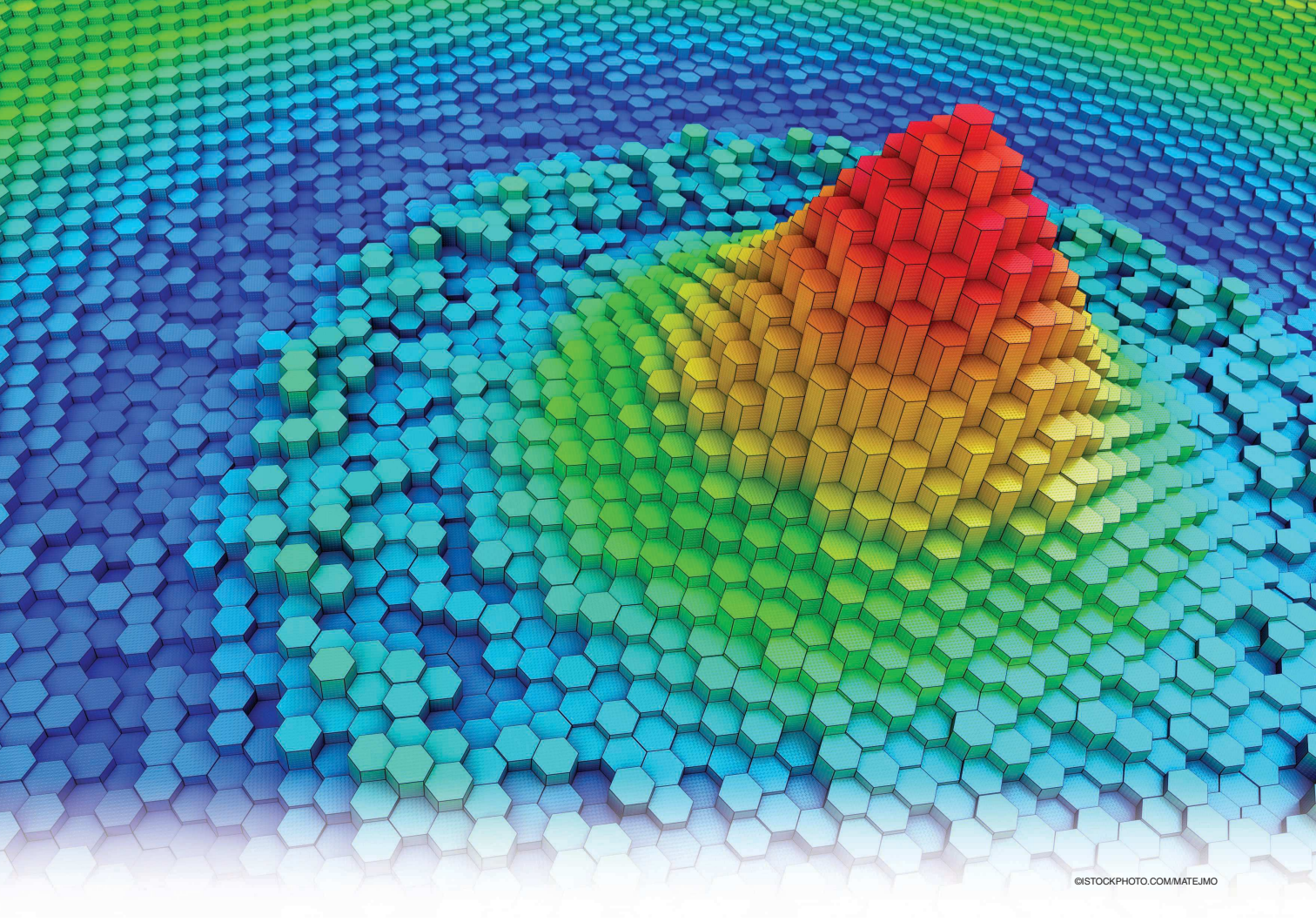


Yuejie Chi and Maxime Ferreira Da Costa

# Harnessing Sparsity Over the Continuum

*Atomic norm minimization for superresolution*

At the core of many sensing and imaging applications, the signal of interest can be modeled as a linear superposition of translated or modulated versions of some template [e.g., a point spread function (PSF) or a Green's function] and the fundamental problem is to estimate the translation or modulation parameters (e.g., delays, locations, or Dopplers) from noisy measurements. This problem is centrally important to not only target localization in radar and sonar, channel estimation in wireless communications, and direction-of-arrival estimation in array signal processing, but also modern imaging modalities such as superresolution single-molecule fluorescence microscopy, nuclear magnetic resonance imaging, and spike localization in neural recordings, among others.

Typically, the temporal or spatial resolution of the acquired signal is limited by the sensing or imaging devices, due to factors such as the numerical aperture of a microscope, the wavelength of the impinging electromagnetic or optical waves, or the sampling

*Digital Object Identifier 10.1109/MSP.2019.2962209*
*Date of current version: 26 February 2020*

rate of an analog-to-digital converter. This resolution limit is well known and often referred to as the *Rayleigh limit* (*RL*) (see "What Is the Rayleigh Limit?"). The performance of matched filtering, or periodogram, which creates a correlation map of the acquired signal against the range of parameters, is limited by the RL, regardless of the noise level.

On the other hand, the desired resolution of parameter estimation can be much higher, a challenge known as *superresolution*. There is a long history of pursuing superresolution algorithms in the signal processing community [1]–[3]. The oldest one probably dates back to de Prony's root-finding method in as early as 1795 [4], and variants of this method that are better suited for noisy data have also been proposed over time (see, e.g., [5]). Subspace methods based on the computation of eigenvector or singular vector decompositions, such as MUSIC [6], ESPRIT [7], and matrix pencil [8], have been another class of popular approaches since their inception in the 1980s. Different forms of maximum likelihood estimators have also been studied extensively [9], [10]. Collectively, these algorithms have superresolution capabilities, that is, they can resolve the parameters of interest at a resolution below the RL when the noise level is sufficiently small.

While a plethora of traditional methods already exists, convex optimization has recently emerged as a compelling framework for performing superresolution, garnering significant attention from multiple communities spanning signal processing, applied mathematics, and optimization. Due to (the relative) tractability of convex analysis and convex optimization, the new framework offers several benefits. First, strong theoretical guarantees are rigorously established to back its performance up, even in the presence of noise and corruptions. Second, it is versatile enough to include prior knowledge into the convex program to handle a wide range of measurement models that are out of the reach of traditional methods. Third, leveraging the rapid progress in large-scale convex optimization opens up the possibility of applying efficient solvers tailored to real-world applications.

The goal of this article is to offer a friendly exposition to atomic norm minimization (ANM) [11] as a canonical convex approach for superresolution. The atomic norm is first proposed in [12] as a general framework for designing tight convex relaxations to promote simple signal decompositions, where one seeks to use a minimal number of atoms to represent a given signal from an atomic set composed of an ensemble of signal atoms. Celebrated convex relaxations such as the $\ell_1$ norm approach for cardinality minimization [13] and the nuclear norm approach for rank minimization [14], can be viewed as particular instances of atomic norms for appropriately defined atomic sets. Specializing the atomic set to a dictionary containing all translations of the template signal over the continuous-valued parameter space, estimating the underlying translation parameters is then equivalent to identifying a sparse decomposition in an infinite-dimensional dictionary.

This key observation allows one to recast superresolution as solving an infinite-dimensional convex program [15], a special form of ANM considered in this article. We first highlight its mathematical formulation through a pedagogical yet useful model of superresolution that amounts to line spectrum estimation, where this infinite-dimensional convex program can be equivalently reformulated as a semidefinite program (SDP). We then demonstrate its versatility by discussing how it can be adapted to address measurement models that traditional methods may not apply easily. Finally, we illustrate its utility in superresolution image reconstruction for single-molecule fluorescence microscopy [16], where the infinite-dimensional convex program can be solved efficiently via tailored solvers.

Throughout this article, we use boldface letters to represent matrices and vectors, e.g., $\boldsymbol{a}$ and $\boldsymbol{A}$. We use $\boldsymbol{A}^{\top}$, $\boldsymbol{A}^{\mathrm{H}}$, $\mathrm{Tr}(\boldsymbol{A})$ to represent the transpose, Hermitian transpose, and trace of $\boldsymbol{A}$, respectively. The conjugate of a complex scalar $a$ is denoted as $a^{*}$. We use $\boldsymbol{A} \succeq 0$ to represent $\boldsymbol{A}$ is positive semidefinite. The matrix toep $(\boldsymbol{u})$ denotes the Hermitian Toeplitz matrix whose first column is equal to $\boldsymbol{u}$, and diag $(\boldsymbol{g})$ denotes the diagonal matrix with diagonal entries given as $\boldsymbol{g}$. The inner product between two matrices $\boldsymbol{X}$ and $\boldsymbol{P}$ is defined as $\langle \boldsymbol{X}, \boldsymbol{P} \rangle = \mathrm{Tr}(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{P})$. Additionally, the notation $f(n) = O(g(n))$ means that there exists a constant $c > 0$ such that $|f(n)| \leq c|g(n)|$.

## What Is the Rayleigh Limit?

The Rayleigh limit (RL) is an empirical criterion characterizing the resolution of an optical system due to diffraction. In a conventional fluorescence microscope, for example, the observed diffraction patterns of two fluorescent point sources become visually more difficult to distinguish as the point sources get closer to each other, as illustrated in Figure S1. They are no longer resolvable when their separation is below the RL.

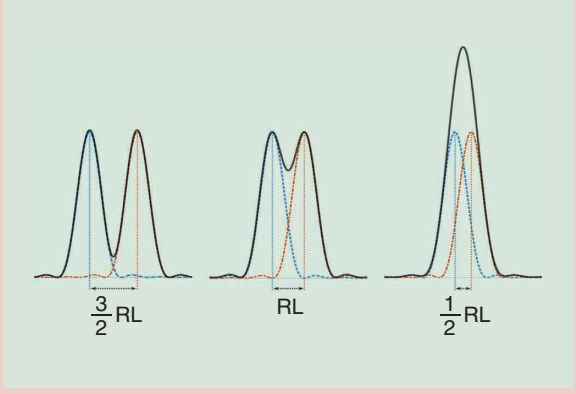


**FIGURE S1.** The combined response for two translated point spread functions under different separations of the point sources. The RL is an indication of the separability of the two sources.

## What is the atomic norm?

An everlasting idea in signal processing is decomposing a signal into a linear combination of judiciously chosen basis vectors and seeking compact and interpretable signal representations that are useful for downstream processing. For example, decomposing time series into sinusoids, speeches and images into wavelets, total system responses into impulse responses, and so on.

To fix ideas, consider the task of representing a signal $\boldsymbol{x}$ in a vector space using atoms from a collection of vectors in $\mathcal{A} = \{\boldsymbol{a}_i\}$ called an *atomic set*. The set $\mathcal{A}$ can contain either a finite or infinite number of atoms. We wish to expand $\boldsymbol{x}$ using the atoms in a form of

$$\boldsymbol{x} = \sum_i c_i \boldsymbol{a}_i, \quad \boldsymbol{a}_i \in \mathcal{A}, \tag{1}$$

where $c_i > 0$ specifies the coefficients of the decomposition. In many applications, the size of $\mathcal{A}$ can be much larger than the dimension of the signal, leading to an overcomplete representation, and there are an infinite number of possibilities to decompose $\boldsymbol{x}$. Which representation, then, shall we pick? Among the many plausible criteria, one meaningful approach is to pursue the Occam's razor principle and seek a parsimonious decomposition of the signal $\boldsymbol{x}$ involving the smallest possible number of atoms in $\mathcal{A}$, i.e., the sparsest solution to (1). The corresponding representation is known as a *sparse representation* [17]. Many real-world signals admit sparse representations for appropriately chosen atomic sets. As a simple example, natural images are approximately sparse by selecting $\mathcal{A}$ as a wavelet frame. Low-rank matrices, another class of signals that have enjoyed wide success in signal processing [18], are sparse with respect to an atomic set $\mathcal{A}$ that is the collection of all unit-norm rank-one matrices.

Given a signal $\boldsymbol{x}$, how do we find its sparse representation in the atomic set $\mathcal{A}$? In general, this problem is nonconvex and can be NP-hard due to the combinatorial aspect of cardinality minimization. The key motivation behind ANM, proposed by Chandrasekaran et al. [12], is to relax the nonconvex sparsity cost by its tight convex surrogate and instead solve the resulting convex relaxation, which is more tractable. This idea is a generalization of the popular $\ell_1$ minimization for sparse vector recovery [19], [20] when $\mathcal{A}$ is a finite set. Therein, one seeks to solve a linear program that minimizes the sum instead of the cardinality of the nonzero coefficients.

To extend the same idea to the case where $\mathcal{A}$ is an arbitrary and possibly infinite-dimensional set, we first take the convex hull of $\mathcal{A}$, denoted as $\mathrm{conv}(\mathcal{A})$, and then define its associated Minkowski functional (or gauge function) as [12]

$$\|\boldsymbol{x}\|_{\mathcal{A}} \triangleq \inf\{t \geq 0 : \boldsymbol{x} \in t \cdot \mathrm{conv}(\mathcal{A})\}, \tag{2}$$

which is the solution to a convex program. When $\mathcal{A}$ is centrally symmetric about the origin, definition (2) leads to a valid norm and is called the *atomic norm* of $\boldsymbol{x}$. Figure 1 illustrates

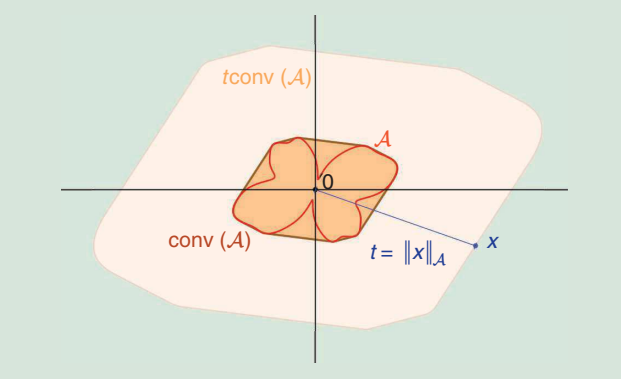


**FIGURE 1.** An atomic set $\mathcal{A}$ (in red) and its convex hull $\mathrm{conv}(\mathcal{A})$ (in orange). The atomic norm of a vector $\boldsymbol{x}$ can be interpreted as the smallest dilation factor $t \geq 0$ such that $\boldsymbol{x}$ belongs to $t\mathrm{conv}(\mathcal{A})$ (in blue).

this concept, where the atomic norm is the smallest nonnegative scaling of $\mathrm{conv}(\mathcal{A})$ until it intersects $\boldsymbol{x}$. Following definition (2), a fundamental geometric property is that the atomic norm ball, i.e., $\{\boldsymbol{x} : \|\boldsymbol{x}\|_{\mathcal{A}} \leq 1\}$, is exactly $\mathrm{conv}(\mathcal{A})$.

More interestingly, consider the case when $\boldsymbol{x}$ lies in an $n$-dimensional vector space. Carathéodory's theorem [21] guarantees that any point in $\mathrm{conv}(\mathcal{A})$ can be decomposed as a convex combination of at most $n+1$ points in $\mathcal{A}$, where $\mathcal{A}$ is not necessarily convex. Therefore, one may rewrite (2) as

$$\|\boldsymbol{x}\|_{\mathcal{A}} = \inf\left\{\sum_i c_i : \boldsymbol{x} = \sum_i c_i \boldsymbol{a}_i,\ c_i > 0,\ \boldsymbol{a}_i \in \mathcal{A}\right\}, \tag{3}$$

as long as the centroid of $\mathrm{conv}(\mathcal{A})$ is the origin. The decomposition $\Sigma_i c_i \boldsymbol{a}_i$ that obtains the infimum is referred to as the *atomic decomposition* of $\boldsymbol{x}$ onto $\mathcal{A}$. It is not hard to see that the atomic norm indeed subsumes the $\ell_1$ norm as a special case but accommodates the more general case where $\mathcal{A}$ can be an infinite-dimensional set.

Several central questions regard how to properly select the atomic set, compute the atomic norm, and find the atomic decomposition, and when does the atomic decomposition coincide with the sparse representation, i.e., the convex relaxation is tight. Clearly, the answers highly depend on the atomic set as well as the signal itself. These questions have been addressed extensively in the study of $\ell_1$-norm minimization for sparse vector recovery [19], [20], [22]. In the context of superresolution, we first address these questions under a simple model that amounts to the classical problem of line spectrum estimation, which has deep connections to systems and control theory.

## A mathematical model of superresolution, equivalent to the line spectrum estimation

We first focus on a simple yet widely applicable model of superresolution that describes the convolution of a sequence of point sources with a PSF that is resolution-limited and is illustrated in Figure 2. Let $x(t)$ be a spike signal given as

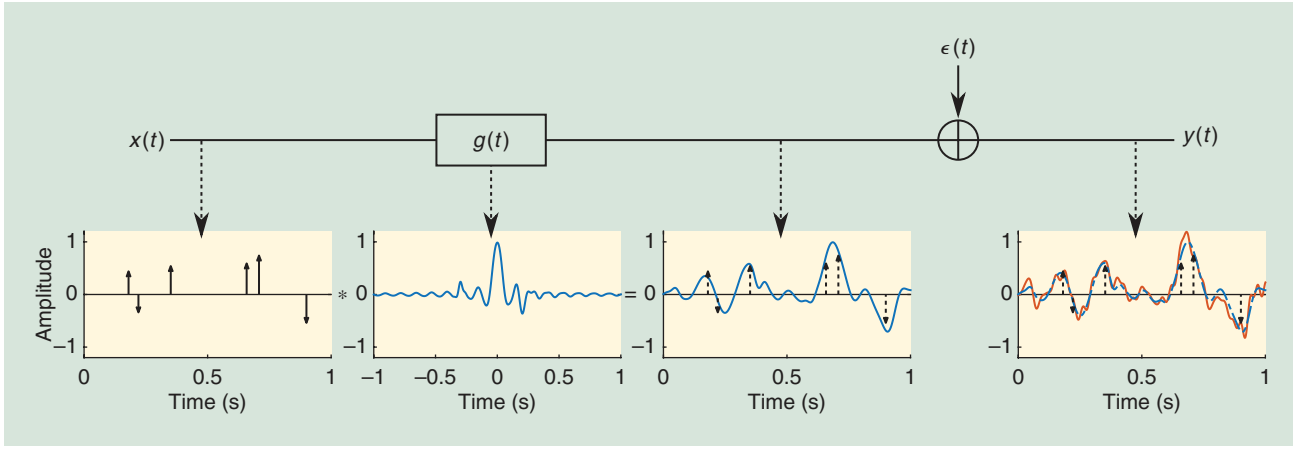


**FIGURE 2.** The mathematical model of superresolution. The spike signal $x(t)$ is convolved with a point spread function $g(t)$, leading to the degradation of its resolution, which is further exacerbated by an additive noise $\epsilon(t)$, producing an output signal $y(t)$.

$$x(t) = \sum_{k=1}^{r} c_k \delta(t - \tau_k). \tag{4}$$

Here, $r$ is the number of spikes, $c_k \in \mathbb{C}$ and $\tau_k \in [0,1)$ are the complex amplitude and delay of the $k$th spike. Without loss of generality, the maximal delay is normalized to 1. Such a spike signal can model many physical phenomena, such as firing times of neurons, locations of fluorescence molecules, and so on. Let $g(t)$ be the PSF whose bandwidth is limited due to the RL. Its Fourier transform $G(f)$ satisfies $G(f) = 0$ whenever $|f| > B/2$ for some bandwidth $B > 0$. Its convolution with $x(t)$, contaminated by an additive noise $\epsilon(t)$, can be written as

$$y(t) = x(t) * g(t) + \epsilon(t) = \sum_{k=1}^{r} c_k g(t - \tau_k) + \epsilon(t),$$

where $*$ denotes the convolution operator. Sampling the Fourier transform of the above equation at the frequencies $\ell = -\lfloor B/2 \rfloor, \ldots, 0, \ldots, \lfloor B/2 \rfloor$, we obtain the measurements

$$Y_\ell = G_\ell \cdot X_\ell + E_\ell = G_\ell \cdot \left( \sum_{k=1}^{r} c_k e^{-j2\pi\ell\tau_k} \right) + E_\ell, \tag{5}$$

where $G_\ell, X_\ell, E_\ell$, and $Y_\ell$ are the Fourier transforms of $g(t)$, $x(t)$, $\epsilon(t)$, and $y(t)$ evaluated at frequency $\ell$, respectively. The total number of samples is $n = 2\lfloor B/2 \rfloor + 1 \approx B$. We write (5) in a vector form as

$$\boldsymbol{y} = \operatorname{diag}(\boldsymbol{g})\boldsymbol{x} + \boldsymbol{\epsilon}, \tag{6}$$

where $\boldsymbol{y} = [Y_\ell], \boldsymbol{g} = [G_\ell], \boldsymbol{x} = [X_\ell]$, and $\boldsymbol{\epsilon} = [E_\ell]$. The problem of superresolution is then to estimate $\{c_k, \tau_k\}_{1 \le k \le r}$ accurately from $\boldsymbol{y}$, without knowing the model order $r$ a priori. Here, the RL is inversely proportional to the bandwidth $B$ and, roughly speaking, is about $1/n$.

When the PSF $g(t)$ is known, one can equalize (5) by multiplying $G_\ell^{-1}$ to both sides, provided that $G_\ell$'s are nonzero. The observation $\boldsymbol{z} = [G_\ell^{-1} Y_\ell]$ relates to $\boldsymbol{x}$ as

$$\boldsymbol{z} = \boldsymbol{x} + \tilde{\boldsymbol{\epsilon}}, \tag{7}$$

where $\tilde{\boldsymbol{\epsilon}}$ is the additive noise. With a slight abuse of notation, we map the index of $\ell$ from $\lfloor B/2 \rfloor, \ldots, -\lfloor B/2 \rfloor$ to $0, \ldots, n-1$ for convenience, and write $\boldsymbol{x}$ as a superposition of complex sinusoids as

$$\boldsymbol{x} = \sum_{k=1}^{r} c_k \boldsymbol{a}(\tau_k), \tag{8}$$

where $\boldsymbol{a}(\tau) \in \mathbb{C}^n$ is a vector defined as

$$\boldsymbol{a}(\tau) = \left[1, e^{j2\pi\tau}, \ldots, e^{j2\pi(n-1)\tau}\right]^\top, \quad \tau \in [0,1). \tag{9}$$

Notably, the above simplified model (7) also amounts to the classical problem of line spectrum estimation, which consists of estimating a mixture of sinusoids (with frequencies $\tau_k \in [0,1)$) from equispaced time samples (sampled at integers $\{0, \ldots, n-1\}$) of the time-domain signal $x_{\mathrm{ls}}(t) = \sum_{k=1}^{r} c_k e^{j2\pi\tau_k t}$. This finds applications in speech processing, power system monitoring, systems identification, and so on. The same model also describes direction-of-arrivals estimation using a uniform linear array, which is studied extensively in the literature of spectrum analysis [1].

## Line spectrum superresolution via ANM

In the absence of noise, one could think of superresolution as estimating the continuous-time spike signal $x(t)$ in (4) from its discrete-time moment measurements $\boldsymbol{x}$ in (8), which are related through

$$\boldsymbol{x} = \int_0^1 \boldsymbol{a}(t)\, \mathrm{d}x(t). \tag{10}$$

One can also think of $x(t)$ as the representation of $\boldsymbol{x}$ over a continuous dictionary

$$\mathcal{A}_0 = \{\boldsymbol{a}(\tau) : \tau \in [0,1)\}, \tag{11}$$

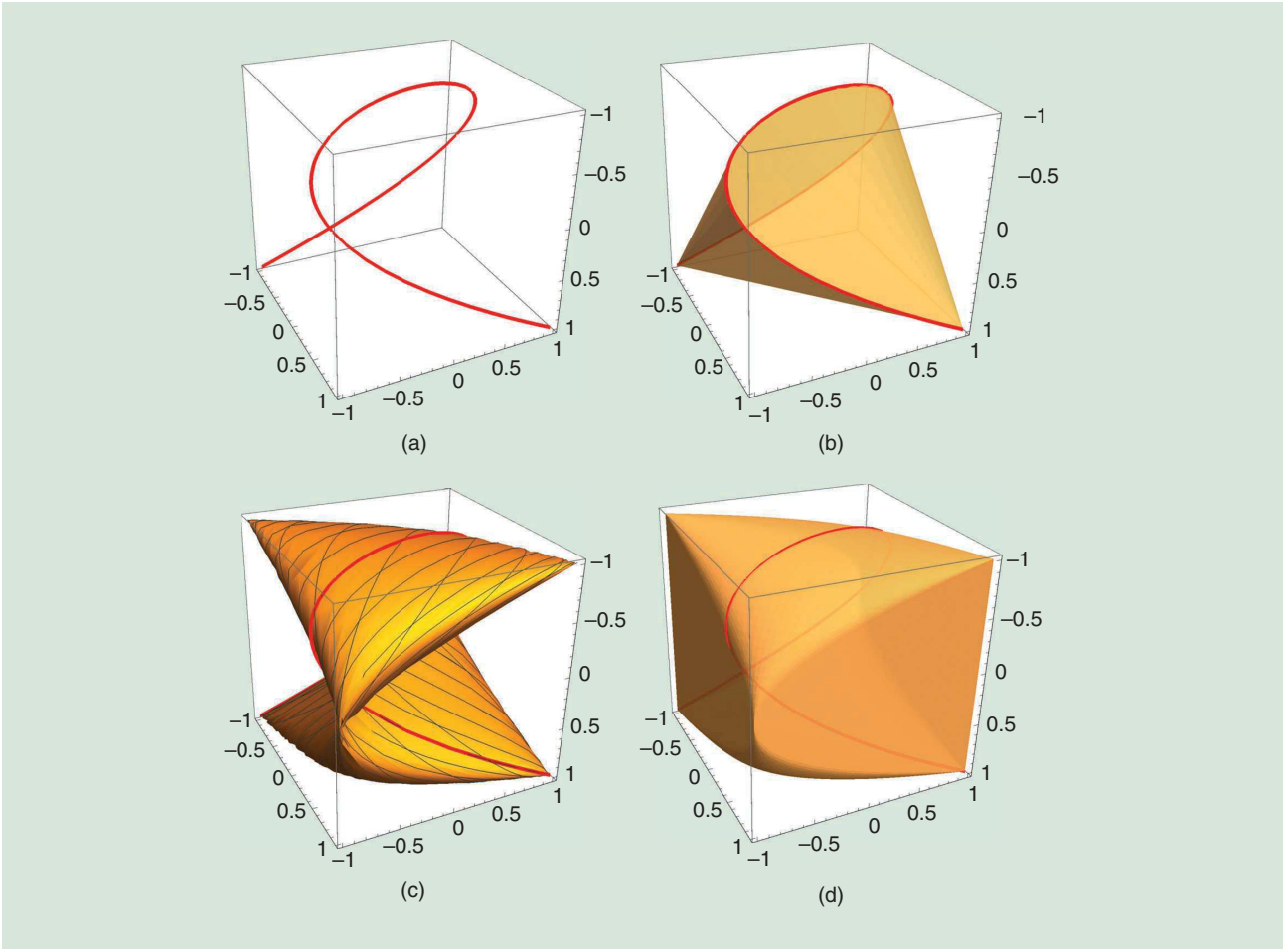


**FIGURE 3.** A visualization of the continuous-valued atomic set for line spectrum superresolution. (a) The moment curve $\mathcal{A}_0$ restricted to three real moments $(\{[\cos(2\pi\tau), \cos(4\pi\tau), \cos(6\pi\tau)]^\top : \tau \in [0,1)\})$ and (b) its convex hull. (c) The phased version of the moment curve $\mathcal{A}_{1D}$ restricted to three real moments $(\{[\cos(2\pi\tau+\phi), \cos(4\pi\tau+\phi), \cos(6\pi\tau+\phi)]^\top : \tau \in [0,1), \phi \in [0,2\pi)\})$ and (d) its convex hull.

which forms a one-dimensional variety of $\mathbb{C}^n$ called the *moment curve*, illustrated in Figure 3(a). It is well known that the convex hull of $\mathcal{A}_0$, illustrated in Figure 3(b), is a body of $\mathbb{C}^n$ that can be parameterized by a set of linear matrix inequalities [23] and has close relationships with the positivity of Hermitian Toeplitz matrices. This fundamental property of the moment curve has many implications in control and signal processing [24], [25] and is key for developing a superresolution theory based on ANM.

It is clearly possible to obtain the same $\boldsymbol{x}$ from different $x(t)$. However, if we impose some sparsity assumption, namely constraining how many spikes are allowed in $x(t)$, this representation can be ensured to be unique. In particular, the representation (8) is unique as long as $r \le \lfloor n/2 \rfloor$ and the support set $\mathcal{T} = \{\tau_k\}_{1 \le k \le r}$ contains distinct elements.

### *Atomic norm for line spectrum superresolution*

To apply the framework of ANM for superresolution, one must first define the atomic set properly. Since the complex amplitudes $c_k$'s can take arbitrary phases, we introduce the following augmented atomic set taking this into account:

$$\mathcal{A}_{1D} = \{e^{j\phi}\boldsymbol{a}(\tau) : \tau \in [0,1), \phi \in [0,2\pi)\}. \tag{12}$$

See an illustration of $\mathcal{A}_{1D}$ and its convex hull in Figure 3(c) and (d). Writing $c_k = |c_k| e^{j\phi_k}$, $\boldsymbol{x}$ can be represented as a positive combination of the atoms in $\mathcal{A}_{1D}$ as $\boldsymbol{x} = \sum_{k=1}^{r} |c_k| e^{j\phi_k} \boldsymbol{a}(\tau_k)$. It is easy to verify that $\mathcal{A}_{1D}$ is centrally symmetric around the origin and consequently it induces an atomic norm over $\mathbb{C}^n$, as defined in (2) and (3). It is worth noting that minimizing the atomic norm of $\boldsymbol{x}$ is equivalent to minimizing the total variation of $x(t)$, i.e.,

$$\min \|x(t)\|_{\mathrm{TV}} \quad \text{s.t.} \quad \boldsymbol{x} = \int_0^1 \boldsymbol{a}(t)\,\mathrm{d}x(t), \tag{13}$$

and both viewpoints are used frequently in the literature.

Remarkably, this atomic norm admits an equivalent SDP characterization, thanks to the following Carathéodory–Fejér–Pisarenko decomposition [26]:

$$\|\boldsymbol{x}\|_{\mathcal{A}} = \inf_{\substack{\boldsymbol{u}\in\mathbb{C}^n \\ t>0}} \left\{ \frac{1}{2n}\mathrm{Tr}(\mathrm{toep}(\boldsymbol{u})) + \frac{1}{2}t : \begin{bmatrix} \mathrm{toep}(\boldsymbol{u}) & \boldsymbol{x} \\ \boldsymbol{x}^{\mathsf{H}} & t \end{bmatrix} \succeq 0 \right\}. \tag{14}$$

## From Bounded Polynomials to Linear Matrix Inequalities

Many applications encountered in signal processing, systems and control theory involve comparing the magnitudes of two real trigonometric polynomials, $R(\tau) = \mathrm{Re}\langle \boldsymbol{a}(\tau), \boldsymbol{r}\rangle$ and $S(\tau) = \mathrm{Re}\langle \boldsymbol{a}(\tau), \boldsymbol{s}\rangle$, e.g., bounding the frequency response of a finite impulse response filter by a desired shape. Although such inequalities, in appearance, require verification over a continuous set of parameters, they can easily be translated into linear matrix inequalities of finite dimension, which are amenable to optimization.

Central to the equivalence is a Gram parametrization of real trigonometric polynomials [24], [S1], [S2], by noticing that every real trigonometric polynomial $R(\tau) = \mathrm{Re}\langle \boldsymbol{a}(\tau), \boldsymbol{r}\rangle$ can be equivalently represented as a quadratic form $R(\tau) = \boldsymbol{a}(\tau)^{\mathrm{H}} \boldsymbol{G} \boldsymbol{a}(\tau)$ for a family of Hermitian matrices $\boldsymbol{G} \in \mathcal{G}(R)$, where $\boldsymbol{G}$ is related to $\boldsymbol{r}$ through the following Gram mapping (Figure S2):

$$\boldsymbol{G} \in \mathcal{G}(R) \Longleftrightarrow \mathrm{Tr}(\boldsymbol{G}) = \mathrm{Re}(r_0),\ \sum_{i=1}^{n-k} G_{i,i+k} = \frac{r_k}{2},\ k = 1, \ldots, n-1. \tag{S1}$$

A remarkable property (e.g., [25, Lemma 4.23]) is that the Gram mapping preserves the partial order between the polynomials and the Hermitian matrices. Let $\boldsymbol{G} \in \mathcal{G}(R)$, then $R(\tau) \le S(\tau)$ holds for every $\tau \in [0, 1)$ if and only if there exists $\boldsymbol{H} \in \mathcal{G}(S)$ such that $\boldsymbol{G} \preceq \boldsymbol{H}$.

As an example, consider the dual norm constraint $\|\boldsymbol{p}\|_{\mathcal{A}}^* \le 1$ in (16), which amounts to upper bounding $R(\tau) = |\langle \boldsymbol{a}(\tau), \boldsymbol{p}\rangle|^2$ by $S(\tau) = 1$. Since $R(\tau) = \boldsymbol{a}(\tau)^{\mathrm{H}} \boldsymbol{p}\boldsymbol{p}^{\mathrm{H}} \boldsymbol{a}(\tau)$, it is clear that $\boldsymbol{p}\boldsymbol{p}^{\mathrm{H}} \in \mathcal{G}(R)$. The constraint holds if and only if their exists a matrix $\boldsymbol{H} \in \mathcal{G}(S)$ satisfying $\boldsymbol{H} \succeq \boldsymbol{p}\boldsymbol{p}^{\mathrm{H}}$. Rewriting this condition using the Schur's complement, as well as expanding the Gram mapping of $S(\tau) = 1$, we can obtain the semidefinite constraint in (17), a consequence also known as the *bounded real lemma*.

### References


[S1] O. Teke and P. P. Vaidyanathan, "On the role of the bounded lemma in the SDP formulation of atomic norm problems," *IEEE Signal Process. Lett.*, vol. 24, no. 7, pp. 972–976, 2017. doi: 10.1109/LSP.2017.2700442.

[S2] M.-D. Choi, T. Y. Lam, and B. Reznick, "Sums of squares of real polynomials," *Proc. Symp. Pure Math.*, vol. 58, no. 2, pp. 103–126, 1995. doi: 10.1090/pspum/058.2/1327293.


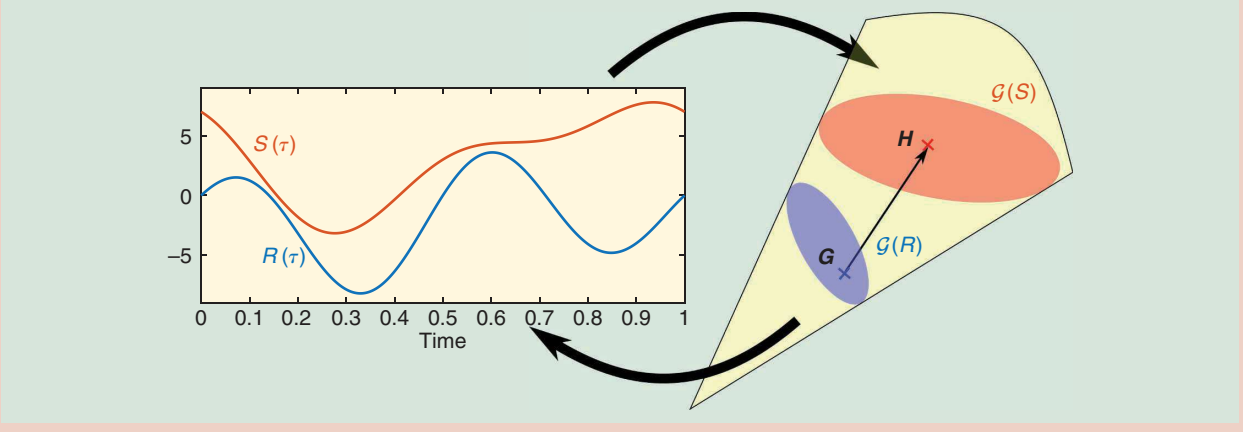


**FIGURE S2.** The Gram mapping for two polynomials satisfying $R(\tau) \le S(\tau)$ over $[0, 1)$, and the corresponding $\boldsymbol{G}$, $\boldsymbol{H}$ in the respective Gram sets such that $\boldsymbol{G} \preceq \boldsymbol{H}$.

Contrary to its abstract definition in (2), the reformulation (14) provides a tractable approach to compute the quantity $\|\boldsymbol{x}\|_{\mathcal{A}}$, which can be accomplished using generic off-the-shelf convex solvers [27]. The Vandermonde decomposition of $\mathrm{toep}(\boldsymbol{u})$, i.e., $\mathrm{toep}(\boldsymbol{u}) = \Sigma_{k=1}^{r'} |c'_k| \boldsymbol{a}(\tau'_k)\boldsymbol{a}(\tau'_k)^{\mathrm{H}}$ can then be used to identify the support $\hat{\mathcal{T}} = \{\tau'_k\}$ of the atomic representation of $\boldsymbol{x}$, as well as the atomic norm $\|\boldsymbol{x}\|_{\mathcal{A}} = \Sigma_{k=1}^{r'} |c'_k|$.

### *Duality and atomic decomposition*

The Lagrangian duality theory marks an important aspect in understanding the atomic norm. The Lagrange dual problem associated with the ANM (2) reads [11]

$$\max_{\boldsymbol{p}} \mathrm{Re}\langle \boldsymbol{x}, \boldsymbol{p}\rangle \quad \text{subject to } \|\boldsymbol{p}\|_{\mathcal{A}}^* \le 1, \tag{15}$$

where the dual atomic norm $\|\boldsymbol{p}\|_{\mathcal{A}}^*$ of a vector $\boldsymbol{p} \in \mathbb{C}^n$ is defined with respect to the atomic set $\mathcal{A}_{1\mathrm{D}}$ as

$$\|\boldsymbol{p}\|_{\mathcal{A}}^* \triangleq \sup_{\boldsymbol{a} \in \mathcal{A}_{1\mathrm{D}}} \mathrm{Re}\langle \boldsymbol{a}, \boldsymbol{p}\rangle = \sup_{\tau \in [0,1)} \underbrace{|\langle \boldsymbol{a}(\tau), \boldsymbol{p}\rangle|}_{P(\tau)}. \tag{16}$$

The last equality of (16) suggests that the dual atomic norm can be interpreted as the supremum of the modulus of a complex trigonometric polynomial $P(\tau) = \langle \boldsymbol{a}(\tau), \boldsymbol{p}\rangle = \Sigma_{\ell=0}^{n-1} p_\ell e^{-j2\pi\ell\tau}$ with coefficients given by the vector $\boldsymbol{p}$. Constraints of this type

are known to be equivalent to linear matrix inequalities involving the positivity of some Hermitian matrices (see "From Bounded Polynomials to Linear Matrix Inequalities"). The dual program (15) can be reformulated into an SDP as

$$\begin{aligned} \max_{\boldsymbol{p}\in\mathbb{C}^n,\boldsymbol{H}\in\mathbb{C}^{n\times n}} \quad & \mathrm{Re}\langle \boldsymbol{x},\boldsymbol{p}\rangle \\ \text{subject to} \quad & \begin{bmatrix} \boldsymbol{H} & \boldsymbol{p} \\ \boldsymbol{p}^{\mathsf{H}} & 1 \end{bmatrix} \succeq 0 \\ & \sum_{i=1}^{n-k} H_{i,i+k} = \delta_k, k = 0,\ldots,n-1, \end{aligned} \tag{17}$$

where $H_{i,j}$ is the $(i,j)$th entry of the matrix $\boldsymbol{H}$ and the indicator function $\delta_k$ equals 1 if $k = 0$ and 0 otherwise.

Another merit of the dual formulation is that the support set of the atomic decomposition can be inferred from the optimal solution $\hat{\boldsymbol{p}}$ to the dual problem (15), by examining the dual polynomial $\hat{P}(\tau) = \langle \boldsymbol{a}(\tau), \hat{\boldsymbol{p}}\rangle$. We identify the spikes as the locations of the extreme values of the modulus of $\hat{P}(\tau)$:

$$\hat{\mathcal{T}} = \{\tau : |\hat{P}(\tau)| = 1\}. \tag{18}$$

This is possible, because, under strong duality, both the primal and the dual problems must share the same optimal objective value, i.e., $\|\boldsymbol{x}\|_{\mathcal{A}} = \sum_{k=1}^{r'} |c'_k|$, where $\boldsymbol{x} = \sum_{k=1}^{r'} c'_k \boldsymbol{a}(\tau'_k)$ is the atomic decomposition of $\boldsymbol{x}$. Consequently, the optimal value of the dual program becomes

$$\mathrm{Re}\langle \boldsymbol{x},\hat{\boldsymbol{p}}\rangle = \mathrm{Re}\left\langle \sum_{k=1}^{r'} c'_k \boldsymbol{a}(\tau'_k), \hat{\boldsymbol{p}} \right\rangle = \mathrm{Re} \sum_{k=1}^{r'} c'^{*}_k \hat{P}(\tau'_k),$$

indicating $\hat{P}(\tau'_k) = \mathrm{sgn}(c'_k) = c'_k/|c'_k|$ whenever the atomic decomposition is nonvanishing at $\tau'_k$. This approach is illustrated in Figure 4 for a length-33 signal with six spikes, where the peaks of $\hat{P}(\tau)$ match the locations of the true spikes, indicating the atomic decomposition perfectly recovers the true sparse representation.

The atomic norm offers an approach for line spectrum superresolution that is drastically different from traditional methods, which rely heavily on the correctness of model order estimation. The dual polynomial approach, in contrast, does not require any prior knowledge of the model order and can estimate the spikes with an infinitesimal precision.

### Exact recovery guarantees

So far, we have explained the algorithmic approach of ANM for line spectrum superresolution. A central question regards understanding whether this convex relaxation is tight or not. More precisely, one would like to identify the conditions under which the estimated support $\hat{\mathcal{T}}$ coincides with the true support $\mathcal{T}$ of the signal $\boldsymbol{x}$ and, correspondingly, the atomic decomposition $\boldsymbol{x} = \sum_{k=1}^{r'} c'_k \boldsymbol{a}(\tau'_k)$ coincides with the sparsest representation $\boldsymbol{x} = \sum_{k=1}^{r} c_k \boldsymbol{a}(\tau_k)$ over the atomic set $\mathcal{A}_{\mathrm{1D}}$.

Such questions were extensively addressed in the context of $\ell_1$ norm minimization, where the atomic set $\mathcal{A}$ has a finite number of elements. The performance guarantees often depend on specific structural properties of $\mathcal{A}$, formalized in the notion of restricted isometry property (RIP) [28] or certain incoherence properties [29]. Unfortunately, these properties do not hold when considering a continuous dictionary such as $\mathcal{A}_{\mathrm{1D}}$, since two atoms $\boldsymbol{a}(\tau)$ and $\boldsymbol{a}(\tau+\delta)$ can be more and more correlated with each other as their separation $\delta$ tends to zero, leading to arbitrarily small RIP or coherence constants.

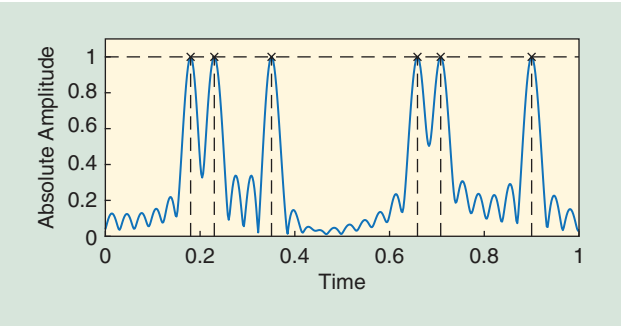


**FIGURE 4.** A spike localization via pinpointing the peaks of the dual polynomial $|\hat{P}(\tau)|$ (in blue) associated with the optimal solution $\hat{\boldsymbol{p}}$ of the dual program (15) for a signal $\boldsymbol{x}$ of length $n = 33$ with six true spikes (in black).

Nonetheless, one could question for which class of signals is the relaxation tight. Leveraging duality theory, the atomic norm approach is tight for a fixed signal $\boldsymbol{x}$, i.e., $\hat{\mathcal{T}} = \mathcal{T}$, as long as there exists a dual certificate $\boldsymbol{p}_\star$, such that $P_\star(\tau) = \langle \boldsymbol{a}(\tau), \boldsymbol{p}_\star\rangle$ satisfies [11]

$$P_\star(\tau_k) = \mathrm{sgn}(c_k), \quad \forall \tau_k \in \mathcal{T}, \tag{19a}$$

$$|P_\star(\tau)| < 1, \quad \forall \tau \notin \mathcal{T}. \tag{19b}$$

In other words, it amounts to finding an $(n-1)$-order trigonometric polynomial that interpolates sign patterns of the spike signal at the spike locations, as well as is bounded in magnitude by 1.

Intuitively, the difficulty of interpolation depends on the separations between the spikes in $\mathcal{T}$ and, more precisely on the minimal separation, or the minimal wrap-around distance between any pair of distinct spikes in $\mathcal{T}$, defined formally as

$$\Delta_{\mathbb{T}}(\mathcal{T}) \triangleq \inf_{\substack{\tau,\tau'\in\mathcal{T},q\in\mathbb{Z} \\ \tau\neq\tau'}} \min |\tau - \tau' + q|. \tag{20}$$

This metric is illustrated in Figure 5 and reflects the periodic behavior of the atom $\boldsymbol{a}(\tau+q) = \boldsymbol{a}(\tau)$ for any integer $q \in \mathbb{Z}$. For instance, if $\mathcal{T} = \{0.1, 0.9\}$, then $\Delta_{\mathbb{T}}(\mathcal{T}) = 0.2$.

A remarkable result, established by Candès and Fernandez-Granda in [11], is that for sufficiently large $n$, a valid certificate can be constructed in a deterministic fashion, as long as the separation condition $\Delta_{\mathbb{T}}(\mathcal{T}) \geq 4/(n-1)$ holds, regardless of the complex amplitudes of the spikes. Furthermore, this result does not make any randomness assumptions on the signal.

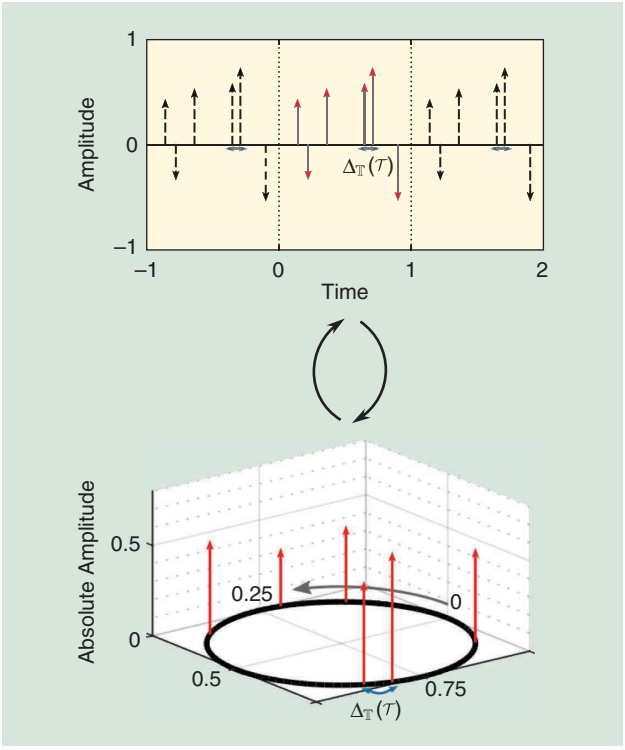


**FIGURE 5.** A representation of the minimal wrap-around distance $\Delta_{\mathbb{T}}(\mathcal{T})$ for a given set of spikes $\mathcal{T}$. The distance corresponds to the minimal gap between any element $\tau \in \mathcal{T}$ and any distinct elements in the aliased set $\mathcal{T} + \mathbb{Z}$.

Later, this separation condition was further improved by Fernandez-Granda [30] to

$$\Delta_{\mathbb{T}}(\mathcal{T}) > \frac{2.52}{n-1}.$$

Conversely, there exist some spike signals with $\Delta_{\mathbb{T}}(\mathcal{T}) < 2/(n-1)$ such that ANM fails to resolve [31].

## Atomic norm denoising

In practice, the observations are corrupted by noise and no estimator can exactly recover the spike signal $x(t)$. This raises a natural question regarding the robustness of the estimate produced by ANM methods. When the noise is additive and the observation $\boldsymbol{z}$ obeys the noisy model (7), it has been proposed to estimate $\boldsymbol{x}$ by searching around the observation $\boldsymbol{z}$ for signals with small atomic norms [32]:

$$\min_{\boldsymbol{x}} \ \frac{1}{2}\| \boldsymbol{x} - \boldsymbol{z} \|_2^2 + \lambda \| \boldsymbol{x} \|_{\mathcal{A}}, \tag{21}$$

where $\lambda > 0$ is a regularization parameter that draws a tradeoff between the fidelity to the observation and the size of the atomic norm. This method, known as *atomic norm denoising*, can be interpreted as a generalization of the celebrated LASSO estimator [33].

When the noise vector $\tilde{\boldsymbol{\epsilon}}$ is composed of independent identically distributed (i.i.d.) complex Gaussian entries with zero mean and variance $\sigma^2$, the mean square error (MSE) of the estimate $\hat{\boldsymbol{x}}$ returned by (21) can be bounded as [32]

$$\frac{1}{n}\| \hat{\boldsymbol{x}} - \boldsymbol{x} \|_2^2 = O\left( \sigma \sqrt{\frac{\log n}{n}} \sum_{k=1}^{r} |c_k| \right)$$

with a high probability by setting $\lambda = \eta\sigma\sqrt{n \log n}$ for some constant $\eta \in (1, \infty)$, e.g., $\eta = 1.2$ in practice. This error rate can be significantly improved when the spikes satisfy the separation condition $\Delta_{\mathbb{T}}(\mathcal{T}) \geq 4/(n-1)$ where with high probability one has [34]

$$\frac{1}{n}\| \hat{\boldsymbol{x}} - \boldsymbol{x} \|_2^2 = O\left( \sigma^2 \frac{r \log n}{n} \right).$$

This last error rate is near-optimal up to some logarithmic factor, since no estimator can achieve an MSE below the rate $O(\sigma^2 r \log(n/r)/n)$ [34].

A more important performance criteria in superresolution concerns the stability of the support estimate $\hat{\mathcal{T}}$, which has been studied in [34]–[38]. When the spikes satisfy the separation condition $\Delta_{\mathbb{T}}(\mathcal{T}) > 5.0018/(n-1)$, and the complex amplitudes of the coefficients $\{c_k\}_{1 \leq k \leq r}$ have approximately the same modulus, then it is established in [36] that the atomic decomposition of the output $\hat{\boldsymbol{x}}$ of (21) is composed of the same number of spikes, i.e., $|\hat{\mathcal{T}}| = |\mathcal{T}| = r$ and that the estimated parameters satisfy

$$|c_k|\,|\hat{\tau}_k - \tau_k| = O\left( \sigma \frac{\sqrt{\log n}}{n^{3/2}} \right), |c_k - \hat{c}_k| = O\left( \sigma \sqrt{\frac{\log n}{n}} \right)$$

with high probability. Altogether, it can be seen that atomic norm denoising achieves near-optimal performance guarantees as long as the spikes are separated by a few times the RL (see "Is the Separation Condition Necessary?").

It is natural to wonder how atomic norm denoising fares compared with classical approaches such as Prony and MUSIC for line spectrum estimation. We examine their ability to resolve close-located spikes with opposite signs, where the reconstruction performance is measured in terms of the MSE of the estimated spike locations $\hat{\mathcal{T}}$, and for different values of separation $\Delta_{\mathbb{T}}(\mathcal{T}) = \alpha/n$. Figure 6 shows the MSE of atomic norm denoising, Prony's method with Cadzow denoising [39], and root-MUSIC [40] with respect to the SNR defined as $\|\boldsymbol{x}\|_2^2/(n\sigma^2)$, benchmarked against the Cramér–Rao bound (CRB), when the separation parameter $\alpha = 2$, 1.75, 1.5, respectively. It is clear that atomic norm denoising outperforms classical approaches and approaches the CRB at a much lower SNR.

## A faster algorithm via ADMM

While the SDP formulation is tractable, its computational complexity is prohibitive when solving large-dimensional problems. Fortunately, it is possible to develop tailored algorithms that are significantly faster. For conciseness, we will discuss one approach based on the alternating direction method of multipliers (ADMM) [32]. The general

## Is the Separation Condition Necessary?

One might wonder if requiring a separation condition makes atomic norm minimization inferior, since many methods do not require such a separation in the noise-free case. However, some form of separation is unavoidable for stable recovery in noisy superresolution, no matter which method is used [S3]. In particular, [S3] shows that when $\Delta_{\mathbb{T}}(\mathcal{T}) < 2/n$, there exists a pair of spike signals $x(t)$ and $x'(t)$ with the same minimal separation, such that no estimator can distinguish them. Figure S3(a) exhibits such a pair of positive spike signals (see [S3] for its construction) with a minimal separation $1.7/n$, where their observations are very close. Figure S3(b) further demonstrates the distance between their observations as the signal dimension increases, for different separation parameter $\alpha$, where $\Delta_{\mathbb{T}}(\mathcal{T}) = \alpha/n$. It is clear that their observations are increasingly indistinguishable as the signal dimension tends to infinity when $\Delta_{\mathbb{T}}(\mathcal{T}) < 2/n$.

### Reference

[S3] A. Moitra, "Super-resolution, extremal functions and the condition number of Vandermonde matrices," in *Proc. 47th Annual ACM Symp. Theory of Computing*, 2015, pp. 821–830. doi: 10.1145/2746539.2746561.

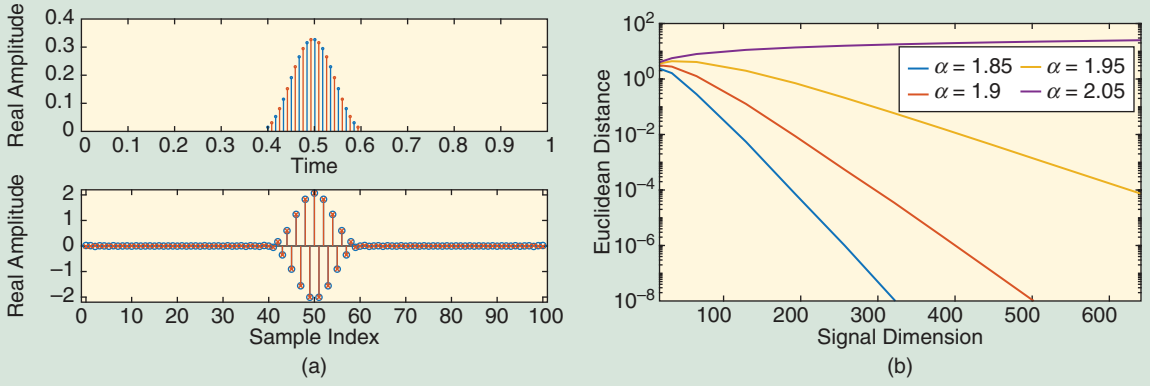


**FIGURE S3.** (a) A pair of spike signal $x(t)$ and $x'(t)$ with a minimal separation $\Delta_{\mathbb{T}}(\mathcal{T}) = 1.7/n$ (top), yet their observations produced according to (10) are almost indistinguishable (bottom). (b) The Euclidean distance between the observations of $x(t)$, $x'(t)$ with a minimal distance $\Delta_{\mathbb{T}}(\mathcal{T}) = \alpha/n$ as a function of the signal length $n$, for different values of the separation parameter $\alpha$.

principle of ADMM is to split the quadratically augmented Lagrangian function of an optimization problem into a sum of separable subfunctions [41]. Each iteration of the algorithm consists of performing independent local minimization on each of those quantities, while ensuring that the feasibility constraints are always satisfied. The iterations run until both primal and dual residuals satisfy a predefined tolerance level.

We take atomic norm denoising (21) as an example, which, in light of (14), can be equivalently rewritten as

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{u},t} \quad & \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{z}\|_2^2 + \frac{\lambda}{2}\Big(\frac{1}{n}\mathrm{Tr}\,(\mathrm{toep}(\boldsymbol{u})) + t\Big) \\ \text{subject to} \quad & \boldsymbol{S} = \begin{bmatrix} \mathrm{toep}\,(\boldsymbol{u}) & \boldsymbol{x} \\ \boldsymbol{x}^{\mathsf{H}} & t \end{bmatrix}, \boldsymbol{S} \succeq 0. \end{aligned}$$

The above program has been "augmented" by introducing an intermediate variable $\boldsymbol{S}$ for the purpose of decoupling the positive semidefinite constraint on the matrix $\boldsymbol{S}$ from the linear constraints on its structure. The augmented Lagrangian $\mathcal{L}$ is given as

$$\begin{aligned} \mathcal{L}(\boldsymbol{x},\boldsymbol{u},t,\boldsymbol{\Sigma},\boldsymbol{S}) = {} & \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{z}\|_2^2 + \frac{\lambda}{2}\Big(\frac{1}{n}\mathrm{Tr}\,(\mathrm{toep}(\boldsymbol{u})) + t\Big) \\ & + \Big\langle \boldsymbol{\Sigma}, \boldsymbol{S} - \begin{bmatrix} \mathrm{toep}\,(\boldsymbol{u}) & \boldsymbol{x} \\ \boldsymbol{x}^{\mathsf{H}} & t \end{bmatrix} \Big\rangle \\ & + \frac{\rho}{2}\Big\| \boldsymbol{S} - \begin{bmatrix} \mathrm{toep}\,(\boldsymbol{u}) & \boldsymbol{x} \\ \boldsymbol{x}^{\mathsf{H}} & t \end{bmatrix} \Big\|_{\mathrm{F}}^2, \end{aligned}$$

where $\boldsymbol{S}$ and $\boldsymbol{\Sigma}$ are $(n+1)$-dimensional Hermitian matrices and $\rho > 0$ is a regularization parameter. The successive update steps to minimize the augmented Lagrangian are given in Algorithm 1. Closed-form solutions can be found for the first update step, yielding a very efficient implementation. The second update is the most costly part, as a projection over the cone of positive semidefinite Hermitian matrices has to be computed. This computation is typically achieved using power methods [42], with a computational complexity of $O(n^3)$ per iteration.

### *Can we discretize?*

It may be worthwhile to pause and compare ANM to other approaches based on convex optimization for superresolution, particularly $\ell_1$ minimization that is widely popular

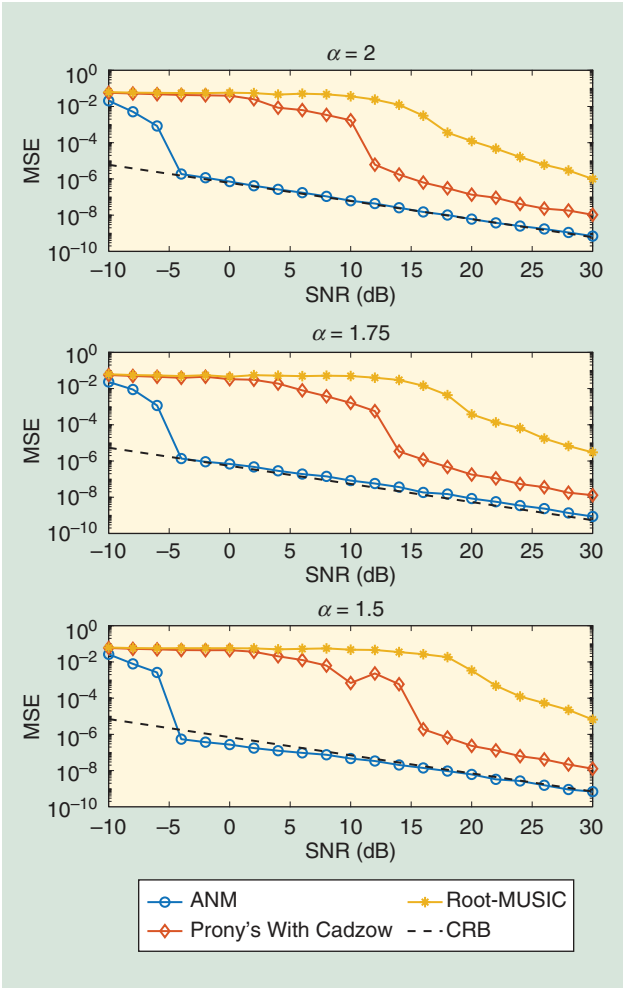


**FIGURE 6.** The MSE of different methods for estimating two spikes with opposite signs separated by $\Delta_{\mathbb{T}}(\mathcal{T}) = \alpha/n$, averaged more than 200 Monte Carlo trials and benchmarked against the CRB for different separation parameter $\alpha = 2, 1.75, 1.5$. Here, the signal length is $n = 101$. SNR: signal-to-noise ratio.

**Algorithm 1. ADMM for atomic norm denoising [32].**

**Input:** Observation $\boldsymbol{z}$; parameters $\lambda, \rho > 0$;
Initialize $j = 0$, and $\boldsymbol{\Sigma}_0, \boldsymbol{S}_0$ to zero matrices
**Repeat** until stopping criteria

$$(\boldsymbol{x}_{j+1}, \boldsymbol{u}_{j+1}, t_{j+1}) \leftarrow \operatorname*{argmin}_{\boldsymbol{x},\boldsymbol{u},t} \mathcal{L}(\boldsymbol{x}, \boldsymbol{u}, t, \boldsymbol{\Sigma}_j, \boldsymbol{S}_j);$$

$$\boldsymbol{S}_{j+1} \leftarrow \operatorname*{argmin}_{\boldsymbol{S} \succeq 0} \mathcal{L}(\boldsymbol{x}_{j+1}, \boldsymbol{u}_{j+1}, t_{j+1}, \boldsymbol{\Sigma}_j, \boldsymbol{S});$$

$$\boldsymbol{\Sigma}_{j+1} \leftarrow \boldsymbol{\Sigma}_j + \rho\left(\boldsymbol{S}_{j+1} - \begin{bmatrix} \operatorname{toep}(\boldsymbol{u}_{j+1}) & \boldsymbol{x}_{j+1} \\ \boldsymbol{x}_{j+1}^{\mathrm{H}} & t_{j+1} \end{bmatrix}\right);$$

$$j \leftarrow j + 1;$$

**Output:** $\boldsymbol{x}_j$

for high-resolution imaging and localization in the recent literature due to compressed sensing (CS) [43], [44].

The $\ell_1$ norm can be seen as a discrete approximation of the atomic norm. Indeed, taking the atomic set $\mathcal{A}_{1D}$, one can pick a desired resolution $Q$ and discretize it as

$$\mathcal{A}_{1D,\text{discrete}} = \left\{ e^{j\phi} \boldsymbol{a}\left(\frac{q}{Q}\right) : q = 0, \ldots, Q-1, \phi \in [0, 2\pi) \right\},$$

and then perform $\ell_1$ minimization over $\mathcal{A}_{1D,\text{discrete}}$. The convex hull of $\mathcal{A}_{1D,\text{discrete}}$ approaches that of $\mathcal{A}_{1D}$ as the discretization gets finer, which suggests the performance of $\ell_1$ minimization over the discretized dictionary approaches that of ANM asymptotically [45]. If the spike signal meets a so-called nondegenerate source condition [46, Definition 2], this approach will return a sparse solution supported on the elements of the discretized $\mathcal{A}_{1D,\text{discrete}}$ surrounding the ground-truth spikes, when the noise is small enough [46], [47]. However, it remains unclear which class of spike signals satisfies the nondegenerate source condition in practice.

However, this discretization may come with several undesired consequences when the grid size $Q$ is finite in practice. The theory of $\ell_1$ minimization only provides exact recovery guarantees when the spikes of $x(t)$ lie on the grid, which is unrealistic. In fact, there is always an inevitable basis mismatch [48] between the spikes represented in the discretized dictionary $\mathcal{A}_{1D,\text{discrete}}$ and the true spikes, no matter how fine the grid is. Perfect recovery is not possible in this situation, even in the absence of noise due to this mismatch. Furthermore, one can find signals whose representations in $\mathcal{A}_{1D,\text{discrete}}$ are not compressible due to spectral leakage and therefore are poorly recovered using $\ell_1$ minimization, e.g., the recovery may contain many spurious spikes. Therefore, cautions are needed to account for such consequences when applying discretization, and efforts to mitigate the basis mismatch have been proposed extensively, e.g., [49] and [50].

## Generalizations of atomic sets

The tool of atomic norms can be extended easily to handle a wide range of scenarios in a unified manner, by properly adjusting the atomic set for signal decompositions such as incorporating prior information and dealing with multidimensional settings and multiple measurement vectors, to illustrate a few.

### *Atomic set for positive spikes*

In some applications, there exist additional information about the spikes, such as the coefficients of the spikes in (8) are positive, i.e., $c_k > 0$. Examples include neural spike sorting, fluorescence microscopy imaging, or covariance-based spectrum estimation for noncoherent sources [1].

In this case, the atomic set reduces to the moment curve $\mathcal{A}_0$ in (11). The induced $\|\boldsymbol{x}\|_{\mathcal{A}}$ is no longer a norm, since $\mathcal{A}_0$ is not centrally symmetric but, nonetheless, similar SDP characterization still holds. To be specific, the dual program now becomes

$$\max_{\boldsymbol{p} \in \mathbb{C}^n} \operatorname{Re}\langle \boldsymbol{x}, \boldsymbol{p} \rangle \quad \text{subject to} \quad \sup_{\tau \in [0,1)} \operatorname{Re}\langle \boldsymbol{a}(\tau), \boldsymbol{p} \rangle \le 1,$$

where the constraint bounds the real part of the trigonometric polynomial $P(\tau) = \langle \boldsymbol{a}(\tau), \boldsymbol{p} \rangle$. Using the Fejér-Riesz theorem (see, e.g., [25, Th. 1.1] and "From Bounded Polynomials to Linear Matrix Inequalities"), this can be equivalently represented as

$$\max_{\boldsymbol{p}\in\mathbb{C}^n,\,\boldsymbol{H}\succeq 0} \quad \mathrm{Re}\langle \boldsymbol{x}, \boldsymbol{p}\rangle$$
$$\text{subject to} \quad \mathrm{Tr}(\boldsymbol{H}) + \mathrm{Re}(p_0) = 1,$$
$$\sum_{i=1}^{n-k} H_{i,i+k} + p_k/2 = 0, \quad k = 1, \ldots, n-1.$$

It is long established [51], [52] that the spikes can be perfectly localized as long as $r \le \lfloor (n-1)/2 \rfloor$, without requiring any separation between the spikes, as long as they are positive. The stability of this approach as well as the implications of nonnegative constraints for other atomic sets are further studied in [53]–[56]. As a comparison, Figure 7 shows the MSE of atomic norm denoising with and without positive constraints, Prony's method with Cadzow denoising [39], and root-MUSIC [40] with respect to the SNR defined as $\|\boldsymbol{x}\|_2^2/(n\sigma^2)$ for resolving two spikes with positive signs, separated by $\Delta_{\mathbb{T}}(\mathcal{T}) = \alpha/n$ for $\alpha = 1, 0.75, 0.5$, respectively. It can be seen that atomic norm denoising still outperforms classical approaches and, in particular, incorporating the positive constraint leads to further improvements.

### *Atomic set for multidimensional spikes*

When the spikes reside in a multidimensional space, one can extend the 1D model in a straightforward manner. Here, we illustrate the setup for the 2D case, where each entry of the signal $\boldsymbol{X}_{\mathrm{2D}} \in \mathbb{C}^{n_1 \times n_2}$ can be expressed as a superposition of $r$ complex sinusoids propagating in two directions:

$$\boldsymbol{X}_{\mathrm{2D}} = \sum_{k=1}^{r} c_k \boldsymbol{a}_1(\tau_{1,k}) \boldsymbol{a}_2(\tau_{2,k})^{\top}, \tag{22}$$

where $c_k$ and $\boldsymbol{\tau}_k = [\tau_{1,k}, \tau_{2,k}]^{\mathsf{T}} \in [0,1)^2$ are the complex amplitude and location of the $k$th spike, and $\boldsymbol{a}_i(\tau)$ is given by (9) with the dimension parameter replaced by $n_i$, $i = 1, 2$. It is natural to define the corresponding atomic set as [11], [57]

$$\mathcal{A}_{\mathrm{2D}} = \{ e^{j\phi} \boldsymbol{a}_1(\tau_1) \boldsymbol{a}_2(\tau_2)^{\top} : \boldsymbol{\tau} \in [0,1)^2, \phi \in [0, 2\pi) \},$$

and the atomic norm according to (2). To localize the spikes, one could similarly study the associated dual problem

$$\max_{\boldsymbol{P}\in\mathbb{C}^{n_1\times n_2}} \mathrm{Re}\langle \boldsymbol{X}, \boldsymbol{P}\rangle \quad \text{subject to } \|\boldsymbol{P}\|_{\mathcal{A}}^* \le 1,$$

where the dual atomic norm can be reinterpreted as the supremum of a bivariate complex trigonometric polynomial $P(\tau_1, \tau_2) = \langle \boldsymbol{a}_1(\tau_1)\boldsymbol{a}_2(\tau_2)^{\top}, \boldsymbol{P}\rangle$ with the matrix $\boldsymbol{P}$ as its coefficients. Again, one can localize the spikes by examining the extremal points of the dual polynomial, which is illustrated in Figure 8. Cautions need to be taken when attempting to solve the dual program in two or higher dimensions, since the bounded real lemma [24], [25] does not hold anymore. Instead, a precise characterization requires solving a hierarchy of sum-of-squares relaxations and, fortunately in practice, the first level usually suffices [24], [57], [58].

The tightness of the ANM approach is closely related to a separation condition analogous to the 1D case [30]. Namely,

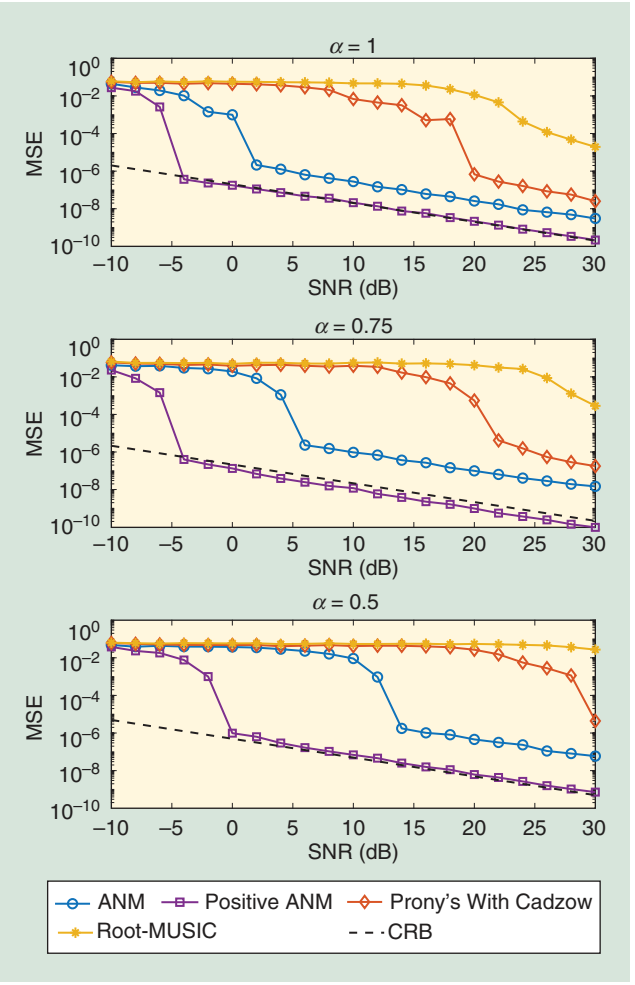


**FIGURE 7.** The MSE of different methods for estimating two spikes with positive signs separated by $\Delta_{\mathbb{T}}(\mathcal{T}) = \alpha/n$, averaged more than 200 Monte Carlo trials and benchmarked against the CRB for different separation parameter $\alpha = 1,\ 0.75,\ 0.5$. Here, the signal length is $n = 101$.

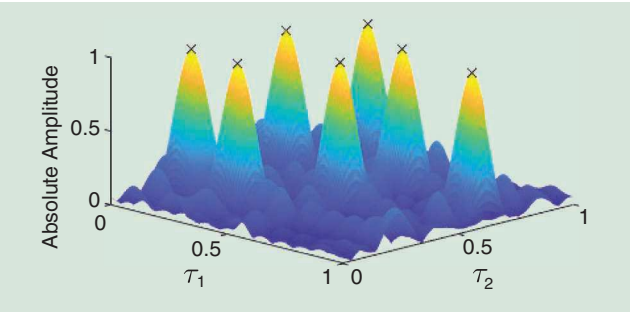


**FIGURE 8.** The spike localization using the dual polynomial approach for 2D spikes via ANM. Here, we set $n_1 = 12$, $n_2 = 10$, and $r = 7$.

the atomic decomposition is unique and exact, as soon as there exists a universal constant $C > 0$ such that the set of spikes $\mathcal{T} = \{\boldsymbol{\tau}_k\}_{1\le k\le r}$ satisfies

$$\Delta_{\mathbb{T}^2}(\mathcal{T}) \triangleq \inf_{\substack{\tau,\tau'\in\mathcal{T}\\ \tau\neq\tau'}} \min_{\boldsymbol{q}\in\mathbb{Z}^2} \|\boldsymbol{\tau} - \boldsymbol{\tau}' + \boldsymbol{q}\|_{\infty} > \frac{C}{\min(n_1, n_2) - 1}.$$

Moreover, if the signal $\boldsymbol{X}_{\text{2D}}$ is real valued, $C = 4.76$ suffices to guarantee exact recovery of the spikes.

### *Atomic set for multiple measurement vectors*

One can collect multiple snapshots of observations, where they share the same spike locations with varying coefficients. Consider $T$ snapshots, stacked in a matrix, $\boldsymbol{X}_{\text{MMV}} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T]$, which is expressed similarly to (8) as

$$\boldsymbol{X}_{\text{MMV}} = \sum_{k=1}^{r} \boldsymbol{a}(\tau_k)\boldsymbol{c}_k^{\top}, \tag{23}$$

where $\boldsymbol{c}_k = [c_{1,k}, \ldots, c_{L,k}] \in \mathbb{C}^T$ is the coefficient of the $k$th spike across the snapshots. Following the recipe of atomic norms, we define the atoms as

$$\boldsymbol{A}(\tau, \boldsymbol{b}) = \boldsymbol{a}(\tau)\boldsymbol{b}^{\top},$$

where $\tau \in [0, 1)$, $\boldsymbol{b} \in \mathbb{C}^T$ with $\|\boldsymbol{b}\|_2 = 1$. The atomic set is defined as

$$\mathcal{A}_{\text{MMV}} = \{\boldsymbol{A}(\tau, \boldsymbol{b}) : \tau \in [0, 1), \|\boldsymbol{b}\|_2 = 1\}.$$

The atomic norm can then be defined following (2), which turns out sharing similar nice SDP characterizations for primal and dual formulations as for the single snapshot model [59]. The atomic norm $\|\boldsymbol{X}_{\text{MMV}}\|_{\mathcal{A}}$ can be written equivalently as

$$\|\boldsymbol{X}_{\text{MMV}}\|_{\mathcal{A}} = \inf_{\substack{\boldsymbol{u} \in \mathbb{C}^n \\ \boldsymbol{W} \in \mathbb{C}^{T \times T}}} \left\{ \frac{1}{2n}\text{Tr}(\text{toep}(\boldsymbol{u})) + \frac{1}{2}\text{Tr}(\boldsymbol{W}) : \begin{bmatrix} \text{toep}(\boldsymbol{u}) & \boldsymbol{X}_{\text{MMV}} \\ \boldsymbol{X}_{\text{MMV}}^{\text{H}} & \boldsymbol{W} \end{bmatrix} \succeq 0 \right\}. \tag{24}$$

A curious comparison can be drawn to the nuclear norm by noticing that one recovers the nuclear norm of $\boldsymbol{X}_{\text{MMV}}$ by replacing the principal block $\text{toep}(\boldsymbol{u})$ in (24) with an arbitrary positive semidefinite matrix. The fact that $\text{toep}(\boldsymbol{u})$ has significantly fewer degrees of freedom ($n$ versus $n^2$) is in parallel to that $\boldsymbol{a}(\tau)$ has significantly fewer degrees of freedom than an arbitrary vector (1 versus $n$).

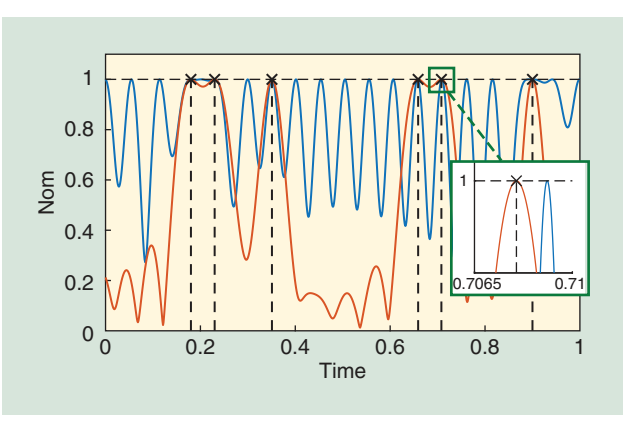


**FIGURE 9.** The dual polynomial using multiple snapshots (in red) successfully localize all of the spikes while the one using a single snapshot (in blue) fails for the same spike signal. Here, $r = 6$, $n = 21$, and $T = 6$.

Again, one can determine the atomic decomposition and localize the spikes by resorting to the dual program in a similar fashion. Figure 9 illustrates how multiple snapshots improve the performance of localization over the single snapshot case when the coefficients across snapshots exhibit some kind of diversity, e.g., generated with i.i.d. complex Gaussian entries.

## Generalizations of measurement models

So far, we have seen that ANM provides a means for super-resolution via convex relaxation in additive Gaussian noise. The framework of convex optimization is quite versatile and can be extended to handle models when the measurements are partially observed, corrupted, contain interfering sources, or even come from unknown modulations. This is an important advantage over classical methods such as MUSIC or ESPRIT, as most of them cannot be extended easily to these variants of models.

### *Compressed spectral sensing*

CS [43], [44] has suggested that it is possible to recover a signal using a number of measurements that is proportional to its degrees of freedom, rather than its ambient dimension. Consider the problem where only a subset of entries of $\boldsymbol{x}$ is observed,

$$\boldsymbol{y}_{\text{CS}} = \boldsymbol{A}_{\text{CS}}\boldsymbol{x},$$

where $\boldsymbol{A}_{\text{CS}} \in \mathbb{C}^{m \times n}$, and $m \ll n$ representing compressive acquisition of the signal $\boldsymbol{x}$. The goal is to recover $\boldsymbol{x}$ and its spectral content from $\boldsymbol{y}_{\text{CS}} \in \mathbb{C}^m$, the compressive measurements. This has applications in wide-band spectrum sensing and cognitive radio [60], for example.

One can easily extend the framework of ANM and recover $\boldsymbol{x}$ by solving the program

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_{\mathcal{A}} \quad \text{subject to} \quad \boldsymbol{y}_{\text{CS}} = \boldsymbol{A}_{\text{CS}}\boldsymbol{x}.$$

When $\boldsymbol{A}_{\text{CS}}$ is a partial observation matrix, namely, a subset of $m$ entries of $\boldsymbol{x}$ is observed uniformly at random, then $\boldsymbol{x}$ can be perfectly recovered with high probability using $m = O(\log^2 n + r \log r \log n)$ measurements as long as $\boldsymbol{x}$ satisfies the separation condition, and with random signs of the coefficients $c_k$'s [15, Th. II.3]. More generally, a broader class of measurement matrices $\boldsymbol{A}_{\text{CS}}$ can be allowed where its rows are drawn independently from some isotropic and incoherent distribution [61], [62], and exact recovery is possible under the same separation condition using a number of measurements on the order of $r$ up to some logarithmic factors. In addition, quantized measurements are further dealt in [63] with theoretical guarantees.

### *Demixing sinusoids and spikes*

Due to sensor failures or malicious environments, the measurements are susceptible to corruptions that can take arbitrary magnitudes. To this end, consider the problem when the observations are contaminated by sparse outliers, where

$$\boldsymbol{y}_{\text{corrupt}} = \boldsymbol{x} + \boldsymbol{s}.$$

Here, $\boldsymbol{s}$ is a sparse vector, where its nonzero entries correspond to corruptions of the observations. The goal is to decompose $\boldsymbol{x}$ and $\boldsymbol{s}$ from $\boldsymbol{y}_{\text{corrupt}}$, a problem intimately related to the uncertainty principle of signal decomposition in [17] and [64] and sparse error correction in CS [65].

Leveraging low-dimensional structures in both $\boldsymbol{x}$ and $\boldsymbol{s}$, we seek $\boldsymbol{x}$ with a small atomic norm and $\boldsymbol{s}$ with a small $\ell_1$ norm that satisfies the observation constraint [66]

$$\min_{\boldsymbol{x},\boldsymbol{s}} \|\boldsymbol{x}\|_{\mathcal{A}} + \lambda \|\boldsymbol{s}\|_1 \quad \text{subject to} \quad \boldsymbol{y}_{\text{corrupt}} = \boldsymbol{x} + \boldsymbol{s},$$

where $\lambda > 0$ is some regularization parameter. As long as the sample size is sufficiently large [66, Th. 2.2] and the spikes satisfy the separation condition, then the above algorithm perfectly localizes the spikes with high probability, even when the corruption amounts to near a constant fraction of the measurements.

### Demixing interfering sources

A scenario of increasing interest is when the observation is composed of a mixture of responses from multiple exciting or transmitting sources, and the goal is to simultaneously separate and localize the sources at a high resolution. For example, an electrode probing the activities in a brain records firing patterns of a few neighboring neurons, each with a distinct PSF. For pedagogical reasons, let us consider a generalization of the model (6) with two interfering sources, where the observation is given as

$$\boldsymbol{y}_{\text{mix}} = \operatorname{diag}(\boldsymbol{g}_1)\boldsymbol{x}_1 + \operatorname{diag}(\boldsymbol{g}_2)\boldsymbol{x}_2,$$

where $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$ correspond to the frequency-domain response of the PSFs, and $\boldsymbol{x}_i = \sum_{k=1}^{r_i} c_{i,k}\boldsymbol{a}(\tau_{i,k}), i = 1,2$. The goal is to separate and recover the spikes in both $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ from $\boldsymbol{y}_{\text{mix}}$, where $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$ are assumed to be known.

Using ANM, one seeks to recover both $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ simultaneously by minimizing the weighted sum of their atomic norms [67]:

$$\begin{aligned} \min_{\boldsymbol{x}_1,\boldsymbol{x}_2} \quad & \|\boldsymbol{x}_1\|_{\mathcal{A}} + \lambda \|\boldsymbol{x}_2\|_{\mathcal{A}} \\ \text{subject to} \quad & \boldsymbol{y}_{\text{mix}} = \operatorname{diag}(\boldsymbol{g}_1)\boldsymbol{x}_1 + \operatorname{diag}(\boldsymbol{g}_2)\boldsymbol{x}_2, \end{aligned}$$

where $\lambda > 0$ is some regularization parameter. Unlike the single-source case, the success of demixing critically depends on how easy it is to tell two PSFs apart; the more similar $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$ are, the harder it is to separate them. A random model can be used to generate dissimilar PSFs, namely it is assumed the entries of $g_i$s are i.i.d. generated from a uniform distribution over the complex circle [67]. The algorithm then succeeds with high probability as long as the sample size is sufficiently large and the spikes within the same signal satisfy the separation condition [67, Th. 2.1], without requiring any separation for spikes coming from different sources.

### Blind superresolution

So far, all algorithms have assumed the PSF as known, which is a reasonable assumption for problems where one can design and calibrate the PSF a priori. In general, one might need to estimate the PSF at the same time, possibly due to the fact that the PSF may drift and needs to be calibrated on the fly during deployment. In this case, we need to revisit (6) and estimate $\boldsymbol{g}$ and $\boldsymbol{x}$ simultaneously from their bilinear measurements,

$$\boldsymbol{y}_{\text{BR}} = \operatorname{diag}(\boldsymbol{g})\boldsymbol{x}.$$

This problem is terribly ill-posed, as the number of unknowns far exceeds the number of observations. One remedy is to exploit additional structures of $\boldsymbol{g}$. For example, if $\boldsymbol{g}$ lies in a known low-dimensional subspace $\boldsymbol{B} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n]^\top \in \mathbb{C}^{n\times d}$ with $d \ll n$, then the degrees of freedom of $\boldsymbol{g}$ is greatly dropped, since one only needs to estimate the coefficient $\boldsymbol{h} \in \mathbb{C}^d$ of $\boldsymbol{g} = \boldsymbol{B}\boldsymbol{h}$ in that subspace, which has a much smaller dimension. Even such, the measurement $\boldsymbol{y}_{\text{BR}}$ is still bilinear in $\boldsymbol{h}$ and $\boldsymbol{x}$, and one cannot directly apply ANM to $\boldsymbol{x}$ as it does not lead to a convex program.

Interestingly, a lifting trick can be applied [68], which rewrites $\boldsymbol{y}_{\text{BR}} = \mathcal{X}(\boldsymbol{Z})$ as linear measurements of a higher-dimensional object $\boldsymbol{Z} = \boldsymbol{x}\boldsymbol{h}^\top \in \mathbb{C}^{n\times d}$ similar to (23):

$$y_{\text{BR},i} = \boldsymbol{b}_i^\top \boldsymbol{h}\boldsymbol{e}_i^\top \boldsymbol{x} = \langle \boldsymbol{x}\boldsymbol{h}^\top, \boldsymbol{e}_i\boldsymbol{b}_i^{\mathsf{H}} \rangle, \quad i = 1, \ldots, n,$$

where $\boldsymbol{e}_i$ is the $i$th standard basis vector. Consequently, one can apply ANM to $\boldsymbol{Z}$ with respect to (24), leading to the algorithm [68]

$$\min_{\boldsymbol{Z}} \|\boldsymbol{Z}\|_{\mathcal{A}} \quad \text{subject to} \quad \boldsymbol{y}_{\text{BR}} = \mathcal{X}(\boldsymbol{Z}).$$

This approach succeeds with high probability; as soon as the sample size is sufficiently large, the spikes are well separated and the PSF satisfies certain incoherence properties [68, Th. 1]. Moreover, it can be further extended to demixing a mixture of sources with unknown PSFs, where each PSF lies in a distinct subspace [69].

## Beyond line spectrum estimation: Superresolution imaging for single-molecule fluorescence microscopy

When the atomic set is composed of a family of complex sinusoidal signals, an exact implementation of the atomic norm in SDP is available. In the most general setting, ANM is an infinite-dimensional convex program whose computation needs to be addressed carefully. Encouragingly, tailored solvers have been proposed and applied successfully to practical applications such as superresolution imaging for single-molecule fluorescence microscopy, which we now present as a case study to show its promise.

### Imaging principle

The development of superresolution fluorescence microscopy, which was awarded the 2014 Nobel Prize in Chemistry, is considered to fundamentally impact biological science and medicine. To date, a partial list of superresolution fluorescence microscopy technologies includes PALM [70], STORM [71], and fPALM [72], which share a similar imaging principle. A very

nice introduction can be found in [73]. While optical microscopy is desirable for imaging complex biological processes in live cells due to its noninvasive nature, due to diffraction limit, which is about hundreds of nanometers, it cannot image detailed internal structures of cells, which are often below 100 nm.

To deal with this challenge, biologists have come up with a clever idea of divide and conquer. To begin, imagine that every point within a cell is equipped with a photoswitchable fluorescent molecule, which means, once excited, the molecule will emit light stochastically over a duration of time to identify its location. This allows one to divide the imaging process into many frames, where in each frame a random and sparse subset of fluorescent molecules (point sources) are activated and localized at a resolution below the diffraction limit using imaging algorithms. The final image is thus obtained by superimposing the localization outcomes of all of the frames. Therefore, the high spatial resolution is achieved by sacrificing the temporal resolution. To speed up the imaging process and improve the temporal resolution, it is desirable to develop localization algorithms that are capable of identifying more fluorescent molecules per frame, which is known as the *emitter density*.

Very interestingly, this imaging principle can be used to reconstruct a 3D biological structure from 2D image frames [74]. One way is to introduce a cylindrical lens to modulate the ellipticity of the PSF based on the depth of the fluorescent object, which can be modeled as a Gaussian pulse with varying ellipticity along the $x$ and $y$ directions,

$$g(x, y \mid z) = \frac{1}{2\pi\sigma_x(z)\sigma_y(z)} e^{-\left[\frac{x^2}{2\sigma_x(z)^2} + \frac{y^2}{2\sigma_y(z)^2}\right]},$$

where $\sigma_x(z)$ and $\sigma_y(z)$ are functions of the depth in the $z$ direction, and can be calibrated in advance. For a 3D scene of point sources,

$$\zeta(x, y, z) = \sum_{i=1}^{r} c_i \delta(x - x_i, y - y_i, z - z_i),$$

its convolution with the PSF $g(x, y \mid z)$ is given as a 2D image in the form of

$$(\zeta * g)(x, y) = \sum_{i=1}^{r} \frac{c_i}{2\pi\sigma_x(z_i)\sigma_y(z_i)} e^{-\left[\frac{(x-x_i)^2}{2\sigma_x(z_i)^2} + \frac{(y-y_i)^2}{2\sigma_y(z_i)^2}\right]}.$$

Therefore, to perform superresolution, one needs to decode simultaneously the ellipticity as well as the location of the PSF, which is much more challenging. In practice, the situation is even more complex, since the continuous spatial function $(\zeta * g)(x, y)$ needs to be further discretized due to pixelization of the detector, leading to a discretized 2D image, $\boldsymbol{\omega} \in \mathbb{R}^{n_1 \times n_2}$, where each entry of $\boldsymbol{\omega}$ corresponds to the integration of $(\zeta * g)(x, y)$ over the area of a pixel. The final image, $\boldsymbol{z}$, counting the number of photons hitting the detector at each pixel, is modeled as a Poisson distribution with rate $\boldsymbol{\omega}$. The whole process is summarized in Figure 10.

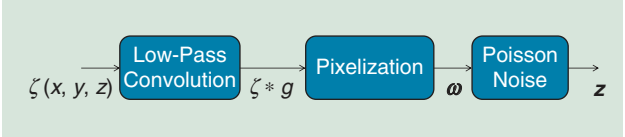


**FIGURE 10.** The mathematical model of 3D imaging in superresolution fluorescence microscopy.

## Applying ANM

Luckily, due to linearity, $\boldsymbol{\omega}$ can be viewed a sparse superposition of atoms that are parameterized by the 3D point sources,

$$\boldsymbol{\omega} = \sum_{i=1}^{r} c_i \boldsymbol{a}(x_i, y_i, z_i) = \int_{x,y,z} \boldsymbol{a}(x, y, z)\, d\zeta(x, y, z),$$

where each atom $\boldsymbol{a}(x, y, z)$ corresponds to the image of a point source at $(x, y, z)$ after convolution and pixelization. The atomic set is then given as $\mathcal{A}_{3D} = \{\boldsymbol{a}(x, y, z) : x, y, z \in \text{imaging range}\}$. The goal is thus to recover $\zeta(x, y, z)$, or the atomic decomposition of $\boldsymbol{\omega}$, from the observation $\boldsymbol{z} \sim \text{Poisson}(\boldsymbol{\omega})$ as accurately as possible.

A natural approach is to seek the sparsest $\boldsymbol{\omega}$ such that the likelihood function of the observation $\boldsymbol{z}$ is maximized. To that end, we consider a constrained maximum likelihood estimation, where we seek to solve $\boldsymbol{\omega}$ via

$$\min_{\boldsymbol{\omega}} -\log p(\boldsymbol{z} \mid \boldsymbol{\omega}) \quad \text{subject to} \quad \|\boldsymbol{\omega}\|_{\mathcal{A}} \le \eta, \tag{25}$$

where $p(\boldsymbol{z} \mid \boldsymbol{\omega})$ is the Poisson likelihood function, $\|\boldsymbol{\omega}\|_{\mathcal{A}}$ is the induced atomic norm with respect to $\mathcal{A}_{3D}$, and $\eta$ is some regularization parameter that may be tuned in practice.

Early efforts such as CSSTORM [75], which are based on $\ell_1$ minimization by directly discretizing the parameter space, suffer from high computational complexity, due to the need of storing and manipulating a large dictionary of atoms, as fine discretization is required along all three spatial dimensions. On the other end, recent algorithmic developments such as alternating descent conditional gradient (ADCG) [76] and CoGEnT [77] solve sparse inverse problems over continuous dictionaries with general convex loss functions at much reduced memory and computation requirements. In a nutshell, the ADCG method is an acceleration of conditional gradient, also known as *Frank–Wolfe,* to solve (25) with a general atomic set $\mathcal{A} = \{\boldsymbol{a}(\theta) : \theta \in \Theta\}$, where $\theta$ is a short-hand notation for the parameter space. In particular, it directly estimates the atomic decomposition of $\boldsymbol{\omega} = \int \boldsymbol{a}(\theta)\, d\zeta(\theta)$.

The standard Frank–Wolfe adds one new atom at every iteration to reduce the negative log-likelihood function; however, it will introduce many spurious atoms and lose sparsity as the iteration increases. To deal with this, ADCG introduces pruning and local refinements with a hope to maintain a sparse representation at all iterations. The detailed procedure of ADCG is given in Algorithm 2. In the $j$th iteration, denote the support and coefficient of the current estimate of $\zeta(\theta)$ as $\mathcal{T}_j$ and $\boldsymbol{c}_j$, and the current estimate of $\boldsymbol{\omega}$ as $\mathcal{A}(\mathcal{T}_j)\boldsymbol{c}_j$. Like Frank–Wolfe, ADCG starts by adding a spike to the estimated support $\mathcal{T}_j$ that maximally correlates with the derivative of $\log p(\boldsymbol{z} \mid \boldsymbol{\omega})$

with respect to $\boldsymbol{\omega}$ at the current estimate, $\mathcal{A}(\mathcal{T}_j)\boldsymbol{c}_j$. Because the spike location will be refined next, in practice, this step can be solved approximately by searching over a coarse grid of $\Theta$ to save computation.

ADCG then deviates from the standard Frank–Wolfe, and tries to improve the updated estimate by performing alternating descent over the coefficient and the support. It iterates between coefficients update via $\ell_1$ minimization, support pruning, and local refinement of the support by holding the coefficients fixed. The last step leverages the fact that $\boldsymbol{a}(\theta)$ is differentiable with respect to $\theta$, and a simple local search via gradient descent allows one to adjust the support to further reduce the loss function. Despite the nonconvexity in this refine step, [76] guarantees the convergence of ADCG with a convergence rate of $O(1/\epsilon)$ to reach $\epsilon$-accuracy in the function value, under some technical assumptions. In practice, the main computational benefits of ADCG are the absence of semidefinite constraints and small memory footprints, making it highly suitable for large-scale implementations.

TVSTORM [78] is a modification of ADCG tailored to solve (26) for 3D image reconstruction with some domain adaptations to speed up implementations. TVSTORM outperforms CSSTORM both in terms of computational time and reconstruction quality. Figure 11(a) shows the diffraction-limited imaging using conventional microscopy; in contrast, the 3D superresolution image reconstructed using TVSTORM in Figure 11(b) is much clearer, where the structure of 3D microtubules can be well resolved with the axial coordinate represented in different colors. Figure 11(c) and (d) compares the reconstruction quality of an enlarged region between TVSTORM and CSSTORM, where TVSTORM provides a visually more smooth reconstruction of the line structures in microtubules. Figure 12 shows that TVSTORM indeed has a higher detection rate and a lower false discovery rate than CSSTORM, while TVSTORM executes much faster.

## Final remarks

In this article, we presented an overview on how to leverage sparsity for continuous parameter estimation via the mathematical concept of atomic norms, which can be regarded as a generalization of the principle of $\ell_1$ norms for discrete model selection. We showcased its application in superresolution from low-pass observations in single-molecule fluorescence microscopy. The appeal of the atomic norm approach is attributed to its elegant mathematical framework, strong performance guarantees, and promises of scalable numerical implementations.

The atomic norm is only one of many possible approaches to exploit sparsity over the continuum. One competitive alternative is structured low-rank matrix optimization [79]–[82]; see [83] for its connections and comparisons with the atomic norm approach.

Another line of work [84]–[86] generalizes the traditional CS to an infinite-dimensional Hilbert space. In addition, sampling theorems are developed for signals with *a* finite rate of innovations together with strategies for perfect reconstruction [87]. More recently, a sparse functional framework has been proposed as a variational approach to handle sparsity over continuous and possibly nonlinear dictionaries. This category of estimators aims to recover functions with minimum support measure subject to the observation constraint [88], [89].

**As a topic still under development, open problems abound for both theoretical and practical performance of optimization-based superresolution in general.**

As a topic still under development, open problems abound for both theoretical and practical performance of optimization-based superresolution in general. We conclude by outlining some exciting future directions.

- *Tight performance analysis in noise*: Existing analyses of atomic norm denoising (21) typically only produce bounds that are tight up to some constant, making them less useful in practice. For example, attempts to benchmark the theoretical bounds against the CRB will be in vain due to the presence of large constants. It is therefore desirable to obtain tight performance bounds such as the one available for matrix denoising [90] that is asymptotically exact.
- *Adaptive selection of regularization parameters*: One benefit of ANM over traditional spectrum estimation approaches is that it can automatically select the model order. However, the choice of the regularization parameter (21) depends on the noise level, which may not be available in practice. How to optimally set the regularization parameters is another problem of great importance; see [91] for some recent developments.
- *Low-rank factorization for SDP formulations of atomic norms*: A popular heuristic to SDP with low-rank solutions is to apply low-rank matrix factorization and solve the corresponding nonconvex optimization problem [92], [93], with the premise of greatly reducing its computational cost. It will be interesting to see if this approach can be

**Algorithm 2. Alternating descent conditional gradient [76].**

**Input:** Observation $\boldsymbol{z}$; Parameter $\eta > 0$;
Initialize $\mathcal{T}_0$ to an empty set, $\boldsymbol{c}_0$ to $\boldsymbol{0}$, and $j = 0$;
**repeat** until stopping criteria
Localize the next spike:

$$\theta_{j+1} \in \operatorname*{argmax}_{\theta \in \Theta} \langle \boldsymbol{a}(\theta_j), \nabla_{\boldsymbol{\omega}} \log p(\boldsymbol{z} \mid \mathcal{A}(\mathcal{T}_j)\boldsymbol{c}_j) \rangle;$$

Update support: $\mathcal{T}_{j+1} \leftarrow \mathcal{T}_j \cup \{\theta_{j+1}\}$;
Refinement: **repeat**
1) Update the amplitudes:

$$\boldsymbol{c}_{j+1} \leftarrow \operatorname*{argmin}_{\|\boldsymbol{c}\|_1 \le \eta} -\log p(\boldsymbol{z} \mid \mathcal{A}(\mathcal{T}_{j+1})\boldsymbol{c});$$

2) Prune support: $\mathcal{T}_{j+1} \leftarrow \text{support}\ (\boldsymbol{c}_{j+1})$;
3) Local descent: improve $\mathcal{T}_{j+1}$ by performing local descent on $-\log p(\boldsymbol{z} \mid \mathcal{A}(\mathcal{T}_{j+1})\boldsymbol{c}_{j+1})$ holding the coefficient $\boldsymbol{c}_{j+1}$ fixed;
$j \leftarrow j + 1$;
**Output:** $(\mathcal{T}_j, \boldsymbol{c}_j)$ and $\boldsymbol{x}_j = \mathcal{A}(\mathcal{T}_j)\boldsymbol{c}_j$.

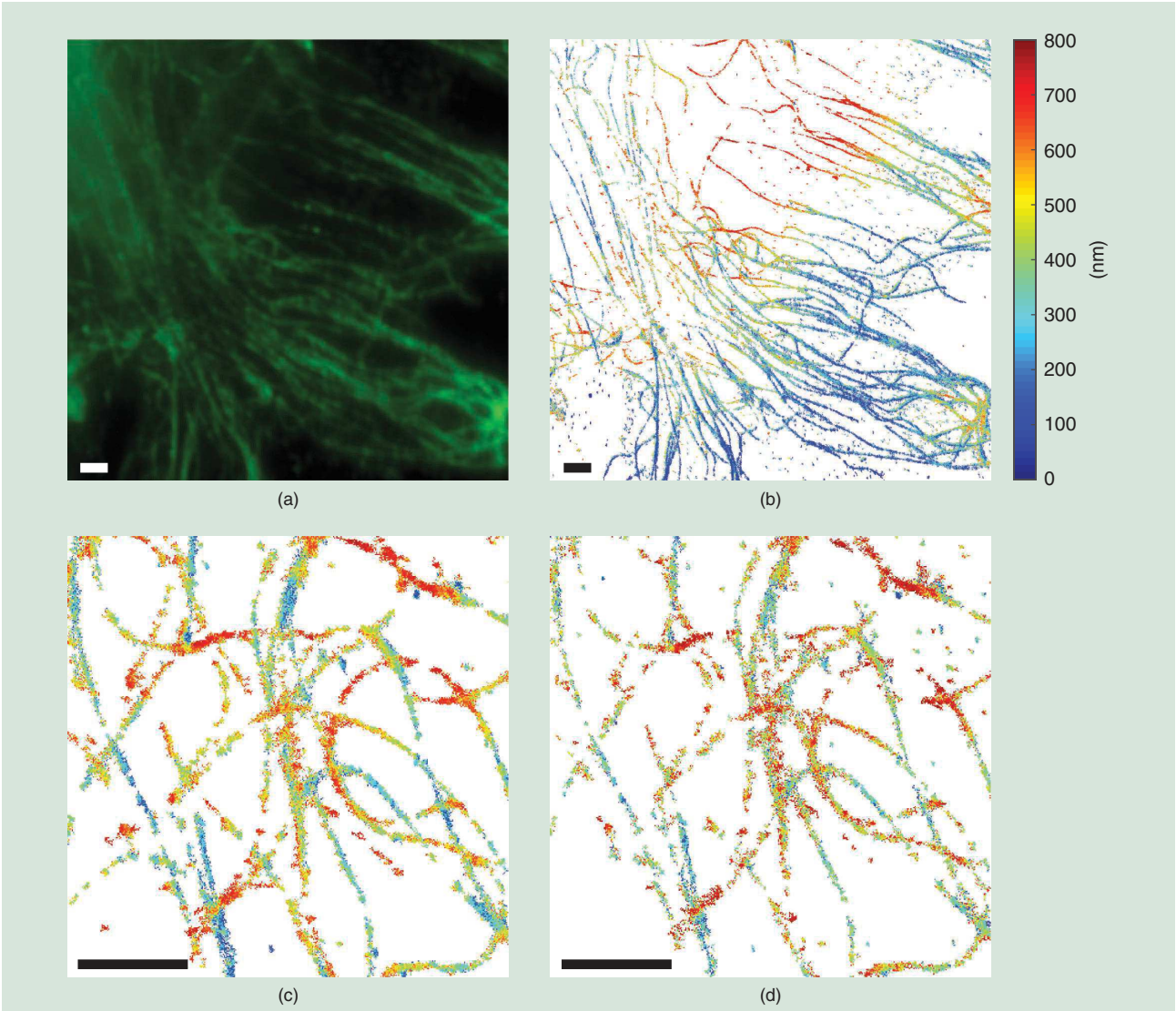


**FIGURE 11.** (a) Diffraction-limited imaging of microtubules using conventional microscopy and (b) superresolution 3D image reconstruction using TVSTORM. A comparison of reconstruction quality between (c) TVSTORM and (d) CSSTORM. (bar: 1.5 $\mu$m). (Source: [78].)

applied to speed up the computation of atomic norms with performance guarantees.

- *Bridging classical and optimization-based approaches*: There are deep connections between traditional (e.g. Prony, MUSIC, and so on) and optimization-based (e.g. ANM and nuclear norm minimization) approaches. Such connections have already been realized, for instance in the early works of Fuchs [94], where he provided an optimization interpretation of the Pisarenko method [95]. Another recent work [96] provided an optimization view to the MUSIC algorithm. It is hopeful that a confluence of past and current ideas will likely deepen our understandings and lead to further algorithmic improvements.
- *ANM for more general measurement models*: While there have been significant advances in the understanding of ANM for line spectrum estimation, its theory and application to other measurement settings require further investigation.
- *Applications in communications, sensing, and imaging*: ANM has recently emerged as a popular approach for many practical applications, such as channel estimation in massive MIMO [97], [98], radar imaging [99], and nuclear magnetic resonance spectroscopy [100]. It is our hope that this article will stir more interest in applying the atomic norm in applications that call for high-resolution parameter estimation.

## Authors

***Yuejie Chi*** (yuejiechi@cmu.edu) received her B.E. (Hon.) degree in electrical engineering from Tsinghua University, Beijing, China, in 2007 and her Ph.D. degree in electrical engineering from Princeton University, New Jersey, in 2012. She is currently an associate professor with the Department of

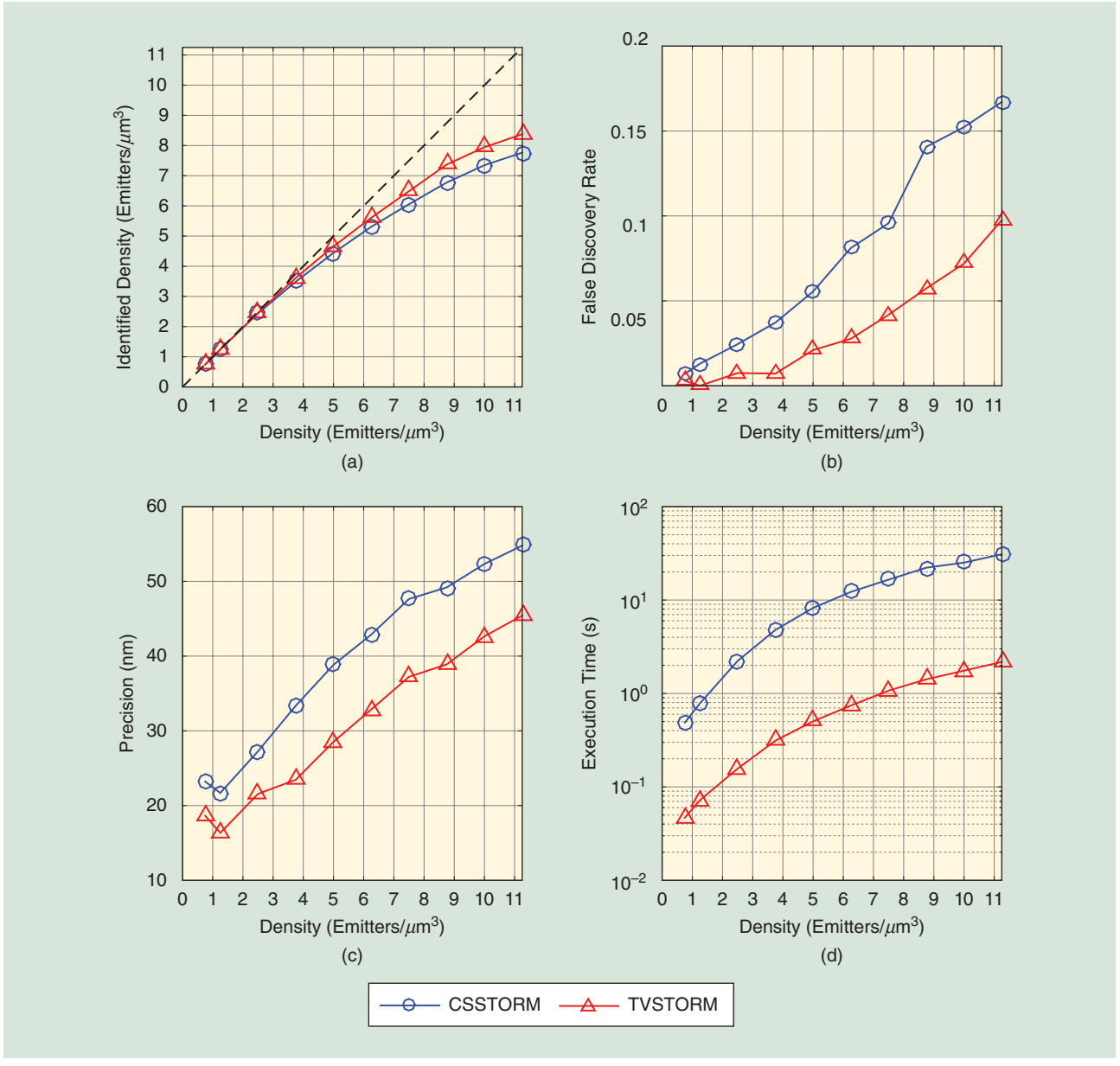


**FIGURE 12.** Comparisons between TVSTORM and CSSTORM for various performance metrics of 3D image reconstruction: (a) identified density, (b) false discovery rate, (c) precision, and (d) execution time with respect to the emitter density. (Source: [78].)

Electrical and Computer Engineering at Carnegie Mellon University, Pittsburgh, Pennsylvania, where she holds the Robert E. Doherty Early Career Development Professorship. She has received multiple awards, and she serves as an associate editor of *IEEE Transactions on Signal Processing*. Her research interests include statistical signal processing, machine learning, and large-scale optimization and their applications in data science, inverse problems, imaging, and sensing systems. She is a Senior Member of the IEEE.

***Maxime Ferreira Da Costa*** (mferreira@cmu.edu) received his M.Sc. degree in signal processing from Imperial College London, United Kingdom, Diplôme d'Ingénieur from Supélec, France, in 2012, and Ph.D. degree in electrical and electronic engineering from Imperial College London in 2018. He is currently a research associate with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania. His research focuses on mathematical signal processing, inverse problems, and optimization. In 2018, he was shortlisted among the finalists for the Jack Keil Wolf Student Paper Award at the IEEE International Symposium on Information Theory. He is a Member of the IEEE.

## Acknowledgments

We would like to thank *IEEE Signal Processing Magazine's* editorial board for early feedback on our white paper and the associate editor and anonymous reviewers for their constructive suggestions that helped to improve the quality of this article. This work is supported in part by the Office of Naval

Research under grants N00014-18-1-2142 and N00014-19-1-2404 and by the National Science Foundation under grants CIF-1826519 and ECCS-1818571.

## References

[1] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*, vol. 1. Englewood Cliffs, NJ: Prentice Hall, 1997.

[2] S. M. Kay and S. L. Marple, "Spectrum analysis: A modern perspective," *Proc. IEEE*, vol. 69, no. 11, pp. 1380–1419, 1981. doi: 10.1109/PROC.1981.12184.

[3] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, 1996. doi: 10.1109/79.526899.

[4] R. Prony, "Essai experimental et analytique, sur les lois de la dilatabilité des fluides de la vapeur de l'alkool, à différentes températures," *J. de l'Ecole Polytechnique Floréal et Plairial*, vol. 1, no. 2, pp. 24–76, 1795.

[5] R. Kumaresan, D. Tufts, and L. L. Scharf, "A Prony method for noisy data: Choosing the signal components and selecting the order in exponential signal models," *Proc. IEEE*, vol. 72, no. 2, pp. 230–233, 1984. doi: 10.1109/PROC.1984.12849.

[6] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, 1986. doi: 10.1109/TAP.1986.1143830.

[7] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, July 1989. doi: 10.1109/29.32276.

[8] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 5, pp. 814–824, May 1990. doi: 10.1109/29.56027.

[9] P. Stoica, R. L. Moses, B. Friedlander, and T. Soderstrom, "Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 3, pp. 378–392, 1989. doi: 10.1109/29.21705.

[10] M. P. Clark and L. L. Scharf, "Two-dimensional modal analysis based on maximum likelihood," *IEEE Trans. Signal Process.*, vol. 42, no. 6, pp. 1443–1452, 1994. doi: 10.1109/78.286959.

[11] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Commun. Pure Appl. Math.*, vol. 67, no. 6, pp. 906–956, 2014. doi: 10.1002/cpa.21455.

[12] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Found. Comput. Math.*, vol. 12, no. 6, pp. 805–849, 2012. doi: 10.1007/s10208-012-9135-7.

[13] S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001. doi: 10.1137/S003614450037906X.

[14] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010. doi: 10.1137/070697835.

[15] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7465–7490, 2013. doi: 10.1109/TIT.2013.2277451.

[16] B. Huang, M. Bates, and X. Zhuang, "Super-resolution fluorescence microscopy," *Annu. Rev. Biochem.*, vol. 78, no. 1, pp. 993–1016, 2009. doi: 10.1146/annurev.biochem.77.061906.092014.

[17] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, 2001. doi: 10.1109/18.959265.

[18] Y. Chen and Y. Chi, "Harnessing structures in big data via guaranteed low-rank matrix estimation: Recent theory and fast algorithms via convex and nonconvex optimization," *IEEE Signal Process. Mag.*, vol. 35, no. 4, pp. 14–31, 2018. doi: 10.1109/MSP.2018.2821706.

[19] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006. doi: 10.1002/cpa.20124.

[20] D. Donoho, "For most large underdetermined systems of linear equations the minimal $\ell_1$-norm solution is also the sparsest solution," *Commun. Pure Appl. Math.*, vol. 59, no. 6, pp. 797–829, 2006. doi: 10.1002/cpa.20132.

[21] C. Carathéodory, "Über den variabilitätsbereich der koeffizienten von potenzreihen, die gegebene werte nicht annehmen," *Math. Ann.*, vol. 64, no. 1, pp. 95–115, 1907. doi: 10.1007/BF01449883.

[22] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Berlin: Springer-Verlag, 2013.

[23] R. Sanyal, F. Sottile, and B. Sturmfels, "Orbitopes," *Mathematika*, vol. 57, no. 2, pp. 275–314, 2011. doi: 10.1112/S002557931100132X.

[24] J.-B. Lasserre, *Moments, Positive Polynomials and Their Applications*, vol. 1. Singapore: World Scientific, 2010.

[25] B. Dumitrescu, *Positive Trigonometric Polynomials and Signal Processing Applications*, vol. 103. Berlin: Springer-Verlag, 2007.

[26] T. T. Georgiou, "The Carathéodory–Fejér–Pisarenko decomposition and its multivariable counterpart," *IEEE Trans. Autom. Control*, vol. 52, no. 2, pp. 212–228, 2007. doi: 10.1109/TAC.2006.890479.

[27] M. Grant and S. Boyd. *CVX: Matlab software for disciplined convex programming*. (2018). [Online]. Available at http://cvxr.com/cvx

[28] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Math.*, vol. 346, nos. 9–10, pp. 589–592, 2008. doi: 10.1016/j.crma.2008.03.014.

[29] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007. doi: 10.1109/TIT.2007.909108.

[30] C. Fernandez-Granda, "Super-resolution of point sources via convex programming," *Inform. Inference*, vol. 5, no. 3, pp. 251–303, 2016. doi: 10.1093/imaiai/iaw005.

[31] M. F. Da Costa and W. Dai, "A tight converse to the spectral resolution limit via convex programming," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, pp. 901–905. doi: 10.1109/ISIT.2018.8437490.

[32] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5987–5999, 2013. doi: 10.1109/TSP.2013.2273443.

[33] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. Roy. Stat. Soc. B, Methodol.*, vol. 58, no. 1, pp. 267–288, 1996. doi: 10.1111/j.2517-6161.1996.tb02080.x.

[34] G. Tang, B. N. Bhaskar, and B. Recht, "Near minimax line spectral estimation," *IEEE Trans. Inf. Theory*, vol. 61, no. 1, pp. 499–512, 2015. doi: 10.1109/TIT.2014.2368122.

[35] C. Fernandez-Granda, "Support detection in super-resolution," in *Proc. 10th Int. Conf. Sampling Theory and Applications (SampTA 2013)*, 2013, pp. 145–148.

[36] Q. Li and G. Tang, "Approximate support recovery of atomic line spectral estimation: A tale of resolution and precision," *Appl. Comput. Harmon. Anal.*, to be published. doi: 10.1016/j.acha.2018.09.005. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S1063520318300824

[37] V. Duval and G. Peyré, "Exact support recovery for sparse spikes deconvolution," *Found. Comput. Math.*, vol. 15, no. 5, pp. 1315–1355, 2015. doi: 10.1007/s10208-014-9228-6.

[38] M. F. Da Costa and Y. Chi, On the stable resolution limit of total variation regularization for spike deconvolution. 2019. [Online]. Available: arXiv:1910.01629.

[39] T. Blu, P. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, "Sparse sampling of signal innovations," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 31–40, Mar. 2008. doi: 10.1109/MSP.2007.914998.

[40] B. D. Rao and K. S. Hari, "Performance analysis of root-MUSIC," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 1939–1949, 1989. doi: 10.1109/29.45540.

[41] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010. doi: 10.1561/2200000016.

[42] G. H. Golub and H. A. van der Vorst, "Eigenvalue computation in the 20th century," *J. Comput. Appl. Math.*, vol. 123, nos. 1–2, pp. 35–65, 2000. doi: 10.1016/S0377-0427(00)00413-1.

[43] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006. doi: 10.1109/TIT.2006.885507.

[44] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289– 1306, 2006. doi: 10.1109/TIT.2006.871582.

[45] G. Tang, B. N. Bhaskar, and B. Recht, "Sparse recovery over continuous dictionaries-just discretize," in *Proc. IEEE Asilomar Conf. Signals, Systems and Computers*, 2013, pp. 1043–1047. doi: 10.1109/ACSSC.2013.6810450.

[46] V. Duval and G. Peyré, "Sparse regularization on thin grids I: The LASSO," *Inverse Probl.*, vol. 33, no. 5, p. 055008, 2017. doi: 10.1088/1361-6420/aa5e12.

[47] V. Duval and G. Peyré, "Sparse spikes super-resolution on thin grids II: The continuous basis pursuit," *Inverse Probl.*, vol. 33, no. 9, p. 095008, 2017. doi: 10.1088/1361-6420/aa7fce.

[48] Y. Chi, L. Scharf, A. Pezeshki, and A. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2182–2195, May 2011. doi: 10.1109/TSP.2011.2112650.

[49] Z. Tan, P. Yang, and A. Nehorai, "Joint sparse recovery method for compressed sensing with structured dictionary mismatches," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 4997–5008, 2014. doi: 10.1109/TSP.2014.2343940.

[50] B. Mamandipoor, D. Ramasamy, and U. Madhow, "Newtonized orthogonal matching pursuit: Frequency estimation over the continuum," *IEEE Trans. Signal Process.*, vol. 64, no. 19, pp. 5066–5081, 2016. doi: 10.1109/TSP.2016.2580523.

[51] D. L. Donoho and J. Tanner, "Sparse nonnegative solution of underdetermined linear equations by linear programming," *Proc. Nat. Academy Sci. U. S. A.*, vol. 102, no. 27, pp. 9446–9451, 2005. doi: 10.1073/pnas.0502269102.

[52] J.-J. Fuchs, "Sparsity and uniqueness for some specific under-determined linear systems," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, vol. 5. Philadelphia: IEEE, 2005, pp. v/729–v/732. doi: 10.1109/ICASSP.2005.1416407.

[53] Q. Denoyelle, V. Duval, and G. Peyré, "Support recovery for sparse super-resolution of positive measures," *J. Fourier Anal. Appl.*, vol. 23, no. 5, pp. 1153–1194, 2017. doi: 10.1007/s00041-016-9502-x.

[54] G. Schiebinger, E. Robeva, and B. Recht, "Superresolution without separation," *Inform. Inference*, vol. 7, no. 1, pp. 1–30, 2017. doi: 10.1093/imaiai/iax006.

[55] A. Eftekhari, J. Tanner, A. Thompson, B. Toader, and H. Tyagi, "Sparse non-negative super-resolution–simplified and stabilised," *Appl. Comput. Harmon. Anal.*, to be published. doi: 10.1016/j.acha.2019.08.004.

[56] V. I. Morgenshtern and E. J. Candes, "Super-resolution of positive sources: The discrete setup," *SIAM J. Imag. Sci.*, vol. 9, no. 1, pp. 412–444, 2016. doi: 10.1137/15M1016552.

[57] Y. Chi and Y. Chen, "Compressive two-dimensional harmonic retrieval via atomic norm minimization," *IEEE Trans. Signal Process.*, vol. 63, no. 4, pp. 1030–1042, 2015. doi: 10.1109/TSP.2014.2386283.

[58] W. Xu, J.-F. Cai, K. V. Mishra, M. Cho, and A. Kruger, "Precise semidefinite programming formulation of atomic norm minimization for recovering d-dimensional ($d \geq 2$) off-the-grid frequencies," in *Proc. IEEE Information Theory and Applications Workshop (ITA)*, 2014, pp. 1–4. doi: 10.1109/ITA.2014.6804267.

[59] Y. Li and Y. Chi, "Off-the-grid line spectrum denoising and estimation with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 64, no. 5, pp. 1257–1269, doi: 10.1109/TSP.2015.2496294.

[60] Z. Tian and G. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, (ICASSP 2007)*, vol. 4. Honolulu, HI: IEEE, 2007, pp. 1357–1360. doi: 10.1109/ICASSP.2007.367330.

[61] E. J. Candès and Y. Plan, "A probabilistic and RIPless theory of compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7235–7254, 2011. doi: 10.1109/TIT.2011.2161794.

[62] R. Heckel and M. Soltanolkotabi, "Generalized line spectral estimation via convex optimization," *IEEE Trans. Inf. Theory*, vol. 64, no. 6, pp. 4001–4023, 2018. doi: 10.1109/TIT.2017.2757003.

[63] H. Fu and Y. Chi, "Quantized spectral compressed sensing: Cramer–Rao bounds and recovery algorithms," *IEEE Trans. Signal Process.*, vol. 66, no. 12, pp. 3268–3279, 2018. doi: 10.1109/TSP.2018.2827326.

[64] D. L. Donoho and P. B. Stark, "Uncertainty principles and signal recovery," *SIAM J. Appl. Math.*, vol. 49, no. 3, pp. 906–931, 1989. doi: 10.1137/0149053.

[65] X. Li, "Compressed sensing and matrix completion with constant proportion of corruptions," *Constr. Approx.*, vol. 37, no. 1, pp. 73–99, 2013. doi: 10.1007/s00365-012-9176-9.

[66] C. Fernandez-Granda, G. Tang, X. Wang, and L. Zheng, "Demixing sines and spikes: Robust spectral super-resolution in the presence of outliers," *Inform. Inference*, vol. 7, no. 1, pp. 105–168, 2017. doi: 10.1093/imaiai/iax005.

[67] Y. Li and Y. Chi, "Stable separation and super-resolution of mixture models," *Appl. Comput. Harmon. Anal.*, vol. 46, no. 1, pp. 1–39, 2019. doi: 10.1016/j.acha.2017.03.003.

[68] Y. Chi, "Guaranteed blind sparse spikes deconvolution via lifting and convex optimization," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 4, pp. 782–794, 2016. doi: 10.1109/JSTSP.2016.2543462.

[69] D. Yang, G. Tang, and M. B. Wakin, "Super-resolution of complex exponentials from modulations with unknown waveforms," *IEEE Trans. Inf. Theory*, vol. 62, no. 10, pp. 5809–5830, 2016. doi: 10.1109/TIT.2016.2586083.

[70] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz et al., "Imaging intracellular fluorescent proteins at nanometer resolution," *Science*, vol. 313, no. 5793, pp. 1642–1645, 2006. doi: 10.1126/science.1127344.

[71] M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," *Nat. Meth.*, vol. 3, no. 10, pp. 793–796, 2006. doi: 10.1038/nmeth929.

[72] S. T. Hess, T. P. Girirajan, and M. D. Mason, "Ultra-high resolution imaging by fluorescence photoactivation localization microscopy," *Biophys. J.*, vol. 91, no. 11, pp. 4258–4272, 2006. doi: 10.1529/biophysj.106.091116.

[73] M. Ehrenberg, "Scientific background on the Nobel Prize in Chemistry 2014," The Nobel Prize, 2014. [Online]. Available: https://www.nobelprize.org/uploads/2018/06/advanced-chemistryprize2014-1.pdf or https://www.nobelprize.org/prizes/chemistry/2014/advanced-information/

[74] B. Huang, W. Wang, M. Bates, and X. Zhuang, "Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy," *Science*, vol. 319, no. 5864, pp. 810–813, 2008. doi: 10.1126/science.1153529.

[75] L. Zhu, W. Zhang, D. Elnatan, and B. Huang, "Faster STORM using compressed sensing," *Nat. Meth.*, vol. 9, no. 7, pp. 721–723, 2012. doi: 10.1038/nmeth.1978.

[76] N. Boyd, G. Schiebinger, and B. Recht, "The alternating descent conditional gradient method for sparse inverse problems," *SIAM J. Optim.*, vol. 27, no. 2, pp. 616–639, 2017. doi: 10.1137/15M1035793.

[77] N. Rao, P. Shah, and S. Wright, "Forward–backward greedy algorithms for atomic norm regularization," *IEEE Trans. Signal Process.*, vol. 63, no. 21, pp. 5798–5811, 2015. doi: 10.1109/TSP.2015.2461515.

[78] J. Huang, M. Sun, J. Ma, and Y. Chi, "Super-resolution image reconstruction for high-density three-dimensional single-molecule microscopy," *IEEE Trans. Comput. Imag.*, vol. 3, no. 4, pp. 763–773, 2017. doi: 10.1109/TCI.2017.2699425.

[79] Y. Chen and Y. Chi, "Robust spectral compressed sensing via structured matrix completion," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6576–6601, Oct. 2014. doi: 10.1109/TIT.2014.2343623.

[80] K. H. Jin, D. Lee, and J. C. Ye, "A general framework for compressed sensing and parallel MRI using annihilating filter based low-rank Hankel matrix," *IEEE Trans. Comput. Imag.*, vol. 2, no. 4, pp. 480–495, 2016. doi: 10.1109/TCI.2016.2601296.

[81] G. Ongie, S. Biswas, and M. Jacob, "Convex recovery of continuous domain piecewise constant images from nonuniform Fourier samples," *IEEE Trans. Signal Process.*, vol. 66, no. 1, pp. 236–250, doi: 10.1109/TSP.2017.2750111.

[82] J.-F. Cai, X. Qu, W. Xu, and G.-B. Ye, "Robust recovery of complex exponential signals from random Gaussian projections via low rank Hankel matrix reconstruction," *Appl. Comput. Harmon. Anal.*, vol. 41, no. 2, pp. 470–490, 2016. doi: 10.1016/j.acha.2016.02.003.

[83] Y. Chi, "Convex relaxations of spectral sparsity for robust super-resolution and line spectrum estimation," in *Proc. Wavelets and Sparsity XVII*, vol. 10394. International Society for Optics and Photonics, 2017, p. 103941G.

[84] G. Puy, M. E. Davies, and R. Gribonval, "Recipes for stable linear embeddings from Hilbert spaces to Rm," *IEEE Trans. Inf. Theory*, vol. 63, no. 4, pp. 2171–2187, 2017. doi: 10.1109/TIT.2017.2664858.

[85] B. Adcock and A. C. Hansen, "Generalized sampling and infinite-dimensional compressed sensing," *Found. Comput. Math.*, vol. 16, no. 5, pp. 1263–1323, 2016. doi: 10.1007/s10208-015-9276-6.

[86] B. Adcock, A. C. Hansen, C. Poon, and B. Roman, "Breaking the coherence barrier: A new theory for compressed sensing," in *Forum of Mathematics, Sigma*, vol. 5, no. e4, 84 pp., 2017. doi: 10.1017/fms.2016.32.

[87] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1417–1428, 2002. doi: 10.1109/TSP.2002.1003065.

[88] L. F. Chamon, Y. C. Eldar, and A. Ribeiro, "Strong duality of sparse functional optimization," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 4739–4743. doi: 10.1109/ICASSP.2018.8461311.

[89] L. F. Chamon, Y. C. Eldar, and A. Ribeiro, Functional nonlinear sparse models. 2018. [Online]. Available: arXiv:1811.00577.

[90] D. Donoho and M. Gavish, "Minimax risk of matrix denoising by singular value thresholding," *Ann. Statist.*, vol. 42, no. 6, pp. 2413–2440, 2014. doi: 10.1214/14-AOS1257.

[91] C. Boyer, Y. De Castro, and J. Salmon, "Adapting to unknown noise level in sparse deconvolution," *Inform. Inference*, vol. 6, no. 3, pp. 310–348, 2017. doi: 10.1093/imai/iaw024.

[92] S. Burer and R. D. Monteiro, "A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization," *Math. Program.*, vol. 95, no. 2, pp. 329–357, 2003. doi: 10.1007/s10107-002-0352-8.

[93] Y. Chi, Y. M. Lu, and Y. Chen, "Nonconvex optimization meets low-rank matrix factorization: An overview," *IEEE Trans. Signal Process.*, vol. 67, no. 20, pp. 5239–5269, 2019. doi: 10.1109/TSP.2019.2937282.

[94] J.-J. Fuchs, "Extension of the Pisarenko method to sparse linear arrays," *IEEE Trans. Signal Process.*, vol. 45, no. 10, pp. 2413–2421, 1997. doi: 10.1109/78.640707.

[95] V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophysical J. Int.*, vol. 33, no. 3, pp. 347–366, 1973. doi: 10.1111/j.1365-246X.1973.tb03424.x.

[96] S. Li, H. Mansour, and M. B. Wakin, An optimization view of MUSIC and its extension to missing data. 2018. Available: arXiv:1806.03511.

[97] J. Deng, O. Tirkkonen, and C. Studer, "MmWave channel estimation via atomic norm minimization for multi-user hybrid precoding," in *Proc. IEEE Wireless Communications and Networking Conf. (WCNC)*, 2018, pp. 1–6. doi: 10.1109/WCNC.2018.8377093.

[98] Y. Tsai, L. Zheng, and X. Wang, "Millimeter-wave beamformed full-dimensional MIMO channel estimation based on atomic norm minimization," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6150–6163, 2018. doi: 10.1109/TCOMM.2018.2864737.

[99] Z. Zhu, G. Tang, P. Setlur, S. Gogineni, M. B. Wakin, and M. Rangaswamy, "Super-resolution in SAR imaging: Analysis with the atomic norm," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 2016, pp. 1–5. doi: 10.1109/SAM.2016.7569639.

[100] J. Ying, H. Lu, Q. Wei, J.-F. Cai, D. Guo, J. Wu, Z. Chen, and X. Qu, "Hankel matrix nuclear norm regularized tensor completion for n-dimensional exponential signals," *IEEE Trans. Signal Process.*, vol. 65, no. 14, pp. 3702–3717, 2017. doi: 10.1109/TSP.2017.2695566.

SP