Convex Relaxations of Spectral Sparsity for Robust Super-Resolution and Line Spectrum Estimation

Yuejie Chi

Department of Electrical and Computer Engineering
Department of Biomedical Informatics
Ohio State University, 2015 Neil Avenue, Columbus, Ohio 43210, USA

ABSTRACT

We consider recovering the amplitudes and locations of spikes in a point source signal from its low-pass spectrum that may suffer from missing data and arbitrary outliers. We first review and provide a unified view of several recently proposed convex relaxations that characterize and capitalize the spectral sparsity of the point source signal without discretization under the framework of atomic norms. Next we propose a new algorithm when the spikes are known *a priori* to be positive, motivated by applications such as neural spike sorting and fluorescence microscopy imaging. Numerical experiments are provided to demonstrate the effectiveness of the proposed approach.

Keywords: spike deconvolution, positivity, semidefinite programming, atomic norms, structured low-rank matrices, outliers, sparsity, line spectrum estimation

1. INTRODUCTION

High-resolution source location is a problem of paramount importance across many domains in science and engineering. Consider a point source signal,

$$x(t) = \sum_{k=1}^{K} c_k \delta(t - \tau_k), \tag{1}$$

where $c_k \in \mathbb{C}$ and $\tau_k \in [0, 1]$ are the amplitude and delay of the kth point source, for $1 \le k \le K$, K is the number of point sources. For many applications such as direction-of-arrival estimation, line spectrum estimation, system identification, neural spike sorting, and microscopy imaging, we observe its low-pass spectrum, given as

$$\boldsymbol{z} = \mathcal{F}x = \sum_{k=1}^{K} c_k \boldsymbol{a}_n(\tau_k) \in \mathbb{C}^n, \tag{2}$$

where $\mathcal{F}x := \int_0^1 \boldsymbol{a}_n(t)dx(t)$, $\boldsymbol{a}_n(\tau) = [1, e^{j\tau}, \ldots, e^{j(n-1)\tau}]^\mathsf{T}$ and n is the signal length. The goal is to localize the point sources, i.e. estimate the nonlinear parameters $\{c_k, \tau_k\}_{k=1}^K$, as precise as possible, from a noisy, subsampled, or even corrupted version of $\boldsymbol{z}$ in a robust and stable manner.

Classical approaches for source localization in statistical signal processing include the Prony's method, ESPRIT [1], MUSIC [2], matrix pencil [3, 4], the Pisarenko's method [5], the Tufts and Kumaresan approach [6], Cadzow's signal enhancement [7], and so on. They exploit the reduced-rank representation of the signal subspace by performing eigenvalue decomposition of the (transformed) data. One important limitation of classical approaches is that any subsampling of the measurement reduces resolution and increases variability. Moreover, many of the approaches are not robust to noise and outliers, and are sensitive to the knowledge of model order.

Recently, a new class of super-resolution approaches based on convex relaxations [8, 9, 10] has been proposed in the literature that achieves competitive performance with classical approaches, and is robust against noise,

Further author information: E-mail: chi.97@osu.edu, Telephone: (614) 297-5097.

missing data and outliers through small modifications, without discretizing the delays over a finite grid [11, 12]. Examples include enhanced matrix completion (EMaC) [8], total variation norm minimization [9], atomic norm minimization (ANM) [10, 13, 14], mosaic matrix completion [15], and so on. These approaches promote the spectral sparsity of the model (2) by embedding the signal into structured low-rank matrix representations.

The goal of the current paper is two-fold. First, we develop a unified view of structured low-rank matrix representations that promote spectral sparsity through the lens of *atomic norms*, by formulating them as atomic norm minimizations arisen from different choices of the *atomic sets* for the Cadzow's signal enhancement. This allows interpreting EMaC [8], also known as structured Hankel matrix completion, as a convex relaxation of the atomic norm minimization approach by Tang et.al. [10] for the class of complex sinusoids. Second, we specialize to the case of positive sources, and propose a new algorithm based on atomic norm minimization to recover positive point sources via semidefinite programming in the presence of arbitrary outliers.

The rest of this paper is organized as follows. Section 2 develops a unified framework for convex relaxations of spectral sparsity under the framework of atomic norm minimization. Section 3 provides a novel semidefinite characterization for positive point sources and its application for robust spike deconvolution under sparse corruptions. Finally, we conclude in Section 4.

Throughout the paper, we use boldface letters to represent matrices and vectors, e.g. $\boldsymbol{A}$ and $\boldsymbol{a}$. Define the Hermitian Toeplitz matrix with $\boldsymbol{z}^{\mathsf{T}}$ being the first row as $\mathcal{T}(\boldsymbol{z})$, where the first entry of $\boldsymbol{z}$, z_1 is real positive. Then the adjoint of $\mathcal{T}$, denoted by $\mathcal{T}^* : \mathbb{C}^{n\times n} \mapsto \mathbb{C}^n$ is defined as $[\mathcal{T}^*(\boldsymbol{M})]_j = \sum_{i=1}^{n-j+1} M_{i,i+j-1}$, for $j = 1, \ldots, n$.

2. ATOMIC NORM MINIMIZATION FOR SIGNAL ENHANCEMENT

The atomic norm is developed by Chandrasekaran et.al. [16] as a general framework to develop convex relaxations for reconstructing signals that are composed of a small number of simple building blocks called *atoms*. Let $\mathcal{A}$ denote the *atomic set*. Assume the signal of interest $\boldsymbol{g}$ can be represented as a linear combination of atoms in $\mathcal{A}$:

$$\boldsymbol{g} = \sum_i \epsilon_i \boldsymbol{\alpha}_i, \quad \boldsymbol{\alpha}_i \in \mathcal{A}.$$

Let $\mathrm{conv}(\mathcal{A})$ be the convex hull of $\mathcal{A}$, and all elements of $\mathcal{A}$ are the extreme points of $\mathrm{conv}(\mathcal{A})$. Define the atomic norm of $\boldsymbol{g}$ as

$$\|\boldsymbol{g}\|_{\mathcal{A}} = \inf\left\{t > 0 : \ \boldsymbol{g} \in t \cdot \mathrm{conv}(\mathcal{A})\right\} = \inf\left\{\sum_i |\epsilon_i| : \ \boldsymbol{g} = \sum_i \epsilon_i \boldsymbol{\alpha}_i, \ \boldsymbol{\alpha}_i \in \mathcal{A}\right\},$$

which can be regarded as a convex relaxation to motivate a sparse representation of $\boldsymbol{g}$ in the atomic set $\mathcal{A}$. Importantly, the dual norm $\|\cdot\|_{\mathcal{A}}^*$ is given as

$$\|\boldsymbol{g}\|_{\mathcal{A}}^* = \sup_{\|\boldsymbol{y}\|_{\mathcal{A}} \le 1} \langle \boldsymbol{g}, \boldsymbol{y}\rangle_{\mathbb{R}} = \sup_{\boldsymbol{\alpha}\in\mathcal{A}} |\langle \boldsymbol{g}, \boldsymbol{\alpha}\rangle|.$$

Several well-studied norms that promote parsimonious representations can be treated as special cases of the atomic norm, for example, the ℓ_1 norm for promoting sparsity and the nuclear norm for promoting low rank.

Defines the signal enhancement [4, 7] as a rearrangement of the signal $\boldsymbol{z}$ into a Hankel matrix*:

$$\mathcal{H}(\boldsymbol{z}) = \begin{bmatrix} z_1 & z_2 & \cdots & x_{n-p+1} \\ z_2 & z_3 & \cdots & x_{n-p+2} \\ \vdots & \vdots & & \vdots \\ z_p & z_{p+1} & \cdots & z_n \end{bmatrix} \in \mathbb{C}^{p\times(n-p+1)}, \tag{3}$$

*More generally, the signal enhancement is a multi-fold Hankel matrix for multi-dimensional models. For notational simplicity, we work with the one-dimensional model in this paper.

where $1 \le p \le n$ is called the *pencil parameter*, and let $q = n - p + 1$ for convenience. In practice it is beneficial to select the pencil parameter p to make $\mathcal{H}(\boldsymbol{x})$ as square as possible. It is easy to check that $\mathcal{H}(\boldsymbol{x})$ admits the following Vandermonde decomposition:

$$\mathcal{H}(\boldsymbol{z}) = \boldsymbol{V}_1 \boldsymbol{D} \boldsymbol{V}_2^{\mathsf{T}} = \sum_{k=1}^{K} c_k \boldsymbol{a}_p(\tau_k) \boldsymbol{a}_q(\tau_k)^{\mathsf{T}}, \tag{4}$$

where $\boldsymbol{D} = \operatorname{diag}\{c_1, c_2, \cdots, c_K\} \in \mathbb{C}^{K \times K}$, $\boldsymbol{V}_1 = \big[\boldsymbol{a}_p(\tau_1), \ldots, \boldsymbol{a}_p(\tau_K)\big] \in \mathbb{C}^{p \times K}$ and $\boldsymbol{V}_2 = \big[\boldsymbol{a}_q(\tau_1), \ldots, \boldsymbol{a}_q(\tau_K)\big] \in \mathbb{C}^{K \times q}$. Therefore,

$$\operatorname{rank}(\mathcal{H}(\boldsymbol{z})) \le K, \tag{5}$$

and the equality holds when the source locations $T = \{\tau_k\}_{k=1}^K$ are distinct. The Hankel matrix $\mathcal{H}(\boldsymbol{z})$ is low rank when $K \ll \min\{p, q\}$, i.e. its rank is much smaller than the ambient dimension. It also plays an important role in the literature of modal analysis [17, Chapter 11], including but not limited to annihilating filtering and linear prediction.

2.1 Convex Relaxations of Spectral Sparsity

It is possible to promote spectral sparsity by properly defining *atomic sets* for $\mathcal{H}(\boldsymbol{z})$, which recovers several recently proposed convex relaxations:

Nuclear norm of $\mathcal{H}(\boldsymbol{z})$***:*** if we let the atomic set be $\mathcal{A}_1 = \{\boldsymbol{u}\boldsymbol{v}^{\mathsf{H}} : \boldsymbol{u} \in \mathbb{C}^p, \boldsymbol{v} \in \mathbb{C}^q, \|\boldsymbol{u}\|_2 = \|\boldsymbol{v}\|_2 = 1\}$, where the atoms are rank-one matrices of unit norm, then the atomic norm of $\mathcal{H}(\boldsymbol{z})$ is equivalent to the nuclear norm of $\mathcal{H}(\boldsymbol{z})$, given as

$$\|\mathcal{H}(\boldsymbol{x})\|_* = \min_{\boldsymbol{W}_1, \boldsymbol{W}_2} \left\{ \frac{1}{2}\operatorname{Tr}(\boldsymbol{W}_1) + \frac{1}{2}\operatorname{Tr}(\boldsymbol{W}_2) \middle| \begin{bmatrix} \boldsymbol{W}_1 & \mathcal{H}(\boldsymbol{x}) \\ \mathcal{H}(\boldsymbol{x})^{\mathsf{H}} & \boldsymbol{W}_2 \end{bmatrix} \succeq 0 \right\}. \tag{6}$$

This gives the nuclear norm minimization studied in [8], known as Enhanced Matrix Completion (EMaC).

Atomic norm of $\mathcal{H}(\boldsymbol{z})$***:*** if we let the atomic set be $\mathcal{A}_2 = \{\frac{1}{\sqrt{pq}} \boldsymbol{a}_p(\tau) \boldsymbol{a}_q(\tau)^{\mathsf{T}} : \tau \in [0, 1)\}$, the atomic norm of $\mathcal{H}(\boldsymbol{z})$ is defined as

$$\|\mathcal{H}(\boldsymbol{z})\|_{\mathcal{A}} := \inf \left\{ \sum_i |\epsilon_i| : \ \mathcal{H}(\boldsymbol{z}) = \sum_i \epsilon_i \frac{1}{\sqrt{pq}} \boldsymbol{a}_p(\tau_i) \boldsymbol{a}_q(\tau_i)^{\mathsf{T}} \right\}. \tag{7}$$

Note that the mapping $\boldsymbol{x} \mapsto \mathcal{H}(\boldsymbol{x})$ is a one-to-one correspondence, and $\mathcal{H}(\boldsymbol{a}_n(\tau)) = \boldsymbol{a}_p(\tau) \boldsymbol{a}_q(\tau)^{\mathsf{T}}$. It is easy to check that (7) is equivalent to the atomic norm of $\boldsymbol{z}$, with the atomic set $\mathcal{A}_3 = \left\{ \frac{1}{\sqrt{pq}} \boldsymbol{a}_n(\tau) : \ \tau \in [0, 1) \right\}$. In other words, (7) equals to the atomic norm proposed in [10] up to a scaling difference:

$$\begin{aligned} \|\mathcal{H}(\boldsymbol{z})\|_{\mathcal{A}} = \|\boldsymbol{z}\|_{\mathcal{A}} &:= \inf \left\{ \sum_i |\epsilon_i| : \ \boldsymbol{z} = \sum_i \epsilon_i \frac{1}{\sqrt{pq}} \boldsymbol{a}_n(\tau_i) \right\} \\ &= \min_{\boldsymbol{u} \in \mathbb{C}^n, t} \left\{ \sqrt{pq} \left(\frac{1}{2n} \operatorname{Tr}(\operatorname{Toep}(\boldsymbol{u})) + \frac{t}{2} \right) \middle| \begin{bmatrix} \operatorname{Toep}(\boldsymbol{u}) & \boldsymbol{z} \\ \boldsymbol{z}^{\mathsf{H}} & t \end{bmatrix} \succeq 0 \right\}. \end{aligned}$$

Since $\mathcal{A}_2 \subset \mathcal{A}_1$, we have the following proposition.

PROPOSITION 2.1. *The nuclear norm minimization* $\|\mathcal{H}(\boldsymbol{z})\|_*$ *is a convex relaxation of the atomic norm minimization* $\|\mathcal{H}(\boldsymbol{z})\|_{\mathcal{A}} = \|\boldsymbol{z}\|_{\mathcal{A}}$.

From the dual perspective, the dual problem of (6) can be written as

$$\hat{\boldsymbol{Y}} = \operatorname{argmax}_{\boldsymbol{Y}} \ \langle \boldsymbol{Y}, \mathcal{H}(\boldsymbol{z}) \rangle_{\mathbb{R}} \quad \text{subject to} \quad \|\boldsymbol{Y}\| \le 1, \tag{8}$$

where $\|\boldsymbol{Y}\| = \sup_{\boldsymbol{u},\boldsymbol{v}:\|\boldsymbol{u}\|_2=1,\|\boldsymbol{v}\|_2=1} |\boldsymbol{u}^{\mathsf{H}}\boldsymbol{Y}\boldsymbol{v}|$ denotes the spectral norm, which is the dual norm of the nuclear norm. By constraining $\boldsymbol{u} = \frac{1}{\sqrt{p}}\boldsymbol{a}_p(\tau)$ and $\boldsymbol{v} = \frac{1}{\sqrt{q}}\boldsymbol{a}_q(\tau)$, we obtain a convex relaxation of (8) as

$$\max_{\boldsymbol{Y}} \langle \boldsymbol{Y}, \mathcal{H}(\boldsymbol{z})\rangle_{\mathbb{R}} \quad \text{subject to} \quad \sup_{\tau\in[0,1)} \left|\left\langle \boldsymbol{Y}, \frac{1}{\sqrt{pq}}\boldsymbol{a}_p(\tau)\boldsymbol{a}_q(\tau)^{\mathsf{T}}\right\rangle\right| \le 1,$$

Since both $\mathcal{H}(\boldsymbol{z})$ and $\boldsymbol{a}_p(\tau)\boldsymbol{a}_q(\tau) = \mathcal{H}(\boldsymbol{a}_n(\tau))$ are Hankel matrices, it is sufficient to consider the solution $\boldsymbol{Y} = \mathcal{H}(\boldsymbol{y})$ as a Hankel matrix, which gives

$$\max_{\boldsymbol{y}} \langle \mathcal{H}(\boldsymbol{y}), \mathcal{H}(\boldsymbol{z})\rangle_{\mathbb{R}} \quad \text{subject to} \quad \sup_{\tau\in[0,1)} \left|\left\langle \mathcal{H}(\boldsymbol{y}), \frac{1}{\sqrt{pq}}\mathcal{H}(\boldsymbol{a}_n(\tau))\right\rangle\right| \le 1.$$

By defining the vector $\tilde{\boldsymbol{y}}$ as $\tilde{y}_i = n_i y_i$, where n_i the number of ones in the matrix $\mathcal{H}(\boldsymbol{e}_i)$, where $\boldsymbol{e}_i$ is the ith standard basis vector, $i = 1, \ldots, n$. Then the above algorithm is equivalent to

$$\max_{\tilde{\boldsymbol{y}}} \langle \tilde{\boldsymbol{y}}, \boldsymbol{z}\rangle_{\mathbb{R}} \quad \text{subject to} \quad \sup_{\tau\in[0,1)} \|\tilde{\boldsymbol{y}}\|_{\mathcal{A}}^{*} \le 1,$$

where $\|\cdot\|_{\mathcal{A}}^{*}$ is the dual norm of $\|\cdot\|_{\mathcal{A}}$. This argument confirms Proposition 2.1.

2.2 Remarks

Although $\|\mathcal{H}(\boldsymbol{z})\|_*$ can be viewed as a convex relaxation of $\|\mathcal{H}(\boldsymbol{z})\|_{\mathcal{A}}$, it admits a larger atomic set therefore can be applied to a larger class of signals. In many applications such as nuclear magnetic resonance imaging [18], it is important to model damping, when the source locations are complex-valued, then the signal model (2) becomes

$$z_i^{\mathrm{d}} = \sum_{k=1}^{K} c_k \zeta_k^{i-1}, \quad i = 1, \ldots, n, \tag{9}$$

where $\zeta_k = \rho_k e^{j2\pi\tau_k}$ with $|\rho_k| \neq 1$. Interestingly, the rank property (5) continues to hold, therefore the nuclear norm minimization of $\mathcal{H}(\boldsymbol{z})$ can still be used. In [19], the nuclear norm minimization of $\mathcal{H}(\boldsymbol{z})$ is further exploited to recover signals with finite rate of innovation. However, the atomic norm $\|\mathcal{H}(\boldsymbol{z})\|_{\mathcal{A}}$ using $\mathcal{A}_2$ as the atomic set is no longer viable for these larger classes of signals.

Even within the class of spectral signals given in (2), it is worth noting that a tighter convex relaxation does not always imply a better performance for a given signal. This is particularly true when the point sources are closely located. In [20], it is shown that there exists a fundamental resolution limit for ANM, such that if the separation between point sources is below certain threshold, there exists signals that cannot be recovered by ANM when their amplitudes have opposite signs. On the other hand, for the success of EMaC, a much milder coherence condition is used in [8], and EMaC might still succeed for signals that are provably not recoverable by ANM. Nonetheless, we note that when the point sources are sufficiently separated, ANM tends to obtain a better performance than EMaC for signals in (2), due to the tightness of the formulation.

3. ROBUST POSITIVE SPIKE DECONVOLUTION

In practice, it is common that the spectrum is corrupted by outliers, for example due to impulsive noise or malicious attacks, and the observation $\boldsymbol{y} \in \mathbb{C}^n$ is given as

$$\boldsymbol{y} = \boldsymbol{z} + \boldsymbol{w}, \tag{10}$$

where, $\boldsymbol{w} \in \mathbb{C}^n$ is the outlier term, which is a sparse vector with $\|\boldsymbol{w}\|_0 = s$, and s is the number of outliers. Provable approaches based on EMaC [8] and ANM [21] are developed to exactly demix $\boldsymbol{z}$ and $\boldsymbol{w}$ by motivating simultaneously the spectral sparsity of $\boldsymbol{z}$ and the sparsity of $\boldsymbol{w}$.

In this section, we specialize to the case of positive sources, i.e. $c_i > 0$, $1 \le i \le k$. It is therefore possible to incorporate a nonnegative constraint to robustly recover the spike signal, via

$$\hat{x} = \mathrm{argmin}_q \|\boldsymbol{y} - \mathcal{F}q\|_1 + \xi\langle q, 1\rangle \quad \text{s.t.} \quad q \ge 0, \tag{11}$$

where $\xi > 0$ is a regularization parameter which can be set as $\xi = \sqrt{n}$ [21]. For a vector $\boldsymbol{r}$, let $\mathcal{F}^*\boldsymbol{r} = \Re\left(\sum_{i=1}^n r_i e^{-j2\pi(i-1)t}\right) = \Re(\boldsymbol{a}_n(t)^{\mathsf{H}}\boldsymbol{r}) := \langle \boldsymbol{r}, \boldsymbol{a}_n(t)\rangle$. Then the dual problem of (11) is given as

$$\max_{\boldsymbol{\eta}\in\mathbb{C}^n} \langle \boldsymbol{\eta}, \boldsymbol{y}\rangle \quad \text{s.t.} \quad \|\boldsymbol{\eta}\|_\infty \le 1, \quad \mathcal{F}^*\boldsymbol{\eta} \le \xi. \tag{12}$$

3.1 Semidefinite Program Characterization

The last constraint in (12) admits an equivalent semidefinite characterization, given below.

PROPOSITION 3.1. *Let $\boldsymbol{\eta} \in \mathbb{C}^n$ and $\epsilon \in \mathbb{R}$,*

$$(\mathcal{F}^*\boldsymbol{\eta})(t) \le \epsilon \quad \forall t \in [0,1] \tag{13}$$

if and only if there exists a Hermitian matrix $\boldsymbol{\Lambda} \succeq 0$ such that

$$\mathcal{T}^*(\boldsymbol{\Lambda}) + \boldsymbol{\eta} = \epsilon \boldsymbol{e}_1,$$

where $\boldsymbol{e}_1 = [1, 0, 0, \ldots, 0]^T$.

Proof. We start by proving the "if" part. Since $\boldsymbol{\Lambda} \succeq 0$, we have for any $t \in [0,1]$,

$$\begin{aligned} 0 \le \boldsymbol{a}_n(t)^{\mathsf{H}}\boldsymbol{\Lambda}\boldsymbol{a}_n(t) &= \langle \boldsymbol{\Lambda}, \boldsymbol{a}_n(t)\boldsymbol{a}_n(t)^{\mathsf{H}}\rangle \\ &= \langle \mathcal{T}^*(\boldsymbol{\Lambda}), \mathcal{T}^*(\boldsymbol{a}_n(t)\boldsymbol{a}_n(t)^{\mathsf{H}})\rangle \\ &= \langle \epsilon\boldsymbol{e}_1 - \boldsymbol{\eta}, \boldsymbol{a}_n(t)\rangle \\ &= \epsilon - \langle \boldsymbol{\eta}, \boldsymbol{a}_n(t)\rangle = \epsilon - \mathcal{F}^*\boldsymbol{\eta}, \end{aligned}$$

yielding $\mathcal{F}^*\boldsymbol{\eta} \le \epsilon$. On the other hand, if (13) holds, then by the Fejér-Riesz Theorem [22], there exists a polynomial $P(t) = \tilde{\boldsymbol{c}}^{\mathsf{H}}\boldsymbol{a}(t)$ such that

$$\epsilon - \mathcal{F}^*\boldsymbol{\eta} = |P(t)|^2 = \boldsymbol{a}_n(t)^{\mathsf{H}}\tilde{\boldsymbol{c}}\tilde{\boldsymbol{c}}^{\mathsf{H}}\boldsymbol{a}_n(t).$$

Pick $\boldsymbol{\Lambda} = \tilde{\boldsymbol{c}}\tilde{\boldsymbol{c}}^{\mathsf{H}}$, then $\boldsymbol{\Lambda} \succeq 0$, and $\langle \mathcal{T}^*(\boldsymbol{\Lambda}), \boldsymbol{a}\rangle = \boldsymbol{a}(t)^{\mathsf{H}}\boldsymbol{\Lambda}\boldsymbol{a}_n(t)$. For all $t \in [0,1]$,

$$\langle \mathcal{T}^*(\boldsymbol{\Lambda}) + \boldsymbol{\eta}, \boldsymbol{a}\rangle = \boldsymbol{a}_n(t)^{\mathsf{H}}\boldsymbol{\Lambda}\boldsymbol{a}_n(t) + \mathcal{F}^*\boldsymbol{\eta} = \epsilon,$$

therefore $\mathcal{T}^*(\boldsymbol{\Lambda}) + \boldsymbol{\eta} = \epsilon\boldsymbol{e}_1$. ∎

Therefore, the problem (12) is equivalent to:

$$\max_{\boldsymbol{\eta}\in\mathbb{C}^n, \boldsymbol{\Lambda}\succeq 0} \langle \boldsymbol{\eta}, \boldsymbol{y}\rangle \quad \text{s.t.} \quad \|\boldsymbol{\eta}\|_\infty \le 1, \quad \mathcal{T}^*(\boldsymbol{\Lambda}) + \boldsymbol{\eta} = \xi\boldsymbol{e}_1. \tag{14}$$

Notably, the dual problem of (14) is given as

$$\hat{\boldsymbol{z}} = \operatorname{argmin}_{\boldsymbol{g}\in\mathbb{C}^n} \|\boldsymbol{y} - \boldsymbol{g}\|_1 + \xi g_1 \quad \text{s.t.} \quad \mathcal{T}(\boldsymbol{g}) \succeq 0, \tag{15}$$

which seeks a low-rank Hermitian PSD Toeplitz matrix $\mathcal{T}(\boldsymbol{g})$ that minimizes a weighted sum of the ℓ_1 norm of the residual as well as the trace of $\mathcal{T}(\boldsymbol{g})$. This suggests that for positive sources, atomic norm minimization is equivalent to trace norm minimization for the PSD Toeplitz matrix constructed by the signal as its first column.

3.2 Spike Localization via the Dual Polynomial

Define the support of x as T, and the support of $\boldsymbol{w}$ as Ω. The existence of the following dual certificate certifies the optimality of (11).

PROPOSITION 3.2. *If there exist $\boldsymbol{r} \in \mathbb{C}^n$ and $Q_{\boldsymbol{r}}(t) = \mathcal{F}^*\boldsymbol{r}$ that satisfy*

$$Q_{\boldsymbol{r}}(\tau_k) = 1, \quad \forall \tau_k \in T, \qquad Q_{\boldsymbol{r}}(t) < 1, \quad \forall t \notin T, \tag{16}$$

$$r_l = \xi^{-1}\operatorname{sgn}(w_l), \quad \forall l \in \Omega, \qquad |r_l| < \xi^{-1}, \quad \forall l \notin \Omega, \tag{17}$$

where $\mathrm{sgn}(\cdot)$ *is understood as the complex sign. Then* x *is unique solution to* (11) *as long as* $k+s \le n$.

Proof. Let $\hat{x} = \sum_{\tau_\ell \in \hat{T}} c'_\ell \delta(t-\tau_\ell)$ be the optimal solution to (11), with $c'_\ell > 0$. Let $\hat{\boldsymbol{w}} = \boldsymbol{y} - \mathcal{F}\hat{x}$. Denote the support of $\hat{x}$ as $\hat{T}$ and the support of $\hat{\boldsymbol{w}}$ as $\hat{\Omega}$. From [21, Lemma C.1], it is known that if $T = T'$ and $\Omega = \Omega'$, then $\hat{x} = x$ as long as $k+s \le n$. For the rest of the proof we assume $T \neq T'$ or $\Omega \neq \Omega'$. Define $h = \hat{x} - x$, then we have

$$\begin{aligned}
\langle Q_{\boldsymbol{r}}(t), h(t)\rangle = \int_0^1 Q_{\boldsymbol{r}}(t)(\hat{x}(t)-x(t))dt &= \sum_{\tau_\ell \in \hat{T}} Q_{\boldsymbol{r}}(\tau_\ell)c'_\ell - \sum_{\tau_k \in T} Q_{\boldsymbol{r}}(\tau_k)c_k \\
&\le \sum_{\tau_\ell \in \hat{T}} c'_\ell - \sum_{\tau_k \in T} c_k \qquad (18)\\
&\le \|\boldsymbol{w}\|_1 - \|\hat{\boldsymbol{w}}\|_1, \qquad (19)
\end{aligned}$$

where (18) follows from (16) and $x, \hat{x}$ are nonnegative, and it becomes strict inequality if $T \neq T'$; (19) follows from the optimality of $\hat{x}$. On the other hand, we have

$$\begin{aligned}
\langle Q_{\boldsymbol{r}}(t), h(t)\rangle &= \int_0^1 \Re\left(\sum_{i=1}^n r_i e^{-j2\pi(i-1)t}\right)(\hat{x}(t)-x(t))dt \qquad (20)\\
&= \Re\left(\sum_{i=1}^n r_i \int_0^1 e^{-j2\pi(i-1)t}(\hat{x}(t)-x(t))dt\right) \qquad (21)\\
&= \langle \boldsymbol{r}, \mathcal{F}(\hat{x}-x)\rangle \\
&= \langle \boldsymbol{r}, \boldsymbol{w}-\hat{\boldsymbol{w}}\rangle \\
&= \|\boldsymbol{w}\|_1 - \langle \boldsymbol{r}_\Omega, \hat{\boldsymbol{w}}_\Omega\rangle - \langle \boldsymbol{r}_{\Omega^c}, \hat{\boldsymbol{w}}_{\Omega^c}\rangle \qquad (22)\\
&\ge \|\boldsymbol{w}\|_1 - \|\boldsymbol{r}_\Omega\|_\infty \|\hat{\boldsymbol{w}}_\Omega\|_1 - \|\boldsymbol{r}_{\Omega^c}\|_\infty \|\hat{\boldsymbol{w}}_{\Omega^c}\|_1 \qquad (23)\\
&\ge \|\boldsymbol{w}\|_1 - \|\hat{\boldsymbol{w}}\|_1, \qquad (24)
\end{aligned}$$

where (22) follows from

$$\langle \boldsymbol{r}, \boldsymbol{w}\rangle = \langle \boldsymbol{r}_\Omega, \boldsymbol{w}_\Omega\rangle = \|\boldsymbol{w}_\Omega\|_1 = \|\boldsymbol{w}\|_1$$

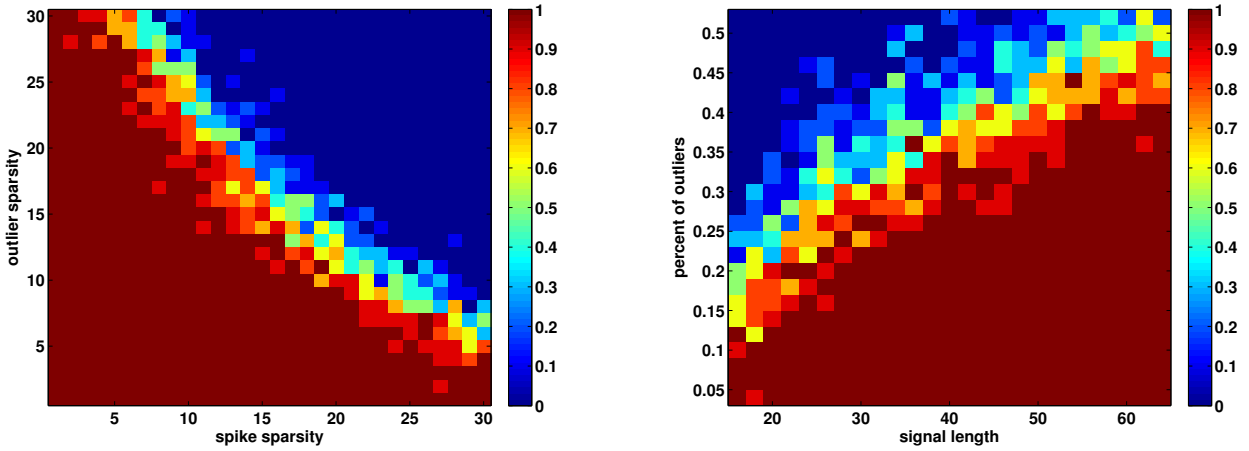
and (23) follows from the Holder's inequality. The last inequality (24) follows from (17), and it becomes a strict inequality if $\Omega \neq \Omega'$. Therefore, combining (19) and (24), we have a contradiction. Therefore $\hat{x} = x$ is the unique optimal solution to (18). ∎

Proposition 3.2 suggests that one can localize the spikes and the outliers from the dual solution of (11), as demonstrated in the numerical example in Fig. 2. Note that, the dual approach does not always produce an exact recovery of the source locations, see a more detailed discussion in [10].

3.3 Numerical Examples

We examine the performance of the proposed algorithm (11). Let $n = 64$. We randomly generate K spikes with their locations and amplitudes both drawn uniformly selected at random from $[0, 1]$. The signal is then corrupted by an s-sparse vector whose support is selected uniformly at random with the nonzero entries drawn from $\mathcal{CN}(0, 20)$. The reconstruction is claimed successful if the normalized reconstruction error satisfies $\|\hat{\boldsymbol{z}} - \boldsymbol{z}\|_2/\|\boldsymbol{z}\|_2 \le 10^{-5}$. The success rate is then calculated by averaging over 10 Monte Carlo simulations. Fig. 1 (a) shows the success rate with respect to (K, s), and (b) shows the success rate with respect to $(n, s/n)$.

We next demonstrate how the support of the outlier and the spike can be localized via the dual approach outlined in Section 3.2. We set $n = 64$, $K = s = 5$. The support of the spikes is fixed as $T = \{0.2, 0.2 + 1/n, 0.315, 0.6, 0.8\}$, where the first two spikes are separated by the so-called Rayleigh limit $1/n$. The amplitudes of the spikes and the outliers are generated in the same manner as earlier. Fig. 2 shows the absolute value of the dual vector $\boldsymbol{r} \in \mathbb{C}^n$ in the top panel, where the support of the outliers can be identified from the indices of the entries of $\boldsymbol{r}$ with unit absolute value. The bottom panel shows the dual polynomial $Q_{\boldsymbol{r}}(t)$, where the support of spikes can be identified from the roots of $Q_{\boldsymbol{r}}(t) = 1$. Note that the dual polynomial is constructed differently from the case without the positivity constraint.



(a) (b)

Figure 1. Success rate of the proposed algorithm (a) with respect to the sparsity of spikes and outliers for a fixed $n = 64$; (b) with respect to the signal length and the percent of outliers for a fixed $K = 6$.

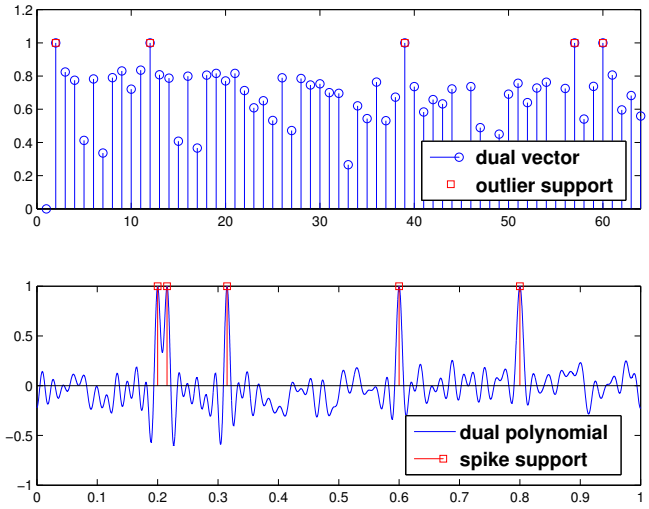


Figure 2. The reconstructed support of outliers and spikes via the dual approach. Top panel: the absolute value of the dual vector and the corresponding true outlier support. Bottom panel: the dual polynomial and the corresponding true spike support.

4. CONCLUSIONS

This paper provides a unified view of structured low-rank representations for super-resolution and line spectrum estimation, using the framework of atomic norms. In particular, the case of positive sources is treated in the application of robust spike deconvolution. Current efforts aim to develop the performance guarantee of the proposed algorithm in Section 3. Looking ahead, despite theoretical appeals, fast algorithms for structured low-rank matrix completion that produce provably accurate estimates are greatly desirable.

ACKNOWLEDGMENTS

This material is based upon work supported in part by the Air Force Office of Scientific Research under award number FA9550-15-1-0205, by the Office of Naval Research under award number N00014-15-1-2387, and by the National Science Foundation under award number CCF-1527456.

REFERENCES

[1] Roy, R. and Kailath, T., "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech and Signal Processing* **37**, 984 –995 (Jul 1989).

[2] Schmidt, R., "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation* **34**(3), 276–280 (1986).

[3] Hua, Y. and Sarkar, T. K., "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Transactions on Acoustics, Speech and Signal Processing* **38**, 814 –824 (may 1990).

[4] Hua, Y., "Estimating two-dimensional frequencies by matrix enhancement and matrix pencil," *IEEE Transactions on Signal Processing* **40**, 2267 –2280 (Sep 1992).

[5] Pisarenko, V. F., "The retrieval of harmonics from a covariance function," *Geophysical Journal International* **33**(3), 347–366 (1973).

[6] Tufts, D. and Kumaresan, R., "Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood," *Proceedings of the IEEE* **70**, 975 – 989 (sept. 1982).

[7] Cadzow, J. A., "Signal enhancement-a composite property mapping algorithm," *Acoustics, Speech and Signal Processing, IEEE Transactions on* **36**(1), 49–62 (1988).

[8] Chen, Y. and Chi, Y., "Robust spectral compressed sensing via structured matrix completion," *Information Theory, IEEE Transactions on* **60**, 6576–6601 (Oct 2014).

[9] Candès, E. J. and Fernandez-Granda, C., "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics* **67**(6), 906–956 (2014).

[10] Tang, G., Bhaskar, B. N., Shah, P., and Recht, B., "Compressed sensing off the grid," *IEEE transactions on information theory* **59**(11), 7465–7490 (2013).

[11] Herman, M. A. and Strohmer, T., "High-resolution radar via compressed sensing," *IEEE transactions on signal processing* **57**(6), 2275–2284 (2009).

[12] Chi, Y., Scharf, L., Pezeshki, A., and Calderbank, A., "Sensitivity to basis mismatch in compressed sensing," *IEEE Transactions on Signal Processing* **59**, 2182–2195 (May 2011).

[13] Chi, Y. and Chen, Y., "Compressive two-dimensional harmonic retrieval via atomic norm minimization," *IEEE Transactions on Signal Processing* **63**(4), 1030–1042 (2015).

[14] Li, Y. and Chi, Y., "Off-the-grid line spectrum denoising and estimation with multiple measurement vectors," *IEEE Transactions on Signal Processing* **64**, 1257–1269 (March 2016).

[15] Cho, M., Cai, J.-F., Liu, S., Eldar, Y. C., and Xu, W., "Fast alternating projected gradient descent algorithms for recovering spectrally sparse signals," in [*2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*], 4638–4642, IEEE (2016).

[16] Chandrasekaran, V., Recht, B., Parrilo, P. A., and Willsky, A. S., "The convex geometry of linear inverse problems," *Foundations of Computational Mathematics* **12**(6), 805–849 (2012).

[17] Scharf, L. L., [*Statistical signal processing*], vol. 98, Addison-Wesley Reading (1991).

[18] Qu, X., Mayzel, M., Cai, J.-F., Chen, Z., and Orekhov, V., "Accelerated nmr spectroscopy with low-rank reconstruction," *Angewandte Chemie International Edition* **54**(3), 852–854 (2015).

[19] Ye, J. C., Kim, J. M., and Jin, K. H., "Compressive sampling using structured low-rank interpolation," *arXiv preprint arXiv:1511.08975* (2015).

[20] Tang, G., "Resolution limits for atomic decompositions via markov-bernstein type inequalities," in [*2015 International Conference on Sampling Theory and Applications (SampTA)*], 548–552 (May 2015).

[21] Fernandez-Granda, C., Tang, G., Wang, X., and Zheng, L., "Demixing sines and spikes: Robust spectral super-resolution in the presence of outliers," *arXiv preprint arXiv:1609.02247* (2016).

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