



Delft University of Technology

Document Version

Final published version

Licence

Dutch Copyright Act (Article 25fa)

Citation (APA)

He, Y., & Joseph, G. (2025). Kronecker-structured Sparse Vector Recovery with Application to IRS-MIMO Channel Estimation. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*. <https://doi.org/10.1109/ICASSP49660.2025.10890014>

Important note

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

Copyright

In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership. Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse

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.

This work is downloaded from Delft University of Technology.

Kronecker-structured Sparse Vector Recovery with Application to IRS-MIMO Channel Estimation

Yanbin He and Geethu Joseph

Signal Processing Systems Group, Delft University of Technology, The Netherlands

Emails: {y.he-1,g.joseph}@tudelft.nl

Abstract—We study the recovery of a sparse vector with a Kronecker structure from an underdetermined linear system with a Kronecker-structured dictionary. This problem arises in several applications, such as the channel estimation of an intelligent reflecting surface-aided wireless system. Existing work only exploits the Kronecker structure in support of the sparse vector and solves the entire linear system jointly with high complexity. Instead, we decompose the original sparse recovery problem into multiple independent subproblems and solve them individually. We obtain the sparse vector as the Kronecker product of individual solutions, retaining its Kronecker structure. Besides, the subproblems exhibit reduced effective measurement noise. Our simulations demonstrate that our method has superior estimation accuracy and runtime compared to the existing work. We attribute the low complexity to the reduced dimensionality of the subproblems and improved accuracy to the denoising effect of the decomposition step.

Index Terms—Basis expansion model, sparse Bayesian learning, singular value decomposition, angular sparsity

I. INTRODUCTION

Compressed sensing (CS) has been extensively applied in diverse domains and its success stems from the inherent redundancies of many real-world signals. With suitable bases, many signals can be represented by a small subset of bases and thus be *sparingly* expressed [1]. Further, CS is a powerful tool for estimating the unknown parameters of non-linear functions by exploiting sparsity using the basis expansion model (BEM) [2]. Specifically, BEM expresses the parametric non-linear function as the product of a *known overcomplete dictionary* of the basis functions and an *unknown sparse coefficient vector*. Here, the basis functions are obtained by sampling the unknown parameter over its range. As only a few functions corresponding to the truth are activated, the coefficient vector is sparse and can be recovered by the CS techniques. We aim to study the sparse recovery problems arising in the context of BEM, applied to the intelligent reflecting surfaces (IRS)-aided wireless communication system for channel estimation.

Beyond the standard CS framework, BEM can exhibit additional structure, such as Kronecker-structured dictionary and Kronecker-structured support of the sparse vector, arising from image processing [3] or wireless communications [4]–[6]. This problem generally takes the following canonical form,

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad (1)$$

where $\mathbf{y} \in \mathbb{C}^{\bar{M} \times 1}$ is the noisy measurement, $\mathbf{H} \in \mathbb{C}^{\bar{M} \times \bar{N}}$ is the dictionary with $\bar{M} < \bar{N}$, $\mathbf{x} \in \mathbb{C}^{\bar{N} \times 1}$ is the unknown

sparse coefficient vector, and \mathbf{n} is the measurement noise. The Kronecker-structured dictionary [3]–[8] can be represented as

$$\mathbf{H} = \mathbf{H}_1 \otimes \mathbf{H}_2 \otimes \cdots \otimes \mathbf{H}_I = \otimes_{i=1}^I \mathbf{H}_i, \quad (2)$$

where $\mathbf{H}_i \in \mathbb{C}^{M_i \times N_i}$ with $\prod_{i=1}^I M_i = \bar{M}$ and $\prod_{i=1}^I N_i = \bar{N}$. Similarly, the support of \mathbf{x} can be expressed as the Kronecker product of the I support vectors of sizes N_1, N_2, \dots, N_I . Here, the goal is to estimate \mathbf{x} given the Kronecker-structured dictionary \mathbf{H} and noisy measurement \mathbf{y} .

The linear inversion problem (1) with Kronecker-structured dictionary (2) has been widely discussed [3]–[8]. One method is based on the orthogonal matching pursuit (OMP), which has low complexity but requires hand-tuning of the stopping criterion. Another approach employs the sparse Bayesian learning (SBL) framework called Kronecker SBL (KroSBL) [4]–[6], [9]. KroSBL assumes a sparse-promoting prior on \mathbf{x} to impose the Kronecker-structured support. It improves the recovery accuracy but at the cost of higher computational complexity.

Further, in applications like intelligent reflecting surface (IRS)-aided wireless channel estimation, BEM can lead to a Kronecker-structured sparse \mathbf{x} , i.e., along with the support, the entries of \mathbf{x} are also Kronecker-structured. For example, in IRS channel estimation [6], [9], BEM utilizes angular sparsity by sampling pre-defined spatial angle grids for the angle-of-departure (AoD) at the transmitter, angle-of-arrival (AoA) at the receiver, and the difference between the AoA and AoD at the IRS. It constitutes an overcomplete dictionary with steering vectors of different combinations of the grid points. Then, the different combinations of AoDs and AoAs result in a Kronecker-structured dictionary and sparse coefficients. Mathematically, the channel estimation takes the form of (1) and (2), and the sparse channel vector is given by

$$\mathbf{x} = \otimes_{i=1}^I \mathbf{x}_i, \quad (3)$$

where $\mathbf{x}_i \in \mathbb{C}^{N_i \times 1}$. However, existing algorithms [6], [9] do not fully exploit prior knowledge of the Kronecker structure in (3). Therefore, we present an efficient method by enforcing (3) to solve (1), with improved accuracy and runtime compared to [6], [9]. Our contributions are two-fold:

- *Decomposition-based algorithm*: We solve the linear inversion problem by first decomposing measurements \mathbf{y} into I sub-vectors $\{\mathbf{y}_i\}_{i=1}^I$ to obtain each \mathbf{x}_i separately, and reconstruct the solution to (1) as $\otimes_{i=1}^I \mathbf{x}_i$.
- *IRS-MIMO channel estimation*: Our method applied to IRS-

MIMO channel estimation indicates that, due to the decomposition step, our algorithm achieves low complexity (with runtime reduced by two orders of magnitude) and implicitly reduces the noise level. This combination leads to better reconstruction performance than existing algorithms.

Overall, our algorithm ensures low complexity, efficient implementation, and enhanced reconstruction performance by enforcing prior knowledge and reducing noise, making it a practical choice for IRS-MIMO channel estimation.

II. KRONECKER-STRUCTURED SPARSE RECOVERY

We study the Kronecker-structured sparse recovery problem of estimating the vector \mathbf{x} in (3) from the measurements \mathbf{y} in (1) and the Kronecker-structured measurement matrix \mathbf{H} in (2). We start with the noiseless set of linear equations and then extend the algorithm to the noisy case. We consider

$$\mathbf{y} = \mathbf{H}\mathbf{x}, \quad (4)$$

with $\mathbf{H} = \otimes_{i=1}^I \mathbf{H}_i$, and $\mathbf{x} = \otimes_{i=1}^I \mathbf{x}_i$. We need the following lemma to devise a decomposition-based algorithm.

Lemma 1. [9, Lemma 4] *Consider a set of linear equations, $(\mathbf{H}_1 \otimes \mathbf{H}_2)(\mathbf{x}_1 \otimes \mathbf{x}_2) = \mathbf{y}_1 \otimes \mathbf{y}_2$ with $\mathbf{y}_1, \mathbf{y}_2 \neq \mathbf{0}$. Solving for $\mathbf{x}_1 \otimes \mathbf{x}_2$ from the linear equations is equivalent to solving $\mathbf{H}_1(\alpha \mathbf{x}_1) = \mathbf{y}_1$ and $\mathbf{H}_2(\alpha^{-1} \mathbf{x}_2) = \mathbf{y}_2$, for any scalar $\alpha \neq 0$ accounting for the scaling ambiguity.*

Here, we can estimate individual vectors \mathbf{x}_1 and \mathbf{x}_2 only up to a scaling ambiguity denoted by α . However, any $\alpha \neq 0$ leads to the same Kronecker product $\mathbf{x}_1 \otimes \mathbf{x}_2 = (\alpha \mathbf{x}_1) \otimes (\alpha^{-1} \mathbf{x}_2)$. A trivial extension of the Lemma 1 to the Kronecker product of I vectors immediately suggests that the noiseless problem (4) can be decomposed into I smaller sparse recovery problems, instead of jointly solving for all vectors $\{\mathbf{x}_i\}_{i=1}^I$.

To elaborate, we first decompose \mathbf{y} into low-dimensional vectors $\{\mathbf{y}_i \in \mathbb{C}^{M_i}\}_{i=1}^I$ such that $\mathbf{y} = \otimes_{i=1}^I \mathbf{y}_i$, to disentangle different \mathbf{x}_i 's into I subproblems $\mathbf{y}_i = \mathbf{H}_i(\alpha_i \mathbf{x}_i)$, for some α_i with $\prod_{i=1}^I \alpha_i = 1$. As mentioned above, α_i 's are not designable parameters but represent the inevitable scaling ambiguity in estimating \mathbf{x}_i 's separately. We find \mathbf{y}_i using $(I-1)$ recursive rank-one decomposition. To circumvent the scaling ambiguity, we add a unit norm constraint to arrive at

$$\bar{\mathbf{y}}_{i-1} = \mathbf{y}_i \otimes \bar{\mathbf{y}}_i \text{ and } \|\mathbf{y}_i\| = 1, \quad \forall i \in [I-1], \quad (5)$$

with $\bar{\mathbf{y}}_0 = \mathbf{y}$, $\bar{\mathbf{y}}_{I-1} = \mathbf{y}_I$, and set $[I-1] := \{1, 2, \dots, I-1\}$. Once $\{\mathbf{y}_i\}_{i=1}^I$ is obtained, we solve I sparse recovery problems given by $\mathbf{y}_i = \mathbf{H}_i(\alpha_i \mathbf{x}_i)$ for $i \in [I]$ to obtain $\{\alpha_i \mathbf{x}_i\}_{i=1}^I$ and the final estimate of $\mathbf{x} = \otimes_{i=1}^I \alpha_i \mathbf{x}_i = \otimes_{i=1}^I \mathbf{x}_i$.

Extending to the noisy setting in (1), the decomposition step of \mathbf{y} minimizes $\|\mathbf{y} - \otimes_{i=1}^I \mathbf{y}_i\|_2$, and we replace (5) with

$$(\mathbf{y}_i, \bar{\mathbf{y}}_i) = \arg \min_{(\mathbf{z} \in \mathbb{C}^{M_i}, \bar{\mathbf{z}}), \|\mathbf{z}\|_2=1} \|\bar{\mathbf{y}}_{i-1} - \mathbf{z} \otimes \bar{\mathbf{z}}\|_2, \quad \forall i \in [I-1], \quad (6)$$

with $\bar{\mathbf{y}}_0 = \mathbf{y}$ and $\bar{\mathbf{y}}_{I-1} = \mathbf{y}_I$. To solve (6), consider the case when $i = 1$. We rearrange \mathbf{y} as $\mathbf{Y} \in \mathbb{C}^{\bar{M}/M_1 \times M_1}$, where \mathbf{y} is the vectorization of \mathbf{Y} . Using $\mathbf{z} \otimes \bar{\mathbf{z}} = \text{vec}(\bar{\mathbf{z}} \mathbf{z}^T)$, we note that (6) with $i = 1$ is a rank-one approximation $\|\mathbf{Y} - \bar{\mathbf{z}} \mathbf{z}^T\|_F$, and \mathbf{z}

Algorithm 1 Decomposition-based sparse recovery

Input: Measurement \mathbf{y} , dictionaries $\{\mathbf{H}_i\}_{i=1}^I$

1: **for** $i = 1, 2, \dots, I$ **do**

2: Solve (6) to obtain \mathbf{y}_i

3: Solve (7) for \mathbf{x}_i using any CS algorithm

4: **end for**

Output: Sparse vector $\mathbf{x} = \otimes_{i=1}^I \mathbf{x}_i$

is the left leading singular vector of \mathbf{Y} , which can be obtained by the singular value decomposition (SVD). Further, (1)-(3) lead to $\mathbf{y} = (\mathbf{H}_1 \mathbf{x}_1) \otimes (\otimes_{i=2}^I \mathbf{H}_i \mathbf{x}_i) + \mathbf{n}$. Due to $\mathbf{z} \otimes \bar{\mathbf{z}} \neq \bar{\mathbf{z}} \otimes \mathbf{z}$, \mathbf{z} and $\bar{\mathbf{z}}$ in (6) naturally match the estimates of $\mathbf{H}_1 \mathbf{x}_1$ and $\otimes_{i=2}^I \mathbf{H}_i \mathbf{x}_i$ (up to scalars). Similarly, (6) can be solved by using SVD recursively without any assumptions. After SVDs, we have I separate sets of noisy linear equations,

$$\mathbf{y}_i = \mathbf{H}_i(\alpha_i \mathbf{x}_i) + \mathbf{n}_i, \quad \forall i \in [I], \quad (7)$$

where \mathbf{n}_i is the noise term. We solve (7) using CS algorithms such as OMP or SBL. The resulting algorithm, denoted as the decomposition-based Sparse Recovery (dSR), is in Algorithm 1. Here, the noise \mathbf{n} in (1) need not have a Kronecker structure. However, our method is effective in noisy scenarios and can even aid in denoising the signal, as discussed next.

A. Denoising Effect

Denoising refers to the noise reduction in the measurements after decomposition, i.e., $\mathbb{E}\{\|\otimes_{i=1}^I \mathbf{y}_i - \mathbf{H}\mathbf{x}\|^2\} < \mathbb{E}\{\|\mathbf{n}\|^2\}$, where $\{\mathbf{y}_i\}_{i=1}^I$ are obtained after the decomposition step (6). To intuitively explain denoising, we reformulate the measurements \mathbf{y} in (1) as a rank-one matrix $(\otimes_{i=2}^I \mathbf{H}_i \mathbf{x}_i)(\mathbf{H}_1 \mathbf{x}_1)^T$ with noise using [10, Eq. (10)],

$$\mathbf{Y} = (\otimes_{i=2}^I \mathbf{H}_i \mathbf{x}_i)(\mathbf{H}_1 \mathbf{x}_1)^T + \mathbf{N}, \quad (8)$$

where $\text{vec}(\cdot)$ denotes vectorization, $\text{vec}(\mathbf{Y}) = \mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$, and $\text{vec}(\mathbf{N}) = \mathbf{n}$. The first step ($i = 1$) of (6), yields $\mathbf{y}_1 \otimes \bar{\mathbf{y}}_1$ as the estimate of $\alpha_1^{-1}(\otimes_{i=2}^I \mathbf{H}_i \mathbf{x}_i)$ and $\alpha_1 \mathbf{H}_1 \mathbf{x}_1$, respectively, for some α_1 , approximating $(\otimes_{i=2}^I \mathbf{H}_i \mathbf{x}_i)(\mathbf{H}_1 \mathbf{x}_1)^T$ by $\bar{\mathbf{y}}_1 \mathbf{y}_1^T$. Therefore, it reduces noise by extracting the rank-one part of the measurement, filtering out higher-rank components that arise due to noise. So, the noise level reduces [11], even after just the first step of (6), as characterized below.

Lemma 2. *Suppose that the noise \mathbf{n} in (1) is zero-mean white Gaussian noise with variance σ^2 . If $\bar{\mathbf{y}}_1$ and \mathbf{y}_1 are obtained from \mathbf{y} using (6), then*

$$\mathbb{E}\{\|\mathbf{y}_1 \otimes \bar{\mathbf{y}}_1 - \mathbf{H}\mathbf{x}\|^2\} \approx \sigma^2(M_1 - 1 + \bar{M}/M_1). \quad (9)$$

Proof. The result is a special case of perturbation analysis of low-rank tensor approximations in [12] by setting the tensor order $R = 2$ and r -ranks as $p_1 = p_2 = 1$ in [12, Eq. (19)]. \square

Lemma 2 states that after the first step of decomposition (6), the noise level $\mathbb{E}\{\|\mathbf{n}\|^2\} = \sigma^2 \bar{M}$ reduces approximately as

$$\frac{\mathbb{E}\{\|\mathbf{y}_1 \otimes \bar{\mathbf{y}}_1 - \mathbf{H}\mathbf{x}\|^2\}}{\mathbb{E}\{\|\mathbf{n}\|^2\}} \approx \frac{M_1 - 1}{\bar{M}} + \frac{1}{M_1} < 1. \quad (10)$$

TABLE I: Demonstration of denoising with $M_i = 10$ for $i = 1, 2, 3$, using the original noisy signal \mathbf{y} , reconstructed signal $\otimes_{i=1}^3 \mathbf{y}_i$ after the decomposition step, and ground truth $\mathbf{H}\mathbf{x}$.

SNR (dB)	5	10	15	20	25	30
$\ \mathbf{y} - \mathbf{H}\mathbf{x}\ _2^2$	34.661	9.553	2.853	0.972	0.304	0.090
$\ \otimes_{i=1}^3 \mathbf{y}_i - \mathbf{H}\mathbf{x}\ _2^2$	0.980	0.272	0.080	0.027	0.009	0.002

TABLE II: Complexity of different schemes. R_{AM} : the number of AM iterations. R_{EM} : the number of EM iterations.

Method	Time Complexity	Space Complexity
AM-KroSBL	$\mathcal{O}(R_{EM}(R_{AM}IN^I + (MN)^I))$	$\mathcal{O}((MN)^I)$
SVD-KroSBL	$\mathcal{O}(R_{EM}(N^{I+1} + (MN)^I))$	$\mathcal{O}((MN)^I)$
dSBL	$\mathcal{O}(R_{EM}N^2MI + M^{I+1})$	$\mathcal{O}(M^I + MN + N^2)$

We further corroborate the overall denoising after $I-1$ steps of (6) by comparing $\|\otimes_{i=1}^I \mathbf{y}_i - \mathbf{H}\mathbf{x}\|$ and $\|\mathbf{y} - \mathbf{H}\mathbf{x}\|$ via simulations in Table I. Here, we generate the noiseless signal $\mathbf{H}\mathbf{x}$ with $I = 3$ and $M_i = 10$ for $i = 1, 2, 3$ using (1). We add zero-mean Gaussian noise \mathbf{n} with different SNRs accordingly. Results show that the noise level is significantly reduced. Moreover, decomposition leads to I low-dimensional sparse recovery problems, shortening runtime, as discussed next.

B. Complexity Analysis

The time and space complexities of Algorithm 1 are $\mathcal{O}(M^{I+1} + IT_{CS})$ and $\mathcal{O}(M^I + MN + S_{CS})$, respectively, assuming $M_i = \mathcal{O}(M)$, $N_i = \mathcal{O}(N)$, and $I \ll M < N$. Here, T_{CS} and S_{CS} are the time and space complexities of the sparse recovery algorithm used with dSR. Also, all sparse recovery subproblems are independent of each other and can be solved in parallel. In that case, the time and space complexity dSR changes to $\mathcal{O}(M^{I+1} + T_{CS})$ and $\mathcal{O}(M^I + IMN + IS_{CS})$, respectively. We consider dSR with SBL for sparse recovery, namely decomposition-based SBL (dSBL), and compare the complexity in Table II with alternating minimization-based KroSBL (AM-KroSBL) and SVD-KroSBL [9].

C. Comparison with SVD-KroSBL

The decomposition step of our algorithm resembles the optimization problem in the SVD-KroSBL algorithm [9]. However, they differ as KroSBL seeks a sparse vector with Kronecker-structured support without considering the Kronecker structure (3) of \mathbf{x} . It relies on Type-II learning to estimate the sparse vector \mathbf{x} , assuming that the measurement noise is Gaussian distributed, $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I})$. KroSBL exploits the Kronecker-structured support of \mathbf{x} by using a Gaussian prior distribution on \mathbf{x} [6] given by $p(\mathbf{x}; \{\gamma_i\}_{i=1}^I) = \mathcal{CN}(\mathbf{0}, \text{diag}(\otimes_{i=1}^I \gamma_i))$, where $\{\gamma_i\}_{i=1}^I$ are the unknown hyperparameters. Then, it solves for the maximum likelihood $\{\gamma_i\}_{i=1}^I$ and uses $\{\gamma_i\}_{i=1}^I$ to retrieve \mathbf{x} . In particular, SVD-KroSBL uses SVD in the r th iteration similar to our dSR as

$$\gamma_i^{(r)} = \arg \min_{\gamma_i: \|\gamma_i\|_2=1, \tilde{\gamma}_i \in \mathbb{R}^{N(I-i)}} \|\tilde{\gamma}_{i-1} - \gamma_i \otimes \tilde{\gamma}_i\|_2, i \in [I-1],$$

where $\tilde{\gamma}_{I-1} = \gamma_I$ and $\tilde{\gamma}_0$ is from the previous iteration [9]. Although this is a decomposition step like (6), they differ in their specific approaches. First, adopting a prior using $\otimes_{i=1}^I \gamma_i$

is to mimic the Kronecker-structured support of the sparse vector (3), and not to decompose the original sparse recovery problem. Second, SVD-KroSBL is iterative where the sparse vector \mathbf{x} is estimated *jointly* using all $\{\gamma_i\}_{i=1}^I$ as in [6, Eq. (18)], which leads to higher complexity, as shown in Table II. Third, adopting $\gamma = \otimes_{i=1}^I \gamma_i$ does not ensure the Kronecker structure (3) in the sparse vector \mathbf{x} . Similarly, if assumption (3) does not hold, our dSR framework fails.

III. CASCADED CHANNEL ESTIMATION FOR A PROTOTYPICAL IRS-AIDED SYSTEM

In this section, we discuss the application of dSR to the problem of IRS-MIMO channel estimation. We consider an uplink narrowband millimeter-wave or terahertz band MIMO system with a T -antenna transmitter mobile station (MS) and an R -antenna receiver base station (BS), served by an L -element uniform linear array IRS.

We start with the channel model in Fig. 1. The channel is the concatenation of the MS-IRS channel \mathbf{H}_{MS} and the IRS-BS channel \mathbf{H}_{BS} . We adopt the geometric model [13]–[15] as

$$\mathbf{H}_{MS} = \sum_{p=1}^{P_{MS}} \sqrt{\frac{LT}{P_{MS}}} \beta_{MS,p} \mathbf{a}_L(\phi_{MS,p}) \mathbf{a}_T(\alpha_{MS})^H \quad (11)$$

$$\mathbf{H}_{BS} = \sum_{p=1}^{P_{BS}} \sqrt{\frac{RL}{P_{BS}}} \beta_{BS,p} \mathbf{a}_R(\alpha_{BS,p}) \mathbf{a}_L(\phi_{BS})^H, \quad (12)$$

where we define $\mathbf{a}_Q(\psi) \in \mathbb{C}^{Q \times 1}$ for an integer Q and angle ψ as $\mathbf{a}_Q(\psi) = 1/\sqrt{Q} [1, e^{j\frac{2\pi d}{\lambda} \cos \psi}, \dots, e^{j\frac{2\pi d}{\lambda} (Q-1) \cos \psi}]^T$. Here, d is the distance between two adjacent elements, and λ is the wavelength. Also, P_{MS} and P_{BS} are the number of spread angles. We denote the p th AoA of the IRS, AoD of the MS, the p th AoA of the BS, and the AoD of the IRS as $\phi_{MS,p}$, α_{MS} , $\alpha_{BS,p}$, and ϕ_{BS} , respectively (see Fig. 1).

For a given IRS configuration $\boldsymbol{\theta} \in \mathbb{C}^{L \times 1}$, the cascaded MS-IRS-BS channel is given by $\mathbf{H}_{BS} \text{diag}(\boldsymbol{\theta}) \mathbf{H}_{MS}$. The i th entry of $\boldsymbol{\theta}$ represents the gain and phase shift by the reflection of the i th IRS element. Our objective is to estimate the cascaded channel $\mathbf{H}_{BS} \text{diag}(\boldsymbol{\theta}) \mathbf{H}_{MS}$ for any given IRS configuration $\boldsymbol{\theta}$.

We assume constant \mathbf{H}_{MS} and \mathbf{H}_{BS} over K coherent slots. We send pilot data $\mathbf{X} \in \mathbb{C}^{T \times K_P}$ spanning K_P time slots repetitively for K_I IRS configurations $\{\boldsymbol{\theta}_k\}_{k=1}^{K_I}$ such that $K = K_I K_P$. Then, the received signal $\mathbf{Y}_k \in \mathbb{C}^{R \times K_P}$ corresponding to the k th configuration $\boldsymbol{\theta}_k$ is $\mathbf{Y}_k = \mathbf{H}_{BS} \text{diag}(\boldsymbol{\theta}_k) \mathbf{H}_{MS} \mathbf{X} + \mathbf{W}_k$ with $\mathbf{W}_k \in \mathbb{C}^{R \times K_P}$ being the noise.

To exploit the channel angular sparsity, we adopt a set of N predefined grid angles $\{\psi_n\}_{n=1}^N$ with $\cos(\psi_n) = 2n/N - 1$ [16]. We collect steering vectors of the grid angles to formulate the BEM dictionaries

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

for any integer $Q > 0$ referring to the number of elements in the array. Then the BEMs of (11) and (12) are

$$\mathbf{H}_{BS} = \mathbf{A}_R \mathbf{g}_R \mathbf{g}_{L,d}^H \mathbf{A}_L^H \text{ and } \mathbf{H}_{MS} = \mathbf{A}_L \mathbf{g}_{L,a} \mathbf{g}_T^H \mathbf{A}_T^H, \quad (13)$$

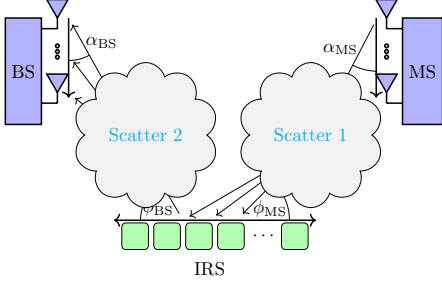


Fig. 1: An illustration of AoAs and AoDs in an IRS-aided uplink channel [6].

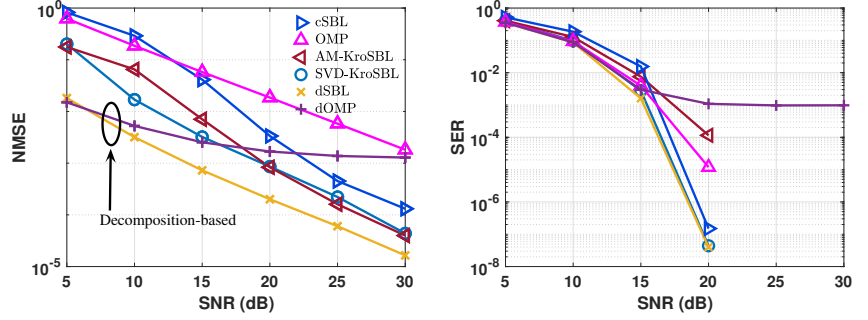


Fig. 2: NMSE and SER performance of different algorithms as functions of SNR with $K_I = 10$ and $K_P = 4$.

with $\mathbf{g}_R, \mathbf{g}_{L,d}, \mathbf{g}_{L,a}, \mathbf{g}_T \in \mathbb{C}^{N \times 1}$ carrying the unknown channel state information, including AoAs/AoDs of the channel and path gains. Rearranging (13) using \mathbf{Y}_k and reorganizing the received signal $\{\mathbf{Y}_k\}_{k=1}^{K_I}$ (see [6] for details), we arrive at

$$\tilde{\mathbf{y}} = (\Phi_L \otimes \Phi_T \otimes \Phi_R)(\mathbf{g}_L \otimes \mathbf{g}_T^* \otimes \mathbf{g}_R) + \tilde{\mathbf{w}} \in \mathbb{C}^{RK \times 1}, \quad (14)$$

where $\Phi_L \in \mathbb{C}^{K_I \times N}$ is the first N columns of $\Theta^T(\mathbf{A}_L^T \odot \mathbf{A}_L^H)^T$, $\Phi_T = \mathbf{X}^T \mathbf{A}_T^*$, and $\Phi_R = \mathbf{A}_R$. Hence, (14) transforms the channel estimation problem into a Kronecker-structured sparse recovery and we can apply Algorithm 1 to estimate the channel represented by $\mathbf{g}_L \otimes \mathbf{g}_T^* \otimes \mathbf{g}_R$. Since there are three dictionaries, i.e., Φ_L , Φ_T , and Φ_R , here $I = 3$.

Extension to multiuser IRS-MIMO is also possible. The shared BS-IRS channel has the same sparse vector \mathbf{g}_R across users, but different user positions and pilot signals \mathbf{X} lead to distinct $\mathbf{g}_L \otimes \mathbf{g}_T^*$ and $\Phi_L \otimes \Phi_T$. We modify (14) as

$$\tilde{\mathbf{y}} = \left\{ \sum_u (\Phi_L \otimes \Phi_T^{(u)})(\mathbf{g}_L^{(u)} \otimes \mathbf{g}_T^{*(u)}) \otimes \Phi_R \mathbf{g}_R \right\} + \tilde{\mathbf{w}},$$

where u is the user index. By decomposing $\tilde{\mathbf{y}} \approx \mathbf{y}_1 \otimes \tilde{\mathbf{y}}_1$, we solve for \mathbf{g}_R and $\mathbf{g}_L^{(u)} \otimes \mathbf{g}_T^{*(u)}$ using linear models $\tilde{\mathbf{y}}_1 = \Phi_R \mathbf{g}_R$ and $\mathbf{y}_1 = \sum_u (\Phi_L \otimes \Phi_T^{(u)})(\mathbf{g}_L^{(u)} \otimes \mathbf{g}_T^{*(u)})$, to obtain the channel estimates for each user.

IV. NUMERICAL RESULTS: IRS-MIMO CHANNEL ESTIMATION

We apply dSR to the IRS-MIMO channel estimation problem using two versions of dSR: decomposition-based OMP (dOMP) and dSBL. For benchmarking, we implement classic SBL (cSBL) [17], OMP [18], AM-KroSBL, and SVD-KroSBL [6]. We use $R = 16$ BS antennas, $T = 6$ MS antennas, $L = 256$ IRS elements, $P_{BS} = P_{MS} = 3$ spread angles, and spacing $d = \lambda/2$. The IRS configurations $\{\theta_k\}_{k=1}^{K_I}$ are selected uniformly from $\{-1/\sqrt{L}, 1/\sqrt{L}\}$ with $K_I = 10$. We send $K_P = 4$ pilot signals for each IRS configuration, resulting in $K = K_I K_P = 40$ pilot signals. A predefined grid with $N = 18$ angles is used in the BEM dictionaries, assuming uniform distribution of AoAs and AoDs. The channel gains $\{\beta_{BS,p}\}_{p=1}^{P_{BS}}$ and $\{\beta_{MS,p}\}_{p=1}^{P_{MS}}$ in (11) and (12) are drawn from the standard complex Gaussian

distribution [19]. The measurement noise is zero-mean white Gaussian. We use three metrics for comparison: normalized mean square error (NMSE) for channel estimation quality [15], [19], [20], symbol error rate (SER) for communication performance [6], [9], and runtime for complexity. Here, NMSE is $\frac{1}{K_I} \sum_{k=1}^{K_I} \frac{\|\mathbf{H}_{BS} \text{diag}(\theta_k) \mathbf{H}_{MS} - \hat{\mathbf{H}}_{BS} \text{diag}(\theta_k) \hat{\mathbf{H}}_{MS}\|_F^2}{\|\mathbf{H}_{BS} \text{diag}(\theta_k) \mathbf{H}_{MS}\|_F^2}$, where the channel estimate is $\hat{\mathbf{H}}_{BS} \text{diag}(\theta_k) \hat{\mathbf{H}}_{MS}$. We compute SER based on 10^6 uncoded 8-QAM symbols decoded using the channel estimate.

Fig. 2 shows that our approach outperforms the existing IRS channel estimation algorithms. Notably, dSBL has the best NMSE and SER. When SNR is low, dOMP outperforms all other methods except dSBL. As the SNR increases, dOMP continues to outperform OMP. Moreover, the under-sampling ratio is $KR/N^I = 40 \times 16/18^3 \approx 10\%$, showing good performance in the lower measurement regime with limited pilot signals. This improved performance can be attributed to two factors: *i*) we explicitly incorporate the Kronecker structure of $\mathbf{g}_L \otimes \mathbf{g}_T^* \otimes \mathbf{g}_R$ in the channel estimation by the decomposition step in dSBL and dOMP, and *ii*) the reduced noise perturbation in $\tilde{\mathbf{y}}$ after the decomposition (denoising).

Furthermore, our approach offers a runtime that is at least two orders of magnitude shorter than existing algorithms. For example, when SNR = 20 dB, the runtime in seconds is 12.595, 5.071, 2.904, 0.006, 0.001, and 0.165 for cSBL, AM-KroSBL, SVD-KroSBL, dSBL, dOMP, and OMP, respectively. Similar trends are observed across other SNR levels, underscoring the low complexity of our dSR framework.

V. CONCLUSION

We studied the sparse recovery problem with a Kronecker-structured dictionary and sparse vector, specifically in the context of channel estimation for IRS-MIMO systems. We first decomposed the problem into independent subproblems and solved them separately, exploiting the Kronecker structure. Our simulations showed that this approach improves the accuracy and runtime compared to the state-of-the-art methods, as the decomposition step leads to dimensionality reduction and denoising. Establishing theoretical guarantees and mathematically analyzing the denoising effect are exciting avenues for future research.

REFERENCES

- [1] K. Kreutz-Delgado, B. Rao, I. Fedorov, and S. Das, "Dictionaries in machine learning," in *Signal Process. Mach. Learn. Theory*, pp. 1073–1159, Elsevier, 2024.
- [2] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: Data mining, inference, and prediction*, vol. 2. Springer, 2009.
- [3] R. Zhao, Q. Wang, J. Fu, and L. Ren, "Exploiting block-sparsity for hyperspectral Kronecker compressive sensing: A tensor-based Bayesian method," *IEEE Trans. Image Process.*, vol. 29, pp. 1654–1668, Oct. 2019.
- [4] 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, pp. 847–859, Apr. 2021.
- [5] 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, pp. 5498–5513, Aug. 2022.
- [6] 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.*, pp. 1–5, Jun. 2023.
- [7] C. F. Caiafa and A. Cichocki, "Computing sparse representations of multidimensional signals using Kronecker bases," *Neural Comput.*, vol. 25, pp. 186–220, Jan. 2013.
- [8] M. F. Duarte and R. G. Baraniuk, "Kronecker compressive sensing," *IEEE Trans. Image Process.*, vol. 21, pp. 494–504, Aug. 2011.
- [9] Y. He and G. Joseph, "Bayesian algorithms for Kronecker-structured sparse vector recovery with application to IRS-MIMO channel estimation," *IEEE Trans. Signal Process.*, 2024.
- [10] K. Schacke, *On the Kronecker product*. Master's thesis, University of Waterloo, 2004.
- [11] T. T. Cai and A. Zhang, "Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics," *Ann. Stat.*, vol. 46, no. 1, pp. 60 – 89, 2018.
- [12] 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.*, pp. 1723–1727, IEEE, 2016.
- [13] Y. You, L. Zhang, M. Yang, Y. Huang, X. You, and C. Zhang, "Structured OMP for IRS-Assisted mmWave channel estimation by exploiting angular spread," *IEEE Trans. Veh. Technol.*, vol. 71, pp. 4444–4448, Apr. 2022.
- [14] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 5, pp. 831–846, 2014.
- [15] P. Wang, J. Fang, H. Duan, and H. Li, "Compressed channel estimation for intelligent reflecting surface-assisted millimeter wave systems," *IEEE Signal Process. Lett.*, vol. 27, pp. 905–909, May 2020.
- [16] Z. Mao, X. Liu, and M. Peng, "Channel estimation for intelligent reflecting surface assisted massive MIMO systems—A deep learning approach," *IEEE Commun. Lett.*, vol. 26, pp. 798–802, Jan. 2022.
- [17] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, pp. 2153–2164, Aug. 2004.
- [18] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inf. Theory.*, vol. 57, no. 7, pp. 4680–4688, 2011.
- [19] 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, pp. 2320–2335, Apr. 2021.
- [20] K. Ardah, S. Gherekhloo, A. L. de Almeida, and M. Haardt, "TRICE: A channel estimation framework for RIS-aided millimeter-wave MIMO systems," *IEEE Signal Process. Lett.*, vol. 28, pp. 513–517, Feb. 2021.