Efficient Super-Resolution Two-Dimensional Harmonic Retrieval with Multiple Measurement Vectors

Zhang, Yu; Wang, Yue; Tian, Zhi; Leus, Geert; Zhang, Gong

DOI
10.1109/TSP.2022.3150964

Publication date
2022

Document Version
Final published version

Published in
IEEE Transactions on Signal Processing

Efficient Super-Resolution Two-Dimensional Harmonic Retrieval With Multiple Measurement Vectors

Yu Zhang, *Student Member, IEEE*, Yue Wang, *Senior Member, IEEE*, Zhi Tian, *Fellow, IEEE*, Geert Leus, *Fellow, IEEE*, and Gong Zhang, *Member, IEEE*

***Abstract*—This paper develops an efficient solution for super-resolution two-dimensional (2D) harmonic retrieval from multiple measurement vectors (MMV). Given the sample covariance matrix constructed from the MMV, a gridless compressed sensing approach is proposed based on the atomic norm minimization (ANM). In the approach, our key step is to perform a redundancy reduction (RR) transformation that effectively reduces the large problem size at hand, without loss of useful frequency information. For uncorrelated sources, the transformed 2D covariance matrices in the RR domain retain a salient structure, which permits a sparse representation over a matrix-form atom set with decoupled 1D frequency components. Accordingly, the decoupled ANM (D-ANM) framework can be applied for super-resolution 2D frequency estimation. Moreover, the resulting RR-enabled D-ANM technique, termed RR-D-ANM, further allows an efficient relaxation under certain conditions, which leads to low computational complexity of the same order as the 1D case. Simulation results verify the advantages of our solutions over benchmark methods, in terms of higher computational efficiency and detectability for 2D harmonic retrieval.**

***Index Terms*—Super-resolution, 2D harmonic retrieval, MMV, RR transformation, D-ANM.**

Manuscript received February 5, 2021; revised July 7, 2021 and December 31, 2021; accepted January 25, 2022. Date of publication February 14, 2022; date of current version March 18, 2022. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ketan Rajawat. This work was supported in part by the U.S. NSF under Grants 1527396, 1939553, 2003211, 2128596, and 2136202, in part by Virginia Research Investment Fund CCI under Grant 223996, in part by the China NSFC under Grants 61871218, 61801211, and 61471191, and in part by the ASPIRE Project, Project 14926 within the OTP Program of NWO-TTW. This work was presented at the 28th European Signal Processing Conference, Amsterdam, The Netherlands, January 2021 [1]. *(Corresponding author: Yue Wang.)*

Yu Zhang is with the Key Laboratory of Radar Imaging and Microwave Photonics, Ministry of Education, Nanjing University of Aeronautics and Astronautics, Nanjing 211100, China, and also with the Electrical and Computer Engineering Department, George Mason University, Fairfax, VA 22030 USA (e-mail: skywalker_zy@163.com).

Yue Wang and Zhi Tian are with the Electrical and Computer Engineering Department, George Mason University, Fairfax, VA 22030 USA (e-mail: ywang56@gmu.edu; ztian1@gmu.edu).

Geert Leus is with the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 2826 CD Delft, The Netherlands (e-mail: g.j.t.leus@tudelft.nl).

Gong Zhang is with the Key Laboratory of Radar Imaging and Microwave Photonics, Ministry of Education, Nanjing University of Aeronautics and Astronautics, Nanjing 211100, China (e-mail: gzhang@nuaa.edu.cn).

Digital Object Identifier 10.1109/TSP.2022.3150964

I. Introduction

TWO-DIMENSIONAL (2D) harmonic retrieval has broad applications in speech processing [2], wireless communications [3], radar systems [4], [5], etc. A number of classical high-resolution methods have been developed, mostly based on statistical analysis of the sample covariance [6]–[12]. These methods usually require a large number of measurements in order to well approximate the desired statistics, which may not work effectively in modern applications of a large problem size and limited sampling resources such as short sensing time and compressive measurements. In recent decades, compressed sensing (CS) based techniques that utilize signal sparsity have attracted much interest and have been widely applied for signal recovery from compressive measurements, based on either a single measurement vector (SMV) [13], [14] or multiple measurement vectors (MMV) [15]–[20]. The CS based methods with MMV are further classified into two categories: sample-based methods (where the observed signals are directly adopted as the multiple inputs to CS techniques) [15], [16], and covariance-based methods (where the covariance matrix of the observed signals is calculated and then fed into subsequent CS recovery operations) [17]–[19]. While the former work well for a single or few snapshots, the latter are computationally much more efficient when the number of MMV is large. Compared with the classical subspace methods [8], [9], the CS approaches with MMV exhibit several advantages: robustness to noise, no requirement on the known number of sources *a priori*, and high sample efficiency.

However, the traditional CS approaches critically hinge on an on-grid assumption that the signal frequency components reside exactly on some predefined grid in the spectral domain. Such on-grid CS solutions offer limited frequency resolution and may suffer from severe performance degradation due to the basis mismatch [21]. To circumvent this issue, a super-resolution technology based on the atomic norm minimization (ANM) principle has been developed for line spectral estimation from compressive measurements [22], [23], for both the SMV and MMV cases [24]–[27]. These ANM techniques utilize the Vandermonde structure of a uniform geometry by enforcing a Toeplitz structure in a semidefinite programming (SDP) based convex optimization problem, which allows super-resolution harmonic retrieval via Vandermonde decomposition.

When the number of MMV is large, an alternative covariance-based gridless technique, termed low-rank structured covariance reconstruction (LRSCR), was proposed for the 1D case [27], which builds on the sample covariance matrix, at computational complexity that is independent of the number of measurement vectors. It is worth noting that the LRSCR jointly incorporates two important properties, i.e., the low-rankness and the Toeplitz structure of the covariance matrix. In contrast, classical structured covariance estimation techniques do not explicitly impose low-rankness [28], [29].

The goal of this paper is to bring the benefits of ANM to 2D harmonic retrieval problems in the presence of MMV, when the problem size is large yet sampling resources are limited. Along this line, the ANM techniques originally developed for 1D scenarios have been extended to 2D scenarios [30]. In vectorization-based ANM (V-ANM), the main idea is to vectorize the 2D data matrix and then cast the two-level Vandermonde structure of the data into a proper SDP formulation under the ANM objective. This technique has been further extended to higher-dimensional harmonic retrieval through generalized Vandermonde decomposition on multi-level Toeplitz matrices [31], and applies to both the SMV and MMV scenarios [32]. Unfortunately, the computational complexity of V-ANM scales exponentially with the dimensionality, and increases drastically as the number of MMV grows. In contrast, the decoupled atomic norm minimization (D-ANM) algorithm [33], by virtue of its frequency decoupling strategy, effectively reduces the computational complexity to be comparable to that of a 1D ANM solution, without loss of optimality. The advantage in computation comes with a stricter yet still mild conditions on the frequency separation and the number of identifiable frequencies [34]. However, D-ANM only applies to the SMV case [33], [34]. To the best of our knowledge, there is still a lack of computationally efficient gridless 2D harmonic retrieval techniques for MMV problems.

To fill this gap, this paper aims at developing decoupling-based 2D harmonic retrieval techniques for the MMV case, by exploiting the structural information of the sample covariance matrix effectively and efficiently. The sample covariance encompasses useful information of all MMV, and can be converted to an SMV form through vectorization. However, it is quite challenging to directly solve this straightforward but naive vectorization-based ANM problem, because the large problem size in the 2D scenario results in unacceptably huge computational complexity. In addition, the frequency resolution would remain the same as that of the SMV case, even in the presence of MMV. To overcome these challenges, we propose an efficient super-resolution 2D harmonic retrieval method with MMV, termed redundancy-reduction (RR) transformation based D-ANM-Relaxation (RR-D-ANM-R). In doing so, our contributions are summarized as follows.

- For 2D harmonic retrieval, given the inherent two-level Toeplitz structure of the covariance matrix in the uncorrelated case, we propose a RR transformation to concisely express the vectorized covariance matrix as a RR vector of much reduced length via linear projection. This RR vector can be sparsely represented by a decoupled atom set, which facilitates its efficient decomposition for 2D harmonic retrieval with a much reduced problem size.
- Based on the RR transformation, we propose a novel formulation for gridless 2D harmonic retrieval with MMV via a customized D-ANM (RR-D-ANM). Further, to make the problem tractable, we equivalently reformulate the original formulation as a solvable convex form based on the existence and uniqueness of the generalized Vandermonde decomposition of the covariance matrix.
- Capitalizing on the strong structured constraint introduced by the RR-enabled D-ANM framework, the RR-D-ANM further allows efficient relaxation as RR-D-ANM-R, which is guaranteed by our theoretical analysis for noise-free and noisy cases. It leads to considerable reduction of complexity at a negligible loss in estimation performance, compared with the RR-D-ANM.
- Based on several structured matrices recovered by our gridless 2D MMV techniques, we design three harmonic retrieval methods for different practical implementation needs. Also, for the noisy case, we provide the construction of a noise-tolerance constraint for the proposed formulations, when the number of MMV is adequately large and the noise statistics are unavailable in practice.

The rest of this paper is organized as follows. Section II first presents the signal model and problem formulation for 2D harmonic retrieval with MMV, and then reviews the relevant prior work on V-ANM, D-ANM and LRSCR. Section III proposes an efficient gridless 2D harmonic retrieval framework based on the RR transformation, under which the RR-enabled D-ANM solutions are developed in the transformed domain. Section IV discusses related design issues. Section V presents simulation results, followed by conclusions in Section VI.

Notations: a, $\boldsymbol{a}$, $\boldsymbol{A}$, $\mathcal{A}$ and $\backslash\mathcal{A}$ denote a scalar, a vector, a matrix, a set and the complement of the set $\mathcal{A}$. $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^H$ are the transpose, conjugate, and conjugate transpose of a vector or matrix. $\mathrm{conv}(\mathcal{A})$ means the convex hull of a set $\mathcal{A}$. $\mathrm{Real}(a)$ and $|a|$ denote the real value and the absolute value of a, respectively. The sign of a is defined as $\mathrm{sign}(a) = \exp(j\phi(a))$ where $\phi(a)$ denotes the phase of a. $\|\boldsymbol{a}\|_0$, $\|\boldsymbol{a}\|_1$ and $\|\boldsymbol{a}\|_2$ are the ℓ_0, ℓ_1 and ℓ_2 norms of $\boldsymbol{a}$. $\mathrm{diag}(\boldsymbol{a})$ generates a diagonal matrix with the diagonal elements constructed from $\boldsymbol{a}$. $\mathrm{vec}(\cdot)$ stacks all the columns of a matrix into a vector and $\mathrm{vecd}(\boldsymbol{A})$ returns a vector consisting of the diagonal elements of $\boldsymbol{A}$ only. $\mathrm{vec}^{-1}(\cdot)$ is the inverse operation of $\mathrm{vec}(\cdot)$. $\boldsymbol{e}_a$ is the vector with only the a-th element being one and zeros elsewhere. $\boldsymbol{I}_a$ is an a-size identity matrix. $\mathcal{S}(\boldsymbol{A})$ indicates a unique mapping from $\boldsymbol{A}$ to a two-level Toeplitz matrix T_{2D}. $\|\boldsymbol{A}\|_F$, $\mathrm{Tr}(\boldsymbol{A})$, $\mathrm{Rank}(\boldsymbol{A})$ are the Frobenius norm, the trace and the rank of $\boldsymbol{A}$. $\otimes$ is the Kronecker product. $\odot$ is the Khatri-Rao product. $\delta_{ij} \in \{1, 0\}$ is the Dirac delta function. $\mathbb{E}\{\cdot\}$ denotes expectation.

II. Signal Model and Prior Art

This section presents the signal model and problem statement of 2D harmonic retrieval with MMV. A brief review of related prior works is provided to highlight their limitations.

A. Signal Model

Consider a 2D harmonic retrieval problem where the signal of interest $\boldsymbol{x}(t) \in \mathbb{C}^{NM}$ is a linear mixture of K 2D sinusoidal components in the form of [1]

$$\boldsymbol{x}(t) = \sum_{i=1}^{K} s_i(t)[\boldsymbol{a}_N(f_{1,i}) \otimes \boldsymbol{a}_M(f_{2,i})] = \sum_{i=1}^{K} s_i(t)\boldsymbol{a}_{2D}(\boldsymbol{f}^i) = (\boldsymbol{A}_N \odot \boldsymbol{A}_M)\boldsymbol{s}(t), \quad t = 1, \ldots, L, \tag{1}$$

where $s_i(t)$ is the complex amplitude of the i-th source, $\boldsymbol{f}^i = [f_{1,i}, f_{2,i}]^T \in (-\frac{1}{2}, \frac{1}{2}]^2$ consists of its digital frequencies along the two orthogonal dimensions, and L is the number of measurement vectors. The 2D manifold vector $\boldsymbol{a}_{2D}(\boldsymbol{f}^i) = \boldsymbol{a}_N(f_{1,i}) \otimes \boldsymbol{a}_M(f_{2,i})$ is made of Vandermonde-structured manifold vectors $\boldsymbol{a}_N(f_{1,i})$ and $\boldsymbol{a}_M(f_{2,i})$ of size N and M, respectively:[1]

$$\boldsymbol{a}_N(f_{1,i}) = [1, \exp(j2\pi f_{1,i}), \ldots, \exp(j2\pi(N-1)f_{1,i})]^T,$$
$$\boldsymbol{a}_M(f_{2,i}) = [1, \exp(j2\pi f_{2,i}), \ldots, \exp(j2\pi(M-1)f_{2,i})]^T. \tag{2}$$

Further, $\boldsymbol{s}(t) = [s_1(t), \ldots, s_K(t)]^T$, $\boldsymbol{f}_n = [f_{n,1}, \ldots, f_{n,K}]^T$ for $n = 1, 2$, $\boldsymbol{A}_N = \boldsymbol{A}_N(\boldsymbol{f}_1) = [\boldsymbol{a}_N(f_{1,1}), \ldots, \boldsymbol{a}_N(f_{1,K})]$, and $\boldsymbol{A}_M = \boldsymbol{A}_M(\boldsymbol{f}_2) = [\boldsymbol{a}_M(f_{2,1}), \ldots, \boldsymbol{a}_M(f_{2,K})]$. In a matrix form, (1) becomes $\boldsymbol{X}_L = (\boldsymbol{A}_N \odot \boldsymbol{A}_M)\boldsymbol{S}_L$, where $\boldsymbol{X}_L = [\boldsymbol{x}(1), \ldots, \boldsymbol{x}(L)] \in \mathbb{C}^{NM \times L}$, $\boldsymbol{S}_L = [\boldsymbol{s}(1), \ldots, \boldsymbol{s}(L)] \in \mathbb{C}^{K \times L}$.

Note that the signal vector $\boldsymbol{x}(t)$ may arise from the vectorization of a matrix $\boldsymbol{X}(t) \in \mathbb{C}^{M \times N}$ via $\boldsymbol{x} = \text{vec}(\boldsymbol{X})$, where

$$\boldsymbol{X}(t) = \sum_{i=1}^{K} s_i(t)\boldsymbol{a}_M(f_{2,i})\boldsymbol{a}_N^T(f_{1,i}) = \boldsymbol{A}_M \text{diag}(\boldsymbol{s}(t))\boldsymbol{A}_N^T. \tag{3}$$

For instance, $\boldsymbol{X}(t)$ can be the samples collected from an $M \times N$ rectangular array at the t-th snapshot, when K plane-wave signals impinge on the array. The equivalence between (1) and (3) derives from the properties of Kronecker and Khatri-Rao products for a diagonal $\boldsymbol{S}$,

$$\text{vec}(\boldsymbol{ASB}) = (\boldsymbol{B}^T \otimes \boldsymbol{A})\text{vec}(\boldsymbol{S}) = (\boldsymbol{B}^T \odot \boldsymbol{A})\text{vecd}(\boldsymbol{S}). \tag{4}$$

In many applications, $\boldsymbol{x}(t)$ is not observed directly, but through subsampling or linear compression via a measurement matrix $\boldsymbol{J} \in \mathbb{C}^{M' \times NM}$ with $M' \leq NM$, such as in bistatic multiple-input and multiple-output (MIMO) radar systems with sparse linear array [5] and in hybrid millimeter-wave massive MIMO systems [20], [37]–[39]. Inflicted by an additive noise $\boldsymbol{n}(t)$, the SMV data $\boldsymbol{y}(t) \in \mathbb{C}^{M'}$ is given by

$$\boldsymbol{y}(t) = \boldsymbol{J}\boldsymbol{x}(t) + \boldsymbol{n}(t). \tag{5}$$

Collecting all $\boldsymbol{y}(t)$'s and $\boldsymbol{n}(t)$'s to form matrices $\boldsymbol{Y}_L$ and $\boldsymbol{N}_L$ respectively, the MMV model is

$$\boldsymbol{Y}_L = \boldsymbol{J}\boldsymbol{X}_L + \boldsymbol{N}_L = \boldsymbol{J}(\boldsymbol{A}_N \odot \boldsymbol{A}_M)\boldsymbol{S}_L + \boldsymbol{N}_L. \tag{6}$$

[1]When non-ideal geometries are encountered, e.g. antenna systems with perturbation, array manifold separation can be applied to retrieve the Vandermonde structure through a Bessel or Fourier approximation [35], [36].

Given random $\boldsymbol{s}(t)$, the desired frequency information lies in the covariance of $\boldsymbol{y}(t)$, defined by

$$\boldsymbol{R}_y = \mathbb{E}\{\boldsymbol{y}(t)\boldsymbol{y}(t)^H\} = \boldsymbol{J}\boldsymbol{R}_x\boldsymbol{J}^H + \boldsymbol{R}_n, \tag{7}$$

where $\boldsymbol{R}_x$ and $\boldsymbol{R}_n$ are the covariance of $\boldsymbol{x}(t)$ and $\boldsymbol{n}(t)$, respectively. Denoting $\boldsymbol{R}_s = \mathbb{E}\{\boldsymbol{s}(t)\boldsymbol{s}(t)^H\}$, we have

$$\boldsymbol{R}_x = \mathbb{E}\{\boldsymbol{x}(t)\boldsymbol{x}(t)^H\} = (\boldsymbol{A}_N \odot \boldsymbol{A}_M)\boldsymbol{R}_s(\boldsymbol{A}_N \odot \boldsymbol{A}_M)^H. \tag{8}$$

This paper considers uncorrelated signal sources, that is, $\mathbb{E}\{s_i(t)s_j^*(t)\} = r_i\delta_{ij}$, $\forall 1 \leq i, j \leq K$. Since $r_i \geq 0$, $\forall i$, $\boldsymbol{R}_s$ is a positive semidefinite (PSD) diagonal matrix:

$$\boldsymbol{R}_s = \text{diag}(\boldsymbol{r}) \succeq 0, \quad \text{where } \boldsymbol{r} = [r_1, \ldots, r_K]^T \geq \boldsymbol{0}. \tag{9}$$

In practice, $\boldsymbol{R}_y$ is approximated by its sample covariance

$$\hat{\boldsymbol{R}}_y = \frac{1}{L}\boldsymbol{Y}_L\boldsymbol{Y}_L^H. \tag{10}$$

The goal of 2D harmonic retrieval is to recover the unknown frequency pairs $\{\boldsymbol{f}^i\}_i$ from either $\boldsymbol{Y}_L$ or $\hat{\boldsymbol{R}}_y$, corresponding to the sample-based or the covariance-based approach.

B. Prior Art

We now review existing approaches for 2D harmonic retrieval, with focus on gridless CS techniques. The limitations of these techniques for the MMV case motivate this work.

1) Statistical Subspace Methods: The 2D frequency estimates can be jointly acquired from the signal subspace of $\boldsymbol{R}_y$, by utilizing the two-level Vandermonde structure exhibited in $\boldsymbol{R}_x$ in (8) [10]–[12]. However, traditional subspace methods require L to be large in order to accurately estimate the sought covariance via sample averaging. For instance, to approximate $\boldsymbol{R}_y$ in (7) by (10), it requires $L \geq M'$ in the worst case when $K(<M')$ is unknown *a priori*. Besides, they may suffer from degraded estimation accuracy in the presence of subsampling or compression, i.e., $M' < NM$. This means that both the number of MMV L and the compressive sample vector length M' have to be sufficiently large. Hence, traditional subspace methods usually experience low sample efficiency.

2) Sample-Based Vectorized 2D ANM (V-ANM): Gridless CS builds on the tenet that the signal of interest can be concisely expressed as a linear combination of a few simple atoms over a known atom set, and atomic norm minimization (ANM) is able to reveal the atomic composition under some conditions. In V-ANM, the signal of interest is $\boldsymbol{x}(t)$ in the SMV case or $\boldsymbol{X}_L$ in the MMV case. Based on (1), a vectorization-based atom set is defined as [30], [31]

$$\mathcal{A}_v = \left\{ \frac{1}{\sqrt{NM}}\boldsymbol{a}_{2D}(\boldsymbol{f}) \,\middle|\, \forall\, \boldsymbol{f} = [f_1, f_2]^T \in (-\tfrac{1}{2}, \tfrac{1}{2}]^2 \right\}. \tag{11}$$

Accordingly, the atomic norm of $\boldsymbol{x}$ (to replace $\boldsymbol{x}(t)$ hereafter) over the atom set $\mathcal{A}_v$ is defined as

$$\|\boldsymbol{x}\|_{\mathcal{A}_v} = \inf\{l > 0 : \boldsymbol{x} \in l\text{conv}(\mathcal{A}_v)\}, \tag{12}$$

which seeks the sparsest (under the ℓ_1-norm measure) decomposition of $\boldsymbol{x}$ over $\mathcal{A}_v$. If the frequencies are adequately separated to meet the separation condition [30], [31], then the atomic decomposition yields the true signal structure in (1), through the following V-ANM formulation

$$\widetilde{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_{\mathcal{A}_v} \text{ s.t. } \|\boldsymbol{y} - \boldsymbol{J}\boldsymbol{x}\|_2^2 \le \mu, \tag{13}$$

where μ is a user-specified parameter for error tolerance.

Calculation of the atomic norm is difficult, particularly when the atom set is of infinite size. Fortunately, $\mathcal{A}_v$ possesses a special structure indicated by the Vandermonde vectors in (2), which leads to a SDP reformulation of V-ANM for computationally tractable solutions. Provided that $K < \min\{N, M\}$ in addition to the separation condition, the SDP-based V-ANM for both SMV ($L = 1$) and MMV is [30]–[32]

$$\begin{aligned}\{\widetilde{\boldsymbol{u}}_{2D}, \widetilde{\boldsymbol{V}}, \widetilde{\boldsymbol{X}}_L\} = \arg\min_{\boldsymbol{u}_{2D}, \boldsymbol{V}, \boldsymbol{X}_L} & \frac{1}{2}(\mathrm{Tr}\left(\mathrm{T}_{2D}(\boldsymbol{u}_{2D})\right) + \mathrm{Tr}(\boldsymbol{V})) \\ \text{s.t.} \quad & \begin{bmatrix} \mathrm{T}_{2D}(\boldsymbol{u}_{2D}) & \boldsymbol{X}_L \\ \boldsymbol{X}_L^H & \boldsymbol{V} \end{bmatrix} \succeq 0 \\ & \|\boldsymbol{Y}_L - \boldsymbol{J}\boldsymbol{X}_L\|_F^2 \le \mu. \end{aligned} \tag{14}$$

Here $\mathrm{T}_{2D}(\boldsymbol{u}_{2D}) \in \mathbb{C}^{NM\times NM}$ is a Hermitian two-level Toeplitz matrix parameterized by $\boldsymbol{u}_{2D} \in \mathbb{C}^{2NM-(N+M)+1}$, which consists of the distinct entries (the conjugate entries are not included) in its first M columns [30]. Once $\widetilde{\boldsymbol{u}}_{2D}$ is acquired from (14), the 2D frequencies can be extracted from $\mathrm{T}_{2D}(\widetilde{\boldsymbol{u}}_{2D})$ via generalized Vandermonde decomposition [31]. Alternatively, traditional 2D subspace methods can be adopted for accurate frequency estimation from the well-conditioned $\mathrm{T}_{2D}(\widetilde{\boldsymbol{u}}_{2D})$ in lieu of the unknown $\boldsymbol{R}_x$.

Note that the V-ANM is a sample-based approach, building on $\boldsymbol{Y}_L$. Moreover, $\mathrm{T}_{2D}(\widetilde{\boldsymbol{u}}_{2D})$ can be constructed even when strong compression is present with $M' \le NM$, and for few snapshots $L \ll NM$. Hence, V-ANM enjoys high sample efficiency. However, its computational cost scales exponentially in the signal dimensionality, because of the multi-level Toeplitz matrix involved. The PSD matrix in the first constraint of (14) is of size $(NM + L) \times (NM + L)$, which results in a complexity order of $\mathcal{O}((NM + L)^{4.5}L^2)$ in SDP implementation [40].

Furthermore, the exact equivalence between the SDP formula in (14) and the original ANM entails an extra condition, that is, $\mathrm{Rank}(\mathrm{T}_{2D}(\boldsymbol{u}_{2D})) < \min\{N, M\}$. This results from the fact that the equivalence between the PSD property of the two-level Toeplitz matrices and the existence and uniqueness of their generalized Vandermonde decomposition only holds under such a condition [31]. Since this constraint cannot be explicitly imposed in the SDP, some checking mechanism has to be applied to the solution of (14), which incurs extra operations [31], [41]. Overall, V-ANM is applicable to problems of small size only, because of its very high computational cost.

3) Sample-Based Decoupled 2D ANM (D-ANM): The disadvantage of V-ANM is overcome by a D-ANM approach [33], [34], which operates on the matrix-form signal $\boldsymbol{X}(t) = \mathrm{vec}^{-1}(\boldsymbol{x}(t))$ in (3). Accordingly, it introduces a matrix-form atom set

$$\mathcal{A}_d = \left\{ \boldsymbol{a}_M(f_2)\boldsymbol{a}_N^H(f_1) \,\middle|\, \forall \boldsymbol{f} = [f_1, f_2]^T \in (-\tfrac{1}{2}, \tfrac{1}{2}]^2 \right\}. \tag{15}$$

Similar to (13), under a certain separation condition [34], the true frequency components of $\boldsymbol{X}$ (to replace $\boldsymbol{X}(t)$ hereafter) correspond to its sparsest decomposition, and can be found by calculating $\|\boldsymbol{X}\|_{\mathcal{A}_d}$ via the following D-ANM [33], [34]:

$$\widetilde{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}} \|\boldsymbol{X}\|_{\mathcal{A}_d} \text{ s.t. } \|\boldsymbol{y} - \boldsymbol{J}\mathrm{vec}(\boldsymbol{X})\|_2^2 \le \mu. \tag{16}$$

Utilizing the inherent Vandermonde structure, (16) is equivalent to the following SDP when $K < \min\{N, M\}$ [33], [34]

$$\begin{aligned}\{\widetilde{\boldsymbol{u}}_1, \widetilde{\boldsymbol{u}}_2, \widetilde{\boldsymbol{X}}\} = \arg \min_{\substack{\boldsymbol{X},\, \boldsymbol{u}_1 \in \mathbb{C}^N, \\ \boldsymbol{u}_2 \in \mathbb{C}^M}} & \frac{1}{2\sqrt{NM}}(\mathrm{Tr}(\mathrm{T}(\boldsymbol{u}_1)) \\ & + \mathrm{Tr}(\mathrm{T}(\boldsymbol{u}_2))) \quad \text{s.t.} \quad \begin{bmatrix} \mathrm{T}(\boldsymbol{u}_2) & \boldsymbol{X} \\ \boldsymbol{X}^H & \mathrm{T}(\boldsymbol{u}_1) \end{bmatrix} \succeq 0 \\ & \|\boldsymbol{y} - \boldsymbol{J}\mathrm{vec}(\boldsymbol{X})\|_2^2 \le \mu, \end{aligned} \tag{17}$$

where $\mathrm{T}(\boldsymbol{u}_n)$, $n = 1, 2$, forms a one-level Hermitian Toeplitz matrix with $\boldsymbol{u}_n$ being its first column. Following (17), $\boldsymbol{f}_n$ can be extracted from the estimated $\mathrm{T}(\widetilde{\boldsymbol{u}}_n)$ individually, $n = 1, 2$, via standard 1D Vandermonde decomposition. While the separation condition of D-ANM is more restrictive than that of V-ANM, D-ANM reduces the complexity order to be comparable to a 1D problem at $\mathcal{O}(N^2M^2(N + M)^{2.5})$, noting that the PSD constraint in (17) is of size $(N + M) \times (N + M)$ only. Remarkably, the decoupling in computation not only retains performance optimality, but also facilitates frequency pairing, because $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ are jointly recovered from the SMV data $\boldsymbol{y}$. However, the D-ANM framework in (17) only works for the SMV case with $\boldsymbol{x}$ collected from a single snapshot. It is not straightforward to directly extend (17) to the MMV case.

4) Covariance-Based LRSCR: For harmonic retrieval, the essence of the ANM approach is to construct a structured matrix $\mathrm{T}_{2D}(\boldsymbol{u}_{2D})$ from the data set $\boldsymbol{Y}_L$ that possesses the desired properties of the idealized signal covariance $\boldsymbol{R}_x$, including being low rank, PSD, and two-level Hermitian Toeplitz. Alternative to sample-based V-ANM, $\boldsymbol{R}_x$ with these properties can be extracted from the sample covariance $\hat{\boldsymbol{R}}_y$, through an LRSCR formulation [42] as follows:

$$\begin{aligned}\widetilde{\boldsymbol{R}}_x = & \arg \min_{\boldsymbol{R}_x = \mathrm{T}_{2D}(\boldsymbol{u}_{2D})} \mathrm{Tr}\left(\boldsymbol{R}_x\right) \\ & \text{s.t.} \left\| \hat{\boldsymbol{R}}_y - \boldsymbol{J}\boldsymbol{R}_x\boldsymbol{J}^H \right\|_F^2 \le \mu' \\ & \boldsymbol{R}_x \succeq 0, \end{aligned} \tag{18}$$

where μ' is similarly defined as μ. Note that the LRSCR in (18) permits a SDP reformulation, similar to the form of (14). For both (18) and its SDP counterpart, the size of any PSD constraint involved is independent of L, which dictates the complexity order of the LRSCR to be $\mathcal{O}((NM)^{4.5})$. This is lower than that of the V-ANM when L is large, but still quite high for practical use in large-size problems when N and/or M go large.

In summary, the aforementioned super-resolution schemes for the MMV case encounter major limitations in terms of either sample efficiency or computational efficiency. The existing D-ANM is efficient in both accounts, but is only applicable to the SMV case and subject to stricter frequency separation conditions. In large-scale problems such as spatial frequency estimation in massive MIMO systems [43], [44], the number of samples is limited due to both real-time processing constraints and sampling costs. Meanwhile, the complexity order has to be amenable for affordable implementation. Therefore, our goal in this work is to develop new efficient super-resolution harmonic retrieval techniques for MMV, by exploiting the mechanism of D-ANM in a covariance-based manner. Moreover, in the existing 2D harmonic retrieval based on the generalized Vandermonde decomposition of the covariance matrix, the number of identifiable sources is limited to be $K < \min\{M, N\}$, which is too conservative, considering the extra information that MMV can provide beyond SMV. Accordingly, we aim to loosen the conservative conditions for guaranteeing exact 2D solutions in the MMV case, by effectively capitalizing on the inherent structure of covariance matrices.

III. Redundancy-Reduction Transformation Based Decoupled Atomic Norm Minimization

This section develops a super-resolution 2D frequency estimation scheme based on the D-ANM scheme, in the presence of MMV. Redundancy-reduction transformation of the signal covariance is introduced, which effectively reduces the computational cost to be comparable to that of a 1D problem.

A. Redundancy-Reduction Transformation

To take advantage of the low complexity of the D-ANM for the SMV case, we turn to $\boldsymbol{r}_x = \text{vec}(\boldsymbol{R}_x) \in \mathbb{C}^{(NM)^2}$ as the structured signal vector of interest. This vector can be approximated by $\text{vec}(\frac{1}{L}\boldsymbol{X}_L\boldsymbol{X}_L^H)$, which retains the useful 2D frequency information in all the columns of $\boldsymbol{X}_L$. Hence, an MMV problem based on $\boldsymbol{X}_L$ can be alternatively solved as an SMV problem based on $\boldsymbol{r}_x$, which is amenable to the D-ANM.

However, the benefit of adopting $\boldsymbol{r}_x$ as the signal of interest comes at the expense of a much enlarged signal length of $(NM)^2$, which incurs a high computational cost when N and/or M go large. To circumvent this issue, we propose a RR transformation to concisely express $\boldsymbol{r}_x$, by removing the redundancy in the entries of $\boldsymbol{r}_x$. The basic idea is to establish a linear mapping that projects the vectorized covariance matrix $\boldsymbol{r}_x$ with repeated entries onto an RR vector $\boldsymbol{z}$ with no repeated entries. A similar thought can be found in our previous work on 1D problems where we establish the linear mapping between the covariance matrices in the compressed and uncompressed domains for compressive cyclic feature detection [45], [46] and DOA estimation [47]. In this work, the RR transformation for the 2D case is much more involved, and critically hinges on the underlying multi-level Vandermonde structures of the manifold vectors that have not been studied under the context of RR, as derived next.

The inherent redundancy in $\boldsymbol{r}_x$ is due to the uncorrelated sources, which yields $\boldsymbol{R}_s = \text{diag}(\boldsymbol{r})$ with only K nonzero entries. By vectorizing both sides of (8), it follows that

$$\boldsymbol{r}_x = \sum_{i=1}^{K} r_i \left(\boldsymbol{a}_N(f_{1,i}) \otimes \boldsymbol{a}_M(f_{2,i})\right)^* \otimes \left(\boldsymbol{a}_N(f_{1,i}) \otimes \boldsymbol{a}_M(f_{2,i})\right). \tag{19}$$

Concerning each summand in (19), an equality arises:

$$\begin{aligned} &\left(\boldsymbol{a}_N(f_{1,i}) \otimes \boldsymbol{a}_M(f_{2,i})\right)^* \otimes \left(\boldsymbol{a}_N(f_{1,i}) \otimes \boldsymbol{a}_M(f_{2,i})\right) \\ &= \boldsymbol{\Psi}(\boldsymbol{a}'_N(f_{1,i}) \otimes \boldsymbol{a}'_M(f_{2,i})), \end{aligned} \tag{20}$$

where $\boldsymbol{a}'_N(f_{1,i}) \in \mathbb{C}^{2N-1}$ and $\boldsymbol{a}'_M(f_{2,i}) \in \mathbb{C}^{2M-1}$ are given by

$$\begin{aligned} \boldsymbol{a}'_N(f_{1,i}) &= \left[e^{-j2\pi(N-1)f_{1,i}}, \ldots, 1, \ldots, e^{j2\pi(N-1)f_{1,i}}\right]^T, \\ \boldsymbol{a}'_M(f_{2,i}) &= \left[e^{-j2\pi(M-1)f_{2,i}}, \ldots, 1, \ldots, e^{j2\pi(M-1)f_{2,i}}\right]^T. \end{aligned} \tag{21}$$

In (20), the matrix $\boldsymbol{\Psi} = (\boldsymbol{I}_N \otimes \boldsymbol{E} \otimes \boldsymbol{I}_M)(\boldsymbol{G}_N \otimes \boldsymbol{G}_M) \in \mathbb{R}^{N^2M^2 \times (2N-1)(2M-1)}$ is the redundancy-reduction (RR) transformation matrix that is determined by N and M only, where $\boldsymbol{E} = \sum_{j=1}^{M} (\boldsymbol{e}_j^T \otimes \boldsymbol{I}_N \otimes \boldsymbol{e}_j) \in \mathbb{R}^{NM \times NM}$ is the commutation matrix, $\boldsymbol{G}_N$ is defined as

$$\boldsymbol{G}_N = \left[\boldsymbol{G}_{N,1}^T, \ldots, \boldsymbol{G}_{N,N}^T\right]^T \in \mathbb{R}^{N^2 \times (2N-1)}, \tag{22}$$

with the i-th block matrix $\boldsymbol{G}_{N,i} = [\boldsymbol{0}_{N\times(N-i)}, \boldsymbol{I}_N, \boldsymbol{0}_{N\times(i-1)}]$, $i = 1, \ldots, N$, and $\boldsymbol{G}_M \in \mathbb{R}^{M^2 \times (2M-1)}$ is defined similarly as (22). The derivation of (20) is provided in Appendix A.

By defining the Kronecker product term in the right hand side of (20) as

$$\boldsymbol{a}'_{2D}(\boldsymbol{f}^i) = \boldsymbol{a}'_N(f_{1,i}) \otimes \boldsymbol{a}'_M(f_{2,i}), \tag{23}$$

(19) can be rewritten as

$$\boldsymbol{r}_x = \boldsymbol{\Psi}\boldsymbol{z}, \text{where } \boldsymbol{z} = \sum_{i=1}^{K} r_i \boldsymbol{a}'_{2D}(\boldsymbol{f}^i) = (\boldsymbol{A}'_N \odot \boldsymbol{A}'_M)\boldsymbol{r} \tag{24}$$

where $\boldsymbol{A}'_N = \boldsymbol{A}'_N(\boldsymbol{f}_1) = [\boldsymbol{a}'_N(f_{1,1}), \ldots, \boldsymbol{a}'_N(f_{1,K})]$ and $\boldsymbol{A}'_M = \boldsymbol{A}'_M(\boldsymbol{f}_2) = [\boldsymbol{a}'_M(f_{2,1}), \ldots, \boldsymbol{a}'_M(f_{2,K})]$, and $\boldsymbol{r} = \text{vecd}(\boldsymbol{R}_s)$. It follows from (4) that

$$\boldsymbol{z} = (\boldsymbol{A}'_N \odot \boldsymbol{A}'_M)\text{vecd}(\boldsymbol{R}_s) = \text{vec}(\boldsymbol{A}'_M \boldsymbol{R}_s \boldsymbol{A}'^T_N) = \text{vec}(\boldsymbol{Z}) \tag{25}$$

where $\boldsymbol{Z} = \boldsymbol{A}'_M \boldsymbol{R}_s \boldsymbol{A}'^T_N$.

Moreover, $\boldsymbol{z}$ can be observed through the vectorized version of $\boldsymbol{R}_y$ in (7) by noting

$$\begin{aligned} \boldsymbol{r}_y = \text{vec}(\boldsymbol{R}_y) &= \text{vec}(\boldsymbol{J}\boldsymbol{R}_x\boldsymbol{J}^H + \boldsymbol{R}_n) = (\boldsymbol{J}^* \otimes \boldsymbol{J})\boldsymbol{r}_x + \boldsymbol{r}_n \\ &= (\boldsymbol{J}^* \otimes \boldsymbol{J})\boldsymbol{\Psi}\boldsymbol{z} + \boldsymbol{r}_n = \boldsymbol{\Gamma}\boldsymbol{z} + \boldsymbol{r}_n \end{aligned} \tag{26}$$

where $\boldsymbol{\Gamma} = (\boldsymbol{J}^* \otimes \boldsymbol{J})\boldsymbol{\Psi} \in \mathbb{C}^{M'^2 \times c}$ with $c = (2N-1)(2M-1)$, and $\boldsymbol{r}_n = \text{vec}(\boldsymbol{R}_n)$ is the noise term.

It is worth noting the importance of (24) and (26) as the equivalent signal and measurement models in the RR domain. According to (21) and (25), the RR vector $\boldsymbol{z}$, or its matrix form $\boldsymbol{Z}$, contains all the harmonic information. Further, $\boldsymbol{z}$ is covariance-based constructed from MMV, but it takes on an SMV form that is crucial for ensuing low-complexity algorithm design. It is the RR transformation matrix $\boldsymbol{\Psi}$ that linearly maps the

original $\boldsymbol{r}_x$ of a large size N^2M^2 to a much smaller vector $\boldsymbol{z}$ of size $(2N-1)(2M-1)$, without loss of any useful information. Further, the dimensionality is reduced as well. As indicated by (19), $\boldsymbol{r}_x$ in the original domain is structurally complex consisting of four nested Vandermonde vectors, which is difficult to tackle. In contract, the RR vector $\boldsymbol{z}$ is modeled to retain the two-level Vandermonde structure in its manifold vector $\boldsymbol{a}'_{2D}(\boldsymbol{f}^i)$ of (23), parameterized by the unknown frequencies of interest. More importantly, compared with (2), (21) suggests two enlarged virtual samplers of length $(2N-1)$ and $(2M-1)$ respectively. In this sense, the unknown frequency pairs can be retrieved through the decomposition of the estimated $\boldsymbol{z}$ or $\boldsymbol{Z}$, at enhanced frequency resolution. With (26), the MMV $\boldsymbol{Y}_L$ is capsulated by the transformed SMV $\hat{\boldsymbol{r}}_y = \text{vec}(\hat{\boldsymbol{R}}_y) = \text{vec}(\frac{1}{L}\boldsymbol{Y}_L\boldsymbol{Y}_L^H)$, which is linearly related to the RR vector $\boldsymbol{z}$. Now, the task boils down to reconstructing the 2D structure of $\boldsymbol{z}$ or $\boldsymbol{Z}$ in the RR domain from $\hat{\boldsymbol{r}}_y$.

B. RR-Based 2D MMV via Customized D-ANM

In the RR domain, the covariance-based $\boldsymbol{z}$ in (24) bears a similar form as that of $\boldsymbol{x}$ in (1), and hence can be mathematically treated as a single sample vector. An intricate difference is that the coefficients $\boldsymbol{r} = [r_1, \ldots, r_K]^T$ in $\boldsymbol{z}$ are nonnegative as indicated by (9), whereas $\boldsymbol{x}$ has complex-valued coefficients. Hence, all the sample-based gridless CS techniques for the simple SMV case, including the 2D V-ANM and D-ANM reviewed in Section II-B, can be applied, with extra care on the nonnegativeness of $\boldsymbol{r}$.

Considering the computational efficiency of D-ANM over V-ANM, we aim to develop a D-ANM solution to extract the structural information of the RR vector $\boldsymbol{z}$. To this end, we inspect its matrix form $\boldsymbol{Z}$ in (25):

$$\boldsymbol{Z} = \text{vec}^{-1}(\boldsymbol{z}) = \sum_{i=1}^{K} r_i \boldsymbol{a}'_M(f_{2,i})\boldsymbol{a}'^T_N(f_{1,i}) = \sum_{i=1}^{K} r_i \boldsymbol{A}'(\boldsymbol{f}^i)$$

where $\boldsymbol{A}'(\boldsymbol{f}^i) = \boldsymbol{a}'_M(f_{2,i})\boldsymbol{a}'^T_N(f_{1,i}) \in \mathbb{C}^{(2M-1)\times(2N-1)}$. It is easy to find that $\boldsymbol{Z}$ has a sparse linear atomic representation over the following matrix-form atom set of infinite size:

$$\mathcal{A}'_d = \left\{ \boldsymbol{A}'(\boldsymbol{f}), \quad \forall \boldsymbol{f} \in (-\tfrac{1}{2}, \tfrac{1}{2}]^2 \right\}. \tag{27}$$

Accordingly, we introduce a new matrix-form atomic norm

$$\|\boldsymbol{Z}\|^+_{\mathcal{A}'_d} = \inf\left\{ \sum_k r_k \,\middle|\, \sum_k r_k \boldsymbol{A}'(\boldsymbol{f}^k),\ \boldsymbol{A}'(\boldsymbol{f}^k) \in \mathcal{A}'_d; r_k \geq 0, \forall k \right\}. \tag{28}$$

Note that this norm of (28) differs from $\|\cdot\|_{\mathcal{A}_d}$ in (16) of the standard D-ANM [33], [34], because of the extra constraint $\boldsymbol{r} \geq \boldsymbol{0}$.

Given $\boldsymbol{Z}$, it is possible to retrieve the components $\{r_k, \boldsymbol{f}^k\}$ of its sparest representation by calculating its atomic norm, which leads to line spectrum estimation. In the presence of noise-inflicted SMV, it boils down to

$$\widetilde{\boldsymbol{Z}} = \arg\min_{\boldsymbol{Z}} \|\boldsymbol{Z}\|^+_{\mathcal{A}'_d} \text{ s.t. } \|\boldsymbol{r}_y - \boldsymbol{\Gamma}\text{vec}(\boldsymbol{Z})\|_2^2 \leq \beta, \tag{29}$$

where β indicates the noise tolerance threshold. This is a customized D-ANM defined on the norm $\|\cdot\|^+_{\mathcal{A}'_d}$, in lieu of $\|\cdot\|_{\mathcal{A}'_d}$ used in the D-ANM formula in (16). To illustrate the intricate difference, we rewrite (29) in an equivalent form:

$$\min_{\boldsymbol{Z}(\boldsymbol{r},\boldsymbol{f}_1,\boldsymbol{f}_2),\boldsymbol{r}} \|\boldsymbol{Z}\|_{\mathcal{A}'_d} = \|\boldsymbol{r}\|_1 \tag{30a}$$

$$\text{s.t.} \quad \|\boldsymbol{r}_y - \boldsymbol{\Gamma}\text{vec}(\boldsymbol{Z})\|_2^2 \leq \beta \tag{30b}$$

$$\boldsymbol{Z} = \boldsymbol{A}'_M(\boldsymbol{f}_2)\text{diag}(\boldsymbol{r})\boldsymbol{A}'^T_N(\boldsymbol{f}_1) \tag{30c}$$

$$\boldsymbol{r} \geq \boldsymbol{0}. \tag{30d}$$

Here (30a) and (30b) are in the same form as (16), (30c) is implicit in the objective function of both (16) and (29), but becomes an explicit constraint because of the new nonnegative constraint (30d). Without (30d), the SDP implementation of the D-ANM in (17) can be used to reformulate (30) into a convex problem. However, due to the extra constraint on $\boldsymbol{r}$ in (30d), this problem becomes intractable. This is because $\boldsymbol{r}$ is intertwined with the other variable $\boldsymbol{Z}$ in the form of (30c).

To solve (30) in a tractable manner, we seek to reformulate $\boldsymbol{r} \geq \boldsymbol{0}$ to an equivalent form with respect to $\boldsymbol{Z}$. To this end, we note from (24) that $\boldsymbol{Z} = \text{vec}^{-1}(\boldsymbol{z})$ in the RR domain is linearly related to $\boldsymbol{R}_x = \text{vec}^{-1}(\boldsymbol{r}_x)$, via

$$\boldsymbol{R}_x = \text{vec}^{-1}(\boldsymbol{r}_x) = \text{vec}^{-1}(\boldsymbol{\Psi}\boldsymbol{z}) = \text{vec}^{-1}\left(\boldsymbol{\Psi}\text{vec}(\boldsymbol{Z})\right). \tag{31}$$

When there exists a unique generalized Vandermonde decomposition of $\boldsymbol{R}_x$, it holds that $\boldsymbol{r} \geq \boldsymbol{0}$ if and only if $\boldsymbol{R}_x \succeq 0$. Fortunately, the desired decomposition property of $\boldsymbol{R}_x$ can be guaranteed, thanks to its connection to $\boldsymbol{Z}$ via (31) and the specific structure of $\boldsymbol{Z}$ in (30c). Specifically, according to D-ANM theory [34], the decomposition of $\boldsymbol{Z}$ in (30c) is unique when $K \leq 2\min\{N, M\} - 2$. Here, $\boldsymbol{R}_x$ represented in the form of (31) is a two-level Toeplitz matrix that is uniquely determined by $\boldsymbol{Z}$ through the proposed RR transformation as a one-to-one mapping between $\boldsymbol{R}_x$ and $\boldsymbol{Z}$. Thus, $\boldsymbol{R}_x$ has a unique generalized Vandermonde decomposition.

Moreover, $\boldsymbol{R}_x \succeq 0$ can be expressed by the following PSD constraint parameterized by $\boldsymbol{Z}$:

$$\text{vec}^{-1}\left(\boldsymbol{\Psi}\text{vec}(\boldsymbol{Z})\right) \succeq 0. \tag{32}$$

Replacing (30d) by (32), and reformulating (30a)-(30c) into a decoupled SDP form as in (17) [34], we reach the following equivalent SDP formula for (29) when $K \leq 2\min\{N, M\} - 2$:

$$\{\widetilde{\boldsymbol{Z}}, \widetilde{\boldsymbol{u}}_N, \widetilde{\boldsymbol{u}}_M\} = \arg\min_{\boldsymbol{Z},\boldsymbol{u}_N,\boldsymbol{u}_M} \frac{1}{2\sqrt{c}}\left(\text{Tr}(\text{T}(\boldsymbol{u}_N)) + \text{Tr}(\text{T}(\boldsymbol{u}_M))\right) \tag{33a}$$

$$\text{s.t.} \begin{bmatrix} \text{T}(\boldsymbol{u}_M) & \boldsymbol{Z} \\ \boldsymbol{Z}^H & \text{T}(\boldsymbol{u}_N) \end{bmatrix} \succeq 0, \tag{33b}$$

$$\left\| \text{vec}(\hat{\boldsymbol{R}}_y) - \boldsymbol{\Gamma}\text{vec}(\boldsymbol{Z}) \right\|_2^2 \leq \beta, \tag{33c}$$

$$\text{vec}^{-1}(\boldsymbol{\Psi}\text{vec}(\boldsymbol{Z})) \succeq 0. \tag{33d}$$

Eq. (33) is the SDP-based D-ANM solution for the MMV case in the RR domain, which we term as RR-D-ANM. It lumps all the measurements $\boldsymbol{Y}_L$ into a single vector $\hat{\boldsymbol{r}}_y$ in a covariance-based manner, and decouples the frequency-dependent variables into two 1D terms without loss of optimality by utilizing the specific structure of $\boldsymbol{Z}$ through D-ANM.

Remark 1: Note that the virtual manifold vectors $\boldsymbol{u}_N$ and $\boldsymbol{u}_M$ have enlarged sizes of $2N-1$ and $2M-1$, compared with the sizes N and M for V-ANM and D-ANM in Section II-B. As a result, (33) can operate under the condition $K \leq 2\min\{N,M\}-2$, which is twice looser than that in [31], [41]. Thus, when $\min\{N,M\}-1 < K \leq 2\min\{N,M\}-2$, the extra checking mechanisms applied in [31], [41] can be skipped if using (33). ■

It is worth noting that although the PSD matrix in (33b) is of size $2(N+M-1)\times 2(N+M-1)$, which is on the same order of size $(N+M)\times(N+M)$ of D-ANM for the SMV case in (17), the extra PSD constraint in (33d) is still of size $MN \times MN$. As a result, the overall complexity order of the proposed RR-D-ANM remains to be comparable to that of the V-ANM and the LRSCR. Next, to reduce the computational complexity, an efficient relaxation of the proposed method is proposed to balance the computational cost and estimation performance.

C. Efficient Relaxation

Since the PSD constraint in (33d) contributes to most of the computational cost, an efficient relaxation of (33) is to drop (33d), which results in the following RR-D-ANM-Relaxation (RR-D-ANM-R)

$$\begin{aligned}&\{\boldsymbol{Z}^\star, \boldsymbol{u}_N^\star, \boldsymbol{u}_M^\star\}\\ &= \arg\min_{\boldsymbol{Z},\boldsymbol{u}_N,\boldsymbol{u}_M} \frac{1}{2\sqrt{c}}\left(\mathrm{Tr}(\mathrm{T}(\boldsymbol{u}_N)) + \mathrm{Tr}(\mathrm{T}(\boldsymbol{u}_M))\right) \qquad (34a)\\ &\quad \text{s.t.} \begin{bmatrix} \mathrm{T}(\boldsymbol{u}_M) & \boldsymbol{Z} \\ \boldsymbol{Z}^H & \mathrm{T}(\boldsymbol{u}_N)\end{bmatrix} \succeq 0, \qquad (34b)\\ &\quad \left\|\mathrm{vec}(\hat{\boldsymbol{R}}_y) - \boldsymbol{\Gamma}\mathrm{vec}(\boldsymbol{Z})\right\|_2^2 \leq \beta. \qquad (34c)\end{aligned}$$

Next, we will show that, under certain conditions, the solution $\boldsymbol{Z}^\star$ obtained through RR-D-ANM-R reaches the ground truth in the noise-free case, and asymptotically approaches the ground truth in the noisy case with high probability.

1) Noise-Free Case: We start from the ideal noise-free case in the absence of any additive noise or finite sample error. According to (26), we have

$$\mathrm{vec}(\hat{\boldsymbol{R}}_y) = \boldsymbol{\Gamma}\mathrm{vec}(\boldsymbol{Z}). \qquad (35)$$

By replacing the error tolerance inequality (34c) with the equality constraint in (35), we have the RR-D-ANM-R for the noise-free case. For this problem, we can state the following theorem providing the fundamental limits of RR-D-ANM-R on sample efficiency when $M = N$.

Theorem 1: Let $\boldsymbol{\Gamma} \in \mathbb{C}^{N'\times(2N-1)^2}$ with $2N-1 \geq 1024$ be a random matrix with rows $\boldsymbol{\gamma}_r^H$ chosen independently from a distribution obeying the isotropy and incoherence properties[2]

$$\mathbb{E}(\boldsymbol{\gamma}_r\boldsymbol{\gamma}_r^H) = \frac{1}{N'}\boldsymbol{I};\ \ \sup_{\boldsymbol{f}} |\langle \boldsymbol{a}'_{2D}(\boldsymbol{f}), \boldsymbol{\gamma}_r\rangle|^2 \leq \frac{(2N-1)^2}{N'}\zeta \quad (36)$$

for some fixed $\zeta \geq 1$. Assume that the signs $\mathrm{sign}(r_i)$ of the complex coefficients $r_i \in \mathbb{C}$ of the unknown $\mathrm{vec}(\boldsymbol{Z}) = \sum_{i=1}^K r_i \boldsymbol{a}'_{2D}(\boldsymbol{f}^i)$, defined as $\mathrm{sign}(r_i) = r_i/|r_i|$, are randomly located on the complex unit circle with its phase $\phi(r_i)$ being symmetrically distributed around 0 between $[-\pi,\pi]$. Also assume that all the frequency pairs $\boldsymbol{f}^i = [f_{1,i}, f_{2,i}]^T$ obey the minimum separation condition

$$\min_{i\neq j}|f_{1,i}-f_{1,j}| \geq \frac{3}{N}, \min_{i\neq j}|f_{2,i}-f_{2,j}| \geq \frac{3}{N}. \qquad (37)$$

Then, the RR-D-ANM-R[3] returns the true $\boldsymbol{Z}$ with a probability at least $1-\delta$, as long as

$$N' \geq C\zeta K \log^2((2N-1)/\delta), \qquad (38)$$

with C being a constant.

Proof: A sketch of proof is provided in Appendix B, which is similar to the proofs of Theorem 1 in [48] and Theorem 6.1 in [34]. □

Theorem 1 provides a design guideline of isotropic and incoherent subsampling matrices. Specifically, the measurement vector length N', which dictates the allowable compression, is lower bounded by a function of N and δ in (38). To construct $\boldsymbol{\Gamma}$ that satisfies (36), we note the definition of $\boldsymbol{\Gamma}$ in (26):

$$\boldsymbol{\Gamma} = (\boldsymbol{J}^* \otimes \boldsymbol{J})\boldsymbol{\Psi} \in \mathbb{C}^{M'^2\times(2N-1)(2M-1)}. \qquad (39)$$

Suppose that a random selection matrix is adopted for subsampling in (6) as $\boldsymbol{J} = \boldsymbol{I}_{\boldsymbol{\Omega}}$ (where $\boldsymbol{I}_{\boldsymbol{\Omega}}$ is generated by randomly keeping only M' rows of $\boldsymbol{I}_{NM}$ with the indices from $\boldsymbol{\Omega} \subset \{1,2,\ldots,NM\}$). Then, $\boldsymbol{\Gamma}$ is also a random selection matrix but with some repeated rows. Let $\boldsymbol{\Gamma}' \in \mathbb{C}^{N'\times(2N-1)(2M-1)}$ with $N' < M'^2$ be a submatrix of $\boldsymbol{\Gamma}$ by keeping only the N' distinct rows of $\boldsymbol{\Gamma}$. Then $\boldsymbol{\Gamma}'$ satisfies the isotropy and incoherence properties in (36). If N' is chosen to further satisfy (38), then we are able to build $\boldsymbol{\Gamma}$ in (39) through an appropriate choice of $\boldsymbol{J}$, which allows RR-D-ANM-R to return the ground truth $\boldsymbol{Z}$ with a probability at least $1-\delta$ in the noise-free case.

2) Noisy Case: We now analyze the performance bound for the estimate of $\boldsymbol{Z}$ under the finite sample effect and the additive white Gaussian noise in (6) satisfying $\mathrm{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$. The noise tolerance constraint in (34c) has to be imposed in RR-D-ANM-R, which can be rewritten equivalently as (c.f. our recent work in [49]):

$$\left\|\hat{\boldsymbol{R}}_y - \boldsymbol{J}\mathcal{S}(\boldsymbol{Z})\boldsymbol{J}^H\right\|_F^2 \leq \beta, \qquad (40)$$

[2]In [48], the authors claim that the assumption $2N-1 \geq 1024$ is made merely for convenience, and that it can be dropped at the cost of a slightly higher constant in the minimum-separation condition.

[3]In the original RR-D-ANM, the coefficients r_i are constrained to be positive real numbers as in (9), which is a special case of the symmetry assumption by viewing the distribution of $\phi(r_i)$ as a delta function centered at 0. Hence, Theorem 1 developed for the RR-D-ANM-R also applies to the RR-D-ANM as a special case, even though the bound (38) can be loose.

where $\mathcal{S}(\cdot)$ defines the mapping $\boldsymbol{R}_x = \mathcal{S}(\boldsymbol{Z})$. Adopting $\boldsymbol{J} = \boldsymbol{I}_{\boldsymbol{\Omega}}$, (40) can be further rewritten as

$$\left\| \hat{\boldsymbol{R}}_y - \mathcal{S}_\Omega(\boldsymbol{Z}) \right\|_F^2 \leq \beta, \tag{41}$$

where $\mathcal{S}_\Omega(\boldsymbol{Z}) = \boldsymbol{I}_{\boldsymbol{\Omega}} \mathcal{S}(\boldsymbol{Z}) \boldsymbol{I}_{\boldsymbol{\Omega}}$ denotes the submatrix sampled from $\mathcal{S}(\boldsymbol{Z})$ with the index set Ω.

To provide theoretical analysis on the error bound for the estimate of $\boldsymbol{Z}$, we reformulate the RR-D-ANM-R into a Lagrangian form as

$$\min_{\boldsymbol{Z}} \tau \|\boldsymbol{Z}\|_{\mathcal{A}'_d} + \frac{1}{2} \left\| \hat{\boldsymbol{R}}_y - \mathcal{S}_\Omega(\boldsymbol{Z}) - \hat{\sigma}^2 \boldsymbol{I}_\Omega \right\|_F^2, \tag{42}$$

where $\tau > 0$ is the regularization parameter and the noise estimate $\hat{\sigma}^2$ is obtained as the smallest eigenvalue[4] of $\hat{\boldsymbol{R}}_y$. Note that (42) is equivalent to the original RR-D-ANM-R with the constraint of (41) in the absence of $\hat{\sigma}^2 \boldsymbol{I}_\Omega$, since their solutions are the same for all possible choices of the regularization parameters [15]. The introduction of the noise estimate $\hat{\sigma}^2 \boldsymbol{I}_\Omega$ is to deduce that the solution of (42) can asymptotically approach the ground truth, as stated by the following theorem.

Theorem 2: Let $\boldsymbol{Z}^*$, σ^2 be the ground truth, and $\hat{\boldsymbol{Z}}$ be the optimal solution to (42), respectively. Set

$$\tau \geq 2(C_1 \mathrm{Tr}(\mathcal{S}_\Omega(\boldsymbol{Z}^*)) + C_2 \sigma^2) M' \sqrt{\frac{\ln L}{L}} \tag{43}$$

for some constant C_1 and C_2. If Ω is a complete sparse ruler such that the unobserved entries of $\mathcal{S}(\boldsymbol{Z})$ can be deduced from the observed ones, then with probability at least $1 - 5L^{-1}$, the solution $\hat{\boldsymbol{Z}}$ to (42) satisfies

$$\left\| \hat{\boldsymbol{Z}} - \boldsymbol{Z}^* \right\|_F \leq \tau \left(8K + \frac{1}{M'} \right). \tag{44}$$

Proof: The proof is provided in Appendix C. □

Note in Theorem 2 that the observation set $\boldsymbol{\Omega}$ is deterministic for a given $\boldsymbol{J} = \boldsymbol{I}_{\boldsymbol{\Omega}}$, and that $8K$ is much larger than $\frac{1}{M'}$. Our algorithm yields reliable estimate of the RR matrix $\boldsymbol{Z}$ as long as the number of measurement vectors L is on the order of $(\mathrm{Tr}(\mathcal{S}_\Omega(\boldsymbol{Z}^*)) + \sigma^2)^2 M'^2 K^2$. The mean squared error (MSE) in (44) diminishes as L increases, and as $L \to \infty$, the ground truth $\boldsymbol{Z}^*$ can be exactly recovered, which means the estimate $\hat{\boldsymbol{Z}}$ of the proposed RR-D-ANM-R is statistically consistent in L. Moreover, since the frequency pair estimates $\{\hat{\boldsymbol{f}}^i\}_i$ can be uniquely determined by the optimal solution of the proposed RR-D-ANM-R (which will be introduced in the next subsection), the estimates $\{\hat{\boldsymbol{f}}^i\}_i$ is also statistically consistent in L. In fact, simulation results will show that even with a relatively small number of MMV, the slight performance loss of the RR-D-ANM-R is acceptable considering its huge complexity reduction, which will be evaluated through complexity analysis and comparison in Section IV-B. Hence, the RR-D-ANM-R offers a nice tradeoff between accuracy and complexity.

Remark 2: There are guidelines on the choice of the noise tolerance threshold β. If the Gaussian noise $\boldsymbol{n}(t)$ in (5) is colored with known distribution $\mathrm{N}(\boldsymbol{0}, \mathrm{diag}(\boldsymbol{\sigma}^2))$, then β in (33c) and (34c) can be set as $\beta = \lambda \|\boldsymbol{\sigma}^4\|_1$, where the scalar λ is chosen such that the inequality constraint (33c) or (34c) holds with high probability [50], [51]. Further, if the number of MMV L is sufficient and $\boldsymbol{Y}$ in (6) is a full or partial observation of $\boldsymbol{X}$, i.e., $\boldsymbol{J}=\boldsymbol{I}_{NM}$ or $\boldsymbol{J}=\boldsymbol{I}_{\boldsymbol{\Omega}}$, then β can be set using the covariance matrix sparse representation method in Proposition 1 of [18]. ■

Remark 3: Although the separation and sampling conditions introduced for RR-D-ANM-R are more restrictive than those of RR-D-ANM, V-ANM and LRSCR, RR-D-ANM-R is the most computationally efficient. Moreover, when the separation and sampling conditions are too strict, the reweighted ANM techniques [51], [52] can be employed in RR-D-ANM-R to enhance the sparsity and resolution. Note in this context that the separation conditions for gridless CS methods are usually more restrictive than those of classical statistical subspace methods, while the former are more sample efficient than the latter. ■

D. Harmonic Retrieval

Given the estimate of $\boldsymbol{Z}$ by solving (33) or (34), the desired $\boldsymbol{R}_x$ can be sequentially estimated from (31). Then, the 2D harmonics can be retrieved from $\boldsymbol{R}_x$ using the Matrix Pencil and Pairing (MaPP) method [31], which is computationally expensive due to the high dimension of $\boldsymbol{R}_x$.

To overcome this problem, we note that both (33) and (34) yield the estimates of not only $\boldsymbol{Z}$, but also $\mathrm{T}(\boldsymbol{u}_M)$ and $\mathrm{T}(\boldsymbol{u}_N)$. These estimates form an augmented matrix in the form of

$$\boldsymbol{R}_A = \begin{bmatrix} \mathrm{T}(\boldsymbol{u}_M) & \boldsymbol{Z} \\ \boldsymbol{Z}^H & \mathrm{T}(\boldsymbol{u}_N) \end{bmatrix}. \tag{45}$$

Next, we develop three efficient harmonic retrieval methods based on $\mathrm{T}(\boldsymbol{u}_M)$ and $\mathrm{T}(\boldsymbol{u}_M)$, $\boldsymbol{Z}$, and $\boldsymbol{R}_A$, respectively.

1) Vandermonde Decomposition Based Harmonic Retrieval: Since $\mathrm{T}(\boldsymbol{u}_M)$ and $\mathrm{T}(\boldsymbol{u}_N)$ contain the unknown frequencies, the Vandermonde decomposition based methods such as MUSIC, ESPRIT [53]–[55] can be employed for 1D harmonic retrieval in each dimension. Subsequently, a pairing operation needs to be done to obtain the frequency pairs [33], [34].

2) $\mathbf{Z}$-Based Harmonic Retrieval: Based on (30c), $\boldsymbol{Z}$ can be regarded as a cross-correlation matrix collected from an array with $\boldsymbol{A}'_M$, $\boldsymbol{A}'^*_N$ and $\mathrm{diag}(\boldsymbol{r})$ being the two manifold matrices and signal correlation matrix respectively. Hence, conventional 2D DOA estimation algorithms based on cross-correlation [56] can be crafted with the proposed RR-D-ANM solutions for harmonic retrieval. In this paper, we adopt the joint-SVD algorithm [56] in simulations.

3) $\mathbf{R}_A$-Based Harmonic Retrieval: According to the decoupled atomic norm theory [33], [34], we have

$$\begin{aligned} \boldsymbol{R}_A &= \begin{bmatrix} \mathrm{T}(\boldsymbol{u}_M) & \boldsymbol{Z} \\ \boldsymbol{Z}^H & \mathrm{T}(\boldsymbol{u}_N) \end{bmatrix} \\ &= \begin{bmatrix} \sqrt{2N-1}\boldsymbol{A}'_M \\ \sqrt{2M-1}\boldsymbol{A}'^*_N \end{bmatrix} \frac{1}{\sqrt{c}} \boldsymbol{R}_s \begin{bmatrix} \sqrt{2N-1}\boldsymbol{A}'_M \\ \sqrt{2M-1}\boldsymbol{A}'^*_N \end{bmatrix}^H \\ &= \boldsymbol{H} \boldsymbol{R}'_s \boldsymbol{H}^H, \end{aligned} \tag{46}$$

[4]The noise power can be well estimated as the smallest eigenvalue of the observed sample covariance, especially when L is large [50].

where $\boldsymbol{H} = [\sqrt{2N-1}\boldsymbol{A}_M'^T, \sqrt{2M-1}\boldsymbol{A}_N'^H]^T$ and $\boldsymbol{R}_s' = \frac{1}{\sqrt{c}}\boldsymbol{R}_s$. Note that $\boldsymbol{R}_A$ in (46) can be regarded as a covariance matrix with $\boldsymbol{H}$ and $\boldsymbol{R}_s'$ being its manifold matrix and signal correlation matrix respectively. Then, some conventional covariance-based 2D DOA estimation algorithms such as 2D-ESPRIT for L-shaped arrays [8], can be incorporated into the proposed RR-D-ANM solutions.

Remark 4: All the reconstructed matrices in (33) and (34) are low rank, and their rank values are equal to the number of frequency pairs. In the three proposed harmonic retrieval methods, either an eigenvalue decomposition (EVD) or singular value decomposition (SVD) is taken on these low-rank matrices. Note that these harmonic retrieval methods are not directly applicable to the original data of interest. It is the main contribution of this work to construct these structured matrices from the original data by the proposed RR-D-ANM or RR-D-ANM-R. Further, from these low-rank matrices, we can determine the number of frequency pairs. This is a standard model order selection problem, for which some mature techniques are available, such as Akaike information criterion (AIC), minimum description length (MDL), etc. A simple strategy is to count the number of eigenvalues or singular values that are larger than a predefined threshold, while a suggested threshold is $0.05\lambda_{max}$ [50], where λ_{max} is the maximum value of the eigenvalues or the singular values. In this sense, the proposed methods can be carried out without prior knowledge of the number of frequency pairs. ■

Remark 5: Note that the three proposed harmonic retrieval methods have different behaviors in terms of estimation accuracy and computational complexity. The Vandermonde decomposition-based method is computationally efficient, but requires an extra pairing operation. The $\boldsymbol{R}_A$-based method is computational expensive relative to the other two. In contrast, the $\boldsymbol{Z}$-based method is not only computationally efficient, but also can achieve automatic pairing. ■

In all, the proposed RR-D-ANM-R is not only computationally efficient, but also amenable to broad applications.

IV. Discussions

This section discusses two issues: one is the choice of the error tolerance parameter and the other is the computational complexity analysis for the proposed techniques.

A. Alternative Error Tolerance Constraint for Sufficient MMV

When noise is present, the noise tolerance constraint in (33c) and (34c), parameterized by a threshold β, needs to be chosen properly for performance guarantee. Remark 2 provides guidelines when the noise statistics is known, and the measurements are collected either in full ($\boldsymbol{J}=\boldsymbol{I}_{NM}$) or through random selection ($\boldsymbol{J}=\boldsymbol{I}_{\boldsymbol{\Omega}}$). However, when $\boldsymbol{J}$ represents other forms of samplers and the noise statistics are unavailable or unreliable in practice, it becomes difficult to choose the value of β. To address this challenge, we next design an alternative error tolerance constraint to replace that of (33c) and (34c), where an alternative user-specified parameter for bounding the observation errors can be set for any form of $\boldsymbol{J}$ and without knowing the noise statistics.

Let us express the noise $\boldsymbol{n}(t)$ in the compressive measurement as $\boldsymbol{n}(t) = \boldsymbol{J}\boldsymbol{n}_0(t)$ for any $\boldsymbol{J}$, where $\boldsymbol{n}_0(t) \in \mathbb{C}^{NM\times 1}$ is the front-end noise prior to compression and can be modeled as circular Gaussian noise with distribution $\mathrm{N}(\boldsymbol{0}, \mathrm{diag}(\boldsymbol{\sigma}_0^2))$. This model subsumes the noncompression case with $\boldsymbol{J}=\boldsymbol{I}_{NM}$. Accordingly, we have $\boldsymbol{Y}_L = \boldsymbol{J}\boldsymbol{X}_{x,n}$ in (6) and $\boldsymbol{R}_y = \boldsymbol{J}\boldsymbol{R}_{x,n}\boldsymbol{J}^H$ in (7), where $\boldsymbol{X}_n = \boldsymbol{X}_L + \boldsymbol{N}_0$ and $\boldsymbol{R}_{x,n} = \boldsymbol{R}_x + \mathrm{diag}(\boldsymbol{\sigma}_0^2)$. Then, similar to (10), we have $\hat{\boldsymbol{R}}_{x,n} = \frac{1}{L}\boldsymbol{X}_n\boldsymbol{X}_n^H$ given L MMV. We seek to quantify the residual error $\hat{\boldsymbol{R}}_{x,n} - \boldsymbol{R}_{x,n}$ caused by both additive noise and finite samples. According to the covariance matrix fitting criterion [29], when the complex amplitude vector $\boldsymbol{s}(t)$ in (1) is a circular Gaussian random variable with zero mean, the vectorization form of this residual error follows an asymptotic normal (AsN) distribution:

$$\mathrm{vec}(\hat{\boldsymbol{R}}_{x,n} - \boldsymbol{R}_{x,n}) \backsim \mathrm{AsN}(\boldsymbol{0}, \frac{1}{L}\boldsymbol{R}_{x,n}^T \otimes \boldsymbol{R}_{x,n}). \tag{47}$$

Accordingly, $\mathrm{vec}(\hat{\boldsymbol{R}}_y - \boldsymbol{R}_y)$, with $\hat{\mathbf{R}}_y$ given in (10), obeys an AsN distribution as well, in the form of $\mathrm{vec}(\hat{\boldsymbol{R}}_y - \boldsymbol{R}_y) \backsim \mathrm{AsN}(\boldsymbol{0}, \frac{1}{L}\boldsymbol{R}_y^T \otimes \boldsymbol{R}_y)$. Per the definition of asymptotic Chi-square ($\mathrm{As}\chi^2$) distributions, we have

$$\left\| (\frac{1}{L}\hat{\boldsymbol{R}}_y^T \otimes \hat{\boldsymbol{R}}_y)^{-\frac{1}{2}} \mathrm{vec}(\hat{\boldsymbol{R}}_y - \boldsymbol{R}_y) \right\|_2^2 \backsim \mathrm{As}\chi^2(M'^2). \tag{48}$$

Thus, the following inequality holds with probability $1-p$:

$$\left\| (\frac{1}{L}\hat{\boldsymbol{R}}_y^T \otimes \hat{\boldsymbol{R}}_y)^{-\frac{1}{2}} \mathrm{vec}(\hat{\boldsymbol{R}}_y - \boldsymbol{R}_y) \right\|_2^2 \leq \eta_p\,, \tag{49}$$

where the user-specified parameter η_p can be uniquely determined from (48) by the degrees of freedom M'^2 and a prefixed allowable deviation probability $p \ll 1$, independent of noise variance $\boldsymbol{\sigma}_0^2$.

Adopting (49) as an alternative error tolerance constraint to replace (33c) and (34c), and replacing $\boldsymbol{R}_y$ in (49) by $\mathrm{vec}^{-1}(\boldsymbol{\Gamma}\mathrm{vec}(\boldsymbol{Z}) + (\boldsymbol{J}^* \otimes \boldsymbol{J})\mathrm{vec}(\mathrm{diag}(\boldsymbol{\sigma}_0^2)))$, the proposed RR-D-ANM and RR-D-ANM-R can be implemented without knowing the noise statistics. In fact, this alternative formulation allows to estimate the noise variance as a byproduct, at the cost of introducing another unknown variable $\boldsymbol{\sigma}_0^2$ into (33) and (34).

B. Computational Complexity

We now analyze the computational complexity of the proposed techniques, compared with other existing 2D gridless methods with MMV. In SDP-based solutions [40], the computational complexity is dictated by the number of decision variables and the sizes of PSD matrices. Specifically, if there are a total of n_1 variables and the size of the PSD matrix is $n_2 \times n_2$, then the SDP can be solved in $\mathcal{O}(n_1^2 n_2^{2.5})$ flops in the worst case [57]. Accordingly, the computational complexity for solving the SDP formulation (14) of the V-ANM is $\mathcal{O}((NM+L)^{4.5}L^2)$. In contrast, the computational complexity of solving our RR-D-ANM (33) is $\mathcal{O}((NM)^{4.5})$, which is much lower than that of (14) especially for a large number of snapshots L. Note that the complexity of our RR-D-ANM remains the same as that of the LRSCR of (18), because the largest PSD constraint of each formula has a comparable size, i.e., $\mathrm{T}_{2D}(\boldsymbol{u}_{2D}) \succeq 0$ in (18) and

TABLE I
COMPLEXITY ORDER OF 2D-MMV GRIDLESS CS METHODS

Solution	Complexity ($N \neq M$)	Complexity ($N=M$)
V-ANM (14)	$\mathcal{O}((NM+L)^{4.5}L^2)$	$\mathcal{O}(N^9L^{6.5})$
LRSCR (18)	$\mathcal{O}((NM)^{4.5})$	$\mathcal{O}(N^9)$
RR-D-ANM (33)	$\mathcal{O}((NM)^{4.5})$	$\mathcal{O}(N^9)$
RR-D-ANM-R (34)	$\mathcal{O}((NM)^2(N+M)^{2.5})$	$\mathcal{O}(N^{6.5})$

$\text{vec}^{-1}(\boldsymbol{\Psi}\text{vec}(\boldsymbol{Z})) \succeq 0$ in (33). Nevertheless, the proposed RR-D-ANM can adopt more efficient frequency-retrieval options such as the $\boldsymbol{Z}$-based method developed in Section III-D, whereas the LRSCR has to be followed by MaPP applied on the large-size $\boldsymbol{R}_x$ [31].

Further, the LRSCR in (18) critically replies on the single PSD constraint $\text{T}_{2D}(\boldsymbol{u}_{2D}) \succeq 0$ to guarantee the desired structure of its solution, and hence cannot remove this constraint that dominates the computational complexity. In contrast, the D-ANM component already enforces the strong structure of the covariance matrix, so that the larger-size PSD constraint (33d) can be omitted to balance computational complexity with performance, leading to the RR-D-ANM-R in (34). With the remaining smaller-size PSD constraint (34b), the computational complexity of the RR-D-ANM-R drops down to $\mathcal{O}((NM)^2(N+M)^{2.5})$, which is 2.5 orders lower than that of the RR-D-ANM and LRSCR, when $N=M$. A comprehensive comparison of the computational complexity among different 2D gridless CS solutions with MMV is summarized in Table I.

Moreover, all the aforementioned methods can be implemented via fast algorithms such as the alternating direction method of multipliers (ADMM) [58] or the iterative Vandermonde decomposition and shrinkage thresholding (IVDST) [59] techniques, to further reduce the computational complexity in algorithm implementation.

V. SIMULATIONS

This section presents numerical results of the proposed gridless 2D harmonic retrieval techniques in MMV scenarios. Existing methods such as conventional MUSIC [60] and the LRSCR of (18) [42] incorporated with the MaPP algorithm for frequency estimation and pairing [31], as well as the Cramer-Rao bound (CRB) under full observations [61] are also simulated as benchmarks. In the conventional MUSIC, the peak searching step is set to be 0.01. In contrast, both the $\boldsymbol{Z}$-based algorithm designed in Section III-D and the MaPP algorithm can be employed for auto-paring the harmonic retrieval for the proposed RR-D-ANM and its relaxation version RR-D-ANM-R. Note that the sample-based ANM has poorer performance than the covariance-based LRSCR, while the LRSCR is more computation-efficient in MMV scenarios since its complexity is independent of the number of MMV [27]. Hence, we only run and compare the LRSCR as the benchmark in simulations and omit the computation-heavy V-ANM. Unless specifically stated, in simulations, we consider the scenarios where the number of MMV is sufficient, the separation and sampling conditions are satisfied for all gridless methods and the number of source signals is known *a priori*. With sufficient MMV, all optimization-based methods apply the alternative error tolerance constraint presented in Section IV-A, thanks to its applicability in the absence of known noise statistics.

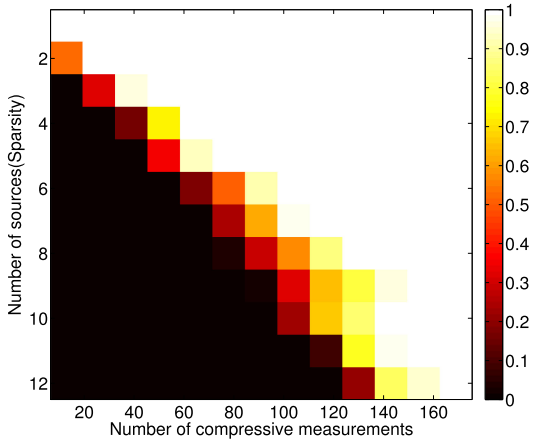


Fig. 1. Phase transition of the proposed RR-D-ANM-R with respect to the number of compressive measurements and the number of sources when $M = N = 7$.

In each trial of the Monte Carlo simulations, the MMV data $\boldsymbol{Y}_L$ is obtained through (6), where the measurement matrix $\boldsymbol{J}$ is a random matrix, the frequencies $\boldsymbol{f}$ are randomly generated under the separation condition, the source signals $\boldsymbol{s}(t)$ are assumed to have same amplitude, and the noise $\boldsymbol{n}(t) = \boldsymbol{J}\boldsymbol{n}_0(t)$ is circular Gaussian noise with $\boldsymbol{n}_0(t)$ satisfying $\text{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_{NM})$. Then, the sample covariance is calculated through (10) as the input to the covariance-based solutions, including MUSIC, LRSCR, RR-D-ANM, and RR-D-ANM-R. The root mean squared error (RMSE) measures the estimation accuracy of 2D harmonic retrieval as $\text{RMSE} = \frac{1}{K}\sum_{k=1}^{K}(\frac{1}{M_t}\sum_{n=1}^{M_t}((\widetilde{f}_{1,k}^n - f_{1,k}^n)^2 + (\widetilde{f}_{2,k}^n - f_{2,k}^n)^2))^{\frac{1}{2}}$, where M_t, $\widetilde{f}_{1,k}^n$ and $\widetilde{f}_{2,k}^n$ denote the number of Monte-Carlo trials, and the estimates of $f_{1,k}^n$ and $f_{2,k}^n$ in the n-th experiment.

A. Efficient Relaxation

First, we evaluate the effectiveness of the proposed RR-D-ANM-R as an efficient relaxation of the original RR-D-ANM. For the noise-free case, we examine the phase transition of the proposed RR-D-ANM-R with respect to the number of compressive measurements and the number of sources. In the simulations, we apply $\boldsymbol{\Gamma}$ in (39) as a random selection matrix and randomly generate the 2D frequencies. For each Monte Carlo trial, it is considered to be a successful recovery of $\boldsymbol{Z}$ from (34), if the normalized mean squared error (NMSE) of the solution $\boldsymbol{Z}^\star$ has $\|\boldsymbol{Z}^\star - \boldsymbol{Z}^*\|_2/\|\boldsymbol{Z}^*\|_2 \leq 10^{-5}$. As shown in Fig. 1, RR-D-ANM-R works well and retrieves the ground truth $\boldsymbol{Z}^*$ with high probability under the satisfied separation and sampling conditions, where the phase transition border corresponds to the fundamental limits of RR-D-ANM-R on sample efficiency revealed by Theorem 1. For the noisy case, Fig. 2 shows that the proposed RR-D-ANM outperforms its relaxation RR-D-ANM-R, which indicates the role of the PSD constraint of the covariance in enhancing the estimation performance. Further,

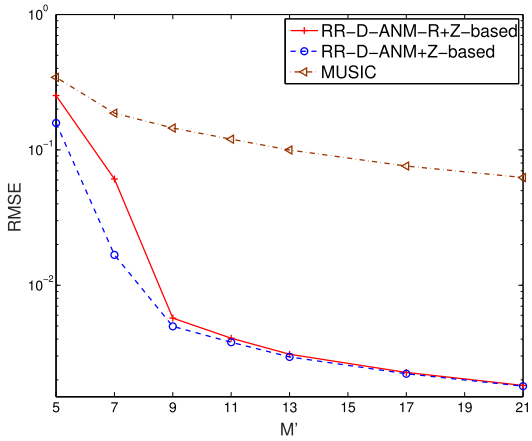


Fig. 2. RMSE vs M' for MUSIC, RR-D-ANM and RR-D-ANM-R, when $M{=}N{=}5$, $L{=}200$, $K{=}3$ and SNR$=$5 dB.

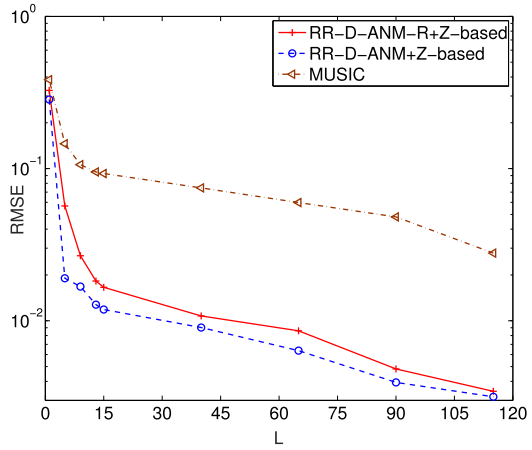


Fig. 3. RMSE vs (insufficient) MMV for MUSIC, RR-D-ANM and RR-D-ANM-R, when $M{=}N{=}5$, $M'{=}15$, $K{=}3$ and SNR$=$10 dB.

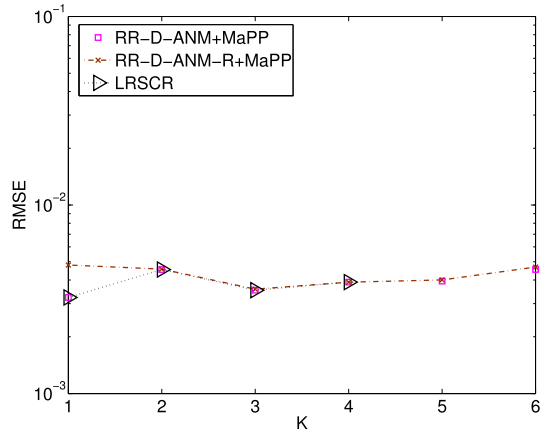


Fig. 4. RMSE vs K for RR-D-ANM, RR-D-ANM-R and LRSCR, when $M{=}N{=}5$, $M'{=}15$, SNR$=$5 dB, $L{=}200$.

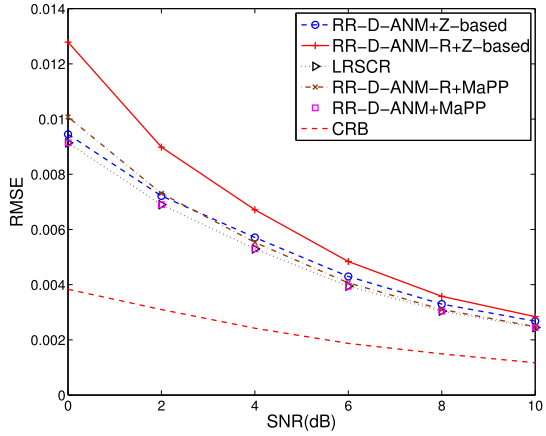


Fig. 5. RMSE vs SNR for RR-D-ANM, RR-D-ANM-R and LRSCR with $M{=}N{=}5$, $M'{=}15$, $K{=}3$ and $L{=}200$.

the performance gap between RR-D-ANM and RR-D-ANM-R diminishes as the number of measurements M' goes large, which reveals a desired tradeoff between accuracy and complexity that can be achieved by RR-D-ANM-R. Classical MUSIC is also tested and always provides the worst estimation performance, which reflects the limitation of classical MUSIC in sample efficiency when $M'{<}NM$. Moreover, Fig. 3 presents the RMSE of these approaches for different numbers of insufficient MMV. In this simulation, instead of using the alternative error tolerance constraint proposed for sufficient MMV, we consider that the noise statistics are known *a priori*, and then the Lagrangian formulations of the proposed RR-D-ANM and RR-D-ANM-R with the regularization parameters set according to Theorem 2 are employed to generate the curves in Fig. 3. It shows similar trends as in Fig. 2 and RR-D-ANM-R can obtain a comparable performance (the gap is less than 0.01) to RR-D-ANM when $L \geq 9$, which means the RR-D-ANM-R can be a good alternative even for small L.

B. Estimation Accuracy

In this subsection, the estimation performance of the proposed methods is evaluated. We first present the RMSE performance of the aforementioned methods with varying number of harmonics. For the LRSCR, when the conservative condition on the number of sources is unsatisfied, i.e., $K > \min\{N, M\} - 1$, the extra checking mechanism is used as proposed in [31]. In Fig. 4, for the curve of LRSCR, after K goes larger than 4, it no longer plots reasonable results of RMSE, given the fact that it cannot be computed sometimes. This is because the checking mechanism outputs the "false" status, which indicates it judges that the generalized Vandermonde decomposition of the estimated covariance matrix from the LRSCR does not exist. Meanwhile, the RR-D-ANM and RR-D-ANM-R both work properly without using an extra checking mechanism as long as $K \leq 2\min\{N, M\} - 2$.

Next, we focus on the case when the conservative condition on the number of sources of the LRSCR is satisfied, i.e., $K \leq \min\{N, M\} - 1$. As shown in Fig. 5, again, the RR-D-ANM outperforms the RR-D-ANM-R, and the gap becomes small as

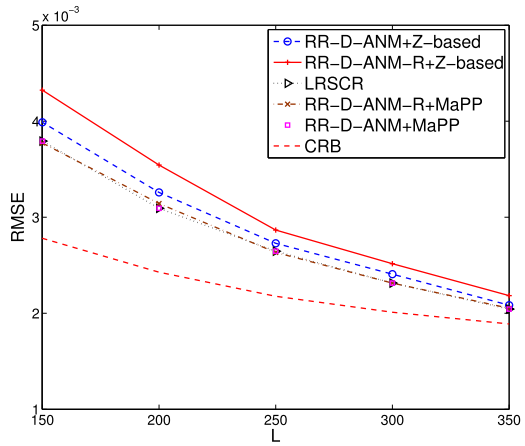


Fig. 6. RMSE vs (sufficient) MMV for RR-D-ANM, RR-D-ANM-R and LRSCR with $M = N = 5$, $M' = 15$, $K = 3$ and SNR $= 10$dB.

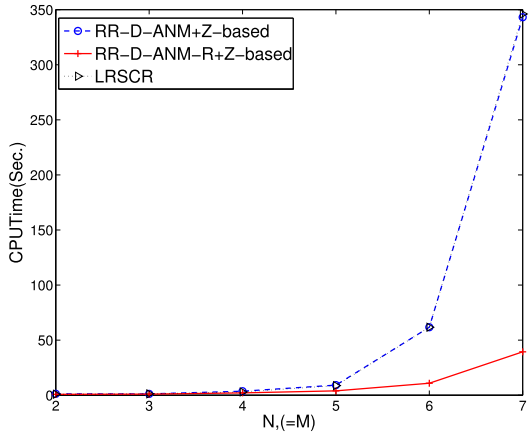


Fig. 7. Computational complexity of RR-D-ANM, RR-D-ANM-R and LRSCR in terms of runtime with L=200, SNR=10dB and M'=$\lfloor \frac{NM}{2} \rfloor$+1.

the SNR increases. Moreover, the LRSCR (which always uses MaPP) obtains a better performance than the RR-D-ANM plus $\boldsymbol{Z}$-based method, but this gap disappears when RR-D-ANM plus MaPP is used, as shown by the overlapping square and triangle marks in Fig. 5. In other words, the LRSCR and the RR-D-ANM lead to the same solutions when the generalized Vandermonde decomposition of the solutions exists. On the other hand, the harmonic retrieval method via MaPP results in a higher computational cost than the $\boldsymbol{Z}$-based method especially as M and N go large. In addition, the gap between the RR-D-ANM-R plus MaPP and RR-D-ANM plus MaPP methods is much smaller than that between the RR-D-ANM-R plus $\boldsymbol{Z}$-based and RR-D-ANM plus $\boldsymbol{Z}$-based methods in Fig. 5. This reflects a tradeoff between the accuracy and complexity of specific 2D harmonic retrieval algorithms, and it can be used as a guideline for the choice among different 2D harmonic retrieval algorithms, thanks to the widely applicability of our proposed methods. In fact, the trivial gap between the two curves of RR-D-ANM-R and RR-D-ANM observed in both Figs. 2 and 3 will diminish when the MaPP method is employed for harmonic retrieval. The CRB under full observations, i.e., $\boldsymbol{J} = \boldsymbol{I}_{NM}$ is also provided as the optimal benchmark for the noncompression case. Note that the gaps between the CRB and the gridless CS methods become small as the signal-to-noise ratio (SNR) increases, which verifies the sample efficiency of the gridless CS methods. Fig. 6 presents the RMSE performance of these approaches for different numbers of sufficient MMV, which shows similar trends as in Fig. 5.

C. Computational Complexity

The computational complexity of the proposed methods and the benchmark methods is tested and compared in this subsection, where all methods use an off-the-shelf SDP-based solver [62]. As shown in Fig. 7, the complexity in terms of runtime[5] of the proposed RR-D-ANM is almost the same as that of the LRSCR. Meanwhile, the complexity of the RR-D-ANM-R stays much lower than the others. These results verify the analysis on complexity order of the different methods presented in Section IV-B. Thus, 1) RR-D-ANM is a nice alternative to the LRSCR when $\min\{N, M\} - 1 \leq K \leq 2\min\{N, M\} - 2$ with small N and M; 2) RR-D-ANM-R offers a good tradeoff between accuracy and complexity as N and M go large regardless of L.

[5]All simulations are conducted in Matlab 2013b on a computer with a 4-core Intel i7-6500 U 2.50 GHz CPU and 8 GB memory.

VI. Conclusion

In this paper, a gridless framework based on the D-ANM technique is proposed to efficiently achieve super-resolution 2D harmonic estimation with MMV. Given the sample covariance matrix collected from MMV, we first establish an RR transformation to linearly map the originally large-size covariance matrix to a small-size RR vector, whose sparse representation enables to reformulate the complex 2D MMV harmonic retrieval as an RR-based D-ANM problem that can be resolved efficiently. In addition, we analyze the computational complexity of the proposed techniques compared with other gridless methods. Simulation results indicate the advantages of our solutions over the existing ones with wider applicability at lower complexity.

Appendix A
Derivation of (20)

Each summand of (19) can be written as

$$\begin{aligned}
&(\boldsymbol{a}_N(f_{1,i}) \otimes \boldsymbol{a}_M(f_{2,i}))^* \otimes (\boldsymbol{a}_N(f_{1,i}) \otimes \boldsymbol{a}_M(f_{2,i})) \\
&= \boldsymbol{a}_N^*(f_{1,i}) \otimes (\boldsymbol{a}_M^*(f_{2,i}) \otimes \boldsymbol{a}_N(f_{1,i})) \otimes \boldsymbol{a}_M(f_{2,i}) \\
&= (\boldsymbol{I}_N \boldsymbol{a}_N^*(f_{1,i})) \otimes (\boldsymbol{E}\,(\boldsymbol{a}_N(f_{1,i}) \otimes \boldsymbol{a}_M^*(f_{2,i}))) \otimes \boldsymbol{a}_M(f_{2,i}) \\
&= (\boldsymbol{I}_N \otimes \boldsymbol{E} \otimes \boldsymbol{I}_M)((\boldsymbol{G}_N \boldsymbol{a}'_N(f_{1,i})) \otimes (\boldsymbol{G}_M \boldsymbol{a}'_M(f_{2,i}))) \\
&= (\boldsymbol{I}_N \otimes \boldsymbol{E} \otimes \boldsymbol{I}_M)(\boldsymbol{G}_N \otimes \boldsymbol{G}_M)(\boldsymbol{a}'_N(f_{1,i}) \otimes \boldsymbol{a}'_M(f_{2,i})) \\
&= \boldsymbol{\Psi}(\boldsymbol{a}'_N(f_{1,i}) \otimes \boldsymbol{a}'_M(f_{2,i})), \qquad (50)
\end{aligned}$$

where the second equation applies the property $\boldsymbol{a} \otimes \boldsymbol{b} = \boldsymbol{K}(\boldsymbol{b} \otimes \boldsymbol{a}), \forall \boldsymbol{a} \in \mathbb{C}^{N\times 1}, \boldsymbol{b} \in \mathbb{C}^{M\times 1}$, and the third equation applies $\boldsymbol{a}^*(f) \otimes \boldsymbol{a}(f) = \boldsymbol{G}\boldsymbol{a}'(f)$.

APPENDIX B
PROOF SKETCH OF THEOREM 1

The proof of Theorem 1 is similar to the proof of Theorem 1 in [48]. The road map of this proof is to guarantee that the dual polynomial of the compression case is close to that of the noncompression case. We first apply the principle of dual polynomial constructions and then prove the existence of their dual certificates. Specifically, we will first characterize properties of a dual polynomial that suffices to certify the optimality and uniqueness of the solutions to the noncompression case and the compression case, and then present a dual construction scheme for these two cases. For the compression case, the dual construction scheme produces a polynomial based on the dual polynomial constructed for the noncompression case. Finally, we provide in Lemma 1 the sufficient conditions on the compressed measurement matrix $\boldsymbol{\Gamma}$ in (26) for guaranteeing the existence of the dual certificate, which concludes the proof of Theorem 1.

A. Optimality Conditions for Dual Polynomial

For the noncompression case, the primal problem of D-ANM can be expressed as

$$\min_{\widetilde{\boldsymbol{Z}}} \left\| \widetilde{\boldsymbol{Z}} \right\|_{\mathcal{A}'_d} s.t.\ \widetilde{\boldsymbol{Z}} = \boldsymbol{Z}, \tag{51}$$

where $\boldsymbol{Z} = \boldsymbol{A}'_N \boldsymbol{R}_s \boldsymbol{A}'^T_N \in \mathbb{C}^{J\times J}$ with $J = 2N-1$. Define the dual norm of $\|\cdot\|_{\mathcal{A}'_d}$ as

$$\left\|\overline{\boldsymbol{Q}}\right\|^*_{\mathcal{A}'_d} = \sup_{\|\boldsymbol{Z}\|_{\mathcal{A}'_d}\le 1} \left\langle \overline{\boldsymbol{Q}}, \boldsymbol{Z} \right\rangle_{\mathcal{R}} = \sup_{\boldsymbol{A}'(\boldsymbol{f})\in\mathcal{A}'_d} \langle \boldsymbol{A}'(\boldsymbol{f}), \boldsymbol{Z}\rangle_{\mathcal{R}}, \tag{52}$$

where the real inner product is defined as $\left\langle \overline{\boldsymbol{Q}}, \boldsymbol{Z} \right\rangle_{\mathcal{R}} = \text{Real}(\left\langle \overline{\boldsymbol{Q}}, \boldsymbol{Z} \right\rangle)$ with the inner product being defined as $\left\langle \overline{\boldsymbol{Q}}, \boldsymbol{Z} \right\rangle = \text{Tr}(\boldsymbol{Z}^H \overline{\boldsymbol{Q}}) = \text{vec}^H(\boldsymbol{Z})\text{vec}(\overline{\boldsymbol{Q}})$. Then, following the standard Lagrangian procedure as in [22], we can obtain the dual problem of D-ANM as

$$\max_{\overline{\boldsymbol{Q}}} \left\langle \overline{\boldsymbol{Q}}, \boldsymbol{Z} \right\rangle_{\mathcal{R}} s.t. \left\|\overline{\boldsymbol{Q}}\right\|^*_{\mathcal{A}'_d} \le 1. \tag{53}$$

For the noncompression case, the optimization problem (51) has a trivial solution, but we can still apply duality to certify the optimality of a particular decomposition. Following standard analysis (see [23], [24]), the optimal solution of (51) is unique if there exists a dual polynomial in the form of

$$\overline{Q}(\boldsymbol{f}) = \left\langle \overline{\boldsymbol{Q}}, \boldsymbol{A}'(\boldsymbol{f}) \right\rangle = \boldsymbol{a}'^H_{2D}(\boldsymbol{f})\overline{\boldsymbol{q}}, \tag{54}$$

with complex coefficients $\overline{\boldsymbol{q}} = \text{vec}(\overline{\boldsymbol{Q}}) \in \mathbb{C}^{J^2}$ satisfying the following set of conditions

$$\begin{aligned} &\overline{Q}(\boldsymbol{f}^i) = \text{sign}(r_i), \forall \boldsymbol{f}^i \in \mathcal{S} \\ &|\overline{Q}(\boldsymbol{f})| < 1, \forall \boldsymbol{f} \notin \mathcal{S}, \end{aligned} \tag{55}$$

where $\mathcal{S} = \{\boldsymbol{f}^1, \ldots, \boldsymbol{f}^K\}$ denotes the unknown set of true frequencies. Moreover, we will construct a dual polynomial satisfying above conditions in the next subsection.

Next, we consider the compression case which can be written in the form

$$\min_{\boldsymbol{Z}} \|\boldsymbol{Z}\|_{\mathcal{A}'_d} s.t. \text{vec}(\boldsymbol{R}_y) = \boldsymbol{\Gamma}\text{vec}(\boldsymbol{Z}). \tag{56}$$

Then, the dual problem of (56) can be achieved via the standard lagrangian procedure [22] as

$$\max_{\boldsymbol{Q}} \langle \boldsymbol{Q}, \boldsymbol{R}_y \rangle_{\mathcal{R}} s.t. \left\| \text{vec}^{-1}(\boldsymbol{\Gamma}^H \text{vec}(\boldsymbol{Q})) \right\|^*_{\mathcal{A}'_d} \le 1. \tag{57}$$

Moreover, following a similar analysis as in [23], [24], we have that $\boldsymbol{Z}$ is the unique optimal solution of (56) if there exists a dual polynomial in the form of

$$\begin{aligned} Q(\boldsymbol{f}) &= \left\langle \text{vec}^{-1}(\boldsymbol{\Gamma}^H \text{vec}(\boldsymbol{Q})), \boldsymbol{A}'(\boldsymbol{f}) \right\rangle \\ &= \left\langle \boldsymbol{\Gamma}^H \boldsymbol{q}, \boldsymbol{a}'_{2D}(\boldsymbol{f}) \right\rangle = \boldsymbol{a}'^H_{2D}(\boldsymbol{f}) \boldsymbol{\Gamma}^H \boldsymbol{q}, \end{aligned} \tag{58}$$

with complex coefficients $\boldsymbol{q} = \text{vec}(\boldsymbol{Q}) \in \mathbb{C}^{N'}$ satisfying the following set of conditions

$$\begin{aligned} &Q(\boldsymbol{f}^i) = \text{sign}(r_i), \forall \boldsymbol{f}^i \in \mathcal{S} \\ &|Q(\boldsymbol{f})| < 1, \forall \boldsymbol{f} \notin \mathcal{S}. \end{aligned} \tag{59}$$

Thus, to prove Theorem 1, we now need to show that such a dual certificate exists under appropriate conditions and this is the subject of the next subsection.

B. Construction of Dual Polynomial

To construct such a dual polynomial satisfying (55), note that $\overline{Q}(\boldsymbol{f})$ is required to be a 2D-trigonometric polynomial, and based on the analysis in [30], [48], the dual polynomial of (54) takes the form

$$\begin{aligned} \overline{Q}(\boldsymbol{f}) = &\sum_{i=1}^K \overline{\alpha}_i \overline{G}(\boldsymbol{f} - \boldsymbol{f}^i) + \sum_{i=1}^K \overline{\beta}_{1i} \overline{G}^{(10)}(\boldsymbol{f} - \boldsymbol{f}^i) \\ &+ \sum_{i=1}^K \overline{\beta}_{2i} \overline{G}^{(01)}(\boldsymbol{f} - \boldsymbol{f}^i), \end{aligned} \tag{60}$$

where $\overline{\alpha}_i, \overline{\beta}_{1i}, \overline{\beta}_{2i}$ are coefficients chosen such that $\overline{Q}(\boldsymbol{f})$ interpolates the points $\text{sign}(r_i)$ at the $\boldsymbol{f}_i$. $\overline{G}(\boldsymbol{f})$ is the 2D Fejér's kernel defined as

$$\overline{G}(\boldsymbol{f}) = K(f_1)K(f_2), \tag{61}$$

where $K(f)$ is the squared Fejér kernel defined as

$$K(f) := \left(\frac{\sin(R\pi f)}{R\sin(\pi f)}\right)^4 = \frac{1}{R}\sum_{k=-P}^{P} g_k e^{j2\pi k f} \tag{62}$$

with $R := \frac{P}{2} + 1$, $J = 2P+1$, and coefficients

$$g_k = \frac{1}{R} \sum_{s=\max(k-R,-R)}^{\min(k+R,R)} \left(1 - \left|\frac{s}{R}\right|\right)\left(1 - \left|\frac{k-s}{R}\right|\right). \tag{63}$$

$\overline{G}^{(i_1 i_2)}(\boldsymbol{f})$ is the partial derivative of $\overline{G}(\boldsymbol{f})$ given by $\overline{G}^{(i_1 i_2)}(\boldsymbol{f}) = \frac{\partial^{i_1}\partial^{i_2}\overline{G}(\boldsymbol{f})}{\partial f_1^{i_1} f_2^{i_2}}$.

Inspired by the construction of $\overline{Q}$, we now design a dual polynomial for the compression case. Specifically, following the analysis in [48], we propose to construct the dual polynomial of (58) as

$$Q(\boldsymbol{f}) = \sum_{i=1}^{K} \alpha_i G_{(00)}(\boldsymbol{f}, \boldsymbol{f}^i) + \sum_{i=1}^{K} \beta_{1i} G_{(10)}(\boldsymbol{f}, \boldsymbol{f}^i) + \sum_{i=1}^{K} \beta_{2i} G_{(01)}(\boldsymbol{f}, \boldsymbol{f}^i), \tag{64}$$

where $G_{\boldsymbol{m}}(\boldsymbol{f}, \boldsymbol{f}^i) = \left\langle \boldsymbol{\Gamma}^H \boldsymbol{\Gamma} \boldsymbol{g}_{\boldsymbol{m}}(\boldsymbol{f}^i), \boldsymbol{a}'_{2D}(\boldsymbol{f}) \right\rangle$ with $\boldsymbol{m} = (m_1\, m_2) \in \{0,1\}^2$, $\boldsymbol{g}(\boldsymbol{f}) = \boldsymbol{g}((f_1, f_2)) = \boldsymbol{g}'(f_1) \otimes \boldsymbol{g}'(f_2)$ with $\boldsymbol{g}'(f) = \frac{1}{R}[g_{-P} e^{j2\pi(-P)f}, \ldots, g_P e^{j2\pi P f}]^T$ and $\boldsymbol{g}_{\boldsymbol{m}}(\boldsymbol{f}) = \boldsymbol{g}_{(m_1 m_2)}(f_1, f_2) = \boldsymbol{g}'_{m_1}(f_1) \otimes \boldsymbol{g}'_{m_2}(f_2)$ being the partial derivative of $\boldsymbol{g}(\boldsymbol{f})$.

Now, with the constructed dual polynomials in (60) for the noncompression case and in (64) for the compression case, we are ready to state the following lemma to conclude the proof of Theorem 1.

Lemma 1: Let $\mathcal{S} = \{\boldsymbol{f}^1, \ldots, \boldsymbol{f}^K\}$ be an arbitrary set of points obeying the minimum separation condition (37). Let $\boldsymbol{u} = [\text{sign}(r_1), \ldots, \text{sign}(r_K)]^T$ be a random vector, whose entries are chosen independently from symmetric distributions on the complex unit circle. Pick $\delta > 0$, and let $\boldsymbol{\Gamma}$ be a random matrix that obeys the concentration inequality

$$\mathbb{P}\left[\frac{\left| \left\langle (\boldsymbol{\Gamma}^H \boldsymbol{\Gamma} - \boldsymbol{I}) \boldsymbol{g}_{\boldsymbol{m}}(\boldsymbol{f}^i), \boldsymbol{a}'^{\boldsymbol{n}}_{2D}(\boldsymbol{f}) \right\rangle \right|}{\mathcal{K}^{\|\boldsymbol{m}+\boldsymbol{n}\|_1}} \geq \frac{c_1}{\sqrt{K \log(\frac{L}{\delta})}} \right] \leq c_3 \frac{\delta}{L^{15}}, \tag{65}$$

for all $\boldsymbol{f}^i, \boldsymbol{f} \in [-0.5, 0.5)^2$, all $\boldsymbol{m}, \boldsymbol{n} \in \{0,1\}^2$ with $\|\boldsymbol{m}\|_1, \|\boldsymbol{n}\|_1 \leq 2$.

Here, $\boldsymbol{a}'^{\boldsymbol{n}}_{2D}(\boldsymbol{f})$ represents the partial derivative of $\boldsymbol{a}'_{2D}(\boldsymbol{f})$ and $\mathcal{K} = \sqrt{|K^{(2)}(0)|} = \sqrt{\frac{4\pi^2(R^2-1)}{3}}$. Moreover, suppose that for $\boldsymbol{f}^i \in [-0.5, 0.5)^2$, there is a constant $\hat{c}$ such that

$$\mathbb{P}\left[\max_{\boldsymbol{f} \in [-0.5, 0.5)^2} \left| \left\langle \boldsymbol{\Gamma} \boldsymbol{g}_{(00)}(\boldsymbol{f}^i), \boldsymbol{\Gamma} \boldsymbol{a}'_{2D}(\boldsymbol{f}) \right\rangle \right| \geq \frac{\hat{c}}{K} L^2 \right] \leq \frac{\delta}{8}. \tag{66}$$

Then, with probability at least $1 - \delta$, there exists a trigonometric polynomial in (58), with complex coefficients $\boldsymbol{q}$ obeying the conditions in (59).

Proof: The proof of Lemma 1 follows similar steps as proposed in the proof of Lemma 1 in [48]. We omit the details here and only list the key steps as follows:

1) We show that under the minimum separation condition in (37), there exists a choice of coefficients $\overline{\alpha}_i, \overline{\beta}_{1i}$ and $\overline{\beta}_{2i}$ such that $\overline{Q}(\boldsymbol{f})$ satisfies the conditions in (55).
2) We show that, with probability at least $1 - \frac{\delta}{2}$, there exists a choice of coefficients α_i, β_{1i} and β_{2i} such that $Q(\boldsymbol{f}^i) = \text{sign}(r_i)$ and the derivative w.r.t. $\boldsymbol{f}^i$ $\nabla Q(\boldsymbol{f}^i) = \boldsymbol{0}$ for all $\boldsymbol{f}^i \in \mathcal{S}$ under the minimum separation condition in (37).
3) Finally, we conclude the proof by showing that with the chosen coefficients with probability at least $1 - \frac{\delta}{2}$, $Q(\boldsymbol{f})$ is close to $\overline{Q}(\boldsymbol{f})$ for all $\boldsymbol{f} \in [-0.5, 0.5)^2 \backslash \mathcal{S}$, which in turn implies the polynomial $Q(\boldsymbol{f})$ obeys $|Q(\boldsymbol{f})| < 1$ uniformly for all $\boldsymbol{f} \in [-0.5, 0.5)^2 \backslash \mathcal{S}$ with probability at least $1 - \frac{\delta}{2}$.

According to the union bound, the two events in the above 2) and 3) happen simultaneously with a probability at least $1 - \delta$. That is, the dual certificate in (59) holds with a probability at least $1 - \delta$, which concludes the proof of Lemma 1. □

Moreover, following the analysis in Section IV-B in [48], we can also show that under the fundamental limits on sample efficiency in (38), the isotropy and incoherence conditions in (36) imply the conditions in (65) and (66) of Lemma 1, which in turn implies the existence of an appropriate dual certificate satisfying the conditions in (59) and then concludes the proof of Theorem 1.

Appendix C
Proof of Theorem 2

To prove Theorem 2, we introduce the following lemma.

Lemma 2: Suppose that $\boldsymbol{z}(t)$ is a Gaussian random process with zero mean and covariance matrix $\boldsymbol{\Sigma}$. Let $\boldsymbol{\Sigma}_L = 1/L \sum_{t=1}^{L} \boldsymbol{z}(l) \boldsymbol{z}^H(l)$ be the sampled covariance matrix and λ_k and $\hat{\lambda}_k$ be the eigenvalue of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_L$, respectively. Then with probability at least $1 - 5L^{-1}$, we have

$$\begin{aligned} \|\boldsymbol{\Sigma}_L - \boldsymbol{\Sigma}\|_F &\leq 2C \text{Tr}(\boldsymbol{\Sigma}) \sqrt{\frac{\ln L}{L}}, \\ |\hat{\lambda}_k - \lambda_k| &\leq \text{Tr}(\boldsymbol{\Sigma}) \sqrt{\frac{\ln L}{L}}, \end{aligned} \tag{67}$$

for all k and some constant C.

Proof: The proof of Lemma 2 can be found in [63]. □

Then, we can state that

$$\begin{aligned} &\left\| \hat{\boldsymbol{R}}_y - \mathcal{S}_\Omega(\boldsymbol{Z}^*) - \hat{\sigma}^2 \boldsymbol{I}_{\boldsymbol{\Omega}} \right\|_F = \left\| \hat{\boldsymbol{R}}_L - \mathcal{S}_\Omega(\boldsymbol{Z}^*) \right\|_F \\ &= \left\| \hat{\boldsymbol{R}}_y - \mathcal{S}_\Omega(\boldsymbol{Z}^*) - \sigma^2 \boldsymbol{I}_{\boldsymbol{\Omega}} + \sigma^2 \boldsymbol{I}_{\boldsymbol{\Omega}} - \hat{\sigma}^2 \boldsymbol{I}_{\boldsymbol{\Omega}} \right\|_F \\ &\leq \left\| \hat{\boldsymbol{R}}_y - \mathcal{S}_\Omega(\boldsymbol{Z}^*) - \sigma^2 \boldsymbol{I}_{\boldsymbol{\Omega}} \right\|_F + \left\| \sigma^2 \boldsymbol{I}_{\boldsymbol{\Omega}} - \hat{\sigma}^2 \boldsymbol{I}_{\boldsymbol{\Omega}} \right\|_F \\ &\leq 2C \text{Tr}(\mathcal{S}_\Omega(\boldsymbol{Z}^*) + \sigma^2 \boldsymbol{I}_{\boldsymbol{\Omega}}) \sqrt{\frac{\ln L}{L}} + |\sigma^2 - \hat{\sigma}^2| \sqrt{M'} \\ &\leq \left(C_1 \text{Tr}(\mathcal{S}_\Omega(\boldsymbol{Z}^*)) + C_2 \sigma^2 \right) \sqrt{\frac{\ln L}{L}} \end{aligned} \tag{68}$$

is satisfied with probability at least $1 - 5L^{-1}$, where $\hat{\boldsymbol{R}}_L = \hat{\boldsymbol{R}}_y - \hat{\sigma}^2 \boldsymbol{I}_{\boldsymbol{\Omega}}$, the first inequality is based on the triangle inequality, and the second and the third inequality is based on Lemma 2. Moreover, denote the model subspace of $\boldsymbol{Z}^*$ spanned by its column and row spaces $\mathcal{U}$ and $\mathcal{V}$ as $\mathcal{M}(\mathcal{U}, \mathcal{V})$ (using $\mathcal{M}$ hereafter for notational simplicity), and its orthogonal complement as $\mathcal{M}^\perp$ [64]. Decompose the error term $\hat{\boldsymbol{Z}} - \boldsymbol{Z}^* = \boldsymbol{H}_1 + \boldsymbol{H}_2$ into two terms satisfying $\boldsymbol{H}_1 \in \mathcal{M}$ with $\text{Rank}(\boldsymbol{H}_1) \leq K$ and $\boldsymbol{H}_2 \in \mathcal{M}^\perp$ [64]. Defining $\tau \triangleq 2(C_1 \text{Tr}(\mathcal{S}_\Omega(\boldsymbol{Z}^*)) + C_2 \sigma^2) M' \sqrt{\frac{\ln L}{L}}$, we have

$$\tau \geq 2M' \left\| \hat{\boldsymbol{R}}_L - \mathcal{S}_\Omega(\boldsymbol{Z}^*) \right\|_F \geq 2 \left\| \hat{\boldsymbol{R}}_L - \mathcal{S}_\Omega(\boldsymbol{Z}^*) \right\|_{\mathcal{A}}^*. \tag{69}$$

Then, applying Lemma 1 of [64], we have

$$\|\boldsymbol{H}_2\|_{\mathcal{A}'_d} \leq 3\,\|\boldsymbol{H}_1\|_{\mathcal{A}'_d}. \tag{70}$$

Further, by the triangle inequality,

$$\begin{aligned}\|\boldsymbol{Z}^*\|_{\mathcal{A}'_d} &= \left\|\boldsymbol{Z}^* - \hat{\boldsymbol{Z}} + \hat{\boldsymbol{Z}}\right\|_{\mathcal{A}'_d} \\ &\leq \left\|\boldsymbol{Z}^* - \hat{\boldsymbol{Z}}\right\|_{\mathcal{A}'_d} + \left\|\hat{\boldsymbol{Z}}\right\|_{\mathcal{A}'_d}\end{aligned} \tag{71}$$

and by the optimality of $\hat{\boldsymbol{Z}}$,

$$\begin{aligned}&\tau\left\|\hat{\boldsymbol{Z}}\right\|_{\mathcal{A}'_d} + \frac{1}{2}\left\|\hat{\boldsymbol{R}}_L - \mathcal{S}_\Omega(\hat{\boldsymbol{Z}})\right\|_F^2 \\ &\leq \tau\,\|\boldsymbol{Z}^*\|_{\mathcal{A}'_d} + \frac{1}{2}\left\|\hat{\boldsymbol{R}}_L - \mathcal{S}_\Omega(\boldsymbol{Z}^*)\right\|_F^2,\end{aligned} \tag{72}$$

which gives

$$\begin{aligned}&\left\|\hat{\boldsymbol{R}}_L - \mathcal{S}_\Omega(\hat{\boldsymbol{Z}})\right\|_F^2 - \left\|\hat{\boldsymbol{R}}_L - \mathcal{S}_\Omega(\boldsymbol{Z}^*)\right\|_F^2 \\ &\leq 2\tau\left(\left\|\boldsymbol{Z}^* - \hat{\boldsymbol{Z}}\right\|_{\mathcal{A}'_d}\right).\end{aligned} \tag{73}$$

Moreover, since

$$\begin{aligned}&\left\|\hat{\boldsymbol{R}}_L - \mathcal{S}_\Omega(\hat{\boldsymbol{Z}})\right\|_F^2 \\ &= \left\|\hat{\boldsymbol{R}}_L - \mathcal{S}_\Omega(\boldsymbol{Z}^*) + \mathcal{S}_\Omega(\boldsymbol{Z}^*) - \mathcal{S}_\Omega(\hat{\boldsymbol{Z}})\right\|_F^2 \\ &= \left\|\hat{\boldsymbol{R}}_L - \mathcal{S}_\Omega(\boldsymbol{Z}^*)\right\|_F^2 + \left\|\mathcal{S}_\Omega(\boldsymbol{Z}^*) - \mathcal{S}_\Omega(\hat{\boldsymbol{Z}})\right\|_F^2 \\ &\quad + 2\left\langle\hat{\boldsymbol{R}}_L - \mathcal{S}_\Omega(\boldsymbol{Z}^*), \mathcal{S}_\Omega(\boldsymbol{Z}^*) - \mathcal{S}_\Omega(\hat{\boldsymbol{Z}})\right\rangle_{\mathcal{R}},\end{aligned} \tag{74}$$

we obtain

$$\begin{aligned}&\left\|\mathcal{S}_\Omega(\hat{\boldsymbol{Z}}) - \mathcal{S}_\Omega(\boldsymbol{Z}^*)\right\|_F^2 \\ &= \left\|\hat{\boldsymbol{R}}_L - \mathcal{S}_\Omega(\hat{\boldsymbol{Z}})\right\|_F^2 - \left\|\hat{\boldsymbol{R}}_L - \mathcal{S}_\Omega(\boldsymbol{Z}^*)\right\|_F^2 \\ &\quad - 2\left\langle\hat{\boldsymbol{R}}_L - \mathcal{S}_\Omega(\boldsymbol{Z}^*), \mathcal{S}_\Omega(\boldsymbol{Z}^*) - \mathcal{S}_\Omega(\hat{\boldsymbol{Z}})\right\rangle_{\mathcal{R}} \\ &\leq 2\tau\left(\left\|\boldsymbol{Z}^* - \hat{\boldsymbol{Z}}\right\|_{\mathcal{A}'_d}\right) \\ &\quad + 2\left\|\hat{\boldsymbol{R}}_L - \mathcal{S}_\Omega(\boldsymbol{Z}^*)\right\|_F\left\|\mathcal{S}_\Omega(\boldsymbol{Z}^*) - \mathcal{S}_\Omega(\hat{\boldsymbol{Z}})\right\|_F \\ &\leq 8\tau\,\|\boldsymbol{H}_1\|_{\mathcal{A}'_d} + \frac{\tau}{M'}\left\|\mathcal{S}_\Omega(\hat{\boldsymbol{Z}}) - \mathcal{S}_\Omega(\boldsymbol{Z}^*)\right\|_F \\ &\leq \tau 8K\,\|\boldsymbol{H}_1\|_F + \frac{\tau}{M'}\left\|\mathcal{S}_\Omega(\hat{\boldsymbol{Z}}) - \mathcal{S}_\Omega(\boldsymbol{Z}^*)\right\|_F \\ &\leq \tau 8K\left\|\hat{\boldsymbol{Z}} - \boldsymbol{Z}^*\right\|_F + \frac{\tau}{M'}\left\|\mathcal{S}_\Omega(\hat{\boldsymbol{Z}}) - \mathcal{S}_\Omega(\boldsymbol{Z}^*)\right\|_F,\end{aligned} \tag{75}$$

where the first inequality is from (73) and the Cauchy-Schwartz inequality, the second inequality is from (69) and (70), and the third inequality is based on $\|\boldsymbol{H}_1\|_{\mathcal{A}'_d} \leq K\|\boldsymbol{H}_1\|_F$ following the definition of D-ANM. Considering that Ω is a complete sparse ruler such that the unobserved entries of $\mathcal{S}(Z)$ can be deduced from the observed ones, we have

$$\left\|\hat{\boldsymbol{Z}} - \boldsymbol{Z}^*\right\|_F \leq \left\|\mathcal{S}_\Omega(\hat{\boldsymbol{Z}}) - \mathcal{S}_\Omega(\boldsymbol{Z}^*)\right\|_F. \tag{76}$$

Hence, substituting (76) into (75), we have

$$\left\|\hat{\boldsymbol{Z}} - \boldsymbol{Z}^*\right\|_F \leq \tau\left(8K + \frac{1}{M'}\right), \tag{77}$$

which concludes the proof.

REFERENCES

[1] Y. Zhang, Y. Wang, Z. Tian, G. Leus, and G. Zhang, "Low-complexity gridless 2D harmonic retrieval via decoupled-ANM covariance reconstruction," in *Proc. 28th Eur. Signal Process. Conf.*, 2021, pp. 1876–1880.

[2] J. C. J. Benesty and Y. Huang, *Microphone Array Signal Processing*, vol. 1. New York, NY, USA: Springer, 2008.

[3] Y. Wang, P. Xu, and Z. Tian, "Efficient channel estimation for massive MIMO systems via truncated two-dimensional atomic norm minimization," in *Proc. IEEE Int. Conf. Commun.*, 2017, pp. 1–6.

[4] F. Wen, X. Xiong, and Z. Zhang, "Angle and mutual coupling estimation in bistatic MIMO radar based on PARAFAC decomposition," *Digit. Signal Process.*, vol. 65, pp. 1–10, 2017.

[5] J. Shi, F. Wen, and T. Liu, "Nested MIMO radar: Coarrays, tensor modeling and angle estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 57, no. 1, pp. 573–585, Feb. 2020.

[6] D. B. Rao and S. Kung, "A state space approach for the 2-D harmonic retrieval problem," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1984, pp. 174–176.

[7] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986.

[8] J. Li and R. T. Compton, "Two-dimensional angle and polarization estimation using the ESPRIT algorithm," *IEEE Trans. Antennas Propag.*, vol. 40, no. 5, pp. 550–555, May 1992.

[9] Y. Hua, "Estimating two-dimensional frequencies by matrix enhancement and matrix pencil," *IEEE Trans. Signal Process.*, vol. 40, no. 9, pp. 2267–2280, Sep. 1992.

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

[11] M. Zoltowski, M. Haardt, and C. P. Mathews, "Closed-form 2D angle estimation with rectangular arrays in element space or beamspace via unitary ESPRIT," *IEEE Trans. Signal Process.*, vol. 44, no. 2, pp. 316–328, Feb. 1996.

[12] M. Haardt, F. Roemer, and G. Del Galdo, "Higher-order SVD-based subspace estimation to improve the parameter estimation accuracy in multidimensional harmonic retrieval problems," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3198–3213, Jul. 2008.

[13] F. Xi, S. Chen, and Z. Liu, "Quadrature compressive sampling for radar signals," *IEEE Trans. Signal Process.*, vol. 62, no. 11, pp. 2787–2802, Jun. 2014.

[14] Y. Wang, Z. Tian, S. Feng, and P. Zhang, "A fast channel estimation approach for millimeter-wave massive MIMO systems," in *Proc. IEEE Glob. Conf. Signal Inf. Process.*, Dec. 2016, pp. 1413–1417.

[15] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

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

[17] P. Stoica, P. Babu, and J. Li, "SPICE: A sparse covariance-based estimation method for array processing," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 629–638, Feb. 2011.

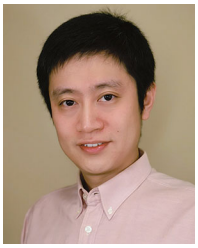
[18] Z. Liu, Z. Huang, and Y. Zhou, "Array signal processing via sparsity-inducing representation of the array covariance matrix," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 49, no. 3, pp. 1710–1724, Jul. 2013.

[19] Y. Wang, Z. Tian, S. Feng, and P. Zhang, "Efficient channel statistics estimation for millimeter-wave MIMO systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2016, pp. 3411–3415.

[20] Y. Zhang, Y. Wang, Z. Tian, G. Leus, and G. Zhang, "Super-resolution spatial channel covariance estimation for hybrid precoding in mmwave massive MIMO," in *Proc. IEEE Glob. Commun. Conf.*, 2019, pp. 1–6.

[21] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2182–2195, May 2011.
[22] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Found. Comput. Math. J. Soc. Foundations Comput. Math.*, vol. 12, no. 6, pp. 805–849, 2012.
[23] 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.
[24] 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.
[25] 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, Dec. 2013.
[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, Oct. 2016.
[27] 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.
[28] T. A. Barton and D. R. Fuhrmann, "Covariance structures for multidimensional data," *Multidimensional Syst. Signal Process.*, vol. 4, no. 2, pp. 111–123, 1993.
[29] H. Li, P. Stoica, and J. Li, "Computationally efficient maximum likelihood estimation of structured covariance matrices," *IEEE Trans. Signal Process.*, vol. 47, no. 5, pp. 1314–1323, May 1999.
[30] 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, Feb. 2015.
[31] Z. Yang, L. Xie, and P. Stoica, "Vandermonde decomposition of multilevel toeplitz matrices with application to multidimensional super-resolution," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3685–3701, Jun. 2016.
[32] J. Steinwandt, F. Roemer, C. Steffens, M. Haardt, and M. Pesavento, "Gridless super-resolution direction finding for strictly non-circular sources based on atomic norm minimization," in *Proc. 50th Asilomar Conf. Signals, Syst., Comput.*, 2016, pp. 1518–1522.
[33] Z. Tian, Z. Zhang, and Y. Wang, "Low-complexity optimization for two-dimensional direction-of-arrival estimation via decoupled atomic norm minimization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2017, pp. 3071–3075.
[34] Z. Zhang, Y. Wang, and Z. Tian, "Efficient two-dimensional line spectrum estimation based on decoupled atomic norm minimization," *Signal Process.*, vol. 163, pp. 95–106, Oct. 2019.
[35] Y. Wang, Y. Zhang, Z. Tian, G. Leus, and G. Zhang, "Angle-based channel estimation with arbitrary arrays," in *Proc. IEEE Glob. Commun. Conf.*, 2019, pp. 1–6.
[36] Y. Wang, Y. Zhang, Z. Tian, G. Leus, and G. Zhang, "Super-resolution channel estimation for arbitrary arrays in hybrid millimeter-wave massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 5, pp. 947–960, Sep. 2019.
[37] J. Zhang and M. Haardt, "Channel estimation and training design for hybrid multi-carrier mmwave massive MIMO systems: The beamspace ESPRIT approach," in *Proc. 25th Eur. Signal Process. Conf.*, 2017, pp. 385–389.
[38] S. Sahnoun, K. Usevich, and P. Comon, "Multidimensional ESPRIT for damped and undamped signals: Algorithm, computations, and perturbation analysis," *IEEE Trans. Signal Process.*, vol. 65, no. 22, pp. 5897–5910, Nov. 2017.
[39] F. Wen, H. C. So, and H. Wymeersch, "Tensor decomposition-based beamspace ESPRIT algorithm for multidimensional harmonic retrieval," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, 2020, pp. 4572–4576.
[40] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Rev.*, vol. 38, no. 1, pp. 49–95, 1996.
[41] W. Xu, J. Cai, K. V. Mishra, M. Cho, and A. Kruger, "Precise semidefinite programming formulation of atomic norm minimization for recovering d-dimensional ($d \geq 2$) off-the-grid frequencies," in *Proc. Inf. Theory Appl. Worksho*, 2014, pp. 1–4.
[42] X. Tian, J. Lei, and L. Du, "A generalized 2-D DOA estimation method based on low-rank matrix reconstruction," *IEEE Access*, vol. 6, pp. 17407–17414, 2018.
[43] Y. Wang and Z. Tian, *Big Data in 5G Encyclopedia of Wireless Networks*. Berlin, Germany: Springer, Mar. 2018. [Online]. Available: https://doi.org/10.1007/978-3-319-32903-1_58-1
[44] Y. Wang, Z. Tian, and X. Cheng, "Enabling technologies for spectrum and energy efficient NOMA-MmWave-MaMIMO systems," *IEEE Wireless Commun.*, vol. 27, no. 5, pp. 53–59, Oct. 2020.
[45] Z. Tian, Y. Tafesse, and B. M. Sadler, "Cyclic feature detection with sub-nyquist sampling for wideband spectrum sensing," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 1, pp. 58–69, Feb. 2012.
[46] D. Romero, D. D. Ariananda, Z. Tian, and G. Leus, "Compressive covariance sensing: Structure-based compressive sensing beyond sparsity," *IEEE Signal Process. Mag.*, vol. 33, no. 1, pp. 78–93, Dec. 2015.
[47] Y. Zhang, G. Zhang, and X. Wang, "Computationally efficient DoA estimation for monostatic MIMO radar based on covariance matrix reconstruction," *Electron. Lett.*, vol. 53, no. 2, pp. 111–113, 2016.
[48] R. Heckel and M. Soltanolkotabi, "Generalized line spectral estimation via convex optimization," *IEEE Trans. Inf. Theory*, vol. 64, no. 6, pp. 4001–4023, Jun. 2018.
[49] Y. Wang, Y. Zhang, Z. Tian, G. Leus, and G. Zhang, "Efficient super-resolution two-dimensional harmonic retrieval via enhanced low-rank structured covariance reconstruction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2020, pp. 5720–5724.
[50] X. Wu, W. Zhu, and J. Yan, "A toeplitz covariance matrix reconstruction approach for direction-of-arrival estimation," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 8223–8237, Sep. 2017.
[51] 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, Feb. 2016.
[52] M.-M. Liu, C.-X. Dong, Y.-Y. Dong, and G.-Q. Zhao, "Superresolution 2D DoA estimation for a rectangular array via reweighted decoupled atomic norm minimization," *Math. Problems Eng.*, vol. 2019, 2019, pp. 1–13.
[53] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and cramer-rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 5, pp. 720–741, May 1989.
[54] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound: Further results and comparisons," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 12, pp. 2140–2150, Dec. 1990.
[55] 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.
[56] J.-F. Gu and P. Wei, "Joint SVD of two cross-correlation matrices to achieve automatic pairing in 2-D angle estimation problems," *IEEE Antennas Wireless Propag. Lett.*, vol. 6, pp. 553–556, Nov. 2007.
[57] Z. Yang, L. Xie, and C. Zhang, "A discretization-free sparse and parametric approach for linear array signal processing," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 4959–4973, Oct. 2014.
[58] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.
[59] Y. Wang and Z. Tian, "IVDST: A fast algorithm for atomic norm minimization in line spectral estimation," *IEEE Signal Process. Lett.*, vol. 25, no. 11, pp. 1715–1719, Nov. 2018.
[60] S. Sekizawa, "Estimation of arrival directions using MUSIC algorithm with a planar array," in *Proc. IEEE Int. Conf. Universal Pers. Commun.*, 1998, vol. 1, pp. 555–559.
[61] H. Yan, J. Li, and G. Liao, "Multitarget identification and localization using bistatic MIMO radar systems," *EURASIP J. Adv. Signal Process.*, vol. 2008, pp. 1–8, 2007.
[62] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," Jan. 2020. [Online]. Available: http://cvxr.com/cvx
[63] F. Bunea and L. Xiao, "On the sample covariance matrix estimator of reduced effective rank population matrices, with applications to fPCA," *Bernoulli*, vol. 21, no. 2, pp. 1200–1230, 2015.
[64] S. N. Negahban, P. Ravikumar, M. J. Wainwright, and B. Yu, "A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers," *Stat. Sci.*, vol. 27, no. 4, pp. 538–557, 2012.

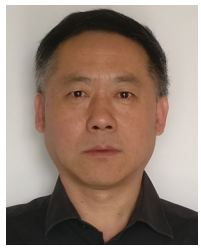
Yu Zhang (Student Member, IEEE) received the B.E. degree in 2013 from the College of Electronics and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China, where he is currently working toward the Ph.D. degree in communication and information system. From 2018 to 2020, he was a Visiting Ph.D. Student with the Electrical and Computer Engineering Department, George Mason University, Fairfax, VA, USA. His research interests include compressive sensing and array signal processing in radar and communications.



Yue Wang (Senior Member, IEEE) received the Ph.D. degree in communication and information system from the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China, in 2011. He is currently a Research Assistant Professor with Electrical and Computer Engineering Department, George Mason University, Fairfax, VA, USA, where he was a Postdoctoral Researcher. Prior to that, he was a Senior Engineer with Huawei Technologies Company, Ltd., China. From 2009 to 2011, he was a Visiting Ph.D. Student with Electrical and Computer Engineering Department, Michigan Technological University, Houghton, MI, USA. His general interests include signal processing, wireless communications, machine learning, and their applications in cyber physical systems. His specific research focuses on compressive sensing, massive MIMO, millimeter-wave communications, cognitive radios, DoA estimation, high-dimensional data analysis, and distributed optimization and learning.

Zhi Tian (Fellow, IEEE) is currently a Professor with the Electrical and Computer Engineering Department of George Mason University, Fairfax, VA, USA, since 2015. Prior to that, she was on the Faculty of Michigan Technological University, Houghton, MI, USA, from 2000 to 2014. She was a Program Director with U.S. National Science Foundation from 2012 to 2014. Her research interests include statistical signal processing, wireless communications, and estimation and detection theory. Her current research interests include compressed sensing for random processes, statistical inference of network data, distributed network optimization and learning, and millimeter-wave communications. She was an IEEE Distinguished Lecturer for both the IEEE Communications Society and the IEEE Vehicular Technology Society. She was an Associate Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and IEEE TRANSACTIONS ON SIGNAL PROCESSING. She was the General Co-Chair of the 2016 IEEE GlobalSIP Conference, and is the Chair of the IEEE Signal Processing Society Big Data Special Interest Group. She is a Member-of-Large of the IEEE Signal Processing Society for the term of 2019–2021.

Geert Leus (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in electrical engineering from Katholieke Universiteit Leuven, Leuven, Belgium, in June 1996 and May 2000, respectively. He is currently a Full Professor with the Faculty of Electrical Engineering, Mathematics and Computer Science of the Delft University of Technology, The Netherlands. He was the recipient of the 2021 EURASIP Individual Technical Achievement Award, a 2005 IEEE Signal Processing Society Best Paper Award, and a 2002 IEEE Signal Processing Society Young Author Best Paper Award. He is a Fellow of EURASIP. Geert Leus was a Member-at-Large of the Board of Governors of the IEEE Signal Processing Society, the Chair of the IEEE Signal Processing for Communications and Networking Technical Committee, and the Editor in Chief of the *EURASIP Journal on Advances in Signal Processing*. He is currently the Chair of the EURASIP Technical Area Committee on Signal Processing for Multisensor Systems and the Editor-in-Chief of *EURASIP Signal Processing*.



Gong Zhang (Member, IEEE) received the Ph.D. degree in electronic engineering from the Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China, in 2002. From 1990 to 1998, he was a Member of Technical Staff with No724 Institute China Shipbuilding Industry Corporation (CSIC), Nanjing, China. Since 1998, he has been with the College of Electronic and Information Engineering, NUAA, where he is currently a Professor. He is a Member of Committee of Electromagnetic Information, Chinese Society of Astronautics (CEI-CSA), a Member of Editorial Board for Journal of Radar, China, and a Senior Member of the Chinese Institute of Electronics. His research interests include signal processing for communications, radar, and sensor processing, statistical and array signal processing, applications of linear algebra and optimization methods in signal processing and communications, estimation and detection theory, sampling theory, classification and recognition, and cooperative and cognitive systems.