



Article Efficient 2D DOA Estimation via Decoupled Projected Atomic Norm Minimization

Mingming Liu, Yangyang Dong *D, Chunxi Dong and Guoqing Zhao

School of Electronic Engineering, Xidian University, Xi'an 710071, China; mingmingliu@stu.xidian.edu.cn (M.L.); chxdong@mail.xidian.edu.cn (C.D.); zhaoguoqingxidian@163.com (G.Z.)

* Correspondence: yydong@xidian.edu.cn

Abstract: This paper presents an efficient two-dimensional (2D) direction of arrival (DOA) estimation method, termed as decoupled projected atomic norm minimization (D-PANM), to solve the angle-ambiguity problem. It first introduces a novel atomic metric via projecting the original atom set onto a smoothing space, based on which we formulate an equivalent semi-definite programming (SDP) problem. Then, two relatively low-complexity decoupled Toeplitz matrices can be obtained to estimate the DOAs. We further exploit the structural information hidden in the newly constructed data to avoid pair matching for the azimuth and elevation angles when the number of sensors is odd, and then propose a fast and feasible decoupled alternating projections (D-AP) algorithm, reducing computational complexity to a great extent. Numerical simulations are performed to demonstrate that the proposed algorithm is no longer restricted by angle ambiguity scenarios, but instead provides a more stable estimation performance, even when multiple signals share the same angles in both azimuth and elevation dimensions. Additionally, it greatly improves the resolution, with control of the computation load compared with the existing atomic norm minimization (ANM) algorithm.

Keywords: two-dimensional DOA estimation; angle ambiguity; atomic norm minimization; pair matching; fast algorithm

1. Introduction

Two-dimensional (2D) direction of arrival (DOA) estimation is an important branch of array signal processing encountered in various applications: radar, wireless communication, sonar, seismology, etc. [1–5]. The core issue of this field is the nonlinear 2D spatial parameters estimation problem. Although numerous algorithms have been devised for 2D DOA estimation to date [6–8], the investigation of fast and effective algorithms with high resolution and precision, utilizing highly limited snapshots with increasingly complex signal scenarios, remains a hot topic.

The current state-of-the-art high-resolution 2D DOA estimation algorithms primarily focus on the subspace-based methods and the sparsity-based ones. The classic subspace algorithms are 2D MUSIC and 2D ESPRIT [9–11]: MUSIC is implemented by employing the orthogonality of the steering vectors and the noise subspace, with a huge computational cost of 2D spectral peak search, and ESPRIT constructs two subspaces of rotation-invariant properties corresponding to two diagonal angular matrices, respectively, thus avoiding a spectral search. Although the subspace methods mentioned above have achieved considerable performance in terms of resolution, theoretically reaching the Cramér–Rao bound (CRB), they heavily rely on a relatively large number of snapshots, an environment with a high signal-to-noise ratio (SNR), and a known source number. Conversely, the sparse reconstruction algorithms are intended to build mathematical models between array observation data and the 2D DOA, followed by a series of optimization steps based on different matching criteria, which no longer require the number of sources as a prior and



Citation: Liu, M.; Dong, Y.; Dong, C.; Zhao, G. Efficient 2D DOA Estimation via Decoupled Projected Atomic Norm Minimization. *Electronics* **2024**, *13*, 846. https:// doi.org/10.3390/electronics13050846

Academic Editor: Yide Wang

Received: 30 December 2023 Revised: 18 February 2024 Accepted: 20 February 2024 Published: 22 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). are robust to noise. Nevertheless, the original sparsity-based algorithms divide the entire spatial directions into discrete grids [12–14], forming a redundant dictionary to formulate the array data, where the grid mismatch problem may occur to a large extent. In view of this, varieties of off-grid algorithms have been proposed, one after the other, to overcome this strict grid limit by introducing quantization errors between divided grids and real values [15–17], and many types of strategies are applied to approximate these errors instead. Unfortunately, these strategies can still hardly express the data precisely when the quantization errors are larger, reducing the accuracy of algorithms greatly. In addition, they all have to face huge computational challenges.

The sparse reconstruction algorithms have opened a new chapter, with the concept of continuous compressed sensing (CCS) first introduced by Candès et al. [18], and the estimated parameters are no longer dependent on the grid, but allowed to take any values. Thus, a theory of super-resolution was born based on the total variation (TV) norm, along with a theoretically minimum interval condition. Due to the CCS presented in a continuous domain, Chandrasekaran et al. [19] and Tang et al. [20] then generalized the theory to the discrete domain and developed an atomic norm metric, which served as the foundation for a range of later algorithms, especially in terms of the parameter estimation, where the atomic norm minimization (ANM) theory made it possible to handle data with a few snapshots – even a single one – while retaining the advantages of sparse algorithms. Meanwhile, Tang et al. [20] argued that the atomic norm could be minimized through equivalent semi-definite programming (SDP), which was further extended to one-dimensional (1D) parameter estimation from complete and incomplete data by Yang et al. [21], and they also completed an intensive study of a reweighting strategy to enhance the resolution greatly [22]. Similar theories have been advanced regarding high-dimensional data by Chi et al. [23] and [24], [25]]. However, the computational complexity begins to increase dramatically as the number of dimensions increases. To handle this, Tian et al. [26] proposed a decoupled atomic norm minimization (DeANM) algorithm expressing the 2D Toeplitz problem into two decoupled 1D matrices. Despite the algorithm displaying a low complexity in terms of computational load, its results are ill-posed when the sources share the same angles in either azimuth or elevation dimension, i.e., they exhibit the problem of angular ambiguity.

In this paper, we consider 2D DOA estimation for a uniform rectangular array (URA) in the case of angle ambiguity and propose an efficient optimization method based on the framework of atomic norm minimization (ANM). Motivated by the idea of spatial smoothing processing [27], we introduce a novel atom metric via a projection operator, which fully exploits the phase elimination property of the SDP problem converted from ANM. The SDP is formulated mainly by two relatively low-complexity decoupled Toeplitz matrices, which is similar to the de-noising covariances in the traditional sense, and hence, the estimations of azimuth and elevation of interest can be efficiently achieved, respectively. This proposed algorithm is, therefore, named decoupled projected atomic norm minimization (D-PANM). In addition, this paper presents a more stable recovery method without pair matching, utilizing the newly constructed atom when the array has an odd number of sensors. This method always provides correct angle pairs, even in complex scenarios where multiple signals share the same angles. Subsequently, a fast implementation of D-PANM, named decoupled alternating projections (D-AP), is presented, which is generalized from the 1D alternating projections (AP) algorithm [28], reducing the computational complexity greatly compared with the most commonly used SDP solver, namely, SDPT3. However, its application is conditional. The simulation results show that the proposed D-PANM is no longer limited by the application scenarios and exhibits a better anti-noise performance than the DeANM algorithm proposed in [26]. Furthermore, compared with the vectorized ANM algorithm [23,24], it provides more effective DOA estimation without pair matching, even when multiple signals share the same azimuth and elevation angles, and it additionally has remarkable advantages in terms of both resolution and computational load.

Our main contributions are summarized as follows:

- We formulate a novel atom norm metric under the framework of ANM via a defined projected operator, which not only follows the decoupled strategy of DeANM to reduce the complexity, but also has the ability to handle the angle-ambiguity problem compared with existing methods.
- We utilize the structural characteristics of the newly constructed data to provide a more stable recovery criterion without pair matching.
- We present a fast implementation of our algorithm to reduce the computational complexity, and employ a joint low-rank projection to improve the convergence rate.
- We further show that our proposed algorithm with a decoupled reweighted strategy has a higher resolution than existing vectorized ANM.

The rest of the paper is organized as follows. Section 2 formulates the 2D DOA estimation model and introduces the problem setup. Section 3 introduces the proposed approach. Section 4 performs numerical simulations to validate the proposed method. Section 5 concludes the paper. For ease of presentation, the main abbreviations used in this paper are provided in Table 1, and a brief overview of the proposed method and the related ANM algorithms is also given in Table 2.

Acronyms	Full Name	Acronyms	Full Name	
1 D	one-dimensional	TV	total variation	
2 D	two-dimensional	AP	alternating projections	
DOA	direction of arrival	MM	majorization-minimization	
URA	uniform rectangular array	SDP	semi-definite programming	
CRB	Cramér–Rao bound	PSD	positive semidefinite	
SNR	signal-to-noise radio	CCS	continuous compressed sensing	
RMSE	root mean squared error	ANM	atomic norm minimization	
AWGN	additive white Gaussian noise	CVX	Matlab software package for convex programming [29]	
MUSIC	multiple signal classification	SDPT3	Matlab software package for semidefinite programming [30]	
ESPRIT	estimation of signal parameters using rotational invariance techniques	MMV	multiple measurement vector	
DeANM	decoupled atomic norm minimization	D-PANM	decoupled projected atomic norm minimization	
vecANM	vectorized atomic norm minimization	D-PRAM	decoupled projected reweighted atomic norm minimization	
vecRAM	vectorized reweighted atomic norm minimization	D-AP	decoupled alternating projections	

Table 1. The abbreviation index.

Table 2. An overview of the proposed method and related algorithms under the framework of ANM.

Algorithms	Main Idea	Advantages	Challenges
ANM(1D) [19,20]	Atom norm minimization (ANM). Semi-definite programming (SDP).	Resolution: high. Number of snapshots: small. Number of sources: not required.	A minimum separation between the parameters is required.

Algorithms	Main Idea	Advantages	Challenges
RAM(1D) [22]	Iteratively carry out ANM(1D) with A reweighting strategy.	Resolution: higher, no limit to minimum separation.	Applied only to 1D data.
vecANM(2D) [23]	2D ANM induced by vectorized atoms. An equivalent SDP.	Retains advantages of the ANM framework for 2D data. Ability to handle angle-ambiguity problem: yes.	Computational cost: huge. Pair matching required: yes.
DeANM(2D) [26]	2D ANM induced by atoms in matrix form. A decoupled strategy for SDP.	Computational cost: low. Number of snapshots: single.	Ability to handle angle-ambiguity problem: no. Anti-noise performance: reduced.
Proposed D-PANM(2D)	2D ANM induced by novel atoms constructed via projection operator. A decoupled strategy for SDP. A recovery criterion without pair matching based on the structural characteristics of the newly constructed data.	Ability to handle angle-ambiguity problem: yes. Anti-noise performance: better than DeANM. Computational cost: lower than vecANM Pair matching required: no. Number of snapshots: single.	Computational cost: having space to improve.
Proposed D-AP(2D)	A fast solver for D-PANM. Decoupled alternating projections. A joint projection.	Computational cost: low.	Anti-noise performance: reduced.
Proposed D-PRAM(2D)	Iteratively carries out D-PANM(2D) with a decoupled reweighting strategy.	Resolution: higher than reweighted version of vecANM, i.e., vecRAM (in Section 4).	Computational cost: increased.

Table 2. Cont.

Notations: Boldface letters stand for vectors and matrices. $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^{-1}$, $(\cdot)^+$, and $E\{\cdot\}$ denote the conjugate, transpose, conjugate transpose, inverse, pseudo-inverse, and statistical expectation, respectively. *trace* (\cdot) , $T_1(\cdot)$, $T(\cdot)$, *rank* (\cdot) , and *conv* (\cdot) represent the trace, the 1D Toeplitz matrix, the block Toeplitz matrix, the rank, and the convex hull, respectively. \mathbb{C} denotes the set of complex numbers. *diag*(**A**) retains the diagonal elements of **A** as a vector, while *DIAG*(**a**) constructs a matrix with vector **a** as the diagonal and zeros elsewhere. $[\cdot]$ and $\lfloor \cdot \rfloor$ are rounded up and down to integers, respectively. *vec*(**A**) indicates the vectorization of the matrix **A**. $\|\cdot\|$ and $\|\cdot\|_F$ stand for \updownarrow_2 and the Frobenius norm, respectively. $\|\cdot\|_{\mathcal{A}}$ represents the atom norm induced by \mathcal{A} . $|\cdot|$ is the amplitude of a complex scalar or the absolute of a real one. \otimes denotes the Kronecker product. **A** \succeq **0** implies that **A** is positive semidefinite (PSD).

2. Signal Model and Problem Statement

Consider a uniform rectangular array (URA) consisting of $N \times M$ sensors with intersensor spacing *d* along the x-direction and z-direction, respectively. There are *I* far-field narrowband uncorrelated signals $\{c_i(t)\}_{i=1}^{I}$ impinging from distinct directions $\{(\theta_i, \varphi_i)\}_{i=1}^{I}$, where θ_i represents the azimuth angle and φ_i is the elevation value. Note

that θ_i is redefined as the angle between the signal and the yoz plane rather than the traditional one (Figure 1).



Figure 1. Uniform rectangular array (URA) configuration.

Therefore, the single-snapshot array output matrix without noise can be expressed as

$$\mathbf{X} = \mathbf{A}_N(\boldsymbol{\theta}) \mathbf{C} \mathbf{A}_M^H(\boldsymbol{\varphi}) = \sum_{i=1}^{I} c_i \mathbf{a}_N(\theta_i) (\mathbf{a}_M^*(\varphi_i))^H$$
(1)

where $\mathbf{A}_N(\mathbf{\theta}) = [\mathbf{a}_N(\theta_1), \mathbf{a}_N(\theta_2), \cdots, \mathbf{a}_N(\theta_I)]$ is the array manifold matrix with the steering vectors $\mathbf{a}_N(\theta_i) = [a_{N,1}(\theta_i), a_{N,2}(\theta_i), \cdots, a_{N,N}(\theta_i)]^T$, having a Vandermonde structure of size *N*, and $\mathbf{A}_M(\mathbf{\varphi}) = [\mathbf{a}_M^*(\varphi_1), \mathbf{a}_M^*(\varphi_2), \cdots, \mathbf{a}_M^*(\varphi_I)]$ is defined similarly, with conjugated Vandermonde steering vectors $\mathbf{a}_M^*(\varphi_i) = [a_{M,1}^*(\varphi_i), a_{M,2}^*(\varphi_i), \cdots, a_{M,M}^*(\varphi_I)]^T$ of size *M*. $\mathbf{C} = diag(c_i)$, where $c_i = |c_i|\xi_i$ with $|c_i|$ being the amplitude of the receiving signal c_i and ξ_i being the phase. Let the spacing *d* be half the wavelength of the signals as usual, and the elements of **X** can then be given by

$$x(n,m) = \sum_{i=1}^{I} c_i e^{-j\pi(n-1)\sin\theta} e^{-j\pi(m-1)\sin\varphi}$$
(2)

In addition, assume that the phases $\{\xi_i\}_{i=0}^{I}$ are i.i.d. samples uniformly drawn from either distribution with mean $E\{\xi_i\} = 0$ or the complex unit circle, which is a necessary condition to guarantee the solutions of the subsequent algorithms in this paper [20,31].

The goal of our paper is to recover $\{\theta_i\}_{i=1}^{I}$ and $\{\varphi_i\}_{i=1}^{I}$ from the receiving data **X**. Note that **X** is a linear combination of a few steering matrices $\mathbf{a}_N(\theta_i) (\mathbf{a}_M^*(\varphi_i))^H$. As is similar to the decoupled atomic norm minimization (DeANM) algorithm in [26], we then utilize the atomic norm to seek the sparsest expression of some defined atoms by treating $\mathbf{a}_N(\theta_i) (\mathbf{a}_M^*(\varphi_i))^H$ as the atom. Specifically, the atom set \mathcal{A}_M and the atomic norm induced by \mathcal{A}_M are defined as

$$\mathcal{A}_{M} = \left\{ \mathbf{a}_{N}(\theta) \left(\mathbf{a}_{M}^{*}(\varphi) \right)^{H}, \theta \in [-\pi, \pi], \varphi \in [-\pi, \pi] \right\}$$
$$= \left\{ \mathbf{A}_{M}(\Theta), \Theta \in [-\pi, \pi] \times [-\pi, \pi] \right\}$$
(3)

$$\|\mathbf{X}\|_{\mathcal{A}_{M}} = \inf\left\{\sum_{i}^{I} |c_{i}| \left| \mathbf{X} = \sum_{i}^{I} c_{i} \mathbf{A}_{M}(\Theta_{i}), \ \mathbf{A}_{M}(\Theta) \in \mathcal{A}_{M} \right.\right\}$$
(4)

where $\Theta = (\theta, \varphi)$, and $\|\cdot\|_{\mathcal{A}_M}$ denotes the atom norm symbol. Next, the atomic norm $\|\mathbf{X}\|_{\mathcal{A}_M}$ is optimized via an equivalent semi-definite programming (SDP) given by

$$\|\mathbf{X}\|_{\mathcal{A}_{M}} = \min_{\mathbf{u}_{\theta}, \mathbf{u}_{\varphi}} \left\{ \frac{1}{2\sqrt{NM}} \left(trace(T_{1}(\mathbf{u}_{\theta})) + trace(T_{1}(\mathbf{u}_{\varphi})) \right) \right\}$$

$$s.t. \begin{bmatrix} T_{1}(\mathbf{u}_{\varphi}) & \mathbf{X}^{H} \\ \mathbf{X} & T_{1}(\mathbf{u}_{\theta}) \end{bmatrix} \succeq \mathbf{0}$$
(5)

where $T_1(\mathbf{u}_{\theta})$ and $T_1(\mathbf{u}_{\varphi})$ denote the 1D Toeplitz matrices, with \mathbf{u}_{θ} and \mathbf{u}_{φ} as their respective first columns, and the $\{\theta_i\}_{i=1}^{I}$ and $\{\varphi_i\}_{i=1}^{I}$ of interest are coded in these two Toeplitz matrices, respectively. Once the optimum solutions $T_1(\hat{\mathbf{u}}_{\theta})$ and $T_1(\hat{\mathbf{u}}_{\varphi})$ are determined, the estimates of $\{\theta_i\}_{i=1}^{I}$ and $\{\varphi_i\}_{i=1}^{I}$ can be obtained accordingly. Unfortunately, when two or more signals impinge from the same direction in either the azimuth or elevation dimension, the optimization results of (5) will be ill-posed because the Toeplitz matrix of the corresponding dimension will be rank-deficient.

3. The Proposed Approach

3.1. The Proposed D-PANM

In order to handle the angle-ambiguity problem mentioned above, [27] adopted a linear projection operator to map the array output matrix to a block Hankle matrix, and then exploited the newly constructed array matrix to estimate the DOAs combined with a traditional matrix pencil approach.

The constructed matrix is a $K \times \overline{K}$ block Hankle matrix, written as

$$\mathbf{X}_{e} = \begin{bmatrix} \mathbf{X}_{0} & \mathbf{X}_{1} & \cdots & \mathbf{X}_{\overline{K}-1} \\ \mathbf{X}_{1} & \mathbf{X}_{2} & \cdots & \mathbf{X}_{\overline{K}} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{X}_{K-1} & \mathbf{X}_{K} & \cdots & \mathbf{X}_{N-1} \end{bmatrix}$$
(6)

where

$$\mathbf{X}_{n} = \begin{bmatrix} x(n,0) & x(n,1) & \cdots & x(n,L-1) \\ x(n,1) & x(n,2) & \cdots & x(n,\overline{L}) \\ \vdots & \vdots & \cdots & \vdots \\ x(n,L-1) & x(n,L) & \cdots & x(n,M-1) \end{bmatrix}$$
(7)

is also an $L \times \overline{L}$ Hankle matrix, with $\overline{L} = M - L + 1$, and $\overline{K} = N - K + 1$. According to [27], the estimation of DOAs achieved the optimal performance when $K = \lceil N/2 \rceil$ and $L = \lceil M/2 \rceil$. In fact, \mathbf{X}_e is a matrix enhanced by applying smoothing processing along each dimension. Here, each column of \mathbf{X}_n is a window segment of the vector sequence $\{x(n,0), x(n,1), \dots, x(n,M-1)\}$, and the parameter *L* is the corresponding sliding window length. \mathbf{X}_e is formed by the window segment of the matrix sequence $\{\mathbf{X}_0, \mathbf{X}_1, \dots, \mathbf{X}_{N-1}\}$, and parameter *K* denotes the sliding window length accordingly.

Then, X_e can be given in the form of steering vectors by

$$\mathbf{X}_{e} = \sum_{i=1}^{I} c_{i} \left(\mathbf{a}_{K}(\theta_{i}) \mathbf{a}_{\overline{K}}^{T}(\theta_{i}) \right) \otimes \left(\mathbf{a}_{L}(\varphi_{i}) \mathbf{a}_{\overline{L}}^{T}(\varphi_{i}) \right)$$
$$= \sum_{i=1}^{I} c_{i} \left(\mathbf{a}_{K}(\theta_{i}) \otimes \mathbf{a}_{L}(\varphi_{i}) \right) \left(\mathbf{a}_{\overline{K}}^{*}(\theta_{i}) \otimes \mathbf{a}_{\overline{L}}^{*}(\varphi_{i}) \right)^{H}$$
(8)

where $\mathbf{a}_{K}(\theta_{i}) = [a_{N,1}(\theta_{i}), a_{N,2}(\theta_{i}), \cdots, a_{N,K}(\theta_{i})]^{T}$ and $\mathbf{a}_{\overline{K}}(\theta_{i}) = [a_{N,1}(\theta_{i}), a_{N,2}(\theta_{i}), \cdots, a_{N,\overline{K}}(\theta_{i})]^{T}$ are obtained from parts of $\mathbf{a}_{N}(\theta_{i})$, respectively, while being acquired in the same way from $\mathbf{a}_{M}(\varphi_{i})$ for $\mathbf{a}_{L}(\varphi_{i}) = [a_{M,1}(\varphi_{i}), a_{M,2}(\varphi_{i}), \cdots, a_{M,L}(\varphi_{i})]^{T}$ and $\mathbf{a}_{\overline{L}}(\varphi_{i}) = [a_{M,1}(\varphi_{i}), a_{M,2}(\varphi_{i}), \cdots, a_{M,L}(\varphi_{i})]^{T}$ and $\mathbf{a}_{\overline{L}}(\varphi_{i}) = [a_{M,1}(\varphi_{i}), a_{M,2}(\varphi_{i}), \cdots, a_{M,L}(\varphi_{i})]^{T}$ and $\mathbf{b}_{\overline{KL}}(\theta_{i}, \varphi_{i}) = \mathbf{a}_{\overline{K}}(\theta_{i}) \otimes \mathbf{a}_{L}(\varphi_{i})$ and $\mathbf{b}_{\overline{KL}}(\theta_{i}, \varphi_{i}) = \mathbf{a}_{\overline{K}}(\theta_{i}) \otimes \mathbf{a}_{L}(\varphi_{i})$, (8) can be concisely expressed as

$$\mathbf{X}_{e} = \sum_{i=1}^{I} c_{i} \mathbf{b}_{KL}(\theta_{i}, \varphi_{i}) \mathbf{b}_{\overline{KL}}^{H}(\theta_{i}, \varphi_{i}) = \sum_{i=1}^{I} |c_{i}| \xi_{i} \mathbf{b}_{KL}(\theta_{i}, \varphi_{i}) \mathbf{b}_{\overline{KL}}^{H}(\theta_{i}, \varphi_{i}) = \mathbf{B}\mathbf{F}^{H}$$
(9)

where $\mathbf{B} = [\mathbf{b}_{KL}(\theta_1, \varphi_1), \mathbf{b}_{KL}(\theta_2, \varphi_2), \cdots, \mathbf{b}_{KL}(\theta_I, \varphi_I)]$, and $\mathbf{F}^H = [\mathbf{f}_1^H, \mathbf{f}_2^H, \cdots, \mathbf{f}_I^H]^T$ with $\mathbf{f}_i = |c_i|\xi_i^* \mathbf{b}_{\overline{KL}}(\theta_i, \varphi_i)$. Using (9), it follows that

$$\mathbf{X}_{e}\mathbf{X}_{e}^{H} = \mathbf{B}\mathbf{F}^{H}\mathbf{F}\mathbf{B}^{H} \tag{10}$$

Note that $\mathbf{F}^{H}\mathbf{F}$ is a diagonally dominant matrix with real values on the diagonal, based on which a new covariance matrix $\mathbf{R}_{1} = \mathbf{B}\Lambda_{1}\mathbf{B}^{H}$ with $\Lambda_{1} = diag(\mathbf{F}^{H}\mathbf{F}) =$ $DIAG([s_{1}^{2}, s_{2}^{2}, \dots, s_{I}^{2}])$ is redefined, where $s_{i}^{2} = \|\mathbf{f}_{i}\|^{2} = \overline{KL}|c_{i}|^{2}$. Obviously, the newly defined \mathbf{R}_{1} displays the same features as the traditional covariance matrix of the Toeplitz structure, and therefore, as long as \mathbf{R}_{1} is given, the azimuth angles $\{\theta_{i}\}_{i=1}^{I}$ can be estimated via extensive existing approaches such as the Vandermonde decomposition [24], the matrix pencil method [32], etc.

On the other hand, we construct another matrix of size $LK \times \overline{LK}$ as follows:

$$\mathbf{X}'_{e} = \sum_{i=1}^{I} c_{i}(\mathbf{a}_{L}(\varphi_{i}) \otimes \mathbf{a}_{K}(\theta_{i})) \left(\mathbf{a}_{\overline{L}}^{*}(\varphi_{i}) \otimes \mathbf{a}_{\overline{K}}^{*}(\theta_{i})\right)^{H}$$
(11)

It is also a block Hankle matrix applied to estimate the elevation angles $\{\varphi_i\}_{i=1}^{l}$ [33], whose analysis procedure is similar to that of the azimuth angles, and (11) can also be rewritten as

$$\mathbf{X}'_{e} = \sum_{i=1}^{I} |c_{i}| \boldsymbol{\xi}_{i} \mathbf{d}_{LK}(\varphi_{i}, \theta_{i}) \mathbf{d}_{\overline{LK}}^{H}(\varphi_{i}, \theta_{i}) = \mathbf{D}\mathbf{G}^{H}$$
(12)

by letting $\mathbf{d}_{LK}(\varphi_i, \theta_i) = \mathbf{a}_L(\varphi_i) \otimes \mathbf{a}_K(\theta_i)$ and $\mathbf{d}_{\overline{LK}}(\varphi_i, \theta_i) = \mathbf{a}_{\overline{L}}^*(\varphi_i) \otimes \mathbf{a}_{\overline{K}}^*(\theta_i)$, where $\mathbf{D} = [\mathbf{d}_{LK}(\varphi_1, \theta_1), \mathbf{d}_{LK}(\varphi_2, \theta_2), \cdots, \mathbf{d}_{LK}(\varphi_I, \theta_I)]$, $\mathbf{G}^H = [\mathbf{g}_1^H, \mathbf{g}_2^H, \cdots, \mathbf{g}_I^H]^T$ with $\mathbf{g}_i = |c_i|\xi_i^*\mathbf{d}_{\overline{LK}}(\varphi_i, \theta_i)$; hence, the diagonally dominant matrix $\mathbf{X}'_e\mathbf{X}'_e^H = \mathbf{D}\mathbf{G}^H\mathbf{G}\mathbf{D}^H$ is obtained. We redefined another new covariance matrix $\mathbf{R}_2 = \mathbf{D}\Lambda_2\mathbf{D}^H$ by retaining $diag(\mathbf{G}^H\mathbf{G}) = DIAG([s_1^2, s_2^2, \cdots, s_I^2])$ as the diagonal matrix Λ_2 , and then $\{\varphi_i\}_{i=1}^I$ could be estimated through an assumed \mathbf{R}_2 in a similar fashion to that outlined in the approaches mentioned earlier.

Inspired by the analysis above, we introduce a permutation matrix:

$$\mathbf{H} = \sum_{j}^{L} \sum_{i}^{K} H_{j,i}^{\overline{L} \times \overline{K}} \otimes H_{i,j}^{\overline{K} \otimes \overline{L}}$$
(13)

where $H_{j,i}^{\overline{L}\times\overline{K}}$ denotes a matrix of size $\overline{L}\times\overline{K}$ with one at the position (j,i) and zeros elsewhere, and $H_{i,j}^{\overline{K}\otimes\overline{L}}$ is defined similarly, except for the size, which is $\overline{K}\times\overline{L}$. Then, we further construct more efficient array observation data $\mathbf{X}_{e}\mathbf{H}^{H}$ rather than utilizing \mathbf{X} or \mathbf{X}_{e} directly, expressed as

$$\begin{aligned} \mathbf{X}_{\varepsilon} \mathbf{H}^{H} &= \sum_{i=1}^{I} c_{i} (\mathbf{a}_{K}(\theta_{i}) \otimes \mathbf{a}_{L}(\varphi_{i})) \left(\mathbf{a}_{\overline{L}}^{*}(\varphi_{i}) \otimes \mathbf{a}_{\overline{K}}^{*}(\theta_{i})\right)^{H} \\ &= \sum_{i=1}^{I} c_{i} \mathbf{b}_{KL}(\theta_{i},\varphi_{i}) \mathbf{d}_{\overline{LK}}^{H}(\varphi_{i},\theta_{i}) \\ &= \mathbf{B} \mathbf{C} \mathbf{D}^{'H} \end{aligned}$$
(14)

where $\mathbf{D}' = [\mathbf{d}_{\overline{LK}}(\varphi_1, \theta_1), \mathbf{d}_{\overline{LK}}(\varphi_2, \theta_2), \cdots, \mathbf{d}_{\overline{LK}}(\varphi_I, \theta_I)]$ has the same structure as the conjugate of **D**, except in terms of size. We denoted the projection operator of the original array receiving matrix **X** onto the new observation data space by

$$\mathcal{P}(\mathbf{X}) = \sum_{i=1}^{I} c_i \mathcal{P}\left(\mathbf{a}_N(\theta_i)(\mathbf{a}_M^*(\varphi_i))^H\right) = \sum_{i=1}^{I} c_i \mathbf{b}_{KL}(\Theta_i) \mathbf{d}_{\overline{LK}}^H(\Theta_i) = \mathbf{BCD'}^H$$
(15)

and the range of \mathcal{P} by $\mathbf{P} = \left\{ \mathbf{E} \in \mathbb{C}^{KL \times \overline{LK}} | \mathbf{E} = \mathcal{P}(\mathbf{X}), \mathbf{X} \in \mathbb{C}^{N \times M} \right\}$, where \mathcal{P} is essentially a smoothing operator that handles the data through multiple sliding windows: $K = \lceil N/2 \rceil$, $L = \lceil M/2 \rceil$, $\overline{K} = N - K + 1$, and $\overline{L} = M - L + 1$.

Next, let us define a new projected atom set as:

$$\mathcal{A}_{\mathcal{P}} = \{\mathcal{P}, (\mathcal{A}_{M})\} = \left\{ \mathbf{b}_{KL}(\Theta) \mathbf{d}_{\overline{LK}}^{H}(\Theta), \Theta \in [-\pi, \pi] \times [-\pi, \pi] \right\}$$
$$= \{\mathbf{A}_{\mathcal{P}}(\Theta), \Theta \in [-\pi, \pi] \times [-\pi, \pi] \}$$
(16)

and hence, the projected atomic norm induced by the convex hull of $\mathcal{A}_{\mathcal{P}}$ is given by

$$\begin{aligned} \|\mathcal{P}(\mathbf{X})\|_{\mathcal{A}_{\mathcal{P}}} &= \inf\{t > 0 | \mathcal{P}(\mathbf{X}) \in tconv(\mathcal{A}_{\mathcal{P}})\} \\ &= \inf\left\{\sum_{i}^{I} |c_{i}| \left| \mathcal{P}(\mathbf{X}) = \sum_{i}^{I} c_{i} \mathbf{A}_{\mathcal{P}}(\Theta_{i}), \ \mathbf{A}_{\mathcal{P}}(\Theta) \in \mathcal{A}_{\mathcal{P}} \right\} \end{aligned}$$
(17)

Note that (17) is an optimization problem seeking the smallest possible combination of the projected atoms from an infinite set $\mathcal{A}_{\mathcal{P}}$. This is not easy to solve directly. As such, we turn to the optimization of an equivalent semi-definite programming (SDP) instead, and propose a decoupled projected atomic norm minimization (D-PANM) problem as described in the proposition below.

Proposition 1. For any $N \times M$ array receiving matrix $\mathbf{X}^0 = \sum_{i=1}^{L} c_i \mathbf{A}_M(\Theta_i)$, consider new observa-

tion data of size $KL \times \overline{LK}$ *as follows:*

$$\mathcal{P}(\mathbf{X}) = \sum_{i}^{I} c_k \mathcal{P}(\mathbf{A}_M(\Theta_i)) = \sum_{i}^{I} c_k \mathbf{A}_{\mathcal{P}}(\Theta_i)$$
(18)

utilizing the projection operator of $\mathbf{X} = \mathbf{X}^0$ onto \mathbf{P} , if the minimum angle distance between $\{\Theta_i\}_{i=1}^{I}$ satisfies

$$\Delta_{\min,\Theta} = \min_{i \neq j} \left\| \Theta_i - \Theta_j \right\|_{-\infty}$$

= min_{i \neq j} \left\{ \left| \theta_i - \theta_j \right|, \left| \varphi_i - \varphi_j \right| \right\} \ge \min\{ \lfloor 4/(N-1) \rfloor, \lfloor 4/(M-1) \rfloor \} (19)

where $|\theta_i - \theta_j|$ and $|\varphi_i - \varphi_j|$ are considered the wrapped distances on the unit circle. Then, the solution to (17) is guaranteed by (18) with at least a possibility of $1 - \delta$, and two Toeplitz matrices with $\{\Theta_i\}_{i=1}^{I}$ coded in can be efficiently achieved via a SDP given by:

$$\|\mathcal{P}(\mathbf{X})\|_{\mathcal{A}_{\mathcal{P}}} = \min_{\mathbf{X}, \mathbf{u}_{b}, \mathbf{u}_{d}} \left\{ \frac{1}{2\sqrt{KL\overline{KL}}} (trace(T(\mathbf{u}_{b})) + trace(T(\mathbf{u}_{d}))) \right\}$$

s.t. $\begin{bmatrix} T(\mathbf{u}_{d}) & \mathcal{P}(\mathbf{X})^{H} \\ \mathcal{P}(\mathbf{X}) & T(\mathbf{u}_{b}) \end{bmatrix} \succeq \mathbf{0}, \mathbf{X} = \mathbf{X}^{0}$ (20)

Since \mathcal{P} is a linear operator satisfying its homogeneity and additivity, the performance guarantees of this proposition, followed by the equivalent SDP problem, can be easily derived from the atom norm theory in [20,26,31]. We omit the details and provide a brief proof of the equivalence (20) in Appendix A.

Remark 1. Given

$$\mathcal{P}(\mathbf{X}) = \sum_{i}^{I} c_{i} \mathbf{A}_{\mathcal{P}}(\Theta_{i}) = \sum_{i}^{I} |c_{i}| \mathbf{A}_{\mathcal{P}}(\Theta_{i}, \xi_{i}) = \sum_{i}^{I} |c_{i}| \mathbf{b}_{KL}\left(\Theta_{i}, \frac{1}{2}\xi_{i}\right) \mathbf{d}_{\overline{LK}}^{H}\left(\Theta_{i}, \frac{1}{2}\xi_{i}\right), \quad (21)$$

the positive semidefinite (PSD) feasible cone can be expressed as follows:

$$\begin{bmatrix} T(\mathbf{u}_d) & \mathcal{P}(\mathbf{X})^H \\ \mathcal{P}(\mathbf{X}) & T(\mathbf{u}_b) \end{bmatrix} = \sum_i^I |c_i| \begin{bmatrix} \mathbf{d}_{\overline{LK}} \left(\Theta_i, \frac{1}{2} \xi_i \right) \\ \mathbf{b}_{KL} \left(\Theta_i, \frac{1}{2} \xi_i \right) \end{bmatrix} \begin{bmatrix} \mathbf{d}_{\overline{LK}} \left(\Theta_i, \frac{1}{2} \xi_i \right) \\ \mathbf{b}_{KL} \left(\Theta_i, \frac{1}{2} \xi_i \right) \end{bmatrix}^H \approx \mathbf{0}.$$
(22)

Then, one has the following two Toeplitz matrices

$$T(\mathbf{u}_b) = \sum_{i}^{I} |c_i| \mathbf{b}_{KL} \left(\Theta_i, \frac{1}{2}\xi_i\right) \mathbf{b}_{KL}^{H} \left(\Theta_i, \frac{1}{2}\xi_i\right) = \sum_{i}^{I} |c_i| \mathbf{b}_{KL} (\Theta_i, 0) \mathbf{b}_{KL}^{H} (\Theta_i, 0)$$
(23)

$$T(\mathbf{u}_d) = \sum_{i}^{I} |c_i| \mathbf{d}_{\overline{LK}} \left(\Theta_i, \frac{1}{2} \xi_i \right) \mathbf{d}_{\overline{LK}}^{H} \left(\Theta_i, \frac{1}{2} \xi_i \right) = \sum_{i}^{I} |c_i| \mathbf{d}_{\overline{LK}} (\Theta_i, 0) \mathbf{d}_{\overline{LK}}^{H} (\Theta_i, 0)$$
(24)

with $\mathbf{u}_b = \sum_{i}^{l} |c_i| \mathbf{b}_{KL}(\Theta_i, 0)$ and $\mathbf{u}_d = \sum_{i}^{l} |c_i| \mathbf{d}_{\overline{LK}}(\Theta_i, 0)$, which implies that $T(\mathbf{u}_b)$ in the PSD cone owns the same column space as the covariance matrix \mathbf{R}_1 defined previously, while $T(\mathbf{u}_d^*)$ also has the same Toeplitz structure as \mathbf{R}_2 , except for the size and the constant differences in terms of the diagonal elements. That is to say, we can efficiently estimate $\{\theta_i\}_{i=1}^{l}$ and $\{\varphi_i\}_{i=1}^{l}$ by exploiting the optimum solutions $T(\hat{\mathbf{u}}_b)$ and $T(\hat{\mathbf{u}}_d^*)$, respectively. It is important to note that the PSD constraint above will eliminate the phase information of multiple measurement data, which here results in the phase eliminations of $\mathcal{P}(\mathbf{X})$ and $\mathcal{P}(\mathbf{X})^H$, respectively, making it possible to separately process the two spaces of $\mathcal{P}(\mathbf{X})$. However, for single-snapshot data, this constraint can help to deal with cases where the rank of the covariance is one, which explains how the framework of ANM can even handle the data with a single snapshot. In addition, the estimators $T(\hat{\mathbf{u}}_b)$ and $T(\hat{\mathbf{u}}_d^*)$ are obtained with a de-noising process; thus, knowledge of the number of sources is not needed, even if the traditional methods are then applied to achieve estimates of $\{\theta_i\}_{i=1}^{l}$ and $\{\varphi_i\}_{i=1}^{l}$.

Remark 2. In particular, the conditions for the minimum angle distances still rely on the size of X (i.e., the lengths N and M) rather than the projection parameters K and L. Furthermore, [23] has shown that the minimum distances can be relaxed to $\min\{\lfloor 1.19/(N-1) \rfloor, \lfloor 1.19/(M-1) \rfloor\}$, which satisfies most practical applications. We conducted experiments to verify these, and these are described in Section 4 of the paper.

In fact, the array receiving data are always corrupted by the additive white Gaussian noise (AWGN) as follows:

$$\mathbf{Y}^0 = \mathbf{X} + \mathbf{N} \tag{25}$$

and hence, the projection of \mathbf{Y}^0 onto *P* is given by

$$\mathcal{P}(\mathbf{Y}^0) = \mathcal{P}(\mathbf{X}) + \mathcal{P}(\mathbf{N})$$
(26)

Nevertheless, $\mathcal{P}(\mathbf{N})$ can hardly be employed directly, because projecting **N** onto **P** will not only increase the computation complexity, but also make the statistically independent noise variables become correlated. Note that \mathcal{P} is a linear mapping, and thus, $\mathcal{P}(\mathbf{N})$ is uniquely determined by the noise data **N**. Thereby, **N** can be considered as the kernel of $\mathcal{P}(\mathbf{N})$, and is an acceptable noise constraint. Combining (17), (20), and Remark 1, two Toeplitz optimization matrices, where $\{\theta_i\}_{i=1}^{I}$ and $\{\varphi_i\}_{i=1}^{I}$ of interest are coded in, are achieved in the presence of noise via

$$\min_{\mathbf{X}} \left\| \mathcal{P}(\mathbf{X}) \right\|_{\mathcal{A}_{\mathcal{P}}} s.t. \left\| \mathbf{Y}^{0} - \mathbf{X} \right\|_{F}^{2} \leq \eta^{2}$$
(27)

or, equivalently,

$$\min_{\mathbf{X},\mathbf{u}_{b},\mathbf{u}_{d}} \left\{ \frac{1}{2\sqrt{KL\overline{KL}}} (trace(T(\mathbf{u}_{b})) + trace(T(\mathbf{u}_{d}))) \right\} \\
s.t. \begin{bmatrix} T(\mathbf{u}_{d}) & \mathcal{P}(\mathbf{X})^{H} \\ \mathcal{P}(\mathbf{X}) & T(\mathbf{u}_{b}) \end{bmatrix} \succeq \mathbf{0}, \left\| \mathbf{Y}^{0} - \mathbf{X} \right\|_{F}^{2} \le \eta^{2}$$
(28)

where η^2 is the noise level known. The equations in (20) or (28) are usually solved based on the interior point method using a SDP solver of the CVX tool, namely, SDPT3 [29,30]. Then, $\{\theta_i\}_{i=1}^{I}$ and $\{\varphi_i\}_{i=1}^{I}$ can be estimated from the optimum solutions $T(\hat{\mathbf{u}}_b)$ and $T(\hat{\mathbf{u}}_d)$ via traditional estimation methods, as mentioned before.

3.2. The Odd-Number Array and Fast Algorithm

This section focuses mainly on the case where both *N* and *M* are odd, and develops a new estimation criterion for $\{\theta_i\}_{i=1}^{I}$ and $\{\varphi_i\}_{i=1}^{I}$ without pair matching [33] after the SDP (20) or (28) is solved, which can bring computational convenience and performance improvements.

Now, we select the optimal parameters K = (N+1)/2 and L = (M+1)/2, where one has $\overline{K} = N - K + 1 = K$ and $\overline{L} = L$. Assume that two estimators $T(\hat{\mathbf{u}}_b)$ and $T(\hat{\mathbf{u}}_d)$ have been obtained, and then the following equations hold:

$$\left(\mathbf{D}'\right)^* = \mathbf{P}\mathbf{B} \ T(\hat{\mathbf{u}}_d^*) = \mathbf{P}T(\hat{\mathbf{u}}_b)\mathbf{P}^H$$
(29)

according to (15), (23), and (24). We denote the eigenvalue decompositions of $T(\hat{\mathbf{u}}_b)$ and $T(\hat{\mathbf{u}}_d)$ by

$$T(\hat{\mathbf{u}}_b) = \mathbf{U}_{bs} \mathbf{\Lambda}_{bs} \mathbf{U}_{bs}^H, \ T(\hat{\mathbf{u}}_d^*) = \mathbf{U}_{ds} \mathbf{\Lambda}_{ds} \mathbf{U}_{ds}^H$$
(30)

where \mathbf{U}_{bs} contains the principal eigenvectors of $T(\hat{\mathbf{u}}_b)$, whose eigenvalues satisfy $\{\sigma_{b,i} \geq \gamma \sigma_{b,\max}\}$. Here, $\sigma_{b,\max}$ denotes the maximum eigenvalue of $T(\hat{\mathbf{u}}_b)$, and γ is a constant and can be fixed at 0.1. Meanwhile, \mathbf{U}_{ds} is obtained in the same way from $T(\hat{\mathbf{u}}_d^*)$, which is reasonable because the minor components of the de-noising estimators $T(\hat{\mathbf{u}}_b)$ and $T(\hat{\mathbf{u}}_d^*)$ have almost near-zero eigenvalues. Then, assuming there are two nonsingular matrices such that

$$\mathbf{U}_{bs} = \mathbf{B}\mathbf{U}_{1}^{-1}, \ \mathbf{U}_{ds} = \left(\mathbf{D}'\right)^{*}\mathbf{U}_{2}^{-1}$$
(31)

we obtain $\mathbf{U}_{ds} = \mathbf{PBU}_2^{-1} = \mathbf{PU}_{bs}\mathbf{U}_1\mathbf{U}_2^{-1}$. Denoting $\mathbf{U}_1\mathbf{U}_2^{-1} \triangleq \mathbf{O}$ leads to $\mathbf{U}_{ds} = \mathbf{PU}_{bs}\mathbf{O}$ and

$$\mathbf{O}^{-1}\mathbf{U}_1 = \mathbf{U}_2 \tag{32}$$

Let us further construct $\mathbf{U}^+_{-bs} \overline{\mathbf{U}}_{bs}$ and $\mathbf{U}^+_{-ds} \overline{\mathbf{U}}_{ds}$, where \mathbf{U}_{-bs} and $\overline{\mathbf{U}}_{bs}$ are selected from \mathbf{U}_{bs} with the last and first *L* rows deleted, respectively, while \mathbf{U}_{-ds} and $\overline{\mathbf{U}}_{ds}$ are constructed from \mathbf{U}_{ds} in the same way with the \overline{K} rows deleted. We then obtain the two following eigenvalue decompositions:

$$\mathbf{U}_{-bs}^{+} \overline{\mathbf{U}}_{bs} = \mathbf{U}_{1} \Gamma \mathbf{U}_{1}^{-1}, \ \mathbf{U}_{-ds}^{+} \overline{\mathbf{U}}_{ds} = \mathbf{U}_{2} \Psi \mathbf{U}_{2}^{-1}$$
(33)

and $\{\theta_i\}_{i=1}^{I}$ and $\{\varphi_i\}_{i=1}^{I}$ are coded in the diagonal eigenvalue matrices Γ and Ψ according to the rotational invariant subspace method [27]. Using (32) and (33), it follows that

$$\boldsymbol{\Gamma} = \mathbf{U}_{1}^{-1} \mathbf{U}_{-bs}^{+} \overline{\mathbf{U}}_{bs} \mathbf{U}_{1}, \ \boldsymbol{\Psi} = \mathbf{U}_{2}^{-1} \mathbf{U}_{-ds}^{+} \overline{\mathbf{U}}_{ds} \mathbf{U}_{2} = \mathbf{U}_{1}^{-1} \mathbf{O} \mathbf{U}_{1}^{+} \overline{\mathbf{U}}_{ds} \mathbf{O}^{-1} \mathbf{U}_{1}.$$
(34)

Since **O** is also a nonsingular matrix, $\mathbf{OU}^+ \overline{\mathbf{U}}_{ds} \mathbf{O}^{-1}$ displays the same eigenvalues as $\mathbf{U}^+ \overline{\mathbf{U}}_{ds}$, i.e., the eigenvalue matrix Ψ . Then, the following linear fitting and a joint eigenvalue decomposition are computed by

$$\alpha \mathbf{U}^{+}_{-bs} \overline{\mathbf{U}}_{bs} + (1-\alpha) \mathbf{U}^{+}_{-ds} \overline{\mathbf{U}}_{ds} = \mathbf{U} \mathbf{D} \mathbf{U}^{H}$$
(35)

where parameter α is introduced to prevent the rank defect when $\theta_i = \varphi_i$. Finally, the eigenvalue matrices can be obtained, adopting the common eigenvector matrix **U** as

$$\Gamma = \mathbf{U}^{H}\mathbf{U}^{+}_{-bs}\overline{\mathbf{U}}_{bs}\mathbf{U}, \ \mathbf{\Psi} = \mathbf{U}^{H}\mathbf{U}^{+}_{-ds}\mathbf{U}_{ds}\mathbf{U}$$
(36)

and the estimation of the DOAs can be achieved directly using $\hat{\Phi} = (diag(\Gamma), diag(\Psi))$ without a pairing step. This new estimation criterion not only quickens the solving process to a certain extent, but also provides the DOA estimation without pair matching, which proves to be more effective even when multiple signals share the same angles in both the azimuth and elevation dimensions (Section 4).

In addition, due to higher temporal complexity of the SDPT3 solver based on the interior point method, we further propose a fast implementation of the SDP (20) or (28), termed as the decoupled alternating projections (D-AP) algorithm, to adequately exploit the structural information about $\mathcal{P}(\mathbf{X})$, which is motivated by [28,34]. Note that the optimization problem (20) displays the following features:

1. The feasible set

$$\mathbf{S} = \begin{bmatrix} T(\mathbf{u}_d) & \mathcal{P}(\mathbf{X})^H \\ \mathcal{P}(\mathbf{X}) & T(\mathbf{u}_b) \end{bmatrix}$$
(37)

is PSD.

- 2. We obtain $T(\mathbf{u}_b) \geq \mathbf{0}$ and $T(\mathbf{u}_d) \geq \mathbf{0}$ using the Schur Complement Lemma [35]; hence, $\lambda_{d,i} \geq 0$ and $\lambda_{b,i} \geq 0$ for all the $i = 1, 2, \dots, I$, where $\lambda_{d,i}$ and $\lambda_{b,i}$ denote the eigenvalues of $T(\mathbf{u}_b)$ and $T(\mathbf{u}_d)$, respectively. Thus, the essence of the objective function in (20) becomes minimizing two \uparrow_1 norms of the eigenvalues. Also, because of the Hermitian features of $T(\mathbf{u}_b)$ and $T(\mathbf{u}_d)$, (20) is equivalent to optimizing two low-rank Toeplitz matrices.
- 3. $T(\mathbf{u}_b)$ and $T(\mathbf{u}_d)$ on the PSD cone have the following relationship:

$$T(\mathbf{u}_d) = \mathbf{P}T(\mathbf{u}_b^*)\mathbf{P}^H \tag{38}$$

Aiming at the analysis above, some important projection operators are defined accordingly:

1. Let *S* be the projection of a $J \times J$ Hermitian matrix **A** onto the PSD subspace, and the range of *S* is the PSD set as

$$\mathcal{S}(\mathbf{A}) = \left\{ \mathbf{B} \in \mathbb{C}^{J \times J} \middle| \mathbf{B} = \sum_{j=1}^{J} \max(\lambda_j, 0) \mathbf{v}_j \mathbf{v}_j^H, \, \mathbf{A} = \sum_{j=1}^{J} \lambda_j \mathbf{v}_j \mathbf{v}_j^H \right\}$$
(39)

where $\mathbf{A} = \sum_{j=1}^{J} \lambda_j \mathbf{v}_j \mathbf{v}_j^H$ is the eigenvalue decomposition with eigenvalues $\{\lambda_j\}_{j=1}^{J}$ and eigenvectors $\{\mathbf{v}_j\}_{j=1}^{J}$.

2. Let \mathcal{L} be the projection of a $J \times J$ Hermitian matrix **A** onto the low-rank set, with a rank of no more than Q, by introducing a threshold parameter τ ,

$$\mathcal{L}(\mathbf{A}) = \left\{ rank(\mathbf{B}) \le Q \middle| \mathbf{B} = \sum_{j=1}^{J} \max(\lambda_j - \tau, 0) \mathbf{v}_j \mathbf{v}_j^H \mathbf{A} = \sum_{j=1}^{J} \lambda_j \mathbf{v}_j \mathbf{v}_j^H \right\}$$
(40)

which aims mainly to make the eigenvalues of **A** sparse, and τ is chosen for a balance between the accuracy of the solution and the convergence rate of the algorithm [36].

3. Let \mathcal{T} be the projection of a $KL \times KL$ matrix **A** onto the two-level Toeplitz subspace, such that

$$\mathcal{T}(\mathbf{A}) = \left\{ \mathcal{T}_{B}(\mathbf{U}) \middle| \begin{array}{l} \mathbf{U}_{K-i+1,j} = \frac{1}{K-i+1} \sum_{\substack{m=1\\m=1}}^{K-i+1} (\mathbf{B}_{j})_{m,m+i-1}, \ i = 1, \dots, K, j = 1, 2, \dots, L \\ \mathbf{U}_{K-i-1,j} = \frac{1}{K+i+1} \sum_{\substack{m=1\\m=1}}^{K+i+1} (\mathbf{B}_{j})_{m-i-1,m}, \ i = -K, \dots, -2, j = 1, 2, \dots, L \\ \mathbf{B}_{j} = \frac{1}{2(L-j+1)} \sum_{t=1}^{L-j+1} \mathbf{A}((t-1)K+1:tK, (t-1)K+(j-1)K+1:tK+(j-1)K) + \\ \mathbf{H}_{i} = \frac{1}{2(L-j+1)} \sum_{t=1}^{L-j+1} \mathbf{A}^{H}((t-1)K+(j-1)K+1:tK+(j-1)K, (t-1)K+1:tK) \end{array} \right\}$$
(41)

where $\mathcal{T}_B(\mathbf{U})$ is an $L \times L$ block Toeplitz projection with each block $T_h(\mathbf{U}_{K:2K-1,|h|+1}, \mathbf{U}_{1:K,|h|+1})$ being a 1D Toeplitz matrix of size $K \times K$, $h = -(L-1), \dots, 0, 1, \dots, (L-1)$ [23], and $\mathbf{U}_{K:2K-1,|h|+1}$ and $\mathbf{U}_{1:K,|h|+1}$ denote the first column and the first row of T_h , respectively.

Then, the D-AP algorithm can be carried out via iteratively projecting the optimization variables **S**, $T(\mathbf{u}_b)$, and $T(\mathbf{u}_d)$ onto the corresponding spaces, which are presented in detail in Table 3, where $T(\mathbf{u}_b)$ and $T(\mathbf{u}_d)$ share a joint low-rank projection utilizing the mathematical relationship (29), improving both the accuracy and the convergence rate of the algorithm to a certain extent.

Table 3. The D-AP algorithm.

A

gorithm 1: D-AP		
1. Initialization: given $n = \overline{KL}$, $\mathbf{Z}^0 = (\mathbf{Y}^0)$, $T^0(\mathbf{u}_b) = (\mathbf{Z}\mathbf{Z}^H)$, $T^0(\mathbf{u}_d) = (\mathbf{Z}^H\mathbf{Z})$		
2. Iteration:		
for j = 1:max iterations		
2.1 $\mathbf{S}^{j-1} = \begin{bmatrix} T^{j-1}(\mathbf{u}_d) & \left(\mathbf{Z}^{j-1}\right)^H \\ \mathbf{Z}^{j-1} & T^{j-1}(\mathbf{u}_b) \end{bmatrix}$, and $\mathbf{S}^j = \left(\mathbf{S}^{j-1}\right)$		
2.2 $T^{j}(\mathbf{u}_{b}) = \mathbf{S}^{j}(n+1:end, n+1:end), T^{j}(\mathbf{u}_{b}) = \mathcal{L}\left(T^{j}(\mathbf{u}_{b})\right) T^{j}(\mathbf{u}_{d}) = \mathbf{P}T^{j}(\mathbf{u}_{b})\mathbf{P}^{H}$		
2.3 cut-off condition: if $\ \mathbf{S}^{j} - \mathbf{S}^{j-1}\ _{F} \leq \delta$, break		
end		
3. Estimation: DOAs $\hat{\Phi}$ are estimated through $T(\hat{\mathbf{u}}_b)$ and $T(\hat{\mathbf{u}}_d)$.		

Remark 3. The initial \mathbf{Y}^0 should be unitized, helping to select the parameter τ in the low-rank projection steps, and an empirical value 0.1 is provided here for the signals in this paper. Note that the low-rank projection in the D-AP algorithm is also a de-noising process in the presence of noise, so this algorithm can also deal with the SDP (28). In addition, the convergence of our proposed D-AP algorithm has been guaranteed, because the range spaces of the projection operators utilized in D-AP are all closed and convex except for the low-rank projection, which is nonconvex, but has been proven to exercise no influence on the convergence of the algorithm [37,38].

Remark 4. The D-AP algorithm is an iterative process that involves alternating projections onto the PSD and two low-rank spaces. Specifically, at each iteration, the PSD projection consists of two steps: the eigenvalue decomposition and matrix multiplication, resulting in a computational complexity of $O((1/2^2)(NM)^3)$. The low-rank projections follow a similar process, but with reduced complexity of $O((3/2^6)(NM)^3)$ due to a joint projection of $T(\mathbf{u}_b)$ and $T(\mathbf{u}_d)$. Assuming that the maximum number of iterations is k_p , the overall complexity is $O(k_p(19/2^6)(N)^6)$ for N = M. Recall that the D-PANM algorithm implemented by the SDPT3 solver relies on calculating the Newton direction by solving a group of linear equations, and its complexity, determined by the size of the PSD constraint, is $O((N)^7)$ for N = M [26]. Consequently, the D-AP algorithm offers a more computationally efficient solution compared with the D-PANM one.

3.3. Higher Resolution and Discussions

In the previous section, we discussed a fast implementation algorithm utilizing two low-rank projections under the framework of the proposed projected atomic norm. Actually, since the atomic norm theory is limited by the conditions of the minimum angle distances (stated in Proposition 1), it is usually substituted by an atomic \uparrow_0 norm [22,39], which will be further transformed into the low-rank problem for optimization.

We define a projected atomic \uparrow_0 norm via the projected atom set $\mathcal{A}_{\mathcal{P}}$, proposed in Section 3.1 as

$$\|\mathcal{P}(\mathbf{X})\|_{\mathcal{A}_{\mathcal{P}},0} = \inf\left\{I \left| \mathcal{P}(\mathbf{X}) = \sum_{i}^{I} c_{i} \mathbf{A}_{\mathcal{P}}(\mathbf{\Phi}_{i}), \ \mathbf{A}_{\mathcal{P}}(\mathbf{\Phi}) \in \mathcal{A}_{\mathcal{P}}\right.\right\}$$
(42)

Similarly, an approximation of $\|\mathcal{P}(\mathbf{X})\|_{\mathcal{A}_{\mathcal{P}},0}$ is allowed as

$$\begin{aligned} \|\mathcal{P}(\mathbf{X})\|_{\mathcal{A}_{\mathcal{P}},0} &= \min_{\mathbf{X},\mathbf{u}_{b},\mathbf{u}_{d}} \left\{ \frac{1}{2} rank(T(\mathbf{u}_{b})) + \frac{1}{2} rank(T(\mathbf{u}_{d})) \right\} \\ &\qquad s.t. \begin{bmatrix} T(\mathbf{u}_{d}) & \mathcal{P}(\mathbf{X})^{H} \\ \mathcal{P}(\mathbf{X}) & T(\mathbf{u}_{b}) \end{bmatrix} \succcurlyeq \mathbf{0}, \ \left\|\mathbf{Y}^{0} - \mathbf{X}\right\|_{F}^{2} \leq \eta^{2} \end{aligned} \tag{43}$$

Unfortunately, the discontinuous problem (43) above is NP-hard. Although the D-AP algorithm proposed earlier can be exploited to solve the low-rank problem, it lacks stability in some case with higher resolution requirements. In order to obtain a more accurate solution, many smooth surrogate functions which simulate the characters of the rank are then adopted for the objective function of (43) instead [40]. Here, we consider two concave trace functions given by

$$f_1(T(\mathbf{u}_b)) = trace\Big((T(\mathbf{u}_b) + \varepsilon_b \mathbf{I})^{-1}\Big), \ f_2(T(\mathbf{u}_d)) = trace\Big((T(\mathbf{u}_d) + \varepsilon_d \mathbf{I})^{-1}\Big)$$
(44)

as the approximations of $rank(T(\mathbf{u}_b))$ and $rank(T(\mathbf{u}_d))$, where ε_b and ε_d are the parameters introduced to avoid the appearances of zero matrices. We then find the local optimum for the new programming problem via the majorization–minimization (MM) algorithm [41], which is an iterative method with the following decoupled weighted optimization problem for each iteration:

$$\min_{\mathbf{u}_{b,j+1},\mathbf{u}_{d,j+1},\mathcal{P}(\mathbf{X})_{j+1}} \left\{ \frac{1}{2} trace \left(\mathbf{W}_{b,j} T(\mathbf{u}_{b}) \right) + \frac{1}{2} trace \left(\mathbf{W}_{d,j} T(\mathbf{u}_{d}) \right) \right\}
s.t. \begin{bmatrix} T(\mathbf{u}_{d}) \quad \mathcal{P}(\mathbf{X})^{H} \\ \mathcal{P}(\mathbf{X}) \quad T(\mathbf{u}_{b}) \end{bmatrix} \geq \mathbf{0}, \left\| \mathbf{Y}^{0} - \mathbf{X} \right\|_{F}^{2} \leq \eta^{2}$$
(45)

with the weighting functions $\mathbf{W}_{b,j} = trace\left(\left(T\left(\mathbf{u}_{b,j}\right) + \varepsilon_{b,j}\mathbf{I}\right)^{-1}\right)$ and $\mathbf{W}_{d,j} = trace\left(\left(T\left(\mathbf{u}_{d,j}\right) + \varepsilon_{d,j}\mathbf{I}\right)^{-1}\right)$. Then, problem (43) can be implemented via iterative optimization (45) until the accuracy condition is met, and this algorithm is termed decoupled projected reweighted atomic norm minimization (D-PRAM), following the same naming convention seen in [22].

The reason for mentioning the above problem in this paper is that we are surprised to find that our proposed projected atom set can not only handle the angle-ambiguity problem under the framework of ANM, but also has a better resolution performance under the framework of atomic \uparrow_0 norm minimization than the existing 2D reweighting atomic norm algorithm induced by a vectorized atom set [22–24]. The vectorized 2D atomic norm optimization problem adopts an atom set given by

$$\mathcal{A}_{V} = \left\{ \mathbf{a}_{M}^{*}(\varphi) \otimes \mathbf{a}_{N}(\theta), \theta \in [-\pi, \pi], \varphi \in [-\pi, \pi] \right\}$$

= $\left\{ \mathbf{a}_{V}(\Theta), \Theta \in [-\pi, \pi] \times [-\pi, \pi] \right\}$ (46)

via vectorization of the matrix set A_M , and its atomic norm, defined similarly to the way stated earlier except for the induced atom, can be computed via the following SDP problem:

$$\|\operatorname{vec}(\mathbf{X})\|_{\mathcal{A}_{v}} = \min_{\mathbf{u}, \operatorname{vec}(\mathbf{X})} \left\{ \frac{1}{2\sqrt{NM}} \operatorname{trace}(T(\mathbf{u})) + \frac{1}{2}t \right\}$$

s.t. $\begin{bmatrix} t & \operatorname{vec}(\mathbf{X})^{H} \\ \operatorname{vec}(\mathbf{X}) & T(\mathbf{u}) \end{bmatrix} \succeq \mathbf{0}, \left\|\operatorname{vec}(\mathbf{Y}^{0}) - \operatorname{vec}(\mathbf{X})\right\|_{2}^{2} \leq \gamma^{2}$ (47)

where $T(\mathbf{u})$ is a $NM \times NM$ block Toeplitz matrix. For ease of comparison, this 2D atomic norm algorithm is termed as vectorized atomic norm minimization (vecANM), and its corresponding reweighted algorithm is referred to as vectorized reweighted atomic norm minimization (vecRAM) according to [22,24], which is a problem optimizing the atomic \uparrow_0 norm induced by the vectorized atom set A_V and similarly formulates the following SDP for each iteration:

$$\min_{\mathbf{u}_{j+1}, vec(\mathbf{X})_{j+1}} \left\{ \frac{1}{2} trace(\mathbf{W}_j T(\mathbf{u})) + \frac{1}{2}t \right\}$$

$$s.t. \begin{bmatrix} t & vec(\mathbf{X})^H \\ vec(\mathbf{X}) & T(\mathbf{u}) \end{bmatrix} \geq \mathbf{0}, \left\| vec(\mathbf{Y}^0) - vec(\mathbf{X}) \right\|_2^2 \leq \gamma^2$$
(48)

where \mathbf{W}_j denotes the weighting function accordingly. Note that (48) here reweights to the single term of the objective function, while the proposed D-PRAM adopts a decoupled reweighting strategy, and we will show that the decoupled strategy achieves a more enhanced performance in terms of resolution in the subsequent simulation. Additionally, the computational load of our proposed algorithms has also been relatively reduced compared with these vectorized 2D atomic norm optimization algorithms. According to the analysis in the previous subsection, the complexity of the vecANM problem (47) is predominantly influenced by the size of the PSD cone, which can reach $\mathcal{O}((NM)^{3.5})$. In contrast, the proposed D-PANM algorithm has a PSD constraint of smaller size $KL \times \overline{LK}$, thereby reducing the complexity to $\mathcal{O}((NM)^{3.5}/2^{3.5})$. For the reweighted versions, i.e., vecRAM and D-PRAM, the computational complexities are $\mathcal{O}(k_{R,1}(NM)^{3.5})$ and $\mathcal{O}(k_{R,2}(NM)^{3.5}/2^{3.5})$, respectively, given the iterations $k_{R,1}$ and $k_{R,2}$. Notably, this implies that D-PRAM holds more noticeable advantages regarding computational efficiency, particularly in scenarios necessitating multiple iterations.

4. Numerical Simulations

We present a series of numerical simulations to illustrate the performance of the proposed D-PANM algorithm and its derived results compared with those of the existing ANM algorithms. All of the methods involved in this section are illustrated briefly in Table 4, along with the computational complexity. Each experiment was based on single-snapshot array data, the signals were generated independently with the same constant magnitudes σ_s and the phases satisfied a randomly uniform distribution from $-\pi$ to π such that $E(\mathbf{CC}^H) = \sigma_s^2 \mathbf{I}$. The array's signal-to-noise radio (SNR) was set to be $10 \log_{10}(\sigma_s^2/\sigma_n^2)$, where σ_n^2 is the covariance of the Gaussian noise. Specifically, the number of signals *I* was not known as a priori, and only the information about the magnitude of noise $NM\sigma_n^2$ was given.

Algorithm	Illustration	Complexity (N = M)
DeANM	Atomic norm minimization induced by matrix atom set [26]	$O(2^{3.5}N^{3.5})$
vecANM	Atomic norm minimization induced by vectorized atom set [23]	$\mathcal{O}(N^7)$
vecRAM	Reweighted version of vecANM [22,24]	$\mathcal{O}(k_{R,2}N^7)$
D-PANM	Proposed decoupled projected atomic norm minimization	$\mathcal{O}\left(\frac{1}{2^{3.5}}N^7\right)$
D-PRAM	Reweighted version of D-PANM	$\mathcal{O}\left(k_{R,1}\frac{1}{2^{3.5}}N^7\right)$
D-AP	Proposed fast implementation of D-PANM with decoupled alternating projections	$\mathcal{O}\left(k_P \frac{19}{2^6} N^6\right)$

Table 4. A brief illustration of proposed algorithms and the comparison ones. ($k_{R,1}$, $k_{R,2}$, and k_P denote the maximum number of iterations in related algorithms).

First, we provide an intuitive example to verify the accuracy of the proposed algorithms when the signals share common angles in both the azimuth and elevation dimensions. In particular, consider a URA with N = M = 9, and I = 7 narrowband signals randomly generated with $\sigma_s^2 = 1$ and directions $(-37.1^\circ, 81.2^\circ)$, $(-19.6^\circ, 34.9^\circ)$, $(5.2^\circ, 19.2^\circ)$, $(5.2^\circ, 70.6^\circ)$, $(20.1^\circ, 55.3^\circ)$, $(38.5^\circ, 19.2^\circ)$, and $(57.3^\circ, 55.3^\circ)$. Assume single-snapshot array data polluted by the noise with $\sigma_n^2 = 1/10^{\text{SNR}/10}$, where SNR = 30dB, and let K = L = (N+1)/2 = 5. Then, 500 Monte Carlo experiments were carried out to estimate the DOAs for each of the algorithms DeANM, vecANM, and D-PANM. Figure 2a–c show the results, respectively, all of which were implemented by the SDPT3 solver. It is apparent that DeANM could hardly obtain correct angles, and vecANM provided incorrect DOA pairs in some runs in such a multi-angle ambiguity scenario. However, the proposed D-PANM with an automatic pairing criterion demonstrated a strong performance. Furthermore, the estimated results of the fast algorithm D-AP proposed in Section 3.2 are described in Figure 2d, which had almost the same recovery performance as the D-PANM did using an SDPT3 solver, but with a fairly small amount of computation. We will give a detailed computational analysis in the following simulation.





Figure 2. Cont.



Figure 2. The estimated DOAs in the angle-ambiguity scenario with N = M = 9, K = L = 5, SNR = 30 dB, and the number of signals I = 7: (a) DeANM; (b) vecANM; (c) proposed D-PANM; (d) proposed D-AP.

Second, we investigate the performance of the algorithms in terms of resolution, adding the reweighted versions of the corresponding algorithms. Let N = M = 9 and the mapping parameters K = L = 5. Suppose there were I = 2 sources impinging onto the array, one of which was set to be the reference with DOAs $(30^{\circ}, 30^{\circ})$, while the other one gradually moved away by a directional step of 0.5° on a scale of 1° to 10° in both the azimuth and elevation dimensions, and that a random fluctuation within a range of $\pm 0.025^{\circ}$ was allowed for each trial. We empirically set the initial reweighting parameters $\varepsilon = \lambda_{max}$ for each algorithm uniformly and reduced them by half in each iteration, where λ_{max} denotes the maximum eigenvalue of the corresponding Toeplitz matrix in the first iteration. One hundred Monte Carlo experiments were carried out, and their results are shown in Figure 3, which compares D-PANM and vecANM with their reweighted algorithms, respectively, where RMSE is the root mean squared error. The results show that D-PANM and vecANM had the same resolution performance, being able to exactly recover two sources mutually separated by 8.5° , and verified the condition for the minimum angle distance argued in Proposition 1: $\sin(8.5^\circ) \ge \min(1.19/(N-1), 1.19/(M-1))$. Meanwhile, D-PRAM (the reweighted version of D-PANM) could efficiently distinguish the sources with a 3° distance while vecRAM (the reweighted version of vecANM) can only distinguish a 6° one, which is to say that our decoupled reweighting strategy adopted in D-PRAM is superior to the one utilized in vecRAM. In addition, the resolution performance of D-AP is also presented in Figure 3 with a deletion of failed runs, denoted by D-APDeleted. Note that, although the curve shows that D-AP is almost unlimited in resolution, it becomes increasing unstable as the angle distance grows closer. To verify this, the success rates of our algorithms D-PRAM and D-AP are shown in Figure 4, where the success rate is set to be the ratio of runs with RMSE $< 10^{-2}$. The results indicate that the performance of D-AP is not stable enough when the angle distances are relatively closer, even though it is generally considered that the implementation method based on AP is capable of handling the low-rank problem. However, its stability becomes guaranteed starting with a larger angle interval of 13°, and the computational complexity benefit remains attractive. By contrast, D-PRAM has a 100% success rate as long as the sources are separated within the minimum allowable range.



Figure 3. Comparison of resolutions for two signals with N = M = 9 and K = L = 5: (a) azimuth distance $\Delta \theta_i$; (b) elevation distance $\Delta \varphi_i$.



Figure 4. Success rates of DOA estimation with respect to the angle interval: (**a**) azimuth distance $\Delta \theta_i$; (**b**) elevation distance $\Delta \varphi_i$.

Furthermore, the mean computational times of the algorithms involved in this section are provided in Figure 5, obtained by a computer with an Inter i7-7700K 4.20 GHz CPU. As shown in the figure, the performance of D-PANM showed improvement in terms of reducing the computational amount compared with vecANM, meanwhile, D-PRAM based on the decoupled reweighted strategy even displayed the same computational complexity as vecANM after multiple iterations, but had a greatly enhanced resolution. Of course, D-AP, as a fast implementation of D-PANM, exhibited reductions orders of magnitude larger in terms of computational complexity. Additionally, it is interesting to find that the curve of DeANM was higher than that of D-AP even though the theoretical computational complexity of DeANM was lower than that of D-AP, which was most likely due to the differences in implementation methods between different algorithms and the time-consuming nature of scheduling the SDPT3 solver. However, there was a downward trend for DeANM as the number of sensors increased.

Finally, we carried out 100 Monte Carlo experiments to validate the performance of the proposed algorithms in the presence of noise. In this simulation, we considered the scenario without angle ambiguity for the convenience of comparing our algorithms and the DeANM algorithm, which can hardly deal with the data when the signals share the same angles in either the azimuth or elevation dimension. In particular, we take a group of signals with distinct directions $(18^\circ, 32^\circ)$, $(38^\circ, 70^\circ)$, and $(50^\circ, 55^\circ)$, and let N = M = 11 and K = L = 6. Assume the array receiving data are polluted by the Gaussian noise, with the SNR varying from 0 to 30dB, and Figure 6 compares the RMSEs of different algorithms with respect to SNR, where CRB denotes the Cramér–Rao bound. The results show that the RMSE curve of the proposed D-PANM kept the same level as that of vecANM, but was lower than that of DeANM, i.e., although the noise constraint in D-PANM was still

better, implying that the smoothing projection process helped to improve the anti-noise performance. In addition, the proposed fast implementation algorithm D-AP exhibited good performance under a larger SNR, but began to degenerate as SNR decreased, mainly due to the influence of the parameters.



Figure 5. Mean computational time vs. the number of sensors.



Figure 6. Noise performance: RMSE vs. SNR with N = M = 11 and K = L = 6.

5. Conclusions

In this study, we developed a 2D DOA estimation method in angle-ambiguity scenarios based on the framework of ANM, and a valid atom set, i.e., the projected atom set $\mathcal{A}_{\mathcal{P}}$, was then constructed taking full advantage of the phase elimination property of the equivalent SDP problem, along with the smoothing idea. Indeed, the D-PANM algorithm induced by it fully retained the benefits of ANM. That is, it is capable of handling the array data with limited snapshots, even a single one, without knowing the source number. Moreover, it not only accurately yielded the estimation of DOA with automatic pairing when two or more signals impinged from the same directions, but also was more robust to the noise compared with ANM of the decoupled type. In addition, the proposed algorithm had a lower computational load compared with the existing vecANM, and its resolution with a decoupled reweighted strategy was superior to that of the comparison algorithms. Furthermore, a D-AP algorithm was also utilized to accelerate the implementation of our problem, which proved effective under many conditions. However, the proposed method is based on the ideal array manifold matrix and is relatively sensitive to signals with vastly different power levels, so a more robust algorithms in a real test environment should be investigated in the future. Moreover, we will consider incorporating the processing method with multiple-snapshot data. Lastly, a characteristic of ANM is that it can deal with incomplete data, so a sparse array or an array with missing elements is also a direction for future research.

Author Contributions: Conceptualization, M.L. and C.D.; methodology, M.L. and Y.D.; software, M.L.; validation, M.L., Y.D. and G.Z.; formal analysis, Y.D.; data curation, G.Z.; writing—original draft preparation, M.L.; writing—review and editing, M.L.; visualization, Y.D.; supervision, C.D.; funding acquisition, Y.D. and C.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 61901332.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Proof of (20) in Proposition 1. Let us introduce the following theorem first before proving this equivalence:

Theorem A1 ([31]). Suppose a 1D multiple measurement vector (MMV) $\mathbf{Y} \in \mathbb{C}^{N \times L}$ is expressed as $\mathbf{Y} = \sum_{i=1}^{I} |c_i| \mathbf{a}(\theta_i) \phi_i$, where $\mathbf{a}(\theta_i)$ denotes the steering vector, $|c_i| > 0$ denotes the amplitude of the receiving signal c_i , and the matrix ϕ_i is the phase of c_i with $\phi_i = |c_i|^{-1} c_i$ and $||\phi_i|| = 1$. Then,

the receiving signal c_i , and the matrix ϕ_i is the phase of c_i with $\phi_i = |c_i| - c_i$ and $||\phi_i|| = 1$. Then, the atom set of MMV data can be defined as

$$\mathcal{A}_{MM} = \{ \mathbf{a}(\theta)\phi, \|\phi\| = 1 \}$$
(A1)

and its atomic norm induced by A_{MM} is given by

$$\|\mathbf{Y}\|_{\mathcal{A}_{MM}} = \inf\left\{\sum_{i}^{I} |c_{i}| \left| \mathbf{Y} = \sum_{i}^{I} |c_{i}| \mathbf{a}(\theta_{i})\phi_{i}, \ \mathbf{a}(\theta_{i})\phi_{i} \in \mathcal{A}_{MM} \right. \right\}$$
(A2)

which equals the following SDP:

$$\|\mathbf{Y}\|_{\mathcal{A}_{MM}} = \min_{\mathbf{W},\mathbf{u}} \left\{ \frac{1}{2\sqrt{N}} [trace(\mathbf{W}) + trace(T_1(\mathbf{u}))] \right\}$$

s.t.
$$\begin{bmatrix} \mathbf{W} & \mathbf{Y}^H \\ \mathbf{Y} & T_1(\mathbf{u}) \end{bmatrix} \succeq \mathbf{0}$$
 (A3)

where $T_1(\mathbf{u})$ denotes the 1D Toeplitz matrix with \mathbf{u} as its first column, and \mathbf{W} is an introduced optimization variable. In addition, the right side of (A3) is denoted by $SDP_M(\mathbf{Y})$.

Next, denote the right side of (20) by $SDP(\mathcal{P}(\mathbf{X}))$, and show $\|\mathcal{P}(\mathbf{X})\|_{\mathcal{A}_{\mathcal{P}}} = SDP(\mathcal{P}, (\mathbf{X}))$ by proving $SDP(\mathcal{P}(\mathbf{X}) \leq \|\mathcal{P}(\mathbf{X})\|_{\mathcal{A}_{\mathcal{P}}}$ and $SDP(\mathcal{P}(\mathbf{X})) \geq \|\mathcal{P}(\mathbf{X})\|_{\mathcal{A}_{\mathcal{P}}}$.

On the one hand, we prove $SDP(\mathcal{P}(\mathbf{X}) \leq ||\mathcal{P}(\mathbf{X})||_{\mathcal{A}_{\mathcal{P}}}$. Suppose that

$$(\mathcal{P}\mathbf{X}) = \sum_{i}^{I} c_k \mathcal{P}(\mathbf{A}_M(\Theta_i)) = \sum_{i}^{I} c_k \mathbf{A}_{\mathcal{P}}(\Theta_i) = \sum_{i}^{I} |c_i| \mathbf{b}_{KL} \left(\Theta_i, \frac{1}{2}\xi_i\right) \mathbf{d}_{LK}^{H} \left(\Theta_i, \frac{1}{2}\xi_i\right)$$
(A4)

is given, then let $\mathbf{u}_b = \sum_{i}^{I} |c_i| \mathbf{b}_{KL}(\Theta_i, 0)$ and $\mathbf{u}_d = \sum_{i}^{I} |c_i| \mathbf{d}_{\overline{LK}}(\Theta_i, 0)$. We have

$$T(\mathbf{u}_{b}) = \sum_{i}^{I} \sqrt{\frac{\overline{KL}}{KL}} |c_{i}| \mathbf{b}_{KL}(\Theta_{i}, 0) \mathbf{b}_{KL}^{H}(\Theta_{i}, 0) = \sum_{i}^{I} \sqrt{\frac{\overline{KL}}{KL}} |c_{i}| \mathbf{b}_{KL}\left(\Theta_{i}, \frac{1}{2}\xi_{i}\right) \mathbf{b}_{KL}^{H}\left(\Theta_{i}, \frac{1}{2}\xi_{i}\right)$$
(A5)
$$T(\mathbf{u}_{d}) = \sum_{i}^{I} \sqrt{\frac{\overline{KL}}{\overline{KL}}} |c_{i}| \mathbf{d}_{\overline{LK}}(\Theta_{i}, 0) \mathbf{d}_{\overline{LK}}^{H}(\Theta_{i}, 0) = \sum_{i}^{I} \sqrt{\frac{\overline{KL}}{\overline{KL}}} |c_{i}| \mathbf{d}_{\overline{LK}}\left(\Theta_{i}, \frac{1}{2}\xi_{i}\right) \mathbf{d}_{\overline{LK}}^{H}\left(\Theta_{i}, \frac{1}{2}\xi_{i}\right)$$
(A6)

Then,

$$\begin{bmatrix} T(\mathbf{u}_d) & \mathcal{P}(\mathbf{X})^H \\ \mathcal{P}(\mathbf{X}) & T(\mathbf{u}_b) \end{bmatrix} = \sum_i^I \frac{1}{\sqrt{KL\overline{KL}}} |c_i| \begin{bmatrix} \sqrt{KL} \mathbf{d}_{\overline{LK}} \left(\Theta_i, \frac{1}{2}\xi_i\right) \\ \sqrt{\overline{KL}} \mathbf{b}_{KL} \left(\Theta_i, \frac{1}{2}\xi_i\right) \end{bmatrix} \begin{bmatrix} \sqrt{KL} \mathbf{d}_{\overline{LK}} \left(\Theta_i, \frac{1}{2}\xi_i\right) \\ \sqrt{\overline{KL}} \mathbf{b}_{KL} \left(\Theta_i, \frac{1}{2}\xi_i\right) \end{bmatrix}^H$$
(A7)

is positive semidefinite (PSD), since $|c_i| \ge 0$ for all *i* hold. In addition, each diagonal element of $T(\mathbf{u}_b)$ is $\sqrt{\frac{KL}{KL}}\sum_{i}^{I} |c_i|$, and the diagonal elements of $T(\mathbf{u}_d)$ is $\sqrt{\frac{KL}{KL}}\sum_{i}^{I} |c_i|$, so the following equations hold:

$$\frac{1}{\sqrt{KL\overline{KL}}}trace(T(\mathbf{u}_b)) = \sum_{i}^{I} |c_i|, \ \frac{1}{\sqrt{KL\overline{KL}}}trace(T(\mathbf{u}_d)) = \sum_{i}^{I} |c_i|.$$
(A8)

Therefore, $SDP(\mathcal{P}(\mathbf{X}))$, as the minimum point, has the following relationship:

$$SDP(\mathcal{P}(\mathbf{X})) \le \frac{1}{2\sqrt{KL\overline{KL}}}(trace(T(\mathbf{u}_b)) + trace(T(\mathbf{u}_d))) = \sum_{i}^{l} |c_i|$$
 (A9)

i.e., $SDP(\mathcal{P}(\mathbf{X})) \leq \|\mathcal{P}(\mathbf{X})\|_{\mathcal{A}_{\mathcal{P}}}.$

On the other hand, we prove $SDP(\mathcal{P}(\mathbf{X})) \ge ||\mathcal{P}(\mathbf{X})||_{\mathcal{A}_{\mathcal{P}}}$ from the results of Theorem A1.

Given

and

$$\mathcal{P}(\mathbf{X}) = \sum_{i}^{I} |c_{i}| \mathbf{b}_{KL} \left(\Theta_{i}, \frac{1}{2} \xi_{i} \right) \mathbf{d}_{LK}^{H} \left(\Theta_{i}, \frac{1}{2} \xi_{i} \right)$$

$$= \sum_{i}^{I} |c_{i}| \left\| \mathbf{d}_{LK}^{H} \left(\Theta_{i}, \frac{1}{2} \xi_{i} \right) \right\| \mathbf{b}_{KL} \left(\Theta_{i}, \frac{1}{2} \xi_{i} \right) \frac{\mathbf{d}_{LK}^{H} \left(\Theta_{i}, \frac{1}{2} \xi_{i} \right)}{\left\| \mathbf{d}_{LK}^{H} \left(\Theta_{i}, \frac{1}{2} \xi_{i} \right) \right\|}$$

$$= \sum_{i}^{I} \sqrt{KL} |c_{i}| \mathbf{b}_{KL} \left(\Theta_{i}, \frac{1}{2} \xi_{i} \right) \frac{\mathbf{d}_{LK}^{H} \left(\Theta_{i}, \frac{1}{2} \xi_{i} \right)}{\left\| \mathbf{d}_{LK}^{H} \left(\Theta_{i}, \frac{1}{2} \xi_{i} \right) \right\|}$$
(A10)

Then, $\left\| \frac{\mathbf{d}_{L\bar{k}}^{H}(\Theta_{i},\frac{1}{2}\xi_{i})}{\|\mathbf{d}_{L\bar{k}}^{H}(\Theta_{i},\frac{1}{2}\xi_{i})\|} \right\| = 1$, and $\mathbf{b}_{KL}\left(\Theta_{i},\frac{1}{2}\xi_{i}\right) \frac{\mathbf{d}_{L\bar{k}}^{H}(\Theta_{i},\frac{1}{2}\xi_{i})}{\|\mathbf{d}_{L\bar{k}}^{H}(\Theta_{i},\frac{1}{2}\xi_{i})\|}$ has the same structure as $\mathbf{a}(\theta)\phi$, which can be considered as the atom of \mathcal{A}_{MM} . Thus, $\|\mathcal{P}(\mathbf{X})\|_{\mathcal{A}_{MM}} = \sum_{i=1}^{L} \sqrt{\overline{KL}}|c_{i}|$,

$$SDP_{M}(\mathcal{P}(\mathbf{X})) = \min_{\mathbf{W}, \mathbf{u}_{b}} \frac{1}{2\sqrt{KL}} [trace(\mathbf{W}) + trace(T(\mathbf{u}_{b}))]$$

$$s.t. \begin{bmatrix} \mathbf{W} & \mathcal{P}(\mathbf{X})^{H} \\ \mathcal{P}(\mathbf{X}) & T(\mathbf{u}_{b}) \end{bmatrix} \succeq \mathbf{0} \qquad (A11)$$

Therefore, one has

$$\|\mathcal{P}(\mathbf{X})\|_{\mathcal{A}_{\mathcal{P}}} = \sum_{i=1}^{I} |c_i| = \frac{1}{\sqrt{\overline{KL}}} \|\mathcal{P}(\mathbf{X})\|_{\mathcal{A}_{MM}}.$$
 (A12)

Combining (20) with (A11), we know that $SDP(\mathcal{P}(\mathbf{X}))$ has the same objective function as $\frac{1}{\sqrt{KT}}SDP_M(\mathcal{P}(\mathbf{X}))$ except for an additional Toeplitz constraint $T(\mathbf{u}_d)$. Thus,

$$SDP(\mathcal{P}(\mathbf{X})) \ge \frac{1}{\sqrt{\overline{KL}}}SDP_M(\mathcal{P}(\mathbf{X})) = \frac{1}{\sqrt{\overline{KL}}} \|\mathcal{P}(\mathbf{X})\|_{\mathcal{A}_{MM}}.$$
 (A13)

According to (A12) and (A13), $SDP(\mathcal{P}(\mathbf{X})) \ge \|\mathcal{P}(\mathbf{X})\|_{\mathcal{A}_{\mathcal{P}}}$. In conclusion, we obtain $\|\mathcal{P}(\mathbf{X})\|_{\mathcal{A}_{\mathcal{P}}} = SDP(\mathcal{P}(\mathbf{X}))$.

References

- Bardhan, S.; Jacob, S. Experimental Observations of Direction of Arrival Estimation Algorithms in a Tank environment for Sonar Applications. In Proceedings of the IEEE International Symposium on Ocean Electronics (IEEE Explore), Kochi, India, 18–20 November 2015.
- Ayllón, D.; Gil-Pita, R.; Rosa-Zurera, M.; Kim, H. Real-time multiple DOA estimation of speech sources in wireless acoustic sensor networks. In Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, Australia, 19–24 April 2015.
- 3. Dudczyk, J.; Rybak, Ł. Application of Data Particle Geometrical Divide Algorithms in the Process of Radar Signal Recognition. *Sensors* **2023**, 23, 8183. [CrossRef]
- 4. Ciuonzo, D. On Time-Reversal Imaging by Statistical Testing. IEEE Signal Process. Lett. 2017, 24, 1024–1028. [CrossRef]
- 5. Devaney, J.A. Time reversal imaging of obscured targets from multistatic data. *IEEE Trans. Antennas Propag.* 2005, 53, 1600–1610. [CrossRef]
- Sakarya, F.A.; Hayes, M.W. Estimating 2-D DOA angles using nonlinear array configurations. *IEEE Trans. Signal Process.* 1995, 43, 2212–2216. [CrossRef]
- Ghobadzadeh, A.; Adve, R. Low-complexity 2D root-MUSIC pairing for an L-shaped array. In Proceedings of the 2017 IEEE Radar Conference (RadarConf), Seattle, WA, USA, 8–12 May 2017; pp. 0957–0962.
- Ciuonzo, D.; Romano, G. Performance analysis of time-reversal MUSIC. *IEEE Trans. Signal Process.* 2015, 63, 2650–2662. [CrossRef]
- Cheng, C.; Hua, Y. Performance analysis of the MUSIC and Pencil-MUSIC algorithms for diversely polarized array. In Proceedings of the Acoustics, Speech, & Signal Processing, on IEEE International Conference IEEE, Adelaide, Australia, 19–22 April 1994.
- 10. Nemirovsky, S.; Doron, M.A. Sensitivity of MUSIC and Root-MUSIC to gain calibration errors of 2D arbitrary array configuration. In Proceedings of the Sensor Array & Multichannel Signal Processing Workshop IEEE, Barcelona, Spain, 18–21 July 2004.
- 11. Van der Veen, A.J.; Ober, P.B. Azimuth and elevation computation in high resolution DOA estimation. *Signal Process. IEEE Trans.* **1992**, 40, 1828–1832. [CrossRef]
- 12. Rauhut, H.; Schnass, K.; Vandergheynst, P. Compressed Sensing and Redundant Dictionaries. *IEEE Trans. Inf. Theory* 2008, 54, 2210–2219. [CrossRef]
- 13. Mallat, S.G.; Zhang, Z. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process. A Publ. IEEE Signal Process. Soc.* **1993**, *12*, 41. [CrossRef]
- 14. Wang, C.; Hu, J.; Zhang, Q.; Yuan, X. An Efficient 2D DOA Estimation Algorithm Based on OMP for Rectangular Array. *Electronics* 2023, *12*, 1634. [CrossRef]
- Yang, Z.; Xie, L.; Zhang, C. Off-Grid Direction of Arrival Estimation Using Sparse Bayesian Inference. *IEEE Trans. Signal Process.* 2013, *61*, 38–43. [CrossRef]
- 16. Jagannath, R.; Hari, K.V.S. Block Sparse Estimator for Grid Matching in Single Snapshot DOA Estimation. *IEEE Signal Process*. *Lett.* **2013**, *20*, 1038–1041. [CrossRef]
- 17. Chen, P.; Cao, Z.; Chen, Z.; Yu, C. Sparse DOD/DOA Estimation in a Bistatic MIMO Radar with Mutual Coupling Effect. *Electronics* **2018**, *7*, 341. [CrossRef]
- 18. Candès, E.J.; Fernandez-Granda, C. Towards a Mathematical Theory of Super-resolution. arXiv 2013, arXiv:1203.5871. [CrossRef]
- 19. Chandrasekaran, V.; Rechr, B.; Parrilo, P.A.; Willsky, A.S. The Convex Geometry of Linear Inverse Problems. *arXiv* 2010, arXiv:1012.0621. [CrossRef]
- 20. Tang, G.; Bhaskar, B.N.; Shah, P.; Recht, B. Compressed Sensing Off the Grid. *IEEE Trans. Inf. Theory* 2013, 59, 7465–7490. [CrossRef]
- 21. Yang, Z.; Xie, L. On Gridless Sparse Methods for Line Spectral Estimation from Complete and Incomplete Data. *IEEE Trans. Signal Process.* **2015**, *63*, 3139–3153. [CrossRef]
- Yang, Z.; Xie, L. Enhancing Sparsity and Resolution via Reweighted Atomic Norm Minimization. *IEEE Trans. Signal Process.* 2016, 64, 995–1006. [CrossRef]

- Chi, Y.; Chen, Y. Compressive Two-Dimensional Harmonic Retrieval via Atomic Norm Minimization. *IEEE Trans. Signal Process.* 2015, 63, 1030–1042. [CrossRef]
- Yang, Z.; Xie, L.; Stoica, P. Vandermonde Decomposition of Multilevel Toeplitz Matrices with Application to Multidimensional Super-Resolution. *IEEE Trans. Inf. Theory* 2016, 62, 3685–3701. [CrossRef]
- 25. Xu, W.; Cai, J.F.; Mishra, K.V.; Cho, M.; Kruger, A. Precise Semidefinite Programming Formulation of Atomic Norm Minimization for Recovering d-Dimensional (D ≥ 2) Off-the-Grid Frequencies. In Proceedings of the 2014 Information Theory and Applications Workshop (ITA), San Diego, CA, USA, 9–14 February 2014.
- 26. Tian, Z.; Zhang, Z.; Wang, Y. Low-complexity optimization for two-dimensional direction-of-arrival estimation via decoupled atomic norm minimization. In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing IEEE, New Orleans, LA, USA, 5–9 March 2017.
- 27. Hua, Y. Estimating two-dimensional frequencies by matrix enhancement and matrix pencil. *IEEE Trans. Signal Process.* **1992**, 40, 2267–2280. [CrossRef]
- Wagner, M.; Gerstoft, P.; Park, Y. Gridless DOA Estimation via Alternating Projections. In Proceedings of the ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019.
- Grant, M. CVX: MATLAB Software for Disciplined Convex Programming; CVX Research: Austin, TX, USA, 2008; Available online: http://cvxr.com/cvx (accessed on 19 February 2024).
- Toh, K.C.; Todd, M.J.; Tütüncü, R.H. SDPT3—A MATLAB Software Package for Semidefinite Programming, Version 3.1. *Optim. Methods Softw.* 1999, 11, 545–581. Available online: https://blog.nus.edu.sg/mattohkc/softwares/sdpt3/ (accessed on 19 February 2024). [CrossRef]
- 31. Yang, Z.; Xie, L. Exact Joint Sparse Frequency Recovery via Optimization Methods. *IEEE Trans. Signal Process. A Publ. IEEE Signal Process. Soc.* 2016, 64, 5145–5157. [CrossRef]
- 32. Hua, Y.; Sarkar, T.K. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Trans Assp.* **1990**, *38*, 814–824. [CrossRef]
- Rouquette, S.; Najim, M. Estimation of Frequencies and Damping Factors by Two-Dimensional ESPRIT Type Methods. *IEEE Trans. Signal Process.* 2001, 49, 237–245. [CrossRef]
- Wang, Y.; Tian, Z. IVDST: A Fast Algorithm for Atomic Norm Minimization in Line Spectral Estimation. *IEEE Signal Process.* Lett. 2018, 25, 2870539. [CrossRef]
- 35. Zhang, F.Z. The Schur Complement and Its Applications; Springer: Greer, SC, USA, 2005.
- 36. Toh, K.; Yun, S. An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. *Pac. J. Optim.* **2010**, *6*, 615–640.
- Bauschke, H.H.; Borwein, J.M. On the convergence of von Neumann's alternating projection algorithm for two sets. *Set-Valued Anal.* 1993, 1, 185–212. [CrossRef]
- 38. Cai, T.; Zhang, A. ROP: Matrix recovery via rank-one projections. Ann. Stats 2013, 43, 102–138. [CrossRef]
- Mohimani, H.; Babaie-Zadeh, M.; Jutten, C. A Fast Approach for Overcomplete Sparse Decomposition Based on Smoothed L-0 Norm. *IEEE Trans. Signal Process.* 2009, 57, 289–301. [CrossRef]
- Fazel, M.; Hindi, H.; Boyd, S.P. Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices. In Proceedings of the 2003 American Control Conference, Denver, CO, USA, 4–6 June 2003.
- 41. Ying, S.; Babu, P.; Palomar, D.P. Majorization-Minimization Algorithms in Signal Processing, Communications, and Machine Learning. *IEEE Trans. Signal Process.* **2016**, *65*, 794–816.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.