Information and Inference: A Journal of the IMA (2023) **12**, 1–55
https://doi.org/10.1093/imaiai/iaad033

Separation-free super-resolution from compressed measurements is possible: an orthonormal atomic norm minimization approach

Jirong Yi, Soura Dasgupta
Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242, USA

Jian-Feng Cai
Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong

Mathews Jacob
Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242, USA

Jingchao Gao
Applied Mathematical and Computational Sciences, University of Iowa, Iowa City, IA 52242, USA

Myung Cho
Department of Electrical and Computer Engineering, Behrend College, Penn State University, Erie, PA, 16563, and California State University, Northridge

and

Weiyu Xu*
Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242, USA
*Corresponding author: weiyu-xu@uiowa.edu

[Received on 3 July 2022; revised on 21 February 2023; accepted on 18 July 2023]

We consider the problem of recovering the superposition of R distinct complex exponential functions from compressed non-uniform time-domain samples. Total variation (TV) minimization or atomic norm minimization was proposed in the literature to recover the R frequencies or the missing data. However, it is known that in order for TV minimization and atomic norm minimization to recover the missing data or the frequencies, the underlying R frequencies are required to be well separated, even when the measurements are noiseless. This paper shows that the Hankel matrix recovery approach can super-resolve the R complex exponentials and their frequencies from compressed non-uniform measurements, regardless of how close their frequencies are to each other. We propose a new concept of orthonormal atomic norm minimization (OANM), and demonstrate that the success of Hankel matrix recovery in separation-free super-resolution comes from the fact that the nuclear norm of a Hankel matrix is an orthonormal atomic norm. More specifically, we show that, in traditional atomic norm minimization, the underlying parameter values *must* be well separated to achieve successful signal recovery, if the atoms are changing continuously with respect to the continuously valued parameter. In contrast, for the OANM, it is possible the OANM is successful even though the original atoms can be arbitrarily close. As a byproduct of this research, we provide one matrix-theoretic inequality of nuclear norm, and give its proof using the theory of compressed sensing.

Keywords: Separation-free super-resolution; orthonormal atomic norm minimization; Hankel matrix; perturbation analysis; nuclear norm.

2000 Math Subject Classification: 34K30; 35K57; 35Q80; 92D25.

1. Introduction

In super-resolution, we are interested in recovering the high-end spectral information of signals from observations of its low-end spectral components [5]. In one setting of super-resolution problems, one aims to recover a superposition of complex exponential functions from time-domain samples. In fact, many problems arising in science and engineering involve high-dimensional signals that can be modelled or approximated by a superposition of a few complex exponential functions. In particular, if we choose the exponential functions to be complex sinusoids, this superposition of complex exponentials models signals in acceleration of medical imaging [25], analog-to-digital conversion [43] and array signal processing [38]. Accelerated nuclear magnetic resonance (NMR) spectroscopy, which is a prerequisite for studying short-lived molecular systems and monitoring chemical reactions in real time, is another application where signals can be modelled or approximated by a superposition of complex exponential functions [19, 22, 34]. How to recover the superposition of complex exponential functions or parameters of these complex exponential functions is of critical importance in these applications.

In this paper, we consider the recovery of superpositions of complex exponentials from linear undersampled measurements. More specifically, let $\boldsymbol{x} \in \mathbb{C}^{2N-1}$ be an unknown vector (which we would like to recover) whose j-th element satisfies

$$\boldsymbol{x}_j = \sum_{k=1}^{R} c_k z_k^j, \qquad j = 0, 1, \ldots, 2N-2, \tag{1.1}$$

where $c_k \in \mathbb{C}$ are complex coefficients and $z_k \in \mathbb{C}$, $k = 1, \ldots, R$, are some unknown complex numbers with $2R \leq 2N-1$ being a positive integer. In other words, $\boldsymbol{x}$ is a superposition of R complex exponential functions. When $z_k = e^{2\pi \imath f_k}$, with $\imath = \sqrt{-1}$ and $k = 1, \ldots, R$, $\boldsymbol{x}$ is a superposition of complex sinusoids. When $z_k = e^{-\tau_k} e^{2\pi \imath f_k}$, with $k = 1, \ldots, R$, and $\tau_k > 0$, $\boldsymbol{x}$ can model signals acquired by NMR spectroscopy in monitoring real-time chemical reactions and studying short-lived molecular systems [3, 4]. The results of this paper apply to general $\tau_k > 0$, but, to simplify presentations, we choose to restrict certain theorems to $\tau_k = 0$. More specifically, $\tau_k > 0$ is allowed in Theorem 3.1 (we do not require knowledge of τ_k to establish recovery guarantees), while τ_k's are assumed to be equal to 0 (thus known) in Theorems 4.1 and 4.3. In our simulations, we also set τ_k to be equal to 0, where we consider the superposition of complex sinusoids.

Since $2R \leq 2N-1$ and often $2R \ll 2N-1$, the degree of freedom to determine $\boldsymbol{x}$ is much less than the ambient dimension $2N-1$. Therefore, one can still recover $\boldsymbol{x}$ by its undersampling [8, 16]. In particular, in undersampling, we assume that $\boldsymbol{x}$ is unknown, and we consider recovering $\boldsymbol{x}$ from its linear measurements

$$\boldsymbol{b} = \mathscr{A}(\boldsymbol{x}), \tag{1.2}$$

where $\mathscr{A}$ is a linear subsampling map from $\mathbb{C}^{2N-1}$ to $\mathbb{C}^M$, $M < 2N-1$. After $\boldsymbol{x}$ is recovered, we can use the single-snapshot MUSIC [24] or Prony's method to recover the parameters z_k.

The problem of recovering $\boldsymbol{x}$ from its linear measurements (1.2) can be solved using compressed sensing[8], by discretizing the dictionary of basis vectors into grid points corresponding to discrete values of z_k. When the parameters f_k's in signals from spectral compressed sensing (or the parameters (f_k, τ_k)'s from signals in accelerated NMR spectroscopy) fall on the grid, compressed sensing is a powerful tool to recover those signals even when the number of samples is far below its ambient dimension

($M \ll 2N-1$) [8, 16]. Nevertheless, the parameters in our problem setting often take continuous values, leading to a continuous dictionary, and may not exactly fall on a grid. The basis mismatch problem between the continuously valued parameters and the grid-valued parameters degrades the performance of conventional compressed sensing [14].

In two seminal papers [5, 41], the authors proposed using total variation (TV) minimization or atomic norm minimization to recover $\boldsymbol{x}$ or to recover z_k, when $z_k = e^{\imath 2\pi f_k}$ with f_k taking continuous values from $[0, 1)$. In these papers, the authors show that TV minimization or atomic norm minimization can recover the continuously valued frequency f_k's in the noise free case. However, as shown in [5, 40, 41], for TV minimization or atomic norm minimization to recover spectrally sparse data or the associated frequencies correctly, it is required that adjacent frequencies be sufficiently separated from each other. For example, for complex exponentials with z_k's taking values on the complex unit circle, it is required that adjacent frequencies $f_k \in [0, 1]$'s be at least $\frac{2}{(2N-1)\times 2\pi}$ apart [40]. This separation condition is necessary for the TV minimization, even if we observe all the $(2N-1)$ data samples, and even if the observations are noiseless. Please notice that the tolerable noise in correctly identifying frequencies to a certain accuracy depends on the frequency separation according to information-theoretic study [28].

This raises a natural question, 'Can we super-resolve the superposition of complex exponentials with *continuously valued* parameters z_k, without requiring frequency separations, from compressed measurements, in noiseless or low-noise measurements?' In this paper, we answer this question in the affirmative. More specifically, we show that a Hankel matrix recovery approach using nuclear norm minimization can super-resolve the superposition of complex exponentials with *continuously valued* parameters z_k from compressed measurements, without requiring frequency separations. This separation-free super-resolution result holds even when we only compressively observe $\boldsymbol{x}$ over a subset $\mathscr{M} \subseteq \{0, ..., 2N-2\}$ with cardinality $|\mathscr{M}| = M$.

In this paper, we give the worst-case and average-case performance guarantees of Hankel matrix recovery in recovering the superposition of complex exponentials. In establishing the worst-case performance guarantees, we establish conditions under which Hankel matrix recovery can recover the underlying complex exponentials, no matter what values the coefficients c_k's of the complex exponentials take. For the average-case performance guarantee, we assume that the phases of the coefficients c_k are uniformly distributed over $[0, 2\pi)$. For both the worst and average cases, we establish that Hankel matrix recovery can super-resolve complex exponentials with continuously valued parameters z_k, no matter how close they are to each other. We further introduce a new concept of orthonormal atomic norm minimization (OANM), and discover that the success of Hankel matrix recovery in separation-free super-resolution comes from the fact that the nuclear norm of the Hankel matrix is an orthonormal atomic norm. In particular, we show that, in traditional atomic norm minimization, for successful signal recovery, the underlying parameters *must* be well separated, if the atoms continuously depend on continuously valued parameters; however, it is possible the OANM can succeed even if the original atoms are arbitrarily close. As a byproduct of this research, we discover one interesting matrix-theoretic inequality of nuclear norm, and give its proof using the theory of compressed sensing.

We remark that some of our results can require high percentage of all the samples for successful recovery using Hankel matrix recovery. For example, in Theorem 3.1, we need $|\mathscr{M}| > 2N - 1 - \frac{N}{2R}$ out of $2N-1$ samples to successfully recover a signal consisting of R atoms. Similarly for Theorems 4.1 and 4.3, we assume the number of samples $|\mathscr{M}| = 2N-2$. We are also only able to show the tightness of Theorem 3.1, for a very special sampling scheme. Even for these special sampling schemes, to the best of our knowledge, our results are the first to show that Hankel matrix recovery can successfully recover superposition of complex exponentials regardless of their frequency or parameter separations. We also want to point out that these theoretical bounds can be conservative, compared with numerical results.

For example, in the experiments demonstrating how frequency separation can influence signal recovery performance, even though we only use $|\mathscr{M}| \approx 51\% \times (2N-1)$ samples, the Hankel matrix recovery approach can already recover faithfully the superposition of a large number of complex sinusoids which have small frequency separations.

1.1 *Comparisons with related works on Hankel matrix recovery and atomic norm minimization*

Low-rank Hankel matrix recovery approaches have been used for recovering parsimonious models in system identifications, control and signal processing. In [26], Markovsky considered low-rank approximations for Hankel structured matrices with applications in signal processing, system identifications and control. Fazel et al. [17, 18] introduced low-rank Hankel matrix recovery via nuclear norm minimization, motivated by applications including realizations and identification of linear time-invariant systems, inferring shapes (points on the complex plane) from moments estimation (which is related to super-resolution with z_k from the complex plane), and moment matrix rank minimization for polynomial optimization. Fazel et al. [18] further designed optimization algorithms to solve the nuclear norm minimization problem for low-rank Hankel matrix recovery. Chen and Chi [11] proposed to use multi-fold Hankel matrix completion for spectral compressed sensing, studied the performance guarantees of spectral compressed sensing via structured multifold Hankel matrix completion and derived performance guarantees of structured matrix completion. However, the results in [11] require that the Dirichlet kernel associated with underlying frequencies satisfies certain incoherence conditions, and these conditions require the underlying frequencies to be well separated from each other. In [15, 44], the authors derived performance guarantees for Hankel matrix completion in system identifications. However, the performance guarantees in [15, 44] require a very specific sampling pattern of fully sampling the upper-triangular part of the Hankel matrix. Moreover, the performance guarantees in [15, 44] require that the parameters z_k be very small (or smaller than 1) in magnitude. In our earlier work [3], we established performance guarantees of Hankel matrix recovery for spectral compressed sensing under Gaussian measurements of $\boldsymbol{x}$. By comparison, this paper considers direct observations of $\boldsymbol{x}$ over a set $\mathscr{M} \subseteq \{0, 1, 2, ..., 2N-2\}$, which is a more relevant sampling model in many applications. In the single-snapshot MUSIC algorithm [24], the Prony's method [33] or the matrix pencil approach [21], one would need the full $(2N-1)$ consecutive samples to perform frequency identifications, while the Hankel matrix recovery approach can work with compressed measurements. When prior information of the locations of the frequencies is available, one can use weighted atomic norm minimization to relax the separation conditions in successful signal recovery [27].

In [39], the authors considered super-resolution without separation using atomic norm minimization, but under the restriction that the coefficients are positive and for a particular set of atoms (point spread functions) $\psi(s,t) = e^{-(s-t)^2}$, where s represents the spatial location in an image and t is the location of a point source of light (t's are the parameters to be estimated). In contrast, the analysis in our paper deals with complex-numbered coefficients, and does not require the coefficients to be positive. In addition, the techniques used for establishing the recovery guarantees in [39] are quite different from ours. In [39], the main idea for establishing the recovery guarantee is constructing a dual certificate, by applying the machinery of Tchebycheff systems to Gaussian point spread functions. By comparison, in this paper, we exploit the null space condition for nuclear-norm-minimization-based signal recovery. This machinery of investigating the null space condition does not require constructing a dual certificate, thus making it possible to show that the Hankel matrix recovery can reconstruct the signals without imposing separation conditions on the underlying atoms.

Additionally, in previous research [30–32], Prony's method and related modified algorithms were studied to recover parameters of signals, which are expressed as sum of exponentials. Specifically, in [31], the authors considered the recovery of signal parameters with equispaced sampled and real-valued data through the approximate Prony's method. In [32], the authors considered the signal parameter estimation problem for the sum of non-increasing exponentials with equispaced sampled data, and introduced the connections among various parameter estimate methods including the Prony's method, the matrix pencil method and the ESPRIT method. In [30], the authors considered the Prony's method with equispaced and non-equispaced sampled data for the signal parameter estimation. In particular, the considered non-equispaced sampled data case in [30] is similar to our problem considered in this paper, which is about signal parameter estimation with missing data. A major difference between [30] and ours is the different methods used in recovering missing entries: in [30], the authors use the method based on interpolation, while we use Hankel matrix completion in this paper. In [28], the author established a phase transition on the cut-off frequency at which noisy super-resolution is possible, and showed that there exists pairs of frequency component hypothesis that cannot be distinguished when the noise is big. However, the results in [28] focus on recovering the frequencies, and do not apply to recovering missing data samples, the recovery of which is shown in this paper to be robust against noises (please see Theorem 3.2). Furthermore, the results in [28] focused on finding worst-case pairs of hard-to-distinguish frequency component hypotheses under worst-case noises, but they do not exclude the possibility of recovering missing data samples or frequencies from randomly distributed noises or magnitude noises with small enough magnitudes.

1.2 *Organization of this paper*

The rest of the paper is organized as follows. In Section 2, we present the problem model, and introduce the Hankel matrix recovery approach. In Section 3, we investigate the worst-case performance guarantees of recovering spectrally sparse signals regardless of frequency separation, using the Hankel matrix recovery approach. In Section 4, we study the Hankel matrix recovery's average-case performance guarantees of recovering spectrally sparse signals regardless of frequency separation. In Section 5, we show that in traditional atomic norm minimization, for successful signal recovery, the underlying parameters *must* be well separated, if the atoms depend on the continuously valued parameters. In Section 6, we introduce the concept of OANM, and show that it is possible that atomic norm minimization is successful even though the original atoms are arbitrarily close. In Section 7, as a byproduct of this research, we provide a new matrix-theoretic inequality of nuclear norm from the theory of compressed sensing. Numerical results are given in Section 8 to validate our theoretical predictions. We conclude our paper in Section 9. The proofs of the technical lemmas are in the appendix.

1.3 *Notations*

We denote the set of complex numbers and real number as $\mathbb{C}$ and $\mathbb{R}$, respectively. We use calligraphic uppercase letters to represent index sets, and use $|\cdot|$ to represent the cardinality of a set. When we use an index set as the subscript of a vector, we refer to the part of the vector over the index set. For example, $\boldsymbol{x}_{\Omega}$ is the part of vector $\boldsymbol{x}$ over the index set Ω. We use $\mathbb{C}_r^{n_1\times n_2}$ to represent the set of matrices from $\mathbb{C}^{n_1\times n_2}$ with rank r. We denote the trace of a matrix $\boldsymbol{X}$ by $\mathrm{Tr}(\boldsymbol{X})$, and denote the real and imaginary parts of a matrix $\boldsymbol{X}$ by $\mathrm{Re}(\boldsymbol{X})$ and $\mathrm{Im}(\boldsymbol{X})$, respectively. The superscripts T and $*$ are used to represent transpose, and conjugate transpose of matrices or vectors. The Frobenius norm, nuclear norm and spectral norm of a matrix are denoted by $\|\cdot\|_F$, $\|\cdot\|_*$ and $\|\cdot\|_2$ (or $\|\cdot\|$), respectively. The notation $\|\cdot\|$ represents the spectral norm if its argument is a matrix, and represents the Euclidean norm if its argument is a vector.

The probability of an event S is denoted by $\mathbb{P}(S)$. A vector or matrix with all its entries zero will be denoted by $\mathbf{0}$. The inner product for matrices is defined as $\langle \boldsymbol{M}, \boldsymbol{N} \rangle = \mathrm{Tr}(\boldsymbol{M}^T \boldsymbol{N})$ when $\boldsymbol{M}$ and $\boldsymbol{N}$ are real valued, or $\langle \boldsymbol{M}, \boldsymbol{N} \rangle = \mathrm{Re}(\mathrm{Tr}(\boldsymbol{M}^* \boldsymbol{N}))$ when $\boldsymbol{M}, \boldsymbol{N}$ are complex valued.

2. Problem statement

We consider the signal model in (1.1) with

$$z_k = e^{\imath 2\pi f_k - \tau_k}, k = 1, 2, \cdots, R, \tag{2.1}$$

where $f_k \in [0, 1)$ are normalized frequencies, $\tau_k > 0$ are the damping terms and the observation set is $\mathscr{M} \subset \{0, 1, \cdots, 2N - 2\}$.

To estimate the continuous parameter f_k in [5, 41], the authors proposed using TV minimization or atomic norm minimization to recover $\boldsymbol{x}$ or to recover the parameter z_k, when $z_k = e^{\imath 2\pi f_k}$ with f_k taking continuous values from $[0, 1)$. However, as shown in [5, 40, 41], in order for atomic norm minimization to recover spectrally sparse data or the associated frequencies correctly, it is necessary that adjacent frequencies be separated far enough from each other. As shown in [40], for complex exponentials with z_k's taking values on the complex unit circle, it is required that adjacent frequencies $f_k \in [0, 1)$ be at least $\frac{2}{2\pi(2N-1)}$ apart. This separation condition is necessary, even if we observe the full $(2N-1)$ data samples, and even if the observations are noiseless. A formal statement of minimal separation of frequencies can be found in Definition 2.1.

DEFINITION 2.1. ([5]) For a frequency subset $\mathscr{F} \subset [0, 1)$ with a group of points, the minimum separation is defined as smallest distance between two arbitrary different elements in $\mathscr{F}$, i.e.

$$\mathrm{dist}(\mathscr{F}) = \inf_{f_i, f_l \in \mathscr{F} f_i \neq f_l} d(f_i, f_l), \tag{2.2}$$

where $d(f_i, f_l)$ is the wrap around distance between two frequencies.

Following the idea of the matrix pencil method in [21] and Enhanced Matrix Completion (EMaC) in [11], we construct a Hankel matrix based on signal $\boldsymbol{x}$. More specifically, define the Hankel matrix $\boldsymbol{H}(\boldsymbol{x}) \in \mathbb{C}^{N \times N}$ by

$$\boldsymbol{H}_{jk}(\boldsymbol{x}) = \boldsymbol{x}_{j+k-2}, \qquad j, k = 1, 2, \ldots, N. \tag{2.3}$$

The expression (1.1) and (2.3) lead to a rank-R decomposition:

$$\boldsymbol{H}(\boldsymbol{x}) = \begin{bmatrix} 1 & \cdots & 1 \\ z_1 & \cdots & z_R \\ \vdots & \vdots & \vdots \\ z_1^{N-1} & \cdots & z_R^{N-1} \end{bmatrix} \begin{bmatrix} c_1 & & \\ & \ddots & \\ & & c_R \end{bmatrix} \begin{bmatrix} 1 & z_1 \cdots & z_1^{N-1} \\ \vdots & \vdots & \vdots \\ 1 & z_R \cdots & z_R^{N-1} \end{bmatrix}.$$

Instead of reconstructing $\boldsymbol{x}$ directly, we reconstruct the rank-R Hankel matrix $\boldsymbol{H}$, subject to the observation constraints. Low-rank matrix recovery has been widely studied in recovering a matrix from incomplete observations [2, 6, 7, 9, 35]. It is well known that minimizing the nuclear norm can lead to a solution of low-rank matrices. We therefore use the nuclear norm minimization to recover the low-rank matrix $\boldsymbol{H}$.

More specifically, for any given $\boldsymbol{x} \in \mathbb{C}^{2N-1}$, let $\boldsymbol{H}(\boldsymbol{x}) \in \mathbb{C}^{N\times N}$ be the corresponding Hankel matrix. We solve the following optimization problem:

$$\min_{\boldsymbol{x}} \|\boldsymbol{H}(\boldsymbol{x})\|_*, \qquad \text{subject to} \quad \mathscr{A}(\boldsymbol{x}) = \boldsymbol{b}, \tag{2.4}$$

where $\|\cdot\|_*$ is the nuclear norm which is the sum of all singular values of a matrix, and $\mathscr{A}$ and $\boldsymbol{b}$ are the linear measurement operator and measurement results. More specifically, $\mathscr{A}$ is a linear undersampling operator and $\boldsymbol{b} = \boldsymbol{x}_{\mathscr{M}}$ is the set of elements from $\boldsymbol{x}$ with indices in $\mathscr{M}$. When there is noise $\boldsymbol{\eta}$ contained in the observation, i.e.

$$\boldsymbol{b} = \mathscr{A}(\boldsymbol{x}) + \boldsymbol{\eta},$$

we solve

$$\min_{\boldsymbol{x}} \|\boldsymbol{H}(\boldsymbol{x})\|_*, \qquad \text{subject to} \quad \|\mathscr{A}(\boldsymbol{x}) - \boldsymbol{b}\|_2 \le \delta, \tag{2.5}$$

where $\delta = \|\boldsymbol{\eta}\|_2$ is the noise level. It is known that the nuclear norm minimization (2.4) can be transformed into a semidefinite program

$$\begin{aligned} &\min_{\boldsymbol{x},\boldsymbol{Q}_1,\boldsymbol{Q}_2} \frac{1}{2}(\mathrm{Tr}(\boldsymbol{Q}_1) + \mathrm{Tr}(\boldsymbol{Q}_2)) \\ &\text{s.t. } \mathscr{A}(\boldsymbol{x}) = \boldsymbol{b}, \\ &\begin{bmatrix} \boldsymbol{Q}_1 & \boldsymbol{H}(\boldsymbol{x})^* \\ \boldsymbol{H}(\boldsymbol{x}) & \boldsymbol{Q}_2 \end{bmatrix} \succeq 0 \end{aligned} \tag{2.6}$$

which can be easily solved with existing convex program solvers such as interior point algorithms.

After successfully recovering all the time samples, we can use the single-snapshot MUSIC algorithm (as discussed in [24]) to identify the underlying frequencies f_k. For the recovered Hankel matrix $\boldsymbol{H}(\boldsymbol{x})$, let its singular value decomposition (SVD) be

$$\boldsymbol{H}(\boldsymbol{x}) = [\boldsymbol{U}_1\ \boldsymbol{U}_2] \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} [\boldsymbol{V}_1\ \boldsymbol{V}_2]^*, \boldsymbol{U}_1, \boldsymbol{V}_1 \in \mathbb{C}^{N\times R}, \tag{2.7}$$

and we define the vector $\boldsymbol{\phi}^N(f)$ and imaging function $J(f)$ as

$$\boldsymbol{\phi}^N(f) = (1, e^{\imath 2\pi f}, ..., e^{\imath 2\pi f(N-1)})^T,\ J(f) = \frac{\|\boldsymbol{\phi}^N(f)\|_2}{\|\boldsymbol{U}_2^*\boldsymbol{\phi}^N(f)\|_2}, f \in [0,1). \tag{2.8}$$

The single-snapshot MUSIC algorithm is given in Algorithm 1. In [24], the authors showed that the MUSIC algorithm can exactly recover all the frequencies by finding the local maximal of $J(f)$. Namely, for undamped signal (1.1) with the set of frequencies $\mathscr{F}$ and $\tau_k = 0$, if $N \ge R$, then '$f \in \mathscr{F}$' is equivalent to '$J(f) = \infty$.'

While it is true that when $R < \frac{N}{2}$, one can use MUSIC to find the frequencies from the first $N-2$ samples and then recover the signals. However, we focus on the nuclear norm for the following reasons:

Algorithm 1 Single-Snapshot MUSIC algorithm 24

1: require: solution $\boldsymbol{x}$, parameter R and N
2: form Hankel matrix $\boldsymbol{Z} \in \mathbb{C}^{N\times N}$
3: SVD $\boldsymbol{Z} = [\boldsymbol{U}_1\ \boldsymbol{U}_2] \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} [\boldsymbol{V}_1\ \boldsymbol{V}_2]^*$ with $\boldsymbol{U}_1 \in \mathbb{C}^{N\times R}$ and $\boldsymbol{\Sigma}_1 \in \mathbb{C}^{R\times R}$
4: compute imaging function $J(f) = \frac{\|\boldsymbol{\phi}^N(f)\|_2}{\|\boldsymbol{U}_2^*\boldsymbol{\phi}^N(f)\|_2}, f \in [0, 1)$
5: get set $\hat{\mathscr{F}}$=$\{\, f : f$ corresponds to R largest local maxima of $J(f)\,\}$

(1) MUSIC does not apply to the case where $R > \frac{N}{2}$ if we use the first $N-2$ samples. However, the nuclear norm minimization can be more powerful than MUSIC in that it can recover more frequencies. For example, our results in Theorem 4.1 show that nuclear norm minimization can recover close to N frequencies, while MUSIC can only recover up to $\frac{N}{2}$ if it uses only the first $N-2$ samples.

(2) In addition, this paper focuses on using nuclear norm minimization, rather than MUSIC. So while the MUSIC can recover the same number ($\frac{N}{2}$) of frequencies *in the worst case* by taking the first $N-2$ samples, our results can still be useful in understanding the power of nuclear norm minimization.

3. Worst-case performance guarantees of separation-free super-resolution

In this section, we provide the worst-case performance guarantees of Hankel matrix recovery for recovering the superposition of complex exponentials. Namely, we provide the conditions under which the Hankel matrix recovery can uniformly recover the superposition of every possible R complex exponentials. Our results show that the Hankel matrix recovery can achieve separation-free super-resolution, even if we consider the criterion of worst-case performance guarantees. Later, we further show that our derived worst-performance guarantees are tight, namely we can find examples where nuclear norm minimization fails to recover the superposition of complex exponentials for a larger R. We start with introducing necessary notations and a few lemmas needed for proofs of our main results.

3.1 *Notations and Useful Lemmas*

Let w_i be the number of elements in the i-th anti-diagonal of any $N \times N$ matrix like $\boldsymbol{H}$, namely,

$$w_i = \begin{cases} i, i = 1, 2, \cdots, N, \\ 2N - i, i = N+1, \cdots, 2N-1. \end{cases} \tag{3.1}$$

Here, we call the $(2N-1)$ anti-diagonals of $\boldsymbol{H}(\boldsymbol{x})$ from the left top to the right bottom as the first anti-diagonal up to the $(2N-1)$-th anti-diagonal. For example, consider a Hankel matrix

$$\begin{bmatrix} h_1 & h_2 & h_3 & h_4 \\ h_2 & h_3 & h_4 & h_5 \\ h_3 & h_4 & h_5 & h_6 \\ h_4 & h_5 & h_6 & h_7 \end{bmatrix}, \tag{3.2}$$

then correspondingly, $w_1 = 1, w_2 = 2, w_3 = 3, w_4 = 4, w_5 = 3, w_6 = 2$ and $w_7 = 1$. We also define w_{min} as

$$w_{min} = \min_{i \in \{0,1,2,\ldots,2N-2\} \setminus \mathscr{M}} w_{i+1}, \tag{3.3}$$

where $\mathscr{M}$ is the observation set.

For a Hankel matrix

$$\boldsymbol{H}_{jk}(\boldsymbol{x}) = \boldsymbol{x}_{j+k-2}, j, k = 1, 2, \cdots, N, \tag{3.4}$$

we define $Ind(j)$ as the set of indices of rows which intersect with the j-th anti-diagonal of the Hankel matrix, and define $\{j : i \in Ind(j)\}$ as the set of indices of all non-zero anti-diagonals which intersect with the i-th row. For example, consider the matrix in (3.2), we have

$$\begin{aligned} &Ind(1) = \{1\}, Ind(2) = \{1, 2\}, Ind(3) = \{1, 2, 3\}, \\ &Ind(4) = \{1, 2, 3, 4\}, \\ &Ind(5) = \{2, 3, 4\}, Ind(6) = \{3, 4\}, Ind(7) = \{4\}, \end{aligned}$$

and

$$\begin{aligned} &\{j : 1 \in Ind(j)\} = \{1, 2, 3, 4\}, \\ &\{j : 2 \in Ind(j)\} = \{2, 3, 4, 5\}, \\ &\{j : 3 \in Ind(j)\} = \{3, 4, 5, 6\}, \\ &\{j : 4 \in Ind(j)\} = \{4, 5, 6, 7\}. \end{aligned}$$

With the setup above, we give the following Theorem 3.1 concerning the worst-case performance guarantee of Hankel matrix recovery, the proof of which is based on the strong null space conditions from [29, 36]. We will specifically use the following Lemma 3.1 from [29].

Lemma 3.1 is called the strong null space condition, and it characterizes the conditions under which the nuclear norm minimization (3.5) can recover any matrix with rank at most R. We use this lemma to show below that the recovery of $\boldsymbol{H}(x)$ from the observation set M is guaranteed by a condition that does not mandate sufficient frequency separation.

LEMMA 3.1. ([29]) Consider an $N \times N$ matrix $\boldsymbol{X}_0$ of rank at most R, and a linear mapping $\mathscr{A}(\boldsymbol{X}_0) = \boldsymbol{b}$. Then the nuclear norm minimization (3.5)

$$\min_{\boldsymbol{X}} \|\boldsymbol{X}\|_*, \qquad \text{subject to} \quad \mathscr{A}(\boldsymbol{X}) = \mathscr{A}(\boldsymbol{X}_0), \tag{3.5}$$

can uniquely and correctly recover every matrix $\boldsymbol{X}_0$ with rank no more than R if, for all non-zero $\boldsymbol{Z} \in \mathscr{N}(\mathscr{A})$,

$$2\|\boldsymbol{Z}\|_{*R} < \|\boldsymbol{Z}\|_*, \tag{3.6}$$

where $\|\boldsymbol{Z}\|_{*R}$ is the sum of the largest R singular values of $\boldsymbol{Z}$, and $\mathcal{N}(\mathscr{A})$ is the null space of $\mathscr{A}$.

When there is noise in the observation data, we have the following results guaranteeing the robustness of the nuclear norm minimization in recovering data.

LEMMA 3.2. Consider an $N \times N$ matrix $\boldsymbol{X}_0$ of rank at most R, and a linear mapping $\mathscr{B}$ such that $\mathscr{B}(\boldsymbol{X}_0 + \boldsymbol{E}) = \boldsymbol{b}$, where $\boldsymbol{E}$ is a perturbation matrix with $\|\boldsymbol{E}\|_* \leq \epsilon$. Then the nuclear norm minimization (3.7)

$$\min_{\boldsymbol{X}} \|\boldsymbol{X}\|_*, \qquad \text{subject to} \quad \mathscr{B}(\boldsymbol{X}) = \boldsymbol{b}, \tag{3.7}$$

will output a $\hat{\boldsymbol{X}}$ such that $\|\boldsymbol{X}_0 - \hat{\boldsymbol{X}}\|_* \leq \frac{3C+1}{C-1}\epsilon$, if for all non-zero $\boldsymbol{Z} \in \mathcal{N}(\mathscr{B})$,

$$(C+1)\|\boldsymbol{Z}\|_{*R} \leq \|\boldsymbol{Z}\|_*, \tag{3.8}$$

where $C > 1$ is a constant, $\|\boldsymbol{Z}\|_{*R}$ is the sum of the largest R singular values of $\boldsymbol{Z}$, and $\mathcal{N}(\mathscr{B})$ is the null space of $\mathscr{B}$.

Proof. We let $\hat{\boldsymbol{X}} = \boldsymbol{X}_0 + \boldsymbol{E} + \boldsymbol{W}$, where $\boldsymbol{W}$ is a non-zero matrix from the null space of $\mathscr{B}$. Then

$$\|\boldsymbol{X}_0\|_* + \|\boldsymbol{E}\|_* \tag{3.9}$$

$$\geq \|\boldsymbol{X}_0 + \boldsymbol{E}\|_* \tag{3.10}$$

$$\geq \|\boldsymbol{X}_0 + \boldsymbol{E} + \boldsymbol{W}\|_* \tag{3.11}$$

$$\geq \|\boldsymbol{X}_0 + \boldsymbol{W}\|_* - \|\boldsymbol{E}\|_* \tag{3.12}$$

$$\geq \sum_{i=1}^{N} |\sigma_i(\boldsymbol{X}_0) - \sigma_i(\boldsymbol{W})| - \|\boldsymbol{E}\|_* \tag{3.13}$$

$$\geq \sum_{i=1}^{R} (\sigma_i(\boldsymbol{X}_0) - \sigma_i(\boldsymbol{W})) + \sum_{i=R+1}^{N} \sigma_i(\boldsymbol{W}) - \|\boldsymbol{E}\|_* \tag{3.14}$$

$$\geq \|\boldsymbol{X}_0\|_* + \|\boldsymbol{W}\|_* - 2\|\boldsymbol{W}\|_{*R} - \|\boldsymbol{E}\|_* \tag{3.15}$$

$$\geq \|\boldsymbol{X}_0\|_* + \frac{C-1}{C+1}\|\boldsymbol{W}\|_* - \|\boldsymbol{E}\|_*, \tag{3.16}$$

where $\boldsymbol{W}$ is a matrix from the null space of $\mathscr{B}$, $\sigma_i(\cdot)$ represents the i-th largest singular value of a matrix and inequality (3.14) is due to $\boldsymbol{X}_0$'s rank being at most R and Lemma 3.3 as shown below. Thus, we have $2\|\boldsymbol{E}\|_* \geq \frac{C-1}{C+1}\|\boldsymbol{W}\|_*$, implying that $\|\boldsymbol{X}_0 - \hat{\boldsymbol{X}}\|_* = \|\boldsymbol{W} + \boldsymbol{E}\|_* \leq \|\boldsymbol{W}\|_* + \|\boldsymbol{E}\|_* \leq \frac{3C+1}{(C-1)}\|\boldsymbol{E}\|_*$. □

LEMMA 3.3. [29] Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two $N \times N$ matrix, then

$$\|\boldsymbol{X} - \boldsymbol{Y}\|_* \geq \sum_{i=1}^{N} |\sigma_i(\boldsymbol{X}) - \sigma_i(\boldsymbol{Y})|,$$

where $\sigma_i(\boldsymbol{X})$ and $\sigma_i(\boldsymbol{Y})$ are, respectively, the i-th largest singular values of $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively.

Because of the two lemmas above, we arrive at the following lemma particularly about the condition for Hankel matrix recovery in super-resolution.

LEMMA 3.4. Consider the signal model (1.1) with z_k defined in (2.1), and the observation set $\mathscr{M} \subseteq \{0, 1, 2, ..., 2N-2\}$. Then the nuclear norm minimization (2.4) will uniquely recover $\boldsymbol{H}(\boldsymbol{x})$, regardless of the (frequency) separation between the R continuously valued (frequency) parameters, if for every non-zero element in the null space $\mathscr{N}(\mathscr{A})$ of the linear mapping $\mathscr{A}$ corresponding to the sampling set $\mathscr{M}$ satisfies the condition in Lemma 3.1. Namely, every non-zero matrix in the set

$$\mathscr{N}(\mathscr{A}) = \{\boldsymbol{Z} \in \mathbb{C}^{N\times N} : \boldsymbol{Z} \text{ is a Hankel matrix, } \boldsymbol{Z}_{j,k} = 0 \text{ if } j+k-2 \in \mathscr{M}\}$$

satisfies the null space condition in Lemma 3.1.

This lemma shows that the guarantee of recovering missing data via Hankel matrix recovery essentially does not require frequency separations of frequencies, but instead depends on the sampling pattern only.

3.2 *Worst-case performance guarantees of separation-free super-resolution*

We are now ready to introduce our main results.

THEOREM 3.1. Consider the signal model (1.1) with z_k defined in (2.1), and the observation set $\mathscr{M} \subseteq \{0, 1, 2, ..., 2N-2\}$. We further define w_i of an $N \times N$ Hankel matrix $\boldsymbol{H}(\boldsymbol{x})$ as in (3.1), and define w_{min} as in (3.3). Then the nuclear norm minimization (2.4) will uniquely recover $\boldsymbol{H}(\boldsymbol{x})$, regardless of the (frequency) separation between the R continuously valued (frequency) parameters, if

$$R < \frac{w_{min}}{2(2N-1-|\mathscr{M}|)}. \tag{3.17}$$

On the one hand, the performance guarantees given in Theorem 3.1 can be conservative: for average-case performance guarantees, even when the number of complex exponentials R is bigger than predicted by Theorem 3.1, the Hankel matrix recovery can still recover the missing data, even though the sinusoids can be very close to each other. On the other hand, the bounds on recoverable sparsity level R given in Theorem 3.1 is tight for worst-case performance guarantees, as shown in the next section. We will show the tightness of recoverable sparsity guaranteed by Theorem 3.1, which is built on the developments in Section 4.1.

Proof of Theorem 3.1: We can change (2.4) to the following optimization problem:

$$\min_{\boldsymbol{B},\boldsymbol{x}} \|\boldsymbol{B}\|_*, \qquad \text{subject to} \quad \boldsymbol{B} = \boldsymbol{H}(\boldsymbol{x}), \ \ \mathscr{A}(\boldsymbol{x}) = \boldsymbol{b}. \tag{3.18}$$

We can see (3.18) as a nuclear norm minimization problem, where the null space of the linear operator (applied to $(\boldsymbol{B},\boldsymbol{x})$) in the constraints of (3.18) is given by $(\boldsymbol{H}(\boldsymbol{z}),\boldsymbol{z})$ such that $\mathscr{A}(\boldsymbol{z}) = 0$.

Using Lemma 3.1, we can see that (3.18) or (2.4) can correctly recover $\boldsymbol{x}$ as a superposition of R complex exponentials if $2\|\boldsymbol{H}(\boldsymbol{z})\|_{*R} < \|\boldsymbol{H}(\boldsymbol{z})\|_*$ holds true for every non-zero $\boldsymbol{z}$ from the null space of $\mathscr{A}$.

From the sampling set $\mathscr{M} \subseteq \{0, 1, 2, ..., 2N-2\}$, the null space of the sampling operator $\mathscr{A}$ is composed of $(2N-1)\times 1$ vectors $\boldsymbol{z}$'s such that $\boldsymbol{z}_{i+1} = 0$ if $i \in \mathscr{M}$. For such a vector $\boldsymbol{z}$, let us denote the element across the i-th anti-diagonal of $\boldsymbol{H}(\boldsymbol{z})$ as a_i, and thus $\boldsymbol{Q} = \boldsymbol{H}(\boldsymbol{z})$ is a Hankel matrix with its $(i+1)$ anti-diagonal element equal to 0 if $i \in \mathscr{M}$. Let $\sigma_1, \cdots, \sigma_N$ be the N singular values of $\boldsymbol{H}(\boldsymbol{z})$ arranged in a descending order. To verify the null space condition for nuclear norm minimization, we would like to find the largest R such that

$$\sigma_1 + \cdots + \sigma_R < \sigma_{R+1} + \cdots + \sigma_N, \tag{3.19}$$

for every non-zero $\boldsymbol{z}$ in the null space of $\mathscr{A}$.

Towards this goal, we first obtain an upper bound for its largest singular value (for a matrix $\boldsymbol{Q}$, we use $\boldsymbol{Q}_{i,:}$ to denote its i-th row vector):

$$\begin{aligned}
\sigma_1 &= \max_{\boldsymbol{u}\in\mathbb{C}^N, \|\boldsymbol{u}\|_2=1} \|\boldsymbol{Q}\boldsymbol{u}\|_2 \\
&= \max_{\boldsymbol{u}\in\mathbb{C}^N, \|\boldsymbol{u}\|_2=1} \sqrt{\sum_{i=1}^{N} |\boldsymbol{Q}_{i,:}\boldsymbol{u}|^2} \\
&= \max_{\boldsymbol{u}\in\mathbb{C}^N, \|\boldsymbol{u}\|_2=1} \sqrt{\sum_{i=1}^{N} \left| \sum_{j\in\{j:\ i\in Ind(j), a_j\neq 0\}} a_j u_{j-i+1} \right|^2},
\end{aligned} \tag{3.20}$$

where $Ind(j)$ is the set of indices of rows which intersect with the j-th anti-diagonal, $\{j : i \in Ind(j), a_j \neq 0\}$ is the set of all non-zero anti-diagonals intersecting with the i-th row and (3.20) is obtained by looking at the N rows. Note that in the calculations above, we only need to pay attention to the non-zero entries in the i-th row of $\boldsymbol{Q}$. From the Cauchy–Schwarz inequality, we have

$$\begin{aligned}
\sigma_1 &\leq \max_{\boldsymbol{u}\in\mathbb{C}^N, \|\boldsymbol{u}\|_2=1} \sqrt{\sum_{i=1}^{N} \left(\sum_{j\in\{j:i\in Ind(j), a_j\neq 0\}} |a_j|^2 \right) \left(\sum_{j\in\{j:i\in Ind(j), a_j\neq 0\}} |u_{j-i+1}|^2 \right)} \\
&= \max_{\boldsymbol{u}\in\mathbb{C}^N, \|\boldsymbol{u}\|_2=1} \sqrt{\sum_{j_1=1}^{2N-1} |a_{j_1}|^2 \sum_{i=1, i\in Ind(j_1)}^{N} \left(\sum_{j_2\in\{j:i\in Ind(j), a_j\neq 0\}} |u_{j_2-i+1}|^2 \right)} \\
&\leq \max_{\boldsymbol{u}\in\mathbb{C}^N, \|\boldsymbol{u}\|_2=1} \sqrt{\sum_{j_1=1}^{2N-1} \left[|a_{j_1}|^2 \left(\sum_{i\in Ind(j_1)} \sum_{j_2\in\{j:i\in Ind(j), a_j\neq 0\}} |u_{j_2-i+1}|^2 \right) \right]}.
\end{aligned} \tag{3.21}$$

Because $\|\boldsymbol{u}\|_2 = 1$ and, for each i, $|u_i|^2$ appears in $\left(\sum_{i\in Ind(j_1)} \sum_{j_2\in\{j:i\in Ind(j),a_j\neq 0\}} |\boldsymbol{u}_{j_2-i+1}|^2\right)$ for no more than $(2N-1-M)$ times, we have

$$\begin{aligned}\sigma_1 &\leq \max_{\boldsymbol{u}\in\mathbb{C}^N,\|\boldsymbol{u}\|_2=1} \sqrt{\sum_{j_1=1}^{2N-1}\left[|a_{j_1}|^2(2N-1-M)\right]} \\ &= \sqrt{\sum_{j\in\{0,1,\ldots,2N-2\}\setminus\mathscr{M}} |a_{j+1}|^2(2N-1-M)}.\end{aligned}$$

Furthermore, summing up the energy of the matrix $\boldsymbol{Q}$, we have

$$\sum_{i=1}^{N}\sigma_i^2 = \sum_{i\in\{0,1,\ldots,2N-2\}\setminus\mathscr{M}} |a_{i+1}|^2 w_{i+1}. \tag{3.22}$$

Since $\frac{\sigma_i}{\sigma_1} \geq \left(\frac{\sigma_i}{\sigma_1}\right)^2$, then for any integer $k \leq N$, we have

$$\begin{aligned}\frac{\sum_{i=1}^{N}\sigma_i}{\sum_{i=1}^{k}\sigma_i} &\geq \frac{\sum_{i=1}^{N}\sigma_i}{k\sigma_1} \\ &\geq \frac{\sum_{i=1}^{N}\sigma_i^2}{k\sigma_1^2} \\ &\geq \frac{\sum_{i\in\{0,1,\ldots,2N-2\}\setminus\mathscr{M}} |a_{i+1}|^2 w_{i+1}}{k\sum_{j\in\{0,1,\ldots,2N-2\}\setminus\mathscr{M}} |a_{j+1}|^2(2N-1-M)} \\ &\geq \frac{\left(\min_{i\in\{0,1,\ldots,2N-2\}\setminus\mathscr{M}} w_{i+1}\right)\cdot\sum_{i\in\{0,1,\ldots,2N-2\}\setminus\mathscr{M}} |a_{i+1}|^2}{k(2N-1-M)\cdot\sum_{j\in\{0,1,\ldots,2N-2\}\setminus\mathscr{M}} |a_{j+1}|^2} \\ &= \frac{\min_{i\in\{0,1,\ldots,2N-2\}\setminus\mathscr{M}} w_{i+1}}{k(2N-1-M)} \\ &= \frac{w_{min}}{k(2N-1-M)}.\end{aligned} \tag{3.23}$$

So if

$$\frac{\min_{i\in\{0,1,\ldots,2N-2\}\setminus\mathscr{M}} w_{i+1}}{R(2N-1-M)} > 2, \tag{3.24}$$

then for every non-zero vector $\boldsymbol{z}$ in the null space of $\mathscr{A}$, and the corresponding Hankel matrix $\boldsymbol{Q} = \boldsymbol{H}(z)$,

$$\sum_{i=1}^{N} \sigma_i > 2\sum_{i=1}^{R} \sigma_i. \tag{3.25}$$

It follows that, for any superposition of $R < \frac{w_{min}}{2(2N-1-M)}$ complex exponentials, we can correctly recover $\boldsymbol{x}$ over the whole set $\{0, 1, ..., 2N-2\}$ using the incomplete sampling set $\mathscr{M}$, regardless of the separations between different frequencies (or between the continuously valued parameters z_k's for damped complex exponentials). □

3.3 *Robustness of Hankel matrix recovery in super-resolution regardless of frequency separation*

We further have the following results of the robustness of the Hankel matrix recovery in super-resolution.

THEOREM 3.2. Consider a ground-truth signal model (1.1) with z_k defined in (2.1), and the observation set $\mathscr{M} \subseteq \{0, 1, 2, ..., 2N-2\}$. Suppose that the ground-truth signal is given by

$$\boldsymbol{x}_j^0 = \sum_{k=1}^{R} c_k z_k^j, \qquad j = 0, 1, \ldots, 2N-2, \tag{3.26}$$

and we denote the ground-truth vector $\boldsymbol{x}^0 = [\boldsymbol{x}_0^0, \boldsymbol{x}_1^0, ..., \boldsymbol{x}_{2N-2}^0]^T$. We further represent the noise-corrupted signal by

$$\boldsymbol{x}_j = \sum_{k=1}^{R} c_k z_k^j + \boldsymbol{e}_j, \qquad j = 0, 1, \ldots, 2N-2, \tag{3.27}$$

where $\boldsymbol{e} = [\boldsymbol{e}_0, \boldsymbol{e}_1, ..., \boldsymbol{e}_{2N-2}]^T$ is the noise vector. We define a partial-observation error vector $\tilde{\boldsymbol{e}} \in \mathbb{C}^{2N-1}$ such that $\tilde{\boldsymbol{e}}_{\mathscr{M}} = \boldsymbol{e}_{\mathscr{M}}$ and $\tilde{\boldsymbol{e}}_{\{0,1,2,...,2N-2\}\setminus\mathscr{M}} = \boldsymbol{0}$. We define $\epsilon = N\|\tilde{\boldsymbol{e}}\|_1$, and denote the measurement results by $\mathscr{A}(\boldsymbol{x}^0) + \tilde{\boldsymbol{e}}_{\mathscr{M}} = \boldsymbol{b}$ or $\mathscr{A}(\boldsymbol{x}^0 + \boldsymbol{e}) = \boldsymbol{b}$.

Then the optimization formulation

$$\min_{\boldsymbol{x}} \|\boldsymbol{H}(\boldsymbol{x})\|_*, \qquad \text{subject to} \quad \mathscr{A}(\boldsymbol{x}) = \boldsymbol{b}, \tag{3.28}$$

will output a $\hat{\boldsymbol{x}}$ such that

$$\|\boldsymbol{H}(\hat{\boldsymbol{x}}) - \boldsymbol{H}(\boldsymbol{x}^0)\|_* \leq \frac{3C+1}{(C-1)}\epsilon,$$

and

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}^0\|_2 \leq \frac{3C+1}{(C-1)}\epsilon,$$

where $C > 1$, regardless of the (frequency) separation between the R continuously valued (frequency) parameters, if

$$R < \frac{w_{min}}{(C+1)(2N-1-|\mathscr{M}|)}. \tag{3.29}$$

Proof. The proof is similar to the proof of Theorem 3.1. Instead of using Lemma 3.1, we use Lemma 3.2.

First of all, we observe that the nuclear norm of the error matrix $\boldsymbol{E} = \boldsymbol{H}(\tilde{\boldsymbol{e}})$ is upper bounded by $N \times \|\tilde{\boldsymbol{e}}\|_1$. This is because the Hankel matrix, which has non-zero elements only in the i-th ($1 \le i \le 2N-1$) anti-diagonal and zeros elsewhere, has its nuclear norm bounded by the number of non-zero elements (which is upper bounded by N) times the magnitude of that non-zero element. By the triangle inequality, we have the nuclear norm of the error matrix $\boldsymbol{E} = \boldsymbol{H}(\tilde{\boldsymbol{e}})$ upper bounded by $N \times \|\tilde{\boldsymbol{e}}\|_1$.

We let $\boldsymbol{X}_0 = \boldsymbol{H}(\boldsymbol{x}^0)$ and $\boldsymbol{E} = \boldsymbol{H}(\tilde{\boldsymbol{e}})$. Then the optimization formulation

$$\min_{\boldsymbol{x}} \|\boldsymbol{H}(\boldsymbol{x})\|_*, \qquad \text{subject to} \quad \mathscr{A}(\boldsymbol{x}) = \boldsymbol{b}, \tag{3.30}$$

is equivalent to

$$\min_{\boldsymbol{X} \in \mathbb{C}^{N \times N}} \|\boldsymbol{X}\|_*, \qquad \text{subject to} \quad \mathscr{B}(\boldsymbol{X}) = \mathscr{B}(\boldsymbol{X}_0 + \boldsymbol{E}), \tag{3.31}$$

where $\mathscr{B}(\cdot)$ is a linear mapping such that its null space is $\{\boldsymbol{H}(\boldsymbol{x}) : \mathscr{A}(\boldsymbol{x}) = \boldsymbol{0}, \boldsymbol{x} \in \mathbb{C}^{2N-1}\}$.

Then we can apply Lemma 3.2 to obtain the error bounds as in this theorem. In fact, following similar proofs as in the proof of Theorem 3.1, we have, if

$$\frac{\min_{i \in \{0,1,\ldots,2N-2\} \setminus \mathscr{M}} w_{i+1}}{R(2N-1-M)} \ge C+1, \tag{3.32}$$

then for every non-zero vector $\boldsymbol{z}$ in the null space of $\mathscr{A}$, and the corresponding Hankel matrix $\boldsymbol{Q} = \boldsymbol{H}(\boldsymbol{z})$, we have

$$\sum_{i=1}^{N} \sigma_i \ge (C+1) \sum_{i=1}^{R} \sigma_i. \tag{3.33}$$

Applying Lemma 3.2, we immediately have

$$\|\boldsymbol{H}(\hat{\boldsymbol{x}}) - \boldsymbol{H}(\boldsymbol{x}^0)\|_* \le \frac{3C+1}{(C-1)} \epsilon.$$

Because

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}^0\|_2 \le \|\boldsymbol{H}(\hat{\boldsymbol{x}}) - \boldsymbol{H}(\boldsymbol{x}^0)\|_F \le \|\boldsymbol{H}(\hat{\boldsymbol{x}}) - \boldsymbol{H}(\boldsymbol{x}^0)\|_*,$$

where $\|\cdot\|$ denotes the Frobenius norm, we have

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}^0\|_2 \leq \frac{3C+1}{C-1}\epsilon.$$

□

3.4 *Implications for robustness in recovering frequencies using recovered data samples*

Let us now consider the recovery of the underlying frequencies. In fact, the robustness of recovering missing data samples translates into the robustness of recovering the frequencies. When the observations are noiseless, for the first time, we have shown that the nuclear norm minimization method can recover the data correctly regardless of frequency separation, and moreover, using the recovered data samples and the Prony's method (or the single snapshot MUSIC algorithm), we can even *exactly* recover the frequency accurately regardless of frequency separation. When the observations are noisy, the Hankel matrix recovery method can guarantee recovering the missing data samples with robustness: regardless of frequency separations, the norms of the recovery error are nicely bounded by the norms of the noise vector, as proved by Theorem 3.2 in this paper. According to theorems 3 and 4 in [24], we know that when the error in the recovered data samples goes to zero, using the single-snapshot MUSIC algorithm, the error in the noise-space correlation function used in the single-snapshot algorithm also goes to zero; and, moreover, there exist local minimizers of the noise-space correlation function converging to the true underlying frequencies. Of course, the ratio of the magnitude of frequency recovery error to the magnitude of noise depends on frequency separations; however, at least this method is robust against noises in the sense that the magnitude of frequency recovery error goes to zero as the noise's magnitude goes to zero. This is very different from atomic norm minimization, which will identify wrong frequencies even with noiseless and complete data if the separation of two frequencies is below a certain threshold [40]. Our results in this paper demonstrate that Hankel matrix recovery method exhibits very different behaviours than the atomic norm minimization, even though both reduce to a semidefinite programming problem for minimizing the nuclear norm of a matrix.

4. Average-case performance guarantees

In this section, we study the performance guarantees for Hankel matrix recovery, when the phases of the coefficients of the R sinusoids are i.i.d. and uniformly distributed over $[0, 2\pi)$. For average-case performance guarantees, we show that we can recover the superposition of a larger number of complex exponentials than Theorem 3.1 offers. We introduce Lemmas 4.1 and 4.2 first, which are used in the derivations of the average-case recovery guarantee stated in Theorems 4.1 and 4.3.

LEMMA 4.1. ([42]) For a sequence of i.i.d. random matrices $\boldsymbol{M}_1, \cdots, \boldsymbol{M}_K$ with dimension $d_1 \times d_2$ and their sum $\boldsymbol{M} = \sum_{i=1}^K \boldsymbol{M}_i$, if $\boldsymbol{M}_i$ satisfies

$$\mathbb{E}[\boldsymbol{M}_i] = \boldsymbol{0}, \|\boldsymbol{M}_i\| \leq L, \forall i = 1, \cdots, K, \tag{4.1}$$

and

$$\nu(\boldsymbol{M}) := \max\left\{\left\|\sum_{i=1}^K \mathbb{E}[\boldsymbol{M}_i\boldsymbol{M}_i^*]\right\|, \left\|\sum_{i=1}^K \mathbb{E}[\boldsymbol{M}_i^*\boldsymbol{M}_i]\right\|\right\}, \tag{4.2}$$

then

$$\mathbb{P}(\|\boldsymbol{M}\| \geq t) \leq (d_1 + d_2) \cdot \exp\left(\frac{-t^2/2}{v(\boldsymbol{M}) + Lt/3}\right), \forall t \geq 0. \tag{4.3}$$

Lemma 4.1, called the Matrix Bernstein Inequality, characterizes the tail behaviour of the spectral norm of a random matrix.

LEMMA 4.2. Let $\boldsymbol{X}_0$ be any $M \times N$ matrix of rank R in $\mathbb{C}^{M\times N}$, and we observe it through a linear mapping $\mathscr{A}(\boldsymbol{X}_0) = \boldsymbol{b}$. We also assume that $\boldsymbol{X}_0$ has an SVD $\boldsymbol{X}_0 = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*$, where $\boldsymbol{U} \in \mathbb{C}^{M\times R}$, $\boldsymbol{V} \in \mathbb{C}^{N\times R}$, and $\boldsymbol{\Sigma} \in \mathbb{C}^{R\times R}$ is a diagonal matrix. Then the nuclear norm minimization (4.4)

$$\min_{\boldsymbol{X}} \|\boldsymbol{X}\|_*, \qquad \text{subject to} \quad \mathscr{A}(\boldsymbol{X}) = \mathscr{A}(\boldsymbol{X}_0), \tag{4.4}$$

correctly and uniquely recovers $\boldsymbol{X}_0$ if, for every non-zero element $\boldsymbol{Q} \in \mathscr{N}(\mathscr{A})$,

$$-|\mathrm{Tr}(\boldsymbol{U}^*\boldsymbol{Q}\boldsymbol{V})| + \|\bar{\boldsymbol{U}}^*\boldsymbol{Q}\bar{\boldsymbol{V}}\|_* > 0, \tag{4.5}$$

where $\bar{\boldsymbol{U}}$ and $\bar{\boldsymbol{V}}$ are such that $[\boldsymbol{U}\ \bar{\boldsymbol{U}}]$ and $[\boldsymbol{V}\ \bar{\boldsymbol{V}}]$ are unitary.

We use Lemma 4.2 as the condition for successful signal recovery through nuclear norm minimization. This lemma is an extension of lemma 13 in [29]. The key difference is that Lemma 4.2 deals with complex-numbered matrices. Moreover, Lemma 4.2 gets rid of the 'iff' claim for the null space condition in lemma 13 of [29], because we find that the condition in lemma 13 of [29] is a sufficient condition for the success of nuclear norm minimization, but not a necessary condition for the success of nuclear norm minimization [46]. We give the proof of Lemma 4.2 in Appendix.

4.1 *Average-case performance guarantees for orthogonal frequency atoms*

THEOREM 4.1. Let us consider the signal model of the superposition of R complex exponentials (1.1) with z_k defined in (2.1) where $\tau_k = 0$. We assume that the R frequencies f_1, f_2,..., and f_R are such that the atoms $(1, e^{\imath 2\pi f_i}, ..., e^{\imath 2\pi f_i(N-1)})^T$, $1 \leq i \leq R$, are orthogonal to each other. We let the observation set be $\mathscr{M} = \{0, 1, 2, ..., 2N-2\} \setminus \{N-1\}$. We assume the phases of coefficients $c_1, \cdots, c_R$ in signal model (1.1) are independent and uniformly distributed over $[0, 2\pi)$. Then the nuclear norm minimization (2.4) will successfully and uniquely recover $\boldsymbol{H}(\boldsymbol{x})$ and $\boldsymbol{x}$, with probability approaching 1 as $N \to \infty$ if

$$R = N - c\sqrt{\log(N)N}, \tag{4.6}$$

where $c > 0$ is a constant.

Proof of Theorem 4.1: We can change (2.4) to the following optimization problem

$$\min_{\boldsymbol{B},\boldsymbol{x}} \|\boldsymbol{B}\|_*, \qquad \text{subject to} \quad \boldsymbol{B} = \boldsymbol{H}(\boldsymbol{x}), \ \ \mathscr{A}(\boldsymbol{x}) = \boldsymbol{b}. \tag{4.7}$$

We can think of (4.7) as a nuclear norm minimization, where the null space of the linear operator (applied to $\boldsymbol{B}$ and $\boldsymbol{x}$) in the constraints of (3.18) is given by $(\boldsymbol{H}(\boldsymbol{z}), \boldsymbol{z})$ such that $\mathscr{A}(\boldsymbol{z}) = 0$.

Then we can see that (4.7) or (2.4) can correctly recover $\boldsymbol{x}$ as a superposition of R complex exponentials if

$$|\mathrm{Tr}(\boldsymbol{V}\boldsymbol{U}^*\boldsymbol{H}(\boldsymbol{z}))| < \|\bar{\boldsymbol{U}}^*\boldsymbol{H}(\boldsymbol{z})\bar{\boldsymbol{V}}\|_*, \tag{4.8}$$

hold true for any $\boldsymbol{z}$ which is a non-zero vector from the null space of $\mathscr{A}$, where $\boldsymbol{H}(\boldsymbol{x}) = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*$ is the SVD of $\boldsymbol{H}(\boldsymbol{x})$ with $\boldsymbol{U} \in \mathbb{C}^{N\times R}$ and $\boldsymbol{V} \in \mathbb{C}^{N\times R}$, and $\bar{\boldsymbol{U}}$ and $\bar{\boldsymbol{V}}$ are such that $[\boldsymbol{U}, \bar{\boldsymbol{U}}]$ and $[\boldsymbol{V}, \bar{\boldsymbol{V}}]$ are unitary.

Without loss of generality, let $f_k = \frac{s_k}{N}$ for $1 \le k \le R$, where s_k's are distinct integers between 0 and $N-1$. Then

$$\boldsymbol{U} = \frac{1}{\sqrt{N}}\begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{\imath 2\pi \frac{s_1}{N}} & e^{\imath 2\pi \frac{s_2}{N}} & \cdots & e^{\imath 2\pi \frac{s_R}{N}} \\ \vdots & \vdots & \ddots & \vdots \\ e^{\imath 2\pi \frac{s_1}{N}(N-1)} & e^{\imath 2\pi \frac{s_2}{N}(N-1)} & \cdots & e^{\imath 2\pi \frac{s_R}{N}(N-1)} \end{bmatrix}\begin{bmatrix} e^{-\imath\theta_1} & 0 & \cdots & 0 \\ 0 & e^{-\imath\theta_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{-\imath\theta_R} \end{bmatrix}, \tag{4.9}$$

and

$$\boldsymbol{V} = \frac{1}{\sqrt{N}}\begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{-\imath 2\pi \frac{s_1}{N}} & e^{-\imath 2\pi \frac{s_2}{N}} & \cdots & e^{-\imath 2\pi \frac{s_R}{N}} \\ \vdots & \vdots & \ddots & \vdots \\ e^{\imath -2\pi \frac{s_1}{N}(N-1)} & e^{-\imath 2\pi \frac{s_2}{N}(N-1)} & \cdots & e^{-\imath 2\pi \frac{s_R}{N}(N-1)} \end{bmatrix}, \tag{4.10}$$

and

$$\boldsymbol{\Sigma} = N\begin{bmatrix} |c_1| & 0 & \cdots & 0 \\ 0 & |c_2| & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & |c_R| \end{bmatrix}, \tag{4.11}$$

where θ_k's ($1 \le k \le R$) are i.i.d. random variables uniformly distributed over $[0, 2\pi)$.

When the observation set $\mathscr{M} = \{0, 1, 2, ..., 2N-2\} \setminus \{N-1\}$, any $\boldsymbol{Q} = \boldsymbol{H}(\boldsymbol{z})$ with $\boldsymbol{z}$ from the null space of $\mathscr{A}$ takes the following form:

$$\boldsymbol{Q} = a\begin{bmatrix} 0 & \cdots & 1 \\ \vdots & \iddots & \vdots \\ 1 & \cdots & 0 \end{bmatrix}, a \in \mathbb{C}. \tag{4.12}$$

Thus

$$|\mathrm{Tr}(\boldsymbol{V}\boldsymbol{U}^*\boldsymbol{Q})| = \left| a\mathrm{Tr}\left(\boldsymbol{U}^* \begin{bmatrix} 0 & \cdots & 1 \\ \vdots & \iddots & \vdots \\ 1 & \cdots & 0 \end{bmatrix} \boldsymbol{V} \right) \right|$$

$$= |a| \left(\frac{1}{\sqrt{N}}\right)^2 \left| \sum_{i=1}^{R} \sum_{t=0}^{N-1} e^{\iota\theta_i} e^{-\iota 2\pi \frac{s_i(N-1-t)}{N}} \times e^{-\iota 2\pi \frac{s_i t}{N}} \right| \tag{4.13}$$

$$= \frac{|a|}{N} \left| \sum_{i=1}^{R} \sum_{t=0}^{N-1} e^{\iota\theta_i} e^{-\iota 2\pi \frac{s_i(N-1)}{N}} \right| \tag{4.14}$$

$$= |a| \left| \sum_{i=1}^{R} e^{\iota\theta_i} e^{-\iota 2\pi \frac{s_i(N-1)}{N}} \right|. \tag{4.15}$$

Notice that random variable $e^{\iota\theta_i} e^{-\iota 2\pi \frac{s_i(N-1)}{N}}$ are mutually independent random variables uniformly distributed over the complex unit circle. We will use the concentration lemma (see for example, [42]) to provide a concentration of measure result for the summation of $e^{\iota\theta_i} e^{-\iota 2\pi \frac{s_i(N-1)}{N}}$.

Applying Lemma 4.1 to the 1×2 matrix composed of the real and imaginary parts of $e^{\iota\theta_i} e^{-\iota 2\pi \frac{s_i(N-1)}{N}}$, with $\nu = \max(R, R) = R$, $L = 1$, $d_1 + d_2 = 3$, we have

$$\mathbb{P}\left(|\mathrm{Tr}(\boldsymbol{V}\boldsymbol{U}^*\boldsymbol{Q})| \geq |a|t\right) \leq 3e^{-\frac{t^2/2}{\nu + Lt/3}} = 3e^{-\frac{t^2/2}{R + t/3}}, \forall t > 0. \tag{4.16}$$

We further notice that, $\bar{\boldsymbol{U}}$'s ($\bar{\boldsymbol{V}}$' s) columns are normalized orthogonal frequency atoms (or their complex conjugates) with frequencies $\frac{l_i}{N}$, with integer l_i's different from the integers s_k's of those R complex exponentials. Thus,

$$\|\bar{\boldsymbol{U}}^*\boldsymbol{Q}\bar{\boldsymbol{V}}\|_* = |a|(N - R) \tag{4.17}$$

since we can always find two unitary matrices $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ such that $\boldsymbol{Q}_1\boldsymbol{Q}\boldsymbol{Q}_2$ becomes an identity matrix without changing the nuclear norm, i.e. $\|\bar{\boldsymbol{U}}^*\boldsymbol{Q}\bar{\boldsymbol{V}}\|_* = \|\bar{\boldsymbol{U}}^*\boldsymbol{Q}_1\boldsymbol{Q}\boldsymbol{Q}_2\bar{\boldsymbol{V}}\|_*$, allowing us to get (4.17).

Pick $t = N - R$, and let $\frac{(N-R)^2/2}{R+(N-R)/3} = c\log(N)$ with c being a positive constant. Solving for R, we obtain $R = N - \left(\frac{2}{3}\log(N) + \sqrt{\frac{4}{9}c^2\log^2(N) + 2c\log(N)N}\right)$, which implies (4.8) holds with probability approaching 1 if $N \to \infty$. This proves our claims. □

4.2 *Tightness of Theorem 3.1*

Building on the discussions above, we now discuss some results on the tightness of the bounds in Theorem 3.1.

We first show that the recoverable sparsity R provided by Theorem 3.1 is tight for $\mathscr{M} = \{0, 1, 2, ..., 2N - 2\} \setminus \{N - 1\}$. For such a sampling set, $w_{min} = N$, and Theorem 3.1 provides a bound $R < \frac{w_{min}}{2} = \frac{N}{2}$.

In fact, we can show that if $R \geq \frac{N}{2}$, we can construct signal examples where the Hankel matrix recovery approach cannot recover the original signal $\boldsymbol{x}$. Consider the signal in Theorem 4.1. We choose the coefficients c_i's such that $e^{\iota\theta_i} e^{-\iota 2\pi \frac{s_i(N-1)}{N}} = 1$, then

$$|\mathrm{Tr}(\boldsymbol{V}\boldsymbol{U}^*\boldsymbol{Q})| = |a| \left| \sum_{i=1}^{R} e^{\iota\theta_i} e^{-\iota 2\pi \frac{s_i(N-1)}{N}} \right| = |a|R, \tag{4.18}$$

and we also have

$$\|\bar{\boldsymbol{U}}^* \boldsymbol{Q} \bar{\boldsymbol{V}}\|_* = |a|(N-R). \tag{4.19}$$

Thus, $|\mathrm{Tr}(\boldsymbol{V}\boldsymbol{U}^*\boldsymbol{Q})| \geq \|\bar{\boldsymbol{U}}^* \boldsymbol{Q} \bar{\boldsymbol{V}}\|_*$ for every $\boldsymbol{Q} = \boldsymbol{H}(\boldsymbol{z})$ with $\boldsymbol{z}$ in the null space of the sampling operator. Thus, the Hankel matrix recovery cannot recover the ground truth $\boldsymbol{x}$. This shows that the prediction by Theorem 3.1 is tight.

According to Theorem 3.1, *in the worst case*, we need a very large number of samples to guarantee successful recovery of signal comprising a small number of atoms. We have established the tightness of Theorem 3.1 for a special sampling scheme, i.e. $\mathscr{M} = \{0, 1, \cdots, 2N-1\} \setminus \{N-1\}$, and even under the restricted sampling patterns considered in this paper, these are the first results in the literature that show successful recovery is guaranteed regardless of frequency separation using nuclear norm minimization in the setting of noiseless measurements.

In fact, we have the following general results suggesting that the tightness of the results in Theorem 3.1 in terms of recovering a low-rank matrix.

THEOREM 4.2. Consider the signal model (1.1) with z_k defined in (2.1), and the observation set $\mathscr{M} \subseteq \{0, 1, 2, ..., 2N-2\}$. We further define w_i of an $N \times N$ Hankel matrix $\boldsymbol{H}(\boldsymbol{x})$ as in (3.1), and define w_{min} as in (3.3). Let $\mathscr{A}(\cdot)$ be a linear mapping of a vector from $\mathbb{C}^{2N-1}$ to $\mathbb{C}^{|\mathscr{M}|}$ as defined by the sampling pattern $\mathscr{M}$.

Then there is a non-zero vector $\mathbf{u} \in \mathbb{C}^{2N-1} \in$ such that $\mathscr{A}(\mathbf{u}) = 0$ and the corresponding Hankel matrix $\boldsymbol{H}(\mathbf{u})$ has a rank of w_{min}; moreover, there is a matrix $\mathbf{Q} \in \mathbb{C}^{N\times N}$ of rank $\lceil w_{min}/2 \rceil$ such that $\|\boldsymbol{H}(\mathbf{u}) + \mathbf{Q}\|_* \leq \|\mathbf{Q}\|_*$.

In other words, let $\mathscr{B}$ be linear mapping from $\mathbb{C}^{N\times N}$ to $\mathbb{C}^{|\mathscr{M}|}$ such that its null space is the same as the set $\{\boldsymbol{H}(\mathbf{u}) \mid \mathscr{A}(\mathbf{u}) = 0\}$, then there exists a low-rank matrix $\mathbf{Q}$ of rank $\lceil w_{min}/2 \rceil$ such that it is not the unique solution to the following nuclear norm minimization problem:

$$\min_{\boldsymbol{X} \in \mathbb{C}^{N\times N}} \|\boldsymbol{X}\|_*, \qquad \text{subject to} \quad \mathscr{B}(\boldsymbol{X}) = \mathscr{B}(\mathbf{Q}). \tag{4.20}$$

Proof. We consider the index i, $0 \leq i \leq (2N-2)$ such that $w_{i+1} = w_{min}$. Then we take the low-rank matrix $\mathbf{Q}$ which has $\lceil w_{min}/2 \rceil$ 1's in its $(i+1)$-th anti-diagonal, and all the other elements in $\mathbf{Q}$ are 0. We take $\mathbf{u} \in \mathbb{C}^N$ such that $\mathbf{u}_i = -1$ at index i but equal to 0 at all the other indices. Then $\boldsymbol{H}(\mathbf{u}) + \mathbf{Q}$ will be a matrix having $w_{min} - \lceil w_{min}/2 \rceil$ '-1's at its $(i+1)$-th diagonal, and have a no bigger nuclear norm than $\mathbf{Q}$, proving this theorem. □

We also want to emphasize that while results in our theorems are conservative, and the practical performance can still be quite encouraging. For example, in Theorem 4.3, the sampling scheme can hardly be regarded as undersampling since we need $\frac{2N-2}{2N-1} = \frac{126}{127} \approx 99\%$ of the samples for successful

recovery. But in the experiments, we actually used about 51% of samples. The experimental results actually imply the possibility of improving the bounds in the recovery guarantees.

4.3 *Average-case performance analysis with arbitrarily close frequency atoms*

In this section, we further show that the Hankel matrix recovery can successfully recover the superposition of complex exponentials with arbitrarily close frequency atoms. In particular, we give average performance guarantees on recoverable sparsity R when frequency atoms are arbitrarily close, and show that the Hankel matrix recovery can deal with much larger recoverable sparsity R for average-case signals than predicted by Theorem 3.1. For the tractability of our theoretical analysis, our results assume that certain frequency atoms are orthogonal to each other; however, in our results, there are always frequency atoms which are arbitrarily close to each other, enforcing small frequency separation in this research problem. We note that in the literature the separation requirement is often defined as the smallest circular separation between frequency atoms. Following this definition, the frequency separations used in this section are still consistent with the definition of small frequency separation, even though some of the frequency atoms are orthogonal to each other. Numerical results suggest that the assumptions that some of the frequency atoms are orthogonal to each other are not necessarily needed; however, removing such assumptions would require some future work.

We consider a signal $\boldsymbol{x}$ composed of R complex exponentials. Among them, $R-1$ orthogonal complex exponentials (without loss of generality, we assume that these $R-1$ frequencies take values $\frac{l_i}{N}$, where l_i's are integers). The other complex exponential $c_{cl}e^{\iota 2\pi f_{cl}j}$ has a frequency arbitrarily close to one of the $R-1$ frequencies, i.e.

$$\boldsymbol{x}_j = \sum_{k=1}^{R-1} c_k e^{\iota 2\pi f_k j} + c_{cl} e^{\iota 2\pi f_{cl} j},\ j \in \{0, 1, \cdots, 2N-2\}, \tag{4.21}$$

where f_{cl} is arbitrarily close to one of the first $(R-1)$ frequencies. For this set-up with arbitrarily close atoms, we have the following average-case performance guarantee.

THEOREM 4.3. Consider the signal model (4.21) with R complex exponentials, where the phases of the coefficients in these complex exponentials are i.i.d. uniformly randomly distributed over $[0, 2\pi)$. The first $(R-1)$ of the complex exponentials are such that the corresponding atom vectors

$$(1, e^{\iota 2\pi f_i},\ ...,\ e^{\iota 2\pi f_i (N-1)})^T$$

are mutually orthogonal, while the R-th exponential has a frequency arbitrarily close to one of the first $(R-1)$ frequencies (in wrap-around distance). Let $c_{min} = \min\{|c_1|, |c_2|, ..., |c_{R-1}|\}$, and define

$$d_{rel} = \frac{|c_{cl}|}{c_{min}} + 1. \tag{4.22}$$

If we have the observation set $\mathscr{M} = \{0, 1, 2, ..., 2N-2\} \setminus \{N-1\}$, and for any constant $c > 0$, if

$$R = N - 4\sqrt{N} d_{rel} + \frac{2}{3} c \log(N) - \sqrt{2cN\log(N) - 8c\sqrt{N} d_{rel}\log(N) + \frac{4}{9}c^2 \log^2(N)}, \tag{4.23}$$

then we can recover the true signal $\boldsymbol{x}$ via Hankel matrix recovery

$$\min_{\boldsymbol{x}} \|\boldsymbol{H}(\boldsymbol{x})\|_*, \text{ subject to } \mathscr{A}(\boldsymbol{x}) = \boldsymbol{b}$$

with high probability as $N \to \infty$, regardless of frequency separations.

Remarks: 1. We can extend this result to cases where neither of the two arbitrarily close frequencies has on-the-grid frequencies, and the Hankel matrix recovery method can recover a similar sparsity; 2. Under a similar number of complex exponentials, with high probability, the Hankel matrix recovery can correctly recover the signal $\boldsymbol{x}$, uniformly over every possible phases of the two complex exponentials with arbitrarily close frequencies; 3. We can extend our results to other sampling sets, but for clarity of presentations, we choose $\mathscr{M} = \{0, 1, 2, ..., 2N-2\} \setminus \{N-1\}$.

The proof of Theorem 4.3 depends on Lemma 4.3, a perturbation result on the polar decomposition of a matrix, which we will present first before introducing the formal proof of Theorem 4.3. Let us consider two matrices $\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$ from $\mathbb{C}_r^{m\times n}$ such that $\boldsymbol{X} = \tilde{\boldsymbol{X}} + \boldsymbol{E}$, where $\boldsymbol{E} \in \mathbb{C}^{m\times n}$. Suppose that the matrix $\boldsymbol{X}$ has SVD

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*, \boldsymbol{U} = [\boldsymbol{U}_1\ \boldsymbol{U}_2] \in \mathbb{C}^{m\times m}, \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \in \mathbb{C}^{m\times n}, \boldsymbol{V} = [\boldsymbol{V}_1\ \boldsymbol{V}_2] \in \mathbb{C}^{n\times n}, \tag{4.24}$$

where $\boldsymbol{U}_1 \in \mathbb{C}^{m\times r}$, $\boldsymbol{V}_1 \in \mathbb{C}^{n\times r}$ and $\boldsymbol{\Sigma}_1 \in \mathbb{C}^{r\times r}$. The polar decomposition of $\boldsymbol{X}$ is given by

$$\boldsymbol{X} = \boldsymbol{P}\boldsymbol{H}, \boldsymbol{P} = \boldsymbol{U}_1\boldsymbol{V}_1^*, \boldsymbol{H} = \boldsymbol{V}_1\boldsymbol{\Sigma}_1\boldsymbol{V}_1^*, \tag{4.25}$$

and the matrix $\boldsymbol{P}$ is the unitary polar factor. Similarly for $\tilde{\boldsymbol{X}}$, we have

$$\tilde{\boldsymbol{X}} = \tilde{\boldsymbol{U}}\tilde{\boldsymbol{\Sigma}}\boldsymbol{V}^*, \tilde{\boldsymbol{U}} = [\tilde{\boldsymbol{U}}_1\ \tilde{\boldsymbol{U}}_2] \in \mathbb{C}^{m\times m}, \tilde{\boldsymbol{\Sigma}} = \begin{bmatrix} \tilde{\boldsymbol{\Sigma}}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \in \mathbb{C}^{m\times n}, \tilde{\boldsymbol{V}} = [\tilde{\boldsymbol{V}}_1\ \tilde{\boldsymbol{V}}_2] \in \mathbb{C}^{n\times n}, \tag{4.26}$$

and the polar decomposition of $\tilde{\boldsymbol{X}}$

$$\tilde{\boldsymbol{X}} = \tilde{\boldsymbol{P}}\tilde{\boldsymbol{H}}, \tilde{\boldsymbol{P}} = \tilde{\boldsymbol{U}}_1\tilde{\boldsymbol{V}}_1^*, \tilde{\boldsymbol{H}} = \tilde{\boldsymbol{V}}_1\tilde{\boldsymbol{\Sigma}}_1\tilde{\boldsymbol{V}}_1^*. \tag{4.27}$$

LEMMA 4.3. ([23]) For matrices $\boldsymbol{X} \in \mathbb{C}_r^{m\times n}$ and $\tilde{\boldsymbol{X}} \in \mathbb{C}_r^{m\times n}$ with SVD defined as (4.24) and (4.26), let σ_r and $\tilde{\sigma}_r$ be the smallest non-zero singular values of $\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$, respectively, then

$$|||\boldsymbol{P} - \tilde{\boldsymbol{P}}||| \leq \left(\frac{2}{\sigma_r + \tilde{\sigma}_r} + \frac{2}{\max\{\sigma_r, \tilde{\sigma}_r\}}\right) |||\boldsymbol{X} - \tilde{\boldsymbol{X}}|||, \tag{4.28}$$

where $|||\cdot|||$ is any unitary invariant norm, and the matrices $\boldsymbol{P}$ and $\tilde{\boldsymbol{P}}$ are defined in (4.25) and (4.27).

Proof of Theorem 4.3: To prove this theorem, we consider a perturbed signal $\tilde{\boldsymbol{x}}$. The original signal $\boldsymbol{x}$ and the perturbed $\tilde{\boldsymbol{x}}$ satisfy

$$\boldsymbol{x}_j = \tilde{\boldsymbol{x}}_j + c_{cl}e^{\iota 2\pi f_{cl}j} - c_{rm}e^{\iota 2\pi f_{rm}j}, j \in \{0, 1, \cdots, 2N-2\}, \tag{4.29}$$

where $\boldsymbol{x}_j$ is the same as in (4.21), and $\tilde{\boldsymbol{x}}$ is defined as

$$\tilde{\boldsymbol{x}}_j = c_{rm} e^{\imath 2\pi f_{rm} j} + \sum_{k=1}^{R-1} c_k e^{\imath 2\pi f_k j}, \tag{4.30}$$

with f_{rm} being a frequency such that its corresponding frequency atom is mutually orthogonal with the atoms corresponding to f_1,..., and f_{R-1}. We further define

$$d_{min} = \min\{|c_1|, \cdots, |c_{R-1}|, |c_{rm}|\}. \tag{4.31}$$

Let us define $\tilde{\boldsymbol{X}} = \boldsymbol{H}(\tilde{\boldsymbol{x}})$, $\boldsymbol{X} = \boldsymbol{H}(\boldsymbol{x})$, and the error matrix $\boldsymbol{E}$ such that

$$\boldsymbol{X} = \tilde{\boldsymbol{X}} + \boldsymbol{E}, \tag{4.32}$$

where

$$\boldsymbol{E} = \begin{bmatrix} c_{cl} - c_{rm} & \cdots & c_{cl} e^{\imath 2\pi f_{cl} \cdot (N-1)} - c_{rm} e^{\imath 2\pi f_{rm} \cdot (N-1)} \\ \vdots & \ddots & \vdots \\ c_{cl} e^{\imath 2\pi f_{cl} \cdot (N-1)} - c_{rm} e^{\imath 2\pi f_{rm} \cdot (N-1)} & \cdots & c_{cl} e^{\imath 2\pi f_{cl} \cdot (2N-2)} - c_{rm} e^{\imath 2\pi f_{rm} \cdot (2N-2)} \end{bmatrix}. \tag{4.33}$$

Both $\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$ are rank-R matrices. For the error matrix $\boldsymbol{E}$, we have

$$\begin{aligned} \|\boldsymbol{E}\|_F &= \left(\sum_{i,k \in \{0,\cdots,N-1\}} |c_{cl} e^{\imath 2\pi f_{cl} \cdot (i+k)} - c_{rm} e^{\imath 2\pi f_{rm} \cdot (i+k)}|^2 \right)^{1/2} \\ &\leq \left(\sum_{i,k \in \{0,\cdots,N-1\}} \left(|c_{cl}| + |c_{rm}| \right)^2 \right)^{1/2} \\ &= \left(N^2 \left(|c_{cl}| + |c_{rm}| \right)^2 \right)^{1/2} \\ &= N \left(|c_{cl}| + |c_{rm}| \right). \end{aligned} \tag{4.34}$$

Following the derivations in Theorem 4.1, $\tilde{\boldsymbol{X}}$ has the following SVD

$$\tilde{\boldsymbol{X}} = \tilde{\boldsymbol{U}} \tilde{\boldsymbol{\Sigma}} \tilde{\boldsymbol{V}}^*, \tilde{\boldsymbol{U}} = [\tilde{\boldsymbol{U}}_1 \ \tilde{\boldsymbol{U}}_2] \in \mathbb{C}^{N \times N}, \tilde{\boldsymbol{V}} = [\tilde{\boldsymbol{V}}_1 \ \tilde{\boldsymbol{V}}_2] \in \mathbb{C}^{N \times N}, \tilde{\boldsymbol{\Sigma}} = \begin{bmatrix} \tilde{\boldsymbol{\Sigma}}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \in \mathbb{C}^{N \times N}, \tag{4.35}$$

where $\tilde{\boldsymbol{U}}_1 \in \mathbb{C}^{N \times R}$, $\tilde{\boldsymbol{\Sigma}}_1 \in \mathbb{C}^{R \times R}$ and $\tilde{\boldsymbol{V}}_1 \in \mathbb{C}^{N \times R}$ are defined as in (4.9), (4.10) and (4.11). The polar decomposition of $\tilde{\boldsymbol{X}}$ using its SVD is given by

$$\tilde{\boldsymbol{X}} = \tilde{\boldsymbol{P}} \tilde{\boldsymbol{H}}, \tilde{\boldsymbol{P}} = \tilde{\boldsymbol{U}}_1 \tilde{\boldsymbol{V}}_1^*, \tilde{\boldsymbol{H}} = \tilde{\boldsymbol{V}}_1 \tilde{\boldsymbol{\Sigma}}_1 \tilde{\boldsymbol{V}}_1^*, \tag{4.36}$$

and the matrix $\tilde{\boldsymbol{P}}$ is called the unitary polar factor. Similarly, for $\boldsymbol{X}$, we have

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}, \boldsymbol{U} = [\boldsymbol{U}_1\ \boldsymbol{U}_2] \in \mathbb{C}^{N\times N}, \boldsymbol{V} = [\boldsymbol{V}_1\ \boldsymbol{V}_2] \in \mathbb{C}^{N\times N}, \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}, \tag{4.37}$$

and its polar decomposition for $\boldsymbol{X}$,

$$\boldsymbol{X} = \boldsymbol{P}\boldsymbol{H}, \boldsymbol{P} = \boldsymbol{U}_1\boldsymbol{V}_1^*, \boldsymbol{H} = \boldsymbol{V}_1\boldsymbol{\Sigma}_1\boldsymbol{V}_1^*. \tag{4.38}$$

Suppose that we try to recover $\boldsymbol{x}$ through the following nuclear norm minimization:

$$\begin{aligned} &\min_{\boldsymbol{x}} \|\boldsymbol{H}(\boldsymbol{x})\|_* \\ &\text{s.t. } \mathscr{A}(\boldsymbol{x}) = \boldsymbol{b}. \end{aligned} \tag{4.39}$$

As in the proof of Theorem 4.1, we analyse the null space condition for successful recovery using Hankel matrix recovery. Towards this, we first bound the unitary polar factor of $\boldsymbol{X}$ through the perturbation theory for polar decomposition.

For our problem, we have $m = n = N, r = R$. Let σ_R be the R-th singular value of $\boldsymbol{X}$. From (4.11) and (4.31), an explicit form for σ_R is

$$\sigma_R = Nd_{min}. \tag{4.40}$$

Then Lemma 4.3 implies that

$$\|\boldsymbol{P} - \tilde{\boldsymbol{P}}\|_F \leq \frac{4}{\sigma_R}\|\boldsymbol{X} - \tilde{\boldsymbol{X}}\|_F. \tag{4.41}$$

From the null space condition for nuclear norm minimization (4.39), we can correctly recover $\boldsymbol{x}$ if

$$|\mathrm{Tr}(\boldsymbol{V}_1\boldsymbol{U}_1^*\boldsymbol{Q})| < \|\bar{\boldsymbol{U}}_1^*\boldsymbol{Q}\bar{\boldsymbol{V}}_1\|_*, \tag{4.42}$$

for every non-zero $\boldsymbol{Q} = \boldsymbol{H}(\boldsymbol{z})$ with $\boldsymbol{z} \in \mathscr{N}(\mathscr{A})$, where $\bar{\boldsymbol{U}}_1$ and $\bar{\boldsymbol{V}}_1$ are such that $[\boldsymbol{U}_1\ \bar{\boldsymbol{U}}_1]$ and $[\boldsymbol{V}_1\ \bar{\boldsymbol{V}}_1]$ are unitary, i.e. $\bar{\boldsymbol{U}}_1 = \boldsymbol{U}_2$ and $\bar{\boldsymbol{V}}_1 = \boldsymbol{V}_2$. Since the observation set $\mathscr{M} = \{0, 1, 2, ..., 2N-2\} \setminus \{N-1\}$, $\boldsymbol{Q}$ takes the form in (4.12) with $a \in \mathbb{C}$, and

$$\|\bar{\boldsymbol{U}}_1^*\boldsymbol{Q}\bar{\boldsymbol{V}}_1\|_* = |a|(N-R). \tag{4.43}$$

Let us define $\Delta \boldsymbol{P} = \tilde{\boldsymbol{P}} - \boldsymbol{P} = \tilde{\boldsymbol{U}}_1\tilde{\boldsymbol{V}}_1^* - \boldsymbol{U}_1\boldsymbol{V}_1^*$, and we have

$$\begin{aligned}
|\mathrm{Tr}(\boldsymbol{V}_1\boldsymbol{U}_1^*\boldsymbol{Q})| &= |<\boldsymbol{Q}, \boldsymbol{U}_1\boldsymbol{V}_1^*>| \\
&\le |<\boldsymbol{Q}, \tilde{\boldsymbol{P}}>| + |<\boldsymbol{Q}, \Delta\boldsymbol{P}>| \\
&= |\mathrm{Tr}(\tilde{\boldsymbol{V}}_1\tilde{\boldsymbol{U}}_1^*\boldsymbol{Q})| + |<\boldsymbol{Q}, \Delta\boldsymbol{P}>| \\
&\le |\mathrm{Tr}(\tilde{\boldsymbol{V}}_1\tilde{\boldsymbol{U}}_1^*\boldsymbol{Q})| + (\|\boldsymbol{Q}\|_F^2\|\|\Delta\boldsymbol{P}\|_F^2)^{1/2} \\
&\le |\mathrm{Tr}(\tilde{\boldsymbol{V}}_1\tilde{\boldsymbol{U}}_1^*\boldsymbol{Q})| + |a|\sqrt{N}\|\Delta\boldsymbol{P}\|_F, \qquad (4.44)
\end{aligned}$$

where we used the Cauchy–Schwartz inequality. Notice that $\boldsymbol{E} = \boldsymbol{X} - \tilde{\boldsymbol{X}}$, and then it follows from (4.41) and (4.34) that

$$\begin{aligned}
|\mathrm{Tr}(\boldsymbol{V}_1\boldsymbol{U}_1^*\boldsymbol{Q})| &\le |\mathrm{Tr}(\tilde{\boldsymbol{V}}_1\tilde{\boldsymbol{U}}_1^*\boldsymbol{Q})| + |a|\frac{4\sqrt{N}}{\tilde{\sigma}_R}\|\boldsymbol{E}\|_F \\
&\le |\mathrm{Tr}(\tilde{\boldsymbol{V}}_1\tilde{\boldsymbol{U}}_1^*\boldsymbol{Q})| + |a|\frac{4\sqrt{N}}{Nd_{min}} \cdot N\left(|c_{cl}| + |c_{rm}|\right) \\
&\le |\mathrm{Tr}(\tilde{\boldsymbol{V}}_1\tilde{\boldsymbol{U}}_1^*\boldsymbol{Q})| + 4|a|\sqrt{N} \cdot \frac{|c_{cl}| + |c_{rm}|}{d_{min}}. \qquad (4.45)
\end{aligned}$$

Thus if

$$|\mathrm{Tr}(\tilde{\boldsymbol{V}}_1\tilde{\boldsymbol{U}}_1^*\boldsymbol{Q})| + 4|a|\sqrt{N} \cdot \frac{|c_{cl}| + |c_{rm}|}{d_{min}} < |a|(N-R), \qquad (4.46)$$

namely,

$$|\mathrm{Tr}(\tilde{\boldsymbol{V}}_1\tilde{\boldsymbol{U}}_1^*\boldsymbol{Q})| < |a|\left(N - R - 4\sqrt{N} \cdot d_{rel}\right), \qquad (4.47)$$

where d_{rel} is defined as $\frac{|c_{cl}|+|c_{rm}|}{d_{min}}$, then solving (4.39) will correctly recover $\boldsymbol{x}$.

The proof of Theorem 4.1 leads to the concentration inequality:

$$\mathbb{P}\left(|\mathrm{Tr}(\tilde{\boldsymbol{V}}\tilde{\boldsymbol{U}}^*\boldsymbol{Q})| \ge |a|t\right) \le 3e^{-\frac{t^2/2}{R+t/3}}, \forall\ t > 0.$$

Taking $t = N - R - 4\sqrt{N}d_{rel}$, we have

$$\begin{aligned}
\frac{t^2/2}{R + t/3} &= \frac{(N - R - 4\sqrt{N}d_{rel})^2/2}{R + (N - R - 4\sqrt{N}d_{rel})/3} \\
&= \frac{3}{2} \cdot \frac{N^2 + R^2 + 16Nd_{rel}^2 - 2NR - 8N\sqrt{N}d_{rel} + 8R\sqrt{N}d_{rel}}{3R + N - R - 4\sqrt{N}d_{rel}}. \qquad (4.48)
\end{aligned}$$

To have successful signal recovery with high probability, we let

$$\frac{3}{2}\cdot\frac{N^2+R^2+16Nd_{rel}^2-2NR-8N\sqrt{N}d_{rel}+8R\sqrt{N}d_{rel}}{3R+N-R-4\sqrt{N}d_{rel}}=c\log(N), \tag{4.49}$$

where $c>0$ is a constant. So we have

$$R^2+sR+r=0, \tag{4.50}$$

where

$$s=8\sqrt{N}d_{rel}-2N-2c_1\log(N), \tag{4.51}$$

$$r=N^2+16Nd_{rel}^2-8\sqrt{N}Nd_{rel}-c_1N\log(N)+4c_1\sqrt{N}d_{rel}\log(N), \tag{4.52}$$

and $c_1=\frac{2}{3}c$.

Since

$$\begin{aligned}
s^2-4r&=(8\sqrt{N}d_{rel}-2N-2c_1\log(N))^2\\
&\quad-4\left(N^2+16Nd_{rel}^2-8\sqrt{N}Nd_{rel}-c_1N\log(N)+4c_1\sqrt{N}d_{rel}\log(N)\right)\\
&=\left(64Nd_{rel}^2+4N^2+4c_1^2\log^2(N)-32\sqrt{N}Nd_{rel}-32c_1\sqrt{N}d_{rel}\log(N)\right)\\
&\quad-\left(4N^2+64Nd_{rel}^2-32\sqrt{N}Nd_{rel}-4c_1N\log(N)+16c_1\sqrt{N}d_{rel}\log(N)\right)\\
&=12c_1N\log(N)-48c_1\sqrt{N}d_{rel}\log(N)+4c_1^2\log^2(N),
\end{aligned}$$

solving the quadratic equation leads to

$$\begin{aligned}
R&=\frac{1}{2}(-s\pm\sqrt{s^2-4r})\\
&=\frac{1}{2}\left(-8Nd_{rel}+2N+2c_1\log(N)\pm\sqrt{12c_1N\log(N)-48c_1\sqrt{N}d_{rel}\log(N)+4c_1^2\log^2(N)}\right)\\
&=N-4\sqrt{N}d_{rel}+c_1\log(N)\pm\sqrt{3c_1N\log(N)-12c_1Nd_{rel}\log(N)+c_1^2\log^2(N)}.
\end{aligned}$$

Plugging in $c_1=\frac{2}{3}c$, we have

$$R=N-4\sqrt{N}d_{rel}+\frac{2}{3}c\log(N)+\sqrt{2cN\log(N)-8c\sqrt{N}d_{rel}\log(N)+\frac{4}{9}c^2\log^2(N)}, \tag{4.53}$$

and

$$R = N - 4\sqrt{N}d_{rel} + \frac{2}{3}c\log(N) - \sqrt{2cN\log(N) - 8c\sqrt{N}d_{rel}\log(N) + \frac{4}{9}c^2\log^2(N)}. \tag{4.54}$$

Notice that we can only take (4.54) since (4.53) can give a R which is greater than N, which is not acceptable.

Since we can freely choose the coefficient c_{rm}, we choose c_{rm} such that

$$|c_{rm}| = c_{min} = \min\{|c_1|, |c_2|, ..., |c_{R-1}|\}.$$

Under such a choice for $|c_{rm}|$, $d_{min} = |c_{rm}| = c_{min}$, leading to $d_{rel} = \frac{|c_{cl}|+|c_{rm}|}{d_{min}} = \frac{|c_{cl}|+c_{min}}{c_{min}}$. This concludes the proof of Theorem 4.3. □

5. Separation is always necessary for the success of atomic norm minimization

In the previous sections, we have shown that the Hankel matrix recovery can recover the superposition of complex exponentials, even though the frequencies of the complex exponentials can be arbitrarily close. In this section, we show that, broadly, the atomic norm minimization must obey a non-trivial resolution limit. This is very different from the behaviour of Hankel matrix recovery. Our results in this section also greatly generalize necessity of resolution limit results in [40], to general continuously parametered dictionary, beyond the dictionary of frequency atoms. Moreover, this new lower bound can work for non-uniform sampling, where some samples at certain time indices can be missing, while the lower bound from [40] applies to uniform sampling. Our new lower bound further applies to samples obtained at non-integer time indices, while the lower bound from [40] does not apply to sampling at non-integer time indices. In terms of technical tools used, our analysis of using matrix analysis is very different from the derivations in [40], which used the Markov–Bernstein type inequalities for finite-degree polynomials.

THEOREM 5.1. Let us consider a dictionary with its atoms parameterized by a continuous-valued parameter $\tau \in \mathbb{C}$. We also assume that each atom $\boldsymbol{a}(\tau)$ belongs to $\mathbb{C}^N$, where N is a positive integer. We assume that the set of all the atoms spans a Q-dimensional subspace in $\mathbb{C}^N$.

Suppose the signal is the superposition of several atoms:

$$\boldsymbol{x} = \sum_{k=1}^{R} c_k \boldsymbol{a}(\tau_k), \tag{5.1}$$

with non-zero $c_k \in \mathbb{C}$ for each k, and R is a positive integer representing the number of active atoms.

Consider any active atoms $\boldsymbol{a}(\tau_{k_1})$ and $\boldsymbol{a}(\tau_{k_2})$. With the other $(R-2)$ active atoms and their coefficients fixed (this includes the case $R = 2$, namely there are only two active atoms in total), if the atomic norm minimization can always identify the two active atoms, and correctly recover the coefficients for $\boldsymbol{a}(\tau_{k_1})$ and $\boldsymbol{a}(\tau_{k_2})$, then the two atoms $\boldsymbol{a}(\tau_{k_1})$ and $\boldsymbol{a}(\tau_{k_2})$ must be well separated such that

$$\|\boldsymbol{a}(\tau_{k_1}) - \boldsymbol{a}(\tau_{k_2})\|_2 \geq 2 \max_{S \geq Q} \max_{\boldsymbol{A} \in \boldsymbol{M}^S} \frac{\sigma_{min}(\boldsymbol{A})}{\sqrt{S}}, \tag{5.2}$$

where S is a positive integer, $\boldsymbol{M}^S$ is the set of matrices with S columns and with each of these S columns corresponding to an atom, and $\sigma_{min}(\cdot)$ is the smallest singular value of a matrix.

Proof of Theorem 5.1: Define the sign of a coefficient c_k as

$$\text{sign}(c_k) = \frac{c_k}{|c_k|}. \tag{5.3}$$

Then according to [41], a necessary condition for the atomic norm to identify the two active atoms, and correctly recover their coefficients is that there exists a dual vector $\boldsymbol{q} \in \mathbb{C}^N$ such that

$$\begin{cases} \boldsymbol{a}(\tau_{k_1})^*\boldsymbol{q} = \text{sign}(c_{k_1}),\ \boldsymbol{a}(\tau_{k_2})^*\boldsymbol{q} = \text{sign}(c_{k_2}) \\ |\boldsymbol{a}(\tau)^*\boldsymbol{q}| \leq 1, \forall \tau \notin \{\tau_{k_1}, \tau_{k_2}\}. \end{cases} \tag{5.4}$$

We pick c_{k_1} and c_{k_2} such that $|\text{sign}(c_{k_1}) - \text{sign}(c_{k_2})| = 2$. Then

$$\|\text{sign}(c_{k_1}) - \text{sign}(c_{k_2})\| = \|\boldsymbol{a}(\tau_{k_1})^*\boldsymbol{q} - \boldsymbol{a}(\tau_{k_2})^*\boldsymbol{q}\|_2 \leq \|\boldsymbol{q}\|_2 \|\boldsymbol{a}(\tau_{k_1}) - \boldsymbol{a}(\tau_{k_2})\|_2, \tag{5.5}$$

Thus

$$\|\boldsymbol{q}\|_2 \geq \frac{2}{\|\boldsymbol{a}(\tau_{k_1}) - \boldsymbol{a}(\tau_{k_2})\|_2}. \tag{5.6}$$

Now we take a group of S atoms, denoted by $\boldsymbol{a}_{select,1}$,..., and $\boldsymbol{a}_{select,S}$, and use them to form the S columns of a matrix $\boldsymbol{A}$. Then

$$\begin{aligned} \sigma_{min}(\boldsymbol{A})\|\boldsymbol{q}\|_2 &\leq \|\boldsymbol{A}^*\boldsymbol{q}\|_2 \\ &= \sqrt{\sum_{j=1}^{S} |\boldsymbol{a}_{select,j}^*\boldsymbol{q}|^2} \\ &\leq \sqrt{S}, \end{aligned} \tag{5.7}$$

where the last inequality comes form (5.4). It follows that

$$\|\boldsymbol{q}\|_2 \leq \frac{\sqrt{S}}{\sigma_{min}(\boldsymbol{A})}. \tag{5.8}$$

Define β as the maximal value of $\frac{\sigma_{min}(\boldsymbol{A})}{\sqrt{S}}$ among all the choices for $\boldsymbol{A}$ and S, namely,

$$\beta = \max_{S \geq Q} \max_{\boldsymbol{A} \in \boldsymbol{M}^S} \frac{\sigma_{min}(\boldsymbol{A})}{\sqrt{S}}. \tag{5.9}$$

TABLE 1 *Comparison between [40] and ours in lower bounds on frequency separations*

N	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
[40]	0.1592	0.0796	0.0531	0.0398	0.0318	0.0265	0.0227	0.0199	0.0177
Our bound	0.1667	0.0628	0.0351	0.0232	0.0167	0.0128	0.0102	0.0084	0.0071

Combining (5.6) and (5.8), we have

$$\|\boldsymbol{a}(\tau_{k_1}) - \boldsymbol{a}(\tau_{k_2})\|_2 \geq 2\beta, \tag{5.10}$$

proving this theorem. □

We would like to remark that, besides applications to frequency atoms, our bounds from Theorem 5.1 on atom separations are quite general, and can be applied to any dictionary of atoms with continuous parameters. The derivations in [40] apply only to dictionary of frequency atoms under complete (without missing samples) uniform sampling at integer time indices. When applied to frequency atoms, our bound from Theorem 5.1 works for non-uniform sampling, where some samples are missing at certain time indices, and works for sampling at non-integer time indices (for example, sampling at time indices $j = \sqrt{1}, \sqrt{2}, \sqrt{3}, ..., \sqrt{2N-2}$; sampling at non-integer times indices may happen because of sampling time jittering), where the bound in [40] does not apply. When further restricted to uniformly sampled-at-integer-time-index frequency atoms, our calculations show that Theorem 5.1 provides similar or even tighter lower bounds on frequency separations than in [40]. In addition, our derivations are novel, and do not resort to the Markov–Bernstein-type inequalities for polynomials as used in [40]. It is possible to further tighten our method of deriving lower bound using matrices, and we will leave it for future work.

In the following calculations, we consider frequency atoms given in the following form (atoms will change for non-uniform sampling or sampling at non-integer time indices),

$$\boldsymbol{a}(f) = [1 \; e^{i2\pi f} \; \cdots \; e^{i2\pi f(2N-2)}]^T \in \mathbb{C}^{2N-1}. \tag{5.11}$$

For two frequencies f_a and f_b, we call $\|\boldsymbol{a}(f_a) - \boldsymbol{a}(f_b)\|$ the atom separation (Theorem 5.1 can directly produce a lower bound on atom separation, while [40] works with frequency separation), and call $|f_a - f_b|_{wrap}$ (wrap-around distance) the frequency separation. We note that, in [40], the author established a lower bound of $\frac{2}{(2N-2)*2\pi}$ for frequency separation (see Theorem 3 in [40]).

5.1 *Uniform sampling (no sample is missing) at integer indices*

In this case, the samples are observed at time indices $j = 0, 1, 2, ..., (2N-2)$. For this sampling pattern, under small N, our bound can be tighter than the lower bound given in [40], and, under large N, the lower bound in [40] is tighter.

Table 1 shows comparison result between [40] and ours on lower bound on frequency separation for $N = 2, ..., 10$.

Lower bound from [40] when $2N - 2 = 2$**:** In this case (three time-domain samples in total), the lower bound on frequency separation in [40] is 0.159. We construct two atoms with frequencies $f_a = 0$ and $f_b = 0.159$, and then calculate the atom separation. This gives an atom separation of 1.94.

New lower bound when $2N-2=2$: To calculate our own lower bound, we consider a set of orthogonal atoms (any set of linearly independent atoms will give us a valid lower bound). The frequencies of these atoms are taken to be $f_1=0, f_2=\frac{1}{3}$ and $f_3=\frac{2}{3}$. Our Theorem 5.1 gives a lower bound of 2.00 for $\|\boldsymbol{a}(f_a)-\boldsymbol{a}(f_b)\|$. To obtain this lower bound based on Theorem 5.1, we take matrix $A=[\boldsymbol{a}(f_1)\ \boldsymbol{a}(f_2)\ \boldsymbol{a}(f_3)]$, and $S=3$ since the three atoms span a three-dimensional linear space. Using these parameters Theorem 5.1 gives $\|a(f_a)-a(f_b)\|\geq 2.00$. This corresponds to a frequency separation $|f_a-f_b|_{wrap}=0.167$. In this case, the bound given by Theorem 5.1 is indeed tighter than the frequency separation bound (0.159) from [40].

Lower bound from [40] when $2N-2=62$: When $2N-2=62$ (totally 63 time-domain samples respectively), the frequency separation requirement in [40] will be 0.0051. We construct two atoms with frequencies $f_a=0$ and $f_b=0.0051$, and then calculate the atom distance. This gives an atom distance of 8.31.

New lower bound when $2N-2=62$: To calculate our own lower bound, we consider a set of orthogonal atoms (any set of linearly independent atoms will give us a valid lower bound). The frequencies of these atoms are taken to be $f_1=0, f_2=\frac{1}{63}, f_3=\frac{2}{63}, \cdots, f_{63}=\frac{62}{63}$. Our Theorem 5.1 gives a bound of 2.00 for $\|\boldsymbol{a}(f_a)-\boldsymbol{a}(f_b)\|$. To obtain this lower bound based on Theorem 5.1, we take matrix $A=[\boldsymbol{a}(f_1)\ \boldsymbol{a}(f_2)\ \cdots\ \boldsymbol{a}(f_{63})]$, and $S=63$ since the 63 atoms span a 63-dimensional linear space. Theorem 5.1 gives $\|a(f_a)-a(f_b)\|\geq 2.00$. This corresponds to a frequency separation $|f_a-f_b|_{wrap}=1.1\times 10^{-3}$, which is less tight than, but comparable with the frequency separation bound from [40].

Lower bound from [40] when $2N-2=126$: When $2N-2=126$ (totally $2N-1=127$ time-domain samples, respectively), the frequency separation required [40] is 0.0025. We simply construct two atoms with frequency $f_a=0$ and $f_b=0.0025$, and then calculate the distance between these two atoms. This gives atom distance of 11.78.

New lower bound when $2N-2=126$: To calculate our own lower bound, we consider a set of orthogonal atoms (any set of linearly independent atoms will give us a valid lower bound). The frequencies of these atoms are taken to be $f_1=0, f_2=\frac{1}{127}, f_3=\frac{2}{127}, \cdots, f_{127}=\frac{126}{127}$. Our Theorem 5.1 gives a bound of 2.00 for $\|\boldsymbol{a}(f_a)-\boldsymbol{a}(f_b)\|$. We take matrix $A=[\boldsymbol{a}(f_1)\ \boldsymbol{a}(f_2)\ \cdots\ \boldsymbol{a}(f_{127})]$, take $S=127$ since the 127 atoms form a 127-dimensional space, and Theorem 5.1 gives $\|a(f_a)-a(f_b)\|\geq 2.00$. This corresponds to a frequency separation $|f_a-f_b|_{wrap}=4\times 10^{-4}$, which is less tight than, but still comparable with the frequency separation bound from [40].

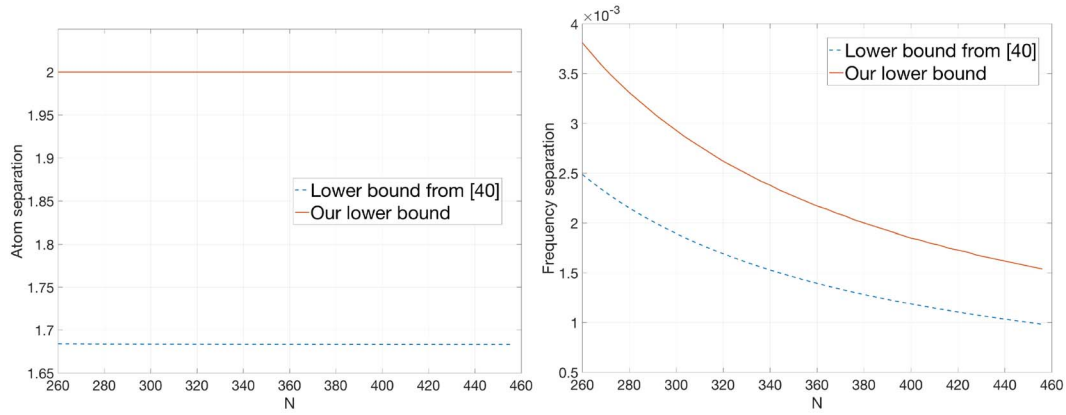
5.2 *Non-uniform sampling (with missing samples) at integer indices*

Theorem 5.1 can still provide valid lower bounds on frequency or atom separations for non-uniform sampling. The lower bound from [40] does not apply to non-uniform sampling directly, but a simple modification can yield a lower bound $\frac{2}{N_{max}\times 2\pi}$ on frequency separation, where N_{max} is the largest observed time index.

As one example, we consider $N=256$ corresponding to $2N-1=511$ total time indices, and we assume that samples are taken at only five time indices: $j=0,1,2,3$ and 124. In this case, the atom at a certain frequency $f\in[0,1)$ is given by

$$\boldsymbol{a}(f)=[1\ e^{i2\pi f}\ \ e^{i4\pi f}\ e^{i6\pi f}\ e^{i2\pi f(124)}]^T\in\mathbb{C}^5. \tag{5.12}$$

Lower bound (modified or unmodified for non-uniform sampling) from [40]: The modified frequency separation $\frac{2}{N_{max}\times 2\pi}$ required by [40] is $2.6*10^{-3}$, where $N_{max}=124$. We simply construct



(a) Lower bounds on atom separation (b) Lower bound on frequency separation

FIG. 1. Comparisons of lower bounds on frequency separations and atom separations

two atoms with frequency $f_a = 0$ and $f_b = 0.0026$, and then calculate the distance between these two atoms. This gives an atom distance of 1.68. A direct application of lower bound $\frac{2}{(2N-2)\times 2\pi}$ will give a frequency separation of 6×10^{-4}, and an atom separation of 0.4817.

Our lower bound from Theorem 5.1: to calculate our own lower bound, we consider a set of atoms (any set of linearly independent atoms will give us a valid lower bound) at frequencies $f_1 = 0, f_2 = \frac{1}{5}, f_2 = \frac{2}{5}, f_3 = \frac{3}{5}, f_5 = \frac{4}{5}$. Our Theorem 5.1 gives a lower bound of 2.00 for $\|\boldsymbol{a}(f_a) - \boldsymbol{a}(f_b)\|_2$ by taking matrix $A = [\boldsymbol{a}(f_1)\ \boldsymbol{a}(f_2)\ \boldsymbol{a}(f_3)\ \boldsymbol{a}(f_4)\ \boldsymbol{a}(f_5)]$, and $S = 5$. This corresponds to a frequency separation $|f_a - f_b|_{wrap} = 3.9 \times 10^{-3}$.

As we can see, for this non-uniform sampling example, our new lower bound is tighter than the lower bounds (both unmodified or modified for non-uniform sampling) from [40] for $N = 256$.

We further consider the same set-up for $N = 256 + 4 \times k$, where $1 \le k \le 50$ is an integer, and the corresponding observation index is $j = 0, 1, 2, 3$ and $124 + 4 \times k$. In this case, the atom at a certain frequency $f \in [0, 1)$ is given by

$$\boldsymbol{a}(f) = [1\ e^{i2\pi f}\ \ e^{i4\pi f}\ e^{i6\pi f}\ e^{i2\pi f(124+4\times k)}]^T \in \mathbb{C}^5. \tag{5.13}$$

Then Fig. 1 shows the lower bound on the frequency separation obtained from [40] and Theorem 5.1, and also shows the lower bound on the atom separation obtained from [40] and Theorem 5.1.

5.3 *Sampling at non-integer indices*

For general sampling at non-integer indices, the lower bound from [40] does not apply, since the proof of the lower bound therein depends on the Markov–Bernstein-type inequalities for (integer-exponent) polynomials. In that case, a trivial lower bound 0 exists for frequency separation.

On the other hand, our lower bound from Theorem 5.1 can still be easily calculated for sampling at non-integer indices. Let us consider $N = 4$, and we observe $2N-1 = 7$ samples. The corresponding time index j for these samples are 0, 1, $2^{1.5}$,..., and $6^{1.5}$. To use Theorem 5.1, instead of optimizing over seven frequencies, we simply randomly selected seven frequencies between 0 and 1 to be frequency atoms for

matrix $\boldsymbol{A}$, and obtained a lower bound 0.1947 for atom separation (notice any matrix $\boldsymbol{A}$ consisting of atoms can give a lower bound in Theorem 5.1). This corresponds to a non-trivial frequency separation of 0.0015. To the best of knowledge, this is the first approach in the literature being able to give a lower bound on frequency separation for such sampling patterns.

6. OANM: Hankel matrix recovery can be immune from atom separation requirements

Our results in the previous sections naturally raise the following question: why can Hankel matrix recovery work without requiring separations between the underlying atoms while it is necessary for the atomic norm minimization to require separations between the underlying atoms? In this section, we introduce the concept of orthonormal atomic norm and its minimization, which explains the success of Hankel matrix recovery in recovering the superposition of complex exponentials regardless of the separations between frequency atoms.

Let us consider a vector $\boldsymbol{w} \in \mathbb{C}^N$, where N is a positive integer. We denote the set of atoms by $\mathscr{AMSET}$, and assume that each atom $\boldsymbol{a}(\tau)$ (parameterized by τ) belongs to $\mathbb{C}^N$. Then the atomic norm is given by [5, 10, 41]

$$\|\boldsymbol{w}\|_{Atomic} = \inf_{s,\tau_k,c_k} \left\{ \sum_{k=1}^{s} |c_k| : \boldsymbol{w} = \sum_{k=1}^{s} c_k \boldsymbol{a}(\tau_k) \right\}. \tag{6.1}$$

We say the atomic norm $\|\boldsymbol{w}\|_{\mathscr{ATOMIC}}$ is an orthonormal atomic norm if, for every $\boldsymbol{w}$,

$$\|\boldsymbol{w}\|_{Atomic} = \sum_{k=1}^{s} |c_k|, \tag{6.2}$$

where $\boldsymbol{w} = \sum_{k=1}^{s} c_k \boldsymbol{a}(\tau_k)$, $\|\boldsymbol{a}(\tau_k)\|_2 = 1$ for every k, and $\boldsymbol{a}(\tau_k)$'s are mutually orthogonal to each other.

In the Hankel matrix recovery, the atom set $\mathscr{AMSET}$ is composed of all the rank-1 matrices in the form $\boldsymbol{uv}^*$, where $\boldsymbol{u}$ and $\boldsymbol{v}$ are unit-norm vectors in $\mathbb{C}^N$ [10, 12]. Let us assume $\boldsymbol{x} \in \mathbb{C}^{2N-1}$. We can see that the nuclear norm of a Hankel matrix is an orthonormal atomic norm of $\boldsymbol{H}(\boldsymbol{x})$:

$$\|\boldsymbol{H}(\boldsymbol{x})\|_* = \sum_{k=1}^{R} |c_k|, \tag{6.3}$$

where $\boldsymbol{H}(\boldsymbol{x}) = \sum_{k=1}^{R} c_k \boldsymbol{u}_k \boldsymbol{v}_k^*$, $\boldsymbol{H}(\boldsymbol{x}) = \boldsymbol{U\Sigma V}^*$ is the SVD of $\boldsymbol{H}(\boldsymbol{x})$, $\boldsymbol{u}_k$ is the k-th column of $\boldsymbol{U}$, and $\boldsymbol{v}_k$ is the k-th column of $\boldsymbol{V}$. This is because the matrices $\boldsymbol{u}_k \boldsymbol{v}_k^*$'s are orthogonal to each other and each of these rank-1 matrices has unit energy.

Let us now further assume that $\boldsymbol{x} \in \mathbb{C}^{2N-1}$ is the superposition of R complex exponentials with $R \leq N$, as defined in (1.1). Then $\boldsymbol{H}(\boldsymbol{x})$ is a rank-R matrix, and can be written as $\boldsymbol{H}(\boldsymbol{x}) = \sum_{k=1}^{R} c_k \boldsymbol{u}_k \boldsymbol{v}_k^*$, where $\boldsymbol{H}(\boldsymbol{x}) = \boldsymbol{U\Sigma V}^*$ is the SVD of $\boldsymbol{H}(\boldsymbol{x})$, $\boldsymbol{u}_k$ is the k-th column of $\boldsymbol{U}$ and $\boldsymbol{v}_k$ is the k-th column of $\boldsymbol{V}$. Even though the original R frequency atoms $\boldsymbol{a}(\tau_k)$'s for $\boldsymbol{x}$ can be arbitrarily close, we can always write $\boldsymbol{H}(\boldsymbol{x})$ as a superposition of R orthonormal atoms $\boldsymbol{u}_k \boldsymbol{v}_k^*$'s from the SVD of $\boldsymbol{H}(\boldsymbol{x})$. Because the original R frequency atoms $\boldsymbol{a}(\tau_k)$'s for $\boldsymbol{x}$ can be arbitrarily close, they can violate the necessary separation condition set forth in (5.2). However, for $\boldsymbol{H}(\boldsymbol{x})$, its composing atoms can be R orthonormal atoms $\boldsymbol{u}_k \boldsymbol{v}_k^*$'s from the SVD of $\boldsymbol{H}(\boldsymbol{x})$. These atoms $\boldsymbol{u}_k \boldsymbol{v}_k^*$'s are of unit energy, and are orthogonal to each other. Thus, these atoms are well

separated and have the opportunity of not violating the necessary separation condition set forth in (5.2). This explains why the Hankel matrix recovery approach can break free from the separation condition which is required for traditional atomic norm minimizations.

7. A matrix-theoretic inequality of nuclear norms and its proof from the theory of compressed sensing

In this section, we present a new matrix-theoretic inequality of nuclear norms, and give a proof of it from the theory of compressed sensing (using nuclear norm minimization). To the best of our knowledge, we have not seen the statement of this inequality of nuclear norms, or its proof elsewhere in the literature.

THEOREM 7.1. Let $\boldsymbol{A} \in \mathbb{C}^{m\times n}$, and $t = \min(m,n)$. Let σ_1, σ_2,..., and σ_t be the singular values of $\boldsymbol{A}$ arranged in descending order, namely

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_t.$$

Let k be any integer such that

$$\sigma_1 + \cdots + \sigma_k < \sigma_{k+1} + \cdots + \sigma_t.$$

Then for any orthogonal projector $\boldsymbol{P}$ onto k-dimensional subspaces in $\mathbb{C}^m$, and any orthogonal projector $\boldsymbol{Q}$ onto k-dimensional subspaces in $\mathbb{C}^n$, we have

$$\|\boldsymbol{PAQ}^*\|_* \leq \|(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{A}(\boldsymbol{I}-\boldsymbol{Q})^*\|_*. \tag{7.1}$$

In particular,

$$\|\boldsymbol{A}_{1:k,1:k}\|_* \leq \|\boldsymbol{A}_{(k+1):m,(k+1):n}\|_*, \tag{7.2}$$

where $\boldsymbol{A}_{1:k,1:k}$ is the submatrix of $\boldsymbol{A}$ with row indices between 1 and k and column indices between 1 and k, and $\boldsymbol{A}_{(k+1):m,(k+1):n}$ is the submatrix of $\boldsymbol{A}$ with row indices between $k+1$ and m and column indices between $k+1$ and n.

Before giving the proof of Theorem 7.1, we will state Lemmas 7.1, 7.2 and 7.3 which will be used in the proof of Theorem 7.1.

LEMMA 7.1. ([29]) Let $\boldsymbol{G}$ and $\boldsymbol{H}$ be two matrices of the same dimension in $\mathbb{C}^{m\times n}$. Then $\sum_{i=1}^{t} |\sigma_i(\boldsymbol{G}) - \sigma_i(\boldsymbol{H})| \leq \|\boldsymbol{G}-\boldsymbol{H}\|_*$, where $t = \min(m,n)$.

Lemma 7.1 actually provides a lower bound for the nuclear norm of the difference between two matrices, and the lower bound is just the sum of the differences of singular values.

LEMMA 7.2. ([1; 20]) For arbitrary matrices $\boldsymbol{X}, \boldsymbol{Y}$ and $\boldsymbol{Z} = \boldsymbol{X} - \boldsymbol{Y} \in \mathbb{C}^{m\times n}$. Let σ_1, σ_2,..., and σ_t ($t = \min\{m,n\}$) be the singular values of a matrix $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ arranged in descending order, namely $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_t$. Let $s_i(\boldsymbol{X},\boldsymbol{Y})$ be the distance between the i-th singular value of $\boldsymbol{X}$ and $\boldsymbol{Y}$, namely,

$$s_i(\boldsymbol{X},\boldsymbol{Y}) = |\sigma_i(\boldsymbol{X}) - \sigma_i(\boldsymbol{Y})|, i = 1,2,\cdots,t. \tag{7.3}$$

Let $s_{[i]}(\boldsymbol{X}, \boldsymbol{Y})$ be the i-th largest value of sequence $s_1(\boldsymbol{X}, \boldsymbol{Y}), s_2(\boldsymbol{X}, \boldsymbol{Y}), \cdots, s_t(\boldsymbol{X}, \boldsymbol{Y})$, then

$$\sum_{i=1}^{k} s_{[i]}(\boldsymbol{X}, \boldsymbol{Y}) \leq \|\boldsymbol{Z}\|_k, \forall k = 1, 2, \cdots, t, \tag{7.4}$$

where $\|\boldsymbol{Z}\|_k$ is defined as $\sum_{i=1}^{k} \sigma_i(\boldsymbol{Z})$.

Different from Lemma 7.1, Lemma 7.2 considers the partial sum of singular values of the difference between two matrices, and similarly, this partial sum can be lower bounded by the partial sum of the difference of singular values.

LEMMA 7.3. Suppose a rank-R matrix $\boldsymbol{X} \in \mathbb{C}^{M \times N}$ admits an SVD $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*$, where $\boldsymbol{\Sigma} \in \mathbb{R}^{R \times R}$ is a diagonal matrix, and $\boldsymbol{U} \in \mathbb{C}^{M \times R}$ and $\boldsymbol{V} \in \mathbb{C}^{N \times R}$ satisfy $\boldsymbol{U}^*\boldsymbol{U} = \boldsymbol{V}^*\boldsymbol{V} = \boldsymbol{I}$.

Define $\mathfrak{S} \subseteq \mathbb{C}^{M \times N}$ as

$$\mathfrak{S} = \left\{ \boldsymbol{U}\boldsymbol{V}^* + \boldsymbol{W} \mid \boldsymbol{U}^*\boldsymbol{W} = \boldsymbol{0},\ \boldsymbol{W}\boldsymbol{V} = \boldsymbol{0},\ \|\boldsymbol{W}\|_2 \leq 1,\ \boldsymbol{W} \in \mathbb{C}^{M \times N} \right\}, \tag{7.5}$$

and define

$$\mathscr{F}\left(\begin{bmatrix} \mathrm{Re}(\boldsymbol{X}) \\ \mathrm{Im}(\boldsymbol{X}) \end{bmatrix} \right) = \|\boldsymbol{X}\|_*.$$

Then we have

$$\mathfrak{H} \equiv \left\{ \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} \;\middle|\; \boldsymbol{\alpha} + \imath\boldsymbol{\beta} \in \mathfrak{S} \right\} = \partial\mathscr{F}\left(\begin{bmatrix} \mathrm{Re}(\boldsymbol{X}) \\ \mathrm{Im}(\boldsymbol{X}) \end{bmatrix} \right). \tag{7.6}$$

Remarks: 1. Lemma 7.3 gives the subdifferential of the nuclear norm of a general complex-numbered matrix (including non-square matrix and non-symmetric matrix); 2. To find the subdifferential of $\|\boldsymbol{X}\|_*$, we need to derive

$$\partial\mathscr{F}\begin{pmatrix} \mathrm{Re}(\boldsymbol{X}) \\ \mathrm{Im}(\boldsymbol{X}) \end{pmatrix},$$

for which we have Lemma 7.3. Note that in our earlier work [3], we have already shown one direction of (7.6), namely

$$\mathfrak{H} = \left\{ \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} \;\middle|\; \boldsymbol{\alpha} + \imath\boldsymbol{\beta} \in \mathfrak{S} \right\} \subseteq \partial\mathscr{F}\left(\begin{bmatrix} \mathrm{Re}(\boldsymbol{X}) \\ \mathrm{Im}(\boldsymbol{X}) \end{bmatrix} \right).$$

Now we show the two sets are indeed equal. The proof of Lemma 7.3 is given in Appendix.

Proof of Theorem 7.1: We first consider the case where all the elements of $\boldsymbol{A}$ are real numbers. Without loss of generality, we consider

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{I}_{k \times k} & \boldsymbol{0}_{k \times (m-k)} \\ \boldsymbol{0}_{(m-k) \times k} & \boldsymbol{0}_{(m-k) \times (m-k)} \end{bmatrix}, \tag{7.7}$$

and

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{I}_{k\times k} & \boldsymbol{0}_{k\times(n-k)} \\ \boldsymbol{0}_{(n-k)\times k} & \boldsymbol{0}_{(n-k)\times(n-k)} \end{bmatrix}. \tag{7.8}$$

Then

$$\boldsymbol{PAQ}^* = \begin{bmatrix} \begin{bmatrix} \boldsymbol{A}_{1,1} & \cdots & \boldsymbol{A}_{1,k} \\ \vdots & \ddots & \vdots \\ \boldsymbol{A}_{k,1} & \cdots & \boldsymbol{A}_{k,k} \end{bmatrix} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}, \tag{7.9}$$

and

$$\begin{aligned} (\boldsymbol{I}-\boldsymbol{P})\boldsymbol{A}(\boldsymbol{I}-\boldsymbol{Q})^* &= \boldsymbol{A} - \boldsymbol{PA} - \boldsymbol{AQ}^* + \boldsymbol{PAQ}^* \\ &= \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \begin{bmatrix} \boldsymbol{A}_{k+1,k+1} & \cdots & \boldsymbol{A}_{k+1,n} \\ \vdots & \ddots & \vdots \\ \boldsymbol{A}_{m,k+1} & \cdots & \boldsymbol{A}_{m,n} \end{bmatrix} \end{bmatrix}. \end{aligned} \tag{7.10}$$

Thus,

$$\|\boldsymbol{PAQ}^*\|_* = \|\boldsymbol{A}_{1:k,1:k}\|_*,\ \|(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{A}(\boldsymbol{I}-\boldsymbol{Q})^*\|_* = \|\boldsymbol{A}_{k+1:m,k+1:n}\|_*. \tag{7.11}$$

We will prove this theorem by contradiction. We first show that if $\sigma_1 + \cdots + \sigma_k < \sigma_{k+1} + \cdots + \sigma_t$ holds for $\boldsymbol{A}$, then for any matrix $\boldsymbol{X} \in \mathbb{C}^{m\times n}$ with rank no more than k, for any positive number $l > 0$, we have

$$\|\boldsymbol{X} + l\boldsymbol{A}\|_* > \|\boldsymbol{X}\|_* \tag{7.12}$$

The proof of (7.12) follows similar arguments as in [29]. Since $\boldsymbol{X}$ has a rank at most k, then $\sigma_i(\boldsymbol{X}) = 0, \forall i > k$. Then we have

$$\begin{aligned} \|\boldsymbol{X} + l\boldsymbol{A}\|_* &\geq \sum_{i=1}^{t} |\sigma_i(\boldsymbol{X}) - \sigma_i(l\boldsymbol{A})| \\ &= \sum_{i=1}^{k} |\sigma_i(\boldsymbol{X}) - \sigma_i(l\boldsymbol{A})| + \sum_{i=k+1}^{t} |\sigma_i(\boldsymbol{X}) - \sigma_i(l\boldsymbol{A})| \\ &\geq \sum_{i=1}^{k} (\sigma_i(\boldsymbol{X}) - \sigma_i(l\boldsymbol{A})) + \sum_{i=k+1}^{t} |\sigma_i(\boldsymbol{X}) - \sigma_i(l\boldsymbol{A})| \end{aligned}$$

$$\geq \sum_{i=1}^{k}(\sigma_i(\boldsymbol{X}) - \sigma_i(l\boldsymbol{A})) + \sum_{i=k+1}^{t} |\sigma_i(l\boldsymbol{A})|$$
$$\geq \sum_{i=1}^{k}\sigma_i(\boldsymbol{X}) + \left(\sum_{i=k+1}^{t}\sigma_i(l\boldsymbol{A}) - \sum_{i=1}^{k}\sigma_i(l\boldsymbol{A})\right)$$
$$> \sum_{i=1}^{k}\sigma_i(\boldsymbol{X}) = \|\boldsymbol{X}\|_*,$$

where, for the first inequality, we used the Lemma 7.2.

We next show that if $\|\boldsymbol{A}_{1:k,1:k}\|_* > \|\boldsymbol{A}_{(k+1):m,(k+1):n}\|_*$, one can construct a matrix $\boldsymbol{X}$ with rank at most k such that

$$\|\boldsymbol{X} + l\boldsymbol{A}\|_* \leq \|\boldsymbol{X}\|_*$$

for a certain $l > 0$. We divide this construction into two cases: when $\boldsymbol{A}_{1:k,1:k}$ has rank equal to k, and when $\boldsymbol{A}_{1:k,1:k}$ has rank smaller than k.

When $\boldsymbol{A}_{1:k,1:k}$ has rank k, we denote its SVD as

$$\boldsymbol{A}_{1:k,1:k} = \boldsymbol{U}_1\boldsymbol{\Sigma}\boldsymbol{V}_1{}^*.$$

Then the SVD of $\begin{bmatrix} \boldsymbol{A}_{1:k,1:k} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$ is given by

$$\begin{bmatrix} \boldsymbol{A}_{1:k,1:k} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{U}_1 \\ \boldsymbol{0} \end{bmatrix} \boldsymbol{\Sigma} \begin{bmatrix} \boldsymbol{V}_1 \\ \boldsymbol{0} \end{bmatrix}^* \tag{7.13}$$

We now construct

$$\boldsymbol{X} = -\begin{bmatrix} \boldsymbol{U}_1 \\ \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_1 \\ \boldsymbol{0} \end{bmatrix}^*.$$

Let us denote

$$\boldsymbol{U}_2 = \begin{bmatrix} \boldsymbol{U}_1 \\ \boldsymbol{0} \end{bmatrix}$$

and

$$\boldsymbol{V}_2 = \begin{bmatrix} \boldsymbol{V}_1 \\ \boldsymbol{0} \end{bmatrix},$$

then the subdifferential of $\|\cdot\|_*$ at $\boldsymbol{X}$ is given by

$$\partial\|\boldsymbol{X}\|_* = \{\boldsymbol{Z} : \boldsymbol{Z} = -\boldsymbol{U}_2\boldsymbol{V}_2^* + \boldsymbol{M}, \text{where } \|\boldsymbol{M}\|_2 \leq 1, \boldsymbol{M}^*\boldsymbol{X} = \boldsymbol{0}, \boldsymbol{X}\boldsymbol{M}^* = \boldsymbol{0}\}. \tag{7.14}$$

For any $\boldsymbol{Z} \in \partial \|\boldsymbol{X}\|_*$,

$$\langle \boldsymbol{Z}, \boldsymbol{A} \rangle = -I_1 + I_2, \tag{7.15}$$

where

$$I_1 = \mathrm{Tr}(\boldsymbol{V}_2 \boldsymbol{U}_2^* \boldsymbol{A}),\ I_2 = \mathrm{Tr}(\boldsymbol{M}^* \boldsymbol{A}). \tag{7.16}$$

Let us partition the matrix $\boldsymbol{A}$ into four blocks:

$$\begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}, \tag{7.17}$$

where $\boldsymbol{A}_{11} \in \mathbb{R}^{k\times k}, \boldsymbol{A}_{12} \in \mathbb{R}^{k\times(n-k)}, \boldsymbol{A}_{21} \in \mathbb{R}^{(m-k)\times k}$ and $\boldsymbol{A}_{21} \in \mathbb{R}^{(m-k)\times(n-k)}$. Then we have

$$\begin{aligned} I_1 &= \mathrm{Tr}(\boldsymbol{V}_2 \boldsymbol{U}_2{}^* \boldsymbol{A}) \\ &= \mathrm{Tr}\left(\begin{bmatrix} \boldsymbol{V}_1 \\ \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_1^* & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} \right) \\ &= \mathrm{Tr}\left(\begin{bmatrix} \boldsymbol{V}_1 \boldsymbol{U}_1^* & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} \right) \\ &= \mathrm{Tr}(\boldsymbol{V}_1 \boldsymbol{U}_1^* \boldsymbol{A}_{11}) = \mathrm{Tr}(\boldsymbol{V}_1 \boldsymbol{U}_1^* \boldsymbol{U}_1 \boldsymbol{\Sigma} \boldsymbol{V}_1^*) \\ &= \sum_{i=1}^{k} \sigma_i(\boldsymbol{A}_{11}) \\ &= \|\boldsymbol{A}_{11}\|_*. \end{aligned} \tag{7.18}$$

Since $\boldsymbol{M}^* \boldsymbol{X} = \boldsymbol{0}$, $\boldsymbol{X}\boldsymbol{M}^* = \boldsymbol{0}$, $\|\boldsymbol{M}\|_2 \le 1$ and $\boldsymbol{X}$ is a rank-k left top corner matrix, we must have

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{M}_{22} \end{bmatrix},$$

where $\boldsymbol{M}_{22}$ is of dimension $(m-k)\times(n-k)$, and $\|\boldsymbol{M}_{22}\|_2 \le 1$.

Then

$$I_2 = \mathrm{Tr}(\boldsymbol{M}^* \boldsymbol{A}) \tag{7.19}$$

$$= \mathrm{Tr}\left(\begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{M}_{22}^* \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} \right) \tag{7.20}$$

$$= \mathrm{Tr}\left(\boldsymbol{M}_{22}^* \boldsymbol{A}_{22}\right) \le \|\boldsymbol{A}_{22}\|_*, \tag{7.21}$$

where the last inequality is from the fact that the nuclear norm is the dual norm of the spectral norm, and $\|\boldsymbol{M}_{22}\|_2 \le 1$.

Thus, we have

$$\langle \boldsymbol{Z}, \boldsymbol{A} \rangle = -I_1 + I_2 \leq -\|\boldsymbol{A}_{11}\|_* + \|\boldsymbol{A}_{22}\|_* < 0, \tag{7.22}$$

because we assume that $\|\boldsymbol{A}_{1:k,1:k}\|_* > \|\boldsymbol{A}_{(k+1):m,(k+1):n}\|_*$. From Theorem 23.4 in [37], we have

$$f'(\boldsymbol{X}; \boldsymbol{A}) = \sup_{\boldsymbol{Z} \in \partial \|\boldsymbol{X}\|_*} \langle \boldsymbol{Z}, \boldsymbol{A} \rangle < 0,$$

where $f'(\boldsymbol{X}; \boldsymbol{A})$ is defined as the one-sided directional derivative in [37]

$$f'(\boldsymbol{X}; \boldsymbol{A}) = \inf_{l>0} \frac{\|\boldsymbol{X} + l\boldsymbol{A}\|_* - \|\boldsymbol{X}\|_*}{l}.$$

This means that, when $\boldsymbol{A}_{11}$ has rank equal to k, we can always find $l > 0$ such that $\|\boldsymbol{X} + l\boldsymbol{A}\|_* - \|\boldsymbol{X}\|_* < 0$, or we can reduce the nuclear norm along direction specified by $\boldsymbol{A}$. Thus, there exists a positive number $l > 0$, such that

$$\|\boldsymbol{X} + l\boldsymbol{A}\|_* \leq \|\boldsymbol{X}\|_*.$$

Let us suppose instead that $\boldsymbol{A}_{11}$ has rank $b < k$. We can write the SVD of $\boldsymbol{A}_{11}$ as

$$\boldsymbol{A}_{11} = \begin{bmatrix} \boldsymbol{U}_1 & \boldsymbol{U}_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_1 & \boldsymbol{V}_3 \end{bmatrix}^*. \tag{7.23}$$

Then we construct

$$\boldsymbol{X} = -\begin{bmatrix} \boldsymbol{U}_1 & \boldsymbol{U}_3 \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_1 & \boldsymbol{V}_3 \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}^*.$$

By going through similar arguments as above (except for taking care of extra terms involving $\boldsymbol{U}_3$ and $\boldsymbol{V}_3$), one can obtain that $\langle \boldsymbol{Z}, \boldsymbol{A} \rangle < 0$ for every $\boldsymbol{Z} \in \partial \|\boldsymbol{X}\|_*$.

In summary, no matter whether $\boldsymbol{A}_{11}$ has rank equal to k or smaller to k, there always exists a positive number $l > 0$, such that$\|\boldsymbol{X} + l\boldsymbol{A}\|_* \leq \|\boldsymbol{X}\|_*$. However, this contradicts (7.12), and we conclude $\|\boldsymbol{A}_{1:k,1:k}\|_* \leq \|\boldsymbol{A}_{(k+1):m,(k+1):n}\|_*$, when $\boldsymbol{A}$ has real-numbered elements.

We further consider the case when $\boldsymbol{A}$ is a complex-numbered matrix. We first derive the subdifferential of $\|\boldsymbol{X}\|_*$ for any complex-numbered matrix $m \times n$ $\boldsymbol{X}$. For any $\boldsymbol{\alpha} \in \mathbb{R}^{m \times n}$ and any $\boldsymbol{\beta} \in \mathbb{R}^{m \times n}$, we define $\mathscr{F} : \mathbb{R}^{2m \times n} \mapsto \mathbb{R}$ as

$$\mathscr{F}\left(\begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix}\right) = \|(\boldsymbol{\alpha} + \imath\boldsymbol{\beta})\|_*. \tag{7.24}$$

For a complex-numbered matrix $\boldsymbol{A}$, we will similarly show that, if $\|\boldsymbol{A}_{1:k,1:k}\|_* > \|\boldsymbol{A}_{(k+1):m,(k+1):n}\|_*$, one can construct a matrix $\boldsymbol{X}$ with rank at most k such that

$$\|\boldsymbol{X} + l\boldsymbol{A}\|_* \leq \|\boldsymbol{X}\|_*$$

for a certain $l > 0$. We divide this construction into two cases: when $\boldsymbol{A}_{1:k,1:k}$ has rank equal to k, and when $\boldsymbol{A}_{1:k,1:k}$ has rank smaller than k.

When $\boldsymbol{A}_{1:k,1:k}$ has rank k, we denote its SVD as

$$\boldsymbol{A}_{1:k,1:k} = \boldsymbol{U}_1 \boldsymbol{\Sigma} \boldsymbol{V}_1{}^*.$$

Then the SVD of $\begin{bmatrix} \boldsymbol{A}_{1:k,1:k} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$ is given by

$$\begin{bmatrix} \boldsymbol{A}_{1:k,1:k} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{U}_1 \\ \boldsymbol{0} \end{bmatrix} \boldsymbol{\Sigma} \begin{bmatrix} \boldsymbol{V}_1 \\ \boldsymbol{0} \end{bmatrix}^*. \tag{7.25}$$

We now construct

$$\boldsymbol{X} = -\begin{bmatrix} \boldsymbol{U}_1 \\ \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_1 \\ \boldsymbol{0} \end{bmatrix}^*.$$

We denote

$$\boldsymbol{U}_2 = \begin{bmatrix} \boldsymbol{U}_1 \\ \boldsymbol{0} \end{bmatrix}, \text{ and } \boldsymbol{V}_2 = \begin{bmatrix} \boldsymbol{V}_1 \\ \boldsymbol{0} \end{bmatrix},$$

then by Lemma 7.3, the subdifferential of $\|\cdot\|_*$ at $\boldsymbol{X}$ is given by

$$\partial \|\boldsymbol{X}\|_* = \{\boldsymbol{Z} : \boldsymbol{Z} = -\boldsymbol{U}_2 \boldsymbol{V}_2^* + \boldsymbol{M}, \text{where } \|\boldsymbol{M}\|_2 \leq 1, \boldsymbol{M}^* \boldsymbol{X} = \boldsymbol{0}, \boldsymbol{X}\boldsymbol{M}^* = \boldsymbol{0}\}. \tag{7.26}$$

Since we define $\langle \boldsymbol{M}, \boldsymbol{N} \rangle = \mathrm{Re}(\mathrm{Tr}(\boldsymbol{M}^* \boldsymbol{N}))$ for complex valued matrices $\boldsymbol{M}, \boldsymbol{N}$, then for any $\boldsymbol{Z} \in \partial \|\boldsymbol{X}\|_*$, we have

$$\langle \boldsymbol{Z}, \boldsymbol{A} \rangle = -I_1 + I_2, \tag{7.27}$$

where

$$I_1 = \mathrm{Re}\left(\mathrm{Tr}(\boldsymbol{V}_2 \boldsymbol{U}_2^* \boldsymbol{A})\right), I_2 = \mathrm{Re}\left(\mathrm{Tr}(\boldsymbol{M}^* \boldsymbol{A})\right). \tag{7.28}$$

Similar to the real-numbered case, let us partition the matrix $\boldsymbol{A}$ into four blocks:

$$\begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}, \tag{7.29}$$

where $\boldsymbol{A}_{11} \in \mathbb{C}^{k\times k}, \boldsymbol{A}_{12} \in \mathbb{C}^{k\times(n-k)}, \boldsymbol{A}_{21} \in \mathbb{C}^{(m-k)\times(n-k)}$ and $\boldsymbol{A}_{21} \in \mathbb{C}^{(m-k)\times(n-k)}$. We still have

$$\begin{aligned}\mathrm{Tr}(\boldsymbol{V}_2\boldsymbol{U}_2{}^*\boldsymbol{A}) &= \mathrm{Tr}\left(\begin{bmatrix}\boldsymbol{V}_1\\ \boldsymbol{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{U}_1^* & \boldsymbol{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{bmatrix}\right)\\ &= \mathrm{Tr}\left(\begin{bmatrix}\boldsymbol{V}_1\boldsymbol{U}_1^* & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{bmatrix}\right)\\ &= \mathrm{Tr}(\boldsymbol{V}_1\boldsymbol{U}_1^*\boldsymbol{A}_{11}) = \mathrm{Tr}(\boldsymbol{V}_1\boldsymbol{U}_1^*\boldsymbol{U}_1\boldsymbol{\Sigma}\boldsymbol{V}_1^*) = \sum_{i=1}^{k}\sigma_i(\boldsymbol{A}_{11}) = \|\boldsymbol{A}_{11}\|_*. \end{aligned} \tag{7.30}$$

So

$$I_1 = \mathrm{Re}\left(\mathrm{Tr}(\boldsymbol{V}_2\boldsymbol{U}_2{}^*\boldsymbol{A})\right) = \|\boldsymbol{A}_{11}\|_*.$$

Since $\boldsymbol{M}^*\boldsymbol{X} = \boldsymbol{0}$, $\boldsymbol{X}\boldsymbol{M}^* = \boldsymbol{0}$, $\|\boldsymbol{M}\|_2 \le 1$ and $\boldsymbol{X}$ is a rank-k left top corner matrix, we must have

$$M = \begin{bmatrix}\boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{M}_{22}\end{bmatrix},$$

where $\boldsymbol{M}_{22}$ is of dimension $(m-k)\times(n-k)$, and $\|\boldsymbol{M}_{22}\|_2 \le 1$. Then we have

$$I_2 = \mathrm{Re}\left(\mathrm{Tr}(\boldsymbol{M}^*\boldsymbol{A})\right) \tag{7.31}$$

$$= \mathrm{Re}\left(\mathrm{Tr}\left(\begin{bmatrix}\boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{M}_{22}^*\end{bmatrix}\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{bmatrix}\right)\right) \tag{7.32}$$

$$= \mathrm{Re}\left(\mathrm{Tr}\left(\boldsymbol{M}_{22}^*\boldsymbol{A}_{22}\right)\right) \le \|\boldsymbol{A}_{22}\|_*, \tag{7.33}$$

where the last inequality is because the nuclear norm is the dual norm of the spectral norm, and $\|\boldsymbol{M}_{22}\|_2 \le 1$.

Thus, we have

$$\langle \boldsymbol{Z}, \boldsymbol{A}\rangle = -I_1 + I_2 \le -\|\boldsymbol{A}_{11}\|_* + \|\boldsymbol{A}_{22}\|_* < 0, \tag{7.34}$$

because we assume that $\|\boldsymbol{A}_{1:k,1:k}\|_* > \|\boldsymbol{A}_{(k+1):m,(k+1):n}\|_*$. Following the same arguments in the real numbered case, there exists a positive number $l > 0$, such that

$$\|\boldsymbol{X} + l\boldsymbol{A}\|_* \le \|\boldsymbol{X}\|_*.$$

Let us suppose instead that $\boldsymbol{A}_{11}$ has rank $b < k$. We can write the SVD of $\boldsymbol{A}_{11}$ as

$$\boldsymbol{A}_{11} = \begin{bmatrix}\boldsymbol{U}_1 & \boldsymbol{U}_3\end{bmatrix}\begin{bmatrix}\boldsymbol{\Sigma} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{V}_1 & \boldsymbol{V}_3\end{bmatrix}^*. \tag{7.35}$$

Then we construct

$$\boldsymbol{X} = -\begin{bmatrix} \boldsymbol{U}_1 & \boldsymbol{U}_3 \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_1 & \boldsymbol{V}_3 \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}^*.$$

By going through similar arguments as above (except for taking care of extra terms involving $\boldsymbol{U}_3$ and $\boldsymbol{V}_3$), one can obtain that $\langle \boldsymbol{Z}, \boldsymbol{A} \rangle < 0$ for every $\boldsymbol{Z} \in \partial \|\boldsymbol{X}\|_*$.

In summary, no matter whether complex-numbered $\boldsymbol{A}_{11}$ has rank equal to k or smaller to k, there always exists a positive number $l > 0$, such that $\|\boldsymbol{X} + l\boldsymbol{A}\|_* \leq \|\boldsymbol{X}\|_*$. However, this contradicts (7.12), and we conclude $\|\boldsymbol{A}_{1:k,1:k}\|_* \leq \|\boldsymbol{A}_{(k+1):m,(k+1):n}\|_*$. □

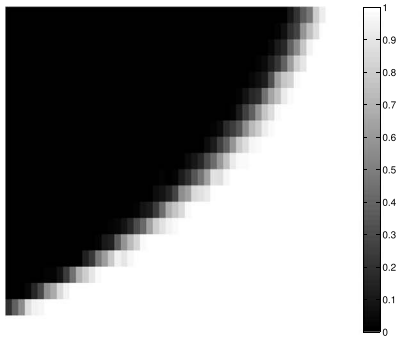
8. Numerical results

In this section, we perform numerical experiments to demonstrate the empirical performance of Hankel matrix recovery, and show its robustness to the separations between atoms. We use superpositions of complex sinusoids as test signals. But we remark that Hankel matrix recovery can also work for superpositions of complex exponentials. We consider the non-uniform sampling of entries studied in [11, 41], where we uniformly randomly observe M entries (without replacement) of $\boldsymbol{x}$ from $\{0, 1, \ldots, 2N-2\}$. By non-uniform sampling, we refer to the following procedure: we randomly select $|\mathscr{M}|=M$ distinct entries of $\boldsymbol{x}$ whose locations are uniformly distributed over $\{0, 1, \cdots, 2N-2\}$. We note that the 'non-uniform sampling' is in contrast to 'uniform sampling' where every entry of $\boldsymbol{x}$ in $\{0, 1, \cdots, 2N-2\}$ is observed. We also consider two signal (frequency) reconstruction algorithms: the Hankel nuclear norm minimization and the atomic norm minimization.

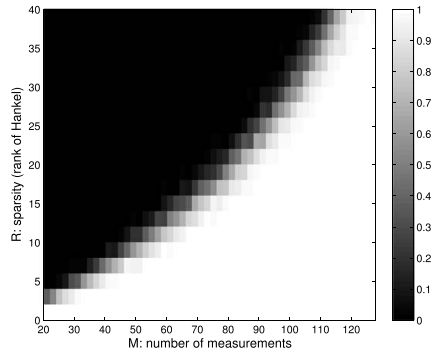
8.1 *Phase transition comparisons between Hankel matrix recovery and atomic norm minimization*

We fix $N = 64$, i.e. the dimension of the ground truth signal $\boldsymbol{x}$ is 127. We conduct experiments under different M and R for different approaches. For each approach with a fixed M and R, we test 100 trials, where each trial is performed as follows. We first generate the true signal $\boldsymbol{x} = [\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{126}]^T$ with $\boldsymbol{x}_j = \sum_{k=1}^{R} c_k e^{t 2\pi f_k j}$ for $j = 0, 1, \ldots, 126$, where f_k are frequencies drawn from the interval $[0, 1)$ uniformly at random, and c_k are complex coefficients satisfying the model $c_k = (1 + 10^{0.5 m_k}) e^{i2\pi\theta_k}$ with m_k and θ_k uniformly randomly drawn from the interval $[0, 1]$. Let the reconstructed signal be represented by $\hat{\boldsymbol{x}}$. If $\frac{\|\hat{\boldsymbol{x}}-\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \leq 10^{-3}$, then we regard it as a successful reconstruction. We also provide the simulation results under the Gaussian measurements of $\boldsymbol{x}$ as in [3].

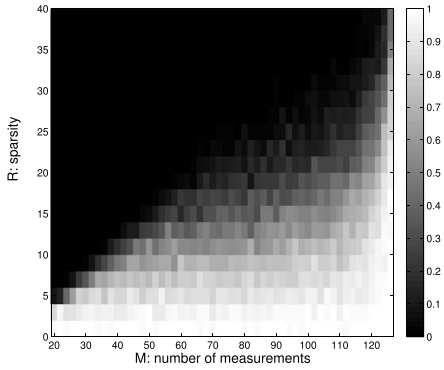
We plot in Fig. 2 the rate of successful reconstruction with respect to different M and R for different approaches. The black and white regions indicate a 0% and 100% of successful reconstruction, respectively, and the grey regions represent successful recovery rate between 0% and 100%. From this figure, we see that the atomic norm minimization still suffers from non-negligible probability of failure even if the number of measurements approach the full 127 samples. The reason is that, since the underlying frequencies are randomly chosen, there is a sizable probability that some frequencies are close to each other. When the frequencies are close to each other violating the atom separation condition, the atomic norm minimization can still fail even if we observe the full 127 samples. By comparison, the Hankel matrix recovery approach experiences a sharper phase transition, and is robust to the frequency separations. We also see that under both the Gaussian projection and the non-uniform sampling models, the atomic norm minimization and the Hankel matrix recovery approach have similar performance.



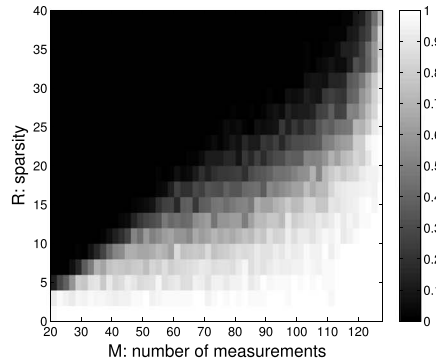
(a) Hankel nuclear norm minimization with random Gaussian projections



(b) Hankel nuclear norm minimization with non-uniform sampling of entries



(c) Atomic norm minimization with random Gaussian projections



(d) Atomic norm minimization with non-uniform sampling of entries

FIG. 2. Performance comparisons between atomic norm minimization and Hankel matrix recovery

8.2 *Robustness of Hankel matrix recovery to small frequency separations*

We further demonstrate the robustness of the Hankel matrix recovery approach to the separations between frequency atoms, as we vary the separations between frequencies. In our first set of experiments, we take $N = 64, |\mathscr{M}| = M = 65$ ($\approx 51\%$ sampling rate) and $R = 8$, and consider noiseless measurements. Again we generate the magnitude of the coefficients as $1 + 10^{0.5p}$, where p is uniformly randomly generated from $[0, 1)$, and the realized magnitudes are 3.1800, 2.5894, 2.1941, 2.9080, 3.9831, 4.0175, 4.1259, 3.6182 in this experiment. The corresponding phases of the coefficients are randomly generated as $2\pi s$, where s is uniformly randomly generated from $[0, 1)$. In this experiment, the realized phases are 4.1097, 5.4612, 5.4272, 4.7873, 1.0384, 0.4994, 3.1975 and 0.5846. The first $R - 1 = 7$ frequencies of exponentials are generated uniformly randomly over $[0, 1)$, and then the last frequency is added in the proximity of the third frequency. In our six experiments, the eighth frequency is chosen such that the frequency separation between the eighth frequency and the third frequency is, respectively,

0.03, 0.01, 0.003, 0.001, 0.0003 and 0.0001. Specifically, in our six experiments, the locations of the eight frequencies are, respectively, {0.3923, 0.9988, 0.3437, 0.9086, 0.6977, 0.0298, 0.4813, 0.3737}, {0.3923, 0.9988, 0.3437, 0.9086, 0.6977, 0.0298, 0.4813, 0.3537}, {0.3923, 0.9988, 0.3437, 0.9086, 0.6977, 0.0298, 0.4813, 0.3467}, {0.3923, 0.9988, 0.3437, 0.9086, 0.6977, 0.0298, 0.4813, 0.3447}, {0.3923, 0.9988, 0.3437, 0.9086, 0.6977, 0.0298, 0.4813, 0.3440} and {0.3923, 0.9988, 0.3437, 0.9086, 0.6977, 0.0298, 0.4813, 0.3438}. Hankel matrix recovery approach gives relative error $\frac{\|\hat{\boldsymbol{x}}-\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} = 8.2556 \times 10^{-9}$, 1.6709×10^{-9}, 2.9204×10^{-9}, 8.9444×10^{-9}, 8.6000×10^{-9} and 2.5026×10^{-9}, respectively. With the recovered data $\hat{\boldsymbol{x}}$, we use the MUSIC algorithm to identify the frequencies. The recovered frequencies for these six cases are, respectively, {0.0298, 0.3437, 0.3737, 0.3923, 0.4813, 0.6977, 0.9086, 0.9988}, {0.0298, 0.3437, 0.3537, 0.3923, 0.4813, 0.6977, 0.9086, 0.9988}, {0.0298, 0.3437, 0.3467, 0.3923, 0.4813, 0.6977, 0.9086, 0.9988}, {0.0298, 0.3437, 0.3447, 0.3923, 0.4813, 0.6977, 0.9086, 0.9988}, {0.0298, 0.3437, 0.3440, 0.3923, 0.4813, 0.6977, 0.9086, 0.9988} and {0.0298, 0.3437, 0.3438, 0.3923, 0.4813, 0.6977, 0.9086, 0.9988}. We illustrate these six cases in Fig. 3, where the peaks of the imaging function $J(f)$ are the locations of the recovered frequencies and the red cross marks indicate where the true frequencies are. We can see the Hankel matrix recovery successfully recovers the missing data and correctly locates the frequencies.

8.3 *Comparisons of Hankel matrix recovery and atomic norm minimization alongside each other: noiseless and noisy observations*

Below, we demonstrate the performance of these two different approaches, i.e. Hankel matrix recovery approach and atomic norm minimization, alongside each other. In both the noiseless-observation case and noisy-observation case (other formulations for noisy cases yield similar results), the atomic norm minimization is modelled as

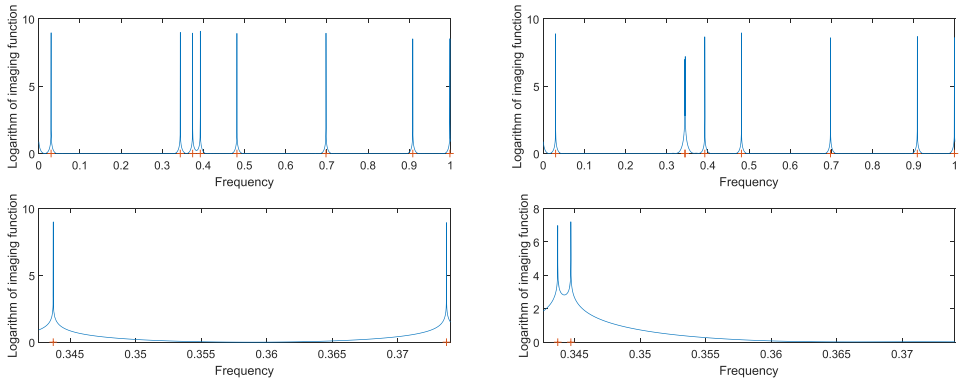
$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_{Atomic}, \quad \text{s.t. } \mathscr{A}(\boldsymbol{x}) = \boldsymbol{b}, \tag{8.1}$$

where $\|\cdot\|_{Atomic}$ is the atomic norm.

Let us first consider noiseless observations. When the frequency separations are 10^{-2}, 3×10^{-2}, 10^{-3}, 3×10^{-3}, 10^{-4} and 3×10^{-4}, respectively, the performance of frequency identification is shown in Fig 3 and 4. For atomic norm minimization, we have used the commonly used dual polynomial method to identify frequency atoms at the frequency locations where the magnitude of the dual polynomial is equal to 1. This means that we only need to identify the locations where the dual polynomial achieves a magnitude of 1 to recover the frequencies.

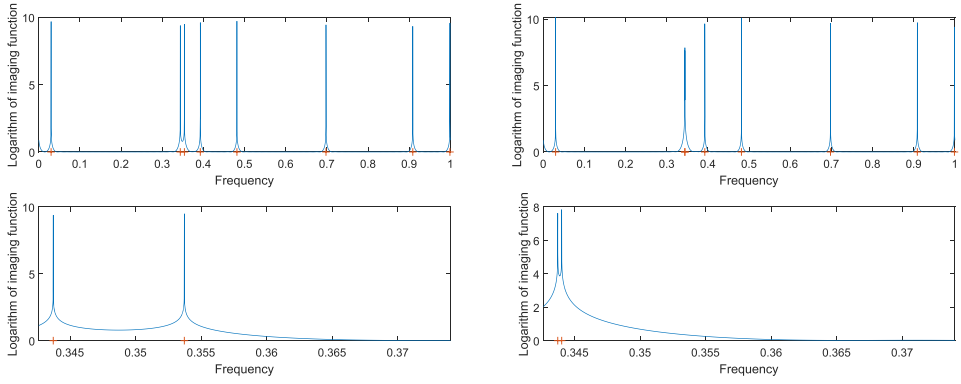
From the results in Fig. 4, we can see that when the frequency separation is big, e.g. 0.03 or 0.01, the atomic norm can identify the frequencies accurately, and the dual polynomial has exactly the same number of peaks as that of the underlying frequencies. Once the frequency separation is below a certain threshold, e.g. when the frequency separation is 0.003, the atomic norm minimization can fail to identify not only the frequencies that are close to each other, but also the frequencies which are far apart. In addition, there are many spurious peaks of the dual polynomial which do not identify the locations of the true frequencies. However, from the results in Fig. 3, the Hankel matrix recovery approach performs much better than the atomic norm minimization. The Hankel matrix recovery approach can still identify frequency atoms even if they are very close to each other, e.g. when the frequency separation is 0.0001 which is much smaller than 0.003.

We also provide quantitative results of signal recovery errors in Table 2, which clearly shows the advantages of Hankel matrix completion when frequency separation is small. We would also like to



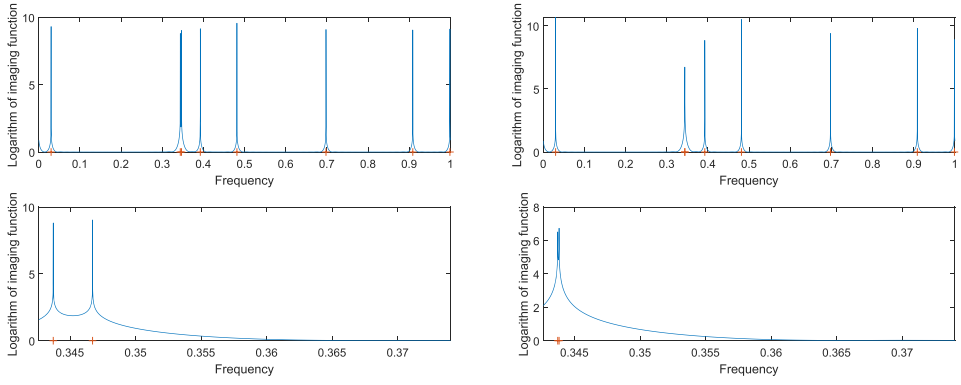
(a) dist$_{\mathscr{F}} = 0.03$.

(b) dist$_{\mathscr{F}} = 0.001$.



(c) dist$_{\mathscr{F}} = 0.01$.

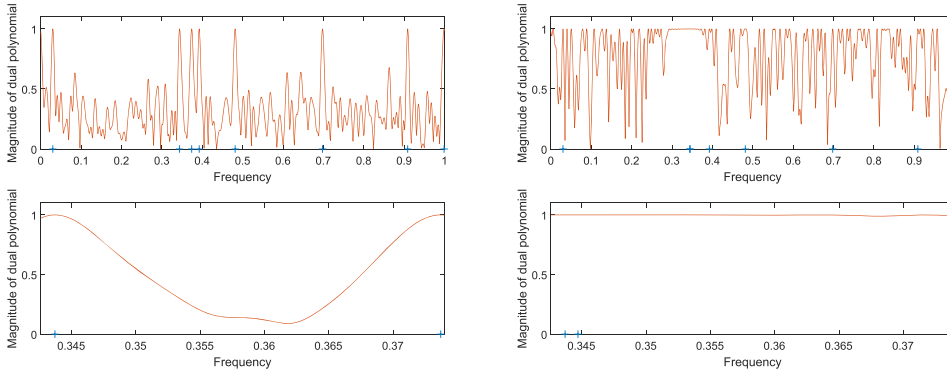
(d) dist$_{\mathscr{F}} = 0.0003$.



(e) dist$_{\mathscr{F}} = 0.003$.

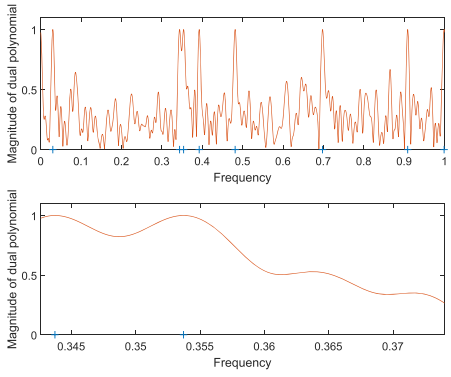
(f) dist$_{\mathscr{F}} = 0.0001$.

FIG. 3. Frequency identification performance with noiseless measurements: Hankel matrix recovery method is used.

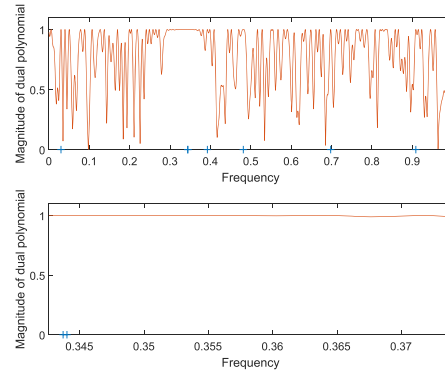


(a) $\mathrm{dist}_{\mathscr{F}} = 0.03$.

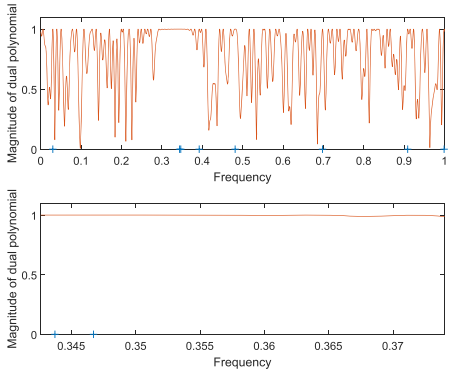
(b) $\mathrm{dist}_{\mathscr{F}} = 0.001$.



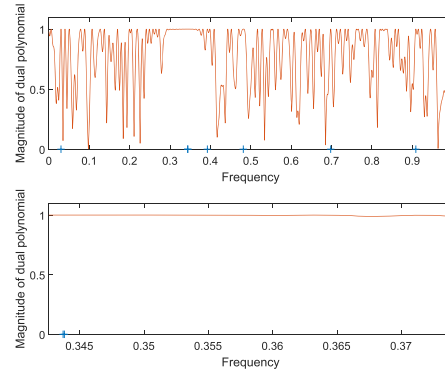
(c) $\mathrm{dist}_{\mathscr{F}} = 0.01$.



(d) $\mathrm{dist}_{\mathscr{F}} = 0.0003$.



(e) $\mathrm{dist}_{\mathscr{F}} = 0.003$.



(f) $\mathrm{dist}_{\mathscr{F}} = 0.0001$.

FIG. 4. Frequency identification performance with noiseless measurements: atomic norm minimization is used. According to [41], the frequencies identified by the atomic norm minimization are the locations where the dual polynomial achieves magnitude 1.

TABLE 2 *Reconstruction performance via Hankel matrix recovery approach (HMC) and atomic norm minimization (ANM).*

	Frequencies	*HMC* $\frac{\|\hat{\boldsymbol{x}}-\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2}$	*ANM* $\frac{\|\hat{\boldsymbol{x}}-\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2}$
Exp. 1	0.3923, 0.3437, 0.0298, 0.3737, 0.6977, 0.4813, 0.9988, 0.9086	8.2556×10^{-9}	8.2777×10^{-7}
Exp. 2	0.4813, 0.0298, 0.3923, 0.9988, 0.3537, 0.6977, 0.3437, 0.9086	1.6709×10^{-9}	5.6747×10^{-6}
Exp. 3	0.4813, 0.0298, 0.3923,0.9988, 0.6977, 0.9086, 0.3467, 0.3437	2.9204×10^{-9}	2.6958×10^{-3}
Exp. 4	0.4813, 0.0298, 0.9086, 0.3923, 0.9988, 0.6977, 0.3447, 0.3437	8.9444×10^{-9}	8.5770×10^{-4}
Exp. 5	0.0298, 0.4813, 0.9086, 0.6977, 0.3923, 0.9988, 0.3440, 0.3437	8.6000×10^{-9}	2.5235×10^{-4}
Exp. 6	0.0298, 0.4813, 0.9086, 0.6977, 0.9988, 0.3923, 0.3438, 0.3437	2.5026×10^{-9}	9.3546×10^{-5}

mention that when the separation is 3×10^{-3}, the atomic norm minimization achieves a reconstruction error $\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2$ of 0.2763 and a relative reconstruction error $\frac{\|\hat{\boldsymbol{x}}-\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2}$ of 2.6958×10^{-3}, while the Hankel matrix recovery can achieve a reconstruction error of 2.993×10^{-7} and a relative reconstruction error of 2.9204×10^{-9}.

We further demonstrate the performance of Hankel matrix recovery under noisy measurements. Again, we consider $N = 64$, $|\mathscr{M}| = 65$ and $R = 8$. The magnitudes of the coefficients of the complex sinusoids is obtained by $1 + 10^{0.5p}$, where p is uniformly randomly generated from $[0, 1)$. The realized magnitudes are 3.1800, 2.5894, 2.1941, 2.9080, 3.9831, 4.0175, 4.1259 and 3.6182, respectively, in our experiment. The phase of the coefficients are obtained by $2\pi s$, where s is uniformly randomly generated from $[0, 1)$. In this example, the realized phases are 4.1097, 5.4612, 5.4272, 4.7873, 1.0384, 0.4994, 3.1975 and 0.5846, respectively. The first $R-1 = 7$ frequencies of exponentials are generated uniformly randomly from $[0, 1)$, and then the last frequency is added with frequency separation 5×10^{-3} from the third frequency. The locations of the realized frequencies are 0.3923, 0.9988, 0.3437, 0.3487, 0.9086, 0.6977, 0.0298 and 0.4813.

We generate the noise vector $\boldsymbol{v} \in \mathbb{C}^{2N-1}$ as $s_1 + \imath s_2$ where each element of $s_1 \in \mathbb{R}^{2N-1}$ and $s_2 \in \mathbb{R}^{2N-1}$ is independently generated from the zero-mean unit-variance standard Gaussian distribution. And we further normalize $\boldsymbol{v}$ such that $\|\boldsymbol{v}\|_2 = 0.1$. In this noisy case, we solve the problem

$$\begin{aligned} &\min_{\boldsymbol{x}} \|\boldsymbol{H}(\boldsymbol{x})\|_* \\ &\text{subject to } \mathscr{A}(\boldsymbol{x}) = \boldsymbol{b}, \end{aligned} \tag{8.2}$$

and

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_{Atomic}, \text{ s.t. } \mathscr{A}(\boldsymbol{x}) = \boldsymbol{b}, \tag{8.3}$$

to get the recovered signal $\hat{\boldsymbol{x}}$. For ANM, the relative reconstruction error is 5.7905×10^{-3}, and the reconstruction error is 6.0706×10^{-1}. For HMC, the relative reconstruction error is 1.0666×10^{-3}, and the reconstruction error is 1.1182×10^{-1}.

We illustrate the locations of the recovered frequencies in Fig. 5. We can see that the Hankel matrix recovery can also provide robust data and frequency recovery under noisy measurements. Similar to the noiseless situation, the separation 5×10^{-3} can be too small for the atomic norm to identify the frequencies successfully, and there are many spurious peaks with magnitude 1. However, the Hankel matrix recovery approach can give relatively accurate identification of frequencies. For noise level $\|\boldsymbol{v}\|_2 = 0.5$ and 1, the recovery performance of ANM and HMC is demonstrated in Fig. 5.

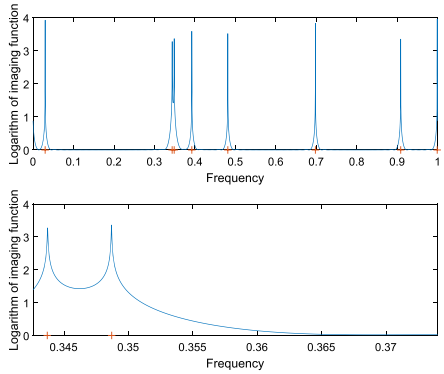
8.4 *Exploring the tightness of proposed theorems*

We now provide empirical evidence for the tightness of the proposed theorems. Although it is challenging to analytically prove the tightness of the bounds in these theorems at this moment, we provide empirical results to show the tightness or the sharpness of our results. We point out that even showing empirically the tightness of Theorem 3.1 (which establish worst-case performance guarantees) is difficult, since there are an infinite number of elements in the null space of the linear mapping for measurements. Hence, in this subsection, we only provide empirical evidences for the tightness of Theorems 4.1 and 4.3.

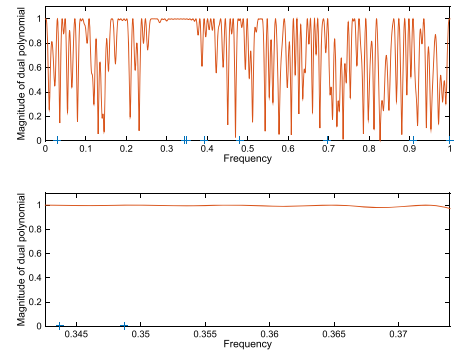
In this new set of experiments, we consider $N = 64$, and the number of atoms R is taken to be ranging from 2 to 64. The magnitude of the coefficient is $1 + 10^{0.5p}$, where p is uniformly generated from $[0, 1)$, and its phase is $2\pi s$ where s is generated uniformly from $[0, 1)$. For each choice of R, we do 100 trials and calculate the corresponding successful recovery rate. A trial is regarded as a success if the decoding algorithm can achieve $\frac{\|\hat{\boldsymbol{x}}-\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \leq 10^{-6}$. We consider the case where only the middle sample of the signal vector of length $(2N - 1)$ is missing, and this is the sampling pattern which appears in Theorems 4.1 and 4.3.

In the first set of experiments, all the frequencies are chosen with frequency separations as non-zero integer multiples of $\frac{1}{127}$, and the corresponding frequency atoms are orthogonal. The decoder observes all the samples except for the middle one, i.e. $|\mathscr{M}| = 126$. The results are presented in Part (a) of Fig. 6. As we can see, the HMC approach achieves far better performance than the ANM when the number of atoms increases and the separations between frequencies become smaller. When the number of frequency atoms is 30, the ANM achieves a successful recovery rate of almost 0, while the HMC approach achieves a successful rate of almost 1. Besides, the curve of HMC is quite sharp in the sense that the success rate drops quickly once the number of atoms exceeds a certain threshold. In Theorem 4.1, we predict that the recoverable number R of atoms approaches N, when N is large enough. The empirical results indeed show that the recoverable number R is around 60, which is very close to $N = 64$. This set of experimental results provide empirical evidence for the sharpness of Theorem 4.1.

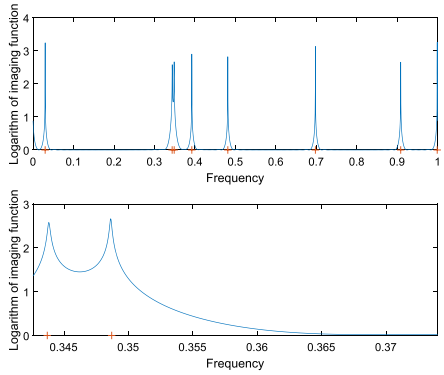
In the second set of experiments, all but one frequencies are chosen such that the corresponding atoms are orthogonal, while the last frequency is chosen such that it is away from the first frequency with a distance of 10^{-4}. All the other configurations are the same as in the experiments for orthogonal atoms. The numerical results for the second set of experiments are shown in Part (b) of Fig. 6. On the one hand, compared with numerical results for orthogonal atoms, we notice that the frequency separations greatly affect the performance of ANM: for example, when the number of frequency atoms is 10, the success rate of ANM under frequency separation 0.0001 is less than 0.5, while the success rate of ANM under separation $\frac{1}{127}$ is about 1. On the other hand, we notice that frequency separations do not affect much the successful recovery rate of the HMC approach, as predicted by our theoretical analysis. In Theorem 3, we analytically showed that the number R of recoverable atoms approaches N asymptotically under



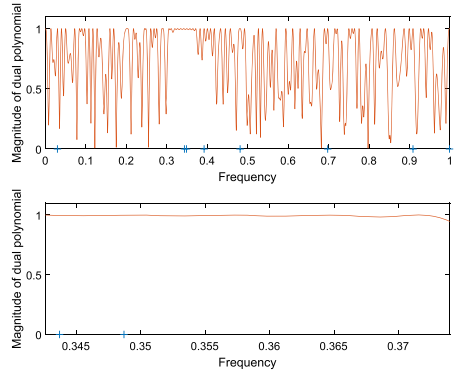
(a) dist$_{\mathscr{F}} = 0.005$ and $\|\boldsymbol{v}\|_2 = 0.1$.



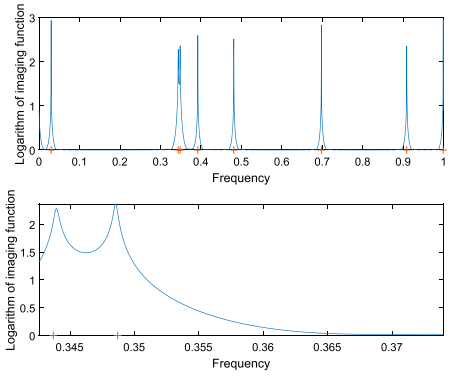
(b) dist$_{\mathscr{F}} = 0.005$ and $\|\boldsymbol{v}\|_2 = 0.1$.



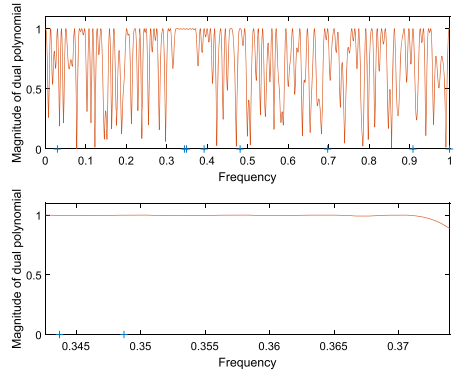
(c) dist$_{\mathscr{F}} = 0.005$ and $\|\boldsymbol{v}\|_2 = 0.5$.



(d) dist$_{\mathscr{F}} = 0.005$ and $\|\boldsymbol{v}\|_2 = 0.5$.



(e) dist$_{\mathscr{F}} = 0.005$ and $\|\boldsymbol{v}\|_2 = 1$.



(f) dist$_{\mathscr{F}} = 0.005$ and $\|\boldsymbol{v}\|_2 = 1$.

FIG. 5. Frequency identification performance with noisy measurements. Left column: Hankel matrix recovery. Right column: atomic norm minimization.

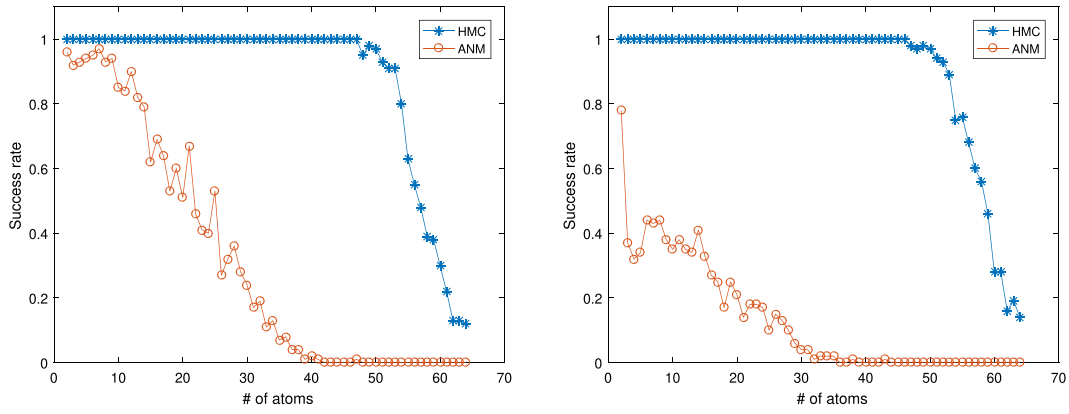


FIG. 6. Successful recovery rate when $N = 64$, and the middle sample of the signal vector of length $2N - 1 = 127$ is not observed. The left figure is for orthogonal atoms with minimum frequency separation no smaller than $\frac{1}{127}$; the right figure is for cases where there are two atoms very close to each other with frequency separation at most 10^{-4}.

the HMC approach, when only the middle sample of the signal vector is missing. The empirical results indeed show that the recoverable number R of frequencies is around 60, which is very close to $N = 64$, even when the frequency separation is small. This set of experimental results provide empirical evidence for the sharpness of Theorem 4.3.

9. Conclusions and future works

In this paper, we have shown, theoretically and numerically, that the Hankel matrix recovery can be robust to frequency separations in super-resolving the superposition of complex exponentials. By comparison, the TV minimization and atomic norm minimization require the underlying frequencies to be well-separated to super-resolve the superposition of complex exponentials even when the measurements are noiseless. In particular, we show that Hankel matrix recovery approach can super-resolve the R frequencies, regardless of how close the frequencies are to each other, from compressed non-uniform measurements. We presented a new concept of OANM, and showed that this concept helps us understand the success of Hankel matrix recovery in separation-free super-resolution. We further show that, in traditional atomic norm minimization, the underlying parameters *must* be well separated so that the signal can be successfully recovered if the atoms are changing continuously with respect to the continuously valued parameters; however, for OANM, it is possible the atomic norm minimization are successful even when the original atoms are arbitrarily close. As a byproduct of this research, we also provide one matrix-theoretic inequality of nuclear norm, and give its proof from the theory of compressed sensing. In future works, it would be interesting to extend the results in this paper to super-resolving the superposition of complex exponentials with higher dimensional frequency parameters [11, 13, 45].

Acknowledgement

Weiyu Xu's research is supported by ECCS-2000425 and ECCS-2133205.

Data Availability Statement

The data underlying this article will be shared on reasonable request to the corresponding author.

References

1. Bhatia, R. (1997) *Matrix analysis*, vol. **169**of *Graduate Texts in Mathematics*. Springer-VerlagNew York.
2. Cai, J., Candès, E. & Shen, Z. (2010) A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.*, **20**, 1956–1982.
3. Cai, J., Qu, X., Xu, W. & Ye, G. (2016) Robust recovery of complex exponential signals from random Gaussian projections via low rank Hankel matrix reconstruction. *Appl. Comput. Harmon. Anal.*, **41**, 470–490.
4. Cai, J., Wang, T. & Wei, K. (2019) Fast and provable algorithms for spectrally sparse signal reconstruction via low-rank Hankel matrix completion. *Appl. Comput. Harmon. Anal.*, **46**, 94–121.
5. Candès, E. & Fernandez-Granda, C. (2014) Towards a mathematical theory of super-resolution. *Comm. Pure Appl. Math.*, **67**, 906–956.
6. Candes, E. & Plan, Y. (2010) Matrix completion with noise. *Proc. IEEE*, **98**, 925–936.
7. Candès, E. & Recht, B. (2009) Exact matrix completion via convex optimization. *Found. Comput. Math.*, **9**, 717–772.
8. Candès, E., Romberg, J. & Tao, T. (2006) Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory*, **52**, 489–509.
9. Candès, E. & Tao, T. (2010) The power of convex relaxation: near-optimal matrix completion. *IEEE Trans. Inf. Theory*, **56**, 2053–2080.
10. Chandrasekaran, V., Recht, B., Parrilo, P. & Willsky, A. (2012) The convex geometry of linear inverse problems. *Found. Comput. Math.*, **12**, 805–849.
11. Chen, Y. & Chi., Y. (2014) Robust spectral compressed sensing via structured matrix completion. *IEEE Trans. Inf. Theory*, **60**, 6576–6601.
12. Chi, Y. (2017) Convex relaxations of spectral sparsity for robust super-resolution and line spectrum estimation., vol. **10394**. SPIE, pp. 314–321.
13. Chi, Y. & Chen, Y. (2014) Compressive two-dimensional harmonic retrieval via atomic norm minimization. *IEEE Trans. Signal Process.*, **63**, 1030–1042.
14. Chi, Y., Scharf, L., Pezeshki, A. & Calderbank, A. (2011) Sensitivity to basis mismatch in compressed sensing. *IEEE Trans. Signal Process.*, **59**, 2182–2195.
15. Dai, L. & Pelckmans, K. (2015) On the nuclear norm heuristic for a Hankel matrix completion problem. *Automatica J. IFAC*, **51**, 268–272.
16. Donoho, D. (2006) Compressed sensing. *IEEE Trans. Inf. Theory*, **52**, 1289–1306.
17. Fazel, M., Hindi, H. & Boyd, S. (2003) Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices. in *Proceedings of the American Control Conference*, vol. **3**, pp. 2156–2162. IEEE.
18. Fazel, M., Pong, T., Sun, D. & Tseng, P. (2013) Hankel matrix rank minimization with applications to system identification and realization. *SIAM J. Matrix Anal. Appl.*, **34**, 946–977.
19. Holland, D., Bostock, M., Gladden, L. & Nietlispach, D. (2011) Fast multidimensional NMR spectroscopy using compressed sensing. *Angew. Chem.*, **123**, 6678–6681.
20. Horn, R. & Johnson, C. (2012) *Matrix analysis*. Cambridge, UK: Cambridge University Press.
21. Hua, Y. & Sarkar, T. (1990) Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Trans. Acoust. Speech Signal Process.*, **38**, 814–824.
22. Kazimierczuk, K. & Orekhov, V. (2011) Accelerated NMR spectroscopy by using compressed sensing. *Angew. Chem.*, **50**, 5556–5559.
23. Li, W. & Sun, W. (2003) New perturbation bounds for unitary polar factors. *SIAM J. Matrix Anal. Appl.*, **25**, 362–372.

24. Liao, W. & Fannjiang, A. (2016) MUSIC for single-snapshot spectral estimation: stability and super-resolution. *Appl. Comput. Harmon. Anal.*, **40**, 33–67.
25. Lustig, M., Donoho, D. & Pauly, J. (2007) Sparse MRI: the application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.*, **58**, 1182–1195.
26. Markovsky, I. (2008) Structured low-rank approximation and its applications. *Automatica J. IFAC*, **44**, 891–909.
27. Mishra, K., Cho, M., Kruger, A. & Xu, W. (2015) Spectral super-resolution with prior knowledge. *IEEE Trans. Signal Process.*, **63**, 5342–5357.
28. Moitra, A. (2015) Super-resolution, extremal functions and the condition number of Vandermonde matrices. *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, 821–830.
29. Oymak, S. & Hassibi, B. (2010) New null space results and recovery thresholds for matrix rank minimization. arXiv preprint arXiv:1011.6326.
30. Plonka, G. & Tasche, M. (2014) Prony methods for recovery of structured functions. *GAMM-Mitteilungen*, **37**, 239–258.
31. Potts, D. & Tasche, M. (2010) Parameter estimation for exponential sums by approximate Prony method. *Signal Process.*, **90**, 1631–1642.
32. Potts, D. & Tasche, M. (2013) Parameter estimation for nonincreasing exponential sums by Prony-like methods. *Linear Algebr. Appl.*, **439**, 1024–1039.
33. Prony, R. (1795) Essai experimental et analytique: Sur les lois de la dilatabilite de fluides elastique et Sur celles de la force expansive de la vapeur de l'alkool, a differentes temperatures. *J. de l'Ecole Polytechnique*, **2**.
34. Qu, X., Mayzel, M., Cai, J., Chen, Z. & Orekhov, V. (2015) Accelerated NMR spectroscopy with low-rank reconstruction. *Angew. Chem.*, **54**, 852–854.
35. Recht, B., Fazel, M. & Parrilo, P. (2010) Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.*, **52**, 471–501.
36. Recht, B., Xu, W. & Hassibi, B. (2011) Null space conditions and thresholds for rank minimization. *Math. Program.*, **127**, 175–202.
37. Rockafellar, R. (2015) *Convex analysis*. Princeton, New Jersey: Princeton University Press.
38. Roy, R. & Kailath, T. (1989) ESPRIT-estimation of signal parameters via rotational invariance techniques. *IEEE Trans. Acoust. Speech Signal Process.*, **37**, 984–995.
39. Schiebinger, G., Robeva, E. & Recht, B. (2018) Superresolution without separation. *Inf. Inference: A Journal of the IMA*, **7**, 1–30.
40. Tang, G. (2015) Resolution limits for atomic decompositions via Markov-Bernstein type inequalities. in *Proceedings of International Conference on Sampling Theory and Applications (SampTA)*, pp. 548–552. IEEE.
41. Tang, G., Bhaskar, B., Shah, P. & Recht, B. (2013) Compressed sensing off the grid. *IEEE Trans. Inf. Theory*, **59**, 7465–7490.
42. Tropp, J. (2015) *An introduction to matrix Concentration inequalities*. Norwell, MA: Now Foundations and Trends.
43. Tropp, J., Laska, J., Duarte, M., Romberg, J. & Baraniuk, R. (2010) Beyond Nyquist: efficient sampling of sparse bandlimited signals. *IEEE Trans. Inf. Theory*, **56**, 520–544.
44. Usevich, K. & Comon, P. (2016) Hankel low-rank matrix completion: performance of the nuclear norm relaxation. *IEEE J. Sel. Top. Signal Process.*, **10**, 637–646.
45. Xu, W., Cai, J., Mishra, K., Cho, M. & Kruger, A. (2014) Precise semidefinite programming formulation of atomic norm minimization for recovering d-dimensional ($d \geq 2$) off-the-grid frequencies. *Proceedings of Information Theory and Applications (ITA) Workshop*. IEEE, pp. 1–4.
46. Yi, J. & Xu, W. (2020) Necessary and sufficient null space condition for nuclear norm minimization in low-rank matrix recovery. *IEEE Trans. Inf. Theory*, **66**, 6597–6604.

Appendix

Proof of Lemma 4.2

Any solution to the nuclear norm minimization must be $\boldsymbol{X}_0 + \boldsymbol{Q}$, where $\boldsymbol{Q}$ is from the null space of $\mathscr{A}$. Suppose that the SVD of $\boldsymbol{X}_0$ is given by

$$\boldsymbol{X}_0 = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^*,$$

where $\boldsymbol{U} \in \mathbb{C}^{M\times R}$, $\boldsymbol{\Lambda} \in \mathbb{C}^{R\times R}$ and $\boldsymbol{V} \in \mathbb{C}^{N\times R}$.

From Lemma 7.3, we know the subdifferential of $\|\cdot\|_*$ at the point $\boldsymbol{X}_0$ is given by

$$\{\boldsymbol{Z}\,|\boldsymbol{Z} = \boldsymbol{U}\boldsymbol{V}^* + \bar{\boldsymbol{U}}\boldsymbol{M}\bar{\boldsymbol{V}}^*, \text{ where } \|\boldsymbol{M}\|_2 \le 1, \boldsymbol{U}^*\bar{\boldsymbol{U}} = \boldsymbol{0}, \bar{\boldsymbol{U}}^*\bar{\boldsymbol{U}} = I, \boldsymbol{V}^*\bar{\boldsymbol{V}} = \boldsymbol{0}, \bar{\boldsymbol{V}}^*\bar{\boldsymbol{V}} = I\}.$$

Then from the property of subdifferential of a convex function, for any $\boldsymbol{Z} = \boldsymbol{U}\boldsymbol{V}^* + \bar{\boldsymbol{U}}\boldsymbol{M}\bar{\boldsymbol{V}}^*$ (with $\|\boldsymbol{M}\|_2 \le 1$) from the subdifferetial of $\|\cdot\|_*$ at the point $\boldsymbol{X}_0$, we have

$$\begin{aligned}
&\|\boldsymbol{X}_0 + \boldsymbol{Q}\|_* \\
&\ge \|\boldsymbol{X}_0\|_* + \langle \boldsymbol{Z}, \boldsymbol{Q}\rangle \\
&= \|\boldsymbol{X}_0\|_* + \mathrm{Re}\left(\mathrm{Tr}\left(\boldsymbol{U}^*\boldsymbol{Q}\boldsymbol{V}\right)\right) + \mathrm{Re}\left(\mathrm{Tr}\left(\bar{\boldsymbol{U}}^*\boldsymbol{Q}\bar{\boldsymbol{V}}\boldsymbol{M}^*\right)\right) \\
&\ge \|\boldsymbol{X}_0\|_* - |\mathrm{Tr}\left(\boldsymbol{U}^*\boldsymbol{Q}\boldsymbol{V}\right)| + \mathrm{Re}\left(\mathrm{Tr}\left(\bar{\boldsymbol{U}}^*\boldsymbol{Q}\bar{\boldsymbol{V}}\boldsymbol{M}^*\right)\right).
\end{aligned}$$

Because we can take any $\boldsymbol{M}$ with $\|\boldsymbol{M}\|_2 \le 1$, when $\boldsymbol{Q} \ne \boldsymbol{0}$, we have

$$\begin{aligned}
\|\boldsymbol{X}_0 + \boldsymbol{Q}\|_* &\ge \|\boldsymbol{X}_0\|_* - |\mathrm{Tr}\left(\boldsymbol{U}^*\boldsymbol{Q}\boldsymbol{V}\right)| + \|\bar{\boldsymbol{U}}^*\boldsymbol{Q}\bar{\boldsymbol{V}}\|_* \\
&> \|\boldsymbol{X}_0\|_*,
\end{aligned}$$

where we used the fact that the dual norm of the spectral norm is the nuclear norm (which also holds for complex-numbered matrices). Thus, $\boldsymbol{X}_0$ is the unique solution to the nuclear norm minimization.

Proof of Lemma 7.3

Proof. We write

$$\boldsymbol{U} = \boldsymbol{\Theta}_1 + \imath\boldsymbol{\Theta}_2, \quad \boldsymbol{V} = \boldsymbol{\Xi}_1 + \imath\boldsymbol{\Xi}_2, \tag{9.1}$$

where $\boldsymbol{\Theta}_1 \in \mathbb{R}^{M\times R}$, $\boldsymbol{\Theta}_2 \in \mathbb{R}^{M\times R}$, $\boldsymbol{\Xi}_1 \in \mathbb{R}^{N\times R}$, and $\boldsymbol{\Xi}_2 \in \mathbb{R}^{N\times R}$. Then, by direct calculation,

$$\boldsymbol{\Theta} \equiv \begin{bmatrix} \boldsymbol{\Theta}_1 & -\boldsymbol{\Theta}_2 \\ \boldsymbol{\Theta}_2 & \boldsymbol{\Theta}_1 \end{bmatrix} \in \mathbb{R}^{2M\times 2R}, \qquad \boldsymbol{\Xi} \equiv \begin{bmatrix} \boldsymbol{\Xi}_1 & -\boldsymbol{\Xi}_2 \\ \boldsymbol{\Xi}_2 & \boldsymbol{\Xi}_1 \end{bmatrix} \in \mathbb{R}^{2N\times 2R} \tag{9.2}$$

satisfy $\boldsymbol{\Theta}^T\boldsymbol{\Theta} = \boldsymbol{\Xi}^T\boldsymbol{\Xi} = \boldsymbol{I}$. Moreover, if we define $\hat{\boldsymbol{\Omega}} = \begin{bmatrix} \mathrm{Re}(\boldsymbol{X}) & -\mathrm{Im}(\boldsymbol{X}) \\ \mathrm{Im}(\boldsymbol{X}) & \mathrm{Re}(\boldsymbol{X}) \end{bmatrix}$, then

$$\hat{\boldsymbol{\Omega}} = \boldsymbol{\Theta}\begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma} \end{bmatrix}\boldsymbol{\Xi}^T \tag{9.3}$$

is an SVD of the real-numbered matrix $\hat{\boldsymbol{\Omega}}$, and the singular values $\hat{\boldsymbol{\Omega}}$ are those of $\boldsymbol{X}$, each repeated twice. Therefore,

$$\mathscr{F}\left(\begin{bmatrix}\mathrm{Re}(\boldsymbol{X})\\ \mathrm{Im}(\boldsymbol{X})\end{bmatrix}\right) = \|\boldsymbol{\Sigma}\|_* = \frac{1}{2}\|\hat{\boldsymbol{\Omega}}\|_*. \tag{9.4}$$

Define a linear operator $\mathscr{E}: \mathbb{R}^{2M\times N} \mapsto \mathbb{R}^{2M\times 2N}$ by

$$\mathscr{E}\left(\begin{bmatrix}\boldsymbol{\alpha}\\ \boldsymbol{\beta}\end{bmatrix}\right) = \begin{bmatrix}\boldsymbol{\alpha} & -\boldsymbol{\beta}\\ \boldsymbol{\beta} & \boldsymbol{\alpha}\end{bmatrix}, \quad \text{with} \quad \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^{M\times N}.$$

By (9.4) and the definition of $\hat{\boldsymbol{\Omega}}$, we obtain

$$\mathscr{F}\left(\begin{bmatrix}\mathrm{Re}(\boldsymbol{X})\\ \mathrm{Im}(\boldsymbol{X})\end{bmatrix}\right) = \frac{1}{2}\left\|\mathscr{E}\left(\begin{bmatrix}\mathrm{Re}(\boldsymbol{X})\\ \mathrm{Im}(\boldsymbol{X})\end{bmatrix}\right)\right\|_*$$

.

From convex analysis and $\hat{\Omega} = \mathscr{E}\left(\begin{bmatrix}\mathrm{Re}(\boldsymbol{X})\\ \mathrm{Im}(\boldsymbol{X})\end{bmatrix}\right)$, the subdifferential of $\mathscr{F}$ is given by

$$\partial\mathscr{F}\left(\begin{bmatrix}\mathrm{Re}(\boldsymbol{X})\\ \mathrm{Im}(\boldsymbol{X})\end{bmatrix}\right) = \frac{1}{2}\mathscr{E}^*\partial\left\|\hat{\boldsymbol{\Omega}}\right\|_*, \tag{9.5}$$

where $\mathscr{E}^*$ is the adjoint of the linear operator $\mathscr{E}$.

On the one hand, the adjoint $\mathscr{E}^*$ is given by, for any $\boldsymbol{\Delta} = \begin{bmatrix}\boldsymbol{\Delta}_{11} & \boldsymbol{\Delta}_{12}\\ \boldsymbol{\Delta}_{21} & \boldsymbol{\Delta}_{22}\end{bmatrix} \in \mathbb{R}^{2M\times 2N}$ with each block in $\mathbb{R}^{M\times N}$,

$$\mathscr{E}^*\boldsymbol{\Delta} = \begin{bmatrix}\boldsymbol{\Delta}_{11} + \boldsymbol{\Delta}_{22}\\ \boldsymbol{\Delta}_{21} - \boldsymbol{\Delta}_{12}\end{bmatrix}. \tag{9.6}$$

On the other hand, since (9.3) provides an SVD of $\hat{\boldsymbol{\Omega}}$,

$$\partial\|\hat{\boldsymbol{\Omega}}\|_* = \left\{\boldsymbol{\Theta}\boldsymbol{\Xi}^T + \boldsymbol{\Delta} \mid \boldsymbol{\Theta}^T\boldsymbol{\Delta} = \boldsymbol{0},\ \boldsymbol{\Delta}\boldsymbol{\Xi} = \boldsymbol{0},\ \|\boldsymbol{\Delta}\|_2 \le 1\right\}. \tag{9.7}$$

Combining (9.5), (9.6), (9.7) and (9.2) yields the subdifferential of $\mathscr{F}(\cdot)$ at $\begin{bmatrix}\mathrm{Re}(\boldsymbol{X})\\ \mathrm{Im}(\boldsymbol{X})\end{bmatrix}$:

$$\begin{aligned}&\partial\mathscr{F}\left(\begin{bmatrix}\mathrm{Re}(\boldsymbol{X})\\ \mathrm{Im}(\boldsymbol{X})\end{bmatrix}\right)\\ &= \left\{\begin{bmatrix}\left(\boldsymbol{\Theta}_1\boldsymbol{\Xi}_1^T + \boldsymbol{\Theta}_2\boldsymbol{\Xi}_2^T + \frac{\boldsymbol{\Delta}_{11}+\boldsymbol{\Delta}_{22}}{2}\right)\\ \left(\boldsymbol{\Theta}_2\boldsymbol{\Xi}_1^T - \boldsymbol{\Theta}_1\boldsymbol{\Xi}_2^T + \frac{\boldsymbol{\Delta}_{21}-\boldsymbol{\Delta}_{12}}{2}\right)\end{bmatrix} \;\middle|\; \boldsymbol{\Delta} = \begin{bmatrix}\boldsymbol{\Delta}_{11} & \boldsymbol{\Delta}_{12}\\ \boldsymbol{\Delta}_{21} & \boldsymbol{\Delta}_{22}\end{bmatrix},\ \boldsymbol{\Theta}^T\boldsymbol{\Delta} = \boldsymbol{0},\ \boldsymbol{\Delta}\boldsymbol{\Xi} = \boldsymbol{0},\ \|\boldsymbol{\Delta}\|_2 \le 1\right\}.\end{aligned} \tag{9.8}$$

We are now ready to show (7.6).

Firstly, we show that any element in $\mathfrak{H} \equiv \left\{ \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} \;\middle|\; \boldsymbol{\alpha} + \imath\boldsymbol{\beta} \in \mathfrak{S} \right\}$ must also be in $\partial\mathscr{F}\left(\begin{bmatrix} \mathrm{Re}(\boldsymbol{X}) \\ \mathrm{Im}(\boldsymbol{X}) \end{bmatrix}\right)$, namely (9.8). In fact, for any $\boldsymbol{W} = \boldsymbol{\Delta}_1 + \imath\boldsymbol{\Delta}_2$ satisfying $\boldsymbol{U}^*\boldsymbol{W} = \boldsymbol{0}, \boldsymbol{W}\boldsymbol{V} = \boldsymbol{0}$ and $\|\boldsymbol{W}\|_2 \leq 1$, we choose $\boldsymbol{\Delta} = \begin{bmatrix} \boldsymbol{\Delta}_1 & -\boldsymbol{\Delta}_2 \\ \boldsymbol{\Delta}_2 & \boldsymbol{\Delta}_1 \end{bmatrix}$. This choice of $\boldsymbol{\Delta}$ satisfies the constraints on $\boldsymbol{\Delta}$ in (9.8). Furthermore, $\boldsymbol{U}\boldsymbol{V}^* + \boldsymbol{W} = (\boldsymbol{\Theta}_1\boldsymbol{\Xi}_1^T + \boldsymbol{\Theta}_2\boldsymbol{\Xi}_2^T + \boldsymbol{\Delta}_1) + \imath(\boldsymbol{\Theta}_2\boldsymbol{\Xi}_1^T - \boldsymbol{\Theta}_1\boldsymbol{\Xi}_2^T + \boldsymbol{\Delta}_2)$. Thus,

$$\mathfrak{H} \subseteq \partial\mathscr{F}\left(\begin{bmatrix} \mathrm{Re}(\boldsymbol{X}) \\ \mathrm{Im}(\boldsymbol{X}) \end{bmatrix}\right). \tag{9.9}$$

Secondly, we show that

$$\partial\mathscr{F}\left(\begin{bmatrix} \mathrm{Re}(\boldsymbol{X}) \\ \mathrm{Im}(\boldsymbol{X}) \end{bmatrix}\right) \subseteq \mathfrak{H}. \tag{9.10}$$

We let $\boldsymbol{\Delta} = \begin{bmatrix} \boldsymbol{\Delta}_{11} & \boldsymbol{\Delta}_{12} \\ \boldsymbol{\Delta}_{21} & \boldsymbol{\Delta}_{22} \end{bmatrix}$ be any matrix satisfying the the constraints on $\boldsymbol{\Delta}$ in (9.8). We claim that $\boldsymbol{W} \doteq \frac{\boldsymbol{\Delta}_{11}+\boldsymbol{\Delta}_{22}}{2} + \imath\frac{\boldsymbol{\Delta}_{21}-\boldsymbol{\Delta}_{12}}{2}$ satisfies $\boldsymbol{U}^*\boldsymbol{W} = \boldsymbol{0}, \boldsymbol{W}\boldsymbol{V} = \boldsymbol{0}$ and $\|\boldsymbol{W}\|_2 \leq 1$.

In fact, from $\boldsymbol{\Theta}^T \begin{bmatrix} \boldsymbol{\Delta}_{11} & \boldsymbol{\Delta}_{12} \\ \boldsymbol{\Delta}_{21} & \boldsymbol{\Delta}_{22} \end{bmatrix} = \boldsymbol{0}$, we have

$$+ \boldsymbol{\Theta}_1^T\boldsymbol{\Delta}_{11} + \boldsymbol{\Theta}_2^T\boldsymbol{\Delta}_{21} = \boldsymbol{0} \tag{9.11}$$

$$+ \boldsymbol{\Theta}_1^T\boldsymbol{\Delta}_{12} + \boldsymbol{\Theta}_2^T\boldsymbol{\Delta}_{22} = \boldsymbol{0} \tag{9.12}$$

$$- \boldsymbol{\Theta}_2^T\boldsymbol{\Delta}_{11} + \boldsymbol{\Theta}_1^T\boldsymbol{\Delta}_{21} = \boldsymbol{0} \tag{9.13}$$

$$- \boldsymbol{\Theta}_2^T\boldsymbol{\Delta}_{12} + \boldsymbol{\Theta}_1^T\boldsymbol{\Delta}_{22} = \boldsymbol{0} \tag{9.14}$$

Thus, we obtain

$$\boldsymbol{U}^*\boldsymbol{W} = (\boldsymbol{\Theta}_1^T - \imath\boldsymbol{\Theta}_2^T)(\frac{\boldsymbol{\Delta}_{11} + \boldsymbol{\Delta}_{22}}{2} + \imath\frac{\boldsymbol{\Delta}_{21} - \boldsymbol{\Delta}_{12}}{2}) \tag{9.15}$$

$$= \boldsymbol{\Theta}_1^T\frac{\boldsymbol{\Delta}_{11} + \boldsymbol{\Delta}_{22}}{2} + \boldsymbol{\Theta}_2^T\frac{\boldsymbol{\Delta}_{21} - \boldsymbol{\Delta}_{12}}{2} + \imath\left(\boldsymbol{\Theta}_1^T\frac{\boldsymbol{\Delta}_{21} - \boldsymbol{\Delta}_{12}}{2} - \boldsymbol{\Theta}_2^T\frac{\boldsymbol{\Delta}_{11} + \boldsymbol{\Delta}_{22}}{2}\right) \tag{9.16}$$

$$= \boldsymbol{0} + \imath\boldsymbol{0} = \boldsymbol{0}, \tag{9.17}$$

where the last two equalities come from adding up (9.11) and (9.14), and subtracting (9.12) from (9.13), respectively.

Similarly from $\begin{bmatrix} \boldsymbol{\Delta}_{11} & \boldsymbol{\Delta}_{12} \\ \boldsymbol{\Delta}_{21} & \boldsymbol{\Delta}_{22} \end{bmatrix} \boldsymbol{\Xi} = \boldsymbol{0}$, we can verify that

$$\boldsymbol{W}\boldsymbol{V} = \boldsymbol{0}.$$

Moreover,

$$\begin{aligned}\|\boldsymbol{W}\|_2 &= \left\|\begin{bmatrix}\frac{\boldsymbol{\Delta}_{11}+\boldsymbol{\Delta}_{22}}{2} & \frac{\boldsymbol{\Delta}_{12}-\boldsymbol{\Delta}_{21}}{2}\\ \frac{\boldsymbol{\Delta}_{21}-\boldsymbol{\Delta}_{12}}{2} & \frac{\boldsymbol{\Delta}_{11}+\boldsymbol{\Delta}_{22}}{2}\end{bmatrix}\right\|_2\\ &= \left\|\frac{1}{2}\begin{bmatrix}\boldsymbol{\Delta}_{11} & \boldsymbol{\Delta}_{12}\\ \boldsymbol{\Delta}_{21} & \boldsymbol{\Delta}_{22}\end{bmatrix}+\frac{1}{2}\begin{bmatrix}\boldsymbol{\Delta}_{22} & -\boldsymbol{\Delta}_{21}\\ -\boldsymbol{\Delta}_{12} & \boldsymbol{\Delta}_{11}\end{bmatrix}\right\|_2\\ &\le \frac{1}{2}\left\|\begin{bmatrix}\boldsymbol{\Delta}_{11} & \boldsymbol{\Delta}_{12}\\ \boldsymbol{\Delta}_{21} & \boldsymbol{\Delta}_{22}\end{bmatrix}\right\|_2+\frac{1}{2}\left\|\begin{bmatrix}\boldsymbol{\Delta}_{22} & -\boldsymbol{\Delta}_{21}\\ -\boldsymbol{\Delta}_{12} & \boldsymbol{\Delta}_{11}\end{bmatrix}\right\|_2\\ &\le \frac{1}{2}+\frac{1}{2}\\ &= 1,\end{aligned}$$

where we used the Jensen's inequality for the spectral norm, and the fact that $1 \ge \left\|\begin{bmatrix}\boldsymbol{\Delta}_{22} & -\boldsymbol{\Delta}_{21}\\ -\boldsymbol{\Delta}_{12} & \boldsymbol{\Delta}_{11}\end{bmatrix}\right\|_2 = \left\|\begin{bmatrix}\boldsymbol{\Delta}_{11} & \boldsymbol{\Delta}_{12}\\ \boldsymbol{\Delta}_{21} & \boldsymbol{\Delta}_{22}\end{bmatrix}\right\|_2$ (which comes from using the variational characterization of spectral norm). This concludes the proof of (9.10).

Combining (9.9) and (9.10), we arrive at (7.6). □