ABSTRACT

Title of Thesis: Performance and Robustness Analysis of Co-Prime and Nested Sampling

Ali Koochakzadeh, Master of Science, 2016

Thesis Directed By: Professor Piya Pal
Department of Electrical and Computer Engineering

Coprime and nested sampling are well known deterministic sampling techniques that operate at rates significantly lower than the Nyquist rate, and yet allow perfect reconstruction of the spectra of wide sense stationary signals. However, theoretical guarantees for these samplers assume ideal conditions such as synchronous sampling, and ability to perfectly compute statistical expectations. This thesis studies the performance of coprime and nested samplers in spatial and temporal domains, when these assumptions are violated.

In spatial domain, the robustness of these samplers is studied by considering arrays with perturbed sensor locations (with unknown perturbations). Simplified expressions for the Fisher Information matrix for perturbed coprime and nested arrays are derived, which explicitly highlight the role of co-array. It is shown that even in presence of perturbations, it is possible to resolve $O(M^2)$ under appropriate conditions on the size of the grid. The assumption of small perturbations leads to a novel "bi-affine" model in terms of source powers and perturbations. The redundancies in the co-array are then exploited to eliminate the nuisance perturbation variable,

and reduce the bi-affine problem to a linear underdetermined (sparse) problem in source powers.

This thesis also studies the robustness of coprime sampling to finite number of samples and sampling jitter, by analyzing their effects on the quality of the estimated autocorrelation sequence. A variety of bounds on the error introduced by such non ideal sampling schemes are computed by considering a statistical model for the perturbation. They indicate that coprime sampling leads to stable estimation of the autocorrelation sequence, in presence of small perturbations. Under appropriate assumptions on the distribution of WSS signals, sharp bounds on the estimation error are established which indicate that the error decays exponentially with the number of samples. The theoretical claims are supported by extensive numerical experiments.

Performance and Robustness Analysis of Co-Prime and Nested Sampling

by

Ali Koochakzadeh

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Masters of Science
2016

Advisory Committee:
Professor Piya Pal, Chair/Advisor
Professor Behtash Babadi
Professor Carol Espy-Wilson

Table of Contents

List of Figures

List of Abbreviations

DOA	Direction of Arrival
ULA	Uniform Linear Array
WSS	Wide Sense Stationary
MUSIC	MUltiple SIgnal Classification
MRA	Minimum Redundancy Array
A/D	Analog to Digital
CRB	Cramér Rao Bound
FIM	Fisher Information Matrix
pdf	Probability Distribution Function
LASSO	Least Absolute Shrinkage and Selection Operator
RMSE	Root Mean Square Error

Chapter 1: Introduction

Parameter estimation is a central problem in statistical signal processing for sensor arrays, and has been studied for many decades [1]. Several problems fall under this general umbrella, two of which constitute the main focus of this thesis: Directions-of-Arrival (DOA) Estimation, and Spectrum Estimation. Recent findings in the field of array processing [2–4] have enabled us to look at these problems from a different perspective, by introducing the concept of co-arrays. Two novel sampling techniques, namely nested and coprime sampling have been proposed, that exploit the co-array geometry to achieve much higher efficiency in terms of required sampling resources, compared to their classical counterparts.

In the context of DOA estimation, uniform spatial sampling geometries such as the Uniform Linear Array (ULA) are well known to be capable of resolving $O(M)$ sources using M sensors [5, 6]. However, by exploiting the geometry of nested [2] and coprime [3] arrays, one can dramatically increase degrees of freedom from $O(M)$ to $O(M^2)$ leading to capability of estimating $O(M^2)$ source directions. In these techniques, the vectorized form of the covariance matrix of received signals is studied, which leads to the concept of virtual co-arrays.

Similarly, co-prime sampling [3, 7] can be used in order to estimate the spec-

trum of wide-sense stationary (WSS) processes at sub-Nyquist rates. Unlike common compressive sensing approaches that use sparsity to enable sampling at sub-Nyquist rates, co-prime sampling only requires the random process to be WSS. For a line spectrum process, co-prime sampling (with coprime numbers M,N) is shown to be capable of recovering $O(MN)$ sinusoidal frequencies [3, 4, 7].

In spite of the huge benefit gained by selecting the location of sensors or sampling instants in coprime/nested fashion, the stability and robustness of these samplers have not been analyzed thoroughly. In this thesis, we examine the robustness issues associated with these samplers under different scenarios. For spatial sampling, we assume that the sensors are not exactly located on their nominal nested or coprime locations. This often happens in practice due to lack of perfect calibration. We study the co-arrays associated with these perturbed arrays, and show that under some mild conditions, it is always possible to resolve both the unknown perturbations, and the source directions, even when the number of sources is more than the number of sensors. We also consider a special case where these perturbations are very small, so that we can use a linear approximation to simplify our model. We will show that this assumption leads to a specific *bi-affine* model, which can be recast as a linear sparse recovery problem. Although the problem of array calibration has been previously studied in the context of arrays, this thesis presents the first results considering the perspective of co-arrays.

In the context of temporal sampling, we study the effect of jitter in the sampling instants, which, in principle, can be thought of as the counterpart to the location errors in the spatial sampling problem. However, there are some subtle

differences between the two problems which will be studied and elaborated in this thesis. We also study the effect of finite number of samples in a special case of spectrum estimation of the moving average random processes using co-prime samplers. The results show that co-prime sampling in the for spectral estimation is robust to finite number of samples as well as sampling jitter.

In this introductory chapter, we will review the problems of DOA estimation (spatial sampling), and spectral estimation (temporal sampling), using the nested and coprime samplers, and show how the concept of co-arrays arise.

Notations: Throughout this thesis, matrices are represented by bold uppercase letters, vectors by bold lowercase letters. The symbol x_i denotes the ith entry of a vector $\mathbf{x}$, while $\mathbf{x}_{a:b}$ represents the (Matlab-style) subvector of $\mathbf{x}$ starting at a and ending at b. Similarly, $\mathbf{X}_{(a:b,c:d)}$ stands for the submatrix of $\mathbf{X}$ consisting of the (a,b)th through (c,d)th elements. The symbols $(.)^*, (.)^T, (.)^H$ represent the conjugate, transpose, and hermitian, respectively. The symbols $\circ$, $\odot$, $\otimes$ stand for Hadamard, (column-wise) Khatri-Rao product, and Kronecker product, respectively. The symbol $\jmath$ denotes the imaginary unit $\sqrt{-1}$. The symbol $\|.\|_F$ represents the matrix Frobenius norm, while $\|.\|_0$, and $\|.\|$ denote the ℓ_0 pseudo-norm and ℓ_2 norm of a vector, respectively. The notation vec(.) represents the vectorized form of a matrix. Re$\{.\}$ returns the real part of a complex valued number.

1.1 DOA estimation

Direction of Arrival (DOA) estimation using an array of antennas is an important problem in array signal processing, which arises in many different scenarios such as target localization, radio astronomy, multi-microphone speech processing, etc [5]. In this Section, we review the recently proposed sparse array geometries, namely nested and coprime arrays, which are capable of localizing $O(M^2)$ sources using only M sensors.

1.1.1 Signal Model

Consider K narrow-band plane waves impinging on a linear antenna array with M sensors where $d_m\lambda/2$ is the distance of the mth antenna from the origin ($m = 1, \cdots, M$), and λ is the carrier wavelength. Let θ_i denote the Direction-of-Arrival (DOA) corresponding to the ith source and the corresponding spatial frequency is $\omega_i = \frac{\lambda}{2}\pi \sin(\theta_i)$, for $1 \leq i \leq K$. The received signal model is given by

$$\mathbf{y}[l] = \mathbf{A}(\boldsymbol{\omega})\mathbf{x}[l] + \mathbf{w}[l], \quad l = 1, 2, \cdots, L$$

where $\mathbf{y}[l] \in \mathbb{C}^M$ denotes the lth time snapshot of the signals received at the M sensors, $\mathbf{x}[l] \in \mathbb{C}^K$ are the unknown source signals and $\mathbf{w}[l] \in \mathbb{C}^M$ is the vector of additive noise, uncorrelated to source signals and satisfying $\mathbb{E}(\mathbf{w}\mathbf{w}^H) = \sigma^2\mathbf{I}$. Here, $\mathbf{A}(\boldsymbol{\omega}) = [\mathbf{a}(\omega_1), \mathbf{a}(\omega_2) \cdots \mathbf{a}(\omega_K)]$ with $\mathbf{a}(\omega_i)$ $(i = 1, \cdots, K)$ representing the steering vector associated with the angle θ_i, where its elements are given by $[\mathbf{a}(\omega_i)]_m = e^{\jmath\omega_i d_m}$.

Let $\mathbf{R}_\mathbf{x}$ be the covariance matrix of the sources, i.e., $\mathbf{R}_\mathbf{x} = \mathbb{E}(\mathbf{x}\mathbf{x}^H)$. Using

these assumptions, the covariance matrix of the received signals can be written as

$$\mathbf{R} := E(\mathbf{y}\mathbf{y}^H) = \mathbf{A}(\boldsymbol{\omega})\mathbf{R}_\mathbf{x}\mathbf{A}^H(\boldsymbol{\omega}) + \sigma^2\mathbf{I}$$

Assuming that the sources are statistically uncorrelated, $\mathbf{R}_\mathbf{x}$ will be a diagonal matrix, i.e., $\mathbf{R}_\mathbf{x} = \text{diag}(\tilde{\boldsymbol{\gamma}})$. It has been shown that certain sparse array geometries such as MRA [8], nested [2], and coprime [3] arrays are capable of recovering more sources than the number of sensors. Since $\mathbf{R}_\mathbf{x}$ is a diagonal matrix, we can write the vectorized form of $\mathbf{R}$ as

$$\mathbf{z} := \text{vec}(\tilde{\mathbf{R}}) = \mathbf{A}_{\text{ca}}(\boldsymbol{\omega})\tilde{\boldsymbol{\gamma}}, \tag{1.1}$$

where $\tilde{\mathbf{R}} := \mathbf{R} - \sigma_n^2\mathbf{I}$, $\mathbf{A}_{\text{ca}}(\boldsymbol{\omega}) = \mathbf{A}^*(\boldsymbol{\omega}) \odot \mathbf{A}(\boldsymbol{\omega})$, $\odot$ denotes the Khatri-Rao product [2]. Since the $((m-1)M+n, i)$-th element of $\mathbf{A}_{\text{ca}}(\boldsymbol{\omega})$ is given by $e^{\jmath\omega_i(d_m - d_n)}$, $1 \leq m, n \leq M$, $1 \leq i \leq D$, the elements of $\mathbf{A}_{\text{ca}}(\boldsymbol{\omega})$ are actually characterized by the difference co-array of the physical array:

$$S_{\text{ca}} = \{d_m - d_n, 1 \leq m, n \leq M\}$$

Therefore, $\mathbf{A}_{\text{ca}}(\boldsymbol{\omega})$ acts as a larger virtual array with sensors located at $d_m - d_n$ for $1 \leq m, n \leq M$. This longer virtual coarray provably helps us to recover $D > M$ sources, if the sensor locations follow some specific structures, such as the nested arrays [2].

Throughout this thesis, the size of the co-array is denoted by $|S_{\text{ca}}|$, and the size of the longest consecutive sequence of integers belonging to S_{ca} is represented by M_{ca}. Next, we will review three main array structures that we will deal with in the sequel. **Uniform Linear Arrays:** Uniform Linear Arrays (ULA) are the

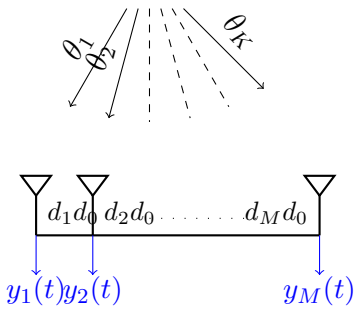


Figure 1.1: Signal Model for DOA Estimation

simplest and the most well studied array geometry. In a ULA, the sensors are placed uniformly on the locations $d_m = m - 1$. The corresponding difference co-array for such an array is then given by

$$S_{\text{ca}}^{\text{ULA}} = \{m | -M + 1 \leq m \leq M - 1\}$$

Therefore, we have $M_{\text{ca}} = 2M - 1 = O(M)$.

Nested Arrays: Assuming M to be even, a nested array [2] consists of $M/2$ sensors arranged as a ULA with spacing of 1 unit (inner ULA), and the remaining $M/2$ sensors on another ULA with spacing of $M/2 + 1$ units (outer ULA). In particular, $d_m = m-1$ (inner ULA), and $d_{m+M/2} = m(M/2+1)$ (outer ULA), for $1 \leq m \leq M/2$. The corresponding difference co-array for such an array is then given by

$$S_{\text{ca}}^{\text{nested}} = \{m | -M^2/4 - M/2 + 1 \leq m \leq M^2/4 + M/2 - 1\}$$

Therefore, we have $M_{\text{ca}} = M^2/4 + M/2 - 1 = O(M^2)$.

Co-prime Arrays: A co-prime array [3] consists $M = 2N_1 + N_2 - 1$ sensors, whose

normalized locations are given by

$$\{(i-1)N_1, 1 \leq i \leq N_2+1\} \cup \{(j-1)N_2, 1 \leq j \leq 2N_1\}$$

where N_1, N_2 are two co-prime numbers, so that $N_1 < N_2$. It is possible to show that the difference co-array corresponding to a co-prime array spans all the integer numbers in the range $-N_1N_2, \cdots, N_1N_2$. In other words, the difference co-array is given by

$$S_{\text{ca}}^{\text{co-prime}} \supseteq \{m| -N_1N_2 \leq m \leq N_1N_2\}$$

Hence, we have $M_{\text{ca}} = 2N_1N_2 + 1 = O(M^2)$.

1.2 Spectrum Estimation

Many signals in real-life applications can be modeled as stochastic processes. A broad range of them fall under the category of Wide Sense Stationary (WSS) processes. A fundamental property of a WSS process $x(t)$ is that $E(x^*(t_1)x(t_2)) = R(t_2 - t_1)$. In other words, the auto-correlation is only a function of difference of sampling instants. By judiciously choosing the sampling instants, it is possible to significantly lower the number of measurements, and yet be able to correctly identify the auto-correlation function. In essence, with little modifications, the co-prime and nested sampling can be used in order to identify the auto-correlation (or equivalently power spectral density) of the stochastic process. In this Section, we review a special case of spectrum estimation using co-prime sampling.

1.2.1 Co-prime sampling and Line Spectrum Estimation

Consider a specific class of WSS signals whose spectrum consists of lines, representing frequencies of sinusoids buried in noise. We briefly review the problem of estimating these frequencies using the concept of co-prime sampling [3]. Consider a signal $x(t)$ composed of K complex sinusoids, i.e., $x(t) = \sum_{k=1}^{K} A_k e^{\jmath(2\pi f_k t + \phi_k)}$. The signal $x(t)$ is WSS if the phases ϕ_k are assumed to be uniformly distributed on $[0, 2\pi]$. The signal is sampled using two A/D converters operating at rates $\frac{1}{MT}$ and $\frac{1}{NT}$, in which $1/T = 2f_{\max}$ is the Nyquist rate, yielding

$$x_1[n] = \sum_{k=1}^{K} A_k e^{\jmath\omega_k Mn + \jmath\phi_k} + w_n \tag{1.2}$$

$$x_2[m] = \sum_{k=1}^{K} A_k e^{\jmath\omega_k Mm + \jmath\phi_k} + w_m, \tag{1.3}$$

where $\omega_k = 2\pi f_k T$. Now, let us construct the vectors $\mathbf{y}_1[l] = [x_1[2Nl]\ x_1[2Nl + 1]\ \ldots\ x_1[2Nl + N - 1]]^T$, $\mathbf{y}_2[l] = [x_2[2Ml + 1]\ x_2[2Ml + 2]\ \ldots\ x_2[2Ml + 2M - 1]]^T$, and $\mathbf{y}[l] = [\mathbf{y}_1[l]^T\ \mathbf{y}_2[l]^T]^T$. Following [3], the autocorrelation matrix of $\mathbf{y}$ can be derived as

$$\mathbf{R}_\mathbf{y} = E\left(\mathbf{y}\mathbf{y}^H\right) = \sum_{k=1}^{K} A_k^2 \mathbf{B}(\omega_k) + \sigma^2 \mathbf{I} \tag{1.4}$$

where $\mathbf{B}(\omega_k) = \mathbf{a}(\omega_k)\mathbf{a}(\omega_k)^H$, $\mathbf{a}(\omega_k) = [\mathbf{a}_M(\omega_k)^T\ \mathbf{a}_N(\omega_k)^T]^T$ with

$$\mathbf{a}_M(w_k) = [1\ e^{\jmath\omega_k M}\ e^{2\jmath\omega_k M}\ \ldots\ e^{\jmath\omega_k M(N-1)}]$$

$$\mathbf{a}_N(w_k) = [e^{\jmath\omega_k N}\ e^{2\jmath\omega_k N}\ \ldots\ e^{\jmath\omega_k N(2M-1)}]$$

One can write the (1.4) in the vectorized form to get

$$\mathrm{vec}(\mathbf{R}_\mathbf{y}) = \sum_{k=1}^{K} \mathbf{b}(\omega_k) A_k^2 + \sigma^2 \mathrm{vec}(\mathbf{I}), \tag{1.5}$$

where $\mathbf{b}(\omega_k) = \mathbf{a}^*(\omega_k) \otimes \mathbf{a}(\omega_k)$, and $\otimes$ denotes the Kronecker product. The elements of $\mathbf{b}(\omega_k)$ have the form $e^{jw_k p}$, where p takes all the integer values between 0 and MN. This gives rise to possibility of detecting $O(MN)$ sinusoids by modifying well known line spectrum estimation algorithms such as MUSIC algorithm [3].

1.3 Outline of the thesis:

This thesis is organized as follows: In Chapter 2, we will consider the problem of DOA estimation in presence of array perturbations. By deriving the Cramér Rao bound, and proving the identifiability, we show that even in presence of co-array perturbations, it is possible to recover $O(M^2)$ number of sources. In Chapter 3, we consider a different perspective of the problem, one in which the perturbations are assumed to be sufficiently small. This assumption helps us to propose an approximate "bi-affine" model. We then show how it is possible to recast this model as a linear sparse problem. Our result shows that for small enough perturbations, it is possible to resolve the array perturbations. Chapter 4 considers the problem of temporal sampling, and studies the effect of jitter and finite samples on the co-prime sampling in the temporal domain.

The content of this thesis is derived from several published conference and journal papers [9–13].

Chapter 2: Performance Analysis of Spatial Nested and Coprime Sampling in Presence of Perturbation

2.1 Introduction

In Section 1.1.1, we showed how the novel array structures can provide us with enhanced an degrees of freedom. Generally, there are two main approaches which can leverage this enhanced degrees of freedom in order to improve the DOA (directions of arrival) estimation. The first approach is subspace-based methods, where typical subspace methods for arrays, such as MUSIC, are performed on the spatially smoothed co-array manifold [3]. In the second approach, a grid-based approximate model is considered and sparsity is exploited to recover the directions of sources. In this case, the range of all possible directions is discretized into a grid, and the DOA estimation problem is then reformulated as a sparse representation problem [4, 14–16]. In this Chapter, we consider the latter approach.

The results established so far are based on an unperturbed array geometry, one in which the sensor locations exactly satisfy the geometrical constraints. Array imperfections, however, are known to severely deteriorate the performance of DOA estimation algorithms [17], [18]. This is mainly due to strong dependence of these al-

gorithms to the availability of an accurately known array manifold. One of the most common forms of these imperfections is uncertainty about the accurate location of the sensors. To address the array imperfections and suppress their effect on DOA estimation, there are many methods proposed in the literature, which are known as *self-calibration* methods. In these methods, the perturbations are modeled as unknown but deterministic parameters, and these parameters are estimated jointly with the directions of the sources. The methods in [18], [19], and [20] resolve the sensor location uncertainty, using eigenstructure-based methods. [21] proposes a unified framework to formulate the array imperfections, and renders a sparse Bayesian perspective for array calibration and DOA estimation. In all of these methods, the array is assumed to be ULA, and none of the self-calibration methods in the literature use the concept of co-arrays.

The materials presented in this chapter have been published in [9, 11, 13].

2.2 Problem Formulation

In this Section we extend the model defined in Section 1.1.1, into the case that we have perturbations in sensor locations, assuming that the sources are located on a predefined grid.

2.2.1 A Grid-Based Model for Co-Array Perturbations

Similar to the model defined in Section 1.1.1, consider an array of M antennas impinged by K narrow-band sources with unknown directions $\boldsymbol{\theta}$. The sensors of the

array are originally designed to be at locations given by the vector $\lambda/2\mathbf{d} \in \mathbb{R}^M$, where λ is the carrier wavelength of the narrowband sources. However due to deformation of the sensor array, the sensors are actually located at $\lambda/2(\mathbf{d}+\boldsymbol{\delta})$, in which $\boldsymbol{\delta} \in \mathbb{R}^M$ is an unknown perturbation vector. Recall that corresponding to each direction of arrival (DOA) θ, we associate a spatial frequency $\omega = \frac{\lambda}{2}\pi \sin\theta$. In this Chapter, we use a grid-based model [4] for the spatial frequencies.

The range $[-\frac{\pi\lambda}{2}, \frac{\pi\lambda}{2})$, is quantized into a certain number of grid points, and we assume that the spatial frequencies indicating the source DOAs lie on this grid. This assumption simplifies much of our later derivations (based on the Cramér Rao bound) since the correlation matrix becomes a linear function of the unknown parameters of interest (signal powers, in our case).

Let $\mathbf{x}[l] \in \mathbb{C}^{N_\theta}$ be a K-sparse vector, containing a non zero subvector $\tilde{\mathbf{x}}[l] \in \mathbb{C}^K$ indicating the lth time sample of the K source signals. The location of the non zero elements of $\tilde{\mathbf{x}}[l]$ reveals the DOA (assuming they are on the grid). The received signal can be written as

$$\mathbf{y}[l] = \mathbf{A}_{\text{grid}}(\boldsymbol{\delta})\mathbf{x}[l] + \mathbf{w}[l], \tag{2.1}$$

in which $\mathbf{A}_{\text{grid}}(\boldsymbol{\delta}) = [\mathbf{a}(\omega_1, \boldsymbol{\delta}), \ldots, \mathbf{a}(\omega_{N_\theta}, \boldsymbol{\delta})]$ is an overcomplete dictionary representing the perturbed grid-based array manifold with ω_k denoting the kth grid point, i.e.,

$$\omega_k = \frac{\lambda}{2} \times \frac{2\pi k}{N_\theta}, k = 0, 1, \cdots, N_\theta - 1 \tag{2.2}$$

The vector $\mathbf{a}(\omega_i, \boldsymbol{\delta}) \in \mathbb{C}^M$ represents the steering vector corresponding to the kth

candidate spatial frequency on the grid and its mth element is given by

$$[\mathbf{a}(\omega_k, \boldsymbol{\delta})]_m = e^{\jmath\omega_k(d_m+\delta_m)}, \quad 1 \leq m \leq M$$

Clearly, this dictionary only depends on the perturbation $\boldsymbol{\delta}$, and the geometry of the array.

Since the sources are assumed to be uncorrelated, following [2], the vectorized form of the covariance matrix of the received signal is given by

$$\mathbf{z} = \mathbf{A}_{\text{ca}}(\boldsymbol{\delta})\boldsymbol{\gamma} + \sigma_w^2 \text{vec}(\mathbf{I}), \tag{2.3}$$

where $\mathbf{A}_{\text{ca}}(\boldsymbol{\delta}) = \mathbf{A}^*_{\text{grid}}(\boldsymbol{\delta}) \odot \mathbf{A}_{\text{grid}}(\boldsymbol{\delta})$ denotes the Khatri-Rao product [4]. The vector $\boldsymbol{\gamma} = [\gamma_1, \cdots, \gamma_{N_\theta}]$ is the diagonal of $\mathbf{R}_{\mathbf{x}} = E(\mathbf{x}\mathbf{x}^H)$. The location of the non zero elements of $\boldsymbol{\gamma}$ coincide with those of $\tilde{\mathbf{x}}[l]$ and reveal the exact DOAs of the source signals. The elements of $\mathbf{A}_{ca}(\boldsymbol{\delta})$ are characterized by the perturbed version of the difference co-array of the physical array. In particular, the $(m + m'M, k)$-th element of $\mathbf{A}_{\text{ca}}(\boldsymbol{\delta})$ is given by $e^{\jmath\omega_k(d_m+\delta_m-d_{m'}-\delta_{m'})}$. Thus each column of $\mathbf{A}_{\text{ca}}(\boldsymbol{\delta})$ is characterized by the perturbed difference co-array:

$$S_{\text{ca}} = \{d_m - d_{m'} + \delta_m - \delta_{m'}, 1 \leq m, m' \leq M\}$$

In the rest of this thesis, without loss of generality, we put the first sensor on the origin, i.e., $d_1 = \delta_1 = 0$. Using this grid-based model, DOA estimation becomes equivalent to recovering a sparse $\boldsymbol{\gamma}$ and identifying its non zero elements (or its support). We can suppress the effect of noise (σ_w^2) by removing the rows in $\mathbf{z}$ and $\mathbf{A}_{\text{ca}}(\boldsymbol{\delta})$ corresponding to $i - j = 0$. We further sort the rows in ascending order with respect to their location in the difference co-array and only keep the positive half to

obtain

$$\mathbf{z}^u = \mathbf{A}^u_{\text{ca}}(\boldsymbol{\delta})\boldsymbol{\gamma} \tag{2.4}$$

where $\mathbf{A}^u_{\text{ca}}(\boldsymbol{\delta}) \in \mathbb{C}^{\frac{|S_{\text{ca}}|-1}{2} \times N_\theta}$, and $|S_{\text{ca}}|$ is the number of distinct elements in the difference co-array.

2.2.2 Number of recoverable sources

The support of sparse $\boldsymbol{\gamma}$ represents the source DOAs. One way to recover the support is to assume that $\boldsymbol{\gamma}$ represents the sparsest vector satisfying (2.4) and solve

$$\min_{\boldsymbol{\gamma}\geq 0, \boldsymbol{\delta}} \|\boldsymbol{\gamma}\|_0 \quad \text{s.t.} \quad \mathbf{z} = \mathbf{A}^u_{ca}(\boldsymbol{\delta})\boldsymbol{\gamma}.$$

The size of the recoverable sparse support (or equivalently, the number of sources, K) in this case, fundamentally depends on the Kruskal Rank of $\mathbf{A}_{\text{ca}}(\boldsymbol{\delta})$.

2.3 Effect of Perturbations: A Cramér Rao Bound Based Study

In this Section, we derive a probabilistic model for the DOA estimation problem in the presence of perturbations. Let us assume that $\mathbf{x}[l], l = 1, 2, \cdots L$ are i.i.d random vectors with normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{R}_\mathbf{x})$, where $\mathbf{R}_\mathbf{x}$ is a diagonal matrix. Let $\mathbf{w}[l]$ be i.i.d Gaussian vectors, independent of the sources signals and distributed as $\mathcal{N}(\mathbf{0}, \sigma_w^2\mathbf{I})$. The received signal is distributed as

$$\mathbf{y}[l] \sim \mathcal{N}(\mathbf{0}, \underbrace{\mathbf{A}_{\text{grid}}(\boldsymbol{\delta})\mathbf{R}_\mathbf{x}\mathbf{A}^H_{\text{grid}}(\boldsymbol{\delta}) + \sigma_w^2\mathbf{I}}_{\mathbf{R}_\mathbf{y}})$$

We will assume the noise power σ_w^2 to be known, since our goal is to understand how the presence of unknown $\boldsymbol{\delta}$ affects the recovery of the desired parameter $\boldsymbol{\gamma}$.

Comment on Notations: For simplicity of notation, we will use $\mathbf{A}_{\text{grid}}, \mathbf{A}_{\text{ca}}, \mathbf{A}_{\text{ca}}^u$ instead of $\mathbf{A}_{\text{grid}}(\boldsymbol{\delta}), \mathbf{A}_{\text{ca}}(\boldsymbol{\delta}), \mathbf{A}_{\text{ca}}^u(\boldsymbol{\delta})$, respectively, in the sequel. Moreover, we will use the notations $\mathbf{A}_{\text{grid},0}, \mathbf{A}_{\text{ca},0}, \mathbf{A}_{\text{ca},0}^u$ to indicate the $\mathbf{A}_{\text{grid}}, \mathbf{A}_{\text{ca}}, \mathbf{A}_{\text{ca}}^u$ evaluated at $\boldsymbol{\delta} = \mathbf{0}$.

2.3.1 Cramér Rao Bound

Singularity of the Fisher Information Matrix implies non existence of a consistent estimator for $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ [22]. Hence, non-singularity of the FIM is a necessary condition for any algorithm to be able to exactly recover $\boldsymbol{\gamma}$ (in the limit as $L \to \infty$). However, it is non trivial to derive explicit conditions relating the array geometry and the range of parameters, for which the FIM is guaranteed to be non singular. In the following, we will conduct a deeper study of the algebraic structure of the perturbed FIM and derive explicit conditions under which such a guarantee will hold. As an important result, we will derive exact conditions on the size of the grid N_θ, size of the co-array M_{ca}, under which, the FIM will be shown to be non singular for almost all values of $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$. We would like to point out the following facts about the results derived in this Chapter:

- **Sparsity Not Assumed:** Although the parameter $\boldsymbol{\gamma}$ may be sparse (if $K < N_\theta$), we *do not impose a sparse prior on the model for deriving the FIM* and hence the guarantees hold regardless of our prior knowledge about the sparsity of $\boldsymbol{\gamma}$. In other words, under the derived conditions, a Maximum Likelihood

method can recover a sparse $\boldsymbol{\gamma}$ (as $L \to \infty$) from an overcomplete observation model (2.1) with $N_\theta > M$, without assuming it to be sparse.

- **Number of sources not assumed to be known:** We also *do not assume knowledge of the number of sources K in deriving the FIM*. Hence, the guarantees hold uniformly for any number of sources, as long as the established conditions are satisfied.

We now turn into deriving the Cramér Rao bound. The probability density function (pdf) of the received signal is given by

$$f(\mathbf{y};\boldsymbol{\psi}) = \frac{1}{\pi^M \det(\mathbf{R_y})} e^{-\mathbf{y}^H \mathbf{R_y}^{-1} \mathbf{y}}. \tag{2.5}$$

where $\boldsymbol{\psi} = [\boldsymbol{\gamma}^T \ \boldsymbol{\delta}^T]^T$ is the vector of parameters and $\mathbf{R_y}$ is a function of $\boldsymbol{\psi}$. The Fisher Information Matrix (FIM) is defined as

$$\mathbf{J}_{ij} = \mathbb{E}\left(\frac{\partial}{\partial \psi_i} \ln f(\mathbf{y};\boldsymbol{\psi}) \frac{\partial}{\partial \psi_j} \ln f(\mathbf{y};\boldsymbol{\psi}) \right). \tag{2.6}$$

Let us denote $\mathbf{W} = \mathbf{R_y}^{-T} \otimes \mathbf{R_y}^{-1}$, and define

$$\mathbf{H}_{\boldsymbol{\delta}} \triangleq [\mathrm{vec}(\mathbf{R}_{\delta_2}) \ \mathrm{vec}(\mathbf{R}_{\delta_3}) \ \ldots \ \mathrm{vec}(\mathbf{R}_{\delta_M})], \mathbf{R}_{\delta_i} \triangleq \frac{\partial \mathbf{R_y}}{\partial \delta_i} \tag{2.7}$$

The following theorem provides necessary and sufficient conditions under which the FIM in (2.6) is non singular:

Theorem 1. *Denoting $\boldsymbol{\psi} = [\boldsymbol{\gamma}^T \ \boldsymbol{\delta}_{2:M}^T]^T$ as the parameters to be estimated, the FIM defined in (2.6) is invertible, iff the matrix $\mathbf{B}$, defined as follows, is full column rank:*

$$\mathbf{B} = [\mathbf{A}_{ca} \ \ \mathbf{H}_{\boldsymbol{\delta}}] \tag{2.8}$$

Proof. For Gaussian distributed random variables with covariance matrix $\mathbf{R_y}$, the Fisher Information Matrix (FIM) can be derived as [23]

$$\frac{1}{L}\mathbf{J}_{ij} = \text{vec}(\frac{\partial \mathbf{R_y}}{\partial \psi_i})^H (\mathbf{R_y}^{-T} \otimes \mathbf{R_y}^{-1}) \text{vec}(\frac{\partial \mathbf{R_y}}{\partial \psi_j}) \tag{2.9}$$

The Fisher information Matrix (FIM) for our model (2.5) can be divided into blocks corresponding to parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ as:

$$\mathbf{J} = \begin{pmatrix} \mathbf{J}_{\boldsymbol{\gamma\gamma}} & \mathbf{J}_{\boldsymbol{\gamma\delta}} \\ \mathbf{J}_{\boldsymbol{\gamma\delta}}^H & \mathbf{J}_{\boldsymbol{\delta\delta}} \end{pmatrix} \tag{2.10}$$

Notice that

$$\text{vec}(\frac{\partial \mathbf{R_y}}{\partial \gamma_i}) = \text{vec}\Big(\mathbf{a}(\omega_i, \boldsymbol{\delta})\mathbf{a}^H(\omega_i, \boldsymbol{\delta})\Big) = \mathbf{a}^*(\omega_i, \boldsymbol{\delta}) \otimes \mathbf{a}(\omega_i, \boldsymbol{\delta})$$

Hence, from (2.9) and (2.7), we obtain

$$\mathbf{J}_{\boldsymbol{\gamma\gamma}} = L\mathbf{A}_{\text{ca}}^H \mathbf{W} \mathbf{A}_{\text{ca}}, \quad \mathbf{J}_{\boldsymbol{\gamma\delta}} = L\mathbf{A}_{\text{ca}}^H \mathbf{W} \mathbf{H}_{\boldsymbol{\delta}}, \quad \mathbf{J}_{\boldsymbol{\delta\delta}} = L\mathbf{H}_{\boldsymbol{\delta}}^H \mathbf{W} \mathbf{H}_{\boldsymbol{\delta}} \tag{2.11}$$

The FIM $\mathbf{J}$ can therefore be expressed as

$$\mathbf{J} = \mathbf{B}^H \mathbf{W} \mathbf{B} \tag{2.12}$$

Since $\mathbf{W} = \mathbf{R_y}^{-T} \otimes \mathbf{R_y}^{-1}$ is positive definite, it follows that $\text{rank}(\mathbf{J}) = \text{rank}(\mathbf{B})$. Hence $\mathbf{J}$ is non singular (i.e. has rank $N_\theta + M - 1$) if and only if $\mathbf{B} \in \mathbb{C}^{N_\theta + M - 1}$ has full column rank.

□

For the unperturbed signal model, the FIM is given by $\mathbf{J}_{\boldsymbol{\gamma\gamma}}$ and the following corollary establishes a necessary and sufficient condition for non singularity of FIM.

Corollary 1. (FIM in absence of perturbation) *The matrix* $\mathbf{J}_{\boldsymbol{\gamma\gamma}}$ *is invertible if and only if* $N_\theta \leq M_{ca}$.

Proof. From (2.11), $\text{rank}(\mathbf{J}_{\boldsymbol{\gamma\gamma}}) = \text{rank}(\mathbf{A}_{\text{ca},0})$, since $\mathbf{W}$ is positive definite. Hence, $\mathbf{J}_{\boldsymbol{\gamma\gamma}}$ is invertible if and only if $\text{rank}(\mathbf{A}_{\text{ca},0}) = N_\theta$. Since $\mathbf{A}_{\text{ca},0}$ has M_{ca} distinct rows which form a Vandermonde submatrix, $\mathbf{A}_{\text{ca},0}$ is full column rank if and only if $N_\theta \leq M_{\text{ca}}$, which concludes the proof. □

Remark 1. (Cramér Rao Bound) *If the FIM is invertible, the Cramér Rao bound can be obtained by computing the inverse of* $\mathbf{J}$. *Moreover, using the Schur complement of* $\mathbf{J}$, *the CRB corresponding to the parameter* $\boldsymbol{\gamma}$ *can be written as follows*

$$\begin{aligned}\frac{1}{L}\left(CRB_{\boldsymbol{\gamma\gamma}}\right)^{-1} &= \frac{1}{L}\left(\mathbf{J}_{\boldsymbol{\gamma\gamma}} - \mathbf{J}_{\boldsymbol{\gamma\delta}}\mathbf{J}_{\boldsymbol{\delta\delta}}^{-1}\mathbf{J}_{\boldsymbol{\gamma\delta}}^{H}\right)\\ &= \mathbf{A}_{ca}^{H}\mathbf{W}\mathbf{A}_{ca} - \mathbf{A}_{ca}^{H}\mathbf{W}\mathbf{H}_{\boldsymbol{\delta}}\left(\mathbf{H}_{\boldsymbol{\delta}}^{H}\mathbf{W}\mathbf{H}_{\boldsymbol{\delta}}\right)^{-1}\mathbf{H}_{\boldsymbol{\delta}}^{H}\mathbf{W}\mathbf{A}_{ca}\\ &= \mathbf{A}_{ca}^{H}\mathbf{W}^{1/2}\Pi^{\perp}_{\mathbf{W}^{1/2}\mathbf{H}_{\boldsymbol{\delta}}}\mathbf{W}^{1/2}\mathbf{A}_{ca},\end{aligned} \tag{2.13}$$

where $\Pi^{\perp}_{\mathbf{W}^{1/2}\mathbf{H}_{\boldsymbol{\delta}}}$ *is the projection matrix onto the null space of* $\mathbf{W}^{1/2}\mathbf{H}_{\boldsymbol{\delta}}$.

2.3.2 Necessary Condition on Size of Grid

The size of the co-array alone dictates the non singularity of FIM in absence of perturbation. However, for a perturbed signal model, it only imposes a necessary condition (not sufficient) on the invertibility of the FIM.

Corollary 2. *If* $N_\theta > |S_{ca}|$, $\mathbf{J}$ *is singular.*

Proof. Since $|S_{\text{ca}}|$ denotes the number of distinct elements in the perturbed co-array, it also represents the number of distinct rows of $\mathbf{A}_{\text{ca}}$. Hence $\text{rank}(\mathbf{A}_{\text{ca}}) \leq |S_{\text{ca}}|$ and when $N_\theta > |S_{\text{ca}}|$, $\mathbf{A}_{\text{ca}}$ is necessarily column rank deficient, implying $\mathbf{B}$ is also column rank deficient. Therefore, by Theorem 1, $\mathbf{J}$ is singular. □

Remark 2. Singularity of the FIM matrix $\mathbf{J}$ implies that there exist no unbiased estimator for $\boldsymbol{\psi}$ with finite variance [22]. The above necessary condition *imposes a restriction on the size of the grid with respect to the size of the co-array.* Recall that, in deriving $\mathbf{J}$, the number of sources K was assumed unknown. Furthermore, $\boldsymbol{\gamma}$ was not even assumed to be sparse; so in principle, the number of unknowns in $\boldsymbol{\gamma}$ is indeed the number of points (N_θ) on the entire grid. Hence, the necessary condition implies an equation-versus-unknown type of bound, where the number of equations are given by the distinct elements of the co-array. *If the grid size N_θ becomes larger than $|S_{ca}|$,* we will necessarily need to impose sparse prior on $\boldsymbol{\gamma}$ for it to be identifiable. *We will further elaborate on this point in Sec. 2.4.*

2.3.3 Sufficient Conditions for Invertible FIM

We now derive sufficient conditions under which the matrix $\mathbf{B}$ is full column rank. Note that $\mathbf{J}$ is a function of $\boldsymbol{\delta}$, and our goal will be to study for what range of values of $\boldsymbol{\delta}$, we can argue its non singularity for *almost all* $\boldsymbol{\gamma} \in \mathbb{R}^{N_\theta}$. We divide our analysis into two cases: In the first scenario, we find sufficient conditions under which $\mathbf{J}$ is invertible for almost all $\boldsymbol{\gamma}$ when $\boldsymbol{\delta} = \mathbf{0}$. Based upon this result, we will argue that under the same conditions, $\mathbf{J}$ will be invertible for almost all $\boldsymbol{\delta} \in \mathbb{R}^M$ as

well.

Notice that studying the non singularity of $\mathbf{J}$ for $\boldsymbol{\delta} = \mathbf{0}$ is fundamentally different from a problem setting where the location of the sensors are known to be not perturbed. We call the latter the "*unperturbed* problem". More precisely, in the *unperturbed* problem, the FIM is equal to the top left block of $\mathbf{J}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}$ of $\mathbf{J}$. Hence, the *unpertubed* problem is identifiable if and only if $\mathbf{J}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}$ is invertible, which simply reduces to $\mathbf{A}_{\mathrm{ca},0}$ being full column-rank. However, for our problem, $\boldsymbol{\delta} = \mathbf{0}$ is not a prior knowledge - rather it is just a particular values of the unknown parameter $\boldsymbol{\delta}$. Therefore, the invertibility of $\mathbf{J}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}$ does not imply the invertibility of $\mathbf{J}$ at $\boldsymbol{\delta} = \mathbf{0}$.

2.3.3.1 Non Singularity of $\mathbf{J}$ at $\boldsymbol{\delta} = \mathbf{0}$

Establishing sufficient conditions under which $\mathbf{J}$ is non singular at $\boldsymbol{\delta} = \mathbf{0}$ requires careful study of the co-array structure of the physical antenna array under question, and the details vary, depending on the array geometry. The following theorems state our main results for Uniform linear array (ULA) and a slightly modified version of the nested array:

Theorem 2. *(***ULA***) For a uniform linear array (ULA) with M sensors, if $N_\theta \leq 2M-2$, $\mathbf{J}|_{\boldsymbol{\delta}=0}$ is invertible for almost all $\boldsymbol{\gamma} \in \mathbb{R}^{N_\theta}$.*

This indicates a rather small grid size for the ULA. However, for the nested array, the grid size (for which $\mathbf{J}|_{\boldsymbol{\delta}=0}$ is guaranteed to be non singular) can be much larger, of $O(M^2)$. To prove this, we use a slightly modified version of the original nested array (assuming M is even), in which the location of the sensors are given

by $d_i\lambda/2$ where

$$d_i = (i-1)d, \quad d_{i+\frac{M}{2}} = \frac{M}{2}id \tag{2.14}$$

for $1 \leq i \leq M/2$. In this case, we can verify that $M_{\text{ca}} = M^2/2 + 1$. For the original nested array [2] (reviewed in Section 1.1.1), $M_{\text{ca}} = M^2/2 + M - 1$. We use this configuration to simplify the proof of the following theorem, which establishes conditions for non singularity of the FIM associated with this modified array:

Theorem 3. *(***Modified Nested Array***) For a modified nested array with M sensors (given by (2.14), with even M), if $N_\theta \leq M^2/2$, $\mathbf{J}|_{\boldsymbol{\delta}=0}$ is invertible for almost all $\boldsymbol{\gamma} \in \mathbb{R}^{N_\theta}$.*

Proof. The proofs can be found in Appendices A.1 and A.2. □

Remark 3. We would like to point out that a slightly stronger result can be established for the original nested array (which has more degrees of freedom, $M_{\text{ca}} = M^2/2 + M - 1$), for which the grid size can be shown to be $N_\theta \leq M_{\text{ca}} - M/2 = M^2/2 + M/2 - 1$. The proof technique will be similar to the one shown in Appendix A.2, with some modifications, which we avoid for ease of exposition.

Remark 4. This result indicates that for grids of size $O(M^2)$ (as long as the size is less than $M^2/2$), $\mathbf{J}$ is guaranteed to be non singular for almost all $\boldsymbol{\gamma}$ *even when we do not know the number of sources K*. This holds for overcomplete grid-based array manifolds $\mathbf{A}$ where the number of grid points can be as large as $O(M^2)$, *without the apriori assumption that the source scene is spare.*

2.3.3.2 Non Singularity of FIM: $\boldsymbol{\delta} \neq \mathbf{0}$

The non singularity of $\mathbf{J}$ for almost all $\boldsymbol{\delta}$ immediately follows from the conditions developed for $\boldsymbol{\delta} = 0$:

Theorem 4. *For ULA and nested array with M sensors, $\mathbf{J}$ is invertible for almost all $\boldsymbol{\delta} \in \mathbb{R}^{M-1}$ and $\boldsymbol{\gamma} \in \mathbb{R}^{N_\theta}$, if $N_\theta \leq 2M-2$ (for ULA) and $N_\theta \leq M^2/2$ (for modified nested array).*

Proof. Since elements of $\mathbf{J}$ are analytic functions of $\boldsymbol{\delta}$, $\det(\mathbf{J})$ is also an analytic function of $\boldsymbol{\delta}$. Therefore, $\det(\mathbf{J})$ has isolated zeros in $\boldsymbol{\delta}$ unless it is trivially zero [24]. However, in Theorems 2, and 3, we have shown that for ULA and modified nested array, $\det(\mathbf{J}) \neq 0$ at $\boldsymbol{\delta} = 0$ as long as $N_\theta \leq 2M - 2$ and $N_\theta \leq M^2/2$ respectively. This rules out the possibility that $\det(\mathbf{J})$ is trivially zero $\forall \boldsymbol{\delta}$. Therefore, the zeros of $\det(\mathbf{J})$ are isolated in $\mathbb{R}^{M-1}$, with a total measure zero. Thus, for almost all $\boldsymbol{\delta} \in \mathbb{R}^{M-1}$ and $\boldsymbol{\gamma} \in \mathbb{R}^{N_\theta}$, $\det(\mathbf{J}) \neq 0$, i.e., $\mathbf{J}$ is invertible. □

Thus, we have established the following key results regarding source localization using perturbed ULA and nested arrays:

- If we do not assume the number of sources K to be known (or, equivalently, do not assume the vector of source powers, $\boldsymbol{\gamma}$, to be sparse), $N_\theta \leq 2M - 2$ is sufficient for $\mathbf{J}$ to be non singular for almost all choices of $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$.

- For nested arrays, under the same assumption of K to be unknown, we can ensure the invertibility of $\mathbf{J}$ for almost all choices of $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ using a much larger overcomplete dictionary, where $N_\theta = O(M^2)$.

2.4 Non Singularity of FIM for Sparse Vectors

The guarantees for non singularity of $\mathbf{J}$ established so far holds for *almost all* choices of $\boldsymbol{\gamma} \in \mathbb{R}^{N_\theta}$. However, they do not ensure non-singularity of $\mathbf{J}$ at a *sparse* $\boldsymbol{\gamma}$, since the set of all sparse $\boldsymbol{\gamma}$ has zero measure in $\mathbb{R}^{N_\theta}$. We therefore need to refine our arguments to make them applicable to sparse $\boldsymbol{\gamma}$ as well. This can be studied for two distinct range of values of N_θ.

2.4.1 Non singularity for small grid size

In this case, we assume that $N_\theta \leq 2M-2$ for ULA, and $N_\theta \leq M^2/2$ for the modified nested array. We show that the Fisher Information Matrix is invertible at almost all sparse $\boldsymbol{\gamma}$. In particular, we have the following result:

Theorem 5. *Assume* $\boldsymbol{\delta} = \mathbf{0}$ *and consider the grid based model* (2.2). *For almost all sparse vectors* $\boldsymbol{\gamma} \in \mathbb{R}^{N_\theta}$ *with* $\|\boldsymbol{\gamma}\|_0 = K, K < N_\theta$, *where* $N_\theta \leq 2M-2$ *for ULA and* $N_\theta \leq \frac{M^2}{2}$ *for modified nested array,* $\mathbf{J}$ *is invertible for almost all* $\boldsymbol{\delta} \in \mathbb{R}^{M-1}$.

Proof. The proof can be found in the Appendix A.3. □

2.4.2 Singularity and Identifiability for $N_\theta > M_{\text{ca}}$

According to Corollary 2, when $N_\theta > |S_{\text{ca}}|$, the Fisher Information Matrix is necessarily singular. As we will show next, the parameter $\boldsymbol{\gamma}$ also becomes non identifiable in this case, and it becomes necessary to assume priors (such as sparsity) on $\boldsymbol{\gamma}$ to render it identifiable.

Definition 1. *Let* $f(\mathbf{y};\boldsymbol{\gamma},\boldsymbol{\delta})$ *be the probability density functions of* $\mathbf{y}$ *parameterized by* $(\boldsymbol{\delta},\boldsymbol{\gamma})$. *The parameters* $(\boldsymbol{\delta},\boldsymbol{\gamma})$ *are identifiable if* $f(\mathbf{y};\boldsymbol{\gamma},\boldsymbol{\delta}) = f(\mathbf{y};\boldsymbol{\gamma}',\boldsymbol{\delta}')$ *implies* $\boldsymbol{\delta} = \boldsymbol{\delta}', \boldsymbol{\gamma} = \boldsymbol{\gamma}'$.

Assuming that $\mathbf{y}$ has zero mean Gaussian distribution, the above definition of identifiability boils down to the uniqueness of the covariance matrix with respect to the parameters. In particular, for our model (2.5), uniqueness of the vectorized covariance matrices implies

$$\mathbf{A}_{\text{ca}}(\boldsymbol{\delta})\boldsymbol{\gamma} = \mathbf{A}_{\text{ca}}(\boldsymbol{\delta}')\boldsymbol{\gamma}' \Leftrightarrow \boldsymbol{\delta} = \boldsymbol{\delta}', \boldsymbol{\gamma} = \boldsymbol{\gamma}' \tag{2.15}$$

We will analyze the consequences of non identifiability for two cases: $\boldsymbol{\delta} = \mathbf{0}$ and $\boldsymbol{\delta} \neq \mathbf{0}$.

1. $\boldsymbol{\delta} = \mathbf{0}$: In this case, $\mathbf{A}_{\text{ca},0}$ is a Vandermonde matrix with M_{ca} distinct rows. One way to ensure identifiability of $\boldsymbol{\gamma}$ is to assume that it is $K-$ sparse (or, equivalently, assume the number of sources to be known). In such a case, $\mathbf{A}_{\text{ca},0}\boldsymbol{\gamma} = \mathbf{z}$ will permit a *unique solution* in $\boldsymbol{\gamma}$, if $K < \frac{\text{k-rank}(\mathbf{A}_{\text{ca}})}{2}$, (see [25]), where k-rank(.) represents the Kruskal rank of a matrix. Owing to the Vandermonde structure of $\mathbf{A}_{\text{ca},0}$, its Kruskal rank is M_{ca}. Hence, in this case, we can ensure identifiability of $\boldsymbol{\gamma}$ for $\boldsymbol{\delta} = 0$, by assuming it to be sparse and ensuring that $K < M_{\text{ca}}/2$.

2. $\boldsymbol{\delta} \neq \mathbf{0}$: Finding an explicit sufficient conditions for identifiability in this case is a nontrivial problem and can be a topic for future research. This is due to the fact that the dictionary $\mathbf{A}_{\text{ca}}$ itself is a function of $\boldsymbol{\delta}$ and it no longer has a

Vandermonde structure, which makes it very difficult to ascertain its Kruskal rank. However, assuming the perturbation to be small, we can obtain sufficient conditions for identifiability that relate $\boldsymbol{\delta}$, $\boldsymbol{\gamma}$ and the smallest singular value of the unperturbed manifold $\mathbf{A}_{\text{ca},0}$, as discussed next.

2.4.3 Sufficient conditions for identifiability for small perturbations

In this section, we derive sufficient conditions for (2.15) to hold, in terms of an upper bound for $\boldsymbol{\delta}$.

Definition 2. *For a vector $\boldsymbol{\delta} \in \mathbb{R}^M$, define $\boldsymbol{\Delta}$ as*

$$\begin{bmatrix} diag(\boldsymbol{\delta}) - \delta_1 \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & diag(\boldsymbol{\delta}) - \delta_2 \mathbf{I} & \cdots \mathbf{0} & \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & diag(\boldsymbol{\delta}) - \delta_M \mathbf{I} \end{bmatrix} \tag{2.16}$$

Also, define $\boldsymbol{\Delta}'$ by replacing the vector $\boldsymbol{\delta}$ with $\boldsymbol{\delta}' \in \mathbb{R}^M$ in (3.2).

Assuming that $\boldsymbol{\delta}$ is small we can write the linear approximation of $\mathbf{A}_{\text{ca}}(\boldsymbol{\delta})$ as

$$(\mathbf{A}_{\text{ca}}(\boldsymbol{\delta}))_{(r-1)M+s,k} \simeq e^{\jmath(d_r - d_s)\omega_k}(1 + (\delta_r - \delta_s)\jmath\omega_k),$$

for $1 \leq r, s \leq M$, which can be also written as

$$\mathbf{A}_{\text{ca}}(\boldsymbol{\delta}) \simeq \mathbf{A}_{\text{ca},0} + \boldsymbol{\Delta}\mathbf{A}_{\text{ca},0}\Upsilon \tag{2.17}$$

where $\Upsilon = \jmath \text{diag}(\omega_1, \cdots, \omega_{N_\theta})$.

For this linearized model, we now proceed to establish sufficient conditions such that (2.15) holds. We assume that the number of sources is known to be at

most K, so that all vectors $\boldsymbol{\gamma}$ in our ambiguity set are at most $K-$sparse. Suppose there exists $\boldsymbol{\delta}' \neq \boldsymbol{\delta}$ and $\boldsymbol{\gamma}' \neq \boldsymbol{\gamma}$ (both at most $K-$ sparse) such that

$$(\mathbf{A}_{\text{ca},0} + \boldsymbol{\Delta}\mathbf{A}_{\text{ca},0}\Upsilon)\boldsymbol{\gamma} = (\mathbf{A}_{\text{ca},0} + \boldsymbol{\Delta}'\mathbf{A}_{\text{ca},0}\Upsilon)\boldsymbol{\gamma}' \tag{2.18}$$

Let S, S' denote the supports of $\boldsymbol{\gamma}, \boldsymbol{\gamma}'$, respectively. Moreover, let $S_1 = S\backslash S', S_2 = S'\backslash S, S_{12} = S \cap S'$, and k_1, k_2, k_{12} be the cardinality of S_1, S_2, S_{12} respectively. Let $\tilde{\mathbf{A}}_{\text{ca},0}$ and $\tilde{\Upsilon}$ be the submatrices of $\mathbf{A}_{\text{ca},0}$ and Υ, respectively, comprised by the columns indexed by $S_1 \cup S_2$. Define $\boldsymbol{\gamma}_\mathbf{i}$ (or $\boldsymbol{\gamma}'_\mathbf{i}$) be the vector comprised by the elements of $\boldsymbol{\gamma}$ (or $\boldsymbol{\gamma}'$) which are indexed by $S_\mathbf{i}$, where $\mathbf{i} = 1, 2, 12$. We can rewrite (2.18) as

$$(\tilde{\mathbf{A}}_{\text{ca},0} + \boldsymbol{\Delta}\tilde{\mathbf{A}}_{\text{ca},0}\tilde{\Upsilon}) \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\gamma}_{12} \\ \mathbf{0} \end{bmatrix} = (\tilde{\mathbf{A}}_{\text{ca},0} + \boldsymbol{\Delta}'\tilde{\mathbf{A}}_{\text{ca},0}\tilde{\Upsilon}) \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\gamma}'_{12} \\ \boldsymbol{\gamma}'_2 \end{bmatrix}$$

which is equivalent to

$$\tilde{\mathbf{A}}_{\text{ca},0} \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\gamma}_{12} - \boldsymbol{\gamma}'_{12} \\ -\boldsymbol{\gamma}'_2 \end{bmatrix} = -\boldsymbol{\Delta}\tilde{\mathbf{A}}_{\text{ca},0}\tilde{\Upsilon} \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\gamma}_{12} \\ \mathbf{0} \end{bmatrix} + \boldsymbol{\Delta}'\tilde{\mathbf{A}}_{\text{ca},0}\tilde{\Upsilon} \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\gamma}'_{12} \\ \boldsymbol{\gamma}'_2 \end{bmatrix} \tag{2.19}$$

Let us assume that each nonzero entry of $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}'$ lie within the range $[\gamma_{\min}\ \gamma_{\max}]$. Moreover, assume that each entry of $\boldsymbol{\delta}$ and $\boldsymbol{\delta}'$ is bounded above by $\delta_{\max}$. We have

$$\|LHS\| \geq \sigma_{\min}(\tilde{\mathbf{A}}_{\text{ca},0})\sqrt{k_1}\gamma_{\min} \tag{2.20}$$

Moreover,

$$\|RHS\| \leq \sigma_{\max}(\Delta)\sigma_{\max}(\tilde{\mathbf{A}}_{\text{ca},0}\tilde{\Upsilon})\sqrt{k_1+k_{12}}\gamma_{\max} \tag{2.21}$$

$$+\sigma_{\max}(\Delta')\sigma_{\max}(\tilde{\mathbf{A}}_{\text{ca},0}\tilde{\Upsilon})\sqrt{k_2+k_{12}}\gamma_{\max} \tag{2.22}$$

in which LHS, RHS refer to the left hand side and right hand side of the equation (2.19), and $\sigma_{\min}(.), \sigma_{\max}(.)$ indicate the smallest and largest singular value of a given matrix, respectively.

Recall that $w_k = \frac{2\pi k}{N_\theta}$. Hence, $\sigma_{\max}(\tilde{\Upsilon}) < 2\pi$. We also have $\sigma_{\max}(\tilde{\mathbf{A}}_{\text{ca},0}\tilde{\Upsilon}) < 2\pi\sigma_{\max}(\tilde{\mathbf{A}}_{\text{ca},0}) < 2\pi\|\tilde{\mathbf{A}}_{\text{ca},0}\|_F < 2\pi M\sqrt{2K}$ and $\sigma_{\max}(\boldsymbol{\Delta}) \leq 2\delta_{\max}$.

Hence, a sufficient condition for identifiability is

$$\|LHS\| > \|RHS\|$$

From (2.20) and (2.21) we can say that one way to ensure $\|LHS\| > \|RHS\|$ is to have

$$\sigma_{\min}(\tilde{\mathbf{A}}_{\text{ca},0})\sqrt{k_1}\gamma_{\min} > 4\pi\delta_{\max}(\sqrt{k_1+k_{12}}+\sqrt{k_2+k_{12}})M\sqrt{2K}\gamma_{\max}$$

which is true if

$$\delta_{\max} < \frac{\sigma_{\min}(\tilde{\mathbf{A}}_{\text{ca},0})}{4\pi M}\frac{\gamma_{\min}}{\sqrt{2K}\gamma_{\max}} \tag{2.23}$$

Considering all possible supports, (2.23) is satisfied if

$$\delta_{\max} < \frac{\tilde{\sigma}_{\min}}{4\pi\sqrt{2K}M}\frac{\gamma_{\min}}{\gamma_{\max}} \tag{2.24}$$

where $\tilde{\sigma}_{\min} = \min \sigma_{\min}(\tilde{\mathbf{A}}_{\text{ca},0})$ over all submatrices $\tilde{\mathbf{A}}_{\text{ca},0}$ with $2K$ columns.

We now summarize this result as the following theorem:

Theorem 6. *Suppose* $|\delta_m| \leq \delta_{\max}$ *for* $m = 1, \cdots, M$, *and* $\delta_{\max}$ *to be small so that we can approximate* $\mathbf{A}_{ca}(\boldsymbol{\delta})$ *as* (3.1). *Moreover, assume that* $\|\boldsymbol{\gamma}\|_0 \leq K$ *and the non-zero elements of* $\boldsymbol{\gamma}$ *lie in the range* $[\gamma_{\min} \ \ \gamma_{\max}]$. *The parameters* $[\boldsymbol{\gamma}, \boldsymbol{\delta}]$ *are identifiable for any grid size* N_θ, *if the maximum perturbation value obeys* (2.24).

2.5 Numerical Studies

In this Section, we evaluate the Cramér Rao bound corresponding to the parameter $\boldsymbol{\gamma}$ in presence of perturbation $\boldsymbol{\delta}$, and compare it with the RMSE of a maximum likelihood solver. Following [1, Section 8.5], the log-likelihood function corresponding to our problem can be written as

$$L(\boldsymbol{\delta}, \boldsymbol{\gamma}) = -[\ln \det \mathbf{R}_{\mathbf{y}} + \mathrm{tr}(\mathbf{R}_{\mathbf{y}}{}^{-1} \hat{\mathbf{R}}_{\mathbf{y}})] \tag{2.25}$$

where $\mathbf{R}_{\mathbf{y}} = \mathbf{A}(\boldsymbol{\delta})\mathrm{diag}(\boldsymbol{\gamma})\mathbf{A}(\boldsymbol{\delta})^H + \sigma_w^2 \mathbf{I}$ and $\hat{\mathbf{R}}_{\mathbf{y}} = \frac{1}{L}\sum_{l=1}^{L} \mathbf{y}[l]\mathbf{y}^H[l]$ is the sample covariance matrix. We also supposed that the noise variance σ_w^2 is known and the signals follow the model defined in (2.5). We find the optimum values for $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ by maximizing $L(\boldsymbol{\delta}, \boldsymbol{\gamma})$ subject to the constraint $\boldsymbol{\gamma} \geq \mathbf{0}$, using `fmincon` function of MATLAB. Notice that the sparsity of $\boldsymbol{\gamma}$ is not utilized to obtain its estimate.

We consider different scenarios with respect to the number of grid points, number of sensors, array structure, and sparsity of the sources. Throughout the simulations, we consider three different arrays: a ULA, a nested array, and a co-prime array (with coprime numbers $N_1 = 4$ and $N_2 = 7$), all with the same number of sensors, $M = 14$ (for coprime array, the number of sensors is $M = 2N_1 + N_2 - 1$). In all cases, we assume the spatial frequencies to lie on a uniform grid with N_θ

grid points. We study the performance for different values of N_θ. The variance of the noise is fixed to be $\sigma_w^2 = 0.1$ in all the simulations. We assume the sensor perturbations to be $\boldsymbol{\delta} = \alpha \boldsymbol{\delta_0}$, where

$$\boldsymbol{\delta_0} = 0.1 \times [0, 1, 3, -1, -3, 1, -4, 2, 6, 9, -3, 4, 5, -7, 4, -1]^T,$$

and α is a scalar, which determines the strength of perturbation. We define RMSE of the maximum likelihood estimator as $\sqrt{\sum_{i=1}^{N_{\text{tests}}} \frac{\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\|^2}{N_\theta N_{\text{tests}}}}$, where $\hat{\boldsymbol{\gamma}}$ is the estimated $\boldsymbol{\gamma}$, and N_{tests} indicates the number of Monte-Carlo simulations for each value of α or L. In all the simulations, $N_{\text{tests}} = 100$. Moreover, in all the plots, CRB is computed from the trace of the Schur complement defined in (2.13).

In the first simulation, we choose $N_\theta = 35$. We study two different cases: in the first setting, we have as many sources as N_θ, all with powers equal to one. In the second case, which we refer to as the *sparse* case, we assume that there are only $K = 4$ active sources with powers equal to one and the rest are zero. In this case, the support of $\boldsymbol{\gamma}$ is given by $S = \{3, 7, 11, 16\}$. As stated earlier, the ML algorithm does not assume $\boldsymbol{\gamma}$ to be a sparse vector *apriori*.

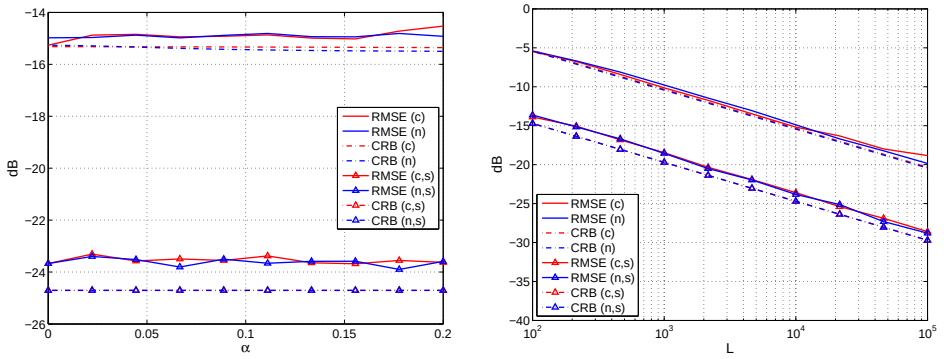
Figure 2.1 shows the performance of the ML estimate and compares it to the Cramér Rao bound for both sparse and non sparse settings. The label "n" indicates the nested array with $M = 14$ antennas, while the label "c" corresponds to the coprime array with $N_1 = 4$ and $N_2 = 7$. Plots which are marked with "s" indicate the case where $\boldsymbol{\gamma}$ is sparse with $K = 4$. We observe that as we increase the number of snapshots L, the accuracy of the ML estimator increases, and its MSE is close to the Cramér Rao bound. Moreover, the *sparse* case possesses better performance

than the non sparse case, although sparsity is not assumed beforehand.

In the second simulation, we compare the CRB of ULA and nested array for different grid sizes. Figure 2.2 depicts the result. We see that in each plot, beyond a certain grid size, the CRB suddenly jumps to very large values, indicating that FIM becomes singular beyond that point. We also observe that the corresponding value for N_θ is much smaller for the ULA than that for the nested array. This supports the fact that nested array is capable of resolving $O(M^2)$ sources even in the presence of perturbations, whereas ULA fails when $N_\theta > 2M - 1 = 27$. In this plot, $\alpha = 0.5$ and $L = 10^4$.

In the third simulation, we examine the probability of FIM being invertible for different number of sensors and grid sizes. For this experiment, we consider N_θ sources located on the grid points, all with powers equal to one. In each trial, we randomly generate a $\boldsymbol{\delta}$ whose entries are uniformly chosen from the range $[-0.5 \;\; 0.5]$ (keeping $\delta_1 = 0$). For every M and N_θ, we count the events for which $\mathbf{J}$ is invertible, and average the result over 100 runs. The result is demonstrated in Figure 2.3. The white pixels represent values of (M, N_θ) for which $\mathbf{J}$ is invertible with high probability. The blue line indicates the value of M as a function of N_θ, below which the FIM evaluated at $\boldsymbol{\delta} = \mathbf{0}$ is nonsingular. This value is computed empirically from the experiments. The red line, however, shows the theoretical bound on N_θ that we derived in Theorems 2, 3 (We used the bound proposed in Remark 3 for nested array). We see that for a ULA, the blue line and the red line match exactly, meaning that the sufficient condition that we derived in Theorem 2 is also necessary. However, there is a small gap between the red and blue lines for the nested array,

indicating the possibility of a gap between necessary and sufficient condition for non singularity of $\mathbf{J}$ at $\boldsymbol{\delta} = \mathbf{0}$. Moreover, we observe that for both ULA and nested array that there is a white area under the red and blue lines which represents the region where $\mathbf{J}\big|_{\boldsymbol{\delta}\neq\mathbf{0}}$ is invertible, although $\mathbf{J}\big|_{\boldsymbol{\delta}=\mathbf{0}}$ is not. This happens due to the fact that the perturbations can slightly increase the rank of $\mathbf{A}_{\text{ca}}(\boldsymbol{\delta})$. Therefore, $\mathbf{J}$ can be invertible even though $\mathbf{A}_{\text{ca},0}$ is not full column rank.



(a) The RMSE of recovered $\boldsymbol{\gamma}$ vs α ($L = 10^4$)

(b) The RMSE of recovered $\boldsymbol{\gamma}$ vs L ($\alpha = 0.1$).

Figure 2.1: The performance of ML compared with CRB in different cases. In these plots, "c", "n" indicate *coprime* and *nested* arrays, respectively. Moreover, "s" indicates the cases where the sources are *sparse* with $K = 5$ (without using this sparsity in solving ML problem or finding the CRB).

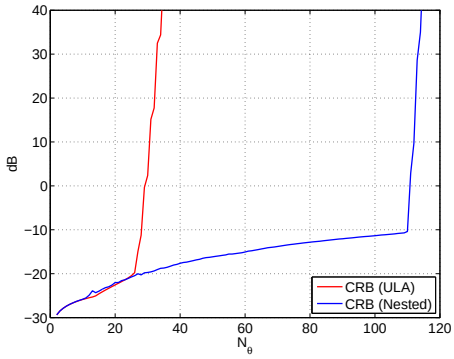
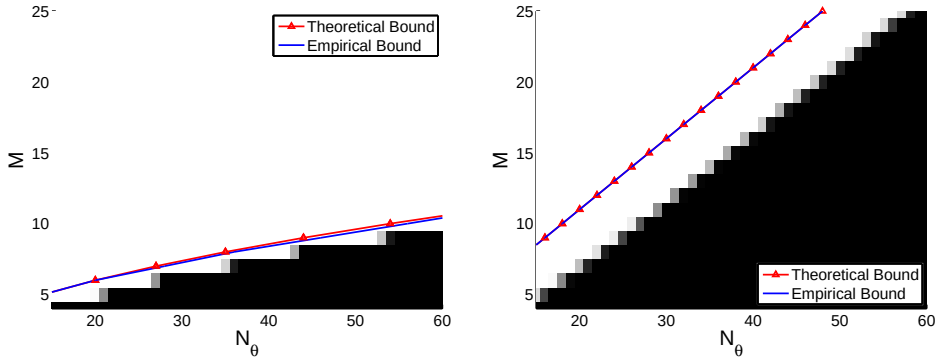


Figure 2.2: The CRB vs N_θ for nested and ULA with $M = 14$ sensors ($\alpha = 0.5, L = 10^5$).



(a) Nested Array (b) ULA

Figure 2.3: The probability of FIM being invertible over random choices of $\boldsymbol{\delta}$ for different number of sensors (M) and different grid sizes (N_θ). The blue line indicates the M above which $\mathbf{J}\big|_{\boldsymbol{\delta}=\mathbf{0}}$ is invertible. The red line shows the theoretical bounds we derived in Theorems 2, 3.

Chapter 3: Joint DOA Estimation and Self-Calibration with perturbed arrays via Bi-Affine Formulation

3.1 Introduction

In this Chapter, we demonstrate that it is possible to jointly estimate the DOAs and perturbation coefficients via a novel bi-affine formulation. We assume that the perturbations $\boldsymbol{\delta}$ are small, so that we can approximate the perturbed coarray manifold using its first order Taylor series expansion. This formulation leads to a "bi-affine" model, which is linear in source powers, and affine in the perturbation variable. We show that it is possible to recover the DOAs even in presence of the nuisance perturbation variables, via clever elimination of variables that exploit the redundancy present in co-arrays. By exploiting the pattern of repeating elements, it is possible to reduce the bi-affine problem to a linear underdetermined (sparse) problem in source powers, which can be efficiently solved using ℓ_1 minimization. We establish precise conditions under which such reduction is possible, for both ULA and a robust version of coprime arrays.

It is worth mentioning that co-array redundancies have been previously shown to be useful in order to resolve phase/gain ambiguity [26]. Their approach builds on

and extends the method in [27], which was originally proposed for a ULA and which requires the number of sources to be less than the number of sensors. However, in this chapter, we consider perturbations in sensor locations, which gives rise to a signal model. Moreover, unlike [26], our method works for non uniform arrays and allows the number of sources to be greater than the number of sensors.

The contents of this chapter have been published in [12].

3.2 Bi-Affine Formulation

3.2.1 The Bi-Affine Model

In this Section, we derive a bi-affine model from the covariance matrix (2.3) corresponding to a sensor array with perturbed locations. Our main assumption is that $\boldsymbol{\delta}$ is small enough so that we can approximate the coarray manifold $\mathbf{A}_{\text{ca}}(\boldsymbol{\delta})$ using the first order Taylor series expansion as follows:

$$(\mathbf{A}_{\text{ca}}(\boldsymbol{\delta}))_{(m'-1)M+m,i} \simeq e^{\jmath(d_m-d'_m)\omega_i}(1+(\delta_m-\delta_{m'})\jmath\omega_i),$$

which can be also written in the matrix form as

$$\mathbf{A}_{\text{ca}}(\boldsymbol{\delta}) \simeq \mathbf{A}_{\text{ca},0} + \boldsymbol{\Delta}\mathbf{A}_{\text{ca},0}\Upsilon \tag{3.1}$$

where $\mathbf{A}_{\text{ca},0}$ denotes the unperturbed co-array manifold, $\Upsilon = \jmath\text{diag}(\omega_1, \cdots, \omega_{N_\theta})$, and $\boldsymbol{\Delta} \in \mathbb{R}^{M^2}$ is given as

$$\begin{bmatrix} \text{diag}(\boldsymbol{\delta}) - \delta_1\mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \text{diag}(\boldsymbol{\delta}) - \delta_2\mathbf{I} & \cdots \mathbf{0} & \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \text{diag}(\boldsymbol{\delta}) - \delta_M\mathbf{I} \end{bmatrix} \tag{3.2}$$

Therefore, (2.4) can be approximated as

$$\mathbf{z} = (\mathbf{A}_{\text{ca},0} + \boldsymbol{\Delta}\mathbf{A}_{\text{ca},0}\Upsilon)\boldsymbol{\gamma} + \sigma_w^2\text{vec}(\mathbf{I}), \tag{3.3}$$

To suppress the effect of noise, we discard the 0th lag of the co-array, and only keep the elements of $\mathbf{z}$ corresponding to the positive half of the co-array to obtain [1]:

$$\mathbf{z}^u = (\mathbf{A}^u_{\text{ca},0} + \boldsymbol{\Delta}^u\mathbf{A}^u_{\text{ca},0}\Upsilon)\boldsymbol{\gamma} \tag{3.4}$$

where $\mathbf{A}^u_{\text{ca},0}$ is the unperturbed co-array manifold with rows corresponding to the positive lags, where we retain repeated rows (that correspond to the same lag in the virtual array). The matrix $\boldsymbol{\Delta}^u$ is constructed from $\boldsymbol{\Delta}$ by retaining only the rows corresponding to those of $\mathbf{A}^u_{\text{ca},0}$. Notice that unlike the approach in [2, 3], we also keep the repeated rows of $\mathbf{A}^u_{\text{ca},0}$, since their corresponding rows in $\mathbf{z}^u$ may not be repeated due to the presence of perturbations. Hence, this redundancy in the rows of $\mathbf{A}^u_{\text{ca},0}$ can help us to get more information on the perturbations.

[1] For ease of exposition, we only consider the positive half of the co-array and demonstrate how to eliminate the spurious variable $\boldsymbol{\delta}$. However, with straightforward modifications to the proposed technique, it is also possible to incorporate the negative half to generate more augmented equations and use them for eliminating $\boldsymbol{\delta}$.

Inspired by the so-called *bilinear* model in the literature, which arises in various problems such as the blind gain and phase calibration (BGPC) problem [28], we call the model given in (3.4) a *"bi-affine"* model, since $\mathbf{z}^u$ is a linear function of $\boldsymbol{\gamma}$ and affine function of $\boldsymbol{\delta}$. Given the covariance matrix $\mathbf{R}_{\mathbf{y}}$ (or equivalently $\mathbf{z}^u$), the goal is to recover the sparse vector $\boldsymbol{\gamma}$ from the bi-affine model defined in (3.3).

3.3 Source Localization: Bi-Affine to Linear Transformation

Under the grid-based model, the DOAs can be estimated from the support of the sparse vector $\boldsymbol{\gamma}$ that is a solution to the bi-affine system of equations (3.4). In general, (3.4) can admit multiple solutions in the variables $(\boldsymbol{\delta}, \boldsymbol{\gamma})$. While the column rank of a matrix describing a linear system of equations determines if it admits a unique solution, to the best of our knowledge, no such general condition exists for a bi-affine (or even bi-linear) system which can be used as a test for existence of unique solution.

In this Section, we will derive a transformation such that *we can extract a linear system of equations in the variable* $\boldsymbol{\gamma}$, from (3.4) by eliminating the variable $\boldsymbol{\delta}$. In particular (3.4) can be reduced to an underdetermined linear system of equations of the form

$$\mathbf{h} = G\boldsymbol{\gamma} \tag{3.5}$$

where $\mathbf{G}$ is a fat matrix, whose size and structure depends on the array geometry. Hence, the bi-affine system of equations will indeed admit a unique solution in $\boldsymbol{\gamma}$ (although, not necessarily in $\boldsymbol{\delta}$) if (3.5) yields a unique $K-$sparse solution. This

transformation of the bi-affine problem into a linear problem will be shown to be possible under appropriate conditions on the array geometry, M (number of sensors), and K (number of sources).

3.3.1 Elimination of Variables Using Co-Array Redundancies

We derive the aforementioned transformation for two different array geometries: uniform linear array (ULA), and a robust version of coprime array (introduced later). The basic idea is to use the pattern of repeated elements in the unperturbed co-array manifold which is specified by the weight function $w(k)$ to equate certain elements of $\mathbf{R_y}$, thereby eliminating $\boldsymbol{\delta}$.

Let R_{mn} denote the (m,n)th element of $\mathbf{R_y}$. Given an integer $k \in S_{\text{ca}}$, we define the following notations:

$$f_k := \sum_{i=1}^{N_\theta} e^{\jmath k\omega_i}\gamma_i \tag{3.6}$$

$$\lambda_k := \sum_{i=1}^{N_\theta} e^{\jmath k\omega_i}\jmath\omega_i\gamma_i \tag{3.7}$$

Using these notations, we can rewrite (3.4) as

$$R_{mn} = f_k + \lambda_k(\delta_m - \delta_n) \tag{3.8}$$

where $1 \leq m,n \leq M$, and $k = d_m - d_n$, $k > 0$.

Notice that f_k and λ_k are themselves linear functions of the unknown sparse vector $\boldsymbol{\gamma}$. If the lag k in the co-array S_{ca} repeats at least twice, (i.e. $w(k) \geq 2$), then this redundancy can be exploited to eliminate variables as follows. If $w(k) \geq 2$, we

must have $d_i - d_j = d_m - d_n = k$, for some $1 \leq i, j, m, n \leq M$. In this case, we have

$$R_{ij} = f_k + \lambda_k(\delta_i - \delta_j)$$

$$R_{mn} = f_k + \lambda_k(\delta_m - \delta_n),$$

The variable f_k can be easily eliminated by subtracting these equations, leading to

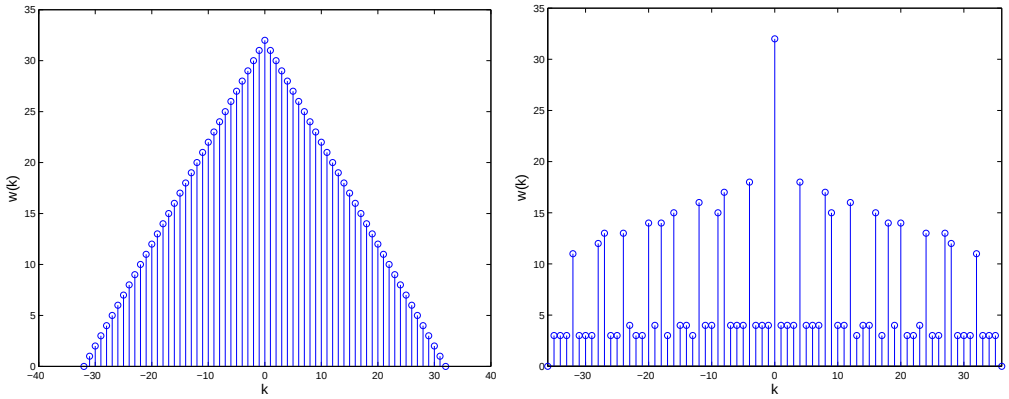
$$R_{ij} - R_{mn} = \lambda_k(\delta_i - \delta_j - \delta_m + \delta_n).$$

A very similar idea can be used (with some additional computations) to eliminate $\boldsymbol{\delta}$ from the M^2 equations of the form (3.8) for both ULA and a robust version of the co-array. Recall that co-array redundancies are also used to calibrate sensors with unknown gain and phase errors [26]. However, since we are concerned with sensor position errors, our signal model fundamentally differs from that considered in [26]. Consequently, we cannot exploit the co-array redundancies in the same way as done in [26]. We need to adopt a slightly more involved approach to eliminate the undesirable variable $\boldsymbol{\delta}$, the details of which depend on the geometry of the physical array.

3.3.2 Uniform Linear Array

In a ULA, we have $d_m - d_n = m - n$. Hence, we can rewrite the equation (3.8) as

$$R_{mn} = f_{m-n} + \lambda_{m-n}(\delta_m - \delta_n) \tag{3.9}$$



(a) ULA (b) Robust-coprime

Figure 3.1: The weight functions corresponding to ULA and robust-coprime array for $M = 32$, $N_1 = 4, N_2 = 9$.

For a given $1 \leq k \leq M-2$, define

$$\bar{r}_k = \sum_{j=2}^{k+1} R_{j,j-1} \tag{3.10}$$

$$\beta_k = \frac{R_{k+2,2} - R_{k+1,1}}{R_{k+2,k+1} - R_{2,1}} \tag{3.11}$$

The following theorem summarizes our main result for the ULA:

Theorem 7. *For a Uniform Linear Array (ULA) containing M antennas with perturbed locations, the bi-affine model (3.4) derived from the signal covariance matrix can be reduced to the form*

$$\mathbf{Cf} = \mathbf{h} \tag{3.12}$$

where $\mathbf{f} \in \mathbb{C}^{M-2} = [f_1, f_2, \cdots, f_{M-2}]^T$ and for every $1 \leq k \leq M-2$, the elements

of $\mathbf{C} \in \mathbb{C}^{(M-3)\times(M-2)}$ *and* $\mathbf{h} \in \mathbb{C}^{M-2}$ *are given by*

$$C_{k-1,1} = -k\beta_k \tag{3.13}$$

$$C_{k-1,k} = 1 \tag{3.14}$$

$$h_{k-1} = R_{k+1,1} - \beta_k \bar{r}_k \tag{3.15}$$

for every $2 \leq k \leq M-2$. *The remaining elements of* $\mathbf{C}$ *are zero. This transformation holds for almost all* $\boldsymbol{\delta} \in \mathbb{R}^M$.

Proof. The proof can be found in A.4. □

Remark 5. *Notice that the elements of* $\mathbf{f}$ *are linear functions of the sparse vector* $\boldsymbol{\gamma}$. *In particular, from* (3.6), *we have*

$$\mathbf{f} = \mathbf{B}_U \boldsymbol{\gamma}$$

where the elements of $\mathbf{B}_U \in \mathbb{C}^{M-2,N_\theta}$ *are given by* $[\mathbf{B}_U]_{m,i} = e^{jm\omega_i}$. *Therefore* (3.12) *can be written as the following system of underdetermined equations (since* $N_\theta \gg M$*)*

$$\mathbf{h} = \mathbf{C}\mathbf{B}_U \boldsymbol{\gamma} \tag{3.16}$$

This system can admit a unique sparse solution, representing the true source powers, if the Kruskal Rank of $\mathbf{C}\mathbf{B}_U$ *is at least* $2K$. *Since kruskal-rank* $(\mathbf{C}\mathbf{B}_U) \leq M-3$, *this implies* $\|\boldsymbol{\gamma}\|_0 = K < (M-3)/2$ *is a sufficient condition for the true sparse* $\boldsymbol{\gamma}$ *to be the unique solution. In practice, by exploiting the fact that* $\boldsymbol{\gamma}$ *is non negative, a larger number of sources may be uniquely recovered. We will study the phase-transition behavior of* ℓ_1 *minimization algorithms to solve* (3.16) *to determine such an empirical relation between* K *and* M.

Remark 6. *The matrix* $\mathbf{C}$ *and the vector* $\mathbf{h}$*, are only functions of the elements of covariance matrix* $\mathbf{R_y}$*, and are not explicit functions of the unknown parameter* $\boldsymbol{\delta}$*. The constructive proofs in Appendices A.4 and A.5, demonstrate the details of eliminating the variable* $\boldsymbol{\delta}$ *from our bi-affine equations.*

3.3.3 Robust Coprime Array

In order to eliminate $\boldsymbol{\delta}$ from the equations (3.4), we need to have $w(k) > 1$. However, in the original coprime array, $w(k) = 1$ for some values of k. Hence, we cannot apply our simplifications. Hence, we will consider an extended version of coprime array, defined as follows.

Definition 3 (Robust Coprime Array)**.** *A robust coprime array contains* $M = 4N_1 + 2N_2 - 2$ *sensors, whose normalized locations are given by*

$$\{(i-1)N_1, 1 \leq i \leq 2N_2 + 1\} \cup \{(j-1)N_2, 1 \leq j \leq 4N_1\}$$

In other words, we extend the usual coprime array by doubling the number of sensors in each ULA. Therefore, we have $M = 4N_1 + 2N_2 - 2$ sensors in the robust coprime array. By adding these additional sensors, we are ensured that every lag between $-N_1N_2, ..., N_1N_2$ is repeated at least twice. We will use these repetitions to resolve the unknown perturbations. Moreover, we will discard the lags beyond the aforementioned range. In the sequel, we will denote $M'_{\text{ca}} = N_1N_2$, representing the number of positive integers in this range.

Notations and Definitions: For the ease of notation, assume every variable with superscript $(.)^{(1)}$ to be associated with the first sub-array (with spacing N_1), and

every variable with superscript $(.)^{(2)}$ to be associated with the second sub-array (with spacing N_2). Therefore, we have $d_i^{(1)} = (i-1)N_1, d_j^{(2)} = (j-1)N_2$, where $1 \leq i \leq 2N_2 + 1$, and $1 \leq j \leq 4N_1$. Moreover, let $R_{ij}^{(12)}$ denote the covariance between the received signal on the ith sensor of first ULA and jth sensor of the second ULA, and $R_{ij}^{(21)} = (R_{ij}^{(12)})^*$. Similarly, $R_{ii'}^{(1)}$ (resp. $R_{jj'}^{(2)}$) denotes the covariance between the received signal on the first (resp. second) sub-array on its ith and i'th (resp. jth and j'th) sensors. This indexing also holds for $\delta_i^{(1)}, \delta_j^{(2)}$. We also define several quantities which will be later used to state our main theorem on robust coprime array. Firstly, define the quantities

$$\bar{r}_i^{(1)} := \sum_{i'=2}^{i} R_{i',i'-1}^{(1)} \tag{3.17}$$

$$\bar{r}_j^{(2)} := \sum_{j'=2}^{j} R_{j',j'-1}^{(2)}. \tag{3.18}$$

$$\alpha := \frac{2\bar{r}_{N_1+1}^{(2)} - \bar{r}_{2N_1+1}^{(2)}}{2\bar{r}_{N_2+1}^{(1)} - \bar{r}_{2N_2+1}^{(1)}} \tag{3.19}$$

We also define the vectors $\boldsymbol{\beta}_{cp} \in \mathbb{C}^{M'_{\text{ca}}}$ and $\mathbf{h}_{cp} \in \mathbb{C}^{M'_{\text{ca}}}$ as follows. From the properties of the difference set of a coprime array, each index k in the range $1 \leq k \leq M'_{\text{ca}}$ is necessarily of one of the following four forms:

$$k = \begin{cases} d_i^{(1)} - d_j^{(2)}, & 1 \leq i \leq N_2, 1 \leq j \leq 2N_1 \\ d_j^{(2)} - d_i^{(1)}, & 1 \leq i \leq N_2, 1 \leq j \leq 2N_1 \\ d_i^{(1)} - d_{i'}^{(1)}, & 1 \leq i, i' \leq N_2 \\ d_j^{(2)} - d_{j'}^{(2)}, & 1 \leq j, j' \leq 2N_1 \end{cases}$$

In other words, k can either be a self-difference of sensors belonging to the same sub-array, or a cross-difference of sensors belonging to different sub-arrays. If k happens

to be both a self-difference and a cross-difference, we consider k as a self-difference within its corresponding sub-array. Moreover, if k happens to be a cross-difference of type $k = d_i^{(1)} - d_j^{(2)}$ for some i, j, and also we have $k = d_{\hat{j}}^{(2)} - d_{\hat{i}}^{(1)}$ for some other $\hat{i}$ and $\hat{j}$, we consider it as a cross-difference of the former type.

The elements of $\boldsymbol{\beta}_{cp}$ and $\mathbf{h}_{cp}$ are then given as follows. Here, the indices i, i', j, and j' vary over the ranges $1 \leq i, i' \leq N_2, 1 \leq j, j' \leq 2N_1$, and $\bar{i} = i + N_2, \bar{j} = j + N_1$.

$$[\boldsymbol{\beta}_{cp}]_k = \begin{cases} \frac{\alpha(R_{\bar{i}\bar{j}}^{(12)} - R_{ij}^{(12)})}{\alpha(\bar{r}_{\bar{i}}^{(1)} - \bar{r}_i^{(1)} - \bar{r}_{N_2+1}^{(1)}) - (\bar{r}_{\bar{j}}^{(2)} - \bar{r}_j^{(2)} - \bar{r}_{N_1+1}^{(2)})}, & k = d_i^{(1)} - d_j^{(2)} \\ -\frac{\alpha(R_{\bar{i}\bar{j}}^{(12)} - R_{ij}^{(12)})^*}{\alpha(\bar{r}_{\bar{i}}^{(1)} - \bar{r}_i^{(1)} - \bar{r}_{N_2+1}^{(1)}) - (\bar{r}_{\bar{j}}^{(2)} - \bar{r}_j^{(2)} - \bar{r}_{N_1+1}^{(2)})} & k = d_j^{(2)} - d_i^{(1)} \\ \frac{R_{k+2,2}^{(1)} - R_{k+1,1}^{(1)}}{R_{k+2,k+1}^{(1)} - R_{2,1}^{(1)}} \quad k \neq N_2, & k = d_i^{(1)} - d_{i'}^{(1)} \\ \frac{R_{k+2,2}^{(2)} - R_{k+1,1}^{(2)}}{R_{k+2,k+1}^{(2)} - R_{2,1}^{(2)}} \quad k \neq N_1, & k = d_j^{(2)} - d_{j'}^{(2)} \end{cases} \tag{3.20}$$

$$[\mathbf{h}_{cp}]_k = \begin{cases} R_{ij}^{(12)} - \beta_k \bar{r}_i^{(1)} + \alpha^{-1} \beta_k \bar{r}_j^{(2)}, & k = d_i^{(1)} - d_j^{(2)} \\ (R_{ij}^{(12)})^* + \beta_k \bar{r}_i^{(1)} - \alpha^{-1} \beta_k \bar{r}_j^{(2)} & k = d_j^{(2)} - d_i^{(1)} \\ R_{i+1,1}^{(1)} - \beta_k \bar{r}_i^{(1)} \quad k \neq N_2, & k = d_i^{(1)} - d_{i'}^{(1)} \\ R_{j+1,1}^{(2)} - \beta_k \bar{r}_j^{(2)} \quad k \neq N_1, & k = d_j^{(2)} - d_{j'}^{(2)} \end{cases} \tag{3.21}$$

Furthermore, $[\mathbf{h}_{cp}]_{N_1} = [\mathbf{h}_{cp}]_{N_2} = 0$. Based upon the above definitions, let us also define a matrix $\mathbf{C}_{cp} \in \mathbb{C}^{M'_{\text{ca}} \times M'_{\text{ca}}}$ such that $[\mathbf{C}_{cp}]_{k,k} = 1, 1 \leq k \leq M'_{\text{ca}}, k \neq N_1, N_2$, and $[\mathbf{C}_{cp}]_{N_1,N_1} = [\mathbf{C}_{cp}]_{N_2,N_2} = 0$. Its remaining entries satisfy:

$$[\mathbf{C}_{cp}]_{k,N_1} = \begin{cases} -(i-1)\beta_k, & k = d_i^{(1)} - d_j^{(2)} \\ (i-1)\beta_k, & k = d_j^{(2)} - d_i^{(1)} \\ -(i-1)\beta_k, & k = d_i^{(1)} - d_{i'}^{(1)}, k \neq N_1 \end{cases} \tag{3.22}$$

$$[\mathbf{C}_{cp}]_{k,N_2} = \begin{cases} \alpha^{-1}(j-1)\beta_k, & k = d_i^{(1)} - d_j^{(2)} \\ -\alpha^{-1}(j-1)\beta_k, & k = d_j^{(2)} - d_i^{(1)} \\ -(j-1)\beta_k, & k = d_j^{(2)} - d_{j'}^{(2)}, k \neq N_2 \end{cases} \tag{3.23}$$

Equipped with the above definitions, we state our main result on robust coprime arrays as the following theorem:

Theorem 8. *For a robust coprime array, the bi-affine formulation (3.4) can be reduced to* $\mathbf{C}_{cp}\mathbf{f}_{cp} = \mathbf{h}_{cp}$*, where* $\mathbf{f}_{cp} = [f_1, f_2, \cdots, f_{M'_{ca}}]^T$ *and* $\mathbf{C}_{cp} \in \mathbb{C}^{M'_{ca} \times M'_{ca}}, \mathbf{h}_{cp} \in \mathbb{C}^{M'_{ca}}$ *are previously defined. This transformation holds for almost all* $\boldsymbol{\delta} \in \mathbb{R}^M$*.*

Proof. The proof can be found in the A.5. □

Remark 7. *Recall that* $\mathbf{f}_{cp} = \mathbf{B}_{cp}\boldsymbol{\gamma}$*, where the elements of* $\mathbf{B}_{cp} \in \mathbb{C}^{M'_{ca}, N_\theta}$ *are given by* $[\mathbf{B}_{cp}]_{m,i} = e^{jm\omega_i}$*. Hence, from Theorem 8, we obtain* $\mathbf{C}_{cp}\mathbf{B}_{cp}\boldsymbol{\gamma} = \mathbf{h}_{cp}$*, which can admit a unique sparse solution in* $\boldsymbol{\gamma}$ *if* $\mathbf{C}_{cp}\mathbf{B}_{cp}$ *has Kruskal rank of* $O(M'_{ca})$*. In future, we will characterize the exact Kruskal rank of* $\mathbf{C}_{cp}\mathbf{B}_{cp}$*. However, in Sec. 4.3, we experimentally show that* ℓ_1 *minimization can resolve larger number of sources for coprime arrays, compared to ULA.*

3.4 Computational Complexity of the Proposed Algorithm

In this Section, we investigate the computational complexity of the proposed algorithm. Our proposed algorithm consists of two stage. In the first stage, we compute the vector $\mathbf{h}$, and the matrix $\mathbf{C}$. We consider the ULA and coprime arrays

separately. In the second stage, a linear sparse problem is solved using the obtained dictionaries from the first stage.

3.4.1 Computational Complexity for the Reduction Stage

3.4.1.1 Computational Complexity for ULA

We can calculate each one of the β_k's in $\mathcal{O}(1)$. Hence, calculating all the β_k's requires $\mathcal{O}(M)$ time. Moreover, computation of $\bar{r}_i$ only involves a summation. Hence, calculating all the $\bar{r}_i$, for $i = 1, \cdots, M$, requires $\mathcal{O}(M)$ time. Therefore, we can compute the matrices $\mathbf{C}$, and $\mathbf{h}$ in $\mathcal{O}(M)$. Finally, since the matrix $\mathbf{C}$ is very sparse and has only two nonzero elements in each row, the product $\mathbf{C}\mathbf{B}_U$ needs the time $O(N_\theta M)$. Therefore, the overall algorithm requires the time $\mathcal{O}(M) + \mathcal{O}(N_\theta M) = \mathcal{O}(N_\theta M)$.

3.4.1.2 Computational Complexity for Robust Coprime Array

In this case, we can calculate $\bar{r}_i^{(1)}$, and $\bar{r}_i^{(2)}$ in $\mathcal{O}(N_1 + N_2) = \mathcal{O}(M)$. Moreover, each element of β_k can be again calculated in $\mathcal{O}(1)$. However, since k can be as large as $N_1 N_2$, it will require the time $\mathcal{O}(N_1 N_2)$ to compute all of the β_k's. Therefore, computing $\mathbf{h}_{cp}$ and $\mathbf{C}_{cp}$ can be done in $\mathcal{O}(N_1 N_2)$. Finally, since the matrix $\mathbf{C}_{cp}$ has $\mathcal{O}(1)$ nonzero elements in each row, computation of the product $\mathbf{C}_{cp}\mathbf{B}_{cp}$ requires the time $\mathcal{O}(N_\theta N_1 N_2)$. Therefore, the overall computations needs $\mathcal{O}(N_1 N_2) + \mathcal{O}(N_\theta N_1 N_2)$ time.

3.4.2 The Overall Computational Complexity

After computing the matrices $\mathbf{C}, \mathbf{h}$, we only need to solve the linear sparse problem. This stage completely depends on the algorithm that we choose. If we use ℓ_1 minimization, it will need $\mathcal{O}\left((N_\theta)^3\right)$ computations. It is also possible to use faster algorithms like Orthogonal Matching Pursuit (OMP), which needs $\mathcal{O}(N_\theta K^2)$ computations [29]. Therefore, the overall computational complexity will be depend on the algorithm that we choose for solving the sparse problem. If we use ℓ_1 minimization, the overall needed time will be $\mathcal{O}\left((N_\theta)^3\right)$, and if we use OMP, the algorithm will need at most $\mathcal{O}\left(N_\theta(K^2+M^2)\right)$ computations.

3.5 Simulations

In this section, we conduct two different sets of numerical experiments to validate our theoretical claims. In all the simulations, we assume $N_\theta = 200$ points on the grid. The DOAs are chosen uniformly between -60° and 60°, and assigned to the closest point on the grid. The perturbations are assumed to be $\boldsymbol{\delta} = \alpha\boldsymbol{\delta}_0$ (Notice that, following the model given in Section 2.2.1, the sensor locations and the perturbations are normalized with respect to half of the wavelength $\lambda/2$). In all the simulations, $\boldsymbol{\delta}_0$ is a fixed vector with $|\boldsymbol{\delta}|_\infty \leq 0.5$, and is drawn from a uniform distribution.

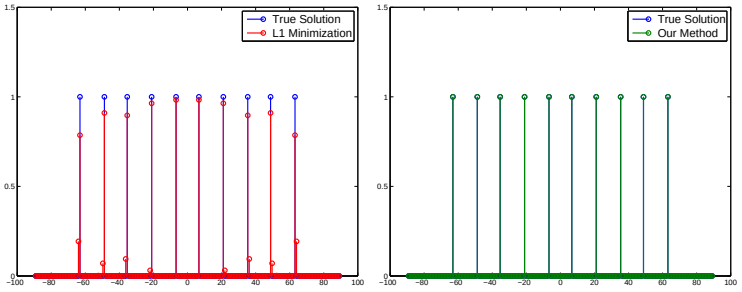
In the first set of simulations, we assume that we know the covariance matrix $\mathbf{R}_\mathbf{y}$, and that the model defined in Sec. 3.2 holds exactly. In this case, we use the approach proposed in the proofs of Theorems 7, and 8 to eliminate the perturba-

tions and recover the source powers. We compare our method against running ℓ_1 minimization on the covariance matrix, assuming that the coarray manifold is unperturbed (which will lead to a basis mismatch). In other words, we compare our approach with the solution of the following problem:

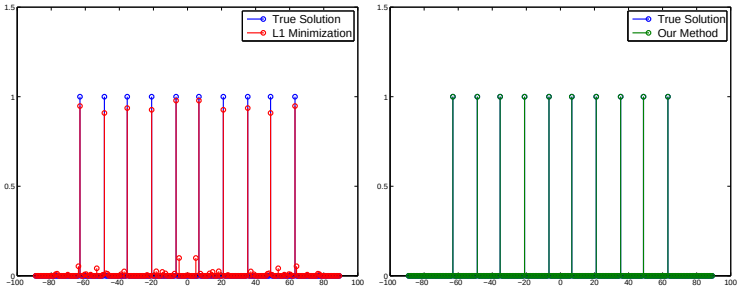
$$\min_{\boldsymbol{\gamma}} \|\boldsymbol{\gamma}\|_1 \quad \text{s.t.} \quad \mathbf{A}^u_{\text{ca},0}\boldsymbol{\gamma} = \mathbf{z}^u$$

As demonstrated in Fig. 3.2, our approach as described in the proofs of Theorems 1 and 2, exactly recovers the true supports for both ULA and coprime arrays when $K = 10 < M = 32$ (The blue line corresponding to the true solution and green line corresponding to our method, match exactly). When $K = 35 > M$, ULA cannot recover the true DOAs, while the robust coprime array perfectly identifies the support (compare Fig. 3.2(f) and Fig. 3.2(h), where $\alpha = 0.5$). In the third experiment, we will empirically study the relationship between K and M (for both ULA and coprime) that ensures perfect recovery of DOAs in the form of a phase transition diagram.

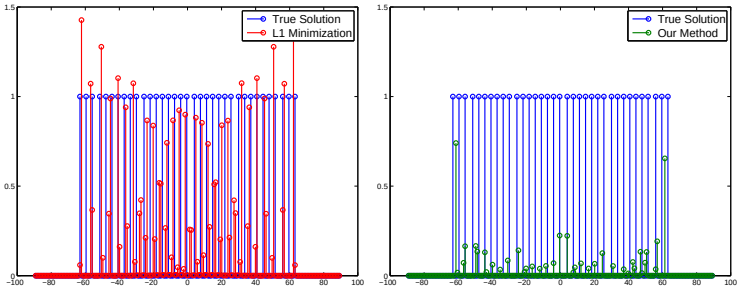
In the second set of experiments, we study the phase transition diagram of ℓ_1 minimization algorithm applied on the linear underdetermined system of equations obtained from Theorems 7, 8. In these simulations, we assume that the covariance matrix is known exactly, i.e., we have infinite number of snapshots. We consider a trial successful if $\|\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}}\|_F \leq \epsilon$, where $\hat{\boldsymbol{\gamma}}$ is the recovered vector of powers, and $\epsilon = 10^{-3}$. The white pixels in the plots of Fig. 3.3 show the problem settings under which performing ℓ_1 minimization on the linear system derived in Theorems 7, and 8 can always recover the true solution. We simulate each case 100 times



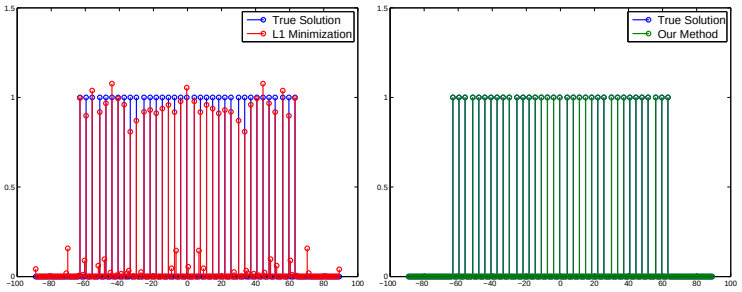
(a) ULA, $K = 10, M = 32$ (b) ULA, $K = 10, M = 32$



(c) Coprime,$K = 10, M = 32$ (d) Coprime,$K = 10, M = 32$

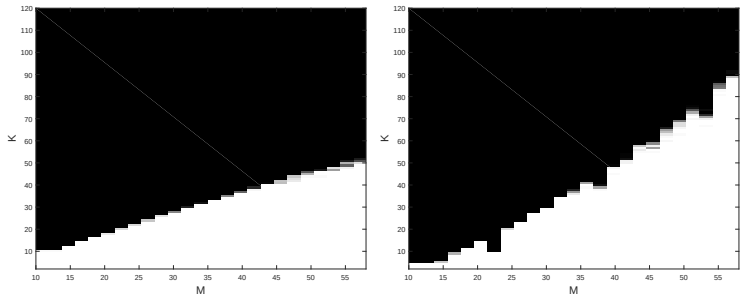


(e) ULA, $K = 35, M = 32$ (f) ULA, $K = 35, M = 32$



(g) Coprime,$K = 35, M = 32$ (h) Coprime,$K = 35, M = 32$

Figure 3.2: Recovered powers using the approach proposed in Theorems 7, 8. In each plot, X-axis shows the directions on the grid in degrees, and Y-axis shows the power corresponding to each direction on the grid.



(a) ULA (b) Coprime

Figure 3.3: Phase transition plots.

and show the probability of success with a gray-scale pixel. In these plots, we only show the cases where M is an even number (because the robust-coprime array can only have even number of sensors). Moreover, for each M, we find N_1, N_2 such that $M = 4N_1 + 2N_2 - 2$ and N_1N_2 is the maximum possible number. For coprime arrays, we find that even in presence of perturbation, white pixels exist in the region where number of sources is greater than number of sensors. This shows that robust coprime array is capable of resolving more sources than the number of sensors, even in the presence of perturbations.

3.6 Conclusion

In this Chapter, we investigated the robustness of coprime arrays to unknown perturbations on the locations of sensors. We assumed that the perturbations are small and developed a bi-affine model in terms of the unknown perturbations and the source powers. We used the redundancies of the difference coarray to eliminate

the nuisance variables, and reduce the bi-affine problem to a linear underdetermined (sparse) problem in source powers, which can be solved using ℓ_1 minimization. We derived this reduction for both ULA and a robust version of coprime arrays. Our simulations showed that if the ideal covariance matrix of the received signals is available, the source powers can be accurately recovered using our proposed approach, thereby validating the theoretical claims in Theorems 7 and 8. We also showed the region (in terms of K and M) under which the bi-affine problem has a unique solution, in the form of a phase transition diagram.

Chapter 4: Performance Analysis of Temporal Sampling

4.1 Introduction

In this Chapter we study the problem of temporal sampling, and examine its robustness to nonideal conditions such as sampling jitter and finite number of snapshots.

The effect of jitter has been extensively studied in the context of uniform sampling [30–32]. Jitters usually occur due to inaccuracies of the system clock of A/D converters at high frequencies. A desirable sampling technique would be one that is tolerant to jitters as well as to additive noise so that small jitters or noise would lead only to relatively small reconstruction errors. This is also known as stability of sampling [33].

However, the existing studies about sampling jitter and perturbations cannot be applied to the co-prime sampling, This is because, in co-prime sampling the goal is to reconstruct the autocorrelation sequence whereas in typical sampling problems the goal is to reconstruct the time domain signal.

The materials presented in this chapter have been published in [10].

4.2 Temporal Coprime Sampling and Effect of Jitter

The guarantees of coprime sampling [3] (see Section 1.2.1 for a short review) are shown to hold for ideal conditions such as long observation time for statistical averaging, and sampling is performed without jitter. We now examine the robustness of coprime sampling to these non ideal conditions and establish how accurately the true autocorrelation sequence can be reconstructed in each case.

4.2.1 Studying the Effect of Finite Samples on Correlation Estimates

Let $R_x(\tau)$ denote the autocorrelation function (ACF) of a wide-sense stationary (WSS) random process $x(t)$. The signal $x(t)$ is sampled with a pair of coprime samplers at the rate $1/MT$ and $1/NT$ $(M < N)$ to obtain the samples $x_M[n] = x(MnT)$ and $x_N[m] = x(NmT)$, where $1/T$ corresponds to the Nyquist rate determined by the power spectrum density of $x(t)$. It can be shown [3] that $\{E[x_M[n]x_N[m], 0 \leq n \leq N-1, 0 \leq m \leq 2M-1\}$ generates correlation values $R_x(kT)$ for all lags $0 \leq k \leq MN-1$. Hence, it is possible to obtain samples of the autocorrelation function at the Nyquist rate $(1/T)$ by using sub-Nyquist samplers operating at M and N times slower than the Nyquist rate.

In this section, we study the effect of computing the correlation with finite samples. The signals obtained from the coprime samplers are:

$$x_1[n,l] = x\left(nMT + 2MNlT\right) \tag{4.1}$$

$$x_2[m,l] = x\left(mNT + 2MNlT\right) \tag{4.2}$$

For the purpose of our analysis we make the following assumptions:

1. **(A1):** $x(nT)$ is assumed to be a zero mean jointly Gaussian WSS process whose autocorrelation $R_x(kT)$ is assumed to be zero for $|k| \geq 2MN$. For instance, a moving-average (MA) process with order less than or equal to $2MN$ would satisfy this criterion. In particular, this implies $x(nT)$ and $x(nT+2MNlT+k)$ are independent for $l, k \neq 0$ (since these variables are jointly Gaussian and uncorrelated).

2. **(A2):** For each n, the random variables $\{x(nT+k), |k| < 2MN\}$ are jointly Gaussian with correlation coefficients given by $R_x(kT), |k| \leq 2MN$.

We will suppress T in our notations. Assumptions (A1) and (A2) imply that $x_1[n,l]$ and $x_2[n,l]$ are jointly Gaussian for each n, l satisfying $E(x_1[n,l]x_2[m,l]) = \mathbf{R_x}(Nm-Mn)$. In practice however, we estimate the autocorrelation sequence using L such observations as

$$\hat{R}_x(Nm-Mn) = \frac{1}{L}\sum_{l=1}^{L} x_1[n,l]x_2[m,l] \tag{4.3}$$

with $0 \leq n \leq N-1$, and $0 \leq m \leq 2M-1$. Our goal is to understand how *perturation and effect of finite samples* jointly influence the estimation of the correlation values using coprime samplers. The following theorem explicitly characterizes such an effect:

Definition 4. *Define the functions ζ_i with respect to parameters a, ρ as*

$$\zeta_i(a,\rho) = \frac{\sqrt{2}|c_i|}{\sqrt{-\rho' + \sqrt{d_i^2 + 4\rho' c_i a}}} e^{-\frac{\sqrt{d_i^2+4\rho' c_i a}-d_i}{2\rho'}}, \tag{4.4}$$

in which i can be either $+$ or $-$, and $c_+ = \rho + a, c_- = a - \rho, d_+ = 1 + \rho^2 + 2a\rho, d_- = 1 + \rho^2 - 2a\rho, \rho' = 1 - \rho^2$.

Theorem 9. *Under assumptions (A1) and (A2) on $x(t)$, for each $k = Nm - Mn, 0 \leq n \leq 2M-1, 0 \leq m \leq N-1$, the autocorrelation $\hat{R}_x[k]$ estimated using L samples, differs from the actual autocorrelation $R_x(k)$ as*

$$P\left(|\hat{R}_x(k) - R_x(k)| > \epsilon\right) \leq \left(\zeta_+(\frac{\epsilon}{R_x(0)}, \frac{R_x(k)}{R_x(0)})\right)^L + \left(\zeta_-(\frac{\epsilon}{R_x(0)}, \frac{R_x(k)}{R_x(0)})\right)^L \tag{4.5}$$

where we have $0 \leq \zeta_+(a,\rho), \zeta_-(a,\rho) < 1$ for $a > 0$.

Proof. Since $R_x(k) = 0, |k| \geq 2MN$ samples, the random variables $z_{mn}[l] \triangleq x_1[n,l]x_2[m,l], l = 1, 2, \cdots L$ are jointly Gaussian and uncorrelated, and hence independent. Also, $E((x_1[n,l])^2) = E((x_2[m,l])^2) = R_x(0)$, and $E(x_1[n,l]x_2[m,l]) = R_x(Nm - Mn)$. Using Chernoff bound on $\hat{R}_x[Nm - Mn] = \frac{1}{L}\sum_{l=1}^{L} z_{mn}[l]$ (See Appendix B), we can obtain the concentration inequality (4.5), in which we substituted a with $\epsilon/R_x(0)$, and ρ with $R_x(Mn - Nm)/R_x(0)$. □

The result shows that under the assumption that $\mathbf{R_x}(k)$ is a finite sequence ($x(t)$ is a MA process), the probability of large deviation of $\hat{R}_x(k)$ from its mean decays exponentially in L. The simulations in Fig. 4.2 experimentally show the tightness of bound for large L.

4.2.2 Robustness to the Perturbation in Sampling Instants

In order to study the effect of random jiter (or perturbation in sampling instants), we consider a specific class of WSS signals whose spectrum consists of lines, representing frequencies of sinusoids buried in noise. We study this case owing to the possibility of obtaining closed form expressions for the contribution of jitter on the estimated autocorrelation sequence. This model is similar to that of DOA estimation using sensor arrays. However, unlike our earlier analysis, (i) we do not consider a grid based model, and (ii) we treat the jitter as a random variable, that takes independent values at each sampling instant.

Recall the model defined in Section 1.2.1, where we have a signal $x(t) = \sum_{k=1}^{K} A_k e^{\jmath(2\pi f_k t+\phi_k)}$, with uniformly distributed phases ϕ_k. The signal is sampled using two A/D converters operating at rates $\frac{1}{MT}$ and $\frac{1}{NT}$, in which $1/T = 2f_{\max}$ is the Nyquist rate, yielding

$$x_1[n] = \sum_{k=1}^{K} A_k e^{\jmath\omega_k Mn+\jmath\phi_k} + z_n \tag{4.6}$$

$$x_2[m] = \sum_{k=1}^{K} A_k e^{\jmath\omega_k Nm+\jmath\phi_k} + z_m, \tag{4.7}$$

where $\omega_k = 2\pi f_k T$.

Now, assume that due to imperfections of the A/D converters, the samples are

picked with a random jitter. The perturbed samples can then be written as

$$\tilde{x}_1[n] = \sum_{k=1}^{K} A_k e^{\jmath\omega_k(Mn+\delta_1[n])+\jmath\phi_k} + z_n \tag{4.8}$$

$$\tilde{x}_2[m] = \sum_{k=1}^{K} A_k e^{\jmath\omega_k(Nm+\delta_2[m])+\jmath\phi_k} + z_m, \tag{4.9}$$

where $\delta_1[n], \delta_2[m]$ are i.i.d random variables distributed uniformly in $[-\frac{\rho}{2}, \frac{\rho}{2}]$. In this case, we can write $\tilde{\mathbf{y}}[k]$ as

$$\tilde{\mathbf{y}}[l] = \sum_{k=1}^{K} \tilde{\mathbf{a}}_k[l] e^{\jmath 2MNl+\jmath\phi_k}. \tag{4.10}$$

Here, $\tilde{\mathbf{a}}_k[l] = \mathbf{a}(\omega_k) \circ \mathbf{p}_k[l]$, where $\circ$ denotes the element-wise product, and $\mathbf{p}_k[l] = [e^{\jmath\omega_k\delta_1[2Nl]}\ e^{\jmath\omega_k\delta_1[2Nl+1]}\ \ldots\ e^{\jmath\omega_k\delta_1[2Nl+N-1]}\ e^{\jmath\omega_k\delta_2[2Ml]}$ $e^{\jmath\omega_k\delta_2[2Ml+1]}\ \ldots\ e^{\jmath\omega_k\delta_2[2Ml+2M-1]}]^T$ is a vector consisting of the samples of $\delta_1[n]$ and $\delta_2[m]$ constructed the same way $\mathbf{y}$ is composed of the elements of $x_1[n]$ and $x_2[m]$. The perturbed autocorrelation matrix is given by the following theorem.

Theorem 10. *In the presence of random jitter in the sampling, the perturbed autocorrelation matrix is given by*

$$\tilde{\mathbf{R}}_{\mathbf{y}} = \sum_{k=1}^{K} A_k^2 \mathbf{B}(\omega_k) \circ \mathbf{E}_k + \sigma^2 \mathbf{I} \tag{4.11}$$

where $\mathbf{E}_k$ is a matrix with ones on the diagonal and $sinc^2(\omega_k\rho/2)$ elsewhere.

Proof. The perturbed autocorrelation matrix is obtained by

$$\tilde{\mathbf{R}}_{\mathbf{y}} = E\left(\tilde{\mathbf{y}}[l]\tilde{\mathbf{y}}^H[l]\right) = E_\delta\left(\sum_{k=1}^{K} A_k^2 \tilde{\mathbf{B}}_k[l] + \sigma^2\mathbf{I}\right),$$

where $\tilde{\mathbf{B}}_k[l] = \tilde{\mathbf{a}}_k[l]\tilde{\mathbf{a}}_k^H[l]$ and $\tilde{\mathbf{B}}_k[l]$ can be written as $\tilde{\mathbf{B}}_k[l] = \mathbf{B}(\omega_k) \circ \mathbf{P}_k[l]$ with $\mathbf{P}_k[l] = \mathbf{p}_k[l]\mathbf{p}_k^H[l]$. The diagonal elements of the matrix $\mathbf{P}_k[l]$ are all 1 and the off

diagonal elements are of the form $e^{\jmath\omega_k\beta_{rs}[l]}, 1 \leq r, s \leq N+2M-1, r \neq s$, in which $\beta_{rs}[l]$ is the difference of two independent random variables with uniform distribution in $[-\frac{\rho}{2}\ \frac{\rho}{2}]$. As a result, the pdf of $\beta_{rs}[l]$ will be a triangular function spanning from $-\rho$ to ρ:

$$f_{\beta_{rs}[l]}(\beta) = \begin{cases} \frac{1}{\rho^2}(\rho+\beta) & \beta < 0 \\ \frac{1}{\rho^2}(\rho-\beta) & \beta \geq 0 \end{cases}$$

Hence, with integration we obtain $E_\delta(e^{\jmath\omega_k\beta_{rs}[l]}) = \left(\frac{\sin(\omega_k\rho/2)}{\omega_k\rho/2}\right)^2 \triangleq \text{sinc}^2(\omega_k\rho/2)$. This leads to (4.11) in which $\mathbf{E}_k$ is a matrix with ones on the diagonal and $\text{sinc}^2(\omega_k\rho/2)$ elsewhere. □

Corollary 3. *Let $\rho\omega_k \ll 1$ for all k. The deviation of perturbed autocorrelation matrix $\tilde{\mathbf{R}}_\mathbf{y}$ from the ideal autocorrelation matrix $\mathbf{R}_\mathbf{y}$ is given by*

$$\|\mathbf{R}_\mathbf{y} - \tilde{\mathbf{R}}_\mathbf{y}\|_F \leq (N+2M-1)\frac{\rho^2}{12}\sqrt{K\sum_{k=1}^{K} A_k^4\omega_k^4} \tag{4.12}$$

Proof. Using (4.11), we obtain $\mathbf{R}_\mathbf{y} - \tilde{\mathbf{R}}_\mathbf{y} = \sum_{k=1}^{K} A_k^2\mathbf{B}(\omega_k) \circ (\mathbf{1} - \mathbf{E}_k)$ in which $\mathbf{1} \in \mathbb{C}^{(N+2M-1)\times(N+2M-1)}$ is an all-ones matrix. Each off-diagonal element of $\mathbf{R}_\mathbf{y} - \tilde{\mathbf{R}}_\mathbf{y}$ can be upper-bounded by

$$\left|\left(\mathbf{R}_\mathbf{y} - \tilde{\mathbf{R}}_\mathbf{y}\right)_{r,s}\right|^2 \leq K\sum_{k=1}^{K} A_k^4\left(1 - \text{sinc}^2(\frac{\rho\omega_k}{2})\right)^2,$$

and the diagonal entries of $\mathbf{R}_\mathbf{y} - \tilde{\mathbf{R}}_\mathbf{y}$ are obviously equal to zero. Hence, we immediately get

$$\|\mathbf{R}_\mathbf{y} - \tilde{\mathbf{R}}_\mathbf{y}\|_F \leq (N+2M-1)\sqrt{K\sum_{k=1}^{K} A_k^4\left(1 - \text{sinc}^2(\frac{\rho\omega_k}{2})\right)^2}$$

Using the assumption $\rho\omega_k \ll 1$ for all k, we can approximate $1 - \text{sinc}^2(\frac{\rho\omega_k}{2})$ as $\frac{\rho^2 w_k^2}{12}$ to obtain (4.12).

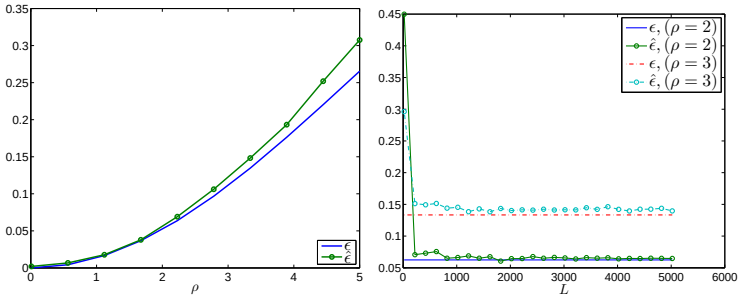
□

4.3 Simulations

In this Section, we study the bounds derived in Section 4.2, and compare them with empirical results. In the first experiment, we study the bound given in equation (4.11). Figure 4.1(a) (left) shows the deviation of the autocorrelation matrix ($\|\mathbf{R_y} - \tilde{\mathbf{R}}_\mathbf{y}\|_F / \|\mathbf{R_y}\|_F$) due to perturbations in sampling instants calculated using equation (4.11). For this experiment, we use an estimated value for $\tilde{\mathbf{R}}_\mathbf{y}$ using L samples. We choose $M = 3, N = 7$, $L = 500$ and consider 10 sinusoids with frequencies uniformly distributed between 10Hz and 200Hz. The deviation is plotted for the *sample* autocorrelation matrix (computed using $L = 500$ samples) and compared against the bound derived in (4.11).The empirical values match with the theoretical bound. The plot in Figure 4.1(b) (right) illustrate this deviation versus the number of samples L, for fixed values of ρ. As we increase L, the empirical values get close to the expected covariance matrices.

In the second experiment, we examine the upper bound given in Theorem 9. We pick the coprime numbers $M = 5, N = 7$. We consider a finite impulse response (FIR) filter with triangular shaped impulse response $h[n]$, non zero between time instants 0 and 29. We generate a WSS moving average (MA) process by passing i.i.d white Gaussian noise through this filter. This choice ensures the autocorrelation

sequence is zero beyond $2MN = 70$ as required by assumption (**A1**). Hence, the bounds of Theorem 9 should be applicable for such a signal, which is validated experimentally. We consider the lag $k = 7$ and set $\epsilon = 0.13 \times R_x(0)$. We conduct $N_{\text{runs}} = 1000$ Monte Carlo simulations to generate different instances of sample autocorrelation $\hat{R}_x(k)$ and empirically estimate the probability of the event $|R_x(k) - \hat{R}_x(k)| > \epsilon$. Figure 4.2 compares the empirical probability with the upperbound given in Theorem 9. The bound becomes tighter for larger values of L, and validates our claim that the error decays exponentially.



(a) Deviation vs ρ

(b) Deviation vs L, for fixed $\rho = 2$, $\rho = 3$.

Figure 4.1: Effect of jitter on autocorrelation estimation with coprime arrays. The symbol ϵ represents the deviation $\|\mathbf{R_y} - \tilde{\mathbf{R}}_\mathbf{y}\|_F / \|\mathbf{R_y}\|_F$ for the ideal value of the perturbed autocorrelation, while $\hat{\epsilon}$ is the same quantity for the estimated autocorrelation, computed using L samples.

4.4 Conclusion

In this Chapter, we studied the effects of additive perturbation and jitter in coprime sampling. We showed that such non idealness in sampling leads to errors in

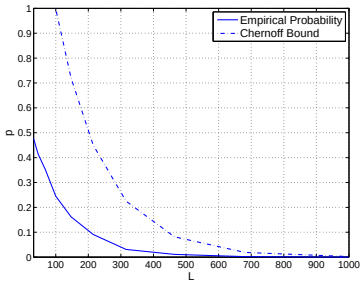


Figure 4.2: The empirical probability of the event $|\hat{R}_x(k) - R_x(k)| > \epsilon R_x(0)$ with $k = 7$ and $\epsilon = 0.13$, compared with the upper bound derived in Theorem 9. $\hat{R}_x(k)$ is the sample autocorrelation sequence of a moving-average process of order 29. Thus, $R_x(k)$ is zero for $k \geq 2MN$.

the estimated correlation which can be bounded under certain mild assumptions on the spectrum of the underlying WSS process. The robustness of coprime sampling is thereby established for a generic class of WSS signals, as well as for line spectrum processes, under small values of the perturbation.

Chapter 5: Conclusion and Future Work

5.1 Concluding Remarks

In this thesis, we studied the effects of perturbations and finite sample on the performance of coprime and nested sensing, in both spatial and temporal domains. For DOA estimation with spatial sensor arrays, the perturbations cause uncertainty in sensor locations and we treated them as unknown deterministic parameters of the problem. We established verifiable conditions under which the FIM is guaranteed to be non singular for such a model. For nested arrays with M sensors, continues to be non singular without assuming sparsity as a prior, as long as the grid size is $O(M^2)$.

We also studied the case where the perturbations are small, and we developed a bi-affine model in terms of the unknown perturbations and the source powers. We used the redundancies of the difference coarray to eliminate the nuisance variables, and reduce the bi-affine problem to a linear underdetermined (sparse) problem in source powers, which can be solved using ℓ_1 minimization. We derived this reduction for both ULA and a robust version of coprime arrays.

In the context of temporal sampling, we studied the effects of random sampling jitters, as well as finite number of samples on the estimated autocorrelation, and

examined the robustness of coprime sampling to these imperfections. Our theoretical guarantees are well supported by numerical experiments.

5.2 Future Work

In future, we will extend the analysis framework to more general parametric spectrum estimation techniques. We will also investigate the trade-off between latency, estimation error and sampling rates. Furthermore, we will consider non Gaussian signal models (which can be applicable for speech processing applications), and extend our current analysis framework.

Appendix A: Proofs

A.1 Proof of Theorem 2

First, let us define some notations which will be used in the proof.

Definition 5. *Let* $\bar{\mathbf{A}}_{ca,0} \in \mathbb{C}^{M_{ca} \times N_\theta}$ *be the matrix constructed from* $\mathbf{A}_{ca,0}$ *by removing the repeated rows and sorting them in order such that first row coresponds to difference of* $-\frac{M_{ca}-1}{2}$ *and last row corresponds to difference of* $\frac{M_{ca}-1}{2}$.

Notice that the difference between $\mathbf{A}^u_{ca,0}$ *and* $\bar{\mathbf{A}}_{ca,0}$ *is that for* $\bar{\mathbf{A}}_{ca,0}$ *we also keep the zero and negative lags of the co-array.*

Recall from Theorem 1 that non singularity of $\mathbf{J}$ is equivalent to $\mathbf{B} = [\mathbf{A}_{\text{ca},0}\ \mathbf{H}_{\boldsymbol{\delta}}]$ being full column rank. Our proof technique involves deriving sufficient conditions under which $\mathbf{B}$ has full column rank. Denote $\tilde{\mathbf{H}}_\delta = [\text{vec}(\mathbf{R}_{\delta_1}))\ \ \text{vec}(\mathbf{R}_{\delta_2})) \cdots\ \ \text{vec}(\mathbf{R}_{\delta_M}))]$. Recall that $\mathbf{H}_{\boldsymbol{\delta}} = (\tilde{\mathbf{H}}_\delta)_{:,2:M}$, i.e., the matrix comprised by the last $M-1$ columns of $\tilde{\mathbf{H}}_\delta$. After establishing that both $\mathbf{H}_{\boldsymbol{\delta}}$ and $\mathbf{A}_{\text{ca},0}$ have full column rank, we establish that there exists no intersection between the column spaces of $\tilde{\mathbf{H}}_\delta$ and $\mathbf{A}_{\text{ca},0}$. Then it directly follows that the column spaces of $\mathbf{H}_{\boldsymbol{\delta}}$ and $Acaz$ do not intersect as well, thereby proving that $\mathbf{B}$ is full column-rank.

Notice that every column of $\tilde{\mathbf{H}}_\delta$ is a vectorized form of the matrices $\mathbf{R}_{\delta_i}$ where

$\mathbf{R}_{\delta_i} \triangleq \frac{\partial \mathbf{R}_\mathbf{y}}{\partial \delta_i} = \mathbf{A}\mathbf{R}_\mathbf{x}\mathbf{D}_{\boldsymbol{\delta}_i}^H + \mathbf{D}_{\boldsymbol{\delta}_i}\mathbf{R}_\mathbf{x}\mathbf{A}^H$, and $\mathbf{D}_{\boldsymbol{\delta}_i} = \frac{\partial \mathbf{A}}{\partial \delta_i}$.

However, the matrix $\mathbf{R}_{\delta_i}$ is only supported on its ith column and its ith row. Hence, vec($\mathbf{R}_{\delta_i}$) is supported only on very specific rows as follows

$$\left(\tilde{\mathbf{H}}_\delta\right)_{q,r} = \begin{cases} \lambda_{r,s} & q = (s-1)M + r, 1 \leq s \leq M, s \neq r \\ -\lambda_{s,r} & q = (r-1)M + s, 1 \leq s \leq M, s \neq r \\ 0 & \text{else} \end{cases} \tag{A.1}$$

Accordingly, $\tilde{\mathbf{H}}_\delta \in \mathbb{C}^{M^2 \times M}$ can be written as

$$\begin{bmatrix}
0 & 0 & 0 & \cdots & 0 & 0 \\
\lambda_{2,1} & -\lambda_{2,1} & 0 & \cdots & 0 & 0 \\
\lambda_{3,1} & 0 & -\lambda_{3,1} & \cdots & 0 & 0 \\
\vdots & \vdots & \cdots & \ddots & \vdots & \vdots \\
\lambda_{M-1,1} & 0 & 0 & \cdots & -\lambda_{M-1,1} & 0 \\
\lambda_{M,1} & 0 & 0 & \cdots & 0 & -\lambda_{M,1} \\
-\lambda_{1,2} & \lambda_{1,2} & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & \cdots & 0 & 0 \\
0 & \lambda_{3,2} & -\lambda_{3,2} & \cdots & 0 & 0 \\
0 & \vdots & \cdots & \ddots & \vdots & \vdots \\
0 & \lambda_{M-1,2} & 0 & \cdots & -\lambda_{M-1,2} & 0 \\
0 & \lambda_{M,2} & 0 & \cdots & 0 & -\lambda_{M,2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
-\lambda_{1,M} & 0 & 0 & \cdots & 0 & \lambda_{1,M} \\
0 & -\lambda_{2,M} & 0 & \cdots & 0 & \lambda_{2,M} \\
0 & 0 & -\lambda_{3,M} & \cdots & 0 & \lambda_{3,M} \\
0 & \vdots & \cdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & -\lambda_{M-1,M} & \lambda_{M-1,M} \\
0 & 0 & 0 & \cdots & 0 & 0
\end{bmatrix}$$

where $\lambda_{r,s} = \sum_{k=1}^{N_\theta} \jmath \gamma_k \frac{2\pi k}{N_\theta} e^{\jmath \frac{2\pi k}{N_\theta}(d_r - d_s)}$, for all $1 \leq r, s \leq M$, $s \neq r$. Moreover, let $\lambda_{(m)} = \lambda_{r,s}$, for $d_r - d_s = m$. It can be verified that, $\lambda_{(m)} = -\lambda^*_{(-m)}$.

Lemma 1. *For almost all* $\boldsymbol{\gamma} \in \mathbb{R}^{N_\theta}$, $\mathbf{H}_{\boldsymbol{\delta}}$ *has full column rank.*

Proof. For almost all $\boldsymbol{\gamma} \in \mathbb{R}^{N_\theta}$ we have $\lambda_{r,s} \neq 0$ for $1 \leq r, s \leq M$. This is because $\lambda_{r,s} = 0$ describes a linear relation between the elements of $\boldsymbol{\gamma}$, which holds only for a set of measure zero in $\mathbb{R}^{N_\theta}$. Thus, looking at the rows 2 through M of $\mathbf{H}_{\boldsymbol{\delta}}$ (or equivalently, at $(\tilde{\mathbf{H}}_\delta)_{2:M,2:M}$), we have a diagonal matrix with nonzero numbers on the diagonal, for almost all $\boldsymbol{\gamma}$. Therefore, $\mathbf{H}_{\boldsymbol{\delta}}$ has rank $M-1$, i.e., $\mathbf{H}_{\boldsymbol{\delta}}$ is full column rank for almost all $\boldsymbol{\gamma}$. □

Observe that

$$\begin{aligned}\lambda_{(m+N_\theta)} &= \sum_{k=1}^{N_\theta} \jmath\gamma_k \frac{2\pi k}{N_\theta} e^{\jmath\frac{2\pi k}{N_\theta}(m+N_\theta)} \\ &= \sum_{k=1}^{N_\theta} \jmath\gamma_k \frac{2\pi k}{N_\theta} e^{\jmath\frac{2\pi k}{N_\theta}m} = \lambda_{(m)}.\end{aligned}$$

Similarly, the $(m+N_\theta)th$ row of $\bar{\mathbf{A}}_{\text{ca},0}$ is equal to its mth row. We will use this fact throughout the proof.

Since $\mathbf{A}_{\text{ca},0}$ is full column rank as long as $N_\theta \leq M_{\text{ca}}$, and $\mathbf{H}_{\boldsymbol{\delta}}$ is full column rank for almost all $\boldsymbol{\gamma}$ (Lemma 1), the only way for $\mathbf{B}$ to be column rank deficient is when there exist *non zero* $\boldsymbol{\alpha} \in \mathbb{C}^{N_\theta}$, $\tilde{\boldsymbol{\beta}} \in \mathbb{C}^{M-1}$ such that

$$\mathbf{A}_{\text{ca},0}\boldsymbol{\alpha} = \mathbf{H}_{\boldsymbol{\delta}}\tilde{\boldsymbol{\beta}}. \tag{A.2}$$

We first show that there exist no $\boldsymbol{\alpha} \neq 0$ and $\boldsymbol{\beta} \neq 0$ such that $\mathbf{A}_{\text{ca},0}\boldsymbol{\alpha} = \tilde{\mathbf{H}}_\delta\boldsymbol{\beta}$. This will, in particular imply non existence of non zero $\boldsymbol{\alpha}$ and $\tilde{\boldsymbol{\beta}}$ such that ((A.2)) holds. We prove this by contradiction, i.e. let us assume there exist nonzero $\boldsymbol{\alpha} \in \mathbb{C}^{N_\theta}$, $\boldsymbol{\beta} \in \mathbb{C}^M$ satisfying $\mathbf{A}_{\text{ca},0}\boldsymbol{\alpha} = \tilde{\mathbf{H}}_\delta\boldsymbol{\beta}$. This means that for every $1 \leq r, s \leq M$ we have

$$\beta_r - \beta_s = \frac{\sum_{k=1}^{N_\theta} \alpha_k e^{\jmath \frac{2\pi k}{N_\theta} m}}{\lambda_{(m)}} := f_{\boldsymbol{\alpha}}(m), \tag{A.3}$$

for $d_r - d_s = m$.

In a ULA, we have $d_i - d_{i-1} = 1$. Thus, $\beta_i - \beta_{i-1} = f_{\boldsymbol{\alpha}}(1)$ for $2 \leq i \leq M$. This implies that

$$\beta_i = \beta_1 + (i-1) f_{\boldsymbol{\alpha}}(1)$$

Upon substitution in (A.2) and using the definition of $f_\alpha(m)$ from (A.3), we obtain

$$\bar{\mathbf{A}}_{\text{ca},0} \boldsymbol{\alpha} = f_{\boldsymbol{\alpha}}(1) \times \begin{bmatrix} -(M-1)\lambda_{(-M+1)} \\ \vdots \\ -\lambda_{(-1)} \\ 0 \\ \lambda_{(1)} \\ \vdots \\ (M-1)\lambda_{(M-1)} \end{bmatrix} := \mathbf{c} \tag{A.4}$$

We notice that $f_{\boldsymbol{\alpha}}(1) \neq 0$. Otherwise, we would have had $\bar{\mathbf{A}}_{\text{ca},0} \boldsymbol{\alpha} = \mathbf{0}$, implying $\boldsymbol{\alpha} = \mathbf{0}$, which contradicts the fact that $\boldsymbol{\alpha} \neq \mathbf{0}$.

For simplicity, we index the rows of (A.4) from $-(M-1)$ to $M-1$. Let $p = -(M-1)$. From (A.3)

$$f_{\boldsymbol{\alpha}}(p) = \frac{\sum_{k=1}^{N_\theta} \alpha_k e^{\jmath 2\pi p k / N_\theta}}{\lambda_{(p)}} = p, \tag{A.5}$$

Notice that $f_{\boldsymbol{\alpha}}(p + N_\theta) = f_{\boldsymbol{\alpha}}(p)$. Also, since $N_\theta \leq 2M - 2$, $p + N_\theta \leq M - 1$ and we

can consider the $(p+N_\theta)$ th row to obtain

$$f_{\boldsymbol{\alpha}}(p+N_\theta) = \frac{\sum_{k=1}^{N_\theta} \alpha_k e^{\jmath 2\pi (p+N_\theta)k/N_\theta}}{\lambda_{(p+N_\theta)}} = p+N_\theta$$
$$\Rightarrow f_{\boldsymbol{\alpha}}(p) = \frac{\sum_{k=1}^{N_\theta} \alpha_k e^{\jmath 2\pi pk/N_\theta}}{\lambda_{(p)}} = p+N_\theta \tag{A.6}$$

However, (A.5) and (A.6) cannot hold at the same time since $p+N_\theta \neq p$. Therefore, the range spaces of $\mathbf{A}_{\text{ca},0}$ and $\mathbf{H}_{\boldsymbol{\delta}}$ do not coincide except for the zero vector. Using the facts that $\mathbf{H}_{\boldsymbol{\delta}}$ is full column rank for almost all $\boldsymbol{\gamma}$, $\mathbf{A}_{\text{ca},0}$ is full column rank for $N_\theta \leq 2M-1$, and the range spaces of $\mathbf{A}_{\text{ca},0}$ and $\mathbf{H}_{\boldsymbol{\delta}}$ do not intersect as long as $N_\theta \leq 2M-2$, we conclude that $\mathbf{B}$ has full column rank for almost all $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ as long as $N_\theta \leq 2M-2$.

A.2 Proof of Theorem 3

Proof. We prove for a slightly modified version of the nested array which is defined in (2.14).

We follow the same lines of proof of Theorem 2 upto (A.3) since the argument upto this point applies to any array geometry. For a nested array, the explicit structure of the vector $\mathbf{H}_{\boldsymbol{\delta}}\boldsymbol{\beta}$ will be quite different and we now examine it more closely. We use the same definition for $f_{\boldsymbol{\alpha}}(m)$ as in (A.3).

Let $f_{\boldsymbol{\alpha}}(m)$ be defined as in (A.3). Assuming M to be an even number, for a modified nested array which is defined in (2.14), we have

$$\beta_i - \beta_{i-1} = f_{\boldsymbol{\alpha}}(1),$$

for $2 \leq i \leq \frac{M}{2}+1$, therefore

$$\beta_i = \beta_1 + (i-1) f_{\boldsymbol{\alpha}}(1). \tag{A.7}$$

Moreover, we have

$$\beta_{\frac{M}{2}+1} - \beta_1 = f_{\boldsymbol{\alpha}}(\frac{M}{2}) \tag{A.8}$$

$$\beta_{\frac{M}{2}+i} - \beta_{\frac{M}{2}+i-1} = \beta_{\frac{M}{2}+1} - \beta_1 = f_{\boldsymbol{\alpha}}(\frac{M}{2}), \tag{A.9}$$

for $i = 2, \cdots, \frac{M}{2}$. Hence, we get

$$\beta_{\frac{M}{2}+i} = \beta_1 + f_{\boldsymbol{\alpha}}(\frac{M}{2}) i.$$

From (A.7) and (A.8), we get $\frac{M}{2} f_{\boldsymbol{\alpha}}(1) = f_{\boldsymbol{\alpha}}(\frac{M}{2})$.

Therefore, we have

$$\mathbf{c} := \bar{\mathbf{A}}_{\mathrm{ca},0} \boldsymbol{\alpha} = f_{\boldsymbol{\alpha}}(1) \times \begin{bmatrix} -\frac{M^2}{4} \lambda_{(-M^2/4)} \\ \vdots \\ -2\lambda_{(-2)} \\ -\lambda_{(-1)} \\ 0 \\ \lambda_{(1)} \\ 2\lambda_{(2)} \\ \vdots \\ \frac{M^2}{4} \lambda_{(M^2/4)} \end{bmatrix}$$

Similar to the proof of Theorem 2, index the rows as $-M^2/4$ to $M^2/4$ and consider the pth row and the $(p+N_\theta)$th row to conclude that (A.5) and (A.6)

cannot hold simultaneously, unless $f_{\boldsymbol{\alpha}}(1) = 0$. Choosing $p = -M^2/4$ is sufficient for our argument, since we only need to ensure that the $(p + N_\theta)$th row falls within range, i.e. $p + N_\theta = -M^2/4 + N_\theta \leq M^2/4$, which obviously holds since we assumed $N_\theta \leq M^2/2$.

□

A.3 Proof of Theorem 5

Proof. Due to Theorem 1, in order to ensure that the FIM is invertible, we only need to show that $\mathbf{B} = [\mathbf{A}_{\text{ca}} \quad \mathbf{H}_{\boldsymbol{\delta}}]$ is full column rank. A sparse $\boldsymbol{\gamma}$ only changes the explicit form of $\mathbf{H}_{\boldsymbol{\delta}}$ since the entries $\lambda_{(m)}$ of $\mathbf{H}_{\boldsymbol{\delta}}$ are now given by:

$$\lambda_{(m)} = \sum_{k=1}^{K} \jmath \gamma_{j_k} \omega_{j_k} e^{\jmath m d \omega_{j_k}},$$

where j_k indicates the index of the kth nonzero element of $\boldsymbol{\gamma}$, and w_{j_k} denotes its corresponding spatial frequency on the grid. As in the proofs of Theorems 2, 3, $\mathbf{B}$ is of full column rank if $\mathbf{A}_{\text{ca}}$ and $\mathbf{H}_{\boldsymbol{\delta}}$ are full column rank and there exist no non zero $\boldsymbol{\alpha}, \tilde{\boldsymbol{\beta}}$ such that $\mathbf{A}_{\text{ca}}\boldsymbol{\alpha} = \mathbf{H}_{\boldsymbol{\delta}}\tilde{\boldsymbol{\beta}}$. The proof for non existence of non zero $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ follow the same lines as earlier. However, we only need to establish conditions for full column rank of $\mathbf{H}_{\boldsymbol{\delta}}$. As argued in the proof of Theorem 2, the structure of $\mathbf{H}_{\boldsymbol{\delta}}$ in (A.1) dictates that it has full column rank $M - 1$ for almost all sparse $\boldsymbol{\gamma}$ with $\|\boldsymbol{\gamma}\|_0 = K$, if $\lambda_{(m)}$ is nonzero for every $0 \leq m \leq M_{\text{ca}}$. Let $\tilde{\boldsymbol{\gamma}} = [\boldsymbol{\gamma}_{j_1}, \cdots, \boldsymbol{\gamma}_{j_K}]$. We see that $\lambda_{(m)} = 0$ for a particular m describes a linear relation between the elements

of $\tilde{\boldsymbol{\gamma}}$:

$$0 = \sum_{k=1}^{K} \jmath \tilde{\gamma}_k \omega_{j_k} e^{\jmath m d \omega_{j_k}}$$

Hence, $\tilde{\boldsymbol{\gamma}}$ has a measure zero in $\mathbb{R}^K$. Hence, for almost all K-sparse $\boldsymbol{\gamma}$, $\lambda_{(m)} \neq 0$. This implies that $\mathbf{H}_{\boldsymbol{\delta}}$ has full column rank, and we can repeat the rest of the proofs of Theorems 2, 3 to establish that $\mathbf{J}$ is non singular for almost all $K-$ sparse $\boldsymbol{\gamma} \in \mathbb{R}^{N_\theta}, \|\boldsymbol{\gamma}\|_0 \leq K$. □

A.4 Proof of Theorem 7

Proof. Let k, i_k be integers such that $1 \leq k \leq M-2$ and $1 \leq i_k \leq M-k-1$. We have

$$R_{i_k+k+1,i_k+k} = f_1 + \lambda_1(\delta_{i_k+k+1} - \delta_{i_k+k}) \tag{A.10}$$

$$R_{i_k+1,i_k} = f_1 + \lambda_1(\delta_{i_k+1} - \delta_{i_k}) \tag{A.11}$$

Similarly,

$$R_{i_k+k+1,i_k+1} = f_k + \lambda_k(\delta_{i_k+k+1} - \delta_{i_k+1}) \tag{A.12}$$

$$R_{i_k+k,i_k} = f_k + \lambda_k(\delta_{i_k+k} - \delta_{i_k}) \tag{A.13}$$

Subtracting (A.10) from (A.11), and also (A.12) from (A.13) we obtain

$$\beta_k = \frac{R_{i_k+k+1,i_k+1} - R_{i_k+k,i_k}}{R_{i_k+k+1,i_k+k} - R_{i_k+1,i_k}} = \frac{\lambda_k}{\lambda_1} \tag{A.14}$$

Here, we assumed that $\boldsymbol{\delta} \in \mathbb{R}^M$ is such that

$$\delta_{i_k+k+1} - \delta_{i_k+k} - \delta_{i_k+1} + \delta_{i_k} \neq 0, \tag{A.15}$$

This can be violated only on a set of measure zero in $\mathbb{R}^M$. Hence, the following results will hold for *almost all* $\boldsymbol{\delta}$.

Notice that for a fixed k, different instances of equation (A.14) corresponding to different values of i_k in the range $1 \leq i_k \leq M-k-1$ are actually identical, as long as the observed covariance matrix is exact, and the assumption (A.15) holds for all i_k. Hence, in the sequel, we will assume $i_k = 1$.

From and (3.9) and (3.10), we can verify that

$$\bar{r}_k = kf_1 + \lambda_1(\delta_{k+1} - \delta_1) \tag{A.16}$$

From (A.13), (A.16) we get, for $2 \leq k \leq M-2$,

$$\beta_k \left(kf_1 - \bar{r}_k\right) = f_k - R_{k+1,1}$$

which can be expressed as

$$C_{k-1,1}f_1 + C_{k-1,k}f_k = h_{k-1}, \quad 2 \leq k \leq M-2 \tag{A.17}$$

where $C_{k-1,1}, C_{k-1,k}$, and h_{k-1} are given in (3.13), (3.14), and (3.15). We can express (A.17) in a more compact and explicit form as

$$\mathbf{Cf} = \mathbf{h} \tag{A.18}$$

in which

$$\mathbf{C} = \begin{bmatrix} C_{1,1} & C_{1,2} & 0 & \cdots & 0 \\ C_{2,1} & 0 & C_{2,3} & \cdots & 0 \\ \vdots & \vdots & \cdots & \ddots & \vdots \\ C_{M-3,1} & 0 & 0 & \cdots & C_{M-3,M-2} \end{bmatrix}$$

$$\mathbf{h} = [h_1 \;\; h_2 \;\; \cdots \;\; h_{M-3}]^T$$

□

A.5 Proof of Theorem 8

Proof. It can be easily verified that

$$\bar{r}_i^{(1)} = (i-1)f_{N_1} + (\delta_i^{(1)} - \delta_1^{(1)})\lambda_{N_1} \tag{A.19}$$

$$\bar{r}_j^{(2)} = (j-1)f_{N_2} + (\delta_j^{(2)} - \delta_1^{(2)})\lambda_{N_2} \tag{A.20}$$

For each lag k, one of the following four possibilities can happen: 1) $k = d_i^{(1)} - d_j^{(2)}$, 2) $k = d_j^{(2)} - d_i^{(1)}$, 3) $k = d_i^{(1)} - d_{i'}^{(1)}$, 4) $k = d_j^{(2)} - d_{j'}^{(2)}$.

1. $k = d_i^{(1)} - d_j^{(2)}$: We consider the case where the lag k is generated taking the difference between the ith sensor from the first sub-array and the jth sensor from the second sub-array.

   Since, we have doubled the number of sensors of each ULA, for each $1 \leq i \leq N_2$, and $1 \leq j \leq 2N_1$ so that $d_i^{(1)} - d_j^{(2)} = N_1 i - N_2 j = k$, the sensors indexed by $\bar{i} = i + N_2, \bar{j} = j + N_1$ also create the same lag k. Therefore, we can rewrite the equations (A.19) and (A.20) for $\bar{i}, \bar{j}$. Subtracting those equations from (A.19) and (A.20), we get

$$\bar{r}_{\bar{i}}^{(1)} - \bar{r}_i^{(1)} = N_2 f_{N_1} + (\delta_{\bar{i}}^{(1)} - \delta_i^{(1)})\lambda_{N_1} \tag{A.21}$$

$$\bar{r}_{\bar{j}}^{(2)} - \bar{r}_j^{(2)} = N_1 f_{N_2} + (\delta_{\bar{j}}^{(2)} - \delta_j^{(2)})\lambda_{N_2} \tag{A.22}$$

We also know that

$$R_{ij}^{(12)} = f_k + (\delta_i^{(1)} - \delta_j^{(2)})\lambda_k \tag{A.23}$$

$$R_{\bar{i}\bar{j}}^{(12)} = f_k + (\delta_{\bar{i}}^{(1)} - \delta_{\bar{j}}^{(2)})\lambda_k \tag{A.24}$$

We know that the (N_2+1)th element of the first ULA, and the (N_1+1)th element of the second ULA happen to be the same sensor on the coprime array (both at the normalized location N_1N_2). For this particular sensor, we can write equations (A.19) and (A.20).

$$\bar{r}_{N_2+1}^{(1)} = N_2 f_{N_1} + (\delta_{N_2+1}^{(1)} - \delta_1^{(1)})\lambda_{N_1} \tag{A.25}$$

$$\bar{r}_{N_1+1}^{(2)} = N_1 f_{N_2} + (\delta_{N_1+1}^{(2)} - \delta_1^{(2)})\lambda_{N_2} \tag{A.26}$$

where $\delta_{N_2+1}^{(1)} = \delta_{N_1+1}^{(2)}$ since they are the same sensor, and also $\delta_1^{(1)} = \delta_1^{(2)}$ for the same reason.

We can write similar equations for $(2N_2+1)$th sensor of first ULA, and $(2N_1+1)$th sensor of second ULA, which again happen to be the same sensor.

$$\bar{r}_{2N_2+1}^{(1)} = 2N_2 f_{N_1} + (\delta_{2N_2+1}^{(1)} - \delta_1^{(1)})\lambda_{N_1} \tag{A.27}$$

$$\bar{r}_{2N_1+1}^{(2)} = 2N_1 f_{N_2} + (\delta_{2N_1+1}^{(2)} - \delta_1^{(2)})\lambda_{N_2}, \tag{A.28}$$

in which $\delta_{2N_2+1}^{(1)} = \delta_{2N_1+1}^{(2)}$.

From (A.25), (A.26), (A.27), (A.28), we get

$$\frac{2\bar{r}_{N_1+1}^{(2)} - \bar{r}_{2N_1+1}^{(2)}}{\lambda_{N_2}} = \frac{2\bar{r}_{N_2+1}^{(1)} - \bar{r}_{2N_2+1}^{(1)}}{\lambda_{N_1}} \tag{A.29}$$

and from equation, we obtain (3.19)

$$\alpha = \frac{\lambda_{N_2}}{\lambda_{N_1}} = \frac{2\bar{r}^{(2)}_{N_1+1} - \bar{r}^{(2)}_{2N_1+1}}{2\bar{r}^{(1)}_{N_2+1} - \bar{r}^{(1)}_{2N_2+1}}$$

From equations (A.23), (A.24) from (A.21), (A.22), we get

$$\begin{aligned}\delta^{(1)}_{\bar{i}} - \delta^{(2)}_{\bar{j}} - \delta^{(1)}_{i} + \delta^{(2)}_{j} &= \frac{\bar{r}^{(1)}_{\bar{i}} - \bar{r}^{(1)}_{i} - N_2 f_{N_1}}{\lambda_{N_1}} \\ &\quad - \frac{\bar{r}^{(2)}_{\bar{j}} - \bar{r}^{(2)}_{j} - N_1 f_{N_2}}{\lambda_{N_2}} \\ &= \frac{R^{(12)}_{\bar{i}\bar{j}} - R^{(12)}_{ij}}{\lambda_k},\end{aligned}$$

whereby (3.20) follows, where $[\boldsymbol{\beta}_{cp}]_k := \frac{\lambda_k}{\lambda_{N_1}}$.

Now, using (A.19), (A.20), we can write

$$\delta^{(1)}_{i} - \delta^{(2)}_{j} = \frac{\bar{r}^{(1)}_{i} - (i-1)f_{N_1}}{\lambda_{N_1}} - \frac{\bar{r}^{(2)}_{j} - (j-1)f_{N_2}}{\alpha\lambda_{N_1}}$$

and

$$R^{(12)}_{ij} = f_k + \beta_k(\bar{r}^{(1)}_{i} - (i-1)f_{N_1} - (\bar{r}^{(2)}_{j} - (j-1)f_{N_2})\alpha^{-1}) \tag{A.30}$$

which is linear in terms of elements of $\mathbf{f}$, and hence it is linear in terms of $\boldsymbol{\gamma}$. By varying the indices i and j in the range $1 \le i \le N_2$ and $1 \le j \le 2N_1$ in (A.30), we obtain the corresponding rows of the system of equations $\mathbf{C}_{cp}\mathbf{f} = \mathbf{h}_{cp}$.

2. $k = d^{(2)}_j - d^{(1)}_i$:

In this case, (A.23), (A.24) should be rewritten as

$$(R^{(12)}_{ij})^* = R^{(21)}_{ji} = f_k + (\delta^{(2)}_{j} - \delta^{(1)}_{i})\lambda_k \tag{A.31}$$

$$(R^{(12)}_{\bar{i}\bar{j}})^* = R^{(21)}_{\bar{j}\bar{i}} = f_k + (\delta^{(2)}_{\bar{j}} - \delta^{(1)}_{\bar{i}})\lambda_k \tag{A.32}$$

We can repeat the math accordingly to get

$$(R_{ij}^{(12)})^* = f_k + \beta_k((\bar{r}_j^{(2)} - (j-1)f_{N_2})\alpha^{-1} - (\bar{r}_i^{(1)} - (i-1)f_{N_1}))$$

3. $k = d_i^{(1)} - d_{i'}^{(1)}$ and $k = d_j^{(2)} - d_{j'}^{(2)}$: These cases can be handled similar to ULA, following the same lines of math as in the proof of Theorem 7.

□

Appendix B: Chernoff Bound

Following [34], the moment generating function of product of two zero mean correlated Gaussian variables with unit variance is given by

$$M_y(t) = \frac{1}{\sqrt{(1-\rho^+ t)(1+\rho^- t)}}, \quad -\frac{1}{\rho^-} < t < \frac{1}{\rho^+} \tag{B.1}$$

in which $Y = X_1 X_2$, and X_1, X_2 are the Gaussian variables, $\rho^+ = 1+\rho$, and $\rho^- = 1-\rho$ where $\rho = E(X_1 X_2)$. Moreover, assuming that we have L independent such Y variables, we can derive a Chernoff bound

$$P(|\frac{1}{L}\sum_{l=1}^{L} Y_l - \rho| > a) \leq (\zeta_+(a,\rho))^L + (\zeta_-(a,\rho))^L\,, \tag{B.2}$$

in which we have

$$\zeta_+(a,\rho) = \inf_{t>0}\left\{e^{-(a+\rho)t} M_y(t)\right\} \tag{B.3}$$

$$\zeta_-(a,\rho) = \inf_{t<0}\left\{e^{-(\rho-a)t} M_y(t)\right\}. \tag{B.4}$$

Taking the derivatives with respect to t, equating with zero subject to the constraints $t_+ > 0$ and $t_- < 0$ and $-\frac{1}{\rho^-} < t_i < \frac{1}{\rho^+}$, we get

$$t_i^* = \pm\frac{-d_i + \sqrt{d_i^2 + 4\rho' c_i a}}{2c_i\rho'} \tag{B.5}$$

in which i can be either $+$ or $-$, and $c_+ = \rho + a, c_- = a - \rho, d_+ = 1 + \rho^2 + 2a\rho, d_- = 1 + \rho^2 - 2a\rho, \rho' = 1 - \rho^2$, and we assign the $+$ sign of the $\pm$ in (B.5) to t_+^* and the $-$ sign to t_-^*.

We substitute t_i^* in (B.3) and (B.4) to get (4.4).

It can be further verified that $\frac{\partial \zeta_i(a,\rho)}{\partial a} < 0$ for $a > 0$, and $\zeta_i(0,\rho) = 1$. Hence, $\zeta_i < 1$ for $a > 0$. Therefore, the RHS of (B.2) decays exponentially with respect to L.

Bibliography

[1] Harry L Van Trees. *Detection, Estimation, and Modulation Theory, Optimum Array Processing*. John Wiley & Sons, 2004.

[2] Piya Pal and PP Vaidyanathan. Nested arrays: a novel approach to array processing with enhanced degrees of freedom. 58(8):4167–4181, 2010.

[3] Piya Pal and PP Vaidyanathan. Coprime sampling and the music algorithm. In *Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (DSP/SPE), 2011 IEEE*, pages 289–294. IEEE, 2011.

[4] Piya Pal and PP Vaidyanathan. Correlation-aware techniques for sparse support recovery. In *Statistical Signal Processing Workshop (SSP), 2012 IEEE*, pages 53–56. IEEE, 2012.

[5] Hamid Krim and Mats Viberg. Two decades of array signal processing research: the parametric approach. *Signal Processing Magazine, IEEE*, 13(4):67–94, 1996.

[6] Ralph O Schmidt. Multiple emitter location and signal parameter estimation. *Antennas and Propagation, IEEE Transactions on*, 34(3):276–280, 1986.

[7] Palghat P Vaidyanathan and Piya Pal. Sparse sensing with co-prime samplers and arrays. *Signal Processing, IEEE Transactions on*, 59(2):573–586, 2011.

[8] S Unnikrishna Pillai, Yeheskel Bar-Ness, and Fred Haber. A new approach to array geometry for improved spatial spectrum estimation. *Proceedings of the IEEE*, 73(10):1522–1524, 1985.

[9] Ali Koochakzadeh and Piya Pal. Sparse source localization in presence of co-array perturbations. In *Sampling Theory and Applications (SampTA), 2015 International Conference on*, pages 563–567. IEEE, 2015.

[10] Ali Koochakzadeh and Piya Pal. On the robustness of co-prime sampling. In *Signal Processing Conference (EUSIPCO), 2015 23rd European*, pages 2825–2829. IEEE, 2015.

[11] Ali Koochakzadeh and Pal Piya. Performance analysis of coprime and nested samplers. *IEEE Transactions on Signal Processing (Under Review)*, 2015.

[12] Ali Koochakzadeh and Pal Piya. Sparse source localization using perturbed arrays via bi-affine modeling. *Signal Processing, Elsevier (Accepted)*, 2016.

[13] Ali Koochakzadeh and Pal Piya. Cramér rao bounds for underdetermined source localization. *IEEE Letters on Signal Processing (Accepted)*, 2016.

[14] Yimin D Zhang, Moeness G Amin, and Braham Himed. Sparsity-based doa estimation using co-prime arrays. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 3967–3971. IEEE, 2013.

[15] Piya Pal and PP Vaidyanathan. On application of lasso for sparse support recovery with imperfect correlation awareness. In *Signals, Systems and Computers (ASILOMAR), 2012 Conference Record of the Forty Sixth Asilomar Conference on*, pages 958–962. IEEE, 2012.

[16] Piya Pal and PP Vaidyanathan. Correlation-aware sparse support recovery: Gaussian sources. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 5880–5884. IEEE, 2013.

[17] LPHK Seymour, CFN Cowan, and PM Grant. Bearing estimation in the presence of sensor positioning errors. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'87.*, volume 12, pages 2264–2267. IEEE, 1987.

[18] Yih-Min Chen, Ju-Hong Lee, C-C Yeh, and Jeich Mar. Bearing estimation without calibration for randomly perturbed arrays. *Signal Processing, IEEE Transactions on*, 39(1):194–197, 1991.

[19] Anthony J Weiss and Benjamin Friedlander. Array shape calibration using sources in unknown locations-a maximum likelihood approach. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 37(12):1958–1966, 1989.

[20] Mats Viberg and A Lee Swindlehurst. A bayesian approach to auto-calibration for parametric array signal processing. *Signal Processing, IEEE Transactions on*, 42(12):3495–3507, 1994.

[21] Zhang-Meng Liu and Yi-Yu Zhou. A unified framework and sparse bayesian perspective for direction-of-arrival estimation in the presence of array imperfections. *Signal Processing, IEEE Transactions on*, 61(15):3786–3798, 2013.

[22] Petre Stoica and Thomas L Marzetta. Parameter estimation problems with singular information matrices. *Signal Processing, IEEE Transactions on*, 49(1):87–90, 2001.

[23] Magnus Jansson, Bo Goransson, and Bjorn Ottersten. A subspace method for direction of arrival estimation of uncorrelated emitter signals. *Signal Processing, IEEE Transactions on*, 47(4):945–956, 1999.

[24] Ravi P Agarwal, Kanishka Perera, and Sandra Pinelas. *An introduction to complex analysis*. Springer Science & Business Media, 2011.

[25] David L Donoho, Michael Elad, and Vladimir N Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *Information Theory, IEEE Transactions on*, 52(1):6–18, 2006.

[26] Keyong Han, Peng Yang, and Arye Nehorai. Calibrating nested sensor arrays with model errors.

[27] Arogyaswami Paulraj and Thomas Kailath. Direction of arrival estimation by eigenstructure methods with unknown sensor gain and phase. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'85.*, volume 10, pages 640–643. IEEE, 1985.

[28] Yanjun Li, Kiryung Lee, and Yoram Bresler. A unified framework for identifiability analysis in bilinear inverse problems with applications to subspace and sparsity models. *arXiv preprint arXiv:1501.06120*, 2015.

[29] Weiyu Xu and Babak Hassibi. Efficient compressive sensing with deterministic guarantees using expander graphs. In *Information Theory Workshop, 2007. ITW'07. IEEE*, pages 414–419. IEEE, 2007.

[30] Mitsuru Shinagawa, Yukio Akazawa, and Tsutomu Wakimoto. Jitter analysis of high-speed sampling systems. *Solid-State Circuits, IEEE Journal of*, 25(1):220–224, 1990.

[31] Harau Kobayashi, Masanao Morimura, Kensuke Kobayashi, and Yoshitaka Onaya. Aperture jitter effects in wideband sampling systems. In *Instrumentation and Measurement Technology Conference, 1999. IMTC/99. Proceedings of the 16th IEEE*, volume 2, pages 880–884. IEEE, 1999.

[32] Selim Saad Awad. Analysis of accumulated timing-jitter in the time domain. *Instrumentation and Measurement, IEEE Transactions on*, 47(1):69–73, 1998.

[33] PP Vaidyanathan. Generalizations of the sampling theorem: Seven decades after nyquist. *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on*, 48(9):1094–1109, 2001.

[34] Leo A Aroian. The probability function of the product of two normally distributed variables. *The Annals of Mathematical Statistics*, pages 265–271, 1947.