Joint Sparsity and Low-Rank Minimization for Reconfigurable Intelligent Surface-Assisted Channel Estimation

Jie Tang, *Senior Member, IEEE*, Xiaoyu Du, Zhen Chen, *Senior Member, IEEE*, Xiuyin Zhang, *Fellow, IEEE*, Daniel Ka Chun So, *Senior Member, IEEE*, Kai-Kit Wong, *Fellow, IEEE*, and Jonathon Chambers, *Fellow, IEEE*

***Abstract*—Reconfigurable intelligent surfaces (RISs) have attracted extensive attention in millimeter wave (mmWave) systems because of the capability of configuring the wireless propagation environment. However, due to the existence of a RIS between the transmitter and receiver, a large number of channel coefficients need to be estimated, resulting in more pilot overhead. In this paper, we propose a joint sparse and low-rank based two-stage channel estimation scheme for RIS-assisted mmWave systems. Specifically, we first establish a low-rank approximation model against the noisy channel, fitting in with the precondition of the compressed sensing theory for perfect signal recovery. To overcome the difficulty of solving the low-rank problem, we propose a trace operator to replace the traditional nuclear norm operator, which can better approximate the rank of a matrix. Furthermore, by utilizing the sparse characteristics of the mmWave channel, sparse recovery is carried out to estimate the RIS-assisted channel in the second stage. Simulation results show that the proposed scheme achieves significant performance gain in terms of estimation accuracy compared to the benchmark schemes.**

***Index Terms*—Channel estimation, reconfigurable intelligent surface, millimeter wave, compressed sensing, sparse and low-rank.**

This work has been supported in part by National Key Research and Development Program of China under Grant 2019YFB1804100, in part by the National Natural Science Foundation of China under Grant 61971194, 62222105, 62001171, in part by the Special Project for Guangxi Science and Technology Bases and Talents under Grant AD21075054, in part by the Open Research Fund of Guangdong Key Laboratory of Aerospace Communication and Networking Technology under Grant 2018B030322004, in part by the International Science and Technology Cooperation Project of Guangzhou (Huangpu) under Grant 2020GH06, in part by Nanning Innovation and Entrepreneurship Leading Talents Yongjiang Plan Funded Project under Grant 2020003, in part by the Natural Science Foundation of Guangdong Province under Grant 2021A1515011966, 2022A1515011189, in part by the Open Research Fund of State Key Laboratory of Integrated Services Networks, Xidian University under Grant ISN23-05, in part by the Open Research Fund of State Key Laboratory of State Key Laboratory of Millimeter Waves, Southeast University under Grant K202411, and in part by the Engineering and Physical Sciences Research Council (EPSRC) under grant EP/V052942/1. (*Corresponding author: Zhen Chen.*)

J. Tang, X. Du, Z. Chen, and X. Zhang are with the School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641, China (e-mail: eejtang@scut.edu.cn; 202120112354@mail.scut.edu.cn; chenz.scut@gmail.com; zhangxiuyin@scut.edu.cn).

D. K. C. So is with the School of Electrical and Electronic Engineering, University of Manchester, Manchester M13 9PL, U.K. (e-mail: d.so@manchester.ac.uk).

K.-K Wong is with the Department of Electronic and Electrical Engineering, University College London, London WC1E 6BT, U.K. (e-mail: kai-kit.wong@ucl.ac.uk).

Jonathon A. Chambers is with the Department of Engineering, University of Leicester, LE1 7RH Leicester, U.K. (e-mail: jonathon.chambers@ncl.ac.uk).

I. Introduction

MILLIMETER wave (mmWave) is regarded as a potential technology for 6th generation (6G) wireless communication to deal with increasingly scarce spectrum resources [1], [2]. However, due to the high operating frequency in the range of 30∼300 GHz, mmWave has the hidden danger of severe path loss and blockages, which limits its application in practical urban cellular systems [3]. A reconfigurable intelligent surface (RIS), with the ability to flexibly configure the wireless transmission environment, has emerged as a promising solution to cope with blockages in mmWave systems. RIS is a planar array composed of numerous passive reflecting elements, which can intelligently reflect incident signals to the desired direction with an adjustable phase shift [4]. In addition, RIS only reflects the incident signal without amplification and decoding, and hence there is no need for additional power consumption. Nevertheless, the promising benefits brought by RIS critically depend on the acquisition of channel state information (CSI), which is a practical challenge due to the following two main reasons [5]. Firstly, the reflecting elements are generally passive, and there is a lack of signal processing capabilities, making traditional transmission training sequence methods inapplicable [6], [7]. Secondly, due to the numerous reflecting elements deployed on a RIS, it is necessary to estimate the large-scale channel matrices, leading to a more complicated estimation process [8], [9].

To overcome the above challenges in RIS-assisted systems, much research has attempted to design excellent estimation algorithms via various signal processing techniques. The classical least squares (LS) and minimum mean squared error (MMSE) methods are frequently used in traditional pilot-based channel estimation due to their low complexity for implementation. Recently, different training protocols have been designed to apply LS and MMSE in RIS-assisted systems [10]–[14]. An LS-based dual-link (BS-RIS-BS) scheme employed pilot signals transmitted from the base station (BS) and subsequently reflected by the RIS to solve the common BS-RIS channel [13]. After obtaining the high-dimensional quasi-static BS-RIS channel, the RIS-user channel could easily be estimated. Inspired by this, the work in [14] designed an anchor-assisted estimation scheme, which employed two dedicated anchor nodes near RIS to estimate the common BS-RIS channel. Based on the obtained BS-RIS channel, the

dynamic RIS-user channel was further estimated effectively via the LS method. In addition to the widely used traditional methods, the booming development of deep learning points to a new direction for channel estimation. As a potent tool to solve nonlinear mapping problems, deep learning technology can learn the approximate mapping function from training data to CSI [15]–[19]. An approximate MMSE solution was obtained in [15] by employing a de-noising convolutional neural network (CNN) to deal with the noisy LS channel matrix. The work in [16] first activated only a part of the RIS elements and estimated the corresponding cascaded channels. The deep neural network (DNN) was then employed to predict other parts based on the obtained results. Furthermore, a new strategy based on matrix decomposition has emerged strongly, which decomposes high dimensional cascaded channels into a set of low dimensional subchannels that are easy to estimate [20]–[23]. By decomposing the three-dimensional cascaded channel into two-dimensional user-RIS and RIS-BS channels, the respective channels were estimated efficiently via parallel factor tensor modeling [20]. The work in [21] modeled the RIS-assisted channel as the keyhole channel and then proposed a singular value decomposition (SVD)-based scheme to estimate a set of rank-one matrices, each of which corresponds to a RIS element.

For RIS-assisted systems operating at the mmWave frequency band, severe path loss and blockages result in a limited number of paths, making the channel exhibit sparse characteristics. Compressed sensing, with the ability to sense the sparsity of channels, has emerged as a powerful method for RIS channel estimation. For example, by finding the sparse representation of cascaded channels through the properties of Khatri-Rao and Kronecker products, the channel estimation was transformed into a sparse signal recovery problem and could be solved efficiently by compressed sensing methods [24]. This work was further advanced in [25] by expanding the mobile station from a single antenna to multiple antennas. Specifically, a two-stage channel estimation scheme was designed, and the atomic norm minimization method was used to estimate the channel parameters sequentially. Similarly, based on the common sparse characteristics shared by all subcarrier channels, the classical orthogonal matching pursuit (OMP) method was employed to solve the channel estimation problem [26]. In addition, other compressed sensing methods such as manifold optimization [27], adaptive grid matching pursuit [28], and iteratively reweighted method [29] have also been employed to cope with RIS channel estimation in mmWave systems. However, several urban environment measurement results reveal that the mmWave channel not only has sparse scattering characteristics but also shows angular spreads of path clusters over the angle-of-departure (AoD) and the angle-of-arrival (AoA) domains [30]–[32]. As demonstrated in [33], the angular spread gives rise to a structured sparsity pattern. Inspired by this, the mmWave channel was further proved to exhibit low-rank characteristics in the presence of angular spread, which can be exploited to improve estimation performance [34].

In this paper, the joint sparse and low-rank characteristics are exploited for channel estimation in the RIS-assisted mmWave system. Based on joint characteristics, we first reconstruct the noisy received signal by low-rank matrix approximation before sparse signal recovery, which is the noise-free requirement of the compressed sensing theory for perfect channel recovery [34]. Specifically, considering the nuclear norm as the convex relaxation of rank function would lead to the results deviating from the true underlying ones [35], [36], an improved trace operator is applied for the low-rank stage to overcome this bias and thus improve the estimation accuracy. The main contributions of this paper are summarized as follows:

- We propose a RIS-assisted massive multiple-input multiple-output (MIMO) framework that combines sparsity and low-rank minimization for channel estimation by leveraging the spatial sparse structure. To be specific, a trace-based rank model is adopted, which is regarded as a low-rank approximation. Due to the powerful recognition of the trace norm and the robust sparse representation abilities of the L1 norm, a joint trace and L1 norm minimization channel estimation problem is formulated to achieve a performance improvement.
- To solve the channel estimation problem, a two-stage algorithm is proposed for RIS-assisted channel estimation. Different from the existing optimization schemes, we propose a trace operator with constantly adjusting coefficients as the cost function to replace the traditional nuclear norm operator, which can better approximate the rank of a matrix. This operator corrects the deviation caused by conventional operators and provides a new choice for the solution of low-rank minimization.
- Extensive simulation results and comparisons are performed to verify the efficiency and robustness of the proposed schemes. Specifically, the proposed joint channel estimator outperforms the conventional schemes in terms of the mean square error (MSE), error vector magnitude (EVM), bit error rate (BER), capacity and success ratio .

The remaining sections of this paper are structured as follows. In Section II, we establish the system model and describe how the conventional sparse recovery scheme works. The mmWave channel model with angular spread is discussed in Section III, where its joint sparse and low-rank characteristics are further elaborated. Based on these joint characteristics, a two-stage channel estimation scheme utilizing an improved rank operator is designed in Section IV, which consists of low-rank matrix approximation and sparse signal recovery. Finally, we reveal the results of our simulation in Section V and draw the conclusion in Section VI.

Notations: Uppercase bold letters represent matrices, and lowercase bold letters represent vectors. For any general matrix $\mathbf{A}$, $\mathbf{A}^T$, $\mathbf{A}^*$ and $\mathbf{A}^H$ represent the transpose, conjugate and conjugate transpose, respectively; $\mathrm{rank}(\mathbf{A})$ and $\mathrm{vec}(\mathbf{A})$ denote rank and vectorization. $\|\mathbf{A}\|_F$, $\|\mathbf{A}\|_1$ and $\|\mathbf{A}\|_2$ are the Frobenius norm, L1 norm and L2 norm of matrix $\mathbf{A}$. For a square matrix $\mathbf{B}$, we set $\mathrm{tr}(\mathbf{B})$ to indicate the trace. $j = \sqrt{-1}$ indicates an imaginary unit. $\mathbf{A} \otimes \mathbf{B}$ represents the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$. $\mathrm{diag}(x_1, \cdots, x_N)$ is the diagonal matrix with $\{x_1, \cdots, x_N\}$ as diagonal elements, $\mathbb{C}^{m \times n}$ means that

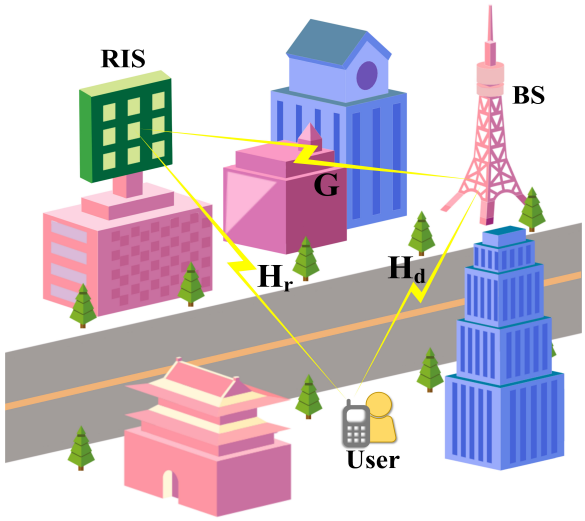

Fig. 1: The RIS-assisted uplink mmWave MIMO system.

the size of the matrix is $m \times n$, and $\boldsymbol{I}$ denotes the identity matrix.

II. System Model

Consider a RIS-assisted uplink mmWave MIMO system (see Fig. 1), which consists of one multi-antenna user, one multi-antenna BS, and one RIS. Suppose there are N_{BS} and N_{US} antennas at the BS and user, respectively, and the RIS is a uniform linear array (ULA) with M reflecting elements. Compared with the conventional scheme, the system introduces the phase shift of RIS to control the transceiver beamforming. At a certain moment t, the transmitter sends a symbol $s(t)$ with a beamforming vector $\boldsymbol{f}(t) \in \mathbb{C}^{N_{US}}$, which reaches the receiver end through the direct channel of user-BS and the cascaded channel of user-RIS-BS. Employing a receive combining vector $\boldsymbol{z}(t) \in \mathbb{C}^{N_{BS}}$, the signals from all antennas are combined together, and the final signal $y(t)$ can be written as

$$y(t)=\boldsymbol{z}^H(t)(\boldsymbol{H}_d+\boldsymbol{G}\boldsymbol{\Theta}\boldsymbol{H}_r)\boldsymbol{f}(t)s(t)+\omega(t), \forall t = 1,\ldots,T, \tag{1}$$

where $\boldsymbol{H}_d \in \mathbb{C}^{N_{BS}\times N_{US}}$, $\boldsymbol{H}_r \in \mathbb{C}^{M\times N_{US}}$, and $\boldsymbol{G} \in \mathbb{C}^{N_{BS}\times M}$ represent user-BS, user-RIS, and RIS-BS channels, respectively. $\boldsymbol{\Theta} = \mathrm{diag}(\beta_1 e^{j\theta_1}, \cdots, \beta_N e^{j\theta_N})$ is the reflection coefficient matrix of the RIS, and $\omega(t) \sim \mathcal{CN}\left(0, \sigma_n^2\right)$ denotes the additive white Gaussian noise.

In a conventional RIS-free mmWave system, since the receiver does not have access to a clean version of the channel matrix $\boldsymbol{H}$, we can only get the noisy version with $\boldsymbol{z}^H\boldsymbol{H}\boldsymbol{f}$. This issue, known as channel subspace sampling restriction, complicates channel estimation [37]. In addition, the introduction of RIS results in more complex channels and larger-scale coefficient matrices, making channel estimation even more complicated. However, the emergence of compressed sensing technology provides us with an effective solution. By utilizing the sparsity of mmWave channels, channel estimation can be expressed as a sparse recovery problem.

Typically, mmWave channels are characterized by the geometric channel model. Taking the direct channel $\boldsymbol{H}_d$ as an example, the model can be expressed as [38]

$$\boldsymbol{H}_d = \sum_{l=1}^{L} \alpha_l \boldsymbol{a}_{dA}(\theta_l)\boldsymbol{a}_{dD}^H(\phi_l), \tag{2}$$

where l represents the l-th path, α_l denotes the relative complex gain, and the total number of paths is L. $\theta_l \in [0, 2\pi]$ and $\phi_l \in [0, 2\pi]$ denote the associated azimuth AoA/AoD. In addition, $\boldsymbol{a}_{dA}(\cdot) \in \mathbb{C}^{N_A}$ and $\boldsymbol{a}_{dD}(\cdot) \in \mathbb{C}^{N_D}$ are array response vectors of the receiver and the transmitter, which can be expressed as

$$\boldsymbol{a}_{dA}(\theta)=\frac{1}{\sqrt{N_A}}[1,e^{j\frac{2\pi}{\lambda}d\sin(\theta)},\cdots,e^{j(N_A-1)\frac{2\pi}{\lambda}d\sin(\theta)}]^T, \tag{3}$$

$$\boldsymbol{a}_{dD}(\phi)=\frac{1}{\sqrt{N_D}}[1,e^{j\frac{2\pi}{\lambda}d\sin(\phi)},\cdots,e^{j(N_D-1)\frac{2\pi}{\lambda}d\sin(\phi)}]^T, \tag{4}$$

where λ denotes the wavelength of the signal, and d denotes the antenna spacing equal to half the wavelength. N_A and N_D are the number of antennas corresponding to AoA and AoD, respectively.

To express the channel estimation problem as a form of sparse recovery, the geometric channel model needs to be transformed into a more compact beamspace MIMO form [39]:

$$\boldsymbol{H}_d = \boldsymbol{A}_{dA}\boldsymbol{H}_v\boldsymbol{A}_{dD}^H, \tag{5}$$

where $\boldsymbol{A}_{dA} \triangleq [\boldsymbol{a}_{dA}(\theta_1), \cdots, \boldsymbol{a}_{dA}(\theta_{N_1})]$ with $(N_1 \geq N_A)$ and $\boldsymbol{A}_{dD} \triangleq [\boldsymbol{a}_{dD}(\phi_1), \cdots, \boldsymbol{a}_{dD}(\phi_{N_2})]$ with $(N_2 \geq N_D)$ are overcomplete matrices corresponding to the steering vectors of pre-discretized AoA and AoD, respectively. $\boldsymbol{H}_v \in \mathbb{C}^{N_1\times N_2}$ denotes a sparse matrix with L nonzero terms that are associated with the channel path gain $\{\alpha_l\}$.

Similarly, the user-RIS and RIS-BS channels can be expressed as

$$\boldsymbol{G} = \boldsymbol{A}_{GA}\boldsymbol{H}_{Gv}\boldsymbol{A}_{GD}^H, \tag{6}$$

$$\boldsymbol{H}_r = \boldsymbol{A}_{rA}\boldsymbol{H}_{rv}\boldsymbol{A}_{rD}^H, \tag{7}$$

$$\begin{aligned} y(t)&=\boldsymbol{z}^H(t)\boldsymbol{H}_d\boldsymbol{f}(t)+\boldsymbol{z}^H(t)\boldsymbol{G}\boldsymbol{\Theta}\boldsymbol{H}_r\boldsymbol{f}(t)+\omega(t)\\ &=\left(\boldsymbol{f}^T(t)\otimes\boldsymbol{z}^H(t)\right)\mathrm{vec}(\boldsymbol{H}_d)+\left(\boldsymbol{f}^T(t)\otimes\boldsymbol{z}^H(t)\right)\mathrm{vec}(\boldsymbol{G}\boldsymbol{\Theta}\boldsymbol{H}_r)+\omega(t)\\ &=\left(\boldsymbol{f}^T(t)\otimes\boldsymbol{z}^H(t)\right)\left[\mathrm{vec}(\boldsymbol{A}_{dA}\boldsymbol{H}_{dv}\boldsymbol{A}_{dD}^H)+\mathrm{vec}(\boldsymbol{A}_{GA}(\boldsymbol{H}_{Gv}\boldsymbol{A}_{GD}^H\boldsymbol{\Theta}\boldsymbol{A}_{rA}\boldsymbol{H}_{rv})\boldsymbol{A}_{rD}^H)\right]+\omega(t)\\ &=\left(\boldsymbol{f}^T(t)\otimes\boldsymbol{z}^H(t)\right)\left[(\boldsymbol{A}_{dD}^*\otimes\boldsymbol{A}_{dA})\,\mathrm{vec}(\boldsymbol{H}_{dv})+(\boldsymbol{A}_{rD}^*\otimes\boldsymbol{A}_{GA})\mathrm{vec}(\boldsymbol{H}_{cv})\right]+\omega(t), \end{aligned} \tag{8}$$

where $\boldsymbol{H}_{Gv}$ and $\boldsymbol{H}_{rv}$ denote the sparse matrix corresponding to user-RIS and RIS-BS channels, $\boldsymbol{A}_{GA}$, $\boldsymbol{A}_{GD}$, $\boldsymbol{A}_{rA}$, and $\boldsymbol{A}_{rD}$ are overcomplete matrices corresponding to the steering vectors of prediscretized AoA and AoD in user-RIS and RIS-BS channels, similar to $\boldsymbol{A}_{dA}$, and $\boldsymbol{A}_{dD}$ in (5)

Substituting (5)-(7) into (1), the received signal model can be written as (8) at the last of the next page, where $\boldsymbol{H}_{Gv}\boldsymbol{A}_{GD}^{H}\boldsymbol{\Theta}\boldsymbol{A}_{rA}\boldsymbol{H}_{rv}$ is combined and expressed as the beamspace form of cascaded channels $\boldsymbol{H}_{cv}$. In addition, the symbol $s(t)$ is set as 1 in the training stage.

Collecting all the measured values $y(t)$ and stacking them into a vector $\boldsymbol{y} \triangleq [y(1), \cdots, y(T)]^T$, we have

$$\begin{aligned}\boldsymbol{y} &= \begin{bmatrix} \boldsymbol{f}^T(1)\otimes \boldsymbol{z}^H(1) \\ \vdots \\ \boldsymbol{f}^T(T)\otimes \boldsymbol{z}^H(T) \end{bmatrix} (\boldsymbol{A}_{dD}^{*}\otimes \boldsymbol{A}_{dA})\,\mathrm{vec}(\boldsymbol{H}_{dv}) \\ &+ \begin{bmatrix} \boldsymbol{f}^T(1)\otimes \boldsymbol{z}^H(1) \\ \vdots \\ \boldsymbol{f}^T(T)\otimes \boldsymbol{z}^H(T) \end{bmatrix} (\boldsymbol{A}_{rD}^{*}\otimes \boldsymbol{A}_{GA})\,\mathrm{vec}(\boldsymbol{H}_{cv}) + \omega \\ &= \boldsymbol{\psi}_1\boldsymbol{h}_1 + \boldsymbol{\psi}_2\boldsymbol{h}_2 + \omega \\ &= \begin{bmatrix} \boldsymbol{\psi}_1 & \boldsymbol{\psi}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{h}_1 \\ \boldsymbol{h}_2 \end{bmatrix} + \omega \\ &= \boldsymbol{\psi}\boldsymbol{h} + \omega, \end{aligned} \tag{9}$$

with

$$\boldsymbol{\psi}_1 = \begin{bmatrix} \boldsymbol{f}^T(1)\otimes \boldsymbol{z}^H(1) \\ \vdots \\ \boldsymbol{f}^T(T)\otimes \boldsymbol{z}^H(T) \end{bmatrix} (\boldsymbol{A}_{dD}^{*}\otimes \boldsymbol{A}_{\ dA}), \tag{10}$$

$$\boldsymbol{\psi}_2 = \begin{bmatrix} \boldsymbol{f}^T(1)\otimes \boldsymbol{z}^H(1) \\ \vdots \\ \boldsymbol{f}^T(T)\otimes \boldsymbol{z}^H(T) \end{bmatrix} (\boldsymbol{A}_{rD}^{*}\otimes \boldsymbol{A}_{GA}), \tag{11}$$

and $\boldsymbol{h}_1 = \mathrm{vec}(\boldsymbol{H}_{dv})$, $\boldsymbol{h}_2 = \mathrm{vec}(\boldsymbol{H}_{cv})$, $\boldsymbol{\psi} = [\boldsymbol{\psi}_1, \boldsymbol{\psi}_2]$, $\boldsymbol{h} = [\boldsymbol{h}_1{}^T, \boldsymbol{h}_2{}^T]^T$.

Making use of the sparse characteristics, the channel estimation problem can be effectively solved by converting it into a sparse recovery problem [40]:

$$\begin{aligned} &\min_{\boldsymbol{h}} \|\boldsymbol{h}\|_1 \\ &s.t.\ \|\boldsymbol{y} - \boldsymbol{\psi}\boldsymbol{h}\|_2 \le \varepsilon, \end{aligned} \tag{12}$$

where ε denotes the accuracy threshold.

According to the theory of compressed sensing, the high dimensional sparse signal $\boldsymbol{h}$ can be successfully recovered from the lower dimensional measurement $\boldsymbol{y}$ in the noiseless environment. Thus it is of vital importance to reconstruct $\hat{\boldsymbol{y}}$ from the noisy received signal $\boldsymbol{y}$ to improve the reliability of sparse signal recovery.

III. Channel Model with Angular Spread

Based on the sparse characteristics, the channel estimation problem was converted into a sparse signal recovery problem in the previous section. However, the mmWave channel also takes the form of angular spread in AoA and AoD domains [31]. The angular spread is caused by scattering clusters, presenting a structured sparsity pattern with each cluster possibly contributing multiple paths [32]. This section will demonstrate that the mmWave channel exhibits joint sparse and low-rank characteristics as the result of asymmetric angular spreads.

Supposing angular spreads in the AoA domain only come from one common AoD, we start with this straightforward scenario to illustrate the low-rank characteristics, and the channel from user to BS can be expressed as

$$\boldsymbol{H}_d = \left(\sum_{i=1}^{I} \alpha_i \boldsymbol{a}_{dA}(\theta - \nu_i)\right) \boldsymbol{a}_{dD}^{H}(\phi), \tag{13}$$

where α_i denotes the gain of the i-th path, and ν_i denotes the offset of the i-th path compared to the mean AoA θ. Obviously, under this circumstance, the rank of matrix $\boldsymbol{H}_d$ is only one.

Extending from this simple case, we further analyze the channel model when two closely spaced AoDs generate angular spreads in the AoA domain,

$$\begin{aligned}\boldsymbol{H}_d = &\left(\sum_{i=1}^{I} \alpha_i \boldsymbol{a}_{dA}(\theta - \nu_i)\right) \boldsymbol{a}_{dD}^{H}(\phi - \varphi_1) \\ &+ \left(\sum_{i=1}^{I} \alpha'_i \boldsymbol{a}_{dA}(\theta - \nu_i)\right) \boldsymbol{a}_{dD}^{H}(\phi - \varphi_2). \end{aligned} \tag{14}$$

Due to the assumption that the two AoDs are close to each other, the corresponding AoA has a similar power angle mode, namely, $\sum_{i=1}^{I} \alpha_i \boldsymbol{a}_{dA}(\theta - \nu_i) = \sum_{i=1}^{I} \alpha'_i \boldsymbol{a}_{dA}(\theta - \nu_i)$. Then, it can be further simplified and expressed as

$$\boldsymbol{H}_d = \left(\sum_{i=1}^{I} \alpha_i \boldsymbol{a}_{dA}(\theta - \nu_i)\right) \left(\sum_{j=1}^{2} \boldsymbol{a}_{dD}^{H}(\phi - \varphi_j)\right). \tag{15}$$

Hence, we can continue to expand the number of AoDs and clusters to characterize the universally applicable geometric channel model as

$$\boldsymbol{H}_d = \sum_{l=1}^{L} \left(\sum_{i=1}^{I} \alpha_{l,i} \boldsymbol{a}_{dA}(\theta_l - \nu_{l,i})\right) \left(\sum_{j=1}^{J} \beta_{l,j} \boldsymbol{a}_{dD}^{H}(\phi_l - \varphi_{l,j})\right), \tag{16}$$

where L is the number of clusters, I is the number of paths, and J can be regarded as the number of AoDs in the clusters. In the l-th cluster, $\alpha_{l,i}$ and $\beta_{l,j}$ are the path gains, θ_l and ϕ_l are the average AoA/AoD, and $\nu_{l,i}$ and $\varphi_{l,j}$ are the corresponding offsets.

Similar to the beamspace form in (5), the geometric channel model could be transformed into the beamspace form with angular spread:

$$\begin{aligned}\boldsymbol{H}_d &= \sum_{l=1}^{L} \boldsymbol{A}_{dA}\boldsymbol{\alpha}_l\boldsymbol{\beta}_l^T\boldsymbol{A}_{dD}^{H} \\ &= \boldsymbol{A}_{dA}\left(\sum_{l=1}^{L} \boldsymbol{\alpha}_l\boldsymbol{\beta}_l^T\right)\boldsymbol{A}_{dD}^{H} \\ &= \boldsymbol{A}_{dA}\boldsymbol{H}_{dv}\boldsymbol{A}_{dD}^{H}, \end{aligned} \tag{17}$$

where $\boldsymbol{\alpha}_l$ and $\boldsymbol{\beta}_l$ are virtual representations over the AoA and AoD domains, and $\boldsymbol{H}_{dv}$ is the virtual beamspace channel of $\boldsymbol{H}_d$.

Since only a tiny piece of the whole angular domain is occupied by the angular spread, both $\boldsymbol{\alpha}_l$ and $\boldsymbol{\beta}_l$ are sparse vectors with only a few non-zero elements centered on the

average AoA and AoD. Thus, the virtual beamspace channel $\boldsymbol{H}_{dv}$ is composed of L sparse matrices. If we assume that $\boldsymbol{\alpha}_l$ and $\boldsymbol{\beta}_l$ contain at most Q non-zero elements, the maximum numbers of non-zero columns and non-zero rows in $\boldsymbol{H}_{dv}$ are both QL, and we have $QL \ll \min\{N_1, N_2\}$. That is, $\boldsymbol{H}_{dv}$ has sparsity, and since $\mathrm{rank}(\boldsymbol{H}_{dv}) = L$, it is clear that $\boldsymbol{H}_{dv}$ has low-rank characteristics. Therefore, the virtual beamspace channel takes on joint sparse and low-rank characteristics. Similarly, the BS-RIS and RIS-user channels also have low-rank virtual beamspace channels. Moreover, in RIS-assisted massive MIMO systems, the number of antennas at the BS and the number of reflecting elements at the RIS are typically larger than the number of antennas at the user. As a result, $\boldsymbol{G\Theta H}_r$ is a low-rank matrix, which has been extensively investigated in mmWave systems.

IV. Two-Stage Channel Estimation Scheme

This section will employ the joint sparse and low-rank characteristics for channel estimation in RIS-assisted mmWave systems. Especially, in addition to the sparse signal recovery, which has been extensively studied, we have previously set up a trace-based low-rank matrix approximation against the noisy channel. This can better meet the precondition of the compressed sensing theory and further improve the channel estimation accuracy.

A. Stage 1: Low-Rank Matrix Approximation

To reconstruct the noisy received signal, we carry out the process of low-rank matrix approximation in the first stage. Aiming at the deviation caused by the traditional nuclear norm minimization method, we propose a trace operator that can approximate the rank of a matrix well.

To begin, we go back to our received signal model (1) and recast it by the low-rank sampling process. Suppose $\mathcal{F}$ and $\mathcal{Z}$ are preprepared randomly for beamforming/receiving vectors $\boldsymbol{f}(t)$ and $\boldsymbol{z}(t)$, where the cardinalities of the sets are $|\mathcal{Z}| = N_Z$ and $|\mathcal{F}| = N_F$. Assume that all the vectors in $\mathcal{Z}$ form a matrix $\boldsymbol{Z} \in \mathbb{C}^{N_{\mathrm{BS}} \times N_{\mathrm{Z}}}$ and that all the vectors in $\mathcal{F}$ form a matrix $\boldsymbol{F} \in \mathbb{C}^{N_{\mathrm{MS}} \times N_F}$. The low-rank matrix sampling model of the received signal can therefore be written as

$$\begin{aligned}\boldsymbol{Y}_{ij} &= (\boldsymbol{Z}^H(\boldsymbol{H}_d + \boldsymbol{G\Theta H}_r)\boldsymbol{F} + \boldsymbol{\Omega})_{ij} \\ &= (\boldsymbol{Z}^H\boldsymbol{H}_d\boldsymbol{F})_{ij} + (\boldsymbol{Z}^H\boldsymbol{G\Theta H}_r\boldsymbol{F})_{ij} + \boldsymbol{\Omega}_{ij}, (i,j) \in \boldsymbol{\Upsilon},\end{aligned} \tag{18}$$

where $\boldsymbol{Y} \triangleq \boldsymbol{Z}^H(\boldsymbol{H}_d + \boldsymbol{G\Theta H}_r)\boldsymbol{F}$ is a low-rank matrix with the rank of L, $\boldsymbol{Y}_{ij}$ represents the ij-th element of $\boldsymbol{Y}$, and $\boldsymbol{\Omega}$ is the noise. $\boldsymbol{\Upsilon}$ represents the observed set.

Then, based on the low-rank matrix sampling model above, we can recover $\hat{\boldsymbol{Y}}$ from the noisy observed value $\boldsymbol{Y}$. Assume that $\boldsymbol{F}$ and $\boldsymbol{Z}$ are square matrices with full-rank, i.e. $N_F = N_{US}$ and $N_Z = N_{BS}$. We can find a low-rank matrix to approximate the original signal [41],

$$\begin{aligned}&\min_{\hat{\boldsymbol{Y}}} \mathrm{rank}(\hat{\boldsymbol{Y}}) \\ &s.t. \left\|\boldsymbol{Y} - \hat{\boldsymbol{Y}}\right\|_F^2 \le \varepsilon_1,\end{aligned} \tag{19}$$

where ε_1 is the precise threshold.

Since the above original problem is not easy to solve directly, many schemes focus on the nuclear norm minimization to obtain the approximation solution [42], [43]. Hence, the original problem (19) can be transformed as

$$\begin{aligned}&\min_{\hat{\boldsymbol{Y}}} \left\|\hat{\boldsymbol{Y}}\right\|_* \\ &s.t. \left\|\boldsymbol{Y} - \hat{\boldsymbol{Y}}\right\|_F^2 \le \varepsilon_1,\end{aligned} \tag{20}$$

where $\left\|\hat{\boldsymbol{Y}}\right\|_*$ denotes the nuclear norm of $\hat{\boldsymbol{Y}}$.

However, since the nuclear norm is a convex relaxation of rank, the solution obtained by minimizing the nuclear norm is usually biased and thus affects channel estimation accuracy [35]. To overcome the deviation, we adopt an improved-rank operator $\mathrm{tr}(\mathbf{P}_\mu(\boldsymbol{\Lambda}))$ to approximate the original low-rank problem [44] and it follows as

$$\mathbf{P}_\mu(\boldsymbol{\Lambda}) = \boldsymbol{\Lambda}(\boldsymbol{\Lambda}^H\boldsymbol{\Lambda} + \mu\boldsymbol{I})^{-1}\boldsymbol{\Lambda}^H, \qquad \mu \ge 0, \tag{21}$$

when the rank of $\boldsymbol{\Lambda}$ is full, we have $\mu = 0$; and when the rank is not full, $\mathrm{tr}(\mathbf{P}_\mu(\boldsymbol{\Lambda}))$ can well approximate $\mathrm{rank}(\boldsymbol{\Lambda})$, as depicted in the following theorem.

Theorem 1: For a matrix $\boldsymbol{\Lambda} \in \mathbb{R}^{M \times N}$, when the parameter satisfies $\mu > 0$, there exists

$$\lim_{\mu \to 0} \mathrm{tr}(\mathbf{P}_\mu(\boldsymbol{\Lambda})) = \mathrm{rank}(\boldsymbol{\Lambda}), \tag{22}$$

where $\mathrm{tr}(\cdot)$ denotes the trace operator. .

Proof: Please refer to the Appendix B for specific proof.

Based on Theorem 1, the original problem could be approximated by the improved rank operator,

$$\begin{aligned}&\min_{\hat{\boldsymbol{Y}}} \mathrm{tr}(\mathbf{P}_\mu(\hat{\boldsymbol{Y}})) \\ &s.t. \left\|\boldsymbol{Y} - \hat{\boldsymbol{Y}}\right\|_F^2 \le \varepsilon_1.\end{aligned} \tag{23}$$

The theorem indicates that when the coefficient μ approaches 0, the improved rank operator $\mathrm{tr}(\mathbf{P}_\mu(\hat{\boldsymbol{Y}}))$ could approach $\mathrm{rank}(\hat{\boldsymbol{Y}})$. Therefore, we set the process of loop iteration to realize the requirement that the coefficient approaches 0.

To be specific, to approach the rank operator as smoothly as possible, we establish an iterative approximation using the trace-based operator $\mathrm{tr}(\mathbf{P}_\mu(\hat{\boldsymbol{Y}}))$ as the cost of minimizing each iteration. The minimum value of each iteration is

$$\min_{\hat{\boldsymbol{Y}}} \left\{ \frac{1}{2}\left\|\boldsymbol{Y} - \hat{\boldsymbol{Y}}\right\|_F^2 + \eta \mathrm{tr}\left(\mathbf{P}_\mu\left(\hat{\boldsymbol{Y}}\right)\right) \right\}. \tag{24}$$

Based on the SVD of $\boldsymbol{Y}$, we have $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{\Sigma}_{\boldsymbol{Y}}\boldsymbol{V}^H$ and define $\boldsymbol{\Phi} = \boldsymbol{U}^H\hat{\boldsymbol{Y}}\boldsymbol{V}$. For unitary matrices $\boldsymbol{V}$ and $\boldsymbol{U}$, there exists $\mathrm{tr}(\mathbf{P}_\mu(\hat{\boldsymbol{Y}})) = \mathrm{tr}\left(\mathbf{P}_\mu\left(\boldsymbol{\Phi}\right)\right)$. Due to the Frobenius norm's unitary invariance, we can derive

$$\begin{aligned} g\left(\hat{\boldsymbol{Y}}\right) &= \tfrac{1}{2}\left\|\boldsymbol{Y} - \hat{\boldsymbol{Y}}\right\|_F^2 + \eta \mathrm{tr}\left(\mathbf{P}_\mu\left(\hat{\boldsymbol{Y}}\right)\right) \\ &= \tfrac{1}{2}\left\|\boldsymbol{U}\left(\boldsymbol{\Sigma}_{\boldsymbol{Y}} - \boldsymbol{\Phi}\right)\boldsymbol{V}^H\right\|_F^2 + \eta \mathrm{tr}\left(\mathbf{P}_\mu\left(\hat{\boldsymbol{Y}}\right)\right) \\ &= \tfrac{1}{2}\left\|\boldsymbol{\Sigma}_{\boldsymbol{Y}} - \boldsymbol{\Phi}\right\|_F^2 + \eta \mathrm{tr}\left(\mathbf{P}_\mu\left(\boldsymbol{\Phi}\right)\right). \end{aligned} \tag{25}$$

By substituting (25), the problem (24) may be recast as

$$\min_{\boldsymbol{\Phi}} \left\{ \frac{1}{2}\left\|\boldsymbol{\Sigma}_{\boldsymbol{Y}} - \boldsymbol{\Phi}\right\|_F^2 + \eta \mathrm{tr}\left(\mathbf{P}_\mu\left(\boldsymbol{\Phi}\right)\right) \right\}. \tag{26}$$

$$\begin{aligned}\mathrm{vec}(\hat{\boldsymbol{Y}}) &= \mathrm{vec}(\boldsymbol{C}_d\boldsymbol{H}_{dv}\boldsymbol{D}_d) + \mathrm{vec}(\boldsymbol{C}_c(\boldsymbol{H}_{Gv}\boldsymbol{A}_{GD}^H\boldsymbol{\Theta}\boldsymbol{A}_{rA}\boldsymbol{H}_{rv})\boldsymbol{D}_c) + \mathrm{vec}(\boldsymbol{\Omega}) \\ &= (\boldsymbol{D}_d^T \otimes \boldsymbol{C}_d)\mathrm{vec}(\boldsymbol{H}_{dv}) + (\boldsymbol{D}_c^T \otimes \boldsymbol{C}_c)\mathrm{vec}(\boldsymbol{H}_{Gv}\boldsymbol{A}_{GD}^H\boldsymbol{\Theta}\boldsymbol{A}_{rA}\boldsymbol{H}_{rv}) + \mathrm{vec}(\boldsymbol{\Omega}) \\ &= (\boldsymbol{D}_d^T \otimes \boldsymbol{C}_d)\mathrm{vec}(\boldsymbol{H}_{dv}) + (\boldsymbol{D}_c^T \otimes \boldsymbol{C}_c)(\boldsymbol{H}_{rv}^T \otimes \boldsymbol{H}_{Gv})\mathrm{vec}(\boldsymbol{A}_{GD}^H\boldsymbol{\Theta}\boldsymbol{A}_{rA}) + \mathrm{vec}(\boldsymbol{\Omega}) \\ &= (\boldsymbol{D}_d^T \otimes \boldsymbol{C}_d)\mathrm{vec}(\boldsymbol{H}_{dv}) + ((\mathrm{vec}(\boldsymbol{A}_{GD}^H\boldsymbol{\Theta}\boldsymbol{A}_{rA}))^T \otimes (\boldsymbol{D}_c^T \otimes \boldsymbol{C}_c))\mathrm{vec}(\boldsymbol{H}_{rv}^T \otimes \boldsymbol{H}_{Gv}) \\ &= \begin{bmatrix} (\boldsymbol{D}_d^T \otimes \boldsymbol{C}_d) & (\mathrm{vec}(\boldsymbol{A}_{GD}^H\boldsymbol{\Theta}\boldsymbol{A}_{rA}))^T \otimes (\boldsymbol{D}_c^T \otimes \boldsymbol{C}_c) \end{bmatrix} \begin{bmatrix} \mathrm{vec}(\boldsymbol{H}_{dv}) \\ \mathrm{vec}(\boldsymbol{H}_{rv}^T \otimes \boldsymbol{H}_{Gv}) \end{bmatrix} + \mathrm{vec}(\boldsymbol{\Omega}) \\ &= \boldsymbol{\psi}\boldsymbol{h} + \mathrm{vec}(\boldsymbol{\Omega}),\end{aligned} \tag{36}$$

For the first portion of the problem (26), the following exists

$$\begin{aligned}\|\boldsymbol{\Sigma}_{\boldsymbol{Y}} - \boldsymbol{\Phi}\|_F^2 &= \|\boldsymbol{\Sigma}_{\boldsymbol{Y}}\|_F^2 + \|\boldsymbol{\Phi}\|_F^2 - 2\mathrm{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{Y}}\boldsymbol{\Phi}^H\right) \\ &\overset{(a)}{\geq} \|\boldsymbol{\Sigma}_{\boldsymbol{Y}}\|_F^2 + \|\boldsymbol{\Sigma}_{\boldsymbol{\Phi}}\|_F^2 - 2\mathrm{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{Y}}\boldsymbol{\Sigma}_{\boldsymbol{\Phi}}{}^H\right) \\ &= \|\boldsymbol{\Sigma}_{\boldsymbol{Y}} - \boldsymbol{\Sigma}_{\boldsymbol{\Phi}}\|_F^2 ,\end{aligned} \tag{27}$$

where operation (a) is based on the von Neumann trace inequality, and $\boldsymbol{\Sigma}_{\boldsymbol{\Phi}}$ denotes a diagonal matrix derived by SVD of $\boldsymbol{\Phi}$ [45].

We can deduce this further and get the following

$$\|\boldsymbol{\Sigma}_{\boldsymbol{Y}} - \boldsymbol{\Phi}\|_F^2 + \eta\mathrm{tr}\left(\mathbf{P}_\mu\left(\boldsymbol{\Phi}\right)\right) \geq \|\boldsymbol{\Sigma}_{\boldsymbol{Y}} - \boldsymbol{\Sigma}_{\boldsymbol{\Phi}}\|_{\mathrm{F}}^2 + \eta\mathrm{tr}\left(\mathbf{P}_\mu\left(\boldsymbol{\Phi}\right)\right), \tag{28}$$

blue where equality holds if $\boldsymbol{\Phi} = \boldsymbol{\Sigma}_{\boldsymbol{\Phi}}$. Based on (28), problem (29) is an effective approximation of (26) as follows

$$\min_{\boldsymbol{\Sigma}_{\boldsymbol{\Phi}}} \left\{ \frac{1}{2}\|\boldsymbol{\Sigma}_{\boldsymbol{Y}} - \boldsymbol{\Sigma}_{\boldsymbol{\Phi}}\|_F^2 + \eta\mathrm{tr}\left(\mathbf{P}_\mu\left(\boldsymbol{\Phi}\right)\right) \right\}. \tag{29}$$

Further, the Frobenius norm and Appendix B allow us to write the problem (29) as follows

$$\begin{aligned}&\min_{\boldsymbol{\Sigma}_{\boldsymbol{\Phi}}} \tfrac{1}{2}\|\boldsymbol{\Sigma}_{\boldsymbol{Y}} - \boldsymbol{\Sigma}_{\boldsymbol{\Phi}}\|_F^2 + \eta\mathrm{tr}\left(\mathbf{P}_\mu\left(\boldsymbol{\Phi}\right)\right) \\ &= \min_{\boldsymbol{\Sigma}_{\boldsymbol{\Phi}}} \tfrac{1}{2}\|\boldsymbol{\Sigma}_{\boldsymbol{Y}} - \boldsymbol{\Sigma}_{\boldsymbol{\Phi}}\|_F^2 + \eta\sum_{i=1}^{L}\tfrac{\sigma_i^2(\boldsymbol{\Phi})}{\sigma_i^2(\boldsymbol{\Phi})+\mu} \\ &= \min_{\boldsymbol{\Sigma}_{\boldsymbol{\Phi}}} \tfrac{1}{2}\|\boldsymbol{\Sigma}_{\boldsymbol{Y}}\|_F^2 + \tfrac{1}{2}\|\boldsymbol{\Sigma}_{\boldsymbol{\Phi}}\|_F^2 - \langle\boldsymbol{\Sigma}_{\boldsymbol{Y}}, \boldsymbol{\Sigma}_{\boldsymbol{\Phi}}\rangle_F + \eta\sum_{i=1}^{L}\tfrac{\sigma_i^2(\boldsymbol{\Phi})}{\sigma_i^2(\boldsymbol{\Phi})+\mu} \\ &= \min_{\sigma_1(\boldsymbol{\Phi}),\cdots,\sigma_L(\boldsymbol{\Phi})} \tfrac{1}{2}\sum_{i=1}^{L}\left(\sigma_i\left(\boldsymbol{\Sigma}_{\boldsymbol{Y}}\right) + \sigma_i\left(\boldsymbol{\Phi}\right)\right)^2 + \eta\sum_{i=1}^{L}\tfrac{\sigma_i^2(\boldsymbol{\Phi})}{\sigma_i^2(\boldsymbol{\Phi})+\mu}.\end{aligned} \tag{30}$$

This implies that we can disentangle the minimizations with regard to $\sigma_1\left(\Phi\right), \cdots, \sigma_L\left(\Phi\right)$ as

$$\min_{\sigma_i(\boldsymbol{\Phi})} \left\{ f\left(\sigma_i\left(\boldsymbol{\Phi}\right)\right) = \frac{1}{2}\left(\sigma_i\left(\boldsymbol{\Sigma}_{\boldsymbol{Y}}\right) + \sigma_i\left(\boldsymbol{\Phi}\right)\right)^2 + \eta\frac{\sigma_i^2\left(\boldsymbol{\Phi}\right)}{\sigma_i^2\left(\boldsymbol{\Phi}\right)+\mu} \right\}. \tag{31}$$

This is a simple scalar minimization problem, which is easy to solve.

As a result, we can write the answer to (19) as

$$\hat{\boldsymbol{Y}} = \boldsymbol{U}\mathrm{diag}(\sigma_1\left(\boldsymbol{\Phi}\right), \cdots, \sigma_L\left(\boldsymbol{\Phi}\right))\boldsymbol{V}^H. \tag{32}$$

Inspired by [46], the initial value of the first loop is set to $\mu = 4\max_i|\sigma_i(\boldsymbol{Y}_\Upsilon)|$. Next, we use $\mu^{(k)} = c\mu^{(k-1)}(0.5 < \mu < 1)$ to approach 0 iteratively, where c is empirically selected to fall between 0.5 and 1.

The improved trace operator-based low-rank approximation algorithm at the first stage of the algorithm is summarized in Algorithm 1.

Algorithm 1 Improved Trace Operator-Based Low-Rank Approximation Scheme.

Input: The observed noisy signal $\boldsymbol{Y}$.
Initialization: $\hat{\boldsymbol{Y}}^{(0)} = \boldsymbol{Y}$, $c = 0.5$, ρ, $k = 1$ and K;
While $k < K$ do
  Iterative regularization: $\boldsymbol{Y}_\Upsilon{}^{(k)} = \hat{\boldsymbol{Y}}^{(k-1)} + \rho(\boldsymbol{Y} - \hat{\boldsymbol{Y}}^{(k-1)})$;
  if $k = 1$
    $\mu = 4\max_i\left|\sigma_i(\boldsymbol{Y}^{(k)})\right|$;
  else
    Update the $\mu^{(k)} = c\mu^{(k-1)}$;
  end if
  Update the $\hat{\boldsymbol{Y}}^{(k)}$ via (23);
  $k = k + 1$;
end while
Output: The reconstructed signal $\hat{\boldsymbol{Y}}$.

B. Stage 2: Sparse Signal Recovery

After completing the reconstruction of $\hat{\boldsymbol{Y}}$, the second stage estimates the channel through sparse signal recovery. Specifically, based on (17), the beamspace model of user-BS, user-RIS, and RIS-BS channels can be written as

$$\begin{aligned}\boldsymbol{H}_d &= \boldsymbol{A}_{dA}\boldsymbol{H}_{dv}\boldsymbol{A}_{dD}^H, \\ \boldsymbol{H}_r &= \boldsymbol{A}_{rA}\boldsymbol{H}_{rv}\boldsymbol{A}_{rD}^H, \\ \boldsymbol{G} &= \boldsymbol{A}_{GA}\boldsymbol{H}_{Gv}\boldsymbol{A}_{GD}^H,\end{aligned} \tag{33}$$

where $\boldsymbol{H}_{dv}$, $\boldsymbol{H}_{rv}$ and $\boldsymbol{H}_{Gv}$ are virtual beamspace vectors, $\boldsymbol{A}_{dA}$, $\boldsymbol{A}_{rA}$, $\boldsymbol{A}_{GA}$ and $\boldsymbol{A}_{dD}^H$, $\boldsymbol{A}_{rD}^H$, $\boldsymbol{A}_{GD}^H$ are composed of the array response vectors corresponding to the AoA and AoD of each channel respectively.

Putting the beamspace model (33) into the sampling model (18), we have

$$\begin{aligned}\hat{\boldsymbol{Y}} &= \boldsymbol{Z}^H(\boldsymbol{A}_{dR}\boldsymbol{H}_{dv}\boldsymbol{A}_{dD}^H)\boldsymbol{F} \\ &\quad + \boldsymbol{Z}^H(\boldsymbol{A}_{GR}\boldsymbol{H}_{Gv}\boldsymbol{A}_{GD}^H)\boldsymbol{\Theta}(\boldsymbol{A}_{rR}\boldsymbol{H}_{rv}\boldsymbol{A}_{rD}^H)\boldsymbol{F} + \boldsymbol{\Omega} \\ &= (\boldsymbol{Z}^H\boldsymbol{A}_{dR})\boldsymbol{H}_{dv}(\boldsymbol{A}_{dD}^H\boldsymbol{F}) \\ &\quad + (\boldsymbol{Z}^H\boldsymbol{A}_{GR})(\boldsymbol{H}_{Gv}\boldsymbol{A}_{GD}^H\boldsymbol{\Theta}\boldsymbol{A}_{rR}\boldsymbol{H}_{rv})(\boldsymbol{A}_{rD}^H\boldsymbol{F}) + \boldsymbol{\Omega} \\ &= \boldsymbol{C}_d\boldsymbol{H}_{dv}\boldsymbol{D}_d \\ &\quad + \boldsymbol{C}_c(\boldsymbol{H}_{Gv}\boldsymbol{A}_{GD}^H\boldsymbol{\Theta}\boldsymbol{A}_{rR}\boldsymbol{H}_{rv})\boldsymbol{D}_c + \boldsymbol{\Omega},\end{aligned} \tag{34}$$

with

$$\begin{aligned}\boldsymbol{C}_d &= \boldsymbol{Z}^H\boldsymbol{A}_{dR}, \\ \boldsymbol{D}_d &= \boldsymbol{A}_{dD}^H\boldsymbol{F}, \\ \boldsymbol{C}_c &= \boldsymbol{Z}^H\boldsymbol{A}_{GR}, \\ \boldsymbol{D}_c &= \boldsymbol{A}_{rD}^H F.\end{aligned} \tag{35}$$

Then, employing the Kronecker product operation and matrix vectorization operator to further reduce the complexity, we obtain (36) shown at the top of last page, with

$$\boldsymbol{\psi} = \left[\begin{array}{cc} (\boldsymbol{D}_d^T \otimes \boldsymbol{C}_d) & (\mathrm{vec}(\boldsymbol{A}_{GD}^H \boldsymbol{\Theta} \boldsymbol{A}_{rA}))^T \otimes (\boldsymbol{D}_c^T \otimes \boldsymbol{C}_c) \end{array} \right], \tag{37}$$

$$\boldsymbol{h} = \left[\begin{array}{c} \mathrm{vec}(\boldsymbol{H}_{dv}) \\ \mathrm{vec}(\boldsymbol{H}_{rv}^T \otimes \boldsymbol{H}_{Gv}) \end{array} \right], \tag{38}$$

where $\boldsymbol{I}$ denotes a identity matrix. Here, channel estimation is converted into a sparse recovery problem,

$$\begin{aligned} &\min_{\boldsymbol{h}} \|\boldsymbol{h}\|_1 \\ &s.t. \left\| \mathrm{vec}(\hat{\boldsymbol{Y}}) - \boldsymbol{\psi}\boldsymbol{h} \right\|_2 \leq \varepsilon_2, \end{aligned} \tag{39}$$

where ε_2 denotes the precise threshold. Based on the relevant literature and previous experience, we made several attempts and finally set the threshold to 0.001.

The conventional compressed sensing-based schemes employ the directly observed data $\boldsymbol{Y}$ to estimate channels, which affects the estimation accuracy due to the existence of noise. However, the proposed scheme reconstructs $\hat{\boldsymbol{Y}}$ before sparse recovery to overcome this problem. The summarized two-stage scheme is shown in Algorithm 2.

Remark: In this paper, our analysis is based on the assumption that the RIS is a ULA. However, for UPA, our method can still be applied, and its simulation is shown in Appendix A.

Algorithm 2 Compressed Sensing-Based Two-Stage Channel Estimation Scheme.

Input: The observed signal $\boldsymbol{Y}$ and the coefficient matrices $\boldsymbol{\psi}$.

1: Recover $\hat{\boldsymbol{Y}}$ based on the **Algorithm 1** via

$$\begin{aligned} &\min_{\hat{\boldsymbol{Y}}} \mathrm{rank}(\hat{\boldsymbol{Y}}) \\ &s.t. \left\| \boldsymbol{Y} - \hat{\boldsymbol{Y}} \right\|_F^2 \leq \varepsilon_1, \end{aligned} \tag{40}$$

2: Estimate $\hat{\boldsymbol{h}}$ via

$$\begin{aligned} &\min_{\boldsymbol{h}} \|\boldsymbol{h}\|_1 \\ &s.t. \left\| \mathrm{vec}(\hat{\boldsymbol{Y}}) - \boldsymbol{\psi}\boldsymbol{h} \right\|_2 \leq \varepsilon_2. \end{aligned} \tag{41}$$

Assuming that in each iteration, the computational complexity can be divided into two part: low-rank matrix approximation and sparse recovery. For low-rank matrix approximation, the SVD of $\mathbf{Y}$ should be considered, and its complexity is $\mathcal{O}(\min(M^2N + N^2M))$; then, the complexity of updating the problem (23) is $\mathcal{O}((M^2N)^3 + 2NM^2)$. For sparse recovery, the corresponding complexity is $\mathcal{O}(\mathcal{K}^3 + 2\mathcal{K})$, where I is the dimension size of ψ. Hence, the complexity of Algorithm 2 can be expressed as,

$$\mathcal{O}(I_{iter}(\min(M^2N+N^2M)+(M^2N)^3+2NM^2+\mathcal{K}^3+2\mathcal{K}))$$

where I_{iter} is the number of iterations.

V. Simulation Result

In this section, we perform simulations and provide numerical results to evaluate the proposed two-stage scheme. Consider a RIS-assisted uplink mmWave MIMO system, as shown in Fig. 1. Assume both the user and BS employ uniform linear array antennas, where the distance of adjacent units is half the signal wavelength. RIS is a panel consisting of M uniform linear arrays, and the reflection coefficient matrix of the RIS is expressed as $\boldsymbol{\Theta} = \mathrm{diag}(\beta_1 e^{j\theta_1}, \cdots, \beta_N e^{j\theta_N})$. In practice, each element of the RIS is usually designed to maximize signal reflection, and thus we set $\beta_i = 1$. θ_N is generated randomly [1]. In the general case, we set $M = 16$, $N_{BS} = 32$ and $N_{US} = 16$. The mmWave channels are generated by the geometric channel in (17). Taking the sparse scattering of mmWave into account, the number of clusters in each transmission link is set to $L = 2$ [34]. For these clusters, we assume the average AoAs and AoDs are 0, and the relative AoA shifts and AoD shifts are randomly generated via an inverse transform sampling-based random variable generator. The beamforming matrix $\boldsymbol{F}$ and receiving combination matrix $\boldsymbol{Z}$ are randomly generated by circularly symmetric complex Gaussian (CSCG) distribution in the experiment. For stage 2, problem (39) is tackled using a fast iterative shrinkage-thresholding algorithm (FISTA) [47]. In addition, the classical LS and OMP methods, as well as one-stage scheme without low-rank approximation and nuclear norm minimization (NNM)-based two-stage scheme [34], are chosen as comparisons. The dictionary of OMP are constructed from the Kronecker product of beamforming and receive combining matrices, i.e $F^T \otimes Z$, to take full advantage of the characteristics of both the transmitter and the receiver. Numerical results derived from the average of 10,000 Monte Carlo experiments.

We select MSE as the metric to assess the accuracy of the channel estimation methods. MSE is defined as the average square sum of the distance between each observed data and the actual value, which is one of the most common measures to reflect the difference between estimators. The formula is shown as follows

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \left(\boldsymbol{h}_i - \hat{\boldsymbol{h}}_i \right)^2, \tag{42}$$

where $\boldsymbol{h}_i$ and $\hat{\boldsymbol{h}}_i$ are the actual and estimated channels respectively. Under the same conditions, the smaller the MSE, the better the estimation scheme.

We first examine the estimation accuracy varying with signal-to-noise ratio (SNR), as illustrated in Fig. 2. It can be observed that the MSEs of all schemes decrease as SNR increases, and the proposed scheme consistently outperforms the others. To be specific, the LS method completely ignores the impact of noise and has poor performance. As a classical compression sensing algorithm, the OMP method makes full use of the sparse characteristics of the channel and can better cope with the noisy environment. Considering the effect of noise on channel recovery, the NNM-based two-stage scheme and the proposed scheme reconstruct the noisy received signal by low-rank matrix approximation, fitting in with the precon-

This article has been accepted for publication in IEEE Transactions on Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2023.3331521

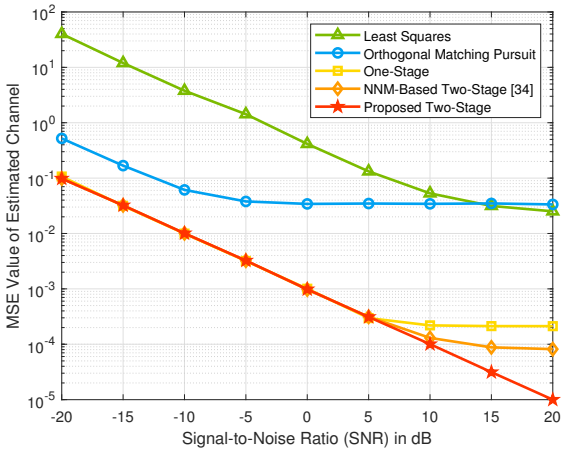

Fig. 2: The MSE performance versus SNR.

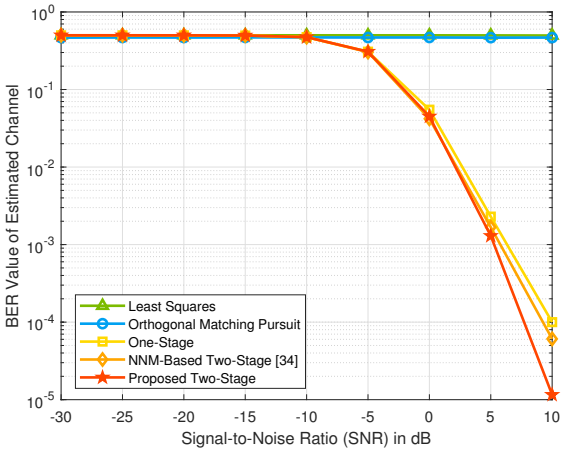

Fig. 4: The BER performance versus SNR.

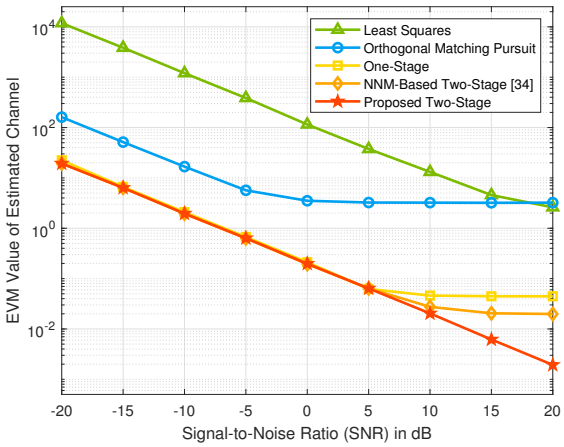

Fig. 3: The EVM performance versus SNR.

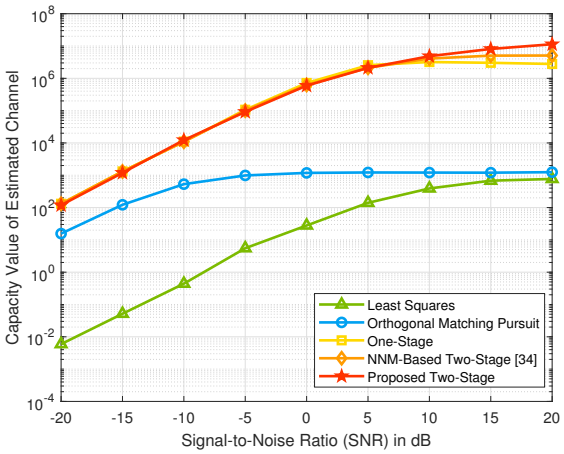

Fig. 5: The capacity performance versus SNR.

dition of the compressed sensing theory, thus bringing more obvious performance improvement. However, the coefficients of the NNM scheme remain unchanged after selection, while the proposed scheme adjusts the coefficients in each iteration. The fixed coefficients limit the flexibility and accuracy of the algorithm because the optimum values of the coefficients may change with the channel parameters. It can be observed that when SNR is high (especially after 10dB), the proposed scheme can deal with inconspicuous noise flexibly, while the performance improvement of the NNM scheme is no longer obvious. Additionally, to ensure that MSE is appropriate as the evaluation index, EVM is employed for verification. The EVM is calculated as the root-mean-square of the ratio of the ideal error-free reference signal power $\hat{\boldsymbol{P}}$ and the actual estimated signal power $\boldsymbol{P}$, expressed as $EVM = \sqrt{\frac{\hat{\boldsymbol{P}}}{\boldsymbol{P}}} \times 100\%$ [48]. As shown in Fig. 3, the variation trend is similar to that of MSE, which indicates the prudence and objectivity of selecting MSE as the metric.

Given that bit error rate (BER) and capacity are pivotal metrics for evaluating channel estimation's effectiveness, as BER mirrors transmission accuracy and capacity gauges achievable data rates, both hinging on precise channel estimation, we now move forward to authenticate their performance. We calculated the change of channel capacity with SNR according to Shannon's formula, as shown in Fig. 4. For convenience, we set the bandwidth to $1Hz$. As SNR increases, the capacity of all methods increases. Consistent with the MSE trend, our algorithm shows greater advantages when the SNR is greater than 10 dB. Fig. 5 illustrates the BER performance in relation to the SNR. After channel estimation for each approach, the beamforming or combining vectors are computed using the technique outlined in [40]. As SNR increases, our proposed method exhibits a decreasing trend in BER, while the performance improvement in other methods is less pronounced. Notably, our proposed method exhibits a constant advantage. These results demonstrate that the proposed scheme can effectively deal with channel estimation in different SNR environments and significantly improve estimation accuracy and performance.

Subsequently, we explore the impact of the pilot sequence length on the estimation performance. The MSE performance in relation to the pilot is delineated in Fig. 6. From the figure, it becomes evident that as the pilot sequence length becomes more extensive, it equips the channel estimation with augmented information, leading to an enhanced accuracy, as reflected by a diminishing MSE. Remarkably, under identical

This article has been accepted for publication in IEEE Transactions on Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2023.3331521

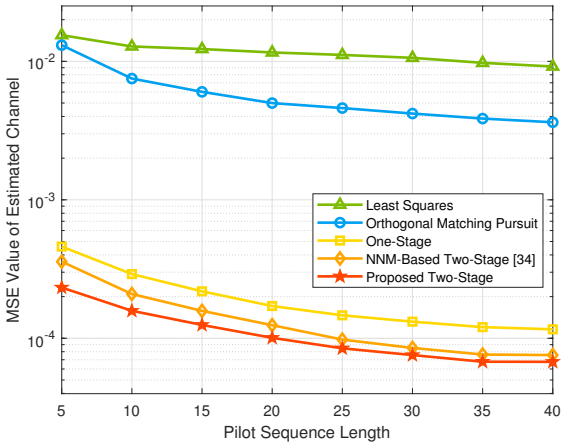

Fig. 6: The MSE performance versus the pilot sequence length.

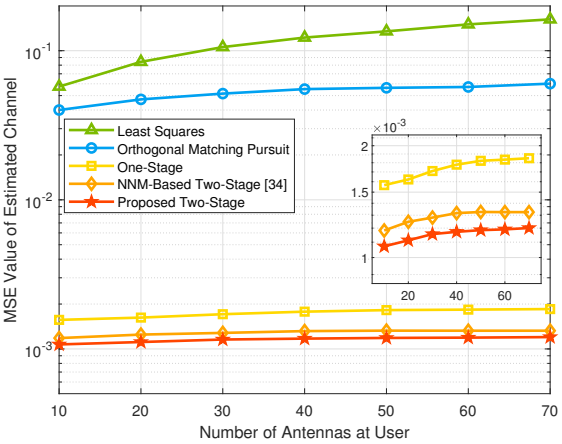

Fig. 8: The MSE performance versus the number of antennas at user.

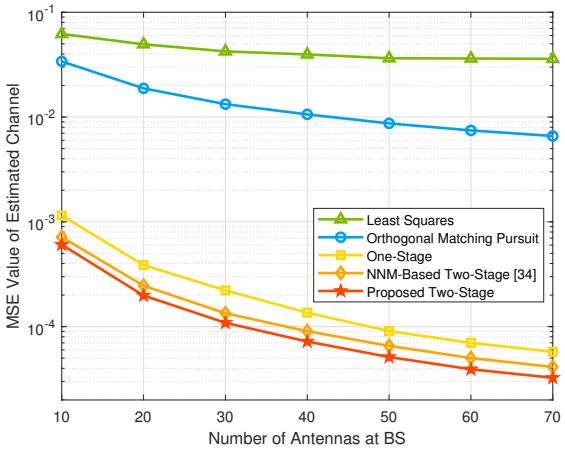

Fig. 7: The MSE performance versus the number of antennas at BS.

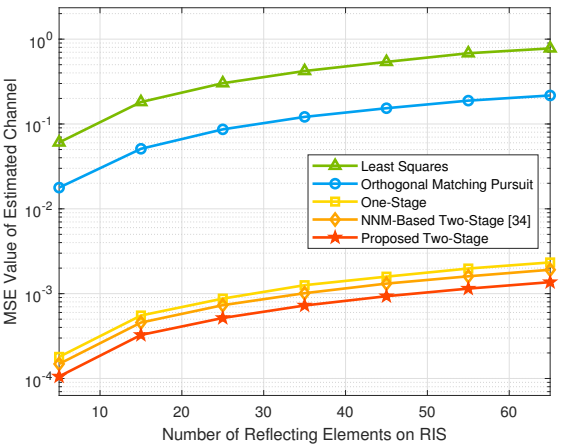

Fig. 9: The MSE performance versus the number of reflecting elements on RIS.

pilot variations, our proposed methodology consistently outperforms other comparison algorithms, showcasing a precision advantage. Further research also indicates that when the length of pilots exceeds 30, the improvement of our proposed scheme in MSE performance gradually slows down, while other methods, such as OMP, require a longer training duration.

Next, the experiments are conducted to see how the numbers of antennas at the BS and user impact the estimation accuracy. We simulate with the SNR of 10dB and a steadily growing antenna number while leaving all other settings unchanged. The numerical results presented in Fig. 7 indicate that increasing the antennas at the BS can reduce MSEs, that is, improve the estimation accuracy. This phenomenon is caused by the improved channel quality due to the increase in the number of antennas. Specifically, the LS method is greatly affected by the noise, which not only shows the worst accuracy but also has the slowest improvement with the increase of antenna numbers. On the contrary, the NNM-based two-stage scheme and the proposed scheme have obvious performance improvement with the increase in the number of antennas. However, it is worth noting that the proposed scheme always maintains the best performance, indicating that the proposed trace operator and the design of the two-stage algorithm are effective and meaningful. Conversely, in the results of Fig. 8, it can be observed that the MSE increases as the number of UE antennas increases. This is because, in the uplink communication system, channel estimation is based on uplink pilot signals sent from UE to BS. Therefore, when the number of UE antennas increases, the variables that need to be estimated in the channel matrix will increase. However, when the number of pilots remains unchanged, the dimension of the received signal will not change, and MSE will become worse. But our proposed method always has optimal performance and strong robustness. Similar situations also appear in Fig. 9 and 10. The increase of the reflection unit makes the channel matrix continuously increase; and more paths also bring more complex channels. It is worth noting that our proposed scheme always has the best performance in terms of precision and robustness.

Given the severe path loss in mmWave communications, the value of SNR before beam alignment is typically low, sometimes below 0dB. In this experiment, we set the SNR to 10dB and -5dB to track the success ratio under the low SNR region. Specifically, when MSE is less than 10^{-3}, it is

This article has been accepted for publication in IEEE Transactions on Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2023.3331521

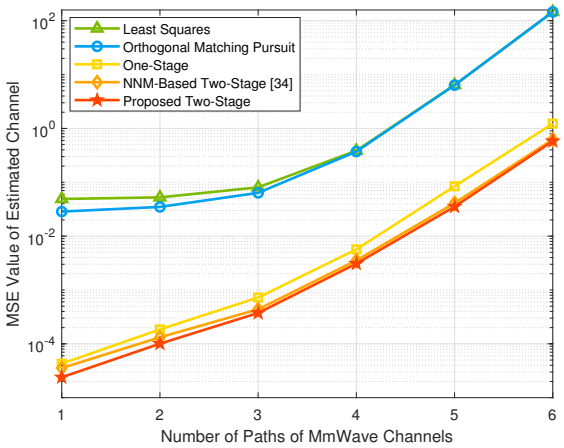

Fig. 10: The MSE performance versus the number of paths in mmWave channels.

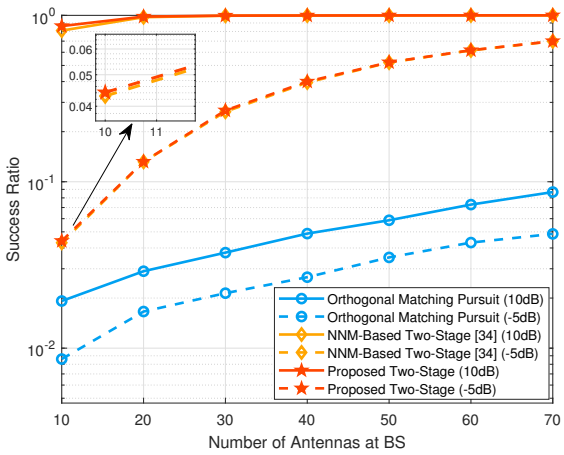

Fig. 11: Success ratio versus the number of antennas at BS.

generally considered to have no error layer so that this trial can be marked as a success. The success ratio is hence defined as the ratio of the number of successes to the total number of trials. It can be observed from the numerical results in Fig. 11 that, with the increase in the number of antennas, the performance of the NNM-based two-stage scheme and the proposed scheme are significantly better than that of the traditional OMP algorithm. In the situation that the SNR is set to 10dB, when the number of antennas is 30, the proposed scheme achieves a high success ratio of 99.64%, while when the number of antennas increases to 70, the success ratio is nearly 100%. The NNM-based schemes also maintain a high success rate but are always lower than the proposed schemes. Even with SNR=-5dB, the proposed scheme still has a success rate of about 0.2 percentage points higher than the NNM-based scheme. We continue to test the change of success ratio with the number of mmWave paths in Fig. 12. When SNR is 10dB, the success ratio of the proposed scheme maintains a high success ratio of 100% and 99.71% when the number of paths is 1 and 2. The consistently high success rate indicates that the proposed method can effectively deal with channel estimation under different communication conditions, even in the severe environment of low SNR and numerous antennas.

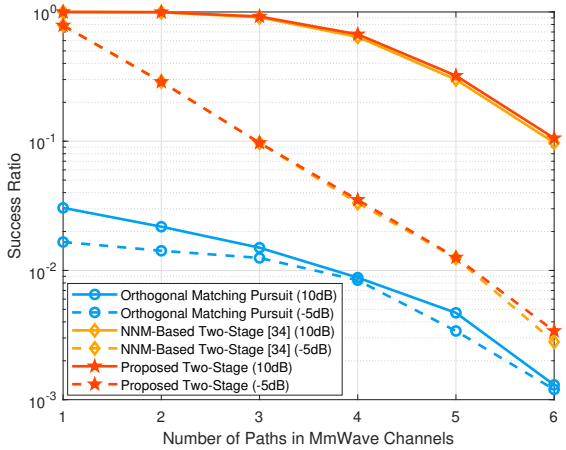

Fig. 12: Success ratio versus the number of paths in mmWave channels.

VI. Conclusion

In this paper, we developed a two-stage channel estimation scheme for RIS-assisted mmWave MIMO systems. Firstly, the RIS-assisted mmWave channel model with angular spread was established, which presented joint sparse and low-rank characteristics. In the first stage, we utilized low-rank characteristics to reconstruct the noisy received signal. Specifically, to solve the low-rank matrix approximation problem, the trace operator was proposed as a replacement since it could approximate the rank operator well. In the second stage, based on the properties of Kronecker products, the channel estimation model was transformed into a sparse signal recovery problem. Simulation results indicated that the proposed scheme could effectively perform accurate channel estimation and was robust for different channel environments.

Appendix A
Simulation of UPA

To verify the performance when RIS is UPA, we provide the simulation as follows. It can be seen from the figure that when RIS is UPA, our channel estimation algorithm still maintains the ideal performance.

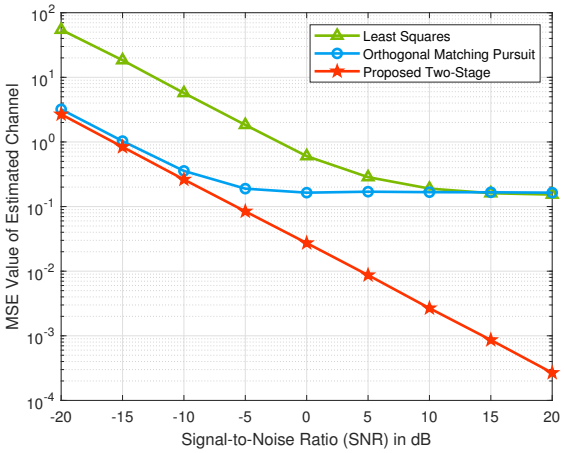

Fig. 13: The MSE performance when RIS is UPA.

APPENDIX B
PROOF OF THE THEOREM 1

Assume $\mathrm{rank}(\boldsymbol{\Lambda}) = r$, and $\boldsymbol{\Lambda}$ is recast by singular value decomposition as the form of $\boldsymbol{\Lambda} = \boldsymbol{U}\Sigma_{\boldsymbol{\Lambda}}\boldsymbol{V}^H = \boldsymbol{U}\begin{pmatrix} \boldsymbol{\Sigma}_r & \mathbf{0}_{r\times(N-r)} \\ \mathbf{0}_{(M-r)\times r} & \mathbf{0}_{(M-r)\times(N-r)} \end{pmatrix}\boldsymbol{V}^H$, where $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthogonal matrices, $\mathbf{0}_{M\times N}$ is $M \times N$ zero matrix and $\boldsymbol{\Sigma}_r = \mathrm{diag}\left\{\sigma_1(\boldsymbol{\Lambda}), \sigma_2(\boldsymbol{\Lambda}), \ldots, \sigma_r(\boldsymbol{\Lambda})\right\}$ denotes diagonal matrix. Based on the above, we can derive

$$\begin{aligned}\boldsymbol{\Lambda}^H\boldsymbol{\Lambda} + \mu\boldsymbol{I} &= \boldsymbol{V}\left(\boldsymbol{\Sigma}_r^H\boldsymbol{\Sigma}_r + \mu\boldsymbol{I}\right)\boldsymbol{V}^H \\ &= \boldsymbol{V}\begin{pmatrix} \boldsymbol{\Sigma}_r^H\boldsymbol{\Sigma}_r + \mu\boldsymbol{I}_{r\times r} & \mathbf{0}_{r\times(N-r)} \\ \mathbf{0}_{(M-r)\times r} & \mu\boldsymbol{I}_{(M-r)\times(N-r)} \end{pmatrix}\boldsymbol{V}^H.\end{aligned}$$

Therefore, we can write the matrix $\mathbf{P}_\mu(\Lambda)$ as

$$\begin{aligned}\mathrm{tr}\left(\mathbf{P}_\mu(\boldsymbol{\Lambda})\right) &= \mathrm{tr}\left(\boldsymbol{\Lambda}\left(\boldsymbol{\Lambda}^H\boldsymbol{\Lambda} + \mu\boldsymbol{I}\right)^{-1}\boldsymbol{\Lambda}^H\right) \\ &= \mathrm{tr}\left(\boldsymbol{U}\boldsymbol{\Sigma}_r\begin{pmatrix} \boldsymbol{\Sigma}_r^H\boldsymbol{\Sigma}_r + \mu\boldsymbol{I}_{r\times r} & \mathbf{0}_{r\times(N-r)} \\ \mathbf{0}_{(M-r)\times r} & \mu\boldsymbol{I}_{(M-r)\times(N-r)} \end{pmatrix}^{-1}\boldsymbol{\Sigma}_r^H\boldsymbol{U}^H\right) \\ &= \mathrm{tr}\left(\begin{pmatrix} \boldsymbol{\Sigma}_r^H\boldsymbol{\Sigma}_r + \mu\boldsymbol{I}_{r\times r} & \mathbf{0}_{r\times(N-r)} \\ \mathbf{0}_{(M-r)\times r} & \mu\boldsymbol{I}_{(M-r)\times(N-r)} \end{pmatrix}^{-1}\begin{pmatrix} \boldsymbol{\Sigma}_r^H\boldsymbol{\Sigma}_r & \mathbf{0}_{r\times(N-r)} \\ \mathbf{0}_{(M-r)\times r} & \mathbf{0}_{(M-r)\times(N-r)} \end{pmatrix}\right) \\ &= \sum_{i=1}^{r}\frac{\sigma_i^2(\boldsymbol{\Lambda})}{\sigma_i^2(\boldsymbol{\Lambda}) + \mu}.\end{aligned}$$

As a result, the preceding calculation led to $\lim_{\mu\to 0}\mathrm{tr}(\mathbf{P}_\mu(\boldsymbol{\Lambda})) = \mathrm{rank}(\boldsymbol{\Lambda})$, and the proof is completed. ■

REFERENCES

[1] D. Zhao, H. Lu, Y. Wang, H. Sun, and Y. Gui, "Joint power allocation and user association optimization for IRS-assisted mmWave systems," *IEEE Transactions on Wireless Communications*, vol. 21, no. 1, pp. 577–590, 2022.

[2] M. Zhang, H. Lu, F. Wu, and C. W. Chen, "NOMA-based scalable video multicast in mobile networks with statistical channels," *IEEE Transactions on Mobile Computing*, vol. 20, no. 6, pp. 2238–2253, 2021.

[3] S. A. Busari, K. M. S. Huq, S. Mumtaz, L. Dai, and J. Rodriguez, "Millimeter-wave massive MIMO communication for future wireless systems: A survey," *IEEE Communications Surveys Tutorials*, vol. 20, no. 2, pp. 836–869, 2018.

[4] Z. Chen, J. Tang, X. Y. Zhang, D. K. C. So, S. Jin, and K.-K. Wong, "Hybrid evolutionary-based sparse channel estimation for IRS-assisted mmWave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 21, no. 3, pp. 1586–1601, 2022.

[5] C. You, B. Zheng, and R. Zhang, "Wireless communication via double IRS: Channel estimation and passive beamforming designs," *IEEE Wireless Communications Letters*, vol. 10, no. 2, pp. 431–435, 2021.

[6] Y. Wang, H. Lu, and H. Sun, "Channel estimation in IRS-enhanced mmWave system with super-resolution network," *IEEE Communications Letters*, vol. 25, no. 8, pp. 2599–2603, 2021.

[7] X. Hu, R. Zhang, and C. Zhong, "Semi-passive elements assisted channel estimation for intelligent reflecting surface-aided communications," *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, pp. 1132–1142, 2022.

[8] B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface assisted multi-user OFDMA: Channel estimation and training design," *IEEE Transactions on Wireless Communications*, vol. 19, no. 12, pp. 8315–8329, 2020.

[9] Z. Chen, J. Tang, X. Y. Zhang, Q. Wu, Y. Wang, D. K. C. So, S. Jin, and K.-K. Wong, "Offset learning based channel estimation for intelligent reflecting surface-assisted indoor communication," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 1, pp. 41–55, 2022.

[10] Z. Wang, L. Liu, and S. Cui, "Channel estimation for intelligent reflecting surface assisted multiuser communications: Framework, algorithms, and analysis," *IEEE Transactions on Wireless Communications*, vol. 19, no. 10, pp. 6607–6620, 2020.

[11] P. Cai, J. Zong, X. Luo, Y. Zhou, S. Chen, and H. Qian, "Downlink channel tracking for intelligent reflecting surface-aided FDD MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 4, pp. 3341–3353, 2021.

[12] J.-M. Kang, "Intelligent reflecting surface: Joint optimal training sequence and refection pattern," *IEEE Communications Letters*, vol. 24, no. 8, pp. 1784–1788, 2020.

[13] C. Hu, L. Dai, S. Han, and X. Wang, "Two-timescale channel estimation for reconfigurable intelligent surface aided wireless communications," *IEEE Transactions on Communications*, vol. 69, no. 11, pp. 7736–7747, 2021.

[14] X. Guan, Q. Wu, and R. Zhang, "Anchor-assisted channel estimation for intelligent reflecting surface aided multiuser communication," *IEEE Transactions on Wireless Communications*, vol. 21, no. 6, pp. 3764–3778, 2022.

[15] N. K. Kundu and M. R. McKay, "A deep learning-based channel estimation approach for MISO communications with large intelligent surfaces," in *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1–6, 2020.

[16] S. Gao, P. Dong, Z. Pan, and G. Y. Li, "Deep multi-stage CSI acquisition for reconfigurable intelligent surface aided MIMO systems," *IEEE Communications Letters*, vol. 25, no. 6, pp. 2024–2028, 2021.

[17] A. M. Elbir and S. Coleri, "Federated learning for channel estimation in conventional and RIS-assisted massive MIMO," *IEEE Transactions on Wireless Communications*, vol. 21, no. 6, pp. 4255–4268, 2022.

[18] S. Liu, Z. Gao, J. Zhang, M. D. Renzo, and M.-S. Alouini, "Deep denoising neural network assisted compressive channel estimation for mmWave intelligent reflecting surfaces," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 9223–9228, 2020.

[19] M. Xu, S. Zhang, C. Zhong, J. Ma, and O. A. Dobre, "Ordinary differential equation-based CNN for channel extrapolation over RIS-assisted communication," *IEEE Communications Letters*, vol. 25, no. 6, pp. 1921–1925, 2021.

[20] L. Wei, C. Huang, G. C. Alexandropoulos, C. Yuen, Z. Zhang, and M. Debbah, "Channel estimation for RIS-empowered multi-user MISO wireless communications," *IEEE Transactions on Communications*, vol. 69, no. 6, pp. 4144–4157, 2021.

[21] S. Eddine Zegrar, L. Afeef, and H. Arslan, "Reconfigurable intelligent surface (RIS): Eigenvalue decomposition-based separate channel estimation," in *2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 1–6, 2021.

[22] H. Liu, X. Yuan, and Y.-J. A. Zhang, "Matrix-calibration-based cascaded channel estimation for reconfigurable intelligent surface assisted multiuser MIMO," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2621–2636, 2020.

[23] Z.-Q. He and X. Yuan, "Cascaded channel estimation for large intelligent metasurface assisted massive MIMO," *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 210–214, 2020.

[24] P. Wang, J. Fang, H. Duan, and H. Li, "Compressed channel estimation for intelligent reflecting surface-assisted millimeter wave systems," *IEEE Signal Processing Letters*, vol. 27, pp. 905–909, 2020.

[25] J. He, H. Wymeersch, and M. Juntti, "Channel estimation for RIS-aided mmWave MIMO systems via atomic norm minimization," *IEEE Transactions on Wireless Communications*, vol. 20, no. 9, pp. 5786–5797, 2021.

[26] Z. Wan, Z. Gao, and M.-S. Alouini, "Broadband channel estimation for intelligent reflecting surface aided mmWave massive MIMO systems," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, pp. 1–6, 2020.

[27] T. Lin, X. Yu, Y. Zhu, and R. Schober, "Channel estimation for intelligent reflecting surface-assisted millimeter wave MIMO systems," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, pp. 1–6, 2020.

[28] C. Jia, J. Cheng, H. Gao, and W. Xu, "High-resolution channel estimation for intelligent reflecting surface-assisted mmWave communications," in *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1–6, 2020.

[29] J. He, M. Leinonen, H. Wymeersch, and M. Juntti, "Channel estimation for RIS-aided mmWave MIMO systems," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, pp. 1–6, 2020.

[30] H. Zhao, R. Mayzus, S. Sun, M. Samimi, J. K. Schulz, Y. Azar, K. Wang, G. N. Wong, F. Gutierrez, and T. S. Rappaport, "28 GHz millimeter wave cellular communication measurements for reflection and penetration loss in and around buildings in New York city," in *2013 IEEE International Conference on Communications (ICC)*, pp. 5163–5167, 2013.

[31] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave

mobile communications for 5G cellular: It will work!," *IEEE Access*, vol. 1, pp. 335–349, 2013.

[32] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1164–1179, 2014.

[33] P. Wang, M. Pajovic, P. V. Orlik, T. Koike-Akino, K. J. Kim, and J. Fang, "Sparse channel estimation in millimeter wave communications: Exploiting joint AoD-AoA angular spread," in *2017 IEEE International Conference on Communications (ICC)*, pp. 1–6, 2017.

[34] X. Li, J. Fang, H. Li, and P. Wang, "Millimeter wave channel estimation via exploiting joint sparse and low-rank structures," *IEEE Transactions on Wireless Communications*, vol. 17, no. 2, pp. 1123–1133, 2018.

[35] J. Guo, Y. Sun, J. Gao, Y. Hu, and B. Yin, "Weighted nuclear norm minimization and its applications to low level vision," *International Journal of Computer Vision*, vol. 121, no. 2, pp. 183–208, 2016.

[36] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave mimo systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, 2014.

[37] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Transactions on Communications*, vol. 61, no. 10, pp. 4391–4403, 2013.

[38] M. Wang, F. Gao, N. Shlezinger, M. F. Flanagan, and Y. C. Eldar, "A block sparsity based estimator for mmWave massive MIMO channels with beam squint," *IEEE Transactions on Signal Processing*, vol. 68, pp. 49–64, 2020.

[39] T. Kim and D. J. Love, "Virtual AoA and AoD estimation for sparse millimeter wave MIMO channels," in *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 146–150, 2015.

[40] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm with application to wavelet-based image deblurring," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 693–696, 2009.

[41] X. Jia, X. Feng, and W. Wang, "Rank constrained nuclear norm minimization with application to image denoising," *Signal Processing*, vol. 129, pp. 1–11, 2016.

[42] C. Sun and R. Dai, "Rank-constrained optimization and its applications," *Automatica*, vol. 82, pp. 128–136, 2017.

[43] E. J. Candes and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2053–2080, 2010.

[44] Z. Chen, Y. Fu, Y. Xiang, and Y. Zhu, "A novel mr image denoising via lrma and nlss," *Signal Processing*, vol. 185, p. 108109, 2021.

[45] L. Mirsky, "A trace inequality of John von Neumann," *Monatshefte für Mathematik*, vol. 79, p. 303–306, 1975.

[46] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed ℓ^0 norm," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 289–301, 2009.

[47] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, pp. 183–202, 2009.

[48] S. Parthasarathy, S. Kumar, R. K. Ganti, S. Kalyani, and K. Giridhar, "Error vector magnitude analysis in generalized fading with co-channel interference," *IEEE Transactions on Communications*, vol. 66, pp. 345–354, 2018.