Resolution Limits for Atomic Decompositions via Markov-Bernstein type Inequalities

Gongguo Tang
Colorado School of Mines
Email: gtang@mines.edu

Abstract—**Atomic norm offers a universal way to regularize ill-posed inverse problems. A fundamental problem in understanding the power of such regularization is determining norm-achieving decompositions. Sufficient conditions for atomic decomposition have been established for many instances of atomic norms. In this work, by using Markov-Bernstein type inequalities, we show that the resolution conditions that appear in these sufficient conditions are almost necessary.**

I. INTRODUCTION

In signal processing, machine learning, and statistics, regularization plays an important role in constructing well-posed inverse problems and producing learning algorithms with good generalization properties [1], [2], [3]. Two most well-studied regularization approaches are based on ℓ_2 and ℓ_1 norms. Despite the increased computational complexity compared to ℓ_2-type regularization, ℓ_1-type regularization has become more popular in the past decade, partly due to its ability to produce sparse models, enabling feature selection and stable recovery with minimal data. Given a set of simple models or atoms, the atomic norm is an abstraction of ℓ_1-type regularization that favors models composed of fewer atoms, generalizing the ℓ_1 norm for finite dictionaries [4] and the nuclear norm for low-rank matrices [5]. Using the notion of descent cones, the authors of [6] argued that the atomic norm is the best possible convex proxy for recovering simple models. In this paper, we aim to further our understanding of atomic norms by studying their ability to resolve the composing atoms of a model.

The problem of atomic decomposition is to determine when a decomposition with respect to a given atomic set achieves the signal's atomic norm. This basic problem is fundamental in understanding the regularization and estimation power of atomic norms. For one thing, it tells us what types of features are selected given full, noise-free data. For another, the dual function constructed to certify a particular decomposition is useful in demonstrating the performance of atomic norm minimization for data corrupted by missing observations, noise, and outliers [7]. For finite atomic sets, it is now well-known that if the atoms satisfy certain incoherence conditions such as the restricted isometry property, then a sparse decomposition achieves the atomic norm [8]. When the atoms are complex sinusoids parameterized by the frequency, Candès and Fernandez-Granda showed that atomic decomposition is solved by atoms with well-separated frequencies [9]. Similar separation conditions also show up when the atoms are rank-one, unit norm tensors [10], translations of a known waveform [11], [12], spherical harmonics [13], and radar signals parameterized by translations and modulations [14].

These sufficient conditions for atomic decompositions suggest that the atomic norm, or ℓ_1-type regularization might have a limitation in resolving atoms: only models/signals composed of sufficiently *different* atoms can be recovered by the atomic norm. The measure of "difference" can be incoherence, orthogonality, or separation in parameters. We can understand such separation conditions by looking at the primal problem of atomic decomposition, which is to minimize the ℓ_1 norm of the coefficients subject to any valid decomposition. When measured by the ℓ_1 norm, including highly correlated atoms in the decomposition is not economical. However,

it is more enlightening to examine the dual problem. Duality theory states that if a decomposition achieves the atomic norm, then there exists a dual function that is uniformly bounded by 1 and interpolates the signs of the decomposition coefficients at the composing atoms. Now suppose two composing atoms with close parameters have coefficients with opposite signs. Then the uniformly bounded dual interpolating function should go from $+1$ to -1 within a very short distance in the parameter space, implying a large derivative of the dual function due to the mean value theorem. This becomes impossible if the dual function's derivative is bounded.

Given a class of functions, Markov-Bernstein type inequalities bound the magnitude of the derivative in terms of the uniform bound on the function value. The classical Bernstein's inequality for trigonometric polynomials of degree at most n states that $|p_n'(\theta)| \leq n \sup_{\theta \in [0, 2\pi)} |p_n(\theta)|$ [15]. When the set of atoms are n equi-spaced samples of complex sinusoids, the dual function is a trigonometric polynomial of degree at most n. We immediately conclude that if two composing atoms of a signal have coefficients with opposite signs, then their frequency parameters should be separated at least by $2/n$. The classical Bernstein's inequality can also be used to establish such resolution limits for atoms paramterized by high-dimensional vectors. For example, consider the atomic decomposition of 3rd order symmetric tensors, whose atoms are parameterized by vectors living on a high-dimensional sphere. For any two composing rank-one tensors $x_i \otimes x_i \otimes x_i$ and $x_j \otimes x_j \otimes x_j$ with opposite coefficient signs, the interpolating dual function becomes a trigonometric polynomial of degree 3 when restricted onto the great circle passing through the two vectors x_i and x_j. Therefore, the angle of these two vectors should be greater than $2/3$.

It is important to point out that the sign pattern of the decomposition's coefficients plays an important role in determine the resolution limit. For a large class of atomic sets, particularly those forming a Chebyshev system , if the signs of a decomposition are all positive, then the decomposition achieves the minimal ℓ_1 norm without any separation requirement. However, in many practical cases, it is necessary to require that a set of atoms are able to achieve the atomic norm regardless of the sign patterns of the coefficients. Therefore, the parameters associated with all the atoms should be well-separated.

The paper is organized as follows. In Section II, we introduce the notions of atoms, atomic norms, and the problem of atomic decomposition. Section IV is devoted to combining duality theory and Markov-Bernstein type inequalities to derive the main resolution limit result. Section V concludes the paper and discusses open problems.

II. ATOMS AND ATOMIC DECOMPOSITIONS

A. Atoms and atomic norms

Consider an atomic set $\mathcal{A} = \{a(\theta) : \theta \in \Theta\}$ whose atoms $a(\theta) \in \mathbb{F}^n$ are parameterized by $\theta \in \Theta \subset \mathbb{R}^d$. Here $\mathbb{F}$ is either the set of real numbers $\mathbb{R}$, or the set of complex numbers $\mathbb{C}$. Throughout the paper, we assume that the components $\{a_i(\theta)\}_{i=1}^n$ of an atom are continuously differentiable functions of θ. For any $z \in \mathbb{F}^n$, its

atomic norm with respect to the atomic set $\mathcal{A}$ is defined as

$$\|z\|_{\mathcal{A}} = \inf \left\{ \sum_j |c_j| : z = \sum_j c_j a(\theta_j), \theta_j \in \Theta \right\}. \quad (1)$$

Depending on the application, the coefficient c_j might be in $\mathbb{R}$ or $\mathbb{C}$.

The atomic norm is the smallest ℓ_1 norm of the coefficients among all possible representations of a signal using atoms in the set $\mathcal{A}$. When Θ is a discrete set with finite cardinality, the definition of the atomic norm reduces to the Basis Pursuit linear program [16]. In this work, we focus on continuously parameterized atomic sets.

Example 1 (Atoms and the associated atomic norms)**.**

1) 1D line spectral estimation and super-resolution [9], [17], [18], [7]. The atoms are trigonometric functions

$$a(\theta) = \begin{bmatrix} 1 & e^{i\theta} & \cdots & e^{in\theta} \end{bmatrix}^T \in \mathbb{C}^{n+1} \quad (2)$$

for $\theta \in [0, 2\pi)$. The corresponding atomic norm can be used to enforce sparsity in the continuous frequency domain. We will see later that the atom (2) can be used to model the problem of resolving point targets from low frequency measurements [9].

2) High-dimensional line spectral estimation [9], [19], [20]. The atom is the direct product of the 1D line spectral atoms:

$$a_{k_1,k_2,\cdots,k_d}(\boldsymbol{\theta}) = e^{i(k_1\theta_1 + k_2\theta_2 + \cdots + k_d\theta_d)}, \quad (3)$$

where the integer observation index $0 \le k_i \le n_i$ and the multi-frequency parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d) \in [0, 2\pi)^d$.

3) Spherical harmonics [21]. This is yet another high-dimensional extension of the 1D line spectral estimation. The atom is parameterized by $\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\|_2 = 1\}$, the unit sphere in $\mathbb{R}^d$, with component given by

$$a_{m,\ell}(x) = Y_m^\ell(x), \ell = 1, 2, \ldots, N_{m,d}. \quad (4)$$

Here $Y_m^\ell(x)$ denotes the ℓth spherical harmonic polynomial of degree m in d variables, and $N_{m,d}$ is the number of such polynomials. When $d = 2$, the spherical harmonic atom reduces to 1D line spectral atom.

4) Matrix factorization [22], [5]. The atomic set

$$\mathcal{A} = \{uv^T : u \in \mathbb{S}^{p-1}, v \in \mathbb{S}^{q-1}\} \quad (5)$$

is the set of rank-1 matrices with unit Frobenius norm. The atomic norm is matrix nuclear norm, which can be computed by solving a semidefinite program.

5) Tensor decomposition [10]. For a mth order symmetric tensor, the atomic set is

$$\mathcal{A} = \{x^{\otimes m} := \underbrace{x \otimes x \otimes \cdots \otimes x}_{m \text{ times}} : x \in \mathbb{S}^{d-1}\} \quad (6)$$

while for a mth order non-symmetric tensor, the atomic set is

$$\mathcal{A} = \{u_1 \otimes u_2 \otimes \cdots \otimes u_m : u_i \in \mathbb{S}^{d_i-1}, i = 1, \ldots, m\}. \quad (7)$$

The associated atomic norms are the symmetric tensor nuclear norm and the tensor nuclear norm.

6) Mixtures of translation-invariant signals [11], [12]. The atoms

$$a(\tau) = [\psi(\frac{t-\tau}{\sigma})]_{t \in T} \quad (8)$$

for $\tau \in [a, b]$, where $\psi(\cdot)$ is a known waveform such as the Gaussian function, sinc function, or Airy function, and T is the set of observation points which could be finite, an interval, or the entire $\mathbb{R}$.

7) Radar signals [14]. The atom is a known waveform $\psi(\cdot)$ translated by τ and modulated by ν:

$$a(\tau, \nu) = [\psi(t-\tau)e^{i2\pi\nu t}]_{t \in T} \quad (9)$$

where T again is the set of observation points.

8) Single-pole linear systems [23]. The atom is the transfer function of a single pole system parameterized by its pole location in $\mathbb{D}_\rho$, the disk on $\mathbb{C}$ of radius $\rho < 1$:

$$a(w) = [\frac{1 - |w|^2}{z - w}]_{z \in \mathcal{Z}}, w \in \mathbb{D}_\rho \quad (10)$$

where $\mathcal{Z}$ is the set of observation points, which could be samples of the unit circle on $\mathbb{C}$.

B. Atomic decomposition and super-resolution

Given an atomic set $\mathcal{A} \subset \mathbb{F}^n$ and any $z \in \mathcal{A}$, the computation of the atomic norm $\|z\|_{\mathcal{A}}$ is equivalent to finding a particular decomposition $z = \sum_j c_j a(\theta_j)$ such that $\|z\|_{\mathcal{A}} = \sum_j |c_j|$. We formalize this in the following definition:

Definition 1. Given an atomic set $\mathcal{A} = \{a(\theta) : \theta \in \Theta\} \subset \mathbb{F}^n$, a finite decomposition $\sum_j c_j a(\theta_j)$ is called an *atomic decomposition* if $\|\sum_j c_j a(\theta_j)\|_{\mathcal{A}} = \sum_j |c_j|$. When all c_js are positive, we also call $\sum_j c_j a(\theta_j)$ an atomic decomposition if it is unique.

Carathéodory' convex hull theorem [24] implies that there exists a finite decomposition that involves at most $n + 1$ atoms. Typically one is interested in finding the decomposition with smallest number of atoms. The definition of the atomic norm does promote sparsity as we can rewrite it as an ℓ_1 minimization problem over $\mathcal{M}(\Theta)$, the set of all signed or complex measures on Θ. That is, the atomic norm $\|z\|_{\mathcal{A}}$ is equal to the optimal value of

$$\underset{\mu \in \mathcal{M}(\Theta)}{\text{minimize}} \|\mu\|_{\text{TV}} \text{ subject to } z = \int_\Theta a(\theta)d\mu \quad (11)$$

Here $\|\cdot\|_{\text{TV}}$ denotes the total variation norm of a signed or complex measure. When the set Θ is finite, the total variation norm reduces to the usual ℓ_1 norm.

The existence of atomic decompositions with at most $n+1$ atoms is equivalent to the existence of a discrete measure $\mu^\star = \sum_j c_j \delta(\theta - \theta_j)$ supported on at most $n + 1$ points in Θ such that $z = \int_\Theta a(\theta)d\mu^\star$. Therefore, the infinite-dimensional optimization formulation (11) of the atomic norm suggests another explanation of atomic decomposition: recovering a discrete measure $\mu^\star$ from its linear measurements given by $z_j = \int_\Theta a_j(\theta)d\mu^\star, j = 1, \ldots, n$. The literature on finite ℓ_1 minimization [4] suggests that this task might be successful if the unknown measure has a small support.

Besides sparsity, several sufficient conditions established in the literature for some of the atoms in Example 1 state that if the support $S(\mu^\star) = \{\theta_j\}$ of the true, unknown measure $\mu^\star$ are well-separated, then $\mu^\star$ is the unique solution to (11) when $z = \int_\theta a(\theta)d\mu^\star$. These sufficient conditions are usually cited as evidence to support the ability of infinite-dimensional ℓ_1 minimization in achieving super-resolution. The concept of super-resolution is particularly clear for the atoms defined in Example 1 1) and 2), where the goal of atomic decomposition is to estimate point targets from low-frequency Fourier measurements. The Rayleigh criterion is often used to argue that the sufficient condition is almost necessary. In this work, we establish a rigorous resolution limit for some common atoms as a consequence of the ℓ_1 nature of the atomic norm.

The separation condition is also reminiscence of the incoherence condition for finite ℓ_1 minimization [4]. Indeed, the separation of the

support $S(\mu^\star)$ ensures that the atoms $\{a(\theta_j)\}$ are uncorrelated.[1]

III. SUFFICIENT CONDITIONS FOR ATOMIC DECOMPOSITIONS

In this section, we collect sufficient conditions for atomic decomposition for some of the atoms listed in Example 1.

A. Sufficient conditions for signed atomic decompositions

For arbitrary combinations of atoms, we have the following theorem, whose proofs are given in [9], [21], [22], [5], [10], [11], [12].

Theorem 1. *In the following, if the corresponding separation condition is satisfied for the given set of atoms, then the measure $\mu^\star$ is the unique solution to the optimization* (11)*, and the corresponding decomposition is the unique atomic decomposition.*

1) *The support of a line spectral measure $\mu^\star = \sum_{j=1}^k c_j \delta(\theta - \theta_j)$ with c_j complex is separated by $\frac{8\pi}{n}$, that is,*

$$\min_{\theta_i \neq \theta_j \in S(\mu^\star)} |\theta_i - \theta_j| > \frac{8\pi}{n}. \tag{12}$$

When c_js are real, the separation can be lowered to $\frac{7.48\pi}{n}$.

2) *The conclusions of part 1) generalize to high-dimensional line spectral atoms defined in (3) with $n_i = n$ if the support of $\mu^\star = \sum_{j=1}^k c_j \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_j)$ satisfies the separation condition*

$$\min_{\boldsymbol{\theta}_i \neq \boldsymbol{\theta}_j} \|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\|_\infty > \frac{c_d}{n} \tag{13}$$

for some constant c_d that depends on the dimension d only.

3) *For spherical harmonic atoms on $\mathbb{S}^2$, separation condition imposed on the support of $\mu^\star = \sum_{j=1}^k c_j \delta(x - x_j)$ is*

$$\min_{x_i, x_j \in \mathbb{S}^2, x_i \neq x_j} \arccos(\langle x_i, x_j \rangle) \geq \frac{c}{m} \tag{14}$$

where c is a numerical constant independent of m.

4) *The matrix decomposition $M = \sum_{j=1}^r c_j u_j v_j^T$ is orthogonal, that is, $\langle u_i, u_j \rangle = 0$ and $\langle v_i, v_j \rangle = 0$ for $i \neq j$. The uniqueness is up to the sign ambiguity.*

5) *The 3rd order symmetric tensor decomposition $T = \sum_{j=1}^r x_j^{\otimes 3}$ is near orthogonal, that is, $\|X^T X - I_r\| \leq 0.0016$ where $X = [x_1, \cdots, x_r]$.*

6) *For mixtures of translation-invariant signals, the observation set $T = \mathbb{R}$ and the waveform $\psi(t)$ satisfies certain technical conditions explained in [11], [12]. The translation parameters of the decomposition $x(t) = \sum_j c_j \psi(\frac{t - \tau_j}{\sigma})$ should be separated by $c\sigma$ for some numerical constant c.*

B. Sufficient conditions for positive atomic combinations

We remark that Theorem 1 holds regardless of the sign patterns of the coefficients $\{c_j\}$. Prior information of the sign pattern can relax the separation requirement, as already illustrated by the case of 1D line spectral atoms, where the critical constant was lowered from 8π to 7.48π when the coefficients are known to be real, *i.e.*, the signs are $\pm$ instead of generic complex phases. In fact, for a large class of atoms, when all the coefficients are known to be positive, no separation is needed for $\mu^\star = \sum_j c_j \delta(\theta - \theta_j)$ with $c_j > 0$ to solve the feasibility problem

$$\text{find } \mu \in \mathcal{M}_+(\Theta) \text{ subject to } z = \int_\Theta a(\theta) d\mu \tag{15}$$

[1]However, a global incoherence condition for all atoms in $\mathcal{A}$ are no longer possible due to the continuity of $a(\theta)$ [7].

where $\mathcal{M}_+(\Theta)$ is the set of (non-negative) measures on Θ, and $z = \sum_j c_j a(\theta_j) = \int_\Theta a(\theta) d\mu^\star$. To make this statement precise, we introduce the following notion of a Chebyshev system:

Definition 2. A system of linearly independent continuous functions $\{u_i(t)\}_{i=0}^n$ defined on $I \subset \mathbb{R}$ is a Chebyshev system if no non-trivial linear combination $u(t) = \sum_{i=0}^n a_i u_i(t)$ with $a_i \in \mathbb{R}$ has more than n distinct zeros in I.

We list some Chebyshev systems. For further examples, see the monograph of M. Krein and A. Nudelman [25] and of S. Karlin and W. Studden [26].

Example 2 (Chebyshev systems)**.**

1) algebraic polynomials: $\{t^i\}_{i=0}^n$ on any interval.
2) trigonometric polynomials: $\{1, \sin(k\theta), \cos(k\theta)\}_{k=1}^n$ on $[0, 2\pi)$.
3) rational functions: $\{\frac{1}{s_i + t}\}_{i=0}^n$ with $s_i > 0$ on $(0, \infty)$.
4) exponentials: $\{e^{\alpha_i t}\}_{i=0}^n$ with $\alpha_i > 0$ on any interval.
5) Gaussian functions: $\{e^{-(s_j - t)^2}\}$ for $s_j > 0$ on $(-\infty, \infty)$.
6) Totally positive kernels: A continuous function $G(t, s)$ defined on $[a, b] \times [c, d]$ is called a totally positive kernel if for any n and any points $(a \leq) t_0 < t_1 < \cdots < t_n (\leq b), (c \leq) s_0 < s_1 < \cdots < s_n (\leq d)$, the determinant $\det([G(t_j, s_k)]_{j,k=0}^n) > 0$. The system $\{u_k(t) = G(t, s_k)\}_{k=0}^n$ is a Chebyshev system if $G(t, s)$ is totally positive. Since the kernels $1/(t+s), e^{ts}, t^s$ and $e^{-(s-t)^2}$ are all totally positive, all previous examples except the trigonometric polynomials are special cases of this example.

We have the following super-resolution result for positive combinations of atoms.

Theorem 2.

1) *Suppose the atom components $\{a_i(\theta)\}_{i=0}^n$ form a Chebyshev system on $[a, b]$. Define $\omega(\theta) = 2$ if $\theta \notin \{a, b\}$ and 1 otherwise. If a decomposition $z = \sum_j c_j a(\theta_j)$ with $c_j > 0$ satisfies $\sum_j \omega(\theta_j) \leq n$, then it uniquely solves* (15)*.*

2) *For the spherical harmonic atom (4) with $d \geq 2$, the decomposition $z = \sum_{j=1}^k c_j a(x_j)$ with $c_j > 0$ is unique if $k \leq m$.*

The proof is based on explicit constructions of dual polynomials according to Proposition 1 introduced in next section.

IV. DUALITY, MARKOV-BERNSTEIN TYPE INEQUALITIES, AND RESOLUTION LIMIT

A. Duality

From the primal formulation of the atomic norm, the resolution limit exists because including highly correlated atoms in a decomposition is not economical when measured by the ℓ_1 norm of the coefficients. However, the argument becomes more apparent if we examine the dual formulation. Regardless of $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, the dual problems of the optimizations (11) and (15) are respectively

$$\underset{q \in \mathbb{F}^n}{\text{maximize}} \ \langle q, z \rangle_\mathbb{R} \text{ subject to } \sup_{\theta \in \Theta} |\langle q, a(\theta) \rangle_\mathbb{R}| \leq 1, \tag{16}$$

$$\underset{q \in \mathbb{F}^n}{\text{maximize}} \ \langle q, z \rangle_\mathbb{R} \text{ subject to } \sup_{\theta \in \Theta} \langle q, a(\theta) \rangle_\mathbb{R} \leq 0, \tag{17}$$

where we used the notation $\langle x, y \rangle_\mathbb{R} = \text{real}(y^* x)$. Define the dual polynomial (function) $Q(\theta) = \langle q, a(\theta) \rangle = \text{real}(a(\theta)^* q), \theta \in \Theta$. The following proposition is the basis for establishing sufficient or necessary conditions for atomic decompositions.

Proposition 1. Suppose $\mu^\star = \sum_{j=1}^k c_j \delta(\theta - \theta_j)$ with support $S = \{\theta_j\}_{j=1}^k$ and $z = \int_\Theta a(\theta) d\mu^\star$.

1) The measure $\mu^\star$ is a solution to (11) if and only if there exists a dual function $Q(\theta) = \langle q, a(\theta) \rangle$ such that

$$\begin{cases} |Q(\theta)| \le 1, & \forall \theta \\ Q(\theta_j) = \text{sign}(c_j), & j = 1, \ldots, k. \end{cases} \quad (18)$$

If in addition that $\{a(\tilde{\theta}_j), \tilde{\theta}_j \in T\}$ are linearly independent, where T is a super set of S (possibly equal), and the condition $|Q(\theta)| \le 1, \forall \theta$ is strengthened to $|Q(\theta)| < 1, \forall \theta \in \Theta/T$, then the solution $\mu^\star$ is unique.

2) Assume the coefficient $c_j > 0$ in the representation of the measure $\mu^\star$. Then it is a solution to (15) if and only if there exists a dual function $Q(\theta) = \langle q, a(\theta) \rangle$ such that

$$\begin{cases} Q(\theta) \le 0, & \forall \theta \\ Q(\theta_j) = 0, & j = 1, \ldots, k. \end{cases} \quad (19)$$

If in addition that $\{a(\tilde{\theta}_j), \tilde{\theta}_j \in T\}$ are linearly independent, where T is a super set of S (possibly equal), and the condition $Q(\theta) \le 0, \forall \theta$ is strengthened to $Q(\theta) < 0, \forall \theta \in \Theta/T$, then the solution $\mu^\star$ is unique.

The sufficient conditions established in Theorems 1 and 2 are all based on constructing a dual polynomial to certify the optimality of $\mu^\star$ according to the sufficiency part of Proposition 1. We offer a proof to Theorem 2 using the sufficiency part of Proposition 1 2).

Proof of Theorem 2:

1) The proof is a direct consequence of [26, Theorem 5.1], where a nonnegative polynomial $u(\theta) = \sum_{i=0}^n \alpha_i a_i(\theta)$ was constructed to vanish precisely at the points $\{\theta_1, \ldots, \theta_k\}$ under the conditions of Theorem 2. The only exception is that when $n = 2m$ and exactly one end point is in S, then $u(\theta)$ might also vanish at the other end point. Therefore, we could take $Q(\theta) = -u(\theta)$ as the dual certificate polynomial, and take $T = S \bigcup \{a, b\}$ in the exception case and $T = S$ otherwise as the superset in part 2) of Proposition 1. Under the Chebyshev system condition, $\{a(\hat{\theta}_i)\}$ are linear independent for any $n + 1$ points $\{\hat{\theta}_i\} \subset \Theta$. Therefore, the measure $\mu^\star$ is the unique solution to (15).

2) The construction of a certifying dual polynomial for spherical harmonic atoms are similar to the one given in [21]. ∎

The necessary part of Proposition, *i.e.*, the existence of a bounded dual polynomial that interpolates the sign of the coefficient vector on the support of $\mu^\star$ when it is an optimal solution, also suggests why there exists a resolution limit when the sign pattern is arbitrary, and why there does not exist one when the coefficients are all positive. Suppose two supporting points θ_i and θ_j have a small distance $\Delta_{i,j} = \|\theta_i - \theta_j\|_2$ and $\text{sign}(c_i) = 1, \text{sign}(c_j) = -1$. If the line segment connecting θ_i and θ_j is contained in Θ, then according to the mean value theorem, there must exist a point $\hat{\theta} \in \Theta$ such that

$$Q(\theta_i) - Q(\theta_j) = \langle \nabla Q(\hat{\theta}), \theta_i - \theta_j \rangle. \quad (20)$$

As a consequence, the gradient of Q at $\hat{\theta}$ is lower bounded:

$$\|\nabla Q(\hat{\theta})\|_2 \ge \frac{|Q(\theta_i) - Q(\theta_j)|}{\Delta_{i,j}} = \frac{|\text{sign}(c_i) - \text{sign}(c_j)|}{\Delta_{i,j}} = \frac{2}{\Delta_{i,j}}. \quad (21)$$

In next section, we introduce Bernstein-Markov type inequalities which bound the size of gradients for classes of uniformly bounded functions. We therefore obtain an upper bound on $\Delta_{i,j}$. When the coefficients are known to be positive, the dual polynomial is not

required to go from 1 to -1, hence the argument does not apply.

B. Markov-Bernstein type inequalities

Markov-Bernstein type inequalities have their origins in the study of algebraic and trigonometric polynomials. The classic Bernstein's inequality asserts that every trigonometric polynomial $Q(\theta)$ with complex coefficients of degree at most n satisfies [15]

$$|Q'(\theta)| \le n \max_{\theta \in [0, 2\pi]} |Q(\theta)|. \quad (22)$$

By defining $Q(\theta) = p(\cos(\theta))$ for any algebraic polynomial $p(t)$ of degree n, we obtain the algebraic Bernstein's inequality [15]:

$$|p'(t)| \le \frac{n}{\sqrt{1 - t^2}} \max_{t \in [-1, 1]} |p(t)|, t \in (-1, 1). \quad (23)$$

Combining this with Shur's inequality [15], we get Markov's inequality for every algebraic polynomial $p(t)$ of degree n [15]:

$$\max_{t \in [-1, 1]} |p'(t)| \le n^2 \max_{t \in [-1, 1]} |p(t)|, t \in [-1, 1]. \quad (24)$$

Given a class of functions $\mathcal{F} = \{f(\theta)\}$ defined on Θ, the Bernstein type inequality is of the form

$$\|\nabla f(\theta)\|_2 \le c_{\mathcal{F}}(\theta) \sup_{\theta \in \Theta} |f(\theta)|, \forall f \in \mathcal{F}, \quad (25)$$

while the Markov type inequality is of the form

$$\sup_{\theta \in \Theta} \|\nabla f(\theta)\|_2 \le C_{\mathcal{F}} \sup_{\theta \in \Theta} |f(\theta)|, \forall f \in \mathcal{F}. \quad (26)$$

Here $c_{\mathcal{F}}(\theta)$ is a function of θ and $C_{\mathcal{F}}$ is a numerical constant, both of which depend only on the function class $\mathcal{F}$. Note that the Bernstein's inequality bounds the size of gradient pointwise, while Markov's inequality bounds the uniform norm of the gradient. When $\mathcal{F}$ is the set of trigonometric polynomials of degree n, we have $c_{\mathcal{F}}(\theta) = C_{\mathcal{F}} = n$.

Besides the classic Markov-Bernstein type inequalities, many others have been developed in the literature such as those for homogeneous and non-homogeneous polynomials defined on a convex body [27], for rational functions [28], for spherical harmonic polynomials [29], and for real exponentials [30].

C. Resolution Limit

We are now ready to state the major theorem on the resolution limit of atomic decompositions:

Theorem 3. *For each of the atomic set $\mathcal{A} = \{a(\theta) : \theta \in \Theta\}$ listed below, we consider a set of parameters $S = \{\theta_1, \theta_2, \ldots, \theta_r\} \subset \Theta$ such that any measure $\mu = \sum_{j=1}^r c_j \delta(\theta - \theta_j)$ supported on S is an atomic decomposition, regardless of the sign pattern of $\{c_j\}_{j=1}^r$.*

1) *For the 1D line spectral atom defined in (2), the minimal separation of points in S satisfies*

$$\Delta = \min_{i \ne j} |\theta_i - \theta_j| \ge 2/n. \quad (27)$$

2) *For multi-dimensional line spectral atoms defined in (3), the separation of points in $S = \{\boldsymbol{\theta}^1, \cdots, \boldsymbol{\theta}^r\}$ satisfies*

$$\Delta = \min_{i \ne j} \sum_{\ell=1}^d n_\ell |\theta_\ell^i - \theta_\ell^j| \ge 2 \quad (28)$$

3) *For the spherical harmonic atom defined in (4), the minimal separation of points in $S = \{x^1, \ldots, x^r\}$ satisfies*

$$\Delta = \min_{i \ne j} \arccos(\langle x^i, x^j \rangle) \ge 2/m. \quad (29)$$

4) *For the mth order symmetric tensor atomic set $\mathcal{A} = \{a(x) = x^{\otimes m} : x \in \mathbb{S}^{d-1}\}$, the minimal separation of points in $S = \{x^1, \ldots, x^r\}$ satisfies*

$$\Delta = \min_{i \neq j} \arccos(|\langle x^i, x^j \rangle|) \geq 2/m. \tag{30}$$

5) *For the mth order non-symmetric tensor atomic set define in (7), the minimal separation of points in $S = \{\boldsymbol{u}^1, \ldots, \boldsymbol{u}^r\}$ with $\boldsymbol{u} = (u_1, \cdots, u_m)$ satisfies*

$$\Delta = \min_{i \neq j} \sum_{\ell=1}^{m} \arccos(|\langle u_\ell^i, u_\ell^j \rangle|) \geq 2. \tag{31}$$

Proof: Without loss of generality, we consider the separation between θ_1 and θ_2. The dual polynomial $Q(\theta) = \langle q, a(\theta) \rangle$ that certifies the optimality of $\mu = \sum_{j=1}^{r} c_j \delta(\theta - \theta_j)$ with $\text{sign}(c_1) = -\text{sign}(c_2) = 1$ is a trigonometric polynomial of degree n. The classic Bernstein's inequality (22) implies that $|Q'(\theta)| \leq n$ since $Q(\theta)$ is uniformly bounded by 1. Invoking (21), we get

$$\Delta_{1,2} = d(\theta_1, \theta_2) \geq 2/n, \tag{32}$$

as desired.

For other parts, one can use the same technique in part 1) by noting that: for part 2), the dual certificate $Q(\boldsymbol{\theta}) = Q(\theta_1^i, \ldots, \theta_\ell, \ldots, \theta_d^i)$ is a trigonometric polynomial of θ_ℓ of degree n_ℓ; for part 3), consider the separation between x^1 and x^2, then the dual polynomial $Q(x) = \langle q, a(x) \rangle$ becomes a trigonometric polynomial of degree m when restricted onto the the great circle that passes x^1 and x^2; part 4) and part 5) is similar to part 2) and part 3), respectively. ∎

Comparing with the sufficient conditions in Theorem 1, we observe that 1) and 3) of Theorem 3 are of the same order. For 2), if all $n_\ell = n$, we can obtain a weaker bound using ℓ_∞ norm

$$\Delta = \min_{i \neq j} \|\boldsymbol{\theta}^i - \boldsymbol{\theta}^j\|_\infty \geq 2/nd, \tag{33}$$

which has the same order in n as the sufficient condition. However, it is not clear whether the $1/d$ dependence on d is optimal or not. The sufficient condition for 3rd order symmetric tensor implies $\min_{i \neq j} \arccos(|\langle x_i, x_j \rangle|) \geq \arccos(0.0016) \approx 1.57 > 2/3$, which is stronger than the necessary condition (30).

V. Conclusions

We conclude by pointing out some implications of the resolution limits of atomic decompositions. One one hand, if the goal is parameter estimation, the established resolution limits imply that atomic norm minimization can not recover closely spaced parameters, which will impact applications where super-resolution is important. On the other hand, if the goal is prediction, as common in machine learning tasks, well-separated parameters can prevent over-fitting and hence improve generalization. We leave making the latter statement precise to future work. Other future directions include investigating alternative regularizers that have better resolution, and effects of incorporating prior information other than the sign patterns.

References

[1] P. J. Bickel, B. Li, A. B. Tsybakov, S. A. van de Geer, B. Yu, T. Valdés, C. Rivero, J. Fan, and A. van der Vaart, "Regularization in statistics," *Test*, vol. 15, no. 2, pp. 271–344, Sep. 2006.

[2] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*. Springer Science & Business Media, Mar. 2000.

[3] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer Science & Business Media, 2000.

[4] E. J. Candes and M. B. Wakin, "An Introduction To Compressive Sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[5] E. J. Candès and B. Recht, "Exact Matrix Completion via Convex Optimization," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[6] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The Convex Geometry of Linear Inverse Problems," *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.

[7] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed Sensing Off the Grid," *Information Theory, IEEE Transactions on*, vol. 59, no. 11, pp. 7465–7490, 2013.

[8] E. J. Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, 2008.

[9] E. J. Candès and C. Fernandez-Granda, "Towards a Mathematical Theory of Super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, Jun. 2014.

[10] G. Tang and P. Shah, "Guaranteed Tensor Decomposition: A Moment Approach," *Submitted to International Conference on Machine Learning*, 2015.

[11] G. Tang and B. Recht, "Atomic decomposition of mixtures of translation-invariant signals," *Preprint*.

[12] T. Bendory, S. Dekel, and A. Feuer, "Robust Recovery of Stream of Pulses using Convex Optimization," *arXiv.org*, Dec. 2014.

[13] ——, "Super-resolution on the Sphere using Convex Optimization," *arXiv.org*, Dec. 2014.

[14] R. Heckel, R. Heckel, V. I. Morgenshtern, V. I. Morgenshtern, and M. Soltanolkotabi, "Super-Resolution Radar," *arXiv.org*, 2014.

[15] P. Borwein and T. Erdélyi, *Polynomials and Polynomial Inequalities*. New York: Springer, 1995.

[16] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[17] G. Tang, B. N. Bhaskar, and B. Recht, "Near Minimax Line Spectral Estimation," *Information Theory, IEEE Transactions on*, vol. 61, no. 1, pp. 499–512, Jan 2014.

[18] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic Norm Denoising With Applications to Line Spectral Estimation," *Signal Processing, IEEE Transactions on*, vol. 61, no. 23, pp. 5987–5999, 2013.

[19] Y. Chi and Y. Chen, "Compressive Two-Dimensional Harmonic Retrieval via Atomic Norm Minimization," *Signal Processing, IEEE Transactions on*, no. 99, pp. 1–1042, 2014.

[20] 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 ≥ 2) off-the-grid frequencies," *Information Theory and Applications Workshop (ITA), 2014*, pp. 1–4, 2014.

[21] T. Bendory, S. Dekel, and A. Feuer, "Exact recovery of Dirac ensembles from the projection onto spaces of spherical harmonics," *arXiv.org*, Dec. 2014.

[22] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, 2010.

[23] P. Shah, B. N. Bhaskar, G. Tang, and B. Recht, "Linear System Identification via Atomic Norm Regularization," in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*. IEEE, 2012, pp. 6265–6270.

[24] A. Barvinok, *A Course in Convexity*. American Mathematical Soc., 2002.

[25] M. G. Krein and A. A. Nudelman, *The Markov moment problem and extremal problems*, ser. Ideas and problems of P. L. Čebyšev and A. A. Markov and their further development. Providence, Rhode Island: American Mathematical Society, 1977, vol. 50.

[26] S. Karlin and W. J. Studden, *Tchebycheff systems: With applications in analysis and statistics*. New York: John Wiley & Sons, 1966.

[27] Y. SARANTOPOULOS, "Bounds on the derivatives of polynomials on Banach spaces," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 110, no. 02, p. 307, Oct. 2008.

[28] P. Borwein, T. Erdélyi, and J. Zhang, "Chebyshev polynomials and Markov-Bernstein type inequalities for rational spaces," *Journal of the London Mathematical Society. Second Series*, vol. 50, no. 3, pp. 501–519, 1994.

[29] O. Kounchev, *Multivariate Polysplines*, ser. Applications to Numerical and Wavelet Analysis. Academic Press, Jun. 2001.

[30] V. Y. Protasov and R. M. Jungers, "Stability of linear switching systems and Markov-Bernstein inequalities for exponents," *arXiv.org*, Jul. 2014.