "Compressive sampling," in *Proc. Int. Congress Math.*, vol. 3, pp. 1433–1452, 2006.

[4] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[5] R. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, pp. 118–121, Jul. 2007.

[6] R. Baraniuk, M. Davenport, R. Devore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constr. Approx*, vol. 28, pp. 253–263, Dec. 2008.

[7] D. Needell and J. A. Tropp, "CoSamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, pp. 301–321, August 2008.

[8] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[9] E. J. Candés and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, pp. 4203–4215, December 2005.

[10] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Communications on Pure and Applied Mathematics*, vol. 63, no. 1, pp. 1–38, 2010.

[11] S. Sarvotham, D. Baron, and R. G. Baraniuk, "Sudocodes–fast measurement and reconstruction of sparse signals," in *IEEE International Symposium on Information Theory*, pp. 2804–2808, IEEE, 2006.

[12] Y. Chi, A. Pezeshki, L. L. Scharf, and R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," in *Proc. 2010 IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP), Dallas, TX*, pp. 3930–3933, Mar. 2010.

[13] L. L. Scharf, E. K. P. Chong, A. Pezeshki, and J. R. Luo, "Sensitivity considerations in compressed sensing," in *Conf. Rec. 45th Annual Asilomar Conf. Signals, Systs., Computs., Pacific Grove, CA,*, pp. 744–748, Nov. 2011.

[14] L. L. Scharf, E. K. P. Chong, A. Pezeshki, and J. R. Luo, "Compressive sensing and sparse inversion in signal processing: Cautionary notes," in *Proc. 7th Workshop on Defence Applications of Signal Processing (DASP), Coolum, Queensland, Australia, Jul. 10-14, 2011.*

[15] P. P. Vaidyanathan and P. Pal, "Sparse sensing with co-prime samplers and arrays," *Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (DSP/SPE)*, vol. 51, pp. 3289–294, Jan 2011.

[16] P. P. Vaidyanathan and P. Pal, "Why does direct-MUSIC on sparse-arrays work?," in *Proc. Asilomar Conference on Signals, Systems and Computers*, pp. 2007–2011, Nov 2013.

[17] D. Tufts, A. Kot, and R. Vaccaro, "The threshold effect in signal processing algorithms which use an estimated subspace," in *SVD and Signal Processing II: Algorithms, Analysis and Applications*, pp. 301–320, New York: Elsevier, 1991.

[18] J. K. Thomas, L. L. Scharf, and D. W. Tufts, "The probability of a subspace swap in the SVD," *IEEE Transactions on Signal Processing*, vol. 43, no. 3, pp. 730–736, 1995.

[19] B. A. Johnson, Y. I. Abramovich, and X. Mestre, "MUSIC, G-MUSIC, and maximum-likelihood performance breakdown," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3944–3958, 2008.

[20] R. Prony, "Essai expérimental et analytique sur les lois de la dilatabilité des fluides élastiques, et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alkool, á différentes températures," *Journal de L'École Polytechnique (Paris)*, vol. 1, no. 2, pp. 24–76, 1795.

[21] R. Kumaresan, L. L. Scharf, and A. K. Shaw, "An algorithm for pole-zero modeling and spectral analysis," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, pp. 637–640, Jun. 1986.

[22] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, pp. 1081–1089, Oct. 1986.

[23] R. Kumaresan and D. Tufts, "Estimating the angles of arrival of multiple plane waves," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-19, pp. 134–139, Jan 1983.

[24] R. Kumaresan, D. Tufts, and L. Scharf, "A Prony method for noisy data: Choosing the signal components and selecting the order in exponential signal models," *Proceedings of the IEEE*, vol. 72, pp. 230–233, Feb 1984.

[25] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, pp. 276–280, Mar 1986.

[26] A. Paulraj, R. Roy, and T. Kailath, "Estimation of signal parameters via rotational invariance techniques - ESPRIT," in *Proc. Asilomar Conference on Signals, Systems and Computers*, Nov 1985.

[27] F. Belloni, A. Richter, and V. Koivunen, "Extension of root-MUSIC to non-ULA array configurations," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 4, pp. 897–900, Toulouse, France, May. 2006.

[28] M. Rubsamen and A. B. Gershman, "Direction-of-arrival estimation for nonuniform sensor arrays: from manifold separation to fourier domain MUSIC methods," *IEEE Transactions on Signal Processing*, vol. 57, pp. 588–599, Feb. 2009.

[29] B. Friedlander and A. J. Weiss, "Direction finding using spatial smoothing with interpolated arrays," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, pp. 574–587, Apr.

[30] B. Friedlander, "The root-MUSIC algorithm for direction finding with interpolated arrays," *Signal Processing*, vol. 30, pp. 15–29, Jun. 1993.

[31] F. Gao and A. B. Gershman, "A generalized ESPRIT approach to direction-of-arrival estimation," *IEEE signal processing letters*, vol. 12, pp. 254–257, Mar. 2005.

[32] L. Coluccio, A. Eisinberg, and G. Fedele, "A prony-like method for non-uniform sampling," *Signal Processing*, vol. 87, pp. 2484–2490, Oct. 2007.

[33] T. Peter, D. Potts, and M. Tasche, "Nonlinear approximation by sums of exponentials and translates," *SIAM Journal on Scientific Computing*, vol. 33, pp. 1920–1947, Aug. 2011.

[34] P. Pakrooh, A. Pezeshki, L. L. Scharf, D. Cochran, and S. D. Howard, "Analysis of Fisher information and the Cramér-Rao bound for nonlinear parameter estimation after compressed sensing," *IEEE Transactions on Signal Processing, Submitted*, Feb. 2015.

[35] P. Pakrooh, L. L. Scharf, and A. Pezeshki, "Threshold effects in parameter estimation from compressed data," *IEEE Transactions on Signal Processing, Submitted*, Apr. 2015.

[36] P. Pakrooh, L. L. Scharf, and A. Pezeshki, "Modal analysis using sparse and co-prime arrays,," *IEEE Transactions on Signal Processing, Submitted*, Apr. 2015.

[37] B. Babadi, N. Kalouptsidis, and V. Tarokh, "Asymptotic achievability of the Cramér-Rao bound for noisy compressive sampling," *IEEE Transactions on Signal Processing*, vol. 57, pp. 1233–1236, Mar. 2009.

[38] R. Niazadeh, M. Babaie-Zadeh, and C. Jutten, "On the achievability of Cramér-Rao bound in noisy compressed sensing," *IEEE Transactions on Signal Processing*, vol. 60, pp. 518–526, Jan. 2012.

[39] D. Ramasamy, S. Venkateswaran, and U. Madhow, "Compressive estimation in AWGN: General observations and a case study," in *Conf. Rec. 46th Annual Asilomar Conf. Signals, Systs., Computs., Pacific Grove, CA,*, pp. 953–957, Nov. 2012.

[40] J. K. Nielsen, M. G. Christensen, and S. H. Jensen, "On compressed sensing and the estimation of continuous parameters from noisy observations," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3609–3612, Kyoto, Japan, Mar. 2012.

[41] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Prentice-Hall, Apr. 1993.

[42] L. L. Scharf and L. T. McWhorter, "Geometry of the Cramér-Rao bound," *Signal Process.*, vol. 31, pp. 301–311, Apr. 1993.

[43] M. Srivastava, "On the complex Wishart distribution," *The Annals of Mathematical Statistics*, pp. 313–315, 1965.

[44] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Transactions on Aerospace and Electronic Systems*, no. 6, pp. 853–863, 1974.

[45] A. T. James, "Distributions of matrix variates and latent roots derived from normal samples," *The Annals of Mathematical Statistics*, pp. 475–501, 1964.

[46] Y. Chikuse, *Statistics on special manifolds*. New York: Springer, 2003.

[47] A. K. Gupta, D. K. Nagar, and E. Bedoya, "Properties of the complex matrix variate Dirichlet distribution," *Scientiae Mathematicae Japonicae*, vol. 66, no. 1, p. 53, 2007.

[48] C. F. Dunkl and D. E. Ramirez, "Computation of the generalized F distribution," *Australian & New Zealand Journal of Statistics*, vol. 43, no. 1, pp. 21–31, 2001.

[49] H. L. Van Trees, *Detection, Estimation, and Modulation Theory: Radar-Sonar Signal Processing and Gaussian Signals in Noise*. Krieger Publishing Co., Inc., 1992.

[50] A. Moffet, "Minimum-redundancy linear arrays," *IEEE Trans. Antennas and Propagation*, vol. 16, pp. 172–175, Mar. 1968.

[51] P. Pakrooh, L. L. Scharf, A. Pezeshki, and Y. Chi, "Analysis of Fisher information and the Cramér-Rao bound for nonlinear parameter estimation after compressed sensing," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6630–6634, Vancouver, BC, Canada, May 2013.

[52] P. Pakrooh, A. Pezeshki, and L. L. Scharf, "Threshold effects in parameter estimation from compressed data," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 997–1000, Dec 2013.

[53] J. H. McClellan and D. Lee, "Exact equivalence of the Steiglitz-McBride iteration and IQML," *IEEE Transactions on Signal Processing*, vol. 39, pp. 509–512, Feb. 1991.

APPENDIX A

Let $m > 2p$ and $z_i^d \neq z_j^d$ for $i \neq j$. Also, let $\mathbf{a} = (a_1, a_2, \ldots, a_p)^T$ be the solution to (4.24). That is,

$$\mathbf{a} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \sum_{n=0}^{N-1} \mathbf{y}^H[n] \mathbf{P}_{\mathbf{A}_0(\boldsymbol{\alpha})} \mathbf{y}[n], \tag{A.1}$$

where $\mathbf{A}_0(\boldsymbol{\alpha})$ denotes an $\mathbf{A}_0$ of the form (4.18), with α_i's replacing a_i's for $i = 1, 2, \ldots, p$. We wish to show that in the noiseless case the roots of the polynomial $A(z) = 1 + \sum_{i=1}^{p} a_i z^{-i}$ are $w_k = z_k^d$ for $k = 1, 2, \ldots, p$ (the dth power of the actual modes).

Without loss of generality we assume that $N = 1$. Let $\mathbf{y} = \mathbf{y}[0]$. In the noiseless case, the minimum value of the objective function $\mathbf{y}^H \mathbf{P}_{\mathbf{A}_0(\boldsymbol{\alpha})} \mathbf{y}$ is zero, and because $\mathbf{A}_0^H(\boldsymbol{\alpha}) \mathbf{A}_0(\boldsymbol{\alpha})$ is always full rank, for the solution vector $\mathbf{a}$, we have $\mathbf{A}_0^H(\mathbf{a}) \mathbf{y} = \mathbf{0}$. Now, in the noiseless case, $\mathbf{y} = \mathbf{V}(\mathbf{z}, \mathbb{I}_s)\mathbf{x}$, where $\mathbf{V}(\mathbf{z}, \mathbb{I}_s)$ is given by (4.16) and $\mathbf{x} = [x_1, x_2, \ldots, x_p]^T$ is the vector of mode weights. Therefore, we have

$$\mathbf{A}_0^H(\mathbf{a}) \mathbf{V}(\mathbf{z}, \mathbb{I}_s) \mathbf{x} = \mathbf{0}, \tag{A.2}$$

which we can reorder to get

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1^d & z_2^d & \cdots & z_p^d \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{(m-p-1)d} & z_2^{(m-p-1)d} & \cdots & z_p^{(m-p-1)d} \end{bmatrix} \begin{bmatrix} x_1 & 0 & \cdots & 0 \\ 0 & x_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_p \end{bmatrix} \begin{bmatrix} A(w_1) \\ A(w_2) \\ \vdots \\ A(w_p) \end{bmatrix} = \mathbf{0}. \tag{A.3}$$

Because the matrix on the left hand side of (A.3) is a full column rank Vandermonde matrix, and the diagonal matrix in the middle is nonsingular (by the assumption of having p actual modes), the above equality holds iff

$$A(w_k) = 0 \quad \text{for } k = 1, \ldots, p. \tag{A.4}$$

APPENDIX B

Let $m > 2p$ and $z_i^d \neq z_j^d$ for $i \neq j$. Also, let $\boldsymbol{\beta}$ be the solution to (4.27) in the noiseless case. Here we use $\boldsymbol{\beta}$ instead of $\hat{\mathbf{b}}$ for the solution to distinguish noiseless and noisy cases. We show $\boldsymbol{\beta}$ solves (4.21) and therefore resolves aliasing.

Again, without loss of generality, we assume $N = 1$. Based on our argument in Appendix A, in the noiseless case we have

$$\mathcal{R} = \{(z_1 e^{j2\pi k_1/d}, z_2 e^{j2\pi k_2/d}, \cdots, z_p e^{j2\pi k_p/d}) \mid 0 \leq k_1, k_2, \ldots, k_p \leq d-1\}, \tag{B.1}$$

where $\{z_i\}_{i=1}^p$ are the actual modes. In this case (4.27) can be rewritten as:

$$\boldsymbol{\beta} = \arg\min_{\boldsymbol{\zeta}} |y_{m-1} + \boldsymbol{\zeta}^T \mathbf{u}|^2 \quad \text{s.t. } \mathbf{V}_p^T(\boldsymbol{\eta})\boldsymbol{\zeta} = -\boldsymbol{\eta}^{\odot M}, \ \boldsymbol{\eta} \in \mathcal{R}, \tag{B.2}$$

where $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_p)$, $\boldsymbol{\eta}^{\odot M} = [\eta_1^M, \eta_2^M, \ldots, \eta_p^M]^T$ and

$$\mathbf{V}_p(\boldsymbol{\eta}) = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \eta_1^d & \eta_2^d & \cdots & \eta_p^d \\ \vdots & \vdots & \ddots & \vdots \\ \eta_1^{(p-1)d} & \eta_2^{(p-1)d} & \cdots & \eta_p^{(p-1)d} \end{bmatrix}. \tag{B.3}$$

In the noiseless case, $\mathbf{u} = \mathbf{V}_p(\mathbf{z})\mathbf{x}$, $y_{m-1} = \sum_{i=1}^p x_i z_i^M$ and $\mathbf{V}_p(\boldsymbol{\eta}) = \mathbf{V}_p(\mathbf{z})$. Therefore, we have

$$\begin{aligned} |y_{m-1} + \boldsymbol{\beta}^T \mathbf{u}|^2 &= |\sum_{i=1}^p x_i z_i^M + \boldsymbol{\beta}^T \mathbf{V}_p(\mathbf{z})\mathbf{x}| \\ &= |\sum_{i=1}^p x_i (z_i^M - \eta_i^M)| \geq 0. \end{aligned} \tag{B.4}$$

Now, because $\boldsymbol{\beta}$ is the solution to (B.2), then in the noiseless case $|y_{m-1} + \boldsymbol{\beta}^T \mathbf{u}|^2 = 0$ and from (B.4) we have $\boldsymbol{\eta}^{\odot M} = \mathbf{z}^{\odot M}$ almost surely. Therefore,

$$\begin{aligned} \boldsymbol{\beta} &= -(\mathbf{V}_P^T(\boldsymbol{\eta}))^{-1} \boldsymbol{\eta}^{\odot M} \\ &= -(\mathbf{V}_P^T(\mathbf{z}))^{-1} \mathbf{z}^{\odot M} \\ &= \mathbf{b}, \end{aligned} \tag{B.5}$$

where $\mathbf{b}$ is the solution to (4.21). ■

APPENDIX C

The Fisher information matrix for the measurement model in (4.2) may be written as

$$\mathbf{J}(\mathbf{z}) = \sum_{n=1}^{N} \mathbf{J}_n(\mathbf{z}), \quad n = 0, 1, \ldots, N-1, \tag{C.1}$$

where $\mathbf{J}_n(\mathbf{z})$ is the Fisher information matrix for the estimation of the modes $\mathbf{z} = [z_1, z_2, \cdots, z_p]^T$ from $\mathbf{y}[n]$ in (4.2). That is,

$$\mathbf{J}_n(\mathbf{z}) = \frac{1}{\sigma^2} \mathbf{G}_n^H(\mathbf{z}) \mathbf{G}_n(\mathbf{z}), \tag{C.2}$$

where $\mathbf{G}_n(\mathbf{z}) = [\mathbf{g}_1[n], \mathbf{g}_2[n], \cdots, \mathbf{g}_p[n]]$, and

$$\mathbf{g}_l[n] = \begin{bmatrix} i_0 z_l^{i_0 - 1} \\ i_1 z_l^{i_1 - 1} \\ \vdots \\ i_{(m-1)} z_l^{i_{(m-1)} - 1} \end{bmatrix} x_l[n] \tag{C.3}$$

is the lth sensitivity vector, for $1 \leq l \leq p$. The CRB for the estimation of the kth mode z_k is the kth diagonal element of $\mathbf{J}^{-1}(\mathbf{z})$.