GRIDLESS DOA ESTIMATION UNDER THE MULTI-FREQUENCY MODEL

Yifan Wu[1]*, Michael B. Wakin*[2]*, and Peter Gerstoft*[1]

[1] University of California, San Diego, La Jolla, CA, USA
[2] Colorado School of Mines, Golden, CO, USA

ABSTRACT

Direction of Arrival (DOA) estimation is widely applied in acoustic source localization. A multi-frequency model is suitable for characterizing the broadband structure in acoustic signals. In this work, we solve the *continuous* (gridless) line spectrum estimation problem by incorporating the multi-frequency model into an atomic norm minimization (ANM) framework. We show that our ANM problem is equivalent to a semi-definite program (SDP) which can be solved by an off-the-shelf SDP solver. We also provide the dual certificate that can certify the optimality of the SDP solution, and we localize the sources by finding the peaks of the norm of the dual polynomial. Numerical results support our theoretical findings and demonstrate the effectiveness of the method.

Index Terms— Atomic norm minimization, DOA estimation, multi-frequency model, semi-definite programming, trigonometric polynomials

1. INTRODUCTION

Direction of Arrival (DOA) estimation is an important topic in array processing, with a wide range of applications in wireless communications, sensor networks, and speech communication. Conventional DOA estimation methods assume a narrowband signal model, and they are successful for narrowband signals (e.g., radar signals) whose bandwidths are significantly less than the central frequency. However, the narrowband assumption may not be valid in acoustic localization scenarios since acoustic signals are typically broadband. Therefore, for such scenarios it is necessary to consider a *multi-frequency model* that decomposes the broadband signal into multiple narrowband signals. Multi-frequency models have been widely applied in acoustic source localization [1–3] when the signal contains a wide range of frequency bins and cannot be characterized by a narrowband model. Grid-based sparse localization approaches for the multi-frequency model have been proposed [2–6] for robustness and aliasing suppression. However, these methods assume that the true DOAs lie on a finite set of grid points, and their performance may degrade if the DOAs fall off the grid.

To overcome the grid mismatch problem, atomic norm minimization (ANM) methods that work on continuous (gridless) dictionaries have been proposed in a variety of contexts [7–18]. ANM extends grid-based ℓ_1 norm minimization to the continuous setting and is commonly applied to solve the line spectrum estimation problem for signals that are sparse in the temporal frequency domain. As is well known, narrowband DOA estimation problems can be mapped into an equivalent problem, where the cosine of each source's DOA maps to a single temporal frequency.

The authors in the pioneering ANM paper [7] worked directly with the continuous (temporal) frequency estimation problem and considered the complete data case. As long as the temporal frequency separation was greater than the minimum separation, exact recovery of the active temporal frequencies was guaranteed. Furthermore, a semidefinite programming (SDP) framework that characterized the ANM problem was presented. Inspired by [7], the authors in [8] studied the continuous temporal frequency estimation based on randomly sampled data for the single measurement vector (SMV) case. The minimum separation condition was relaxed in [11]. ANM for multiple measurement vectors (MMVs) was studied in [10,13]. In [12], the author considered a super-resolution problem that had a similar setup to [7] except that the point spread function was assumed to be unknown. Based on the assumption that the point spread function was stationary and lived in a known subspace, the lifting trick was applied, and the problem was formulated using ANM. The model was generalized to non-stationary point spread functions in [14]. The sample complexity of modal analysis with random temporal compression was established in [15]. ANM for 2D temporal frequency estimation was studied in [9]. In [16], the authors proposed a reweighted ANM framework, which can enhance the sparsity and achieve super-resolution. ANM was also recently applied in digital beamforming [19], adaptive interference cancellation [20], time-varying channel estimation [21], denoising [22, 23], and blind demodulation [24, 25]. We refer readers to [26] for a comprehensive overview of ANM and its applications. However, all of these prior works implicitly used a narrowband signal assumption and are not appropriate to broadband DOA estimation.

In this paper, we consider the *continuous* (gridless) line spectrum estimation problem under the multi-frequency model. It turns out that this problem can be reformulated as a certain continuous ANM problem so that super-resolution can be achieved. This ANM problem can be further characterized by a trigonometric polynomial problem that was extensively discussed in [27]. By applying the bounded real lemma for vector trigonometric polynomials in [27], we equivalently formulate our ANM problem as an SDP problem that can be solved by an off-the-shelf SDP solver (such as CVX [28]). In addition, we provide the dual certificate that certifies the optimality of the atomic decomposition. Numerical results support our theory and demonstrate the effectiveness of the method.

2. SIGNAL MODEL

We make the following assumptions for the signal model:

1. There are N_m sensors forming a uniform linear array (ULA) with array spacing d.
2. There are K active sources impinging on the array from unknown directions of arrival (DOAs) θ.
3. Each source has N_f active temporal frequency components, each at a multiple of a fundamental frequency f_0, i.e., $f \in \{1, 2, \ldots, N_f\}$ and $f f_0 \in \{f_0, 2f_0, \ldots, N_f f_0\}$.

4. We suppose $\frac{2\pi f_0 d}{c} \leq \pi$ or $d \leq \frac{c}{2f_0}$ holds, where c is the speed of propagation. Note that this is only a technical assumption for simplifying the theoretical analysis and it is not required in the practical applications. We also notice that this is the maximum separation for the fundamental frequency to avoid grating lobe (aliasing). For higher frequencies (i.e. $f \geq 2$), the grating lobe will still exist. This gives the multi-frequency DOA problem an interesting structure in that the data at each higher frequency is aliased; such aliasing is not considered in conventional narrowband ANM papers.

Based on the above assumptions, we absorb the constant parameters d, f_0, and c into a scaled DOA parameter $w = w(\theta) := \frac{2\pi f_0 d \cos(\theta)}{c} \in [-\pi, \pi]$. We often refer to w simply as the DOA.

For each frequency index $f \in \{1, 2, \ldots, N_f\}$, let $\mathbf{y}_f \in \mathbb{C}^{N_m}$ denote the received signal across the N_m sensors. Stacking all of these into a matrix, the full set of received data is denoted by $\mathbf{Y} := [\mathbf{y}_1 \ldots \mathbf{y}_{N_f}] \in \mathbb{C}^{N_m \times N_f}$. Summing over the K active DOAs, gives

$$\mathbf{Y} = \sum_w [x_w(1)\mathbf{a}(1, w) \ldots x_w(N_f)\mathbf{a}(N_f, w)], \tag{1}$$

where $\mathbf{a}(f, w) := \frac{1}{\sqrt{N_m}}[1 \ldots e^{-jwf(N_m-1)}]^T = \frac{1}{\sqrt{N_m}}[z^{f\cdot 0} \ldots z^{f\cdot(N_m-1)}]^T \in \mathbb{C}^{N_m}$ is the (normalized) array manifold vector (steering vector) corresponding to frequency bin f and DOA w, and $x_w(f)$ is the signal amplitude for the f-th frequency bin (z is defined as e^{-jw}). Our goal is to identify the K active DOAs w from the data matrix $\mathbf{Y}$ appearing in (1).

Define a matrix $\mathbf{X} \in \mathbb{C}^{N_m \times N_f}$ as follows

$$\mathbf{X} := \sum_w c_w \mathbf{A}(w) \circledast \mathbf{x}_w^T, \tag{2}$$

where $\mathbf{A}(w) := [\mathbf{a}(1, w) \ldots \mathbf{a}(N_f, w)] \in \mathbb{C}^{N_m \times N_f}$ and $\mathbf{x}_w := [x_w(1) \ldots x_w(N_f)]^T \in \mathbb{C}^{N_f}$. Without loss of generality, $\|\mathbf{x}_w\|_2 = 1$ is assumed and the coefficient c_w is used to absorb any other scaling of the amplitude vector. In (2), $\circledast$ is the operator defined as $[\mathbf{A} \circledast \mathbf{b}^T] := [\mathbf{a}_1 b(1) \ldots \mathbf{a}_{N_f} b(N_f)] \in \mathbb{C}^{N_m \times N_f}$ where $\mathbf{A} = [\mathbf{a}_1 \ldots \mathbf{a}_{N_f}] \in \mathbb{C}^{N_m \times N_f}$, and $\mathbf{b} = [b(1) \ldots b(N_f)]^T \in \mathbb{C}^{N_f}$. Finally, we define the identity operator $\mathcal{L} : \mathbb{C}^{N_m \times N_f} \to \mathbb{C}^{N_m \times N_f}$ such that $\mathbf{Y} = \mathcal{L}(\mathbf{X})$.

3. METHODOLOGY

3.1. Atomic Norm Minimization

To efficiently represent tensors of the form (2), we define the following atomic set ($d = \frac{c}{2f_0}$ is assumed here, and this assumption will be relaxed to $d \leq \frac{c}{2f_0}$ in the later discussions):

$$\mathcal{A} := \{\mathbf{A}(w) \circledast \mathbf{x}_w^T : \ w \in [-\pi, \pi], \mathbf{x}_w \in \mathbb{C}^{N_f}, \|\mathbf{x}_w\|_2 = 1\}. \tag{3}$$

From (2), $\mathbf{X}$ is a sparse combination of K atoms from $\mathcal{A}$ since only a few directions have active sources. ANM provides a framework for identifying such sparse combinations in continuously parameterized dictionaries. In our case, the dictionary $\mathcal{A}$ parameterized by the continuous DOA w. Thus, by applying ANM on $\mathbf{Y} = \mathcal{L}(\mathbf{X})$, grid-free source localization can be achieved.

We propose the following ANM-based optimization framework:

$$\min_{\mathbf{X}} \quad \|\mathbf{X}\|_{\mathcal{A}} \quad \text{s.t.} \qquad \mathbf{Y} = \mathcal{L}(\mathbf{X}), \tag{4}$$

where the atomic norm is defined as $\|\mathbf{X}\|_{\mathcal{A}} := \inf\{t \geq 0 | \mathbf{X} \in t \cdot \text{conv}(\mathcal{A})\} = \{\sum_w c_w | \mathbf{X} = \sum_w c_w \mathbf{A}(w) \circledast \mathbf{x}_w^T, c_w \geq 0\}$. Based on the definition of the dual norm, we express $\|\mathbf{X}\|_{\mathcal{A}}$ in terms of the dual atomic norm $\|\mathbf{Q}\|_{\mathcal{A}}^*$ as follows

$$\begin{aligned} \|\mathbf{X}\|_{\mathcal{A}} &:= \sup_{\|\mathbf{Q}\|_{\mathcal{A}}^* \leq 1} \langle \mathbf{Q}, \mathbf{X} \rangle_{\mathbb{R}} = \sup_{\|\mathbf{Q}\|_{\mathcal{A}}^* \leq 1} \langle \mathcal{L}(\mathbf{Q}), \mathcal{L}(\mathbf{X}) \rangle_{\mathbb{R}} \\ &= \sup_{\|\mathbf{Q}\|_{\mathcal{A}}^* \leq 1} \langle \mathcal{L}(\mathbf{Q}), \mathbf{Y} \rangle_{\mathbb{R}}. \end{aligned} \tag{5}$$

Now, we consider the constraint $\|\mathbf{Q}\|_{\mathcal{A}}^* \leq 1$ (note that $\|\mathbf{x}_w\|_2 = 1$ will be used in the following derivation, and $\mathbf{Q} := [\mathbf{q}_1 \ldots \mathbf{q}_{N_f}] \in \mathbb{C}^{N_m \times N_f}$ is the dual variable):

$$\begin{aligned} \|\mathbf{Q}\|_{\mathcal{A}}^* &:= \sup_{\|\mathbf{X}\|_{\mathcal{A}} \leq 1} \langle \mathbf{Q}, \mathbf{X} \rangle_{\mathbb{R}} = \sup_{\substack{\mathbf{x}_w \\ w}} \langle \mathbf{Q}, \mathbf{A}(w) \circledast \mathbf{x}_w^T \rangle_{\mathbb{R}} \\ &= \sup_{\substack{x_w(f) \\ w}} \sum_{f=1}^{N_f} \langle \mathbf{q}_f, x_w(f)\mathbf{a}(f, w) \rangle_{\mathbb{R}} \\ &= \sup_{\substack{x_w(f) \\ w}} \text{Re}\Big(\sum_{f=1}^{N_f} x_w(f)\mathbf{q}_f^H \mathbf{a}(f, w)\Big) \\ &\overset{(a)}{=} \sup_{\substack{\mathbf{x}_w \\ w}} \text{Re}(\mathbf{x}_w^H \psi(\mathbf{Q}, \mathbf{a}(w))) = \sup_{\substack{\mathbf{x}_w \\ w}} |\mathbf{x}_w^H \psi(\mathbf{Q}, \mathbf{a}(w))| \\ &\overset{(b)}{=} \sup_w \|\psi(\mathbf{Q}, \mathbf{a}(w))\|_2 \overset{(c)}{=} \sup_{w \in [-\pi, \pi]} \sigma_{\max}(\mathbf{h}(w)) \leq 1, \end{aligned} \tag{6}$$

where (a) follows by the definition $\psi(\mathbf{Q}, \mathbf{a}(w)) := [\mathbf{q}_1^H \mathbf{a}(1, w) \ldots \mathbf{q}_{N_f}^H \mathbf{a}(N_f, w)]^T \in \mathbb{C}^{N_f}$ (this is called the dual polynomial [7], [8]), and (b) follows from the definition of the operator norm. For (c), we define the vector polynomial $\mathbf{h}(w)$ as follows. Each entry of $\psi(\mathbf{Q}, \mathbf{a}(w))$ is the inner product between a slice of $\mathbf{Q}$ and an array manifold vector with *different parameters*, which serves as an obstacle for us to find an efficient polynomial representation. To overcome this challenge, we define $\tilde{\mathbf{a}}(w) := \frac{1}{\sqrt{N}}[1 \quad e^{-jw1} \ldots e^{-jwN_f(N_m-1)}]^T \in \mathbb{C}^N$, where $N := N_f(N_m - 1) + 1$. Then there exists a coefficient vector $\mathbf{u}_f \in \mathbb{C}^N$ (a zero-padded version of $\mathbf{q}_f$) such that $\mathbf{u}_f^H \tilde{\mathbf{a}}(w) = \mathbf{q}_f^H \mathbf{a}(f, w) \quad \forall f \in \{1, \ldots, N_f\}$. Hence,

$$\begin{aligned} \psi(\mathbf{Q}, \mathbf{a}(w)) &:= [\mathbf{q}_1^H \mathbf{a}(1, w) \ldots \mathbf{q}_{N_f}^H \mathbf{a}(N_f, w)]^T \\ &= \begin{bmatrix} \mathbf{u}_1^H \\ \vdots \\ \mathbf{u}_{N_f}^H \end{bmatrix} \tilde{\mathbf{a}}(w) = \mathbf{H}\tilde{\mathbf{a}}(w) := \mathbf{h}(w), \end{aligned} \tag{7}$$

where the last line follows by defining $\mathbf{H} := [\mathbf{u}_1 \ldots \mathbf{u}_{N_f}]^H \in \mathbb{C}^{N_f \times N}$. Denoting the columns of $\mathbf{H}$ as $\mathbf{H} := [\mathbf{h}_0 \ldots \mathbf{h}_{N-1}]$, we have $\mathbf{h}(w) = \sum_{k=0}^{N-1} \mathbf{h}_k e^{-jwk} = \sum_{k=0}^{N-1} \mathbf{h}_k z^k$. Stacking these columns, we define $\bar{\mathbf{H}} := [\mathbf{h}_0^T \ldots \mathbf{h}_{N-1}^T]^T \in \mathbb{C}^{N \cdot N_f}$. Finally, inspired by [27], we have the following proposition.

Proposition 3.1 *Let $\mathbf{h}(w)$ be as defined in (7). Then $\|\mathbf{Q}\|_{\mathcal{A}}^* = \sup_{w \in [-\pi, \pi]} \sigma_{\max}(\mathbf{h}(w)) \leq 1$ holds if and only if there exists a matrix $\mathbf{Q}_0 \in \mathbb{C}^{N \times N} \succeq 0$ such that*

$$\sum_{i=1}^{N-k} \mathbf{Q}_{i,i+k} = \delta_k = \begin{cases} 1, & k = 0, \\ 0, & k = 1, \ldots, N-1, \end{cases} \tag{8}$$

and such that

$$\begin{bmatrix} \mathbf{Q}_0 & \widehat{\mathbf{H}} \\ \widehat{\mathbf{H}}^H & \mathbf{I}_{N_f} \end{bmatrix} \succeq 0. \tag{9}$$

Proof Construct the Hermitian trigonometric polynomial $R(w) := 1 - \mathbf{H}^H(w)\mathbf{H}(w)$. Note $\mathbf{h}^H(w)\mathbf{h}(w) = \|\mathbf{h}(w)\|_2^2 = \sigma_{\max}^2(\mathbf{h}(w))$. Therefore, $R(w) \geq 0$ if and only if $\sigma_{\max}(\mathbf{h}(w)) \leq 1$.

Now we consider the condition for $R(w) \geq 0$. Define $R_0(w) = 1$ and $R_1(w) = \mathbf{h}^H(w)\mathbf{h}(w)$, thus $R(w) = R_0(w) - R_1(w)$. Based on Lemma 4.25 in [27], $R_0(w) \geq R_1(w)$ if and and only if $\mathbf{Q}_0 \succeq \mathbf{Q}_1$, where $\mathbf{Q}_0 \in \mathbb{C}^{N\times N}$ and $\mathbf{Q}_1 \in \mathbb{C}^{N\times N}$ are Gram matrices associated with $R_0(w)$ and $R_1(w)$, respectively.

We further observe $\mathbf{h}(w) = [\mathbf{h}_0 \dots \mathbf{h}_{N-1}][z^0 \dots z^{N-1}]^T = \widehat{\mathbf{H}}^H[z^0 \dots z^{N-1}]^T := \widehat{\mathbf{H}}^H\mathbf{z}$. Then, the Gram matrix representation of $R_1(w)$ can be obtained as follows

$$R_1(w) = \mathbf{h}^H(w)\mathbf{h}(w) = \mathbf{z}^H\widehat{\mathbf{H}}\widehat{\mathbf{H}}^H\mathbf{z}. \tag{10}$$

Hence, $\mathbf{Q}_1 = \widehat{\mathbf{H}}\widehat{\mathbf{H}}^H$ is the Gram matrix associated with $R_1(w)$. Since $\mathbf{Q}_0 \succeq \mathbf{Q}_1$, we have $\mathbf{Q}_0 \succeq \widehat{\mathbf{H}}\widehat{\mathbf{H}}^H$. Based on Schur complement condition, (9) will be hold.

On the other hand, based on the Gram matrix representation of $R_0(w) = \mathbf{z}^H\mathbf{Q}_0\mathbf{z}$, we have

$$1 = R_0(w) = \mathrm{Tr}(R_0(w)) = \mathrm{Tr}(\mathbf{z}^H\mathbf{Q}_0\mathbf{z}) = \mathrm{Tr}(\mathbf{z}\mathbf{z}^H\mathbf{Q}_0). \tag{11}$$

Since $\mathbf{z}\mathbf{z}^H = \sum_{k=-(N-1)}^{N-1} \mathbf{\Theta}_k z^{-k}$ ($\mathbf{\Theta}_k \in \mathbb{R}^{N\times N}$ is the matrix such that all non-zero entries only lie in the k-th diagonoal), then

$$1 = R_0(w) = \mathrm{Tr}\Big(\sum_{k=-(N-1)}^{N-1} \mathbf{\Theta}_k\mathbf{Q}_0 z^{-k}\Big). \tag{12}$$

To ensure (12) is satisfied, (8) needs to be held. This completes the proof. □

3.2. Equivalent SDP

From Proposition 3.1 and (5)-(6) , we have the following SDP that is equivalent to (4):

$$\begin{aligned} &\max_{\mathbf{Q},\mathbf{Q}_0} \langle \mathcal{L}(\mathbf{Q}), \mathbf{Y}\rangle_{\mathbb{R}} \\ &\text{s.t.} \begin{bmatrix} \mathbf{Q}_0 & \widehat{\mathbf{H}} \\ \widehat{\mathbf{H}}^H & \mathbf{I}_{N_f} \end{bmatrix} \succeq 0 \\ &\sum_{i=1}^{N-k} \mathbf{Q}_{i,i+k} = \delta_k. \end{aligned} \tag{13}$$

3.3. Dual Certificate

The dual polynomial $\psi(\mathbf{Q}, \mathbf{a}(w))$ that was introduced in (7) serves as a certificate for the optimality of (4). Specifically, we have the following theorem.

Theorem 3.2 *Let $\mathcal{W}$ denote a collection of source directions w; its cardinality $|\mathcal{W}|$ equals the number of sources. Then $\mathbf{X} = \sum_{w\in\mathcal{W}} c_w\mathbf{A}(w) \circledast \mathbf{x}_w^T$ ($\|\mathbf{x}_w\|_2 = 1$) is the unique atomic decomposition such that $\|\mathbf{X}\|_{\mathcal{A}} = \sum_{w\in\mathcal{W}} |c_w|$ if the following two conditions are satisfied:*
(1) There exists $\mathbf{Q}$ such that the dual polynomial $\psi(\mathbf{Q}, \mathbf{a}(w))$ satisfies

$$\begin{cases} \psi(\mathbf{Q}, \mathbf{a}(w)) = \mathrm{sign}(c_w^*)\mathbf{x}_w & \forall w \in \mathcal{W} \\ \|\psi(\mathbf{Q}, \mathbf{a}(w))\|_2 < 1 & \forall w \notin \mathcal{W}, \end{cases} \tag{14}$$

where $\mathrm{sign}(c_w^*) := \frac{c_w^*}{|c_w^*|}$.
(2) $\{\mathbf{A}(w) \circledast \mathbf{x}_w^T : w \in \mathcal{W}\}$ is a linearly independent set.

Proof First, notice that if (14) is satisfied, based on (6), we have $\|\mathbf{Q}\|_{\mathcal{A}}^* \leq 1$. Then,

$$\begin{aligned} \|\mathbf{X}\|_{\mathcal{A}} &\geq \|\mathbf{X}\|_{\mathcal{A}} \cdot \|\mathbf{Q}\|_{\mathcal{A}}^* \overset{(a)}{\geq} \langle \mathbf{Q}, \mathbf{X}\rangle_{\mathbb{R}} \\ &= \langle \mathbf{Q}, \sum_{w\in\mathcal{W}} c_w\mathbf{A}(w) \circledast \mathbf{x}_w^T\rangle_{\mathbb{R}} = \sum_{w\in\mathcal{W}} \mathrm{Re}[c_w\langle \mathbf{Q}, \mathbf{A}(w) \circledast \mathbf{x}_w^T\rangle] \\ &= \sum_{w\in\mathcal{W}} \sum_{f=1}^{N_f} \mathrm{Re}[c_w\langle \mathbf{q}_f, x_w(f)\mathbf{a}(f,w)\rangle] \\ &= \sum_{w\in\mathcal{W}} \mathrm{Re}[c_w\langle \mathbf{x}_w, \psi(\mathbf{Q}, \mathbf{a}(w))\rangle] \\ &\overset{(b)}{=} \sum_{w\in\mathcal{W}} \mathrm{Re}[c_w \cdot \mathrm{sign}(c_w^*)\|\mathbf{x}_w\|_2^2] = \sum_{w\in\mathcal{W}} |c_w| \overset{(c)}{\geq} \|\mathbf{X}\|_{\mathcal{A}}, \end{aligned} \tag{15}$$

where (a) is based on Hölder's inequality [29], (b) follows because if $w \in \mathcal{W}$, then $\psi(\mathbf{Q}, \mathbf{a}(w)) = \mathrm{sign}(c_w^*)\mathbf{x}_w$ based on (14), and (c) follows from the definition of the atomic norm. Hence, $\|\mathbf{X}\|_{\mathcal{A}} = \langle \mathbf{Q}, \mathbf{X}\rangle_{\mathbb{R}} = \sum_{w\in\mathcal{W}} |c_w|$.

For the uniqueness, suppose there exists another decomposition $\mathbf{X} = \sum_{w'} c_{w'}\mathbf{A}(w') \circledast \mathbf{x}_{w'}^T$ which satisfies $\|\mathbf{X}\|_{\mathcal{A}} = \sum_{w'} |c_{w'}|$. There must exist $w' \notin \mathcal{W}$ contributing to $\mathbf{X}$ due to the mutual linear independence for the atoms. Therefore, we have the following contradiction:

$$\begin{aligned} &\sum_{w'} |c_{w'}| = \|\mathbf{X}\|_{\mathcal{A}} = \langle \mathbf{Q}, \mathbf{X}\rangle_{\mathbb{R}} = \langle \mathbf{Q}, \sum_{w} c_{w'}\mathbf{A}(w') \circledast \mathbf{x}_{w'}^T\rangle_{\mathbb{R}} \\ &= \sum_{w'} \mathrm{Re}[c_{w'}\langle \mathbf{x}_{w'}, \psi(\mathbf{Q}, \mathbf{a}(w'))\rangle] \\ &= \sum_{w'\in\mathcal{W}} \mathrm{Re}[c_{w'}\langle \mathbf{x}_{w'}, \psi(\mathbf{Q}, \mathbf{a}(w'))\rangle] + \sum_{w'\notin\mathcal{W}} \mathrm{Re}[c_{w'}\langle \mathbf{x}_{w'}, \psi(\mathbf{Q}, \mathbf{a}(w'))\rangle] \\ &\overset{(a)}{<} \sum_{w'\in\mathcal{W}} |c_{w'}| + \sum_{w'\notin\mathcal{W}} |c_{w'}| = \sum_{w'} |c_{w'}|, \end{aligned} \tag{16}$$

where (a) is because of (14). Therefore, the atomic decomposition which satisfies $\|\mathbf{X}\|_{\mathcal{A}} = \sum_{w\in\mathcal{W}} |c_w|$ must be unique. □

Remark 2 This theorem indicates that as long as (14) and the linear independence condition are satisfied, the optimality of the atomic decomposition is guaranteed. In (14), we have $\|\psi(\mathbf{Q}, \mathbf{a}(w))\|_2 = 1$ for $w \in \mathcal{W}$ and $\|\psi(\mathbf{Q}, \mathbf{a}(w))\|_2 < 1$ for $w \notin \mathcal{W}$. Therefore, $\|\psi(\mathbf{Q}, \mathbf{a}(w))\|_2$ will achieve the maximum value at the ground-truth DOAs. To localize the DOAs, we only need to find a collection of w associated with the peaks of $\|\psi(\mathbf{Q}, \mathbf{a}(w))\|_2$. By solving ((13) via CVX [28], we can obtain $\mathbf{Q}$. By using the dual solution $\mathbf{Q}$ to construct the dual polynomial $\psi(\mathbf{Q}, \mathbf{a}(w))$ based on (7), we can evaluate the ℓ_2 norm of $\psi(\mathbf{Q}, \mathbf{a}(w))$ for arbitrary DOA w and localize the sources with the maximum ℓ_2 norm.

Remark 3 The aliasing present in the multi-frequency measurements makes further analysis of the vector-valued dual polynomial challenging, since for each $f \geq 2$, row f of $\psi(\mathbf{Q}, \mathbf{a}(w))$ is necessarily periodic in w with period $2\pi/f$. We are exploring this interesting structure in our ongoing analysis.

4. NUMERICAL EXAMPLES

We suppose $K = 3$ incoherent sources have DOAs $\theta = 80.4°, 88.6°$, and $92.5°$ ($90°$ is considered broadside). Assume $c = 340$ m/s, $f_0 = 100$ Hz, a uniform linear array with $N_m = 12$ sensors and spacing $d = \frac{c}{2f_0}$, and no noise is present. The number of frequency

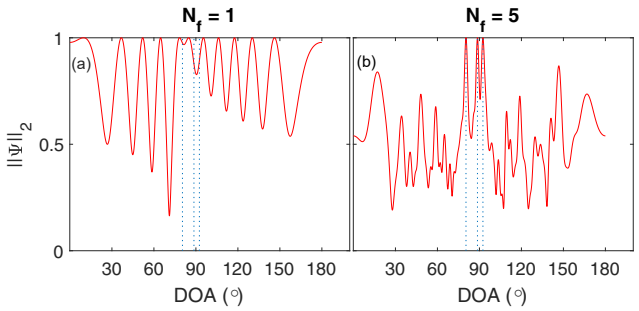

Fig. 1. The ℓ_2 norm of the dual polynomial vector for (a) $N_f = 1$ (only 100 Hz is used) and (b) $N_f = 5$ ($\{100, 200, 300, 400, 500\}$ Hz are all used). $K = 3$, $\theta = \{80.4°, 88.6°, 92.5°\}$, $N_m = 12$, $f_0 = 100$ Hz, $d = \frac{c}{2f_0}$, and $\mathbf{x}_w \sim \mathcal{CN}(0, 1)$. No noise is present.

bins $N_f = 5$ and the (temporal) frequencies of the sources are $\{100, 200, 300, 400, 500\}$ Hz. The amplitude vectors $\mathbf{x}_w$ of the 3 sources are randomly generated with standard complex normal distribution $\mathcal{CN}(0, 1)$ and then normalized so that $\|\mathbf{x}_w\|_2 = 1$. All $c_w = 1$. We solve (13) by CVX [28] and construct the dual polynomial as (7). The peaks of the norm of the dual polynomial in Fig. 1(b) exactly capture the source locations.

The recovery of the amplitude vector $\mathbf{x}_w$ is considered in Fig. 2. In (14), since $c_w = 1$, the dual polynomial evaluated in the ground-truth DOAs should exactly capture $\mathbf{x}_w$. We plot the (modulus) dual polynomial for the 3 sources under each frequency and the power for each source at the estimated DOAs. This verifies Theorem 3.2 and indicates the dual polynomial can be used to estimate $\mathbf{x}_w$.

We also study the effects of the number of frequency bins N_f, on the DOA estimation. For the above setup, we consider the ℓ_2 norm of the dual polynomial for $N_f = 1$ (only 100 Hz) and $N_f = 5$ (5 frequencies are used), respectively (see Fig. 1). When $N_f = 1$, we see both false positives and false negatives. In contrast, when $N_f = 5$, the peaks exactly capture the DOAs. When $N_f = 1$, although there is no aliasing in the data, the dual polynomial fails to capture the sources since the sources are not well-separated. When $N_f = 5$, the multi-frequency data contains aliasing, however, since this aliasing occurs at a different multiple for each frequency f, the norm of the dual polynomial vector can be used for localizing the sources. This demonstrates the benefits of multi-frequency ANM.

The root mean square error (RMSE) of the DOAs versus N_f is evaluated in Fig. 3. In this noise-free example, $N_m = 9$ and there are K = 3 or 6 sources with DOAs uniformly generated from a 1° angular grid between $[0°, 180°]$ with at least 15° separation between DOAs. c, f_0, and d are as previous example. The amplitude $\mathbf{x}_w$ has identical entries (each entry is $1/\sqrt{N_f}$). N_f varies from 1 to 6 and the frequencies are $\{1 \cdot 100, \ldots, N_f \cdot 100\}$ Hz. 100 Monte Carlo trials are averaged, and in each trial the DOAs are different. The RMSE for each trial is computed as $\sqrt{\frac{1}{K}\sum_{k=1}^{K}(\hat{\theta}_{km} - \theta_{km})^2}$, where $\hat{\theta}$, and θ are sorted estimated DOAs, and sorted ground-truth DOAs. We compare the RMSE with multi-frequency sparse Bayesian learning (SBL) [6] with 1° angular grid. Therefore, there is no grid mismatch for SBL. From Fig. 3 (b), both methods improve as N_f increases due to the presence of more data. The improvement for ANM is partially due to the coherent structure originating from the identical entries in $\mathbf{x}_w$. The identical entries in $\mathbf{x}_w$ allow the dual polynomial to satisfy (14) even for aliasing cases, while for other arbitrary $\mathbf{x}_w$, (14) is typically violated. Based on Theorem 3.2, the

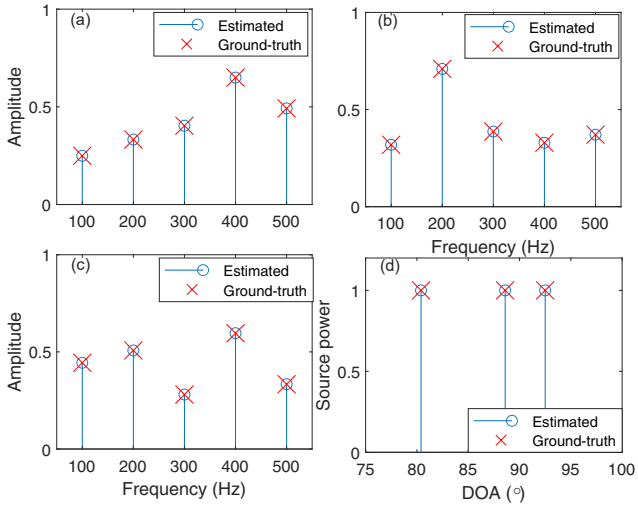

Fig. 2. (a)–(c): Recovery performance of the modulus of $\mathbf{x}_w$ for the three sources; (d) Source power estimation from the dual polynomial evaluated at the estimated DOAs. Same setup as Fig. 1 (b).

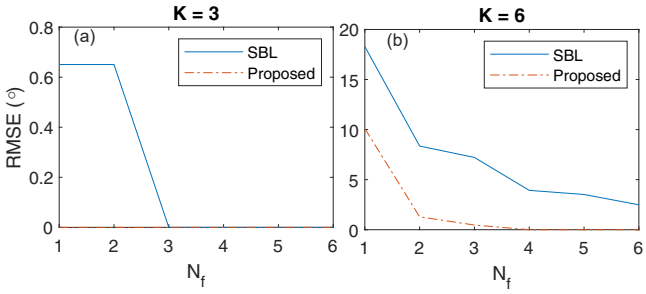

Fig. 3. RMSE (°) versus N_f for randomly generated DOAs with at least 15° separation. $K = 3$ (a) or 6 (b), $N_m = 9$, $N_f = \{1, ..., 6\}$, $MC = 100$, and $\mathbf{x}_w$ has uniform entries. c, f_0, and d are the same as Fig. 1. No noise is present.

algorithm is guaranteed to succeed if (14) is satisfied, thus ANM can perfectly resolve all sources if $\mathbf{x}_w$ has identical entries.

5. CONCLUSION

In this paper, we adapt the ANM framework to the multi-frequency model so that it can support *continuous* parameter estimation. The ANM is reformulated as an SDP problem based on the bounded real lemma so that the ANM becomes a computationally tractable problem. The dual certificate is also given, which makes it possible to localize the sources by finding the positions of peaks in the norm of the dual polynomial. Numerical results support our theory and show the effectiveness of the method. Additional analysis will be included in our future paper.

6. REFERENCES

[1] N. Antonello, E. De Sena, M. Moonen, P. A. Naylor, and T. van Waterschoot, "Joint acoustic localization and dereverberation through plane wave decomposition and sparse regularization,"

IEEE/ACM Trans. Audio, Speech, Lang. Process., vol. 27, no. 12, pp. 1893–1905, 2019.

[2] K. L. Gemba, S. Nannuru, P. Gerstoft, and W. S. Hodgkiss, "Multi-frequency sparse Bayesian learning for robust matched field processing," *J. Acoust. Soc. Am.*, vol. 141, no. 5, pp. 3411–3420, 2017.

[3] K. L. Gemba, S. Nannuru, and P. Gerstoft, "Robust ocean acoustic localization with sparse Bayesian learning," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 1, pp. 49–60, 2019.

[4] Z. Tang, G. Blacquiere, and G. Leus, "Aliasing-free wideband beamforming using sparse signal representation," *IEEE Trans. Wirel. Signal Process.*, vol. 59, no. 7, pp. 3464–3469, 2011.

[5] Z. M. Liu, Z. T. Huang, and Y. Y. Zhou, "An efficient maximum likelihood method for direction-of-arrival estimation via sparse Bayesian learning," *IEEE Trans. Wireless Comm.*, vol. 11, no. 10, pp. 1–11, 2012.

[6] S. Nannuru, K. L. Gemba, P. Gerstoft, W. S. Hodgkiss, and C. F. Mecklenbräuker, "Sparse Bayesian learning with multiple dictionaries," *Signal Process.*, vol. 159, pp. 159–170, 2019.

[7] 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.

[8] 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.

[9] 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, 2014.

[10] 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, 2015.

[11] C. Fernandez-Granda, "Super-resolution of point sources via convex programming," *Inf. Inference, J. IMA*, vol. 5, no. 3, pp. 251–303, 2016.

[12] 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.

[13] Z. Yang and L. Xie, "Exact joint sparse frequency recovery via optimization methods," *IEEE Trans. Signal Process.*, vol. 64, no. 19, pp. 5145–5157, 2016.

[14] 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.

[15] S. Li, D. Yang, G. Tang, and M. B. Wakin, "Atomic norm minimization for modal analysis from random and compressed samples," *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1817–1831, 2018.

[16] Z. Yang and L. Xie, "Enhancing sparsity and resolution via reweighted atomic norm minimization," *IEEE Trans. Signal Process.*, vol. 64, no. 4, pp. 995–1006, 2015.

[17] M. Wagner, Y. Park, and P. Gerstoft, "Gridless doa estimation and root-music for non-uniform linear arrays," *IEEE Trans. Signal Process.*, vol. 69, pp. 2144–2157, 2021.

[18] A. Xenaki and P. Gerstoft, "Grid-free compressive beamforming," *J. Acoust. Soc. Am.*, vol. 137, pp. 1923–1935, 2015.

[19] S. Li, P. Nayeri, and M. B. Wakin, "Digital beamforming robust to time-varying carrier frequency offset," *arXiv preprint arXiv:2103.04948*, 2021.

[20] S. Li, D. Gaydos, P. Nayeri, and M. B. Wakin, "Adaptive interference cancellation using atomic norm minimization and denoising," *IEEE Antennas Wirel. Propag. Lett.*, vol. 19, no. 12, pp. 2349–2353, 2020.

[21] J. Li and U. Mitra, "Improved atomic norm based time-varying multipath channel estimation," *IEEE Trans. Commun.*, 2021.

[22] 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.

[23] S. Li, M. B. Wakin, and G. Tang, "Atomic norm denoising for complex exponentials with unknown waveform modulations," *IEEE Trans. Inf. Theory*, vol. 66, no. 6, pp. 3893–3913, 2019.

[24] Y. Xie, M. B. Wakin, and G. Tang, "Simultaneous sparse recovery and blind demodulation," *IEEE Trans. Signal Process.*, vol. 67, no. 19, pp. 5184–5199, 2019.

[25] Y. Xie, M. B. Wakin, and G. Tang, "Support recovery for sparse signals with unknown non-stationary modulation," *IEEE Trans. Signal Process.*, vol. 68, pp. 1884–1896, 2020.

[26] Y. Chi and M. F. Da Costa, "Harnessing sparsity over the continuum: Atomic norm minimization for superresolution," *IEEE Signal Process. Mag.*, vol. 37, no. 2, pp. 39–57, 2020.

[27] B. Dumitrescu, *Positive trigonometric polynomials and signal processing applications*, vol. 103, Springer, 2017.

[28] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," 2014.

[29] W. H. Yang, "On generalized Hölder inequality," 1991.