Compressive Recovery of 2-D Off-Grid Frequencies

Yuejie Chi[†], Yuxin Chen[‡]
[†]Department of Electrical and Computer Engineering, The Ohio State University
[‡]Department of Electrical Engineering, Stanford University
Email: chi.97@osu.edu, yxchen@stanford.edu

***Abstract*—Estimation of two-dimensional frequencies arises in many applications such as radar, inverse scattering, and wireless communications. In this paper, we consider retrieving continuous-valued two-dimensional frequencies in a mixture of r complex sinusoids from a random subset of its n regularly-spaced time-domain samples. We formulate an atomic norm minimization program that, with high probability, guarantees perfect recovery from $\mathcal{O}(r \log r \log n)$ samples under a mild frequency separation condition. We propose to solve the atomic norm minimization via semidefinite programming, and validate the proposed algorithm via numerical examples. Our work extends the framework proposed by Tang et. al. [1] for line spectrum estimation to multi-dimensional spectrum estimation.**

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

I. Introduction

The problem of estimating two-dimensional (2D) frequencies arises in many signal processing applications. With certain transmitted waveforms, one can interpret one dimension as time delays and the other as Doppler shifts, and use it to characterize signal propagation from a transmitter to a receiver. In radar and sonar systems, each pair of time delay and Doppler shift specifies the location and speed of a scatter, hence retrieving these parameters is of great importance for target localization and tracking [2]. Another application concerns channel sensing in wireless communications, where estimating accurate channel state information is crucial for coherent detection to maintain 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 [3], and each path can be identified by a triple of time delay, Doppler shift and attenuation.

A natural goal in various applications is to minimize the number of samples required for frequency retrieval. In wireless communications, training pilots are transmitted and extracted from the received signal in order to estimate the channel. The smaller the number of pilots, the higher the data rate. Conventional methods are often based on linear least-squares estimators [4], where the number of samples has to be greater than the dimensionality of the signal space determined by the maximal time delay and Doppler shift. Another conventional approach is based on parametric representation, which relies on methods such as 2D unitary ESPRIT [5], Clark and Scharf's IQML method [6], the Matrix Enhancement Matrix Pencil (MEMP) method [7], etc, to directly estimate 2D frequencies. However, these approaches often require equi-spaced time-domain samples. Besides, their performance highly rely on prior knowledge of the model order (i.e. the number of sinusoids), and many of these methods are sensitive to model mismatch and noise.

Recent advance in compressed sensing (CS) [8], [9] posits that it is possible to recover a spectrally-sparse signal from a small number of time-domain samples. Specifically, consider a time-domain signal of dimension $n = n_1 \times n_2$ composed of r 2D sinusoids. If the frequencies of the sinusoids (approximately) lie on the DFT grid over a unit disk $[0, 1] \times [0, 1]$, the signal of interest becomes a sparse signal in the DFT basis. It is then shown that the signal can be recovered from a random subset of time-domain samples with sample size $\mathcal{O}(r \log n)$, via efficient ℓ_1-minimization [10]. The success of CS has witnessed a large body of work that enables sub-Nyquist sampling, particularly for compressive channel sensing [11], [12], high-resolution radar [13], [14], and multi-user detection [15]. However, when approximating continuous-valued frequencies on a discrete grid, the signal cannot be exactly sparse due to spectral leakage of the off-grid frequencies. This effect is studied in great details in [16], which often results in significant performance degradation of conventional CS algorithms (e.g. basis pursuit [17]).

Recently, Candès and Fernandez-Granda [18] proposed to recover the continuous-valued frequencies with infinite precision using total-variation norm minimization, from a small number of *consecutive* time-domain samples. The authors showed that if the frequencies are separated at least by $c_d/(r \log r \log n)$ for some small constant c_d depending on the frequency dimension d, then it is possible to recover the frequencies using $\mathcal{O}(r \log r \log n)$ samples; the approach is also robust to noise [19]. With about the same number of samples, Tang *et. al.* [1] greatly relaxed the separation condition to $4/n$ in the one-dimensional setting, by sampling *randomly* its time-domain samples with the aid of atomic norm minimization. The atomic norm was proposed by Chandrasekaran *et. al.* [20] as a general recipe of 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 one-sparse vectors, nuclear norm minimization for low-rank matrix completion where the atoms are rank-one matrices, and so on. It is worth noting that the atomic norm for spectrally-sparse signals is equivalent to the total-variation norm studied in [18], [19]. On the other hand, the current authors approached the harmonic retrieval problem via structured matrix completion [21], [22], by performing nuclear

norm minimization over multi-fold Hankel matrices. This approach requires slightly larger sample size ($\mathcal{O}(r\log^2 n)$) to guarantee perfect recovery, provided that the signal model enjoys a few incoherence properties.

In this paper, we extend the approach by Tang *et. al.* [1] to higher dimensions by instantiating the case in 2D models. We demonstrate that with a sample size of $\mathcal{O}(r\log r\log n)$, the proposed atomic norm minimization algorithm is guaranteed to work with high probability under a similar mild frequency separation condition. The proof is inspired by [1] and [18], that is, to construct a dual polynomial certifying the optimality of the solution for the corresponding convex optimization problem. We then propose to solve the atomic norm minimization problem via semidefinite programming, which can be solved efficiently using off-the-shelf solvers. However, unlike the line spectrum case [1], the equivalence between the atomic norm minimization and our proposed semidefinite program is not guaranteed in general. This is partly due to the fact that the Caratheodory's theorem [23] does not hold in higher dimensions. Instead, we validate the effectiveness of the proposed semidefinite program as well as its stability against noise through numerical examples. The proposed algorithm and analysis immediately extend to an even higher-dimensional frequency model.

The rest of the paper is organized as follows. In Section II, we formulate the problem and review related literature. Section III proposes the atomic norm minimization problem and presents its performance guarantee. Section IV formulates a semidefinite program to approximate the original atomic norm minimization. Section V provides numerical experiments that validate our algorithm. Finally we conclude in Section VI.

II. Problem Formulation and Related Work

A. Problem Formulation

Consider an $n_1\times n_2$ data matrix $\boldsymbol{X}_0$ where each entry $x_{\boldsymbol{k}} = x_{k_1,k_2}$ can be expressed as a sum of r complex sinusoids,

$$x_{\boldsymbol{k}} = x_{k_1,k_2} = \sum_{i=1}^{r} d_i y_i^{k_1} z_i^{k_2} = \frac{1}{\sqrt{n_1 n_2}}\sum_{i=1}^{r} d_i e^{j2\pi \boldsymbol{f}_i^T \boldsymbol{k}}, \quad (1)$$

where $\boldsymbol{k} = [k_1,k_2] \in J := \{0,\dots,n_1-1\}\times\{0,\dots,n_2-1\}$, d_i is complex,

$$y_i = \frac{1}{\sqrt{n_1}} e^{j2\pi f_{1i}}, \text{ and } z_i = \frac{1}{\sqrt{n_2}} e^{j2\pi f_{2i}}$$

for each $1\le i\le r$. Let $\mathcal{F} = \{\boldsymbol{f}_i = (f_{1i}, f_{2i}) \in [0,1]^2,\ 1\le i\le r\}$ be the set of *distinct* frequencies. We can write $\boldsymbol{X}_0$ in a matrix form as follows:

$$\boldsymbol{X}_0 = \boldsymbol{Y}\boldsymbol{D}\boldsymbol{Z}^T, \quad (2)$$

where $\boldsymbol{Y}$ is given as

$$\boldsymbol{Y} = [\boldsymbol{a}(f_{11}),\dots,\boldsymbol{a}(f_{1r})] \in \mathbb{C}^{n_1\times r}, \quad (3)$$

with $\boldsymbol{a}(f_{1i}) = [1, y_i, \dots, y_i^{n_1-1}]^T$, $\boldsymbol{Z}$ is given as

$$\boldsymbol{Z} = [\boldsymbol{b}(f_{21}),\dots,\boldsymbol{b}(f_{2r})] \in \mathbb{C}^{n_2\times r}, \quad (4)$$

with $\boldsymbol{b}(f_{2i}) = [1, z_i, \dots, z_i^{n_2-1}]^T$, and $\boldsymbol{D}$ is a diagonal matrix given as

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

Let the vectorized data matrix $\boldsymbol{x}_0 = \text{vec}(\boldsymbol{X}_0^T)$ be

$$\begin{aligned}\boldsymbol{x}_0 = (\boldsymbol{Y}\otimes\boldsymbol{Z})\boldsymbol{d} &= \sum_{i=1}^{r} d_i \boldsymbol{a}(f_{1i})\otimes\boldsymbol{b}(f_{2i}) \\ &= \sum_{i=1}^{r} d_i \boldsymbol{c}(f_{1i}, f_{2i}), \end{aligned} \quad (6)$$

where $\otimes$ denotes Kronecker product, $\|\boldsymbol{c}(\boldsymbol{f}_i)\|_2 = 1$ and

$$\boldsymbol{c}(\boldsymbol{f}_i) = \boldsymbol{c}(f_{1i}, f_{2i}) = \boldsymbol{a}(f_{1i})\otimes\boldsymbol{b}(f_{2i}) \in \mathbb{C}^{n_1 n_2}.$$

In this paper we assume that we observe a random subset of entries of $\boldsymbol{X}_0$. Specifically, denote by $\Omega \subset J$ the set of locations such that x_{k_1,k_2} are observed if and only if $(k_1,k_2)\in\Omega$. We define $\mathcal{P}_\Omega(\boldsymbol{X}) = \boldsymbol{P}\odot\boldsymbol{X}$, where $\boldsymbol{P} = [p_{k_1,k_2}] \in \{0,1\}^{n_1\times n_2}$ as a binary mask matrix where $p_{k_1,k_2} = 1$ if $(k_1,k_2)\in\Omega$ and $p_{k_1,k_2} = 0$ otherwise, and $\odot$ denotes point-wise product. Without ambiguity we use the same notation Ω, J and $\mathcal{P}_\Omega$ to denote the set of observed entries, all entries, and the binary mask matrix of the vectorized signal $\boldsymbol{x}_0$, respectively.

We assume that each entry is observed uniformly at random with probability $p = \frac{|\Omega|}{n_1 n_2} \in [0,1]$. Our goal is signal reconstruction, i.e. to recover the unobserved entries of the original data matrix $\boldsymbol{X}_0$. We note that the frequencies $\mathcal{F}$ can be recovered using conventional approaches such as the MEMP method [7] from the data matrix once it is reconstructed.

B. Conventional CS Approach

CS represents $\boldsymbol{x}_0$ as a sparse signal in a pre-determined basis. We discretize the 2D plane $[0,1]\times[0,1]$ with grid points $\boldsymbol{\omega} = [\omega_1,\omega_2] \in \Omega_d$, where $\omega_1 \in \{0,\dots,(n_1-1)/n_1\}$ and $\omega_2 \in \{0,\dots,(n_2-1)/n_2\}$. The vectorized signal $\boldsymbol{x}_0$ can be approximately represented as a sparse signal as

$$\boldsymbol{x}_0 \approx \sum_{i=1}^{r} \tilde{d}_i \boldsymbol{c}(\boldsymbol{\omega}_i), \quad (7)$$

where $\tilde{d}_i$ is the amplitude of the mode $\boldsymbol{\omega}_i$. Write the basis as

$$\boldsymbol{F} = [\boldsymbol{c}(\boldsymbol{\omega})] = \boldsymbol{F}_1\otimes\boldsymbol{F}_2 \in \mathbb{C}^{n_1 n_2\times n_1 n_2},$$

where $\boldsymbol{F}_1$ and $\boldsymbol{F}_2$ are DFT matrices of dimension n_1 and n_2, respectively. CS suggests that we could perfectly recover $\boldsymbol{x}_0$ via ℓ_1-minimization as follows

$$\min \quad \|\tilde{\boldsymbol{d}}\|_1 \quad \text{subject to} \quad \mathcal{P}_\Omega(\boldsymbol{F}\tilde{\boldsymbol{d}}) = \mathcal{P}_\Omega(\boldsymbol{x}_0),$$

where $\tilde{\boldsymbol{d}} = [\tilde{d}_1,\dots,\tilde{d}_{n_1 n_2}]$ is the reconstructed sparse representation of $\boldsymbol{x}_0$ in $\boldsymbol{F}$. The major issue with the above approach is that the frequencies $\boldsymbol{f}_i$ never lie perfectly on the grid, since it is determined by the scattering nature of physics. Hence there is an unavoidable mismatch between the true frequencies and the discrete grid, and the performance of sparse recovery algorithms can degrade considerably without further processing. In this paper we adopt a different approach

and try to recover the frequencies directly without assuming a fine grid.

III. Atomic Norm Minimization for 2D Frequency Recovery

The atomic norm is proposed in [20] as a general recipe of proposing 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 one-sparse vectors for sparse recovery, rank-one matrices for low-rank matrix completion, and so on.

In the case of 2D spectral analysis, it is straightforward to define the atomic norm for a signal $\boldsymbol{x}$ as [1]

$$\|\boldsymbol{x}\|_{\mathcal{A}} = \inf\left\{\sum_i |d_i| \Big| \boldsymbol{x} = \sum_i d_i \boldsymbol{c}_i(\boldsymbol{f}_i)\right\}, \tag{8}$$

which is obtained by convexifying the atomic set of 2D complex sinusoids $\mathcal{A} = \{\boldsymbol{c}(\boldsymbol{f}) | \boldsymbol{f} \in [0,1]^2\}$. We aim to solve the following atomic norm minimization problem

$$\min_{\boldsymbol{x}} \quad \|\boldsymbol{x}\|_{\mathcal{A}} \quad \text{subject to } \mathcal{P}_\Omega(\boldsymbol{x}) = \mathcal{P}_\Omega(\boldsymbol{x}_0). \tag{9}$$

This approach is considered in [1] for line spectral estimation and is shown to allow perfect frequency recovery from a random subset of $\mathcal{O}(r \log r \log n)$ samples with high probability. Here, we aim to extend the approach to 2D frequency reconstruction. We refer the readers to [20] and [1] for detailed discussion about the atomic norm.

The following theorem shows that the atomic norm minimization (15) recovers the data matrix $\boldsymbol{X}_0$ perfectly under mild separation conditions, provided that the number of samples is $\mathcal{O}(r \log r \log n)$. The proof can be found in [24]. For simplicity we will assume $n_1 = n_2 = n^{1/2}$.

Theorem 1. *Let $n_1 = n_2 = n^{1/2}$. Suppose we observe samples of a data matrix in* (1) *with frequencies $\boldsymbol{f}_i = [f_{1i}, f_{2i}] \in [0,1]^2, 1 \le i \le r$ on the index set $T \subset J$ of size $|\Omega| = m$ uniformly at random. If the signs of d_i's are i.i.d. and uniformly generated from the complex unit circle, and the minimum separation between frequencies satisfies*

$$\Delta_{\min} \triangleq \min_{i \neq j} \|\boldsymbol{f}_i - \boldsymbol{f}_j\|_\infty \ge \frac{c_2}{\lfloor (n^{1/2}-1)/4 \rfloor} \tag{10}$$

for some small universal constant c_2, then there exists a numerical constant C such that if

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

then $\boldsymbol{X}$ is the unique optimizer to (9) *with probability at least $1-\delta$.*

Remark 1. Theorem 1 asserts that if the frequencies are minimally separated as in (10), and if r is not too small, the recovery is exact as soon as m is $\mathcal{O}(r \log r \log n)$ with high probability. This result is order-wise equivalent to the performance guarantee for line spectrum estimation in [1].

Remark 2. We compare Theorem 1 with conventional methods such as ESPRIT. ESPRIT can recover the underlying frequencies from $\Theta(r)$ *equally spaced* samples of the data matrix $\boldsymbol{X}_0$. The number of samples doesn't depend on the dimension of $\boldsymbol{X}_0$ but only the degrees of freedom. The proposed algorithm (9) assumes random subsampling of the data matrix $\boldsymbol{X}_0$, accommodating more versatile sampling methods compared with ESPRIT. This is accomplished at a price of a higher number of samples. Moreover, in the noise-free setting as described in Theorem 1, ESPRIT allows recovery without a separation condition like (10). However, a separation condition is necessary for stable reconstruction when noise is present, as discussed in [18], [25].

Remark 3. We compare Theorem 1 with the result in CS. CS assumes a discretized grid Ω_d of the 2D plane $[0,1]^2$ for the data model as discussed in (7). When the frequencies in $\mathcal{F}$ are indeed on the grid, CS allows recovery of r complex sinusoids from a number of $\mathcal{O}(r \log n)$ samples. The proposed algorithm (9) can be regarded as a remedy of CS for targets off the grid with slightly more samples.

IV. Heuristic Semidefinite Program

Theorem 1 indicates that solving the atomic norm minimization problem (9) admits perfect signal recovery from only a small number of its samples. However, it remains to be seen how to efficiently solve (9). Unfortunately, the exact semidefinite programming characterization of atomic norm minimization for line spectrum estimation, as proposed in [1], cannot be extended to 2D in the most general sense. This arises due to the fundamental difficulty of generalizing the Caratheodory's theorem [23] beyond the 1D model. Nonetheless, in this section we propose a semidefinite program to approximately solve (9) which exhibits excellent empirical performance in Section V.

Given a $(2n_1-1) \times (2n_2-1)$ matrix $\boldsymbol{T}$, we first define an $n_1 \times n_1$ block Toeplitz matrix $\boldsymbol{T}_{\mathcal{S}}$ with respect to $\boldsymbol{T}$ as

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

where each block $\boldsymbol{T}_l$, $-n_1 < l < n_1$ is an $n_2 \times n_2$ Toeplitz matrix defined as

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

We use $\boldsymbol{T}_{\mathcal{S}}$ 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 $(\boldsymbol{a}(f_{1i})\boldsymbol{a}(f_{1i})^*) \otimes (\boldsymbol{b}(f_{2i})\boldsymbol{b}(f_{2i})^*)$ is a two-fold block Toeplitz matrix. We first give the following proposition.

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

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

and the corresponding optimal value as $\|\boldsymbol{x}\|_{\mathcal{T}}$. *We have* $\|\boldsymbol{x}\|_{\mathcal{A}} \geq \|\boldsymbol{x}\|_{\mathcal{T}}$. *Moreover, if* $\hat{\boldsymbol{T}}_{\mathcal{S}}$ *can be written as*

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

where

$$\boldsymbol{V} = [\boldsymbol{c}(f_{11}, f_{21}), \ldots, \boldsymbol{c}(f_{1r}, f_{2r})],$$
$$\boldsymbol{\Sigma} = diag([\sigma_1, \ldots, \sigma_r]),$$

and σ_i*'s are real positive, 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

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

where $\boldsymbol{T}_{\mathcal{S}} = \sum_i |d_i| \boldsymbol{c}(\boldsymbol{f}_i)\boldsymbol{c}(\boldsymbol{f}_i)^*$, therefore

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

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

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

then we have $\hat{\boldsymbol{T}}_{\mathcal{S}} \succeq 0$ and $\hat{\boldsymbol{T}}_{\mathcal{S}} \succeq t^{-1}\boldsymbol{x}\boldsymbol{x}^*$. If we can write $\hat{\boldsymbol{T}}_{\mathcal{S}} = \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}^*$, then $\boldsymbol{x} = \boldsymbol{V}\boldsymbol{d}$ for some vector $\boldsymbol{d}$. Let $\boldsymbol{q}$ be any vector such that $\boldsymbol{V}^*\boldsymbol{q} = \mathrm{sign}(\boldsymbol{d})$, where $\mathrm{sign}(\boldsymbol{d})$ is the sign vector of $\boldsymbol{d}$, then $\hat{\boldsymbol{T}}_{\mathcal{S}} \succeq t^{-1}\boldsymbol{x}\boldsymbol{x}^*$ implies

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

This further leads to

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

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

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

$$\begin{aligned} \min_{\boldsymbol{T},\boldsymbol{x},t} \quad & \frac{1}{2}\mathrm{tr}\,(\boldsymbol{T}_{\mathcal{S}}) + \frac{1}{2}t \\ \text{subject to} \quad & \begin{bmatrix} \boldsymbol{T}_{\mathcal{S}} & \boldsymbol{x} \\ \boldsymbol{x}^* & t \end{bmatrix} \succeq 0, \quad \mathcal{P}_{\Omega}(\boldsymbol{x}) = \mathcal{P}_{\Omega}(\boldsymbol{x}_0). \end{aligned} \tag{15}$$

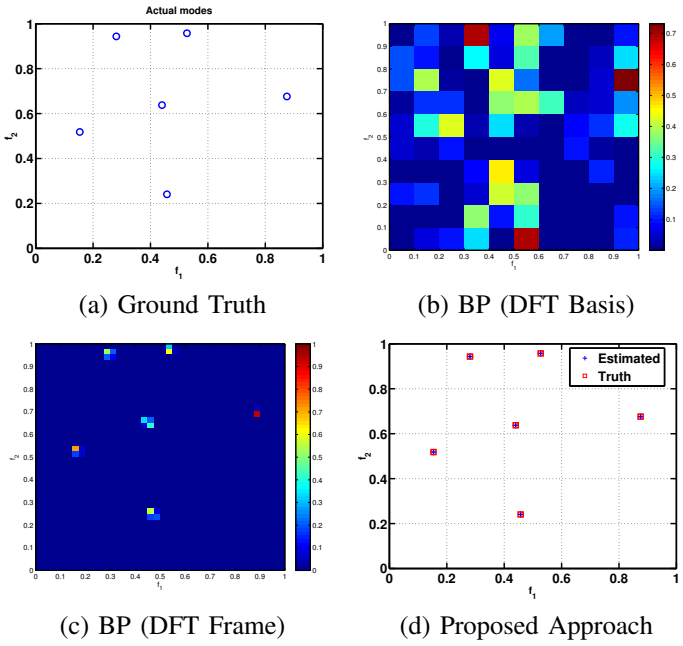
Unlike the line spectral estimation algorithm proposed in [1], the semidefinite program formulation (15) is not formally equivalent to (9) in general.

V. Numerical Simulations

We present numerical examples to examine the performance of the proposed algorithm (15). In the first example, let $n_1 = n_2 = 10$. We randomly generated $r = 6$ frequency pairs in $[0, 1]^2$, with

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

where the coefficient of each mode was generated with constant magnitude one and a random phase from $\mathcal{U}[0, 2\pi]$. The actual frequency locations are depicted in Fig. 1 (a). Each entry was observed with probability $p = m/(n_1 n_2)$, where $m = 30$. We then solved (15) using CVX [26]. 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) illustrates 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 [7] from the data matrix recovered from (15) 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)

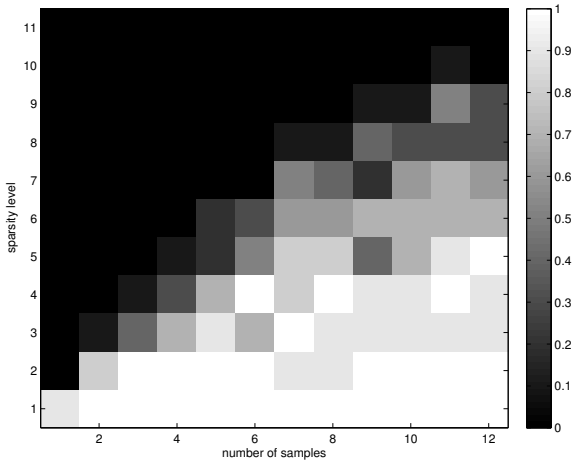
(c) BP (DFT Frame) (d) Proposed Approach

Fig. 1. The recovered frequencies when $n_1 = n_2 = 10$ from $m = 30$ measurements using difference methods. (a) Ground truth. (b) BP with DFT basis. (c) BP with oversampled DFT frame. (d) Atomic norm minimization.

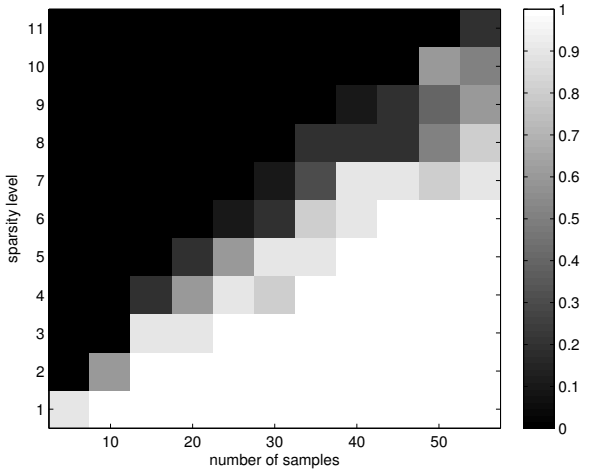
We also examined the phase transition of the proposed algorithm (15). Let $n_1 = n_2 = 8$. For each pair of m and the number of modes r, we repeated 10 Monte Carlo experiments, where in each experiment we randomly generated r complex sinusoids (a) without any separation assumption or (b) under a separation condition of $\Delta_{\min} = 1.5/n_1$ and m samples. The recovery was claimed successful if the normalized mean squared error (NMSE) error $\|\boldsymbol{x}^\star - \boldsymbol{x}\|_2/\|\boldsymbol{x}\|_2 \leq 10^{-5}$, where $\boldsymbol{x}^\star$ was the reconstructed data. Fig. 2 illustrates the empirical success probability for each pair of m and r, with the color of each cell reflecting the success rate, with Fig. 2 (a) for random modes and (b) for random modes satisfying a separation condition. Fig. 2 (b) has a much better phase transition compared with (a), indicating that the number of samples grows approximately linearly with respect to r, which confirms our theoretical analysis.

VI. Conclusions

In this paper we address the problem of retrieving 2D continuous-valued frequencies of a signal from a random



(a) Random locations



(b) $\Delta_{\min} = 1.5/n_1$

Fig. 2. Phase transition plots when $n_1 = n_2 = 8$, where mode locations are (a) generated randomly without imposing a separation condition; (b) generated randomly with a separation condition $\Delta_{\min} = 1.5/n_1$. The success rate in each cell is calculated by averaging over 10 Monte Carlo runs.

subset of its regularly-spaced time samples. We formulate an atomic norm minimization problem, and show that with a sampling size of $\mathcal{O}(r \log r \log n)$ it is sufficient to guarantee perfect frequency recovery with high probability under a mild separation condition. Our approach can be extended to higher-dimension model using a multi-fold block Toeplitz matrix constructed similar to (14), and a similar semidefinite program can be proposed. Future works include examinations of robustness of the proposed algorithm with respect to noise and sparse outliers.

ACKNOWLEDGEMENTS

Y. Chi was partially supported by a grant from the Simons Foundation 276631.

REFERENCES

[1] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," July 2012. [Online]. Available: http://arxiv.org/abs/1207.6053

[2] M. Skolnik, "Radar handbook," 1970.

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

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

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

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

[11] 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.

[12] 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.

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

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

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

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

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

[18] E. Candes and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," March 2012. [Online]. Available: http://arxiv.org/abs/1203.5871

[19] E. J. Candes and C. Fernandez-Granda, "Super-resolution from noisy data," *Arxiv 1211.0290*, November 2012.

[20] 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.

[21] Y. Chen and Y. Chi, "Spectral compressed sensing via structured matrix completion," *arXiv preprint arXiv:1304.4610*, 2013.

[22] ——, "Robust spectral compressed sensing via structured matrix completion," *arXiv preprint arXiv:1304.8126*, 2013.

[23] C. Carathéodory and L. Fejér, "Uber den zusammenghang 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.

[24] Y. Chi and Y. Chen, "Compressive recovery of 2-d off-grid frequencies via atomic norm minimization," *In preparation*.

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

[26] M. Grant, S. Boyd, and Y. Ye, "Cvx: Matlab software for disciplined convex programming," *Online accessiable: http://stanford. edu/˜ boyd/cvx*, 2008.