UNIVERSITY OF CALIFORNIA SAN DIEGO

Sampling for Underdetermined Linear and Multilinear Inverse Problems: role of geometry and statistical priors

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Electrical Engineering (Signal & Image Processing)

by

Ali Koochakzadeh

Committee in charge:

Professor Piya Pal, Chair
Professor Bhaskar Rao
Professor Behrouz Touri
Professor Rayan Saab
Professor Lawrence Saul

2020

The dissertation of Ali Koochakzadeh is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California San Diego

2020

DEDICATION

To my family and my friends who have been supportive throughout these years.

EPIGRAPH

We should not pretend to understand the world only by the intellect. The judgement of the intellect is only part of the truth. —Carl Jung

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Piya Pal for her support as my PhD advisor and chair of my committee. Throughout this journey I have learned a lot from her, and I cannot thank her enough for her guidance, and patience.

The contents in this chapter are reprint of the material published in "IEEE Signal Processing Letters" (Sec. 2.1 and Sec. 2.5), "IEEE Transactions on Signal Processing" (Sec. 2.2), "IEEE ICASSP 2017 Conference" (Sec. 2.3), and "Elsevier Journal on Digital Signal Processing" (Sec. 2.4).

The work presented in Sec. 2.1 was supported in part by "University of Maryland, College Park" and in part by the "Department of Defense". Sec. 2.2 was supported in part by the "Department of Defense", in part by the "NSF CAREER Award (ECCS 1553954)", in part by "the University of Maryland, College Park", and in part by the "University of California, San Diego". Sec. 2.3 was supported in part by "NSF CAREER award (ECCS 1553954)", and in part by the "University of California, San Diego". The work presented in Sec. 2.5 was supported by in part by the "Office of Naval Research grant (ONR N00014-18-1-2038)", and in part by the "National Science Foundation (NSF CAREER ECCS 1700506)".

Chapter 3 contains the reprint of the material published in "IEEE Transactions on Signal Processing" (Sec. 3.1), "IEEE ICASSP 2018" (Sec. 3.2), "SPIE 2018 Conference on Compressive Sensing VII: From Diverse Modalities to Big Data Analytics" (Sec. 3.3), "2019 53rd Asilomar Conference on Signals, Systems, and Computers" (Sec. 3.4).

The work in Sec. 3.1 was supported by in part by the "National Science Foundation under CAREER Award ECCS 1553954", in part by the "Office of Naval Research under Grant N00014-18-1-2038", and in part by the "University of California, San Diego". Sec. 3.2 was supported in part by "NSF CAREER award (ECCS 1553954)", and in part by the "University of California, San Diego". Sec 3.4 was supported by the "Office of Naval Research grant (ONR N00014-18-1-2038)", and the "University of California, San Diego".

Finally, the contents in Chapter 4 are reprints of the material published in "IEEE Data Science Workshop 2019" (Sec. 4.2), and "IEEE Sensor Array and Multichannel Signal Processing (SAM) 2020" (Sec. 4.3). Sec. 4.4 contains the material submitted to "IEEE Journal on Selected Topics in Signal Processing (JSTSP)".

The work in Sec. 4.2 was supported in parts by the "Office of Naval Research grant (ONR N00014-18-1-2038)", and the "University of California, San Diego". The work in Sec. 4.3 and 4.4 was supported by in part by the "Office of Naval Research grant (ONR N00014-18-1-2038)", and in part by the "National Science Foundation (NSF CAREER ECCS 1700506)".

VITA

2009-2014	B. S. in Electrical Engineering, Sharif University of Technology, Iran.
2014-2016	M. S. in Electrical Engineering, University of Maryland, College Park, USA.
2016-2020	Ph. D. in Electrical Engineering (Signal & Image Processing), University of California San Diego.

ABSTRACT OF THE DISSERTATION

Sampling for Underdetermined Linear and Multilinear Inverse Problems: role of geometry and statistical priors

by

Ali Koochakzadeh

Doctor of Philosophy in Electrical Engineering (Signal & Image Processing)

University of California San Diego, 2020

Professor Piya Pal, Chair

Estimating the underlying parameters of a statistical signal from noisy observations is a central problem in signal processing, with a wide variety of applications in many different fields such as machine learning, source localization, channel estimation for modern millimeter-wave (mmWave) communication systems, etc. Classical algorithms for source localization guarantee recovery of only $K = O(M)$ sources using a Uniform Linear Array equipped with M antennas. Recently, it has been shown that using certain non-uniform array designs, such as coprime and nested arrays, once certain correlation priors are assumed, it is possible to break this limit and go all the way up to $K = O(M^2)$ sources. This thesis sheds more light on this phenomena, and

more general cases of under-determined inverse problems in both linear, and non-linear settings. We show that for linear inverse problems, not only CRB exists for the case that $K = O(M^2)$, for certain non-uniform arrays, but also it continues to exist even if the antenna locations are perturbed due to physical deformation of the device, or only a compressed version of the measurements are available. For a more general class of linear inverse problems, we show that in presence of certain correlation priors, one can recover sparse vectors of sparsity $K = O(M^2)$, where the probability of detecting a wrong support for the sparse vector decays to zero exponentially fast as more and more temporal snapshots are obtained. We show these results hold for a variety of different statistical models, namely Gaussian sources, bounded sources, when the measurement matrices are equi-angular tight frames, and finally for the case that the measurements are obtained in adaptively. This thesis also considers multilinear inverse problems, namely tensor decompositions as well as certain non-convex problems with applications in millimeter-wave (mmWave) communication systems. We propose tensor decomposition algorithms for channel estimation for mmWave communication systems equipped with hybrid analog/digital beamforming, for cases such as multi-carrier Single-Input/Multiple-Output and single-carrier Multiple-Input/Multiple-Output, where we show the immense benefit gained by posing certain commonly considered statistical assumptions on the channel parameters, which leads to a provable increased identifiability compared to the existing algorithms for mmWave channel estimation.

Chapter 1

Introduction

Parameter estimation, i.e., estimating the defining parameters of a statistical signal from noisy observations is a central problem in signal processing, which has been studied for decades [KV96], and has broad applications in a wide variety of fields, such as machine learning, array signal processing, data mining, spectrum estimation, channel estimation for modern millimeter-wave (mmWave) communication systems, etc. Classical algorithms have been developed for many different problem models such as the MUSIC algorithm for Direction-of-Arrival estimation, where the goal is to find the directions of far-field narrow-band sources from the measurements obtained by an array of antennas. Recently, it has been shown that in many cases, one can benefit from the mathematics of the problem model in order to strategize the sampling in a judicious manner, in order to be able to estimate the parameters defining the problem model, in a much more efficient way, compared to the classical approaches.

Examples of such sampling strategies are co-prime[PV11], and nested sampling [PV10] which has been shown to be successful to recover $O(M^2)$ sources from only M antennas, unlike the classical approaches that could only recovery $O(M)$ sources using M antennas. The design of these novel array architectures is inspired solely by the mathematics of the problem and certain statistical assumptions. Similar ideas have been also used to design efficient temporal samplers,

which use statistical properties of certain wide sense stationary processes. This thesis sheds more light on the properties of these non-uniform sampling strategies, in order to estimate parameters associated with linear, and multi-linear inverse problems.

This thesis is concerned with three main questions:

1. Can we provably identify $K = O(M^2)$ uncorrelated sources, by showing existence of Cramér Rao bounds, even in presence of array imperfections such as sensor location errors? (Addressed in Chapter 2).

2. In a Multiple Measurements Vector (MMV) problem with correlation priors, can we identify supports of size $O(M^2)$? How does the probability of detecting a wrong support decay, as we get more and more measurements? (Addressed in Chapter 3).

3. In applications such as channel estimation for mmWave communication, how many channel paths can we identify once we make certain statistical assumptions on the channel parameters? What is the role of non-uniform array, and hybrid beamforming? (Addressed in Chapter 4).

In Chapter 2, we consider the problem of parameter estimation for linear inverse problems, from different perspectives. Sec. 2.1 addresses this problem in the context of Direction-of-Arrival estimation, and provides fundamental lower bounds, namely Cramér Rao bounds (CRB) on estimation error for the parameters. The analysis provided in Sec. 2.1 show that a certain property associated with array geometry, namely difference co-array, plays a central role on determining the number of sources that can be recovered. For the first time, we show that the CRB exists even if $K = O(M^2)$ sources are available. Existence of CRB is a fundamental property which guarantees that one can come up with algorithms that can find unbiased estimates of the parameters of the problem. Sec. 2.2 and 2.4 focus on a variant of DoA estimation problem, where the antennas are not exactly located on their nominal locations, due to manufacturing imperfections, etc. In Sec. 2.2, we show that even if the array suffers from perturbations, the CRB still exists as long

as the difference co-array associated with the antenna array satisfies certain properties. Sec. 2.4 proposes an algorithm to find the source locations as well as the array perturbations, by making the assumption that the perturbations are small enough. Sec. 2.5 presents the Cramér Rao bounds for compressed uniform and non-uniform arrays, and shows even in presence of a compression matrix, we can come up with sufficient conditions under which CRB exists. For both uniform and non-uniform arrays, Sec. 2.5 shows that CRB exists for almost all $N \times M$ compression matrices (which arise in applications such as channel estimation for mmWave communication systems equipped with hybrid analog/digital beamforming), even if the number of sources are as large as $K = \min(M, O(N^2))$ for uniform arrays, and $K = \min(O(M^2), O(N^2))$ for nonuniform arrays such as nested arrays. Sec. 2.3 considers the asymptotic properties of the CRB, and specifically considers the case where the number of antennas tend to infinity. We show that even if we have infinite number of antennas, the estimation error of the source powers cannot go below a certain threshold. Hence, showing a fundamental limit on the estimation error even with infinite number of spatial samples.

Chapter 3, however, focuses on the asymptotic properties of increasing the number of temporal samples. In particular, this chapter assumes a grid-based model, where the sources are located on a grid, which is equivalent to a Multiple Measurement Model (MMV) problem. In chapter 3 we shows that as long as the sampling matrix satisfies certain properties, in terms of Kruskal rank of its self Khatri-Rao product, the MMV model is fundamentally able to find sparse signals with sparsity as large as $K = O(M^2)$, such that the probability of detecting a wrong *support*[1] decays to zero, exponentially fast as more and more temporal measurements are collected. Chapter 3 proves this result under different assumptions. Sec. 3.1 considers the case that the sources have a Gaussian distribution with equal powers. By considering an exhaustive search decoder, the results in Sec. 3.1 show that reliable support recovery is possible even if $K = O(M^2)$, where M is the size of each measurement vector. Sec. 3.3 shows a similar result,

[1]The index set corresponding to the non-zero elements. A more precise definition is given in Chapter 3.

through proposing a greedy algorithm, for the case that the sources have unequal powers, but the measurement matrix is an equiangular tight frame. Sec. 3.2 proves as long as $N = O(M^2)$, with N being the length of the unknown sparse vectors, and the sources have a bounded distribution, one can still reliably recover the support using a greedy algorithm. Finally, Sec. 3.4 shows that the correlation aware support recovery can be extended to adaptive settings, where the measurement matrix is designed as more and more measurements are obtained, assuming that the measurements are obtained from uncorrelated sources. The results in Sec. 3.4 also guarantee reliable recovery of supports of size $K = O(M^2)$.

Chapter 4 contains non-convex algorithms, and tensor decomposition algorithms with applications in mmWave channel estimation. Sec. 4.1 presents a non-convex algorithm for beam-pattern design based on Wirtinger flow. Sec. 4.2 presents an algorithm for Canonical Polyadic Decomposition of Overcomplete constrained tensors, where for a constrained fourth order tensor of size $M \times N \times M \times N$, we show that it is possible to recover $O(M^2N)$ tensor factors. The constrains assumed in Sec. 4.2 are closely related to difference set considered in Chapter 2. Sec. 4.3 proposes a tensor decomposition algorithm for mmWave channel estimation, which is capable of recovering $L = O(N_T N_R)$ channel paths, for a single-carrier Multiple-Input/Multiple-Output system with N_T Radio-Frequency (RF) chains in the transmitter and N_R RF-chains in the receiver. Finally, Sec. 4.4, based on the ideas presented in Sec. 4.2 proposes tensor decomposition algorithms for mmWave channel estimation in a multi-carrier Single-Input/Multiple-Output system equipped with OFDM modulation, which are capable of recovering $L = O(NK^2)$ channel paths, with N being the number of RF-chains of the base station, and K being the number of frequency bins used for training. The main idea in Sec. 4.4 is to design the frequency bins used for the pilots in a non-uniform fashion, that not only enables efficient channel training, but also one can perform communication simultaneously with channel estimation, thereby reducing the channel training overhead.

Chapter 2

Linear Inverse Problems: A Cramér Rao Bound Based Study

Direction of Arrival (DOA) estimation is a central problem in antenna array processing, that arises in diverse scenarios such as target localization, tracking and interference suppression in passive and active radar, radio astronomy, multi-microphone speech processing, and so forth [KV96]. The Cramér-Rao bound (CRB) provides a fundamental lower bound on the estimation error (mean squared error) of any unbiased DOA estimator, and hence it can be used as a universal tool to assess the performance of DOA estimation algorithms. Over the last three decades, CRB for DOA estimation have been extensively studied and simplified for different array signal models [SN89, SN90b, SN90a, JGO99a]. However, almost all existing derivations assume *an overdetermined signal model*, where the number of sources (say, K) is smaller than the number (say, M) of sensors. As explained later in this chapter, this is partly because most array signal models are based on a uniform linear array (ULA), that is incapable of resolving more sources than sensors.

Non-uniform array sparse geometries such as nested [PV10] and coprime arrays [PV11], however, have been shown to be capable of localizing $O(M^2)$ sources with only M antennas.

This is achieved by exploiting their difference co-arrays (or virtual arrays) [HK90b], which provably contain ULA segments with $O(M^2)$ elements. However, existing results mostly establish identifiability of $K \geq M$ DOA parameters (assuming ideal covariance matrix is available), and there is a pressing need for analyzing the estimation performance of nested or coprime arrays for underdetermined DOA estimation, in presence of noise and/or limited time snapshots. The first step towards such a performance analysis would be to develop CRB for the corresponding signal models. Existing CRB derivations are mostly applicable in the regime $K < M$, and they do not indicate if it is at all possible to compute the CRB when $K \geq M$.

This chapter focuses on deriving CRB for underdetermined source localization problem. We show that for a cleverly designed non-uniform linear array, the CRB can exist even if $K \geq M$, and establish necessary as well as sufficient conditions for existence of CRB. We also consider other variants of source localization problem, and their corresponding CRB. Section 2.2, focuses on the CRB for source localization in presence of sensor location errors, where we show that even in presence of unknown perturbations in the sensor locations, the CRB can still exists for the regime that $K \geq M$.

Section 2.3 considers the asymptotic properties of the CRB, and specifically considers the case where the number of antennas tend to infinity, while the number of sources is kept fixed. We show that even if we have infinite number of antennas, the estimation error of the source powers cannot go below a certain threshold. Hence, showing a fundamental limit on the estimation error even with infinite number of spatial samples.

In Section 2.5 we provide the CRB for a compressed sparse array, where it is assumed that instead of having direct access to the antennas, we can only obtain a lower dimensional (compressed) set of measurements. Such scenarios arise in applications such as millimeter-wave communication systems, which we consider in Chapter 4. We show that even in presence of such compression schemes, the CRB still exists, for almost all compression matrices, even if the number of sources exceed the smaller dimension of the compression matrix.

2.1 Cramér-Rao Bounds for Underdetermined Source Localization

Although Cramér-Rao bounds for array processing have been studied for decades [SN89, SN90b, SN90a, JGO99a], the case of underdetermined DOA estimation (when there are fewer sources than the number of physical sensors) has received little attention. In this Section, our goal is to bridge this gap by deriving exact conditions on the geometry of the array manifold under which the *FIM is guaranteed to be non-singular even in the underdetermined setting*, which is equivalent to the existence of CRB. Under these conditions, the CRB will provide a fundamental lower bound on the performance of any algorithm that aims to localize *more sources than sensors*, such as those proposed in [CY15, PV10, ZQA14].

2.1.1 Signal Model and Stochastic Cramér-Rao Bound

Signal Model

Consider K narrowband plane waves impinging on a linear array with M sensors, where the mth sensor is located at a distance of d_m from the origin (normalized with respect to the carrier wavelength). Let θ_i denote the Direction-of-Arrival (DOA) of the ith source. The received signal model is then given by

$$\mathbf{y}(l) = \mathbf{A}(\Theta)\mathbf{s}(l) + \mathbf{n}(l), \quad l = 1, 2, \cdots, L$$

where $\mathbf{y}(l) \in \mathbb{C}^M$ denotes the lth time snapshot of the signals received at the M sensors, $\mathbf{s}(l) \in \mathbb{C}^D$ are the unknown source signals and $\mathbf{n}(l)$ is the vector of additive noise at the M sensors. Here, $\mathbf{A}(\Theta) = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2) \cdots \mathbf{a}(\theta_K)]$ with $\mathbf{a}(\theta_k)$ $(d = 1, \cdots, D)$ representing the steering vector corresponding to the angle θ_k, and its elements are given by $[\mathbf{a}(\theta_k)]_m = e^{j\pi d_m \sin\theta_k}$.

Two types of assumptions on the source signals $\mathbf{s}(l)$ are commonly used for deriving

the CRBs [SN90b]: (i) the conditional model assumption (CMA), and (ii) the unconditional model assumption (UMA) [SN90b]. Under CMA, the source signals $\{\mathbf{s}(l)\}_{l=1}^{L}$ are assumed to be deterministic unknown parameters. However, under UMA, $\{\mathbf{s}(l)\}_{l=1}^{L}$ are random vectors, often assumed to be i.i.d realizations of a Gaussian distribution. We will show that *under the UMA, it is possible derive CRB even in the underdetermined setting (i.e. $K \geq M$) for suitable arrays, with certain additional assumptions on the correlation of* $\mathbf{s}(l)$. The main assumptions are given by

(A1) The source signal vectors $\{\mathbf{s}(l)\}_{l=1}^{L}$ are *pairwise uncorrelated* i.i.d Gaussian random vectors, distributed as $\mathbf{s}(l) \sim \mathcal{N}(\mathbf{0}, \Lambda)$, where Λ is a *diagonal* matrix with $\Lambda = \text{diag}(\sigma_1^2, \cdots, \sigma_K^2)$, and σ_i^2 is the power of the ith source signal.

(A2) The noise $\mathbf{n}(l)$ is spatially and temporally white Gaussian process, uncorrelated from the source signals, i.e., $\mathbf{n}(l) \sim \mathcal{N}(\mathbf{0}, \sigma_n^2\mathbf{I}), E(\mathbf{n}(l)\mathbf{s}^H(l)) = \mathbf{0}$. The noise power σ_n^2 is assumed to be known.

Assumption (A1) will be a *crucial factor enabling the recovery of more sources than sensors for suitable arrays*. Under (A1-A2), the received signal vectors $\{\mathbf{y}(l)\}_{l=1}^{L}$ are i.i.d Gaussian random variables, distributed as

$$\mathbf{y}(l) \sim \mathcal{N}(\mathbf{0}, \mathbf{A}(\Theta)\Lambda\mathbf{A}^H(\Theta) + \sigma_n^2\mathbf{I}) \tag{2.1}$$

The $2K$ unknown parameters describing the distribution of the received signal are

$$\Theta = [\sin\theta_1, \sin\theta_2, \cdots, \sin\theta_K]^T,$$
$$\mathbf{p} = [\sigma_1^2, \sigma_2^2, \cdots, \sigma_K^2]^T$$

We denote the vector of unknown parameters as $\psi = [\Theta^T, \mathbf{p}^T]^T$. We will study the Fisher Information Matrix (FIM) for estimating ψ from the received signal $\mathbf{y}(l), l = 1, 2, \cdots L$ distributed according to (2.1) to establish precise conditions under which the CRB can be computed in the

underdetermined setting ($K \geq M$).

Unconditional Cramér-Rao Bound

Consider L i.i.d realizations of a random vector $\mathbf{y} \in \mathbb{C}^M$ distributed as $\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}(\psi))$. Notice that the covariance matrix $\mathbf{R}(\psi)$ is parameterized by ψ. For the given signal model, $\mathbf{R}(\psi) = \mathbf{A}(\Theta)\Lambda\mathbf{A}^H(\Theta) + \sigma_n^2\mathbf{I}$. In this case, the entries of (unconditional) FIM $\mathbf{J}(\psi)$ are given by the general formula

$$\frac{1}{L}[\mathbf{J}(\psi)]_{m,n} = \text{vec}^H(\frac{\partial \mathbf{R}(\psi)}{\partial \Psi_m})\mathbf{W}(\psi)\text{vec}(\frac{\partial \mathbf{R}(\psi)}{\partial \Psi_n}), \tag{2.2}$$

where $\mathbf{W}(\psi) = \left(\mathbf{R}^{-T}(\psi) \otimes \mathbf{R}^{-1}(\psi)\right)$ [SLG01]. This representation of the FIM leads to the following (matrix) form of the CRB derived in [SN90a]:

$$CRB(\Theta) = \frac{\sigma^2}{2L}\{\text{Re}(\mathbf{D}^H\Pi_{\mathbf{A}}^{\perp}\mathbf{D}) \circ (\Lambda\mathbf{A}^H{\mathbf{R}_{\mathbf{y}}}^{-1}\mathbf{A}\Lambda)^T\}^{-1} \tag{2.3}$$

where $\mathbf{D} = [\frac{\partial \mathbf{a}(\theta_1)}{\partial \theta_1}\ \frac{\partial \mathbf{a}(\theta_2)}{\partial \theta_2}\ \cdots\ \frac{\partial \mathbf{a}(\theta_K)}{\partial \theta_K}]$. In deriving (2.3), the *FIM has been assumed to be invertible*, an assumption which is largely true in the regime $K < M$ considered in [SN90a, SLG01]. However, in this Section, we will re-evaluate the rank of $\mathbf{J}(\psi)$ in the regime $K \geq M$, and develop explicit conditions (dictated by the array geometry), under which it is possible to compute the CRB. However, the expression of the CRB may not be simplified to the form (2.3) when $K \geq M$.

2.1.2 Non-Singularity of Unconditional FIM and Cramér-Rao Bound

In this section, we study the algebraic structure of $\mathbf{J}(\psi)$ to understand when it is possible to derive CRB in the underdetermined setting. Under the assumption (A1), we rewrite the FIM which was earlier derived in [SLG01, JGO99a], and then conduct a deeper analysis on $\mathbf{J}(\psi)$, highlighting the role of "co-array" as a significant factor in determining the rank of $\mathbf{J}(\psi)$.

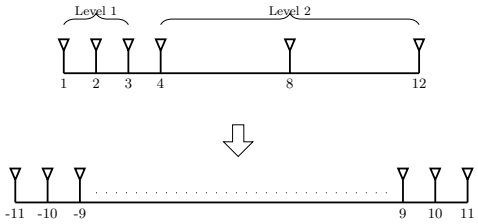


Figure 2.1: A 2-level nested array with $M = 6$ (top) and its difference co-array (bottom).

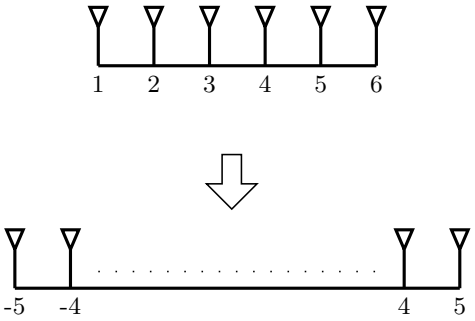


Figure 2.2: A ULA with $M = 6$ (top) and its difference co-array (bottom).

Co-Array and Khatri-Rao Product

Given a set $\mathcal{S} = \{d_m, 1 \leq m \leq M\}$ of sensor positions, the difference co-array associated with $\mathcal{S}$ is given by

$$\mathcal{S}_d = \{d_m - d_n, 1 \leq m, n \leq M\}$$

Notice that the number of distinct elements in $\mathcal{S}_d$ (denoted as $|S_k|$) is completely determined by the array geometry. It is well known that for a uniform linear array (ULA), $|\mathcal{S}_d| = 2M - 1$ whereas that for nested, coprime, and minimum redundancy arrays (MRA) is $O(M^2)$ [PV10]. Fig. 2.1 (resp. 2.2) the geometries of nested array (resp. ULA) along with its corresponding difference co-arrays.

Given the array manifold matrix $\mathbf{A}(\Theta)$, we denote its Khatri-Rao product $\mathbf{A}_{ca}(\Theta) \in \mathbb{C}^{M^2 \times K}$ as $\mathbf{A}_{ca}(\Theta) = \mathbf{A}^*(\Theta) \odot \mathbf{A}(\Theta)$ where $\odot$ represents column-wise Kronecker product between two matrices with same number of columns. There is a direct connection between $\mathbf{A}_{ca}(\Theta)$ and $\mathcal{S}_{ca}$

since the elements of $\mathbf{A}_{ca}(\Theta)$ are given by

$$[\mathbf{A}_{ca}(\Theta)]_{k,d} = e^{j\pi \sin\theta_d (d_m - d_n)}, \quad k = (n-1)M + m \tag{2.4}$$

Each distinct row of $\mathbf{A}_{ca}(\Theta)$ corresponds to an element in S_{ca} and hence $\mathbf{A}_{ca}(\Theta)$ corresponds to the manifold of a *virtual array* whose sensors are located at S_{ca}.

We also define a related matrix $\mathbf{B}(\Theta) \in \mathbb{C}^{M^2 \times K}$ whose elements are given by

$$[\mathbf{B}(\psi)]_{k,d} = j\pi(d_m - d_n) e^{j\pi \sin\theta_k (d_m - d_n)} \sigma_k^2, \tag{2.5}$$

where $k = (n-1)M + m$. It can be easily verified that $\mathbf{B}(\Theta)$ contains the same number of repeated rows as $\mathbf{A}_{ca}(\Theta)$ and they occur at identical indices. Moreover, we have $\text{rank}(\mathbf{A}_{ca}(\Theta)), \text{rank}(\mathbf{B}(\Theta)) \leq \min(K, |S_d|)$.

Non-singularity of FIM when $K \geq M$

We now state our main result regarding the non-singularity of $\mathbf{J}(\psi)$ by explicitly accounting for the role played by the co-array. We make the following additional assumption on the array geometry:

(A3) S_k contains the consecutive integers between 0 and M_d (and their negatives).

Many arrays such as ULA, nested and coprime satisfy this assumption, where $M_d = M - 1$ for ULA and $M_d = O(M^2)$ for nested and coprime arrays. In general, we have $2M_d + 1 \leq |S_k|$. The following theorem states the main result concerning the invertibility of $\mathbf{J}(\psi)$, and derives conditions under which the CRB exists. For ease of representation, we suppress the notations ψ and Θ in the following derivations, hoping that they will be clear from the context.

Theorem 1. *Under assumptions (A1-A3), if $2K \leq 2M_d + 1$, the FIM $\mathbf{J}(\psi)$ is invertible and the*

CRB for Θ *is given by*

$$CRB(\Theta) = \frac{1}{L}\left(\mathbf{B}^H \mathbf{W}^{1/2} \Pi^{\perp}_{\mathbf{W}^{1/2}\mathbf{A}_{ca}} \mathbf{W}^{1/2} \mathbf{B}\right)^{-1} \tag{2.6}$$

where $\Pi^{\perp}_{\mathbf{X}} \triangleq \mathbf{I} - \mathbf{X}(\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H$ *denotes the projection matrix onto the orthogonal complement of range of* $\mathbf{X}$*. When* $2K > |S_k|$*,* $\mathbf{J}(\psi)$ *is necessarily singular.*

Proof. Following [JGO99a], the FIM can be written as

$$\mathbf{J}(\psi) = L\mathbf{G}^H \mathbf{W} \mathbf{G}$$

where $\mathbf{G} = [\mathbf{B} \quad \mathbf{A}_{ca}]$. Since $\mathbf{W} = \mathbf{R}^{-T} \otimes \mathbf{R}^{-1}$ is positive definite (because $\mathbf{R}$ is positive definite), it is clear that $\mathbf{J}$ is non singular if and only if $\text{rank}(\mathbf{G}) = 2K$. From (2.4) and (2.5), notice that each of the matrices $\mathbf{A}_{ca}$ and $\mathbf{B}$ contains$|\mathcal{S}_k|$ distinct rows and they occupy the same indices in both matrices. Hence, $\text{rank}(\mathbf{G}) \leq \min\{2K, |S_k|\}$ and $2K \leq |S_k|$ is a necessary condition for $\mathbf{J}(\psi)$ to be non singular. To show the sufficient condition for FIM to be invertible, let $\mathbf{A}^{(u)}_{ca} \in \mathbb{C}^{(2M_d+1)\times K}$ and $\mathbf{B}^{(u)} \in \mathbb{C}^{(2M_d+1)\times K}$ denote the submatrices of $\mathbf{A}_{ca}$ and $\mathbf{B}$ that contain the distinct rows corresponding to the consecutive lags in the range $-M_d, \cdots, M_d$. Then, it can be verified that, for $1 \leq m \leq 2M_d + 1$,

$$[\mathbf{A}^{(u)}_{ca}]_{m,d} = e^{j\pi(m-M_d-1)\sin\theta_k}$$
$$[\mathbf{B}^{(u)}]_{m,d} = j\pi\sigma_k^2(m - M_d - 1)e^{j\pi(m-M_d-1)\sin\theta_k}$$

Since $2K \leq 2M_d + 1$, let us consider $2K \times 2K$ submatrices $\mathbf{A}_1 \in \mathbb{C}^{2K\times 2K}$ and $\mathbf{B}_1 \in \mathbb{C}^{2K\times 2K}$ of $\mathbf{A}^{(u)}_{ca}$ and $\mathbf{B}^{(u)}$ respectively such that, for $1 \leq m \leq 2K$,

$$[\mathbf{A}_1]_{m,d} = [\mathbf{A}^{(u)}_{ca}]_{m+1+M_d-D,k}$$
$$[\mathbf{B}_1]_{m,k} = [\mathbf{B}^{(u)}]_{m+1+M_d-K,k}$$

Clearly, $\mathbf{G}_1 = [\mathbf{A}_1 \quad \mathbf{B}_1]$ consists of $2K$ rows from $\mathbf{G}$ and $\mathbf{G}$ is non singular if $\mathbf{G}_1$ has rank $2K$. It is easy to show that indeed $\text{rank}(\mathbf{G}_1) = 2K$, by contradiction. Let us assume that $\mathbf{G}_1$ is rank deficient. Then $\exists \mathbf{v} \in \mathbb{C}^{2K}, \mathbf{v} = [v_1, v_2, \cdots v_{2K}]^T$ such that $\mathbf{A}_1^T \mathbf{v} = 0,$ and $\mathbf{B}_1^T \mathbf{v} = 0$. Define polynomials

$$f(x) \triangleq \sum_{m=1}^{2K} v_m x^{m-1}, h(x) \triangleq x^{-K+1} f(x),$$

and let $f'(x), h'(x)$ be their derivatives with respect to x. Then, $\mathbf{A}_1^T \mathbf{v} = 0$ implies that $h(x_k) = 0$, for $x_k = e^{j\omega_k}$, thereby $f(x_k) = 0$. Similarly, $\mathbf{B}_1^T \mathbf{v} = 0$ implies $x_k h'(x_k) = 0$, which is true only if $f'(x_k) = 0$. Hence, the polynomial $f(x)$ is such that both $f(x)$ and $f'(x)$ have zeros given by $x = e^{j\omega_1}, \cdots, e^{j\omega_K}$, implying these are double zeros for $f(x)$. This implies that $f(x)$ has $2K$ (double) zeros. However, the degree of $f(x)$ is $2K-1$, implying that it can only have $2K-1$ zeros, thereby contradicting our assumption that $\mathbf{G}_1$ is rank deficient. Hence, if $2K \leq 2M_d + 1$, $\mathbf{G}$ has rank $2K$ and $\mathbf{J}(\psi)$ is non singular. In this case, following [SLG01, JGO99a], the CRB for Θ is given by (2.6), which can be computed as the Schur complement of the top $K \times K$ block of $\mathbf{J}(\psi)$.

□

Remark 1. *Antenna arrays with the same number (M) of elements, but with different geometries, can have drastically different values for M_d. Therefore, the geometry of the array is the key factor that determines if the CRB exists in the underdetermined ($K \geq M$) regime. For a uniform linear array (ULA) with M antennas, $M_d = M-1$ and $|S_k| = 2M_d + 1 = 2M - 1$. Hence, from Theorem* 1*, the corresponding FIM is necessarily singular if $K \geq M$, implying that no unbiased estimator with finite variance exists [SM01a]. This explains why most traditional DOA estimation algorithms (that implicitly assume a ULA geometry) require $K < M$. A nested array [PV10] with M sensors, however, satisfies $M_d = M^2/4 + M/2 - 1$ and $S_d = 2M_d + 1$. Hence, it is possible to compute CRB for such an array even in the underdetermined setting ($K \geq M$) as long as $K \leq M^2/4 + M/2 - 1$.*

Table 2.1: CRB expressions for the signal model given by (2.1)

Problem Setting	CRB
Underdetermined/Determined $(K \geq M)$	(2.6)
Overdetermined $(K < M)$	(2.6), can be simplified to (2.3)

Remark 2. *For nested and ULA,* $|S_k| = 2M_d + 1$. *Hence* $2K \leq 2M_d + 1$ *is* both necessary and sufficient *for* $\mathbf{J}(\psi)$ *to be invertible for these arrays. However, for coprime arrays, due to presence of holes,* $|S_k| > 2M_d + 1$ *and hence* there is a gap between the necessary and sufficient conditions. *This gap will be demonstrated in our numerical simulations.*

We now summarize the CRB expressions for different problem settings in Table 2.1. In particular, note that the CRB expression for the determined signal model ($K = M$) is also given by (2.6). When $K < M$, (2.6) can be further simplified to (2.3).

Conditional versus Unconditional CRB

We have established that the unconditional FIM can be non singular even when $K \geq M$. We now consider the conditional signal model which has been extensively studied in [SN89, SN90b]. In [VP12], the conditional signal model was used to derive CRB for a coprime array when there are $K = 2$ sources. However, we will now show that the conditional model is an inappropriate choice for analyzing the performance of these arrays (whose difference set contains $O(M^2)$ elements) *when there are more sources than sensors*, since the corresponding CRB does not exist. To see this, recall that under the CMA, the received signal vectors $\{\mathbf{y}[l]\}_{l=1}^{L}$ are i.i.d and distributed as

$$\mathbf{y}(l) \sim \mathcal{N}(\mathbf{A}(\Theta)\mathbf{s}(l), \sigma_n^2\mathbf{I}) \tag{2.7}$$

Assuming the noise power σ_n^2 is known, there are $2LK + K$ unknown (real valued) parameters associated with this model, given by $\psi = [\{\mathrm{Re}(\mathbf{s}^T[l]]\}_{l=1}^{L}, \{\mathrm{Im}(\mathbf{s}^T[l]]\}_{l=1}^{L}, \Theta^T,]^T$. The corresponding

FIM is given by [SN89, Eqn (E.9)]

$$\mathbf{J}_C(\psi) = \begin{bmatrix} \mathbf{H}_s & \mathbf{X} \\ \mathbf{X}^H & \Gamma \end{bmatrix} \tag{2.8}$$

where $\mathbf{X}, \Gamma$ are appropriately defined matrices (see Appendix E in [SN89]) and $\mathbf{H}_b \in \mathbf{R}^{2KL \times 2KL}$ is a block diagonal matrix consisting of L blocks given by

$$\mathbf{H}_s = \begin{bmatrix} \mathbf{H}_b & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mathbf{H}_b \end{bmatrix} \tag{2.9}$$

Here $\mathbf{H}_b \in \mathbf{R}^{2K \times 2K}$ is given by $\mathbf{H}_b = \begin{bmatrix} \mathrm{Re}(\mathbf{H}) & -\mathrm{Im}(\mathbf{H}) \\ \mathrm{Im}(\mathbf{H}) & \mathrm{Re}(\mathbf{H}) \end{bmatrix}$ where

$$\mathbf{H} = \frac{2}{\sigma^2} \mathbf{A}^H(\Theta)\mathbf{A}(\Theta) \tag{2.10}$$

Hence, from (2.8), we can say that the FIM $\mathbf{J}_C(\psi)$ is non singular only if $\mathbf{H}_s$ is invertible which is true if and only if $\mathbf{H}$ is non singular. However, from (2.10), $\mathbf{H}$ is a $K \times K$ matrix which is non singular if and only if $\mathbf{A}(\Theta) \in \mathbb{C}^{M \times K}$ has rank K, which is true only if $K \leq M$.

Theorem 2. *Consider the Conditional Signal Model, where the L signal vectors received at an array of M sensors are distributed as (2.7). A necessary condition for the associated Fisher Information Matrix* $\mathbf{J}_c(\psi)$*, given by (2.8), to be invertible, is* $K \leq M$.

Hence, under CMA, the CRB exists only if $K \leq M$, i.e., there are fewer sources than sensors. When $K > M$, $\mathbf{J}_C(\psi)$ is necessarily singular, implying that no unbiased estimator with finite variance exists [SM01a]. This is true irrespective of the array geometry. Hence, under the CMA, it is not possible identify more sources than sensors, even when sparse arrays such as

coprime and nested, are used.

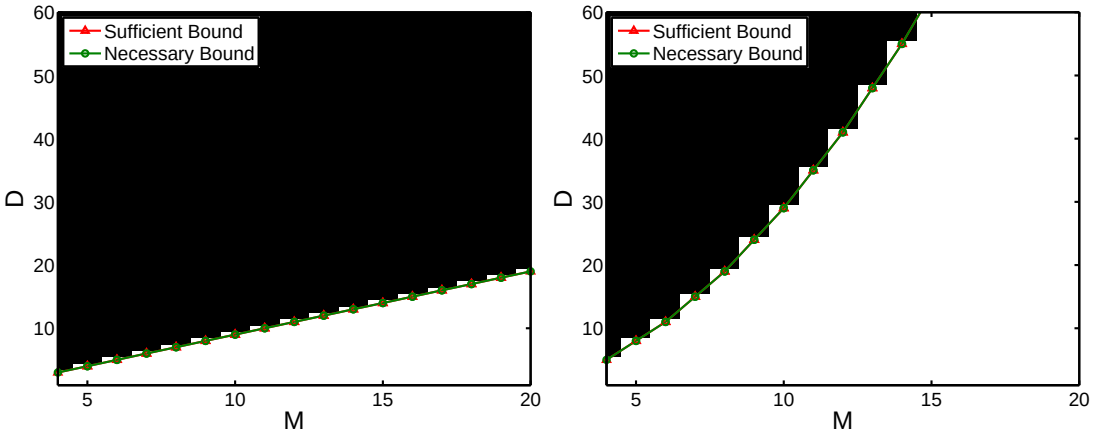
2.1.3 Simulations

We now conduct some proof-of-concept simulations to validate our claims and compare the Cramér-Rao bounds of ULA and different sparse arrays both in the overdetermined ($K < M$) and underdetermined ($K \geq M$) regimes. Firstly, we study the phase transition diagrams for non-singularity of FIM for ULA, nested and coprime arrays. The sources are located uniformly on the range $[-\pi/3 \ \ \pi/3]$, all with powers equal to 1. For each value of M and K, the nonsingularity of FIM is tested and white pixels show the values of K and M where FIM is invertible, while black pixels denote a singular FIM. For the coprime array, given M, the coprime numbers N_1, N_2 are chosen such that $2N_1 + N_2 - 2 = M$, $N_1 < N_2$, and $M_d = N_1 N_2$ attains the maximum possible value. Here, we assume $L = 1000$, and $\sigma_n^2 = 0.05$. The necessary and sufficient bounds derived in Theorem 1 are also overlaid. As illustrated in Figs. 2.3a, 2.3b, for the ULA and nested array, the necessary and sufficient conditions match exactly. However, in Fig. 2.3c we observe a gap between the two bounds, which is due to the fact that $2M_d + 1$ can be smaller than $|S_k|$.

In the second set of simulations, we evaluate the CRB (calculated from the trace of the $CRB(\Theta)$) corresponding to ULA, coprime and nested arrays for both $K < M$ and $K \geq M$. We assume $L = 1000$, $M = 12, N_1 = 3, N_2 = 7$. The CRB, as a function of SNR, is plotted in Fig. 2.4. The plots indicate that the nested array has the best performance compared to the other arrays.

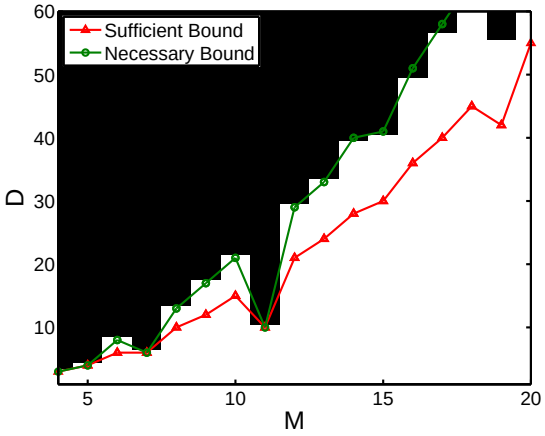
2.1.4 Conclusion

In this Section, we considered an underdetermined signal model (more sources than sensors) and derived sufficient and necessary conditions for the associated Fisher Information Matrix to be non-singular. We established that the Conditional Model (CM) is unsuitable for deriving the CRB in the underdetermined setting. We derived closed form expressions for the



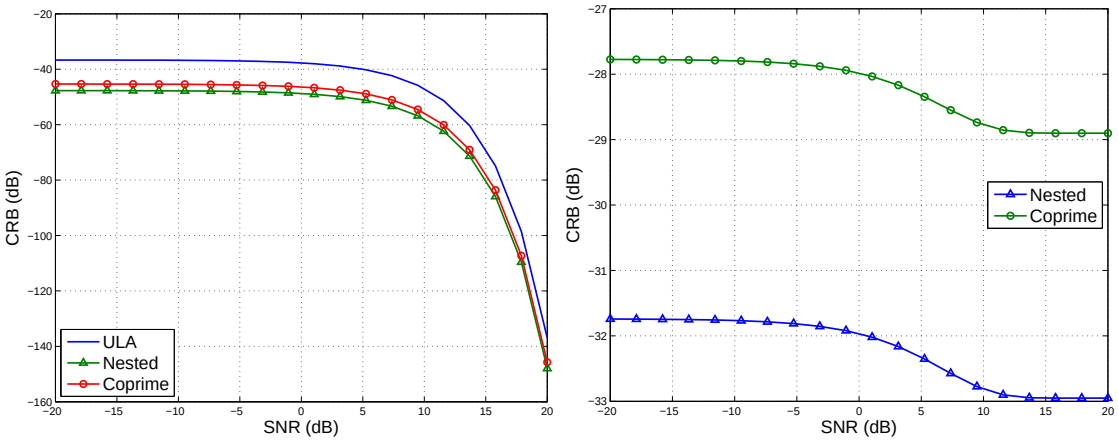
(a) ULA

(b) Nested



(c) Coprime

Figure 2.3: Phase transition for non-singularity of the Fisher Information Matrix.



(a) CRB vs SNR, $K = 5$ (b) CRB vs SNR, $K = 18$

Figure 2.4: The CRB for different arrays, as a function of SNR.

Cramér-Rao bounds for underdetermined DOA estimation and numerically compared the CRB for different array geometries such as ULA, nested and coprime arrays. We also verified our theoretical bounds through numerical simulations and empirical phase transition diagram.

2.2 Performance of Uniform and Sparse Non-Uniform Samplers In Presence of Modeling Errors: A Cramér-Rao Bound Based Study

In this section, we study the performance of coprime and nested samplers in non-ideal setting, where assumptions such as such as synchronous sampling, and ability to perfectly compute statistical expectations are violated. Using a general grid-based signal model that applies to both spatial and temporal line spectrum estimation, the effect of perturbations in sampling instants is evaluated by deriving fundamental Cramér-Rao Bounds (CRB) for line spectrum estimation with perturbed samplers. For the first time, simplified expressions for the Fisher Information matrix for perturbed coprime and nested samplers are derived, which explicitly highlight the role of coarray. We show that even in presence of perturbations, it is possible to resolve $O(M^2)$ spectral lines under appropriate conditions on the size of the grid. The effect of finite data on the CRB is also studied, and necessary and sufficient conditions are derived to ensure that the CRB decreases monotonically to zero with the number of measurements, even when there are more sources than sensors. Finally, the theoretical results derived in this Section are supported by extensive numerical experiments.

2.2.1 Line Spectrum Estimation and Sampling Perturbation: Fundamentals

In this Section, we consider a general signal model for line spectrum estimation (which can be used to represent both spatial and temporal line spectrum) and study the effect of sampling perturbations. These perturbations represent (i) sensor location errors in the context of spatial sampling, and (ii) jitter in the case of temporal sampling.

A Unified Grid-Based Model for Line Spectrum Estimation

We consider a line spectrum process, whose power spectrum consists of K spectral lines (which can be spatial, or temporal, as explained later) contaminated with additive white noise, uncorrelated with the signals. We assume a grid-based model for line spectrum estimation such that the spectral lines lie on a predefined grid of N_θ points, each of which represent a candidate frequency $\omega_n, 0 \leq n \leq N_\theta - 1$. Assuming that we collect M samples of this line spectrum process, acquired at the sampling instants $d_1, d_2, \cdots, d_M$, the mth sample of the observed signal can be written as

$$y_m = \sum_{i=1}^{N_\theta} e^{j2\pi d_m i/N_\theta} x_i + w_m, \quad 1 \leq m \leq M \tag{2.11}$$

Here $w_m, 1 \leq m \leq M$ represent M samples of white Gaussian noise, uncorrelated with the signal. The vector $\mathbf{x} = [x_1, x_2, \cdots, x_{N_\theta}]^T$ is a *sparse* vector with K non zero elements (i.e. $\|\mathbf{x}\|_0 = K$) representing (possibly complex) amplitudes of the K spectral lines. When the non zero elements of $\mathbf{x}$ are random variables, it is well known that (2.12) represents a wide-sense stationary (WSS) line spectrum process if and only if the non zero elements of $\mathbf{x}$ are also *statistically uncorrelated* [Vai07]. We will use this assumption throughout this Section. Defining $\mathbf{y} = [y_1, y_2, \cdots, y_M]^T$ and a matrix $\mathbf{A}_{\text{grid},0} \in \mathbb{C}^{M,N_\theta}$ such that $[\mathbf{A}_{\text{grid},0}]_{m,n} = e^{j\omega_n d_m}$ and $\mathbf{w} = [w_1, w_2, \cdots, w_M]^T$, we can represent (2.11) as

$$\mathbf{y} = \mathbf{A}_{\text{grid},0}\mathbf{x} + \mathbf{w} \tag{2.12}$$

Here, $\mathbf{A}_{\text{grid},0}$ is a possibly overcomplete dictionary (with $N_\theta > M$) that only depends on the sampling instants and the grid size N_θ. The general model (2.12) for line spectrum can be further modified and interpreted differently depending on whether the underlying line spectrum process is spatial or temporal.

1. **Spatial Line Spectrum Process:** The spatial line spectrum estimation problem is closely associated with that of estimating the directions-of-arrival (DOA) of narrow-band sources of radiation impinging on an array of sensors. Specific instances of this problem include target

localization and tracking in radar and sonar systems, speaker localization with microphone arrays, and neural source localization in medical imaging. In these cases, M denotes the number of sensors, and K represents the number of sources. The spatial sampling instant $\lambda/2d_m$ indicates the location of the mth sensor, where λ is the carrier wavelength of the impinging waves. The frequencies associated with spectrum lines in this scenario are determined by the direction of arrival of the sources. In particular, corresponding to each direction of arrival (DOA) $\theta \in [-\pi/2, \pi/2]$, we associate a spatial frequency $\omega = \pi \sin\theta$. We discretize the range ($[0, 2\pi)$) of possible values of ω into N_θ points, where the nth grid point is given by

$$\omega_n = \frac{2\pi n}{N_\theta}, n = 0, 1, \cdots, N_\theta - 1 \tag{2.13}$$

A key distinguishing feature of the signal model for spatial line spectrum estimation is that we acquire samples in *both space and time*, by observing a number of temporal snapshots (say, L) at each of the M sensors. This leads to the following space-time signal model for spatial spectrum estimation, which is a more generalized form of (2.12):

$$\mathbf{y}[l] = \mathbf{A}_{\text{grid},0}\mathbf{x}[l] + \mathbf{w}[l], \quad 1 \leq l \leq L \tag{2.14}$$

Here $\mathbf{y}[l]$ denotes the lth time snapshot of the signal received at the array of M antennas. It is typically assumed that the L temporal samples are statistically independent. [JGO99b, SN90c].

2. **Temporal Line Spectrum Process:**

It is well known that a line spectrum process in temporal domain can be represented as

$$s(t) = \sum_{k=1}^{K} x_k e^{j2\pi f_k t} \tag{2.15}$$

where $E(x_i x_j^*) = 0$ for $i \neq j$, and $E(x_i) = 0$ [Vai07, Theorem 7.3], and $\{f_k\}_{k=1}^K$ represent the

frequencies of the spectral lines. We assume that the signal is sampled using a periodic non uniform sampler [AL12] with period M_pT seconds, where $1/T = 2\max_k f_k$ is the Nyquist rate. We collect a total of ML samples (M and L are integers) over M_pLT seconds, by dividing the total acquisition time into L blocks of M_pT seconds (each block corresponds to one period of the periodic sampler), and collecting M samples in each block. The rate of such a sampler is therefore given by $\frac{M}{M_pT}$Hz. The sampling instants within the lth block are given by

$$t_{m,l} = (M_p(l-1) + d_m)T, 1 \leq l \leq L, 1 \leq m \leq M$$

We further assume that the location d_MT of the last sample in each block satisfies $(d_M + 1)T = M_pT$. As will be evident later, for a uniform sampler $M_p = d_M + 1 = M$, and its rate is $M/M_pT = 1/T$ Hz (Nyquist rate). However, for a nested sampler, $M_p = d_M + 1 = \frac{M}{2}(\frac{M}{2} + 1)$ and hence it operates at a sub-Nyquist rate of $\frac{2}{(\frac{M}{2}+1)T}$Hz.

To facilitate a unified Cramér-Rao Bound based analysis, we will henceforth assume that the non-zero amplitudes x_k in (2.15) are slowly time varying functions of time. In particular, each non-zero amplitude is a function of time (denoted explicitly as $x_k(t)$), that assumes a constant value of $x_k[l]$ over the lth block of length M_pT seconds, and changes independently from one block to another. In particular,

$$x_k(t) = x_k[l], \quad (l-1)M_pT \leq t < lM_pT$$

where $\{x_k[l]\}_{l=1}^{L}$ are zero-mean indpendent random variables. Such assumption of independence across blocks is a standard practice in spectral analysis using Bartlett type of spectrum estimators [SM05]. The samples of $s(t)$ in (2.15), contaminated with white

measurement noise, can therefore be represented as

$$y_m[l] = \sum_{k=1}^{K} x_k[l] e^{j2\pi T f_k d_m} + w_m[l] \tag{2.16}$$

Here $w_m[l]$ is the additive white noise uncorrelated with the signal, and $E(x_i[l]x_k^*[l']) = \gamma_k \delta[l-l']\delta[i-k]$ where γ_k represents the power of the kth source, and $\delta[.]$ is the Kronecker delta function. Defining an equivalent digital frequency $\omega = 2\pi fT$ and assuming that the K spectral lines lie on a grid of N_θ candidate frequencies $\omega_n, 0 \leq n \leq N-1$ with

$$\omega_n = \frac{2\pi n}{N_\theta}, 0 \leq n \leq N_\theta - 1,$$

we obtain the same model (2.14) as spatial line spectrum estimation introduced earlier. Notice that (2.14) represents an M dimensional vector valued wide-sense stationary (WSS) process [Vai07] such that

$$E(\mathbf{y}[l]) = 0, \quad E(\mathbf{y}[l]\mathbf{y}^H[l-k]) = \mathbf{R}_\mathbf{y}\delta[k]$$

Uniform and non-Uniform Sampling for line spectrum estimation

Our goals in this section are to study how perturbation and finite number of samples affect the quality of line spectrum estimators. To this end, we will consider two sampling strategies: (i) uniform Nyquist sampling and (ii) non uniform sparse Sub-Nyquist sampling, special instances of which are nested and coprime sampling. For uniform sampling, the sampling instants satisfy $d_m = m-1, 1 \leq m \leq M$. For nested sampling, assuming that M is an even number, we have

$$d_m = \begin{cases} m-1 & 1 \leq m \leq M/2 \\ (M/2+1)(m-M/2)-1 & M/2+1 \leq m \leq M \end{cases}$$

For coprime sampling, let us consider coprime numbers N_1 and N_2 ($N_1 < N_2$) such that $M = 2N_1 + N_2 - 1$. Then, for coprime sampling, the sampling instants d_m satisfy

$$d_m = \begin{cases} N_2(m-1) & 1 \leq m \leq 2N_1 \\ N_1(m-2N_1) & 2N_1+1 \leq m \leq M \end{cases}$$

In the context of spatial sampling, it is well known that with M spatial samples (or sensors), nested and coprime sampling can resolve $O(M^2)$ spectral lines, whereas uniform sampling can identify only $O(M)$ such lines [PV10, VP11]. In the context of temporal sampling, nested and coprime sampling operate at $O(M)$ times slower rates compared to uniform Nyquist sampling, and yet they can successfully resolve the spectral lines without aliasing [PV11]. The guarantees for nested and coprime sampling have so far been derived under idealistic assumptions such as a perfectly calibrated sensor array, or ignoring the effect of random jitter on temporal sampling. We will now evaluate the performance of these samplers in the presence of perturbations in the sampling instants that can result from calibration errors in sensor arrays, and random jitter in temporal sampling.

Effect of Perturbation in sampling instants

It is generally known that perturbations can seriously affect the performance of line spectrum estimation algorithms. For non uniform samplers such as nested and coprime arrays, the *difference co-array* has been shown to play a key role in their ability to resolve more sources than the number of sensors. Although perturbation of sensor locations in antenna arrays has been studied since decades [CLYM91, WF89, VS94, LS15a], their effect on the difference co-arrays of sparse samplers, and the ability to resolve $O(M^2)$ sources with M sensors, is much less understood. We aim to bridge such a gap by explicitly characterizing the role of perturbation in the co-array domain and establishing conditions for identifiability of DOAs in presence of perturbation.

For both uniform and non-uniform sampling, the perturbed sampling instants are given by $d_m + \delta_m, 1 \leq m \leq M$ where δ_m represents the *unknown* perturbation. For rest of this Section, without loss of generality, we will assume that the first sample is at the origin of our reference, i.e., $d_1 = \delta_1 = 0$. For example, in spatial spectrum estimation, this means that the location of other sensors are measured relative to that of the first sensor. For spatial spectrum sampling, using (2.14), the signal received at a perturbed antenna array can be written as

$$\mathbf{y}[l] = \mathbf{A}_{\text{grid}}(\delta)\mathbf{x}[l] + \mathbf{w}[l], 1 \leq l \leq L \tag{2.17}$$

where $\delta \in \mathbb{R}^{M-1}$ is defined as

$$\delta = [\delta_2, \delta_3, \cdots \delta_M]^T \tag{2.18}$$

and $\mathbf{A}_{\text{grid}}(\delta)$ is obtained from $\mathbf{A}_{\text{grid}}$ by replacing d_m with $d_m + \delta_m$. Note that (2.17) also represents the signal model for temporal line spectrum process in presence of sampling jitter (where δ represents the unknown jitter), with $L = 1$.

The Co-Array Model Since the non-zero elements of $\mathbf{x}$ are assumed to be uncorrelated, the vectorized form of the covariance matrix of the signal $\mathbf{y}$ in (2.17) is given by

$$\mathbf{z} = \text{vec}\Big(E(\mathbf{y}\mathbf{y}^H)\Big) = \mathbf{A}_{\text{ca}}(\delta)\gamma + \sigma_w^2 \,\text{vec}(\mathbf{I}), \tag{2.19}$$

where $\mathbf{A}_{\text{ca}}(\delta) = \mathbf{A}^*_{\text{grid}}(\delta) \odot \mathbf{A}_{\text{grid}}(\delta) \in \mathcal{C}^{M^2 \times N_\theta}$ denotes the Khatri-Rao product [PV12a]. The vector $\gamma = [\gamma_1, \cdots, \gamma_{N_\theta}]$ is the diagonal of $\mathbf{R}_{\mathbf{x}} = E(\mathbf{x}\mathbf{x}^H)$. The location of the non zero elements of γ coincide with those of $\mathbf{x}[l]$ and exactly reveal the frequencies of the spectral lines for the grid-based model. The elements of $\mathbf{A}_{\text{ca}}(\delta)$ are characterized by the perturbed version of the difference co-array. In particular, the $(m + m'M, k)$-th element of $\mathbf{A}_{\text{ca}}(\delta)$ is given by $e^{j\omega_k(d_m + \delta_m - d_{m'} - \delta_{m'})}$.

Thus each column of $\mathbf{A}_{\text{ca}}(\delta)$ is a function of the *perturbed difference set*:

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

We can suppress the effect of noise (σ_w^2) by removing the rows in $\mathbf{z}$ and $\mathbf{A}_{\text{ca}}(\delta)$ corresponding to the difference 0 in S_{ca}. We further sort the rows in ascending order with respect to their locations in the difference set and only retain the elements corresponding to the positive half to obtain

$$\mathbf{z}^u = \mathbf{A}_{\text{ca}}^u(\delta)\gamma \tag{2.20}$$

where $\mathbf{A}_{\text{ca}}^u(\delta) \in \mathbb{C}^{\frac{|S_{\text{ca}}|-1}{2} \times N_\theta}$, and $|S_{\text{ca}}|$ is the number of distinct elements in the difference set. We use the notation M_{ca} to denote the number of elements in the unperturbed difference set, i.e.

$$M_{\text{ca}} = |S_{\text{ca}}|_{(\delta=0)} \tag{2.21}$$

Number of Recoverable Spectral Lines Recall that the support of sparse γ represents the frequencies of spectral lines. One way to recover the support is to assume that the true γ represents the sparsest solution to (2.20) and solve

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

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

In Absence of Perturbation: In absence of perturbation ($\delta = 0$), the sparse vector γ can be uniquely recovered if $\text{Supp}(\gamma) \leq M_{ca}/2$, where M_{ca} is given by (2.21). For uniform sampling, the

unperturbed difference set is given by

$$S_{\text{ca}}^{\text{U}} = \{n \big| n = -M, \ldots, M\}$$

Hence, for uniform sampling, we have $M_{\text{ca}} = O(M)$. For nested sampling [PV10] with even number of sensors and two levels of nesting,

$$S_{\text{ca}}^{\text{Nested}} = \{n \big| n = -M'_{\text{ca}}, \ldots, M'_{\text{ca}}; M'_{\text{ca}} = \frac{M^2}{4} + \frac{M}{2} - 1\}$$

and hence $M_{\text{ca}} = 2M'_{\text{ca}} + 1 = O(M^2)$.

Perturbation and Size of Support: The presence of perturbation has a non trivial effect on the size of the recoverable support. By a simple equations-versus-unknown argument, it can be seen that for nested samplers, the number of distinct equations in (2.20) is still $O(M^2)$, whereas the number of unknowns is $M + K$, since we have M unknown values for the perturbation and K unknown values for the spectral lines. Since the number of unknowns only increases by $O(M)$, it may be still possible to identify $K = O(M^2)$ spectral lines in presence of perturbations. In this Section, we present a more formal study of the effect of perturbation, based on the Cramér Rao bound, in which we study conditions under which the Fisher Information Matrix (FIM) is guaranteed to be non singular, even in presence of perturbation. The analysis is not based on any particular algorithm, rather on the received signal model itself, which alone determines the Fisher Information Matrix.

2.2.2 Effect of Perturbations: A Cramér Rao Bound Based Study

Recall that under the grid-based model, the sampled line spectrum process is given by

$$\mathbf{y}[l] = \mathbf{A}_{\text{grid}}(\delta)\mathbf{x}[l] + \mathbf{w}[l], \quad 1 \leq l \leq L \tag{2.22}$$

We assume that $\mathbf{x}[l], l = 1, 2, \cdots, L$ are i.i.d random vectors following the normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{R}_{\mathbf{x}})$, where $\mathbf{R}_{\mathbf{x}}$ is a diagonal matrix satisfying $\mathbf{R}_{\mathbf{x}} = \text{diag}(\gamma)$. 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 signal $\mathbf{y}[l]$ is therefore distributed as

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

In the sequel, we will assume the noise power σ_w^2 to be known, since our goal is to understand how the presence of unknown δ affects the recovery of the desired parameter γ.

Remark 3. *In line spectrum estimation, a typical assumption is that $x_k = C_k e^{j\phi_k}$, where $C_k > 0$ and $\phi_k\big|_{k=1}^{K}$ are i.i.d random phases, uniformly distributed in the range $[0, 2\pi)$. Although this model also leads to a line spectrum process, for the ease of exposition, we assume that x_k are normally distributed as stated above. This leads to a Gaussian model for the overall process* $\mathbf{y}$ *and the associated Fisher Information Matrix becomes analytically tractable.*

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 $\mathbf{A}_{\text{grid}}(\delta)$, $\mathbf{A}_{\text{ca}}(\delta)$, $\mathbf{A}_{\text{ca}}^u(\delta)$, 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 $\delta = \mathbf{0}$.

Cramér Rao Bound

Singularity of the Fisher Information Matrix implies non existence of a consistent estimator for γ and δ [SM01a], unless the parameters satisfy certain constraints. Hence, non-singularity

of the FIM is a necessary condition for any algorithm to be able to exactly recover γ (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 δ and γ. We would like to point out the following facts about the results derived in this Section:

- **Sparsity Not Assumed:** Although the parameter γ 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 γ. In other words, under the derived conditions, a Maximum Likelihood method can recover a sparse γ (as $L \to \infty$) from an overcomplete observation model (2.12) with $N_\theta > M$, without assuming it to be sparse.

- **Number of spectral lines not assumed to be known:** We also *do not assume knowledge of the number of spectral lines, 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 to deriving the desired Cramér Rao bound. The probability density function (pdf) of the received signal is given by

$$f(\mathbf{y};\psi) = \frac{1}{\pi^M \det(\mathbf{R_y})} e^{-\mathbf{y}^H \mathbf{R_y}^{-1} \mathbf{y}}. \tag{2.23}$$

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

$$\mathbf{J}_{ij} = \mathbb{E}\left(\frac{\partial}{\partial \psi_i} \ln f(\mathbf{y};\psi) \frac{\partial}{\partial \psi_j} \ln f(\mathbf{y};\psi) \right). \tag{2.24}$$

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

$$\mathbf{H} \triangleq [\text{vec}(\mathbf{R}_{\delta_2})\ \text{vec}(\mathbf{R}_{\delta_3})\ \ldots\ \text{vec}(\mathbf{R}_{\delta_M})], \mathbf{R}_{\delta_i} \triangleq \frac{\partial \mathbf{R_y}}{\partial \delta_i} \tag{2.25}$$

The following theorem provides necessary and sufficient conditions under which the FIM in (2.24) is non singular:

Theorem 3. *Denoting $\psi = [\gamma^T\ \delta^T]^T$ as the parameters to be estimated, the FIM defined in (2.24) is invertible, iff the matrix* $\mathbf{B}$*, defined as follows, is full column rank:*

$$\mathbf{B} = [\mathbf{A}_{\text{ca}}\ \ \mathbf{H}] \tag{2.26}$$

Proof. When the observed signal consists of L i.i.d zero mean Gaussian random vectors with covariance matrix $\mathbf{R_y}$, the corresponding Fisher Information Matrix (FIM) can be derived as [JGO99b]

$$\frac{1}{L}\mathbf{J}_{ij} = \text{vec}(\frac{\partial \mathbf{R_y}}{\partial \psi_i})^H (\mathbf{R_y}^{-T} \otimes \mathbf{R_y}^{-1})\, \text{vec}(\frac{\partial \mathbf{R_y}}{\partial \psi_j}) \tag{2.27}$$

The Fisher information Matrix (FIM) for our model (2.23) can be divided into blocks corresponding to parameters γ and δ as:

$$\mathbf{J} = \begin{pmatrix} \mathbf{J}_{\gamma\gamma} & \mathbf{J}_{\gamma\delta} \\ \mathbf{J}_{\gamma\delta}^H & \mathbf{J}_{\delta\delta} \end{pmatrix} \tag{2.28}$$

Notice that

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

Hence, from (2.27) and (2.25), we obtain

$$\mathbf{J}_{\gamma\gamma} = L\mathbf{A}_{\text{ca}}^{H}\mathbf{W}\mathbf{A}_{\text{ca}},\ \mathbf{J}_{\gamma\delta} = L\mathbf{A}_{\text{ca}}^{H}\mathbf{W}\mathbf{H},\ \mathbf{J}_{\delta\delta} = L\mathbf{H}^{H}\mathbf{W}\mathbf{H} \tag{2.29}$$

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

$$\mathbf{J} = L\mathbf{B}^{H}\mathbf{W}\mathbf{B} \tag{2.30}$$

Since $\mathbf{W} = {\mathbf{R}_{\mathbf{y}}}^{-T} \otimes {\mathbf{R}_{\mathbf{y}}}^{-1}$ 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 $\mathbf{J}_{\gamma\gamma}$ and the following corollary establishes a sufficient condition for non singularity of the FIM.

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

Proof. From (2.29), $\text{rank}(\mathbf{J}_{\gamma\gamma}) = \text{rank}(\mathbf{A}_{\text{ca},0})$, since $\mathbf{W}$ is positive definite. Hence, $\mathbf{J}_{\gamma\gamma}$ 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 $N_\theta \leq M_{\text{ca}}$, which concludes the proof. □

Remark 4. *If the array manifold is such that $\mathbf{A}_{\text{ca},0}$ contains a Vandemonde matrix, and other rows of $\mathbf{A}_{\text{ca},0}$ are only repetitions of rows from that Vandermonde matrix (in other words, the co-array does not contain any "holes"), the condition $N_\theta \leq M_{\text{ca}}$ also becomes necessary for $\mathbf{J}_{\gamma\gamma}$ to be invertible. This happens for uniform and nested samplers, but not for coprime samplers.*

Remark 5. (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 γ can be written as follows

$$\begin{aligned}\frac{1}{L}\left(CRB_{\gamma\gamma}\right)^{-1} &= \frac{1}{L}\left(\mathbf{J}_{\gamma\gamma}-\mathbf{J}_{\gamma\delta}\mathbf{J}_{\delta\delta}^{-1}\mathbf{J}_{\gamma\delta}^{H}\right)\\ &= \mathbf{A}_{\text{ca}}^{H}\mathbf{W}\mathbf{A}_{\text{ca}}-\mathbf{A}_{\text{ca}}^{H}\mathbf{W}\mathbf{H}\left(\mathbf{H}^{H}\mathbf{W}\mathbf{H}\right)^{-1}\mathbf{H}^{H}\mathbf{W}\mathbf{A}_{\text{ca}}\\ &= \mathbf{A}_{\text{ca}}^{H}\mathbf{W}^{1/2}\Pi_{\mathbf{W}^{1/2}\mathbf{H}}^{\perp}\mathbf{W}^{1/2}\mathbf{A}_{\text{ca}},\end{aligned} \tag{2.31}$$

where $\Pi_{\mathbf{W}^{1/2}\mathbf{H}}^{\perp}$ is the projection matrix onto the null space of $\mathbf{W}^{1/2}\mathbf{H}$.

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 $N_\theta > |S_{\text{ca}}|$, $\mathbf{J}$ 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 3, $\mathbf{J}$ is singular. □

Remark 6. Singularity of the FIM matrix $\mathbf{J}$ implies that there exists no unbiased estimator for ψ with finite variance [SM01a]. 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, γ was not even assumed to be sparse; so in principle, the number of unknowns in γ 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 γ for it to be identifiable. *We will*

further elaborate on this point in Sec. 2.2.3.

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 δ, and our goal will be to study for what range of values of δ, we can argue its non singularity for *almost all* $\gamma \in \mathbb{R}^{N_\theta}$. We divide our analysis into two cases: In the first scenario, we find sufficient conditions under which $\mathbf{J}$ is invertible for almost all γ when $\delta = \mathbf{0}$. Based upon this result, we will argue that under the same conditions, $\mathbf{J}$ will be invertible for almost all $\delta \in \mathbb{R}^{M-1}$ as well.

Notice that studying the non singularity of $\mathbf{J}$ for $\delta = \mathbf{0}$ 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 $\mathbf{J}_{\gamma\gamma}$ of $\mathbf{J}$. Hence, the *unpertubed* problem is identifiable if and only if $\mathbf{J}_{\gamma\gamma}$ is invertible, which simply reduces to $\mathbf{A}_{\text{ca},0}$ being full column-rank. However, in our case, since δ is an unknown parameter, the invertibility of $\mathbf{J}_{\gamma\gamma}$ does not imply the invertibility of $\mathbf{J}$ at $\delta = \mathbf{0}$.

Non Singularity of J at $\delta = \mathbf{0}$ Establishing sufficient conditions under which $\mathbf{J}$ is non singular at $\delta = \mathbf{0}$ 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 sampling and a slightly modified version of nested sampling:

Theorem 4. *(***Uniform Sampling***) For uniform sampling, if $N_\theta \leq 2M-2$, $\mathbf{J}|_{\delta=0}$ is invertible for almost all $\gamma \in \mathbb{R}^{N_\theta}$.*

This implies a rather small grid size for the ULA. However, for nested samplers, the grid size (for which $\mathbf{J}|_{\delta=0}$ is guaranteed to be non singular) can be as large as $O(M^2)$. To prove this, we use a slightly modified version of the original nested sampler (assuming M is even), where the

sampling instants are given by d_i where

$$d_i = (i-1), \quad d_{i+\frac{M}{2}} = \frac{M}{2} i \tag{2.32}$$

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 [PV10], $M_{\text{ca}} = M^2/2 + M - 1$. 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 5. *(***Modified Nested Sampling***) For a modified nested sampler (given by* (2.32)*, with even M), if* $N_\theta \leq M^2/2$, $\mathbf{J}|_{\delta=0}$ *is invertible for almost all* $\gamma \in \mathbb{R}^{N_\theta}$.

Proof. The proofs can be found in Appendices 2.6.1 and 2.6.2. □

Remark 7. 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 2.6.2, with some modifications, which we avoid for ease of exposition.

Remark 8. 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 γ *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 sparse*.

Non Singularity of FIM: $\delta \neq \mathbf{0}$ The non singularity of $\mathbf{J}$ for almost all δ immediately follows from the conditions developed for $\delta = 0$:

Theorem 6. *For uniform and nested sampling,* $\mathbf{J}$ *is invertible for almost all* $\delta \in \mathbb{R}^{M-1}$ *and* $\gamma \in \mathbb{R}^{N_\theta}$*, if* $N_\theta \leq 2M - 2$ *(for uniform sampler) and* $N_\theta \leq M^2/2$ *(for modified nested sampler).*

Proof. Since elements of $\mathbf{J}$ are analytic functions of δ, $\det(\mathbf{J})$ is also an analytic function of δ. Therefore, as long as as $\det(\mathbf{J})$ is not trivially zero for all δ, $\det(\mathbf{J})$ has isolated zeros in δ [APP11]. However, in Theorems 4, and 5, we have shown that for uniform and modified nested sampling, $\det(\mathbf{J}) \neq 0$ at $\delta = 0$ 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 $\forall \delta$. Therefore, the zeros of $\det(\mathbf{J})$ are isolated in $\mathbb{R}^{M-1}$, which has a total measure of zero. Thus, for almost all $\delta \in \mathbb{R}^{M-1}$ and $\gamma \in \mathbb{R}^{N_\theta}$, $\det(\mathbf{J}) \neq 0$, 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 spectral lines K to be known (or, equivalently, do not assume the vector of source powers, γ, to be sparse), $N_\theta \leq 2M-2$ is sufficient for $\mathbf{J}$ to be non singular for almost all choices of γ and δ.
- For nested sampliers, under the same assumption of K to be unknown, we can ensure the invertibility of $\mathbf{J}$ for almost all choices of γ and δ using a much larger overcomplete dictionary, where $N_\theta = O(M^2)$.

2.2.3 Non Singularity of FIM for Sparse Vectors

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

Non singularity for small grid size

In this case, we assume that $N_\theta \leq 2M-2$ for uniform samplers, and $N_\theta \leq M^2/2$ for the modified nested sampler. We show that the Fisher Information Matrix is invertible at almost all

sparse γ. In particular, we have the following result:

Theorem 7. *Assume* $\delta = \mathbf{0}$ *and consider the grid based model* (2.13)*. For almost all sparse vectors* $\gamma \in \mathbb{R}^{N_\theta}$ *with* $\|\gamma\|_0 = K, K < N_\theta$*, where* $N_\theta \leq 2M-2$ *for uniform samplers and* $N_\theta \leq \frac{M^2}{2}$ *for modified nested samplers,* $\mathbf{J}$ *is invertible for almost all* $\delta \in \mathbb{R}^{M-1}$*.*

Proof. The proof can be found in the Appendix 2.6.3. □

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 γ also becomes non identifiable in this case, and it becomes necessary to assume priors (such as sparsity) on γ to render it identifiable.

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

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 (2.23), uniqueness of the vectorized covariance matrices implies

$$\mathbf{A}_{\text{ca}}(\delta)\gamma = \mathbf{A}_{\text{ca}}(\delta')\gamma' \Leftrightarrow \delta = \delta', \gamma = \gamma' \tag{2.33}$$

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

1. $\delta = \mathbf{0}$: In this case, $\mathbf{A}_{\text{ca},0}$ contains a Vandermonde matrix with M_{ca} distinct rows. One way to ensure identifiability of γ is to assume that it is $K-$ sparse (or, equivalently, assume the number of sources to be known). In such a case, $\mathbf{A}_{\text{ca},0}\gamma = \mathbf{z}$ will permit a *unique solution* in γ, if $K < \frac{\text{k-rank}(\mathbf{A}_{\text{ca}})}{2}$, (see [DET06]), 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 γ for $\delta = 0$, by assuming it to be sparse and ensuring that $K < M_{\text{ca}}/2$.

2. $\delta \neq \mathbf{0}$: Finding 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 δ 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 δ, γ and the smallest singular value of the unperturbed manifold $\mathbf{A}_{\text{ca},0}$, as discussed next.

Sufficient conditions for identifiability for small perturbations

In this section, we derive sufficient conditions for (2.33) to hold, in terms of an upper bound for δ.

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

$$\begin{bmatrix} \text{diag}(\delta) - \delta_1 \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \text{diag}(\delta) - \delta_2 \mathbf{I} & \cdots \mathbf{0} & \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \text{diag}(\delta) - \delta_M \mathbf{I} \end{bmatrix} \tag{2.34}$$

Also, define Δ' by replacing the vector δ with $\delta' \in \mathbb{R}^M$ in (2.34).

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

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

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

$$\mathbf{A}_{\text{ca}}(\delta) \simeq \mathbf{A}_{\text{ca},0} + \Delta\mathbf{A}_{\text{ca},0}\Upsilon \tag{2.35}$$

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

For this linearized model, we now proceed to establish sufficient conditions such that (2.33) holds. We assume that the number of sources is known to be at most K, so that all vectors γ in our ambiguity set are at most $K-$sparse. Suppose there exists $\delta' \neq \delta$ and $\gamma' \neq \gamma$ (both at most $K-$ sparse) such that

$$(\mathbf{A}_{\text{ca},0} + \Delta\mathbf{A}_{\text{ca},0}\Upsilon)\gamma = (\mathbf{A}_{\text{ca},0} + \Delta'\mathbf{A}_{\text{ca},0}\Upsilon)\gamma' \tag{2.36}$$

Let S, S' denote the supports of γ, γ', 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 $\gamma_{\mathbf{i}}$ (or $\gamma'_{\mathbf{i}}$) be the vector comprised by the elements of γ (or γ') which are indexed by $S_{\mathbf{i}}$, where $\mathbf{i} = 1, 2, 12$. We can rewrite (2.36) as

$$(\tilde{\mathbf{A}}_{\text{ca},0} + \Delta\tilde{\mathbf{A}}_{\text{ca},0}\tilde{\Upsilon}) \begin{bmatrix} \gamma_1 \\ \gamma_{12} \\ \mathbf{0} \end{bmatrix} = (\tilde{\mathbf{A}}_{\text{ca},0} + \Delta'\tilde{\mathbf{A}}_{\text{ca},0}\tilde{\Upsilon}) \begin{bmatrix} \mathbf{0} \\ \gamma'_{12} \\ \gamma'_2 \end{bmatrix}$$

which is equivalent to

$$\tilde{\mathbf{A}}_{\text{ca},0} \begin{bmatrix} \gamma_1 \\ \gamma_{12} - \gamma'_{12} \\ -\gamma'_2 \end{bmatrix} = -\Delta\tilde{\mathbf{A}}_{\text{ca},0}\tilde{\Upsilon} \begin{bmatrix} \gamma_1 \\ \gamma_{12} \\ \mathbf{0} \end{bmatrix} + \Delta'\tilde{\mathbf{A}}_{\text{ca},0}\tilde{\Upsilon} \begin{bmatrix} \mathbf{0} \\ \gamma'_{12} \\ \gamma'_2 \end{bmatrix} \tag{2.37}$$

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

$$\|LHS\| \geq \sigma_{\min}(\tilde{\mathbf{A}}_{\text{ca},0})\sqrt{k_1}\gamma_{\min} \tag{2.38}$$

Moreover,

$$\|RHS\| \leq \sigma_{\max}(\Delta)\sigma_{\max}(\tilde{\mathbf{A}}_{\text{ca},0}\tilde{\Upsilon})\sqrt{k_1+k_{12}}\gamma_{\max} \tag{2.39}$$

$$+\sigma_{\max}(\Delta')\sigma_{\max}(\tilde{\mathbf{A}}_{\text{ca},0}\tilde{\Upsilon})\sqrt{k_2+k_{12}}\gamma_{\max} \tag{2.40}$$

in which LHS, RHS refer to the left hand side and right hand side of the equation (2.37), 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 $\sigma_{\max}(\Delta) \leq 2\delta_{\max}$.

Hence, a sufficient condition for identifiability is

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

From (2.38) and (2.39) 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

$$\delta_{\max} < \frac{\sigma_{\min}(\tilde{\mathbf{A}}_{\text{ca},0})}{4\pi M}\frac{\gamma_{\min}}{\sqrt{2K}\gamma_{\max}} \tag{2.41}$$

Considering all possible supports, (2.41) is satisfied if

$$\delta_{\max} < \frac{\tilde{\sigma}_{\min}}{4\pi\sqrt{2K}M}\frac{\gamma_{\min}}{\gamma_{\max}} \tag{2.42}$$

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 8. *Suppose* $|\delta_m| \leq \delta_{\max}$ *for* $m = 1, \cdots, M$*, and* $\delta_{\max}$ *to be small so that we can approximate* $\mathbf{A}_{\text{ca}}(\delta)$ *as* (2.35)*. Moreover, assume that* $\|\gamma\|_0 \leq K$ *and the non-zero elements of* γ *lie in the range* $[\gamma_{\min} \ \ \gamma_{\max}]$*. The parameters* $[\gamma, \delta]$ *are identifiable for any grid size* N_θ*, if the maximum perturbation value obeys* (2.42)*.*

2.2.4 Effect of Finite Temporal Samples on Cramér-Rao Bound

Given the unified signal model (2.14) for spatial and temporal line spectrum estimation, we now study the performance of sparse (such as nested) and uniform samplers as a function of the number (L) of temporal snapshots. For temporal line spectrum, we fix the number of samples (M) in each block, and study the behavior of the CRB as we increase the total number of blocks L. As argued earlier, since the support of γ reveals the source directions (on the grid), successful recovery of γ also ensures successful recovery of source directions for the grid based model.

Recall from (2.14) that for spatial spectrum estimation, we consider L time snapshots of the signal received at an array of M sensors. Similarly, for temporal line spectrum estimation, we collect a total of ML measurements over L blocks, with M samples in each block. For both cases, we have the following measurement model:

$$\mathbf{y}[l] = \mathbf{A}_{\text{grid},0}\mathbf{x}[l] + \mathbf{w}[l], 1 \leq l \leq L$$

We assume that the L temporal snapshots (or equivalently, the measurements across L blocks) are independent. In particular, we assume that $\{\mathbf{y}[l]\}_{l=1}^{L}$ are L i.i.d random vectors distributed as

$$\mathbf{y}[l] \overset{\text{i.i.d}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{A}_{\text{grid},0}\mathbf{R}_{\mathbf{x}}\mathbf{A}_{\text{grid},0}^{H} + \sigma_w^2\mathbf{I}), \tag{2.43}$$

where $\mathbf{R}_{\mathbf{x}} = \text{diag}(\boldsymbol{\gamma})$. Since our goal in this section is to study the effect of L on the Cramér Rao bound, we consider an unperturbed model (i.e. $\boldsymbol{\delta} = 0$), and hence, the FIM is a function of only the parameter $\boldsymbol{\gamma}$. From (2.28), the FIM for $\boldsymbol{\gamma}$ is readily given by

$$\mathbf{J}_{\boldsymbol{\gamma}\boldsymbol{\gamma}} = L\left(\mathbf{A}_{\text{ca},0}^{H}\mathbf{W}\mathbf{A}_{\text{ca},0}\right) \tag{2.44}$$

where $\mathbf{A}_{\text{ca},0}$ and $\mathbf{W}$ are not functions of L. Therefore, we have the following theorem on the CRB of $\boldsymbol{\gamma}$ as a function of L:

Theorem 9. *Consider the measurement model (2.43), where L either represents the number of temporal snapshots collected at an array of M sensors (for spatial line spectrum estimation), or the number of blocks or periods (with M samples in each period) over which measurements are collected using a periodic non-uniform sampler (for temporal line spectrum estimation). Assuming that the measurements* $\mathbf{y}[l], 1 \leq l \leq L$ *are distributed as (2.43), the CRB of* $\boldsymbol{\gamma}$ *exists, and is proportional to* $1/L$ *if and only if* $N_{\theta} \leq M_{\text{ca}}$.

Proof. The CRB exists if and only if $\mathbf{J}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}$ is non-singular, which is true if and only if $N_{\theta} \leq M_{\text{ca}}$ (see Corollary 1). In this case, $\mathbf{J}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1} = \frac{1}{L}\left(\mathbf{A}_{\text{ca},0}^{H}\mathbf{W}\mathbf{A}_{\text{ca},0}\right)^{-1}$ which establishes that the CRB is proportional to $1/L$. □

Since the CRB goes to zero as $L \to \infty$ (as long as $N_{\theta} \leq M_{\text{ca}}$), it also implies that asymptotically in L, the Maximum Likelihood Estimator of $\boldsymbol{\gamma}$ (which asymptotically attains the CRB) will correctly identify $\boldsymbol{\gamma}$ and therefore recover the source directions as well (by identifying the support of $\boldsymbol{\gamma}$).

Role of Array Geometry: The array geometry plays a role in ensuring the existence of the CRB, since M_{ca} is completely determined by the geometry. Hence, for a ULA, the CRB exists if and only if $N_\theta \leq 2M-1$ whereas for nested array, the CRB exists if and only if $N_\theta \leq M^2/2+M-1$. However, in the regime $N_\theta \leq M_{\text{ca}}$, the CRB decays to 0 with $O(1/L)$ *irrespective of the array geometry*.

Need for independence across blocks: For line spectrum estimation, the CRB behaves differently with respect to M (number of samples within a block) and L (the number of blocks). The assumption of independence across blocks is critical to ensure that the CRB decays to 0 as $L \to \infty$ (at the rate $1/L$), when M is held constant. However, the CRB does not necessarily decay to zero if we instead let $M \to \infty$ (and hold L constant). This is because within a block, the M samples are highly correlated, which can lead to saturation of the CRB at a non zero value as $M \to \infty$. We will exhibit this effect in our numerical simulations.

2.2.5 Experimental Results

We conduct two sets of experiments to 1) Study the effect of perturbed sampling on Cramér Rao bounds for line spectrum estimation with sparse and uniform samplers, and 2) Study the behavior of Cramér Rao Bounds with increasing the number of temporal samples for temporal spectrum estimation.

Perturbation in Spatial Line Spectrum Estimation

We evaluate the Cramér Rao bound corresponding to the parameter γ in presence of perturbation δ, and compare it with the RMSE of a maximum likelihood estimator for jointly estimating γ and δ. Following [VT04, Section 8.5], the log-likelihood function corresponding to our problem can be written as

$$L(\delta,\gamma) = -[\ln\det \mathbf{R}_{\mathbf{y}} + \text{tr}(\mathbf{R}_{\mathbf{y}}{}^{-1}\hat{\mathbf{R}}_{\mathbf{y}})] \tag{2.45}$$

where $\mathbf{R}_{\mathbf{y}} = \mathbf{A}_{\text{grid}}(\delta)\,\text{diag}(\gamma)\mathbf{A}_{\text{grid}}(\delta)^H + \sigma_w^2\mathbf{I}$ 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 assume that the noise variance σ_w^2 is known. We find the optimum values for γ and δ by maximizing $L(\delta,\gamma)$ subject to the constraint $\gamma \geq \mathbf{0}$, using `fmincon` function of MATLAB. Notice that the sparsity of γ is not exploited since we do not use a regularized log likelihood function.

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, 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 noise is fixed at $\sigma_w^2 = 0.1$ in all the experiments. We assume the spatial perturbations to be $\delta = \alpha\delta_{\mathbf{0}}$, where

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

and α is a scalar, which determines the strength of the perturbation. We define RMSE of the maximum likelihood estimator as $\sqrt{\sum_{i=1}^{N_{\text{tests}}} \frac{\|\hat{\gamma}-\gamma\|^2}{N_\theta N_{\text{tests}}}}$, where $\hat{\gamma}$ is the estimated γ, 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 (2.31).

In the first simulation, we choose $N_\theta = 35$. We consider two scenarios: in the first setting, we have as many sources as the grid size (i.e. $K = N_\theta$), all with unit power. In the second case, (which we will refer to as the *sparse* setting), we assume that there are only $K = 4$ active sources with powers equal to one. In this case, the support of γ is given by $S = \{3,7,11,16\}$. As stated earlier, the ML algorithm does not assume γ to be a sparse vector *apriori*. Fig. 2.5 shows the

RMSE of the ML estimate relative to the Cramér Rao bound for both sparse and non-sparse settings. The label "n" indicates nested array while the label "c" corresponds to the coprime array, both with $M = 14$ sensors. Furthermore the label "s" indicate the case where γ is sparse with $K = 4$. Since $N_\theta > 2M - 1$, the CRB for ULA does not exist and is not plotted. It can be seen that the CRB for both coprime and nested arrays do not show significant variation over the range of α considered. However, the CRB for $K = 4$ (sparse setting) is significantly lower than that for $K = N_\theta$ (non-sparse setting) However, as we increase the number of snapshots L, the RMSE of the ML estimator can be seen to decrease as expected, and its MSE approaches the corresponding Cramér Rao bound. Moreover, the CRB for the *sparse* setting is smaller than that for the non-sparse model, although sparsity is not exploited as a prior. In Fig. 2.6, we compare the CRB for ULA, nested and coprime arrays as a function of α when $N_\theta = 25$. Since $N_\theta = 25 < 2M - 2$, the CRB for ULA exists and can be compared against that for nested and coprime arrays. It can be seen that the coprime array exhibits the lowest CRB over this range of α whereas the ULA has the highest CRB.

Our experiments so far indicate that in the regime $N_\theta < 2M - 1$, the CRB for both ULA and sparse arrays are very close to each other. However, the main distinction between these samplers is in the way their CRBs behave as the *grid size* N_θ *increases*. Our second experiment demonstrates this, where we compare the CRB of ULA, nested and coprime arrays as a function of N_θ (keeping M constant) and depict the result in Fig. 2.7. Here, $\alpha = 0.5$ and $L = 10^2$. It can be seen that beyond a certain grid size, the CRB corresponding to each type of array suddenly jumps to very large values, indicating that the FIM becomes singular beyond that point. We also observe that the value of N_θ at which this happens, is much smaller for the ULA than that for the nested and coprime arrays. 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 $N_\theta > 2M - 1 = 27$. Also, the FIM for coprime arrays becomes singular at a smaller grid size compared to a nested array with same number of sensors.

In the third simulation, we examine the probability of FIM being invertible for different number of sensors and grid sizes for ULA and the nested array. For this experiment, we again consider $K = N_\theta$ sources located on the grid points, all with powers equal to one. In each trial, we randomly generate a δ 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 2.8. 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 $\delta = \mathbf{0}$ 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 4, 5 (We used the bound proposed in Remark 7 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 4 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 $\delta = \mathbf{0}$. 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 $\mathbf{J}\big|_{\delta \neq \mathbf{0}}$ is invertible, although $\mathbf{J}\big|_{\delta = \mathbf{0}}$ is not. This happens due to the fact that the perturbations can slightly increase the rank of $\mathbf{A}_{\text{ca}}(\delta)$. Therefore, $\mathbf{J}$ can be invertible even though $\mathbf{A}_{\text{ca},0}$ is not full column rank.

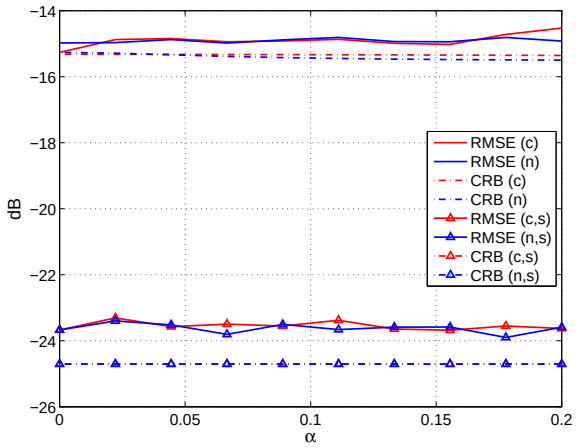
Effect of Finite Temporal Samples on CRB

We now study the CRB as functions of both L (number of independent blocks or snapshots) and M (the number of samples within each block) and demonstrate that it behaves quite differently with respect to these two quantities. The behavior of the CRB with respect to L is depicted in Fig. 2.5 (b), which shows that as we increase L, the CRB monotonically decays to zero at the rate $1/L$, verifying the claim in Theorem 9. We next investigate the effect of increasing the number (M) of measurements in each block, while keeping L fixed at $L = 1$. We choose $N_\theta = 40$, and

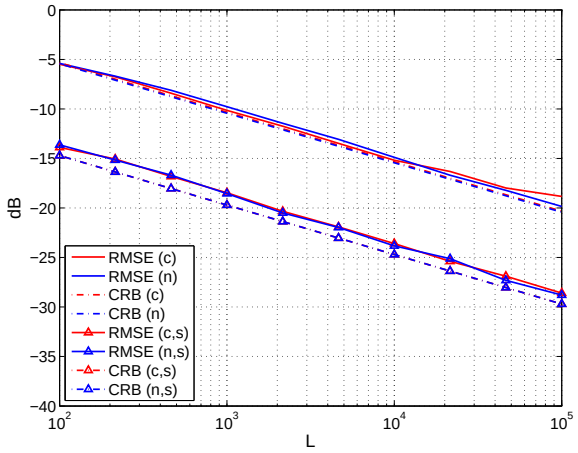
$\sigma_w^2 = 0.25$. The power of the active sources are assumed to be equal to 1, and they are uniformly located in the range $[-\frac{\pi}{6} \quad \frac{\pi}{6}]$. In Fig. 2.9, we separately plot the CRBs corresponding to the zero and the non-zero elements of γ, as functions of M. As can be seen, the CRB corresponding to the zero and non-zero elements of γ exhibit completely different behaviors. For the chosen parameters, the CRB for the zero elements monotonically decrease as we increase M. However, the CRB corresponding to the non-zero elements do not decrease below a certain non-zero limit (which is 1 in this case) and they essentially *saturate*. This means that the CRB corresponding to non-zero powers remain strictly lower bounded by a positive constant even as $M \to \infty$. Hence, by increasing the size of the blocks (for temporal line spectrum estimation), or the number of antennas (for spatial line spectrum estimation), it is not possible to reliably estimate the parameter γ. We will analytically characterize this interesting behavior in future.

2.2.6 Conclusion

In this section, 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 are treated 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, the FIM continues to be non singular as long as the grid size is $O(M^2)$. We separately considered the case of sparse sources and established sufficient conditions for identifiability even when the FIM becomes singular. We also studied the effect of finite number of temporal samples on the CRB, for both temporal and spatial line spectral estimation, and showed that the CRB converges to zero as we increase the number of snapshots for spatial line spectrum estimation (or equivalently, the number of blocks for temporal line spectrum estimation).



(a) The RMSE of recovered γ vs α ($L = 10^4$)



(b) The RMSE of recovered γ vs L ($\alpha = 0.1$).

Figure 2.5: The RMSE of ML estimate 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).

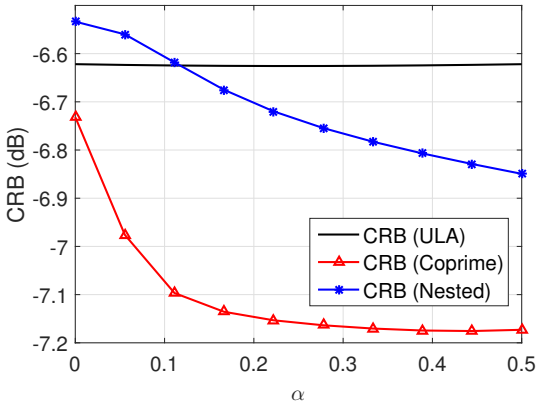


Figure 2.6: The CRB corresponding to ULA, nested and coprime arrays vs α, where $L = 10^2$, $M = 14$, and $N_\theta = 25$.

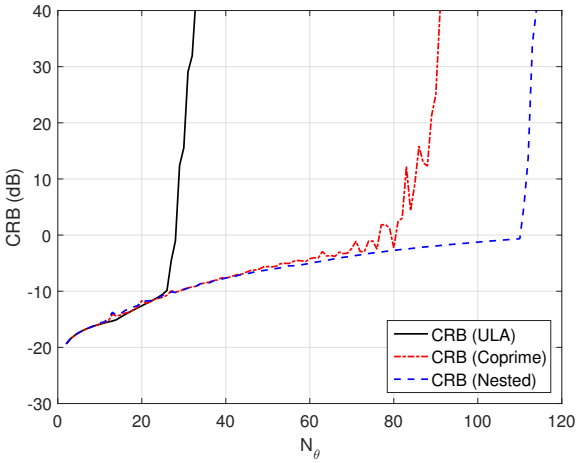
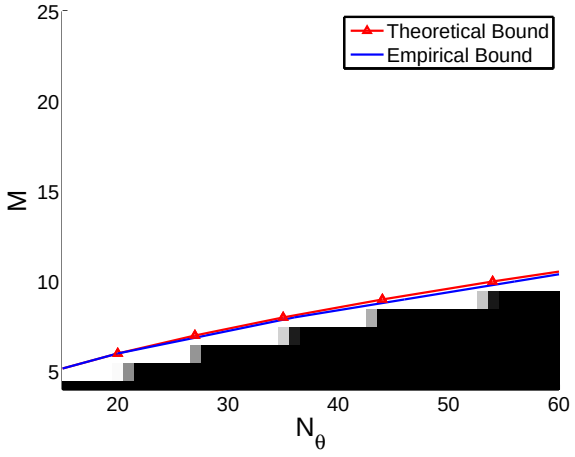
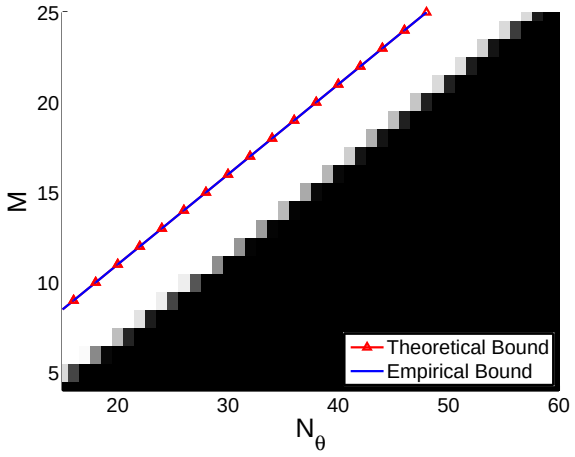


Figure 2.7: The CRB vs N_θ for nested, co-prime (with co-prime numbers $4, 7$) and ULA with $M = 14$ sensors ($\alpha = 0.5, L = 10^2$).

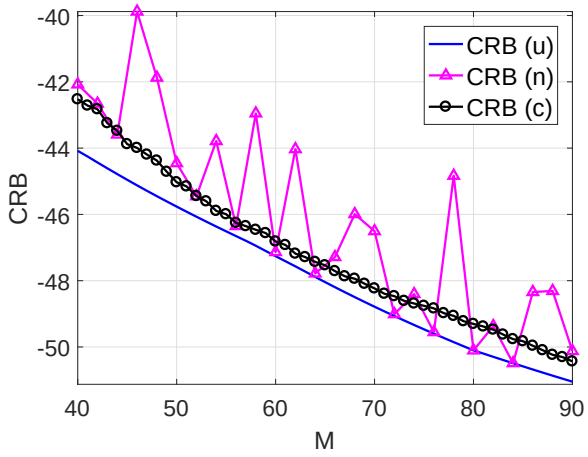


(a) Nested Array

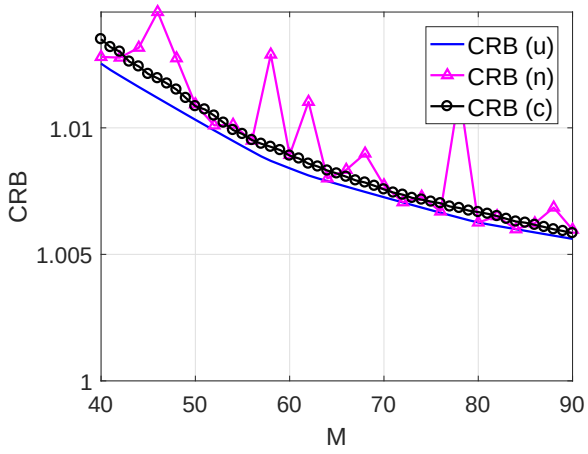


(b) ULA

Figure 2.8: The probability of FIM being invertible over random choices of δ for different number of sensors (M) and different grid sizes (N_θ). The blue line indicates the M above which $\mathbf{J}\big|_{\delta=\mathbf{0}}$ is invertible. The red line shows the theoretical bounds we derived in Theorems 4, 5.



(a) Zero Elements



(b) Nonzero Elements

Figure 2.9: The CRB for different samplers as a function of M, when the number of blocks (or snapshots) is held constant at $L = 1$. Here, "(u)", "(n)", and "(c)" denote ULA, nested, and coprime arrays respectively.

2.3 Saturation of the Cramér Rao Bound

In this section, we show that the CRB exhibits saturation with respect to the number of spatial measurements, i.e., it can be lower bounded by a non-negative quantity that does not go to zero even when the number of spatial measurements tends to infinity. Moreover, the CRB corresponding to the nonzero and zero elements of the sparse hyper-parameter can exhibit different behaviors. While the CRB for the non-zero elements always *saturate* regardless of the type of dictionary, saturation of the CRB for zero elements provably happens when the dictionary has normalized columns. For an unnormalized dictionary, singular values of certain sub-dictionaries determine if saturation can happen, prompting future research into this interesting phenomenon.

We consider a more general setting compared to the Sections 2.3, such that our problem model can be also applicable to the more general problem of Sparse Bayesian Learning (SBL) [Tip01, TF$^+$03, WR04, WR07], and is not restricted to Direction-of-Arrival estimation type of problems. Here, we briefly describe this more general problem setting:

Sparse Bayesian Learning (SBL) constitutes an important family of Bayesian algorithms where the goal is to estimate a sparse signal $\mathbf{x} \in \mathbb{F}^N$, from compressed measurement $\mathbf{y} \in \mathbb{F}^M$ acquired as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w} \tag{2.46}$$

Here $\mathbf{A} \in \mathbb{F}^{M \times N}$ ($M < N$) denotes an underdetermined dictionary and $\mathbf{w} \in \mathbb{F}^{M \times 1}$ denotes the additive noise. Throughout this section, $\mathbb{F}$ can be either the set of real ($\mathbb{R}$) or complex ($\mathbb{C}$) numbers. Unlike traditional Compressed Sensing algorithms [C$^+$06, Don06] that only exploit the sparsity of $\mathbf{x}$ to solve the ill-posed problem (2.46), SBL algorithms impose a suitable prior distribution on $\mathbf{x}$ (that also models its sparsity) and computes the corresponding posterior estimate. Alongside recovering $\mathbf{x}$, SBL algorithms also allow estimation of certain hyper-parameters characterizing

the prior distribution of $\mathbf{x}$ that crucially control its sparsity as well as correlation structure [ZR11].

The authors in [PM13] investigated fundamental performance limits of the SBL framework by deriving appropriate Cramér Rao Bounds (CRB) on the mean-squared error (MSE) of SBL estimators for $\mathbf{x}$ and associated hyper-parameters. However, the analytical behavior of these bounds as a function of number of measurements M, have not been investigated so far. Of particular interest would be to understand if increasing the number of measurements M enables us to estimate the hyper-parameters with proportionately decreasing MSE that converge to 0? As we will show in this Section, the answer is a surprising "no", implying that even when M goes to infinity, the CRB does not decrease below a certain positive quantity, and it essentially *saturates*. Hence, even with infinite measurements, no *unbiased* estimator exists that can *exactly* recover the hyper-parameter. We mathematically characterize this *saturation* behavior of the CRB corresponding to both zero and nonzero elements of the sparse hyper-parameter.

Related Work. Cramér-Rao Bounds for estimating sparse signals in presence of noise have been derived in [BKT09, BHE10]. However, these results do not consider a stochastic model (or prior) for $\mathbf{x}$ and hence cannot be applied for analyzing SBL. In [PM13], for the first time, CRB expressions for the SBL framework were derived, assuming different statistical models. In [PV14a], the authors proved that the sparse hyper-parameter can be identifiable even when the number of non-zero elements of $\mathbf{x}$ exceed M, and derived corresponding CRB expressions. This paper conducts further analysis of the CRB for sparse hyper-parameters, and mathematically justifies its saturation behavior (which manifests differently for the non-zero and zero elements).

2.3.1 Statistical Model for SBL

Assume that the signal $\mathbf{x}$ is a random vector distributed as $\mathbf{x} \sim \mathbb{F}\mathcal{N}(\mathbf{0}, \mathbf{P})$ where $\mathbf{P} = \text{diag}(p_1, p_2, \cdots, p_N)$, and we denote $\mathbf{p} = [p_1, \cdots, p_N]^T$ as the vector of hyper-parameters, representing the power of the elements of $\mathbf{x}$. The vector $\mathbf{p}$ is assumed to be sparse where $\mathcal{S}$ denotes the support of $\mathbf{p}$, with $|\mathcal{S}| = K$, i.e., $\mathbf{p}$ contains only K nonzero numbers. Furthermore, let $\mathbf{w}$ represent

white Gaussian noise with distribution $\mathbb{F}\mathcal{N}(\mathbf{0}, \sigma_w^2\mathbf{I})$, which is uncorrelated with $\mathbf{x}$. We henceforth assume that the noise variance σ_w^2 is also known. Under these assumptions, one can write the probability density function (pdf) of $\mathbf{y}$ as

$$p_{\mathbf{Y};\mathbf{p}}(\mathbf{y};\mathbf{P}) = [(2\pi)^M(\det \mathbf{R})]^{-\eta_{\mathbb{F}}} e^{-\eta_{\mathbb{F}}\mathbf{y}^H\mathbf{R}^{-1}\mathbf{y}} \tag{2.47}$$

where $\mathbf{R}$ is the covariance matrix of the random variable $\mathbf{y}$, given by

$$\mathbf{R} = \mathbf{A}\mathbf{P}\mathbf{A}^H + \sigma_w^2\mathbf{I},$$

and $\eta_{\mathbb{F}} = 1$ for $\mathbb{F} = \mathbb{C}$, and $\eta_{\mathbb{F}} = \frac{1}{2}$ for $\mathbb{F} = \mathbb{R}$.

In SBL, both the signal $\mathbf{x}$, and the hyper-parameter $\mathbf{p}$ can be recovered by respectively solving the so-called Type I, and Type II estimation problems. In this Section, we are primarily interested in estimating the underlying hyper-parameter $\mathbf{p}$ that characterize the signal distribution (as well as its sparsity).

Review of Cramér Rao Bounds For Hyperparameter Estimation

It is well known that the Cramér Rao Bound (CRB) serves as a fundamental lower bound on the Mean Squared Error (MSE) of *any* unbiased estimator for a (deterministic) parameter. In [PM13], various CRB expressions (such as Hybrid, Bayesian and Marginalized CRBs) are derived under different statistical assumptions and models. Since our goal in this Section is to analyze the CRB for the hyperparameter $\mathbf{p}$, we consider the marginalized CRB (MCRB) for $\mathbf{p}$ derived in [PM13]. The authors in [PM13] also show that among all CRB expressions, the MCRB provides the tightest lower bound.

The MCRB for $\mathbf{p}$ can be derived using the marginalized distribution of $\mathbf{y}$ given by (2.47) and assuming that the noise power σ_w^2 is known. In this case, $\mathbf{p}$ is the only unknown parameter characterizing $p_{\mathbf{Y};\mathbf{p}}(\mathbf{y};\mathbf{P})$, and the corresponding Fisher Information Matrix (FIM) $\mathbf{J}$ can be shown

to be of the form [PV14a]

$$\mathbf{J} = \eta_{\mathbb{F}} \mathbf{A}_{\text{ca}}^H \mathbf{W} \mathbf{A}_{\text{ca}} \tag{2.48}$$

where $\mathbf{A}_{\text{ca}} = \mathbf{A}^* \odot \mathbf{A}$, $\mathbf{W} = \mathbf{R}^{-T} \otimes \mathbf{R}^{-1}$. The matrix $\mathbf{A}_{\text{ca}}$ denotes the Khatri-Rao product (or column-wise Kronecker product) of the dictionary $\mathbf{A}$ and crucially controls important properties of $\mathbf{J}$. By considering the rank of $\mathbf{A}_{\text{ca}}$, the authors in [PV14a] have been able to provide the following necessary and sufficient condition under which the MCRB for $\mathbf{p}$ exists:

Theorem 10. *[PV14a] The FIM* $\mathbf{J}$ *given in* (2.48) *is non-singular if and only if* $N = \text{rank}(\mathbf{A}_{\text{ca}})$.

Hence, as long as $N = \text{rank}(\mathbf{A}_{\text{ca}})$ (which can imply $N = O(M^2)$ for certain dictionaries), the CRB exists and can be used to lower bound the MSE of any unbiased estimate of $\mathbf{p}$.

2.3.2 Saturation of the MCRB

For many overdetermined estimation problems ($N \leq M$), the CRB typically converges to 0 asymptotically as the number of measurements $M \to \infty$, implying that the parameter can be estimated with zero MSE (as $M \to \infty$) using appropriate estimators (such as Maximum Likelihood Estimator). However, we will now show that the MCRB for SBL (that typically involves a compressive measurement model with $N > M$) can saturate at a value strictly bounded away from zero, even when $M \to \infty$. This behavior implies that it is not possible to find an unbiased estimator that can recover $\mathbf{p}$ with zero MSE as the number of measurements grows infinitely large. In this regard, we will show that the *non-zero and zero* elements of $\mathbf{p}$ exhibit different saturation behavior as follows:

(i) CRB of Non-Zero Elements: For all values of N and M and all choices of the dictionary $\mathbf{A}$, the CRB corresponding to the nonzero elements of $\mathbf{p}$ always exhibit a saturation effect, meaning that we can find a lowerbound for the CRB (in terms of M) that tends to a strictly positive limit as $M \to \infty$.

(ii) CRB of Zero Elements: The CRB corresponding to the zero elements of $\mathbf{p}$ can be lower bounded by a *non zero quantity* (even when $M \to \infty$) as long as the columns of the dictionary $\mathbf{A}$ are normalized. If the columns of $\mathbf{A}$ are not normalized, saturation of the CRB is shown to be determined by the singular values of certain submatrices of $\mathbf{A}$.

Saturation of the CRB Corresponding to Nonzero Elements

Let $\mathbf{C} = \mathbf{J}^{-1}$, where the ith diagonal element of $\mathbf{C}$ provides a lower bound on the MSE of any unbiased estimator for $[\mathbf{p}]_i$, i.e., given any unbiased estimate $\hat{\mathbf{p}}(\mathbf{y})$ (which is a function of only the measurement $\mathbf{y}$) of $\mathbf{p}$, we have

$$\mathbb{E}_{\mathbf{y}}|[\mathbf{p}]_i - [\hat{\mathbf{p}}(\mathbf{y})]_i|^2 \geq [\mathbf{C}]_{ii}$$

The following theorem shows that if $i \in \mathcal{S}$ (i.e., $[\mathbf{p}]_i > 0$), then $[\mathbf{C}]_{ii}$ is strictly bounded away from zero.

Theorem 11. *Consider the model (2.46), where the measurement* $\mathbf{y}$ *is distributed according to (2.47). If* $N = rank(\mathbf{A}_{\text{ca}})$*, the CRB corresponding to the unknown parameter* $\mathbf{p}$ *satisfies*

$$[\mathbf{C}]_{ii} \geq \eta_{\mathbb{F}}^{-1} p_i^2, \quad i \in \mathcal{S}$$

Proof. Since $N = \text{rank}(\mathbf{A}_{\text{ca}})$, $\mathbf{J}$ is invertible and the CRB exists. Following the analysis in [SM93] and (2.48), it can be shown that ith diagonal element of the $\mathbf{C}$ can be written as

$$[\mathbf{C}]_{ii}^{-1} = \eta_{\mathbb{F}} \|\Pi^{\perp}_{\mathbf{W}^{1/2}\mathbf{A}_{\text{ca}}^{(-i)}} \mathbf{W}^{1/2} \mathbf{a}_{\text{ca}}^{(i)}\|^2$$

where $\mathbf{a}_{\text{ca}}^{(i)} = \mathbf{a}_i^* \otimes \mathbf{a}_i$,

$$\mathbf{A}_{\text{ca}}^{(-i)} = [\mathbf{a}_{\text{ca}}^{(1)}, \mathbf{a}_{\text{ca}}^{(2)}, \cdots, \mathbf{a}_{\text{ca}}^{(i-1)}, \mathbf{a}_{\text{ca}}^{(i+1)}, \cdots, \mathbf{a}_{\text{ca}}^{(N)}].$$

In other words, $\mathbf{A}_{\text{ca}}^{(-i)}$ contains a total of $N-1$ columns that excludes the column $\mathbf{a}_{\text{ca}}^{(i)}$. Furthermore, given any matrix $\mathbf{B}$ with full column rank, $\Pi_{\mathbf{B}}^{\perp} = \mathbf{I} - \mathbf{B}(\mathbf{B}^H\mathbf{B})^{-1}\mathbf{B}^H$ denotes the projection onto the orthogonal complement of range space of $\mathbf{B}$. Therefore, we can write

$$\begin{aligned}\eta_{\mathbb{F}}^{-1}[\mathbf{C}]_{ii}^{-1} &= (\mathbf{a}_{\text{ca}}^{(i)})^H \Big(\mathbf{W} - \\ &\qquad \mathbf{W}\mathbf{A}_{\text{ca}}^{(-i)}((\mathbf{A}_{\text{ca}}^{(-i)})^H\mathbf{W}\mathbf{A}_{\text{ca}}^{(-i)})^{-1}(\mathbf{A}_{\text{ca}}^{(-i)})^H\mathbf{W}\Big)\mathbf{a}_{\text{ca}}^{(i)} \\ &\leq (\mathbf{a}_{\text{ca}}^{(i)})^H\mathbf{W}\mathbf{a}_{\text{ca}}^{(i)} = |\mathbf{a}_i^H\mathbf{R}^{-1}\mathbf{a}_i|^2. \end{aligned} \tag{2.49}$$

where the last equality can be verified using algebraic properties of the Kronecker product. Since $i \in \mathcal{S}$, we can decompose $\mathbf{R}$ as

$$\mathbf{R} = \tilde{\mathbf{A}}_i\tilde{\mathbf{P}}_i\tilde{\mathbf{A}}_i^H + p_i\mathbf{a}_i\mathbf{a}_i^H + \sigma_w^2\mathbf{I}$$

where $\tilde{\mathbf{A}}_i$ is comprised of columns of $\mathbf{A}$ indexed by $\mathcal{S} \setminus i$, and $\tilde{\mathbf{P}}_i$ is a diagonal matrix composed of the corresponding elements of $\mathbf{p}$. Let us denote $\mathbf{R}_i := p_i\mathbf{a}_i\mathbf{a}_i^H + \sigma_w^2\mathbf{I}$. Using Woodbury's matrix identity [Woo50], we have

$$\mathbf{R}^{-1} = \mathbf{R}_i^{-1} - \mathbf{R}_i^{-1}\tilde{\mathbf{A}}_i(\tilde{\mathbf{P}}_i^{-1} + \tilde{\mathbf{A}}_i^H\mathbf{R}_i^{-1}\tilde{\mathbf{A}}_i)^{-1}\tilde{\mathbf{A}}_i^H\mathbf{R}_i^{-1}. \tag{2.50}$$

We can further use the Sherman-Morison [SM50] formula to get

$$\mathbf{R}_i{}^{-1} = \sigma_w^{-2}\mathbf{I} - \frac{\sigma_w^{-4}p_i\mathbf{a}_i\mathbf{a}_i^H}{1+\sigma_w^{-2}\|\mathbf{a}_i\|^2p_i}$$

Therefore,

$$\mathbf{a}_i^H\mathbf{R}_i{}^{-1}\mathbf{a}_i = \frac{\|\mathbf{a}_i\|^2}{\sigma_w^2} - \frac{\sigma_w^{-4}\|\mathbf{a}_i\|^4p_i}{1+\sigma_w^{-2}\|\mathbf{a}_i\|^2p_i} = \frac{\sigma_w^{-2}\|\mathbf{a}_i\|^2}{1+\sigma_w^{-2}\|\mathbf{a}_i\|^2p_i}$$

Using (2.49), (2.50), and the fact that

$$\mathbf{a}_i^H \mathbf{R}_i^{-1} \tilde{\mathbf{A}}_i (\tilde{\mathbf{P}}_i^{-1} + \tilde{\mathbf{A}}_i^H \mathbf{R}_i^{-1} \tilde{\mathbf{A}}_i)^{-1} \tilde{\mathbf{A}}_i^H \mathbf{R}_i^{-1} \mathbf{a}_i \geq 0,$$

we conclude that

$$\begin{aligned}
[\mathbf{C}]_{ii} &\geq \frac{\eta_{\mathbb{F}}^{-1}}{|\mathbf{a}_i^H \mathbf{R}^{-1} \mathbf{a}_i|^2} \\
&\geq \frac{\eta_{\mathbb{F}}^{-1}}{|\mathbf{a}_i^H \mathbf{R}_i^{-1} \mathbf{a}_i|^2} = \eta_{\mathbb{F}}^{-1} \left(\frac{\sigma_w^2 + \|\mathbf{a}_i\|^2 p_i}{\|\mathbf{a}_i\|^2} \right)^2 && (2.51) \\
&\geq \eta_{\mathbb{F}}^{-1} p_i^2 && (B_{nz})
\end{aligned}$$

where the label B_{nz} stands for the bound for nonzero entries. The bounds for zero entries (B_{z1}, B_{z2}) will be studied later. □

Remark 9. *The theorem indicates that for all admissible values of M, N and sparsity K, the CRB corresponding to the non zero elements of* $\mathbf{p}$ *is strictly greater than* 0*, regardless of the structure of the dictionary. This happens in both overdetermined* $(N \leq M)$ *and underdetermined settings* $(N > M)$*, implying that the non zero elements of* $\mathbf{p}$ *cannot be estimated with zero MSE even when* $M \to \infty$.

Remark 10. *For the special case when* $\mathbf{A}^H \mathbf{A} = M\mathbf{I}$ *(which holds only if* $N \leq M$*), the authors in [PM13] show that the inequality in* (2.51) *holds with equality. Our result generalizes this observation for any dictionary* $\mathbf{A}$ *and for all values of M and N.*

Lower Bounds on the CRB Corresponding to Zero Elements

We will now show that for $i \notin \mathcal{S}$ (i.e. $[\mathbf{p}]_i = 0$), saturation of the CRB may or may not happen, depending on the structure of the dictionary $\mathbf{A}$ and the normalization of its columns.

Saturation Effect for Normalized Dictionaries Let $\mathbf{A}$ be a dictionary with normalized columns such that

$$\|\mathbf{a}_i\|_2^2 = c, \quad 1 \leq i \leq N$$

where c is a constant that does not depend on M or N. In this case, the CRB corresponding to the zero elements of $\mathbf{p}$ will saturate, as given by the following theorem:

Theorem 12. *Consider the model (2.46), where the measurement* $\mathbf{y}$ *is distributed according to (2.47), and the columns of* $\mathbf{A}$ *are normalized such that* $\|\mathbf{a}_i\|_2^2 = c, 1 \leq i \leq N$ *where c is a universal constant that does not depend on M or N. If* $N = rank(\mathbf{A}_{\text{ca}})$*, the CRB corresponding to the unknown parameter* $\mathbf{p}$ *satisfies*

$$[\mathbf{C}]_{ii} \geq \frac{\sigma_w^4 \eta_{\mathbb{F}}^{-1}}{c^2} \quad i \notin \mathcal{S}$$

Proof. Similar to the proof of Theorem 11, we use Woodbury's matrix identity on $\mathbf{R}^{-1}$, but in a different form. In particular, we can write

$$\mathbf{R}^{-1} = \sigma_w^{-2}(\mathbf{I} - \tilde{\mathbf{A}}(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}} + \sigma_w^2\tilde{\mathbf{P}}^{-1})^{-1}\tilde{\mathbf{A}}^H)$$

where $\tilde{\mathbf{A}}$ is the matrix comprised of columns of $\mathbf{A}$ indexed by $\mathcal{S}$, and $\tilde{\mathbf{P}}$ is a diagonal matrix containing only the non zero elements of $\mathbf{p}$. Since $i \notin \mathcal{S}$, we have

$$\begin{aligned}\mathbf{a}_i^H \mathbf{R}_i{}^{-1}\mathbf{a}_i &= \sigma_w^{-2}\mathbf{a}_i^H(\mathbf{I} - \tilde{\mathbf{A}}(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}} + \sigma_w^2\mathbf{P}^{-1})^{-1}\tilde{\mathbf{A}}^H)\mathbf{a}_i \\ &\leq \sigma_w^{-2}\|\mathbf{a}_i\|^2\end{aligned}$$

which follows from the fact that

$$\mathbf{a}_i^H \tilde{\mathbf{A}} (\tilde{\mathbf{A}}^H \tilde{\mathbf{A}} + \sigma_w^2 \tilde{\mathbf{P}}^{-1})^{-1} \tilde{\mathbf{A}}^H \mathbf{a}_i \geq 0.$$

Using (2.49), for $i \notin \mathcal{S}$, we can always write

$$[\mathbf{C}]_{ii} \geq \frac{\sigma_w^4 \eta_{\mathbb{F}}^{-1}}{\|\mathbf{a}_i\|^4} = \frac{\sigma_w^4 \eta_{\mathbb{F}}^{-1}}{c^2} \tag{B_{z1}}$$

□

Notice that the first inequality in (B_{z1}) provides a valid lower bound for any dictionary $\mathbf{A}$ (regardless of normalization of columns). However, for unnormalized dictionaries, if $\|\mathbf{a}_i\|^2$ grows monotonically with M, the lower bound $\frac{\sigma_w^4 \eta_{\mathbb{F}}^{-1}}{\|\mathbf{a}_i\|^4}$ in (B_{z1}) converges to a trivial value of 0 (as $M \to \infty$) which does not shed any light into the asymptotic behavior of the CRB.

Lower Bound for Unnormalized Dictionaries, and $K \geq M$ To better understand the behavior of CRB for dictionaries with unnormalized columns, we consider a special case when $K \geq M$ and the non-zero hyper-parameters are all equal to p, i.e, $[\mathbf{p}]_i = p$ for $i \in \mathcal{S}$, and $[\mathbf{p}]_i = 0$ for $i \notin \mathcal{S}$. We further assume that $\tilde{\mathbf{A}}$ has full row rank M (which is possible since $K \geq M$). Consider the singular value decomposition of $\tilde{\mathbf{A}}\tilde{\mathbf{A}}^H$ as $\tilde{\mathbf{A}}\tilde{\mathbf{A}}^H = \mathbf{U}\Sigma\mathbf{U}^H$ where $\Sigma = \text{diag}(\sigma_1, \cdots, \sigma_M)$, and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_M > 0$. Thus, we can write

$$\mathbf{R} = p\tilde{\mathbf{A}}\tilde{\mathbf{A}}^H + \sigma_w^2 \mathbf{I} = \mathbf{U}[p\Sigma + \sigma_w^2 \mathbf{I}]\mathbf{U}^H$$

Therefore, we have

$$\mathbf{R}^{-1} = \mathbf{U}\Gamma\mathbf{U}^H$$

where $\Gamma = \text{diag}(\frac{1}{p\sigma_1+\sigma_w^2}, \frac{1}{p\sigma_2+\sigma_w^2}, \cdots, \frac{1}{p\sigma_M+\sigma_w^2})$. Hence,

$$\mathbf{a}_i^H \mathbf{R}^{-1} \mathbf{a}_i \leq \sigma_{\max}(\mathbf{R}^{-1}) \|\mathbf{a}_i\|^2 = \frac{\|\mathbf{a}_i\|^2}{p\sigma_M + \sigma_w^2}$$

with $\sigma_{\max}(.)$ denoting the maximum singular value of a matrix. Using (2.49), we get

$$[\mathbf{C}]_{ii} \geq \eta_{\mathbb{F}}^{-1} \frac{(p\sigma_{\min}^2(\tilde{\mathbf{A}}) + \sigma_w^2)^2}{\|\mathbf{a}_i\|^4} \tag{B_{z2}}$$

where $\sigma_M = \sigma_{\min}^2(\tilde{\mathbf{A}})$, and $\sigma_{\min}(.)$ indicates the smallest nonzero singular value of a matrix.

Proposition 1. *If* $\frac{\sigma_{\min}(\tilde{\mathbf{A}})}{\|\mathbf{a}_i\|} = O(1)$ *(i.e. does not scale with M or N),* $[\mathbf{C}]_{ii}$ *in* (B_{z2}) *for* $i \notin \mathcal{S}$*, will be bounded below by a positive quantity as* $M \to \infty$*.*

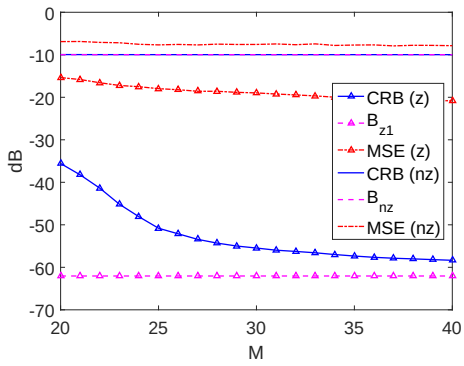
Therefore, when the columns $\mathbf{A}$ are not normalized, saturation may or may not happen. This depends on the asymptotic behavior of the smallest singular value of $\tilde{\mathbf{A}}$ with respect to its column norm, as we increase the sizes M, N and K.

2.3.3 Simulations

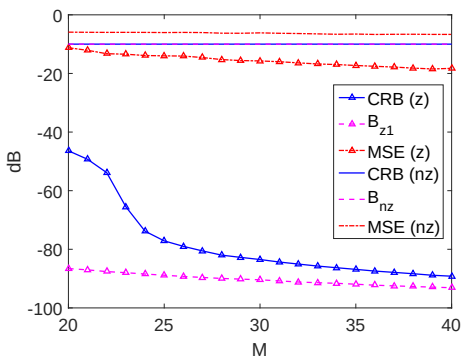
We conduct numerical experiments to examine the behavior of the CRB for zero and nonzero elements of $\mathbf{p}$, as we increase the size of dictionary $\mathbf{A}$. We generate a matrix $\mathbf{A_0}$ with i.i.d. standard normal entries, and let $\mathbf{A}$ be a submatrix of $\mathbf{A_0}$ by choosing the first M (resp. N) rows (resp. columns) of $\mathbf{A_0}$. In each simulation, we generate $\mathbf{p}$ such that the support corresponding to a smaller sparsity level K is a subset of the support corresponding to the larger value of K. We consider $L = 20$ i. i. d. realizations of the vector $\mathbf{x}$ with the same support. This essentially scales the CRB values by a factor of $\frac{1}{L}$ and does not affect our analysis, yet it can slightly improve the performance of our estimator (discussed later), whose error is compared with CRB. The noise variance is assumed to be $\sigma_w = 0.05$, and all the nonzero values of $\mathbf{p}$ are equal to 1.

We consider three different experimental settings, and for each case, we consider both

normalized and unnormalized $\mathbf{A}$. In "Setting 1", we fix $N = 100$, $K = 10$ and increase M. In "Setting 2", we also let K and N grow as we increase M, such that $K = \lfloor \frac{M}{4} \rfloor$, $N = 4M$. "Setting 3" differs from "Setting 2" only in the fact that K can be larger than M. In particular, we let $K = 2M$, $N = 4M$. In all cases, we let the starting M to be $M = 20$, to ensure nonsingularity of the FIM. The experimental results for each scenario are plotted in Figure 2.11, where we compare the CRB with the lowerbounds established in this Section. In Fig. 2.10 (a,b) we also show the Mean Square Error (MSE) of the Maximum Likelihood (ML) estimator, and compare it with the CRB corresponding to Setting 1. The MSE of the ML estimator is computed by averaging over 2000 Monte-Carlo simulations for each M. We observe that the saturation effect always happens for both zero and non-zero elements when the dictionary is normalized, thereby validating our theoretical claims. When $\mathbf{A}$ is not normalized, the CRB corresponding to the zero elements seem to decrease monotonically for Settings 1 and 2 (where we have $K < M$). In future, we will explore the behavior of the CRB of zero elements in greater detail.

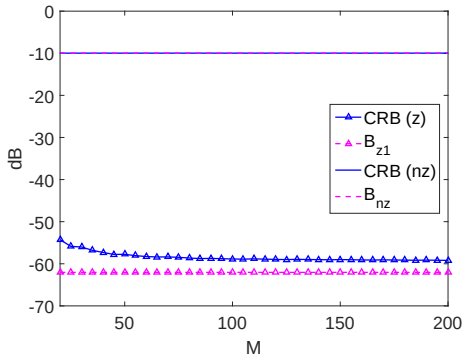


(a) Setting 1, Normalized

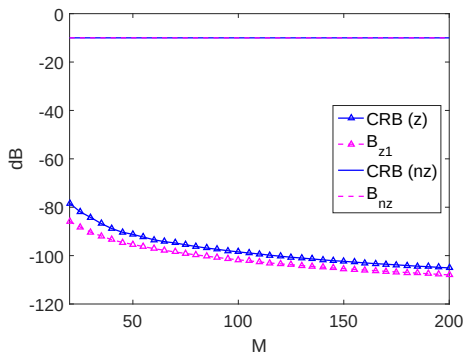


(b) Setting 1, Unnormalized

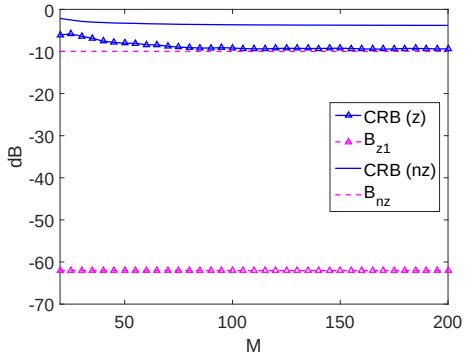
Figure 2.10: Comparisons between the CRB, the lower bounds established in this Section (indicated by their corresponding labels), and the MSE of ML algorithm. The label "(z)" indicates zero elements and "(nz)" represents nonzero elements.



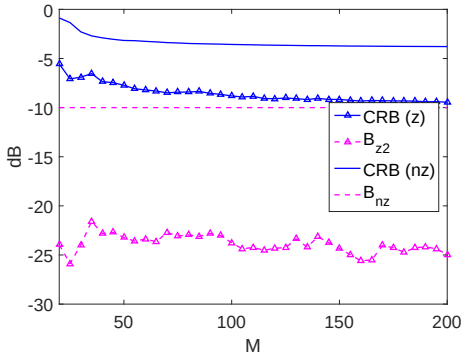
(c) Setting 2, Normalized



(d) Setting 2, Unnormalized



(e) Setting 3, Normalized



(f) Setting 3, Unnormalized

Figure 2.11: Comparison between the CRB, and the lower bounds, established in this Section. The labels in this Figure are the same as those in Fig. 2.10.

2.3.4 Conclusion

We considered the Marginalized Cramér Rao bound associated with hyper-parameter estimation in Sparse Bayesian Learning. We showed that the CRB corresponding to the nonzero elements is always bounded below by a positive quantity which does not go to zero as we increase the number measurements, thereby exhibiting saturation. However, for the zero elements, saturation of the CRB may or may not happen, depending on the column norm as well as the algebraic structure of the dictionary. We will further investigate this phenomena in future by deriving suitable upper bounds for the CRB corresponding to zero elements.

2.4 Sparse Source Localization Using Perturbed Arrays via Bi-Affine Modeling

Array imperfections such as gain and/or phase error, perturbations in sensor locations, and mutual coupling, can significantly degrade the performance of DOA estimation algorithms [SCG87], [CLYM91]. This is mainly due to the strong dependence of these algorithms on the accurate knowledge of the underlying array manifold. In this Section, similar to Sec. 2.2, we consider the sensor location error as the only imperfection associated with the physical array, i.e., we assume that the sensor locations are perturbed from their nominal positions. The problem of DOA estimation using such perturbed arrays has been well studied for more than two decades. Existing approaches mostly treat the perturbations as unknown but deterministic parameters, and then estimate these parameters jointly with the DOAs. Classical methods such as [CLYM91], [WF89], [PK85] [VS94], resolve array uncertainties using eigenstructure-based methods, or variants of the maximum-likelihood approach. Recently, [LZ13] proposed a unified framework for different kind of array imperfections, and proposed a Bayesian approach for array calibration and DOA estimation. However, these approaches mostly work for an overdetermined signal model (fewer sources than sensors), primarily because many of them consider a uniform linear array.

In recent times, the problem of blind gain and phase calibration (BGPC) has been formulated as a bilinear problem [LLB15], which in turn, can be recast as a convex optimization problem, using the idea of "lifting" [CSV13, ARR14b, LS15b]. However, such a formulation does not consider the concept of co-array, and, hence their guarantees are not applicable for an underdetermined signal model where the number of sources can possibly be $O(M^2)$.

In contrast, the authors in [HYN15], studied the effect of co-array geometry on the BGPC problem and proposed a new self-calibration algorithm for nested arrays in presence of gain/phase errors. Their approach builds on and extends the method in [PK85], which was

originally proposed for a ULA. However, in this Section, we consider perturbations in sensor locations, which gives rise to a signal model, which is distinctly different from that considered in [HYN15]. In BGPC problems, the gain and/or phase of the sensors are unknown, and the goal is to resolve both unknown gain and/or phase and the DOAs. In our case, we assume that the phase and gain of the signals received from the sensors are ideal, but the location of the sensors are perturbed. We will compare the signal model defined for gain/phase error, which has been studied in [HYN15], against sensor location error in Section 2.4.1, and establish important differences between them.

Since the self calibration algorithm developed in [HYN15] cannot be directly applied to our case, we follow a different approach in this Section. We assume that the perturbations are small, so that we can approximate the 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 a clever elimination of variables. By exploiting the pattern of repeating elements, it is possible to reduce the said 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.

Notation: Throughout this Section, matrices are represented by upper case bold letters, and vectors by lower case bold letters. The symbol x_i represents the ith entry of a vector $\mathbf{x}$. The symbol j denotes the imaginary unit $\sqrt{-1}$. The symbols $(.)^*, (.)^T, (.)^H$ stand for the conjugate, transpose, and hermitian, respectively. The symbols $\circ$, $\odot$, $\otimes$ represent the Hadamard product, Khatri-Rao product, and Kronecker product, respectively. The symbol $\|.\|_F$ denotes the matrix Frobenius norm and $\text{vec}(.)$ represents the vectorized form of a matrix.

2.4.1 Signal Model for Gain/Phase Error vs Location Errors

Consider a linear array of M antennas impinged by K narrow-band sources with unknown directions of arrival (DOA) $\theta \in \mathbb{R}^K, \theta = [\theta_1, \theta_2, \cdots, \theta_K]^T$. Let $\mathbf{y}[l] \in \mathbb{C}^M$ be the vector of signals received by the M antennas, $\mathbf{x}[l] \in \mathbb{C}^K$ represent the emitted signals from K sources, and $\mathbf{w}[l]$ be the additive noise (all corresponding to the lth time snapshot). The source signals are assumed to be zero mean, and pairwise uncorrelated, and the noise vector is zero mean, i.i.d. with variance σ_w^2, and uncorrelated from the signal. We do not make any specific assumptions on the distribution of the signal or noise.

The sensors are designed to be at the nominal locations $\tilde{d}_1, \tilde{d}_2, \cdots, \tilde{d}_M$, where $\tilde{d}_m \in \mathbb{R}$ for $1 \leq m \leq M$, and $\tilde{d}_m = Dd_m$. Here, $d_m \in \mathbb{Z}$, and D is the minimum inter-element spacing of the array, which is typically chosen to be $D = \lambda/2$, λ being the carrier wavelength of the narrowband sources. Note that d_m are the normalized sensor locations (and $\tilde{d}_m$ are the actual sensor locations). In the sequel, we will use the normalized locations as we introduce the perturbed array model. In this Section, we consider two different array geometries: uniform linear array (ULA), and coprime array. In a ULA, we have $d_m = m - 1$, for $m = 1, \cdots, M$. A coprime array, however, is comprised of two different ULAs with spacings N_1 and N_2, where N_1 and N_2 are coprime numbers. We will review the coprime arrays in more detail in Sec. 2.4.1. To simplify the notations, we designate a spatial frequency $\omega_i = \frac{2\pi D}{\lambda} \sin\theta_i$ corresponding to each direction of arrival θ_i for $1 \leq i \leq K$. Choosing $D = \lambda/2$, we have $\omega_i = \pi \sin\theta_i$. Also, let $\omega = [\omega_1, \cdots, \omega_K]^T$ be the vector of spatial frequencies associated with the K sources.

Let $\zeta \in \mathbb{C}^M$ be a vector of unknown parameters associated with array imperfections, such as gain/phase, or sensor location errors. The received samples at the time instant l can be written as

$$\mathbf{y}[l] = \mathbf{A}(\omega, \zeta)\mathbf{x}[l] + \mathbf{w}[l] \tag{2.52}$$

in which $\mathbf{A}(\omega, \zeta) = [\mathbf{a}(\omega_1, \zeta), \ldots, \mathbf{a}(\omega_K, \zeta)]$ denotes the array manifold, and $\mathbf{a}(\omega_i, \zeta) \in \mathbb{C}^M$ is the

steering vector for the ith source. In the absence of array imperfections ($\boldsymbol{\zeta} = \mathbf{0}$), the mth element of the steering vector corresponding to direction θ_i is given by $a_m(\omega_i, \mathbf{0}) = e^{jd_m\omega_i}$. In the following subsections, we will first review the concept of a virtual array by considering the covariance matrix for the unperturbed problem [PV10]. Subsequently, we will discuss and distinguish the signal models corresponding to two different kinds of array imperfections: (i) gain/phase error, and (ii) sensor location error.

Virtual array in the absence of array imperfections

In the absence of array imperfections ($\boldsymbol{\zeta} = \mathbf{0}$), we can write the covariance matrix of the received signals as

$$\mathbf{R}_{\mathbf{y}} = E(\mathbf{y}\mathbf{y}^H) = \mathbf{A}_0(\boldsymbol{\omega})\mathbf{R}_{\mathbf{x}}\left(\mathbf{A}_0(\boldsymbol{\omega})\right)^H + \sigma_w^2\mathbf{I} \tag{2.53}$$

where $\mathbf{R}_{\mathbf{x}} = E(\mathbf{x}\mathbf{x}^H)$ is the covariance matrix of the sources, and $\mathbf{A}_0(\boldsymbol{\omega}) = \mathbf{A}(\boldsymbol{\omega}, \mathbf{0})$. Assuming that the sources are uncorrelated, i.e., $\mathbf{R}_{\mathbf{x}}$ is diagonal, following [PV10] the vectorized form of the covariance matrix can be written as

$$\mathbf{z} = \mathbf{A}_{\mathrm{KR},0}(\boldsymbol{\omega})\tilde{\mathbf{p}} + \sigma_w^2\,\mathrm{vec}(\mathbf{I}), \tag{2.54}$$

where $\mathbf{A}_{\mathrm{KR},0}(\boldsymbol{\omega}) = \mathbf{A}_0(\boldsymbol{\omega})^* \odot \mathbf{A}_0(\boldsymbol{\omega})$ is the difference co-array, $\tilde{\mathbf{p}} = [p_1, p_2, \cdots, p_K]$ is the diagonal of $\mathbf{R}_{\mathbf{x}}$, and $\mathbf{z} = \mathrm{vec}(\mathbf{R}_{\mathbf{y}})$. The $(m + (m'-1)M, i)$-th element of $\mathbf{A}_{\mathrm{KR},0}(\boldsymbol{\theta})$ is given by $e^{j\omega_i(d_m - d_{m'})}$. Therefore, each column of $\mathbf{A}_{\mathrm{KR},0}(\boldsymbol{\omega})$ is characterized by the difference co-array:

$$S_{\mathrm{ca}} = \{d_m - d_{m'}, 1 \le m, m' \le M\}$$

Define $M_{\mathrm{ca}} = |S_{\mathrm{ca}}|$ to be the number of distinct elements in the unperturbed virtual coarray, and also let $M'_{\mathrm{ca}} = \frac{M_{\mathrm{ca}}-1}{2}$ be the size of the positive half of the coarray.

Weight Function: Let $w(k)$ be the number of repetitions of the lag k in the difference coarray, i.e., number of pairs of indices (i, j) such that $d_i - d_j = k$. The function $w(k)$ is known as the weight function of the coarray [PV10]. In later sections, we will make extensive use of $w(k)$ to eliminate the unwanted perturbation variable in our proposed biaffine model. We will consider array geometries such that $w(k) > 1$ for an adequate number of lags k contained in S_{ca}. The weight functions corresponding to ULA and robust coprime array (defined later) are illustrated in Figure 2.12, in which $N_1 = 4, N_2 = 9, M = 32$.

ULA and its Coarray For a ULA, we have $d_m = m - 1$, with $1 \leq m \leq M$. The coarray corresponding to a ULA is given by

$$S_{\text{ca}}^{\text{ULA}} = \{m \big| m = -(M-1), \ldots, M-1\}$$

Therefore, for a ULA we have $M_{\text{ca}} = 2M - 1$, and $M'_{\text{ca}} = M - 1$.

Coprime Array and its Coarray In a coprime array, we choose two coprime numbers N_1, N_2 ($N_1 < N_2$), and place the sensors on two ULAs with inter-element spacings N_1, and N_2. In the original coprime array introduced in [PV11], the first ULA is comprised of N_2 sensors located at the locations $d_n^{(1)} = nN_1$, and the second ULA has $2N_1$ sensors which are located at the $d_m^{(2)} = mN_2$, where $0 \leq n \leq N_2 - 1, 1 \leq m \leq 2N_1 - 1$, and we have $d_i = d_i^{(1)}$ for $1 \leq i \leq N_2$, and $d_j = d_{j-N_2}^{(2)}$ for $N_2 + 1 \leq j \leq 2N_1 + N_2 - 1$. This choice leads to a virtual array such that

$$S_{\text{ca}}^{\text{Coprime}} \supset \{n \big| n = -N_1N_2, \ldots, N_1N_2\}$$

Hence, $M'_{\text{ca}} > N_1N_2$, and $M_{\text{ca}} > 2N_1N_2 - 1$. In other words, the virtual coarray has $O(N_1N_2)$ elements, although the physical array has only $M = 2N_1 + N_2 - 1$ sensors. This increased degrees of freedom makes it possible to resolve more sources than the available number of sensors. In the

following sections, we will see what happens to the coarray when we have imperfections in the coprime array.

Perturbed Sensor Array and its Virtual Array

Assume that the sensors are not located at their nominal positions, and their perturbed locations are given by $d_1+\delta_1, d_2+\delta_2, \cdots, d_M+\delta_M$, in which $\delta \in \mathbb{R}^M, \delta = [\delta_1, \delta_2, \cdots, \delta_M]^T$ is an unknown perturbation vector. In this case, we have $a_m(\omega_i, \delta) = e^{j(d_m+\delta_m)\omega_i}$, from which we can rewrite the array manifold $\mathbf{A}(\omega, \delta) \in \mathbb{C}^{M\times K}$ as

$$\mathbf{A}(\omega, \delta) = \mathbf{A}_0(\omega) \circ \mathbf{P}(\omega, \delta), \tag{2.55}$$

where $\mathbf{A}_0(\omega) = \mathbf{A}(\omega, \mathbf{0})$ is the unperturbed array manifold, and $\mathbf{P}(\omega, \delta) \in \mathbb{C}^{M\times K}$ is a perturbation matrix whose (m,i)th element is given by $e^{j\omega_i\delta_m}$. The covariance matrix of the received signals can therefore be written as

$$\mathbf{R}_\mathbf{y} = E(\mathbf{y}\mathbf{y}^H) = \mathbf{A}(\omega, \delta)\mathbf{R}_\mathbf{x}(\mathbf{A}(\omega, \delta))^H + \sigma_w^2\mathbf{I}, \tag{2.56}$$

and similar to Sec. 2.4.1, we can write the vectorized form of the covariance matrix to get

$$\mathbf{z} = \mathbf{A}_{\text{KR}}(\omega, \delta)\tilde{\mathbf{p}} + \sigma_w^2 \text{vec}(\mathbf{I}), \tag{2.57}$$

The $(m+(m'-1)M, k)$-th element of $\mathbf{A}_{\text{KR}}(\omega, \delta)$ is given by $e^{j\omega_k(d_m+\delta_m-d_{m'}-\delta_{m'})}$. The difference coarray corresponding to the perturbed array is then

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

Grid based model: In this Section, we consider a grid-based model for the DOAs [PV12a, KP15],

where the range of possible directions $[-\pi/2, \pi/2]$, is quantized into N_θ grid points. We can then construct a *grid based* array manifold $\mathbf{A}_{\text{grid}}(\delta) \in \mathbb{C}^{M \times N_\theta}$, where each column of this matrix is a steering vector corresponding to a particular direction on the grid. The grid-based co-array manifold is given by $\mathbf{A}_{\text{ca}}(\delta) = \mathbf{A}_{\text{grid}}(\delta)^* \odot \mathbf{A}_{\text{grid}}(\delta)$. *Clearly, this co-array manifold only depends on* δ*, and the structure of the array*. We can now rewrite (2.57) as

$$\mathbf{z} = \mathbf{A}_{\text{ca}}(\delta)\mathbf{p} + \sigma_w^2 \,\text{vec}(\mathbf{I}), \tag{2.58}$$

where the non-zero elements of $\mathbf{p} \in \mathbb{C}^{N_\theta}$ are equal to the corresponding elements of $\tilde{\mathbf{p}}$. The locations of the non zero elements of $\mathbf{p}$ can be used to recover the DOAs.

The model derived in this subsection for sensor location errors is distinctly different from the model based on gain/phase error, which we discuss next.

Covariance matrix with unknown sensor gain & phase

The problem of co-array based DOA estimation with phase and gain errors has been thoroughly studied in [HYN15] for ULA and nested arrays. In this Section, we will repeat some of the results from [HYN15] to distinguish it from the problem considered in this Section. The received signal at an array with gain and phase errors is given by

$$\mathbf{y}[l] = \Phi\Psi\mathbf{A}(\omega)\mathbf{x}[l] + \mathbf{w}[l]$$

where $\Phi = \text{diag}(e^{j\phi_1}, \cdots, e^{j\phi_M})$, $\Psi = \text{diag}(\psi_1, \cdots, \psi_M)$ are the respective phase and gain errors, in which $\psi_m > 0$ for $1 \leq m \leq M$. We can stack the vectors $\mathbf{y}[l]$ for $l = 1, \cdots, L$ into the columns of $\mathbf{Y} \in \mathbb{C}^{M \times L}$ (and similarly for $\mathbf{X}$ and $\mathbf{W}$) to get

$$\mathbf{Y} = \Lambda\mathbf{A}(\omega)\mathbf{X} + \mathbf{W}, \tag{2.59}$$

in which $\Lambda = \Phi\Psi$. This problem is known as the Blind Gain and Phase Calibration (BGPC), which has been studied recently in [LLB15, LS15b]. However, these results do not cast the problem in coarray domain, and hence do not address the possibility of resolving more sources than the number of sensors.

The authors in [HYN15], on the other hand, cast the BGPC problem in the co-array domain by using properties of the covariance matrix $\mathbf{R_y}$ for nested arrays and ULA. Let us first write the covariance matrix of the received signal

$$\mathbf{R_y} = \Phi\Psi\mathbf{A}(\omega)\mathbf{R_x}\left(\mathbf{A}(\omega)\right)^H\Psi^H\Phi^H + \sigma_w^2\mathbf{I} \tag{2.60}$$

The (m,n)th element of $\mathbf{R_y}$ is then given by

$$R_{m,n} = \psi_m\psi_n e^{j(\phi_m-\phi_n)}\left(\sum_{k=1}^{K} p_k e^{j(d_m-d_n)\omega_k} + \delta_{m,n}\sigma_w^2\right)$$

with $\delta_{m,n} = 1$ if $m = n$, and $\delta_{m,n} = 0$ otherwise.

Remark 11. *Comparing the covariance matrices corresponding to sensor location errors, and gain/phase errors, which are respectively given by* (2.56) *and* (2.60)*, we notice some important differences. In* (2.56)*, the array imperfections appear in the form of the matrix* $\mathbf{P}(\omega,\delta)$ *which* forms a Hadamard product with the array-manifold*. However, in* (2.60)*, the gain and phase errors are captured in the matrix* Λ *which* gets multiplied to the array manifold matrix*. Due to this, the phase error in BGPC problems* is independent of the directions of arrival of sources*, whereas, in presence of location errors, the phase error in each entry of the array manifold is a function of the DOAs. Therefore, the approach in [HYN15] cannot be directly used to resolve the sensor location errors in our case.*

2.4.2 Formulation as a Bi-Affine problem

The Bi-Affine Model

In this Section, we derive a bi-affine model from the covariance matrix (2.56) corresponding to a sensor array with perturbed locations. Our main assumption is that δ is small enough so that we can approximate the coarray manifold $\mathbf{A}_{\text{ca}}(\delta)$ using the first order Taylor series expansion as follows:

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

which can be also written in the matrix form as

$$\mathbf{A}_{\text{ca}}(\delta) \simeq \mathbf{A}_{\text{ca},0} + \Delta\mathbf{A}_{\text{ca},0}\Upsilon \tag{2.61}$$

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

$$\begin{bmatrix} \text{diag}(\delta)-\delta_1\mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \text{diag}(\delta)-\delta_2\mathbf{I} & \cdots\mathbf{0} & \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \text{diag}(\delta)-\delta_M\mathbf{I} \end{bmatrix} \tag{2.62}$$

Therefore, (2.58) can be approximated as

$$\mathbf{z} = (\mathbf{A}_{\text{ca},0} + \Delta\mathbf{A}_{\text{ca},0}\Upsilon)\mathbf{p} + \sigma_w^2\,\text{vec}(\mathbf{I}), \tag{2.63}$$

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]:

$$\mathbf{z}^u = (\mathbf{A}^u_{\text{ca},0} + \Delta^u \mathbf{A}^u_{\text{ca},0} \Upsilon)\mathbf{p} \tag{2.64}$$

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 Δ^u is constructed from Δ by retaining only the rows corresponding to those of $\mathbf{A}^u_{\text{ca},0}$. Notice that unlike the approach in [PV11, PV10], 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.

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 [LLB15], we call the model given in (2.64) a *"bi-affine"* model, since $\mathbf{z}^u$ is a linear function of $\mathbf{p}$ and affine function of δ. Given the covariance matrix $\mathbf{R}_\mathbf{y}$ (or equivalently $\mathbf{z}^u$), the goal is to recover the sparse vector $\mathbf{p}$ from the bi-affine model defined in (2.63).

2.4.3 Source Localization: Bi-Affine to Linear Transformation

Under the grid-based model, the DOAs can be estimated from the support of the sparse vector $\mathbf{p}$ that is a solution to the bi-affine system of equations (2.64). In general, (2.64) can admit multiple solutions in the variables $(\delta, \mathbf{p})$. 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.

[1] For ease of exposition, we only consider the positive half of the co-array and demonstrate how to eliminate the spurious variable δ. 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 δ.

In this Section, we will derive a transformation such that *we can extract a linear system of equations in the variable* $\mathbf{p}$, from (2.64) by eliminating the variable δ. In particular (2.64) can be reduced to an underdetermined linear system of equations of the form

$$\mathbf{h} = \mathbf{G}\mathbf{p} \tag{2.65}$$

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 $\mathbf{p}$ (although, not necessarily in δ) if (2.65) 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).

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 $\mathbf{R}_{\mathbf{y}}$, thereby eliminating δ.

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

$$f_k := \sum_{i=1}^{N_\theta} e^{jk\omega_i} p_i \tag{2.66}$$

$$\lambda_k := \sum_{i=1}^{N_\theta} e^{jk\omega_i} j\omega_i p_i \tag{2.67}$$

Using these notations, we can rewrite (2.64) as

$$R_{mn} = f_k + \lambda_k(\delta_m - \delta_n) \tag{2.68}$$

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 $\mathbf{p}$. 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

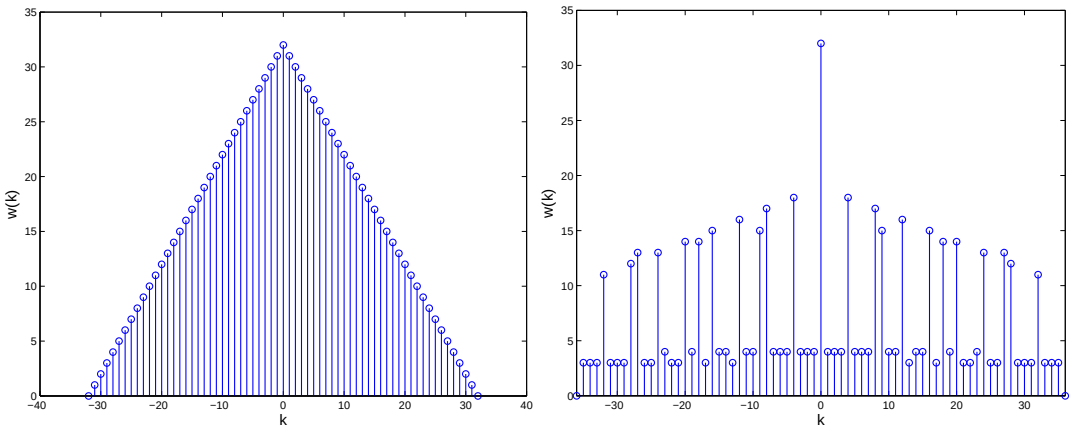
$$R_{ij} - R_{mn} = \lambda_k(\delta_i - \delta_j - \delta_m + \delta_n).$$

A very similar idea can be used (with some additional computations) to eliminate δ from the M^2 equations of the form (2.68) 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 [HYN15]. However, since we are concerned with sensor position errors, our signal model fundamentally differs from that considered in [HYN15]. Consequently, as discussed in Sec. 2.4.1, we cannot exploit the co-array redundancies in the same way as done in [HYN15]. We need to adopt a slightly more involved approach to eliminate the undesirable variable δ, the details of which depend on the geometry of the physical array.

Uniform Linear Array

In a ULA, we have $d_m - d_n = m - n$. Hence, we can rewrite the equation (2.68) as

$$R_{mn} = f_{m-n} + \lambda_{m-n}(\delta_m - \delta_n) \tag{2.69}$$



(a) ULA (b) Robust-coprime

Figure 2.12: The weight function 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

$$\bar{r}_k = \sum_{j=2}^{k+1} R_{j,j-1} \tag{2.70}$$

$$\beta_k = \frac{R_{k+2,2} - R_{k+1,1}}{R_{k+2,k+1} - R_{2,1}} \tag{2.71}$$

The following theorem summarizes our main result for the ULA:

Theorem 13. *For a Uniform Linear Array (ULA) containing M antennas with perturbed locations, the bi-affine model* (2.64) *derived from the signal covariance matrix can be reduced to the form*

$$\mathbf{Cf} = \mathbf{h} \tag{2.72}$$

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*

$$C_{k-1,1} = -k\beta_k \tag{2.73}$$

$$C_{k-1,k} = 1 \tag{2.74}$$

$$h_{k-1} = R_{k+1,1} - \beta_k \bar{r}_k \tag{2.75}$$

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

Proof. The proof can be found in 2.6.4. □

Remark 12. *Notice that the elements of* $\mathbf{f}$ *are linear functions of the sparse vector* $\mathbf{p}$*. In particular, from* (2.66)*, we have*

$$\mathbf{f} = \mathbf{B}_U \mathbf{p}$$

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* (2.72) *can be written as the following system of underdetermined equations (since* $N_\theta \gg M$*)*

$$\mathbf{h} = \mathbf{C}\mathbf{B}_U \mathbf{p} \tag{2.76}$$

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* $\|\mathbf{p}\|_0 = K < (M-3)/2$ *is a sufficient condition for the true sparse* $\mathbf{p}$ *to be the unique solution. In practice, by exploiting the fact that* $\mathbf{p}$ *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* (2.76) *to determine such an empirical relation between* K *and* M*.*

Remark 13. *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* δ*. The constructive proofs in Appendices A,B demonstrate the details of eliminating the variable* δ *from our bi-affine equations.*

Robust Coprime Array

In order to eliminate δ from the equations (2.64), we need to have $w(k) > 1$. As evident from Fig. 2.12b, that in the original coprime array, $w(k) = 1$ for some values of k, due to which 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

$$\bar{r}_i^{(1)} := \sum_{i'=2}^{i} R_{i',i'-1} \tag{2.77}$$

$$\bar{r}_j^{(2)} := \sum_{j'=2}^{j} R_{j',j'-1}. \tag{2.78}$$

$$\alpha := \frac{2\bar{r}_{N_1+1}^{(2)} - \bar{r}_{2N_1+1}^{(2)}}{2\bar{r}_{N_2+1}^{(1)} - \bar{r}_{2N_2+1}^{(1)}} \tag{2.79}$$

We also define the vectors $\beta_{cp} \in \mathbb{C}^{M'_{\text{ca}}}$ 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 β_{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 $\bar{i} = i + N_2, \bar{j} = j + N_1$.

$$[\beta_{cp}]_k = \begin{cases} \frac{\alpha(R_{i\bar{j}}^{(12)} - R_{ij}^{(12)})}{\alpha(\bar{r}_{\bar{i}}^{(1)} - \bar{r}_i^{(1)} - \bar{r}_{N_2+1}^{(1)}) - (\bar{r}_{\bar{j}}^{(2)} - \bar{r}_j^{(2)} - \bar{r}_{N_1+1}^{(2)})}, & k = d_i^{(1)} - d_j^{(2)} \\ -\frac{\alpha(R_{\bar{i}\bar{j}}^{(12)} - R_{ij}^{(12)})^*}{\alpha(\bar{r}_{\bar{i}}^{(1)} - \bar{r}_i^{(1)} - \bar{r}_{N_2+1}^{(1)}) - (\bar{r}_{\bar{j}}^{(2)} - \bar{r}_j^{(2)} - \bar{r}_{N_1+1}^{(2)})} & k = d_j^{(2)} - d_i^{(1)} \\ \frac{R_{k+2,2}^{(1)} - R_{k+1,1}^{(1)}}{R_{k+2,k+1}^{(1)} - R_{2,1}^{(1)}} \quad k \neq N_2, & k = d_i^{(1)} - d_{i'}^{(1)} \\ \frac{R_{k+2,2}^{(2)} - R_{k+1,1}^{(2)}}{R_{k+2,k+1}^{(2)} - R_{2,1}^{(2)}} \quad k \neq N_1, & k = d_j^{(2)} - d_{j'}^{(2)} \end{cases} \tag{2.80}$$

$$[\mathbf{h}_{cp}]_k = \begin{cases} R_{ij}^{(12)} - \beta_k \bar{r}_i^{(1)} + \alpha^{-1} \beta_k \bar{r}_j^{(2)}, & k = d_i^{(1)} - d_j^{(2)} \\ (R_{ij}^{(12)})^* + \beta_k \bar{r}_i^{(1)} - \alpha^{-1} \beta_k \bar{r}_j^{(2)} & k = d_j^{(2)} - d_i^{(1)} \\ R_{i+1,1}^{(1)} - \beta_k \bar{r}_i^{(1)} \quad k \neq N_2, & k = d_i^{(1)} - d_{i'}^{(1)} \\ R_{j+1,1}^{(2)} - \beta_k \bar{r}_j^{(2)} \quad k \neq N_1, & k = d_j^{(2)} - d_{j'}^{(2)} \end{cases} \tag{2.81}$$

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:

$$[\mathbf{C}_{cp}]_{k,N_1} = \begin{cases} -(i-1)\beta_k, & k = d_i^{(1)} - d_j^{(2)} \\ (i-1)\beta_k, & k = d_j^{(2)} - d_i^{(1)} \\ -(i-1)\beta_k, & k = d_i^{(1)} - d_{i'}^{(1)}, k \neq N_1 \end{cases} \tag{2.82}$$

$$[\mathbf{C}_{cp}]_{k,N_2} = \begin{cases} \alpha^{-1}(j-1)\beta_k, & k = d_i^{(1)} - d_j^{(2)} \\ -\alpha^{-1}(j-1)\beta_k, & k = d_j^{(2)} - d_i^{(1)} \\ -(j-1)\beta_k, & k = d_j^{(2)} - d_{j'}^{(2)}, k \neq N_2 \end{cases} \tag{2.83}$$

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

Theorem 14. *For a robust coprime array, the bi-affine formulation* (2.64) *can be reduced to* $\mathbf{C}_{cp}\mathbf{f}_{cp} = \mathbf{h}_{cp}$*, where* $\mathbf{f}_{cp} = [f_1, f_2, \cdots, f_{M'_{\text{ca}}}]^T$ *and* $\mathbf{C}_{cp} \in \mathbb{C}^{M'_{\text{ca}} \times M'_{\text{ca}}}, \mathbf{h}_{cp} \in \mathbb{C}^{M'_{\text{ca}}}$ *are previously defined. This transformation holds for almost all* $\delta \in \mathbb{R}^M$.

Proof. The proof can be found in the 2.6.5. □

Remark 14. *Recall that* $\mathbf{f}_{cp} = \mathbf{B}_{cp}\mathbf{p}$*, where the elements of* $\mathbf{B}_{cp} \in \mathbb{C}^{M'_{\text{ca}}, N_\theta}$ *are given by* $[\mathbf{B}_{cp}]_{m,i} = e^{jm\omega_i}$*. Hence, from Theorem* 2*, we obtain* $\mathbf{C}_{cp}\mathbf{B}_{cp}\mathbf{p} = \mathbf{h}_{cp}$*, which can admit a unique sparse solution in* $\mathbf{p}$ *if* $\mathbf{C}_{cp}\mathbf{B}_{cp}$ *has kruskal rank of* $O(M'_{\text{ca}})$*. In future, we will characterize the exact kruskal rank of* $\mathbf{C}_{cp}\mathbf{B}_{cp}$*. However, in Sec. 2.4.5, we experimentally show that* ℓ_1 *minimization can resolve larger number of sources for coprime arrays, compared to ULA.*

2.4.4 Iterative Algorithm for finite snapshots and noise

While our main results show that it is fundamentally possible to eliminate the nuisance variable δ and solve for DOAs, they are derived under the assumption that the ideal covariance matrix $\mathbf{R_Y}$ is available. In practice however, we can only estimate $\mathbf{R_Y}$ using a finite number of snapshots. For the estimated covariance matrix, the technique for variable elimination in the proofs of Theorems 1 and 2 may not be robust (although it works perfectly for the ideal covariance matrix). This prompts us to use an iterative algorithm (derived earlier in [KP15] to jointly estimate $\mathbf{p}$ and δ in presence of finite snapshots. In order to recover δ, we need to assume that one of sensors locations is exactly known. This is because of the fact that the perturbed

1: $\delta^{(0)} \leftarrow \mathbf{0}$, $i \leftarrow 0$
2: **repeat**
3: $\quad \mathbf{p}^{(i)} \leftarrow \arg\min_{\mathbf{p}} \left\{ \|\mathbf{A}_{\text{ca}}(\delta^{(i)})\mathbf{p} - \mathbf{z}\|_F^2 \right\}$ s.t. $\|\mathbf{p}\|_0 \leq K$
4: $\quad \delta^{(i+1)} \leftarrow \arg\min_{\delta} \left\{ \|\mathbf{A}_{\text{ca}}(\delta)\mathbf{p}^{(i)} - \mathbf{z}\|_F^2 \right\}$
5: $\quad i \leftarrow i+1$
6: **until** Convergence
7: $\hat{\delta} \leftarrow \delta^{(i)}$, $\hat{\mathbf{p}} \leftarrow \mathbf{p}^{(i)}$

Figure 2.13: Algorithm for jointly estimating the perturbations and the source directions

coarray manifold only depends on the pairwise differences of the sensor locations, i.e., adding a fixed bias to all the sensor locations would not change the $\mathbf{A}(\delta)$ and $\mathbf{z}$. Hence, without loss of generality, we assume $\delta_1 = 0$.

In [KP15], we proposed an iterative approach based on alternating minimization which is summarized in Fig. 2.13. The minimization with respect to $\mathbf{p}$ is performed by solving a sparse recovery problem which could be done using LASSO or OMP. In order to update the vector δ, in the second stage of the iterative algorithm, an error function defined as $E(\delta) = \|\mathbf{A}_{\text{ca}}(\delta)\mathbf{p} - \mathbf{z}\|_F^2$ is minimized through running T iterations of gradient descent

$$\delta \leftarrow \delta - \mu \frac{\partial E(\delta)}{\partial \delta}, \tag{2.84}$$

in which $\mu > 0$ is the step size parameter and the mth component of $\frac{\partial E(\delta)}{\partial \delta}$ for $m \geq 2$ is obtained by

$$\frac{\partial E(\delta)}{\partial \delta_m} = \text{tr}\left(\left(-2\mathbf{z}\mathbf{p}^T + 2\mathbf{A}_{\text{ca}}(\mathbf{0})\mathbf{p}\mathbf{p}^T \right) \frac{\partial \mathbf{P}_{\text{ca}}(\delta)}{\partial \delta_m} \right), \tag{2.85}$$

and $\frac{\partial E(\delta)}{\partial \delta_1} = 0$, where $\mathbf{P}_{\text{ca}}(\delta)$ is given by $\mathbf{P}^*(\omega, \delta) \odot \mathbf{P}(\omega, \delta)$ for ω evaluated on the grid points.

Remark 15. *In our earlier paper [KP15], we have shown that the algorithm summarized in Fig. 2.13 is guaranteed to converge to a local minimum. We showed that the convergence is subject to success of both the LASSO and the gradient descent stages. For the LASSO stage to succeed, the*

sparsity of $\mathbf{p}$ *should be high enough so that the recovering a unique sparse solution be guaranteed. For the gradient descent stage to succeed, we need to have* $\mu < \frac{2}{C}$ *where*

$$C = 2\pi M\sqrt{2M}\|-\mathbf{z}\mathbf{p}^T + \mathbf{A}_{\text{ca}}(\mathbf{0})\mathbf{p}\mathbf{p}^T\|.$$

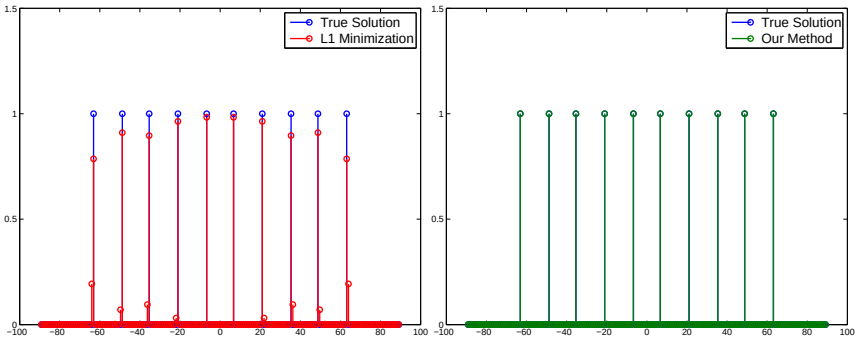
2.4.5 Simulations

In this section, we conduct three 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 $\delta = \alpha\delta_0$ (Notice that, following the model given in Section 2.4.1, the sensor locations and the perturbations are normalized with respect to half of the wavelength $\lambda/2$). In the first and second sets of simulations, δ_0 is a fixed vector with $|\delta|_\infty \leq 0.5$, and is drawn from a uniform distribution. However, for the third set of simulations, we choose δ_0 randomly in each trial.

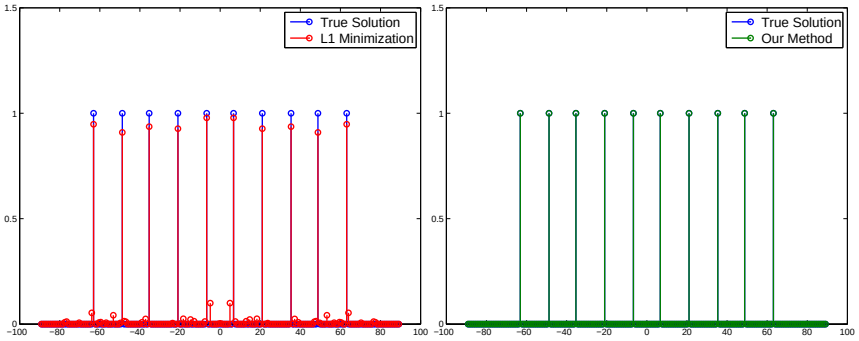
In the first set of simulations, we assume that we know the covariance matrix $\mathbf{R}_\mathbf{y}$, and that the model defined in Sec. 2.4.2 holds exactly. In this case, we use the approach proposed in the proofs of Theorems 13, and 14 to eliminate the perturbations and recover the source powers. 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:

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

As demonstrated in Fig. 2.14 and 2.15, 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,



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

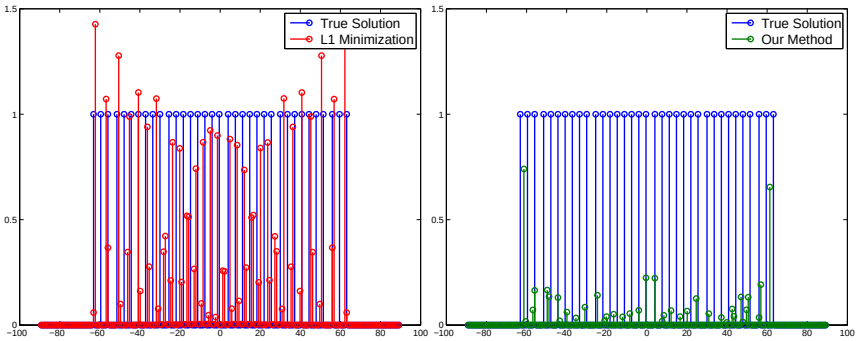


(c) Coprime, $K = 10, M = 32$ (d) Coprime, $K = 10, M = 32$

Figure 2.14: Recovered powers using the approach proposed in Theorems 13, 14. 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.

match exactly). When $K = 35 > M$, ULA cannot recover the true DOAs, while the robust coprime array perfectly identifies the support (compare Fig. 2.15b and Fig. 2.15d, 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.

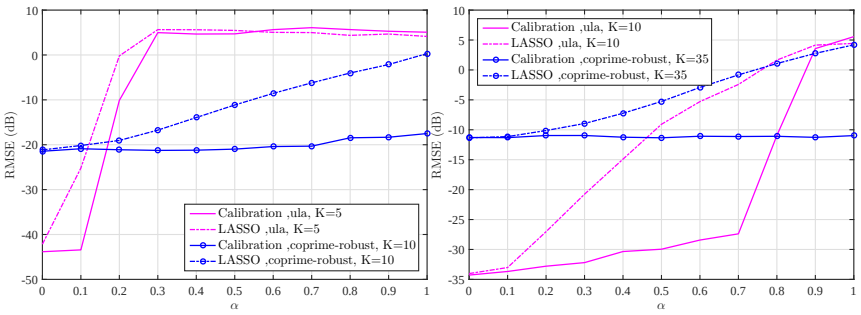
In the second set of simulations, we use an estimated covariance matrix using a finite number of snapshots. In our simulations, we consider both ULA and robust coprime array geometries. For the robust coprime array with $M = 16$, we pick $N_1 = 2, N_2 = 5$, and for $M = 32$ we choose $N_1 = 4, N_2 = 9$, which correspond to Fig. 2.16a, and Fig. 2.16b, respectively. We use the algorithm described in Sec. 2.4.4, which is labeled as "Calibration" in the figure. "LASSO" refers to naively running LASSO on the vectorized covariance matrix assuming an unperturbed



(a) ULA, $K = 35, M = 32$ (b) ULA, $K = 35, M = 32$

(c) Coprime, $K = 35, M = 32$ (d) Coprime, $K = 35, M = 32$

Figure 2.15: Recovered powers using the approach proposed in Theorems 13, 14. 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) $M = 16$ (b) $M = 32$

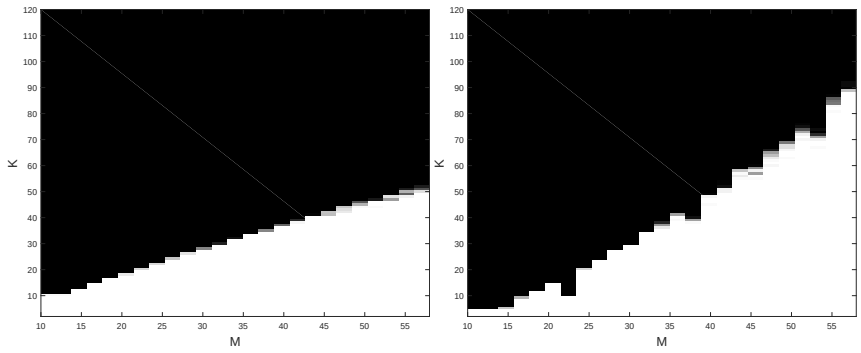
Figure 2.16: RMSE of the recovered sources vs perturbations for different arrays and different number of sensors and sources.

dictionary, i.e., solving the following problem:

$$\min_{\mathbf{p}} \|\mathbf{p}\|_1 + \lambda \|\mathbf{A}^u_{\text{ca},0}\mathbf{p} - \mathbf{z}^u\|_2^2$$

in which λ is a regularization parameter. For each value of α, we show the average value of root-mean-square error (RMSE) over $N_{\text{tests}} = 100$ Monte-Carlo runs, where RMSE $\triangleq \sqrt{\sum_{i=1}^{N_{\text{tests}}} \frac{\|\hat{\mathbf{p}}-\mathbf{p}\|^2}{N_{\text{tests}}}}$, and $\hat{\mathbf{p}}$ is the recovered vector of source powers. We observe that proposed iterative algorithm can recover source powers and is able to resolve the perturbations. We noticed in our simulations that the proposed approach (calibration) may sometimes perform worse than the simple LASSO (for instance, when $\alpha > 0.7$ in Fig. 2.16b). We observed this happens because the iterative algorithm finds a local optimum instead of the global optimum, thereby failing to recover the true source powers. A theoretical analysis of this phenomenon will be a topic of future research.

In the third 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 13, 14. 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 $\|\mathbf{p} - \hat{\mathbf{p}}\|_F \leq \varepsilon$, where $\hat{\mathbf{p}}$ is the recovered vector of powers, and $\varepsilon = 10^{-3}$. The white pixels in the plots of Fig. 2.17 show the problem settings under which performing ℓ_1 minimization on the linear system derived in



(a) ULA (b) Coprime

Figure 2.17: Phase transition plots

Theorems 13, and 14 can always recover the true solution. We simulate each case 100 times 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.

2.4.6 Conclusion

In this Section, 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 13 and 14. 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. When only a finite number of snapshots are available, we proposed an iterative algorithm to jointly recover perturbations and source powers, which shows satisfactory performance in our simulations.

2.5 Compressed Arrays and Hybrid Channel Sensing: A Cramér-Rao Bound Based Analysis

So far in this chapter, we have shown that sparse array geometries such as nested [PV10] and coprime arrays [PV11] can offer significant benefits over the conventional and widely-used Uniform Linear Arrays (ULA), such as the ability to resolve $O(M^2)$ sources with M sensors [PV10], smaller Cramér-Rao bounds [KP16a], higher Direction-of-Arrival (DOA) estimation accuracy [PV10], higher resolution [PV10, KQP18, QP19], and resilience to sensor location errors [KP16b, KP17]. As we will also see in Chapter 4, in mmWave communication, the problem of channel estimation turns out to be related to the problem of DOA estimation since mmWave channels exhibit spatial sparsity with only a small number of scattering paths whose angles of departure/arrival and gains parameterize the channel model [HC16, GWH17, AEALH13, EARAS$^+$14, PH18]. Deploying a large number of antennas lead to higher data rates and improved channel estimation accuracy, but at the cost of increased hardware complexity (due to a proportionately larger number of front-end Radio frequency chains), power consumption and cost. Since sparse arrays can provide high DOA estimation accuracy with much fewer antenna elements compared to a ULA, they offer a promising way to reduce (or compress) the number of physical antennas and thereby the overall system complexity and cost. With this point of view, the authors in [HC16] have shown that by subselecting a sparse coprime array (with $O(\sqrt{M})$ antennas) from a much larger ULA with M antennas using a coprime selection matrix (which can be implemented using analog switches), it is possible to preserve the statistical information about the channel encoded in the received signals, and yet drastically reduce the number of spatial measurements.

Antenna subselection is a special case of deploying a more general compression matrix $\mathbf{W} \in \mathbb{C}^{N \times M}, N < M$ (sometimes also called a beamforming matrix, in the context of mmWave communication) at the output of the antenna array in order to reduce the overall system complexity

[WLP09, GZC18, SCL17]. In this case, for a given number of antennas, the array geometry plays a very important role in determining the achievable compression. The authors in [GZC18] studied the combined benefits gained by using a sparse receiving array and a complex valued compression matrix, which is referred to as a Compressed Sparse Array (CSA). They derived *necessary conditions in terms of N and M* for the existence of the Cramér-Rao Bound for DOA estimation with CSA, which offers insights into the degree of compression that may be achieved by sparse arrays. However, these necessary conditions do not guarantee that the CRB will indeed exist in these compressive regimes, and hence there is a need for *tight sufficient conditions for the existence of CRB*. From a practical point of view, sufficient conditions are also important to guide system design, by suitably choosing M, N and array geometry. In [KP16a], we had derived the first *necessary and sufficient conditions* for the existence of the CRB for sparse arrays when the number of sources exceeds the number of sensors. However, the presence of a complex valued compression matrix $\mathbf{W}$ poses significant technical challenge in translating the proof techniques for sparse arrays to a compressed sparse array.

Contributions: In this section, we derive the first sufficient conditions for the existence of CRB using properties of zero sets of multivariate polynomials (Lemma 1) which help us deal with the compression matrix in an ingenious way [SS94, GR09, JStB01, PV14b]. For almost all complex $\mathbf{W}$, we show that the CRB is guaranteed to exist in different regimes of compression, that depend on the array geometry. Our sufficient conditions are tight, since they agree with the necessary conditions from [GZC18] and provably show the benefit of sparse compressed arrays over compressed ULAs.

Notations: The symbols $(.)^T, (.)^H, (.)^*$ stand for transpose, Hermitian, and complex conjugate, respectively. The symbols $\otimes$, $\odot$ denote Kronecker, and (column-wise) Khatri-Rao product, respectively. Vectors, matrices, and sets are shown with letters such as $\mathbf{A}, \mathbf{a}$, and $\mathcal{A}$, respectively. $|\mathcal{A}|$ denotes cardinality of a set $\mathcal{A}$, and $\mathcal{A}\backslash\mathcal{B}$ stands for set subtraction.

2.5.1 Signal Model and Review of Compressed Arrays

Consider a linear (possibly non-uniform) array with M antennas located at $\frac{\lambda}{2}\{d_1, d_2, \cdots, d_M\}$, where λ is the carrier wavelength of narrowband signals received by the array and d_m are typically selected as integers or rational numbers. The tth temporal snapshot of the signal vector $\mathbf{z}(t) \in \mathbb{C}^M$ at the output of the array, corrupted by additive noise $\mathbf{n}(t) \in \mathbb{C}^M$, is given by

$$\mathbf{z}(t) = \mathbf{A}(\theta)\mathbf{x}(t) + \mathbf{n}(t) \tag{2.86}$$

The model (2.86) arises in various applications. In Direction-of-Arrival (DOA) estimation with narrowband radars, $\theta = [\theta_1, \theta_2, \cdots, \theta_K]$ represent the DOAs of K narrowband far-field sources with waveforms given by the vector $\mathbf{x}(t) = [x_1(t), \cdots, x_K(t)]^T \in \mathbb{C}^K$. In mmWave communication, the channel exhibits sparse scattering that consists of a few (or clusters of) multipath components, where $\theta_i, 1 \leq i \leq K$ represents the angle of arrival of the ith multipath component, and $x_i(t)$ is the corresponding time-varying path gain [2] [PPYH17, AMGPH14]. In either case, $\mathbf{A}(\theta) = [\mathbf{a}(\theta_1), \cdots, \mathbf{a}(\theta_K)] \in \mathbb{C}^{M \times K}$ where $\mathbf{a}(\theta_i) \in \mathbb{C}^M$ is the array steering vector corresponding to the angle θ_i, with elements given by $[\mathbf{a}(\theta_i)]_m = e^{j\frac{2\pi}{\lambda} d_m \sin\theta_i}$, $j = \sqrt{-1}$.

In mmWave hybrid communication systems [ALH15, AEALH13, PH18], the signals at the M antennas are linearly combined through a network of analog circuitry (denoted by a matrix $\mathbf{W} \in \mathbb{C}^{N \times M}$) to produce a lower dimensional/compressed measurement vector $\mathbf{y}(t) \in \mathbb{C}^N$ with $N < M$

$$\mathbf{y}(t) = \mathbf{W}\mathbf{z}(t) = \mathbf{W}\mathbf{A}(\theta)\mathbf{x}(t) + \mathbf{W}\mathbf{n}(t) \tag{2.87}$$

It is well-known that for an uncompressed array ($\mathbf{W} = \mathbf{I}$), non-uniform array geometries (such as nested array) can resolve $K = O(M^2)$ sources by utilizing the structure of the covariance

[2] This prototype mmWave channel model can be extended to consider angular spread, and clusters with correlated multipath [HC16, AGPHJ18].

matrix of the data $\mathbf{y}(t)$, while uniform arrays (such as ULAs) can only resolve $K = O(M)$ sources [PV10]. However, for a compressed array with $N < M$, the number of identifiable sources is *dependent on the choice of both* $\mathbf{W}$ *and array geometry*.

Necessary Conditions for Existence of Cramér Rao Bounds when $N < M$

Suppose $\mathbf{x}(t) \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \mathbf{P})$, where $\mathbf{P}$ is a diagonal matrix with $\mathbf{p} \in \mathbb{R}^K$ being its diagonal elements (assuming that the multipath gains are statistically uncorrelated [PH16, GZC18, PH18, AGPHJ18]) and $\mathbf{n}(t) \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. The analytic form of the Fisher Information Matrix (FIM) corresponding to the parameters $\theta, \mathbf{p}, \sigma^2$ was recently studied in [GZC18] using derivations similar to [KP16a]:

Theorem 15. *Assume that we collect* T *independent snapshots of the compressive measurements* $\mathbf{y}(t), t = 1, 2, \cdots, T$, $\mathbf{y}(t) \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \mathbf{R}_\mathbf{y})$, *where* $\mathbf{R}_\mathbf{y} = \mathbf{W}\left(\mathbf{A}(\theta)\mathbf{P}\mathbf{A}(\theta)^H + \sigma^2\mathbf{I}\right)\mathbf{W}^H$ *is the covariance matrix of* $\mathbf{y}(t)$. *The Fisher Information Matrix (FIM) corresponding to the parameters* $\psi = [\theta^T, \mathbf{p}^T, \sigma^2]^T \in \mathbb{R}^{2K+1}$ *is given by* [3]

$$\mathbf{J}(\psi) = T\mathbf{G}^H\mathbf{H}\mathbf{G} \tag{2.88}$$

where $\mathbf{H} = \mathbf{R}_\mathbf{y}^{-T} \otimes \mathbf{R}_\mathbf{y}^{-1}$, $\mathbf{G} \in \mathbb{C}^{N_{RF}^2 \times (2K+1)}$ *is given as* $\mathbf{G} = (\mathbf{W}^* \otimes \mathbf{W})[\mathbf{A}^* \odot \mathbf{A} \;\; \mathbf{B} \;\; \tilde{\mathbf{e}}_M]$, *such that* $\tilde{\mathbf{e}}_M = vec(\mathbf{I}_M)$ *and* $\mathbf{B}, \mathbf{A}^* \odot \mathbf{A} \in \mathbb{C}^{M^2 \times K}$ *are given as*

$$[\mathbf{B}]_{\tilde{m},k} = j\pi(d_i - d_j)e^{j\pi(d_i - d_j)\sin\theta_k}\cos\theta_k p_k$$
$$[\mathbf{A}^* \odot \mathbf{A}]_{\tilde{m},k} = e^{j\pi(d_i - d_j)\sin\theta_k}$$

where $\tilde{m} = (j-1)M + i$, *and* $1 \le i, j \le M$.

Let $\mathcal{D}$ denote the difference set of sensor locations, defined as $\mathcal{D} = \frac{\lambda}{2}\{d_m - d_n, 1 \le$

[3] We suppress the dependence of matrices $\mathbf{G}, \mathbf{H}, \mathbf{B}, \mathbf{A}$ to the parameters ψ for notational simplification.

$m,n \leq M\}$. Suppose the physical array is such that $\mathcal{D}$ is a ULA with M_d elements, satisfying $\mathcal{D} = \{p\frac{\lambda}{2}, -M_d \leq p \leq M_d\}$. If $\mathbf{W} = \mathbf{I}$ (i.e., the measurements are directly collected at the output of the antenna array, without the mixing matrix $\mathbf{W}$), it is shown in [LV17, KP16a] that

$$\text{rank}(\mathbf{J}(\psi)) = \text{rank}(\mathbf{G}) = 2K+1, \quad \text{if } K \leq M_d \tag{2.89}$$

Invertibility of the FIM is an important consideration for parameter estimation. It is well-known that if the FIM is not invertible, there will not exist an unbiased estimator for the parameters $\mathbf{p}, \theta, \sigma^2$ with finite variance [SM01b, KP16a]. The authors in [GZC18] derived the following *necessary conditions* for the FIM to be invertible (or equivalently, sufficient conditions for the FIM to be non-invertible) [4]:

Corollary 3. *[GZC18] If $N^2 > |\mathcal{D}|$, for any choice of distinct K DOAs, the FIM in* (2.88) *is singular if $K > |\mathcal{D}|/2$. If $N^2 \leq |\mathcal{D}|$, then FIM is singular if $K > \lfloor (N^2-1)/2 \rfloor$.*

However, Corollary 3 *does not provide sufficient conditions in terms of K, M and N, under which the FIM will be guaranteed to be invertible*. Obtaining such a sufficient condition is important, since it will enable us to operate in the correct parametric regime (in terms of N and M) for localizing K DOAs. We bridge this important gap by highlighting the roles of two well-known array geometries: ULA and nested arrays.

2.5.2 Sufficient Conditions for Existence of CRB for Compressed Arrays

In this section, we develop sufficient conditions (in terms of K, N and M) for the existence of CRB for compressed arrays. Let us first state the following preliminary lemma:

Lemma 1. *[SS94, GR09, JStB01] Let $f : \mathbb{R}^N \to \mathbb{R}$ be an analytic function. If there exists an element $\mathbf{x_0} \in \mathbb{R}^N$ such that $f(\mathbf{x}) \neq 0$, then the set $\{\mathbf{x} : f(\mathbf{x}) = 0\}$ has Lebesgue measure zero.*

[4] With slight adjustments to account for having $2K+1$ parameters instead of $2K+2$ parameters in [GZC18]

Equipped with Lemma 1, we are ready to state the following important lemma, which tests invertibility of the FIM using properties of the zero-set of a real-valued multivariate polynomial.

Lemma 2. *Given an antenna array whose antennas are located on $d_1, d_2, \cdots, d_M$, if there exists a matrix $\mathbf{W}_0 \in \mathbb{C}^{N\times M}$ for which the FIM defined in* (2.88) *with $\mathbf{W} = \mathbf{W}_0$ is nonsingular, then for almost all $\mathbf{W} \in \mathbb{C}^{N\times M}$ (except possibly for a set of measure zero), the FIM will continue to be invertible (i.e., the CRB exists). Moreover, if $2K+1 > N^2$, the FIM is necessarily singular.*

Proof. When $2K+1 > N^2$, the matrix $\mathbf{G}$ becomes a fat matrix (more columns than the number of rows), hence $\mathbf{J}$ will necessarily be a singular matrix. The sufficient condition for invertibility of $\mathbf{J}$ can be proved as follows: we can write the matrix $\mathbf{G}$ in terms of its real and imaginary parts as $\mathbf{G} = \mathbf{G}_r + j\mathbf{G}_i$. If $\mathbf{v} \in \mathcal{N}(\mathbf{G})$, and $\mathbf{v} = \mathbf{v}_r + j\mathbf{v}_i$ then [5] it is straightforward to show that $\mathbf{v} \in \mathcal{N}(\mathbf{G})$ is equivalent to $\tilde{\mathbf{G}}\tilde{\mathbf{v}} = \mathbf{0}$, where

$$\tilde{\mathbf{G}} = \begin{bmatrix} \mathbf{G}_r & -\mathbf{G}_i \\ \mathbf{G}_i & \mathbf{G}_r \end{bmatrix}, \text{ and } \tilde{\mathbf{v}} = \begin{bmatrix} \mathbf{v}_r \\ \mathbf{v}_i \end{bmatrix}$$

Therefore, $\mathbf{G}$ is full column rank if and only if $\det(\tilde{\mathbf{G}}^T\tilde{\mathbf{G}}) \neq 0$. It is easy to see that $\det(\tilde{\mathbf{G}}^T\tilde{\mathbf{G}})$ is a polynomial in terms of the elements of $\mathbf{W}_r, \mathbf{W}_i$, i.e, the real part and imaginary parts of $\mathbf{W}$. Hence, $f(\mathbf{W}_r, \mathbf{W}_i) = \det(\tilde{\mathbf{G}}^T\tilde{\mathbf{G}})$ is an analytic function with $2MN$ real-valued variables. Hence, if we are able to show that $f(\mathbf{W}_r, \mathbf{W}_i)$ is a nontrivial polynomial, then based on the result from Lemma 1 the set $\{[\mathbf{W}_r, \mathbf{W}_i] | f(\mathbf{W}_r, \mathbf{W}_i) = 0\}$ has measure zero. To show that $f(\mathbf{W}_r, \mathbf{W}_i)$ is a nontrivial polynomial, it suffices to find one example $\mathbf{W}_{0,r}$, $\mathbf{W}_{0,i}$ which guarantees $f(\mathbf{W}_{0,r}, \mathbf{W}_{0,i}) \neq 0$. We let $\mathbf{W}_{0,r}$ (resp. $\mathbf{W}_{0,i}$) to be the real (resp. imaginary) part of $\mathbf{W}_0$. □

Hence, if we can construct a specific $\mathbf{W}_0 \in \mathbb{C}^{N\times M}$ such that the FIM $\mathbf{J}(\psi)|_{\mathbf{W}=\mathbf{W}_0}$ is non-singular, then the FIM is also guaranteed to be non-singular for *almost all choices of* $\mathbf{W} \in \mathbb{C}^{N\times M}$. The construction of $\mathbf{W}_0$ is highly dependent on the array geometry and reveal different dependence

[5] Here $\mathcal{N}(.)$ denotes the null-space of a matrix.

of K on N and M. In the following, we explicitly construct such a $\mathbf{W}_0$ separately for ULA and nested array.

Existence of CRB for Standard Uniform Linear Arrays

For a ULA, the sensor locations satisfy $d_m = m, 1 \leq m \leq M$. The following theorem shows that as long as $K \leq \min(M, O(N^2))$, the CRB exists for almost all compression matrices $\mathbf{W} \in \mathbb{C}^{N\times M}$.

Theorem 16. *Consider a ULA with $d_m = m$, $1 \leq m \leq M$. Let $N_0 = \lfloor 2\sqrt{M+1} - 1 \rfloor$. For such an array, the FIM $\mathbf{J}(\psi)$ given by* (2.88) *is non-singular for almost all $\mathbf{W} \in \mathbb{C}^{N\times M}$ (except possibly for a set of measure zero), if $K \leq N_2(N_1+1) - 1$ where N_1, N_2 are integers given by*

$$(N_1, N_2) = \left(\lfloor \frac{\min(N, N_0)}{2} \rfloor, \lceil \frac{\min(N, N_0)}{2} \rceil \right) \tag{2.90}$$

Proof. The theorem follows from Lemma 2 provided we can construct a $\mathbf{W}_0^{(N)} \in \mathbb{C}^{N\times M}$ in the regimes specified by (2.90) so that $\mathbf{J}(\psi)|_{\mathbf{W}=\mathbf{W}_0^{(N)}}$ is invertible. We construct $\mathbf{W}_0^{(N)}$ by considering the following two cases:

- Case 1: $N \leq N_0$: Notice that in this case $N = N_1 + N_2$. We construct $\mathbf{W}_0^{(N)}$ as follows:

$$[\mathbf{W}_0^{(N)}]_{n,m} = \begin{cases} 1, m = n, 1 \leq n \leq N_1 \\ 1, m = (N_1+1)(n-N_1), N_1+1 \leq n \leq N_1+N_2 \\ 0, \text{otherwise.} \end{cases} \tag{2.91}$$

  Notice that $\mathbf{W}_0^{(N)}$ acts as a selection matrix that selects a subset $\mathcal{S}_{\text{NA}} \subseteq \{1, \cdots, M\}$ of size $|\mathcal{S}_{\text{NA}}| = N$, such that

$$\mathcal{S}_{\text{NA}} = \frac{\lambda}{2}\{1, 2, \cdots, N_1\} \cup \{m(N_1+1)\frac{\lambda}{2}, 1 \leq m \leq N_2\} \tag{2.92}$$

$\mathcal{S}_{\text{NA}}$ can be identified as a nested array with N_1+N_2 sensors [PV10]. The difference set of such an array is given by

$$\mathcal{D}_{\text{NA}} = \frac{\lambda}{2}\{-M_d, \cdots, M_d : M_d = (N_1+1)N_2 - 1\} \tag{2.93}$$

Define $\mathbf{A}_0(\theta) = \mathbf{W}_0^{(N)}\mathbf{A}(\theta)$. Then, it is easy to verify that

$$[\mathbf{A}_0(\theta)]_{m,k} = e^{j\pi z_m \sin(\theta_k)}, 1 \le k \le K, z_m \in \mathcal{S}_{\text{NA}}.$$

Hence, $\mathbf{A}_0(\theta)$ represents the array manifold of a nested array with N sensors (N_1 sensors in inner and N_2 sensors in outer ULAs). Given $\mathbf{W}_0$, the compressed measurement vectors $\mathbf{y}_0(t) \triangleq \mathbf{y}(t)|_{\mathbf{W}=\mathbf{W}_0^{(N)}}$ are given by (2.86):

$$\mathbf{y}_0(t) = \mathbf{W}_0^{(N)}\mathbf{A}(\theta)\mathbf{x}(t) + \mathbf{W}_0^{(N)}\mathbf{n}(t) = \mathbf{A}_0(\theta)\mathbf{x}(t) + \mathbf{n}_0(t) \tag{2.94}$$

where $\mathbf{n}_0(t)$ is a subvector of $\mathbf{n}(t)$ and therefore it continues to be i.i.d. Gaussian white noise, distributed as $\mathbf{n}_0(t) \sim \mathbb{C}\mathcal{N}(0, \sigma_n^2\mathbf{I}_N)$. Therefore, (2.94) represents the measurement model of the signal received at a nested array with $N_1+N_2 = N$ sensors, contaminated with additive white Gaussian noise, and $\mathbf{J}_0(\psi) \triangleq \mathbf{J}(\psi)|_{\mathbf{W}=\mathbf{W}_0^{(N)}}$ represent the Fisher Information Matrix (FIM) corresponding to the measurements $\mathbf{y}_0(t), t = 1, 2, \cdots, T$. Since $\mathcal{D}_{\text{NA}}$ is a ULA with $2M_d+1$ elements, we can use (2.89) to conclude that $\mathbf{J}_0(\psi)$ is non singular as long as $K \le M_d = (N_1+1)N_2 - 1$.

- Case 2, $N > N_0$: In this case, $N_1+N_2 = N_0$. Let $\Delta_N = N - N_0$ and let $\overline{\mathcal{S}_{\text{NA}}}$ be an ordered set [6] defined as $\overline{\mathcal{S}_{\text{NA}}} = \frac{\lambda}{2}\{1, 2, \cdots, M\} \backslash \mathcal{S}_{\text{NA}}$, where $\mathcal{S}_{\text{NA}}$ is given by (2.92). We construct $\mathbf{W}_0^{(N)}$ as $\mathbf{W}_0^{(\text{aug})} = [(\mathbf{W}_0^{(N_0)})^T \ \ \mathbf{C}_0^T]^T$ where $\mathbf{W}_0^{(N_0)} \in \mathbb{R}^{N_0 \times M}$ is constructed according to (2.91), and

[6] Elements of $\overline{\mathcal{S}_{\text{NA}}}$ are indexed in ascending order.

$\mathbf{C}_0 \in \mathbb{R}^{\Delta_N \times M}$ is constructed as

$$[\mathbf{C}_0]_{n,m} = \begin{cases} 1, & m = [\overline{\mathcal{S}_{\text{NA}}}]_n \\ 0, & \text{otherwise.} \end{cases}$$

Here, the notation $[\overline{\mathcal{S}_{\text{NA}}}]_n$ indicates the nth element (in ascending order) of the set $\overline{\mathcal{S}_{\text{NA}}}$. As before, let $\mathbf{A}_0^{(\text{aug})} \triangleq \mathbf{W}_0^{(\text{aug})}\mathbf{A}(\theta)$. Then, $\mathbf{A}_0^{(\text{aug})}(\theta)$ can be identified as the manifold of an "augmented" nested array, with $[\mathbf{A}_0^{(\text{aug})}]_{m,k} = e^{jz_m \sin(\theta_k)}$, $z_m \in \mathcal{S}_{\text{aug-NA}}$ where $\mathcal{S}_{\text{aug-NA}}$ is the set of sensor locations of this augmented nested array given by $\mathcal{S}_{\text{aug-NA}} = \mathcal{S}_{\text{NA}} \cup \mathcal{C}$ with $\mathcal{C} = \{[\overline{\mathcal{S}_{\text{NA}}}]_n : 1 \leq n \leq \Delta_n\}$. It can be verified that the difference set of $\mathcal{S}_{\text{aug-NA}}$ is a ULA given by $\mathcal{D}_{\text{aug-NA}} = \frac{\lambda}{2}\{-M_d, \cdots, M_d : M_d = N_2(N_1+1)-1\}$. As before, $\mathbf{y}_0^{(\text{aug})}(t) \triangleq \mathbf{y}(t)|_{\mathbf{W}=\mathbf{W}_0^{(\text{aug})}}$ represents the measurement model of an augmented nested array with sensors at $\mathcal{S}_{\text{aug-NA}}$, contaminated with white Gaussian noise, and $\mathbf{J}_0(\psi) = \mathbf{J}(\psi)|_{\mathbf{W}=\mathbf{W}_0^{(\text{aug})}}$ is the corresponding FIM. Since $\mathcal{D}_{\text{aug-NA}}$ is again a ULA with $2M_d+1$ elements, we can use (2.89) to conclude that $\mathbf{J}_0(\psi)$ is non singular as long as $K \leq M_d = N_2(N_1+1)-1$.

□

Remark 16. *It is possible to improve the upper-bound provided in Theorem 16 (up to constant scaling factors), if we use enhanced versions of the nested array, such as [YSYC16, ZWKZ19, RDL+20] to construct* $\mathbf{W}_0^{(N)}$. *We plot such a bound derived using the array structure proposed in [RDL+20] in our simulations.*

Existence of CRB for Modified Nested Arrays

We consider a specific class of nested arrays with M antennas (henceforth referred to as modified nested array, or mod-NA), where we adjust the number of elements and inter-element

spacings as a function of N and M. Let $n_1 = \lfloor \frac{N}{2} \rfloor$. We choose the sensor locations as

$$\mathcal{S}_{\text{mod-NA}} = \frac{\lambda}{4}\{1,2,\cdots,2n_1\} \cup \frac{\lambda}{2}\{n_1 m, 2 \leq m \leq (M-2n_1+1)\}$$

Notice that the spacing in the inner ULA is $\lambda/4$, which is half the standard spacing of $\lambda/2$. For this nested array, the following theorem shows a construction of $\mathbf{W}_0$ such that $\mathbf{J}(\psi)|_{\mathbf{W}=\mathbf{W}_0}$ is non-singular.

Theorem 17. *Consider an antenna array whose sensor locations are given by the set $\mathcal{S}_{mod\text{-}NA}$. For such an array, the FIM $\mathbf{J}(\psi)$ given by* (2.88) *is non-singular for almost all $\mathbf{W} \in \mathbb{C}^{N\times M}$ (except possibly for a set of measure zero) whenever $N \leq \frac{2M}{3}$, and $K \leq n_1(N-n_1+1)-1$.*

Proof. Similar to the proof of Theorem 16, it is sufficient to find a matrix $\mathbf{W}_0 \in \mathbb{C}^{N\times M}$ such that the FIM in (2.88) is non-singular. Lemma 2 will then imply that $\mathbf{J}(\psi)$ will be non-singular for almost all $\mathbf{W} \in \mathbb{C}^{N\times M}$. We construct $\mathbf{W}_0$ as follows:

$$[\mathbf{W}_0]_{n,m} = \begin{cases} 1, m = 2n, & 1 \leq n \leq n_1 \\ 1, m = n_1 + n, & n_1 + 1 \leq n \leq N \\ 0, & \text{otherwise.} \end{cases} \tag{2.95}$$

Noice that $N \leq \frac{2M}{3} \Rightarrow N + n_1 \leq M$. Hence, the Nth (or last) row of $\mathbf{W}_0$ contains a single 1 in the $(N+n_1)$th column, while ensuring that $N+n_1 \leq M$. Let $\mathbf{A}_0 \triangleq \mathbf{W}_0\mathbf{A}(\theta)$. Then, $\mathbf{A}_0(\theta)$ can be identified as the manifold of a nested array with N antennas whose locations are given by the set

$$\mathcal{S}_{\text{NA}}^{(N)} = \frac{\lambda}{4}\{2,4,\cdots,2n_1\} \cup \frac{\lambda}{2}\{2n_1, 3n_1, \cdots, (N-n_1+1)n_1\}$$

It is easy to see that the difference set of $\mathcal{S}_{\text{NA}}^{(N)}$ is a standard ULA given by $\mathcal{D}_{\text{NA}}^{(N)} = \frac{\lambda}{2}\{-M_d, \cdots, M_d : M_d = n_1(N-n_1+1)-1\}$. Hence, $\mathbf{y}_0(t) \triangleq \mathbf{y}(t)|_{\mathbf{W}=\mathbf{W}_0}$ represents the measurement model of an N-element nested array with sensors at $\mathcal{S}_{\text{NA}}^{(N)}$, contaminated with white Gaussian noise, and $\mathbf{J}_0(\psi) =$

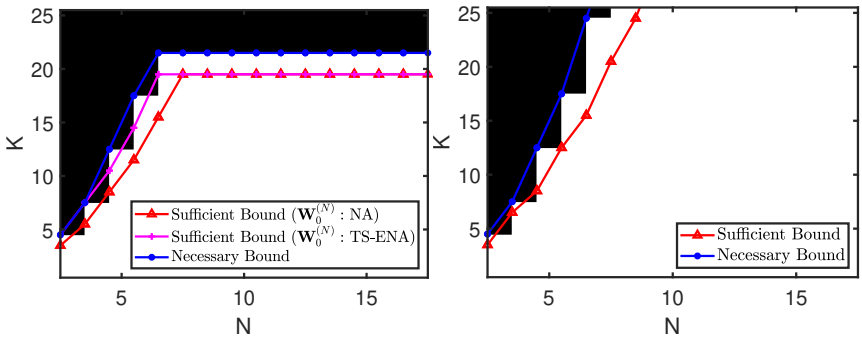
$\mathbf{J}(\psi)|_{\mathbf{W}=\mathbf{W}_0}$ is the corresponding FIM. Since $\mathcal{D}_{\text{NA}}$ is again a standard ULA with $2M_d+1$ elements, we can use (2.89) to conclude that $\mathbf{J}_0(\psi)$ is non singular as long as $K \leq M_d = n_1(N-n_1+1)-1$.

□

Remark 17. Compressed Nested Array versus ULA: *Notice that* $n_1(N-n_1+1)-1$ *scales as* $\Theta(N^2)$. *Hence, Theorem 17 dictates that for a compressed nested array with* $N < 2M/3$, *the FIM is invertible for almost all* $\mathbf{W}$ *as long as* $K = O(N^2)$. *In particular, if N scales linearly with M (i.e.* $N = \alpha M$ *for some constant* $\alpha < 2/3$*), then the FIM is nonsingular even when* $K = O(M^2)$. *On the other hand, for a compressed ULA, Theorem 16 dictates that the FIM is nonsingular as long as* $K = \min\left(O(N^2), M\right)$. *In particular, if N scales linearly with M, the FIM is non singular only if* $K < M$. *Hence our results show that for almost all choices of* $\mathbf{W}$, *even after compression, nested arrays potentially allow localization of* $K = O(M^2)$ *sources as long as N scales linearly with M.*

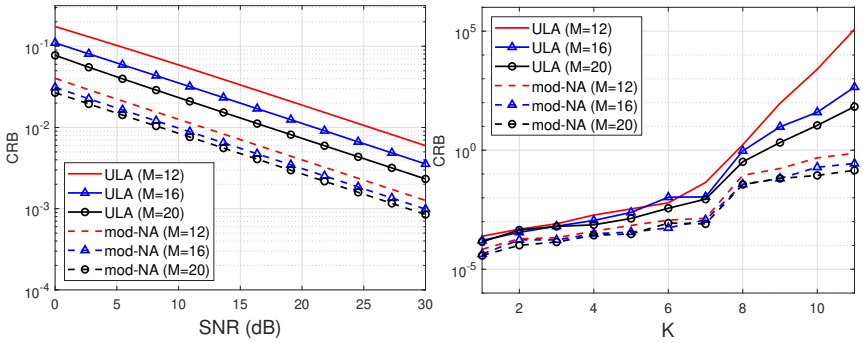
2.5.3 Simulations

We consider a ULA and a modified nested array with $M = 22$ antennas. For each value of K the sources are located uniformly on the range $[-\frac{(K-1)\pi}{2K}, \frac{(K-1)\pi}{2K}]$, with equal power $p_k = 1$, $k = 1, \cdots, K$ and noise power $\sigma^2 = 0.1$, with SNR defined as $\text{SNR} = 10\log_{10}(\frac{\frac{\sum_{k=1}^{K} p_k}{K}}{\sigma^2})$. We assume that $\mathbf{W}_{n,m} = e^{j\phi_{nm}}$ where ϕ_{nm} are i.i.d random variables distributed uniformly on $[0, 2\pi]$. In Figure (2.18a) and (2.18b), for each pair of (N, K), we plot the empirical probability with which the FIM in (2.88) is invertible, with white indicating probability 1 and black indicating 0. In each run, the necessary and sufficient bounds derived in Theorems 16, 17 are also overlaid, which agree with the empirical plots. In Figure 2.18a, we show the sufficient bound derived from nested array, as suggested by Theorem 16 (denoted as $\mathbf{W}_0^{(N)}$: NA). We also overlay a sufficient bound that is derived by choosing $\mathbf{W}_0^{(N)}$ in the proof of Theorem 16 to generate a two-sided extended nested array (TS-ENA) [RDL$^+$20] (denoted as $\mathbf{W}_0^{(N)}$: TS-ENA). This has the same order-wise scaling but better constants than the nested array.



(a) ULA (b) Modified nested array

Figure 2.18: Phase-transition for nonsingularity of FIM in (2.88) for different array geometries.



(a) CRB vs SNR (b) CRB vs K

Figure 2.19: CRB of compressed ULA and compressed modified nested array, as a function of (a) SNR and (b) K values.

In Figure 2.19, we plot the CRB as a function of SNR and K. We assume $N = 6, M \in \{12, 16, 20\}$, and fix the compression matrix $\mathbf{W}$. In Figure 2.19a, we assume that $K = 4$ and the CRB is evaluated for both ULA and modified nested arrays. As the SNR increases, the CRB decreases, and is always smaller for the compressed nested array configuration. In Figure 2.19b, where we assume SNR $= 40dB$, we see that as the number of sources increases, the CRB also increases. Moreover, we can observe a jump in the CRB in the overcomplete regime where $K > N$. A more detailed study of the CRB versus SNR and K can be an interesting topic of future research.

2.5.4 Conclusion

We developed the first results that derive sufficient conditions for the existence of the CRB for compressed arrays, by explicitly highlighting the role of array geometry. We studied the role of compression in both uniform and sparse antenna arrays, and established sufficient conditions under which the Cramér Rao bound is guaranteed to exist for a generic $N \times M$ compression matrix $\mathbf{W}$. For a uniform linear array, the FIM will be invertible if $K = \min(O(N^2), O(M))$. When suitably designed sparse arrays (such as nested arrays) are used, one is able to further improve this to $K = \min(O(N^2), O(M^2))$.

2.6 Appendix

2.6.1 Proof of Theorem 4

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

Definition 4. *Let $\bar{\mathbf{A}}_{\text{ca},0} \in \mathbb{C}^{M_{\text{ca}} \times N_\theta}$ be the matrix constructed from $\mathbf{A}_{\text{ca},0}$ by removing the repeated rows and sorting them in order such that first row coresponds to difference of $-\frac{M_{\text{ca}}-1}{2}$ and last row corresponds to difference of $\frac{M_{\text{ca}}-1}{2}$.*

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

Recall from Theorem 3 that non singularity of $\mathbf{J}$ is equivalent to $\mathbf{B} = [\mathbf{A}_{\text{ca},0}\ \mathbf{H}]$ 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 $\mathbf{H} = (\tilde{\mathbf{H}}_\delta)_{:,2:M}$, i.e., the matrix comprised by the last $M-1$ columns of $\tilde{\mathbf{H}}_\delta$. After establishing that both $\mathbf{H}$ 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 $\mathbf{H}$ 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 $\mathbf{R}_{\delta_i} \triangleq \frac{\partial \mathbf{R}_\mathbf{y}}{\partial \delta_i} = \mathbf{A}\mathbf{R}_\mathbf{x}\mathbf{D}^H_{\delta_i} + \mathbf{D}_{\delta_i}\mathbf{R}_\mathbf{x}\mathbf{A}^H$, and $\mathbf{D}_{\delta_i} = \frac{\partial \mathbf{A}}{\partial \delta_i}$.

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

$$\left(\tilde{\mathbf{H}}_\delta\right)_{q,r} = \begin{cases} \lambda_{r,s} & q = (s-1)M + r, 1 \le s \le M, s \ne r \\ -\lambda_{s,r} & q = (r-1)M + s, 1 \le s \le M, s \ne r \\ 0 & \text{else} \end{cases} \tag{2.96}$$

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 $\lambda_{r,s} = \sum_{k=1}^{N_\theta} j\gamma_k \frac{2\pi k}{N_\theta} e^{j\frac{2\pi k}{N_\theta}(d_r - d_s)}$, 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 3. *For almost all* $\gamma \in \mathbb{R}^{N_\theta}$, $\mathbf{H}$ *has full column rank.*

Proof. For almost all $\gamma \in \mathbb{R}^{N_\theta}$ 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 γ, which holds only for a set of measure zero in $\mathbb{R}^{N_\theta}$. Thus, looking at the rows 2 through M of $\mathbf{H}$ (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 γ. Therefore, $\mathbf{H}$ has rank $M-1$, i.e., $\mathbf{H}$ is full column rank for almost all γ. □

Observe that

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

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 $\mathbf{H}$ is full column rank for almost all γ (Lemma 3), the only way for $\mathbf{B}$ to be column rank deficient is when there exist *non zero* $\alpha \in \mathbb{C}^{N_\theta}$, $\tilde{\beta} \in \mathbb{C}^{M-1}$ such that

$$\mathbf{A}_{\text{ca},0}\alpha = \mathbf{H}\tilde{\beta}. \tag{2.97}$$

We first show that there exist no $\alpha \neq 0$ and $\beta \neq 0$ such that $\mathbf{A}_{\text{ca},0}\alpha = \tilde{\mathbf{H}}_\delta\beta$. This will, in particular imply non existence of non zero α and $\tilde{\beta}$ such that ((2.97)) holds. We prove this by contradiction, i.e. let us assume there exist nonzero $\alpha \in \mathbb{C}^{N_\theta}$, $\beta \in \mathbb{C}^M$ satisfying $\mathbf{A}_{\text{ca},0}\alpha = \tilde{\mathbf{H}}_\delta\beta$. This means that for every $1 \leq r,s \leq M$ we have

$$\beta_r - \beta_s = \frac{\sum_{k=1}^{N_\theta} \alpha_k e^{j\frac{2\pi k}{N_\theta}m}}{\lambda_{(m)}} := f_\alpha(m), \tag{2.98}$$

for $d_r - d_s = m$.

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

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

Upon substitution in (2.97) and using the definition of $f_\alpha(m)$ from (2.98), we obtain

$$\bar{\mathbf{A}}_{\text{ca},0}\alpha = f_\alpha(1) \times \begin{bmatrix} -(M-1)\lambda_{(-M+1)} \\ \vdots \\ -\lambda_{(-1)} \\ 0 \\ \lambda_{(1)} \\ \vdots \\ (M-1)\lambda_{(M-1)} \end{bmatrix} := \mathbf{c} \tag{2.99}$$

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

For simplicity, we index the rows of (2.99) from $-(M-1)$ to $M-1$. Let $p = -(M-1)$. From (2.98)

$$f_\alpha(p) = \frac{\sum_{k=1}^{N_\theta} \alpha_k e^{j2\pi pk/N_\theta}}{\lambda_{(p)}} = p, \tag{2.100}$$

Notice that $f_\alpha(p + N_\theta) = f_\alpha(p)$. 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

$$f_\alpha(p+N_\theta) = \frac{\sum_{k=1}^{N_\theta} \alpha_k e^{j2\pi(p+N_\theta)k/N_\theta}}{\lambda_{(p+N_\theta)}} = p+N_\theta$$
$$\Rightarrow f_\alpha(p) = \frac{\sum_{k=1}^{N_\theta} \alpha_k e^{j2\pi pk/N_\theta}}{\lambda_{(p)}} = p+N_\theta \tag{2.101}$$

However, (2.100) and (2.101) cannot hold at the same time since $p+N_\theta \neq p$. Therefore, the range spaces of $\mathbf{A}_{\text{ca},0}$ and $\mathbf{H}$ do not coincide except for the zero vector. Using the facts that $\mathbf{H}$ is full column rank for almost all γ, $\mathbf{A}_{\text{ca},0}$ is full column rank for $N_\theta \leq 2M-1$, and the range spaces of $\mathbf{A}_{\text{ca},0}$ and $\mathbf{H}$ do not intersect as long as $N_\theta \leq 2M-2$, we conclude that $\mathbf{B}$ has full column rank for almost all γ and δ as long as $N_\theta \leq 2M-2$.

2.6.2 Proof of Theorem 5

Proof. We prove for a slightly modified version of the nested array which is defined in (2.32).

We follow the same lines of proof of Theorem 4 up to (2.98) since the argument upto this point applies to any array geometry. For a nested array, the explicit structure of the vector $\mathbf{H}\beta$ will be quite different and we now examine it more closely. We use the same definition for $f_\alpha(m)$ as in (2.98).

Let $f_\alpha(m)$ be defined as in (2.98). Assuming M to be an even number, for a modified nested array which is defined in (2.32), we have

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

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

$$\beta_i = \beta_1 + (i-1) f_\alpha(1). \tag{2.102}$$

Moreover, we have

$$\beta_{\frac{M}{2}+1} - \beta_1 = f_\alpha(\frac{M}{2}) \tag{2.103}$$

$$\beta_{\frac{M}{2}+i} - \beta_{\frac{M}{2}+i-1} = \beta_{\frac{M}{2}+1} - \beta_1 = f_\alpha(\frac{M}{2}), \tag{2.104}$$

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

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

From (2.102) and (2.103), we get $\frac{M}{2} f_\alpha(1) = f_\alpha(\frac{M}{2})$.

Therefore, we have

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

Similar to the proof of Theorem 4, 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 (2.100) and (2.101) cannot hold simultaneously, unless $f_\alpha(1) = 0$. 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 $N_\theta \leq M^2/2$.

□

2.6.3 Proof of Theorem 7

Proof. Due to Theorem 3, in order to ensure that the FIM is invertible, we only need to show that $\mathbf{B} = [\mathbf{A}_{\text{ca}}\mathbf{H}]$ is full column rank. A sparse γ only changes the explicit form of $\mathbf{H}$ since the entries $\lambda_{(m)}$ of $\mathbf{H}$ are now given by:

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

where j_k indicates the index of the kth nonzero element of γ, and w_{j_k} denotes its corresponding spatial frequency on the grid. As in the proofs of Theorems 4, 5, $\mathbf{B}$ is of full column rank if $\mathbf{A}_{\text{ca}}$ and $\mathbf{H}$ are full column rank and there exist no non zero $\alpha, \tilde{\beta}$ such that $\mathbf{A}_{\text{ca}}\alpha = \mathbf{H}\tilde{\beta}$. The proof for non existence of non zero α and β follow the same lines as earlier. However, we only need to establish conditions for full column rank of $\mathbf{H}$. As argued in the proof of Theorem 4, the structure of $\mathbf{H}$ in (2.96) dictates that it has full column rank $M-1$ for almost all sparse γ with $\|\gamma\|_0 = K$, if $\lambda_{(m)}$ is nonzero for every $0 \leq m \leq M_{\text{ca}}$. Let $\tilde{\gamma} = [\gamma_{j_1}, \cdots, \gamma_{j_K}]$. We see that $\lambda_{(m)} = 0$ for a particular m describes a linear relation between the elements of $\tilde{\gamma}$:

$$0 = \sum_{k=1}^{K} j\tilde{\gamma}_k\omega_{j_k}e^{jm\omega_{j_k}}$$

Hence, $\tilde{\gamma}$ has a measure zero in $\mathbb{R}^K$. Hence, for almost all K-sparse γ, $\lambda_{(m)} \neq 0$. This implies that $\mathbf{H}$ has full column rank, and we can repeat the rest of the proofs of Theorems 4, 5 to establish that $\mathbf{J}$ is non singular for almost all $K-$ sparse $\gamma \in \mathbb{R}^{N_\theta}, \|\gamma\|_0 \leq K$. □

2.6.4 Proof of Theorem 13

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

$$R_{i_k+k+1,i_k+k} = f_1 + \lambda_1(\delta_{i_k+k+1} - \delta_{i_k+k}) \tag{2.105}$$

$$R_{i_k+1,i_k} = f_1 + \lambda_1(\delta_{i_k+1} - \delta_{i_k}) \tag{2.106}$$

Similarly,

$$R_{i_k+k+1,i_k+1} = f_k + \lambda_k(\delta_{i_k+k+1} - \delta_{i_k+1}) \tag{2.107}$$

$$R_{i_k+k,i_k} = f_k + \lambda_k(\delta_{i_k+k} - \delta_{i_k}) \tag{2.108}$$

Subtracting (2.105) from (2.106), and also (2.107) from (2.108) we obtain

$$\beta_k = \frac{R_{i_k+k+1,i_k+1} - R_{i_k+k,i_k}}{R_{i_k+k+1,i_k+k} - R_{i_k+1,i_k}} = \frac{\lambda_k}{\lambda_1} \tag{2.109}$$

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

$$\delta_{i_k+k+1} - \delta_{i_k+k} - \delta_{i_k+1} + \delta_{i_k} \neq 0, \tag{2.110}$$

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

Notice that for a fixed k, different instances of equation (2.109) 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 (2.110) holds for all i_k. Hence, in the sequel, we will assume $i_k = 1$.

From and (2.69) and (2.70), we can verify that

$$\bar{r}_k = kf_1 + \lambda_1(\delta_{k+1} - \delta_1) \tag{2.111}$$

From (2.108), (2.111) we get, for $2 \leq k \leq M-2$,

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

which can be expressed as

$$C_{k-1,1} f_1 + C_{k-1,k} f_k = h_{k-1}, \quad 2 \leq k \leq M-2 \tag{2.112}$$

where $C_{k-1,1}, C_{k-1,k}$, and h_{k-1} are given in (2.73), (2.74), and (2.75). We can express (2.112) in a more compact and explicit form as

$$\mathbf{Cf} = \mathbf{h} \tag{2.113}$$

in which

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

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

□

2.6.5 Proof of Theorem 14

Proof. It can be easily verified that

$$\bar{r}_i^{(1)} = (i-1)f_{N_1} + (\delta_i^{(1)} - \delta_1^{(1)})\lambda_{N_1} \tag{2.114}$$

$$\bar{r}_j^{(2)} = (j-1)f_{N_2} + (\delta_j^{(2)} - \delta_1^{(2)})\lambda_{N_2} \tag{2.115}$$

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 $1 \le i \le N_2$, and $1 \le j \le 2N_1$ 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 (2.114) and (2.115) for $\bar{i}, \bar{j}$. Subtracting those equations from (2.114) and (2.115), we get

$$\bar{r}_{\bar{i}}^{(1)} - \bar{r}_i^{(1)} = N_2 f_{N_1} + (\delta_{\bar{i}}^{(1)} - \delta_i^{(1)})\lambda_{N_1} \tag{2.116}$$

$$\bar{r}_{\bar{j}}^{(2)} - \bar{r}_j^{(2)} = N_1 f_{N_2} + (\delta_{\bar{j}}^{(2)} - \delta_j^{(2)})\lambda_{N_2} \tag{2.117}$$

   We also know that

$$R_{ij}^{(12)} = f_k + (\delta_i^{(1)} - \delta_j^{(2)})\lambda_k \tag{2.118}$$

$$R_{\bar{i}\bar{j}}^{(12)} = f_k + (\delta_{\bar{i}}^{(1)} - \delta_{\bar{j}}^{(2)})\lambda_k \tag{2.119}$$

   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 $N_1 N_2$). For this particular sensor, we can write equations (2.114) and (2.115).

$$\bar{r}^{(1)}_{N_2+1} = N_2 f_{N_1} + (\delta^{(1)}_{N_2+1} - \delta^{(1)}_1)\lambda_{N_1} \tag{2.120}$$

$$\bar{r}^{(2)}_{N_1+1} = N_1 f_{N_2} + (\delta^{(2)}_{N_1+1} - \delta^{(2)}_1)\lambda_{N_2} \tag{2.121}$$

where $\delta^{(1)}_{N_2+1} = \delta^{(2)}_{N_1+1}$ since they are the same sensor, and also $\delta^{(1)}_1 = \delta^{(2)}_1$ 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.

$$\bar{r}^{(1)}_{2N_2+1} = 2N_2 f_{N_1} + (\delta^{(1)}_{2N_2+1} - \delta^{(1)}_1)\lambda_{N_1} \tag{2.122}$$

$$\bar{r}^{(2)}_{2N_1+1} = 2N_1 f_{N_2} + (\delta^{(2)}_{2N_1+1} - \delta^{(2)}_1)\lambda_{N_2}, \tag{2.123}$$

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

From (2.120), (2.121), (2.122), (2.123), we get

$$\frac{2\bar{r}^{(2)}_{N_1+1} - \bar{r}^{(2)}_{2N_1+1}}{\lambda_{N_2}} = \frac{2\bar{r}^{(1)}_{N_2+1} - \bar{r}^{(1)}_{2N_2+1}}{\lambda_{N_1}} \tag{2.124}$$

and from equation, we obtain (2.79)

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

From equations (2.118), (2.119) from (2.116), (2.117), we get

$$\begin{aligned}\delta_{\bar{i}}^{(1)} - \delta_{\bar{j}}^{(2)} - \delta_i^{(1)} + \delta_j^{(2)} &= \frac{\bar{r}_{\bar{i}}^{(1)} - \bar{r}_i^{(1)} - N_2 f_{N_1}}{\lambda_{N_1}} \\ &\quad - \frac{\bar{r}_{\bar{j}}^{(2)} - \bar{r}_j^{(2)} - N_1 f_{N_2}}{\lambda_{N_2}} \\ &= \frac{R_{\bar{i}\bar{j}}^{(12)} - R_{ij}^{(12)}}{\lambda_k},\end{aligned}$$

whereby (2.80) follows, where $[\beta_{cp}]_k := \frac{\lambda_k}{\lambda_{N_1}}$.

Now, using (2.114), (2.115), we can write

$$\delta_i^{(1)} - \delta_j^{(2)} = \frac{\bar{r}_i^{(1)} - (i-1)f_{N_1}}{\lambda_{N_1}} - \frac{\bar{r}_j^{(2)} - (j-1)f_{N_2}}{\alpha\lambda_{N_1}}$$

and

$$R_{ij}^{(12)} = f_k + \beta_k(\bar{r}_i^{(1)} - (i-1)f_{N_1} - (\bar{r}_j^{(2)} - (j-1)f_{N_2})\alpha^{-1}) \tag{2.125}$$

which is linear in terms of elements of $\mathbf{f}$, and hence it is linear in terms of $\mathbf{p}$. By varying the indices i and j in the range $1 \le i \le N_2$ and $1 \le j \le 2N_1$ in (2.125), we obtain the corresponding rows of the system of equations $\mathbf{C}_{cp}\mathbf{f} = \mathbf{h}_{cp}$.

2. $k = d_j^{(2)} - d_i^{(1)}$:

In this case, (2.118), (2.119) should be rewritten as

$$(R_{ij}^{(12)})^* = R_{ji}^{(21)} = f_k + (\delta_j^{(2)} - \delta_i^{(1)})\lambda_k \tag{2.126}$$

$$(R_{\bar{i}\bar{j}}^{(12)})^* = R_{\bar{j}\bar{i}}^{(21)} = f_k + (\delta_{\bar{j}}^{(2)} - \delta_{\bar{i}}^{(1)})\lambda_k \tag{2.127}$$

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 1.

□

2.7 Acknowledgements

The contents in this chapter are reprint of the material published in "IEEE Signal Processing Letters" (Sec. 2.1 and Sec. 2.5), "IEEE Transactions on Signal Processing" (Sec. 2.2), "IEEE ICASSP 2017 Conference" (Sec. 2.3), and "Elsevier Journal on Digital Signal Processing" (Sec. 2.4).

The work presented in Sec. 2.1 was supported in part by "University of Maryland, College Park" and in part by the "Department of Defense". Sec. 2.2 was supported in part by the "Department of Defense", in part by the "NSF CAREER Award (ECCS 1553954)", in part by "the University of Maryland, College Park", and in part by the "University of California, San Diego". Sec. 2.3 was supported in part by "NSF CAREER award (ECCS 1553954)", and in part by the "University of California, San Diego". The work presented in Sec. 2.5 was supported by in part by the "Office of Naval Research grant (ONR N00014-18-1-2038)", and in part by the "National Science Foundation (NSF CAREER ECCS 1700506)".

Chapter 3

Sparse Support Recovery for Underdetermined Linear Problems

In chapter 2, we considered the problem of source localization, in the setting that the source locations were *not* assumed to be on a predefined grid. This Section, however, considers a different viewpoint. Here, we consider a more general problem setting that can have applications beyond source localization. In order to connect the contents of this chapter to source localization however, it is enough to assume the measurement matrix $\mathbf{A}$ in this chapter, is a discrete version of $\mathbf{a}(\theta)$ in Chapter 2, i.e., $\mathbf{A} = [\mathbf{a}(\theta_1), \cdots, \mathbf{a}(\theta_N)]$ for a predefined grid $\theta_1, \cdots, \theta_N$. In the following, we define the problem setting considered in this chapter:

Consider a set of L jointly sparse vectors[1] $\mathbf{x}[l] \in \mathbb{F}^N, \qquad l = 0, \cdots, L-1$, which share a common support $\mathcal{S}$ of size $K < N$,

$$\text{Supp}(\mathbf{x}[l]) = \mathcal{S}, \ \forall\, 0 \leq i \leq L-1, \ |\mathcal{S}| = K.$$

[1] Depending on the context, $\mathbb{F}$ denotes the real or complex field, i.e., $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$.

We obtain linear measurements of $\mathbf{x}[l]$ (contaminated with additive noise $\mathbf{w}[l]$) given by

$$\mathbf{y}[l] = \mathbf{A}\mathbf{x}[l] + \mathbf{w}[l], \quad l = 1, \cdots, L. \tag{3.1}$$

Here $\mathbf{A} \in \mathbb{F}^{M \times N}$ is the measurement matrix with $M < N$ and therefore, $\mathbf{y}[l]$ represent a set of compressed measurements of $\mathbf{x}[l], l = 1, 2, \cdots L$. This observation model lies at the heart of compressive sensing and sparse signal recovery [C$^+$06, CT05, Don06, TG07b, WR04, GR97, CRT06, CW08] where the goal is to reconstruct the sparse signal $\mathbf{x}[l]$ given the noisy compressed data $\mathbf{y}[l]$. Over the past decade this problem has received immense attention, and a large number of computationally efficient algorithms have been developed to provably recover the sparse signals $\mathbf{x}[l]$ from $M = O(K \log(N/K))$ compressive measurements [GR97, CRT06, CT07, TG07b, DM09, NT09].

In its most basic form, the measurement model (3.1) only consists of $L = 1$ measurement vector, which is also known as the Single Measurement Vector (SMV) problem. The SMV problem is known to have a unique K-sparse solution, with minimum ℓ_0 norm, if and only if $K \leq \text{k-rank}(\mathbf{A})/2$, where k-rank$(.)$ denotes the Kruskal rank [FR13]. When $L > 1$, (3.1) is known as the Multiple Measurement Vector (MMV) problem. It has been shown that by collecting $L > 1$ measurement vectors, one can recover signals with larger sparsity, compared to that for the SMV problem. In particular, the MMV problem is known to have a unique jointly sparse solution with sparsity K, if $K \leq (\text{k-rank}(\mathbf{A}) + \text{rank}(\mathbf{Y}))/2$ [DE12]. Several algorithms have been developed for the MMV model that tries to exploit the benefit offered by the rank of the measurements $\mathbf{Y}$ and recover larger sparse supports [CREKD05, CH05, CH06, TGS06, BCHJ14, REC04, LBJ12, ME08]. However, theoretical guarantees for these algorithms are limited to the regime $K < M$.

Of particular interest in many different applications is to detect the *location/indices* of the common nonzero elements of $\mathbf{x}[l]$, (i.e., the set $\mathcal{S}$), as they contain physically meaningful information. For instance, in a grid based model for Directions-of-Arrival (DOA) estimation

problem [MCW05, XHYB12, PV13], the location of nonzero elements determines the directions of the impinging narrow-band waves from point sources. In compressive spectrum sensing for cognitive radio, the sparse support represents the occupied frequency bands on which the communication needs to be avoided by the cognitive radio agents [TG07a]. In sparse linear regression [LS07, LW11, NW08], the support indicates the set of parameters that contribute to the data. In these cases, it is imperative to analyze the performance of algorithms that directly aim at recovering the support without first estimating the sparse signal vectors. This problem of *sparse support recovery* (or model selection in statistics community), has received a great deal of attention over the past few years and several computationally tractable algorithms have been proposed [Wai09b, FRG09, CP09]. Concurrently, a significant body of work is dedicated towards characterizing fundamental performance limits of sparse support recovery [Wai09a, JKR11, AAS17, SC17], often using information-theoretic tools. In the following subsection, we will review the key results on sparse support recovery.

The performance of sparse support recovery has been extensively analyzed by several researchers. These results mostly provide necessary and sufficient conditions under which support recovery will succeed. In the following, we will review some of these results:

Algorithm-Specific Guarantees and Upper Bounds on Probability of Error The techniques to derive upper bounds on the probability of error in support recovery are usually constructive, i.e., a specific support recovery algorithm is first proposed (sometimes computationally intractable), and then its performance is analyzed. For $L = 1$, (i.e., SMV model), Fletcher et al. [FRG09] considered a "maximum correlation" detector, which despite its simplicity, shows that $\Omega(K\log(N-K))$ measurements indeed suffice to ensure perfect support recovery. Exhaustive-search decoders, which are computationally intractable, are considered in [Wai09a, Rad11, ASZ10, TN10] which also indicate a similar sample complexity of $M = \Omega(K\log\frac{N}{K})$. A body of work also uses various tools from information theory such as joint

typicality decoder [AT10], rate-distortion theory [RG13, FRG07], belief propagation and list decoding [PDM09], capacity of multiple access communication channels [JKR11] to provide achievability and/or converse results on successful sparse support recovery. Recently, the authors in [AAS17, SC17] unify sparse support recovery problems by considering linear, nonlinear and probabilistic models, and provide optimal sample complexity in terms of necessary and sufficient number of measurements. However, a common feature of most of these results is that the sparse signal $\mathbf{x}[l]$ is modeled as a (unknown) deterministic quantity and statistical priors on $\mathbf{x}[l]$ (such as its correlation structure) are not fully exploited.

In contrast, Tang et al [TN10] considered an MMV model and assumed that the signals are generated by *uncorrelated* Gaussian sources with equal powers. They proposed an exhaustive search Maximum Likelihood (ML) decoder, which searches for all possible support sets and finds the one that maximizes the likelihood function. They derived upper bounds on the probability of error of such a detector, in terms of M, L, N and K. Although the (non zero) signals are assumed to be uncorrelated, the derivation of the upper bound does not fully exploit this structure. Hence, their results do not guarantee successful support detection when $K > M$. Recently, [PYL17] considered a similar problem setting, and using a typical set decoder, they ensured reliable support recovery as long as $K = O(M)$.

Necessary Conditions on Support Recovery In order to find lower bounds on probability of error, (or fundamental performance limits of sparse support recovery algorithms), a body of work [Wai09a, TN10, AAS17, SC17, JKR11, ASZ10, AT10] has used various versions of Fano's inequality, which provide information theoretic lower bounds for any support detection algorithm. These results indicate that for deterministic signal models (and for the SMV model with statistical priors on $\mathbf{x}[l]$), it is necessary to have $M = \Omega(K \log N)$ to ensure vanishing probability of error of any support recovery algorithm. Although Fano's inequality has proved to be useful in SMV model to derive tight bounds for necessary number of measurements for reliable support recovery,

these bounds may not be tight for MMV model with specific correlation structures. Specifically, the authors in [TN10] use Fano's inequality to derive lower bounds for the problem of support recovery in an MMV model where the non zero signals are statistically uncorrelated. However, as the authors point out, their lower bound on the probability of error is loose and can become negative for large L. In this chapter, we address this issue by considering a more fundamental condition called *covariance identifiability*, to establish performance limits for sparse support recovery in such MMV models.

Recovering Supports of Size $K > M$ All of the algorithms discussed so far guarantee successful recovery of supports of size $K = O(M)$. However, in MMV models, under certain deterministic and statistical assumptions on $\mathbf{x}[l]$, one can recover signals with sparsity K larger than M. This has been so far demonstrated in two distinct lines of work. The authors in [BKDM14] show that Sparse Bayesian Learning is in fact capable of recovering more sources than the number of measurements ($K > M$), in a scenario where the measurements are assumed to be noiseless, and the non zero rows of $\mathbf{X} = [\mathbf{x}[0], \mathbf{x}[1], \cdots, \mathbf{x}[L-1]]$ are orthogonal. They also provide empirical results that show the possibility of recovering supports of size $K > M$, when the measurements are contaminated with noise.

On the other hand, our prior work [PV15] has shown that when the non zero elements of $\mathbf{x}[l]$ are assumed to be statistically uncorrelated, one can recover supports of size $K = O(M^2)$, for cleverly designed measurement matrices. In an ideal setting, where the exact covariance matrix of the measurements $\mathbf{y}[l]$ is available (by allowing $L \to \infty$), we guaranteed that it is possible to recover supports of size $K = O(M^2)$ by solving a suitable l_1 minimization problem. For finite L, we proposed a new algorithm, known as Correlation-Aware LASSO (or Co-LASSO) which uses the correlation structure of the data to recover the supports of size $K > M$. However, no theoretical guarantees currently exist that can characterize the probability with which support recovery succeeds for finite L when $K > M$, and ensure that this probability goes to 1 as $L \to \infty$.

In this chapter, we bridge this gap and make the following contributions.

3.1 Correlation Aware Support Recovery

In this section, we provide new probabilistic guarantees for recovering the common support of jointly sparse vectors in Multiple Measurement Vector (MMV) models. In recent times, Bayesian approaches for sparse signal recovery (such as Sparse Bayesian Learning, Correlation-Aware LASSO) have shown preliminary evidence that under appropriate conditions (such as access to ideal covariance matrix of the measurements, or certain restrictive orthogonality condition on the signals), it is possible to recover supports of size (K) larger than the dimension (M) of each measurement vector. However, no results exist that characterize the probability with which this can be achieved for finite number of measurement vectors (L). this Section bridges this gap by formulating the support recovery problem in terms of a multiple hypothesis testing framework. Chernoff-type upper bounds on the probability of error are established, and new sufficient conditions are derived that guarantee its exponential decay with respect to L even when $K = O(M^2)$. Our sufficient conditions are based on the properties of the so-called Khatri-Rao product of the measurement matrix, and reveal the importance of sampler design. Negative results are also established indicating that when K exceeds a certain threshold (in terms of M), there will exist a class of measurement matrices for which any support recovery algorithm will fail. Using results from geometric probability, we characterize the probability with which a randomly generated measurement matrix will belong to this class and show that this probability tends to 1 asymptotically in the size (N) of the sparse vectors.

Contributions

1. Using the same signal model from [PV15, TN10] that assumes the non-zero elements of $\mathbf{x}[l]$ to be statistically uncorrelated, we consider a Maximum Likelihood based support detector. We show that when N, M, K are kept fixed, the probability of error of this detector goes to zero exponentially fast in L, as long as a fundamental condition called *covariance identifiability* is satisfied. Conversely, when covariance identifiability fails, the error of any

support detection algorithm will not converge to 0 even when $L \to \infty$.

2. We show that the condition $K < \text{k-rank}(\mathbf{A}^* \odot \mathbf{A})/2$, is sufficient for the covariance identifiability condition to be satisfied. Since k-rank$(\mathbf{A}^* \odot \mathbf{A})$ can be as large as $O(M^2)$ for suitably designed measurement matrices, our results suggest that one can recover supports of size $K = O(M^2)$ for such matrices.

3. Using results from geometric probability, we also show that when $K \geq M^2 + M + 2$, with high probability, there exists measurement matrices (referred to as ambiguous measurement matrices in this Section) for which the covariance identifiability condition will be violated. We exactly characterize this probability, and also provide simplified lower bounds, which tend to 1 as $N \to \infty$.

Notations

Throughout this section, matrices are represented by bold uppercase letters, vectors by bold lowercase letters. The italic letters like $\mathcal{S}$ denote index sets containing integers in increasing order. The symbol x_i denotes the ith entry of a vector $\mathbf{x}$. The notation $\mathbf{A}_{\mathcal{S}}$ (resp. $\mathbf{x}_{\mathcal{S}}$) represents the submatrix (resp. subvector) of $\mathbf{A}$ (resp. $\mathbf{x}$) whose columns (elements) are indexed by $\mathcal{S}$. The symbols $\odot$ and $\otimes$ represent the Khatri-Rao and Kronecker products, respectively. Moreover, $(.)^H$ and $(.)^*$ denote the Hermitian and complex conjugate operators. Other notations should be clear from the context. The notations $\mathbb{R}_+$ (resp. $\mathbb{R}_{++}$) denote the set of nonnegative (resp. positive) real numbers.

Some Basic Definitions:

In this section, we will review some basic definitions which will be useful throughout the paper:

Definition 5. ***Kruskal Rank*** *of a matrix* $\mathbf{X}$*, denoted by* k-rank($\mathbf{X}$)*, is the maximum integer k, such that any set of k columns of* $\mathbf{X}$ *is linearly independent [Kru77].*

Definition 6. ***Khatri-Rao*** *product of two matrices* $\mathbf{A} \in \mathbb{F}^{M_1 \times N}$*, and* $\mathbf{B} \in \mathbb{F}^{M_2 \times N}$ *(with the same number of columns) is defined as*

$$\mathbf{A} \odot \mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{b}_1, \mathbf{a}_2 \otimes \mathbf{b}_2, \cdots, \mathbf{a}_N \otimes \mathbf{b}_N]$$

where $\mathbf{a}_i$ *(resp.* $\mathbf{b}_i$*) indicate the ith column of the matrix* $\mathbf{A}$ *(resp.* $\mathbf{B}$*), and* $\otimes$ *indicates the Kronecker product.*

Example 1. *Let* $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2]$*, where* $\mathbf{a}_1 = [1,2]^T$*, and* $\mathbf{a}_2 = [3,4]^T$*. Moreover, let* $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2]$*, where* $\mathbf{b}_1 = [5,6]^T$*, and* $\mathbf{a}_2 = [7,8]^T$*. Then, we have*

$$\mathbf{A} \odot \mathbf{B} = \begin{bmatrix} 5 & 21 \\ 6 & 24 \\ 10 & 28 \\ 12 & 32 \end{bmatrix}$$

3.1.1 Problem Formulation

We adopt the signal model defined earlier in (3.1). We further make the following assumptions:

- (**A1**) The vectors $\mathbf{x}_{\mathcal{S}}[l] \in \mathbb{F}^K$, consisting of the nonzero elements of $\mathbf{x}[l]$, $0 \leq l \leq L-1$, are i.i.d. Gaussian random vectors distributed as

$$\mathbf{x}_{\mathcal{S}}[l] \sim \mathbb{F}\mathcal{N}(0, \mathbf{P})$$

where $\mathbf{P} = \text{diag}(\mathbf{p})$ is a diagonal matrix, and $\mathbf{p} \in \mathbb{R}_{++}^{K\times 1}$ is the vector of source powers comprised of positive entries. A diagonal covariance matrix implies that the nonzero elements of $\mathbf{x}_{\mathcal{S}}[l]$ are uncorrelated.

- (**A2**) The additive noise $\mathbf{w}[l] \in \mathbb{F}^M$ is white, and independent of $\mathbf{x}[l]$, distributed as

$$\mathbf{w}[l] \sim \mathbb{F}\mathcal{N}(0, \sigma^2\mathbf{I}).$$

  Here, σ^2 is the power of the noise, and assumed to be known *apriori*.

- (**A3**) We assume that $\mathbf{A} \in \mathbb{F}^{M\times N}$ is a fixed and known measurement matrix.

In assumptions (**A1**) (resp. (**A2**)), when $\mathbb{F} = \mathbb{C}$, we further assume that $\mathbf{x}_{\mathcal{S}}[l]$ (resp. $\mathbf{w}[l]$) follow *circularly symmetric* complex Gaussian distribution, with the specified mean and covariance matrices.

The assumption (**A1**) is widely used in Bayesian compressed sensing [JXC08], especially in sparse Bayesian learning (SBL)[WR04, PV14a], and correlation aware sparse estimation [PV13]. Stacking the columns of $\mathbf{y}[l]$ (resp. $\mathbf{x}[l], \mathbf{w}[l]$) into the matrix $\mathbf{Y}$ (resp. $\mathbf{X}$, $\mathbf{W}$), one can equivalently write (3.1) as

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{W}.$$

where $\mathbf{Y} \in \mathbb{F}^{M\times L}, \mathbf{X} \in \mathbb{F}^{N\times L}, \mathbf{W} \in \mathbb{F}^{M\times L}$. Under the assumptions (**A1**) and (**A2**), $\mathbf{Y}$ follows a matrix variate Gaussian distribution as

$$\mathbf{Y}|\mathcal{S} \sim \mathbb{F}\mathcal{N}(\mathbf{0}, \mathbf{R}_{\mathcal{S}} \otimes \mathbf{I}_L)$$

where $\mathbf{R}_{\mathcal{S}} = \mathbf{A}_{\mathcal{S}}\mathbf{P}\mathbf{A}_{\mathcal{S}}^H + \sigma^2\mathbf{I}$, with the probability density function (pdf) given by

$$f(\mathbf{Y}|\mathcal{S}) = \frac{1}{((\pi/\eta_{\mathbb{F}})^M|\mathbf{R}_{\mathcal{S}}|)^{\eta_{\mathbb{F}}L}} \exp[-\eta_{\mathbb{F}} \operatorname{tr}(\mathbf{Y}^H\mathbf{R}_{\mathcal{S}}^{-1}\mathbf{Y})] \tag{3.2}$$

where $\eta_{\mathbb{F}} = 1$ for $\mathbb{F} = \mathbb{C}$, and $\eta_{\mathbb{F}} = 1/2$ for $\mathbb{F} = \mathbb{R}$.

The goal of this Section is to study the problem of recovering $\mathcal{S}$ given $\mathbf{Y}$, under assumptions (**A1**), (**A2**), and (**A3**). Let $\psi(\mathbf{Y})$ be any detector which knows K, and chooses one of $Q := \binom{N}{K}$ possible support sets. For such a support detector the average probability of error in detecting the correct support can be written as

$$p_e = \sum_{i=1}^{Q} \mathbb{P}(\mathcal{S}_i)\mathbb{P}(\psi(\mathbf{Y}) \neq \mathcal{S}_i|\mathcal{S}_i),$$

where $\mathcal{S}_i$, $i = 1, \cdots, Q$, denotes the ith candidate support. For simplicity of notations, we denote $\mathbf{A}_{\mathcal{S}_i}$ by $\mathbf{A}_i$, where $i = 1, \cdots, Q$. Moreover, let

$$\mathbf{R}_i = \mathbf{A}_i\mathbf{P}\mathbf{A}_i^H + \sigma^2\mathbf{I}.$$

denote the covariance matrix correpsonding to the support $\mathcal{S}_i$.

Unlike previous works, [TN10, Wai09a, AAS17], which assume $K \leq M$, we consider the problem of support recovery in the regime $K > M$, and characterize p_e for this regime. Our goal is not to derive the tightest bounds on p_e, but to derive fundamental conditions under which $p_e \to 0$, exponentially fast with respect to L, and characterize the corresponding error exponent. Our results are based on a fundamental notion of *covariance identifiability*, and role of certain Khatri-Rao products of $\mathbf{A}$, as discussed in the next section.

3.1.2 Identifiability of Covariance Matrices & Role of Khatri-Rao Product

In this Section, we are interested in knowing if it is possible to identify supports of size $K = O(M^2)$ using the data model defined in Section 3.1.1. Given $\mathbf{A}$ and $\mathbf{p}$, the distribution $f(\mathbf{Y}|\mathcal{S})$ of the measurements is parameterized by the support $\mathcal{S}$. We are interested in the notion of covariance identifiability defined as follows:

Definition 7. ***Covariance Identifiability:*** *For a given* $\mathbf{P} = \text{diag}(\mathbf{p})$*, and a measurement matrix* $\mathbf{A}$*, we have covariance identifiability if the covariance matrices* $\mathbf{R}_i = \mathbf{A}_i\mathbf{P}\mathbf{A}_i^H + \sigma^2\mathbf{I}$*, and* $\mathbf{R}_j = \mathbf{A}_j\mathbf{P}\mathbf{A}_j^H + \sigma^2\mathbf{I}$ *satisfy*

$$\mathbf{R}_i = \mathbf{R}_j \Leftrightarrow \mathcal{S}_i = \mathcal{S}_j, \ \ \forall i, j \in \{1, \cdots, Q\}. \tag{3.3}$$

Remark 18. If (3.3) fails to hold for some $(\hat{i}, \hat{j})$, then the two densities $f(\mathbf{Y}|\mathcal{S}_{\hat{i}}) = f(\mathbf{Y}|\mathcal{S}_{\hat{j}})$ become identical, and as a simple argument will later show, any detector will fail to distinguish between $\mathcal{S}_{\hat{i}}$ and $\mathcal{S}_{\hat{j}}$.

It can be seen that covariance identifiablity is determined by properties of the measurement matrix $\mathbf{A}$. In this Section, we are interested in characterizing the set of measurement matrices $\mathbf{A}$, for which covariance identifiability is violated. To this end, we define the notion of *ambiguous* $\mathbf{A}$ as follows:

Definition 8. *A matrix* $\mathbf{A}$ *is said to be ambiguous (in the context of support recovery), if there exists a vector* $\mathbf{p} \in \mathbb{R}_{++}^K$*, and supports* $\mathcal{S}_i$*,* $\mathcal{S}_j$*,* $\mathcal{S}_i \neq \mathcal{S}_j$*, such that the covariance matrix* $\mathbf{R}_i = \mathbf{A}_i \,\text{diag}(\mathbf{p})\mathbf{A}_i^H + \sigma^2\mathbf{I}$ *is identical to* $\mathbf{R}_j = \mathbf{A}_j \,\text{diag}(\mathbf{p})\mathbf{A}_j^H + \sigma^2\mathbf{I}$*. We denote the set of all ambiguous* $\mathbf{A}$ *by*

$$I_{\mathbf{A}} = \{\mathbf{A} : \exists \mathbf{p} \in \mathbb{R}_{++}^K, i, j \big| \ \mathcal{S}_i \neq \mathcal{S}_j, \ s.t. \ \mathbf{R}_i = \mathbf{R}_j\} \tag{3.4}$$

It is clear that if the measurement matrix $\mathbf{A} \in I_{\mathbf{A}}$, there will exist a $\mathbf{p} \in \mathbb{R}_{++}^K$, such that the

covariance identifiability will fail, for this $(\mathbf{A}, \mathbf{p})$ pair. We now derive a necessary condition for a measurement matrix $\mathbf{A}$ to be ambiguous.

Lemma 4. *If* $K < \frac{\text{k-rank}(\mathbf{A}^* \odot \mathbf{A})}{2}$, *then* $\mathbf{A} \notin I_{\mathbf{A}}$.

Proof. We prove by contradiction. Suppose that $\mathbf{A} \in I_{\mathbf{A}}$. Then, there must exist $\mathbf{p}$, $\mathcal{S}_i$, and $\mathcal{S}_j$, with $\mathcal{S}_i \neq \mathcal{S}_j$ such that

$$\mathbf{A}_i \operatorname{diag}(\mathbf{p}) \mathbf{A}_i^H = \mathbf{A}_j \operatorname{diag}(\mathbf{p}) \mathbf{A}_j^H. \tag{3.5}$$

This can be rewritten in vectorized form as

$$\mathbf{B}_i \mathbf{p} = \mathbf{B}_j \mathbf{p}. \tag{3.6}$$

where $\mathbf{B} := \mathbf{A}^* \odot \mathbf{A} \in \mathbb{F}^{M^2 \times N}$, and $\mathbf{B}_i$ (resp. $\mathbf{B}_j$) indicate the submatrices of $\mathbf{B}$ whose columns are indexed by the set $\mathcal{S}_i$ (resp. $\mathcal{S}_j$), i.e., $\mathbf{B}_i = \mathbf{A}_i^* \odot \mathbf{A}_i \in \mathbb{F}^{M^2 \times K}$, and $\mathbf{B}_j = \mathbf{A}_j^* \odot \mathbf{A}_j \in \mathbb{F}^{M^2 \times K}$. Moreover, let $\mathbf{B}_{i \cap j}$, $\mathbf{B}_{i \setminus j}$, $\mathbf{B}_{j \setminus i}$, be the submatrices of $\mathbf{B}$ whose columns are indexed by the sets $\mathcal{S}_i \cap \mathcal{S}_j, \mathcal{S}_i \setminus \mathcal{S}_j, \mathcal{S}_j \setminus \mathcal{S}_i$. Similarly, define $\mathbf{p}_{i \cap j}$, $\mathbf{p}_{i \setminus j}$, $\mathbf{p}_{j \setminus i}$. One can rewrite (3.6) as

$$[\mathbf{B}_{i \setminus j} | \mathbf{B}_{i \cap j}] \begin{bmatrix} \mathbf{p}_{i \setminus j} \\ \mathbf{p}_{i \cap j} \end{bmatrix} = [\mathbf{B}_{j \setminus i} | \mathbf{B}_{i \cap j}] \begin{bmatrix} \mathbf{p}'_{j \setminus i} \\ \mathbf{p}'_{i \cap j} \end{bmatrix} \tag{3.7}$$

where

$$\begin{bmatrix} \mathbf{p}'_{j \setminus i} \\ \mathbf{p}'_{i \cap j} \end{bmatrix} = \Pi \begin{bmatrix} \mathbf{p}_{i \setminus j} \\ \mathbf{p}_{i \cap j} \end{bmatrix}$$

for some permutation matrix $\Pi \in \mathbb{R}^{K\times K}$. From (3.7), we get

$$\underbrace{[\mathbf{B}_{i\cap j}|\mathbf{B}_{i\setminus j}|\mathbf{B}_{j\setminus i}]}_{\mathbf{B}_{i\cup j}} \underbrace{\begin{bmatrix} \mathbf{p}_{i\cap j}-\mathbf{p}'_{i\cap j} \\ \mathbf{p}_{i\setminus j} \\ -\mathbf{p}'_{j\setminus i} \end{bmatrix}}_{\tilde{\mathbf{p}}} = \mathbf{0} \tag{3.8}$$

Since $\mathcal{S}_i \neq \mathcal{S}_j$, $\tilde{\mathbf{p}}$ cannot be a zero vector. This implies that $\mathbf{B}_{i\cup j}$ is column rank deficient. Since the number of columns of $\mathbf{B}_{i\cup j}$ is given by $|\mathcal{S}_i \cup \mathcal{S}_j|$, and $|\mathcal{S}_i \cup \mathcal{S}_j| \leq 2K$, we have

$$\text{rank}(\mathbf{B}_{i\cup j}) \leq \min(M^2, 2K) = 2K \tag{3.9}$$

The last inequality in (3.9) is true because we have already assumed that $K < \frac{\text{k-rank}(\mathbf{A}^*\odot\mathbf{A})}{2}$, and since k-rank$(\mathbf{A}^* \odot \mathbf{A}) \leq$ rank$(\mathbf{A}^* \odot \mathbf{A}) \leq M^2$, it follows that $K < M^2/2$. Inequality (3.9) contradicts the assumption that $K < \frac{\text{k-rank}(\mathbf{B})}{2}$, hence $\mathbf{A} \notin I_{\mathbf{A}}$. □

3.1.3 Recovering Support of Size $K = O(M^2)$ Using Multiple Hypothesis Testing Framework

In this section, we will consider the same maximum likelihood decoder as [TN10], and analyze its probability of error. Although [TN10] uses the same assumption (**A1**), it ensures $p_e \to 0$ (as $L \to \infty$), only in the regime $K \leq M$. To the best of our knowledge, no existing result ensures exponentially vanishing probability of error (w.r.t. L), when $K > M$. We now bridge this gap by explicitly showing the role of the Khatri-Rao product of $\mathbf{A}$. We prove that it is possible to detect supports of size $K = O(M^2)$ with a probability of error that decays to zero exponentially fast w.r.t. L. Similar to [TN10], for the remainder of this section, we will assume that all signals have equal powers, i.e., $\mathbf{p} = p\mathbf{1}_K$, where $\mathbf{1}_K \in \mathbb{R}^K$ is the all-ones vector. Furthermore, we assume that the detector knows the ratio of source power to noise power (SNR) p/σ^2.

Support Detection via Multiple Hypothesis Testing

In this section, we cast the joint support recovery problem under the multiple hypothesis testing framework. We first review the binary hypothesis testing framework to illustrate how supports of size $K = O(M^2)$ can be identified, and then extend this result to the multiple hypothesis testing framework.

Binary Hypothesis Tesing In a binary hypothesis testing problem, one aims to decide which of two candidate distributions generated the data $\mathbf{Y}$. In a binary support recovery problem, these candidate distributions are characterized by two possible supports $\mathcal{S}_0$ and $\mathcal{S}_1$. In this case, we can write the two hypotheses as

$$\begin{cases} H_0 : \mathbf{Y} \sim f(\mathbf{Y}|\mathcal{S}_0) \\ H_1 : \mathbf{Y} \sim f(\mathbf{Y}|\mathcal{S}_1) \end{cases}$$

where $\mathbf{Y}|\mathcal{S}_i$ has a matrix variate Gaussian distribution (from assumption (**A1**)):

$$\mathbf{Y}|\mathcal{S}_i \sim \mathbb{F}\mathcal{N}(0, \mathbf{R}_i \otimes \mathbf{I}_L),$$

where $i = 0, 1$. Let $\psi(\mathbf{Y})$ be the output of a detector, which either returns $\mathcal{S}_0$ or $\mathcal{S}_1$. The corresponding probability of error is given by

$$p_e = \mathbb{P}(\mathcal{S}_0)P_{10} + \mathbb{P}(\mathcal{S}_1)P_{01} \tag{3.10}$$

where $P_{10} = \mathbb{P}(\psi(\mathbf{Y}) = \mathcal{S}_1|\mathcal{S}_0)$, and $P_{01} = \mathbb{P}(\psi(\mathbf{Y}) = \mathcal{S}_0|\mathcal{S}_1)$. We further assume that $\mathbb{P}(\mathcal{S}_0) = \mathbb{P}(\mathcal{S}_1) = \frac{1}{2}$.

Multiple Hypothesis Testing In this case, the detector chooses one out of $Q = \binom{N}{K}$ possible supports, which can be written as the following multiple hypothesis testing problem:

$$\begin{cases} H_1 : \mathbf{Y} \sim \mathbb{F}\mathcal{N}(0, \mathbf{R}_1 \otimes \mathbf{I}_L) \\ H_2 : \mathbf{Y} \sim \mathbb{F}\mathcal{N}(0, \mathbf{R}_2 \otimes \mathbf{I}_L) \\ \quad \vdots \\ H_Q : \mathbf{Y} \sim \mathbb{F}\mathcal{N}(0, \mathbf{R}_Q \otimes \mathbf{I}_L) \end{cases} \tag{3.11}$$

Here $\mathbf{R}_q = p\mathbf{A}_q\mathbf{A}_q^H + \sigma^2\mathbf{I}$, for $q = 1, \cdots, Q$. Assuming equal probabilities on all the hypotheses, i.e., $\mathbb{P}(\mathcal{S}_i) = \frac{1}{Q}$ for all $i = 1, \cdots, Q$, we consider the following maximum likelihood detector

$$\hat{\psi}(\mathbf{Y}) = \arg \max_{1 \le i \le Q} \mathbb{P}(\mathbf{Y}|\mathcal{S}_i) \tag{3.12}$$

The conditional probability of error for such a detector can be written as

$$p_{e|H_i} = \mathbb{P}(\hat{\psi}(\mathbf{Y}) \neq \mathcal{S}_i|\mathcal{S}_i) \tag{3.13}$$

and the average probability of error equals

$$p_e = \frac{1}{Q}\sum_{i=1}^{Q} p_{e|\mathcal{S}_i} \tag{3.14}$$

Review of Tang's Results [TN10]

In [TN10], Tang et al. analyzed the above multiple hypothesis testing framework and produced the following upper bound

Theorem 18. *[TN10, Theorem 1] If $M \ge 2K$, and $\bar{\lambda} > 4[K(N-K)]^{\frac{1}{\eta_{\mathbb{F}}L}}$, then the probability of*

error p_e, for the detector given by (3.12)*, is bounded by*

$$p_e \leq \frac{1}{2} \frac{\frac{K(N-K)}{(\bar{\lambda}/4)^{\eta_{\mathbb{F}} L}}}{1 - \frac{K(N-K)}{(\bar{\lambda}/4)^{\eta_{\mathbb{F}} L}}} \tag{3.15}$$

where $\bar{\lambda} = \min_{i \neq j} \bar{\lambda}_{ij}$, $\bar{\lambda}_{ij}$ is the geometric mean of the eigenvalues of $\mathbf{R}_i^{1/2} \mathbf{R}_j^{-1} \mathbf{R}_i^{1/2}$ which are greater than one, $\eta_{\mathbb{C}} = 1$, and $\eta_{\mathbb{R}} = 1/2$.

Since, for large L, the upper bound (3.15) is proportional to $e^{-L \log(\bar{\lambda}/4)}$, this result shows that the probability of error decays exponentially fast as L goes to infinity, provided $K \leq \frac{M}{2}$ and $\bar{\lambda}$ is larger than a given threshold [TN10].

Although this result guarantees exponential decay for p_e, it is limited only to the regime $K \leq M$, and does not consider the case $K > M$. Moreover, according to this result, in order to ensure exponential decay for p_e, one requires $\bar{\lambda}$ to be greater than some threshold which depends on L. However, as we will show in Section 3.1.3, it is possible to have exponentially decaying p_e even for $K = O(M^2)$, and our sufficient conditions only depend on the structure of the measurement matrix.

New Upper Bounds on Probability of Error (valid for $K > M$)

We now derive new upper bounds for probability of error, which are valid even if $K > M$.

Binary Hypothesis Testing for Support Recovery Let us first consider the binary hypothesis testing framework, introduced in Section 3.1.3. Similar to [TN10], we consider the following detector $\hat{\psi}(\mathbf{Y})$:

$$\hat{\psi}(\mathbf{Y}) = \begin{cases} \mathcal{S}_0, & \ell(\mathbf{Y}) < 0 \\ \mathcal{S}_1, & \ell(\mathbf{Y}) \geq 0 \end{cases}$$

where $\ell(\mathbf{Y}) = \log \frac{f(\mathbf{Y}|\mathcal{S}_1)}{f(\mathbf{Y}|\mathcal{S}_0)}$. One way to find an upper bound on the probability of error for $\hat{\psi}(\mathbf{Y})$ is to use Chernoff bound on the test statistic $\ell(\mathbf{Y})$. Let $\mu(s)$ be the logarithm of the moment generating function of $\ell(\mathbf{Y})$ given by

$$\begin{aligned}\mu(s) &= \log \int_{-\infty}^{\infty} e^{s\ell(\mathbf{Y})} f(\mathbf{Y}|\mathcal{S}_0) d\mathbf{Y} \\ &= \log \int_{-\infty}^{\infty} [f(\mathbf{Y}|\mathcal{S}_1)]^s [f(\mathbf{Y}|\mathcal{S}_0)]^{1-s} d\mathbf{Y}\end{aligned} \tag{3.16}$$

Then, one can write the Chernoff bound for p_e as

$$p_e \leq \frac{1}{2} e^{\mu(\hat{s})} \leq \frac{1}{2} e^{\mu(s)} \tag{3.17}$$

where $\hat{s} = \arg\min_{0 \leq s \leq 1} \mu(s)$, and $0 \leq s \leq 1$. Plugging the distribution functions of $f(\mathbf{Y}|H_0)$, and $f(\mathbf{Y}|H_1)$ into (3.16), one can get [TN10]

$$\mu(s) = -\eta_{\mathbb{F}} L \log |s\mathbf{H}^{1-s} + (1-s)\mathbf{H}^{-s}|$$

where $\mathbf{H} = \mathbf{R}_0^{1/2} \mathbf{R}_1^{-1} \mathbf{R}_0^{1/2}$.

Although finding the tightest bound (i.e., optimum $\hat{s}$) leads to an intractable expression, one can still get an upper bound on the probability of error by letting $s = \frac{1}{2}$ [TN10]. We define:

$$\gamma_{01} := -\frac{\mu(1/2)}{\eta_{\mathbb{F}} L} = \log |\frac{1}{2}\mathbf{H}^{\frac{1}{2}} + \frac{1}{2}\mathbf{H}^{-\frac{1}{2}}|. \tag{3.18}$$

We can further simplify (3.18) as

$$\gamma_{01} = \sum_{i=1}^{M} \log(\frac{\sqrt{\lambda_i} + 1/\sqrt{\lambda_i}}{2}) \tag{3.19}$$

where λ_i is the ith eigenvalue of $\mathbf{H}$. Therefore, it follows that

$$p_e \leq \frac{1}{2} e^{-\eta_{\mathbb{F}} \gamma_{01} L} \tag{3.20}$$

From (3.18) it can be seen that γ_{01} is independent of L. Therefore, if $\gamma_{01} > 0$, then (3.20) implies that p_e decays to zero exponentially fast in L. We now state two important lemmas which specify conditions under which $\gamma_{01} > 0$.

Lemma 5. *$\gamma_{01} > 0$ if and only if there exists some $i \in \{1, \cdots, M\}$ for which $\lambda_i \neq 1$.*

Proof. For every positive x, we have

$$x + \frac{1}{x} \geq 2,$$

and equality holds if and only if $x = 1$. Hence, in equation (3.19) we observe that $\log(\frac{\sqrt{\lambda_i}+1/\sqrt{\lambda_i}}{2}) \geq 0$ for every $i \in \{1, \cdots, M\}$. Moreover, $\gamma_{01} = 0$ if and only if $\log(\frac{\sqrt{\lambda_i}+1/\sqrt{\lambda_i}}{2}) = 0$ for every i. This is true if and only if $\lambda_i = 1$ for every $i \in \{1, \cdots, M\}$. Equivalently, $\gamma_{01} > 0$ if and only if there is at least one $i \in \{1, \cdots, M\}$ for which $\lambda_i \neq 1$. □

Lemma 6. *$\gamma_{01} = 0$ if and only if $\mathbf{R}_0 = \mathbf{R}_1$.*

Proof. If $\mathbf{R}_0 = \mathbf{R}_1$, $\mathbf{H} = \mathbf{I}$, implying $\gamma_{01} = 0$ from (3.18). Conversely, if $\gamma_{01} = 0$, from Lemma 5 we must have $\lambda_i = 1$ for all $i \in \{1, \cdots, M\}$. This implies

$$\mathbf{R}_0^{1/2} \mathbf{R}_1^{-1} \mathbf{R}_0^{1/2} = \mathbf{I}$$

which is true if and only if $\mathbf{R}_0 = \mathbf{R}_1$. □

Multiple Hypothesis Testing In this case, the conditional probability of error is given by (3.13). Using the Chernoff bound and the union bound, we get

$$p_{e|H_j} \leq \frac{1}{2} \sum_{\substack{i=1 \\ i \neq j}}^{Q} e^{\mu(s;\mathcal{S}_i,\mathcal{S}_j)} \tag{3.21}$$

where $\mu(s;\mathcal{S}_i,\mathcal{S}_j)$ is the moment generating function of the log-likelihood function $\ell_{ij}(\mathbf{Y}) = \log \frac{f(\mathbf{Y}|\mathcal{S}_i)}{f(\mathbf{Y}|\mathcal{S}_j)}$ under the hypothesis that $\mathcal{S}_j$ is the true support. Let

$$\gamma_{ij} = \sum_{k=1}^{M} \log\left(\frac{\sqrt{\lambda_k^{(i,j)}} + \frac{1}{\sqrt{\lambda_k^{(i,j)}}}}{2}\right)$$

where $\lambda_k^{(i,j)}$ is the kth eigenvalue of $\mathbf{R}_i^{1/2}\mathbf{R}_j^{-1}\mathbf{R}_i^{1/2}$. Comparing with equations (3.18) and (3.19), it can be seen that $\mu(\frac{1}{2};\mathcal{S}_i,\mathcal{S}_j) = -\eta_{\mathbb{F}} L \gamma_{ij}$. Therefore, we can rewrite (3.21) as

$$p_{e|H_j} \leq \frac{1}{2} \sum_{\substack{i=1 \\ j \neq i}}^{Q} e^{-\eta_{\mathbb{F}} \gamma_{ij} L} \leq \frac{Q}{2} e^{-\eta_{\mathbb{F}} \gamma_j L}$$

where $\gamma_j = \min_{i \neq j} \gamma_{ij}$. Hence, the average probability of error, defined in (3.14), can be bounded as

$$p_e \leq \frac{Q}{2} e^{-\eta_{\mathbb{F}} \gamma L}$$

where

$$\gamma = \min_j \gamma_j. \tag{3.22}$$

Using Stirling's approximation [Cor09]

$$\binom{N}{K} \leq \left(\frac{eN}{K}\right)^K,$$

we have

$$p_{e|H_j} \leq \frac{1}{2} e^{K \log \frac{eN}{K} - \gamma_j \eta_{\mathbb{F}} L}, \tag{3.23}$$

and

$$p_e \leq \frac{1}{2} e^{K \log \frac{eN}{K} - \gamma \eta_{\mathbb{F}} L}. \tag{3.24}$$

Therefore, when K and N are held constant, p_e converges to zero as $L \to \infty$, provided $\gamma > 0$. The following lemma connects the condition $\gamma > 0$ to covariance identifiability, defined in Section 3.1.2.

Lemma 7. *For given* $\mathbf{A}$ *and* $\mathbf{p}$*, the covariance identifiability condition, given by* (3.3)*, is equivalent to the condition* $\gamma > 0$*.*

Proof. Having $\gamma > 0$ is equivalent to

$$\gamma_{ij} > 0 \quad \forall i \neq j,\ i, j \in \{1, \cdots, Q\}$$

By the application of Lemma 6, it can be readily seen that $\gamma_{ij} > 0$ for all $i \neq j$, only if $\mathbf{R}_i \neq \mathbf{R}_j$. Therefore, the condition $\gamma > 0$ implies the *covariance identifiability* condition (3.3). Conversely, if covariance identifiability holds, there does not exist any $i \neq j$, such that $\mathbf{R}_i = \mathbf{R}_j$. Using Lemma 6, we conclude that $\gamma_{ij} > 0$ for all $i \neq j$, implying $\gamma_j > 0$, and hence $\gamma > 0$. □

Remark 19. *The decay rate* γ *given in* (3.24) *depends on the measurement matrix* $\mathbf{A}$ *(in particular, on M and N), the sparsity K, as well as the signal to noise ratio* p/σ^2*. In Theorem 19, we will derive explicit conditions on* $\mathbf{A}$ *which ensure* $\gamma > 0$*, thereby leading to a vanishing probability of error with respect to L. In general, it is non-trivial to obtain a simplified and tighter lower bound*

on γ which depends on more relatable properties of $\mathbf{A}$ such as its Restricted Isometry Property (RIP). The authors in [TN10] partially simplified the expression for γ only in the regime $K < M$, and for $K = 1$, they were able to relate the decay rate γ to the RIP of $\mathbf{A}$. In the regime $K > M$ (which is of interest in this Section), establishing such a result is an open problem for future research.

Sufficient Conditions for Successful Support Recovery

In this section, we show that the covariance identifiability condition can be related to the Kruskal rank of the Khatri-Rao product ($\mathbf{A}^* \odot \mathbf{A} \in \mathbb{F}^{M^2 \times N}$) of the measurement matrix, using which we establish sufficient conditions that guarantee exponential decay of p_e with respect to L even when $K > M$. Our main result is given by the following theorem:

Theorem 19. *Consider the maximum likelihood detector* (3.12) *for the joint support recovery problem* (3.11). *For any $\delta > 0$, the probability of error (p_e) of this detector satisfies $p_e \leq \delta$ provided*

$$K < \frac{\text{k-rank}(\mathbf{A}^* \odot \mathbf{A})}{2}$$

and

$$L \geq \frac{1}{\eta_{\mathbb{F}} \gamma} (\log \frac{1}{2\delta} + K \log \frac{eN}{K}). \tag{3.25}$$

Proof. The result follows from the application of Lemma 4, Lemma 7, and the upper bound (3.24). □

Remark 20. An immediate consequence of Theorem 19 is that if K, N and M are held constant, while maintaining $K < \frac{\text{k-rank}(\mathbf{A}^* \odot \mathbf{A})}{2}$, the probability of error decays to zero exponentially fast in L.

Remark 21. *Notice that the parameter γ depends on the dimensions M, N and the sparsity K. If we also allow M, N and K to scale, the scaling of L should be such that it satisfies the inequality* (3.25). *Obtaining an explicit expression for how L should scale as a function of M,N and K is non trivial, since it is not easy to analytically characterize γ as a function of M,N, and K. However, in Section 3.1.5, we will study the behavior of γ with respect to M,N, and K by conducting numerical simulations.*

It is shown in [PV12a, PV13] that the Kruskal rank of $\mathbf{B} = \mathbf{A}^* \odot \mathbf{A} \in \mathbb{F}^{M^2 \times N}$ can be as large as $O(M^2)$ for some cleverly designed measurement matrices $\mathbf{A}$ (such as nested [PV10], coprime [PV11] and minimum redundancy array manifolds). For such matrices Theorem 19 ensures that supports of size $K = O(M^2)$ can be recovered with exponentially vanishing probability of error.

3.1.4 Characterization of Ambiguous Measurement Matrices and Failure of Support Recovery

In this section, we will characterize the class $I_{\mathbf{A}}$ of ambiguous measurement matrices (defined in (3.4)), and understand their implications on support recovery. We will first show that if covariance identifiability is violated, then the probability of error (of any detector) is bounded below by a positive quantity that does not go to zero as $L \to \infty$. From our discussion in Section 3.1.2 we know that covariance identifiability is violated if $\mathbf{A} \in I_{\mathbf{A}}$. We will characterize the set $I_{\mathbf{A}}$ by determining the probability with which a randomly generated $\mathbf{A}$ belongs to $I_{\mathbf{A}}$, and show that this probability goes to one, as $N \to \infty$, in the regime $2N \geq K \geq M^2 + M + 2$.

Covariance Nonidentifiability and Failure of Support Recovery

The following lemma shows that if the identifiability condition (3.3) is violated, the average probability of error will never go to zero, even when $L \to \infty$. We state this result as the following lemma.

Lemma 8. *Given* $\mathbf{A} \in \mathbb{F}^{M\times N}$ *and* $\mathbf{p} \in \mathbb{R}^K$*, suppose there exist two supports* $S_{\hat{i}}, S_{\hat{j}}$ *(*$\hat{i} \neq \hat{j}$*) such that the covariance matrices* $\mathbf{R}_{\hat{i}} = \mathbf{A}_{\hat{i}}\text{diag}(\mathbf{p})\mathbf{A}_{\hat{i}}^H + \sigma^2\mathbf{I}$ *and* $\mathbf{R}_{\hat{j}} = \mathbf{A}_{\hat{j}}\text{diag}(\mathbf{p})\mathbf{A}_{\hat{j}}^H + \sigma^2\mathbf{I}$ *become identical. Then the average probability of error of any support detector is lower bounded as*

$$p_e \geq \frac{1}{Q}.$$

Proof. We will prove this result for a genie-aided support detector $\hat{\mathcal{S}}^g$ which has the extra knowledge of the true source powers $\mathbf{p}$. The detector $\hat{\mathcal{S}}^g$ clearly cannot perform worse than any other detector $\hat{\mathcal{S}}$ that does not know $\mathbf{p}$ (since, $\hat{\mathcal{S}}^g$ can choose to ignore this extra information about $\mathbf{p}$). Hence, the probability of error p_e of any detector $\hat{\mathcal{S}}$ satisfies

$$p_e \geq p_e^g,$$

where p_e^g denotes the probability of error for the genie-aided detector. Now, if there exists a pair $(\mathcal{S}_{\hat{i}}, \mathcal{S}_{\hat{j}})$ for which $\mathbf{R}_{\hat{i}} = \mathbf{R}_{\hat{j}}$, then we have

$$f(\mathbf{Y}|\mathcal{S}_{\hat{i}}) = f(\mathbf{Y}|\mathcal{S}_{\hat{j}}) \qquad \forall \mathbf{Y} \in \mathbb{F}^{M\times L}. \tag{3.26}$$

For the genie-aided detector $\hat{\mathcal{S}}^g$, the probability of error is given by

$$\begin{aligned} p_e^g &= \sum_{i=1}^{Q} \mathbb{P}(\hat{\mathcal{S}}^g \neq \mathcal{S}_i|\mathcal{S}_i)\mathbb{P}(\mathcal{S}_i) \\ &= \frac{1}{Q}\sum_{i=1}^{Q} \mathbb{P}(\hat{\mathcal{S}}^g \neq \mathcal{S}_i|\mathcal{S}_i) \\ &= \frac{1}{Q}\sum_{\substack{i=1 \\ i\neq\hat{i},\hat{j}}}^{Q} \mathbb{P}(\hat{\mathcal{S}}^g \neq \mathcal{S}_i|\mathcal{S}_i) + \frac{1}{Q}p_e^{(\hat{i},\hat{j})} \end{aligned}$$

where

$$p_e^{(\hat{i},\hat{j})} = \mathbb{P}(\hat{\mathcal{S}}^g \neq \mathcal{S}_{\hat{i}} | \mathcal{S}_{\hat{i}}) + \mathbb{P}(\hat{\mathcal{S}}^g \neq \mathcal{S}_{\hat{j}} | \mathcal{S}_{\hat{j}}). \tag{3.27}$$

Any genie-aided detector is typically a function of the measurements $\mathbf{Y}$ with the extra provision that in case of a tie between two candidate supports, it breaks ties randomly. The most general form of such a detector can be written as a function of $\mathbf{Y}$ and another random variable θ (possibly dependent on $\mathbf{Y}$) [2] as $\hat{\mathcal{S}}^g = \psi^g(\mathbf{Y};\theta)$. Accounting for the error incurred by such random tie-breaks, the probability $\mathbb{P}(\hat{\mathcal{S}}^g \neq \mathcal{S}_{\hat{j}} | \mathcal{S}_{\hat{j}})$ can be written as

$$\begin{aligned}
\mathbb{P}(\hat{\mathcal{S}}^g \neq \mathcal{S}_{\hat{j}} | \mathcal{S}_{\hat{j}}) &= \mathbb{E}_{\theta,\mathbf{Y}|\mathcal{S}_{\hat{j}}} \left[\mathbb{1}(\psi^g(\mathbf{Y};\theta) \neq \mathcal{S}_{\hat{j}}) \right] \\
&\overset{(a)}{=} \mathbb{E}_{\mathbf{Y}|\mathcal{S}_{\hat{j}}} \left\{ \mathbb{E}_{\theta|\mathbf{Y},\mathcal{S}_{\hat{j}}} \left[\mathbb{1}(\psi^g(\mathbf{Y};\theta) \neq \mathcal{S}_{\hat{j}}) \right] \right\} \\
&\overset{(b)}{=} \mathbb{E}_{\mathbf{Y}|\mathcal{S}_{\hat{j}}} \left\{ \mathbb{E}_{\theta|\mathbf{Y}} \left[\mathbb{1}(\psi^g(\mathbf{Y};\theta) \neq \mathcal{S}_{\hat{j}}) \right] \right\} \\
&\overset{(c)}{=} \mathbb{E}_{\mathbf{Y}|\mathcal{S}_{\hat{i}}} \left\{ \mathbb{E}_{\theta|\mathbf{Y}} \left[\mathbb{1}(\psi^g(\mathbf{Y};\theta) \neq \mathcal{S}_{\hat{j}}) \right] \right\} \\
&= \mathbb{P}(\hat{\mathcal{S}}^g \neq \mathcal{S}_{\hat{j}} | \mathcal{S}_{\hat{i}})
\end{aligned} \tag{3.28}$$

where $\mathbb{1}[.]$ is the indicator function, (a) follows from law of iterated expectations, (b) follows from the fact that given $\mathbf{Y}$, θ is independent of $\mathcal{S}_{\hat{i}}$, i.e the choice of randomization for breaking ties solely depends on the data $\mathbf{Y}$, and (c) follows from the fact that the conditional distributions

[2] As a concrete example, if $\hat{\mathcal{S}}^g$ is a genie-aided maximum-likelihood detector, then depending on $\mathbf{Y}$, the log-likelihood function $l(\mathbf{Y};\mathcal{S})$ can have one or multiple global optima (say, q optima, where $1 < q \leq Q = \binom{N}{K}$). When the latter happens and the detector is such that it breaks ties evenly, θ will be a discrete uniform random variable taking up values $1, \cdots, q$, each with equal probability of $1/q$, and its probability density function (for such a $\mathbf{Y}$) will be given by $g(\theta|\mathbf{Y}) = \frac{1}{q}\sum_{i=1}^{q} \delta(\theta - i)$, where $\delta(.)$ is the Dirac-delta function. When there is a unique global optimum, there are no ties to be broken, and θ will be a deterministic quantity (i.e., its density will consist of a single Dirac delta centered around the unique optimum).

of $\mathbf{Y}$, given $\mathcal{S}_{\hat{i}}$ and $\mathcal{S}_{\hat{j}}$, are identical (according to (3.26)). Therefore, from (3.27) we have

$$\begin{aligned} p_e^{(\hat{i},\hat{j})} &= \mathbb{P}(\hat{\mathcal{S}}^g \neq \mathcal{S}_{\hat{i}} | \mathcal{S}_{\hat{i}}) + \mathbb{P}(\hat{\mathcal{S}}^g \neq \mathcal{S}_{\hat{j}} | \mathcal{S}_{\hat{j}}) \\ &\overset{(d)}{=} \mathbb{P}(\hat{\mathcal{S}}^g \neq \mathcal{S}_{\hat{i}} | \mathcal{S}_{\hat{i}}) + \mathbb{P}(\hat{\mathcal{S}}^g \neq \mathcal{S}_{\hat{j}} | \mathcal{S}_{\hat{i}}) \\ &\geq \mathbb{P}(\{\hat{\mathcal{S}}^g \neq \mathcal{S}_{\hat{i}}\} \cup \{\hat{\mathcal{S}}^g \neq \mathcal{S}_{\hat{j}}\} | \mathcal{S}_{\hat{i}}) \\ &= 1 - \mathbb{P}(\{\hat{\mathcal{S}}^g = \mathcal{S}_{\hat{i}}\} \cap \{\hat{\mathcal{S}}^g = \mathcal{S}_{\hat{j}}\} | \mathcal{S}_{\hat{i}}) \end{aligned}$$

Here, (d) follows from (3.28). However, after incorporating random tie-breaks, the output of the detector $\hat{\mathcal{S}}^g$ cannot be simultaneously equal to two different supports $\mathcal{S}_{\hat{i}}$ and $\mathcal{S}_{\hat{j}}$, implying $\mathbb{P}(\{\hat{\mathcal{S}}^g = \mathcal{S}_{\hat{i}}\} \cap \{\hat{\mathcal{S}}^g = \mathcal{S}_{\hat{j}}\} | \mathcal{S}_{\hat{i}}) = 0$. Therefore, we conclude that that $p_e^{(\hat{i},\hat{j})} \geq 1$. Hence, $p_e \geq p_e^g \geq \frac{1}{Q}$. □

Hence, if covariance identifiability is violated, no detector (even genie-aided) can successfully recover the true support even when $L \to \infty$.

Characterization of Ambiguous Measurement Matrices

From Section 3.1.2, we know that if $\mathbf{A} \in I_{\mathbf{A}}$, then there exists $\mathbf{p}$ for which the covariance identifiability is violated, for that $\mathbf{A}$ and $\mathbf{p}$. In this subsection, we will focus on characterizing the set $I_{\mathbf{A}}$.

Lemma 9. *$\mathbf{A} \in I_{\mathbf{A}}$ if and only if there exists $\mathcal{S}_i, \mathcal{S}_j$ such that the null space $N(\mathbf{A}_i^* \odot \mathbf{A}_i - \mathbf{A}_j^* \odot \mathbf{A}_j)$ contains a positive vector $\mathbf{p}$.*

Proof. The proof follows from the definition of $I_{\mathbf{A}}$ in (3.4) and the fact that $\text{vec}(\mathbf{R}_i) = (\mathbf{A}_i^* \odot \mathbf{A}_i)\mathbf{p} + \sigma^2 \text{vec}(\mathbf{I})$. □

Lemma 9 dictates that in order to characterize $I_{\mathbf{A}}$, we need to understand when does the null space of $\mathbf{A}_i^* \odot \mathbf{A}_i - \mathbf{A}_j^* \odot \mathbf{A}_j \in \mathbb{F}^{M^2 \times N}$ contain a positive vector. This question has a direct connection to the following elegant result in geometric probability originally proved by Wendel in 1962 [Wen62].

Theorem 20. *[Wen62] Suppose $\mathbf{x}_1, \cdots, \mathbf{x}_n$ are i.i.d random points in $\mathbb{R}^d$ such that their distribution is symmetric about origin, and with probability one all subsets of d points are linearly independent. Then*

$$\mathbb{P}(\mathbf{0} \notin conv\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}) = \frac{1}{2^{n-1}} \sum_{k=0}^{d-1} \binom{n-1}{k},$$

where $conv\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ denotes the convex hull generated by the points $\mathbf{x}_1, \cdots, \mathbf{x}_n$.

Lemma 10. *The condition $\mathbf{0} \in conv\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ is equivalent to the following:*

$$\exists \mathbf{p} \in \mathbb{R}^n_+, \mathbf{p} \neq \mathbf{0}, \text{ such that } \mathbf{p} \in N([\mathbf{x}_1|\mathbf{x}_2|\cdots|\mathbf{x}_n]).$$

Proof. If $\mathbf{0} \in \text{conv}\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$, then

$$\exists \tilde{\mathbf{p}} \in \mathbb{R}^n_+ : \sum_{i=1}^n \tilde{p}_i \mathbf{x}_i = \mathbf{0}, \sum_{i=1}^n \tilde{p}_i = 1.$$

We clearly have $[\mathbf{x}_1|\mathbf{x}_2|\cdots|\mathbf{x}_n]\tilde{\mathbf{p}} = \mathbf{0}$. Conversely, let $\mathbf{p} \in N([\mathbf{x}_1|\mathbf{x}_2|\cdots|\mathbf{x}_n])$, such that $\mathbf{p} \in \mathbb{R}^n_+$, $\mathbf{p} \neq \mathbf{0}$. Then, one can construct $\tilde{\mathbf{p}}$ as $\tilde{\mathbf{p}} = \mathbf{p}/(\sum_{i=1}^n p_i)$, such that $\sum_{i=1}^n \tilde{p}_i \mathbf{x}_i = 0$, and $\sum_{i=1}^n \tilde{p}_i = 1$, implying that $\mathbf{0} \in \text{conv}\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$. □

Wendel's theorem (Theorem 20) together with Lemma 10, provides the probability with which the null space of certain i.i.d. random matrices contains nonnegative vectors $\mathbf{p} \in \mathbb{R}^n_+$. We now provide a slightly stronger version of this result, by computing the probability with which the null space of such matrices contains a *strictly* positive vector $\mathbf{p} \in \mathbb{R}^n_{++}$. This probability turns out to be identical to that in Theorem 20.

Corollary 4. ***(Corollary to Wendel's Theorem)*** *Let $\mathbf{X} \in \mathbb{R}^{d \times n}$ be a matrix with i.i.d. columns, such that their distribution is symmetric about origin, and with probability one all subsets of d*

columns are linearly independent (i.e., k-rank$(\mathbf{X}) = d$ *w.p. 1). Then*

$$\mathbb{P}(\exists \mathbf{p} \in \mathbb{R}^n_{++}, \mathbf{X}\mathbf{p} = \mathbf{0}) = 1 - \frac{1}{2^{n-1}} \sum_{k=0}^{d-1} \binom{n-1}{k}. \tag{3.29}$$

Proof. Proof can be found in Appendix 3.5.1. □

Using Lemma 10 and Corollary 4, we now show that if the entries of $\mathbf{A}$ are i.i.d. random variables, then we can obtain a lower bound on the probability that $\mathbf{A}$ is ambiguous.

Theorem 21. *Let $\mathbf{A} \in \mathbb{R}^{M \times N}$ be a real valued matrix whose entries are chosen independently from a continuous distribution over $\mathbb{R}$. If $K \geq M^2 + M + 2$, and $N \geq 2K$, then $\mathbb{P}(\mathbf{A} \in I_{\mathbf{A}})$ is at least*

$$1 - e^{-\lfloor \frac{N}{2K} \rfloor (\frac{2K - M^2 - M}{2})(\frac{K-1}{2K-M^2-M} - \log(\frac{K-1}{2K-M^2-M}) - 1)}. \tag{3.30}$$

Proof. See appendix 3.5.2. □

Remark 22. One implication of Theorem 21 is that as N goes to infinity (with $K \geq M^2 + M + 2$, $N \geq 2K$), with probability approaching one, it is possible to find pathological cases such that the identifiability condition (3.3) is violated. As $N \to \infty$, for almost all matrices $\mathbf{A}$ there will exist a pair $(\hat{i}, \hat{j})$ with $1 \leq \hat{i} < \hat{j} \leq \binom{N}{K}$, such that $N(\mathbf{A}_{\hat{i}} \odot \mathbf{A}_{\hat{i}} - \mathbf{A}_{\hat{j}} \odot \mathbf{A}_{\hat{j}})$ contains a positive vector $\hat{\mathbf{p}}_{\mathbf{A}}$. If the distribution of the data is characterized by such $\mathbf{A}, \hat{\mathbf{p}}_{\mathbf{A}}, \mathcal{S}_{\hat{i}}$, i.e.,

$$\mathbf{Y}|\mathcal{S}_{\hat{i}} \sim \mathcal{N}\left(\mathbf{0}, (\mathbf{A}_{\hat{i}} \mathrm{diag}(\hat{\mathbf{p}}_{\mathbf{A}}) \mathbf{A}_{\hat{i}}^H + \sigma^2 \mathbf{I}) \otimes \mathbf{I}_L\right),$$

then by the result of Theorem 21 and Lemma 8, the probability of error of any support detector $\psi(\mathbf{Y})$ will not go to zero, even if the number of snapshots L tends to infinity.

3.1.5 Numerical Experiments

In this section, we will validate our theoretical results through numerical simulations. We will study our upper bounds on the probability of error, and our probabilistic results on existence of ambiguous measurement matrices $\mathbf{A} \in I_{\mathbf{A}}$ that lead to non-identifiable covariance matrices.

Upper Bound on Probability of Error

In this experiment, we compare the performance of the exhaustive search based maximum likelihood (ML) detector (defined in (3.12)), Sparse Bayesian Learning (SBL) algorithm [WR04], and the upper bound on the performance of the multiple hypothesis testing framework, given by (3.23). We choose $M = 4, 5$ and $N = 20$, [3] and consider two cases for K: in the first case (Figures 3.1a, 3.1c) $K < M$ and in the second case (Figures 3.1b, 3.1d), $K > M$. For each value of K, we fix the true support that generates the data $\mathbf{Y}$, compute the empirical conditional probability of error for the two algorithms (given this support), and compare them with the conditional upper bound in (3.23). For the SBL algorithm, the detected support is chosen as the indices corresponding to the K largest elements of the estimated hyperparameter $\boldsymbol{\gamma}$ containing the signal powers. For all the experiments, we use the same measurement matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$, (which is randomly generated from an i.i.d standard normal distribution and then held fixed for all the simulations). Moreover, we assume that all sources have powers equal to 1 (i.e. $\mathbf{p} = \mathbf{1}$), and the noise power is $\sigma^2 = 10^{-2}$. The empirical probability of error for the detectors are computed by running 2000 Monte-Carlo simulations.

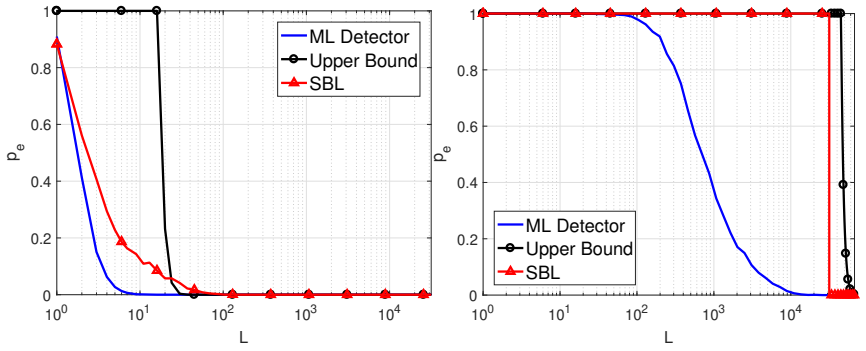
In Figure 3.1, we compare the performance of the detectors against the aforementioned upper bound, as a function of L, for different values of M and K. Here, "ML Detector" refers to the maximum likelihood detector introduced in equation (3.12), "Upper Bound" stands for the upper bound derived for the conditional probability of error for the maximum likelihood

[3] Since we use exhaustive search ML detector in our simulations, we restrict ourselves to smaller values of M and N.

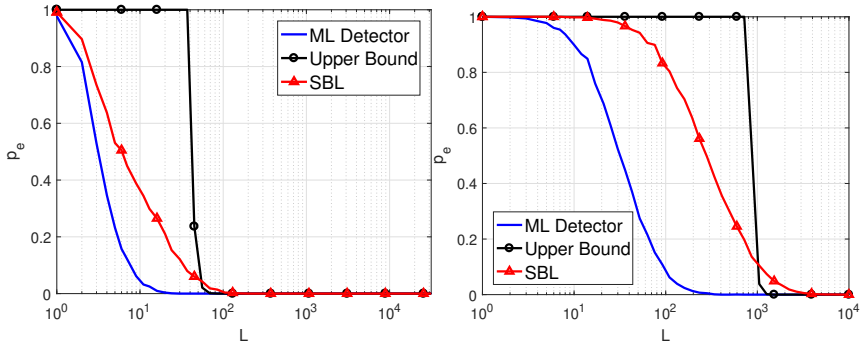
detector, derived in (3.23), and "SBL" indicates the probability of error for the SBL-based support detector. As we can observe, for both $K < M$ and $K > M$, the optimal detector as well as the SBL-based detector can reliably recover the support as L increases. We also observe that the multiple hypothesis testing framework shows a better performance, compared to the SBL-based detector since the former performs an exhaustive search over all possible supports of size K. We also notice that increasing the number of measurements M (even from $M = 4$ to $M = 5$) has a significant effect on the probability of error in the regime $K > M$, whereas the effect is not so pronounced for $K < M$. Finally, the "Upper Bound" shown in Figure 3.1 is only a valid upper bound for the ML detector, and does not necessarily provide an upper bound for the SBL algorithm. This is clear from Figure 3.1a where the error probability of SBL actually exceeds this upper bound in a small interval around $10 < L < 100$ (also Figure 3.1d, around $L = 10^3$). Analyzing the performance of the SBL algorithm requires more careful investigation, and is beyond the scope of this Section.

Failure of Support Recovery

In this experiment, we choose $M = 3$, $N = 17$ and compare the performance of the exhaustive search ML and SBL detectors as in Section 3.1.5. However, we now let $K = 8 > (M^2 + M)/2$. In this case, with probability given in (3.61), there exists a positive vector $\mathbf{p} > \mathbf{0}$ such that the identifiability condition (3.3) is violated, for two given supports $\mathcal{S}_1, \mathcal{S}_2$. We find such a $\mathbf{p}$, by first generating a random measurement matrix $\mathbf{A}$ with i.i.d standard Gaussian entries, and



(a) $K = 3 < M = 4$ (b) $K = 11 > M = 4$



(c) $K = 4 < M = 5$ (d) $K = 11 > M = 5$

Figure 3.1: Probability of error for exhaustive search based Maximum Likelihood (ML) detector and SBL algorithm, as a function of L. Here, $N = 20$. The upper bound (3.23) on the probability of error for the ML detector is also plotted for reference.

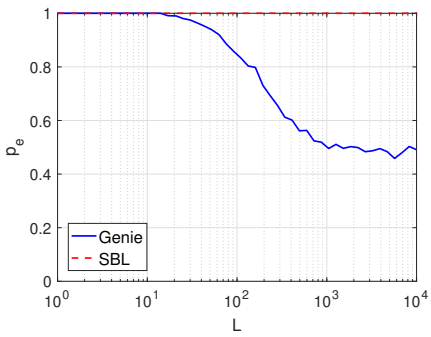


Figure 3.2: Probability of error for the ML detector and SBL algorithm as a function of L. Here, $M = 3, N = 17, K = 8 > \frac{M^2+M}{2}$. The probability of error does not go to zero as L increases.

then solving the following feasibility problem

$$\begin{aligned} &\text{find } \mathbf{p} \\ &\text{subject to} \\ &\mathbf{p} > 0 \\ &(\mathbf{A}_1 \odot \mathbf{A}_1 - \mathbf{A}_2 \odot \mathbf{A}_2)\mathbf{p} = \mathbf{0} \end{aligned} \tag{3.31}$$

where $\mathcal{S}_1 = \{1, \cdots, K\}$ and $\mathcal{S}_2 = \{K+1, \cdots, 2K\}$. If (3.31) turns out to be infeasible, we generate another random measurement matrix $\mathbf{A}$, and repeat solving the feasibility problem (3.31), until we find an $\mathbf{A} \in I_{\mathbf{A}}$ and its corresponding $\mathbf{p}$. We observe that in this regime, the probability of error for both the detectors remain bounded away from zero. As depicted in Figure 3.2, "SBL" always fails to detect the true support. Even the probability of error for "Genie"-aided detector [4] (that knows the true value of power of the K sources $\mathbf{p}$), does not go to zero for large L.

[4] For definition of *genie* refer to the proof of Lemma 8 in Section 3.1.4

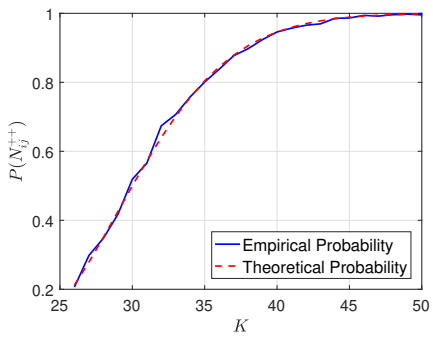


Figure 3.3: Empirical and theoretical probability of the event N_{12}^{++}, for $M = 5, N \geq 2K$.

Existence of Ambiguous Measurement Matrix

In this experiment, we conduct Monte-Carlo simulations to empirically compute the probability of the event N_{12}^{++}

$$N_{12}^{++} = \{\exists \mathbf{p} > \mathbf{0} : (\mathbf{A}_1 \odot \mathbf{A}_1 - \mathbf{A}_2 \odot \mathbf{A}_2)\mathbf{p} = \mathbf{0}\}$$

The goal is to verify that this empirical probability matches the theoretical value given by the RHS of (3.58), which will validate our corollary to Wendel's theorem. We choose two disjoint supports $\mathcal{S}_1$ and $\mathcal{S}_2$ ($\mathcal{S}_1 \cap \mathcal{S}_2 = \emptyset$) and generate $\mathbf{A}_1$ and $\mathbf{A}_2$ with i.i.d standard Gaussian entries. In order to determine if the event N_{12}^{++} occurs, we solve the feasibility problem (3.31).

We compute the probability with which N_{12}^{++} occurs by counting the number of times (3.31) returns a solution (feasibility test) out of 2000 Monte-Carlo runs. The corresponding probability is plotted in Figure 3.3, as a function of K. In this simulation, $M = 5$, and $N \geq 2K$ (notice that the probability of the event N_{12}^{++} is not a function of N.). We observe that the plots corresponding to the empirical probability of the event N_{12}^{++}, and the theoretical probability from (3.61) match perfectly.

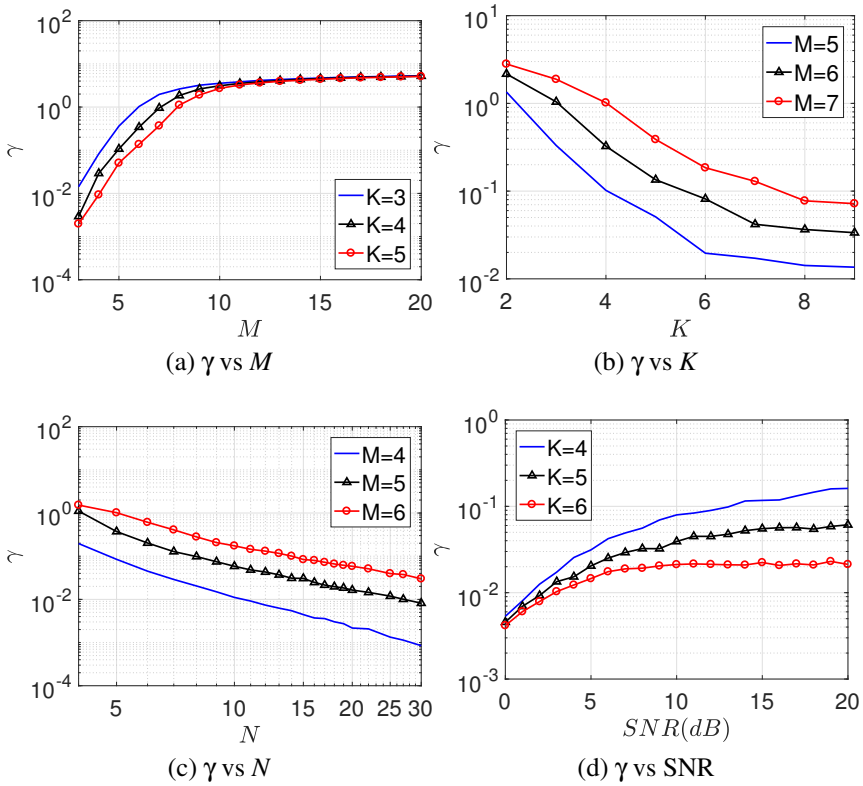
Scaling of γ with respect to dimensions and SNR

In this experiment, we numerically study the behavior of the error exponent γ, defined

in (3.22), as functions of M, N, K and signal-to-noise ratio (SNR) p/σ^2. In all experiments, we assume the elements of $\mathbf{A}$ to be i.i.d. according to $\mathbb{R}\mathcal{N}(0,1/M)$, and for each realization of $\mathbf{A}$, we compute γ following (3.22), and plot the average value of γ over 200 runs for each problem setting. Figures 3.4a, 3.4b, 3.4c, and 3.4d show the behavior of γ as functions of M, K, N, and SNR, respectively. In Figures 3.4a, 3.4b, and 3.4d, we have $N = 10$, in Figure 3.4c, we have $K = 3$, in Figure 3.4d, we have $M = 5$, and in Figures 3.4a, 3.4c, 3.4b, $SNR = 20dB$. We observe that γ exhibits an increasing trend with respect to M and SNR, which means that for larger M or SNR, the probability of error in (3.24) will have a faster decay rate with respect to L. This is consistent with the intuition that for larger spatial samples (M) or SNR, one can attain same probability of error with fewer temporal samples (L). Moreover, we also observe that γ is a decreasing function of K and N. With respect to N, from the log-log plot in 3.4c, it seems that γ decays approximately as $1/N^{\varepsilon}$, for some $\varepsilon > 0$. Analytically characterizing the behavior of γ with respect to M, N, K, and SNR can be an interesting topic for future research.

3.1.6 Conclusion

We provided new probabilistic guarantees for recovering the common support of jointly sparse vectors. We formulated the support recovery problem in terms of a multiple hypothesis testing framework, and derived Chernoff-type upper bounds on the probability of detecting a wrong support. We established sufficient conditions under which the probability of error is guaranteed to have an exponential decay with respect to the number of measurements L, even when $K = O(M^2)$ (where K denotes the sparsity, and M is the size of each measurement vector). Our sufficient conditions are based on properties of the Khatri-Rao product of the measurement matrix, and also indicates the role of sampler design. We also established that when K is larger than a certain threshold (in terms of M), there will exist a class of measurement matrices for which the probability of error of any support recovery algorithm will be bounded away from zero, even when $L \to \infty$. We characterized the probability with which a randomly generated



(a) γ vs M (b) γ vs K

(c) γ vs N (d) γ vs SNR

Figure 3.4: Characterization of the behavior of γ as functions of M, N, K, and SNR.

measurement matrix will belong to this class, by leveraging results from geometric probability, and showed that this probability tends to 1 asymptotically in the size (N) of the sparse vectors.

3.2 Non-Asymptotic Guarantees for Correlation-Aware Support Detection

In Section 3.2, using the same signal model as [TN10], we showed that it is possible to recover supports of size $K = O(M^2)$ for appropriate measurement matrices, as long as the non-zero signals have equal power and the detector knows K. In this section, we relax both conditions and show that is possible to recover supports of size $K = O(M^2)$ even for sources with unequal power, and without the knowledge of K. Unlike [TN10, KQP18], we do not impose specific distribution on the measurements and only assume them to be bounded real-valued random variables. Using a simple least squares estimate of the source powers, followed by hard-thresholding, we are able to recover sparse supports of size $K > M$, with overwhelming probability (with respect to L).

3.2.1 Signal Model

In this Section, we consider the MMV model introduced in (3.1) with L measurement vectors. We make the following statistical assumptions on the signal and noise:

(A1) Non-zero elements of the signal $\mathbf{x}[l]$ are uncorrelated, i.e. $\mathbb{E}(\mathbf{x}[l]\mathbf{x}[l]^H) = \mathbf{P}$, where $\mathbf{P} = \text{diag}(p_1, \cdots, p_N)$ is a diagonal matrix, and $\{\mathbf{x}[l]\}_{l=0}^{L-1}$ are independent and identically distributed (i.i.d.) random vectors. Moreover $p_i = 0, i \neq \mathcal{S}$, consistent with the fact that the L vectors $\mathbf{x}[l], 0 \leq l \leq L-1$ share a common support $\mathcal{S}$.

(A2) Signal $\mathbf{x}[l]$ and noise $\mathbf{w}[l]$ are uncorrelated, i.e, $\mathbb{E}(\mathbf{x}[l]\mathbf{w}[l]^H) = \mathbf{0}$.

(A3) The noise $\mathbf{w}$ is white, i.e. $\mathbb{E}(\mathbf{w}[l]\mathbf{w}[l]^H) = \sigma^2\mathbf{I}$, and $\{\mathbf{w}[l]\}_{l=0}^{L-1}$ are i.i.d. random variables. We assume σ^2 is known.

(A4) The signal and noise are *bounded random variables*, i.e., $\|\mathbf{x}[l]\|_2 \leq C_x$, $\|\mathbf{w}[l]\|_2 \leq C_w$, where $C_x, C_w > 0$ are positive constants.

(A5) The measurement matrix $\mathbf{A}$ satisfies $\text{rank}(\mathbf{A}^* \odot \mathbf{A}) = N$.

Assumptions (**A1-A3**) are typical in the context of Sparse Bayesian Learning (SBL) [WR04], line spectrum estimation and so forth. However, unlike SBL, (**A4**) further enforces the signal and noise to be bounded random variables. This assumption simplifies the error analysis of our proposed detector in the regime $K > M$, and ensures that the error decays exponentially fast in L. Unlike SBL, we do not consider any particular distribution for the measurements. The following remark immediately follows from the assumption (**A4**):

Remark 23. *Under assumption (***A4***), we have* $\|\mathbf{y}\| \leq C_y$

$$C_y = \sigma_{\max}(\mathbf{A})C_x + C_w \tag{3.32}$$

where $\sigma_{\max}(\mathbf{A})$ *denotes the maximum singular value of the measurement matrix* $\mathbf{A}$.

Remark 24. *Based on assumptions (***A1-A3***), one can write the covariance matrix of the measurement vectors as*

$$\mathbf{R} := \mathbb{E}(\mathbf{y}[l]\mathbf{y}[l]^H) = \mathbf{A}\mathbf{P}\mathbf{A}^H + \sigma^2\mathbf{I}.$$

The vectorized form of the covariance matrix can be written as

$$\text{vec}(\mathbf{R}) = (\mathbf{A}^* \odot \mathbf{A})\mathbf{p} + \sigma^2 \text{vec}(\mathbf{I})$$

where $\mathbf{p} = [p_1, \cdots, p_N]^T$ is a sparse vector with support $\mathcal{S}$. The goal of support recovery in MMV models is to detect the common support $\mathcal{S}$ from the measurements $\mathbf{Y} = [\mathbf{y}[0], \cdots, \mathbf{y}[L-1]]$. Let $\psi(\mathbf{Y})$ denote a detector that returns a candidate support. The probability of detecting a wrong

support, given $\mathcal{S}$ is the true support, can be expressed as

$$p_{e|\mathcal{S}} = \mathbb{P}(\psi(\mathbf{Y}) \neq \mathcal{S}|\mathcal{S})$$

It has been empirically demonstrated that SBL is capable of detecting supports of size larger than M, but no theoretical guarantees exist. In this Section, we propose a simple detector $\psi_{\mathrm{LS}}(\mathbf{Y})$ (that *does not know the support size K*), and compute upper bounds on the probability of error $p_{e|\mathcal{S}}$ of this detector. Before presenting our results, we review existing results that consider support recovery in the regime $K > M$ but only provide partial guarantees.

3.2.2 Review of Correlation-Aware Techniques for Recovering Supports of Size $K = O(M^2)$

In compressed sensing, existing guarantees for sparse support recovery are mostly relevant in the regime $K < M$. The only algorithms, which, under certain restrictive assumptions, theoretically or experimentally show possibility of recovering supports of size $K > M$, are Sparse Bayesian Learning [WR04, BKDM14], and Correlation-Aware LASSO (Co-LASSO) [PV15]. We now briefly review these results and elaborate more on the role of correlation awareness in recovering supports of size $K = O(M^2)$.

1. *Sparse Bayesian Learning:* The authors in [BKDM14] show that the MSBL algorithm is capable of recovering supports of size $K > M$ under the following assumptions: 1) The measurements are assumed to be noiseless. 2) Non-zero rows of $\mathbf{X} = [\mathbf{x}[0], \mathbf{x}[1], \cdots, \mathbf{x}[L-1]]$ are orthogonal. Although these conditions may not be satisfied in practice, their numerical results show that even under a noisy setting MSBL is able to recover supports of size $K > M$.

2. *Correlation-Aware Support Recovery:* In our earlier work in [PV15], we showed that if we

have access to the exact covariance matrix $\mathbf{R}$ (which happens when $L \to \infty$), then, under assumptions **(A1-A3)** and **(A5)**, it is possible to recover sparse supports of size $K = O(M^2)$, by solving the following ℓ_1 minimization problem:

$$\min_{\mathbf{p} \geq \mathbf{0}} \|\mathbf{p}\|_1 \qquad \text{subject to} \qquad (\mathbf{A}^* \odot \mathbf{A})\mathbf{p} = \text{vec}(\mathbf{R})$$

For finite L, we can only compute an estimate of $\mathbf{R}$. In this case, we proposed a variation of LASSO [PV15] namely (Co-LASSO) for joint support recovery, and showed that it can recover $\mathcal{S}$ as long as $K < \frac{1}{2}(1 + \frac{1}{\mu^2})$. Here $\mu \leq 1$ is the mutual coherence of $\mathbf{A}$ defined as

$$\mu = \max_{i \neq j} \frac{|\mathbf{a}_i^H \mathbf{a}_j|}{\|\mathbf{a}_i\|_2 \|\mathbf{a}_j\|_2}$$

This result showed that by merely exploiting the lack of correlation between sparse signals, one can recover larger supports compared to traditional coherence-based guarantees in compressed sensing (which require $K < \frac{1}{2}(1 + \frac{1}{\mu})$) [Tib96]. However, in presence of finite L, these guarantees are rather weak and only apply in the regime $K < M$.

3. *Existence of Cramér Rao Bound, when $N = O(M^2)$:* In past work [PV14a], we showed that the Cramér Rao Bound (CRB) for estimating source powers in a MMV model (3.1) exist, even when $K = O(M^2)$, as long as $\text{rank}(\mathbf{A} \odot \mathbf{A}) = N$. This condition is obeyed by almost all choices of $\mathbf{A}$ if $N \leq \frac{M^2+M}{2}$. In this setting, as $L \to \infty$, the CRB goes to zero at the rate $1/L$. Since Maximum Likelihood (ML) Estimates asymptotically attain the CRB, this automatically shows that MSBL can recover the vector $\mathbf{p}$ as $L \to \infty$ since it solves a maximum likelihood problem.

Most of aforementioned results provide asymptotic guarantees (i.e. when $L \to \infty$). No non-asymptotic guarantees currently exist for support recovery in the regime $K = O(M^2)$ that can ensure $p_{e|\mathcal{S}}$ decays exponentially fast in L. In the next section, we will address this issue by

proposing a detector which is based on solving a simple least squares problem followed by a thresholding step.

3.2.3 A Least Squares Thresholding Based Support Detector

We propose the following simple detector based on least-squares method:

$$\psi(\mathbf{Y};\tau,\mathbf{A},\sigma^2) = \{i|\hat{p}_i \geq \tau, \hat{\mathbf{p}} = \phi(\mathbf{Y};\mathbf{A},\sigma^2)\} \tag{3.33}$$

where τ is a predefined threshold, and

$$\phi(\mathbf{Y};\mathbf{A},\sigma^2) = (\mathbf{A}^* \odot \mathbf{A})^\dagger \operatorname{vec}(\hat{\mathbf{R}} - \sigma^2\mathbf{I}) \tag{3.34}$$

is the least square estimator of the vector of source powers $\mathbf{p}$, where $\hat{\mathbf{R}}$ denotes the sample covariance matrix, defined as

$$\hat{\mathbf{R}} = \frac{1}{L}\sum_{l=1}^{L} \mathbf{y}[l]\mathbf{y}[l]^H \tag{3.35}$$

Inspite of its simplicity, we will now show that this detector can recover supports of size $K = O(M^2)$ with overwhelming probability. [5] We first state some preliminary lemmas:

Lemma 11. *The estimator* (3.34) *is unbiased, i.e* $\mathbb{E}(\hat{\mathbf{p}}) = \mathbf{p}$.

[5] Although the MMV model is underdetermined ($N > M$), under assumption (**A5**), $\mathbf{A}^* \odot \mathbf{A}$ is tall and has full column-rank. Hence it is reasonable to estimate $\mathbf{p}$ using least squares method. Assumption (**A5**) continues to hold in the regime $M < N < (M^2+M)/2$ for almost all $\mathbf{A} \in \mathbb{R}^{M\times N}$, and it serves as a necessary condition for existence of CRB [PV15, PV14a].

Proof. Let $\hat{\mathbf{p}} = \phi(\mathbf{Y}; \mathbf{A}, \sigma^2)$. We have

$$\begin{aligned}
\mathbb{E}(\hat{\mathbf{p}}) &= \mathbb{E}((\mathbf{A}^* \odot \mathbf{A})^\dagger \operatorname{vec}(\hat{\mathbf{R}} - \sigma^2 \mathbf{I})) \\
&= (\mathbf{A}^* \odot \mathbf{A})^\dagger \operatorname{vec}(\mathbb{E}(\hat{\mathbf{R}}) - \sigma^2 \mathbf{I})) \\
&= (\mathbf{A}^* \odot \mathbf{A})^\dagger \operatorname{vec}(\mathbf{R} - \sigma^2 \mathbf{I})) \qquad (3.36) \\
&= (\mathbf{A}^* \odot \mathbf{A})^\dagger \operatorname{vec}(\mathbf{A}\mathbf{P}\mathbf{A}^H) \\
&= (\mathbf{A}^* \odot \mathbf{A})^\dagger (\mathbf{A}^* \odot \mathbf{A})\mathbf{p} \\
&= \mathbf{p} \qquad (3.37)
\end{aligned}$$

where (3.36) follows from the fact that $\mathbb{E}(\hat{\mathbf{R}}) = \frac{1}{L}\sum_{l=1}^{L} \mathbb{E}(\mathbf{y}[l]\mathbf{y}[l]^H) = \mathbf{R}$, and (3.37) holds due to assumption **(A5)**. □

Lemma 12. *The estimator* (3.34) *can be also be written as*

$$\hat{p}_i = \frac{1}{L} \sum_{l=0}^{L-1} \sum_{j=1}^{N} b_{ij} (|\mathbf{a}_j^H \mathbf{y}[l]|^2 - \sigma^2 \|\mathbf{a}_j\|^2) \qquad (3.38)$$

where $\mathbf{B} := [b_{ij}] = \left((\mathbf{A}^* \odot \mathbf{A})^H (\mathbf{A}^* \odot \mathbf{A})\right)^{-1}$.

Proof. Following the definition of matrix $\mathbf{B}$, one can write the estimator (3.33) as

$$\begin{aligned}
\hat{\mathbf{p}} &= \mathbf{B}(\mathbf{A}^* \odot \mathbf{A})^H \operatorname{vec}(\hat{\mathbf{R}} - \sigma^2 \mathbf{I}) \\
&= \mathbf{B}\mathbf{J}^H (\mathbf{A}^* \otimes \mathbf{A})^H \operatorname{vec}(\hat{\mathbf{R}} - \sigma^2 \mathbf{I}) \\
&= \mathbf{B}\mathbf{J}^H \operatorname{vec}(\mathbf{A}^H (\hat{\mathbf{R}} - \sigma^2 \mathbf{I})\mathbf{A}) \\
&= \mathbf{B} \operatorname{diag}(\mathbf{A}^H (\hat{\mathbf{R}} - \sigma^2 \mathbf{I})\mathbf{A}) \qquad (3.39) \\
&= \frac{1}{L} \sum_{l=0}^{L-1} \mathbf{B} \operatorname{diag}(\mathbf{A}^H (\mathbf{y}[l]\mathbf{y}[l]^H - \sigma^2 \mathbf{I})\mathbf{A}) \qquad (3.40)
\end{aligned}$$

where $\mathbf{J} \in \mathbb{R}^{N^2 \times N}$ is an appropriate column selection matrix, $\operatorname{diag}(\mathbf{X})$ (with matrix argument $\mathbf{X}$)

returns a column vector containing the diagonal entries of the matrix $\mathbf{X}$. The equation (3.39) follows by exploiting the structure of the matrix $\mathbf{J}$, and (3.40) follows from the definition of $\hat{\mathbf{R}}$ in (3.35) and changing the order of summations. □

To facilitate our analysis, for the rest of this Section, we will further assume that $\mathbb{F} = \mathbb{R}$, i.e., all random variables and the measurement matrix $\mathbf{A}$ are real valued.

Lemma 13. *Given any $i \in \{1, \cdots, N\}$ and $\eta > 0$, it holds that*

$$\mathbb{P}(|\hat{p}_i - p_i| > \eta) \leq 2e^{-\beta_i L \eta^2}$$

where β_i is a constant (specified in the proof).

Proof. Using the result of Lemma 12, we can write

$$\hat{p}_i = \frac{1}{L} \sum_{l=0}^{L-1} (z_l^{(i)} - \hat{e}_i)$$

where $\hat{e}_i = \sum_{j=1}^{N} b_{ij} \sigma^2 \|\mathbf{a}_j\|^2$, and

$$z_l^{(i)} = \sum_{j=1}^{N} b_{ij} |\mathbf{a}_j^H \mathbf{y}[l]|^2.$$

Next, we show that each $|z_l^{(i)}|$ is bounded. We have

$$|z_l^{(i)}| \leq \sum_{j=1}^{N} |b_{ij}| |\mathbf{a}_j^H \mathbf{y}[l]|^2 \tag{3.41}$$

$$\leq \sum_{j=1}^{N} |b_{ij}| \|\mathbf{a}_j\|^2 \|\mathbf{y}[l]\|^2 \tag{3.42}$$

$$\leq C_y^2 \sum_{j=1}^{N} |b_{ij}| \|\mathbf{a}_j\|^2 := C_z^{(i)} \tag{3.43}$$

From Lemma 11 we know that $\mathbb{E}(\hat{p}_i) = p_i$. Therefore, using Hoeffding Inequality [Hoe63], we

obtain

$$\mathbb{P}(|\hat{p}_i - p_i| > \eta) \leq 2e^{-\frac{L\eta^2}{2(C_z^{(i)})^2}}$$

which concludes the proof by choosing $\beta_i = \frac{1}{2(C_z^{(i)})^2}$. □

Equipped with the above lemmas, we are now ready to state our main result:

Theorem 22. *Under assumptions* **(A1-A5)***, the probability of error $p_{e|\mathcal{S}}$ of the detector* (3.33) *with $\tau = \frac{p_{\min}}{2}$ is upper bounded as*

$$p_{e|\mathcal{S}} \leq e^{-\beta p_{\min}^2 L/4 + \log(2N))}$$

where $p_{\min} := \min_{i \in \mathcal{S}} p_i$, and $\beta = \min_i \frac{1}{2(C_z^{(i)})^2}$, with $C_z^{(i)}$ given by (3.43)*.*

Proof. For the detector specified by (3.33), consider any threshold τ such that $\tau < p_{\min}$. In this case, the probability of detecting a wrong support (given $\mathcal{S}$ is the true support) can be written as

$$\begin{aligned}
p_{e|\mathcal{S}} &= \mathbb{P}(\psi(\mathbf{Y}; \tau, \mathbf{A}, \sigma^2) \neq \mathcal{S} | \mathcal{S}) \\
&= \mathbb{P}\left(\bigcup_{i \in \mathcal{S}} \{\hat{p}_i < \tau\} \cup \bigcup_{i \notin \mathcal{S}} \{\hat{p}_i > \tau\} \right) && (3.44) \\
&\leq \sum_{i \in \mathcal{S}} \mathbb{P}(\hat{p}_i < \tau) + \sum_{i \notin \mathcal{S}} \mathbb{P}(\hat{p}_i > \tau) && (3.45) \\
&\leq \sum_{i \in \mathcal{S}} \mathbb{P}(|\hat{p}_i - p_i| > p_i - \tau) + \sum_{i \notin \mathcal{S}} \mathbb{P}(|\hat{p}_i| > \tau) && (3.46)
\end{aligned}$$

where (3.45) follows from the union bound, and (3.46) follows from the fact that $\hat{p}_i \leq \tau$ is

equivalent to $\hat{p}_i - p_i \leq \tau - p_i$, which implies $|\hat{p}_i - p_i| \geq p_i - \tau$ [6] Using Lemma 13, we have

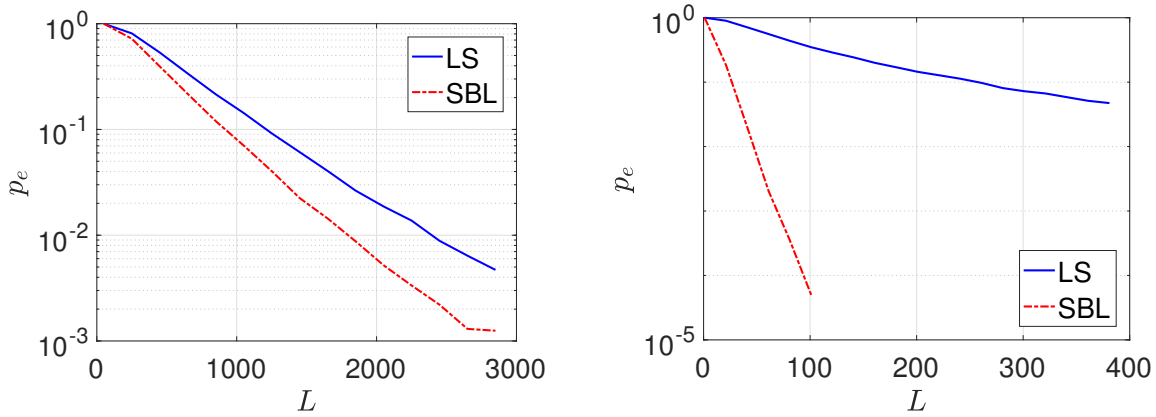
$$\begin{aligned} p_{e|\mathcal{S}} &\leq \sum_{i \in \mathcal{S}} 2e^{-\beta_i L(p_i-\tau)^2} + \sum_{i \notin \mathcal{S}} 2e^{-\beta_i L \tau^2} \\ &\leq 2Ke^{-\beta L(p_{\min}-\tau)^2} + 2(N-K)e^{-\beta L \tau^2} \end{aligned}$$

where $\beta = \min_i \beta_i$. Substituting $\tau = \frac{p_{\min}}{2}$ concludes the proof. □

3.2.4 Simulations

We now numerically validate that it is possible to obtain exponentially decaying probability of error for support recovery in the regime $K > M$. To this end, we consider two algorithms: i) MSBL [BKDM14], and ii) the proposed detector in (3.33). We consider a fixed measurement matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$, $M = 7, N = 21$. The ith nonzero element of $\mathbf{x}[l]$ is chosen from the uniform distribution over $[-\sqrt{3p_i}, \sqrt{3p_i}]$ (which will ensure that $\mathbb{E}(x_i^2[l]) = p_i$). The elements of the noise vector $w_i[l]$ are i.i.d. and uniformly distributed in the range $[-\sqrt{3}\sigma, \sqrt{3}\sigma]$. We let $p_i = 1$, for $i \in \mathcal{S}$, and $\sigma = 0.1$. For the proposed least square detector, we set the threshold $\tau = \frac{1}{2}$. We use the same threshold τ to detect the support using MSBL. Fig. 3.5 shows the probability of error of both detectors as a function of L (log scale), for both $K > M$ and $K < M$. It is clear that the slope for both detectors is linear in L, indicating an exponential decay of $p_{e|\mathcal{S}}$ with respect to L. It can also be seen that MSBL has a better error exponent compared to the least squares detector, in both the regimes. It will be of interest in future to analyze the performance of MSBL, and characterize this error exponent.

[6] Since $\tau < p_{\min}$, we have $\tau - p_i < 0$ for all $i \in \mathcal{S}$



(a) $K = 17 > M, N = 21$ (b) $K = 6 < M, N = 21$

Figure 3.5: Probability of error of both detectors ("LS" denotes the proposed least squares detector, and "SBL" denotes the detector based on Sparse Bayesian Learning algorithm.)

3.2.5 Conclusion

In this Section, we considered the problem of joint support recovery of sparse signals in multiple measurement vector (MMV) models. For the first time, we provided non-asymptotic guarantees for recovering supports of size $K = O(M^2)$, where M is the size of each measurement vector. Our detector is based on a simple least square estimator of source powers, followed by a hard thresholding operation. Assuming the sparse signals to be statistically uncorrelated bounded random variables, we can ensure that the probability of detecting a wrong support approaches zero exponentially fast in L even when $K > M$. This result holds for appropriately designed measurement matrices whose Khatri-Rao products satisfy certain rank constraints.

3.3 A Greedy Approach for Correlation-Aware Sparse Support Recovery

In Section 3.1, we showed that under same assumptions as [TN10], it is possible to have *exponentially decaying* probability of error as $L \to \infty$, even when $K > M$, as long as $K = O(M^2)$. Moreover, we also showed negative results indicating that when $K = \Omega(M^2)$, there exist problem instances for which, the probability of error is bounded away from zero, even for $L \to \infty$. In Section 3.2, we showed that if the signal and noise are assumed to be bounded random variables, then recovering $K = O(M^2)$ with exponentially decaying p_e (in terms of L) is possible, using a support detector which performs a least squares inversion followed by a hard-thresholding step.

In this section, we show the possibility of $K > M$ under the setting that the sources are Gaussian (unbounded), and may have unequal powers. We show that for certain measurement matrices (such as equiangular tight frames), recovering supports of size $K > M$, is possible. Furthermore, if equiangular tight frames of size $N = O(M^2)$ exist, we can also provide guarantees for the support recovery problem when $K = O(M^2)$. However, the existence of such equiangular tight frames turns out to be an open problem for general M [Zau11].

3.3.1 Problem Formulation

We consider the measurement model (3.1) and make the following statistical assumptions on signals $\mathbf{x}[l]$, $\mathbf{w}[l]$, and the measurement matrix $\mathbf{A}$:

- **(A1)** The signals $\mathbf{x}[l]$ are identically and independently distributed (i.i.d.) according to normal distribution $\mathbb{F}\mathcal{N}(0,\mathbf{P})$[7], where $\mathbf{P} = \text{diag}(\mathbf{p})$ is a diagonal matrix, and $\mathbf{p}$ is a sparse vector with support $\mathcal{S}$.

- **(A2)** We assume the noise to be i.i.d. Gaussian with $\mathbb{F}\mathcal{N}(\mathbf{0},\sigma^2\mathbf{I})$, with known σ^2.

[7] In the case, $\mathbb{F} = \mathbb{C}$, the notation $\mathbb{C}\mathcal{N}(.,.)$ refers to circularly symmetric complex Gaussian distribution

- **(A3)** We further assume that the signal and noise are uncorrelated, i.e., $\mathbb{E}(\mathbf{x}[l]\mathbf{w}[l]^H) = \mathbf{0}$.

- **(A4)** The columns of $\mathbf{A}$ have equal norms, i.e., $\|\mathbf{a}_i\|_2 = c$, for some constant c, and all $1 \leq i \leq N$.

Under **(A1)**-**(A3)**, $\mathbf{y}[l]$ ($l = 0, 1, \cdots, L-1$) are i.i.d. vectors according to $\mathbb{F}\mathcal{N}(\mathbf{0}, \mathbb{R})$, where

$$\mathbf{R} = \mathbf{APA}^H + \sigma^2\mathbf{I}.$$

The goal of support recovery from MMV models is to recover $\mathcal{S}$ from $\mathbf{Y} = [\mathbf{y}[1], \cdots, \mathbf{y}[L]]$. Let $\hat{\mathcal{S}} = \psi(\mathbf{Y})$, for a support detector $\psi(.)$. Corresponding to a support detector $\psi(.)$, one can define the probability of error as $p_e = \mathbb{P}(\hat{\mathcal{S}} \neq \mathcal{S})$. In the following Section, we will propose a greedy support detection algorithm, and derive an upper bound for the corresponding probability of error. We will develop conditions under which the probability of error can go to zero (even for $K > M$), exponentially fast in L.

3.3.2 Proposed Support Detector

Let $\hat{\mathbf{R}} = \mathbf{YY}^H/L$ be the sample covariance matrix of the received signals. We consider the following greedy support detector, assuming knowledge of σ^2:

1. Compute ρ_i as

$$\rho_i := \mathbf{a}_i^H(\hat{\mathbf{R}} - \sigma^2\mathbf{I})\mathbf{a}_i, i = 1, \cdots, N.$$

2. Let $j_1, j_2, \cdots, j_N$ be such that $\rho_{j_1} \geq \rho_{j_2} \geq \cdots \geq \rho_{j_N}$. Choose $\hat{\mathcal{S}}$ to be

$$\hat{\mathcal{S}} = \{j_1, j_2, \cdots, j_K\}.$$

One can think of ρ_i as the "correlation" of $\mathbf{a}_i^* \otimes \mathbf{a}_i$ with the vectorized form of $\hat{\mathbf{R}} - \sigma^2\mathbf{I}$. This support detector simply computes all these "correlation" values, and reports the detected support $\hat{\mathcal{S}}$ as the indices corresponding to the K largest values. This can be thought of performing a simple matching pursuit on the vectorized form of the sample covariance matrix.

Analyzing the Probability of Error

The probability of detecting a wrong support using the proposed detector is given by

$$p_e = \mathbb{P}(\exists i \in \mathcal{S}, j \notin \mathcal{S} : \rho_i < \rho_j) = \mathbb{P}(\bigcup_{i\in\mathcal{S}, j\notin\mathcal{S}} \{\rho_i < \rho_j\})$$

By the application of union bound, we get

$$p_e \leq \sum_{i\in\mathcal{S}} \sum_{j\notin\mathcal{S}} \mathbb{P}(\rho_i < \rho_j)$$

Hence, we only need to find an upper bound on $\mathbb{P}(\rho_i < \rho_j)$. In order to do so, let us first take a closer look at the quantity $\rho_i - \rho_j$. Since **(A4)** holds, we have

$$\begin{aligned}
\rho_i - \rho_j &= \frac{1}{L}\operatorname{tr}(\mathbf{a}_i^H \mathbf{Y}\mathbf{Y}^H \mathbf{a}_i) - \frac{1}{L}\operatorname{tr}(\mathbf{a}_j^H \mathbf{Y}\mathbf{Y}^H \mathbf{a}_j) - \sigma^2(\mathbf{a}_i^H\mathbf{a}_i - \mathbf{a}_j^H\mathbf{a}_j) \\
&= \frac{1}{L}\operatorname{tr}(\mathbf{Y}^H \mathbf{a}_i\mathbf{a}_i^H \mathbf{Y}) - \frac{1}{L}\operatorname{tr}(\mathbf{Y}^H \mathbf{a}_j\mathbf{a}_j^H \mathbf{Y}) \\
&= \frac{1}{L}\sum_{\ell=1}^{L} \mathbf{y}_\ell^H(\mathbf{a}_i\mathbf{a}_i^H - \mathbf{a}_j\mathbf{a}_j^H)\mathbf{y}_\ell
\end{aligned}$$

Let $\mathbf{z}_\ell$ be i.i.d random variables distribued as $\mathbb{F}\mathcal{N}(0, \mathbf{I}_M)$. Then, we have

$$\begin{aligned}
\mathbb{P}(\rho_i < \rho_j) &= \mathbb{P}(\frac{1}{L}\sum_{\ell=1}^{L} \mathbf{y}_\ell^H(\mathbf{a}_i\mathbf{a}_i^H - \mathbf{a}_j\mathbf{a}_j^H)\mathbf{y}_\ell < 0) \\
&= \mathbb{P}(\sum_{\ell=1}^{L} \mathbf{z}_\ell^H \mathbf{R}^{1/2}(\mathbf{a}_i\mathbf{a}_i^H - \mathbf{a}_j\mathbf{a}_j^H)\mathbf{R}^{1/2}\mathbf{z}_\ell < 0)
\end{aligned}$$

Therefore, we have

$$\mathbb{P}(\rho_i < \rho_j) = \mathbb{P}(\mathbf{z}^H(\mathbf{I}_M \otimes \mathbf{Q}_{ij})\mathbf{z} < 0)$$

where $\mathbf{z} \sim \mathcal{N}(0, \mathbf{I}_{ML})$, and

$$\mathbf{Q}_{ij} = \mathbf{R}^{1/2}(\mathbf{a}_i\mathbf{a}_i^H - \mathbf{a}_j\mathbf{a}_j^H)\mathbf{R}^{1/2}.$$

Notice that the matrices $\mathbf{Q}_{ij}$ have rank at most 2. We let $\lambda_1^{(i,j)} \geq \lambda_2^{(i,j)}$ be the nonzero eigenvalues of $\mathbf{Q}_{ij}$. The following Lemma provides an upper bound for the probability $\mathbb{P}(\rho_i < \rho_j)$, computed based on Chernoff bound.

Lemma 14. *As long as*

$$\lambda_1^{(i,j)} > 0 > \lambda_2^{(i,j)} \text{ and } \lambda_1^{(i,j)} > |\lambda_2^{(i,j)}|, \tag{3.47}$$

it holds that

$$\mathbb{P}(\rho_i < \rho_j) \leq e^{-\gamma_{ij}L}$$

where $\gamma_{ij} = \kappa_{\mathbb{F}} \log \frac{\lambda_1^{(i,j)} + |\lambda_2^{(i,j)}|}{2\sqrt{\lambda_1^{(i,j)}|\lambda_2^{(i,j)}|}}$, *where* $\kappa_{\mathbb{C}} = 2$, *and* $\kappa_{\mathbb{R}} = 1$.

By applying union bound on all pairs of $i \in \mathcal{S}$ and $j \notin \mathcal{S}$, we get the following Corollary:

Corollary 5. *If the condition* (3.47) *holds for every* $i \in \mathcal{S}$, *and* $j \notin \mathcal{S}$, *it follows that*

$$p_e \leq \sum_{i \in \mathcal{S}} \sum_{j \notin \mathcal{S}} e^{-\gamma_{ij}L}$$

In order to obtain a sufficient condition for exponential decay of p_e with respect to L, we need to characterize the condition (3.47). In the following subsection, we will establish sufficient

conditions on the measurement matrix $\mathbf{A}$, which will guarantee that the condition (3.47) holds for every $i \in \mathcal{S}, j \notin \mathcal{S}$.

Conditions for Exponential Decay

In this section, we will study the condition (3.47) more closely, and derive sufficient conditions under which (3.47) holds, implying that p_e will have exponentially decaying probability of error. The following Lemma establishes an equivalent condition for (3.47):

Lemma 15. *For every $i \in \mathcal{S}$, $j \notin \mathcal{S}$, the condition*

$$\lambda_1^{(i,j)} > 0 > \lambda_2^{(i,j)}, \lambda_1^{(i,j)} > |\lambda_2^{(i,j)}|$$

holds if and only if $\mathbf{a}_i^H \mathbf{R} \mathbf{a}_i > \mathbf{a}_j^H \mathbf{R} \mathbf{a}_j$

Proof. In order to simplify our notations, we use auxiliary notations $\mathbf{u} := \mathbf{R}^{1/2}\mathbf{a}_i$, $\mathbf{v} := \mathbf{R}^{1/2}\mathbf{a}_j$, and the notations $\lambda_1 := \lambda_1^{(i,j)}, \lambda_2 := \lambda_2^{(i,j)}$.

We can write $\mathbf{Q}_{ij} = \mathbf{u}\mathbf{u}^H - \mathbf{v}\mathbf{v}^H$. We have $\mathrm{tr}(\mathbf{Q}_{ij}) = \|\mathbf{u}\|^2 - \|\mathbf{v}\|^2 = \lambda_1 + \lambda_2$, and $\mathrm{tr}(\mathbf{Q}_{ij}^2) = \|\mathbf{u}\|^4 + \|\mathbf{v}\|^4 - 2|\mathbf{u}^H\mathbf{v}|^2 = \lambda_1^2 + \lambda_2^2$. It follows that $\lambda_1\lambda_2 = |\mathbf{u}^H\mathbf{v}|^2 - \|\mathbf{u}\|^2\|\mathbf{v}\|^2$. Hence, the eigenvalues should satisfy the following equation:

$$\lambda^2 - (\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2)\lambda + |\mathbf{u}^H\mathbf{v}|^2 - \|\mathbf{u}\|^2\|\mathbf{v}\|^2 = 0$$

Therefore, we have

$$\begin{aligned}
\lambda_1^{(i,j)} &= \frac{1}{2}\left(\|\mathbf{u}\|^2 - \|\mathbf{v}\|^2 + \Delta\right) \\
\lambda_2^{(i,j)} &= \frac{1}{2}\left(\|\mathbf{u}\|^2 - \|\mathbf{v}\|^2 - \Delta\right) \\
\Delta &= \sqrt{(\|\mathbf{u}\|^2 - \|\mathbf{v}\|^2)^2 + 4(\|\mathbf{u}\|^2\|\mathbf{v}\|^2 - |\mathbf{u}^H\mathbf{v}|^2)}
\end{aligned}$$

Consequently, we have $\lambda_1^{(i,j)} \geq |\lambda_2^{(i,j)}|$, if and only if $\|\mathbf{u}\| \geq \|\mathbf{v}\|$, which concludes the proof. □

The following Lemma, gives an even simpler equivalent condition, under the setting where the columns of $\mathbf{A}$ are unit-norm:

Lemma 16. *If the columns of* $\mathbf{A}$ *have unit-norm, the condition* (3.47) *is equivalent to*

$$p_i + \sum_{\substack{k=1 \\ k \neq i}}^{K} |\mu_{ik}|^2 p_k > \sum_{k=1}^{K} |\mu_{jk}|^2 p_k \tag{3.48}$$

where $\mu_{ij} = \mathbf{a}_i^H \mathbf{a}_j$.

Proof. The proof follows from simple algebraic manipulation of the condition $\mathbf{a}_i^H \mathbf{R} \mathbf{a}_i > \mathbf{a}_j^H \mathbf{R} \mathbf{a}_j$. □

One can observe that if all the coherence values μ_{ij} are equal to some constant μ, then the inequality (3.48) will automatically hold. However, the property $\mu_{ij} = \mu$ for $i \neq j$ coincides with the definition of equiangular frames:

Definition 9. *If for a matrix* $\mathbf{A} \in \mathbb{F}^{M \times N}$*, it holds that*

- $\|\mathbf{a}_i\| = 1$, *for all* $1 \leq i \leq N$.
- $|\mathbf{a}_i^H \mathbf{a}_j| = \mu$, *for all* $1 \leq i < j \leq N$.

$\mathbf{A}$ *is called an equiangular frame.*

Therefore, we can state the following Lemma:

Lemma 17. *If* $\mathbf{A}$ *is an equiangular frame, then the condition* (3.47) *holds.*

Subsequently, we can state the following theorem:

Theorem 23. *If* $\mathbf{A}$ *is an equiangular frame, it holds that*

$$p_e \leq e^{-\gamma L + \log(K(N-K))}$$

where

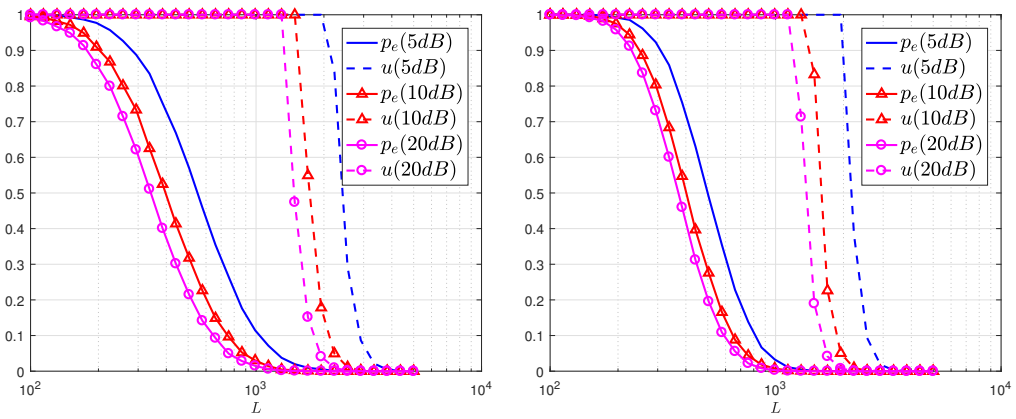
$$\gamma = \min_{i \in \mathcal{S}, j \notin \mathcal{S}} \gamma_{ij}$$

Remark 25. *We can observe that theorem 23 can hold even if* $K > M$*, and sources have unequal powers. In particular, the theorem only relies on* $\mathbf{A}$ *being an equiangular frame, and does not directly require any other conditions on K, N, and M.*

Remark 26. *Theorem 23 states that as long as an equiangular frame of size* $M \times N$ *exists, it is possible to do support recovery for any* $K \leq N$*. Therefore, possibility of recovering supports of size* $K = O(M^2)$ *is subject to existence of equiangular frames with dimensions* $N = O(M^2)$*. This turns out to be an open problem, and the definite answer for a general M is not known [FM15].*

3.3.3 Simulations

In this section, we provide numerical simulations showing the performance of the proposed greedy support detector. In our simulations, we let $\mathbf{A}$ to be an equiangular tight frame, for certain M, N. We let the source powers corresponding to the nonzero elements of $\mathbf{p}$ to be chosen i.i.d. from uniform distribution in range $[p_{\min}, p_{\max}]$. Figure 3.6 shows the empirical probability of error (averaged over 10000 runs) as well as the theoretical upper bound derived in Theorem 23, with respect to L, for different signal-to-noise ratios (SNR).



(a) $M = 63, N = 126, K = 100$ (b) $M = 122, N = 244, K = 150$

Figure 3.6: Empirical probability of error for the proposed support detector (labeled as "p_e"), and the upperbound given in Theorem 23 (labeled as "u"). The plots are shown for different values of SNR. We have $p_{\min} = 0.1, p_{\max} = 1.1$. In both cases, the matrix $\mathbf{A}$ is an equiangular tight frame constructed by Paley's conference matrix [Ren07].

3.3.4 Conclusion

In this section, we considered the problem of sparse support recovery and proposed a simple greedy support detection approach. We provided upper bounds for the probability of error of this detection approach and showed that when the measurement matrix is an *equiangular frame*, support recovery with exponentially decaying probability of error in L is possible. Moreover, our guarantees can potentially hold even if $K > M$, or even when the sources have unequal powers. We also concluded that if equiangular frames of size $N = O(M^2)$ exist (which is an open problem) then support recovery for $K = O(M^2)$ is also possible.

3.4 A Sequential Approach for Sparse Support Recovery using Correlation Priors

In Sections 3.1, 3.2, and 3.3 we considered batch support recovery setting where the measurement matrix is assumed to be fixed. In this section, we consider an adaptive setting where the goal is to improve the signal-to-noise (SNR) ratio of the measurements by adaptively designing the measurement matrix (similar to [MN14, HBCN12, ST06]), as more and more measurements are collected. Adaptive algorithms also have a lower computational cost in most cases, as these algorithms are able to identify the support of the sparse vector as soon as *enough* number of measurements are collected [MN14], while the batch algorithms may require solving a convex optimization problem which can be computationally costly [Bar07, MN14].

However, it has been shown that in both batch and adaptive support recovery algorithms, which use only a single measurement vector (SMV), having $M = O(k \log N)$ measurements is necessary and sufficient [AAS17, Wai09a] to successfully identify the support, where M is the size of the measurement vector, and k, N are the sparsity and the size of the unknown sparse vector, respectively. This means that when only a single measurement vector is available, identifying supports of size $k > M$ is not possible. However, in presence of multiple measurement vectors (MMV), where the sparse signals are assumed to have a common support, under certain additional statistical priors on the measurement vectors, namely the nonzero elements of sparse vectors being generated by uncorrelated sources, in Sections 3.1, 3.2, and 3.3 we showed that in the batch setting, it is possible to recover supports of size $k > M$, and even $k = O(M^2)$ provided that the measurement matrix satisfies certain algebraic properties [KQP18, KP18d, KP18b].

In this Section, we will consider the problem of sparse support recovery in presence of multiple measurements where we also have ability to adaptively design the measurement matrix as groups of multiple measurement vectors are collected. This problem can have applications in channel identification for mmWave communication systems. By providing an adaptive algorithm,

we will show that using certain statistical priors on the sparse signals, namely assuming they are spatially and temporally uncorrelated, it is possible to recover the sparsity pattern using $M = O(\sqrt{k}\log N)$ measurements. This implies that recovering supports of size $k = O(M^2)$ is practical in the adaptive setting which matches with the already known result for the batch setting, provided in Section 3.1. We will also establish non-asymptotic guarantees for our proposed algorithm indicating that the probability of detecting a wrong support decays to zero, exponentially fast, as the number of measurement vectors collected in each block increases.

3.4.1 Problem Formulation

We slightly modify the batch problem setting considered in Sections 3.1, 3.2, and 3.3, and instead assume an adaptive setting. Let us assume that at the t-th iteration of the adaptive sensing scheme, we are able to observe L independent measurement vectors, where each vector has size M_t and is obtained via the following model

$$\mathbf{Y}_t = \mathbf{A}_t \mathbf{X}_t + \mathbf{Z}_t$$

where $\mathbf{Y}_t \in \mathbb{R}^{M_t \times L}$, $\mathbf{X}_t = [\mathbf{x}_1^{(t)}, \cdots, \mathbf{x}_L^{(t)}] \in \mathbb{R}^{N \times L}$, and $\mathbf{Z}_t = [\mathbf{z}_1^{(t)}, \cdots, \mathbf{z}_L^{(t)}]$ is the additive noise. The vectors $\{\mathbf{x}_l^{(t)}\}_{t=1,l=1}^{T,L}$ are assumed to have the a shared support $\mathcal{S}$, i.e.,

$$i \notin \mathcal{S} \Rightarrow [\mathbf{x}_l^{(t)}]_i = 0$$

The measurement matrix $\mathbf{A}_t \in \mathbb{R}^{M_t \times N}$ can be designed based on the previously collected measurements $\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_{t-1}$. Our goal is to reliably identify the support set $\mathcal{S}$ given all the measurements $\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_T$, where T is a stopping time where the algorithm decides the collected measurements are sufficient to make a reliable decision. Notice that over the course of the adaptive sensing scheme, we will collect $M = \sum_{t=1}^{T} M_t$ row vector measurements of size $1 \times L$.

We want to use as few measurements as possible to be able to identify the support set $\mathcal{S}$. In order to do so, we make fairly moderate statistical assumptions on the signals $\mathbf{X}_t$ and noise $\mathbf{W}_t$ which are practical in many settings, and will later use these assumptions to design our algorithm:

(A1) Non-zero elements of the signal $\mathbf{x}_l^{(t)}$ are uncorrelated, i.e. $\mathbb{E}(\mathbf{x}^{(t)}(\mathbf{x}^{(t)})^T) = \mathbf{P}$, where $\mathbf{P} = \text{diag}(p_1, \cdots, p_N)$ is a diagonal matrix, and $\{\mathbf{x}_l^{(t)}\}_{l=1,t=1}^{L,T}$ are independent and identically distributed (i.i.d.) zero mean random vectors. Moreover $p_i = 0, i \neq \mathcal{S}$, consistent with the fact that the L vectors $\{\mathbf{x}_l^{(t)}\}_{l=1,t=1}^{L,T}$ share a common support $\mathcal{S}$.

(A2) Signal $\mathbf{x}_l^{(t)}$ and noise $\mathbf{x}_l^{(t)}$ are uncorrelated, i.e, $\mathbb{E}(\mathbf{x}_l^{(t)}(\mathbf{w}_l^{(t)})^T) = \mathbf{0}$.

(A3) The noise $\mathbf{w}$ is white, i.e. $\mathbb{E}(\mathbf{w}_l^{(t)}(\mathbf{w}_l^{(t)})^T) = \sigma^2\mathbf{I}$, and $\{\mathbf{w}_l^{(t)}\}_{l=1,t=1}^{L,T}$ are i.i.d. random zero mean random variables. We assume σ^2 is known.

(A4) The signal and noise are real-valued zero mean Gaussian random variables, i.e., for every $t = 1, \cdots, T$, $l = 1, \cdots, L$ we have

$$\mathbf{x}_l^{(t)} \sim \mathcal{N}(\mathbf{0}, \mathbf{P})$$
$$\mathbf{w}_l^{(t)} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_{M_t})$$

where $\mathbf{I}_{M_t}$ is an identity matrix of size $M_t \times M_t$.

3.4.2 Proposed Algorithm

Let $\mathbf{y}_{(i)} \in \mathbb{R}^{1\times L}$ be the ith row of the matrix of measurements $\mathbf{Y}_t$ (where we suppressed the dependence on t in the notation to avoid notational overhead). We have

$$\mathbf{y}_{(i)} = \mathbf{a}_{(i)}\mathbf{X}_t + \mathbf{w}_{(i)}$$

where $\mathbf{a}_{(i)} \in \mathbb{R}^{1\times N}$ is the ith row of the measurement matrix, and $\mathbf{w}_{(i)}$ is the ith row of $\mathbf{W}_t$. Now, let $\mathbf{y}_{(j)}$ to be another row of $\mathbf{Y}_t$, where $i \neq j$. Using the assumptions **(A1)**, **(A2)**, **(A3)** it holds that

$$\begin{aligned}\rho_{ij} := \mathbb{E}(\mathbf{y}_{(i)}\mathbf{y}_{(j)}^T) &= \mathbf{a}_{(i)}\mathbf{P}\mathbf{a}_{(j)}^T \\ &= (\mathbf{a}_{(i)} \circ \mathbf{a}_{(j)})\mathbf{p} \in \mathbb{R}\end{aligned} \tag{3.49}$$

where $\circ$ represents elements-wise Hadamard product.

For simplicity, we assume that the adaptive sensing scheme designs the measurement matrix $\mathbf{A}_t$ using only ones and zeros. More general cases can be considered in future. Let $\mathbf{r} \in \mathbb{N}^{N_r}$ be a vector of integers whose elements belong to the set $\{1,2,\cdots,N\}$. Let the rth row of $\mathbf{A}_t$ be such that $\mathbf{a}_{(r)} = \mathbf{1}_\mathbf{r} \in \{0,1\}^{1\times N}$ where the n-th element of $\mathbf{1}_\mathbf{r}$ equals

$$[\mathbf{1}_\mathbf{r}]_n = \begin{cases} 1 & n \in \mathbf{r} \\ 0 & n \notin \mathbf{r} \end{cases}$$

where for notational simplicity we have treated $\mathbf{r}$ as a set composed of the elements of $\mathbf{r}$. Similarly, define $\mathbf{c} \in \mathbb{N}^{N_c}$ and correspondingly the cth row of $\mathbf{A}_t$ to be $\mathbf{a}_{(c)} = \mathbf{1}_\mathbf{c}$. Based on these assumptions, we have

$$\mathbf{a}_{(r)} \circ \mathbf{a}_{(c)} = \mathbf{1}_{\mathbf{r}\cap\mathbf{c}}$$

where the operator $\cap$ treats the vectors $\mathbf{r},\mathbf{c}$ as sets and finds their intersection. Using the fact that the nonzero elements of $\mathbf{p}$ are all positive, we can easily verify that

$$\rho_{rc} = 0 \Leftrightarrow p_k = 0 \ \ \forall k \in \mathbf{r}\cap\mathbf{c}.$$

This property can help us to design an adaptive algorithm to identify the support set $\mathcal{S}$ using the statistical assumptions **(A1)**-**(A4)**, and carefully designing the measurement matrices $\mathbf{A}_t$. In

practice, given the finite block length L, it is not possible to compute ρ_{ij} accurately. Instead, we can estimate them using sample covariance formulation:

$$\hat{\rho}_{ij}^{(t)} = \frac{1}{L}\sum_{l=1}^{L}[\mathbf{Y}_t]_{il}[\mathbf{Y}_t]_{jl} \tag{3.50}$$

where we denoted the dependence on t for clarity of notation.

We are now equipped to propose our adaptive support recovery algorithm. Let $\mathcal{W}_t' \subseteq \{1,2,\cdots,N\}$ be the set of indices for which the adaptive algorithm, up to iteration t, has not made a decision whether they belong to the support $\mathcal{S}$ or not, and $\hat{\mathcal{S}}$ be the set of of support detected by the algorithm. Clearly, $\mathcal{W}_1' = \{1,2,\cdots,N\}$, and we have $\hat{\mathcal{S}} = \{\}$ initially. In order to design the measurement matrix $\mathbf{A}_t$, we use two different schemes to divide the sets $\mathcal{W}_t'$ into at most 2^i mutually exclusive collectively exhaustive sets, where i indicates the bisection *level* of our bisection schemes, and initially we set $i = 1$. The bisection schemes are as follows:

1. For $\mathbf{r} \in \mathbb{N}^{N_r}$ define the operator:

$$\mathbf{r}^{(1)}, \mathbf{r}^{(2)} \leftarrow \text{bisect}^{(\text{range})}(\mathbf{r})$$

such that if $\mathbf{r} = [r_1, r_2, \cdots, r_{N_r}]^T$, then

$$\mathbf{r}^{(1)} = \mathbf{r}(1 : \lfloor \frac{N_r}{2} \rfloor) \tag{3.51}$$

$$\mathbf{r}^{(2)} = \mathbf{r}(\lfloor \frac{N_r}{2} \rfloor + 1 : N_r) \tag{3.52}$$

2. For $\mathbf{c} \in \mathbb{N}^{N_c}$ define the operator:

$$\mathbf{c}^{(1)}, \mathbf{c}^{(2)} \leftarrow \text{bisect}^{(\text{even-odd})}(\mathbf{c})$$

such that if $\mathbf{c} = [c_1, c_2, \cdots, c_{N_c}]^T$, then

$$\mathbf{c}^{(1)} = \mathbf{c}(1:2:\text{end}) \tag{3.53}$$

$$\mathbf{c}^{(2)} = \mathbf{c}(2:2:\text{end}) \tag{3.54}$$

where we have used the MATLAB indexing in equations (3.51), (3.52), (3.53), and (3.54). Now for each vector $\mathbf{r}$, we define the rows of the rth row of the measurement matrix to be $\mathbf{a}_{(r)} = \mathbf{1}_{\mathcal{W}_t'(\mathbf{r})}$. Here, the notation $\mathcal{W}_t'(\mathbf{r})$ stands for a subset of $\mathcal{W}_t'$ indexed by the elements of $\mathbf{r}$, i.e., $\mathcal{W}_t'(\mathbf{r}) = \{w_{r_1}, w_{r_2}, \cdots, w_{r_{N_r}}\}$, where $\mathcal{W}_t' = \{w_1, \cdots, w_{|\mathcal{W}_t'|}\}$ is assumed to be an ordered set. We make similar definitions for $\mathcal{W}_t'(\mathbf{c})$, $\mathbf{a}_{(c)}$ for any vector $\mathbf{c} \in \mathbb{N}^{N_c}$.

To clarify these notations and definitions, we provide the following example:

Example 2. *Let $N = 8$, and consider the set $\mathcal{W}_1 = \{1,2,3,4,5,6,7,8\}$. After performing the two bisection schemes, we have $\mathcal{R} = \{\mathbf{r}_1, \mathbf{r}_2\} = \{[1,2,3,4],[5,6,7,8]\}$, $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2\} = \{[1,3,5,7],[2,4,6,7]\}$ such that*

$$\mathbf{a}_{(r_1)} = \mathbf{1}_{\mathcal{W}_1(\mathbf{r}_1)} = [1,1,1,1,0,0,0,0]$$

$$\mathbf{a}_{(r_2)} = \mathbf{1}_{\mathcal{W}_1(\mathbf{r}_2)} = [0,0,0,0,1,1,1,1]$$

$$\mathbf{a}_{(c_1)} = \mathbf{1}_{\mathcal{W}_1(\mathbf{c}_1)} = [1,0,1,0,1,0,1,0]$$

$$\mathbf{a}_{(c_2)} = \mathbf{1}_{\mathcal{W}_1(\mathbf{c}_2)} = [0,1,0,1,0,1,0,1]$$

where $\mathbf{a}_{(r_1)}, \mathbf{a}_{(r_2)}, \mathbf{a}_{(c_1)}, \mathbf{a}_{(c_2)}$ are rows of the measurement matrix $\mathbf{A}_1 \in \mathbb{R}^{4\times 8}$. We can assign row indices $r_1 = 1, r_2 = 2, c_1 = 3, c_2 = 4$. We can also observe that

$$\mathbf{a}_{(r_1)} \circ \mathbf{a}_{(c_1)} = \mathbf{1}_{\mathbf{r}_1 \cap \mathbf{c}_1} = [1,0,1,0,0,0,0,0]$$

$$\mathbf{a}_{(r_2)} \circ \mathbf{a}_{(c_1)} = \mathbf{1}_{\mathbf{r}_2 \cap \mathbf{c}_1} = [0,0,0,0,1,0,1,0]$$

If we have $\mathbf{p} = [1,0,0,0,0,1,0,0]^T$*, it follows that* $(\mathbf{a}_{(r_1)} \circ \mathbf{a}_{(c_1)})\mathbf{p} = 1$*, while* $(\mathbf{a}_{(r_2)} \circ \mathbf{a}_{(c_1)})\mathbf{p} = 0$*. From the latter equation we can directly infer that* $p_5 = p_7 = 0$*, only using the fact that the elements of* $\mathbf{p}$ *are nonnegative. We will see how these intuitions can help us to develop our adaptive support recovery algorithm. Notice that if we were to use the original rows of* $\mathbf{A}_1$*, i.e.,* $\mathbf{a}_{(x)}\mathbf{p}$*, where* $\mathbf{x} = 1,2,3,4$*, instead of their pair-wise Hadamard products, none of them would help us make a decision about any of the elements of* $\mathbf{p}$*, since all* $\mathbf{a}_{(x)}\mathbf{p}$ *are nonzero.*

As we observed in the above example, simply computing the Hadamard product terms (such as $\mathbf{a}_{(r_1)} \circ \mathbf{a}_{(c_1)}$), which arises naturally from the cross-correlation terms in (3.49), we can infer more accurate information about support of $\mathbf{p}$, than what we would have obtained by linear measurements from the original rows of $\mathbf{A}_1$. However, if in the above example we had $\mathbf{p} = [1,1,0,0,1,1,0,0]^T$ even any of the terms $\mathbf{a}_{(r)} \circ \mathbf{a}_{(c)}$ with $r \in \{1,2\}$, $c \in \{3,4\}$ would not be enough to help us to infer any information about the support of $\mathbf{p}$ as $(\mathbf{a}_{(r)} \circ \mathbf{a}_{(c)})\mathbf{p}$ would have been nonzero for all combinations of r, c.

In this case, we would need to increase the level of bisection of the original measurement matrix $\mathbf{A}_1$. This can be done by performing the bisection operations recursively, and obtaining the sets $\mathcal{R} = \{[1,2],[3,4],[5,6],[7,8]\}$, and $\mathcal{C} = \{[1,5],[3,7],[2,6],[4,8]\}$. Using these new sets, we can design a new measurement matrix $\mathbf{A}_2 \in \mathbb{R}^{8\times 8}$ which can accurately infer the support of $\mathbf{p}$ by considering the terms $\mathbf{a}_{(r)} \circ \mathbf{a}_{(c)}$.

Our adaptive support recovery algorithm is presented in Algorithm 1. We start with $i = 1$, which means the number of levels of bisection equals one. At each iteration, we compute the sets $\mathcal{R}, \mathcal{C}$ based on the levels of bisections needed at that particular iteration. Using these sets, we can design the rows of the measurement matrix $\mathbf{A}_t$, and take the measurements. Once the measurements are obtained, we estimate the cross-correlation terms $\hat{\rho}_{rc}^{(t)}$ using (3.50). If the estimated $\hat{\rho}_{rc}^{(t)}$ is close to zero, we can infer that $p_i = 0$ for all $i \in \mathbf{r} \cap \mathbf{c}$. Otherwise, some elements of $\mathbf{r} \cap \mathbf{c}$ can belong to the support $\mathcal{S}$. We can only be sure that which element belongs to $\mathcal{S}$ if $|\mathbf{r} \cap \mathbf{c}| = 1$. Once we make decisions on the elements of $\mathcal{W}_t$, we can remove those elements from

$\mathcal{W}_t'$. This is done by the operation $\mathcal{W}_t'(\mathbf{r} \cap \mathbf{c}) \leftarrow \{\}$ in Algorithm 1, which removes all the elements of $\mathcal{W}_t'$ indexed by the elements in the set $\mathbf{r} \cap \mathbf{c}$. If at iteration t, we have made enough number of decisions so that $|\mathcal{W}_{t+1}'| \leq |\mathcal{W}_t'|/2$, we do not need to increase the level of bisection i. However, if the observations that we made were not informative enough, we need to increase i and so that we can obtain more informative measurements in the next iterations. The algorithm stops when all the elements of the support are discovered, or the set $\mathcal{W}_t'$ becomes empty.

The following theorem establishes the maximum number of iterations needed to detect the support:

Theorem 24. *The total number of measurements required by Algorithm 1 is at most* $C\sqrt{k}(\log_2 N)$*, where* C *is a universal constant.*

Proof. The proof is provided in Section 3.5.4. □

In practice, due to the finite block length L, the support detection algorithm may make erroneous decisions. The following theorem establishes a non-asymptotic bound on the probability with which the proposed support recovery algorithm detects a wrong support. Our bound shows that as $L \to \infty$, this probability of error goes to zero exponentially fast.

Theorem 25. *Under assumptions* ***(A1)-(A5)****, probability with which Algorithm 1 detects a wrong support, i.e.,* $\mathbb{P}(\hat{\mathcal{S}} \neq \mathcal{S})$*, is upper bounded by*

$$1 - C_1 k \log N \exp\left(-\min\left\{ \frac{9L\varepsilon^2}{256C_2^2(2p_{\max}^2 k^2 + \sigma^4)}, \frac{3L\varepsilon}{16C_2\sqrt{2(2p_{\max}^2 k^2 + \sigma^4)}} \right\} \right)$$

where $C_1, C_2 > 0$ *are universal constants, and* $p_{\max}$ *is the maximum element of the* $\mathbf{p}$ *vector.*

Proof. The proof is provided in Section 3.5.5. □

Algorithm 1 Correlation-aware Adaptive Support Recovery

$\mathcal{W}_1 \leftarrow \{1, \cdots, N\}$
$\hat{\mathcal{S}} \leftarrow \{\}$
$i \leftarrow 1, \quad t \leftarrow 1$
while $|\hat{\mathcal{S}}| < k$ **and** $\mathcal{W}_t \neq \emptyset$ **do**
  $\mathcal{R} \leftarrow \{[1, 2, \cdots, |\mathcal{W}_t|]\}$
  $\mathcal{C} \leftarrow \{[1, 2, \cdots, |\mathcal{W}_t|]\}$
  for $j \in \{1, 2, \cdots, i\}$ **do**
    for $\mathbf{r} \in \mathcal{R}$ **do**
      Replace $\mathbf{r}$ in $\mathcal{R}$ with $\mathbf{r}^{(1)}, \mathbf{r}^{(2)} \leftarrow \text{bisect}^{(\text{range})}(\mathbf{r})$
    end for
    for $\mathbf{c} \in \mathcal{C}$ **do**
      Replace $\mathbf{c}$ in $\mathcal{C}$ with $\mathbf{c}^{(1)}, \mathbf{c}^{(2)} \leftarrow \text{bisect}^{(\text{even-odd})}(\mathbf{c})$
    end for
  end for
  Remove empty vectors from sets $\mathcal{R}, \mathcal{C}$.
  for $\mathbf{r} \in \mathcal{R}$: Let $\mathbf{a}_{(r)} = \mathbf{1}_{\mathcal{W}_t(\mathbf{r})}$.
  for $\mathbf{c} \in \mathcal{R}$: Let $\mathbf{a}_{(c)} = \mathbf{1}_{\mathcal{W}_t(\mathbf{c})}$.
  Take measurements $\mathbf{Y}_t = \mathbf{A}_t \mathbf{X}_t + \mathbf{Z}_t$.
  $\mathcal{W}_{t+1} \leftarrow \mathcal{W}_t$
  for $\mathbf{r} \in \mathcal{R}$ **do**
    for $\mathbf{c} \in \mathcal{C}$ **do**
      Compute $\hat{\rho}_{rc}^{(t)}$ using (3.50).
      if $|\hat{\rho}_{rc}^{(t)}| < \varepsilon$ **then**
        $\mathcal{W}_{t+1}(\mathbf{r} \cap \mathbf{c}) \leftarrow \{\}$
      else if $|\mathbf{r} \cap \mathbf{c}| = 1$ **then**
        $\hat{S} \leftarrow \hat{S} \cup \mathcal{W}_t(\mathbf{r} \cap \mathbf{c})$
        $\mathcal{W}_{t+1}(\mathbf{r} \cap \mathbf{c}) \leftarrow \{\}$
      end if
    end for
  end for
  if $|\mathcal{W}_{t+1}| > |\mathcal{W}_t|/2$ **then**
    $i \leftarrow i + 1$
  end if
  $t \leftarrow t + 1$
end while

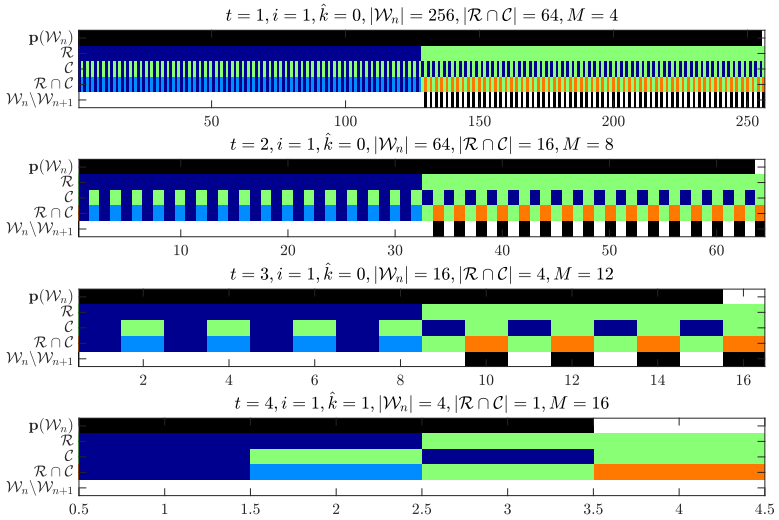
3.4.3 Simulations

In this section, we provide our numerical simulations to examine the performance of our proposed adaptive support recovery algorithm.

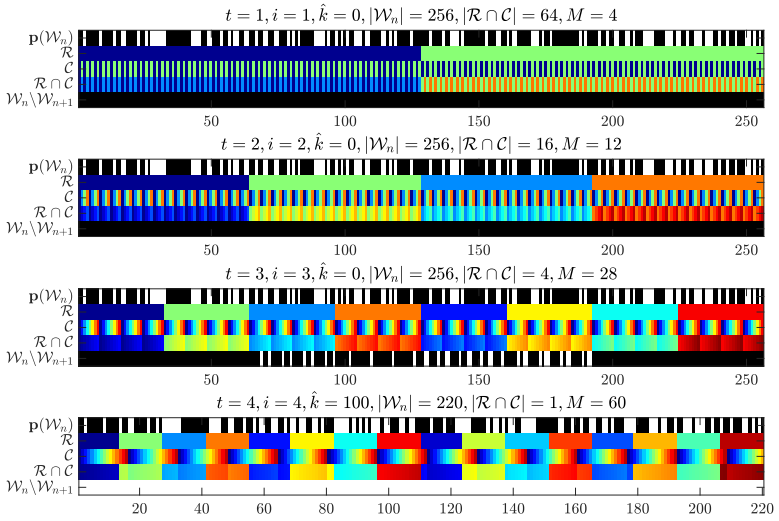
In the first experiment, we demonstrate the iterations of our algorithm and show how the sets $\mathcal{R}, \mathcal{C}, \mathcal{W}_t, \hat{\mathcal{S}}$ evolve as the algorithm goes through its iterations. In Figure 3.7 we show the iterations of our adaptive algorithm in a noiseless setting, for the cases $k = 1, k = 100$. The sets in $\mathcal{R}, \mathcal{C}$ are showed in a color-coded format, i.e., the points corresponding to the same color belong to the same set. For example, the first row of Figure 3.7a shows how the two bisection schemes $\text{bisect}^{(\text{range})}(\mathbf{r})$, $\text{bisect}^{(\text{even-odd})}(\mathbf{c})$ divide the index sets in $\mathcal{W}_1$ into two sets. Further levels of bisections are also shown in later iterations as more colors are used, indicating more sets in $\mathcal{R}, \mathcal{C}$. The row $\mathcal{W}_t \backslash \mathcal{W}_{t+1}$ shows that which elements of $\mathcal{W}_t$ will be removed for the next iteration of the algorithm.

In Figure 3.7a, where $k = 1$, at each iteration the adaptive algorithm removes roughly 3/4th of the elements of $\mathcal{W}_t$, i.e., $\mathcal{W}_{t+1} \leq \mathcal{W}_t/4$. Hence, i remains to be equal to one. In the second example, depicted in Figure 3.7b, we have $k = 100$. The algorithm initially is not able to make any decisions on the elements of $\mathcal{W}_1$. Therefore, the size of $|\mathcal{W}_t|$ remains the same upto iteration 3, while at each iteration the algorithm increases the level of bisection to increase the granularity of the measurement matrix. At iteration $t = 5$, the algorithm is able to make some decisions on the elements of $\mathcal{W}_t$. Finally, at 7th iteration, the algorithm is able to detect all the elements of the support. It is worth noting that the overall number of measurements collected in Figure 3.7b is $M = 60$ which is less than the sparsity $k = 100$.

In the second set of experiments, we perform Monte-Carlo simulations to study the probability of error with respect to the number of measurements L in each block. We perform 20000 simulations for each case, and compute the empirical probability of detecting a wrong support. Figure 3.8 shows this probability. We let $N = 640$, the sources to have equal power $p = 1$, the noise variance to be $\sigma^2 = 0.01$, and we choose $\varepsilon = 0.45L$ for all the simulations. As



(a) $k = 1$



(b) $k = 100$

Figure 3.7: The sets $\mathcal{R}, \mathcal{C}, \hat{\mathcal{S}}$ as Algorithm 1 goes through its iterations.

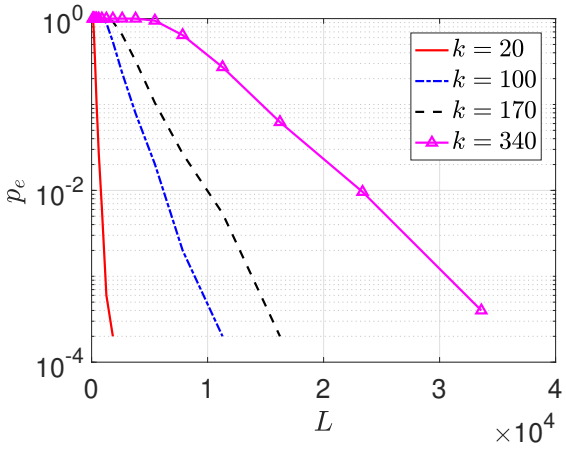


Figure 3.8: Probability of detecting a wrong support p_e with respect to the number of measurement vectors in each block L.

we can observe, the plots show a linear behavior in logarithmic scale which indicates exponential decay with respect to L.

Moreover, the probability of error decays faster for smaller k. Studying the behavior of the error exponent, i.e., the slope of lines in Figure 3.8 with respect to k, M, N can be a topic of future research.

3.4.4 Conclusion

In this section, we considered the problem of adaptive support recovery of jointly sparse signals in presence of certain statistical priors on the sparse signals. We proposed an adaptive algorithm, with low computational complexity, which is able to recover supports of size k with at most $O(\sqrt{k}\log N)$ measurements, where N is the size of the sparse signals, under certain practical statistical assumptions. We also established that the probability of detecting a wrong support using the proposed adaptive algorithm goes to zero exponentially fast as more and more measurement vectors are collected.

3.5 Appendix

3.5.1 Proof of Corollary 4

Preliminary Lemma

Before proving Corollary 4, we state the following lemma:

Lemma 18. *Suppose* $\mathbf{X} \in \mathbb{R}^{d\times n}$ $(n > d)$ *has full Kruskal rank, i.e.,* k-rank$(\mathbf{X}) = d$. *If there exists a vector* $\mathbf{p} \in \mathbb{R}^n_+$ *satisfying* $\mathbf{Xp} = \mathbf{0}$, *then there exists a vector* $\hat{\mathbf{p}} \in \mathbb{R}^n_{++}$ *(i.e.,* $\hat{\mathbf{p}}$ *has strictly positive entries) such that* $\mathbf{X}\hat{\mathbf{p}} = \mathbf{0}$.

Proof. We prove that $\hat{\mathbf{p}}$ exists via induction. In particular, we show that if there exists a vector $\mathbf{p}^{(m)} \in \mathbb{R}^n_+$ satisfying $\mathbf{Xp}^{(m)} = \mathbf{0}$, one can find $\mathbf{p}^{(m+1)} \in \mathbb{R}^n_+$ such that $\|\mathbf{p}^{(m+1)}\|_0 = \|\mathbf{p}^{(m)}\|_0 + 1$, and $\mathbf{Xp}^{(m+1)} = \mathbf{0}$. By induction, this concludes that $\hat{\mathbf{p}}$ exists such that $\|\hat{\mathbf{p}}\|_0 = n$, and $\mathbf{X}\hat{\mathbf{p}} = \mathbf{0}$.

We know that there exists a $\mathbf{p} \in \mathbb{R}^n_+$, with $\mathbf{Xp} = \mathbf{0}$, and we make this $\mathbf{p}$ our basis for induction, i.e, $\mathbf{p}^{(0)} = \mathbf{p}$.

Now let us assume that there exists $\mathbf{p}^{(m)} \in \mathbb{R}^n_+$ with $s^{(m)}$ nonzeros, i.e., $\|\mathbf{p}^{(m)}\|_0 = s^{(m)}$, and $\mathbf{Xp}^{(m)} = \mathbf{0}$. Let $i_1, \cdots, i_{s^{(m)}}$ denote the indices nonzero entries of $\mathbf{p}^{(m)}$. Since k-rank$(\mathbf{X}) = d$, we have $s^{(m)} \geq d+1$ (otherwise one can find d linearly dependent columns in $\mathbf{X}$, which contradicts the fact that k-rank$(\mathbf{X}) = d$). Moreover, $\mathbf{X}$ being full Kruskal rank also implies that $\mathbf{x}_{i_1}, \cdots, \mathbf{x}_{i_{s^{(m)}}}$ contains d linearly independent vectors. Hence, any column $\mathbf{x}_q$ ($q \notin \{i_1, \cdots, i_{s^{(m)}}\}$) is linearly dependent on $\mathbf{x}_{i_1}, \cdots, \mathbf{x}_{i_{s^{(m)}}}$. This implies that there exists $\mathbf{c} \in \mathbb{R}^{s^{(m)}}$ such that

$$\sum_{k=1}^{s^{(m)}} c_k \mathbf{x}_{i_k} = \mathbf{x}_q \tag{3.55}$$

Moreover, since $\mathbf{Xp}^{(m)} = \mathbf{0}$, we have

$$\sum_{k=1}^{s^{(m)}} p^{(m)}_{i_k} \mathbf{x}_{i_k} = \mathbf{0} \tag{3.56}$$

Combining (3.55) and (3.56), for any $\alpha \in \mathbb{R}$, we get

$$\sum_{k=1}^{s^{(m)}} (p_{i_k}^{(m)} - \alpha c_k)\mathbf{x}_{i_k} + \alpha \mathbf{x}_q = \mathbf{0}. \tag{3.57}$$

We define a new vector $\mathbf{p}^{(m+1)} \in \mathbb{R}^n$ with entries

$$\begin{cases} p_{i_k}^{(m+1)} = p_{i_k}^{(m)} - \alpha c_k, & 1 \leq k \leq s^{(m)}, \\ p_q^{(m+1)} = \alpha, & \\ p_i^{(m+1)} = 0, & i \notin \{i_1, \cdots, i_{s^{(m)}}, q\}. \end{cases}$$

From (3.57), we have that $\mathbf{X}\mathbf{p}^{(m+1)} = \mathbf{0}$. Let $p_{\min}^{(m)} := \min_{1 \leq k \leq s^{(m)}} p_{i_k}^{(m)}$ and $\|\mathbf{c}\|_1 := \sum_{i=1}^{s^{(m)}} |c_i|$. One can easily check that choosing $\alpha = p_{\min}^{(m)} / \|\mathbf{c}\|_1 > 0$ will ensure that $p_{i_k}^{(m+1)} > 0$ for all $k = 1, 2, \cdots, s^{(m)}$. Therefore, $\mathbf{p}^{(m+1)}$ has positive entries at the indices $i_1, \cdots, i_{s^{(m)}}, q$ and hence $\|\mathbf{p}^{(m+1)}\|_0 = s^{(m)} + 1$, which concludes the proof. □

Proof of Corollary 4

We are now ready to state the proof of Corollary 4:

Proof. Define the following sets

$$G = \{\mathbf{X} \in \mathbb{R}^{d \times n} : \text{k-rank}(\mathbf{X}) = d\}$$

$$N^{++} = \{\mathbf{X} \in \mathbb{R}^{d \times n} : \exists \mathbf{p} \in \mathbb{R}^n_{++}, \mathbf{X}\mathbf{p} = \mathbf{0}\}$$

$$N^{+} = \{\mathbf{X} \in \mathbb{R}^{d \times n} : \exists \mathbf{p} \in \mathbb{R}^n_{+}, \mathbf{X}\mathbf{p} = \mathbf{0}\}$$

By definition of the sets G, N^+ and N^{++}, it can be readily seen that

$$\mathbb{P}(\mathbf{X} \in G \cap N^{++}) = \mathbb{P}(\exists \mathbf{p} \in \mathbb{R}^n_{++}, \mathbf{X}\mathbf{p} = \mathbf{0}),$$

and

$$\mathbb{P}(\mathbf{X} \in G \cap N^{+}) = \mathbb{P}(\exists \mathbf{p} \in \mathbb{R}^{n}_{+}, \mathbf{X}\mathbf{p} = \mathbf{0}).$$

From Wendel's theorem, we have

$$\mathbb{P}(\exists \mathbf{p} \in \mathbb{R}^{n}_{+}, \mathbf{X}\mathbf{p} = \mathbf{0}) = 1 - \frac{1}{2^{n-1}} \sum_{k=0}^{d-1} \binom{n-1}{k}.$$

We will now show that the sets $G \cap N^{+}$, and $G \cap N^{++}$ are equal, thereby concluding the proof. From definitions, it is clear that $N^{++} \subseteq N^{+}$. Hence, $N^{++} \cap G \subseteq N^{+} \cap G$. Conversely, consider any $\mathbf{X} \in G \cap N^{+}$. For such matrix $\mathbf{X}$, there exists $\mathbf{p} \in \mathbb{R}^{n}_{+}$ such that $\mathbf{X}\mathbf{p} = \mathbf{0}$, and k-rank$(\mathbf{X}) = d$. By Lemma 18, we know that for this particular matrix $\mathbf{X}$, there also exists a $\hat{\mathbf{p}} \in \mathbb{R}^{n}_{++}$ such that $\mathbf{X}\hat{\mathbf{p}} = \mathbf{0}$. Therefore, $\mathbf{X} \in G \cap N^{++}$. Hence, $G \cap N^{+} \subseteq G \cap N^{++}$, implying $G \cap N^{+} = G \cap N^{++}$. $\square$

3.5.2 Proof of Theorem 21

Preliminary Lemmas and Definitions

Before presenting the proof of theorem 21, we state some preliminary definitions and lemmas.

Definition 10. *For any $\mathbf{a} \in \mathbb{R}^{M}$, define $\mathbf{a}^{\otimes 2} := \mathbf{a} \otimes \mathbf{a}$. Also, let $\overline{\mathbf{a} \otimes \mathbf{a}}$ (or equivalently $\overline{\mathbf{a}^{\otimes 2}}$) be a $\overline{M} := \frac{M^2+M}{2}$ dimensional vector which contains the upper triangular elements (including the diagonal) of the matrix $\mathbf{a}\mathbf{a}^{T}$. Moreover, for any $\mathbf{A} \in \mathbb{R}^{M \times N}$, define the $\overline{M} \times N$ dimensional matrix*

$$\overline{\mathbf{A} \odot \mathbf{A}} = [\overline{\mathbf{a}_1^{\otimes 2}} | \overline{\mathbf{a}_2^{\otimes 2}} | \cdots | \overline{\mathbf{a}_N^{\otimes 2}}].$$

We now state two lemmas which will be useful in our proof.

Lemma 19. *Suppose* $\mathbf{A},\mathbf{B} \in \mathbb{R}^{M\times n}$, $n \geq \overline{M} := \frac{M^2+M}{2}$ *be random matrices whose elements are drawn independently from continuous distributions over* $\mathbb{R}$. *Let* $\overline{\mathbf{C}} = \overline{\mathbf{A}\odot\mathbf{A}} - \overline{\mathbf{B}\odot\mathbf{B}}$. *With probability one, it holds that* k-rank$(\overline{\mathbf{C}}) = \overline{M}$.

Proof. We follow a proof technique similar to that of Theorem 4 in [PV15]. In order to show that $\overline{\mathbf{C}}$ has full Kruskal rank with probability one, we show that every collection of $\overline{M}$ columns of $\overline{\mathbf{C}}$ has full column rank with probability one over $\mathbb{R}^{M\overline{M}}$. By the application of union bound over all possible $\binom{n}{\frac{M^2+M}{2}}$ sets, we conclude that k-rank$(\overline{\mathbf{C}}) = \overline{M}$, with probability one over $\mathbb{R}^{Mn}$. Consider a set $\mathcal{S} \subseteq \{1,\cdots,n\}$ with $|\mathcal{S}| = \overline{M}$, and let $\overline{\mathbf{C}}_{\mathcal{S}} \in \mathbb{R}^{\overline{M}\times\overline{M}}$ denote the submatrix of $\overline{\mathbf{C}}$, whose columns are indexed by the set $\mathcal{S}$ (similarly define $\mathbf{A}_{\mathcal{S}},\mathbf{B}_{\mathcal{S}}$). Consider the function

$$f(\mathbf{A}_{\mathcal{S}},\mathbf{B}_{\mathcal{S}}) = \det(\underbrace{\overline{\mathbf{A}_{\mathcal{S}}\odot\mathbf{A}_{\mathcal{S}}} - \overline{\mathbf{B}_{\mathcal{S}}\odot\mathbf{B}_{\mathcal{S}}}}_{\overline{\mathbf{C}}_{\mathcal{S}}})$$

Because the elements in $\mathbf{A}_{\mathcal{S}},\mathbf{B}_{\mathcal{S}}$ are drawn independently from continuous distributions, the function $f(\mathbf{A}_{\mathcal{S}},\mathbf{B}_{\mathcal{S}})$ is a multivariate polynomial in $2M\overline{M}$ variables, and therefore is analytic in $\mathbb{R}^{2M\overline{M}}$. Using the properties of analytic functions, if $f(\mathbf{A}_{\mathcal{S}},\mathbf{B}_{\mathcal{S}})$ is not an identically zero polynomial, its zero set will have measure zero on $\mathbb{R}^{2M\overline{M}}$. Therefore, it remains to show that $f(\mathbf{A}_{\mathcal{S}},\mathbf{B}_{\mathcal{S}})$ is not a trivial polynomial. One way to do that is to find a specific point $(\mathbf{A}_0,\mathbf{B}_0)$ such that $f(\mathbf{A}_0,\mathbf{B}_0) \neq 0$. We follow the same construction for $\mathbf{A}_0$ as the one proposed in the proof of Theorem 4 in [PV15]. We repeat this construction here for completeness: Divide the $\overline{M}$ columns of $\mathbf{A}_0$ into M unequal groups $\{J_m\}_{m=1}^{M}$ such that $|J_m| = M-m+1$. Relabel the columns of $\mathbf{A}_0$ as $\{j_m^k, 1 \leq k \leq M-m+1\}$, and construct the elements of $\mathbf{A}_0$ in the columns indexed by J_m (for all $m = 1,\cdots,M$) as

$$[\mathbf{A}_0]_{i,j_m^k} = \begin{cases} 1 & \text{if } i = m, \text{and } k = 1 \\ 1 & \text{if } (i = m \text{ or } i = m+k-1), \text{and } k > 1 \\ 0 & \text{otherwise} \end{cases}$$

As shown in [PV15], the matrix $\overline{\mathbf{A}_0 \odot \mathbf{A}_0}$, through permuting rows and columns, can be transformed into an upper triangular matrix, with diagonal entries equal to 1. We also let $\mathbf{B}_0 = \mathbf{0}$. Therefore, $f(\mathbf{A}_0, \mathbf{B}_0) = \det(\overline{\mathbf{A}_0 \odot \mathbf{A}_0} - \overline{\mathbf{B}_0 \odot \mathbf{B}_0}) = 1$. This implies that $f(\mathbf{A}_{\mathcal{S}}, \mathbf{B}_{\mathcal{S}})$ is a nontrivial polynomial in $\mathbb{R}^{2M\bar{M}}$, and hence with probability one $f(\mathbf{A}_{\mathcal{S}}, \mathbf{B}_{\mathcal{S}}) \neq 0$. This implies $\overline{\mathbf{C}}_{\mathcal{S}}$ is full-rank, thereby concluding the proof. □

Remark 27. Lemma 19 implies that the columns of $\overline{\mathbf{A} \odot \mathbf{A}} - \overline{\mathbf{B} \odot \mathbf{B}}$ are i.i.d. points in $\mathbb{R}^{\frac{M^2+M}{2}}$ such that with probability one, every set of $\frac{M^2+M}{2}$ points are linearly independent.

Lemma 20. *[Wor94] (Chernoff bound) When $k > n/2$, it holds that*

$$\frac{1}{2^n} \sum_{i=k}^{n} \binom{n}{i} \leq e^{-k(\frac{n}{2k} - 1 - \log(\frac{n}{2k}))}$$

Proof of Theorem 21

Equipped with above lemmas and definitions, we now proceed to the proof of Theorem 21:

Proof. Let S_i, S_j be index sets such that $\mathcal{S}_i \cap \mathcal{S}_j = \emptyset$, $1 \leq i < j \leq \binom{N}{K}$. Define the event N_{ij}^{++} as

$$N_{ij}^{++} = \{\exists \mathbf{p} \in \mathbb{R}_{++}^{K} : (\mathbf{A}_i \odot \mathbf{A}_i - \mathbf{A}_j \odot \mathbf{A}_j)\mathbf{p} = \mathbf{0}\}.$$

Notice that the condition $(\mathbf{A}_i \odot \mathbf{A}_i - \mathbf{A}_j \odot \mathbf{A}_j)\mathbf{p} = \mathbf{0}$ is equivalent to $(\overline{\mathbf{A}_i \odot \mathbf{A}_i} - \overline{\mathbf{A}_j \odot \mathbf{A}_j})\mathbf{p} = \mathbf{0}$, since $\overline{\mathbf{A}_i \odot \mathbf{A}_i}$, following the definition given in Definition 10, only removes the identical rows from $\mathbf{A}_i \odot \mathbf{A}_i$. Since the columns of $\mathbf{A}$ are i.i.d., we can conclude that the columns of $\overline{\mathbf{A} \odot \mathbf{A}}$ are also i.i.d. Since, $\mathcal{S}_i \cap \mathcal{S}_j = \emptyset$, this implies that the columns of $\overline{\mathbf{A}_i \odot \mathbf{A}_i} - \overline{\mathbf{A}_j \odot \mathbf{A}_j}$ are i.i.d. as well. Moreover, the distribution of the columns of $\overline{\mathbf{A}_i \odot \mathbf{A}_i} - \overline{\mathbf{A}_j \odot \mathbf{A}_j}$ is symmetric about zero.

Since $\mathbf{A}_i$ and $\mathbf{A}_j$ have i.i.d. columns, we can use Lemma 19 (with $n := K, \mathbf{A} := \mathbf{A}_i, \mathbf{B} := \mathbf{A}_j$), to conclude that all subsets of $\overline{M}$ columns of $\overline{\mathbf{A}_i \odot \mathbf{A}_i} - \overline{\mathbf{A}_j \odot \mathbf{A}_j}$ are linearly independent, with probability one. Hence, the K columns of $\overline{\mathbf{A}_i \odot \mathbf{A}_i} - \overline{\mathbf{A}_j \odot \mathbf{A}_j}$ are i.i.d. points in $\mathbb{R}^{\frac{M^2+M}{2}}$

such that their distribution is symmetric about zero, and every subset of $\overline{M}$ points are linearly independent, with probability one. Therefore, they satisfy the conditions for our corollary to Wendel's theorem. We can directly use Lemma 10 and Corollary 4 (with $d := \overline{M}$, $n := K$) to compute the probability of the event N_{ij}^{++} as

$$\mathbb{P}(N_{ij}^{++}) = 1 - \frac{1}{2^{K-1}} \sum_{i=0}^{\overline{M}-1} \binom{K-1}{i} \tag{3.58}$$

$$\geq 1 - e^{-(K-\overline{M})(\frac{K-1}{2(K-\overline{M})} - \log(\frac{K-1}{2(K-\overline{M})})-1)} \tag{3.59}$$

where the inequality (3.59) follows from Lemma 20, assuming that $K > 2\overline{M}+1$.

We are now ready to characterize the probability that $\mathbf{A} \in I_{\mathbf{A}}$. Define the set

$$\overline{I}_{\mathbf{A}} = \{\mathbf{A} : \exists \mathbf{p} \in \mathbb{R}_{++}^{K}, i, j \,\big|\, \mathcal{S}_i \cap \mathcal{S}_j = \emptyset, \\ (\mathbf{A}_i \odot \mathbf{A}_i - \mathbf{A}_j \odot \mathbf{A}_j)\mathbf{p} = \mathbf{0}\}$$

The difference between $I_{\mathbf{A}}$ and $\overline{I}_{\mathbf{A}}$ is that in $\overline{I}_{\mathbf{A}}$ we restrict ourselves to disjoint supports. It is clear that

$$\mathbb{P}(\mathbf{A} \in I_{\mathbf{A}}) \geq \mathbb{P}(\mathbf{A} \in \overline{I}_{\mathbf{A}}) \tag{3.60}$$

and hence, any lower bound on $\mathbb{P}(\mathbf{A} \in \overline{I}_{\mathbf{A}})$ serves as a lower bound on $\mathbb{P}(\mathbf{A} \in I_{\mathbf{A}})$. We first derive a loose lower bound on $\mathcal{P}(\mathbf{A} \in \overline{I}_{\mathbf{A}})$, by only considering two disjoint supports $\mathcal{S}_1 = \{1, \cdots, K\}$ and $\mathcal{S}_2 = \{K+1, \cdots, 2K\}$. Then we have

$$\mathbb{P}(\mathbf{A} \in \overline{I}_{\mathbf{A}}) = \mathbb{P}\Big(\bigcup_{\mathcal{S}_i \cap \mathcal{S}_j = \emptyset} N_{ij}^{++}\Big) \geq \mathbb{P}(N_{12}^{++}) \\ = 1 - \frac{1}{2^{K-1}} \sum_{i=0}^{\overline{M}-1} \binom{K-1}{i}. \tag{3.61}$$

Hence, (3.61) serves as a (loose) lower bound on $\mathcal{P}(\mathbf{A} \in \overline{I}_{\mathbf{A}})$. A tighter lower bound can be found as follows. Let $N' = \lfloor \frac{N}{2K} \rfloor 2K$. Consider partitioning the indices $\{1, \cdots, N'+2K\}$ as

$$\mathbb{N}_{[1,K]}, \mathbb{N}_{[K+1,2K]}, \cdots, \mathbb{N}_{[N'+K+1,N'+2K)]}$$

where $\mathbb{N}_{[a,b]} = \{a, a+1, \cdots, b\}$, for $a, b \in \mathbb{N}$, $a \leq b$. Now consider the following set consisting of tuples of the form $(\mathbb{N}_{a,a+K}, \mathbb{N}_{a+K+1,a+2K})$:

$$\begin{aligned} I_{K,N} = \Big\{ & \left(\mathbb{N}_{[1,K]}, \mathbb{N}_{[K+1,2K]}\right), \\ & \left(\mathbb{N}_{[2K+1,3K]}, \mathbb{N}_{[3K+1,4K]}\right), \cdots, \\ & \left(\mathbb{N}_{[N'+1,N'+K]}, \mathbb{N}_{[N'+K+1,N'+2K]}\right) \Big\} \end{aligned}$$

Each tuple in $I_{K,N}$ represents a pair of disjoint supports of size K. The probability $\mathbb{P}(\mathbf{A} \in \overline{I}_{\mathbf{A}})$ can be bounded as

$$\begin{aligned} \mathbb{P}(\mathbf{A} \in \overline{I}_{\mathbf{A}}) &= \mathbb{P}\Big(\bigcup_{\mathcal{S}_i \cap \mathcal{S}_j = \emptyset} N_{ij}^{++} \Big) \\ &\geq \mathbb{P}\Big(\bigcup_{(\mathcal{S}_i, \mathcal{S}_j) \in I_{K,N}} N_{ij}^{++} \Big) \\ &= 1 - \mathbb{P}\Big(\bigcap_{(\mathcal{S}_i, \mathcal{S}_j) \in I_{K,N}} \bar{N}_{ij}^{++} \Big) \\ &= 1 - \prod_{(\mathcal{S}_i, \mathcal{S}_j) \in I_{K,N}} \mathbb{P}(\bar{N}_{ij}^{++}) \qquad (3.62) \\ &= 1 - \left(\frac{1}{2^{K-1}} \sum_{i=0}^{\frac{1}{2}M(M+1)-1} \binom{K-1}{i} \right)^{\lfloor \frac{N}{2K} \rfloor} \qquad (3.63) \end{aligned}$$

where $\bar{N}_{ij}^{++}$ denotes the complement of the event N_{ij}^{++}, the equality (3.62) follows from the fact that the events N_{ij}^{++} are independent, since all the supports $(\mathcal{S}_i, \mathcal{S}_j) \in I_{K,N}$ are disjoint, and (3.63)

follows from (3.58). By using the lower bound in (3.59), we can derive the bound given in (3.30), and conclude the proof.

□

3.5.3 A preliminary lemma

Lemma 21. *[Zaj18] Let $\mathbf{g} \in \mathbb{R}^d$, be a vector of random Gaussian variables distributed as $\mathcal{N}(0,\mathbf{I})$, and $\mathbf{A} \in \mathbb{R}^{d\times d}$ then*

$$\mathbb{P}(|\mathbf{g}^T\mathbf{A}\mathbf{g} - \mathbb{E}(\mathbf{g}^T\mathbf{A}\mathbf{g})| > t) \leq 2exp\left(-\min\left\{\frac{9t^2}{512C_1^2\|\mathbf{A}\|_F^2}, \frac{3t}{32C_1\|\mathbf{A}\|}\right\}\right)$$

where $C_1 > 0$ is a universal constant.

Lemma 22. *Let $x_1,\cdots,x_L$, $y_1,\cdots,y_L$ be random Gaussian variables such that each pair (x_i,y_i) is distributed as $\mathcal{N}(0,\Sigma)$, where $\Sigma \in \mathbb{R}^{2\times 2}$, then*

$$\mathbb{P}(|\mathbf{g}^T\mathbf{A}\mathbf{g}| > t) \leq 2exp\left(-\min\left\{\frac{9Lt^2}{128C_1^2\|\Sigma\|_F^2}, \frac{3Lt}{16C_1\|\Sigma\|_F}\right\}\right)$$

Proof. We can write

$$x_i y_i = \mathbf{z}_i^T\mathbf{J}\mathbf{z}_i$$

where

$$\mathbf{J} = \begin{bmatrix} 0 & 1/2 \\ 1/2 & 0 \end{bmatrix}, \mathbf{z}_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix}$$

Moreover, $\mathbf{z}_i$ can be expressed as $\mathbf{z}_i = \Sigma^{1/2}\mathbf{g}_i$, where $\mathbf{g}_i \sim \mathcal{N}(0,\mathbf{I}_2)$. Therefore, $x_i y_i = \mathbf{g}_i^T\Sigma^{1/2}\mathbf{J}\Sigma^{1/2}\mathbf{g}_i$,

and

$$\frac{1}{L}x_i y_i = \frac{1}{L}\mathbf{g}^T(\mathbf{I}_L \otimes (\Sigma^{1/2}\mathbf{J}\Sigma^{1/2}))\mathbf{g}$$

where $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{2L})$. Therefore, we can use $\mathbf{A} := (\mathbf{I}_L \otimes (\Sigma^{1/2}\mathbf{J}\Sigma^{1/2}))$ in Lemma 21. We have

$$\begin{aligned}\|\mathbf{A}\|_F^2 &= \frac{1}{L}\|\Sigma^{1/2}\mathbf{J}\Sigma^{1/2}\|_F^2 \\ &= \frac{1}{L}\mathrm{tr}(\Sigma^{1/2}\mathbf{J}\Sigma\mathbf{J}\Sigma^{1/2}) \\ &= \frac{1}{L}\mathrm{tr}(\Sigma\mathbf{J}\Sigma\mathbf{J}) = \frac{1}{4L}\|\Sigma\|_F^2\end{aligned}$$

and

$$\begin{aligned}\|\mathbf{A}\| &= \frac{1}{L}\|\mathbf{I}_L\|\|\Sigma^{1/2}\mathbf{J}\Sigma^{1/2}\| \\ &\leq \frac{1}{L}\|\Sigma^{1/2}\mathbf{J}\Sigma^{1/2}\|_F \leq \frac{1}{2L}\|\Sigma\|_F\end{aligned}$$

$$\mathbb{P}(|\mathbf{g}^T\mathbf{A}\mathbf{g}| > t) \leq 2\exp\left(-\min\left\{\frac{9Lt^2}{128C_1^2\|\Sigma\|_F^2}, \frac{3Lt}{16C_1\|\Sigma\|_F}\right\}\right)$$

□

3.5.4 Proof of Theorem 24

Proof. Let w_n be the size of the set $\mathcal{W}$ at the nth iteration of the algorithm, and i_n be the counter i in Algorithm 1 at nth iteration, where it holds that $0 \leq i_n \leq n$. Of all the pairs $\mathbf{r} \cap \mathbf{c}$ for $\mathbf{r} \in \mathcal{R}$ and $\mathbf{c} \in \mathcal{C}$ at most K of them can be nonempty. The number of elements in each set $\mathbf{r}$ is at most

$\lceil \frac{w_n}{2^{i_n}} \rceil$, and the number of elements in each set $\mathbf{r} \cap \mathbf{c}$ is at most $\lceil \frac{w_n}{2^{2i_n}} \rceil$. Therefore,

$$w_{n+1} \leq \min(w_n, K\lceil \frac{w_n}{2^{2i_n}} \rceil),$$

and we have $w_1 = N$. At the nth iteration, we make at most 2^{n+1} measurements. The algorithm stops when the number of elements in the sets $\mathbf{r} \cap \mathbf{c}$ is at most one, which means $\lceil \frac{w_n}{2^{2i_n}} \rceil = 1$. We denote the stopping iteration as $n = n_t$.

We show that after some iteration $\hat{n}$, for all $\hat{n} \leq n \leq n_t$ we necessarily have

$$w_{n+1} < \frac{w_n}{2} \tag{3.64}$$

A sufficient condition for (3.64) is that

$$K\lceil \frac{w_n}{2^{2i_n}} \rceil \leq \frac{w_n}{2}$$

Since $n < n_t$, we necessarily have $\frac{w_n}{2^{2i_n}} > 1$, hence we can use the fact that $\lceil x \rceil \leq 2x$, for $x \geq 1$. Therefore, we can derive a more relaxed sufficient condition as

$$2K\frac{w_n}{2^{2i_n}} \leq \frac{w_n}{2} \tag{3.65}$$

Notice that (3.65) implies (3.64) for $n < n_t$. From (3.65) we get $i_n \geq \frac{1}{2}\log_2(4K)$. Since we always have $i_n \leq n$, we conclude that for $n \geq \hat{n} = \lceil \frac{1}{2}\log_2(4K) \rceil$, (3.64) necessarily holds. The number of measurements upto the iteration $\hat{n} =$ is

$$2\sum_{i=1}^{\hat{n}} 2^{i_n} \leq 2\sum_{i=1}^{\hat{n}} 2^n = 4\sum_{n=0}^{\hat{n}-1} 2^n = 4(2^{\hat{n}} - 1) \leq 4(2^{\frac{1}{2}\log_2(2K)} + 1) \leq 8\sqrt{4K}$$

Since for all iterations $n \geq \hat{n}$, (3.64) holds, based on the Algorithm 1, we have $i_n = \hat{n}$ for $n \geq \hat{n}$. Using (3.64), and the fact that $w_{\hat{n}} \leq N$, we get $w_n \leq \lceil N2^{-(n-\hat{n})} \rceil$, for all $n \geq \hat{n}$. The algorithm

stops at iteration n_t which satisfies $\frac{w_{n_t}}{2^{2i_{n_t}}} \leq 1$. A sufficient condition is that

$$\frac{2N2^{-(n_t-\hat{n})}}{2^{2\hat{n}}} \leq 1$$

which means $i_n = \lceil \frac{1}{2}\log_2 N \rceil$, i.e.,

$$\frac{2^{-n_t+1+\log_4 N}}{2^{\hat{n}}} \leq 1$$

where $n_t = \lceil 1+\log_2 N - \hat{n} \rceil$.

The number of measurements from iteration $\hat{n}$ to the final iteration n_t is

$$\begin{aligned} 2^{\hat{n}}(n_t - \hat{n} + 1) &\leq (\lceil 1+\log_2 N - \hat{n} \rceil - \hat{n} + 1)2^{\lceil \frac{1}{2}\log_2(4K) \rceil} \\ &\leq 4\sqrt{4K}\log_2 N \end{aligned}$$

Hence, the overall number of measurements is at most $4\sqrt{4K}(\log_2 N + 2)$. □

3.5.5 Proof of Theorem 25

Proof. A sufficient condition for the adaptive scheme provided in Algorithm 1 to fail is that at any stage of the algorithm we mistakenly remove indices corresponding to the set $\mathcal{W}(\mathbf{c}_i \cap \mathbf{r}_j)$. This can happen if we have $|\rho_{ij}| < \varepsilon$ while $\mathbf{p}_{\mathcal{W}(\mathbf{c}_i \cap \mathbf{r}_j)}$ has nonzero elements. We use the tail bound provided in Lemma 22, with

$$\mathbb{P}(|\rho_{ij}| > \varepsilon) \leq 2\exp\Bigg(-\min\Bigg\{ \frac{9L\varepsilon^2}{256C_1^2(2p_{\max}^2K^2+\sigma^4)}, \\ \frac{3L\varepsilon}{16C_1\sqrt{2(2p_{\max}^2K^2+\sigma^4)}} \Bigg\} \Bigg)$$

By employing the union bound and the fact that the total number of making a decision is

$$\sum_{n=1}^{\hat{n}} 2^{2n} + (n_1 - \hat{n})2^{2\hat{n}} = \frac{4}{3}(2^{2\hat{n}} - 1) + (\lceil 1 + \log_2 N - \hat{n} \rceil - \hat{n})2^{2\hat{n}}$$
$$\leq (\frac{8}{3} + 4(2 + \log_2 N))\sqrt{4K} = (4(\frac{8}{3} + \log_2 N))\sqrt{4K}$$

the probability of success is at least

$$1 - (8(\frac{8}{3} + \log_2 N))\sqrt{4K}\exp\left(-\min\left\{\frac{9L\varepsilon^2}{256C_1^2(2p_{\max}^2K^2 + \sigma^4)}, \frac{3L\varepsilon}{16C_1\sqrt{2(2p_{\max}^2K^2 + \sigma^4)}}\right\}\right)$$

□

3.6 Acknowledgements

The contents in this chapter are reprint of the material published in "IEEE Transactions on Signal Processing" (Sec. 3.1), "IEEE ICASSP 2018" (Sec. 3.2), "SPIE 2018 Conference on Compressive Sensing VII: From Diverse Modalities to Big Data Analytics" (Sec. 3.3), "2019 53rd Asilomar Conference on Signals, Systems, and Computers" (Sec. 3.4).

The work in Sec. 3.1 was supported by in part by the "National Science Foundation under CAREER Award ECCS 1553954", in part by the "Office of Naval Research under Grant N00014-18-1-2038", and in part by the "University of California, San Diego". Sec. 3.2 was supported in part by "NSF CAREER award (ECCS 1553954)", and in part by the "University of California, San Diego". Sec 3.4 was supported by the "Office of Naval Research grant (ONR N00014-18-1-2038)", and the "University of California, San Diego".

Chapter 4

Tensor Decompositions and Non-Convex Algorithms with Applications in mmWave Communication Systems

In this chapter, we present tensor decomposition algorithms, as well as non-convex algorithms with applications in mmWave communication systems. In the following, we will give a brief introduction on a specific type of tensor decomposition, namely Canonical Polyadic decomposition, as well as a review on channel estimation for mmWave communication systems.

Canonical Polyadic Decomposition Decomposition of tensors (d-dimensional arrays, $d > 2$) into minimal number of rank-1 tensors, referred to as Canonical Polyadic decomposition (also CANDECOMP/PARAFAC), is an important problem that arises accross diciplines such as signal processing, machine learning, psychometrics, chemometrics, etc [KB09, SDLF$^+$17]. It has been established that unlike matrices, tensors can have CP ranks (the number of rank-1 tensors in CP decomposition) that can far exceed the individual dimensions of the tensor, and in most cases decomposition of such tensors are known to be NP-hard [HL13]. Nevertheless, it has been shown

that under some mild conditions, CP decomposition can be unique, even without assuming any structure on the tensor factors.

Although most of the literature on tensor decomposition is focused on unstructured tensors [Bro97, SDLF$^+$17], there is also a significant body of literature which focuses on tensors that have extra structure [SDL13, SDL17a, GBB$^+$16]. For example, tensors whose factors are Vandermonde matrices arise in applications such as array signal processing [SBG00], and multidimensional harmonic retrieval [SDL13, SDL17a, SDL17b]. Taking advantage of these structures can lead to easily implementable algorithms which are capable of recovering tensors of rank higher than the dimensions. For example, [SDL17a] has shown that for a dth order tensor consisting of $d-1$ Vandermonde factors of dimensions $N_1, N_2, \cdots, N_{d-1}$, upto $N_1 N_2 \cdots N_{d-1}$ factors can be recovered using only linear algebraic operations.

Recent results in Direction-of-Arrival (DOA) estimation [PV10, PV11] show that certain non-uniform array geometries such as nested [PV10], coprime [PV11], [QP14], and minimum redundancy arrays [Mof68] can be useful to detect more sources than the number of sensors. In particular, these geometries are capable of recovering $O(M^2)$ sources using only M antennas. In such array geometries, because of non-uniformity of the array, the array manifold matrix is no longer Vandermonde, but a sub-selected rows of a Vandermonde matrix, henceforth called *Vandermonde-like*. However, the arrays are designed in such a way that the self-Khatri-Rao product of these matrices have a Vandermonde submatrix. The Khatri-Rao structure arises naturally under the assumption that the sources are uncorrelated. The idea is also generalized into $2q$th order moments of a $2q$th order nested array in [PV12c], where recovery of $O(M^{2q})$ uncorrelated non-Gaussian sources is guaranteed using only M antennas.

In section 4.2, we present an algorithm for decomposition of structured overcomplete tensors. In Sec. 4.4 we will show how we can use this algorithm in order to perform channel estimation for mmWave communication systems.

Channel Estimation for mmWave Communication Systems Communication over millimeter-wave (mmWave) frequencies (30GHz-300GHz), which is a key enabler of 5th generation of wireless communication systems, is a challenging task due to several physical limitations such as high propagation loss, directivity, and sensitivity to blockage [NLJ+15, PK11, RSP+14]. One of the well-known solutions to overcome some of these challenges is to increase spatial diversity by deploying a massive number of antennas (e.g., 100s), at the base-station (BS), and possibly multiple antennas at the mobile-station (MS) [RSP+14]. However, deployment of such a huge number of antennas can dramatically complicate the radio-frequency (RF) circuitry of these systems, leading to an increased power consumption, and manufacturing cost. To mitigate these problems, a well-known solution is to use a hybrid analog/digital beamforming strategy, which helps us reduce the number of required RF-chains. In such architectures, N RF-chains are connected to M antennas ($N \ll M$) through a network of simpler RF circuitry such as analog phase-shifters and/or switches [MRH+17].

In order to ensure a reliable communication, it is crucial to obtain an accurate estimate of the channel. A key property of mmWave channels is that they can be characterized by a few number of channel paths [NLJ+15, XGJ16, AH16, PAGH19], even though the channel matrix itself can be very large. Each channel path is characterized by a few parameters such as gain, angle-of-arrival (AoA), angle-of-departure (AoD), delay, and doppler-shift[1]. Hence, in order to estimate the channel, it suffices to estimate these parameters [PAGH19, XGJ16]. Many papers have already addressed the problem of channel estimation in different scenarios [PH16, PAGH19, RTC15, ZH17, QFSY18, QFS19, ALH15, LGL14, AMGPH14, EARAS+14], including single-carrier [PH16, QFS19, AEALH14] versus multi-carrier cases [PAGH19, ZH17, ZFY+17, AdA17], or high-mobility [RTC15] versus low-mobility environments [ZH17]. In this chapter, we consider the case where the communication system operates on multiple carriers in a low-mobility environment.

[1]Depending on the assumptions made in the problem model, the effect of some of these parameters can be negligible.

In Sections 4.3 and 4.4, we will show how we can design tensor decomposition algorithms which can help us estimate the mmWave communication channels.

4.1 Beam-Pattern Design for Hybrid Beamforming using Wirtinger Flow

In a wireless mmWave communication system, in order to maximize the signal-to-noise ratio (SNR) at the receiver, beamforming vectors at the transmitter and receiver are designed carefully using the Channel State Information (CSI). However, in mmW systems, a large number of antennas at both transmitter and receiver makes it impractical to accurately estimate the large channel matrix, as it requires a large communication overhead. On the other hand, due to high directivity of mmW channels, and availability of a sufficiently large number of antennas, it is possible to efficiently implement beamforming using feedback assisted beam alignment [HKL$^+$13, SCL$^+$15a]. In such scenarios, a common codebook is available at both transmitter and receiver. The codebook contains a set of codewords, each corresponding to a particular predefined beampattern [SCL15b, SCL17].

A major challenge in mmW communication systems is that owing to the large number of antennas, dedicating a separate radio frequency (RF) chain (consisting of analog-to-digital converters, modulators, etc) to individual antenna is extremely costly and power-hungry [PK11]. Therefore, unlike conventional MIMO systems, it is not practical to implement an all-digitial beamformer. In order to mitigate this problem, alternative architectures have been proposed. A common practice is to use a hybrid digital/analog beamforming strategy [MRH$^+$17, SY16, HIR$^+$14, AMGPH14, BBS13, SB10, EARAS$^+$14, RMRGPH16, NLLNH17, LWY$^+$17], where instead of purely digital processing, a combination of digital and analog beamforming is employed. Since the analog circuitry in a hybrid beamforming system is implemented using only phase-shifters [MRH$^+$17], it poses certain non-convex constraints on the design of a hybrid beamformer, making it a challenging optimization problem.

A body of past work on hybrid beamforming is dedicated to maximizing data transmission rate, and computationally efficient algorithms have been proposed [RMRGPH16, NLLNH17,

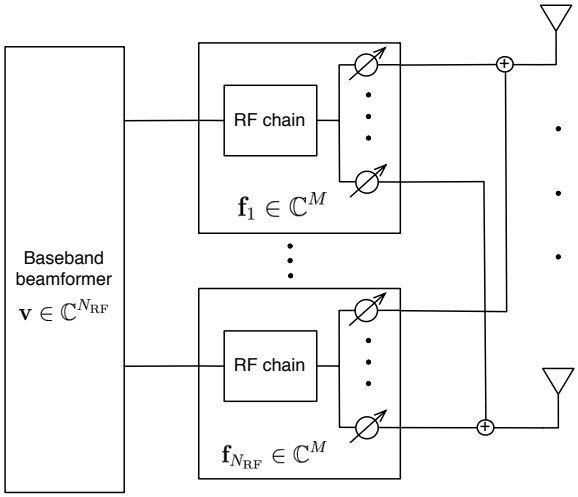


Figure 4.1: Hybrid beamforming

MRH$^+$17]. In contrast, in this Section, we focus on a different optimization problem where the goal is to approximate a given beam-pattern, for which we develop a new non-convex algorithm. Our algorithm performs a greedy initialization followed by a gradient descent procedure. Since the optimization variables, i.e., the baseband weights, and the phase-shifter coefficients are complex quantities, we use Wirtinger calculus to compute the derivatives. Wirtinger calculus has shown to be successful in other applications such as solving quadratic equations and phase-retrieval [CLS15]. The proposed algorithm can be used in codebook design, and channel sounding stages of a mmW communication system, and works for arbitrary array geometries (including non uniform arrays such as nested and coprime arrays [PV10, PV11]). Our simulations show that compared to previous optimization techniques for beam-pattern design in hybrid beamforming [SCL15b, SCL17], our algorithm has lower computational complexity.

Notations: Throughout this Section, matrices and vectors are represented by bold uppercase and bold lowercase letters, respectively. The symbol $(.)^T, (.)^H, (.)^*, (.)^\dagger$ represent the matrix transpose, matrix Hermitian, complex conjugate, and Moore-Penrose psuedo-inverse, respectively. The symbols $\otimes$, $\odot$, $\circ$ stand for Kronecker, and (column-wise) Khatri-Rao product, and elementwise Hadamard product, respectively. The rest of the notations are clear from the context.

4.1.1 System Model

Consider a millimeter wave (mmW) antenna array, consisting of M antennas, which can act as transmitter and/or receiver in a massive MIMO wireless communication system. As depicted in 4.1, we assume that the antenna array is equipped with a hybrid-beamforming hardware, consisting of a digital baseband beamformer, and an analog beamforming circuitry. The analog circuit consists of $N_{\text{RF}} < M$ RF chains, which are connected to the physical antennas through a network of analog phase-shifters, where each RF-chain is connected to an antenna through a phase-shifter and an adder. The analog phase shifters connected to each RF chain are controlled by the unit-norm beamsteering vectors $\mathbf{f}_1, \cdots, \mathbf{f}_{N_{\text{RF}}} \in \mathcal{B}_M$, where

$$\mathcal{B}_M = \{\mathbf{w} \in \mathbb{C}^M : |w_i|^2 = 1/M, 1 \leq i \leq M\}.$$

Moreover, the digital baseband beamformer is controlled by the digital beamsteering vector $\mathbf{v} \in \mathbb{C}^{N_{\text{RF}}}$. Hence, the overall hybrid beamsteering vector is given by $\mathbf{c} := \mathbf{F}\mathbf{v}$, where $\mathbf{F} = [\mathbf{f}_1, \cdots, \mathbf{f}_{N_{\text{RF}}}]$. We also assume that the hybrid beamformer should satisfy the unit norm constraint $\|\mathbf{c}\| = 1$, i.e., $\|\mathbf{F}\mathbf{v}\| = 1$, due to the transmit power constraint of the beamformer.

In this Section, we consider a one dimensional linear antenna array of arbitrary geometry. Assume that the antenna array is designed for a carrier wavelength of λ, and the antennas are located at $d_1, \cdots, d_M$. In this case, the array steering vector $\mathbf{a}(\theta) \in \mathbb{C}^M$ can be written as

$$\mathbf{a}(\theta) = [e^{j\frac{2\pi}{\lambda}d_1 \sin\theta}, e^{j\frac{2\pi}{\lambda}d_2 \sin\theta}, \cdots, e^{j\frac{2\pi}{\lambda}d_M \sin\theta}]^T,$$

In a uniform linear array (ULA), the locations are chosen to be $d_1 = 0, d_2 = \frac{\lambda}{2}, \cdots, d_M = \frac{\lambda}{2}(M-1)$. In this Section, the antennas need not be located uniformly, and our algorithm can be applied to nonuniform arrays (such as nested and coprime arrays) as well.

The beampattern generated by the hybrid antenna array is of the following form

$$t(\theta;\mathbf{F},\mathbf{v}) = |\mathbf{a}(\theta)^H \mathbf{F}\mathbf{v}|^2.$$

In this Section, our goal is to design the hybrid beamformer (i.e., the analog phase shifters $\mathbf{F}$, and the digital beamformer $\mathbf{v}$), such that the resulting beam-pattern $t(\theta;\mathbf{F},\mathbf{v})$ is as close as possible to a desired beampattern $p(\theta)$. Towards this end, we consider the same optimization problem as in [SCL15b, SCL17]:

$$\begin{aligned} \underset{\mathbf{F}\in\mathbb{C}^{M\times N_{\mathrm{RF}}},\mathbf{v}\in\mathbb{C}^{N_{\mathrm{RF}}},\alpha\in\mathbb{R}}{\text{minimize}} \quad & \int_{-\pi}^{\pi} |\alpha t(\theta;\mathbf{F},\mathbf{v}) - p(\theta)|^2 d\theta \\ \text{subject to} \quad & \mathbf{f}_1,\cdots,\mathbf{f}_{N_{\mathrm{RF}}} \in \mathcal{B}_M, \\ & \|\mathbf{F}\mathbf{v}\| = 1 \end{aligned} \tag{4.1}$$

where α is a scaling factor. We consider a more tractable version of (4.1), by quantizing the directions θ into a finite grid of N points: $\{\theta_1,\cdots,\theta_N\}$. Define the vectors $\mathbf{t},\mathbf{p}$ whose elements are given by $t_i = t(\theta_i;\mathbf{F},\mathbf{v})$, and $p_i = p(\theta_i)$. In this case, one can write

$$\mathbf{t} = |\mathbf{A}^H \mathbf{F}\mathbf{v}|^{\circ 2},$$

where the notation $|.|^{\circ 2}$ indicates taking element-wise squared absolute value of a vector, and $\mathbf{A} = [\mathbf{a}(\theta_1),\cdots,\mathbf{a}(\theta_N)]$. In order to incorporate the constraint $\|\mathbf{F}\mathbf{v}\| = 1$, one can define the auxiliary variable $\mathbf{u}$, and solve the following optimization problem:

$$\begin{aligned} (\hat{\mathbf{F}},\hat{\mathbf{u}}) = \quad & \underset{\substack{\mathbf{F}\in\mathbb{C}^{M\times N_{\mathrm{RF}}} \\ \mathbf{u}\in\mathbb{C}^{N_{\mathrm{RF}}},\alpha\in\mathbb{R}^{++}}}{\text{argmin}} \quad \|\alpha\frac{|\mathbf{A}^H\mathbf{F}\mathbf{u}|^{\circ 2}}{\|\mathbf{F}\mathbf{u}\|_2^2} - \mathbf{p}\|^2 \\ & \text{subject to} \quad \mathbf{f}_1,\cdots,\mathbf{f}_{N_{\mathrm{RF}}} \in \mathcal{B}_M, \end{aligned} \tag{4.2}$$

The optimal $\mathbf{v}$ is then given as $\hat{\mathbf{v}} = \hat{\mathbf{u}}/\|\hat{\mathbf{F}}\hat{\mathbf{u}}\|$. Both the constraints and the objective function of (4.2) are nonconvex, rendering the problem (4.2) a nonconvex optimization problem. In the following section, we will propose a non-convex algorithm based on the Wirtinger Flow framework [CLS15, CC15]. The method consists of suitable initialization followed by a series of gradient descent iterations, where the complex gradients are derived using Wirtinger calculus [KD09, HG07].

4.1.2 Proposed Algorithm: Lifting Aided Wirtinger Flow

In order to solve the optimization problem (4.2), we propose a greedy approach to first find initial estimates for the variables $\mathbf{F}, \mathbf{u}$. Once the initial estimates are found, we then proceed by performing gradient descent iterations to arrive at a local minimum. Since the optimization is performed with respect to complex variables, we use Wirtinger calculus in matrix variables to find the derivatives [HG07].

Initialization Based on Lifting

In this subsection, our goal is to find a suitable initialization point for the variables $\mathbf{F}, \mathbf{u}$. To this end, we first ignore the fact that the overall beamsteering vector is decomposable as a product of $\mathbf{F}$ and $\mathbf{v}$, and instead, consider a relaxed version of (4.2) as follows:

$$\underset{\mathbf{c}\in\mathbb{C}^M, \alpha\in\mathbb{R}^{++}}{\text{minimize}} \quad \|\alpha|\mathbf{A}^H\mathbf{c}|^{\circ 2} - \mathbf{p}\|_2^2 \quad \text{subject to } \|\mathbf{c}\| = 1 \tag{4.3}$$

where $\mathbf{c}$ can be any unit-norm vector in $\mathbb{C}^M$, and need not follow the hybrid beamforming constraints. Notice that (4.3) is still a nonconvex problem in $\mathbf{c}$, and hard to solve. Using properties of Kronecker and Hadamard product, one can verify that $|\mathbf{A}^H\mathbf{c}|^{\circ 2} = (\mathbf{A}^* \odot \mathbf{A})^H(\mathbf{c}^* \otimes \mathbf{c})$. Using the "lifting" trick (which has been popularized for phase-retrieval [CLS15] and blind deconvolution

[ARR14a]), one can define a new variable $\mathbf{D} = \alpha \mathbf{c}\mathbf{c}^H$ and rewrite (4.3) as

$$\begin{aligned} \underset{\mathbf{D} \in \mathbb{C}^{M \times M}}{\text{minimize}} \quad & \|(\mathbf{A}^* \odot \mathbf{A})^H \operatorname{vec}(\mathbf{D}) - \mathbf{p}\|_2^2 \\ \text{subject to} \quad & \operatorname{rank}(\mathbf{D}) = 1, \quad \mathbf{D} \succeq 0 \end{aligned} \tag{4.4}$$

Problem (4.4) is still nonconvex due to the constraint $\operatorname{rank}(\mathbf{D}) = 1$. We can further relax this rank constraint and use the nuclear norm of $\mathbf{D}$ as a convex surrogate to obtain a relaxed version of (4.4) as follows

$$\begin{aligned} \underset{\mathbf{D} \in \mathbb{C}^{M \times M}}{\text{minimize}} \quad & \|(\mathbf{A}^* \odot \mathbf{A})^H \operatorname{vec}(\mathbf{D}) - \mathbf{p}\|_2^2 + \lambda \operatorname{tr}(\mathbf{D}) \\ & \mathbf{D} \succeq 0 \end{aligned} \tag{4.5}$$

where $\lambda > 0$ is a regularization parameter. Since (4.5) is a semi-definite program (SDP), it can be efficiently solved by off-the-shelf solvers such as SDPT3 [TTT99]. Let $\hat{\mathbf{D}}$ denote the optimal solution to (4.5). We can find the closest rank-1 approximation of $\hat{\mathbf{D}}$ by performing its singular value decomposition, and retaining the dominant singular vector [EY36]. Hence, an approximate solution for (4.3) is given as $\hat{\mathbf{c}}_0 = \eta_1$, and $\alpha_0 = \sigma_1$, where σ_1 is the largest singular value of $\hat{\mathbf{D}}$ and η_1 is the corresponding singular vector.

The vector $\hat{\mathbf{c}}_0$ can serve as an initial guess for the unconstrained beamforming vector $\mathbf{c}$, wheres α_0 is used as an initialization for α. In order to find initial estimates of $\mathbf{F}$, and $\mathbf{v}$ from $\hat{\mathbf{c}}_0$, we proceed by performing the greedy iterations given in Algorithm 2 (lines 6–11). Here, the notation $\mathcal{P}_\phi(\mathbf{x})$ is defined as follows

$$[\mathcal{P}_\phi(\mathbf{x})]_i = \frac{1}{\sqrt{M}} e^{j \measuredangle x_i}$$

where $\measuredangle(.)$ returns the phase of a complex number. Moreover, the notation $\Pi_{\mathbf{X}} = \mathbf{X}\mathbf{X}^\dagger$ indicates

the projection matrix onto the range space of the matrix $\mathbf{X}$, while mat$(.)$ rearranges the elements of an $M^2 \times 1$ vector into an $M \times M$ matrix, and svd$(.)$ denotes the singular value decomposition of a matrix.

Algorithm 2 Lifting Aided Greedy Initialization Algorithm

1: Given $\mathbf{p} \in \mathbb{R}^N$, solve (4.5).
2: $[\mathbf{U}_{\hat{\mathbf{d}}}, \Sigma_{\hat{\mathbf{d}}}, \mathbf{V}_{\hat{\mathbf{d}}}] = \text{svd}(\text{mat}(\hat{\mathbf{d}}))$
3: $\eta_1 := [\mathbf{U}_{\hat{\mathbf{d}}}]_{:,1}, \sigma_1 = [\Sigma_{\hat{\mathbf{d}}}]_{1,1}$
4: $\hat{\mathbf{c}}_0 = \eta_1, \alpha_0 = \sigma_1$
5: $\tilde{\mathbf{F}}^{(1)} \leftarrow \mathcal{P}_\phi(\hat{\mathbf{c}}_0)$
6: **for** i=2 to N_{RF} **do**
7: $\quad \tilde{\mathbf{f}}^{(i)} \leftarrow \hat{\mathbf{c}}_0 - \Pi_{\tilde{\mathbf{F}}^{(i-1)}} \hat{\mathbf{c}}_0$
8: $\quad \tilde{\mathbf{F}}^{(i)} \leftarrow [\tilde{\mathbf{F}}^{(i)}, \mathcal{P}_\phi(\tilde{\mathbf{f}}^{(i)})]$
9: **end for**
10: $\tilde{\mathbf{u}} \leftarrow (\tilde{\mathbf{F}}^{(N_{\text{RF}})})^\dagger \hat{\mathbf{c}}_0$
11: **return** $(\tilde{\mathbf{F}}^{(N_{\text{RF}})}, \tilde{\mathbf{u}})$

Parameterized Wirtinger Flow Based Iterations

After we find an initial guess for $\mathbf{F}, \mathbf{u}$, and α using Algorithm 2, we use a gradient descent based approach to find a local optimum for (4.2). However, since the objective function in (4.2) is a function of complex valued variables, we need to use Wirtinger calculus to compute the derivatives.

In order to enforce the constraints $\mathbf{f}_1, \cdots, \mathbf{f}_{N_{\text{RF}}} \in \mathcal{B}_M$, we introduce an equivalent parameterization of $\mathbf{F}$ as follows. Let Φ be an auxiliary variable which parameterizes $\mathbf{F}$ as

$$\mathbf{F}(\Phi) = \frac{1}{\sqrt{M}} e^{\circ(j\Phi)}$$

Here $e^{\circ(.)}$ represents applying the exponential function element-wise on a matrix. We also define the following functions:

$$\begin{aligned}
\mathbf{q}(\Phi) &= \mathbf{A}^H \mathbf{F}(\Phi)\mathbf{u} \\
\mathbf{t}(\Phi, \mathbf{u}, \mathbf{u}^*) &= |\mathbf{q}(\Phi)|^{\circ 2} \\
g(\Phi, \mathbf{u}, \mathbf{u}^*) &= \|\mathbf{F}(\Phi)\mathbf{u}\|^2 \\
\mathbf{h}(\Phi, \mathbf{u}, \mathbf{u}^*) &= \frac{\mathbf{t}(\Phi, \mathbf{u}, \mathbf{u}^*)}{g(\Phi, \mathbf{u}, \mathbf{u}^*)} \\
r(\Phi, \mathbf{u}, \mathbf{u}^*, \alpha) &= \|\alpha \mathbf{h}(\Phi, \mathbf{u}, \mathbf{u}^*) - \mathbf{p}\|^2
\end{aligned}$$

where $r(\Phi, \mathbf{u}, \mathbf{u}^*, \alpha)$ can be easily verified to be the objective function of the optimization problem (4.2). The goal is to calculate the complex derivatives of the functions $r(\Phi, \mathbf{u}, \mathbf{u}^*, \alpha)$ with respect to its parameters, denoted as $\mathcal{D}_\Phi r, \mathcal{D}_\mathbf{u} r, \mathcal{D}_{\mathbf{u}^*} r, \mathcal{D}_\alpha r$. According to the theory of optimization of functions with complex valued matrix parameters, the directions where the function $r(\Phi, \mathbf{u}, \mathbf{u}^*, \alpha)$ has the maximum rate of decay with respect to $\text{vec}(\Phi)$, and $\mathbf{u}$ are given by $(\mathcal{D}_{\Phi^*} r)^T$, and $(\mathcal{D}_{\mathbf{u}^*} r)^T$, respectively. Hence, we obtain the following update rules for gradient descent:

$$\begin{aligned}
\text{vec}(\Phi) &\leftarrow \text{vec}(\Phi) - \mu(\mathcal{D}_\Phi r)^T \\
\text{vec}(\mathbf{u}) &\leftarrow \text{vec}(\mathbf{u}) - \mu(\mathcal{D}_{\mathbf{u}^*} r)^T \\
\alpha &\leftarrow \alpha - \mu \mathcal{D}_\alpha r
\end{aligned}$$

A careful reader can refer to [HG07] to find the exact definition of complex derivatives. The following Lemma gives closed forms for the derivatives $\mathcal{D}_\Phi r, \mathcal{D}_{\mathbf{u}^*} r, \mathcal{D}_\alpha r$:

Lemma 23. *The derivatives $\mathcal{D}_{\Phi}r, \mathcal{D}_{\mathbf{u}^*}r$, can be computed by the following expressions:*

$$\begin{aligned}
\mathcal{D}_{\Phi}r &= -4(\alpha\mathbf{h}-\mathbf{p})^T Im\left\{\left(\frac{\mathrm{diag}(\mathbf{q}^*)(\mathbf{u}^T\otimes\mathbf{A}^H)}{\|\mathbf{F}\mathbf{u}\|^2}\right.\right.\\
&\qquad\left.\left.-\frac{|\mathbf{q}|^{\circ 2}}{\|\mathbf{F}\mathbf{u}\|^4}\mathrm{vec}(\mathbf{F}^*\mathbf{u}^*\mathbf{u}^T)^T\right)\mathrm{diag}(\mathrm{vec}(\mathbf{F}))\right\}\\
\mathcal{D}_{\mathbf{u}^*}r &= 2(\alpha\mathbf{h}-\mathbf{p})^T\left(\frac{\mathrm{diag}(\mathbf{q}^*)\mathbf{A}^H\mathbf{F}}{\|\mathbf{F}\mathbf{u}\|^2}-\frac{|\mathbf{q}|^{\circ 2}}{\|\mathbf{F}\mathbf{u}\|^4}\mathrm{vec}(\mathbf{F}^T\mathbf{F}^*\mathbf{u}^*)^T\right)^*\\
\mathcal{D}_{\alpha}r &= -\mathbf{h}^H\mathbf{p}+\alpha\|\mathbf{h}\|^2
\end{aligned}$$

Proof. The proof follows from application of chain rule, and derivatives of linear and quadratic functions of complex matrices from [HG07]. The detailed derivation of the given expressions is omitted due to lack of space. □

4.1.3 Simulations

In the first set of simulations, we consider a codebook design problem similar to [SCL15b, SCL17]. Following [SCL15b], we consider two cases: $B = 2$ and $B = 3$, where 2^B denotes the number of codewords. The qth codeword $(\mathbf{F}_q, \mathbf{v}_q)$ is designed such that the corresponding beam-pattern is equal to

$$\mathbf{p} = \frac{2^B}{M}(\mathbf{e}_q \otimes \mathbf{1}_L) \tag{4.6}$$

where $L = 50$, $N = L2^B$, and $\mathbf{e}_q \in \mathbb{R}^{2^B}$ denotes the qth standard basis vector, and $\mathbf{1}_L \in \mathbb{R}^L$ represents the vector of all ones. Such a beampattern could be useful during the channel sounding stage in a massive MIMO communication system [SCL15b]. In all plots, we assume we have $N_{\mathrm{RF}} = 8$ available RF-chains. Fig. 4.2 (resp. Fig. 4.3) compares the codebook of size $2^B = 4$ (resp. $2^B = 8$) generated by our algorithm with that of [SCL15b] (we search over the set $\mathcal{G}_M^N$, defined in [SCL15b], with $M = 8, N = 3$). We see that in all cases, our proposed codebook shows a sharper roll-off compared to [SCL15b]. Also, for cases that the beampatterns need to be more

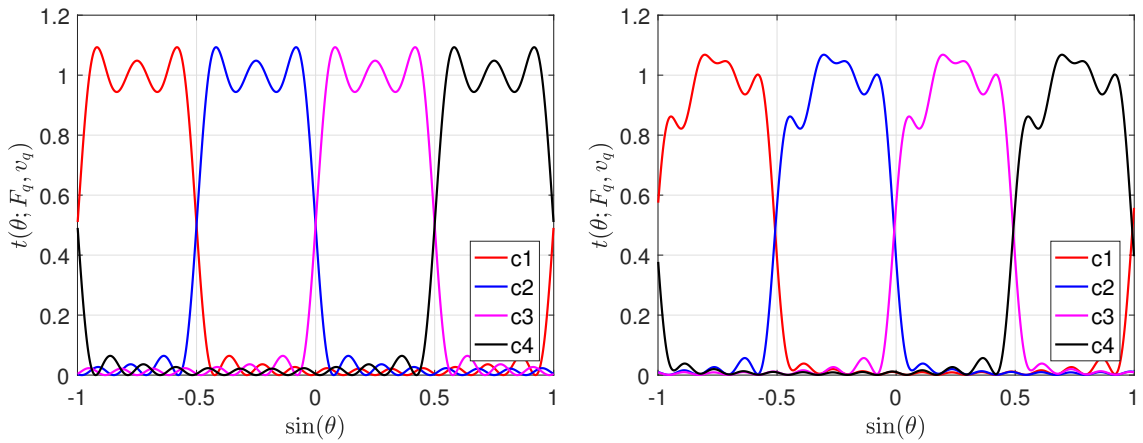
flat over the passband (i.e, Fig. 4.2 (c), (d)), our algorithm generates a cleaner beampattern with less ripple.

Remark 28. *In terms of computational complexity, a key distinction between the algorithm in [SCL15b, SCL17], and our proposed algorithm is that in [SCL15b, SCL17], the authors perform an exhaustive search over all possible ways that one can assign phases to the elements of the vector* γ*, such that* $\mathbf{p} = \gamma \circ \gamma^*$*, by quantizing the phases into* N_q *levels for all possible* L *nonzero elements in* $\mathbf{p}$*, which has the complexity at least* N_q^{L-1} *(although they use a more tractable set* $\mathcal{G}_M^N$*, the complexity is still exponential). This exhaustive search can have prohibitive complexity as the number of nonzero elements of* $\mathbf{p}$ *increases, making it intractable to solve for beampatterns with large number of nonzero elements. However, in this Section, we do not require any exhaustive search, making it possible to solve for beampatterns with arbitrarily large number of nonzero elements. One of the computational bottlenecks of our proposed algorithm is the initialization step, which requires solving an SDP, which has the complexity at least* $\Omega(M^3)$ *[NJS13]. In future, we will investigate how we can reduce the complexity of our algorithm, and also provide convergence analysis for the proposed Wirtinger Flow iterations.*

Remark 29. *Another distinction of our proposed method from that in [SCL15b, SCL17] is that the beam pattern* $\mathbf{p}$ *can be defined by any arbitrary vector, as long as it satisfies certain power constraints. However, the beampattern in [SCL15b, SCL17] can only be of the form given in* (4.6)*.*

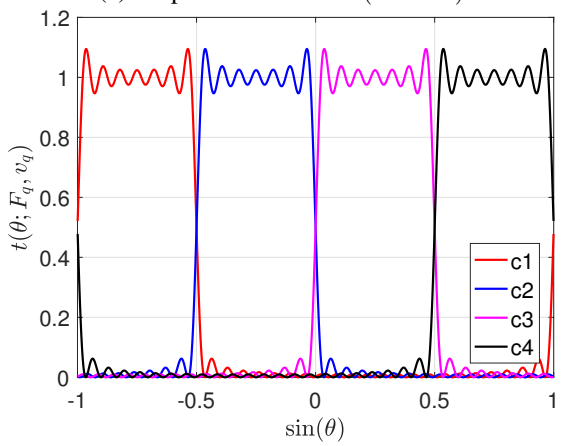
4.1.4 Conclusion

In this Section, we proposed a beam-pattern design algorithm for hybrid beamforming. We followed the same problem formulation as in [SCL15b], but proposed a novel non-convex algorithm based on Wirtinger flow framework. Our optimization procedure is based on carefully finding an initialization point by solving a suitable convex optimization problem (using ideas

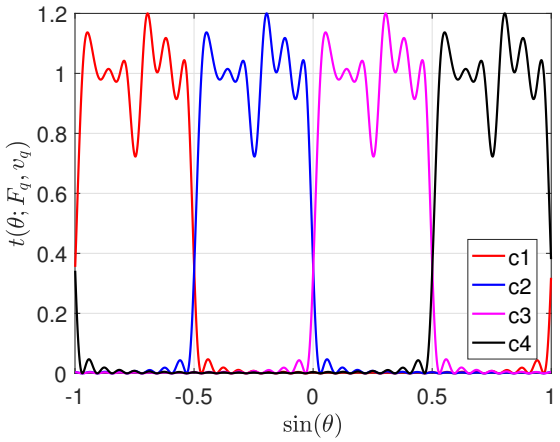


(a) Proposed Codebook $(M = 16)$

(b) Codebook in [SCL15b] $(M = 16)$



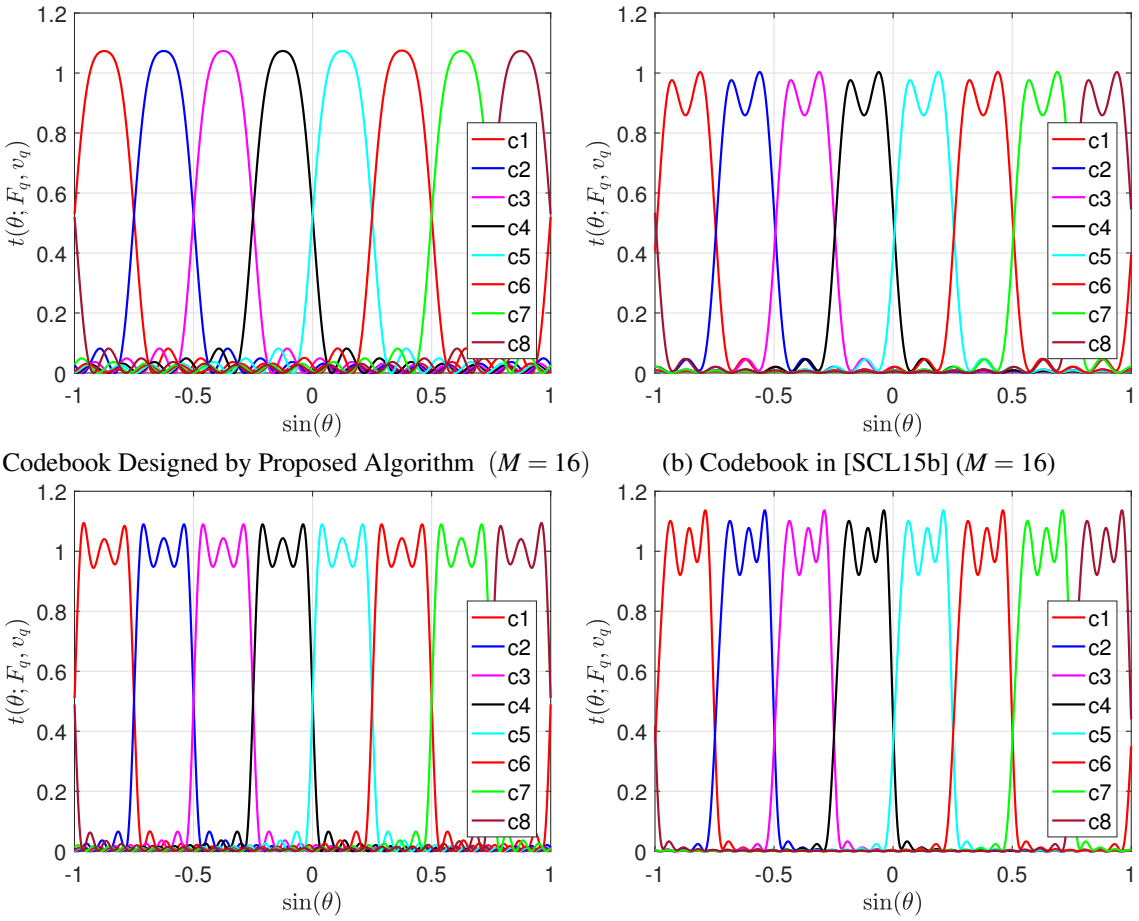
(c) Proposed Codebook $(M = 32)$



(d) Codebook in [SCL15b] $(M = 32)$

Figure 4.2: Codebook Designed by Proposed Algorithm versus those designed by [SCL15b], for a codebook of size $2^B = 4$.

from lifting), followed by gradient descent. The gradients are calculated using Wirtinger calculus, since the objective function is in terms of complex valued matrices. Unlike the algorithm in [SCL15b], our algorithm does not have exponential computational complexity in terms of the size of the beamwidth and can be used for arbitrary beampattern synthesis with arrays of any geometry.



(a) Codebook Designed by Proposed Algorithm $(M = 16)$ (b) Codebook in [SCL15b] $(M = 16)$

(c) Codebook Designed by Proposed Algorithm $(M = 32)$ (d) Codebook in [SCL15b] $(M = 32)$

Figure 4.3: Codebook Designed by Proposed Algorithm versus those designed by [SCL15b], for a codebook of size $2^B = 8$.

4.2 Canonical Polyadic (CP) Decomposition of Structured Semi-Symmetric Fourth-Order Tensors

In this section, we consider decomposition of tensors which have *Vandermonde-like* (corresponding to sub-selected rows of Vandermonde matrices, according to geometries such as nested sampling [PV10]) factors as well as unstructured factors. We explain the model in detail in Section 4.2.1. Such tensors can arise in certain DOA estimation applications using non-uniform arrays where the sources are assumed to transmit uncorrelated symbols with a CDMA (code-division multiple access) encoding. We show that using purely linear algebraic algorithms, it is possible to recover $O(M^2N)$ factors, from a $M \times N \times M \times N$ tensor, where M is the dimension of the structured factors, and N is the dimension of the unstructured factors. Our proposed method uses the properties of difference co-arrays to *lift* the given tensor into a higher dimensional tensor, and then employs an ESPRIT-type algorithm (inspired by [SDL17a]), which uses shift-invariance property of the Vandermonde matrices in order to decompose the tensor factors. Our result differs from [SDL17a] due to the following reasons: (i) we consider semi-symmetric tensors where two of the Vandermonde-like factors are complex conjugate of one another, whereas guarantees of [SDL17a] hold for distinct factors, (ii) our factors are not Vandermonde but consist of sub-selected rows of Vandermonde matrices, (iii) two of our factors are unstructured, compared to one unstructured factor of [SDL17a].

4.2.1 Problem Definition

We consider a semi-symmetric fourth order tensor of dimensions $M \times N \times M \times N$, whose factors are assumed to be constrained in the first two dimensions and unconstrained in the other two dimensions. The tensor is semi-symmetric in the sense that the two structured (as well as unstructured) factors are complex conjugate of one another. In particular, we assume that

$\mathcal{T} \in \mathbb{C}^{M\times N\times M\times N}$ has a decomposition as

$$\mathcal{T} = \sum_{k=1}^{r} p_k \mathbf{a}_k \otimes \mathbf{b}_k \otimes \mathbf{a}_k^* \otimes \mathbf{b}_k^*, \tag{4.7}$$

where $p_k \in \mathbb{R}$, $\|\mathbf{b}_k\|_2 = 1$, $\|\mathbf{a}_k\|_2 = \sqrt{M}$. We assume that the factors $\mathbf{b}_k \in \mathbb{C}^N$ can be any (generic) unit-norm unconstrained vector. However, the factors $\mathbf{a}_k \in \mathbb{C}^M$ are assumed to be sub-selected rows of Vandermonde vectors, such that the ith element of $\mathbf{a}_k$ is given as

$$[\mathbf{a}_k]_i = e^{jz_i\omega_k} \tag{4.8}$$

where $z_i \in \mathbb{Z}$ are integers, and $j = \sqrt{-1}$. Correspondingly, define $\mathbf{A} \in \mathbb{C}^{M\times r}, \mathbf{B} \in \mathbb{C}^{N\times r}$, as

$$\mathbf{A} = [\mathbf{a}_1, \cdots, \mathbf{a}_r]$$
$$\mathbf{B} = [\mathbf{b}_1, \cdots, \mathbf{b}_r]$$

Such a Vandermonde-like structure for matrix $\mathbf{A}$ can arise in applications such as array signal processing with non-uniform arrays [PV10, PV11].

4.2.2 Khatri-Rao Product and Role of Difference Sets

In this subsection, we will review the concept of difference sets which will be later used in Section 4.2.3 in order to construct a higher dimensional tensor $\tilde{\mathcal{T}}$ by rearranging and repeating the entries of $\mathcal{T}$ in a certain way.

For integer numbers $z_1, z_2, \cdots, z_M$, define the indexed-difference set as [PV12c]

$$\mathcal{D}_I = \{(i, j, z_i - z_j), 1 \le i, j \le M\}$$

Moreover, define $\tilde{M}$ to be the largest integer such that the differences $z_i - z_j$ span the range of

consecutive integers $-(\tilde{M}-1),\cdots,\tilde{M}-1$.

The difference sets arise naturally in the exponents of the matrix $\mathbf{A}^* \odot \mathbf{A}$ (interested reader can refer to [PV10]). We define the matrix $\tilde{\mathbf{A}} \in \mathbb{C}^{\tilde{M}\times r}$ to be a Vandermonde submatrix of $\mathbf{A}^* \odot \mathbf{A}$ corresponding to the largest consecutive nonnegative differences in $\mathcal{D}_I$. In particular, we let the columns of $\tilde{\mathbf{A}}$ to be

$$[\tilde{\mathbf{a}}_k]_m = e^{j(m-1)\omega_k}, \;\; 1 \leq m \leq \tilde{M}-1 \tag{4.9}$$

and $\tilde{\mathbf{A}} = [\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2, \cdots, \tilde{\mathbf{a}}_r] \in \mathbb{C}^{\tilde{M}\times r}$.

In the context of array signal processing, this corresponds to the largest (nonnegative) ULA in the virtual co-array, which is used to perform co-array MUSIC [PV10]. Compared to the difference set introduced in [PV12c, PV10], $\mathcal{D}_I$ also keeps track of the pairs of indices (i,j) which give rise to each particular difference $(z_i - z_j)$. In Section 4.2.3, we will use these indices in order to *lift* the tensor $\mathcal{T}$ to a higher dimensional tensor.

For particular array geometries, it has been shown that one can achieve $\tilde{M} = O(M^2)$. The well-known designs which can achieve $\tilde{M} = O(M^2)$ include nested arrays [PV10], coprime arrays [PV11], minimum redundancy arrays [Mof68], etc. For example, a two-level nested array with an even number of antennas (even M) has the following structure [PV10]:

$$z_i = \begin{cases} i, & 1 \leq i \leq \dfrac{M}{2} \\ (\dfrac{M}{2}+1)i, & \dfrac{M}{2}+1 \leq i \leq M \end{cases}$$

It turns out that for such an array geometry, we have $\tilde{M} = \frac{M^2}{4} + \frac{M}{2} - 1$.

4.2.3 Proposed Algorithm

In this section, we will show that although the factor $\mathbf{A}$ is not exactly a Vandermonde

matrix, it is possible to *lift* the tensor $\mathcal{T}$ to a higher dimensional tensor $\tilde{\mathcal{T}}$, whose decomposition contains Vandermonde factors. We then employ an ESPRIT-type algorithm (similar to [SDL17a]) on the *lifted* tensor $\tilde{\mathcal{T}}$ to find the CP decomposition.

Lifting the tensor to a higher dimension

As we observed in Section 4.2.2, if the locations $z_1, \cdots, z_M$ are carefully designed, the size of the difference set $\mathcal{D}_I$ can be as large as $O(M^2)$. We will see how this can help us in decomposing $O(M^2N)$ factors from the tensor $\mathcal{T}$. We define this *lifting* operation as follows:

Definition 11. *Given a tensor $\mathcal{T} \in \mathbb{C}^{M\times N\times M\times N}$, the 4th order tensor* $\textsc{TensorToep}^{(\mathcal{D}_I)}(\mathcal{T}) \in \mathbb{C}^{\tilde{M}\times N\times \tilde{M}\times N}$ *is constructed as*

$$[\textsc{TensorToep}^{(\mathcal{D}_I)}(\mathcal{T})]_{p,n_1,q,n_2} = \mathcal{T}_{m_1,n_1,m_2,n_2}$$

where $(m_1, m_2, p-q) \in \mathcal{D}_I$, $1 \le n_1, n_2 \le N$, $1 \le m_1, m_2 \le M$. In presence of noise, we may average over the redundancies in $\mathcal{D}_I$.

Since depending on the design of $z_1, \cdots, z_M$, we can have $\tilde{M} \gg M$, the tensor $\textsc{TensorToep}^{(\mathcal{D}_I)}(\mathcal{T})$ can potentially have much larger dimensions than $\mathcal{T}$. This will enable us to decompose tensors of much higher ranks. The following Lemma shows the decomposition corresponding to this *lifted* tensor:

Lemma 24. *It holds that for $\mathcal{T}$ defined in* (4.7)

$$\textsc{TensorToep}^{(\mathcal{D}_I)}(\mathcal{T}) = \sum_{k=1}^{r} p_k \tilde{\mathbf{a}}_k \otimes \mathbf{b}_k \otimes \tilde{\mathbf{a}}_k^* \otimes \mathbf{b}_k^*$$

where $\tilde{\mathbf{a}}_k \in \mathbb{C}^{\tilde{M}}$ is the kth column of $\tilde{\mathbf{A}}$ defined in (4.9).

Proof. Proof follows from applying definition 11 to the entries of $\mathcal{T}$ defined in (4.7). □

If we did not use the structure of the difference set, it would still be possible to recover $r = O(MN)$ factors using a variant of the algorithm proposed in [DLCC07] (given certain genericity assumptions hold). However, as we will see in Section 4.2.3, using the shift-invariance property of the Vandermonde matrices, it is possible to decompose tensors of rank $r = O(M^2N)$.

We finish this section by providing one more definition, which will be used to rearrange the elements of the tensor $\text{TENSORTOEP}^{(\mathcal{D}_I)}(\mathcal{T})$ into a square matrix:

Definition 12. *Given $\tilde{\mathcal{T}}$, let $\text{FLATTEN}(\tilde{\mathcal{T}}) \in \mathbb{C}^{\tilde{M}N \times \tilde{M}N}$ be such that its (i,j)th element is given as*

$$[\text{FLATTEN}(\tilde{\mathcal{T}})]_{i,j} = [\tilde{\mathcal{T}}]_{m,n_1,t,n_2}$$

such that $i = m + (n_1 - 1)\tilde{M}$, $j = t + (n_2 - 1)\tilde{M}$, $1 \leq m,t \leq \tilde{M}, 1 \leq n_1, n_2 \leq N$.

ESPRIT-type Algorithm for Tensor Decomposition

In this Section, we will employ an ESPRIT-type algorithm (similar to [SDL17a]) which uses the shift-invariance property of Vandermonde matrices in order to find the tensor decomposition for $\text{TENSORTOEP}^{(\mathcal{D}_I)}(\mathcal{T})$.

Using the definitions given in Section 4.2.1, define $\tilde{\mathbf{T}} \in \mathbb{C}^{\tilde{M}N \times \tilde{M}N}$ to be such that $\tilde{\mathbf{T}} = \text{FLATTEN}(\text{TENSORTOEP}^{(\mathcal{D}_I)}(\mathcal{T}))$. Following the definitions of the operations $\text{FLATTEN}(.)$, and $\text{TENSORTOEP}^{(\mathcal{D}_I)}(.)$, and Lemma 24, we have that

$$\tilde{\mathbf{T}} = (\mathbf{B} \odot \tilde{\mathbf{A}})\mathbf{P}(\mathbf{B} \odot \tilde{\mathbf{A}})^H$$

where $\mathbf{P} = \text{diag}(p_1, \cdots, p_r)$. Let $\mathbf{U}\Sigma\mathbf{V}^H$ be the singular value decomposition of $\tilde{\mathbf{T}}$. If the matrix $\mathbf{B} \odot \tilde{\mathbf{A}}$ has full column rank, there exists a nonsingular matrix $\mathbf{F} \in \mathbb{C}^{r \times r}$ such that

$$\mathbf{UF} = \mathbf{B} \odot \tilde{\mathbf{A}} \tag{4.10}$$

The following Corollary shows the conditions under which $\mathbf{B} \odot \tilde{\mathbf{A}}$ has full column rank:

Corollary 6. *(Corollary to [SDL13, Lemma III.1]) Let* $\mathbf{A}$ *be sub-selected rows of a Vandermonde matrix,* $\tilde{\mathbf{A}}$ *be defined as in* (4.9)*, and* $\mathbf{B} \in \mathbb{C}^{N \times r}$*. The matrix* $\mathbf{B} \odot \tilde{\mathbf{A}}$ *has rank* $\min(\tilde{M}N, r)$ *for almost all* $\mathbf{A}$*,* $\mathbf{B}$*.*

Our goal is to find $\mathbf{F}$ such the columns of $\mathbf{UF}$ are vectorized form of rank-1 matrices of the form $\tilde{\mathbf{a}}(\omega)\mathbf{b}^T$ (as in RHS of (4.10)), where $[\tilde{\mathbf{a}}(\omega)]_m = e^{j(m-1)\omega}$. It turns out that such rank-1 matrices satisfy the following row-shift-invariance property:

Definition 13. *[SDL17a] For a matrix* $\mathbf{X} \in \mathbb{C}^{N_1 \times N_2}$*, we say it has row-shift-invariance property if*

$$\underline{\mathbf{X}} = \alpha \overline{\mathbf{X}} \tag{4.11}$$

for some $\alpha \in \mathbb{C}$*,* $\overline{\mathbf{X}}, \underline{\mathbf{X}} \in \mathbb{C}^{(N_1-1) \times N_2}$ *are the matrices obtained by removing the first, and last row of* $\mathbf{X}$*, respectively.*

One can verify that if $\mathbf{X} = \tilde{\mathbf{a}}(\omega)\mathbf{b}^T$, then we have $\alpha = e^{-j\omega}$ in (4.11). We will use this property of the columns of $\mathbf{UF}$ in order to identify the matrix $\mathbf{F}$.

We follow a procedure similar to ESPRIT algorithm [RK89, SDL17a, SDL17b]. Let $\mathbf{U}^{(1)}, \mathbf{U}^{(\tilde{M})} \in \mathbb{C}^{(\tilde{M}-1)N \times r}$ be such that their ith column are given as

$$\begin{aligned} \overline{\mathbf{u}}_i^{(1)} &= \text{VEC}\big(\overline{\text{UNVEC}(\mathbf{u}_i)}\big), \\ \overline{\mathbf{u}}_i^{(\tilde{M})} &= \text{VEC}\big(\underline{\text{UNVEC}(\mathbf{u}_i)}\big) \end{aligned}$$

where $\mathbf{u}_i$ denotes the ith column of $\mathbf{U}$, $\text{UNVEC}(.)$ rearranges the elements of $\mathbf{u}_i$ into a $\tilde{M} \times N$ matrix, and $\text{VEC}(.)$ rearranges the elements of a $(\tilde{M}-1) \times N$ matrix into a vector. Here, the notations $\overline{\mathbf{X}}$ and $\underline{\mathbf{X}}$ denote the removal of first and last rows of matrix $\mathbf{X}$.

Using the row-shift-invariance property of the columns of $\mathbf{B} \odot \tilde{\mathbf{A}}$ (as discussed in Definition

13), we have $\underline{\tilde{\mathbf{A}}} = \overline{\tilde{\mathbf{A}}}\mathbf{D}$, where $\mathbf{D} = \text{diag}(e^{-j\omega_1}, \cdots, e^{-j\omega_r})$. Therefore, it holds that

$$\mathbf{B} \odot \underline{\tilde{\mathbf{A}}} = (\mathbf{B} \odot \overline{\tilde{\mathbf{A}}})\mathbf{D}$$

However, from definition of $\mathbf{F}$ and $\mathbf{U}^{(1)}, \mathbf{U}^{(\tilde{M})}$ we have

$$\mathbf{B} \odot \underline{\tilde{\mathbf{A}}} = \mathbf{U}^{(1)}\mathbf{F}$$
$$\mathbf{B} \odot \overline{\tilde{\mathbf{A}}} = \mathbf{U}^{(\tilde{M})}\mathbf{F}$$

which means that $\mathbf{U}^{(1)}\mathbf{F} = \mathbf{U}^{(\tilde{M})}\mathbf{F}\mathbf{D}$. Assuming that $\mathbf{B} \odot \overline{\tilde{\mathbf{A}}}$ is full-column rank (which holds for almost all $\mathbf{A}, \mathbf{B}$ if $r \leq (\tilde{M}-1)N$, due to Corollary 6), we can find $\mathbf{F}$ from the following eigenvalue decomposition

$$(\mathbf{U}^{(\tilde{M})})^{\dagger}\mathbf{U}^{(1)} = \mathbf{F}\mathbf{D}\mathbf{F}^{-1}$$

Once the matrix $\mathbf{F}$ is computed, we can find the vectors $\tilde{\mathbf{a}}_i$ (resp. $\mathbf{b}_i$ upto global phase ambiguity) by considering the largest left (resp. right) singular vectors of the matrices $\text{UNVEC}(\tilde{\mathbf{U}}\mathbf{f}_i)$, $1 \leq i \leq r$, with $\mathbf{f}_i$ denoting the ith column of $\mathbf{F}$). Using the fact that the first entry of $\tilde{\mathbf{a}}_k$ is always equal to one, $\|\tilde{\mathbf{a}}_k\| = \sqrt{\tilde{M}}$, $\|\mathbf{b}_k\| = 1$, p_k can be computed. We summarize our result as the following theorem:

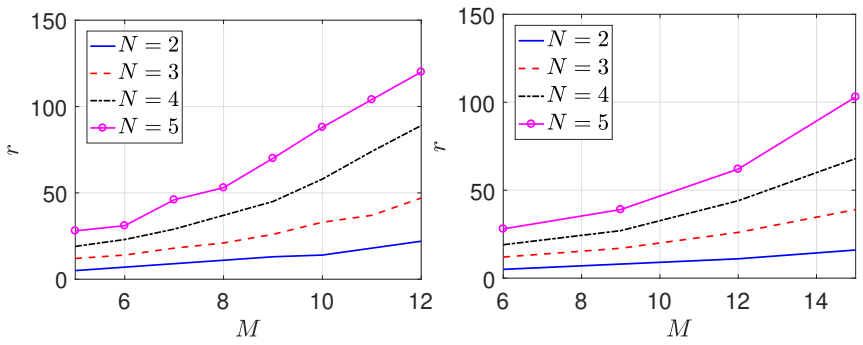
Theorem 26. *Let* $\mathbf{A}, \mathbf{B}$ *be such that* $\mathbf{B} \odot \overline{\tilde{\mathbf{A}}} \in \mathbb{C}^{(\tilde{M}-1)N \times r}$ *is full column rank (which holds for almost all* $\mathbf{A}, \mathbf{B}$ *if* $r \leq (\tilde{M}-1)N$*). Then, the decomposition* (4.7) *can uniquely be found using a linear algebraic algorithm (upto global phase ambiguity for all entries of each* $\mathbf{b}_k$*).*

Remark: Compared to our result proposed in [KP18c], which considers a more general case of $(2q+2d)$th order tensors (with $2q$ structured and $2d$ unstructured factors), the algorithm proposed in this Section is much simpler and is tailored to the specific case of $d = q = 1$. We defer proposing a more unifying framework to the future.

4.2.4 Simulations

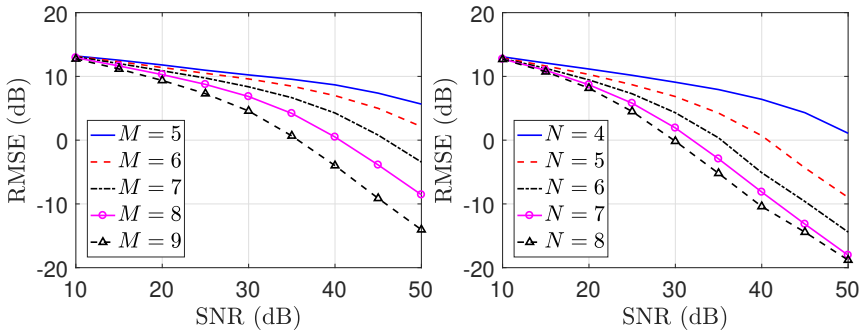
In this section, we will provide numerical simulations to examine the performance of our proposed algorithm. We consider 4th order tensors with CP-decomposition given as in (4.7). In the first set of simulations, we consider two different sparse array geometries, namely nested and coprime arrays. We assume that the frequencies corresponding to structured factors $\mathbf{a}_i$ are given as $\omega_i = \pi \sin(\theta_i)$, such that θ_i, $1 \leq i \leq r$ lies on a predefined uniform grid. The elements of the non-structured factors $\mathbf{b}_i$ are chosen i.i.d from circularly symmetric complex Gaussian distribution with zero mean and unit variance. We define the recovery error as $\varepsilon^2 := \|\mathbf{A} - \hat{\mathbf{A}}\hat{\Pi}\|_F^2/M + \|\mathbf{B} - \hat{\mathbf{B}}\hat{\Pi}\hat{\Lambda}\|_2^2/N$, where $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$ are the recovered matrices containing the factors of the tensor, and $\hat{\Pi}$ is an appropriate permutation matrix which minimizes ε, and $\hat{\Lambda}$ is an appropriate diagonal matrix which resolves global phase ambiguity. If for given dimensions M, N and rank r, we obtain $\varepsilon \leq 10^{-4}$ over 10 different runs, we denote that as a "successful" tensor decomposition for those M, N, r. The plots in Figure 4.4 show the maximum rank r for which "successful" recovery can be attained for a given M, and N. In Figure 4.4a, the structured factors are chosen such that z_m lie on nested array, while in Figure 4.4b, z_m follow a coprime geometry with coprime numbers $N_1 = M/3, N_2 = M/3 + 1$ (only evaluated for M such that $M/3$ is an integer).

In our second experiment, we consider tensor decomposition in presence of additive noise. We define signal-to-noise ratio (SNR) to be $10\log_{10}(\sum_{i=1}^{r} p_i/\sigma^2)$, where σ^2 denotes the variance of the additive circularly symmetric complex Gaussian noise to the original tensor $\mathcal{T}$. The factors $\mathbf{a}_i, \mathbf{b}_i$ are generated similar to the first experiment, and only the nested array geometry is used. In this case, the operation $\textsc{TensorToep}^{(\mathcal{D}_I)}(.)$ is slightly modified so that the the tensor elements are avaraged over redundancies in the difference set $\mathcal{D}_I$ (an interested reader can refer to the concept of weight functions in [PV10]). The plots in Figure 4.5, show the average recovery error ε (denoted as RMSE) in log-scale, with respect to the SNR. As we can see, increasing the SNR leads to more accuracy of the recovered factors, and the overall trend of the RMSE with respect



(a) Nested array (b) Coprime array

Figure 4.4: Maximum tensor rank that our proposed algorithm can successfully decompose, in a noiseless setting, for 4th order tensors whose structured factors have nested or coprime geometry.



(a) $M = 8$ (b) $N = 5$

Figure 4.5: The average RMSE of the proposed algorithm in presence of additive noise, with respect to SNR.

to SNR shows the stability of our tensor decomposition algorithm with respect to the noise. We defer theoretical analysis of the stability of our algorithm to future. In all simulations, $r = 14$ is fixed. As we can see, larger M, and N can lead to more accurate recovery.

4.2.5 Conclusion

We considered Canonical Polyadic (CP) decomposition of fourth order tensors which are assumed to have both structured and unstructured factors. Motivated by applications in array signal processing for nonuniform arrays, the structured factors were assumed to be sub-selected

rows of Vandermonde matrices. We showed that by employing linear algebraic techniques and the properties of difference co-array in nonuniform arrays, it is possible to recover $O(M^2N)$ tensor factors from a $M \times N \times M \times N$ tensor, with M being the dimension of the structured factors and N corresponding to unstructured ones. Numerical simulations were provided to examine the performance of our proposed method.

4.3 Channel Estimation for Hybrid MIMO Communication with (Non-)Uniform Linear Arrays via Tensor Decomposition

In this Section, we develop a novel gridless channel estimation algorithm which converts the channel estimation problem into a Canonical Polyadic tensor Decomposition (CPD) problem, for which a long-established literature with powerful theoretical as well as experimental results exists [DL06, SPF14]. Unlike the other existing tensor based approaches for mmWave channel estimation which use multi-carrier measurements [ZFY$^+$17, AdA17, AdAH19, PAGH19], specific designs for the pilot signals [QFS19], or dual-polarized antenna arrays [QFSY18], we only use measurements from a single carrier and take advantage of statistical properties of the channel to reframe the channel estimation problem as an instance of a CPD problem. We also show that it is possible to estimate mmWave channels for communication systems whose antenna arrays have a non-uniform geometry. Non-uniform arrays can result from either non-uniform placement of the antennas, or due to switching-off subsets of transmit and receive antennas of a large uniform linear array (through a process of antenna selection) in order to to save power. In the context of array signal processing, it has been established that non-uniform arrays with specific designs can have huge benefits over uniform arrays that employ the same number of antennas [PV10, PV11]. In this Section, we utilize the difference co-array structure of non-uniform antennas and show that even if both the transmitter and the receiver use non-uniform arrays, it is possible to estimate channels comprising of a maximum of $L = O(N_T N_R)$ channel paths, where N_T, N_R are the number of RF-chains in the transmitter and the receiver, respectively. Through our numerical simulations, we further show that non-uniform arrays achieve a smaller error for channel estimation compared to linear arrays, mainly due to the enlarged size of their difference co-arrays.

4.3.1 Signal Model

We consider a single user mmWave massive MIMO communication system, where the transmitter is equipped with M_T antennas and N_T RF-chains, and the receiver has M_R antennas and N_R RF-chains, where $N_T \ll M_T, N_R \ll M_R$. The RF-chains are connected to the antennas through a network of phase-shifters. We define $\mathbf{F} \in \mathbb{C}^{N_T \times M_T}$ to be the matrix of phase-shifters for the transmitter, whose elements are unimodular whose elements satisfy $\|F_{i,j}\| = 1/\sqrt{M_T}, \forall i, j$. Also, let $\mathbf{W} \in \mathbb{C}^{N_R \times M_R}$ to be matrices of analog phase-shifters for the receiver, such that $\|W_{i,j}\| = 1/\sqrt{M_R}, \forall i, j$.

The channel sensing is performed over several blocks of time, known as fading blocks. Within each block, the channel is assumed to remain constant. During the qth fading block we assume that the mmWave channel is characterized as [XGJ16]

$$\mathbf{H}(q) = \mathbf{A}_R(\theta)\Lambda(q)\mathbf{A}_T^T(\Phi)$$

where $\mathbf{A}_R(\theta) \in \mathbb{C}^{M_R \times L}, \mathbf{A}_T(\Phi) \in \mathbb{C}^{M_T \times L}$ are the array manifold matrices corresponding to the receiver and the transmitter respectively, which are assumed to remain constant during all the sensing blocks. The array manifold matrix for the receiver is defined as $\mathbf{A}_R(\theta) = [\mathbf{a}_R(\theta_1), \cdots, \mathbf{a}_R(\theta_L)]$, where the mth element of the steering vector $\mathbf{a}_R(\theta_i)$ equals $e^{j\pi \sin(\theta_i) d_m^{(R)}}$ ($j = \sqrt{-1}$). Here, $d_m^{(R)}, 1 \leq m \leq M_R$ denote the sensor locations for the receiver antenna array, normalized with respect to $\bar{\lambda}/2$, where $\bar{\lambda}$ is the carrier wavelength. We make similar definitions for the transmitter antenna array, with $[\mathbf{a}_T(\phi_i)]_m = e^{j\pi \sin(\phi_i) d_m^{(T)}}$. The matrix of channel gains $\Lambda(q) = \text{diag}(\lambda(q)) \in \mathbb{C}^{L \times L}$ is assumed to be a diagonal matrix comprising of channel path gains $\lambda(q)$ at the qth fading block. Moreover, we assume that $\lambda(q)$ follow a circularly symmetric Gaussian distribution $\lambda(q) \sim \mathcal{N}(\mathbf{0}, \mathbf{P})$, where $\mathbf{P} \in \mathbb{R}^{L \times L}$ is the covariance matrix of the channel gains. The channel gains are also assumed to be statistically uncorrelated, i.e. $\mathbf{P} = \text{diag}(\mathbf{p})$ is a diagonal matrix, which is a common assumption for mmWave channels in many scenarios [PH16, PH18, AGPHJ18].

Pilot-Aided Channel Sensing: Preliminaries

We assume a similar methodology for pilot-aided channel sensing as [GWH17]. For completeness, we briefly review this procedure from [GWH17] before describing our proposed algorithm in Sec. 4.3.2. Let $\phi(t)$ be a pulse-shaping function of duration T, such that $\int_0^T |\phi(t)|^2 dt = 1$, and $s_n(t) = \sqrt{\frac{E}{N_T}}\phi(t-(n-1)T)$ be the pilot signal used by the nth RF chain at the transmitter at time t, with the property that

$$\int_0^{N_T T} s_{n_1}(t) s_{n_2}(t) dt = \frac{E}{N_T}\delta_{n_1,n_2}$$

with $\delta_{n_1,n_2} = 1$ if $n_1 = n_2$ and $\delta_{n_1,n_2} = 0$ otherwise. Here E is the transmit power equally allocated for each pilot. During each fading block, the transmitter transmits N_T pilot waveforms $s_1(t), \cdots, s_{N_T}(t)$. The receiver properly applies matched filters on the signals received on its RF-chains. The output of the matched filter corresponding to nth pilot waveform, and qth fading block will be [GWH17]:

$$\mathbf{y}_n(q) = \frac{E}{N_T}\mathbf{W}\mathbf{H}(q)(\mathbf{F}_{[n,:]})^T + \mathbf{z}_n(q)$$

where $\mathbf{z}_n(q) \in \mathbb{C}^{N_T}$ is the additive noise with circularly symmetric Gaussian distribution $\mathcal{N}(\mathbf{0}, \sigma^2\mathbf{W}\mathbf{W}^H)$, and $\mathbf{F}_{[n,:]}$ denotes the nth row of the matrix $\mathbf{F}$. Stacking all the vectors $\mathbf{y}_n(q)$ into a longer vector $\mathbf{y}(q) = [\mathbf{y}_1(q)^T, \mathbf{y}_2(q)^T, \cdots, \mathbf{y}_{N_T}(q)^T]^T \in \mathbb{C}^{N_T N_R}$, the covariance matrix $\mathbf{R}_\mathbf{y} = \mathbb{E}(\mathbf{y}(q)\mathbf{y}(q)^H)$ is given by

$$\mathbf{R}_\mathbf{y} = \mathbf{B}\mathbf{P}\mathbf{B}^H + \sigma^2\mathbf{I}_{N_T \times N_T} \otimes \left(\mathbf{W}\mathbf{W}^H\right) \tag{4.12}$$

where $\mathbf{B} := (\mathbf{F}\mathbf{A}_T) \odot (\mathbf{W}\mathbf{A}_R)$. The covariance matrix $\mathbf{R}_\mathbf{y}$ can be estimated using the sample covariance matrix:

$$\mathbf{R}_\mathbf{y} \simeq \frac{1}{T} \sum_{q=1}^{T} \mathbf{y}(q)\mathbf{y}(q)^H, \tag{4.13}$$

where T is the number of channel fading blocks used for training. The goal is to identify the channel parameters θ_i, ϕ_i, for $i = 1, 2, \cdots, L$ by utilizing the structure of the covariance matrix. It is to be noted that owing to the presence of the compressive (fat) matrices $\mathbf{W}, \mathbf{F}$, standard DoA estimation algorithms such as MUSIC, ESPRIT cannot be directly applied.

4.3.2 Channel Estimation with Co-Array via CP Decomposition

In this section we propose a novel channel estimation technique inspired by CPD algorithms, that proceeds in two stages: (i) Identify the factors $\mathbf{W}\mathbf{A}_R$ and $\mathbf{F}\mathbf{A}_T$ using fourth-order CPD methods, and (ii) Identify the AoAs θ and AoDs Φ by utilizing the co-array structure of the transmit and receive antenna arrays. Assuming that the noise variance σ^2 is known, we can rearrange[2] the elements of $\mathbf{R}_\mathbf{y} - \sigma^2 \mathbf{I}_{N_T \times N_T} \otimes (\mathbf{W}\mathbf{W}^H)$ into a fourth order tensor $\mathcal{R} \in \mathbb{C}^{N_R \times N_T \times N_R \times N_T}$, given by

$$\mathcal{R} = \sum_{i=1}^{L} p_i \alpha_i \circ \beta_i \circ \alpha_i^* \circ \beta_i^* \tag{4.14}$$

where $\alpha_i = \mathbf{W}\mathbf{a}_R(\theta_i), \beta_i = \mathbf{F}\mathbf{a}_T(\phi_i)$, and $\circ$ indicates outer product. Evidently (4.14) represents a Canonical Polyadic (CP) decomposition of the tensor $\mathcal{R}$ where α_i and β_i are the CP factors. In the next subsection, we first propose an algorithm that can recover these factors $\alpha_i, \beta_i, i = 1, 2, \cdots L$ of the tensor $\mathcal{R}$ by performing Canonical Polyadic Decomposition (CPD). In Sec 4.3.2, we will show how θ and Φ can be recovered from α_i and β_i by utilizing the difference co-array geometry. By using a combination of these algebraic methods, we can potentially recover $L = O(N_T N_R)$ channel paths, without the need to consider multi-carrier signals (as suggested by [ZH17]), or a

[2] We let $\mathcal{R}_{t_1,t_1,r_2,t_2} = [\mathbf{R}_\mathbf{y} - \sigma^2 \mathbf{I}_{N_T \times N_T} \otimes (\mathbf{W}\mathbf{W}^H)]_{\tilde{i},\tilde{j}}$, where $\tilde{i} = (t_1 - 1)N_R + r_1, \tilde{j} = (t_2 - 1)N_T + r_2$, $1 \le r_1, r_2 \le N_R, 1 \le t_1, t_2 \le N_T$.

time-consuming search over a 2D grid [GWH17].

Recovery of CP Factors by Rank-1 Test

Let $\mathbf{U}\Gamma\mathbf{U}^H = \mathbf{R}_\mathbf{y}$ be the singular value decomposition of the covariance matrix $\mathbf{R}_\mathbf{y}$. If the CP decomposition (4.14) is unique (we will address the issue of uniqueness later in Theorem 27), there exists a unitary matrix $\mathbf{Q} \in \mathbb{C}^{L\times L}$ such that $\tilde{\mathbf{U}}\mathbf{Q} = \mathbf{B}$ where $\mathbf{Q}\mathbf{Q}^H = \mathbf{I}$, and $\tilde{\mathbf{U}} = \mathbf{U}\Gamma^{1/2}$. One of the key properties of the columns of $\mathbf{B}$ is that each column is a vectorized form of a *rank one* matrix, namely the ith column can be written as the vectorized form of the rank one matrix $\alpha_i\beta_i^T = (\mathbf{W}\mathbf{a}_R(\theta_i))(\mathbf{F}\mathbf{a}_T(\phi_i))^T$, for $i = 1,\cdots,L$. Hence, one way to find the CP decomposition of (4.14) is to find a matrix $\mathbf{Q}$ such that this property is satisfied. This can be done by using an algebraic rank-one test [DL06, DL14, KSP18, KP19b]:

$$C_2(\text{unvec}(\tilde{\mathbf{U}}\mathbf{q}_i) = 0$$

where $\text{unvec}(\tilde{\mathbf{U}}\mathbf{q}_i) \in \mathbb{C}^{N_R\times N_T}$ rearranges the elements of $\tilde{\mathbf{U}}\mathbf{q}_i$ into a matrix, and the notation $C_2(\mathbf{X}) \in \mathbb{C}^{\binom{N_R}{2}\times\binom{N_T}{2}}$ represents a matrix comprised of all 2×2 minors of a matrix $\mathbf{X} \in \mathbb{C}^{N_R\times N_T}$. Using the bilinearity of the map $C_2(.)$, and performing algebraic manipulations similar to [DL06, DLCC07, KSP18], it is possible to show that $\mathbf{Q}$ can be found by the procedure explained in Algorithm 3 (Steps 3-5). Finally α_i and β_i can be recognized as the top left and top right singular vectors respectively of $\text{unvec}(\tilde{\mathbf{U}}\mathbf{q}_i)$.

The following theorem states a sufficient condition under which the tensor decomposition presented in Algorithm 3 is able to uniquely identify the tensor factors α_i, β_i.

Theorem 27. *If for a choice of* $\mathbf{W},\mathbf{F},\theta,\Phi$*, the matrices* $(\mathbf{F}\mathbf{A}_T(\Phi))\odot(\mathbf{W}\mathbf{A}_R(\theta))$ *and* $C_2(\mathbf{F}\mathbf{A}_T(\Phi))\odot C_2(\mathbf{W}\mathbf{A}_R(\theta))$ *are both full-column rank, then the linear algebraic algorithm in 3 can uniquely find the factors* $\alpha_i = \mathbf{W}\mathbf{A}_R(\theta_i)$*, and* $\beta_i = \mathbf{F}\mathbf{a}_T(\phi_i)$*,* $i = 1,2,\cdots,L$*.*

Proof. The proof can be derived using the uniqueness condition provided in Section 2.1 of

Algorithm 3 Tensor Based Channel Estimation

1: Estimate the tensor $\mathcal{R}$, using (4.13).
2: Compute eigenvalue decomposition $\mathbf{R} = \mathbf{U}\Lambda\mathbf{U}^H$, and let $\tilde{\mathbf{U}} = \mathbf{U}\Lambda^{1/2}$. Also, let $\mathbf{U}^{(i)} = \text{unvec}(\tilde{\mathbf{u}}_i)$, with $\tilde{\mathbf{u}}_i$ being the ith column of $\tilde{\mathbf{U}}$.
3: Let the columns of matrix $\mathbf{G} \in \mathbb{C}^{\binom{N_R}{2}\binom{N_T}{2} \times L^2}$ be indexed by $(i-1)L+j$ with $1 \le i, j \le L$, and assign

$$[\mathbf{G}]_{((m,n,p,q)),(i-1)L+j} = \tilde{\mathbf{U}}^{(i)}_{m,n}\tilde{\mathbf{U}}^{(j)}_{p,q} - \tilde{\mathbf{U}}^{(i)}_{p,n}\tilde{\mathbf{U}}^{(j)}_{m,q}$$

where the notation $((m,n,p,q))$ represents a unique integer in the range $1 \le ((m,n,p,q)) \le \binom{N_R}{2}\binom{N_T}{2}$ which is assigned to every tuple m,n,p,q, where $1 \le m < n \le N_R$, and $1 \le p < q \le N_T$.
4: Compute L right singular vectors $\mathbf{v}_1, \cdots, \mathbf{v}_L$ of $\mathbf{G}$ corresponding to L smallest singular values. Form the matrices $\mathbf{V}_i = \text{unvec}(\mathbf{v}_i) \in \mathbf{C}^{L \times L}$, $i = 1, \cdots, L$.
5: Form the tensor $\mathcal{V} \in \mathbb{C}^{L \times L \times L}$ whose frontal slices are $\mathbf{V}_1, \cdots, \mathbf{V}_L$. Find the CP decomposition of $\mathcal{V}$ through Generalized Eigenvalue Decomposition [DL14, DL06], and store the CP factors in matrices $\mathbf{D}, \mathbf{Q}, \mathbf{Q}^T \in \mathbb{C}^{L \times L}$.
6: Find top left and right singular vector of $\text{unvec}(\tilde{\mathbf{U}}\mathbf{q}_i)$, and denote them as $\hat{\alpha}_i$, $\hat{\beta}_i$, which serve as estimates of α_i and β_i respectively.
7: Compute θ_i, ϕ_i using the procedure explained in Section 4.3.2.
8: Find the source powers by solving the least squares problem (4.18).

[DL06]. □

Remark 30. *A necessary condition for $C_2(\mathbf{FA}_T) \odot C_2(\mathbf{WA}_R)$ to be full column rank is $\binom{L}{2} \leq \binom{N_T}{2}\binom{N_R}{2}$ and $L \leq N_T N_R$. Hence, Algorithm 3 is potentially able to recover a maximum of $L = O(N_T N_R)$ channel paths.*

Recovering Channel Parameters from Tensor Factors Utilizing Difference Co-Array

In this section, we present a novel algorithm inspired from ESPRIT that can recover θ and Φ from $\alpha_i, \beta_i, i = 1, 2, \cdots L$ by exploiting the rich geometry of the difference co-array of sparse transmit and receive arrays. In particular, we cast the angle estimation problem in terms of the *difference co-array* of these arrays, and utilize the shift-invariance property of the Uniform Linear Array (ULA) segment of the difference co-array. Utilizing the difference co-arrays of cleverly designed non-uniform arrays are known to offer significant benefits over the traditional uniform linear array based direction-of-arrival estimation, including the ability to resolve $O(M^2)$ sources using M antennas [PV10, PV11].

Definition 14. *Given an antenna array with elements located at $d_1, \cdots, d_M$ (normalized w.r.t. $\bar{\lambda}/2$), the difference co-array is defined as $\mathcal{D} = \{d_i - d_j | 1 \leq i, j \leq M\}$.*

Let the sparse receive array be such that its difference co-array $\mathcal{D}^{(R)}$ is a continuous set of integers in the range $-M_{\text{ca}}^{(R)}, \cdots, M_{\text{ca}}^{(R)}$. For ULA, $M_{\text{ca}}^{(R)} = M_R - 1$ and for nested array, $M_{\text{ca}}^{(R)} = M_R^2/4 + M_R/2 - 1$. In the following, we explain our approach for estimating θ_i from α_i. A similar technique can be used to recover ϕ_i from β_i. Assuming that Algorithm 1 identified α_i and β_i correctly (upto an inherent scaling ambiguity of $\tilde{\alpha}_i$ and $\tilde{\beta}_i$ respectively), we have $\hat{\alpha}_i = \tilde{\alpha}_i \mathbf{W}\mathbf{a}_R(\theta_i)$. Therefore,

$$\alpha_i^* \otimes \alpha_i = |\tilde{\alpha}_i|^2 (\mathbf{W}^* \otimes \mathbf{W})(\mathbf{a}_R^*(\theta_i) \otimes \mathbf{a}_R(\theta_i))$$

We can identify $\mathbf{a}_R^*(\theta_i) \otimes \mathbf{a}_R(\theta_i)$ as the steering vector of a *virtual* array whose elements are located at the difference co-array $\mathcal{D}^{(R)}$. Since we assumed that the set $\mathcal{D}^{(R)}$ is comprised of consecutive integers, there exists a selection matrix $\Pi^{(R)} \in \mathbb{R}^{M_R^2 \times (2M_{\text{ca}}^{(R)}+1)}$ consisting of ones and zeros such that $\hat{\alpha}_i^* \otimes \hat{\alpha}_i = |\tilde{\alpha}_i|^2 (\mathbf{W}^* \otimes \mathbf{W}) \Pi^{(R)} \tilde{\mathbf{a}}_R(\theta_i)$ where $\tilde{\mathbf{a}}_R(\theta_i)$ corresponds to the largest consecutive part of $\mathbf{a}_R^*(\theta_i) \otimes \mathbf{a}_R(\theta_i)$, i.e., the steering vector corresponding to an antenna array whose elements are located on $-M_{\text{ca}}, \cdots, M_{\text{ca}}$. Let $\mathbf{V}$ be basis for the nullspace $N((\mathbf{W}^* \otimes \mathbf{W})\Pi^{(R)})$, and let $\mathbf{v}_0 = ((\mathbf{W}^* \otimes \mathbf{W})\Pi^{(R)})^\dagger (\hat{\alpha}^* \otimes \hat{\alpha})$. Then it holds that

$$|\tilde{\alpha}_i|^2 \tilde{\mathbf{a}}_R(\theta_i) = \mathbf{V}\mathbf{c} + \mathbf{v}_0 \tag{4.15}$$

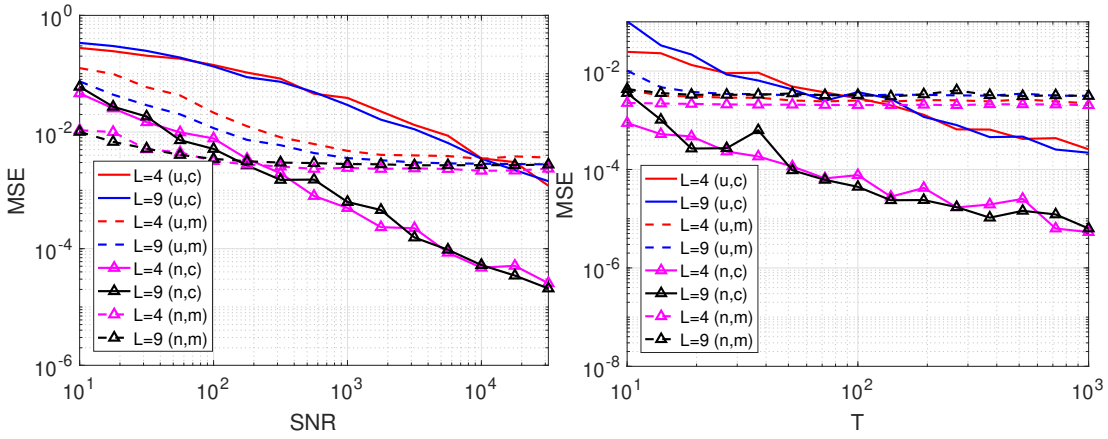
for some unknown $\mathbf{c}$. Since $\tilde{\mathbf{a}}_R(\theta_i)$ is a Vandermonde vector, it has the following shift invariance property:

$$\overline{\tilde{\mathbf{a}}_R(\theta_i)} = e^{j\pi \sin(\theta_i)} (\underline{\tilde{\mathbf{a}}_R(\theta_i)}) \tag{4.16}$$

where the notations $\overline{(.)}, \underline{(.)}$ indicate removal of first and last rows, respectively. Define $\mathbf{V}_1 = [\mathbf{v}_0, \mathbf{V}]$ and $\mathbf{c}_1 = [1, \mathbf{c}^T]$. Combining (4.15), (4.16), we get

$$\overline{\mathbf{V}_1}\mathbf{c}_1 = e^{j\pi \sin(\theta_i)} \underline{\mathbf{V}_1}\mathbf{c}_1 \tag{4.17}$$

which means that $\mathbf{c}_1$ is a generalized eigenvector of the matrices $\overline{\mathbf{V}_1}, \underline{\mathbf{V}_1}$ and $e^{j\pi \sin(\theta_i)}$ is the corresponding generalized eigenvalue. We can therefore perform generalized eigenvalue decomposition of $\overline{\mathbf{V}_1}, \underline{\mathbf{V}_1}$ and estimate θ_i as the phase of the generalized eigenvalue closest to unit circle. Similar procedure can be performed on tensor factors $\beta_i = \mathbf{F}\mathbf{a}_T(\phi_i)$ to obtain the AoDs ϕ_i. Once the AoAs and AoDs are estimated we can find the powers corresponding to the channel paths by



(a) MSE vs SNR (b) MSE vs T

Figure 4.6: The mean square error (MSE) w.r.t SNR and number of fading blocks, for different number of channel paths and array geometries. The notation (x,y) in the legends of the plots indicate the array type ($x \in \{u,n\}$, 'u' stands for ULA and 'n' indicates nested array) and algorithm ($y \in \{c,m\}$, 'c' denotes CPD (our algorithm), 'm' stands for beamspace MUSIC [GWH17].

solving the least squares problem:

$$\min_{\mathbf{p}} \|\text{vec}(\mathcal{R}) - (\mathbf{B}^* \odot \mathbf{B})\mathbf{p}\| \tag{4.18}$$

which has the closed form $\mathbf{p} = (\mathbf{B}^* \odot \mathbf{B})^{\dagger}\text{vec}(\mathcal{R})$, with $(.)^{\dagger}$ indicating pseudo-inverse of a matrix.

4.3.3 Numerical Experiments

In this section, we present our numerical experiments to examine the performance of the proposed algorithm. We consider a mmWave MIMO channel with $L \in \{4,9\}$ Angle-of-Departures Φ and Angle-of-Arrivals θ located uniformly in the ranges $[\frac{-5\pi}{12}, \frac{4\pi}{12}]$, and $[\frac{-4\pi}{12}, \frac{5\pi}{12}]$ respectively. We assume all powers are equal to 1. The SNR is defined to be $\text{SNR} = \frac{E}{N_T\sigma^2}$, and we choose $E = 0.5$. The transmitter and receiver are both equipped with $N_T = N_R = 5$ RF-chains and $M_T = M_R = 10$ antennas. We consider both ULA and nested array geometries in the simulations.

In our experiments, we study the effect of SNR and number of channel fading blocks

T on the accuracy of the recovered AoAs and AoDs, using our algorithm and the beamspace MUSIC algorithm presented in [GWH17]. The total error shown in plot 4.6 is defined as $\|[\hat{\theta}^T, \hat{\Phi}^T] - [\theta^T, \Phi^T]\|^2/(2K)$ and is averaged over 1000 runs. The phase-shifters matrices $\mathbf{W}, \mathbf{F}$ have unimodular elements whose phases are drawn from uniform distribution in $[0, 2\pi]$ and kept fixed for every 20 runs. The number of grid-points used for the beamspace MUSIC is chosen to be 1000. We used the library TensorLab3.0 [VDS$^+$16] to implement the proposed CPD, and use Simultaneous Generalized Schur Decomposition (SGSD) for simultaneous matrix diagonalization[3]. In order to estimate the AoAs and AoDs from the tensor factors, we use the procedure explained in Sec 4.3.2 for moderate and high values of SNR and T. Owing to noise sensitivity of the generalized eigenvalue decomposition at low SNR (or small T), we determine the angles by solving the following optimization problem

$$\min_{\theta_i} \| \frac{\mathbf{W}\mathbf{a}_R(\theta_i)}{\|\mathbf{W}\mathbf{a}_R(\theta_i)\|} - \hat{\alpha}_i \| \tag{4.19}$$

We solve (4.19) by performing a simple 1-d search (similar to [ZFY$^+$17]) over 1000 grid points, which is still much faster than the 2-d search required by beamspace MUSIC [GWH17].

As we can observe in Figure 4.6, as the SNR increases the average error for our algorithm as well as beamspace MUSIC [GWH17] decreases. Although for smaller SNRs, our algorithm has a worse performance than beamspace MUSIC, which is caused by sensitivity of CPD and simultaneous matrix diagonalization to noise, this effect is greatly reduced as the SNR increases, and our algorithm achieves a higher performance than that of [GWH17]. The low SNR performance can be potentially improved using adaptive sensing based ideas for transmit beamforming from [KSP19, CRJ19] . We can also see that the beamspace MUSIC cannot achieve a higher accuracy beyond a certain point due to the finite number of grid-points. Increasing the size of grid can be extremely costly as the beamspace MUSIC algorithm relies on a 2D search

[3]This algorithm is capable of finding CPD for tensors whose ranks exceeds two of its dimensions, which is the case in our simulations.

We can also observe that using a nested array geometry can achieve a lower error compared to ULA with the same number of antennas. Studying the effect of array geometry on the estimation error of the channel parameters will be a topic of future research.

4.3.4 Conclusion

We proposed a tensor based approach for estimation of channel parameters for a mmWave MIMO communication system. Unlike many existing algorithms, our approach does not require assumption of a pre-defined grid for the angles-of-arrival and angles-of-departure which can cause grid-mismatch errors. Utilizing the rich co-array geometry of sparse arrays employed at the transmitter and receiver, we proposed an algebraic technique to perform tensor decomposition, and derived sufficient conditions under which we are able to recover all the channel paths. In particular, we showed that a maximum of $L = O(N_T N_R)$ channel paths can be identifiable where N_R and N_T are the number of RF chains on the receiver and the transmitter, respectively. Compared to algorithms such as Beamspace MUSIC our approach has a lower computational complexity as it avoids two-dimensional grid search.

4.4 Tensor Decomposition for Multi-Carrier mmWave Channel Estimation with Correlation Priors

Due to high dimensionality of the data acquired for estimation of mmWave channels, as we also observed in Section 4.3, it is natural to convert the problem into a tensor decomposition problem, by carefully rearranging the received signals from the antennas into a higher order tensor. In the multiple-carrier case, many results already exist which use the Canonical-Polyadic (CP) decomposition of higher order tensors in order to estimate the channel parameters [QFS19, ZH17, ZGZ$^+$17, ZFY$^+$17, PAGH19]. It is well-known that CP decomposition can be unique even if the tensor rank is much larger than the smallest dimension [DL06, DLCC07, DDL13, BCMT10, RSH12]. Hence, these algorithms are able to estimates channels which have a larger number of channel paths [ZH17]. A key benefit of these results is that unlike the approaches in [ALH15, AMGPH14, EARAS$^+$14], they do not need the assumption of a pre-defined grid for the AoAs and AoDs.

However, most of these results do not fully exploit statistical properties of the channel. Specifically, when the gains associated with channel paths are uncorrelated, which can often be the case [PH16, GZC18, PH18, AGPHJ18], few results currently exist [KP20, PH18, HC16], for either single-carrier or multi-carrier channel estimation. The authors in [HC16], for example, utilize the correlation structure of a single-carrier mmWave channel, to come up with judicious ways of designing compression matrices in a hybrid beamforming communication system equipped with M uniformly located antennas, which can reduce the number of RF-chains to $O(\sqrt{M})$, while preserving the statistical information about the channel encoded in the received signals. Their design of the compression matrix is based on certain non-uniform array geometries such as coprime arrays [PV11] (and similar array geometries such as [PV10, LV16, QZA15, Mof68]), which are known to enjoy much higher degrees of freedom than their classical counterparts such as uniform linear arrays.

Though the results in [HC16] shed light on the benefits provided by utilizing correlation priors on the channel gains in a single carrier communication system, existing algorithms for channel estimation of multi-carrier systems, such as [ZH17, PAGH19] do not fully exploit the role of correlation priors. The authors in [PAGH19] consider a SIMO multi-carrier system, and show how the received samples on the RF-chains of the BS can be rearranged into a third order tensor whose CP decomposition determines the channel parameters. Although they derive upper bounds on the number of channel paths to guarantee that the CP decomposition is unique, their algorithm which is based on alternating minimization is not always guaranteed to identify the CP factors [PAGH19]. The authors in [ZH17], on the other hand, consider a MIMO multi-carrier mmWave system, and show once particular design (DFT beamforming) for the analog beamformer are assumed, one can provably recover $L = \min(NN_T, K)$ channel paths (where N_T is the number of RF chains of the MS), when K frequency bins are dedicated for pilot signals.

In this Section, we develop new tensor decomposition algorithms with theoretical guarantees that can identify significantly larger number of channel paths compared to existing techniques reviewed earlier. In particular, we show that in an OFDM system, if the frequency bins are carefully located in a non-uniform fashion, we can provably recover up to $L = \min(O(N^2), O(M), O(K^2))$ channel paths by considering certain tensor flattenings that capture the geometry of the difference-set of the non-uniformly placed pilots. At the expense of a higher computational complexity, we also show that the identiability results can be further improved to $L = O(NK^2)$ by lifting the tensor to a higher dimension. Note that these guarantees do not require a non-uniform array and continue to hold even when we deploy a ULA.

The contributions of this section are threefold: 1) Similar to [GZC18, AGPHJ18, PH16], we use correlation priors on the channel parameters, but unlike previous methods, we show that these priors combined with judiciously chosen frequency bins for pilots, can guarantee the recovery of up to $L = O(NK^2)$ paths, by developing linear algebraic based CCP decomposition algorithms. Moreover, this non-uniform selection of frequency bins for pilots enables us to

dedicate the other unused frequency bins for communication purposes, thereby reduce the communication overhead for channel training (See Sec. 4.4.2, and 4.4.3 for details).

2) Even though our proposed algorithms can provably find the CCP decomposition, due to existence of the beamforming matrix $\mathbf{W}$, recovering certain channel parameters, namely Angle-of-Arrivals (AoA) can be still nontrivial and the solution can be ambiguous. To the best of our knowledge, for the first time we propose easy to check conditions (in terms of $\mathbf{W}$ and the true AoAs) under which such ambiguities will not arise for a given $\mathbf{W}$ (See Sec. 4.4.4 for details).

3) We further provide a first order perturbation analysis which sheds light on the sensitivity of our proposed algorithms to small perturbations caused by an inaccurate estimate of the covariance matrix of the received signals (See Sec. 4.4.2 for details).

Notations: The symbols $(.)^T, (.)^H, (.)^*, (.)^\dagger$ stand for matrix transpose, matrix Hermitian, complex conjugate, pseudo-inverse, respectively, and $j = \sqrt{-1}$. The symbols $\otimes$, $\odot$ represent Kronecker, and (column-wise) Khatri-Rao product, respectively. Matrices (resp. vectors) are shown with boldface uppercase (resp. lowercase) characters such as $\mathbf{A}$ (resp. $\mathbf{b}$). Tensors are shown using italic uppercase characters such as $\mathcal{T}$. The remaining notations are clear from the context.

4.4.1 Problem Model

Consider a SIMO communication system where each MS has a single antenna and the BS is equipped with M antennas and $N \ll M$ RF-chains. We assume that the communication system uses a multi-carrier CP-OFDM modulation with FFT length of N_{fft}, Cyclic Prefix (CP) length of N_{cp}, and sampling period $T_s = 1/(N_{\text{fft}}\Delta_f)$, where Δ_f denotes the subcarrier spacing. For each OFDM symbol, only K frequency bins are used for channel training $(K \leq N_{\text{fft}})$, and the remaining bins could be used for communication. We assume that the pilots are located on i_1th, i_2th, $\cdots$, i_Kth frequency bins, and unlike conventional methods such as [ZPHN16], we do not assume $i_1, \cdots, i_K$ to be equally spaced.

The frequency domain representation of a frequency selective mmWave channel is given

as [ABB+07, HGPR+16]

$$\mathbf{h}(f) = \sum_{\ell=1}^{L} g_\ell \mathbf{a}(\theta_\ell) e^{-j2\pi\tau_\ell f},$$

where τ_ℓ represents channel path delay, θ_ℓ is angle-of-arrival, g_ℓ is the channel gain corresponding the ℓth path, and $\mathbf{a}(\theta)$ is the array steering vector corresponding to the antenna array. For the rest of this Section, we assume that the antennas are located uniformly, with inter-element spacing of $\lambda/2$, with λ being the carrier wave-length[4]. Hence, we have $\mathbf{a}(\theta) = [1, e^{j\pi\sin(\theta)}, \ldots, e^{j(M-1)\pi\sin(\theta)}]^T$.

We assume that θ_ℓ, $\ell = 1, \cdots, L$ remain constant during the channel estimation process, while g_ℓ can vary from symbol to symbol, but remains constant over each symbol. Moreover, we assume that $N_{\text{cp}} > (\max\{\tau_\ell\}_{\ell=1}^{L} - \min\{\tau_\ell\}_{\ell=1}^{L})/T_s$, and $N_{\text{cp}} < N_{\text{fft}}$. Equipped with these assumptions, the sampled channel transfer function at k-th pilot subcarrier and t-th symbol is given by [ZH17, GHDW16]

$$\mathbf{h}_k[t] = \sum_{\ell=1}^{L} g_\ell[t] \mathbf{a}(\theta_\ell) e^{-j2\pi\Delta_f \tau_\ell i_k}.$$

Assuming that the MS transmits the pilot signals $s_k[t] \in \mathbb{C}$ at the kth frequency bin, and tth symbol, the signal received at the output of the RF-chains at the BS can be written as

$$\mathbf{y}_k[t] = \mathbf{W}\mathbf{h}_k[t] s_k[t] + \mathbf{W}\mathbf{n}_k[t]$$

where $\mathbf{n}_k[t]$ is the additive noise on the antennas, and $\mathbf{W} \in \mathbb{C}^{N \times M}$ is a hybrid beamforming matrix, with $N < M$. Assuming that pilots are unimodular [PH16, PH18] (i.e., $|s_k[t]| = 1$), we can compute

$$\bar{\mathbf{y}}_k[t] = s_k^*[t]\mathbf{y}_k[t] = \mathbf{W}\mathbf{h}_k[t] + \mathbf{W}\bar{\mathbf{n}}_k[t],$$

[4]Non-uniform arrays can potentially provide further benefits in terms of improving the number of identifiable channel paths L, which can be a topic of future research.

where $\bar{\mathbf{n}}_k[t] = s_k^*[t]\mathbf{n}_k[t]$. Define $\mathbf{z}[t] = [\bar{\mathbf{y}}_1^T[t], \cdots, \bar{\mathbf{y}}_K^T[t]]^T \in \mathbb{C}^{NK\times 1}$. Then it holds that

$$\mathbf{z}[t] = (\mathbf{W}\otimes\mathbf{I}_K)\left(\sum_{\ell=1}^{L} g_\ell[t]\mathbf{a}(\theta_\ell)\otimes\mathbf{f}(\tau_\ell)\right) + (\mathbf{W}\otimes\mathbf{I}_K)\tilde{\mathbf{n}}[t]$$

with $\mathbf{f}(\tau_\ell) = [e^{-2j\pi\Delta_f\tau_\ell i_1}, \cdots, e^{-2j\pi\Delta_f\tau_\ell i_K}]^T$, and $\tilde{\mathbf{n}}[t] = [\bar{\mathbf{n}}_1^T[t], \cdots, \bar{\mathbf{n}}_K^T[t]]^T \in \mathbb{C}^{MK\times 1}$.

Now let us define $\mathbf{g}[t] = [g_1[t], \cdots, g_\ell[t]]^T \in \mathbb{C}^L$ be the vector containing gains corresponding to the channel paths at tth symbol. We assume that these gains are uncorrelated across different channel paths, i.e., $\mathbb{E}(\mathbf{g}[t]\mathbf{g}^H[t]) = \mathbf{P}$ where $\mathbf{P}$ is a diagonal matrix with $\mathbf{P} = \text{diag}(\mathbf{p})$, which is a practical assumption in many settings [PH16, PH18, AGPHJ18, GZC18]. Moreover, we make the assumption that $\mathbb{E}(\mathbf{g}[t]\mathbf{n}^H[t]) = \mathbf{0}$, and $\mathbb{E}(\mathbf{n}[t]\mathbf{n}^H[t]) = \sigma^2\mathbf{I}$. Hence, the covariance matrix $\mathbf{R} = \mathbb{E}(\mathbf{z}[t]\mathbf{z}^H[t])$ is given as $\mathbf{R} = \tilde{\mathbf{R}} + \sigma^2(\mathbf{W}\mathbf{W}^H\otimes\mathbf{I}_K)$ where

$$\tilde{\mathbf{R}} = (\mathbf{W}\otimes\mathbf{I}_K)(\mathbf{A}\odot\mathbf{F})\mathbf{P}(\mathbf{A}\odot\mathbf{F})^H(\mathbf{W}\otimes\mathbf{I}_K)^H \tag{4.20}$$

Here, $\mathbf{F} = [\mathbf{f}(\tau_1), \ldots, \mathbf{f}(\tau_L)] \in \mathbb{C}^{K\times L}$, and $\mathbf{A} = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_L)] \in \mathbb{C}^{M\times L}$. We further make the assumption that σ^2 is known. Hence, if the covariance matrix $\mathbf{R}$ is available, we can directly find $\tilde{\mathbf{R}}$ by subtracting $\sigma^2(\mathbf{W}\mathbf{W}^H\otimes\mathbf{I}_K)$ from $\mathbf{R}$. In practice, we only have access to a finite (say T) number of received symbols $\mathbf{z}[1], \mathbf{z}[2], \ldots, \mathbf{z}[T]$. Hence, we can only obtain an estimate of $\mathbf{R}$. One way to estimate $\mathbf{R}$ is through computing the sample covariance matrix

$$\hat{\mathbf{R}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{z}[t]\mathbf{z}[t]^H, \tag{4.21}$$

using which we can estimate $\tilde{\mathbf{R}}$ as $\hat{\tilde{\mathbf{R}}} = \hat{\mathbf{R}} - \sigma^2(\mathbf{W}\mathbf{W}^H\otimes\mathbf{I}_K)$. In this Section, we will show how the matrix $\tilde{\mathbf{R}}$ can be rearranged into a fourth order tensor, and by computing its Constrained Canonical Polyadic (CCP) decomposition, one is able to recover the channel parameters. Before presenting our algorithms, first let us review some of the key concepts in tensor decomposition, as well as difference co-arrays.

Preliminary Definitions on Tensors

A p-th order tensor is a multi-dimensional array $\mathcal{T} \in \mathbb{C}^{N_1 \times N_2 \times \cdots \times N_p}$ indexed by $i_1, \cdots, i_p$, such that $1 \leq i_1 \leq N_1, \cdots, 1 \leq i_p \leq N_p$. In general, tensors can be represented by lower dimensional factors, and many different approaches have been developed for tensor decomposition [SDLF$^+$17]. In this Section, we consider a specific form of tensor decomposition, namely Constrained Canonical Polyadic (CCP) decomposition [SDL16a], where the tensor factors are restricted to certain predefined sets $\mathbb{G}^{(q)}$, for $q = 1, 2, \cdots, p$:

Definition 15. *Let the sets $\mathbb{G}^{(1)}, \mathbb{G}^{(2)}, \ldots, \mathbb{G}^{(q)}$ be such that $\mathbb{G}^{(q)} \subseteq \mathbb{C}^{N_q}$, for $q = 1, 2, \cdots, p$.* ***Constrained CP-Rank*** *of $\mathcal{T} \in \mathbb{C}^{N_1 \times N_2 \times \cdots \times N_p}$, herein denoted as $crank_{\{\mathbb{G}^{(q)}\}_{q=1}^{p}}(\mathcal{T})$, is the smallest integer L for which there exist $\mathbf{a}_\ell^{(1)} \in \mathbb{G}^{(1)}, \cdots, \mathbf{a}_\ell^{(p)} \in \mathbb{G}^{(p)}$, $\ell = 1, 2, \cdots, L$ such that*

$$\mathcal{T} = \sum_{i=1}^{L} \mathbf{a}_\ell^{(1)} \otimes \mathbf{a}_\ell^{(2)} \otimes \cdots \otimes \mathbf{a}_\ell^{(p)}$$

For example, if some of the sets $\mathbb{G}^{(q)}$ restrict the tensor factors to be only of the form of Vandermonde vectors, then the constrained CP-decomposition can be related to a multi-dimensional harmonic retrieval problem [SDL16a]. In this Section, we assume that some[5] of the sets $\mathbb{G}^{(q)}$ have the following form

$$\mathbb{V}_{\mathbf{i}} = \{[e^{ji_1\omega}, e^{ji_2\omega}, \cdots, e^{ji_K\omega}]^T : \forall \omega \in [-\pi, \pi)\} \tag{4.22}$$

where $\mathbf{i} = [i_1, \cdots, i_K]$ and i_k are integers, $k = 1, \cdots, K$. The set $\mathbb{V}_{\mathbf{i}}$ can be thought of as a set containing steering vectors for all possible directions in $[-\pi, \pi)$ corresponding to a non-uniform antenna array. In the next section, we highlight properties of $\mathbb{V}_{\mathbf{i}}$ with regards to certain non-uniform choices of $i_1, \cdots, i_K$.

[5] As we will see later in Sec. 4.4.1, the other factors are constrained using the set $\mathbb{B}_{\mathbf{W}}$ defined in (4.25).

Khatri-Rao Product and Role of Difference Sets

In this section, we will review the concept of difference sets which will be later used in our proposed algorithms in Sec. 4.4.2 and 4.4.3. Let us begin by defining the difference set corresponding to a set of integers:

Definition 16. *For integer numbers $i_1, i_2, \cdots, i_K$, define the difference set as [PV10]*

$$\mathbb{D}_{\mathbf{i}} = \{i_{k_1} - i_{k_2}, 1 \leq k_1, k_2 \leq K\}$$

Moreover, define K' to be the largest integer such that the differences $i_{k_1} - i_{k_2}$ span the range of consecutive integers $-K', \cdots, K'$, i.e.,

$$K' = \max \ \ K_1 \ \ s.t. \ \ \{-K_1, \cdots, K_1\} \in \mathbb{D}_{\mathbf{i}}$$

Finally, denote $\mathbb{U}_{\mathbf{i}} = \{-K', \cdots, K'\}$.

Based on our definition of matrix the $\mathbf{F} \in \mathbb{C}^{K \times L}$, in Sec. 4.4.1, one can see that difference sets would arise naturally in the exponents of the matrix $\mathbf{F}^* \odot \mathbf{F}$ (interested reader can refer to [PV10] to get more familiar with the concept of co-arrays). In Sec. 4.4.2 and 4.4.3 we will show how difference co-arrays can be used in order to perform channel estimation when $L > \max(N, F)$.

Let $\tilde{K} = 2K' + 1 = |\mathbb{U}_{\mathbf{i}}|$. We define the matrix $\tilde{\mathbf{F}} \in \mathbb{C}^{\tilde{K} \times L}$ to be a Vandermonde submatrix of $\mathbf{F}^* \odot \mathbf{F}$ corresponding to the largest consecutive differences $-K', \cdots, K'$ in $\mathbb{D}_{\mathbf{i}}$. In particular, we let the columns of $\tilde{\mathbf{F}}$ to be

$$[\tilde{\mathbf{f}}_\ell]_{k+K'+1} = e^{-j2\pi\Delta_f k\tau_\ell}, \ \ -K' \leq k \leq K' \tag{4.23}$$

and $\tilde{\mathbf{F}} = [\tilde{\mathbf{f}}_1, \tilde{\mathbf{f}}_2, \cdots, \tilde{\mathbf{f}}_\ell] \in \mathbb{C}^{\tilde{K} \times L}$.

Since $\tilde{\mathbf{F}}$ is obtained by subselecting certain rows from $\mathbf{F}$, one can always find a row-

selection matrix $\Pi_{\mathbb{U}_\mathbf{i}} \in \{0,1\}^{\tilde{K}\times K^2}$ such that $\tilde{\mathbf{F}} = \Pi_{\mathbb{U}_\mathbf{i}}(\mathbf{F}^* \odot \mathbf{F})$.

For particular choices of $i_1, \cdots, i_K$, such as nested arrays [PV10], coprime arrays [PV11], and minimum redundancy arrays [Mof68], it has been shown that one can achieve $\tilde{K} = O(K^2)$. As an example, a two-level nested array with an even K has the following geometry [PV10]:

$$i_k = \begin{cases} k, & 1 \le k \le \frac{K}{2} \\ (\frac{K}{2}+1)k, & \frac{K}{2}+1 \le i \le K \end{cases}$$

It turns out that for such choices of $i_1, \cdots, i_K$, we have $\tilde{K} = \frac{K^2}{2} + K - 1$.

Formulating Channel Estimation as a 4-th order Constrained Tensor Decomposition Problem

The elements of the matrix $\tilde{\mathbf{R}}$, defined in (4.20) can be rearranged into a fourth order tensor of the form[6]

$$\mathcal{R} = \sum_{\ell=1}^{L} p_\ell \mathbf{f}(\tau_\ell) \otimes (\mathbf{W}\mathbf{a}(\theta_\ell)) \otimes \mathbf{f}^*(\tau_\ell) \otimes (\mathbf{W}\mathbf{a}(\theta_\ell))^*. \tag{4.24}$$

Therefore, $\mathcal{R}$ has a constrained CP decomposition of $\text{crank}_{\{\mathbb{G}^{(q)}\}_{q=1}^4}(\mathcal{R}) \le L$, with $\mathbb{G}^{(1)} = \mathbb{V}_\mathbf{i}, \mathbb{G}^{(3)} = \mathbb{V}_\mathbf{i}^*$, and $\mathbb{G}^{(2)} = \mathbb{B}_\mathbf{W}$, $\mathbb{G}^{(4)} = \mathbb{B}_\mathbf{W}^*$, where $\mathbb{V}_\mathbf{i}$ is defined in (4.22), and

$$\mathbb{B}_\mathbf{W} = \{\mathbf{W}\mathbf{a}(\theta) : \forall \theta \in [\frac{-\pi}{2}, \frac{\pi}{2})\}. \tag{4.25}$$

and the notation $(.)^*$ takes conjugate from all elements in a set. In Sec. 4.4.2, and 4.4.3, we propose two different algorithms for decomposition of tensor $\mathcal{R}$, which will lead to sufficient conditions under which $\text{crank}_{\{\mathbb{G}^{(q)}\}_{q=1}^4}(\mathcal{R}) = L$, and the CCP decomposition is unique. In each of these algorithms we consider a different form of flattening of the tensor $\mathcal{R}$ into matrices, leading

[6]This rearrangement can be done by $\mathcal{R}_{j_1,j_3,j_2,j_4} = \mathbf{R}_{(j_3-1)K+j_1,(j_4-1)K+j_2}$, where $1 \le j_1, j_2 \le K, 1 \le j_3, j_4 \le N$.

to different guarantees in terms of number of channel paths L that can be recovered.

4.4.2 Algorithm 1: Correlation Aware Channel Estimation using *ESPRIT* (COACH-ESPRIT-I)

Let us consider the following flattening[7] of the tensor $\mathcal{R}$:

$$\mathbf{R}^{(1)} = (\mathbf{F}^* \odot \mathbf{F})\mathbf{P}\left((\mathbf{WA})^* \odot (\mathbf{WA})\right)^T \tag{4.26}$$

Based on the discussion in Sec. 4.4.1, the matrix $\mathbf{F}^* \odot \mathbf{F}$ contains a ULA section $\tilde{\mathbf{F}}$ which is obtained by $\tilde{\mathbf{F}} = \Pi_{\mathbb{U}_\mathbf{i}}\mathbf{F}$, where $\Pi_{\mathbb{U}_\mathbf{i}}$ is a selection matrix that only depends on the structure of the difference set $\{i_{k_1} - i_{k_2} : 1 \leq k_1, k_2 \leq K\}$. Hence, from (4.26) we get

$$\Pi_{\mathbb{U}_\mathbf{i}}\mathbf{R}^{(1)} = \tilde{\mathbf{F}}\mathbf{B}^T \tag{4.27}$$

where $\mathbf{B} = \left((\mathbf{WA})^* \odot (\mathbf{WA})\right)\mathbf{P}$. Equation (4.27) is closely connected to the one-dimensional harmonic retrieval problem, which has been widely studied in the literature [SDL16a, SDL16b, JStB01]. One of the well-known approaches to solve (4.27) is the ESPRIT algorithm [RK89]. In the following, we show how we can find the channel parameters $\theta_\ell, \tau_\ell, p_\ell$ by using the idea of ESPRIT algorithm.

Let $\mathbf{U}\Sigma\mathbf{V}^H$ be the truncated singular value decomposition[8] (t-SVD) of the matrix $\mathbf{R}^{(1)}$. We assume that both the matrices $\mathbf{F}^* \odot \mathbf{F}$ and $(\mathbf{WA})^* \odot (\mathbf{WA})$ have full column rank (we will later in Sec. 4.4.2 show under what conditions these assumptions hold). Therefore, there exists a

[7] This flattening can be obtain by letting $\mathbf{R}^{(1)}_{j_1+(j_2-1)K, j_3+(j_4-1)K} = \mathcal{R}_{j_1,j_3,j_2,j_4}$, where $1 \leq j_1, j_2 \leq K, 1 \leq j_3, j_4 \leq N$.

[8] Here, $\mathbf{U} \in \mathbb{C}^{K^2 \times L}$, $\mathbf{V} \in \mathbb{C}^{N^2 \times L}$ are left and right singular vectors corresponding to nonzero singular values of $\mathbf{R}^{(1)}$, and $\Sigma \in \mathbb{R}^{L \times L}$ is the diagonal matrix containing nonzero singular values.

nonsingular matrix $\mathbf{E}$ such that

$$\mathbf{UE} = \mathbf{F}^* \odot \mathbf{F}$$

In order to find $\mathbf{E}$, we use the property that the ULA section corresponding to the co-array $\mathbf{F}^* \odot \mathbf{F}$ satisfies the shift invariance property. In other words, it holds that

$$\overline{\mathbf{\Pi}_{\mathbb{U}_\mathbf{i}}}(\mathbf{F}^* \odot \mathbf{F}) = \underline{\mathbf{\Pi}_{\mathbb{U}_\mathbf{i}}}(\mathbf{F}^* \odot \mathbf{F})\mathbf{\Lambda}$$

where $\mathbf{\Lambda} = \text{diag}(e^{-j2\pi\Delta_f\tau_1}, \cdots, e^{-j2\pi\Delta_f\tau_L})$, and the notations $\overline{(.)}, \underline{(.)}$ stand for removal of the first and last row of a matrix. Therefore, $\mathbf{E}$ must satisfy the following equation:

$$\overline{\mathbf{\Pi}_{\mathbb{U}_\mathbf{i}}}\mathbf{UE} = \underline{\mathbf{\Pi}_{\mathbb{U}_\mathbf{i}}}\mathbf{UE}\mathbf{\Lambda}.$$

Assuming that $\underline{\mathbf{\Pi}_{\mathbb{U}_\mathbf{i}}}(\mathbf{F}^* \odot \mathbf{F})$ is full column rank (sufficient conditions established in Sec. 4.4.2), we can find $\mathbf{E}$ from the following eigenvalue decomposition (EVD):

$$(\underline{\mathbf{\Pi}_{\mathbb{U}_\mathbf{i}}}\mathbf{U})^\dagger \overline{\mathbf{\Pi}_{\mathbb{U}_\mathbf{i}}}\mathbf{U} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^{-1} \tag{4.28}$$

Using the eigenvalues $\mathbf{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_L)$, we can immediately find the channel path delays $\hat{\tau}_\ell = -\text{Im}(\log(\lambda_\ell)/(2\pi\Delta_f))$, using which we can form the matrix $\hat{\mathbf{F}}$, with elements $[\hat{\mathbf{F}}]_{k,\ell} = e^{-j2\pi\Delta_f\hat{\tau}_\ell i_k}$. The matrix $\mathbf{B}$ can be estimated by solving the following least squares problem:

$$\hat{\mathbf{B}} = \arg\min_{\mathbf{B}} \|(\hat{\mathbf{F}}^* \odot \hat{\mathbf{F}})\mathbf{B}^T - \mathbf{R}^{(2)}\|_F$$

which has the closed form $\hat{\mathbf{B}}^T = (\hat{\mathbf{F}}^* \odot \hat{\mathbf{F}})^\dagger \mathbf{R}^{(2)}$.

So far, we have been able to estimate the matrix $\mathbf{B} = ((\mathbf{WA})^* \odot (\mathbf{WA}))\,\mathbf{P}$. In order to find

the parameters θ_ℓ, we propose to solve the following optimization problem:

$$\hat{\theta}_\ell = \arg\max_{\theta} \frac{|\hat{\mathbf{b}}_\ell^H((\mathbf{Wa}(\theta))^* \otimes (\mathbf{Wa}(\theta)))|}{\|(\mathbf{Wa}(\theta))^* \otimes (\mathbf{Wa}(\theta))\|} \tag{4.29}$$

Once $\hat{\theta}_\ell$ are recovered, we can easily find the parameters $\hat{p}_\ell$, by evaluating the function in (4.29) at $\hat{\theta}_\ell$, for $\ell = 1, \cdots, L$. The steps of this algorithm are summarized in Algo. 4.

Algorithm 4 COACH-ESPRIT-I

1: Compute t-SVD: $\mathbf{U}\Sigma\mathbf{V}^H = \mathbf{R}^{(1)}$.
2: Compute EVD: $(\underline{\Pi_{\mathbb{U}_\mathbf{i}}}\mathbf{U})^\dagger \overline{\Pi_{\mathbb{U}_\mathbf{i}}}\mathbf{U} = \mathbf{E}\Lambda\mathbf{E}^{-1}$
3: $\hat{\tau}_\ell = -\mathrm{Im}(\log(\lambda_\ell)/(2\pi))$
4: Let $[\hat{\mathbf{F}}]_{k,\ell} = e^{-j2\pi\Delta_f \hat{\tau}_\ell i_k}$, $1 \le k \le K, 1 \le \ell \le L$.
5: Compute $\hat{\mathbf{B}}^T = (\hat{\mathbf{F}}^* \odot \hat{\mathbf{F}})^\dagger \mathbf{R}^{(2)}$
6: Solve (4.29) to find $\hat{\theta}_\ell$ for $1 \le \ell \le L$.
7: $\hat{p}_\ell = \frac{|\hat{\mathbf{b}}_\ell^H((\mathbf{Wa}(\hat{\theta}_\ell))^* \odot (\mathbf{Wa}(\hat{\theta}_\ell)))|}{\|(\mathbf{Wa}(\hat{\theta}_\ell))^* \odot (\mathbf{Wa}(\hat{\theta}_\ell))\|}$

Identifiability Results

It can be verified from the derivation of algorithm COACH-ESPRIT-I that whenever the matrices $\underline{\Pi_{\mathbb{U}_\mathbf{i}}}(\mathbf{F}^* \odot \mathbf{F})$ and $(\mathbf{WA})^* \odot (\mathbf{WA})$ have full column rank it is possible to uniquely identify the factors $\mathbf{F}$ and $\mathbf{WA}$. The following theorem states this result:

Theorem 28. *If the matrices* $\underline{\Pi_{\mathbb{U}_\mathbf{i}}}(\mathbf{F}^* \odot \mathbf{F})$ *and* $(\mathbf{WA})^* \odot (\mathbf{WA})$ *are both full column rank, then the tensor* $\mathcal{R}$ *given in* (4.24) *has a unique Constrained CP decomposition with constraints* $\{\mathbb{V}_\mathbf{i}, \mathbb{C}^N, \mathbb{V}_\mathbf{i}^*, \mathbb{C}^N\}$*, and Steps 1-5 of the algorithm* COACH-ESPRIT-I *can recover it.*

Proof. Since $\underline{\Pi_{\mathbb{U}_\mathbf{i}}}(\mathbf{F}^* \odot \mathbf{F})$ has full column rank, a nonsingular matrix $\mathbf{E}$ satisfying (4.28) exists and hence Λ can be identified by the eigenvalues of $(\underline{\Pi_{\mathbb{U}_\mathbf{i}}}(\mathbf{F}^* \odot \mathbf{F}))^\dagger \overline{\Pi_{\mathbb{U}_\mathbf{i}}}(\mathbf{F}^* \odot \mathbf{F})$. Once Λ is recovered, we can identify $\mathbf{F}$. Since $\mathbf{F}^* \odot \mathbf{F}$ has full column rank, $\mathbf{B}$ can be uniquely found by solving $(\mathbf{F}^* \odot \mathbf{F})\mathbf{B}^T = \mathbf{R}^{(2)}$. This shows that $\mathbf{Wa}(\theta_\ell)$ for $\ell = 1, 2, \cdots, L$ can be uniquely identified. □

In the following Theorem, we provide conditions in terms of the dimensions L,N,M such that the matrix $(\mathbf{WA})^* \odot (\mathbf{WA})$ is full column rank for almost all choices of $\mathbf{W} \in \mathbb{C}^{N\times M}$:

Theorem 29. *The matrix* $(\mathbf{WA})^* \odot (\mathbf{WA})$ *is full column rank for almost all* $\mathbf{W} \in \mathbb{C}^{N\times M}$ *(except for possibly a set of measure zero), as long as* $L \leq 2N_2(N_1+1)+1$ *where* N_1, N_2 *are integers given by*

$$(N_1,N_2) = \left(\lfloor \frac{\min(N,N_0)}{2} \rfloor, \lceil \frac{\min(N,N_0)}{2} \rceil \right) \tag{4.30}$$

and $N_0 = \lfloor 2\sqrt{M+1} - 1 \rfloor$.

Proof. In [KP] we have provided a proof for a more general case: nonsingularity of Fisher Information Matrix for direction of arrival estimation in presence of a compression matrix, we showed that $\mathbf{G} = (\mathbf{WA})^* \odot (\mathbf{WA})$ is full column rank, for almost all $\mathbf{W} \in \mathbb{C}^{N\times M}$. Hence, the premise of theorem 29 is a simpler case of what we proved in [KP]. An interested reader might refer to [KP] for a more detailed proof. Here, we present a brief sketch of the proof.

Through a similar argument to [KP, Lemma 1], it can be shown that if there exists a matrix $\mathbf{W}_0 \in \mathbb{C}^{N\times M}$ such that $(\mathbf{W}_0\mathbf{A})^* \odot (\mathbf{W}_0\mathbf{A})$ is full column rank, then for almost all $\mathbf{W} \in \mathbb{C}^{N\times M}$, $(\mathbf{WA})^* \odot (\mathbf{WA})$ is full column rank. Therefore, it remains to construct a matrix $\mathbf{W}_0$ for which $(\mathbf{W}_0\mathbf{A})^* \odot (\mathbf{W}_0\mathbf{A})$ is full column rank. In [KP], we showed that this is possible by considering a nested array [PV10] structure, with N_1 elements in the inner ULA, and N_2 elements in the outer ULA, where N_1, N_2 are determined in the statement of the theorem. □

Combining the results of Theorems 28, and 29 we get the following Corollary, regarding maximum number of CCP factors identified by COACH-ESPRIT-I.

Corollary 7. *If the generators* $e^{-j2\pi\Delta_f\tau_1}, \cdots, e^{-j2\pi\Delta_f\tau_L}$ *are distinct, as long as*

$$L \leq \min(2N_2(N_1+1)+1, \tilde{K}-1),$$

with N_1*, and* N_2 *given in* (4.30)*, the tensor* $\mathcal{R}$ *given in* (4.24) *has a unique Constrained CP*

decomposition with constraints $\{\mathbb{V}_{\mathbf{i}}, \mathbb{C}^N, \mathbb{V}_{\mathbf{i}}^*, \mathbb{C}^N\}$*, for almost all* $\mathbf{W} \in \mathbb{C}^{N\times M}$ *and the algorithm* COACH-ESPRIT-I *can recover the factors* $\mathbf{f}(\tau_\ell)$*, and* $\mathbf{Wa}(\theta_\ell)$*, for* $\ell = 1, 2, \cdots, L$*.*

Remark 31. *If* $\mathbf{i}$ *is chosen such that* $\tilde{K} = O(K^2)$*, the result of Corollary 7 indicates that the algorithm* COACH-ESPRIT-I *can work as long as* $L = \min(O(K^2), O(N^2), O(M))$*. In Sec. 4.4.3, we show how we can achieve* $L = O(K^2 N)$*, by using a different rearrangement of the tensor.*

Remark 32. *In [PAGH19], the authors considered the estimation of channel parameters by decomposition of a tensor of dimensions* $K \times N \times T$*. While the tensor can potentially have a unique decomposition as long as* $K = O(K+N+T)$ *(due to Kruskal's theorem [Kru76]), their proposed approach, which is based on Alternating Least Squares [KB09] is not necessarily guaranteed to find the exact tensor decomposition even in the noiseless case. In [ZH17], the authors considered a MIMO communication system equipped with* $M_T > 1$ *antennas in the transmitter and* $N_T < M_T$ *RF-chains, and showed how channel parameters can be identified by decomposing a* $N_T \times N \times K$ *tensor. While their proposed tensor decomposition algorithm is guaranteed to work when specific cases for the beamforming matrix* $\mathbf{W}$ *are considered (such as DFT beamforming), they can only recover* $L = \min(K, NN_T)$ *channel paths.*

In contrast, the algorithms presented in this Section can potentially recovery up to $L = O(K^2 N)$ *channel paths. Extension of the results in this Section to the MIMO case can be a topic of future research.*

First Order Perturbations

In practice, the covariance matrix $\mathbf{R}$ can only be approximated using finite number of samples, using estimates such as sample covariance matrix $\hat{\hat{\mathbf{R}}}$ given in (4.21). Therefore, the algorithm COACH-ESPRIT-I has access only to a perturbed version $\hat{\mathbf{R}}^{(1)}$ of the matrix $\mathbf{R}^{(1)}$. In this section, we provide a first order perturbation analysis of the algorithm, and will show how

the estimated $\hat{\theta}_\ell, \hat{\tau}_\ell$ will behave with respect to small perturbations in $\mathbf{R}^{(1)}$. Let[9]

$$\hat{\mathbf{R}}^{(1)} = \mathbf{R}^{(1)} + \Delta_{\mathbf{R}^{(1)}}$$

Theorem 30. *Let* $[\mathbf{U}\ \mathbf{U}_n] \in \mathbb{C}^{K^2 \times K^2}$ *(resp.* $[\mathbf{V}\ \mathbf{V}_n] \in \mathbb{C}^{N^2 \times N^2}$*) be the left (resp. right) singular vectors of* $\mathbf{R}^{(1)}$ *with* $\mathbf{U} \in \mathbb{C}^{K^2 \times L}$ *(resp.* $\mathbf{V} \in \mathbb{C}^{N^2 \times L}$*) being the left (resp. right) singular vectors corresponding to the nonzero singular values of* $\mathbf{R}^{(1)}$*, and* $\Sigma \in \mathbb{R}^{L \times L}$ *be the diagonal matrix containing nonzero singular values. Then, up to the first order of approximation in* $O(\|\Delta_{\mathbf{R}^{(1)}}\|)$ *we have*

$$\delta_{\tau_\ell} = Im\left(\tilde{\mathbf{e}}_\ell^T(\underline{\Pi_{\mathbb{U}_\mathbf{i}}}\mathbf{U})^\dagger(\overline{\Pi_{\mathbb{U}_\mathbf{i}}}/\lambda_\ell - \underline{\Pi_{\mathbb{U}_\mathbf{i}}})\Delta_\mathbf{U}\mathbf{e}_\ell)/(2\pi\Delta_f)\right) \tag{4.31}$$

where $\tilde{\mathbf{e}}_\ell$ *is the* ℓ*th row of* $\mathbf{E}^{-1}$*,* $\mathbf{e}_\ell$ *is the* ℓ*th column of* $\mathbf{E}$*,* λ_ℓ *is the* ℓ*th diagonal element of* Λ*, with* $\mathbf{E}$*,* Λ *defined in Sec. 4.4.2, and*

$$\Delta_\mathbf{U} = \mathbf{U}_n\mathbf{U}_n^H\Delta_{\mathbf{R}^{(1)}}\mathbf{V}\Sigma^{-1} \tag{4.32}$$

Moreover, it holds that

$$\delta_{\theta_\ell} = -\frac{\frac{\delta_{\dot{u}}(\theta_\ell)}{u(\theta_\ell)} - \frac{\delta_u(\theta_\ell)\dot{u}(\theta_\ell)}{u^2(\theta_\ell)}}{\frac{u(\theta_\ell)\ddot{u}(\theta_\ell)-(\dot{u}(\theta_\ell))^2}{u^2(\theta_\ell)} - \frac{v(\theta_\ell)\ddot{v}(\theta_\ell)-(\dot{v}(\theta_\ell))^2}{v^2(\theta_\ell)}} \tag{4.33}$$

[9]Throughout this section, for all other matrices we also use a similar notation (i.e., for a matrix $\mathbf{X}$, $\hat{\mathbf{X}} = \mathbf{X} + \Delta_\mathbf{X}$ shows the perturbed version of matrix $\mathbf{X}$). Similarly, for a vector $\mathbf{x}$, we show the perturbed version as $\hat{\mathbf{x}} = \mathbf{x} + \delta_\mathbf{x}$, and finally the notation $\hat{x} = x + \delta_x$ shows perturbed version of a scalar x. Herein, all perturbations are computed up to the first order $O(\|\Delta_\mathbf{R}^{(1)}\|)$

where

$$\begin{aligned}
\delta_u(\theta_\ell) &= 2Re(\delta_{\mathbf{b}_\ell}^H \mathbf{b}_\ell)\|\mathbf{b}_\ell\|_2^2 \\
\delta_{\dot{u}}(\theta_\ell) &= 2Re((\delta_{\mathbf{b}_\ell}^H \mathbf{b}_\ell)^*(\mathbf{b}_\ell^H \dot{\mathbf{b}}_\ell) + (\delta_{\mathbf{b}_\ell}^H \dot{\mathbf{b}}_\ell)\|\mathbf{b}_\ell\|_2^2) \\
u(\theta_\ell) &= \|\mathbf{b}_\ell\|_2^4, \;\; \dot{u}(\theta_\ell) = 2Re(\mathbf{b}_\ell^H \dot{\mathbf{b}}_\ell)\|\mathbf{b}_\ell\|^2 \\
\ddot{u}(\theta_\ell) &= 2\|\mathbf{b}_\ell^H \dot{\mathbf{b}}_\ell\|^2 + 2Re(\mathbf{b}_\ell^H \ddot{\mathbf{b}}_\ell)\|\mathbf{b}_\ell\|_2^2 \\
v(\theta_\ell) &= \|\mathbf{b}_\ell\|_2^2, \;\; \dot{v}(\theta_\ell) = 2Re(\dot{\mathbf{b}}_\ell^H \mathbf{b}_\ell) \\
\ddot{v}(\theta_\ell) &= 2\|\ddot{\mathbf{b}}_\ell\|^2 + 2Re(\mathbf{b}_\ell^H \ddot{\mathbf{b}}_\ell)
\end{aligned}$$

Here, $\mathbf{b}_\ell = (\mathbf{W}^* \otimes \mathbf{W})\mathbf{a}_{kr}(\theta_\ell)$, $\mathbf{a}_{kr}(\theta_\ell) = \mathbf{a}^*(\theta_\ell) \otimes \mathbf{a}(\theta_\ell)$, $\dot{\mathbf{b}}_\ell = (\mathbf{W}^* \otimes \mathbf{W})\dot{\mathbf{a}}_{kr}(\theta_\ell)$, $\ddot{\mathbf{b}}_\ell = (\mathbf{W}^* \otimes \mathbf{W})\ddot{\mathbf{a}}_{kr}(\theta_\ell)$, $\dot{\mathbf{a}}_{kr}(\theta) = \frac{d\mathbf{a}_{kr}(\theta)}{d\theta}$, *and* $\ddot{\mathbf{a}}_{kr}(\theta) = \frac{d^2\mathbf{a}_{kr}(\theta)}{d\theta^2}$.

Proof. The proof is provided in Appendix 4.5.1. □

4.4.3 Algorithm 2: COACH-ESPRIT-II

The algorithm COACH-ESPRIT-I is capable of recovering $\min(O(K^2), O(N^2), O(M))$ CCP factors (or equivalent number of channel paths). However, if $N \ll K$, the number of RF-chains can become a bottle-neck on the number of channel paths L that we are able to recover. In this section, we present another algorithm which is based on a different way of rearranging the elements of the tensor $\mathcal{R}$ into a matrix, and is potentially able to recover $L = O(K^2N)$. This type of rearrangement uses the difference set corresponding to the pilots $\mathbb{D}_\mathbf{i} = \{i_{k_1} - i_{k_2} : 1 \leq k_1, k_2 \leq K\}$.

Lifting the tensor to a higher dimension

As we observed in Sec. 4.4.1, if the locations $i_1, \cdots, i_K$ are judiciously designed, the size of the difference set $\mathbb{D}_\mathbf{i}$ can be as large as $O(K^2)$. We will show how we can use the difference set $\mathbb{D}_\mathbf{i}$ in order to rearrange the tensor $\mathcal{R}$ into a higher dimensional tensor. Before doing that, let us

first define the following indexed version of difference set:

Definition 17. ***Indexed difference set:*** *For integer numbers $i_1, i_2, \cdots, i_K$, define the indexed difference set as*

$$\mathbb{D}_{\mathbf{i}}^{\mathbb{I}} = \{(k_1, k_2, i_{k_1} - i_{k_2}), 1 \leq k_1, k_2 \leq K\}$$

Using this definition, we are now able to define the following *lifting* operator, which can help us decompose upto $O(K^2N)$ channel paths:

Definition 18. *Given a tensor $\mathcal{R} \in \mathbb{C}^{K \times N \times K \times N}$, the 4th order tensor $\textsc{TensorToep}^{(\mathbb{D}_{\mathbf{i}}^{\mathbb{I}})}(\mathcal{R}) \in \mathbb{C}^{\tilde{K} \times N \times \tilde{K} \times N}$ is constructed as*

$$[\textsc{TensorToep}^{(\mathbb{D}_{\mathbf{i}}^{\mathbb{I}})}(\mathcal{R})]_{p,n_1,q,n_2} = \mathcal{R}_{k_1,n_1,k_2,n_2}$$

where $(k_1, k_2, p-q) \in \mathbb{D}_{\mathbf{i}}^{\mathbb{I}}$, $1 \leq n_1, n_2 \leq N$, $1 \leq m_1, m_2 \leq K$. In presence of noise, we may average over the redundancies in $\mathbb{D}_{\mathbf{i}}^{\mathbb{I}}$, i.e.,

$$[\textsc{TensorToep}^{(\mathbb{D}_{\mathbf{i}}^{\mathbb{I}})}(\mathcal{R})]_{p,n_1,q,n_2} = \frac{\sum_{(k_1,k_2) \in \mathbb{D}_{\mathbf{i}}^{(p,q)}} \mathcal{R}_{k_1,n_1,k_2,n_2}}{|\mathbb{D}_{\mathbf{i}}^{(p,q)}|}$$

where $\mathbb{D}_{\mathbf{i}}^{(p,q)} = \{(k_1, k_2) : (k_1, k_2, p-q) \in \mathbb{D}_{\mathbf{i}}^{\mathbb{I}}\}$.

As we showed in Sec. 4.4.1, depending on the choice of $i_1, \cdots, i_K$, we can have $\tilde{K} \gg K$, the tensor $\textsc{TensorToep}^{(\mathbb{D}_{\mathbf{i}}^{\mathbb{I}})}(\mathcal{R})$ can potentially have much larger dimensions than $\mathcal{R}$. This will enable us to decompose tensors of much higher ranks. The following Lemma shows the decomposition corresponding to this *lifted* tensor:

Lemma 25. *[KP19a] It holds that for $\mathcal{R}$ defined in* (4.24)

$$\textsc{TensorToep}^{(\mathbb{D}_{\mathbf{i}}^{\mathbb{I}})}(\mathcal{R}) = \sum_{\ell=1}^{L} p_\ell \tilde{\mathbf{f}}_\ell \otimes (\mathbf{W}\mathbf{a}_\ell) \otimes \tilde{\mathbf{f}}_\ell^* \otimes (\mathbf{W}\mathbf{a}_\ell)^*$$

where $\tilde{\mathbf{f}}_\ell \in \mathbb{C}^{\tilde{K}}$ *is the* ℓ*th column of* $\tilde{\mathbf{F}}$ *defined in* (4.23)*, and* $\mathbf{a}_\ell$ *is a shorthand for* $\mathbf{a}(\theta_\ell)$.

Proof. Proof follows from applying definition 18 to the entries of $\mathcal{R}$ defined in (4.24). □

In Sec. 4.2, we proposed an algorithm for decomposition of tensors of the form $\text{TENSORTOEP}^{(\mathbb{D}^{\mathbb{I}}_{\mathbf{i}})}(\mathcal{R})$. Here, we briefly review this algorithm. Let us first make the following definition:

Definition 19. *Given* $\tilde{\mathcal{R}} \in \mathbb{C}^{\tilde{K}\times N\times\tilde{K}\times N}$*, let* $\text{FLATTEN}(\tilde{\mathcal{R}}) \in \mathbb{C}^{\tilde{K}N\times\tilde{K}N}$ *be such that its* (i,j)*th element is given as*

$$[\text{FLATTEN}(\tilde{\mathcal{T}})]_{i,j} = [\tilde{\mathcal{T}}]_{k_1,n_1,k_2,n_2}$$

such that $i = k_1 + (n_1-1)\tilde{K}$*,* $j = k_2 + (n_2-1)\tilde{K}$*,* $1 \le k_1,k_2 \le \tilde{K}, 1 \le n_1,n_2 \le N$.

Now define $\mathbf{R}^{(2)} \in \mathbb{C}^{\tilde{M}N\times\tilde{M}N}$ to be such that $\mathbf{R}^{(2)} = \text{FLATTEN}(\text{TENSORTOEP}^{(\mathbb{D}^{\mathbb{I}}_{\mathbf{i}})}(\mathcal{R}))$. Following the definitions of the operations $\text{FLATTEN}(.)$, and $\text{TENSORTOEP}^{(\mathbb{D}^{\mathbb{I}}_{\mathbf{i}})}(.)$, and Lemma 25, we have that

$$\mathbf{R}^{(2)} = (\mathbf{WA}\odot\tilde{\mathbf{F}})\mathbf{P}(\mathbf{WA}\odot\tilde{\mathbf{F}})^H$$

Let $\mathbf{U\Sigma V}^H$ be the truncated singular value decomposition (t-SVD) of $\mathbf{R}^{(2)}$. Upon assuming that the matrix $(\mathbf{WA})\odot\tilde{\mathbf{F}}$ has full column rank[10], there exists a nonsingular matrix $\mathbf{E} \in \mathbb{C}^{L\times L}$ such that

$$\mathbf{UE} = (\mathbf{WA})\odot\tilde{\mathbf{F}} \tag{4.34}$$

One of the key properties of the columns of $(\mathbf{WA})\odot\tilde{\mathbf{F}}$ is that they are vectorized form of rank-1 matrices of the form $\tilde{\mathbf{f}}(\tau)\mathbf{c}_\ell^T$, where $[\tilde{\mathbf{f}}(\tau)]_{k+K'+1} = e^{-j2\pi\Delta_f k\tau}$, and $\mathbf{c}_\ell = \mathbf{Wa}(\theta_\ell)$. Such rank-1

[10]Conditions under which such an assumption is valid are provided in Sec. 4.4.3

matrices satisfy the following row-shift-invariance property:

Definition 20. *[KP19a, SDL16a] For a matrix $\mathbf{X} \in \mathbb{C}^{\tilde{K} \times N}$, we say it has row-shift-invariance property if*

$$\overline{\mathbf{X}} = \alpha \underline{\mathbf{X}} \tag{4.35}$$

for some $\alpha \in \mathbb{C}$, $\overline{\mathbf{X}}, \underline{\mathbf{X}} \in \mathbb{C}^{(\tilde{K}-1) \times N}$ are the matrices obtained by removing the first, and last row of $\mathbf{X}$, respectively.

It can be easily verified that for $\mathbf{X} = \tilde{\mathbf{f}}(\omega)\mathbf{c}_\ell^T$, we have $\alpha = e^{-j2\pi\Delta_f \tau}$ in (4.35). We define $\mathbf{U}_1, \mathbf{U}_2 \in \mathbb{C}^{(\tilde{K}-1)N \times L}$ to be such that their ℓth column are given as

$$[\mathbf{U}_1]_\ell = \text{VEC}(\overline{\text{UNVEC}(\mathbf{u}_\ell)}), \tag{4.36}$$

$$[\mathbf{U}_2]_\ell = \text{VEC}(\underline{\text{UNVEC}(\mathbf{u}_\ell)}) \tag{4.37}$$

where $\mathbf{u}_\ell$ denotes the ℓth column of $\mathbf{U}$, $\text{UNVEC}(.)$ rearranges the elements of $\mathbf{u}_\ell$ into a $\tilde{K} \times N$ matrix, and $\text{VEC}(.)$ rearranges the elements of a $(\tilde{K}-1) \times N$ matrix into a vector.

Using the row-shift-invariance property of the columns of $\mathbf{WA} \odot \tilde{\mathbf{F}}$ (as discussed in Definition 20), we have $\overline{\tilde{\mathbf{F}}} = \underline{\tilde{\mathbf{F}}}\Lambda$, where $\Lambda = \text{diag}(e^{-j2\pi\Delta_f\tau_1}, \cdots, e^{-j2\pi\Delta_f\tau_\ell})$. Therefore, it holds that

$$\mathbf{WA} \odot \overline{\tilde{\mathbf{F}}} = (\mathbf{WA} \odot \underline{\tilde{\mathbf{F}}})\Lambda$$

Hence, upon making the assumption that $(\mathbf{WA}) \odot \overline{\tilde{\mathbf{F}}}$ is full-column rank[11] we can follow a ESPRIT-based approach similar to Sec. 4.4.2, and [RK89, SDL16a, SDL16b, KP19a] to obtain the parameters τ_ℓ. Estimation of the parameters θ_ℓ, p_ℓ is very straightforward, and along with other steps of this algorithm are summarized in Algo. 5.

[11] Exact conditions are presented in Sec. 4.4.3.

Algorithm 5 COACH-ESPRIT-II

1: Compute $\mathbf{R}^{(2)} = \text{FLATTEN}(\text{TENSORTOEP}^{(\mathbb{D}_{\mathbf{i}}^{\mathbb{I}})}(\mathcal{R}))$.
2: Compute t-SVD: $\mathbf{U}\Sigma\mathbf{V}^H = \mathbf{R}^{(2)}$.
3: Compute $\mathbf{U}_1, \mathbf{U}_2$ using (4.36), (4.37).
4: Compute EVD: $\mathbf{U}_1^{\dagger}\mathbf{U}_2 = \mathbf{E}\Lambda\mathbf{E}^{-1}$
5: $\hat{\tau}_\ell = -\text{Im}(\log(\lambda_\ell)/(2\pi\Delta_f))$
6: Let $\tilde{\mathbf{U}}_\ell\tilde{\Sigma}_\ell\tilde{\mathbf{V}}_\ell^H = \text{UNVEC}(\mathbf{U}\mathbf{e}_\ell)$.
7: Estimate $\hat{\theta}_\ell$ from

$$\hat{\theta_\ell} = \arg\max_{\theta} \frac{|\tilde{\mathbf{u}}_{\ell,1}^H \mathbf{W}\mathbf{a}(\theta)|}{\|\mathbf{W}\mathbf{a}(\theta)\|}$$

where $\tilde{\mathbf{u}}_{\ell,1}$ is first column of $\tilde{\mathbf{U}}_\ell$.
8: Form the matrix $[\hat{\tilde{\mathbf{F}}}]_{k+K'+1,\ell} = e^{-j2\pi k\Delta_f\hat{\tau}_\ell}$, $-K' \le k \le K', 1 \le \ell \le L$.
9: Form the matrix $[\hat{\mathbf{A}}]_{m,\ell} = e^{j\pi m\sin(\hat{\theta}_\ell)}$, $1 \le m \le M, 1 \le \ell \le L$.
10: Compute $\hat{\mathbf{p}} = \text{diag}\left((\mathbf{W}\hat{\mathbf{A}} \odot \hat{\tilde{\mathbf{F}}})^{\dagger}\mathbf{R}^{(2)}((\mathbf{W}\hat{\mathbf{A}} \odot \hat{\tilde{\mathbf{F}}})^H)^{\dagger}\right)$, where diag(.) puts the diagonal elements of a matrix into a vector.

Remark 33. *As we will show in the following, the algorithm* COACH-ESPRIT-II *is capable of recovering $O(NK^2)$ channel paths. However, compared to Algorithm 4, this algorithm has a higher computational complexity, due to the fact that $\mathbf{R}^{(2)}$ can potentially have a much higher dimensions than $\mathbf{R}^{(1)}$. If $L \le \min(O(M), O(K^2), O(N^2))$, one can still use algorithm* COACH-ESPRIT-I*, with a lower computational complexity.*

Identifiability Results

In this subsection, we show conditions under which the tensor decomposition algorithm proposed in Sec. 4.4.3 is able to recover the constrained CP decomposition.

Theorem 31. *If the matrix $(\mathbf{WA}) \odot (\overline{\Pi_{\mathbb{U}_{\mathbf{i}}}}(\mathbf{F}^* \odot \mathbf{F}))$ has full column rank then the tensor $\mathcal{R}$ has a unique Constrained CP decomposition with constraints $\{\mathbb{V}_{\mathbf{i}}, \mathbb{C}^N, \mathbb{V}_{\mathbf{i}}^*, \mathbb{C}^N\}$, and the algorithm* COACH-ESPRIT-II *can recover it.*

Proof. Since $(\mathbf{WA}) \odot \overline{\tilde{\mathbf{F}}}$ has full column rank, a nonsingular matrix $\mathbf{E}$ satisfying Line 4 of Algo. 5 exists and hence $\mathbf{E}$ and Λ can be identified by the computing EVD of $\mathbf{U}_1^{\dagger}\mathbf{U}_2$. Once Λ is recovered,

we can identify $\mathbf{F}$. Since $\mathbf{WA} \odot \tilde{\mathbf{F}} = \mathbf{UE}$, we can also identify $\mathbf{WA}$. In particular, the ℓth column of $\mathbf{WA}$, i.e., $\mathbf{Wa}(\theta_\ell)$ can be recovered by considering the top left singular vector of $\text{UNVEC}(\mathbf{Ue}_\ell)$, as described in Line 6, and 7 of Algo. 5. $\square$

Remark 34. *A necessary condition for* $\overline{\tilde{\mathbf{F}}} \odot (\mathbf{WA})$ *to be full column rank is that* $L \leq (\tilde{K}-1)N$. *We conjecture that this is also a sufficient condition for almost all* $\mathbf{W}$, τ, θ, *however, the proof is beyond the scope of this Section.*

4.4.4 On identifying θ from $\mathbf{Wa}(\theta) = \alpha\mathbf{b}$

Under suitable conditions, Theorems 28 and 31 ensure that COACH-ESPRIT-I (Algo. 4) and COACH-ESPRIT-II (Algo. 5) can identify the tensor factors $\mathbf{f}(\tau_\ell)$, and $\mathbf{Wa}(\theta_\ell)$ ($\ell = 1, \cdots, L$) correctly. Given $\mathbf{b}_\ell = \mathbf{Wa}(\theta_\ell)$, Line 6 of COACH-ESPRIT-I and Line 7 of COACH-ESPRIT-II can estimate θ_ℓ as

$$\hat{\theta}_\ell = \arg\max_{\theta} \frac{|(\mathbf{b}_\ell \otimes \mathbf{b}_\ell)^H ((\mathbf{Wa}(\theta))^* \otimes (\mathbf{Wa}(\theta)))|}{\|(\mathbf{Wa}(\theta))^* \otimes (\mathbf{Wa}(\theta))\|}, \tag{4.38}$$

and

$$\hat{\theta}_\ell = \arg\max_{\theta} \frac{|\mathbf{b}_\ell^H \mathbf{Wa}(\theta)|}{\|\mathbf{Wa}(\theta)\|}, \tag{4.39}$$

respectively. We are interested in knowing when $\hat{\theta}_\ell$ matches the true θ_ℓ. To understand this, first notice that, in absence of noise, the solution to both (4.38) and (4.39) satisfy $\mathbf{Wa}(\hat{\theta}_\ell) = \alpha\mathbf{b}_\ell$, for some $\alpha \in \mathbb{C}$. Given $\mathbf{W}$ and $\mathbf{b}_\ell$, define the set

$$\mathbb{S}_{\mathbf{W},\mathbf{b}_\ell} = \{\theta \in [-\pi, \pi) | \mathbf{Wa}(\theta) = \alpha\mathbf{b}_\ell, \text{for some } \alpha \in \mathbb{C}\}.$$

It follows that

$$\hat{\theta}_\ell = \theta_\ell \text{ if and only if } \mathbb{S}_{\mathbf{W},\mathbf{b}_\ell} = \{\theta_\ell\}$$

where θ_ℓ is the true AoA of the ℓth channel path. In other words, if $\mathbf{b}_\ell$ is recovered successfully (which is guaranteed by Theorems 28 and 31), then COACH-ESPRIT-I and COACH-ESPRIT-II will also successfully recover θ_ℓ from $\mathbf{b}_\ell$ if and only if $\mathbb{S}_{\mathbf{W},\mathbf{b}_\ell}$ is a singleton. It is however, nontrivial to ascertain when this happens, primarily because $\mathbf{W}$ is a fat matrix with a nontrivial null-space, and $\mathbf{a}(\theta)$ has a specific geometry imparted by the array. The condition under which $\mathbb{S}_{\mathbf{W},\mathbf{b}_\ell}$ is a singleton is obviously dependent on $\mathbf{W}$ and has been characterized only for specific designs of $\mathbf{W}$ [AG05, QFS19]. We review the results in [AG05] in the following subsection.

Existing results on identifiability of θ_ℓ

The authors in [AG05] consider the identifiability issues associated with beamspace processing algorithms (such as beamspace MUSIC [AG05], or beamspace ESPRIT [XSRK94]). Here, we review some of the interesting results from [AG05] which also guarantees $\mathbb{S}_{\mathbf{W},\mathbf{b}_\ell} = \{\theta_\ell\}$.

Proposition 2. *[AG05] Suppose*

$$\mathcal{N}(\mathbf{W}) = span\{\mathbf{a}(\phi_1), \cdots, \mathbf{a}(\phi_{M-N})\},$$

Then, if $\theta_0 \notin \{\phi_1, \cdots, \phi_{M-N}\}$*, we have* $\mathbf{Wa}(\theta) = \alpha\mathbf{Wa}(\theta_0)$ *if and only if* $\theta = \theta_0$.

Although this result exactly characterizes the ambiguous set for which the problem $\mathbf{Wa}(\theta) = \alpha\mathbf{b}$ does not have a unique solution in θ, the result holds only for very specific $\mathbf{W}$ matrices whose null-space is span of vectors of the form $\mathbf{a}(\phi)$. This in general might not be the case, and in many scenarios a more general beamforming matrices $\mathbf{W}$ could be used [NZL17, KP18a].

A verifiable condition to test identifiability of θ_ℓ given W and $\mathbf{b}_\ell$

In this section, we will provide a more general condition which could be checked for any given matrix $\mathbf{W}$. Our approach is based on testing for the existence of a rank-1 matrix in a given subspace, and is inspired by the ideas in [BVD$^+$18], and the shift-invariance property of Vandermonde matrices. Let us first make a few definitions:

Definition 21. *For a vector $\mathbf{x} \in \mathbb{C}^M$, define the following $(M-1)\times 2$ Hankel matrix as*

$$\mathcal{H}_2(\mathbf{x}) = \begin{bmatrix} x_1 & x_2 \\ x_2 & x_3 \\ \vdots & \vdots \\ x_{M-1} & x_M \end{bmatrix}.$$

Definition 22. *For a matrix $\mathbf{X} \in \mathbb{C}^{M\times N}$, define the compound matrix $C_2(\mathbf{X}) \in \mathbb{C}^{\binom{M}{2}\times\binom{N}{2}}$ to be a matrix comprised of all 2×2 minors of $\mathbf{X}$. In particular, the $(\frac{(m_2-1)(m_2-2)}{2}+m_1, \frac{(n_2-1)(n_2-2)}{2}+n_1)$th element of $C_2(\mathbf{X})$ equals*

$$X_{m_1,n_1}X_{m_2,n_2} - X_{m_1,n_2}X_{m_2,n_1},$$

where $1 \le m_1 < m_2 \le M, 1 \le n_1 < n_2 \le N$. It is well known that rank$(\mathbf{X}) \le 1$ if and only if $C_2(\mathbf{X}) = 0$.

Given the vector $\mathbf{b}_\ell$, we define $\mathbf{v}_\ell = \mathbf{W}^\dagger \mathbf{b}_\ell$ and $\tilde{\mathbf{V}}_\ell = [\mathbf{v}_\ell, \mathbf{V}]$, where $\mathbf{V} \in \mathbb{C}^{M\times r}$ $(r \ge M-N)$ is a basis for $\mathcal{N}(\mathbf{W})$ (i.e., the nullspace of $\mathbf{W}$).

Define $\tilde{\Psi}(\mathbf{W}, \mathbf{b}_\ell) \in \mathbb{C}^{\binom{M-1}{2\times(r+1)^2}}$ to be such that its $(\frac{(n-1)(n-2)}{2}+m, \frac{q_2(q_2-1)}{2}+q_1)$th element

equals

$$\beta_{q_1,q_2}\Big(\tilde{V}^{(\ell)}_{m,q_1}\tilde{V}^{(\ell)}_{n+1,q_2}-\tilde{V}^{(\ell)}_{m+1,q_1}\tilde{V}^{(\ell)}_{n,q_2} + \tilde{V}^{(\ell)}_{m,q_2}\tilde{V}^{(\ell)}_{n+1,q_1}-\tilde{V}^{(\ell)}_{m+1,q_2}\tilde{V}^{(\ell)}_{n,q_1}\Big), \tag{4.40}$$

where $1 \le q_1 \le q_2 \le r+1, 1 \le m < n \le M-1$, $\tilde{V}^{(\ell)}_{s_1,s_2}$ is the (s_1, s_2)th element of $\tilde{\mathbf{V}}_\ell$, and $\beta_{q_1,q_2} = 1$ if $q_1 \neq q_2$, and $\beta_{q_1,q_2} = \frac{1}{2}$ if $q_1 = q_2$.

Equipped with these notations and definitions, in the following Theorem we state a verifiable sufficient condition in terms of $\mathbf{W}$ and $\mathbf{b}_\ell$ under which the set $\mathbb{S}_{\mathbf{W},\mathbf{b}_\ell}$ is a singleton:

Theorem 32. *Given* $\mathbf{b}_\ell$*, and* $\mathbf{W}$*, the set* $\mathbb{S}_{\mathbf{W},\mathbf{b}_\ell}$ *is a singleton if dim*$(\mathcal{N}(\tilde{\psi}(\mathbf{W},\mathbf{b}_\ell))) = 1$*, where* $\tilde{\psi}(\mathbf{W},\mathbf{b}_\ell)$ *is defined in* (4.40)*.*

Proof. We prove by contradiction. Suppose $\exists\bar{\theta}_\ell \neq \theta_\ell$ such that $\bar{\theta}_\ell \in \mathbb{S}_{\mathbf{W},\mathbf{b}_\ell}$. This implies

$$\mathbf{W}\mathbf{a}(\theta_\ell) = \alpha\mathbf{b}_\ell, \text{and } \mathbf{W}\mathbf{a}(\bar{\theta}_\ell) = \bar{\alpha}\mathbf{b}_\ell$$

for some $\alpha, \bar{\alpha} \in \mathbb{C}$. Therefore, we have $\mathbf{a}(\theta_\ell) = \tilde{\mathbf{V}}_\ell\mathbf{c}$, and $\tilde{\mathbf{V}}_\ell\bar{\mathbf{c}}$ for some $\mathbf{c}, \bar{\mathbf{c}} \in \mathbb{C}^{r+1}$, where the first element of $\mathbf{c}$ (resp. $\bar{\mathbf{c}}$) equals α (resp. $\bar{\alpha}$).

Notice that since $\mathbf{a}(\theta)$ is the array steering vector of a uniform linear array, it is easy to verify that the Hankel matrix $\mathcal{H}_2(\mathbf{a}(\theta)) \in \mathbb{C}^{(M-1)\times 2}$ is rank one, where $\mathcal{H}_2(.)$ is defined in Definition 21. Hence, solving the following problem

$$\text{find } \mathbf{c} \in \mathbb{C}^{r+1} \text{ s.t. } \text{rank}(\mathcal{H}_2(\mathbf{V}_\ell\mathbf{c})) = 1, [\mathbf{V}_\ell]^T_{1,:}\mathbf{c} = 1$$

is equivalent to finding the vector $\mathbf{c} \in \mathbb{C}^{r+1}$ that satisfies $\mathbf{a}(\theta) = \tilde{\mathbf{V}}_\ell\mathbf{c}$, where $[\mathbf{V}_\ell]^T_{1,:}$ denotes the first row of $\mathbf{V}_\ell$, and the condition $[\mathbf{V}_\ell]^T_{1,:}\mathbf{c} = 1$ ensures that $\mathbf{c}$ is scaled correctly. Using Definition

22, the condition $\text{rank}(\mathcal{H}_2(\mathbf{V}_\ell \mathbf{c})) = 1$ is equivalent to having

$$C_2(\mathcal{H}_2(\mathbf{V}_\ell \mathbf{c})) = 0 \tag{4.41}$$

Equation (4.41) implies

$$[\mathcal{H}_2(\tilde{\mathbf{V}}_\ell \mathbf{c})]_{m,1}[\mathcal{H}_2(\tilde{\mathbf{V}}_\ell \mathbf{c})]_{n,2} - [\mathcal{H}_2(\tilde{\mathbf{V}}_\ell \mathbf{c})]_{m,2}[\mathcal{H}_2(\tilde{\mathbf{V}}_\ell \mathbf{c})]_{n,1} = 0$$

Using the fact that $\mathcal{H}_2(\mathbf{V}_\ell \mathbf{c}) = \sum_{q=1}^{r+1} c_q \mathcal{H}_2(\tilde{\mathbf{v}}_q^{(\ell)})$ (where $\tilde{\mathbf{v}}_q^{(\ell)}$ is the ℓth column of $\tilde{\mathbf{V}}_\ell$, and c_q is the qth element of $\mathbf{c}$) we get

$$\sum_{q_1=1}^{r+1} \sum_{q_2=1}^{r+1} c_{q_1} c_{q_2} \psi^{(m,n)}(\tilde{\mathbf{v}}_{q_1}^{(\ell)}, \tilde{\mathbf{v}}_{q_2}^{(\ell)}) = 0$$

where

$$\begin{aligned} \psi^{(m,n)}(\tilde{\mathbf{v}}_{q_1}^{(\ell)}, \tilde{\mathbf{v}}_{q_2}^{(\ell)}) = {} & [\mathcal{H}_2(\tilde{\mathbf{v}}_{q_1}^{(\ell)})]_{m,1}[\mathcal{H}_2(\tilde{\mathbf{v}}_{q_2}^{(\ell)})]_{n,2} \\ & - [\mathcal{H}_2(\tilde{\mathbf{v}}_{q_1}^{(\ell)})]_{m,2}[\mathcal{H}_2(\tilde{\mathbf{v}}_{q_2}^{(\ell)})]_{n,1} \end{aligned}$$

Or equivalently,

$$\psi(\mathbf{W}, \mathbf{b}_\ell)(\mathbf{c} \otimes \mathbf{c}) = \mathbf{0} \tag{4.42}$$

where the $(\frac{(n-1)(n-2)}{2} + m, q_1 + (r+1)(q_2 - 1))$th element of $\psi(\mathbf{W}, \mathbf{b}_\ell) \in \mathbb{C}^{\binom{M-1}{2} \times (r+1)^2}$ is equal to $\psi^{(m,n)}(\tilde{\mathbf{v}}_{q_1}^{(\ell)}, \tilde{\mathbf{v}}_{q_2}^{(\ell)})$, for $1 \leq m < n \leq M-1$, $1 \leq q_1, q_2 \leq r+1$.

We also know that $\mathbf{c} \otimes \mathbf{c}$ is the vectorized form of the symmetric matrix $\mathbf{c}\mathbf{c}^T$. Therefore, $\mathbf{c} \in \mathcal{S}_{r+1}$, where $\mathcal{S}_{r+1}$ denotes the subspace corresponding to vectorized form of $(r+1) \times (r+1)$

symmetric matrices. Hence, $\mathbf{c}\otimes\mathbf{c}\in\mathcal{N}(\psi(\mathbf{W},\mathbf{b}_\ell))\cap\mathcal{S}_{r+1}$. This is equivalent to having

$$\tilde{\psi}(\mathbf{W},\mathbf{b}_\ell)\text{uptri}(\mathbf{c}\otimes\mathbf{c})=\mathbf{0} \tag{4.43}$$

where the matrix $\tilde{\psi}(\mathbf{W},\mathbf{b}_\ell)$ is defined in (4.40), and $\text{uptri}(\mathbf{c}\otimes\mathbf{c})\in\mathbb{C}^{\binom{r+2}{2}}$ is a vector containing the upper-triangular (including the diagonal) elements of $\mathbf{c}\mathbf{c}^T$. Notice that if equation (4.43) holds for $\mathbf{c}$, it also holds for any $\beta\mathbf{c}$, where $\beta\in\mathbb{C}$. In order to resolve this scaling ambiguity, we can use the fact that $[\mathbf{V}_\ell]_{1,:}^T\mathbf{c}=1$. Hence, solving $\mathbf{a}(\theta)=\tilde{\mathbf{V}}_\ell\mathbf{c}$ is equivalent to finding $\mathbf{c}$ that satisfies (4.43) and is scaled such that $[\mathbf{V}_\ell]_{1,:}^T\mathbf{c}=1$.

Following a similar procedure for $\bar{\theta}$, it holds that

$$\tilde{\psi}(\mathbf{W},\mathbf{b}_\ell)\text{uptri}(\bar{\mathbf{c}}\otimes\bar{\mathbf{c}})=\mathbf{0}. \tag{4.44}$$

for some $\bar{\mathbf{c}}\in\mathbb{C}^{r+1}$, such that $\mathbf{a}(\bar{\theta})=\tilde{\mathbf{V}}\bar{\mathbf{c}}$. However, since $\dim\mathcal{N}(\tilde{\psi}(\mathbf{W},\mathbf{b}_\ell))=1$, from (4.43) and (4.44), we must have

$$\text{uptri}(\mathbf{c}\otimes\mathbf{c})=\gamma\text{uptri}(\bar{\mathbf{c}}\otimes\bar{\mathbf{c}})$$

for some nonzero scalar $\gamma\in\mathbb{C}$. Therefore, it holds that $c_q^2=\gamma\bar{c}_q^2$, i.e.,[12] $\mathbf{c}=\pm\sqrt{\gamma}\bar{\mathbf{c}}$. Using the fact that $[\mathbf{V}_\ell]_{1,:}^T\mathbf{c}=[\mathbf{V}_\ell]_{1,:}^T\bar{\mathbf{c}}=1$, we conclude $\mathbf{c}=\bar{\mathbf{c}}$. Hence, $\mathbf{a}(\theta_\ell)=\mathbf{a}(\bar{\theta}_\ell)$, which implies $\theta_\ell=\bar{\theta}_\ell$. Therefore, $\mathbb{S}_{\mathbf{W},\mathbf{b}_\ell}=\{\theta_\ell\}$.

□

Remark 35. *The algorithm presented in this section can also be extended to non-uniform linear arrays, whose difference co-array contains a uniform linear array (such as nested [PV10], and coprime [PV11] arrays), although the details can be nontrivial. However, we leave this derivation for future.*

[12] Square root of a complex number $\gamma=r_\gamma e^{j\phi_\gamma}$ is defined as $\sqrt{\gamma}=\sqrt{r_\gamma}e^{j\phi_\gamma/2}$, where $\gamma_r\in\mathbb{R}$, $\gamma_r>0$ and $\gamma_c\in[0,2\pi)$.

4.4.5 Numerical Experiments

In this section, we present our numerical experiments to evaluate the performance of the proposed algorithms. We consider a SIMO mm-Wave communication system with $N = 8$ RF-chains and $M = 16$ uniformly located antennas with inter-element spacing of $\lambda/2$. We use $K = 8$ frequency bins allocated to the pilots, located in a non-uniform fashion based on the nested array geometry. In particular, we use the frequency bins $\mathbf{i} = [1,2,3,4,9,14,19]$, which yields a difference co-array of size $\tilde{K} = 37$.

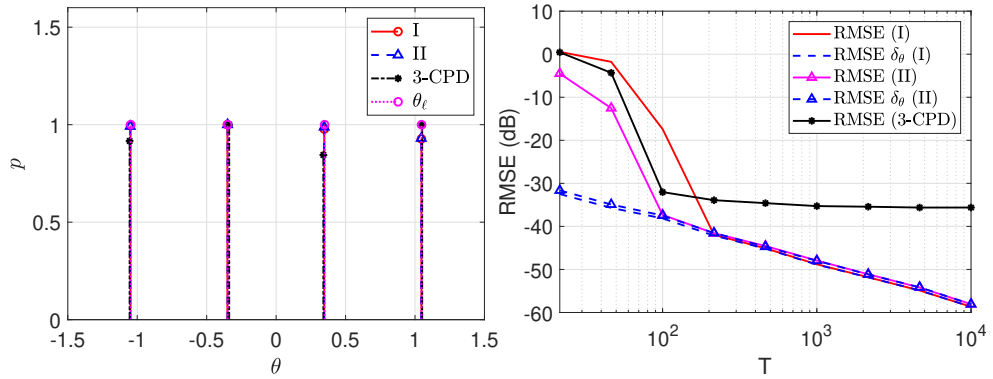
For the first set of experiments, the AoAs θ_ℓ corresponding to the channel paths are located uniformly in the range $[-\frac{\pi}{3}, \frac{\pi}{3}]$, for $\ell = 1, \cdots, L$. The channel path delays (τ_ℓ/Δ_f modulo 2π) are chosen uniformly in the range $[-\frac{2\pi}{5}, \frac{2\pi}{5}]$, and the powers are set to $p_\ell = 1$, for $\ell = 1, \cdots, L$. For all simulations in this experiment, we use a randomly generated $\mathbf{W}$ matrix with unimodular elements, and $\mathbf{W}$ is kept fixed for all simulations. We examine the performance of the algorithms COACH-ESPRIT-I and COACH-ESPRIT-II (denoted as "I", and "II" in Fig. 4.7), as well as the algorithm presented in [PAGH19] (denoted as "3-CPD" in Fig. 4.7), where the time snapshots of the measurements from the RF-chains are directly used and CP decomposition (CPD) is performed to estimate the AoAs corresponding to the channel paths. The CPD is implemented using ALS algorithm in TensorLab3.0 package [VDDL16]. Figures 4.7a, 4.7c, and 4.7e show the recovered $\hat{\mathbf{p}}$ (normalized w.r.t $\max_\ell \hat{p}_\ell$ for better visibility) and $\hat{\theta}$ using Algorithms COACH-ESPRIT-I and COACH-ESPRIT-II, where $T = 500$ and $\sigma = 2\sqrt{L}$. We observe that the algorithm COACH-ESPRIT-II can find the correct AoAs for all cases $L = 4, 9, 12$. The algorithm COACH-ESPRIT-I fails when $L = 12$, and the Algorithm in [PAGH19] can only recover the true AoAs when $K = 4$.

Figures 4.7b, 4.7d, and 4.7f show the Root Mean Squared Error (RMSE), defined as $\text{RMSE (dB)} = 10\log_{10} \frac{\sum_{g=1}^{N_{\text{runs}}} \|\hat{\theta}^{(g)} - \theta\|}{N_{\text{runs}}}$, corresponding to the algorithms COACH-ESPRIT-I and COACH-ESPRIT-II, with $\hat{\theta}^{(g)}$ denoting the output of the algorithms at gth run, for $N_{\text{runs}} = 100$ Monte-Carlo runs. The RMSE is calculated with respect to the number of snapshots T. We also

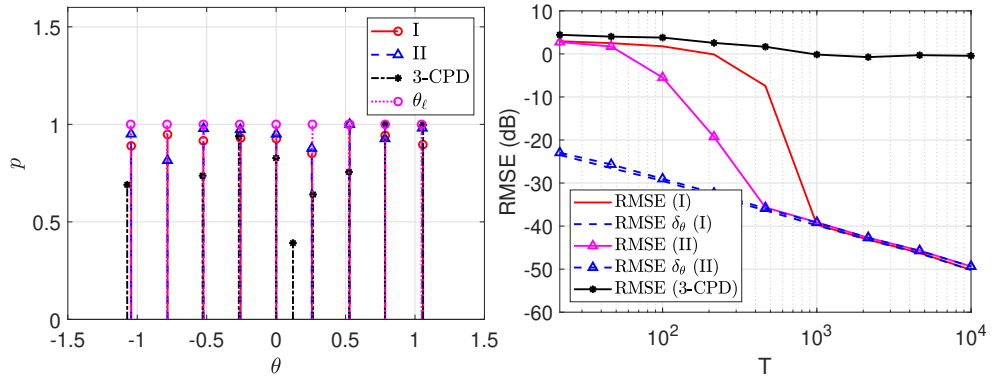
plot the RMSE corresponding to the first order perturbations of the estimated parameters δ_{θ_i}, using the formulation derived in Sec. 4.4.2 (and similar calculations for Algorithm COACH-ESPRIT-II are also derived, however, we skipped those derivations in the paper for brevity). We observe that for the cases where the number of channel paths are small ($L = 4 < N$), the Algorithm in [PAGH19] is able to find the AoAs with a relatively good accuracy. However, its performance does not improve with increasing T. On the other hand, when the number of channel paths is larger than N, i.e., $L = 9, L = 12$ both of our proposed algorithms perform much better than [PAGH19], and the algorithm COACH-ESPRIT-II shows the best performance, while requiring fewer number of snapshots than that of COACH-ESPRIT-I in order to achieve the performance predicted by the RMSE of δ_θ. In the second set of experiments, we show the phase transition plots corresponding to the algorithms COACH-ESPRIT-I, and COACH-ESPRIT-II. The white pixels indicate the regions where the algorithm is able to exactly recover the channel parameters in a noiseless setting. We can see that the plots also match with theoretical bounds derived in Sec. 4.4.2 and 4.4.3. In Figures 4.8a, 4.8b, 4.8c and 4.8d we keep $N = 3$, $N = 10$, $K = 3$, $K = 10$ fixed, respectively. In all figures, we have $M = 20$. In all plots of Fig. 4.8, we use the algorithm used in the constructive proof of Theorem 32 to identify the channel paths θ.

4.4.6 Conclusion

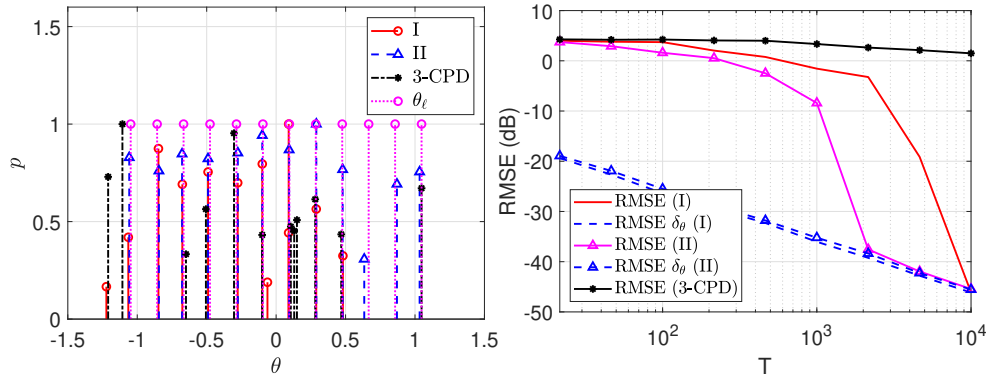
We considered the problem of channel estimation for SIMO multi-carrier communication systems, in presence of correlation priors on the channel paths. We showed that once the pilots used for channel training are judiciously placed on certain non-uniform locations, through proposing two linear algebraic based tensor decomposition algorithms, we are able to recover L channel paths that can be as large as $\min(O(N^2), O(M), O(K^2))$, and can further improve it to $L = O(NK^2)$, with N, M, and K being the number of RF-chains, number of antennas, and number of pilot frequency bins used, respectively. Hence, unlike many of the existing algorithms that can only work when $L = O(N)$, our proposed approach can handle cases where $L \gg N$. We



(a) $\hat{\mathbf{p}}$ and $\hat{\theta}$, $L = 4$ (b) RMSE vs T, $L = 4$

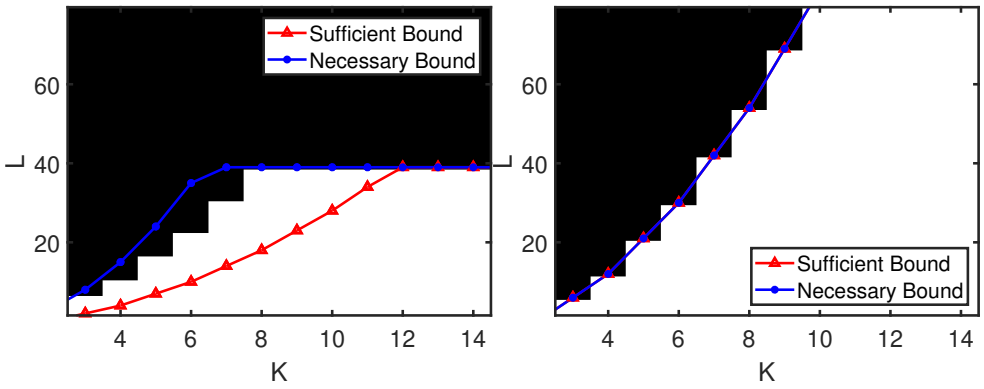


(c) $\hat{\mathbf{p}}$ and $\hat{\theta}$, $L = 9$ (d) RMSE vs T, $L = 9$

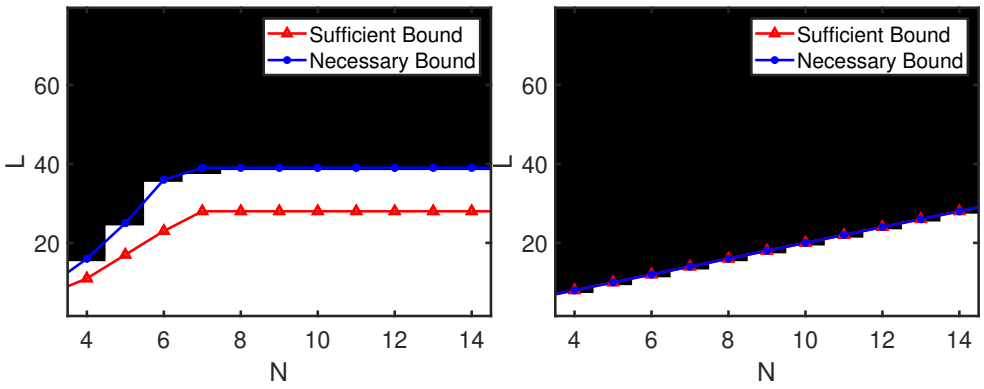


(e) $\hat{\mathbf{p}}$ and $\hat{\theta}$, $L = 12$ (f) RMSE vs T, $L = 12$

Figure 4.7: Comparison between performance of algorithms COACH-ESPRIT-I, COACH-ESPRIT-II, and [PAGH19]. Figures (a), (c), (e): recovered $\hat{\theta}_\ell$, $\hat{p}_\ell$. Figures (b), (d), (f) RMSE as a function T for different number of channel paths L.



(a) COACH-ESPRIT-I K vs L (b) COACH-ESPRIT-II K vs L



(c) COACH-ESPRIT-I N vs L (d) COACH-ESPRIT-II N vs L

Figure 4.8: Phase transition corresponding to the algorithms COACH-ESPRIT-I, and COACH-ESPRIT-II.

also analyzed the performance of the proposed algorithms in the case that we have access to a perturbed version of the covariance matrix, due to finite number of snapshots, and show how the estimates of the channel parameters will change, when we have small perturbations. Finally, we numerically established the superior performance of our proposed algorithms.

4.5 Appendix

4.5.1 Proof of Theorem 30

Proof. Similar to the analysis in [LLV93, LLM08] the perturbed left singular vectors of $\hat{\mathbf{R}}^{(1)}$ are given as $\hat{\mathbf{U}} = \mathbf{U} + \Delta_{\mathbf{U}}$, where $\Delta_{\mathbf{U}}$ is given in (4.32).

Upon computing the perturbations of eigenvalues of $(\underline{\Pi_{\mathbb{U}_{\mathbf{i}}}\mathbf{U}})^{\dagger}\overline{\Pi_{\mathbb{U}_{\mathbf{i}}}}\mathbf{U} = \mathbf{E}\Lambda\mathbf{E}^{-1}$, and applying the results from [LLV93] we derive equation (4.31).

We now proceed into finding the perturbations in the recovered AoAs $\hat{\theta}_{\ell}$ up to the first order. For brevity, let us define $\mathbf{F}^{\odot 2} = \mathbf{F}^{*} \odot \mathbf{F}$. It is straightforward to show that

$$\Delta_{\mathbf{F}^{\odot 2}} = j\pi \mathrm{diag}(\mathbf{i} \otimes \mathbf{1}_K - \mathbf{1}_K \otimes \mathbf{i})\mathbf{F}^{\odot 2}\mathrm{diag}(\delta_{\tau})$$

where $\mathbf{i} = [i_1, i_2, \cdots, i_K]^T$. Using first order perturbations of pseudo-inverses [Wed73], we have

$$\Delta_{(\mathbf{F}^{\odot 2})^{\dagger}} = ((\mathbf{F}^{\odot 2})^H\mathbf{F}^{\odot 2})^{-1}\Delta^H_{\mathbf{F}^{\odot 2}}\Pi^{\perp}_{\mathbf{F}^{\odot 2}} - (\mathbf{F}^{\odot 2})^{\dagger}\Delta_{\mathbf{F}^{\odot 2}}(\mathbf{F}^{\odot 2})^{\dagger}$$

where $\Pi^{\perp}_{\mathbf{F}^{\odot 2}} = \mathbf{I}_{K^2} - \mathbf{F}^{\odot 2}((\mathbf{F}^{\odot 2})^H\mathbf{F}^{\odot 2})^{-1}(\mathbf{F}^{\odot 2})^H$. Define $\mathbf{B}^{\odot 2} := (\mathbf{WA})^{*} \odot (\mathbf{WA})$, then it holds that

$$\Delta_{\mathbf{B}^{\odot 2}} = (\Delta_{(\mathbf{F}^{\odot 2})^{\dagger}}\mathbf{R}^{(1)} + (\mathbf{F}^{\odot 2})^{\dagger}\Delta_{\mathbf{R}^{(1)}})^T$$

Now, given the perturbations in $\Delta_{\mathbf{B}^{\odot 2}} = [\delta_{\mathbf{b}_1}, \cdots, \delta_{\mathbf{b}_L}]$, our goal is to find the perturbations corresponding to the recovered $\hat{\theta}_{\ell} = \theta_{\ell} + \delta_{\theta_{\ell}}$, up to the first order. For notational brevity, we first make a few definitions: Let $\mathbf{a}_{\mathrm{kr}}(\theta) = \mathbf{a}^{*}(\theta) \otimes \mathbf{a}(\theta)$, $\dot{\mathbf{a}}_{\mathrm{kr}}(\theta) = \frac{d\mathbf{a}_{\mathrm{kr}}(\theta)}{d\theta}$, and $\ddot{\mathbf{a}}_{\mathrm{kr}}(\theta) = \frac{d^2\mathbf{a}_{\mathrm{kr}}(\theta)}{d\theta^2}$. Define

$\hat{\mathbf{b}}_\ell$ to be the ith column of $\mathbf{B}^{\odot 2} + \Delta_{\mathbf{B}^{\odot 2}}$. Let $\mathbf{W}^{\otimes 2} = \mathbf{W}^* \otimes \mathbf{W}$, and

$$\hat{u}(\theta) = |\hat{\mathbf{b}}_\ell^H \mathbf{W}^{\otimes 2} \mathbf{a}_{\mathrm{kr}}(\theta)|^2$$
$$v(\theta) = \|\mathbf{W}^{\otimes 2} \mathbf{a}_{\mathrm{kr}}(\theta)\|_2$$

Define $\hat{g}(\theta) = \log(\hat{u}(\theta)/v(\theta))$, where $\hat{u}(\theta)/v(\theta)$ is the objective function in (4.29), when $\mathbf{b}_\ell$ is perturbed to $\hat{\mathbf{b}}_\ell$. Our goal is to approximate $\hat{\theta}_\ell = \arg\max_\theta \hat{g}(\theta)$. Let $\theta_\ell = \arg\max_\theta g(\theta)$. We have $\frac{d\hat{g}(\theta)}{d\theta}|_{\theta=\hat{\theta}_\ell} = 0$. Assuming that $\hat{\theta}_\ell - \theta_\ell = \delta_{\theta_\ell}$ is small enough, we can approximate $\frac{d\hat{g}(\theta)}{d\theta}|_{\theta=\hat{\theta}_\ell}$ as

$$0 = \frac{d\hat{g}(\theta)}{d\theta}|_{\theta=\hat{\theta}_\ell} = \frac{d\hat{g}(\theta)}{d\theta}|_{\theta=\theta_\ell} + (\hat{\theta}_\ell - \theta_\ell)\frac{d^2\hat{g}(\theta)}{d\theta^2}|_{\theta=\theta_\ell}$$

Hence, $\delta_{\theta_\ell} = \hat{\theta}_\ell - \theta_\ell = -\frac{\dot{\hat{g}}(\theta_\ell)}{\ddot{\hat{g}}(\theta_\ell)}$, where the notation $\dot{\hat{g}}(\theta_\ell)$ stands for $\frac{d\hat{g}(\theta)}{d\theta}|_{\theta=\theta_\ell}$, and $\ddot{\hat{g}}(\theta_\ell)$ denotes $\frac{d^2\hat{g}(\theta)}{d\theta^2}|_{\theta=\theta_\ell}$. Using the notations $\hat{g}(\theta_\ell) = g(\theta_\ell) + \delta_g(\theta_\ell)$, $\dot{\hat{g}}(\theta_\ell) = \dot{g}(\theta_\ell) + \delta_{\dot{g}}(\theta_\ell)$, and $\ddot{\hat{g}}(\theta_\ell) = \ddot{g}(\theta_\ell) + \delta_{\ddot{g}}(\theta_\ell)$, we can write

$$\delta_{\theta_\ell} = -\frac{\dot{g}(\theta_\ell) + \delta_{\dot{g}}(\theta_\ell)}{\ddot{g}(\theta_\ell) + \delta_{\ddot{g}}(\theta_\ell)} \approx -\frac{\delta_{\dot{g}}(\theta_\ell)}{\ddot{g}(\theta_\ell)}$$

where we used the fact that $\dot{g}(\theta_\ell) = 0$, as θ_ℓ is the optimum value for $g(\theta)$. Therefore, we only need to compute the quantities $\delta_{\dot{g}}(\theta_\ell)$ and $\ddot{g}(\theta_\ell)$. Defining the notations $\dot{u}, \dot{\hat{u}}, \dot{v}, \ddot{u}, \ddot{\hat{u}}, \ddot{v}$ corresponding to $\hat{u}, u, v$, we can write the expansion of $\hat{u}$ upto the first order in $\delta_{\mathbf{b}_\ell}$ as

$$\hat{u}(\theta) \approx |\mathbf{b}_\ell^H \mathbf{W}^{\otimes 2} \mathbf{a}_{\mathrm{kr}}(\theta)|^2 + \delta_u(\theta)$$

where $\delta_u(\theta) = 2\mathrm{Re}((\delta_{\mathbf{b}_\ell}^H \mathbf{W}^{\otimes 2} \mathbf{a}_{\mathrm{kr}}(\theta))^*(\mathbf{b}_\ell^H \mathbf{W}^{\otimes 2} \mathbf{a}_{\mathrm{kr}}(\theta)))$. Moreover, the first order derivatives of

$\hat{u}, v$ with respect to θ can be computed as

$$\dot{\hat{u}}(\theta) \approx \overbrace{2\mathrm{Re}((\mathbf{b}_\ell^H \mathbf{W}^{\otimes 2}\mathbf{a}_{\mathrm{kr}}(\theta))^* \mathbf{b}_\ell^H \mathbf{W}^{\otimes 2}\dot{\mathbf{a}}_{\mathrm{kr}}(\theta))}^{\dot{u}(\theta)} + \delta_{\dot{u}}(\theta)$$
$$\dot{v}(\theta) = 2\mathrm{Re}((\mathbf{W}^{\otimes 2}\dot{\mathbf{a}}_{\mathrm{kr}}(\theta))^H (\mathbf{W}^{\otimes 2}\mathbf{a}_{\mathrm{kr}}(\theta)))$$

where upto the first order in $\delta_{\mathbf{b}_\ell}$ we have

$$\begin{aligned}\delta_{\dot{u}}(\theta) = {} & 2\mathrm{Re}((\delta_{\mathbf{b}_\ell}^H \mathbf{W}^{\otimes 2}\mathbf{a}_{\mathrm{kr}}(\theta))^* (\mathbf{b}_\ell^H \mathbf{W}^{\otimes 2}\dot{\mathbf{a}}_{\mathrm{kr}}(\theta))) \\ & + 2\mathrm{Re}((\mathbf{b}_\ell^H \mathbf{W}^{\otimes 2}\mathbf{a}_{\mathrm{kr}}(\theta))^* (\delta_{\mathbf{b}_\ell}^H \mathbf{W}^{\otimes 2}\dot{\mathbf{a}}_{\mathrm{kr}}(\theta)))\end{aligned}$$

Finally, the second derivatives of $u(\theta), v(\theta)$ can be computed as

$$\begin{aligned}\ddot{u}(\theta) = {} & 2\|\mathbf{b}_\ell^H \mathbf{W}^{\otimes 2}\dot{\mathbf{a}}_{\mathrm{kr}}(\theta)\|^2 \\ & + 2\mathrm{Re}((\mathbf{b}_\ell^H \mathbf{W}^{\otimes 2}\ddot{\mathbf{a}}_{\mathrm{kr}}(\theta))^* \mathbf{b}_\ell^H \mathbf{W}^{\otimes 2}\mathbf{a}_{\mathrm{kr}}(\theta)) \\ \ddot{v}(\theta) = {} & 2\|\mathbf{W}^{\otimes 2}\ddot{\mathbf{a}}_{\mathrm{kr}}(\theta)\|^2 + 2\mathrm{Re}((\mathbf{W}^{\otimes 2}\ddot{\mathbf{a}}_{\mathrm{kr}}(\theta_\ell))^H \mathbf{W}^{\otimes 2}\mathbf{a}_{\mathrm{kr}}(\theta)).\end{aligned}$$

From the definition of $\hat{g}$, we have

$$\dot{\hat{g}}(\theta) = \frac{\dot{\hat{u}}(\theta)}{\hat{u}(\theta)} - \frac{\dot{v}(\theta)}{v(\theta)} = \frac{\dot{u}(\theta) + \delta_{\dot{u}}(\theta)}{u(\theta) + \delta_u(\theta)} - \frac{\dot{v}(\theta)}{v(\theta)}$$

which up to first order of approximation in δ_{θ_ℓ} we have $\dot{\hat{g}}(\theta) \approx \dot{g}(\theta) + \delta_{\dot{g}}(\theta)$, where

$$\delta_{\dot{g}}(\theta) = \frac{\delta_{\dot{u}}(\theta)}{u(\theta)} - \frac{\delta_u(\theta)\dot{u}(\theta)}{u^2(\theta)}$$

It can also be verified that

$$\ddot{g}(\theta) = \frac{u(\theta)\ddot{u}(\theta) - (\dot{u}(\theta))^2}{u^2(\theta)} - \frac{v(\theta)\ddot{v}(\theta) - (\dot{v}(\theta))^2}{v^2(\theta)}.$$

which leads to (4.33). □

4.6 Acknowledgements

The contents in this chapter are reprints of the material published in “IEEE Data Science Workshop 2019” (Sec. 4.2), and “IEEE Sensor Array and Multichannel Signal Processing (SAM) 2020” (Sec. 4.3). Sec. 4.4 contains the material submitted to “IEEE Journal on Selected Topics in Signal Processing (JSTSP)”.

The work in Sec. 4.2 was supported in parts by the “Office of Naval Research grant (ONR N00014-18-1-2038)”, and the “University of California, San Diego”. The work in Sec. 4.3 and 4.4 was supported by in part by the “Office of Naval Research grant (ONR N00014-18-1-2038)”, and in part by the “National Science Foundation (NSF CAREER ECCS 1700506)”.

Bibliography

[AAS17] C. Aksoylar, G. K. Atia, and V. Saligrama. Sparse signal processing with linear and nonlinear observations: A unified shannon-theoretic approach. *IEEE Transactions on Information Theory*, 63(2):749–776, Feb 2017.

[ABB+07] Peter Almers, Ernst Bonek, A Burr, Nicolai Czink, Mérouane Debbah, Vittorio Degli-Esposti, Helmut Hofstetter, Pekka Kyösti, Dave Laurenson, Gerald Matz, et al. Survey of channel and radio propagation models for wireless mimo systems. *EURASIP Journal on Wireless Communications and Networking*, 2007(1):019070, 2007.

[AdA17] Daniel C Araújo and André LF de Almeida. Tensor-based compressed estimation of frequency-selective mmwave mimo channels. In *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 1–5. IEEE, 2017.

[AdAH19] Khaled Ardah, Andre LF de Almeida, and Martin Haardt. Low-complexity millimeter wave csi estimation in mimo-ofdm hybrid beamforming systems. In *WSA 2019; 23rd International ITG Workshop on Smart Antennas*, pages 1–5. VDE, 2019.

[AEALH13] Ahmed Alkhateeb, Omar El Ayach, Geert Leus, and Robert W Heath. Hybrid precoding for millimeter wave cellular systems with partial channel knowledge. In *2013 Information Theory and Applications Workshop (ITA)*, pages 1–5. IEEE, 2013.

[AEALH14] Ahmed Alkhateeb, Omar El Ayach, Geert Leus, and Robert W Heath. Channel estimation and hybrid precoding for millimeter wave cellular systems. *IEEE Journal of Selected Topics in Signal Processing*, 8(5):831–846, 2014.

[AG05] Ali Nasiri Amini and Tryphon T Georgiou. Avoiding ambiguity in beamspace processing. *IEEE Signal Processing Letters*, 12(5):372–375, 2005.

[AGPHJ18] Anum Ali, Nuria González-Prelcic, and Robert W Heath Jr. Spatial covariance estimation for millimeter wave hybrid systems using out-of-band information. *arXiv preprint arXiv:1804.11204*, 2018.

[AH16] Ahmed Alkhateeb and Robert W Heath. Frequency selective hybrid precoding for limited feedback millimeter wave systems. *IEEE Transactions on Communications*, 64(5):1801–1818, 2016.

[AL12] Dyonisius Dony Ariananda and Geert Leus. Compressive wideband power spectrum estimation. *IEEE Transactions on Signal Processing*, 60(9):4775–4789, 2012.

[ALH15] Ahmed Alkhateeb, Geert Leus, and Robert W Heath. Compressed sensing based multi-user millimeter wave systems: How many measurements are needed? In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2909–2913. IEEE, 2015.

[AMGPH14] Ahmed Alkhateeb, Jianhua Mo, Nuria Gonzalez-Prelcic, and Robert W Heath. Mimo precoding and combining solutions for millimeter-wave systems. *IEEE Communications Magazine*, 52(12):122–131, 2014.

[APP11] Ravi P Agarwal, Kanishka Perera, and Sandra Pinelas. *An introduction to complex analysis*. Springer Science & Business Media, 2011.

[ARR14a] A. Ahmed, B. Recht, and J. Romberg. Blind deconvolution using convex programming. *IEEE Transactions on Information Theory*, 60(3):1711–1732, March 2014.

[ARR14b] Arif Ahmed, Benjamin Recht, and Justin Romberg. Blind deconvolution using convex programming. *IEEE Transactions on Information Theory*, 60(3):1711–1732, 2014.

[ASZ10] S. Aeron, V. Saligrama, and M. Zhao. Information theoretic bounds for compressed sensing. *IEEE Transactions on Information Theory*, 56(10):5111–5130, Oct 2010.

[AT10] M. Akcakaya and V. Tarokh. Shannon-theoretic limits on noisy compressive sampling. *IEEE Transactions on Information Theory*, 56(1):492–504, Jan 2010.

[Bar07] Richard G Baraniuk. Compressive sensing. *IEEE signal processing magazine*, 24(4), 2007.

[BBS13] John Brady, Nader Behdad, and Akbar M Sayeed. Beamspace mimo for millimeter-wave communications: System architecture, modeling, analysis, and measurements. *IEEE Transactions on Antennas and Propagation*, 61(7):3814–3827, 2013.

[BCHJ14] J. D. Blanchard, M. Cermak, D. Hanle, and Y. Jing. Greedy algorithms for joint sparse recovery. *IEEE Transactions on Signal Processing*, 62(7):1694–1704, April 2014.

[BCMT10] Jerome Brachat, Pierre Comon, Bernard Mourrain, and Elias Tsigaridas. Symmetric tensor decomposition. *Linear Algebra and its Applications*, 433(11-12):1851–1872, 2010.

[BHE10] Zvika Ben-Haim and Yonina C Eldar. The cramér-rao bound for estimating a sparse parameter vector. *IEEE Transactions on Signal Processing*, 58(6):3384–3389, 2010.

[BKDM14] O. Balkan, K. Kreutz-Delgado, and S. Makeig. Localization of more sources than sensors via jointly-sparse bayesian learning. *IEEE Signal Processing Letters*, 21(2):131–134, Feb 2014.

[BKT09] Behtash Babadi, Nicholas Kalouptsidis, and Vahid Tarokh. Asymptotic achievability of the cramér–rao bound for noisy compressive sampling. *IEEE Transactions on Signal Processing*, 57(3):1233–1236, 2009.

[Bro97] Rasmus Bro. Parafac. tutorial and applications. *Chemometrics and intelligent laboratory systems*, 38(2):149–171, 1997.

[BVD+18] Martijn Boussé, Nico Vervliet, Ignat Domanov, Otto Debals, and Lieven De Lathauwer. Linear systems with a canonical polyadic decomposition constrained solution: Algorithms and applications. *Numerical Linear Algebra with Applications*, 25(6):e2190, 2018.

[C+06] Emmanuel J Candès et al. Compressive sampling. In *Proceedings of the international congress of mathematicians*, volume 3, pages 1433–1452. Madrid, Spain, 2006.

[CC15] Yuxin Chen and Emmanuel Candes. Solving random quadratic systems of equations is nearly as easy as solving linear systems. In *Advances in Neural Information Processing Systems*, pages 739–747, 2015.

[CFG14] Emmanuel J Candès and Carlos Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on pure and applied Mathematics*, 67(6):906–956, 2014.

[CH05] Jie Chen and Xiaoming Huo. Sparse representations for multiple measurement vectors (mmv) in an over-complete dictionary. In *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, volume 4, pages iv/257–iv/260 Vol. 4, March 2005.

[CH06] J. Chen and X. Huo. Theoretical results on sparse representations of multiple-measurement vectors. *IEEE Transactions on Signal Processing*, 54(12):4634–4643, Dec 2006.

[CLS15] Emmanuel J Candes, Xiaodong Li, and Mahdi Soltanolkotabi. Phase retrieval via wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory*, 61(4):1985–2007, 2015.

[CLYM91] Yih-Min Chen, Ju-Hong Lee, C-C Yeh, and Jeich Mar. Bearing estimation without calibration for randomly perturbed arrays. *IEEE Transactions on Signal Processing*, 39(1):194–197, 1991.

[Cor09] Thomas H Cormen. *Introduction to algorithms*. MIT press, 2009.

[CP09] Emmanuel J Candès and Yaniv Plan. Near-ideal model selection by ℓ_1 minimization. *The Annals of Statistics*, 37(5A):2145–2177, 2009.

[CREKD05] S. F. Cotter, B. D. Rao, Kjersti Engan, and K. Kreutz-Delgado. Sparse solutions to linear inverse problems with multiple measurement vectors. *IEEE Transactions on Signal Processing*, 53(7):2477–2488, July 2005.

[CRJ19] Sung-En Chiu, Nancy Ronquillo, and Tara Javidi. Active learning and csi acquisition for mmwave initial alignment. *IEEE Journal on Selected Areas in Communications*, 37(11):2474–2489, 2019.

[CRT06] E. J. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, Feb 2006.

[CSV13] Emmanuel J Candes, Thomas Strohmer, and Vladislav Voroninski. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274, 2013.

[CT05] E. J. Candes and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, Dec 2005.

[CT07] Emmanuel Candès and Terence Tao. The dantzig selector: Statistical estimation when p is much larger than n. *The Annals of Statistics*, pages 2313–2351, 2007.

[CW08] E. J. Candes and M. B. Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, March 2008.

[CY15] Jin Ho Choi and Chang D Yoo. Underdetermined high-resolution doa estimation: A 2ρth-order source-signal/noise subspace constrained optimization. *IEEE Transactions on Signal Processing*, 63(7):1858–1873, 2015.

[DDL13] Ignat Domanov and Lieven De Lathauwer. On the uniqueness of the canonical polyadic decomposition of third-order tensors—part i: Basic results and uniqueness of one factor matrix. *SIAM Journal on Matrix Analysis and Applications*, 34(3):855–875, 2013.

[DE12] M. E. Davies and Y. C. Eldar. Rank awareness in joint sparse recovery. *IEEE Transactions on Information Theory*, 58(2):1135–1146, Feb 2012.

[DET06] David L Donoho, Michael Elad, and Vladimir N Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on Information Theory*, 52(1):6–18, 2006.

[DL06] Lieven De Lathauwer. A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization. *SIAM journal on Matrix Analysis and Applications*, 28(3):642–666, 2006.

[DL14] Ignat Domanov and Lieven De Lathauwer. Canonical polyadic decomposition of third-order tensors: Reduction to generalized eigenvalue decomposition. *SIAM Journal on Matrix Analysis and Applications*, 35(2):636–660, 2014.

[DLCC07] Lieven De Lathauwer, Josphine Castaing, and Jean-Franois Cardoso. Fourth-order cumulant-based blind identification of underdetermined mixtures. *IEEE Transactions on Signal Processing*, 55(6):2965–2973, 2007.

[DM09] W. Dai and O. Milenkovic. Subspace pursuit for compressive sensing signal reconstruction. *IEEE Transactions on Information Theory*, 55(5):2230–2249, May 2009.

[Don06] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, April 2006.

[EARAS+14] Omar El Ayach, Sridhar Rajagopal, Shadi Abu-Surra, Zhouyue Pi, and Robert W Heath. Spatially sparse precoding in millimeter wave mimo systems. *IEEE transactions on wireless communications*, 13(3):1499–1513, 2014.

[EY36] Carl Eckart and Gale Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936.

[FM15] Matthew Fickus and Dustin G Mixon. Tables of the existence of equiangular tight frames. *arXiv preprint arXiv:1504.00253*, 2015.

[FR13] Simon Foucart and Holger Rauhut. *A mathematical introduction to compressive sensing*, volume 1. Birkhäuser Basel, 2013.

[FRG07] A. K. Fletcher, S. Rangan, and V. K. Goyal. On the rate-distortion performance of compressed sensing. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, volume 3, pages III–885–III–888, April 2007.

[FRG09] A. K. Fletcher, S. Rangan, and V. K. Goyal. Necessary and sufficient conditions for sparsity pattern recovery. *IEEE Transactions on Information Theory*, 55(12):5758–5772, Dec 2009.

[GBB+16] José Henrique de M Goulart, Maxime Boizard, Rémy Boyer, Gérard Favier, and Pierre Comon. Tensor cp decomposition with structured factor matrices: Algorithms and performance. *IEEE Journal of Selected Topics in Signal Processing*, 10(4):757–769, 2016.

[GHDW16] Zhen Gao, Chen Hu, Linglong Dai, and Zhaocheng Wang. Channel estimation for millimeter-wave massive mimo with hybrid precoding over frequency-selective fading channels. *IEEE Communications Letters*, 20(6):1259–1262, 2016.

[GR97] I. F. Gorodnitsky and B. D. Rao. Sparse signal reconstruction from limited data using focuss: a re-weighted minimum norm algorithm. *IEEE Transactions on Signal Processing*, 45(3):600–616, Mar 1997.

[GR09] Robert Clifford Gunning and Hugo Rossi. *Analytic functions of several complex variables*, volume 368. American Mathematical Soc., 2009.

[GWH17] Ziyu Guo, Xiaodong Wang, and Wei Heng. Millimeter-wave channel estimation based on 2-d beamspace music method. *IEEE Transactions on Wireless Communications*, 16(8):5384–5394, 2017.

[GZC18] M. Guo, Y. D. Zhang, and T. Chen. Doa estimation using compressed sparse array. *IEEE Transactions on Signal Processing*, 66(15):4133–4146, Aug 2018.

[HBCN12] Jarvis Haupt, Richard Baraniuk, Rui Castro, and Robert Nowak. Sequentially designed compressed sensing. In *2012 IEEE Statistical Signal Processing Workshop (SSP)*, pages 401–404. IEEE, 2012.

[HC16] Saeid Haghighatshoar and Giuseppe Caire. Massive mimo channel subspace estimation from low-dimensional projections. *IEEE Transactions on Signal Processing*, 65(2):303–318, 2016.

[HG07] A. Hjorungnes and D. Gesbert. Complex-valued matrix differentiation: Techniques and key results. *IEEE Transactions on Signal Processing*, 55(6):2740–2746, June 2007.

[HGPR+16] Robert W Heath, Nuria Gonzalez-Prelcic, Sundeep Rangan, Wonil Roh, and Akbar M Sayeed. An overview of signal processing techniques for millimeter wave mimo systems. *IEEE journal of selected topics in signal processing*, 10(3):436–453, 2016.

[HIR+14] Shuangfeng Han, Chih Lin I, C. Rowell, Z. Xu, Sen Wang, and Zhengang Pan. Large scale antenna system with hybrid digital and analog beamforming structure. In *2014 IEEE International Conference on Communications Workshops (ICC)*, pages 842–847, June 2014.

[HK+90a] Ralph T Hoctor, Saleem Kassam, et al. The unifying role of the coarray in aperture synthesis for coherent and incoherent imaging. *Proceedings of the IEEE*, 78(4):735–752, 1990.

[HK90b] Ralph T Hoctor and Saleem A Kassam. The unifying role of the coarray in aperture synthesis for coherent and incoherent imaging. *Proceedings of the IEEE*, 78(4):735–752, 1990.

[HKL+13] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh. Millimeter wave beamforming for wireless backhaul and access in small cell networks. *IEEE Transactions on Communications*, 61(10):4391–4403, October 2013.

[HL13] Christopher J Hillar and Lek-Heng Lim. Most tensor problems are np-hard. *Journal of the ACM (JACM)*, 60(6):45, 2013.

[HL16] Yonghee Han and Jungwoo Lee. Two-stage compressed sensing for millimeter wave channel estimation. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 860–864. IEEE, 2016.

[Hoe63] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

[HYN15] Keyong Han, Peng Yang, and Arye Nehorai. Calibrating nested sensor arrays with model errors. *IEEE Transactions on Antennas and Propagation*, 63(11):4739–4748, 2015.

[JGO99a] Magnus Jansson, Bo Goransson, and Bjorn Ottersten. A subspace method for direction of arrival estimation of uncorrelated emitter signals. *IEEE Transactions on Signal Processing*, 47(4):945–956, 1999.

[JGO99b] Magnus Jansson, Bo Goransson, and Bjorn Ottersten. A subspace method for direction of arrival estimation of uncorrelated emitter signals. *IEEE Transactions on Signal Processing*, 47(4):945–956, 1999.

[JKR11] Y. Jin, Y. H. Kim, and B. D. Rao. Limits on support recovery of sparse signals via multiple-access communication techniques. *IEEE Transactions on Information Theory*, 57(12):7877–7892, Dec 2011.

[JStB01] Tao Jiang, Nicholas D Sidiropoulos, and Jos MF ten Berge. Almost-sure identifiability of multidimensional harmonic retrieval. *IEEE Transactions on Signal Processing*, 49(9):1849–1859, 2001.

[JXC08] S. Ji, Y. Xue, and L. Carin. Bayesian compressive sensing. *IEEE Transactions on Signal Processing*, 56(6):2346–2356, June 2008.

[KB09] Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

[KD09] Ken Kreutz-Delgado. The complex gradient operator and the cr-calculus. *arXiv preprint arXiv:0906.4835*, 2009.

[KP] Ali Koochakzadeh and Piya Pal. Compressed arrays and hybrid channel sensing: Acram'er-rao bound based analysis. *Submitted to IEEE Signal Processing Letters*.

[KP15] Ali Koochakzadeh and Piya Pal. Sparse source localization in presence of co-array perturbations. *Proc. of 2015 Sampling Theory and Applications (SampTA)*, 2015.

[KP16a] Ali Koochakzadeh and Piya Pal. Cramér–rao bounds for underdetermined source localization. *IEEE Signal Processing Letters*, 23(7):919–923, 2016.

[KP16b] Ali Koochakzadeh and Piya Pal. Performance of uniform and sparse non-uniform samplers in presence of modeling errors: A cramér-rao bound based study. *IEEE Transactions on Signal Processing*, 65(6):1607–1621, 2016.

[KP17] Ali Koochakzadeh and Piya Pal. Sparse source localization using perturbed arrays via bi-affine modeling. *Digital Signal Processing*, 61:15–25, 2017.

[KP18a] Ali Koochakzadeh and Piya Pal. Beam-pattern design for hybrid beamforming using wirtinger flow. In *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 1–5. IEEE, 2018.

[KP18b] Ali Koochakzadeh and Piya Pal. A greedy approach for correlation-aware sparse support recovery. In *Compressive Sensing VII: From Diverse Modalities to Big Data Analytics*, volume 10658, page 106580C. International Society for Optics and Photonics, 2018.

[KP18c] Ali Koochakzadeh and Piya Pal. Mixed factor structured tensor decomposition via solving quadratic equations. In *Accepted in Asilomar Conference on Signals, Systems, and Computers*, 2018.

[KP18d] Ali Koochakzadeh and Piya Pal. Non-asymptotic guarantees for correlation-aware support detection. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3181–3185. IEEE, 2018.

[KP19a] A. Koochakzadeh and P. Pal. Canonical polyadic (cp) decomposition of structured semi-symmetric fourth-order tensors. In *2019 IEEE Data Science Workshop (DSW)*, pages 305–309, 2019.

[KP19b] Ali Koochakzadeh and Piya Pal. On flattening of symmetric tensors and identification of latent factors. In *2019 53rd Annual Conference on Information Sciences and Systems (CISS)*, pages 1–2. IEEE, 2019.

[KP20] Ali Koochakzadeh and Piya Pal. Channel estimation for hybrid mimo communication with (non-) uniform linear arrays via tensor decomposition. In *2020 IEEE 11th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pages 1–5. IEEE, 2020.

[KQP18] A. Koochakzadeh, H. Qiao, and P. Pal. On fundamental limits of joint sparse support recovery using certain correlation priors. *IEEE Transactions on Signal Processing*, pages 1–1, 2018.

[Kru76] Joseph B Kruskal. More factors than subjects, tests and treatments: an indeterminacy theorem for canonical decomposition and individual differences scaling. *Psychometrika*, 41(3):281–293, 1976.

[Kru77] Joseph B Kruskal. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear algebra and its applications*, 18(2):95–138, 1977.

[KSP18] Ali Koochakzadeh, Pulak Sarangi, and Piya Pal. Mixed factor structured tensor decomposition via solving quadratic equations. In *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, pages 647–651. IEEE, 2018.

[KSP19] Ali Koochakzadeh, Pulak Sarangi, and Piya Pal. A sequential approach for sparse support recovery using correlation priors. In *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, pages 586–590. IEEE, 2019.

[KV96] Hamid Krim and Mats Viberg. Two decades of llrray signal processing research. *IEEE signal processing magazine*, 1996.

[LBJ12] K. Lee, Y. Bresler, and M. Junge. Subspace methods for joint sparse recovery. *IEEE Transactions on Information Theory*, 58(6):3613–3641, June 2012.

[LGL14] Junho Lee, Gye-Tae Gil, and Yong H Lee. Exploiting spatial sparsity for estimating channels of hybrid mimo systems in millimeter wave communications. In *2014 IEEE Global Communications Conference*, pages 3326–3331. IEEE, 2014.

[LLB15] 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.

[LLM08] Jun Liu, Xiangqian Liu, and Xiaoli Ma. First-order perturbation analysis of singular vectors in singular value decomposition. *IEEE Transactions on Signal Processing*, 56(7):3044–3049, 2008.

[LLV93] Fu Li, Hui Liu, and Richard J Vaccaro. Performance analysis for doa estimation algorithms: unification, simplification, and observations. *IEEE Transactions on Aerospace and Electronic Systems*, 29(4):1170–1184, 1993.

[LS07] E. G. Larsson and Y. Selen. Linear regression with a sparse parameter vector. *IEEE Transactions on Signal Processing*, 55(2):451–460, Feb 2007.

[LS15a] Shuyang Ling and Thomas Strohmer. Self-calibration and biconvex compressive sensing. *arXiv preprint arXiv:1501.06864*, 2015.

[LS15b] Shuyang Ling and Thomas Strohmer. Self-calibration and biconvex compressive sensing. *arXiv preprint arXiv:1501.06864*, 2015.

[LV16] Chun-Lin Liu and PP Vaidyanathan. Super nested arrays: Linear sparse arrays with reduced mutual coupling—part i: Fundamentals. *IEEE Transactions on Signal Processing*, 64(15):3997–4012, 2016.

[LV17] Chun-Lin Liu and PP Vaidyanathan. Cramér–rao bounds for coprime and other sparse arrays, which find more sources than sensors. *Digital Signal Processing*, 61:43–61, 2017.

[LW11] Po-Ling Loh and Martin J Wainwright. High-dimensional regression with noisy and missing data: Provable guarantees with non-convexity. In *Advances in Neural Information Processing Systems*, pages 2726–2734, 2011.

[LWY+17] Nanxi Li, Zaixue Wei, Hongwen Yang, Xin Zhang, and Dacheng Yang. Hybrid precoding for mmwave massive mimo systems with partially connected structure. *IEEE Access*, 5:15142–15151, 2017.

[LZ13] 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. *IEEE Transactions on Signal Processing*, 61(15):3786–3798, 2013.

[MCW05] D. Malioutov, M. Cetin, and A. S. Willsky. A sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE Transactions on Signal Processing*, 53(8):3010–3022, Aug 2005.

[ME08] M. Mishali and Y. C. Eldar. Reduce and boost: Recovering arbitrary sets of jointly sparse vectors. *IEEE Transactions on Signal Processing*, 56(10):4692–4702, Oct 2008.

[MN14] Matthew L Malloy and Robert D Nowak. Near-optimal adaptive compressed sensing. *IEEE Transactions on Information Theory*, 60(7):4001–4012, 2014.

[Mof68] Alan Moffet. Minimum-redundancy linear arrays. *IEEE Transactions on antennas and propagation*, 16(2):172–175, 1968.

[MRH+17] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda. Hybrid beamforming for massive mimo: A survey. *IEEE Communications Magazine*, 55(9):134–141, 2017.

[NJS13] Praneeth Netrapalli, Prateek Jain, and Sujay Sanghavi. Phase retrieval using alternating minimization. In *Advances in Neural Information Processing Systems*, pages 2796–2804, 2013.

[NLJ+15] Yong Niu, Yong Li, Depeng Jin, Li Su, and Athanasios V Vasilakos. A survey of millimeter wave communications (mmwave) for 5g: opportunities and challenges. *Wireless networks*, 21(8):2657–2676, 2015.

[NLLNH17] Duy HN Nguyen, Long Bao Le, Tho Le-Ngoc, and Robert W Heath. Hybrid mmse precoding and combining designs for mmwave multiuser systems. *IEEE Access*, 5:19167–19181, 2017.

[NT09] Deanna Needell and Joel A Tropp. Cosamp: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321, 2009.

[NW08] Sahand Negahban and Martin J Wainwright. Joint support recovery under high-dimensional scaling: Benefits and perils of $\ell_{1,\infty}$-regularization. In *Proceedings of the 21st International Conference on Neural Information Processing Systems*, pages 1161–1168. Curran Associates Inc., 2008.

[NZL17] Song Noh, Michael D Zoltowski, and David J Love. Multi-resolution codebook and adaptive beamforming sequence design for millimeter wave beam alignment. *IEEE Transactions on Wireless Communications*, 16(9):5689–5701, 2017.

[PAGH19] S. Park, A. Ali, N. González-Prelcic, and R. W. Heath. Spatial channel covariance estimation for hybrid architectures based on tensor decompositions. *IEEE Transactions on Wireless Communications*, pages 1–1, 2019.

[PDM09] H. V. Pham, W. Dai, and O. Milenkovic. Sublinear compressive sensing reconstruction via belief propagation decoding. In *2009 IEEE International Symposium on Information Theory*, pages 674–678, June 2009.

[PH16] Sungwoo Park and Robert W Heath. Spatial channel covariance estimation for mmwave hybrid mimo architecture. In *Signals, Systems and Computers, 2016 50th Asilomar Conference on*, pages 1424–1428. IEEE, 2016.

[PH18] Sungwoo Park and Robert W Heath. Spatial channel covariance estimation for the hybrid mimo architecture: A compressive sensing-based approach. *IEEE Transactions on Wireless Communications*, 17(12):8047–8062, 2018.

[PK85] 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.

[PK11] Z. Pi and F. Khan. An introduction to millimeter-wave mobile broadband systems. *IEEE Communications Magazine*, 49(6):101–107, June 2011.

[PM13] Ranjitha Prasad and Chandra R Murthy. Cramér-rao-type bounds for sparse bayesian learning. *IEEE Transactions on Signal Processing*, 61(3):622–632, 2013.

[PPYH17] Sungwoo Park, Jeonghun Park, Ali Yazdan, and Robert W Heath. Exploiting spatial channel covariance for hybrid precoding in massive mimo systems. *IEEE Transactions on Signal Processing*, 65(14):3818–3832, 2017.

[PV10] P. Pal and P. P. Vaidyanathan. Nested arrays: A novel approach to array processing with enhanced degrees of freedom. *IEEE Transactions on Signal Processing*, 58(8):4167–4181, Aug 2010.

[PV11] P. Pal and P. P. Vaidyanathan. Coprime sampling and the music algorithm. In *2011 Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE)*, pages 289–294, Jan 2011.

[PV12a] P. Pal and P. P. Vaidyanathan. Correlation-aware techniques for sparse support recovery. In *2012 IEEE Statistical Signal Processing Workshop (SSP)*, pages 53–56, Aug 2012.

[PV12b] P. Pal and P. P. Vaidyanathan. On application of lasso for sparse support recovery with imperfect correlation awareness. In *2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, pages 958–962, Nov 2012.

[PV12c] Piya Pal and PP Vaidyanathan. Multiple level nested array: An efficient geometry for $2q$ th order cumulant based array processing. *IEEE Transactions on Signal Processing*, 60(3):1253–1269, 2012.

[PV13] P. Pal and P. P. Vaidyanathan. Correlation-aware sparse support recovery: Gaussian sources. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 5880–5884, May 2013.

[PV14a] P. Pal and P. P. Vaidyanathan. Parameter identifiability in sparse bayesian learning. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1851–1855, May 2014.

[PV14b] Piya Pal and PP Vaidyanathan. Pushing the limits of sparse support recovery using correlation information. *IEEE Transactions on Signal Processing*, 63(3):711–726, 2014.

[PV15] P. Pal and P. P. Vaidyanathan. Pushing the limits of sparse support recovery using correlation information. *IEEE Transactions on Signal Processing*, 63(3):711–726, Feb 2015.

[PYL17] S. Park, N. Y. Yu, and H. N. Lee. An information-theoretic study for joint sparsity pattern recovery with different sensing matrices. *IEEE Transactions on Information Theory*, 63(9):5559–5571, Sept 2017.

[QFS19] Cheng Qian, Xiao Fu, and Nikolaos D Sidiropoulos. Algebraic channel estimation algorithms for fdd massive mimo systems. *IEEE Journal of Selected Topics in Signal Processing*, 13(5):961–973, 2019.

[QFSY18] Cheng Qian, Xiao Fu, Nicholas D Sidiropoulos, and Ye Yang. Tensor-based channel estimation for dual-polarized massive mimo systems. *IEEE Transactions on Signal Processing*, 66(24):6390–6403, 2018.

[QP14] Heng Qiao and Piya Pal. Generalized nested sampling for compression and exact recovery of symmetric toeplitz matrices. In *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 443–447. IEEE, 2014.

[QP19] Heng Qiao and Piya Pal. Guaranteed localization of more sources than sensors with finite snapshots in multiple measurement vector models using difference co-arrays. *IEEE Transactions on Signal Processing*, 67(22):5715–5729, 2019.

[QZA15] Si Qin, Yimin D Zhang, and Moeness G Amin. Generalized coprime array configurations for direction-of-arrival estimation. *IEEE Transactions on Signal Processing*, 63(6):1377–1390, 2015.

[Rad11] K. Rahnama Rad. Nearly sharp sufficient conditions on exact sparsity pattern recovery. *IEEE Transactions on Information Theory*, 57(7):4672–4679, July 2011.

[RDL+20] Shiwei Ren, Wentao Dong, Xiangnan Li, Weijiang Wang, and Xiaoran Li. Extended nested arrays for consecutive virtual aperture enhancement. *IEEE Signal Processing Letters*, 27:575–579, 2020.

[REC04] B. D. Rao, K. Engan, and S. Cotter. Diversity measure minimization based method for computing sparse solutions to linear inverse problems with multiple measurement vectors. In *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages ii–369–72 vol.2, May 2004.

[Ren07] Joseph M Renes. Equiangular tight frames from paley tournaments. *Linear Algebra and its Applications*, 426(2-3):497–501, 2007.

[RG13] G. Reeves and M. C. Gastpar. Approximate sparsity pattern recovery: Information-theoretic lower bounds. *IEEE Transactions on Information Theory*, 59(6):3451–3465, June 2013.

[RK89] Richard Roy and Thomas Kailath. Esprit-estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on acoustics, speech, and signal processing*, 37(7):984–995, 1989.

[RMRGPH16] Cristian Rusu, Roi Mendez-Rial, Nuria González-Prelcic, and Robert W Heath. Low complexity hybrid precoding strategies for millimeter wave communication systems. *IEEE Transactions on Wireless Communications*, 15(12):8380–8393, 2016.

[RSH12] Florian Roemer, Carola Schroeter, and Martin Haardt. A semi-algebraic framework for approximate cp decompositions via joint matrix diagonalization and generalized unfoldings. In *2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, pages 2023–2027. IEEE, 2012.

[RSP+14] W. Roh, J. Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar. Millimeter-wave beamforming as an enabling technology for 5g cellular communications: theoretical feasibility and prototype results. *IEEE Communications Magazine*, 52(2):106–113, February 2014.

[RTC15] Xiang Ren, Meixia Tao, and Wen Chen. Compressed channel estimation with position-based ici elimination for high-mobility simo-ofdm systems. *IEEE Transactions on Vehicular Technology*, 65(8):6204–6216, 2015.

[SB10] Akbar Sayeed and Nader Behdad. Continuous aperture phased mimo: Basic theory and applications. In *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, pages 1196–1203. IEEE, 2010.

[SBG00] Nicholas D Sidiropoulos, Rasmus Bro, and Georgios B Giannakis. Parallel factor analysis in sensor array processing. *IEEE transactions on Signal Processing*, 48(8):2377–2388, 2000.

[SC17] J. Scarlett and V. Cevher. Limits on support recovery with probabilistic models: An information-theoretic framework. *IEEE Transactions on Information Theory*, 63(1):593–620, Jan 2017.

[SCG87] LPHK Seymour, CFN Cowan, and PM Grant. Bearing estimation in the presence of sensor positioning errors. In *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 12, pages 2264–2267, 1987.

[SCL+15a] J. Song, J. Choi, S. G. Larew, D. J. Love, T. A. Thomas, and A. A. Ghosh. Adaptive millimeter wave beam alignment for dual-polarized mimo systems. *IEEE Transactions on Wireless Communications*, 14(11):6283–6296, Nov 2015.

[SCL15b] J. Song, J. Choi, and D. J. Love. Codebook design for hybrid beamforming in millimeter wave systems. In *2015 IEEE International Conference on Communications (ICC)*, pages 1298–1303, June 2015.

[SCL17] Jiho Song, Junil Choi, and David J Love. Common codebook millimeter wave beam design: Designing beams for both sounding and communication with

uniform planar arrays. *IEEE Transactions on Communications*, 65(4):1859–1872, 2017.

[SDL13] Mikael Sørensen and Lieven De Lathauwer. Blind signal separation via tensor decomposition with vandermonde factor: Canonical polyadic decomposition. *IEEE Transactions on Signal Processing*, 61(22):5507–5519, 2013.

[SDL16a] Mikael Sørensen and Lieven De Lathauwer. Multidimensional harmonic retrieval via coupled canonical polyadic decomposition—part i: Model and identifiability. *IEEE Transactions on Signal Processing*, 65(2):517–527, 2016.

[SDL16b] Mikael Sørensen and Lieven De Lathauwer. Multidimensional harmonic retrieval via coupled canonical polyadic decomposition—part ii: Algorithm and multirate sampling. *IEEE Transactions on Signal Processing*, 65(2):528–539, 2016.

[SDL17a] Mikael Sørensen and Lieven De Lathauwer. Multidimensional harmonic retrieval via coupled canonical polyadic decomposition—part i: Model and identifiability. *IEEE Transactions on Signal Processing*, 65(2):517–527, 2017.

[SDL17b] Mikael Sørensen and Lieven De Lathauwer. Multidimensional harmonic retrieval via coupled canonical polyadic decomposition—part ii: Algorithm and multirate sampling. *IEEE Transactions on Signal Processing*, 65(2):528–539, 2017.

[SDLF+17] Nicholas D Sidiropoulos, Lieven De Lathauwer, Xiao Fu, Kejun Huang, Evangelos E Papalexakis, and Christos Faloutsos. Tensor decomposition for signal processing and machine learning. *IEEE Transactions on Signal Processing*, 65(13):3551–3582, 2017.

[SLG01] Petre Stoica, Erik G Larsson, and Alex B Gershman. The stochastic crb for array processing: A textbook derivation. *IEEE Signal Processing Letters*, 8(5):148–150, 2001.

[SM50] Jack Sherman and Winifred J Morrison. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *The Annals of Mathematical Statistics*, 21(1):124–127, 1950.

[SM93] Louis L Scharf and LT McWhorter. Geometry of the cramér–rao bound. *Signal Processing*, 31(3):301–311, 1993.

[SM01a] Petre Stoica and Thomas L Marzetta. Parameter estimation problems with singular information matrices. *IEEE Transactions on Signal Processing*, 49(1):87–90, 2001.

[SM01b] Petre Stoica and Thomas L Marzetta. Parameter estimation problems with singular information matrices. *IEEE Transactions on Signal Processing*, 49(1):87–90, 2001.

[SM05] Petre Stoica and Randolph L Moses. *Spectral analysis of signals*, volume 452. Pearson Prentice Hall Upper Saddle River, NJ, 2005.

[SN89] Petre Stoica and Arye Nehorai. Music, maximum likelihood, and cramer-rao bound. *IEEE Transactions on Acoustics, speech, and signal processing*, 37(5):720–741, 1989.

[SN90a] Petre Stoica and Arye Nehorai. Music, maximum likelihood, and cramer-rao bound: further results and comparisons. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(12):2140–2150, 1990.

[SN90b] Petre Stoica and Arye Nehorai. Performance study of conditional and unconditional direction-of-arrival estimation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(10):1783–1795, 1990.

[SN90c] Petre Stoica and Arye Nehorai. Performance study of conditional and unconditional direction-of-arrival estimation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(10):1783–1795, 1990.

[SPF14] Nikos D Sidiropoulos, Evangelos E Papalexakis, and Christos Faloutsos. A parallel algorithm for big tensor decomposition using randomly compressed cubes (paracomp). In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2014.

[SR17] Shu Sun and Theodore S Rappaport. Millimeter wave mimo channel estimation based on adaptive compressed sensing. In *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*, pages 47–53. IEEE, 2017.

[SS94] P Stoica and T Soderstrom. Parameter identifiability problem in signal processing. *IEE Proceedings-Radar, Sonar and Navigation*, 141(3):133–136, 1994.

[ST06] Suvrit Sra and Joel A Tropp. Row-action methods for compressed sensing. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, volume 3, pages III–III. IEEE, 2006.

[SY16] F. Sohrabi and W. Yu. Hybrid digital and analog beamforming design for large-scale antenna arrays. *IEEE Journal of Selected Topics in Signal Processing*, 10(3):501–513, April 2016.

[TBSR13] Gongguo Tang, Badri Narayan Bhaskar, Parikshit Shah, and Benjamin Recht. Compressed sensing off the grid. *IEEE transactions on information theory*, 59(11):7465–7490, 2013.

[TF+03] Michael E Tipping, Anita C Faul, et al. Fast marginal likelihood maximisation for sparse bayesian models. In *AISTATS*, 2003.

[TG07a] Z. Tian and G. B. Giannakis. Compressed sensing for wideband cognitive radios. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, volume 4, pages IV–1357–IV–1360, April 2007.

[TG07b] J. A. Tropp and A. C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 53(12):4655–4666, Dec 2007.

[TGS06] Joel A Tropp, Anna C Gilbert, and Martin J Strauss. Algorithms for simultaneous sparse approximation. part i: Greedy pursuit. *Signal Processing*, 86(3):572–588, 2006.

[Tib96] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[Tip01] Michael E Tipping. Sparse bayesian learning and the relevance vector machine. *Journal of machine learning research*, 1(Jun):211–244, 2001.

[TN10] G. Tang and A. Nehorai. Performance analysis for sparse support recovery. *IEEE Transactions on Information Theory*, 56(3):1383–1399, March 2010.

[TTT99] Kim-Chuan Toh, Michael J Todd, and Reha H Tütüncü. Sdpt3—a matlab software package for semidefinite programming, version 1.3. *Optimization methods and software*, 11(1-4):545–581, 1999.

[Vai07] P. P. Vaidyanathan. The theory of linear prediction. *Synthesis Lectures On Signal Processing*, 2(1):1–184, 2007.

[VDDL16] Nico Vervliet, Otto Debals, and Lieven De Lathauwer. Tensorlab 3.0—numerical optimization strategies for large-scale constrained and coupled matrix/tensor factorization. In *2016 50th Asilomar Conference on Signals, Systems and Computers*, pages 1733–1738. IEEE, 2016.

[VDS+16] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, and L. De Lathauwer. Tensorlab 3.0, Mar. 2016. Available online.

[VP11] P. P. Vaidyanathan and P. Pal. Sparse sensing with co-prime samplers and arrays. *IEEE Transactions on Signal Processing*, 59(2):573–586, Feb 2011.

[VP12] PP Vaidyanathan and Piya Pal. Direct-music on sparse arrays. In *2012 International Conference on Signal Processing and Communications (SPCOM)*, pages 1–5. IEEE, 2012.

[VS94] Mats Viberg and A Lee Swindlehurst. A bayesian approach to auto-calibration for parametric array signal processing. *IEEE Transactions on Signal Processing*, 42(12):3495–3507, 1994.

[VT04] Harry L Van Trees. *Detection, Estimation, and Modulation Theory, Optimum Array Processing*. John Wiley & Sons, 2004.

[Wai09a] M. J. Wainwright. Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting. *IEEE Transactions on Information Theory*, 55(12):5728–5741, Dec 2009.

[Wai09b] M. J. Wainwright. Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1 -constrained quadratic programming (lasso). *IEEE Transactions on Information Theory*, 55(5):2183–2202, May 2009.

[Wed73] Per-Åke Wedin. Perturbation theory for pseudo-inverses. *BIT Numerical Mathematics*, 13(2):217–232, 1973.

[Wen62] James G Wendel. A problem in geometric probability. *Math. Scand*, 11:109–111, 1962.

[WF89] Anthony J Weiss and Benjamin Friedlander. Array shape calibration using sources in unknown locations-a maximum likelihood approach. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(12):1958–1966, 1989.

[WLP09] Ying Wang, Geert Leus, and Ashish Pandharipande. Direction estimation using compressive sampling array processing. In *Statistical Signal Processing, 2009. SSP'09. IEEE/SP 15th Workshop on*, pages 626–629. IEEE, 2009.

[Woo50] Max A Woodbury. Inverting modified matrices. *Memorandum report*, 42:106, 1950.

[Wor94] Thomas Worsch. *Lower and upper bounds for (sums of) binomial coefficients*. Univ. Karlsruhe, Fakultät für Informatik, 1994.

[WR04] D. P. Wipf and B. D. Rao. Sparse bayesian learning for basis selection. *IEEE Transactions on Signal Processing*, 52(8):2153–2164, Aug 2004.

[WR07] David P Wipf and Bhaskar D Rao. An empirical bayesian strategy for solving the simultaneous sparse approximation problem. *IEEE Transactions on Signal Processing*, 55(7):3704–3716, 2007.

[XGJ16] Hongxiang Xie, Feifei Gao, and Shi Jin. An overview of low-rank channel estimation for massive mimo systems. *IEEE Access*, 4:7313–7321, 2016.

[XHYB12] D. Xu, N. Hu, Z. Ye, and M. Bao. The estimate for doas of signals using sparse recovery method. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2573–2576, March 2012.

[XSRK94] Guanghan Xu, Seth D Silverstein, Richard H Roy, and Thomas Kailath. Beamspace esprit. *IEEE Transactions on Signal Processing*, 42(2):349–356, 1994.

[YSYC16] Minglei Yang, Lei Sun, Xin Yuan, and Baixiao Chen. Improved nested array with hole-free dca and more degrees of freedom. *Electronics Letters*, 52(25):2068–2070, 2016.

[ZAH13] Yimin D Zhang, Moeness G Amin, and Braham Himed. Sparsity-based doa estimation using co-prime arrays. In *Proc. of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),*, pages 3967–3971, 2013.

[Zaj18] Krzysztof Zajkowski. Bounds on tail probabilities for quadratic forms in dependent sub-gaussian random variables. *arXiv preprint arXiv:1809.08569*, 2018.

[Zau11] Gerhard Zauner. *Grundzüge einer nichtkommutativen Designtheorie*. PhD thesis, PhD thesis, University of Vienna, 1999. Published in English translation: Zauner, G. Quantum designs: foundations of a noncommutative design theory. Int. J. Quantum Inf. 9, 2011.

[ZFY+17] Zhou Zhou, Jun Fang, Linxiao Yang, Hongbin Li, Zhi Chen, and Rick S Blum. Low-rank tensor decomposition-aided channel estimation for millimeter wave mimo-ofdm systems. *IEEE Journal on Selected Areas in Communications*, 35(7):1524–1538, 2017.

[ZGZ+17] Chengwei Zhou, Yujie Gu, Yimin D Zhang, Zhiguo Shi, Tao Jin, and Xidong Wu. Compressive sensing-based coprime array direction-of-arrival estimation. *IET Communications*, 11(11):1719–1724, 2017.

[ZH17] Jianshu Zhang and Martin Haardt. Channel estimation for hybrid multi-carrier mmwave mimo systems using three-dimensional unitary esprit in dft beamspace. In *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 1–5. IEEE, 2017.

[ZPHN16] Jianshu Zhang, Ivan Podkurkov, Martin Haardt, and Adel Nadeev. Channel estimation and training design for hybrid analog-digital multi-carrier single-user massive mimo systems. In *WSA 2016; 20th International ITG Workshop on Smart Antennas*, pages 1–8. VDE, 2016.

[ZQA14] Yimin D Zhang, Si Qin, and Moeness G Amin. Doa estimation exploiting coprime arrays with sparse sensor spacing. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2267–2271. IEEE, 2014.

[ZR11] Zhilin Zhang and Bhaskar D Rao. Sparse signal recovery with temporally correlated source vectors using sparse bayesian learning. *IEEE Journal of Selected Topics in Signal Processing*, 5(5):912–926, 2011.

[ZWKZ19] Zhi Zheng, Wen-Qin Wang, Yangyang Kong, and Yimin D Zhang. Misc array: A new sparse array design achieving increased degrees of freedom and reduced

mutual coupling effect. *IEEE Transactions on Signal Processing*, 67(7):1728–1741, 2019.