UNIVERSITY OF CALIFORNIA SAN DIEGO

Structured Sub-Nyquist Sampling with Applications in Compressive Toeplitz Covariance Estimation, Super-Resolution and Phase Retrieval

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

in

Electrical Engineering
(Signal and Image Processing)

by

Heng Qiao

Committee in charge:

Professor Piya Pal, Chair
Professor Ery Arias-Castro
Professor Massimo Franceschetti
Professor Bhaskar D. Rao
Professor Rayan Saab

2019

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

_____

_____

_____

_____

_____

Chair

University of California San Diego

2019

DEDICATION

To my family.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I am greatly indebted to my advisor, Professor Piya Pal, for her selfless support and guidance during my Ph.D. life. I deeply thank her for leading me into this profession and help me discover the beauty of rigorous scientific research. Without her support, I might have left academia years back. Besides guiding me to write papers, Prof. Pal cultivated my scientific ethics and set a high standard for my future academic career. I learned from her that the fundamental goal of research is more about understanding rather than just accumulating papers. And the best papers are those that elegantly explain the universal principles of core problems in the fields.

I would like to thank Prof. Bhaskar Rao, Prof. Massimo Franceschetti, Prof. Ery Arias-Castro, and Prof. Rayan Saab for serving on my committee. I have learned a lot from their lectures, papers and feedback during my PhD progress. I benefit a lot from the outstanding academic environment of UCSD.

I would like to express my thanks to my lab mates Ali KoochakZadeh, Pulak Sarangi, Mehmet Can Hucumenoglu and Sina Shahsavari. I have gained many helpful suggestions from the discussion and collaborations with them.

Finally, I want to thank my wife and parents for their unconditional support in all senses and understanding of my career choice. This dissertation is dedicated to them.

Chapter 2, in part, is a reprint of the material as it appears in following papers:

- H. Qiao and P. Pal, "Gridless Line Spectrum Estimation and Low-Rank Toeplitz Matrix Compression Using Structured Samplers: A Regularization-Free Approach", *IEEE Transactions on Signal Processing*, vol. 65, no. 9, pp. 2211-2226, May 2017.

- H. Qiao and P. Pal, "On Maximum Likelihood Methods For Localizing More Sources than Sensors", *IEEE Signal Processing Letters*, vol. 24, no. 5, pp. 703-706, 2017.

- H. Qiao and P. Pal, "Generalized Nested Sampling for Compressing Low Rank Toeplitz Matrices", *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 1844-1848, Nov. 2015.

- H. Qiao and P. Pal, "Stable Compressive Low Rank Toeplitz Covariance Estimation Without Regularization", in the *Proceedings of Asilomar Conference on Signals, Systems and Computers*, 2016.

- H. Qiao and P. Pal, "Generalized Nested Sampling for Compression and Exact Recovery of Symmetric Toeplitz Matrix," in the *Proceedings of IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Atlanta, GA, USA, 2014.

The dissertation author was the primary investigator and author of these papers.

Chapter 3, in part, is a reprint of the material as it appears in the papers:

- H. Qiao and P. 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*, vol. 67, no. 22, pp. 5715-5729, Nov. 2019.

- H. Qiao and P. Pal, "A Non-Convex Approach to Non-Negative Super-Resolution: Theory and Algorithm", in the *Proceedings of 44th International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, May 2019.

- H. Qiao and P. Pal, " On modulus of continuity for noisy positive super-resolution", in the *Proceedings of 43nd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2018)*, Calgary, Canada.

- H. Qiao and P. Pal, "Performance Limits of Covariance-Driven Super Resolution Imaging", in the *Proceedings of Asilomar Conference on Signals, Systems and Computers*, 2017.

- H. Qiao and P. Pal,"Unified Analysis of Co-Array Interpolation for Direction-of-Arrival Estimation," in the *Proceedings of 42nd International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, March 2017.

The dissertation author was the primary investigator and author of these papers.

Chapter 4, in part, is a reprint of the material as it appears in the papers:

- H. Qiao and P. Pal, " Sparse Phase Retrieval with Near Minimal Measurements: A Structured Sampling Based Approach", in the *Proceedings of 41st IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, 2016.

- H. Qiao and P. Pal, "Sparse Phase Retrieval Using Partial Nested Fourier Samplers," in the *Proceedings of IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Orlando, FL, USA, 2015.

The dissertation author was the primary investigator and author of these papers.

Chapter 5, in part, is a reprint of the material as it appears in the papers:

- H. Qiao, M. C. Hucumenoglu and P. Pal, "Compressive Kriging Using Multi-Dimensional Generalized Nested Sampling", in the *Proceedings of Asilomar Conference on Signals, Systems and Computers*, 2018.

- H. Qiao and P. Pal, "Understanding the Role of Positive Constraints in Sparse Bilinear Problems", in the *Proceedings of 7th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Dec. 2017.

The dissertation author was the primary investigator and author of these papers.

VITA

2012	Bachelor of Engineering, Tsinghua University, China
2016	Master of Science, University of Maryland, College Park
2019	Doctor of Philosophy, University of California San Diego

PUBLICATIONS

H. Qiao and P. 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*, vol. 67, no. 22, pp. 5715-5729, Nov. 2019.

Ali KoochakZadeh, H. Qiao and P. Pal, "On Fundamental Limits of Joint Sparse Support Recovery Using Certain Correlation Priors, *IEEE Transactions on Signal Processing*, vol. 66, no. 17, pp. 4612-4625, Sep. 2018.

H. Qiao and P. Pal, "Gridless Line Spectrum Estimation and Low-Rank Toeplitz Matrix Compression Using Structured Samplers: A Regularization-Free Approach, *IEEE Transactions on Signal Processing*, vol. 65, no. 9, pp. 2211-2226, May 2017.

H. Qiao and P. Pal, "On Maximum Likelihood Methods For Localizing More Sources than Sensors, *IEEE Signal Processing Letters*, vol. 24, no. 5, pp. 703-706, May 2017.

H. Qiao and P. Pal, "Generalized Nested Sampling for Compressing Low Rank Toeplitz Matrices, *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 1844-1848, Nov. 2015.

H. Qiao and P. Pal, "A Non-Convex Approach to Non-Negative Super-Resolution: Theory and Algorithm", in the *Proceedings of 44th International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, May 2019.

H. Qiao, M. C. Hucumenoglu and P. Pal, "Compressive Kriging Using Multi-Dimensional Generalized Nested Sampling, in the *Proceedings of Asilomar Conference on Signals, Systems and Computers*, 2018.

H. Qiao and P. Pal, " On modulus of continuity for noisy positive super-resolution, in the *Proceedings of 43nd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2018)*, Calgary, Canada.

H. Qiao and P. Pal, "Understanding the Role of Positive Constraints in Sparse Bilinear Problems", in the *Proceedings of 7th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Dec. 2017. **[Nominated for Best Student Paper Award Competition]**

P. Sarangi, H. Qiao and P. Pal, On the Role of Sampling and Sparsity in Phase Retrieval for Optical Coherence Tomography, in the *Proceedings of 7th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Dec. 2017.

H. Qiao and P. Pal, "Performance Limits of Covariance-Driven Super Resolution Imaging", in the *Proceedings of Asilomar Conference on Signals, Systems and Computers*, 2017.

H. Qiao and P. Pal, "Multiple Hypothesis Testing for Dynamic Support Recovery," in the *18th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2017.

H. Qiao and P. Pal,"Unified Analysis of Co-Array Interpolation for Direction-of-Arrival Estimation, in the *Proceedings of 42nd International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, March 2017. **[Best Student Paper Award, First Place]**

H. Qiao and P. Pal, "Stable Compressive Low Rank Toeplitz Covariance Estimation Without Regularization, in the *Proceedings of Asilomar Conference on Signals, Systems and Computers*, 2016.

H. Qiao and P. Pal, "Finite Sample Analysis Covariance Compression Using Structured Samplers, in the *Proceedings 9th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Rio de Janeiro, Brazil, 2016.

H. Qiao and P. Pal, " Sparse Phase Retrieval with Near Minimal Measurements: A Structured Sampling Based Approach", in the *Proceedings of 41st IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, 2016.

H. Qiao and P. Pal, "Sparse Phase Retrieval Using Partial Nested Fourier Samplers, in the *Proceedings of IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Orlando, FL, USA, 2015.

H. Qiao and P. Pal, "Generalized Nested Sampling for Compression and Exact Recovery of Symmetric Toeplitz Matrix, in the *Proceedings of IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Atlanta, GA, USA, 2014.

ABSTRACT OF THE DISSERTATION

Structured Sub-Nyquist Sampling with Applications in Compressive Toeplitz Covariance Estimation, Super-Resolution and Phase Retrieval

by

Heng Qiao

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

University of California San Diego, 2019

Professor Piya Pal, Chair

Sub-Nyquist sampling has received a huge amount of interest in the past decade. In classical compressed sensing theory, if the measurement procedure satisfies a particular condition known as Restricted Isometry Property (RIP), we can achieve stable recovery of signals of low-dimensional intrinsic structures with an order-wise optimal sample size. Such low-dimensional structures include sparse and low rank for both vector and matrix cases. The main drawback of conventional compressed sensing theory is that random measurements are required to ensure the RIP property. However, in many applications such as imaging and array signal processing,

applying independent random measurements may not be practical as the systems are deterministic. Moreover, random measurements based compressed sensing always exploits convex programs for signal recovery even in the noiseless case, and solving those programs is computationally intensive if the ambient dimension is large, especially in the matrix case.

The main contribution of this dissertation is that we propose a deterministic sub-Nyquist sampling framework for compressing the structured signal and come up with computationally efficient algorithms. Besides widely studied sparse and low-rank structures, we particularly focus on the cases that the signals of interest are stationary or the measurements are of Fourier type. The key difference between our work from classical compressed sensing theory is that we explicitly exploit the second-order statistics of the signals, and study the equivalent quadratic measurement model in the correlation domain. The essential observation made in this dissertation is that a difference/sum coarray structure will arise from the quadratic model if the measurements are of Fourier type. With these observations, we are able to achieve a better compression rate for covariance estimation, identify more sources in array signal processing or recover the signals of larger sparsity.

In this dissertation, we will first study the problem of Toeplitz covariance estimation. In particular, we will show how to achieve an order-wise optimal compression rate using the idea of sparse arrays in both general and low-rank cases. Then, an analysis framework of super-resolution with positivity constraint is established. We will present fundamental robustness guarantees, efficient algorithms and applications in practices. Next, we will study the problem of phase-retrieval for which we successfully apply the sparse array ideas by fully exploiting the quadratic measurement model. We achieve near-optimal sample complexity for both sparse and general cases with practical Fourier measurements and provide efficient and deterministic recovery algorithms. In the end, we will further elaborate on the essential role of non-negative constraint in underdetermined inverse problems. In particular, we will analyze the nonlinear co-array interpolation problem and develop a universal upper bound of the interpolation error. Bilinear

problem with non-negative constraint will be considered next and the exact characterization of the ambiguous solutions will be established for the first time in literature. At last, we will show how to apply the nested array idea to solve real problems such as Kriging. Using spatial correlation information, we are able to have a stable estimate of the field of interest with fewer sensors than classic methodologies. Extensive numerical experiments are implemented to demonstrate our theoretical claims.

Chapter 1

Introduction

1.1 Background and Motivations

Sub-Nyquist sampling has been intensely studied in the past decades along with the emergence of ”Big Data” [20, 21]. The research interest comes from the fact that there is a widening gap between the volume of available data and the resources for storing, communicating and processing. Big data arises in many applications ranging from sensor networks, computer vision, artificial intelligence, surveillance system, genomics, web search, video streaming, social networks, to name a few. However, one common feature of these applications is that the underlying information of interest has a structure of much lower dimension compared to the ambient size of the raw data. Two typical low-dimensional structures considered in literature are sparse and low-rank. For example, a medical imaging device will collect a sequence of two-dimensional images which are corrupted by the blurring kernel and/or additive noise. The ultimate goal is to recover the locations of particular molecules rather than the time-variant signal intensities. This key observation provides us the opportunity to design efficient sampling strategies with much fewer samples by exploiting certain prior knowledge of the sources.

The classical compressed sensing theory was motivated by the prior work on decomposing

a signal in a sparse way with respect to a particular over-complete dictionary, and considered an opposite problem that how to design a measurement matrix to compress a sparse signal without information loss [22–26]. The key analysis tools are variants of concentration bounds which will ensure that the measurement process will satisfy a property known as Restricted Isometry Property (RIP) with high probability. Similar ideas have been later extended to matrix completion [51] and sparse/low rank matrix compression [28, 29, 89]. Because of the dependence on sophisticated probabilistic tools, the measurement procedures in all these work are based on random sampling, which may not be practical in many applications. Instead, in this dissertation, we will mainly study underdetermined inverse problems with deterministic sampling strategies. One of the advantages of deterministic sampling is that it allows us to exploit the statistical prior knowledge of the signals, which cannot be efficiently utilized by random sampling. As demonstrated in literature, certain statistical structures are known a priori. For example, in array signal processing, the income signals from different targets are statistically uncorrelated. Using this kind of prior knowledge, we are able to derive theoretical guarantees of stable sparse recovery with fewer measurements or sensors. Moreover, efficient algorithms can be designed.

As a high-level summary, our work in this dissertation differs from previous literature in following ways:

- We avoid the use of random sampling by explicitly exploiting the statistical properties of the signals and certain quadratic structures of the measurement models.
- Using the statistical structures of the signal, we design deterministic samplers based on the idea of sparse arrays, with which we can achieve order-wise optimal sample complexity and come up with efficient algorithms.
- Our analysis will be applicable to a family of recovery algorithms including both convex and non-convex programs. The performance guarantees are also universal and derived from the fundamental structures of the signals and models.

- We highlight the role of non-negative constraints (positivity/positive semidefinite) in sparse recovery. These constraints are not subspace-based as widely used sparse and low-rank constraints. We show that non-negative constraints are essential in robustness analysis and allows for relaxed conditions and simpler algorithms.

1.2 Compressive Toeplitz Covariance Sketching and Quadratic Samplers

In classical compressed sensing theory [96], the compressive measurement model can be written as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$$

where $\mathbf{y} \in \mathbb{C}^M$ represents low dimensional linear measurement of a high dimensional sparse vector $\mathbf{x} \in \mathbb{C}^N$ ($N \gg M$) using the measurement matrix $\mathbf{A} \in \mathbb{C}^{M,N}$. The number of non-zero elements of $\mathbf{x}$, denoted by $\|\mathbf{x}\|_0 = s$ is typically small, i.e. $s \ll N$. Given $\mathbf{y}$, $\mathbf{x}$ is typically reconstructed using the following l_1 minimization:

$$\begin{aligned} &\min_{\mathbf{z}} \|\mathbf{z}\|_1 \\ &s.t. \quad \mathbf{y} = \mathbf{A}\mathbf{x} \end{aligned} \tag{1.1}$$

By invoking certain isometric properties of high dimensional random linear operators [24], the original high dimensional signal can be successfully recovered from its low-dimensional measurement using l_1 minimization [22, 23]. In particular, for a wide class of random $\mathbf{A}$ with i.i.d entries, it can be shown that $M = O(s \log(N/s))$ measurements suffice for perfectly reconstructing $\mathbf{x}$ with overwhelming probability (that grows to 1 exponentially with N).

In many applications, however, our goal is to infer certain statistics of interest from high dimensional signals (such as stock price series over several months). In such cases, it may not be necessary to reconstruct the original signal and compression can be obtained without requiring the signal to be sparse. Furthermore, the physics of the problem can impose structures on the ensuing acquisition system, leading to the possibility of "structured sampling" strategies. Also often, one can make informed assumptions about the nature of randomness, or statistical distribution of the data (which is frequently done in statistical signal processing) that can be judiciously exploited by the sampling technique. Standard compressive sensing techniques, that heavily rely on sparsity of representation, and use linear random projections for taking measurements, may turn out to be either inapplicable, or sub-optimal in such settings. We will illustrate this in the context of compressive covariance sketching, where the goal is to infer the covariance matrix parameterizing the distribution of high dimensional signals, from their compressed sketch.

In many signal and information processing tasks, (such as spectral estimation and source localization), the covariance matrix $\mathbf{R}_{\mathbf{x}} = \mathbb{E}(\mathbf{x}\mathbf{x}^H)$ of the (zero-mean) high dimensional random signal $\mathbf{x}$ is used for subsequent estimation/detection tasks. However, owing to its large size, it may be impractical to store and/or communicate $\mathbf{R}_{\mathbf{x}}$ (or its estimate). Instead, if we acquire compressive linear measurements of $\mathbf{x}$ as $\mathbf{y} = \mathbf{A}\mathbf{x}$, the covariance matrix $\mathbf{R}_{\mathbf{y}}$ of $\mathbf{y}$ *now acts as a compressive sketch of* $\mathbf{R}_{\mathbf{x}}$ which can be effectively stored and/or processed. The high dimensional covariance matrix $\mathbf{R}_{\mathbf{x}}$ and its compressive sketch $\mathbf{R}_{\mathbf{y}}$ are related as

$$\mathbf{R}_{\mathbf{y}} = \mathbf{A}\mathbf{R}_{\mathbf{x}}\mathbf{A}^H \tag{1.2}$$

Notice that $\mathbf{R}_{\mathbf{y}}$ and $\mathbf{R}_{\mathbf{x}}$ are still linearly related, and in the most general setting, this linear map is equal to the Kronecker product $\mathbf{A}^* \otimes \mathbf{A}$. This is seen more clearly using the following vectorized

form:

$$\text{vec}(\mathbf{R_y}) = \Big(\mathbf{A}^* \otimes \mathbf{A}\Big)\text{vec}(\mathbf{R_x})$$

Hence, each element of $\mathbf{R_y}$ is a *quadratic function of the elements of* $\mathbf{A}$. The key idea in compressive covariance sensing is to design the linear operator $\mathbf{A}$ such that its aforementioned quadratic form possess certain desirable properties which can be exploited to *reconstruct* $\mathbf{R_x}$ *from an optimal number of measurements*. It is to be noted that Kronecker products of measurement matrices have been studied and analyzed for compressed sensing and sketching of images and other matrices [27, 28]. More recently, the performance of nuclear norm based compressive covariance estimation algorithms has been studied using random $\mathbf{A}$ with i.i.d entries. However, when the covariance matrix is highly structured, a direct application of these results will produce sub-optimal number of measurements. In other words, by carefully exploiting the specific structure of $\mathbf{R_x}$ (such as its positive semidefinite property), it may be possible to achieve a greater degree of compression via clever design of *structured deterministic* $\mathbf{A}$. In this thesis, we will assume $\mathbf{R_x}$ to be a Toeplitz structured covariance matrix, and derive an *optimal structured sampling strategy* (inspired from prior work on nested arrays [45]) that can provably perform exact and stable reconstruction of $\mathbf{R_x}$ from its compressed sketch, acquired using an optimal number of measurements, which is *only a function of the rank of* $\mathbf{R_x}$.

1.3 Super Resolution and Support Recovery with Non Negative Constraint

The common goal of super-resolution and support recovery is to estimate the locations of the non-zero entries from a few noisy low-passed measurements. In previous literature, the structures exploited to solve this underdetermined system are sparse [139, 141, 143, 145] and

non-negative constraint [61, 148]. In particular, the measurement model in both continuous and discrete cases can be expressed in a unified way as

$$\mathbf{y} = \mathbf{Q}\mathbf{x} + \mathbf{n} \tag{1.3}$$

where $\mathbf{n}$ represents the additive noise and $\mathbf{Q}, \mathbf{x}$ have different meanings according to the context

- **Continuous Case:** $\mathbf{Q}\mathbf{x} \in \mathbb{C}^k$ are the first few Fourier-series coefficients and $\mathbf{x} \in \mathbb{C}^k$ consists of the amplitudes of k spikes. The goal is to recover the amplitudes $\mathbf{x}$ and gridless support $\boldsymbol{\omega} = \{\omega_0, \cdots, \omega_{k-1}\}$ from $\mathbf{y}$.
- **Discrete Case:** $\mathbf{Q} \in \mathbb{C}^{M \times N}$ is a partial DFT matrix and $\mathbf{x} \in \mathbb{C}^N$ is a sparse vector with $\|\mathbf{x}\|_0 = k$. The objective is to recover the support as well as the non-zero entries of $\mathbf{x}$.

The problems of super-resolution and support recovery are closely connected and similar analysis tools can be exploited to provide robust guarantees of estimating $\mathbf{x}$ in both cases.

The major contribution of this dissertation to the theory of super-resolution/support-recovery is that we show it is possible to have robust estimate even in the regime $k \geq M$ by exploiting the idea of sparse array. No analysis has been done in this regime in literature. The application of sparse array is natural as the measurement matrix $\mathbf{Q}$ is of Fourier type. A difference co-array will arise in the correlation domain such that $\mathbf{Q}$ (in the correlation domain) will have more rows than the physical number of measurements. This will allow for the stable recovery of larger support.

We will highlight the role of non-negativity constraint on $\mathbf{x}$ which arises if we only focus on the support recovery and exploit the correlation information of the signal ($\mathbf{x}$ will represent the signal powers in this case). Furthermore, with the non-negative constraint, we can formulate a non-convex program to promote the sparsity of the recovered support. The key idea is that a non-negative vector can be written as a entry-wise square of another vector with the same support. We propose an iterative algorithm to efficiently solve this non-convex program. As another

contribution of independent interest, we provide a benchmark named Modulus of Continuity (MC) for evaluating different convex/non-convex super-resolution algorithms with non-negative constraint. By computing the MC of super-resolution with non-negative constraint, we show that non-negative constraint will ensure a lower error bound compared with general case.

1.4 Phase Retrieval and Structured Fourier Samplers

Besides covariance compression, quadratic Samplers also arise in a famous problem from high resolution optical imaging, namely that of phase retrieval. It finds extensive application in many areas of imaging science, such as X-ray crystallography, diffraction imaging, molecular imaging and high resolution microscopy, astronomical imaging, to name a few. The goal is to recover an unknown signal (or an image) from the magnitude of its Fourier measurements. It arises from the fact that detectors often are unable to measure the phase of incident optical wave, whereby much of the structural information contained in the image may be lost. The (noiseless) measurement model for phase retrieval can be represented as

$$y_i = |\langle \mathbf{a}_i, \mathbf{x} \rangle|, i = 1, 2, \cdots, M \tag{1.4}$$

Here $\mathbf{x} \in \mathbb{C}^N$ is the unknown signal of interest and $y_i, 1 \leq i \leq M$ represent M intensity measurements acquired using the measurement vectors $\mathbf{a}_i, i = 1, 2 \cdots, M$. The problem of phase retrieval has received great attention across scientific and engineering communities [75, 77, 100, 105], both due to fundamental mathematical questions on the number of necessary and sufficient measurements (i.e. relation between M and N) and the need for developing robust algorithms that can successfully recover $\mathbf{x}$ (upto a trivial global phase ambiguity) from $y_i, i = 1, 2, \cdots M$. The problem of Fourier phase retrieval (i.e . when $\{\mathbf{a}_i\}_{i=1}^M$ represent columns of a DFT matrix) is particularly elusive, since the presence of multiple spectral factors make the problem fundamentally ill-posed (to be elaborated later in Chapter 4). In recent times, there

have been attempts at resolving this ambiguity by using sparsity as a prior [104], using coded diffraction masks [164], or using STFT [72]. However, these methods are often sub-optimal in terms of the number of measurements required to ensure perfect reconstruction.

In this dissertation, we will develop a new Fourier-based measurement system (again, inspired from nested arrays) that can perform phase retrieval with provably near-minimal number of measurements. A key idea is to realize that the non-linear measurement model for phase retrieval can be recast in the following form

$$y_i^2 = \left(\mathbf{a}_i^T \otimes \mathbf{a}_i^H\right) \text{vec}(\mathbf{x}\mathbf{x}^H) \tag{1.5}$$

It can be seen that y_i^2 is *a linear function* of the matrix $\mathbf{x}\mathbf{x}^H$ and this equivalent linear map actually consists of *quadratic products* of the measurement vector $\mathbf{a}_i$. We can actually view (1.5) as a special case of covariance sketching, where $\mathbf{x}\mathbf{x}^H$ represents a *rank*-1 *covariance matrix*. This formulation will help us exploit ideas from covariance estimation using nested samplers to design highly efficient Fourier-based measurement vectors $\mathbf{a}_i$ for phase retrieval.

1.5 Dissertation Structure and Contributions

This dissertation is organized as follows:

1.5.1 Compressive Toeplitz Covariance Estimation and Generalized Nested Sampler

In Chapter 2, we address the problem of compressive Toeplitz covariance estimation using Generalized Nested Sampler (GNS). We will show how to exploit the difference coarray structure embedded in the quadratic measurement model, and achieve the order-wise optimal sample complexity. The contributions are summarized as follows

- In the noiseless case, we obtain order-wise optimal sample complexity using GNS for the general or low rank case respectively. Our recovery algorithm is deterministic and has much lower computational complexity compared to previous methods, especially in the low rank case.

- In the noisy case, we provide a universal robustness analysis of Toeplitz covariance estimation which is independent of particular algorithms. This analysis relies on the Vandermonde decomposition of Toeplitz covariance matrices and exploits the tools in the literature of line spectrum estimation.

- For low-rank Toeplitz covariance matrix estimation, our algorithm does not involve any penalty such as nuclear norm to promote low rank structure. Instead, we propose a simple two step approach and make predictions to estimate the original high-dimensional covariance matrix. We demonstrate that positive semidefinite Toeplitz constraint alone will guarantee a stable recovery. Our algorithm is computationally efficient as we do not solve a convex program of the ambient signal dimension.

- We analyze the non-convex maximum likelihood (ML) algorithm for estimating low rank Toeplitz covariance matrix with finite noisy measurements. By analyzing three equivalent ML formulations, we show exact recovery is guaranteed with GNS measurements.

1.5.2 Super Resolution and Support Recovery with Non Negative Constraint

In Chapter 3, we discuss the closely-related problems of super-resolution and support recovery. For the first time in literature, we introduce the idea of sparse array to recover more sources than sensors. We consider both continuous and discrete settings. We particularly highlight the role of non-negativity in the robustness analysis. We make following significant contributions

- Motivated by the Fourier structure of the measurement matrix $\mathbf{Q}$, we exploit the correlation information and apply the idea of sparse array to demonstrate the possibility of recovering larger support than the number of measurements. In the correlation domain, $\mathbf{x}$ will represent the signal powers and is naturally non-negative. Our recovery algorithms will enforce the non-negative constraint.

- We develop a universal bound on the error of estimating $\mathbf{x}$. The analysis is applicable to both convex/non-convex algorithms as long as the non-negativity constraint is enforced and a separation condition of the true support is satisfied .

- With the universal upper bound of the estimation error, we provide a computationally efficient algorithm to recover the support by simply hard-thresholding the estimated signal powers. Our analysis does not require the prior knowledge of noise and signal powers.

- We try to understand how non-negativity constraint is helpful on stable recovery by formulating an equivalent non-convex optimization program. Such formulation directly relies on the non-negative property, and is shown to further promote sparsity than convex programs. We provide a reweighted iterative algorithm to solve the non-convex program.

- We study a particular super-resolution problem known as fluorescence microscopy. Motivated by the underlying physics, we exploit the facts that the sources are statistically uncorrelated and the point spread function can be approximated by a Gaussian function. After applying the sparse array idea, we successfully show that more sources than sensors can be localized by utilizing a novel sum coarray structure in the correlation domain.

1.5.3 Phase Retrieval and Structured Fourier Samplers

In Chapter 4, we study the problem of phase retrieval using structured Fourier samplers. Compared to previous relevant work, we have made several significant contributions:

- We propose a novel Fourier-type sampler called partial nested Fourier sampler (PNFS) that can exploit the quadratic measurement model of phase retrieval. A difference coarray structure arises which can decompose the coupled correlation lags to guarantee uniqueness of recovery. Since the sampling scheme is deterministic, our recovery algorithm is also deterministic and computationally efficient compared with conventional convex methods with random measurements.

- Using PNFS, for the first time in literature, we achieve $O(s \log N)$ sample complexity for stable sparse phase retrieval.

Finally, in Chapter 5, we discuss two important extensions of our theoretical work in previous chapters, and further emphasize the usefulness of non-negativity constraint. In particular, we will present a unified analysis of co-array interpolation, which is applicable to a broad family of algorithms with positive semidefinite constraint. This work won the *best student paper award (first place)* at IEEE ICASSP 2017 in New Orleans, USA.

Then, we study the bilinear problem with non-negative constraint. For the first time in literature, an exact analysis of the set of ambiguous solutions is developed, especially in the case with more variables than equations. This work was nominated for the best student paper award at IEEE CAMSAP 2019.

In the end, we will apply the nested array idea to the problem of Kriging and provide a robustness guarantee based on total least-squares method. The field of interest can be stably estimated with fewer sensors as a benefit of exploiting the spatial correlation information and sparse array structure.

Chapter 2

Compressive Covariance Sketching and Quadratic Samplers

This chapter considers the problem of compressively sampling wide sense stationary random vectors with a low rank Toeplitz covariance matrix. Certain families of structured deterministic samplers are shown to efficiently compress a high dimensional Toeplitz matrix of size $N \times N$, producing a compressed sketch of size $O(\sqrt{r}) \times O(\sqrt{r})$.The reconstruction problem can be cast as that of line spectrum estimation, whereby, in absence of noise, Toeplitz matrices of any size N can be exactly recovered from compressive sketches of size $O(\sqrt{r}) \times O(\sqrt{r})$, no matter how large N is. In presence of noise and finite data, the line spectrum estimation algorithm is combined with a novel denoising technique that only exploits a positive semidefinite (PSD) Toeplitz constraint to denoise the compressed sketch using a simple least-squares minimization framework. A major advantage of the algorithm is that it does not require any regularization parameter. It also enjoys lower computational complexity owing to its ability to *predict the unobserved entries* of the low rank Toeplitz matrix. Explicit bounds on the reconstruction error are established and it is shown that the PSD constraint on the denoiser is sufficient to ensure stable reconstruction from a sketch of size $O(\sqrt{r}) \times O(\sqrt{r})$. Extensive simulations demonstrate

that the proposed algorithm provides better performance over random samplers and algorithms that use nuclear norm based regularizers.

2.1 Introduction

Estimation of second-order statistics (or correlation) of high-dimensional data plays a central role in modern statistical analysis and information processing. The covariance matrix acts as a sufficient statistic in many signal processing problems [33, 34]. It also provides a compact summary of a large dataset, and is used for dimensionality reduction. A popular example is that of principal component analysis [35, 36] where the second-order statistics of the data are used to project the data along the dominant eigenvectors, thereby attaining dimension reduction. The inverse covariance matrix also plays an important role in many applications related to classification of Gaussian data and establishing independence relations in exploratory data analysis and testing [37]. Owing to its large dimension, it may not be always possible to store and/or reliably communicate the entire high dimensional covariance matrix. Hence, it is crucial to obtain a compressive sketch of the covariance matrix which can be efficiently stored and transmitted. The topic of *compressive covariance sampling* [38–40,42,43], is receiving increasing attention, where the goal is to compress and reconstruct the high dimensional covariance matrix using so-called *covariance samplers*. In general, it is not possible to design a compressive sampler unless the correlation matrix exhibits some low dimensional structure that allows compression. Typical structural assumptions include that of sparsity, low rank and stationarity of data (which imposes a Toeplitz structure on the covariance matrix) [38, 40].

2.1.1 Related Work

The problem of obtaining a sketch of the covariance matrix by compressively sampling the underlying random process has been recently investigated in a number of works [28,38–40,42]. In

[28], a high dimensional covariance matrix $\mathbf{\Sigma} \in \mathbb{R}^{N\times N}$ is sketched using rank-one measurements where $\mathbf{\Sigma}$ is assumed to exhibit distributed sparsity. The required sample size for compressing sparse covariance matrices is proved to be $O(\sqrt{N}\log N)$. When the covariance matrix exhibits a Toeplitz structure, compressive covariance sensing becomes equivalent to compressive power spectrum estimation, which has been investigated in [39, 40, 42–45]. The common theme in this body of work is the use of deterministic sub-Nyquist samplers (often inspired from the idea of difference-sets [32]) for compressively sampling WSS signals. Such samplers can compress Toeplitz matrices of size $N \times N$ using a sketch of size $O(\sqrt{N}) \times O(\sqrt{N})$. The work in [46], considers a cyclostationary signal model for which the number of measurements is shown to be $O(\sqrt{N})$. In [48], the authors consider the estimation of Toeplitz covariance matrix via Maximum Likelihood methods. However, the results are asymptotic and no stability result is available for a noisy and finite snapshot model.

In this chapter, besides Toeplitz structure, we also exploit low rank of the covariance matrix that allows further compression over what is possible by exploiting the Toeplitz structure alone. In many practical problems across scientific and engineering disciplines, the signal of interest often has a spectrally sparse represenation, i.e., its power spectrum can be modeled as a superposition of a few spikes [109]. By Carathéodory's theorem, the covariance matrix of such signals possess low rank positive semidefinite (PSD) Toeplitz structure, and they arise in a large number of practical applications such as direction finding for radar, sonar and astronomical imaging [49, 110–112, 139], neural source localization using sensor arrays (e.g. EEG/MEG) [113–115], source localization and inverse scattering in seismic imaging [116, 117, 233] and so forth. In these problems, low rank is typically attributed to the presence of only a few sources or scatterers compared to the number of physical sensors. [1] A key feature of our reconstruction algorithm is that unlike sparsity enforcing techniques [119, 120] which discretize the parameter space into a grid with finite points

[1] Owing to the presence of noise, the covariance matrix of the received signal in these applications is not strictly low rank, rather it is a sum of low rank PSD Toeplitz matrix, and a scaled diagonal matrix, where the scaling factor denotes noise power. We will show in Sec. 2.4 that the proposed sampling and reconstruction framework continues to be applicable for such *full rank positive definite* Toeplitz matrices as well.

and assumes the parameters to lie on this grid, we allow the spectral lines to take up any value continuously within a range. In this regard, our algorithm belongs to a class of "discretization-free" or "gridless line spectrum estimation" techniques which encompass classical methods such as MUSIC, covariance fitting algorithms based on Maximum Likelihood criterion [48, 121], as well as more recent techniques based on nuclear and atomic norm minimization [57, 59].

The terminology "gridless" is also used in recent literature [122, 123, 127, 128]. The authors in [122, 127] consider gridless extension of the SPICE algorithm [120] and its connection to regularizer based methods. In [123], a denoising based MUSIC algorithm for DOA estimation of more sources than sensors is proposed to combat finite snapshot estimation errors. In [128], algorithmic implementations of atomic norm based algorithms are discussed. However, these works do not establish any stability analysis or explicit error bounds for gridless methods as a function of the number of snapshots.

2.2 Preliminaries for Low Rank Toeplitz Recovery

We begin by introducing our measurement model and then review a key property of low rank Toeplitz matrices that we will exploit throughout the paper.

2.2.1 Model Description

Consider a sequence of high dimensional zero-mean random vectors $\{\mathbf{x}_p\}_{p=-\infty}^{\infty}$ of dimension N (N is a large integer), whose covariance matrix is given by $E(\mathbf{x}_p\mathbf{x}_p^T) \triangleq \mathbf{T} \in \mathbb{R}^{N\times N}$. We compressively sample the data using a sampling matrix $\mathbf{A}_s \in \mathbb{R}^{M\times N}$, $M \ll N$ to obtain $\mathbf{y}_p = \mathbf{A}_s\mathbf{x}_p$ where M is treated as sample size to be minimized throughout the paper. The covariance matrix of $\{\mathbf{y}_p\}_{p=-\infty}^{\infty}$ is given by

$$\mathbf{R}_{\mathbf{Y}} = \mathrm{E}[\mathbf{y}_p\mathbf{y}_p^H] = \mathbf{A}_s\mathbf{T}\mathbf{A}_s^T \tag{2.1}$$

Instead of the larger covariance matrix $\mathbf{T}$, we store and/or transmit the compressed covariance matrix $\mathbf{R_Y} \in \mathbb{R}^{M\times M}$. This paper focuses on the special case when the vectors $\mathbf{x}_p$ are wide-sense stationary, whereby its covariance matrix $\mathbf{T} \in \mathbb{R}^{N\times N}$ is a Toeplitz matrix, satisfying $[\mathbf{T}]_{m,n} = [\mathbf{T}]_{m+k,n+k} = t_{|m-n|}, \quad \forall m, n, k$. The goal of this paper is to design the sampling matrix $\mathbf{A}_s$ to obtain the compressed sketch $\mathbf{R_Y}$ and develop a reconstruction algorithm to recover $\mathbf{T}$ from $\mathbf{R_Y}$ under the assumption that $\mathbf{T}$ is *Toeplitz and low rank*.

2.2.2 Low Rank Toeplitz Matrix and Vandermonde Decomposition Lemma

Our proposed sampling scheme and recovery algorithms are fundamentally based on the famous Carathéodory's theorem [56, 64, 65] that provides an explicit algebraic structure of $\mathbf{T}$ in terms of a Vandermonde matrix:

Theorem 1 *A positive semidefinite Toeplitz matrix $\mathbf{T} \in \mathbb{R}^{N\times N}$ with rank $r < N$ has the following decomposition:*

$$\mathbf{T} = \mathbf{V}_N \mathbf{D} \mathbf{V}_N^{\mathrm{H}} \tag{2.2}$$

where $\mathbf{V}_N \in \mathbb{C}^{N\times r} = [\mathbf{v}_N(f_1), \mathbf{v}_N(f_2), \cdots, \mathbf{v}_N(f_r)]$ and each column $\mathbf{v}_N(f_i)$ is defined as

$$[\mathbf{v}_N(f_i)]_k = e^{j2\pi f_i(k-1)} \quad f_i \in (-1/2, 1/2],\ 1 \le k \le N \tag{2.3}$$

The matrix $\mathbf{D} \in \mathbb{R}^{r\times r}$ is diagonal with positive entries $\{d_1, d_2, \cdots, d_r\}$.

Remark 1. *The Vandermonde decomposition lemma is also true for complex valued low rank PSD Toeplitz matrices. However, we present it for real valued $\mathbf{T}$ which is the focus of current paper.*

The decomposition (2.2) allows us to deduce similar factorization for all leading principals

of $\mathbf{T}$. In particular, we have the following corollary

Corollary 1 *For all $1 \leq n \leq N$, we have the following decomposition of $\mathbf{T}_{(n)}$*

$$\mathbf{T}_{(n)} = \mathbf{V}_n \mathbf{D} \mathbf{V}_n^{\mathrm{H}} \tag{2.4}$$

where the columns of $\mathbf{V}_n \in \mathbb{C}^{n \times r}$ are defined in the same way as (2.3).

The degrees of freedom of a matrix is defined as the minimum number of real numbers needed to represent it. Using Carathéodory's theorem, the degrees of freedom of a rank r Toeplitz matrix is given by

Corollary 2 *A PSD Toeplitz matrix $\mathbf{T} \in \mathbb{R}^{N \times N}$ with rank r, has at most $2r$ degrees of freedom (DOF), characterized by the real numbers $\{f_i, d_i\}_{i=1}^{r}$ given by (2.3).*

Two important remarks follow:

- The DOF of a rank $r < N$ Toeplitz matrix is *completely independent* of the ambient dimension N. We will exploit this property to propose a recovery technique that has significantly lower complexity than the nuclear norm minimization framework of [58, 59].
- Any (order-wise) optimal sketching method should produce a sketch $\mathbf{R}_{\mathbf{Y}}$ of size $O(\sqrt{r}) \times O(\sqrt{r})$, i.e., it should contain $O(r)$ measurements of $\mathbf{T}$. The proposed sampling and reconstruction scheme will be shown to be order-wise optimal.

2.3 Near Optimal Compression and Recovery of Low-Rank Toeplitz Matrices

The Vandermonde decomposition lemma dictates that a rank r PSD Toeplitz matrix can be compressed by simply retaining its $n \times n$ principal $\mathbf{T}_{(n)}$ where $n = O(r)$. The possibility of

compressing and reconstructing a $n \times n$ Toeplitz matrix simply by exploiting the redundancies in its entries has been explored in [39,40,42,43] where the sampling matrix $\mathbf{A}_s$ is constructed using a minimum redundancy sampler or a sparse ruler [32]. The size of the optimally compressed covariance matrix is $O(\sqrt{n}) \times O(\sqrt{n})$ and it retains all n distinct entries of $\mathbf{T}_{(n)}$. However, one disadvantage of using sparse rulers is that there are no closed form expressions for the sampling set, or the exact size of the sketch. We recently proposed another structured deterministic sampler, namely, the Generalized Nested Sampler (GNS) [2, 10] that ensures perfect reconstruction of $\mathbf{T}_{(n)}$ from a compressed sketch of size $O(\sqrt{n}) \times O(\sqrt{n})$. An advantage of GNS is that closed form expressions for the sampling matrix $\mathbf{A}_s$ and the size of the sketch can be derived for almost any n.

The use of random samplers for compressing Toeplitz matrices has also been considered in [40], and it is shown that with probability 1, they attain the same order wise compression (i.e. $O(\sqrt{n})$) as sparse rulers. However, these samplers usually lead to a dense measurement matrix $\mathbf{A}_s$ while sparse rulers or GNS yield highly sparse $\mathbf{A}_s$ which can require less storage space and allow faster computations. We next introduce the principles behind a GNS and discuss how it can be used for low-rank Toeplitz compression. We would like to reiterate that in principle, the GNS can be replaced by a sparse ruler or minimum redundancy sampler, without any loss in generality of the derived results.

2.3.1 Review of Generalized Nested Sampler (GNS)

The Generalized Nested Sampler (GNS) was first introduced in [10] and further developed in [2]. Following [10], we review some key properties in this section. A GNS is defined in terms of two integer-valued functions $\Theta(N)$ and $\Gamma(N)$.

Definition 1 *For any integer $L \geq 6$, define $\Theta(L)$ and $\Gamma(L)$ as integers such that*

$$\Theta(L) = \lfloor \sqrt{L + \frac{1}{4}} - \frac{1}{2} \rfloor \quad \Gamma(L) = 1 + L - \Theta^2(L) \tag{2.5}$$

Here $\lfloor . \rfloor$ denotes the floor operation.

Given an integer L, a GNS can be defined as a measurement matrix as follows:

Definition 2 *For any integer $L \geq 6$, define the effective Generalized Nested Sampling matrix $\mathbf{A}_{GNS}^{L} \in \mathbb{R}^{M \times L}$, with $M = \Gamma(L) + \Theta(L) - 1$, as*

$$[\mathbf{A}_{GNS}^{L}]_{i,j} = \begin{cases} 1 & \text{if } i = j,\ 1 \leq i \leq \Gamma(L) \\ 1 & \text{if } j = (i - \Gamma(L))\Theta(L) + i, \\ & \quad \Gamma(L) < i \leq M \\ 0 & \text{Elsewhere} \end{cases} \tag{2.6}$$

Remark 2. *As a simple observation, from (2.5), we always have $M < L$ for $L \geq 6$. Also, M is of order $O(\sqrt{L})$. The specific case of $L = M^2/4 + M/2 - 1$ was introduced as "Nested Array" in [45]. As discussed in [10], the fundamental idea behind GNS or other sparse ruler type sampler is to exploit the difference set of the sampling indices. In particular, each row of $\mathbf{A}_s$ contains a single 1 and let $c(i)$ denote the index of the column containing it. Then, the (i,j)th entry of $\mathbf{R}_{\mathbf{Y}}$ corresponds to $t_{c(i)-c(j)}$. The length of smallest range over which (i,j) should be chosen so that $\{c(i) - c(j)\}$ spans all integers from 0 to $L-1$ is $O(\sqrt{L})$ and the GNS shows a constructive way to select $c(i)$ over this range.*

The following result from [10] shows how to compress a $N \times N$ Toeplitz matrix without assuming it to be low rank.

Lemma 1 *A real symmetric Toeplitz matrix $\mathbf{T} \in \mathbb{R}^{N \times N}$ can be exactly recovered from its compressed measurement $\mathbf{R}_{\mathbf{Y}} = \mathbf{A}_{GNS}^{N} \mathbf{T} (\mathbf{A}_{GNS}^{N})^{T}$ where $\mathbf{A}_{GNS}^{N} \in \mathbb{R}^{M \times N}$ is a Generalized Nested Sampling Matrix given by (2.6).*

To illustrate how GNS works, we show an example of small dimension. Let $\mathbf{T}$ be a real PSD Toeplitz matrix of dimension $N = 6$ with first column $[t_0, t_1, \cdots, t_5]^T$. Then $\mathbf{A}_{GNS}^N$ is given by

$$\mathbf{A}_{GNS}^N = \left[\begin{array}{ccc:ccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hdashline 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \tag{2.7}$$

Then the compressed sketch $\mathbf{R_Y}$ is given by

$$\mathbf{R_Y} = \left[\begin{array}{ccc:c} t_0 & t_1 & t_2 & t_5 \\ t_1 & t_0 & t_1 & t_4 \\ t_2 & t_1 & t_0 & t_3 \\ \hdashline t_5 & t_4 & t_3 & t_0 \end{array}\right] \tag{2.8}$$

Then obviously, we can recover $\mathbf{T}$ from observation $\mathbf{R_Y}$.

Compression Using Structure Alone: It is worth noting that the row or column size M of the compressed matrix $\mathbf{R_Y}$ is $O(\sqrt{N})$. This shows that GNS can compress a $N \times N$ Toeplitz matrix $\mathbf{T}$ by entirely exploiting its structure, even when it is not necessarily low rank. As an immediate consequence of Lemma 1, we have following corollary on recovering the $n \times n$ principal $\mathbf{T}_{(n)}$ of $\mathbf{T}$.

Corollary 3 *For any $6 \leq n \leq N$, $\mathbf{T}_{(n)}$ can be exactly recovered from its compressive sketch $\mathbf{R_Y} = \mathbf{A}_s \mathbf{T} \mathbf{A}_s^T$ where the measurement matrix $\mathbf{A}_s \in \mathbb{R}^{(\Gamma(n)+\Theta(n)-1)\times N}$ is given by $\mathbf{A}_s$ defined as*

$$\mathbf{A}_s = [\mathbf{A}_{GNS}^n, \mathbf{0}] \tag{2.9}$$

where $\mathbf{A}_{GNS}^n \in \mathbb{R}^{(\Gamma(n)+\Theta(n)-1)\times n}$ is a GNS defined as (2.6).

2.3.2 Sampling and Reconstruction in absence of noise

We now propose an end-to-end sampling and reconstruction scheme for low rank PSD Toeplitz matrices in noiseless case using GNS as a representative example of an order-wise optimal sampler. In principle, GNS can also be replaced by a sparse-ruler type sampler [39, 40].

1. **Compression:** Given a sequence of high dimensional WSS data $\mathbf{x}_p \in \mathbb{R}^N$ with Toeplitz covariance matrix $\mathbf{T}$ having rank $r < N$, obtain compressed measurements

$$\mathbf{y}_p \in \mathbb{R}^{\Gamma(r+q)+\Theta(r+q)-1}$$

as

$$\mathbf{y}_p = \mathbf{A}_s \mathbf{x}_p, \quad \mathbf{A}_s = [\mathbf{A}_{GNS}^{r+q}, \mathbf{0}] \tag{2.10}$$

Here $q \geq 1$ and $\mathbf{A}_{GNS}^{r+q} \in \mathbb{R}^{(\Gamma(r+q)+\Theta(r+q)-1)\times(r+q)}$ is a GNS sampler. Compute the covariance of the compressed measurements to obtain the required sketch

$$\mathbf{R}_\mathbf{Y} \in \mathbb{R}^{(\Gamma(r+q)+\Theta(r+q)-1)\times(\Gamma(r+q)+\Theta(r+q)-1)}$$

of $\mathbf{T}$ as $\mathbf{R}_\mathbf{Y} \triangleq E(\mathbf{y}_p \mathbf{y}_p^T) = \mathbf{A}_s \mathbf{T} \mathbf{A}_s^T$ From the structure of $\mathbf{A}_s$ in (2.10), it can be readily observed that $\mathbf{R}_\mathbf{Y} = \mathbf{A}_{GNS}^{r+q} \mathbf{T}_{(r+q)} (\mathbf{A}_{GNS}^{r+q})^T$

2. **Reconstruction:** Given $\mathbf{R}_\mathbf{Y}$ obtained from the compression stage, we proceed to reconstruct $\mathbf{T}$ as follows:

   (a) Recover $\mathbf{T}_{(r+q)}$ from $\mathbf{R}_\mathbf{Y}$. This is possible as dictated by Corollary 3.

   (b) Noticing that $\mathbf{T}_{(r+q)}$ is a rank deficient (rank r) PSD Toeplitz matrix for $q \geq 1$, let $\{f_i, d_i\}, i = 1, 2, \cdots, r$ be the parameters describing its parametric decomposition

(2.4). Recover $\{f_i, d_i\}$ using MUSIC and least-square (LS) according to (2.4).

(c) Given $\{f_i, d_i\}$, recover $\mathbf{T}$ using its Vandermonde decomposition (2.2).

Fig. 2.1 shows the pictorial depiction of the end-to-end compression and reconstruction system.

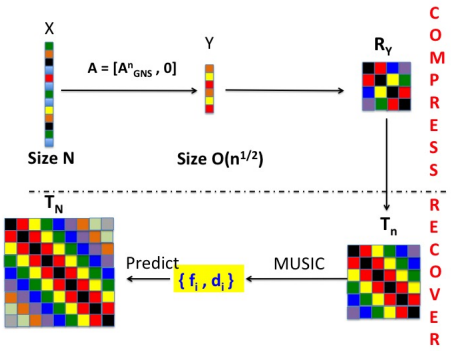


Figure 2.1: GNS based sampling and reconstruction of low rank PSD Toeplitz matrix in absence of noise.

2.3.3 Efficient Reconstruction Algorithm via Linear Prediction

In [2], we relate the low-rank Toeplitz covariance estimation problem to linear prediction of line spectrum process. Consider a zero mean scalar WSS process $z[n]$ whose autcorrelation matrix of size N is given by $\mathbf{T}$. In particular, defining $\mathbf{z}_N[n] = [z[n], z[n-1], \cdots z[n-N+1]]^T$, we have

$$\mathbf{T} = E(\mathbf{z}_N[n]\mathbf{z}_N[n]^T) \tag{2.11}$$

Now we know that $\mathbf{T}_{(r+1)}$ is of rank r, and hence has a 1 dimensional null space. Without losing generality, let $\mathbf{b} \in \text{Null}(\mathbf{T}_{(r+1)})$ and to make it unique, we normalize the first entry of $\mathbf{b}$ to

1, i.e., $b_1 = 1$. Then,

$$\begin{aligned}
&\mathbf{T}_{(r+1)}\mathbf{b} = 0 \Rightarrow \mathbf{b}^T\mathbf{T}_{(r+1)}\mathbf{b} = 0 \\
&\Rightarrow \mathbf{b}^T \mathrm{E}\left\{\mathbf{z}_{r+1}[n](\mathbf{z}_{r+1}[n])^T\right\}\mathbf{b} = 0 \\
&\Leftrightarrow \mathrm{E}\left\{|\mathbf{b}^T\mathbf{z}_{r+1}[n]|^2\right\} = 0 \Leftrightarrow \mathbf{b}^T\mathbf{z}_{r+1}[n] \equiv 0
\end{aligned} \tag{2.12}$$

Since $b_1 = 1$, (2.12) implies

$$z[n] = -\sum_{i=1}^{r} b_{i+1} z[n-i] \tag{2.13}$$

Notice that (2.13) describes an order r linear predictor for the WSS process $z[n]$. In fact, it also shows that $z[n]$ is a predictable process, which can be completely predicted from its past r samples. The filter coefficients $-b_i, i = 2, 3, \cdots r+1$ actually correspond to the optimum order-r linear predictor (since it attains 0 prediction error) and can be obtained by solving (for instance, by using Levinson Durbin Algorithm)

$$\mathbf{T}_{(r)}\tilde{\mathbf{b}} = \mathbf{t}_r \tag{2.14}$$

where $\mathbf{t}_r = [t_1, t_2, \cdots t_r]^T$, and $[\tilde{\mathbf{b}}]_i = -b_{i+1}, i = 1, 2, \cdots r$. Notice that $\mathbf{T}_{(r)}$ is full rank and $\mathbf{T}_{(r)}$ and $\mathbf{t}_r$ are completely known since GNS has already recovered $\mathbf{T}_{(r+1)}$. The equation (2.13) immediately leads to an iterative approach to estimate t_i. We can write

$$\begin{aligned}
&\mathrm{E}\{z[n]z[n-r-k]\} = -\sum_{i=1}^{r} b_{i+1}\mathrm{E}\{z[n-i]z[n-r-k]\} \\
&\Rightarrow t_{r+k} = -\sum_{i=1}^{r} b_{i+1} t_{r+k-i}
\end{aligned} \tag{2.15}$$

Since GNS provides us with $t_i, 1 \leq i \leq r$, all the other entries of $\mathbf{T}$ can be recovered in an

Table 2.1: GNS for Linear Prediction

Input: $z[n]$ *Output:* LP coefficients $\tilde{\mathbf{b}}$
(a) Construct a GNS matrix $\mathbf{A_s} \in \mathbb{R}^{M \times N}$ with $N = r + 1$. This implies $M = O(\sqrt{r})$
(b) Obtain compressed measurements $\mathbf{y}[n] = \mathbf{A_s}\mathbf{z}_{r+1}[n]$
(c) Compute the compressed correlation matrix $\mathbf{R_Y} = E(\mathbf{y}[n]\mathbf{y}^T[n])$
(d) Obtain $\mathbf{T}_{(r+1)}$ from $\mathbf{R_Y}$ (using Lemma 1).
(e) Compute $\tilde{\mathbf{b}}$ using $\mathbf{T}_{(r+1)}$.

iterative way using (2.15), whereby the original Toeplitz matrix is exactly recovered.

Consider the problem of linear prediction of a scalar WSS process $z[n]$ whose autocorrelation matrix $\mathbf{T} \in \mathbb{R}^{N \times N}$ is of rank r for every $N \geq r$. From (2.13), it is easily seen that $z[n]$ is a predictable process, each value of which can be exactly predicted from its past r samples. The linear prediction coefficients $\tilde{\mathbf{b}}$ are obtained by solving (2.14) where $\mathbf{T}_{(r)}$ is full rank. Typically $\mathbf{T}_{(r+1)}$ is computed by uniformly sampling $z[n]$ so that $\mathbf{T}_{(r+1)} = E(\mathbf{z}_{r+1}[n]\mathbf{z}_{r+1}^T[n])$ where $\mathbf{z}_{r+1}[n] \in \mathbb{R}^{r+1}$ contains $r + 1$ consecutive samples of $z[n]$. However, the properties of GNS can be leveraged to compute $\mathbf{T}_{(r+1)}$ more efficiently using $O(\sqrt{r})$ non uniform samples of $z[n]$ as dictated by the sampling matrix of the GNS. The steps of this procedure are described in Table 2.1.

2.4 Stable Reconstruction with noisy finite measurements: A Regularizer-Free Approach

In this section, we consider a finite number (L) of noisy compressed measurements as

$$\mathbf{y}_p = \mathbf{A}_s\mathbf{x}_p + \mathbf{n}_p \quad 1 \leq p \leq L \tag{2.16}$$

where $\{\mathbf{n}_p\}$ are independent zero-mean white Gaussian noise vectors with covariance $\sigma_n \mathbf{I}_M$. We propose a novel algorithm that generalizes the noiseless reconstruction technique proposed in Sec. 2.3.2, by adding a simple denoiser prior to performing MUSIC. The main idea is to solve a constrained least squares minimization problem (with positive semidefinite and Toeplitz constraints) to denoise the n noisy measurements ($n = O(r)$) and use them to predict the remaining $N-n$ entries of $\mathbf{T}$. We show that this algorithm leads to provably stable reconstruction with finite noisy measurements.

2.4.1 Vandermonde Decomposition of Positive Definite Toeplitz Matrices

Our algorithm is based on the following elegant Vandermonde decomposition lemma from [68] that holds for any positive *definite* Toeplitz matrix:

Lemma 2 *[68] A positive (semi) definite Toeplitz matrix $\mathbf{T} \in \mathbb{R}^{N \times N}$ has the following decomposition :*

$$\mathbf{T} = \mathbf{V}_{N \times N'} \mathbf{D} \mathbf{V}_{N \times N'}^H + \sigma \mathbf{I}_N \tag{2.17}$$

Here σ is the smallest singular value of $\mathbf{T}$, $\mathbf{I}_N \in \mathbb{R}^{N \times N}$ is the identity matrix and $\mathbf{D} = diag(d_1, d_2, \cdots, d_{N'})$ with $d_i > 0$. The matrix

$$\mathbf{V}_{N \times N'} \in \mathbb{C}^{N \times N'} = [\mathbf{v}_N(f_1), \mathbf{v}_N(f_2), \cdots, \mathbf{v}_N(f_{N'})]$$

is a Vandermonde matrix satisfying $N' < N$ with the ith column $\mathbf{v}_N(f_i)$ given by (2.3).

Remark 3. *Lemma 18 is actually a more general version of (2.2) and it applies to both positive definite and semidefinite Toeplitz matrices, using appropriate values of σ and N'. When $\mathbf{T}$ is low rank, (5.30) reduces to (2.2) with $\sigma = 0, N' = rank(\mathbf{T})$ If $\mathbf{T}$ is positive definite, $N' = N - m_\sigma$*

where m_σ is the multiplicity of its smallest singular value σ. In both cases, the frequencies in the representations (5.30) and (2.2) are uniquely determined and can be found using a harmonic retrieval algorithm such as MUSIC [149].

2.4.2 Regularization Free Reconstruction from Finite Noisy Measurements

Consider the following noisy compressive sketch (of size $O(\sqrt{n}) \times O(\sqrt{n})$, where $n \geq r+1$) of $\mathbf{T}$ obtained from the sample covariance corresponding to the measurements (2.16)

$$\begin{aligned} \tilde{\mathbf{R}}_{\mathbf{Y}} &= \frac{1}{L}\sum_{l=1}^{L} \mathbf{y}_l \mathbf{y}_l^T = \mathbf{A}_s \mathbf{T}_{\mathbf{1}} \mathbf{A}_s^T + \mathbf{W}^{(L)} \\ \mathbf{T}_{\mathbf{1}} &= \mathbf{T} + \sigma_n \mathbf{I}_N \end{aligned} \tag{2.18}$$

Here $\mathbf{W}^{(L)}$ denotes the finite snapshot estimation error associated with the empirical covariance matrix $\tilde{\mathbf{R}}_{\mathbf{Y}}$. As before, the sampling matrix $\mathbf{A}_s$ is an order-wise optimal structured sampler (such as the GNS in (2.9), designed for a specified $n \geq r+1$) and we use the fact that $\mathbf{A}_s \mathbf{I}_N \mathbf{A}_s^T = \mathbf{I}_M$. From properties of GNS (or sparse ruler), we construct a vector $\tilde{\mathbf{t}}_{(n)} \in \mathbb{R}^n$ by selecting n appropriate entries of $\tilde{\mathbf{R}}_{\mathbf{Y}}$ that [2]

$$\tilde{\mathbf{t}}_{(n)} = \mathbf{t}_{1(n)} + \mathbf{w}_{(n)}^{(L)} \tag{2.19}$$

where $\mathbf{t}_{1(n)} = \mathbf{t}_{(n)} + \sigma_n \mathbf{e}_{1(n)}$ and $\mathbf{e}_{1(n)} = [1, 0, \cdots, 0]^T$.

In order to recover $\mathbf{T}$ from the compressive noisy measurements $\tilde{\mathbf{t}}_{(n)}$, we propose a new

[2] Recall that $\mathbf{R}_{\mathbf{Y}}$ contains a total of n distinct entries which are distributed across the matrix, with repetitions. Instead of directly selecting the corresponding n distinct entries from $\tilde{\mathbf{R}}_{\mathbf{Y}}$, it is also possible to compute an average corresponding to each entry over the number of times it repeats. This will lead to a weighted least squares objective in (2.20) instead of a simple least squares, and Steps 2 and 3 will continue to hold. The analysis framework developed in this paper will be applicable for such a setting as well, with straightforward modifications.

recovery algorithm which is described in Table 1.

The algorithm consists of three steps: (i) Denoising (ii) Parameterization, and (iii) Prediction. The denoising step finds a denoised estimate $\mathbf{t}^{\#}_{(n)}$ of $\mathbf{t}_{1(n)}$, by solving a simple least squares problem under Toeplitz and Positive semidefinite (PSD) constraints. Under these constraints, it is guaranteed that $\mathcal{T}(\mathbf{t}^{\#}_{(n)}) \succeq \mathbf{0}$ Therefore, according to Lemma 18, $\mathcal{T}(\mathbf{t}^{\#}_{(n)})$ has a Vandermonde decomposition given by (2.21). The second step (Parameterization) crucially utilizes this representation to estimate the associated parameters $\sigma^{\#}, n', f_i^{\#}$ and $d_i^{\#}$. The final step then comprises of predicting the remaining $N - n$ entries of $\mathbf{T}$ by using these parameters.

Remark 4. *The PSD constraint in the denoising stage, and the Vandermonde decomposition dictated by Lemma 18 are crucial ingredients in the proposed algorithm. In fact the PSD constraint ensures that we can apply Lemma 18 to the denoised estimate $\mathbf{t}^{\#}_{(n)}$ and use the corresponding parametric representation to predict the remaining $N - n$ entries.*

Remark 5. *Notice that Step 3 in Table 1 essentially extends the PSD Toeplitz matrix $\mathcal{T}(\mathbf{t}^{\#}_{(n)})$ (of size $n \times n$), to a bigger PSD Toeplitz matrix $\mathbf{T}^{\#}_{(N)}$ of size $N \times N$, such that the $n \times n$ principal of $\mathbf{T}^{\#}_{(N)}$ is equal to $\mathcal{T}(\mathbf{t}^{\#}_{(n)})$. If $\sigma^{\#} > 0$, or equivalently, if $\mathcal{T}(\mathbf{t}^{\#}_{(n)})$ is positive definite (which typically happens in presence of noise), it is possible to find two or more $N \times N$ Toeplitz covariance matrices such that $\mathcal{T}(\mathbf{t}^{\#}_{(n)})$ corresponds to the $n \times n$ principal of both [64, 66, 68]. The proposed algorithm finds* one such representation. *On the other hand, if $\sigma^{\#} = 0$, there is only one way to extend it into a PSD Toeplitz matrix of larger dimension N.*

2.4.3 Sampling Requirements with Infinite Snapshots

In order to analyze the performance of the algorithm proposed in Table 2.2, we first show that with infinite snapshots, it recovers the desired PSD Toeplitz covariance matrix $\mathbf{T}$ and noise power σ_n with only $n \geq r + 1$ measurements. Compared to the random sampler proposed

Table 2.2: Low Rank PSD Toeplitz Matrix Recovery In Presence of Error

Input: Noisy measurements $\tilde{\mathbf{t}}_{(n)} \in \mathbb{R}^n$ satisfying (2.19)
Output: Estimate of the entries $t_i, i = 0, 1, \cdots, N-1$ of the PSD Toeplitz matrix $\mathbf{T} \in \mathbb{R}^{N \times N}$ and noise power σ_n

Step 1: Denoising
Obtain denoised estimate $\mathbf{t}^{\#}_{(n)}$ of $\mathbf{t}_{1(n)}$ as

$$\begin{aligned} &\mathbf{t}^{\#}_{(n)} \triangleq \arg \min_{\mathbf{u} \in \mathbb{R}^n} \|\tilde{\mathbf{t}}_{(n)} - \mathbf{u}\|_2 \\ &s.t. \, \mathcal{T}(\mathbf{u}) \succeq 0 \end{aligned} \tag{2.20}$$

where $\mathcal{T}(\mathbf{u})$ denotes a real symmetric Toeplitz matrix whose first column is $\mathbf{u}$.
Step 2: Parameterization
Let the Vandermonde decomposition of $\mathcal{T}(\mathbf{t}^{\#}_{(n)})$ be:

$$\mathcal{T}(\mathbf{t}^{\#}_{(n)}) = \mathbf{V}^{\#}_{n \times n'} \mathbf{D}^{\#} (\mathbf{V}^{\#}_{n \times n'})^H + \sigma^{\#} \mathbf{I}_n \tag{2.21}$$

where $\mathbf{V}^{\#}_{n \times n'} = [\mathbf{v}_n(f^{\#}_1), \mathbf{v}_n(f^{\#}_2), \cdots, \mathbf{v}_n(f^{\#}_{n'})]$ is a Vandermonde matrix of size $n \times n'$ parameterized by frequencies $\{f^{\#}_1, f^{\#}_2, \cdots, f^{\#}_{n'}\}$ and $D^{\#} = diag(d^{\#}_1, d^{\#}_2, \cdots, d^{\#}_{n'})$ where $d^{\#}_i > 0$. The parameters in the above representation can be determined as follows:

- Compute $\sigma^{\#}$ as the minimum singular value of $\mathbf{T}^{\#}_{(n)}$
- Compute n' as the integer $n - m_{\sigma^{\#}}$ where $m_{\sigma^{\#}}$ is the multiplicity of $\sigma^{\#}$.
- Compute the frequencies $\{f^{\#}_1, f^{\#}_2, \cdots, f^{\#}_{n'}\}$ in the representation (2.21). The MUSIC algorithm, for instance, can be used to exactly recover $\{f^{\#}_1, f^{\#}_2, \cdots, f^{\#}_{n'}\}$.
- Given the frequencies, $\sigma^{\#}$ and n', the amplitudes $d^{\#}_i$ are computed as the least squares solution to the system of linear equations (in $d^{\#}_i$) given by (2.21).

Step 3: Prediction Predict the remaining $N - n$ entries of $\mathbf{T}$ as

$$t^{\#}_m = \sum_{i=1}^{n'} d^{\#}_i e^{j2\pi f^{\#}_i m}, \quad n \leq m \leq N-1$$

The estimate of $\mathbf{T}$ is therefore given by $\mathbf{T}^{\#}_{(N)} = \mathcal{T}(\mathbf{t}^{\#}_{(N)})$ and the estimate of σ_n is $\sigma^{\#}$.

in [38] that requires $O(r\text{polylog}N)$ samples with infinite snapshots, the number of measurements required by our algorithm is near-minimal, with no dependence on the ambient dimension N.

Theorem 2 *When $L \to \infty$, the solution of the algorithm proposed in Table 1 satisfies $\mathbf{T} = \mathbf{T}^{\#}$ and $\sigma_n = \sigma^{\#}$ if $n \geq r+1$.*

Proof. As $L \to \infty$, $\mathbf{W}^{(L)} \to \mathbf{0}$ and $\tilde{\mathbf{t}}_{(n)} \to \mathbf{t}_{1(n)}$. Therefore, with infinite snapshots, $\mathbf{t}^{\#}_{(n)} = \mathbf{t}_{1(n)}$ becomes the unique global minimizer of (2.20). If $n \geq r+1$, $\mathcal{T}(\mathbf{t}^{\#}_{(n)})$ is full rank and σ_n is the smallest singular value. The frequencies $\{f_i\}_{i=1}^r$ can be uniquely determined using the MUSIC algorithm since $\mathcal{T}(\mathbf{t}^{\#}_{(n)}) - \sigma_n \mathbf{I}_N$ is of rank r. Given $\{f_i\}_{i=1}^r$, the amplitudes $\{d_i\}_{i=1}^r$ can also be uniquely recovered by least-square since $\mathbf{V}^{\#}_{n \times n'}$ is a Vandermonde matrix with full-column rank. Since the representation (2.2) of $\mathbf{T}$ consists of these same frequencies and amplitudes, Step 3 of the proposed algorithm exactly recovers $\mathbf{T}$. Therefore, with infinite snapshots, we can exactly recover $\mathbf{T}$ when $n \geq r+1$. This implies that the dimension (M) of the compressed covariance matrix satisfies $M = O(\sqrt{r})$, which is order-wise optimal.

2.4.4 Stability Analysis with Finite Noiseless Snapshots

We first analyze the performance of the proposed algorithm as a function of the number of snapshots (L) in absence of additive noise. The stability analysis in [57] is also performed under a similar noiseless setting by only considering the effect of finite snapshots. We divide the total error in N entries of the Toeplitz matrix into two parts (i) estimation error, that represents the error in the first n entries, and (ii) prediction error, which is the error in the last $N-n$ entries. While it is straightforward to bound the estimation error directly using the denoising step, establishing a bound on the prediction error is more involved. We will make novel use of properties of trigonometric polynomials, such as those developed in [59, 141, 142] to formulate an explicit bound on the prediction error. Note that the analysis technique in [59, 141, 142] was developed for Total Variation norm minimization/atomic norm minimization - however, our algorithm *does not*

use any regularizer. In this case, the ideal covariance matrix of $\mathbf{y}_p$ is given by $\mathbf{R_Y} = \mathbf{A}_s\mathbf{T}\mathbf{A}_s^T$, and the estimated covariance matrix (using L samples) is given by $\tilde{\mathbf{R}}_\mathbf{Y} = \frac{1}{L}\sum_{p=1}^{L}\mathbf{y}_p\mathbf{y}_p^T$. Here $\tilde{\mathbf{R}}_\mathbf{Y}$ satisfies (2.18) with $\mathbf{W}^{(L)}$ representing the approximation error due to finite snapshot averaging. As before, using the properties of GNS or sparse ruler, we can extract n appropriate entries of $\tilde{\mathbf{R}}_\mathbf{Y}$ to obtain (2.19). In order to statistically characterize $\mathbf{w}_{(n)}$ in terms of L, we assume that $\mathbf{x}_p, 1 \leq p \leq L$ represent L i.i.d. Gaussian random vectors distributed as $\mathbf{x}_p \sim \mathcal{N}(\mathbf{0}, \mathbf{T})$ and therefore, the measurements $\mathbf{y}_p$ are distributed as $\mathbf{y}_p \sim \mathcal{N}(\mathbf{0}, \mathbf{R_Y})$. We now invoke the following large deviation bound from [129]

Lemma 3 *(Proposition A.3 in [129]) Let $\{\mathbf{y}_p\}_{p=1}^{L}$ be zero mean i.i.d Gaussian random vectors distributed as $\mathbf{y}_p \sim \mathcal{N}(\mathbf{0}, \mathbf{R_Y})$. Then,*

$$\mathbb{P}\left\{\|\mathbf{R_Y} - \tilde{\mathbf{R}}_\mathbf{Y}\|_F \geq \frac{\text{trace}(\mathbf{R_Y})}{\sqrt{L}}\right\} \leq 2e^{-2c_1\sqrt{L}} \tag{2.22}$$

where c_1 is a positive universal constant.

It can be easily verified that $\|\mathbf{w}_{(n)}^{(L)}\|_2 \leq \|\mathbf{R_Y} - \tilde{\mathbf{R}}_\mathbf{Y}\|_F$. Hence, Lemma 17 implies that with probability $1 - O(e^{-2c_1\sqrt{L}})$, the estimation error $\mathbf{w}_{(n)}^{(L)}$ satisfies

$$\|\mathbf{w}_{(n)}^{(L)}\|_2 \leq \frac{Mt_0}{\sqrt{L}} \tag{2.23}$$

where $M = O(\sqrt{n})$.

Firstly, the estimation error associated with the denoising step is given by the following lemma.

Lemma 4 *The solution $\mathbf{t}_{(n)}^{\#}$ to (2.20) satisfies*

$$\|\mathbf{t}_{(n)}^{\#} - \mathbf{t}_{(n)}\|_2 \leq 2\|\mathbf{w}_{(n)}^{(L)}\|_2 \tag{2.24}$$

Proof. Since the true $\mathbf{t}_{(n)}$ is a feasible solution to (2.20), we have $\|\mathbf{t}^{\#}_{(n)} - \mathbf{t}_{(n)}\|_2 \leq 2\|\mathbf{t}_{(n)} - \tilde{\mathbf{t}}_{(n)}\|_2 = 2\|\mathbf{w}^{(L)}_{(n)}\|_2$ which establishes (2.24). Before establishing the bound on prediction error, we will need to define several quantities, and state two lemmas from [59] that will be used in our proof. Notice that $\mathbf{t}_{(n)}$ and $\mathbf{t}^{\#}_{(n)}$ satisfy

$$\mathbf{t}_{(n)} = \int_{-1/2}^{1/2} \mathbf{v}_n(f)\mu(df) \quad \mathbf{t}^{\#}_{(n)} = \int_{-1/2}^{1/2} \mathbf{v}_n(f)\mu^{\#}(df) \tag{2.25}$$

where $\mu, \mu^{\#}$ are positive finite measures given by

$$\begin{aligned} \mu &= \sum_{i=1}^{r} d_i \delta(f - f_i) \\ \mu^{\#} &= \sum_{i=1}^{n'} d^{\#}_i \delta(f - f^{\#}_i) + \frac{\sigma^{\#}}{n} \sum_{i=1}^{n} \delta(f - \frac{(i-1)}{n}) \end{aligned} \tag{2.26}$$

The difference measure is defined by $\nu = \mu^{\#} - \mu$. Let $\rho(\hat{f}_1, \hat{f}_2)$ denote the *wraparound* distance function for distinct frequencies $\hat{f}_1, \hat{f}_2 \in (-1/2, 1/2]$ [141]. Following [59], we define neighborhoods $\mathcal{N}_l$ around each true frequency f_l as $\mathcal{N}_l = \{f \in (-1/2, 1/2] : \rho(f, f_l) \leq 0.16/n\}$ as well as a far region $\mathcal{F} \triangleq (-1/2, 1/2] \setminus \bigcup_{l=1}^{r} \mathcal{N}_l$. Let $\mathcal{P}_{\mathfrak{F}}$ denote the projection of any measure onto the true frequency support $\mathfrak{F} = \{f_1, f_2, \cdots, f_r\}$ and $\|\cdot\|_{TV}$ be the Total Variation (TV) norm [142]. Furthermore, let

$$\begin{aligned} I^l_0 &:= \left| \int_{\mathcal{N}_l} \nu(df) \right| \quad I^l_1 := n \left| \int_{\mathcal{N}_l} (f - f_l)\nu(df) \right| \\ I^l_2 &:= \frac{n^2}{2} \int_{\mathcal{N}_l} (f - f_l)^2 |\nu|(df) \quad I_i := \sum_{l=1}^{r} I^l_i, \text{ for } i = 0, 1, 2 \end{aligned} \tag{2.27}$$

We will also make use of the following two lemmas from [59] in order to derive the desired bound

on the prediction error.

Lemma 5 *(Lemma 3 and Theorem 4 in [59]) Consider the measurement model (2.19) and the representations for* $\mathbf{t}_{(n)}$ *and* $\mathbf{t}^{\#}_{(n)}$ *as given by (2.25). If the true frequencies* $\{f_l\}_{l=1}^r$ *satisfies* $\min_{p\neq q}\rho(f_p,f_q) > 4/n$ *and* $n > 256$*, then there exists a trigonometric polynomial* $Q(f)$ *such that*

$$\begin{aligned}
&\|\mathcal{P}_{\mathfrak{F}}(\nu)\|_{TV} = \int_{-1/2}^{1/2} Q(f)\mathcal{P}_{\mathfrak{F}}(\nu)(df) \\
&|Q(f)| \leq 1 - \frac{C_a}{2}n^2(f-f_l)^2 \quad f \in \mathcal{N}_l, 1 \leq l \leq r \qquad (2.28)\\
&|Q(f)| \leq 1 - C_b \quad f \in \mathcal{F} \\
&\left|\int_{-1/2}^{1/2} Q(f)\nu(df)\right| \leq \frac{C_c r\xi}{n} \qquad (2.29)
\end{aligned}$$

where C_a, C_b, C_c *are positive constants and* $\xi \triangleq \sup_{f\in(-1/2,1/2]} |\langle \mathbf{v}_n(f), \mathbf{t}^{\#}_{(n)} - \mathbf{t}_{(n)}\rangle|$

Lemma 6 *(Lemma 2 in [59]) Consider the measurement model (2.19) and the representations for* $\mathbf{t}_{(n)}$ *and* $\mathbf{t}^{\#}_{(n)}$ *as given by (2.25). If the frequencies* $\{f_l\}_{l=1}^r$ *satisfy* $\min_{p\neq q}\rho(f_p,f_q) > 4/n$ *and* $n > 256$*, then there exist positive constants* $\tilde{c}_1, \tilde{c}_2$ *such that*

$$\begin{aligned}
I_0 &\leq \tilde{c}_1\left(\frac{r\xi}{n} + I_2 + \int_{\mathcal{F}} |\nu|(df)\right) \\
I_1 &\leq \tilde{c}_2\left(\frac{r\xi}{n} + I_2 + \int_{\mathcal{F}} |\nu|(df)\right) \qquad (2.30)
\end{aligned}$$

where $\xi \triangleq \sup_{f\in(-1/2,1/2]} |\langle \mathbf{v}_n(f), \mathbf{t}^{\#}_{(n)} - \mathbf{t}_{(n)}\rangle|$

Notice that Lemma 6 establishes upper bounds on I_0 and I_1 in terms of I_2 and $\int_{\mathcal{F}} |\nu|(df)$. We next prove a key lemma which, in turn, bounds the quantities I_2 and $\int_{\mathcal{F}} |\nu|(df)$ in terms of the measurement noise.

Lemma 7 *Consider the measurement model (2.19) and the representations for* $\mathbf{t}_{(n)}$ *and* $\mathbf{t}^{\#}_{(n)}$ *as*

given by (2.25). If the true frequencies $\{f_l\}_{l=1}^r$ *satisfy* $\min_{p\neq q}\rho(f_p,f_q) > 4/n$ *and* $n > 256$*, then, there exist positive constants* c_1, c_2 *such that*

$$I_2 + \int_{\mathcal{F}} |\nu|(df) \leq c_1\left(\frac{c_2 r\xi}{n} + |t_0^{\#} - t_0|\right) \tag{2.31}$$

where $\xi \triangleq \sup_{f\in(-1/2,1/2]} |\langle \mathbf{v}_n(f), \mathbf{t}_{(n)}^{\#} - \mathbf{t}_{(n)}\rangle|$

Proof. Since $\{f_l\}_{l=1}^r$ satisfy $\min_{p\neq q}\rho(f_p,f_q) > 4/n$, from Lemma 5, there exists a polynomial $Q(f)$ such that

$$\|\mathcal{P}_{\mathfrak{F}}(\nu)\|_{TV} = \int_{-1/2}^{1/2} Q(f)\mathcal{P}_{\mathfrak{F}}(\nu)(df) \tag{2.32}$$

Using triangle inequality and recalling that $\mathcal{P}_{\mathfrak{F}}$ denotes the projection onto $\mathfrak{F}$, we have

$$\begin{aligned}
\|\mathcal{P}_{\mathfrak{F}}(\nu)\|_{TV} &\leq \left|\int_{-1/2}^{1/2} Q(f)\nu(df)\right| + \left|\int_{\mathfrak{F}^c} Q(f)\nu(df)\right| \\
&\leq \frac{C_c r\xi}{n} + \sum_{f_l\in\mathfrak{F}}\left|\int_{\mathcal{N}_l\setminus f_l} Q(f)\nu(df)\right| + \left|\int_{\mathcal{F}} Q(f)\nu(df)\right| \\
&\leq \frac{C_c r\xi}{n} + \sum_{f_l\in\mathfrak{F}}\left(\int_{\mathcal{N}_l\setminus f_l} |\nu|(df) - C_a I_2^l\right) + (1-C_b)\int_{\mathcal{F}}|\nu|(df) \\
&= \frac{C_c r\xi}{n} + \|\mathcal{P}_{\mathfrak{F}^c}(\nu)\|_{TV} - C_a I_2 - C_b\int_{\mathcal{F}}|\nu|(df)
\end{aligned} \tag{2.33}$$

where we have used (2.29) and (2.28) for the second and third inequalities. We also have

$$\begin{aligned}
\|\mu^{\#}\|_{TV} &= \|\mu+\nu\|_{TV} = \|\mu + \mathcal{P}_{\mathfrak{F}}(\nu)\|_{TV} + \|\mathcal{P}_{\mathfrak{F}^c}(\nu)\|_{TV} \\
&\geq \|\mu\|_{TV} - \|\mathcal{P}_{\mathfrak{F}}(\nu)\|_{TV} + \|\mathcal{P}_{\mathfrak{F}^c}(\nu)\|_{TV}
\end{aligned} \tag{2.34}$$

Combining (2.34) and (2.33), we get

$$C_a I_2 + C_b \int_{\mathcal{F}} |\nu|(df) \leq \frac{C_c r \xi}{n} + \|\mu^{\#}\|_{TV} - \|\mu\|_{TV} \tag{2.35}$$

A key observation is to notice that both $\mu, \mu^{\#}$ as given by (2.26) are positive measures. Therefore, we have

$$\|\mu\|_{TV} = t_0, \quad \|\mu^{\#}\|_{TV} = t_0^{\#} \tag{2.36}$$

Using (2.36) in (2.35), we finally obtain

$$\begin{aligned} C_a I_2 + C_b \int_{\mathcal{F}} |\nu|(df) &\leq \frac{C_c r \xi}{n} + t_0^{\#} - t_0 \\ I_2 + \int_{\mathcal{F}} |\nu|(df) &\leq \frac{1}{\min\{C_a, C_b\}} \left(\frac{C_c r \xi}{n} + |t_0^{\#} - t_0| \right) \end{aligned} \tag{2.37}$$

which completes the proof.

Equipped with Lemmas 5, 6 and 7, we present our main result regarding the prediction error in the following theorem:

Theorem 3 *Consider L i.i.d measurements $\{\mathbf{y}_p\}_{p=1}^{L}$ (acquired via GNS or sparse ruler) distributed as $\mathbf{y}_p \sim \mathcal{N}(\mathbf{0}, \mathbf{A}_s \mathbf{T} \mathbf{A}_s^T)$ and let $\tilde{\mathbf{R}}_{\mathbf{Y}} = \frac{1}{L} \sum_{p=1}^{L} \mathbf{y}_p \mathbf{y}_p^T$ be the estimated covariance matrix, from which the observation model (2.19) is derived, where $\mathbf{t}_{(n)}$ has the line spectrum representation (2.25). Let $\mathbf{T}^{\#}$ be the estimate of $\mathbf{T}$ obtained from the algorithm proposed in Table 1 and $t_m^{\#}$ denote the mth entry in the first row (or column) of $\mathbf{T}^{\#}$. Then, with probability $1 - O(e^{-2c_1\sqrt{L}})$*

$$\frac{1}{n} \|\mathbf{t}_{(n)}^{\#} - \mathbf{t}_{(n)}\|_2 \leq \frac{2Mt_0}{n\sqrt{L}} \tag{2.38}$$

Furthermore, if $n > 256$, and if the frequencies $\{f_1, f_2, \cdots, f_r\}$ satisfy $\min_{p \neq q} \rho(f_p, f_q) > 4/n$ then, with probability $1 - O(e^{-2c_1\sqrt{L}})$, the prediction error $|t_m^{\#} - t_m|$, for $n \leq m \leq N-1$, can

be bounded as

$$
\begin{aligned}
&|t_m^{\#} - t_m| \\
&\leq \left(\gamma_1 + \frac{\gamma_2 \pi m}{n} + \frac{\gamma_3 \pi^2 m^2}{n^2}\right)\left(\frac{\gamma_4 r}{\sqrt{n}} + 2\right)\frac{M t_0}{\sqrt{L}}
\end{aligned} \tag{2.39}
$$

where $c_1, \gamma_1, \gamma_2, \gamma_3, \gamma_4$ *are positive constants and* $M = O(\sqrt{n})$.

Proof. The bound (2.38) follows directly from Lemma 4 and (2.23). To establish the bound (2.39) on the prediction error, observe that for any $n \leq m \leq N-1$, we have

$$
\begin{aligned}
&|t_m^{\#} - t_m| = |\int_{-1/2}^{1/2} e^{j2\pi m f} \nu(df)| \\
&\leq \int_{\mathcal{F}} |\nu|(df) + \sum_{l=1}^{r} \left| \int_{\mathcal{N}_l} e^{j2\pi m f} \nu(df) \right|
\end{aligned} \tag{2.40}
$$

where (2.40) follows from triangle inequality. Using Taylor's theorem for each neighborhood N_l around f_l we get,

$$
\begin{aligned}
&\left| \int_{\mathcal{N}_l} e^{j2\pi m f} \nu(df) \right| \leq \left| \int_{\mathcal{N}_l} \nu(df) \right| \\
&+2\pi m \left| \int_{\mathcal{N}_l} (f - f_l)\nu(df) \right| + 2\pi^2 m^2 \int_{\mathcal{N}_l} (f - f_l)^2 |\nu|(df) \\
&= I_0^l + \frac{2\pi m}{n} I_1^l + \frac{4\pi^2 m^2}{n^2} I_2^l
\end{aligned} \tag{2.41}
$$

Therefore, (2.40) can be simplified as

$$
\begin{aligned}
&|t_m^{\#} - t_m| \leq \int_{\mathcal{F}} |\nu|(df) + I_0 + \frac{2\pi m}{n} I_1 + \frac{4\pi^2 m^2}{n^2} I_2 \\
&\leq \left(\tilde{c}_1 + \frac{2\pi m}{n}\tilde{c}_2 + 1\right) \int_{\mathcal{F}} |\nu|(df) \\
&+ \left(\tilde{c}_1 + \frac{2\pi m}{n}\tilde{c}_2 + \frac{4\pi^2 m^2}{n^2}\right) I_2 + (\tilde{c}_1 + \tilde{c}_2 \frac{2\pi m}{n}) \frac{r\xi}{n}
\end{aligned} \tag{2.42}
$$

where the inequality (2.42) follows from Lemma 6. Since $n \leq m$, we further have

$$\begin{aligned}
&|t_m^{\#} - t_m| \\
&\leq \left(\tilde{c}_1 + \frac{2\pi m}{n}\tilde{c}_2 + \frac{4\pi^2 m^2}{n^2}\right)\left(I_2 + \int_{\mathcal{F}} |\nu|(df)\right) \\
&+(\tilde{c}_1 + \tilde{c}_2\frac{2\pi m}{n})\frac{r\xi}{n} \\
&\leq \left(\bar{c}_1 + \frac{\bar{c}_2\pi m}{n} + \frac{\bar{c}_3\pi^2 m^2}{n^2}\right)\left(\frac{\bar{c}_4 r\xi}{n} + |t_0^{\#} - t_0|\right) \qquad (2.43)
\end{aligned}$$

where we used Lemma 7 to derive (2.43) and $\{\bar{c}_i\}_{i=1}^4$ are positive constants. Using Cauchy-Schwartz inequality, we can bound ξ as

$$\xi \leq \|\mathbf{v}_n(f)\|_2 \|\mathbf{w}_{(n)}^{(L)}\|_2 = \sqrt{n}\|\mathbf{w}_{(n)}^{(L)}\|_2 \leq \frac{\sqrt{n}Mt_0}{\sqrt{L}} \qquad (2.44)$$

Finally, from (2.20), it follows that

$$|t_0^{\#} - t_0| \leq \|\mathbf{t}_{(n)}^{\#} - \mathbf{t}_{(n)}\|_2 \leq 2\frac{Mt_0}{\sqrt{L}} \qquad (2.45)$$

The proof completes by substituting (2.44) and (2.45) in (2.43).

Theorem 3 provides a finite-sample bound on the prediction error and shows that as $L \to \infty$, it is possible to perfectly recover $\mathbf{T}$ using the proposed prediction based framework.

Remark 6. *In [57], the authors derived a related bound on the covariance estimation error (as a function of L) by considering a noiseless model, and sampling the entire covariance matrix. This corresponds to the scenario when $n = N$ and no prediction is necessary. Hence, we can compare the error bound in [57] with the bound on estimation error (2.38) from Step 1 of our proposed algorithm. According to [57],*

$$\frac{1}{n}\|\mathbf{t}_{(n)}^{\#} - \mathbf{t}_{(n)}\|_2 \leq c\lambda\sqrt{\frac{r}{n}} \qquad (2.46)$$

where λ is the regularization parameter and c is a universal constant. For large enough L, the parameter λ in Theorem 4 of [57] is lower bounded as

$$\lambda \geq c' \sqrt{M t_0 \|\mathbf{R_Y}\|} \sqrt{\frac{\log(Ln)}{L}} \tag{2.47}$$

where c' is another constant and $\|\cdot\|$ is the operator norm. From (2.46) and (2.47), the tightest upper bound on $\frac{1}{n}\|\mathbf{t}^{\#}_{(n)} - \mathbf{t}_{(n)}\|_2$ is given by

$$\frac{1}{n}\|\mathbf{t}^{\#}_{(n)} - \mathbf{t}_{(n)}\|_2 \leq c_1 \frac{\sqrt{M t_0 \|\mathbf{R_Y}\|}}{\sqrt{nL}} \sqrt{r \log(Ln)} \tag{2.48}$$

Since $M < n$, we have $\frac{M}{n} < \sqrt{\frac{M}{n}}$. In addition, due to the structure of $\mathbf{A}_s$, the diagonal entries of $\mathbf{R_Y}$ are all t_0, and $\|\mathbf{R_Y}\| \geq t_0$. Hence the bound (2.38) is tighter than (2.48). Our numerical experiments will further validate this fact.

2.4.5 Stability Analysis in Presence of Noise

We now consider the effect of additive noise in (2.19). Using Lemma 17, with probability $1 - O(e^{-2c_1\sqrt{L}})$, $\mathbf{w}^{(L)}_{(n)}$ satisfies

$$\|\mathbf{w}^{(L)}_{(n)}\|_2 \leq \frac{M(t_0 + \sigma_n)}{\sqrt{L}} \tag{2.49}$$

where $M = O(\sqrt{n})$.

Theorem 3 can be used to derive noisy finite-snapshot guarantees with some modifications. Notice that $\tilde{\mathbf{t}}_{1(n)}$ from (2.19) now serves as the input for the algorithm proposed in Table 1. The solution $\mathbf{t}^{\#}_{(n)}$ from Step 1 satisfies $\mathcal{T}(\mathbf{t}^{\#}_{(n)}) = \mathbf{V}^{\#}_{n,n'}\mathbf{D}^{\#}(\mathbf{V}^{\#}_{n,n'})^H + \sigma^{\#}\mathbf{I}_n$ where $\mathbf{V}^{\#}_{n,n'}\mathbf{D}^{\#}(\mathbf{V}^{\#}_{n,n'})^H$ is a low rank PSD Toeplitz matrix which can also be verified to be real-valued (since $\mathbf{t}^{\#}_{(n)}$ is real). Let $\hat{\mathbf{t}}^{\#}_{(n)}$ denote the first row (or column) of this matrix, i.e. $\mathcal{T}(\hat{\mathbf{t}}^{\#}_{(n)}) = \mathbf{V}^{\#}_{n,n'}\mathbf{D}^{\#}(\mathbf{V}^{\#}_{n,n'})^H$ Notice

that $\hat{\mathbf{t}}^{\#}_{(n)}$ (and not $\mathbf{t}^{\#}_{(n)}$) is the estimate of $\mathbf{t}_{(n)}$. We can now derive bounds on the observed and prediction error as follows:

$$\begin{aligned} \|\mathbf{t}^{\#}_{(n)} - \mathbf{t}_{1(n)}\|_2 \quad &= \|\hat{\mathbf{t}}^{\#}_{(n)} - \mathbf{t}_{(n)} + (\sigma^{\#} - \sigma_n)\mathbf{e}_{1(n)}\|_2 \\ &\geq \|\hat{\mathbf{t}}^{\#}_{(n)} - \mathbf{t}_{(n)}\|_2 - |\sigma^{\#} - \sigma_n| \end{aligned} \tag{2.50}$$

On the other hand,

$$\begin{aligned} \|\mathbf{t}^{\#}_{(n)} - \mathbf{t}_{1(n)}\|_2 \quad &\leq \|\mathbf{t}^{\#}_{(n)} - \tilde{\mathbf{t}}_{1(n)}\|_2 + \|\tilde{\mathbf{t}}_{1(n)} - \mathbf{t}_{1(n)}\|_2 \\ &\leq 2\|\tilde{\mathbf{t}}_{1(n)} - \mathbf{t}_{1(n)}\|_2 = 2\|\mathbf{w}^{(L)}_{(n)}\|_2 \end{aligned} \tag{2.51}$$

where the second inequality follows from the fact that $\mathbf{t}^{\#}_{(n)}$ is the minimizer of (2.20). Combining (3.68) and (2.51), obtain the following bound on the estimation error (of first n entries):

$$\|\hat{\mathbf{t}}^{\#}_{(n)} - \mathbf{t}_{(n)}\|_2 \leq 2\|\mathbf{w}^{(L)}_{(n)}\|_2 + |\sigma^{\#} - \sigma_n| \tag{2.52}$$

We need some additional notations to derive the prediction error bound. Similar to (2.25), $\hat{\mathbf{t}}^{\#}_{(n)}$ satisfies

$$\hat{\mathbf{t}}^{\#}_{(n)} = \int_{-1/2}^{1/2} \mathbf{v}_n(f)\hat{\mu}^{\#}(df), \quad \hat{\mu}^{\#} = \sum_{i=1}^{n'} d_i^{\#}\delta(f - f_i^{\#}) \tag{2.53}$$

Defining $\hat{\nu} = \hat{\mu}^{\#} - \mu$ as the difference measure, it can be verified that Lemmas 5, 6 and 7 continue to hold, using $\hat{\nu}$ as the difference measure instead of ν, and replacing $\mathbf{t}^{\#}_{(n)}$ and $t_0^{\#}$ with $\hat{\mathbf{t}}^{\#}_{(n)}$ and $\hat{t}_0^{\#}$ (the zeroth entry of $\hat{\mathbf{t}}^{\#}_{(n)}$) respectively. Furthermore we have

$$\begin{aligned} \sup_{f\in(-1/2,1/2]} |\langle \mathbf{v}_n(f), \hat{\mathbf{t}}^{\#}_{(n)} - \mathbf{t}_{(n)}\rangle| &\leq \sqrt{n}\|\hat{\mathbf{t}}^{\#}_{(n)} - \mathbf{t}_{(n)}\|_2 \\ &\leq \sqrt{n}\Big(2\|\mathbf{w}^{(L)}_{(n)}\|_2 + |\sigma^{\#} - \sigma_n|\Big) \end{aligned} \tag{2.54}$$

where the last inequality follows from (2.52). Similarly, we have

$$|\hat{t}_0^{\#} - t_0| \leq \|\hat{\mathbf{t}}_{(n)}^{\#} - \mathbf{t}_{(n)}\|_2 \leq 2\|\mathbf{w}_{(n)}^{(L)}\|_2 + |\sigma^{\#} - \sigma_n| \tag{2.55}$$

Following the same steps as in the proof of Theorem 3 (for deriving (2.43)), and using (2.55) and (2.54), we obtain the following upper bound on prediction error as

$$\begin{aligned} &|\hat{t}_m^{\#} - t_m| \leq \\ &\left(\bar{c}_1 + \frac{\bar{c}_2 \pi m}{n} + \frac{\bar{c}_3 \pi^2 m^2}{n^2}\right)\left(\frac{\bar{c}_4 r}{\sqrt{n}} + 1\right)(2\|\mathbf{w}_{(n)}^{(L)}\|_2 + |\sigma^{\#} - \sigma_n|) \end{aligned} \tag{2.56}$$

Finally, using (2.49), we summarize the bounds on the estimation and prediction error as

$$\begin{aligned} &\|\hat{\mathbf{t}}_{(n)}^{\#} - \mathbf{t}_{(n)}\|_2 \leq \eta \\ &|\hat{t}_m^{\#} - t_m| \leq \left(\bar{c}_1 + \frac{\bar{c}_2 \pi m}{n} + \frac{\bar{c}_3 \pi^2 m^2}{n^2}\right)\left(\frac{\bar{c}_4 r}{\sqrt{n}} + 1\right)\eta, m \geq n \end{aligned} \tag{2.57}$$

where $\eta \triangleq 2\|\mathbf{w}_{(n)}^{(L)}\|_2 + |\sigma^{\#} - \sigma_n|$ and with probability $1 - O(e^{-2c_1\sqrt{L}})$, $\mathbf{w}_{(n)}^{(L)}$ satisfies (2.49).

Remark7. *Unlike the result in Theorem 3, the error bound (2.57) is* implicit *since it is itself a function of the error $|\sigma^{\#} - \sigma_n|$ in estimating the noise power σ_n. It is non-trivial to explicitly express this error as a function of finite snapshots, which is beyond the scope of this paper. However, in the limit $L \to \infty$, since $\tilde{\mathbf{R}}_{\mathbf{Y}} \to \mathbf{A}_s \mathbf{T} \mathbf{A}_s^T + \sigma_n \mathbf{I}_M$, following our previous argument, it holds that $\mathbf{t}_{1(n)}$ will be the unique minimizer of (2.20), implying $\sigma^{\#} = \sigma_n$. Hence as $L \to \infty$, both prediction and estimation error will become zero, and it will be possible to exactly recover $\mathbf{T}$.*

2.4.6 Advantages of the Proposed Algorithm

Compared to recently proposed compressive Toeplitz covariance estimation algorithms [38,57] that are based on nuclear norm minimization framework, or algorithms derived from Maximum Likelihood (ML) estimation [48, 121], the proposed algorithm enjoys unique advantages which are discussed as follows:

1. **Absence of Regularization Parameter:** The PSD Toeplitz constraint in (2.20) ensures stable reconstruction using a simple least squares denoising (a fact also noted in [70, 71]), without the need for any regularizer. In contrast, the method proposed in [38] requires knowledge of noise power (or an upper bound) while [57] needs a tuning parameter for atomic-norm based regularization. Our algorithm can be especially advantageous over these techniques since tuning parameters are often sensitive to uncertainties in our knowledge of noise power. Although the ML based methods [121] also do not require any regularization parameter, their performance can only be analyzed in the asymptotic regime. In contrast, the analysis framework for our method permits non-asymptotic theoretical analysis of the reconstruction error, even for non-Gaussian models.

2. **Stable Reconstruction with Near-minimal Sample Size:** Theorem 2 shows that in absence of noise, the proposed algorithm only requires $n \geq r+1$ samples for perfect recovery of $\mathbf{T}$. In presence of noise, the number of measurements required for stable reconstruction is indirectly provided by the separation condition in Theorem 3. In particular, in order to satisfy the separation condition, we need $n > 4r$ measurements, and this implies that the size of the sketch $\tilde{\mathbf{R}}_{\mathbf{Y}}$ is $O(\sqrt{r}) \times O(\sqrt{r})$, which *does not depend on the ambient dimension* N. On the other hand, the random sampling based approach in [38] requires $O(r\text{polylog}N)$ samples, whereas the structured sampling based approach in [57] and the ML based approach using sparse arrays in [121] use a sketch of size $O(\sqrt{N}) \times O(\sqrt{N})$ to recover a Toeplitz PSD matrix of size $N \times N$ (which is sub-optimal when $r \ll N$).

3. **Low Computational Complexity:** The algorithms in [38, 57] as well as the ML based technique in [121] attempt to recover the entire $N \times N$ matrix $\mathbf{T}$ by solving a single Semidefinite Program (SDP), whose problem size (and the computational complexity) scales with N. In contrast, we only need to solve an SDP (2.20) with n variables, and perform an eigenvalue decomposition (Step 2 in Table 1) on a matrix of size $n \times n$. The complexity of both steps scale only with $n = O(r)$, and is independent of the ambient dimension N. Hence our approach requires fewer measurements and has lower complexity (especially when $r \ll N$) than those proposed in [38, 57] and [121].

4. **Theoretical Guarantees with Finite Samples:** One of our key contributions over the ML based algorithms in [121] is that we are able to analytically characterize the reconstruction error and provide explicit upper bounds that continue to hold for a finite number (L) of samples. On the other hand, the algorithm in [48, 121] is shown to be asymptotically equivalent to Maximum Likelihood Estimation for Gaussian signal models. Its performance in presence of finite snapshots has not been analyzed. Furthermore, the algorithms in [48, 121] are all derived by assuming a Gaussian measurement model, whereas our results in Theorem 3 hold for any distribution on the signal as long as $\mathbf{w}_{(n)}^{(L)}$ can be upper bounded.

5. **Separation Condition and Prediction v/s Estimation Error:** Theorem 2 dictates that the proposed algorithm can exactly reconstruct a low rank Toeplitz PSD $\mathbf{T}$ using $n \geq r + 1$ measurements *without requiring a separation condition on the frequenices*. This is because we exploit the positive semidefiniteness of $\mathbf{T}$ which, in turn, dictates that the coefficients d_i are positive. The fact that the "separation condition" can be avoided when the coefficients d_i in the line spectrum are positive has been noted in [58, 61–63] and our algorithm further corroborates this observation. In presence of noise, we have established different bounds for the estimation error (given in (2.38)) and prediction error (given by (2.39)). The former is obtained by directly using triangle inequality, whereas the latter result is the first of

its kind. Another important distinction between the two bounds is that (2.38) does not require a separation condition, whereas it is needed for establishing (2.39).The reason is that for prediction, we need to estimate the frequencies that parameterize $\mathbf{T}$ (which is not necessary for just denoising the observed entries). Existing results in noisy line spectrum estimation [56, 58, 59, 141, 142] also seem to require a "separation condition" for developing error bounds on the estimated frequencies. In [62], similar results as in [59] have been obtained without explicitly assuming separation condition, but requiring the dual polynomial to satisfy a so-called *Quadratic Isolation Condition* (QIC). Another closely related idea is that of Rayleigh regularity [61] which does not lead to a strict separation condition on the frequencies. It is however, non trivial to extend this analysis for bounding the error in frequency estimates. Since it is presently unclear what kind of separation is fundamentally necessary for frequency estimation in presence of noise, in this paper, we still assume the specific form of the separation condition as used in [58, 59, 142], and leave the general case as an open problem for future research.

2.5 ML Methods for Localizing More Sources than Sensors

In previous sections, we have studied the problem of compressive Toeplitz covariance matrix estimation using GNS. Because of the Vandermonde decomposition (2.2), recovering low-rank Toeplitz covariance matrix can be related to the problem of direction-of-arrival (DOA) estimation in array signal processing. In particular, the frequencies correspond to the incoming directions of the sources and the diagonal entries of $\mathbf{D}$ represent the signal powers. In the next, we will formally establish this relationship, and show more sources than sensors can be detected using GNS.

2.5.1 Signal Model, Direct MUSIC and Co-Array MUSIC

Consider a linear array of M antennas where $z_i\lambda/2, i = 1, 2, \cdots M$ denotes the distance of the ith antenna from the the origin. Here z_i is an integer and λ is the carrier wavelength of narrowband signals received by the array. The antenna receives $D^\star$ narrowband sources simultaneously from directions $\mathbf{\Psi} = [\psi_1, \psi_2, \cdots, \psi_{D^\star}]$. The k th time sample of the received signal at the M antennas can be expressed as

$$\mathbf{y}[k] = \mathbf{A}(\boldsymbol{\theta}^\star)\mathbf{s}[k] + \boldsymbol{n}[k] \tag{2.58}$$

Here $\mathbf{A}(\boldsymbol{\theta}^\star) \in \mathbb{C}^{M\times D^\star}$ is the array manifold matrix given by $[\mathbf{A}(\boldsymbol{\theta}^\star)]_{m,n} = e^{j2\pi z_m\theta_n^\star} = e^{j\pi z_m \sin\psi_n}$, $\mathbf{s}[k] \in \mathbb{C}^{D^\star}$ denotes the vector of zero-mean wide sense stationary (WSS) source signals with a *diagonal covariance matrix* $\mathbf{R}_{ss} = \text{diag}(p_1, p_2, \cdots, p_{D^\star})$ and $\mathbf{n}[k]$ denotes the additive white noise with power σ_n^2, uncorrelated with $\mathbf{s}[k]$. The covariance matrix of the received signal is given by

$$\mathbf{R}_{yy} = E(\mathbf{y}[k]\mathbf{y}^H[k]) = \underbrace{\mathbf{A}(\boldsymbol{\theta}^\star)\mathbf{R}_{ss}\mathbf{A}^H(\boldsymbol{\theta}^\star)}_{\mathbf{R}_{sig}} + \sigma_n^2\mathbf{I}$$

A central problem in array processing is to estimate $\boldsymbol{\theta}^\star$ given time samples of the vector $\mathbf{y}[k]$. Traditional DOA estimation algorithms operate in the regime $D^\star < M$, in which case the matrix $\mathbf{R}_{sig} \in \mathbb{C}^{M\times M}$ is low-rank (rank $D^\star$). Classical subspace-based methods (such as MUSIC) identify the null-space of $\mathbf{R}_{sig}$ directly from the data covariance matrix $\mathbf{R}_{yy}$ and exploit its structure to estimate the DOAs. This has been referred to as "direct MUSIC" in recent literature, in order to distinguish it from "Co-array MUSIC" which we review next.

Given a set of antenna locations (normalized with respect to $\lambda/2$) $\{z_i, i = 1, \cdots, M\}$, its "difference co-array" is the set S_{ca} of all pairwise differences: $S_{ca} = \{z_i - z_j, 1 \le i, j \le M\}$. For certain sparse arrays such as the nested array, S_{ca} contains $M_{ca} \triangleq M^2/4 + M/2$

distinct consecutive integers (and their negatives). The difference co-array of the nested array is therefore a uniform linear array (ULA) with $O(M^2)$ elements. Let $\mathbf{A}_{nested}(\boldsymbol{\theta}^\star) \in \mathbb{C}^{M\times D^\star}$ and $\mathbf{A}_{ca}(\boldsymbol{\theta}^\star) \in \mathbb{C}^{M_{ca}\times D^\star}$ respectively denote the array manifolds corresponding to the nested array and the positive half of its difference co-array, where $[\mathbf{A}_{ca}(\boldsymbol{\theta}^\star)]_{m,n} = e^{j2\pi m\theta_n^\star} = e^{j\pi m \sin\psi_n}, 1 \leq m \leq M_{ca}, 1 \leq n \leq D^\star$. Then, it can be easily verified that [10, 45]

$$\mathbf{A}_{nested}(\boldsymbol{\theta}^\star) = \mathbf{S}\mathbf{A}_{ca}(\boldsymbol{\theta}^\star) \tag{2.59}$$

where $\mathbf{S} \in \mathbb{R}^{M\times M_{ca}}$ is a row-selection matrix

$$[\mathbf{S}]_{i,j} = \begin{cases} 1 & 1 \leq i = j \leq M/2 \\ 1 & \frac{M}{2}+1 \leq i \leq M, j = (\frac{M}{2}+1)(i-\frac{M}{2}) \\ 0 & \text{otherwise} \end{cases}$$

The covariance matrix $\mathbf{R}_{yy}$ of the signals received at a nested array satisfies

$$\mathbf{R}_{yy} = \mathbf{S}\Big(\underbrace{\mathbf{A}_{ca}(\boldsymbol{\theta}^\star)\mathbf{R}_{ss}\mathbf{A}_{ca}^H(\boldsymbol{\theta}^\star)}_{\mathbf{T}_{ca}} + \sigma_n^2\mathbf{I}\Big)\mathbf{S}^T \tag{2.60}$$

Here $\mathbf{T}_{ca}$ represents the (noiseless) covariance matrix of the signals, *as though it was received at the virtual difference co-array* with M_{ca} elements. Since S_{ca} is a ULA, $\mathbf{T}_{ca}$ is Toeplitz structured. As long as $D^\star < M_{ca}$, $\mathbf{T}_{ca}$ is rank-deficient and the DOAs $\boldsymbol{\theta}^\star$ can be uniquely identified by using the MUSIC algorithm on $\mathbf{T}_{ca}$. This is also known as "Co-array MUSIC " [130, 131].

In practice, however, we do not have access to $\mathbf{R}_{yy}$ and can only estimate it from a finite number (say, L) snapshots as $\hat{\mathbf{R}}_{yy} = \frac{1}{L}\sum_{k=1}^{L}\mathbf{y}[k]\mathbf{y}^H[k]$. Obtaining a reliable estimate of $\mathbf{T}_{ca}$ from $\hat{\mathbf{R}}_{yy}$ becomes a challenging task, especially when $D^\star > M$ [131]. In recent times, nuclear norm and atomic-norm based regularized least square algorithms have been proposed to estimate $\mathbf{T}_{ca}$ from $\hat{\mathbf{R}}_{yy}$ [57, 59]. In this paper, we focus on certain constrained Maximum Likelihood (ML)

methods for estimating $\mathbf{T}_{ca}$ from $\hat{\mathbf{R}}_{yy}$. Unlike [57, 59], this framework is regularizer-free, and we show that it provably produces the ML estimate of the DOAs $\boldsymbol{\theta}^\star$ as well *without assuming the number of sources $D^\star$ to be known*.

2.5.2 Constrained Maximum Likelihood Co-Array Covariance and DOA estimation with Finite Snapshots

Under the Gaussian stochastic signal model [130], the L snapshots $\mathbf{y}[k], 1 \leq k \leq L$ of the signal received at a nested array with M antennas are assumed to be i.i.d random vectors distributed as

$$\mathbf{y}[k] \sim \mathcal{N}\Big(\mathbf{0}, \mathbf{A}_{nested}(\boldsymbol{\theta}^\star)\mathbf{R}_{ss}\mathbf{A}^H_{nested}(\boldsymbol{\theta}^\star) + \sigma_n^2\mathbf{I}\Big) \tag{2.61}$$

Since $\mathbf{T}_{ca}$ is a positive semidefinite (PSD) Toeplitz matrix, we consider the following constrained Maximum Likelihood problem for estimating $\mathbf{T}_{ca}$

$$\min_{\mathbf{T}\in\mathbb{C}^{M_{ca}\times M_{ca}}} \mathcal{L}(\mathbf{T}) \quad \text{s.t } \mathbf{T} \succeq \mathbf{0}, \quad \mathbf{T} \text{ is Toeplitz} \tag{2.62}$$

and $\mathcal{L}(\mathbf{T}) = \log\det(\mathbf{S}\mathbf{T}\mathbf{S}^T) + \text{Trace}\Big((\mathbf{S}\mathbf{T}\mathbf{S}^T)^{-1}\hat{\mathbf{R}}_{yy}\Big)$. Let $\hat{\mathbf{T}}$ be a global optimum of (2.62). Then $\hat{\mathbf{T}}$ is guaranteed to be a PSD Toeplitz matrix, and therefore it permits the following Vandermonde Decomposition, owing to Carathéodory's Lemma [59, 68]

$$\hat{\mathbf{T}} = \mathbf{A}_{ca}(\boldsymbol{\theta'})\text{diag}(\mathbf{p}')\mathbf{A}^H_{ca}(\boldsymbol{\theta'}) + \sigma'\mathbf{I} \tag{2.63}$$

Here $\mathbf{A}_{ca}(\boldsymbol{\theta'}) \in \mathbb{C}^{M_{ca}\times D'}$ is a *tall* Vandermonde matrix satisfying $D' < M_{ca}$ (with elements given by $[\mathbf{A}_{ca}(\boldsymbol{\theta'})]_{m,n} = e^{j2\pi m\theta'_n}$), and $\sigma' \geq 0$. The representation (2.63) is unique in $(D', \boldsymbol{\theta'}, \mathbf{p}', \sigma')$ [68]. Moreover, since $D' < M_{ca}$, $\mathbf{A}_{ca}(\boldsymbol{\theta'})$ is column-rank deficient and $\boldsymbol{\theta}'$ can be *uniquely identified* using the MUSIC algorithm [149] on $\hat{\mathbf{T}}_{ca} = \hat{\mathbf{T}} - \sigma'\mathbf{I}$. We denote the overall framework

for ML estimation of $\mathbf{T}_{ca}$, followed by MUSIC, as "Co-array ML-MUSIC" or (Co-MLM).

Since $D' < M_{ca}$ and M_{ca} can be as large as $O(M^2)$, it is possible to estimate as many as $O(M^2)$ sources with only M physical sensors. We now show that $\boldsymbol{\theta}'$ also serves as the Maximum Likelihood (ML) estimate of $\boldsymbol{\theta}^\star$ when the number of sources is considered to be an unknown integer-valued parameter in the ML objective.

2.5.3 ML DOA estimation with Unknown Source Number

Under the stochastic signal model (2.58), Maximum Likelihood estimate of the DOAs are obtained by solving the following optimization problem

$$\min_{\boldsymbol{\theta}\in(-1/2,1/2]^{D^\star},\mathbf{p}\in\mathbb{R}^{D^\star}_{++},\sigma\geq 0} \mathcal{L}_{ML}(\boldsymbol{\theta},\mathbf{p},\sigma) \tag{2.64}$$

where

$$\begin{aligned}\mathcal{L}_{ML}(\boldsymbol{\theta},\mathbf{p},\sigma) = \log\det(&\mathbf{A}_{nested}(\boldsymbol{\theta})\text{diag}(\mathbf{p})\mathbf{A}^H_{nested}(\boldsymbol{\theta})\\ &+\sigma\mathbf{I}) + \text{Trace}\Big((\mathbf{A}_{nested}(\boldsymbol{\theta})\text{diag}(\mathbf{p})\mathbf{A}^H_{nested}(\boldsymbol{\theta}) +\\ &\sigma\mathbf{I})^{-1}\hat{\mathbf{R}}_{yy}\Big)\end{aligned} \tag{2.65}$$

Comparing (2.62), (2.60) and (2.64), we obtain

$$\mathcal{L}_{ML}(\boldsymbol{\theta},\mathbf{p},\sigma) = \mathcal{L}(\mathbf{T})_{\mathbf{T}=\mathbf{A}_{ca}(\boldsymbol{\theta})\text{diag}(\mathbf{p})\mathbf{A}^H_{ca}(\boldsymbol{\theta})+\sigma\mathbf{I}} \tag{2.66}$$

Note that in (2.64), the number of sources $D^\star$ *has been assumed to be known*. This is a common practice in traditional ML estimation, where the model order (or the number of sources) is either assumed to be known, or estimated using the AIC or MDL criteria [132, 133].

When the number of sources ($D^\star$) is unknown, one may treat it as an unknown (integer-valued) parameter D and *jointly minimize the negative log likelihood function* (2.65) with respect

to D and the other parameters:

$$\min_{\substack{D\in\mathbb{Z}^+,\\ 1\le D\le M_{ca}-1}} \min_{\substack{\boldsymbol{\theta}\in(-1/2,1/2]^D,\\ \mathbf{p}\in\mathbb{R}^D_{++},\sigma\ge 0}} \mathcal{L}_{ML}(D,\boldsymbol{\theta},\mathbf{p},\sigma) \tag{2.67}$$

Solving (2.67) with respect to D, $\boldsymbol{\theta}$, $\mathbf{p}$ and σ is equivalent to *jointly finding the maximum likelihood estimates for the number of sources, and the DOAs*. Notice that the number of sources in (2.67) is allowed to be larger than M. At first glance, (2.67) may appear to be a mixed discrete-continuous optimization problem (since D is a positive integer and $\boldsymbol{\theta}, \mathbf{p}$ and σ are real valued parameters). However, the following theorem establishes a direct connection between the ML DOA estimate obtained from (2.67) and $\boldsymbol{\theta'}$ obtained from (Co-MLM) algorithm. For the remainder of the paper, we will assume that $\hat{\mathbf{R}}_{yy}$ is full rank which is valid when L is large enough. [3]

Theorem 4 *Assume that the empirical covariance matrix $\hat{\mathbf{R}}_{yy}$ is full rank. Let $\hat{\mathbf{T}}$ be a global minimum of (2.62) and $(D', \boldsymbol{\theta'}, \mathbf{p}', \sigma')$ be the parameters associated with its Carathéodory representation (2.63), obtained from the (Co-MLM) algorithm. Then $(D', \boldsymbol{\theta'}, \mathbf{p}', \sigma')$ represents a global minimum of the ML problem (2.67).*

Proof. We prove by contradiction. If $(D', \boldsymbol{\theta'}, \mathbf{p}', \sigma')$ does not represent a global minimum of (2.67), there exists $(\tilde{D}, \tilde{\boldsymbol{\theta}}, \tilde{\mathbf{p}}, \tilde{\sigma}) \neq (D', \boldsymbol{\theta'}, \mathbf{p}', \sigma')$ such that

$$\mathcal{L}_{ML}(\tilde{D}, \tilde{\boldsymbol{\theta}}, \tilde{\mathbf{p}}, \tilde{\sigma}) < \mathcal{L}_{ML}(D', \boldsymbol{\theta'}, \mathbf{p}', \sigma') \tag{2.68}$$

Using (2.63), construct a PSD Toeplitz matrix $\tilde{\mathbf{T}} = \mathbf{A}_{ca}(\tilde{\boldsymbol{\theta}})\text{diag}(\tilde{\mathbf{p}})\mathbf{A}_{ca}^H(\tilde{\boldsymbol{\theta}}) + \tilde{\sigma}\mathbf{I}$. Then, (2.68), (2.59) and (2.66) imply that $\mathcal{L}(\tilde{\mathbf{T}}) < \mathcal{L}(\hat{\mathbf{T}})$. Since $\tilde{\mathbf{T}}$ is feasible for (2.62), this contradicts the claim that $\hat{\mathbf{T}}$ is a global minimum of (2.62), and proves the theorem. Theorem 4 establishes that the solution of (Co-MLM) is guaranteed to be a global minimum of the ML problem (2.67).

[3] This assumption will imply that the unconstrained ML objective $\mathcal{L}_{gen}(\mathbf{R}) = \log\det(\mathbf{R}) + \text{Trace}(\mathbf{R}^{-1}\hat{\mathbf{R}}_{yy})$ attains a finite global minimum value over the set of positive semidefinite matrices $\mathbf{R}$ [134]. Hence, the objective functions of (2.62) and (2.67) remain finite at their respective global minima, and we can compare their values.

This is true for every finite value of L (such that $\hat{\mathbf{R}}_{yy}$ is full rank), as well as when the number of sources exceeds the number of sensors, i.e. $D^\star > M$.

2.5.4 Exact Recovery with Orthogonal Source Waveforms

We have shown that D' and $\boldsymbol{\theta}'$ obtained from the (Co-MLM) algorithm are also the ML estimates of $D^\star$ and $\boldsymbol{\theta}^\star$. However, with finite number (L) of snapshots, usually $D' \neq D^\star$ and $\boldsymbol{\theta}' \neq \boldsymbol{\theta}^\star$. We will now show that under certain conditions on the source waveforms, it is possible to ensure exact recovery of the source DOAs, i.e. $D' = D^\star$ and $\boldsymbol{\theta}' = \boldsymbol{\theta}^\star$ even with finite number of snapshots. [4] We make the following assumptions:

1. (A1) The measurements are noiseless, i.e. $\mathbf{n} = \mathbf{0}$ in (2.58), and the number ($D^\star$) of sources satisfies $M \leq D^\star < M_{ca}$.

2. (A2) The source signals are orthogonal [135] , i.e. for $\mathbf{p}^\star \in \mathbb{R}^{D^\star}_{++}$,

$$\frac{1}{L}\mathbf{X}_s\mathbf{X}_s{}^H = \text{diag}(\mathbf{p}^\star), \mathbf{X}_s \triangleq [\mathbf{s}[1], \mathbf{s}[2], \cdots, \mathbf{s}[L]]$$

Theorem 5 *Under assumptions (A1-A2), there exists a unique global minimum $\hat{\mathbf{T}}$ of (2.62), and the parameters $(D', \boldsymbol{\theta}', \mathbf{p}', \sigma')$ associated with its Carathéodory Representation (2.63) satisfy $D' = D^\star, \boldsymbol{\theta}' = \boldsymbol{\theta}^\star, \mathbf{p}' = \mathbf{p}^\star$.*

Proof. Recall that $\hat{\mathbf{R}}_{yy}$ is assumed to be full rank, implying that it is the unique minimum of $\min_{\mathbf{R}\succeq 0} \mathcal{L}_{gen}(\mathbf{R}) = \log\det(\mathbf{R}) + \text{Trace}(\mathbf{R}^{-1}\hat{\mathbf{R}}_{yy})$ [134, 136]. Moreover, due to (A1), $\hat{\mathbf{R}}_{yy} = \mathbf{S}\mathbf{T}^\star\mathbf{S}^T$ where $\mathbf{T}^\star$ is a Toeplitz matrix:

$$\mathbf{T}^\star = \mathbf{A}_{ca}(\boldsymbol{\theta}^\star)\text{diag}(\mathbf{p}^\star)\mathbf{A}_{ca}^H(\boldsymbol{\theta}^\star) \tag{2.69}$$

[4] A related result on recovering more sources than sensors was reported in the context of SBL by using sparse representation over a known dictionary and analyzing the derivatives at the global minima. However, our framework does not use a known dictionary (i.e. $\mathbf{A}_{ca}(\boldsymbol{\theta}^\star)$ contains unknown parameters) or sparsity; rather we exploit the Toeplitz structure and Vandermonde decomposition imposed by the nested array geometry.

Comparing $\mathcal{L}_{gen}(\mathbf{R})$ and (2.62), we can see that $\mathbf{T}^\star$ is also a global minimum of (2.62). Since $\hat{\mathbf{R}}_{yy}$ is the unique global minimum of $\mathcal{L}_{gen}(\mathbf{R})$, it follows that any other global minimum $\mathbf{T}^*$ of (2.62) must also satisfy $\hat{\mathbf{R}}_{yy} = \mathbf{S}\mathbf{T}^*\mathbf{S}^T$. Due to the properties of nested arrays [10, 45], for any Toeplitz matrices $\mathbf{T}_1$ and $\mathbf{T}_2$

$$\mathbf{S}\mathbf{T}_1\mathbf{S}^T = \mathbf{S}\mathbf{T}_2\mathbf{S}^T \Rightarrow \mathbf{T}_1 = \mathbf{T}_2 \tag{2.70}$$

This implies that $\mathbf{T}^* = \mathbf{T}^\star$ and (2.62) has a unique global minimum at $\mathbf{T} = \mathbf{T}^\star$. Finally, since $D^\star < M_{ca}$, the Carathéodory representation of $\mathbf{T}^\star$ is unique and given by (2.69), implying that $D' = D^\star, \boldsymbol{\theta}' = \boldsymbol{\theta}^\star, \mathbf{p}' = \mathbf{p}^\star$.

Remark 8. *Theorem 5 shows that in spite of being highly non-convex, (Co-MLM) algorithm can exactly recover the true DOAs if the source waveforms are orthogonal, even when the number of sources* is unknown and exceeds the number of sensors.

2.6 Numerical Results

In this section, we show simulation results supporting the practical applicability of the proposed compression and recovery algorithms for low rank PSD Toeplitz matrices, and numerically compare its performance with related works in [38, 57, 121]. For given rank $r \leq N$, we generate the real PSD Toeplitz matrix by invoking the Vandermonde decomposition (2.2). The frequencies and amplitudes are generated randomly while satisfying the separation condition. In particular, for a given r, we generate r equispaced frequencies on $[-1/2, 1/2]$ (ensuring that both positive and negative frequencies occur in pair so that $\mathbf{T}$ is real). The amplitudes d_i are generated as i.i.d random variables following a uniform distribution on [0,1]. The same amplitude d_i is used for both f_i and $-f_i$. We further normalize the entries of $\mathbf{T}$ so that the diagonal entries satisfy $t_0 = 1$. The recovered matrix is denoted as $\mathbf{T}^\#$. We use the GNS matrix $\mathbf{A}_s$ defined in (2.9) to

compress $\mathbf{T}$. All simulations are executed using the CVX Toolbox for MATLAB.

2.6.1 Phase Transition with Infinite Noiseless Snapshots

Fig. 2.2 shows the noiseless phase transition plot of the probability of successfully recovering $\mathbf{T}$ from its compressed sketch, for different choices of the rank r and the sampled size n. As a reference, we also show the theoretical lower bound and it is obvious that the simulation results agree with this bound perfectly. In particular, GNS based compression, along with the proposed reconstruction algorithm can perfectly recover $\mathbf{T}$ as soon as $n \geq r + 1$. The phase transition exhibits *a perfectly linear behavior, which is in agreement with the fundamental compression limit of rank r Toeplitz matrices* .

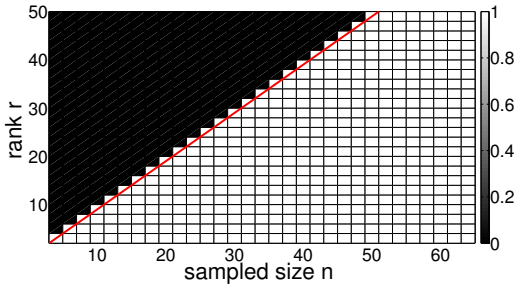


Figure 2.2: Phase transition plot for the proposed GNS based compression and MUSIC based reconstruction of $\mathbf{T}$. A trial is declared successful if $\|\mathbf{t}-\mathbf{t}^{\#}\|_2/N \leq 0.001$. White cells indicate success while black denote failure. The red line represents $n = r + 1$ and the result is averaged over 50 runs. $N = 113$.

We also compare the results with random sampling based compression and recovery of Toeplitz matrices, as proposed in [38]. The sampling model for our method is different from that in [38] (in particular, [38] uses rank 1 measurement matrices). For fairness of comparison, we fix the value of n and simulate the measurement model in [38] by collecting n measurements. This ensures that the reconstruction algorithms for *both approaches use the compressed sketch of same size*. Fig. 2.3 shows the noiseless phase transition for the nucelar norm based recovery algorithm from [38]. Comparing Fig. 2.2 and Fig. 2.3, it is obvious that the proposed method has tighter

transition boundary and larger success region. The underlying reason for this difference is that we transform the matrix completion problem into spectrum detection problem and the Vandermonde decomposition theorem gives us deterministic guarantees with minimum possible measurement size, thereby leading to the sharp phase transition. The non-linear shape of the transition region in Fig. 2.3 is due to the nature of random sampling used in [38], for which the number of required measurements needed for a given r is strictly larger than that for our method.

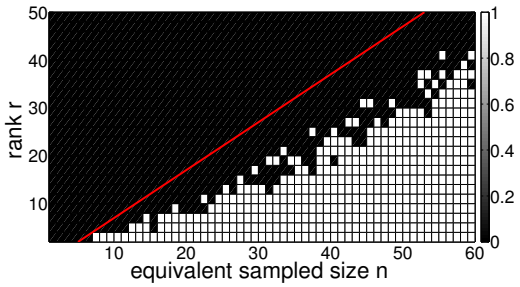


Figure 2.3: Phase transition plot for method in [38] and nuclear-norm minimization based reconstruction of $\mathbf{T}$. A trial is declared successful if $\|\mathbf{t}-\mathbf{t}^{\#}\|_2/N \leq 0.001$. White cells indicate success while black denote failure. The red line represents $n = r + 3$ and the result is averaged over 100 runs. $N = 113$.

2.6.2 Study of Prediction and Total Error

We next evaluate the performance of the proposed method in presence as a function of number measurement size (n), SNR, number of snapshots, and compare them with related works in [38, 57]. In particular, we compare the following algorithms:

- *proposed*: This refers to the proposed reconstruction algorithm described in Table 1.
- *nuclear-psd*: This refers to the nuclear norm minimization framework in [57]

$$\begin{aligned} \mathbf{t}^{\#} &= \arg\min_{\mathbf{t}\in\mathbb{R}^N} \frac{1}{2}\|\tilde{\mathbf{R}}_{\mathbf{Y}} - \mathcal{A}_{\text{det}}(\mathcal{T}(\mathbf{t}))\|_F^2 + \lambda \text{Trace}(\mathcal{T}(\mathbf{t})) \\ s.t. \quad & \mathcal{T}(\mathbf{t}) \succeq 0 \end{aligned} \tag{2.71}$$

where $\mathcal{A}_{\text{det}}(\cdot)$ represents a deterministic sampler and $\tilde{\mathbf{R}} \in \mathbb{R}^{M\times M}$ represents the compressed sample covariance matrix of size $M \times M$. In [57], $\mathcal{A}_{\text{det}}(\cdot)$ corresponded to a sparse ruler that sampled the *entire covariance matrix*, implying $M = O(\sqrt{N})$ and there was no need for prediction. In this paper, we will use a GNS instead of sparse ruler (so that the proposed algorithm and (2.71) have the same sampling operator) and consider different values for M under different experimental settings.

- *random*: This represents the covariance compression/reconstruction framework of [38] using random samplers. In particular, the algorithm in [38] solves the following convex problem:

$$\begin{aligned} \mathbf{T}^{\#}_{(N)} = \arg \min_{\mathbf{T}\in\mathbb{R}^{N\times N}} \text{Trace}(\mathbf{T}) \\ s.t. \quad \mathbf{T} \succeq 0, \text{ Toeplitz} \\ \|\mathbf{z} - \mathcal{A}_{\text{random}}(\mathbf{T})\|_2 \leq \varepsilon \end{aligned} \tag{2.72}$$

  where $\mathbf{z} \in \mathbb{R}^{M_r}$ denotes M_r measurements of the high dimensional covariance matrix $\mathbf{T}$ acquired using so-called rank-1 *random measurements* as proposed in [38]. The corresponding sampling operator is denoted by $\mathcal{A}_{\text{random}}(\cdot)$. Notice that the proposed algorithm and (2.71) use the same measurement model which samples a certain number of entries of $\mathbf{T}$. On the other hand, the measurement model in (2.72) is different, where each measurement contains a random linear combination of all the entries of $\mathbf{T}$. However, for fair comparison, in this subsection, we provide the *same measurement size* to our algorithm and [38]. In particular, this implies $n = M_r$ (recall that n denotes the number of distinct entries of $\mathbf{T}$ that are observed using GNS based measurement model).

In this subsection, we compare the performance of our algorithm against (2.71) and (2.72) under two different settings. In the first setting, we assume that the acquired sketch $\tilde{\mathbf{R}}_{\mathbf{Y}}$ is a perturbed

version of the error-free measurements, i.e.

$$\tilde{\mathbf{R}}_{\mathbf{Y}} = \mathbf{A}_s \mathbf{T} \mathbf{A}_s^T + \mathbf{W} \tag{2.73}$$

where $\mathbf{W}$ represents a perturbation matrix with bounded entries. This measurement model with bounded perturbations was used in the numerical experiments of [38], and we consider a similar setting to be consistent with [38]. The regularization parameter λ for [57] is found through exhaustive search to ensure best performance. Similarly, in (2.72), we assume knowledge of the upper bound on the perturbation errors. In contrast, our algorithm *does not need* any regularization parameter nor the knowledge of the perturbation error bound.

In the second setting, we consider the measurement model (2.18) where $\mathbf{W}^{(L)}$ represents the noisy finite snapshot estimation error, with L denoting the number of snapshots, and σ_n representing the noise power.

Setting 1: Perturbed Measurements

Under the perturbed measurement model, as before, we can extract n entries $\tilde{\mathbf{t}}_{(n)}$ from $\tilde{\mathbf{R}}_{\mathbf{Y}}$ such that $\tilde{\mathbf{t}}_{(n)} = \mathbf{t}_{(n)} + \mathbf{w}_{(n)}$ where $\mathbf{w}_{(n)}$ denotes the corresponding perturbation vector. In this case, we can define a Signal-to-Noise Ratio (SNR) as $SNR \triangleq 10 \log_{10} \frac{\sum_{i=1}^{n} [\mathbf{t}_{(n)}]_i^2}{\sum_{i=1}^{n} [\mathbf{w}_{(n)}]_i^2}$ The entries of $\mathbf{w}_{(n)}$ are assumed to be i.i.d random variables following the uniform distribution on $\alpha[0, 1]$, where α is a normalizing constant adjusted to different values of the SNR as defined above. The normalized total error of reconstruction is defined as $\epsilon = \frac{\|\mathbf{t} - \mathbf{t}^{\#}\|_2}{\|\mathbf{t}\|_2}$ where $\mathbf{t}^{\#}$ is the estimate of $\mathbf{t}$ (the first row/column of $\mathbf{T}$). Similarly, the normalized prediction error is defined as $\epsilon_{pred} = \frac{\|\mathbf{t}_{(-n)} - \mathbf{t}^{\#}_{(-n)}\|_2}{\|\mathbf{t}_{(-n)}\|_2}$ where $\mathbf{t}_{(-n)} = [t_n, t_{n+1}, \cdots, t_{N-1}]^T$ and $\mathbf{t}^{\#}_{(-n)} = [t^{\#}_n, t^{\#}_{n+1}, \cdots, t^{\#}_{N-1}]^T$.

We further assume that all algorithms use the *same number of measurements*, denoted by n. This means that the proposed algorithm as well as (2.71) effectively sample the first n (perturbed) entries of $\mathbf{T}$, whereas for (2.72), we assume $M_r = n$. When $n < N$, the algorithms

in (2.72) and (2.71) implicitly perform "extrapolation/prediction" in order to find the optimal estimate of $\mathbf{T}$. In Fig. 2.4, we study the normalized prediction error ϵ_{pred} and the normalized total error ϵ of the aforementioned algorithms as a function of SNR. It can be seen that the proposed method outperforms the algorithms in [38, 57], in terms of both the prediction error and the total normalized error.

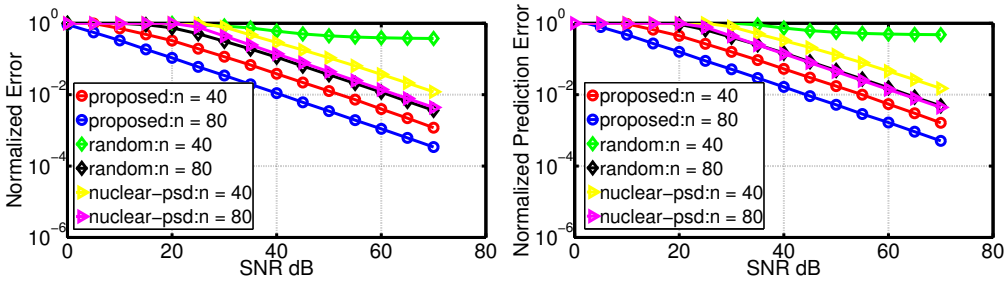


Figure 2.4: Estimation error of different algorithms as functions of SNR. (Top) Normalized error ϵ v/s SNR. (Bottom) Normalized prediction error v/s SNR. The results are averaged over 100 runs. Here $N = 110$, $r = 30$.

We also study the prediction error as a function of the sampled size n. Fig. 2.5 shows the normalized total error and prediction error as a function of sampled size n. It can be seen that the average prediction error decreases with increasing n, and increases as rank and noise power increase. In all cases, the proposed method shows better performance over *random* and *nuclear-psd*. The absence of any tuning parameter, and exploitation of the exact parametric representation of low rank PSD Toeplitz matrices (for both compression and reconstruction) are potential factors behind the superior performance of our algorithm.

Setting 2:Finite Noisy Snapshots

We next consider the noisy measurement model (2.16) consisting of a finite number (L) of snapshots. We study the error of reconstruction for two values of n: (i) $n = 30$ (ii) $n = N = 42$. In the latter case, no prediction is necessary, and we only only execute Step 1 (denoising) of our proposed algorithm. Fig. 2.6 and Fig. 2.7 show the performance of all the algorithms as functions

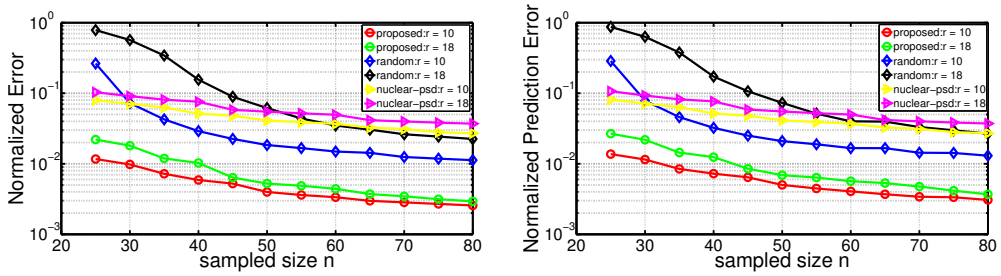


Figure 2.5: Estimation error of different algorithms as functions of n. (Top) Normalized error ϵ v/s n. (Bottom) Normalized prediction error v/s n. The results are averaged over 100 runs. Here $N = 110$, SNR = 50dB.

of L for both scenarios. It can be seen that the proposed algorithm continues to outperform (2.72) and (2.71). This shows that simple least squares based estimation with PSD and Toeplitz constraints provides the best performance in this setting.

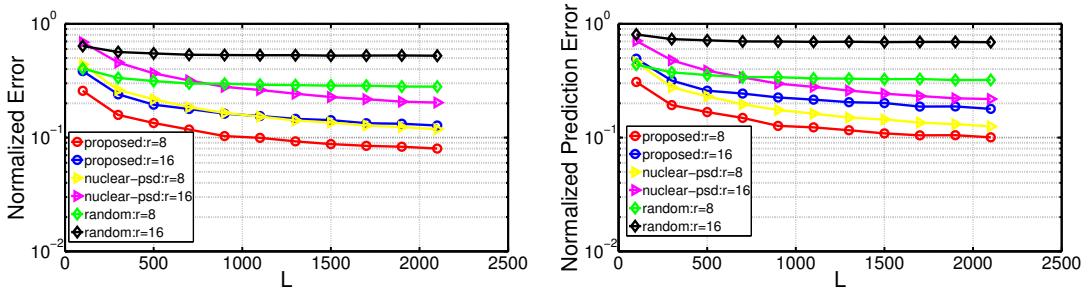


Figure 2.6: Estimation error of different algorithms as functions of L. (Top) Normalized error ϵ v/s L. (Bottom) Normalized prediction error v/s L. The results are averaged over 200 runs. Here $n = 30, N = 42, \sigma_n = 0.09$.

2.6.3 Approximate Low Rank

In practice, $\mathbf{T}$ may not be low rank but can be approximated by a low rank matrix. We study the robustness of the proposed method in such a setting when the entries of $\mathbf{T}$ can no longer be represented as a sum of complex exponentials. We generated an approximately low rank $\mathbf{T}$ by adding a small diagonal loading factor to a low rank PSD Toeplitz matrix. In Fig. 2.8, we

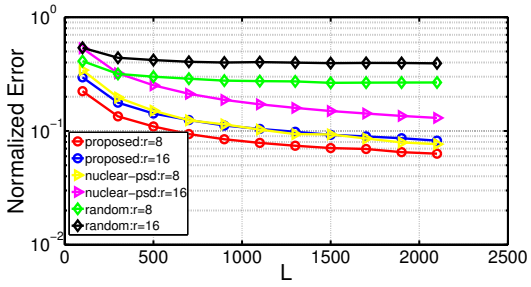


Figure 2.7: Normalized error ϵ of different algorithms with $n = N = 42$. $\sigma_n = 0.09$. Averaged over 200 runs.

study the performance of proposed method for such an approximately low rank $\mathbf{T}$ as a function of sampled dimension n and compare it with *nuclear-psd* and *random*. The proposed method exhibits robustness to violation of the low rank assumption and its performance improves with increasing n.

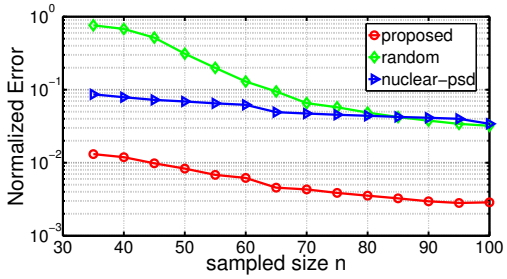


Figure 2.8: Recovery performance of different algorithms when $\mathbf{T}$ is approximately low rank Toeplitz. The matrix $\mathbf{T}$ is approximately of rank 30, with ambient dimension $N = 110$. Here SNR is 50 dB and the results are averaged over 50 runs.

2.6.4 Comparison with Maximum Likelihood based method

We now compare the performance of our algorithm with the SPA algorithm developed in [121], which is derived from the Maximum Likelihood method and is also regularization-free. Since the SPA algorithm [121] assumes that the measurements $\{\mathbf{y}_p\}$ are Gaussian random variables, we also generate both the data $\{\mathbf{x}_p\}$ and noise $\{\mathbf{n}_p\}$ as i.i.d zero mean Gaussian random variables, with respective covariance matrices given by $\mathbf{T}$ and $\sigma_n\mathbf{I}$. Additionally, we

adopt the same definition of SNR as in [121]: $SNR = 10\log_{10}\frac{\min_{1\leq i\leq r}(d_i)}{\sigma_n}$ The normalized mean-square-error (N-MSE) is defined as

$$\text{N-MSE} = \frac{1}{K}\sum_{k=1}^{K}\frac{\|\hat{\mathbf{t}}_k^{\#} - \mathbf{t}\|_2^2}{\|\mathbf{t}\|_2^2} \tag{2.74}$$

where $\hat{\mathbf{t}}_k^{\#}$ is the estimate of $\mathbf{t}$ (the first column of $\mathbf{T}$) for kth Monte-Carlo run.

In Fig. 2.9, we compare the N-MSE as a function of SNR for different values of rank r and number of snapshots L. It can be seen that for the same value of $n = 30$, SPA performs better for smaller value of r, whereas our algorithm outperforms SPA when $r = 28$, which corresponds to a maximally compressed setting. Additionally, the performance of both algorithms improve as L increases. We also compare the frequency estimation performance of both algorithms by reproducing a similar plot from [121] (corresponding to sparse arrays) in Fig. 2.10 which shows the estimated values of frequencies and powers over several Monte Carlo runs. [5]. Both methods exhibit similar frequency reconstruction performance. Unlike SPA, although the proposed method is not provably asymptotically Maximum Likelihood for Gaussian signals, its frequency reconstruction performance seems to be similar to SPA, and for near-critically compressed measurements, it can even outperform SPA in terms of covariance estimation error.

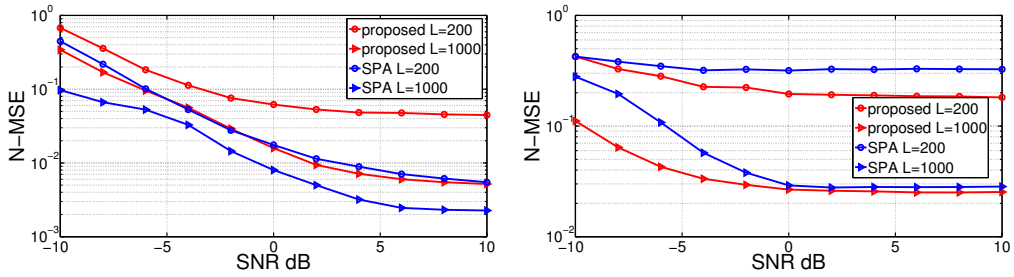


Figure 2.9: Normalized mean-square-error (N-MSE) as a function of SNR, $N = 40$, $n = 30$. The results are averaged over 100 Monte Carlo runs. (Top) $r = 10$ (Bottom) $r = 28$.

[5] We only show the non-negative frequencies for convenience

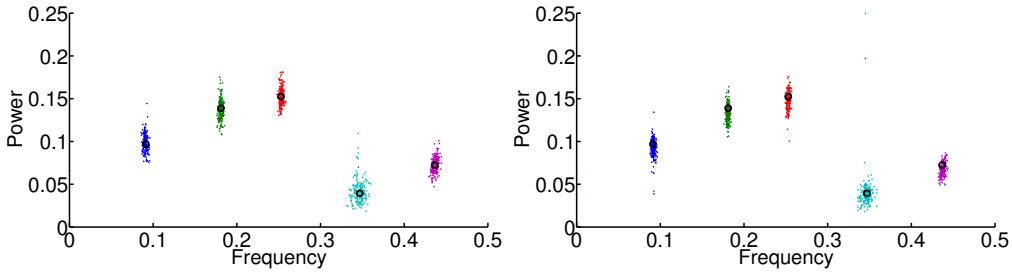


Figure 2.10: Estimates of frequencies and power ($\{f_i, d_i\}$) produced by SPA and the proposed algorithm. Here, $r = 10$, $n = 12$, $N = 40$, $L = 500$ and we consider 200 Monte Carlo runs. (Top) Proposed method. (Bottom) SPA [121]

.

2.6.5 Study on Frequency Separation

Recall that the stability of prediction in Theorem 3 is established under a separation condition on the frequencies, that is inversely proportional to n. We now study the effect of frequency separation as a function of n on the reconstruction error. We generate $\mathbf{T}$ of size $N = 20$, with four symmetric frequencies (two positive and two negative) where one of the positive frequencies is fixed at 0.1 and the other is located at $0.1 + \Delta$. We consider two values for the frequency separation Δ: (i) $\Delta = 0.02$ and (ii) $\Delta = 0.2$. For each value of Δ, we compare the performance of our algorithm against (2.72) and (2.71).

For each of the following experiments, we assume that (2.71) samples all of the N entries of $\mathbf{T}$, (implying $M = O(\sqrt{N})$). As stated earlier, this same setting was used in the numerical experiments of [57]. Similarly, we assume that (2.72) uses $M_r = N$ measurements. However, for our proposed algorithm, we use two different values of n (for each Δ), in order to understand how changing n affects the quality of prediction for small and large Δ.

In the first experiment, we assume $n = 12$, implying that the number of measurements for the proposed algorithm is smaller than both *nuclear-psd* and *random*, and we need to predict the remaining $N - n = 8$ entries of $\mathbf{T}$. Fig. 2.11 shows the performance of all the algorithms for $\Delta = 0.02$ and $\Delta = 0.2$. Here, N-MSE denotes the covariance estimation error given by (2.74) and F-MSE is the mean squared error for frequency estimation defined as F-MSE $=$

$\frac{1}{K}\sum_{k=1}^{K}\sum_{i=1}^{r}|f_{i,k}-f_{i,k}^{\#}|^2$ where K is the number of Monte-Carlo runs and $f_{i,k}^{\#}$ is the estimate of f_i in kth run. It can be seen that for $\Delta = 0.02$ (small frequency separation), (2.71) outperforms the proposed algorithm. However, for a larger separation of $\Delta = 0.2$, the proposed algorithm shows similar performance as (2.71), although it uses smaller number of measurements.

In the second experiment, we assume that the proposed algorithm also samples the entire matrix $\mathbf{T}$, (i.e. $n = N$) and therefore uses the same number of measurements as *nuclear-psd* and *random*. Fig. 2.12 shows the corresponding performance of all the algorithms in this setting. It can be seen that proposed algorithm outperforms both *nuclear-psd* and *random* in terms of N-MSE, *regardless of the separation between frequencies*. From these experiments, it can be concluded that when the proposed algorithm uses fewer measurements than *nuclear-psd*, its performance depends on the separation Δ between frequencies. For small Δ, *nuclear-psd* outperforms the proposed method, but for larger Δ, both algorithms exhibit similar N-MSE for different values of L. On the other hand, when both algorithms use the same number of measurements, the proposed algorithm outperforms *nuclear-psd* in terms of N-MSE, *regardless of frequency separation*.

2.6.6 Computational Complexity

Finally, we compare the computational complexity of the proposed method with *nuclear-psd* and *random*. Fig. 2.13 shows the run-time of these algorithms as we increase the problem size N. We simulated all algorithms on a Dell OptiPlex 7020 desktop with Intel(R) Core(TM) i7-4790 CPU @ 3.60GHz, and 16 GB Memory, using the CVX toolbox for MATLAB, and on the same dataset. Since the problem size (number of unknown variables) of the proposed algorithm is $O(n)$, rather than N, the complexity of our method is smaller than the other algorithms. Moreover, our complexity *does not grow with* N. This may turn out to be especially advantageous in the high dimensional setting when N is very large.

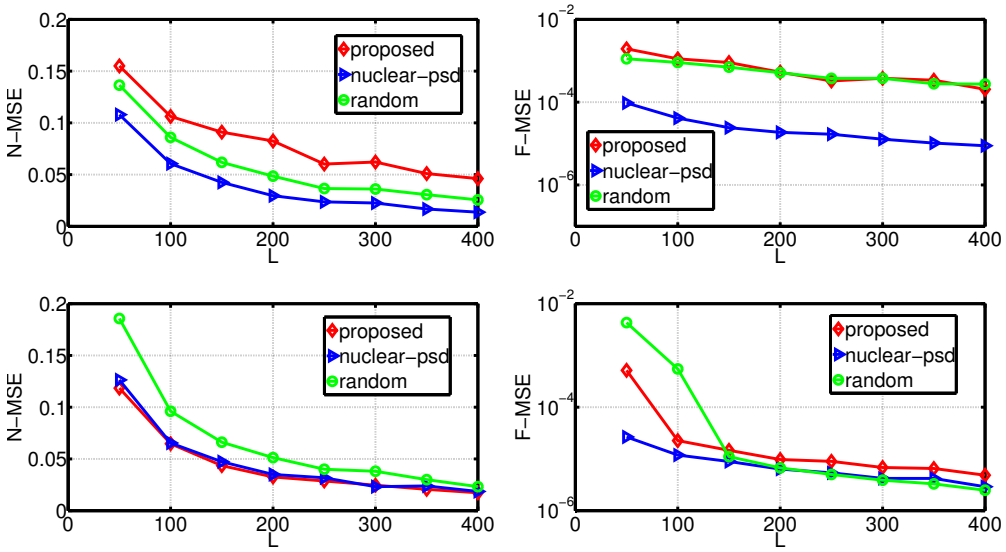


Figure 2.11: N-MSE and F-MSE as a function of the number of snapshots (L) for two different values of Δ: (Top) $\Delta = 0.02$ and the frequencies in the positive half are $[0.1, 0.12]$ (Bottom) $\Delta = 0.2$ and the frequencies in the positive half are $[0.1, 0.3]$. Here, *random* uses N rank-one random measurements and similarly, *nuclear-psd* samples the entire $N \times N$ matrix $\mathbf{T}$. The proposed algorithm only observes $\mathbf{T}_{(n)}$ with $n = 12$ and uses less measurements than *random* and *nuclear-psd*. Here, $N = 20$, $\sigma_n = 0.09$, $r = 4$ and the results are averaged over 200 runs.

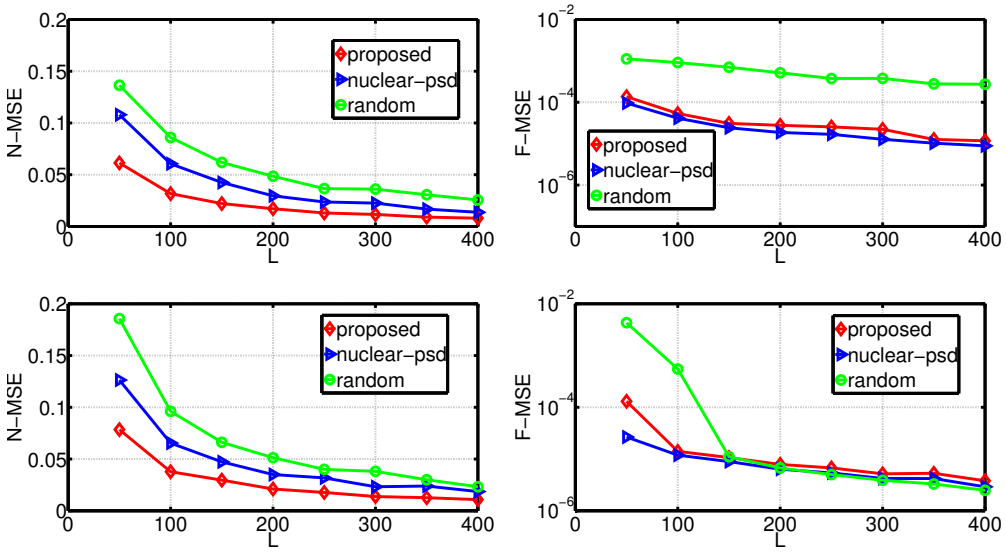


Figure 2.12: N-MSE and F-MSE as a function of the number of snapshots (L) for two different values of Δ: (Top) $\Delta = 0.02$ and the frequencies in the positive half are $[0.1, 0.12]$ (Bottom) $\Delta = 0.2$ and the frequencies in the positive half are $[0.1, 0.3]$. The experimental settings are identical to that in Fig. 2.11 except that in this case, the proposed algorithm also observes the entire matrix with $n = N$.

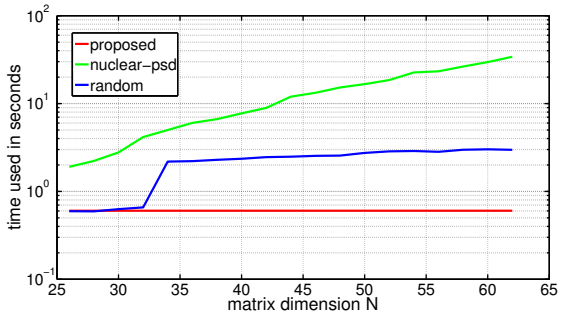


Figure 2.13: Comparison of run-times of the proposed method and the nuclear norm based recovery algorithms in [38, 57]. Here, $r = 10$, SNR $= 20$ dB and $n = 20$. The run-time is averaged over 100 runs.

2.6.7 Numerical Performance of Coarray ML-MUSIC

We compare the performance of the (Co-MLM) framework with several other algorithms that attempt to estimate the co-array covariance matrix $\mathbf{T}_{ca}$ from $\hat{\mathbf{R}}_{yy}$. Since the ML problem (2.62) is highly non-convex, we adopt the idea of Extended Invariance Principle proposed in [137], which provides an approximation of the ML objective given sufficient number of snapshots [4,121]. In particular, we solve the following convex problem

$$\begin{aligned} &\min_{\mathbf{T},\mathbf{X}} \text{Trace}(\mathbf{X}) + \text{Trace}(\hat{\mathbf{R}}_{yy}^{-1}\mathbf{S}\mathbf{T}\mathbf{S}^T) \\ &\begin{bmatrix} \mathbf{X} & \hat{\mathbf{R}}_{yy}^{1/2} & \\ \hat{\mathbf{R}}_{yy}^{1/2} & \mathbf{S}\mathbf{T}\mathbf{S}^T & \\ & & \mathbf{T} \end{bmatrix} \succeq 0 \\ &\mathbf{T} \text{ is Toeplitz} \end{aligned} \tag{2.75}$$

We compare (2.75) with the nuclear norm based algorithm in [57] (henceforth referred as *nuclear-psd*) and the Co-array MUSIC algorithm of [45, 138]. Consider a nested array with $M = 6$ sensors located at $\{1, 2, 3, 4, 8, 12\}$. It receives signals from $D^\star = 10 > M$ narrowband sources whose spatial frequencies are equi-spaced in the range $[-0.4, 0.4]$. The sources are assumed to be of equal power, i.e. $p_1^\star = p_2^\star = \cdots = p_{D^\star}^\star = 5$. We first compute an estimate $\hat{\mathbf{T}}$ of the co-array covariance matrix using the three algorithms, and then apply the MUSIC algorithm on $\hat{\mathbf{T}}$ to estimate the DOAs. The smallest singular value of $\hat{\mathbf{T}}$ serves as an estimate of the noise power. After estimating the DOAs, the source powers can be obtained from $\hat{\mathbf{T}}$ using the Least Squares method. A typical MUSIC spectrum for all three algorithms is shown in Fig. 5.2. It can be seen that for this realization, (Co-MLM) correctly identifies the DOAs, whereas the other two algorithms exhibit spurious peaks.

We also study the Mean-Squared-Error (MSE) of DOA estimates for all three algorithms. Following [4, 121], to compute the MSE, we use the estimated DOAs that correspond to the $D^\star$

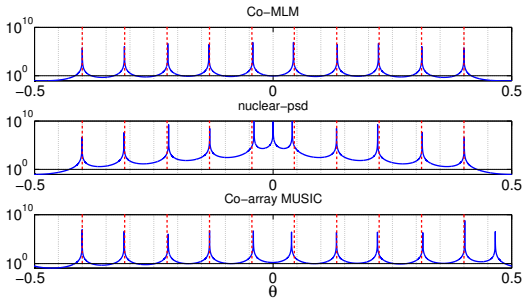


Figure 2.14: The log-spectrum for MUSIC algorithm of all three methods. $L = 200$, $\sigma_n = 0.5$. The red dotted lines indicate the true DOAs.

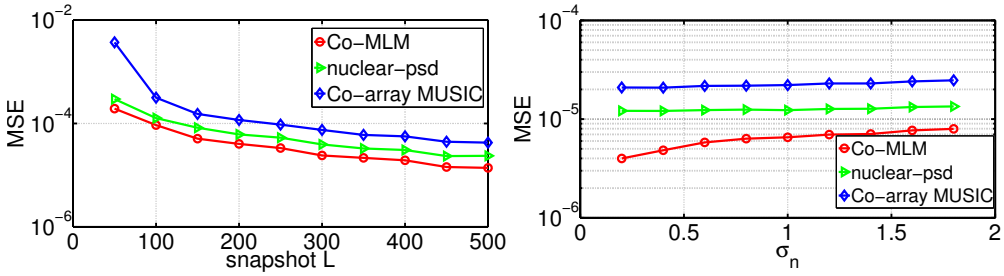


Figure 2.15: MSE of DOA estimates for the three algorithms (Left) MSE versus L. Here, $\sigma_n = \sqrt{5}$. The plots are averaged over 300 runs. (Right) MSE versus σ_n with $L = 800$, averaged over 4000 runs.

largest values of the estimated source powers. Fig. 2.15 shows the MSE as a function of σ_n and L. For both cases, (Co-MLM) outperforms the other two algorithms, and the gap in MSE is larger for smaller L.

2.7 Conclusion

In this chapter, we considered the problem of compressing and recovering a low rank PSD Toeplitz matrix using minimal number of measurements. As a major contribution of this work, we showed that it is possible to recover a rank r PSD Toeplitz matrix from a sketch of size $O(\sqrt{r}) \times O(\sqrt{r})$, which is order-wise optimal. The sketch can be obtained by using a newly proposed class of structured sampler, namely, the Generalized Nested Sampler (GNS), or sparse ruler based sampling techniques. In absence of noise, these structured samplers provably outperform random sampling where the number of required measurements exhibits a logarithmic dependence on ambient dimension N. We further reformulated the reconstruction problem in terms of line spectrum estimation and studied the performance of gridless techniques, such as MUSIC, for recovering $\mathbf{T}$ from its sketch produced by the GNS. By using the Vandermonde decomposition of PSD Toeplitz matrices, we show that a simple least squares denoising with PSD constraints suffice to guarantee stable reconstruction of a $N \times N$ Toeplitz covariance matrix of rank r from a sketch of size $O(\sqrt{r}) \times O(\sqrt{r})$. The proposed method is regularization-free and has low complexity. In the presence of noise and finite snapshots, we developed an explicit bound on the prediction error in terms of r, noise power and the observation length n.

Motivated by the Vandermonde decomposition, we also studied the maximum likelihood problem for estimating the DOAs of $O(M^2)$ sources that exploits the co-array structure of a nested array with M antennas. We showed that a certain constrained ML framework (with PSD and Toeplitz constraints) for estimating the co-array covariance matrix provably produces a ML estimate for the DOAs when the number of sources is assumed to be unknown. In addition, under

a certain orthogonality condition on the source signal waveforms, the PSD Toeplitz constrained ML algorithm can exactly recover the true DOAs even with a finite number of temporal snapshots.

The numerical simulations validated the theoretical claims established in this chapter Future work will be directed towards understanding the need for separation condition for frequency estimation in presence of noise, and establishing optimal error bounds.

Chapter 2, in part, is a reprint of the material as it appears in following papers:

- H. Qiao and P. Pal, "Gridless Line Spectrum Estimation and Low-Rank Toeplitz Matrix Compression Using Structured Samplers: A Regularization-Free Approach, *IEEE Transactions on Signal Processing*, vol. 65, no. 9, pp. 2211-2226, May 2017.

- H. Qiao and P. Pal, "On Maximum Likelihood Methods For Localizing More Sources than Sensors, *IEEE Signal Processing Letters*, vol. 24, no. 5, pp. 703-706, 2017.

- H. Qiao and P. Pal, "Generalized Nested Sampling for Compressing Low Rank Toeplitz Matrices, *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 1844-1848, Nov. 2015.

- H. Qiao and P. Pal, "Stable Compressive Low Rank Toeplitz Covariance Estimation Without Regularization, in the *Proceedings of Asilomar Conference on Signals, Systems and Computers*, 2016.

- H. Qiao and P. Pal, "Generalized Nested Sampling for Compression and Exact Recovery of Symmetric Toeplitz Matrix, in the *Proceedings of IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Atlanta, GA, USA, 2014.

The dissertation author was the primary investigator and author of these papers.

Chapter 3

Super-Resolution and Support Recovery with Non-Negative Constraint

This chapter simultaneously considers the problems of super-resolution and support-recovery with non-negative constraint. The two problems have similar measurement models and analysis tools in the proof of stable recovery. In the following, we will first discuss super-resolution and provide an intuitive explanation of the usefulness of non-negative constraint. Then, we will study the problem of joint support recovery with multiple measurement vector model. The non-negative constraint naturally arises as we explicitly exploit the correlation information of the jointly sparse signals. We establish a universal analysis framework that is independent of the chosen objective functions in the optimization program. The universal upper bound on the estimation error is shown to be tight by explicitly constructing an example. As another important contribution, we propose a computationally efficient algorithm for exactly recover the support by simply hard-thresholding the estimated signal power. Next, we try to further generalize the universal analysis framework by introducing the concept of modulus of continuity. This quantity fundamentally bounds the estimation error and is only dependent on the measurement model. We show that non-negativity constraint will give a better modulus of continuity in terms

of an amplification factor compared to the general case. Finally, we will discuss a particular super-resolution problem known as fluorescence microscopy. Sparse array idea is shown to be useful to identify more sources than sensors by exploiting the Gaussian point spread function and the statistical property of the point sources.

3.1 Introduction

The problem of super-resolution is fundamental across imaging applications such as astronomy [150], medical imaging [151], microscopy [152] and radar [153]. In these systems, the resolution of the captured image is always limited by the physical measurement process which necessitates the use of sophisticated signal processing techniques to retrieve finer details that are apparently lost.

The problem of super-resolution with noisy measurements was analyzed in the pioneering work by Donoho [139] and further developed in recent works [141,142] where total-variation (TV) and l_1 norm based convex algorithms were used for promoting sparse structure in super-resolution reconstruction. The analysis technique of [141, 142] involves an explicit construction of a certain dual polynomial (based on the Fejér kernel), whose properties can be exploited to analyze the performance of convex super-resolution algorithms for noisy line spectrum estimation [59] and low-rank Toeplitz covariance estimation [4].

More recently, the role of positive constraints in super resolution was analyzed in [61] by imposing a new notion of Rayleigh regularity on the underlying signal. Using the same dual polynomial as [141, 142], the authors in [61] established stability guarantees for a simple l_1 norm based denoising problem with non-negative constraint. In another recent work [148], the author established robust recovery guarantees of positive streams of spikes by imposing strong structural constraints on the admissible blurring kernel. It should be noted that existing analysis of noisy super-resolution focus on solving *convex* problems. In this chapter, we show that positive

constraints on the unknown target signal can be exploited in a suitable way alongside its sparsity, leading to a *non-convex* super-resolution problem which minimizes the $l_{1/2}$ quasinorm of the signal. Such $l_q(0 < q \leq 1)$ norm based non-convex constrained optimization problems and corresponding algorithms to (approximately) solve them, have been studied in recent literature [154–157, 163]. However, in order to establish stable reconstruction guarantees, these techniques either rely on exploiting certain properties of the measurement matrix such as the Restricted Isometry Property (RIP) [96]) [156, 158–160, 163–166], or Kruskal rank [155]. However, the measurement matrix arising in super-resolution imaging is a *deterministic rank-deficient* matrix composed of DFT matrices for which RIP cannot be established. Besides, $l_{1/2}$ quasinorm is not differentiable and recent advances in non-convex gradient descent based algorithms are inapplicable [86, 158, 164].

The problem of joint sparse support recovery from Multiple Measurement Vectors (MMV), is of significant interest in compressive sensing and sparse signal recovery [16, 135, 147, 168, 170–174]. The goal is to identify the common support (denoted by $\mathcal{S}$) of a set of L vectors $\mathbf{x}[l] \in \mathbb{C}^N, 1 \leq l \leq L$ using L (compressed) measurement vectors $\mathbf{y}[l] \in \mathbb{C}^M, 1 \leq l \leq L$ ($M < N$) acquired using a common measurement matrix $\mathbf{A} \in \mathbb{C}^{M \times N}$ and contaminated with additive noise $\mathbf{w}[l]$

$$\mathbf{y}[l] = \mathbf{A}\mathbf{x}[l] + \mathbf{w}[l], \quad \text{Supp}(\mathbf{x}[l]) = \mathcal{S}, |\mathcal{S}| = s \tag{3.1}$$

The MMV model is originally inspired by measurement models in sensor array processing [172, 174] where $\mathbf{y}[l]$ denotes the lth temporal sample (or snapshot) of the measurements collected at an array of M sensors. The MMV model arises in a wide range of applications such as EEG/MEG source localization [188, 189], DOA estimation [184, 185, 199, 200], sub-Nyquist sampling [41, 194], channel estimation [195] and MRI [196], to name a few.

The MMV problem has been widely studied in the past decade, giving rise to a large

number of computationally efficient algorithms that exploit the joint sparsity of $\{\mathbf{x}[l]\}_{l=1}^{L}$ as well as the rank of the measurement matrix $\mathbf{Y}_L = [\mathbf{y}[1], \mathbf{y}[2], \cdots, \mathbf{y}[L]]$ for better performance and/or recovery guarantees [170–173, 181, 201]. On the other hand, rigorous guarantees for support recovery in MMV models (in terms of achievability as well as converse results) have been developed using information theoretic tools [143, 145], exhaustive search decoders [146], joint typicality decoder [144], and recent unification of such guarantees for linear and non-linear measurement models [143, 202]. However, a common feature of most of these results is that their theoretical guarantees are applicable only when the size of the support is less than M, and cannot be applied in the regime $s > M$. A main reason 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 by appropriate design of the measurement matrix. We now review the role of sparse arrays on the design of the measurement matrix $\mathbf{A}$. Such arrays have drawn considerable attention in recent times due to their ability to recover supports of size $s > M$.

3.1.1 Difference Co-Arrays, Correlation-Awareness and Localization of More Sources than Sensors

Sparse sensor arrays such as nested [45] and coprime arrays [17] are gaining increasing attention owing to many associated benefits over the conventional and widely-used Uniform Linear Array (ULA) [41, 130, 203]. One of the striking features of these arrays is that their difference co-array [204] contains an ULA segment of length $\Theta(M^2)$. This difference co-array controls the correlation structure of the measurements, and can be utilized to resolve more sources than sensors. In recent times, clever extensions of these arrays such as super nested arrays [198] and generalized coprime arrays [199] have been proposed in literature to combat mutual coupling and avoid small inter-element spacings. Besides applications in Direction-of-Arrival (DOA) estimation and super-resolution imaging [7, 175], the idea of sparse array as a sampling operator has intimate connections to the important problem of compressive covariance

sensing [4, 18, 40, 41, 130]. Statistical performance of sparse arrays in estimation of more source directions than sensors has been recently analyzed in terms of Cramér-Rao bounds [130, 169, 179], mean squared error (MSE) of estimators based on the Co-array MUSIC algorithm [130] and (nonlinear) Least Squares algorithms [203].

However, these results are asymptotic in nature (i.e. tight only when $L \to \infty$) and are either loose or inapplicable when L is finite. Moreover, they also require the number of sources to be known apriori, or assume certain stringent conditions on correct identification of the signal subspace, which may be unrealistic in practice (or at best, very difficult to ensure with finite L). On the other hand, recent advances in analysis of super-resolution algorithms based on Total Variation (TV) norm and atomic norm minimization [1] cannot ensure absence of spurious spike estimates and cannot be directly utilized to characterize the MSE of source localization [59, 141].

Sparse Arrays and MMV: As pointed out in [135, 172], the MMV model has its origins in sensor array and multichannel signal processing. Inspired by their success in array processing, sparse arrays have been successfully used in MMV models in recent times to empirically demonstrate the possibility of *recovering more sources than sensors in applications ranging from radar, sonar, satellite based navigation (GNSS) to super-resolution optical imaging* [16, 135, 175, 176, 183, 186, 187, 190]. Contrary to standard MMV algorithms [173, 201], these techniques aim to directly recover the sparse support (bypassing the recovery of the unknown vectors $\mathbf{x}[l]$) from the data correlation matrix $\frac{1}{L}\sum_{l=1}^{L}\mathbf{y}[l]\mathbf{y}^H[l]$ captured by a sparse array, under the assumption that the non-zero elements of $\mathbf{x}[l]$ are statistically uncorrelated. The idea of using such correlation priors on $\mathbf{x}[l]$ for source localization has been of continued interest in MMV models for over a decade [120, 174, 180, 181, 184, 185, 189, 205], giving rise to powerful Bayesian algorithms for sparse signal recovery, such as the M-SBL framework [181, 188, 197].

However, theoretical conditions under which these algorithms can recover supports of size larger than M with finite L, have so far only been partially developed in two distinct lines of work.

[1] Contrary to MMV models that assume sources to be located on a grid [120, 147, 174, 175, 180], these algorithms assume a "gridless" model for source locations.

In the first line of work, the authors of [135] make a very interesting observation that as long as the rows of the matrix $\mathbf{X}_L = [\mathbf{x}[1], \mathbf{x}[2], \cdots, \mathbf{x}[L]]$ are orthogonal, M-SBL can identify supports of size larger than M. However, such a deterministic condition does not lead to meaningful statistical guarantees on support recovery. In the second line of work [16], we proposed solving a new convex optimization problem called "Correlation-Aware LASSO" (or Co-LASSO) that utilizes the structure of the data correlation matrix $\frac{1}{L}\sum_{l=1}^{L} \mathbf{y}[l]\mathbf{y}^H[l]$ to recover the joint support of more sources than sensors. Although Co-LASSO provably recovers supports of size $s = \Theta(M^2)$ when the exact correlation matrix is available (i.e. when $L \to \infty$), establishing such guarantees for finite L has been an *open problem* in the regime $s > M$ [16, 101, 147]. To address this issue, we recently cast the correlation-aware support recovery problem in terms of a multiple hypothesis testing framework [168] (where each hypothesis corresponds to one of $\binom{N}{s}$ possible supports), and derived non-asymptotic bounds on the probability of error in the regime $s = \Theta(M^2)$ as a function of L. Although this result applies to general measurement matrices (including sparse arrays), the algorithm for support recovery involves testing $\binom{N}{s}$ hypotheses which can become computationally intractable even for moderate N and s. In this paper, we overcome this drawback by analyzing a *computationally efficient* framework (which is closely related to our convex Co-LASSO algorithm from [16]) for support recovery in the regime $s = \Theta(M^2)$, and developing *non-asymptotic* guarantees on the probability of support recovery, as a function of L. Our guarantees do not require the knowledge of s and the source powers can be unequal.

3.1.2 The Modulus of Continuity

In this chapter, we consider the same problem setting as in [61] where the goal is to reconstruct a *sparse non-negative* discrete signal from low-frequency measurements. In contrast to previous works, our goal is to perform a unified analysis of positive super-resolution independent of particular algorithms. To achieve this, we revisit the concept of Modulus of Continuity (MC) [139, 140] which essentially provides an upper bound on the error of any algorithm,

simply by leveraging the structure of signals. We study the explicit role of positivity on the Modulus of Continuity and our results show that the scaling factor of MC can be improved from $O(N^3)$ [139, 140] to $O(N^{2.5})$ where N is the dimension of the underlying signal. This improvement is due to positive constraints imposed on the desired signal, which is not considered in [139].

3.1.3 Notations

Throughout this chapter, scalars, vectors and matrices are denoted by lowercase, boldface lowercase, and boldface uppercase alphabets respectively. The Kruskal rank of a matrix $\mathbf{X}$ (the maximum integer k such that any k columns of $\mathbf{X}$ are linearly independent) is denoted by $\text{Krank}(\mathbf{X})$. The notation $\text{vec}(\mathbf{X})$ denotes the column-wise vectorization of a matrix $\mathbf{X}$. The notation $\mathbf{X} \odot \mathbf{Y}$ represents the Khatri-Rao product (column-wise Kronecker product) of matrices $\mathbf{X}, \mathbf{Y}$ with the same number of columns.

3.2 Understanding the Role of Non Negativity Constraint in Super Resolution

3.2.1 Problem Formulation

The goal of discrete positive super-resolution [61, 148, 206] is to reconstruct a signal (or image) $\mathbf{x}^\star \in \mathbb{R}^N$ from measurement $\mathbf{y} \in \mathbb{C}^N$ of the form [61]

$$\mathbf{y} = \mathbf{Q}\mathbf{x}^\star + \mathbf{w} \qquad \mathbf{x}^\star \geq \mathbf{0} \tag{3.2}$$

where $\mathbf{x}^\star$ is a sparse vector with non-negative entries, $\mathbf{w}$ is the measurement noise and $\mathbf{Q} \in \mathbb{C}^{N \times N}$ is the measurement matrix. Here, $\mathbf{Q}$ represents a low-pass filter such that $\mathbf{y}$ only retains the

low-frequency components of $\mathbf{x}^\star$, and the high-frequency components are lost. This imparts the following special structure to $\mathbf{Q}$ [61],

$$\mathbf{Q} = \mathbf{F}_N^H \mathbf{\Lambda}_n \mathbf{F}_N \tag{3.3}$$

where $\mathbf{F}_N \in \mathbb{C}^N$ is given by $[\mathbf{F}_N]_{k,l} = \frac{1}{\sqrt{N}} e^{-j2\pi kl/N}$, $-N/2+1 \leq k \leq N/2$, $0 \leq l \leq N-1$ and $\mathbf{\Lambda}_n = \text{diag}([\lambda_{-N/2+1}, \cdots, \lambda_{N/2}])$ with

$$\lambda_k = \begin{cases} 1, & k = -\frac{n-1}{2}, \cdots, \frac{n-1}{2} \\ 0, & \text{otherwise} \end{cases}$$

[2]. Hence, $\mathbf{Q}$ only collects the n low-frequency DFT coefficients of $\mathbf{x}^\star$. The goal of super-resolution is to accomplish the difficult task of recovering the *lost high frequency components* of $\mathbf{x}^\star$ by utilizing its sparsity.

In recent efforts to solve the positive superresolution problem *with provable guarantees*, the authors in [61] proposed the following convex optimization problem to estimate sparse non-negative $\mathbf{x}^\star$

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{Q}\mathbf{x}\|_1 \quad s.t. \quad \mathbf{x} \geq \mathbf{0} \quad (P_{\text{den}})$$

Inspite of its simple formulation, (P_{den}) is quite effective in finding $\mathbf{x}^\star$ with provable guarantees. In fact, it is shown that if $\mathbf{x}^\#$ is an optimal solution to (P_{den}), then the l_1 norm of the estimation error $\|\mathbf{x}^\# - \mathbf{x}^\star\|_1$ gets amplified by a factor of $(\frac{N}{n-1})^2$ where $\frac{N}{n-1}$ is the so-called super-resolution factor (SRF).

Notice that the formulation (P_{den}) does not explicitly enforce any sparsity penalty on $\mathbf{x}$, and only uses the prior that it is non-negative. If we assume that $\|\mathbf{w}\|_1 \leq \delta_1$, we can further

[2]For ease of presentation, we assume that the ambient dimension N is even and n is odd [61]

promote sparsity by using the l_1 norm of $\mathbf{x}$ as a convex surrogate for its sparsity [96]. This will be equivalent to adding a non-negative constraint to the convex super-resolution problem proposed in [141, 142]:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad s.t. \quad \|\mathbf{y} - \mathbf{Q}\mathbf{x}\|_p \leq \delta_p, \mathbf{x} \geq \mathbf{0} \quad (P_1)$$

where p is usually chosen as $p = 1, 2$. Although (P_1) is reminiscent of standard l_1 minimization problem from compressed sensing, conventional analysis tools such as Restricted Isometry Property (RIP) [96] or neighborly-polytope conditions [207, 208] are inapplicable in this case. This is because $\mathbf{Q}$ is a *deterministic* rank-deficient matrix composed of DFT matrices, for which neither RIP nor neighborly-polytope properties can be readily established. The problem (P_1) without the positivity constraint and for $p = 1$ was analyzed in [141,142] using a different analysis technique that constructs a certain dual certificate in the form of a trigonometric polynomial, and obtained similar error bounds as [61].

3.2.2 Motivation for using $l_{1/2}$ quasinorm in positive super-resolution

As a simple fact, any non-negative vector $\mathbf{x}$ can be represented as $\mathbf{x} = \mathbf{h} \circ \mathbf{h}$, where $\circ$ represents the Hadamard product. Thus, the convex l_1 norm minimization problem (P_1) can be equivalently rewritten in terms of $\mathbf{h}$ as

$$\begin{aligned} &\min_{\mathbf{h}} \|\mathbf{h}\|_2^2 \qquad (\tilde{P}_1) \\ &s.t. \quad \|\mathbf{y} - \mathbf{Q}(\mathbf{h} \circ \mathbf{h})\|_p \leq \delta_p, \mathbf{h} \geq \mathbf{0} \end{aligned}$$

Without loss of generality, we can assume $\mathbf{h}$ is also non-negative. The formulation $(\tilde{P}_1)$ has convex objective and non-convex constraints. Clearly, $(\tilde{P}_1)$ is equivalent to the convex problem (P_1) due to a one-to-one mapping between $\mathbf{h}$ and $\mathbf{x}$, and the optimal $\mathbf{h}$ has the same support as

the optimal $\mathbf{x}$. As evident from $(\tilde{P}_1)$, minimizing l_1 norm of $\mathbf{x}$ is equivalent to minimizing the l_2 norm of $\mathbf{h}$. A natural question to ask is: what happens if we enforce sparsity of $\mathbf{h}$ by replacing its l_2 norm with l_1 norm in the objective function? In other words, we consider the following problem

$$\min_{\mathbf{h}} \|\mathbf{h}\|_1 \quad s.t. \quad \|\mathbf{y} - \mathbf{Q}(\mathbf{h} \circ \mathbf{h})\|_p \leq \delta_p, \mathbf{h} \geq \mathbf{0} \quad (\tilde{P}_2)$$

Using $\mathbf{x} = \mathbf{h} \circ \mathbf{h}$, $(\tilde{P}_2)$ can be rewritten as

$$\min_{\mathbf{x}} \|\mathbf{x}\|_{\frac{1}{2}} \quad s.t. \quad \|\mathbf{y} - \mathbf{Q}\mathbf{x}\|_p \leq \delta_p, \mathbf{x} \geq \mathbf{0} \quad (P_2)$$

where we use the fact that $\|\mathbf{x}\|_{\frac{1}{2}}^{0.5} = \|\mathbf{h}\|_1$. The problem (P_2) minimizes the non-convex $l_{1/2}$ quasinorm of $\mathbf{x}$ over a convex feasible set. It is well known that minimizng the $l_{1/2}$ quasinorm favors even sparser solutions over minimizing l_1 norm [162, 163, 209, 210]. While $l_{1/2}$ quasinorm minimization has been explored and analyzed as a better alternative to l_1 norm for promoting sparsity, the corresponding theoretical guarantees (which are based on RIP) [160, 163, 165, 166] do not apply to $\mathbf{Q}$ which represents a low-pass filter in super-resolution imaging. We bridge this gap by first proposing an iterative reweighted l_1 norm minimization algorithm (for approximating the $l_{1/2}$ quasinorm) and developing theoretical guarantees under which this algorithm can provide stable solution in presence of noise.

3.2.3 Iterative Algorithm to approximate $l_{1/2}$ quasinorm minimization

Since (P_2) is non-convex and has non-differentiable objective function, recent advances in non-convex gradient descent based algorithms [86, 158, 159, 164] are not applicable. Inspired by [163], we propose an iterative reweighted l_1 norm minimization algorithm to solve (P_2), by explicitly enforcing positivity of the desired signal.

Table 3.1: Algorithm 1: Non-negative Reweighted l_1 Minimization

Input: Noisy measurements $\mathbf{y}$, parameters δ_p and $\varepsilon > 0$
Output: An estimate $\mathbf{x}^{\#}$ of $\mathbf{x}^{\star}$.

1. **Initialization:** An initial feasible guess $\mathbf{x}_0$ such that $\|\mathbf{y} - \mathbf{Q}\mathbf{x}_0\|_p \leq \delta_p, \mathbf{x}_0 \geq \mathbf{0}$, and a sequence of non-increasing positive numbers $\{\epsilon_n\}$ such that $\lim_{n\to\infty} \epsilon_n = 0$.

2. **Iteration:** Given $\mathbf{x}_n$, obtain $\mathbf{x}_{n+1}$ as

$$\mathbf{x}_{n+1} = \arg\min_{\mathbf{z}\in\mathbb{R}^N} \sum_{i=0}^{N-1} \frac{z_i}{(x_{n,i} + \epsilon_n)^{\frac{1}{2}}} \qquad (P_3)$$
$$s.t. \quad \|\mathbf{y} - \mathbf{Q}\mathbf{z}\|_p \leq \delta_p, \quad \mathbf{z} \geq \mathbf{0}$$

3. **Stopping Criterion:** Stop when $\|\mathbf{x}_n - \mathbf{x}_{n+1}\|_1 \leq \varepsilon$. Return $\mathbf{x}_{n+1}$ as the estimate of $\mathbf{x}^{\star}$.

The problem (P_3) in Algorithm 1 can be identified as a reweighted l_1 minimization problem, where the weights are given by $(x_{n,i} + \epsilon_n)^{-0.5}, i = 0, \cdots, N-1$. [3] The motivation stems from prior works in Majorization-Minimization (MM) algorithms that iteratively minimizes simple (possibly convex) surrogates for a given objective function [157]. In our case, we want to minimize the non-convex $l_{1/2}$ quasinorm $g(\mathbf{z}) \triangleq \|\mathbf{z} + \epsilon\|_{1/2}^{0.5} = \sum_{i=0}^{N-1} \sqrt{z_i + \epsilon}$, for $\mathbf{z} \geq \mathbf{0}$. We instead iteratively minimize the first-order linear approximation of $g(\mathbf{z})$ at $\mathbf{z} = \mathbf{x}_n$, giving rise to the following formulation

$$\min_{\mathbf{z}\geq\mathbf{0}} \Big\{ g(\mathbf{x}_n) + \sum_{i=0}^{N-1} \frac{1}{2} \frac{z_i - x_{n,i}}{\sqrt{x_{n,i} + \epsilon}} \Big\}, \text{ s.t. } \|\mathbf{y} - \mathbf{Q}\mathbf{z}\|_p \leq \delta_p, \tag{3.4}$$

Here, $\sum_{i=0}^{N-1} \frac{z_i}{\sqrt{x_{n,i}+\epsilon}}$ can be identified as the weighted l_1 norm of non-negative $\mathbf{z}$, implying that (3.4) identical to (P_3).

[3] The positive parameter ϵ_n is used to avoid zero denominator [157].

3.2.4 Analysis of $l_{1/2}$ minimization: Convergence and Error Bound

A main contribution of our paper is to analyze Algorithm 1 given the special structure of the low-pass filter $\mathbf{Q}$, and develop explicit bounds on the estimation error $\|\mathbf{x}^\star - \mathbf{x}^\#\|_1$. Although Algorithm 1 does not solve a convex problem, we show that the error bound behaves similar to the convex non-negative superresolution algorithm proposed in [61] and gets amplified by SRF^2.

We begin by defining the set of signals obyeing separation condition [4, 59].

Definition 3 *(Set of Non-Negative Signals Obeying Separation Condition) Given N and n, the set Δ^+_{sep} is given by*

$$\{\mathbf{x} \in \mathbb{C}^N, \mathbf{x} \geq \mathbf{0} \mid \rho(\frac{k}{N}, \frac{l}{N}) \geq \frac{4}{n-1} \quad \forall k \neq l \in \mathit{supp}(\mathbf{x})\}$$

where $\rho(\cdot,\cdot)$ is a wrap-around distance function [142] such that for $\forall \mu_1, \mu_2 \in [0,1]$, we have $\rho(\mu_1,\mu_2) \triangleq \min(|\mu_1 - \mu_2|, |\mu_1 + 1 - \mu_2|, |\mu_2 + 1 - \mu_1|)$

The following theorem shows that the sequence of iterates produced by Algorithm 1 has a converging subsequence, and whenever $\mathbf{x}^\star \in \Delta^+_{\text{sep}}$, the limit of this convergent subsequence produces a stable estimate of $\mathbf{x}^\star$ (and in particular, exactly recovers $\mathbf{x}^\star$ in absence of noise).

Theorem 6 *Given any non-increasing positive sequence $\{\epsilon_n\}$ and a feasible initial point $\mathbf{x}_0$, the solution sequence $\{\mathbf{x}_n\}$ of Algorithm 1 has a convergent subsequence which converges to a feasible point $\mathbf{x}^\#$ of (P_2). Furthermore, if $\mathbf{x}^\star \in \Delta^+_{sep}$, the limit $\mathbf{x}^\#$ obeys*

$$\|\mathbf{x}^\# - \mathbf{x}^\star\|_1 \leq C \left(\frac{N}{n-1}\right)^2 \delta_1 \tag{3.5}$$

where C is a positive constant.

Proof. Notice that

$$\begin{aligned}
&\sum_{i=0}^{N-1}(x_{n+1,i}+\epsilon_{n+1})^{1/2}\\
&\overset{(a)}{\leq}\sum_{i=0}^{N-1}\frac{(x_{n+1,i}+\epsilon_n)^{1/2}}{(x_{n,i}+\epsilon_n)^{1/4}}(x_{n,i}+\epsilon_n)^{1/4}\\
&\overset{(b)}{\leq}\left[\sum_{i=0}^{N-1}\frac{(x_{n+1,i}+\epsilon_n)}{(x_{n,i}+\epsilon_n)^{1/2}}\right]^{1/2}\left[\sum_{i=0}^{N-1}(x_{n,i}+\epsilon_n)^{1/2}\right]^{1/2}\\
&\overset{(c)}{\leq}\left[\sum_{i=0}^{N-1}\frac{(x_{n,i}+\epsilon_n)}{(x_{n,i}+\epsilon_n)^{1/2}}\right]^{1/2}\left[\sum_{i=0}^{N-1}(x_{n,i}+\epsilon_n)^{1/2}\right]^{1/2}\\
&=\sum_{i=0}^{N-1}(x_{n,i}+\epsilon_n)^{1/2}
\end{aligned}$$

where (a) is due to $\epsilon_{n+1}\leq\epsilon_n$, (b) follows from Hölder's inequality and (c) is true because $\mathbf{x}_{n+1}$ is the optimal solution of (P_3). We further use the fact that [163]

$$\|\mathbf{x}_n\|_\infty\leq\left[\sum_{i=0}^{N-1}(x_{n,i}+\epsilon_n)^{1/2}\right]^2\leq\left[\sum_{i=0}^{N-1}(x_{0,i}+\epsilon_0)^{1/2}\right]^2$$

This shows that the sequence $\{\mathbf{x}_n\}$ is bounded, and thus it has a converging subsequence. Additionally, the feasible set of (P_3) is the intersection of non-negative orthant and closed l_p ball (where $p=1,2$) and hence any cluster point of $\{\mathbf{x}_n\}$ will be feasible [211].

To prove the second part, we use the following fact about $\mathbf{Q}$ from [61]. Let $\mathbf{v}=\mathbf{x}^{\#}-\mathbf{x}^{\star}$, and $\mathcal{T}_{\mathbf{v}}=\{l|v_l<0,0\leq l\leq N-1\}$. If $\mathbf{x}^{\star}\in\Delta_{sep}^{+}$, there exists $\mathbf{q}\in\mathbb{R}^N$ and $c\left(\frac{n-1}{N}\right)^2\leq\eta<1$ where $c=0.0036$, such that $\mathbf{Qq}=\mathbf{q}$ and [61]

$$q_l=-\eta\quad\text{If } l\in\mathcal{T}_{\mathbf{v}};\quad\eta<q_l<1-\eta\quad\text{otherwise}$$

Given the existence of such a $\mathbf{q}$, we have

$$
\begin{aligned}
&|\mathbf{q}^T\mathbf{v}| = |(\mathbf{Q}\mathbf{q})^T\mathbf{v}| = |\mathbf{q}^T\mathbf{Q}\mathbf{v}| \leq \|\mathbf{q}\|_\infty \|\mathbf{Q}\mathbf{v}\|_1 \\
&\leq (1-\eta)\|\mathbf{Q}\mathbf{x}^{\#} - \mathbf{Q}\mathbf{x}^{\star}\|_1 \\
&\leq (1-\eta)\left(\|\mathbf{Q}\mathbf{x}^{\#} - \mathbf{y}\|_1 + \|\mathbf{Q}\mathbf{x}^{\star} - \mathbf{y}\|_1\right) \leq 2(1-\eta)\delta_1
\end{aligned}
$$

On the other hand, we also have

$$
|\mathbf{q}^T\mathbf{v}| = |\sum_{l=0}^{N-1} q_l v_l| = \sum_{l=0}^{N-1} q_l v_l \geq \eta\|\mathbf{v}\|_1
$$

The proof completes by using $\eta = c\left(\frac{n-1}{N}\right)^2$.

3.3 Background on Sparse Arrays and Correlation-Aware Support Recovery

We consider the MMV model introduced in (3.1). Following [16, 168, 181, 182], we make the following assumptions on the correlation of $\mathbf{X}_L$:

- **[(A1)] Uncorrelated Source Signals.** The source signals $\mathbf{x}[l], 1 \leq l \leq L$ are assumed to be zero-mean, independent and identically distributed (i.i.d.) random vectors. Furthermore, the non-zero elements of $\mathbf{x}[l]$ are assumed to be statistically uncorrelated, i.e.,

$$
\mathbb{E}(\mathbf{x}[l]) = \mathbf{0}, 1 \leq i \leq L
$$

$$
\mathbb{E}\Big([\mathbf{x}[l]]_i [\mathbf{x}[m]]_k^*\Big) = p_i^\star \delta(i-k)\delta(l-m), \quad i,k \in \mathcal{S}
$$

where $p_i^\star$ is the power of ith component, $\delta(\cdot)$ is the Kronecker delta function and * denotes conjugate.

- **[(A2)] Noise.** The additive noise vectors $\{\mathbf{w}[l]\}_{l=1}^L$ are statistically uncorrelated with $\mathbf{X}_L$ and distributed as

$$\mathbf{w}[l] \overset{i.i.d}{\sim} \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$$

  In our analysis, σ^2 is assumed to be *unknown*.

- **[(A3)] Measurement Matrix.** Throughout the paper, the measurement matrix $\mathbf{A}$ is assumed to be a fixed deterministic matrix, representing the array manifold of a sparse array (as described later). The theoretical guarantees of this paper will be developed for such a deterministic $\mathbf{A}$. [4]

3.3.1 Fundamental Limits of Correlation-Aware Support Recovery from MMV

The goal in MMV problems is to detect the common support $\mathcal{S}$ from $\mathbf{Y}_L$ [16, 147, 182]. Most MMV algorithms achieve this by first recovering $\mathbf{X}_L$ and then detecting its common support [171, 174, 201]. The need to exactly recover $\mathbf{X}_L$ first automatically restricts the size of recoverable support to $|\mathcal{S}| < M$ [170–173]. In particular, it is well-known that $\mathbf{X}_L$ can be exactly recovered if [172, 173]

$$s < \frac{\text{Krank}(\mathbf{A}) + \text{rank}(\mathbf{Y}_L)}{2} \tag{3.6}$$

[4] Our results deviate from usual guarantees in compressed sensing which rely on the measurement matrix $\mathbf{A}$ being random and following appropriate distributions [96].

Since Krank$(\mathbf{A}) \leq M$ and rank$(\mathbf{Y}_L) \leq M$, (3.6) implies that $s < M$. However, a recent line of work [16, 135, 168] has shown that by exploiting correlation priors on $\mathbf{X}_L$ (given by Assumption **(A1)**), it is possible to recover $\mathcal{S}$ from the correlation matrix $\mathbf{R_{yy}} = \mathbb{E}(\frac{1}{L}\mathbf{Y}_L\mathbf{Y}_L^H)$ of the measurements *without having to first estimate* $\mathbf{X}_L$. The idea behind such "correlation-aware" techniques for support recovery is also closely related to an important body of research on compressive covariance sampling [4, 16, 38, 40, 41] where the goal is to recover the covariance matrix of $\mathbf{X}_L$ from its compressed sketch $\mathbf{R_{yy}}$. The main idea behind these methods is to design the measurement matrix appropriately so that it becomes possible to recover supports of size as large as $s = \Theta(M^2)$. To see this, notice that Assumption **(A1)** implies

$$\mathbf{R_{yy}} = \mathbf{A}\text{diag}(\mathbf{p}^\star)\mathbf{A}^H + \sigma^2\mathbf{I} \tag{3.7}$$

where $\mathbf{p}^\star \in \mathbb{R}^N$ is a *sparse vector* satisfying

$$[\mathbf{p}^\star]_i = \begin{cases} p_i^\star, & i \in \mathcal{S} \\ 0 & \text{otherwise} \end{cases}$$

In other words, the support of $\mathbf{p}^\star$ is exactly $\mathcal{S}$. Hence, we can recover $\mathcal{S}$ by first finding a sparse $\mathbf{p}^\star$ that satisfies (3.7). Assuming we know [5] the noise power σ^2, this amounts to solving the following l_0 quasi-norm minimization problem

$$\min_{\mathbf{z} \succeq \mathbf{0}} \|\mathbf{z}\|_0 \quad s.t \quad (\mathbf{A}^* \odot \mathbf{A})\mathbf{z} = \text{vec}(\mathbf{R_{yy}} - \sigma^2\mathbf{I}) \qquad (P0)_{Co}$$

where we used the fact that vec$(\mathbf{R_{yy}}) = (\mathbf{A}^* \odot \mathbf{A})\mathbf{p}^\star + \sigma^2vec(\mathbf{I})$. Using standard results from compressed sensing [96], it follows that the solution to $(P0)_{Co}$ will be unique, and identical to $\mathbf{p}^\star$

[5] This assumption will be relaxed later.

if

$$s \leq \frac{1}{2}\text{Krank}(\mathbf{A}^* \odot \mathbf{A}) \tag{3.8}$$

Hence, the Kruskal-rank of $\mathbf{A}^* \odot \mathbf{A}$ (and not of $\mathbf{A}$) controls the maximum size of support that can be recovered by solving $(P0)_{Co}$. Since the matrix $\mathbf{A}^* \odot \mathbf{A}$ is of size $M^2 \times N$, (i.e. it has the same number of columns as $\mathbf{A}$ but its number of rows is M times larger), its Kruskal-rank can be as large as $\Theta(M^2)$. To understand what kind of measurement matrices permit such large Kruskal-rank, we turn to the scenario when $\mathbf{A}$ represents the manifold of a suitable sparse array, such as a nested or coprime array [17, 45].

3.3.2 Sparse Arrays and Difference Sets

In many imaging problems using sensor arrays, the measurement matrix $\mathbf{A}$ has a Fourier structure, since the acquired measurements naturally correspond to Fourier transform of the underlying signal of interest [23, 175, 193, 194]. For example, in DOA estimation and super-resolution microscopy, [6] the location of each point source of interest (which can represent a far-field target or a molecule) can be mapped to a spatial frequency $\omega \in [0, 2\pi)$. By discretizing the range of spatial frequencies into N bins as $\omega_n = 2\pi n/N, 0 \leq n \leq N-1$, one obtains the measurement matrix $\mathbf{A}$ as [7, 61, 120, 175, 176, 199, 200]

$$[\mathbf{A}]_{m,n} = e^{-jd_m \frac{2\pi n}{N}}, \quad 1 \leq m \leq M, 0 \leq n \leq N-1 \tag{3.9}$$

where d_m is an integer denoting the location of the mth sensor (or pixel), normalized with respect to the wavelength of the impinging narrowband waveform. This discretized model for source localization has been widely used, especially in conjunction with MMV models

[6] For ease of exposition, we consider a one-dimensional source localization problem, although the results can be directly extended to two dimensions.

[16, 120, 175, 177, 183, 199, 200].

Under assumption $(\mathbf{A1})$, we know from (3.8) that the Kruskal-rank of $\mathbf{A}_{\text{KR}} = \mathbf{A}^* \odot \mathbf{A}$ dictates the size of recoverable support. Upon close inspection, it can be verified that $\mathbf{A}_{\text{KR}}$ is given by

$$[\mathbf{A}_{\text{KR}}]_{(m-1)M+l,n} = e^{j(d_m - d_l)\frac{2\pi n}{N}}, \quad 1 \leq m, l \leq M \tag{3.10}$$

Comparing (3.9) and (3.10), each column of $\mathbf{A}_{\text{KR}}$ can be thought of as the steering vector (corresponding to a candidate spatial frequency) of a *virtual* array (also called the difference co-array) whose elements are located at positions given by the following difference set [32, 41, 45, 199, 204]

$$\mathbb{D}_{\text{diff}} \triangleq \{d_m - d_l, 1 \leq m, l \leq M\} \tag{3.11}$$

The structure of $\mathbb{D}_{\text{diff}}$ and the number of its distinct elements directly control the Kruskal-rank of $\mathbf{A}_{\text{KR}}$, and therefore the largest size of recoverable support [16]. If we carefully design the set of sensor location $\mathbb{S} = \{d_m\}_{m=1}^{M}$, then it is possible to ensure that its difference set $\mathbb{D}_{\text{diff}}$ has $\Theta(M^2)$ distinct *consecutive integers*. In other words, the virtual array (whose element locations are given by $\mathbb{D}_{\text{diff}}$) will have a ULA segment with $\Theta(M^2)$ elements. This is precisely the main idea behind the design of sparse arrays. Sparse arrays such as minimum redundancy arrays [204], nested arrays [45, 124] , coprime arrays [17], and their generalizations and extensions [199] are carefully designed so that their difference sets satisfy this property. We now briefly review nested and coprime arrays and compare their difference sets with the popular ULA.

1. **ULA**: For a uniform linear array,

$$d_i = i - 1, \quad 1 \leq i \leq M.$$

It can be easily seen that in this case

$$\mathbb{D}_{\text{diff}} = \{m, -M+1 \leq m \leq M-1\}$$

and hence $\mid \mathbb{D}_{\text{diff}} \mid = 2M-1$. In fact, it can be verified that in this case $\text{Krank}(\mathbf{A}_{\text{KR}}) = 2M-1$. Hence, (3.8) implies the following upper bound on the size of the recoverable support $s < M$. This is consistent with the popular belief that the number of resolvable directions is ultimately upper-bounded by the number of sensors or antennas.

2. **Nested Array**: For a nested array [45], the set of sensor locations is given by

$$\begin{aligned} &\{d_i\}_{i=1}^{M} = \{m-1, 1 \leq m \leq M/2\} \cup \\ &\{(M/2+1)n - 1, 1 \leq n \leq M/2\} \end{aligned} \tag{3.12}$$

The difference set is

$$\begin{aligned} &\mathbb{D}_{\text{diff}} = \\ &\{m, -M^2/4 - M/2 + 1 \leq m \leq M^2/4 + M/2 - 1\} \end{aligned}$$

Since $\mathbb{D}_{\text{diff}}$ contains *all consecutive integers* within the above range, $\mathbf{A}_{\text{KR}}$ is a Vandermonde matrix with $\frac{M^2}{2} + M - 1$ distinct rows. Hence,

$$\text{Krank}(\mathbf{A}_{\text{KR}}) = \frac{M^2}{2} + M - 1$$

Hence the Kruskal-rank of $\mathbf{A}_{\text{KR}}$ for nested arrays is as large as $\Theta(M^2)$, whereas that for ULA is $2M-1$.

3. **Coprime Array**: A coprime array is an array of $M = 2M_1 + M_2 - 1$ sensors, where M_1

and M_2 are coprime integers. The locations of the sensors are given by

$$\{d_i\}_{i=1}^{M} = \\ \{M_1 m, 0 \leq m \leq M_2 - 1\} \cup \{M_2 n, 1 \leq n \leq 2M_1 - 1\}.$$

The difference set in this case contains all integers in the range

$$\{n, -M_1 M_2 \leq n \leq M_1 M_2\}.$$

Thus, the rows of $\mathbf{A}_{\text{KR}}$ contain a Vandermonde submatrix with $2M_1M_2 + 1$ distinct rows, implying

$$\text{Krank}(\mathbf{A}_{\text{KR}}) \geq 2M_1M_2 + 1.$$

For $M_2 = M_1 + 1$, we have $M = 3M_1$ and $\text{Krank}(\mathbf{A}_{\text{KR}}) = \Theta(M^2)$.

From the above discussions, it is clear that for nested and coprime arrays, we can find an integer M_{ca} such that (i) $\mathbb{D}_{\text{diff}}$ contains all *consecutive integers* in the range $-M_{ca}$ and M_{ca}, *and* (ii) M_{ca} is as large as $\Theta(M^2)$. The same is true for generalizations of nested and coprime arrays, such as super nested [198], and generalized coprime arrays [199]. More generally, for all these arrays, $\mathbb{D}_{\text{diff}}$ contains a ULA segment given by the set

$$\mathbb{U} = \{n, -M_{ca} \leq n \leq M_{ca}\}, \quad M_{ca} = \Theta(M^2) \tag{3.13}$$

This prompts us to unify the treatment of nested arrays, coprime arrays (and their extensions) by defining a family of "Order-Optimal Sparse Arrays (OOSA)":

Definition 4 (OOSA) *An array is an OOSA if its difference set $\mathbb{D}_{\text{diff}}$ contains a subset $\mathbb{U}$ of the form (3.13) with $M_{ca} = \Theta(M^2)$.*

The criterion for any array to be an OOSA is that its difference set should contain a ULA segment of size $\Theta(M^2)$. Since the maximum size of $\mathbb{D}_{\text{diff}}$ can only be $\Theta(M^2)$, these arrays are called "Order-Optimal" since the ULA segment $\mathbb{U}$ is of maximum possible size order-wise with respect to M.

Remark9. Virtual Array Manifold corresponding to $\mathbb{U}$: *For the remainder of this paper, we will assume that* $\mathbf{A}$ *represents the measurement matrix corresponding to an OOSA (which includes nested, coprime and their extensions). From (3.9) and the structure of the difference set (3.11) of any OOSA, it can be seen that* $\mathbf{A}_{KR} \in \mathbb{C}^{M^2 \times N}$ *contains a Vandermonde sub-matrix* $\mathbf{A}_{KR}^{\mathbb{U}} \in \mathbb{C}^{2M_{ca}+1 \times N}$ *with* $2M_{ca}+1$ *rows, whose elements are given by*

$$[\mathbf{A}_{KR}^{\mathbb{U}}]_{m+M_{ca},n} = e^{-j2\pi mn/N}$$
$$-M_{ca} \leq m \leq M_{ca}, 0 \leq n \leq N-1 \tag{3.14}$$

In other words $\mathbf{A}_{\text{KR}}^{\mathbb{U}}$ represents the *virtual array manifold* corresponding to the largest central ULA segment $\mathbb{U}$ in the difference set of an OOSA. Fig. 3.1 shows non-negative halves of $\mathbb{D}_{\text{diff}}$ and $\mathbb{U}$ (denoted $\mathbb{D}_{\text{diff}}^{+}$ and $\mathbb{U}^{+}$ respectively) for ULA, nested and coprime arrays.

3.3.3 Limitations in Existing Guarantees With Finite Number of Measurement Vectors when $|\mathcal{S}| > M$

It is clear from the above discussion that an OOSA can ensure unique support recovery in the regime $s = \Theta(M^2)$ via solving $(P0)_{Co}$. However, this requires *exact knowledge* of $\mathbf{R}_{\mathbf{yy}}$, which is only available as $L \to \infty$. In practice, with a finite number (L) of measurement vectors, we use the sample covariance matrix $\hat{\mathbf{R}}_{\mathbf{yy}} \triangleq \frac{1}{L}\mathbf{Y}_L\mathbf{Y}_L^H$ as an estimate of $\mathbf{R}_{\mathbf{yy}}$. As a practical and computationally tractable approach to correlation-aware support recovery, we

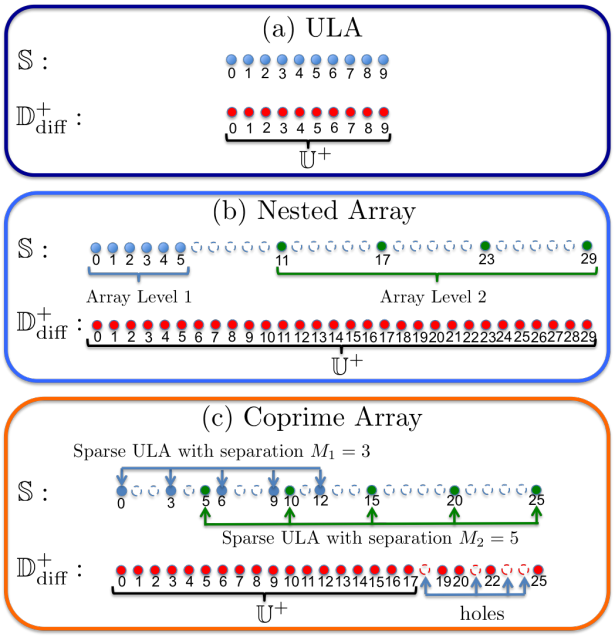


Figure 3.1: Examples of ULA, nested array ($M = 10$) and coprime array ($M_1 = 3, M_2 = 5$) with their difference co-arrays (non-negative part) respectively.

proposed to solve the following convex problem, also referred to as "correlation-aware" LASSO (or, Co-LASSO) [16]

$$\min_{\mathbf{z} \succeq \mathbf{0}} \frac{1}{2} \|\hat{\mathbf{r}}_{\mathbf{y}} - \mathbf{A}_{\text{KR}} \mathbf{z}\|_2^2 + \lambda \|\mathbf{z}\|_1 \quad \text{(Co-LASSO)}$$

where $\hat{\mathbf{r}}_{\mathbf{y}} \triangleq \text{vec}(\hat{\mathbf{R}}_{\mathbf{yy}})$. The first term in the objective function of (Co-LASSO) captures the mismatch between ideal covariance matrix and its estimate, while the second term promotes sparsity by replacing l_0 quasi-norm with its convex surrogate l_1 norm. While LASSO is a well-studied problem that with provable guarantees for support recovery in the regime $|\mathcal{S}| < M$ [101], *it is currently unknown if (Co-LASSO) can provably recover support of size $s > M$ with high probability in L.* Existing analysis of Co-LASSO is based on the mutual coherence of $\mathbf{A}_{\text{KR}}$ (Theorem 7 in [16]), which scales as $O(1/M)$. As a consequence, current non-asymptotic guarantees for (Co-LASSO) are only valid in the regime $s < M$ and they fail to predict its performance when $s = \Theta(M^2)$. *This is a major current limitation of existing analysis of*

correlation-aware support recovery with finite data. Evidently, the limitation is due to the fact that these guarantees rely on mutual coherence, which is known to yield sub-optimal sample complexity. The limitation can be potentially overcome if one can establish Restricted Isometry Property (RIP) [96] of the Khatri-Rao product matrix $\mathbf{A}^* \odot \mathbf{A}$ in the regime $s > M$. However, establishing the RIP of $\mathbf{A}_{\text{KR}}$ when $\mathbf{A}$ is a *deterministic matrix* representing the array manifold of a sparse array is *a challenging open problem*, since construction of deterministic matrices satisfying RIP even in the well-studied regime $s \leq M$ is currently unknown [96].

3.4 The Role of Non-Negativity and Universal Recovery Guarantees with Order-Optimal Sparse Arrays (OOSA)

In this chapter, we will overcome the above mentioned bottleneck by utilizing the fact that besides being sparse, $\mathbf{p}^\star$ is also *non-negative*. By utilizing (i) the non-negativity of $\mathbf{p}^\star$ (or positivity of its non-zero elements) and (ii) the geometry of sparse arrays as captured by the matrix $\mathbf{A}_{\text{KR}} = \mathbf{A}^* \odot \mathbf{A}$, we will show that it is possible to obtain universal guarantees for stable reconstruction of $\mathbf{p}^\star$ (as a function of L) that will be obeyed by a large class of correlation-aware support recovery algorithms (to be introduced later) in the regime of interest $s = \Theta(M^2)$.

3.4.1 Why Positivity Alone Suffices

In order to understand why the positivity of $\mathbf{p}^\star$ can alone ensure its exact reconstruction from $M = \Omega(\sqrt{s})$ measurements (without even explicitly using its sparsity), we first revisit the case when we have perfect knowledge of the covariance matrix $\mathbf{R}_{\mathbf{yy}} = \mathbb{E}\Big(\frac{1}{L}\mathbf{Y}_L\mathbf{Y}_L^H\Big)$ of the measurements. As discussed earlier, the non-negative l_0 quasi-norm minimization problem $(P0)_{Co}$ from [16] can uniquely recover $\mathbf{p}^\star$ with support as large as $s = \Theta(M^2)$, when $\mathbf{A}$ represents the manifold of a suitable sparse array. We now strengthen this result from [16] by showing

that in this case, there is actually no need to minimize the l_0 quasi-norm, (or, explicitly enforce sparsity) to recover $\mathbf{p}^\star$. As long as $s \leq M_{ca}$, non-negativity of $\mathbf{p}^\star$ alone is sufficient to uniquely recover it from $\mathbf{R_{yy}}$ *even when the noise power* σ^2 *is unknown*, as given by the following result

Theorem 7 *Suppose we are given the exact covariance matrix* $\mathbf{R_{yy}} = \mathbb{E}(\frac{1}{L}\mathbf{Y}_L\mathbf{Y}_L^H)$. *Assuming that the noise power* σ^2 *is unknown, it is possible to exactly recover* $\mathbf{p}^\star$ *(using a variant of the Co-array MUSIC algorithm [126]) as long as* $s \leq M_{ca}$.

Proof. The proof uses an algorithm similar to Co-array MUSIC [45, 126] to guarantee unique recovery. Please See Appendix 3.11.1 for details.

Remark 10. *When* $N > M^2$ *(i.e. when* $\mathbf{A}_{KR}$ *is a fat matrix),* $vec(\mathbf{R_{yy}}) = \mathbf{A}_{KR}\mathbf{z} + \sigma^2 vec(\mathbf{I})$ *represents an underdetermined system of equations in* $\mathbf{z}$ *with possibly infinite solutions. However, Theorem 7 shows that by utilizing the fact that* $\mathbf{p}^\star$ *is non-negative, it is possible to exactly recover it even when the noise power is unknown. Therefore, there is no need to minimize the* l_0 *quasi-norm in* $(P0)_{Co}$*, or explicitly utilize the sparsity of* $\mathbf{p}^\star$*. Since for an OOSA,* $M_{ca} = \Theta(M^2)$*, we can uniquely recover supports of size as large as* $\Theta(M^2)$.

Remark 11. *The fact that non-negative constraints (or more generally, conic constraints) can lead to unique signal recovery in noiseless underdetermined problems has been previously utilized in compressed sensing [100, 207, 208], super-resolution imaging [61] and phase retrieval [212]. However, these results do not exploit the structure of the correlation matrix* $\mathbf{R_{yy}}$ *of the measurements (especially when acquired using a sparse array), and therefore cannot guarantee the recovery of supports of size* $s > M$*. In contrast, our result shows that using sparse arrays and the non-negativity of* $\mathbf{p}^\star$*, it is possible to uniquely recover the support from* $\mathbf{R_{yy}}$ *(without minimizing any objective) even when* $s = \Theta(M^2)$.

3.4.2 Stable Recovery in the regime $|\mathcal{S}| > M$ with estimated covariance matrix: Preliminaries

Theorem 7 shows that it is possible to exactly recover $\mathbf{p}^\star$ even when $s \gg M$ assuming we have *exact knowledge of the correlation matrix* $\mathbf{R_{yy}}$. In practice however, we can only *estimate* $\mathbf{R_{yy}}$ from a finite number of (L) of multiple measurement vectors (MMV) $\mathbf{y}[1], \cdots \mathbf{y}[L]$. An important question therefore is

"Can the positivity of the non-zero elements of $\mathbf{p}^\star$ *still be exploited in a useful way to derive performance guarantees for support recovery algorithms in the regime* $s = \Theta(M^2)$*?"*

A main contribution of this paper is to show that this is indeed possible, *even without explicitly utilizing the sparsity of* $\mathbf{p}^\star$. We first model the deviation of $\hat{\mathbf{R}}_{\mathbf{yy}}$ from the ideal covariance matrix $\mathbf{R_{yy}}$ as follows. Let $\hat{\mathbf{r}}_{\mathbf{y}} = \text{vec}(\hat{\mathbf{R}}_{\mathbf{yy}}) \in \mathbb{C}^{M^2}$ and $\mathbf{r}_{\mathbf{y}} \triangleq \text{vec}\Big(\mathbf{R_{yy}} - \sigma^2\mathbf{I}\Big) = \mathbf{A}_{\text{KR}}\mathbf{p}^\star$. Define

$$\Delta_{\mathbf{r}} \triangleq \hat{\mathbf{r}}_{\mathbf{y}} - \mathbf{r}_{\mathbf{y}} = \underbrace{\text{vec}(\hat{\mathbf{R}}_{\mathbf{yy}} - \mathbf{R_{yy}})}_{\Delta_f} + \sigma^2 \text{vec}(\mathbf{I}) \tag{3.15}$$

where Δ_f represents the finite sample correlation estimation error and the second term accounts for the *unknown* noise power σ^2. We will first develop stability guarantees in terms of $\Delta_{\mathbf{r}}$, and in Sec.3.5.1, we will utilize the statistical model of Δ_f (in terms of L) and develop probabilistic guarantees as a function of L.

Extracting the ULA segment from Difference Co-Array: From Remark 9 and (3.14), we know that one can construct a sub-vector $\mathbf{r}_{\mathbf{y}}^{\mathbb{U}} \in \mathbb{C}^{2M_{ca}+1}$ such that $\mathbf{r}_{\mathbf{y}}^{\mathbb{U}} = \mathbf{A}_{\text{KR}}^{\mathbb{U}}\mathbf{p}^\star$. We can extract the same $2M_{ca}+1$ indices from $\hat{\mathbf{r}}_{\mathbf{y}}$ to obtain the vector $\hat{\mathbf{r}}_{\mathbf{y}}^{\mathbb{U}}$ such that

$$\hat{\mathbf{r}}_{\mathbf{y}}^{\mathbb{U}} = \mathbf{r}_{\mathbf{y}}^{\mathbb{U}} + \Delta_{\mathbf{r}}^{\mathbb{U}} = \mathbf{A}_{\text{KR}}^{\mathbb{U}}\mathbf{p}^\star + \Delta_{\mathbf{r}}^{\mathbb{U}} \tag{3.16}$$

where $\Delta_{\mathbf{r}}^{\mathbb{U}}$ is composed of the corresponding indices of $\Delta_{\mathbf{r}}$.

Remark 12. *In this paper, we have only exploited the central ULA segment $\mathbb{U}$ of the difference coarray $\mathbb{D}_{diff}$. Consequently, the size of the recoverable support is governed by M_{ca} which represents the size of $\mathbb{U}$. For certain sparse arrays, such as coprime array, $\mathbb{D}_{diff}$ contains holes. In recent literature [191, 192], it has been shown that the full degrees of freedom of $\mathbb{D}_{diff}$ can be exploited for DOA estimation, if we allow array motion, or solve a l_1 regularized LASSO for on-grid source detection. Another option is to perform co-array interpolation, for which we developed a unified analysis framework in [13]. In future, we plan to utilize these results to exploit the full degrees of freedom of $\mathbb{D}_{diff}$ for support estimation.*

3.4.3 Universal Upper Bounds on Error with Non Negative Constraint when $|\mathcal{S}| > M$

Given any $\mathbf{p}^{\#} \in \mathbb{C}^N$ such that $\mathbf{p}^{\#} \succeq \mathbf{0}$, we will now derive a *universal upper bound* on the estimation error $\|\mathbf{p}^\star - \mathbf{p}^{\#}\|_2$ in terms of $\|\mathbf{A}^{\mathbb{U}}_{\text{KR}}(\mathbf{p}^\star - \mathbf{p}^{\#})\|_2$. To begin, we define a set of *non-negative sparse signals* whose supports obey a certain "separation condition". Such separation condition (without non-negative constraint) has been utilized in recent times for analyzing the performance of super-resolution imaging algorithms [4, 59, 61, 141].

Definition 5 *(Set of Non-negative Signals Obeying Separation Condition) Given N and M_{ca}, define the set $\mathcal{P}^+_{sep}$*

$$\mathcal{P}^+_{sep} \triangleq$$
$$\{\mathbf{p} \in \mathbb{C}^N \mid \mathbf{p} \succeq \mathbf{0}, \phi(\frac{k}{N}, \frac{l}{N}) \geq \frac{2}{M_{ca}}, \forall k \neq l \in \mathit{Supp}(\mathbf{p})\}$$

where $\phi(\cdot,\cdot) : \mathbb{R}^2 \to \mathbb{R}_+$ is a wrap-around distance function [141] such that for $\forall \mu_1, \mu_2 \in [0,1]$

$$\phi(\mu_1, \mu_2) \triangleq \min(|\mu_1 - \mu_2|, |\mu_1 + 1 - \mu_2|, |\mu_2 + 1 - \mu_1|)$$

The set $\mathcal{P}_{\text{sep}}^{+}$ consists of non-negative sparse signals such that the indices belonging to their support are separated by at least $\frac{2N}{M_{ca}}$. [7] We present our main result in the following theorem, which holds for any $\mathbf{p}^{\star} \in \mathcal{P}_{\text{sep}}^{+}$. Our proof (presented in Appendix 3.11.2) utilizes a key Lemma from [61] where it was used to analyze a specific l_1-norm minimization algorithm. The separation condition used in [61] automatically restricted the sparsity to be smaller than M (i.e. $s < M$). Theorem 25 significantly generalizes this result by showing that it can be used to derive performance guarantees for a larger class of algorithms, even in the regime $s > M$ by exploiting the difference co-array of an OOSA.

Theorem 8 *Suppose* $\mathbf{p}^{\star} \in \mathcal{P}_{sep}^{+}$, $M_{ca} \geq 128$ [8] *and* $N \geq 3.03(2M_{ca}+1)$. *Given any* $\mathbf{p}^{\#} \succeq \mathbf{0}$, *we have*

$$\|\mathbf{p}^{\#} - \mathbf{p}^{\star}\|_1 \leq \frac{1-\rho}{\rho} \|\mathbf{A}_{KR}^{\mathbb{U}}(\mathbf{p}^{\#} - \mathbf{p}^{\star})\|_2 \tag{3.17}$$

where $\rho = 0.0295\left(\frac{M_{ca}}{N}\right)^2$.

Proof. See Appendix 3.11.2 Notice that Theorem 25 only requires $\mathbf{p}^{\star}$ to obey the separation condition; there is no constraint on the other vector $\mathbf{p}^{\#}$ except requiring it to be non-negative. We now show that the upper bound (3.17) can be used to bound the estimation error of *the following class of optimization problems*, where $\mathbf{p}^{\#}$ is interpreted as the optimal solution to the respective problem.

[7] The separation condition can be relaxed to the more general Rayleigh Regularity condition [61]. However, for ease of exposition, in this paper we assume that $\mathbf{p}^{\star}$ obeys separation condition.

[8] For nested array, this holds if $M \geq 22$.

3.4.4 Stability Guarantees for A Class of Correlation-Aware Algorithm

As direct application of Theorem 25, consider the following class of optimization problems for estimating $\mathbf{p}^\star$.

$$\min_{\mathbf{z}} f(\mathbf{z}) \quad (P_{\text{Co-den}})$$
$$\|\hat{\mathbf{r}}_{\mathbf{y}}^{\mathbb{U}} - \mathbf{A}_{\text{KR}}^{\mathbb{U}}\mathbf{z}\|_2 \leq \epsilon, \quad \mathbf{z} \succeq \mathbf{0}$$

The non-negative constraint is natural since the optimization variable $\mathbf{z}$ corresponds to the unknown signal powers. The parameter $\epsilon > 0$ captures the effect of (finite sample) covariance estimation error. [9] The following corollary to Theorem 25 shows how (3.17) can be used to obtain stability guarantees for $(P_{\text{Co-den}})$.

Corollary 4 *Suppose the objective $f(\mathbf{z})$ of the optimization problem $(P_{\text{Co-den}})$ is a continuous function and $\epsilon \geq \|\Delta_{\mathbf{r}}^{\mathbb{U}}\|_2$. Assume the conditions of Theorem 25 hold. If $\mathbf{p}^{\#}$ is a solution to $(P_{\text{Co-den}})$, then*

$$\|\mathbf{p}^{\#} - \mathbf{p}^\star\|_1 \leq \frac{1-\rho}{\rho}(\|\Delta_{\mathbf{r}}^{\mathbb{U}}\|_2 + \epsilon) \tag{3.18}$$

Proof. Since $\epsilon \geq \|\Delta_{\mathbf{r}}^{\mathbb{U}}\|_2$, the feasible set of $(P_{\text{Co-den}})$ contains $\mathbf{p}^\star$ and is non-empty. Moreover, since the objective function $f(\mathbf{z})$ of $(P_{\text{Co-den}})$ is continuous and the constraint set $\{\mathbf{z} \succeq \mathbf{0}, \|\hat{\mathbf{r}}_{\mathbf{y}}^{\mathbb{U}} - \mathbf{A}_{\text{KR}}^{\mathbb{U}}\mathbf{z}\|_2 \leq \epsilon\}$ is compact, the minimizer $\mathbf{p}^{\#}$ exists [211]. The result then directly follow from Theorem 25 via the triangle inequality

$$\|\mathbf{A}_{\text{KR}}^{\mathbb{U}}(\mathbf{p}^\star - \mathbf{p}^{\#})\|_2 \leq \|\hat{\mathbf{r}}_{\mathbf{y}}^{\mathbb{U}} - \mathbf{A}_{\text{KR}}^{\mathbb{U}}\mathbf{p}^\star\|_2 + \|\hat{\mathbf{r}}_{\mathbf{y}}^{\mathbb{U}} - \mathbf{A}_{\text{KR}}^{\mathbb{U}}\mathbf{p}^{\#}\|_2$$

[9] We call this optimization framework $(P_{\text{Co-den}})$ (where "Co-den" is abbreviation for "correlation-aware denoising") since it resembles the popular basis pursuit denoising [26], except that $\hat{\mathbf{r}}_{\mathbf{y}}^{\mathbb{U}}$ is the estimated correlation of the measurements.

Remark 13. *Notice that* $(P_{\text{Co-den}})$ *actually represents a* class of optimization problems *which can be either convex or non-convex depending on the choice of* $f(\mathbf{z})$*. The stability guarantee from Corollary 4 applies to the estimate produced by each optimization problem from this class. In fact, the bound (3.18) is independent of the specific choice of* $f(z)$ *and actually applies to any point* $\mathbf{p}^{\#}$ *belonging to the feasible set. However, as demonstrated by the simulations in Sec. 3.9.2, the actual performance of* $(P_{\text{Co-den}})$ *will depend on the choice of* $f(\mathbf{z})$*. In the future, we would like to tailor the universal upper bound in Corollary 4 to specific choices of the objective function* $f(\mathbf{z})$*.*

Remark 14. Exact Reconstruction and Size of Support: *In absence of noise (i.e.* $\sigma = 0$*) and correlation estimation error (i.e.* $\Delta_f = 0$*), Corollary 4 ensures exact recovery of* $\mathbf{p}^{\star}$*. In this case, the only assumption is that* $\mathbf{p}^{\star}$ *should obey the separation condition, which in turn determines the size of the recoverable support as follows. The separation condition dictates that any two indices in* $\mathcal{S}$ *are separated by at least* $2N/M_{ca}$*. This implies that the size of the support must satisfy* $|\mathcal{S}| \leq M_{ca}/2$*. Since* $M_{ca} = \Theta(M^2)$ *for an OOSA, Corollary 4 indicates that it is possible to recover well-separated supports of size* $\Theta(M^2)$*.*

Remark 15. On Error Amplification and Stability: *The upper bound (3.18) on the estimation error* $\|\mathbf{p}^{\#} - \mathbf{p}^{\star}\|_1$ *is an amplified version of the covariance estimation error* $\|\Delta_{\mathbf{r}}^{\mathbb{U}}\|_2$*, where the amplification factor is* $\approx (1/\rho) = c(N/M_{ca})^2$*,* c *being a universal constant. Such amplification of noise during signal reconstruction has been recently observed/analyzed in the context of super-resolution imaging [61], where the goal is to recover an* N*-dimensional sparse signal (with both high and low frequency components) from only* K low-frequency *DFT coefficients. The noise amplification factor (NAF) in such cases is known to scale as* $(N/K)^2$ *[61, 139, 141]. From the structure of* $\mathbf{A}_{KR}^{\mathbb{U}}$ *as given by (3.14), we can easily identify the vector* $\mathbf{A}_{KR}^{\mathbb{U}}\mathbf{p}^{\star}$ *as the* $K = 2M_{ca} + 1$ *low-frequency DFT coefficients (symmetrically located around* 0*) of* $\mathbf{p}^{\star}$*. Hence our error bound is consistent with known results from super-resolution image reconstruction. In Sec. 3.7, we will*

further show that the amplification factor in (3.17) is tight with respect to N, *by constructing specific* $\mathbf{p}^\star$ *and* $\mathbf{p}^\#$ *that attain this bound.*

3.5 Universal Support Recovery via Thresholding and Finite Snapshot Guarantees

Theorem 25 shows that it is possible to perform stable reconstruction of $\mathbf{p}^\star$ from the estimated correlation matrix, by exploiting its non-negativity. However, in many applications (such as direction-of-arrival in radar, source localization using EEG measurements), it is the support $\mathcal{S}$ of $\mathbf{p}^\star$ that carries the information of interest. Hence, it is crucial to understand how the error bound in (3.17) can be utilized to overcome this weakness and ensure support recovery in the regime $|\mathcal{S}| = \Theta(M^2)$. Notice that (3.18) is derived in terms of the covariance estimation error $\|\Delta_{\mathbf{r}}^{\mathbb{U}}\|_2$. Since we estimate the covariance matrix using a finite number L of snapshots, $\Delta_{\mathbf{r}}$ is a random variable whose distribution depends on L. Hence, the guarantees of Theorem 25 are actually *conditioned on* $\Delta_{\mathbf{r}}$. In this section, we will develop probabilistic guarantees on recovering $\mathcal{S}$ in the regime $|\mathcal{S}| > M$, by utilizing the distribution of $\Delta_{\mathbf{r}}$, which will explicitly reveal the role of the number of measurement vectors L. To make the analysis tractable, we make the following additional assumption on the distribution of $\mathbf{y}[l]$ similar to [130, 181, 182]:

- **[(A4)]** The measurements $\{\mathbf{y}[l]\}_{l=1}^{L}$ are zero mean i.i.d complex Gaussian random vectors distributed as $\mathbf{y}[l] \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_{\mathbf{yy}})$.

3.5.1 Simple Hard Thresholding For Support Recovery with Finite L and Unknown σ^2 when $|\mathcal{S}| > M$

To recover $\mathcal{S}$ from the estimate $\mathbf{p}^{\#}$, we propose the following simple element-wise Hard-Thresholding operator $\mathcal{H}_T : \mathbb{R}^N \to \mathbb{R}^N$ (with threshold $T \in \mathbb{R}_+$)

$$[\mathcal{H}_T(\mathbf{x})]_i = \begin{cases} x_i & \text{If } x_i > T \\ 0 & \text{If } x_i \leq T \end{cases}$$

Define

$$p_{\min} \triangleq \min_{i \in \mathcal{S}} p_i^\star \tag{3.19}$$

The following theorem provides explicit conditions on p_{min}, L, as well as the choices of parameters ϵ in $(P_{\text{Co-den}})$ and T, which ensure recovery of $\mathcal{S}$ (with high probability in L) by applying the Hard-Thresholding Operator $\mathcal{H}_T$ on the solution of $(P_{\text{Co-den}})$. The novel contribution of the theorem is that the guarantees continue to hold in the regime $M < |\mathcal{S}| \leq M_{ca}$, where $M_{ca} = \Theta(M^2)$ for any OOSA.

Theorem 9 *Consider the MMV model (3.1) satisfying the assumptions* **(A1-A4)**. *Further assume that* $\mathbf{p}^\star \in \mathcal{P}_{sep}^+$, $M_{ca} \geq 128$, $N \geq 3.03(2M_{ca}+1)$ *and*

$$p_{\min} > 4\frac{1-\rho}{\rho}\sigma^2 \tag{3.20}$$

$$L > \frac{32M^2(\|\mathbf{p}^\star\|_1 + \sigma^2)^2}{(p_{min} - 4\frac{1-\rho}{\rho}\sigma^2)^2}\left(\frac{1-\rho}{\rho}\right)^2 \tag{3.21}$$

If $\mathbf{p}^{\#}$ *is a solution to* $(P_{\text{Co-den}})$ *(where* $f(\mathbf{z})$ *can be any continuous function), then, with probability*

at least $1 - 2e^{-2c\sqrt{L}}$*, we have*

$$\mathit{Supp}\Big(\mathcal{H}_T(\mathbf{p}^{\#})\Big) = \mathcal{S} \tag{3.22}$$

provided we choose the parameter ϵ *of* $(P_{\text{Co-den}})$*, and the threshold* T *from the following ranges:*

$$\sigma^2 + \frac{\sqrt{2}M(\|\mathbf{p}^\star\|_1 + \sigma^2)}{\sqrt{L}} \leq \epsilon < \left(\frac{\rho}{1-\rho}\right)\frac{p_{\min}}{4}, \tag{3.23}$$

and

$$\begin{aligned}&\left(\frac{1-\rho}{\rho}\right)(\epsilon + \frac{\sqrt{2}M(\|\mathbf{p}^\star\|_1 + \sigma^2)}{\sqrt{L}} + \sigma^2) \leq T \\ &< p_{min} - \left(\frac{1-\rho}{\rho}\right)(\epsilon + \frac{\sqrt{2}M(\|\mathbf{p}^\star\|_1 + \sigma^2)}{\sqrt{L}} + \sigma^2)\end{aligned} \tag{3.24}$$

The constants ρ *and* c *are specified in Theorem 25 and Lemma 17 (Appendix 3.11.3) respectively.*

Proof. The proof combines the universal error bound (3.17) with certain concentration inequalities on the finite-sample estimation error $\Delta_{\mathbf{r}}$. Please refer to Appendix 3.11.3 for details.

Remark 16. Guaranteed Support Recovery with finite L **when** $|\mathcal{S}| > M$**:** *Theorem 9 shows that as long as* L *is larger than the threshold (3.21) and the minimum source power* $p_{\min}$ *is large enough compared to the noise power* σ^2 *(as given by (3.20)), it is possible to* exactly recover *supports that obey the* separation condition*. In particular, such supports can be recovered by first solving* $(P_{\text{Co-den}})$ *(with any continuous objective* $f(\mathbf{z})$*) with the parameter* ϵ *tuned to be in the range (3.23), followed by hard-thresholding with the threshold* T *chosen from the range (3.24). The size of the recoverable support is implicitly controlled by the* separation condition *and as discussed in Sec.3.4.3, the size can be as large as* $s \leq M_{ca} = \Theta(M^2)$*. Hence, for any OOSA (such as nested and coprime arrays), it is indeed possible to recover supports of size* $\Theta(M^2)$ *with*

overwhelming probability (that increases to 1 *exponentially in* $\sqrt{L}$*). Theorem 9 thereby settles an open question from [16] where such finite snapshot guarantees could only be developed for supports of size no larger than* $O(M)$*. Moreover, the guarantees in [16] were only established for a specific algorithm (Co-LASSO), whereas the analysis framework in this paper applies to a larger family of optimization problems, with convex as well as non-convex objectives. Finally, we also extend our recent non-asymptotic support recovery guarantees from [168] (that used multiple hypothesis testing) by showing that support recovery in the regime* $s > M$ *is possible with* computationally efficient *algorithms, without assuming the sources to be equi-power, or knowing the sparsity* s*.*

3.6 Modulus of Continuity: A Universal Benchmark for Evaluating Super-Resolution Algorithms

In this section, we will perform a unified analysis of positive super-resolution independent of particular algorithms. We revisit the concept of Modulus of Continuity (MC) [139, 140] which essentially provides an upper bound on the error of any algorithm, simply by leveraging the structure of signals. We consider following discrete measurement model

$$\mathbf{y} = \mathbf{Q}\mathbf{x}^\star + \mathbf{w} \qquad \mathbf{x}^\star \geq \mathbf{0} \tag{3.25}$$

where $\mathbf{x}^\star \in \mathbb{R}^N$ is sparse with positive non-zero entries, and $\mathbf{w}$ is the measurement noise. In the context of super-resolution, the measurements $\mathbf{y}$ only retain low-frequency components of the signal $\mathbf{x}^\star$. Following the notations in [61], $\mathbf{Q}$ is defined by

$$\mathbf{Q} = \mathbf{F}_N^H \mathbf{\Lambda}_n \mathbf{F}_N \tag{3.26}$$

where $\mathbf{F}_N \in \mathbb{C}^N$ is given by $[\mathbf{F}_N]_{k,l} = \frac{1}{\sqrt{N}} e^{-j2\pi kl/N}$, $-N/2+1 \leq k \leq N/2$, $0 \leq l \leq N-1$ and $\mathbf{\Lambda}_n = \text{diag}([\lambda_{-N/2+1}, \cdots, \lambda_{N/2}])$ with

$$\lambda_k = \begin{cases} 1, & k = -\frac{n-1}{2}, \cdots, \frac{n-1}{2} \\ 0, & \text{otherwise} \end{cases}$$

We assume N is even and n is odd. Intuitively, $\mathbf{Q}$ only collects n low-frequency coefficients of the DFT of $\mathbf{x}^\star$. This model popularly arises in discrete super-resolution problems with positive constraints [61, 148, 206]. Let $\mathbf{x}^\#$ be any estimate of $\mathbf{x}^\star$. Our goal in this paper is to address following question:

(Q): How to obtain a universal upper bound on the estimation error $\|\mathbf{x}^\star - \mathbf{x}^\#\|_2$ in terms of $\|\mathbf{w}\|_2$, that will be obeyed by any algorithm? Can we improve this bound by constraining $\mathbf{x}^\#$ to be non-negative?

3.6.1 Modulus of Continuity and Universal Bounds

In order to address (Q), the authors in [61, 139] have used the following notion of Modulus of Continuity (MC):

Definition 6 *Let $\mathcal{X}^\star, \mathcal{X}^\# \subset \mathbb{R}^N$ be classes of signals, $\|.\|_p$ be the $p-$norm, and $\mathbf{Q}$ be a linear operator. Then, the modulus of continuity is defined as*

$$MC(\mathbf{Q}, \mathcal{X}^\star, \mathcal{X}^\#, p) = \sup_{\substack{\mathbf{x}_1 \in \mathcal{X}^\#, \mathbf{x}_2 \in \mathcal{X}^\star \\ \mathbf{x}_1 \neq \mathbf{x}_2}} \frac{\|\mathbf{x}_1 - \mathbf{x}_2\|_p}{\|\mathbf{Q}(\mathbf{x}_1 - \mathbf{x}_2)\|_p} \tag{3.27}$$

The sets $\mathcal{X}^\star, \mathcal{X}^\#$ capture desired structures of the signal of interest, such as sparsity, positivity etc. In estimation problems, $\mathcal{X}^\star$ often represents a class to which the true signal belongs, and

$\mathcal{X}^{\#}$ represents the feasible set to which the estimator belongs. In most cases, either $\mathcal{X}^{\#} = \mathcal{X}^{\star}$ or $\mathcal{X}^{\star} \subset \mathcal{X}^{\#}$. In order to see how MC fundamentally controls the estimation error of any algorithm, we first need to define admissible estimates as follows:

Definition 7 *Consider the measurement model (3.40) with $\|\mathbf{w}\|_p \leq \epsilon$ and $\mathbf{x}^{\star} \in \mathcal{X}^{\star}$. Any estimate $\mathbf{x}^{\#}$ of $\mathbf{x}^{\star}$ is said to be admissible if*

$$\mathbf{x}^{\#} \in \mathcal{X}^{\#}, \quad \|\mathbf{y} - \mathbf{Q}\mathbf{x}^{\#}\|_p \leq \epsilon$$

The quantity $MC(\mathbf{Q}, \mathcal{X}^{\star}, \mathcal{X}^{\#}, p)$ then provides an upper bound on the error of any admissible estimate as follows:

Lemma 8 *Consider the model (3.40) with $\|\mathbf{w}\|_p \leq \epsilon$, and suppose $\mathbf{x}^{\#}$ is any admissible estimator of $\mathbf{x}^{\star}$. Then,*

$$\|\mathbf{x}^{\star} - \mathbf{x}^{\#}\|_p \leq 2\epsilon MC(\mathbf{Q}, \mathcal{X}^{\star}, \mathcal{X}^{\#}, p)$$

Remark 17. *The Modulus of Continuity therefore determines a universal upper bound on the estimation error $\|\mathbf{x}^{\star} - \mathbf{x}^{\#}\|_p$. The value of $MC(\mathbf{Q}, \mathcal{X}^{\star}, \mathcal{X}^{\#}, p)$ is algorithm-independent and only depends on the choices of $\mathcal{X}^{\star}, \mathcal{X}^{\#}, \mathbf{Q}$ and the choice of the norm. However, exact computation of $MC(\mathbf{Q}, \mathcal{X}^{\star}, \mathcal{X}^{\#}, p)$ is a challenging task, which was first studied in the pioneering work by [139] in the context of super-resolution reconstruction of spike signals from low-frequency measurements and further developed in recent work on discrete positive super-resolution [61]. We will review this result by introducing the following class of signals that obey a separation condition [61].*

Definition 8 *(Set of Signals Obeying Separation Condition) Given N and n, the set Δ_{sep} is given*

by

$$\Delta_{sep} \triangleq \{\mathbf{x} \in \mathbb{C}^N \mid \rho(\frac{k}{N}, \frac{l}{N}) \geq \frac{4}{n-1} \quad \forall k \neq l \in supp(\mathbf{x})\}$$

where $\rho(\cdot,\cdot)$ is a wrap-around distance function [142] such that for $\forall \mu_1, \mu_2 \in [0,1]$

$$\rho(\mu_1, \mu_2) \triangleq \min(|\mu_1 - \mu_2|, |\mu_1 + 1 - \mu_2|, |\mu_2 + 1 - \mu_1|)$$

Additionally, the set Δ_{sep}^{+} is given by

$$\Delta_{sep}^{+} \triangleq \{\mathbf{x} \in \Delta_{sep}, \mathbf{x} \geq \mathbf{0}\}$$

If we assume $\mathcal{X}^\star = \mathcal{X}^\# = \Delta_{\text{sep}}$, then the following result provides an explicit upper bound on $MC(\mathbf{Q}, \mathcal{X}^\star, \mathcal{X}^\#, 2)$ in terms of n and N:

Lemma 9 *[61, 139] Let $\mathcal{X}^\star = \mathcal{X}^\# = \Delta_{sep}$, and let $\mathbf{Q}$ be given by (3.26). Then,*

$$MC(\mathbf{Q}, \Delta_{sep}, \Delta_{sep}, 2) \leq C(n)N^3 \tag{3.28}$$

where $C(n)$ is a function of only n (independent of N) implicitly defined in [139].

Given the measurement model (3.40), the goal of super-resolution is to reconstruct the N DFT coefficiens of sparse $\mathbf{x}^\star$ (or equivalently, the signal $\mathbf{x}^\star$) from observations that only preserve the lowest $n < N$ frequency components. If we assume that both the true signal and its estimate $\mathbf{x}^\#$ belong to Δ_{sep} (i.e., they satisfy the separation condition), then Lemma 9 and Lemma 8 show that given n, the estimation error grows as $O(N^3)$.

However, in practice, it is difficult to develop algorithms that can actually constrain $\mathbf{x}^\#$ to belong to Δ_{sep}. [10] In recent work [61], the authors developed an l_1 minimization framework

[10] Partly because Δ_{sep} is a non-convex set

for super resolution reconstruction, where they only constrained the estimate $\mathbf{x}^{\#}$ to be positive and developed algorithm-specific error bound (with respect to l_1 norm of the error). Inspired by this work, we will develop a new bound for $MC(\mathbf{Q}, \mathcal{X}^{\star}, \mathcal{X}^{\#}, 2)$ where $\mathcal{X}^{\star}$ imposes minimum separation as well as positivity on the true signal, whereas $\mathcal{X}^{\#}$ only imposes a positive constraint on $\mathbf{x}^{\#}$. Our analysis will show that this bound grows as $O(N^{2.5})$ and is therefore tighter than (3.28).

3.6.2 New Bound on Modulus of Continuity for Positive Super-Resolution and Applications

Our new upper bound for the modulus of continuity is based on a recent result from [4] for continuous Direction-of-Arrival estimation. For any vector $\mathbf{x}$, let $\text{Toep}(\mathbf{x})$ denote the Hermitian Toeplitz matrix with $\mathbf{x}$ as the first column. Consider $\mathbf{r}^{\star} \in \mathbb{C}^K$ such that

$$\text{Toep}(\mathbf{r}^{\star}) = \sum_{i=1}^{D} \mathbf{a}(\theta_i)\mathbf{a}^H(\theta_i)d_i \tag{3.29}$$

where $D < K$, $d_i > 0$, $\theta_i \in [0, 1]$ and

$$\mathbf{a}(\theta_i) = [1, e^{-j2\pi\theta_i}, \cdots, e^{-j2\pi(K-1)\theta_i}]^T$$

It can be easily seen that $\text{Toep}(\mathbf{r}^{\star}) \succeq \mathbf{0}$. We invoke the following result from [4]:

Theorem 10 *[4] Let $\mathbf{r}^{\star}$ be given by (3.29), and $\mathbf{r}^{\#} \in \mathbb{C}^K$ be any vector such that Toep$(\mathbf{r}^{\#}) \succeq \mathbf{0}$. If the frequencies $\{\theta_i\}_{i=1}^{D}$ satisfy the separation condition*

$$\min_{l \neq m} \rho(\theta_l, \theta_m) > \frac{2}{h}$$

and $h > 128$*, then there exist positive constants* $\bar{c}_1, \bar{c}_2, \bar{c}_3, \bar{c}_4$ *such that for* $h \leq k < K$

$$\begin{aligned} &|r_k^\star - r_k^\#| \\ &\leq \left(\bar{c}_1 + \frac{\bar{c}_2 \pi k}{h} + \frac{\bar{c}_3 \pi^2 k^2}{h^2}\right)\left(\frac{\bar{c}_4 D}{\sqrt{h}} + 1\right) \|\mathbf{r}_h^\star - \mathbf{r}_h^\#\|_2 \end{aligned} \tag{3.30}$$

where $\mathbf{r}_h^\star = [r_0^\star, \cdots, r_{h-1}^\star]^T, \mathbf{r}_h^\# = [r_0^\#, \cdots, r_{h-1}^\#]^T$.

Equipped with Theorem 10, the main result of this paper is given by

Theorem 11 *Let* $\mathcal{X}^\star, \mathcal{X}^\#$ *be chosen as*

$$\mathcal{X}^\star = \Delta_{sep}^+, \quad \mathcal{X}^\# = \mathbb{R}_+^N \triangleq \left\{\mathbf{x} \in \mathbb{R}^N : \mathbf{x} \geq \mathbf{0}\right\}$$

Furthermore, let the matrix $\mathbf{Q}$ *be given by (3.26). If* $n > 256$*, the Modulus of Continuity is upper bounded as*

$$MC(\mathbf{Q}, \mathcal{X}^\star, \mathcal{X}^\#, 2) \leq \sqrt{2 + (N - n + 1)\beta(n, N)} \tag{3.31}$$

where

$$\beta(n, N) \triangleq \left(\bar{c}_1 + \frac{\bar{c}_2 \pi N}{n+1} + \frac{\bar{c}_3 \pi^2 N^2}{(n+1)^2}\right)^2 \left(\bar{c}_4 \sqrt{\frac{n+1}{8}} + 1\right)^2$$

Here $\bar{c}_1, \bar{c}_2, \bar{c}_3, \bar{c}_4$ *are the same constants as in Theorem 10.*

Proof. For $\forall \mathbf{x}^\star \in \mathcal{X}^\star, \forall \mathbf{x}^\# \in \mathcal{X}^\#$, $\mathbf{x}^\star \neq \mathbf{x}^\#$, $\mathbf{F}_N \mathbf{x}^\star$ and $\mathbf{F}_N \mathbf{x}^\#$ are symmetric and we can define $\mathbf{r}^\star, \mathbf{r}^\# \in \mathbb{C}^{\frac{N}{2}+1}$ as

$$r_i^\star = [\mathbf{F}_N \mathbf{x}^\star]_{i+\frac{N}{2}-1}, \quad r_i^\# = [\mathbf{F}_N \mathbf{x}^\#]_{i+\frac{N}{2}-1}, \quad 0 \leq i \leq \frac{N}{2}$$

Since $\mathbf{x}^\star, \mathbf{x}^\# \geq \mathbf{0}$, it follows that [4]

$$\text{Toep}(\mathbf{r}^\star) \succeq \mathbf{0} \qquad \text{Toep}(\mathbf{r}^\#) \succeq \mathbf{0}$$

Moreover $\mathbf{r}^\star$ has the form $\mathbf{r}^\star = \sum_{k=1}^{\|\mathbf{x}^\star\|_0} \mathbf{a}(\theta_k)\mathbf{a}^H(\theta_k)x_k^\star$ where $\theta_k = k/N, k \in \text{supp}(\mathbf{x}^\star)$. Since $\mathbf{x}^\star \in \Delta_{\text{sep}}$, this implies $\rho(\theta_k, \theta_l) \geq \frac{4}{n-1}$. Hence, Theorem 10 applies (by replacing K, h, and D with $\frac{N}{2}+1$, $\frac{n+1}{2}$, and $\|\mathbf{x}^\star\|_0$ respectively)

$$\sum_{i=\frac{n+1}{2}}^{\frac{N}{2}} (r_i^\star - r_i^\#)^2 \leq \frac{N-n+1}{2}\|\mathbf{r}^\star_{\frac{n+1}{2}} - \mathbf{r}^\#_{\frac{n+1}{2}}\|_2^2 \cdot$$
$$\left(\bar{c}_1 + \frac{\bar{c}_2\pi N}{n+1} + \frac{\bar{c}_3\pi^2N^2}{(n+1)^2}\right)^2 \left(\bar{c}_4\sqrt{\frac{n+1}{8}} + 1\right)^2$$

where we use the fact that $\|\mathbf{x}^\star\|_0 \leq \frac{n+1}{4}$ owing to the separation condition. Also note that

$$\|\mathbf{F}_N(\mathbf{x}^\star - \mathbf{x}^\#)\|_2^2 \leq 2\|\mathbf{r}^\star - \mathbf{r}^\#\|_2^2$$
$$\leq 2\|\mathbf{r}^\star_{\frac{n+1}{2}} - \mathbf{r}^\#_{\frac{n+1}{2}}\|_2^2 \left(1 + \frac{N-n+1}{2}\beta(n,N)\right)$$

Moreover,

$$\|\mathbf{Q}(\mathbf{x}^\star - \mathbf{x}^\#)\|_2^2 = \|\mathbf{\Lambda}_n\mathbf{F}_N(\mathbf{x}^\star - \mathbf{x}^\#)\|_2^2$$
$$= 2\|\mathbf{r}^\star_{\frac{n+1}{2}} - \mathbf{r}^\#_{\frac{n+1}{2}}\|_2^2 - (r_0^\star - r_0^\#)^2 \geq \|\mathbf{r}^\star_{\frac{n+1}{2}} - \mathbf{r}^\#_{\frac{n+1}{2}}\|_2^2$$

This implies

$$\frac{\|\mathbf{x}^\star - \mathbf{x}^\#\|_2^2}{\|\mathbf{Q}(\mathbf{x}^\star - \mathbf{x}^\#)\|_2^2} = \frac{\|\mathbf{F}_N(\mathbf{x}^\star - \mathbf{x}^\#)\|_2^2}{\|\mathbf{Q}(\mathbf{x}^\star - \mathbf{x}^\#)\|_2^2} \leq \frac{\|\mathbf{F}_N(\mathbf{x}^\star - \mathbf{x}^\#)\|_2^2}{\|\mathbf{r}^\star_{\frac{n+1}{2}} - \mathbf{r}^\#_{\frac{n+1}{2}}\|_2^2}$$
$$\leq 2\left(1 + \frac{N-n+1}{2}\beta(n,N)\right)$$

thereby proving the theorem.

3.6.3 Comparison with Lemma 9

Our result in Theorem 25 significantly differs from the result in Lemma 9 in the following ways:

- In Lemma 9, the signal classes $\mathcal{X}^\star, \mathcal{X}^\#$ are identical, while in Theorem 25, they are different. In particular, $\mathcal{X}^\star$ (the set to which the true signal belongs) contains all non-negative vectors that satisfy separation condition, whereas $\mathcal{X}^\#$ (the set to which the estimate belongs) simply contains all non-negative vectors. Such distinction of $\mathcal{X}^\star, \mathcal{X}^\#$ enables better analysis of practical estimation algorithms since it is difficult for an algorithm to actually impose the constraint $\mathbf{x}^\# \in \Delta_{\text{sep}}$.

- The upper bound on MC in Lemma 9 is given by

$$MC(\mathbf{Q}, \Delta_{\text{sep}}, \Delta_{\text{sep}}, 2) \leq C(n)N^3 \tag{3.32}$$

  On the other hand, Theorem 25 shows that

$$MC(\mathbf{Q}, \Delta^+_{\text{sep}}, \mathbb{R}^N_+, 2) \lesssim O(\frac{\sqrt{N-n}N^2}{n^{1.5}}) \tag{3.33}$$

  When N is large (and n is fixed), Lemma 9 shows that the upper bounded is $O(N^3)$ while Theorem 25 suggests that this can be tightened to $O(N^{2.5})$. This 0.5 improvement in the exponent with respect to N is mainly due to the introduction of positive constraints. To the best of our knowledge, this improvement is the first result of its kind.

3.6.4 Unified Analysis of Specific Algorithms

In Lemma 8, we have shown that $MC(\mathbf{Q}, \mathcal{X}^\star, \mathcal{X}^\#, 2)$ provides an upper bound on the estimation error of any algorithm that produces an admissible estimate. To illustrate this further, we study the following three convex problems to estimate $\mathbf{x}^\star$ from the measurement model (3.40).

$$\begin{aligned}
&\text{find} \quad \mathbf{z} \geq \mathbf{0} \quad s.t \quad \|\mathbf{y} - \mathbf{Q}\mathbf{z}\|_2 \leq \varepsilon \qquad (\textbf{Algo-F}) \\
&\min \|\mathbf{z}\|_1 \quad s.t. \quad \|\mathbf{y} - \mathbf{Q}\mathbf{z}\|_2 \leq \varepsilon, \mathbf{z} \geq \mathbf{0} \quad (\textbf{Algo-}l_1) \\
&\min \|\mathbf{y} - \mathbf{Q}\mathbf{z}\|_2 \quad s.t \quad \mathbf{z} \geq \mathbf{0} \qquad (\textbf{Algo-}l_2)
\end{aligned}$$

Applying Theorem 25 and Lemma 8, we have the following unified analysis of the preceding algorithms

Corollary 5 *Consider the noisy measurement model (3.40) with $\|\mathbf{w}\|_2 \leq \epsilon$. Suppose the true signal satisfies $\mathbf{x}^\star \in \Delta^+_{sep}$. Let $\mathbf{x}^\#_F, \mathbf{x}^\#_1, \mathbf{x}^\#_2$ be the optimal solutions of $(\textbf{Algo-F}), (\textbf{Algo-}l_1)$, and $(\textbf{Algo-}l_2)$ respectively. If $n > 256$, we have*

$$\begin{aligned}
&\max\{\|\mathbf{x}^\star - \mathbf{x}^\#_F\|_2, \|\mathbf{x}^\star - \mathbf{x}^\#_1\|_2, \|\mathbf{x}^\star - \mathbf{x}^\#_2\|_2\} \\
&\leq 2\epsilon\sqrt{2 + (N - n + 1)\beta(n, N)}. \qquad (3.34)
\end{aligned}$$

3.7 Achievability of the Universal Upper Bound

Our guarantees on support recovery builds on the universal upper bound (3.17) on the distance between $\mathbf{p}^\star \in \mathcal{P}^+_{sep}$ (obeying the separation condition) and any *non-negative* vector $\mathbf{p}^\#$ that serves as an estimate of $\mathbf{p}^\star$. As discussed in Remark 15 and Theorem 25, for large grid sizes,

(in particular, when $N/(2M_{ca}+1) \geq 3.03$) the universal upper bound essentially scales as

$$\|\mathbf{p}^\star - \mathbf{p}^\#\|_1 \leq A(N,M)\|\mathbf{A}_{\text{KR}}^{\mathbb{U}}(\mathbf{p}^\star - \mathbf{p}^\#)\|_2 \tag{3.35}$$

where

$$A(N,M) = 34\frac{N^2}{M_{ca}^2}$$

represents the amplification factor, i.e., the factor by which $\|\mathbf{A}_{\text{KR}}^{\mathbb{U}}(\mathbf{p}^\star - \mathbf{p}^\#)\|_2$ gets amplified in the reconstruction/estimation error $\|\mathbf{p}^\star - \mathbf{p}^\#\|_1$. In this section, we investigate if this exponent (i.e. 2) of N in $A(N,M)$ is tight or achievable *assuming M is constant*. Notice that (3.17) is valid for any non-negative pair of signals $(\mathbf{p}^\star, \mathbf{p}^\#)$ as long as $\mathbf{p}^\star$ also obeys the separation condition. Hence, in order to show tightness, we need to construct a specific pair of vectors $(\mathbf{p}_1^\star, \mathbf{p}_1^\#)$ with $\mathbf{p}_1^\star, \mathbf{p}_1^\# \succeq \mathbf{0}$ and $\mathbf{p}_1^\star \in \mathcal{P}_{sep}^+$, such that

$$\|\mathbf{p}_1^\star - \mathbf{p}_1^\#\|_2 \geq CN^2\|\mathbf{A}_{\text{KR}}^{\mathbb{U}}(\mathbf{p}_1^\star - \mathbf{p}_1^\#)\|_2 \tag{3.36}$$

where C is a constant (possibly depending only on M which is assumed to be constant). Establishing such a *lower bound* that also scales quadratically in N, will establish that the exponent of N in the upper bound (3.35) is indeed tight. In the following theorem, we show that such $\mathbf{p}_1^\star, \mathbf{p}_1^\#$ satisfying (3.36) indeed exist.

Theorem 12 *There exist $\mathbf{p}_1^\star \in \mathcal{P}_{sep}^+$ and $\mathbf{p}_1^\# \succeq \mathbf{0}$ such that whenever $M_{ca} \geq 128$ and $N \geq 3.03(2M_{ca}+1)$, we have*

$$\|\mathbf{p}_1^\star - \mathbf{p}_1^\#\|_1 \leq C_1(M)N^2\|\mathbf{A}_{KR}^{\mathbb{U}}(\mathbf{p}_1^\star - \mathbf{p}_1^\#)\|_2$$

and

$$\|\mathbf{p}_1^\star - \mathbf{p}_1^\#\|_1 \geq C_2(M)N^2\|\mathbf{A}_{KR}^{\mathbb{U}}(\mathbf{p}_1^\star - \mathbf{p}_1^\#)\|_2 \tag{3.37}$$

where $C_1(M)$ and $C_2(M)$ are constants (which are functions of M) given by

$$C_1(M) = \frac{34}{M_{ca}^2}, \quad C_2(M) = \frac{1}{KM_{ca}^{2.5}}$$

with $K = \pi^2\sqrt{\frac{7}{15}} + 0.1973\pi^3$.

Proof. The upper bound is due to Theorem 25. For the lower bound, please refer to Appendix 3.11.4. The explicit construction of $\mathbf{p}_1^\star$ and $\mathbf{p}_1^\#$ shows that the exponent of N in $A(N, M)$ is indeed tight and for fixed M

$$\frac{\|\mathbf{p}_1^\star - \mathbf{p}_1^\#\|_1}{\|\mathbf{A}_{\text{KR}}^{\mathbb{U}}(\mathbf{p}_1^\star - \mathbf{p}_1^\#)\|_2} = \Theta(N^2) \tag{3.38}$$

The construction of $\mathbf{p}_1^\star$ and $\mathbf{p}_1^\#$ also sheds further light into what kind source power and its estimate may result in the worst-case estimation error. In the proof of Theorem 12, we construct $\mathbf{p}_1^\star$ as a single non-negative spike at index l and hence it automatically satisfies the separation condition. The vector $\mathbf{p}_1^\#$, on the other hand, consists of two non-negative spikes symmetrically placed on either side of the true spike at indices $l-1$ and $l+1$. This represents a classical scenario where two very closely spaced spikes can be misdetected as a single spike, or when a single spike splits into two closely spaced neighboring spikes [15]. Under these scenarios, Theorem 12 shows that the estimation error indeed grows as $\Theta(N^2)$.

3.8 Covariance-Driven Super Resolution Imaging: Application of Coarray in Fluorescence Microscopy

Optical super-resolution microscopy overcomes Abbe's diffraction limit in diffractive optical imaging [216,217]. Two popular microscopic methods are PALM [218] and STORM [219] where the key idea is to image only a fraction of fluorophores once a time. The point sources can be switched between a fluorescent and a dark state by using different laser pulses. A drawback of PALM and STORM is that tens of thousands of diffraction limited images are required and the corresponding total exposure time can be as long as several minutes [219, 220]. To reduce the total acquisition time, another super-resolution microscopy algorithm named SOFI was proposed in [221, 222] where the basic idea is to utilize the temporal correlation structure of photons from different emitters. For Gaussian point-spread-functions (PSF), the width of the equivalent PSF in correlation domain is decreased by a factor of $\sqrt{2}$, provided the fluctuations of point sources are uncorrelated. The resolution can be further improved if higher order statistics (HOS) of the measurements is exploited. However, more frames are needed to estimate the HOS and weak sources can be severely masked by strong ones [220, 222]. Consequently, SOFI cannot achieve the same resolution levels as PALM or STORM.

In a very recent work [220], the authors proposed a new algorithm called SPARCOM to overocme the drawbacks of SOFI. To exploit the sparsity structure in the images, the authors assume the point sources to lie on a high-resolution grid. To enforice sparsity, a LASSO problem was formulated in the correlation domain. This method is motivated by earlier work on correlation aware LASSO [16] and the main advantage over aforementioned methods is that it can achieve high resolution as STORM and PALM while the data acquisition time is significantly faster. For large field of view and high target resolution, the computational complexity of SPARCOM can be prohibitive owing to large grid size.

In this paper, we propose a new algorithm called spatially smoothed SOFI-MUSIC (SS-

SOFI) by exploiting the correlations among point sources in a novel manner. In contrast to SPARCOM, the sources are not required to lie on known grid and SOFI-MUSIC is computationally more efficient. For Gaussian PSF, we show that a novel sum co-array structure emerges owing to the correlation structure of the measurements. Motivated by our earlier works on sparse arrays [124, 125], we show that it is possible to resolve more sources than sensors by carefully placing the physical sensors and building a suitable large covariance matrix.

3.8.1 Problem Formulation

Consider L fluctuating point sources with fluorescence source distribution at time t given by [220, 221]:

$$J(\mathbf{r}, t) = \sum_{k=0}^{L-1} \delta(\mathbf{r} - \mathbf{r}_k^\star) s_k(t) \tag{3.39}$$

where $\delta(\cdot)$ is the Dirac Delta function. Here, $\mathbf{r}_k^\star = [x_k^\star, y_k^\star]^T \in \mathbb{R}^2$ denotes the location of kth source and $s_k(t)$ represents its time-dependent brightness. The measured intensity at any position $\mathbf{r}$ is the convolution of $J(\mathbf{r}, t)$ and the point spread function $u(\mathbf{r})$ of the microscope, which is typically assumed to be known [220, 221]. The measurement at the mth sensor (located at $\tilde{\mathbf{r}}_m$), is given by

$$f(\tilde{\mathbf{r}}_m, t) = \sum_{k=0}^{L-1} u(\tilde{\mathbf{r}}_m - \mathbf{r}_k^\star) s_k(t) \tag{3.40}$$

Our measurement model is similar to [220–222], and we also make the following statistical assumptions as in [220]:

A1 The locations $\{\mathbf{r}_k^\star\}_{k=0}^{L-1}$ are fixed over the acquisition time.

A2 The brightness functions $\{s_k(t)\}_{k=0}^{L-1}$ are wide sense stationary, with $\mathbb{E}\{s_k(t)\} = s_k^\star$ and $\mathrm{Cov}(\mathbf{s}(t), \mathbf{s}(t)) = \mathbf{P}$ with $\mathbf{s}(t) = [s_0(t), \cdots, s_{L-1}(t)]^T$.

3.8.2 Review of SOFI

In [221, 222], the authors assume that the point sources are statistically uncorrelated, implying that $\mathbf{P}$ is diagonal with $\mathbf{P} = \text{diag}(p_0, \cdots, p_{L-1})$. The PSF $u(\mathbf{r})$ is assumed to be a Gaussian kernel, given by

$$u(\mathbf{r}) = e^{-\alpha\|\mathbf{r}\|_2^2} \tag{3.41}$$

Consequently, the correlation of the measurement at the mth sensor (with location $\tilde{\mathbf{r}}_m$) is given by

$$\rho_m = \text{Cov}(f(\tilde{\mathbf{r}}_m), f(\tilde{\mathbf{r}}_m)) = \sum_{k=0}^{L-1} u^2(\tilde{\mathbf{r}}_m - \mathbf{r}_k^\star)p_k \tag{3.42}$$

Comparing the correlation representation (3.42) and the physical measurements (3.40), we can say that $\tilde{u}(\mathbf{r})$ is the equivalent point spread function for second-order statistics is

$$\tilde{u}(\mathbf{r}) = u^2(\mathbf{r}) = e^{-2\alpha\|\mathbf{r}\|_2^2}$$

Notice that the width of the "new" PSF is reduced by a factor of $\sqrt{2}$ and the resolution is accordingly increased. In general, given the nth order statistics of the physical measurements, the resolution can be improved by a factor of $\sqrt{n}$. However, this straightforward method has following drawbacks in practice [220]

- Computing higher order statistics of measurements requires a larger number of frames and decreases its temporal resolution.
- As indicated by (3.42), for the new PSF, the weak emitters are masked even more by the stronger ones. This is known as dynamic range expansion and it worsens for higher order statistics.

3.8.3 Review of SPARCOM

In general, SOFI cannot achieve the same resolution as STORM and PALM, since a large number of temporal snapshots is needed to realize the effect of the equivalent PSF $u^2(\mathbf{r})$. To remedy this drawback, the authors of [220] proposed the idea of correlation-aware LASSO [16] to recover the source locations. Suppose M^2 sensors are placed on a $M \times M$ *low-resolution grid* Ω_l with spacing Δ_l and the point sources are assumed to lie on a $N \times N$ *high-resolution grid* Ω_h with separation $\Delta_h \ll \Delta_l$. We further assume $\Omega_l \subset \Omega_h$. Under this setting, the temporal measurements (3.40) can be rewritten as

$$f(m\Delta_l, n\Delta_l, t) = \sum_{i,l=0}^{N-1} u(m\Delta_l - i\Delta_h, n\Delta_l - l\Delta_h) s_{il}(t) \tag{3.43}$$

where L out of $\{s_{il}(t)\}_{i,l=0}^{N-1}$ are non-zero. For efficient implementation, the authors of [220] computed the Fourier transform of the physical measurements:

$$\begin{aligned} &Y(k_m, k_n, t) = \\ &\sum_{[i,l]\in\Omega_h} s_{il}(t) \sum_{[\hat{m},\hat{n}]\in\Omega_l} u(\hat{m}-i, \hat{n}-l) e^{-j2\pi\frac{k_m\hat{m}}{N}} e^{-j2\pi\frac{k_n\hat{n}}{N}} \end{aligned} \tag{3.44}$$

where $0 \le k_m, k_n \le M-1$. Let $\mathbf{y}(t)$, $\mathbf{s}(t)$ be the column-wise vectorized forms of $\mathbf{Y}(t) = [y(k_m, k_n, t)] \in \mathbb{C}^{M\times M}$, and $\mathbf{S}(t) = [s_{il}(t)] \in \mathbb{R}^{N\times N}$ respectively. It is shown in [220] that

$$\mathbf{y}(t) = \mathbf{A}\mathbf{s}(t) \quad \mathbf{R}_y = \text{Cov}(\mathbf{y}(t), \mathbf{y}(t)) = \mathbf{A}\tilde{\mathbf{P}}\mathbf{A}^H$$

where $\mathbf{A} \in \mathbb{C}^{M^2\times N^2}$ is a suitably defined matrix that can be derived from (3.44). When the sources are uncorrelated, $\tilde{\mathbf{P}}$ is diagonal and the source locations are estimated by solving following

Correlation-aware LASSO [16]

$$\min_{\mathbf{x}\geq\mathbf{0}} \lambda\|\mathbf{x}\|_1 + \frac{1}{2}\|\hat{\mathbf{R}}_y - \sum_{l=1}^{N^2} \mathbf{a}_l\mathbf{a}_l^H x_l\|_F^2 \quad \text{(SPARCOM)}$$

Here $\mathbf{a}_l$ is the l-th column of $\mathbf{A}$ and $\hat{\mathbf{R}}_y$ is the empirical covariance of $\mathbf{y}(t)$ given by

$$\hat{\mathbf{R}}_y = \frac{1}{T}\sum_{t=0}^{T-1}(\mathbf{y}(t) - \bar{\mathbf{y}})(\mathbf{y}(t) - \bar{\mathbf{y}})^H$$

where $\bar{\mathbf{y}}$ is the empirical mean. The key idea in [220] is based on the correlation-aware sparse recovery framework proposed in [16]. However, SPARCOM has following disadvantages in practice:

- In [218, 219], the field of view is a few microns while the target resolution level is tens of nanometres. Thus, to obtain comparable performance, the dimension N should be $\sim 10^2$ or 10^3. Thus, $\mathbf{A}$ will be a huge complex-valued matrix and solving (SPARCOM) may be both memory and computationally inefficient.

- To reveal the sparsity, the sources are assumed to be on a prescribed $N \times N$ dense grid Ω_h while in practical imaging [218, 219], the point sources are often clustered, and not necessarily located on a uniform grid. Increasing N will not eliminate grid mismatch, and will worsen the computational complexity.

- In compressed sensing theory [96], the regularization parameter λ is dependent on the noise or approximation error. In practice, it is not easy to find the optimal λ without partial knowledge of ground truth and/or noise level.

3.8.4 Gridless Covariance Driven Super-resolution Imaging

In this paper, the locations of L sources and sensors are not restricted to lie on a known grid. Suppose we have M sensors located at $\{\tilde{\mathbf{r}}_m\}_{m=0}^{M-1}$. From (3.40), the vector of measurements is given by

$$\mathbf{f}(t) = \mathbf{U}\mathbf{s}(t) \qquad [\mathbf{U}]_{m,k} = u(\tilde{\mathbf{r}}_m - \mathbf{r}_k^\star) \tag{3.45}$$

The covariance of $\mathbf{f}(t)$ can be expressed as

$$\mathbf{R}_f \triangleq \mathbb{E}[\mathbf{f}(t) - \mathbb{E}(\mathbf{f}(t))][\mathbf{f}(t) - \mathbb{E}(\mathbf{f}(t))]^T = \mathbf{U}\mathbf{P}\mathbf{U}^T$$

In this case, the empirical covariance matrix using T samples is given by

$$\hat{\mathbf{R}}_f = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{f}(t) - \bar{\mathbf{f}})(\mathbf{f}(t) - \bar{\mathbf{f}})^T = \mathbf{U}\hat{\mathbf{P}}\mathbf{U}^T$$

where $\bar{\mathbf{f}} = 1/T\sum_{t=1}^{T}\mathbf{f}(t)$. Inspired by Direction-of-Arrival (DOA) estimation algorithms [47], we propose to use the MUSIC [149] algorithm to identify the source locations $\{\mathbf{r}_k^\star\}_{k=0}^{L-1}$. We name this variant of MUSIC for super-resolution imaging as SOFI-MUSIC since it is based on the statistical assumptions in [221, 222].

MUSIC is a gridless algorithm and its success depends on the algebraic structure of $\mathbf{U}$. In classical 1D narrow-band DOA estimation, $\mathbf{U}$ is typically Vandermonde structured. However, this is not true for 2D non-uniform sensor arrays used in SOFI-MUSIC [47]. To guarantee the uniqueness of recovery, we first need to define the following notion of identifiability.

Definition 9 *(Identifiability) Consider the measurement model (3.40) acquired at an array of sensors located at $\{\tilde{\mathbf{r}}_m\}_{m=0}^{M-1}$. A set of L sources with locations $\{\mathbf{r}_k^\star\}_{k=0}^{L-1}$ are said to be identifiable with this sensor array if $\mathbf{U}$ satisfies the following conditions:*

- $\mathbf{U}$ *is full column rank*
- *For any* $\hat{\mathbf{r}} \notin \{\mathbf{r}_k^\star\}_{k=0}^{L-1}$, *define* $\hat{\mathbf{u}} = [u(\tilde{\mathbf{r}}_0 - \hat{\mathbf{r}}), \cdots, u(\tilde{\mathbf{r}}_{M-1} - \hat{\mathbf{r}})]^T$. *Then* $[\mathbf{U}, \hat{\mathbf{u}}]$ *is full column rank.*

Remark 18. *In above definition, the locations of sensors and sources are fixed. In ID however, if the sensors are on a uniform linear array (ULA), then* $\mathbf{U}$ *is Vandermonde structured regardless of the source locations. Thus, identifiability is ensured uniformly for any collection of* L *sources, as long as* $M \geq L + 1$. *As discussed in [125], the identifiability of 2D DOA estimation depends on both the sensor and source locations which is fundamentally different from 1D case with Vandermonde structure.*

We can now state a general theorem about recovery guarantees of SOFI-MUSIC algorithm:

Theorem 13 *If* $\mathbf{U}$ *satisfies the identifiability condition from Def. 9 and* $\mathbf{P}$ *is full rank, the source locations* $\{\mathbf{r}_k^\star\}_{k=0}^{L-1}$ *can be exactly recovered via SOFI-MUSIC.*

Proof. Suppose $\mathbf{U}$ satisfies the identifiability property. Then we have, $L \leq M - 1$ and the range spaces $\mathcal{R}(\mathbf{R}_f) = \mathcal{R}(\mathbf{U}) = \mathcal{N}^\perp(\mathbf{R}_f)$ given $\mathbf{P}$ is full rank. For any $\hat{\mathbf{r}}$, define $\hat{\mathbf{u}} = [u(\tilde{\mathbf{r}}_0 - \hat{\mathbf{r}}), \cdots, u(\tilde{\mathbf{r}}_{M-1} - \hat{\mathbf{r}})]^T$. Then $\hat{\mathbf{u}} \in \mathcal{R}(\mathbf{U})$ if and only if $\hat{\mathbf{r}} \in \{\mathbf{r}_k^\star\}_{k=0}^{L-1}$. Thus the source locations can be exactly recovered as the nulls in the spectrum of the MUSIC algorithm [47].

3.8.5 Novel Sum Co-array Structure with Gaussian PSF

The identifiability condition in Def. 9 implies that SOFI-MUSIC can identify no more than $M - 1$ sources with M sensors. This is the source localization limit in classical DOA estimation [47] which suggests that the number of sources should be less than the number of sensors. However, we now show that this limitation can be overcome if the sources are uncorrelated and the PSF is a 2D Gaussian. The idea is to exploit the structure of a certain sum co-array. This is by our earlier line of work [4, 16, 45, 124, 125], which suggest that it is

possible to localize more sources than sensors using non-uniform arrays with extended difference co-arrays. However, these results cannot be directly used here due to the presence of a Gaussian PSF. Interestingly, we will show that the covariance matrix of the measurements in this case permits a novel decomposition of the Gaussian PSF, revealing a sum co-array, and allowing localization of more sources.

Note that since $\mathbf{P}$ is diagonal, we can write

$$\mathbf{v}_f \triangleq \text{vec}(\mathbf{R}_f) = \underbrace{(\mathbf{U} \odot \mathbf{U})}_{\mathbf{U}_{\text{KR}}} \mathbf{p} \tag{3.46}$$

where $\mathbf{p} = [p_0, \cdots, p_{L-1}]^T$ and $\odot$ denotes the Khatri-Rao product (column-wise Kronecker product of $\mathbf{U}$). Specifically, we have

$$\begin{aligned} [\mathbf{U}_{\text{KR}}]_{m_1 M + m_2, k} &= e^{-\alpha(\|\tilde{\mathbf{r}}_{m_1} - \mathbf{r}_k^\star\|_2^2 + \|\tilde{\mathbf{r}}_{m_2} - \mathbf{r}_k^\star\|_2^2)} \\ &= e^{-\alpha(\|\tilde{\mathbf{r}}_{m_1}\|_2^2 + \|\tilde{\mathbf{r}}_{m_2}\|_2^2)} e^{2\alpha\langle \tilde{\mathbf{r}}_{m_1} + \tilde{\mathbf{r}}_{m_2}, \mathbf{r}_k^\star \rangle} e^{-2\alpha\|\mathbf{r}_k^\star\|_2^2} \end{aligned} \tag{3.47}$$

Notice that the first term in the exponent is $-\alpha(\|\tilde{\mathbf{r}}_{m_1}\|_2^2 + \|\tilde{\mathbf{r}}_{m_2}\|_2^2)$ which is only a function of known sensor locations. Similarly, the third term $-2\alpha\|\mathbf{r}_k^\star\|_2^2$ is only a function of the unknown source locations $\mathbf{r}_k^\star$. The second term $2\alpha\langle \tilde{\mathbf{r}}_{m_1} + \tilde{\mathbf{r}}_{m_2}, \mathbf{r}_k^\star \rangle$ is an inner-product between the source location $\mathbf{r}_k^\star$ and the elements of the following set

$$\mathbb{S}_{sum} \triangleq \{\tilde{\mathbf{r}}_{m_1} + \tilde{\mathbf{r}}_{m_2}, 0 \leq m_1, m_2 \leq L-1\} \tag{3.48}$$

The set $\mathbb{S}_{sum}$ contains pair-wise sum of sensor locations and it represents the so-called sum co-array. We will now show how we can perform source localization by leveraging the structure of the sum co-array.

First, define diagonal matrices $\mathbf{D}_s \in \mathbb{R}^{M^2 \times M^2}, \mathbf{D}_p \in \mathbb{R}^{L \times L}$ as

$$\mathbf{D}_s \triangleq \begin{bmatrix} \ddots & & \\ & e^{\alpha(\|\tilde{\mathbf{r}}_{m_1}\|_2^2 + \|\tilde{\mathbf{r}}_{m_2}\|_2^2)} & \\ & & \ddots \end{bmatrix}$$

$$\mathbf{D}_p \triangleq \begin{bmatrix} \ddots & & \\ & e^{-2\alpha\|\mathbf{r}_k^\star\|_2^2} & \\ & & \ddots \end{bmatrix}$$

Since $\mathbf{D}_s$ is known, we can construct $\tilde{\mathbf{v}}_f$ from $\mathbf{v}_f$ as

$$\tilde{\mathbf{v}}_f \triangleq \mathbf{D}_s \mathbf{v}_f$$

From (3.46), it can be easily seen that

$$\tilde{\mathbf{v}}_f = \mathbf{U}_{sum} \mathbf{D}_p \mathbf{p} \tag{3.49}$$

The elements of $\mathbf{U}_{sum}$ are characterized by the source locations and the sum co-array $\mathbb{S}_{sum}$ as

$$[\mathbf{U}_{sum}]_{m_1 M + m_2, k} = e^{2\alpha \langle \tilde{\mathbf{r}}_{m_1} + \tilde{\mathbf{r}}_{m_2}, \mathbf{r}_k^\star \rangle} \tag{3.50}$$

The vector $\tilde{\mathbf{p}} = \mathbf{D}_p \mathbf{p}$ consists of unknown source parameters. We can now think of performing MUSIC on $\tilde{\mathbf{v}}_f$ to localize more sources, provided $\mathbf{U}_{sum}$ satisfiesthe identifiability condition in Def. 9.

3.8.6 Spatial Smoothing based 2D SOFI-MUSIC

The expression (3.49) is similar to the measurement model (3.45) of a single snapshot DOA estimation problem, where the sources are captured by the sum co-array $\mathbf{U}_{sum}$ instead of the physical array $\mathbf{U}$. We follow similar ideas as in [124, 125, 223, 224] and construct a PSD matrix of suitable rank on which we can apply SOFI-MUSIC algorithm.

For ease of exposition, we will assume that the physical array is a rectangular separable 2D nested array [124]. In this case, the sum co-array $\mathbb{S}_{sum}$ has locations given by

$$\mathbb{S}_{sum} = \{[i,j]^T, 0 \leq i \leq I-1, 0 \leq j \leq J-1\}$$

In Fig.3.2, we give an example of physical nested array and its sum co-array. For a 2D nested array with $O(I+J)$ sensors, $\mathbb{S}_{sum}$ is a uniform rectangular array with IJ elements.

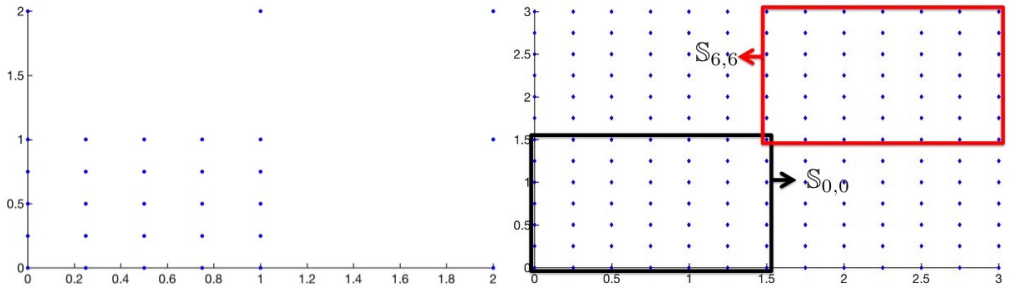


Figure 3.2: (Left) Physical sensor locations. (Right) Sum co-array locations

The sum co-array consists of overlapping subarrays as follows

$$\mathbb{S}_{m,n}(i,j) = \mathbb{S}_{0,0}(i,j) + [m,n]^T$$

where $\mathbb{S}_{0,0}$ is the fundamental subarray given by

$$\mathbb{S}_{0,0}(i,j) = [i,j]^T, \qquad 0 \leq i < I_1, 0 \leq j < J_1$$

It can be easily seen that a total of $N_m N_n$ shifted subarrays can be obtained from $\mathbb{S}_{sum}$ where $N_m = I + 1 - J_1, N_n = J + 1 - J_1$. Corresponding to each subarray, we can partition $\tilde{\mathbf{v}}_f$ into $N_m N_n$ overlapping vectors $\tilde{\mathbf{v}}_f^{(l)}, l = 0, \cdots, N_m N_n - 1$ and define a smoothed covariance matrix as

$$\mathbf{R}_{sum} = \sum_{l=0}^{N_m N_n - 1} \tilde{\mathbf{v}}_f^{(l)} \tilde{\mathbf{v}}_f^{(l)T}$$

It can be shown [124, 125] that if $I_1 = \frac{I+1}{2}, J_1 = \frac{J+1}{2}$, $\mathbf{R}_{sum}$ has the form

$$\mathbf{R}_{sum} = (\mathbf{S}_{0,0} \mathbf{P} \mathbf{S}_{0,0}^T)^2 \tag{3.51}$$

where $\mathbf{P} = \text{diag}(p_0, \cdots, p_{L-1})$ and $\mathbf{S}_{0,0}$ is specified by

$$[\mathbf{S}_{0,0}]_{J_1 i + j, k} = e^{2\alpha\left(x_k^\star i + y_k^\star j\right)}$$

with $\mathbf{r}_k^\star = [x_k^\star, y_k^\star]^T$.

From (3.51), it can seen that the fundamental subarray $\mathbb{S}_{0,0}$ (and equivalently $\mathbf{S}_{0,0}$) controls the performance of 2D MUSIC algorithm. Similar to Theorem 13, we have the following guarantee for resolving more sources via spatial smoothing with 2D nested array:

Theorem 14 *Assuming* $\mathbf{P}$ *is diagonal, the source locations* $\{\mathbf{r}_k^\star\}_{k=0}^{L-1}$ *can be exactly recovered via spatially smoothed SOFI-MUSIC if* $\mathbf{S}_{0,0}$ *satisfies the identifiability condition in Def. 9.*

3.8.7 Experiments

We generate synthetic data similar to [220], to verify the SOFI-MUSIC algorithm. In the experiments, the PSF is assumed to be Gaussian with $\alpha = 100$ in (3.41). There are 9 point sources in the ground truth and 36 sensors are placed on a 6×6 grid with separation 0.16. The

correlation matrix $\mathbf{P} = 10\mathbf{I}$ for uncorrelated sources. In Fig. 3.3, we compare the performances of SOFI-MUSIC and SPARCOM. It can be seen that SOFI-MUSIC outperforms SPARCOM in terms of the reconstructed source locations.

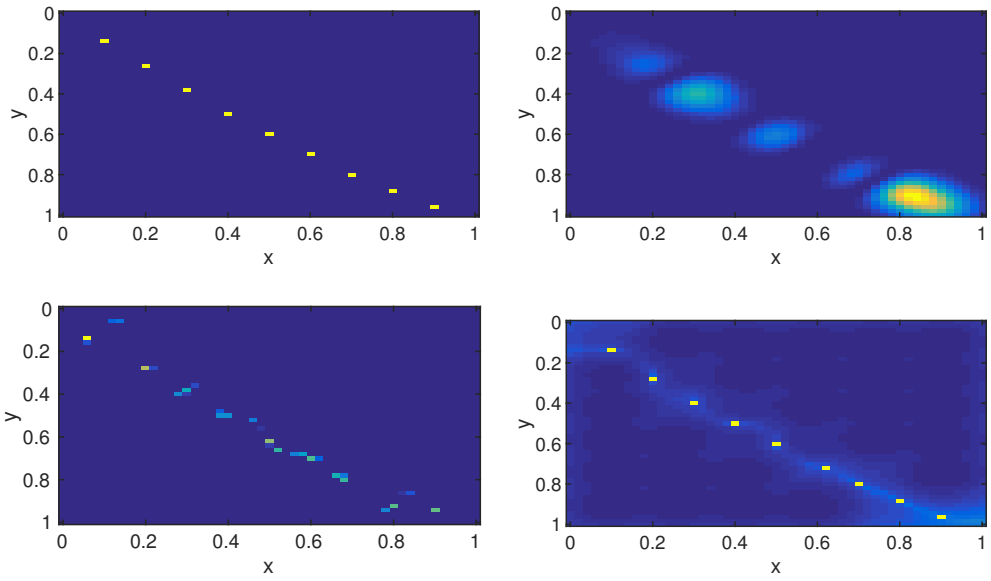


Figure 3.3: (Upper Left) Ground truth of emitter locations. (Upper Right) Diffraction-limited images obtained using Gaussian PSF averaged over 50 frames (Bottom Left) SPARCOM with $\lambda = 10^{-4}$ (Bottom Right) The log-spectrum of SOFI-MUSIC

3.9 Numerical Results

3.9.1 Non-convex Algorithms for Super-Resolution with Non-negative Constraint

We now conduct numerical experiments to demonstrate that the proposed reweighted iterative algorithm for approximating $l_{1/2}$ quasinorm minimization can produce better estimate of $\mathbf{x}^\star$ both in terms of sparsity and smaller estimation error. We choose $N = 64$, $n = 21$, and the true sparsity is set at $\|\mathbf{x}^\star\|_0 = 6$. The non-zero entries of $\mathbf{x}^\star$ are produced by first generating uniform

random variables between 1 and 2 and then normalizing the entries so that $\|\mathbf{x}^\star\|_1 = 1$. Similarly, the measurement noise $\mathbf{w}$ is produced by generating complex standard Gaussian random variables and then normalizing $\mathbf{w}$ such that $\|\mathbf{w}\|_1 = \delta_1$. We will compare the performance of different algorithms by varying δ_1. To implement Algorithm 1, we set the stopping parameter to $\varepsilon = 0.001$ and select $\epsilon_n = \frac{10^{-4}}{n}$.

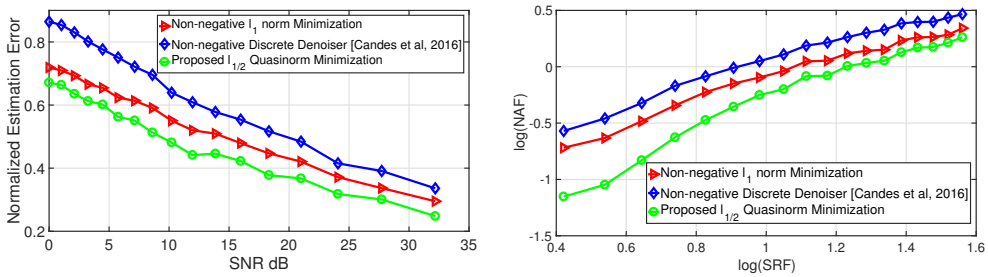


Figure 3.4: (Left) Comparative performance of different algorithms as a function of Signal-to-Noise ratio (SNR). The results are averaged over 1200 Monte Carlo runs. (Right) Noise Amplification Factor (NAF) of (P_{den}), (P_1) and proposed $l_{1/2}$ minimization, as a function of Super-Resolution Factor (SRF) $\frac{N}{n-1}$. The results are averaged over 800 runs.

In Fig. 3.4, we compare the performance of (P_{den}), (P_1), and (P_2) [11] by varying δ_1 which represents the l_1 norm of the noise $\mathbf{w}$. The Signal-to-Noise ratio (SNR) is defined as $20\log(\frac{\|\mathbf{x}^\star\|_1}{\|\mathbf{w}\|_1})$. It can be seen that the proposed algorithm produces the smallest normalized estimation error $\|\mathbf{x}^\# - \mathbf{x}^\star\|_1/\|\mathbf{x}^\star\|_1$ and outperforms both (P_{den}) and (P_1).

We finally demonstrate the reconstruction quality of the Algorithm 1 for 2D super-resolution. The proposed algorithm can be readily extended to two dimensions by choosing $\mathbf{Q}$ as a 2D DFT matrix. Fig. 3.5 shows the performance of Algorithm 1 and (P_1) on synthetic $2D$ data. The ground truth is a sparse $N \times N$ image where $N = 24$. We generate the low-frequency measurements by only retaining the 49 low frequency DFT coefficients. We further normalize $\mathbf{w}$ so that $\|\mathbf{w}\|_1 = 0.1$. It can be clearly seen that Algorithm 1 exactly recovers the true support while (P_1) produces several false peaks. This further corroborates the fact that the proposed $l_{1/2}$

[11] we choose $p = 1$ for solving (P_{den}) and (P_1)

minimization framework indeed favors and identifies sparser solutions.

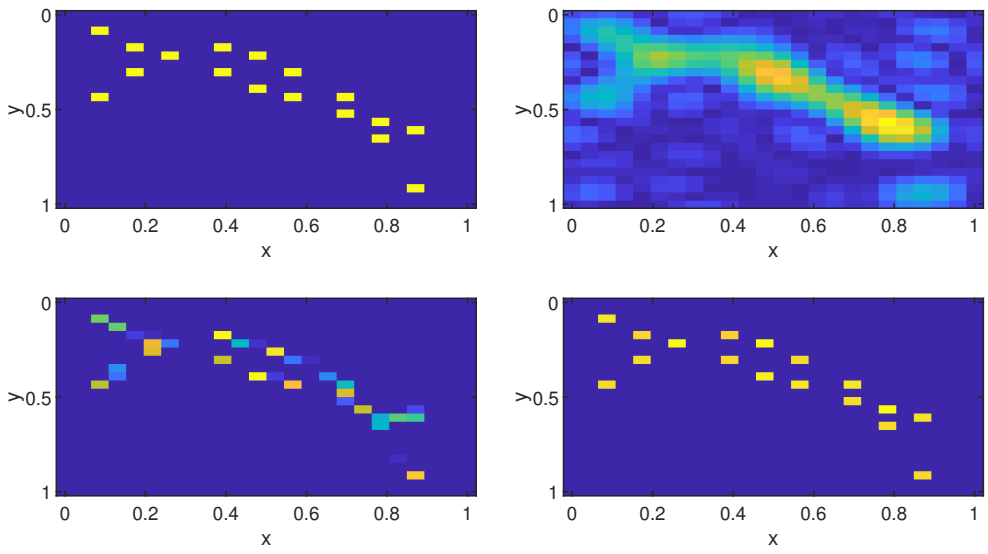


Figure 3.5: (Top Left) Ground truth image with positive emitters. (Top Right) Measured image consisting of only low frequency components. (Bottom Left) Estimate produced by solving convex problem (P_1) (Bottom Right) Estimate produced by the proposed iterative $l_{1/2}$ minimization algorithm.

3.9.2 Simulations for Joint Support Recovery Using Difference Co-Arrays

We now perform a series of numerical experiments to study the performance of sparse arrays for support recovery in the regime $s > M$. We implement the proposed correlation-aware support recovery framework where $\mathbf{A}$ represents the array manifold of a nested array (an OOSA) given by (3.12).The joint support $\mathcal{S}$ is generated uniformly at random obeying the separation condition in Def. 5. The non-zero elements of $\mathbf{p}^\star$ are assumed to be all equal to 10 (i.e. $p_{\min} = 10$) and the noise power is assumed to be $\sigma = 1$. We consider two metrics for evaluating performance: (i) Power Estimation Error, and (ii) Success Rate of Support Recovery.

Similar to [170–172, 182], we claim that support recovery is successful if the support of the s largest non-zero entries of $\mathbf{p}^{\#}$ is identical to the true support $\mathcal{S}$. Notice that $(P_{\text{Co-den}})$ actually

represents a class of optimization problems, which can be convex or non-convex depending on the choice of $f(\mathbf{z})$. In this paper, we consider the following three specific instances of this family

- Case 1: $f(\mathbf{z}) = \|\mathbf{z}\|_1$. We call this problem $(P1_{\text{Co-den}})$ and implement it using standard CVX package.
- Case 2: $f(\mathbf{z}) = \|\mathbf{z}\|_0$. We call this problem $(P0_{\text{Co-den}})$. Since it is a non-convex problem, we approximate it using a reweighted l_1 minimization heuristic [157].
- Case 3: $f(\mathbf{z}) = \|\mathbf{z}\|_{1/2}$. We call this $(P_{1/2,\text{Co-den}})$. This is a non-convex quasi-norm minimization problem with non-negative constraints, which we solve using a majorization-minimization based heuristic [14].

Power Estimation Error and the Universal Upper Bound

We first compare the normalized power estimation error $\|\mathbf{p}^{\#} - \mathbf{p}^{\star}\|_1/\|\mathbf{p}^{\star}\|_1$(denoted by NE) for each of $(P1_{\text{Co-den}})$, $(P0_{\text{Co-den}})$ and $(P_{1/2,\text{Co-den}})$ against the theoretical upper bound from (3.18). For this experiment, we let $M = 12$, $N = 200$ and $s = 20$. We vary the number of snapshots L and compute the empirical mean covariance perturbation error $\|\Delta_{\mathbf{r}}\|_2$ for each L by averaging over 100 runs. Fig. 3.6 shows the mean normalized power estimation error as a function of the (empirical) mean covariance perturbation $\|\Delta_{\mathbf{r}}\|_2$, establishing the validity of the upper bound (3.18) for both convex and non-convex problems.

Comparison of Support Recovery as a function of L and s

We compare the success rate of $(P1_{\text{Co-den}})$ (using nested arrays), against two conventional MMV algorithms: MMV-BP [201] and RA-ORMP [172]. [12] Similar to the experimental settings in [172] [13], we implement these algorithms either using a random Gaussian $\mathbf{A}$ with i.i.d zero-mean

[12] We implemented MMV-BP using the SPGL1 solver for MATLAB available at http://www.cs.ubc.ca/ mpf/spgl1/.

[13] Since MMV-BP and RA-ORMP do not utilize correlation of the measurements, they are unable to exploit the difference co-array of sparse arrays and hence are unsuitable to be used with such measurement matrices.

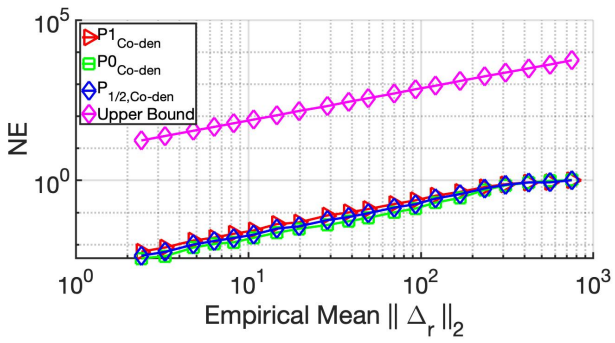
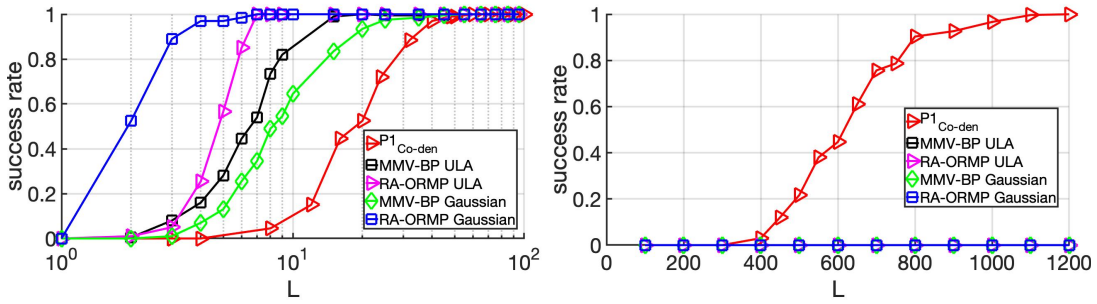


Figure 3.6: Mean Normalized Power Estimation Error and the Upper Bound (3.18) as a function of $\|\Delta_{\mathbf{r}}\|_2$. Here $M = 12$, $N = 200$, $s = 20$, and $\mathbf{A}$ is generated using a nested array with $M_{ca} = 41$. The results are averaged over 100 runs.

unit-variance elements, or the manifold of a uniform linear array (ULA) with elements given by $\mathbf{A}_{m,n} = e^{-j2\pi mn/N}, 1 \leq m \leq M, 0 \leq n \leq N-1$. Let $M = 24, N = 400$. Fig. 3.7(a) shows the success rate of the algorithms as a function of L in two distinct regimes (a) $s = 8 < M$ and (b) $s = 72 > M$. When $s < M$, Fig. 3.7(a) shows that RA-ORMP with random Gaussian measurement matrix $\mathbf{A}$ has the highest success rate for a given L. On the other hand, when $s > M$, MMV-BP, and RA-ORMP completely fail to localize the sources since they do not use the correlation structure of the data. However, $(P1_{\text{Co-den}})$ successfully recovers the sources in this regime with probability > 0.9 when $L > 800$. Notice that the number of snapshots (L) needed to ensure success rate of 1 is higher in this case than when $s < M$. Fig. 3.8 shows the success rate of all the algorithms as a function of s for a fixed $L = 100$. As expected MMV-BP and RA-ORMP are unable to recover supports of size $s > M = 24$ while $(P1_{\text{Co-den}})$ is able to recover supports of size significantly larger than 25.

Comparison with Vector Approximate Message Passing

Very recently, the authors of [213] proposed a new sparse signal reconstruction algorithm called Vector Approximate Message Passing (VAMP). Compared to traditional Approximate Message Passing [214], VAMP is applicable to a larger class of measurement matrices $\mathbf{A}$. In this section, we conduct numerical experiments to compare $(P1_{\text{Co-den}})$ with VAMP, especially in the



(a) (b)

Figure 3.7: Probability of successful support recovery as a function of L. Here, $M = 24, N = 400$, and the results are averaged over 300 runs. (a) $s = 8 < M = 24$. (b) $s = 72 > M = 24$.

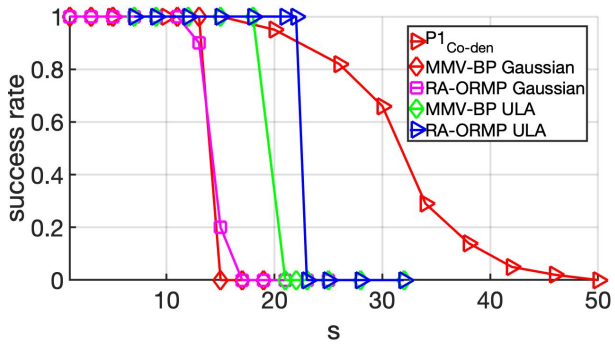


Figure 3.8: Probability of successful support recovery as a function of sparsity s. Here, $M = 24, N = 300, L = 100$. For nested array, we have $M_{ca} = 155$. Results are averaged over 100 runs.

regime $s > M$. We implemented VAMP using the code from the website of the authors of [213] and used similar simulation settings as suggested by the demo code. The non-zero entries of signal and noise are generated as zero-mean i.i.d Gaussian random variables. The nonzero signal powers are assumed to be equal to p, while the noise power is denoted by σ^2. We implemented VAMP with two types of measurement matrices: (i) $\mathbf{A}$ representing a nested array and constructed according to (3.9), and (ii) $\mathbf{A}$ with i.i.d Gaussian entries as defined in the demo code of [213]. VAMP returns an estimate for $\mathbf{X}_L$ and we estimate the support by retaining the indices of s rows of this estimate with the largest l_2 norm. We consider two simulation settings:

- **Case 1:** We use i.i.d Gaussian $\mathbf{A}$ for VAMP and choose $N = 200, p = 10000, \sigma^2 = 1, L = 1000$.
- **Case 2:** We use $\mathbf{A}$ corresponding to a nested array for VAMP and choose $N = 200, p = 1000, \sigma^2 = 1, L = 100$.

For both cases, we implement $(P1_{\text{Co-den}})$ using nested arrays. The probability of successful support recovery as a function of sparsity s is plotted in Fig. 3.9. From the simulations, it appears that when $\mathbf{A}$ corresponds to a nested array, VAMP fails to detect support in the regime $s > M = 150$. On the other hand, when $\mathbf{A}$ is an i.i.d. Gaussian matrix, both algorithms are able to recover supports of size larger than M. For example, both $(P1_{\text{Co-den}})$ and VAMP can recover supports of size larger than 100 with $M = 36$ and $M = 80$ measurements respectively.

Phase Transition

We study phase transition of $(P1_{\text{Co-den}})$, MMV-BP and RA-ORMP by plotting their success rates as functions of both M and s. As before, we implement $(P1_{\text{Co-den}})$ using nested arrays, while we use a ULA measurement matrix for MMV-BP and RA-ORMP. We fix $N = 600, L = 2000$. For each (M, s) pair, we compute the success rate over 100 Monte Carlo runs. Fig. 3.10 shows the average success rate with respect to the number of sensors M and number of

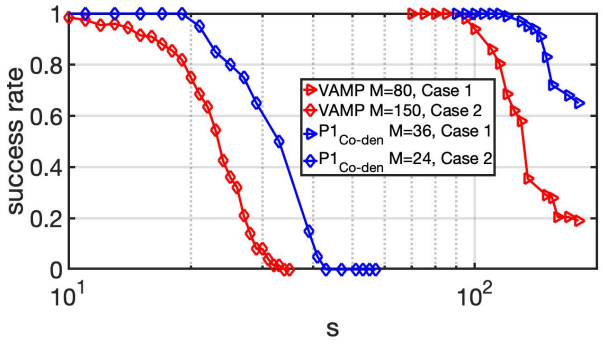


Figure 3.9: Comparative probability of successful support recovery for $(P1_{\text{Co-den}})$ and VAMP as a function of sparsity s. Results are averaged over 100 runs.

sources s, where black pixels represent zero success rate and white pixels denote 100% success rate. For comparison, we also overlay the curve $s = 0.18M^2$ on the phase transition plot of $(P1_{\text{Co-den}})$ which roughly matches the phase transition boundary and validates our claim that it is possible to localize $s = \Theta(M^2)$ sources with nested arrays. The phase transition regions of MMV-BP and RA-ORMP, on the other hand, show that the number of resolvable sources scales linearly with M, matching their theoretical performance limits [172, 201].

Achievability of Upper Bound

In Theorem 12, we showed that for fixed M, there exist $\mathbf{p}_1^\star$ and $\mathbf{p}_1^\#$ such that (3.38) holds. In this section, we demonstrate this scaling with respect to N by plotting $\|\mathbf{A}_{\text{KR}}^{\mathbb{U}}(\mathbf{p}_1^\star - \mathbf{p}_1^\#)\|_2$ against N for fixed $\|\mathbf{p}_1^\star - \mathbf{p}_1^\#\|_1$. We generate $\mathbf{p}_1^\star$ and $\mathbf{p}_1^\#$ according to the proof of Theorem 12, such that $\mathbf{p}_1^\star$ contains a single spike and $\mathbf{p}_1^\#$ consists of two neighboring equal power spikes given by (3.68). Fig. 3.11 clearly shows that $\|\mathbf{A}_{\text{KR}}^{\mathbb{U}}(\mathbf{p}_1^\star - \mathbf{p}_1^\#)\|_2$ indeed scales as $\Theta(N^2)$.

Performance of Other "Correlation-Aware" Algorithms for MMV models

In our final set of experiments, we demonstrate the superior performance of a family of related "correlation-aware" algorithms for MMV models, namely M-SBL [181], M-FOCUSS [174] and SPICE [120] that also utilize the correlation of the measurements, and are potentially

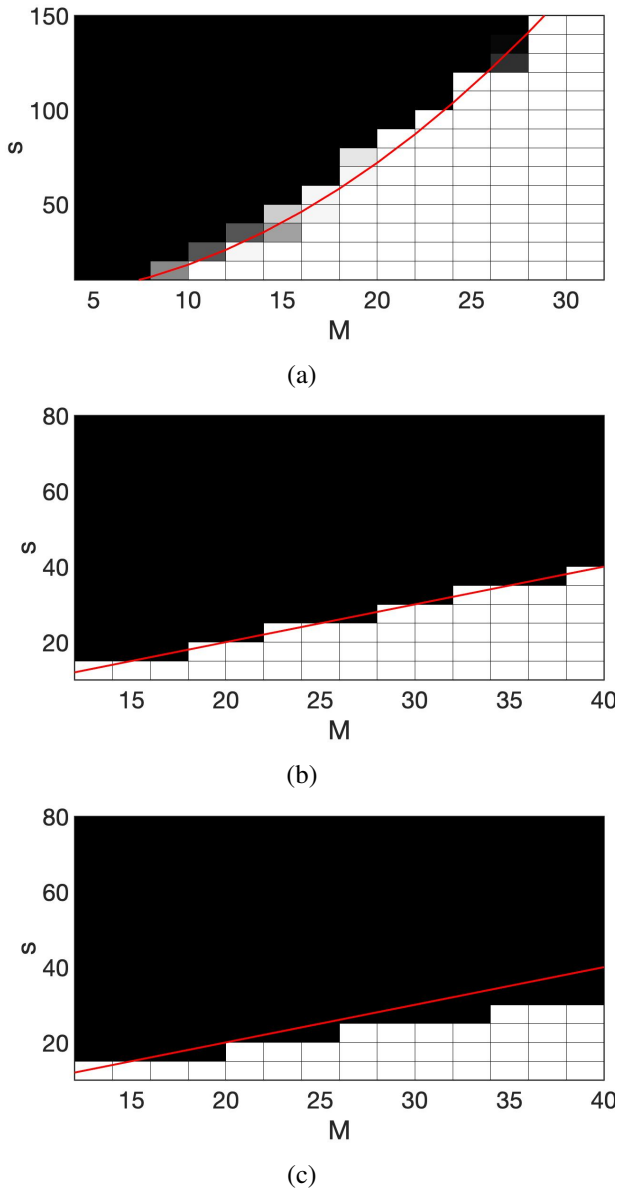


Figure 3.10: Phase transition of success rate as function of sparsity s and number of measurements M: (a) $(P1_{\text{Co-den}})$, (b) RA-ORMP, and (c) MMV-BP. White pixels indicate perfect recovery and black pixels denote total failure. Here $L = 2000, N = 600$ and the results are averaged over 50 runs. The overlaid red curve represents $s = 0.18M^2$ in (a) and $s = M$ in (b) and (c).

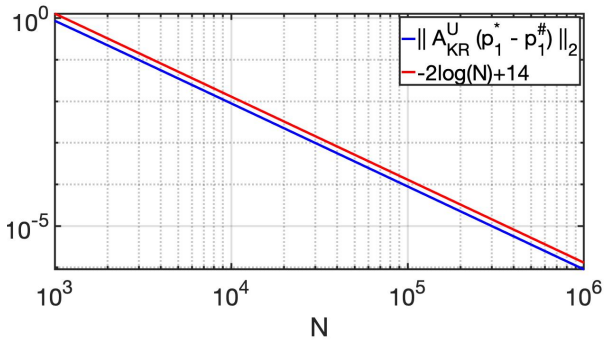


Figure 3.11: Log-log plot of $\|\mathbf{A}^{\mathbb{U}}_{\text{KR}}(\mathbf{p}_1^\star - \mathbf{p}_1^\#)\|_2$ as a function of N, for fixed $\|\mathbf{p}_1^\star - \mathbf{p}_1^\#\|_1 = 2$. Here $\mathbf{p}_1^\star$ and $\mathbf{p}_1^\#$ are constructed according to the proof of Theorem 12. The plot is overlaid with the function $-2\log N + 14$. The two plots share the same slope of -2, demonstrating the order-wise tightness of (3.17) with respect to N.

capable of resolving more sources than sensors when used with OOSA. However, these algorithms are non-convex and difficult to analyze, especially with finite L. Nonetheless, they show excellent empirical performance with nested arrays, as demonstrated in Fig 3.12. We fix $M = 24, N = 300, L = 100$, and plot the success rate of these three algorithms as a function of L, for both ULA and nested array (with same number of sensors). It can be seen that MSBL, MFOCUSS, and SPICE have significantly higher success rate with nested arrays, and cannot resolve more sources than sensors with ULA. Hence, Fig. 3.12 empirically demonstrates that exploiting the difference co-array of OOSA can help these algorithms identify more sources than sensors with finite measurements. A rigorous *non-asymptotic* analysis of this behavior is a very interesting problem for future research.

3.10 Conclusion

In this chapter we analyzed the problem of super-resolution where the desired signal is both sparse and non-negative. We proposed a constrained non-convex $l_{1/2}$ quasinorm minimization problem to promote sparsity in the reconstructed signal. Such a formulation naturally stems from exploiting non-negative constraints on the signal. Although $l_{1/2}$ quasinorm is non-convex and

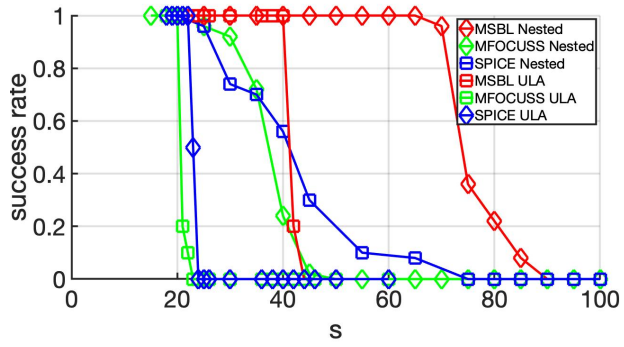


Figure 3.12: Success rate of M-SBL, M-FOCUSS and SPICE as a function of sparsity s. Here $M = 24, N = 300, L = 100$. For nested array, $M_{ca} = 155$. The results are averaged over 200 runs.

non-differentiable and the measurement matrix does not satisfy RIP, the stability of the solution can still be guaranteed by constructing appropriate dual certificates. An iterative reweighted l_1 minimization algorithm is proposed to approximate the $l_{1/2}$ quasinorm and simulations show that it has better performance than l_1 norm minimization, in terms of both accuracy and sparsity of the solution.

By exploiting the difference co-array of order-optimal sparse arrays, we developed probabilistic guarantees on exact support recovery in the regime $s > M$, as a function of the number of snapshots L, that is applicable for a large class of "correlation-aware" support recovery algorithms. The guarantees involve developing upper bounds on the estimation error of source powers for any algorithm in this class, which was shown to scale quadratically with the dimension N. As a result of independent interest, we showed that such quadratic dependence is tight (for a given number M of sensors).

In future, we will extend the analysis framework of this paper to other non-convex "correlation-aware" algorithms such as M-SBL which exhibited excellent empirical performances in our simulations. Another interesting direction is to extend our analysis to the "gridless" case by integrating results from [215] to characterize the performance of subspace based algorithms that exploit the co-array structure. Finally, the proof of achievability in Sec. 3.7 can be generalized to understand how the locations of spikes fundamentally affect estimation error.

Chapter 3, in part, is a reprint of the material as it appears in the papers:

- H. Qiao and P. 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*, vol. 67, no. 22, pp. 5715-5729, Nov. 2019.

- H. Qiao and P. Pal, "A Non-Convex Approach to Non-Negative Super-Resolution: Theory and Algorithm", in the *Proceedings of 44th International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, May 2019.

- H. Qiao and P. Pal, " On modulus of continuity for noisy positive super-resolution, in the *Proceedings of 43nd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2018)*, Calgary, Canada.

- H. Qiao and P. Pal, "Performance Limits of Covariance-Driven Super Resolution Imaging", in the *Proceedings of Asilomar Conference on Signals, Systems and Computers*, 2017.

- H. Qiao and P. Pal,"Unified Analysis of Co-Array Interpolation for Direction-of-Arrival Estimation, in the *Proceedings of 42nd International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, March 2017.

The dissertation author was the primary investigator and author of these papers.

3.11 Appendices

3.11.1 Proof of Theorem 1

Since $\mathbf{A}$ represents an OOSA, $\mathbf{A}_{\text{KR}} = \mathbf{A}^* \odot \mathbf{A}$ contains a Vandermonde sub-matrix $\mathbf{A}_{\text{KR}}^{\mathbb{U}}$ given by (3.14). We further select the last $M_{ca} + 1$ rows of $\mathbf{A}_{\text{KR}}^{\mathbb{U}}$, and collect them in a matrix

$\mathbf{V} \in \mathbb{C}^{M_{ca}+1 \times N}$ given by

$$\mathbf{V}_{m,n} = e^{-j2\pi mn/N}, 0 \leq m \leq M_{ca}, 0 \leq n \leq N-1$$

Let $\mathbf{v_y} \triangleq \text{vec}(\mathbf{R_{yy}})$. From (3.16), we can extract an appropriate sub-vector of $\mathbf{v_y}$, viz. $\mathbf{v_y^{\mathbb{U}^+}} \in \mathbb{C}^{M_{ca}+1}$ such that

$$\mathbf{v_y^{\mathbb{U}+}} = \mathbf{r_y^{\mathbb{U}^+}} + \sigma^2 \mathbf{e}_1 = \mathbf{V}_{\mathcal{S}} \mathbf{p}^{\star}_{\mathcal{S}} + \sigma^2 \mathbf{e}_1 \tag{3.52}$$

where $\mathbf{e}_1 = [1, 0, \cdots, 0]^T \in \mathbb{R}^{M_{ca}+1}$, the matrix $\mathbf{V}_{\mathcal{S}} \in \mathbb{C}^{M_{ca}+1 \times s}$ consists of the columns of $\mathbf{V}$ indexed by $\mathcal{S}$, and $\mathbf{p}^{\star}_{\mathcal{S}}$ is the vector of non-zero entries of $\mathbf{p}^{\star}$. Given $\mathbf{R_{yy}}$, the Co-array MUSIC algorithm [126] produces an estimate $\hat{\mathcal{S}}$ of the support $\mathcal{S}$ as follows:

Step 1: Extract the vector $\mathbf{v_y^{\mathbb{U}^+}}$ from $\mathbf{R_{yy}}$ and construct a Hermitian Toeplitz matrix $\mathbf{T} \in \mathbb{C}^{(M_{ca}+1) \times (M_{ca}+1)}$ with $\mathbf{v_y^{\mathbb{U}^+}}$ as the first column.

Step 2: Let $\hat{\sigma}^2$ be the smallest singular value of $\mathbf{T}$, and $\mathbf{U} \in \mathbb{C}^{M_{ca}+1 \times K}$ be an orthonormal basis for Null$(\mathbf{T} - \hat{\sigma}^2 \mathbf{I})$ where $K = \dim(\text{Null}(\mathbf{T} - \hat{\sigma}^2 \mathbf{I}))$. Construct the set $\hat{\mathcal{S}} \subset [0, N-1]$ such that

$$\hat{\mathcal{S}} = \{i \in [0, N-1] \quad \text{s.t. } \|\mathbf{U}^H \mathbf{v}_i\|_2 = 0\}$$

Step 3: Return $\hat{\mathcal{S}}$ and $\hat{\sigma}$ as estimates for $\mathcal{S}$ and σ.

We show that as long as $s \leq M_{ca}$, we will have

$$\hat{\mathcal{S}} = \mathcal{S}, \quad \hat{\sigma} = \sigma$$

Firstly, it can be verified that

$$\mathbf{T} = \mathbf{V}_{\mathcal{S}} \text{diag}(\mathbf{p}_{\mathcal{S}}^{\star}) \mathbf{V}_{\mathcal{S}}^{H} + \sigma^2 \mathbf{I} \tag{3.53}$$

Since $\mathbf{p}_{\mathcal{S}}^{\star} > \mathbf{0}$ and $s \leq M_{ca}$, the smallest singular value of $\mathbf{T}$ is σ^2, implying that $\hat{\sigma} = \sigma$ and $\text{rank}(\mathbf{T} - \hat{\sigma}^2 \mathbf{I}) = s$. Therefore, from (3.53), we have

$$\mathbf{V}_{\mathcal{S}}^{H} \mathbf{u} = \mathbf{0}, \quad \forall \mathbf{u} \in \text{Null}(\mathbf{T} - \hat{\sigma}^2 \mathbf{I})$$

which implies that $\mathcal{S} \subseteq \hat{\mathcal{S}}$. Suppose there exists $k_0 \in \hat{\mathcal{S}}, k_0 \notin \mathcal{S}$. Then, we must have

$$[\mathbf{V}_{\mathcal{S}}, \mathbf{v}_{k_0}]^{H} \mathbf{u} = \mathbf{0}, \forall \mathbf{u} \in \text{Null}(\mathbf{T} - \hat{\sigma}^2 \mathbf{I}) \tag{3.54}$$

Now $[\mathbf{V}_{\mathcal{S}}, \mathbf{v}_{k_0}] \in \mathbb{C}^{(M_{ca}+1) \times (s+1)}$ is a row-Vandermonde matrix and hence its $(s+1)$ columns must be linearly independent since $s \leq M_{ca}$. However, (3.54) implies that these $s+1$ columns should also be orthogonal to $\text{Null}(\mathbf{T} - \hat{\sigma}^2 \mathbf{I})$. Since $\text{rank}(\mathbf{T} - \hat{\sigma}^2 \mathbf{I}) = s$, the orthogonal complement of $\text{Null}(\mathbf{T} - \hat{\sigma}^2 \mathbf{I})$ has dimension s and cannot contain $s+1$ independent vectors. Hence such a k_0 cannot exist, implying $\hat{\mathcal{S}} = \mathcal{S}$. Once the support and σ are correctly recovered, the vector $\mathbf{p}_{\mathcal{S}}^{\star}$ can be recovered as the unique solution to the system of equations (3.52), since $\mathbf{V}_{\mathcal{S}}$ is full-column rank.

3.11.2 Proof of Theorem 2

Consider the matrix $\tilde{\mathbf{A}}_{\text{KR}}^{\mathbb{U}} \in \mathbb{C}^{N \times N}$ obtained by zero-padding $\mathbf{A}_{\text{KR}}^{\mathbb{U}}$ as follows:

$$\tilde{\mathbf{A}}_{\text{KR}}^{\mathbb{U}} = [\mathbf{0}_{(\lceil N/2 \rceil - M_{ca} - 1) \times N}^{T}, (\mathbf{A}_{\text{KR}}^{\mathbb{U}})^{T}, \mathbf{0}_{(\lfloor N/2 \rfloor - M_{ca}) \times N}^{T}]^{T} \tag{3.55}$$

It can be easily verified that $\tilde{\mathbf{A}}_{\text{KR}}^{\mathbb{U}} = \mathbf{D}_N \mathbf{F}_N$ where the elements of $\mathbf{F}_N \in \mathbb{C}^{N\times N}$ are given by $[\mathbf{F}_N]_{k,l} = e^{-j2\pi kl/N}, -\lceil N/2\rceil + 1 \leq k \leq \lfloor N/2\rfloor, 0 \leq l \leq N-1$ and

$$\mathbf{D}_N = \text{diag}(\hat{d}_{-\lceil N/2\rceil+1}, \cdots \hat{d}_{\lfloor N/2\rfloor})$$

is a diagonal matrix with

$$\hat{d}_m = \begin{cases} 1, & -M_{ca} \leq m \leq M_{ca} \\ 0, & \text{otherwise} \end{cases}$$

We utilize the following lemma (Lemma 1 in [61]) adapted to our measurement model.

Lemma 10 *[61] If $\mathbf{p}^\star \in \mathcal{P}_{sep}^+$, $M_{ca} \geq 128$ and $N \geq 3.03(2M_{ca}+1)$, then for any non-negative vector $\hat{\mathbf{p}} \in \mathbb{R}^N$ and $\mathbf{h} = \hat{\mathbf{p}} - \mathbf{p}^\star$, there exists $\mathbf{q} = [q_0, \cdots, q_{N-1}]^T \in \mathbb{R}^N$ such that*

$$\tilde{\mathbf{A}}_{KR}^{\mathbb{U}}\mathbf{q} \;=\; \mathbf{F}_N\mathbf{q}, \quad \|\mathbf{q}\|_\infty \leq 1$$

where $\rho = 0.0295\left(\frac{M_{ca}}{N}\right)^2$ and

$$q_l = 0 \;\; \textit{if}\; h_l < 0\,;\; q_l > 2\rho \;\; \textit{if}\; h_l \geq 0$$

Given the existence of the vector $\mathbf{q}$ from Lemma 10, define $\tilde{\mathbf{q}} \triangleq \mathbf{q} - \rho\mathbf{1}$. Let $\mathbf{h} = \mathbf{p}^\# - \mathbf{p}^\star$. Then, we have

$$|\langle \tilde{\mathbf{q}}, \mathbf{h}\rangle| = |\sum_{l=0}^{N-1} \tilde{q}_l h_l| = \sum_{l=0}^{N-1} |\tilde{q}_l||h_l| \geq \rho\|\mathbf{h}\|_1 \tag{3.56}$$

On the other hand, using properties of $\mathbf{q}$ from Lemma 10, it follows that

$$\tilde{\mathbf{A}}^{\mathbb{U}}_{\text{KR}}(\mathbf{q}-\rho\mathbf{1}) = \mathbf{F}_N(\mathbf{q}-\rho\mathbf{1})$$

Then, we have

$$\begin{aligned}
&|\langle\tilde{\mathbf{q}},\mathbf{h}\rangle| = |\langle\frac{1}{N}\mathbf{F}_N^H\tilde{\mathbf{A}}^{\mathbb{U}}_{KR}\tilde{\mathbf{q}},\mathbf{h}\rangle| = |\langle\frac{1}{N}\mathbf{F}_N^H\mathbf{D}_N\mathbf{F}_N\tilde{\mathbf{q}},\mathbf{h}\rangle| \\
&= |\langle\tilde{\mathbf{q}},\frac{1}{N}\mathbf{F}_N^H\tilde{\mathbf{A}}^{\mathbb{U}}_{\text{KR}}\mathbf{h}\rangle| \leq \|\tilde{\mathbf{q}}\|_2\|\frac{1}{N}\mathbf{F}_N^H\tilde{\mathbf{A}}^{\mathbb{U}}_{\text{KR}}\mathbf{h}\|_2 \\
&= \|\tilde{\mathbf{q}}\|_2\frac{1}{\sqrt{N}}\|\tilde{\mathbf{A}}^{\mathbb{U}}_{\text{KR}}\mathbf{h}\|_2 \leq (1-\rho)\|\mathbf{A}^{\mathbb{U}}_{\text{KR}}\mathbf{h}\|_2 \qquad (3.57)
\end{aligned}$$

where we used the facts that $1-\rho > \rho$, $\mathbf{F}_N^H\mathbf{F}_N = N$, $\|\tilde{\mathbf{q}}\|_\infty \leq 1-\rho$ and $\|\mathbf{A}^{\mathbb{U}}_{\text{KR}}\mathbf{h}\|_2 = \|\tilde{\mathbf{A}}^{\mathbb{U}}_{\text{KR}}\mathbf{h}\|_2$. Combining (3.56) and (3.57), we have

$$\|\mathbf{h}\|_1 \leq \frac{1-\rho}{\rho}\|\mathbf{A}^{\mathbb{U}}_{\text{KR}}\mathbf{h}\|_2$$

which completes the proof.

3.11.3 Proof of Theorem 9

Let $\mathbf{p}^{\#}$ be a solution to $(P_{\text{Co-den}})$. If $\epsilon \geq \|\Delta^{\mathbb{U}}_{\mathbf{r}}\|_2$, we can use the bound (3.18) in Corollary 4 to obtain

$$\|\mathbf{p}^{\#}_{\mathcal{S}^c}\|_1 \leq \frac{1-\rho}{\rho}(\|\Delta^{\mathbb{U}}_{\mathbf{r}}\|_2+\epsilon), \qquad (3.58)$$

$$\|\mathbf{p}^{\star}_{\mathcal{S}} - \mathbf{p}^{\#}_{\mathcal{S}}\|_1 \leq \frac{1-\rho}{\rho}(\|\Delta^{\mathbb{U}}_{\mathbf{r}}\|_2+\epsilon) \qquad (3.59)$$

If we choose $T \geq \frac{1-\rho}{\rho}(\|\Delta^{\mathbb{U}}_{\mathbf{r}}\|_2+\epsilon)$, then from (3.59), it follows that $[\mathcal{H}_T(\mathbf{p}^{\#})]_i = 0$ if $i \in \mathcal{S}^c$. This implies that the support of $\mathcal{H}_T(\mathbf{p}^{\#})$ will be contained in $\mathcal{S}$. On the other hand, suppose

$p_{\min} > 2\frac{1-\rho}{\rho}(\|\Delta_{\mathbf{r}}^{\mathbb{U}}\|_2 + \epsilon)$ and we choose T in the range $T < p_{\min} - \frac{1-\rho}{\rho}(\|\Delta_{\mathbf{r}}^{\mathbb{U}}\|_2 + \epsilon)$. Then, from (3.59), it follows that

$$[\mathbf{p}_{\mathcal{S}}^{\#}]_i \geq p_{min} - \frac{1-\rho}{\rho}(\|\Delta_{\mathbf{r}}^{\mathbb{U}}\|_2 + \epsilon) > T \quad i \in \mathcal{S},$$

implying that the support of $\mathcal{H}_T(\mathbf{p}^{\#})$ will contain $\mathcal{S}$. Summarizing, if

$$p_{\min} > 2\frac{1-\rho}{\rho}(\|\Delta_{\mathbf{r}}^{\mathbb{U}}\|_2 + \epsilon), \quad \epsilon \geq \|\Delta_{\mathbf{r}}^{\mathbb{U}}\|_2 \tag{3.60}$$

and the threshold T is selected from the range

$$\frac{1-\rho}{\rho}(\|\Delta_{\mathbf{r}}^{\mathbb{U}}\|_2 + \epsilon) \leq T < p_{\min} - \frac{1-\rho}{\rho}(\|\Delta_{\mathbf{r}}^{\mathbb{U}}\|_2 + \epsilon) \tag{3.61}$$

we can ensure that $\text{Supp}\Big(\mathcal{H}_T(\mathbf{p}^{\#})\Big) = \mathcal{S}$. Hence, if $p_{\min}$ is large enough and T is chosen from a suitable range, it is possible to exactly recover $\mathcal{S}$ via hard-thresholding. Notice that the results so far are conditioned on the covariance estimation error $\Delta_{\mathbf{r}}$. We now make these guarantees probabilistic by utilizing appropriate concentration bound on $\Delta_{\mathbf{r}}$ that exploits the Gaussian distribution of the measurements $\mathbf{Y}_L$. Since $\Delta_{\mathbf{r}}$ is a complex random variable (as $\mathbf{A}$ is complex), we first represent each complex measurement vector $\mathbf{y}[l]$ in terms of real and imaginary parts as $\mathbf{y}[l] = \mathbf{y}_{\mathbb{R}}[l] + j\mathbf{y}_{\mathbb{I}}[l]$. Define

$$\mathbf{v}[l] = [\mathbf{y}_{\mathbb{R}}^T[l], \mathbf{y}_{\mathbb{I}}^T[l]]^T$$

Since $\{\mathbf{y}[l]\}_{l=1}^L$ are i.i.d complex Gaussian vectors, $\mathbf{v}[l]$ are i.i.d and real-valued Gaussian vectors. Define the true and empirical covariance matrices of $\mathbf{v}[l]$ as

$$\mathbf{R}_{\mathbf{vv}} = \mathbb{E}\Big(\mathbf{v}[l]\mathbf{v}^T[l]\Big) \quad \hat{\mathbf{R}}_{\mathbf{vv}} = \frac{1}{L}\sum_{l=1}^{L}\mathbf{v}[l]\mathbf{v}^T[l]$$

It can be easily verified that

$$\|\mathbf{R_{yy}} - \hat{\mathbf{R}}_{\mathbf{yy}}\|_F \leq \sqrt{2}\|\mathbf{R_{vv}} - \hat{\mathbf{R}}_{\mathbf{vv}}\|_F$$
$$\text{trace}(\mathbf{R_{yy}}) = \text{trace}(\mathbf{R_{vv}}) \tag{3.62}$$

We now invoke the following concentration bound on the covariance estimation error of *real-valued* Gaussian vectors.

Lemma 11 *(Proposition A.3 in [129]) Let $\{\mathbf{z}[l]\}_{l=1}^{L}$ be zero mean i.i.d real Gaussian random vectors distributed as $\mathbf{z}[l] \sim \mathcal{N}(\mathbf{0}, \mathbf{R_z})$. Then,*

$$\mathbb{P}\left\{\|\mathbf{R_z} - \hat{\mathbf{R}}_{\mathbf{z}}\|_F \geq \frac{\text{trace}(\mathbf{R_z})}{\sqrt{L}}\right\} \leq 2e^{-2c\sqrt{L}} \tag{3.63}$$

where $\hat{\mathbf{R}}_{\mathbf{z}} = \frac{1}{L}\sum_{l=1}^{L}\mathbf{z}[l]\mathbf{z}[l]^T$ and c is a positive universal constant.

Proof. (of Theorem 9) Since $\mathbf{v}[l]$ are real zero-mean i.i.d Gaussian vectors with covariance $\mathbf{R_{vv}}$, Lemma 17 implies

$$\mathbb{P}\left\{\|\mathbf{R_{vv}} - \hat{\mathbf{R}}_{\mathbf{vv}}\|_F \geq \frac{\text{trace}(\mathbf{R_{vv}})}{\sqrt{L}}\right\} \leq 2e^{-2c\sqrt{L}} \tag{3.64}$$

Recall from (3.15) that $\|\Delta_f\|_2 = \|\mathbf{R_{yy}} - \hat{\mathbf{R}}_{\mathbf{yy}}\|_F$. Using (3.62) and (5.25), it follows that

$$\mathbb{P}\left\{\|\Delta_f\|_2 \geq \frac{\sqrt{2}\text{trace}(\mathbf{R_{yy}})}{\sqrt{L}}\right\} = \tag{3.65}$$
$$\mathbb{P}\left\{\|\mathbf{R_{yy}} - \hat{\mathbf{R}}_{\mathbf{yy}}\|_F \geq \frac{\sqrt{2}\text{trace}(\mathbf{R_{yy}})}{\sqrt{L}}\right\} \leq 2e^{-2c\sqrt{L}}$$

Notice that $\mathrm{trace}(\mathbf{R_{yy}}) = \mathrm{trace}(\mathbf{A}\mathrm{diag}(\mathbf{p}^\star)\mathbf{A}^H + \sigma^2\mathbf{I}) = M(\|\mathbf{p}^\star\|_1 + \sigma^2)$. If (3.21) holds, we can say that

$$\frac{\sqrt{2}\mathrm{trace}(\mathbf{R_{yy}})}{\sqrt{L}} < \frac{\rho}{1-\rho}\frac{p_{min}}{4} - \sigma^2 \tag{3.66}$$

Since $\|\Delta_{\mathbf{r}}^{\mathbb{U}}\|_2 \leq \sigma^2 + \|\Delta_f\|_2$, the concentration bound (3.65) implies that, with probability exceeding $1 - 2e^{-2c\sqrt{L}}$,

$$\|\Delta_{\mathbf{r}}^{\mathbb{U}}\|_2 < \frac{\sqrt{2}\mathrm{trace}(\mathbf{R_{yy}})}{\sqrt{L}} + \sigma^2 < \frac{\rho}{1-\rho}\frac{p_{min}}{4} \tag{3.67}$$

Since ϵ and T are selected according to (3.23) and (3.24) respectively, it can be easily seen that (3.67) ensure that the sufficient conditions (3.20) and (3.21) hold with probability exceeding $1-2e^{-2c\sqrt{L}}$. Hence, we conclude that $\mathrm{Supp}(\mathcal{H}_T(\mathbf{p}^\#)) = \mathcal{S}$ with probability at least $1-2e^{-2c\sqrt{L}}$.

3.11.4 Proof of Theorem 12

Let l be an integer in the range $0 < l < N-1$. We construct $\mathbf{p}_1^\star \in \mathbb{R}^N$ and $\mathbf{p}_1^\# \in \mathbb{R}^N$ as follows

$$[\mathbf{p}_1^\star]_n = \frac{\eta}{2}\delta[n-l], [\mathbf{p}_1^\#]_n = \frac{\eta}{4}(\delta[n-l-1] + \delta[n-l+1])$$

where $\delta[n]$ denotes the Kronecker delta function. Obviously, $\mathbf{p}_1^\star$ satisfies the separation condition as it contains only one spike. Notice that $\|\mathbf{p}_1^\star - \mathbf{p}_1^\#\|_1 = \eta$. Define $\mathbf{s} = [s_{-M_{ca}}, \cdots, s_{M_{ca}}]^T \triangleq \mathbf{A}_{\mathrm{KR}}^{\mathbb{U}}(\mathbf{p}_1^\star - \mathbf{p}_1^\#)$. From the definition of $\mathbf{A}_{\mathrm{KR}}^{\mathbb{U}}$, we have

$$s_m = \sum_{k=0}^{N-1} e^{-j2\pi mk/N}([\mathbf{p}_1^\star]_k - [\mathbf{p}_1^\#]_k), \quad -M_{ca} \leq m \leq M_{ca}$$

We can further simplify s_m as

$$\begin{aligned} s_m &= \frac{\eta}{2}\sum_{k=0}^{N-1} e^{-j2\pi mk/N}(\delta_l - \frac{1}{2}\delta_{l-1} - \frac{1}{2}\delta_{l+1}) \\ &= \eta \sin^2(\frac{\pi m}{N}) e^{-j2\pi ml/N} \end{aligned}$$

Then, $\|\mathbf{s}\|_2^2 = \sum_{m=-M_{ca}}^{M_{ca}} |s_m|^2 = \eta^2 \sum_{m=-M_{ca}}^{M_{ca}} \sin^4(\frac{\pi m}{N})$

Using Taylor series expansion with Lagrange remainder, we have

$$\sin^4(\frac{\pi m}{N}) = \frac{\pi^4 m^4}{N^4} + \frac{f^{(6)}(\xi)}{6!}\frac{\pi^6 m^6}{N^6} \quad |m| \leq M_{ca}$$

where $f(x) = \sin^4(x)$ and $0 < \xi < \frac{\pi m}{N}$. Thus,

$$\begin{aligned} \|\mathbf{s}\|_2^2 &= \eta^2 \sum_{m=-M_{ca}}^{M_{ca}} \frac{\pi^4 m^4}{N^4} + \frac{f^{(6)}(\xi)}{6!}\frac{\pi^6 m^6}{N^6} \\ &= \frac{\eta^2\pi^4}{N^4}\left(\frac{2}{5}M_{ca}^5 + M_{ca}^4 + \frac{2}{3}M_{ca}^3 - \frac{1}{3}M_{ca}\right) \\ &+\eta^2 \sum_{m=-M_{ca}}^{M_{ca}} \frac{f^{(6)}(\xi)}{6!}\frac{\pi^6 m^6}{N^6} \\ &\leq \frac{\eta^2\pi^4}{N^4}\left(\frac{2}{5}M_{ca}^5 + M_{ca}^4 + \frac{2}{3}M_{ca}^3 - \frac{1}{3}M_{ca}\right) + c_1^2\eta^2\frac{M_{ca}^7}{N^6} \end{aligned}$$

for any constant c_1 such that $c_1^2 \geq \max_{\xi\in[0,\frac{\pi M_{ca}}{N}]} 2\pi^6 \left|\frac{f^{(6)}(\xi)}{6!}\right|$. For $M_{ca} \geq 16$, it can be verified that

$$\frac{2}{5}M_{ca}^5 + M_{ca}^4 + \frac{2}{3}M_{ca}^3 - \frac{1}{3}M_{ca} \leq \frac{M_{ca}^4}{15}7M_{ca}$$

Using the fact that for positive numbers α, β, $\sqrt{\alpha+\beta} \leq \sqrt{\alpha} + \sqrt{\beta}$, it follows that

$$\|\mathbf{A}_{\text{KR}}^{\mathbb{U}}(\mathbf{p}_1^\star - \mathbf{p}_1^\#)\|_2 \leq \eta\Big(\frac{\pi^2 M_{ca}^2}{\sqrt{15}N^2}\sqrt{7M_{ca}} + \frac{cM_{ca}^3}{N^3}\sqrt{M_{ca}}\Big) \tag{3.68}$$

Therefore, we finally have

$$\|\mathbf{p}_1^{\#} - \mathbf{p}_1^{\star}\|_1 = \eta \geq \frac{1}{\left(\pi^2\sqrt{\frac{7}{15}}M_{ca}^{2.5} + cM_{ca}^{2.5}\frac{M_{ca}}{N}\right)} N^2 \|\mathbf{A}_{\text{KR}}^{\mathbb{U}}(\mathbf{p}_1^{\star} - \mathbf{p}_1^{\#})\|_2 \tag{3.69}$$

It can be verified that whenever $\frac{M_{ca}}{N} \leq 0.25$,

$$\max_{\xi \in [0, \frac{\pi M_{ca}}{N}]} 2\pi^6 \left|\frac{f^{(6)}(\xi)}{6!}\right| \leq 1.43\pi^6 \tag{3.70}$$

Since $N > 3.03(2M_{ca}+1)$ (i.e.$\frac{M_{ca}}{N} \leq 0.165$), we can choose c_1 as $c_1 = \sqrt{1.43\pi^6} = 1.1958\pi^3$. Using this value of c_1 and the fact that $\frac{M_{ca}}{N} \leq 0.165$, (3.37) follows from (3.69) where the constant K is given by $K = \pi^2\sqrt{\frac{7}{15}} + 0.165c_1 = \pi^2\sqrt{\frac{7}{15}} + 0.1973\pi^3$.

Chapter 4

Phase Retrieval and Structured Fourier Sampler

The problem of phase retrieval is considered in this chapter, where the measurement vectors are deterministic and Fourier-like. Under some mild assumptions, $O(s\log N)$ measurements are proved to be sufficient to recover an s-sparse complex vector of dimension N from its phaseless measurement via convex programming. The key contribution is to show that unlike existing work in sparse phase retrieval, the so-called "collision-free" condition is not needed for the proposed approach, and hence, there is no upper bound on s for which the sparse vector can be recovered. Even for non sparse complex data, the number of measurements needed by this approach almost attains the lower bound conjectured in current literature. The algorithms developed in this work are based on a newly introduced class of Fourier samplers, namely Partial Nested Fourier Samplers, which can naturally avoid the "collision-free" condition by performing a novel decoupling of quadratic terms arising in the phaseless measurements.

4.1 Introduction

Reconstruction of signals from magnitudes of linear samples has recently received great research interest. In many problems, the measurement scheme is such that only the amplitude of the linear measurement is kept and the phase information is lost. The problem of recovering the original signal from these "phaseless" measurements is known as *phase retrieval* or *phaseless reconstruction*. The earliest work in phase retrieval [75] draws its motivation from speech processing where the authors show that it is possible to reconstruct signals up to some minor ambiguities only from magnitude measurements. Besides speech processing, phase retrieval has applications in X-ray imaging, crystallography, electron-microscopy, coherence theory, quantum mechanics, differential geometry and other fields [76, 77].

Two types of measurement models are primarily studied in literature: Fourier based, and random. In [77, 81–83, 92], the authors consider the Fourier measurement model for acquiring phaseless measurements. We will propose novel variations of the Fourier measurement model to establish the claims of this paper. In [70, 79, 84, 97], the authors use random measurements to exploit ideas from compressed sensing for performing phase retrieval. The phase retrieval problem is converted to a low-rank matrix recovery problem in [77] which allows convex programming based solutions. It is shown that under some mild conditions, the convex approach produces the desired solution even with Fourier measurements [79]. In [75, 76], the authors consider abstract measurement models for deducing universal lower bound (sufficient conditions on the number of measurements).

The number of phaseless measurements can be significantly reduced if the underlying $\mathbf{x}$ is s-sparse (i.e. with s non zero elements). Using standard DFT measurement vectors, the phase retrieval problem is often transformed into that of recovering $\mathbf{x}$ from its autocorrelation function (ACF) [84]. Following [85], the authors impose the so-called *collision-free* [84] condition on $\mathbf{x}$ so that the ACF is guaranteed to be sparse. Under this condition, $O(s^2)$ Fourier measurements

are proved to be sufficient for uniquely recovering $\mathbf{x}$ up to a global phase. In [81, 82], the authors are able to avoid the collision-free property by employing a combinatorial recovery algorithm. However, the sparsity s cannot be larger than $O(N^{1/3})$. In [171], $O(s\log(N/s))$ is proved to be sufficient for sparse phase retrieval when random measurements are used. In [87], the authors follow the convex formulation of [79] and show that $O(s^2\log(N/s^2))$ is enough for phase retrieval. Finally, in [93], the authors adopt the idea of [164] and prove that $O(s\log N)$ Fourier measurements are sufficient for recovering a sparse $\mathbf{x}$ after it is modulated using four different masks. The interested readers may refer to [78], for an overview of the current state-of-art in the phase retrieval algorithms.

In this chapter, we introduce a new design of Fourier measurement vectors, namely the Partial Nested Fourier Sampler (PNFS), drawing inspiration from our past and current work in nested sampling and its extensions [2, 10, 45]. As will be demonstrated, the idea of partial nested sampling is highly effective for the phase retrieval problem since it naturally allows decoupling of terms arising in the equivalent quadratic measurement model. We show that $O(s\log N)$ PNFS measurements are sufficient for recovering $\mathbf{x}$ up to global phase factor by solving a suitable l_1 programming problem. Unlike [84], the PNFS can completely avoid the need for a collision-free condition on $\mathbf{x}$ and hence there is no restriction on the maximum size of the sparse support in our framework. In contrast to [93], we do not need masks to modulate $\mathbf{x}$ and our algorithm may be easier to be implemented in practice. Furthermore, for a non-sparse complex $\mathbf{x}$, the PNFS needs only $4N - 5$ Fourier measurements using a simple reconstruction scheme, that comes very close to the universal lower bound conjectured in current literature [76].

Then, we further develop the theory of PNFS for sparse phase retrieval by proposing a randomized version of the basic PNFS, namely the R-PNFS. By using a certain decoupling property of the R-PNFS, along with a new cancellation based algorithm (that effectively cancels out certain unwanted quadratic terms in the autocorrelation of the signal), we are able to demonstrate that $O(s \log N)$ measurements are sufficient to recover the sparse signal with probability 1.

We also prove that the proposed algorithm is stable in presence of bounded noise, and present numerical simulations to validate the theoretical claims.

4.2 Problem Setting and Related Work

4.2.1 Problem Setting

Let $\mathbf{x} \in \mathbb{H}^N$ be an unknown vector of interest where the field $\mathbb{H}$ can either be $\mathbb{R}$ or $\mathbb{C}$, and $\mathcal{F} = \{\mathbf{f}_1, \mathbf{f}_2, \cdots, \mathbf{f}_M\}$ be a set of M measurement vectors where $\mathbf{f}_i \in \mathbb{H}^N$. To be precise, we have the following general measurement model

$$y_i = |< \mathbf{x}, \mathbf{f}_i >| \quad 1 \leq i \leq M \tag{4.1}$$

where $\mathbf{y} = [y_1, y_2, \cdots, y_M]^T \in \mathbb{R}_+^M$. Hence, we only collect a set of magnitude (or energy) measurements and we lose the phase information of $< \mathbf{x}, \mathbf{f}_i >$. The key question in phase retrieval is: *Can we recover* $\mathbf{x}$ *from phaseless measurements* $\mathbf{y}$*?* If the answer is positive, we will say that the measurement vectors $\mathcal{F}$ are successful at phase retrieval [76]. Before answering this question, it is straightforward to notice that there is some intrinsic ambiguity in recovering unknown $\mathbf{x}$ from (5.1). When $\mathbb{H} = \mathbb{R}$, we cannot distinguish $\pm\mathbf{x}$ as they provide the same measurements. Similarly, if $\mathbb{H} = \mathbb{C}$, we will not be able to separate $\mathbf{x}$ from its ambiguities $\tilde{\mathbf{x}}$ where $\tilde{\mathbf{x}} = c\mathbf{x}$ with $|c| = 1$. So, $\mathbf{x}$ can only be recovered in sense of $\mathbb{H} \setminus \mathbb{T}$ where $\mathbb{T} = \{\pm 1\}$ for $\mathbb{H} = \mathbb{R}$ and $\mathbb{T}$ denotes the unit circle in the complex plane if $\mathbb{H} = \mathbb{C}$. Hence $\mathbf{x}$ can be at best recovered upto a global phase ambiguity.

As pointed out in [77], the phase retrieval task is tightly related to that of recovering low-rank matrices using quadratic measurements. Particularly, we have

$$y_i^2 = \mathbf{f}_i^H \mathbf{x}\mathbf{x}^H \mathbf{f}_i \quad 1 \leq i \leq M \tag{4.2}$$

Since $\mathbf{x}\mathbf{x}^H$ is rank 1, phase retrieval can be interpreted as a rank-one matrix recovery problem given measurement $\mathbf{y}$ [77, 79]. Equivalently, the quadratic form (4.2) can be written as

$$y_i^2 = \left(\mathbf{f}_i^T \otimes \mathbf{f}_i^H\right) \text{Vec}\left(\mathbf{x}\mathbf{x}^H\right) \tag{4.3}$$

We will use this equivalent representation of the phaseless measurements throughout this paper. We will mainly focus on the case where $\mathbf{f}_i$ are complex Fourier-type [77, 81, 84, 92] measurement vectors, and $\mathbf{x}$ will be assumed to be s-sparse with $\mathbb{H} = \mathbb{C}$.

4.2.2 Sufficient Condition with Generic Measurements

In [75, 76], sufficient and necessary conditions have been discussed for both non-sparse and sparse phase retrieval problems. In particular, we have the following lemma from [76].

Lemma 12 *[76] Let $\mathbb{H} = \mathbb{C}$. A set $\mathcal{F}$ with $M \geq 4s - 2$ generically chosen vectors in $\mathbb{C}^N$ succeeds in phase retrieval of s-sparse vector $\mathbf{x}$.*

By *generic*, we mean an open dense subset of the set of all $M-$element frames in $\mathbb{H}^N$ [75]. To the best of our knowledge, the necessary bound for complex case has not yet been proved. It is conjectured that $M \geq 4N - 4$ (or $M \geq 4s - 2$) is necessary for recovering arbitrary $\mathbf{x}$ (or s-sparse) in the complex case [76, 78]. *It should be noticed that the generic measurement vectors $\mathcal{F}$ considered in [75, 76] may not be Fourier.* Hence these necessary and sufficient conditions do not necessarily apply to Fourier measurements.

4.2.3 Related Work with Fourier Measurements

We first define a general model for Fourier measurement vector as follows.

Definition 10 *(**General Fourier Measurement**:) A General Fourier Measurement (GFM) vector*

is defined as

$$\mathbf{f}_i = \alpha \left[z_i^{n_1}, z_i^{n_2}, \cdots, z_i^{n_N}\right]^T \tag{4.4}$$

where z_i is on the unit circle in complex plane, α is a normalizing constant and $\mathcal{N} = \{n_1, n_2, \cdots, n_N\}$ are non-negative integers.

Typically, two general kinds of Fourier measurement vectors are considered in literature. In [77, 81, 92, 93, 95, 164], the measurement vectors are drawn from conventional DFT matrix. Specifically, in this case, $\mathcal{N} = \{0, 1, \cdots, N-1\}$, $\alpha = N^{-1/2}$ and $z_i = e^{j2\pi(i-1)/N}$. On the other hand, in [84, 85], the authors transform the phase retrieval problem to that of recovering $\mathbf{x}$ from its autocorrelation vector $\mathbf{r_x} \in \mathbb{C}^{2N-1}$ defined as

$$[\mathbf{r_x}]_l = \sum_{k=\max\{1,1-l\}}^{\min\{N,N-l\}} x_k \bar{x}_{k+l} \quad 0 \leq |l| \leq N-1 \tag{4.5}$$

where $\bar{x}$ denotes the complex conjugate of x. An important observation in this regard is *even when $\mathbf{x}$ is s-sparse, $\mathbf{r_x}$ may not be sparse*. In order to impose a required sparsity on $\mathbf{r_x}$ (so that l_1 minimization can be applied on $\mathbf{r_x}$ to recover the support) the so-called *collision-free condition* is proposed.

Definition 11 (Collision Free Condition) *[84, 85] A sparse vector $\mathbf{x}$ has collision free property if for pairs of distinct entries $(p, q), (m, n)$ in the support of $\mathbf{x}$, $p - q \neq m - n$ unless $(p, q) = (m, n)$.*

Given any set of distinct integers of size s, we have a total of $(s^2 - s)/2$ positive pair-wise differences. So if the dimension of $\mathbf{x}$ is N, the upper bound for s such that an s-sparse $\mathbf{x}$ satisfies collision-free condition is given by $s \leq \chi_N \triangleq (1 + \sqrt{8N-7})/2$. With collision-free property, $\mathbf{r_x}$ is guaranteed to have sparsity $s^2 - s + 1$ [84] and compressed sensing based recovery algorithms can be used with $M = O(s^2)$ measurements. In a later section, we show by simulations that in

practice, collision-free property holds only for a small range of values for s, and the upper limit can be very small compared to N. This prompted us to find another way to reconstruct a sparse $\mathbf{x}$ (without any restriction on the sparsity s) given the quadratic measurement model (4.3) and develop the main results of this chapter.

4.3 Nested Fourier Measurements and Phase Retrieval

4.3.1 Nested Fourier Measurement and Decoupling

We know that for a sparse $\mathbf{x}$, its autocorrelation sequence $\mathbf{r_x}$ may not be sparse. One way to enforce sparsity on $\mathbf{r_x}$ is to impose a *collision-free* condition (see Def. 11) on the indices of support. However, the collision-free condition only holds for small values of s. As a major contribution of this paper, we now propose a Fourier type measurement model namely the Partial Nested Fourier Sampler (PNFS), built upon the nested sampling idea in [2,10,45], that completely avoids the need for a collision-free condition and provides good performance guarantees.

Definition 12 (Partial Nested Fourier Sampler:) *We define a Partial Nested Fourier Sampler (PNFS) as a special form of GFM vector defined in (4.4) where* $\mathcal{N} = \{1, 2, \cdots, N-1, 2N-2\}$, $\alpha = (4N-5)^{-1/4}$ *and* $z_i = e^{j2\pi(i-1)/(4N-5)}$.

Substituting this choice of $\mathbf{f}_i$ in (4.3) and combining identical columns, we have

$$y_i^2 = \frac{1}{\sqrt{4N-5}} \left[z_i^{-(2N-3)}, \cdots, z_i^{-1}, 1, z_i^1, \cdots, z_i^{2N-3} \right] \tilde{\mathbf{x}} \tag{4.6}$$

where $\tilde{\mathbf{x}} \in \mathbb{C}^{4N-5}$ is the corresponding rearranged version of $\text{Vec}(\mathbf{x}\mathbf{x}^H)$ with following form

$$[\tilde{\mathbf{x}}]_m = \begin{cases} \sum_{k=1}^{N} |x_k|^2 & m = 0 \\ \sum_{k=1}^{N-1-m} x_k \bar{x}_{k+m} & m = 1, 2, \cdots, N-2 \\ x_{2N-2-m} \bar{x}_N & N-1 \leq m \leq 2N-3 \\ \overline{[\tilde{\mathbf{x}}]_{-m}} & m < 0 \end{cases} \tag{4.7}$$

where we re-number the indices of $\tilde{\mathbf{x}}$ in range $[-2N+3, 2N-3]$ for simplicity and clearance.

Decoupling Effect And Basics of Recovery: The most important property of PNFS is that for $|m| \geq N-1$, $[\tilde{\mathbf{x}}]_m$ only consists of single terms instead of a sum. Moreover, each of these terms has a constant factor x_N. **The important advantage of decoupled products is that if x_N is nonzero, the sparsity of the sub-vector consisting of $\tilde{\mathbf{x}}$ for $|m| \geq N-1$ reveals the support of $\mathbf{x}$.** In addition, for $s \geq 2$, $[\tilde{\mathbf{x}}]_m$ will vanish for $|m| \geq N-1$ if and only if $x_N = 0$. However, without any prior knowledge about the support of $\mathbf{x}$, there is no guarantee that x_N is nonzero and for this reason, we define the following column-permuted version of the PNFS sampling vector $\mathbf{f}_i$ as

$$\mathbf{f}_i^{(l)} = \frac{1}{\sqrt[4]{4N-5}} \left[z_i^1, z_i^2, \cdots, z_i^{N-1}, z_i^{2N-2} \right] \mathbf{\Pi}^{(l)} \tag{4.8}$$

where $z_i = e^{j2\pi(i-1)/(4N-5)}$ and $\mathbf{\Pi}^{(l)}$ is a permuting matrix such that the vector $\mathbf{x}^{(l)} = \mathbf{\Pi}^{(l)}\mathbf{x}$ satisfies $[\mathbf{x}^{(l)}]_l = x_N, [\mathbf{x}^{(l)}]_N = x_l, [\mathbf{x}^{(l)}]_i = x_i, i \neq l, N$. The basic idea of using the permuted PNFS vector is that for some l, we can ensure that $[\mathbf{x}^{(l)}]_N$ is non zero. For that choice of l, we can then recover $\tilde{\mathbf{x}}^{(l)}$ from measurements $y_i^{(l)}$ with model (5.1), and use the decoupled entries

(guaranteed to be non zero since $[\mathbf{x}^{(l)}]_N \neq 0$) to estimate the support of $\mathbf{x}^{(l)}$ (or equivalently of $\mathbf{x}$) and the corresponding non zero elements (upto a global phase ambiguity). For each l, we collect $\tilde{M}$ phaseless measurements $y_i^{(l)}, i = 1, 2, \cdots, \tilde{M}$ using the permuted PNFS vector (4.8) and obtain

$$\tilde{\mathbf{y}}^{(l)} = \mathbf{Z}\tilde{\mathbf{x}}^{(l)} \tag{4.9}$$

where $[\tilde{\mathbf{y}}^{(l)}]_i = (y_i^{(l)})^2, [\mathbf{Z}]_{i,m} = \frac{1}{\sqrt{4N-5}} e^{j2\pi \frac{(i-1)m}{4N-5}}, 1 \leq i \leq \tilde{M}, \quad -2N+3 \leq m \leq 2N-3$. It is easy to see that $\mathbf{Z}$ is invertible if $\tilde{M} = 4N-5$ and $\tilde{\mathbf{x}}^{(l)}$ can be recovered from $\tilde{\mathbf{y}}^{(l)}$.

4.3.2 Iterative Algorithm

We now describe the details of an iterative algorithm that uses the permuted PNFS vectors iteratively to find a non zero entry of $\mathbf{x}$. Noting that $[\tilde{\mathbf{x}}^{(l)}]_m$ will be all zero for $|m| \geq N-1$ if and only if the last entry is non zero given $s \geq 2$, the algorithm starts with $l = N$ and reduces l in each step until it finds a non zero x_l. It then successively recovers $\tilde{\mathbf{x}}^{(l)}$ and $\mathbf{x}$ upto a global phase. Table 1 summarizes the algorithm

4.3.3 Performance Guarantees of the Iterative Algorithm

In this section, we will show $\mathbf{x}^{\#}$ is a valid estimation of $\mathbf{x}$ provided the iterative algorithm is feasible. Obviously, the number of measurements needed is determined by the smallest index l_{min} such that $x_{l_{min}}$ is non zero in Table 1. We define best case for $l_{min} = N$ and worst case for $l_{min} = s$. Formally, we have the following result

Theorem 15 *Let $\mathbf{x} \in \mathbb{C}^N$ be s-sparse with $s \geq 3$. The estimate $\mathbf{x}^{\#}$ produced by the iterative algorithm described in Table 1 is equal to $\mathbf{x}$ (in the sense of $\mathbb{C} \setminus \mathbb{T}$) if the total number of phaseless measurements M equals $4N-5$ for the best case and $(N-s+1)(4N-5)$ for the worst case.*

Table 4.1: Iterative Algorithm for Phase Retrieval using PNFS

Input: data $\mathbf{x}$ *Output:* estimation $\mathbf{x}^{\#}$

1. *Initialization:* $l = N$
2. *Loop:*
   (a) **Step S1:** Using the permuted PNFS vectors (4.8), obtain $4N-5$ phaseless measurements

   $$y_i^{(l)} = | < \mathbf{x}, \mathbf{f}_i^{(l)} > |, i = 1, 2, \cdots 4N-5$$

   Using (4.9), recover $\tilde{\mathbf{x}}^{(l)} = \mathbf{Z}^{-1}\tilde{\mathbf{y}}^{(l)}$

   (b) **Step S2:** If $[\tilde{\mathbf{x}}^{(l)}]_m = 0, \forall |m| \geq N-1$, declare $x_l = 0$. Assign $l \to l-1$ and go back to Step **S1**.
   If $[\tilde{\mathbf{x}}^{(l)}]_m \neq 0$ for some m with $|m| \geq N-1$, proceed to the recovery stage.
3. *Recovery:*
   (a) Choose $m^* \in \{1, 2, \cdots, N-2\}$ such that $[\tilde{\mathbf{x}}^{(l)}]_{m^*} \neq 0$ and compute

   $$|x_N^{(l)}| = \sqrt{[\tilde{\mathbf{x}}^{(l)}]_{m^*}/\beta}$$

   & $\beta = \sum_{k=1}^{N-1-m^*} [\tilde{\mathbf{x}}^{(l)}]_{2N-2-k} \overline{[\tilde{\mathbf{x}}^{(l)}]}_{2N-2-k-m^*}$
   (b) Obtain estimate $\mathbf{x}^{\#}$ as

   $$[\mathbf{x}^{\#}]_q = \begin{cases} \left(\frac{[\tilde{\mathbf{x}}^{(l)}]_{2N-2-q}}{|x_N^{(l)}|}\right) & q \neq \{l, N\} \\ |x_N^{(l)}| & q = l \\ \frac{[\tilde{\mathbf{x}}^{(l)}]_{2N-2-l}}{|x_N^{(l)}|} & q = N \end{cases}$$

Proof. It is easy to see that if $\mathbf{x}$ has sparsity s, the total number of iterations required by the algorithm is 1 in the best case (i.e. when $x_N \neq 0$) and $N - s + 1$ in the worst case. In each iteration, we collect $4N - 5$ phaseless measurements $y_i^{(l)}$, and hence we need to collect $4N - 5$ measurements in the best case and $(4N-5)(N-s+1)$ measurements in the worst case. The final step is then to show the correctness of this algorithms in recovering $\mathbf{x}$ upto a global phase. We prove correctness for the case when the algorithm terminates in 1 step (i.e. when x_N is non zero) since the proof remains identical for other cases just by exchanging l and N. The first idea in the proof is to show the existence of m^* such that $[\tilde{\mathbf{x}}^{(l)}]_{m^*} \neq 0$. Denote $\check{\mathbf{x}} = [x_1, x_2, \cdots, x_{N-1}]^T$ and let $\mathbf{r}_{\check{\mathbf{x}}} \in \mathbb{C}^{2N-3}$ be the autocorrelation vector of $\check{\mathbf{x}}$. Suppose m^* does not exist, implying $[\tilde{\mathbf{x}}]_m = 0$ for $1 \leq |m| \leq N-2$. Hence, $[\mathbf{r}_{\check{\mathbf{x}}}]_n = \gamma\delta(n)$ where $\gamma = [\tilde{\mathbf{x}}]_0 - |x_N|^2$ and $\delta(n)$ is Kronecker delta. This means that $\hat{\mathbf{r}}_{\check{\mathbf{x}}}(e^{j\omega}) \triangleq \sum_{n=-N+2}^{N-2} [\mathbf{r}_{\check{\mathbf{x}}}]_n e^{-j\omega n}$ is an all-pass filter. However, $\hat{\mathbf{r}}_{\check{\mathbf{x}}}(e^{j\omega}) = |\hat{\check{\mathbf{x}}}(e^{j\omega})|^2$ where $\hat{\check{\mathbf{x}}}(e^{j\omega}) \triangleq \sum_{n=-N+2}^{N-2} [\check{\mathbf{x}}]_n e^{-j\omega n}$. This implies $\hat{\check{\mathbf{x}}}(e^{j\omega})$ is also an all-pass filter. Since $\hat{\check{\mathbf{x}}}(e^{j\omega})$ is an FIR filter, this is not possible unless we have [98]

$$[\check{\mathbf{x}}]_n = \lambda\delta(n - n_0) \tag{4.10}$$

for some n_0 satisfying $1 \leq n_0 \leq N-1$ and λ is a constant. However, since $s \geq 3$, $\check{\mathbf{x}}$ has at least two non zero entries which contradicts (4.10). Therefore, the existence of m^* is guaranteed. It is then easy to see that $\mathbf{x}^{\#}$ is equal to $\mathbf{x}$ in sense of $\mathbb{C} \setminus \mathbb{T}$. In particular, assuming $l_{min} = N$, we have $[\mathbf{x}^{\#}]_N = \sqrt{[\tilde{\mathbf{x}}]_{m^*}/\beta} = |x_N|$. Now, for $1 \leq q \leq N-1$, from (4.7), we have $[\tilde{\mathbf{x}}]_{2N-2-q} = x_q \bar{x}_N$. Therefore, $[\mathbf{x}^{\#}]_q = \frac{[\tilde{\mathbf{x}}]_{2N-2-q}}{|x_N|} = cx_q$ where $c = \bar{x}_N/|x_N|$ is the global phase term. Note that this iterative algorithm imposes no upper bound on s. We also have following corollary for non-sparse $\mathbf{x}$.

Corollary 6 *If $\mathbf{x}$ is not sparse (i.e. $s = N$), the number of measurements needed for recovering $\mathbf{x}$ is $M = 4N - 5$.*

4.4 Phase Retrieval with Prior Information on Support

It is straightforward to see that Theorem 15 is not efficient in recovering a sparse $\mathbf{x}$ since the iterative algorithm does not fully exploit the sparsity of $\mathbf{x}$. However, the result of Theorem 15 is still important for the following reasons. Firstly, Theorem 15 does not require any prior knowledge of Supp($\mathbf{x}$) and it is valid for all $s \geq 3$ (unlike the algorithms requiring *collision-free* requirement). Secondly, for non-sparse $\mathbf{x}$, $4N-5$ measurements needed by Theorem 15 is consistent with the universal lower bound $4N-4$ conjectured in literature for recovering arbitrary $\mathbf{x}$. Above all, the structure of (4.7) reveals the possibility for using l_1 minimization on the sub-vector consisting of decoupled elements.

In practice, we may assume some mild prior knowledge about the support of $\mathbf{x}$. For instance, in [93], the authors assume the first entry of $\mathbf{x}$ is nonzero. Similarly, in this section, *we assume to know any one index in the support of* $\mathbf{x}$. Without loss of generality, we may assume that x_N is nonzero (since, by using a permuting matrix for the known entry, we can ensure x_N to be non zero).

4.4.1 A Cancellation Based Approach for Sparse Recovery

It can be seen from (4.7) that the subvector of $\tilde{\mathbf{x}}$ indexed by $1 \leq |m| \leq N-2$ is not necessarily sparse. We propose *a novel variation of the basic PNFS to cancel out this "non-sparse" segment* and produce a difference-vector that is guaranteed to have the same sparsity as that of $\mathbf{x}$. Assuming that x_N is nonzero, we propose to use two sets of measurements, $\tilde{\mathbf{y}}, \tilde{\mathbf{y}}' \in \mathbb{C}^{\tilde{M}}$ as

$$[\tilde{\mathbf{y}}]_i = |<\mathbf{x}, \mathbf{f}_i>|^2 \tag{4.11}$$

$$[\tilde{\mathbf{y}}']_i = |<\mathbf{x}, \mathbf{f}_i^{'}>|^2 \tag{4.12}$$

where $\mathbf{f}_i$ denotes the PNFS vector (as in Def. 12) and $\mathbf{f}_i^{'}$ is defined as

$$\mathbf{f}_i^{'} = \frac{1}{\sqrt[4]{4N-5}} \left[z_i^1, z_i^2, \cdots, z_i^{N-1}, 0 \right] \tag{4.13}$$

where $z_i = e^{j2\pi(i-1)/(4N-5)}$. Denoting $\hat{\mathbf{y}} = \tilde{\mathbf{y}} - \tilde{\mathbf{y}}'$, we have

$$\hat{\mathbf{y}} = \mathbf{Z}\hat{\mathbf{x}} \tag{4.14}$$

where

$$[\hat{\mathbf{x}}]_m = \begin{cases} |x_N|^2 & m = 0 \\ 0 & m = 1, 2, \cdots, N-2 \\ x_{2N-2-m}\bar{x}_N & m = N-1, \cdots, 2N-3 \\ \overline{[\hat{\mathbf{x}}]_{-m}} & m < 0 \end{cases}$$

and $\mathbf{Z} \in \mathbb{C}^{\tilde{M}, 4N-5}$ defined as in (4.9). Notice that $\tilde{\mathbf{x}}$ has sparsity $2s-1$ and support of $\mathbf{x}$ (except the Nth entry) is identical to that of the subvector of $\tilde{\mathbf{x}}$ indexed by $m = N-1, \cdots, 2N-3$. We next discuss how large $\tilde{M}$ should be for perfectly recovering $\hat{\mathbf{x}}$ from $\hat{\mathbf{y}}$.

4.4.2 Number of Measurements

We can recover $\hat{\mathbf{x}}$ by solving the l_1 minimization:

$$\min_{\theta} \|\theta\|_1 \quad \text{subject to } \hat{\mathbf{y}} = \mathbf{Z}\theta \quad (\mathbf{P1})$$

The vector $\mathbf{x}$ can then be recovered from the solution of (**P1**). The following theorem establishes the total number of measurements sufficient to recover $\mathbf{x}$ from the proposed cancellation based approach.

Theorem 16 *Let $\mathbf{x} \in \mathbb{C}^N$ be a sparse vector with s non zero elements and $x_N \neq 0$. Suppose we construct the difference measurement vector $\hat{\mathbf{y}}$ as in (4.14) using $\tilde{M}$ pairs of sampling vectors $\{\mathbf{f}_{i_k}, \mathbf{f}'_{i_k}\}_{k=1}^{\tilde{M}}$ where indices i_k are selected uniformly at random between 1 and $4N-5$. Then $\mathbf{x}$ can be recovered (in sense of $\mathbb{C} \setminus \mathbb{T}$) by solving (P1) if $\tilde{M} = Cs\log N$ for some constant C.*

Proof. Since the indices $\{i_k\}_{k=1}^{\tilde{M}}$ are chosen uniformly at random between 1 and $4N-5$ to construct the measurement vectors $\mathbf{f}'_{i_k}$ and $\mathbf{f}_{i_k}$, the resulting matrix $\mathbf{Z}$ in (4.14) consists of $\tilde{M}$ rows of a DFT matrix (of dimension $4N-5$) which are chosen uniformly at random. Well known results from compressed sensing using random Fourier matrices [22, 84] guarantee that the solution to (**P1**) will be $\hat{\mathbf{x}}$ with high probability, provided the total number of measurements $2\tilde{M}$ satisfies $\tilde{M} = C\hat{s}\log N$ where $\hat{s} = 2s-1$ is the sparsity of $\hat{\mathbf{x}}$. After recovering $\hat{\mathbf{x}}$, $x_N^{\#} = |x_N|$ can be obtained by observing $[\hat{\mathbf{x}}]_0$. The remaining entries can then be estimated as $x_q^{\#} = [\hat{\mathbf{x}}]_{2N-2-q}/|x_N^{\#}|$. The validity of estimation $\mathbf{x}^{\#}$ can be proved in the same way as in proof of Theorem 15.

4.5 Sparse Phase Retrieval Using Randomized PNFS

We introduce a randomized version of the PNFS for sparse phase retrieval as follows:

Definition 13 *(**Randomized PNFS**) A Randomized PNFS (R-PNFS) consists of measurement vectors*

$$\mathbf{f}_i^{\text{(R-PNFS)}} = [\mathbf{I}_{N,N} \quad \mathbf{v}]\, \mathbf{f}_i^{(N+1)}$$

where $\mathbf{v} \in \mathbb{C}^N$ is a random vector with independent entries, and $\mathbf{f}_i^{(N+1)}$ is defined in Def.12 for dimension $N+1$.

Given the unknown signal $\mathbf{x}^\star \in \mathbb{C}^N$, the phaseless measurement obtained using a R-PNFS vector can be expressed as

$$y_i = \left| \left(\mathbf{f}_i^{\text{R-PNFS}} \right)^H \mathbf{x}^\star \right|^2 + n_i = \left| \mathbf{f}_i^{(N+1)H} \begin{bmatrix} \mathbf{x}^\star \\ \mathbf{v}^H \mathbf{x}^\star \end{bmatrix} \right|^2 + n_i \tag{4.15}$$

The basic idea of R-PNFS is to concatenate an extra element $x_{N+1} = \mathbf{v}^H \mathbf{x}^\star$ to form the vector $\mathbf{x} = [\mathbf{x}^{\star T} \quad x_{N+1}]^T$, and then measure $\mathbf{x}$ using PNFS for dimension $N+1$. Since the elements of $\mathbf{v}$ are independent random variables, it follows that the last entry of $\mathbf{x}$ satisfies $x_{N+1} \neq 0$ with probability 1. This enables us to devise an efficient cancellation based algorithm for sparse phase retrieval as follows.

4.5.1 A Cancellation Based Algorithm for R-PNFS

We measure a sparse $\mathbf{x}^\star$ (with s non zero elements) using two sets of PNFS samplers, and perform sparse recovery on the difference between the two measurements. This enables us to "cancel" out certain non-zero terms in the autocorrelation of $\mathbf{x}^\star$ and retain only "decoupled terms" (singletons) which have a maximum sparsity of $2s+1$. We begin by introducing a second sampling vector $\tilde{\mathbf{f}}_i^{(N+1)} \in \mathbb{C}^N$ as

$$\tilde{\mathbf{f}}_i^{(N+1)} = [\mathbf{I}_{N,N} \quad \mathbf{0}]\, \mathbf{f}_i^{(N+1)}$$

This sampler can be thought of as a masked version of the PNFS sampler defined in Def. 12. Following are the main steps of the algorithm:

1. Collect two sets of (noisy) phaseless measurements $\mathbf{y}^{(1)}, \mathbf{y}^{(2)} \in \mathbb{C}^{\tilde{M}}$ as

$$\begin{aligned} \mathbf{y}_i^{(1)} &= \left| \left(\mathbf{f}_i^{\text{R-PNFS}} \right)^H \mathbf{x}^\star \right|^2 + \mathbf{n}_i^{(1)} \\ \mathbf{y}_i^{(2)} &= \left| \left(\tilde{\mathbf{f}}_i^{(N+1)} \right)^H \mathbf{x}^\star \right|^2 + \mathbf{n}_i^{(2)} \end{aligned} \tag{4.16}$$

   We assume the noise is bounded, i.e. $|\mathbf{n}_i^{(k)}| \leq \eta, k = 1, 2$. Notice that we collect a total of $M = 2\tilde{M}$ measurements.

2. Compute the difference measurement $\boldsymbol{\Delta}\mathbf{y} = \mathbf{y}^{(1)} - \mathbf{y}^{(2)}$. The *key step is to notice that*

$$\boldsymbol{\Delta}\mathbf{y} = \mathbf{Z}\hat{\mathbf{x}} + \boldsymbol{\Delta}\mathbf{n} \tag{4.17}$$

   where the unknown vector $\hat{\mathbf{x}} \in \mathbf{C}^{4N-1}$ consists only of "decoupled" quadratic terms (singletons of the form $\bar{x}_{N+1}x_i, i = 1, 2, \cdots, N$) given by

$$[\hat{\mathbf{x}}]_m = \begin{cases} |x_{N+1}|^2 & m = 0 \\ 0 & m = 1, 2, \cdots, N-1 \\ x_{2N-m}\bar{x}_{N+1} & m = N, \cdots, 2N-1 \\ \overline{[\hat{\mathbf{x}}]}_{-m} & m < 0 \end{cases}$$

   Since $x_{N+1} = \mathbf{v}^H\mathbf{x}^\star$ where $\mathbf{v}$ is a random vector with independent entries, it holds that $x_{N+1} \neq 0$ with probability 1. Hence $\hat{\mathbf{x}}$ has exactly $2s+1$ non zero elements. We also have $\boldsymbol{\Delta}\mathbf{n} = \mathbf{n}^{(1)} - \mathbf{n}^{(2)}$, and the matrix $\mathbf{Z} \in \mathbb{C}^{\tilde{M}, 4N-1}$ is a partial DFT matrix with $[\mathbf{Z}]_{i,k} = \frac{1}{\sqrt{4N-1}} e^{j2\pi \frac{n_i k}{4N-1}}$.

3. Obtain an estimate of $\hat{\mathbf{x}}$ as the solution to the following l_1-minimization problem:

$$\min_{\theta} \|\theta\|_1 \quad \text{subject to } \|\mathbf{\Delta y} - \mathbf{Z}\theta\|_2 \leq \eta\sqrt{\tilde{M}} \quad (\mathbf{P1})$$

4. Given the solution $\hat{\mathbf{x}}^{\#}$ to (**P1**), the estimate for each entry of $\mathbf{x}^{\star}$ is given by $x_q^{\#} = [\hat{\mathbf{x}}^{\#}]_{2N-q}/|\sqrt{[\hat{\mathbf{x}}^{\#}]_0}|$ for $1 \leq q \leq N$ and $x_{N+1}^{\#} = |\sqrt{[\hat{\mathbf{x}}^{\#}]_0}|$.

4.5.2 Stability of Noisy Phase Retrieval with R-PNFS

To analyze the performance of the proposed algorithm, we use the following lemma from [96] which is tailored for the form (**P1**):

Lemma 13 *[96] Consider a sparse $\hat{\mathbf{x}} \in \mathbb{C}^{4N-1}$ with $2s+1$ non zero elements and $\mathbf{Z} \in \mathbb{C}^{\tilde{M},4N-1}$ be the DFT matrix with $\tilde{M}$ rows whose indices are chosen uniformly at random from $[0, 4N-2]$. If $\tilde{M} \geq c_0(2s+1)\log(4N-1)\log(\varepsilon^{-1})$, then with probability at least $1-\varepsilon$, the solution $\hat{\mathbf{x}}^{\#}$ of (**P1**) satisfies*

$$\|\hat{\mathbf{x}} - \hat{\mathbf{x}}^{\#}\|_2 \leq c_1\sqrt{2s+1}\eta \tag{4.18}$$

where c_0, c_1 are universal constants.

Theorem 17 *Given a sparse $\mathbf{x}^{\star} \in \mathbb{C}^N$ (with s non zeros), and the measurement vector $\mathbf{v} \in \mathbb{C}^N$, consider the measurement model (4.16) where the indices n_i of $\mathbf{f}_i^{(N+1)}, i = 1, 2\cdots, M$ are chosen uniformly at random from $[0, 4N-2]$. If $\tilde{M} \geq c_0(2s+1)\log(4N-1)\log(\varepsilon^{-1})$ and*

$|x_{N+1}|^2 > c_1\sqrt{2s+1}\eta$, *with probability at least* $1-\varepsilon$, *the estimates* $x_q^{\#}$ *of* $x_q^\star, 1 \le q \le N$, *satisfy*

$$\sum_{q=1}^{N} |x_q^\star - e^{j\phi_0} x_q^{\#}| \le \frac{c_1\sqrt{(2s+1)(4N-1)}}{\sqrt{|x_{N+1}|^2 - c_1\sqrt{2s+1}\eta}}\eta + \|\mathbf{x}^\star\|_1 \left(\frac{1}{\sqrt{1 - c_1 \frac{\sqrt{2s+1}\eta}{|x_{N+1}|^2}}} - 1 \right) \tag{4.19}$$

where $x_{N+1} = \mathbf{v}^H\mathbf{x}^\star$, $\phi_0 = arg_{\phi\in[0,2\pi)} x_{N+1}/|x_{N+1}|$, *and* c_0, c_1 *are universal constants given in Lemma 13.*

Proof. According to the proposed algorithm, the estimate for each entry of $\mathbf{x}^\star$ is given by $x_q^{\#} = [\hat{\mathbf{x}}^{\#}]_{2N-q}/x_{N+1}^{\#}$ for $1 \le q \le N$. Then, we have

$$\begin{aligned} |x_q^\star - e^{j\phi_0} x_q^{\#}| &= |\frac{x_{N+1}}{|x_{N+1}|}\left(\frac{[\hat{\mathbf{x}}]_{2N-q}}{|x_{N+1}|} - \frac{[\hat{\mathbf{x}}^{\#}]_{2N-q}}{|x_{N+1}^{\#}|} \right)| \\ &\le \beta \frac{\epsilon_{2N-q}}{|x_{N+1}|} + |1-\beta||x_q^\star| \end{aligned} \tag{4.20}$$

where $\epsilon_{2N-q} \triangleq |[\hat{\mathbf{x}}]_{2N-q} - [\hat{\mathbf{x}}^{\#}]_{2N-q}|$ and $\beta = \frac{|x_{N+1}|}{|x_{N+1}^{\#}|}$. It follows that

$$\begin{aligned} \sum_{q=1}^{N} |x_q^\star - e^{j\phi_0} x_q^{\#}| &\le \beta \frac{\sum_{q=1}^{N} \epsilon_{2N-q}}{|x_{N+1}|} + |1-\beta| \sum_{q=1}^{N} |x_q^\star| \\ &\le \beta \frac{\|\hat{\mathbf{x}} - \hat{\mathbf{x}}^{\#}\|_1}{|x_{N+1}|} + |1-\beta|\|\mathbf{x}^\star\|_1 \end{aligned} \tag{4.21}$$

Since $\|\hat{\mathbf{x}} - \hat{\mathbf{x}}^{\#}\|_1 \le \sqrt{4N-1}\|\hat{\mathbf{x}} - \hat{\mathbf{x}}^{\#}\|_2$, Lemma 13 gives us

$$\begin{aligned} &\sum_{q=1}^{N} |x_q^\star - e^{j\phi_0} x_q^{\#}| \\ &\le c_1\beta \frac{\sqrt{(2s+1)(4N-1)}\eta}{|x_{N+1}|} + |1-\beta|\|\mathbf{x}^\star\|_1 \end{aligned} \tag{4.22}$$

Since $[\hat{\mathbf{x}}]_0 = |x_{N+1}|^2$, it follows from Lemma 1 that $|1 - \frac{1}{\beta^2}| \leq c_1 \frac{\sqrt{2s+1}\eta}{|x_{N+1}|^2}$. If $|x_{N+1}|^2 > c_1\sqrt{2s+1}\eta$, we have

$$1 - c_1 \frac{\sqrt{2s+1}\eta}{|x_{N+1}|^2} \leq \frac{1}{\beta^2} \leq 1 + c_1 \frac{\sqrt{2s+1}\eta}{|x_{N+1}|^2} \tag{4.23}$$

The proof completes by plugging (4.23) in (4.22).

Remark 1. In absence of noise, setting $\eta = 0$ in (4.19) implies exact recovery of $\mathbf{x}^\star$ with a global phase ambiguity ϕ_0 which is explicitly given. This is achieved using a total of $M = 2\tilde{M}$ measurements, where $\tilde{M} = O(s \log N)$. Hence, our algorithm recovers $\mathbf{x}^\star$ with an order-wise minimal (up to a logarithmic factor) number of measurements.

Remark 2. Unlike "lifting" based approaches [77, 90], our method is based on l_1-minimization with only $O(N)$ variables. This implies significant computational saving and allows faster implementation.

4.6 Numerical Results

4.6.1 Phase Retrieval with Prior Knowledge

In Fig. 4.1, we simulate the probability of "no-collision" as a function of sparsity s with fixed ambient data dimension $N = 10000$. For each run and sparsity s, we randomly choose the support and test whether it satisfies the collision-free property. The probability of collision-free is computed by averaging over 2000 such random runs. It can be seen that for s much smaller than the ideal upper bound $\chi_N \approx 142$, the probability of "no-collision" goes to zero. As a consequence, the ACF based method has undesirable restriction on the allowable sparsity. To validate the theoretical claims of this paper, we now focus on Theorem 16 because the correctness of Theorem 15 can be easily verified in a deterministic way. We choose $N = 150$ and vary

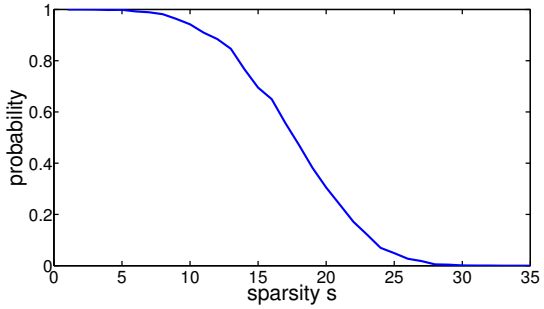


Figure 4.1: The probability of "no-collision" as a function of sparsity s. The ambient dimension is $N = 10000$ and the result is averaged over 2000 runs.

the sparsity s. Nonzero part of $\mathbf{x}$ is generated with complex Gaussian distribution. We assume $x_N \neq 0$ and obtain the estimate $\mathbf{x}^{\#}$ of $\mathbf{x}$ by constructing the difference vector $\hat{\mathbf{y}}$ and solving (**P1**) with CVX tool for MATLAB. The global phase ambiguity is $\rho = x_N/x_N^{\#}$. Using ρ we can compute the entry-wise estimation error as $|x_i - \rho x_i^{\#}|$ for $1 \leq i \leq N$. In Fig. 4.2, we show the phase transition plot of Theorem 16. We declare the recovery is successful if $|x_i - \rho x_i^{\#}| \leq 10^{-6}$ for all entries and failed otherwise. For each pair (M, s), we generate random s-sparse complex signal $\mathbf{x}$ and compute the probability of success over 100 runs. As a reference, we also draw the line representing $3s\log N$ where the choice of 3 is for ease of comparison. We find that the simulation matches the claim given in the theorem regarding the sufficient measurement size being $O(s\log N)$.

4.6.2 Phase Retrieval with R-PNFS

We consider a complex valued signal $\mathbf{x}^{\star} \in \mathbb{C}^N$ with s non zero elements, and $\|\mathbf{x}^{\star}\|_2 = 1$. Both the nonzero indices and amplitudes are generated randomly.

The phase transition plots of the proposed method for both noiseless and noisy signal models is depicted in Fig.4.3. Here $N = 100$. In the noiseless setting, for each M and s, we declare success if $\max_q |x_q^{\star} - e^{j\phi_0} x_q^{\#}| < 10^{-6}$. For the noisy model, we assume the entry wise noise to be upper bounded by $\epsilon = 0.01$ and plot the reconstruction error $\frac{1}{N}\sum_{q=1}^{N} |x_q^{\star} - e^{j\phi_0} x_q^{\#}|$. We also superpose the line corresponding to $M = 3s \log N$ to demonstrate that the proposed

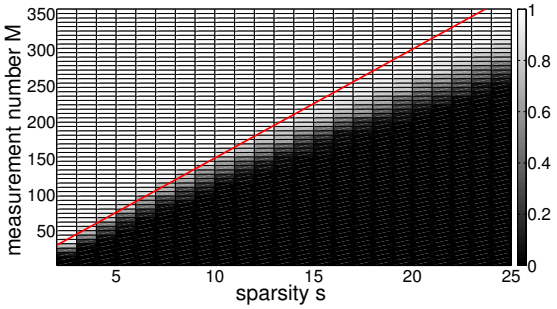


Figure 4.2: The phase transition plot for Theorem 16. $M = 2\tilde{M}$ is the total number of measurements needed and $N = 150$. The red line represents $3s\log N$. The color bar denotes probability of success from 0 to 1. The white cells denote successful recoveries (i.e. $|x_i - \rho x_i^{\#}| \leq 10^{-6}$ for all entries) and black cells denote failures.The results are averaged over 100 runs.

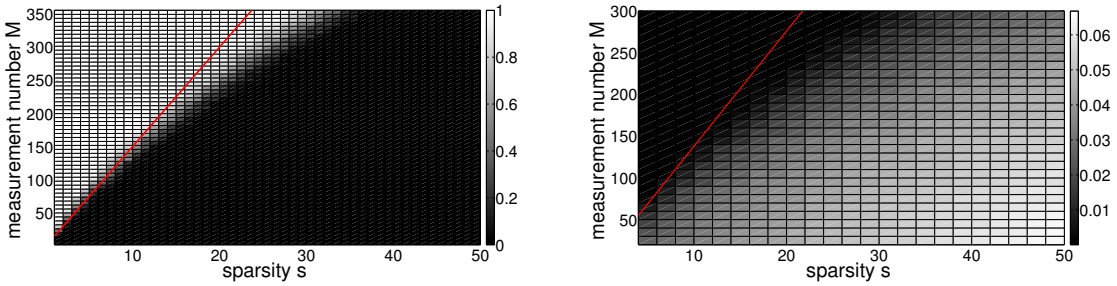


Figure 4.3: (Left) Phase transition for noiseless case, averaged over 100 runs with $N = 150$. White and black boxes denote success rates of 1 and 0 respectively. (Right) Phase transition for noisy case averaged over 50 runs with $N = 100$, and entry-wise noise bounded by 0.01. Each box denotes $\frac{1}{N}\sum_{q=1}^{N}|x_q^{\star} - e^{j\phi_0}x_q^{\#}|$. The red line represents $M = 3s\log N$ for both.

approach recovers the true $\mathbf{x}^\star$ with $M = O(s \log N)$ measurements.

In Fig. 4.4, we show an example of sparse phase retrieval using the proposed R-PNFS sampler and cancellation based algorithm. Here $N = 350, s = 6, M = 100$. It can be seen that the proposed technique recovers the true $\mathbf{x}^\star$ faithfully up to a global phase ambiguity, the value of which is easily obtained from the complex plane representation.

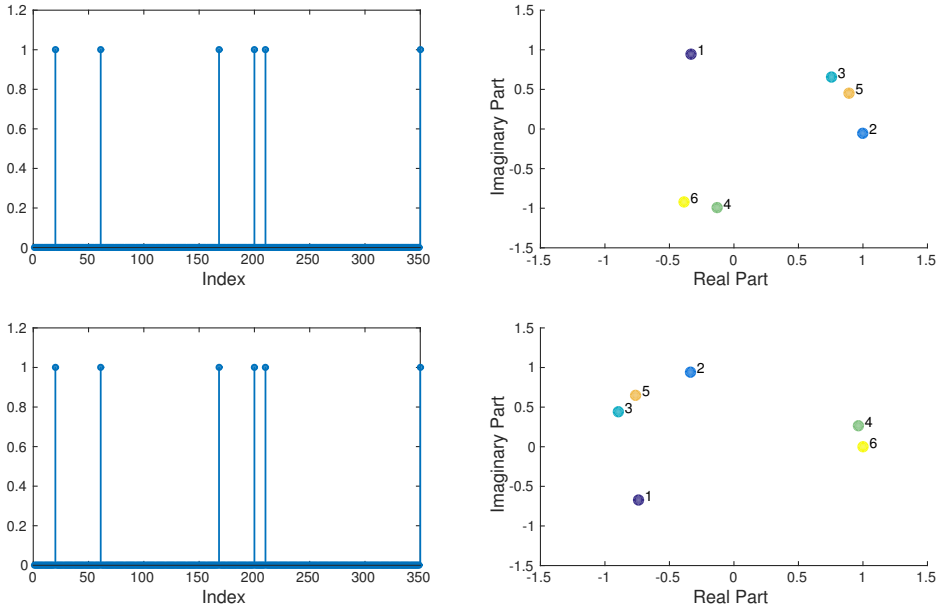


Figure 4.4: (Top left) Amplitudes of the original data. (Top right) The complex plane representation of the nonzero part of the original data.(Bottom left) Amplitudes of the recovered data. (Bottom right) The complex plane representation of the recovered data. Here, N=350, s=6 and M = 100.

4.7 Conclusion

We consider the sparse phase retrieval problem using a newly introduced class of Partial Nested Fourier Sampler (PNFS). The PNFS avoids the need for the so-called "collision-free" condition required by most existing work in sparse phase retrieval and proposes a simple and effective algorithm for reconstructing the sparse signal using only $O(s \log N)$ measurements. A

key contribution of this chapter is to exploit and extend the idea of nested sampling to design general Fourier measurements for phase retrieval which is simple and applicable for most cases. If no prior knowledge is available, we proposed a new structured sampling scheme, namely the Randomized Partial Nested Fourier Sampler (R-PNFS), along with a novel cancellation based algorithm which can provably recover sparse complex valued signals from their amplitude measurements. The proposed technique requires only $M = O(s \log N)$ measurements which is near-optimal compared to the underlying degree of freedom of the sparse signal. We also showed that under mild conditions, the approach is stable to bounded measurement noise.

Chapter 4, in part, is a reprint of the material as it appears in the papers:

- H. Qiao and P. Pal, " Sparse Phase Retrieval with Near Minimal Measurements: A Structured Sampling Based Approach", in the *Proceedings of 41st IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, 2016.

- H. Qiao and P. Pal, "Sparse Phase Retrieval Using Partial Nested Fourier Samplers, in the *Proceedings of IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Orlando, FL, USA, 2015.

The dissertation author was the primary investigator and author of these papers.

Chapter 5

Coarray Interpolation, Sparse Bilinear Problem with Non-Negative Constraint and Nested Array based Kriging

5.1 Introduction

In this chapter, we will discuss three separate but implicitly related problems. Array interpolation is of great importance in many applications especially in the case of missing data and failing sensor. Previous study on array interpolation is mainly in physical domain and an optimal linear mapping is constructed to interpolate the physical sensors. However, coarray interpolation is in the correlation domain that the second-order statistic of the signal is explicitly exploited. Moreover, coarray interpolation is by nature a non-linear mapping of the physical measurements. The key constraint for stable coarray interpolation is the positive semidefinite of the corresponding Toeplitz matrix.

We will next study the problem of sparse bilinear problem with non-negative constraint. For the first time in literature, we show that it is possible to estimate the true solutions up to a

global scaling factor when there are more unknowns than the number of equations. We exactly characterize the set of ambiguous solutions and state the necessary and sufficient conditions for exact recovery.

in the end, we will consider the application of nested array idea to an important geostatistical problem, Kriging. By assuming the field of interest is stationary, we relate the compressive Kriging to Toeplitz covariance matrix estimation using nested array. We propose a robustness analysis based on total least squares.

The common underlying characteristic of these problems is the non-negative constraint. By studying these three different problems, we again demonstrate the fact that non-negative constraints are crucial for stable recoveries in variant inverse problems.

5.2 Unified Analysis of Co-Array Interpolation with Application in Direction-of-Arrival Estimation

This section considers the problem of co-array interpolation for direction-of-arrival (DOA) estimation with sparse non-uniform arrays. By utilizing the much longer difference co-array associated with these arrays, it is possible to perform DOA estimation of more sources than sensors. Since the co-array may contain holes (or missing lags), interpolation algorithms have been proposed to fully utilize the remaining elements of the co-array beyond that captured in the contiguous ULA segment. However, the quality and stability of interpolation performed by such algorithms (especially in presence of modeling errors) have not been analyzed. This paper provides a unified analysis of co-array interpolation algorithms to bound the interpolation error in terms of modeling errors. The results are universal in the sense that they can be applied to analyze any algorithm that utilizes the positive semidefinite (PSD) structure of the interpolated covariance matrix. The general framework is then applied to analyze specific algorithms and simulations are conducted to study their interpolation errors.

5.2.1 Introduction

Direction-of-arrival (DOA) estimation of energy-emitting sources is a central problem arising in diverse applications such as radar, sonar, medical imaging and communications [49]. Sparse non-uniform arrays such as nested, coprime and minimum redundancy arrays are known to offer distinct advantages over traditionally used Uniform Linear Arrays (ULA) owing to their ability to resolve more sources than sensors [17, 45, 204]. The basic idea is to create a longer virtual difference co-array [17] by judicious array design, whose degrees-of-freedom (DOF) can be exploited by well-designed algorithms such as Co-array MUSIC [19, 88].

For many non uniform arrays (such as coprime arrays), the difference co-array is not continuous and has holes or missing lags. Since co-array MUSIC algorithms are capable of only exploiting the DOF of a continuous ULA segment of the co-array, several array interpolation techniques such as positive definite Toeplitz completion [225], co-array interpolation/extrapolation [226, 227], and nuclear norm minimization [228] have been proposed to interpolate the correlation values at the missing lags and use the interpolated co-array for DOA estimation.

In this section, we propose a unified framework for analyzing co-array interpolation algorithms by developing an explicit upper bound on the interpolation error, in terms of the measurement error. Our analysis framework is very generic and can be applied to *any algorithm that utilizes the positive semidefinite (PSD) structure* of the interpolated covariance matrix. As special cases, we use this general framework to develop algorithm-specific error bounds for the algorithms in [225, 228]. Our results establish stability of these interpolation algorithms with respect to modeling errors (such as that due to finite snapshot averaging) and demonstrate that perfect interpolation is possible as the error decays to zero.

Related Work. While the performance of traditional array interpolation techniques have been analyzed in the past in terms of bias and mean squared error [177, 178], these methods are primarily based on interpolating the *physical array* using linear transforms, and cannot be used for *co-array interpolation* since the co-array is a non-linear function (Kronecked product) of the physical array.

On the other hand, interpolation algorithms that directly work in the co-array domain [225, 228], have not been analyzed. In this paper, we bridge this gap by providing a unified analysis of co-array interpolation algorithms. Our analysis is based on recently developed tools from super resolution theory and positive semidefinite Toeplitz covariance compression [4, 6, 59, 141, 142].

5.2.2 Co-array based signal model and need for interpolation

Consider D narrowband statistically uncorrelated sources impinging on a linear sensor array from directions $\bar{\theta}_i, 1 \leq i \leq D$. The array contains K sensors with the kth sensor located at $z_k d$, where z_k is an integer and $d = \lambda/2$ (λ being the carrier wavelength of the narrowband sources). The signals received at the K sensors are given by

$$\mathbf{x} = \sum_{i=1}^{D} c_i \mathbf{a}_{\mathbb{S}}(\theta_i) + \mathbf{n}_{\mathbb{S}} \tag{5.1}$$

where c_i denotes the amplitude of each source (assumed to be zero-mean random variables) and $\mathbf{a}_{\mathbb{S}}(\theta_i) \in \mathbb{C}^K$ represents the steering vector corresponding to the normalized DOA $\theta_i \in \mathbb{T} = [-1/2, 1/2]$, which is given by $\theta_i = (d/\lambda)\sin\bar{\theta}_i$ ($\bar{\theta}_i$ being the DOA satisfying $\bar{\theta}_i \in [-\pi/2, \pi/2]$). The steering vector satisfies $[\mathbf{a}_{\mathbb{S}}(\theta_i)]_k = [e^{j2\pi z_k \theta_i}]$. Here, $\mathbf{n}_{\mathbb{S}}$ represents zero-mean additive noise at K sensors, statistically uncorrelated with the source amplitudes c_i. The statistical assumptions on source signal and noise are summarized as [49, 228]

$$\mathbb{E}[c_i^* c_j] = \sigma_i^2 \delta_{i,j}, \mathbb{E}[c_i^* \mathbf{n}_{\mathbb{S}}] = \mathbf{0}, \mathbb{E}[\mathbf{n}_{\mathbb{S}} \mathbf{n}_{\mathbb{S}}^H] = \sigma^2 \mathbf{I}$$

Under the above assumptions, the correlation matrix $\mathbf{R}_{\mathbb{S}} \in \mathbb{C}^{K \times K}$ of the received signals is given by

$$\mathbf{R}_{\mathbb{S}} = \sum_{i=1}^{D} \sigma_i^2 \mathbf{a}_{\mathbb{S}}(\theta_i) \mathbf{a}_{\mathbb{S}}^H(\theta_i) + \sigma^2 \mathbf{I} \tag{5.2}$$

Denoting $\mathbb{S} = \{z_k, 1 \leq k \leq K\}$ as the set of sensor positions (normalized with respect to d), its difference co-array is defined as [45]

$$\mathbb{D} \triangleq \{z_k - z_j | z_k, z_j \in \mathbb{S}\}$$

Let us associate a vector $\mathbf{a}_{\mathbb{D}}$ with this difference set, such that $[\mathbf{a}_{\mathbb{D}}(\theta_i)]_m = [e^{j2\pi n_m \theta_i}], n_m \in \mathbb{D}$. The vectorized version of (5.2), after removal of repeated rows, is given by

$$\mathbf{r}_{\mathbb{D}} = \sum_{i=1}^{D} \sigma_i^2 \mathbf{a}_{\mathbb{D}}(\theta_i) + \sigma^2 \mathbf{e}_0 \tag{5.3}$$

where $\mathbf{e}_0$ has zero entries everywhere except at the location corresponding to the lag 0 [88, 228]. Due to similarities between (5.3) and (5.1), we can treat (5.3) as the signal received at *a virtual sensor array with sensors positions given by* $\mathbb{D}$.

Depending on the geometry of $\mathbb{S}$, the difference co-array $\mathbb{D}$ may be continuous (i.e. it can itself be a uniform linear array or a ULA, consisting of only consecutive integers), or it may contain holes. The former is known as a fully augmentable array and the latter is called a partially augmentable array [225]. Following [228], we associate the following two uniform linear arrays $\mathbb{U}$ and $\mathbb{V}$ with $\mathbb{D}$ as follows:

Definition 14 *[228] Let $\mathbb{U}$ be the maximum ULA contained in $\mathbb{D}$ such that*

$$\mathbb{U} = \{m | \{-|m|, \cdots, -1, 0, 1, \cdots, |m|\} \subseteq \mathbb{D}\}$$

and $\mathbb{V}$ be the smallest ULA containing $\mathbb{D}$ such that $\mathbb{V} = \{m | \min(\mathbb{D}) \leq m \leq \max(\mathbb{D})\}$.

As an example, let $\mathbb{S} = \{0, 1, 2, 6\}$, then $\mathbb{D} = \{-6, -5, -4, -2, -1, 0, 1, 2, 4, 5, 6\}$, $\mathbb{U} = \{-2, -1, 0, 1, 2\}$ and $\mathbb{V} = \{-6, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6\}$. For fully augmentable arrays, we have $\mathbb{U} = \mathbb{D} = \mathbb{V}$. Examples include ULA, nested array [45] and minimum redundancy

array [204]. On the other hand, for partially augmentable arrays, we have $\mathbb{U} \subset \mathbb{D} \subset \mathbb{V}$. Coprime array [17] is an example of partially augmentable array.

5.2.3 Co-Array MUSIC for Partially Augmentable Arrays

Co-array based DOA estimation algorithms (such as co-array MUSIC [19]) can utilize the degrees of freedom (given by the cardinality) in the virtual ULA segment $\mathbb{U}$ contained in $\mathbb{D}$, and for well-designed arrays, it is possible to resolve more sources than sensors. For nested and coprime arrays with K sensors, $|\mathbb{U}| = O(K^2)$ and hence it is possible to resolve $O(K^2)$ sources using only K sensors.

For partially augmentable arrays, the virtual ULA $\mathbb{V}$ is strictly larger than $\mathbb{U}$. However, co-array MUSIC [19] cannot directly utilize the DOF in $\mathbb{V}$ since certain entries of $\mathbb{V}$ do not appear in $\mathbb{D}$. To address this issue, *a preprocessing step based on interpolation* has been suggested [225,228]. Similar to $\mathbf{r}_{\mathbb{D}}$, let $\mathbf{r}_{\mathbb{U}}$ be the sub-vector of $\mathbf{r}_{\mathbb{D}}$, containing the correlation values evaluated at lags given by $\mathbb{U}$, and $\mathbf{r}_{\mathbb{V}}$ be a vector that consists of correlation values at lags given by the set $\mathbb{V}$. Co-array MUSIC can be applied on $\mathbf{r}_{\mathbb{V}}$ to fully exploit the DOF of partially augmentable arrays (provided $\mathbf{r}_{\mathbb{V}}$ can be estimated using $\mathbf{r}_{\mathbb{D}}$).

5.2.4 Interpolation Algorithms

Let $\tilde{\mathbf{r}}_{\mathbb{D}}$ and $\tilde{\mathbf{r}}_{\mathbb{U}}$ be the corresponding estimates of $\mathbf{r}_{\mathbb{D}}$ and $\mathbf{r}_{\mathbb{U}}$ computed using finite number of snapshots. In particular, $\tilde{\mathbf{r}}_{\mathbb{U}}$ is a subvector of $\tilde{\mathbf{r}}_{\mathbb{D}}$ and

$$\tilde{\mathbf{r}}_{\mathbb{D}} = \mathbf{r}_{\mathbb{D}} + \mathbf{w}_{\mathbb{D}} \tag{5.4}$$

where $\mathbf{w}_{\mathbb{D}}$ captures the finite-snapshot estimation error. We now briefly describe two algorithms, one based on maximum entropy method [225], and the other based on nuclear norm minimization [228] that aim to estimate $\mathbf{r}_{\mathbb{V}}$ from $\tilde{\mathbf{r}}_{\mathbb{D}}$. For convenience, we denote $\mathbb{U}^+, \mathbb{V}^+, \mathbb{D}^+$ as the non-

negative subsets of $\mathbb{U}, \mathbb{V}, \mathbb{D}$ respectively, and let $\mathcal{T}(\mathbf{v})$ be the Hermitian symmetric Toeplitz matrix with $\mathbf{v}$ as the first column.

(a) Maximum Entropy Method:

In [225], the authors used maximum entropy (ME) as a criterion for extrapolation of correlation at lags in $\mathbb{V}$ outside the range of $\mathbb{D}$. The algorithm consists of two steps. Firstly, given $\tilde{\mathbf{r}}_{\mathbb{U}^+}$, it aims to find the closest positive semidefinite (PSD) Toeplitz matrix $\mathcal{T}(\mathbf{r}^{ME}_{\mathbb{U}^+})$ fitting the data as follows:

$$\begin{aligned} &\mathbf{r}^{ME}_{\mathbb{U}^+} = \arg\min_{\mathbf{x}_{\mathbb{U}^+}} \|\mathbf{x}_{\mathbb{U}^+} - \tilde{\mathbf{r}}_{\mathbb{U}^+}\|_2 \quad (\textbf{ME-1}) \\ &s.t. \quad \mathcal{T}(\mathbf{x}_{\mathbb{U}^+}) \succeq 0 \end{aligned} \tag{5.5}$$

In the next step, the vector of autocorrelation values $\mathbf{r}^{ME}_{\mathbb{V}^+}$ (extrapolated at lags in $\mathbb{V}^+$), is computed as

$$\begin{aligned} &\mathbf{r}^{ME}_{\mathbb{V}^+} = \arg\max_{\mathbf{x}_{\mathbb{V}^+}} \det(\mathcal{T}(\mathbf{x}_{\mathbb{V}^+})) \quad (\textbf{ME-2}) \\ &s.t. \quad [\mathcal{T}(\mathbf{x}_{\mathbb{V}^+})]_{n,1} = [\mathbf{r}^{ME}_{\mathbb{U}^+}]_n, n \in \mathbb{U}^+ \\ &\qquad \|[\mathcal{T}(\mathbf{x}_{\mathbb{V}^+})]_{\mathbb{D}^+\setminus\mathbb{U}^+,1} - \tilde{\mathbf{r}}_{\mathbb{D}^+\setminus\mathbb{U}^+}\|_2 \leq \epsilon_1, \\ &\qquad \mathcal{T}(\mathbf{x}_{\mathbb{V}^+}) \succeq 0 \end{aligned} \tag{5.6}$$

Here ϵ_1 is a parameter that can be tuned to ensure non-empty feasible set. In particular it can be made equal to $\|\mathbf{w}_{\mathbb{D}^+\setminus\mathbb{U}^+}\|_2$. Co-array MUSIC can be finally applied on $\mathcal{T}(\mathbf{r}^{ME}_{\mathbb{V}^+})$ to perform DOA estimation using the DOF of $\mathbb{V}$. Notice that the ME method utilizes PSD constraint in both steps so that the Toeplitz matrix constructed using the extrapolated values continues to be PSD.

(b) Nuclear Norm Minimization:

In [228], the authors assumed that the desired covariance matrix $\mathcal{T}(\mathbf{r}_{\mathbb{V}^+})$ exhibits low rank and proposed to minimize its nuclear norm (as a surrogate for rank) to perform interpolation. In its original form, the algorithm does not impose any PSD constraint on the solution. Since

our analysis framework will explicitly use PSD constraint, we consider the following modified problem instead which uses PSD constraint:

$$\begin{aligned}
\mathbf{r}_{\mathbb{V}^+}^{NUC} &= \arg\min_{\hat{\mathbf{x}}_{\mathbb{V}}} \|\mathcal{T}(\hat{\mathbf{x}}_{\mathbb{V}^+})\|_* \quad \textbf{(NUC-PSD)} \\
s.t. \quad & \|[\mathcal{T}(\hat{\mathbf{x}}_{\mathbb{V}^+})]_{\mathbb{U}^+,1} - \tilde{\mathbf{r}}_{\mathbb{U}^+}\|_2 \leq \epsilon \\
& \|[\mathcal{T}(\hat{\mathbf{x}}_{\mathbb{V}^+})]_{\mathbb{D}^+\setminus\mathbb{U}^+,1} - \tilde{\mathbf{r}}_{\mathbb{D}^+\setminus\mathbb{U}^+}\|_2 \leq \tilde{\epsilon} \\
& \mathcal{T}(\hat{\mathbf{x}}_{\mathbb{V}^+}) \succeq 0
\end{aligned}$$

where $\epsilon, \tilde{\epsilon}$ are parameters (dependent on estimation error) to ensure that (**NUC-PSD**) feasible (i.e. so that the true solution is contained in the feasible set).

5.2.5 A Unified Analysis of Extrapolation Error

For notational simplicity, let r_n denote the nth entry of $\mathbf{r}_{\mathbb{V}^+}$. Then, using the representation (5.3), the desired value of r_n is given by

$$r_n = \sum_{i=1}^{D} \sigma_i^2 e^{j2\pi n\theta_i} + \sigma^2\delta(n) \quad n \in \mathbb{V}^+$$

Let $\mathbf{r}_{\mathbb{V}^+}^{\#}$ denote *any estimate* of $\mathbf{r}_{\mathbb{V}^+}$ such that

$$\mathcal{T}(\mathbf{r}_{\mathbb{V}^+}^{\#}) \succeq \mathbf{0} \tag{5.7}$$

Notice that this automatically implies $\mathcal{T}(\mathbf{r}_{\mathbb{U}^+}^{\#}) \succeq \mathbf{0}$. We now present a fundamental result, upper bounding the extrapolation error $|r_n^{\#} - r_n|$ (for any missing or unobserved lag n outside the range of $\mathbb{D}$) in terms of the estimation error in the correlation values for the lags in $\mathbb{D}$. The proof follows from a closely related lemma in [4,6]. The stability analysis requires a separation condition on the true directions and we define $\rho(\cdot,\cdot)$ as the distance function in wraparound manner over $\mathbb{T}$ [141].

Theorem 18 *Let* $\mathbf{r}_{\mathbb{V}^+}^{\#}$ *denote* any estimate *of* $\mathbf{r}_{\mathbb{V}^+}$ *such that (5.7) holds. If the true DOAs* $\{\theta_l\}_{l=1}^{D}$ *in (5.2) satisfy*

$$\min_{p \neq q} \rho(\theta_p, \theta_q) > 4/|\mathbb{U}^+| \tag{5.8}$$

and $|\mathbb{U}^+| > 256$*, then there exist positive constants* $\bar{c}_1, \bar{c}_2, \bar{c}_3, \bar{c}_4$ *such that for* $|\mathbb{U}^+| \leq n < |\mathbb{V}^+|$

$$\begin{aligned} &|r_n - r_n^{\#}| \\ &\leq \left(\bar{c}_1 + \frac{\bar{c}_2 \pi n}{|\mathbb{U}^+|} + \frac{\bar{c}_3 \pi^2 n^2}{|\mathbb{U}^+|^2}\right)\left(\frac{\bar{c}_4 D \xi}{|\mathbb{U}^+|} + [\mathbf{r}_{\mathbb{U}^+}^{\#}]_0 - [\mathbf{r}_{\mathbb{U}^+}]_0\right) \\ &\leq \left(\bar{c}_1 + \frac{\bar{c}_2 \pi n}{|\mathbb{U}^+|} + \frac{\bar{c}_3 \pi^2 n^2}{|\mathbb{U}^+|^2}\right)\left(\frac{\bar{c}_4 D \xi}{|\mathbb{U}^+|} + \|\mathbf{r}_{\mathbb{U}^+}^{\#} - \mathbf{r}_{\mathbb{U}^+}\|_2\right) \end{aligned} \tag{5.9}$$

where $\xi \triangleq \sup_{\theta \in \mathbb{T}} |\langle \mathbf{a}_{\mathbb{U}^+}(\theta), \mathbf{r}_{\mathbb{U}^+}^{\#} - \mathbf{r}_{\mathbb{U}^+} \rangle|$

Remark 1. Notice that the above bound on extrapolation error holds *irrespective of any specific algorithm used* as long as the algorithm enforces the PSD constraint (5.7).

Remark 2. Theorem 26 indicates that the upper bound on extrapolation error bound (on missing lags) is controlled by estimation error of the correlation supported on the observed set $\mathbb{U}^+$. Depending on the algorithm used, the extrapolation error can be magnified by a factor of $O(n^2/|\mathbb{U}^+|^2)$ (with respect to the estimation error on the observed entries in $\mathbb{U}$). A similar quadratic scaling between high frequency reconstruction error and low frequency observation error has also been reported in [61, 141, 142] for super-resolution imaging using TV-norm based reconstruction.

5.2.6 Analysis of Specific Extrapolation Algorithms

We now apply the result from Theorem 26 to perform a unified analysis of the extrapolation algorithms presented earlier.

Analysis of Maximum Entropy Method

In this case, $\mathbf{r}_{\mathbb{V}^+}^{\#} = \mathbf{r}_{\mathbb{V}^+}^{ME}$ and (ME-2) ensures that $\mathcal{T}(\mathbf{r}_{\mathbb{V}}^{ME}) \succeq \mathbf{0}$. Hence Theorem 26 applies, and we have following result:

Theorem 19 *If the true DOAs $\{\theta_l\}_{l=1}^{D}$ satisfy*

$$\min_{p \neq q} \rho(\theta_p, \theta_q) > 4/|\mathbb{U}^+|$$

*and $|\mathbb{U}^+| > 256$, then there exist positive constants $\bar{c}_1, \bar{c}_2, \bar{c}_3, \bar{c}_4$ such that for $|\mathbb{U}^+| \leq n < |\mathbb{V}^+|$ and $n \notin \mathbb{D}^+$, the solution $\mathbf{r}_{\mathbb{V}^+}^{ME}$ to (***ME-2***) satisfies*

$$\begin{aligned} & |r_n - r_n^{ME}| \\ & \leq \left(\bar{c}_1 + \frac{\bar{c}_2 \pi n}{|\mathbb{U}^+|} + \frac{\bar{c}_3 \pi^2 n^2}{|\mathbb{U}^+|^2}\right) \left(\frac{\bar{c}_4 D}{\sqrt{|\mathbb{U}^+|}} + 1\right) \|\mathbf{r}_{\mathbb{U}^+}^{\#} - \mathbf{r}_{\mathbb{U}^+}\|_2 \\ & \leq 2\left(\bar{c}_1 + \frac{\bar{c}_2 \pi n}{|\mathbb{U}^+|} + \frac{\bar{c}_3 \pi^2 n^2}{|\mathbb{U}^+|^2}\right) \left(\frac{\bar{c}_4 D}{\sqrt{|\mathbb{U}^+|}} + 1\right) \|\mathbf{w}_{\mathbb{U}^+}\|_2 \end{aligned} \tag{5.10}$$

where $\mathbf{w}_{\mathbb{U}^+}$ denotes the finite-snapshot estimation error (supported on $\mathbb{U}^+$) as given in (5.4).

Proof. By triangle inequality, we have

$$\xi \leq \|\mathbf{a}_{\mathbb{U}^+}(\theta)\|_2 \|\mathbf{r}_{\mathbb{U}^+}^{\#} - \mathbf{r}_{\mathbb{U}^+}\|_2 = \sqrt{|\mathbb{U}^+|} \|\mathbf{r}_{\mathbb{U}^+}^{\#} - \mathbf{r}_{\mathbb{U}^+}\|_2 \tag{5.11}$$

From (**ME-1**), we have

$$\begin{aligned} \|\mathbf{r}_{\mathbb{U}^+}^{\#} - \mathbf{r}_{\mathbb{U}^+}\|_2 &\leq \|\mathbf{r}_{\mathbb{U}^+}^{\#} - \tilde{\mathbf{r}}_{\mathbb{U}^+}\|_2 + \|\tilde{\mathbf{r}}_{\mathbb{U}^+} - \mathbf{r}_{\mathbb{U}^+}\|_2 \\ &\leq 2\|\tilde{\mathbf{r}}_{\mathbb{U}^+} - \mathbf{r}_{\mathbb{U}^+}\|_2 = 2\|\mathbf{w}_{\mathbb{U}^+}\|_2 \end{aligned} \tag{5.12}$$

Since $\mathcal{T}(\mathbf{r}_{\mathbb{V}^+}^{ME}) \succeq \mathbf{0}$ and the separation condition is satisfied, we know from Theorem 26 that (5.9)

holds. The proof then follows by substituting (5.11) and (5.12) in (5.9).

Nuclear Norm Minimization

In this case, $\mathbf{r}_{\mathbb{V}^+}^{\#} = \mathbf{r}_{\mathbb{V}^+}^{NUC}$ and the PSD constraint ensures that $\mathcal{T}(\mathbf{r}_{\mathbb{V}^+}^{NUC}) \succeq \mathbf{0}$. Then Theorem 26 applies, leading to the following results

Theorem 20 *If the true DOAs* $\{\theta_l\}_{l=1}^{D}$ *satisfy*

$$\min_{p \neq q} \rho(\theta_p, \theta_q) > 4/|\mathbb{U}^+|$$

and $|\mathbb{U}^+| > 256$*, then there exist positive constants* $\bar{c}_1, \bar{c}_2, \bar{c}_3, \bar{c}_4$ *such that for* $|\mathbb{U}^+| \leq n < |\mathbb{V}^+|$ *and* $n \notin \mathbb{D}^+$*,*

$$\begin{aligned} &|r_n - r_n^{NUC}| \\ &\leq \left(\bar{c}_1 + \frac{\bar{c}_2 \pi n}{|\mathbb{U}^+|} + \frac{\bar{c}_3 \pi^2 n^2}{|\mathbb{U}^+|^2} \right) \frac{\bar{c}_4 D}{\sqrt{|\mathbb{U}^+|}} (\epsilon + \|\mathbf{w}_{\mathbb{U}^+}\|_2) \end{aligned} \tag{5.13}$$

Proof. For any feasible PSD Toeplitz matrix, we have

$$\|\mathcal{T}(\mathbf{x}_{\mathbb{V}^+})\|_* = N[\mathbf{x}_{\mathbb{V}^+}]_0$$

where $[\mathbf{x}_{\mathbb{V}^+}]_0$ denotes the entry corresponding to the zero lag. Therefore, the minimizing solution $\mathbf{r}_{\mathbb{V}^+}^{NUC}$ of (**NUC-PSD**) satisfies $[\mathbf{r}_{\mathbb{V}^+}^{NUC}]_0 \leq [\mathbf{r}_{\mathbb{V}^+}]_0$. Also, notice that

$$\begin{aligned} &\|\mathbf{r}_{\mathbb{U}^+}^{NUC} - \mathbf{r}_{\mathbb{U}^+}\|_2 \leq \|\mathbf{r}_{\mathbb{U}^+}^{NUC} - \tilde{\mathbf{r}}_{\mathbb{U}^+}\|_2 + \|\tilde{\mathbf{r}}_{\mathbb{U}^+} - \mathbf{r}_{\mathbb{U}^+}\|_2 \\ &\leq \epsilon + \|\mathbf{w}_{\mathbb{U}^+}\|_2 \end{aligned}$$

The proof completes by applying the triangle inequality (5.11) on ξ and using these results in (5.9). Letting $\epsilon = \|\mathbf{w}_{\mathbb{U}^+}\|_2$ ensures a non-empty feasible set in (**NUC-PSD**) and the bound in (5.13) becomes proportional to $\|\mathbf{w}_{\mathbb{U}^+}\|_2$. Hence, using the universal result presented in Theorem 26, we have derived upper bounds on the extrapolation error corresponding to (**ME-2**) and (**NUC-PSD**), explicitly in terms of the finite snapshot error $\mathbf{w}_{\mathbb{U}^+}$. These results indicate that the extrapolation error for these algorithms goes to zero asymptotically in the number of snapshots (since $\|\mathbf{w}_{\mathbb{U}^+}\|_2$ tends to zero with increasing snapshots), indicating stability of extrapolation. To the best of our knowledge, Theorems 19,20 present the first results on stability of both algorithms in terms of extrapolation error.

5.2.7 Numerical Results

We follow the same experimental setting as in [228] which uses a coprime array with sensors located at $\mathbb{S} = [0, 3, 5, 6, 9, 10,\ 12, 15, 20, 25]$. In this case, $|\mathbb{D}| = 43, |\mathbb{U}| = 35$ and $|\mathbb{V}| = 51$. For a given number (D) of sources, we generate the true DOAs as $\theta_i = -0.4 + 0.8(n-1)/(D-1)$ for $1 \leq n \leq D$ [228]. We choose both signal and noise powers to be 1 (i.e. SNR of 0 dB). We estimate the correlation matrix at the output of the coprime array by averaging over L snapshots as

$$\tilde{\mathbf{R}}_{\mathbb{S}} = \frac{1}{L}\sum_{i=1}^{L} \mathbf{x}_i \mathbf{x}_i^H$$

Hence, the error $\mathbf{w}_{\mathbb{D}}$ in (5.4) is due to finite snapshot averaging. We study the interpolation error of (**ME-2**) and (**NUC-PSD**) as a function of L. Let $\mathbf{r}^{\#}_{\mathbb{V}^+}$ be the estimate of $\mathbf{r}_{\mathbb{V}^+}$ obtained from either algorithm. The normalized interpolation error is defined as

$$NMSE_{int} = \frac{1}{\|\mathbf{r}_{\mathbb{V}^+\setminus\mathbb{D}^+}\|_2^2} E\Big(\|\mathbf{r}^{\#}_{\mathbb{V}^+\setminus\mathbb{D}^+} - \mathbf{r}_{\mathbb{V}^+\setminus\mathbb{D}^+}\|_2^2\Big) \tag{5.14}$$

In Fig. 5.1, we plot the $NMSE_{int}$ (averaged over 100 Monte Carlo runs) as a function of L for both algorithms corresponding to different number of sources. The interpolation error decreases monotonically with increasing L, indicating stability of reconstruction. It can also been that (**NUC-PSD**) performs better than (**ME-2**), especially for larger L.

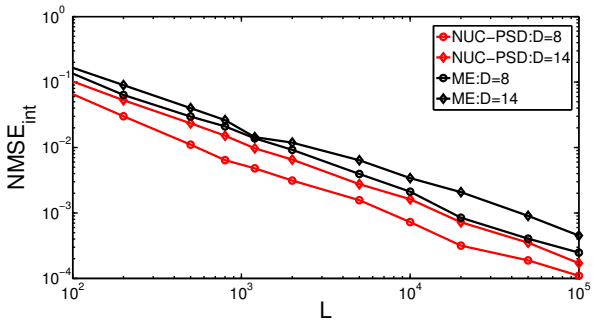


Figure 5.1: $NMSE_{int}$ (averaged over 100 runs) as a function of L for (**NUC-PSD**) and (**ME**) algorithms.

In Fig. 5.2, we compare the MUSIC spectra obtained by applying co-array MUSIC algorithm on $\mathcal{T}(\tilde{\mathbf{r}}_{\mathbb{U}^+})$ (i.e. the correlation matrix corresponding to only the contiguous ULA segment $\mathbb{U}^+$) and on $\mathcal{T}(\mathbf{r}_{\mathbb{V}^+}^{NUC})$ (correlation matrix interpolated using (**NUC-PSD**)). It can be seen that the quality of DOA estimation can be improved by using the full interpolated co-array $\mathbb{V}^+$ instead of using only the contiguous ULA segment $\mathbb{U}^+$.

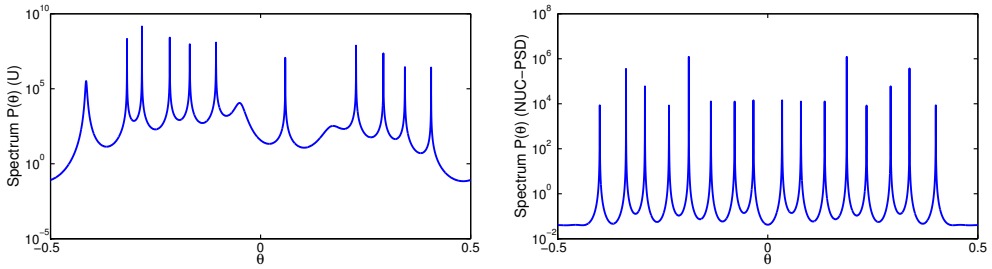


Figure 5.2: MUSIC Spectrum obtained by using co-array MUSIC algorithm on (Left) $\mathcal{T}(\tilde{\mathbf{r}}_{\mathbb{U}^+})$, and (Right) $\mathcal{T}(\mathbf{r}_{\mathbb{V}^+}^{NUC})$, interpolated using (**NUC-PSD**) algorithm. Here, D = 16, L = 50.

5.2.8 Conclusion

In this section, we analyzed the problem of co-array extrapolation that allows us to estimate correlation values at missing lags (or holes) in the co-array of partially augmentable arrays. We provided a universal upper bound on the extrapolation error for these missing correlation values, in terms of the estimation error corresponding to the contiguous ULA segment of the co-array. This bound is universal in the sense that it is obeyed by any extrapolation algorithm that exploits the PSD constraint on the autocorrelation matrix. Using this unified framework, we analyzed the performance of two extrapolation algorithms and established the stability of extrapolation (with respect to finite-snapshot error). Their performance is further illustrated through numerical experiments.

5.3 Understanding the Role of Positive Constraints in Sparse Bilinear Problems

This section considers a certain class of sparse bilinear problems that arises in blind spike deconvolution and array calibration. Existing works based on the idea of lifting attempt to solve this problem by imposing subspace constraints to reduce the number of unknowns, and developing probabilistic guarantees. The goal of this section is to understand the role of positivity in sparse bilinear problems and develop deterministic guarantees for exact support recovery without additional subspace constraints on the unknown quantities. We derive the necessary and sufficient number of measurements needed to exactly recover the location of two spikes simply by exploiting positivity of the spike amplitudes and the blurring kernel. This is the first result of its kind. An interesting consequence of this analysis is to show that exploitation of positivity can lead to exact support recovery even when the number of equations is fewer than the number of unknowns. The theoretical claims are verified through simulations and compared against lifting

based algorithms.

5.3.1 Introduction

Sparse spikes deconvolution is an important class of bilinear problems that finds application in single-molecule microscopy [229, 230], radar signal processing [231], nuclear magnetic resonance spectroscopy [232] and inverse scattering in seismic imaging [233]. The fundamental goal therein is to localize a set of underlying point sources using measurements that are generated via convolution with an unknown point spread function (PSF).

Blind spike deconvolution is an ill-posed problem [31, 234] and it becomes necessary to impose appropriate constraints or priors in order to guarantee exact recovery up to a scaling factor. In [234], the authors introduce the idea of lifting [77, 79] for blind deconvolution. In this case, the PSF is assumed to belong to a known lower-dimensional subspace and the problem is recast as a low rank matrix reconstruction problem [235]. In [31, 236], the spiking signal is assumed to contain only K non-zero elements (corresponding to spike locations) and the PSF is assumed to lie in a P dimensional subspace. Atomic norm minimization [59] is used in [31] to show that $M = O(K^2P^2)$ noiseless measurements are sufficient for exact recovery. In [236], the authors consider the problem of blind array calibration which is equivalent to the blind spike deconvolution problem in Fourier domain, and $M = O(KP)$ measurements are proved to be sufficient.

In contrast to recent works that use subspace constraints to reduce the number of unknowns, the focus of this section is to understand how positive constraints can help solve the ill-posed blind spike deconvolution problem. The role of positivity in solving undetermined linear problems (such as those arising in compressed sensing) has been well investigated [207, 208, 237–240], leading to reduced number of measurements, and deterministic guarantees on support recovery. In [60, 61], the authors study the problem of super-resolution and assume that the unknown spike signal is positive. In this case, the separation conditions on the spikes [141, 142] can be relaxed.

Positivity is also widely observed in real applications. For example, in fluorescence microscopy, both the PSF and the sparse signal are positive [220, 241].

Although the role of positivity in sparse *linear* problems is well understood, to the best of our knowledge, there are no corresponding results for *bilinear* problems. This work provides the first result of its kind to show how positivity alone (without any additional subspace constraints) can help resolve ambiguities in ill-posed bilinear problems. We consider a blind deconvolution problem consisting of two spikes and derive the necessary and sufficient number of measurements to guarantee exact spike detection. Unlike the probabilistic guarantees of [31,234,236], our results are deterministic, and more relevant for practical scenarios. Another significant consequence of our analysis is to show that positivity can ensure exact spike detection even when the number of equations (or measurements) is smaller than the number of unknowns. The analysis framework of [31, 234, 236, 242, 243] will fail in such a setting since they use subspace constraints to reduce the number of unknowns with respect to the number of equations.

5.3.2 Signal Model for Blind Spike Deconvolution: From Bilinear to Linear

Consider the measurement model

$$\mathbf{y} = \text{diag}(\mathbf{g}^\star)\mathbf{A}\mathbf{p}^\star \tag{5.15}$$

where $\mathbf{A} \in \mathbb{C}^{M\times 2N}$ is an overcomplete DFT dictionary with columns

$$\mathbf{a}_n = [1, e^{-j2\pi\frac{n}{2N}}, \cdots, e^{-j2\pi\frac{(M-1)n}{2N}}]^T, -(N-1) \leq n \leq N$$

, $\mathbf{g}^\star \in \mathbb{R}^M$, and $\mathbf{p}^\star \in \mathbb{C}^{2N}$ is a sparse vector with $K \ll N$ non-zero elements. Each column of $\mathbf{A}$ corresponds to one of $2N$ on-grid (normalized) frequencies

$$\mathcal{F}_N = \{n/(2N), -(N-1) \leq n \leq N\}$$

and the indices of the non-zero elements of $\mathbf{p}^\star$ determine which frquencies contribute to the measurement. This model arises frequently in a wide variety of problems including blind spike deconvolution, and array calibration [31, 236]. In the context of blind spike deconvolution, $\mathbf{y}$ represents the frequency domain measurements of a blurred spike train with K spikes whose locations are revealed by the support of sparse $\mathbf{p}^\star$, and $\mathbf{g}^\star$ represents M low frequency measurements of the blurring kernel [142]. For array calibration, $\mathbf{y}$ represents the measurements received at an array with M antennas, $\mathbf{g}^\star$ represents the unknown gain of each antenna, and the support of $\mathbf{p}^\star$ reveals the direction-of-arrival (DOA) of point sources illuminating the array.

5.3.3 Bilinear Model and Equivalent Linear Formulation

The signal model (5.15) is a *bilinear function* of $\mathbf{p}^\star$ and $\mathbf{g}^\star$. Bilinear problems have received great attention in recent times, and fundamental results on bilinear identifiability, as well as algorithms (such as those based on the idea of lifting) for solving $\mathbf{g}^\star$ and $\mathbf{p}^\star$ have been developed [30, 234, 242–244]. However, in addition to the sparsity of $\mathbf{p}^\star$, these results constrain $\mathbf{g}^\star$ to lie in some known subspace. In sharp contrast to these results, we will investigate the role of positivity in bilinear problems and understand if it is at all possible to guarantee perfect recovery of the spiking instants, without imposing any subspace constraint on $\mathbf{g}^\star$. In particular, we will assume $\mathbf{p}^\star, \mathbf{g}^\star \geq \mathbf{0}$, and determine the number of measurements M needed to guarantee exact recovery of the *support of* $\mathbf{p}^\star$. Our result is the first of its kind to show that blind spike deconvolution is possible with positive constraint, and specify the minimum number of measurements necessary and sufficient for exact support recovery.

Without loss of generality, we can assume $g_i \neq 0$ (otherwise the ith measurement is 0 and can be discarded). Using $h_i = 1/g_i$, the bilinear model (5.15) can be linearized as

$$\text{diag}(\mathbf{h}^\star)\mathbf{y} = \mathbf{A}\mathbf{p}^\star$$

In order to recover the support of $\mathbf{p}^\star$, we now consider the following combinatorial problem that enforces positivity of $\mathbf{p}$ and $\mathbf{g}$:

$$\begin{aligned} &\min_{\mathbf{p},\mathbf{h}} \|\mathbf{p}\|_0 \quad (\mathbf{P}_0) \\ &s.t. \quad \text{diag}(\mathbf{h})\mathbf{y} = \mathbf{A}\mathbf{p} \\ &\quad \mathbf{p} \geq \mathbf{0}, \mathbf{h} \geq \boldsymbol{\epsilon} \end{aligned}$$

where $\boldsymbol{\epsilon} > 0$ is chosen to avoid trivial solution $\mathbf{h} = \mathbf{p} = \mathbf{0}$. The main advantage of formulation $(\mathbf{P}_0)$ over (5.15) is that it can be relaxed to a linear problem if $\|\cdot\|_0$ is replaced by $\|\cdot\|_1$. In this paper, we will establish necessary and sufficient conditions for exactly recovering the support of $\mathbf{p}^\star$ in terms of M. The key idea is to show how positivity can help remove the ambiguities in support recovery.

5.3.4 Identifiability Analysis with Positive Constraint

In this section, we develop conditions under which the solution to $(\mathbf{P}_0)$ will reveal the true support. For ease of exposition, we will consider the special case of $K = 2$ (representing two on-grid frequencies) and leave the more general case for future. The case of $K = 2$ frequencies is quite interesting in its own right since it is often used for studying resolution limits of spike detection and DOA estimation. As we will show, even this relatively simpler case requires a detailed characterization of ambiguity sets and provides brand new insights into the role of positivity in resolving the ambiguities of bilinear problem $(\mathbf{P}_0)$.

5.3.5 Single Frequency ($K = 1$) and Role of Positivity

We first consider the simplest case where the model (5.15) consists of $K = 1$ on-grid frequency $\omega^\star \in \{\frac{n}{2N}\}_{n=-(N-1)}^{N}$ and amplitude $p^\star > 0$. Let $\mathbf{h}^\star$ be the true PSF with $\mathbf{h}^\star \geq \boldsymbol{\epsilon}$. Notice that the solution of $(\mathbf{P}_0)$, being the sparsest solution, will also consist of only 1 frequency. The frequency of the solution to $(\mathbf{P}_0)$ can only belong to the following set

$$\mathbb{S}_1 \triangleq \left\{ \omega \in \mathcal{F}_N \;\middle|\; \exists \mathbf{h} \geq \boldsymbol{\epsilon}, p > 0, \frac{pe^{-j2\pi i\omega}}{h_i} = \frac{p^\star e^{-j2\pi i\omega^\star}}{h_i^\star}, i = 0, \cdots, M-1 \right\}$$

The following lemma shows that $\mathbb{S}_1 = \{\omega^\star\}$ as long as we have at least $M = 2$ measurements.

Theorem 21 *Consider the signal model (5.15) consisting of $K = 1$ frequency $\omega^\star$ with amplitude $p^\star > 0$. Then $\mathbb{S}_1 = \{\omega^\star\}$ if $M \geq 2$.*

Proof. Suppose $\mathbb{S}_1$ contains a spurious frequency $\hat{\omega} \neq \omega^\star$. Then we must have

$$\frac{e^{-j2\pi i\omega^\star}}{e^{-j2\pi i\hat{\omega}}} = \frac{h_i^\star}{\hat{h}_i} > 0 \quad i = 0, \cdots, M-1$$

For $i = 1$, we have $e^{j2\pi(\hat{\omega}-\omega^\star)} > 0$, which implies $2\pi(\hat{\omega} - \omega^\star) = 2k\pi, k \in \mathbb{Z}$. Since $|\hat{\omega} - \omega^\star| < 1$, $k = 0$ is the only feasible solution and $\hat{\omega} = \omega^\star$. This simple but intuitive result shows that by simply using the positivity of $\mathbf{h}^\star$ and $\mathbf{p}^\star$, we can exactly identify a single frequency with $M = 2$ measurements.

5.3.6 Two Symmetric Frequencies with Equal Power

We now consider the more interesting case of two frequencies with equal positive power, $p_1^\star = p_2^\star = p^\star$. We first assume that the two frequencies are symmetric with respect to 0, given by $\omega^\star$ and $-\omega^\star$, and then extend the result to the more general case of two asymmetric frequencies. As earlier, our analysis is based on characterizing the solution set of $(\mathbf{P}_0)$. Since $K = 2$, the

solution to $(\mathbf{P}_0)$ can consist of either one, or two frequencies, which can respectively belong to the following two sets

$$\mathbb{S}_{2,1}^{sym}(\omega^\star) \triangleq \Bigg\{ \omega \in \mathcal{F}_N \;\Big|\; \exists \mathbf{h} \geq \boldsymbol{\epsilon}, p > 0$$
$$\frac{pe^{-j2\pi i\omega}}{h_i} = \frac{p^\star(e^{-j2\pi i\omega^\star} + e^{j2\pi i\omega^\star})}{h_i^\star}, i = 0, \cdots, M-1 \Bigg\}$$
$$\mathbb{S}_{2,2}^{sym}(\omega^\star) \triangleq \Bigg\{ (\omega_1, \omega_2) \in \mathcal{F}_N \times \mathcal{F}_N, \Big| \exists \mathbf{h} \geq \boldsymbol{\epsilon}, p_1, p_2 > 0,$$
$$\frac{p_1 e^{-j2\pi i\omega_1} + p_2 e^{-j2\pi i\omega_2}}{h_i} = \frac{p^\star(e^{-j2\pi i\omega^\star} + e^{j2\pi i\omega^\star})}{h_i^\star}$$
$$\omega_1 < \omega_2, i = 0, 1, \cdots, M-1 \Bigg\}$$

Notice that $\omega^\star \neq 0$ (since we have two sources) and we assume $0 < \omega^\star < 1/2$. Our goal is to derive a condition on M such that the solution of $(\mathbf{P}_0)$ consists of exactly two spikes $(-\omega^\star, \omega^\star)$. As a first step, we show that $\mathbb{S}_{2,1}^{sym}(\omega^\star)$ is an empty set:

Theorem 22 *Consider the signal model (5.15) consisting of $K = 2$ on-grid frequencies $(-\omega^\star, \omega^\star)$ satisfying $-\frac{1}{2} < -\omega^\star < 0 < \omega^\star < \frac{1}{2}$ and equal amplitudes $p_1^\star = p_2^\star = p^\star > 0$. If $M \geq 1 + \max\{\lceil \frac{1}{4\omega^\star} \rceil, \lceil \frac{1}{2-4\omega^\star} \rceil\}$, then $\mathbb{S}_{2,1}^{sym}(\omega^\star) = \emptyset$*

Proof. We prove by contradiction. Suppose $\mathbb{S}_{2,1}^{sym}(\omega^\star)$ is non-empty and $\hat{\omega} \in \mathbb{S}_{2,1}^{sym}$. Then, from the definition of the set $\mathbb{S}_{2,1}^{sym}(\omega^\star)$,

$$\cos(2\pi i\omega^\star)e^{j2\pi i\hat{\omega}} = \frac{h_i^\star p}{2h_i p^\star} > 0 \quad i = 0, \cdots, M-1 \tag{5.16}$$

For $i = 1$, (5.16) implies that $\hat{\omega} \in \{0, \frac{1}{2}\}$. We now show that both choices will lead to a contradiction.

We first consider $\hat{\omega} = 0$. In this case, since $\omega^\star < \frac{1}{2}$, it follows that $\frac{1}{4\omega^\star} \leq \lceil \frac{1}{4\omega^\star} \rceil \leq \frac{3}{4\omega^\star}$. Since $M \geq \lceil \frac{1}{4\omega^\star} \rceil$, we can find a positive integer $\hat{i} = \lceil \frac{1}{4\omega^\star} \rceil$ such that $\cos(2\pi\hat{i}\omega^\star) \leq 0$. This will contradict (5.16) for $i = \hat{i} < M$.

Now consider $\hat{\omega} = \frac{1}{2}$. In this case, the condition (5.16) can be rewritten as

$$\cos(2\pi i\omega^\star)e^{j\pi i} = \cos[2\pi i(\frac{1}{2} - \omega^\star)] > 0, 0 \leq i \leq M-1 \tag{5.17}$$

Defining $\tilde{\omega} \triangleq \frac{1}{2} - \omega^\star$, we have $0 < \tilde{\omega} < \frac{1}{2}$. If $\frac{1}{4} \leq \tilde{\omega} < \frac{1}{2}$, we have $\cos(2\pi\tilde{\omega}) \leq 0$ which contradicts (5.17) for $i = 1$. On the other hand, if $0 < \tilde{\omega} < \frac{1}{4}$, then $\cos(2\pi\hat{i}\tilde{\omega}) \leq 0$, again contradicting (5.17) for $i = \hat{i} = \lceil \frac{1}{4\tilde{\omega}} \rceil$. Summarizing, if $M \geq 1 + \max\{\lceil \frac{1}{4\omega^\star} \rceil, \lceil \frac{1}{2-4\omega^\star} \rceil\}$, there does not exist any $\hat{\omega} \in \mathbb{S}_{2,1}^{sym}(\omega^\star)$.

Note that $\max\{\lceil \frac{1}{4\omega^\star} \rceil, \lceil \frac{1}{2-4\omega^\star} \rceil\} \leq \lceil \frac{N}{2} \rceil$ and the equality holds for $\omega^\star = \frac{1}{2N}$ or $\frac{N-1}{2N}$. Hence, Theorem 22 implies that $M \geq 1 + \lceil \frac{N}{2} \rceil$ is sufficient to ensure $\mathbb{S}_{2,1}^{sym} = \emptyset$, regardless of the value of $\omega^\star$. The following corollary shows that this value of M is also necessary.

Corollary 7 *If $M \leq \lceil \frac{N}{2} \rceil$, then $\mathbb{S}_{2,1}^{sym}(\frac{1}{2N})$ is non-empty.*

Proof. We will show that $\mathbb{S}_{2,1}^{sym}(\omega^\star)$ contains $\hat{\omega} = 0$ in this case. If $M \leq \lceil \frac{N}{2} \rceil$, then

$$\cos(2\pi\frac{i}{2N}) > 0 \quad i = 0, 1, \cdots, \lceil \frac{N}{2} \rceil - 1$$

Given $\mathbf{h}^\star \geq \boldsymbol{\epsilon}$ and $p^\star > 0$, choose a $\hat{p}$ satisfying $\hat{p} > 2p^\star$ and construct $\hat{\mathbf{h}}$ as

$$\hat{h}_k = \frac{\hat{p}h_k^\star}{2p^\star \cos(2\pi\frac{k}{2N})} \quad 0 \leq k \leq M-1$$

It can be easily verified that $\hat{\mathbf{h}} \geq \boldsymbol{\epsilon}$. From the definition of the set $\mathbb{S}_{2,1}^{sym}(\frac{1}{2N})$ it follows that $\hat{\omega} = 0 \in \mathbb{S}_{2,1}^{sym}(\frac{1}{2N})$ We now show that the set $\mathbb{S}_{2,2}^{sym}(\omega^\star)$ only contains $(-\omega^\star, \omega^\star)$. Our result is based on the following lemma, which we state without proof.

Lemma 14 *Let $-\frac{1}{2} < \hat{\omega}_1 < \hat{\omega}_2 \leq \frac{1}{2}$ and $\hat{p}_1, \hat{p}_2 > 0$. Define $\xi_k = \hat{p}_1 e^{-j2\pi k\hat{\omega}_1} + \hat{p}_2 e^{-j2\pi k\hat{\omega}_2}$. Then $\hat{p}_1 = \hat{p}_2, \hat{\omega}_1 = -\hat{\omega}_2$ or $\hat{\omega}_1 = 0, \hat{\omega}_2 = \frac{1}{2}$ if ξ_1, ξ_2 are real valued.*

The following theorem shows that $\mathbb{S}_{2,2}^{sym}(\omega^\star)$ does not contain any spurious frequencies

other than $(-\omega^\star, \omega^\star)$.

Theorem 23 *Consider the signal model (5.15) consisting of $K = 2$ on-grid frequencies $(-\omega^\star, \omega^\star)$ satisfying $-\frac{1}{2} < -\omega^\star < 0 < \omega^\star < \frac{1}{2}$ and equal amplitudes $p_1^\star = p_2^\star = p^\star > 0$. Assuming $N \geq 4$, then $\mathbb{S}_{2,2}^{sym}(\omega^\star) = \{(-\omega^\star, \omega^\star)\}$ for all $\omega^\star \in \{\frac{n}{2N}\}_{n=1}^{N-1}$ if $M \geq 1 + \lceil \frac{N}{2} \rceil$.*

Proof. Due to limitations in space, we briefly describe the basic idea of the proof, and leave the details in the future full journal paper. Lemma 14 can be used to argue that any frequency pair belonging to $\mathbb{S}_{2,2}^{sym}(\omega^\star)$ is symmetric. Suppose that $\mathbb{S}_{2,2}^{sym}(\omega^\star)$ contains a frequency pair $(-\hat{\omega}, \hat{\omega})$. As earlier, we prove by contradiction by showing the existence of an integer $\hat{i} \leq \lceil \frac{N}{2} \rceil$ such that $\cos(2\pi\omega^\star\hat{i})\cos(2\pi\hat{\omega}\hat{i}) \leq 0$. From the definition of $\mathbb{S}_{2,2}^{sym}(\omega^\star)$, this actually implies that $\hat{\omega} \notin \mathbb{S}_{2,2}^{sym}(\omega^\star)$ for any $\omega^\star \in \{\frac{n}{2N}\}_{n=1}^{N-1}$. Combining Theorems 22, 23 and Corollary 7, we establish the main result on exact support recovery (for $K = 2$) by solving $(\mathbf{P}_0)$.

Theorem 24 *Consider the signal model (5.15) consisting of $K = 2$ on-grid frequencies $(-\omega^\star, \omega^\star)$ satisfying $-\frac{1}{2} < -\omega^\star < 0 < \omega^\star < \frac{1}{2}$ and equal amplitudes $p_1^\star = p_2^\star = p^\star > 0$. Assuming $N \geq 4$, then the solution $\hat{\mathbf{p}}$ to $(\mathbf{P}_0)$ satisfies Support$(\hat{\mathbf{p}})$ = Support$(\mathbf{p}^\star)$ if and only if $M \geq 1 + \lceil \frac{N}{2} \rceil$.*

Remark 19. *For two symmetric spikes with equal power, the measurement $\mathbf{y}$ is real valued. In this case, there are M equations and $M + 2$ unknowns ($\mathbf{h}^\star \in \mathbb{R}^M$, $p^\star \in \mathbb{R}$ and $\omega^\star \in \mathbb{R}$). Since the number of equations is fewer than the number of unknowns, standard identifiability analysis based on dimension counting [242, 243] will fail to gaurantee exact recovery (even probabilistically). In fact, the number of unknowns in [242, 243] is always assumed to be fewer than the number of equations, by imposing suitable subspace constraints on $\mathbf{h}$. Additionally, the results in [242, 243] hold with probability 1, while the analysis of this paper is deterministic. In contrast, the key contribution of this paper is to show that* exact recovery of support is guaranteed with positive constraint *even the number of equations is strictly smaller than the number of unknowns.*

Remark 20. *If $\mathbf{h}$ is known, the solution of $(\mathbf{P}_0)$ is guaranteed to recover the true frequencies if*

$M \geq 4$ *[96]. However, when* $\mathbf{h}$ *is unknown, Theorem 24 suggests that at least* $M = 1 + \lceil \frac{N}{2} \rceil$ *measurements are needed to uniquely recover the support of two symmetric spikes. This is due to the ambiguity caused by unknown* $\mathbf{h}$ *and will be further demonstrated via numerical simulations.*

5.3.7 Extension to Asymmetric Spikes with Equal Power

Finally, we generalize the above analysis to asymmetric spikes. Assume the true spikes are located at $-\frac{1}{2} < \omega_1^\star < \omega_2^\star \leq \frac{1}{2}$ with equal power $p^\star$. As before, the frequencies associated with the solution to $(\mathbf{P}_0)$ can belong to one of the following two sets

$$\begin{aligned}
\mathbb{S}_{2,1}^{asym} \triangleq \Bigg\{ & \omega \in \mathcal{F}_N \mid \exists \mathbf{h} \geq \boldsymbol{\epsilon}, p > 0 \\
& \frac{p e^{-j2\pi i\omega}}{h_i} = \frac{p^\star (e^{-j2\pi i\omega_1^\star} + e^{-j2\pi i\omega_2^\star})}{h_i^\star}, i = 0, 1, \cdots, M-1 \Bigg\} \\
\mathbb{S}_{2,2}^{asym} \triangleq \Bigg\{ & (\omega_1, \omega_2) \in \mathcal{F}_N \times \mathcal{F}_N | \exists \mathbf{h} \geq \boldsymbol{\epsilon}, p_1, p_2 > 0, \\
& \frac{p_1 e^{-j2\pi i\omega_1} + p_2 e^{-j2\pi i\omega_2}}{h_i} = \frac{p^\star (e^{-j2\pi i\omega_1^\star} + e^{-j2\pi i\omega_2^\star})}{h_i^\star} \\
& \omega_1 < \omega_2, i = 0, 1, \cdots, M-1 \Bigg\}
\end{aligned}$$

Following previous analysis, we can first convert the measurement with asymmetric frequencies into an equivalent form with symmetric frequencies and argue that $\mathbb{S}_{2,2}^{asym} = \{(\omega_1^\star, \omega_2^\star)\}$.

5.3.8 Experiments

In this section, we demonstrate our theoretical claims with different numerical experiments, and compare with lifting based methods in [31,236]. We choose the true frequencies from the grid points $\{\frac{n}{2N}\}_{n=-N+1}^{N}$. We use $p^\star = 1$ and $\boldsymbol{\epsilon} = \mathbf{1}$ in $(\mathbf{P}_0)$ for all the experiments. To compare with AtomicLift [31] and SparseLift [236], we generate random subspaces of dimension L (spanned by the columns of $\mathbf{B} \in \mathbb{R}^{M \times L}$ with uniformly distributed entries in range $(0, 1)$) and construct

$\mathbf{g}^\star = \mathbf{B}\boldsymbol{\lambda}^\star$ where $\boldsymbol{\lambda}^\star = \frac{\mathbf{1}}{M}$ satisfies $\mathbf{h}^\star \geq \boldsymbol{\epsilon}$. We choose $N = 17$ and the support of $\mathbf{p}^\star$ is selected by $(-1, 1)$.

Since $(\mathbf{P}_0)$ is not convex, we solve it using exhaustive search over the grid of $2N$ points and considering all possible frequency pairs (when $K = 2$). Indeed, the aim of this paper is not to develop an efficient algorithm, but to understand if it is at all possible to recover true frequencies using only positivity of $\mathbf{h}^\star$ and $\mathbf{p}^\star$. The recovery is claimed to be successful if this search only yields the true frequency pair. For AtomicLift and SparseLift, we use the same criterion in terms of normalized error for successful recovery as described in [31, 236] respectively. As discussed earlier, when $\mathbf{h}$ is known, the support of $2-$sparse $\mathbf{p}$ can be uniquely determined when $M \geq 4$. However, for unknown $\mathbf{h}$, we have shown in Theorem 24 that at least $\lceil \frac{N}{2} \rceil + 1$ measurements are needed. In Fig.5.3, we demonstrate this fact by choosing true support $(-1, 1)$ and exhaustively searching the feasible set of $(\mathbf{P}_0)$ for all possible supports of size 1 or 2. The probability of success shows a sharp transition exactly happens at $M = \lceil \frac{N}{2} \rceil + 1$ as indicated by Theorem 22 and 23. Note that the performance of both AtomicLift and SparseLift algorithms show degraded performance when compared against this threshold. In fact their performances depend on the subspace dimension L, while our results are independent of knowledge of $\mathbf{B}$. Moreover, the lifting based algorithms cannot work when the number of unknowns approaches the number of equations (i.e. as L approaches M) and AtomicLift almost always fails.

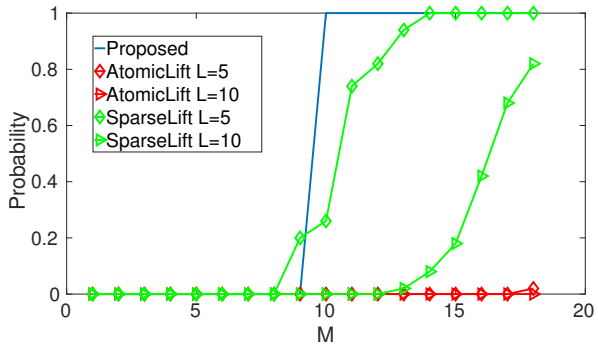


Figure 5.3: Probability of success vs. the number of measurements M.

5.3.9 Conclusion

In this section, we investigated into the role of positivity in sparse bilinear problems that arise in blind spike deconvolution and blind array calibration. In contrast to existing results that assume subspace constraints to reduce the number of unknowns, we merely exploit positivity of the unknown quantities to develop explicit bounds on the number of measurements that guarantee exact support recovery. Our analysis shows that positivity is a powerful constraint that can lead to exact support recovery even when there are more unknowns than the number of equations. Additionally, our theoretical claims are deterministic and hence better applicable to real problems. In future, we will extend our analysis to guarantee exact recovery of an arbitrary number of spikes (beyond $K = 2$) using results on the algebraic variety of trigonometric polynomials.

5.4 Compressive Kriging Using Multi-Dimensional Generalized Nested Sampling

This section considers the problem of Kriging in statistical geophysics where the goal is to interpolate the value of a physical quantity of interest at an unsampled spatial location by exploiting the correlation between measurements collected from a (possibly limited) number of suitably spaced sensors. We introduce the idea of two-dimensional nested array in Kriging to reduce the number of sensors deployed in the field and enable efficient interpolation. Using spatial stationarity (which is a standard assumption in Kriging), we leverage the so-called "difference set" of nested arrays to enable robust spatial compression. Additionally, we can also make predictions beyond the field of view by constructing a virtual covariance matrix of larger dimension from the covariance matrix of the physical measurements. We present robustness analysis based on total least squares and concentration bounds of sample covariance.

5.4.1 Introduction

Estimating the value of a physical quantity (such as temperature, pressure, etc) at an unsampled spatial location using measurements collected from neighboring sensors, is a central problem in geostatistics with applications in oil exploration, seismic data analysis and environmental monitoring [245]. Given data collected from a limited number of sensors distributed over a spatial region of interest, the goal of Kriging is to predict the value of a missing data point at an unknown sensor location, via a probabilistic approach that heavily exploits the correlation structure in the measurements. The basic idea in Kriging is similar to prediction of space-time series, where one computes the best linear function of the observed data, based on prior knowledge of the second order statistics.

Given the limited number of sensors, it is important to determine an appropriate placement that allows efficient interpolation. In prior works on sensor placement [246, 247], the spatial quantity of interest has been modeled as Gaussian processes [248] and near-optimal sensor placement strategies have been developed in terms of entropy or mutual information criterion with the prior knowledge of the covariance kernel. In this paper, we will assume that the measurements are spatially stationary, i.e., the correlation between the measurements collected at any two locations only depends on the difference between their coordinates. However, how the correlation depends on the spatial differences, is not assumed to be known. This is a standard assumption for Kriging-based problems. We will show that such spatial stationarity can be readily utilized to develop near-optimal sensor placement strategies based on the idea of "difference sets". Building on our prior works on nested array geometry, we will demonstrate how to reduce the number of physical sensors by exploiting such multi-dimensional difference sets [124, 125]. Additionally, we will show that predictions can be made beyond the field of view under certain conditions by exploiting the idea of Toeplitz covariance extrapoolation [4, 61].

Our contributions in this section can be summarized as follows

- Under the assumption of spatial stationarity, we show that the number of predictable

locations is determined by the distinct mutual location differences of the sensors. Using two dimensional nested array geometry, we can reduce the number of sensors for a certain number of mutual differences.

- We will further relate the Kriging problem to Toeplitz matrix estimation and show that prediction beyond the field of view is possible if certain conditions hold. This is possible because extrapolation of positive semidefinite Toeplitz matrix can be made which allows more mutual differences than those from measurements.

5.4.2 Problem Formulation

In the following, we first introduce a particular mathematical model for Kriging, called "universal Kriging" [245]. Let $\mathcal{X} \subset \mathbb{R}^n$ denote a spatial region of interest, and $z(\mathbf{x}), \mathbf{x} \in \mathcal{X}$ be the function we are interested in estimating. Under probabilistic model of Kriging, the function $z(\mathbf{x})$ is assumed to be a random process of the form

$$z(\mathbf{x}) = m(\mathbf{x}) + y(\mathbf{x}) \tag{5.18}$$

where $m(\mathbf{x})$ is an unknown deterministic mean function, and $y(\mathbf{x})$ denotes a spatially stationary zero-mean random fluctuation, i.e.

$$\mathbb{E}\Big(y(\mathbf{x}_1)y(\mathbf{x}_2)\Big) = r(\|\mathbf{x}_1 - \mathbf{x}_2\|_2) \tag{5.19}$$

where $r(\cdot)$ is a deterministic function over $\mathbb{R}^n$. This implies that the correlation between the values of $y(\mathbf{x})$ at any two locations only depends on the difference between their coordinates. It should be noted that we will not assume the function $t(\cdot)$ is known. Instead, we will estimate the correlation function from the measurements. The mean function $m(\mathbf{x})$ is modeled as a linear

combination of known basis functions $\{f^l(\mathbf{x})\}$ [249]

$$m(\mathbf{x}) = \sum_{l=0}^{L} a_l f^l(\mathbf{x}) \tag{5.20}$$

where a_l are fixed but unknown coefficients. A standard choice for $f^l(\mathbf{x})$ in one dimension is monomials of the form $f^l(x) = x^l$.

Given measurements $z(\mathbf{x}_i), i = 1, 2, \cdots, N$, we extrapolate the value of $z(\cdot)$ at a given spatial point $\mathbf{x}^\star$ of interest by computing a linear combination of these N measurements. In particular, we solve for optimal coefficients $\lambda_i^\star, 0 \leq i \leq N$ of linear combination as

$$\begin{aligned} \boldsymbol{\lambda}^\star &= \arg\min_{\boldsymbol{\lambda}} \mathbb{E}\left[z(\mathbf{x}^\star) - (\sum_{i=1}^{N} \lambda_i z(\mathbf{x}_i) + \lambda_0) \right]^2 \\ s.t. \qquad & m(\mathbf{x}^\star) = \sum_{i=1}^{N} \lambda_i m(\mathbf{x}_i) + \lambda_0 \end{aligned} \tag{5.21}$$

where $\{\mathbf{x}_i\}_{i=1}^N$ are the locations of the physical sensors. The constraint in the above optimization, which is also known as the universal conditions [249], ensures that the extrapolated value of $z(\mathbf{x}^\star)$ is unbiased. Note that it is due to the constraint of unknown mean function $m(\mathbf{x})$ that universal Kriging is different from Linear Minimum Mean Squared Error (LMMSE) Estimator. The optimal coefficients can be computed as the solution to the following system of equations [245]

$$\left\{ \begin{array}{ll} \sum_j \lambda_j r_{ij} + \sum_l \mu_l f^l(\mathbf{x}_i) = r_i^\star & i = 1, \cdots, N \\ \sum_i \lambda_i f^l(\mathbf{x}_i) = f^l(\mathbf{x}^\star) & l = 0, \cdots, L \end{array} \right\}. \tag{5.22}$$

where $r_{ij} = \mathbb{E}(y(\mathbf{x}_i)y(\mathbf{x}_j))$, $r_i^\star = \mathbb{E}(y(\mathbf{x}_i)y(\mathbf{x}^\star))$ and $\{\mu_l\}$ are Lagrange multipliers.

In this paper, we first study the simple case that the mean function is zero, for which the

optimal linear combination coefficients $\boldsymbol{\lambda}^\star$ is given by

$$\mathbf{R}\boldsymbol{\lambda}^\star = \mathbf{r}^\star \tag{5.23}$$

where $\mathbf{R}_{ij} = r_{ij}$ and $\mathbf{r}^\star = [r_1^\star, \cdots, r_N^\star]^T$. The equation (5.23) reduces to the celebrated Wiener-Kolmogorov linear prediction equation [250].

5.4.3 Two Dimensional Nested Array and Robust Compressive Kriging

Suppose $\mathbf{x}^\star$ is a spatial point at which the function $\mathbf{y}(\mathbf{x}^\star)$ needs to be estimated. The correlation vector $\mathbf{r}^\star$ is not known a priori if no assumption is made on the underlying random field. However, if spatial stationarity holds, we can identify $\mathbf{r}^\star$ from $\mathbf{R}$ if the mutual differences $\mathbb{D}^\star \triangleq \{\|\mathbf{x}_i - \mathbf{x}^\star\|_2\}_{i=1}^N$ is a subset of $\mathbb{D}_{\mathbf{R}} \triangleq \{\|\mathbf{x}_i - \mathbf{x}_j\|_2\}_{i,j=1}^N$.

Thus, to predict as many spatial points as possible, we expect the difference set $\mathbb{D}_{\mathbf{R}}$ to contain enough distinct elements for fixed N. In the following, we will utilize a special sensor array geometry called two-dimensional nested array to achieve this [124, 125].

5.4.4 Two Dimensional Nested Array

Two-dimensional nested array is a multidimensional extension of one-dimensional sparse arrays that include minimum redundancy array (MRA) [204], nested array [45] and Co-prime array [17]. As discussed later, the advantage of two dimensional nested array is that the size of mutual difference set $\mathbb{D}_{\mathbf{R}}$ can be $O(N^2)$, which in turn may allow prediction of more spatial points of interest.

In the next, we will follow [124] to introduce the main concepts of two dimensional nested array. Given a $D \times D$ nonsingular matrix $\mathbf{V}$, the D dimensional lattice generated by $\mathbf{V}$ is defined

as

$$\text{LAT}(\mathbf{V}) = \{\mathbf{t} : \mathbf{t} = \mathbf{V}\mathbf{n}\}$$

where $\mathbf{n} \in \mathbb{N}^D$ is an integer vector. The matrix $\mathbf{V}$ is known as the generator of the lattice. The fundamental parallelepiped (FPD) of the lattice LAT$(\mathbf{V})$ is defined as

Definition 15 *The FPD of* $\mathbf{V}$ *in* D *dimensions is defined as the set of all vectors of the form*

$$\{\mathbf{V}\mathbf{x}, \mathbf{x} \in [0,1)^D\}$$

Conceptually, $FPD(\mathbf{V})$ consists of all points contained in the parallelepiped whose sides are given by the columns of $\mathbf{V}$. In Fig.5.4, we show an example of $FPD(\mathbf{V})$. The volume of $FPD(\mathbf{V})$ is given by $|\det(\mathbf{V})|$.

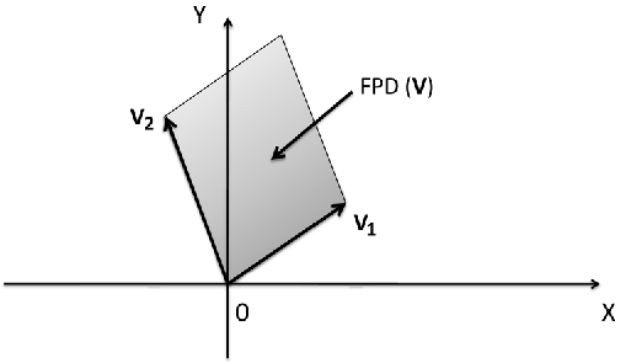


Figure 5.4: An example of $FPD(\mathbf{V})$ for $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2]$ [124]

The two-dimensional nested array consists of two lattices: LAT$(\mathbf{N}^{(s)})$ and LAT$(\mathbf{N}^{(d)})$. The two lattice generators $\mathbf{N}^{(s)}$ and $\mathbf{N}^{(d)}$ is related as

$$\mathbf{N}^{(s)} = \mathbf{N}^{(d)}\mathbf{P}$$

where $\mathbf{P}$ is a 2×2 integer matrix. By the definition of lattice, we can see that LAT$(\mathbf{N}^{(s)})$ is

a sublattice of LAT($\mathbf{N}^{(d)}$) and sparser. We can generate arbitrary sublattice of LAT($\mathbf{N}^{(d)}$) by choosing proper $\mathbf{P}$. In Fig., we present a particular pair of lattices.

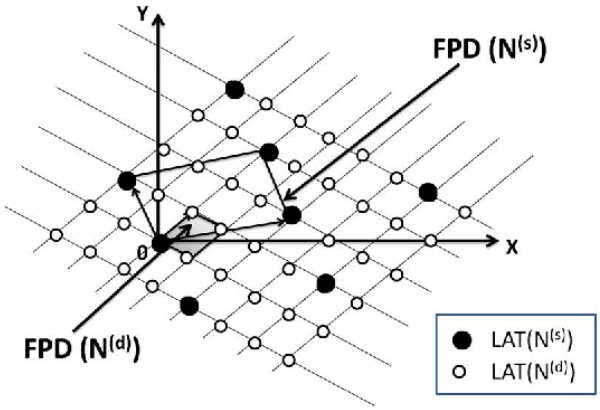


Figure 5.5: A pair of lattces LAT($\mathbf{N}^{(s)}$) and LAT($\mathbf{N}^{(d)}$), where $\mathbf{N}^{(s)} = \mathbf{N}^{(d)}\mathbf{P}$. It can be seen that LAT($\mathbf{N}^{(s)}$) is a sublattice of LAT($\mathbf{N}^{(d)}$) [124]

Now we are ready to state the two-dimensional nested array configuration proposed in [124]. Particularly, the sensor locations $\mathbb{S}$ is given by

$$\mathbb{S} = \{\mathbf{N}^{(d)}\mathbf{n}^{(d)}, \mathbf{n}^{(d)} \in FPD(\mathbf{P})\} \bigcup$$
$$\{\mathbf{N}^{(s)}[k_1, k_2]^T, 0 \leq k_1 \leq N_1^{(s)} - 1, 0 \leq k_2 \leq N_2^{(s)-1}\}$$

The first part of $\mathbb{S}$ is on the dense lattice LAT($\mathbf{N}^{(d)}$) where the number of sensors is given by $N^{(d)} = |\det(\mathbf{P})|$. The second is from the sparse lattice LAT($\mathbf{N}^{(s)}$) with $N^{(s)} = N_1^{(s)}N_2^{(s)}$ elements. It is shown in [124] that the difference set of $\mathbb{S}$ contain $N^{(d)}N^{(s)}$ distinct points on the dense lattice LAT($\mathbf{N}^{(d)}$).

Remark 21. *Compared to entropy or mutual information based methods [246, 247], the two dimensional nested array configuration is deterministic and easy to deploy on the field $\mathcal{X}$. As revealed later, the nested array configuration also avoids the problem of entropy based method that tend to place sensors on the boundary.*

5.4.5 Robustness Analysis of Compressive Kriging

If ideal covariance $\mathbf{R}$ is available and $\mathbf{r}^\star$ can be identified from $\mathbf{R}$, the optimal linear combination coefficients are given by

$$\lambda^\star = \mathbf{R}^\dagger \mathbf{r}^\star \tag{5.24}$$

where $\mathbf{R}^\dagger$ is the pseudo-inverse of $\mathbf{R}$. In practice however, we only collect a finite number of snapshots for each sensor, from which we compute the sample covariance

$$\hat{\mathbf{R}} = \frac{1}{L}\sum_{l=1}^{L} \mathbf{y}_l \mathbf{y}_l^T$$

where $\mathbf{y}_l = [y_l(\mathbf{x}_1), \cdots, y_l(\mathbf{x}_N)]^T$ is the lth snapshot of all N sensors. As in ideal case, we can construct the finite estimate $\hat{\mathbf{r}}$ of $\mathbf{r}^\star$ from $\hat{\mathbf{R}}$. To be general, we assume

$$\hat{\mathbf{R}} = \mathbf{R} + \Delta\mathbf{R} \qquad \hat{\mathbf{r}} = \mathbf{r}^\star + \Delta\mathbf{r}$$

Since we have estimation error on both sides of (5.23), we will use Total Least Squares to provide robustness guarantee of estimating $\lambda^\star$. In particular, from [251], we have the following result

Lemma 15 *Let $\lambda^\star$ be the solution given by (5.24). Assume rank($\mathbf{R}$) = rank($\hat{\mathbf{R}}$), $\|\Delta\mathbf{R}\| \leq \epsilon\|\mathbf{R}\|$ and $\|\Delta\mathbf{r}\|_2 \leq \epsilon\|\mathbf{r}^\star\|_2$. If $\kappa\epsilon < 1$, then*

$$\frac{\|\hat{\lambda} - \lambda^\star\|_2}{\|\lambda^\star\|} \leq \kappa\epsilon\left(1 + \frac{2}{1-\kappa\epsilon}\right)$$

where $\hat{\lambda} = \hat{\mathbf{R}}^\dagger\hat{\mathbf{r}}$ and $\kappa = \|\mathbf{R}\|\|\mathbf{R}^\dagger\|$

To apply Lemma 15, we assume the perturbation is small enough such that the rank of

sample covariance $\hat{\mathbf{R}}$ is consistent with that of $\mathbf{R}$. Additionally, there is a trade-off between the condition number κ and the allowable perturbation ϵ.

To bound the perturbation of $\hat{\mathbf{R}}$ and the associated $\hat{\mathbf{r}}$, we will borrow results on multivariate concentration inequalities from high-dimensional statistics. First, we will bound the perturbation of $\hat{\mathbf{R}}$ in operator norm using the following lemma adapted from Theorem 4.7.1 and Exercise 4.7.3 in [252].

Lemma 16 *[252] Let $\{\mathbf{y}_l\}_{l=1}^{L}$ be zero mean i.i.d Gaussian random vectors distributed as $\mathbf{y}_l \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ and sample covariance is given by $\hat{\mathbf{R}}$. Then for any $u \geq 0$,*

$$\|\hat{\mathbf{R}} - \mathbf{R}\| \leq C \left(\sqrt{\frac{N+u}{L}} + \frac{N+u}{L} \right) \|\mathbf{R}\|$$

with probability at least $1 - 2e^{-u}$ and C is a positive constant.

Since we estimate $\hat{\mathbf{r}}$ from $\hat{\mathbf{R}}$, the perturbation error $\|\Delta\mathbf{r}\|_2$ is bounded by $\|\Delta\mathbf{R}\|_F$. The following lemma gives a probabilistic bound on $\|\mathbf{R} - \hat{\mathbf{R}}\|_F$.

Lemma 17 *(Proposition A.3 in [129]) Let $\{\mathbf{y}_l\}_{l=1}^{L}$ be zero mean i.i.d Gaussian random vectors distributed as $\mathbf{y}_l \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$. Then,*

$$\mathbb{P}\left\{ \|\mathbf{R} - \hat{\mathbf{R}}\|_F \geq \frac{\mathrm{trace}(\mathbf{R})}{\sqrt{L}} \right\} \leq 2e^{-2c_1\sqrt{L}} \tag{5.25}$$

where $\hat{\mathbf{R}}$ is sample covariance and c_1 is a positive universal constant.

Remark 22. *As revealed by Lemma 16, the relative perturbation ratio has an upper bound of order $\sqrt{\frac{N+u}{L}} + \frac{N+u}{L}$. For large L and fixed N, the upper bound decreases as $O(\frac{1}{\sqrt{L}})$. On the other hand, we know that with high probability, the perturbation $\|\Delta\mathbf{r}\|_2$ is upper bounded by $\frac{\mathrm{trace}(\mathbf{R})}{\sqrt{L}} = \frac{\sum_{i=1}^{N}\sigma_i^2}{\sqrt{L}}$ where $\sigma_i^2 = \mathbb{E}(y(\mathbf{x}_i)y(\mathbf{x}_i))$. Thus, given L large enough, we can make $\kappa\epsilon < 1$ and apply Lemma 15.*

To sum up, we have the following result on the robustness of (5.24) with finite (L) snapshots.

Theorem 25 *Assume $\{\mathbf{y}_l\}_{l=1}^L$ be zero mean i.i.d Gaussian random vectors distributed as $\mathbf{y}_l \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ and $\mathbf{R}$ full rank. Let κ be the condition number of $\mathbf{R}$. Given any $u \geq 0$, if*

$$L > \max\left\{ \frac{4(N+u)}{(1+\sqrt{1+\frac{4}{\kappa C}})^2}, \left(\frac{\textit{trace}(\mathbf{R})\kappa}{\|\mathbf{r}^\star\|_2} \right)^2 \right\} \tag{5.26}$$

then with probability at least $1 - 2e^{-u} - 2e^{-2c_1\sqrt{L}}$, we have

$$\frac{\|\hat{\lambda} - \lambda^\star\|_2}{\|\lambda^\star\|} \leq \kappa\epsilon \left(1 + \frac{2}{1-\kappa\epsilon} \right)$$

where

$$\epsilon = \max\left\{ C\sqrt{\frac{N+u}{L}} + \frac{N+u}{L}, \frac{\textit{trace}(\mathbf{R})}{\sqrt{L}\|\mathbf{r}^\star\|_2} \right\} \tag{5.27}$$

and C and c_1 are constants from Lemma 16 and Lemma 17.

Proof. From Lemma 16 and Lemma 17, we know that with probability at least $1 - 2e^{-u} - 2e^{-2c_1\sqrt{L}}$,

$$\|\Delta\mathbf{R}\| \leq C\left(\sqrt{\frac{N+u}{L}} + \frac{N+u}{L} \right) \|\mathbf{R}\| \tag{5.28}$$

$$\|\Delta\mathbf{R}\|_F \leq \frac{\text{trace}(\mathbf{R})}{\sqrt{L}} \tag{5.29}$$

hold simultaneously. Additionally, if (5.26) hold and we choose ϵ as in (5.27), we have

$$\|\Delta\mathbf{R}\| \leq \epsilon\|\mathbf{R}\| \qquad \|\Delta\mathbf{R}\|_F \leq \epsilon\|\mathbf{r}^\star\|_2$$

and $\epsilon\kappa < 1$. Note that the first inequality implies

$$\|\Delta\mathbf{R}\| < \frac{\|\mathbf{R}\|}{\kappa} = s_{min}(\mathbf{R})$$

where $s_{min}(\mathbf{R})$ is the smallest singular value of $\mathbf{R}$. Thus $\hat{\mathbf{R}}$ has the same rank as $\mathbf{R}$. Hence, we can apply Lemma 15 and the proof completes.

5.4.6 Toeplitz Covariance Extrapolation and Prediction Beyond Boundary

To solve $\hat{\lambda}$, we need the difference lags in $\mathbf{r}^\star$ to be contained in $\mathbf{R}$. In previous section, we showed that the use of two-dimensional nested array will allow prediction of more spatial points of interest. However, the maximum difference lag is restricted by the range of sensor locations $\mathbb{S}$. In this section, we will briefly discuss the possibility of predicting locations outside the range of the difference set $\mathbb{D}_{\mathbb{S}}$ of $\mathbb{S}$. This is due to the recent work in [4] on Toeplitz covariance matrix extrapolation. For positive semidefinite Toeplitz matrix, we have following important result

Lemma 18 *[4, 68] A positive (semi) definite Toeplitz matrix $\mathbf{T} \in \mathbb{C}^{N\times N}$ has the following unique decomposition :*

$$\mathbf{T} = \mathbf{V}_{N\times N'}\mathbf{D}\mathbf{V}^H_{N\times N'} + \sigma\mathbf{I}_N \tag{5.30}$$

Here σ is the smallest singular value of $\mathbf{T}$, $\mathbf{I}_N \in \mathbb{R}^{N\times N}$ is the identity matrix and $\mathbf{D} = diag(d_1, d_2, \cdots, d_{N'})$ with $d_i > 0$. The matrix $\mathbf{V}_{N\times N'} \in \mathbb{C}^{N\times N'} = [\mathbf{v}_N(f_1), \mathbf{v}_N(f_2), \cdots, \mathbf{v}_N(f_{N'})]$ is a Vandermonde matrix satisfying $N' < N$ with the ith column $\mathbf{v}_N(f_i)$ given by

$$[\mathbf{v}_N(f_i)]_k = e^{j2\pi f_i k} \quad 0 \le k \le N-1$$

It is shown in [4] that if the associated frequencies $\{f_i\}$ satisfy a separation condition, we can have stable extrapolation of $\mathbf{T}$ to include more difference lags. Particularly, we have following result

Theorem 26 *Let $\tilde{\mathbf{r}} \in \mathbb{C}^N$ denote any Hermitian Toeplitz matrix such that $\mathcal{T}(\tilde{\mathbf{r}}) \succeq \mathbf{0}$. If the angles $\{\theta_l\}_{l=1}^D$ associated with low rank $\mathcal{T}(\mathbf{r})$ satisfy*

$$\min_{p \neq q} \rho(\theta_p, \theta_q) > 4/m \tag{5.31}$$

and $N > m > 256$, then there exist positive constants $\bar{c}_1, \bar{c}_2, \bar{c}_3, \bar{c}_4$ such that for $m \leq n < N$

$$\begin{aligned} &|r_n - \tilde{r}_n| \\ &\leq \left(\bar{c}_1 + \frac{\bar{c}_2 \pi n}{m} + \frac{\bar{c}_3 \pi^2 n^2}{m^2}\right) \left(\frac{\bar{c}_4 D \xi}{m} + ([\tilde{\mathbf{r}}_m]_0 - [\mathbf{r}_n]_0)^+\right) \end{aligned}$$

where $\xi \triangleq \sup_{\theta \in \mathbb{T}} |\langle \mathbf{a}_m(\theta), \tilde{\mathbf{r}}_m - \mathbf{r}_m \rangle|$ and $\rho(\cdot)$ is wrap-around distance function.

However, Theorem 26 cannot be readily applied here as $\mathbf{R}$ is not Toeplitz structured. If the sensor locations form a uniform rectangular array, the covariance matrix $\mathbf{R}$ is multi-level Toeplitz structured [253]. In the future work, we will find a particular unfolding of $\mathbf{R}$ so that we can apply Theorem 26 to allow predictions beyond the boundary.

5.4.7 Experiments

In this section, we will show a sample configuration of two-dimensional nested array and the performance of (5.24) with finite snapshots. In Fig.5.6, we choose $\mathbf{N}^{(d)} = [-1, 0; 0, -1]$, $\mathbf{P} = [-2, 0; 0, -2]$ and $N_1^{(s)} = N_2^{(s)} = 3$. It is clear from the example that the sensors are not concentrated on the boundary and the difference set contains a rectangular array with more

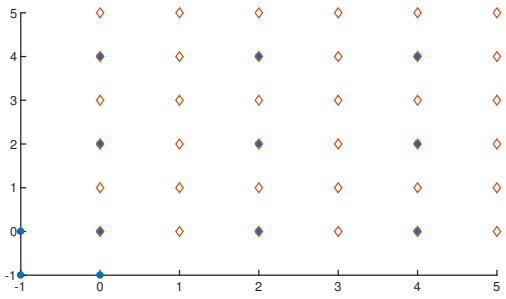


Figure 5.6: A sample configuration of two-dimensional nested array. Blue solid dots represent the sensor locations and red circles denote the difference set in positive orthant.

difference lags than the physical sensor locations $\mathbb{S}$.

The sample covariances $\hat{\mathbf{R}}$ and $\hat{\mathbf{r}}$ are computed from L snapshots of a zero-mean spatially stationary Gaussian process. In Fig.5.7, we show the relative estimation error $\|\hat{\lambda} - \lambda^\star\|_2 / \|\lambda^\star\|_2$ as a function of L. The relative error decreases as L increases.

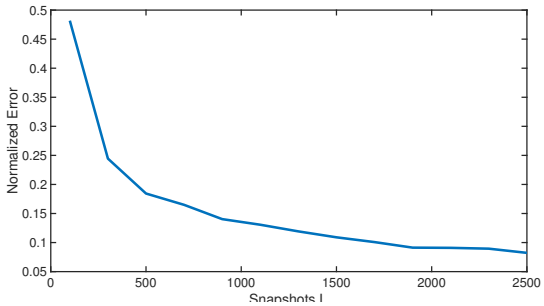


Figure 5.7: Relative estimation error as a function of snapshot number L, averaged over 100 runs.

5.4.8 Conclusion

In this section, we analyzed the problem of co-array extrapolation that allows us to estimate correlation values at missing lags (or holes) in the co-array of partially augmentable arrays. We provided a universal upper bound on the extrapolation error for these missing correlation values, in terms of the estimation error corresponding to the contiguous ULA segment of the co-array.

This bound is universal in the sense that it is obeyed by any extrapolation algorithm that exploits the PSD constraint on the autocorrelation matrix. Using this unified framework, we analyzed the performance of two extrapolation algorithms and established the stability of extrapolation (with respect to finite-snapshot error). Their performance is further illustrated through numerical experiments.

Chapter 5, in part, is a reprint of the material as it appears in the papers:

- H. Qiao, M. C. Hucumenoglu and P. Pal, "Compressive Kriging Using Multi-Dimensional Generalized Nested Sampling, in the *Proceedings of Asilomar Conference on Signals, Systems and Computers*, 2018.

- H. Qiao and P. Pal, "Understanding the Role of Positive Constraints in Sparse Bilinear Problems", in the *Proceedings of 7th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Dec. 2017.

The dissertation author was the primary investigator and author of these papers.

Bibliography

[1] Heng Qiao and P. 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*, to appear.

[2] Heng Qiao, and Piya Pal, "Generalized Nested Sampling for Compressing Low Rank Toeplitz Matrices", *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 1844-1848, Nov. 2015.

[3] Heng Qiao and P. Pal, "On Maximum Likelihood Methods For Localizing More Sources than Sensors", *IEEE Signal Processing Letters*, vol. 24, no. 5, pp. 703-706, 2017.

[4] Heng Qiao and P. Pal, "Gridless Line Spectrum Estimation and Low-Rank Toeplitz Matrix Compression Using Structured Samplers: A Regularization-Free Approach", *IEEE Transactions on Signal Processing*, vol. 65, no. 9, pp. 2211-2226, May 2017.

[5] Ali Koochakzadeh, H. Qiao and P. Pal, "On Fundamental Limits of Joint Sparse Support Recovery Using Certain Correlation Priors", *IEEE Transactions on Signal Processing*, vol. 66, no. 17, pp. 4612-4625, Sep. 2018.

[6] Heng Qiao and Piya Pal, "Stable Compressive Low Rank Toeplitz Covariance Estimation Without Regularization", accepted by *Asilomar Conference on Signals, Systems, and Computers*, 2016.

[7] H. Qiao and P. Pal, "Performance Limits of Covariance-Driven Super Resolution Imaging", in *Proc. Asilomar Conf. on Signals, Syst., and Comput.*, 2017.

[8] H. Qiao, M. C. Hucumenoglu and P. Pal, "Compressive Kriging Using Multi-Dimensional Generalized Nested Sampling, in the *Proceedings of Asilomar Conference on Signals, Systems and Computers*, 2018.

[9] Heng Qiao and Piya Pal, "Finite Sample Analysis of Covariance Compression Using

Structured Samplers", *The Ninth IEEE Sensor Array and Multichannel Signal Processing Workshop*, Rio de Janeiro, Brazil, july 2016.

[10] Heng Qiao and Piya Pal, "Generalized Nested Sampling for Compression and Exact Recovery of Symmetric Toeplitz Matrices", IEEE GlobalSIP Symposium on Information Processing for Big Data, Atlanta, Dec. 3-5, 2014.

[11] Heng Qiao and Piya Pal, "Sparse Phase Retrieval Using Partial Nested Fourier Samplers", accepted by IEEE GlobalSIP Symposium on Information Processing for Big Data, Orlando, Dec. 14-16, 2015.

[12] Heng Qiao and Piya Pal, "Sparse Phase Retrieval with Near Minimal Measurements: A Structured Sampling Based Approach", In: *IEEE International Conference on Accoustics, Speech and Signal Processing (ICASSP)*, 2016.

[13] H. Qiao and P. Pal, "Unified Analysis of Co-Array Interpolation for Direction-of-Arrival Estimation," in Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, USA, March 2017

[14] H. Qiao and P. Pal, "A non-convex approach to non-negative super-resolution: theory and algorithm", in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Brighton, UK, May 2019.

[15] H. Qiao and P. Pal, "Understanding the Role of Positive Constraints in Sparse Bilinear Problems", in *Proc. IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2017.

[16] P. Pal and P. P. Vaidyanathan, "Pushing the limits of sparse support recovery using correlation information", *IEEE Transactions on Signal Processing*, vol. 63, no. 3, pp. 711-726, Feb. 2015.

[17] P. P. Vaidyanathan and P. Pal, "Sparse sensing with co-prime samplers and arrays", *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 573-586, Feb. 2011.

[18] P. Pal, "Correlation awareness in low-rank models: sampling, algorithms, and fundamental limits", *IEEE Signal Processing Magazine*, vol. 35, no. 4, pp. 56-71, July 2018.

[19] P. Pal and P. P. Vaidyanathan, "Coprime sampling and the music algorithms", *Proc. IEEE Digital Signal Process. Signal Process. Educ. Workshop*, pp. 289-294, Sedona, AZ, Jan. 2011.

[20] D. Donoho, “50 years of Data Science”, Sep. 2015.

[21] R. G. Baraniuk, “More is less: signal processing and the data deluge,” *Science*, vol. 331, no. 6018, pp. 717-719, Feb. 2011.

[22] E. J. Candès, J. Romberg and T. Tao, “Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information”, *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489-509, Feb. 2006.

[23] E. J. Candès and T. Tao, “Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?”, *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406-5425, Dec. 2006.

[24] E. J. Candès and T. Tao, “Decoding by Linear Programming”, *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 42034215, Dec. 2005.

[25] D. L. Donoho, “Compressed Sensing”, *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289-1306, April 2006.

[26] S. S. Chen, D. L. Donoho and M. A. Saunders, “Atomic decomposition by basis pursuit”, *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33-61, 1998.

[27] M. F. Duarte and R. G. Baraniuk, “Kronecker Compressive Sensing,” IEEE Trans. Image Processing, vol. 21, no. 2, pp. 494-504, Aug. 2011.

[28] G. Dasarathy, P. Shah, D. N. Bhaskar and R. D. Nowak, “Sketching Sparse Matrices, Covariances, and Graphs via Tensor Products”, *IEEE Trans. Inf. Theory*, vol. 61, no. 3, pp. 1373-1388, Nov. 2015.

[29] R. Kueng, H. Rauhut and U. Terstiege, “Low rank matrix recovery from rank one measurements”, *Applied and Computational Harmonic Analysis*, to appear.

[30] S. Ling and T. Strohmer, “Self-Calibration and Biconvex Compressive Sensing”, *Inverse Problems*, vol. 31, no. 11, 115002, Sep. 2015.

[31] Y. Chi, “Guaranteed Blind Sparse Spikes Deconvolution via Lifting and Convex Optimization”, *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 4, pp. 782-794, June 2016.

[32] J. Leech, “On the representation of 1,2,...,n by differences,” *J. London Math. Soc.*, vol. s1-3, no. 2, pp. 160-169, 1956.

[33] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79-86, March 1951.

[34] L. L. Scharf, "Statistical signal processing: Detection, Estimation, and Time Series Analysis," *Prentice Hall*, 1991.

[35] B. Moore,"Principal component analysis in linear systems: Controllability, observability, and model reduction," *IEEE Transactions on Automatic Control*, vol. 26, no. 1, pp. 17-32, Feb. 1981.

[36] N. Kambhatla, and T. K. Leen."Dimension reduction by local principal component analysis." *Neural Computation* vol. 9, no. 7, pp. 1493-1516, Oct. 1997.

[37] P. J. Bickel and E. Levina, "Regularized estimation of large covariance matrices," *The Annals of Statistics*, vol. 36, no. 1, pp.199-227, Feb. 2008.

[38] Y. Chen, Y. Chi and A. Goldsmith, "Exact and stable covariance estimation from quadratic sampling via convex programming," *IEEE Trans. Inf. Theory*, vol. 61, no. 7, pp. 4034-4059, July 2015.

[39] D. Romero, and G. Leus, "Compressive covariance sampling." *Information Theory and Applications Workshop (ITA)*, pp. 1-8, Feb. 2013.

[40] D. Romero, R. Lopez-Valcarce, and G. Leus, "Compression limits for random vectors with linearly parameterized second-order statistics." *IEEE Trans. Inf. Theory*, vol. 61, no. 3, pp. 1410-1425, March 2015.

[41] D. Romero, D. D. Arianandа, Z. Tian and G. Leus, "Compressive Covariance Sensing: Structure-based compressive sensing beyond sparsity", *IEEE Signal Processing Magazine*, vol. 33, no. 1, pp. 78-93, Jan. 2016.

[42] D. D. Arianandа and G. Leus, "Compressive wideband power spectrum estimation, *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4775-4789, Sep. 2012.

[43] G. Leus and Z. Tian, "Recovering second-order statistics from compressive measurements, in *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pp. 337-340, Dec. 2011.

[44] D. Cohen, and Y. C. Eldar, "Sub-Nyquist sampling for power spectrum sensing in cognitive radios: A unified approach." *IEEE Transactions on Signal Processing*, vol. 62, no.15, pp. 3897-3910, Aug. 2014.

[45] Piya Pal and P. P. Vaidyanathan, "Nested arrays: a novel approach to array processing with enhanced degrees of freedom", *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4167-4181, Aug. 2010.

[46] Z. Tian, Y. Tafesse and B. M. Sadler, "Cyclic feature detection with sub-nyquist sampling for wideband spectrum sensing," *IEEE J. Selected Topics in Signal Proc.*, vol. 6, no. 1, pp. 58-69, Dec. 2012.

[47] P. Stoica and A. Nehorai, "MUSIC, Maximum Likelihood, and Cramer-Rao Bound", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 5, pp. 720-741, May 1989.

[48] H. Li, P. Stoica and J. Li, "Computationally Efficient Maximum Likelihood Estimation of Structured Covariance Matrices", *IEEE Transactions on Signal Processing*, vol. 47, no. 5, pp. 1314-1323, May 1999.

[49] H. Krim and M. Viberg,"Two decades of array signal processing: The parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67 -94, July 1996.

[50] M. Fazel, H. Hindi, and S. P. Boyd, "A rank minimization heuristic with application to minimum order system approximation", *Proc. of American Control Conference, 2001*, vol. 6, pp. 4734-4739, 2001.

[51] E. Candes and B. Recht, "Exact Matrix Completion via Convex Optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717-772, Dec. 2009.

[52] Z. Ding and L. Qiu, "Blind MIMO Channel Identification From Second Order Statistics Using Rank Deficient Channel Convolution Matrix," *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 535-544, Feb. 2003.

[53] E. Candes and T. Tao, "The Power of Convex Relaxation: Near-Optimal Matrix Completion," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2053-2080, May 2010

[54] D. Gross, "Recovering low-rank matrices from few coefficients in any basis," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp.1548-1566, March 2011.

[55] Y. Plan and R. Vershynin, "Robust 1-bit Compressed Sensing and Sparse Logistic Regression: A Convex Programming Approach," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 482-494, Jan. 2013.

[56] G. Tang, B. N. Bhaskar, P. Shah and B. Recht, "Compressed Sensing Off the Grid," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7465 -7490, Nov. 2013.

[57] Y. Li and Y. Chi, "Off-the-Grid Line Spectrum Denoising and Estimation with Multiple Measurement Vectors", *IEEE Transactions on Signal Processing*, vol. 64, no. 5, pp. 1257-1269, March 2016.

[58] B. N. Bhaskar, G. Tang and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5987-5999, Apr. 2012.

[59] G. Tang, B. N. Bhaskar and B. Recht, "Near minimax line spectral estimation", *IEEE Trans. Inf. Theory*, vol. 61, no. 1, pp. 499 -512, Jan. 2015.

[60] G. Schiebinger, E. Robeva and B. Recht, " Superresolution without Separation," *arXiv preprint arXiv:1506.03144v2*, 2015.

[61] V. I. Morgenshtern and E. J. Candès, "Super-Resolution of Positive Sources: the Discrete Setup," *arXiv preprint arXiv:1504.00717v1*, 2015.

[62] J-M. Azais, Y. De Castro, and F. Gamboa, "Spike detection from inaccurate samplings," *Applied and Computational Harmonic Analysis*, vol. 38, no. 2, pp. 177-195, March 2015.

[63] Y. De Castro and F. Gamboa, "Exact reconstruction using Beurling minimal extrapolation," *J. Math. Anal. Appl.*, vol. 395, no. 1, pp. 336-354, Nov. 2012.

[64] U. Grenander and G. Szego, "Toeplitz forms and their applications," *Chelsea Pub.Co.*, 1984.

[65] M. Jansson and B. Ottersten, "Structured Covariance Matrix Estimation:A Parametric Approach," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 5, pp. 3172-3175, 2000.

[66] J. Makhoul, "Linear Prediction: A Tutorial Review", *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561-580, April 1975.

[67] P. Stoica and K. C. Sharman, "Maximum Likelihood Methods for Direction-of-Arrival Estimation", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, no. 7, pp. 1132-1143, July 1990.

[68] P. Stoica and R. Moses, "Spectral analysis of signals", Prentice-Hall, NJ: Upper Saddle River, 2005.

[69] S. N. Negahban, P. Ravikumar, M. J. Wainwright and B. Yu, " A unified framework for

high-dimensional analysis of m-estimators with decomposable regularizers," *Statistical Science*, vol. 27, no. 4, pp. 538-557, 2012.

[70] E. Candès and X. Li, "Solving Quadratic Equations via PhaseLift When There Are About as Many Equations as Unknowns," *Found. Comput. Math.*, vol. 14, no. 5, pp. 1017-1026, Oct. 2014.

[71] M. Kabanava, R. Kueng, H. Rauhut and U. Terstiege, "Stable low-rank matrix recovery via null space properties", *arXiv preprint arXiv:1507.07184v4*, 2015.

[72] Y. C. Eldar, P. Sidorenko, D. G. Mixon, S. Barel and O. Cohen, "Sparse Phase Retrieval from Short-Time Fourier Measurements', *IEEE Signal Processing Letters*, vol. 22, no. 5, pp. 638-642, Oct. 2014.

[73] E. Candès and C. Fernandez-Granda,"Super-resolution from noisy data", *J. Fourier Anal. Appl.*, vol. 19, no. 6, pp. 1229-1254, 2013.

[74] T. Bäckström, "Vandermonde Factorization of Toeplitz Matrices and Applications in Filtering and Warping", *IEEE Transactions on Signal Processing*, vol. 61, no. 24, pp. 6257-6263, Dec. 2013.

[75] R. Balan, P. Casazza and D. Edidin, "On signal reconstruction without phase," *Appl. Comput. Harmon. Anal.*, vol. 20, no. 3, pp. 345-356, May 2006.

[76] Y. Wang and Z. Xu, "Phase retrieval for sparse signals," *Appl. Comput. Harmon. Anal.*, vol. 37, no. 3, pp. 531-544, Nov. 2014.

[77] E. Candès, Y. C. Eldar, T. Strohmer and V. Voroninski, "Phase retrieval via matrix completion," *SIAM J. Imaging Sci.* vol. 6, no. 1, pp. 199-225, 2013.

[78] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao and M. Segev, "Phase retrieval with application to optical imaging: A contemporary overview," *IEEE Signal Processing Magazine*, vol. 32, no. 3, pp. 87-109, May 2015.

[79] E. Candès, T. Strohmer and V. Voroninski, "PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming, "vol. 66, no. 8, pp. 1241-1274, Aug. 2013.

[80] E. Candès, X. Li and M. Soltanolkotabi, "Phase retrieval via Wirtinger Flow: Theory and Algorithms," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1985 - 2007, April 2015.

[81] K. Jaganathan, S. Oymak and B. Hassibi, "On robust phase retrieval for sparse signals," *Proc. 50th Ann. Allerton Conf. Commun., Control, Comput.*, pp. 794-799 2012.

[82] K. Jaganathan, S. Oymak and B. Hassibi, "Recovery of sparse 1-D signals from the magnitudes of their Fourier transform," *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 1473-1477 2012.

[83] K. Jaganathan, S. Oymak and B. Hassibi, "Sparse phase retrieval: Uniqueness guarantees and recovery algorithms," *arXiv preprint arXiv: 1311.2745*, Nov. 2013.

[84] H. Ohlsson and Y. C. Eldar, "On conditions for uniqueness in sparse phase retrieval," In: *IEEE International Conference on Accoustics, Speech and Signal Processing (ICASSP)*, 2014.

[85] J. Ranieri, A. Chebira, Y. M. Lu and M. Vetterli, "Phase retrieval for sparse signals: Uniqueness conditions," *arXiv preprint arXiv: 1308.3058*, Aug. 2013.

[86] P. Natrapalli, P. Jain and S. Sanghavi, "Phase retrieval using alternating minimization," *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, pp.2796-2804, 2013.

[87] H. Ohlsson, A. Yang, R. Dong and S. Sastry, "CPRL–An Extension of Compressive Sensing to the Phase Retrieval Problem,' *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, pp.1367-1375, 2012.

[88] X. Li and V. Voroninski, "Sparse signal recovery from quadratic measurements via convex programming," *SIAM Journal on Mathematical Analysis*, vol.45, no.5, pp.3019-3033, 2013.

[89] S. Oymak, A. Jalali, M. Fazel, Y. Eldar and B. Hassibi, "Simultaneously structured models with applications to sparse and low-rank matrices," *IEEE Trans. Inf. Theory*, vol. 63, no. 4, pp. 1043 - 1055, Feb. 2015.

[90] S. Bahmani and J. Romberg, "Efficient Compressive Phase Retrieval with Constrained Sensing Vectors," *arXiv preprint arXiv: 1507.08254v1*, July 2015.

[91] Y. C. Eldar and S. Mendelson, "Phase retrieval: Stability and recovery guarantees," *Appl. Comput. Harmon. Anal.*, vol. 36, no. 3, pp. 473-494, May 2014.

[92] P. Walk and P. Jung, "Stable recovery from the magnitude of symmetrized Fourier measurements,"In: *IEEE International Conference on Accoustics, Speech and Signal Processing (ICASSP)*, 2014.

[93] C. Yapar, V. Pohl and H. Boche, "Fast compressive phase retrieval from Fourier measurements," In: *IEEE International Conference on Accoustics, Speech and Signal Processing (ICASSP)*, 2015.

[94] E. Candès, X. Li and M. Soltanolkotabi, "Phase retrieval from coded diffraction patterns," *to appear in Appl. Comput. Harmon. Anal.*, 2014.

[95] Y. Wang, "Minimal frames for phase retrieval," In: *Workshop of phaseless recovery*, College Park, 2013.

[96] A. Foucart and H. Rauhut, "A Mathematical Introduction to Compressive Sensing," *Applied and Numerical Harmonic Analysis*, Birkhäuser, 2013.

[97] E. Candès, X. Li and M. Soltanolkotabi, "Phase retrieval via Wirtinger Flow: Theory and Algorithms," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1985 - 2007, April 2015.

[98] P. P. Vaidyanathan, "Multirate Systems and Filter Banks," 1993: Prentice-Hall

[99] A. S. Bandeira, J. Cahill, D. G. Mixon and A. A. Nelson, "Saving phase: Injectivity and stability for phase retrieval," *Applied and Computational Harmonic Analysis*, vol. 37, no. 1, pp. 106-125 July 2014.

[100] J. R. Fienup, "Phase retrieval algorithms: A personal tour [Invited]," *Appl. Opt.*, vol. 52, pp. 45-56, 2013.

[101] J. J. Fuchs, "Recovery of exact sparse representations in the presence of noise", *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3601-3608, Oct. 2005.

[102] J. R. Fienup and J. C. Dainty, "Phase retrieval and image reconstruction for astronomy," in *Image Recovery: Theory and Application*, H. Stark, Ed. New York, NY, USA: Academic, 1987, pp. 231 - 275.

[103] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, pp. 237-246, 1972.

[104] K. Huang, Y. C. Eldar and N. D. Sidiropoulos, "Phase Retrieval from 1D Fourier Measurements: Convexity, Uniqueness, and Algorithms", *arXiv preprint arXiv: 1603.05215v1*, March 2016.

[105] Y. Shechtman, A. Beck and Y. C. Eldar, "GESPAR: Efficient phase retrieval of sparse signals," *IEEE Transactions on Signal Processing*, vol. 62, no. 4, pp. 928-938, Feb. 2014.

[106] D. S. Weller, A. Pnueli, G. Divon, O. Radzyner, Y. C. Eldar and J. A. Fessler, "Undersampled Phase Retrieval With Outliers", *IEEE Transactions on Computational Imaging*, vol. 1, no. 4, pp. 247-258, Dec. 2015.

[107] P. Schniter and S. Rangan, "Compressive Phase Retrieval via Generalized Approximate Message Passing", *IEEE Transactions on Signal Processing*, vol. 63, no. 4, pp. 1043-1055, Feb. 2015.

[108] S. Mukherjee and C. S. Seelamantula, "Fienup Algorithm With Sparsity Constraints: Application to Frequency-Domain Optical-Coherence Tomography", *IEEE Transactions on Signal Processing*, vol. 62, no. 18, pp. 4659-4672, Sep. 2014.

[109] Y. Chen and Y. Chi, "Robust Spectral Compressed Sensing via Structured Matrix Completion", *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6576-6601, Oct. 2014.

[110] M. Wax and A. Leshem, "Joint Estimation of Time Delays and Directions of Arrival of multiple Reflections of a Known Signal", *IEEE Transactions on Signal Processing*, vol. 45, no. 10, pp. 2477-2484, Oct. 1997.

[111] A. Leshem and A. J. van der Veen, "Radio-astronomical imaging in the presence of strong radio interference", *IEEE Trans. Inf. Theory*, vol. 46, no. 5, pp. 1730-1747, Aug. 2000.

[112] L. C. Potter, E. Ertin, J. T. Parker and M. Cetin, "Sparsity and compressed sensing in radar imaging", *Proc. IEEE*, vol. 98, no. 6, pp. 1006-1020, Jun. 2010.

[113] J. C. Mosher, P. S. Lewis and R. M. Leahy, "Multiple Dipole Modeling and Localization from Spatio-Temporal MEG Data", *IEEE Trans. Biomedical Engineering*, vol. 39, no. 6, pp. 541-557, June 1992.

[114] K. G. Oweiss and D. J. Anderson, "A new technique for blind source separation using subband subspace analysis in correlated multichannel signal environments", *Proc. IEEE ICASSP*, 2001.

[115] H. M. Huizenga, J. C. de Munck, L. J. Waldorp and R. P. P. P. Grasman, "Spatiotemporal EEG/MEG source analysis based on a parametric noise covariance model", *IEEE Trans. Biomedical Engineering*, vol. 49, no. 6, pp. 533-539, June 2002.

[116] S. Miron, N. Le Bihan and J. I. Mars, "Vector-Sensor MUSIC for Polarized Seismic Sources Localization", *EURASIP Journal on Applied Signal Processing*, vol. 2005, no. 1, pp. 74-84, Jan. 2005.

[117] S. Miron, N. Le Bihan and J. I. Mars, "Quaternion-MUSIC for Vector-Sensor Array Processing", *IEEE Transactions on Signal Processing*, vol. 54, no. 4, pp. 1218-1229, April 2006.

[118] L. Borcea, G. Papanicolaou, C. Tsogka and J. Berryman, "Imaging and time reversal in random media", *Inverse Problems*, vol. 18, no. 5, pp. 1247-1279, 2002.

[119] D. Malioutov, M. Cetin and A. Willsky, "Sparse signal reconstruction persepective for source localization with sensor arrays", *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010-3022, Aug. 2005.

[120] P. Stoica, P. Babu and J. Li, "SPICE: A sparse covariance-based estimation method for array processing", *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 629-638, Feb. 2011.

[121] Z. Yang, L. Xie and C. Zhang, "A discretization-free sparse and parametric approach for linear array signal processing", *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 4959 - 4973, Oct. 2014.

[122] Z. Yang and L. Xie, "On gridless sparse methods for line spectral estimation from complete and incomplete data", *IEEE Transactions on Signal Processing*, vol. 63, no. 12, pp. 3139-3153, June 2015.

[123] P. Pal and P. P. Vaidyanathan,"A grid-less approach to underdetermined direction of arrival estimation via low rank matrix denoising", *IEEE Signal Processing Letters*, vol. 21, no. 6, pp. 737-741, June 2014.

[124] P. Pal and P. P. Vaidyanathan, "Nested Arrays in Two Dimensions, Part I: Geometrical Considerations", *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4694-4705, Sep. 2012.

[125] P. Pal and P. P. Vaidyanathan, "Nested Arrays in Two Dimensions, Part II: Applications in Two Dimensional Array Processing", *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4706-4718, Sep. 2012.

[126] P. Pal and P. P . Vaidyanathan, "Coprime sampling and the music algorithm", in *Proc. 2011 Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE)*, Jan. 2011.

[127] P. Stoica, G. Tang, Z. Yang and D. Zachariah, "Gridless compressive-sensing methods for frequency estimation: points of tangency and links to basics", *Proc. of the 22nd European Signal Processing Conference (EUSIPCO)*, 2014.

[128] A. Panahi, M. Viberg and B. Hassibi, "A numerical implementation of gridless compressed sensing", *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015.

[129] F. Bunea and L. Xiao, "On the sample covariance matrix estimator of reduced effective rank population matrices, with applications to fPCA," *Bernoulli*, vol. 21, no. 2, pp. 1200-1230, 2015.

[130] M. Wang and A. Nehorai, "Coarrays, MUSIC, and the Cramér Rao bound", *IEEE Transactions on Signal Processing*, vol. 65, no.4, pp. 933-946, Feb. 2017.

[131] Y. I. Abramovich, D. A. Gray, A. Y. Gorokhov and N. K. Spencer, "Positive-Definite Toeplitz Completion in DOA Estimation for Nonuniform Linear Antenna Arrays - Part I: Fully Augmentable Arrays", *IEEE Transactions on Signal Processing*, vol. 46, no. 9, pp. 2458 - 2471, Sep. 1998.

[132] Q. Cheng, P. Pal, M. Tsuji and Y. Hua, "An MDL Algorithm for Detecting More Sources Than Sensors Using Outer-Products of Array Output", *IEEE Transactions on Signal Processing*, vol. 62, no. 24, pp. 6438-6453, Dec. 2014.

[133] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules", *IEEE Signal Proc. Mag.*, vol. 21, no. 4, pp. 36-47, July 2004.

[134] J. P. Burg, D. G. Luenberger and D. L. Wenger, "Estimation of Structured Covariance Matrices", *Proc. IEEE*, vol. 70, no. 9, pp. 963-974, Sep. 1982.

[135] O. Balkan, K. K. Delgado and S. Makeig, "Localization of More Sources Than Sensors via Jointly-Sparse Bayesian Learning", *IEEE Signal Processing Lett.*, vol. 21, no. 2, pp. 131-134, Feb. 2014.

[136] T. W. Anderson and I. Olkin, "Maximum-likelihood estimation of the parameters of a multivariate normal distribution", *Linear Algebra and its Applications*, vol. 70, pp. 147-171, Oct. 1985.

[137] P. Stoica and T. Söderström, "On reparameterization of loss functions used in estimation and the invariance principles", *Signal Process.*, vol. 17, no. 4, pp. 383-387, Aug. 1989.

[138] C.-L. Liu and P. Vaidyanathan, "Remarks on the spatial smoothing step in coarray MUSIC", *IEEE Signal Processing Lett.*, vol. 22, no. 9, pp. 1438-1442, Sep. 2015.

[139] D. L. Donoho, "Superresolution via sparsity constraints", *SIAM J. Math. Anal.*, vol. 23, no. 5, pp. 1309-1331, Sep. 1992.

[140] D. L. Donoho, "Statistical Estimation and Optimal Recovery", *Ann. Statist.*, vol. 22, no. 1, pp. 238-270, 1994.

[141] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution", *Commun. Pure Appl. Math.*, vol. 67, no. 6, pp. 906-956, June 2014.

[142] E. J. Candès and C. Fernandez-Granda,"Super-resolution from noisy data", *J. Fourier Anal. Appl.*, vol. 19, no. 6, pp. 1229-1254, 2013.

[143] J. Scarlett and V. Cevher, "Limits on support recovery with probabilistic models: An information-theoretic framework", *IEEE Trans. Inf. Theory*, vol. 63, no. 1, pp. 593-620, Jan. 2017.

[144] M. Akcakaya and V. Tarokh, "Shannon-theoretic limits on noisy compressive sampling", *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 492-504, Jan. 2010.

[145] M. J. Wainwright, "Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting", *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5728-5741, Dec. 2009.

[146] G. Tang and A. Nehorai, "Performance analysis for sparse support recovery", *IEEE Trans. Inf. Theory*, vol. 56, no.3, pp. 1383-1399, March 2010.

[147] Y. Jin and B. D. Rao, "Support Recovery of Sparse Signals in the Presence of Multiple Measurement Vectors", *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 3139-3157, May 2013.

[148] T. Bendory, "Robust Recovery of Positive Stream of Pulses", *IEEE Transactions on Signal Processing*, vol. 65, no. 8, pp. 2114-2122, April 2017.

[149] R. O. Schmidt, "Multiple emitter location and signal parameter estimation", *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276-280, March 1986.

[150] K. G. Puschmann and F. Kneer, "On super-resolution in astronomical imaging", *Astronomy and Astrophysics*, vol. 436, no. 1, pp. 373-378, 2005.

[151] H. Greenspan, "Super-resolution in medical imaging", *Comput. J.*, vol. 52, no. 1, pp. 43-63, 2009.

[152] C. W. McCutchen, "Superresolution in microscopy and the Abbe resolution limit", *J. Opt. Soc. Am.*, vol. 57, no. 10, pp. 1190-1192, 1967.

[153] R. Heckel, V. I. Morgenshtern and M. Soltanolkotabi, "Super-resolution radar", *Information and Inference: A Journal of the IMA*, vol. 5, no. 1, pp. 22-75, March. 2016.

[154] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A reweighted minimum norm algorithm", *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 600 - 616, March 1997.

[155] Md. M. Hyder and K. Mahata, "An improved smoothed l_0 approximation algorithm for sparse representation" *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2194 - 2205, April 2010.

[156] T. Blumensath and M. E. Davies, "Normalized iterative hard thresholding: guaranteed stability and performance" *IEEE J. Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 298-309, April 2010.

[157] E. J. Candès, M. B. Wakin and S. P. Boyd, "Enhancing sparsity by reweighted l_1 minimization", *J. Fourier Anal. Appl.*, vol. 14, pp. 877-905, 2008.

[158] R. Ge, J . D. Lee and T. Ma, "Matrix completion has no spurious local minima", in Advances in Neural Information Processing Systems (NIPS), 2016.

[159] C. Ma, K. Wang, Y. Chi and Y. Chen,"Implicit Regularization in Nonconvex Statistical Estimation: Gradient Descent Converges Linearly for Phase Retrieval, Matrix Completion, and Blind Deconvolution", *arXiv:*1711.10467, 2017.

[160] I. Daubechies, R. Devore, M. Fornasier and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery" *Communications on Pure and Applied Mathematics*, vol. LXIII, 0001-0038, 2010.

[161] M. S. Lobo, M. Fazel and S. Boyd, "Portfolio optimization with linear and fixed transaction costs", *Ann. Oper. Res.*, vol. 152, no. 1, pp. 341-365, July 2007.

[162] K. Mohan and M. Fazel, "Iterative Reweighted Algorithms for Matrix Rank Minimization", *Journal of Machine Learning Research*, vol. 13, pp. 3441-3473, Nov. 2012.

[163] S. Foucart and M. Lai, "Sparsest solutions of underdetermined linear systems via l_p-minimization for $0 < q \leq 1$, *Appl. Comput. Harmon. Anal.*, vol. 26, pp. 395-407, 2009.

[164] E. J. Candès, X. Li and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: theory and algorithms", *IEEE Trans. Info. Theory*, vol. 61, no. 4, pp. 1985-2007, 2015.

[165] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization", *IEEE Signal Processing Letters*, vol. 14, no. 10, pp. 707-710, Oct. 2007.

[166] M. Fornasier, H. Rauhut and R. Ward, "Low-rank matrix recovery via iteratively reweighted least squares minimization", *SIAM J. Optim.*, vol. 21, no. 4, pp. 1614-1640, 2011.

[167] A. Koochakzadeh and P. Pal, "Cramér-Rao bounds for underdetermined source localization", *IEEE Signal Processing Letters*, vol. 23, no. 7, pp. 919-923, July 2016.

[168] Ali Koochakzadeh, H. Qiao and P. Pal, "On Fundamental Limits of Joint Sparse Support Recovery Using Certain Correlation Priors", *IEEE Transactions on Signal Processing*, vol. 66, no. 17, pp. 4612-4625, Sep. 2018.

[169] C. L. Liu and P. P. Vaidyanathan, "Cramér-Rao bounds for coprime and other sparse arrays, which find more sources than sensors", *Digit. Signal Process.*, vol. 61, pp. 43-61, 2017.

[170] K. Lee, Y. Bresler and M. Junge, "Subspace Methods for Joint Sparse Recovery", *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3613-3641, June 2012.

[171] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces", *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302-5316, Nov. 2011.

[172] M. E. Davies and Y. C. Eldar, "Rank Awareness in Joint Sparse Recovery", *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1135-1146, Feb. 2012.

[173] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors", *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4634-4643, Dec. 2006.

[174] S. F. Cotter, B. D. Rao, E. Kjersti and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors", *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2477 - 2488, July 2005.

[175] O. Solomon, Y. C. Eldar, M. Mutzafi and M. Segev, "SPARCOM: Sparsity Based Super-Resolution Correlation Microscopy", *SIAM J. Imaging Sci.*, vol. 12, no. 1, pp. 392-419, 2019.

[176] O. Solomon, M. Mutzafi, M. Segev and Y. C. Eldar, "Sparsity-based super-resolution microscopy from correction information", *Optics Express*, vol. 26, no. 14, pp. 18238-18269, 2018.

[177] P. Hyberg, M. Jansson and B. Ottersten, "Array Interpolation and Bias Reduction", *IEEE Transactions on Signal Processing*, vol. 52, no. 10, pp. 2711-2720, Oct. 2004.

[178] P. Hyberg, M. Jansson and B. Ottersten, "Array Interpolation and DOA MSE Reduction", *IEEE Transactions on Signal Processing*, vol. 53, no. 12, pp. 4464-4471, Dec. 2005.

[179] A. Koochakzadeh and P. Pal, "Cramér-Rao bounds for underdetermined source localization", *IEEE Signal Processing Letters*, vol. 23, no. 7, pp. 919-923, July 2016.

[180] J. Zheng and M. Kaveh, "Sparse spatial spectral estimation: a covariance fitting algorithm, performance and regularization", *IEEE Transactions on Signal Processing*, vol. 61, no. 11, pp. 2767-2777, June 2013.

[181] D. P. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem", *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3704-3716, July 2007.

[182] Z. Zhang and B. D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning", *IEEE J. Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 912-926, Sept. 2011.

[183] Y. D. Zhang, M. G. Amin and B. Himed, "Sparsity-based DOA estimation using co-prime arrays", in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013.

[184] K. Liu and Y. D. Zhang, "Coprime array-based DOA estimation in unknown nonuniform noise environment", *Digital Signal Processing*, vol. 79, pp. 66-74, 2018.

[185] M. Guo, Y. D. Zhang and T. Chen, "DOA estimation using compressed sparse array", *IEEE Transactions on Signal Processing*, vol. 66, no. 15, pp. 4133-4146, Aug. 2018.

[186] Sung-En Chiu and B. Rao, "Correlation learning on joint support recovery for more sources than measurements", in *Proc. IEEE 9th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, July 2016.

[187] S. Nannuru, A. Koochakzadeh, Kay L. Gemba, P. Pal and P. Gerstoft, "Sparse Bayesian learning for beamforming using sparse linear arrays", *J. Acoust. Soc. Am.*, vol. 144, no. 5, pp. 2719-2729, Nov. 2018.

[188] D. P. Wipf, J. P. Owen, H. T. Attias, K. Sekihara and S. S. Nagarajan, "Robust Bayesian estimation of the location, orientation, and time course of multiple correlated neural sources using MEG", *NeuroImage*, vol. 49, no. 1, pp. 641-655, Jan. 2010.

[189] Y. Zhang, G. Zhou , J. Jin , Q. Zhao, X. Wang and A. Cichocki, "Sparse Bayesian Classification of EEG for BrainComputer Interface", *IEEE Trans. Neural Networks and Learning Systems*, vol. 27, no. 11, pp. 2256-2267, Nov. 2016.

[190] M. G. Amin, X. Wang, Y. D. Zhang, F. Ahmad and E. Aboutanios, "Sparse Arrays and Sampling for Interference Mitigation and DOA Estimation in GNSS", *Proceedings of the IEEE*, vol. 104, no. 6, pp. 1302-1317, June 2016.

[191] G. Qin, M. G. Amin and Y. D. Zhang, "DOA estimation Exploiting Sparse Array Motions, *IEEE Transactions on Signal Processing*, vol. 67, no. 11, pp. 3013 - 3027, June. 2019.

[192] M. Guo, Y. D. Zhang and T. Chen, "DOA estimation using compressed sparse array, *IEEE Transactions on Signal Processing*, vol. 66, no. 15, pp. 4133-4146, Aug. 2018.

[193] J. Romberg, "Imaging via compressive sampling", *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 14-20, March 2008.

[194] O. Bar-Ilan and Y. C. Eldar, "Sub-Nyquist Radar via Doppler Focusing", *IEEE Transactions on Signal Processing*, vol. 62, no. 7, pp. 1796-1811, April 2014.

[195] W. U. Bajwa, J. Haput, A. M. Sayeed and R. Novak, "Compressed channel sensing: A new approach to estimating sparse multipath channels", *Proceedings of IEEE*, vol. 98, no. 6, pp. 1058-1076, June 2010.

[196] M. Lustig, D. L. Donoho, J. M. Santos and J. M. Pauly, "Compressed sensing MRI", *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 72-82, March 2008.

[197] N. Hu, B. Sun, Y. Zhang, J. Dai, J. Wang and C. Chang, "Underdetermined DOA Estimation Method for Wideband Signals Using Joint Nonnegative Sparse Bayesian Learning", *IEEE Signal Processing Letters*, vol. 24, no. 5, pp. 535-539, May 2017.

[198] Chun-Lin Liu and P. P. Vaidyanathan, "Super nested arrays: linear sparse arrays with reduced mutual coupling-Part I: Fundamentals", *IEEE Transactions on Signal Processing*, vol. 64, no. 15, pp. 3997-4012, Aug. 2016

[199] S. Qin, Y. D. Zhang and M. G. Amin, "Generalized coprime array configurations for direction-of-arrival estimation", *IEEE Transactions on Signal Processing*, vol. 63, no. 6, pp. 1377-1390, March 2015.

[200] D. Malioutov, M. Cetin, and A. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays", *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010-3022, Aug. 2005.

[201] J. A. Tropp, "Algorithms for simultaneous sparse approximation Part II: convex relaxation", *Signal Processing*, vol. 86, pp. 589-602, 2006.

[202] C. Aksoylar, G. K. Atia and V. Saligrama, "Sparse signal processing with linear and nonlinear observations: a unified Shannon-Theoretic approach", *IEEE Trans. Inf. Theory*, vol. 63, no. 2, pp. 749-776, Feb. 2017.

[203] S. Sedighi, R. B. S. Mysore, S. Maleki and B. Ottersten, "Consistent least squares estimator for co-array-based doa estimation", in *Proc. IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, July 2018.

[204] A. Moffet, "Minimum-redundancy linear arrays", *IEEE Trans. Antennas and Propagation*, vol. 16, no. 2, pp. 172-175, March 1968.

[205] J. Kim, O. Lee and J. Ye, "Compressive MUSIC: Revisiting the link between compressive sensing and array signal processing", *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 278-301, Jan. 2012.

[206] Q. Denoyelle, V. Duval and G. Peyré, "Support Recovery for Sparse Super-Resolution of Positive Measures", *J. Fourier Anal. Appl.*, vol. 23, no. 5, pp. 1153-1194, 2017.

[207] D. L. Donoho and J. Tanner, "Sparse nonnegative solution of underdetermined linear equations by linear programming", *Proc. Natl. Acad. Sci.*, vol 102, no. 27, pp. 9446-9451, July 2005.

[208] M. Wang, W. Xu and A. Tang, "A Unique 'Nonnegative' Solution to an Underdetermined System: From Vectors to Matrices", *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 1007-1016, Mar. 2011.

[209] Y. Xie, S. Gu, Y. Liu, W. Zuo, W. Zhang and L. Zhang, "Weighted schatten p-norm minimization for image denoising and Background subtraction", *IEEE Transactions on Signal Processing*, vol. 25, no. 10, pp. 4842-4857, Oct. 2016.

[210] M. E. Davies and R. Gribonval, "Restricted isometry constants where l^p sparse recovery can fail for $0 < p \leq 1$", *IEEE Trans. Info. Theory*, vol. 55, no. 5, pp. 2203-2214, Feb. 2009.

[211] W. Rudin, "Principles of mathematical analysis", McGraw-Hill Education, 1976.

[212] L. Demanet and P. Hand," Stable optimizationless recovery from phaseless linear measurements", *Journal of Fourier Analysis and Applications*, vol. 20, no. 1, pp. 199-221, 2014.

[213] S. Rangan, P. Schniter and A. K. Fletcher, "Vector Approximate Message Passing", *IEEE Trans. Inf. Theory*, to appear.

[214] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing", *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18914-18919, Nov. 2009.

[215] W. Li, W. Liao and A. Fannjiang, "Super-resolution limit of the ESPRIT algorithm", *arXiv: 1905.03782*, 2019.

[216] M. Born and E. Wolf, "Principles of optics: electromagnetic theory of propagation, interference and diffraction of light", CUP Archive, 2000.

[217] E. Abbe, "Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung," 9(1), pp. 413-418, 1873.

[218] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz and H. F. Hess, "Imaging intracellular fluorescent proteins at nanometer resolution," *Science*, vol. 313, no. 5793, pp. 1642-1645, 2006.

[219] M. J. Rust, M. Bates and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical resconstruction microscopy (STORM)," *Nature Methods*, vol. 3, no. 10, pp. 793-795, 2006.

[220] O. Solomon, M. Mutzafi, M. Segev and Y. C. Eldar, "Sparsity based super-resolution optical imaging using correlation information", *Proc. of IEEE Intl. Conf. on Acoustic, Speech and Sig. Proc. (ICASSP)*, pp. 6215-6219, New Orleans, LA, 2017.

[221] T. Dertinger, R. Colyer, R. Vogel, J. Enderlein and S. Weiss, "Achieving increased resolution and more pixels with Superresolution Optical Fluctuation Imaging (SOFI)," *Optics express*, vol. 18, no. 18, pp. 18875-18885, 2010.

[222] T. Dertinger, R. Colyer, G. lyer, S. Weiss and J. Enderlein,"Fast, background-free, 3D super-resolution optical fluctuation imaging (SOFI)", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 52, pp. 22287-22292, 2009.

[223] T.-J. Shan, M. Wax and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 33, no. 4, pp. 806-811, Aug. 1985.

[224] Y.-M.Chen, "On spatial smoothing for two-dimensional direction-of-arrival estimation of coherent signals", *IEEE Transactions on Signal Processing*, vol. 45, no. 7, pp. 1689-1696, Jul. 1997.

[225] Y. I. Abramovich, N. K. Spencer and A. Y. Gorokhov, "Positive-Definite Toepitz Completion in DOA Estimation for Nonuniform Linear Antenna Arrays–Part II: Partially Augmentable Arrays", *IEEE Transactions on Signal Processing*, vol. 47, no. 6, pp. 1502-1521, June 1999.

[226] E. BouDaher, F. Ahmad and M. G. Amin, "Sparsity-Based Extrapolation for Direction-of-Arrival Estimation Using Co-Prime Arrays", *SPIE Commercial+ Scientific Sensing and Imaging*, 98570L-98570L-6, 2016.

[227] T. E. Tuncer, T. K. Yasar and B. Friedlander, "Direction of arrival estimation for nonuniform linear arrays by using array interpolation," *Radio Science*, vol. 42, no. 4, 2007.

[228] C. L. Liu, P. P. Vaidyanathan and P. Pal, "Coprime Coarray Interpolation for DOA Estimation via Nuclear Norm Minimization", *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2016.

[229] B. Huang, W. Wang, M. Bates, and X. Zhuang, "Threedimensional super-resolution imaging by stochastic optical reconstruction microscopy", *Science*, vol. 319, no. 5864, pp. 810-813, 2008.

[230] S. Quirin, S. Pavani, and R. Piestun, "Optimal 3d singlemolecule localization for superresolution microscopy with aberrations and engineered point spread functions", *Proceedings of the National Academy of Sciences,* vol. 109, no. 3, pp. 675-679, 2012.

[231] R. Heckel, V. Morgenshtern, and M. Soltanolkotabi, "Super-resolution radar", *Information and Inference: A Journal of the IMA*, vol. 5, no. 1, pp. 22-75, Mar. 2016.

[232] X. Qu, M. Mayzel, J.-F. Cai, Z. Chen, V. Orekhov, "Accelerated NMR spectroscopy with low-rank reconstruction", *Angew. Chem., Int. Ed.*, vol. 54, no. 3, pp. 852-854, Jan. 2015.

[233] L. Borcea, G. Papanicolaou, C. Tsogka, J. Berryman, "Imaging and time reversal in random media", *Inverse Problems*, vol. 18, no. 5, pp. 1247-1279, Aug. 2002.

[234] A. Ahmed, B. Recht and J. Romberg, "Blind deconvolution using convex programming", *IEEE Trans. Inf. Theory*, vol. 60, no. 3, pp. 1711-1732, Mar. 2014.

[235] M. A. Davenport and J. Romberg, "An Overview of Low-Rank Matrix Recovery From Incomplete Observations", *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 4, pp. 608-622, June 2016.

[236] S. Ling and T. Strohmer, "Self-calibration and biconvex compressive sensing", *Inverse Problems*, vol.31, no. 11, 115002, 2015.

[237] M. Wang, W. Xu and A. Tang, "On the Performance of Sparse Recovery Via l_p-Minimization $(0 \leq p \leq 1)$", *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7255-7278, Nov. 2011.

[238] M. Slawski and M. Hein, "Non-negative least squares for high-dimensional linear models: Consistency and sparse recovery without regularization", *Electronic Journal of Statistics*, vol. 7, pp. 3004-3056, 2013.

[239] D. L. Donoho and J. Tanner, "Counting the Faces of Randomly-Projected Hypercubes and Orthants, with Applications", *Discrete Comput. Geom.*, vol. 43, no. 3, pp. 522-541, April 2010.

[240] A. M. Bruckstein, M. Elad and M. Zibulevsky, "On the Uniqueness of Nonnegative Sparse Solutions to Underdetermined Systems of Equations", *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 4813-4820, Nov. 2008.

[241] H. Kirshner, F. Aguet, D. Sage and M. Unser, "3-D psf fitting for fluorescence microscopy: implementation and localization application", *Journal of microscopy*, vol. 249, no. 1, pp. 13-25, 2013.

[242] Y. Li , K. Lee and Y. Bresler, "Identifiability in Blind Deconvolution With Subspace or Sparsity Constraints", *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 4266-4275, July 2016.

[243] Y. Li, K. Lee and Y. Bresler, "Identifiability in Bilinear Inverse Problems with Applications to Subspace or Sparsity-Constrained Blind Gain and Phase Calibration", *IEEE Trans. Inf. Theory*, vol. 63, no. 2, pp. 822-842, Feb. 2017.

[244] C. Bilen, G. Puy, R. Gribonval and L. Daudet, "Convex optimization approaches for blind sensor calibration using sparsity", *IEEE Transactions on Signal Processing*, vol. 62, no. 18, pp. 4847-4856, Sep. 2014.

[245] Jean-Paul Chilès and P. Delfiner, "Geostatistics-Modeling Spatial Uncertainty", A John Wiley & Sons, 2012.

[246] A. Krause, A. Singh, C. Guestrin, "Near-Optimal Sensor Placements in Gaussian Processes: Theory, Efficient Algorithms and Empirical Studies", *Journal of Machine Learning Research* vol. 9, pp. 235-284, 2008.

[247] A. Krause, C. Guestrin, A. Gupta and J. Kleinberg. "Near-optimal Sensor Placements: Maximizing Information while Minimizing Communication Cost", *Proc. 5th International Conference on Information Processing in Sensor Networks*, ISPN, 2006.

[248] W. F. Caselton and J. V. Zidek, "Optimal monitoring networks designs", *Statistics and Probability Letters*, vol. 2, no. 4, pp. 223 - 227, 1984.

[249] G. Matheron, "Le krigeage universel", Cahiers du Centre de Morphologie Mathematique de Fontainebleau, Fasc. 1, Ecole des Mines de Paris, 1969.

[250] T. Kailath, A. H. Sayed and B. Hassibi, "Linear Estimation", Prentice Hall, 2000.

[251] M. Wei, "The perturbation of consistent least squares", *Linear Algebra and Its Applications*, vol. 112, pp. 231-245, Jan. 1989.

[252] R. Vershynin, "High-Dimensional Probability An Introduction with Applications in Data Science", Cambridge University Press, 2018.

[253] Z. Yang, L. Xie and P. Stoica, "Vandermonde Decomposition of Multilevel Toeplitz Matrices With Application to Multidimensional Super-Resolution", *IEEE Trans. Information Theory*, vol. 62, no. 6, pp. 3685-3701, June 2016.