An Optimization View of MUSIC and Its Extension to Missing Data

Shuang Li, Hassan Mansour, and Michael B. Wakin

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. In this work, we offer a novel optimization-based perspective on the classical MUSIC algorithm that could lead to future developments and understanding. In particular, we propose an algorithm for spectral estimation that involves searching for the peaks of the dual polynomial corresponding to a certain nuclear norm minimization (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. Simulation results also indicate that the proposed algorithms significantly outperform some relevant existing methods in frequency estimation of damped exponentials.

Index Terms—Spectral estimation, MUSIC algorithm, nuclear norm minimization, optimization, low-rank matrix completion.

I. 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 imaging [3], [4] and modal analysis [5], [6]. It is well known that the frequencies and damping factors can be identified by the classical spectrum estimation approaches, such as Prony’s method [7], the Matrix Pencil method [8], and the Multiple Signal Classification (MUSIC) algorithm [9], [10], 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 appealing alternative [11] and results in the partial observations (missing data) discussed in this work.

The MUSIC algorithm, which is widely used in signal processing [12], [13], was first proposed by Schmidt as an improvement to Pisarenko’s method [9]. 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 [14]. 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 [15]. 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 [16] 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 [17], [18], low-rank approximation [19], [20] system identification [21], [22], and image denoising [23], [24]. 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 [17]. 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 [25], [26]. 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) considered in many existing works can only handle frequency estimation in undamped sinusoids [25], [27], [15], [28].

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 believe this offers a novel optimization-based perspective on the MUSIC algorithm that could lead to future developments and understanding. 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. 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 recover the spectral parameters with high probability by searching for the locations in the damping-frequency plane where the \( \ell_2 \)-norm of the dual polynomial achieves 1, as long as the number of uniform samples is larger than the number of spectral parameters. For MD-MUSIC, our theorem shows that we can perfectly identify the spectral parameters with high probability by searching for the locations in the damping-frequency plane where the \( \ell_2 \)-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. Moreover, we quantify how the sample complexity depends on the true spectral parameters.

The remainder of this paper is organized as follows. In Section II, we introduce both the SMV and MMV settings considered in this paper. In Section III, we review the classical MUSIC algorithm as well as its variants. In Section IV, 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 V. In Section VI, 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 VII.

II. 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 “\(^t\)” to denote row vectors, and superscripts “\(^T\)” and “\(^H\)” to denote transpose and conjugate transpose, respectively.

A. Single Measurement Vector (SMV) setting

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

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

where \(\{c_k\}, \{r_k\}, \{f_k\}\) and \(e(t)\) are the unknown 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 coefficients \(c_k > 0\), and \(e(t) \sim \mathcal{CN}(0, \sigma^2)\).

B. Multiple Measurement Vector (MMV) setting

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

\[
y^o(t) = x^o(t) + e^o(t), \quad x^o(t) = \sum_{k=1}^{K} c_k r_k^t e^{j2\pi f_k t} \phi_k^T, \quad (2)
\]

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

Suppose we take \(M\) uniform samples and arrange \(y_m^o = y^o(m)\) as the \(m\)-th row of a data matrix \(Y \in \mathbb{C}^{M \times N}\). Define \(X^* \triangleq [(x_0^o)^T (x_1^o)^T \cdots (x_{M-1}^o)^T]^T\) and \(E \triangleq [(e_0^o)^T (e_1^o)^T \cdots (e_{M-1}^o)^T]^T\) as the noiseless data matrix and observation noise matrix, respectively. Then, we have

\[
Y = [(y_0^o)^T (y_1^o)^T \cdots (y_{M-1}^o)^T]^T
= \sum_{k=1}^{K} c_k a(r_k, f_k) \phi_k^T + E
= A_x f D \Phi^T + E
= X^* + E
\]

with

\[
a(r, f) \triangleq \begin{cases} \sqrt{\frac{1-r^2}{1-r^2\pi^2}} [1 e^{j2\pi f 1} \ldots e^{j2\pi f (M-1)}]^T, & r < 1, \\ \frac{1}{\sqrt{M}} [1 e^{j2\pi f 1} \ldots e^{j2\pi f (M-1)}]^T, & r = 1, \end{cases}
\]

1Note that we abbreviate \(a(r, f)\) to \(a(f)\) when \(r = 1\).
which has the autocorrelation matrix posing an estimate of the autocorrelation matrix known frequencies $R$ to approximate

Then, the following sample autocorrelation matrix can be used as

\[ y_n = x_n + e_n = \sum_{k=1}^{K} \bar{c}_{n,k} a(r_k, f_k) + e_n, \quad n = 1, \ldots, N, \quad (5) \]

where $\bar{c}_{n,k} = \bar{c}_k \phi_{n,k}$ with $\phi_{n,k}$ being the $(n,k)$-th entry of $\Phi$. 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.

III. PRIOR WORK

In this section, we review the classical MUSIC algorithm [9], [10] as well as its two variants, Damped MUSIC (DMUSIC) [29] and MUSIC adapted to missing data with Gaussian white noise (denoted as MN-MUSIC) [16].

A. MUltiple SIgnal Classification (MUSIC) algorithm

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 $y(t) \in \mathbb{C}^M$ as

\[ y(t) \triangleq [y(t) \ y(t+1) \ \cdots \ y(t+M-1)]^\top, \quad (6) \]

which has the autocorrelation matrix

\[ R_y \triangleq \mathbb{E}\{y(t)y(t)\}^\top \]. \]

The classical MUSIC algorithm aims to identify the unknown frequencies \{\textit{f}_k\} by constructing (and then decomposing) an estimate of the autocorrelation matrix $R_y$ without damping, namely, in the case where all $r_k = 1$ [30]. This requires $M > K$. Specifically, consider a full set of uniform observations \{\textit{y}(t)\} with $t = 0, 1, \ldots, L-1$, for some $L > M$. Then, the following sample autocorrelation matrix can be used to approximate $R_y$:

\[ \hat{R}_y = \frac{1}{L - M + 1} \sum_{t=0}^{L-M} y(t)y(t)^\top. \quad (7) \]

Let $[\hat{U}_s \ \hat{U}_n]$ denote the orthonormal eigenvectors of $\hat{R}_y$. In particular, suppose $\hat{U}_s \in \mathbb{C}^{M \times K}$ (signal space) and $\hat{U}_n \in \mathbb{C}^{M \times (M-K)}$ (noise space) are associated with the $K$ largest eigenvalues and the $M-K$ smallest eigenvalues of $\hat{R}_y$, 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 $\hat{R}_y$ as in (7) and its eigenvectors $\hat{U}_n$
3: compute the pseudospectrum: $1/\|\hat{U}_n^\top a(f)\|_2^2$, where $a(f)$ is defined in (4) with $r = 1$
4: localize the $K$ largest local maxima of pseudospectrum to get $\hat{f}_k$
5: return $\hat{f}_k$
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 $y(t)$ in (6) can be expressed as

\[ y(t) = \sum_{k=1}^{K} c_k e^{j2\pi f_k t} a(f_k) + e(t) = A_f c(t) + e(t) \]

with

\[ A_f = [a(f_1), a(f_2), \cdots, a(f_K)], \]

\[ c(t) = \sqrt{M} [c_1 e^{j2\pi f_1 t}, c_2 e^{j2\pi f_2 t}, \cdots, c_K e^{j2\pi f_K t}]^\top, \]

\[ e(t) = [e(t), e(t+1), \cdots, e(t+M-1)]^\top, \]

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

\[ R_y = \mathbb{E}\{y(t)y(t)^\top\} = A_f R_c A_f^\top + \sigma^2 I_M \]

if $c(t)$ is uncorrelated with $e(t)$. Here, $R_c \triangleq \mathbb{E}\{c(t)c(t)^\top\}$ is the autocorrelation matrix of $c(t)$ and $I_M$ denotes the $M \times M$ identity matrix. Note that the coefficients \{\textit{c}_k\} may be uncorrelated ($R_c$ is diagonal) or may contain completely correlated pairs ($R_c$ is singular). We are interested in the first case, namely, $R_c$ is diagonal, and positive definite since $c_k > 0$, for $k = 1, \ldots, K$. On the other hand, the rank of $A_f$ is $K$ when all the frequencies \{\textit{f}_k\} are distinct and $M \geq K$. It follows that the rank of $A_f R_c A_f^\top$ is $K$. Let \{\textit{\lambda}_m\}, $m = 1, \ldots, M$ denote the non-increasing eigenvalues of $A_f R_c A_f^\top$. Then, we have

\[ \lambda_{K+1} = \cdots = \lambda_M = 0. \]

As a consequence, the determinant of $A_f R_c A_f^\top$ is

\[ \det(A_f R_c A_f^\top) = \det(R_y - \sigma^2 I_M) = 0, \]

which implies that

\[ \lambda_m^\prime = \sigma^2, \quad m = K + 1, \ldots, M, \]

where \textit{\lambda}_m^\prime is the \textit{m}-th non-increasing eigenvalue of $R_y$.Denoting \textit{u}_m as the \textit{m}-th eigenvector of $R_y$ corresponding to eigenvalue \textit{\lambda}_m^\prime, we have

\[ R_y u_m = \lambda_m^\prime u_m, \quad m = 1, \ldots, M. \quad (8) \]

As is stated in [10], in general, $R_c$ will be “merely” positive definite to reflect the arbitrary degrees of pair-wise correlations occurring between the coefficients.
Replacing $\mathbf{R}_m = \mathbf{A}_f \mathbf{R_c} \mathbf{A}_f^H + \sigma^2 \mathbf{I}_M$ into the above equation (8), we have

$$\mathbf{A}_f \mathbf{R_c} \mathbf{A}_f^H \mathbf{u}_m = (\lambda_m^H - \sigma^2) \mathbf{u}_m = 0 \quad \text{or} \quad \mathbf{A}_f^H \mathbf{u}_m = 0$$

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

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 [14]. 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

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

(9)

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

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

with

$$\mathbf{a}(f) = \frac{1}{\sqrt{M}} \left[ e^{j2\pi f_1} \cdots e^{j2\pi (M-1)f} \right]^T$$

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 $\mathbf{R}(f)$ 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

$$\hat{\mathbf{R}}_y = \frac{1}{L - M + 1} \mathcal{H}_y \mathcal{H}_y^H.$$

Thus, the eigenvectors of $\hat{\mathbf{R}}_y$ 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) MMV MUSIC via data matrix: The MUSIC algorithm is also widely used in MMV problems [35], [36], [37]. Given a multiple measurement matrix $\mathbf{Y} = [y_1, \ldots, y_N]$ (see Section III-C), 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 identify the frequency parameters by localizing the peaks of the imaging function. In particular, denote

$$\mathbf{Y} = [\mathbf{U}_s \mathbf{U}_n] \Sigma [\mathbf{V}_s \mathbf{V}_n]^H$$

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

$$J(f) = \frac{1}{\|\mathbf{U}_n^H \mathbf{a}(f)\|_2}.$$

B. 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 [29] 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

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

with $\mathbf{a}(r, f)$ defined in (4). The intuition behind DMUSIC is that the Hankel matrix in (9) can be rewritten as

$$\mathcal{H}_y = \mathbf{A}_r \mathbf{D}_c (\mathbf{A}_f^N)^T + \mathcal{H}_c,$$

where $\mathcal{H}_c$ is a Hankel matrix formulated with $\{c(t)\}$, $t = 0, 1, \ldots, L - 1$, 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 $c_k^2 (\frac{1}{1 - r_k^2})\sqrt{1 - r_k^2}$. $\mathbf{A}_r$ and $\mathbf{A}_f^N$ are Vandermonde matrices defined as

$$\mathbf{A}_r \triangleq [\mathbf{a}(r_1, f_1), \ldots, \mathbf{a}(r_K, f_K)],$$

$$\mathbf{A}_f^N \triangleq [\mathbf{a}_N(r_1, f_1), \ldots, \mathbf{a}_N(r_K, f_K)],$$

with

$$\mathbf{a}(r, f) \triangleq \sqrt{\frac{1 - r_k^2}{1 - r_k^2 M}} \left[ r_k e^{j2\pi f_1}, \ldots, r_k^{M-1} e^{j2\pi (M-1)f} \right]^T,$$

$$\mathbf{a}_N(r, f) \triangleq \sqrt{\frac{1 - r_k^2}{1 - r_k^2 N}} \left[ r_k e^{j2\pi f_1}, \ldots, r_k^{N-1} e^{j2\pi (N-1)f} \right]^T.$$  (11)

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

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

One can show that the range spaces of $\mathcal{H}_y$, $\mathbf{A}_r$, and $\mathbf{U}_s$ are all equal when there is no noise. Then, $\mathbf{a}(r, f)$ 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 $\mathbf{a}(r, f)$ 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 $\mathbf{a}(r, f)$ that are most nearly orthogonal to the noise space $\mathbf{U}_n$. Indeed, Hankel structure has been widely used in a variety of algorithms for spectral estimation in the literature [31], [32], [33], [34].
C. 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 [16]. The authors consider the MMV setting as introduced in Section III-A3. More precisely, consider an observed $M \times N$ matrix $Y = [y_1, \cdots, y_N]$, where $y_n$ is defined in (5) and repeated as follows

$$y_n = \sum_{k=1}^{K} c_{n,k} a(f_k) + e_n, \quad n = 1, \ldots, N,$$

with $r = 1$ since undamped signals are considered in [16].

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

$$Y_{\Omega_{i,j}} = \begin{cases} Y_{ij}, & (i,j) \in \Omega, \\ 0, & \text{else}. \end{cases}$$

Then in MN-MUSIC, an SVD is directly performed on $Y_\Omega$ to get the signal space matrix $U_s$, which contains the left singular vectors of $Y_\Omega$ corresponding to the $K$ largest singular values. Finally, the frequencies are estimated by finding the peaks of

$$\|U_s^H a(f)\|^2_2,$$

which is essentially same as in Sections III-A and III-B.

IV. MAIN RESULTS

In this section we outline a connection between MUSIC and low-rank 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 [15], [27]. Moreover, a low-rank Hankel matrix recovery problem has recently been considered for damped spectral analysis [38]; that work involves solving the NNM (12) and (16) with an extra Hankel constraint on $X$. 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.

A. 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 $X^* = A_{rf} D_v \Phi^*$ or its noisy version $Y$ (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 III-B, this Hankel matrix $H_x$ can be decomposed as $H_x \triangleq A_{rf} D_v (A_{rf}^H)^\top$ and is of rank $K$ when there is no noise. One can observe that both $X^*$ and $H_x$ are low-rank matrices and have the same type of decompositions. Therefore, the analysis on $X^*$ can also be applied to $H_x$, which implies that the algorithms we build using $X^*$ in the MMV scenario also work for the SMV scenario.

Assume that $X^*$ is given and $K \ll M, N$, note that $X^*$ in (3) is low rank. Inspired by the low-rank property of $X^*$ and the dual analysis that is commonly used in atomic norm minimization (ANM) [25], [27], let us consider the following nuclear norm minimization (NNM)

$$\hat{X} = \arg \min_{X} \|X\|_* \quad \text{s.t.} \quad X = X^*.$$ (12)

Although this problem has a trivial solution (namely, $\hat{X} = X^*$), it is interesting because we can compute the corresponding dual feasible point $Q$, 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 $a^*(t)$ in (2). In particular, the Lagrange function is given as

$$L(X, Q) = \|X\|_* + \langle X^* - X, Q \rangle_R$$

$$= \|X\|_* - \langle X, Q \rangle_R + \langle X^*, Q \rangle_R,$$

with $Q$ being the dual variable. $\langle \cdot, \cdot \rangle_R$ is defined as the real inner product, i.e.,

$$\langle X^*, Q \rangle_R = \text{Re}(\langle X^*, Q \rangle) = \text{Re}(\text{Tr}(Q^H X^*))$$

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

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

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

$$\partial \|X\|_* = \partial \|X^*\|_*$$

$$= \{ Z : Z = U_X V^H_{X^*} + W, U^H_X W = 0, W V_{X^*} = 0, \|W\| \leq 1 \}$$

since $\hat{X} = X = X^*$. Note that $X^* = U_X S_X V^H_{X^*}$ is a truncated SVD of $X^*$ with $U_X, S_X, V^H_{X^*} \in \mathbb{R}^{M \times K}$ and $V_{X^*} \in \mathbb{C}^{N \times K}$. We can also construct a

$$Q \in \partial \|X\|_*$$

by letting $0 \in \frac{\partial L(Q, X)}{\partial X}$ according to the zero-gradient condition in the Karush-Kuhn-Tucker (KKT) conditions [39]. Finally, we have

$$Q = U_X V^H_{X^*},$$ (13)

by choosing $W = 0$. Given a dual feasible point $Q = U_X V^H_{X^*}$, we define the dual polynomial as

$$Q(r, f) \triangleq Q^H a(r, f),$$ (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 $\|Q(r, f)\|_2$ 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 [15], one needs the minimum separation \( \Delta_f \), which is defined in Theorem IV.2, to be on the order of \( \frac{1}{M} \) even for the full data noiseless setting.)

**Theorem IV.1.** Let \( \mathcal{RF} \) denote the set of the true damping factor and frequency pairs, i.e.,

\[
\mathcal{RF} = \{(r_1, f_1), \ldots, (r_K, f_K)\}.
\]

Given the full data matrix \( \mathbf{X}^* \) as in (3), compute its truncated SVD \( \mathbf{X}^* = \mathbf{U}_X \cdot \mathbf{S}_X \cdot \mathbf{V}_X^H \). With \( \mathbf{Q} \) as in (13), the dual polynomial defined in (14) satisfies

\[
\|\mathbf{Q}(r_k, f_k)\|_2 = 1, \quad \forall \ (r_k, f_k) \in \mathcal{RF},
\]

\[
\|\mathbf{Q}(r, f)\|_2 < 1, \quad \forall \ (r, f) \notin \mathcal{RF},
\]

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

The proof of Theorem IV.1 is given in Section V-A.

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 \( \mathbf{X}^* \). 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

1: \textbf{procedure} \textsc{Input}(\( \mathbf{X}^* \in \mathbb{C}^{M \times N} \))
2: \hspace{1em} compute truncated SVD of \( \mathbf{X}^*: \mathbf{X}^* = \mathbf{U}_X \cdot \mathbf{S}_X \cdot \mathbf{V}_X^H \).
3: \hspace{1em} form the dual feasible point: \( \mathbf{Q} = \mathbf{U}_X \cdot \mathbf{V}_X^H \).
4: \hspace{1em} form the dual polynomial: \( \mathbf{Q}(r, f) = \mathbf{Q}^H \mathbf{a}(r, f) \).
5: \hspace{1em} localize the places where \( \|\mathbf{Q}(r, f)\|_2 \leq 1 \) to get \( (\hat{r}_k, \hat{f}_k) \).
6: \textbf{return} \( (\hat{r}_k, \hat{f}_k) \).
7: \textbf{end procedure}

Note that Algorithm 2 is essentially equivalent to the MUSIC (in the undamped case) and DMUSIC (in the damped case) algorithms outlined in Section III. This is due to the fact that \( \|\mathbf{Q}(r, f)\|_2 = \|\mathbf{U}_X^H \mathbf{a}(r, f)\|_2 \). 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 \( \|\mathbf{Q}(r, f)\|_2 \leq 1 \). 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 \( \mathbf{Q} = \mathbf{U}_X \cdot \mathbf{V}_X^H \) of NNM induces a dual polynomial that satisfies \( \|\mathbf{Q}(r, f)\|_2 = \|\mathbf{U}_X^H \mathbf{a}(r, f)\|_2 \). On the other hand, the dual feasible point of ANM formulations does not admit the structure \( \mathbf{Q} = \mathbf{U}_X \cdot \mathbf{V}_X^H \), in general.

Finally, consider the case when the given data matrix \( \mathbf{Y} \) contains some additive white Gaussian noise, i.e.,

\[
\mathbf{Y} = \mathbf{X}^* + \mathbf{E}
\]

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 \( \lambda \) 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.

**B. 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 \( \mathbf{X}^* \) 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 \( \Omega \), i.e,

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

Notice that recovering the missing entries of the matrix \( \mathbf{X}^* \) reduces to a matrix completion problem [17], commonly formulated via the following NNM

\[
\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \|\mathbf{X}\|_* \quad \text{s.t.} \quad \mathbf{X}_\Omega = \mathbf{X}_\Omega^*, \tag{16}
\]

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

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

The dual problem of (17) is given by

\[
\max_{\mathbf{Q}} (\mathbf{Q}_\Omega, \mathbf{X}_\Omega) \mathcal{R} \\
\text{s.t.} \quad \|\mathbf{Q}\|_2 \leq 1, \quad \mathbf{Q}_\Omega^* = 0.
\]

Therefore, we can define the dual polynomial as

\[
\mathbf{Q}(r, f) = \mathcal{Q}^H \mathbf{a}(r, f),
\]

where \( \mathbf{Q} \) is the dual solution. Similar to Theorem IV.1, the following Theorem guarantees that we can identify the true \( r_k \)’s and \( f_k \)’s by localizing the places where \( \|\mathbf{Q}(r, f)\|_2 \) achieves 1.

**Theorem IV.2.** Suppose \( \mathbf{X}^* \) 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 \( \mathbf{X}_\Omega^* \), suppose

\[
|\Omega| \geq c_1 M N \log^4(MN)
\]
for some numerical constants \( c_1 > 0 \) and \( c_2 \). Here,
\[
\mu_1 \geq \max \left\{ \frac{M}{\mathcal{L}(M, r, f)} - \frac{\mu_2}{\sigma^2_{\min}(\Phi)} \right\}.
\]
denotes an incoherence parameter with
\[
\mu_2 \triangleq \max_{1 \leq n \leq N} \left( \sum_{k=1}^{K} |\phi_{nk}|^2 \right) \frac{N}{K},
\]
\( \mathcal{L}(M, r, f) \) is a function of \( M, r, f \) and defined as
\[
\mathcal{L}(M, r, f) \triangleq \min_{1 \leq k \leq K} \left[ \frac{c_2}{\Delta_f} \right],
\]
with \( c_2 \) being a constant and
\[
\gamma_M(r_k) \triangleq \begin{cases} \frac{2^M - 1}{2 \log(\tau)} & \text{if } r_k < 1, \\ \frac{r_k}{\frac{1}{r_k}} & \text{if } r_k = 1, \end{cases}
\]
and \( \Delta_f \) denotes the minimal separation between true frequencies, where \( |f_k - f_l| \) is the wrap-around distance on the unit circle. Then, \( X^* \) is the unique solution of (16) with probability at least \( 1 - (MN)^{-2} \). Given the recovered full data matrix \( X^* \), Theorem IV.1 guarantees the perfect recovery of all the \( (r_k, f_k) \) pairs.

The proof for Theorem IV.2 relies on some of the results in paper [40]. However, those results do not extend directly to the damped exponential case. Rather than use a certain incoherence property as in [40], 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 exploit the minimal singular value of \( \tilde{A}_r^H \tilde{A}_r \) with the function \( \mathcal{L}(M, r, f) \) by exploiting the Vandermonde structure of \( \tilde{A}_r \) [41], instead of giving an incoherence property depending on just the minimal singular value of \( \tilde{A}_r^H \tilde{A}_r \) as in [40]. Note that \( \mathcal{L}(M, r, f) = M - \frac{2c_\Delta}{\Delta_f} \) is on the order of \( M \) when there is no damping (i.e., \( r = 1 \)) and the frequencies are well separated (\( \Delta_f = O(1/M^2) \)).

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 \( \hat{X} \). Interestingly, however, as we demonstrate in Section VI, it is sometimes possible with MD-MUSIC to perfectly recover the \( r_k \)'s and \( f_k \)'s even when exact recovery of \( X^* \) fails. This implies that MD-MUSIC is actually more powerful than the alternative approach mentioned above. We leave the analysis of this phenomenon (parameter recovery without exact data matrix recovery) for future work.

V. PROOFS

A. Proof for Theorem IV.1

Denote a truncated SVD of \( A_r \) as \( A_r^* = U_{A_r} S_{A_r} V_{A_r}^H \). Since \( Q = X^* \cdot X^* \), we have
\[
\| Q^H a(r, f) \|_2^2 = \| U_{X}^H a(r, f) \|_2^2 = a(r, f)^H U_{X}^H a(r, f) = a(r, f)^H U_{X}^H S_X V_X^H V_X S_X^H U_{X}^H a(r, f) = a(r, f)^H X^* (X^*)^H a(r, f) = a(r, f)^H A_r^* F_r^H D_r^* \Phi^T (\Phi^T)^H D_r^{-1} A_r^* a(r, f) = a(r, f)^H A_r^* F_r^H D_r^* a(r, f) = a(r, f)^H A_r^* F_r^H D_r^* a(r, f) = a(r, f)^H A_r^* F_r^H D_r^* a(r, f) = a(r, f)^H U_{A_r} S_{A_r} V_{A_r}^H a(r, f) = \left( P_{U_{A_r}} a(r, f) \right)^H a(r, f),
\]
where \( P_{U_{A_r}} a(r, f) \) is the orthogonal projection of \( a(r, f) \) onto the range space of \( U_{A_r} \), i.e., \( \mathcal{R}(U_{A_r}) \). Note that the third equality is obtained by plugging in \( I = S_X \cdot V_X^H V_X S_X^H \), while the fifth equality is obtained by plugging in \( X^* = A_r^* D_r \Phi^T \) and \( (X^*)^T = (\Phi^T)^H D_r^{-1} A_r^* \). Also, note that \( \Phi^T (\Phi^T)^H = I \).
when $\Phi \in \mathbb{C}^{N \times K}$ is of rank $K$, which gives the sixth equality. The seventh equality holds due to

$$A_{r_f}^\dagger A_{r_f} = U_{A_{r_f}} S_{A_{r_f}} V_{A_{r_f}}^H V_{A_{r_f}}^{-1} U_{A_{r_f}}^H = U_{A_{r_f}} U_{A_{r_f}}^H.$$  

- For $\forall (r_k, f_k) \in \mathcal{R} \mathcal{F}$, we have $a(r_k, f_k) \in \mathcal{R}(U_{A_{r_f}})$, which implies

$$P_{U_{A_{r_f}}}(a(r_k, f_k)) = a(r_k, f_k).$$

Therefore, we have

$$\| Q^H a(r_k, f_k) \|_2^2 = \langle a(r_k, f_k), a(r_k, f_k) \rangle = 1.$$  

- For $\forall (r, f) \notin \mathcal{R} \mathcal{F}$, if we have $a(r, f) \notin \mathcal{R}(U_{A_{r_f}})$, which implies

$$P_{U_{A_{r_f}}}(a(r, f)) = a(r, f) - P_{U_{A_{r_f}}}(a(r, f)),$$

we would then have

$$\| Q^H a(r, f) \|_2^2 = \langle a(r, f), a(r, f) \rangle - \langle P_{U_{A_{r_f}}}(a(r, f)), a(r, f) \rangle < \langle a(r, f), a(r, f) \rangle = 1.$$  

Thus, we only need to show $a(r, f) \notin \mathcal{R}(U_{A_{r_f}})$ for $\forall (r, f) \notin \mathcal{R} \mathcal{F}$. Define a Vandermonde matrix $A_{r_f}^v \in \mathbb{C}^{M \times K}$ as

$$A_{r_f}^v \triangleq [a^v(r_1, f_1), \cdots, a^v(r_K, f_K)]$$  

(18)

with $a^v(r, f) \triangleq [1, e^{j2\pi f1}, \cdots, e^{j2\pi f(M-1)}]^T$, which is the unnormalized version of $a(r, f)$. Then, $A_{r_f}$ is the column-normalized version of $A_{r_f}^v$. Assuming $M \geq K$, it follows that the first $K$ rows of $A_{r_f}^v$ form a square Vandermonde matrix, denoted as $A_{r_f}^v$, whose determinant is given by [43, 44]

$$\det(A_{r_f}^v) = \prod_{1 \leq i<k \leq K} (r_i e^{j2\pi f_k} - r_k e^{j2\pi f_i}).$$

Then, $\text{rank}(A_{r_f}) = \text{rank}(A_{r_f}^v) = K$ if $M \geq K$ and $(r_i, f_i) \neq (r_k, f_k)$ for all $i \neq k$. Similarly, we have

$$\text{rank}([A_{r_f}, a(r, f)]) = K + 1,$$

i.e., $a(r, f) \notin \mathcal{R}(A_{r_f}) = \mathcal{R}(U_{A_{r_f}})$ if $(r, f) \notin \mathcal{R} \mathcal{F}$, $M \geq K$ and all the $(r_k, f_k)$ pairs in $\mathcal{R} \mathcal{F}$ are distinct. This completes the proof of Theorem IV.1.

B. Proof for Theorem IV.2

Define

$$D_{c, M} \triangleq \sqrt{M} \text{diag}([c_1, c_2, \cdots, c_K]),$$

$$\bar{A}_{r_f} \triangleq \frac{1}{\sqrt{M}} A_{r_f}^v,$$

where $A_{r_f}^v$ is the unnormalized Vandermonde matrix and defined in (18). Observe that the transpose of the noiseless data matrix $X_r^T \triangleq X^T \Phi D_{c, M} \bar{A}_{r_f} \Phi^{-1}$ can be viewed as the block Hankel matrix $X_c$ introduced in [40], but with $k_1 = N$ and $k_2 = 1$. Define $P_T$ as the projection operator that acts on the tangent space of $X_r^T$. Denote a truncated SVD of $X_r^T$ as $X_r^T = U_{X_r} S_{X_r} V_{X_r}^H$. Then, [40, Lemma 1] can be adapted to provide us sufficient conditions that are used to guarantee the unique optimality of $X_r^T$. In particular, we can set $A$ and $A_{r_f}$ as in [40, Lemma 1] as the identity operator and the random sampling operator $P_{\Omega}$, respectively. Therefore, we need the following condition

$$\left\| P_T - \frac{NM}{|\Omega|} P_T P_{\Omega} P_T \right\| \leq \frac{1}{2}.$$  

(19)

to hold.

Next, we verify that the above condition (19) holds with high probability under certain conditions. Define $A_{(n,m)} \in \mathbb{R}^{N \times M}$ as a matrix with the $(n, m)$-th entry being 1 and others being 0. We first quantify the projection of $A_{(n,m)}$ onto the subspace $T$, the tangent space of $X_r^T$. In particular, we have the following lemma which utilizes a quite different incoherence property than the one used in [40, Lemma 2].

Lemma V.1. For some constant $\mu$, if

$$\sigma_{\min}(\Phi H \Phi) \geq \frac{\mu_2}{\mu_1} \text{ and } \sigma_{\min}(\bar{A}_{r_f}^H \bar{A}_{r_f}) \geq \frac{1}{\mu_1},$$  

then

$$\| U X_r^T \Phi H \Phi A_{(n,m)} \|_F^2 \leq \frac{\mu_1 c_8 K}{N M},$$

$$\| A_{(n,m)} V X_r^T \Phi H \Phi \|_F^2 \leq \frac{\mu_2 c_8 K}{N M},$$

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 $c_8 \triangleq \max(N, M)$ and $\mu_2 \triangleq \max_{1 \leq n \leq N}(\sum_{k=1}^{K} |\phi_k|^2)^{\frac{N}{2}}$. It follows that

$$\| P_T A_{(n,m)} \|_F^2 \leq \| U X_r^T \Phi H \Phi A_{(n,m)} \|_F^2 + \| A_{(n,m)} V X_r^T \Phi H \Phi \|_F^2 \leq \frac{2\mu_1 c_8 K}{N M}.$$  

(21)

Proof. Note that $U X_r^T$ (V $X_r^T$) and $\Phi$ ($\bar{A}_{r_f}$) determine the same column (row) space of $X_r^T$. In particular, we have

$$U X_r^T \Phi H \Phi A_{(n,m)} = \Phi (\Phi H \Phi)^{-1} \Phi^H,$$

$$V X_r^T \Phi H \Phi A_{(n,m)} = \bar{A}_{r_f}^H \bar{A}_{r_f} \bar{A}_{r_f} \bar{A}_{r_f}^{-1} \bar{A}_{r_f}^H,$$

which implies

$$\| U X_r^T U_{X_r}^H A_{(n,m)} \|_F^2 = \| \Phi (\Phi H \Phi)^{-1} \Phi^H A_{(n,m)} \|_F^2 = \| (\Phi H \Phi)^{-1} (\Phi H A_{(n,m)}) \|_F^2 \leq \| (\Phi H \Phi)^{-1} \| \| \Phi H A_{(n,m)} \|_F^2 = \frac{1}{\sigma_{\min}(\Phi H \Phi)} \| \Phi H A_{(n,m)} \|_F^2$$

and

$$\| A_{(n,m)} V X_r^T V_{X_r}^H \|_F^2 = \| A_{(n,m)} \bar{A}_{r_f}^H \bar{A}_{r_f} \bar{A}_{r_f}^{-1} \bar{A}_{r_f}^H \|_F^2 = \| (\bar{A}_{r_f}^H \bar{A}_{r_f})^{-1} A_{(n,m)} \|_F^2 \| \bar{A}_{r_f} \|_F^2 \leq \frac{1}{\sigma_{\min}(\bar{A}_{r_f}^H \bar{A}_{r_f})} \| A_{(n,m)} \bar{A}_{r_f} \|_F^2.$$
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 \( A_{(n,m)} \in \mathbb{R}^{N \times M} \) is a matrix with the \((n,m)\)-th entry being 1 and all others being 0. Therefore, we can bound \( \| \Phi^H A_{(n,m)} \|_F^2 \) and \( \| A_{(n,m)} \tilde{A}_{r,f} \|_F^2 \) with
\[
\| \Phi^H A_{(n,m)} \|_F^2 = \sum_{k=1}^{K} |\phi_{nk}|^2 \leq \frac{\mu_2}{N} K,
\]
\[
\| A_{(n,m)} \tilde{A}_{r,f} \|_F^2 = \sum_{k=1}^{K} \frac{1}{M} r_{2m}^k \leq \frac{K}{M}.
\]

Define \( c_s \triangleq \max(N, M) \). Then, if
\[
\sigma_{\min}(\Phi^H \Phi) \geq \frac{\mu_2}{\mu_1}, \quad \sigma_{\min}(\tilde{A}_{r,f}^H \tilde{A}_{r,f}) \geq \frac{1}{\mu_1},
\]
we can get
\[
\| U_{X^*}^H U_{X^*}^H A_{(n,m)} \|_F^2 \leq \frac{\mu_1 K}{N M} = \frac{\mu_1 K M}{N M} \leq \frac{\mu_1 K c_s}{N M},
\]
\[
\| A_{(n,m)} V_{X^*} V_{X^*}^H \|_F^2 \leq \frac{\mu_1 K}{M} = \frac{\mu_1 K N}{N M} \leq \frac{\mu_1 K c_s}{N M}.
\]

Then, we obtain (21).

Similar to Lemma 3 in [40], we would then have that condition (19) 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 Theorem IV.2 follows the corresponding proof steps for Theorem 1 in [40]. This yields Theorem IV.2, which is similar to Theorem 1 in [40] but with different incoherence properties (20).

To obtain these different incoherence properties, we bound the minimum nonzero singular value of \( \Phi \) and \( \tilde{A}_{r,f} \). It follows from Theorem 5 of [41] that
\[
\sigma_{\min}(\tilde{A}_{r,f}^H \tilde{A}_{r,f}) = \sigma_{\min}(\tilde{A}_{r,f}) = \frac{1}{M} \sigma_{\min}^2(\tilde{A}_{r,f}) \geq \frac{1}{M} \mathcal{L}(M, r, f),
\]
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
\[
\gamma_M(r_k) \triangleq \begin{cases} \frac{r_k^{2M-1}}{2 \log(r_k)}, & r_k < 1, \\ M, & r_k = 1, \ldots, K. \end{cases}
\]

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 (\( \Delta_f = O(\frac{1}{M}) \)).

![Figure 1. Noiseless full data: the first three columns in data matrix \( X^* \).](image)

To satisfy the two assumptions in (20), we can let
\[
\frac{1}{M} \mathcal{L}(M, r, f) \geq \frac{1}{\mu_1}, \quad \text{and} \quad \mu_1 \geq \frac{\mu_2}{\sigma_{\min}^2(\Phi)},
\]
that is,
\[
\mu_1 \geq \max \left\{ \frac{M}{\mathcal{L}(M, r, f)} - \frac{\mu_2}{\sigma_{\min}^2(\Phi)} \right\}.
\]

VI. NUMERICAL SIMULATIONS

A. 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 \( X^* \) is then generated as
\[
X^* = A_{r,f} D_c \Phi^T,
\]
where \( A_{r,f} \) and \( D_c \) are generated according to their definition in Section II, \( \Phi \) 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 \( X^* \) are shown in Figure 1. Given the above data matrix \( X^* \), we then use Algorithm 2 to identify all the \( r_k \)'s and \( f_k \)'s. Figure 2 displays a surface plot of \( \| Q(r, f) \|_2 \) 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 3. We set the regularization parameter as \( \lambda = 3.9558 \) and then solve the nuclear norm denoising program (15). As is shown in Figure 4, we observe that the \((r, f)\) pairs can still be estimated by localizing the peaks of \( \| Q(r, f) \|_2 \). In particular, the estimated damping ratios and frequencies are given as \( \tilde{r}_1 = 0.92, \tilde{r}_2 = 0.98, \tilde{r}_3 = 0.79 \) and \( \tilde{f}_1 = 0.1, \tilde{f}_2 = 0.4, \tilde{f}_3 = 0.8 \). Note that we leave the corresponding theoretical guarantees for future work.
In some cases where the data matrix recovery is not perfectly indicated in Theorem IV.2. We also notice a similar behavior appearing in parameters recovery. Figure 7 (c) again indicates that we can still successfully recover the parameters in some cases where the data matrix recovery is not perfectly recovered.

We perform 20 trials in this part of simulation. It can be seen in Figure 7 that the minimal number of measurements needed for perfect data matrix recovery does scale roughly linearly with K, as indicated in Theorem IV.2. We also notice a similar behavior appearing in parameters recovery. Figure 7 (c) again indicates that we can still successfully recover the parameters in some cases where the data matrix recovery is not perfectly recovered.

We notice that the data matrix $X^\star$ is also well recovered in this case. In particular, we define the relative recovery error of data matrix as $\text{RelErr} \triangleq \frac{\|X_k^\star - X_k\|_F}{\|X_k\|_F}$, where $X_k^\star$ and $X_k$ 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 6, 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 $X^\star$. We leave the analysis of this phenomenon for future work.

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 $F = 0.05 : 0.05 : 0.95$ and a damping ratio set $R = 0.94 : 0.0025 : 1.5$. Denote $(\hat{r}, \hat{f})$ and $(r^\star, f^\star)$ as the recovered parameters and true parameters, respectively. We consider the parameter recovery to be a success if

$$\max_{1 \leq k \leq K}(|\hat{r}_k - r^\star_k|) \leq 10^{-5}, \quad \max_{1 \leq k \leq K}(|\hat{f}_k - f^\star_k|) \leq 10^{-5}. \quad (22)$$

Similarly, we consider the data matrix recovery to be a success if the relative recovery error

$$\frac{\|\hat{X} - X^\star\|_F}{\|X^\star\|_F} \leq 10^{-5}.$$ 

We perform 20 trials in this part of simulation. It can be seen in Figure 7 that the minimal number of measurements needed for perfect data matrix recovery does scale roughly linearly with $K$, as indicated in Theorem IV.2. We also notice a similar behavior appearing in parameters recovery. Figure 7 (c) again indicates that we can still successfully recover the parameters in some cases where the data matrix recovery is not perfectly recovered.

C. Data coherence

In this section, we conduct two numerical experiments to examine the influence of the minimum frequency separation $\Delta f$ and the matrix $\Phi$ on the performance of missing data matrix recovery. The standard literature on matrix completion [42] relates the recoverability of a matrix $X^\star$ to its coherence, defined as

$$\mu^\star_0 \triangleq \max \{\mu^\star_1(X^\star), \mu^\star_2(X^\star)\}$$

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 $r = 0.01 \pi \leq 0.94$. 

---

3Note that CVX [45] can return the estimated data matrix $\hat{X}$ as well as the dual solution $Q$ by solving the SDP form in (17).
with
\[
\mu_1^*(\mathbf{X}^*) \triangleq \frac{M}{K} \max_{1 \leq m \leq M} \| \mathbf{U}_m^T \mathbf{e}_m^c \|^2_2,
\]
\[
\mu_2^*(\mathbf{X}^*) \triangleq \frac{N}{K} \max_{1 \leq n \leq N} \| \mathbf{V}_n \mathbf{e}_n^c \|^2_2,
\]
where \( \mathbf{X}^* = \mathbf{U}_X \mathbf{S}_X \mathbf{V}_H \), is a truncated SVD of \( \mathbf{X}^* \), and \( \mathbf{e}_m^c \in \mathbb{R}^M \) and \( \mathbf{e}_n^c \in \mathbb{R}^N \) 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 \( \Delta f \). We generate \( \Phi \in \mathbb{C}^{N \times K} \) using normalized columns from a discrete Fourier matrix, which implies \( \mu_2^* = 1 \) and ensures that \( \mu_0^* = \mu_1^* \). 10⁴ trials are performed in this experiment. Other settings are the same as in Section VI-B. It is shown in Figure 8 that the coherence parameter \( \mu_0^* \) decreases as the minimum frequency separation \( \Delta f \) increases, which also explains why the probability of successful data matrix recovery increases as \( \Delta f \) increases.

In the second experiment, we examine the influence of the matrix \( \Phi \) on the performance of missing data recovery. We change \( N = 10 \) to make sure that the coherence parameter \( \mu_0^* \) is not too large. By fixing \( \Delta f = 1/M \), we have \( \mu_1^* = 1 \) and thus \( \mu_0^* = \mu_2^* \). We again generate \( \Phi \in \mathbb{C}^{N \times K} \) using normalized 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 9 shows that the coherence parameter \( \mu_0^* \) increases as \( \phi_{1,1} \) increases, which also explains why the probability of successful data matrix recovery decreases as \( \phi_{1,1} \) increases.

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 Theorem IV.2 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 Theorem IV.2.

D. Comparison with existing algorithms

In this section, we implement a series of experiments to compare our proposed algorithms with three existing methods:
With a full SVD of the data matrix $\hat{X}$ from NNM. In particular, we set the estimated number of frequencies $\hat{K}$ as the number of singular values of $\hat{X}$ that are not less that $0.1\sigma_{\text{max}}$, where $\sigma_{\text{max}}$ is the maximal singular value of $\hat{X}$. Then, we use this estimated $\hat{K}$ when we implement the MUSIC algorithm.

Table I
Comparison of the “one-step” MD-MUSIC algorithm and the “two-step” NNM+MUSIC for parameter recovery. We present the probability of successful recovery over 100 trials.

|                  | 10% missing | 20% missing | 30% missing | 40% missing |
|------------------|-------------|-------------|-------------|-------------|
| MD-MUSIC         | 100%        | 100%        | 98%         | 97%         |
| NNM+MUSIC        | 70%         | 68%         | 69%         | 60%         |

2) MN-MUSIC: Next, we compare our proposed MD-MUSIC algorithm with the NNM-MUSIC algorithm introduced in Section III-C [16] in a scenario where 20% of the noiseless data entries are missing. We observe that the NNM-MUSIC algorithm never successfully recovers the frequencies and damping ratios since it performs an SVD directly on the missing data.\(^7\)

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

$$\min_{X, u, W} \frac{1}{2M} \text{Tr}(\text{Toep}(u)) + \frac{1}{2} \text{Tr}(W)$$

s.t. $\begin{bmatrix} \text{Toep}(u) & X \\ X^H & W \end{bmatrix} \succeq 0$, $X_\Omega = X_\Omega^*$

where $\text{Toep}(u)$ is a Hermitian Toeplitz matrix with the vector $u$ being its first column. We use $X = X^*$ instead of $X_\Omega = X_\Omega^*$ 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 $\ell_2$-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 VI-A with $N = 10$ and with a variety of $M$ and $\Delta f$. We

\(^6\)A similar idea has also been considered in [38].

\(^7\)No results are shown since MN-MUSIC never recovers successfully.
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 VI-C, we then let $f_2 = f_1 + \Delta f$ with various values of $\Delta f$. The simulation results are given in Figure 10. It can be seen that the NN-MUSIC algorithm significantly outperforms ANM and can always recover the frequencies exactly, as indicated in Theorem IV.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 11, 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 11(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 11 (a, c) to test with a larger value of $\Delta f$. Other parameters used in this part are the same as in the full data experiments.

VII. 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. 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 and NSF CAREER grant CCF–1149225.

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. Qi, 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] L. C. Potter, E. Ertin, J. T. Parker, and M. Cetin, “Sparsity and compressed sensing in radar imaging,” Proceedings of the IEEE, vol. 98, no. 6, pp. 1006–1020, 2010.
[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, 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] R. de Prony, “Essai expérimental et analytique sur les lois de la dilabilité et sur celles de la force expansive de la vapeur de leau et de la vapeur de la lavooll,a différentes températures,” J. de l’Ecole Polytechnique, vol. 1, no. 22, pp. 24–76, 1795.
[8] 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.
[9] R. Schmidt, A signal subspace approach to multiple emitter location and spectral estimation. Stanford University, 1981.
[10] R. Schmidt, “Multiple emitter location and signal parameter estimation,” IEEE Transactions on Antennas and Propagation, vol. 34, no. 3, pp. 276–280, 1986.
[11] F. Marvasti, Nonuniform sampling: theory and practice. Springer Science & Business Media, 2012.
P. Stoica and R. L. Moses, "Introduction to spectral analysis," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3571–3586, 2008.

Y. Li, K. Liu, and J. Razavilar, "Improved parameter estimation schemes for damped sinusoidal signals," *IEEE Transactions on Signal Processing*, vol. 64, no. 19, pp. 5145–5157, 2016.

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.

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

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.

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.

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

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.

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.

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.

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.

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.

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.

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.

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.

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.

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

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.

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.

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.

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.