ABSTRACT

Title of thesis: Nested Sampling and its Applications in Stable Compressive Covariance Estimation and Phase Retrieval with Near-Minimal Measurements

Heng Qiao, Master of Science, 2016

Thesis directed by: Professor Piya Pal
Department of Electrical and Computer Engineering

Compressed covariance sensing using quadratic samplers is gaining increasing interest in recent literature. Covariance matrix often plays the role of a sufficient statistic in many signal and information processing tasks. However, owing to the large dimension of the data, it may become necessary to obtain a compressed sketch of the high dimensional covariance matrix to reduce the associated storage and communication costs. Nested sampling has been proposed in the past as an efficient sub-Nyquist sampling strategy that enables perfect reconstruction of the autocorrelation sequence of Wide-Sense Stationary (WSS) signals, as though it was sampled at the Nyquist rate. The key idea behind nested sampling is to exploit properties of the difference set that naturally arises in quadratic measurement model associated with covariance compression. In this thesis, we will focus on developing novel versions of nested sampling for low rank Toeplitz covariance estimation, and phase retrieval, where the latter problem finds many applications in high resolution optical imaging, X-ray crystallography and molecular imaging.

The problem of low rank compressive Toeplitz covariance estimation is first shown to be fundamentally related to that of line spectrum recovery. In absence if noise, this connection can be exploited to develop a particular kind of sampler called the Generalized Nested Sampler (GNS), that can achieve optimal compression rates. In presence of bounded noise, we develop a regularization-free algorithm that provably leads to stable recovery of the high dimensional Toeplitz matrix from its order-wise minimal sketch acquired using a GNS. Contrary to existing TV-norm and nuclear norm based reconstruction algorithms, our technique does not use any tuning parameters, which can be of great practical value.

The idea of nested sampling idea also finds a surprising use in the problem of phase retrieval, which has been of great interest in recent times for its convex formulation via PhaseLift, By using another modified version of nested sampling, namely the Partial Nested Fourier Sampler (PNFS), we show that with probability one, it is possible to achieve a certain conjectured lower bound on the necessary measurement size. Moreover, for sparse data, an l1 minimization based algorithm is proposed that can lead to stable phase retrieval using order-wise minimal number of measurements.

Nested Sampling and its Applications in Stable Compressive Covariance Estimation and Phase Retrieval with Near-Minimal Measurements

by

Heng Qiao

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

Advisory Committee:
Professor Piya Pal, Chair/Advisor
Professor Rama Chellappa
Professor Min Wu

Table of Contents

List of Tables

List of Figures

List of Abbreviations

$\otimes$	Kronecker product
$Vec(\cdot)$	column-wise vectorization
$O(\cdot)$	$a = O(b) \Leftrightarrow \left\lvert\frac{a}{b}\right\rvert \leqslant c$ for some positive constant c
GNS	Generalized Nested Sampler
PSD	Positive Semidefinite
SDP	Semidefinite Programming
DOF	Degree of Freedom
TV	Total Variation
PNFS	Partial Nested Fourier Sampler
R-PNFS	Randomized Partial Nested Fourier Sampler

Chapter 1: Introduction

Technological advances in device physics, microelectronics, signal processing and computing have contributed to the emergence of "Big Data" over the last decade [1, 2]. Big data is routinely encountered in sensor networks, genomics, remote sensing, imaging, particle physics, web search, social networks, and so forth. This has led to a widening gap between the volume of available data and the capabilities for storing, communicating and processing them efficiently and reliably. Fortunately however, the amount of information buried in the data acquired by sensors in most scenarios is substantially lower compared to the number of raw samples acquired. Additionally, such data can also exhibit well- defined structures, often imposed by the generative physical model. For example, a radar system collects large number of samples of 3 dimensional data across fast time, slow time and space, but the ultimate goal is to find three parameters (range, velocity and direction) of a few targets of interest. This key observation has led to the possibility of novel sampling strategies and design of sensing systems that can directly capture the information using far fewer samples.

1.1 Compressive Sensing and Linear Measurement Model

In recent times, compressed sensing has been popularized as an efficient tool for big-data processing, where the essential idea is to exploit the fact that many real-life signals of interest (such as images and videos) have sparse representations over known bases (i.e. they can be represented using only a few non-zero numbers with respect to known basis vectors) [3, 4, 6]. Such signals can then be sub-sampled using a *random linear projections* for efficient storage and/or communication. 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 [5], the original high dimensional signal can be successfully recovered from its low-dimensional measurement using l_1 minimization [3, 4]. 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).

1.2 Compressive Covariance Sketching and Quadratic Samplers

In many applications, however, the goal is to infer quantities of interest from high dimensional signals (such as recognize and track an object from a video sequence, or identify point sources of radiation from signals collected at an array of imaging sensors). In such cases, it *reconstruction of the original signal may be unnecessary and compression may be attained even without invoking sparsity.* 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 $\mathbf{R_x} = E(\mathbf{x}\mathbf{x}^H)$ 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 $\mathbf{R_x}$ (or its estimate). Instead, if we acquire compressive linear measurements of $\mathbf{x}$ as $\mathbf{y} = \mathbf{A}\mathbf{x}$, the covariance matrix $\mathbf{R_y}$ of $\mathbf{y}$ *now acts as a compressive sketch of* $\mathbf{R_x}$

which can be effectively stored and/or processed. The high dimensional covariance matrix $\mathbf{R_x}$ and its compressive sketch $\mathbf{R_y}$ are related as

$$\mathbf{R_y} = \mathbf{A}\mathbf{R_x}\mathbf{A}^H \tag{1.2}$$

Notice that $\mathbf{R_y}$ and $\mathbf{R_x}$ 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 [7, 8]. 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 [29]) 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 Phase Retrieval and Quadratic Samplers

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

$$y_i = |\langle \mathbf{a}_i, \mathbf{x} \rangle|, i = 1, 2, \cdots, M \tag{1.3}$$

Here $\mathbf{x} \in \mathbb{C}^N$ is the unknown signal of interest and $y_i, 1 \leqslant i \leqslant M$ 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 [86, 59, 90, 61], 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 2). In recent times, there have been attempts at resolving this ambiguity by using sparsity as a prior [89], using coded diffraction masks [80], or using STFT [56]. However, these methods are often sub-optimal in terms of the number of measurements required to ensure perfect reconstruction.

In this thesis, 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

$$y_i^2 = \left(\mathbf{a}_i^T \otimes \mathbf{a}_i^H\right) \mathrm{vec}(\mathbf{x}\mathbf{x}^H) \tag{1.4}$$

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 (1.4) 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.

1.4 Contributions and Organization of The Dissertation

The contributions of this thesis can be summarized as following:

1. **Optimum Low Rank Toeplitz Covariance Compression Using Structured Samplers:** We consider the problem compressing a low rank (say,

with rank r) high dimensional Toeplitz covariance matrix (of size $N \times N$, where N is large), and develop a new structured sampling strategy, namely the Generalized Nested Sampler (GNS), using which the high dimensional matrix can be exactly recovered from its noiseless sketch of size $M \times M$, as long as $M = O(\sqrt{r})$. It can be shown that the size of the sketch is optimal since it attains the degrees of freedom associated with low rank Toeplitz covariance matrix. Our sampling strategy fully exploits the positive semidefiniteness as well as the Toeplitz structure of the covariance matrix, and outperforms random samplers where the number of measurements typically involve an extra polylog(N) factor. We also show that GNS is stable in presence of bounded noise.

2. **Low Complexity and Regularizer-Free Reconstruction:** We also develop a new reconstruction algorithm for recovering high dimensional Toeplitz structured $\mathbf{R_x}$ from its compressed sketch $\mathbf{R_Y}$ acquired using a GNS. Our algorithm enjoys several advantages over state-of-the art nuclear norm based recovery algorithms. Firstly, our algorithm exploits the parametric representation of Toeplitz PSD matrices and this leads to a significantly lower computational complexity over nuclear norm minimization techniques. Secondly, our algorithm exploits the positive semidefinite property of covariance matrices to proposed a novel pre-processing step using constrained least square denoising, that completely avoids the need for any regularization parameter. This offers great advantage over algorithms that use nuclear-norm regulariz-

ers since the choice of regularization parameter can be quite sensitive to our knowledge of the noise power. We analytically characterize the performance of our algorithm and show that it leads to stable covariance estimation from its order-wise minimal sketch.

We report out results on low rank Toeplitz covariance compression and reconstruction in Chapter 2.

3. **Fourier Phase Retrieval with near-minimal number of measurements:** In 2, we consider the problem of Fourier-based phase retrieval and design a new sampling technique, namely Partial Nested Fourier Samplers (PNFS) that can overcome the inherent ambiguity of standard Fourier sampling. In absence of noise, we show that our sampler can recover almost all N dimensional image using only $4N - 5$ intensity measurements, which has an interesting connection with a certain existing conjecture regarding the number of measurements necessary for exact phase retrieval. When the underlying signal is sparse, the PNFS can be modified and combined with l_1 minimization algorithm to ensure recovery of sparse signals from only $O(slog(N/s)$ intensity measurements, which has not been so far possible for Fourier based phase retrieval.

The content of this thesis is derived from several published and submitted conference and journal papers [13, 14, 17, 75, 76, 15, 16].

Chapter 2: Low Rank Toeplitz Covariance Estimation

In this chapter, we will focus on compressing and recovering Toeplitz structured covariance matrix with near-minimal number of measurements. Since Toeplitz matrix consists of repeated entries, we can apply nested array idea [29] to extract distinct entries by designing particular sampler. We will discuss the idea of nested array sampling and illustrate it with some simple examples. Then we will define a particular kind of sampler called Generalized Nested Sampler (GNS) that is powerful in compressing Toeplitz covariance which is the focus of next chapter.

2.1 Overview and Prior Art

Estimation of second-order statistics (or correlation) of high-dimensional data plays a central role in modern statistical analysis and information processing. The covariance often acts as a sufficient statistic in many signal processing problems [18, 19]. The covariance matrix also provides a compact summary of a large dataset, and is used for dimension reduction. A popular example is that of principal component analysis [20, 21] 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 important role in many applications re-

lated to classification of Gaussian data and establishing independence relations in exploratory data analysis and testing [22].

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* [23, 24, 25, 26, 27], is receiving increasing attention, where the goal is to compress and reconstruct the high dimensional covariance matrix using so-called *quadratic 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) [23, 25].

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 [23, 24, 25, 26, 8]. In [8], a high dimensional covariance matrix $\mathbf{\Sigma} \in \mathbb{R}^{N,N}$ is sketched using quadratic measurements where $\mathbf{\Sigma}$ is assumed to exhibit distributed sparsity. The required sample size for compressing sparse covariacne 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 [27, 26, 24, 25, 28, 29]. The common

theme in this body of work is the use of deterministic sub-Nyquist samplers (often inspired from the idea of difference-sets [12]) 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 [30], considers a cyclostationary signal model for which the number of measurements is shown to be $O(\sqrt{N})$. In [31], the authors consider the estimation of Toeplitz covariance matrix via Maximum Likelihood methods and the results hold only in asymptotic sense and no stability result is discussed for finite sample case. These results in literature do not consider low rank Toeplitz matrices and the reconstruction framework is quite different from Least-Square (LS) based approach proposed in this thesis.

In this thesis, besides Toeplitz structure, we also exploit low rank of the covariance matrix that allows further compression over what is possible by exploiting only the Toeplitz structure. Low rank positive semidefinite (PSD) Toeplitz matrices arise in applications such as direction finding in radar and astronomical imaging systems, where the low rank is typically attributed to the presence of only a few sources or scatterers compared to the number of physical sensors [32]. Another closely related application is that of LTI dynamical system realization from convex time domain constraints on its impulse response. In such a case, the problem reduces to finding a low rank Hankel structured matrix (very closely related to the Toeplitz structure) describing the LTI system [33].

Much of the existing work on low rank matrix recovery from compressive measurements [34, 35, 36, 37, 38, 39] uses random sampling along with a nuclear norm minimization heuristic to establish performance guarantees. The number of mea-

surements in such cases is shown to be of the general form $O(r^{\alpha}N^{\beta}\text{polylog}(N)), \alpha, \beta > 0$. These methods however do not consider compression of low rank Toeplitz matrices. The authors in [23] provide a unified analysis for compressing and recovering low rank Toeplitz covariance matrix (again, using a nuclear norm heuristic) and shows that $O(r\text{polylog}N)$ measurements are sufficient for compressing N-dimensional Toeplitz covariance matrices with rank r. Recently, in [40], the authors consider the problem of line spectrum estimation from multiple measurement vector models (MMV) compressed using deterministic sparse samplers, and propose a nuclear norm minimization technique to recover the frequencies. However, the size of the compressed covariance matrix (obtained from the MMV model) in [40] is $O(\sqrt{N}) \times O(\sqrt{N})$ since the sampling scheme does not exploit low rank. Furthermore the recovery guarantees cannot be easily extended to the case when the compressed sketch is of size $O(\sqrt{r}) \times O(\sqrt{r}), r \ll N$.

Our work stands in sharp contrast to random sampling based approaches since we use a *deterministic structured sampler*. Hence, we cannot use the existing tools for analyzing the performance of nuclear norm minimization algorithms (which heavily reply on random sampling for showing existence of necessary dual certificates, or proving RIP of suitable sampling operators). For compressing low rank PSD Toeplitz matrices, we use a newly proposed deterministic sampling scheme called Generalized Nested Sampling (GNS) [13, 14]. Compared with existing sparse ruler type samplers [24, 25, 28], GNS provides a closed-form expression for the sampling matrix for any dimension. Moreover, unlike [24, 25, 28, 40], the size of the compressed sketch is $O(\sqrt{r}) \times O(\sqrt{r})$. The proposed reconstruction technique is also very different from

existing body of work in low rank matrix recovery. We transform the problem of low rank Toeplitz matrix recovery to that of line spectrum estimation, and use the recently developed analysis tools in frequency domain [41, 42, 39] for establishing performance guarantees for noiseless as well as noisy recovery. As a consequence of the positive semidefinite property of the covariance matrix, our analysis framework can avoid the need for a separation condition on the true frequencies in noiseless case, which is a central assumption in [42, 41, 43]. The fact that separation condition can be avoided for positive sources has been discussed in many recent works [43, 44, 45, 46, 47]. In this paper, we propose a parameter-free algorithm based on LS-denoising and prove the proposed algorithm is stable if separation condition on true frequencies is satisfied as in literature.

2.2 Nested Array Sampling and Generalized Nested Sampler

2.2.1 Difference Set

The key idea of nested array sampling is the use of difference set brought by the quadratic measurement model. Consider the general quadratic model

$$\mathbf{Y} = \mathbf{A}\mathbf{X}\mathbf{B}^T \tag{2.1}$$

where $\mathbf{X} \in \mathbb{R}^{N\times N}$, $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{M\times N}$. We have

$$[\mathbf{Y}]_{i,j} = \sum_{1\leqslant m,n\leqslant N} [\mathbf{A}]_{i,m}[\mathbf{X}]_{m,n}[\mathbf{B}]_{j,n} \tag{2.2}$$

The difficulty of quadratic model is that all entries of $\mathbf{X}$ are coupled together in each measurement $[\mathbf{Y}]_{i,j}$. One simple way to fix this is to design $\mathbf{A}, \mathbf{B}$ such that there is

one-to-one mapping from $\mathbf{Y}$ to $\mathbf{X}$. In another word, we hope following relationship holds

$$[\mathbf{Y}]_{i,j} = [\mathbf{A}]_{i,f(i)}[\mathbf{X}]_{f(i),g(j)}[\mathbf{B}]_{j,g(j)} \tag{2.3}$$

where $f(\cdot), g(\cdot)$ are index mappings specified by the samplers design. In practice, we may choose $\mathbf{B} \equiv \mathbf{A}$, then the entry of $\mathbf{X}$ selected in $[\mathbf{Y}]_{i,j}$ is $[\mathbf{X}]_{f(i),f(j)}$. When $\mathbf{X}$ is Toeplitz structured (see Chapter 2), the entry selected will be determined by the difference $f(i) - f(j)$.

However, since the range of i, j are both up-limited by the sample size M, the difference $f(i) - f(j)$ will also be bounded from both sides as we require $f(\cdot)$ to be a one-to-one mapping function. Then we naturally hope that the difference set $\{f(i) - f(j)\}$ will span as large consecutive range as possible for $1 \leqslant i, j \leqslant M$ and consequently more entries of $\mathbf{X}$ can be sampled.

The study of difference set can be dated back to early works [12]. It has been proved that to cover the range $[1, N]$, we need at least $O(\sqrt{N})$ integers. As a simple example, consider an index set $\mathcal{N} = \{1, 2, 3, 4\}$, its difference set $D_{\mathcal{N}}$ is given by

$$D_{\mathcal{N}} = \{-3, -2, -1, 0, 1, 2, 3\} \tag{2.4}$$

With the number of integers fixed, consider another index set $\mathcal{N}' = \{1, 2, 3, 6\}$, its difference set is following

$$D_{\mathcal{N}'} = \{-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5\} \tag{2.5}$$

Obviously, $\mathcal{N}'$ is better than $\mathcal{N}$ in the sense of representing larger consecutive integer range. So, it is preferable to define mapping

$$f(i) \Leftrightarrow [\mathcal{N}']_i \tag{2.6}$$

which can sample more entries of $\mathbf{X}$ for $1 \leqslant i, j \leqslant 4$.

To the best of our knowledge, there is no deterministic way to construct optimal index set $\mathcal{N}$ for any given sample size M. What is known in literature is that the optimal rate is $O(\sqrt{N})$. As one of the major contributions of this thesis, we define a particular kind of sampler based on nested array sampling idea which works for almost every dimension N and theoretically achieve $O(\sqrt{N})$ sampling rate.

2.2.2 Generalized Nested Sampler

The Generalized Nested Sampler (GNS) was first introduced in [13] and further developed in [14]. Following [13], 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 $K \geqslant 6$, define $\Theta(K)$ as the maximum integer θ such that*

$$\theta^2 + \theta \leqslant K \tag{2.7}$$

$$\Gamma(K) = 1 + K - \Theta^2(K)$$

A GNS can be defined as a measurement matrix for any integer K as follows:

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

$$[\mathbf{A}^K_{GNS}]i,j = \begin{cases} 1 & \text{if } i = j,\ 1 \leqslant i \leqslant \Gamma(K) \\ 1 & \text{if } j = (i - \Gamma(K))\Theta(K) + i, \\ & \quad \Gamma(K) < i \leqslant M \\ 0 & \text{Elsewhere} \end{cases} \tag{2.8}$$

The key motivation of GNS is to give a unified algorithm for general dimension that can be used to establish a one-to-one mapping as discussed in previous section. The following lemma shows that the definition (2.8) indeed constructs such a mapping.

Lemma 1. *Given an integer $N \geqslant 6$, there exists a set of integers, Ω, of the following form, such that every integer from 0 to $N-1$ can be expressed as a difference of two elements in Ω.*

$$\begin{aligned} \Omega = \{&1, 2, \cdots, \Gamma(N), \\ &\Gamma(N) + \Delta(N), \Gamma(N) + 2\Delta(N), \cdots, \Gamma(N) + n\Delta(N)\} \end{aligned} \tag{2.9}$$

where $\Delta(N) = \Theta(N) + 1$ and $n = \Theta(N) - 1$

Proof. First noting that for $N \geqslant 6$, $\Theta(N) \geqslant 2$ and $\Gamma(N) + n\Delta = \Gamma(N) + \Theta(N)^2 - 1 = N$ by (2.8). Since $\Gamma(N) \geqslant \Delta(N) = \Theta(N) + 1$, we can express any integer ζ between 0 and $N-1$, in the form $\zeta = \Gamma(N) + l\Delta(N) - m$ where $0 \leqslant l \leqslant n$ and

$0 \leqslant m \leqslant \Gamma(N)$. If $m = 0$, we consider l in the range $0 \leqslant l \leqslant n-1$ and express ζ as $\zeta = \Gamma(N) + (l+1)\Delta(N) - \Delta(N)$ where $\Gamma(N) + (l+1)\Delta(N)$ and $\Delta(N)$ are in Ω. If $1 \leqslant m \leqslant \Gamma(N)$, noting that $\Gamma(N) + l\Delta(N)$ and m are in Ω, it can be concluded that ζ can be expressed by difference of two elements in Ω. □

As a simple example, we consider the case $N = 6$, then

$$\mathbf{A}_{GNS}^{N} = \left[\begin{array}{ccc:ccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hdashline 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \tag{2.10}$$

and the associated set Ω is given by

$$\Omega = \{1, 2, 3, 6\} \tag{2.11}$$

which is already discussed in earlier section.

If we use $\mathbf{A}_{GNS}^{N}$ in (2.1), the measurements $\mathbf{Y}$ will contain and only contain all the entries $[\mathbf{X}]_{m,n}$ satisfying $m - n \in D_{\Omega}$. Obviously, a large portion of $\mathbf{X}$ is ignored by using such sparse structured sampling matrix $\mathbf{A}_{GNS}^{N}$ but it may be already sufficient for sampling some highly structured matrix like Toeplitz covariance matrix, which is discussed in detail in later sections.

2.3 Preliminaries for Low Rank Toeplitz Recovery

A first contribution of our work is to show that *certain structured deterministic samplers* can provably lead to the optimum recovery guarantees in terms of the

number of measurements needed to recover the high dimensional Toeplitz matrix. This is possible due to two reasons: (i) Toeplitz matrices (irrespective of rank) are highly structured objects, and deterministic samplers can be designed to completely exploit their structure (ii) Low rank PSD Toeplitz matrices possess additional algebraic properties which the proposed deterministic sampler can exploit. We will first introduce our measurement model and review a key property of low rank Toeplitz matrices that we will exploit throughout the paper.

2.3.1 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 $E(\mathbf{x}_p\mathbf{x}_p^T) \triangleq \mathbf{T} \in \mathbb{R}^{N,N}$. We compressively sample the data using a sampling matrix $\mathbf{A_s} \in \mathbb{R}^{M,N}$, $M \ll N$ 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

$$\mathbf{R_Y} = \mathrm{E}[\mathbf{y}_p\mathbf{y}_p^T] = \mathbf{A_s}\mathbf{T}\mathbf{A_s}^T \tag{2.12}$$

Instead of the larger covariance matrix $\mathbf{T}$, we store and/or transmit the compressed covariance matrix $\mathbf{R_Y} \in \mathbb{R}^{M,M}$. This paper focuses on the special case when the vectors $\mathbf{x}_p$ are wide-sense stationary, whereby its covariance matrix $\mathbf{T} \in \mathbb{R}^{N,N}$ 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*.

2.3.2 Low Rank Toeplitz Matrix and Vandermonde Decomposition Lemma

Our proposed sampling scheme and recovery algorithms are fundamentally based on the famous Caratheodory's theorem [39, 48, 49] that provides an explicit algebraic structure of $\mathbf{T}$ in terms of a Vandermonde matrix:

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

$$\mathbf{T} = \mathbf{V}_N \mathbf{D} \mathbf{V}_N^{\mathrm{H}} \tag{2.13}$$

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

$$[\mathbf{v}_N(f_i)]_k = e^{j2\pi f_i(k-1)} \quad f_i \in (-1/2, 1/2],\ 1 \leqslant k \leqslant N \tag{2.14}$$

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

Remark: 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 (2.13) allows us to a deduce similar factorization for all leading principals of $\mathbf{T}$. In particular, we have the following corollary

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

$$\mathbf{T}_{(n)} = \mathbf{V}_n \mathbf{D} \mathbf{V}_n^{\mathrm{H}} \tag{2.15}$$

where the columns of $\mathbf{V}_n \in \mathbb{C}^{n,r}$ are defined in the same way as (2.14).

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

Corollary 2. *A PSD Toeplitz matrix $\mathbf{T} \in \mathbb{R}^{N,N}$ 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 (2.14).*

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 [42, 41].

- Any (order-wise) optimal sketching method should produce a sketch $\mathbf{R_Y}$ 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.

2.3.3 Application of Generalized Nested Sampler

In earlier sections, we introduce the nested array sampling idea and Generalized Nested Sampler (GNS). Due to the special structure of Toeplitz covariance matrix, GNS is well suitable to Toeplitz matrix sampling.

To illustrate how GNS works, we show a small example. 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

$$\mathbf{A}_{GNS}^{N} = \left[\begin{array}{ccc:ccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hdashline 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \tag{2.16}$$

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

$$\mathbf{R_Y} = \left[\begin{array}{ccc:c} t_0 & t_1 & t_2 & t_5 \\ t_1 & t_0 & t_1 & t_4 \\ t_2 & t_1 & t_0 & t_3 \\ \hdashline t_5 & t_4 & t_3 & t_0 \end{array}\right] \tag{2.17}$$

Then obviously, we can recover $\mathbf{T}$ from observation $\mathbf{R_Y}$. It can be seen that $\mathbf{R_Y}$ is itself structured when $\mathbf{A}_{GNS}^{N}$ is applied. Particularly, the diagonal blocks will also be Toeplitz which is illustrated in Fig 2.1. The result is formalized in Theorem 2.

Theorem 2. *A real symmetric Toeplitz matrix $\mathbf{T} \in R^{N,N}$, $N \geqslant 6$, can be recovered from its compressed quadratic measurement $\mathbf{R_Y} = \mathbf{ATA'}$ where $\mathbf{A} \in R^{M,N}$ is a Generalized Nested Sampling Matrix given by (2.8).*

Proof. Each entry of $\mathbf{R_Y}$ can be expressed as

$$\mathbf{R_Y}_{i,j} = \sum_{p}\sum_{q} \mathbf{A}_{i,p}\mathbf{T}_{p,q}\mathbf{A}_{j,q} = \sum_{p}\sum_{q} \mathbf{A}_{i,p}t_{|p-q|}\mathbf{A}_{j,q} \tag{2.18}$$

From the definition of $\mathbf{A}$, for each i or j, we have one and only one entry $\mathbf{A}_{i,p(i)}$ or $\mathbf{A}_{j,q(j)}$ which is nonzero. Here, we use notations $p(i)$ and $q(j)$ to express the dependency on i and j explicitly.

Then, $\mathbf{R_Y}_{i,j} = t_{|p(i)-q(j)|}$ and $p(i), q(j)$ are both in the set Ω defined in (2.9). From Lemma 1, we know for any $N \geqslant 6$, any integer from 0 to $N-1$ could be expressed by two elements in Ω, equivalently, $p(i)$ and $q(j)$ could be found that any entry t_n in $\omega_{\mathbf{T}}$ can be expressed as $t_n = t_{|p(i)-q(j)|} = \mathbf{R_Y}_{i,j}$. And the Toeplitz matrix $\mathbf{T}$ is exactly recovered in the sense that each entry in $\omega_{\mathbf{T}}$ can be exactly found in the measurement $\mathbf{R_Y}$ where the position of $t_n, 0 \leqslant n \leqslant N-1$ is (i,j) such that $\mathbf{A}_{i,p(i)} = \mathbf{A}_{j,q(j)} = 1$ and $n = |p(i) - q(j)|$. □

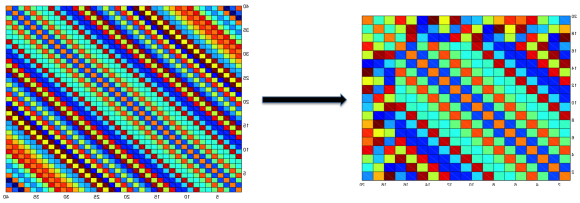

Figure 2.1: The structure of $\mathbf{R_Y}$ when $\mathbf{A}^N_{GNS}$ is applied.

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* $1 \leqslant n \leqslant N$, $\mathbf{T}_{(n)}$ *can be exactly recovered from its compressive sketch* $\mathbf{R_Y} = \mathbf{A}_s \mathbf{T} \mathbf{A}_s^T$ *where the measurement matrix* $\mathbf{A}_s \in \mathbb{R}^{\Gamma(n)+\Theta(n)-1,N}$ *is given by* $\mathbf{A_s}$

$$\mathbf{A_s} = [\mathbf{A}^n_{GNS}, \mathbf{0}] \tag{2.19}$$

where $\mathbf{A}^n_{GNS} \in \mathbb{R}^{\Gamma(n)+\Theta(n)-1,n}$ *is a GNS defined as (2.8).*

2.4 Near Optimal Compression and Recovery of Low-Rank Toeplitz Matrices without Noise

The Vandermonde decomposition lemma dictates that a rank r PSD Toeplitz matrix can be compressed by simply retaining its $n \times n$ principal minor $\mathbf{T}_{(n)}$ where $n = O(r)$. However, $\mathbf{T}_{(n)}$, being a real-valued Toeplitz matrix, contains only n distinct entries. This leads to the possibility of further compressing $\mathbf{T}_{(n)}$ using a suitable sampler. The possibility of compressing and reconstructing a $n \times n$ Toeplitz matrix simply by exploiting the redundancies in its entries has been addressed in [24, 25, 26, 27] where the sampling matrix $\mathbf{A}_s$ is constructed using a minimum redundancy sampler or a sparse ruler [12]. The size of the optimally compressed covariance matrix is $O(\sqrt{n}) \times O(\sqrt{n})$ and it retains all n distinct entries of $\mathbf{T}_{(n)}$ from which $\mathbf{T}_{(n)}$ can be perfectly reconstructed in absence of noise. 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) [13, 14] 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 [25], 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 a highly sparse $\mathbf{A}_s$ which can require less storage space and allow faster computations.

For noiseless case, there are two different ways to recover $\mathbf{T}$ exactly. One way is to use the idea of linear prediction [50] which is developed in [14] and the other is more based on the Vandermonde decomposition of $\mathbf{T}$ or equivalently the associated frequencies and amplitudes. We will cover both methods in following sections.

2.5 Sampling and Reconstruction Scheme via Vandermonde Decomposition

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 [24, 25].

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

$$\mathbf{y_p} = \mathbf{A}_s\mathbf{x}_p, \quad \mathbf{A}_s = [\mathbf{A}_{GNS}^{r+q}, \mathbf{0}] \tag{2.20}$$

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

$\mathbf{R_Y} \in \mathbb{R}^{\Gamma(r+q)+\Theta(r+q)-1,\Gamma(r+q)+\Theta(r+q)-1}$of $\mathbf{T}$ as

$$\mathbf{R_Y} \triangleq E(\mathbf{y_p}\mathbf{y_p}^T) = \mathbf{A}_s\mathbf{T}\mathbf{A}_s^T$$

From the structure of $\mathbf{A}_s$ in (2.20), it can be readily observed that

$$\mathbf{R_Y} = \mathbf{A}_{GNS}^{r+q}\mathbf{T}_{(r+q)}(\mathbf{A}_{GNS}^{r+q})^T \tag{2.21}$$

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

   (a) Recover $\mathbf{T}_{(r+q)}$ from $\mathbf{R_Y}$. 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 $q \geqslant 1$, let $\{f_i, d_i\}, i = 1, 2, \cdots, r$ be the parameters describing its parametric decomposition (2.15). Recover $\{f_i, d_i\}$ using MUSIC and least-square (LS) according to (2.15).

   (c) Given $\{f_i, d_i\}$, recover $\mathbf{T}$ using its Vandermonde decomposition (2.13).

Fig. 2.2 shows the pictorial depiction of the end-to-end compression and reconstruction system.

The proposed scheme enjoys several advantages, such as requiring the order-wise optimum number of measurements to compress a rank r Toeplitz matrix, and employing a lower complexity reconstruction procedure compared to existing methods, which can be especially attractive for high dimensional (large N) problems. We next elaborate on these points.

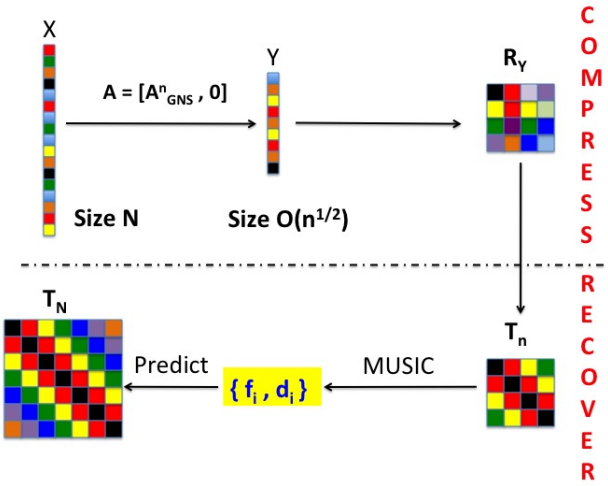

Figure 2.2: GNS based sampling and reconstruction of low rank PSD Toeplitz matrix.

2.5.1 Optimum Compression

GNS or sparse-ruler ased compression strategy produces a sketch $\mathbf{R_Y}$ of size $(\Gamma(r+1)+\Theta(r+1)-1)\times(\Gamma(r+1)+\Theta(r+1)-1)$. Since $\Gamma(r+1), \Theta(r+1) = O(\sqrt{r})$, we need $O(r)$ numbers to represent/store $\mathbf{R_Y}$. Recall that the degree of freedom of a rank r Toeplitz matrix is at most $2r$. Hence, the GNS is *order-wise optimal, requiring only* $O(r)$ *measurements to compress* $\mathbf{T}$. The same comment applied to sparse rulers. Another crucial property of GNS is that the size of the compressed matrix $\mathbf{R_Y}$ is *independent of the ambient high dimension* N *of* $\mathbf{T}$. This stands in sharp contrast to the random sampling based compression scheme suggested in [23] where the total number of measurements in the compressed sketch is of the form $O(r\text{poly}\log N)$. Therefore there is a logarithmic factor depending on N, which shows that random sampling is not order-wise optimal, and cannot attain the DOF

of $\mathbf{T}$. This is because random sampling is not *structure-aware* and hence cannot fully exploit the inherent redundancies among the entries of the Toeplitz matrix. The compressed sketch produced by the GNS however, attains the DOF (upto a constant) by fully exploiting the structure of $\mathbf{T}$ and its size *does not scale* with N. Hence, we can compress Toeplitz matrices of arbitrarily large size N but fixed rank r, using a sketch of constant (with respect to N) size of $O(r)$.

2.5.2 Low-complexity Reconstruction and Power of Prediction

Given the sketch $\mathbf{R}_{\mathbf{Y}}$, the reconstruction scheme proceeds in two stages: it first extracts the parameters $\{f_i, d_i\}$ from $\mathbf{T}_{(r+1)}$ recovered from $\mathbf{R}_{\mathbf{Y}}$. Then, it *predicts the remaining entries of* $\mathbf{T}$ using the Vandermonde decomposition (2.13). This prediction based scheme offers a powerful advantage in terms of reducing the computational complexity of the overall reconstruction. We make this advantage explicit by comparing with the reconstruction algorithm proposed in [23]. The authors in [23] obtain a sketch $\mathbf{Y} = \mathcal{A}(\mathbf{T})$ of $\mathbf{T}$ where $\mathcal{A}(\cdot)$ is a random linear map, and propose to reconstruct $\mathbf{T}$ as

$$\min_{\mathbf{X} \in \mathbb{R}^{N \times N}} \|\mathbf{X}\|_* \quad (N1)$$
$$\text{subject to} \qquad \mathbf{Y} = \mathcal{A}(\mathbf{X}), \quad \mathbf{X} \text{ is PSD Toeplitz}$$

The problem $(N1)$ thus *recovers the entire matrix* $\mathbf{T}$ in a single-shot and the problem size is directly proportional to N. This has the following disadvantages:

1. The complexity of $(N1)$ scales with N and can become prohibitive for very large N.

2. The problem $(N1)$ needs to be *re-solved* for each size N. In other words, even if we recover a principal $\mathbf{T}_{(n)}$ by solving $(N1)$, it cannot be directly used to compute a larger principal $\mathbf{T}_{(n')}$ $(n' > n)$, and one needs to solve $(N1)$ again for the new problem size n'.

The proposed reconstruction method however offers the following advantages that overcome both of these limitations:

1. *Low Complexity Reconstruction*: Just as our compressed sketch is independent of the ambient dimension N, the key step of our reconstruction scheme (where we recover $\{f_i, d_i\}$) also requires a much smaller problem size (compared to $(N1)$) that scales only with r. Thus our reconstruction scheme has an overall complexity that is completely independent of N and this can offer substantial computational saving especially when N is very large.

2. *Power of Prediction:* Given estimates of $f_i, d_i, i = 1, 2, \cdots, r$, our algorithm can recover Toeplitz matrices of any size by using the Vandermonde decomposition (2.13). This is because we can perfectly predict (in absence of noise and other additive errors) the remaining entries of $\mathbf{T}$ once we know its $(r + 1) \times (r + 1)$ leading principal $\mathbf{T}_{(r+1)}$. Therefore, we *solve for* $\{f_i, d_i\}$ *only once* and use them to predict $\mathbf{T}$ of any size. Hence, if we want to recover a larger principal minor $\mathbf{T}_{n'}$ $(n' > n)$, we just use (2.13) to compute $n' - n$ additional values.

2.6 Stable Recovery of Low Rank Toeplitz Covariance in Presence of Bounded Noise: A Parameter Free Approach

In previous section, we review the GNS sampler and show the algorithm for noiseless case. $M = O(\sqrt{r})$ is proved to be sufficient for exact recovery via MUSIC and (2.13). However, if noise is present, the algorithm proposed earlier cannot work for stable estimation. In this section, we will discuss the algorithm with noisy measurements and prove the stability. It will be shown that $M = O(\sqrt{n})$ is still sufficient for stable recovery where $n \ll N$ and the computational complexity is much lower than existent methods in literature. In addition, the algorithm is parameter free in the sense that there is no regularization [42, 40] and noise power needs not to be estimated [23]. Moreover, MUSIC algorithm is still exploited for frequency estimation but the stability analysis is done for finite sample case as compared to ML-based methods [31, 51].

2.6.1 A Parameter Free Algorithm with Noisy Measurements

The results developed so far guarantee perfect recovery of rank r PSD Toeplitz matrix $\mathbf{T}$ of any size from its *noiseless* compressed sketch $\mathbf{R_Y}$ in (2.12) of size $O(\sqrt{r}) \times O(\sqrt{r})$. In practice, we will almost always encounter additive measurement noise. Furthermore, instead of $\mathbf{R_Y}$, we can only compute its estimate $\hat{\mathbf{R}}_\mathbf{Y}$ as $\hat{\mathbf{R}}_\mathbf{Y} = \frac{1}{L}\sum_{p=1}^{L} \mathbf{y_p}\mathbf{y_p}^T$. We can model the effect of both noise and finite sample

averaging as an additive error term $\mathbf{W}$. In other words, we assume to observe

$$\hat{\mathbf{R}}_{\mathbf{Y}} = \mathbf{A}_s \mathbf{T} \mathbf{A}_s^T + \mathbf{W} \tag{2.22}$$

where noise matrix $\mathbf{W}$ is bounded and $\mathbf{A}_s$ constructed by (2.19) for specified $n \ll N$.

The first step of our proposed reconstruction scheme is to estimate $\mathbf{t}_{(n)}$ from $\hat{\mathbf{R}}_{\mathbf{Y}}$. Let t_i denote the ith entry of $\mathbf{t}_{(n)}$. Notice that each $t_i, i = 0, 1, \cdots n-1$ can appear more than once in $\mathbf{A}_s \mathbf{T} \mathbf{A}_s^T$ and there are many ways to retrieve noisy measurements of $\mathbf{t}_{(n)}$. In this paper, we arbitrary pick only one entry from $\hat{\mathbf{R}}_{\mathbf{Y}}$ for each t_i without averaging. By doing this, we only need to store and communicate n samples $\hat{\mathbf{t}}_{(n)}$.

In Sec.2.5, MUSIC algorithm can be used to to recover $\{f_i, d_i\}$ exactly which is based on the low rank fact and perfect knowledge of $\mathbf{t}_{(n)}$ for $n \geqslant r+1$. However, in presence of noise, low rank property is lost in general and (2.13) or (2.15) is not readily applicable. It is well known that Vandermonde decomposition of positive definite Toeplitz covariance is not unique [58]. In this paper, we use the elegant representation of any PSD Toeplitz matrices from [52].

Lemma 2. *[52] A positive semidefinite Toeplitz matrix $\mathbf{T} \in \mathbb{R}^{N,N}$ has the following decomposition :*

$$\mathbf{T} = \mathbf{V}_N \mathbf{D} \mathbf{V}_N^H + \sigma \mathbf{I}_N \tag{2.23}$$

where σ is the smallest singular value of $\mathbf{T}$ and $\mathbf{I}_N$ is identity matrix. $\mathbf{V}_N \in \mathbb{C}^{N,N'} = [\mathbf{v}_N(f_1), \mathbf{v}_N(f_2), \cdots, \mathbf{v}_N(f_{N'})]$ with $N' < N$ and each column $v_N(f_i)$ is defined by (2.14) and $\mathbf{D} = diag(d_1, d_2, \cdots, d_{N'})$ contains the corresponding positive amplitudes.

Remark: For low rank case, (2.23) reduces to (2.13) and $N' = rank(\mathbf{T})$. If $\mathbf{T}$ is positive definite, $N' = N - m_\sigma$ where m_σ is the multiplicity of σ. For both cases, the frequencies are uniquely determined and can be computed by MUSIC or other methods.

In [41, 42], the authors solve the following atomic norm based denoising method for an estimation $\tilde{\mathbf{t}}_{(n)}$ of $\mathbf{t}_{(n)}$ given the noisy measurements $\hat{\mathbf{t}}_{(n)}$.

$$\tilde{\mathbf{t}}_{(n)} = \arg\min_{\mathbf{z}} \frac{1}{2}\|\hat{\mathbf{t}}_{(n)} - \mathbf{z}\|_2^2 + \tau\|\mathbf{z}\|_{\mathcal{A}^n} \tag{2.24}$$

where $\|\cdot\|_{\mathcal{A}^n}$ is atomic norm and τ is a tuning parameter to control the weight of atomic norm and dependent on the noise power. The estimated frequencies are achieved by solving the dual problem.

However, it is not easy to choose parameter τ since noise power may not be known. Inspired by the works in [54, 55], we show in following sections that PSD constraint is enough for stable reconstruction via least square denoising. In addition, we do not need to estimate the unknown noise power and tune the parameter for regularization.

The algorithm for structured compression and stable recovery is summarized in Table 1.The whole algorithm proposed in this paper consists of two separate steps. Firstly, we compress the large covariance $\mathbf{T}$ with GNS for a $n \ll N$ and extract minimum number of noisy measurements to further reduce the storage/transmission load. Secondly, we do LS-denoising and make predictions for a PSD estimation of dimension N via MUSIC and representation (2.23).

Input: Noisy sketch $\hat{\mathbf{R}}_{\mathbf{Y}} \in \mathbb{R}^{n,n}$ observed by (2.22)
Output: Estimates of the entries $t_i, i = 0, 1, \cdots, N-1$ of $\mathbf{T}$.

1. *Redundancy Reduction:* Given the ideal repetitive pattern [13], extract noisy measurements $\hat{\mathbf{t}}_{(n)}$ without averaging that

$$\hat{\mathbf{t}}_{(n)} = \mathbf{t}_{(n)} + \mathbf{w}_{(n)} \tag{2.25}$$

where $\mathbf{t}_{(n)} = [t_0, t_1, \cdots, t_{n-1}]^T$ and bounded noise $\|\mathbf{w}_{(n)}\|_\infty \leqslant \varepsilon$.

2. *Denoising:* Obtain denoised estimation $\mathbf{t}^{\#}_{(n)}$ as

$$\begin{aligned} &\mathbf{t}^{\#}_{(n)} \triangleq \min_{\mathbf{u}\in\mathbb{R}^n} \|\hat{\mathbf{t}}_{(n)} - \mathbf{u}\|_2 \\ &s.t.\, \mathcal{T}(\mathbf{u}) \succeq 0 \end{aligned} \tag{2.26}$$

where $\mathcal{T}(\mathbf{u})$ denotes the real symmetric Toeplitz matrix with first column $\mathbf{u}$.

3. *Parameterization :*

Given $\mathbf{t}^{\#}_{(n)}$, from Lemma 2, we have

$$\mathbf{T}^{\#}_{(n)} = \mathcal{T}(\mathbf{t}^{\#}_{(n)}) = \mathbf{V}_n \mathbf{D}^{\#} \mathbf{V}_n^H + \sigma^{\#} \mathbf{I}_n \tag{2.27}$$

with frequencies $\{f_1^{\#}, f_2^{\#}, \cdots, f_{\tilde{r}}^{\#}\}$ and positive amplitudes $\{d_1^{\#}, d_2^{\#}, \cdots, d_{\tilde{r}}^{\#}\}$.

4. **Prediction:**

- If $\sigma^{\#} = 0$, the recovery of $\mathbf{T}$ is given by $\mathbf{T}^{\#} = \mathbf{V}_N \mathbf{D}^{\#} \mathbf{V}_N^H$ for frequencies $\{f_i^{\#}\}_{i=1}^{\tilde{r}}$ and dimension N.
- If $\sigma^{\#} > 0$, the recovery of $\mathbf{T}$ is given by $\mathbf{T}^{\#} = \mathbf{V}_N \mathbf{D}^{\#} \mathbf{V}_N^H + \sigma^{\#} \mathbf{I}_N$ for frequencies $\{f_i^{\#}\}_{i=1}^{\tilde{r}}$ and dimension N.

Table 2.1: Low Rank PSD Toeplitz Matrix Recovery In Presence of Bounded Noise

Remark: We exploit MUSIC algorithm for obtaining (2.27) and the PSD

constraint in (2.26) is critical for applying Lemma 2.

Remark: If $\sigma^{\#} > 0$ or equivalently, $\mathbf{T}^{\#}_{(n)}$ is positive definite, it should be noted that there are more than one way to extend it into dimension N keeping the PSD property. For example, let Toeplitz matrices $\mathbf{X}, \mathbf{X}_1, \mathbf{X}_2$ defined as

$$\mathbf{X} = \mathbf{I}_2, \ \mathbf{X}_1 = \mathbf{I}_3, \ \mathbf{X}_2 = \mathcal{T}([1, 0, 0.1]^T)$$

It can be easily verified that $\mathbf{X}, \mathbf{X}_1, \mathbf{X}_2$ are all positive definite and $\mathbf{X}_1, \mathbf{X}_2$ are both valid extensions of $\mathbf{X}$. Obviously, $\mathbf{X}_1, \mathbf{X}_2$ have different Carathéodory parameterizations.

On the other hand, if $\sigma^{\#} = 0$, there is only one way to extend it into a PSD Toeplitz matrix of larger dimension N. This is due to the fact that line spectrum WSS process is linearly predictable [14, 50].

2.6.2 Near Optimal Performance in Noiseless Case

In Sec.2.5 as well as in [13, 14], it has been shown that $M = O(\sqrt{r})$ is sufficient for exact recovery. Still, the algorithm in Table 1 achieves near optimal sample complexity $M = O(\sqrt{r})$ and there is no separation condition necessary for noiseless case since only MUSIC is applied. Particularly, we have following result.

Theorem 3. *If* $\mathbf{W} = \mathbf{0}$*, then* $\mathbf{T} = \mathbf{T}^{\#}$ *if and only if* $n \geqslant r + 1$*.*

Proof. If $\mathbf{W} = 0$, then $\tilde{\mathbf{t}}_{(n)} = \mathbf{t}_{(n)}$ and the only solution of (2.26) is $\mathbf{t}^{\#}_{(n)} = \mathbf{t}_{(n)}$. If $n \geqslant r + 1$, the frequencies $\{f_i\}_{i=1}^r$ are uniquely determined by MUSIC via representation (2.15) where $\mathbf{T}_{(n)}$ is rank deficient. The amplitudes $\{d_i\}_{i=1}^r$ can also be uniquely

recovered given $\{f_i\}_{i=1}^r$ by least-square since $\mathbf{V}_n$ is a Vandermonde matrix. If $n \leqslant r$, $\mathbf{T}^{\#}$ will be full rank since $\mathcal{T}(\mathbf{t}_{(n)}^{\#})$ is positive definite and $\sigma^{\#} > 0$. □

2.6.3 Stability Analysis in Presence of Bounded Noise

In this section, we will analyze the stability of the proposed algorithm in Table 1. The analysis is done from the view of frequency domain and the key tools are developed in works [42, 43, 57]. The key idea is based on the fact that PSD Toeplitz matrix is associated with a line spectrum and the estimation error of such line spectrum can be controlled by the denosing step (2.26). And the prediction error is in turn a function of the spectrum estimation error.

First, the estimation error of the denoising phase is given in following lemma. As in noiseless case, there is no separation condition is needed on the frequencies.

Lemma 3. *Let* $\mathbf{t}_{(n)}^{\#}$ *be the solution of (2.26), then we have*

$$\|\mathbf{t}_{(n)}^{\#} - \mathbf{t}_{(n)}\|_2 \leqslant 2\|\mathbf{w}_{(n)}\|_2 \leqslant 2\sqrt{n}\varepsilon \tag{2.28}$$

Proof. The result is straightforward with triangle inequality by noting that true $\mathbf{t}_{(n)}$ is feasible and noise model (2.25). □

In presence of noise, $\mathcal{T}(\mathbf{t}_{(n)}^{\#})$ will in general be full rank and we will always apply (2.23) rather than (2.13). It should be noted that $\{\mathbf{v}_n(f)\}$ defined in Theorem 2 is a frame over $\mathbb{R}^n$ for $f \in (-1/2, 1/2]$. Given $\mathbf{T}, \mathbf{T}^{\#}$, there exist finite measures

$\mu, \mu^{\#}$ such that

$$\mathbf{t}_{(n)} = \int_{-1/2}^{1/2} \mathbf{v}_n(f)\mu(df) \quad \mathbf{t}_{(n)}^{\#} = \int_{-1/2}^{1/2} \mathbf{v}_n(f)\mu^{\#}(df) \tag{2.29}$$

The measures $\mu, \mu^{\#}$ may not be unique and one possible construction is given by

$$\begin{aligned} \mu &= \sum_{i=1}^{r} d_i \delta(f - f_i) \\ \mu^{\#} &= \sum_{i=1}^{\hat{r}} d_i^{\#} \delta(f - f_i^{\#}) + \frac{\sigma^{\#}}{n} \sum_{i=1}^{n} \delta(f - \frac{(i-1)}{n}) \end{aligned} \tag{2.30}$$

where $\delta(\cdot)$ is Dirac function and $\{f_i^{\#}, d_i^{\#}\}_{i=1}^{\hat{r}}$ are frequencies and positive amplitudes associated with $\mathbf{T}^{\#}$ of which $\sigma^{\#}$ is the smallest singular value, all are uniquely defined as in Lemma 2. Both $\mu, \mu^{\#}$ are positive and following analysis is based on the construction in (2.30).

Given the representation in frequency domain, we could apply the tools in [42, 57] which are used to bound the spectrum estimation error. To proceed, we need first introduce some notations. We define a distance function $\rho(\hat{f}_1, \hat{f}_2)$ for distinct frequencies $\hat{f}_1, \hat{f}_2 \in (-1/2, 1/2]$ in a wraparound manner [43]. The difference measure is defined as $\nu = \mu^{\#} - \mu$. Define neighborhood $\mathcal{N}_i$ around each true frequency f_l by $\mathcal{N}_l = \{f \in (-1/2, 1/2] : \rho(f, f_l) \leqslant 0.16/n\}$ and also far region $\mathcal{F} \triangleq (-1/2, 1/2] \backslash \bigcup_{l=1}^{r} \mathcal{N}_l$. Define $\mathcal{P}_{\mathfrak{F}}$ be the projection of any measure onto the true frequency support $\mathfrak{F} = \{f_1, f_2, \cdots, f_r\}$.$\|\cdot\|_{TV}$ is total variation norm [57]. We also

define

$$
\begin{aligned}
I_0^l &:= \left| \int_{\mathcal{N}_l} \nu(df) \right| \\
I_1^l &:= n \left| \int_{\mathcal{N}_l} (f - f_l) \nu(df) \right| \\
I_2^l &:= \frac{n^2}{2} \int_{\mathcal{N}_l} (f - f_l)^2 |\nu|(df) \\
I_i &:= \sum_{l=1}^{r} I_i^l, \text{ for } i = 0, 1, 2
\end{aligned} \tag{2.31}
$$

We need following two lemmas from [42] to bound the spectrum estimation error. Contrary to noiseless case and Lemma 3, we need separation condition on the true frequencies $\{f_i\}_{i=1}^r$.

Lemma 4. *[42, 57] 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

$$
\begin{aligned}
&\|\mathcal{P}_{\mathfrak{F}}(\nu)\|_{TV} = \int_{-1/2}^{1/2} Q(f) \mathcal{P}_{\mathfrak{F}}(\nu)(df) \\
&|Q(f)| \leqslant 1 - \frac{C_a}{2} n^2 (f - f_l)^2 \quad f \in \mathcal{N}_l, 1 \leqslant l \leqslant r \\
&|Q(f)| \leqslant 1 - C_b \quad f \in \mathcal{F} \qquad (2.32) \\
&\left| \int_{-1/2}^{1/2} Q(f) \nu(df) \right| \leqslant \frac{C_c r \xi}{n} \qquad (2.33)
\end{aligned}
$$

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 5. *[42] 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 exist positive constants* $\tilde{c}_1, \tilde{c}_2$ *such that*

$$\begin{aligned} I_0 &\leqslant \tilde{c}_1 \left(\frac{r\xi}{n} + I_2 + \int_{\mathcal{F}} |\nu|(df) \right) \\ I_1 &\leqslant \tilde{c}_2 \left(\frac{r\xi}{n} + I_2 + \int_{\mathcal{F}} |\nu|(df) \right) \end{aligned} \tag{2.34}$$

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

From Lemma 5, I_2 and $\int_{\mathcal{F}} |\nu|(df)$ are key values needing to be bounded which is done in following lemma.

Lemma 6. *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 exist positive constants* c_1, c_2 *such that*

$$I_2 + \int_{\mathcal{F}} |\nu|(df) \leqslant c_1 \left(\frac{c_2 r\xi}{n} + |t_0^{\#} - t_0| \right) \tag{2.35}$$

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

Proof. Since separation condition on $\mathfrak{F}$ is satisfied, from Lemma 4, there exist polynomial $Q(f)$ such that

$$\|\mathcal{P}_{\mathfrak{F}}(\nu)\|_{TV} = \int_{-1/2}^{1/2} Q(f) \mathcal{P}_{\mathfrak{F}}(\nu)(df) \tag{2.36}$$

With triangle inequality and noting $\mathcal{P}_{\mathfrak{F}}$ is projection onto $\mathfrak{F}$, we have

$$\begin{aligned}
\|\mathcal{P}_{\mathfrak{F}}(\nu)\|_{TV} &\leqslant \left|\int_{-1/2}^{1/2} Q(f)\nu(df)\right| + \left|\int_{\mathfrak{F}^c} Q(f)\nu(df)\right| \\
&\leqslant \frac{C_c r\xi}{n} + \sum_{f_l\in\mathfrak{F}} \left|\int_{\mathcal{N}_l\setminus f_l} Q(f)\nu(df)\right| + \left|\int_{\mathcal{F}} Q(f)\nu(df)\right| \\
&\leqslant \frac{C_c r\xi}{n} + \sum_{f_l\in\mathfrak{F}} \left(\int_{\mathcal{N}_l\setminus f_l} |\nu|(df) - C_a I_2^l\right) \\
&+(1-C_b)\int_{\mathcal{F}} |\nu|(df) \\
&= \frac{C_c r\xi}{n} + \|\mathcal{P}_{\mathfrak{F}^c}\|_{TV} - C_a I_2 - C_b \int_{\mathcal{F}} |\nu|(df)
\end{aligned} \tag{2.37}$$

where we use (2.33) for second and third inequality.

Next, noting that

$$\begin{aligned}
&\|\mu^{\#}\|_{TV} = \|\mu+\nu\|_{TV} = \|\mu+\mathcal{P}_{\mathfrak{F}}\|_{TV} + \|\mathcal{P}_{\mathfrak{F}^c}\|_{TV} \\
&\geqslant \|\mu\|_{TV} - \|\mathcal{P}_{\mathfrak{F}}\|_{TV} + \|\mathcal{P}_{\mathfrak{F}^c}\|_{TV}
\end{aligned} \tag{2.38}$$

Then we have

$$C_a I_2 + C_b \int_{\mathcal{F}} |\nu|(df) \leqslant \frac{C_c r\xi}{n} + \|\mu^{\#}\|_{TV} - \|\mu\|_{TV} \tag{2.39}$$

Noting that both $\mu, \mu^{\#}$ constructed from(2.30) are positive measures, we have

$$\begin{aligned}
C_a I_2 + C_b \int_{\mathcal{F}} |\nu|(df) &\leqslant \frac{C_c r\xi}{n} + t_0^{\#} - t_0 \\
I_2 + \int_{\mathcal{F}} |\nu|(df) &\leqslant \frac{1}{\min\{C_a, C_b\}} \left(\frac{C_c r\xi}{n} + |t_0^{\#} - t_0|\right)
\end{aligned} \tag{2.40}$$

and the proof completes. □

Remark. It should be noted that we use the fact both $\mu, \mu^{\#}$ are positive measures in (2.39) and then we can equivalently express the TV norm in terms of

entry value. And this partially explain why PSD constraint is enough and atomic norm is not necessary as needed for general case [42].

Now, we are ready to prove the main result of this paper. In Lemma 3, we have already given the estimation error for observed part which is not based on separation condition. However, for predicted part, we have to use the analysis tools in frequency domain to represent prediction error in terms of spectrum estimation error,i.e, I_0, I_1, I_2 and $\int_{\mathcal{F}} |\nu|(df)$. And then separation condition is needed as in previous lemmas.

Theorem 4. *Suppose the noisy measurements are specified by (2.22) and (2.25) with uniform noise bound ε. Let $\mathbf{T}^{\#}$ be the recovered PSD Toeplitz matrix by solving (2.26) and predicting following the recovery algorithm, then*

$$\frac{1}{n}\|\mathbf{t}^{\#}_{(n)} - \mathbf{t}_{(n)}\|_2 \leqslant \frac{2\varepsilon}{\sqrt{n}} \tag{2.41}$$

If the true frequencies $\{f_1, f_2, \cdots, f_r\}$ satisfy

$$\min_{p \neq q} \rho(f_p, f_q) > 4/n$$

and $n > 256$, we have

$$|t^{\#}_m - t_m| \leqslant \left(\gamma_1 + \frac{\gamma_2 \pi m}{n} + \frac{\gamma_3 \pi^2 m^2}{n^2}\right)\left(\gamma_4 r + \sqrt{n}\right)\varepsilon \tag{2.42}$$

where $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ are positive constants and $n \leqslant m \leqslant N-1$.

Proof. (2.41) is from Lemma 3 and no separation condition is necessary.

For any $n \leqslant m \leqslant N-1$, we have

$$\begin{aligned} |t_m^{\#} - t_m| &= |\int_{-1/2}^{1/2} e^{j2\pi mf}\nu(df)| \\ &\leqslant \int_{\mathcal{F}} |\nu|(df) + \sum_{l=1}^{r} \left| \int_{\mathcal{N}_l} e^{j2\pi mf}\nu(df) \right| \end{aligned} \tag{2.43}$$

where triangle inequality is used.

With Taylor's theorem around each f_l, it follows that

$$\begin{aligned} &\left| \int_{\mathcal{N}_l} e^{j2\pi mf}\nu(df) \right| \leqslant \left| \int_{\mathcal{N}_l} \nu(df) \right| \\ &+2\pi m \left| \int_{\mathcal{N}_l} (f-f_l)\nu(df) \right| + 2\pi^2 m^2 \int_{\mathcal{N}_l} (f-f_l)^2 |\nu|(df) \\ &= I_0^l + \frac{2\pi m}{n} I_1^l + \frac{4\pi^2 m^2}{n^2} I_2^l \end{aligned} \tag{2.44}$$

Then, (2.43) can be simplified as

$$\begin{aligned} &|t_m^{\#} - t_m| \leqslant \int_{\mathcal{F}} |\nu|(df) + I_0 + \frac{2\pi m}{n} I_1 + \frac{4\pi^2 m^2}{n^2} I_2 \\ &\leqslant \left(\tilde{c}_1 + \frac{2\pi m}{n}\tilde{c}_2 + 1 \right) \int_{\mathcal{F}} |\nu|(df) \\ &+ \left(\tilde{c}_1 + \frac{2\pi m}{n}\tilde{c}_2 + \frac{4\pi^2 m^2}{n^2} \right) I_2 \\ &+ (\tilde{c}_1 + \tilde{c}_2 \frac{2\pi m}{n}) \frac{r\xi}{n} \end{aligned} \tag{2.45}$$

where we use Lemma 5. Since $n \leqslant m$, we finally have

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

where we use Lemma 6.

To bound ξ, we apply Cauchy-Schwartz inequality that

$$\xi \leqslant \|\mathbf{v}_n(f)\|_2 \|\mathbf{w}_{(n)}\|_2 = \sqrt{n}\|\mathbf{w}_{(n)}\|_2 \leqslant n\epsilon$$

Finally, we always have $|t_0^{\#} - t_0| \leqslant \|\mathbf{t}_{(n)}^{\#} - \mathbf{t}_{(n)}\|_2 \leqslant 2\sqrt{n}\varepsilon$ and the proof completes. □

The following remarks apply to our derivation of the prediction error:

1. From (2.43), the prediction error is a trigonometric polynomial with finite period due to ν is a finite measure. Consequently, the prediction error will not go to infinity when m approaching infinity and there is a global upper bound.

2. Since the separation condition should be satisfied, we can choose $n > 4r$ where r is the rank of $\mathbf{T}$ as well as the number of true frequencies. And the sample size M can be still $O(\sqrt{r})$ which may be much smaller than N and near optimal.

3. **Separation Condition and Prediction v/s Observed Error:** Notice that we have different bounds for the observation error (given in (2.41)) and prediction error (given by (2.42)). 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 (2.41) does not require a separation condition, whereas it is needed for establishing (2.42). 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 [42, 41, 39, 43, 57] seem to require a "separation condition" for developing error bounds on the estimated frequencies. In [46], similar results as in [42] have been obtained without explicitly assuming separation condition, but requiring the dual polynomial to satisfy a so-called *Quadratic Isolation Condition* (QIC). So far, it is still not clear whether separation condition is necessary for satisfying QIC. Another closely related idea is that of Rayleigh regularity [45] 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. It can be a question of future interest to see if we can derive tighter error bounds for predicted entries using this condition (instead of the current minimum separation criterion). 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 [41, 42, 57], and leave the general case as an open problem for future research.

2.6.4 Comparison with Nuclear Norm Based Recovery using Structured Samplers

Compression and reconstruction of low-rank Toeplitz structured covariance matrices has been recently studied in [40] in the context of line spectrum estimation from MMV models. The signal model introduced in [40] assumes the compressed measurements $\mathbf{y}_p$ to be partial observations of $\mathbf{x}_p$, i.e. $\mathbf{y}_p = \mathbf{x}_{\Xi,p}$ where $\Xi \subset \{0, 1, \cdots, N-1\}$

denotes the rows of $\mathbf{x}_p$ that are sampled. Let $\mathcal{T}(\mathbf{t}_N) \in \mathbb{R}^{N\times N}$ denote the real-valued Toeplitz structured covariance matrix of $\mathbf{x}_p$ where $\mathbf{t}_N$ denotes the first row. The estimated covariance matrix of the compressed data is denoted by $\hat{\mathbf{R}}_{\mathbf{Y}} = \frac{1}{L}\sum_{p=1}^{L} \mathbf{y}_p \mathbf{y}_p^T$ which satisfies $E(\hat{\mathbf{R}}_{\mathbf{Y}}) = \mathcal{P}_{\Xi}(\mathcal{T}(\mathbf{t}_N))$, where $\mathbf{t}_N$ is the first column of $\mathbf{T}$ and $\mathcal{P}_{\Xi}$ is a selection matrix that only preserves the submatrix composed of rows and columns indexed by Ξ. Equivalently, we can also think of $\hat{\mathbf{R}}_{\mathbf{Y}}$ as a corrupted version of $\mathcal{P}_{\Xi}(\mathcal{T}(\mathbf{t}_N))$. The authors in [40] propose to recover $\mathbf{t}_N$ by solving and analyzing the following nuclear norm minimization problem

$$\min_{\mathbf{z}\in\mathbb{R}^N} \frac{1}{2}\|\mathcal{P}_{\Xi}(\mathcal{T}(\mathbf{z})) - \hat{\mathbf{R}}_{\mathbf{Y}}\|_F^2 + \lambda\|\mathcal{T}(\mathbf{z})\|_* \tag{2.46}$$

where $\|\cdot\|_*$ denotes the nuclear norm of a matrix. There are important differences between this approach, and the compression/recovery framework proposed in this paper that are worth highlighting:

1. **Number of Measurements and Complexity:** The performance guarantees for (2.46) are derived under two choices of the sampling set Ξ that correspond to *structured samplers*: (i) full observations, i.e., $\Xi = \{0, 1, \cdots, N-1\}$ and (ii) Ξ corresponding to a sparse ruler. In this case the size of Ξ is $O(\sqrt{N})$. Although the true covariance matrix $\mathcal{T}(\mathbf{t}_N)$ is assumed to be low rank ($r < N$), the size of the compressed covariance sketch $\hat{\mathbf{R}}_{\mathbf{Y}}$ for both choices *has no dependence on the rank* r *and is at least as large as* $O(\sqrt{N}) \times O(\sqrt{N})$. This also implies that the problem size (and the computational complexity) of (2.46) scales with N, since it aims to recover the entire row $\mathbf{t}_N \in \mathbb{R}^N$ of the Toeplitz matrix.

In contrast, the proposed GNS-based compressor produces a sketch of size $O(\sqrt{r}) \times O(\sqrt{r})$ in noiseless case, and the key step in our reconstruction scheme involves solving an SDP (for atomic norm minimization) that scales only with r, and is independent of the ambient dimension N. Hence our approach requires fewer measurements and has lower complexity than that proposed in [40].

2. **Performance Guarantees of observed v/s predicted entries:** Denoting $\mathbf{t}_N^{\#}$ as the solution to (2.46), the error bound $\|\mathbf{t}_N^{\#} - \mathbf{t}_N\|$ in [40] is derived under the aforementioned choices of Ξ, *both of which ensure that all N entries of $\mathbf{t}_N$ are observed at least once.* However, in the proposed approach, we only observe $n = O(r)$ entries of $\mathbf{t}_N$ and predict the remaining ones. If we choose Ξ to be of size $O(\sqrt{r})$ (i.e., we only observe $O(r)$ noise corrupted sentries of $\mathbf{t}_N$), and try to reconstruct the entire vector $\mathbf{t}_N$ using (2.46), we cannot obtain an explicit bound on the error $\|\mathbf{t}_N^{\#} - \mathbf{t}_N\|$ by following the analysis framework developed in [40, 53], since bounds on the error in the unobserved $N-n$ entries of $\mathbf{t}_N$ cannot be easily computed in such a case. The performance guarantees of (2.46) in this setting, is a question of future interest, which may require us to relate the error in the n observed entries to that in remaining $N - n$ entries of $\mathbf{t}_N$ via the parametric representation of $\mathcal{T}(\mathbf{t}_N)$. However, we will numerically compare its performance with the proposed algorithm is previous section.

   On the other hand, we can directly compare the "observed error" $\mathbf{e}_{(n)}$ (i.e.

estimation error in the first n entries of $\mathbf{t}_N$) produced by the LS-denoising approach with the estimation error of the following modified version of (2.46) that recovers only the first n entries of $\mathbf{t}_N$:

$$\min_{\mathbf{z}\in\mathbb{R}^n} \frac{1}{2}\|\mathcal{P}_{\Xi_{(n)}}(\mathcal{T}(\mathbf{z})) - \hat{\mathbf{R}}_Y\|_F^2 + \lambda\|\mathcal{T}(\mathbf{z})\|_* \tag{2.47}$$

Here, $\Xi_{(n)}$ corresponds to the sampling pattern of a GNS sampler or sparse ruler of size $O(\sqrt{n})$, *that selects entries from the* $n\times n$ *principal minor of* $\mathcal{T}(\mathbf{t}_N)$ *(with some repetitions)*, and we recover the estimate $\mathbf{t}_{(n)}^{\#}$ of $\mathbf{t}_{(n)}$ (instead of $\mathbf{t}_N$). In this case, we can use the analysis technique of [40] (which is based upon the analysis framework for M-estimators in [53]) to bound the error $\|\mathbf{t}_{(n)} - \mathbf{t}_{(n)}^{\#}\|_2$. From [40, 53], if $\lambda \geqslant 2\|\mathcal{P}_{\Xi_{(n)}}(\mathcal{T}(\mathbf{t}_{(n)})) - \hat{\mathbf{R}}_{\mathbf{Y}}\|$, then the estimation error (when $\Xi_{(n)}$ represents a sparse ruler or GNS), is given by

$$\|\mathbf{t}_{(n)} - \mathbf{t}_{(n)}^{\#}\|^2 \leqslant C\lambda^2 nr \tag{2.48}$$

The choice of λ depends on the specific sampler used but a lower bound can be computed as follows. When $\Xi_{(n)}$ corresponds to sparse ruler or GNS, we observe n entries of $\mathcal{T}(\mathbf{t}_{(n)})$ at least once, implying $\|\mathcal{P}_{\Xi_{(n)}}(\mathcal{T}(\mathbf{t}_n)) - \hat{\mathbf{R}}_{\mathbf{Y}}\|_F \geqslant \sqrt{n\varepsilon^2}$ where ε is the upper bound for entry-wise noise as introduced in (2.25). Then λ satisfies

$$\begin{aligned}\lambda &\geqslant 2\|\mathcal{P}_{\Xi_{(n)}}(\mathcal{T}(\mathbf{t}_{(n)})) - \hat{\mathbf{R}}_{\mathbf{Y}}\| \\ &\geqslant \frac{2}{\sqrt{n}}\|\mathcal{P}_{\Xi_{(n)}}(\mathcal{T}(\mathbf{t}_{(n)})) - \hat{\mathbf{R}}_{\mathbf{Y}}\|_F = 2\varepsilon\end{aligned} \tag{2.49}$$

From (2.48) and (2.49), we have the following "best-case" upper bound on the

estimation error (over all choices of λ)

$$\|\mathbf{t}_{(n)} - \mathbf{t}_{(n)}^{\#}\|_2^2 \leqslant C' nr\varepsilon^2 \tag{2.50}$$

where C' is a constant.

We can compare (2.50) with the bound in Theorem 4 obtained from simple LS-denoising with PSD constraint. If rank r can be treated as constant then (2.50) is of the same order as (2.42). It is to be noted that (2.42) represents a worst-case or most pessimistic upper bound (computed using the worst case value of $\|\mathbf{w}_{(n)}\|_2$) whereas (2.50) is a best-case upper bound over all choices of λ. Moreover, both bounds do not require *any separation condition* on the frequencies $f_i, i = 1, 2, \cdots r$.

At last, it should be noted that PSD constraint is not used when the authors of [40] analyze the recovery performance (see Appendix D therein). In the following section, we will numerically show that even PSD constraint is added to (2.46), the proposed algorithm in this paper provides better performance in sense of estimation error when the total number of measurements are the same.

2.7 Numerical Results

In previous sections, we have discussed the theoretical results on low rank Toeplitz covariance matrix estimation. In noiseless case, optimal compression rate can be achieved with respect to the DOF of rank r Toeplitz covariance that only $r+1$ entries of original $\mathbf{T}$ are sufficient for exact recovery. If the noise is present,

the algorithm proposed in this thesis has much lower computational complexity compared to existing methods in literature since we only need to sample a submatrix $\mathbf{T}_{(n)}$ where $n \ll N$. In addition, the proposed algorithm is parameter free compared to those in [42, 23, 57, 43] which is advantageous in real practices since it may not be possible to estimate the noise-power dependent parameters.

In this section, we will implement extensive numerical experiments to demonstrate our theoretical claims made in earlier sections. And we will show these simulations results for noiseless and noisy cases respectively. In addition, we will compare the proposed algorithm with other methods in litarature.

2.7.1 Exact Recovery via Vandermonde Decomposition

Fig. 2.3 shows the 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 coupled with MUSIC based recovery can perfectly recover $\mathbf{T}$ as soon as $n \geqslant r + 1$. The phase transition exhibits *a perfectly linear behavior, which is in agreement with the fundamental compression limit of rank r Toeplitz matrices* .

We compare the proposed method with the random sampling based compression and recovery of Toeplitz matrices, proposed in [23]. The sampling model for our method is different from that in [23]. For fairness of comparison, we fix the value of n and simulate the measurement model in [23] by collection n measure-

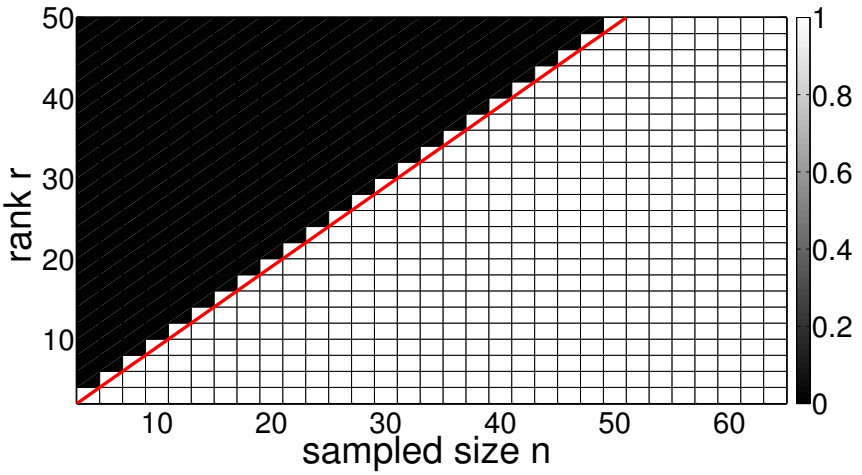

Figure 2.3: Phase transition plot for the proposed GNS based compression and MUSIC based reconstruction of $\mathbf{T}$. A trial is declared successful if $\|\mathbf{t}_{(N)} - \mathbf{t}^{\#}_{(N)}\|_2 / N \leqslant 0.001$. White cells indicate success while black denote failure. The red line represents $n = r + 1$ and the result is averaged over 50 runs. $N = 113$.

ments. This ensures that the reconstruction algorithms for *both approaches use the compressed sketch of same size*. Fig. 2.4 shows the phase transition for the approach in [23]. Comparing Fig. 2.3 and Fig. 2.4, 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. 2.4 is due to the nature of random sampling used in [23], for which the number of required measurements needed for a given r is strictly larger than that for our method.

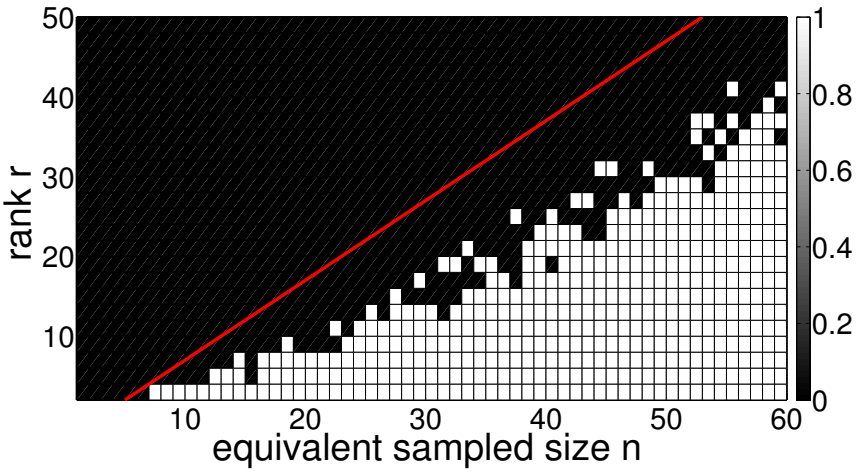

Figure 2.4: Phase transition plot for method in [23] and nuclear-norm minimization based reconstruction of $\mathbf{T}$. A trial is declared successful if $\|\mathbf{t}_{(N)} - \mathbf{t}^{\#}_{(N)}\|_2/N \leqslant 0.001$. White cells indicate success while black denote failure. The red line represents $n = r + 3$ and the result is averaged over 100 runs. $N = 113$.

2.7.2 Observed and Prediction Error In Presence of Noise

We next evaluate the performance of the proposed method in presence of bounded noise, and compare it with related works in [23, 40]. In particular, we compare the following algorithms:

- *proposed*: This represents the proposed reconstruction algorithm described in Table 1.
- *nuclear-psd*: This represents the algorithm in [40] with PSD constraint. We use GNS for compressing the MMV model described in [40] and use two versions of the nuclear norm minimization algorithm : one for recovering the sampled submatrix $\mathbf{T}_{(n)}$, and the other for recovering the entire matrix $\mathbf{T}$. The specific version will become clear depending on the context.
- *CCG*: This represents the compression/reconstruction framework of [23] using random samplers. Although the sampling scheme is different from our method, we assume that the method in [23] collects n measurements, which match the total number of entries in our compressed sketch.

We numerically choose regularization parameters λ (for [40]) to ensure the best performance.

For noisy case, we define the Signal-to-Noise Ratio (SNR) as

$$SNR \triangleq 10 \log \frac{\sum_{i=1}^{n} t_i^2}{\sum_{i=1}^{n} w_i^2} \tag{2.51}$$

where $\mathbf{w}$ and $\mathbf{t}$ are noise and signal vectors of same length n. And the normalized estimation error is defined as

$$\epsilon = \frac{\|\mathbf{t} - \mathbf{t}^{\#}\|_2}{\|\mathbf{t}\|_2} \tag{2.52}$$

where $\mathbf{t}^{\#}$ is the estimation of $\mathbf{t}$ which is the first column of $\mathbf{T}$. Similarly, the normalized prediction error is defined as

$$\epsilon_{pred} = \frac{\|\mathbf{t}_{(-n)} - \mathbf{t}^{\#}_{(-n)}\|_2}{\|\mathbf{t}_{(-n)}\|_2} \tag{2.53}$$

where $\mathbf{t}_{(-n)} = [t_n, t_{n+1}, \cdots, t_{N-1}]^T$ and $\mathbf{t}^{\#}_{(-n)} = [t^{\#}_n, t^{\#}_{n+1}, \cdots, t^{\#}_{N-1}]^T$.

In Fig. 2.5, we study the prediction error ϵ_{pred} and the total error ϵ of the aforementioned algorithms as a function of SNR. It can be seen that, proposed method outperforms algorithms in [23, 40]. Particularly, since the sampler in [23] is generated randomly, all entries of $\mathbf{T}$ are effectively sampled as compared to GNS sampling matrix $\mathbf{A}_s$ which only sample a submatrix. However, the proposed method gives better performance in both total normalized error and prediction error.

We further study the prediction error for the proposed method as a function of the sampled size n. This also represents the scenario when we may overestimate the rank r and oversample the measurements (i.e. $n > r$). Fig. 2.6 shows the normalized total error and prediction error as a function of sampled size n. It can be seen that generally, the average prediction error decreases for increasing n, and increases for increasing rank and noise power. And the proposed method provides better performance than other two alternatives. It should be also notated that there is a threshold for proposed method which around the rank r which corresponds to the noiseless case.

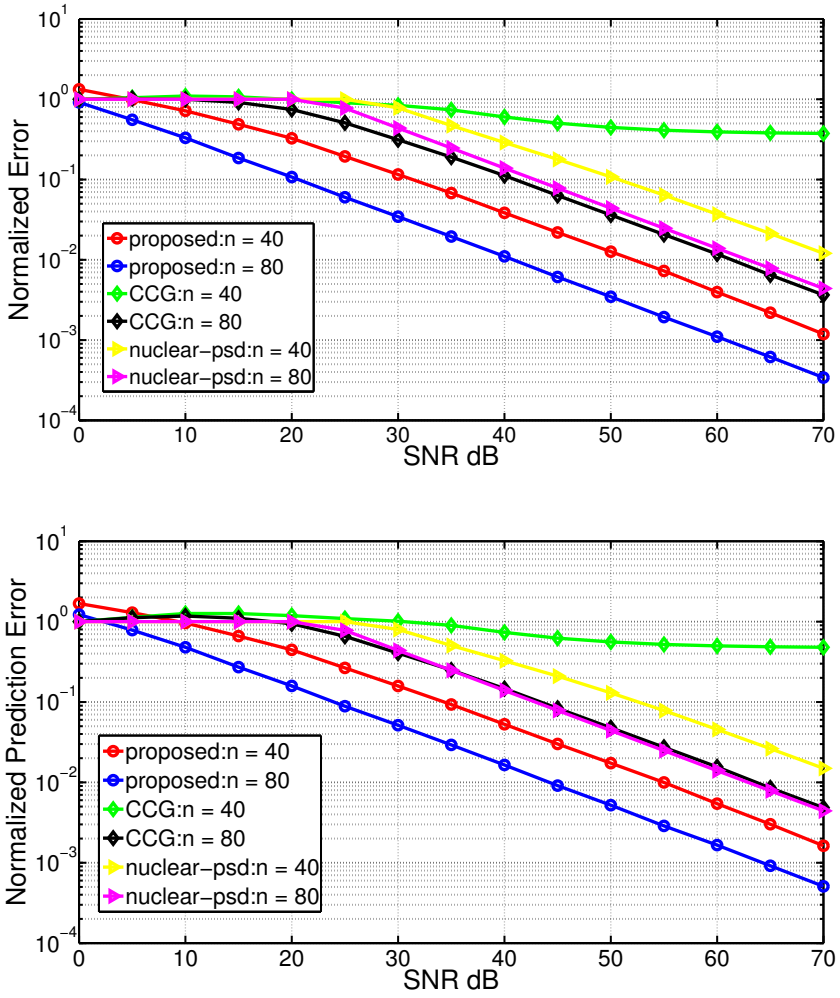

Figure 2.5: Estimation error of different algorithms as functions of SNR of. (Top) Normalized error ϵ v/s SNR. (Bottom) Prediction error v/s SNR. The results are averaged over 100 runs. Here $N = 110$, $r = 30$.

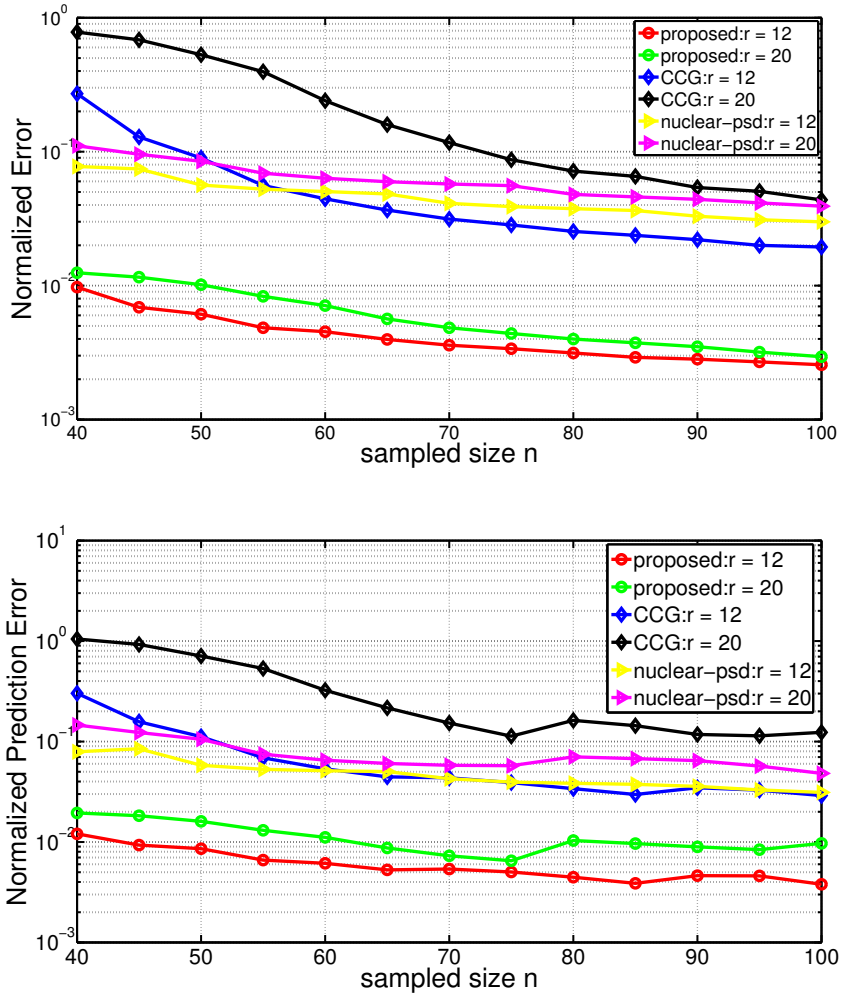

Figure 2.6: Estimation error of different algorithms as functions of n of. (Top) Normalized error ϵ v/s n. (Bottom) Prediction error v/s n. The results are averaged over 100 runs. Here $N = 110$, SNR $= 50$dB.

2.7.3 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.2.7, we 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 the method in [23]. The proposed method exhibits robustness to violation of the low rank assumption and its performance improves with increasing n.

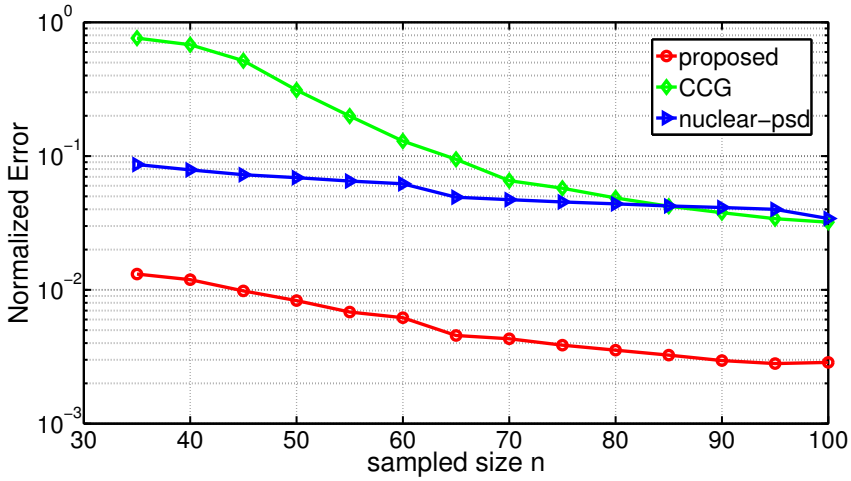

Figure 2.7: Recovery performance of proposed method and other methods in [23, 40] when $\mathbf{T}$ is approximately low rank Toeplitz. The matrix $\mathbf{T}$ is approximately rank 30 with ambient dimension $N = 110$. Here SNR is 50 dB and the results are averaged over 50 runs.

2.7.4 Frequency Estimation Performance

With bounded noise, the $\mathcal{T}(\mathbf{t}^{\#}_{(n)})$ given by (2.26) will be full rank. Consequently, via (2.23), there will be alias frequencies when $n > r$. For the proposed method, we will treat the r frequencies with largest amplitudes as the recovered frequencies. Since $\mathbf{T}^{\#}$ is real, the frequencies appear in conjugate pairs and we will show the frequency region $[0, 1/2]$ without losing generality. In Fig.2.8, we show the recovered the frequencies for different n and SNR along with the true frequencies.

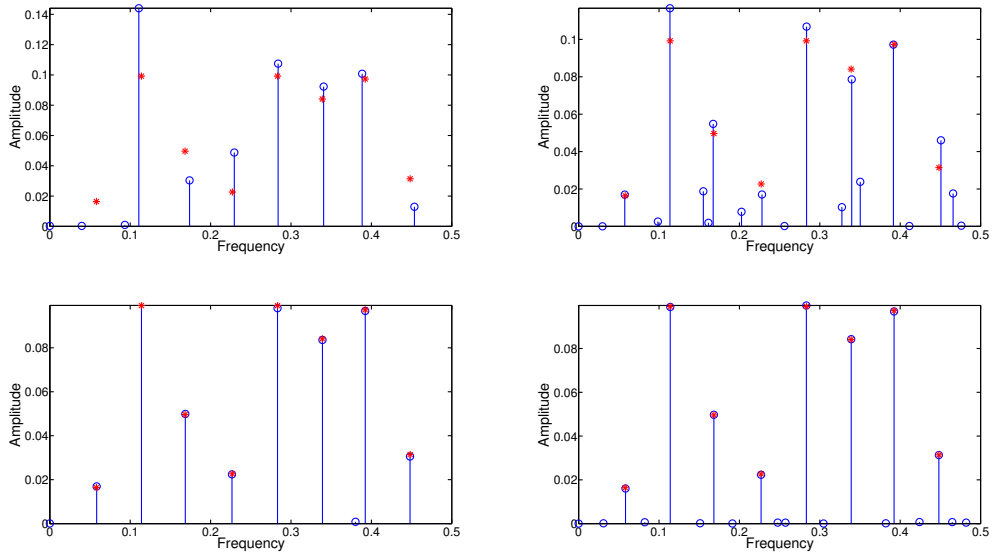

Figure 2.8: Recovered frequencies (blue) and true frequencies (red). (Top) SNR = 10 dB, $r = 16$. Left: $n = 20$, Right: $n = 40$. (Bottom) SNR = 40 dB, $r = 16$. Left: $n = 20$, Right: $n = 40$.

Let $\{\tilde{f}_i\}_{i=1}^{r}$ be the recovered frequencies corresponding to largest amplitudes, the frequency estimation error is defined as

$$\epsilon_f = \sqrt{\sum_{i=1}^{r}(f_i - \tilde{f}_i)^2} \tag{2.54}$$

where both true and recovered frequencies are ordered in same manner. In Fig.2.9, we show the averaged frequency estimation error as a function of SNR and for different n. It can be seen that proposed algorithm provides accurate estimation of frequencies. For the same SNR, frequency error is larger when n increases since more alias frequencies occur by computing (2.23). And the estimation error decreases when SNR increases.

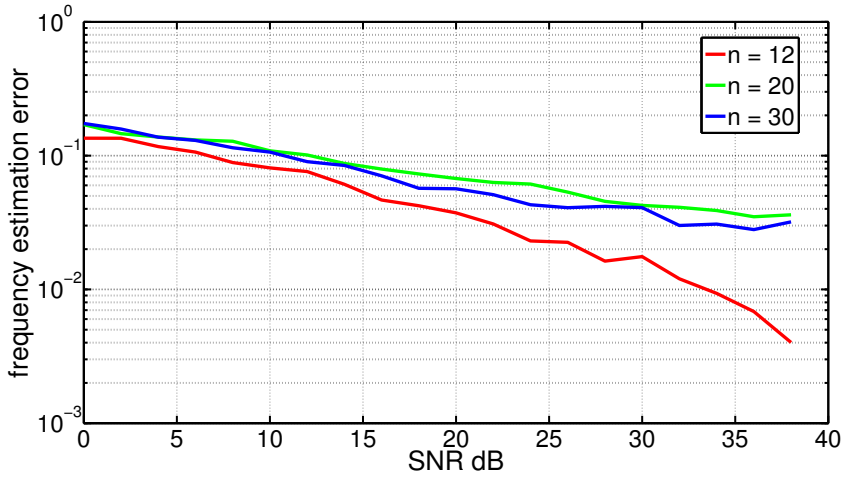

Figure 2.9: Frequency estimation plot for proposed algorithm. The matrix $\mathbf{T}$ is rank 10 with ambient dimension $N = 50$. Results are averaged over 500 runs.

2.7.5 Computational Complexity

Finally, we compare the computational complexity of the proposed method with *nuclear* and *CCG*. Fig. 2.10 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.

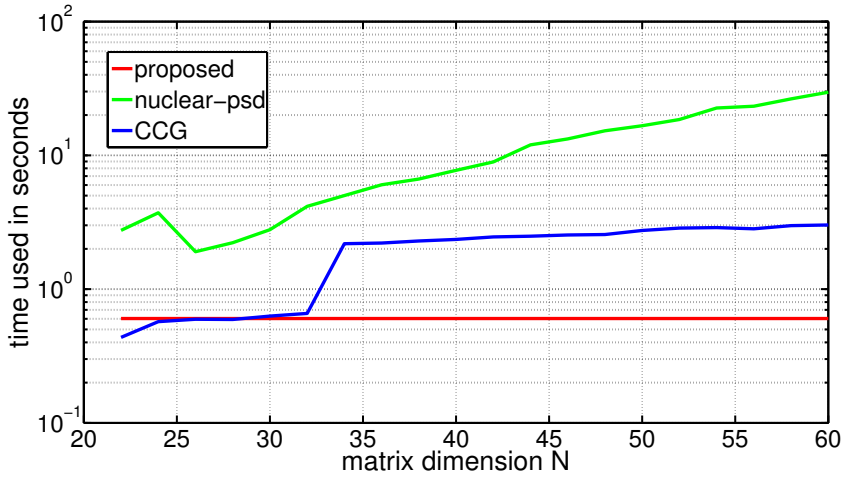

Figure 2.10: Comparison of run-times of the proposed method and the nuclear norm based recovery algorithms in [23, 40]. Here, $r = 10$, SNR $= 20$ dB and $n = 20$. The run-time is averaged over 100 runs.

Chapter 3: Randomized PNFS and Sparse Phase Retrieval with Noise

3.1 Introduction

The problem of reconstructing an unknown signal (up to a global phase) from its phaseless measurements has been studied for decades owing to its wide applications in many areas of imaging science such as X-ray crystallography, diffraction imaging and molecular imaging, and so forth [59, 60, 61]. The problem can be studied under various settings by considering a real or complex signal model, with or without sparsity constraints. A comprehensive review of existing measurement strategies and reconstruction algorithms for phase retrieval is provided in [62].

A central goal in phase retrieval problems is to develop an effective measurement strategy and a recovery algorithm which can provably recover the unknown signal with *minimal* number of measurements. In early works [86, 87, 88], the algorithms are iterative and start with a phase guess. The estimat is then updated repeatedly between the measurement domain and spatial domain. Recent approaches based on the elegant idea of "lifting" can provably recover (non-sparse) signals of dimension N using $O(N)$ or $O(N \log N)$ measurements [61, 63, 54, 83], by solving an appropriate convex problem in the "lifted" variable.

3.1.1 Related Work

Recovering a sparse signal from its phaseless measurements with near optimal number of measurements (which is $O(s)$ up to a logarithmic factor) is a challenging problem that has received much attention in recent times [65, 66, 67, 68]. In fact, it becomes necessary to impose a sparse prior on the unknown signal to ensure its unique recovery, when Fourier measurements are used. An l_1-minimization-based approach for sparse phase retrieval is proposed in [68], which requires $O(s^2)$ measurements along with an additional Collision-Free-Condition [69] on the autocorrelation of the unknown signal. In [65, 66, 67], the authors use a graph-decoding based approach which requires the sparsity to be at most $O(\sqrt{N})$. Recent iterative approaches using alternating minimization also require the number of measurements to grow quadratically in s [70]. Another iterative algorithm based on Fienup's work [86] is proposed in [93]. In [90], the authors develop a greedy algorithm for sparse phase retrieval. In [91], convex programming with $l1$ constraint is used and outliers of measurement are cosidered additional to noise. The authors of [92] solve the sparse phase retrieval with the idea of approximate message passing. In all [93, 90, 91, 92], no stable result is theoretically established and sufficient number of measurements for stable recovery is not given. In addition, for real signal and Fourier measurements, these works need cross-validation to find the optimal estimation and get rid of ambiguity given the true signal, which may not be possible in real applications. In [71, 23, 72], the problem of sparse phase retrieval is cast as a joint low-rank+sparse matrix recovery problem which minimizes the weighted linear combination of the

nuclear norm and l_1 norm of an appropriate lifted variable. However, as pointed out in [73], convex optimization based techniques for such simultaneously structured models (low rank+sparse) will necessarily require the number of measurements to be at least quadratic in s. Very recently, concurrent with our own work on Partial Nested Fourier Sampler (PNFS), a promising approach to overcome this limitation has been suggested in [74], where by using constrained measurement vectors and a two-step recovery algorithm, the authors can guarantee unique solution to the sparse phase retrieval problem using only $O(s\log(N/s))$ measurements.

3.1.2 Our Contributions

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 [29, 13, 14]. 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. Unlike [68], 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 [79], 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 [60]. 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.

3.2 Problem Setting and Fourier Based Phase Retrieval

Let $\mathbf{x} \in \mathbb{C}^N$ be a complex valued vector which may be sparse. Given M sampling vectors $\mathbf{f}_i \in \mathbb{C}^N, i = 1, 2, \cdots, M$, we obtain M noisy magnitude measurements as [63, 77, 23]

$$y_i = |\langle \mathbf{x}, \mathbf{f}_i \rangle|^2 + n_i \tag{3.1}$$

where n_i denotes the additive noise. The fundamental objective of phase retrieval problem is to recover $\mathbf{x}$ from $y_i, i = 1, 2, \cdots, M$. It is well known that for complex $\mathbf{x}$, we can only recover $\mathbf{x}$ up to a global phase ambiguity [59]. We can equivalently express the measurements as

$$y_i = \left(\mathbf{f_i}^T \otimes \mathbf{f_i}^H\right) \mathrm{Vec}\left(\mathbf{x}\mathbf{x}^H\right) + n_i \tag{3.2}$$

where $\otimes$ denotes the Kronecker product and $\mathrm{Vec}(\cdot)$ is the column-wise vectorized form of a matrix.

3.2.1 Limitations of Fourier Sampling based phase retrieval

Sparse phase retrieval problem was first studied in the context of Fourier measurement vectors [78, 65, 68, 79]. In the Fourier based phase retrieval problem, we collect measurements $y_i, 1 \leqslant i \leqslant M$ using an (oversampled) Fourier sampling vector [62, 61, 78, 65, 80, 79, 81]

$$\mathbf{f}_i = \alpha \left[1, z_i, z_i^2 \cdots, z_i^{N-1}\right]^T$$

Here $z_i = e^{j2\pi i/M}$ where $M \geqslant N$ is the oversampling factor.

It is readily seen that the Fourier based phase retrieval problem is equivalent to recovering a zero padded (if $M > N$) $\mathbf{x}$ from its autocorrelation sequence. This problem has an inherent ambiguity since two distinct finite length signals $x_1[n]$ and $x_2[n]$ (with same length N) can exhibit identical autocorrelation. This can be seen from the fact that the polynomial $X_1(z)\bar{X}_1(1/\bar{z})$ (denoting the $z-$ transform of the autocorrelation of $x_1[n]$ and $\bar{\cdot}$ is conjugate) can be decomposed into two spectral factors of same length in more than one way. To remove this ambiguity, it is necessary to impose additional priors on the signal $\mathbf{x}$. In [89], the authors avoid the inherent ambiguity of spectral factorization by adding one additional entry to make the signal minimum-phase. The number of measurements needed in [89] is $2N$ but sparsity can not be exploited.

Sparsity as a prior: A popular prior knowledge used in recent literature is that $\mathbf{x} \in \mathbb{C}^N$ is sparse with $s < N$ non zero elements. However, even with sparse priors,

it is non trivial to ensure unique recovery of $\mathbf{x}$ from its autocorrelation, since the autocorrelation may not be sparse. To remedy this, a "Collision-Free Condition" (CFC) is further imposed in literature[68, 69] and restated in Def.3. Under this condition, for $s \neq 6$, $\mathbf{x}$ can be uniquely recovered from M Fourier measurements where $M \geqslant s^2 - s + 1$, and M is a prime integer [68, 62].

Definition 3. (Collision-Free Condition) *[69, 68] 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* $(p, q) = (m, n)$.

Drawbacks: A major drawback of CFC is that it imposes an upper bound on the sparsity of $\mathbf{x}$ that we can only recover sufficiently sparse vectors whose sparsity can be at most $s = O(\sqrt{N})$. In practice, the no-collision property may only hold for even smaller values of s as experimentally validated in Fig.3.1 . Secondly, even with CFC, the l_1 minimization based recovery algorithm proposed in [68] requires $M = O(s^2)$ measurements, which is larger than the degrees of freedom in a sparse $\mathbf{x}$.

3.3 Nested Fourier Measurements and Phase Retrieval without Noise

3.3.1 Nested Fourier Measurement and Decoupling

As a major contribution of this thesis, we now propose a Fourier type measurement model namely the Partial Nested Fourier Sampler (PNFS), built upon the nested

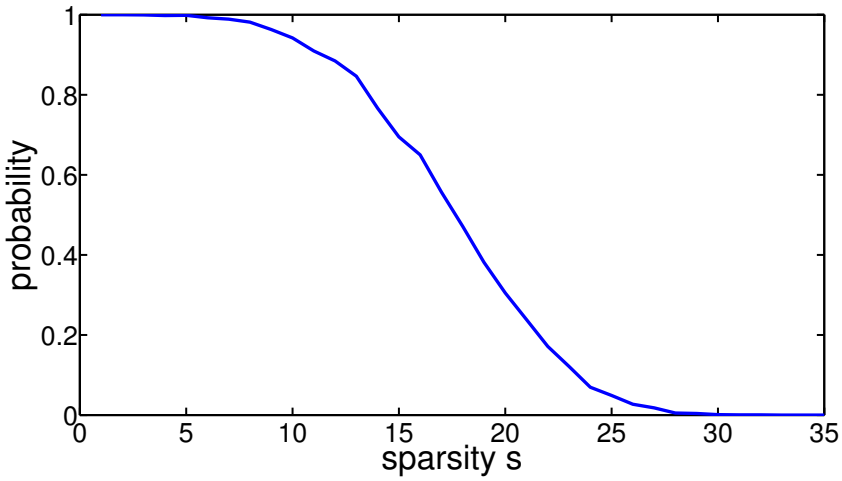

Figure 3.1: The probability of "no-collision" as a function of sparsity s. The ambient dimension is $N = 10000$ and the result is averaged over 2000 runs.

sampling idea in [29, 13, 14], that completely avoids the need for a collision-free condition and provides good performance guarantees.

We first define the following general model for Fourier measurement vectors:

Definition 4. *(**General Fourier Measurement**:) A General Fourier Measurement (GFM) vector is defined as*

$$\mathbf{f}_i = \alpha \left[z_i^{n_1}, z_i^{n_2}, \cdots, z_i^{n_N}\right]^T \tag{3.3}$$

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.

Definition 5. (Partial Nested Fourier Sampler:) *We define a Partial Nested Fourier Sampler (PNFS) as a special form of GFM vector defined in (3.3) 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 (3.3) and combining identical columns, in noiseless case, we have

$$y_i = \frac{1}{\sqrt{4N-5}} \Big[z_i^{-(2N-3)}, \cdots, z_i^{-1}, 1, z_i^1, \cdots, z_i^{2N-3} \Big] \tilde{\mathbf{x}} \tag{3.4}$$

where $\tilde{\mathbf{x}} \in \mathbb{C}^{4N-5}$ is the corresponding rearranged version of $\mathrm{Vec}(\mathbf{x}\mathbf{x}^H)$ with following form

$$[\tilde{\mathbf{x}}]_m = \begin{cases} \sum_{k=1}^{N} |x_k|^2 & m = 0 \\ \sum_{k=1}^{N-1-m} x_k \bar{x}_{k+m} & m = 1, 2, \cdots, N-2 \\ x_{2N-2-m} \bar{x}_N & N-1 \leqslant m \leqslant 2N-3 \\ \overline{[\tilde{\mathbf{x}}]_{-m}} & m < 0 \end{cases} \tag{3.5}$$

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 $|m| \geqslant N-1$, $[\tilde{\mathbf{x}}]_m$ 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 $|m| \geqslant N-1$ reveals the support of x.** In addition, for $s \geqslant 2$, $[\tilde{\mathbf{x}}]_m$ will vanish for $|m| \geqslant N-1$ 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

$$\mathbf{f}_i^{(l)} = \frac{1}{\sqrt[4]{4N-5}} \left[z_i^1, z_i^2, \cdots, z_i^{N-1}, z_i^{2N-2} \right] \mathbf{\Pi}^{(l)} \tag{3.6}$$

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 $y_i^{(l)}$, 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 (3.6) and obtain

$$\mathbf{y}^{(l)} = \mathbf{Z}\tilde{\mathbf{x}}^{(l)} \tag{3.7}$$

where $[\mathbf{y}^{(l)}]_i = y_i^{(l)}, [\mathbf{Z}]_{i,m} = \frac{1}{\sqrt{4N-5}} e^{j2\pi \frac{(i-1)m}{4N-5}}, 1 \leqslant i \leqslant \tilde{M}, \quad -2N+3 \leqslant m \leqslant 2N-3$. It is easy to see that $\mathbf{Z}$ is invertible if $\tilde{M} = 4N-5$ and $\tilde{\mathbf{x}}^{(l)}$ can be recovered from $\tilde{\mathbf{y}}^{(l)}$.

3.3.2 Iterative Algorithm

We now describe the details of a simple iterative algorithm that uses the permuted PNFS vectors iteratively to find a non zero entry of $\mathbf{x}$. Assuming sparsity $s \geqslant 2$, note that $[\tilde{\mathbf{x}}^{(l)}]_m$ will be all zero for $|m| \geqslant N-1$ if and only if the last entry is zero. Hence, the proposed 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

---

Input: data $\mathbf{x}$ *Output:* estimation $\mathbf{x}^{\#}$

1. *Initialization:* $l = N$
2. *Loop:*
   (a) **Step S1:** Using the permuted PNFS vectors (3.6), obtain $4N-5$ phaseless measurements

   $$y_i^{(l)} = |\langle \mathbf{x}, \mathbf{f}_i^{(l)} \rangle|^2, i = 1, 2, \cdots 4N-5$$

   Using (3.7), recover $\tilde{\mathbf{x}}^{(l)} = \mathbf{Z}^{-1}\mathbf{y}^{(l)}$

   (b) **Step S2:** If $[\tilde{\mathbf{x}}^{(l)}]_m = 0, \forall |m| \geqslant N-1$, 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 $|m| \geqslant N-1$, 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

   $$|x_N^{(l)}| = \sqrt{\gamma / [\tilde{\mathbf{x}}^{(l)}]_{m*}}$$

   & $\gamma = \sum_{k=1}^{N-1-m^*} [\tilde{\mathbf{x}}^{(l)}]_{2N-2-k} \overline{[\tilde{\mathbf{x}}^{(l)}]}_{2N-2-k-m*}$

   (b) Obtain estimate $\mathbf{x}^{\#}$ as

$$[\mathbf{x}^{\#}]_q = \begin{cases} \left(\frac{[\tilde{\mathbf{x}}^{(l)}]_{2N-2-q}}{|x_N^{(l)}|} \right) & q \neq \{l, N\} \\ |x_N^{(l)}| & q = l \\ \frac{[\tilde{\mathbf{x}}^{(l)}]_{2N-2-l}}{|x_N^{(l)}|} & q = N \end{cases}$$

Table 3.1: Iterative Algorithm for Phase Retrieval using PNFS

3.3.3 Accuracy of the Iterative Algorithm

In this section, we will show $\mathbf{x}^{\#}$ exactly recovers $\mathbf{x}$ in absence of noise. 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. In terms of the minimum number of required measurements, it is clear that the best case occurs when $l_{min} = N$ and the worst case happens for $l_{min} = s$. Formally, we have the following result

Theorem 5. *Let $\mathbf{x} \in \mathbb{C}^N$ be s-sparse with $s \geqslant 3$. The estimate $\mathbf{x}^{\#}$ produced by the iterative algorithm described in Table 1 is equal to $\mathbf{x}$ (in the sense of $\mathbb{C}\backslash\mathbb{T}$) 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.*

Proof. 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 $\breve{\mathbf{x}} = [x_1, x_2, \cdots, x_{N-1}]^T$ and let $\mathbf{r}_{\breve{\mathbf{x}}} \in \mathbb{C}^{2N-3}$ be the autocorrelation vector of $\breve{\mathbf{x}}$. Suppose m^* does not exist, implying $[\tilde{\mathbf{x}}]_m = 0$ for $1 \leqslant |m| \leqslant N-2$. Hence, $[\mathbf{r}_{\breve{\mathbf{x}}}]_n = \gamma\delta(n)$ where $\gamma = [\tilde{\mathbf{x}}]_0 - |x_N|^2$ and $\delta(n)$ is Kronecker delta. This means that $\hat{\mathbf{r}}_{\breve{\mathbf{x}}}(e^{j\omega}) \triangleq \sum_{n=-N+2}^{N-2}[\mathbf{r}_{\breve{\mathbf{x}}}]_n e^{-j\omega n}$ is an all-pass filter. However, $\hat{\mathbf{r}}_{\breve{\mathbf{x}}}(e^{j\omega}) = |\hat{\breve{\mathbf{x}}}(e^{j\omega})|^2$ where $\hat{\breve{\mathbf{x}}}(e^{j\omega}) \triangleq \sum_{n=-N+2}^{N-2}[\breve{\mathbf{x}}]_n e^{-j\omega n}$. This implies $\hat{\breve{\mathbf{x}}}(e^{j\omega})$ 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 [84]

$$[\check{\mathbf{x}}]_n = \lambda\delta(n - n_0) \tag{3.8}$$

for some n_0 satisfying $1 \leqslant n_0 \leqslant N-1$ and λ is a constant. However, since $s \geqslant 3$, $\check{\mathbf{x}}$ has at least two non zero entries which contradicts (3.8). Therefore, the existence of m^* is guaranteed.

It is then easy to see that $\mathbf{x}^{\#}$ is equal to $\mathbf{x}$ in sense of $\mathbb{C}\backslash\mathbb{T}$. In particular, assuming $l_{min} = N$, we have $[\mathbf{x}^{\#}]_N = \sqrt{\gamma/[\tilde{\mathbf{x}}^{(l)}]_{m^*}} = |x_N|$. Now, for $1 \leqslant q \leqslant N-1$, from (3.5), 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 4. *If* $\mathbf{x}$ *is nowhere vanishing (i.e.* $s = N$*), the number of measurements needed for recovering* $\mathbf{x}$ *is* $M = 4N - 5$.

Connection to $4N-4$ **Conjecture:** The interesting part of Corollary 4 is that $4N-4$ is a well known conjectured lower bound for complex phase retrieval [62]. Let $M^*(N)$ be the size of a set of measurements, then if successful phase retrieval holds for all $\mathbf{x} \in \mathbb{C}^N$, to the best of our knowledge, the lower bound on the number of measurements needed for phase retrieval is given by [85]

$$M^*(N) \geqslant \begin{cases} 4N - 2\alpha(N-1) - 3 & \text{for all } N \\ 4N - 2\alpha(N-1) - 2 & \text{if } N \text{ is odd and} \\ & \alpha(N-1) = 2 \bmod 4 \\ 4N - 2\alpha(N-1) - 1 & \text{if } N \text{ is even and} \\ & \alpha(N-1) = 3 \bmod 4 \end{cases}$$

where $\alpha(N-1)$ denotes the number of 1's in the binary representation of $N-1$ which is always no less than 1. As a quick observation, the lower bound is itself upper bounded by $4N-5$ where $\alpha(N-1)$ is 1, and $4N-5$ is the exact lower bound for $N = 2^p + 1$ for $p \geqslant 1$.

3.4 Randomized PNFS and Phase and Phase Retrieval with Noise

One of the major reasons for the inefficiency of the iterative algorithm in Theorem 5 is the number of measurements spent towards finding the non-zero pivot entry, $x_N^{(l)}$. We now introduce a randomized version of PNFS for sparse phase retrieval which requires only $O(slogN)$ measurements to ensure phase retrieval with high probability.

Definition 6. *(***Randomized PNFS***) A Randomized PNFS (R-PNFS) consists of measurement vectors*

$$\mathbf{f}_i^{(R\text{-}PNFS)} = [\mathbf{I}_{N,N} \quad \mathbf{v}]\, \mathbf{f}_i^{(N+1)}$$

where $\mathbf{v} \in \mathbb{C}^N$ *is a random vector with independent entries, and* $\mathbf{f}_i^{(N+1)}$ *is defined in (5) 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

$$y_i = \left| \left(\mathbf{f}_i^{\text{R-PNFS}} \right)^H \mathbf{x}^\star \right|^2 + n_i = \left| \mathbf{f}_i^{(N+1)^H} \begin{bmatrix} \mathbf{x}^\star \\ \\ \mathbf{v}^H \mathbf{x}^\star \end{bmatrix} \right|^2 + n_i \tag{3.9}$$

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.

3.4.1 A Cancellation Based Algorithm for R-PNFS

The main idea behind reducing the number of measurements for sparse retrieval using R-PNFS is to measure a sparse $\mathbf{x}^\star$ (with s non zero elements) using two sets of R-PNFS samplers, and perform sparse recovery on the difference between these 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 (5). 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

$$\begin{aligned} \mathbf{y}_i^{(1)} &= \left| \left(\mathbf{f}_i^{\text{R-PNFS}} \right)^H \mathbf{x}^\star \right|^2 + n_i^{(1)} \\ \mathbf{y}_i^{(2)} &= \left| \left(\tilde{\mathbf{f}}_i^{(N+1)} \right)^H \mathbf{x}^\star \right|^2 + n_i^{(2)} \end{aligned} \tag{3.10}$$

   We assume the noise is bounded, i.e. $|\mathbf{n}_i^{(k)}| \leqslant \eta, k = 1, 2$. 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*

$$\boldsymbol{\Delta}\mathbf{y} = \mathbf{Z}\hat{\mathbf{x}} + \boldsymbol{\Delta}\mathbf{n} \tag{3.11}$$

   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 $[\mathbf{Z}]_{i,k} = \frac{1}{\sqrt{4N-1}} e^{j2\pi \frac{m_i k}{4N-1}}$.

3. Obtain an estimate of $\hat{\mathbf{x}}$ as the solution to the following l_1-minimization problem:

$$\min_{\theta} \|\theta\|_1 \quad \text{subject to } \|\mathbf{\Delta y} - \mathbf{Z}\theta\|_2 \leqslant \eta\sqrt{\tilde{M}} \quad \textbf{(P1)}$$

4. Given the solution $\hat{\mathbf{x}}^{\#}$ to (**P1**), the estimate for each entry of $\mathbf{x}^{\star}$ is given by $x_q^{\#} = [\hat{\mathbf{x}}^{\#}]_{2N-q}/x_{N+1}^{\#}$ for $1 \leqslant q \leqslant N$ and $x_{N+1}^{\#} = |\sqrt{[\hat{\mathbf{x}}^{\#}]_0}|$.

3.4.2 Stability of Noisy Phase Retrieval with R-PNFS

To analyze the performance of the proposed algorithm, we use the following lemma from [82] which is tailored for the form (**P1**):

Lemma 7. *[82] 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 M rows whose indices are chosen uniformly at random from $[0, 4N-2]$. If $\tilde{M} \geqslant c_0(2s+1)\log(4N-1)\log(\varepsilon^{-1})$, then with probability at least $1-\varepsilon$, the solution $\hat{\mathbf{x}}^{\#}$ of (**P1**) satisfies*

$$\|\hat{\mathbf{x}} - \hat{\mathbf{x}}^{\#}\|_2 \leqslant c_1\sqrt{2s+1}\eta \tag{3.12}$$

where c_0, c_1 are universal constants.

Theorem 6. *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 (3.10) where the indices m_i of $\mathbf{f}_i^{(N+1)}, i = 1, 2\cdots, M$ are chosen uniformly at random from $[0, 4N-2]$. If $\tilde{M} \geqslant c_0(2s+1)\log(4N-1)\log(\varepsilon^{-1})$ and $|x_{N+1}|^2 > c_1\sqrt{2s+1}\eta$, with probability at least*

$1-\varepsilon$, *the estimates* $x_q^{\#}$ *of* $x_q^{\star}, 1 \leqslant q \leqslant N$, *satisfy*

$$\sum_{q=1}^{N} |x_q^{\star} - e^{j\phi_0} x_q^{\#}| \leqslant \frac{c_1\sqrt{(2s+1)(4N-1)}}{\sqrt{|x_{N+1}|^2 - c_1\sqrt{2s+1}\eta}}\eta + \\ + \|\mathbf{x}^{\star}\|_1 \left(\frac{1}{\sqrt{1 - c_1 \frac{\sqrt{2s+1}\eta}{|x_{N+1}|^2}}} - 1 \right) \tag{3.13}$$

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 7.*

Proof. For each $1 \leqslant q \leqslant N$, we have

$$\begin{aligned} x_q^{\star} - e^{j\phi_0} x_q^{\#} &= \frac{[\hat{\mathbf{x}}]_{2N-q}}{\bar{x}_{N+1}} - \frac{x_{N+1}}{|x_{N+1}|} \frac{[\hat{\mathbf{x}}^{\#}]_{2N-q}}{x_{N+1}^{\#}} \\ &= \frac{x_{N+1}}{|x_{N+1}|} \frac{[\hat{\mathbf{x}}]_{2N-q}}{|x_{N+1}|} - \frac{x_{N+1}}{|x_{N+1}|} \frac{[\hat{\mathbf{x}}^{\#}]_{2N-q}}{x_{N+1}^{\#}} \\ &= \frac{x_{N+1}}{|x_{N+1}|} \left(\frac{[\hat{\mathbf{x}}]_{2N-q}}{|x_{N+1}|} - \frac{[\hat{\mathbf{x}}]_{2N-q}}{x_{N+1}^{\#}} + \right. \\ &\left. \frac{[\hat{\mathbf{x}}]_{2N-q}}{|x_{N+1}^{\#}|} - \frac{[\hat{\mathbf{x}}^{\#}]_{2N-q}}{x_{N+1}^{\#}} \right) \\ &= \frac{x_{N+1}}{|x_{N+1}|} \left(\frac{[\hat{\mathbf{x}}]_{2N-q}}{x_{N+1}^{\#}} - \frac{[\hat{\mathbf{x}}^{\#}]_{2N-q}}{x_{N+1}^{\#}} \right) + \\ &\frac{x_{N+1}}{|x_{N+1}|} \left(\frac{[\hat{\mathbf{x}}]_{2N-q}}{|x_{N+1}|} - \frac{[\hat{\mathbf{x}}]_{2N-q}}{x_{N+1}^{\#}} \right) \\ &= \frac{x_{N+1}}{|x_{N+1}|} \left(\frac{[\hat{\mathbf{x}}]_{2N-q}}{x_{N+1}^{\#}} - \frac{[\hat{\mathbf{x}}^{\#}]_{2N-q}}{x_{N+1}^{\#}} \right) + \\ &\left(1 - \frac{|x_{N+1}|}{x_{N+1}^{\#}} \right) x_q^{\star} \end{aligned} \tag{3.14}$$

Using simple triangle inequality, we have

$$|x_q^{\star} - e^{j\phi_0} x_q^{\#}| \leqslant \beta \frac{\epsilon_{2N-q}}{|x_{N+1}|} + |1-\beta||x_q^{\star}| \tag{3.15}$$

and

$$\sum_{q=1}^{N} |x_q^\star - e^{j\phi_0} x_q^\#| \leqslant \beta \frac{\sum_{q=1}^{N} \epsilon_{2N-q}}{|x_{N+1}|} + |1-\beta| \sum_{q=1}^{N} |x_q^\star| \tag{3.16}$$

where $\epsilon_{2N-q} \triangleq |[\tilde{\mathbf{x}}]_{2N-q} - [\tilde{\mathbf{x}}^\#]_{2N-q}|$ and $\beta = \frac{|x_{N+1}|}{x_{N+1}^\#}$. We have

$$\begin{aligned} &\sum_{q=1}^{N} |x_q^\star - e^{j\phi_0} x_q^\#| \leqslant \beta \frac{\sum_{q=1}^{N} \epsilon_{2N-q}}{|x_{N+1}|} + |1-\beta| \sum_{q=1}^{N} |x_q^\star| \\ &\leqslant \beta \frac{\|\hat{\mathbf{x}} - \hat{\mathbf{x}}^\#\|_1}{|x_{N+1}|} + |1-\beta| \|\mathbf{x}^\star\|_1 \end{aligned} \tag{3.17}$$

Since $\|\hat{\mathbf{x}} - \hat{\mathbf{x}}^\#\|_1 \leqslant \sqrt{4N-1} \|\hat{\mathbf{x}} - \hat{\mathbf{x}}^\#\|_2$, Lemma 7 gives us

$$\begin{aligned} &\sum_{q=1}^{N} |x_q^\star - e^{j\phi_0} x_q^\#| \\ &\leqslant c_1 \beta \frac{\sqrt{(2s+1)(4N-1)}\eta}{|x_{N+1}|} + |1-\beta| \|\mathbf{x}^\star\|_1 \end{aligned} \tag{3.18}$$

Since $[\hat{\mathbf{x}}]_0 = |x_{N+1}|^2$, it follows from Lemma 7 that $|1 - \frac{1}{\beta^2}| \leqslant c_1 \frac{\sqrt{2s+1}\eta}{|x_{N+1}|^2}$. If $|x_{N+1}|^2 > c_1\sqrt{2s+1}\eta$, we have

$$1 - c_1 \frac{\sqrt{2s+1}\eta}{|x_{N+1}|^2} \leqslant \frac{1}{\beta^2} \leqslant 1 + c_1 \frac{\sqrt{2s+1}\eta}{|x_{N+1}|^2} \tag{3.19}$$

The proof completes by plugging (3.19) in (3.18). □

3.5 Numerical Results

In this section, we will perform numerical experiments to validate the theoretical claims made in this chapter. We will separately consider non-sparse and sparse signal models for phase retrieval and for each case, we will compare with existing

state-of-the art methods in literature. We define the normalized error (minimized over all possible phase ambiguities) for any estimate $\mathbf{x}^{\#}$ of $\mathbf{x}$ as

$$\epsilon = \min_{\phi \in (0,2\pi]} \frac{\|\mathbf{x} - e^{j\phi}\mathbf{x}^{\#}\|_2}{\|\mathbf{x}\|_2}$$

3.5.1 Non-sparse Phase Retrieval

From Theorem 5 and Corollary 4, we have shown that $4N-5$ PNFS measurements are sufficient for exact recovery in absence of noise. In [89], the authors solve the phase retrieval problem by adding one additional entry to the original data so that the extended data has minimum phase property. Then, given the Fourier measurements, the auto-correlation function can be computed and the minimum-phase solution can be found uniquely. However, the number of measurements needed is lower bounded by $2N$ *regardless of the sparsity*. We compare our proposed algorithm with that in [89] for non-sparse data. In addition, we also compare with random measurement based method in [23] which uses Semidefinite Programming (SDP) and nuclear norm regularizer for rank minimization.

In Fig. 3.2, we generate real Gaussian random data. It should be noted that the proposed method and the one in [89] can also work for complex data but the SDP based method in [23] is discussed only for real data. It can be seen that the proposed method outperforms the other two methods in terms of normalized estimation error.

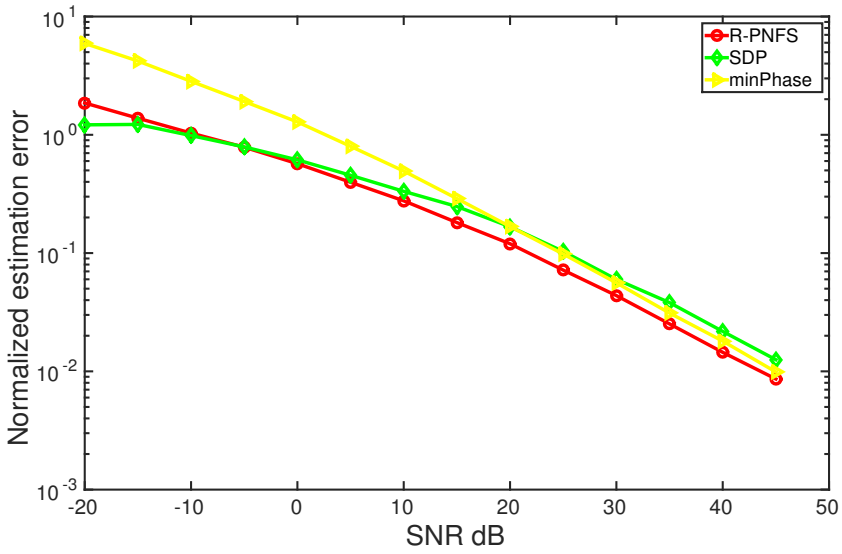

Figure 3.2: Performance comparison for non-sparse real data for proposed R-PNFS method and those in [89, 23]. Data dimension is $N = 30$ with $M = 4N - 1$ and the results are averaged over 50 runs.

3.5.2 Sparse Phase Retrieval

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.

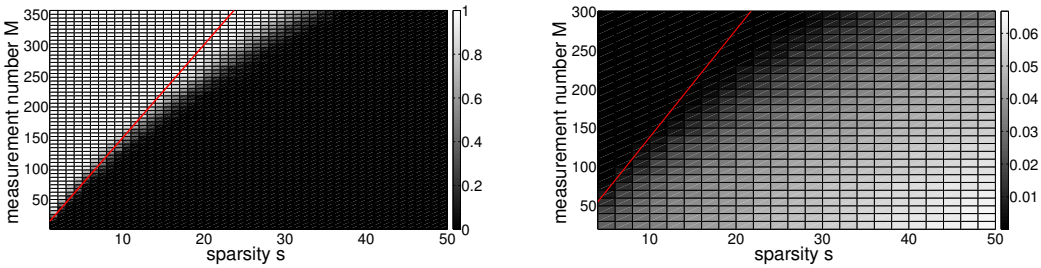

Figure 3.3: (Left) Phase transition for noiseless case, averaged over 100 runs with $N = 150$. White and black boxes denote success rates of 1 and 0 respectively. (Right) Phase transition for noisy case averaged over 50 runs with $N = 100$, and entry-wise noise bounded by 0.01. Each box denotes $\frac{1}{N}\sum_{q=1}^{N} |x_q^\star - e^{j\phi_0} x_q^\#|$. The red line represents $M = 3s \log N$ for both.

The phase transition plots of the proposed method for both noiseless and noisy signal models is depicted in Fig.3.3. 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 approach recovers the true $\mathbf{x}^\star$ with $M = O(s \log N)$ measurements.

In Fig. 3.4, 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.

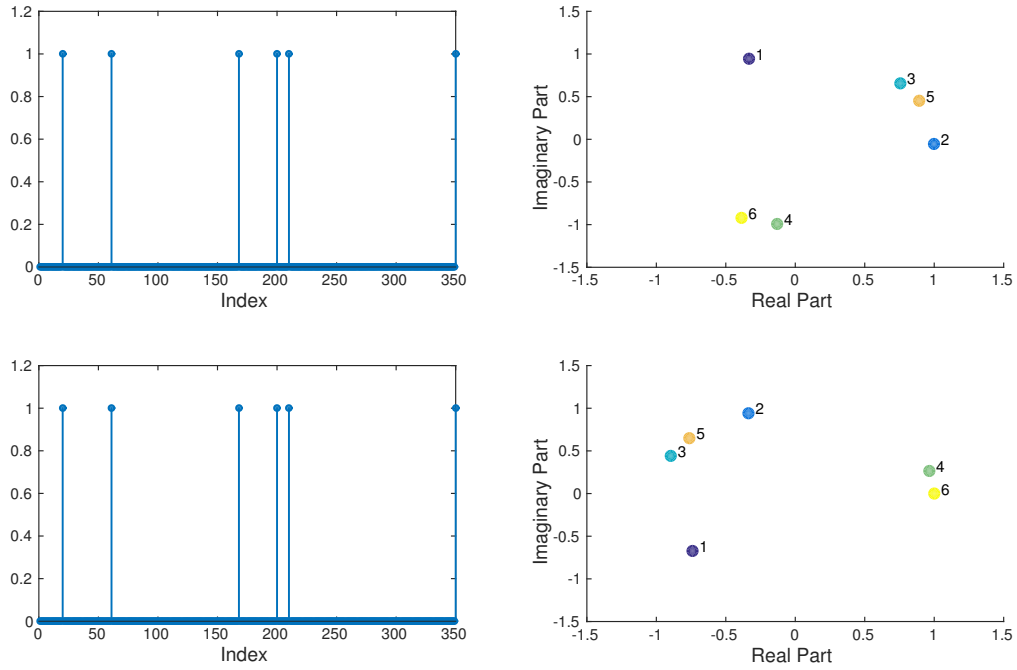

Figure 3.4: (Top left) Amplitudes of the original data. (Top right) The complex plane representation of the nonzero part of the original data.(Bottom left) Amplitudes of the recovered data. (Bottom right) The complex plane representation of the recovered data. Here, N=350, s=6 and M = 100.

In [90], the authors propose a greedy algorithm called "GESPAR" for sparse phase retrieval of real data with Fourier measurements. The estimate is found not only by searching all possible phase in range $(0, 2\pi]$ but also trying different shifted versions of the true data, which may not be possible in practice. However, we still make comparisons with GESPAR in Fig.3.5 assuming that it is allowed to perform the shift operations. It should be noted that GESPAR always need $4N$ measure-

ments which may not be desirable for small sparsity s. It can be clearly seen that proposed method outperforms GESPAR with the same number of measurements and SNR.

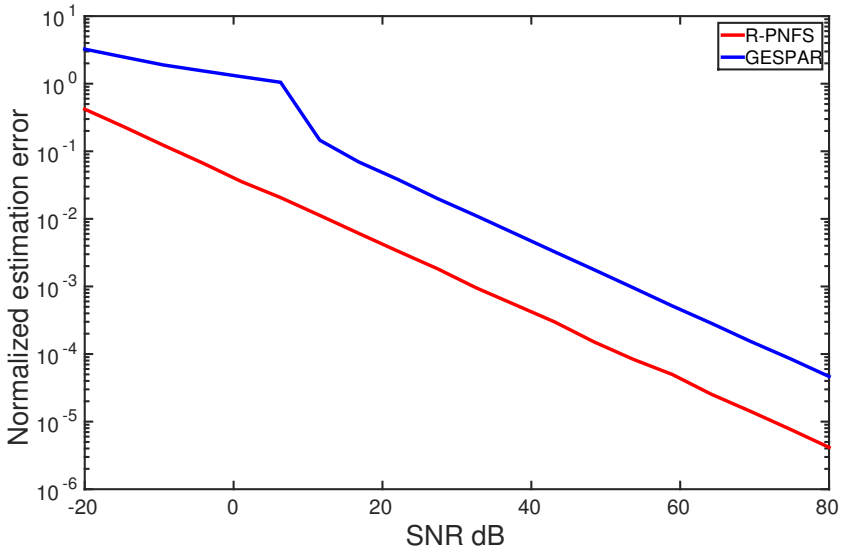

Figure 3.5: Performance comparison for parse real data for proposed R-PNFS method and GESPAR [90]. Data dimension is $N = 64$ with $s = 8$ The results are averaged over 50 runs.

Chapter 4: Conclusion

In this thesis, we considered compressive quadratic measurement models for two signal processing problems, namely that of covariance estimation and phase retrieval, and demonstrated how the idea of nested sampling can used to achieve optimal compression in both cases.

For compressive covariance estimation, we focused on the class of low rank Toeplitz covariance matrices and introduced a new structured sampler, namely the Generalized Nested Sampler (GNS) for compressing such matrices. 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. 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 linear prediction and line spectrum estimation respectively and studied the performance of gridless techniques, such as MUSIC, for recovering $\mathbf{T}$ from its sketch produced by the GNS. In absence of noise, we established exact recovery of a rank-r Toeplitz matrix of *any size* by Carathéodory's representation of low rank Toeplitz matrix and perfect recovery is possible without assuming a separation condition between

the frequencies parameterizing the low rank PSD Toeplitz matrix. In the presence of bounded noise, we propose a parameter-free algorithm for recovering $\mathbf{T}$. We use LS-denoising and make predictions for estimating whole high-dimensional matrix. This novel prediction based algorithm is based on the representation of positive definite Toeplitz covariance in [52]. We developed an explicit bound on the prediction error in terms of r, noise power and the observation length n. The numerical simulations validated the theoretical claims established in this paper and show that proposed algorithm provides better performance than existing algorithms in literature. Future work will be directed towards understanding the need for separation condition for frequency estimation in presence of noise, and establishing optimal error bounds.

We next considered the problem of phase retrieval using Fourier measurements, and introduced a new Fourier-based sampler, namely the Partial Nested Fourier Sampler (PNFS) that is capable of overcoming inherent ambiguities of Fourier based phase retrieval. For non-sparse signals, we established that $4N-5$ PNFS measurements are sufficient for exact recovery in absence of noise. It should be noted that $4N-4$ is a conjectured lower bound for general phase retrieval. For sparse signals, a randomized version of PNFS and a novel cancellation-based algorithm are proposed which only require $M = O(s \log N)$ measurements for stable estimation based. Unlike other algorithms, the proposed framework works for almost all sparsity levels and its estimation error is smaller than greedy algorithms as shown by simulation result.

In future, we will develop a unified analytical framework for analyzing the performance of general regularizer-free algorithms for covariance compression and

understand if they are capable of achieving the Cramér-Rao bounds. Our results can also be applied to Sparse Bayesian Learning where covariance estimation is a key step towards estimating a sparse signal. Finally, we will explore the possibility of implementing the proposed PNFS sampler using holographic techniques, and integrating them with random coded masks.

Bibliography

[1] D. Donoho, "50 years of Data Science", Sep. 2015.

[2] R. G. Baraniuk, "More is less: signal processing and the data deluge," *Science*, vol. 331, no. 6018, pp. 717-719, Feb. 2011.

[3] 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.

[4] 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.

[5] E. J. Candès and T. Tao, "Decoding by Linear Programming", *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 42034215, Dec. 2005.

[6] D. L. Donoho, "Compressed Sensing", *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289-1306, April 2006.

[7] M. F. Duarte and R. G. Baraniuk, "Kronecker Compressive Sensing," IEEE Trans. Image Processing, vol. 21, no. 2, pp. 494-504, Aug. 2011.

[8] 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.

[9] R. Kueng, H. Rauhut and U. Terstiege, "Low rank matrix recovery from rank one measurements", *Applied and Computational Harmonic Analysis*, to appear.

[10] S. Ling and T. Strohmer, "Self-Calibration and Biconvex Compressive Sensing", *Inverse Problems*, vol. 31, no. 11, 115002, Sep. 2015.

[11] Y. Chi, "Guaranteed Blind Sparse Spikes Deconvolution via Lifting and Convex Optimization", *arXiv: 1506.02751v3*, 2016.

[12] J. Leech, "On the representation of 1,2,...,n by differences," *J. London Math. Soc.*, vol. s1-3, no. 2, pp. 160-169, 1956.

[13] 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.

[14] 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.

[15] Heng Qiao and Piya Pal, "Stable Compressive Low Rank Toeplitz Covariance Estimation Without Regularization", accepted by *Asilomar Conference on Signals, Systems, and Computers*, 2016.

[16] 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.

[17] Heng Qiao, and Piya Pal, "Gridless Line Spectrum Estimation and Low-Rank Toeplitz Matrix Compression Using Structured Samplers: A Regularization-Free Approach," under revision.

[18] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79-86, March 1951.

[19] L. L. Scharf, "Statistical signal processing: Detection, Estimation, and Time Series Analysis," *Prentice Hall*, 1991.

[20] 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.

[21] N. Kambhatla, and T. K. Leen."Dimension reduction by local principal component analysis." *Neural Computation* vol. 9, no. 7, pp. 1493-1516, Oct. 1997.

[22] 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.

[23] 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.

[24] D. Romero, and G. Leus, "Compressive covariance sampling." *Information Theory and Applications Workshop (ITA)*, pp. 1-8, Feb. 2013.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] E. Candes and B. Recht, "Exact Matrix Completion via Convex Optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717-772, Dec. 2009.

[35] 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.

[36] 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

[37] 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.

[38] 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.

[39] 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.

[40] 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.

[41] 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.

[42] 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.

[43] E. 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.

[44] G. Schiebinger, E. Robeva and B. Recht, " Superresolution without Separation," *arXiv preprint arXiv:1506.03144v2*, 2015.

[45] V. I. Morgenshtern and E. J. Candès, "Super-Resolution of Positive Sources: the Discrete Setup," *arXiv preprint arXiv:1504.00717v1*, 2015.

[46] 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.

[47] 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.

[48] U. Grenander and G. Szego, "Toeplitz forms and their applications," *Chelsea Pub.Co.*, 1984.

[49] 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.

[50] J. Makhoul, "Linear Prediction: A Tutorial Review", *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561-580, April 1975.

[51] 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.

[52] P. Stoica and R. Moses, "Spectral analysis of signals", Prentice-Hall, NJ: Upper Saddle River, 2005.

[53] 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.

[54] 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.

[55] M. Kabanava, R. Kueng, H. Rauhut and U. Terstiege, "Stable low-rank matrix recovery via null space properties", *arXiv preprint arXiv:1507.07184v4*, 2015.

[56] 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.

[57] E. Candès and C. Fernandez-Granda,"Super-resolution from noisy data", *J. Fourier Anal. Appl.*, vol. 19, no. 6, pp. 1229-1254, 2013.

[58] 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.

[59] 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.

[60] Y. Wang and Z. Xu, "Phase retrieval for sparse signals," *Appl. Comput. Harmon. Anal.*, vol. 37, no. 3, pp. 531-544, Nov. 2014.

[61] 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.

[62] 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.

[63] 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.

[64] 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.

[65] 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.

[66] 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.

[67] K. Jaganathan, S. Oymak and B. Hassibi, "Sparse phase retrieval: Uniqueness guarantees and recovery algorithms," *arXiv preprint arXiv: 1311.2745*, Nov. 2013.

[68] 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.

[69] J. Ranieri, A. Chebira, Y. M. Lu and M. Vetterli, "Phase retrieval for sparse signals: Uniqueness conditions," *arXiv preprint arXiv: 1308.3058*, Aug. 2013.

[70] 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.

[71] 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.

[72] 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.

[73] 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.

[74] S. Bahmani and J. Romberg, "Efficient Compressive Phase Retrieval with Constrained Sensing Vectors," *arXiv preprint arXiv: 1507.08254v1*, July 2015.

[75] 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.

[76] 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.

[77] 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.

[78] 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.

[79] 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.

[80] E. Candès, X. Li and M. Soltanolkotabi, "Phase retrieval from coded diffraction patterns," *to appear in Appl. Comput. Harmon. Anal.*, 2014.

[81] Y. Wang, "Minimal frames for phase retrieval," In: *Workshop of phaseless recovery*, College Park, 2013.

[82] A. Foucart and H. Rauhut, "A Mathematical Introduction to Compressive Sensing," *Applied and Numerical Harmonic Analysis*, Birkhäuser, 2013.

[83] 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.

[84] P. P. Vaidyanathan, "Multirate Systems and Filter Banks," 1993: Prentice-Hall

[85] 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.

[86] J. R. Fienup, "Phase retrieval algorithms: A personal tour [Invited]," *Appl. Opt.*, vol. 52, pp. 45-56, 2013.

[87] 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.

[88] 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.

[89] 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.

[90] 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.

[91] 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.

[92] 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.

[93] 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.