Compressive Two-Dimensional Harmonic Retrieval via Atomic Norm Minimization

Yuejie Chi, *Member, IEEE*, and Yuxin Chen, *Student Member, IEEE*

Abstract—This paper is concerned with estimation of two-dimensional (2-D) frequencies from partial time samples, which arises in many applications such as radar, inverse scattering, and super-resolution imaging. Suppose that the object under study is a mixture of r continuous-valued 2-D sinusoids. The goal is to identify all frequency components when we only have information about a random subset of n regularly-spaced time samples. We demonstrate that under some mild spectral separation condition, it is possible to exactly recover all frequencies by solving an atomic norm minimization program, as long as the sample complexity exceeds the order of $r \log r \log n$. We then propose to solve the atomic norm minimization via a semidefinite program and provide numerical examples to justify its practical ability. Our work extends the framework proposed by Tang *et. al.* for line spectrum estimation to 2-D frequency models.

Index Terms—atomic norm, basis mismatch, continuous-valued frequency recovery, sparsity

I. INTRODUCTION

The problem of estimating two-dimensional (2-D) spectrum is encountered in a variety of signal processing applications. For instance, the multi-dimensional frequency model naturally arises in several operational scenarios in multiple-input multiple-output (MIMO) radars [2], where multiple components of each frequency correspond respectively to the direction of arrival, direction of departure, and Doppler shift of a scatter. Retrieving these parameters is of great importance for localization and tracking of targets [3]. A second application concerns channel sensing in wireless communications, where accurate estimation of channel state information is crucial for coherent detection in order to ensure high data rate. Physical arguments and a growing body of experimental evidence suggest that the number of significant paths in a wireless channel is typically small [4], [5]. Each path is specified by a triple of time delay, Doppler shift and attenuation, and can be mapped to a multi-dimensional frequency. Another example is super-resolution imaging [6], where any 2-D point source translates into a 2-D complex sinusoid after passing through a Fourier imaging system.

One of the essential goals in various applications is to minimize the number of samples required to recover the underlying frequencies. Take wireless communications as an example, where training pilots are transmitted and extracted from the received signal to estimate the channel. The smaller the number of pilots, the higher the data rate. Conventional channel estimation methods are often based on linear least-squares estimators [7], which requires the sample size to be greater than the dimensionality of the signal space determined by the maximal time delay and Doppler shift. To reduce the required sample size, conventional approaches are often based on parametric representation, which directly estimate 2-D frequencies via super-resolution methods such as 2-D unitary ESPRIT [8], 2-D MUSIC [9], Clark and Scharf's IQML method [10], the Matrix Enhancement Matrix Pencil (MEMP) method [11], etc. However, many of these approaches require equi-spaced time-domain samples. They also rely on prior knowledge on the model order – the number of sinusoids. Moreover, these methods are often sensitive to model order mismatch and noise.

Pioneered by the work of Candès et. al. [12] and Donoho [13], Compressive Sensing (CS) suggests that it is possible to recover a spectrally sparse signal from highly incomplete time-domain samples. Specifically, consider a time-domain signal of ambient dimension $n = n_1 \times n_2$, composed of r distinct 2-D complex sinusoids. If the frequencies of the sinusoids lie approximately on the fine DFT grid of the normalized frequency plane $[0, 1) \times [0, 1)$, the signal of interest can be sparsely represented over the DFT basis. It has been demonstrated that the signal can be recovered from a random subset of time-domain samples with a sample size of $\mathcal{O}(r \log n)$ [14] via basis pursuit [15] or greedy pursuit [16]. The success of CS has inspired a large body of algorithm and system design enabling sub-Nyquist sampling, notably for compressive channel sensing [17], [18], high-resolution radar [19], [20], and multi-user detection [21]. Caution needs to be exercised, however, when approximating the continuous-valued frequencies over a discrete (DFT) grid, since the signal of interest often contains off-the-grid components and might not enjoy a good sparse approximation over the discrete basis. This effect has been studied in great details in [22], revealing considerable performance degradation of conventional CS algorithms when applied to off-the-grid signals. Several possible remedies have been suggested ever since, see for example [23]–[27]. Nevertheless, a grid is still assumed in these remedies and therefore the continuous-valued frequencies cannot be recovered perfectly. In addition, there seems to be little theoretical understanding of these approaches.

Several recent works have been proposed to deal directly with continuous-valued frequencies without imposing a dis-

Y. Chi is with Department of Electrical and Computer Engineering and Department of Biomedical Informatics, The Ohio State University. Email: chi.97@osu.edu. Y. Chen is with Department of Electrical Engineering, Stanford University. Email: yxchen@stanford.edu.

crete dictionary. In the pioneer work of [28], Candès and Fernandez-Granda proves that perfect frequency extrapolation is possible from partial low-end time samples by solving a total-variation minimization program. This analyses therein readily extend to multi-dimensional frequency models. Tang et. al. [29] investigates the problem of 1-D spectral estimation when one is given randomly observed time-domain samples, and proves that atomic norm minimization [30] succeeds with $\mathcal{O}(r \log r \log n)$ samples, assuming that the wrap-around distance between distinct frequencies is at least $4/n$. More precisely, the atomic norm proposed by Chandrasekaran et. al. [30] is a general recipe for developing convex optimization solutions for model selection, where the goal is to minimize the number of selected atoms for a given parsimonious model. Many well-known problems can be treated as a special case of atomic norm minimization, including ℓ_1-minimization for sparse recovery where the atoms are unit-norm one-sparse vectors, nuclear norm minimization for low-rank matrix completion where the atoms are unit-norm rank-one matrices, and so on. For spectrally sparse signals, the atoms are Vandermonde vectors with a continuous-valued frequency in $[0, 1)$. It is worth noting that the atomic norm for spectrally sparse signals is equivalent to the total-variation norm adopted in [28], [31]. Another line of work has approached the multi-dimensional harmonic retrieval problem via Enhanced Matrix Completion (EMaC) [32], [33], namely, to perform nuclear norm minimization over multi-fold Hankel matrices constructed from the time-domain samples. This algorithm is guaranteed to work from $\mathcal{O}(r\text{polylog}n)$ random samples, provided that the signal model obeys some mild incoherence properties.

In this paper, we extend the atomic norm minimization approach by Tang *et. al.* [29] to 2-D frequency models. When the sample size exceeds the order of $r \log r \log n$, the proposed atomic norm minimization algorithm is guaranteed to perfectly recover all 2-D frequency components with high probability, under a mild frequency separation condition. The proof is inspired by [29] and [28], that is, to construct a dual polynomial certifying the optimality of the solution to the corresponding convex program. We then propose to solve the atomic norm minimization problem via semidefinite programming (SDP), which can be performed tractably using off-the-shelf SDP solvers. However, unlike the case in 1-D model [29], the equivalence between the atomic norm minimization and our proposed SDP is not guaranteed, primarily beacuse the Caratheodory's theorem [34] does not hold in higher dimensions. Instead, we validate the effectiveness of the proposed SDP through numerical examples and its noise robustness is also examined. After the conference version [1] of this paper was published, Xu *et. al.* developed a precise SDP characterization [35] of the 2-D atomic norm minimization based on the theory of positive trigonometric polynomials [36], where our proposed SDP can be regarded as a first-order relaxation in their sum-of-squares relaxation hierarchy.

The rest of the paper is organized as follows. In Section II, we formulate the problem and review related literature. Section III presents the proposed atomic norm minimization algorithm along with its performance guarantee, whose proof is deferred to the Appendix. Section IV introduces a semidefinite

program to approximate the original atomic norm minimization. Numerical experiments are supplied in Section V to validate the practical applicability of our algorithm. Finally we conclude in Section VI with a summary of our findings. Throughout the paper, we use $(\cdot)^\top$ and $(\cdot)^*$ to denote the transpose and the conjugate transpose, respectively.

II. PROBLEM FORMULATION AND RELATED WORK

A. Problem Formulation

Without loss of generality, consider a 2-D square data matrix $\boldsymbol{X}^\star$ of size $n = (4M + 1) \times (4M + 1)$, where $J = \{-2M, \ldots, 0, \ldots, 2M\} \times \{-2M, \ldots, 0, \ldots, 2M\}$ denotes the union of the indices of $\boldsymbol{X}^\star$. This assumption is imposed to simplify the development of the theoretical guarantees, and can be removed with little modifications, see [29] for a similar treatment. Each entry of $\boldsymbol{X}^\star$ can be expressed as a superposition of r complex sinusoids observed at the time index $\boldsymbol{k} = [k_1, k_2] \in J$, i.e.

$$x_{\boldsymbol{k}}^\star = x_{k_1, k_2}^\star = \frac{1}{(4M+1)} \sum_{i=1}^{r} d_i e^{j2\pi \boldsymbol{f}_i^\top \boldsymbol{k}}, \qquad (1)$$

where d_i represents the complex amplitude associated with each $1 \le i \le r$. Let $\Omega = \{\boldsymbol{f}_i = (f_{1i}, f_{2i}) \in [0, 1) \times [0, 1),\ 1 \le i \le r\}$ be the set of distinct frequencies. For notational simplicity, we introduce the following unit-norm atoms:

$$\begin{cases} \boldsymbol{a}(f_{1i}) &= \frac{1}{\sqrt{(4M+1)}} \left[y_i^{-2M}, \ldots, 1, \ldots, y_i^{2M} \right]^\top, \\ \boldsymbol{a}(f_{2i}) &= \frac{1}{\sqrt{(4M+1)}} \left[z_i^{-2M}, \ldots, 1, \ldots, z_i^{2M} \right]^\top, \end{cases}$$

where $y_i = e^{j2\pi f_{1i}}$, and $z_i = e^{j2\pi f_{2i}}$. This allows us to write $\boldsymbol{X}^\star$ in a matrix form as follows

$$\boldsymbol{X}^\star = \boldsymbol{Y}\boldsymbol{D}\boldsymbol{Z}^\top, \qquad (2)$$

where $\boldsymbol{Y}$ is given by

$$\boldsymbol{Y} = [\boldsymbol{a}(f_{11}), \ldots, \boldsymbol{a}(f_{1r})] \in \mathbb{C}^{(4M+1) \times r}, \qquad (3)$$

$$\boldsymbol{Z} = [\boldsymbol{a}(f_{21}), \ldots, \boldsymbol{a}(f_{2r})] \in \mathbb{C}^{(4M+1) \times r}, \qquad (4)$$

and

$$\boldsymbol{D} = \text{diag}([d_1, d_2, \cdots, d_r]) = \text{diag}(\boldsymbol{d}) \in \mathbb{C}^{r \times r}. \qquad (5)$$

Denote by $\boldsymbol{x}^\star = \text{vec}((\boldsymbol{X}^\star)^\top) \in \mathbb{C}^{(4M+1)^2}$ the vectorized data matrix, then one has

$$\boldsymbol{x}^\star = (\boldsymbol{Y} \otimes \boldsymbol{Z})\boldsymbol{d} = \sum_{i=1}^{r} d_i \boldsymbol{a}(f_{1i}) \otimes \boldsymbol{a}(f_{2i})$$

$$= \sum_{i=1}^{r} d_i \boldsymbol{c}(\boldsymbol{f}_i), \qquad (6)$$

where $\otimes$ represents Kronecker product, and

$$\boldsymbol{c}(\boldsymbol{f}_i) = \boldsymbol{c}(f_{1i}, f_{2i}) := \boldsymbol{a}(f_{1i}) \otimes \boldsymbol{a}(f_{2i}) \in \mathbb{C}^{(4M+1)^2}.$$

satisfying $\|\boldsymbol{c}(\boldsymbol{f}_i)\|_2 = 1$.

In this paper, we assume that m entries of $\boldsymbol{X}^\star$ are observed uniformly at random. Specifically, denote by $T \subset J$ as the index set such where $x_{k_1, k_2}^\star$ are observed if and only iff $(k_1, k_2) \in T$. Define the operator $\mathcal{P}_T$ such that $\mathcal{P}_T(\boldsymbol{M})$

represents the orthogonal projection of $\boldsymbol{M}$ onto the subspace of matrices supported on T. We shall abuse the notation, without ambiguity, to let T, J and $\mathcal{P}_T$ represent the set of observed entries, all entries, and the observation operator with respect to the vectorized signal $\boldsymbol{x}^\star$ as well.

The primary focus of this paper is to recover the unobserved entries of the original data matrix $\boldsymbol{X}^\star$. We note that the frequencies Ω can also be recovered using conventional approaches such as the MEMP method [11] once the data matrix is recovered.

B. Conventional CS Approach

To apply conventional CS paradigms, we represents $\boldsymbol{x}^\star$ as a sparse signal in a pre-determined basis by discretizing the 2-D plane $[0,1) \times [0,1)$ with grid points $\boldsymbol{t} = [t_1, t_2] \in \Omega_d$, where $t_1, t_2 \in \left\{0, \ldots, \frac{4M}{4M+1}\right\}$. Write the resulting DFT basis as

$$\boldsymbol{F} = [\boldsymbol{c}(\boldsymbol{t})|_{\boldsymbol{t} \in \Omega_d}] = \boldsymbol{F}_1 \otimes \boldsymbol{F}_1 \in \mathbb{C}^{(4M+1)^2 \times (4M+1)^2},$$

where $\boldsymbol{F}_1$ is a DFT matrix of dimension $(4M+1) \times (4M+1)$. The vectorized signal $\boldsymbol{x}^\star$ can then be represented using $\boldsymbol{F}$ as

$$\boldsymbol{x}^\star = \boldsymbol{F}\widetilde{\boldsymbol{d}}, \tag{7}$$

where $\widetilde{\boldsymbol{d}}$ is approximately sparse. CS suggests that we could recover $\boldsymbol{x}^\star$ using the ℓ_1-minimization as

$$\min_{\boldsymbol{d}} \quad \|\boldsymbol{d}\|_1 \quad \text{subject to} \quad \mathcal{P}_T(\boldsymbol{F}\boldsymbol{d}) = \mathcal{P}_T(\boldsymbol{x}^\star),$$

where the minimizer is returned as an estimate of $\widetilde{\boldsymbol{d}}$. The major issue with the above approach is that the frequencies $\boldsymbol{f}_i$ never lie perfectly on the grid Ω_d, resulting in inevitable mismatch issue between the true frequencies and the discrete grid. It has been demonstrated in [22] that the performance of sparse recovery algorithms can degenerate considerably. In this paper we will adopt a different approach and attempt to recover the frequencies directly without imposing a grid.

III. ATOMIC NORM MINIMIZATION FOR 2-D HARMONIC RETRIEVAL

The atomic norm is proposed in [30] as a general recipe of designing convex optimization solutions for model selection, by convexifying the atomic set of the parsimonious models. The atomic set of a signal model is defined as the simplest building blocks of the signal, such as unit-norm one-sparse vectors for sparse recovery, unit-norm rank-one matrices for low-rank matrix completion, and so on. Interested readers are referred to [30] for a detailed discussion about the atomic norm. In the case of 2-D harmonic retrieval, it is straightforward to define the atomic set as the collection of all normalized 2-D complex sinusoids:

$$\mathcal{A} := \{\boldsymbol{c}(\boldsymbol{f})|\boldsymbol{f} \in [0,1) \times [0,1)\},$$

and the atomic norm for a signal $\boldsymbol{x}$ as

$$\|\boldsymbol{x}\|_{\mathcal{A}} := \inf_{\substack{\boldsymbol{f}_i \in [0,1) \times [0,1) \\ d_i \in \mathbb{C}}} \left\{\sum_i |d_i| \Big| \boldsymbol{x} = \sum_i d_i \boldsymbol{c}_i(\boldsymbol{f}_i)\right\}. \tag{8}$$

This is obtained by convexifying the atomic representation of $\boldsymbol{x}$ using the smallest number of 2-D frequency spikes:

$$\|\boldsymbol{x}\|_{\mathcal{A},0} = \inf_{\substack{\boldsymbol{f}_i \in [0,1) \times [0,1) \\ d_i \in \mathbb{C}}} \left\{s \Big| \boldsymbol{x} = \sum_{i=1}^s d_i \boldsymbol{c}_i(\boldsymbol{f}_i)\right\}.$$

The above definition generalizes the atomic norm for 1-D harmonic signals in [29] and allows one to accommodate higher dimensions. Given partial observations of $\boldsymbol{x}^\star$ (or equivalently $\mathcal{P}_T(\boldsymbol{x}^\star)$), we attempt recovery via the following atomic norm minimization program

$$\hat{\boldsymbol{x}} = \operatorname{argmin}_{\boldsymbol{x}} \|\boldsymbol{x}\|_{\mathcal{A}} \quad \text{subject to} \quad \mathcal{P}_T(\boldsymbol{x}) = \mathcal{P}_T(\boldsymbol{x}^\star), \tag{9}$$

namely, to seek a signal with minimal atomic norm satisfying the observation constraints. This approach is adopted in [29] for line spectrum estimation when the set of atoms is $\mathcal{A} = \{\boldsymbol{a}(f)|f \in [0,1)\}$. In [29], it is shown that a random subset containing $\mathcal{O}(r \log r \log n)$ samples can ensure exact frequency recovery under a mild frequency separation condition.

The following theorem establishes similar performance guarantees hold in the 2-D case, namely, the proposed algorithm (9) recovers the true data $\boldsymbol{x}^\star$ perfectly under a properly defined separation condition, provided that the sample complexity exceeds the order of $r \log r \log n$.

Theorem 1. *Let $M \geq 256$. Suppose that we observe samples of a data matrix $\boldsymbol{X}^\star$ in (1) on the index set $T \subset J$ of size $|T| = m$ uniformly at random, where $\boldsymbol{f}_i \in [0,1) \times [0,1)$. Suppose that the signs of d_i's are i.i.d. and uniformly drawn from $\{+1, -1\}$, and the minimum separation between $\boldsymbol{f}_i$'s satisfies*

$$\Delta_{\min} \triangleq \min_{i \neq j} \|\boldsymbol{f}_i - \boldsymbol{f}_j\|_\infty$$

$$= \min_{i \neq j} \max\{|f_{1i} - f_{1j}|, |f_{2i} - f_{2j}|\} \geq \frac{1.19}{M}, \tag{10}$$

where $|f_{1i} - f_{1j}|$, $|f_{2i} - f_{2j}|$ are the wrap-around distances on the unit circle. Then there exists a numerical constant $C > 0$ such that if

$$m \geq C \max\left\{\log^2 \frac{M}{\delta}, r \log \frac{r}{\delta} \log \frac{M}{\delta}\right\}, \tag{11}$$

then the solution to (9) is exact and unique with probability at least $1 - \delta$. The same results hold with a different constant in (10) when the signs of d_i's are i.i.d. uniformly generated on the complex unit circle.

The proof can be found in Appendix B. Theorem 1 suggests that as long as the frequencies are minimally separated as in (10), the recovery via atomic norm minimization is exact once m is on the order of $\max\{\log^2 n, r \log r \log n\}$. This orderwise bound agrees with the performance guarantee for line spectrum estimation as derived in [29].

We compare Theorem 1 with conventional subspace methods such as ESPRIT. ESPRIT is able to recover the underlying frequencies from $\Theta(r)$ *consecutive* samples of the data matrix $\boldsymbol{X}^\star$. The number of samples required for exact recovery depends only on the underlying degrees of freedom irrespective of the ambient dimension of $\boldsymbol{X}^\star$. In contrast, the proposed

algorithm (9) assumes random subsampling of the data matrix $\boldsymbol{X}^\star$ and requires a slightly higher sample complexity about the order of $r \log r \log n$. Moreover, in the absence of noise, ESPRIT allows recovery without imposing a separation condition like (10). Note, however, that a separation condition is necessary when noise is present, as detailed in [28], [37]. We will demonstrate through numerical examples that the proposed algorithm (9) is stable under noisy observations as well.

We also compare Theorem 1 with standard results in CS [14]. When the frequencies in Ω are indeed on the DFT grid, CS allows recovery of r complex sinusoids from a number $\mathcal{O}(r \log(n/r))$ of samples. The proposed algorithm (9) can be regarded as a remedy of CS for targets off the grid with slightly larger sample complexity.

IV. APPROXIMATE SEMIDEFINITE PROGRAM TO SOLVE ATOMIC NORM MINIMIZATION

Theorem 1 indicates that solving the atomic norm minimization problem (9) allows perfect recovery of the data matrix from only a small number of its time samples. However, a natural question arises as to how to solve (9) in a tractable manner. Unfortunately, the exact semidefinite programming characterization of atomic norm minimization in the line spectrum case, as proposed in [29], cannot be extended straightforwardly to 2-D models, due to the fundamental difficulty of generalizing the classical Caratheodory's theorem [34] to higher dimensions. Nonetheless, in this section we propose a semidefinite program to approximately solve (9), which exhibits excellent empirical performance in Section V. We also provide a sufficient condition when the proposed semidefinite program returns the solution to (9).

We describe the algorithm in the general case when the dimension of $\boldsymbol{X}^\star$ is $n_1 \times n_2$, which is not necessarily square. We will still assume that $\boldsymbol{X}^\star$ satisfies (2), but slightly abuse notations by letting

$$\boldsymbol{Y} = [\boldsymbol{a}(f_{11}), \ldots, \boldsymbol{a}(f_{1r})]$$

with $\boldsymbol{a}(f_{1i}) = [1, e^{j2\pi f_{1i}}, \ldots, e^{j2\pi(n_1-1)f_{1i}}]^\top$, and

$$\boldsymbol{Z} = [\boldsymbol{b}(f_{21}), \ldots, \boldsymbol{b}(f_{2r})],$$

with $\boldsymbol{b}(f_{2i}) = [1, e^{j2\pi f_{2i}}, \ldots, e^{j2\pi(n_2-1)f_{1i}}]^\top$, wherever they are clear from context.

Before presenting the algorithm, we first define a matrix enhancement using two-fold Toeplitz structures. Given a $(2n_1-1) \times (2n_2-1)$ matrix $\boldsymbol{T} = [x_{l_1,l_2}]$ with $-n_1 < l_1 < n_1$ and $-n_2 < l_2 < n_2$, we define an $n_1 \times n_1$ block Toeplitz matrix $\mathcal{S}(\boldsymbol{T})$ from $\boldsymbol{T}$ as

$$\mathcal{S}(\boldsymbol{T}) = \begin{bmatrix} \boldsymbol{T}_0 & \boldsymbol{T}_{-1} & \cdots & \boldsymbol{T}_{-(n_1-1)} \\ \boldsymbol{T}_1 & \boldsymbol{T}_0 & \cdots & \boldsymbol{T}_{-(n_1-2)} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{T}_{n_1-1} & \boldsymbol{T}_{n_1-2} & \cdots & \boldsymbol{T}_0 \end{bmatrix}, \quad (12)$$

where each block $\boldsymbol{T}_l$ $(-n_1 < l < n_1)$ is an $n_2 \times n_2$ Toeplitz matrix defined from the lth row of $\boldsymbol{T}$:

$$\boldsymbol{T}_l = \begin{bmatrix} x_{l,0} & x_{l,-1} & \cdots & x_{l,-(n_2-1)} \\ x_{l,1} & x_{l,0} & \cdots & x_{l,-(n_2-2)} \\ \vdots & \vdots & \vdots & \vdots \\ x_{l,n_2-1} & x_{l,n_2-2} & \cdots & x_{l,0} \end{bmatrix}.$$

We use $\mathcal{S}(\boldsymbol{T}) \in \mathbb{C}^{n_1 n_2 \times n_1 n_2}$ to represent the corresponding two-fold block Toeplitz matrix constructed from $\boldsymbol{T}$. It is straightforward to verify that for any i, an atom in the form of $(\boldsymbol{a}(f_{1i})\boldsymbol{a}(f_{1i})^*) \otimes (\boldsymbol{b}(f_{2i})\boldsymbol{b}(f_{2i})^*)$ forms a two-fold block Toeplitz matrix. The following proposition presents a semidefinite program that allows approximation of the atomic norm $\|\boldsymbol{x}\|_\mathcal{A}$.

Proposition 1. *Let $\boldsymbol{X}$ be an $n_1 \times n_2$ matrix and $\boldsymbol{x} = vec(\boldsymbol{X}^\top)$. Denote*

$$\{\hat{\boldsymbol{T}}, \hat{t}\} = \mathrm{argmin}_{\boldsymbol{T},t} \left\{ \frac{1}{2}\mathrm{tr}\left(\mathcal{S}(\boldsymbol{T})\right) + \frac{1}{2}t \,\middle|\, \begin{bmatrix} \mathcal{S}(\boldsymbol{T}) & \boldsymbol{x} \\ \boldsymbol{x}^* & t \end{bmatrix} \succeq 0 \right\}, \tag{13}$$

and let $\|\boldsymbol{x}\|_\mathcal{T}$ the objective value under $\{\hat{\boldsymbol{T}}, \hat{t}\}$. Then we have

$$\|\boldsymbol{x}\|_\mathcal{A} \geq \|\boldsymbol{x}\|_\mathcal{T}.$$

Furthermore, if $\mathcal{S}(\hat{\boldsymbol{T}})$ can be written as

$$\mathcal{S}(\hat{\boldsymbol{T}}) = \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}^*, \tag{14}$$

where

$$\boldsymbol{V} = [\boldsymbol{c}(\boldsymbol{f}_1), \ldots, \boldsymbol{c}(\boldsymbol{f}_r)],$$
$$\boldsymbol{\Sigma} = \mathrm{diag}([\sigma_1, \ldots, \sigma_r]),$$

with σ_i's being real and positive values, then $\|\boldsymbol{x}\|_\mathcal{A} = \|\boldsymbol{x}\|_\mathcal{T}$.

Proof. Let $\boldsymbol{x} = \sum_i d_i \boldsymbol{c}(\boldsymbol{f}_i)$, where $d_i = |d_i|e^{j\theta_i}$, then

$$\sum_i |d_i| \begin{bmatrix} \boldsymbol{c}(\boldsymbol{f}_i) \\ e^{-j\theta_i} \end{bmatrix} \begin{bmatrix} \boldsymbol{c}(\boldsymbol{f}_i) \\ e^{-j\theta_i} \end{bmatrix}^* = \begin{bmatrix} \mathcal{S}(\boldsymbol{T}) & \boldsymbol{x} \\ \boldsymbol{x}^* & \sum_i |d_i| \end{bmatrix} \succeq 0,$$

where $\mathcal{S}(\boldsymbol{T}) = \sum_i |d_i| \boldsymbol{c}(\boldsymbol{f}_i)\boldsymbol{c}(\boldsymbol{f}_i)^*$. This indicates that

$$\|\boldsymbol{x}\|_\mathcal{T} \leq \frac{1}{2}\mathrm{tr}\left(\mathcal{S}(\boldsymbol{T})\right) + \frac{1}{2}\sum_i |d_i|$$
$$= \sum_i |d_i| = \|\boldsymbol{x}\|_\mathcal{A}.$$

Moreover, if the optimal $\hat{\boldsymbol{T}}$ and $\hat{t}$ satisfy

$$\begin{bmatrix} \mathcal{S}(\hat{\boldsymbol{T}}) & \boldsymbol{x} \\ \boldsymbol{x}^* & \hat{t} \end{bmatrix} \succeq 0,$$

then we have $\mathcal{S}(\hat{\boldsymbol{T}}) \succeq 0$ and $\mathcal{S}(\hat{\boldsymbol{T}}) \succeq \hat{t}^{-1}\boldsymbol{x}\boldsymbol{x}^*$ by Schur complement condition. If we can write $\mathcal{S}(\hat{\boldsymbol{T}}) = \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}^*$, then $\boldsymbol{x}$ falls within the column space of $\mathcal{S}(\hat{\boldsymbol{T}})$ or, equivalently, $\boldsymbol{x} = \boldsymbol{V}\boldsymbol{d}$ for some vector $\boldsymbol{d}$. Let $\boldsymbol{q}$ be any vector such that $\boldsymbol{V}_2^*\boldsymbol{q} = \mathrm{sign}(\boldsymbol{d})$, where $\mathrm{sign}(\boldsymbol{d})$ is the sign vector of $\boldsymbol{d}$, then

$$\mathrm{tr}(\boldsymbol{\Sigma}) = \boldsymbol{q}^*\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}^*\boldsymbol{q} \geq \frac{1}{\hat{t}}\boldsymbol{q}^*\boldsymbol{V}\boldsymbol{d}\boldsymbol{d}^*\boldsymbol{V}^*\boldsymbol{q} = \frac{1}{\hat{t}}\left(\sum_i |d_i|\right)^2.$$

This implies that

$$\frac{1}{2}\text{tr}\left(\mathcal{S}(\hat{\boldsymbol{T}})\right) + \frac{1}{2}\hat{t} = \frac{1}{2}\text{tr}\left(\boldsymbol{\Sigma}\right) + \frac{1}{2}\hat{t} \geq \sqrt{\text{tr}\left(\boldsymbol{\Sigma}\right)\hat{t}}$$
$$\geq \sum_i |d_i| \geq \|\boldsymbol{x}\|_{\mathcal{A}}, \qquad (15)$$

which is equivalent to $\|\boldsymbol{x}\|_{\mathcal{T}} \geq \|\boldsymbol{x}\|_{\mathcal{A}}$. Therefore we have $\|\boldsymbol{x}\|_{\mathcal{T}} = \|\boldsymbol{x}\|_{\mathcal{A}}$. $\qquad\square$

We propose to approximate the atomic norm minimization algorithm in (9) via the following semidefinite program

$$\{\hat{\boldsymbol{T}}, \hat{\boldsymbol{x}}, \hat{t}\} = \text{argmin}_{\boldsymbol{T}, \boldsymbol{x}, t} \quad \frac{1}{2}\text{tr}\left(\mathcal{S}(\boldsymbol{T})\right) + \frac{1}{2}t \qquad (16)$$
$$\text{subject to} \quad \begin{bmatrix} \mathcal{S}(\boldsymbol{T}) & \boldsymbol{x} \\ \boldsymbol{x}^* & t \end{bmatrix} \succeq 0,$$
$$\mathcal{P}_T(\boldsymbol{x}) = \mathcal{P}_T(\boldsymbol{x}^\star).$$

Unlike the 1-D algorithm proposed in [29], since it is not guaranteed to write $\mathcal{S}(\hat{\boldsymbol{T}})$ into a form of (14), the semidefinite program formulation (16) is in general not guaranteed to be equivalent to (9).

Although the equivalence between (16) and (9) is not ensured, we can establish that if rank$(\mathcal{S}(\boldsymbol{T}))$ is not greater than $\min\{n_1, n_2\}$ for certain matrix $\boldsymbol{T}$ (in general it could be as large as $n_1 n_2$), it can indeed be written uniquely in the form of (14). This is characterized in the following proposition, whose proof is deferred to Appendix H.

Proposition 2. *If* $r = \text{rank}(\mathcal{S}(\boldsymbol{T})) \leq \min\{n_1, n_2\}$, *then* $\mathcal{S}(\boldsymbol{T})$ *is PSD if and only if it can be represented as* (14).

From the above proposition, it is straightforward that if the solution to (16) satisfies rank$(\mathcal{S}(\hat{\boldsymbol{T}})) \leq \min\{n_1, n_2\}$, the semidefinite characterization of (16) is exact.

Remark 1. The dual problem of (16) can be written as

$$\max_{\boldsymbol{\nu}, \boldsymbol{Q}} \text{Re}(\langle \boldsymbol{\nu}, \boldsymbol{x}^\star \rangle), \text{ subject to } \begin{bmatrix} \boldsymbol{Q} & \boldsymbol{\nu} \\ \boldsymbol{\nu}^* & 1 \end{bmatrix} \succeq 0,$$
$$\langle \boldsymbol{Q}, \boldsymbol{\Theta}_{\boldsymbol{k}} \rangle = \delta_{\boldsymbol{k}},$$
$$\boldsymbol{\nu}_{T^c} = \boldsymbol{0},$$

where $\boldsymbol{\Theta}_{\boldsymbol{k}}$, $\boldsymbol{k} = (k_1, k_2)$ is the Knonecker product of the $n_1 \times n_1$ symmetric Toeplitz matrix generated by the k_1-th standard basis vector, and the $n_2 \times n_2$ symmetric Toeplitz matrix generated by the k_2-th standard basis vector; and $\delta_{\boldsymbol{k}} = 1$ if $\boldsymbol{k} = (0,0)$, and $\delta_{\boldsymbol{k}} = 0$ otherwise. This is exactly the first-order relaxation in sum-of-squares relaxation hierarchy proposed in [35] for the precise SDP characterization of (19). Therefore, one can also employ the checking mechanism proposed in [35] to determine if (16) is exact.

V. NUMERICAL SIMULATIONS

We present numerical examples to verify the performance of the proposed algorithm (16) for a data matrix $\boldsymbol{X}^\star$ of size $n_1 \times n_2$. In the first example, let $n_1 = n_2 = 10$. We randomly generated $r = 6$ frequency pairs in $[0, 1) \times [0, 1)$, with

$$\Omega = [(0.1537, 0.5181), (0.2810, 0.9436), (0.4401, 0.6377),$$
$$(0.5271, 0.9577), (0.4574, 0.2407), (0.8754, 0.6761)],$$

where the coefficient of each frequency was generated with constant magnitude one and a random phase from $\mathcal{U}[0, 2\pi]$. In typical applications of interest such as radar or channel estimation, these frequency pairs correspond to delay, Doppler and amplitudes of the scatters. The actual frequency locations are depicted in Fig. 1 (a). Each entry in $\boldsymbol{X}^\star$ was observed with probability $p = m/(n_1 n_2)$, with $m = 30$, which can be collected using the sub-Nyquist sampling framework described in [38]. We then implemented (16) using CVX [39]. Notice that the number of unknown parameters was $3r = 18$. Fig. 1 (b) shows the recovered frequency locations using basis pursuit (BP) by assuming the signal is sparse in a DFT basis, and Fig. 1 (c) shows the recovered frequency locations using BP by assuming the signal is sparse in a DFT frame oversampled by a factor of 4. Finally, the recovered frequency locations using MEMP [11] from the data matrix recovered from (16) are depicted in Fig. 1 (d), superimposed on the ground truth. The reconstruction is perfect using the proposed approach when the data is noise-free.

(a) Ground Truth

(b) BP (DFT Basis)

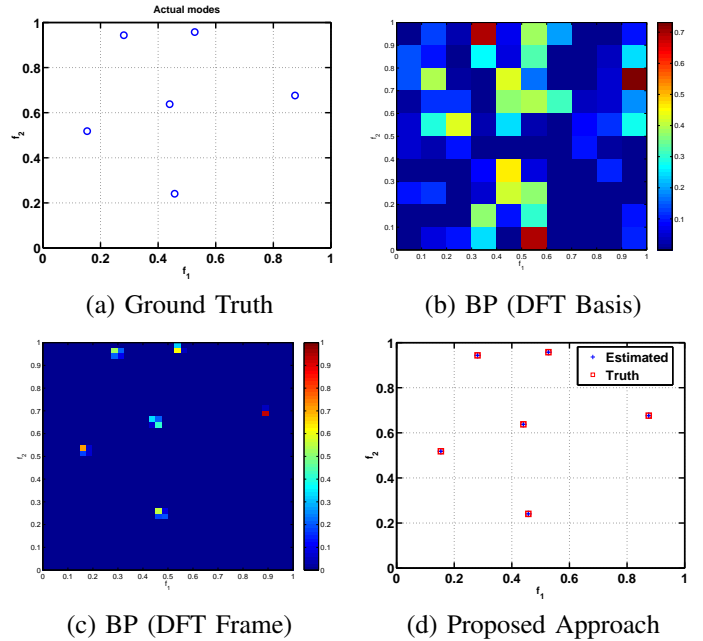(c) BP (DFT Frame)

(d) Proposed Approach

Fig. 1. The recovered signal in the frequency domain for $n_1 = n_2 = 10$ from $m = 30$ measurements.

We also examine the phase transition of the proposed algorithm (16). Let $n_1 = n_2 = 8$. For each pair of m and the number of modes r, we ran 10 experiments, where in each experiment r complex sinusoids (a) are generated randomly, or (b) generated randomly until a separation condition of $\Delta_{\min} = 1.5/n_1$ is satisfied. The recovery was claimed successful if the normalized mean squared error (NMSE) error $\|\hat{\boldsymbol{x}} - \boldsymbol{x}^\star\|_2 / \|\boldsymbol{x}^\star\|_2 \leq 10^{-3}$, where $\hat{\boldsymbol{x}}$ was the reconstructed data. Fig. 2 shows the success rate for each pair of m and r, with the grayscale of each cell reflecting the empirical rate of success, for the two cases described above respectively in (a) and (b). Fig. 2 (b) has a much sharper phase transition compared with (a), indicating that the number of samples grows approximately linearly with respect to r when the separation condition is imposed, in line with our theoretical

analysis.

We further compare the proposed algorithm with the EMaC algorithm proposed in [33] by setting the pencil parameters therein to be $k_1 = 4$ and $k_2 = 5$ respectively, which yields a two-fold Hankel matrix of size 20×20 to be completed. Fig. 2 (c) shows the success rate of EMaC for each pair of m and r under the same condition as Fig. 2 (a) when the frequencies are randomly generated. Our numerical examples also indicate that unlike the atomic norm approach, the phase transition curve of EMaC is insensitive to the separation condition. While the EMaC algorithm yields a much sharper phase transition than that of the proposed algorithm for randomly generated frequencies, the range of its recoverable (m, r) is much smaller, partially due to the small dimensionality of the relevant two-fold Hankel matrix when the data size is small.

We further examine the performance of (16) in the presence of noise. The noisy data was generated as

$$\boldsymbol{X}_{\text{noisy}} = \boldsymbol{X}^\star + \sigma \boldsymbol{N},$$

where $\boldsymbol{X}^\star$ was generated in the same way as in Fig. 2 with $r = 3$ different frequencies, and $\boldsymbol{N}$ was standard additive white Gaussian noise (AWGN) with each entry i.i.d. from $\mathcal{N}(0, 1)$. The signal-to-noise ratio is defined as $\text{SNR} = 10 \log_{10} \frac{m}{n\sigma^2}$, which has been scaled with respect to the number of samples. The proposed algorithm was modified to incorporate noise as

$$\{\hat{\boldsymbol{T}}, \hat{\boldsymbol{x}}, t\} = \operatorname{argmin}_{\boldsymbol{T}, \boldsymbol{x}, t} \frac{1}{2} \text{tr}\left(\mathcal{S}(\boldsymbol{T})\right) + \frac{1}{2} t \qquad (17)$$

$$\text{subject to} \quad \begin{bmatrix} \mathcal{S}(\boldsymbol{T}) & \boldsymbol{x} \\ \boldsymbol{x}^* & t \end{bmatrix} \succeq 0, \quad \|\mathcal{P}_T(\boldsymbol{x} - \boldsymbol{x}^\star)\|_2 \leq \sqrt{m}\sigma.$$

Fig. 3 illustrates the NMSE against SNR under different sample complexity. When $m = 20$, there is a possibility of failure that $\boldsymbol{x}^\star$ is not perfectly recovered even without noise, so the NMSE is relatively large under all SNRs. When $m = 30, 40$, and 50, it is with high possibility that $\boldsymbol{x}^\star$ is perfectly recovered without noise. When this was the case, the performance degenerated gracefully as the SNR decreases. The performance also improved when the number of samples m increases, but the gain was not as significant as long as it is above certain threshold.

VI. CONCLUSIONS

In this paper we explore estimation of 2-D frequency components of a spectrally sparse signal, when we are given a random subset of its regularly spaced samples. We formulate an atomic norm minimization problem, and show that a sample size of $\mathcal{O}(r \log r \log n)$ is sufficient to guarantee perfect frequency recovery, provided that a mild separation condition is satisfied. Our work can be extended to an arbitrary higher dimension and a similar semidefinite program can be proposed using a multi-fold block Toeplitz matrix constructed similar to (14). Finally, it remains to be seen how to develop more efficient first-order algorithms in solving the semidefinite program (16), as generic SDP solvers based on interior point methods are limited to small-dimensionality problems.
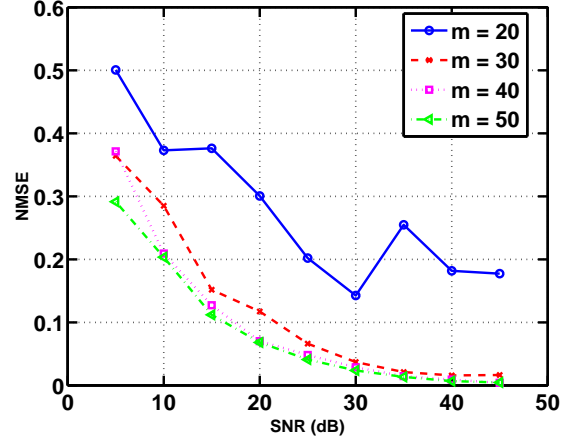

Fig. 3. NMSE vs SNR when $n_1 = n_2 = 8$, where frequency locations are generated randomly with a separation condition $\Delta_{\min} = 1.5/n_1$ for different numbers of samples.

APPENDIX A
USEFUL LEMMAS

We first present a few useful inequalities that will be used in the proofs.

Lemma 1. *(Noncommutative Bernstein's inequality) [40] Let $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_L$ be independent zero-mean symmetric random matrices of dimension $d \times d$. Suppose $\sigma^2 = \left\| \sum_{k=1}^{L} \mathbb{E}[\boldsymbol{X}_k \boldsymbol{X}_k^\top] \right\|$ and $\|\boldsymbol{X}_k\| \leq B$ almost surely for all k. Then for any $0 < t < \sigma^2/B$,*

$$\Pr\left[\left\| \sum_{k=1}^{L} \boldsymbol{X}_k \right\| > t \right] \leq 2d \exp\left(-\frac{3t^2}{8\sigma^2} \right). \qquad (18)$$

Lemma 2. *(Talagrand's concentration inequality) [41] Let $\{Y_j\}$ be a finite sequence of independent random variables taking values in a Banach space and V be defined as*

$$V = \sup_{h \in \mathcal{H}} \sum_j h(Y_j)$$

for a countable family of real valued functions $\mathcal{H}$. Assume that $|h| \leq B$ and $\mathbb{E}[h(Y_j)] = 0$ for all $h \in \mathcal{H}$ and every j. Then for all $t > 0$,

$$\Pr\left(|V - \mathbb{E}[V]| > t\right)$$
$$\leq 16 \exp\left(-\frac{t}{KB} \log\left(1 + \frac{Bt}{\sigma^2 + B\mathbb{E}[\bar{V}]} \right) \right),$$

where $\sigma^2 = \sup_{h \in \mathcal{H}} \sum_j \mathbb{E}[h^2(Y_j)]$, $\bar{V} = \sup_{h \in \mathcal{H}} \left| \sum_j h(Y_j) \right|$, and K is a numerical constant.

Lemma 3. *(Hoeffding's inequality) [42] Let the components of $\boldsymbol{u} \in \mathbb{C}^n$ be sampled i.i.d. from a symmetric distribution on the complex unit circle, $\boldsymbol{w} \in \mathbb{C}^n$, and t be a positive real number. Then*

$$\Pr\left(|\langle \boldsymbol{u}, \boldsymbol{w} \rangle| \geq t\right) \leq 4 \exp\left(-\frac{t^2}{4\|\boldsymbol{w}\|_2^2} \right).$$
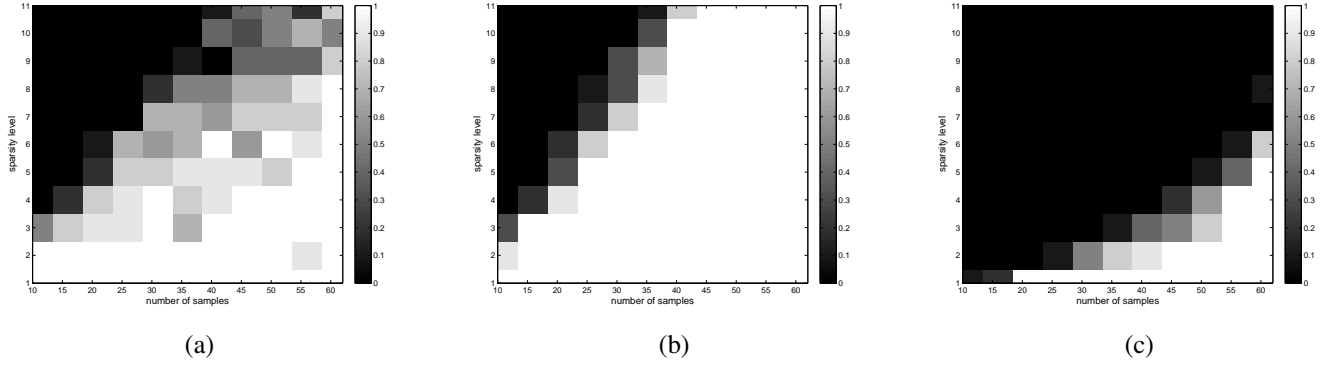
Fig. 2. Phase transition plots when $n_1 = n_2 = 8$: (a) the proposed algorithm with randomly generated frequencies; (b) the proposed algorithm with randomly generated frequencies satisfying a separation condition $\Delta_{\min} = 1.5/n_1$; (c) the EMaC algorithm [33] with randomly generated frequencies. The success rate is calculated by averaging over 10 runs.

APPENDIX B
PROOF OF THEOREM 1

This section is dedicated to the proof of Theorem 1 when the signs of d_i's are randomly drawn from $\{+1, -1\}$. The proof is similar for the case where d_i's are complex-valued, following the discussions in [28, Section 1.3]. A road map of the proof is given below. We will first characterize properties of a dual polynomial that suffices to certify the optimality and uniqueness of the solution to (9), and then present a randomized dual construction scheme. Specifically, the construction scheme produces a polynomial by randomizing the dual polynomial in [28] constructed for the full-observation case. Finally, we will show that this random polynomial satisfies the optimality and uniqueness conditions with high probability.

A. Optimality Conditions for Dual Polynomial

The dual norm of $\|\boldsymbol{x}\|_{\mathcal{A}}$ is defined as

$$\|\boldsymbol{q}\|_{\mathcal{A}}^* = \sup_{\|\boldsymbol{x}\|_{\mathcal{A}} \leq 1} \mathrm{Re}(\langle \boldsymbol{q}, \boldsymbol{x} \rangle) = \sup_{\boldsymbol{f} \in [0,1) \times [0,1)} |\langle \boldsymbol{q}, \boldsymbol{c}(\boldsymbol{f}) \rangle|,$$

where $\boldsymbol{f} = (f_1, f_2)$. As a result, the dual problem associated with (9) is given by

$$\max_{\boldsymbol{q}} \mathrm{Re}(\langle \boldsymbol{q}, \boldsymbol{x}^\star \rangle) \quad \text{subject to } \|\boldsymbol{q}\|_{\mathcal{A}}^* \leq 1, \ \boldsymbol{q}_{T^c} = \boldsymbol{0}, \quad (19)$$

where $T^c = J \backslash T$ is the complement set of T. Following standard analysis (see [28]), the optimal solution of (9) is unique if there exists a dual polynomial

$$Q(\boldsymbol{f}) = \sum_{\boldsymbol{k} \in J} q_{\boldsymbol{k}} e^{-j2\pi \boldsymbol{f}^\top \boldsymbol{k}}$$

satisfying the following set of conditions

$$q_{\boldsymbol{k}} = 0, \quad \forall \boldsymbol{k} \notin T, \quad (20a)$$
$$Q(\boldsymbol{f}_i) = \mathrm{sign}(d_i), \quad \forall \boldsymbol{f}_i \in \Omega, \quad (20b)$$
$$|Q(\boldsymbol{f})| < 1, \quad \forall \boldsymbol{f} \notin \Omega, \quad (20c)$$

where $\mathrm{sign}(\cdot)$ represents the complex sign. In the sequel we will produce a dual polynomial satisfying the conditions (20a)-(20c) with high probability.

B. Fejr's Kernel

In [28], the dual polynomial is constructed from the squared Fejr's kernel [43], which is defined in the 1-D setting as

$$K(f) = \frac{1}{M} \sum_{k=-2M}^{2M} g_M(k) e^{-j2\pi fk} = \left[\frac{\sin(\pi M f)}{M \sin(\pi f)} \right]^4,$$

for $f \in [0, 1)$, where

$$g_M(k) = \frac{1}{M} \sum_{s=\max(k-M,-M)}^{\min(k+M,M)} \left(1 - \left| \frac{s}{M} \right| \right) \left(1 - \left| \frac{k-s}{M} \right| \right).$$

Two important features of $K(f)$ are worth mentioning: 1) it is nonnegative, and 2) it exhibits rapid decay to zero as f grows. We note that $\|g_M\|_\infty = \max_k |g_M(k)| \leq 1$, which will be useful in later analysis.

In the 2-D setting, the corresponding Fejr's kernel is defined as

$$K(\boldsymbol{f}) = K(f_1) K(f_2) = \frac{1}{M^2} \sum_{\boldsymbol{k} \in J} g_M(k_1) g_M(k_2) e^{-j2\pi \boldsymbol{f}^\top \boldsymbol{k}}$$

for $\boldsymbol{f} = [f_1, f_2] \in [0, 1) \times [0, 1)$. Let $K^{(i_1 i_2)}(\boldsymbol{f})$ be the partial derivative of $K(\boldsymbol{f})$ given by

$$K^{(i_1 i_2)}(\boldsymbol{f}) = \frac{\partial^{i_1} \partial^{i_2} K(\boldsymbol{f})}{\partial f_1^{i_1} f_2^{i_2}}.$$

C. Construction of Dual Polynomial

Following the argument in [29, Section IV-B], it is sufficient to consider a Bernoulli observation model such that each entry in J is observed with probability

$$p = \frac{m}{M^2}.$$

Specifically, we assign an i.i.d. Bernoulli random variable $\delta_{\boldsymbol{k}}$ to indicate whether the $\boldsymbol{k}$th entry is observed, which satisfies

$$\mathrm{Pr}(\delta_{\boldsymbol{k}} = 1) = p = \frac{m}{M^2} < 1. \quad (21)$$

Define a randomized 2-D Fejr's kernel as

$$K_{\mathrm{r}}(\boldsymbol{f}) = \frac{1}{M^2} \sum_{\boldsymbol{k} \in J} g_M(k_1) g_M(k_2) \delta_{\boldsymbol{k}} e^{-j2\pi \boldsymbol{f}^\top \boldsymbol{k}}, \quad (22)$$

where $\delta_{\boldsymbol{k}}$ is defined in (21). Let $K_{\mathrm{r}}^{(i_1 i_2)}(\boldsymbol{f})$ be the partial derivative of $K_{\mathrm{r}}(\boldsymbol{f})$ as

$$K_{\mathrm{r}}^{(i_1 i_2)}(\boldsymbol{f}) = \frac{\partial^{i_1} \partial^{i_2} K_{\mathrm{r}}(\boldsymbol{f})}{\partial f_1^{i_1} f_2^{i_2}}$$

for any $i_1, i_2 = 0, 1, 2$. Then their expected values with respect to $\delta_{\boldsymbol{k}}$ can be computed as

$$\mathbb{E}\left[K_{\mathrm{r}}^{(i_1 i_2)}(\boldsymbol{f})\right] = p K^{(i_1 i_2)}(\boldsymbol{f}).$$

We propose to construct the dual polynomial of (19) as

$$Q(\boldsymbol{f}) = \sum_{i=1}^{r} \alpha_i K_{\mathrm{r}}(\boldsymbol{f} - \boldsymbol{f}_i) + \sum_{i=1}^{r} \beta_{1i} K_{\mathrm{r}}^{(10)}(\boldsymbol{f} - \boldsymbol{f}_i) +$$
$$\sum_{i=1}^{r} \beta_{2i} K_{\mathrm{r}}^{(01)}(\boldsymbol{f} - \boldsymbol{f}_i), \qquad (23)$$

i.e. a superposition of the randomized Fejr's kernel and its first-order partial derivatives at the frequencies in $\mathcal{F}$.

To establish Theorem 1, we need to verify that $Q(\boldsymbol{f})$ in the form of (23) satisfies the hypotheses (20a)-(20c) with high probability. Apparently, Condition (20a) is satisfied by the randomized construction scheme. The next step is to tweak the interpolation coefficients $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_r]^\top$, $\boldsymbol{\beta}_1 = [\beta_{11}, \ldots, \beta_{1r}]^\top$ and $\boldsymbol{\beta}_2 = [\beta_{21}, \ldots, \beta_{2r}]^\top$ to satisfy (20b). Specifically, let the (ℓ, i)th entry of $\boldsymbol{E}_{i_1 i_2}$ be

$$(\boldsymbol{E}_{i_1 i_2})_{\ell i} = Q^{(i_1 i_2)}(\boldsymbol{f}_\ell - \boldsymbol{f}_i),$$

where $Q^{(i_1 i_2)}(\boldsymbol{f})$ represents the partial derivative of $Q(\boldsymbol{f})$. Choose the coefficients $\boldsymbol{\alpha}$, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ such that

$$\begin{bmatrix} \boldsymbol{E}_{00} & \kappa \boldsymbol{E}_{10} & \kappa \boldsymbol{E}_{01} \\ -\kappa \boldsymbol{E}_{10} & -\kappa^2 \boldsymbol{E}_{20} & -\kappa^2 \boldsymbol{E}_{11} \\ -\kappa \boldsymbol{E}_{01} & -\kappa^2 \boldsymbol{E}_{11} & -\kappa^2 \boldsymbol{E}_{02} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \kappa^{-1} \boldsymbol{\beta}_1 \\ \kappa^{-1} \boldsymbol{\beta}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \quad (24)$$

where[1] $\kappa^{-1} = \sqrt{K''(0)}$, and $\boldsymbol{v} \in \mathbb{R}^r$ obeys $v_k = \mathrm{sign}(d_k)$. Denote by $\boldsymbol{E}$ the matrix on the left-hand side of the above equation

$$\boldsymbol{E} = \begin{bmatrix} \boldsymbol{E}_{00} & \kappa \boldsymbol{E}_{10} & \kappa \boldsymbol{E}_{01} \\ -\kappa \boldsymbol{E}_{10} & -\kappa^2 \boldsymbol{E}_{20} & -\kappa^2 \boldsymbol{E}_{11} \\ -\kappa \boldsymbol{E}_{01} & -\kappa^2 \boldsymbol{E}_{11} & -\kappa^2 \boldsymbol{E}_{02} \end{bmatrix} \in \mathbb{C}^{3r \times 3r},$$

whose expected value is given by $\mathbb{E}[\boldsymbol{E}] = p\bar{\boldsymbol{E}}$. Here, $\bar{\boldsymbol{E}}$ denotes

$$\bar{\boldsymbol{E}} = \begin{bmatrix} \bar{\boldsymbol{E}}_{00} & \kappa \bar{\boldsymbol{E}}_{10} & \kappa \bar{\boldsymbol{E}}_{01} \\ -\kappa \bar{\boldsymbol{E}}_{10} & -\kappa^2 \bar{\boldsymbol{E}}_{20} & -\kappa^2 \bar{\boldsymbol{E}}_{11} \\ -\kappa \bar{\boldsymbol{E}}_{01} & -\kappa^2 \bar{\boldsymbol{E}}_{11} & -\kappa^2 \bar{\boldsymbol{E}}_{02} \end{bmatrix}$$

with each sub-block defined with the (ℓ, i)th entry being

$$(\bar{\boldsymbol{E}}_{i_1 i_2})_{\ell i} = \bar{Q}^{(i_1 i_2)}(\boldsymbol{f}_\ell - \boldsymbol{f}_i),$$

where $\bar{Q}^{(i_1 i_2)}(\boldsymbol{f}) := \mathbb{E}[Q^{(i_1 i_2)}(\boldsymbol{f})]$. In order to find a solution to (24), one first needs to demonstrate that $\boldsymbol{E}$ is invertible. To this end, we begin by presenting the following lemma, whose proof is given in Appendix C.

Lemma 4. *Under the conditions of Theorem 1, one has*

$$\|\boldsymbol{I} - \bar{\boldsymbol{E}}\| \le 0.1982.$$

[1] $K''(0)$ is the second-order derivative of $K(f)$ at $f = 0$.

Lemma 4 immediately implies that $\bar{\boldsymbol{E}}$ is invertible,

$$\|\bar{\boldsymbol{E}}\| \le 1 + \|\boldsymbol{I} - \bar{\boldsymbol{E}}\| \le 1.1982, \qquad (25)$$

and

$$\left\|\bar{\boldsymbol{E}}^{-1}\right\| \le \frac{1}{1 - \|\boldsymbol{I} - \bar{\boldsymbol{E}}\|} \le 1.2472.$$

The matrix $\boldsymbol{E}$ can then be expressed as

$$\boldsymbol{E} = \frac{1}{M^2} \sum_{\boldsymbol{k} \in J} g_M(k_1) g_M(k_2) \delta_{\boldsymbol{k}} \boldsymbol{e}_{\boldsymbol{k}} \boldsymbol{e}_{\boldsymbol{k}}^*, \qquad (26)$$

where $\boldsymbol{e}_{\boldsymbol{k}}$ is given by

$$\boldsymbol{e}_{\boldsymbol{k}} = \begin{bmatrix} 1 \\ j2\pi\kappa k_1 \\ j2\pi\kappa k_2 \end{bmatrix} \otimes \begin{bmatrix} e^{-j2\pi \boldsymbol{f}_1^\top \boldsymbol{k}} \\ \vdots \\ e^{-j2\pi \boldsymbol{f}_r^\top \boldsymbol{k}} \end{bmatrix} \in \mathbb{C}^{3r}. \qquad (27)$$

Similarly, one can write $\bar{\boldsymbol{E}}$ as

$$\bar{\boldsymbol{E}} = \frac{1}{M^2} \sum_{\boldsymbol{k} \in J} g_M(k_1) g_M(k_2) \boldsymbol{e}_{\boldsymbol{k}} \boldsymbol{e}_{\boldsymbol{k}}^*. \qquad (28)$$

We will establish that the spectral norm of $p^{-1}\boldsymbol{E}$ can be well controlled, as stated in the following lemma. The proof is deferred to Appendix D.

Lemma 5. *Let $0 \le \delta \le 1$. If $\Delta_{\min} \ge \frac{1.19}{M}$, and $m \ge c_1 r \log\left(\frac{6r}{\delta}\right)$ for some positive constant $c_1 > 0$, then with probability at least $1 - \delta$,*

$$\left\|p^{-1}\boldsymbol{E} - \bar{\boldsymbol{E}}\right\| \le \frac{1}{4}.$$

Denote by $\mathcal{E}_1$ the event $\mathcal{E}_1 = \left\{\|p^{-1}\boldsymbol{E} - \bar{\boldsymbol{E}}\| \le \frac{1}{4}\right\}$, which occurs with probability $\Pr(\mathcal{E}_1) \ge 1 - \delta$. Conditional on $\mathcal{E}_1$, one has

$$\left\|\boldsymbol{I} - p^{-1}\boldsymbol{E}\right\| \le \|\boldsymbol{I} - \bar{\boldsymbol{E}}\| + \|p^{-1}\boldsymbol{E} - \bar{\boldsymbol{E}}\| < 1, \qquad (29)$$

revealing the invertibility of $\boldsymbol{E}$. Writing $\boldsymbol{E}^{-1} = [\boldsymbol{L} \quad \boldsymbol{R}]$ for some $\boldsymbol{L} \in \mathbb{C}^{3r \times r}$, we have

$$\left\|\boldsymbol{L} - p^{-1}\bar{\boldsymbol{L}}\right\| \le \frac{1}{2} p^{-1} \left\|\bar{\boldsymbol{E}}^{-1}\right\|^2, \quad \|\boldsymbol{L}\| \le 2p^{-1}\left\|\bar{\boldsymbol{E}}^{-1}\right\|,$$

where $\bar{\boldsymbol{L}} = p\mathbb{E}[\boldsymbol{L}]$ follows from [29, Corollary IV.5]. The interpolation coefficients can thus be written as

$$\begin{bmatrix} \boldsymbol{\alpha} & \kappa^{-1}\boldsymbol{\beta}_1 & \kappa^{-1}\boldsymbol{\beta}_2 \end{bmatrix}^\top = \boldsymbol{L}\boldsymbol{v}. \qquad (30)$$

With this choice, (20b) is satisfied trivially.

D. Verification of (20c)

What remains to be estalibshed is Condition (20c). We will first show that it holds on a regular grid $\Omega_{\mathrm{g}} \subset [0,1) \times [0,1)$, and then extends it to the continuous domain.

To proceed, define $\boldsymbol{w}^{(i_1 i_2)}(\boldsymbol{f})$ as

$$
\boldsymbol{w}^{(i_1 i_2)}(\boldsymbol{f}) = \kappa^{i_1+i_2}
\begin{bmatrix}
K_{\mathrm{r}}^{(i_1 i_2)}(\boldsymbol{f} - \boldsymbol{f}_1)^* \\
\vdots \\
K_{\mathrm{r}}^{(i_1 i_2)}(\boldsymbol{f} - \boldsymbol{f}_r)^* \\
\kappa K_{\mathrm{r}}^{(i_1+1,i_2)}(\boldsymbol{f} - \boldsymbol{f}_1)^* \\
\vdots \\
\kappa K_{\mathrm{r}}^{(i_1+1,i_2)}(\boldsymbol{f} - \boldsymbol{f}_r)^* \\
\kappa K_{\mathrm{r}}^{(i_1,i_2+1)}(\boldsymbol{f} - \boldsymbol{f}_1)^* \\
\vdots \\
\kappa K_{\mathrm{r}}^{(i_1,i_2+1)}(\boldsymbol{f} - \boldsymbol{f}_r)^*
\end{bmatrix}
$$

$$
= \frac{1}{M^2} \sum_{\boldsymbol{k} \in J} (j2\pi\kappa)^{i_1+i_2} k_1^{i_1} k_2^{i_2} g_M(k_1) g_M(k_2) e^{j2\pi \boldsymbol{f}^\top \boldsymbol{k}} \delta_{\boldsymbol{k}} \boldsymbol{e}_{\boldsymbol{k}},
\tag{31}
$$

then the derivatives of the dual polynomial $Q(\boldsymbol{f})$ in (23) can be written as

$$
\kappa^{i_1+i_2} Q^{(i_1 i_2)}(\boldsymbol{f}) = (\boldsymbol{w}^{(i_1 i_2)}(\boldsymbol{f}))^* \boldsymbol{L} \boldsymbol{v}.
$$

Define the mean of $\boldsymbol{w}^{(i_1 i_2)}(\boldsymbol{f})$ such that

$$
\mathbb{E}\left[\boldsymbol{w}^{(i_1 i_2)}(\boldsymbol{f})\right] = p\bar{\boldsymbol{w}}^{(i_1 i_2)}(\boldsymbol{f}).
$$

We have

$$
\begin{aligned}
&(\boldsymbol{w}^{(i_1 i_2)}(\boldsymbol{f}))^* \boldsymbol{L} \boldsymbol{v} \\
&= (\bar{\boldsymbol{w}}^{(i_1 i_2)}(\boldsymbol{f}))^* \bar{\boldsymbol{L}} \boldsymbol{v} + p(\bar{\boldsymbol{w}}^{(i_1 i_2)}(\boldsymbol{f}))^* [\boldsymbol{L} - p^{-1}\bar{\boldsymbol{L}}] \boldsymbol{v} + \\
&\quad [\boldsymbol{w}^{(i_1 i_2)}(\boldsymbol{f}) - p\bar{\boldsymbol{w}}^{(i_1 i_2)}(\boldsymbol{f})]^* \boldsymbol{L} \boldsymbol{v} \\
&= \kappa^{i_1+i_2} \bar{Q}^{(i_1 i_2)}(\boldsymbol{f}) + I_1^{(i_1 i_2)}(\boldsymbol{f}) + I_2^{(i_1 i_2)}(\boldsymbol{f}),
\end{aligned}
$$

where

$$
I_1^{(i_1 i_2)}(\boldsymbol{f}) = p(\bar{\boldsymbol{w}}^{(i_1 i_2)}(\boldsymbol{f}))^* [\boldsymbol{L} - p^{-1}\bar{\boldsymbol{L}}] \boldsymbol{v},
$$

$$
I_2^{(i_1 i_2)}(\boldsymbol{f}) := [\boldsymbol{w}^{(i_1 i_2)}(\boldsymbol{f}) - p\bar{\boldsymbol{w}}^{(i_1 i_2)}(\boldsymbol{f})]^* \boldsymbol{L} \boldsymbol{v}.
$$

We first need to establish that $I_1^{(i_1 i_2)}(\boldsymbol{f})$ and $I_2^{(i_1 i_2)}(\boldsymbol{f})$ can be controlled uniformly over all $\boldsymbol{f}$. To this end, we apply similar techniques adopted in [29], which first bound $I_1^{(i_1 i_2)}(\boldsymbol{f})$ and $I_2^{(i_1 i_2)}(\boldsymbol{f})$ on a regular grid Ω_{g}, and then extend the result to all frequencies. The following lemma quantifies the perturbations on a regular grid, whose proof is deferred to Appendix E.

Lemma 6. *Suppose $\Delta_{\min} \geq \frac{1.19}{M}$. For a regular grid Ω_g, there exists a numerical constant C such that if*

$$
m \geq C \frac{1}{\epsilon^2} \max\left\{\log^2 \frac{|\Omega_g|}{\delta}, r \log \frac{r}{\delta} \log \frac{|\Omega_g|}{\delta}\right\},
\tag{32}
$$

then $\sup_{\boldsymbol{f} \in \Omega_g} |I_1^{(i_1 i_2)}(\boldsymbol{f})| \leq \epsilon/2$ and $\sup_{\boldsymbol{f} \in \Omega_g} |I_2^{(i_1 i_2)}(\boldsymbol{f})| \leq \epsilon/2$ for $i_1, i_2 = 0,1,2,3$ with probability at least $1 - \delta$.

Following Lemma 6, we immediately show that the event

$$
\mathcal{E}_2 = \left\{ \sup_{\boldsymbol{f} \in \Omega_{\mathrm{g}}} \kappa^{i_1+i_2} \left|Q^{(i_1 i_2)}(\boldsymbol{f}) - \bar{Q}^{(i_1 i_2)}(\boldsymbol{f})\right| \leq \epsilon, \right.
$$
$$
\left. 0 \leq i_1, i_2 \leq 3 \right\}.
$$

occurs with probability at least $1 - \delta$ on the grid Ω_{g}.

We will first extend $\mathcal{E}_2$ to the whole continuous domain by the following lemma whose proof can be found in Appendix F.

Lemma 7. *Suppose that $\Delta_{\min} \geq \frac{1.19}{M}$. There exists a numerical constant $C > 0$ such that if m satisfies (32) for some constant C, then*

$$
\kappa^{i_1+i_2} \left|Q^{(i_1 i_2)}(\boldsymbol{f}) - \bar{Q}^{(i_1 i_2)}(\boldsymbol{f})\right| \leq \epsilon, \quad i_1, i_2 = 0,1,2,3
\tag{33}
$$

for $\boldsymbol{f} \in [0,1) \times [0,1)$ with probability at least $1 - \delta$.

Finally, we can establishes (20c) through the following lemma, where the proof is supplied in Appendix G.

Lemma 8. *Suppose that $\Delta_{\min} \geq \frac{1.19}{M}$. There exists a universal constant $C > 0$ such that if*

$$
m \geq C \max\left\{\log^2 \frac{M}{\delta}, r \log \frac{r}{\delta} \log \frac{M}{\delta}\right\},
\tag{34}
$$

then with probability at least $1 - \delta$, one has $|Q(\boldsymbol{f})| < 1$ for $\boldsymbol{f} \notin \Omega$.

Combining the above lemmas, we have successfully constructed a dual polynomial $Q(\boldsymbol{f})$ when m satisfies (34), completing the proof of Theorem 1.

APPENDIX C
PROOF OF LEMMA 4

Proof. When $\Delta_{\min} \geq 1.19/M$, using the result in [28, Proof of Lemma C.2], we have

$$
\begin{aligned}
\|\boldsymbol{I} - \bar{\boldsymbol{E}}_0\|_\infty &\leq 4.854 \cdot 10^{-2}, \\
\|\kappa\bar{\boldsymbol{E}}_{10}\|_\infty &\leq 4.2580 \cdot 10^{-2} \\
\|\kappa^2\bar{\boldsymbol{E}}_{11}\|_\infty &\leq 4.7905 \cdot 10^{-2}, \\
\|\boldsymbol{I} - \kappa^2\bar{\boldsymbol{E}}_{20}\|_\infty &\leq 0.1076,
\end{aligned}
$$

where $\|\cdot\|_\infty$ is the matrix infinity norm, i.e. the maximum absolute row sum. Since $\boldsymbol{I} - \bar{\boldsymbol{E}}$ is symmetric and its diagonal entries are all zero, by the Gershgorin's circle theorem [44],

$$
\begin{aligned}
\|\boldsymbol{I} - \bar{\boldsymbol{E}}\| &\leq \|\boldsymbol{I} - \bar{\boldsymbol{E}}\|_\infty \\
&\leq \max\Big\{ \|\boldsymbol{I} - \bar{\boldsymbol{E}}_0\|_\infty + \|\kappa\bar{\boldsymbol{E}}_{01}\|_\infty + \|\kappa\bar{\boldsymbol{E}}_{10}\|_\infty, \\
&\quad \|\kappa\bar{\boldsymbol{E}}_{10}\|_\infty + \|\kappa^2\bar{\boldsymbol{E}}_{11}\|_\infty + \|\boldsymbol{I} - \kappa^2\bar{\boldsymbol{E}}_{20}\|_\infty \Big\} \leq 0.1982.
\end{aligned}
$$

$\square$

APPENDIX D
PROOF OF LEMMA 5

Proof. First, write $\boldsymbol{E} - p\bar{\boldsymbol{E}} = \sum_{\boldsymbol{k} \in J} \boldsymbol{G}_{\boldsymbol{k}}$, where

$$
\boldsymbol{G}_{\boldsymbol{k}} = \frac{1}{M^2} g_M(k_1) g_M(k_2)(\delta_{\boldsymbol{k}} - p)\boldsymbol{e}_{\boldsymbol{k}}\boldsymbol{e}_{\boldsymbol{k}}^*
$$

is a random zero-mean self-adjoint matrix. We would like to apply Lemma 1. Since $\mathbb{E}\boldsymbol{G}_{\boldsymbol{k}} = 0$ and

$$
\begin{aligned}
B := \|\boldsymbol{G}_{\boldsymbol{k}}\| &= \frac{1}{M^2} \|g_M(k_1) g_M(k_2)(\delta_{\boldsymbol{k}} - p)\boldsymbol{e}_{\boldsymbol{k}}\boldsymbol{e}_{\boldsymbol{k}}^*\| \\
&\leq \frac{1}{M^2} \|g_M(k_1)\|_\infty \|g_M(k_2)\|_\infty \|\boldsymbol{e}_{\boldsymbol{k}}\|^2 \tag{35} \\
&\leq \frac{1}{M^2} 27r, \tag{36}
\end{aligned}
$$

where (35) follows from $|\delta_{\boldsymbol{k}} - p| \leq \{1 - p, p\} < 1$, (36) follows from $\|g_M\|_\infty \leq 1$ and

$$\|\boldsymbol{e_k}\|^2 \leq r(1 + 2 \max_{|k_1| \leq 2M} (2\pi k_1 \kappa)^2) \leq 27r$$

for $M \geq 4$ where the last inequality follows from $|2\pi k_1 \kappa|^2 \leq 13$ in [29]. And

$$\sigma^2 := \frac{1}{M^4} \left\| \sum_{\boldsymbol{k} \in J} g_M^2(k_1) g_M^2(k_2) \|\boldsymbol{e_k}\|^2 \boldsymbol{e_k} \boldsymbol{e_k}^* \mathbb{E}\left[(\delta_{\boldsymbol{k}} - p)^2\right] \right\|$$

$$\leq \frac{27rp(1-p)}{M^4} \left\| \sum_{\boldsymbol{k} \in J} g_M^2(k_1) g_M^2(k_2) \boldsymbol{e_k} \boldsymbol{e_k}^* \right\|$$

$$\leq \frac{27rp(1-p)}{M^2} \|g_M(k_1)\|_\infty \|g_M(k_2)\|_\infty \|\bar{\boldsymbol{E}}\|$$

$$\leq \frac{27rp(1-p)}{M^2} 1.1982$$

$$\leq \frac{33rp}{M^2},$$

where the last inequality follows from (25). Applying Lemma 1 and setting $t = \frac{1}{4}p$, one obtains

$$\Pr\left[\|p^{-1}\boldsymbol{E} - \bar{\boldsymbol{E}}\| > \frac{1}{4}\right] \leq 6r \exp\left(-\frac{m}{c_1}\right) \leq \delta \qquad (37)$$

for some constant c_1, leading to $m \geq c_1 r \log\left(\frac{6r}{\delta}\right)$. $\qquad \square$

APPENDIX E
PROOF OF LEMMA 6

Proof. We first write $\|\boldsymbol{w}^{(i_1 i_2)}(\boldsymbol{f}) - p\bar{\boldsymbol{w}}^{(i_1 i_2)}(\boldsymbol{f})\|_2$ as

$$\boldsymbol{w}^{(i_1 i_2)}(\boldsymbol{f}) - p\bar{\boldsymbol{w}}^{(i_1 i_2)}(\boldsymbol{f})$$

$$= \frac{1}{M^2} \cdot \sum_{\boldsymbol{k} \in J} (j2\pi\kappa)^{i_1+i_2} k_1^{i_1} k_2^{i_2} g_M(k_1) g_M(k_2)$$

$$\cdot e^{j2\pi \boldsymbol{f}^\top \boldsymbol{k}} (\delta_{\boldsymbol{k}} - p) \boldsymbol{e_k}$$

$$\triangleq \sum_{\boldsymbol{k} \in J} \boldsymbol{Y}_{\boldsymbol{k}}^{(i_1 i_2)},$$

where

$$\boldsymbol{Y}_{\boldsymbol{k}}^{(i_1 i_2)} = \frac{1}{M^2} (j2\pi\kappa)^{i_1+i_2} k_1^{i_1} k_2^{i_2}$$

$$\cdot g_M(k_1) g_M(k_2) e^{j2\pi \boldsymbol{f}^\top \boldsymbol{k}} (\delta_{\boldsymbol{k}} - p) \boldsymbol{e_k}.$$

Write

$$\|\boldsymbol{w}^{(i_1 i_2)}(\boldsymbol{f}) - p\bar{\boldsymbol{w}}^{(i_1 i_2)}(\boldsymbol{f})\|_2 = \sup_{\boldsymbol{h}: \|\boldsymbol{h}\|=1} \sum_{\boldsymbol{k} \in J} \langle \boldsymbol{Y}_{\boldsymbol{k}}^{(i_1 i_2)}, \boldsymbol{h} \rangle_\mathbb{R},$$

$$= \sup_{\boldsymbol{h}: \|\boldsymbol{h}\|=1} \sum_{\boldsymbol{k} \in J} h(\boldsymbol{Y}_{\boldsymbol{k}}^{(i_1 i_2)}),$$

where $h(\boldsymbol{Y}_{\boldsymbol{k}}^{(i_1 i_2)}) = \langle \boldsymbol{Y}_{\boldsymbol{k}}^{(i_1 i_2)}, \boldsymbol{h} \rangle_\mathbb{R}$. To apply Lemma 2, we compute

$$|h(\boldsymbol{Y}_{\boldsymbol{k}}^{(i_1 i_2)})| \leq \frac{1}{M^2} |2\pi k_1 \kappa|^{i_1} |2\pi k_2 \kappa|^{i_2} \|g_M(k_1)\|_\infty$$

$$\cdot \|g_M(k_2)\|_\infty \|\boldsymbol{e_k}\|_2$$

$$\leq \frac{1}{M^2} 4^{i_1+i_2+2} \sqrt{r},$$

and

$$\mathbb{E}\left[\left\| \boldsymbol{w}^{(i_1 i_2)}(\boldsymbol{f}) - p\bar{\boldsymbol{w}}^{(i_1 i_2)}(\boldsymbol{f}) \right\|_2^2\right]$$

$$\leq \sum_{\boldsymbol{k} \in J} \frac{1}{M^4} |2\pi k_1 \kappa|^{2i_1} |2\pi k_2 \kappa|^{2i_2}$$

$$\cdot g_M^2(k_1) g_M^2(k_2) p(1-p) \|\boldsymbol{e_k}\|_2^2$$

$$\leq \frac{mr}{M^4} 4^{2(i_1+i_2)+3},$$

when $M \geq 4$. Then from Jensen's inequality,

$$\mathbb{E}\left[\|\boldsymbol{w}^{(i_1 i_2)}(\boldsymbol{f}) - p\bar{\boldsymbol{w}}^{(i_1 i_2)}(\boldsymbol{f})\|_2\right] \leq \frac{\sqrt{mr}}{M^2} 2^{2(i_1+i_2)+3}.$$

We can then upper bound

$$\mathbb{E}\left[h^2(\boldsymbol{Y}_{\boldsymbol{k}}^{(i_1 i_2)})\right] \leq \frac{4^{2i_1+2i_2}}{M^4} \|g_M(k_1)\|_\infty \|g_M(k_2)\|_\infty$$

$$\cdot \mathbb{E}\left[(\delta_{\boldsymbol{k}} - p)^2\right] \left| \langle \sqrt{g_M(k_1) g_M(k_2)} \boldsymbol{e_k}, \boldsymbol{h} \rangle \right|^2$$

$$\leq \frac{4^{2i_1+2i_2}}{M^4} p \left| \langle \sqrt{g_M(k_1) g_M(k_2)} \boldsymbol{e_k}, \boldsymbol{h} \rangle \right|^2,$$

thereafter following similar argument as in [29, Proof of Lemma IV.6],

$$\sum_{\boldsymbol{k} \in J} \mathbb{E}\left[h^2(\boldsymbol{Y}_{\boldsymbol{k}}^{(i_1 i_2)})\right] \leq 4^{2(i_1+i_2)} \frac{m}{M^4} \|\bar{\boldsymbol{E}}\|$$

$$\leq 2^{4(i_1+i_2)+1} \frac{m}{M^4},$$

where we have used $p = m/M^2$ and $\|\bar{\boldsymbol{E}}\| \leq 2$. By Lemma 2, if we let

$$\bar{\sigma}_{i_1 i_2}^2 = \max\left\{ 2^{4(i_1+i_2)+1} \frac{m}{M^4}, \frac{r\sqrt{m}}{M^4} 2^{4(i_1+i_2)+7} \right\}$$

$$= 2 \cdot 16^{(i_1+i_2)} \frac{m}{M^2} \max\left\{ 1, 2^4 \frac{r}{\sqrt{m}} \right\}.$$

and

$$a \leq \begin{cases} \sqrt{2} m^{1/4} & \text{if } 2^4 r/\sqrt{m} \geq 1, \\ \frac{\sqrt{2}}{16} \sqrt{\frac{m}{s}} & \text{otherwise,} \end{cases}$$

then we have

$$\|\boldsymbol{w}^{(i_1 i_2)}(\boldsymbol{f}) - p\bar{\boldsymbol{w}}^{(i_1 i_2)}(\boldsymbol{f})\|_2 \leq \frac{\sqrt{mr}}{M^2} 2^{2(i_1+i_2)+3} + a\bar{\sigma}_{i_1 i_2},$$

with probability at least $1 - 64e^{-\gamma a^2}$. Consequently

$$\left\| [\boldsymbol{w}^{(i_1 i_2)}(\boldsymbol{f}) - p^2\bar{\boldsymbol{w}}^{(i_1 i_2)}(\boldsymbol{f})]^* \boldsymbol{L} \right\|_2$$

$$\leq \|\boldsymbol{L}\| \|\boldsymbol{w}^{(i_1 i_2)}(\boldsymbol{f}) - p^2\bar{\boldsymbol{w}}^{(i_1 i_2)}(\boldsymbol{f})\|$$

$$\leq 2p^{-1} \|\bar{\boldsymbol{E}}^{-1}\| \left(\frac{\sqrt{mr}}{M^2} 2^{2(i_1+i_2)+3} + a\bar{\sigma}_{i_1 i_2} \right)$$

$$\leq 4 \left(2^{2(i_1+i_2)+3} \sqrt{\frac{r}{m}} + \frac{M^2}{m} a\bar{\sigma}_{i_1 i_2} \right) \triangleq \lambda_{i_1 i_2}.$$

To bound $I_2^{(i_1 i_2)}(\boldsymbol{f})$ on the set $\boldsymbol{f} \in \Omega_g$ we use Hoeffding's inequality and the union bound (see [29, proof of Lemma IV.8]), which gives

$$\Pr\left(\sup_{\boldsymbol{f} \in \Omega_g} |I_2^{(i_1 i_2)}(\boldsymbol{f})| \leq \epsilon \right) \leq 9|\Omega_g| e^{-\frac{\epsilon^2}{4\lambda_{i_1 i_2}^2}} + 64|\Omega_g| e^{-\gamma a^2}$$

$$+ \Pr(\mathcal{E}_1^c), \qquad (38)$$

Therefore we can write $\boldsymbol{S}_0$ as

$$\boldsymbol{S}_0 = \boldsymbol{V}_3 \mathrm{diag}([|\boldsymbol{o}_1^*\tilde{\boldsymbol{\sigma}}|^2, \ldots, |\boldsymbol{o}_r^*\tilde{\boldsymbol{\sigma}}|^2])\boldsymbol{V}_3^*, \qquad (43)$$

and $\boldsymbol{V}_3 \in \mathbb{C}^{n_1 \times r}$ is a Vandermonde matrix with the ith column $\boldsymbol{a}(f_{2i}) = [1, z_i, \ldots, z_i^{n_1-1}]^T$. Then $\mathcal{S}(\boldsymbol{T})$ can be written as

$$\mathcal{S}(\boldsymbol{T}) = \sum_{i=1}^{r} |\boldsymbol{o}_i^*\tilde{\boldsymbol{\sigma}}|^2 (\boldsymbol{a}(f_{1i}) \otimes \boldsymbol{b}(f_{2i}))(\boldsymbol{a}(f_{1i}) \otimes \boldsymbol{b}(f_{2i}))^*$$
$$= \sum_{i=1}^{r} |\boldsymbol{o}_i^*\tilde{\boldsymbol{\sigma}}|^2 \boldsymbol{c}(\boldsymbol{f}_i)\boldsymbol{c}(\boldsymbol{f}_i)^*, \qquad (44)$$

where $\sigma_i = |\boldsymbol{o}_i^*\tilde{\boldsymbol{\sigma}}|^2$. Therefore we have established Proposition 2. $\qquad\square$

REFERENCES

[1] Y. Chi and Y. Chen, "Compressive recovery of 2-D off-grid frequencies," in *Proceedings of Asilomar Conference on Signals, Systems, and Computers*, Nov. 2013.

[2] D. Nion and N. Sidiropoulos, "Tensor algebra and multidimensional harmonic retrieval in signal processing for MIMO radar," *Signal Processing, IEEE Transactions on*, vol. 58, no. 11, pp. 5693–5705, Nov 2010.

[3] M. Skolnik, *Radar handbook*. New York: McGraw-Hill, 1970.

[4] A. M. Sayeed and B. Aazhang, "Joint multipath-doppler diversity in mobile wireless communications," *IEEE Transaction on Communications*, no. 1, pp. 123–132, Jan. 1999.

[5] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[6] M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," *Nature methods*, vol. 3, no. 10, pp. 793–796, 2006.

[7] Y. Chi, A. Gomaa, N. Al-Dhahir, and A. Calderbank, "Training signal design and tradeoffs for spectrally-efficient multi-user MIMO-OFDM systems," *Wireless Communications, IEEE Transactions on*, vol. 10, no. 7, pp. 2234–2245, 2011.

[8] M. Haardt, M. Zoltowski, C. Mathews, and J. Nossek, "2D unitary ESPRIT for efficient 2D parameter estimation," in *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on*, vol. 3. IEEE, 1995, pp. 2096–2099.

[9] Y. Hua, "A pencil-music algorithm for finding two-dimensional angles and polarizations using crossed dipoles," *Antennas and Propagation, IEEE Transactions on*, vol. 41, no. 3, pp. 370–376, 1993.

[10] M. Clark and L. Scharf, "Two-dimensional modal analysis based on maximum likelihood," *Signal Processing, IEEE Transactions on*, vol. 42, no. 6, pp. 1443–1452, 1994.

[11] Y. Hua, "Estimating two-dimensional frequencies by matrix enhancement and matrix pencil," *IEEE Transactions on Signal Processing*, vol. 40, no. 9, pp. 2267–2280, Sep 1992.

[12] E. Candés, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[13] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 1289–1306, Feb 2006.

[14] E. Candes and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse problems*, vol. 23, no. 3, p. 969, 2007.

[15] S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.

[16] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[17] W. Bajwa, J. Haupt, G. Raz, and R. Nowak, "Compressed channel sensing," in *Information Sciences and Systems, 2008. CISS 2008. 42nd Annual Conference on*. IEEE, 2008, pp. 5–10.

[18] C. R. Berger, S. Zhou, J. C. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: from subspace methods to compressed sensing," *IEEE Trans. Signal Processing*, vol. 58, no. 3, pp. 1708–1721, 2010.

[19] M. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *Signal Processing, IEEE Transactions on*, vol. 57, no. 6, pp. 2275–2284, 2009.

[20] R. Baraniuk and P. Steeghs, "Compressive radar imaging," in *IEEE Radar Conference*, 2007, pp. 128–133.

[21] Y. Chi, Y. Xie, and R. Calderbank, "Compressive demodulation of mutually interfering signals," *arXiv preprint arXiv:1303.3904*, 2013.

[22] Y. Chi, L. L. Scharf, A. Pezeshki, and R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Trans. Signal Processing*, May 2011.

[23] A. Fannjiang and H.-C. Tseng, "Compressive radar with off-grid and extended targets," *arXiv preprint arXiv:1209.6399*, 2012.

[24] M. F. Duarte and R. G. Baraniuk, "Spectral compressive sensing," *Applied and Computational Harmonic Analysis*, 2012.

[25] Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse bayesian inference," *Signal Processing, IEEE Transactions on*, vol. 61, no. 1, pp. 38–43, 2013.

[26] Z. Tan, P. Yang, and A. Nehorai, "Joint sparse recovery method for compressed sensing with structured dictionary mismatch," *arXiv preprint arXiv:1309.0858*, 2013.

[27] L. Hu, Z. Shi, J. Zhou, and Q. Fu, "Compressed sensing of complex sinusoids: An approach based on dictionary refinement," *Signal Processing, IEEE Transactions on*, vol. 60, no. 7, pp. 3809–3822, 2012.

[28] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, 2013.

[29] G. Tang, B. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *Information Theory, IEEE Transactions on*, vol. 59, no. 11, pp. 7465–7490, 2013.

[30] V. Chandrasekaran, B. Recht, P. Parrilo, and A. Willsky, "The convex algebraic geometry of linear inverse problems," in *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*. IEEE, 2010, pp. 699–703.

[31] E. J. Candès and C. Fernandez-Granda, "Super-resolution from noisy data," *Journal of Fourier Analysis and Applications*, vol. 19, no. 6, pp. 1229–1254, 2013.

[32] Y. Chen and Y. Chi, "Spectral compressed sensing via structured matrix completion," in *International Conference on Machine Learning*, Atlanta, GA, Jun 2013.

[33] ——, "Robust spectral compressed sensing via structured matrix completion," *IEEE Transactions on Information Theory*, vol. 60, no. 10, pp. 6576–6601, Oct 2014.

[34] C. Carathéodory and L. Fejér, "Uber den zusammenhang der extemen von harmonischen funktionen mit ihren koeffizienten und uber den picard-landauschen satz," *Rendiconti del Circolo Matematico di Palermo*, vol. 32, pp. 218–239, 1911.

[35] W. Xu, J.-F. Cai, K. V. Mishra, M. Cho, and A. Kruger, "Precise semidefinite programming formulation of atomic norm minimization for recovering d-dimensional ($d \geq 2$) off-the-grid frequencies," *arXiv preprint arXiv:1312.0485*, 2013.

[36] B. Dumitrescu, *Positive trigonometric polynomials and signal processing applications*. Springer, 2007.

[37] D. Slepian, "Prolate spheroidal wave functions, Fourier analysis and uncertainty," *Bell Syst. Tech. J*, vol. 57, no. 5, pp. 1371–1429, 1978.

[38] Y. Chi, "Sparse mimo radar via structured matrix completion," in *Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE*. IEEE, 2013, pp. 321–324.

[39] M. Grant, S. Boyd, and Y. Ye, "CVX: Matlab software for disciplined convex programming," *Online accessible: http://stanford. edu/~boyd/cvx*, 2008.

[40] J. Tropp, "User-friendly tail bounds for sums of random matrices," *Foundations of Computational Mathematics*, vol. 12, no. 4, pp. 389–434, 2012.

[41] M. Talagrand, "Concentration of measure and isoperimetric inequalities in product spaces," *Publications Mathematiques de l'IHES*, vol. 81, no. 1, pp. 73–205, 1995.

[42] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 13–30, 1963.

[43] J. Taylor, "Aggregate dynamics and staggered contracts," *The Journal of Political Economy*, pp. 1–23, 1980.

[44] S. Gershgorin, "Uber die abgrenzung der eigenwerte einer matrix," *Izv. Akad. Nauk. USSR Otd. Fiz.-Mat. Nauk*, no. 6, pp. 749–754, 1931.

[45] A. Schaeffer, "Inequalities of a. markoff and s. bernstein for polynomials and related functions," *Bull. Amer. Math. Soc*, vol. 47, pp. 565–579, 1941.

[46] L. Gurvits and H. Barnum, "Largest separable balls around the maximally mixed bipartite quantum state," *Physical Review A*, vol. 66, no. 6, p. 062311, 2002.

Yuejie Chi (S'09-M'12) received the Ph.D. degree in Electrical Engineering from Princeton University in 2012, and the B.E. (Hon.) degree in Electrical Engineering from Tsinghua University, Beijing, China, in 2007. Since September 2012, she has been an assistant professor with the department of Electrical and Computer Engineering and the department of Biomedical Informatics at The Ohio State University. She is the recipient of the IEEE Signal Processing Society Young Author Best Paper Award in 2013 and the Best Paper Award at the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) in 2012. She received the Ralph E. Powe Junior Faculty Enhancement Award from Oak Ridge Associated Universities in 2014, a Google Faculty Research Award in 2013, the Roberto Padovani scholarship from Qualcomm Inc. in 2010, and an Engineering Fellowship from Princeton University in 2007. She has held visiting positions at Colorado State University, Stanford University and Duke University, and interned at Qualcomm Inc. and Mitsubishi Electric Research Lab. Her research interests include high-dimensional data analysis, statistical signal processing, machine learning and their applications in communications, networks, imaging and bioinformatics.

Yuxin Chen (S'09) received the B.S. in Microelectronics with High Distinction from Tsinghua University in 2008, the M.S. in Electrical and Computer Engineering from the University of Texas at Austin in 2010, and the M.S. in Statistics from Stanford University in 2013. He is currently a Ph.D. candidate in the Department of Electrical Engineering at Stanford University. His research interests include information theory, compressed sensing, network science and high-dimensional statistics.