MITSUBISHI ELECTRIC RESEARCH LABORATORIES
https://www.merl.com

Recovery Analysis of Damped Spectrally Sparse Signals and Its Relation to MUSIC

Li, Shuang; Mansour, Hassan; Wakin, Michael

TR2021-008 March 11, 2021

Abstract

One of the classical approaches for estimating the frequencies and damping factors in a spectrally sparse signal is the MUltiple SIgnal Classification (MUSIC) algorithm, which exploits the low-rank structure of an autocorrelation matrix. Low rank matrices have also received considerable attention recently in the context of optimization algorithms with partial observations, and nuclear norm minimization (NNM) has been widely used as a popular heuristic of rank minimization for low-rank matrix recovery problems. On the other hand, it has been shown that NNM can be viewed as a special case of atomic norm minimization (ANM), which has achieved great success in solving line spectrum estimation problems. However, as far as we know, the general ANM (not NNM) considered in many existing works can only handle frequency estimation in undamped sinusoids. In this work, we aim to fill this gap and deal with damped spectrally sparse signal recovery problems. In particular, inspired by the dual analysis used in ANM, we offer a novel optimization-based perspective on the classical MUSIC algorithm and propose an algorithm for spectral estimation that involves searching for the peaks of the dual polynomial corresponding to a certain NNM problem, and we show that this algorithm is in fact equivalent to MUSIC itself. Building on this connection, we also extend the classical MUSIC algorithm to the missing data case. We provide exact recovery guarantees for our proposed algorithms and quantify how the sample complexity depends on the true spectral parameters. In particular, we provide a parameter-specific recovery bound for low-rank matrix recovery of jointly sparse signals rather than use certain incoherence properties as in existing literature. Simulation results also indicate that the proposed algorithms significantly outperform some relevant existing methods (e.g., ANM) in frequency estimation of damped exponentials. Spectral estimation, nuclear norm minimization, atomic norm minimization, MUSIC algorithm, low-rank matrix completion

Information and Inference: A Journal of the IMA

© 2021 MERL. This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.

Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, Cambridge, Massachusetts 02139

Recovery Analysis of Damped Spectrally Sparse Signals and Its Relation to MUSIC

Shuang Li, Hassan Mansour, and Michael B. Wakin *

August 19, 2020

Abstract

One of the classical approaches for estimating the frequencies and damping factors in a spectrally sparse signal is the MUltiple SIgnal Classification (MUSIC) algorithm, which exploits the low-rank structure of an autocorrelation matrix. Low-rank matrices have also received considerable attention recently in the context of optimization algorithms with partial observations, and nuclear norm minimization (NNM) has been widely used as a popular heuristic of rank minimization for low-rank matrix recovery problems. On the other hand, it has been shown that NNM can be viewed as a special case of atomic norm minimization (ANM), which has achieved great success in solving line spectrum estimation problems. However, as far as we know, the general ANM (not NNM) considered in many existing works can only handle frequency estimation in undamped sinusoids. In this work, we aim to fill this gap and deal with damped spectrally sparse signal recovery problems. In particular, inspired by the dual analysis used in ANM, we offer a novel optimization-based perspective on the classical MUSIC algorithm and propose an algorithm for spectral estimation that involves searching for the peaks of the dual polynomial corresponding to a certain NNM problem, and we show that this algorithm is in fact equivalent to MUSIC itself. Building on this connection, we also extend the classical MUSIC algorithm to the missing data case. We provide exact recovery guarantees for our proposed algorithms and quantify how the sample complexity depends on the true spectral parameters. In particular, we provide a parameter-specific recovery bound for low-rank matrix recovery of jointly sparse signals rather than use certain incoherence properties as in existing literature. Simulation results also indicate that the proposed algorithms significantly outperform some relevant existing methods (e.g., ANM) in frequency estimation of damped exponentials. Spectral estimation, nuclear norm minimization, atomic norm minimization, MUSIC algorithm, low-rank matrix completion.

1 Introduction

In this paper, we consider the problem of identifying the frequencies and damping factors contained in a spectrally sparse signal, namely, a superposition of a few complex sinusoids with damping, either from a complete set of uniform samples (which we refer to as full observations) or from a random set of partial uniform samples (which we refer to as the missing data case). This kind of signal arises in many applications, such as nuclear magnetic resonance spectroscopy [1, 2], radar processing [3, 4], modal analysis [5, 6], and electric motor fault detection [7]. It is well known that the frequencies and damping factors can be identified by the classical spectrum estimation approaches, such as Prony's method [8], the Matrix Pencil method [9], and the MUltiple SIgnal Classification (MUSIC) algorithm [10, 11], when full observations are available. However, in many real-world applications, obtaining such full observations with high speed uniform sampling is of high cost and technically prohibitive. Lower-rate, nonuniform sampling can be an

*SL and MBW are with the Department of Electrical Engineering, Colorado School of Mines, Golden, CO 80401. Email: {shuangli,mwakin}@mines.edu. HM is with Mitsubishi Electric Research Laboratories, Cambridge, MA 02139. Email: mansour@merl.com.

appealing alternative [12, 6, 13, 14, 15] and results in the partial observations (missing data) discussed in this work.

The MUSIC algorithm, which is widely used in signal processing [16, 17], was first proposed by Schmidt as an improvement to Pisarenko's method [10]. MUSIC exploits the low-rank structure of an autocorrelation matrix, which is divided into the noise subspace and signal subspace via an eigenvalue decomposition. The spectral parameters are then identified by searching for the zeros of a noise-space correlation function [18]. The MUSIC algorithm can be used either for spectral analysis of one signal (the single measurement vector, or SMV, problem) or for multiple measurement vector (MMV) problems involving joint sparse frequency estimation [14]. However, a limitation of these classical spectral estimation methods is that they are not compatible with the random sampling or compression protocols that can be used to reduce the front-end sampling burden. One recent work [19] does adapt the MUSIC algorithm to the setting with noisy missing data, the authors provide asymptotic theoretical guarantees on the performance of a singular value decomposition (SVD) on the noisy partially observed data matrix. In contrast, in this work we consider two settings—the (noiseless and noisy) full observation case and the noiseless missing data case—and establish non-asymptotic theoretical guarantees for our proposed algorithms.

We focus on both the SMV and MMV settings in this paper. Samples of the spectrally sparse vector-valued signal (MMV setting) considered in this work can be arranged into a low-rank matrix while samples of the spectrally sparse scalar-valued signal (SMV setting) can be used to form a Hankel matrix, which is also a low-rank matrix. Low-rank matrices have received considerable attention recently in the context of optimization algorithms with partial observations. In particular, low-rank matrix recovery from missing data appears in many practical problems such as matrix completion [20, 21], low-rank approximation [22, 23], system identification [24, 25], and image denoising [26, 27]. A common approach for recovering a low-rank matrix is known as rank minimization. However, rank minimization problems are, in general, NP-hard. Fortunately, a popular heuristic of rank minimization problems, nuclear norm minimization (NNM), performs very well in low-rank matrix recovery when certain conditions on the measurement system are satisfied [20]. Recently, it has been shown that NNM for low-rank matrix recovery can be viewed as a special case of atomic norm minimization (ANM) when the atoms are composed of rank one matrices [28, 29]. ANM is a general optimization framework for decomposing structured signals and matrices into sparse combinations of continuously-parameterized atoms from some dictionary, and one of the primary successes of ANM has been in solving the line spectrum estimation problem in both the complete and missing data cases. Most of the theory for ANM in line spectrum estimation has relied on insight gained from analyzing the dual solution to the ANM problem. However, as far as we know, the general ANM (not NNM) formulation considered in many existing works can only handle frequency estimation in undamped sinusoids [28, 30, 14, 13, 31]. This is due to the existence of an SDP form for ANM when there is no damping contained in the signals. In this work, we aim to fill this gap and identify both the frequencies and damping factors contained in a spectrally sparse signal.

The fact that NNM is a special case of ANM suggests that ANM-type dual analysis can *also* be used for NNM. In particular, in this paper, we propose an algorithm for spectral estimation that involves searching for the peaks of the dual polynomial corresponding to the NNM problem. We name this algorithm NN-MUSIC (nuclear norm minimization view of MUSIC), and we highlight the fact that in the full observation case, NN-MUSIC is in fact *equivalent* to MUSIC itself. We also provide one such development in this paper: unlike classical MUSIC, the NN-MUSIC algorithm can be naturally generalized to the missing data case, and so we also propose and analyze such a Missing Data MUSIC (MD-MUSIC) algorithm in this paper. MD-MUSIC is not equivalent to first using NNM to complete the missing data and second running conventional MUSIC on the full data matrix; rather, it involves extracting frequency estimates directly from the NNM dual polynomial, and we demonstrate that it can succeed even when "two-step" algorithms fail. Both NN-MUSIC and MD-MUSIC can deal with damped sinusoids. Our simulations also illustrate the advantage of these two proposed algorithms over ANM in frequency estimation of damped sinusoids.

Using our analytical framework, we also provide exact recovery guarantees for both NN-MUSIC and MD-MUSIC. For NN-MUSIC, our theorem indicates that we can perfectly identify the spectral parameters by searching for the locations in the damping-frequency plane where the ℓ_2-norm of the dual polynomial achieves 1, as long as the true spectral parameters are distinct from each other and the number of uniform samples is larger than the number of spectral parameters. For MD-MUSIC, our theory shows that we can perfectly identify the spectral parameters with high probability by searching for the locations in the damping-frequency plane where the ℓ_2-norm of the dual polynomial achieves 1 if the number of random samples is sufficiently large, the true spectral parameters are distinct from each other, and the number of uniform samples (from which the random samples are drawn) is larger than the number of spectral parameters. It is even possible to get perfect parameter recovery without exact data recovery. Moreover, we provide a parameter-specific recovery bound for low-rank matrix recovery of jointly sparse signals, that is, we quantify how the sample complexity depends on the true spectral parameters rather than use certain incoherence properties as in existing literature.

The remainder of this paper is organized as follows. In Section 2, we introduce both the SMV and MMV settings considered in this paper. In Section 3, we review the classical MUSIC algorithm as well as its variants. In Section 4, we offer a novel optimization-based perspective on the MUSIC algorithm by highlighting the fact that the proposed NN-MUSIC algorithm is equivalent to MUSIC in the full observation case. We also generalize it to the missing data case and propose the MD-MUSIC algorithm to support the idea that this connection between NNM and MUSIC could lead to future developments and understanding. The proofs for theoretical guarantees are presented in Section 5. In Section 6, we explore the recovery performance of the proposed NN-MUSIC and MD-MUSIC algorithms with numerical simulations. Finally, we conclude this work and discuss future directions in Section 7.

2 Signal Models

We are interested in identifying the frequencies and damping factors contained in a spectrally sparse signal, which can be a scalar-valued signal in the SMV setting or a vector-valued signal in the MMV setting. We first introduce the SMV and MMV settings that are considered in this work. Throughout this work, we use superscript "$\circ$" to denote row vectors, and superscripts "$\top$" and "H" to denote transpose and conjugate transpose, respectively.

2.1 Single Measurement Vector (SMV) setting

In the SMV setting, a scalar-valued, continuous-time signal is assumed to have the form

$$y(t) = x(t) + e(t), \quad x(t) = \sum_{k=1}^{K} c_k r_k^t e^{j2\pi f_k t}, \tag{1}$$

where $\{c_k\}$, $\{r_k\}$, $\{f_k\}$ and $e(t)$ are the unknown complex coefficients, damping ratios, frequency parameters, and additive observation noise, respectively. Such signals appear in many applications, such as radar, sonar, and communications. Without loss of generality, we assume the frequencies $\{f_k\}$ belong to the interval $[0, 1)$, the damping ratios $\{r_k\}$ belong to the interval $[0, 1]$, the complex coefficients $c_k \neq 0$, and $e(t) \sim \mathcal{CN}(0, \sigma^2)$.

2.2 Multiple Measurement Vector (MMV) setting

In the MMV setting, we consider a vector-valued signal $\boldsymbol{y}^\circ(t) \in \mathbb{C}^{1\times N}$, which is a superposition of K damped sinusoids with additive observation noise $\boldsymbol{e}^\circ(t) \in \mathbb{C}^{1\times N}$. More precisely,

$$\boldsymbol{y}^\circ(t) = \boldsymbol{x}^\circ(t) + \boldsymbol{e}^\circ(t), \quad \boldsymbol{x}^\circ(t) = \sum_{k=1}^{K} c_k r_k^t e^{j2\pi f_k t} \boldsymbol{\phi}_k^\top, \tag{2}$$

with $c_k \neq 0$, $f_k \in [0, 1)$ and $r_k \in [0, 1]$ being the k-th complex coefficient, frequency, and damping factor, respectively. Here, each $\boldsymbol{\phi}_k \in \mathbb{C}^N$ is a normalized vector ($\|\boldsymbol{\phi}_k\|_2 = 1$) that can be viewed as the mode shape in modal analysis problems [5, 6].

Suppose we take M uniform samples and arrange $\boldsymbol{y}_m^\circ = \boldsymbol{y}^\circ(m)$ as the m-th row of a data matrix $\mathbf{Y} \in \mathbb{C}^{M\times N}$. Define $\mathbf{X}^\star \triangleq \left[(\boldsymbol{x}_0^\circ)^\top \; (\boldsymbol{x}_1^\circ)^\top \; \cdots \; (\boldsymbol{x}_{M-1}^\circ)^\top\right]^\top$ and $\mathbf{E} \triangleq \left[(\boldsymbol{e}_0^\circ)^\top \; (\boldsymbol{e}_1^\circ)^\top \; \cdots \; (\boldsymbol{e}_{M-1}^\circ)^\top\right]^\top$ as the noiseless data matrix and observation noise matrix, respectively. Then, we have

$$\mathbf{Y} = \left[(\boldsymbol{y}_0^\circ)^\top \; (\boldsymbol{y}_1^\circ)^\top \; \cdots \; (\boldsymbol{y}_{M-1}^\circ)^\top\right]^\top = \sum_{k=1}^{K} \widetilde{c}_k \boldsymbol{a}(r_k, f_k)\boldsymbol{\phi}_k^\top + \mathbf{E} = \mathbf{A}_{rf}\mathbf{D}_{\widetilde{c}}\mathbf{\Phi}^\top + \mathbf{E} = \mathbf{X}^\star + \mathbf{E} \tag{3}$$

with[1]

$$\boldsymbol{a}(r,f) \triangleq \begin{cases} \sqrt{\frac{1-r^2}{1-r^{2M}}}\left[1 \; re^{j2\pi f 1} \cdots r^{M-1}e^{j2\pi f(M-1)}\right]^\top, & r<1, \\ \frac{1}{\sqrt{M}}\left[1 \; e^{j2\pi f 1} \; \cdots \; e^{j2\pi f(M-1)}\right]^\top, & r=1, \end{cases} \tag{4}$$

and

$$\widetilde{c}_k \triangleq \begin{cases} c_k\sqrt{\frac{1-r_k^{2M}}{1-r_k^2}}, & r_k < 1, \\ c_k\sqrt{M}, & r_k = 1, \end{cases} \qquad k = 1, \ldots, K.$$

In addition, we define $\mathbf{A}_{rf} \triangleq [\boldsymbol{a}(r_1, f_1), \cdots, \boldsymbol{a}(r_K, f_K)]$, $\mathbf{D}_{\widetilde{c}} \triangleq \mathrm{diag}([\widetilde{c}_1, \cdots, \widetilde{c}_K])$, and $\mathbf{\Phi} \triangleq [\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_K]$.

Let $\boldsymbol{y}_n$, $\boldsymbol{x}_n$ and $\boldsymbol{e}_n$ denote the n-th column of $\mathbf{Y}$, $\mathbf{X}^\star$ and $\mathbf{E}$, respectively. It can be seen that

$$\boldsymbol{y}_n = \boldsymbol{x}_n + \boldsymbol{e}_n = \sum_{k=1}^{K} \check{c}_{n,k}\boldsymbol{a}(r_k, f_k) + \boldsymbol{e}_n, \; n = 1, \ldots, N, \tag{5}$$

where $\check{c}_{n,k} = \widetilde{c}_k\phi_{n,k}$ with $\phi_{n,k}$ being the (n,k)-th entry of $\mathbf{\Phi}$. In this model, the observed data consists of N observed length-M signals, each comprised of K damped sinusoids. The N signals share the same set of unknown frequencies and damping factors, but each has a unique set of coefficients.

3 Prior Work

In this section, we review the classical MUSIC algorithm [10, 11] as well as its two variants, Damped MUSIC (DMUSIC) [32] and MUSIC adapted to missing data with Gaussian white noise (denoted as MN-MUSIC) [19].

3.1 MUltiple SIgnal Classification (MUSIC) algorithm

3.1.1 SMV MUSIC via autocorrelation matrix

By sampling the scalar-valued, continuous-time signal $y(t)$, defined in (1), at M equally spaced times, one can define a vector $\boldsymbol{y}(t) \in \mathbb{C}^M$ as

$$\boldsymbol{y}(t) \triangleq [y(t) \; y(t+1) \; \cdots \; y(t+M-1)]^\top, \tag{6}$$

which has the autocorrelation matrix $\mathbf{R}_y \triangleq \mathbb{E}\{\boldsymbol{y}(t)\boldsymbol{y}(t)^H\}$.

The classical MUSIC algorithm aims to identify the unknown frequencies $\{f_k\}$ by constructing (and then decomposing) an estimate of the autocorrelation matrix $\mathbf{R}_y$ without damping, namely, in the case where all $r_k = 1$ [33]. This requires $M > K$. Specifically, consider a full set of uniform observations $\{y(t)\}$ with $t = 0, 1, \ldots, L-1$, for some $L > M$. Then, the following sample autocorrelation matrix can be used to approximate $\mathbf{R}_y$:

$$\widehat{\mathbf{R}}_y = \frac{1}{L-M+1}\sum_{t=0}^{L-M} \boldsymbol{y}(t)\boldsymbol{y}(t)^H. \tag{7}$$

[1]Note that we abbreviate $\boldsymbol{a}(r, f)$ to $\boldsymbol{a}(f)$ when $r = 1$.

Let $[\widehat{\mathbf{U}}_s\ \widehat{\mathbf{U}}_n]$ denote the orthonormal eigenvectors of $\widehat{\mathbf{R}}_y$. In particular, suppose $\widehat{\mathbf{U}}_s \in \mathbb{C}^{M\times K}$ (signal space) and $\widehat{\mathbf{U}}_n \in \mathbb{C}^{M\times(M-K)}$ (noise space) are associated with the K largest eigenvalues and the $M-K$ smallest eigenvalues of $\widehat{\mathbf{R}}_y$, respectively. Then, we summarize the classical MUSIC algorithm in Algorithm 1.

Algorithm 1 MUSIC

1: **procedure** INPUT($\{y(t)\}_{t=0}^{L-1},\ K$)
2: compute the autocorrelation matrix $\widehat{\mathbf{R}}_y$ as in (7) and its eigenvectors $\widehat{\mathbf{U}}_n$
3: compute the pseudospectrum: $1/\|\widehat{\mathbf{U}}_n^H \boldsymbol{a}(f)\|_2^2$, where $\boldsymbol{a}(f)$ is defined in (4) with $r=1$
4: localize the K largest local maxima of pseudospectrum to get $\widehat{f}_k$
5: **return** $\widehat{f}_k$
6: **end procedure**

The intuition behind the MUSIC algorithm comes from the fact that, as a consequence of the scalar-valued signal model in (1), the vector-valued signal $\boldsymbol{y}(t)$ in (6) can be expressed as

$$\boldsymbol{y}(t) = \sum_{k=1}^{K} \sqrt{M} c_k e^{j2\pi f_k t} \boldsymbol{a}(f_k) + \boldsymbol{e}(t) = \mathbf{A}_f \boldsymbol{c}(t) + \boldsymbol{e}(t)$$

with

$$\begin{aligned}
\mathbf{A}_f &= [\boldsymbol{a}(f_1),\ \boldsymbol{a}(f_2),\ \cdots, \boldsymbol{a}(f_K)],\\
\boldsymbol{c}(t) &= \sqrt{M}[c_1 e^{j2\pi f_1 t},\ c_2 e^{j2\pi f_2 t},\ \cdots, c_K e^{j2\pi f_K t}]^\top,\\
\boldsymbol{e}(t) &= [e(t),\ e(t+1),\ \cdots, e(t+M-1)]^\top,
\end{aligned}$$

where $\boldsymbol{a}(f)$ is defined in (4) with $r=1$. Then, the autocorrelation matrix becomes

$$\mathbf{R}_y = \mathbb{E}\{\boldsymbol{y}(t)\boldsymbol{y}(t)^H\} = \mathbf{A}_f \mathbf{R}_c \mathbf{A}_f^H + \sigma^2 \mathbf{I}_M$$

if $\boldsymbol{c}(t)$ is uncorrelated with $\boldsymbol{e}(t)$. Here, $\mathbf{R}_c \triangleq \mathbb{E}\{\boldsymbol{c}(t)\boldsymbol{c}(t)^H\}$ is the autocorrelation matrix of $\boldsymbol{c}(t)$ and $\mathbf{I}_M$ denotes the $M\times M$ identity matrix. Note that the coefficients $\{c_k\}$ may be uncorrelated ($\mathbf{R}_c$ is diagonal) or may contain completely correlated pairs ($\mathbf{R}_c$ is singular). We are interested in the first case, namely, $\mathbf{R}_c$ is diagonal and positive definite since $c_k \neq 0$.[2] On the other hand, the rank of $\mathbf{A}_f$ is K when all the frequencies $\{f_k\}$ are distinct and $M \geq K$. It follows that the rank of $\mathbf{A}_f \mathbf{R}_c \mathbf{A}_f^H$ is K. Let $\{\lambda_m\},\ m=1,\ldots,M$ denote the non-increasing eigenvalues of $\mathbf{A}_f \mathbf{R}_c \mathbf{A}_f^H$. Then, we have $\lambda_{K+1} = \cdots = \lambda_M = 0$. As a consequence, the determinant of $\mathbf{A}_f \mathbf{R}_c \mathbf{A}_f^H$ is $\det(\mathbf{A}_f \mathbf{R}_c \mathbf{A}_f^H) = \det(\mathbf{R}_y - \sigma^2 \mathbf{I}_M) = 0$, which implies that $\lambda_m^y = \sigma^2,\ m = K+1,\ldots,M$, where λ_m^y is the m-th non-increasing eigenvalue of $\mathbf{R}_y$. Denoting $\boldsymbol{u}_m$ as the m-th eigenvector of $\mathbf{R}_y$ corresponding to eigenvalue λ_m^y, we have

$$\mathbf{R}_y \boldsymbol{u}_m = \lambda_m^y \boldsymbol{u}_m,\ m = 1,\ldots,M. \tag{8}$$

Replacing $\mathbf{R}_y = \mathbf{A}_f \mathbf{R}_c \mathbf{A}_f^H + \sigma^2 \mathbf{I}_M$ into the above equation (8), we have

$$\mathbf{A}_f \mathbf{R}_c \mathbf{A}_f^H \boldsymbol{u}_m = (\lambda_m^y - \sigma^2)\boldsymbol{u}_m = \mathbf{0} \ \text{ or } \ \mathbf{A}_f^H \boldsymbol{u}_m = \mathbf{0}$$

when $\lambda_m^y = \sigma^2$, or equivalently, $m = K+1,\ldots,M$. Then, $\boldsymbol{a}(f)$, which is defined in (4), is orthogonal to $\boldsymbol{u}_m,\ m = K+1,\ldots,M$ (columns of $\widehat{\mathbf{U}}_n$), when $f = f_k,\ k = 1,\ldots,K$. Therefore, we can identify the frequencies by localizing the K peaks of the pseudospectrum $1/\|\widehat{\mathbf{U}}_n^H \boldsymbol{a}(f)\|_2^2$.

[2] As is stated in [11], in general, $\mathbf{R}_c$ will be "merely" positive definite to reflect the arbitrary degrees of pair-wise correlations occurring between the coefficients.

3.1.2 SMV MUSIC via Hankel matrix

As an alternative to the above autocorrelation matrix, a certain Hankel matrix can also be used in the MUSIC algorithm [18].[3] In particular, from the same full set of uniform observations $\{y(t)\}$ with $t = 0, 1, \ldots, L-1$, one can formulate the Hankel matrix

$$\mathcal{H}_y = \begin{bmatrix} y(0) & y(1) & \cdots & y(N-1) \\ y(1) & y(2) & \cdots & y(N) \\ \vdots & \vdots & & \vdots \\ y(M-1) & y(M) & \cdots & y(L-1) \end{bmatrix} \tag{9}$$

for some positive integers M and N satisfying $M + N = L + 1$. Then define the noise-space correlation function $R(f)$ and imaging function $J(f)$ as

$$R(f) = \|\mathbf{U}_n^H \boldsymbol{a}(f)\|_2, \quad J(f) = \frac{1}{\|\mathbf{U}_n^H \boldsymbol{a}(f)\|_2}$$

with $\boldsymbol{a}(f) = \frac{1}{\sqrt{M}}[1 \; e^{j2\pi f 1} \; \cdots \; e^{j2\pi f(M-1)}]^\top$ as defined in (4). Here, $\mathbf{U}_n$ spans the noise subspace and contains the left singular vectors of $\mathcal{H}_y$ corresponding to the $M - K$ smallest singular values. The frequencies can then be estimated by identifying the K local minima of the noise-space correlation function $R(f)$ or the K local maxima of the imaging function $J(f)$.

Note that the sample autocorrelation matrix in (7) and the Hankel matrix in (9) are related by $\widehat{\mathbf{R}}_y = \frac{1}{L-M+1}\mathcal{H}_y\mathcal{H}_y^H$. Thus, the eigenvectors of $\widehat{\mathbf{R}}_y$ are the same as the left singular vectors of $\mathcal{H}_y$ up to a unitary transform. Therefore, the MUSIC algorithm based on the autocorrelation matrix and the Hankel matrix are equivalent since the imaging function $J(f)$ is equivalent to the pseudospectrum in Algorithm 1.

3.1.3 MMV MUSIC via data matrix

The MUSIC algorithm is also widely used in MMV problems [38, 39, 40]. Given a multiple measurement matrix $\mathbf{Y} = [\boldsymbol{y}_1, \cdots, \boldsymbol{y}_N]$ (see Section 3.3), one can directly compute an SVD of $\mathbf{Y}$ to obtain the noise space $\mathbf{U}_n$ from the left singular vectors of $\mathbf{Y}$ and then identify the frequency parameters by localizing the peaks of the imaging function. In particular, denote $\mathbf{Y} = [\mathbf{U}_s \; \mathbf{U}_n] \; \boldsymbol{\Sigma} \; [\mathbf{V}_s \; \mathbf{V}_n]^H$ as an SVD of the data matrix $\mathbf{Y}$. For the same reason, one can estimate the frequencies by finding the peaks of the imaging function $J(f) = \frac{1}{\|\mathbf{U}_n^H \boldsymbol{a}(f)\|_2}$.

3.2 Damped MUSIC (DMUSIC)

In the general model of (1), the complex-valued sinusoids are damped and decay over time. For this more general case, the DMUSIC algorithm introduced in [32] aims to estimate *both* the frequencies $\{f_k\}$ and damping ratios $\{r_k\}$ directly using the rank-deficiency and Hankel properties of (9). Similar to classical MUSIC, DMUSIC involves constructing the noise subspace matrix $\mathbf{U}_n$ by computing an SVD of the Hankel matrix $\mathcal{H}_y$. Then, the (r_k, f_k) pairs are identified by finding the peaks of the imaging function

$$J(r, f) = \frac{1}{\|\mathbf{U}_n^H \boldsymbol{a}(r, f)\|_2} \tag{10}$$

with $\boldsymbol{a}(r, f)$ defined in (4).

The intuition behind DMUSIC is that the Hankel matrix in (9) can be rewritten as

$$\mathcal{H}_y = \mathbf{A}_{rf}\mathbf{D}_c(\mathbf{A}_{rf}^N)^\top + \mathcal{H}_e,$$

[3] Indeed, Hankel structure has been widely used in a variety of algorithms for spectral estimation in the literature [34, 35, 36, 37].

where $\mathcal{H}_e$ is a Hankel matrix formulated with $\{e(t)\}$, $t = 0, \ldots, L-1$, and $\mathbf{D}_c$ is a diagonal matrix with diagonal entries being the scaled coefficients c_k. Precisely, the k-th diagonal entry of $\mathbf{D}_c$ is $\frac{c_k\sqrt{(1-r_k^{2M})(1-r_k^{2N})}}{(1-r_k^2)}$. $\mathbf{A}_{rf}$ and $\mathbf{A}_{rf}^N$ are Vandermonde matrices defined as

$$\mathbf{A}_{rf} \triangleq [\boldsymbol{a}(r_1, f_1), \cdots, \boldsymbol{a}(r_K, f_K)], \quad \mathbf{A}_{rf}^N \triangleq [\boldsymbol{a}_N(r_1, f_1), \cdots, \boldsymbol{a}_N(r_K, f_K)],$$

with

$$\boldsymbol{a}(r,f) \triangleq \sqrt{\frac{1-r_k^2}{1-r_k^{2M}}}[1 \; re^{j2\pi f 1} \cdots r^{M-1}e^{j2\pi f(M-1)}]^\top,$$
$$\boldsymbol{a}_N(r,f) \triangleq \sqrt{\frac{1-r_k^2}{1-r_k^{2N}}}[1 \; re^{j2\pi f 1} \cdots r^{N-1}e^{j2\pi f(N-1)}]^\top. \quad (11)$$

Note that we add a subscript "N" in (11) to distinguish $\boldsymbol{a}_N(r,f) \in \mathbb{C}^N$ from $\boldsymbol{a}(r,f) \in \mathbb{C}^M$. When $M, N \geq K$ and all the (r_k, f_k) pairs are distinct, $\mathbf{A}_{rf}$ and $\mathbf{A}_{rf}^N$ are full rank. Then, $\mathcal{H}_x \triangleq \mathbf{A}_{rf}\mathbf{D}_c(\mathbf{A}_{rf}^N)^\top$ is of rank K. Now, consider the case when there is no noise, i.e., $\mathcal{H}_y = \mathcal{H}_x$. Denote an SVD of $\mathcal{H}_y$ as

$$\mathcal{H}_y = [\mathbf{U}_s \; \mathbf{U}_n]\boldsymbol{\Sigma}\begin{bmatrix} \mathbf{V}_s^H \\ \mathbf{V}_n^H \end{bmatrix}.$$

One can show that the range spaces of $\mathcal{H}_y$, $\mathbf{A}_{rf}$, and $\mathbf{U}_s$ are all equal when there is no noise. Then, $\boldsymbol{a}(r,f)$ is orthogonal to the columns of $\mathbf{U}_n$ when $(r,f) = (r_k, f_k)$, $k = 1, \ldots, K$. If noise exists, the orthogonal relationship between $\boldsymbol{a}(r,f)$ and $\mathbf{U}_n$ no longer holds. However, one can identify all the (r_k, f_k) pairs by finding the peaks of the imaging function defined in (10), that is, searching for $\boldsymbol{a}(r,f)$ that are most nearly orthogonal to the noise space $\mathbf{U}_n$.

3.3 MN-MUSIC for missing and noisy data

The classical MUSIC algorithm has also been adapted to the missing data case with Gaussian white noise (denoted as MN-MUSIC) for applications such as direction of arrival (DOA) estimation [19]. The authors consider the MMV setting as introduced in Section 3.1.3. More precisely, consider an observed $M \times N$ matrix $\mathbf{Y} = [\boldsymbol{y}_1, \cdots, \boldsymbol{y}_N]$, where $\boldsymbol{y}_n$ is defined in (5) and repeated as follows

$$\boldsymbol{y}_n = \sum_{k=1}^{K} \breve{c}_{n,k}\boldsymbol{a}(f_k) + \boldsymbol{e}_n, \quad n = 1, \ldots, N,$$

with $r = 1$ since undamped signals are considered in [19].

Assume we partially observe the entries of $\mathbf{Y}$ with i.i.d. Bernoulli randomly sampled locations $\Omega \subset \{1, \ldots, M\} \times \{1, \ldots, N\}$. Let $\mathbf{Y}_\Omega$ be the projection matrix of $\mathbf{Y}$ on the index set Ω, i.e

$$\mathbf{Y}_{\Omega ij} = \begin{cases} \mathbf{Y}_{ij}, & (i,j) \in \Omega, \\ 0, & \textit{else}. \end{cases}$$

Then in MN-MUSIC, an SVD is directly performed on $\mathbf{Y}_\Omega$ to get the signal space matrix $\mathbf{U}_s$, which contains the left singular vectors of $\mathbf{Y}_\Omega$ corresponding to the K largest singular values. Finally, the frequencies are estimated by finding the peaks of $\|\mathbf{U}_s^H\boldsymbol{a}(f)\|_2^2$, which is essentially same as in Sections 3.1 and 3.2.

4 Main Results

In this section we outline a connection between MUSIC and low-rank matrix optimization using nuclear norm minimization (NNM), and based on this connection we propose an extension of MUSIC that is appropriate for the missing data case. Our interest in NNM here is specifically due to its connection with MUSIC. There are, of course, alternative low-rank optimization problems that can also be used for spectral analysis. Among these, atomic norm minimization (ANM) has been proposed and analyzed for solving the undamped line spectrum estimation problem in both the full and missing data cases [14, 30]. Moreover, a low-rank Hankel matrix recovery problem has recently been considered for damped spectral analysis [41]; that work involves solving the NNM (12) and (16) with an extra Hankel constraint on $\mathbf{X}$. While these alternative frameworks have some benefits, we believe that our work sheds light on a more fundamental problem, given the considerable attention that MUSIC has received over the last several decades. This understanding may lead to new developments for MUSIC and other optimization algorithms for spectral analysis in the future.

4.1 Optimization connection to MUSIC in the full data case

In this section, we consider both the SMV and MMV settings. Given a set of uniform samples from the signal model (1) in the SMV setting and the data matrix $\mathbf{X}^\star = \mathbf{A}_{rf}\mathbf{D}_{\widetilde{c}}\mathbf{\Phi}^\top$ or its noisy version $\mathbf{Y}$ (3) in the MMV setting, our goal is to identify the frequencies $\{f_k\}$ and damping factors $\{r_k\}$. Note that in the SMV setting, we can construct a Hankel matrix as in (9). As is shown in Section 3.2, this Hankel matrix $\mathcal{H}_x$ can be decomposed as $\mathcal{H}_x \triangleq \mathbf{A}_{rf}\mathbf{D}_c(\mathbf{A}_{rf}^N)^\top$ and is of rank K when there is no noise. One can observe that both $\mathbf{X}^\star$ and $\mathcal{H}_x$ are low-rank matrices and have the same type of decompositions. Therefore, the analysis on $\mathbf{X}^\star$ can also be applied to $\mathcal{H}_x$, which implies that the algorithms we build using $\mathbf{X}^\star$ in the MMV scenario also work for the SMV scenario.

Assume that $\mathbf{X}^\star$ is given and $K \ll M, N$, note that $\mathbf{X}^\star$ in (3) is low rank. Inspired by the low-rank property of $\mathbf{X}^\star$ and the dual analysis that is commonly used in atomic norm minimization (ANM) [28, 30], let us consider the following nuclear norm minimization (NNM)

$$\widehat{\mathbf{X}} = \arg\min_{\mathbf{X}} \ \|\mathbf{X}\|_* \quad \text{s.t.} \ \mathbf{X} = \mathbf{X}^\star. \tag{12}$$

Although this problem has a trivial solution (namely, $\widehat{\mathbf{X}} = \mathbf{X}^\star$), it is interesting because we can compute the corresponding dual feasible point $\mathbf{Q}$, which is a solution of the dual problem, via the Lagrange function of (12) and thus identify the frequencies and damping factors that are contained in the spectrally sparse signal $\boldsymbol{x}^\circ(t)$ in (2). In particular, the Lagrange function is given as

$$\mathcal{L}(\mathbf{X},\mathbf{Q}) = \|\mathbf{X}\|_* + \langle \mathbf{X}^\star - \mathbf{X}, \mathbf{Q}\rangle_{\mathbb{R}} = \|\mathbf{X}\|_* - \langle \mathbf{X}, \mathbf{Q}\rangle_{\mathbb{R}} + \langle \mathbf{X}^\star, \mathbf{Q}\rangle_{\mathbb{R}},$$

with $\mathbf{Q}$ being the dual variable. $\langle\cdot,\cdot\rangle_{\mathbb{R}}$ is defined as the real inner product, i.e.,

$$\langle \mathbf{X}^\star, \mathbf{Q}\rangle_{\mathbb{R}} = \mathrm{Re}(\langle \mathbf{X}^\star, \mathbf{Q}\rangle) = \mathrm{Re}(\mathrm{Tr}(\mathbf{Q}^H\mathbf{X}^\star))$$

with $\mathrm{Tr}(\cdot)$ denoting the trace of a matrix. Then, the subgradient of $\mathcal{L}(\mathbf{X},\mathbf{Q})$ with respect to $\mathbf{X}$ is

$$\frac{\partial \mathcal{L}(\mathbf{X},\mathbf{Q})}{\partial \mathbf{X}} = \partial\|\mathbf{X}\|_* - \mathbf{Q},$$

where $\partial\|\mathbf{X}\|_*$ is the subdifferential of the nuclear norm and given as

$$\partial\|\mathbf{X}\|_* = \partial\|\mathbf{X}^\star\|_* = \{\mathbf{Z}: \ \mathbf{Z} = \mathbf{U}_{X^\star}\mathbf{V}_{X^\star}^H + \mathbf{W}, \ \mathbf{U}_{X^\star}^H\mathbf{W} = \mathbf{0}, \ \mathbf{W}\mathbf{V}_{X^\star} = \mathbf{0}, \ \|\mathbf{W}\| \le 1\}$$

since $\widehat{\mathbf{X}} = \mathbf{X} = \mathbf{X}^\star$. Here, we use $\|\mathbf{W}\|$ to denote the spectral norm of the matrix $\mathbf{W}$. Note that $\mathbf{X}^\star = \mathbf{U}_{X^\star}\mathbf{S}_{X^\star}\mathbf{V}_{X^\star}^H$ is a truncated SVD of $\mathbf{X}^\star$ with $\mathbf{U}_{X^\star} \in \mathbb{C}^{M\times K}$, $\mathbf{S}_{X^\star} \in \mathbb{R}^{K\times K}$ and $\mathbf{V}_{X^\star} \in \mathbb{C}^{N\times K}$. We can also construct a $\mathbf{Q} \in \partial\|\mathbf{X}\|_*$ by letting $\mathbf{0} \in \frac{\partial\mathcal{L}(\mathbf{Q},\mathbf{X})}{\partial\mathbf{X}}$ according to the zero-gradient condition in the Karush-Kuhn-Tucker (KKT) conditions [?]. Finally, we have

$$\mathbf{Q} = \mathbf{U}_{X^\star}\mathbf{V}_{X^\star}^H + \mathbf{W} \tag{13}$$

with $\mathbf{U}_{X^\star}^H\mathbf{W} = \mathbf{0}$, $\mathbf{W}\mathbf{V}_{X^\star} = \mathbf{0}$, $\|\mathbf{W}\| \le 1$. Note that the dual solution to (12) is not unique. In particular, one can verify that any $\mathbf{Q} \in \partial\|\mathbf{X}^\star\|_*$ is a dual solution since $\langle \mathbf{X}^\star, \mathbf{Q}\rangle_{\mathbb{R}} = \|\mathbf{X}^\star\|_*$.[4]

Given a dual feasible point $\mathbf{Q} = \mathbf{U}_{X^\star}\mathbf{V}_{X^\star}^H + \mathbf{W}$, we define the dual polynomial as

$$\mathcal{Q}(r,f) \triangleq \mathbf{Q}^H \boldsymbol{a}(r,f), \tag{14}$$

which is inspired by the dual analysis in ANM. The following theorem guarantees that we can identify the true r_k's and f_k's by localizing the places where $\|\mathcal{Q}(r,f)\|_2$ achieves 1. Moreover, it also indicates that one does not need a separation condition in this full data noiseless setting. (In some previous work on optimization-based spectral estimation [14], one needs the minimum separation Δ_f, which is defined in Corollary 4.1, to be on the order of $\frac{1}{M}$ even for the full data noiseless setting.)

Theorem 4.1. *Let $\mathcal{RF}$ denote the set of the true damping factor and frequency pairs, i.e.,*

$$\mathcal{RF} = \{(r_1,f_1),\cdots,(r_K,f_K)\}.$$

Given the full data matrix $\mathbf{X}^\star$ as in (3), compute its truncated SVD $\mathbf{X}^\star = \mathbf{U}_{X^\star}\mathbf{S}_{X^\star}\mathbf{V}_{X^\star}^H$. For all $\mathbf{Q} \in \partial\|\mathbf{X}^\star\|_*$ with $\|\mathbf{W}\| < 1$, the dual polynomial defined in (14) satisfies

$$\|\mathcal{Q}(r_k,f_k)\|_2 = 1,\ \forall\ (r_k,f_k) \in \mathcal{RF},$$
$$\|\mathcal{Q}(r,f)\|_2 < 1,\ \forall\ (r,f) \notin \mathcal{RF},$$

if $M \ge K+1$, all the (r_k,f_k) pairs in $\mathcal{RF}$ are distinct, and $\mathbf{\Phi} \in \mathbb{C}^{N\times K}$ is of rank K.

The proof of Theorem 4.1 is given in Section 5.1. Note that for the case when $\|\mathbf{W}\| = 1$, one may get $\|\mathcal{Q}(r,f)\|_2 = 1$ for some $(r,f) \notin \mathcal{RF}$, i.e., having some false peaks when checking the norm of the dual polynomial. To remove these false peaks, one can solve the following least squares problem. Denote $\{(\widehat{r}_k,\widehat{f}_k)\}_{k=1}^{\widehat{K}}$ as the damping factor and frequency pairs estimated by localizing the places where $\|\mathcal{Q}(r,f)\|_2 = 1$. (To localize the places where $\|\mathcal{Q}(r,f)\|_2 = 1$, one can use the *findpeaks* function in Matlab.) Define $\widehat{\mathbf{A}}_{rf} \triangleq [\boldsymbol{a}(\widehat{r}_1,\widehat{f}_1),\cdots,\boldsymbol{a}(\widehat{r}_{\widehat{K}},\widehat{f}_{\widehat{K}})]$. Then the least squares problem

$$\min_{\mathbf{\Phi}_c} \|\mathbf{X}^\star - \widehat{\mathbf{A}}_{rf}\mathbf{\Phi}_c\|_F$$

can be used to remove the false peaks. The $(\widehat{r}_k,\widehat{f}_k)$ pairs corresponding to zero rows of $\mathbf{\Phi}_c$ can be viewed as false estimates.

Based on the above analysis, we propose the following algorithm, named NN-MUSIC (nuclear norm minimization view of MUSIC algorithm), to estimate the damping factors $\{r_k\}$ and frequencies $\{f_k\}$ of the damped sinusoids from the data matrix $\mathbf{X}^\star$. Note that the step with the highest computational cost is the SVD step, and this needs to be performed only once.

Algorithm 2 NN-MUSIC

procedure INPUT($\mathbf{X}^\star \in \mathbb{C}^{M\times N}$)
  compute truncated SVD of $\mathbf{X}^\star$: $\mathbf{X}^\star = \mathbf{U}_{X^\star}\mathbf{S}_{X^\star}\mathbf{V}_{X^\star}^H$
  form the dual feasible point: $\mathbf{Q} = \mathbf{U}_{X^\star}\mathbf{V}_{X^\star}^H$
  form the dual polynomial: $\mathcal{Q}(r,f) = \mathbf{Q}^H\boldsymbol{a}(r,f)$
  localize the places where $\|\mathcal{Q}(r,f)\|_2 = 1$ to get $(\widehat{r}_k,\widehat{f}_k)$
  return $(\widehat{r}_k,\widehat{f}_k)$
end procedure

[4]When using CVX with the default solver SDPT3 to solve the SDP form of the NNM problem (12), we observe that it returns both the primal solution and a minimum norm dual solution, namely, $\mathbf{Q} = \mathbf{U}_{X^\star}\mathbf{V}_{X^\star}^H$, due to the use of a Conjugate Gradient based algorithm [?, ?].

Note that Algorithm 2 is essentially equivalent to the MUSIC (in the undamped case) and DMUSIC (in the damped case) algorithms outlined in Section 3. This is due to the fact that $\|\mathcal{Q}(r,f)\|_2 = \|\mathbf{U}_{X^\star}^H \boldsymbol{a}(r,f)\|_2$.[5] When there is no noise, the DMUSIC algorithm and its variants characterize the spectral parameters by locating the zeros of a noise-space correlation function or the peaks of the imaging function, and the proposed NN-MUSIC algorithm identifies the spectral parameters by localizing the (r,f) pairs where $\|\mathcal{Q}(r,f)\|_2$ achieves 1. While MUSIC has been classically understood from an algebraic perspective (owing to its closed form), we believe the derivation of NN-MUSIC offers a novel optimization-based perspective on MUSIC that could lead to future developments and understanding.

We also stress that this connection to MUSIC is unique to NNM and does not apply in general to ANM. In particular, the connection arises specifically because the dual feasible point $\mathbf{Q} = \mathbf{U}_{X^\star}\mathbf{V}_{X^\star}^H$ of NNM induces a dual polynomial that satisfies $\|\mathcal{Q}(r,f)\|_2 = \|\mathbf{U}_{X^\star}^H \boldsymbol{a}(r,f)\|_2$. On the other hand, the dual feasible point of ANM formulations does not admit the structure $\mathbf{Q} = \mathbf{U}_{X^\star}\mathbf{V}_{X^\star}^H$, in general.

Finally, consider the case when the given data matrix $\mathbf{Y}$ contains some additive white Gaussian noise, i.e., $\mathbf{Y} = \mathbf{X}^\star + \mathbf{E}$ with $\mathbf{E}$ denoting the measurement noise. Then, we can solve the following nuclear norm denoising program

$$\min_{\mathbf{X}} \ \frac{1}{2}\|\mathbf{Y} - \mathbf{X}\|_F^2 + \lambda\|\mathbf{X}\|_*, \tag{15}$$

where λ is a regularization parameter. As is shown in the simulation, we can estimate the (r,f) pairs by localizing the peaks of the norm of the corresponding dual polynomial. We leave the robust performance analysis of this framework for future work. In particular, it would be interesting to characterize the parameter estimation performance of program (15) in terms of the signal-to-noise ratio and the separation of the true frequencies, similar to the analysis in [42].

4.2 Extension to the missing data case

Unlike the classical formulation of MUSIC, the optimization-based derivation of NN-MUSIC allows it to be naturally extended to the missing data case. In particular, assume that we partially observe the entries of the full data matrix $\mathbf{X}^\star$ in (3) with uniformly random sampled locations $\Omega \subset \{1,\ldots,M\} \times \{1,\ldots,N\}$. Let $\mathbf{X}_\Omega = \mathcal{P}_\Omega(\mathbf{X})$ be the projection matrix of $\mathbf{X}$ on the index set Ω, i.e

$$\mathbf{X}_{\Omega ij} = \mathcal{P}_{(i,j)}(\mathbf{X}) = \begin{cases} \mathbf{X}_{ij}, & (i,j) \in \Omega, \\ 0, & else. \end{cases}$$

Notice that recovering the missing entries of the matrix $\mathbf{X}^\star$ reduces to a matrix completion problem [20], commonly formulated via the following NNM

$$\widehat{\mathbf{X}} = \arg\min_{\mathbf{X}} \ \|\mathbf{X}\|_* \quad \text{s.t.} \ \mathbf{X}_\Omega = \mathbf{X}_\Omega^\star, \tag{16}$$

which can be solved by the corresponding semi-definite program (SDP)

$$\min_{\mathbf{X},\mathbf{T},\mathbf{D}} \frac{1}{2}\operatorname{Tr}(\mathbf{T}) + \frac{1}{2}\operatorname{Tr}(\mathbf{D}) \quad \text{s.t.} \ \begin{bmatrix} \mathbf{T} & \mathbf{X} \\ \mathbf{X}^H & \mathbf{D} \end{bmatrix} \succeq 0, \ \mathbf{X}_\Omega = \mathbf{X}_\Omega^\star. \tag{17}$$

The dual problem of (17) is given by

$$\max_{\mathbf{Q}} \ \langle \mathbf{Q}_\Omega, \mathbf{X}_\Omega^\star \rangle_{\mathbb{R}} \quad \text{s.t.} \ \|\mathbf{Q}_\Omega\|_2 \le 1. \tag{18}$$

Therefore, we can define the dual polynomial as $\mathcal{Q}(r,f) \triangleq \mathbf{Q}^H \boldsymbol{a}(r,f)$, where $\mathbf{Q}$ is the dual solution. Similar to Theorem 4.1, the following theorem guarantees that we can identify the true r_k's and f_k's by localizing the places where $\|\mathcal{Q}(r,f)\|_2$ achieves 1.

[5] This fact is key to building the connection between MUSIC and nuclear norm minimization. This is why we construct $\mathbf{Q} = \mathbf{U}_{X^\star}\mathbf{V}_{X^\star}^H$ explicitly in line 3 of Algorithm 2; this corresponds to a choice of $\mathbf{W} = \mathbf{0}$ in (13). Note that other constructions of $\mathbf{Q}$ with $\mathbf{W} \neq \mathbf{0}$ will still work for localizing the (r,f) pairs but will not equal $\|\mathbf{U}_X^H a(r,f)\|_2$.

Theorem 4.2. *Suppose $\mathbf{X}^\star$ is a data matrix of the form (3) and all the (r_k, f_k) pairs are distinct. Given the uniformly partial random observed data matrix $\mathbf{X}^\star_\Omega$. For any data matrix $\widehat{\mathbf{X}}$ obtained by solving the NNM problem (16), we denote $\mathbf{U}_{\widehat{\mathbf{X}}}\mathbf{S}_{\widehat{\mathbf{X}}}\mathbf{V}_{\widehat{\mathbf{X}}}^H$ as a truncated SVD of $\widehat{\mathbf{X}}$. Denote $\mathbf{Q}_\Omega$ as the projection matrix of the dual solution $\mathbf{Q}$ on the index set Ω. Then, $\mathbf{Q}_\Omega$ has the form $\mathbf{Q}_\Omega = \mathbf{U}_{\widehat{\mathbf{X}}}\mathbf{V}_{\widehat{\mathbf{X}}}^H + \mathbf{W}$ with $\mathbf{U}_{\widehat{\mathbf{X}}}^H\mathbf{W} = \mathbf{0}$, $\mathbf{W}\mathbf{V}_{\widehat{\mathbf{X}}} = \mathbf{0}$, and $\|\mathbf{W}\| \le 1$ for some matrix $\mathbf{W}$. Here, we consider dual solutions with $\mathbf{Q} = \mathbf{Q}_\Omega$.[6] If the range space of $\widehat{\mathbf{X}}$ contains the range space of $\mathbf{X}^\star$ (i.e., $\mathcal{R}(\widehat{\mathbf{X}}) \supseteq \mathcal{R}(\mathbf{X}^\star)$) and $\|\mathbf{W}\| < 1$[7], the dual polynomial $\mathcal{Q}(r,f) = \mathbf{Q}^H \boldsymbol{a}(r,f)$ satisfies*

$$\begin{aligned}
&\|\mathcal{Q}(r,f)\|_2 = 1, \quad \forall\ (r,f) \in \mathcal{RF}, \ \text{or} \ \forall\ (r,f) \notin \mathcal{RF} \ \text{with} \ \boldsymbol{a}(r,f) \in \mathcal{R}(\mathbf{U}_{\widehat{\mathbf{X}}}),\\
&\|\mathcal{Q}(r,f)\|_2 < 1, \quad \forall\ (r,f) \notin \mathcal{RF} \ \text{with} \ \boldsymbol{a}(r,f) \notin \mathcal{R}(\mathbf{U}_{\widehat{\mathbf{X}}}).
\end{aligned}$$

The above theorem is proved in Section 5.2. Note that one can even localize the true (r, f) pairs from the dual polynomial when we do not have perfectly recovered data matrix (i.e., $\widehat{\mathbf{X}} \neq \mathbf{X}^\star$). For the case when we do have perfect data recovery, i.e., $\widehat{\mathbf{X}} = \mathbf{X}^\star$, we can further conclude that $\|\mathcal{Q}(r,f)\|_2 = 1$ only when $(r,f) \in \mathcal{RF}$ and $\|\mathcal{Q}(r,f)\|_2 < 1$ as long as $(r,f) \notin \mathcal{RF}$. Moreover, we can also quantify sample complexity needed for perfect data recovery in terms of the explicit parameters such as damping ratios and frequencies instead of some incoherence property. We summarize these results in the following corollary.

Corollary 4.1. *Suppose $\mathbf{X}^\star$ is a data matrix of the form (3) and all the (r_k, f_k) pairs are distinct. Given the uniformly partial random observed data matrix $\mathbf{X}^\star_\Omega$, suppose $|\Omega| \ge c_1\mu_1 c_s K \log^4(MN)$ for some numerical constants $c_1 > 0$ and $c_s \triangleq \max\{M, N\}$. Here, $\mu_1 \ge \max\left\{\frac{M}{\mathcal{L}(M,r,f)}, \frac{\mu_2}{\sigma^2_{\min}(\mathbf{\Phi})}\right\}$ denotes an incoherence parameter with $\mu_2 \triangleq \max_{1\le n\le N}\left(\sum_{k=1}^K |\phi_{nk}|^2\right)\frac{N}{K}$, $\mathcal{L}(M,r,f)$ is a function of M, r, f and defined as $\mathcal{L}(M,r,f) \triangleq \min\limits_{1\le k\le K} \frac{1}{r_k}\left[\gamma_M(r_k) - \frac{c_2}{\Delta_f}(1 + r_k^{2M})\right]$ with c_2 being a constant and*

$$\gamma_M(r_k) \triangleq \begin{cases} \frac{r_k^{2M}-1}{2\log(r_k)}, & r_k < 1,\\ M, & r_k = 1,\end{cases}$$

and $\Delta_f \triangleq \min_{k\neq l} |f_k - f_l|$ denotes the minimum separation between true frequencies, where $|f_k - f_l|$ is the wrap-around distance on the unit circle. Then, $\mathbf{X}^\star$ is the unique solution of (16) with probability at least $1 - (MN)^{-2}$. Moreover, when $\widehat{\mathbf{X}} = \mathbf{X}^\star$ is the unique solution to (16), the dual solution $\mathbf{Q}$ has the form $\mathbf{Q} = \mathbf{U}_{\mathbf{X}^\star}\mathbf{V}_{\mathbf{X}^\star}^H + \mathbf{W}$, where $\mathbf{U}_{X^\star}^H\mathbf{W} = \mathbf{0}$, $\mathbf{W}\mathbf{V}_{X^\star} = \mathbf{0}$, and $\|\mathbf{W}\| \le 1$. In this case, if $\|\mathbf{W}\| < 1$, then the dual polynomial satisfies

$$\begin{aligned}
\|\mathcal{Q}(r_k,f_k)\|_2 = 1,\ \forall\ (r_k,f_k) \in \mathcal{RF},\\
\|\mathcal{Q}(r,f)\|_2 < 1,\ \forall\ (r,f) \notin \mathcal{RF}.
\end{aligned}$$

The proof for Corollary 4.1 relies on some of the results in [46]. However, those results do not extend directly to the damped exponential case. Rather than use a certain incoherence property as in [46], we incorporate the damping ratios $\{r_k\}$ into the signal and develop theoretical guarantees that explicitly depend on the parameters, i.e., damping ratios and minimum frequency separation. In particular, we explicitly bound the minimal singular value of $\widetilde{\mathbf{A}}_{rf}^H\widetilde{\mathbf{A}}_{rf}$ with the function $\mathcal{L}(M,r,f)$ by exploiting the Vandermonde structure of $\widetilde{\mathbf{A}}_{rf}$ [47], instead of giving an incoherence property depending on just the minimal singular value of

[6]Recall that we observe CVX with the default solver SDPT3 always returns a dual solution with minimal energy, so the entries of $\mathbf{Q}$ outside the index Ω are always 0, i.e., we have $\mathbf{Q} = \mathbf{Q}_\Omega$.

[7]Same as in the full data case, when $\|\mathbf{W}\| = 1$, one may get $\|\mathcal{Q}(r,f)\|_2 = 1$ for some $(r,f) \notin \mathcal{RF}$ with $\boldsymbol{a}(r,f) \notin \mathcal{R}(\mathbf{U}_{\widehat{\mathbf{X}}})$, i.e., having some false peaks when checking the norm of dual polynomial. Again, one can remove these false peaks by solving a least squares problem.

$\widetilde{\mathbf{A}}_{rf}^H \widetilde{\mathbf{A}}_{rf}$ as in [46]. Note that $\mathcal{L}(M,r,f) = M - \frac{2c_2}{\Delta_f}$ is on the order of M when there is no damping (i.e., $r=1$) and the frequencies are well separated ($\Delta_f = \mathcal{O}(\frac{1}{M})$). In the case where $r<1$, the sample complexity $|\Omega|$ scales inversely with $\mathcal{L}(M,r,f) = \min_{1\le k\le K} \frac{1}{r_k}\left[\frac{r_k^{2M}-1}{2\log(r_k)} - \frac{c_2}{\Delta_f}(1+r_k^{2M})\right]$, which increases monotonically with r_k when the constant c_2 is sufficiently small. Therefore, it can be seen that the sample complexity $|\Omega|$ decreases if the minimum frequency separation Δ_f increases or the the damping ratio r_k increases. Please see Section 5.3 for details and Section 6.3 for supporting experiments.

Note that the set Ω is chosen uniformly at random from all subsets of $\{1,\dots,M\}\times\{1,\dots,N\}$ with a given cardinality $|\Omega|$. Since the columns of $\mathbf{\Phi}\in\mathbb{C}^{N\times K}$ are assumed to be normalized, we have $1\le\mu_2\le N$ according to the definition of μ_2. In particular, $\mu_2=1$ when all the entries of $\mathbf{\Phi}$ have magnitude $\frac{1}{\sqrt{N}}$ and $\mu_2=N$ when $\mathbf{\Phi}$ has a row containing all 1's with all other rows being 0. Due to the normalized columns in $\mathbf{\Phi}$, we also have that $\sigma_{\min}^2(\mathbf{\Phi})\le 1$. Therefore, it can be seen from the above theorem that when there is no damping (or only light damping, i.e., r_k is close to 1) and the frequencies are well separated, and μ_2 is close to 1, we can bound μ_1 by a constant and thus the number of measurements needed for perfect recovery is comparable to best case bounds for rank-K matrix completion. Specifically, state-of-the-art bounds [43] for low-rank matrix completion from uniform random samples involve a dependence on a certain coherence parameter (equal to the maximum leverage score of the matrix); when this coherence parameter is small, the sample complexity is $|\Omega| = \mathcal{O}(\max(M,N)K)$ (up to logarithmic factors). The significance to Corollary 4.1 is that the sample complexity is not stated in terms of the matrix coherence; rather, the dependence on the damping ratios and minimum frequency separation is explicitly revealed. We also note that this is quite distinct from the work [46], in which the theoretical guarantees are built on a different incoherence property rather than the explicit parameters such as frequencies.

Inspired by Algorithm 2 and the above analysis, we propose the following Missing Data MUSIC algorithm, named MD-MUSIC, to identify the damping factors $\{r_k\}$ and frequencies $\{f_k\}$ from the partially observed data matrix $\mathbf{X}_\Omega^\star$. Note that any off-the-shelf SDP solver could be used to solve the SDP in (17)[8].

Algorithm 3 MD-MUSIC

```
procedure INPUT(X*_Ω ∈ C^{M×N})
    compute X̂ and Q by solving the SDP (17)
    form the dual polynomial: Q(r, f) = Q^H a(r, f)
    localize the places where ‖Q(r, f)‖_2 = 1 to get (r̂_k, f̂_k)
    return X̂ and (r̂_k, f̂_k)
end procedure
```

Finally, we note that one could also consider an alternative approach wherein one first solves the NNM problem in (17) and then uses Algorithm 2 to identify the r_k's and f_k's using $\widehat{\mathbf{X}}$. Interestingly, however, as we demonstrate in Theorem 4.2, it is sometimes possible with MD-MUSIC to perfectly recover the r_k's and f_k's even when exact recovery of $\mathbf{X}^\star$ fails. This implies that MD-MUSIC is actually more powerful than the alternative approach mentioned above. We also conduct simulations to further present this interesting phenomenon (parameter recovery without exact data matrix recovery) in Section 6.

[8] Primal-dual algorithms that are used in solvers such as CVX can return both the data matrix $\widehat{\mathbf{X}}$ and the dual solution $\mathbf{Q}$ of (17) and (18), respectively.

5 Proofs

5.1 Proof for Theorem 4.1

Denote a truncated SVD of $\mathbf{A}_{rf}$ as $\mathbf{A}_{rf} = \mathbf{U}_{A_{rf}}\mathbf{S}_{A_{rf}}\mathbf{V}_{A_{rf}}^H$. We first consider the case when $\mathbf{W} = \mathbf{0}$, i.e., $\mathbf{Q} = \mathbf{U}_{X^\star}\mathbf{V}_{X^\star}^H$. Note that

$$\begin{aligned}
&\|\mathbf{Q}^H\boldsymbol{a}(r,f)\|_2^2 = \|\mathbf{U}_{X^\star}^H\boldsymbol{a}(r,f)\|_2^2 = \boldsymbol{a}(r,f)^H\mathbf{U}_{X^\star}\mathbf{U}_{X^\star}^H\boldsymbol{a}(r,f) = \boldsymbol{a}(r,f)^H\mathbf{U}_{X^\star}\mathbf{S}_{X^\star}\mathbf{V}_{X^\star}^H\mathbf{V}_{X^\star}\mathbf{S}_{X^\star}^{-1}\mathbf{U}_{X^\star}^H\boldsymbol{a}(r,f)\\
=&\boldsymbol{a}(r,f)^H\mathbf{X}^\star(\mathbf{X}^\star)^\dagger\boldsymbol{a}(r,f) = \boldsymbol{a}(r,f)^H\mathbf{A}_{rf}\mathbf{D}_{\widetilde{c}}\mathbf{\Phi}^\top\left(\mathbf{\Phi}^\top\right)^\dagger\mathbf{D}_{\widetilde{c}}^{-1}\mathbf{A}_{rf}^\dagger\boldsymbol{a}(r,f) = \boldsymbol{a}(r,f)^H\mathbf{A}_{rf}\mathbf{A}_{rf}^\dagger\boldsymbol{a}(r,f)\\
=&\boldsymbol{a}(r,f)^H\mathbf{U}_{A_{rf}}\mathbf{U}_{A_{rf}}^H\boldsymbol{a}(r,f) = \left\langle \mathcal{P}_{\mathbf{U}_{A_{rf}}}(\boldsymbol{a}(r,f)), \boldsymbol{a}(r,f)\right\rangle,
\end{aligned}$$

where $\mathcal{P}_{\mathbf{U}_{A_{rf}}}(\boldsymbol{a}(r,f)) \triangleq \mathbf{U}_{A_{rf}}\mathbf{U}_{A_{rf}}^H\boldsymbol{a}(r,f)$ is defined as the orthogonal projection of $\boldsymbol{a}(r,f)$ onto the range space of $\mathbf{U}_{A_{rf}}$, i.e., $\mathcal{R}(\mathbf{U}_{A_{rf}})$. Note that the third equality is obtained by plugging in $\mathbf{I} = \mathbf{S}_{X^\star}\mathbf{V}_{X^\star}^H\mathbf{V}_{X^\star}\mathbf{S}_{X^\star}^{-1}$ while the fifth equality is obtained by plugging in $\mathbf{X}^\star = \mathbf{A}_{rf}\mathbf{D}_{\widetilde{c}}\mathbf{\Phi}^\top$ and $(\mathbf{X}^\star)^\dagger = (\mathbf{\Phi}^\top)^\dagger\mathbf{D}_{\widetilde{c}}^{-1}\mathbf{A}_{rf}^\dagger$. Also, note that $\mathbf{\Phi}^\top\left(\mathbf{\Phi}^\top\right)^\dagger = \mathbf{I}$ when $\mathbf{\Phi} \in \mathbb{C}^{N\times K}$ is of rank K, which gives the sixth equality. The seventh equality holds due to $\mathbf{A}_{rf}\mathbf{A}_{rf}^\dagger = \mathbf{U}_{A_{rf}}\mathbf{S}_{A_{rf}}\mathbf{V}_{A_{rf}}^H\mathbf{V}_{A_{rf}}\mathbf{S}_{A_{rf}}^{-1}\mathbf{U}_{A_{rf}}^H = \mathbf{U}_{A_{rf}}\mathbf{U}_{A_{rf}}^H$.

- For all $(r_k, f_k) \in \mathcal{RF}$, we have $\boldsymbol{a}(r_k,f_k) \in \mathcal{R}(\mathbf{U}_{A_{rf}})$, which implies $\mathcal{P}_{\mathbf{U}_{A_{rf}}}(\boldsymbol{a}(r_k,f_k)) = \boldsymbol{a}(r_k,f_k)$. Therefore, we have $\|\mathbf{Q}^H\boldsymbol{a}(r_k,f_k)\|_2^2 = \langle\boldsymbol{a}(r_k,f_k), \boldsymbol{a}(r_k,f_k)\rangle = 1$.
- For all $(r,f) \notin \mathcal{RF}$, if we have $\boldsymbol{a}(r,f) \notin \mathcal{R}(\mathbf{U}_{A_{rf}})$, which implies $\mathcal{P}_{\mathbf{U}_{A_{rf}}}(\boldsymbol{a}(r,f)) = \boldsymbol{a}(r,f) - \mathcal{P}_{\mathbf{U}_{A_{rf}}^\perp}(\boldsymbol{a}(r,f))$, we would then have

$$\|\mathbf{Q}^H\boldsymbol{a}(r,f)\|_2^2 = \langle\boldsymbol{a}(r,f), \boldsymbol{a}(r,f)\rangle - \left\langle \mathcal{P}_{\mathbf{U}_{A_{rf}}^\perp}(\boldsymbol{a}(r,f)), \boldsymbol{a}(r,f)\right\rangle < \langle\boldsymbol{a}(r,f), \boldsymbol{a}(r,f)\rangle = 1.$$

Thus, we only need to show $\boldsymbol{a}(r,f) \notin \mathcal{R}(\mathbf{U}_{A_{rf}})$ for all $(r,f) \notin \mathcal{RF}$. Define a Vandermonde matrix $\mathbf{A}_{rf}^v \in \mathbb{C}^{M\times K}$ as

$$\mathbf{A}_{rf}^v \triangleq [\boldsymbol{a}^v(r_1,f_1), \cdots, \boldsymbol{a}^v(r_K,f_K)] \tag{19}$$

with $\boldsymbol{a}^v(r,f) \triangleq [1\ re^{j2\pi f1}\ \cdots\ r^{M-1}e^{j2\pi f(M-1)}]^\top$, which is the unnormalized version of $\boldsymbol{a}(r,f)$. Then, $\mathbf{A}_{rf}$ is the column-normalized version of $\mathbf{A}_{rf}^v$. Assuming $M \geq K$, it follows that the first K rows of $\mathbf{A}_{rf}^v$ form a square Vandermonde matrix, denoted as $\mathbf{A}_{rf}^K$, whose determinant is given by [44, 45]

$$\det(\mathbf{A}_{rf}^K) = \prod_{1\leq i<k\leq K}(r_k e^{j2\pi f_k} - r_i e^{j2\pi f_i}).$$

Then, $\operatorname{rank}(\mathbf{A}_{rf}) = \operatorname{rank}(\mathbf{A}_{rf}^v) = K$ if $M \geq K$ and $(r_i, f_i) \neq (r_k, f_k)$ for all $i \neq k$. Similarly, we have

$$\operatorname{rank}([\mathbf{A}_{rf}|\boldsymbol{a}(r,f)]) = K+1,$$

i.e., $\boldsymbol{a}(r,f) \notin \mathcal{R}(\mathbf{A}_{rf}) = \mathcal{R}(\mathbf{U}_{\mathbf{A}_{rf}})$ if $(r,f) \notin \mathcal{RF}$, $M \geq K+1$ and all the (r_k,f_k) pairs in $\mathcal{RF}$ are distinct.

It remains to show the case when $\mathbf{W} \neq \mathbf{0}$. In particular, we have $\|\mathcal{Q}(r,f)\|_2^2 = \|\mathbf{U}_{X^\star}^H\boldsymbol{a}(r,f)\|_2^2 + \|\mathbf{W}^H\boldsymbol{a}(r,f)\|_2^2$.

- For all $(r_k,f_k) \in \mathcal{RF}$, we have $\boldsymbol{a}(r_k,f_k) \in \mathcal{R}(\mathbf{U}_{A_{rf}}) = \mathcal{R}(\mathbf{U}_{X^\star})$, and so together with $\mathbf{U}_{X^\star}^H\mathbf{W} = \mathbf{0}$, we have $\|\mathcal{Q}(r_k,f_k)\|_2^2 = \|\mathbf{U}_{X^\star}^H\boldsymbol{a}(r_k,f_k)\|_2^2 = 1$.

- For all $(r,f)\notin\mathcal{RF}$, we have $\boldsymbol{a}(r,f)\notin\mathcal{R}(\mathbf{U}_{A_{rf}})=\mathcal{R}(\mathbf{U}_{X^\star})$ as shown above. Denote $\mathbf{W}=\mathbf{U}_{X^\star}^{\perp}\mathbf{S}_W{\mathbf{V}_{X^\star}^{\perp}}^H$ as a truncated SVD of $\mathbf{W}$ with $\mathbf{U}_{X^\star}^H\mathbf{U}_{X^\star}^{\perp}=\mathbf{0}$, $\mathbf{V}_{X^\star}^H\mathbf{V}_{X^\star}^{\perp}=\mathbf{0}$ and $\|\mathbf{S}_W\|=\|\mathbf{W}\|\le 1$. In the case where $\|\mathbf{S}_W\|=\|\mathbf{W}\|<1$, we have

$$\begin{aligned}\|\mathcal{Q}(r,f)\|_2^2&=\|\mathbf{U}_{X^\star}^H\boldsymbol{a}(r,f)\|_2^2+\|\mathbf{S}_W{\mathbf{U}_{X^\star}^{\perp}}^H\boldsymbol{a}(r,f)\|_2^2\le\|\mathbf{U}_{X^\star}^H\boldsymbol{a}(r,f)\|_2^2+\|\mathbf{S}_W\|^2\|{\mathbf{U}_{X^\star}^{\perp}}^H\boldsymbol{a}(r,f)\|_2^2\\&<\|\mathbf{U}_{X^\star}^H\boldsymbol{a}(r,f)\|_2^2+\|{\mathbf{U}_{X^\star}^{\perp}}^H\boldsymbol{a}(r,f)\|_2^2\le 1.\end{aligned}$$

This completes the proof of Theorem 4.1.

5.2 Proof for Theorem 4.2

Unlike in Corollary 4.1, we now focus on the case when the data matrix is not perfectly recovered, i.e., $\widehat{\mathbf{X}}\neq\mathbf{X}^\star$. Again, with some fundamental Lagrange analysis, we can conclude that the dual solution $\mathbf{Q}=\mathbf{Q}_\Omega$ with minimal energy belongs to the subdifferential of $\|\widehat{\mathbf{X}}\|_*$. Then, we have $\mathbf{Q}=\mathbf{Q}_\Omega=\mathbf{U}_{\widehat{\mathbf{X}}}\mathbf{V}_{\widehat{\mathbf{X}}}^H+\mathbf{W}$ with $\mathbf{U}_{\widehat{\mathbf{X}}}^H\mathbf{W}=\mathbf{0}$, $\mathbf{W}\mathbf{V}_{\widehat{\mathbf{X}}}=\mathbf{0}$, and $\|\mathbf{W}\|\le 1$ for some matrix $\mathbf{W}$. Note that $\mathbf{Q}=\mathbf{Q}_\Omega$ is a sparse matrix with zero entries on the complement index of Ω, so $\mathbf{W}$ should be a non-zero matrix. Recall that $\mathbf{X}^\star=\mathbf{A}_{rf}\mathbf{D}_{\widetilde{c}}\mathbf{\Phi}^\top$ with $\mathbf{A}_{rf}=[\boldsymbol{a}(r_1,f_1),\cdots,\boldsymbol{a}(r_K,f_K)]$. Then, we have

$$\boldsymbol{a}(r_k,f_k)\in\mathcal{R}(\mathbf{A}_{rf})=\mathcal{R}(\mathbf{X}^\star)\subseteq\mathcal{R}(\widehat{\mathbf{X}})=\mathcal{R}(\mathbf{U}_{\widehat{\mathbf{X}}}),\quad k=1,\ldots,K.$$

- For all $(r_k,f_k)\in\mathcal{RF}$, we have $\boldsymbol{a}(r_k,f_k)\in\mathcal{R}(\mathbf{U}_{\widehat{\mathbf{X}}})$, and so together with $\mathbf{U}_{\widehat{\mathbf{X}}}^H\mathbf{W}=\mathbf{0}$, we have $\|\mathcal{Q}(r_k,f_k)\|_2^2=\|\mathbf{U}_{\widehat{\mathbf{X}}}^H\boldsymbol{a}(r_k,f_k)\|_2^2=1$.
- For all $(r,f)\notin\mathcal{RF}$, we have $\boldsymbol{a}(r,f)\notin\mathcal{R}(\mathbf{A}_{rf})=\mathcal{R}(\mathbf{X}^\star)$ as shown in Section 5.1. (1) If $\boldsymbol{a}(r,f)\in\mathcal{R}(\mathbf{U}_{\widehat{\mathbf{X}}})$, we still have $\|\mathcal{Q}(r,f)\|_2^2=1$ as above. This could result in false estimation of (r,f) pairs when we localize the places where $\|\mathcal{Q}(r,f)\|_2$ achieves 1. (2) If $\boldsymbol{a}(r,f)\notin\mathcal{R}(\mathbf{U}_{\widehat{\mathbf{X}}})$, denote $\mathbf{W}=\mathbf{U}_{\widehat{\mathbf{X}}}^{\perp}\mathbf{S}_W{\mathbf{V}_{\widehat{\mathbf{X}}}^{\perp}}^H$ as a truncated SVD of $\mathbf{W}$ with $\mathbf{U}_{\widehat{\mathbf{X}}}^H\mathbf{U}_{\widehat{\mathbf{X}}}^{\perp}=\mathbf{0}$, $\mathbf{V}_{\widehat{\mathbf{X}}}^H\mathbf{V}_{\widehat{\mathbf{X}}}^{\perp}=\mathbf{0}$ and $\|\mathbf{S}_W\|=\|\mathbf{W}\|\le 1$. Recall that $\mathbf{W}\neq\mathbf{0}$. In the case where $\|\mathbf{S}_W\|=\|\mathbf{W}\|<1$, we have

$$\begin{aligned}\|\mathcal{Q}(r,f)\|_2^2&=\|\mathbf{U}_{\widehat{\mathbf{X}}}^H\boldsymbol{a}(r,f)\|_2^2+\|\mathbf{S}_W{\mathbf{U}_{\widehat{\mathbf{X}}}^{\perp}}^H\boldsymbol{a}(r,f)\|_2^2\le\|\mathbf{U}_{\widehat{\mathbf{X}}}^H\boldsymbol{a}(r,f)\|_2^2+\|\mathbf{S}_W\|^2\|{\mathbf{U}_{\widehat{\mathbf{X}}}^{\perp}}^H\boldsymbol{a}(r,f)\|_2^2\\&<\|\mathbf{U}_{\widehat{\mathbf{X}}}^H\boldsymbol{a}(r,f)\|_2^2+\|{\mathbf{U}_{\widehat{\mathbf{X}}}^{\perp}}^H\boldsymbol{a}(r,f)\|_2^2\le 1.\end{aligned}$$

Thus, we finish the proof of Theorem 4.2.

5.3 Proof for Corollary 4.1

Define $\mathbf{D}_{cM}\triangleq\sqrt{M}\,\mathrm{diag}([c_1,\ c_2,\ \cdots,\ c_K])$, $\widetilde{\mathbf{A}}_{rf}\triangleq\frac{1}{\sqrt{M}}\mathbf{A}_{rf}^v$, where $\mathbf{A}_{rf}^v$ is the unnormalized Vandermonde matrix and defined in (19). Observe that the transpose of the noiseless data matrix $\mathbf{X}_\top^\star\triangleq{\mathbf{X}^\star}^\top=\mathbf{\Phi}\mathbf{D}_{\widetilde{c}}\mathbf{A}_{rf}^\top=\mathbf{\Phi}\mathbf{D}_{cM}\widetilde{\mathbf{A}}_{rf}^\top$ can be viewed as the block Hankel matrix $\mathbf{X}_e$ introduced in [46], but with $k_1=N$ and $k_2=1$. Define $\mathcal{P}_T$ as the projection operator that acts on the tangent space of $\mathbf{X}_\top^\star$. Denote a truncated SVD of $\mathbf{X}_\top^\star$ as $\mathbf{X}_\top^\star=\mathbf{U}_{\mathbf{X}_\top^\star}\mathbf{S}_{\mathbf{X}_\top^\star}\mathbf{V}_{\mathbf{X}_\top^\star}^H$. Then, [46, Lemma 1] can be adapted to provide us sufficient conditions that are used to guarantee the unique optimality of $\mathbf{X}^\star$. In particular, we can set $\mathcal{A}$ and $\mathcal{A}_\Omega$ as in [46, Lemma 1] as the identity operator and the random sampling operator $\mathcal{P}_\Omega$, respectively. Therefore, we need the following condition

$$\left\|\mathcal{P}_T-\frac{NM}{|\Omega|}\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T\right\|\le\frac{1}{2}. \tag{20}$$

Next, we verify that the above condition (20) holds with high probability under certain conditions. Define $\mathbf{A}_{(n,m)}\in\mathbb{R}^{N\times M}$ as a matrix with the (n,m)-th entry being 1 and others being 0. We first quantify the projection of $\mathbf{A}_{(n,m)}$ onto the subspace T, the tangent space of $\mathbf{X}_\top^\star$. In particular, we have the following lemma which utilizes a quite different incorence property than the one used in [46, Lemma 2].

Lemma 5.1. *For some constant μ_1, if*

$$\sigma_{\min}(\mathbf{\Phi}^H\mathbf{\Phi}) \geq \frac{\mu_2}{\mu_1} \quad \text{and} \quad \sigma_{\min}(\widetilde{\mathbf{A}}_{rf}^H\widetilde{\mathbf{A}}_{rf}) \geq \frac{1}{\mu_1}, \tag{21}$$

then

$$\|\mathbf{U}_{\mathbf{X}_\top^\star}\mathbf{U}_{\mathbf{X}_\top^\star}^H\mathbf{A}_{(n,m)}\|_F^2 \leq \frac{\mu_1 c_s K}{NM}, \quad \|\mathbf{A}_{(n,m)}\mathbf{V}_{\mathbf{X}_\top^\star}\mathbf{V}_{\mathbf{X}_\top^\star}^H\|_F^2 \leq \frac{\mu_1 c_s K}{NM}$$

hold for any $(n,m) \in [N]\times[M]$ with $[N] \triangleq \{1,2,\ldots,N\}$ and $[M] \triangleq \{1,2,\ldots,M\}$. We have defined $c_s \triangleq \max\{N,M\}$ and $\mu_2 \triangleq \max_{1\leq n\leq N}\left(\sum_{k=1}^K|\phi_{nk}|^2\right)\frac{N}{K}$. It follows that

$$\|\mathcal{P}_T(\mathbf{A}_{(n,m)})\|_F^2 \leq \|\mathbf{U}_{\mathbf{X}_\top^\star}\mathbf{U}_{\mathbf{X}_\top^\star}^H\mathbf{A}_{(n,m)}\|_F^2 + \|\mathbf{A}_{(n,m)}\mathbf{V}_{\mathbf{X}_\top^\star}\mathbf{V}_{\mathbf{X}_\top^\star}^H\|_F^2 \leq \frac{2\mu_1 c_s K}{NM}. \tag{22}$$

Proof. Note that $\mathbf{U}_{\mathbf{X}_\top^\star}$ ($\mathbf{V}_{\mathbf{X}_\top^\star}$) and $\mathbf{\Phi}$ ($\widetilde{\mathbf{A}}_{rf}$) determine the same column (row) space of $\mathbf{X}_\top^\star$. In particular, we have

$$\mathbf{U}_{\mathbf{X}_\top^\star}\mathbf{U}_{\mathbf{X}_\top^\star}^H = \mathbf{\Phi}(\mathbf{\Phi}^H\mathbf{\Phi})^{-1}\mathbf{\Phi}^H, \quad \mathbf{V}_{\mathbf{X}_\top^\star}\mathbf{V}_{\mathbf{X}_\top^\star}^H = \widetilde{\mathbf{A}}_{rf}(\widetilde{\mathbf{A}}_{rf}^H\widetilde{\mathbf{A}}_{rf})^{-1}\widetilde{\mathbf{A}}_{rf}^H,$$

which implies

$$\begin{aligned}\|\mathbf{U}_{\mathbf{X}_\top^\star}\mathbf{U}_{\mathbf{X}_\top^\star}^H\mathbf{A}_{(n,m)}\|_F^2 =& \|\mathbf{\Phi}(\mathbf{\Phi}^H\mathbf{\Phi})^{-1}\mathbf{\Phi}^H\mathbf{A}_{(n,m)}\|_F^2 = \left\langle(\mathbf{\Phi}^H\mathbf{\Phi})^{-1}\mathbf{\Phi}^H\mathbf{A}_{(n,m)}, \mathbf{\Phi}^H\mathbf{A}_{(n,m)}\right\rangle \\ \leq& \|(\mathbf{\Phi}^H\mathbf{\Phi})^{-1}\|\|\mathbf{\Phi}^H\mathbf{A}_{(n,m)}\|_F^2 = \frac{1}{\sigma_{\min}(\mathbf{\Phi}^H\mathbf{\Phi})}\|\mathbf{\Phi}^H\mathbf{A}_{(n,m)}\|_F^2\end{aligned}$$

and

$$\begin{aligned}\|\mathbf{A}_{(n,m)}\mathbf{V}_{\mathbf{X}_\top^\star}\mathbf{V}_{\mathbf{X}_\top^\star}^H\|_F^2 =& \|\mathbf{A}_{(n,m)}\widetilde{\mathbf{A}}_{rf}(\widetilde{\mathbf{A}}_{rf}^H\widetilde{\mathbf{A}}_{rf})^{-1}\widetilde{\mathbf{A}}_{rf}^H\|_F^2 = \left\langle\mathbf{A}_{(n,m)}\widetilde{\mathbf{A}}_{rf}(\widetilde{\mathbf{A}}_{rf}^H\widetilde{\mathbf{A}}_{rf})^{-1}, \mathbf{A}_{(n,m)}\widetilde{\mathbf{A}}_{rf}\right\rangle \\ \leq& \|(\widetilde{\mathbf{A}}_{rf}^H\widetilde{\mathbf{A}}_{rf})^{-1}\|\|\mathbf{A}_{(n,m)}\widetilde{\mathbf{A}}_{rf}\|_F^2 = \frac{1}{\sigma_{\min}(\widetilde{\mathbf{A}}_{rf}^H\widetilde{\mathbf{A}}_{rf})}\|\mathbf{A}_{(n,m)}\widetilde{\mathbf{A}}_{rf}\|_F^2.\end{aligned}$$

Define $\mu_2 \triangleq \max_{1\leq n\leq N}\left(\sum_{k=1}^K|\phi_{nk}|^2\right)\frac{N}{K}$. Note that $1 \leq \mu_2 \leq N$. Recall that $\mathbf{A}_{(n,m)} \in \mathbb{R}^{N\times M}$ is a matrix with the (n,m)-th entry being 1 and all others being 0. Therefore, we can bound $\|\mathbf{\Phi}^H\mathbf{A}_{(n,m)}\|_F^2$ and $\|\mathbf{A}_{(n,m)}\widetilde{\mathbf{A}}_{rf}\|_F^2$ with

$$\|\mathbf{\Phi}^H\mathbf{A}_{(n,m)}\|_F^2 = \sum_{k=1}^K|\phi_{nk}|^2 \leq \mu_2\frac{K}{N}, \quad \|\mathbf{A}_{(n,m)}\widetilde{\mathbf{A}}_{rf}\|_F^2 = \sum_{k=1}^K\frac{1}{M}r_k^{2m} \leq \frac{K}{M}.$$

Define $c_s \triangleq \max(N,M)$. Then, if

$$\sigma_{\min}(\mathbf{\Phi}^H\mathbf{\Phi}) \geq \frac{\mu_2}{\mu_1}, \quad \sigma_{\min}(\widetilde{\mathbf{A}}_{rf}^H\widetilde{\mathbf{A}}_{rf}) \geq \frac{1}{\mu_1},$$

we can get

$$\begin{aligned}\|\mathbf{U}_{\mathbf{X}_\top^\star}\mathbf{U}_{\mathbf{X}_\top^\star}^H\mathbf{A}_{(n,m)}\|_F^2 \leq \frac{\mu_1 K}{N} = \frac{\mu_1 KM}{NM} \leq \frac{\mu_1 Kc_s}{NM}, \\ \|\mathbf{A}_{(n,m)}\mathbf{V}_{\mathbf{X}_\top^\star}\mathbf{V}_{\mathbf{X}_\top^\star}^H\|_F^2 \leq \frac{\mu_1 K}{M} = \frac{\mu_1 KN}{NM} \leq \frac{\mu_1 Kc_s}{NM}.\end{aligned}$$

Then, we obtain (22). ∎

Similar to Lemma 3 in [46], we would then have that condition (20) holds with probability at least $1-(NM)^{-4}$ if $|\Omega| \geq c_1 \mu_1 c_s K \log(NM)$, where $c_1 \geq 0$ is a constant.

The remaining proof for Corollary 4.1 follows the corresponding proof steps for Theorem 1 in [46]. This yields Corollary 4.1, which is similar to Theorem 1 in [46] but with different incoherence properties (21).

To obtain these different incoherence properties, we bound the minimum nonzero singular value of $\mathbf{\Phi}$ and $\widetilde{\mathbf{A}}_{rf}$. It follows from Theorem 5 of [47] that

$$\sigma_{\min}(\widetilde{\mathbf{A}}_{rf}^H \widetilde{\mathbf{A}}_{rf}) = \sigma_{\min}^2(\widetilde{\mathbf{A}}_{rf}) = \frac{1}{M}\sigma_{\min}^2(\mathbf{A}_{rf}^v) \geq \frac{1}{M}\mathcal{L}(M,r,f),$$

where $\mathcal{L}(M,r,f)$ is defined as

$$\mathcal{L}(M,r,f) \triangleq \min_{1\leq k\leq K} \frac{1}{r_k}\left[\gamma_M(r_k) - \frac{c_2}{\Delta_f}(1+r_k^{2M})\right]$$

with c_2 being a constant and

$$\gamma_M(r_k) \triangleq \begin{cases} \frac{r_k^{2M}-1}{2\log(r_k)}, & r_k < 1, \\ M, & r_k = 1, \end{cases} \qquad k = 1, \ldots, K.$$

Note that $\mathcal{L}(M,r,f) = M - \frac{2c_2}{\Delta_f}$ is on the order of M when there is no damping (i.e., $r=1$) and the frequencies are well separated ($\Delta_f = O(\frac{1}{M})$). To satisfy the two assumptions in (21), we can let $\frac{1}{M}\mathcal{L}(M,r,f) \geq \frac{1}{\mu_1}$, and $\mu_1 \geq \frac{\mu_2}{\sigma_{\min}^2(\mathbf{\Phi})}$, that is, $\mu_1 \geq \max\left\{\frac{M}{\mathcal{L}(M,r,f)}, \frac{\mu_2}{\sigma_{\min}^2(\mathbf{\Phi})}\right\}$.

With some fundamental Lagrange analysis as in Section 4.1, we can also conclude that the optimal dual solution $\mathbf{Q}$ belongs to the subdifferential of $\|\widehat{\mathbf{X}}\|_*$, where $\widehat{\mathbf{X}}$ is the optimal solution of (16). In the case of exact recovery, i.e., $\widehat{\mathbf{X}} = \mathbf{X}^\star$, the dual solution $\mathbf{Q}$ has the form $\mathbf{Q} = \mathbf{U}_{\mathbf{X}^\star}\mathbf{V}_{\mathbf{X}^\star}^H + \mathbf{W}$, where $\mathbf{U}_{X^\star}^H \mathbf{W} = \mathbf{0}$, $\mathbf{W}\mathbf{V}_{X^\star} = \mathbf{0}$, and $\|\mathbf{W}\| \leq 1$. Then, we are still able to identify the true r_k's and f_k's by localizing the places where $\|\mathcal{Q}(r,f)\|_2 = \|\mathbf{Q}^H \boldsymbol{a}(r,f)\|_2$ achieves 1 with $\mathbf{Q}$ being the optimal dual solution of (18) and $\|\mathbf{W}\| < 1$ in the case when $\widehat{\mathbf{X}} = \mathbf{X}^\star$. Thus, we finish the proof of Corollary 4.1.

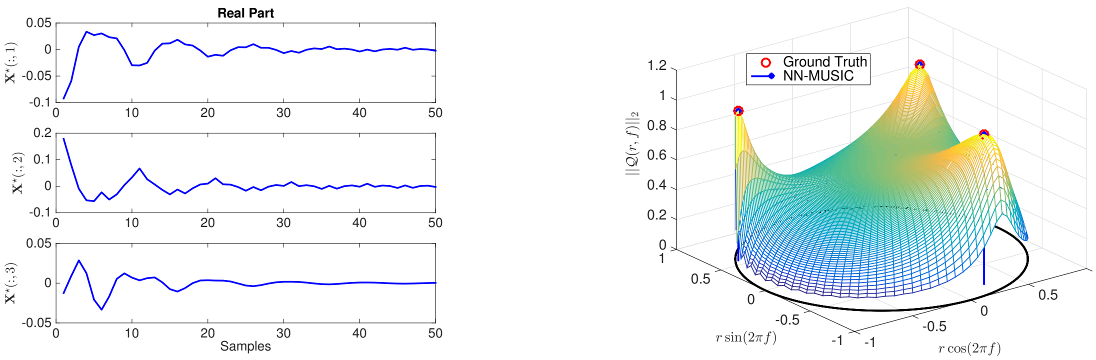
6 Numerical Simulations

6.1 Full data case

In this experiment, we use synthetic data to test the proposed Algorithm 2 with $K=3$. The true r_k's and f_k's are set as $r_1 = 0.92,\ r_2 = 0.98,\ r_3 = 0.85$ and $f_1 = 0.1,\ f_2 = 0.4,\ f_3 = 0.8$. We set $M = N = 50$. The data matrix $\mathbf{X}^\star$ is then generated as $\mathbf{X}^\star = \mathbf{A}_{rf}\mathbf{D}_{\widetilde{c}}\mathbf{\Phi}^\top$, where $\mathbf{A}_{rf}$, and $\mathbf{D}_{\widetilde{c}}$ are generated according to their definition in Section 2, $\mathbf{\Phi}$ is generated as a Gaussian random matrix with normalized columns, and the c_k's are set as K Gaussian random numbers with zero mean and unit variance. The first three columns of $\mathbf{X}^\star$ are shown in Figure 1 (a). Given the above data matrix $\mathbf{X}^\star$, we then use Algorithm 2 to identify all the r_k's and f_k's. Figure 1 (b) displays a surface plot of $\|\mathcal{Q}(r,f)\|_2$ and indicates that Algorithm 2 identifies all the r_k's and f_k's perfectly.

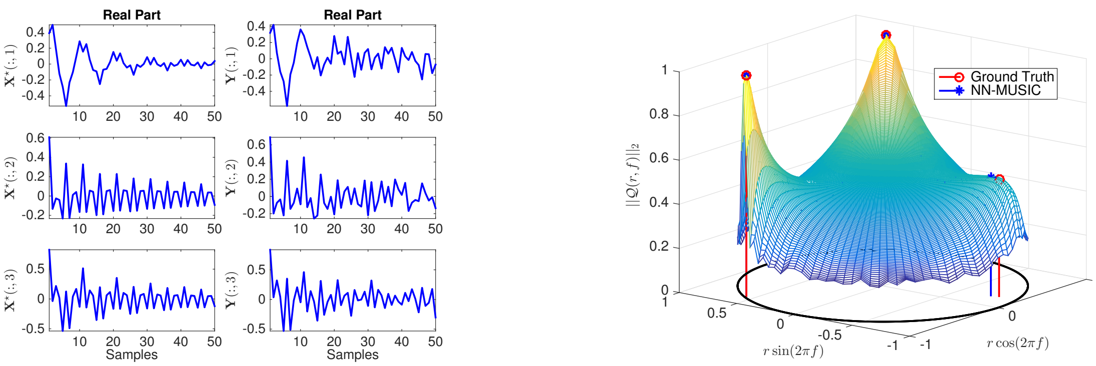
Next, as a demonstration, we repeat the above experiment but with additive white Gaussian noise with variance $\sigma = 0.1$ (SNR = 9.8806dB). The noiseless data and noisy data are shown in Figure 2 (a). We set the regularization parameter[9] as $\lambda = 0.2\sigma\sqrt{4MN\log(M)} = 3.9558$ and then solve the nuclear norm denoising program (15). As is shown in Figure 2 (b), we observe that the (r,f) pairs can still be estimated by localizing the peaks of $\|\mathcal{Q}(r,f)\|_2$. In particular, the estimated damping ratios and frequencies are given as $\widehat{r}_1 = 0.92,\ \widehat{r}_2 = 0.98,\ \widehat{r}_3 = 0.79$ and $\widehat{f}_1 = 0.1,\ \widehat{f}_2 = 0.4,\ \widehat{f}_3 = 0.8$. Note that we leave the corresponding theoretical guarantees for future work.

[9] Here, the regularization parameter is set according to [6].



(a) (b) 40% missing

Figure 1: Noiseless full data: (a) The first three columns in data matrix $\mathbf{X}^\star$. (b) The blue lines correspond to the locations where $\|\mathcal{Q}(r,f)\|_2$ achieves 1 while the red circles correspond to the true r_k's and f_k's. They coincide because the recovery is perfect.



(a) (b) 40% missing

Figure 2: Noisy full data: (a) Left: noiseless data $\mathbf{X}^\star$. Right: noisy data $\mathbf{Y}$. (b) The blue lines correspond to the peaks of $\|\mathcal{Q}(r,f)\|_2$ while the red lines mark the position of the true r_k's and f_k's.

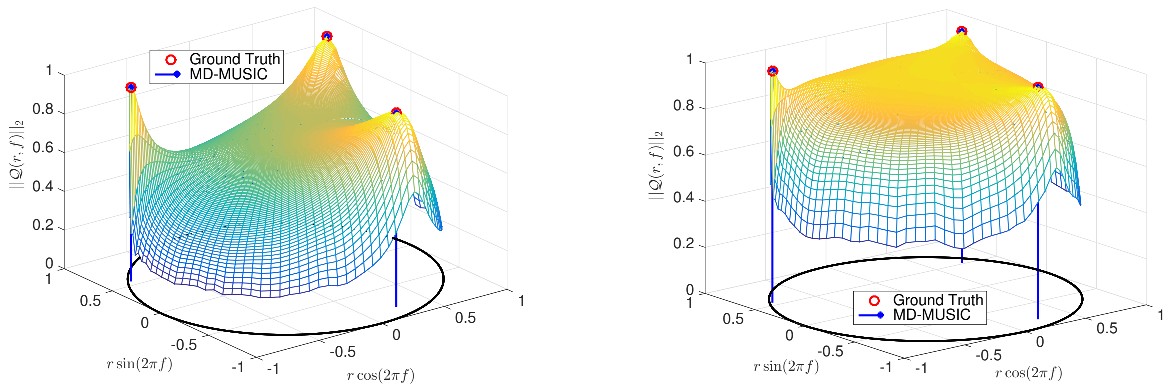
6.2 Missing data case

We repeat the above experiments with missing data, namely we identify the damping factors and frequencies from the given partially observed data matrix $\mathbf{X}^\star_\Omega$ by solving the NNM problem in (16).[10] All parameters are set same as in the noise-free setting of Section 6.1. After generating the full data matrix $\mathbf{X}^\star$, we randomly remove 20% and 40% of its entries. We then use Algorithm 3 to identify all the r_k's and f_k's from the partial data. Figure 3 indicates that Algorithm 3 identifies all the r_k's and f_k's perfectly.

We notice that the data matrix $\mathbf{X}^\star$ is also well recovered in this case. In particular, we define the relative recovery error of data matrix as $\text{RelErr} \triangleq \frac{\|\widehat{\mathbf{X}}-\mathbf{X}^\star\|_F}{\|\mathbf{X}^\star\|_F}$, where $\mathbf{X}^\star$ and $\widehat{\mathbf{X}}$ denote the true full data matrix and the recovered data matrix via NNM. In particular, we have $\text{RelErr} = 4.7487 \times 10^{-10}$ when 20% of the data is missing and $\text{RelErr} = 3.5190 \times 10^{-8}$ when 40% of the data is missing. Moreover, as is shown in Figure 4, we also observe that in some cases, the r_k's and f_k's can be perfectly recovered even if we do not perfectly recover $\mathbf{X}^\star$, which further supports our Theorem 4.2.

Finally, we investigate the minimal number of measurements needed for perfect recovery with various numbers K of spectral components. We set $M = 70$ and $N = 50$. For each value of K, we randomly pick K frequencies and damping ratios from a frequency set $\mathcal{F} = 0.05 : 0.05 : 0.95$ and a damping ratio

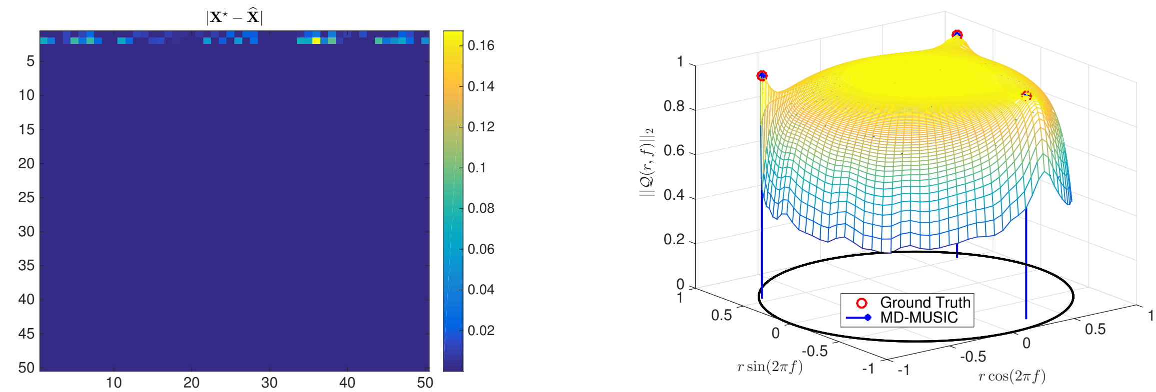
[10] Note that CVX [?] can return the estimated data matrix $\widehat{\mathbf{X}}$ as well as the dual solution $\mathbf{Q}$ by solving the SDP form in (17).



(a) 20% missing

(b) 40% missing

Figure 3: Noiseless missing data: the blue lines correspond to the locations where $\|\mathcal{Q}(r, f)\|_2$ achieves 1 while the red circles correspond to the true r_k's and f_k's. They coincide because the recovery is perfect.



(a) data reconstruction error

(b) 40% missing

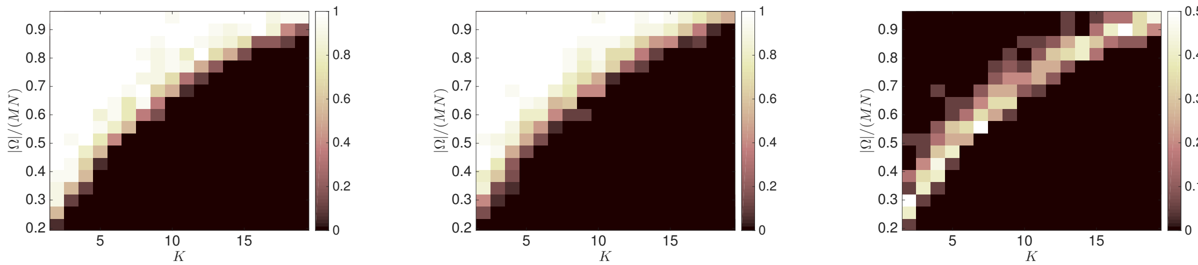
Figure 4: Noiseless missing data: in some cases, the r_k's and f_k's can be perfectly recovered even when $\mathbf{X}^\star$ is not. In this example, we have RelErr $= 0.0412$.

set $\mathcal{R} = 0.94 : 0.0025 : 1$.[11] Denote $(\widehat{r}, \widehat{f})$ and $(r^\star, f^\star)$ as the recovered parameters and true parameters, respectively. We consider the parameter recovery to be a success if

$$\max_{1\le k\le K}(|\widehat{r}_k - r_k^\star|) \le 10^{-5}, \quad \max_{1\le k\le K}(|\widehat{f}_k - f_k^\star|) \le 10^{-5}. \tag{23}$$

Similarly, we consider the data matrix recovery to be a success if the relative recovery error $\frac{\|\widehat{\mathbf{X}} - \mathbf{X}^\star\|_F}{\|\mathbf{X}^\star\|_F} \le 10^{-5}$. We perform 20 trials in this part of simulation. It can be seen in Figure 5 that the minimal number of measurements needed for perfect data matrix recovery does scale roughly linearly with K, as indicated in Corollary 4.1. We also notice a similar behavior appearing in parameters recovery. Figure 5 (c) again indicates that we can still successfully recover the parameters in some cases where the data matrix is not perfectly recovered.

[11] We choose 0.94 as the lowest damping ratio since we want to keep at least 1% energy at the end of uniform sampling. Therefore, we have $r = 0.01^{\frac{1}{M}} \le 0.94$.



(a) (r, f) (b) $\mathbf{X}^\star$ (c) probability difference

Figure 5: Probability of successful recovery for (a) parameters (r, f) and (b) data matrix $\mathbf{X}^\star$. (c) presents the probability of successful recovery for parameters minus the probability of successful recovery for the data matrix.

6.3 Data coherence

In this section, we conduct three numerical experiments to examine the influence of the minimum frequency separation Δ_f, the matrix $\mathbf{\Phi}$, and the damping ratio r_k on the performance of missing data matrix recovery. The standard literature on matrix completion [43] relates the recoverability of a matrix $\mathbf{X}^\star$ to its *coherence*, defined as $\mu_0^\star \triangleq \max\{\mu_1^\star(\mathbf{X}^\star), \mu_2^\star(\mathbf{X}^\star)\}$ with

$$\mu_1^\star(\mathbf{X}^\star) \triangleq \frac{M}{K} \max_{1\le m\le M} \left\|\mathbf{U}_{\mathbf{X}^\star}^\top \boldsymbol{e}_m^c\right\|_2^2, \quad \mu_2^\star(\mathbf{X}^\star) \triangleq \frac{N}{K} \max_{1\le n\le N} \left\|\mathbf{V}_{\mathbf{X}^\star}^\top \boldsymbol{e}_n^c\right\|_2^2,$$

where $\mathbf{X}^\star = \mathbf{U}_{X^\star}\mathbf{S}_{X^\star}\mathbf{V}_{X^\star}^H$ is a truncated SVD of $\mathbf{X}^\star$, and $\boldsymbol{e}_m^c \in \mathbb{R}^M$ and $\boldsymbol{e}_n^c \in \mathbb{R}^N$ denote canonical basis vectors.

In the first experiment, we examine the influence of minimum frequency separation on the performance of missing data recovery with $M = 50$, $N = 30$, $K = 2$ and $|\Omega| = 450$, i.e., 70% of the data are missing. To simplify the experiment, we set $r_1 = r_2 = 1$ and $c_1 = c_2 = 1$. We fix $f_1 = 0.1$ and let $f_2 = f_1 + \Delta_f$ with various values of the minimum frequency separation Δ_f. We generate $\mathbf{\Phi} \in \mathbb{C}^{N\times K}$ using normalized columns from a discrete Fourier matrix, which implies $\mu_2^\star = 1$ and ensures that $\mu_0^\star = \mu_1^\star$. 10^4 trials are performed in this experiment. Other settings are the same as in Section 6.2. It is shown in Figure 6 (a, b) that the coherence parameter $\mu_0^\star$ decreases as the minimum frequency separation Δ_f increases, which also explains why the probability of successful data matrix recovery increases as Δ_f increases.

In the second experiment, we examine the influence of the matrix $\mathbf{\Phi}$ on the performance of missing data recovery. We change $N = 10$ to make sure that the coherence parameter $\mu_0^\star$ is not too large. By fixing $\Delta_f = 1/M$, we have $\mu_1^\star = 1$ and thus $\mu_0^\star = \mu_2^\star$. We again generate $\mathbf{\Phi} \in \mathbb{C}^{N\times K}$ using columns from the discrete Fourier matrix, but we then replace its first entry $\phi_{1,1}$ with scalars in the range of $[1, 10]$ and then normalize its columns. Other settings are same as the first experiment. We conduct 500 trials in this experiment. Figure 6 (c, d) shows that the coherence parameter $\mu_0^\star$ increases as $\phi_{1,1}$ increases, which also explains why the probability of successful data matrix recovery decreases as $\phi_{1,1}$ increases.

In the third experiment, we examine the influence of the damping ratio on the performance of missing data recovery. We fix $\Delta_f = 1/M$ and repeat the first experiment with various values of r_2. We conduct 100 trials in this experiment. As shown in Figure 6 (e, f), the coherence parameter $\mu_0^\star$ decreases as r_2 increases, which also explains why the probability of successful data matrix recovery increases as r_2 increases. This is to be expected since exponentials with smaller damping ratio are transient and their contribution tends to fade quickly from the measured data. Moreover, Figure 6 (f) again indicates that we can successfully recover the parameters in some cases where the data matrix is not perfectly recovered.

These numerical experiments give a sense of how spectral parameters influence the coherence, and thus, recoverability of the data matrix. We stress again, however, that the significance of Corollary 4.1 is that the sample complexity is not stated in terms of the matrix coherence (which may be difficult to immediately relate to the more tangible signal parameters); rather, the dependence on the damping ratios and minimum frequency separation is explicitly revealed in Corollary 4.1.

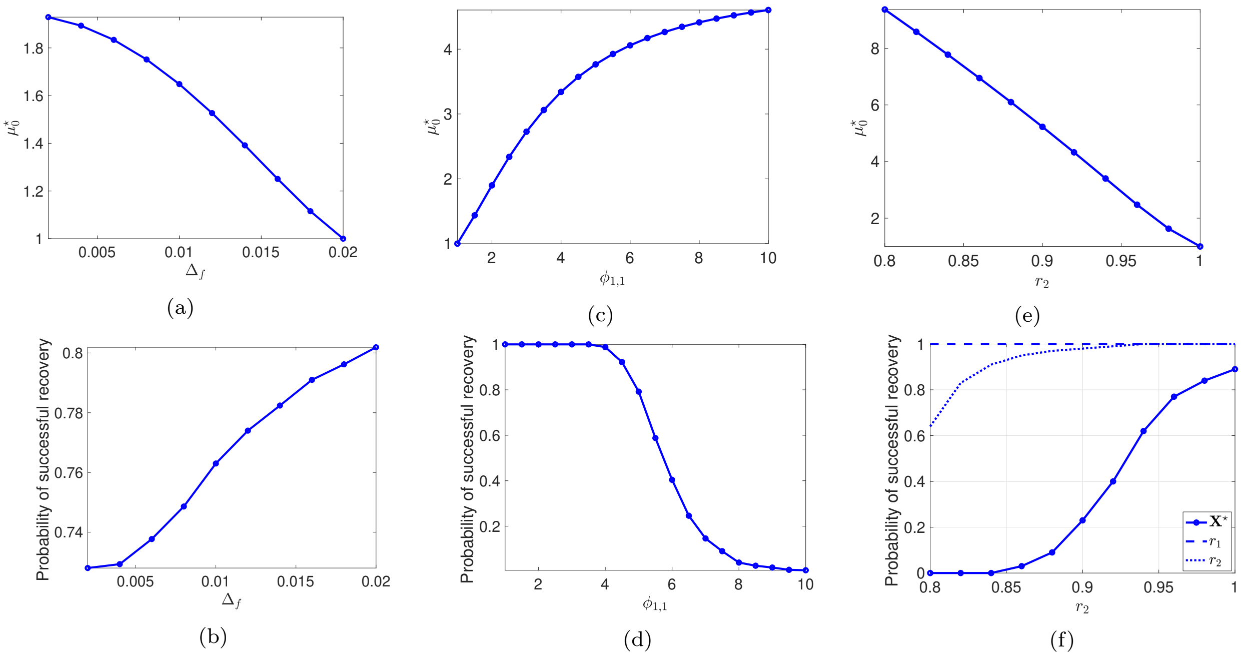


Figure 6: Influence of minimum frequency separation Δ_f, the matrix $\mathbf{\Phi}$, and damping ratio r_2 on the performance of missing data recovery: (a) coherence parameter of the data matrix $\mathbf{X}^\star$ and (b) probability of successful data matrix recovery with respect to the minimum frequency separation Δ_f. (c) coherence parameter of the data matrix $\mathbf{X}^\star$ and (d) probability of successful data matrix recovery with respect to a various of matrix $\mathbf{\Phi}$. (e) coherence parameter of the data matrix $\mathbf{X}^\star$ and (f) probability of successful data matrix and damping ratio recovery with respect to the second damping ratio r_2.

6.4 Comparison with existing algorithms

In this section, we implement a series of experiments to compare our proposed algorithms with three existing methods: 1) NNM+MUSIC/ESPRIT, 2) MN-MUSIC, and 3) ANM. We define successful parameter recovery as in (23).

6.4.1 NNM+MUSIC and NNM+ESPRIT

We use NNM+MUSIC to denote an alternative approach wherein one first solves the NNM problem in (17) to get $\widehat{\mathbf{X}}$ and then uses Algorithm 2 (or, equivalently, MUSIC) to identify the r_k's and f_k's from $\widehat{\mathbf{X}}$.[12] We use NNM+ESPRIT to denote a similar two-stage approach but with MUSIC replaced by the Estimation of Signal Parameters via Rotation Invariance Techniques (ESPRIT) algorithm. The true r_k's and f_k's are set same as in Section 6.1. We set $M = 50$ and $N = 20$. To show the advantage of our MD-MUSIC over NNM+MUSIC and NNM+ESPRIT, we present the probability of successful parameter recovery (defined in (23)) in Table 1. In the "two-step" algorithms NNM+MUSIC and NNM+ESPRIT, we use the true K as the number of frequencies when we implement the MUSIC or ESPRIT algorithm even though it is unknown and needs to estimated in practice.

6.4.2 MN-MUSIC

Next, we compare our proposed MD-MUSIC algorithm with the MN-MUSIC algorithm introduced in Section 3.3 [19] in a scenario where 20% of the noiseless data entries are missing. We observe that the MN-MUSIC

[12] A similar idea has also been considered in [41].

Table 1: Comparison of the "one-step" MD-MUSIC algorithm and the "two-step" NNM+MUSIC and NNM+ESPRIT algorithms for parameter recovery. We present the probability of successful recovery over 1000 trials.

	10% missing	20% missing	30% missing	40% missing
MD-MUSIC	99.6%	96.9%	90.8%	78.8%
NNM+MUSIC	99.4%	96.3%	85.0%	56.2%
NNM+ESPRIT	98.8%	89.9%	66.6%	29.5%

algorithm never successfully recovers the frequencies and damping ratios since it performs an SVD directly on the missing data.[13]

6.4.3 ANM

Finally, we compare the proposed NN-MUSIC and MD-MUSIC algorithms with ANM in the full and missing data cases, respectively. In ANM, we solve the following SDP

$$\min_{\mathbf{X},\boldsymbol{u},\mathbf{D}} \frac{1}{2M}\operatorname{Tr}(\operatorname{Toep}(\boldsymbol{u})) + \frac{1}{2}\operatorname{Tr}(\mathbf{D}) \quad \text{s.t.} \begin{bmatrix} \operatorname{Toep}(\boldsymbol{u}) & \mathbf{X} \\ \mathbf{X}^H & \mathbf{D} \end{bmatrix} \succeq 0,\ \mathbf{X}_\Omega = \mathbf{X}^\star_\Omega,$$

where $\operatorname{Toep}(\boldsymbol{u})$ is a Hermitian Toeplitz matrix with the vector $\boldsymbol{u}$ being its first column. We use $\mathbf{X} = \mathbf{X}^\star$ instead of $\mathbf{X}_\Omega = \mathbf{X}^\star_\Omega$ in the full data case. Similar to NN-MUSIC and MD-MUSIC, given the dual solution of the above SDP, we then formulate a dual polynomial and localize the places where the ℓ_2-norm of the dual polynomial achieves 1 to extract the estimated frequencies. Since ANM can only recover frequencies, we only compare the accuracy of estimated frequencies in this section. All the simulation results presented in this section are an average over 100 trials.

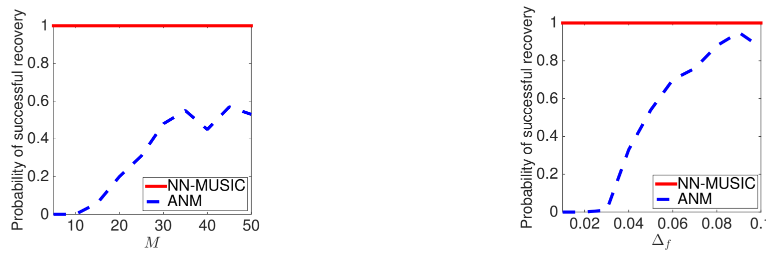
In the full data case, we repeat the first experiment in Section 6.1 with $N = 10$ and with a variety of M and Δ_f. We firstly fix $\Delta_f = 0.06$ and set the true frequency and damping pairs as $(r_1, f_1) = (0.86, 0.1)$, $(r_2, f_2) = (0.92, 0.16)$, and $(r_3, f_3) = (0.98, 0.8)$. Then, we compare NN-MUSIC and ANM with a variety of M. Next, we fix $M = 20$, $r_1 = 0.92$, $r_2 = 0.98$, and $f_1 = 0.1$. Similar as in Section 6.3, we then let $f_2 = f_1 + \Delta_f$ with various values of Δ_f. The simulation results are given in Figure 7. It can be seen that the NN-MUSIC algorithm significantly outperforms ANM and can always recover the frequencies exactly, as indicated in Theorem 4.1. This is because our data contains damping, which is not modeled in ANM.

In the missing data case, we randomly remove 20% or 40% of the data entries. We repeat the above two experiments with these partially observed data matrices to compare MD-MUSIC and ANM. As shown in Figure 8, MD-MUSIC still outperforms ANM significantly in most cases due to its ability to handle damped signals. We also observe that ANM can have a higher probability of successful recovery once the number of observed entries is too small, as shown in Figure 8 (c). However, the success probability in this case is still significantly less than 1. Note that we have changed f_2 from 0.16 to 0.2 in Figure 8 (a, c) to test with a larger value of Δ_f. Other parameters used in this part are the same as in the full data experiments.

7 Conclusion

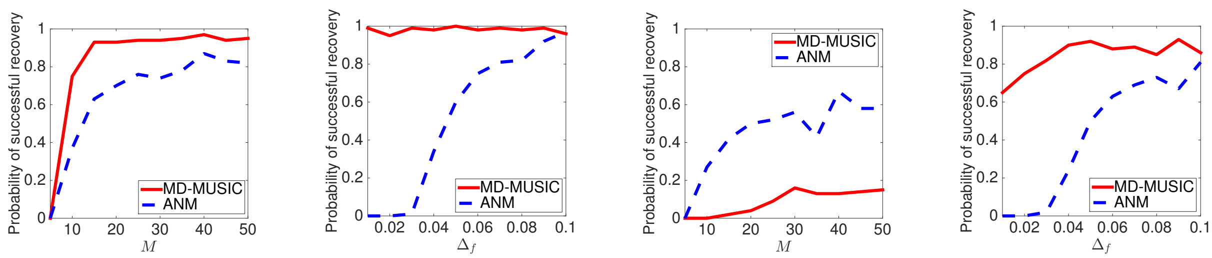
In this work, we provide a convex optimization view for the classical MUSIC algorithm in spectral estimation with damping. In particular, we build a connection between NNM and the classical MUSIC algorithm, which inspires us to propose a new algorithm, named MD-MUSIC, for the missing data field. Theoretical results are provided to guarantee the proposed algorithms. In particular, it is possible to get exact parameter recovery with the MD-MUSIC algorithm even when we do not have perfect data recovery. Moreover, for

[13]No results are shown since MN-MUSIC never recovers successfully.



(a) (b)

Figure 7: Comparison of NN-MUSIC and ANM in the noiseless full data case, with damped exponentials. (a) Probability of successful frequency recovery as a function of M, with fixed $\Delta_f = 0.06$ and $K = 3$. (b) Probability of successful frequency recovery as a function of Δ_f, with fixed $M = 20$ and $K = 2$.



(a) 20% missing (b) 20% missing (c) 40% missing (d) 40% missing

Figure 8: Comparison of MD-MUSIC and ANM in the noiseless missing data case and damped exponentials, with (a, b) 20% or (c, d) 40% of the entries randomly removed. (a), (c) Probability of successful frequency recovery as a function of M, with fixed $\Delta_f = 0.1$ and $K = 3$. (b), (d) Probability of successful frequency recovery as a function of Δ_f, with fixed $M = 20$ and $K = 2$.

the missing data case, we also quantify how the sample complexity depends on the true spectral parameters rather than use certain incoherence properties as in existing literature. Meanwhile, numerical simulations indicate that the proposed algorithms work very well and significantly outperform some relevant existing methods in frequency estimation of damped exponentials. We leave the robust performance analysis on noisy data for future work.

Acknowledgement

MW and SL were supported by NSF grant CCF–1409258, NSF CAREER grant CCF–1149225, and NSF grant CCF-1704204.

References

[1] S. Umesh and D. W. Tufts, "Estimation of parameters of exponentially damped sinusoids using fast maximum likelihood estimation with application to nmr spectroscopy data," *IEEE Transactions on Signal Processing*, vol. 44, no. 9, pp. 2245–2259, 1996.

[2] X. Qu, M. Mayzel, J.-F. Cai, Z. Chen, and V. Orekhov, "Accelerated nmr spectroscopy with low-rank reconstruction," *Angewandte Chemie International Edition*, vol. 54, no. 3, pp. 852–854, 2015.

[3] Y. Xie, S. Li, G. Tang, and M. B. Wakin, "Radar signal demixing via convex optimization," in *2017 22nd International Conference on Digital Signal Processing (DSP)*, pp. 1–5, IEEE, 2017.

[4] Z. Zhu and M. B. Wakin, "On the dimensionality of wall and target return subspaces in through-the-wall radar imaging," in *The 4th International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa)*, pp. 110–114, IEEE, 2016.

[5] J. Y. Park, M. B. Wakin, and A. C. Gilbert, "Sampling considerations for modal analysis with damping," in *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2015*, vol. 9435, p. 94350U, International Society for Optics and Photonics, 2015.

[6] S. Li, D. Yang, G. Tang, and M. B. Wakin, "Atomic norm minimization for modal analysis from random and compressed samples," *IEEE Transactions on Signal Processing*, vol. 66, no. 7, pp. 1817–1831, 2018.

[7] Y. Xie, D. Liu, H. Mansour, and P. T. Boufounos, "Robust parameter estimation of contaminated damped exponentials," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5500–5504, IEEE, 2020.

[8] R. de Prony, "Essai expérimental et analytique sur les lois de la dilabilité et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alkool,a différentes températures," *J. de l'Ecole Polytechnique*, vol. 1, no. 22, pp. 24–76, 1795.

[9] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 5, pp. 814–824, 1990.

[10] R. Schmidt, *A signal subspace approach to multiple emitter location and spectral estimation.* Stanford University, 1981.

[11] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[12] F. Marvasti, *Nonuniform sampling: theory and practice.* Springer Science & Business Media, 2012.

[13] Y. Li and Y. Chi, "Off-the-grid line spectrum denoising and estimation with multiple measurement vectors," *IEEE Transactions on Signal Processing*, vol. 64, no. 5, pp. 1257–1269, 2016.

[14] Z. Yang and L. Xie, "Exact joint sparse frequency recovery via optimization methods," *IEEE Transactions on Signal Processing*, vol. 64, no. 19, pp. 5145–5157, 2014.

[15] J. Y. Park, M. B. Wakin, and A. C. Gilbert, "Modal analysis with compressive measurements," *IEEE Transactions on Signal Processing*, vol. 62, no. 7, pp. 1655–1670, 2014.

[16] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 1055–1096, 1982.

[17] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, 1996.

[18] W. Liao and A. Fannjiang, "Music for single-snapshot spectral estimation: Stability and super-resolution," *Applied and Computational Harmonic Analysis*, vol. 40, no. 1, pp. 33–67, 2016.

[19] R. T. Suryaprakash and R. R. Nadakuditi, "The performance of music-based doa in white noise with missing data," in *2012 IEEE Statistical Signal Processing Workshop (SSP)*, pp. 800–803, IEEE, 2012.

[20] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, p. 717, 2009.

[21] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[22] J. Gillard and A. Zhigljavsky, "Optimization challenges in the structured low rank approximation problem," *Journal of Global Optimization*, vol. 57, no. 3, pp. 733–751, 2013.

[23] V. Larsson and C. Olsson, "Convex low rank approximation," *International Journal of Computer Vision*, vol. 120, no. 2, pp. 194–214, 2016.

[24] M. Fazel, H. Hindi, and S. P. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *American Control Conference, 2001. Proceedings of the 2001*, vol. 6, pp. 4734–4739, IEEE, 2001.

[25] M. Fazel, T. K. Pong, D. Sun, and P. Tseng, "Hankel matrix rank minimization with applications to system identification and realization," *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 3, pp. 946–977, 2013.

[26] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2862–2869, 2014.

[27] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 4729–4743, 2014.

[28] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.

[29] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Foundations of Computational mathematics*, vol. 12, no. 6, pp. 805–849, 2012.

[30] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE transactions on information theory*, vol. 59, no. 11, pp. 7465–7490, 2013.

[31] S. Li, M. B. Wakin, and G. Tang, "Atomic norm denoising for complex exponentials with unknown waveform modulations," *IEEE Transactions on Information Theory*, vol. 66, no. 6, pp. 3893–3913, 2020.

[32] Y. Li, K. Liu, and J. Razavilar, "Improved parameter estimation schemes for damped sinusoidal signals," *Electrical Engineering Department and Institute for Systems Research, University of Maryland at College Park*, 1999.

[33] P. Stoica and R. L. Moses, *Introduction to spectral analysis*, vol. 1. Prentice hall Upper Saddle River, NJ, 1997.

[34] H. H. Yang and Y. Hua, "On rank of block hankel matrix for 2-d frequency detection and estimation," *IEEE Transactions on Signal Processing*, vol. 44, no. 4, pp. 1046–1048, 1996.

[35] L. L. Scharf and B. Friedlander, "Toeplitz and hankel kernels for estimating time-varying spectra of discrete-time random processes," *IEEE Transactions on Signal Processing*, vol. 49, no. 1, pp. 179–189, 2001.

[36] F. Andersson, M. Carlsson, J.-Y. Tourneret, and H. Wendt, "A new frequency estimation method for equally and unequally spaced data," *IEEE Transactions on Signal Processing*, vol. 62, no. 21, pp. 5761–5774, 2014.

[37] F. Andersson and M. Carlsson, "Fixed-point algorithms for frequency estimation and structured low rank approximation," *Applied and Computational Harmonic Analysis*, vol. 46, no. 1, pp. 40–65, 2019.

[38] W. Liao, "Music for joint frequency estimation: stability with compressive measurements," in *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 414–418, IEEE, 2014.

[39] M.-D. Iordache, J. M. Bioucas-Dias, A. Plaza, and B. Somers, "Music-csr: Hyperspectral unmixing via multiple signal classification and collaborative sparse regression," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 7, pp. 4364–4382, 2014.

[40] J. M. Kim, O. K. Lee, and J. C. Ye, "Compressive music: Revisiting the link between compressive sensing and array signal processing," *IEEE Transactions on Information Theory*, vol. 58, no. 1, pp. 278–301, 2012.

[41] W. Xu, J. Yi, S. Dasgupta, J.-F. Cai, M. Jacob, and M. Cho, "Separation-free super-resolution from compressed measurements is possible: an orthonormal atomic norm minimization approach," in *2018 IEEE International Symposium on Information Theory (ISIT)*, pp. 76–80, IEEE, 2018.

[42] Q. Li and G. Tang, "Approximate support recovery of atomic line spectral estimation: A tale of resolution and precision," *Applied and Computational Harmonic Analysis*, 2018.

[43] Y. Chen, S. Bhojanapalli, S. Sanghavi, and R. Ward, "Completing any low-rank matrix, provably," *The Journal of Machine Learning Research*, vol. 16, no. 1, pp. 2999–3034, 2015.

[44] A. Klinger, P. Joseph, R. von Randow, C. Brookes, C. Green, J. Adams, L. W. Cohen, L. Bers, G. Albert, R. Tittle, *et al.*, "Classroom notes," *The American Mathematical Monthly*, vol. 74, no. 5, pp. 571–588, 1967.

[45] D. Kalman, "The generalized vandermonde matrix," *Mathematics Magazine*, vol. 57, no. 1, pp. 15–21, 1984.

[46] Y. Chen and Y. Chi, "Robust spectral compressed sensing via structured matrix completion," *IEEE Transactions on Information Theory*, vol. 60, no. 10, pp. 6576–6601, 2014.

[47] C. Aubel and H. Bölcskei, "Vandermonde matrices with nodes in the unit disk and the large sieve," *Applied and Computational Harmonic Analysis*, 2017.