

2D DOA Estimation of Coherent Signals Exploiting Forward-Backward Covariance Tensor

Saidur R. Pavel, *Student Member, IEEE*, Yimin D. Zhang, *Fellow, IEEE*,
and Shunqiao Sun, *Senior Member, IEEE*

Abstract—Coherent signals commonly arise in applications such as wireless communication and radar sensing, necessitating decorrelation before applying conventional direction-of-arrival (DOA) estimation methods. For two-dimensional (2D) sensor arrays, the covariance entries of the received signals form a tensor, which is rank-deficient when the impinging signals are coherent. This paper develops two novel decorrelation strategies to address the rank deficiency issue in 2D DOA estimation. The first strategy reconstructs a decorrelated tensor by rearranging a slice of the coherent covariance tensor and, to enhance dimensionality and increase the degrees of freedom (DOFs) of the array, integrating a backward slice. The second strategy employs forward backward spatial smoothing (FBSS) with an optimal sensor arrangement. In particular, we optimize four FBSS parameters: the numbers of overlapping subarrays along the X and Y axes, and the numbers of antennas in each subarray along the X and Y axes, in order to maximize the DOFs for a given array size. Furthermore, we derive a closed-form expression for the number of achievable DOFs. Simulation results confirm that the proposed methods outperform existing techniques in terms of DOF enhancement and DOA estimation accuracy.

Index Terms—Direction-of-arrival estimation, coherent signals, tensor reconstruction, spatial smoothing, degree of freedom.

I. INTRODUCTION

DIRECTION-of-arrival (DOA) estimation, which determines the spatial spectra of incoming electromagnetic waves, is a fundamental technology in sensor array signal processing with broad applications in, e.g., wireless communication, radar, radio astronomy, and remote sensing [1]–[7]. Various DOA estimation methods have been developed, including subspace-based approaches [8], [9], beamforming techniques [10]–[12], and sparsity-inducing methods [13]–[18]. Among these, subspace-based methods, such as MUltiple SIgnal Classification (MUSIC) [8] and Estimation of Signal Parameters via Rotational Invariant Techniques (ESPRIT) [9], are popularly used due to their ability to achieve high-resolution DOA estimation with low complexity. For linear arrays, these methods estimate signal DOAs by exploiting the eigenstructure of the covariance matrix. However, they

were originally developed to handle uncorrelated signals with a full-rank source covariance matrix. When coherent signals are present, these methods fail because the source covariance matrix becomes rank-deficient.

In practice, coherent signals often arise due to factors such as multipath propagation in wireless communications, low-angle reflections in radar sensing, and intelligent jamming in defense applications. In this case, estimating the DOAs of coherent signals is challenging because the covariance matrix becomes rank-deficient. Therefore, decorrelating the covariance matrix to restore its rank becomes essential before applying subspace-based DOA estimation methods. Several algorithms have been developed to handle coherent signals, including the generalized MUSIC [19] and those based on subspace-fitting [20], [21] and maximum likelihood [22]. However, these methods involve multi-dimensional searches, making them computationally expensive. The well-known spatial smoothing (SS) for uniform linear array (ULA) [23] mitigates signal coherence and restores the rank of covariance matrix by dividing the entire array into multiple overlapping subarrays and averaging their covariance matrices. However, this approach reduces the number of degrees of freedom (DOFs) to approximately half the number of sensors.

To address this limitation, forward-backward spatial smoothing (FBSS) is introduced in [24] to incorporate the complex conjugates of the backward subarrays alongside forward subarrays, thus increasing the number of available DOFs to two-thirds of the number of sensors. The work in [25] further analyzes the antenna size requirements for ULAs from a Hadamard product perspective, relating the array size to the number of sources, the rank of the source covariance matrix, and the coherence structure of the sources. To improve the angular resolution of coherent signals, improved spatial smoothing (ISS) was developed in [26], [27] based on quadratic spatial smoothing, which is then further improved by an enhanced spatial smoothing (ESS) method [28].

A more computationally efficient alternative to SS was developed in [29], where a single row of the covariance matrix is arranged into a Toeplitz structure. This approach restores the rank of the covariance matrix regardless of signal coherence and achieves DOFs similar to SS. The Toeplitz-based method is further improved in [30], [31] by expanding the dimension of the signal subspace through the utilization of both forward and backward vectors. Building on these ideas, [32] extends the Toeplitz-based approaches to detect mixed coherent and uncorrelated sources. By constructing multiple Toeplitz matrices from both the rows and columns of the

The work of S. R. Pavel and Y. D. Zhang was supported in part by the National Science Foundation (NSF) under Grant ECCS-2236023 and in part by the Air Force Office of Scientific Research (AFOSR) under Grant FA9550-23-1-0255. The work of S. Sun was supported in part by NSF under Grants CCF-2153386 and ECCS-2340029, and Alabama Transportation Institute (ATI).

S. R. Pavel and Y. D. Zhang are with the Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA 19122, USA (emails: pavel.saidur@temple.edu, ydzhang@temple.edu).

S. Sun is with the Department of Electrical and Computer Engineering, The University of Alabama, Tuscaloosa, AL 35487, USA (email: shunqiao.sun@ua.edu).

covariance matrix, this method achieves DOFs exceeding two-thirds of the number of sensors.

Most methods for handling coherent signals are designed for linear arrays, which detect only one-dimensional (typically azimuth) angles. Many practical applications require two-dimensional (2D) antenna arrays capable of capturing both azimuth and elevation angular information [33]–[36]. A 2D SS scheme was proposed in [37], where the upper bound on the required number of subarrays for estimating a given number of coherent sources was discussed. Extending this idea, [38] analyzed the necessary and sufficient conditions for detecting \tilde{K} coherent signals. This approach requires a URA of size $2\tilde{K} \times 2\tilde{K}$. The paper further introduced a minimal array based on an L -shaped antenna arrangement with L -shaped subarray grouping, which reduces the number of required antennas to $\tilde{K}^2 + 4\tilde{K} - 2$. An FBSS technique for a $3\tilde{K}/2 \times 3\tilde{K}/2$ URA [39] detects \tilde{K} coherent signals. In [40], a block Hankel matrix is formed from the observed data samples to enhance the rank and estimate two-dimensional frequencies, while [41] also employs block Hankel matrix formation to decorrelate the coherent covariance matrix, achieving improved performance compared with SS-based approaches.

Tensor modeling of 2D array signals effectively captures their inherent structural properties, and tensor decomposition techniques leverage these properties for signal DOA estimation. Commonly used tensor decomposition methods include canonical polyadic decomposition (CPD) [42], [43], Tucker decomposition [44], and high-order singular value decomposition (HOSVD) [45]. In [46], multiple-invariance sensor array processing is linked to parallel factor (PARAFAC) analysis for DOA estimation and identifiability analysis. Tensor-based approaches generally provide superior estimation performance over matrix-based techniques when handling signals with dimensionality greater than two, as they more effectively exploit the multi-dimensional structure of the data [47]–[49]. A coarray tensor DOA estimation approach is developed in [50] for a 2D coprime arrays. However, these methods mainly handle uncorrelated sources. To address coherent sources, [51] introduces a Tensor-MODE approach that leverages HOSVD for multi-dimensional harmonic retrieval in the presence of coherent signals, demonstrating improved performance over tensor eigenvector-based methods [52], which fail under coherent conditions. Tensor-based SS techniques have been investigated in [53], [54] for MIMO configurations, and in [49] for polarized vector-sensor URA arrays. However, these SS schemes require multiple tensor computations, leading to reduced decorrelation effectiveness and increased computational cost. To overcome these limitations, [55] proposes a decorrelation method that directly reconstructs the covariance tensor by exploiting its structural properties. This approach reconstructs the covariance tensor by arranging its slices into a Toeplitz structure, followed by CPD to estimate the steering vectors and signal DOAs. This concept was extended to sparse arrays for coherent [56] and mixed coherent and uncorrelated scenarios [57]. However, these methods have limited DOFs mainly due to two reasons: 1) the exclusive use of the forward covariance tensor, and 2) the relatively small dimension of the steering vectors corresponding to the four-dimensional (4D)

covariance tensor.

In this paper, we develop two decorrelation schemes to handle rank-deficient sample covariance tensors computed from coherent signals received at a URA. The first approach introduces a frowrad-backward tensor reconstruction (FBTR) strategy, in which a decorrelated covariance tensor is constructed by rearranging slices of the coherent covariance tensor. Specifically, a particular slice of the covariance tensor and its flipped conjugate counterpart are rearranged and effectively fused to produce the decorrelated tensor. This results in a covariance tensor with a higher dimension than the existing method reported in [55], thereby increasing the number of DOFs. Moreover, this strategy has lower computational complexity than SS-based methods as it only requires slicing and rearranging operations without additional arithmetic calculations. Furthermore, it achieves improved DOA estimation performance over SS-based methods and matrix based approaches by effectively exploiting the structural characteristics of multi-dimensional signals through CPD. The second scheme is based on SS, where multiple covariance matrices are computed from subarrays of the antenna array. Note that, for a URA, subarrays can be formed in both azimuth and elevation dimensions. We derive the optimal array configuration by determining the number of subarrays along each axis and the number of antennas in each subarray. These four parameters are optimized to minimize the number of antennas for a given number of DOFs using Karush–Kuhn–Tucker (KKT) conditions to obtain a real-valued solution. Since practical implementations require integer solutions, we develop an efficient optimization scheme that refines the real-valued solution through a local search. After determining the optimal array configuration, the covariance matrices along with their flipped and conjugated versions are computed from the subarrays and averaged to obtain a decorrelated covariance matrix. By combining the optimal array configuration with FBSS, our approach achieves a higher number of DOFs for 2D arrays compared to existing methods.

The main contributions of this paper are summarized as follows:

- We optimize the sensor placement in a URA for FBSS by jointly optimizing four key parameters, namely, the number of subarrays and the number of sensors in each subarray along both the X (azimuth) and Y (elevation) dimensions. The results maximize the achievable DOFs for a given number of antennas. To the best of our knowledge, this is the first work that introduces flexible array configurations that are required to achieve a specified number of DOFs while ensuring the required azimuth and elevation array apertures.
- Building on the optimized array arrangement, we propose two decorrelation strategies for handling coherent impinging signals: FBTR and FBSS. By integrating optimal array arrangement and FB processing, these methods offer a higher number of DOFs compared to existing approaches.
- We derive closed-form expressions for the number of detectable coherent signals under both decorrelation strategies. For FBTR, the number of DOFs is expressed in

terms of the number of antennas along the azimuth and elevation dimensions, while for FBSS, it is formulated based on the total number of antennas and the minimum number of antennas along either the azimuth or elevation axis required to maintain a minimum required aperture.

- Through DOF analysis, we demonstrate that the FBSS method, when combined with an optimal array arrangement, provides the highest number of DOFs for 2D coherent signal scenarios. On the other hand, the FBTR strategy yields superior DOA estimation performance by effectively exploiting the structural properties of high-dimensional signals through tensor decomposition, e.g., CPD, although it provides fewer DOFs. Furthermore, by matricizing the decorrelated covariance tensor obtained via FBTR, referred to as FBTR-mat, offers a computationally efficient decorrelation scheme that achieves DOFs comparable to FBSS with the optimal array.

The rest of the paper is organized as follows: Section II introduces the preliminaries, including notations and tensor operations. Section III presents the signal model. Section IV describes the structured tensor-based decorrelation strategy, while Section V presents the SS-based decorrelation strategy. Section VI provides an analysis of the achieved DOFs and optimal array designs. Section VII presents simulation results and, finally, Section VIII concludes the paper.

II. PRELIMINARIES

Notations: We use lower-case bold characters (e.g., \mathbf{a}), upper-case bold characters (e.g., \mathbf{A}), and upper-case calligraphic bold characters (e.g., \mathcal{A}) to denote vectors, matrices, and tensors, respectively. In particular, \mathbf{I} and \mathcal{I} respectively denote the identity matrix and the identity tensor of a proper dimension. $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^*$ respectively represent the transpose, Hermitian, and conjugation operation. In addition, \circ represents the outer product, \odot stands for the Hadamard product, and \otimes denotes the Khatri-Rao product. $\mathbb{E}[\cdot]$ stands for the statistical expectation operator. $\mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$ describes the complex space with the specified dimension, $[\cdot]_{\sqcup_i}$ indicates tensor concatenation along the i th dimension, and $\mathcal{A}_1 \times_{i_1}^{i_2} \mathcal{A}_2$ denotes the tensor contraction along the i_1 th dimension of \mathcal{A}_1 and the i_2 th dimension of \mathcal{A}_2 . $\text{Diag}(\cdot)$ denotes a diagonal matrix where the elements of a vector form the diagonal entries, while $\text{squeeze}(\cdot)$ denotes a matrix or tensor with the same elements as the input, but with dimensions of length 1 removed. $\mathbf{1}\{A\}$ denotes the indicator function, which equals to 1 if an event A is true and 0 otherwise. Finally, $\lfloor \cdot \rfloor$, $\lceil \cdot \rceil$, and $\text{round}(\cdot)$ represents the floor and ceiling operations, and rounding to the nearest integer respectively, on a scalar.

Tensor reshaping: For an N -dimensional tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$, $\langle \mathcal{A} \rangle_{\mathbb{P}_1, \dots, \mathbb{P}_J}$ reshapes \mathcal{A} into a J -dimensional tensor of size $\prod_{p_1 \in \mathbb{P}_1} I_{p_1} \times \prod_{p_2 \in \mathbb{P}_2} I_{p_2} \times \dots \times \prod_{p_J \in \mathbb{P}_J} I_{p_J}$, with \mathbb{P}_j being a partition of $\{1, 2, \dots, N\}$. In particular, the element $\mathcal{A}(i_1, \dots, i_N)$ is mapped to the (k_1, \dots, k_J) th element of the reshaped tensor as $\langle \mathcal{A} \rangle_{\mathbb{P}_1, \dots, \mathbb{P}_J}(k_1, \dots, k_J) = \mathcal{A}(i_1, \dots, i_N)$, where $k_j = 1 + \sum_{r=1}^{\lfloor \mathbb{P}_j \rfloor} (i_{p_{jr}} - 1) \prod_{s=1}^{r-1} I_{p_{js}}$.

CPD and tensor canonical polyadic (CP) rank: CPD factorizes a high-order tensor into a linear combination of rank-

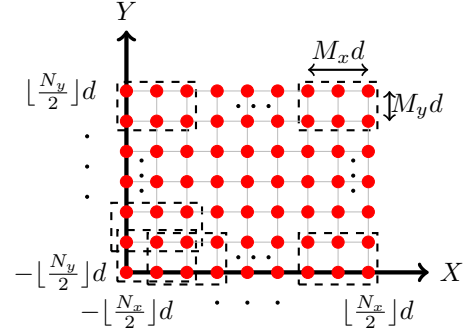


Fig. 1: URA configuration.

1 tensor components. The CPD of an N -dimensional tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$ is expressed as

$$\mathcal{A} = \sum_{r=1}^R \eta_r \mathbf{a}_1(r) \circ \mathbf{a}_2(r) \cdots \circ \mathbf{a}_N(r) \triangleq \llbracket \boldsymbol{\eta}; \mathbf{A}_1; \mathbf{A}_2; \cdots \mathbf{A}_N \rrbracket, \quad (1)$$

where R is a positive integer, $\mathbf{a}_n(r) \in \mathbb{C}^{I_n}$ is a CP factor, $\mathbf{A}_n = [\mathbf{a}_n(1), \mathbf{a}_n(2), \dots, \mathbf{a}_n(R)] \in \mathbb{C}^{I_n \times R}$ denotes the corresponding factor matrix for $n = 1, 2, \dots, N$, and $\boldsymbol{\eta} = [\eta_1, \eta_2, \dots, \eta_R]^T$ is a vector of scalar coefficients.

The CP rank of tensor \mathcal{A} is defined as the smallest number of rank-1 tensors required to reconstruct \mathcal{A} , i.e., the smallest integer R in Eq. (1).

III. SIGNAL MODEL

Consider a URA \mathbb{S} , as depicted in Fig. 1, consisting of N_x omnidirectional sensors located in the X axis and N_y sensors in the Y axis. Assuming both N_x and N_y to be odd integers, the sensor locations can be expressed as

$$\mathbb{S} = \{(x_{\mathbb{S}}, y_{\mathbb{S}}) | x_{\mathbb{S}} \in [-\lfloor N_x/2 \rfloor d, \lfloor N_x/2 \rfloor d], y_{\mathbb{S}} \in [-\lfloor N_y/2 \rfloor d, \lfloor N_y/2 \rfloor d]\}, \quad (2)$$

where $d = \lambda/2$ is the inter-element spacing with λ denoting the wavelength. The total number of array sensors is $N = N_x N_y$. The case where N_x and N_y are even is discussed in Appendix A, provided in the Supplement.

A. Structural Property of Uncorrelated Covariance Tensor

In this subsection, we review the structural properties of the covariance tensor for uncorrelated impinging signals. These properties provide essential insights for the subsequent discussion on decorrelating the covariance tensor corresponding to coherent signals.

Consider K narrowband uncorrelated far-field signals impinging on the array \mathbb{S} from directions (ϕ_k, θ_k) , $k = 1, 2, \dots, K$, where $\phi_k \in [-\pi, \pi]$ and $\theta_k \in [0, \pi]$, respectively, denote the azimuth and elevation angles. 2D angular information is embedded in each snapshot received by the array \mathbb{S} . Therefore, the array received signal at time t is modeled as

$$\mathbf{X}_u(t) = \sum_{k=1}^K s_k(t) \mathbf{a}(\mu_k) \circ \mathbf{a}(\nu_k) + \mathbf{N}(t) \in \mathbb{C}^{N_x \times N_y}, \quad (3)$$

where $\mu_k = \sin(\theta_k) \cos(\phi_k)$, $\nu_k = \sin(\theta_k) \sin(\phi_k)$, and $\mathbf{a}(\mu_k)$ and $\mathbf{a}(\nu_k)$ are the steering vectors along the X and Y axes, respectively expressed as

$$\begin{aligned} \mathbf{a}(\mu_k) &= [e^{-j\pi(-\lfloor \frac{N_x}{2} \rfloor)\mu_k}, \dots, e^{-j\pi(\lfloor \frac{N_x}{2} \rfloor)\mu_k}]^T \in \mathbb{C}^{N_x}, \\ \mathbf{a}(\nu_k) &= [e^{-j\pi(-\lfloor \frac{N_y}{2} \rfloor)\nu_k}, \dots, e^{-j\pi(\lfloor \frac{N_y}{2} \rfloor)\nu_k}]^T \in \mathbb{C}^{N_y}. \end{aligned} \quad (4)$$

In addition, $\mathbf{N}(t)$ is the independent and identically distributed additive white Gaussian noise (AWGN) matrix.

The 4D covariance tensor $\mathcal{R}_u \in \mathbb{C}^{N_x \times N_y \times N_x \times N_y}$ of matrix $\mathbf{X}_u(t)$ is obtained as

$$\begin{aligned} \mathcal{R}_u &= \mathbb{E}\{\mathbf{X}_u(t) \circ \mathbf{X}_u^*(t)\} \\ &= \sum_{k=1}^K \sigma_k^2 \mathbf{a}(\mu_k) \circ \mathbf{a}(\nu_k) \circ \mathbf{a}^*(\mu_k) \circ \mathbf{a}^*(\nu_k) + \sigma_n^2 \mathcal{I}, \end{aligned} \quad (5)$$

where σ_k^2 is power of the k th signal and σ_n^2 is the noise variance.

By utilizing a total number of T snapshots, the received signal matrices $\mathbf{X}_u(t)$, $t = 1, 2, \dots, T$, are concatenated along the temporal dimension to obtain the following three-dimensional data tensor,

$$\begin{aligned} \mathcal{X}_u &= [\mathbf{X}_u(1), \mathbf{X}_u(2), \dots, \mathbf{X}_u(T)]_{\sqcup_3} \\ &= \sum_{k=1}^K \mathbf{a}(\mu_k) \circ \mathbf{a}(\nu_k) \circ \mathbf{s}_k + \mathcal{N} \in \mathbb{C}^{N_x \times N_y \times T}, \end{aligned} \quad (6)$$

where $\mathbf{s}_k = [s_k(1), s_k(2), \dots, s_k(T)]^T$ is the signal waveform vector for the k th signal and \mathcal{N} is the noise tensor. In this case, the sample covariance tensor is estimated as

$$\hat{\mathcal{R}}_u = \frac{1}{T} \mathcal{X}_u \times_3 \mathcal{X}_u^*. \quad (7)$$

The uncorrelated covariance tensor \mathcal{R}_u admits a rank- K CP model and exhibits the tensorial Hermitian Toeplitz property, i.e.,

$$\begin{aligned} \mathcal{R}_u(m, n, m', n') &= \sum_{k=1}^K \sigma_k^2 e^{-j\pi(m-m')\mu_k} e^{-j\pi(n-n')\nu_k} \\ &\quad + \sigma_n^2 \delta_{(m,n,m',n')} \\ &= \mathcal{R}_u(m + c_1, n + c_2, m' + c_1, n' + c_2) \\ &= \mathcal{R}_u^*(m' + c_1, n' + c_2, m + c_1, n + c_2), \end{aligned} \quad (8)$$

where c_1 and c_2 are integer constants and

$$\delta_{(m,n,m',n')} = \begin{cases} 1, & m = m' \text{ and } n = n', \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

It is observed in Eq. (8) that a specific slice of the uncorrelated covariance tensor, such as $\mathcal{R}_u(m, :, m', :)$, exhibits the Toeplitz property, i.e., $\mathcal{R}_u(m, n, m', n') = \mathcal{R}_u(m, n + c_2, m', n' + c_2)$. However, it may not exhibit the Hermitian property, i.e., $\mathcal{R}_u(m, n, m', n')$ is not necessarily equal to $\mathcal{R}_u^*(m, n', m', n)$.

When coherent signals are present, this tensorial Toeplitz property no longer holds and, as a result, decorrelation is necessary. This is considered in the following subsection.

B. Signal Model of the Coherent Signals

Now assume that the impinging K signals are mutually coherent to each other. Taking the first signal $s_1(t)$ as the reference, the k th coherent signal at time instant t can be expressed as

$$s_k(t) = \alpha_k s_1(t), \quad (10)$$

for $k = 1, 2, \dots, K$, where α_k represents a complex attenuation factor of the k th signal with respect to the reference signal with $\alpha_1 = 1$. The array received signal tensor for the coherent case can be expressed as

$$\mathcal{X} = \sum_{k=1}^K \alpha_k \mathbf{a}(\mu_k) \circ \mathbf{a}(\nu_k) \circ \mathbf{s}_1 + \mathcal{N} \in \mathbb{C}^{N_x \times N_y \times T}, \quad (11)$$

where $\mathbf{s}_1 = [s_1(1), s_1(2), \dots, s_1(T)]^T$ is the signal waveform vector for the reference signal.

The 4D forward covariance tensor, $\mathcal{R}^{(f)} \in \mathbb{C}^{N_x \times N_y \times N_x \times N_y}$, of matrix $\mathbf{X}(t)$ for the coherent signal case is obtained as

$$\begin{aligned} \mathcal{R}^{(f)} &= \sigma_s^2 \sum_{k=1}^K \sum_{k'=1}^K \alpha_k^* \alpha_{k'} \mathbf{a}(\mu_{k'}) \circ \mathbf{a}(\nu_{k'}) \circ \mathbf{a}^*(\mu_k) \circ \mathbf{a}^*(\nu_k) \\ &\quad + \sigma_n^2 \mathcal{I}, \end{aligned} \quad (12)$$

where $\sigma_s^2 = \mathbb{E}[s_1(t)s_1^*(t)]$ is the power of the reference signal. Due to the presence of the cross-terms, it is clear in Eq. (12) that the covariance tensor $\mathcal{R}^{(f)}$ suffers from a rank-deficiency problem and needs to be appropriately preprocessed to facilitate DOA estimation.

IV. DECORRELATION OF COVARIANCE TENSOR BASED ON STRUCTURED TENSOR RECONSTRUCTION

A. Forward-Backward Tensor Reconstruction Strategy

In this section, we proposed an effective decorrelation strategy for the coherent covariance tensor using a structured tensor reconstruction approach. The covariance tensor in Eq. (12) does not exhibit the tensorial Toeplitz property due to the presence of cross-correlations between the mutually coherent sources, resulting in rank deficiency. To address this problem, we extend the approach proposed in [55], which utilizes only the forward covariance tensor, to incorporate both forward and backward covariance tensors, thereby increasing the dimensionality of the decorrelated covariance tensor and, subsequently, enhancing the DOFs.

Express a particular (m, n, m', n') th element of the forward coherent covariance tensor $\mathcal{R}^{(f)}$ in Eq. (12) as

$$\begin{aligned} \mathcal{R}^{(f)}(m, n, m', n') &= \sigma_s^2 \sum_{k'=1}^K \alpha_{k'} e^{-j\pi m \mu_{k'}} e^{-j\pi n \nu_{k'}} \\ &\quad \cdot \sum_{k=1}^K \alpha_k^* e^{j\pi m' \mu_k} e^{j\pi n' \nu_k} + \sigma_n^2 \delta_{(m,n,m',n')} \\ &= b_{(m,n)} \sum_{k=1}^K \alpha_k^* e^{j\pi m' \mu_k} e^{j\pi n' \nu_k} \\ &\quad + \sigma_n^2 \delta_{(m,n,m',n')}, \end{aligned} \quad (13)$$

where $m, m' \in [-\lfloor \frac{N_x}{2} \rfloor, \lfloor \frac{N_x}{2} \rfloor]$ and $n, n' \in [-\lfloor \frac{N_y}{2} \rfloor, \lfloor \frac{N_y}{2} \rfloor]$. The term $b_{(m,n)} = \sigma_s^2 \sum_{k'=1}^K \alpha_{k'} e^{-j\pi m \mu_{k'}} e^{-j\pi n \nu_{k'}}$ depends only on the first 2D indices, i.e., m and n , of the covariance tensor. This tensor contains cross-covariance terms among the impinging signals, resulting in rank deficiency. Additionally, it does not satisfy the tensorial Toeplitz property, as described in Section III-A. To address these issues, we develop the proposed structured tensor reconstruction-based decorrelation strategy, as described below.

We first obtain the backward covariance tensor through flipped sensor ordering and complex conjugation as

$$\mathcal{R}^{(b)}(m, n, m', n') = \left(\mathcal{R}^{(f)}(-m, -n, -m', -n') \right)^* \quad (14)$$

By fixing the first two indices of the covariance tensors to specific values, e.g., (m, n) , we obtain the following forward and backward covariance matrices as

$$\begin{aligned} \mathbf{R}^{(f)} &= \text{squeeze} \left(\mathcal{R}^{(f)}(m, n, :, :) \right) \in \mathbb{C}^{N_x \times N_y}, \\ \mathbf{R}^{(b)} &= \text{squeeze} \left(\mathcal{R}^{(b)}(m, n, :, :) \right) \in \mathbb{C}^{N_x \times N_y}. \end{aligned} \quad (15)$$

Using these matrices, we construct a decorrelated tensor $\mathcal{D} \in \mathbb{C}^{\tilde{N}_x \times \tilde{N}_y \times \tilde{N}_x \times \tilde{N}_y}$, where the $(\tilde{m}, :, \tilde{m}', :)$ th slice of \mathcal{D} is obtained by arranging the rows of $\mathbf{R}^{(f)}$ and $\mathbf{R}^{(b)}$ in a Toeplitz structure. Note that the row indices of both $\mathbf{R}^{(f)}$ and $\mathbf{R}^{(b)}$ are between $-\lfloor \frac{N_x}{2} \rfloor$ and $\lfloor \frac{N_x}{2} \rfloor$.

To construct the decorrelated covariance tensor \mathcal{D} , the approach in [55] employs only the forward covariance matrix $\mathbf{R}^{(f)}$. Specifically, a second and fourth dimensional slice $\mathcal{D}(:, \tilde{n}, :, \tilde{n}')$ is obtained by extracting the $(-\tilde{n} + \tilde{n}')$ th column of $\mathbf{R}^{(f)}$ and arranging it in a Toeplitz form. The 0th element of $\mathbf{R}^{(f)}$ is placed at the top, followed by the remaining elements to form a complete Toeplitz structure, as shown in Fig. 2(a).

In contrast, the proposed method incorporates both the forward and backward covariance matrices $\mathbf{R}^{(f)}$ and $\mathbf{R}^{(b)}$. In this case, the slice $\mathcal{D}(:, \tilde{n}, :, \tilde{n}')$ is constructed by placing the $-p$ th element of the $(-\tilde{n} + \tilde{n}')$ th column at the top and forming a Toeplitz structure using $\mathbf{R}^{(f)}$ for indices satisfying $\tilde{m} \leq \lfloor \frac{N_x}{2} \rfloor - p$. Beyond this range, the corresponding column of $\mathbf{R}^{(b)}$ is used to complete the slice, as shown in Fig. 2(b). This joint utilization increases the slice dimension from $\lfloor \frac{N_x+1}{2} \rfloor \times \lfloor \frac{N_x+1}{2} \rfloor$ to $(\lfloor \frac{N_x+1}{2} \rfloor + p) \times (\lfloor \frac{N_x+1}{2} \rfloor + p)$.

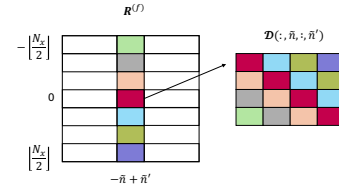
In general, a particular $(\tilde{m}, \tilde{n}, \tilde{m}', \tilde{n}')$ th element of tensor \mathcal{D} can be obtained as

$$\begin{aligned} &\mathcal{D}(\tilde{m}, \tilde{n}, \tilde{m}', \tilde{n}') \\ &= \begin{cases} \mathcal{R}^{(f)}(m, n, -\tilde{m} + \tilde{m}' - p, -\tilde{n} + \tilde{n}'), & \tilde{m} \leq \lfloor \frac{N_x}{2} \rfloor - p, \\ \mathcal{R}^{(b)}(m, n, -\tilde{m} + \tilde{m}' + p, -\tilde{n} + \tilde{n}') = \\ \left(\mathcal{R}^{(f)}(-m, -n, \tilde{m} - \tilde{m}' - p, \tilde{n} - \tilde{n}') \right)^*, & \tilde{m} > \lfloor \frac{N_x}{2} \rfloor - p, \end{cases} \quad (16) \end{aligned}$$

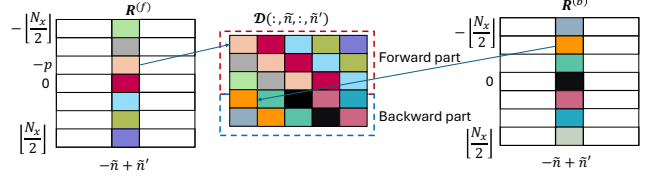
where $\tilde{m}, \tilde{m}' \in [0, \lfloor \frac{N_x}{2} \rfloor + p]$ and $\tilde{n}, \tilde{n}' \in [0, \lfloor \frac{N_y}{2} \rfloor]$. The reconstructed tensor \mathcal{D} satisfies the following rank property.

Theorem 1. *Tensor \mathcal{D} satisfies the tensorial Toeplitz property and serves as a rank- K decorrelated covariance tensor.*

Proof. According to the mapping rule described in Eq. (16), a particular element of tensor \mathcal{D} can be expressed as



(a) Forward only



(b) Forward backward

Fig. 2: Construction of a slice $\mathcal{D}(:, \tilde{n}, :, \tilde{n}')$, (a): Forward only mapping: only $\mathbf{R}^{(f)}$ is used to construct the slice, (b): Forward backward mapping: Both $\mathbf{R}^{(f)}$ and $\mathbf{R}^{(b)}$ are used to construct the slice. In this example $N_x = 7$ is considered with $p = 1$.

$$\mathcal{D}(\tilde{m}, \tilde{n}, \tilde{m}', \tilde{n}') = b_{(m,n)}^{fb} \sum_{k=1}^K \gamma_k^{fb} e^{-j\pi(\tilde{m}-\tilde{m}')\mu_k} e^{-j\pi(\tilde{n}-\tilde{n}')\nu_k}, \quad (17)$$

where

$$\begin{aligned} &b_{(m,n)}^{fb} \\ &= \begin{cases} b_{(m,n)}^{(1)} = \sigma_s^2 \sum_{k'=1}^K \alpha_{k'} e^{-j\pi m \mu_{k'}} e^{-j\pi n \nu_{k'}}, & \tilde{m} \leq \lfloor \frac{N_x}{2} \rfloor - p, \\ b_{(m,n)}^{(2)} = \sigma_s^2 \sum_{k'=1}^K \alpha_{k'}^* e^{-j\pi m \mu_{k'}} e^{-j\pi n \nu_{k'}}, & \tilde{m} > \lfloor \frac{N_x}{2} \rfloor - p, \end{cases} \quad (18) \end{aligned}$$

and

$$\gamma_k^{fb} = \begin{cases} \gamma_k = \alpha_k^* e^{-j\pi p \mu_k}, & \tilde{m} \leq \lfloor \frac{N_x}{2} \rfloor - p, \\ \gamma_k = \alpha_k e^{j\pi p \mu_k}, & \tilde{m} > \lfloor \frac{N_x}{2} \rfloor - p. \end{cases} \quad (19)$$

Tensor \mathcal{D} described in Eq. (17) exhibits a tensorial Toeplitz structure, i.e., it satisfies $\mathcal{D}(\tilde{m}, \tilde{n}, \tilde{m}', \tilde{n}') = \mathcal{D}(\tilde{m} + c_1, \tilde{n} + c_2, \tilde{m}' + c_1, \tilde{n}' + c_2)$. Additionally, each slice of \mathcal{D} also demonstrates a Toeplitz structure. From Eq. (17), we can reformulate \mathcal{D} as

$$\begin{aligned} \mathcal{D} &= b_{(m,n)}^{fb} \sum_{k=1}^K \gamma_k^{fb} \mathbf{g}(\mu_k) \circ \mathbf{g}(\nu_k) \circ \mathbf{g}^*(\mu_k) \circ \mathbf{g}^*(\nu_k) \\ &= b_{(m,n)}^{fb} [\boldsymbol{\gamma}^{fb}; \mathbf{G}(\mu); \mathbf{G}(\nu); \mathbf{G}^*(\mu); \mathbf{G}^*(\nu)] \\ &\in \mathbb{C}^{(\lfloor \frac{N_x+1}{2} \rfloor + p) \times (\lfloor \frac{N_y+1}{2} \rfloor) \times (\lfloor \frac{N_x+1}{2} \rfloor + p) \times (\lfloor \frac{N_y+1}{2} \rfloor)}, \end{aligned} \quad (20)$$

where $\boldsymbol{\gamma}^{fb} = [\gamma_1^{fb}, \dots, \gamma_K^{fb}]^T \in \mathbb{C}^K$, $\mathbf{g}(\mu_k) = [1, e^{-j\pi \mu_k}, \dots, e^{-j\pi(\lfloor \frac{N_x}{2} \rfloor + p)\mu_k}]^T \in \mathbb{C}^{\lfloor \frac{N_x+1}{2} \rfloor + p}$ and $\mathbf{g}(\nu_k) = [1, e^{-j\pi \nu_k}, \dots, e^{-j\pi(\lfloor \frac{N_y}{2} \rfloor)\nu_k}]^T \in \mathbb{C}^{\lfloor \frac{N_y+1}{2} \rfloor}$ act as steering vectors for the k th coherent signal along the X and Y axes, respectively, and $\mathbf{G}(\mu) = [\mathbf{g}(\mu_1), \dots, \mathbf{g}(\mu_K)]$ and $\mathbf{G}(\nu) = [\mathbf{g}(\nu_1), \dots, \mathbf{g}(\nu_K)]$ are the CP factor matrices. From Eq. (20), it is observed that the reconstructed covariance tensor \mathcal{D} is represented as the sum of only K outer products, effectively eliminating cross-terms among the impinging signals compared to the coherent covariance tensor in Eq. (12). Therefore,

\mathcal{D} admits a rank- K CP model and serves as the decorrelated covariance tensor of the array. \square

B. DOA Estimation via CPD

1) *DOA estimation*: The decorrelated covariance tensor \mathcal{D} in Eq. (20) can alternatively be expressed as

$$\begin{aligned} \mathcal{D} &= b_{(m,n)}^{fb} \sum_{k=1}^K (\mathbf{g}(\mu_k) \circ \mathbf{g}(\nu_k) \circ \mathbf{g}^*(\mu_k) \circ \mathbf{g}^*(\nu_k)) \times_1 \mathbf{\Gamma}^{fb} \\ &= b_{(m,n)}^{fb} \left[\tilde{\mathbf{G}}(\mu); \mathbf{G}(\nu); \mathbf{G}^*(\mu); \mathbf{G}^*(\nu) \right], \end{aligned} \quad (21)$$

where $\mathbf{\Gamma}^{fb} = \text{diag}(\gamma)$. Since the varying values of γ_k^{fb} are absorbed into the first factor matrix via the mode-1 tensor-matrix product, the first factor matrix $\tilde{\mathbf{G}}(\mu)$ becomes contaminated and no longer represents the steering vectors. In contrast, the remaining factor matrices are unaffected and thus remain reliable for DOA estimation. The tensor \mathcal{D} is decomposed via CPD to obtain the factor matrices $\hat{\mathbf{G}}(\mu)$, $\hat{\mathbf{G}}(\nu)$, $\hat{\mathbf{G}}^*(\mu)$ and $\hat{\mathbf{G}}^*(\nu)$. The k th columns of $\hat{\mathbf{G}}^*(\mu)$ and $\hat{\mathbf{G}}^*(\nu)$, denoted as $\hat{\mathbf{g}}^*(\mu_k)$ and $\hat{\mathbf{g}}^*(\nu_k)$, are used to estimate μ_k and ν_k according to

$$\mu_k = \frac{1}{\pi(\lfloor \frac{N_x+1}{2} \rfloor + p)} \sum_{i=1}^{\lfloor \frac{N_x+1}{2} \rfloor + p} \frac{\angle \hat{\mathbf{g}}^*(\mu_k)(i+1)}{\hat{\mathbf{g}}^*(\mu_k)(i)} \quad (22)$$

and

$$\nu_k = \frac{1}{\pi \lfloor \frac{N_y+1}{2} \rfloor} \sum_{i=1}^{\lfloor \frac{N_x+1}{2} \rfloor} \frac{\angle \hat{\mathbf{g}}^*(\nu_k)(i+1)}{\hat{\mathbf{g}}^*(\nu_k)(i)}, \quad (23)$$

where $\hat{\mathbf{g}}^*(\mu_k)(i)$ and $\hat{\mathbf{g}}^*(\nu_k)(i)$ denote the i th elements of $\hat{\mathbf{g}}^*(\mu_k)$ and $\hat{\mathbf{g}}^*(\nu_k)$, respectively. Finally, the azimuth and elevation angles are estimated as $\hat{\phi}_k = \arctan(\frac{\hat{\mu}_k}{\hat{\nu}_k})$ and $\theta_k = \arcsin(\sqrt{\hat{\mu}_k^2 + \hat{\nu}_k^2})$.

2) *Identifiability in CPD*: The uniqueness of the estimated factor matrices is guaranteed through CPD when the following condition is satisfied

$$\kappa(\hat{\mathbf{G}}(\mu)) + \kappa(\hat{\mathbf{G}}(\nu)) + \kappa(\hat{\mathbf{G}}^*(\mu)) + \kappa(\hat{\mathbf{G}}^*(\nu)) \geq 2K + 3, \quad (24)$$

where $\kappa(\hat{\mathbf{G}}(\mu)) = \kappa(\hat{\mathbf{G}}^*(\mu)) = \min(\lfloor \frac{N_x+1}{2} \rfloor + p, K)$ and $\kappa(\hat{\mathbf{G}}(\nu)) = \kappa(\hat{\mathbf{G}}^*(\nu)) = \min(\lfloor \frac{N_y+1}{2} \rfloor, K)$. From this condition, the maximum number of coherent signals that can be detected is

$$\text{DOF}_{\text{FBTR}} = \left\lfloor \frac{N_x+1}{2} \right\rfloor + \left\lfloor \frac{N_y+1}{2} \right\rfloor - 2 + p. \quad (25)$$

The forward-only tensor reconstruction (FTR) developed in [55] is able to detect at most $\lfloor \frac{N_x+1}{2} \rfloor + \lfloor \frac{N_y+1}{2} \rfloor - 2$ sources, whereas the proposed approach can detect p additional sources, thereby providing an improvement in the maximum number of resolvable coherent signals.

C. Analysis of the Number of Degrees of Freedom

In this subsection, we derive an expression for the number of DOFs achieved using the FBTR. Note that, the proposed approach described in Eq. (16) obtains a higher dimension of \mathcal{D} and can resolve p more sources compared to the method

developed in [55] because we used both forward and backward covariances.

From the above discussion, it is evident that the term p plays a crucial role in dimensionality increment and, consequently, in the enhancement of the number of DOFs. The following lemma quantitatively determines the optimum value of p that maximizes the number of DOFs. Note that Lemma 1 considers only the X dimension as p is defined along that dimension.

Lemma 1. For N_x antennas in the X dimension, the optimum value of p that maximizes the number of DOFs is given by

$$p = \begin{cases} \lfloor \frac{N_x+1}{6} \rfloor, & N_x \text{ is odd,} \\ \lfloor \frac{N_x+4}{6} \rfloor, & N_x \text{ is even.} \end{cases} \quad (26)$$

Proof. Consider a slice of the decorrelated tensor, $\text{squeeze}(\mathcal{D}(:, \tilde{n}, :, \tilde{n}'))$, in which the second and fourth dimensional indices are fixed. In the forward-only SS case, where $p = 0$ and N_x is an odd number, the slice has $\lfloor \frac{N_x+1}{2} \rfloor$ rows and $\lfloor \frac{N_x+1}{2} \rfloor$ columns derived from the forward matrix. As p increases, the number of columns increases by p to become $\lfloor \frac{N_x+1}{2} \rfloor + p$, whereas the number of rows contributed by the forward matrix decreases by p to become $\lfloor \frac{N_x+1}{2} \rfloor - p$. To maintain a square structure for the slice, the backward matrix fills in the remaining rows, resulting in an increased overall dimension for the slice, hence an increased number of DOFs. The optimum value of p is the maximum value to maintain a square structure of the slice. This condition can be expressed as

$$2 \left(\left\lfloor \frac{N_x+1}{2} \right\rfloor - p \right) \geq \left(\left\lfloor \frac{N_x+1}{2} \right\rfloor + p \right), \quad (27)$$

which results

$$p \leq \left\lfloor \frac{N_x+1}{6} \right\rfloor. \quad (28)$$

As a result, the maximum value that p can take is

$$p = \left\lfloor \frac{N_x+1}{6} \right\rfloor. \quad (29)$$

In a similar fashion, it is proved in Appendix A of the Supplement that $p = \lfloor \frac{N_x+4}{6} \rfloor$ when N_x is even. \square

For all integer values of N_x , whether odd or even, the DOF expressions from Eq. (25) can be simplified as

$$\text{DOF}_{\text{FBTR}} = \left\lfloor \frac{N_x+1}{2} \right\rfloor + \left\lfloor \frac{N_y+1}{2} \right\rfloor - 2 + \left\lfloor \frac{N_x}{6} \right\rfloor + u_1, \quad (30)$$

where

$$u_1 = \begin{cases} 0, & N_x \equiv 0, 1, 3 \pmod{6}, \\ 1, & N_x \equiv 2, 4, 5 \pmod{6}. \end{cases} \quad (31)$$

To achieve even more DOFs, we matricize tensor \mathcal{D} , resulting in a higher dimension of steering vectors along the X and Y axes as

$$\begin{aligned} \mathbf{D} &\triangleq \langle \mathcal{D} \rangle_{\{1,2\},\{3,4\}} \\ &= \mathbf{J} \odot b_{(m,n)}^{(1)} \mathbf{G} \mathbf{\Gamma} \mathbf{G}^H + (\mathbf{1} - \mathbf{J}) \odot b_{(m,n)}^{(2)} \mathbf{G} \mathbf{\Gamma}^H \mathbf{G}^H, \end{aligned} \quad (32)$$

where $\mathbf{G} = \mathbf{G}(\mu) \otimes \mathbf{G}(\nu) \in \mathbb{C}^{\tilde{N} \times K}$, $\mathbf{\Gamma} = \text{Diag}(\gamma_1, \dots, \gamma_K) \in \mathbb{C}^{K \times K}$. In addition, $\mathbf{J} \in \{0, 1\}^{\tilde{N} \times \tilde{N}}$ is a selection matrix,

where the first \check{N} rows are filled with ones and the remaining $\check{N} - \check{N}$ rows are filled with zeros, expressed as

$$\mathbf{J} = [\mathbf{1}_{\check{N} \times \check{N}}^T \quad \mathbf{0}_{(\check{N}-\check{N}) \times \check{N}}^T]^T \quad (33)$$

with $\check{N} = (\lfloor \frac{N_x+1}{2} \rfloor + p)(\lfloor \frac{N_y+1}{2} \rfloor)$ and $\check{N} = (\lfloor \frac{N_x+1}{2} \rfloor - p)(\lfloor \frac{N_y+1}{2} \rfloor)$.

Since \mathbf{D} is full rank provided that all the μ and ν values are distinct, the maximum number of sources it can detect is

$$\text{DOF}_{\text{FBTR-mat}} = \left\lfloor \frac{N_x+1}{2} \right\rfloor \left\lfloor \frac{N_y+1}{2} \right\rfloor + p \left\lfloor \frac{N_y+1}{2} \right\rfloor - 1, \quad (34)$$

where p is given in (26). As such, this method can detect $p \lfloor \frac{N_y+1}{2} \rfloor$ more sources than the FTR as developed in [55]. Note that, the value of p increases with increase of N_x , i.e., for a larger array size, the difference in the number of DOFs between the proposed method and the approach developed in [55] becomes more pronounced.

For all integer values of N_x , whether odd or even, from Eq. (34) can be expressed as

$$\text{DOF}_{\text{FBTR-mat}} = \left\lfloor \frac{N_y+1}{2} \right\rfloor \left(4 \left\lfloor \frac{N_x}{6} \right\rfloor + u_2 \right) - 1, \quad (35)$$

where

$$u_2 = \begin{cases} 0, & N_x \equiv 0 \pmod{6}, \\ 1, & N_x \equiv 1 \pmod{6}, \\ 2, & N_x \equiv 2 \text{ or } 3 \pmod{6}, \\ 3, & N_x \equiv 4 \pmod{6}, \\ 4, & N_x \equiv 5 \pmod{6}. \end{cases} \quad (36)$$

The derivation is provided in Appendix A of the Supplement.

Compared to the FTR developed in [55], the proposed approach offers a larger number of DOFs. In particular, for DOA estimation using CPD, it can resolve $\lfloor \frac{N_x}{6} \rfloor + u_1$ additional sources compared to [55]. Furthermore, when the covariance tensor is matricized and MUSIC is applied for DOA estimation in both cases, the proposed method achieves $(\lfloor \frac{N_x}{6} \rfloor + u_2)(\lfloor \frac{N_y+1}{2} \rfloor)$ more DOFs than the method in [55].

It is noted that the above DOF expression is derived under the assumption that all μ and ν values are distinct. When multiple DOAs share the same ν and μ values, additional conditions are required to maintain the rank for applying the MUSIC algorithm. Consider the case where $\eta_x - 1$ sources share the same ν value, denoted as ν_0 . Because the rank of matrix \mathbf{D} is determined by $\mathbf{G} = \mathbf{G}(\mu) \otimes \mathbf{G}(\nu)$, a relevant submatrix of \mathbf{G} that contains these repeated ν values can be written as

$$\mathbf{G}_s = (\mathbf{I} \otimes \mathbf{g}(\nu_0)) [\mathbf{g}(\mu_{k,1}), \mathbf{g}(\mu_{k,2}), \dots, \mathbf{g}(\mu_{k,(\eta_x-1)})], \quad (37)$$

which has rank $\min(\lfloor \frac{N_x+1}{2} \rfloor + p, \eta_x - 1)$. To avoid rank loss within this group of identical ν values and to enable spectrum estimation using MUSIC by satisfying the extended linear independence property [38], it is required to satisfy $\lfloor \frac{N_x+1}{2} \rfloor + p \geq \eta_x$. Similarly, to accommodate a group of $\eta_y - 1$ identical μ 's, it is required that $\lfloor \frac{N_y+1}{2} \rfloor \geq \eta_y$.

V. DECORRELATION OF THE COVARIANCE TENSOR USING FBSS

In this section, we describe the SS-based processing to decorrelate the covariance tensor. In this approach, the antenna array is divided into multiple overlapping subarrays as depicted in Fig. 1. Each subarray consists of M_x sensors along the X direction and M_y sensors along the Y direction. The total number of subarrays is $L_x L_y$, where

$$L_x = N_x - M_x + 1 \quad \text{and} \quad L_y = N_y - M_y + 1 \quad (38)$$

represent the numbers of subarrays in the X and Y directions, respectively. The covariance tensors of all subarrays, along with their flipped and conjugated counterparts, are then averaged to obtain the decorrelated covariance tensor.

In the following two subsections, we describe the mechanism of forward and backward smoothing approaches.

A. Forward Smoothing

Consider the covariance matrix corresponding to the $(1, 1)$ th subarray as the reference, expressed as

$$\begin{aligned} \mathbf{R}_{(1,1)}^{(f)} &= \mathbf{A} \mathbf{R}_s \mathbf{A}^H + \sigma_n^2 \mathbf{I} \\ &= \sigma_s^2 \mathbf{A} \boldsymbol{\alpha} \boldsymbol{\alpha}^H \mathbf{A}^H + \sigma_n^2 \mathbf{I} \in \mathbb{C}^{M_x M_y \times M_x M_y}, \end{aligned} \quad (39)$$

where $\mathbf{A} = [\mathbf{a}_s(\mu_1) \otimes \mathbf{a}_s(\nu_1), \dots, \mathbf{a}_s(\mu_K) \otimes \mathbf{a}_s(\nu_K)] \in \mathbb{C}^{M_x M_y \times K}$ is the array manifold matrix with

$$\mathbf{a}_s(\mu_k) = [e^{(-\lfloor \frac{N_x}{2} \rfloor) u_k}, \dots, e^{(-\lfloor \frac{N_x}{2} \rfloor + M_x - 1) u_k}]^T \in \mathbb{C}^{M_x}, \quad (40)$$

$$\mathbf{a}_s(\nu_k) = [e^{(-\lfloor \frac{N_y}{2} \rfloor) v_k}, \dots, e^{(-\lfloor \frac{N_y}{2} \rfloor + M_y - 1) v_k}]^T \in \mathbb{C}^{M_y} \quad (41)$$

are the steering vectors of the reference subarray along the X and Y directions, respectively, with $u_k = e^{-j\pi\mu_k}$ and $v_k = e^{-j\pi\nu_k}$, $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_K]^T$, and $\mathbf{R}_s = \mathbb{E}\{\mathbf{s}(t)\mathbf{s}^H(t)\} = \sigma_s^2 \boldsymbol{\alpha} \boldsymbol{\alpha}^H$ is the source covariance matrix. In a similar fashion, the forward covariance matrix for the (l_x, l_y) th subarray can be expressed as

$$\mathbf{R}_{(l_x, l_y)}^{(f)} = \sigma_s^2 \mathbf{A} \boldsymbol{\Phi}_{l_x, l_y} \boldsymbol{\alpha} \boldsymbol{\alpha}^H \boldsymbol{\Phi}_{l_x, l_y}^H \mathbf{A}^H + \sigma_n^2 \mathbf{I}, \quad (42)$$

for $l_x = 1, \dots, L_x$ and $l_y = 1, \dots, L_y$, where $\boldsymbol{\Phi}_{l_x, l_y} = \text{Diag}([e^{-j\pi\gamma_{l_x, l_y}^1}, \dots, e^{-j\pi\gamma_{l_x, l_y}^K}])$ with $\gamma_{l_x, l_y}^k = (l_x - 1)\mu_k + (l_y - 1)\nu_k$. The forward covariance matrix can be obtained as

$$\mathbf{R}^{(f)} = \frac{1}{L_x L_y} \sum_{l_x=1}^{L_x} \sum_{l_y=1}^{L_y} \mathbf{R}_{(l_x, l_y)}^{(f)} = \mathbf{A} \mathbf{R}_s^{(f)} \mathbf{A}^H + \sigma_n^2 \mathbf{I} \quad (43)$$

with

$$\mathbf{R}_s^{(f)} = \frac{\sigma_s^2}{L_x L_y} \sum_{l_x=1}^{L_x} \sum_{l_y=1}^{L_y} \boldsymbol{\Phi}_{l_x, l_y} \boldsymbol{\alpha} \boldsymbol{\alpha}^H \boldsymbol{\Phi}_{l_x, l_y}^H = \frac{\sigma_s^2}{L_x L_y} \mathbf{C} \mathbf{C}^H, \quad (44)$$

where

$$\mathbf{C} = [\boldsymbol{\Phi}_{1,1} \boldsymbol{\alpha}, \dots, \boldsymbol{\Phi}_{L_x, L_y} \boldsymbol{\alpha}] = \mathbf{D} \mathbf{V} \in \mathbb{C}^{K \times L_x L_y}. \quad (45)$$

The rank of the source covariance matrix for the forward smoothing, $\mathbf{R}_s^{(f)}$, is same as the rank of \mathbf{C} . Also, since $\mathbf{C} = \mathbf{D} \mathbf{V}$ and \mathbf{D} is a diagonal square matrix with nonzero

diagonal entries and is thus full rank, the rank of \mathbf{C} is the same as the rank of \mathbf{V} . Matrix \mathbf{V} can be expressed as

$$\mathbf{V} = \begin{bmatrix} e^{-j\pi\gamma_{1,1}^1} & e^{-j\pi\gamma_{1,2}^1} & \dots & e^{-j\pi\gamma_{L_x L_y}^1} \\ e^{-j\pi\gamma_{1,1}^2} & e^{-j\pi\gamma_{1,2}^2} & \dots & e^{-j\pi\gamma_{L_x L_y}^2} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-j\pi\gamma_{1,1}^K} & e^{-j\pi\gamma_{1,2}^K} & \dots & e^{-j\pi\gamma_{L_x L_y}^K} \end{bmatrix}, \quad (46)$$

which is of rank $\min(K, L_x L_y)$ for distinct μ and ν due to its Vandermonde-like structure. As a result, the source covariance matrix $\mathbf{R}_s^{(f)}$ associated with the forward smoothed covariance matrix is non-singular and can detect K coherent sources if

$$L_x L_y \geq K. \quad (47)$$

If $\eta_x - 1$ DOAs share the same $\nu = \nu_0$, then for those sources $\gamma_{l_x, l_y}^k = (l_x - 1)\mu_k + (l_y - 1)\nu_0$, so variation of γ_{l_x, l_y}^k of across l_y is only a common scalar and the corresponding rows of \mathbf{V} depend on l_x in a Vandermonde manner. Consequently, the submatrix of \mathbf{V} formed by these rows has rank $\min(\eta_x - 1, L_x)$. To avoid rank loss of \mathbf{V} , $L_x \geq \eta_x - 1$ is required. Similarly, if $(\eta_y - 1)$ DOAs has same μ values, then $L_y \geq \eta_y - 1$.

B. Backward Smoothing

In a similar fashion, the backward covariance matrix for the (l_x, l_y) th subarray can be expressed as

$$\mathbf{R}_{(l_x, l_y)}^{(b)} = \mathbf{A}\Phi_{l_x, l_y} \mathbf{R}_{\bar{s}} \Phi_{l_x, l_y}^H \mathbf{A}^H + \sigma_n^2 \mathbf{I}, \quad (48)$$

where $\mathbf{R}_{\bar{s}}$ is the source covariance matrix for the backward smoothing and can be expressed as

$$\begin{aligned} \mathbf{R}_{\bar{s}} &= \sigma_s^2 \Phi_{-(N_x-1), -(N_y-1)} \boldsymbol{\alpha}^* \boldsymbol{\alpha}^T \Phi_{-(N_x-1), -(N_y-1)}^H \\ &= \sigma_s^2 \boldsymbol{\delta} \boldsymbol{\delta}^H, \end{aligned} \quad (49)$$

where $\boldsymbol{\delta} = \Phi_{-(N_x-1), -(N_y-1)} \boldsymbol{\alpha}^*$. Then, the backward smoothed source covariance matrix $\mathbf{R}_s^{(b)}$ can be obtained by averaging the source covariance matrices of all subarrays, expressed as

$$\mathbf{R}_s^{(b)} = \frac{\sigma_s^2}{L_x L_y} \sum_{l_x=1}^{L_x} \sum_{l_y=1}^{L_y} \Phi_{l_x, l_y} \boldsymbol{\delta} \boldsymbol{\delta}^H \Phi_{l_x, l_y}^H = \frac{\sigma_s^2}{L_x L_y} \mathbf{E} \mathbf{E}^H, \quad (50)$$

where $\mathbf{E} = [\Phi_{1,1} \boldsymbol{\delta}, \dots, \Phi_{L_x, L_y} \boldsymbol{\delta}] = \mathbf{F} \mathbf{V}$ with $\mathbf{F} = \text{Diag}(\boldsymbol{\delta})$. With a similar argument as the forward SS case, the rank of the backward smoothed source covariance matrix $\mathbf{R}_s^{(b)}$ is $\min(K, L_x L_y)$. Therefore, if backward spatial smoothing is used alone, K coherent sources can be detected provided that the necessary condition

$$L_x L_y \geq K, \quad (51)$$

and the sufficient condition is $L_x \geq \eta_x - 1$ and $L_y \geq \eta_y - 1$.

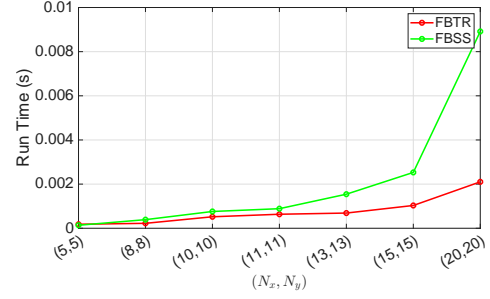


Fig. 3: Runtime comparison for FBTR vs FBSS.

C. Forward-Backward Smoothing

In this subsection, we demonstrate that, by combining the both forward and backward smoothing, the number of required subarrays can be reduced by half. The FB smoothed source covariance matrix is obtained by taking average of the forward and backward source covariance matrices as

$$\mathbf{R}_s^{(fb)} = \frac{\sigma_s^2}{2L_x L_y} (\mathbf{C} \mathbf{C}^H + \mathbf{E} \mathbf{E}^H) = \frac{\sigma_s^2}{2L_x L_y} \mathbf{G} \mathbf{G}^H, \quad (52)$$

where \mathbf{G} is given by

$$\mathbf{G} = [\mathbf{C} \ \mathbf{E}] = [\mathbf{D} \mathbf{V} \ \mathbf{F} \mathbf{V}] \in \mathbb{C}^{K \times 2L_x L_y}. \quad (53)$$

It is clear that $\text{rank}(\mathbf{R}_s^{(fb)}) = \text{rank}(\mathbf{G}) = \min(K, 2L_x L_y)$. As such, by combining the forward and backward processing,

$$K \leq 2L_x L_y \quad (54)$$

coherent sources can be detected. That is, the total number of subarrays is required to be at least half the number of coherent sources, which is the necessary condition. And the sufficient condition for detecting K coherent sources, which shares same ν values for $\eta_x - 1$ DOAs and same μ values for $\eta_y - 1$ DOAs is $L_x \geq \frac{\eta_x - 1}{2}$ and $L_y \geq \frac{\eta_y - 1}{2}$.

D. Computational Complexity Analysis

In this subsection, we compare the computational complexity of the FBTR and the FBSS methods. After the coherent covariance tensor is computed, the computational complexity for the decorrelation process of the two methods is as follows:

- 1) For the FBTR strategy, the decorrelation process consists of two main steps, namely
 - Forming matrices $\mathbf{R}^{(f)}$ and $\mathbf{R}^{(b)}$ by fixing the first two indices of the coherent covariance tensor \mathcal{R} ;
 - Constructing a slice of the decorrelated covariance tensor, $\mathcal{D}(\tilde{m}, :, \tilde{m}', :)$, using the rows of $\mathbf{R}^{(f)}$ and $\mathbf{R}^{(b)}$, as described in Eq. (16).

These operations involve only slicing and reshaping of the tensor, without requiring any additional arithmetic computations.

- 2) For the FBSS method, obtaining the forward and backward covariance matrices $\mathbf{R}^{(f)} \in \mathbb{C}^{M_x M_y \times M_x M_y}$ and $\mathbf{R}^{(b)} \in \mathbb{C}^{M_x M_y \times M_x M_y}$ requires $L_x L_y$ matrix additions. This results in a computational complexity of $\mathcal{O}(L_x L_y M_x^2 M_y^2)$ additions.

Fig. 3 illustrates the MATLAB runtime required for decorrelation using the FBTR and FBSS strategies. The simulations

were performed on a MacBook Pro equipped with an Apple M1 chip and 16 GB of RAM. It is clear that the FBTR is significantly more computationally efficient than the FBSS. In particular, as the number of antennas increases, the runtime gap becomes more significant because FBSS require summations over higher-dimensional data, whereas the TR-based methods involve no arithmetic operations for decorrelation.

VI. DOF ANALYSIS AND OPTIMAL ARRAY DESIGN

A. Optimization Problem for DOF Analysis

In order to use the FB smoothed covariance matrix in a subspace-based DOA estimation algorithm, e.g., MUSIC, to detect K signals, the necessary condition is

$$M_x M_y \geq K + 1, \quad (55)$$

while the sufficient condition is $M_x \geq K + 1$ and $M_y \geq K + 1$. Based on these necessary and sufficient conditions of subarray size and the number of subarrays, at least $(9/4)K^2$ sensors are required when employing forward-backward SS with a URA. The work in [38] further proposes a minimal array structure using an L-shape array with L-shape subarray grouping, which satisfies the above conditions and reduces the number of required sensors to $K^2 + 4K - 2$ for detecting K coherent sources. This design provides the optimal solution for the worst-case scenario, i.e., when either all μ_k 's or all ν_k 's are identical. However, in many practical cases, only a subset of μ_k 's or ν_k 's are identical. In such scenarios, the minimal L-shape array of [38] becomes over-designed, leading to more sensors than necessary. Moreover, due to the inherent L-shape geometry, backward smoothing cannot be exploited, and thus the full benefit of forward-backward smoothing is not realized.

In this paper, we generalize the concept by formulating an optimization problem to jointly determine the subarray size and the number of subarrays in a URA. Unlike the worst-case design [38], our approach introduces bounds on these parameters to explicitly account for the maximum allowable number of identical μ_k 's or ν_k 's among coherent sources. As a result, the proposed method requires that the subarray size and the number of subarrays satisfy the necessary conditions (54) and (55), together with additional constraints that allow for a specific number of μ or ν equality. Specifically, if several groups of coherent sources share the same ν_k 's with the maximum group size $\eta_x - 1$, and similarly several groups share the same μ_k 's with the maximum group size $\eta_y - 1$, then the following conditions must hold: $M_x \geq \eta_x + 1$, $M_y \geq \eta_y + 1$, $L_x \geq (\eta_x - 1)/2$, and $L_y \geq (\eta_y - 1)/2$. By exploiting this flexibility, the proposed optimization framework avoids overdesign, enables the use of full forward-backward smoothing, and achieves significant reduction in the required number of sensors compared to the worst-case design in [38]. Then, the total number of sensors of the URA is

$$\begin{aligned} N &= (L_x + M_x - 1)(L_y + M_y - 1) \\ &= (L_x L_y + M_x M_y + 1) - (L_x + L_y + M_x + M_y) + \\ &\quad (L_x M_y + L_y M_x). \end{aligned} \quad (56)$$

To use a minimum number of sensors for detecting K sources, the following objective function needs to be minimized:

$$\begin{aligned} \min_{L_x, L_y, M_x, M_y} \quad & f = -(L_x + L_y + M_x + M_y) \\ & + (L_x M_y + L_y M_x) \\ \text{s.t.} \quad & L_x L_y \geq \frac{K}{2}, \quad M_x M_y \geq K + 1, \\ & L_x \geq \zeta_x, \quad L_y \geq \zeta_y, \quad M_x \geq \eta_x, \quad M_y \geq \eta_y, \\ & L_x, L_y, M_x, M_y \in \mathbb{Z}_+, \end{aligned} \quad (57)$$

where $\eta_x \geq 2$ and $\eta_y \geq 2$ are predefined constants that represent one greater than the maximum number of identical ν 's and μ 's, respectively, and also determine the subarray apertures, while $\zeta_x = \frac{\eta_x - 1}{2}$ and $\zeta_y = \frac{\eta_y - 1}{2}$.

B. Relaxed Real-Valued Solutions

Because the optimization problem (57) involves integer variables and thus is difficult to directly tackle, we first obtain a real-valued solution and then perform numerical optimization for integer solutions.

The real-valued solution for (57) considering $\eta_y \leq \eta_x$ is obtained in Appendix B of the Supplement and is expressed as

$$\begin{aligned} (L_x, L_y, M_x, M_y) &= \\ & \begin{cases} (\zeta_x, \zeta_y, \eta_x, \eta_y), & \text{if } \zeta_x \zeta_y > \frac{C-1}{2}, \\ (l_x, l_y, \eta_x, \eta_y), & \text{if } \zeta_x \zeta_y \leq \frac{C-1}{2} \text{ and } \\ & \eta_x \eta_y > C, \\ \left(\sqrt{\frac{K(C-\eta_y)}{2\eta_y(\eta_y-1)}}, \sqrt{\frac{K\eta_y(\eta_y-1)}{2(C-\eta_y)}}, \frac{C}{\eta_y}, \eta_y \right), & \text{if } \eta_x \eta_y \leq C, \end{cases} \end{aligned} \quad (58)$$

with

$$(l_x, l_y) = \begin{cases} \left(\sqrt{\frac{K(\eta_x-1)}{2(\eta_y-1)}}, \sqrt{\frac{K(\eta_y-1)}{2(\eta_x-1)}} \right), & \text{if } \sqrt{\frac{K(\eta_x-1)}{2(\eta_y-1)}} \geq \zeta_x \text{ and } \\ & \sqrt{\frac{K(\eta_y-1)}{2(\eta_x-1)}} \geq \zeta_y, \\ \left(\frac{K}{2\zeta_y}, \zeta_y \right), & \text{otherwise,} \end{cases} \quad (59)$$

where $C = K + 1$. For $\eta_y > \eta_x$, the solution for the case $\eta_x \eta_y \leq C$ is symmetric to that of $\eta_y \leq \eta_x$. Note that, in the sequel, we will use L_x, L_y, M_x, M_y, C to denote the real-valued solutions, and $\bar{L}_x, \bar{L}_y, \bar{M}_x, \bar{M}_y, \bar{C}$ for their integer counterparts.

We can also find the number of DOFs, i.e., $K = C - 1$ from a given number of sensors N , by substituting the optimal solution from Eq. (58) into Eq. (56) as

$$C = \begin{cases} 1 + 2 \left(\sqrt{N} - 2\sqrt{\zeta_x \zeta_y} \right)^2, & \text{if } \zeta_x \zeta_y \leq \frac{C-1}{2}, \eta_x \eta_y \geq C, L_x \geq \zeta_x, \\ & \text{and } L_y \geq \zeta_y, \\ 1 + \frac{2}{3}N - 4\zeta_x \zeta_y, & \text{if } \zeta_x \zeta_y \leq \frac{C-1}{2}, \eta_x \eta_y \geq C, \text{ and} \\ & \text{otherwise,} \\ 1 + \frac{(2N+1)^2}{8N}, & \text{if } C \geq \eta_x \eta_y \text{ and } \eta_y = 2, \\ -\frac{\beta - \sqrt{\beta^2 - 4\alpha\gamma}}{2\alpha}, & \text{if } C \geq \eta_x \eta_y \text{ and } \eta_y > 2, \end{cases} \quad (60)$$

where $\alpha = \left(\frac{1}{\eta_y} - \frac{1}{2}\right)^2$, $\beta = 2\left(\frac{1}{\eta_y} - \frac{1}{2}\right)b - 2N$, and $\gamma = b^2 + 2N$, $b = N + \eta_y - \frac{3}{2}$.

C. Determination of Optimal Integer Solutions

In practice, integer solutions are required for L_x, L_y, M_x, M_y , and N . Denote $\bar{L}_x, \bar{L}_y, \bar{M}_x, \bar{M}_y, \bar{C}$, and \bar{N} as their integer solutions. A 4D brute-force search for the integer combination $(\bar{L}_x, \bar{L}_y, \bar{M}_x, \bar{M}_y)$ for a large array would be extremely computationally expensive. A naïve alternative approach is

Algorithm 1: Algorithm for finding integer solutions $(\bar{L}_x, \bar{L}_y, \bar{M}_x, \bar{M}_y)$

Input : $C, L_x, L_y, \eta_x, \eta_y, \delta$
Output: $(\bar{L}_x, \bar{L}_y, \bar{M}_x, \bar{M}_y), \bar{C},$ and \bar{N}

- 1 Set $L_{x,\min} = \lceil \frac{\eta_x - 1}{2} \rceil, L_{y,\min} = \lceil \frac{\eta_y - 1}{2} \rceil,$
 $M_{x,\min} = \lceil \eta_x \rceil, M_{y,\min} = \lceil \eta_y \rceil$
- 2 Construct integer search neighborhoods around the parameters spanning $\pm\delta$ as
 $\mathcal{L}_x = \max(L_{x,\min}, \text{round}(L_x) - \delta) : \text{round}(L_x) + \delta,$
and similarly for $\mathcal{L}_y, \mathcal{M}_x, \mathcal{M}_y.$
- 3 Initialize candidate list $\mathcal{S} = \emptyset$
- 4 **for each** $\bar{L}_x \in \mathcal{L}_x, \bar{L}_y \in \mathcal{L}_y, \bar{M}_x \in \mathcal{M}_x, \bar{M}_y \in \mathcal{M}_y$
do
- 5 $N_q = (\bar{L}_x + \bar{M}_x - 1)(\bar{L}_y + \bar{M}_y - 1)$
- 6 $C_{\text{cand}} = \min(\bar{M}_x \bar{M}_y, 2\bar{L}_x \bar{L}_y + 1)$
- 7 Append $(N_q, C_{\text{cand}}, \bar{L}_x, \bar{L}_y, \bar{M}_x, \bar{M}_y)$ to $\mathcal{S}.$
- 8 Sort \mathcal{S} lexicographically in ascending order by
keys $\{1\{N_q >$
 $N\}, |N_q - N|, -C_{\text{cand}}, -\bar{L}_x, -\bar{M}_x, \bar{L}_y, \bar{M}_y\}$
- 9 Set the first element after sorting as
 $(\bar{L}_x, \bar{L}_y, \bar{M}_x, \bar{M}_y, \bar{C}, \bar{N})$
- 10 **end for**
- 11 **return** $(\bar{L}_x, \bar{L}_y, \bar{M}_x, \bar{M}_y, \bar{C}, \bar{N})$

to round up the real-valued solutions as $\bar{L}_x = \lceil L_x \rceil, \bar{L}_y = \lceil L_y \rceil, \bar{M}_x = \lceil M_x \rceil, \bar{M}_y = \lceil M_y \rceil.$ This approach incurs negligible computational complexity. However, the resulting integer solutions are oversized and therefore lead to a larger-than-necessary number of antennas, as we demonstrate later.

In the following, we present an efficient approach to obtain a near-optimal integer solution with a low computational complexity. This approach uses the real-valued KKT solutions as a starting point and applies a local search and rounding procedure. In particular, we introduce a local search parameter δ to define the size of the search region, resulting in a search space of at most $(2\delta + 1)^4$ integers. A typical value of δ may range between 1 and 2. As such, the overall complexity is significantly lower than that of a brute-force approach.

In summary, to determine the number of DOFs for a total N antennas, we first obtain the real-valued solutions for (L_x, L_y, M_x, M_y) and C using Eqs. (58) and (60). Using these solutions as the initial value and presetting a search parameter δ , we then employ Algorithm 1 to find an integer solution for $(\bar{L}_x, \bar{L}_y, \bar{M}_x, \bar{M}_y)$ along with \bar{C} . Note that $\bar{C} \leq C$ holds due to the stricter constraints for the integer solutions.

D. Cramer-Rao Bound (CRB) for 2D DOA Estimation

In this subsection, we derive the CRB expression for 2D DOA estimation, which serves as a performance bound for comparing the DOA estimation accuracy of different methods in the sequel. The CRB expression is derived under the assumptions that the columns of the array manifold matrix are linearly independent and the noise components are spatially and temporarily uncorrelated. For a sufficiently large number of snapshots T , the CRB can be expressed as [58]

$$\text{CRB} = \mathbf{F}^{-1}, \quad (61)$$

where \mathbf{F} denotes the Fisher information matrix, expressed as

$$\mathbf{F} = \frac{2T}{\sigma_n^2} \left(\text{Re} \left[\mathbf{D}^H \mathbf{P}_\perp \mathbf{D} \odot \mathbf{R}_s^T \right] \right), \quad (62)$$

$\mathbf{P}_\perp = \mathbf{I} - \mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H$ is the projection matrix, \mathbf{R}_s is the source covariance matrix, and \mathbf{D} is the matrix containing derivatives of steering vectors with respect to the parameters of interest.

In the underlying DOA estimation problem, the parameters of interest are azimuth and elevation angles (ϕ, θ) for K sources. Define the matrices \mathbf{D}_θ and \mathbf{D}_ϕ as

$$\begin{aligned} \mathbf{D}_\theta &= \left[\frac{\partial \mathbf{a}(\mu_1, \nu_1)}{\partial \theta_1}, \dots, \frac{\partial \mathbf{a}(\mu_K, \nu_K)}{\partial \theta_K} \right], \\ \mathbf{D}_\phi &= \left[\frac{\partial \mathbf{a}(\mu_1, \nu_1)}{\partial \phi_1}, \dots, \frac{\partial \mathbf{a}(\mu_K, \nu_K)}{\partial \phi_K} \right], \end{aligned} \quad (63)$$

where $\mathbf{a}(\mu_k, \nu_k) = \mathbf{a}(\mu_k) \otimes \mathbf{a}(\nu_k).$ Stacking these matrices yields $\mathbf{D} = [\mathbf{D}_\theta \ \mathbf{D}_\phi],$ which results in

$$\mathbf{D}^H \mathbf{P}_\perp \mathbf{D} = \begin{bmatrix} \mathbf{D}_\theta^H \mathbf{P}_\perp \mathbf{D}_\theta & \mathbf{D}_\theta^H \mathbf{P}_\perp \mathbf{D}_\phi \\ \mathbf{D}_\phi^H \mathbf{P}_\perp \mathbf{D}_\theta & \mathbf{D}_\phi^H \mathbf{P}_\perp \mathbf{D}_\phi \end{bmatrix}. \quad (64)$$

Therefore, the Fisher information matrix becomes

$$\mathbf{F} = \frac{2T}{\sigma_n^2} \begin{bmatrix} \mathbf{F}_\theta & \mathbf{F}_{\theta\phi} \\ \mathbf{F}_{\phi\theta} & \mathbf{F}_\phi \end{bmatrix}, \quad (65)$$

where

$$\begin{aligned} \mathbf{F}_\theta &= \text{Re} \left[\mathbf{D}_\theta^H \mathbf{P}_\perp \mathbf{D}_\theta \odot \mathbf{R}_s^T \right], \\ \mathbf{F}_\phi &= \text{Re} \left[\mathbf{D}_\phi^H \mathbf{P}_\perp \mathbf{D}_\phi \odot \mathbf{R}_s^T \right], \\ \mathbf{F}_{\theta\phi} &= \text{Re} \left[\mathbf{D}_\theta^H \mathbf{P}_\perp \mathbf{D}_\phi \odot \mathbf{R}_s^T \right], \\ \mathbf{F}_{\phi\theta} &= \text{Re} \left[\mathbf{D}_\phi^H \mathbf{P}_\perp \mathbf{D}_\theta \odot \mathbf{R}_s^T \right]. \end{aligned} \quad (66)$$

Finally, the CRB can be obtained by substituting these results to Eq. (61).

VII. SIMULATION RESULTS*A. Optimal Array Configuration*

In the first example, we consider a URA with $N = 25$ antennas and $\eta_x = \eta_y = 3,$ implying that up to two DOAs may share the same μ or ν value. To determine the optimal array configuration, we first compute the real-valued solution $C = K + 1,$ which is found to be $C = 12.8643,$ and obtain $(L_x, L_y, M_x, M_y) = (3.1229, 1.8995, 4.2881, 3)$ according to Eqs. (60) and (58), as listed in Table I. However, since integer solutions are required for these parameters, we utilize the real-valued $C, L_y,$ and η_y in Algorithm 1 with local search parameter $\delta = 2$ to obtain the integer solutions as $(\bar{L}_x, \bar{L}_y, \bar{M}_x, \bar{M}_y) = (2, 3, 4, 3)$ and $\bar{C} = 12.$ Note in this case that $\bar{C} = \lfloor C \rfloor,$ and the number of DOFs is thus $K = 11.$ Using this solution, the numbers of antennas in the X- and Y-axis directions are $\bar{N}_x = \bar{L}_x + \bar{M}_x - 1 = 5$ and $\bar{N}_y = \bar{L}_y + \bar{M}_y - 1 = 5,$ respectively. Under this configuration, the FBTR-mat approach yields the same number of DOFs, i.e., 11, as obtained from Eq. (35). Moreover, $\lfloor (N_x + 1)/2 \rfloor + p = 4$ and $\lfloor (N_y + 1)/2 \rfloor = 3,$ supporting up to three identical ν values but at most two identical μ values. We consider 11 DOAs

Method	Subarray arrangement (L_x, L_y, M_x, M_y)	Number of antennas required	Number of DOFs
Solution for $N = 25$ antennas with $\eta_x = \eta_y = 3$			
KKT solution for real-valued case	(3.1229, 1.8995, 4.2881, 3)	25	11.8643
Round up using ceiling function	(4, 2, 5, 3)	40	14
Optimal integer solutions using Algorithm 1	(2, 3, 4, 3)	25	11
Solution for $N = 49$ antennas with $\eta_x = \eta_y = 4$			
KKT solution for real-valued case	(4.0668, 2.6408, 5.6199, 4)	49	21.4795
Round up using ceiling function	(5, 3, 6, 4)	60	23
Optimal integer solutions using Algorithm 1	(3, 4, 5, 4)	49	19

TABLE I: Comparison of integer solutions and KKT solutions.

Number of antennas	Array arrangement (N_x, N_y)	Subarray arrangement (L_x, L_y, M_x, M_y)	Method	Number of DOFs
$\eta_x = \eta_y = 4$				
90	(18, 5)	NA	FBTR	13
90	(18, 5)	NA	FBTR-mat	35
90	(18, 5)	NA	FTR [55]	10
90	(18, 5)	NA	FTR-mat [55]	26
90	(18, 5)	(9, 2, 10, 4)	FBSS	36
$\eta_x = \eta_y = 6$				
90	(10, 9)	NA	FBTR	10
90	(10, 9)	NA	FBTR-mat	34
90	(10, 9)	NA	FTR [55]	8
90	(10, 9)	NA	FTR-mat [55]	24
90	(10, 9)	(5, 4, 6, 6)	FBSS	35
90	L shape	L shape grouping	Minimal array [38]	7

TABLE II: Comparison of the number of DOFs.

Method	$\eta_x = \eta_y$	Number of Antennas required
Proposed optimal array	2	6
	3	12
	4	25
	5	36
Minimal array	5	30

TABLE III: Comparison of the number of optimal array vs minimal array for 4 sources.

with azimuth–elevation pairs $(90^\circ, 30^\circ)$, $(0^\circ, 30^\circ)$, $(45^\circ, 45^\circ)$, $(-36^\circ, 21^\circ)$, $(-30^\circ, 43^\circ)$, $(36^\circ, 53^\circ)$, $(18^\circ, 31^\circ)$, $(54^\circ, 47^\circ)$, $(4^\circ, 59^\circ)$, $(-18^\circ, 73^\circ)$, $(-52^\circ, 67^\circ)$. Among these, the DOAs, $(90^\circ, 30^\circ)$ and $(45^\circ, 45^\circ)$ share the same ν value of 0.5, while $(0^\circ, 30^\circ)$ and $(45^\circ, 45^\circ)$ share the same μ value of 0.5. Figs. 4(a) and 5(a) show the MUSIC spectra for this configuration using FBSS and FBTR-mat, respectively, where both methods correctly detect all 11 sources, equal to the available DOFs. However, when three identical μ or ν values are present, both methods fail, as shown in Figs. 4(b) and 5(b). In this case, two spurious peaks appear at $(-45^\circ, 45^\circ)$ and $(-60, 90)$, which also correspond to $\mu = 0.5$ in the repeated- μ group.

To handle larger multiplicities, e.g., $\eta_x = \eta_y = 4$, the continuous and integer solutions are $(L_x, L_y, M_x, M_y) = (\bar{L}_x, \bar{L}_y, \bar{M}_x, \bar{M}_y) = (2, 2, 4, 4)$, respectively, with $(N_x, N_y) = (5, 5)$, resulting in $K = 8$ DOFs. In this case, the FBSS method successfully detects all sources,

as shown in Fig. 4(c), while the FBTR-mat still produces two false peaks associated with identical μ values, since $\lfloor (N_y + 1)/2 \rfloor = 3$, allowing only up to two identical μ 's.

When η_x and η_y are reduced, the number of achievable DOFs increases. Considering from pure DOF perspective, if we allow η_x and η_y to be 1, we obtain the highest number of $\lfloor \frac{2}{3} \cdot 25 \rfloor = 16$ DOFs for the uniform linear array case [24]. However, in such a case, the array loses its ability to resolve elevation angles. As such, an appropriate value of η_y should be specified to maintain the elevation subarray aperture.

We further consider a URA with $N = 49$ antennas designed for $\eta_x = \eta_y = 4$. Here, FBSS supports up to three identical μ or ν values. The integer solution is $(\bar{L}_x, \bar{L}_y, \bar{M}_x, \bar{M}_y) = (3, 4, 5, 4)$ with $(N_x, N_y) = (7, 7)$, yielding 19 DOFs for both FBSS and FBTR-mat. In this configuration, $\lfloor (N_x + 1)/2 \rfloor + p = 5$ and $\lfloor (N_y + 1)/2 \rfloor = 4$, confirming support for up to three identical μ or four identical ν groups. To form 19 sources, we extend the previous 11 DOAs by adding $(50^\circ, 25^\circ)$, $(-56^\circ, 33^\circ)$, $(-26^\circ, 65^\circ)$, $(-8^\circ, 17^\circ)$, $(-14^\circ, 57^\circ)$, $(-6^\circ, 37^\circ)$, $(30^\circ, 90^\circ)$, $(0^\circ, 30^\circ)$ to make 19 sources. Figs. 4(d) and 5(d) illustrate that both the FBSS and FBTR-mat approaches successfully detect all sources, with the latter resulting a cleaner spectrum.

Fig. 6 compares the proposed FBTR with CPD and FTR method developed in [55]. For a 25-antenna array with $(N_x, N_y) = (5, 5)$, the proposed method achieves 5 DOFs according to Eq. (30), whereas the method in [55] can resolve at most 4 sources. As shown in Figs. 6(a) and 6(b), the proposed method successfully detects all 5 sources, while the method in [55] fails since the number of sources exceeds its DOF limit. Figs. 6(c) and 6(d) illustrate the case with 7 sources with $(N_x, N_y) = (7, 7)$. Here again, the proposed method detects all signals, whereas the method in [55] fails due to the number of sources being beyond its DOF capability.

B. Comparison of DOFs

Table II compares the number of DOFs among various decorrelation methods. For this comparison, we consider an array of 90 antennas with two different arrangements. The first arrangement is for $\eta_y = 4$, and so, $(\bar{L}_x, \bar{L}_y, \bar{M}_x, \bar{M}_y) = (9, 2, 10, 4)$, resulting in $(\bar{N}_x, \bar{N}_y) = (18, 5)$ obtained from Algorithm 1. In this case, the FTR method [55] achieves 10 DOFs. In contrast, the proposed FBTR strategy attains 13 DOFs as computed from Eq. (30). Furthermore, by matricizing the reconstructed tensor, the number of DOFs is increased to

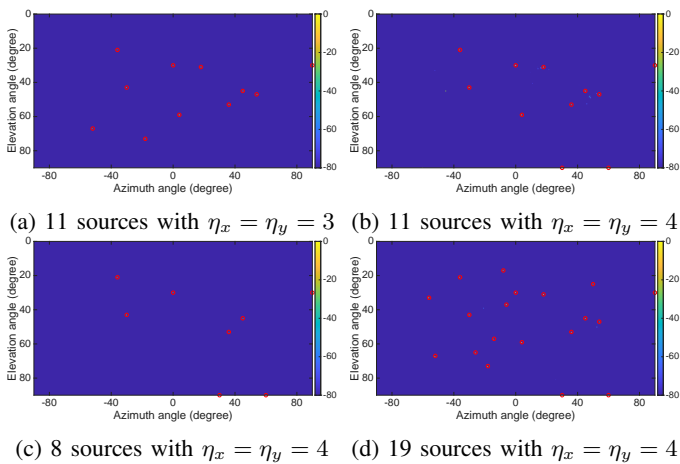


Fig. 4: FBSS-based method

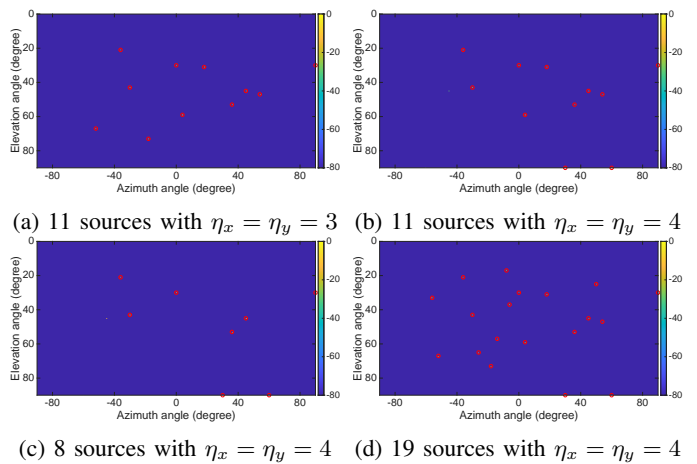


Fig. 5: FBTR-mat-based method

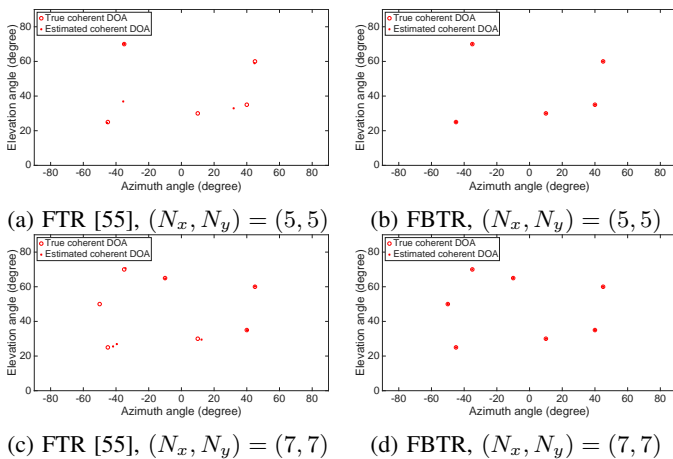
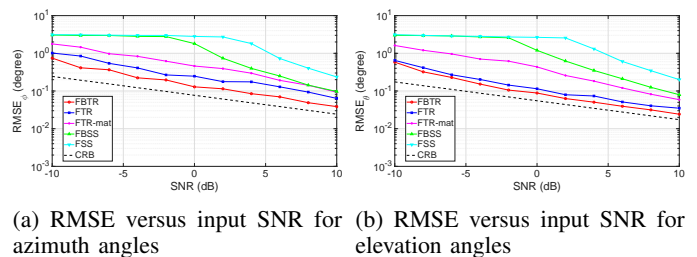


Fig. 6: DOF comparisons of tensor reconstruction-based approaches.

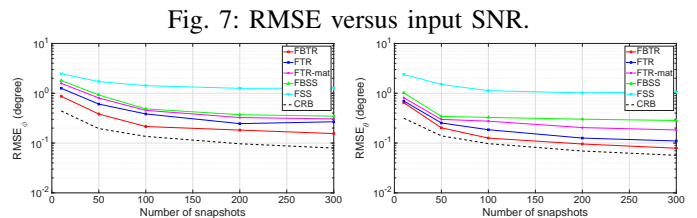
35 using Eq. (35), whereas the method in [35] can detect up to 26 sources. Thus, incorporating backward slices into the tensor reconstruction technique provides an additional 3 DOFs for the CPD-based tensor decomposition approach and 9 additional DOFs for the matricized tensor reconstruction compared with the forward-only counterpart. For the same configuration, the FBSS achieves 36 DOFs.

For a larger value of $\eta_x = \eta_y = 6$, the number of DOFs decreases relative to the $\eta_x = \eta_y = 4$ case. Nevertheless, the proposed method still achieves a higher number of DOFs compared with the method in [55]. In all these scenarios, the minimal array configuration for SS [38] achieves only 7 DOFs, which is significantly lower than the proposed methods due to its worst-case design and the inability to exploit the backward strategy.

To further compare the proposed optimal array with the minimal array [38], we consider 4 sources with different values of η_x and η_y , as listed in Table III. The minimal array is designed solely for the worst case, i.e., $\eta_x = \eta_y = 5$, where all 4 sources have identical μ and ν values. In this case, it requires 30 antennas, whereas the proposed optimal configuration requires 36 antennas. However, for smaller η_x and η_y , the proposed array uses fewer antennas than the



(a) RMSE versus input SNR for azimuth angles (b) RMSE versus input SNR for elevation angles



(a) RMSE versus number of snapshots for azimuth angles (b) RMSE versus number of snapshots for elevation angles

Fig. 7: RMSE versus input SNR.

minimal array. For example, when $\eta_x = \eta_y = 4$, the proposed array requires only 25 antennas.

In summary, the proposed method achieves a higher number of DOFs than existing methods. In particular, the tensor reconstruction-based approaches provide larger DOFs compared with the method in [55]. For SS-based methods, the minimal array in [38] requires the fewest sensors when all sources share identical μ and ν values. However, when a reduced number of identical μ and ν values are allowed, the FBSS with the proposed optimal array design may require fewer sensors than the minimal array.

C. Estimation Accuracy Analysis

In this subsection, we compare the DOA estimation accuracy of both proposed methods. We consider $N = 49$ antennas with $\eta_y = 4$ resulting in the sensor arrangement $(\bar{L}_x, \bar{L}_y, \bar{M}_x, \bar{M}_y) = (3, 4, 5, 4)$. The number of antennas is specified by $(\bar{N}_x, \bar{N}_y) = (7, 7)$. The DOA estimation performance is evaluated by computing the root-mean-squared error (RMSE).

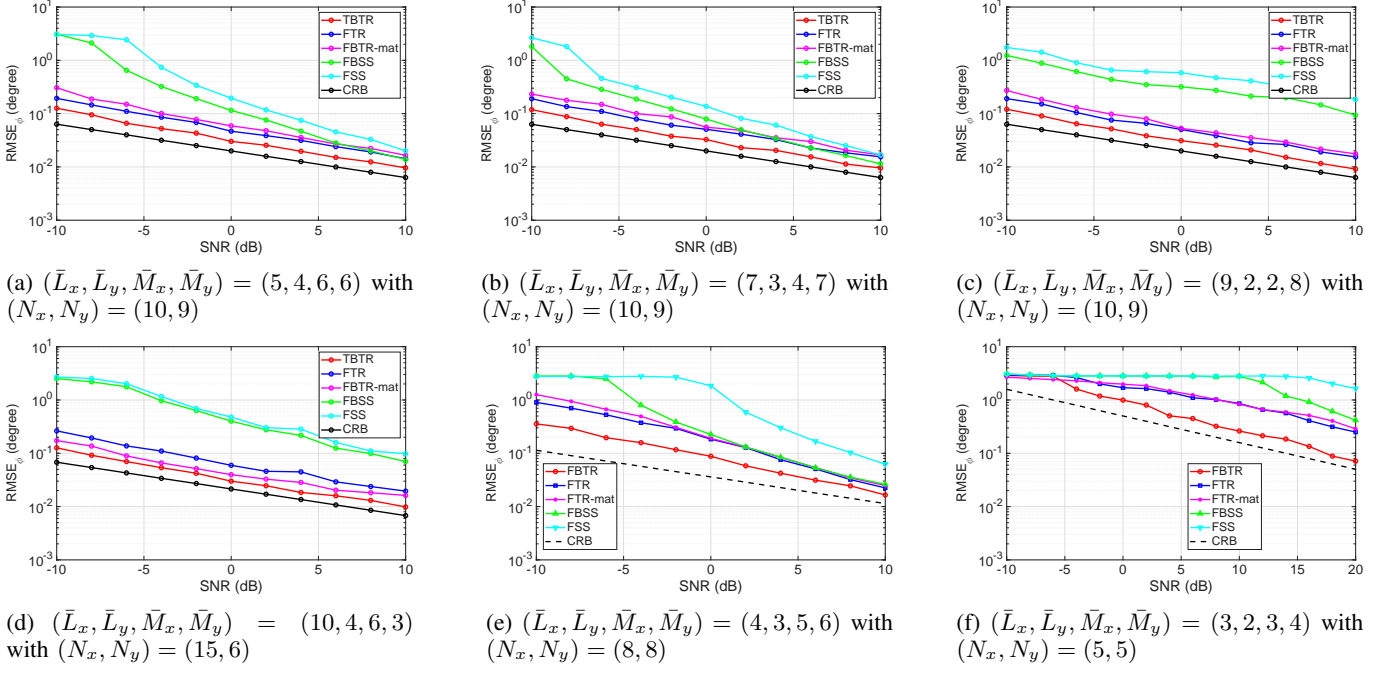


Fig. 9: RMSE versus input SNR for azimuth angles.

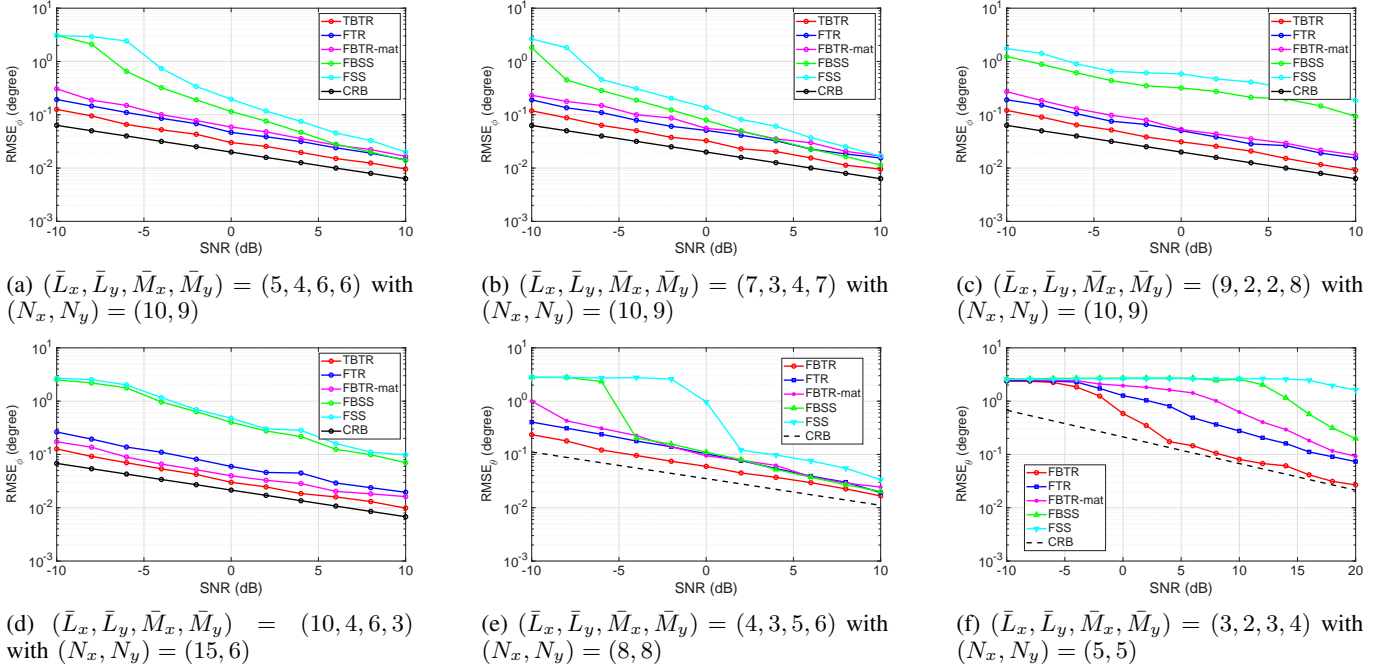


Fig. 10: RMSE versus input SNR for elevation angles.

Fig. 7 depicts the RMSE values against the input SNR. We consider three coherent signals with azimuth and elevation angle pairs $(30^\circ, 45^\circ)$, $(6^\circ, 57^\circ)$, $(42^\circ, 51^\circ)$, and their μ and ν pairs are $(0.61, 0.35)$, $(0.83, 0.088)$, $(0.58, 0.52)$. 1,000 snapshots are assumed, and a total of 100 Monte Carlo trials are performed to compute the RMSE values. From Fig. 7, it is observed that the tensor reconstruction approaches provide significantly better performance than the SS-based approaches across all SNR regions, with the improvement being particularly evident at lower SNRs. This is because

the structural characteristics of the high-dimensional signal tensor are more effectively exploited by tensor decomposition methods such as CPD. The FBTR-mat also achieves lower RMSE values compared with the SS-based approaches. Moreover, the proposed FBTR achieves better performance than the FTR decomposition in [55], due to the enhanced decorrelation capability provided by backward processing. Similarly, FBSS outperforms forward-only SS (FSS). When the input SNR is 10 dB, the RMSE values for azimuth estimation are 0.0385, 0.0636, 0.0931, 0.0968, and 0.2359

for the FBTR, FTR, FBTR-mat, FBSS, and FSS, respectively. For elevation estimation, the corresponding RMSE values are 0.0242, 0.0349, 0.0589, 0.0797, and 0.2007.

Figs. 8(a) and 8(b) show the RMSE values versus the number of snapshots for azimuth and elevation angles, respectively, where the number of snapshots is varied from 10 to 300 with the input SNR fixed at 5 dB. Similar to Fig. 7, tensor reconstruction approaches outperform the existing methods, with the proposed FBTR achieving lowest RMSE values compared to the other methods.

We also compared the performance of the different methods against the CRB as a theoretical lower bound. Note that the CRB expression in Eq. (61) is derived for the underlying URA and does not require the sources to be uncorrelated because coherence is captured through the source covariance matrix. Specifically, the CRB depends only on the array geometry, the source covariance matrix before decorrelation, the DOAs, and the noise power, all of which are common to the decorrelation strategies considered in this paper. From Figs. 6 and 7, it can be observed that the proposed FBTR method remains consistently closer to the CRB than the other methods.

To further evaluate the generalizability and robustness of the proposed methods, we conduct additional simulations by 1) fixing the total number of antennas while varying the array geometry, and 2) varying both the total number of antennas and their arrangements. For each antenna arrangement, we generated random DOAs for three sources. Figs. 9 and 10 illustrate the RMSE versus input SNR for azimuth and elevation angle estimation, respectively. The first four plots (Figs. 9(a)–9(d) and 10(a)–10(d)) all use 90 antennas. We randomly select four geometries: $(L_x, L_y, M_x, M_y) = (5, 4, 6, 6)$, $(7, 3, 4, 7)$, $(8, 2, 2, 8)$, and $(10, 4, 6, 3)$. The first three arrangements yield $(N_x, N_y) = (10, 9)$, while the last arrangement gives $(N_x, N_y) = (15, 6)$. In all cases, the proposed FBTR method consistently outperforms the other approaches, and overall the tensor-based methods outperform the spatial smoothing-based methods, especially at low SNR. The performance gap is particularly pronounced for the third and fourth configurations with $(8, 2, 2, 8)$ and $(10, 4, 6, 3)$ because their subarrays have a very small aperture in one of the two dimensions (e.g., $M_x = 2$ in the third arrangement and $M_y = 3$ in the fourth arrangement), thus restricting their capability to effectively capture 2D angular information. In contrast, the tensor-based methods do not rely on subarray formation and thus are not affected by such aperture imbalance. The last two plots, Figs. 9(e)–9(f) and 10(e)–10(f), use different numbers of antennas, $(\bar{L}_x, \bar{L}_y, \bar{M}_x, \bar{M}_y) = (4, 3, 5, 6)$ and $(3, 2, 3, 4)$, which yield $(N_x, N_y) = (8, 8)$ and $(5, 5)$, respectively. In these cases, the proposed FBTR method again outperforms all other existing methods and remains closest to the CRB.

D. Connection between Number of DOFs and DOA Estimation Accuracy

In this subsection, we discuss the connection between the number of DOFs and DOA estimation accuracy. Since no single universal relationship exists between them, we separately examine the following cases:

1) *Tensor versus matrix-based methods for a fixed array configuration*: Tensor reconstruction-based methods such as FBTR and FTR offer better DOA estimation accuracy, as illustrated in Figs. 7–10. This is because tensor modeling of the higher-dimensional signal preserves its structural characteristics, and tensor decomposition techniques such as CPD exploit this structure to estimate the DOAs. In contrast, matrix-based approaches such as FBSS and FBTR-mat merge the dimensional information, i.e., the steering vectors, thereby losing the structural characteristics.

In terms of DOFs, tensor reconstruction-based approaches provide fewer DOFs than matrix-based methods, since the latter merge the steering vectors from different dimensions into enlarged composite vectors, thereby increasing the overall DOFs. This distinction in DOF enhancement is evident when comparing Eqs. (25) and (34). For the tensor-based FBTR method, the total number of DOFs is obtained by adding the DOFs along the X and Y axes, whereas the matrix-based FBTR-mat formulation yields a multiplicative combination of these DOFs.

2) *Forward-only versus forward-backward case for a fixed array configuration and decorrelation strategy*: For a fixed array configuration and a fixed tensor- or matrix-modeling framework, forward-backward approaches provide both a higher number of DOFs and improved estimation accuracy compared to their forward-only counterparts, as shown in Figs. 7–10. This advantage arises because forward-backward methods, such as FBTR, utilize larger effective tensor slices than FTR, while FBSS employs larger effective subarrays than FSS. These enlarged dimensions increase the number of DOFs and improve the DOA estimation accuracy.

3) *Different antenna configuration for a fixed number of antennas and fixed decorrelation strategy*: For a fixed number of antennas, the solution to the optimization problem in Eq. (57) yields the maximum achievable number of DOFs. From the DOA estimation accuracy perspective, the performance of the TR-based methods remains robust as long as the array configuration is not extremely unbalanced, i.e., when one dimension has many sensors while the other has very few. In contrast, SS-based methods require reasonably balanced subarray dimensions, and highly unbalanced subarray configurations lead to noticeable performance degradation as shown in Figs. 9(c), 9(d), 10(c), and 10(d). This degradation occurs because balanced arrays offer sufficient aperture in both dimensions, enabling more effective capture of both azimuth and elevation information.

4) *Complementarity of the proposed strategies*: In all cases, forward-backward strategies should be adopted because they improve both the DOFs and the DOA estimation accuracy. For a fixed number of antenna elements, when the number of coherent sources is relatively small, such as localizing a few coherent sources or a few dominant propagation paths with high accuracy in radar or wireless communication, the FBTR method is more suitable. In contrast, when the number of sources is larger, such as in resolving many multipath components in rich scattering wireless environments, FBSS or FBTR-mat with the proposed optimal array arrangement becomes more appropriate.

VIII. CONCLUSION

This paper addressed the rank deficiency problem of the covariance tensor obtained from a URA for mutually coherent sources, which poses a significant challenge for conventional DOA estimation methods. To overcome this issue, we proposed two decorrelation strategies, one based on FBTR and the other based on FBSS. For the FBSS approach, we derived an optimal array arrangement that specifies both the number of antennas in each subarray and the number of subarrays along the X and Y axes, maximizing the number of DOFs for a given number of antennas. This optimal configuration is applied to both proposed strategies to achieve the highest number of DOFs for a particular number of antennas. Both strategies incorporate forward and backward processing to enhance the DOFs. When combined with the optimal array arrangement, the FBSS approach achieves the highest number of DOFs among existing methods for 2D coherent signals. In contrast, the FBTR method provides improved DOA estimation performance by exploiting tensor decomposition technique. Furthermore, the matricized FBTR, i.e., FBTR-mat offers a more computationally efficient decorrelation process than FBSS while achieving a similar number of DOFs.

REFERENCES

- [1] D. H. Johnson, *Array Signal Processing: Concepts and Techniques*. Prentice-Hall, 1993.
- [2] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. Wiley, 2002.
- [3] S. van der Tol, B. D. Jeffs, and A.-J. van der Veen, "Self-calibration for the LOFAR radio astronomical array," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4497–4510, 2007.
- [4] T. E. Tuncer and B. Friedlander, *Classical and Modern Direction-of-Arrival Estimation*. Academic Press, 2009.
- [5] M. G. Amin, X. Wang, Y. D. Zhang, F. Ahmad, and E. Aboutanios, "Sparse array and sampling for interference mitigation and DOA estimation in GNSS," *Proc. IEEE*, vol. 104, no. 6, pp. 1302–1317, June 2016.
- [6] S. Sun and Y. D. Zhang, "4D automotive radar sensing for autonomous vehicles: A sparsity-oriented approach," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 4, pp. 879–891, 2021.
- [7] C. Zhou, Y. Gu, Y. D. Zhang, and Z. Shi, "Sparse array interpolation for direction-of-arrival estimation," in *Sparse Arrays for Radar, Sonar, and Communications*, M. G. Amin, Ed. Wiley-IEEE Press, 2024.
- [8] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propagat.*, vol. 34, no. 3, pp. 276–280, 1986.
- [9] 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, 1989.
- [10] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 5, no. 2, pp. 4–24, 1988.
- [11] J. Litva and T. K. Lo, *Digital Beamforming in Wireless Communications*. Artech House, 1996.
- [12] J. Li and P. Stoica, *Robust Adaptive Beamforming*. Wiley, 2005.
- [13] 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, 2005.
- [14] Z.-M. Liu, Z.-T. Huang, and Y.-Y. Zhou, "An efficient maximum likelihood method for direction-of-arrival estimation via sparse bayesian learning," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 1–11, 2012.
- [15] Y. D. Zhang, M. G. Amin, and B. Himed, "Sparsity-based doa estimation using co-prime arrays," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2013, pp. 3967–3971.
- [16] S. Qin, Y. D. Zhang, and M. G. Amin, "Generalized coprime array configurations for direction-of-arrival estimation," *IEEE Trans. Signal Process.*, vol. 63, no. 6, pp. 1377–1390, 2015.
- [17] C. Zhou, Y. Gu, X. Fan, Z. Shi, G. Mao, and Y. D. Zhang, "Direction-of-arrival estimation for coprime array via virtual array interpolation," *IEEE Trans. Signal Process.*, vol. 66, no. 22, pp. 5956–5971, 2018.
- [18] S. Liu, Z. Mao, Y. D. Zhang, and Y. Huang, "Rank minimization-based toeplitz reconstruction for doa estimation using coprime array," *IEEE Commun. Lett.*, vol. 25, no. 7, pp. 2265–2269, 2021.
- [19] M. Zoltowski and F. Haber, "A vector space approach to direction finding in a coherent multipath environment," *IEEE Trans. Antennas Propagat.*, vol. 34, no. 9, pp. 1069–1079, 1986.
- [20] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Signal Process.*, vol. 39, no. 5, pp. 1110–1121, 1991.
- [21] M. Viberg, B. Ottersten, and T. Kailath, "Detection and estimation in sensor arrays using weighted subspace fitting," *IEEE Trans. Antennas Propagat.*, vol. 39, no. 11, pp. 2436–2449, 1991.
- [22] Z. Yang and X. Chen, "Maximum likelihood direction-of-arrival estimation via rank-constrained ADMM," in *Proc. CIE Int. Conf. Radar*, Haikou, China, 2021, pp. 2376–2380.
- [23] J. E. Evans, J. R. Johnson, and D. Sun, "Application of advanced signal processing techniques to angle of arrival estimation in ATC navigation and surveillance systems," *MIT Lincoln Lab. Tech. Rep.*, 1982.
- [24] S. U. Pillai and B. H. Kwon, "Forward/backward spatial smoothing techniques for coherent signal identification," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 1, pp. 8–15, 1989.
- [25] Z. Yang, P. Stoica, and J. Tang, "Source resolvability of spatial-smoothing-based subspace methods: A Hadamard product perspective," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2543–2553, 2019.
- [26] W. Du and R. L. Kirlin, "Improved spatial smoothing techniques for doa estimation of coherent signals," *IEEE Trans. Signal Process.*, vol. 39, no. 5, pp. 1208–1210, 2002.
- [27] M. Dong, S. Zhang, X. Wu, and H. Zhang, "A high resolution spatial smoothing algorithm," in *Proc. Int. Symp. Microw., Antenna, Propagat., EMC Techno. Wireless Commun.*, 2007, pp. 1031–1034.
- [28] J. Pan, M. Sun, Y. Wang, and X. Zhang, "An enhanced spatial smoothing technique with ESPRIT algorithm for direction of arrival estimation in coherent scenarios," *IEEE Trans. Signal Process.*, vol. 68, pp. 3635–3643, 2020.
- [29] F.-M. Han and X.-D. Zhang, "An ESPRIT-like algorithm for coherent DOA estimation," *IEEE Antennas Wireless Propagat. Lett.*, vol. 4, pp. 443–446, 2005.
- [30] S. R. Pavel, Y. D. Zhang, and B. Himed, "Structured decorrelation of covariance matrix for DOA estimation of coherent signals," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2025, pp. 2087–2091.
- [31] Y.-H. Choi, "ESPRIT-based coherent source localization with forward and backward vectors," *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 6416–6420, 2010.
- [32] S. R. Pavel and Y. D. Zhang, "Direction-of-arrival estimation of mixed coherent and uncorrelated signals," *IEEE Signal Process. Lett.*, vol. 31, pp. 2180–2184, 2024.
- [33] P. Heidenreich, A. M. Zoubir, and M. Rubsamen, "Joint 2-D DOA estimation and phase calibration for uniform rectangular arrays," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4683–4693, 2012.
- [34] J. Li, X. Zhang, and H. Chen, "Improved two-dimensional DOA estimation algorithm for two-parallel uniform linear arrays using propagator method," *Signal Process.*, vol. 92, no. 12, pp. 3032–3038, 2012.
- [35] Z. Zheng and S. Mu, "Two-dimensional DOA estimation using two parallel nested arrays," *IEEE Commun. Lett.*, vol. 24, no. 3, pp. 568–571, 2019.
- [36] S. Qin, Y. D. Zhang, and M. G. Amin, "Improved two-dimensional DOA estimation using parallel coprime arrays," *Signal Process.*, vol. 172, no. 107428, pp. 1–9, 2020.
- [37] C.-C. Yeh, J.-H. Lee, and Y.-M. Chen, "Estimating two-dimensional angles of arrival in coherent source environment," *IEEE Trans. Acoust., speech, and signal process.*, vol. 37, no. 1, pp. 153–155, 2002.
- [38] Y.-M. Chen, "On spatial smoothing for two-dimensional direction-of-arrival estimation of coherent signals," *IEEE Trans. Signal Process.*, vol. 45, no. 7, pp. 1689–1696, 1997.
- [39] H. Yi and X. Zhou, "On 2D forward-backward spatial smoothing for azimuth and elevation estimation of coherent signals," in *Proc. IEEE Antennas Propagat. Society Int. Symp.*, vol. 2, 2005, pp. 80–83.
- [40] Y. Hua, "Estimating two-dimensional frequencies by matrix enhancement and matrix pencil," *IEEE Trans. Signal Process.*, vol. 40, no. 9, pp. 2267–2280, 1992.

- [41] F.-J. Chen, S. Kwong, and C.-W. Kok, "Esprit-like two-dimensional DOA estimation for coherent signals," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 3, pp. 1477–1484, 2010.
- [42] J. D. Carroll and J.-J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, 1970.
- [43] R. A. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multimodal factor analysis," *UCLA Work. Papers Phonetics*, vol. 16, pp. 1–84, 1970.
- [44] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.
- [45] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, 2000.
- [46] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Trans. Signal Process.*, vol. 48, no. 8, pp. 2377–2388, 2000.
- [47] M. Boizard, G. Ginolhac, F. Pascal, S. Miron, and P. Forster, "Numerical performance of a tensor MUSIC algorithm based on HOSVD for a mixture of polarized sources," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2013, pp. 1–5.
- [48] C.-L. Liu and P. Vaidyanathan, "Tensor MUSIC in multidimensional sparse arrays," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, 2015, pp. 1783–1787.
- [49] M.-Y. Cao, X. Mao, X. Long, and L. Huang, "Tensor approach to DOA estimation of coherent signals with electromagnetic vector-sensor array," *Sensors*, vol. 18, no. 4320, pp. 1–22, 2018.
- [50] H. Zheng, C. Zhou, Z. Shi, Y. Gu, and Y. D. Zhang, "Coarray tensor direction-of-arrival estimation," *IEEE Trans. Signal Process.*, vol. 71, pp. 1128–1142, 2023.
- [51] F. Wen and H. C. So, "Tensor-MODE for multi-dimensional harmonic retrieval with coherent sources," *Signal Process.*, vol. 108, pp. 530–534, Mar. 2015.
- [52] W. Sun, H. C. So, F. K. W. Chan, and L. Huang, "Tensor approach for eigenvector-based multi-dimensional harmonic retrieval," *IEEE Trans. Signal Process.*, vol. 61, no. 13, pp. 3378–3388, 2013.
- [53] X. Wang, W. Wang, J. Liu, Q. Liu, and B. Wang, "Tensor-based real-valued subspace approach for angle estimation in bistatic MIMO radar with unknown mutual coupling," *Signal Process.*, vol. 116, pp. 152–158, Nov. 2015.
- [54] Y. Lin, S. Jin, M. Matthaiou, and X. You, "Tensor-based channel estimation for millimeter wave MIMO-OFDM with dual-wideband effects," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4218–4232, 2020.
- [55] H. Zheng, C. Zhou, Z. Shi, and Y. Gu, "Structured tensor reconstruction for coherent DOA estimation," *IEEE Signal Process. Lett.*, vol. 29, pp. 1634–1638, 2022.
- [56] M. S. R. Pavel, Y. D. Zhang, and B. Himed, "Tensor reconstruction-based sparse array interpolation for 2-D DOA estimation of coherent signals," in *Proc. IEEE Radar Conf.*, 2024, pp. 1–6.
- [57] S. R. Pavel, Y. D. Zhang, S. Sun, and A. L. F. de Almeida, "Tensor reconstruction-based sparse array 2-D DOA estimation of mixed coherent and uncorrelated signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2024, pp. 12 876–12 880.
- [58] 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, 2002.

Supplementary Material for “2D DOA Estimation of Coherent Signals Exploiting Forward-Backward Covariance Tensor”

Saidur R. Pavel, *Student Member, IEEE*, Yimin D. Zhang, *Fellow, IEEE*, and Shunqiao Sun, *Senior Member, IEEE*

APPENDIX A

TENSOR RECONSTRUCTION STRATEGY FOR EVEN NUMBER OF SENSOR ALONG THE AXES

In this appendix, we investigate the cases in which N_x and N_y are both even. Then, the sensor locations are expressed as

$$\mathbb{S} = \{(x_{\mathbb{S}}, y_{\mathbb{S}}) | x_{\mathbb{S}} \in [-N_x/2, N_x/2 - 1]d, y_{\mathbb{S}} \in [-N_y/2, N_y/2 - 1]d\}. \quad (67)$$

Note that it is also possible for one axis to have an even number of antennas while the other axis has an odd number. In such scenarios, the reconstruction strategy for the even axis follows the approach developed for the even case, whereas the strategy for the odd axis follows the odd case.

A particular (m, n, m', n') th element of the forward covariance tensor, i.e., $\mathcal{R}^{(f)}(m, n, m', n')$ can be expressed similarly as Eq. (13) for $m, m' \in [-\frac{N_x}{2}, \frac{N_x}{2} - 1]$ and $n, n' \in [-\frac{N_y}{2}, \frac{N_y}{2} - 1]$. The backward covariance tensor is expressed as

$$\mathcal{R}^{(b)}(m, n, m', n') = (\mathcal{R}^{(f)}(- (m + 1), - (n + 1), - (m' + 1), - (n' + 1)))^*. \quad (68)$$

From $\mathcal{R}^{(f)}$ and $\mathcal{R}^{(b)}$, we obtain the forward and backward matrices $\mathbf{R}^{(f)}$ and $\mathbf{R}^{(b)}$ similarly as Eq. (15).

Using these matrices, we construct a decorrelated tensor \mathcal{D} , where the $(:, \tilde{n}, :, \tilde{n}')$ th slice of \mathcal{D} is obtained by arranging the rows of $\mathbf{R}^{(f)}$ and $\mathbf{R}^{(b)}$ in a Toeplitz structure. Note that the row indices of both $\mathbf{R}^{(f)}$ and $\mathbf{R}^{(b)}$ are between $-\frac{N_x}{2}$ and $\frac{N_x}{2} - 1$.

Fig. 11 illustrates the proposed FBTR-based decorrelation process, which incorporates both the forward and backward covariance matrices $\mathbf{R}^{(f)}$ and $\mathbf{R}^{(b)}$. In this case, the slice $\mathcal{D}(:, \tilde{n}, :, \tilde{n}')$ is constructed by placing the $-p$ th element of the $(-\tilde{n} + \tilde{n}')$ th column at the top and forming a Toeplitz structure using $\mathbf{R}^{(f)}$ for indices satisfying $\tilde{m} \leq \frac{N_x}{2} - p$. Beyond this range, the corresponding column of $\mathbf{R}^{(b)}$ is used to complete the slice. By doing so, the dimension of \mathcal{D} is $(\frac{N_x}{2} + p) \times \frac{N_y}{2} \times (\frac{N_x}{2} + p) \times \frac{N_y}{2}$, which follows the same expression for the odd case, i.e., $(\lfloor \frac{N_x+1}{2} \rfloor + p) \times \lfloor \frac{N_y+1}{2} \rfloor \times (\lfloor \frac{N_x+1}{2} \rfloor + p) \times \lfloor \frac{N_y+1}{2} \rfloor$. The value of p for the even case is obtained from Lemma 2.

Lemma 2. For even number of antennas N_x , the optimum value of p that maximizes the number of DOFs is $\lfloor \frac{N_x}{6} + \frac{2}{3} \rfloor$.

Proof. Consider a slice of the decorrelated tensor for fixed second and fourth dimensional indices, e.g., squeeze($\mathcal{D}(:, \tilde{n},$

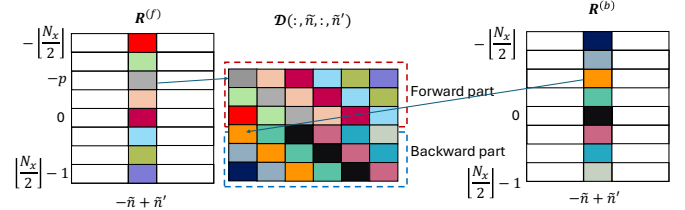


Fig. 11: Construction of a slice $\mathcal{D}(:, \tilde{n}, :, \tilde{n}')$ for even N_x with Forward backward mapping. In this example, $N_x = 8$ is considered with $p = 2$.

$:, \tilde{n}')$). In the forward-only case, where $p = 0$ and N_x is even, the slice has $\frac{N_x}{2}$ columns and $\frac{N_x}{2} + 1$ rows available from the forward matrix. As p increases, the number of columns increases by p to $\frac{N_x}{2} + p$, while the number of rows contributed by the forward matrix decreases by p to $\frac{N_x}{2} + 1 - p$. The optimum value of p is the maximum value to maintain square structure of the slice. This condition can be expressed as

$$2 \left(\frac{N_x}{2} + 1 + p \right) \geq \left(\frac{N_x}{2} - p \right), \quad (69)$$

which results

$$p \leq \frac{N_x}{6} + \frac{2}{3}. \quad (70)$$

As a result, the maximum value p can have is

$$p = \left\lfloor \frac{N_x}{6} + \frac{2}{3} \right\rfloor. \quad (71)$$

□

For both odd and even N_x and N_y cases, the number of DOFs is obtained from Eq. (34) as

$$\begin{aligned} \text{DOF}_{\text{FBTR-mat}} &= \left\lfloor \frac{N_x + 1}{2} \right\rfloor \left\lfloor \frac{N_y + 1}{2} \right\rfloor + p \left\lfloor \frac{N_y + 1}{2} \right\rfloor - 1 \\ &= \left\lfloor \frac{N_y + 1}{2} \right\rfloor v - 1, \end{aligned} \quad (72)$$

where $v = \lfloor \frac{N_x+1}{2} + p \rfloor$ and p is given as

$$p = \begin{cases} \lfloor \frac{N_x+1}{6} \rfloor, & N_x \text{ is odd,} \\ \lfloor \frac{N_x+4}{6} \rfloor, & N_x \text{ is even.} \end{cases} \quad (73)$$

Proposition: The DOF expression from Eq. (72) for both even and odd N_x and N_y can be simplified as

$$\text{DOF}_{\text{FBTR-mat}} = \left\lfloor \frac{N_y + 1}{2} \right\rfloor \left(4 \left\lfloor \frac{N_x}{6} \right\rfloor + u \right) - 1, \quad (74)$$

where

$$u = \begin{cases} 0, & N_x \equiv 0 \pmod{6}, \\ 1, & N_x \equiv 1 \pmod{6}, \\ 2, & N_x \equiv 2 \text{ or } 3 \pmod{6}, \\ 3, & N_x \equiv 4 \pmod{6}, \\ 4, & N_x \equiv 5 \pmod{6}. \end{cases} \quad (75)$$

Proof. We consider the following cases based on the divisibility of 6, and denote $z = \lfloor \frac{N_x}{6} \rfloor$.

1) $N_x \equiv 0 \pmod{6}$: In this, we can write $N_x = 6z$, thus $p = \lfloor \frac{N_x+4}{6} \rfloor = \lfloor z + \frac{4}{6} \rfloor = z$ and $\lfloor \frac{N_x+1}{2} \rfloor = \lfloor 3z + \frac{1}{2} \rfloor = 3z$. Therefore,

$$v = \left\lfloor \frac{N_x + 1}{2} + p \right\rfloor = 4z. \quad (76)$$

2) $N_x \equiv 1 \pmod{6}$: In this, we have $N_x = 6z + 1$. As such, $p = \lfloor \frac{N_x+1}{6} \rfloor = \lfloor z + \frac{1}{3} \rfloor = z$ and $\lfloor \frac{N_x+1}{2} \rfloor = \lfloor \frac{6z+2}{2} \rfloor = 3z + 1$. Therefore,

$$v = \left\lfloor \frac{N_x + 1}{2} + p \right\rfloor = 4z + 1. \quad (77)$$

3) $N_x \equiv 2 \pmod{6}$: In this case, we obtain $N_x = 6z + 2$. Therefore, $p = \lfloor \frac{N_x+4}{6} \rfloor = \lfloor \frac{6z+6}{6} \rfloor = z + 1$ and $\lfloor \frac{N_x+1}{2} \rfloor = \lfloor \frac{6z+3}{2} \rfloor = 3z + 1$. As a result,

$$v = \left\lfloor \frac{N_x + 1}{2} + p \right\rfloor = 4z + 2. \quad (78)$$

4) $N_x \equiv 3 \pmod{6}$: In this case, we have $N_x = 6z + 3$. As such, $p = \lfloor \frac{N_x+1}{6} \rfloor = \lfloor \frac{6z+4}{6} \rfloor = z$ and $\lfloor \frac{N_x+1}{2} \rfloor = \lfloor \frac{6z+4}{2} \rfloor = 3z + 2$. Therefore,

$$v = \left\lfloor \frac{N_x + 1}{2} + p \right\rfloor = 4z + 2. \quad (79)$$

5) $N_x \equiv 4 \pmod{6}$: In this case, we obtain $N_x = 6z + 4$. Therefore, $p = \lfloor \frac{N_x+4}{6} \rfloor = \lfloor \frac{6z+8}{6} \rfloor = z + 1$ and $\lfloor \frac{N_x+1}{2} \rfloor = \lfloor \frac{6z+5}{2} \rfloor = 3z + 2$. As a result,

$$v = \left\lfloor \frac{N_x + 1}{2} + p \right\rfloor = 4z + 3. \quad (80)$$

6) $N_x \equiv 5 \pmod{6}$: In this case, we can write $N_x = 6z + 5$. As such, $p = \lfloor \frac{N_x+1}{6} \rfloor = \lfloor \frac{6z+6}{6} \rfloor = z + 1$ and $\lfloor \frac{N_x+1}{2} \rfloor = \lfloor \frac{6z+6}{2} \rfloor = 3z + 3$. Therefore, we obtain

$$v = \left\lfloor \frac{N_x + 1}{2} + p \right\rfloor = 4z + 4. \quad (81)$$

□

APPENDIX B

SOLVING THE OPTIMIZATION PROBLEM EQ. (57)

We first relax the integer constraint in Eq. (57) as

$$\begin{aligned} \min_{L_x, L_y, M_x, M_y} \quad & f = -(L_x + L_y + M_x + M_y) \\ & + (L_x M_y + L_y M_x) \\ \text{s.t.} \quad & -L_x L_y \leq -\frac{K}{2}, \quad -M_x M_y \leq K + 1, \\ & -L_x \leq -\zeta_x, \quad -L_y \leq -\zeta_y, \\ & -M_x \leq -\eta_x, \quad -M_y \leq -\eta_y, \end{aligned} \quad (82)$$

The Lagrangian for the problem (82) is obtained as

$$\begin{aligned} \mathcal{L} = & -(L_x + L_y + M_x + M_y) + (L_x M_y + L_y M_x) \\ & + \lambda_P (P - L_x L_y) + \lambda_C (C - M_x M_y) + \lambda_{l_x} (\zeta_x - L_x) \\ & + \lambda_{l_y} (\zeta_y - L_y) + \lambda_{m_x} (\eta_x - M_x) + \lambda_{m_y} (\eta_y - M_y), \end{aligned} \quad (83)$$

where $P = \frac{K}{2}$, $C = K + 1$, and λ 's are the Lagrangian multipliers corresponding to the inequality constraints. By letting $\frac{\partial \mathcal{L}}{\partial L_x} = \frac{\partial \mathcal{L}}{\partial L_y} = \frac{\partial \mathcal{L}}{\partial M_x} = \frac{\partial \mathcal{L}}{\partial M_y} = 0$, we can obtain the following Karush–Kuhn–Tucker (KKT) conditions:

$$M_y - 1 - \lambda_P L_y - \lambda_{l_x} = 0, \quad (84a)$$

$$M_x - 1 - \lambda_P L_x - \lambda_{l_y} = 0, \quad (84b)$$

$$L_y - 1 - \lambda_C M_y - \lambda_{m_x} = 0, \quad (84c)$$

$$L_x - 1 - \lambda_C M_x - \lambda_{m_y} = 0, \quad (84d)$$

$$\lambda_P (P - L_x L_y) = 0, \quad (84e)$$

$$\lambda_C (C - M_x M_y) = 0, \quad (84f)$$

$$\lambda_{l_x} (\zeta_x - L_x) = 0, \quad (84g)$$

$$\lambda_{l_y} (\zeta_y - L_y) = 0, \quad (84h)$$

$$\lambda_{m_x} (\eta_x - M_x) = 0, \quad (84i)$$

$$\lambda_{m_y} (\eta_y - M_y) = 0, \quad (84j)$$

$$L_x \geq \zeta_x, L_y \geq \zeta_y, M_x \geq \eta_x, M_y \geq \eta_y, \quad (84k)$$

$$L_x L_y \geq P, M_x M_y \geq C, \quad (84l)$$

$$\lambda_1, \lambda_2, \lambda_2, \lambda_2 \geq 0. \quad (84m)$$

Considering the complementary slackness of the λ 's, we have the following cases:

1) Case 1 ($\lambda_P = \lambda_C = 0$): In this case, $\lambda_{l_x} = M_y - 1 > 0$ and $\lambda_{l_y} = M_x - 1 > 0$ from Eqs. (84a) and (84b), which implies $L_x = \zeta_x$ and $L_y = \zeta_y$ from Eqs. (84g) and (84h). Moreover, $\lambda_{m_x} = L_y - 1$ and $\lambda_{m_y} = L_x - 1$ from Eqs. (84c) and (84d). Considering the four possible cases of equality or inequality between $(\lambda_{m_x}, \lambda_{m_y})$, the corresponding solutions are $(\zeta_x, \zeta_y, \max(\eta_x, \frac{C}{\eta_y}), \eta_y)$, $(\zeta_x, \zeta_y, \eta_x, \max(\eta_y, \frac{C}{\eta_x}))$, and $(\zeta_x, \zeta_y, \eta_x, \eta_y)$. However these solutions are equivalent to $(\zeta_x, \zeta_y, \eta_x, \eta_y)$.

Proof. The feasibility condition requires $\zeta_x \zeta_y \geq \frac{K}{2}$. By definition, $\zeta_x = \frac{\eta_x - 1}{2}$ and $\zeta_y = \frac{\eta_y - 1}{2}$. Therefore,

$$\eta_x \eta_y = (2\zeta_x + 1)(2\zeta_y + 1) = 4\zeta_x \zeta_y + 2(\zeta_x + \zeta_y) + 1. \quad (85)$$

Since $\zeta_x \zeta_y \geq \frac{K}{2}$, we obtain

$$\eta_x \eta_y \geq 2K + 2(\zeta_x + \zeta_y) + 1 = 2C + 2(\zeta_x + \zeta_y) - 1 \geq 2C. \quad (86)$$

Now, $\eta_x \eta_y \geq 2C$ implies $\eta_x \geq \frac{2C}{\eta_y} \geq \frac{C}{\eta_y}$, hence $\max(\eta_x, \frac{C}{\eta_y}) = \eta_x$. Similarly, it follows that $\max(\eta_y, \frac{C}{\eta_x}) = \eta_y$. □

2) Case 2 ($\lambda_P = 0$ and $\lambda_C \neq 0$): No solution exists in this branch. In this case, $\lambda_{l_x} = M_y - 1 > 0$ and $\lambda_{l_y} = M_x - 1 > 0$ from Eq. (84a). From Eqs. (84g) and (84h), we have $L_x = \zeta_x$ and $L_y = \zeta_y$, while $\lambda_C = 0$ implies $M_x M_y = C = K + 1$ from Eq. (84f). The feasibility condition requires $M_x \geq \eta_x$ and $M_y \geq \eta_y$, which gives

$M_x M_y \geq \eta_x \eta_y$. By definition, $\zeta_x = \frac{\eta_x - 1}{2}$ and $\zeta_y = \frac{\eta_y - 1}{2}$, so

$$\eta_x \eta_y = (2\zeta_x + 1)(2\zeta_y + 1) = 4\zeta_x \zeta_y + 2(\zeta_x + \zeta_y) + 1. \quad (87)$$

Since the feasibility condition also requires $\zeta_x \zeta_y \geq \frac{K}{2}$ and $\zeta_x, \zeta_y \geq 0.5$, it follows that

$$\eta_x \eta_y \geq 2K + 3 = 2C + 1 \geq C. \quad (88)$$

Thus, $M_x M_y \geq \eta_x \eta_y > C$, which contradicts $M_x M_y = C$. Hence, no feasible solution exists in this case.

- 3) Case 3 ($\lambda_P \neq 0$ and $\lambda_C = 0$): From Eqs. (84c) and (84d), we have $\lambda_{mx} = L_y - 1$ and $\lambda_{my} = L_x - 1$, and from Eq. (84e), $L_x L_y = P$.

For $\lambda_{lx} \neq 0$ and $\lambda_{ly} \neq 0$, considering the equality and inequality to zero of $(\lambda_{mx}, \lambda_{my})$, the solutions are $(L_x, L_y, M_x, M_y) = (\zeta_x, \zeta_y, \eta_x, \eta_y)$, $(\zeta_x, \zeta_y, \max(\eta_x, \frac{C}{\eta_y}), \eta_y)$, and $(\zeta_x, \zeta_y, \eta_x, \max(\eta_y, \frac{C}{\eta_x}))$. However these solutions are equivalent to $(\zeta_x, \zeta_y, \eta_x, \eta_y)$.

For $\lambda_{lx} \neq 0$ and $\lambda_{ly} = 0$, the solutions are $(\zeta_x, \frac{P}{\zeta_x}, \max(\eta_x, \frac{C}{\eta_y}), \eta_y)$ and $(\zeta_x, \frac{P}{\zeta_x}, \eta_x, \max(\eta_y, \frac{C}{\eta_x}))$.

For $\lambda_{lx} = 0$ and $\lambda_{ly} \neq 0$, the solutions are $(\frac{P}{\zeta_y}, \zeta_y, \max(\eta_x, \frac{C}{\eta_y}), \eta_y)$ and $(\frac{P}{\zeta_y}, \zeta_y, \eta_x, \max(\eta_y, \frac{C}{\eta_x}))$. However, these solutions are equivalent to $(\zeta_x, \frac{P}{\zeta_x}, \eta_x, \eta_y)$ and $(\frac{P}{\zeta_y}, \zeta_y, \eta_x, \eta_y)$.

Proof. Consider $(L_x, L_y, M_x, M_y) = (\zeta_x, \frac{P}{\zeta_x}, \max(\eta_x, \frac{C}{\eta_y}), \eta_y)$. This solution is obtained when $\lambda_{my} \neq 0$, $\lambda_{lx} \neq 0$, and $\lambda_{mx} = \lambda_{ly} = 0$. Since $\lambda_{mx} = 0$ implies $L_y = 1$. Also, $L_x = \zeta_x$ and $L_x L_y = P$ give $\zeta_x = P$, leading to $\eta_x = 2\zeta_x + 1 = 2P + 1 = C$. Because $\eta_y \geq 2$, we have $\frac{C}{\eta_y} \leq C = \eta_x$, and thus $\max(\eta_x, \frac{C}{\eta_y}) = \eta_x$. Hence, the solution reduces to $(\zeta_x, \frac{P}{\zeta_x}, \eta_x, \eta_y)$. A similar argument holds for the other case. \square

Finally, for $\lambda_{lx} = 0$ and $\lambda_{ly} = 0$, using $L_y = \frac{M_y - 1}{\lambda_P}$, $L_x = \frac{M_x - 1}{\lambda_P}$, and $L_x L_y = P$ from Eqs. (84a), (84b), and (84e), and considering the equality and inequality to zero of $(\lambda_{mx}, \lambda_{my})$, the solutions are $(L_x, L_y, M_x, M_y) = (1, P, \eta_x, P(\eta_x - 1) + 1)$, $(P, 1, P(\eta_y - 1) + 1, \eta_y)$, and $(\sqrt{\frac{P(\eta_x - 1)}{\eta_y - 1}}, \sqrt{\frac{P(\eta_y - 1)}{\eta_x - 1}}, \eta_x, \eta_y)$.

It can be shown that $(\zeta_x, \frac{P}{\zeta_x}, \eta_x, \eta_y)$ and $(\frac{P}{\zeta_y}, \zeta_y, \eta_x, \eta_y)$ yield the same value of the objective function. Without loss of generality, we select $(\frac{P}{\zeta_y}, \zeta_y, \eta_x, \eta_y)$. The solutions $(1, P, \eta_x, P(\eta_x - 1) + 1)$ and $(P, 1, P(\eta_y - 1) + 1, \eta_y)$ provide larger values of the objective function compared with $(\frac{P}{\zeta_y}, \zeta_y, \eta_x, \eta_y)$ and can therefore be discarded. Hence, in this case, the unique solution is $(\frac{P}{\zeta_y}, \zeta_y, \eta_x, \eta_y)$ and $(\sqrt{\frac{P(\eta_x - 1)}{\eta_y - 1}}, \sqrt{\frac{P(\eta_y - 1)}{\eta_x - 1}}, \eta_x, \eta_y)$.

- 4) Case 4 ($\lambda_P \neq 0$ and $\lambda_C \neq 0$): This case implies $L_x L_y = P$ and $M_x M_y = C$. Based on the equality or inequality to zero of λ_{lx} , λ_{ly} , λ_{mx} , and λ_{my} , we got the following cases.

- a) All of them are zeros:

Substituting $\lambda_{lx} = \lambda_{ly} = \lambda_{mx} = \lambda_{my} = 0$ into Eqs. (84a)–(84d) and subtracting Eq. (84b) from Eq. (84a) yields

$$M_y - M_x = \lambda_P (L_y - L_x). \quad (89)$$

Similarly, subtracting Eq. (84d) from Eq. (84c), we get

$$L_y - L_x = \lambda_C (M_y - M_x). \quad (90)$$

Substituting Eq. (89) into Eq. (90), we have

$$L_y - L_x = \lambda_P \lambda_C (L_y - L_x). \quad (91)$$

This leads to two possibilities, i.e., either $\lambda_P \lambda_C = 1$ or $L_y = L_x$.

For $\mu_1 \mu_2 = 1$:

Adding Eqs. (84a) and (84b), we obtain

$$(M_x + M_y) - \lambda_P (L_x + L_y) = 2, \quad (92)$$

that is,

$$-\lambda_C (M_x + M_y) + (L_x + L_y) = -2\lambda_C. \quad (93)$$

Similarly, adding Eqs. (84c) and (84d), we get

$$-\lambda_C (M_x + M_y) + (L_x + L_y) = 2. \quad (94)$$

Comparing the above two equations, we find that $\lambda_C = -1$, which contradicts with $\lambda_C > 0$. Hence, it is not feasible.

For $L_x = L_y$:

From Eq. (89), this also implies $M_x = M_y$. Applying these conditions to Eq. (84e), we find:

$$L_x = L_y = \sqrt{\frac{K}{2}}, \quad (95)$$

and from Eq. (84f), we obtain:

$$M_x = M_y = \sqrt{C}. \quad (96)$$

Therefore, the solution is $(L_x, L_y, M_x, M_y) = (\sqrt{\frac{K}{2}}, \sqrt{\frac{K}{2}}, \sqrt{C}, \sqrt{C})$.

- b) Three of them are zeros:

- i) $\lambda_{lx} = \lambda_{ly} = \lambda_{mx} = 0$, and $\lambda_{my} \neq 0$:

From Eq. (84j), we find that $M_y = \eta_y$, which implies that $M_x = \frac{C}{\eta_y}$. From Eqs. (84a) and (84e), we obtain $\lambda_P = \frac{\eta_y - 1}{L_y} = \frac{2(\eta_y - 1)L_x}{K}$. From Eqs. (84b) and (84f), we have $\lambda_P = \frac{M_x - 1}{L_x} = (\frac{C}{\eta_y} - 1)/L_x$. Comparing these two expressions yields

$$\frac{2(\eta_y - 1)L_x}{K} = \frac{C - \eta_y}{L_x}, \quad (97)$$

i.e.,

$$L_x^2 = \frac{K(C - \eta_y)}{2\eta_y(\eta_y - 1)}, \quad (98)$$

which results in $L_x = \sqrt{\frac{K(C - \eta_y)}{2\eta_y(\eta_y - 1)}}$. As a result,

$$L_y = \frac{K}{2L_x} = \sqrt{\frac{K\eta_y(\eta_y - 1)}{2(C - \eta_y)}}. \quad (99)$$

Therefore, the solution is $(L_x, L_y, M_x, M_y) = (\sqrt{\frac{K(C - \eta_y)}{2\eta_y(\eta_y - 1)}}, \sqrt{\frac{K\eta_y(\eta_y - 1)}{2(C - \eta_y)}}, \frac{C}{\eta_y}, \eta_y)$.

Due to symmetry, it can be shown that, for the case $\lambda_{lx} = \lambda_{ly} = \lambda_{my} = 0$ and $\lambda_{mx} \neq 0$, the solution is $(L_x, L_y, M_x, M_y) = (\sqrt{\frac{K\eta_x(\eta_x-1)}{2(C-\eta_x)}}, \sqrt{\frac{K(C-\eta_x)}{2\eta_x(\eta_x-1)}}, \eta_x, \frac{C}{\eta_x})$.

ii) $\lambda_{ly} = \lambda_{mx} = \lambda_{my} = 0$, and $\lambda_{lx} \neq 0$:

This case implies $L_x = \zeta_x$ and $L_y = \frac{P}{\zeta_x}$ from Eqs. (84g) and (84e). From Eq. (84c), we obtain

$$\lambda_C = \frac{L_y-1}{M_y} = \frac{(\frac{P}{\zeta_x}-1)M_x}{C}, \text{ while from Eq. (84d), } \lambda_C = \frac{\zeta_x-1}{M_x}.$$

Equating these two expressions of λ_C yields $M_x = \sqrt{\frac{C(\zeta_x-1)}{\frac{K}{2\zeta_x}-1}}$ and $M_y = \sqrt{\frac{C(\frac{K}{2\zeta_x}-1)}{\zeta_x-1}}$.

Therefore the solution is $(L_x, L_y, M_x, M_y) = (\zeta_x, \frac{K}{2\zeta_x}, \sqrt{\frac{C(\zeta_x-1)}{\frac{K}{2\zeta_x}-1}}, \sqrt{\frac{C(\frac{K}{2\zeta_x}-1)}{\zeta_x-1}})$. This solution, however, is not feasible.

Proof. Here $M_x^2 = \frac{C(\zeta_x-1)}{\frac{K}{2\zeta_x}-1} = \frac{2(K+1)\zeta_x(\zeta_x-1)}{K-2\zeta_x}$. Since the feasibility condition requires $M_x \geq \eta_x$ and $M_y \geq \eta_y$, we have $C = M_x M_y \geq \eta_x \eta_y = 2(2\zeta_x+1)$, because $\eta_y \geq 2$. Therefore $K \geq 4\zeta_x+1$. Plugging this value into the expression for M_x^2 yields

$$\begin{aligned} M_x^2 &\leq \frac{2(4\zeta_x+2)\zeta_x(\zeta_x-1)}{4\zeta_x+1-2\zeta_x} = 4\zeta_x(\zeta_x-1) \\ &= (2\zeta_x+1)^2 - (8\zeta_x+1) < (2\zeta_x+1)^2 = \eta_x^2, \end{aligned} \quad (100)$$

which is infeasible. \square

Due to symmetry, for the case $\lambda_{lx} = \lambda_{mx} = \lambda_{my} = 0$, and $\lambda_{ly} \neq 0$, the solution is $(L_x, L_y, M_x, M_y) = (\frac{K}{2\zeta_y}, \zeta_y, \sqrt{\frac{C(\frac{K}{2\zeta_y}-1)}{\zeta_y-1}}, \sqrt{\frac{C(\zeta_y-1)}{\frac{K}{2\zeta_y}-1}})$, which is also infeasible.

c) Two of them are zeros:

i) $\lambda_{lx} \neq 0, \lambda_{ly} \neq 0, \lambda_{mx} = 0$ and $\lambda_{my} = 0$:

From Eqs. (84g) and (84h), we obtain $L_x = \zeta_x$ and $L_y = \zeta_y$. From Eq. (84c), $\lambda_C = \frac{L_y-1}{M_y} = \frac{(\zeta_y-1)M_x}{C}$, while from Eq. (84d), $\lambda_C = \frac{\zeta_x-1}{M_x}$. Equating these two expressions of λ_C

gives $M_x = \sqrt{\frac{C(\zeta_x-1)}{(\zeta_y-1)}}$ and $M_y = \sqrt{\frac{C(\zeta_y-1)}{(\zeta_x-1)}}$. Therefore, the solution is $(L_x, L_y, M_x, M_y) = (\zeta_x, \zeta_y, \sqrt{\frac{C(\zeta_x-1)}{(\zeta_y-1)}}, \sqrt{\frac{C(\zeta_y-1)}{(\zeta_x-1)}})$. This solution is infeasible since

$$\begin{aligned} M_x M_y &\geq \eta_x \eta_y = 4\zeta_x \zeta_y + 2(\zeta_x + \zeta_y) + 1 \\ &= 2K + 2(\zeta_x + \zeta_y) + 1 \\ &= C + K + 2(\zeta_x + \zeta_y) > C, \end{aligned} \quad (101)$$

which contradicts the condition $M_x M_y = C$, and therefore the solution is not feasible.

ii) $\lambda_{lx} = 0, \lambda_{ly} = 0, \lambda_{mx} \neq 0$, and $\lambda_{my} \neq 0$:

From Eqs. (84i) and (84j), we find that $M_x = \eta_x$ and $M_y = \eta_y$. This is a special case of Case (b1) when $\eta_x \eta_y = C$, thus results in the same solution.

For all other cases, it is straightforward to obtain the solutions are one of the followings

$(\zeta_x, \zeta_y, \eta_x, \eta_y), (\zeta_x, \zeta_y, \eta_x, \frac{C}{\eta_x}), (\zeta_x, \zeta_y, \frac{C}{\eta_y}, \eta_y), (\zeta_x, \frac{K}{2\zeta_x}, \eta_x, \eta_y)$, and $(\frac{K}{2\zeta_y}, \zeta_y, \eta_x, \eta_y)$. Therefore in this case, the unique solutions are considering $\eta_y \geq \eta_x$ is $(\sqrt{\frac{K(C-\eta_y)}{2\eta_y(\eta_y-1)}}, \sqrt{\frac{K\eta_y(\eta_y-1)}{2(C-\eta_y)}}, \frac{C}{\eta_y}, \eta_y)$

and $(\sqrt{\frac{K}{2}}, \sqrt{\frac{K}{2}}, \sqrt{C}, \sqrt{C})$. Between these two solutions, it can be shown that, the solution $(\sqrt{\frac{K(C-\eta_y)}{2\eta_y(\eta_y-1)}}, \sqrt{\frac{K\eta_y(\eta_y-1)}{2(C-\eta_y)}}, \frac{C}{\eta_y}, \eta_y)$ yields a lower value of the objective function than $(\sqrt{\frac{K}{2}}, \sqrt{\frac{K}{2}}, \sqrt{C}, \sqrt{C})$, and therefore the latter can be discarded.

Proof. Minimizing the objective function is equivalent to minimizing

$$g = (L_x - 1)(M_y - 1) + (L_y - 1)(M_x - 1). \quad (102)$$

First, compute the value of g for the first solution. Define $A = \eta_y - 1$ and $B = \frac{C}{\eta_y} - 1$. In this case, $L_x L_y = P$, and it follows that

$$\frac{L_x}{L_y} = \frac{B}{A}, \quad (103)$$

which gives $L_x = \frac{PB}{A}$ and $L_y = \frac{PA}{B}$.

Therefore, the objective function for the first solution is

$$g_1 = A(L_x - 1) + B(L_y - 1) = 2\sqrt{PAB} - (A + B). \quad (104)$$

Note that $A+B = \eta_y + \frac{C}{\eta_y} - 2 \geq 2(\sqrt{C}-1)$, since $\eta_y + \frac{C}{\eta_y} \geq 2\sqrt{C}$ by the AM-GM inequality. Furthermore, $AB = C - \eta_y - \frac{C}{\eta_y} + 1 \leq C + 1 - 2\sqrt{C} = (\sqrt{C}-1)^2$, which implies $\sqrt{AB} \leq \sqrt{C}-1$. Substituting these bounds into g_1 yields

$$\begin{aligned} g_1 &\leq 2\sqrt{P}(\sqrt{C}-1) - 2(\sqrt{C}-1) \\ &= 2(\sqrt{C}-1)(\sqrt{P}-1), \end{aligned} \quad (105)$$

which is exactly the value of the objective function for the second solution, g_2 . Hence $g_1 \leq g_2$.

Therefore, in this case it is sufficient to retain only the first solution. \square

Therefore, all solutions can be summarized as follows

$$(L_x, L_y, M_x, M_y) = \begin{cases} (\zeta_x, \zeta_y, \eta_x, \eta_y), & \text{if } \zeta_x \zeta_y > \frac{C-1}{2}, \\ (l_x, l_y, \eta_x, \eta_y), & \text{if } \zeta_x \zeta_y \leq \frac{C-1}{2} \text{ and } \eta_x \eta_y > C, \\ \left(\sqrt{\frac{K(C-\eta_y)}{2\eta_y(\eta_y-1)}}, \sqrt{\frac{K\eta_y(\eta_y-1)}{2(C-\eta_y)}}, \frac{C}{\eta_y}, \eta_y\right), & \text{if } \eta_x \eta_y \leq C, \end{cases} \quad (106)$$

with

$$(l_x, l_y) = \begin{cases} \left(\sqrt{\frac{K(\eta_x-1)}{2(\eta_y-1)}}, \sqrt{\frac{K(\eta_y-1)}{2(\eta_x-1)}}\right), & \text{if } \sqrt{\frac{K(\eta_x-1)}{2(\eta_y-1)}} \geq \zeta_x \text{ and } \sqrt{\frac{K(\eta_y-1)}{2(\eta_x-1)}} \geq \zeta_y, \\ \left(\frac{K}{2\zeta_y}, \zeta_y\right), & \text{otherwise.} \end{cases} \quad (107)$$