

Non-Uniform Array and Frequency Spacing for Regularization-Free Gridless DOA

Yifan Wu¹, Michael B. Wakin², *Fellow, IEEE*, and Peter Gerstoft³, *Fellow, IEEE*

Abstract—Gridless direction-of-arrival (DOA) estimation with multiple frequencies can be applied in acoustics source localization problems. We formulate this as an atomic norm minimization (ANM) problem and derive an equivalent regularization-free semi-definite program (SDP) thereby avoiding regularization bias. The DOA is retrieved using a Vandermonde decomposition on the Toeplitz matrix obtained from the solution of the SDP. We also propose a fast SDP program to deal with non-uniform array and frequency spacing. For non-uniform spacings, the Toeplitz structure will not exist, but the DOA is retrieved via irregular Vandermonde decomposition (IVD), and we theoretically guarantee the existence of the IVD. We extend ANM to the multiple measurement vector (MMV) cases and derive its equivalent regularization-free SDP. Using multiple frequencies and the MMV model, we can resolve more sources than the number of physical sensors for a uniform linear array. Numerical results demonstrate that the regularization-free framework is robust to noise and aliasing, and it overcomes the regularization bias.

Index Terms—Atomic norm minimization, multiple frequencies, Vandermonde decomposition, DOA estimation.

I. INTRODUCTION

DIRECTION-OF-ARRIVAL (DOA) estimation is an important topic in sensor array processing [1] that has a broad range of applications in wireless communication [2], radar [3], remote sensing, etc. Conventional DOA estimation methods (e.g. multiple signal classification (MUSIC) [4], and estimation of signal parameters via rotational invariant techniques (ESPRIT) [5]) are mainly developed for narrowband signals. In the past few decades, some wideband DOA estimation methods have been proposed [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16]. Recently proposed methods based on sparse recovery and a multi-frequency model [8], [10] have demonstrated superior per-

formance in wideband DOA estimation problems. Before introducing the contributions of this paper, we review the relevant prior works.

A. Related Work

1) *Wideband DOA Estimation and Multiple Frequencies:* Wideband signal DOA estimation has been studied for decades [6], [11], [12], [13], [14]. In [6], a subspace-based wideband DOA estimation approach, the incoherent signal subspace method (ISSM), was proposed. The coherent signal subspace method (CSSM) [11] led to improved performance compared to ISSM. A broadband spatial-spectrum estimation approach [12] overcame the peak bias and source spectral content sensitivity from CSSM. Variants of CSSM, such as the weighted average of signal subspaces method [13] and the test of orthogonality of projected subspaces method [14] were also proposed. Recently, some wideband DOA estimation methods based on sparse recovery have also been developed [7], [8], [9], [10], [15], [16], [17], [18], [19]. These sparsity-based methods have demonstrated superior performance compared to conventional methods and generally require much fewer samples.

The multi-frequency model [7], [8], [9], [10], [15], [16], [20] has shown success in modeling wideband signals. The multi-frequency model uses N_f (rather than 1) temporal frequency bins in a frequency set $\{F_1, \dots, F_{N_f}\}$ to characterize a wideband signal. All these frequencies are used for estimation, as opposed to using a single frequency under the narrowband model. One challenge for multi-frequency processing is aliasing [8], [10], which will be present when the receiver spacing is greater than the half wavelength of the highest frequency. The performance of a DOA estimation method may degrade significantly in the presence of aliasing. In [15], the authors present an aliasing-free DOA estimation method based on sparse signal recovery. In [8], [9], [16], wideband signal DOA estimation based on sparse Bayesian learning (SBL) with multiple frequencies is proposed [8] and applied to matched field processing [16] and robust ocean acoustic localization [9]. A DOA estimation method based on low-rank structured matrix completion for sparse arrays under the multi-frequency model is proposed in [20]. A joint localization and dereverberation method based on sparse regularization is also proposed in [7] for room source localization and tracking.

2) *Atomic Norm Minimization (ANM):* ANM was initially proposed in [21] as a general framework for promoting sparse signal decompositions. The main benefit of ANM is that it

Manuscript received 31 August 2023; revised 15 January 2024; accepted 26 March 2024. Date of publication 8 April 2024; date of current version 23 April 2024. This work was supported in part by the NSF under Grant CCF-1704204 and Grant CCF-2203060, and in part by the Office of Naval Research (ONR) under Grant N00014-21-1-2267. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Pia Addabbo. (Corresponding author: Yifan Wu.)

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

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

Digital Object Identifier 10.1109/TSP.2024.3386018

TABLE I
SURVEY OF SDPs USED IN [10] AND THIS WORK

	[10]	This Work
Assumption	ULA and uniform frequency	NUA and NUF
Procedure	Dual SDP $\rightarrow$ Polynomial rooting	Primal SDP $\rightarrow$ IVD
Model	SMV	MMV
Noise-free SDPs	Dual uniform (20) (SDP equivalent to ANM)	Dual uniform (13) (SDP equivalent to MMV-MF ANM)
	Fast dual (27)–(28) (extension of (20); nonuniform not tested)	Fast dual (17), extension of (13); accommodates NUA/NUF
	Full-dimension primal (29) (dual of (20); uniform case)	Fast primal (19) (dual of (17); accommodates NUA/NUF)
		Full-dimension primal (24) (dual of (13); uniform case)
Noisy SDPs	Dual uniform (21), robust version of (20)	Dual uniform (15) (robust version of (13))

overcomes the grid mismatch error that plagues grid-based methods. The pioneering ANM paper [22] proposed an optimization-based continuous (temporal) frequency estimation method and provided a theoretical guarantee when full data are available. The authors in [23] studied continuous temporal frequency estimation based on randomly sampled data for the single measurement vector (SMV) case. ANM for multiple measurement vectors (MMVs) under the uniform (or equispaced) time samples (analogous to a uniform linear array, or ULA) setup was studied in [24], [25], [26], and it was extended to the non-uniform array (NUA) setting in [27]. It was also extended to multiple frequencies for wideband DOA estimation in [10], [28], [29]. The sample complexity of modal analysis with random temporal compression was established in [30]. See [31] for a comprehensive overview of ANM and its applications.

3) *Lifting*: A “lifting trick” was applied to an ANM problem for point spread function (PSF) estimation in [32]. This allowed a bilinear inverse problem to be transformed into a linear inverse problem by assuming the PSF lies in a known low-dimensional space. This lifting idea was also applied to biconvex compressive sensing problems in [33] to formulate them as convex programs.

4) *Non-Uniform Array and More Sources Than Sensors*: An NUA enables the possibility to resolve more sources than the number of physical sensors. Early works involving NUA include the minimum redundancy array (MRA) [34], and the minimum holes array (MHA) [35]. For a given number of sources, MRA and MHA require an extensive search through all possible sensor combinations to find the optimal design. Recently, a new structure of NUA, known as a co-prime array [36], was developed. The co-prime array has a closed-form expression for the sensor positions so that the exhaustive search over the sensor combinations is avoided. The nested array [37] and co-array [38] based approaches were also proposed to detect more sources than the number of sensors.

An alternative way to resolve more sources than sensors is to use fourth-order cumulants [39], [40]. However, this approach is limited to non-Gaussian sources. In [41], with the help of the Khatri-Rao (KR) product and assuming quasi-stationary sources, it was shown that one can identify up to $2N - 1$ sources using an N -element ULA without computing higher-order statistics. Unfortunately, the quasi-stationary assumption does not apply to stationary sources.

B. Our Contributions

In previous work [10], we developed a gridless DOA estimation method for the multi-frequency model based on ANM. This was formulated as a semi-definite program (SDP) problem so that ANM is solved using off-the-shelf SDP solvers, e.g. CVX [42]. The DOAs are retrieved by finding the roots of the dual polynomial. The dual polynomial served as a certificate for the optimality and an interpolation method that constructed the dual certificate was presented.

In this work, we propose a wideband DOA estimation framework that significantly expands the applicability from [10]. Our contribution is summarized in the following respects (see also Table I).

1) *Regularization-Free Framework*: An SDP that is equivalent to ANM is formulated in [10] based on the dual atomic norm and the definition of the dual polynomial in the noise-free case. When noise is present, a common strategy in ANM works is giving some tolerance to the constraints using regularization in the SDP [23], [25], [43]. In addition to the challenges from the noise, [10] shows that for array spacing above half a wavelength of the highest frequency and with multiple sources, the performance may degrade remarkably due to a phenomenon termed near collision [10]. To mitigate near collisions, an $\ell_{1,2}$ regularization term is added. Although these regularization terms prevent failures due to noise or near collisions, they lead to bias. The performance of ANM degrades due to such bias compared to the competing method SBL [44], especially at a low signal-to-noise ratio (SNR).

Although most ANM methods promote robustness to noise by adding a regularization term, [27] demonstrates that it is possible to deal with the noise by solving a noise-free optimization problem. In [27], the authors propose a two-step DOA estimation approach. The first step is to apply the alternating projection (AP) algorithm to solve a *noise-free* optimization problem and obtain a matrix with an irregular Toeplitz structure. The second step computes an irregular Vandermonde decomposition (achieved by generalized root-MUSIC) to retrieve the DOAs from the irregular Toeplitz matrix. Although the optimization problem solved in the first step does not have explicit robustness to the noise, the second step enables the method to work in noisy cases. This method effectively avoids the explicit bias and the non-trivial effort required to tune the regularization parameter.

Inspired by [27], we formulate the dual problem of the *noise-free* SDP in [10, eq. (20)] *without regularization*. This problem

is again an SDP and we deem it as the *primal domain SDP* (since [10] formulates its SDP in the dual domain). Both the primal and dual SDPs involve a certain lifting mapping which we define. Though different from [32], this lifting embeds the problems in a higher dimensional space where it is more natural to combine all frequencies.

Solving this SDP gives a Toeplitz matrix, and the DOAs are further retrieved by Vandermonde decomposition of this Toeplitz matrix. One computational method for Vandermonde decomposition is root-MUSIC [45], and it has robustness to both noise and near collisions. Therefore, with the help of “post-SDP processing” (root-MUSIC), regularization is avoided and no prior knowledge of the noise is needed. At low SNRs in simulation, the method can achieve a better performance than competing methods, and it approaches the Cramér-Rao bound (CRB) [46], [47] for Gaussian noise.

2) *Non-Uniform Frequencies and Irregular Vandermonde Decomposition:* We also develop a fast SDP for the primal domain SDP. The fast SDP is derived based on the dual problem of the fast algorithm in [10]. The fast algorithm can not only improve the speed but also can extend the method to the non-uniform frequency (NUF) case. In this case, the DOAs are encoded in a matrix with an *irregular Toeplitz* structure. We apply the irregular Vandermonde decomposition (IVD) [27] to this matrix to retrieve the DOAs. Furthermore, we provide a theoretical guarantee for the existence of the IVD which is not shown in [27]. While it is mentioned in [10] that the fast dual algorithm proposed therein can be applied to the NUF case, this is not tested in [10], and our experiments (see Fig. 5) indicate that the fast primal method is more effective.

3) *Multiple Snapshots and More Sources Than Sensors:* The method in [10] is developed under the SMV case. Prior works show that MMVs can give improved performance [24], [25], [48]. That motivates us to extend the framework in [10] to the MMV case. In the MMV setting, the received signal from the sensor array is a three-dimensional tensor (sensors $\times$ snapshots $\times$ frequencies). Based on the signal model, we formulate the corresponding ANM problem and derive the SDP (in the dual domain) that is equivalent to the ANM. The dual problem of the SDP is then derived to obtain the SDP in the primal domain. The purpose of the primal SDP is to enhance the robustness to the noise and near collision without regularization.

The multi-frequency setup also enables resolving more sources than sensors case in the ULA setting. The maximum number of uniquely identifiable sources in an N_M -element ULA is $N_M - 1$ [48, Sec. 11.2.3] for the single-frequency case. Co-prime array techniques [36] can break through such a limit with a carefully designed array structure, enabling the resolution of more sources than the number of sensors. We show that it is possible to resolve more sources than sensors *with a ULA under the multi-frequency model*. The physical intuition is that multiple frequencies increase the diversity of the harmonics and these “new harmonics” can serve as extra “virtual sensors” in a large virtual array. Due to this intrinsic property, it is possible to break through such a bottleneck in the ULA setup. In many practical scenarios, the array geometry is

fixed and ULA is one of the most commonly used arrays. This result has a practical impact and demonstrates the benefit of multi-frequency processing.

In summary, the framework proposed is superior to [10] in terms of generality, practicality, performance, and complexity. Our work also demonstrates the possibility of resolving more sources than sensors *under the ULA setup* which is an important merit of the multi-frequency model. This work also significantly extends our preliminary conference paper [49]. It considers the non-uniform array (NUA) case and includes more extensive theoretical analysis and numerical experiments. It also discusses the connection between atomic ℓ_0 norm minimization and rank minimization.

C. Notation

Throughout the paper, the following notation is adopted. Boldface letters are used to represent matrices and vectors. Conventional notations $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^*$, $\langle \cdot, \cdot \rangle_{\mathbb{R}}$, and $\langle \cdot, \cdot \rangle$ 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. $\|\cdot\|_p$, $\|\cdot\|_F$, and $\|\cdot\|_{\text{HS}}$ are used to express vector ℓ_p norm, matrix Frobenius norm, and Hilbert-Schmidt norm for the tensor (for a 3D tensor $\|\mathcal{A}\|_{\text{HS}} = \sqrt{\sum_{ijk} |a_{ijk}|^2}$). For a Hermitian matrix $\mathbf{A}$, $\mathbf{A} \succeq 0$ means $\mathbf{A}$ is a positive semidefinite matrix. The imaginary unit is denoted by $j = \sqrt{-1}$.

II. PRELIMINARIES

When multiple snapshots are available, DOA estimation methods can have improved performance [24], [25], [48]. In this section, we extend the SMV multi-frequency ANM framework for gridless DOA estimation from [10] to the MMV setting; we refer to the resulting framework as the *MMV-MF model*. This model will help us explore the possibility of having more sources than the sensors in Sec. V.

A. Assumptions

The following assumptions are made for the array configuration and signal model:

- 1) The sensors comprise a linear array with positions drawn from a uniform grid $\{0, 1, \dots, N_M - 1\} \cdot d$, where d is the sensor spacing unit. We let $\mathcal{M} \subseteq \{0, 1, \dots, N_M - 1\}$ denote the indices of the actual sensors; the resulting positions are thus $\{m \cdot d | m \in \mathcal{M}\}$. We define $N_m := |\mathcal{M}| \leq N_M$ as the number of sensors. When all sensors are present, $N_m = N_M$, and we have a *uniform linear array (ULA)* case. When only some sensors are present, $N_m < N_M$, and we have a *nonuniform array (NUA)* case.
- 2) The sources have temporal frequency components drawn from a uniform grid $\{1, \dots, N_F\} \cdot F_1$, where F_1 is the spacing between frequencies. Let $\lambda_1 := c/F_1$ denote the wavelength corresponding to F_1 , where c is the propagation speed. We assume $\lambda_1 = 2d$ where d is the sensor spacing unit above; equivalently, $d = \frac{c}{2F_1}$. This spacing is for simplifying the derivation and can be relaxed to any

$d \leq \frac{\lambda_1}{2}$ (see [10] for details). We let $\mathcal{F} \subseteq \{1, \dots, N_F\}$ denote the indices of the active source frequencies; the resulting frequencies are thus $\{f \cdot F_1 | f \in \mathcal{F}\}$ and the wavelengths are $\{\lambda_1/f | f \in \mathcal{F}\}$. We define $N_f := |\mathcal{F}| \leq N_F$ to be the number of active source frequencies. When all frequencies are active, $N_f = N_F$, and we refer to this as the *uniform frequency* case. When only some frequencies are active, $N_f < N_F$, and we refer to this as the *nonuniform frequency (NUF)* case.

- 3) Suppose there are N_l snapshots (time samples) received by each sensor. The source amplitude for the f -th frequency ($f \in \mathcal{F}$) is $\mathbf{x}_w(f) = [x_w^{(1)}(f) \dots x_w^{(N_l)}(f)]^T \in \mathbb{C}^{N_l}$.
- 4) There are K active uncorrelated sources impinging on the array from unknown directions of arrival (DOAs) θ , or in directional cosines

$$w := F_1 d \cos(\theta) / c = \cos(\theta) / 2. \quad (1)$$

B. MMV-MF Model

We begin by considering the case of a ULA with uniform frequencies, i.e., $N_m = N_M$ and $N_f = N_F$. (We incorporate the NUA and NUF cases in Section IV.) The received signals can be arranged into a tensor $\mathcal{Y} \in \mathbb{C}^{N_M \times N_l \times N_F}$ (sensors $\times$ snapshots $\times$ frequencies) with the following structure:

$$\mathcal{Y} = \mathcal{X} + \mathcal{N} \quad (2)$$

$$\begin{aligned} \mathcal{X} &= \sum_w c_w [\mathbf{a}(1, w) \mathbf{x}_w^T(1) | \dots | \mathbf{a}(N_F, w) \mathbf{x}_w^T(N_F)] \\ &= \sum_w c_w \mathbf{A}(w) * \mathbf{X}_w^T \end{aligned} \quad (3)$$

where $\mathbf{a}(f, w) = [1 e^{-j2\pi w f} \dots e^{-j2\pi w f(N_M-1)}]^T = [1 z^f \dots z^{f(N_M-1)}]^T \in \mathbb{C}^{N_M}$ ($z := e^{-j2\pi w}$) is the array manifold vector for the f -th frequency. $\mathcal{N} \in \mathbb{C}^{N_M \times N_l \times N_F}$ denotes additive Gaussian uncorrelated noise in (2). Denote $\mathbf{A}(w) = [\mathbf{a}(1, w) \dots \mathbf{a}(N_F, w)] \in \mathbb{C}^{N_M \times N_F}$ and $\mathbf{X}_w = [\mathbf{x}_w(1) \dots \mathbf{x}_w(N_F)]^T \in \mathbb{C}^{N_F \times N_l}$. $\mathbf{A}(w) * \mathbf{X}_w^T$ is the “reshaped Khatri-Rao product” defined as $[\mathbf{A}(w) * \mathbf{X}_w^T]_{:,f} := \mathbf{a}(f, w) \mathbf{x}_w^T(f)$ ($f = 1, \dots, N_F$). When $N_l = 1$, the above matches the SMV model in [10]. We assume $\|\mathbf{X}_w\|_F = 1$, as the coefficient c_w can be used to absorb any other scaling of the norm of the source amplitude via the product $c_w \mathbf{X}_w$.

Finally, we define

$$N = N_F(N_M - 1) + 1, \quad (4)$$

noting that $N_F(N_M - 1)$ appears in the largest exponent of any array manifold vector used in the MMV-MF model. Consequently, N will determine the size of certain SDP formulations such as (13).

C. Collision and Near Collision

A challenge for multi-frequency processing is the risk of a phenomenon known as *collision*, which occurs when, at some frequencies, the array manifold vectors for two DOAs coincide

due to aliasing. Two DOAs w_1 and w_2 are said to have a *collision* in the f -th frequency if [10, eq. (46)]

$$\mathbf{a}(f, w_1) = \mathbf{a}(f, w_2). \quad (5)$$

Such a collision occurs whenever w_1 and w_2 satisfy [10, eq. (47)]

$$|w_1 - w_2| = \frac{k}{f} \quad (f \in \mathcal{F}, f > 1). \quad (6)$$

A *near collision* is said to occur when [10, eq. (50)]

$$|w_1 - w_2| \approx \frac{k}{f} \quad (f \in \mathcal{F}, f > 1). \quad (7)$$

D. Irregular Vandermonde and Toeplitz Matrices

Define some integer-valued vector $\gamma = [\gamma_1 \dots \gamma_{N_\gamma}]^T \in \mathbb{Z}^{N_\gamma}$, complex-valued vector $\mathbf{z} = [z_1 \dots z_{N_z}]^T \in \mathbb{C}^{N_z}$, and $\mathbf{w}(\gamma, \mathbf{z}) := [z^{\gamma_1} \dots z^{\gamma_{N_\gamma}}]^T$. For arbitrary dimensions N_γ and N_z , an *irregular Vandermonde matrix* of size $N_\gamma \times N_z$ is a matrix having the form [27, eq. (25)]

$$\begin{aligned} \mathbf{W} &= \mathbf{W}(\gamma, \mathbf{z}) = [\mathbf{z}^{\gamma_1} \dots \mathbf{z}^{\gamma_{N_\gamma}}]^T \\ &= [\mathbf{w}(\gamma, z_1) \dots \mathbf{w}(\gamma, z_{N_z})]. \end{aligned} \quad (8)$$

Note that when the entries of γ form an arithmetic progression, specifically $\gamma = [0 \dots N_\gamma - 1]^T$, $\mathbf{W}(\gamma, \mathbf{z})$ forms a regular Vandermonde matrix.

An (N_γ, N_z) -*irregular Toeplitz matrix* is any matrix $\mathbf{T} \in \mathbb{C}^{N_\gamma \times N_\gamma}$ that can be constructed from an irregular Vandermonde matrix as follows [27, eq. (27)]:

$$\mathbf{T} = \mathbf{W}(\gamma, \mathbf{z}) \mathbf{D} \mathbf{W}(\gamma, \mathbf{z})^H, |\mathbf{z}| = 1, \quad (9)$$

where $\gamma \in \mathbb{Z}^{N_\gamma}$ and $\mathbf{z} \in \mathbb{C}^{N_z}$, and where $\mathbf{D} \in \mathbb{R}^{N_z \times N_z}$ is a diagonal matrix. We refer to (9) as an *irregular Vandermonde decomposition (IVD)*. Note that any $N_\gamma \times N_\gamma$ positive semi-definite regular Toeplitz matrix $\mathbf{T}$ with rank N_z has a regular Vandermonde decomposition of the form (9) in which $\gamma \in \mathbb{Z}^{N_\gamma}$ is an arithmetic progression.

III. ATOMIC NORM MINIMIZATION FOR MMV-MF

In this section, we formulate the atomic norm minimization problem for the MMV-MF model. Then, we derive an equivalent SDP that makes the proposed framework computationally feasible. We note that the ANM we derive has multiple frequencies while the ANM in [24], [25] operates at a single frequency. The multi-frequency model can be applied to wideband signals, while their MMV model can only be applied to narrowband signals.

Define the atomic set

$$\mathcal{A} = \{\mathbf{A}(w) * \mathbf{X}_w^T \mid w \in [-1/2, 1/2], \|\mathbf{X}_w\|_F = 1\}. \quad (10)$$

The atomic norm of a tensor $\mathcal{X} \in \mathbb{C}^{N_M \times N_l \times N_F}$ is defined as $\|\mathcal{X}\|_{\mathcal{A}} := \inf \{\sum_w |c_w| \mid \mathcal{X} = \sum_w c_w \mathbf{A}(w) * \mathbf{X}_w^T, \|\mathbf{X}_w\|_F = 1\}$. The *atomic norm minimization (ANM)* problem for the noise-free case can be expressed as

$$\min_{\mathcal{X}} \|\mathcal{X}\|_{\mathcal{A}} \quad \text{s.t.} \quad \mathcal{Y} = \mathcal{X}. \quad (11)$$

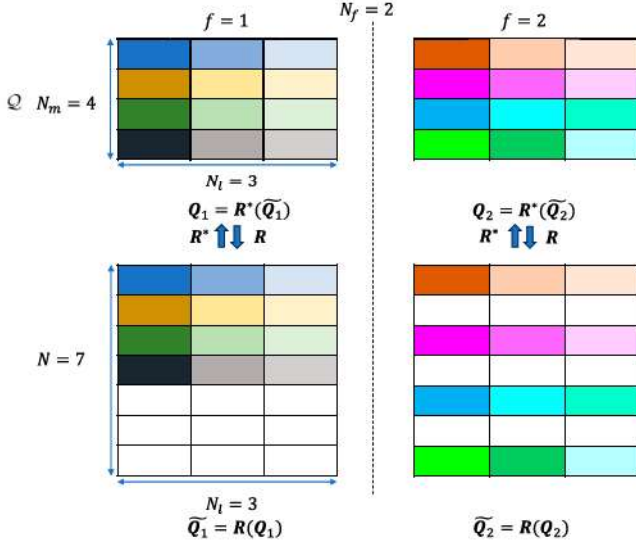


Fig. 1. Demonstration for the lifting mapping $\mathcal{R}(\cdot)$ and its adjoint mapping $\mathcal{R}^*(\cdot)$. $N_m = N_M = 4$, $N_l = 3$, $N_f = N_F = 2$, $N = (N_M - 1)N_F + 1 = 7$.

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

$$\min_{\mathcal{X}} \|\mathcal{X}\|_{\mathcal{A}} \quad \text{s.t.} \quad \|\mathcal{Y} - \mathcal{X}\|_{\text{HS}} \leq \eta. \quad (12)$$

The following proposition guarantees that (11) is equivalent to an SDP problem.

Proposition III.1: Problem (11) is equivalent to the following SDP problem

$$\begin{aligned} \max_{\mathcal{Q}, \mathbf{P}_0} \langle \mathcal{Q}, \mathcal{Y} \rangle_{\mathbb{R}} \quad \text{s.t.} \quad & \begin{bmatrix} \mathbf{P}_0 & \tilde{\mathcal{Q}} \\ \tilde{\mathcal{Q}}^H & \mathbf{I}_{N_l N_F} \end{bmatrix} \succeq 0, \\ & \sum_{i=1}^{N-k} \mathbf{P}_0(i, i+k) = \delta_k, \tilde{\mathcal{Q}} = [\mathcal{R}(\mathbf{Q}_1) \dots \mathcal{R}(\mathbf{Q}_{N_F})], \end{aligned} \quad (13)$$

where $\mathcal{Q} = [\mathbf{Q}_1 | \dots | \mathbf{Q}_{N_F}] \in \mathbb{C}^{N_M \times N_l \times N_F}$ is the dual variable, $\mathbf{P}_0 \in \mathbb{C}^{N \times N}$, $\tilde{\mathcal{Q}} = [\tilde{\mathbf{Q}}_1 \dots \tilde{\mathbf{Q}}_{N_F}] \in \mathbb{C}^{N \times N_l N_F}$, and $\tilde{\mathbf{Q}}_f = \mathcal{R}(\mathbf{Q}_f) : N_M \times N_l \rightarrow N \times N_l$ is a mapping defined as

$$\mathcal{R}(\mathbf{Q}_f)(i, l) = \begin{cases} \mathbf{Q}_f(m, l) & \text{for } (i, l) = (f(m-1)+1, l) \\ 0 & \text{otherwise.} \end{cases} \quad (14)$$

Proof: See Appendix A.

Fig. 1 demonstrates the mapping $\mathcal{R}$. Across all frequencies, $\mathcal{R} : N_M \times N_l \times N_F \rightarrow N \times N_l N_F$ is a linear mapping and can be expressed as a tall binary matrix multiply $\text{vec}(\tilde{\mathcal{Q}}) = \mathbf{R} \text{vec}(\mathcal{Q})$. The transpose of the matrix $\mathbf{R}$ describes the behavior of the adjoint operator $\mathcal{R}^* : N \times N_l N_F \rightarrow N_M \times N_l \times N_F$, which is also demonstrated in Fig. 1.

To provide intuition for the role of $\mathcal{R}$, recall from (3) that the array manifold vectors in the MMV-MF model are frequency-dependent, and so the rows of different slices of $\mathcal{X}$ correspond to different space-frequency products. After lifting the

dual variable tensor $\mathcal{Q}$ to a higher-dimensional space, however, every row corresponds to the same space-frequency product $f(m-1)$, allowing $\tilde{\mathcal{Q}}$ to play a similar role in the SDP to the dual variables in more conventional ANM formulations.

In the noisy case, the equivalent SDP of (12) is the regularized version of (13):

$$\begin{aligned} \max_{\mathcal{Q}, \mathbf{P}_0} \langle \mathcal{Q}, \mathcal{Y} \rangle_{\mathbb{R}} - \eta \|\mathcal{Q}\|_{\text{HS}} \quad \text{s.t.} \quad & \begin{bmatrix} \mathbf{P}_0 & \tilde{\mathcal{Q}} \\ \tilde{\mathcal{Q}}^H & \mathbf{I}_{N_l N_F} \end{bmatrix} \succeq 0, \\ & \sum_{i=1}^{N-k} \mathbf{P}_0(i, i+k) = \delta_k, \tilde{\mathcal{Q}} = [\mathcal{R}(\mathbf{Q}_1) \dots \mathcal{R}(\mathbf{Q}_{N_F})], \end{aligned} \quad (15)$$

where η depends on the noise level and is the same as in (12).

IV. REGULARIZATION-FREE SDP AND FAST ALGORITHM

In the previous section, we obtained an SDP that is equivalent to ANM. This SDP relies on the dual norm and dual polynomial (see Appendix A for the dual norm (39) and dual polynomial (40)), and so we deem the SDP in (13) as the *dual SDP*. We now derive the dual problem of the SDP in Sec. III; we deem this as the *primal SDP*. The benefit of the primal SDP is that it is regularization-free and it thus avoids regularization bias in (15). In numerical experiments, this primal SDP is inherently robust to noise and near collisions. Further, we derive a fast, reduced-dimension version of the primal SDP. The fast program improves the speed, and more importantly, it relaxes the requirements that the sensor positions and temporal frequencies be uniform.

A. Non-Uniform Array (NUA) and Non-Uniform Frequency (NUF) Settings

We focused on the ULA and uniform frequency case. However, in general, the array spacing and frequency may not be uniform. Thus we generalize the proposed framework to NUA and NUF cases.

Recall that $\mathcal{F} \subseteq \{1, \dots, N_F\}$ denotes the indices of the active source frequencies, with $N_f := |\mathcal{F}| \leq N_F$ denoting the number of active frequencies. The nonuniform frequency (NUF) case corresponds to the scenario where $N_f < N_F$, i.e., only some of the frequencies are active. Similarly, $\mathcal{M} \subseteq \{0, 1, \dots, N_M - 1\}$ denotes the indices of the sensors, with $N_m := |\mathcal{M}| \leq N_M$ denoting the number of sensors. The nonuniform array (NUA) case corresponds to the scenario where $N_m < N_M$, i.e., only some sensors are present.

Recall that every exponent in an array manifold vector from the MMV-MF model involves a product of one temporal frequency and one sensor position. To capture all such products in the nonuniform setting, we define a spatial-frequency index set $\mathcal{U}$ as follows:

$$\mathcal{U} = \{m \cdot f | m \in \mathcal{M}, f \in \mathcal{F}\}. \quad (16)$$

The cardinality of this set $N_u := |\mathcal{U}| \leq N$, with N is defined in (4). In many settings, $N_u \ll N$. In later sections, we see that the size of the fast SDP depends on N_u , and its complexity is greatly reduced compared to the original SDP.

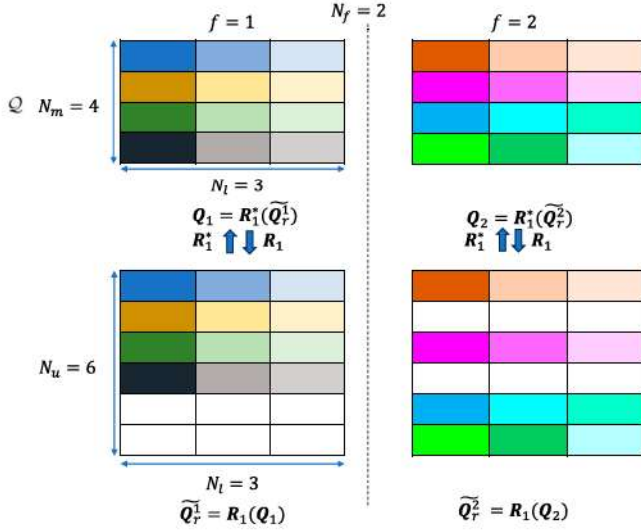


Fig. 2. Demonstration for the $\mathcal{R}_1(\cdot)$ mapping and its adjoint mapping $\mathcal{R}_1^*(\cdot)$. $N_m = N_M = 4$, $N_l = 3$, $N_f = N_F = 2$, $\mathcal{U} = \{0, 1, 2, 3, 4, 6\}$, $N_u = |\mathcal{U}| = 6$.

B. Fast Dual SDP for the NUA and NUF Case

Proposition III.1 gives the SDP for the ULA and uniform frequency case. We generalize the SDP to the NUA and NUF cases. The SDP is not only more general but also can reduce the complexity in the ULA and uniform frequency case. Inspired by the fast algorithm in [10, Sec. III-F], the SDP in this section is considered the fast algorithm for MMV.

For NUA and NUF, the measurement tensor $\mathcal{Y} \in \mathbb{C}^{N_m \times N_l \times N_f}$ and the SDP in Proposition III.1 is generalized as

$$\max_{\mathcal{Q}, \mathbf{P}_{r0}} \langle \mathcal{Q}, \mathcal{Y} \rangle_{\mathbb{R}} \quad \text{s.t.} \quad \begin{bmatrix} \mathbf{P}_{r0} & \tilde{\mathbf{Q}}_r \\ \tilde{\mathbf{Q}}_r^H & \mathbf{I}_{N_l N_f} \end{bmatrix} \succeq 0, \\ \sum_{\mathcal{U}_j - \mathcal{U}_i = k} \mathbf{P}_{r0}(i, j) = \delta_k, \tilde{\mathbf{Q}}_r = [\mathcal{R}_1(\mathbf{Q}_1) \dots \mathcal{R}_1(\mathbf{Q}_{N_f})], \quad (17)$$

where $\mathcal{Q} = [\mathbf{Q}_1 | \dots | \mathbf{Q}_{N_f}] \in \mathbb{C}^{N_m \times N_l \times N_f}$ is the dual variable, $\mathbf{P}_{r0} \in \mathbb{C}^{N_u \times N_u}$, $\tilde{\mathbf{Q}}_r = [\tilde{\mathbf{Q}}_r^1 \dots \tilde{\mathbf{Q}}_r^{N_f}] \in \mathbb{C}^{N_u \times N_l N_f}$ ($\tilde{\mathbf{Q}}_r^f = \mathcal{R}_1(\mathbf{Q}_f) \in \mathbb{C}^{N_u \times N_l}$), and $\mathcal{R}_1(\mathbf{Q}_f) : N_m \times N_l \rightarrow N_u \times N_l$ is a mapping that pads zeros to the extra entries defined as

$$\mathcal{R}_1(\mathbf{Q}_f)(r, l) = \begin{cases} \mathbf{Q}_f(m, l) & \text{for } (\mathcal{U}_r, l) = (f \cdot (m-1), l) \\ 0 & \text{otherwise.} \end{cases} \quad (18)$$

Fig. 2 demonstrates the $\mathcal{R}_1(\cdot)$ mapping. We note that any rows of $\mathbf{Q}_f$ which would have remained all-zero under the operator $\mathcal{R}(\cdot)$ (corresponding to unused space-frequency products) are simply omitted in $\mathcal{R}_1(\cdot)$.

Comparing (18) with (14), these two mappings pad zeros for the same input $\mathbf{Q}_f$ to obtain the output matrix with a different dimension. The $\mathcal{R}$ mapping defined in (14) maps a matrix with N_M rows into one with N rows, while $\mathcal{R}_1$ defined in (18) omits the unused products of temporal frequency and sensor position, mapping a matrix with N_m rows into one with only N_u rows.

This not only gives a lower-dimensional formulation (the size of $\mathbf{P}_{r0}$ decreases from $N \times N$ to $N_u \times N_u$), but it naturally accommodates the NUA and NUF settings. Still, (17) can be applied to the ULA and uniform frequency case, where N_u will often be somewhat smaller than N .

C. Fast Primal SDP for the NUA and NUF Case

In this section, we derive the dual problem of (17), yielding a fast primal SDP that is regularization-free and naturally accommodates the NUA and NUF settings.

Proposition IV.1: The dual problem of (17) is given by

$$\min_{\mathbf{W}, \mathbf{u}, \tilde{\mathbf{Y}}} [\text{Tr}(\mathbf{T}(\mathbf{u})) + \text{Tr}(\mathbf{W})] \\ \text{s.t.} \quad \begin{bmatrix} \mathbf{T}(\mathbf{u}) & \tilde{\mathbf{Y}} \\ \tilde{\mathbf{Y}}^H & \mathbf{W} \end{bmatrix} \succeq 0, \mathbf{Y}_f = \mathcal{R}_1^*(\tilde{\mathbf{Y}}_f), f \in \mathcal{F}, \quad (19)$$

where $\tilde{\mathbf{Y}} \in \mathbb{C}^{N_u \times N_l N_f}$, $\mathbf{W} \in \mathbb{C}^{N_l N_f \times N_l N_f}$, $\mathbf{Y}_f \in \mathbb{C}^{N_m \times N_l}$ is the slice of the received signal tensor $\mathcal{Y}$ corresponding to frequency f , and $\tilde{\mathbf{Y}}_f \in \mathbb{C}^{N_u \times N_l}$ comes from taking the N_l columns of $\tilde{\mathbf{Y}}$ corresponding to frequency f . $\mathcal{R}_1^*(\cdot) : N_u \times N_l \rightarrow N_m \times N_l$ is the adjoint mapping of $\mathcal{R}_1$.

Proof: Consider the Lagrangian given by

$$\begin{aligned} \mathcal{L}(\mathcal{Q}, \mathbf{P}_{r0}, \mathbf{U}_r, \Lambda_1, \Lambda_2, \Lambda_3, \Lambda_Q, \mathbf{v}) \\ = \langle \mathcal{Q}, \mathcal{Y} \rangle_{\mathbb{R}} - \left\langle \begin{bmatrix} \Lambda_1 & \Lambda_2 \\ \Lambda_2^H & \Lambda_3 \end{bmatrix}, \begin{bmatrix} \mathbf{P}_{r0} & \mathbf{U}_r \\ \mathbf{U}_r^H & \mathbf{I}_{N_l N_f} \end{bmatrix} \right\rangle_{\mathbb{R}} \\ - \sum_{k=0}^{N-1} v_k \left(\delta_k - \sum_{\mathcal{U}_j - \mathcal{U}_i = k} \mathbf{P}_{r0}(i, j) \right) \\ - \sum_{f \in \mathcal{F}} \langle \Lambda_Q^f, \mathbf{U}_r^f - \mathcal{R}_1(\mathbf{Q}^f) \rangle_{\mathbb{R}} \\ = \sum_{f \in \mathcal{F}} [\langle \mathbf{Q}_f, \mathbf{Y}_f \rangle_{\mathbb{R}} + \langle \Lambda_Q^f, \mathcal{R}_1(\mathbf{Q}_f) \rangle_{\mathbb{R}}] - [\langle \mathbf{P}_{r0}, \Lambda_1 \rangle_{\mathbb{R}} \\ + 2 \langle \Lambda_2, \mathbf{U}_r \rangle_{\mathbb{R}} + \text{Tr}(\Lambda_3)] - \mathbf{v}_0 + \langle \mathbf{P}_{r0}, \mathbf{T}(\mathbf{v}) \rangle_{\mathbb{R}} \\ - \sum_{f \in \mathcal{F}} \langle \Lambda_Q^f, \mathbf{U}_r^f \rangle_{\mathbb{R}}. \end{aligned} \quad (20)$$

Note that we use the following fact during the derivation: $\sum_{k=0}^{N-1} v_k \sum_{\mathcal{U}_j - \mathcal{U}_i = k} \mathbf{P}_{r0}(i, j) = \langle \mathbf{P}_{r0}, \mathbf{T}(\mathbf{v}) \rangle_{\mathbb{R}}$, where $\mathbf{T} : N \times 1 \rightarrow N_u \times N_u$ is explicitly defined as (note $*$ denotes complex conjugate)

$$\mathbf{T}(\mathbf{v})(i, j) := \begin{cases} v_{\mathcal{U}_j - \mathcal{U}_i} & \mathcal{U}_j - \mathcal{U}_i \geq 0 \\ v_{\mathcal{U}_i - \mathcal{U}_j}^* & \mathcal{U}_j - \mathcal{U}_i < 0. \end{cases} \quad (21)$$

We provide an example of this mapping in Sec. IV-E.

The dual matrix $\begin{bmatrix} \Lambda_1 & \Lambda_2 \\ \Lambda_2^H & \Lambda_3 \end{bmatrix}$ associated with the inequality constraint $\begin{bmatrix} \mathbf{P}_{r0} & \mathbf{U}_r \\ \mathbf{U}_r^H & \mathbf{I}_{N_f} \end{bmatrix} \succeq 0$ needs to be a PSD matrix to ensure that the inner product between these two matrices is non-negative so that the optimal value for the dual problem gives a lower bound for the primal problem.

The dual function is

$$\begin{aligned} g(\Lambda_1, \Lambda_2, \Lambda_3, \Lambda_Q, \mathbf{v}) \\ = \inf_{\mathcal{Q}, \mathbf{P}_{r0}, \mathbf{U}_r} \mathcal{L}(\mathcal{Q}, \mathbf{P}_{r0}, \mathbf{U}_r, \Lambda_1, \Lambda_2, \Lambda_3, \Lambda_Q, \mathbf{v}) \\ \text{s.t. } \begin{bmatrix} \Lambda_1 & \Lambda_2 \\ \Lambda_2^H & \Lambda_3 \end{bmatrix} \succeq 0. \end{aligned} \quad (22)$$

The infimum of $\mathcal{L}$ in (20) over $\mathcal{Q}$ is thereby $\inf_{\mathcal{Q}} J(\mathcal{Q}) := \sum_{f \in \mathcal{F}} [\langle \mathbf{Q}_f, \mathbf{Y}_f \rangle_{\mathbb{R}} + \langle \Lambda_Q^f, \mathcal{R}_1^*(\mathbf{Q}_f) \rangle_{\mathbb{R}}] = \sum_{f \in \mathcal{F}} [\langle \mathbf{Q}_f, \mathbf{Y}_f \rangle_{\mathbb{R}} + \langle \mathbf{Q}_f, \mathcal{R}_1^*(\Lambda_Q^f) \rangle_{\mathbb{R}}] = \sum_{f \in \mathcal{F}} \langle \mathbf{Q}_f, \mathbf{Y}_f + \mathcal{R}_1^*(\Lambda_Q^f) \rangle_{\mathbb{R}}$. The infimum of $J(\mathcal{Q})$ is bounded only if $\mathbf{Y}_f = -\mathcal{R}_1^*(\Lambda_Q^f)$ for any $f \in \mathcal{F}$. Similarly, the infimum of $\mathcal{L}$ over $\mathbf{P}_{r0}$ is bounded only if $\mathbf{T}(\mathbf{v}) = \Lambda_1 \succeq 0$. The infimum of $\mathcal{L}$ over $\mathbf{U}_r$ is bounded only if $\Lambda_Q^f = -2\Lambda_2^f$. Considering $2\Lambda_2^f = \tilde{\mathbf{Y}}_f$, then we must have $\mathbf{Y}_f = -\mathcal{R}_1^*(\Lambda_Q^f) = \mathcal{R}_1^*(2\Lambda_2^f) = \mathcal{R}_1^*(\tilde{\mathbf{Y}}_f)$.

Considering $\Lambda_3 = \frac{1}{2}\mathbf{W}$ and $\mathbf{v} = \frac{1}{2}\mathbf{u}$, the dual function becomes $-\frac{1}{2}\text{Tr}(\mathbf{W}) - \frac{1}{2}\text{Tr}(\mathbf{T}(\mathbf{u}))$.

Therefore, the fast program in the primal domain is given by (19). $\square$

1) *SMV Setup*: The fast program (19) can not only improve the execution time in the uniform cases, but it naturally accommodates the NUA and NUF cases as well. (19) can also be adapted to the SMV setup (i.e. $N_l = 1$). In that case, the received signal $\mathbf{Y}$ will reduce to an $N_m \times N_f$ matrix and (19) will reduce to

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{u}, \tilde{\mathbf{Y}}} [\text{Tr}(\mathbf{T}(\mathbf{u})) + \text{Tr}(\mathbf{W})] \\ \text{s.t. } \begin{bmatrix} \mathbf{T}(\mathbf{u}) & \tilde{\mathbf{Y}} \\ \tilde{\mathbf{Y}}^H & \mathbf{W} \end{bmatrix} \succeq 0, \mathbf{Y}_f = \mathcal{R}_1^*(\tilde{\mathbf{Y}}_f), f \in \mathcal{F}, \end{aligned} \quad (23)$$

where $\tilde{\mathbf{Y}} \in \mathbb{C}^{N_u \times N_f}$, $\mathbf{W} \in \mathbb{C}^{N_f \times N_f}$, $\mathbf{Y}_f \in \mathbb{C}^{N_m \times 1}$ is the column of the received signal $\mathbf{Y}$ corresponding to frequency f , and $\tilde{\mathbf{Y}}_f \in \mathbb{C}^{N_u \times 1}$ comes from taking the column of $\tilde{\mathbf{Y}}$ corresponding to frequency f .

2) *Comparison to Full-Dimension Primal SDP*: Recall that (13) is the dual SDP for the ULA and uniform frequency setting. A significant difference between (13) and (17) lies in the dimensions of the matrices in the PSD constraint and the equality constraint. Following the same procedure in this section, the dual SDP of (13) can be obtained, yielding the following full-dimension primal SDP for the ULA and uniform-frequency case:

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{u}, \tilde{\mathbf{Y}}_N} [\text{Tr}(\text{Toep}(\mathbf{u})) + \text{Tr}(\mathbf{W})] \\ \text{s.t. } \begin{bmatrix} \text{Toep}(\mathbf{u}) & \tilde{\mathbf{Y}}_N \\ \tilde{\mathbf{Y}}_N^H & \mathbf{W} \end{bmatrix} \succeq 0, \mathbf{Y}_f = \mathcal{R}^*(\tilde{\mathbf{Y}}_{Nf}), f = 1, \dots, N_F, \end{aligned} \quad (24)$$

where $\text{Toep}(\cdot) : N \times 1 \rightarrow N \times N$ is the Toeplitz operator that maps a vector to a self-adjoint Toeplitz matrix. $\tilde{\mathbf{Y}}_N \in \mathbb{C}^{N \times N_l N_F}$, $\mathcal{R}^*(\cdot) : N \times N_l \rightarrow N_M \times N_l$ is the adjoint mapping of $\mathcal{R}(\cdot)$, and $\tilde{\mathbf{Y}}_{Nf} \in \mathbb{C}^{N \times N_l}$ is taking N_l columns from $\tilde{\mathbf{Y}}_N$ (from the $(f-1) \cdot N_l + 1$ -th to the $f \cdot N_l$ -th column).

Compared to (19), a main difference is that $\mathbf{T}(\mathbf{u}) \in \mathbb{C}^{N_u \times N_u}$ in (19) is changed to $\text{Toep}(\mathbf{u}) \in \mathbb{C}^{N \times N}$.

D. Existence of Irregular Vandermonde Decomposition (IVD)

The full-dimension primal SDP in (24) has an interesting connection to the SDPs from the ANM literature which involve trace minimization of a (regular) Toeplitz matrix [23], [25]. In ANM problems that involve trace minimization of a regular Toeplitz matrix, one typically computes the Vandermonde decomposition of the resulting Toeplitz matrix to extract the frequencies/DOAs. Indeed, as we discuss further in Section VI, trace minimization serves as a convex relaxation of rank minimization, and a formal connection can be established between rank minimization and finding the sparsest decomposition in the atomic set $\mathcal{A}$.

In contrast, the fast primal SDPs (19) and (23) derived in the previous section involve trace minimization not of a Toeplitz matrix but rather a matrix of the form $\mathbf{T}(\mathbf{u})$. (See Sec. IV-E for an illustration of the structure of $\mathbf{T}(\mathbf{u})$.) However, as we establish in Theorem IV.2 below, there is an important connection between $\mathbf{T}(\mathbf{u})$ and Toeplitz matrices: $\mathbf{T}(\mathbf{u})$ is guaranteed to be an irregular Toeplitz matrix, and therefore is guaranteed to have an IVD. This inspires our proposed method for extracting DOA information from $\mathbf{T}(\mathbf{u})$, which we outline in Section IV-F.

Theorem IV.2: For any $\mathbf{u}$ such that $\text{Toep}(\mathbf{u})$ is PSD, $\mathbf{T}(\mathbf{u}) \in \mathbb{C}^{N_u \times N_u}$ is an (N_u, K) -irregular Toeplitz matrix, where $K = \text{rank}(\text{Toep}(\mathbf{u}))$. Specifically, $\mathbf{T}(\mathbf{u})$ has an IVD of the form (9), where $\gamma = [\mathcal{U}_1, \dots, \mathcal{U}_{N_u}]^T$.

Proof: First, let $P_{\mathcal{U}} : \mathbb{C}^N \rightarrow \mathbb{C}^{N_u}$ denote a linear restriction operator that selects only the entries in a vector corresponding to the positions indexed by $\mathcal{U}$.

Now, consider $\text{Toep}(\mathbf{u}) \in \mathbb{C}^{N \times N}$, and observe that $\mathbf{T}(\mathbf{u}) \in \mathbb{C}^{N_u \times N_u}$ can be obtained by a mapping from $\text{Toep}(\mathbf{u})$ as follows: $\mathbf{T}(\mathbf{u}) := P_{\mathcal{U}} \text{Toep}(\mathbf{u}) P_{\mathcal{U}}^H$. Since $\text{Toep}(\mathbf{u})$ is PSD, it is guaranteed to have a Vandermonde decomposition of the form [48, Theorem 11.5]:

$$\text{Toep}(\mathbf{u}) = \mathbf{V}(\mathbf{z}) \mathbf{D} \mathbf{V}(\mathbf{z})^H \quad (25)$$

where $\mathbf{V}(\mathbf{z}) \in \mathbb{C}^{N \times K}$ is a Vandermonde matrix parameterized by $\mathbf{z}$ with $|\mathbf{z}| = 1$, and $\mathbf{D} \in \mathbb{R}^{K \times K}$ is a diagonal matrix with positive diagonals. Hence,

$$\begin{aligned} \mathbf{T}(\mathbf{u}) &= P_{\mathcal{U}} \text{Toep}(\mathbf{u}) P_{\mathcal{U}}^H = P_{\mathcal{U}} \mathbf{V}(\mathbf{z}) \mathbf{D} \mathbf{V}(\mathbf{z})^H P_{\mathcal{U}}^H \\ &= (P_{\mathcal{U}} \mathbf{V}(\mathbf{z})) \mathbf{D} (\mathbf{V}(\mathbf{z})^H P_{\mathcal{U}}^H) \\ &= \mathbf{W}(\gamma, \mathbf{z}) \mathbf{D} \mathbf{W}(\gamma, \mathbf{z})^H, \end{aligned} \quad (26)$$

where $\mathbf{W}(\gamma, \mathbf{z}) := P_{\mathcal{U}} \mathbf{V}(\mathbf{z})$ will be an irregular Vandermonde matrix of the form (8) with $\gamma = [\mathcal{U}_1 \dots \mathcal{U}_{N_u}]^T$. Therefore, $\mathbf{T}(\mathbf{u})$ is an (N_u, K) -irregular Toeplitz matrix. $\square$

E. An Example for $\mathbf{T}(\mathbf{v})$

We demonstrate the structure of $\mathbf{T}(\mathbf{v})$ in the following example. Consider $\mathcal{M} = \{0, 1, 3, 4\}$ and $\mathcal{F} = \{1, 3, 4\}$. Therefore, $N_m = |\mathcal{M}| = 4$, $N_f = |\mathcal{F}| = 3$, $\mathcal{U} = \{0, 1, 3, 4, 9, 12, 16\}$, $N_M = 5$, $N_F = 4$, $N = (N_M - 1)N_F + 1 = 17$, and $N_u =$

$|\mathcal{U}| = 7$. For $\mathbf{v} = [v_0 \dots v_{16}]^T$, $\mathbf{T}(\mathbf{v}) \in \mathbb{C}^{N_u \times N_u}$ can be expressed as

$$\mathbf{T}(\mathbf{v}) = \begin{bmatrix} v_0 & v_1 & v_3 & v_4 & v_9 & v_{12} & v_{16} \\ v_1^* & v_0 & v_2 & v_3 & v_8 & v_{11} & v_{15} \\ v_3^* & v_2^* & v_0 & v_1 & v_6 & v_9 & v_{13} \\ v_4^* & v_3^* & v_1^* & v_0 & v_5 & v_8 & v_{12} \\ v_9^* & v_8^* & v_6^* & v_5^* & v_0 & v_3 & v_7 \\ v_{12}^* & v_{11}^* & v_9^* & v_8^* & v_3^* & v_0 & v_4 \\ v_{16}^* & v_{15}^* & v_{13}^* & v_{12}^* & v_7^* & v_4^* & v_0 \end{bmatrix}.$$

Note v_{10} and v_{14} do not appear in $\mathbf{T}(\mathbf{v})$. In Theorem IV.2, we show that for any $\mathbf{v}$ such that $\text{Toep}(\mathbf{v})$ is PSD, $\mathbf{T}(\mathbf{v})$ is guaranteed to be an irregular Toeplitz matrix. In this case, $\mathbf{T}(\mathbf{v})$ is guaranteed to have an IVD: $\mathbf{T}(\mathbf{v}) = \mathbf{W}(\gamma, \mathbf{z})\mathbf{D}\mathbf{W}(\gamma, \mathbf{z})^H$, where $\mathbf{D} = \text{diag}(d_1, \dots, d_K)$ with $K = \text{rank}(\text{Toep}(\mathbf{v}))$, and where $\gamma = [0 \ 1 \ 3 \ 4 \ 9 \ 12 \ 16]^T$.

F. DOA Extraction

After solving the fast primal SDP (19) by an off-the-shelf SDP solver (e.g., CVX [42]) and obtaining $\mathbf{u}$, we propose to extract the DOAs by exploiting the IVD of the irregular Toeplitz $\mathbf{T}(\mathbf{u}) = \mathbf{W}(\gamma, \mathbf{z})\mathbf{D}\mathbf{W}(\gamma, \mathbf{z})^H$. Although this factorization is not computed explicitly as part of solving the SDP, its existence provides a means to estimate the entries of $\mathbf{z}$, each corresponding to a point on the unit circle whose complex angle encodes a DOA.

Let $\mathbf{T}(\mathbf{u})$ denote an (N_u, K) -irregular Toeplitz matrix that has an IVD of the form $\mathbf{T}(\mathbf{u}) = \mathbf{W}(\gamma, \mathbf{z})\mathbf{D}\mathbf{W}(\gamma, \mathbf{z})^H$, where $\gamma = [\mathcal{U}_1 \dots \mathcal{U}_{N_u}]^T$. Consider the eigen-decomposition of $\mathbf{T}(\mathbf{u})$:

$$\mathbf{T}(\mathbf{u}) = \mathbf{U}_S \mathbf{\Lambda}_S \mathbf{U}_S^H + \mathbf{U}_N \mathbf{\Lambda}_N \mathbf{U}_N^H, \quad (27)$$

where $\mathbf{\Lambda}_S \in \mathbb{C}^{K \times K}$ is a diagonal matrix containing the K largest eigenvalues of $\mathbf{T}(\mathbf{u})$, $\mathbf{U}_S \in \mathbb{C}^{N_u \times K}$ contains the corresponding eigenvectors, and $\mathbf{\Lambda}_N \in \mathbb{C}^{(N_u-K) \times (N_u-K)}$ and $\mathbf{U}_N \in \mathbb{C}^{N_u \times (N_u-K)}$ contain the remaining (zero) eigenvalues and corresponding eigenvectors. $\mathbf{U}_S$ and $\mathbf{U}_N$ are known as the *signal* and *noise subspaces*, respectively.

For $z \in \mathbb{C}$, define the irregular null spectrum $\tilde{D}(z)$ of $\mathbf{T}(\mathbf{u})$ as [27, eq. (29)]

$$\tilde{D}(z) = \mathbf{w}(\gamma, z)^H \mathbf{U}_N \mathbf{U}_N^H \mathbf{w}(\gamma, z) = \mathbf{w}(\gamma, z)^H \mathbf{G} \mathbf{w}(\gamma, z), \quad (28)$$

where $\mathbf{G} = \mathbf{U}_N \mathbf{U}_N^H$. The behavior of the irregular null spectrum is plotted in Fig. 3.

Since $\mathbf{G} \perp \mathbf{W}(\gamma, \mathbf{z})$ and $|\mathbf{z}| = 1$, the DOAs encoded in $\mathbf{z}$ are associated to the K roots of $\tilde{D}(z)$ on the unit circle. [27] suggests that the local minima of $\tilde{D}(z)$ evaluated on the unit circle give DOA estimates with similar accuracy as those given by the actual roots. Therefore, $\mathbf{z}$ is estimated as [27, eq. (43)]

$$\hat{\mathbf{z}} = \arg \min_{|z|=1}^k \tilde{D}(z), \quad k = 1, \dots, K \quad (29)$$

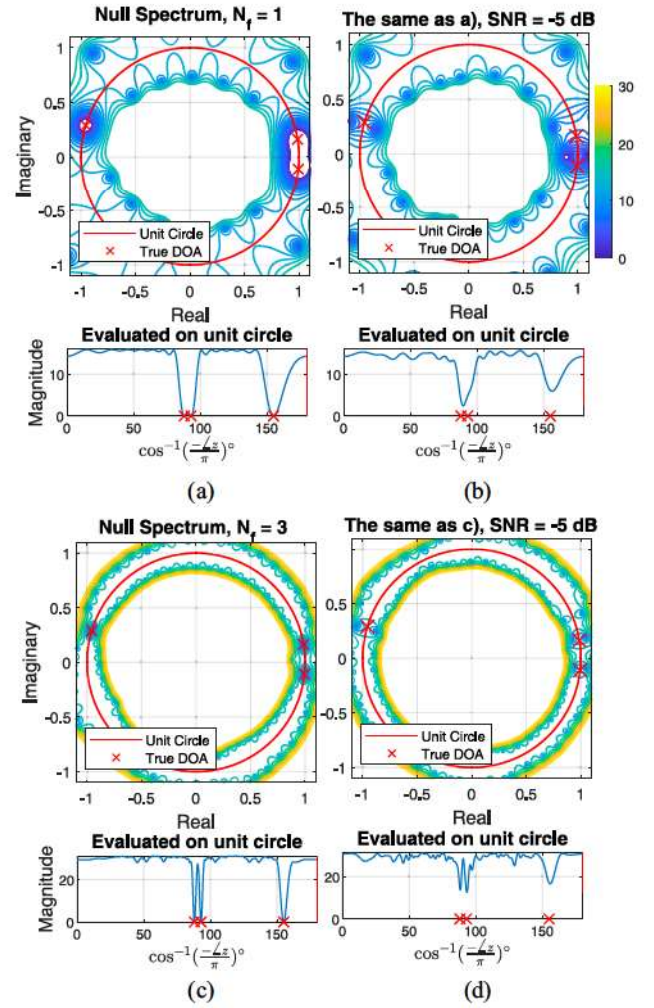


Fig. 3. Null spectrum contours (dB) using $N_M = 16$, $K = 3$, $N_L = 5$. For a) - b) $N_F = 1$, and for c) - d) $N_F = 3$. The frequency set is $\{100, \dots, N_F \cdot 100\}$ Hz and ULA is applied. DOAs are $[88, 93, 155]^\circ$, marked by red x's, the red line marks the complex unit circle. a) and c): null spectrum from noise-free measurement. b) and d): null spectrum from SNR = -5 dB measurement.

where $\arg \min_z^k$ denotes the argument, z , which produces the k th smallest local minima. The DOAs $\hat{\theta}$, $\hat{w}$, and $\hat{z}$ are estimated by

$$\hat{\theta} = \cos^{-1} \left(-\frac{\angle \hat{z}}{\pi} \right), \quad \hat{w} = \frac{-\angle \hat{z}}{2\pi}, \quad \hat{z} = e^{-j\pi \cos(\hat{\theta})}. \quad (30)$$

In summary, we first solve the SDP (19) via an off-the-shelf SDP solver (e.g., CVX [42]). After $\mathbf{u}$ is obtained, the DOAs can be retrieved by computing the irregular null spectrum $\tilde{D}(z)$ of $\mathbf{T}(\mathbf{u})$ and following the steps mentioned in this section. The implementation details of the proposed method are summarized in Algorithm 1.

V. MORE SOURCES THAN SENSORS FOR ULA

Many prior works have demonstrated the possibility of resolving more sources than the number of array sensors based on special array geometries such as MRA [34], [50], co-prime arrays [36], and nested array [37]. However, for single-frequency

Algorithm 1 Regularization-free DOA estimation**Input:** $\mathcal{Y} \in \mathbb{C}^{N_m \times N_l \times N_f}$, K **Initialization:**Solve (19) by CVX and obtain $\mathbf{u}$ Obtain $\mathbf{T}(\mathbf{u})$ based on (21) $[\mathbf{U}, \Lambda] = \text{eig}(\mathbf{T}(\mathbf{u}))$ $\mathbf{U}_N = \mathbf{U}(:, K+1 : N_u)$ $\mathbf{G} = \mathbf{U}_N \mathbf{U}_N^H$ Obtain $\tilde{D}(z)$ based on (28) $\hat{\mathbf{z}} = \text{find}(\arg \min(\tilde{D}(z)), |z| = 1)$ $\hat{\theta} \leftarrow 180 - \text{acosd}(\text{angle}(\hat{\mathbf{z}}/\pi))$ **Output:** $\hat{\theta}$

ULA, the maximum number of resolvable sources is $N_M - 1$ [48, Sec 11.2.3]. In this section, we will demonstrate the possibility of resolving more sources than sensors *under the ULA setup* if multiple frequencies are available. We primarily solve (24) and follow the procedures in Algorithm 1 to retrieve the DOAs. In our multi-frequency ANM configuration, it can resolve up to $N - 1 = (N_M - 1)N_F$ sources as $\text{Toep}(\mathbf{u}) \in \mathbb{C}^{N \times N}$ and $\mathbf{U}_N$ exists only if $K \leq N - 1$. The reason for using (24) instead of (19) is that (19) can resolve up to $N_u - 1$ sources and (24) has the potential to resolve more sources than (19) because $\text{Toep}(\mathbf{u})$ in (24) has a higher dimension than $\mathbf{T}(\mathbf{u})$ in (19). This idea was also demonstrated in [51] for coprime frequencies, though the method in [51] used grid-based DOA estimation.

The key observation for the multi-frequency model is that these frequencies increase the diversity of the harmonics. These extra harmonics serve as “virtual” sensors in the array, and they bring about an enhanced degree of freedom. For example, consider a ULA with $N_M = 4$ sensors and $N_F = 5$ uniform frequencies. Therefore, it can resolve up to $(N_M - 1)N_F = 15$ sources. The SDP problem (19) can be interpreted as a structured covariance matrix estimation problem ($\mathbf{T}(\mathbf{u})$ can be interpreted as the covariance matrix). We notice this covariance matrix is in a higher dimension, which corresponds to our intuition that there are more sensors in our “virtual” array.

As an example, suppose we have $N_M = 4$ sensors, $N_F = 5$ frequencies ($\{100, \dots, 500\}$ Hz), $N_l = 1$ noise-free snapshot, and $K = 10, 11, 12, 13, 14, 15$ sources with uniform and deterministic across frequencies. For $K = 10, 12$, and 15 , the DOAs are generated as the uniform distribution in the cosine domain (i.e., the DOAs are $\lfloor \cos^{-1}(-1 + 2([1 : K] - 0.5)/K) \rfloor$). For $K = 11$, we pick up the last 11 sources in the $K = 12$ case. For $K = 13$, we pick up the middle 13 sources in the $K = 15$ case, and for $K = 14$, we pick up the middle 14 sources. We plot the estimated DOAs for ANM. From Fig. 4, we can see our ANM can resolve up to $(N_M - 1)N_F = 15$ sources.

VI. RANK MINIMIZATION AND ATOMIC ℓ_0 NORM MINIMIZATION

In this section, we highlight the connection between rank minimization and atomic ℓ_0 norm minimization. More specifically, atomic ℓ_0 norm minimization can be interpreted as a covariance matrix estimation approach where low-rankness and

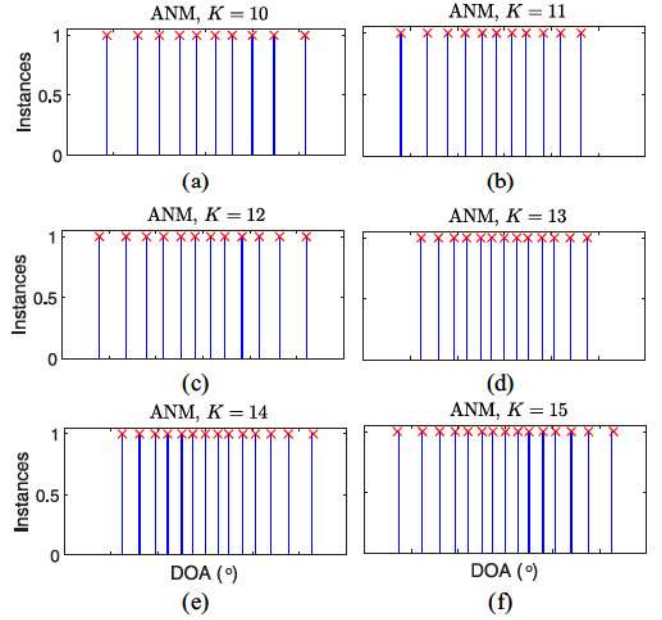


Fig. 4. Estimated and true DOAs for ANM (“x” indicates the true DOAs and the blue vertical line indicates the estimated DOAs). $N_M = 4$, $N_F = 5$, $N_l = 1$, and $K = 10, 11, 12, 13, 14, 15$. The RMSEs of ANM under $K = 10, 11, 12, 13, 14, 15$ are 0.005° , 0.16° , 0.20° , 0.04° , 0.27° , and 0.27° .

Toeplitz structure are explicitly enforced [23], [25], [52]. However, atomic ℓ_0 norm minimization is non-convex and may not be computationally feasible. By considering ANM, the convex relaxation of the atomic ℓ_0 norm, we obtain (19) and (23) as trace minimization problems that are computationally feasible and in which the low-rankness and Toeplitz structure are implicitly enforced. In this way, we can understand the benefits of ANM compared to conventional covariance matrix estimation using the sample covariance matrix. Before describing the equivalence, we review the definitions of the covariance matrix and the sample covariance matrix.

We assume $N_l = 1$, noise-free measurement, uniform frequency and ULA setup, and full dimensional SDP in this section, and our discussion serves as a means to interpret (23).

A. Covariance Matrix Estimation

Suppose $\tilde{\mathbf{Y}} \in \mathbb{C}^{N \times N_F}$ is noise-free and defined as

$$\tilde{\mathbf{Y}} = \sum_{k=1}^K \mathbf{z}_k \mathbf{x}_k^H = [\mathbf{z}_1 \dots \mathbf{z}_K][\mathbf{x}_1 \dots \mathbf{x}_K]^H = \mathbf{Z}\mathbf{X} \quad (31)$$

where $\mathbf{z}_k := [z_k^0 \dots z_k^{N-1}]^T \in \mathbb{C}^N$, $\mathbf{x}_k = [x_k^{(1)} \dots x_k^{(N_F)}]^T \in \mathbb{C}^{N_F}$, $\mathbf{Z} := [\mathbf{z}_1 \dots \mathbf{z}_K] \in \mathbb{C}^{N \times K}$, and $\mathbf{X} := [\mathbf{x}_1 \dots \mathbf{x}_K]^H \in \mathbb{C}^{K \times N_F}$. Note that $\mathbf{Z}$ is a Vandermonde matrix and that $\tilde{\mathbf{Y}}$ satisfies

$$\mathbf{Y} = \mathcal{R}^*(\tilde{\mathbf{Y}}). \quad (32)$$

The sample covariance matrix $\hat{\mathbf{R}}_{\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}}$ and covariance matrix $\mathbf{R}_{\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}}$ are defined as

$$\hat{\mathbf{R}}_{\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}} = \frac{1}{N_f} \tilde{\mathbf{Y}} \tilde{\mathbf{Y}}^H, \quad (33)$$

$$\mathbf{R}_{\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}} = \mathbb{E}[\hat{\mathbf{R}}_{\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}}] = \mathbf{Z} \Lambda_X \mathbf{Z}^H \quad (34)$$

where $\Lambda_X := \frac{1}{N_f} \mathbb{E}[\mathbf{X}\mathbf{X}^H]$ is a diagonal matrix due to uncorrelated sources. Note that $\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}$ is a Toeplitz Hermitian matrix in the noise-free case, $\mathbf{Z}\Lambda_X\mathbf{Z}^H$ is its Vandermonde decomposition, and the DOAs are encoded in the Vandermonde matrix $\mathbf{Z}$. The essence of some classical DOA estimation approaches (e.g. MUSIC, and ESPRIT) lies in Vandermonde decomposition of the estimated covariance matrix.

However, $\tilde{\mathbf{Y}}$ is not fully observed in our problem as only $\mathbf{Y}$, the image of the $\mathcal{R}^*$ mapping is accessible. To further obtain $\tilde{\mathbf{Y}}$, the lifting mapping $\mathcal{R}$ needs to be applied. Note if we apply $\mathcal{R}^*$ first and then $\mathcal{R}$ on a matrix, we may not obtain the same matrix as the white entries in Fig. 1 cannot be recovered after the $\mathcal{R}^*$ mapping. Therefore, the covariance matrix of $\tilde{\mathbf{Y}}$ must be estimated by solving a convex optimization problem.

B. Connection Between Rank Minimization and Atomic ℓ_0 Norm Minimization

As described in Sec. IV-F, after solving the SDP (23) and obtaining $\mathbf{u}$, DOAs are extracted by computing the IVD of $\mathbf{T}(\mathbf{u})$. In light of the discussion in Section VI-A, then, (23) can be interpreted as a covariance matrix estimation problem where $\mathbf{T}(\mathbf{u})$ serves as an estimate for a covariance matrix that contains the DOA information. In this section, we discuss this connection more deeply.

From (34), the true covariance matrix in the noise-free case $\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}$ has three important properties: (1) Toeplitz and Hermitian; (2) PSD; (3) low-rank (its rank is K (number of sources) and is usually much smaller than its size N_u). A commonly used estimate for the covariance matrix is the *sample covariance matrix* $\hat{\mathbf{R}}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}$ defined in (33), which is PSD. However, this estimate does not promote the Toeplitz structure of the covariance matrix. This limitation is overcome by the SDP formulation in (23). The irregular Toeplitz structure is enforced in $\mathbf{T}(\mathbf{u})$. Meanwhile, (23) also promotes low-rank structure, a fact that warrants more discussion.

The atomic ℓ_0 norm of an $N \times N_F$ matrix $\tilde{\mathbf{Y}}$ is defined as

$$\|\tilde{\mathbf{Y}}\|_{A,0} := \inf \left\{ K \mid \tilde{\mathbf{Y}} = \sum_{k=1}^K c_k \mathbf{z}_k \mathbf{x}_k^H, c_k > 0 \right\} \quad (35)$$

where $\mathbf{z}_k := [z_k^0 \dots z_k^{N-1}]^T \in \mathbb{C}^N$ such that $|z_k| = 1$ and $\mathbf{x}_k = [x_k^{(1)} \dots x_k^{(N_F)}]^T \in \mathbb{C}^{N_F}$ such that $\|\mathbf{x}_k\|_2 = 1$.

The following proposition establishes an equivalence between the atomic ℓ_0 norm and rank minimization.

Proposition VI.1 ([48, Theorem 11.13]): For any $N \times N_F$ matrix $\tilde{\mathbf{Y}}$ with an atomic decomposition of the form (35) (which includes any $\tilde{\mathbf{Y}}$ satisfying (31)), $\|\tilde{\mathbf{Y}}\|_{A,0}$ is equal to the optimal value of the following rank minimization problem:

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{u}} \quad & \text{rank}(\text{Toep}(\mathbf{u})) \\ \text{s.t.} \quad & \begin{bmatrix} \text{Toep}(\mathbf{u}) & \tilde{\mathbf{Y}} \\ \tilde{\mathbf{Y}}^H & \mathbf{W} \end{bmatrix} \succeq 0. \end{aligned} \quad (36)$$

Remark: To summarize our intuition, the proposition above indicates that (23), which is essentially a reduced-dimension convex relaxation of (36), will promote both low-rankness and

Toeplitz structure and therefore yields a favorable covariance matrix estimation that reveals the sparse decomposition of the DOAs and is consistent with the observed data.

VII. NUMERICAL RESULTS

We use numerical experiments to examine the performance of the method. In this section, N_F and N_f are used to denote the number of frequencies for the uniform and non-uniform frequency sets, respectively. N_M and N_m are used to denote the number of sensors for the uniform and non-uniform array spacing sets, respectively. For each experiment and trial, K DOAs are generated, and $c_w = 1$ for each source unless otherwise specified. The source amplitude $\mathbf{X}_w$ is complex Gaussian. N_l snapshots are collected. The uniform frequency set is defined as $\{1, \dots, N_F\} \cdot F_1$ (F_1 is the minimum frequency). The array spacing for ULA is $\frac{\lambda_1}{2}$ where λ_1 is the wavelength for the minimum frequency in the frequency set. The noise for each frequency and each snapshot is randomly generated from the complex Gaussian distribution $\mathcal{CN}(0, \sigma^2)$ and then scaled to fit the desired signal-to-noise ratio (SNR) defined as

$$\text{SNR} = 20 \log_{10} \frac{\|\mathcal{X}\|_{\text{HS}}}{\|\mathcal{N}\|_{\text{HS}}}. \quad (37)$$

In the Monte-Carlo experiments, $MC = 100$ trials are executed to compute the root mean square error (RMSE) defined as

$$\text{RMSE} = \sqrt{\frac{1}{MC} \sum_{m=1}^{MC} \left[\min \left(\frac{1}{K} \sum_{k=1}^K (\hat{\theta}_{mk} - \theta_{mk})^2, 10^2 \right) \right]}, \quad (38)$$

where $\hat{\theta}_{mk}$, and θ_{mk} are (sorted) estimated DOAs, and (sorted) ground-truth DOAs for the k th DOA and m th trial. A maximum threshold of 10° is used to penalize the incorrect DOA estimates. We compare the proposed method with the multi-frequency sparse Bayesian learning (SBL) [8]. The Cramér-Rao bound (CRB) [46, Eq. (121)] for the multi-frequency model is computed for reference.

A. Robustness to Aliasing/Collision

We first examine the robustness of aliasing/collision. Suppose $K = 3$ sources impinge in a ULA with $N_M = 16$ sensors. The source amplitudes are complex Gaussian and the DOAs are randomly generated from a uniform distribution with range $[15^\circ, 165^\circ]$ with minimum separation $4/N_M$ in the cosine domain, a choice inspired by [10, Theorem 4.2] (here the separation is in the cosine domain while in [10] it is in the w domain). We consider $N_F = 2$ or 4 under the single-snapshot and uniform frequency case ($N_l = 1$). All frequencies other than the fundamental frequency will have the risk of aliasing/collision. We solve the SDP program (23) by CVX [42] and apply the root-MUSIC (Vandermonde decomposition) to retrieve the DOAs.

From Fig. 5(a) and 5(b), the primal ANM (ANM P) is more robust to the aliasing than SBL. It also overcomes the collision issues for the dual ANM (ANM D) [10]. Moreover, the primal ANM does not need any hyper-parameter tuning and it avoids the bias from the regularization terms.

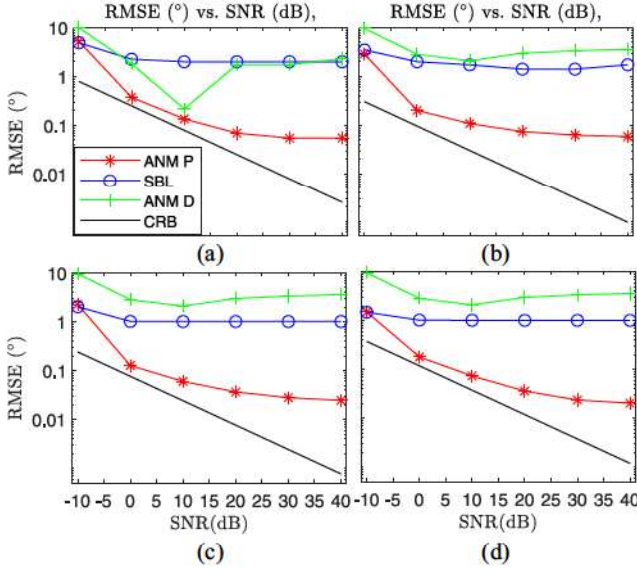


Fig. 5. RMSE (°) versus SNR. $N_M = 16$ ULA with $d = \lambda_{100}/2$. $N_L = 1$, and $K = 3$ sources with randomly generated DOAs from $[15^\circ, 165^\circ]$ with minimum separation $4/N_M$ in the cosine domain. (a): $N_F = 2$ with frequency set $\{100, 200\}$ Hz; (b) $N_F = 4$ with frequency set $\{100, 200, 300, 400\}$ Hz; (c) $N_F = 4$ with frequency set $\{100, 200, 300, 500\}$ Hz; (d) $N_F = 4$ with frequency set $\{200, 300, 400, 500\}$ Hz. The proposed primal SDP program (ANM P) and the dual SDP program [10] (ANM D) as well as SBL and CRB are shown. All source amplitudes are complex Gaussian with unit variance and $c_w = 1$.

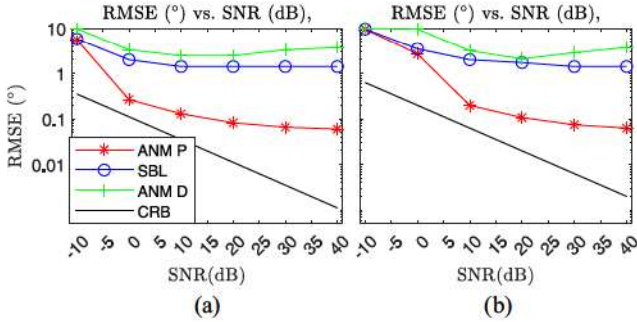


Fig. 6. Same setup as Fig. 5(b), but the c_w of the three sources are (a) 1, $\sqrt{2}$, $\sqrt{3}$ and (b) 1, 2, 4.

B. Non-Uniform c_w

In the previous section, the c_w is uniform across all sources. In this section, we examine the case when c_w of each source is different. From Fig. 6, the proposed method (ANM-P) achieves almost the same performance as the uniform case and could be applied when the c_w of each source is different.

C. Non-Uniform Frequency Cases

We examine the performance under the non-uniform frequency set. In this case, $N_f = 4$, and the frequency set is $\{100, 200, 300, 500\}$ Hz and $\{200, 300, 400, 500\}$ Hz. Other conditions are the same as in Sec. VII-A. Fig. 5(c) and 5(d) demonstrates the effectiveness of the proposed method under the non-uniform frequency case. We see superior performance to the fast dual algorithm proposed in [10].

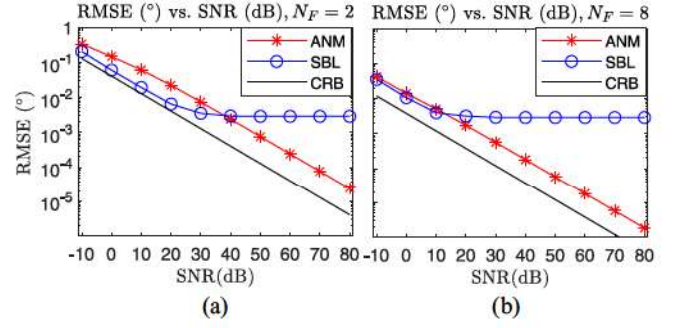


Fig. 7. RMSE (°) versus SNR for MMV setup. $N_M = 16$ ULA with $d = \lambda_{100}/2$. $K = 3$ DOAs at $[88^\circ, 93^\circ, 155^\circ] + \epsilon$ where ϵ is the random offsets from a uniform distribution $[0, 1]$. $N_L = 20$. (a): $N_F = 2$ with frequency set $\{100, 200\}$ Hz; (b) $N_F = 8$ with frequency set $\{100, \dots, 800\}$ Hz.

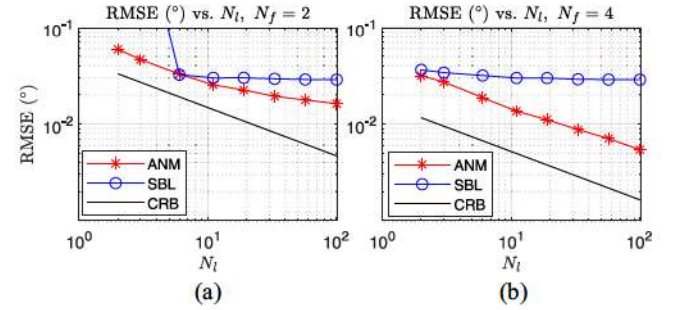


Fig. 8. RMSE (°) versus N_L for MMV setup. $N_M = 16$ ULA with $d = \lambda_{100}/2$. $K = 3$ DOAs at $[88^\circ, 93^\circ, 155^\circ] + \epsilon$ where ϵ is the random offsets from a uniform distribution $[0, 1]$. SNR = 20 dB. (a): $N_F = 2$ with frequency set $\{100, 200\}$ Hz; (b) $N_F = 4$ with frequency set $\{100, 200, 300, 400\}$ Hz.

D. MMV Case

We examine the performance of ANM under the MMV setup. We consider the case $N_L = 20$, and $K = 3$ DOAs at $[88, 93, 155]^\circ + \epsilon$ where ϵ is a three dimensional random vector with uniform distribution from $[0, 1]$. Fig. 7 demonstrates the superior performance of ANM in the high SNR region, and it follows the trend of CRB.

We then examine the performance of ANM with varying numbers of snapshots N_L for SNR = 20 dB, and the other setup as Fig. 7. From Fig. 8, ANM follows the trend of CRB and outperforms SBL. In addition, comparing Fig. 8(a) with Fig. 8(b), ANM performs better with higher N_F , which demonstrates the benefits of multi-frequency processing.

E. The Effect of Multiple Frequencies

We study the performance of the method under varying N_F in Fig. 9. From Fig. 9(a), the estimation error of ANM generally goes down with increasing N_F , the only exception is $N_F = 7$, where it increases by roughly 0.01° . To understand that, the true and the aliasing DOAs are in Fig. 9(b). It can be seen that the DOAs 93° and 155° nearly collide with each other at frequency 700 Hz. Referring to (1) and (7), this can be understood as $w_2 = 1/2 \cdot \cos(93^\circ)$, $w_3 = 1/2 \cdot \cos(155^\circ)$, $|w_2 - w_3| \approx 3/7$, and there is a near collision in frequency $f = 7$ (i.e. $7 \cdot 100 = 700$ Hz). There are other intersection points between the solid

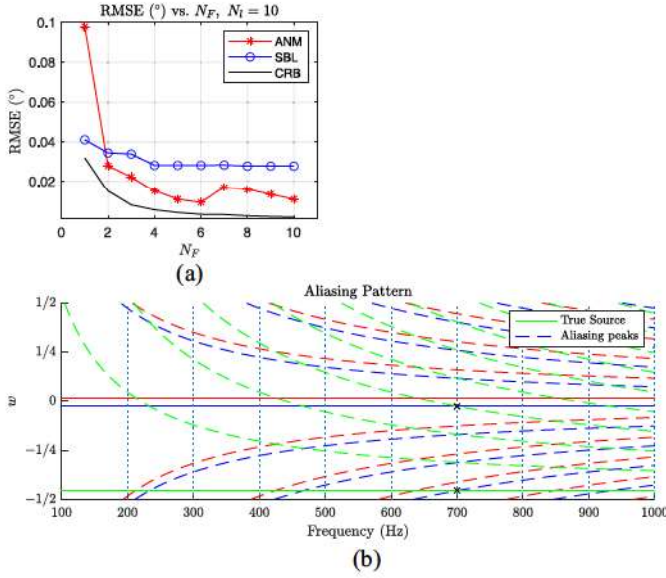


Fig. 9. (a) RMSE (°) versus N_F and (b) aliasing pattern for MMV setup. $N_M = 16$ ULA with $d = \lambda_{100}/2$. $N_l = 10$. The frequency set is $\{100, \dots, N_F \cdot 100\}$ Hz. $K = 3$ DOAs at $[88^\circ, 93^\circ, 155^\circ] + \epsilon$ where ϵ is the random offsets from a uniform distribution $[0, 1]$. SNR = 20 dB. In (b), the true (solid) and the aliasing DOAs (dashed) are shown, with true DOAs $[88^\circ$ (red), 93° (blue), 155° (green)].

and dashed lines but none of them lie in any frequency that belongs to the frequency set. That explains why the error increases when N_F increases from 6 to 7.

F. Co-Prime Array and More Sources than Sensors

We examine an $N_m = 6$ co-prime array, a particular example of the non-uniform array (NUA). A co-prime array involves two ULAs with spacing $M_1 d$ and $M_2 d$. M_1 and M_2 are co-prime integers and their greatest common divisor is 1. The first ULA has M_2 sensors and the second ULA has $2M_1$ sensors. Since the first sensor is shared, there are $N_m = 2M_1 + M_2 - 1$ sensors in the array. In this example, we consider $M_1 = 2$, $M_2 = 3$. $N_f = 3$ and the non-uniform frequency set is $\{100, 300, 400\}$ Hz. $d = \lambda_{100}/2$. The first ULA is $[0, 2d, 4d]$ and the second ULA is $[0, 3d, 6d, 9d]$. The entire co-prime array is $[0, 2d, 3d, 4d, 6d, 9d]$. $N_l = 50$, SNR = 20 dB, and $K = 7$ DOAs with at $\{45^\circ, 60^\circ, 75^\circ, 90^\circ, 105^\circ, 120^\circ, 140^\circ\}$. Note, $K > N_m$ in this case.

From Fig. 10, the proposed method resolves more DOAs than sensors in the NUA case, while SBL fails in this case and has a high RMSE (The maximum RMSE is 10° as the maximum threshold of the RMSE for one trial is 10° based on (38)).

Further, we examine the case when there are more DOAs than sensors under the ULA setup. We have already demonstrated that in Sec. V under a noise-free and uniform amplitude setup. Here, we consider a more practical case when there is noise and the amplitude is random. From Fig. 11, ANM can resolve 6 DOAs when only $N_M = 4$ physical sensors are available under the noisy and non-uniform amplitude case and it achieves lower RMSE performance than SBL.

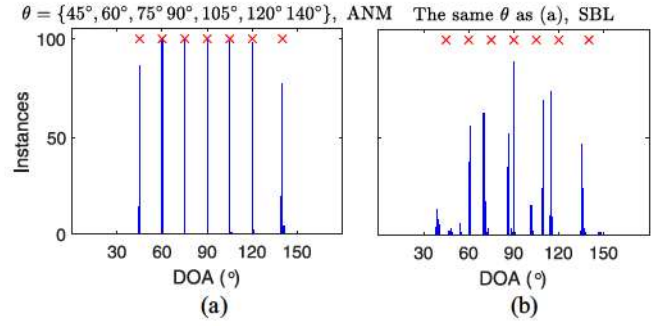


Fig. 10. Histogram for the estimated DOAs for (a) ANM, and (b) SBL under the co-prime array with $N_m = 6$ (sensor locations are $[0, 2, 3, 4, 6, 9]$), $N_f = 3$ ($[100, 300, 400]$ Hz), $N_l = 50$, SNR = 20 dB, and $K = 7$. The RMSE for ANM is 0.2° and for SBL 8.6° .

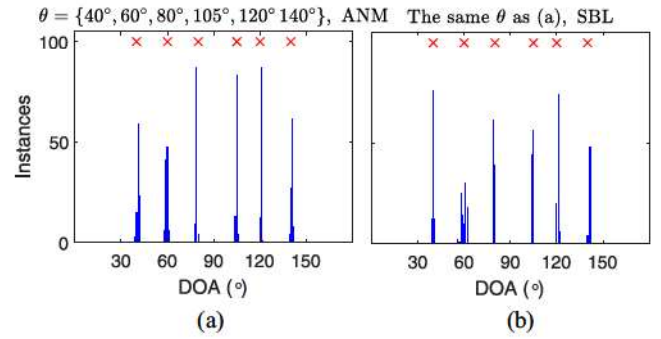


Fig. 11. Histogram for the estimated DOAs for an ULA (a) ANM, and (b) SBL. $N_M = 4$, $N_F = 3$ ($[100, 200, 300]$ Hz), $N_l = 50$, SNR = 20 dB and $K = 6$. The RMSE for ANM is 0.90° , and for SBL 1.10° .

Although we only demonstrated the co-prime array as an important example of NUA, the proposed method can be applied to any NUA satisfying the assumptions in Sec. II-A1.

G. Practical Test

We consider a case with $N_F = 20$ frequencies, $N_l = 10$ snapshots, and $N_M = 20$ sensors. In previous examples, the number of frequencies is small, but in practical cases, there may be many more frequencies in the wideband signal. We compare the performance to the SBL with high resolution 0.01° . From Fig. 12, our method can deal with such a practical case with lower RMSE than the high-resolution SBL.

H. Robustness to Imperfect Estimation of K

In the previous examples, we used perfect knowledge of the number of sources K . In practice, K can be estimated using any of a variety of classical source estimation methods (e.g. Gerschgorin disk method). However, it is possible that this estimate may not be exactly correct. In the following example, we demonstrate robust results when a larger K is assumed (i.e., $K_{\text{est}} \geq K$).

The setup is the same as in Fig. 4 except $K_{\text{est}} = 15$. From Fig. 13, the proposed method can still capture the K true sources for all cases in this example. Therefore, it has robustness to the case when K is overestimated. In practice, if K is

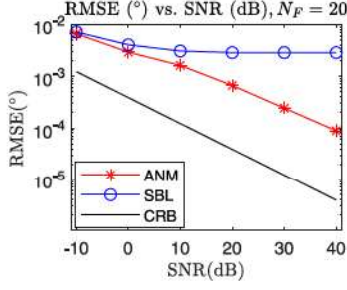


Fig. 12. RMSE (°) versus SNR for MMV setup. $N_M = 20$ ULA with $d = \lambda_{100}/2$. $K = 3$ DOAs at $[88^\circ, 93^\circ, 155^\circ] + \epsilon$ where ϵ is the random offsets from a uniform distribution $[0, 1]$. $N_l = 10$. (a): $N_F = 20$ with frequency set $\{100, \dots, 2000\}$ Hz.

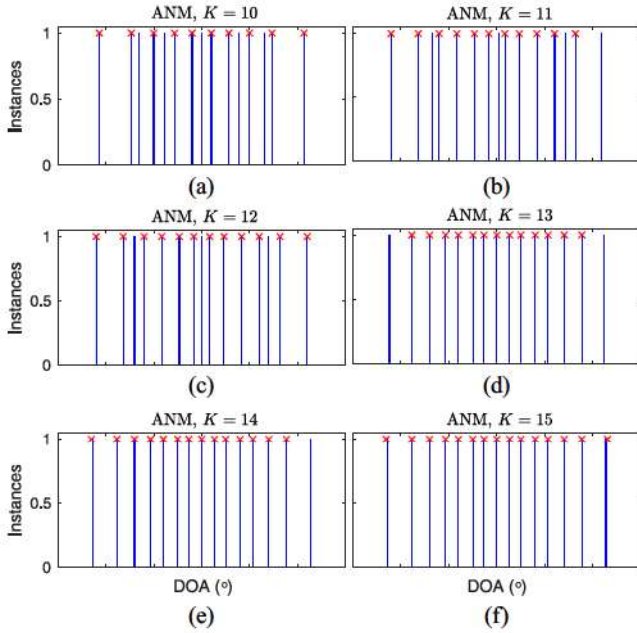


Fig. 13. Estimated and true DOAs for ANM ("x" indicates the true DOAs and the blue vertical line indicates the estimated DOAs). The same setup as Fig. 4 except K is estimated as 15 for all cases.

TABLE II
AVERAGE CPU RUN TIME (s) UNDER THE
SAME SETUP AS FIG. 5(b)

	-10 dB	10 dB	30 dB
Proposed	1.1	1.1	1.0
ANM [10]	3.8	3.8	3.7
SBL	7.0	4.9	4.8

unknown, we can feed a reasonably large K to the proposed method and it may still be possible to resolve all the true sources.

I. Complexity

We compare the average CPU time of the proposed method with the dual ANM [10] and SBL (with a 0.01° grid) over 100 trials with the setup used in Fig. 5(b) with SNR values of -10, 10, and 30 dB. In these results, presented in Table II, the

proposed method is at least 3 times faster than both the dual ANM and SBL.

VIII. CONCLUSION

This paper proposes a gridless DOA estimation method based on *regularization-free* SDP and Vandermonde decomposition. We further extend this framework to MMV, NUA, and non-uniform frequency cases. Under the NUA and non-uniform frequency case, the Toeplitz structure will not hold. However, we demonstrate the possibility of using IVD in these cases, and the existence of IVD is theoretically guaranteed. With the help of multiple frequencies, the method can resolve more sources than the number of physical sensors under the ULA setup. Therefore, multi-frequency processing can reduce the number of physical sensors and increase the maximum resolvable sources. Numerical results demonstrate the proposed framework is robust to noise and aliasing and can achieve a superior performance under the MMV, NUA, and NUF setup.

APPENDIX

A. Proof for Proposition III.1

Proof: The primal atomic norm $\|\mathcal{X}\|_{\mathcal{A}}$ is expressed in terms of the dual atomic norm $\|\mathcal{Q}\|_{\mathcal{A}}^*$ as

$$\|\mathcal{X}\|_{\mathcal{A}} = \sup_{\|\mathcal{Q}\|_{\mathcal{A}}^* \leq 1} \langle \mathcal{Q}, \mathcal{X} \rangle_{\mathbb{R}} = \sup_{\|\mathcal{Q}\|_{\mathcal{A}}^* \leq 1} \langle \mathcal{Q}, \mathcal{Y} \rangle_{\mathbb{R}}, \quad (39)$$

where the last equality is only for the noise-free case. For any dual variable $\mathcal{Q}$, we can define the *dual polynomial matrix* $\Psi(\mathcal{Q}, w) \in \mathbb{C}^{N_F \times N_l}$ as

$$\Psi(\mathcal{Q}, w) := [\mathbf{Q}_1^H \mathbf{a}(1, w) \dots \mathbf{Q}_{N_F}^H \mathbf{a}(N_F, w)]^T. \quad (40)$$

Since each frequency has different array manifold vectors, it is difficult to express $\Psi(\mathcal{Q}, w)$ as a matrix multiplication of $\mathcal{Q}$ and a vector. To construct a homogeneous representation for $\Psi(\mathcal{Q}, w)$, we will leverage $\mathbf{z} := [z^0 \dots z^{N-1}]^T \in \mathbb{C}^N$ ($z = z(w) := e^{-j2\pi w}$), an ensemble of the array manifold, and the matrix $\tilde{\mathbf{Q}}_f \in \mathbb{C}^{N \times N_l}$ defined as follows [10, eq. (14)]

$$\tilde{\mathbf{Q}}_f(i, l) = \begin{cases} \mathbf{Q}_f(m, l) & \text{for } (i, l) = (f \cdot (m-1) + 1, l) \\ 0 & \text{otherwise,} \end{cases} \quad (41)$$

or $\tilde{\mathbf{Q}}_f = \mathcal{R}(\mathbf{Q}_f)$. With the help of $\tilde{\mathbf{Q}}_f$ and $\mathbf{z}$, $\Psi(\mathcal{Q}, w)$ has the representation

$$\Psi(\mathcal{Q}, w) = [\tilde{\mathbf{Q}}_1^H \mathbf{z} \dots \tilde{\mathbf{Q}}_{N_F}^H \mathbf{z}]^T. \quad (42)$$

Now, we consider $\|\mathcal{Q}\|_{\mathcal{A}}^*$, which appears in the constraint in (39). Recalling that $\|\mathbf{X}_w\|_F = 1$, we have a similar derivation to [10, eq. (17)]:

$$\begin{aligned} \|\mathcal{Q}\|_{\mathcal{A}}^* &:= \sup_{\|\mathcal{X}\|_{\mathcal{A}} \leq 1} \langle \mathcal{Q}, \mathcal{X} \rangle_{\mathbb{R}} = \sup_{\|\mathcal{X}\|_{\mathcal{A}} \leq 1} \langle \mathcal{Q}, \mathbf{A}(w) * \mathbf{X}_w^T \rangle_{\mathbb{R}} \\ &= \sup_{\mathbf{X}_w} \text{Tr} \left[\sum_{f=1}^{N_F} \mathbf{Q}_f^H \mathbf{a}(f, w) \mathbf{x}_w^T(f) \right] \\ &= \sup_{\mathbf{X}_w} \text{Tr}[\Psi^H \mathbf{X}_w] = \sup_w \|\Psi(\mathcal{Q}, w)\|_F. \end{aligned} \quad (43)$$

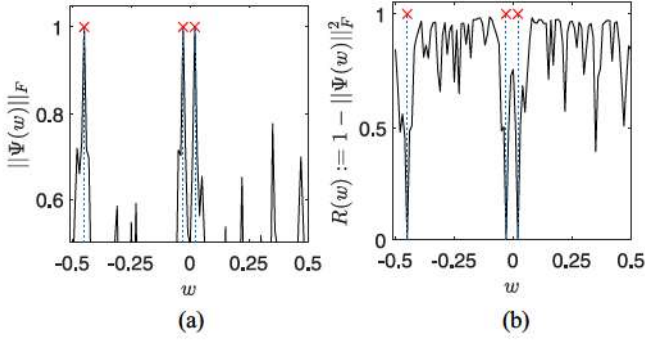


Fig. 14. Dual polynomial visualization. A ULA with $N_M = 16$ sensors and spacing $d = c/2F_1$ ($F_1 = 100$ Hz) is used. $N_F = 5$, $N_l = 5$, $\theta = [87.7076^\circ, 93.4398^\circ, 154.1581^\circ]$, and $w = [0.02, -0.03, -0.45]$. (a) $\|\Psi(\mathbf{Q}, w)\|_F$ versus w ; (b) $R(w)$ versus w .

Using (43), the condition $\|\mathbf{Q}\|_{\mathcal{A}}^* \leq 1$ can be equivalently formulated as an SDP constraint. Construct a similar polynomial as in [10, eq. (23)]:

$$\begin{aligned} R(w) &:= 1 - \|\Psi(\mathbf{Q}, w)\|_F^2 = 1 - \text{Tr}[\Psi^H(\mathbf{Q}, w)\Psi(\mathbf{Q}, w)] \\ &= 1 - \text{Tr}\left(\sum_{f=1}^{N_F} \tilde{\mathbf{Q}}_f^H \mathbf{z} \mathbf{z}^H \tilde{\mathbf{Q}}_f\right) = 1 - \sum_{f=1}^{N_F} \mathbf{z}^H \tilde{\mathbf{Q}}_f \tilde{\mathbf{Q}}_f^H \mathbf{z}. \end{aligned} \quad (44)$$

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

Now, suppose there exists a matrix $\mathbf{P}_0 \in \mathbb{C}^{N \times N}$ such that the constraints in (13) hold. We must argue that $R(w) \geq 0$ and therefore $\|\mathbf{Q}\|_{\mathcal{A}}^* \leq 1$ for all w . Consider the expression $\mathbf{z}^H \mathbf{P}_0 \mathbf{z}$ and note that

$$\mathbf{z}^H \mathbf{P}_0 \mathbf{z} = \text{Tr}(\mathbf{z} \mathbf{z}^H \mathbf{P}_0) = \sum_{k=-(N-1)}^{N-1} r_k z^{-k} \quad (45)$$

where $r_k = \sum_{i=1}^{N-k} \mathbf{P}_0(i, i+k)$ for $k \geq 0$ and $r_k = r_{-k}^*$ for $k < 0$. Since $\sum_{i=1}^{N-k} \mathbf{P}_0(i, i+k) = \delta_k$ holds, we can conclude that $\mathbf{z}^H \mathbf{P}_0 \mathbf{z} = \mathbf{z}^0 = 1$. Define $\mathbf{P}_1 := \sum_{f=1}^{N_F} \tilde{\mathbf{Q}}_f \tilde{\mathbf{Q}}_f^H = \tilde{\mathbf{Q}} \tilde{\mathbf{Q}}^H$ and substitute this fact into $R(w)$. We have

$$R(w) = \mathbf{z}^H \mathbf{P}_0 \mathbf{z} - \mathbf{z}^H \mathbf{P}_1 \mathbf{z} = \mathbf{z}^H (\mathbf{P}_0 - \mathbf{P}_1) \mathbf{z}. \quad (46)$$

Since $\begin{bmatrix} \mathbf{P}_0 & \tilde{\mathbf{Q}} \\ \tilde{\mathbf{Q}}^H & \mathbf{I}_{N_l N_F} \end{bmatrix} \succeq 0$, its Schur complement $\mathbf{P}_0 - \tilde{\mathbf{Q}} \mathbf{I}_{N_l N_F}^{-1} \tilde{\mathbf{Q}}^H = \mathbf{P}_0 - \mathbf{P}_1 \succeq 0$, 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 need to argue that there exists a matrix $\mathbf{P}_0 \in \mathbb{C}^{N \times N} \succeq 0$ such that the constraints in (13) hold. Since $R(w) \leq 0$, $1 \geq \mathbf{z}^H \mathbf{P}_1 \mathbf{z}$, where $\mathbf{P}_1 := \sum_{f=1}^{N_F} \tilde{\mathbf{Q}}_f \tilde{\mathbf{Q}}_f^H = \tilde{\mathbf{Q}} \tilde{\mathbf{Q}}^H$. From [53, Lemma 4.25] and the fact that 1 and $\mathbf{z}^H \mathbf{P}_1 \mathbf{z}$ are both 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 $\sum_{i=1}^{N-k} \mathbf{P}_0(i, i+k) = \delta_k$ hold. The matrix $\begin{bmatrix} \mathbf{P}_0 & \tilde{\mathbf{Q}} \\ \tilde{\mathbf{Q}}^H & \mathbf{I}_{N_l N_F} \end{bmatrix}$ has Schur complement $\mathbf{P}_0 - \tilde{\mathbf{Q}} \mathbf{I}_{N_l N_F}^{-1} \tilde{\mathbf{Q}}^H = \mathbf{P}_0 - \mathbf{P}_1 \succeq 0$, and therefore this matrix is positive semi-definite. This concludes the proof. $\square$

REFERENCES

- [1] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. Hoboken, NJ, USA: Wiley, 2002.
- [2] Y. Chen, L. Yan, C. Han, and M. Tao, "Millidegree-level direction-of-arrival estimation and tracking for Terahertz ultra-massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 869–883, Feb. 2022.
- [3] C. Vasaneli et al., "Calibration and direction-of-arrival estimation of millimeter-wave radars: A practical introduction," *IEEE Antennas Propag. Mag.*, vol. 62, no. 6, pp. 34–45, Dec. 2020.
- [4] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.
- [5] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 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, Aug. 1984.
- [7] 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, Dec. 2019.
- [8] 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, Jun. 2019.
- [9] 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, Mar. 2019.
- [10] Y. Wu, M. B. Wakin, and P. Gerstoft, "Gridless DOA estimation with multiple frequencies," *IEEE Trans. Signal Process.*, vol. 71, pp. 417–432, 2023.
- [11] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 33, no. 4, pp. 823–831, Aug. 1985.
- [12] 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, Jul. 1988.
- [13] 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, Oct. 2001.
- [14] 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, Jun. 2006.
- [15] Z. Tang, G. Blaquiere, and G. Leus, "Aliasing-free wideband beamforming using sparse signal representation," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3464–3469, Jul. 2011.
- [16] 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.
- [17] J. Zhang, N. Hu, M. Bao, X. Li, and W. He, "Wideband DOA estimation based on block FOCUSS with limited samples," in *Proc. IEEE Glob. Conf. Signal Inform. Process.*, 2013, pp. 634–637.
- [18] 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, Jan. 2016.
- [19] C. Liu, Y. V. Zakharov, and T. Chen, "Broadband underwater localization of multiple sources using basis pursuit de-noising," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1708–1717, Apr. 2012.
- [20] S. Zhang, A. Ahmed, Y. D. Zhang, and S. Sun, "Enhanced DOA estimation exploiting multi-frequency sparse array," *IEEE Trans. Signal Process.*, vol. 69, pp. 5935–5946, 2021.
- [21] 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.
- [22] E. J. Candes and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Commun. Pure Appl. Math.*, vol. 67, no. 6, pp. 906–956, 2014.
- [23] 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, Nov. 2013.
- [24] 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, Mar. 2016.

- [25] Z. Yang and L. Xie, "Exact joint sparse frequency recovery via optimization methods," *IEEE Trans. Signal Process.*, vol. 64, no. 19, pp. 5145–5157, Oct. 2016.
- [26] 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, Apr. 2019.
- [27] 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.
- [28] Y. Wu, M. B. Wakin, and P. Gerstoft, "Gridless DOA estimation under the multi-frequency model," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2022, pp. 5982–5986.
- [29] Y. Jiang, D. Li, X. Wu, and W.-P. Zhu, "A gridless wideband DOA estimation based on atomic norm minimization," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop (SAM)*, Piscataway, NJ, USA: IEEE Press, 2020, pp. 1–5.
- [30] 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, Apr. 2018.
- [31] 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, Mar. 2020.
- [32] 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, Jun. 2016.
- [33] S. Ling and T. Strohmer, "Self-calibration and biconvex compressive sensing," *Inverse Problems*, vol. 31, no. 11, 2015, Art. no. 115002.
- [34] A. Moffet, "Minimum-redundancy linear arrays," *IEEE Trans. Antennas Propag.*, vol. 16, no. 2, pp. 172–175, Mar. 1968.
- [35] G. S. Bloom and S. W. Golomb, "Applications of numbered undirected graphs," in *Proc. IEEE*, vol. 65, no. 4, pp. 562–570, Apr. 1977.
- [36] P. P. Vaidyanathan and P. Pal, "Sparse sensing with co-prime samplers and arrays," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 573–586, Feb. 2011.
- [37] P. Pal and P. P. Vaidyanathan, "Nested arrays: A novel approach to array processing with enhanced degrees of freedom," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4167–4181, Aug. 2010.
- [38] M. Wang and A. Nehorai, "Coarrays, music, and the Cramér–Rao bound," *IEEE Trans. Signal Process.*, vol. 65, no. 4, pp. 933–946, Feb. 2016.
- [39] M. C. Dogan and J. M. Mendel, "Applications of cumulants to array processing. I. Aperture extension and array calibration," *IEEE Trans. Signal Process.*, vol. 43, no. 5, pp. 1200–1216, May 1995.
- [40] P. Chevalier, L. Albera, A. Ferréol, and P. Comon, "On the virtual array concept for higher order array processing," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1254–1271, Apr. 2005.
- [41] W.-K. Ma, T.-H. Hsieh, and C.-Y. Chi, "DOA estimation of quasi-stationary signals via Khatri-Rao subspace," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2009, pp. 2165–2168.
- [42] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," 2014.
- [43] A. Xenaki and P. Gerstoft, "Grid-free compressive beamforming," *J. Acoust. Soc. Amer.*, vol. 137, no. 4, pp. 1923–1935, 2015.
- [44] P. Gerstoft, C. F. Mecklenbräuker, A. Xenaki, and S. Nannuru, "Multisnapshot sparse Bayesian learning for DOA," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1469–1473, Oct. 2016.
- [45] B. D. Rao and K. S. Hari, "Performance analysis of root-MUSIC," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 37, no. 12, pp. 1939–1949, Dec. 1989.
- [46] 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. 175101–175124, 2020.
- [47] Y. Liang, W. Cui, Q. Shen, W. Liu, and H. Wu, "Cramér–Rao bound for DOA estimation exploiting multiple frequency pairs," *IEEE Signal Process. Lett.*, vol. 28, pp. 1210–1214, 2021.
- [48] Z. Yang, J. Li, P. Stoica, and L. Xie, "Sparse methods for direction-of-arrival estimation," in *Academic Press Library in Signal Processing*, vol. 7. Amsterdam, The Netherlands: Elsevier, 2018, pp. 509–581.
- [49] Y. Wu, M. B. Wakin, P. Gerstoft, and Y. Park, "Non-uniform frequency spacing for regularization-free gridless DOA," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2024, pp. 9291–9295.
- [50] C. Zhou, Y. Gu, X. Fan, Z. Shi, G. Mao, and Y. D. Zhang, "Direction-of-arrival estimation for coprime array via virtual array interpolation," *IEEE Trans. Signal Process.*, vol. 66, no. 22, pp. 5956–5971, Nov. 2018.
- [51] S. Qin, Y. D. Zhang, M. G. Amin, and B. Himed, "DOA estimation exploiting a uniform linear array with multiple co-prime frequencies," *Signal Process.*, vol. 130, pp. 37–46, Jan. 2017.
- [52] S. Liu, Z. Mao, Y. D. Zhang, and Y. Huang, "Rank minimization-based Toeplitz reconstruction for DoA estimation using coprime array," *IEEE Commun. Lett.*, vol. 25, no. 7, pp. 2265–2269, Jul. 2021.
- [53] B. Dumitrescu, *Positive Trigonometric Polynomials and Signal Processing Applications*, vol. 103. New York, NY, USA: Springer-Verlag, 2017.



Yifan Wu received the M.S. and Ph.D. degrees from the University of California, San Diego, in 2021, and 2023, respectively. He is currently a Postdoctoral Researcher with Colorado School of Mines. His research interests include statistical signal processing, machine learning, tensor decomposition techniques, and optimization theory.



Michael B. Wakin (Fellow, IEEE) received the B.S. degree in electrical engineering, the B.A. degree in mathematics, in 2000 (*summa cum laude*), and the M.S. and Ph.D. degrees in electrical engineering, in 2002 and 2007, respectively, all from Rice University. He is a Professor of electrical engineering with Colorado School of Mines. He was an NSF Mathematical Sciences Postdoctoral Research Fellow with Caltech from 2006 to 2007, an Assistant Professor with the University of Michigan from 2007 to 2008, and a Ben L. Fryrear Associate Professor with Mines from 2015 to 2017. His research interests include signal and data processing using sparse, low-rank, and manifold-based models. In 2007, he shared Hershel M. Rich Invention Award from the Rice University for the design of a single-pixel camera based on compressive sensing. In 2008, he received the DARPA Young Faculty Award for his research in compressive multi-signal processing for environments such as sensor and camera networks. In 2012, he received the NSF CAREER Award for research into dimensionality reduction techniques for structured data sets. In 2014, he received the Excellence in Research Award for his research as a Junior Faculty Member with Mines. He is a recipient of the Best Paper Award and the Signal Processing Magazine Best Paper Award from IEEE Signal Processing Society. He has served as an Associate Editor for both IEEE SIGNAL PROCESSING LETTERS and IEEE TRANSACTIONS ON SIGNAL PROCESSING, and as a Senior Area Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING.



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