



Delft University of Technology

## Efficient Off-Grid Bayesian Parameter Estimation for Kronecker-Structured Signals

He, Yanbin; Joseph, Geethu

**DOI**

[10.1109/TSP.2025.3583895](https://doi.org/10.1109/TSP.2025.3583895)

**Publication date**

2025

**Document Version**

Final published version

**Published in**

IEEE Transactions on Signal Processing

**Citation (APA)**

He, Y., & Joseph, G. (2025). Efficient Off-Grid Bayesian Parameter Estimation for Kronecker-Structured Signals. *IEEE Transactions on Signal Processing*, 73, 2616-2630.

<https://doi.org/10.1109/TSP.2025.3583895>

**Important note**

To cite this publication, please use the final published version (if applicable).

Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights.

We will remove access to the work immediately and investigate your claim.

**Green Open Access added to [TU Delft Institutional Repository](#)  
as part of the Taverne amendment.**

More information about this copyright law amendment  
can be found at <https://www.openaccess.nl>.

Otherwise as indicated in the copyright section:  
the publisher is the copyright holder of this work and the  
author uses the Dutch legislation to make this work public.

# Efficient Off-Grid Bayesian Parameter Estimation for Kronecker-Structured Signals

Yanbin He , Graduate Student Member, IEEE and Geethu Joseph , Senior Member, IEEE

**Abstract**—This work studies the problem of jointly estimating unknown parameters from Kronecker-structured multidimensional signals, which arises in applications like intelligent reflecting surface (IRS)-aided channel estimation. Exploiting the Kronecker structure, we decompose the estimation problem into smaller, independent subproblems across each dimension. Each subproblem is posed as a sparse recovery problem using basis expansion and solved using a novel off-grid sparse Bayesian learning (SBL)-based algorithm. Additionally, we derive probabilistic error bounds for the decomposition, quantify its denoising effect, and provide convergence analysis for off-grid SBL. Our simulations show that applying the algorithm to IRS-aided channel estimation improves accuracy and runtime compared to state-of-the-art methods through the low-complexity and denoising benefits of the decomposition step and the high-resolution estimation capabilities of off-grid SBL.

**Index Terms**—Sparse Bayesian learning, higher-order SVD, intelligent reflecting surface, channel estimation, basis expansion.

## I. INTRODUCTION

MULTIDIMENSIONAL signals arise in several engineering applications such as image processing [2], [3], [4] and wireless communications [5], [6], [7]. In these contexts, the data is represented as a function of different dimensions, each conveying a specific physical quantity. For example, in the uplink narrowband intelligent reflecting surface (IRS)-aided system, the received signal at the base station (BS) from the mobile station (MS) is a function of angle-of-departure (AoD) at MS, the difference of angle-of-arrival (AoA) and AoD at the IRS, and AoA at BS [8]. Considering the angular domain of each array as a separate dimension, this signal is multidimensional [7], [8]. This structure is captured by the Kronecker product [9], leading to the fundamental model,

$$\bar{\mathbf{y}} = \mathbf{y}_1 \otimes \mathbf{y}_2 \otimes \cdots \otimes \mathbf{y}_I + \bar{\mathbf{n}} = \otimes_{i=1}^I \mathbf{y}_i + \bar{\mathbf{n}}, \quad (1)$$

Received 29 November 2024; revised 7 May 2025 and 17 June 2025; accepted 19 June 2025. Date of publication 30 June 2025; date of current version 10 July 2025. An earlier version of this paper was presented in part at International Conference on Acoustics, Speech, and Signal Processing, 2025 [DOI: 10.1109/ICASSP49660.2025.10890014]. The associate editor coordinating the review of this article and approving it for publication was Xiao Fu. (Corresponding author: Geethu Joseph.)

The authors are with the Signal Processing Systems Group, Electrical Engineering, Mathematics, and Computer Science Faculty, Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: y.he-1@tudelft.nl; g.joseph@tudelft.nl).

Digital Object Identifier 10.1109/TSP.2025.3583895

where  $\otimes_{i=1}^I \mathbf{y}_i \in \mathbb{C}^{\bar{M}}$  is an  $I$ -dimensional signal,  $\mathbf{y}_i \in \mathbb{C}^{M_i}$  represents the signal in each dimension,  $\bar{\mathbf{n}}$  is the noise,  $\bar{M} = \prod_{i=1}^I M_i$ , and  $\otimes$  is the Kronecker product. Each  $\mathbf{y}_i$  encapsulates the signal in the corresponding dimension (e.g., AoA and AoD) and is expressed as a weighted sum of nonlinear parametric functions,

$$\mathbf{y}_i = \sum_{s=1}^{S_i} \mathbf{h}_i(\bar{\psi}_{i,s}) \bar{x}_{i,s}, \quad (2)$$

with parameters  $\bar{\psi}_{i,s}$  and weights  $\bar{x}_{i,s}$ , for  $s = 1, 2, \dots, S_i$ , where  $\mathbf{h}_i(\cdot) \in \mathbb{C}^{M_i}$  is the nonlinear function. In the IRS-aided system example, the parameter  $\bar{\psi}_{i,s}$  can be AoAs or AoDs, the nonlinear function  $\mathbf{h}_i$  is related to the steering vector, and  $\bar{x}_{i,s}$  represents the path gain corresponding to each AoA or AoD. Thus, the channel estimation problem reduces to estimating all  $\bar{\psi}_{i,s}$ 's and  $\bar{x}_{i,s}$ 's from the received signal  $\bar{\mathbf{y}}$ , where  $S_i$ 's are also unknown. Hence, this paper focuses on the general problem of estimating the parameters and weights  $\{\bar{\psi}_{i,s}, \bar{x}_{i,s}\}_{s=1}^{S_i}$  from measurements  $\bar{\mathbf{y}}$  and the function  $\mathbf{h}_i(\cdot)$  for all  $I$  dimensions.

A popular approach for parameter estimation from (1) and (2) is multidimensional basis expansion model (BEM) [2], [3], [4], [8], [10], [11]. It evaluates the nonlinear function over pre-sampled grids of unknown parameters in each dimension to express the multidimensional signal as the product of a known overcomplete Kronecker-structured dictionary of the basic functions and an unknown sparse coefficient vector as

$$\bar{\mathbf{y}} = (\otimes_{i=1}^I \mathbf{H}_i) \mathbf{x} + \bar{\mathbf{n}}. \quad (3)$$

Here,  $\bar{\mathbf{y}} \in \mathbb{C}^{\bar{M}}$  is the measurement,  $\mathbf{H}_i \in \mathbb{C}^{M_i \times N_i}$  is the overcomplete basis for  $\mathbf{h}_i(\cdot)$ , and  $\otimes_{i=1}^I \mathbf{H}_i \in \mathbb{C}^{\bar{M} \times \bar{N}}$  is the overall dictionary with  $\bar{M} < \prod_{i=1}^I N_i = \bar{N}$ . Also,  $\mathbf{x} \in \mathbb{C}^{\bar{N}}$  is the unknown sparse vector, and  $\bar{\mathbf{n}}$  is the measurement noise. The model in (2) leads to a Kronecker-structure in  $\mathbf{x}$

$$\mathbf{x} = \otimes_{i=1}^I \mathbf{x}_i, \quad (4)$$

where  $\mathbf{x}_i \in \mathbb{C}^{N_i}$  is the weight in the  $i$ th dimension. The overcomplete dictionary makes  $\mathbf{x}_i$  sparse, with only  $S_i$  nonzero entries corresponding to the true parameters. Thus, estimating parameters and coefficients is a sparse recovery problem.

Solving (3) for a sparse vector  $\mathbf{x}$  with Kronecker-structured support has been discussed in [3], [7], [10], [11]. A greedy method, Kronecker- orthogonal matching pursuit (OMP), generalizes the traditional OMP to multidimensional BEM [3]. It has low complexity but requires hand-tuning of a sensitive stopping threshold [7]. Another approach with improved

accuracy and no hand-tuning relies sparse Bayesian learning (SBL). It adopts a fictitious Gaussian prior on the sparse vector with a Kronecker-structured covariance matrix to enforce a Kronecker-structured support [7], [10], [11]. Although Kronecker-SBL (KroSBL) can be readily applied to our problem, it has two main drawbacks, as elaborated below.

First, KroSBL does not fully exploit prior knowledge (4). While KroSBL employs a Kronecker-structured covariance matrix, the variance only determines whether an entry is nonzero; in effect, KroSBL only exploits the Kronecker structure of the support vector. Consequently, KroSBL directly estimates the high-dimensional vector  $\mathbf{x}$ . Although the state-of-the-art KroSBL algorithm employs some complexity reduction techniques [7], it still faces high overall complexity due to the high dimensionality of the Kronecker product. So, we seek a method that can better exploit the Kronecker structure while significantly reducing complexity compared to KroSBL.

Second, KroSBL relies on multidimensional BEM, formulating the dictionary using predefined grids. However, the true parameters may fall off these grids, causing a *grid mismatch* issue [12], [13], which can degrade the estimation performance. A popular way to address mismatch is to estimate the unknown parameters by optimizing the SBL cost function. However, this is challenging due to the *nonlinearity* of the SBL cost function. One approach is to use linearization methods that initialize the unknown parameters using pre-sampled values from the range of interest, locally approximate the nonlinear cost function using linearization techniques such as Taylor expansion [12], [14], [15], and solve for the first-order coefficients. However, such an approximation is only valid locally [16], [17]. Initialization with finer sampled values can only alleviate but not fully eliminate this issue [16], [17], [18], and it also induces higher-dimension dictionary with worse coherence condition. Alternatively, some works turn to marginal likelihood maximization by isolating the contribution of each variable to the likelihood, and then sequentially selecting the candidate that maximizes this isolated contribution [16], [19], [20], [21]. This greedy strategy starts with an empty dictionary and progressively adds the top-contributing candidates. So it is closely related to matching pursuit algorithms [20] and can suffer performance degradation as the number of unknowns increases [22]. The framework in [23] can also be leveraged to jointly construct Kronecker-structured dictionary and sparse vector. However, the variational Bayesian-based approach in [23] assumes unconstrained columns in dictionary, while the column of our dictionary is constrained by column function  $\mathbf{h}_i(\cdot)$ . In addition to the above off-grid methods, *gridless* methods, which operate directly in the continuous domain without discretization [13], are also studied in the literature. Example approaches include atomic norm-based methods [24], [25], [26] and variational Bayesian line spectral estimation techniques [27], [28]. However, atomic norm methods require a structured column function  $\mathbf{h}_i(\cdot)$  to admit a Vandermonde decomposition for parameter recovery, while the distribution approximation used in variational Bayesian approaches is valid when  $\mathbf{h}_i(\cdot)$  consists of complex sinusoidal components. Thus, they cannot be trivially generalized to different forms

of  $\mathbf{h}_i(\cdot)$ . The drawbacks and limitations of existing multidimensional BEM, off-grid SBL algorithms, and gridless methods motivate novel approaches to solving our parameter estimation problem.

We aim to develop a method for estimating the parameters and weights  $\{\bar{\psi}_{i,s}, \bar{x}_{i,s}\}_{s=1}^{S_i}$  for all  $I$  dimensions using  $\bar{\mathbf{y}}$  in (1) and (2) with three key features: (i) utilizing the Kronecker structure in (1); (ii) overcoming the grid mismatch of linearization and marginal likelihood optimization; and (iii) achieving lower complexity compared to KroSBL. Our algorithm follows the BEM paradigm using SBL and enjoys the theoretical guarantees. Our main contributions are as follows:

- *Decomposition-Based Algorithm*: We present two methods to decompose the measurement  $\bar{\mathbf{y}}$  into multiple low-dimensional measurements, better utilizing the prior information of the Kronecker structure in Sec. II-A. It transforms the joint multidimensional unknown parameters estimation into multiple separate subproblems in each dimension, leading to reduced complexity.
- *Off-grid Algorithm*: We use BEM for parameters estimation in each dimension and cast it into a sparse vector recovery problem solved using the expectation-maximization (EM)-based SBL in Sec. II-B. We further incorporate a grid optimization step in the EM iterations to address grid mismatch, implemented via alternating minimization. This approach fills gaps in prior work by enabling optimization without the linearization approximations used in [12], [14], [15] and jointly updating all variables, avoiding the greedy selection strategies of [16], [19], [20], [21].
- *Algorithm Analyses and Extensions*: We study the decomposition step and the iterative grid optimization in Sec. III. We theoretically quantify the error bound of the decomposition step in the presence of noise and the denoising effect which we attribute to the better estimation performance. We discuss the convergence property of our algorithm. We also explore potential extensions and applicability of our decomposition strategy to alternative measurement structures that emerge in various practical scenarios in Sec. III-D.
- *Application*: In Sec. IV, we analyze the signal model of a prototypical IRS-aided wireless communication system and explain the implementation of our algorithm for uplink cascaded IRS channel estimation.
- *Numerical Results*: We evaluate our schemes in three scenarios in Sec. V. The first scenario highlights the computational efficiency and denoising benefits of the decomposition method. The second scenario demonstrates the high-resolution estimation capabilities of off-grid SBL. The third scenario focuses on IRS channel estimation, showcasing improved accuracy and reduced runtime, driven by the combined effects of decomposition and off-grid SBL.

In short, our algorithm estimates parameters from Kronecker-structured multidimensional signals, tackling grid mismatch and high complexity through two key techniques: decomposition and off-grid SBL. These techniques are of independent interest and can be applied separately, depending on the specific

signal model. Compared to our preliminary work [1], we make several novel contributions here. Apart from the singular value decomposition (SVD)-based approach in [1], we introduce higher-order singular value decomposition (HOSVD)-based approach to the decomposition step and present its probabilistic error bounds, including a decomposition error bound and a characterization of denoising. Additionally, we develop an off-grid SBL algorithm, addressing the issue of grid mismatch and analyzing its convergence properties. Further, we explore extensions of our strategy to alternative measurement structures that emerge in various practical scenarios, and present extensive numerical evaluations of our two-step solution applied to IRS channel estimation.

*Notation:* Boldface small letters denote vectors, boldface capital letters denote matrices, and calligraphic letters denote tensors. We use  $[I]$  to denote the set  $\{1, 2, \dots, I\}$  and the symbols  $\otimes$ ,  $\odot$ ,  $\circ$ , and  $\times_i$  to denote Kronecker, Khatri-Rao, tensor outer product, and tensor  $i$ th mode product, respectively.

*Tensor preliminaries:* We present tensor fundamentals including outer product, tensor matricization, tensor mode product, HOSVD, and multilinear rank. The outer product of vectors  $\{\mathbf{t}_i \in \mathbb{C}^{M_i}\}_{i=1}^I$  yields a rank-one tensor  $\mathcal{R} := \circ_{i=1}^I \mathbf{t}_i \in \mathbb{C}^{M_1 \times \dots \times M_I}$  with entries  $[\mathcal{R}]_{m_1, \dots, m_I} = \prod_{i=1}^I [\mathbf{t}_i]_{m_i}$ . The  $i$ th matricization matrix  $\mathbf{T}_{(i)}$  of tensor  $\mathcal{T}$  is defined as

$$[\mathbf{T}_{(i)}]_{m_i, j} = [\mathcal{T}]_{m_1, m_2, \dots, m_I}, \quad \forall i \in [I],$$

with  $m_i \in [M_i]$ , and  $j = 1 + \sum_{\ell=1, \ell \neq i}^I \left( \prod_{p=1, p \neq i}^{\ell-1} M_p \right) (m_\ell - 1)$ . Matrix  $\mathbf{T}_{(i)}$  matricizes  $\mathcal{T}$  along its  $i$ th dimension and has dimension  $M_i \times \prod_{\ell=1, \ell \neq i}^I M_\ell$  with the  $(m_i, j)$ -th entry given by the  $(m_1, \dots, m_I)$ -th entry of  $\mathcal{T}$ . Specifically, for rank one tensor  $\mathcal{R}$ , its  $i$ th matricization matrix is [29]

$$\mathbf{R}_{(i^*)} = \mathbf{t}_{i^*} \left( \left( \otimes_{i=I}^{i^*+1} \mathbf{t}_i \right) \otimes \left( \otimes_{i=i^*-1}^1 \mathbf{t}_i \right) \right)^\top,$$

for any index  $i^* \in [I]$  which we use to avoid confusion with iteration index  $i$  in the Kronecker product. The  $i$ th mode product of a matrix  $\mathbf{D}_i \in \mathbb{C}^{N_i \times M_i}$  with  $\mathcal{T}$  is denoted by

$$\mathcal{M} = \mathcal{T} \times_i \mathbf{D}_i,$$

where  $\mathcal{M} \in \mathbb{C}^{M_1 \times \dots \times M_{i-1} \times N_i \times M_{i+1} \times \dots \times M_I}$ . The  $i$ th matricization matrix of  $\mathcal{M}$ , i.e.,  $\mathbf{M}_{(i)}$  is given by

$$\mathbf{M}_{(i)} = \mathbf{D}_i \mathbf{T}_{(i)}.$$

For a tensor  $\mathcal{T} \in \mathbb{C}^{M_1 \times \dots \times M_I}$ , its HOSVD is given by

$$\mathcal{T} = \mathcal{C} \times_1 \mathbf{E}_1 \times_2 \mathbf{E}_2 \dots \times_I \mathbf{E}_I,$$

where  $\mathbf{E}_i \in \mathbb{C}^{M_i \times M_i}$  is the left singular matrix of  $\mathbf{T}_{(i)}$ . Tensor  $\mathcal{C}$  is the core tensor of  $\mathcal{T}$ , obtained as  $\mathcal{C} = \mathcal{T} \times_1 \mathbf{E}_1^\mathsf{H} \times_2 \mathbf{E}_2^\mathsf{H} \dots \times_I \mathbf{E}_I^\mathsf{H}$ . If we denote the rank of  $\mathbf{T}_{(i)}$  as  $R_i$  for  $i \in [I]$ , then tensor  $\mathcal{T}$  is said to have a multilinear rank  $(R_1, \dots, R_I)$ .

## II. OFF-GRID SPARSE RECOVERY ALGORITHM FOR KRONECKER-STRUCTURED MEASUREMENTS

In this section, we study the parameter estimation problem with Kronecker-structured measurements. The signal model is

$$\bar{\mathbf{y}} = \otimes_{i=1}^I \bar{\mathbf{H}}_{i, \bar{\psi}_i} \bar{\mathbf{x}}_i + \bar{\mathbf{n}} = \otimes_{i=1}^I \mathbf{y}_i + \bar{\mathbf{n}}, \quad (5)$$

where the noise term in  $\bar{\mathbf{n}}$  need not be Kronecker-structured. For  $i \in [I]$ , the matrix  $\bar{\mathbf{H}}_{i, \bar{\psi}_i}$  is parameterized by  $\bar{\psi}_i := [\bar{\psi}_{i,1}, \dots, \bar{\psi}_{i,S_i}]^\top \in \mathbb{R}^{S_i}$  as follows,

$$\bar{\mathbf{H}}_{i, \bar{\psi}_i} := [\mathbf{h}_i(\bar{\psi}_{i,1}) \quad \dots \quad \mathbf{h}_i(\bar{\psi}_{i,S_i})] \in \mathbb{C}^{M_i \times S_i},$$

where  $\mathbf{h}_i \in \mathbb{C}^{M_i}$  is a *known* and *continuous* column function. The scalar  $S_i$  is the number of unknowns in  $\bar{\mathbf{H}}_{i, \bar{\psi}_i}$ . We assume  $\bar{\psi}_{i,s} \in [\psi_{i,1}, \psi_{i,r}]$ , a *known* compact range of the unknown parameters, and the goal is to estimate  $\bar{\psi}_i$  and  $\bar{\mathbf{x}}_i$  from (5).

To ensure identifiability of  $\bar{\psi}_i$ , we assume  $\mathbf{h}_i(\psi_p) \neq \mathbf{h}_i(\psi_q)$  for any  $\psi_p \neq \psi_q$ . Identifiability of  $\bar{\mathbf{x}}_i$  is limited by the Kronecker structure, i.e., for scalars  $\{\alpha_i\}_{i=1}^I$  with  $\prod_{i=1}^I \alpha_i = 1$ , the set of vectors  $\{\bar{\mathbf{x}}_i\}_{i=1}^I$  and  $\{\alpha_i \bar{\mathbf{x}}_i\}_{i=1}^I$  both result in  $\bar{\mathbf{y}}$  when combined with a given noise vector  $\bar{\mathbf{n}}$ . However, in many applications (e.g., channel estimation [5], [8]), the goal is to recover the solution up to a scaling factor, as we later elaborate in Sec. IV. Therefore, we aim to jointly obtain  $\bar{\psi}_i$  and the coefficient  $\bar{\mathbf{x}}_i$  up to scaling ambiguities, given *i*) measurement  $\bar{\mathbf{y}}$ , *ii*) vector function  $\mathbf{h}_i$ , and *iii*) range  $[\psi_{i,1}, \psi_{i,r}]$  for  $i \in [I]$ .

We devise a two-step solution: the first step decomposes (5) into  $I$  subproblems, each estimating  $\bar{\psi}_i$  and  $\bar{\mathbf{x}}_i$ , and the second step solves these subproblems using an off-grid approach.

### A. Step 1: Decomposition-Based Algorithm

To develop the decomposition algorithm, we use Lemma 1 for the noiseless set of linear equations,  $\bar{\mathbf{y}} = \otimes_{i=1}^I \bar{\mathbf{H}}_{i, \bar{\psi}_i} \bar{\mathbf{x}}_i$ .

*Lemma 1* [7, Lemma 4]: Consider linear equations  $\mathbf{y}_1 \otimes \mathbf{y}_2 = (\mathbf{H}_1 \otimes \mathbf{H}_2)(\mathbf{x}_1 \otimes \mathbf{x}_2) \neq \mathbf{0}$ . Solving for  $\mathbf{x}_1 \otimes \mathbf{x}_2$  from the equations is equivalent to solving for  $\mathbf{x}_1$  and  $\mathbf{x}_2$  from  $\mathbf{H}_1(\alpha \mathbf{x}_1) = \mathbf{y}_1$  and  $\mathbf{H}_2(\alpha^{-1} \mathbf{x}_2) = \mathbf{y}_2$ , for any scalar  $\alpha \neq 0$ .

Lemma 1 indicates that we can estimate individual vectors  $\mathbf{x}_1$  and  $\mathbf{x}_2$ , up to a scaling ambiguity  $\alpha$ . Therefore, if  $\bar{\mathbf{y}}$  is split into  $I$  low-dimensional vectors  $\{\hat{\mathbf{y}}_i \in \mathbb{C}^{M_i}\}_{i=1}^I$ , then (5) in the noiseless case ( $\bar{\mathbf{n}} = \mathbf{0}$ ) can be decomposed into  $I$  subproblems, each with ambiguity  $\{\alpha_i \neq 0\}_{i=1}^I$  with  $\prod_{i=1}^I \alpha_i = 1$ . This approach allows solving for  $\bar{\mathbf{x}}_i$  individually, rather than jointly. We now discuss the decomposition of  $\bar{\mathbf{y}}$  into low-dimensional vectors  $\hat{\mathbf{y}}_i$ 's, in both noiseless and noisy cases.

*1) Noiseless Setting and HOSVD:* In the noiseless case, we aim to find  $\{\hat{\mathbf{y}}_i\}_{i=1}^I$  so that  $\bar{\mathbf{y}} = \mathbf{y} = \otimes_{i=1}^I \hat{\mathbf{y}}_i$ . This can be achieved using HOSVD applied to the tensor representation of  $\bar{\mathbf{y}}$ . Using (1),  $\bar{\mathbf{y}}$  can be represented as an  $I$ th order tensor<sup>1</sup>  $\mathcal{Y} = \circ_{i=1}^I \mathbf{y}_i \in \mathbb{C}^{M_1 \times \dots \times M_I}$ , where  $\circ$  is the tensor outer product. Its  $i^*$ th mode matricization for any  $i^* \in [I-1]$  is

$$\mathbf{Y}_{(i^*)} = \mathbf{y}_{i^*} \left( \left( \otimes_{i=I}^{i^*+1} \mathbf{y}_i \right) \otimes \left( \otimes_{i=i^*-1}^1 \mathbf{y}_i \right) \right)^\top, \quad (6)$$

where  $\mathbf{y}_{i^*} \in \mathbb{C}^{M_{i^*}}$  is the  $i^*$ th component in the Kronecker product  $\otimes_{i=1}^I \mathbf{y}_i$ , and  $(\cdot)^\top$  is the transpose operator. We use the index  $i^*$  in (6) and the subsequent discussion to avoid confusion with the iteration index  $i$  in the Kronecker product expression. Now, an estimate  $\hat{\mathbf{y}}_{i^*}$  of  $\mathbf{y}_{i^*}$  up to scaling ambiguities is the left leading singular vector  $\mathbf{e}_{i^*}$  of the rank-one matrix  $\mathbf{Y}_{(i^*)}$ ,

<sup>1</sup>From [29],  $\text{vec}(\mathcal{Y}) = \text{vec}(\circ_{i=1}^I \mathbf{y}_i) = \otimes_{i=1}^I \mathbf{y}_i$  where the subscript is descending. For simplicity, we use ascending subscripts in the tensor outer product, resulting in  $\mathcal{Y}$  and  $\bar{\mathbf{y}}$  containing identical entries, albeit reordered.

i.e.,  $\hat{\mathbf{y}}_{i^*} = \mathbf{e}_{i^*}$ , for  $i^* \in [I-1]$ . For  $i^* = I$ , the estimate  $\hat{\mathbf{y}}_I$  is  $\mathbf{e}_I$  multiplied by the leading singular value of  $\mathbf{Y}_{(I)}$ , ensuring  $\otimes_{i=1}^I \hat{\mathbf{y}}_i = \otimes_{i=1}^I \mathbf{y}_i$ . The decomposition is called the HOSVD, assuming a multilinear rank of  $(1, \dots, 1)$  due to the Kronecker structure [29], [30], [31].

2) *Noisy Case and Truncated HOSVD*: Extending to the noisy setting, the decomposition step becomes

$$\{\hat{\mathbf{y}}_i\}_{i=1}^I = \arg \min_{\{\mathbf{z}_i \in \mathbb{C}^{M_i}\}_{i=1}^I} \|\bar{\mathbf{y}} - \otimes_{i=1}^I \mathbf{z}_i\|_2, \quad (7)$$

where  $\|\cdot\|_2$  is the vector  $\ell_2$  norm. We see that (7) is the same as seeking a tensor  $\hat{\mathcal{Y}} = \otimes_{i=1}^I \mathbf{z}_i$  with multilinear rank  $(1, \dots, 1)$  from measurement tensor  $\bar{\mathcal{Y}}$  obtained from  $\bar{\mathbf{y}}$  as

$$\min_{\hat{\mathcal{Y}}} \|\bar{\mathcal{Y}} - \hat{\mathcal{Y}}\|_{\text{F}} \text{ s.t. multilinear rank of } \hat{\mathcal{Y}} \text{ is } (1, \dots, 1), \quad (8)$$

where  $\|\cdot\|_{\text{F}}$  is the Frobenius norm. Unlike the noiseless case, here the  $i$ th mode matricization  $\bar{\mathbf{Y}}_{(i)}$  of  $\bar{\mathcal{Y}}$  is not rank-one due to noise. We solve (8) through the truncated HOSVD, where only the left leading singular vector is selected. We obtain  $\hat{\mathcal{Y}} = \xi \times_1 \mathbf{e}_1 \dots \times_I \mathbf{e}_I$  and  $\xi = \bar{\mathbf{y}} \times_1 \mathbf{e}_1^H \dots \times_I \mathbf{e}_I^H$ , where  $\mathbf{e}_i$  is the left leading singular vector of the  $i$ th mode matricization  $\bar{\mathbf{Y}}_{(i)}$  of  $\bar{\mathcal{Y}}$  for  $i \in [I]$  [32]. Here, operator  $\times_i$  is the  $i$ th tensor mode product and  $(\cdot)^H$  is the conjugate transpose. Then, a solution to (7) is  $\hat{\mathbf{y}}_i = \mathbf{e}_i$  for  $i \in [I-1]$  and  $\hat{\mathbf{y}}_I = \xi \mathbf{e}_I$ .

3) *A Low-complexity Approximation*: When  $I$  and  $M_i$  are large, HOSVD can become computationally intensive due to SVD needed to obtain  $\mathbf{e}_i$  for  $i \in [I-1]$ . Hence, we offer a low-complexity method using recursive SVD-based rank-one approximations,

$$(\hat{\mathbf{y}}_i, \bar{\mathbf{y}}_i) = \arg \min_{(\mathbf{z}_i \in \mathbb{C}^{M_i}, \bar{\mathbf{z}}_i), \|\mathbf{z}_i\|_2=1} \|\bar{\mathbf{y}}_{i-1} - \mathbf{z}_i \otimes \bar{\mathbf{z}}_i\|_2, \quad (9)$$

for  $i \in [I-1]$  where  $\bar{\mathbf{y}}_0 = \bar{\mathbf{y}}$  and  $\bar{\mathbf{y}}_{I-1} = \hat{\mathbf{y}}_I$ . For example, we consider the case when  $i = 1$ . We rearrange  $\bar{\mathbf{y}}$  as  $\bar{\mathbf{Y}} \in \mathbb{C}^{\bar{M}/M_1 \times M_1}$  where  $\text{vec}(\bar{\mathbf{Y}}) = \bar{\mathbf{y}}$ . Since  $\mathbf{z}_i \otimes \bar{\mathbf{z}}_i = \text{vec}(\bar{\mathbf{z}}_i \mathbf{z}_i^T)$ , (9) is equivalent to a rank-one approximation that minimizes  $\|\bar{\mathbf{Y}} - \bar{\mathbf{z}}_i \mathbf{z}_i^T\|_{\text{F}}$ , and  $\hat{\mathbf{y}}_i$  is the leading singular vector of  $\bar{\mathbf{Y}}$ .

Compared to HOSVD, here, the problem dimension decreases with  $i$  as  $\bar{\mathbf{z}}_i \in \mathbb{C}^{\prod_{j>i} M_j}$ , and the overall complexity is dominated by the first step, i.e.,  $i = 1$ . Besides, in the noiseless case, (9) and HOSVD yield the same solution.

Combining the decomposition step for  $\mathbf{y}$  obtaining  $\{\hat{\mathbf{y}}_i\}_{i=1}^I$  with Lemma 1, we break down the original  $\bar{M}$ -dimensional problem into  $I$  subproblems of dimensions  $\{M_i\}_{i=1}^I$ ,

$$\hat{\mathbf{y}}_i = \bar{\mathbf{H}}_{i, \bar{\psi}_i} \bar{\mathbf{x}}_i + \bar{\mathbf{n}}_i, \quad i \in [I], \quad (10)$$

which can be solved in parallel. Here, we assume  $\alpha_i = 1$  without loss of generality, as we seek solutions up to a scaling factor. The decomposition better exploits the Kronecker structure in the measurements, aiding denoising (see Sec. III-A) and reducing the complexity (see Sec. III-D). Before presenting these analyses, we first develop an algorithm to estimate  $\bar{\psi}_i$  and  $\bar{\mathbf{x}}_i$  from (10) for a given  $i$ .

### B. Step 2: Off-Grid SBL-Based Estimation Algorithm

In each dimension, the subproblem takes the general form of  $\hat{\mathbf{y}} = \bar{\mathbf{H}}_{\bar{\psi}} \bar{\mathbf{x}} + \bar{\mathbf{n}}$  for a nonlinear function  $\mathbf{h}$  when entries of  $\bar{\psi}$  belong to  $[\psi_1, \psi_r]$ , where we drop the dimension index  $i$ . We adopt BEM by discretizing  $[\psi_1, \psi_r]$  with a set of variables  $\psi \in \mathbb{R}^N$  with the  $n$ th variable  $\psi_n$ , forming the dictionary

$$\mathbf{H}(\psi) := [\mathbf{h}(\psi_1) \quad \dots \quad \mathbf{h}(\psi_N)] \in \mathbb{R}^{M \times N}.$$

This leads to the BEM with coefficient vector  $\mathbf{x}$  as

$$\hat{\mathbf{y}} = \mathbf{H}(\psi) \mathbf{x} + \mathbf{n}, \quad (11)$$

Only a few entries of  $\psi$  correspond to the true parameters  $\bar{\psi}$ , making  $\mathbf{x}$  sparse. However, solving (11) by fixing  $\psi$  at some predefined grid points, as in standard sparse recovery, leads to grid mismatch. Thus, we treat  $\psi$  as variable and jointly estimate  $\psi$  and sparse  $\mathbf{x}$  using the SBL framework. Then, the nonzero entries of  $\mathbf{x}$  and their corresponding  $\psi_n$ 's are estimates of  $\bar{\mathbf{x}}$  and  $\bar{\psi}$ , respectively.

SBL adopts a fictitious zero mean complex Gaussian distribution  $\mathcal{CN}(\mathbf{x} | \mathbf{0}, \Gamma)$  as prior on the sparse vector  $\mathbf{x}$  with an unknown diagonal covariance matrix  $\Gamma$ . Let  $\Gamma = \text{diag}(\gamma) \in \mathbb{R}^{N \times N}$  with the diagonal entries  $\gamma \in \mathbb{R}^N$ . We assume Gaussian noise  $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I})$  with unknown variance  $\sigma^2$ . Using type II ML, we first estimate the hyperparameters  $\gamma$ ,  $\psi$ , and  $\sigma^2$ , and then, the estimate of  $\mathbf{x}$  is  $\arg \max_{\mathbf{x}} p(\mathbf{x} | \hat{\mathbf{y}}; \gamma, \psi, \sigma^2)$ . The ML estimates of the hyperparameters are

$$\min_{\gamma \geq \mathbf{0}, \psi \in [\psi_1, \psi_r]^N, \sigma^2 > 0} \mathcal{L}(\gamma, \psi, \sigma^2), \quad (12)$$

where  $\gamma \geq \mathbf{0}$  indicates that the entries of  $\gamma$  are nonnegative. Using the SBL priors, negative log-likelihood function  $\mathcal{L}$  is

$$\mathcal{L}(\gamma, \psi, \sigma^2) = \log p(\hat{\mathbf{y}}; \gamma, \psi, \sigma^2) = \log |\Sigma_y| + \text{tr}(\hat{\mathbf{y}}^H \Sigma_y^{-1} \hat{\mathbf{y}}),$$

where  $\Sigma_y = \sigma^2 \mathbf{I}_M + \mathbf{H}(\psi) \Gamma \mathbf{H}(\psi)^H$  and  $\text{tr}(\cdot)$  is the trace operator. We resort to the EM algorithm to solve (12). Specifically, the  $r$ th iteration of EM is

$$\begin{aligned} \text{E-step: } & Q(\gamma, \psi, \sigma^2) \\ & := \mathbb{E}_{\mathbf{x} | \hat{\mathbf{y}}; \{\gamma, \psi, \sigma^2\}^{(r)}} \{\log p(\hat{\mathbf{y}}, \mathbf{x}; \gamma, \psi, \sigma^2)\}, \\ \text{M-step: } & \{\gamma, \psi, \sigma^2\}^{(r+1)} = \arg \max_{\gamma > \mathbf{0}, \psi \in [\psi_1, \psi_r]^N, \sigma^2 > 0} Q(\gamma, \psi, \sigma^2). \end{aligned}$$

Here,  $\gamma > \mathbf{0}$  means the entries of  $\gamma$  should be positive to avoid degenerate distributions. Further, we note that

$$\begin{aligned} Q(\gamma, \psi, \sigma^2) &= \mathbb{E}_{\mathbf{x} | \hat{\mathbf{y}}; \{\gamma, \psi, \sigma^2\}^{(r)}} \{\log p(\hat{\mathbf{y}} | \mathbf{x}; \psi, \sigma^2)\} \\ &+ \mathbb{E}_{\mathbf{x} | \hat{\mathbf{y}}; \{\gamma, \psi, \sigma^2\}^{(r)}} \{\log p(\mathbf{x}; \gamma)\}, \end{aligned}$$

and thus, the optimization problem in the M-step is separable in  $\gamma$  and  $\{\psi, \sigma^2\}$ . The optimization problem in  $\gamma$  is

$$\gamma^{(r+1)} = \arg \min_{\gamma > \mathbf{0}} \log |\text{diag}(\gamma)| + (\mathbf{d}^{(r)})^T \gamma^{-1} = \mathbf{d}^{(r)}, \quad (13)$$

with  $\mathbf{d}^{(r)} = \text{diag}(\Sigma_x + \mu_x \mu_x^H)$  where  $\text{diag}(\cdot)$  returns the diagonal entries of the matrix input, and  $\gamma^{-1}$  representing element-wise inversion. Here,  $\mu_x$  and  $\Sigma_x$  are the mean and variance of conditional distribution  $p(\mathbf{x} | \hat{\mathbf{y}}; \{\gamma, \psi, \sigma^2\}^{(r)})$ , respectively,

$$\mu_x = (\sigma^{-2})^{(r)} \Sigma_x \mathbf{H}(\psi^{(r)})^H \hat{\mathbf{y}},$$

$$\Sigma_x = \left[ (\sigma^{-2})^{(r)} \mathbf{H}(\psi^{(r)})^H \mathbf{H}(\psi^{(r)}) + \text{diag}(\gamma^{(r)})^{-1} \right]^{-1}. \quad (14)$$

Optimizing  $\psi$  and  $\sigma^2$  in the M-step yields

$$\{\psi, \sigma^2\}^{(r+1)} = \arg \min_{\psi \in [\psi_1, \psi_r]^N, \sigma^2 > 0} N \log \sigma + \frac{1}{2\sigma^2} g(\psi),$$

where  $g(\psi) := \|\hat{\mathbf{y}} - \mathbf{H}(\psi)\mu_x\|_2^2 + \text{tr}(\Sigma_x \mathbf{H}(\psi)^H \mathbf{H}(\psi))$ , is independent of  $\sigma^2$ . Given  $\psi^{(r+1)}$ , we update  $\sigma^2$  as

$$(\sigma^2)^{(r+1)} = g(\psi^{(r+1)})/N. \quad (15)$$

Further,  $\psi^{(r+1)} = \arg \min_{\psi} g(\psi)$  simplifies to

$$\begin{aligned} \psi^{(r+1)} = \arg \min_{\psi \in [\psi_1, \psi_r]^N} & \text{tr}(\mathbf{H}(\psi) \Sigma \mathbf{H}(\psi)^H) \\ & - 2\text{Re}\{\text{tr}(\mathbf{M} \mathbf{H}(\psi))\}, \end{aligned} \quad (16)$$

where  $\Sigma := \Sigma_x + \mu_x \mu_x^H$  and  $\mathbf{M} = \mu_x \hat{\mathbf{y}}^H$ . We use the alternating minimization method to solve (16), where we alternatively optimize one entry of  $\psi$  while keeping all others fixed. The  $t$ th iteration of the alternating minimization method updates the  $n$ th variable  $\psi_{n^*}$  by minimizing

$$f_{n^*}(\psi_{n^*}) = 2\text{Re}\{\mathbf{v}_{n^*}^H \mathbf{h}(\psi_{n^*})\} + \Sigma_{n^*, n^*} \mathbf{h}(\psi_{n^*})^H \mathbf{h}(\psi_{n^*}).$$

Here,  $\mathbf{v}_{n^*} \in \mathbb{C}^M$  is defined as

$$\begin{aligned} \mathbf{v}_{n^*} = & \sum_{n=1}^{n^*-1} \Sigma_{n^*, n} \mathbf{h}(\psi_n^{(r, t+1)}) \\ & + \sum_{n=n^*+1}^N \Sigma_{n^*, n} \mathbf{h}(\psi_n^{(r, t)}) - \mathbf{M}_{n^*, :}^H, \end{aligned}$$

where  $\mathbf{M}_{n^*, :}$  is the  $n^*$ th row of  $\mathbf{M}$ , and  $\Sigma_{n^*, n^*}$  is the  $(n^*, n^*)$ th entry of  $\Sigma$ . We use index  $n^*$  to avoid confusion with iteration index  $n$ . Hence, the  $t$ th alternative minimization iterate  $\psi_n^{(r, t+1)}$  in the  $r$ th EM iteration for index  $n^* \in [N]$  is

$$\psi_{n^*}^{(r, t+1)} = \arg \min_{\frac{\psi_{n^*}^{(r, t)} + \psi_{n^*-1}^{(r, t)}}{2} \leq \psi \leq \frac{\psi_{n^*}^{(r, t)} + \psi_{n^*+1}^{(r, t)}}{2}} f_{n^*}(\psi). \quad (17)$$

The assumption  $\mathbf{h}_i(\psi_p) \neq \mathbf{h}_i(\psi_q)$  for any  $\psi_p \neq \psi_q$  ensures the solution identifiability of problem (17). Although  $f_{n^*}(\psi)$  in (17) is a nonlinear function of  $\psi$ , we avoid local first-order approximations [18] by using the alternating minimization. It translates the problem (16) into simple *one dimension* subproblems (17). We solve (17) using a simple (one-dimensional) grid search, preserving accuracy and ensuring easy implementation. Our off-grid SBL (OffSBL) and the overall decomposition-based SBL (dSBL) are outlined in Algorithms 1 and 2, respectively.

So far, we discussed our dSBL framework. Next, we discuss how the recovered signals  $\{\mathbf{x}_i\}_{i=1}^I$  are used to infer coefficients  $\{\bar{\mathbf{x}}_i\}_{i=1}^I$ , unknown parameters  $\{\psi_i\}_{i=1}^I$ , and  $\{S_i\}_{i=1}^I$ . Through automatic relevance determination mechanism [33], SBL promotes sparsity in  $\mathbf{x}_i$  by learning a sparse hyperparameter  $\gamma_i$  for  $i \in [I]$ . We note that  $\gamma_i$  is the hyperparameter involved in (12) where the subscript  $i$  referring to the  $i$ th subproblem (10) is dropped. Then, the support of  $\mathbf{x}_i$ , indicated by the significant

---

**Algorithm 1:** OffSBL.

---

**Input:** Measurement  $\hat{\mathbf{y}}$ , the number of initial grids  $N$ , the range  $[\psi_1, \psi_r]$ , and thresholds  $\epsilon_1, \epsilon_2 < 1$

- 1: Set  $r = 0$ , initialize  $\gamma^{(0)} = \mathbf{1}$ , and initialize  $\psi^{(0)}$  with uniform samples  $\{\theta_n\}_{n=1}^N$  from  $[\psi_1, \psi_r]$ .
- 2: **repeat**
- 3:     Compute  $\mu_x$  and  $\Sigma_x$  using (14)
- 4:     Update  $\gamma^{(r+1)}$  using (13)
- 5:     Set  $t = 0$ , initialize  $\psi^{(r, t)} = \psi^{(r)}$
- 6:     **repeat**
- 7:         Obtain  $\psi_{n^*}^{(r, t+1)}$  by solving (17), for  $n^* \in [N]$
- 8:         Set  $t = t + 1$
- 9:     **until**  $\|\psi^{(t+1)} - \psi^{(t)}\| < \epsilon_1$
- 10:    Let  $\psi^{(r+1)} = \psi^{(r, t)}$
- 11:    Obtain  $(\sigma^2)^{(r+1)}$  using (15) and set  $r = r + 1$
- 12: **until**  $\|\gamma^{(r+1)} - \gamma^{(r)}\|_2 / \|\gamma^{(r)}\|_2 < \epsilon_2$

**Output:** Estimated  $\mathbf{x} = \mu_x$ .

---

**Algorithm 2:** Decomposition-Based SBL (dSBL).

---

**Input:** Measurement  $\bar{\mathbf{y}}$ , dictionaries  $\mathbf{H}_i$  for  $i \in [I]$

- 1: **for**  $i \in [I]$  **do**
- 2:     Solve (8) or (9) to obtain  $\hat{\mathbf{y}}_i$
- 3:     Solve (10) for  $\mathbf{x}_i$  using Algorithm 1
- 4: **end for**

**Output:** For  $i \in [I]$ ,  $\bar{\mathbf{x}}_i$  corresponds to the nonzero entries of  $\mathbf{x}_i$  and  $\bar{\psi}_i$  corresponds to the support of  $\mathbf{x}_i$ .

---

peaks in amplitude, reveals the estimated active components  $\bar{\psi}_i$ , while its values are the estimates of  $\bar{\mathbf{x}}_i$ ; the number of estimated active components is  $S_i$ . Due to computational limits, we cap the number of EM iterations, so the inactive components are not exactly zero but a very small value. We threshold  $\gamma_i$  (e.g.,  $10^{-4}$ ) to prune the values below the threshold and retain the other values for estimating  $\{\bar{\mathbf{x}}_i\}_{i=1}^I$  and  $\{\bar{\psi}_i\}_{i=1}^I$ , as explained before.

### III. THEORETICAL ANALYSIS AND EXTENSIONS

This section analyzes our dSBL algorithm, covering the decomposition error bound and denoising effect of HOSVD and convergence results of OffSBL. We then present the complexity analysis of dSBL, demonstrating its computational advantage compared to other methods. Finally, we discuss extensions of our algorithms to other similar signal models.

#### A. Analysis of Decomposition-Based Algorithm

We start with the decomposition error bound, where we quantify the error between the decomposed vectors  $\{\hat{\mathbf{y}}_i\}_{i=1}^I$  and the true signal components  $\{\mathbf{y}_i = \mathbf{H}_{i, \bar{\psi}_i} \bar{\mathbf{x}}_i\}_{i=1}^I$ . We measure the error as the angle between  $\mathbf{y}_i$  and  $\hat{\mathbf{y}}_i$ , accounting for scaling ambiguity in the decomposition step.

*Theorem 1 (Decomposition Accuracy):* Let  $\bar{\mathbf{y}} = \otimes_{i=1}^I \mathbf{y}_i + \bar{\mathbf{n}} \in \mathbb{R}^{\bar{M}}$  be the noisy measurement as given in (5), where  $\mathbf{y}_i \in$

$\mathbb{R}^{M_i}$  and  $\bar{\mathbf{n}}$  has independent zero-mean Gaussian entries with variance  $\sigma_t^2$ . Suppose the signal satisfies

$$\lambda^2 = \|\otimes_{i=1}^I \mathbf{y}_i\|_2^2 / \sigma_t^2 \geq C_{\text{gap}} \left( \sqrt{\bar{M}} + \max_{1 \leq i \leq I} M_i \right),$$

for a large constant  $C_{\text{gap}} > 0$ . Then, there exist constants  $c, C > 0$  such that with probability at least  $1 - C \exp\{-cM_i\}$ ,

$$\sin(\vartheta_i) \leq C \left( \sqrt{M_i} \lambda^{-1} + \sqrt{\bar{M}} \lambda^{-2} \right), \quad i \in [I],$$

where  $\vartheta_i := \arccos(|\mathbf{y}_i^\top \hat{\mathbf{y}}_i| / (\|\mathbf{y}_i\|_2 \|\hat{\mathbf{y}}_i\|_2))$  is the angle between  $\mathbf{y}_i$  and its estimate  $\hat{\mathbf{y}}_i$  obtained by solving (7) and (8).

*Proof:* See Appendix A.  $\square$

We note that  $\lambda/\bar{M}$  represents the signal-to-noise ratio (SNR) of our measurement model, and a higher SNR (i.e., a larger  $\lambda$ ) improves estimation accuracy, as expected. As  $\lambda$  goes to  $\infty$  (noiseless case), the error bounds approach zero. Conversely, when the signal strength is insufficient to meet the required condition, there is no consistent estimator for  $\mathbf{y}_i$ 's [31].

While the decomposition accuracy reflects how well  $\hat{\mathbf{y}}_i$  aligns with the true signal  $\mathbf{y}_i$ , we can also access the noise level after decomposition. We next quantify the denoising effect of the decomposition step, which refers to noise reduction in the measurements, i.e.,  $\|\otimes_{i=1}^I \hat{\mathbf{y}}_i - \otimes_{i=1}^I \mathbf{y}_i\|_2^2$  is expected to be smaller than  $\mathbb{E}\{\|\bar{\mathbf{n}}\|_2^2\} = \sigma_t^2 \bar{M}$ , as summarized in the following result.

**Theorem 2 (Denoising Effect):** Let  $\bar{\mathbf{y}} = \otimes_{i=1}^I \mathbf{y}_i + \bar{\mathbf{n}} \in \mathbb{R}^{\bar{M}}$  denote the noisy measurement as in (5), where  $\mathbf{y}_i \in \mathbb{R}^{M_i}$  and  $\bar{\mathbf{n}}$  has independent zero-mean Gaussian entries with variance  $\sigma_t^2$ . Let  $\hat{\mathbf{y}}_i$  denote the estimate of  $\mathbf{y}_i$  obtained by solving (7) and (8). Then, the estimates satisfy

$$\begin{aligned} & \|\otimes_{i=1}^I \hat{\mathbf{y}}_i - \otimes_{i=1}^I \mathbf{y}_i\|_2 \\ & \leq \sigma_t \left( 2 \sum_{i=1}^I \left[ 3\sqrt{M_i} + \sqrt{\bar{M}/M_i} \right] + 1 + 2\sqrt{\max_{1 \leq j \leq I} M_j} \right), \end{aligned} \quad (18)$$

with probability exceeding  $1 - 3 \sum_{i=1}^I e^{-M_i}$ . Moreover,

$$\mathbb{E} \left\{ \|\otimes_{i=1}^I \hat{\mathbf{y}}_i - \otimes_{i=1}^I \mathbf{y}_i\|_2^2 \right\} \approx \sigma_t^2 \left( \sum_{i=1}^I M_i + 1 - I \right). \quad (19)$$

*Proof:* See Appendix B.  $\square$

To gain insights from Theorem 2, suppose that  $M_i = \mathcal{O}(M)$  for  $i \in [I]$  for some value  $M$ . Then, (18) shows that the noise level  $\|\otimes_{i=1}^I \hat{\mathbf{y}}_i - \otimes_{i=1}^I \mathbf{y}_i\|_2^2$  in the decomposed signal is dominated by the term  $\mathcal{O}(\sigma_t^2 \bar{M}/M_i) = \mathcal{O}(M^{I-1} \sigma_t^2)$  for  $I \geq 2$ . Then, in the general situation with  $I \geq 2$ , compared to the noisy signal  $\bar{\mathbf{y}}$ , the noise level reduces from  $\mathcal{O}(M^I \sigma_t^2)$  to  $\mathcal{O}(M^{I-1} \sigma_t^2)$  after HOSVD. Besides, from (19), the average noise level can be approximated as  $\mathcal{O}(M^I \sigma_t^2)$  when  $M > I$ . Specifically, the noise level  $\mathbb{E}\{\|\bar{\mathbf{n}}\|^2\} = \sigma_t^2 \bar{M}$  reduces approximately by

$$\frac{\mathbb{E} \left\{ \|\otimes_{i=1}^I \hat{\mathbf{y}}_i - \otimes_{i=1}^I \mathbf{y}_i\|_2^2 \right\}}{\mathbb{E}\{\|\bar{\mathbf{n}}\|_2^2\}} \approx \frac{\sum_{i=1}^I M_i + 1 - I}{\bar{M}} < 1.$$

The probabilistic bound in Theorem 2 can also be interpreted as an error bound for HOSVD. Consider the simplest case

$I = 2$  and fix  $\bar{M} = M_1 M_2$  and  $\sigma_t$ . The upper bound in (18) (with respect to  $\sigma_t$ ) can be bounded from below as

$$\begin{aligned} & 8(\sqrt{M_1} + \sqrt{M_2}) + 1 + 2\sqrt{\max\{M_1, M_2\}} \\ & \geq 8 \left[ \min_{\substack{M_1, M_2 \\ M_1 M_2 = \bar{M}}} \sqrt{M_1} + \sqrt{M_2} \right] + 1 + 2\bar{M}^{1/4} = 18\bar{M}^{1/4} + 1, \end{aligned}$$

where equality is achieved when  $M_1 = M_2$ . Therefore, the bound in (18) is minimized when the vectors  $\mathbf{y}_i$ 's have the same size.

We present the next corollary on the low complexity approximation, obtained by setting  $I = 2$  and  $M_2 = \bar{M}/M_1$  in Theorem 2, noting that it has the same first step as HOSVD.

**Corollary 1:** Under the assumptions of Theorem 2, if  $\bar{\mathbf{y}}_1$  and  $\hat{\mathbf{y}}_1$  are obtained from  $\bar{\mathbf{y}}$  using the low complexity approximation (9), with probability at least  $1 - 3(e^{-M_1} + e^{-\bar{M}/M_1})$ ,

$$\|\hat{\mathbf{y}}_1 \otimes \bar{\mathbf{y}}_1 - \bar{\mathbf{y}}\|_2 \leq \sigma_t \left( 1 + 18\sqrt{\max\{M_1, \bar{M}/M_1\}} \right),$$

$$\text{and } \mathbb{E} \left\{ \|\hat{\mathbf{y}}_1 \otimes \bar{\mathbf{y}}_1 - \bar{\mathbf{y}}\|_2^2 \right\} \approx \sigma_t^2 (M_1 + \bar{M}/M_1 - 1).$$

Corollary 1 shows that the first step of low complexity approximation also aids denoising. To intuitively see this, we reorganize  $\otimes_{i=1}^I \mathbf{y}_i$  into a rank-one matrix  $\mathbf{Y} := (\otimes_{i=2}^I \mathbf{y}_i) \mathbf{y}_1^\top$

$$\bar{\mathbf{Y}} = \mathbf{Y} + \bar{\mathbf{N}} = (\otimes_{i=2}^I \mathbf{y}_i) \mathbf{y}_1^\top + \bar{\mathbf{N}},$$

with  $\text{vec}(\bar{\mathbf{Y}}) = \bar{\mathbf{y}}$ ,  $\text{vec}(\mathbf{Y}) = \otimes_{i=1}^I \mathbf{y}_i$ , and  $\text{vec}(\bar{\mathbf{N}}) = \bar{\mathbf{n}}$ . Here, the noise term is unstructured  $\bar{\mathbf{N}}$  and typically has full rank. The first step ( $i = 1$ ) of (9), yields  $\hat{\mathbf{y}}_1$  and  $\bar{\mathbf{y}}_1$ , which estimate  $\mathbf{Y}$  as  $\bar{\mathbf{y}}_1 \hat{\mathbf{y}}_1^\top$ . Comparing this estimate with  $\bar{\mathbf{Y}}$ , we observe that  $\bar{\mathbf{y}}_1 \hat{\mathbf{y}}_1^\top$  preserves the rank-one structure of the signal. It discards the components that violate the rank-one constraint, which are often attributed to noise, effectively performing denoising. However, a drawback of the low-complexity approximation is that the error from one step can propagate to the subsequent steps, as later computations depend on estimates from the previous steps. In contrast, the error in HOSVD is independent in each subspace, leading to a better decomposition but with a higher computation cost.

### B. Analysis of OffSBL Algorithm

This section discusses the convergence results for OffSBL in Algorithm 1. We note that OffSBL is a two-level iterative algorithm, where the outer EM iteration is given by (14) followed by (13) and (15), and the inner loop updates the grid points via (17). We first provide the guarantees for the inner loop for a given  $r$ th EM iteration.

**Lemma 2 (Convergence of Update):** Consider the alternating update (17) in the  $r$ th EM iteration. If  $\sup_{\psi \in [\psi_1, \psi_2]} \|\mathbf{h}(\psi)\|_2^2 < \infty$ , the sequence  $\{g(\psi^{(r,t)})\}_{t=0}^\infty$  is non-increasing and convergent. Its iterate  $\{\psi^{(r,t)}\}_{t=0}^\infty$  adopts at least one limit point.

*Proof:* First, we note that there always exists an optimal solution to (17) due to the continuity of the function  $f_{n^*}$ . The extreme value theorem states that  $f_{n^*}(\psi)$  must reach a minimum at least once within the closed and bounded constraint set. The solvability of (17) further indicates

$$g(\psi^{(r,t+1)}) \leq g(\psi^{(r,t)}). \quad (20)$$

Further, the cost function  $g(\psi)$  is bounded from below as

$$\begin{aligned} g(\psi) &= \|\hat{\mathbf{y}} - \mathbf{H}(\psi)\mu_x\|_2^2 + \text{tr}(\Sigma_x \mathbf{H}(\psi)^\text{H} \mathbf{H}(\psi)) \\ &\geq -2\text{Re}(\text{tr}\{\mathbf{M}\mathbf{H}(\psi)\}) \\ &\geq -\text{tr}\{\mathbf{M}^\text{H} \mathbf{M}\} - \text{tr}\{\mathbf{H}^\text{H}(\psi) \mathbf{H}(\psi)\} \\ &\geq -\text{tr}\{\mathbf{M}^\text{H} \mathbf{M}\} - N \sup_{\psi \in [\psi_1, \psi_r]} \|\mathbf{h}(\psi)\|_2^2 > -\infty. \end{aligned}$$

Thus, by the monotone convergence theorem, the sequence  $\{g(\psi^{(r,t)})\}_{t=0}^\infty$  converges. The monotonicity (20) also ensures that  $\{\psi^{(r,t)}\}_{t=0}^\infty$  belongs to the sublevel set of  $\psi^{(r,0)}$ . Since  $g(\psi)$  is continuous, its sublevel sets are compact, implying that the sequence  $\{\psi^{(r,t)}\}_{t=0}^\infty$  adopts at least one limit point.  $\square$

We next prove the convergence of our OffSBL algorithm. For the convergence result, we assume the zero-mean Gaussian noise  $\mathbf{n}$  in (11) has a known variance  $\sigma^2 > 0$  and present the convergence based on iterates  $\{\gamma^{(r)}, \psi^{(r)}\}_{r=0}^\infty$ , i.e., we do not update  $\sigma^2$  via (15), but use the true value of  $\sigma^2$  in Algorithm 1. This assumption simplifies deriving a lower bound for the negative log-likelihood function (12), which is challenging when  $\sigma^2 = 0$  or treated as a variable. Moreover, assuming  $\sigma^2 > 0$  is standard in SBL analysis [34], [35] and the noiseless settings corresponds the limit where  $\sigma^2 \rightarrow 0$ . Thus, the results below are asymptotically applicable to the noiseless case.

*Theorem 3 (Convergence Property of OffSBL):* Consider the problem (11) where  $\mathbf{n}$  is zero-mean Gaussian noise with known variance  $\sigma^2 > 0$ , solved using OffSBL in Algorithm 1. If  $\sup_{\psi \in [\psi_1, \psi_r]} \|\mathbf{h}(\psi)\|_2^2 < \infty$ , the cost function sequence  $\{\mathcal{L}(\gamma^{(r)}, \psi^{(r)})\}_{r=0}^\infty$  converges monotonically to some value  $\mathcal{L}^*$ , and the sequence  $\{\gamma^{(r)}, \psi^{(r)}\}_{r=0}^\infty$  converges to a set  $\mathcal{S}^*$  with  $\mathcal{L}(\gamma, \psi) = \mathcal{L}^*$  for any  $\{\gamma, \psi\} \in \mathcal{S}^*$ .

*Proof:* We have  $Q(\gamma^{(r)}, \psi^{(r)}) \leq Q(\gamma^{(r+1)}, \psi^{(r+1)})$  in the  $r$ th EM iteration due to Lemma 2. Then, OffSBL is a generalized EM algorithm and the sequence  $\{\mathcal{L}(\gamma^{(r)}, \psi^{(r)})\}_{r=0}^\infty$  is nonincreasing [36]. Also,  $\mathcal{L}(\gamma, \psi)$  is bounded from below as

$$\mathcal{L}(\gamma, \psi) = \log |\Sigma_y| + \text{tr}(\hat{\mathbf{y}}^\text{H} \Sigma_y^{-1} \hat{\mathbf{y}}) \geq \log |\Sigma_y| \geq M \log \sigma^2,$$

where  $\Sigma_y^{-1} = (\sigma^2 \mathbf{I}_M + \mathbf{H}(\psi) \Gamma \mathbf{H}(\psi)^\text{H})^{-1}$  is positive definite for  $\gamma \geq \mathbf{0}$  and  $\psi \in [\psi_1, \psi_r]^N$ . The last inequality is because the eigenvalues of  $\Sigma_y$  are lower bounded by  $\sigma^2$  [7]. Thus, by the monotone convergence theorem, the sequence  $\{\mathcal{L}(\gamma^{(r)}, \psi^{(r)})\}_{r=0}^\infty$  converges to some value  $\mathcal{L}^*$ .

Further, the function  $\mathcal{L}$  is a coercive function of  $\gamma$  and continuous in both  $\gamma$  and  $\psi$  [34, Lemma 3]. Consequently, its sublevel sets are compact. The nonincreasing  $\{\mathcal{L}(\gamma^{(r)}, \psi^{(r)})\}_{r=0}^\infty$  indicates that  $\{\gamma^{(r)}, \psi^{(r)}\}_{r=0}^\infty$  adopts at least one limit point in the level set of  $\mathcal{L}^*$ , i.e.,  $\mathcal{S}^*$ , which completes the proof.  $\square$

We conclude by noting that although the properties of limit points of the EM iterations are unknown, OffSBL offers a sequence over which the negative log-likelihood  $\mathcal{L}$  is nonincreasing, aligning with the problem (12). We also empirically observe that OffSBL iterates also converge.

### C. Algorithm Complexity

For simplicity and interpretability, we assume  $M_i = \mathcal{O}(M)$ ,  $N_i = \mathcal{O}(N)$ , and  $I < M < N$ , where  $N_i$  is the number of

grids adopted in the BEM of (10). The time complexity of dSBL with HOSVD is  $\mathcal{O}(R_{\text{EM}} N^2 M I + I M^{I+1})$  while low-complexity approximation has  $\mathcal{O}(R_{\text{EM}} N^2 M I + M^{I+1})$ . Here,  $R_{\text{EM}}$  denotes the number of EM iterations. The difference between HOSVD and low complexity approximation-based schemes is roughly of order  $I$ . Meanwhile, both have similar space complexity, i.e.,  $\mathcal{O}(M^I + MN + N^2)$ . Also, all sparse recovery subproblems are independent of each other and can thus be solved in parallel. In that case, the time complexity changes to  $\mathcal{O}(M^{I+1} + R_{\text{EM}} N^2 M)$  for low-complexity approximation based and  $\mathcal{O}(I M^{I+1} + R_{\text{EM}} N^2 M)$  for HOSVD based, while the space complexity becomes  $\mathcal{O}(M^I + I M N + I N^2)$ . For comparison, the time complexity of AM- and SVD-KroSBL is  $\mathcal{O}(R_{\text{EM}} (R_{\text{AM}} I N^I + (MN)^I))$  and  $\mathcal{O}(R_{\text{EM}} (N^{I+1} + (MN)^I))$ , respectively, and the space complexity is  $\mathcal{O}((MN)^I)$  for both. Here,  $R_{\text{AM}}$  denotes the number of AM iterations in AM-KroSBL. Therefore, both the time and space complexities of our algorithm are several orders less than the state-of-the-art KroSBL methods.

### D. Extensions to Similar Structures

We reiterate that the decomposition and the OffSBL algorithms we presented are general algorithmic techniques and can also be applied independently. For instance, for estimation tasks involving Kronecker-structured signals, if there is no grid mismatch and the parameters  $\bar{\psi}_i$  lie on a discrete set, one can combine the decomposition algorithm with any sparse recovery algorithm. Similarly, OffSBL is a stand-alone off-grid algorithm for conventional linear inversion problems (i.e.,  $I = 1$ ). Moreover, these techniques can be extended to other parameter estimation models, as discussed next.

1) *Superposition of Kronecker-structured Data:* In several wideband orthogonal frequency division multiplexing multiple input multiple output (MIMO) systems [5], [6], [37], measurements  $\bar{\mathbf{y}}$  takes the form

$$\bar{\mathbf{y}} = \sum_{u=1}^U \left( \otimes_{i=1}^I \bar{\mathbf{H}}_{u,i, \bar{\psi}_{u,i}} \bar{\mathbf{x}}_{u,i} \right) + \bar{\mathbf{n}}. \quad (21)$$

where the special case of  $U = 1$  reduces to (5). Here, we can rewrite (21) using the tensor form as

$$\mathcal{Y} = \sum_{u=1}^U \circ_{i=1}^I \left( \bar{\mathbf{H}}_{u,i, \bar{\psi}_{u,i}} \bar{\mathbf{x}}_{u,i} \right) + \bar{\mathcal{N}}.$$

Thus, each factor  $\bar{\mathbf{H}}_{u,i, \bar{\psi}_{u,i}} \bar{\mathbf{x}}_{u,i}$  for  $u = [U]$  and  $i \in [I]$  can be obtained by tensor canonical polyadic decomposition under mild conditions for uniqueness of the decomposition [29], followed by OffSBL for unknown parameter estimation.

2) *Non-Kronecker-structured Sparse Vector:* Some applications [2], [7], [10] lead to the measurement model

$$\bar{\mathbf{y}} = \left( \otimes_{i=1}^I \bar{\mathbf{H}}_{i, \bar{\psi}_i} \right) \bar{\mathbf{x}} + \bar{\mathbf{n}},$$

where the coefficient vector  $\bar{\mathbf{x}}$  is not Kronecker-structured. Here, direct decomposition of  $\bar{\mathbf{y}}$  cannot be applied, but we can use the Kronecker-structured dictionary in multidimensional BEMs as  $\bar{\mathbf{H}} = \otimes_{i=1}^I \bar{\mathbf{H}}_i(\bar{\psi}_{(i)})$ . Here, vector  $\bar{\psi}_{(i)}$  collects all

the variables in the  $i$ th BEM dictionary matrix  $\mathbf{H}_i$ . Then, we arrive at the sparse vector problem

$$\bar{\mathbf{y}} = \otimes_{i=1}^I \mathbf{H}_i (\psi_{(i)}) \mathbf{x} + \bar{\mathbf{n}}, \quad (22)$$

which can be solved using the OffSBL algorithm. Specifically, we adopt a fictitious Gaussian prior distribution with covariance matrix  $\Gamma = \text{diag}(\gamma) \in \mathbb{R}^{\bar{N} \times \bar{N}}$  on  $\mathbf{x}$ . Then the estimates of  $\gamma$ ,  $\{\psi_{(i)}\}_{i=1}^I$  and  $\sigma^2$  are determined by the type II ML with EM algorithm. The hyperparameter  $\gamma$  and the noise variance  $\sigma^2$  can be similarly obtained using (13) and (15). We can exploit the Kronecker structure in (22) to simplify the alternating minimization of OffSBL for  $\{\psi_{(i)}\}_{i=1}^I$ . The EM update step for  $\{\psi_{(i)}\}_{i=1}^I$  is equivalent to minimizing

$$g\left(\{\psi_{(i)}\}_{i=1}^I\right) = \text{tr}\left(\mathbf{H}\Sigma\mathbf{H}^H\right) - 2\text{Re}\{\text{tr}(\mathbf{M}\mathbf{H})\}.$$

Let  $\psi_{i^*,n^*}$  be the  $n^*$ th variable of the  $i^*$ th BEM dictionary matrix  $\mathbf{H}_{i^*}$ . Then, the alternating minimization step optimizes  $\psi_{i^*,n^*}$  with other  $\psi_{i,n}$ 's being fixed, as in OffSBL. To this end, for any  $i^*$ , we can reorder the Kronecker product as

$$\mathbf{H} = \otimes_{i=1}^I \mathbf{H}_i = \mathbf{P}_{i^*} (\mathbf{H}_{i^*} \otimes \mathbf{S}_{i^*}) \mathbf{Q}_{i^*},$$

where  $\mathbf{S}_{i^*} = (\otimes_{i=i^*+1}^I \mathbf{H}_i) \otimes (\otimes_{i=1}^{i^*-1} \mathbf{H}_i)$  is independent of  $\mathbf{H}_{i^*}$  and  $\mathbf{P}_{i^*}$  and  $\mathbf{Q}_{i^*}$  are the corresponding permutation matrices [38]. Thus, the update step for  $\psi_{i^*,n^*}$  minimizes

$$\begin{aligned} f_{i^*,n^*}(\psi_{i^*,n^*}) &= \text{tr}\left(\left[\left(\mathbf{H}_{i^*}^H \mathbf{H}_{i^*}\right) \otimes \left(\mathbf{S}_{i^*}^H \mathbf{S}_{i^*}\right)\right] \mathbf{Q}_{i^*} \Sigma \mathbf{Q}_{i^*}^H\right) \\ &\quad + \text{tr}([\mathbf{H}_{i^*} \otimes \mathbf{S}_{i^*}] \mathbf{Q}_{i^*} \mathbf{M} \mathbf{P}_{i^*}) \\ &= 2\text{Re}\{\mathbf{v}_{i^*,n^*}^H \mathbf{h}_{i^*,n^*}\} + c_{i^*,n^*} \|\mathbf{h}_{i^*,n^*}\|_2^2 + \rho_{i^*,n^*}, \end{aligned}$$

where  $\mathbf{v}_{i^*,n^*}$ ,  $c_{i^*,n^*}$  and  $\rho_{i^*,n^*}$  are independent of  $\psi_{i^*,n^*}$ . Thus,  $f_{i^*,n^*}(\psi_{i^*,n^*})$  can be efficiently minimized with respect to  $\psi_{i^*,n^*}$  using a grid search. Also, alternating minimization sequentially updates  $\psi_{i^*,n^*}$  for different values of  $i^*$  and  $n^*$ , unlike dSBL where parallel updates are possible. Consequently, OffSBL incurs a higher computational cost here.

#### IV. APPLICATION: CHANNEL ESTIMATION FOR IRS-AIDED MIMO SYSTEM

In this section, we explore the application of our algorithm to cascaded channel estimation in an IRS-assisted MIMO system. We consider an uplink MIMO millimeter-wave system with a transmitter MS with  $T$  antennas, a receiver BS with  $R$  antennas, and a uniform linear array IRS with  $L$  elements. Let  $\Phi_{\text{MS}} \in \mathbb{C}^{L \times T}$  and  $\Phi_{\text{BS}} \in \mathbb{C}^{R \times L}$  denote the geometric narrowband MS-IRS and IRS-BS channel, respectively,

$$\Phi_{\text{MS}} = \sum_{p=1}^{P_{\text{MS}}} \sqrt{\frac{LT}{P_{\text{MS}}}} \beta_{\text{MS},p} \mathbf{a}_L(\phi_{\text{MS},p}) \mathbf{a}_T(\alpha_{\text{MS}})^H, \quad (23)$$

$$\Phi_{\text{BS}} = \sum_{p=1}^{P_{\text{BS}}} \sqrt{\frac{RL}{P_{\text{BS}}}} \beta_{\text{BS},p} \mathbf{a}_R(\alpha_{\text{BS},p}) \mathbf{a}_L(\phi_{\text{BS}})^H, \quad (24)$$

where we define the steering vector  $\mathbf{a}_Q(\psi) \in \mathbb{C}^Q$  for an integer  $Q$  and angle  $\psi$  as

$$\mathbf{a}_Q(\psi) = 1/\sqrt{Q} [1, e^{j\frac{2\pi\Delta}{\eta} \cos \psi}, \dots, e^{j\frac{2\pi\Delta}{\eta} (Q-1) \cos \psi}]^T$$

Here,  $\Delta$  is the distance between two adjacent elements, and  $\eta$  is the wavelength. We denote the number of rays in the scatter as  $P_{\text{MS}}$  and  $P_{\text{BS}}$ . The angles  $\phi_{\text{MS},p}$ ,  $\alpha_{\text{MS}}$ ,  $\alpha_{\text{BS},p}$ , and  $\phi_{\text{BS}}$  denote the  $p$ th AoA of the IRS, and the AoD of the MS, the  $p$ th AoA of the BS, and the AoD of the IRS, respectively (see [8, Fig. 1]). Then, the cascaded MS-IRS-BS channel can be expressed as  $\Phi_{\text{BS}} \text{diag}(\omega) \Phi_{\text{MS}}$  for any IRS configuration  $\omega \in \mathbb{C}^L$ . The  $l$ th entry of  $\omega$  represents the gain and phase shift due to the  $l$ th IRS element. Our goal is to estimate the cascaded channel  $\Phi_{\text{BS}} \text{diag}(\omega) \Phi_{\text{MS}}$ , which is a function of angles  $\phi_{\text{MS},p}$ ,  $\alpha_{\text{MS}}$ ,  $\alpha_{\text{BS},p}$ , and  $\phi_{\text{BS}}$ , for a given  $\omega$ .

We estimate the channel using pilot data transmitted over  $K$  time slots. We choose  $K_1$  IRS configurations, and for each configuration, transmit pilot data  $\mathbf{G} \in \mathbb{C}^{T \times K_P}$  over  $K_P$  time slots, where  $K = K_1 K_P$ . For the  $k$ th configuration  $\omega_k$ , the received signal  $\mathbf{Y}_k = \Phi_{\text{BS}} \text{diag}(\omega_k) \Phi_{\text{MS}} \mathbf{G} + \mathbf{N}_k \in \mathbb{C}^{R \times K_P}$  where  $\mathbf{N}_k \in \mathbb{C}^{R \times K_P}$  is the noise. Using (23) and (24), we get

$$\begin{aligned} \mathbf{Y}_k &= \zeta \sqrt{L} \mathbf{A}_{R,\text{BS}} \beta_{\text{BS}} \mathbf{a}_L(\phi_{\text{BS}})^H \text{diag}(\omega_k) \\ &\quad \times \mathbf{A}_{L,\text{MS}} \beta_{\text{MS}} \mathbf{a}_T(\alpha_{\text{MS}})^H \mathbf{G} + \mathbf{N}_k, \end{aligned}$$

where  $\mathbf{A}_{R,\text{BS}} \in \mathbb{C}^{R \times P_{\text{BS}}}$  and  $\beta_{\text{BS}} \in \mathbb{C}^{P_{\text{BS}}}$  have  $\mathbf{a}_R(\alpha_{\text{BS},p})$  and  $\beta_{\text{BS},p}$  as their  $p$ th column and entry, respectively. Similarly,  $\mathbf{A}_{L,\text{MS}} \in \mathbb{C}^{R \times P_{\text{MS}}}$  and  $\beta_{\text{MS}} \in \mathbb{C}^{P_{\text{MS}}}$  have  $\mathbf{a}_L(\alpha_{\text{MS},p})$  and  $\beta_{\text{MS},p}$  as their  $p$ th column and entry, respectively. Also,  $\zeta := \sqrt{\frac{LRT}{P_{\text{MS}}P_{\text{BS}}}}$ . Vectorizing  $\mathbf{Y}_k$  and using the properties of the Khatri-Rao product [39, Lemma A1] leads to (see [8] for detailed algebraic simplifications)

$$\begin{aligned} \mathbf{y}_k &= \zeta \sqrt{L} (\mathbf{A}_{L,\text{MS}} \beta_{\text{MS}} \mathbf{a}_T(\alpha_{\text{MS}})^H \mathbf{G})^\top \\ &\quad \odot (\mathbf{A}_{R,\text{BS}} \beta_{\text{BS}} \mathbf{a}_L(\phi_{\text{BS}})^H) \omega_k + \mathbf{n}_k \\ &= [\omega_k^\top \mathbf{A}_{L,\text{I}} \beta_{\text{MS}}] [\zeta \mathbf{G}^\top \mathbf{a}_T^*(\alpha_{\text{MS}})] \otimes [\mathbf{A}_{R,\text{BS}} \beta_{\text{BS}}] + \mathbf{n}_k, \end{aligned}$$

where  $\mathbf{A}_{L,\text{I}} \in \mathbb{C}^{L \times P_{\text{MS}}}$  whose  $p$ th column is  $\mathbf{a}_L(\phi_{\text{MS},p} - \phi_{\text{BS}})$ ,  $\odot$  is the Khatri-Rao product, and  $(\cdot)^*$  is conjugate. Here, the channel is given by

$$\begin{aligned} \text{vec}(\Phi_{\text{BS}} \text{diag}(\omega) \Phi_{\text{MS}}) \\ = [\omega^\top \mathbf{A}_{L,\text{I}} \beta_{\text{MS}} \zeta \mathbf{a}_T^*(\alpha_{\text{MS}})] \otimes [\mathbf{A}_{R,\text{BS}} \beta_{\text{BS}}]. \end{aligned} \quad (25)$$

Combining the received signals obtained for the  $K_1$  configurations, we obtain  $\bar{\mathbf{y}} \in \mathbb{C}^{RK}$

$$\bar{\mathbf{y}} = [\Omega^\top \mathbf{A}_{L,\text{I}} \beta_{\text{MS}}] \otimes [\zeta \mathbf{G}^\top \mathbf{a}_T^*(\alpha_{\text{MS}})] \otimes [\mathbf{A}_{R,\text{BS}} \beta_{\text{BS}}] + \bar{\mathbf{n}}, \quad (26)$$

where  $\Omega \in \mathbb{C}^{L \times K_1}$  with the  $k$ th column as  $\omega_k$ . Therefore, from (25), the channel estimation problem reduces to recovering  $\mathbf{A}_{L,\text{I}} \beta_{\text{MS}}$ ,  $\mathbf{a}_T(\alpha_{\text{MS}})$ , and  $\mathbf{A}_{R,\text{BS}} \beta_{\text{BS}}$  from  $\bar{\mathbf{y}}$  up to a scaling factor, given that  $\mathbf{G}$  and  $\Omega$  are known.

Comparing (26) with (5), we see the signal model here is the Kronecker product of three terms, i.e.,  $I = 3$ . The unknown parameters are AoA or AoD given by  $\bar{\psi}_1 \in \mathbb{R}^{P_{\text{MS}}}$ ,  $\bar{\psi}_2 = \alpha_{\text{MS}} \in \mathbb{R}$ , and  $\bar{\psi}_3 \in \mathbb{R}^{P_{\text{BS}}}$  where the  $p$ th entry of  $\bar{\psi}_1$  and  $\bar{\psi}_3$  are  $\phi_{\text{MS},p} - \phi_{\text{BS}}$  and  $\alpha_{\text{BS},p}$ , respectively. Also,  $\psi_{i,1} = 0$  and  $\psi_{i,r} = \pi$  for all values of  $i$ . The basis functions are related to the steering vectors as

$$\mathbf{h}_1(\psi) = \Omega^\top \mathbf{a}_L(\psi), \quad \mathbf{h}_2(\psi) = \mathbf{G}^\top \mathbf{a}_T(\psi), \quad \mathbf{h}_3(\psi) = \mathbf{a}_R(\psi).$$

Correspondingly, we have  $\bar{x}_1 = \beta_{\text{MS}}$ ,  $\bar{x}_2 = \zeta$ , and  $\bar{x}_3 = \beta_{\text{BS}}$ . The channel estimation problem is now translated into estimating  $\{\bar{\psi}_i\}_{i=1}^3$  and  $\{\bar{x}_i\}_{i=1}^3$  from the noisy measurement  $\bar{y}$ , where our dSBL (Algorithm 1) can be applied.

Using the decomposition step (Line 2) of Algorithm 1, we first decompose  $\bar{y}$  into three vectors,  $\hat{y}_1 \in \mathbb{C}^{K_1}$ ,  $\hat{y}_2 \in \mathbb{C}^{K_P}$ , and  $\hat{y}_3 \in \mathbb{C}^R$ , corresponding to the three terms in the Kronecker product in (26). We then use the dictionary for the steering vectors for a given integer  $Q$  as

$$\mathbf{A}_Q(\psi) = [\mathbf{a}_Q(\psi_1) \quad \mathbf{a}_Q(\psi_2) \quad \dots \quad \mathbf{a}_Q(\psi_N)] \in \mathbb{C}^{Q \times N}, \quad (27)$$

where  $\psi$  captures the unknown angles. This formulation leads to the following three problems similar to (11),

$$\hat{y}_i = \mathbf{H}_i(\psi_{(i)})\mathbf{x}_i + \mathbf{n}_i, \quad i = 1, 2, 3, \quad (28)$$

where  $\mathbf{H}_1(\psi_{(1)}) = \Omega^\top \mathbf{A}_L(\psi_{(1)})$ ,  $\mathbf{H}_2(\psi_{(2)}) = \mathbf{G}^\top \mathbf{A}_T(\psi_{(2)})$ , and  $\mathbf{H}_3(\psi_{(3)}) = \mathbf{A}_R(\psi_{(3)})$ . We solve them via OffSBL, initializing  $\psi_{(i)}$  by sampling the angular domain using  $N$  grid angles  $\{\theta_n\}_{n=1}^N$  such that  $\cos(\theta_n) = 2(n-1)/N - 1$ . OffSBL provides estimates  $(\hat{\psi}_{(i)}, \hat{\mathbf{x}}_i)$  of  $(\psi_{(i)}, \bar{\mathbf{x}}_i)$ , for  $i = 1, 2, 3$ . Finally, using (25) we compute the channel estimate for a given configuration  $\omega$  as

$$[\omega^\top \mathbf{A}_L(\hat{\psi}_{(1)})\hat{\mathbf{x}}_1 \mathbf{A}_T(\hat{\psi}_{(2)})\hat{\mathbf{x}}_2] \otimes [\mathbf{A}_R(\hat{\psi}_{(3)})\hat{\mathbf{x}}_3].$$

Here, the channel estimate is not affected by scaling ambiguity.

We reiterate that OffSBL is a standalone off-grid algorithm applicable to various linear inversion problems (i.e.,  $I = 1$ ). One notable example is direction-of-arrival estimation, where  $K$  far-field narrowband signals impinge on a uniform array with  $L$  elements ( $K < L$ ), resulting in the received signal

$$\hat{y} = \mathbf{A}_L(\psi)\mathbf{x} + \mathbf{n},$$

where the BEM dictionary  $\mathbf{A}_L(\psi)$  defined in (27) has steering vectors as its columns,  $\psi$  captures the AoAs, and  $\mathbf{x}$  is the sparse vector whose support corresponds to the true AoAs. This formulation is similar to  $i = 3$  in (28) for the IRS-aided channel estimation problem. In several other applications, the BEM dictionary takes the form  $\mathbf{B}\mathbf{A}_L(\psi)$ . Specifically, for  $i = 1$  and  $i = 2$  in (28),  $\mathbf{B}$  corresponds to  $\Omega^\top$  and  $\mathbf{G}^\top$ , respectively. In other cases,  $\mathbf{B}$  can represent beamformers [5], [6], [40], combiners [5], IRS configurations [41], pilot data [6], or random matrices [42].

## V. PERFORMANCE EVALUATION

We present numerical results to compare our algorithm with the state-of-the-art methods<sup>2</sup>. We present three sets of results. The first two demonstrate the decomposition step and OffSBL for parameter estimation. The third shows the combined results for IRS-aided channel estimation.

<sup>2</sup>Our code is found at <https://github.com/YanbinHe/JournalDecomOffGrid>. HOSVD is implemented using Tensorlab [43].

### A. Decomposition-Based Sparse Vector Recovery

In this section, we highlight the advantages of the decomposition step in reducing computational complexity and enhancing the denoising effect. We focus on recovering the Kronecker-structured sparse vector (4) using a multidimensional BEM (3) in the on-grid setting, without requiring the OffSBL algorithm. By combining the decomposition step with SBL, we demonstrate the benefits of this approach. We compare our method's performance with other methods that do not use decomposition, such as classical SBL [44], classical OMP, AM-KroSBL, and SVD-KroSBL [8]. Specifically, AM- and SVD-KroSBL only consider the Kronecker-structured support of the sparse vector and do not exploit the Kronecker structure in the nonzero entries as in (4).

We set  $I = 3$  with  $M_i = M$ , and  $N_i = 12$  for  $i = 1, 2, 3$  in (3) and (4). So, we have  $\mathbf{H} = \otimes_{i=1}^3 \mathbf{H}_i$  and  $\mathbf{x} = \otimes_{i=1}^3 \mathbf{x}_i$  with  $\mathbf{x}_i \in \mathbb{R}^{12}$ . The columns of  $\mathbf{H}_i \in \mathbb{C}^{M \times 12}$  for  $i = 1, 2, 3$  are the steering vectors evaluated by the grids  $\{\theta_n\}_{n=1}^{12}$  defined in Sec. IV. There are four nonzeros in each  $\mathbf{x}_i$ , whose positions are uniformly chosen from the grids and amplitudes are uniformly drawn from  $[0.5, 1.5]$ . Here, measurement level  $M$  is set to be  $\{8, 9, 10\}$ , labeled as Low, Medium, and High measurement case, controlling the number of measurements  $\bar{M} = M^3$  and the undersampling ratio  $M^3/N^3$ . We adopt the additive white Gaussian noise with zero mean whose variance is determined by SNR (dB) =  $10 \log_{10} \mathbb{E}\{\|\mathbf{H}\mathbf{x}\|_2^2/\|\mathbf{n}\|_2^2\}$  of  $\{5, 10, 15, 20, 25, 30\}$ . Three metrics are considered for performance evaluation: normalized mean squared error (NMSE), support recovery rate (SRR), and run time. Here, we define

$$\text{NMSE} = \mathbb{E} \left\{ \frac{\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2}{\|\mathbf{x}\|_2^2} \right\}, \quad \text{SRR} = \frac{|\text{supp}(\hat{\mathbf{x}}) \cap \text{supp}(\mathbf{x})|}{|\text{supp}(\hat{\mathbf{x}}) \cup \text{supp}(\mathbf{x})|},$$

where  $\mathbf{x}$  is the ground truth and  $\hat{\mathbf{x}}$  is the estimated vector. We limit the number of iterations for the SBL-based methods (dSBL, cSBL, AM-KroSBL, and SVD-KroSBL) to 200 and prune small entries in hyperparameters for faster convergence.

The denoising effect of the decomposition step is shown in Table I. Here, we compare the noise levels of *a*) the original noisy signal  $\mathbf{y}$  in (3), *b*) the signal after decomposition  $\hat{\mathbf{y}}_h := \otimes_{i=1}^3 \hat{\mathbf{y}}_i$  where  $\hat{\mathbf{y}}_i$ 's are obtained through (7) or (8), *c*) the signal after decomposition  $\hat{\mathbf{y}}_1 := \otimes_{i=1}^3 \hat{\mathbf{y}}_i$  where  $\hat{\mathbf{y}}_i$ 's are obtained through (9), and *d*) the result of (19). It can be seen that the noise level for both low complexity approximation and HOSVD is significantly reduced after decomposition. Low complexity approximation can reduce even more noise. It also closely matches the result in (19), validating our claim on the denoising effect discussed in Sec. III-A.

Fig. 1 shows that with higher SNR and more measurements, all algorithms yield better NMSE and SRR, as expected. Our dSBL algorithm outperforms other methods in NMSE and has the best SRR performance in most cases, demonstrating the efficacy of the decomposition idea. In contrast to the SVD-KroSBL algorithm that uses Kronecker-structured support, dSBL achieves superior NMSE by using the additional Kronecker structure in nonzero entries explicitly enforced via

TABLE I

ILLUSTRATION OF DENOISING WITH  $M = 10$ , USING THE NOISY SIGNAL  $\mathbf{y}$ , RECONSTRUCTED SIGNAL  $\hat{\mathbf{y}}_h = \otimes_{i=1}^3 \hat{\mathbf{y}}_i$  USING HOSVD, RECONSTRUCTED SIGNAL  $\hat{\mathbf{y}}_l = \otimes_{i=1}^3 \hat{\mathbf{y}}_i$  USING LOW COMPLEXITY APPROXIMATION, AND GROUND TRUTH  $\mathbf{y}_0$

| SNR   | 5 dB    | 10 dB  | 15 dB  | 20 dB  | 25 dB  | 30 dB  |
|---|---------|--------|--------|--------|--------|--------|
| Noise level                                 |         |        |        |        |        |        |
| $\ \mathbf{y} - \mathbf{y}_0\ _2^2$         | 34.6609 | 9.5529 | 2.8528 | 0.9720 | 0.3044 | 0.0895 |
| $\ \hat{\mathbf{y}}_h - \mathbf{y}_0\ _2^2$ | 0.9910  | 0.2741 | 0.0798 | 0.0267 | 0.0087 | 0.0023 |
| $\ \hat{\mathbf{y}}_l - \mathbf{y}_0\ _2^2$ | 0.9798  | 0.2720 | 0.0796 | 0.0267 | 0.0087 | 0.0023 |
| From (19)                                   | 0.9286  | 0.2580 | 0.0774 | 0.0263 | 0.0082 | 0.0024 |

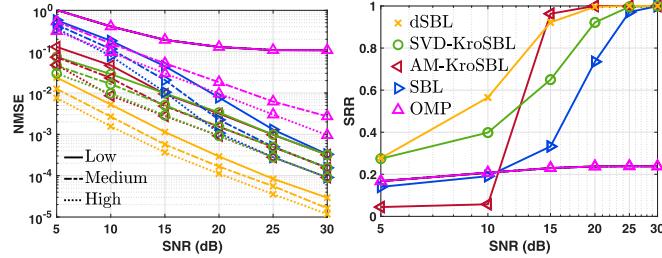


Fig. 1. NMSE and SRR of different algorithms as functions of SNR. For the SRR curves, we choose the low measurement case to avoid clutter.

(4) through the decomposition step. The relatively lower performance of AM-KroSBL is attributed to its slow convergence, given that we fix the number of EM iterations, as pointed out in [7]. The lower SRR and NMSE observed in the low SNR regime are due to small nonzero values in the estimate at locations where the ground truth is zero. We only include the low measurement case in the SRR result in Fig. 1, since all three measurement regimes exhibit similar trends; the others were omitted for clarity and better illustration.

Finally, Table II demonstrates that dSBL requires two-order less run time than the other competing algorithms, corroborating the computational advantage of our decomposition.

### B. Off-Grid Parameters Estimation

In this section, we apply OffSBL to the unknown parameters estimation problem. The model we consider here is the case in (28) with  $i = 1$ , where the goal is to estimate angles  $\psi_{(i)}$  and coefficients  $\mathbf{x}_i$ . The column function is  $\mathbf{h}_1(\psi) = \Omega^\top \mathbf{a}_L(\psi)$  with  $\Omega \in \mathbb{C}^{L \times M}$ ,  $L = 256$ , and  $M$  being the number of measurements. Here,  $M$  is  $\{20, 25, 30, 35, 40, 45, 50\}$  and controls the undersampling ratio defined as  $\frac{M}{N}$  with  $N = 180$ . The matrix  $\Omega$  is randomly generated, whose entries take the form  $e^{j\phi}$  where  $\phi$  is drawn from a uniform distribution on  $[0, \pi]$ . We set the number of unknown parameters (angles)  $S_1$  to be  $\{2, 4, 6\}$ , and the angles are drawn sequentially from a uniform distribution on  $[-0.9, 0.9]$  ensuring a minimal separation of 0.1. The coefficients are drawn from  $\mathcal{CN}(0, 2)$ .

We use three benchmarks: *i*) classical (on-grid) SBL, *ii*) off-grid sparse Bayesian inference (OGSBI) using the first-order Taylor expansion [14], and *iii*) light-weight sequential SBL (LWSSBL), a state-of-the-art off-grid method using marginal likelihood optimization [21]. In our simulations, we do not

TABLE II  
RUNTIME OF DIFFERENT SCHEMES IN SECONDS

| SNR        | 5 dB  | 10 dB | 15 dB | 20 dB | 25 dB | 30 dB |
|------------|-------|-------|-------|-------|-------|-------|
| OMP        | 0.599 | 0.602 | 0.605 | 0.603 | 0.604 | 0.603 |
| cSBL       | 8.961 | 7.470 | 6.111 | 5.552 | 5.397 | 5.318 |
| AM-KroSBL  | 8.528 | 8.516 | 7.249 | 5.424 | 4.520 | 4.093 |
| SVD-KroSBL | 4.534 | 3.360 | 2.840 | 2.668 | 2.627 | 2.608 |
| dSBL       | 0.009 | 0.005 | 0.004 | 0.004 | 0.004 | 0.004 |

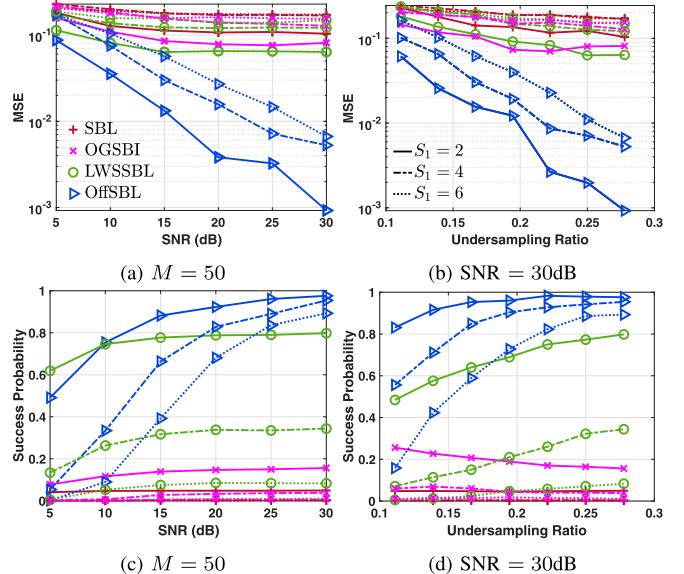


Fig. 2. MSE And success probability as a function of SNR and undersampling ratio  $\frac{M}{N}$  for unknown parameter estimation with a varying number of unknown parameters  $S_1$  and  $N = 180$ .

provide the number of unknowns  $S_1$  to all algorithms, but only an upper bound  $\bar{S}$  of the number of unknowns. In practice, we only solve the problem (17) for the variables corresponding to  $\bar{S}$  largest peaks of the hyperparameter  $\gamma$  instead of all variables. OffSBL estimates the noise variance using (15). SBL and OGSBI can also estimate the noise variance, while noise variance estimation for LWSSBL is not discussed [21]. So for LWSSBL, we set noise variance estimate as  $0.1\|\bar{\mathbf{y}}\|_2^2/M$  as in [21]. We choose  $\text{SNR} = 10 \log_{10} \mathbb{E}\{\|\bar{\mathbf{H}}_{\bar{\psi}} \bar{\mathbf{x}}\|_2^2/\|\bar{\mathbf{n}}\|_2^2\}$  as  $\{5, 10, 15, 20, 25, 30\}$  in dB. The performance metrics are mean squared error (MSE) and success probability, where

$$\text{MSE} = \mathbb{E} \left\{ \frac{1}{S_1} \sum_{s=1}^{S_1} (\bar{\psi}_s - \hat{\psi}_s)^2 \right\},$$

with expectation taken over  $10^3$  independent trials. Here,  $\bar{\psi}_s$  and  $\hat{\psi}_s$  denote the true value and the estimation, respectively. The success probability is defined as the fraction of trials with MSE smaller than  $10^{-6}$ .

We compare MSE and recovery probability for different SNRs and undersampling ratios in Fig. 2. We see that higher SNR and more measurements facilitate all algorithms, except OGSBI in Fig. 2(b) and 2(d). This is because OGSBI cannot effectively optimize the grid points in this setting, as we show later in Fig. 3. Among all candidates, our OffSBL has the

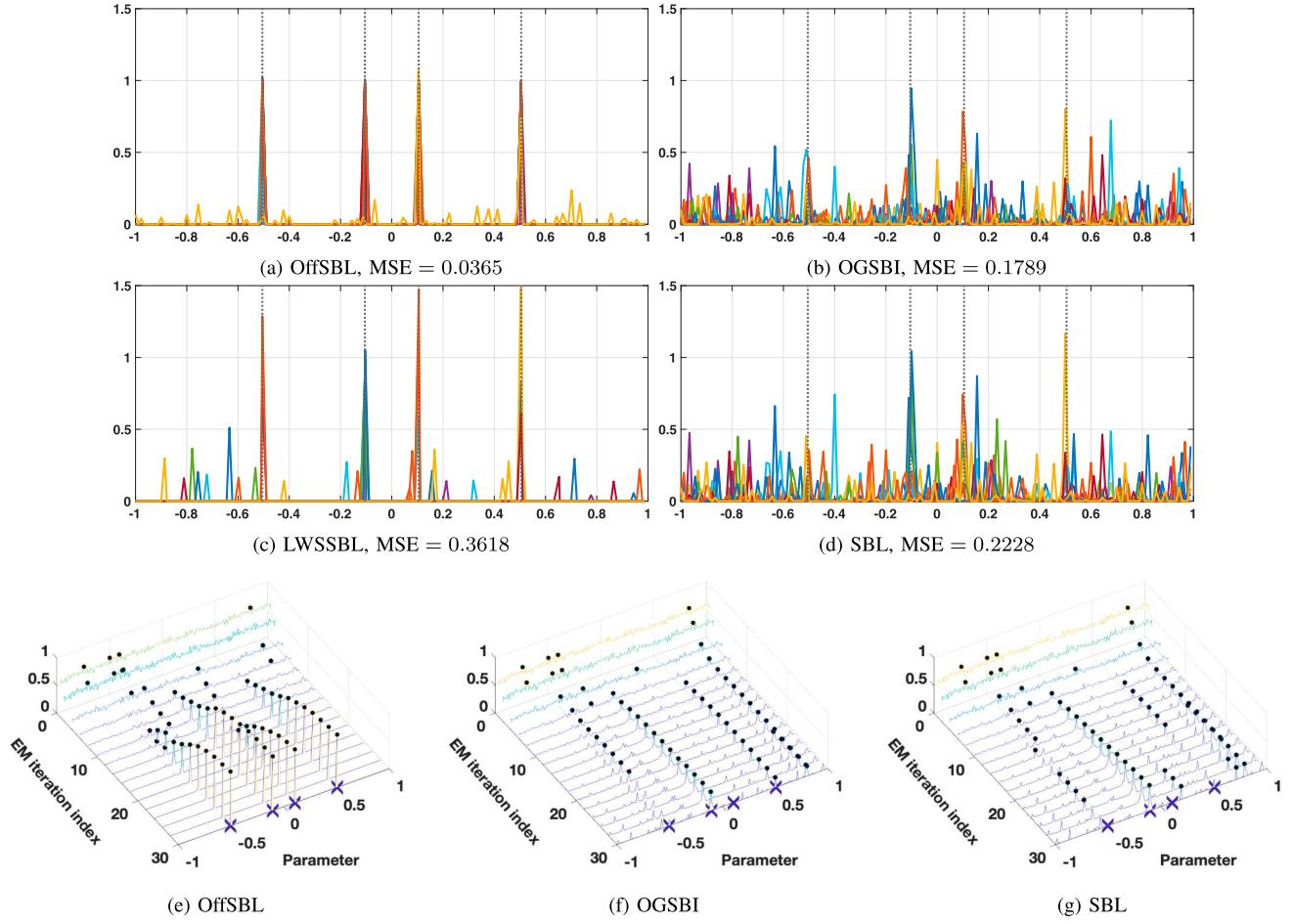


Fig. 3. Pseudospectrum  $\gamma$  of different algorithms in the worst-case scenario analysis. (a)–(d): pseudospectrum  $\gamma^{(1000)}$  of the 1000th EM iteration. Different colors mean different independent realizations of gaussian noise. (e)–(g): dynamic evolution of pseudospectrum  $\{\gamma^{(r)}\}_{1 \leq r \leq 30}$  for the EM iterations of EM-based SBL algorithms. Black dot: grid points chosen to be updated. Cross: true values. Although SBL operates as an on-grid method, we still include the grid points that correspond to the top four peaks, solely for comparison purposes.

best performance in both MSE and recovery probability in most cases. An exception is SNR = 5dB where LWSSBL has a higher success probability. However, LWSSBL often produces larger errors when it fails, making OffSBL superior in MSE.

In Fig. 3, we present a worst-case scenario study. We set  $M = 60$  measurements and SNR = 30dB. The unknowns are  $[-0.5050, -0.1050, 0.1050, 0.5050]$ , shown as vertical dashed lines in Fig. 3(a)–3(d). These values are intentionally selected to be midway between two grids to create a challenging case for grid optimization. All the coefficients are set to one. We provide the number of unknowns to all the algorithms but not the noise variance. We perform  $10^3$  EM iterations to facilitate the convergence of all algorithms. The input of all algorithms is the same noiseless signal but with ten independent Gaussian noise realizations. We plot the final pseudospectrum (hyperparameter  $\gamma^{(1000)}$ ) after  $10^3$  EM iterations for ten noise realizations in Fig. 3(a)–3(d) with different colors.

Comparing the different algorithms, we note that OffSBL consistently recovers all parameters, with minimal amplitude spikes appearing in the pseudospectrum corresponding to parameters other than the true values. LWSSBL also recovers the unknowns but with a lower success probability. LWSSBL

TABLE III  
RUNTIME IN SECOND FOR UNKNOWN PARAMETER ESTIMATION WITH  
 $S_1 = 2$  AND SNR = 30 DB

| $M$    | 20    | 25    | 30    | 35    | 40    | 45    | 50    |
|--------|-------|-------|-------|-------|-------|-------|-------|
| OffSBL | 0.602 | 0.546 | 0.539 | 0.541 | 0.477 | 0.504 | 0.522 |
| SBL    | 0.131 | 0.137 | 0.161 | 0.170 | 0.171 | 0.181 | 0.191 |
| LWSSBL | 0.032 | 0.033 | 0.035 | 0.037 | 0.038 | 0.041 | 0.043 |
| OGSBI  | 0.250 | 0.269 | 0.274 | 0.286 | 0.290 | 0.305 | 0.315 |

exhibits more peaks at parameters other than the true values, implying that it is more prone to being misled by incorrect columns in the dictionary due to its greedy nature. In contrast, while OffSBL takes longer to reach the final result (see Table III), evaluating all columns rather than proceeding greedily reduces the risk of being misled by incorrect columns.

Further, there is little difference between OGSBI and the on-grid benchmark SBL, indicating that the first-order approximation is less effective in this case. However, OGSBI has some improvement over SBL as reflected by a lower MSE. These findings also highlight that algorithms relying on on-grid SBL for rough estimates and then refining peaks are likely to fail, as on-grid SBL often doesn't provide a reliable starting point, with peaks rarely matching the true parameters. This is likely due

to the dictionary's structure, which takes the form  $\Omega^T A_L(\psi)$  for some integer  $L$ . When  $\Omega^T \in \mathbb{C}^{M \times L}$  has fewer rows than columns ( $M < L$ ), the compression effect from multiplication by  $\Omega^T$  can lead to information loss, creating a challenging setting for off-grid sparse recovery [45]. However, in many applications, such as IRS channel estimation, where the value of  $M$  represents the number of time slots,  $M$  is typically limited. Thus, integrating grid updates into the EM iteration, as implemented in OffSBL, is essential.

Fig. 3(e)–3(g) present the pseudospectrum  $\{\gamma^{(r)}\}_{1 \leq r \leq 30}$  for OffSBL, OGSBI, and SBL, along with the grid points that are updated dynamically throughout the EM iteration. Although SBL is an on-grid method, we pinpoint the grids of the top four peaks. All algorithms start from the same  $\gamma$ . Our OffSBL demonstrates superior optimization of grid points, identifying the correct values and amplitudes, whereas SBL and OGSBI do not reveal the true parameters. The pseudospectrum highlights the effectiveness of our grid adjustment.

### C. IRS-Aided Wireless Channel Estimation

We focus on the IRS-aided channel estimation problem, as described in Sec. IV. Here, we first use the decomposition step and then turn to the BEM and apply OffSBL separately for  $i = 1, 2, 3$  in (28). Thus, the channel estimation scheme can be viewed as a collective evaluation of the decomposition and OffSBL. For benchmarking, we apply the same decomposition step, and then solve (28) using the same algorithms as in Sec. V-B. For simplicity, we denote the problem (28) with  $i = 1, 2, 3$  as P1, P2, and P3, respectively.

For IRS-aided channel estimation, we use  $R = 16$  BS antennas,  $T = 6$  MS antennas,  $L = 256$  IRS elements. We consider only one path between the BS and IRS [46], [47], [48], [49], as the IRS is typically mounted in locations with fewer obstacles [50], [51], and the line-of-sight path is generally much stronger than the other paths. Therefore, we take  $P_{\text{BS}} = 1$  and  $P_{\text{MS}} = 3$ . The IRS configuration entries  $\{\omega_k\}_{k=1}^{K_1}$  are  $1/\sqrt{L}e^{j\phi}$  where  $\phi$  is drawn uniformly randomly from  $[0, \pi]$  with  $K_1 = 40$ . We send  $K_P = 20$  pilot signals for each IRS configuration. Our OffSBL algorithm uses the dictionaries in P1, P2, and P3 with  $N_1 = 180$ ,  $N_2 = 50$ , and  $N_3 = 50$ , respectively. The other algorithms set  $N_1 = 180$ ,  $N_2 = 150$ , and  $N_3 = 150$  grid points. The channel gains  $\beta_{\text{BS}}$  and  $\{\beta_{\text{MS},p}\}_{p=1}^{P_{\text{MS}}}$  in (23) and (24) are drawn from standard complex Gaussian [52]. We randomly draw  $\alpha_{\text{MS}}$ ,  $\{\phi_{\text{MS},p}\}_{p=1}^{P_{\text{MS}}}$ ,  $\phi_{\text{BS}}$ , and  $\alpha_{\text{BS}}$  from uniform distribution in  $[0.3, 0.5]$ ,  $[-0.2, 0.2]$ ,  $[0.3, 0.5]$ , and  $[0, 0.5]$ , respectively. We also assume that the angles are separated by at least 0.07. We opt for SNR  $\{-5, 0, 5, 10, 15, 20\}$  in dB. Also, additional channel estimation metrics are NMSE and symbol error rate (SER) over  $10^6$  16-QAM symbols decoded using the estimated channel, where

$$\frac{1}{K_1} \sum_{k=1}^{K_1} \frac{\|\mathbf{H}_{\text{BS}} \text{diag}(\omega_k) \mathbf{H}_{\text{MS}} - \tilde{\mathbf{H}}_{\text{BS}} \text{diag}(\omega_k) \tilde{\mathbf{H}}_{\text{MS}}\|_{\text{F}}^2}{\|\mathbf{H}_{\text{BS}} \text{diag}(\omega_k) \mathbf{H}_{\text{MS}}\|_{\text{F}}^2}$$

with  $\tilde{\mathbf{H}}_{\text{BS}} \text{diag}(\omega_k) \tilde{\mathbf{H}}_{\text{MS}}$  being the channel estimate.

We first examine the angle estimation results in Fig. 4. It can be seen that our OffSBL can achieve the best performance in

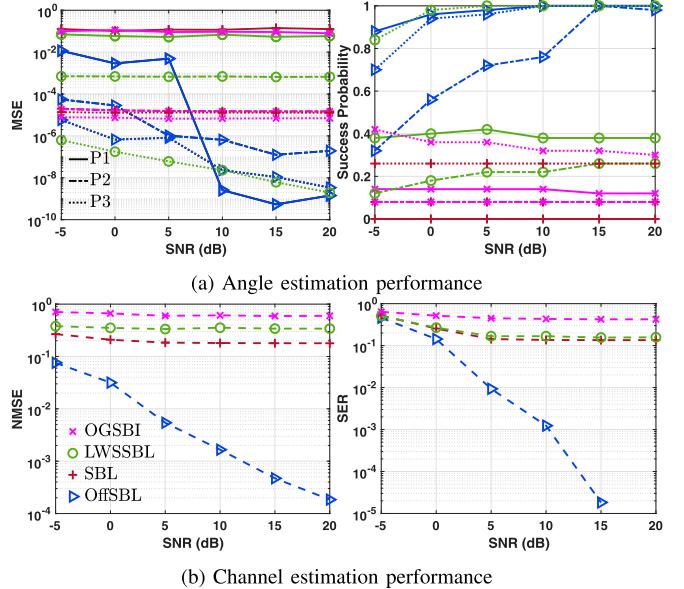


Fig. 4. IRS-Aided channel estimation. (a): MSE and recovery success probability for angle estimation in P1, P2, and P3. (b): NMSE of IRS-aided channel estimation and SER of different algorithms as functions of SNR.

solving P1 and P2 except for the low SNR case for P2. P3 reduces to the normal DoA estimation problem, where LWSSBL exhibits superior recovery ability. However, at higher SNRs, our algorithm is able to achieve comparable performance. As evident from the NMSE and SER plots, OffSBL consistently recovers the true angles and accurately retrieves the coefficients, leading to the best NMSE and SER. Although other algorithms can perform well in solving P2 and P3, the significant recovery errors in P1 affect the overall accuracy of channel retrieval.

## VI. CONCLUSION

We addressed the joint estimation of unknown parameters and coefficients from Kronecker-structured measurements. Leveraging the Kronecker structure, we decomposed the problem into smaller independent subproblems. Each subproblem was solved with EM-based SBL integrated with a novel grid optimization method to reduce grid mismatch. We provided a theoretical analysis of the error bound for the decomposition step and established the algorithm's convergence. Our decomposition step also reduces the noise level in the measurements, which was also analyzed theoretically. Numerical results showed that the decomposition step reduces complexity, while the grid optimization improves accuracy. Future work can analyze the resolution of OffSBL and extend it to the recovery of sparse tensors with ranks greater than one.

## APPENDIX A PROOF OF THEOREM 1

We define  $\mathbf{P}_U$  as the projection matrix onto the column space of a given matrix  $\mathbf{U}$  and  $\mathbf{U}_{\perp}$  as the projection onto its orthogonal subspace. Also,  $\|\cdot\|$  is the matrix spectral norm. We need the below lemma for the proof.

*Lemma 3* [53, Supplement Sec. 1.2]: Suppose  $\mathbf{X} \in \mathbb{R}^{p_1 \times p_2}$  is a rank- $r$  matrix and  $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$  where the entries of  $\mathbf{Z} \in \mathbb{R}^{p_1 \times p_2}$  follow a zero mean Gaussian distribution with unit variance. We denote  $\mathbf{V}, \hat{\mathbf{V}} \in \mathbb{R}^{p_2 \times r}$  as the matrix of the right singular vectors of  $\mathbf{X}$  and the matrix of the top  $r$  right singular vectors of  $\mathbf{Y}$ , respectively. Suppose the  $r$ th right singular value  $\sigma_r^2$  of  $\mathbf{X}$  satisfies  $\sigma_r^2 \geq C_{\text{gap}} \sigma^2(\sqrt{p_1 p_2} + p_2)$  for some large constant  $C_{\text{gap}} > 0$ . Then, for all  $x \geq 0$ , there exist constants  $C, c > 0$  such that

$$\begin{aligned} \mathbb{P}\{\|\mathbf{P}_{\mathbf{Y}\mathbf{V}}\mathbf{Y}\mathbf{V}_{\perp}\| \leq x\} &\geq 1 - C \exp\{-c(\sigma_r^2 + p_1)\} \\ &\quad - C \exp\left\{Cp_2 - c \min\left(x^2, x\sqrt{\sigma_r^2 + p_1}\right)\right\}, \end{aligned}$$

and with probability exceeding  $1 - C \exp\{-c\sigma_r^4/(\sigma_r^2 + p_1)\}$ ,

$$\|\sin \Theta(\hat{\mathbf{V}}, \mathbf{V})\|^2 \leq C(\sigma_r^2 + p_1)\sigma_r^{-4}\|\mathbf{P}_{\mathbf{Y}\mathbf{V}}\mathbf{Y}\mathbf{V}_{\perp}\|^2,$$

Here,  $\Theta(\hat{\mathbf{V}}, \mathbf{V}) = \text{diag}(\arccos(\sigma_1), \dots, \arccos(\sigma_r))$  where  $\sigma_1 \geq \dots \geq \sigma_r \geq 0$  are the singular values of  $\mathbf{V}^T \hat{\mathbf{V}}$ .

We prove Theorem 1 for  $i = 1$ , and  $i = 2, \dots, I$  follow similarly. Also, we consider the decomposition of  $\bar{\mathbf{y}}/\sigma_t$  instead of  $\bar{\mathbf{y}}$ . This scaling does not alter the subspaces obtained after decomposition but ensures that the noise entries follow a zero-mean, unit-variance Gaussian distribution, as in Lemma 3.

For  $i = 1$ , the true and estimated subspaces are spanned by  $\mathbf{y}_1/\|\mathbf{y}_1\|$  and  $\hat{\mathbf{y}}_1/\|\hat{\mathbf{y}}_1\|$ , respectively. The first mode matricization of the tensor  $\mathcal{Y}$ , as defined in (6), is  $\mathbf{Y}_{(1)}$ . Setting  $r = 1, p_1 = \bar{M}/M_1$  and  $p_2 = M_1$ , and consequently,  $\sigma_r = \lambda$  in Lemma 3, we derive

$$\begin{aligned} \sin^2 \vartheta_1 &= \|\sin \Theta(\hat{\mathbf{y}}_1/\|\hat{\mathbf{y}}_1\|, \mathbf{y}_1/\|\mathbf{y}_1\|)\| \\ &\leq C(\lambda^2 + \bar{M}/M_1)\lambda^{-4}\|\mathbf{P}_{\mathbf{Y}_{(1)}^T \mathbf{y}_1} \mathbf{Y}_{(1)}^T \mathbf{y}_{1\perp}\|^2, \end{aligned} \quad (29)$$

with probability at least  $1 - C e^{-c \frac{\lambda^4}{\lambda^2 + p_1}}$ .

Further, we bound  $\|\mathbf{P}_{\mathbf{Y}_{(1)}^T \mathbf{y}_1} \mathbf{Y}_{(1)}^T \mathbf{y}_{1\perp}\|^2$  using Lemma 3 by setting  $x = \sqrt{\bar{C}M_1}$  where  $\bar{C} < C/c < C_{\text{gap}}$ ,

$$\begin{aligned} \mathbb{P}\{\|\mathbf{P}_{\mathbf{Y}\mathbf{V}}\mathbf{Y}\mathbf{V}_{\perp}\| \leq \bar{C}\sqrt{M_1}\} &\geq 1 - C \exp\{-c(\sigma_r^2 + p_1)\} \\ &\quad - C \exp\left\{CM_1 - c \min\left(\bar{C}M_1, \sqrt{\bar{C}M_1(\sigma_r^2 + p_1)}\right)\right\}, \end{aligned} \quad (30)$$

Then, we simplify the right-hand side of (30) using

$$\sigma_r^2 + p_1 \geq \sigma_r^2 = \lambda^2 \geq C_{\text{gap}}(\sqrt{p_1 M_1} + M_1) \geq C_{\text{gap}} M_1. \quad (31)$$

Then, (30) is simplified as

$$\begin{aligned} \mathbb{P}\{\|\mathbf{P}_{\mathbf{Y}\mathbf{V}}\mathbf{Y}\mathbf{V}_{\perp}\| \leq \bar{C}\sqrt{M_1}\} &\geq 1 - C \exp\{-cC_{\text{gap}}M_1\} \\ &\quad - C \exp\left\{\left[C - c \min\left(\bar{C}, \sqrt{\bar{C}C_{\text{gap}}}\right)\right] M_1\right\} \geq 1 - C e^{-\bar{c}M_1}, \end{aligned}$$

for some constant  $\bar{c}$ . So, (29) and the union bound implies

$$\sin^2 \vartheta_1 \leq \frac{C(\lambda^2 + \bar{M}/M_1)}{\lambda^4} \bar{C}M_1 = \frac{C\bar{C}M_1}{\lambda^2} + \frac{C\bar{C}\bar{M}}{\lambda^4}, \quad (32)$$

with probability exceeding  $1 - C e^{-c \frac{\lambda^4}{\lambda^2 + p_1}} - C e^{-\bar{c}M_1}$ .

Furthermore, since  $\lambda^2 \geq C_{\text{gap}}(\sqrt{p_1 M_1} + M_1)$ , with  $\bar{C} = \min(C_{\text{gap}}, C_{\text{gap}}^2)$ , we derive

$$\lambda^2 \geq \sqrt{\bar{C}p_1 M_1} + \bar{C}M_1 \geq \frac{\bar{C}M_1 + \sqrt{\bar{C}^2 M_1^2 + 4\bar{C}p_1 M_1}}{2}.$$

Then,  $\lambda^4 - \bar{C}M_1\lambda^2 - \bar{C}p_1 M_1 \geq 0$  because  $\lambda^2$  is greater than both roots of the quadratic function in  $\lambda^2$ . Thus, we deduce  $\frac{\lambda^4}{\lambda^2 + p_1} \geq \bar{C}M_1$ . So, from (32), we arrive at the desired result,

$$\mathbb{P}\left\{\sin \vartheta_1 \leq \frac{\sqrt{C\bar{C}M_1}}{\lambda} + \frac{\sqrt{C\bar{C}\bar{M}}}{\lambda^2}\right\} \geq 1 - 2C e^{-\max(c\bar{C}, \bar{c})M_1}.$$

## APPENDIX B PROOF OF THEOREM 2

Our proof uses the following lemmas.

*Lemma 4* [31, Lemma 6]: Suppose  $\mathbf{X}, \mathbf{Z}$  are two matrices, and the projection matrix orthogonal to the subspace spanned by the leading  $r$  left singular vectors of  $\mathbf{X} + \mathbf{Z}$  is  $\mathbf{U}$ . Then,  $\|\mathbf{U}\mathbf{X}\|_{\text{F}} \leq 2\sqrt{r}\|\mathbf{Z}\|$ , where  $r$  is the rank of  $\mathbf{X}$ .

*Lemma 5* [54, Corollary 5.35]: Let  $\mathbf{Z} \in \mathbb{R}^{p_1 \times p_2}$  whose entries are independent Gaussian random variables with zero mean unit variance. Then, for any  $x \geq 0$ , the matrix satisfies  $\|\mathbf{Z}\| \geq \sqrt{p_1} + \sqrt{p_2} + x$ , with probability less than  $2e^{-x^2/2}$ .

*Lemma 6* [55, Lemma 8.1]: Suppose  $X$  satisfies the non-central  $\chi_d^2(\nu)$  distribution with  $d$  degrees of freedom and non-centrality parameter  $\nu$ . Then, for all  $x > 0$ , it satisfies  $X \geq (d + \nu) + 2\sqrt{(d + 2\nu)x} + 2x$  with probability less than  $e^{-x}$ .

*Lemma 7*: [30] Consider tensors  $\mathcal{Y}, \mathcal{X}$ , and  $\mathcal{Z}$ , such that  $\mathcal{Y} = \mathcal{X} \times_1 \mathbf{A}_1 \cdots \times_I \mathbf{A}_I$ , where  $\mathcal{X} = \mathcal{Z} \times_1 \mathbf{B}_1 \cdots \times_I \mathbf{B}_I$ , for any compatible matrices  $\{\mathbf{A}_i, \mathbf{B}_i\}_{i=1}^I$ . Then,  $\mathcal{Y} = \mathcal{Z} \times_1 (\mathbf{A}_1 \mathbf{B}_1) \cdots \times_I (\mathbf{A}_I \mathbf{B}_I)$ .

Setting the tensor order  $R = I$  and  $r$ -ranks as  $p_i = 1$  for  $i \in [I]$  in [32, Eq. (19)] leads to (19). To prove the probabilistic bound, we note that using tensor notation,  $\|\otimes_{i=1}^I \hat{\mathbf{y}}_i - \otimes_{i=1}^I \mathbf{y}_i\|_2 = \|\hat{\mathcal{Y}} - \mathcal{Y}\|_{\text{F}}$ , where  $\hat{\mathcal{Y}}$  is HOSVD output and  $\mathcal{Y} = \mathcal{Y} + \bar{\mathcal{N}}$ , with  $\mathcal{Y}, \mathcal{Y}$  and  $\bar{\mathcal{N}}$  being the measurement, the noiseless signal, and noise tensors, respectively, from (5).

HOSVD reconstructs  $\hat{\mathcal{Y}} = \xi \times_1 \mathbf{e}_1 \cdots \times_I \mathbf{e}_I$ . Here,  $\mathbf{e}_i$ , the leading left singular vectors of the  $i$ th mode matricization of  $\hat{\mathcal{Y}}$  given by  $\mathbf{Y}_{(i)} + \mathbf{N}_{(i)}$ , with  $\mathbf{Y}_{(i)}$  and  $\mathbf{N}_{(i)}$  are the  $i$ th mode matricization of  $\mathcal{Y}$  and  $\bar{\mathcal{N}}$ , respectively. Also,  $\xi = \bar{\mathcal{Y}} \times_1 \mathbf{e}_1^T \cdots \times_I \mathbf{e}_I^T$ , as the signal is real. Lemma 7 implies

$$\hat{\mathcal{Y}} = (\mathcal{Y} + \bar{\mathcal{N}}) \times_1 (\mathbf{e}_1 \mathbf{e}_1^T) \times_2 \cdots \times_I (\mathbf{e}_I \mathbf{e}_I^T) = \bar{\mathcal{N}}_{(I)} + \mathcal{Y}_{(I)},$$

where we define  $\bar{\mathcal{N}}_{(I)} = \bar{\mathcal{N}} \times_1 (\mathbf{e}_1 \mathbf{e}_1^T) \cdots \times_I (\mathbf{e}_I \mathbf{e}_I^T)$  and  $\mathcal{Y}_{(i)} = \mathcal{Y} \times_1 (\mathbf{e}_1 \mathbf{e}_1^T) \cdots \times_i (\mathbf{e}_i \mathbf{e}_i^T)$  with  $\mathcal{Y}_{(0)} = \mathcal{Y}$ . Therefore,

$$\begin{aligned} \|\otimes_{i=1}^I \hat{\mathbf{y}}_i - \otimes_{i=1}^I \mathbf{y}_i\|_2 &= \|\hat{\mathcal{Y}} - \mathcal{Y}\|_{\text{F}} \\ &\leq \|\mathcal{Y}_{(I)} - \mathcal{Y}\|_{\text{F}} + \|\bar{\mathcal{N}}_{(I)}\|_{\text{F}}. \end{aligned} \quad (33)$$

To bound the first term in (33), let  $\mathbf{P}_{i\perp}$  be the projection matrix orthogonal to  $\mathbf{e}_i$  so that  $\mathcal{Y} = \mathcal{Y} \times_1 (\mathbf{P}_{1\perp} + (\mathbf{e}_1 \mathbf{e}_1^T))$ , leading to

$$\begin{aligned} \mathcal{Y} &= \mathcal{Y}_{(0)} \times_1 \mathbf{P}_{1\perp} + \mathcal{Y}_{(1)} \\ &= \mathcal{Y}_{(0)} \times_1 \mathbf{P}_{1\perp} + \mathcal{Y}_{(1)} \times_2 \mathbf{P}_{2\perp} + \mathcal{Y}_{(2)} \end{aligned}$$

$$= \sum_{i=1}^I \mathcal{Y}_{(i-1)} \times_i \mathbf{P}_{i\perp} + \mathcal{Y}_{(I)}.$$

Therefore, using triangle inequality, we obtain

$$\begin{aligned} \|\mathcal{Y}_{(I)} - \mathcal{Y}\|_{\text{F}} &\leq \sum_{i=1}^I \|\mathcal{Y}_{(i-1)} \times_i \mathbf{P}_{i\perp}\|_{\text{F}} \leq \sum_{i=1}^I \|\mathcal{Y} \times_i \mathbf{P}_{i\perp}\|_{\text{F}} \\ &= \sum_{i=1}^I \|\mathbf{P}_{i\perp} \mathbf{Y}_{(i)}\|_{\text{F}} \leq \sum_{i=1}^I 2\|\mathbf{N}_{(i)}\|. \end{aligned} \quad (34)$$

The last step follows from Lemma 4, as  $\mathbf{P}_{i\perp}$  is the projection matrix orthogonal to  $\mathbf{e}_i$  and the rank of  $\mathbf{Y}_{(i)}$  is 1 from (6). Also, Lemma 5 with  $x = \sqrt{2M_i}$  and  $\mathbf{Z} = \sigma_t^{-1} \mathbf{N}_{(i)}$  implies that with probability at least  $1 - 2e^{-M_i}$

$$\begin{aligned} \|\sigma_t^{-1} \mathbf{N}_{(i)}\| &\leq \sqrt{M_i} + \sqrt{\bar{M}/M_i} + \sqrt{2M_i} \\ &\leq 3\sqrt{M_i} + \sqrt{\bar{M}/M_i}. \end{aligned}$$

From (34), with probability exceeding  $1 - 2 \sum_{i=1}^I e^{-M_i}$ ,

$$\|\mathcal{Y}_{(I)} - \mathcal{Y}\|_{\text{F}} \leq 2\sigma_t \sum_{i=1}^I \left[ 3\sqrt{M_i} + \sqrt{\bar{M}/M_i} \right]. \quad (35)$$

Next, we bound the second term in (33). We note that  $\|\sigma_t^{-1} \bar{\mathcal{N}}_{(I)}\|_{\text{F}}^2$  is an 1-dimensional projection of a zero mean unit variance Gaussian tensor and follows  $\chi_1^2(0)$  [56, Supplement Sec. C.4]. Lemma 6 states that for any  $x > 0$   $\sigma_t^{-1} \|\bar{\mathcal{N}}_{(I)}\|_{\text{F}} \leq \sqrt{1 + 2\sqrt{x} + 2x} \leq 1 + 2\sqrt{x}$  with probability exceeding  $1 - e^{-x}$ . Setting  $x = \max_{1 \leq i \leq I} M_i$  and combining with (33) and (35) using the union bound yields (18) with probability exceeding  $1 - 2 \sum_{i=1}^I e^{-M_i} - e^{-\max_{1 \leq i \leq I} M_i} \geq 1 - 3 \sum_{i=1}^I e^{-M_i}$ . Hence, the proof is complete.

## REFERENCES

- Y. He and G. Joseph, "Kronecker-structured sparse vector recovery with application to IRS-MIMO channel estimation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Apr. 2025, pp. 1–5.
- C. F. Caiafa and A. Cichocki, "Block sparse representations of tensors using Kronecker bases," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal, Process.*, Mar. 2012, pp. 2709–2712.
- C. F. Caiafa and A. Cichocki, "Computing sparse representations of multidimensional signals using Kronecker bases," *Neural Comput.*, vol. 25, no. 1, pp. 186–220, Jan. 2013.
- C. F. Caiafa and A. Cichocki, "Multidimensional compressed sensing and their applications," *Wiley Interdiscip. Rev.: Data Min. Knowl. Discov.*, vol. 3, no. 6, pp. 355–380, Oct. 2013.
- Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and R. S. Blum, "Low-rank tensor decomposition-aided channel estimation for millimeter wave MIMO-OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1524–1538, Jul. 2017.
- R. Wang, H. Ren, C. Pan, G. Zhou, and J. Wang, "Tensor decomposition-based time varying channel estimation for mmWave MIMO-OFDM systems," Aug. 2024, *arXiv:2403.02942*.
- Y. He and G. Joseph, "Bayesian algorithms for Kronecker-structured sparse vector recovery with application to IRS-MIMO channel estimation," *IEEE Trans. Signal Process.*, vol. 73, pp. 142–157, 2025.
- Y. He and G. Joseph, "Structure-aware sparse Bayesian learning-based channel estimation for intelligent reflecting surface-aided MIMO," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal, Process.*, Jun. 2023, pp. 1–5.
- G. Ortiz-Jiménez, M. Coutino, S. P. Chepuri, and G. Leus, "Sparse sampling for inverse problems with tensors," *IEEE Trans. Signal Process.*, vol. 67, no. 12, pp. 3272–3286, Jun. 2019.
- W.-C. Chang and Y. T. Su, "Sparse Bayesian learning based tensor dictionary learning and signal recovery with application to MIMO channel estimation," *IEEE J. Sel. Top. Signal Process.*, vol. 15, no. 3, pp. 847–859, Apr. 2021.
- X. Xu, S. Zhang, F. Gao, and J. Wang, "Sparse Bayesian learning based channel extrapolation for RIS assisted MIMO-OFDM," *IEEE Trans. Commun.*, vol. 70, no. 8, pp. 5498–5513, Aug. 2022.
- H. Zhu, G. Leus, and G. B. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2002–2016, May 2011.
- G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7465–7490, Nov. 2013.
- Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse Bayesian inference," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 38–43, Jan. 2012.
- K. You, W. Guo, T. Peng, Y. Liu, P. Zuo, and W. Wang, "Parametric sparse Bayesian dictionary learning for multiple sources localization with propagation parameters uncertainty," *IEEE Trans. Signal Process.*, vol. 68, pp. 4194–4209, 2020.
- Y. Mao, Q. Guo, J. Ding, F. Liu, and Y. Yu, "Marginal likelihood maximization based fast array manifold matrix learning for direction of arrival estimation," *IEEE Trans. Signal Process.*, vol. 69, pp. 5512–5522, 2021.
- J. Dai, A. Liu, and V. K. Lau, "FDD massive MIMO channel estimation with arbitrary 2D-array geometry," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2584–2599, May 2018.
- M. Ibrahim, F. Römer, R. Alieiev, G. D. Galdo, and R. S. Thomä, "On the estimation of grid offsets in CS-based direction-of-arrival estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal, Process.*, 2014, pp. 6776–6780.
- A. Faul and M. Tipping, "Analysis of sparse Bayesian learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 14, pp. 383–389, Dec. 2001.
- S. E. Ament and C. P. Gomes, "Sparse Bayesian learning via stepwise regression," in *Proc. Int. Conf. Mach. Learn.*, Jul. 2021, pp. 264–274.
- R. R. Pote and B. D. Rao, "Light-weight sequential SBL algorithm: An alternative to OMP," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Jun. 2023, pp. 1–5.
- A. Lin, A. H. Song, B. Bilgic, and D. Ba, "Covariance-free sparse Bayesian learning," *IEEE Trans. Signal Process.*, vol. 70, pp. 3818–3831, 2022.
- C. K. Thomas and D. Slock, "Space alternating variational estimation and Kronecker structured dictionary learning," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Piscataway, NJ, USA: IEEE Press, 2019, pp. 5556–5560.
- T. Jia, H. Liu, C. Gao, and J. Yan, "Bayesian direction of arrival estimation using atomic norm minimization with prior knowledge," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 60, no. 5, pp. 5742–5755, Oct. 2024.
- Z. Yang and L. Xie, "Frequency-selective Vandermonde decomposition of Toeplitz matrices with applications," *Signal Process.*, vol. 142, pp. 157–167, Jul. 2018.
- Z. Yang and L. Xie, "On gridless sparse methods for line spectral estimation from complete and incomplete data," *IEEE Trans. Signal Process.*, vol. 63, no. 12, pp. 3139–3153, Jun. 2015.
- M.-A. Badiu, T. L. Hansen, and B. H. Fleury, "Variational Bayesian inference of line spectra," *IEEE Trans. Signal Process.*, vol. 65, no. 9, pp. 2247–2261, Jan. 2017.
- J. Zhu, Q. Zhang, P. Gerstoft, M.-A. Badiu, and Z. Xu, "Gridless variational Bayesian line spectral estimation with multiple measurement vectors," *Signal Process.*, vol. 161, pp. 155–164, Mar. 2019.
- A. Cichocki et al., "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, Mar. 2015.
- 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, Apr. 2000.
- A. Zhang and D. Xia, "Tensor SVD: Statistical and computational limits," *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 7311–7338, May 2018.
- E. R. Balda, S. A. Cheema, J. Steinwandt, M. Haardt, A. Weiss, and A. Yeredor, "First-order perturbation analysis of low-rank tensor approximations based on the truncated HOSVD," in *Proc. Asilomar Conf. Signals Syst. Comput.*, Nov. 2016, pp. 1723–1727.

- [33] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Jun. 2001.
- [34] G. Joseph and C. R. Murthy, "On the convergence of a Bayesian algorithm for joint dictionary learning and sparse recovery," *IEEE Trans. Signal Process.*, vol. 68, pp. 343–358, Nov. 2019.
- [35] S. Khanna and C. R. Murthy, "On the support recovery of jointly sparse Gaussian sources via sparse Bayesian learning," *IEEE Trans. Inf. Theory*, vol. 68, no. 11, pp. 7361–7378, Jun. 2022.
- [36] C. J. Wu, "On the convergence properties of the EM algorithm," *Ann. Statist.*, vol. 11, pp. 95–103, Mar. 1983.
- [37] D. C. Araújo, A. L. De Almeida, J. P. Da Costa, and R. T. de Sousa, "Tensor-based channel estimation for massive MIMO-OFDM systems," *IEEE Access*, vol. 7, pp. 42133–42147, Mar. 2019.
- [38] C. F. Van Loan, "The ubiquitous Kronecker product," *J. Comput. Appl. Math.*, vol. 123, nos. 1–2, pp. 85–100, Nov. 2000.
- [39] C. R. Rao, "Estimation of heteroscedastic variances in linear models," *J. Am. Stat. Assoc.*, vol. 65, no. 329, pp. 161–172, Apr. 1970.
- [40] R. Schroeder, J. He, H. Djelouat, and M. Juntti, "Low-complexity near-field channel estimation for hybrid RIS assisted systems," Apr. 2024, *arXiv:2404.17411*.
- [41] J. Wang, J. Fang, and H. Li, "Intelligent reflecting surface-assisted NLOS sensing via tensor decomposition," in *Proc. EUSIPCO*, Aug. 2024, pp. 1–5.
- [42] Y. Wang, G. Leus, and A. Pandharipande, "Direction estimation using compressive sampling array processing," in *IEEE Workshop Stat. Signal Process.*, Aug. 2009, pp. 626–629.
- [43] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, and L. De Lathauwer, "Tensorlab 3.0." 2016. [Online]. Available: <https://www.tensorlab.net>
- [44] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.
- [45] M. Guo, Y. D. Zhang, and T. Chen, "DOA estimation using compressed sparse array," *IEEE Trans. Signal Process.*, vol. 66, no. 15, pp. 4133–4146, Jun. 2018.
- [46] D. Dampahalage, K. S. Manosha, N. Rajatheva, and M. Latva-Aho, "Supervised learning based sparse channel estimation for RIS aided communications," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, May 2022, pp. 8827–8831.
- [47] J. He, M. Leinonen, H. Wymeersch, and M. Juntti, "Channel estimation for RIS-aided mmWave MIMO systems," in *Proc. IEEE Glob. Commun. Conf.*, Feb. 2020, pp. 1–6.
- [48] Z. Wan, Z. Gao, F. Gao, M. Di Renzo, and M.-S. Alouini, "Terahertz massive MIMO with holographic reconfigurable intelligent surfaces," *IEEE Trans. Commun.*, vol. 69, no. 7, pp. 4732–4750, Mar. 2021.
- [49] L. Yashvanth and C. R. Murthy, "Cascaded channel estimation for distributed IRS aided mmWave massive MIMO systems," in *Proc. IEEE Glob. Commun. Conf.*, Dec. 2022, pp. 717–723.
- [50] B. Zheng, C. You, W. Mei, and R. Zhang, "A survey on channel estimation and practical passive beamforming design for intelligent reflecting surface aided wireless communications," *IEEE Commun. Surv. Tut.*, vol. 24, no. 2, pp. 1035–1071, Feb. 2022.
- [51] Y. Liu et al., "Reconfigurable intelligent surfaces: Principles and opportunities," *IEEE Commun. Surv. Tut.*, vol. 23, no. 3, pp. 1546–1577, May 2021.
- [52] Y. Lin, S. Jin, M. Matthaiou, and X. You, "Channel estimation and user localization for IRS-assisted MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2320–2335, Apr. 2021.
- [53] T. T. Cai and A. Zhang, "Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics," *Ann. Statist.*, vol. 46, no. 1, pp. 60–89, Feb. 2018.
- [54] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," Nov. 2010, *arXiv:1011.3027*.
- [55] L. Birgé, "An alternative point of view on Lepski's method," *Lecture Notes-Monograph Ser.*, pp. 113–133, Jul. 2001.
- [56] A. Zhang and R. Han, "Optimal sparse singular value decomposition for high-dimensional high-order data," *J. Am. Statist. Assoc.*, vol. 114, no. 528, pp. 1708–1725, Mar. 2019.