Structure-Aware Sparse Bayesian Learning-Based Channel Estimation for Intelligent Reflecting Surface-Aided MIMO

He, Yanbin; Joseph, Geethu

DOI
10.1109/ICASSP49357.2023.10095932

Publication date
2023

Document Version
Final published version

Published in
Proceedings of the ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-6327-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICASSP49357.2023.10095932

STRUCTURE-AWARE SPARSE BAYESIAN LEARNING-BASED CHANNEL ESTIMATION FOR INTELLIGENT REFLECTING SURFACE-AIDED MIMO

Yanbin He and Geethu Joseph

Department of Microelectronics, Delft University of Technology, Delft, The Netherlands
Emails: {y.he-1, g.joseph}@tudelft.nl.

ABSTRACT

This paper presents novel cascaded channel estimation techniques for an intelligent reflecting surface-aided multiple-input multiple-output system. Motivated by the channel angular sparsity at higher frequency bands, the channel estimation problem is formulated as a sparse vector recovery problem with an inherent Kronecker structure. We solve the problem using the sparse Bayesian learning framework which leads to a non-convex optimization problem. We offer two solution techniques to the problem based on alternating minimization and singular value decomposition. Our simulation results illustrate the superior performance of our methods in terms of accuracy and run time compared with the existing works.

***Index Terms*—** Cascaded channel, Kronecker product, compressed sensing, structured sparsity, alternating minimization, singular value decomposition

1. INTRODUCTION

An intelligent reflecting surface (IRS) is a digitally controlled meta-surface containing a large number of passive reflecting elements. By reconfiguring the reflection coefficient of each element, IRS controls the wireless channel to improve the coverage and capacity of the communication system [1–4]. However, to enhance the channel properties via IRS, obtaining accurate channel state information is inevitable. Therefore, in this paper, we address the uplink channel estimation problem for an IRS-aided multi-input multi-output (MIMO) system by exploiting the intrinsic channel structure.

Related works: Early works on channel estimation for IRS-aided communication systems focused on unstructured channel models [5], employing least squares or linear minimum mean square error estimators [6]. However, in higher frequency bands (for example, millimeter wave or terahertz band) both mobile station (MS)-IRS and IRS-base stations (BS) channels exhibit strong sparsity in the angular domain [6]. This observation motivated the IRS-aided channel estimation algorithms to explore the intrinsic sparsity of the channel, reducing the pilot overhead [6]. Recent estimators further enhanced the accuracy by accounting for additional structures along with sparsity. Some examples are clustered sparsity structure in the angular domain [7] and joint sparsity in a multiuser setting [2, 8]. Most studies use orthogonal matching pursuit (OMP)-based methods. Despite low complexity, their heuristic nature leads to inferior channel estimation accuracy compared to other sparsity-driven approaches. An alternative approach is the iterative reweighted method-based sparse channel estimation [9], but it does not incorporate any additional signal structure. [10] presents a sparse Bayesian learning (SBL) scheme to handle the inherent Kronecker structure of the cascaded BS-IRS-MS sparse channel. However, the derivation of the SBL algorithm relied on several approximations leading to a suboptimal estimation accuracy [11]. Hence, we seek a novel channel estimator that exploits the Kronecker-sparse structure of the cascaded channel and offers improved estimation accuracy and complexity.

Contributions: Our contributions are two novel SBL channel estimation algorithms for an IRS-aided system:

- *Alternating minimization (AM)-based:* This method solves the underlying optimization problem of the SBL algorithm exactly using the AM procedure, inheriting the convergence property of the SBL algorithm.
- *Singular value decomposition (SVD)-based:* The second method uses a simple approximation to obtain the SBL algorithm. However, the resulting algorithm is faster and more accurate than the state-of-the-art Kronecker-SBL [10].

Overall, we derive two SBL-based channel estimators that exploit the Kronecker-sparse structure, leading to improved pilot overhead. The algorithms can be of independent interest because Kronecker-sparse structure naturally arises in a basis expansion problem with multiple unknown parameters.

2. CASCADED CHANNEL ESTIMATION PROBLEM

Consider an uplink MIMO millimeter-wave/terahertz band system with an MS with M antennas, a BS with B antennas, and a uniform linear array IRS with L elements. We assume that the line-of-sight (LOS) path between the BS and MS is blocked, and the LOS paths between the BS and IRS and

the IRS and MS are much stronger than the non-LOS paths. Let $\boldsymbol{H}_{\rm MS} \in \mathbb{C}^{L\times M}$ and $\boldsymbol{H}_{\rm BS} \in \mathbb{C}^{B\times L}$ denote the MS-IRS and IRS-BS channels, respectively. We assume a narrowband fading channel following the Saleh-Valenzuela model [7]:

$$\boldsymbol{H}_{\rm MS} = \sum_{p=1}^{P_{\rm MS}} \sqrt{\frac{LM}{P_{\rm MS}}} \beta_{{\rm MS},p} \boldsymbol{a}_L(\phi_{{\rm MS},p}) \boldsymbol{a}_M(\alpha_{\rm MS})^{\mathsf{H}} \quad (1)$$

$$\boldsymbol{H}_{\rm BS} = \sum_{p=1}^{P_{\rm BS}} \sqrt{\frac{BL}{P_{\rm BS}}} \beta_{{\rm BS},p} \boldsymbol{a}_B(\alpha_{{\rm BS},p}) \boldsymbol{a}_L(\phi_{\rm BS})^{\mathsf{H}}, \quad (2)$$

where $P_{\rm MS}$ and $P_{\rm BS}$ are the number of rays. Also, for any integer Q and angle ψ, steering vector $\boldsymbol{a}_Q(\psi) \in \mathbb{C}^{Q\times 1}$ is

$$\boldsymbol{a}_Q(\psi) = \frac{1}{\sqrt{Q}} \begin{bmatrix} 1 & e^{j2\pi\delta\cos\psi} \cdots e^{j2\pi(Q-1)\delta\cos\psi} \end{bmatrix}^{\mathsf{T}}. \quad (3)$$

Here, we assume half-wavelength spacing, i.e., $\delta = 0.5$. The angles $\phi_{{\rm MS},p}$, $\alpha_{\rm MS}$, $\alpha_{{\rm BS},p}$, and $\phi_{\rm BS}$ denote the p-th angle of arrival (AoA) of the IRS, and the angle of departure (AoD) of the MS, the p-th AoA of the BS, and the AoD of the IRS, respectively (see Fig. 1). Therefore, the cascaded MS-IRS-BS channel is given by $\boldsymbol{H}_{\rm BS}\,{\rm diag}(\boldsymbol{\theta})\boldsymbol{H}_{\rm MS}$ for the IRS configuration $\boldsymbol{\theta} \in \mathbb{C}^{L\times 1}$. Here, the i-th entry of $\boldsymbol{\theta}$ represents the gain and phase shift due to the i-th IRS element. We aim to estimate the cascaded channel for any $\boldsymbol{\theta}$ given by $\boldsymbol{H}_{\rm BS}\,{\rm diag}(\boldsymbol{\theta})\boldsymbol{H}_{\rm MS}$. This problem arises, for example, in beamforming problem to obtain the IRS configuration that maximizes the channel gain $\|\boldsymbol{H}_{\rm BS}\,{\rm diag}(\boldsymbol{\theta})\boldsymbol{H}_{\rm MS}\|_F^2$ [12]. However, ${\rm vec}(\boldsymbol{H}_{\rm BS}\,{\rm diag}(\boldsymbol{\theta})\boldsymbol{H}_{\rm MS}) = (\boldsymbol{H}_{\rm MS}^{\mathsf{T}} \odot \boldsymbol{H}_{\rm BS})\boldsymbol{\theta}$, where $\odot$ is the Khatri-Rao product. Therefore, the cascaded channel estimation is equivalent to estimating $\boldsymbol{H}_{\rm MS}^{\mathsf{T}} \odot \boldsymbol{H}_{\rm BS}$.

To estimate the channel, we send pilot symbols over K time slots over which $\boldsymbol{H}_{\rm MS}$ and $\boldsymbol{H}_{\rm BS}$ are assumed to be constant. We choose $K_{\rm I} < K$ IRS configurations, and for each configuration, we transmit pilot data $\boldsymbol{X} \in \mathbb{C}^{M\times K_{\rm P}}$ over $K_{\rm P}$ time slots such that $K = K_{\rm I}K_{\rm P}$. Hence, the received signal $\boldsymbol{Y}_k \in \mathbb{C}^{B\times K_{\rm P}}$ corresponding to the k-th configuration $\boldsymbol{\theta}_k$ is

$$\boldsymbol{Y}_k = \boldsymbol{H}_{\rm BS}\,{\rm diag}(\boldsymbol{\theta}_k)\boldsymbol{H}_{\rm MS}\boldsymbol{X} + \boldsymbol{W}_k, \quad (4)$$

where $\boldsymbol{W}_k \in \mathbb{C}^{B\times K_{\rm P}}$ is the additive white Gaussian noise with zero mean and known variance σ^2. Our objective is to estimate $\boldsymbol{H}_{\rm MS}^{\mathsf{T}} \odot \boldsymbol{H}_{\rm BS}$, using the data model in (4) and the knowledge of $\boldsymbol{X}$ and $\{\boldsymbol{Y}_k, \boldsymbol{\theta}_k\}_{k=1}^{K_{\rm I}}$. The estimation task is challenging because of (i) highly structured unknowns $\boldsymbol{H}_{\rm MS}$ and $\boldsymbol{H}_{\rm BS}$ entangled with the known quantities $\boldsymbol{\theta}_k$ and $\boldsymbol{X}$; and (ii) the scaling ambiguity due to the product form $\boldsymbol{H}_{\rm BS}\,{\rm diag}(\boldsymbol{\theta}_k)\boldsymbol{H}_{\rm MS}$. In particular, if an algorithm estimates $\boldsymbol{H}_{\rm MS}$ and $\boldsymbol{H}_{\rm BS}$ separately, it cannot distinguish two solutions $(\boldsymbol{H}_{\rm MS}, \boldsymbol{H}_{\rm BS})$ and $(1/q\boldsymbol{H}_{\rm MS}, q\boldsymbol{H}_{\rm BS})$, for any $q \neq 0$. The following section presents our estimation algorithm.

3. CHANNEL ESTIMATION ALGORITHMS

This section formulates the channel estimation task as a sparse recovery problem exploring angular sparsity and derives new Bayesian algorithms using the SBL framework.

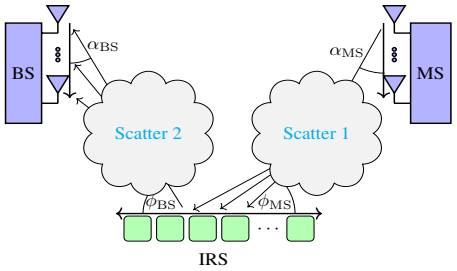

Fig. 1: An illustration of AoAs and AoDs in the uplink channel of an IRS-aided system.

3.1. Sparse Recovery Formulation

The first step to formulate the sparse recovery problem is to disentangle $\boldsymbol{\theta}_k$ from the unknowns $\boldsymbol{H}_{\rm MS}$ and $\boldsymbol{H}_{\rm BS}$. For this, we vectorize both sides of (4) to obtain

$$\bar{\boldsymbol{y}}_k = \big((\boldsymbol{X}^{\mathsf{T}}\boldsymbol{H}_{\rm MS}^{\mathsf{T}}) \odot \boldsymbol{H}_{\rm BS}\big)\boldsymbol{\theta}_k + \bar{\boldsymbol{w}}_k \in \mathbb{C}^{BK_{\rm P}\times 1}, \quad (5)$$

So, the received data $\bar{\boldsymbol{Y}} \in \mathbb{C}^{BK_{\rm P}\times K_{\rm I}}$ for the IRS configurations $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{K_{\rm I}}] \in \mathbb{C}^{L\times K_{\rm I}}$ is

$$\bar{\boldsymbol{Y}} = [\bar{\boldsymbol{y}}_1, \ldots, \bar{\boldsymbol{y}}_{K_{\rm I}}] = \big((\boldsymbol{X}^{\mathsf{T}}\boldsymbol{H}_{\rm MS}^{\mathsf{T}}) \odot \boldsymbol{H}_{\rm BS}\big)\boldsymbol{\Theta} + \bar{\boldsymbol{W}}, \quad (6)$$

where $\bar{\boldsymbol{W}} = [\bar{\boldsymbol{w}}_1, \ldots, \bar{\boldsymbol{w}}_{K_{\rm I}}] \in \mathbb{C}^{BK_{\rm P}\times K_{\rm I}}$. Next, we leverage angular sparsity in the channel matrices $\boldsymbol{H}_{\rm MS}$ and $\boldsymbol{H}_{\rm BS}$. For this, we apply the basis expansion model by discretizing the angular domain using a set of N grid angles $\{\psi_n\}_{n=1}^N$ such that $\cos(\psi_n) = 2n/N - 1$ [13]. Then, (1) and (2) reduce to

$$\boldsymbol{H}_{\rm BS} = \boldsymbol{A}_B \boldsymbol{g}_{\rm B} \boldsymbol{g}_{\rm L,d}^{\mathsf{H}} \boldsymbol{A}_L^{\mathsf{H}} \quad \text{and} \quad \boldsymbol{H}_{\rm MS} = \boldsymbol{A}_L \boldsymbol{g}_{\rm L,a} \boldsymbol{g}_{\rm M}^{\mathsf{H}} \boldsymbol{A}_M^{\mathsf{H}}, \quad (7)$$

where for any integer $Q > 0$, using (3), we define

$$\boldsymbol{A}_Q = \begin{bmatrix} \boldsymbol{a}_Q(\psi_1) & \boldsymbol{a}_Q(\psi_2) & \ldots & \boldsymbol{a}_Q(\psi_N) \end{bmatrix} \in \mathbb{C}^{Q\times N}. \quad (8)$$

Also, $\boldsymbol{g}_{\rm B}, \boldsymbol{g}_{\rm L,d}, \boldsymbol{g}_{\rm L,a}, \boldsymbol{g}_{\rm M} \in \mathbb{C}^{N\times 1}$ are the unknown sparse channel representations. The non-uniform grid points in the angular domain help to reduce the computational complexity of the estimation algorithm, which is discussed in (11). We also note that we use the same grid angle set $\{\psi_n\}_{n=1}^N$ for the AoAs and AoDs of the two channels for simplicity, but our framework and algorithm can also handle different grid sets.

Combining (7) and (6), and using the properties of the Khatri-Rao product [14, Lemma A1], we disentangle the unknown sprase vectors from the known matrices as follows:

$$\bar{\boldsymbol{Y}} = \big[(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{A}_M^*) \otimes \boldsymbol{A}_B\big]\big[(\boldsymbol{g}_{\rm L,a}\boldsymbol{g}_{\rm M}^{\mathsf{H}})^{\mathsf{T}} \otimes (\boldsymbol{g}_{\rm B}\boldsymbol{g}_{\rm L,d}^{\mathsf{H}})\big] \times (\boldsymbol{A}_L^{\mathsf{T}} \odot \boldsymbol{A}_L^{\mathsf{H}})\boldsymbol{\Theta} + \bar{\boldsymbol{W}}, \quad (9)$$

where $\otimes$ denotes the Kronecker product. Using Kronecker product's mixed-product property, we vectorize (9) to derive

$$\tilde{\boldsymbol{y}} = (\tilde{\boldsymbol{\Phi}}_{\rm L} \otimes \boldsymbol{\Phi}_{\rm M} \otimes \boldsymbol{\Phi}_{\rm B})(\boldsymbol{g}_{\rm L,a} \otimes \boldsymbol{g}_{\rm L,d} \otimes \boldsymbol{g}_{\rm M}^* \otimes \boldsymbol{g}_{\rm B}) + \tilde{\boldsymbol{w}}, \quad (10)$$

where $\tilde{\boldsymbol{\Phi}}_{\rm L} = \boldsymbol{\Theta}^{\mathsf{T}}(\boldsymbol{A}_L^{\mathsf{T}} \odot \boldsymbol{A}_L^{\mathsf{H}})^{\mathsf{T}}$, $\boldsymbol{\Phi}_{\rm M} = \boldsymbol{X}^{\mathsf{T}}\boldsymbol{A}_M^*$, and $\boldsymbol{\Phi}_{\rm B} = \boldsymbol{A}_B$. Further, we note that the only distinct columns of $\tilde{\boldsymbol{\Phi}}_{\rm L}$ are its first N columns [12]. Hence, removing the redundant columns to reduce the dimension of the representation, we get

$$\tilde{\boldsymbol{y}} = (\boldsymbol{\Phi}_{\rm L} \otimes \boldsymbol{\Phi}_{\rm M} \otimes \boldsymbol{\Phi}_{\rm B})\boldsymbol{g} + \tilde{\boldsymbol{w}} = \tilde{\boldsymbol{H}}\boldsymbol{g} + \tilde{\boldsymbol{w}} \in \mathbb{C}^{BK\times 1}, \quad (11)$$

where $\boldsymbol{\Phi}_{\rm L} \in \mathbb{C}^{K_{\rm I}\times N}$ is the submatrix formed by the first N columns of $\tilde{\boldsymbol{\Phi}}_{\rm L}$ and $\tilde{\boldsymbol{H}} = \boldsymbol{\Phi}_{\rm L} \otimes \boldsymbol{\Phi}_{\rm M} \otimes \boldsymbol{\Phi}_{\rm B} \in \mathbb{C}^{BK\times N^3}$. Also,

we define $\boldsymbol{g} = \boldsymbol{g}_{\mathrm{L}} \otimes \boldsymbol{g}_{\mathrm{M}}^{*} \otimes \boldsymbol{g}_{\mathrm{B}} \in \mathbb{C}^{N^3 \times 1}$ with $\boldsymbol{g}_{\mathrm{L}} \in \mathbb{C}^{N \times 1}$ being the scaled version of the first N entries of $\boldsymbol{g}_{\mathrm{L,a}} \otimes \boldsymbol{g}_{\mathrm{L,d}}$. Hence, (11) translates the channel estimation problem into a sparse vector recovery problem with unknown $\boldsymbol{g}$. Using $\boldsymbol{g}$ and (7), we obtain the product term in the channel as

$$\mathrm{vec}(\boldsymbol{H}_{\mathrm{MS}}^{\mathsf{T}} \odot \boldsymbol{H}_{\mathrm{BS}}) = (\boldsymbol{\Phi}_{\mathrm{A}} \otimes \boldsymbol{A}_{M}^{*} \otimes \boldsymbol{\Phi}_{\mathrm{B}})\boldsymbol{g}, \quad (12)$$

where $\boldsymbol{\Phi}_{\mathrm{A}}$ is the first N columns of $(\boldsymbol{A}_{L}^{\mathsf{T}} \odot \boldsymbol{A}_{L}^{\mathsf{H}})^{\mathsf{T}}$. Finally, the cascaded channel for a given IRS configuration $\boldsymbol{\theta}$ is computed as $(\boldsymbol{H}_{\mathrm{MS}}^{\mathsf{T}} \odot \boldsymbol{H}_{\mathrm{BS}})\boldsymbol{\theta}$. Thus, the rest of this section is devoted to derive an algorithm to estimate Kronecker-sparse $\boldsymbol{g}$ in (11).

3.2. Kronecker-Sparse Bayesian Learning Algorithms

Inspired by the SBL framework [15], we impose a fictitious sparsity promoting zero-mean Gaussian prior [15] (with unknown covariance) on the sparse vector $\boldsymbol{g}$. In our setting, to mimic the Kronecker structure, we construct the covariance matrix as $\mathrm{diag}(\otimes_{j=1}^{3}\boldsymbol{\gamma}_j)$ where the vectors $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3 \in \mathbb{R}^{N \times 1}$ are the unknown hyperparameters corresponding to the low-dimensional sparse vectors $\boldsymbol{g}_{\mathrm{L}}, \boldsymbol{g}_{\mathrm{M}}^{*}$, and $\boldsymbol{g}_{\mathrm{B}}$, respectively. Specifically, we assume

$$p(\boldsymbol{g}; \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3) = \mathcal{CN}(\boldsymbol{0}, \mathrm{diag}(\boldsymbol{\gamma})) \text{ with } \boldsymbol{\gamma} = \otimes_{j=1}^{3}\boldsymbol{\gamma}_j. \quad (13)$$

Then, we use Type-II maximum likelihood (ML) estimation, i.e, we first estimate the hyperparameters $\{\boldsymbol{\gamma}_j\}_{j=1}^{3}$, and using them, the estimate of $\boldsymbol{g}$ is the maximum point of $p(\boldsymbol{g}|\boldsymbol{y}; \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3)$. The ML estimates of $\{\boldsymbol{\gamma}_j\}_{j=1}^{3}$ are obtained by maximizing the likelihood $p(\tilde{\boldsymbol{y}}; \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3, \sigma^2)$ with respect to them. However, this maximization problem does not admit a closed form solution, and therefore, we resort to the Expectation-Maximization (EM) algorithm [15–17]. The EM algorithm iterates between the E-step that provides a lower bound of the log-likelihood and the M-step which maximizes the bound. Specifically, the r-th iteration of EM is

$$\textbf{E-step: } Q(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3 | \boldsymbol{\gamma}^{(r-1)}) = \mathbb{E}_{\boldsymbol{g}|\tilde{\boldsymbol{y}}; \boldsymbol{\gamma}^{(r-1)}}\{\log[p(\tilde{\boldsymbol{y}}, \boldsymbol{g}; \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3)]\}, \quad (14)$$

$$\textbf{M-step: } \{\boldsymbol{\gamma}_1^{(r)}, \boldsymbol{\gamma}_2^{(r)}, \boldsymbol{\gamma}_3^{(r)}\} = \underset{\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3}{\arg\max}\, Q(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3 | \boldsymbol{\gamma}^{(r-1)}), \quad (15)$$

where $\boldsymbol{\gamma}^{(r)} = \otimes_{j=1}^{3}\boldsymbol{\gamma}_j^{(r)}$ is the r-th iterate. Further, we have

$$p(\tilde{\boldsymbol{y}}, \boldsymbol{g}; \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3) \propto p(\tilde{\boldsymbol{y}}|\boldsymbol{g}) p(\boldsymbol{g}; \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3). \quad (16)$$

Thus, from (13), the M-step can be simplified as

$$\underset{\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3}{\arg\min} \log|\,\mathrm{diag}(\boldsymbol{\gamma})| + \boldsymbol{d}^{\mathsf{T}}\boldsymbol{\gamma}^{-1} \quad \text{s.t. } \boldsymbol{\gamma} = \otimes_{j=1}^{3}\boldsymbol{\gamma}_j, \quad (17)$$

where $(\cdot)^{-1}$ is the element-wise inversion, and we define $\boldsymbol{d} = \mathrm{diag}(\boldsymbol{\Sigma}_{\mathrm{g}} + \boldsymbol{\mu}_{\mathrm{g}}\boldsymbol{\mu}_{\mathrm{g}}^{\mathsf{H}})$. Here, $\boldsymbol{\mu}_{\mathrm{g}}$ and $\boldsymbol{\Sigma}_{\mathrm{g}}$ are the mean and variance of conditional distribution $p(\boldsymbol{g}|\tilde{\boldsymbol{y}}; \boldsymbol{\gamma}^{(r-1)})$:

$$\boldsymbol{\mu}_{\mathrm{g}} = \sigma^{-2}\boldsymbol{\Sigma}_{\mathrm{g}}\tilde{\boldsymbol{H}}^{\mathsf{H}}\tilde{\boldsymbol{y}},\ \boldsymbol{\Sigma}_{\mathrm{g}} = \left[\sigma^{-2}\tilde{\boldsymbol{H}}^{\mathsf{H}}\tilde{\boldsymbol{H}} + (\boldsymbol{\Gamma}^{(r-1)})^{-1}\right]^{-1}, \quad (18)$$

with $\boldsymbol{\Gamma}^{(r-1)} = \mathrm{diag}(\boldsymbol{\gamma}^{(r-1)})$. We present two novel ways to solve (17): AM-based and SVD-based, as discussed below.

AM-based Kronecker SBL (AM-KroSBL): The AM-KroSBL solves (17) by setting the gradient of the objective function with respect to the optimization variables to zero which gives

$$\boldsymbol{\gamma}_j = N^{-2}[(\otimes_{l=1}^{j-1}\boldsymbol{\gamma}_l^{-1}) \otimes \boldsymbol{I} \otimes (\otimes_{l=j+1}^{3}\boldsymbol{\gamma}_l^{-1})]^{\mathsf{T}}\boldsymbol{d}, \quad j = 1, 2, 3. \quad (19)$$

The AM-KroSBL alternatively updates $\boldsymbol{\gamma}_1$, $\boldsymbol{\gamma}_2$, and $\boldsymbol{\gamma}_3$ using (19) until converge. Also, to resolve the scaling ambiguity, i.e., $\boldsymbol{\gamma} = \otimes_{j=1}^{3}\boldsymbol{\gamma}_j = \alpha^{-1}\boldsymbol{\gamma}_1 \otimes \alpha\boldsymbol{\gamma}_2 \otimes \boldsymbol{\gamma}_3$, we normalize $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ to have unit norm. As estimating $\boldsymbol{g}$ only needs $\boldsymbol{\gamma}$, normalization causes no issue but reduces solution space. Since the M-step is solved exactly, the algorithm inherits the convergence property of the EM algorithm [18]. However, due to an (inner) iterative step in the M-step, AM-KroSBL is not computationally efficient. So, we next present a non-iterative method based on SVD.

SVD-based Kronecker SBL (SVD-KroSBL): In this method, we solve (17) without the constraint $\boldsymbol{\gamma} = \otimes_{j=1}^{3}\boldsymbol{\gamma}_j$ and then project the solution to the constraint set. Specifically, we have

$$\underset{\boldsymbol{\gamma}}{\arg\min} \log|\,\mathrm{diag}(\boldsymbol{\gamma})| + \boldsymbol{d}^{\mathsf{T}}\boldsymbol{\gamma}^{-1} = \boldsymbol{d}. \quad (20)$$

To project the solution to the constraint set, we solve for $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3$ that minimizes $\|\boldsymbol{d} - \otimes_{j=1}^{3}\boldsymbol{\gamma}_j\|$. We further approximate this optimization problem as two rank-1 approximations solved using SVD:

$$\underset{\substack{\boldsymbol{\gamma}_1, \tilde{\boldsymbol{\gamma}} \\ \|\boldsymbol{\gamma}_1\|=1}}{\arg\min} \|\boldsymbol{d} - \mathrm{vec}(\tilde{\boldsymbol{\gamma}}\boldsymbol{\gamma}_1^{\mathsf{T}})\|,\ \underset{\substack{\boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3 \\ \|\boldsymbol{\gamma}_2\|=1}}{\arg\min} \|\tilde{\boldsymbol{\gamma}} - \mathrm{vec}(\boldsymbol{\gamma}_3\boldsymbol{\gamma}_2^{\mathsf{T}})\|, \quad (21)$$

where we use the fact that $\mathrm{vec}(\tilde{\boldsymbol{\gamma}}\boldsymbol{\gamma}_1^{\mathsf{T}}) = \boldsymbol{\gamma}_1 \otimes \tilde{\boldsymbol{\gamma}}$ and the unit-norm constraints resolve the scaling ambiguity.

3.3. Complexity Reduction and Analysis

SBL is known to be computationally inefficient due to matrix inversion in (18). [11] introduced a technique to reduce the algorithm complexity using the Kronecker structure. We next present a novel technique to further improve the complexity, which can applied to both AM-KroSBL and SVD-KroSBL. Specifically, invoking the matrix inversion lemma and the mixed product property of the Kroncker product and using the definition $\tilde{\boldsymbol{H}} = \boldsymbol{\Phi}_{\mathrm{L}} \otimes \boldsymbol{\Phi}_{\mathrm{M}} \otimes \boldsymbol{\Phi}_{\mathrm{B}}$, we rewrite (18) as

$$\boldsymbol{\Sigma}_{\mathrm{g}} = \boldsymbol{\Gamma}^{(r-1)} - \boldsymbol{\Gamma}^{(r-1)}\tilde{\boldsymbol{H}}^{\mathsf{H}}(\sigma^2\boldsymbol{I} + (\boldsymbol{\Phi}_{\mathrm{L}}\boldsymbol{\Gamma}_1^{(r-1)}\boldsymbol{\Phi}_{\mathrm{L}}^{\mathsf{H}}) \otimes (\boldsymbol{\Phi}_{\mathrm{M}}\boldsymbol{\Gamma}_2^{(r-1)}\boldsymbol{\Phi}_{\mathrm{M}}^{\mathsf{H}}) \otimes (\boldsymbol{\Phi}_{\mathrm{B}}\boldsymbol{\Gamma}_3^{(r-1)}\boldsymbol{\Phi}_{\mathrm{B}}^{\mathsf{H}}))^{-1}\tilde{\boldsymbol{H}}\boldsymbol{\Gamma}^{(r-1)}, \quad (22)$$

where $\boldsymbol{\Gamma}_j^{(r-1)} = \mathrm{diag}(\boldsymbol{\gamma}_j^{(r-1)})$, for $j = 1, 2, 3$. Let the eigenvalue decomposition of the three matrices in (22) be $\boldsymbol{\Phi}_{\mathrm{L}}\boldsymbol{\Gamma}_1^{(r-1)}\boldsymbol{\Phi}_{\mathrm{L}}^{\mathsf{H}} = \boldsymbol{U}_1\boldsymbol{\Pi}_1\boldsymbol{U}_1^{\mathsf{H}}$, $\boldsymbol{\Phi}_{\mathrm{M}}\boldsymbol{\Gamma}_2^{(r-1)}\boldsymbol{\Phi}_{\mathrm{M}}^{\mathsf{H}} = \boldsymbol{U}_2\boldsymbol{\Pi}_2\boldsymbol{U}_2^{\mathsf{H}}$, and $\boldsymbol{\Phi}_{\mathrm{B}}\boldsymbol{\Gamma}_3^{(r-1)}\boldsymbol{\Phi}_{\mathrm{B}}^{\mathsf{H}} = \boldsymbol{U}_3\boldsymbol{\Pi}_3\boldsymbol{U}_3^{\mathsf{H}}$. Then, we derive

$$\boldsymbol{\Sigma}_{\mathrm{g}} = \boldsymbol{\Gamma}^{(r-1)}(\boldsymbol{I} - \tilde{\boldsymbol{H}}^{\mathsf{H}}\boldsymbol{U}(\sigma^2\boldsymbol{I} + \boldsymbol{\Pi})^{-1}\boldsymbol{U}^{\mathsf{H}}\tilde{\boldsymbol{H}}\boldsymbol{\Gamma}^{(r-1)}), \quad (23)$$

where $\boldsymbol{U} = \otimes_{j=1}^{3}\boldsymbol{U}_j$, and $\boldsymbol{\Pi} = \otimes_{j=1}^{3}\boldsymbol{\Pi}_j$. Combining (18) and (23), the posterior mean $\boldsymbol{\mu}_{\mathrm{g}}$ is

$$\boldsymbol{\mu}_{\mathrm{g}} = \sigma^{-2}(\boldsymbol{\Gamma}^{(r-1)}\tilde{\boldsymbol{H}}^{\mathsf{H}}\tilde{\boldsymbol{y}} - \boldsymbol{\Gamma}^{(r-1)}\tilde{\boldsymbol{H}}^{\mathsf{H}}\boldsymbol{U}(\sigma^2\boldsymbol{I} + \boldsymbol{\Pi})^{-1}\boldsymbol{U}^{\mathsf{H}}\tilde{\boldsymbol{H}}\boldsymbol{\Gamma}^{(r-1)}\tilde{\boldsymbol{H}}^{\mathsf{H}}\tilde{\boldsymbol{y}}). \quad (24)$$

The overall complexity of the different algorithms are summarized in Table 1. Here, $R_{\rm EM}$ is the number of EM iterations that varies across algorithms, and $R_{\rm AM}$ is the number of alternating iterations in (19). We see that the proposed schemes have lower complexity compared to the KroSBL.

In short, we present two SBL algorithms which differ from the state-of-the-art KroSBL [10, 11] in two ways: the new complexity reduction technique in the E-step and the AM and SVD based-solutions in the M-step. Comparing our two schemes, AM-KroSBL enjoys the convergence guarantee while SVD-KroSBL is more efficient in practice (see Fig. 2).

Table 1: Time complexity of different versions of KroSBL

Method	Complexity
AM-KroSBL	$\mathcal{O}\big(R_{\rm EM}(R_{\rm AM}N^3 + N^3KB)\big)$
SVD-KroSBL	$\mathcal{O}\big(R_{\rm EM}(N^4 + N^3KB)\big)$
KroSBL in [10, 11]	$\mathcal{O}\big(R_{\rm EM}N^6\big)$

Algorithm 1: Our KroSBL Channel Estimation

Data: Received signal $\tilde{\boldsymbol{y}}$, Pilot signal $\boldsymbol{X}$, IRS configuration $\boldsymbol{\Theta}$, noise power σ^2

Parameters: Threshold ϵ and iterations $R_{\max}$

Initialization: $\boldsymbol{\gamma}_1^{(-1)} = \boldsymbol{\gamma}_2^{(-1)} = \boldsymbol{\gamma}_3^{(-1)} = \mathbf{1}$; $\boldsymbol{\mu}_{\rm g}^{(0)} = \mathbf{0}$, $\boldsymbol{\mu}_{\rm g}^{(-1)} = \mathbf{1}$; and $r = 0$.

while $\|\boldsymbol{\mu}_{\rm g}^{(r)} - \boldsymbol{\mu}_{\rm g}^{(r-1)}\|_2 > \epsilon$ and $r < R_{\max}$ **do**

  Define $\boldsymbol{\Gamma}_j^{(r-1)} = \mathrm{diag}(\boldsymbol{\gamma}_j^{(r-1)}), \forall j = 1, 2, 3$

  Compute $\boldsymbol{\Sigma}_{\rm g}$ and $\boldsymbol{\mu}_{\rm g}$ using (23) and (24)

  Obtain $\{\boldsymbol{\gamma}_j^{(r)}\}_{j=1}^3$ either using (19) and normalizing $\{\boldsymbol{\gamma}_j^{(r)}\}_{j=1}^2$ (AM-KroSBL), or using (21) (SVD-KroSBL).

  Update iteration number $r \leftarrow r + 1$

end

Compute $\boldsymbol{H}_{\rm MS}^{\mathsf{T}} \odot \boldsymbol{H}_{\rm BS}$ from (12) with $\boldsymbol{g} = \boldsymbol{\mu}_{\rm g}^{(r)}$

Result: Cascaded channel function $(\boldsymbol{H}_{\rm MS}^{\mathsf{T}} \odot \boldsymbol{H}_{\rm BS})\boldsymbol{\theta}$

4. NUMERICAL SIMULATION

Our simulation setting is as follows. We choose $B = 16$ BS antennas, $M = 6$ MS antennas, and $L = 256$ IRS elements. Each entry of the IRS configurations $\{\boldsymbol{\theta}_k\}_{k=1}^{K_{\rm I}}$ is uniformly drawn from $\{-1/\sqrt{N}, 1/\sqrt{N}\}$ with $K_{\rm I} = 4, 10$. For each IRS configuration, we send $K_{\rm P} = 6$ pilot signals. The number of grid angles is $N = 18$ and all AoDs/AoAs are drawn uniformly from the grid angles. Further, the channel gains $\{\beta_{{\rm BS},p}\}_{p=1}^{P_{\rm B}}$ and $\{\beta_{{\rm MS},p}\}_{p=1}^{P_{\rm M}}$ in (1) and (2) are drawn from $\mathcal{CN}(0, 1)$ [19]. We use four performance metrics: normalized mean squared error (NMSE), support recovery rate (SRR), run time, and symbol error rate (SER). Here, NMSE is given as $\frac{1}{K_{\rm I}}\sum_{k=1}^{K_{\rm I}} \frac{\|\boldsymbol{H}_{\rm BS}\,\mathrm{diag}(\boldsymbol{\theta}_k)\boldsymbol{H}_{\rm MS} - \tilde{\boldsymbol{H}}_{\rm BS}\,\mathrm{diag}(\boldsymbol{\theta}_k)\tilde{\boldsymbol{H}}_{\rm MS}\|_F^2}{\|\boldsymbol{H}_{\rm BS}\,\mathrm{diag}(\boldsymbol{\theta}_k)\boldsymbol{H}_{\rm MS}\|_F^2}$, with $\tilde{\boldsymbol{H}}_{\rm BS}\,\mathrm{diag}(\boldsymbol{\theta}_k)\tilde{\boldsymbol{H}}_{\rm MS}$ being the reconstructed channel, and SRR is $\frac{|\mathrm{supp}(\tilde{\boldsymbol{g}}) \cap \mathrm{supp}(\boldsymbol{g})|}{|\mathrm{supp}(\tilde{\boldsymbol{g}}) - \mathrm{supp}(\boldsymbol{g})| + |\mathrm{supp}(\boldsymbol{g})|}$, with $\tilde{\boldsymbol{g}}$ being the estimate of $\boldsymbol{g}$ and $\mathrm{supp}(\cdot)$ representing the set of indices of the nonzero entries of a vector. We use the classic SBL [15], OMP, and the KroSBL in [11] as benchmarks.

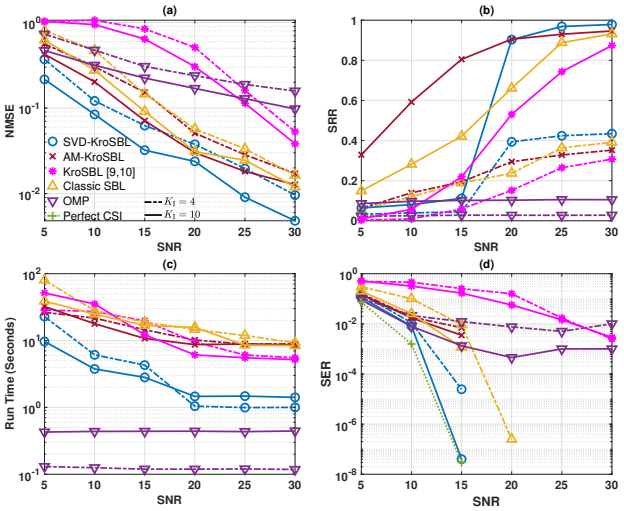

Fig. 2: Comparison of NMSE, SRR, run time, and SER of AM-KroSBL and SVD-KroSBL with the competing schemes as a function SNR when $K_{\rm I} = 4, 10$, $K_{\rm P} = 6$ and $N = 18$.

From Fig. 2, we observe that both SVD-KroSBL and AM-KroSBL outperform the other schemes in terms of NMSE, SRR and SER. Especially in the low SNR and low overhead (quantified by $K_{\rm I}$) regimes, our algorithms have the best NMSE. From Fig. 2 (b), in the low SNR regime, the AM-KroSBL has the best SRR, while the SVD-KroSBL has the optimal performance in the high SNR case for both low and high overhead cases. We observe that in the low SNR regime, SVD-KroSBL outputs a sparse vector with many small terms, leading to a low SRR. But since the sparse vector is dominated by large values on the correct support, NMSE is still low. Further, Fig. 2 (c) indicates that SVD-KroSBL, compared with other SBL-based methods, has one order less run time. The high run time of AM-KroSBL is expected due to the inner loop in the M-step, yet its run time is comparable to the classic SBL and KroSBL but with better NMSE. Finally, Fig. 2 (d) shows that when $K_{\rm I} = 4$, our schemes possess better SER than others. In contrast, when $K_{\rm I} = 10$, only SVD-KroSBL has lower SER compared to the existing schemes and approaches the oracle scheme with perfect CSI.

5. CONCLUSION

In this paper, we studied the channel estimation for IRS-aided MIMO system, exploiting the Kronecker sparse structure in the angular domain. We presented two novel SBL-based channel estimation methods with superior performance over the state-of-the-art methods. Our AM-KroSBL enjoys strong convergence guarantees while the SVD-KroSBL is suitable for practical applications owing to its low run time. Handling off-grid angle mismatch and extending our algorithms to multi-user case are interesting avenues for future work.

6. REFERENCES

[1] Qingqing Wu, Shuowen Zhang, Beixiong Zheng, Changsheng You, and Rui Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313–3351, May 2021.

[2] Xiuhong Wei, Decai Shen, and Linglong Dai, "Channel estimation for RIS assisted wireless communications—part I: Fundamentals, solutions, and future opportunities," *IEEE Commun. Lett.*, vol. 25, no. 5, pp. 1398–1402, May 2021.

[3] Emil Björnson, Henk Wymeersch, Bho Matthiesen, Petar Popovski, Luca Sanguinetti, and Elisabeth de Carvalho, "Reconfigurable intelligent surfaces: A signal processing perspective with wireless applications," *IEEE Signal Process. Mag.*, vol. 39, no. 2, pp. 135–158, Mar. 2022.

[4] Chu Li, Aydin Sezgin, and Zhu Han, "On the Impact of Oscillator Phase Noise in an Irs-assisted MISO TDD System," in *Proc. Int. ITG Workshop Smart Antennas (WSA)*. VDE, 2021, pp. 1–6.

[5] A Lee Swindlehurst, Gui Zhou, Rang Liu, Cunhua Pan, and Ming Li, "Channel estimation with Reconfigurable Intelligent Surfaces–A general framework," *Proc. IEEE*, Sept. 2022.

[6] Beixiong Zheng, Changsheng You, Weidong Mei, and Rui Zhang, "A survey on channel estimation and practical passive beamforming design for intelligent reflecting surface aided wireless communications," *IEEE Commun. Surv. Tutor.*, vol. 24, no. 2, pp. 1035–1071, 2022.

[7] You You, Li Zhang, Minhua Yang, Yongming Huang, Xiaohu You, and Chuan Zhang, "Structured OMP for IRS-Assisted mmWave channel estimation by exploiting angular spread," *IEEE Trans. Veh. Technol.*, vol. 71, no. 4, pp. 4444–4448, Apr. 2022.

[8] Gui Zhou, Cunhua Pan, Hong Ren, Petar Popovski, and A Lee Swindlehurst, "Channel estimation for RIS-aided multiuser millimeter-wave systems," *IEEE Trans. Signal Process.*, vol. 70, pp. 1478–1492, 2022.

[9] Jiguang He, Markus Leinonen, Henk Wymeersch, and Markku Juntti, "Channel estimation for RIS-aided mmWave MIMO systems," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*. IEEE, 2020, pp. 1–6.

[10] Xiaowen Xu, Shun Zhang, Feifei Gao, and Jiangzhou Wang, "Sparse Bayesian learning based channel extrapolation for RIS assisted MIMO-OFDM," *IEEE Trans. Commun.*, vol. 70, no. 8, pp. 5498–5513, Aug. 2022.

[11] Wen-Che Chang and Yu 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.

[12] Peilan Wang, Jun Fang, Huiping Duan, and Hongbin Li, "Compressed channel estimation for intelligent reflecting surface-assisted millimeter wave systems," *IEEE Signal Process. Lett.*, vol. 27, pp. 905–909, 2020.

[13] Zhendong Mao, Xiqing Liu, and Mugen Peng, "Channel estimation for intelligent reflecting surface assisted massive MIMO systems—A deep learning approach," *IEEE Commun. Lett.*, vol. 26, no. 4, pp. 798–802, 2022.

[14] C Radhakrishna Rao, "Estimation of heteroscedastic variances in linear models," *J. Am. Stat. Assoc.*, vol. 65, no. 329, pp. 161–172, 1970.

[15] David P Wipf and Bhaskar D Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.

[16] Lu Wang, Lifan Zhao, Susanto Rahardja, and Guoan Bi, "Alternative to extended block sparse Bayesian learning and its relation to pattern-coupled sparse Bayesian learning," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2759–2771, May 2018.

[17] Zhilin Zhang and Bhaskar D Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE J. Sel. Top. Signal Process.*, vol. 5, no. 5, pp. 912–926, Sept. 2011.

[18] Igor Fedorov, *Structured Learning with Scale Mixture Priors*, Ph.D. Dissertation, ECE Dept., UC San Diego, 2018.

[19] Yuxing Lin, Shi Jin, Michail Matthaiou, and Xiaohu 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.