JOINT SPARSITY RECOVERY FOR SPECTRAL COMPRESSED SENSING

Yuejie Chi

Department of Electrical and Computer Engineering
The Ohio State University
Columbus, Ohio 43210

ABSTRACT

Compressed Sensing (CS) is an effective approach to reduce the required number of samples for reconstructing a sparse signal in an a priori basis, but may suffer severely from the issue of basis mismatch. In this paper we study the problem of simultaneously recovering multiple spectrally-sparse signals that are supported on the same frequencies lying arbitrarily on the unit circle. We propose an atomic norm minimization problem, which can be regarded as a continuous counterpart of the discrete CS formulation and be solved efficiently via semidefinite programming. Through numerical experiments, we show that the number of samples per signal may be further reduced by harnessing the joint sparsity pattern of multiple signals.

Index Terms— compressed sensing, basis mismatch, atomic norm, joint sparsity

1. INTRODUCTION

Compressed Sensing (CS) [1, 2] asserts a signal can be recovered from a small number of linear measurements if it can be regarded as a sparse or compressible signal in an a priori basis. An important consequence is its application in analog-to-digital conversion and spectrum estimation, where the signal of interest is spectrally sparse, composing of a small number r of frequency components. It is shown that if the frequencies of a signal all lie on the DFT grid, the signal of length n can then be recovered exactly from a random subset of $O(r \log n)$ samples with high probability [3]. However, in reality the frequencies of a signal never lie on a grid, no matter how fine the grid is; rather, they are continuous-valued and determined by the mother nature, e.g. a point spread function. Performance degeneration of CS algorithms is observed and studied in [4–6] when a "basis mismatch" between the actual frequencies and the assumed grid occurs, and many algorithms have been proposed to mitigate the basis mismatch effect [7, 8].

More recently, several approaches have been proposed to assume away the need for an a priori basis while maintaining the capabilities of subsampling in CS with rigorous performance guarantees. One recent approach is via atomic norm minimization [9], which is a general recipe for designing convex solutions to parsimonious model selection. It has been successfully applied to recover a spectrally-sparse signal from a subset of its consecutive samples [10] or randomly selected samples [11]. In particular, Tang et. al. showed that a spectrally-sparse signal can be recovered from $O(r \log n \log r)$ random samples with high probability when the frequencies satisfy a mild separation condition [11]. This framework has also been extended to handle higher-dimensional frequencies [12]. Another approach is proposed in [13, 14], where the problem is reformulated into a structured matrix completion inspired by matrix pencil [15]. For this approach, it is shown that $O(r \log^2 n)$ randomly selected samples are sufficient to guarantee perfect recovery with high probability under some mild incoherence conditions and the approach is amenable to higher-dimensional frequencies. With slightly more samples, both approaches can recover off-the-grid frequencies at an arbitrary precision. We refer interested readers for respective papers for details.

In this paper, we study the problem of simultaneously recovering multiple spectrally-sparse signals that are supported on the same frequencies lying arbitrarily on the unit circle. With Multiple measurement Vectors (MMV) in a CS framework, it is shown to further reduce the required number of samples [16–21] by harnessing the joint sparsity pattern of different signals. Our proposed algorithm can be regarded as a continuous counterpart of the MMV model in CS, and is based on atomic norm minimization which can be solved efficiently using semidefinite programming. We characterize a dual certificate for the optimality of the proposed algorithm. Comparisons are given for joint sparse recovery between conventional CS algorithms and the proposed algorithm through numerical experiments, which indicate that the number of samples per signal can be reduced by harnessing the joint sparsity pattern of multiple signals using atomic norm minimization.

The rest of the paper is organized as below. Section 2 describes the problem formulation. Section 3 formulates the atomic norm minimization problem for joint sparse recovery. Section 4 gives numerical experiments and we conclude in Section 5. Throughout the paper, we use bold letters to denote matrices and vectors, and unbolded letters to denote scalars. The transpose is denoted by $(\cdot)^T$, the complex conjugate is denoted by $(\cdot)^*$, and the trace is denoted by $\text{Tr}(\cdot)$.

2. PROBLEM FORMULATION AND BACKGROUNDS

Let $\boldsymbol{x} = [x_1, \ldots, x_n]^T \in \mathbb{C}^n$ be a spectrally-sparse signal with r distinct frequency components, which can be written as

$$\boldsymbol{x} = \sum_{k=1}^{r} c_k \boldsymbol{a}(f_k) \triangleq \boldsymbol{V}\boldsymbol{c}, \tag{1}$$

where each atom $\boldsymbol{a}(f)$ is defined as

$$\boldsymbol{a}(f) = \frac{1}{\sqrt{n}} \left[1, e^{j2\pi f}, \ldots, e^{j2\pi f(n-1)} \right]^T, \quad f \in [0, 1). \tag{2}$$

The Vandermonde matrix $\boldsymbol{V}$ is given as $\boldsymbol{V} = [\boldsymbol{a}(f_1), \ldots, \boldsymbol{a}(f_r)] \in \mathbb{C}^{n \times r}$, and the coefficient vector $\boldsymbol{c} = [c_1, \ldots, c_r]^T \in \mathbb{C}^r$. The set of

frequencies $\mathcal{F} = \{f_k\}_{k=1}^r$ can lie anywhere on the unit circle, such that f_k is continuously valued in $[0, 1)$.

In an MMV model, we consider L signals, stacked in a matrix, $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L]$, where each $\boldsymbol{x}_l \in \mathbb{C}^n$, $l = 1, \ldots, L$, shares the same set of frequencies and is composed of the form of (1) as

$$\boldsymbol{x}_l = \sum_{k=1}^r c_{k,l} \boldsymbol{a}(f_k) = \boldsymbol{V} \boldsymbol{c}_l, \tag{3}$$

with $\boldsymbol{c}_l = [c_{1,l}, \ldots, c_{r,l}]^T$. Hence $\boldsymbol{X}$ can be expressed as

$$\boldsymbol{X} = \boldsymbol{V} \boldsymbol{C}, \tag{4}$$

where $\boldsymbol{C} = \begin{bmatrix} \boldsymbol{c}_1 & \cdots & \boldsymbol{c}_L \end{bmatrix} \in \mathbb{C}^{r \times L}$. Assume we observe the *same* subset of entries of each $\boldsymbol{x}_l$, and denote the location set of observed entries by Ω, with $m = |\Omega|$. When there is no noise, the observations can be given as

$$\boldsymbol{Z}_\Omega = \mathcal{P}_\Omega(\boldsymbol{X}) \in \mathbb{C}^{m \times L},$$

where $\mathcal{P}_\Omega$ is a projection operator that only preserves the rows of $\boldsymbol{X}$ indexed by Ω.[1]

The traditional CS approach assumes that $\boldsymbol{x}_l$'s are sparse in an a priori determined DFT basis or DFT frame $\boldsymbol{F} \in \mathbb{C}^{n \times d}$ $(d \geq n)$, and they share the same sparsity pattern. Hence, the signal $\boldsymbol{X}$ is *modeled* as

$$\boldsymbol{X} = \boldsymbol{F} \boldsymbol{\Theta}, \tag{5}$$

where $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d]^* \in \mathbb{C}^{d \times L}$, and the number of nonzero rows of $\boldsymbol{\Theta}$ is small. Define the group sparsity ℓ_1 norm of $\boldsymbol{\Theta}$ as $\|\boldsymbol{\Theta}\|_{2,1} = \sum_{i=1}^d \|\boldsymbol{\theta}_i\|_2$. A convex optimization algorithm [17] that motivates group sparsity can be posed to solve the MMV model as

$$\hat{\boldsymbol{\Theta}} = \mathrm{argmin}_{\boldsymbol{\Theta}} \ \|\boldsymbol{\Theta}\|_{2,1} \quad \text{s.t.} \quad \boldsymbol{F}_\Omega \boldsymbol{\Theta} = \boldsymbol{Z}_\Omega, \tag{6}$$

where $\boldsymbol{F}_\Omega$ is the subsampled DFT basis or DFT frame on Ω. The signal then is recovered as $\hat{\boldsymbol{X}} = \boldsymbol{F} \hat{\boldsymbol{\Theta}}$. However, when the frequencies $\mathcal{F}$ are off-the-grid, the model (5) becomes highly inaccurate due to spectral leakage along the Dirichlet kernel, making (6) degrades significantly in performance. We will compare against this conventional approach with our proposed algorithm in Section 4.

3. ATOMIC NORM MINIMIZATION FOR MMV MODELS

In this section we develop the atomic norm minimization algorithm for solving the MMV model with spectrally-sparse signals. We first define an atom as

$$\boldsymbol{A}(f, \boldsymbol{b}) = \boldsymbol{a}(f) \boldsymbol{b}^*, \tag{7}$$

where $f \in [0, 1)$, $\boldsymbol{b} \in \mathbb{R}^L$ satisfying $\|\boldsymbol{b}\|_2 = 1$, and the set of atoms as $\mathcal{A} = \{\boldsymbol{A}(f, \boldsymbol{b}) | f \in [0, 1), \|\boldsymbol{b}\|_2 = 1\}$. Define

$$\|\boldsymbol{X}\|_{\mathcal{A},0} = \inf_r \left\{ \boldsymbol{X} = \sum_{k=1}^r c_k \boldsymbol{A}(f_k, \boldsymbol{b}_k), c_k \geq 0 \right\}.$$

as the smallest number of atoms to describe $\boldsymbol{X}$. Our goal is thus to minimize $\|\boldsymbol{X}\|_{\mathcal{A},0}$ that satisfies the observation, given as

$$\min \|\boldsymbol{X}\|_{\mathcal{A},0} \quad \text{s.t.} \quad \boldsymbol{Z}_\Omega = \mathcal{P}_\Omega(\boldsymbol{X}). \tag{8}$$

It is easy to show that $\|\boldsymbol{X}\|_{\mathcal{A},0}$ can be represented equivalently as

$$\|\boldsymbol{X}\|_{\mathcal{A},0} = \inf_{\boldsymbol{u}, \boldsymbol{W}} \left\{ \mathrm{rank}(\mathrm{toep}(\boldsymbol{u})) \middle| \begin{bmatrix} \mathrm{toep}(\boldsymbol{u}) & \boldsymbol{X} \\ \boldsymbol{X}^* & \boldsymbol{W} \end{bmatrix} \succ \boldsymbol{0} \right\},$$

---
[1] We remark that, this restriction of fixing the observation pattern across different signals, is actually unnecessary for the proposed algorithm.

where $\mathrm{toep}(\boldsymbol{u})$ is the Toeplitz matrix generated by $\boldsymbol{u}$. Hence (8) is NP-hard. We will alternatively consider the convex relaxation of $\|\boldsymbol{X}\|_{\mathcal{A},0}$, defining the atomic norm [9] of $\boldsymbol{X}$ as

$$\|\boldsymbol{X}\|_{\mathcal{A}} = \inf \{ t > 0 : \ \boldsymbol{X} \in t \,\mathrm{conv}(\mathcal{A}) \}$$

$$= \inf \left\{ \sum_k c_k \middle| \boldsymbol{X} = \sum_k c_k \boldsymbol{A}(f_k, \boldsymbol{b}_k), c_k \geq 0 \right\}. \tag{9}$$

The atomic norm of a single vector $\boldsymbol{x}_l$ defined in [11] becomes a special case of (9) for $L = 1$. We propose to solve the following atomic norm minimization algorithm:

$$\hat{\boldsymbol{X}} = \mathrm{argmin}_{\boldsymbol{X}} \ \|\boldsymbol{X}\|_{\mathcal{A}} \quad \text{s.t.} \quad \boldsymbol{Z}_\Omega = \mathcal{P}_\Omega(\boldsymbol{X}). \tag{10}$$

3.1. Semidefinite Program Characterization

We now prove the following equivalent semidefinite program (SDP) characterization of $\|\boldsymbol{X}\|_{\mathcal{A}}$, showing that (10) can be solved efficiently using off-the-shelf SDP solvers.

Theorem 1. *The atomic norm $\|\boldsymbol{X}\|_{\mathcal{A}}$ can be written equivalently as*

$$\|\boldsymbol{X}\|_{\mathcal{A}} = \inf_{\boldsymbol{u} \in \mathbb{C}^n, \boldsymbol{W} \in \mathbb{C}^{L \times L}} \left\{ \frac{1}{2} \mathrm{Tr}(\mathrm{toep}(\boldsymbol{u})) + \frac{1}{2} \mathrm{Tr}(\boldsymbol{W}) \middle| \right.$$
$$\left. \begin{bmatrix} \mathrm{toep}(\boldsymbol{u}) & \boldsymbol{X} \\ \boldsymbol{X}^* & \boldsymbol{W} \end{bmatrix} \succ \boldsymbol{0} \right\}.$$

Proof. Denote the value of the right hand side as $\|\boldsymbol{X}\|_{\mathcal{T}}$. Suppose that $\boldsymbol{X} = \sum_{k=1}^r c_k \boldsymbol{a}(f_k) \boldsymbol{b}_k^*$, there exists a vector $\boldsymbol{u}$ such that

$$\mathrm{toep}(\boldsymbol{u}) = \sum_{k=1}^r c_k \boldsymbol{a}(f_k) \boldsymbol{a}(f_k)^*,$$

by the Vandermonde decomposition lemma [22]. It is obvious that

$$\begin{bmatrix} \mathrm{toep}(\boldsymbol{u}) & \boldsymbol{X} \\ \boldsymbol{X}^* & \sum_{k=1}^r c_k \boldsymbol{b}_k \boldsymbol{b}_k^* \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{k=1}^r c_k \boldsymbol{a}(f_k) \boldsymbol{a}(f_k)^* & \sum_{k=1}^r c_k \boldsymbol{a}(f_k) \boldsymbol{b}_k^* \\ \sum_{k=1}^r c_k \boldsymbol{b}_k \boldsymbol{a}(f_k)^* & \sum_{k=1}^r c_k \boldsymbol{b}_k \boldsymbol{b}_k^* \end{bmatrix}$$

$$= \sum_{k=1}^r c_k \begin{bmatrix} \boldsymbol{a}(f_k) \\ \boldsymbol{b}_k \end{bmatrix} \begin{bmatrix} \boldsymbol{a}(f_k)^* & \boldsymbol{b}_k^* \end{bmatrix} \succ 0,$$

and

$$\frac{1}{2} \mathrm{Tr}(\mathrm{toep}(\boldsymbol{u})) + \frac{1}{2} \mathrm{Tr}(\boldsymbol{W}) = \sum_{k=1}^r c_k = \|\boldsymbol{X}\|_{\mathcal{A}},$$

therefore $\|\boldsymbol{X}\|_{\mathcal{T}} \leq \|\boldsymbol{X}\|_{\mathcal{A}}$. On the other hand, suppose that for any $\boldsymbol{u}$ and $\boldsymbol{W}$ that satisfy

$$\begin{bmatrix} \mathrm{toep}(\boldsymbol{u}) & \boldsymbol{X} \\ \boldsymbol{X}^* & \boldsymbol{W} \end{bmatrix} \succ \boldsymbol{0},$$

with $\mathrm{toep}(\boldsymbol{u}) = \boldsymbol{V} \boldsymbol{D} \boldsymbol{V}^*$, $\boldsymbol{D} = \mathrm{diag}(d_i)$, $d_i > 0$. It follows that $\boldsymbol{X}$ is in the range of $\boldsymbol{V}$, hence $\boldsymbol{X} = \boldsymbol{V} \boldsymbol{B}$ with the columns of $\boldsymbol{B}^T$ given by $\boldsymbol{b}_i$. Since $\boldsymbol{W}$ is also PSD, $\boldsymbol{W}$ can be written as $\boldsymbol{W} = \boldsymbol{B}^* \boldsymbol{E} \boldsymbol{B}$ where $\boldsymbol{E}$ is also PSD. We now have

$$\begin{bmatrix} \mathrm{toep}(\boldsymbol{u}) & \boldsymbol{X} \\ \boldsymbol{X}^* & \boldsymbol{W} \end{bmatrix} = \begin{bmatrix} \boldsymbol{V} & \\ & \boldsymbol{B}^* \end{bmatrix} \begin{bmatrix} \boldsymbol{D} & \boldsymbol{I} \\ \boldsymbol{I} & \boldsymbol{E} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}^* & \\ & \boldsymbol{B} \end{bmatrix} \succ \boldsymbol{0},$$

which yields

$$\begin{bmatrix} \boldsymbol{D} & \boldsymbol{I} \\ \boldsymbol{I} & \boldsymbol{E} \end{bmatrix} \succ \boldsymbol{0}$$

and $\boldsymbol{E} \succ \boldsymbol{D}^{-1}$ by the Schur complement lemma. Now observe

$$\mathrm{Tr}(\boldsymbol{W}) = \mathrm{Tr}(\boldsymbol{B}^* \boldsymbol{E} \boldsymbol{B}) \geq \mathrm{Tr}(\boldsymbol{B}^* \boldsymbol{D}^{-1} \boldsymbol{B})$$
$$= \mathrm{Tr}(\boldsymbol{D}^{-1} \boldsymbol{B} \boldsymbol{B}^*) = \sum_i d_i^{-1} \|\boldsymbol{b}_i\|^2.$$

Therefore,

$$\frac{1}{2}\mathrm{Tr}(\mathrm{toep}(\boldsymbol{u})) + \frac{1}{2}\mathrm{Tr}(\boldsymbol{W}) = \frac{1}{2}\mathrm{Tr}(\boldsymbol{D}) + \frac{1}{2}\mathrm{Tr}(\boldsymbol{W})$$
$$\geq \sqrt{\mathrm{Tr}(\boldsymbol{D}) \cdot \mathrm{Tr}(\boldsymbol{W})}$$
$$\geq \sqrt{\left(\sum_i d_i\right)\left(\sum_i d_i^{-1}\|\boldsymbol{b}_i\|^2\right)}$$
$$\geq \sum \|\boldsymbol{b}_i\| \geq \|\boldsymbol{X}\|_{\mathcal{A}},$$

which gives $\|\boldsymbol{X}\|_{\mathcal{T}} \geq \|\boldsymbol{X}\|_{\mathcal{A}}$. Therefore, $\|\boldsymbol{X}\|_{\mathcal{T}} = \|\boldsymbol{X}\|_{\mathcal{A}}$. $\qquad\square$

From Theorem 1, we can now equivalently write (10) as the following SDP:

$$\hat{\boldsymbol{X}} = \operatorname{argmin}_{\boldsymbol{X}} \inf_{\boldsymbol{u}, \boldsymbol{W}} \frac{1}{2}\mathrm{Tr}(\mathrm{toep}(\boldsymbol{u})) + \frac{1}{2}\mathrm{Tr}(\boldsymbol{W}) \qquad (11)$$
$$\text{s.t.} \begin{bmatrix} \mathrm{toep}(\boldsymbol{u}) & \boldsymbol{X} \\ \boldsymbol{X}^* & \boldsymbol{W} \end{bmatrix} \succ \boldsymbol{0}, \ \boldsymbol{Z}_\Omega = \mathcal{P}_\Omega(\boldsymbol{X}).$$

3.2. Dual Certification

An important question is when the algorithm (10) admits perfect recovery. To this end, we study the dual problem of (10). Denote the optimal solution of (10) as $\boldsymbol{X}^\star$. Let $\boldsymbol{Y} \in \mathbb{C}^{n \times L}$, define $\langle \boldsymbol{Y}, \boldsymbol{X} \rangle = \mathrm{Tr}(\boldsymbol{X}^* \boldsymbol{Y})$, and $\langle \boldsymbol{Y}, \boldsymbol{X} \rangle_{\mathbb{R}} = \mathrm{Re}(\langle \boldsymbol{Y}, \boldsymbol{X} \rangle)$. The dual norm of $\|\boldsymbol{X}\|_{\mathcal{A}}$ can be defined as

$$\|\boldsymbol{Y}\|_{\mathcal{A}}^* = \sup_{\|\boldsymbol{X}\|_{\mathcal{A}} \leq 1} \langle \boldsymbol{Y}, \boldsymbol{X} \rangle_{\mathbb{R}}$$
$$= \sup_{f \in [0,1), \|\boldsymbol{b}\|_2 = 1} \langle \boldsymbol{Y}, \boldsymbol{a}(f)\boldsymbol{b}^* \rangle_{\mathbb{R}}$$
$$= \sup_{f \in [0,1), \|\boldsymbol{b}\|_2 = 1} |\langle \boldsymbol{b}, \boldsymbol{Y}^* \boldsymbol{a}(f) \rangle|$$
$$= \sup_{f \in [0,1)} \|\boldsymbol{Y}^* \boldsymbol{a}(f)\|_2 \triangleq \sup_{f \in [0,1)} \|\boldsymbol{Q}(f)\|_2,$$

where $\boldsymbol{Q}(f) = \boldsymbol{Y}^* \boldsymbol{a}(f)$ is a length-L vector with each entry a polynomial in f. The dual problem of (10) can thus be written as

$$\max_{\boldsymbol{Y}} \langle \boldsymbol{Y}_\Omega, \boldsymbol{X}_\Omega^\star \rangle_{\mathbb{R}} \quad \text{s.t.} \quad \|\boldsymbol{Y}\|_{\mathcal{A}}^* \leq 1, \boldsymbol{Y}_{\Omega^c} = 0, \qquad (12)$$

where $\boldsymbol{Y}_{\Omega^c}$ denotes the projection of $\boldsymbol{Y}$ on the rows denoted by the location set $\Omega^c = \{1, \dots, n\} \backslash \Omega$. Let $(\boldsymbol{X}, \boldsymbol{Y})$ be primal-dual feasible to (10) and (12), we have $\langle \boldsymbol{Y}, \boldsymbol{X} \rangle_{\mathbb{R}} = \langle \boldsymbol{Y}, \boldsymbol{X}^\star \rangle_{\mathbb{R}}$. Strong duality holds since Slater's condition holds, and it implies that the solutions of (10) and (12) equal if and only if $\boldsymbol{Y}$ is dual optimal and $\boldsymbol{X}$ is primal optimal [23]. Using strong duality we can obtain a dual certification to the optimality of the solution of (10).

Proposition 1. *The solution of* (10) $\hat{\boldsymbol{X}} = \boldsymbol{X}^\star$ *is the unique optimizer if there exists* $\boldsymbol{Q}(f) = \boldsymbol{Y}^* \boldsymbol{a}(f)$ *such that*

$$\begin{cases} (C1): & \boldsymbol{Q}(f_k) = \boldsymbol{b}_k, \quad \forall f_k \in \mathcal{F}, \\ (C2): & \|\boldsymbol{Q}(f)\|_2 < 1, \quad \forall f \notin \mathcal{F}, \\ (C3): & \boldsymbol{Y}_{\Omega^c} = 0. \end{cases} \qquad (13)$$

Proof. First, any $\boldsymbol{Y}$ satisfying (13) is dual feasible. We have

$$\|\boldsymbol{X}^\star\|_{\mathcal{A}} \geq \|\boldsymbol{X}^\star\|_{\mathcal{A}} \|\boldsymbol{Y}\|_{\mathcal{A}}^*$$
$$\geq \langle \boldsymbol{Y}, \boldsymbol{X}^\star \rangle_{\mathbb{R}} = \langle \boldsymbol{Y}, \sum_{k=1}^r c_k \boldsymbol{a}(f_k)\boldsymbol{b}_k^* \rangle_{\mathbb{R}}$$
$$= \sum_{k=1}^r \mathrm{Re}\left(c_k \langle \boldsymbol{Y}, \boldsymbol{a}(f_k)\boldsymbol{b}_k^* \rangle\right)$$
$$= \sum_{k=1}^r \mathrm{Re}\left(c_k \langle \boldsymbol{b}_k, \boldsymbol{Q}(f_k) \rangle\right)$$
$$= \sum_{k=1}^r \mathrm{Re}\left(c_k \langle \boldsymbol{b}_k, \boldsymbol{b}_k \rangle\right) = \sum_{k=1}^r c_k \geq \|\boldsymbol{X}^\star\|_{\mathcal{A}}.$$

Hence $\langle \boldsymbol{Y}, \boldsymbol{X}^\star \rangle_{\mathbb{R}} = \|\boldsymbol{X}^\star\|_{\mathcal{A}}$. By strong duality we have $\boldsymbol{X}^\star$ is primal optimal and $\boldsymbol{Y}$ is dual optimal.

For uniqueness, suppose $\hat{\boldsymbol{X}}$ is another optimal solution which has support outside $\mathcal{F}$. It is trivial to justify if $\hat{\boldsymbol{X}}$ and $\boldsymbol{X}^\star$ have the same support, they must coincide since the set of atoms with frequencies in $\mathcal{F}$ is independent. Let $\hat{\boldsymbol{X}} = \sum_k \hat{c}_k \boldsymbol{a}(\hat{f}_k)\hat{\boldsymbol{b}}_k^*$. We then have

$$\langle \boldsymbol{Y}, \hat{\boldsymbol{X}} \rangle_{\mathbb{R}} = \sum_{\hat{f}_k \in \mathcal{F}} \mathrm{Re}\left(\hat{c}_k \langle \hat{\boldsymbol{b}}_k, \boldsymbol{Q}(\hat{f}_k) \rangle\right) + \sum_{\hat{f}_l \notin \mathcal{F}} \mathrm{Re}\left(\hat{c}_l \langle \hat{\boldsymbol{b}}_l, \boldsymbol{Q}(\hat{f}_l) \rangle\right)$$
$$< \sum_{\hat{f}_k \in \mathcal{F}} \hat{c}_k + \sum_{\hat{f}_l \notin \mathcal{F}} \hat{c}_l = \|\hat{\boldsymbol{X}}\|_{\mathcal{A}},$$

which contradicts strong duality. Therefore the optimal solution of (10) is unique. $\qquad\square$

Proposition 1 offers a way to certify the optimality of (10) as long as we can find a dual polynomial $\boldsymbol{Q}(f)$ that satisfies (13). This is left for future work.

4. NUMERICAL EXPERIMENTS

In this section, we evaluation the performance of the proposed algorithm (10), showing that the number of measurements per signal may be reduced as we increase the number of signals L for achieving the same performance.

4.1. Phase transition when varying the number of signals

Let $n = 64$ and $m = 32$. In each Monte Carlo experiment, we randomly generate a spectrally-sparse signal with r frequencies randomly located in $[0, 1)$ that satisfies a separation condition $\Delta = \min_{k \neq l} |f_k - f_l| \geq 1/n$. This separation condition is slightly weaker than the condition asserted in [11] to guarantee the success of (10) with high probability for $L = 1$. For each frequency component, we randomly generate the amplitude for each signal from the standard complex Gaussian distribution $\mathcal{CN}(0, 1)$. We run (11) using CVX [24] and calculate the reconstruction normalized mean squared error (NMSE) as $\|\hat{\boldsymbol{X}} - \boldsymbol{X}^\star\|_F / \|\boldsymbol{X}^\star\|_F$ where $\boldsymbol{X}^\star$ is the ground truth. The experiment is claimed successful if NMSE $\leq 10^{-5}$. For each pair of r and L, we run a total of 50 Monte Carlo experiments and output the average success rate. Fig. 2 shows the success rate of reconstruction versus the sparsity level r for $L = 1$, 2, and 3 respectively. As we increase L, the success rate becomes higher for the same sparsity level.

Fig. 1 shows the reconstructed dual polynomial for a randomly generated spectrally-sparse signal with $r = 10$ when $L = 1$ and $L =$
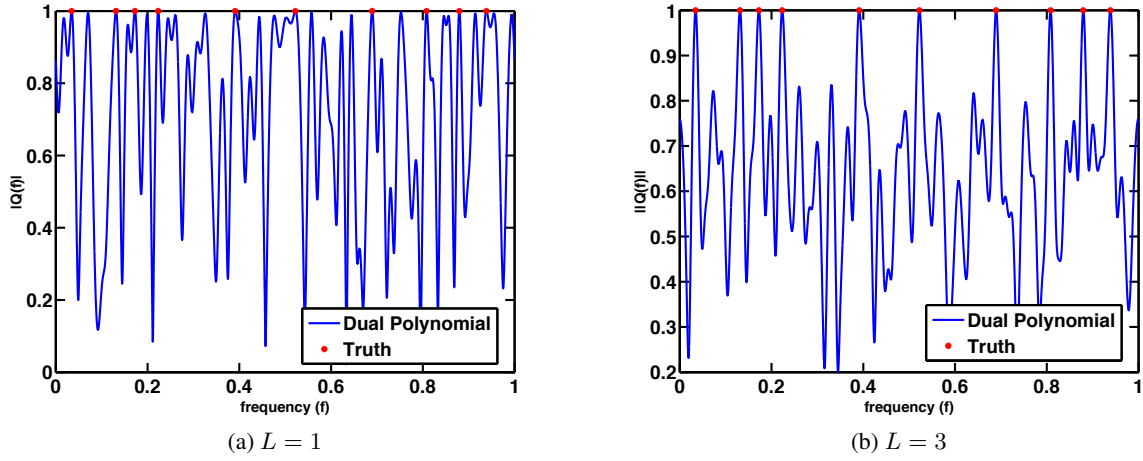
(a) $L = 1$

(b) $L = 3$

Fig. 1. The reconstructed dual polynomial for a randomly generated spectrally-sparse signal with $r = 10$, $n = 64$, and $m = 32$: (a) $L = 1$, (b) $L = 3$.
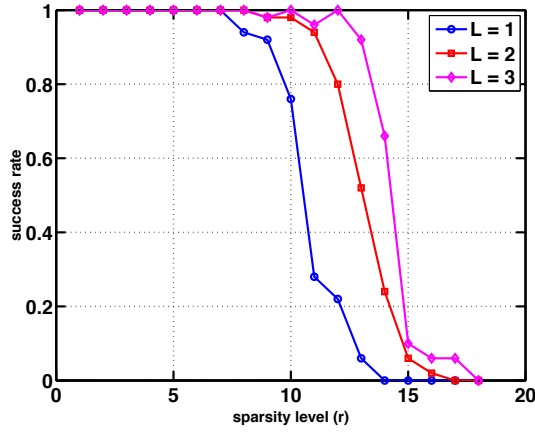

Fig. 2. Success rate of reconstruction versus the sparsity level r for $L = 1, 2, 3$ when $n = 64$, $m = 32$ and the frequencies are generated satisfying a separation condition $\Delta \geq 1/n$.

(a) $L = 1$

(b) $L = 3$

Fig. 3. The NMSE of reconstruction versus the number of samples per signal m for a randomly generated spectrally-sparse signal satisfying a separation condition $\Delta \geq 1/n$ with $r = 8$, $n = 64$: (a) $L = 1$, (b) $L = 3$.

3 respectively. It can be seen that although the algorithm achieves perfect recovery for both cases, the reconstructed dual polynomial has a much better localization property when multiple signals are present.

4.2. Comparison of CS and the proposed algorithm

Let $n = 64$. We randomly generate $r = 10$ frequencies on the unit circle $[0, 1)$ satisfying a separation condition $\Delta \geq 1/n$ as in the previous setup. Fig. 3 shows the NMSE of the reconstructed signal using the CS-MMV algorithm (6) with a DFT basis, a DFT frame with an oversampling factor $c = 2$, a DFT frame with an oversampling factor $c = 4$, and the proposed atomic norm minimization algorithm (10) for (a) $L = 1$ and (b) $L = 3$. The atomic norm minimization algorithm outperforms (6) even when the reconstruction is not exact at smaller values of m. The CS algorithm (6) can never achieve exact recover since a randomly generated frequency is always off the grid, while the recovery of the atomic norm algorithm (10) is exact after
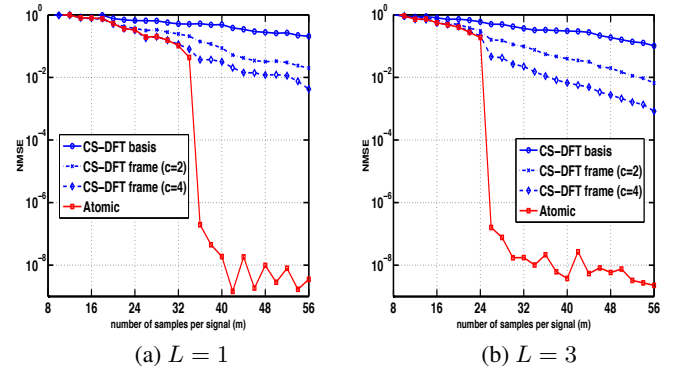
m exceeds a certain threshold for success. As we increase the number of signals L, both CS algorithms and the atomic norm algorithm improve their performance.

5. CONCLUSIONS

In this paper we study the problem of simultaneously recovering multiple spectrally-sparse signals that are supported on the same frequencies lying arbitrarily on the unit circle. We propose an atomic norm minimization problem, and solve it efficiently via semidefinite programming. Through numerical experiments, we show that the number of samples per signal may be further reduced by harnessing the joint sparsity pattern of multiple signals. The proposed atomic norm minimization algorithm can also be applied when the observation patterns are different across different signals. Future work is to develop theoretical guarantees of the proposed algorithm and examine its performance in the presence of noise. Another interesting direction is to study MMV extensions of the off-the-grid approach in [13, 14].

6. REFERENCES

[1] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[2] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289 –1306, April 2006.

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

[4] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *Signal Processing, IEEE Transactions on*, vol. 59, no. 5, pp. 2182–2195, 2011.

[5] L. L. Scharf, E. K. Chong, A. Pezeshki, and J. R. Luo, "Sensitivity considerations in compressed sensing," in *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on*. IEEE, 2011, pp. 744–748.

[6] P. Pakrooh, L. L. Scharf, A. Pezeshki, and Y. Chi, "Analysis of fisher information and the cramer-rao bound for nonlinear parameter estimation after compressed sensing," in *ICASSP*, 2013.

[7] 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, Jan 2013.

[8] A. Fannjiang and W. Liao, "Coherence pattern-guided compressive sensing with unresolved grids," *SIAM Journal on Imaging Sciences*, vol. 5, no. 1, pp. 179–202, 2012.

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

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

[11] 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

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

[13] Y. Chen and Y. Chi, "Spectral compressed sensing via structured matrix completion," *International Conference on Machine Learning (ICML)*, 2013.

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

[15] 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.

[16] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. part i: Greedy pursuit," *Signal Processing*, vol. 86, no. 3, pp. 572–588, 2006.

[17] J. A. Tropp, "Algorithms for simultaneous sparse approximation. part ii: Convex relaxation," *Signal Processing*, vol. 86, no. 3, pp. 589–602, 2006.

[18] K. Lee, Y. Bresler, and M. Junge, "Subspace methods for joint sparse recovery," *Information Theory, IEEE Transactions on*, vol. 58, no. 6, pp. 3613–3641, 2012.

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

[20] M. E. Davies and Y. C. Eldar, "Rank awareness in joint sparse recovery," *Information Theory, IEEE Transactions on*, vol. 58, no. 2, pp. 1135–1146, 2012.

[21] M. Mishali and Y. C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *Signal Processing, IEEE Transactions on*, vol. 56, no. 10, pp. 4692–4702, 2008.

[22] 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.

[23] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

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