

Gridless DOA Estimation with Multiple Frequencies

Yifan Wu, Michael B. Wakin, *Fellow, IEEE*, and Peter Gerstoft, *Fellow, IEEE*

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 paper, the *continuous* (gridless) DOA estimation problem with multiple frequencies is considered. This problem is formulated as an atomic norm minimization (ANM) problem. The ANM problem is equivalent to a semi-definite program (SDP) which can be solved by an off-the-shelf SDP solver. The dual certificate condition is provided to certify the optimality of the SDP solution so that the sources can be localized by finding the roots of a polynomial. We also construct the dual polynomial to satisfy the dual certificate condition and show that such a construction exists when the source amplitude has a uniform magnitude. In multi-frequency ANM, spatial aliasing of DOAs at higher frequencies can cause challenges. We discuss this issue extensively and propose a robust solution to combat aliasing. Numerical results support our theoretical findings and demonstrate the effectiveness of the proposed method.

Index Terms—Atomic norm minimization, DOA estimation, multiple frequency model, trigonometric polynomials.

I. INTRODUCTION

LINE spectrum estimation is a fundamental problem in signal processing, and has many applications in direction-of-arrival (DOA) estimation in sensor array processing [1], wideband channel estimation [2], and modern imaging modalities [3]. In line spectrum estimation, the observed signal $x[n]$ is a superposition of K complex sinusoids (i.e. $x[n] = \sum_{k=1}^K c_k e^{-j2\pi f_k n}$) and the goal is estimating the frequencies f_k of these K sinusoids. An important application of line spectrum estimation is DOA estimation [1]. For DOA estimation, we have K plane waves from angles $\{\theta_1, \dots, \theta_K\}$ impinging on an array with N_m sensors. Due to different propagation delays to each sensor, the received data is a sum of K spatial sinusoid vectors $[1 \dots e^{-j\frac{2\pi f_0(N_m-1)d \cos \theta_k}{c}}]^T$ ($k \in \{1, \dots, K\}$) parameterized by the plane wave directions θ_k (f_0 is a temporal frequency). Our goal is to estimate the K DOAs (θ_k) based on the received data. The cosine of each DOA linearly maps to a single spatial frequency $\frac{2\pi f_0 d \cos \theta_k}{c}$ of the sinusoid, and once the spatial frequencies are estimated, the DOA can be retrieved. Many line spectrum estimation methods as multiple signal classification (MUSIC) [4], and estimation of signal parameters via rotational invariant techniques (ES-PRIT) [5], have been used for narrow band signals.

Unfortunately, the aforementioned methods cannot be applied in wideband DOA estimation problems such as ocean acoustics localization and speaker localization. Wideband signal DOA estimation has been studied for decades [6]–[10].

A subspace-based wideband DOA estimation approach, incoherent signal subspace method [6], was proposed with later improvement in the coherent signal subspace method (CSSM) [7]. A broadband spatial-spectrum estimation approach [8] overcame the peak bias and source spectral content sensitivity from CSSM. Variants of CSSM, such as the weighted average of signal subspaces method [9], and the test of orthogonality of projected subspaces method [10] were also proposed. Recently, some wideband DOA estimation methods based on sparse recovery have also been developed [11]–[14]. These sparsity-based methods have demonstrated superior performance compared to conventional methods.

The multi-frequency (or multi-dictionary) model [14]–[18] has shown success in modeling wideband signals. The multi-frequency model uses N_f (rather than 1) temporal frequency bins in a frequency set $\mathcal{F} = \{f_1, \dots, f_{N_f}\}$ to characterize a wideband signal. These frequencies are then used for estimation, as opposed to using a single frequency under the narrowband model. The multi-frequency model was used for ocean acoustics localization [18]. Most of the existing methods assume that the true spatial frequencies lie on a finite set of grid points, and their performance may degrade if the true spatial frequencies 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 [19]–[32]. ANM extends grid-based, sparsity-promoting 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. ANM was initially proposed in [19], which provides a general recipe for finding convex solutions to promote sparse decompositions, where one seeks to represent a given signal based on a minimal number of atoms from an atomic set composed of an ensemble of signal atoms. The ANM framework overcomes the grid mismatch issue and can achieve potentially infinite precision. However, all prior ANM works used a narrowband assumption and are not applicable for wideband DOA estimation.

A. Related Work

1) *Multiple Frequencies*: Multiple frequencies decompose a wideband signal into multiple narrowband signals and therefore are widely applied in acoustics source localization [15], [16], [18] when the signal contains a wide range of frequency bins and cannot be characterized by a narrowband model. Some grid-based sparse localization approaches for the multiple frequencies were proposed [14], [15], [17], [18], [33] for robustness and aliasing suppression.

2) *Atomic Norm Minimization*: ANM was initially proposed in [19] as a general framework for promoting sparse signal decompositions. The pioneering ANM paper [20] worked

Yifan Wu and Peter Gerstoft are with University of California, San Diego, La Jolla, CA, 92093, USA (e-mails: {yiw062, pgerstoft}@ucsd.edu).

Michael B. Wakin is with Colorado School of Mines, Golden, CO, 80401, USA (e-mail: mwakin@mines.edu)

Manuscript received February 8, 2023

directly with the continuous (temporal) frequency estimation problem and considered the complete data case. As long as the temporal frequency separation was greater than a certain minimum separation, exact recovery of the active temporal frequencies was guaranteed. Furthermore, a semidefinite programming (SDP) framework that characterized the ANM problem was presented. The authors in [21] studied continuous temporal frequency estimation based on randomly sampled data for the single measurement vector (SMV) case. The minimum separation condition was relaxed in [24]. ANM for multiple measurement vectors (MMVs) was studied in [23], [26], [30]. In [25], the author considered a super-resolution problem that had a similar setup to [20] 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 [27]. The sample complexity of modal analysis with random temporal compression was established in [28]. ANM for 2D temporal frequency estimation was studied in [22]. In [29], the authors proposed a reweighted ANM framework, which enhances the sparsity and achieves super-resolution. An atomic norm for DOA estimation under gain-phase noise [34] was proposed to mitigate the artifacts for electromagnetic signals. ANM was also recently applied in digital beamforming [35], [36], adaptive interference cancellation [37], denoising [38], [39], and blind demodulation [40], [41]. We refer readers to [42] for a comprehensive overview of ANM and its applications.

Our multi-frequency problem is different from the MMV problems [23], [26], [30] extensively studied in the past few years. Although both our work and MMVs fall under the general topic of multi-channel line spectrum estimation, the temporal frequencies in each channel are different in our problem while they are the same in MMVs. Therefore, each channel is modulated with a different sinusoid while this *heterogeneous* modulation is absent in MMVs. This heterogeneous modulation leads to several challenges for theoretical analysis. First, it makes it difficult to derive an equivalent SDP problem based on the Vandermonde decomposition as has been done in many prior ANM works. Second, under our setup, each frequency other than the first will experience spatial aliasing of the DOAs. This leads to potential collisions or near collisions of the DOAs which are challenging to resolve. Thus, although having multiple frequencies does provide more data, one must ensure that aliasing does not undermine this benefit. These challenges make our problem more difficult to analyse than MMV problems. We will elaborate on these two challenges and our solutions in Sec. I-B.

B. Our Contributions

In this work, we extend ANM to the multi-frequency framework so that it can be used for DOA estimation with wideband signals. Our contributions are summarized as follows:

(1) Formulate an equivalent SDP problem. Although ANM itself is a convex optimization problem, it is not directly solvable due to an infinite number of optimization parameters.

Therefore, it is critical to find a computationally feasible solution that equivalently characterizes the ANM problem. Several prior works showed that certain ANM problems could be equivalently characterized by SDPs [21], [23], [26]. The derivation of an SDP problem typically relies on a Vandermonde decomposition, and equivalence with the ANM can be proved by showing that the SDP solution is both an upper and a lower bound for the ANM [21], [23], [26]. Unfortunately, this commonly used technique cannot be applied in our case due to the heterogeneous temporal frequencies across different channels. In [35], [43], certain SDPs were derived using the Vandermonde decomposition, but only the lower bound for the ANM problem could be guaranteed. In this work, we derive an equivalent SDP based on the bounded real lemma for trigonometric polynomials [44]. This equivalent SDP will provide a computationally feasible solution for the ANM when multiple frequencies are considered. We also explain how our SDP is the dual to a minor adaptation of the SDP proposed in [45] for line spectrum estimation with harmonics.

(2) Provide the dual certificate condition. We derive a dual certificate condition that can be used to certify the optimal atomic decomposition. In particular, the DOAs of the sources are localized with the help of the dual polynomial arising from the ANM optimization problem. As long as the dual polynomial satisfies the dual certificate condition, the frequencies can be localized by finding the roots of a polynomial. Therefore, the dual certificate condition not only provides a theoretical guarantee for the optimality, but also offers a method for the DOA estimation.

(3) Construct the dual polynomial that satisfies the dual certificate condition. In cases where we can prove the existence of a dual polynomial that satisfies the dual certificate condition, then the optimality and therefore exact DOA estimation are guaranteed. If the array spacing $d \leq \frac{c}{2N_f f_0} = \frac{\lambda_{N_f}}{2}$, spatial aliasing would be fully avoided for all of the temporal frequencies, and it may be possible to construct a valid dual polynomial under a mild separation assumption on the source directions. In such a case, the success of the algorithm is guaranteed.

The dual polynomial is developed our model for arbitrary spacing d . A larger aperture $(N_m - 1)d$ with greater d may improve spatial resolution but introduces spatial aliasing. If the spacing $d = \frac{c}{2f_0} = \frac{\lambda_1}{2}$, spatial aliasing is present in all but the first frequency. This spacing necessarily creates periodicity in all but the first frequency of the vector-valued dual polynomial. Such periodicity brings the risk of creating ambiguity in the source direction. More specifically, after spatial aliasing, when two source directions coincide at one frequency, we refer to this as *collision*. Collision may happen in multiple frequency bins, and it becomes more likely for great N_f . Most ANM works need well-separated harmonics to work [20], [21], [23], [26]. However, in a multi-frequency scenario, one must consider the separations for DOAs across all frequencies. Assuming collisions and near collisions are thus avoided and under some additional assumptions about the source amplitudes, we guarantee that there exists a dual polynomial satisfying the dual certificate condition.

(4) Implementation. We propose a fast implementation so that the SDP has a reduced size. This fast implementation also extends the approach to an arbitrary set of frequencies. Numerical results show that the dual polynomial still serves as a precise indicator for the DOAs. Hence, in terms of the DOA estimation, the algorithm succeeds even when collisions are present.

Finally, our work is inspired by recent advances in ANM for super-resolution, but significantly deviates from the existing MMV works. This work significantly extends our previous ICASSP paper [46]. It includes additional analysis for the dual polynomial construction, aliasing and collision, and provides a fast algorithm and extensive simulations. This paper is the first work that extends ANM to multiple frequencies so that it can be adapted to *gridless* DOA estimation for wideband signals via convex programming.

C. Notations and Organization

Boldface letters represents matrices and vectors. Conventional notations $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^*$, $\langle \cdot \rangle_{\mathbb{R}}$, and $\langle \cdot \rangle$ stand for matrix/vector transpose, Hermitian transpose, complex conjugate, real inner product, and inner product, respectively. $\text{Tr}(\cdot)$ is used to represent the trace of a matrix. $\|\cdot\|_p$ and $\|\cdot\|_F$ are used to express vector ℓ_p norm and matrix Frobenius norm. For a Hermitian matrix $\mathbf{A}$, $\mathbf{A} \succeq 0$ means $\mathbf{A}$ is a positive semidefinite (PSD) matrix. $\odot$ stands for the Hadamard product. The $\ell_{1,2}$ norm of a matrix $\mathbf{A} = [\mathbf{a}_1 \dots \mathbf{a}_N]$ is defined as $\|\mathbf{A}\|_{1,2} := \sum_{i=1}^N \|\mathbf{a}_i\|_2$. The imaginary unit is denoted by $j = \sqrt{-1}$.

The rest of the paper is organized as follows. Sec. II introduces the signal model and the assumptions. The equivalent SDP and the dual certificate condition are derived in Sec. III. Sec. IV constructs the dual polynomial that satisfies the dual certificate condition and also analyses the collision and near collision issues. Sec. V presents some numerical examples to support and demonstrate theoretical findings. Finally, Sec. VI concludes the paper.

II. SIGNAL MODEL

A. Assumptions and Model Framework

1) Assumptions: The following assumptions are made for the array configuration and 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 an integer multiple f of a fundamental frequency f_0 , i.e., $f \in \{1, \dots, N_f\}$ and $ff_0 \in \{f_0, \dots, N_f f_0\}$. This is only a technical assumption to simplify the analysis, and our method can be applied in any frequency set with the fast algorithm proposed in Sec. III-F.
- 4) Suppose $d \leq \frac{c}{2f_0}$ holds (or, equivalently, $\frac{2\pi f_0 d}{c} \leq \pi$), where c is the speed of propagation. We also notice that $d = \frac{c}{2f_0}$ is the maximum separation to avoid

spatial aliasing at the fundamental frequency. For higher frequencies (i.e. $f \geq 2$), aliasing will still exist. Such aliasing is not considered in conventional narrowband ANM papers. It is possible to develop the method with $d = \frac{c}{2N_f f_0}$ so that aliasing can be completely avoided in all frequencies.

2) Multiple Frequencies: Based on the above assumptions, we absorb the constant parameters d , f_0 , and c into a scaled DOA parameter $w = w(\theta) := \frac{f_0 d \cos(\theta)}{c} \in [-f_0 d/c, f_0 d/c]$. Henceforth, w is simply referred as the DOA.

For each temporal frequency $ff_0 \in \{1, \dots, N_f\} \cdot f_0$, let $\mathbf{y}_f \in \mathbb{C}^{N_m}$ denote the received signal across the N_m sensors. $\mathbf{y}_f$ can be expressed as a sum of K spatial sinusoid vectors, with the k -th vector having spatial frequency $f w(\theta_k)$. Importantly, the spatial frequency depends on both the temporal frequency ff_0 and the DOA $w(\theta_k)$. To better illustrate these effects, we refer the reader to Fig. 1. Suppose $N_f = 3$, $N_m = 5$, and the input signal (top row) is a complex sinusoid with temporal frequency ff_0 . The spatial samples obtained from the sensors (red) will be sampled sinusoids (bottom row) with different spatial frequencies that depend on both the temporal frequency and the DOA.

Stacking all of the data from the N_f frequencies into a matrix, the full set of received data is denoted by $\mathbf{Y} := [\mathbf{y}_1 \dots \mathbf{y}_{N_f}] \in \mathbb{C}^{N_m \times N_f}$. Summing over the K active DOAs, we write

$$\mathbf{Y} = \mathbf{X} + \mathbf{W}, \quad (1)$$

where

$$\begin{aligned} \mathbf{X} &:= \sum_w c_w [x_w(1)\mathbf{a}(1, w) \dots x_w(N_f)\mathbf{a}(N_f, w)] \\ &= \sum_w c_w \mathbf{A}(w) \otimes \mathbf{x}_w^T, \end{aligned} \quad (2)$$

$\mathbf{a}(f, w) := [1 \dots e^{-j2\pi f w(N_m-1)}]^T \in \mathbb{C}^{N_m}$ is the array manifold vector (steering vector) corresponding to the f -th frequency bin and DOA w , $x_w(f)$ is the signal amplitude for the f -th frequency bin, and $\mathbf{W} := [\mathbf{w}_1, \dots, \mathbf{w}_{N_f}] \in \mathbb{C}^{N_m \times N_f}$ is additive Gaussian uncorrelated noise. $\mathbf{x}_w := [x_w(1) \dots x_w(N_f)]^T \in \mathbb{C}^{N_f}$ is a collection of N_f amplitudes corresponding to the same DOA, $\mathbf{A}(w) := [\mathbf{a}(1, w) \dots \mathbf{a}(N_f, w)] \in \mathbb{C}^{N_m \times N_f}$, and $\otimes$ is the Khatri-Rao product defined as $[\mathbf{A}(w) \otimes \mathbf{x}_w^T] := [\mathbf{a}(1, w)x_w(1) \dots \mathbf{a}(N_f, w)x_w(N_f)] \in \mathbb{C}^{N_m \times N_f}$. We assume that $\|\mathbf{x}_w\|_2 = 1$; the coefficient c_w absorbs any other scaling of the source amplitudes $c_w \mathbf{x}_w$. Our goal is to identify the K active DOAs w from the data matrix $\mathbf{Y}$.

In the following sections, we primarily develop the optimization methods within the noise-free model, i.e., where $\mathbf{W} = \mathbf{0}$. We describe, however, how the optimization problem is modified if noise is present (see (10) and (21)).

B. Mapping Operator

In this section, we will define some mapping operators that help us set up our method. Define

$$\begin{aligned} \mathbf{z} = \mathbf{z}(w) &:= [1 \ e^{-j2\pi w} \dots e^{-j2\pi w N_f(N_m-1)}]^T \\ &= [z^0 \ z^1 \dots z^{N-1}]^T \in \mathbb{C}^N \end{aligned} \quad (3)$$

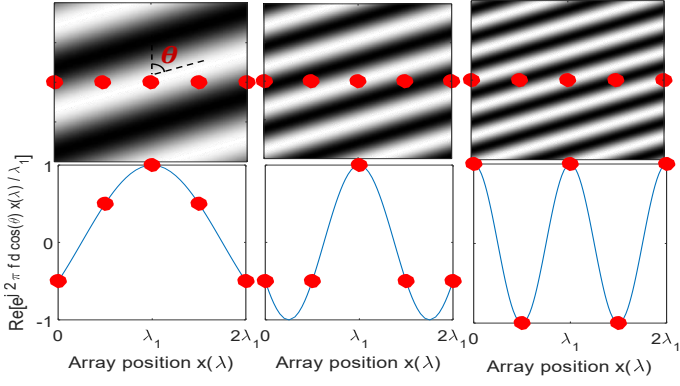


Fig. 1. Multi-frequency data on array with $N_m = 5$ sensors. Top row: time snapshot of propagating plane wave with angle of arrival θ and temporal frequency (left to right) $f_0, 2f_0, 3f_0$. Bottom row: array data are samples of a spatial sinusoid whose spatial frequency depends on the temporal frequency and DOA. Only the real part of the array data is shown.

that collects all possible complex exponentials from the array manifolds in all frequencies, where $N := N_f(N_m - 1) + 1$ and $z = z(w) := e^{-j2\pi w}$. The intuition for defining the $\mathbf{z}$ notation will be explained after the definition of the dual polynomial vector (13). Introduce $\mathbf{Z} = \mathbf{Z}(w) := [\mathbf{z} \dots \mathbf{z}] \in \mathbb{C}^{N \times N_f}$ and define $\mathbf{X}' \in \mathbb{C}^{N \times N_f}$ as

$$\mathbf{X}' := \left(\sum_w c_w \mathbf{Z} \otimes \mathbf{x}_w^T \right). \quad (4)$$

Then, we define the $\mathcal{R}$ operator that maps $\mathbf{X}'$ to $\mathbf{X}$ as

$$\mathbf{X} = \mathcal{R}(\mathbf{X}') \Rightarrow \mathbf{X}(i, j) = \mathbf{X}'(1 + (i - 1)j, j), \quad (5)$$

where $\mathcal{R} : N \times N_f \rightarrow N_m \times N_f$ is a mapping that selects N_m elements from the N elements in each column of $\mathbf{X}'$. We demonstrate the mapping in Fig. 2. Note that $\mathbf{A}(w)$ can be represented in terms of $\mathbf{Z}$ by using the $\mathcal{R}$ operator as

$$\mathbf{A}(w) = \mathcal{R}(\mathbf{Z}). \quad (6)$$

Note that in [45, (8)], an analogous mapping operator is introduced in the context of the line spectrum estimation problem with harmonics. In [45, (9)], the transformation is applied in the signal space and enables the formulation of an SDP problem in the primal domain. In our paper, $\mathcal{R}$ is applied to the coefficient matrix $\mathbf{H}$ (see (15)) and that enables us to formulate an SDP problem in the dual domain.

III. METHODOLOGY

A. Atomic Norm Minimization (ANM)

To efficiently represent matrices of the form (2), we define the atomic set

$$\mathcal{A} := \{\mathbf{A}(w) \otimes \mathbf{x}_w^T : w \in [-f_0 d/c, f_0 d/c], \|\mathbf{x}_w\|_2 = 1\}. \quad (7)$$

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 $\mathcal{A}$ is parameterized by the continuous DOA w .

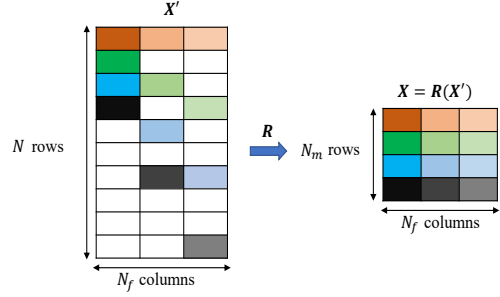


Fig. 2. Compaction of matrix $\mathbf{X}'$ to $\mathbf{X}$ by mapping $\mathcal{R} : N \times N_f \rightarrow N_m \times N_f$ defined in (5).

In the noise-free case, to identify the K active DOAs w from the data matrix $\mathbf{Y}$, we propose the following ANM-based optimization framework:

$$\min_{\mathbf{X}} \|\mathbf{X}\|_{\mathcal{A}} \quad \text{s.t.} \quad \mathbf{Y} = \mathbf{X}, \quad (8)$$

where the atomic norm is defined as

$$\begin{aligned} \|\mathbf{X}\|_{\mathcal{A}} &:= \inf \{t \geq 0 \mid \mathbf{X} \in t \cdot \text{conv}(\mathcal{A})\} \\ &= \inf \left\{ \sum_w |c_w| \mid \mathbf{X} = \sum_w c_w \mathbf{A}(w) \otimes \mathbf{x}_w^T \right\}. \end{aligned} \quad (9)$$

When noise is present, we modify the optimization problem to relax the equality constraint:

$$\min_{\mathbf{X}} \|\mathbf{X}\|_{\mathcal{A}} \quad \text{s.t.} \quad \|\mathbf{Y} - \mathbf{X}\|_F \leq \eta, \quad (10)$$

where η depends on the noise level.

It is not obvious how to obtain DOAs directly based on (8) (and (10)), as one of the solutions is $\mathbf{Y}$ itself. In the following sections, we develop an equivalent optimization problem for computing the atomic decomposition of $\mathbf{Y}$, which enables determining the DOAs via the dual polynomial.

B. Dual Atomic Norm and Dual Polynomial

Let $\|\mathbf{X}\|$ be a matrix norm. The associate *dual norm*, denoted $\|\mathbf{Q}\|^*$, is defined as [47, Appendix A.1.6],

$$\|\mathbf{Q}\|^* := \sup_{\|\mathbf{X}\| \leq 1} \langle \mathbf{Q}, \mathbf{X} \rangle_{\mathbb{R}}. \quad (11)$$

Also note that the dual of the dual norm is the primal norm.

Now we apply (11) to the atomic norm. The primal atomic norm $\|\mathbf{X}\|_{\mathcal{A}}$ is expressed in terms of the dual atomic norm $\|\mathbf{Q}\|_{\mathcal{A}}^*$ (where $\mathbf{Q} := [\mathbf{q}_1 \dots \mathbf{q}_{N_f}] \in \mathbb{C}^{N_m \times N_f}$ is the dual variable) as

$$\|\mathbf{X}\|_{\mathcal{A}} := \sup_{\|\mathbf{Q}\|_{\mathcal{A}}^* \leq 1} \langle \mathbf{Q}, \mathbf{X} \rangle_{\mathbb{R}} = \sup_{\|\mathbf{Q}\|_{\mathcal{A}}^* \leq 1} \langle \mathbf{Q}, \mathbf{Y} \rangle_{\mathbb{R}}, \quad (12)$$

where the last equality is only for the noise-free case (see the constraint in (8)).

For any dual variable $\mathbf{Q}$, we define the corresponding *dual polynomial vector* $\psi(\mathbf{Q}, w) \in \mathbb{C}^{N_f}$ as

$$\begin{aligned} \psi(\mathbf{Q}, w) &:= [\mathbf{q}_1^H \mathbf{a}(1, w) \dots \mathbf{q}_{N_f}^H \mathbf{a}(N_f, w)]^T \\ &= \left[\sum_{m=1}^{N_m} q_1^*(m) z^{(m-1)} \dots \sum_{m=1}^{N_m} q_{N_f}^*(m) z^{N_f \cdot (m-1)} \right]^T. \end{aligned} \quad (13)$$

Note that each entry in $\psi(\mathbf{Q}, w)$ is a polynomial in z . The dual polynomial will be useful for setting up the dual certificate condition and extracting the DOA (see Sec. III-D and Sec. III-E). However, since each frequency has different array manifold vectors, it is difficult to express $\psi(\mathbf{Q}, w)$ as a matrix product of $\mathbf{Q}$ and a vector. To construct a homogeneous representation for $\psi(\mathbf{Q}, w)$, we will leverage $\mathbf{z}$, an ensemble of the array manifold, and the matrix $\mathbf{H} \in \mathbb{C}^{N \times N_f}$ defined in terms of $\mathbf{Q}$ as follows ($m = \{1, \dots, N_m\}$, $f = \{1, \dots, N_f\}$)

$$\mathbf{H}(i, f) = \begin{cases} \mathbf{Q}(m, f) & \text{for } (i, f) = (f \cdot (m-1) + 1, f) \\ 0 & \text{otherwise,} \end{cases} \quad (14)$$

or $\mathbf{H} = \mathcal{R}^*(\mathbf{Q})$, where $\mathcal{R}^* : N_m \times N_f \rightarrow N \times N_f$ is the adjoint mapping of $\mathcal{R}$. Note the relationship between $\mathbf{Q}$ and $\mathbf{H}$ can be alternatively expressed as

$$\mathbf{Q} = \mathcal{R}(\mathbf{H}). \quad (15)$$

With the help of $\mathbf{H}$ and $\mathbf{z}$, $\psi(\mathbf{Q}, w)$ has the homogeneous representation

$$\psi(\mathbf{Q}, w) = \mathbf{H}^H \mathbf{z}. \quad (16)$$

Now, we consider $\|\mathbf{Q}\|_{\mathcal{A}}^*$, which appears in a constraint in (12). Recalling that $\|\mathbf{x}_w\|_2 = 1$, we have

$$\begin{aligned} \|\mathbf{Q}\|_{\mathcal{A}}^* &:= \sup_{\|\mathbf{X}\|_{\mathcal{A}} \leq 1} \langle \mathbf{Q}, \mathbf{X} \rangle_{\mathbb{R}} = \sup_{\|\mathbf{X}\|_{\mathcal{A}} \leq 1} \text{Re}[\text{Tr}(\mathbf{Q}^H \mathbf{X})] \\ &= \sup_{\mathbf{x}_w} \text{Re}[\text{Tr}(\mathbf{Q}^H \mathbf{A}(w) \otimes \mathbf{x}_w^T)] \\ &= \sup_{x_w(f)} \text{Re} \left(\sum_{f=1}^{N_f} x_w(f) \mathbf{q}_f^H \mathbf{a}(f, w) \right) \\ &\stackrel{(a)}{=} \sup_{\mathbf{x}_w} \text{Re}(\mathbf{x}_w^H \psi(\mathbf{Q}, w)) = \sup_{\mathbf{x}_w} |\mathbf{x}_w^H \psi(\mathbf{Q}, w)| \\ &\stackrel{(b)}{=} \sup_w \|\psi(\mathbf{Q}, w)\|_2 = \sup_w \|\mathbf{H}^H \mathbf{z}\|_2 \end{aligned} \quad (17)$$

where (a) follows by the definition of the dual polynomial vector and (b) follows from the definition of the operator norm.

Using (17), the condition $\|\mathbf{Q}\|_{\mathcal{A}}^* \leq 1$ can be equivalently formulated as an SDP constraint. To simplify the theoretical analysis, we assume $d = \frac{c}{2f_0}$ and thus $w \in [-1/2, 1/2]$ here. We however notice that the “if” part can be generalized to any $d \leq \frac{c}{2f_0}$.

Proposition 3.1: Let $\psi(\mathbf{Q}, w)$ be as defined in (13) and $w \in [-1/2, 1/2]$. Then $\|\mathbf{Q}\|_{\mathcal{A}}^* \leq 1$ holds if and only if there exists a matrix $\mathbf{P}_0 \in \mathbb{C}^{N \times N} \succeq 0$ such that

$$\sum_{i=1}^{N-k} \mathbf{P}_0(i, i+k) = \delta_k = \begin{cases} 1, & k=0, \\ 0, & k=1, \dots, N-1, \end{cases} \quad (18)$$

and such that

$$\begin{bmatrix} \mathbf{P}_0 & \mathbf{H} \\ \mathbf{H}^H & \mathbf{I}_{N_f} \end{bmatrix} \succeq 0. \quad (19)$$

Proof See Appendix A. $\square$

C. SDP Formulations of ANM Problems

1) *Noise-free ANM:* In the noise-free case, based on Proposition 3.1 and (12), we have an SDP that is equivalent to (8):

$$\begin{aligned} \max_{\mathbf{Q}, \mathbf{P}_0} \langle \mathbf{Q}, \mathbf{Y} \rangle_{\mathbb{R}} \quad \text{s.t.} \quad & \begin{bmatrix} \mathbf{P}_0 & \mathbf{H} \\ \mathbf{H}^H & \mathbf{I}_{N_f} \end{bmatrix} \succeq 0, \\ & \sum_{i=1}^{N-k} \mathbf{P}_0(i, i+k) = \delta_k, \mathbf{H} = \mathcal{R}^*(\mathbf{Q}), \end{aligned} \quad (20)$$

where the dual variable $\mathbf{Q} \in \mathbb{C}^{N_m \times N_f}$, and $\mathbf{H}$ is related to $\mathbf{Q}$ as in (14).

2) *Robust ANM:* To make ANM robust to noise and near collisions (see (50)), we use the following alternative to (20):

$$\begin{aligned} \max_{\mathbf{Q}, \mathbf{P}_0} \langle \mathbf{Q}, \mathbf{Y} \rangle_{\mathbb{R}} - \eta \|\mathbf{Q}\|_F - \lambda \|\mathbf{Q}\|_{1,2} \quad \text{s.t.} \quad & \begin{bmatrix} \mathbf{P}_0 & \mathbf{H} \\ \mathbf{H}^H & \mathbf{I}_{N_f} \end{bmatrix} \succeq 0, \\ & \sum_{i=1}^{N-k} \mathbf{P}_0(i, i+k) = \delta_k, \mathbf{H} = \mathcal{R}^*(\mathbf{Q}), \end{aligned} \quad (21)$$

where the term $\eta \|\mathbf{Q}\|_F$ suppresses noise [25, (15)] [36, (34), and App. D]. The value of η is the same as in (10) [25, [36]. Based on similar arguments to [36, App. D], (21) with $\lambda = 0$ is the dual problem of (10). We further add an $\ell_{1,2}$ regularization term to suppress near collisions. The $\ell_{1,2}$ regularization term $\lambda \|\mathbf{Q}\|_{1,2}$ promotes column sparsity, and it reduces the contributions from the “bad frequencies”. Near collision is a phenomenon that arises in our multi-frequency ANM model, and it is introduced in Sec. IV-D1. For the noise-free data, one may set $\eta = 0$, and for the near-collisions-free data, one may set $\lambda = 0$.

D. Dual Certificate

The dual polynomial $\psi(\mathbf{Q}, w)$ introduced in (13) serves as a certificate for the optimality of (8) and can therefore be used to extract the unknown DOAs. Specifically, we have the following dual certificate theorem, which is inspired by [21, Proposition II.4]. To ensure uniqueness, a linear independence assumption is added.

Theorem 3.2: Define $\mathcal{W} := \{w_1, \dots, w_K\}$ as a collection of DOAs with cardinality K . Then $\mathbf{X} = \sum_{w \in \mathcal{W}} c_w \mathbf{A}(w) \otimes \mathbf{x}_w^T$ ($\|\mathbf{x}_w\|_2 = 1$) 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 vector $\psi(\mathbf{Q}, w)$ satisfies

$$\begin{cases} \psi(\mathbf{Q}, w) = \text{sign}(c_w^*) \mathbf{x}_w & \forall w \in \mathcal{W} \\ \|\psi(\mathbf{Q}, w)\|_2 < 1 & \forall w \notin \mathcal{W}, \end{cases} \quad (22)$$

where $\text{sign}(c_w^*) := \frac{c_w^*}{|c_w^*|}$.

(2) $\{\mathbf{A}(w) \otimes \mathbf{x}_w^T : w \in \mathcal{W}\}$ is a linearly independent set.

Proof See Appendix B. $\square$

E. DOA Extraction

Based on Theorem 3.2, we know if (22) is satisfied, the optimality is guaranteed. In (22), $\|\psi(\mathbf{Q}, w)\|_2 = 1$ for $w \in \mathcal{W}$. After solving the SDPs (20)–(21) by CVX [48], the optimal

dual variables $\mathbf{Q}$ (and thus $\mathbf{H}$) are obtained. Then, the $\mathbf{D}$ is retrieved by finding the roots for $R(w)$ defined in (57).

Based on (16), $R(z)$ has the polynomial representation

$$R(z) = 1 - \mathbf{z}^H \mathbf{P}_1 \mathbf{z} = 1 - \sum_{i=-(N-1)}^{(N-1)} r_i z^i, \quad (6)$$

where $\mathbf{P}_1 := \mathbf{H}\mathbf{H}^H$ and $r_k := \sum_{i=1}^{N-k} \mathbf{P}_1(i, i+k)$. Indeed $R(w)$ is a polynomial with degree $2(N-1)$. The roots $\hat{z}$ be obtained, and $\hat{w}$ is retrieved by locating the roots of R on the unit circle (see Fig. 3 (c)):

$$\hat{w} = \left\{ -\frac{\angle \hat{z}}{2\pi} \middle| R(\hat{z}) = 0, |\hat{z}| = 1 \right\}. \quad (7)$$

Note $\angle \hat{z} = -2\pi \hat{w} = -\frac{2\pi f_0 d}{c} \cos \theta = \frac{2\pi f_0 d}{c} \cos(\pi - \theta)$. θ is therefore estimated by

$$\hat{\theta} = \pi - \cos^{-1} \left(\frac{\angle \hat{z}}{2\pi f_0 d / c} \right). \quad (8)$$

The implementation details for the proposed algorithm summarized in Algorithm 1.

Algorithm 1 Gridless DOA estimation algorithm

Input: $\mathbf{Y} \in \mathbb{C}^{N_m \times N_f}$, d, f_0, c, K, η (for noisy case), λ (for near collision case)

Initialization:

(For noisy or near collision case) Solve (21) by CVX to obtain $\mathbf{H}$

(Otherwise) Solve (20) by CVX and obtain $\mathbf{H}$ from $\mathbf{Q}$

$\mathbf{P}_1 \leftarrow \mathbf{H}\mathbf{H}^H$

$N \leftarrow N_f(N_m - 1) + 1$

while $-(N-1) \leq k \leq (N-1)$ **do**

$r_k \leftarrow \sum_{i=1}^{N-k} \mathbf{P}_1(i, i+k)$

end while

$\mathbf{r} \leftarrow [-r_{-(N-1)} \cdots -r_{(N-1)}]$

$\mathbf{r}(N) \leftarrow \mathbf{r}(N) + 1$

$\mathbf{roots} \leftarrow \text{roots}(\mathbf{r})$

$[\text{dist}, \text{ind}] \leftarrow \text{sort}(\text{abs}(1 - \text{abs}(\mathbf{roots})))$

$\mathbf{roots_sort} \leftarrow \mathbf{roots}(\text{ind})$

$\mathbf{roots_unique} \leftarrow \mathbf{roots_sort}(1 : 2 : 2K)$

$\hat{\theta} \leftarrow 180 - \text{acosd}(\text{angle}(\mathbf{roots_unique}) / (f_0 d / c))$

Output: $\hat{\theta}$

F. Fast Algorithm

We notice that many rows in the matrix $\mathbf{H}$ are all zero, yet they contribute to the size of the SDP constraint in (19). This inspires us to come up with a fast algorithm which only includes the non-zero rows of $\mathbf{H}$ in the SDP constraint. This fast algorithm generalizes the method to any frequency set.

In particular, consider a frequency set $\mathcal{F} = \{F_1, \dots, F_{N_f}\} \cdot f_0$ with integers $F_1, \dots, F_{N_f}$ and define $\mathcal{U} = \{m \cdot f | m \in \{0, \dots, N_m - 1\}, f \in \{F_1, \dots, F_{N_f}\}\}$ with cardinality N_u . The ratio of N/N_u in Fig. 4(a) shows a factor of 2 in savings for large N_m and N_f which gives up to a factor of 30 savings in CPU time (Fig. 4 (b)). Assume the entries in $\mathcal{U}$ are sorted in ascending order. The matrix

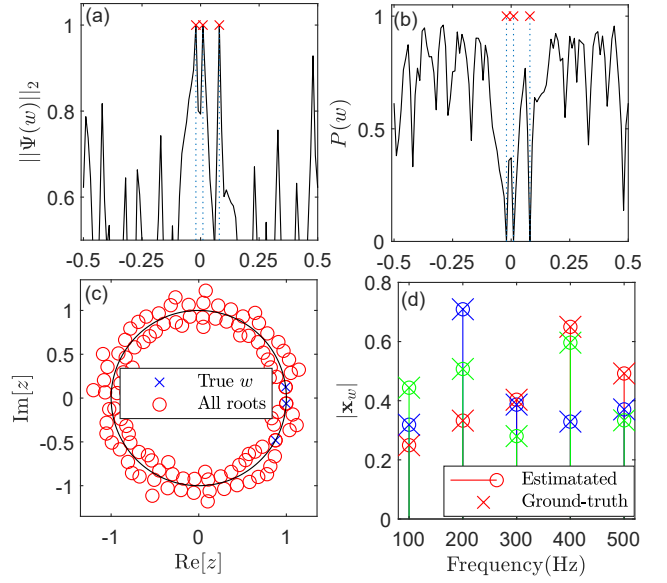


Fig. 3. DOA extraction through the dual polynomial. An ULA with $N_m = 12$ sensors, and spacing $d = c/2f_0$ is used. $N_f = 5$. $\theta = [80.7931^\circ, 88.854^\circ, 92.2924^\circ]$, and $w = [0.08, 0.01, -0.02]$. (a) $\|\psi(\mathbf{Q}, w)\|_2$ versus w ; (b) $P(w)$ versus w ; (c) Roots for $P(w)$; (d) Amplitude estimation for each frequency (three colors are used to indicate different sources).

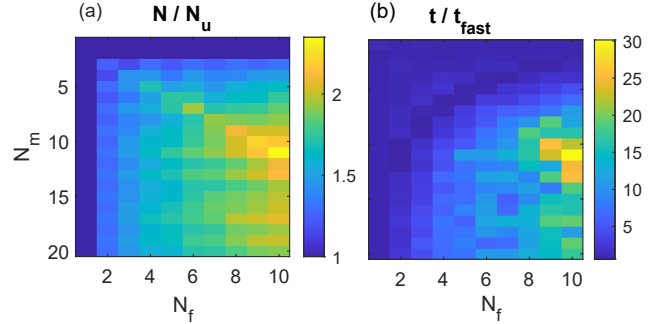


Fig. 4. (a) N/N_u ; (b) t/t_{fast} , where t and t_{fast} are CPU times for (20) and the fast program, respectively.

$\mathbf{H}_r \in \mathbb{C}^{N_u \times N_f}$ with a reduced number of rows can be expressed in terms of $\mathbf{Q}$ as

$$\mathbf{H}_r(r, f) = \begin{cases} \mathbf{Q}(m, f) & \text{for } (\mathcal{U}_r, f) = (f \cdot (m-1) + 1, f) \\ 0 & \text{otherwise;} \end{cases} \quad (26)$$

note r is the index of $\mathcal{U}_r = f \cdot (m-1) + 1$. We have the following proposition for an SDP with reduced dimension.

Proposition 3.3: Let $\psi(\mathbf{Q}, w)$ be as defined in (13). Then $\|\mathbf{Q}\|_{\mathcal{A}}^* \leq 1$ holds if there exists a matrix $\mathbf{P}_{r0} \in \mathbb{C}^{N_u \times N_u} \succeq 0$ such that

$$\sum_{\substack{i,j \\ \mathcal{U}_i - \mathcal{U}_j = k}} \mathbf{P}_{r0}(i, j) = \delta_k = \begin{cases} 1, & k = 0, \\ 0, & k = 1, \dots, N-1 \end{cases} \quad (27)$$

and such that

$$\begin{bmatrix} \mathbf{P}_{r0} & \mathbf{H}_r \\ \mathbf{H}_r^H & \mathbf{I}_{N_f} \end{bmatrix} \succeq 0, \quad (28)$$

where $\mathbf{H}_r$ is defined in (26).

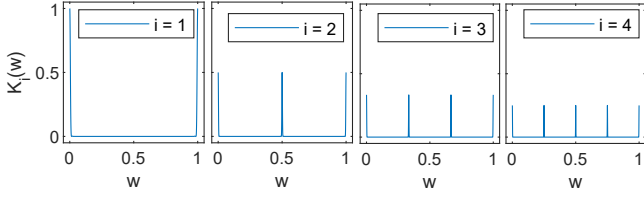


Fig. 5. Visualization of $K_i(w)$ for $i \in \{1, 2, 3, 4\}$.

The proof is in the Appendix C. $\square$

We therefore propose fast alternatives to (20) and (21) by incorporating the reduced dimension SDP constraint. Note that in Proposition 3.1, we theoretically guaranteed the equivalence between (20) and (8). However, we only guarantee the “if” part in Proposition 3.3. Nevertheless, it turns out that the fast algorithm achieves promising performance in the empirical experiments while greatly reducing the computational complexity. The empirical improvement in computational complexity is up to a factor of 30 (see Fig. 4 (b)). We apply the fast algorithms throughout Sec. V.

G. Dual SDP Problem

Based on [45, (16)], we consider the dual problem of (20). The dual problem of (20) is also an SDP, and it can be expressed as

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{u}, \tilde{\mathbf{Y}}} & [\text{Tr}(\mathbf{W}) + \text{Tr}(\text{Toep}(\mathbf{u}))] \\ \text{s.t.} & \begin{bmatrix} \text{Toep}(\mathbf{u}) & \tilde{\mathbf{Y}} \\ \tilde{\mathbf{Y}}^H & \mathbf{W} \end{bmatrix} \succeq 0, \mathbf{Y} = \mathcal{R}(\tilde{\mathbf{Y}}), \end{aligned} \quad (29)$$

where $\mathbf{W} \in \mathbb{C}^{N_f \times N_f}$, $\mathbf{u} \in \mathbb{C}^N$, $\tilde{\mathbf{Y}} \in \mathbb{C}^{N \times N_f}$, and $\text{Toep}(\mathbf{u})$ is a $N \times N$ Toeplitz matrix with the first column $\mathbf{u}$.

The derivation of the dual problem is provided in App. D. After solving (29), the DOAs are retrieved using the Vandermonde decomposition of $\text{Toep}(\mathbf{u})$ [45] and the root-MUSIC procedure. Since both (20) and (29) are strictly feasible, strong duality holds. Therefore, the optimal values for (20) and (29) must be the same. Note the matrix associated with the PSD constraint for both problems are $(N + N_f) \times (N + N_f)$. We can solve either one of them for DOA estimation.

IV. DUAL POLYNOMIAL CONSTRUCTION

In Theorem 3.2, a *sufficient* condition for optimal atomic decomposition was given. In this section, for certain scenarios, we show that it is possible to construct a dual certificate satisfying (22). This implies the success of the DOA estimation algorithm in the noise-free setting.

Following from [20], we consider an alternative, symmetric index set $\{-2M, \dots, 2M\}$ (modified from $\{0, \dots, N_m - 1\}$), where $M = \frac{N_m - 1}{4}$. Constructing a dual certificate satisfying the requisite properties (22) using the original index set is equivalent to constructing a “modulated” dual polynomial $\psi(w)$ (note that $\psi(w)$ is different from the $\psi(\mathbf{Q}, w)$ defined in Sec. III) on the symmetric index set satisfying

$$\begin{cases} \psi(w) = \text{sign}(c_w^*) \bar{\mathbf{x}}_w & \forall w \in \mathcal{W} \\ \|\psi(w)\|_2 < 1 & \forall w \notin \mathcal{W}, \end{cases} \quad (30)$$

where $\bar{\mathbf{x}}_w(i) := \mathbf{x}_w(i) \cdot e^{-j2\pi w i \frac{N_m - 1}{2}}$, $\forall i \in \{1, \dots, N_f\}$. Note $|\bar{\mathbf{x}}_w(i)| = |\mathbf{x}_w(i)| \cdot e^{-j2\pi w i \frac{N_m - 1}{2}} = |\mathbf{x}_w(i)|$, and $|\psi(w)(i)| \cdot e^{j2\pi w i \frac{N_m - 1}{2}} = |\psi(w)(i)|$. Therefore, as long as $\psi(w)$ (associated with the new index set $\{-2M, \dots, 2M\}$) satisfies (30), $\bar{\psi}(w) := \psi(w) \odot [e^{j2\pi w \frac{N_m - 1}{2}} \dots e^{j2\pi w N_f \frac{N_m - 1}{2}}]^T$ (associated with the original index set) must satisfy (22). Indeed, $\|\psi(w)\|_2 = \|\bar{\psi}(w)\|_2$ and $\bar{\psi}(w) = \text{sign}(c_w^*) \mathbf{x}_w$ for $\forall w \in \mathcal{W}$. In this section, we will construct $\psi(w)$ that satisfies (30).

In addition, $w \in [0, 1)$ is assumed in this section. Due to the periodicity of the kernel, it is equivalent to consider $w \in [-1/2, 1/2]$ as $w \in [0, 1)$. This assumption indicates that $d = \frac{c}{2f_0}$ needs to be assumed for the following analysis.

A. Interpolation Kernel

Inspired by [20], we leverage the i -th order squared Fejér kernel $K_i(w)$ for the dual polynomial construction:

$$\begin{aligned} K_i(w) &:= \frac{1}{iM} \sum_{k=-2M}^{2M} g_M(k) e^{-j2\pi k w \cdot i} \\ &= \frac{1}{i} \left[\frac{\sin(\pi(M+1)wi)}{(M+1)\sin(\pi wi)} \right]^4, \end{aligned} \quad (31)$$

where

$$g_M(k) = \frac{1}{M} \sum_{t=\max\{k-M, -M\}}^{\min\{k+M, M\}} \left(1 - \frac{|t|}{M}\right) \left(1 - \frac{|k-t|}{M}\right). \quad (32)$$

$K_i(w)$, $i \in \{1, 2, 3, 4\}$ is shown in Fig. 5. When $i = 1$, $K_i(w)$ corresponds to the classical kernel used for the dual polynomial construction in [20], [21], [23], [26], [30]. When i increases, the period of the kernel reduces to $1/i$. Therefore, the periodic copies appears in the visible region $[0, 1)$, and will potentially bring about aliasing for the localization. In addition, note that the amplitude of $K_i(w)$ shrinks to $1/i$, which will cancel the scaling factor i of $K_i'(w)$.

We summarize some useful facts for $K_i(w)$

$$K_i(w) = \frac{1}{i} K_1(iw) \quad K_i'(w) = K_1'(iw) \quad K_i''(w) = i K_1''(iw). \quad (33)$$

B. Dual Polynomial Construction by Interpolation Kernel

We construct the dual polynomial vector $\psi(w) \in \mathbb{C}^{N_f}$ as follows

$$\psi(w) := \begin{bmatrix} \sum_{w_k \in \mathcal{W}} [\alpha_{k,1} K_1(w - w_k) + \beta_{k,1} K_1'(w - w_k)] \\ \vdots \\ \sum_{w_k \in \mathcal{W}} [\alpha_{k,N_f} K_{N_f}(w - w_k) + \beta_{k,N_f} K_{N_f}'(w - w_k)] \end{bmatrix}, \quad (34)$$

where $K_i'(w - w_k)$ is the first order derivative for $K_i(w - w_k)$.

The constructed dual polynomial in (34) is valid if there exists $\alpha_{k,i}$ and $\beta_{k,i}$ ($i = 1, \dots, N_f$) that satisfy (22). To satisfy (22), for each frequency, we must have [20]

$$\begin{bmatrix} \mathbf{D}_{i,0} & \mathbf{D}_{i,1} \\ \mathbf{D}_{i,1} & \mathbf{D}_{i,2} \end{bmatrix} \begin{bmatrix} \alpha_{1i} \\ \vdots \\ \alpha_{Ki} \\ \beta_{1i} \\ \vdots \\ \beta_{Ki} \end{bmatrix} = \begin{bmatrix} \text{sign}(c_w^*) \bar{\mathbf{x}}_{w_1}(i) \\ \vdots \\ \text{sign}(c_w^*) \bar{\mathbf{x}}_{w_K}(i) \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{c}_i \\ \mathbf{0}_K \end{bmatrix}, \quad (35)$$

where $(K_i^{(l)})$ is the l -th order derivative of K_i

$$[\mathbf{D}_{i,l}]_{mn} := K_i^{(l)}(w_m - w_n), \quad m, n \in \{1, \dots, K\}, l \in \{0, 1, 2\}, \quad (36)$$

and $\mathbf{c}_i := [\text{sign}(c_w^*) \bar{\mathbf{x}}_{w_1}(i) \dots \text{sign}(c_w^*) \bar{\mathbf{x}}_{w_K}(i)]^T \in \mathbb{C}^K$. $\boldsymbol{\psi}(w)$ can be expressed as

$$\boldsymbol{\psi}(w) = \left[\sum_{k=1}^K \mathbf{c}_1(k) \dots \sum_{k=1}^K \mathbf{c}_{N_f}(k) \right]^T. \quad (37)$$

One sufficient condition to ensure the existence for $\alpha_{k,i}$ and $\beta_{k,i}$ ($i = \{1, \dots, N_f\}$) is that

$$\mathbf{K}_i := \begin{bmatrix} \mathbf{D}_{i,0} & \mathbf{D}_{i,1} \\ \mathbf{D}_{i,1} & \mathbf{D}_{i,2} \end{bmatrix} \in \mathbb{C}^{2K \times 2K} \quad (38)$$

is invertible for any $i \in \{1, \dots, N_f\}$, which means $\text{rank}(\mathbf{K}_i) = 2K$. Then, the solution to (35) is uniquely determined by inverting $\mathbf{K}_i$. Unfortunately, the invertibility of $\mathbf{K}_i$ may not be guaranteed in general.

C. Single Source Analysis

We begin with single source analysis ($K = 1$). For one source, there is no separation condition or risk of collision to consider in the analysis. The constructed $K_i(w)$ is guaranteed to satisfy (22) as stated in the theorem.

Theorem 4.1: Suppose $K = 1$ (DOA is w_1), and $\bar{\mathbf{x}}_{w_1}(i) \neq 0$ for $\forall i \in \{1, \dots, N_f\}$. We then have

$$\begin{cases} \boldsymbol{\psi}(w) = \text{sign}(c_w^*) \bar{\mathbf{x}}_w & w = w_1 \\ \|\boldsymbol{\psi}(w)\|_2 < 1 & \forall w \neq w_1. \end{cases} \quad (39)$$

Proof Since $K = 1$, (35) reduces to

$$\begin{bmatrix} K_i(0) & K_i'(0) \\ K_i'(0) & K_i''(0) \end{bmatrix} \begin{bmatrix} \alpha_{1i} \\ \beta_{1i} \end{bmatrix} = \begin{bmatrix} 1/i & 0 \\ 0 & K_i''(0) \end{bmatrix} \begin{bmatrix} \alpha_{1i} \\ \beta_{1i} \end{bmatrix} = \begin{bmatrix} \text{sign}(c_w^*) \bar{\mathbf{x}}_{w_1}(i) \\ 0 \end{bmatrix}. \quad (40)$$

Hence $\alpha_{1i} = i \cdot \text{sign}(c_w^*) \bar{\mathbf{x}}_{w_1}(i)$ and $\beta_{1i} = 0$. Furthermore,

$$\|\boldsymbol{\psi}(w)\|_2^2 = \sum_{i=1}^{N_f} |\alpha_{1i} K_i(w - w_1)|^2. \quad (41)$$

When $w = w_1$, $\boldsymbol{\psi}(w) = [\alpha_{11} K_1(0) \dots \alpha_{1N_f} K_{N_f}(0)]^T = \text{sign}(c_w^*) [\bar{\mathbf{x}}_{w_1}(1) \dots \bar{\mathbf{x}}_{w_1}(N_f)]^T = \text{sign}(c_w^*) \bar{\mathbf{x}}_w$ and $\|\boldsymbol{\psi}(w)\|_2^2 = \|\bar{\mathbf{x}}_{w_1}\|_2^2 = 1$.

For $w \neq w_1$, suppose $\alpha_{1i} = i \cdot \text{sign}(c_w^*) \bar{\mathbf{x}}_{w_1}(i) \neq 0$ for $\forall i \in \{1, \dots, N_f\}$, and notice that $K_i(w - w_1) < K_i(0) = 1/i$. Therefore

$$\|\boldsymbol{\psi}(w)\|_2^2 = \sum_{i=1}^{N_f} |\alpha_{1i} K_i(w - w_1)|^2 < \sum_{i=1}^{N_f} |\alpha_{1i} K_i(0)|^2 = 1. \quad (42)$$

Therefore, (39) must hold. $\square$

D. Multiple Source Analysis

The analysis is now extended to multiple source cases. For the existing ANM based methods, if there is more than one source, a minimum separation condition is assumed [20], [21], [23], [26]. However, in our signal model, we have to consider the potential for aliasing and collisions (see Sec. IV-D1).

We first define the separation of $\mathcal{W}$ for the i -th frequency $\Delta(\mathcal{W}^i)$ as the closest wrap-around distance between two distinct DOAs w_m, w_n

$$\Delta(\mathcal{W}^i) := \inf_{w_m, w_n \in \mathcal{W}} \min\{|w_m - w_n| \bmod 1, 1 - (|w_m - w_n| \bmod 1)\}. \quad (43)$$

Note that although $|w_m - w_n| \in [0, 1)$, for $i \geq 2$, $i|w_m - w_n|$ can be greater than 1. Due to the periodicity of the interpolation kernel, we keep only the fractional part of $i|w_m - w_n|$ in the definition of the separation. We first introduce the concepts of aliasing and collision before our analysis.

1) *Aliasing and Collision:* *Aliasing.* Because of the wrap-around nature of $\mathbf{a}(i, w)$, when $d > \frac{\lambda_{N_f}}{2}$ there will be aliasing peaks in the higher frequencies. Aliasing can happen even for the single source case. Specifically, based on [1], if the temporal frequency $f \cdot f_0$ satisfies

$$f \cdot f_0 \geq \frac{c}{d} \frac{1}{1 + |\cos(\theta)|}, \quad (44)$$

then aliasing peaks enter into the visible region $[-1/2, 1/2]$ and that frequency experiences *aliasing*. When $d = c/(2f_0)$ and $\theta \in [0^\circ, 180^\circ]$, aliasing happens for all $f \geq 2$. In addition to the peak associated with the ground-truth DOA w , there are aliasing peaks with DOAs $\bar{w} = w \pm \frac{k}{f}$, ($k < f, k \in \mathbb{N}_+$). It can be shown that

$$\mathbf{a}(f, w) = \mathbf{a}(f, \bar{w}). \quad (45)$$

Aliasing happens for the single frequency beamforming [1] provided that the temporal frequency is high enough. In [17], the authors demonstrate that multiple frequencies can overcome aliasing for conventional beamforming (CBF) and sparse Bayesian learning (SBL) methods.

Collision. One consequence of aliasing is the possibility of *collision* of multiple DOAs. Collision occurs when one DOA lies exactly in the positions of the aliasing peaks of another source. Formally, suppose there are $K = 2$ distinct DOAs w_1 and w_2 ($w_1, w_2 \in [-1/2, 1/2]$). w_1 and w_2 are said to have *collision* in the i -th frequency if

$$\mathbf{a}(i, w_1) = \mathbf{a}(i, w_2). \quad (46)$$

Such collision occurs whenever w_1 and w_2 satisfy

$$|w_1 - w_2| = \frac{k}{i} \quad (i \in \{2, \dots, N_f\}, k < i, k \in \mathbb{N}_+). \quad (47)$$

When collision occurs in the i -th frequency bin, it is verified that the pi -th ($p \geq 2, p \in \mathbb{N}_+$) frequency bins also have collision.

For CBF and SBL, collision may bring about ambiguities in the source power (and amplitude) estimation as these two sources share the same array manifold vector.

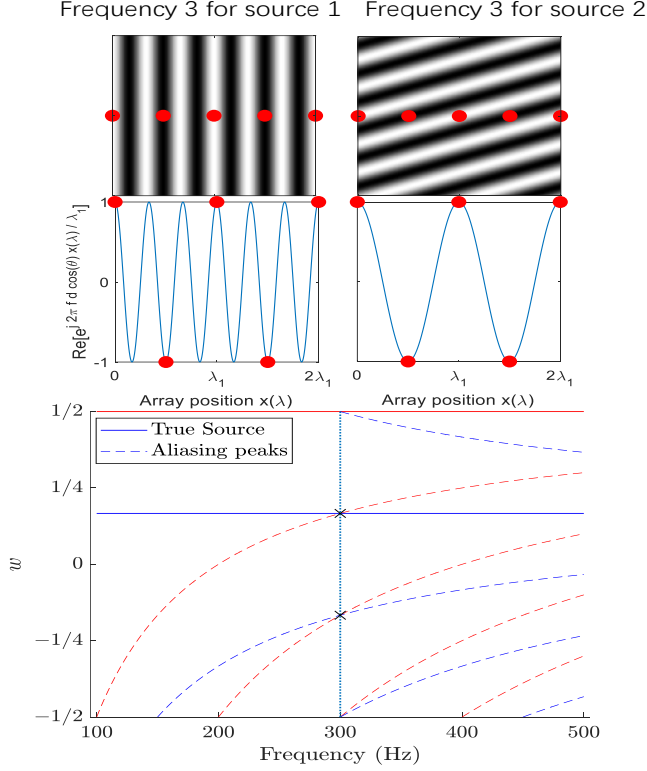


Fig. 6. Collision demonstration. $K = 2$, $N_f = 5$, $w_1 = 1/2$, $w_2 = 1/6$. (a–b) are the same as Fig. 1. (c) Red lines indicate w_1 and blue lines indicate w_2 for the true sources (solid), and the aliased signal (dashed). Collision occurs at 300 Hz.

As an example, let $N_f = 5$, $f_0 = 100$ Hz, $w_1 = 1/2$, and $w_2 = 1/6$. Then $\mathbf{a}(3, w_1) = \mathbf{a}(3, w_2)$ and so these two sources collide in the third frequency bin. As Fig. 6 (a)–(b) illustrate, the spatial samples obtained from all sensors are the same at that frequency. In addition, collision can be interpreted as the intersection of the true DOA of one source and the aliasing peaks of another source. In Fig. 6 (c), it is clear the collision exists in the third frequency (300 Hz).

For our ANM problem, if (47) is satisfied, based on (46), we must have

$$\mathbf{q}_i^H \mathbf{a}(i, w_1) = \mathbf{q}_i^H \mathbf{a}(i, w_2). \quad (48)$$

Based on the definition of the dual polynomial in (13), the i -th entry of $\psi(\mathbf{Q}, w_1)$ and $\psi(\mathbf{Q}, w_2)$ must therefore be equal. This serves as an additional constraint for the dual polynomial. We refer to (47) as the *exact collision* case. Collisions complicate the construction of a dual polynomial that satisfies the optimality condition (see (48)). However, we observe that in the numerical experiments, the method still works in the presence of exact collisions (See Fig. 8 (a)).

2) *Case Classification*: With multiple sources, depending on the true DOAs, we have three possible cases:

- Case 1: There exists an *exact collision*. An exact collision in the i -th frequency is defined as

$$|w_m - w_n| = \frac{k}{i} \quad (i \in \{1, \dots, N_f\}, k < i). \quad (49)$$

for some DOAs w_m, w_n . For example, suppose $w_1 = 1/2$, $w_2 = 1/6$, $N_f = 6$. Since $|w_1 - w_2| = 1/3$, the third frequency has collision. Indeed, as shown in Fig. 6, the spatial samples obtained from all sensors are the same in the third frequency. Notice also that $|w_1 - w_2| = 2/6 = 1/3$, so the sixth frequency also has collision.

- Case 2: There exists a *near collision*. A near collision in the i -th frequency is defined as

$$|w_m - w_n| = \frac{k}{i} \pm \epsilon \quad (i \in \{1, \dots, N_f\}, k < i), \quad (50)$$

for some w_m, w_n for sufficiently small $\epsilon > 0$. The upper bound of ϵ is proportional to $1/N_m$. For example, suppose $w_1 = 1/4$, $w_2 = 0.001$, $N_f = 6$, and the minimum separation condition $\Delta_{\min} = 0.01$. Then $|w_1 - w_2| = 1/4 - 0.001 = 1/4 - \epsilon$ with $\epsilon = 0.001 < \Delta_{\min}$. Therefore, the fourth frequency has a near collision.

- Case 3: There are no collisions or near collisions across all N_f frequencies. For example, suppose $w_1 = 1/4$, $w_2 = 1/10$, $N_f = 6$, and $\Delta_{\min} = 0.01$. It can be easily shown that there is no collision or near collision for any $i \in \{1, \dots, N_f\}$.

3) *Case 1 and 2 Study*: For Case 1 and 2, an analytical guarantee is hard to obtain due to the singularity of $\mathbf{K}_i$. We list some properties for Case 1 in the Appendix E. Although an analytical guarantee is hard to obtain, we find the method (20) can perform well in Case 1 (See Fig. 8 (a)). However, directly solving (20) for Case 2 does not give a satisfactory performance (See Fig. 7). To resolve the near collision issue in (21), we proposed a robust solution in (21). The robust solution applies $\ell_{1,2}$ regularization to nullify the contribution from the near collision frequencies. The numerical examples (see Fig. 7) demonstrate the effectiveness of the $\ell_{1,2}$ regularization in suppressing near collisions.

4) *Case 3 Analysis*: For Case 3, there is no collision and therefore the theoretical analysis becomes tractable. Under a uniform amplitude assumption, we draw analytical conclusions on $\|\psi(w)\|_2$ in Theorem 4.2.

Theorem 4.2: If the amplitude is uniform across frequencies for each source (i.e. $|\mathbf{x}_w(1)| = \dots = |\mathbf{x}_w(N_f)| = 1/\sqrt{N_f}$ for all $w \in \mathcal{W}$), $\Delta(\mathcal{W}^i) \geq 4/(N_m - 1)$ and $N_m \geq 257$, then $\|\psi(w)\|_2 < 1$ for $w \notin \mathcal{W}$.

Remark The assumptions on the uniform amplitudes and the number of sensors are made to facilitate the proof and may not be necessary in practice. Intuitively, the uniform amplitude assumption prevents certain frequency bins from dominating the source amplitudes, which in the extreme case could transform the multi-frequency model into the single-frequency model. The assumption on the number of sensors is used to bound the Fejér kernel. Note also that the separation assumption implicitly implies an upper bound for the source number K .

Proof See Appendix H and the following paragraphs.

With the first K constraints in (35), the constructed $\psi(w)$ automatically satisfies the first equality condition in (22) as $\psi(w)$ satisfies (37). However, we also need to show that with the last K equality constraints in (35), the constructed $\psi(w)$ satisfies the second inequality condition in (22) (i.e.

$\|\psi(w)\|_2 < 1$), and we prove Theorem 4.2 to guarantee that. Inspired by [20], to bound $\|\psi(w)\|_2$, α and β in (35) need to be bounded first. To simplify the derivation, we prove the case when $K = 2$ in the following sections. The result can be generalized to $K > 2$ with the same reasoning.

Supposing that $K = 2$, (35) is simplified as a 4×4 system of equations. Note that $i = 1$ is the classical case [20], [21]. Since collision is absent in this case, the matrix $\mathbf{K}_i$ defined in (38) is invertible (for detailed reasoning, see Appendix G). Therefore, the solution for (35) is *uniquely* identified as

$$\begin{bmatrix} \alpha_{1i} \\ \alpha_{2i} \\ \beta_{1i} \\ \beta_{2i} \end{bmatrix} = \begin{bmatrix} \mathbf{D}_{i,0} & \mathbf{D}_{i,1} \\ \mathbf{D}_{i,1} & \mathbf{D}_{i,2} \end{bmatrix}^{-1} \begin{bmatrix} \text{sign}(c_w^*)\bar{\mathbf{x}}_{w_1}(i) \\ \text{sign}(c_w^*)\bar{\mathbf{x}}_{w_2}(i) \\ 0 \\ 0 \end{bmatrix} \quad (51)$$

$$= \begin{bmatrix} \mathbf{S}_i^{-1} \\ -\mathbf{D}_{i,2}^{-1}\mathbf{D}_{i,1}\mathbf{S}_i^{-1} \end{bmatrix} \begin{bmatrix} \text{sign}(c_w^*)\bar{\mathbf{x}}_{w_1}(i) \\ \text{sign}(c_w^*)\bar{\mathbf{x}}_{w_2}(i) \end{bmatrix}.$$

where the Schur complement $\mathbf{S}_i := \mathbf{D}_{i,0} - \mathbf{D}_{i,1}\mathbf{D}_{i,2}^{-1}\mathbf{D}_{i,1}$.

Define $\alpha_i := [\alpha_{1i} \ \alpha_{2i}]^T$ and $\beta_i := [\beta_{1i} \ \beta_{2i}]^T$. The following lemma gives upper bounds for $\|\alpha_i\|_\infty$ and $\|\beta_i\|_\infty$.

Lemma 4.3: If $\Delta(\mathcal{W}^i) \geq 4/(N_m - 1) = 1/M$ and $N_m \geq 257$ (or $f_c := 2M \geq 128$), then

$$(1) \|\alpha_i\|_\infty \leq i \cdot 1.008824 \text{ and } \|\beta_i\|_\infty \leq \frac{3.294 \times 10^{-2}}{f_c}. \quad (52)$$

(2) If the amplitude is uniform across frequencies for each source (i.e. $|\mathbf{x}_w(1)| = \dots = |\mathbf{x}_w(N_f)| = 1/\sqrt{N_f}$ for all $w \in \{w_1, w_2\}$), we further have

$$\|\alpha_i\|_\infty \leq \frac{i \cdot 1.008824}{\sqrt{N_f}}, \|\beta_i\|_\infty \leq \frac{3.294 \times 10^{-2}}{f_c \sqrt{N_f}}. \quad (53)$$

Proof See Appendix F for (1). The proof for (2) is similar to that of Lemma 4.3 with the additional condition $\left\| \begin{bmatrix} \text{sign}(c_w^*)\bar{\mathbf{x}}_{w_1}(i) \\ \text{sign}(c_w^*)\bar{\mathbf{x}}_{w_2}(i) \end{bmatrix} \right\|_\infty = \frac{1}{\sqrt{N_f}}$. $\square$

Now that the upper bounds for $\|\alpha_i\|_\infty$ and $\|\beta_i\|_\infty$ have been obtained, $\|\psi(w)\|_2$ can be further bounded. The remaining steps for bounding $\|\psi(w)\|_2$ are available in Appendix H.

V. NUMERICAL RESULTS

A. Case Studies

We evaluate our method for the 3 cases mentioned in Sec. IV-D. The noisy case is also evaluated.

The simulation setup for the following examples is K incoherent sources have DOAs $\theta = \{\theta_1, \dots, \theta_K\}$ (90° is considered broadside). Assume $c = 340$ m/s, $f_0 = 100$ Hz, a uniform linear array with N_m sensors and spacing $d = \frac{c}{2f_0}$. The temporal frequencies of the sources are $\{1, \dots, N_f\} \cdot f_0$ 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 $\|\mathbf{x}_w\|_2 = 1$. In Fig. 8, 100 realizations are evaluated and in each realization, $\mathbf{x}_w$ will be different. We plot the distribution of the DOA estimation of these realizations in the histogram. All $c_w = 1$. The noise for each frequency $\mathbf{w}_f$ is randomly generated from the distribution

$\mathcal{CN}(0, \sigma^2)$ and then scaled to fit the desired signal-to-noise ratio (SNR) defined as

$$\text{SNR} = 20 \log_{10} \frac{\|\mathbf{X}\|_F}{\|\mathbf{W}\|_F}. \quad (54)$$

This setup is applied in all of the examples in the Sec. V-A unless otherwise specified.

1) *The Dual Polynomial for Case 2:* For Case 2, if $d = \frac{c}{2f_0}$, then all of the frequencies other than the first frequency will have the risk of near collision. To overcome this issue, robust ANM (see (21)) needs to be employed to suppress the near collision. An alternative way to suppress the collision is to choose a smaller spacing $d = \frac{c}{2N_f f_0}$ so that the collision can be completely avoided for all frequencies. These two collision suppression methods will be examined. Suppose there are $K = 2$ incoherent sources. In this case, if $N_f \geq 2$, then the $2n$ -th (n is any positive integer) frequency will have the near collision. The dual polynomial for different N_f , λ , d , and θ (λ is the regularization hyper-parameter in (21)) can be seen in Fig. 7. For the regularization parameter λ , we empirically choose it proportional to N_f (i.e. $\lambda = k \cdot N_f$, with $k = 0.125$ in particular for Fig. 7). The intuition behind is that for more frequencies, the near collision is more likely to happen. However, since the regularization can bring bias, a smaller λ is more favorable in practice.

From Fig. 7 (a), if we only solve (20) without regularization, numerous spurious peaks are an obstacle for identifying source positions. However, with regularization, the dual polynomial peaks become precise indicators for the source positions (See Fig. 7(b-c)). When $N_f = 6$, the near collision frequencies are the 2nd, 4th, and 6th frequencies. Fig. 7 (d) demonstrates the success of choosing a smaller spacing $d = \frac{c}{2N_f f_0}$ in collision suppression without regularization. However, there are potential limitations for smaller spacing. Comparing Fig. 7 (e) and (f), the smaller aperture cannot resolve the close sources while the larger aperture can. Thus, although the smaller aperture can avoid the collision, it has lower spatial resolution. We leave the theoretical analysis for choosing the regularization hyperparameter λ as future work.

2) *Case 1, 3, and Noisy cases:* The histograms for these cases are plotted in Fig. 8. Since $|w_1 - w_2| = |w_2 - w_3| = 0.25$ and $|w_1 - w_3| = 0.5$, there are collisions in both the second and fourth frequencies. From Fig. 8 (a), all of the instances in the histogram are nevertheless concentrated in the ground-truth positions, which shows the proposed method can capture the ground-truth positions accurately and has the robustness to the exact collisions. The robustness to the exact collisions is attributed to the combination of multiple frequencies. For the collision frequencies, these two sources are essentially one source since they share the same array manifolds for these frequencies (see (46)) and they are mixed coherently, which makes it difficult to separate them. For the non-collision frequencies, the two sources are well-separated. Therefore, if we combine all N_f frequencies, the two peaks associated with the DOAs still stand out as long as there exists non-collision frequencies. To demonstrate Case 3, we compare the single-frequency ($N_f = 1$, see Fig. 8 (b)) and multi-frequency ($N_f = 5$, see Fig. 8 (c)) scenarios. When $N_f = 1$, there are

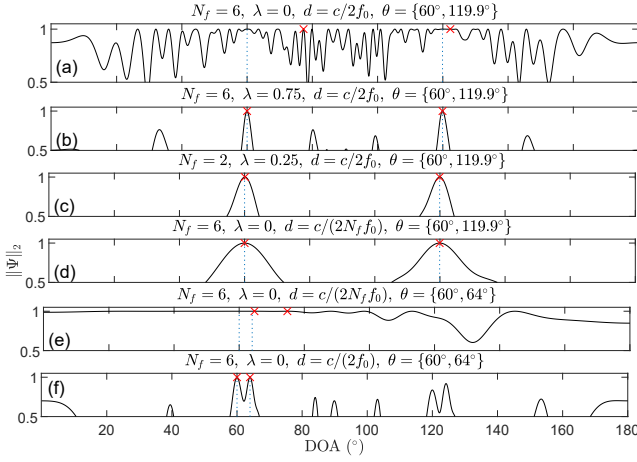


Fig. 7. $\|\psi(\mathbf{Q}, \mathbf{w})\|_2$ versus DOA θ for Case 2. $N_m = 12$, $f_0 = 100$ Hz, $d = \frac{c}{2f_0}$, and $\mathbf{x}_w \sim \mathcal{CN}(0, 1)$, $K = 2$. “x” indicates the peak, and the dashed lines indicate the ground-truth DOAs.

many bins that lie in the undesired positions. In contrast, when $N_f = 5$, the bins are mostly concentrated in the ground-truth positions. This example demonstrates the potential benefits of multi-frequency ANM. In Fig. 8 (d), the setup is identical to that in Fig. 8 (c) except the noise is present. For the noisy case, the empirical value of η is chosen as [25]

$$\eta = \sigma/2 \cdot \sqrt{N_m N_f + 2\sqrt{N_m N_f}}.$$

From Fig. 8 (d), the proposed method captures the source positions accurately in the noisy cases.

B. DOA Estimation Performance Evaluation

To comprehensively evaluate the performance of the proposed method, we conduct Monte Carlo experiments. In all of the experiments in this section, each point represents $MC = 100$ trials, and the root mean square error (RMSE) and mean absolute error (MAE) are computed as

$$\text{RMSE} = \sqrt{\frac{1}{MC} \sum_{m=1}^{MC} \left[\frac{1}{K} \sum_{k=1}^K (\hat{\theta}_{mk} - \theta_{mk})^2 \right]}. \quad (55)$$

$$\text{MAE} = \frac{1}{MC} \sum_{m=1}^{MC} \left(\frac{1}{K} \sum_{k=1}^K |\hat{\theta}_{mk} - \theta_{mk}| \right), \quad (56)$$

where $\hat{\theta}_{mk}$, and θ_{mk} are (sorted) estimated DOAs, and (sorted) ground-truth DOAs for the k -th source and m -th trial. A maximum threshold of 10° was used to penalize the incorrect DOA estimates (see below). c , f_0 , d , and the temporal frequencies are the same as those in Sec. V-A. We also compare the proposed method (ANM) with the multi-frequency sparse Bayesian learning (SBL) [17] and Cramér-Rao bound (CRB) [49, Eq. (119)]. For SBL, the spatial angle is discretized into grids with 0.5° between the adjacent grid points. Although there are many DOA estimation methods, very few of them have been developed for the multiple-frequency model. Therefore, only SBL and CRB are included for reference.

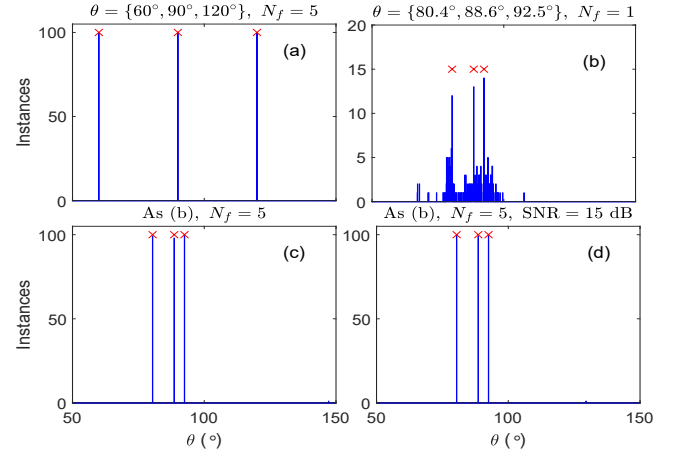


Fig. 8. Histogram of the estimated DOA $\hat{\theta}$ for 100 realizations with true DOAs (x). $N_m = 12$, $f_0 = 100$ Hz, $d = \frac{c}{2f_0}$, and $\mathbf{x}_w \sim \mathcal{CN}(0, 1)$, $K = 3$. For each realization, $\mathbf{x}_w$ will be different. No noise is present except for (d) where SNR is 15 dB.

1) *DOA Estimation under Varying SNR*: We first examine the robustness of ANM to noise. The performance of each algorithm under $d = \frac{\lambda_{N_f}}{2}$ is detailed in Fig. 9. Notice that in this setup, there will be no aliasing or collision. Therefore, we can turn off the $\ell_{1,2}$ regularization in (21). The proposed algorithm outperforms SBL in the high SNR cases. At low SNRs, SBL achieves a better performance since it can estimate the noise power. Note for the SBL with limited 0.5° separation, the achievable accuracy for RMSE is 0.125° . In addition, it turns out that SBL has no failure trials ($\text{RMSE} > 10^\circ$ is defined as failure) starting from $\text{SNR} = 0, -5, -5$, and -10 dB for $N_f = 1, 2, 4$, and 8 . For ANM, the same happens for $\text{SNR} = 0, 0, -5$, and -5 dB. Therefore, for both SBL and ANM, the performance improves in the low SNR region, which demonstrates the enhanced robustness to noise for the multi-frequency processing.

We then change the spacing to $d = \frac{\lambda_1}{2}$ (See Fig. 10). In this case, aliasing and possible collisions will be present when $N_f \geq 2$. However, if more frequencies are available, such ambiguities can be potentially suppressed [17]. For that reason, we only consider the case with 8 frequencies from $100, \dots, 800$ Hz. In Fig. 9 the frequencies were $12.5, \dots, 100$ Hz, the aperture is here a factor 8 larger in Fig. 10. Although the error stops to decrease for ANM in the high SNR region due to the bias from the regularization, the performance still improves in the low SNR region if more frequencies are available. In addition, compared with Fig. 9 (d), the performance of ANM improves when SNR is between 0 to 20 dB, and that demonstrates the benefits of larger apertures.

2) *DOA Estimation under Varying K*: We examine the DOA estimation performance under varying numbers of sources (K) in this section. Both the real flat (Fig. 11 (a)) and complex random amplitude source (Fig. 11 (b)) are tested under noise-free conditions. DOA is an integer randomly generated from a uniform distribution between $[0^\circ, 180^\circ]$. Therefore, there is no grid mismatch issue for SBL. For the real and flat amplitude case ($\mathbf{x}_w = 1/\sqrt{N_f} \cdot \mathbf{1}_{N_f}$), ANM

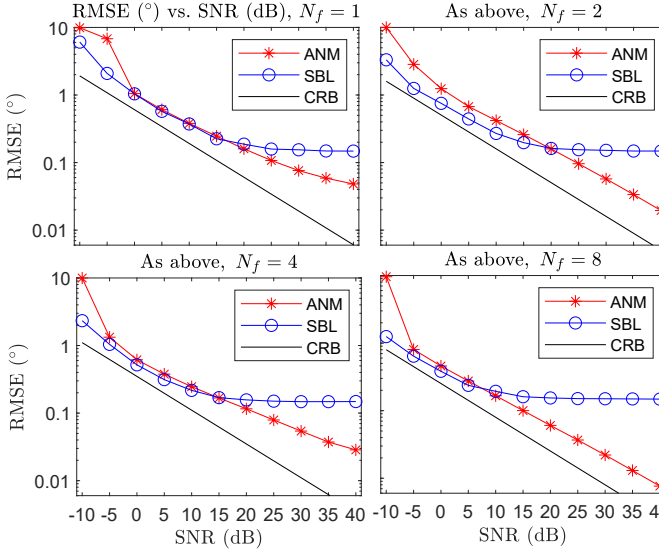


Fig. 9. RMSE (°) vs. SNR for $d = \frac{\lambda N_f}{2} = \frac{c}{2N_f f_0}$. $N_m = 15$, $K = 3$, $f_0 = 100$ Hz, and the frequency set is $\{1, \dots, N_f\} \cdot f_0$ Hz. The $\ell_{1,2}$ regularization parameter $\lambda = 0$ for all plots. Each point represents 100 trials. The DOAs for each trial are randomly generated between $[10^\circ, 170^\circ]$ with a minimum angular separation $4/N_m$. $\mathbf{x}_w \sim \mathcal{CN}(0, 1)$.

will be immune to collisions (or near collisions) since the fundamental constraint (48) and the dual certificate condition (22) can be satisfied simultaneously. Therefore, the optimality is guaranteed and perfect DOA estimation is expected. In the complex random amplitude case, since near collisions affect the performance of ANM, robust ANM (see (21)) is applied. From Fig. 11 (b), the DOA estimation error increases when the complex amplitude is applied for both methods. ANM (and robust ANM) still outperforms SBL for both real and complex amplitudes even if there is no grid mismatch for SBL. Fig. 11 (b) also demonstrates the effectiveness of robust ANM for suppressing near collisions. Because of the presence of near collisions in the complex amplitude case, more frequencies do not necessarily bring about better performance for ANM.

3) *DOA estimation under Varying DOA separation*: Finally, we study the DOA estimation performance under different DOA separations. Since the amplitude is real and flat, ANM is immune to near collisions. From Fig. 11 (c), SBL has the same estimation error for all DOA separations and N_f . That error is entirely from the grid mismatch. However, the proposed gridless approach overcomes this issue and achieves exact DOA estimation.

VI. CONCLUSIONS

The ANM framework is extended to support continuous parameter estimation across multiple frequencies. ANM is initially formulated as an equivalent SDP problem based on the bounded real lemma so that the ANM becomes computationally tractable. In addition, the dual certificate condition is derived. With the help of the dual certificate condition, the optimality can be certified, and the DOAs are identified by finding the roots of a polynomial. We also construct the

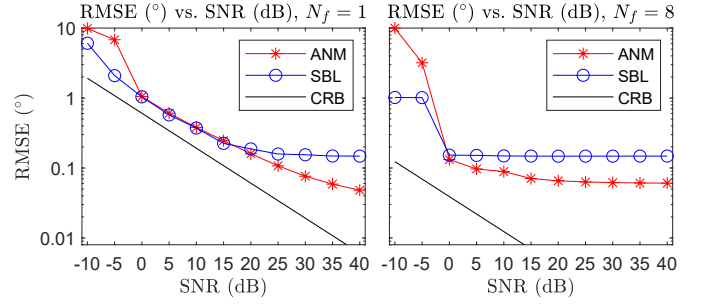


Fig. 10. RMSE (°) vs. SNR for $d = \frac{\lambda_1}{2} = \frac{c}{2f_0}$. $N_m = 15$, $K = 3$, $f_0 = 100$ Hz, and the frequency set is $\{1, \dots, N_f\} \cdot f_0$ Hz. The $\ell_{1,2}$ regularization parameter $\lambda = 0.6$ for (b). Each point represents 100 trials. The DOAs for each trial are randomly generated between $[10^\circ, 170^\circ]$ with a minimum angular separation $4/N_m$. $\mathbf{x}_w \sim \mathcal{CN}(0, 1)$.

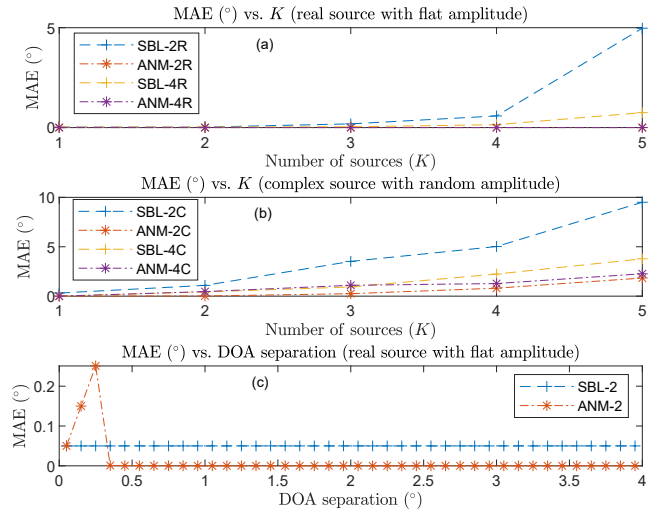


Fig. 11. MAE (°) vs. K (a–b) and DOA separation (c). $N_m = 15$. Each point represents 100 trials, and no noise is present. For (a–b), $N_f = \{2, 4\}$. For (b), robust ANM (21) is used. For (c), $N_f = 2$, $K = 2$, and $\mathbf{x}_w = 1/\sqrt{N_f} \cdot \mathbf{1}_{N_f}$. The first DOA is $90^\circ - \text{DOA separation}$, and the second DOA is $90^\circ + \text{DOA separation}$. The grid resolution for SBL is 0.1° .

dual certificate and show that a valid construction exists when the source amplitude has a uniform magnitude. Based on our signal model, the higher frequencies may have the risk of collision or near collision. These two cases are extensively studied and a robust ANM method with regularization is proposed for near collision suppression. The numerical results demonstrate the effectiveness of the proposed method.

ACKNOWLEDGEMENT

This work is supported by NSF Grant CCF-1704204, NSF Grant CCF-2203060, and Office of Naval Research (ONR) Grant N00014-21-1-2267.

REFERENCES

- [1] H. L. Van Trees, *Optimum array processing: Part IV of detection, estimation, and modulation theory*. John Wiley & Sons, 2004.
- [2] K. Venugopal, A. Alkhateeb, N. G. Prelcic, and R. W. Heath, "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1996–2009, 2017.

- [3] Y. Wei, Y. Li, Z. Ding, Y. Wang, T. Zeng, and T. Long, "SAR parametric super-resolution image reconstruction methods based on admm and deep neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10 197–10 212, 2021.
- [4] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, 1986.
- [5] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, and Signal Process.*, vol. 37, no. 7, pp. 984–995, 1989.
- [6] M. Wax, T.-J. Shan, and T. Kailath, "Spatio-temporal spectral analysis by eigenstructure methods," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 4, pp. 817–827, 1984.
- [7] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wideband sources," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 4, pp. 823–831, 1985.
- [8] K. M. Buckley and L. J. Griffiths, "Broad-band signal-subspace spatial-spectrum (BASS-ALE) estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 7, pp. 953–964, 1988.
- [9] E. D. Di Claudio and R. Parisi, "WAVES: Weighted average of signal subspaces for robust wideband direction finding," *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2179–2191, 2001.
- [10] Y.-S. Yoon, L. M. Kaplan, and J. H. McClellan, "TOPS: New DOA estimator for wideband signals," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 1977–1989, 2006.
- [11] J. Zhang, N. Hu, M. Bao, X. Li, and W. He, "Wideband DOA estimation based on block FOCUSS with limited samples," in *IEEE GlobalSIP*, 2013, pp. 634–637.
- [12] L. Wang, L. Zhao, G. Bi, C. Wan, L. Zhang, and H. Zhang, "Novel wideband DOA estimation based on sparse Bayesian learning with Dirichlet process priors," *IEEE Trans. Signal Process.*, vol. 64, no. 2, pp. 275–289, 2015.
- [13] C. Liu, Y. V. Zakharov, and T. Chen, "Broadband underwater localization of multiple sources using basis pursuit denoising," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1708–1717, 2011.
- [14] Z. Tang, G. Blacquiere, and G. Leus, "Aliasing-free wideband beamforming using sparse signal representation," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3464–3469, 2011.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] C. Fernandez-Granda, "Super-resolution of point sources via convex programming," *Inf. Inference, J. IMA*, vol. 5, no. 3, pp. 251–303, 2016.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] 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.
- [30] Z. Yang, J. Tang, Y. C. Eldar, and L. Xie, "On the sample complexity of multichannel frequency estimation via convex optimization," *IEEE Trans. Inf. Theory*, vol. 65, no. 4, pp. 2302–2315, 2018.
- [31] 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.
- [32] Y. Park and P. Gerstoft, "Gridless sparse covariance-based beamforming via alternating projections including co-prime arrays," *J. Acoust. Soc. Am.*, vol. 151, no. 6, pp. 3828–3837, 2022.
- [33] 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.
- [34] P. Chen, Z. Chen, Z. Cao, and X. Wang, "A new atomic norm for DOA estimation with gain-phase errors," *IEEE Trans. Signal Process.*, vol. 68, pp. 4293–4306, 2020.
- [35] S. Li, P. Nayeri, and M. B. Wakin, "Digital beamforming robust to time-varying carrier frequency offset," *arXiv preprint arXiv:2103.04948*, 2021.
- [36] A. Xenaki and P. Gerstoft, "Grid-free compressive beamforming," *J. Acoust. Soc. Am.*, vol. 137, pp. 1923–1935, 2015.
- [37] 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.
- [38] 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.
- [39] 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.
- [40] 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.
- [41] —, "Support recovery for sparse signals with unknown non-stationary modulation," *IEEE Trans. Signal Process.*, vol. 68, pp. 1884–1896, 2020.
- [42] 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.
- [43] J. W. Helland, *Atomic Norm Algorithms for Blind Spectral Super-resolution Problems*. Colorado School of Mines, 2019.
- [44] B. Dumitrescu, *Positive trigonometric polynomials and signal processing applications*. Springer, 2017, vol. 103.
- [45] Z. Yang, "A gridless sparse method for super-resolution of harmonics," in *IEEE EUSIPCO*, 2017, pp. 2096–2100.
- [46] Y. Wu, M. B. Wakin, and P. Gerstoft, "Gridless DOA estimation under the multi-frequency model," in *IEEE ICASSP*, 2022, pp. 5982–5986.
- [47] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [48] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," 2014.
- [49] Y. Liang, W. Liu, Q. Shen, W. Cui, and S. Wu, "A review of closed-form Cramér-Rao bounds for DOA estimation in the presence of Gaussian noise under a unified framework," *IEEE Access*, vol. 8, pp. 175 101–175 124, 2020.

APPENDIX

A. Proof for Proposition 3.1

Construct the Hermitian trigonometric polynomial

$$R(w) := 1 - \|\mathbf{H}^H \mathbf{z}\|_2^2 = 1 - \mathbf{z}^H \mathbf{H} \mathbf{H}^H \mathbf{z}. \quad (57)$$

From (17), we know that $\|\mathbf{Q}\|_{\mathcal{A}}^* \leq 1$ holds if and only if $R(w) \geq 0$ for all $w \in [-1/2, 1/2]$.

First, suppose there exists a matrix $\mathbf{P}_0 \in \mathbb{C}^{N \times N} \succeq 0$ such that (18) and (19) hold. We must argue that $R(w) \geq 0$ for all w . Consider the expression $\mathbf{z}^H \mathbf{P}_0 \mathbf{z}$ and note that

$$\mathbf{z}^H \mathbf{P}_0 \mathbf{z} = \text{Tr}(\mathbf{z}^H \mathbf{P}_0 \mathbf{z}) = \text{Tr}(\mathbf{z} \mathbf{z}^H \mathbf{P}_0) = \sum_{k=-(N-1)}^{N-1} r_k z^{-k},$$

where $r_k = \sum_{i=1}^{N-k} \mathbf{P}_0(i, i+k)$ for $k \geq 0$ and $r_k = r_{-k}^*$ for $k < 0$. From (18), we conclude that $\mathbf{z}^H \mathbf{P}_0 \mathbf{z} = z^0 = 1$. Substituting this into $R(w)$ and defining $\mathbf{P}_1 := \mathbf{H} \mathbf{H}^H$ gives

$$R(w) = \mathbf{z}^H \mathbf{P}_0 \mathbf{z} - \mathbf{z}^H \mathbf{P}_1 \mathbf{z} = \mathbf{z}^H (\mathbf{P}_0 - \mathbf{P}_1) \mathbf{z}.$$

Since the matrix in (19) is PSD, its Schur complement $\mathbf{P}_0 - \mathbf{H} \mathbf{I}_{N_f}^{-1} \mathbf{H}^H = \mathbf{P}_0 - \mathbf{P}_1 \succeq 0$, and so $R(w) \geq 0$ for all $w \in [-1/2, 1/2]$.

Next, suppose $R(w) \geq 0$ for all $w \in [-1/2, 1/2]$. We must argue that there exists a matrix $\mathbf{P}_0 \in \mathbb{C}^{N \times N} \succeq 0$ such that (18) and (19) hold. Since $R(w) \geq 0$, $1 \geq \mathbf{z}^H \mathbf{P}_1 \mathbf{z}$, where we have again defined $\mathbf{P}_1 := \mathbf{H} \mathbf{H}^H \succeq 0$. From [44, Lemma 4.25] and the fact that 1 and $\mathbf{z}^H \mathbf{P}_1 \mathbf{z}$ are univariate trigonometric polynomials, it follows that there exists $\mathbf{P}_0 \succeq \mathbf{P}_1$ such that $1 = \mathbf{z}^H \mathbf{P}_0 \mathbf{z}$ and (18) hold. The matrix in (19) has Schur complement $\mathbf{P}_0 - \mathbf{H} \mathbf{I}_{N_f}^{-1} \mathbf{H}^H = \mathbf{P}_0 - \mathbf{P}_1 \succeq 0$, and therefore (19) holds. $\square$

B. Proof for Theorem 3.2

First, notice that if (22) is satisfied, based on (17), we have $\|\mathbf{Q}\|_{\mathcal{A}}^* \leq 1$. Then,

$$\begin{aligned} \|\mathbf{X}\|_{\mathcal{A}} &\geq \|\mathbf{X}\|_{\mathcal{A}} \cdot \|\mathbf{Q}\|_{\mathcal{A}}^* \stackrel{(a)}{\geq} \langle \mathbf{Q}, \mathbf{X} \rangle_{\mathbb{R}} = \text{Re}[\text{Tr}(\mathbf{Q}^H \mathbf{X})] \\ &= \sum_{w \in \mathcal{W}} \text{Re}[\text{Tr}(c_w \mathbf{Q}^H \mathbf{A}(w) \otimes \mathbf{x}_w^T)] \\ &= \sum_{w \in \mathcal{W}} \sum_{f=1}^{N_f} \text{Re}[c_w \mathbf{q}_f^H x_w(f) \mathbf{a}(f, w)] = \sum_{w \in \mathcal{W}} \text{Re}[c_w \mathbf{x}_w^H \boldsymbol{\psi}(\mathbf{Q}, w)] \\ &\stackrel{(b)}{=} \sum_{w \in \mathcal{W}} \text{Re}[c_w \text{sign}(c_w^*) \|\mathbf{x}_w\|_2^2] = \sum_{w \in \mathcal{W}} |c_w| \stackrel{(c)}{\geq} \|\mathbf{X}\|_{\mathcal{A}}, \end{aligned} \quad (58)$$

where (a) is based on Hölder's inequality, (b) follows because if $w \in \mathcal{W}$, then $\boldsymbol{\psi}(\mathbf{Q}, w) = \text{sign}(c_w^*) \mathbf{x}_w$ based on (22), and (c) follows from the definition of the atomic norm (9) as the infimum of the combination coefficients. Hence, $\|\mathbf{X}\|_{\mathcal{A}} = \langle \mathbf{Q}, \mathbf{X} \rangle_{\mathbb{R}} = \sum_{w \in \mathcal{W}} |c_w|$.

For uniqueness, suppose there exists another decomposition $\mathbf{X} = \sum_{w'} c_{w'} \mathbf{A}(w') \otimes \mathbf{x}_{w'}^T$ 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 of the atoms. Therefore, we have the contradiction:

$$\begin{aligned} \sum_{w'} |c_{w'}| &= \|\mathbf{X}\|_{\mathcal{A}} = \langle \mathbf{Q}, \mathbf{X} \rangle_{\mathbb{R}} = \sum_{w'} \text{Re}[c_{w'} \langle \mathbf{x}_{w'}, \boldsymbol{\psi}(\mathbf{Q}, w') \rangle] \\ &= \sum_{w' \in \mathcal{W}} \text{Re}[c_{w'} \mathbf{x}_{w'}^H \boldsymbol{\psi}(\mathbf{Q}, w')] + \sum_{w' \notin \mathcal{W}} \text{Re}[c_{w'} \mathbf{x}_{w'}^H \boldsymbol{\psi}(\mathbf{Q}, w')] \\ &\stackrel{(a)}{<} \sum_{w' \in \mathcal{W}} |c_{w'}| + \sum_{w' \notin \mathcal{W}} |c_{w'}| = \sum_{w'} |c_{w'}|, \end{aligned} \quad (59)$$

where (a) is because of (22). Therefore, the atomic decomposition which satisfies $\|\mathbf{X}\|_{\mathcal{A}} = \sum_{w \in \mathcal{W}} |c_w|$ must be unique. $\square$

C. Proof for Proposition 3.3

Construct the Hermitian trigonometric polynomial

$$R(w) := 1 - \|\mathbf{H}_r^H \mathbf{z}_r\|_2^2 = 1 - \mathbf{z}_r^H \mathbf{H}_r \mathbf{H}_r^H \mathbf{z}_r. \quad (60)$$

From (17), we know that $\|\mathbf{Q}\|_{\mathcal{A}}^* \leq 1$ holds if and only if $R(w) \geq 0$ for all w .

First, suppose there exists a matrix $\mathbf{P}_{r0} \in \mathbb{C}^{N_u \times N_u} \succeq 0$ such that (27) and (28) hold. We must argue that $R(w) \geq 0$ for all w . Consider the expression $\mathbf{z}_r^H \mathbf{P}_{r0} \mathbf{z}_r$ and note that

$$\begin{aligned} \mathbf{z}_r^H \mathbf{P}_{r0} \mathbf{z}_r &= \text{Tr}(\mathbf{z}_r^H \mathbf{P}_{r0} \mathbf{z}_r) = \text{Tr}(\mathbf{z}_r \mathbf{z}_r^H \mathbf{P}_{r0}) = \sum_{k=-(N-1)}^{N-1} r_k z^{-k} \\ r_k &= \sum_{i,j, \mathcal{U}_j - \mathcal{U}_i = k} \mathbf{P}_{r0}(i, j) \end{aligned} \quad (61)$$

for $k \geq 0$ and $r_k = r_{-k}^*$ for $k < 0$. From (27), we then conclude that $\mathbf{z}_r^H \mathbf{P}_{r0} \mathbf{z}_r = z^0 = 1$. Substituting this fact into $R(w)$ and defining $\mathbf{P}_{r1} := \mathbf{H}_r \mathbf{H}_r^H$, we have

$$R(w) = \mathbf{z}_r^H \mathbf{P}_{r0} \mathbf{z}_r - \mathbf{z}_r^H \mathbf{P}_{r1} \mathbf{z}_r = \mathbf{z}_r^H (\mathbf{P}_{r0} - \mathbf{P}_{r1}) \mathbf{z}_r. \quad (62)$$

Since the matrix in (28) is PSD, its Schur complement $\mathbf{P}_{r0} - \mathbf{H}_r \mathbf{I}_{N_f}^{-1} \mathbf{H}_r^H = \mathbf{P}_{r0} - \mathbf{P}_{r1} \succeq 0$, and so $R(w) \geq 0$ for all w . $\square$

D. The Derivation of the Dual Problem of (20)

Consider the Lagrangian of (20) given by

$$\begin{aligned} \mathcal{L}(\mathbf{Q}, \mathbf{P}_0, \mathbf{H}, \boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2, \boldsymbol{\Lambda}_3, \boldsymbol{\Lambda}_Q, \mathbf{v}) &= \\ \langle \mathbf{Q}, \mathbf{Y} \rangle_{\mathbb{R}} - \left\langle \begin{bmatrix} \boldsymbol{\Lambda}_1 & \boldsymbol{\Lambda}_2 \\ \boldsymbol{\Lambda}_2^H & \boldsymbol{\Lambda}_3 \end{bmatrix}, \begin{bmatrix} \mathbf{P}_0 & \mathbf{H} \\ \mathbf{H}^H & \mathbf{I}_{N_f} \end{bmatrix} \right\rangle_{\mathbb{R}} \\ &- \sum_{k=0}^{N-1} v_k (\delta_k - \sum_{j-i=k} \mathbf{P}_0(i, j)) - \langle \boldsymbol{\Lambda}_Q, \mathbf{H} - \mathcal{R}^*(\mathbf{Q}) \rangle_{\mathbb{R}} \\ &= \langle \mathbf{Q}, \mathbf{Y} \rangle_{\mathbb{R}} + \langle \boldsymbol{\Lambda}_Q, \mathcal{R}^*(\mathbf{Q}) \rangle_{\mathbb{R}} - [\langle \mathbf{P}_0, \boldsymbol{\Lambda}_1 \rangle_{\mathbb{R}} + 2\langle \boldsymbol{\Lambda}_2, \mathbf{H} \rangle_{\mathbb{R}} + \text{Tr}(\boldsymbol{\Lambda}_3)] \\ &- \mathbf{v}_0 + \langle \mathbf{P}_0, \text{Toep}(\mathbf{v}) \rangle_{\mathbb{R}} - \langle \boldsymbol{\Lambda}_Q, \mathbf{H} \rangle_{\mathbb{R}}. \end{aligned} \quad (63)$$

The derivation uses: $\sum_{k=0}^{N-1} v_k \sum_{j-i=k} \mathbf{P}_0(i, j) = \langle \mathbf{P}_0, \text{Toep}(\mathbf{v}) \rangle_{\mathbb{R}}$. Further, the dual matrix $\begin{bmatrix} \boldsymbol{\Lambda}_1 & \boldsymbol{\Lambda}_2 \\ \boldsymbol{\Lambda}_2^H & \boldsymbol{\Lambda}_3 \end{bmatrix}$ associated with the inequality constraint $\begin{bmatrix} \mathbf{P}_0 & \mathbf{H} \\ \mathbf{H}^H & \mathbf{I}_{N_f} \end{bmatrix} \succeq 0$ is an PSD matrix to ensure the inner product between these two matrices is non-negative, whereby the optimal value for the dual problem gives a lower bound for the primal problem.

The dual function is

$$\begin{aligned} g(\boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2, \boldsymbol{\Lambda}_3, \boldsymbol{\Lambda}_Q, \mathbf{v}) &= \inf_{\mathbf{Q}, \mathbf{P}_0, \mathbf{H}} \mathcal{L}(\mathbf{Q}, \mathbf{P}_0, \mathbf{H}, \boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2, \boldsymbol{\Lambda}_3, \boldsymbol{\Lambda}_Q, \mathbf{v}) \\ \text{s.t. } \begin{bmatrix} \boldsymbol{\Lambda}_1 & \boldsymbol{\Lambda}_2 \\ \boldsymbol{\Lambda}_2^H & \boldsymbol{\Lambda}_3 \end{bmatrix} &\succeq 0. \end{aligned} \quad (64)$$

The infimum of $\mathcal{L}$ over $\mathbf{Q}$ is thereby $\inf_{\mathbf{Q}} J(\mathbf{Q}) := [\langle \mathbf{Q}, \mathbf{Y} \rangle_{\mathbb{R}} + \langle \boldsymbol{\Lambda}_Q, \mathcal{R}^*(\mathbf{Q}) \rangle_{\mathbb{R}}] = [\langle \mathbf{Y}, \mathbf{Q} \rangle_{\mathbb{R}} + \langle \mathcal{R}(\boldsymbol{\Lambda}_Q), \mathbf{Q} \rangle_{\mathbb{R}}] = \langle \mathbf{Y} + \mathcal{R}(\boldsymbol{\Lambda}_Q), \mathbf{Q} \rangle_{\mathbb{R}}$. The infimum of $J(\mathbf{Q})$ is bounded only if $\mathbf{Y} = -\mathcal{R}(\boldsymbol{\Lambda}_Q)$. Similarly, the infimum of $\mathcal{L}$ over $\mathbf{P}_0$ is bounded only if $\text{Toep}(\mathbf{v}) = \boldsymbol{\Lambda}_1 \succeq 0$. The infimum of $\mathcal{L}$ over $\mathbf{H}$ is bounded only if $\boldsymbol{\Lambda}_Q = -2\boldsymbol{\Lambda}_2$. Consider $2\boldsymbol{\Lambda}_2 = \tilde{\mathbf{Y}}$, then we must have $\mathbf{Y} = -\mathcal{R}(\boldsymbol{\Lambda}_Q) = \mathcal{R}(2\boldsymbol{\Lambda}_2) = \mathcal{R}(\tilde{\mathbf{Y}})$.

Consider $\mathbf{A}_3 = \frac{1}{2}\mathbf{W}$, and $\mathbf{v} = \frac{1}{2}\mathbf{u}$, the dual function becomes $-\frac{1}{2}\text{Tr}(\mathbf{W}) - \frac{1}{2}\text{Tr}(\text{Toep}(\mathbf{u}))$. The dual problem is

$$\begin{aligned} \max_{\mathbf{W}, \mathbf{u}, \tilde{\mathbf{Y}}} & -\frac{1}{2}[\text{Tr}(\mathbf{W}) + \text{Tr}(\text{Toep}(\mathbf{u}))] \\ \text{s.t.} & \begin{bmatrix} \text{Toep}(\mathbf{u}) & \tilde{\mathbf{Y}} \\ \tilde{\mathbf{Y}}^H & \mathbf{W} \end{bmatrix} \succeq 0, \mathbf{Y} = \mathcal{R}(\tilde{\mathbf{Y}}), \end{aligned} \quad (65)$$

which is equivalent to (29). $\square$

E. Properties for Exact Collision

1) $\mathbf{K}_i$ is Singular: First observe that $\mathbf{K}_i$ in (38) is singular. We also recognize the periodicity of $K_i(w)$. Since $K_i(w) = K_i(w + k/i)$ ($k < i, i \in \{1, \dots, N_f\}$), k/i is the period for $K_i(w)$. In addition, k/i is also the period for $K'_i(w)$ and $K''_i(w)$. Without loss of generality, we assume there exists collision between w_1 and w_2 (i.e. $|w_1 - w_2| = \frac{k}{i}$), then

$$\begin{aligned} K_i(0) &= K_i(w_1 - w_2) = K_i(w_2 - w_1) = 1, \\ K'_i(0) &= K'_i(w_1 - w_2) = K'_i(w_2 - w_1) = 0, \text{ and} \\ K''_i(0) &= K''_i(w_1 - w_2) = K''_i(w_2 - w_1). \end{aligned} \quad (66)$$

The first and second row of $\mathbf{K}_i$ are

$$\begin{aligned} [K_i(w_1 - w_1) \dots K_i(w_1 - w_K) \dots K'_i(w_1 - w_1) \dots K'_i(w_1 - w_K)] \\ [K_i(w_2 - w_1) \dots K_i(w_2 - w_K) \dots K'_i(w_2 - w_1) \dots K'_i(w_2 - w_K)]. \end{aligned} \quad (67)$$

Note that $K_i(w_2 - w_j) = K_i(w_1 - w_j - (w_1 - w_2)) = K_i(w_1 - w_j)$ and $K'_i(w_2 - w_j) = K'_i(w_1 - w_j - (w_1 - w_2)) = K'_i(w_1 - w_j)$ for any j . Thus, the first two rows are identical. $\mathbf{K}_i$ is hence rank-deficient and singular.

However, the singularity of $\mathbf{K}_i$ does not imply that the solution to the system of equations (35) does not exist. If $[\text{sign}(c_w^*)x_{w_1}(i) \dots \text{sign}(c_w^*)x_{w_K}(i) \ 0 \dots 0]^T := \hat{\mathbf{x}}_i$ lies in the range space of $\mathbf{K}_i$, the solution of (35) exists but non-unique. Among the infinite number of solutions, we choose the Moore-Penrose pseudoinverse solution $\mathbf{K}_i^\dagger \hat{\mathbf{x}}_i$.

2) *Recovery for the Coefficients not Possible:* Here, we discuss the possibility of recovering the coefficients under the collision condition. Although it is possible to localize the sources, the recovery of the coefficients $\hat{c}_k \hat{\mathbf{x}}_k$ is not possible due to the fundamental limit in (46).

The DOAs are localized by finding the peak of the dual polynomial vector under the collision condition. For the estimated DOAs $(\hat{w}_1, \dots, \hat{w}_K)$, (2) gives

$$\mathbf{X} = \sum_{k=1}^K \hat{c}_k \mathbf{A}(\hat{w}_k) \otimes \hat{\mathbf{x}}_k^T = \sum_{k=1}^K \mathbf{A}(\hat{w}_k) \otimes \tilde{\mathbf{x}}_k^T, \quad (68)$$

where $\tilde{\mathbf{x}}_k := \hat{c}_k \hat{\mathbf{x}}_k$. Since $\mathbf{Y} = \mathbf{X} = [\mathbf{y}_1 \dots \mathbf{y}_{N_f}]$, the entries in $\tilde{\mathbf{x}}_k$ are recovered by solving $\mathbf{y}_f = \sum_{k=1}^K \mathbf{a}(f, \hat{w}_k) \tilde{\mathbf{x}}_k = [\mathbf{a}(f, \hat{w}_1) \dots \mathbf{a}(f, \hat{w}_K)] [\tilde{\mathbf{x}}_1(f) \dots \tilde{\mathbf{x}}_K(f)]^T$ ($f = 1, \dots, N_f$).

However, when $f = i$, $\mathbf{a}(i, \hat{w}_1) = \mathbf{a}(i, \hat{w}_2)$ from (46). Then, $\mathbf{a}(i, \hat{w}_1) \tilde{\mathbf{x}}_1(i) + \mathbf{a}(i, \hat{w}_2) \tilde{\mathbf{x}}_2(i) = \mathbf{a}(i, \hat{w}_1) [\tilde{\mathbf{x}}_1(i) + \tilde{\mathbf{x}}_2(i)]$. Therefore, we have to decouple $\tilde{\mathbf{x}}_1(i)$ and $\tilde{\mathbf{x}}_2(i)$ based on their sum, which is impossible.

F. Proof for Lemma 4.3

From below (30), $|\bar{\mathbf{x}}_w(i)| = |\mathbf{x}_{w_2}(i)|$, we have

$$\begin{aligned} \|\alpha_i\|_\infty &= \left\| \mathbf{S}_i^{-1} \begin{bmatrix} \text{sign}(c_w^*) \bar{\mathbf{x}}_{w_1}(i) \\ \text{sign}(c_w^*) \bar{\mathbf{x}}_{w_2}(i) \end{bmatrix} \right\|_\infty \\ &\leq \|\mathbf{S}_i^{-1}\|_\infty \left\| \begin{bmatrix} \text{sign}(c_w^*) \mathbf{x}_{w_1}(i) \\ \text{sign}(c_w^*) \mathbf{x}_{w_2}(i) \end{bmatrix} \right\|_\infty \leq \|\mathbf{S}_i^{-1}\|_\infty, \end{aligned} \quad (69)$$

$$\begin{aligned} \|\beta_i\|_\infty &\leq \left\| \mathbf{D}_{i,2}^{-1} \mathbf{D}_{i,1} \mathbf{S}_i^{-1} \begin{bmatrix} \text{sign}(c_w^*) \mathbf{x}_{w_1}(i) \\ \text{sign}(c_w^*) \mathbf{x}_{w_2}(i) \end{bmatrix} \right\|_\infty \\ &\leq \|\mathbf{D}_{i,2}^{-1} \mathbf{D}_{i,1} \mathbf{S}_i^{-1}\|_\infty \leq \|\mathbf{D}_{i,2}^{-1}\|_\infty \|\mathbf{D}_{i,1}\|_\infty \|\mathbf{S}_i^{-1}\|_\infty. \end{aligned} \quad (70)$$

$\|\mathbf{S}_i^{-1}\|_\infty$ is bounded as

$$\begin{aligned} \|\mathbf{S}_i^{-1}\|_\infty &= \|(\mathbf{D}_{i,0} - \mathbf{D}_{i,1} \mathbf{D}_{i,2}^{-1} \mathbf{D}_{i,1})^{-1}\|_\infty \\ &\leq 1/(1 - \|\mathbf{I} - (\mathbf{D}_{i,0} - \mathbf{D}_{i,1} \mathbf{D}_{i,2}^{-1} \mathbf{D}_{i,1})\|_\infty) \\ &\leq 1/[1 - (\|\mathbf{I} - \mathbf{D}_{i,0}\|_\infty + \|\mathbf{D}_{i,1}\|_\infty^2 \|\mathbf{D}_{i,2}^{-1}\|_\infty)]. \end{aligned} \quad (71)$$

Inspired by the proof of [20, Lemma 2.2], the bounds for $\|\mathbf{I} - \mathbf{D}_{i,0}\|_\infty$, $\|\mathbf{D}_{i,1}\|_\infty$, and $\|K''_i(0)\mathbf{I} - \mathbf{D}_{i,2}\|_\infty$ are established (define $d_0 := 6.253 \times 10^{-3}$, $d_1 := 7.639 \times 10^{-2}$, $d_2 := 1.053$, $d_3 := 11/32\pi^2$, where d_0, d_1 , and d_2 are empirical [20] and d_3 is analytical):

$$\begin{aligned} \|\mathbf{I} - \mathbf{D}_{i,0}\|_\infty &\leq \left\| \mathbf{I} - \frac{1}{i} \mathbf{I} \right\|_\infty + \left\| \frac{1}{i} \mathbf{I} - \mathbf{D}_{i,0} \right\|_\infty = 1 - \frac{1}{i} + |K_i(w_1 - w_2)| \\ &= 1 - \frac{1}{i} + \frac{1}{i} |K_1(i(w_1 - w_2))| \leq 1 + \frac{d_0 - 1}{i}, \\ \|\mathbf{D}_{i,1}\|_\infty &= |K'_i(w_1 - w_2)| = |K'_1(i(w_1 - w_2))| \leq d_1 f_c, \\ \|K''_i(0)\mathbf{I} - \mathbf{D}_{i,2}\|_\infty &= |K''_i(w_1 - w_2)| = |K''_1(i(w_1 - w_2))| \leq i d_2 f_c^2, \\ |K''_i(0)| &= \frac{i\pi^2 f_c(f_c + 4)}{3} \geq \frac{i\pi^2 f_c^2}{3} + \frac{4i\pi^2 f_c^2}{3 \cdot 128} = i \cdot d_3 f_c^2. \end{aligned} \quad (72)$$

Therefore, $\|\mathbf{D}_{i,2}^{-1}\|_\infty$ is bounded as follows ($d_4 := 1/(d_3 - d_2) = 0.4275$)

$$\|\mathbf{D}_{i,2}^{-1}\|_\infty \leq \frac{1}{|K''_i(0)| - \|K''_i(0)\mathbf{I} - \mathbf{D}_{i,2}\|_\infty} \leq \frac{1}{i(d_3 - d_2)f_c^2} = \frac{d_4}{i f_c^2}. \quad (73)$$

Then, following (69) and (70), the bounds for $\|\alpha_i\|_\infty$ and $\|\beta_i\|_\infty$ are (define $c_\alpha := 1.008824$, and $c_\beta := 3.294 \times 10^{-2}$):

$$\begin{aligned} \|\alpha_i\|_\infty &\leq \|\mathbf{S}_i^{-1}\|_\infty \leq \frac{i}{1 - d_0 - d_1^2 d_4} := i \cdot c_\alpha, \\ \|\beta_i\|_\infty &\leq \|\mathbf{D}_{i,2}^{-1}\|_\infty \|\mathbf{D}_{i,1}\|_\infty \|\mathbf{S}_i^{-1}\|_\infty \leq \frac{d_1 d_4}{f_c(1 - d_0 - d_1^2 d_4)} := \frac{c_\beta}{f_c}. \end{aligned} \quad (74)$$

$\square$

G. Invertibility of $\mathbf{K}_i$

Using the Schur complement, $\mathbf{K}_i$ is invertible if $\mathbf{D}_{i,2}$ and the Schur complement $\mathbf{S}_i := \mathbf{D}_{i,0} - \mathbf{D}_{i,1} \mathbf{D}_{i,2}^{-1} \mathbf{D}_{i,1}$ are both invertible. To show that, we use the fact that a Hermitian matrix $\mathbf{M}$ is invertible if $\|\mathbf{I} - \mathbf{M}\|_\infty < 1$ [20, eq. (2.12)].

We begin with $\mathbf{D}_{i,2}$. Notice $|K''_i(0)| = i|K''_1(0)| = \frac{i\pi^2 f_c(f_c + 4)}{3}$. Therefore, based on (72),

$$\left\| \mathbf{I} - \frac{\mathbf{D}_{i,2}}{K''_i(0)} \right\|_\infty = \frac{\|\mathbf{D}_{i,2}\|_\infty}{|K''_i(0)|} \leq \frac{i d_2 f_c^2}{i\pi^2 f_c(f_c + 4)/3} < 1, \quad (75)$$

which implies that $\frac{\mathbf{D}_{i,2}}{K_{i,2}''(0)}$ is invertible. Hence, $\mathbf{D}_{i,2}$ is also invertible. We then consider the invertibility of $\mathbf{S}_i$. Based on the triangle inequality,

$$\|\mathbf{I} - \mathbf{S}_i\|_\infty \leq \|\mathbf{I} - \mathbf{D}_{i,0}\|_\infty + \|\mathbf{D}_{i,1}\|_\infty^2 \|\mathbf{D}_{i,2}^{-1}\|_\infty. \quad (76)$$

Hence, to show $\|\mathbf{I} - \mathbf{S}_i\|_\infty < 1$, $\|\mathbf{I} - \mathbf{D}_{i,0}\|_\infty$, $\|\mathbf{D}_{i,1}\|_\infty$, and $\|\mathbf{D}_{i,2}^{-1}\|_\infty$ need to be bounded.

Plugging in the bounds in (72), and (73), we have

$$\|\mathbf{I} - \mathbf{S}_i\|_\infty \leq 1 + \frac{d_0 + d_1^2 d_4 - 1}{i} = 1 + \frac{8.747 \times 10^{-3} - 1}{i} < 1, \quad (77)$$

which implies that $\mathbf{S}_i$ is invertible. $\square$

H. Proof for Theorem 4.2

For simplicity, we assume $K = 2$ in this section. But the theorem can be generalized to $K \geq 2$ if the separation condition is satisfied. Based on the assumption $|\mathbf{x}_{w_1}(i)| = |\mathbf{x}_{w_2}(i)| = 1/\sqrt{N_f}$ for $\forall i \in \{1, \dots, N_f\}$, as long as each entry in the constructed dual polynomial vector satisfies $|\psi^i(w; w_1, w_2)| < 1/\sqrt{N_f}$, then $\|\psi(w)\|_2 < 1$. Therefore, the bounds in Lemma 4.3 (2) further indicate $|\psi^i(w; w_1, w_2)|$ (denote $c_\alpha := 1.008824$, $c_\beta := 3.294 \times 10^{-2}$, $c := \frac{1}{\sqrt{N_f}}$)

$$\begin{aligned} |\psi^i(w; w_1, w_2)| &= \left| \sum_{k \in \{1,2\}} \alpha_{k,i} K_i(w - w_k) + \beta_{k,i} K_i'(w - w_k) \right| \\ &\leq \|\alpha_i\|_\infty \sum_{k \in \{1,2\}} |K_i(w - w_k)| + \|\beta_i\|_\infty \sum_{k \in \{1,2\}} |K_i'(w - w_k)| \\ &\leq c[c_\alpha \sum_{k \in \{1,2\}} \frac{|K_1(i(w - w_k))|}{i} + \frac{c_\beta}{f_c} \sum_{k \in \{1,2\}} |K_1'(i(w - w_k))|] \\ &= c[c_\alpha \sum_{k \in \{1,2\}} |K_1(i(w - w_k))| + \frac{c_\beta}{f_c} \sum_{k \in \{1,2\}} |K_1'(i(w - w_k))|] \\ &= c[\sum_{k \in \{1,2\}} c_\alpha |K_1(i(w - w_k) \bmod 1)| + \frac{c_\beta}{f_c} |K_1'(i(w - w_k) \bmod 1)|] \end{aligned} \quad (78)$$

When $i = 1$,

$$|\psi^1(w; w_1, w_2)| \leq c[c_\alpha \sum_{k \in \{1,2\}} |K_1(w - w_k)| + \frac{c_\beta}{f_c} \sum_{k \in \{1,2\}} |K_1'(w - w_k)|]. \quad (79)$$

We show $c[c_\alpha \sum_{k \in \{1,2\}} |K_1(w - w_k)| + \frac{c_\beta}{f_c} \sum_{k \in \{1,2\}} |K_1'(w - w_k)|] < \frac{1}{\sqrt{N_f}}$ by applying [20, Lemma 2.3 and 2.4]. We consider both the near and far regions. The near region $\mathcal{T}_{\text{near}}$ and far region $\mathcal{T}_{\text{far}}$ are defined as $\mathcal{T}_{\text{near}} := \cup_{k=1}^2 [w_k - \nu, w_k + \nu]$ and $\mathcal{T}_{\text{far}} := [0, 1] \setminus \mathcal{T}_{\text{near}}$, where $\nu = \frac{0.1649}{f_c}$.

For $\mathcal{T}_{\text{far}}$, based on [20, Lemma 2.4]

$$\begin{aligned} c_\alpha \sum_{k \in \{1,2\}} |K_1(w - w_k)| + \frac{c_\beta}{f_c} \sum_{k \in \{1,2\}} |K_1'(w - w_k)| \\ \leq 0.99992 < 1. \end{aligned} \quad (80)$$

Therefore,

$$|\psi^1(w; w_1, w_2)| \leq c[c_\alpha \sum_{k \in \{1,2\}} |K_1(w - w_k)| + \frac{c_\beta}{f_c} \sum_{k \in \{1,2\}} |K_1'(w - w_k)|] < \frac{1}{\sqrt{N_f}}$$

If $i > 1$, the only difference between the last line of (78) and the right hand side of (79) is the dilation of K_1 and K_1' . This

indicates the i -th entry is a special case for $i = 1$. Therefore, $|\psi^i(w; w_1, w_2)| < c = 1/\sqrt{N_f}$ will also hold for $i > 1$. Hence, in $\mathcal{T}_{\text{far}}$, $\|\psi(w)\|_2 < 1$ for $w \notin \mathcal{W}$.

For $\mathcal{T}_{\text{near}}$, inspired by the proof in [20, Lemma 2.3], we show the strict concavity of $|\psi^i(w; w_1, w_2)|$. We have

$$\begin{aligned} \psi_R^i(w) \psi_R^{i''}(w) + |\psi^{i'}(w)|^2 + |\psi_I^i(w)| |\psi_I^{i''}(w)| \\ \leq -9.291 \times 10^{-2} (i f_c / \sqrt{N_f})^2 < 0 \end{aligned} \quad (81)$$

and

$$\begin{aligned} \frac{d^2 |\psi^i|(w)}{dw^2} &= - \frac{(\psi_R^i(w) \psi_R^{i'}(w) + \psi_I^i(w) \psi_I^{i'}(w))^2}{|\psi^i(w)|^3} \\ &+ \frac{\psi_R^i(w) \psi_R^{i''}(w) + |\psi^{i'}(w)|^2 + |\psi_I^i(w)| |\psi_I^{i''}(w)|}{|\psi^i(w)|} < 0. \end{aligned} \quad (82)$$

Since $\psi^{i'}(w_1) = \psi^{i'}(w_2) = 0$, local strict concavity will imply $|\psi^i(w; w_1, w_2)| < 1/\sqrt{N_f}$ in $\mathcal{T}_{\text{near}}$. $\square$



Yifan Wu received the M.S. from the University of California, San Diego in 2021. He is now pursuing Ph.D. degree in the University of California, San Diego. His research interest includes statistical signal processing, machine learning, and optimization theory.



Michael B. Wakin (Fellow, IEEE) is a Professor of Electrical Engineering at the Colorado School of Mines. Dr. Wakin received a Ph.D. in electrical engineering in 2007 from Rice University. He was an NSF Mathematical Sciences Postdoctoral Research Fellow at Caltech from 2006-2007, an Assistant Professor at the University of Michigan from 2007-2008, and a Ben L. Fryrear Associate Professor at Mines from 2015-2017. His research interests include signal and data processing using sparse, low-rank, and manifold-based models.

In 2008, Dr. Wakin received the DARPA Young Faculty Award for his research in compressive multi-signal processing for environments such as sensor and camera networks. In 2012, Dr. Wakin received the NSF CAREER Award for research into dimensionality reduction techniques for structured data sets. Dr. Wakin is a recipient of the Best Paper Award and the Signal Processing Magazine Best Paper Award from the IEEE Signal Processing Society. He has served as an Associate Editor for IEEE Signal Processing Letters and IEEE Transactions on Signal Processing, and he is currently a Senior Area Editor for IEEE Transactions on Signal Processing.



Peter Gerstoft (Fellow, IEEE) received the Ph.D. from the Technical University of Denmark, Lyngby, Denmark, in 1986. Since 1997, he has been with the University of California, San Diego. His current research interests are signal processing and machine learning applied to acoustic, seismic, and electromagnetic signals. For more information see <http://noiselab.ucsd.edu>.