Two-Stage Channel Estimation for Reconfigurable Intelligent Surface-Assisted mmWave Systems

Jie Tang*, Xiaoyu Du*, Zhen Chen*, Xiuyin Zhang*, Kai-Kit Wong†, and Jonathon Chambers‡
*School of Electronic and Information Engineering, South China University of Technology, Guangzhou, China
†Department of Electronic and Electrical Engineering, University College London, London, UK
‡School of Engineering, University of Leicester, Leicester, UK
Email: eejtang@scut.edu.cn, dxy293953@163.com, chenz@scut.edu.cn, zhangxiuyin@scut.edu.cn,
k.wong@ee.ucl.ac.uk, and Jonathon.Chambers@le.ac.uk.

***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 channel 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 RIS-assisted channels 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, mmWave, compressed sensing, sparse and low-rank.**

I. Introduction

Millimeter wave (mmWave) is regarded as a potential technology for sixth-generation (6G) wireless communication to deal with increasingly scarce spectrum resources [1]. 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 [2]. Fortunately, 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 the incident signal to the desired direction with an adjustable phase shift [3]. 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 [4]. Firstly, the reflecting elements are generally passive, and there is a lack of signal processing capabilities, making traditional transmission training sequence methods inapplicable. Secondly, due to the vast number of reflecting elements deployed on a RIS, it is necessary to estimate the large-scale channel matrices, leading to a more complicated estimation process.

To overcome the above challenges in RIS-assisted systems, much research has attempted to design excellent estimation algorithms via various signal processing techniques, such as least squares (LS), minimum mean squared error (MMSE), deep learning, and matrix decomposition [5]. Especially, for the 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 potent tool for RIS channel estimation. By finding the sparse representation of cascaded channels, the channel estimation problem can be transformed into a sparse signal recovery problem and solved effectively by compressed sensing methods [6]. Inspired by this, our previous work has developed a hybrid multi-objective evolutionary paradigm and achieved high-resolution channel estimation effectively [3]. 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 [7]. Further, the mmWave channel was proved to exhibit joint sparse and low-rank characteristics in the presence of angular spreads, which can be exploited to improve estimation performance [8].

In this paper, we propose a RIS-assisted mmWave massive multiple-input multiple-output (MIMO) framework that combines sparsity and low-rank minimization for channel estimation by leveraging the spatial sparse structure. Specifically, we first set up a low-rank approximation model to reconstruct the noisy observed signal so as to satisfy the precondition of compressed sensing theory [8]. Different from the existing low-rank optimization schemes, we propose a trace operator as the cost function to replace the traditional nuclear norm operator, which can better approximate the rank of a matrix. Due to the powerful recognition of the trace norm and robust sparse representation abilities of the L1 norm, a joint trace and L1 norm minimization channel estimation scheme is formulated to achieve a performance improvement. Simulation results 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) and success ratio.

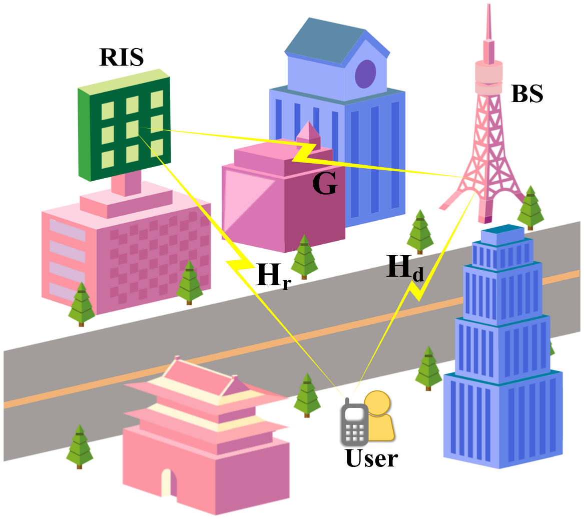

Fig. 1. The RIS-assisted uplink mmWave MIMO system.

II. System Model

Consider a RIS-assisted uplink mmWave MIMO system, as indicated in Fig. 1, which consists of one multi-antenna user, one multi-antenna base station (BS), and one RIS. Suppose there are N_{BS} and N_{US} antennas at the BS and user, respectively, and the RIS is equipped with $M \times M$ reflecting elements. 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, 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 the training stage, the symbol $s(t)$ is generally set as 1.

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 [9]. In addition, the introduction of RIS results in more complex channels and larger-scale channel 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, the channel estimation problem can be expressed as a sparse recovery problem, which can be solved easily. Furthermore, the mmWave channel also takes the form of angular spread in the AoA and AoD domains. The angular spread is caused by scattering clusters, presenting a structured sparsity pattern with each cluster possibly contributing multiple paths. Therefore, we will derive the channel model with angular spreads and discuss the joint sparse and low-rank characteristics.

Supposing the 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 model from user to BS can express as

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

where I is the number of paths, α_i denotes the gain of the i-th path, ν_i denotes the offset of the i-th path compared to the mean AoA θ, and ϕ is the AoD. In addition, $\boldsymbol{a}_{dA} \in \mathbb{C}^{N_A}$ and $\boldsymbol{a}_{dD} \in \mathbb{C}^{N_D}$ 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. 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,

$$\begin{aligned}\boldsymbol{H}_d = &\left(\sum_{i=1}^{I} \alpha_i \boldsymbol{a}_{dA}(\theta - \nu_i)\right) \boldsymbol{a}_{dD}^H(\phi - \varphi_1) \\ &+ \left(\sum_{i=1}^{I} \alpha'_i \boldsymbol{a}_{dA}(\theta - \nu_i)\right) \boldsymbol{a}_{dD}^H(\phi - \varphi_2).\end{aligned} \tag{5}$$

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

$$\boldsymbol{H}_d = \left(\sum_{i=1}^{I} \alpha_i \boldsymbol{a}_{dA}(\theta - \nu_i)\right) \left(\sum_{j=1}^{2} \boldsymbol{a}_{dD}^H(\phi - \varphi_j)\right). \tag{6}$$

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

$$\boldsymbol{H}_d=\sum_{l=1}^{L}\left(\sum_{i=1}^{I} \alpha_{l,i} \boldsymbol{a}_{dA}(\theta_l - \nu_{l,i})\right)\left(\sum_{j=1}^{J} \beta_{l,j} \boldsymbol{a}_{dD}^H(\phi_l - \varphi_{l,j})\right), \tag{7}$$

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.

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

$$\begin{aligned}\boldsymbol{H}_d &= \sum_{l=1}^{L} \boldsymbol{A}_{dA}\boldsymbol{\alpha}_l\boldsymbol{\beta}_l^T\boldsymbol{A}_{dD}^H\\ &= \boldsymbol{A}_{dA}(\sum_{l=1}^{L} \boldsymbol{\alpha}_l\boldsymbol{\beta}_l^T)\boldsymbol{A}_{dD}^H\\ &= \boldsymbol{A}_{dA}\boldsymbol{H}_{dv}\boldsymbol{A}_{dD}^H,\end{aligned} \tag{8}$$

where $\boldsymbol{\alpha}_l$ and $\boldsymbol{\beta}_l$ are virtual representations over the AoA and AoD domains, $\boldsymbol{A}_{dA} \triangleq [\boldsymbol{a}_{dA}(\theta_1),\cdots,\boldsymbol{a}_{dA}(\theta_{N_1})]$ 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, and $\boldsymbol{H}_{dv}$ is defined as 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 sparse characteristics, 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.

III. 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 observed value, we first carry out the process of low-rank matrix approximation. 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 two 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 [8]

$$\begin{aligned}\boldsymbol{Y}_{ij}&=(\boldsymbol{Z}^H(\boldsymbol{H}_d+\boldsymbol{G\Theta H}_r)\boldsymbol{F}+\boldsymbol{\Omega})_{ij}\\ &=(\boldsymbol{Z}^H\boldsymbol{H}_d\boldsymbol{F})_{ij}+(\boldsymbol{Z}^H\boldsymbol{G\Theta H}_r\boldsymbol{F})_{ij}+\boldsymbol{\Omega}_{ij},(i,j)\in\boldsymbol{\Upsilon},\end{aligned} \tag{9}$$

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 $\boldsymbol{Y}$. $\boldsymbol{\Omega}$ is the noise. $\boldsymbol{\Upsilon}$ represents the observed set, and we have $|\boldsymbol{\Upsilon}| = T$.

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

$$\begin{aligned}&\min_{\hat{\boldsymbol{Y}}} \mathrm{rank}(\hat{\boldsymbol{Y}})\\ &s.t.\left\|\boldsymbol{Y}-\hat{\boldsymbol{Y}}\right\|_F^2 \leq \varepsilon_1,\end{aligned} \tag{10}$$

where $\mathrm{rank}(\cdot)$ is the operator to calculate the rank of a matrix, $\|\cdot\|_F$ represents the Frobenius norm, and 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.

However, the solution obtained by minimizing the nuclear norm is usually biased and thus affects channel estimation accuracy. In order to overcome the deviation, we propose an improved rank operator $\mathrm{tr}(\mathbf{P}_\mu(\boldsymbol{\Lambda}))$ based on the projection matrix as

$$\mathbf{P}_\mu(\boldsymbol{\Lambda})=\boldsymbol{\Lambda}(\boldsymbol{\Lambda}^H\boldsymbol{\Lambda}+\mu\boldsymbol{I})^{-1}\boldsymbol{\Lambda}^H, \quad \mu \geq 0, \tag{11}$$

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

$$\lim_{\mu\to 0} \mathrm{tr}(\mathbf{P}_\mu(\boldsymbol{\Lambda})) = \mathrm{rank}(\boldsymbol{\Lambda}), \tag{12}$$

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

Proof: Please refer to the appendix for specific proof.

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

$$\begin{aligned}&\min_{\hat{\boldsymbol{Y}}} \mathrm{tr}(\mathbf{P}_\mu(\hat{\boldsymbol{Y}}))\\ &s.t.\left\|\boldsymbol{Y}-\hat{\boldsymbol{Y}}\right\|_F^2 \leq \varepsilon_1.\end{aligned} \tag{13}$$

The theorem indicates that when the coefficient μ approaches 0, the improved rank operator $\mathrm{tr}(\mathbf{P}_\mu(\hat{\boldsymbol{Y}}))$ can 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

$$\min_{\hat{\boldsymbol{Y}}} \left\{\frac{1}{2}\left\|\boldsymbol{Y}-\hat{\boldsymbol{Y}}\right\|_F^2 + \eta\mathrm{tr}\left(\mathbf{P}_\mu\left(\hat{\boldsymbol{Y}}\right)\right)\right\}. \tag{14}$$

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

$$\begin{aligned} g\left(\hat{\boldsymbol{Y}}\right) &= \tfrac{1}{2}\left\|\boldsymbol{Y}-\hat{\boldsymbol{Y}}\right\|_F^2 + \eta\mathrm{tr}\left(\mathbf{P}_\mu\left(\hat{\boldsymbol{Y}}\right)\right) \\ &= \tfrac{1}{2}\left\|\boldsymbol{U}\left(\boldsymbol{\Sigma}_{\boldsymbol{Y}}-\boldsymbol{\Phi}\right)\boldsymbol{V}^{\boldsymbol{H}}\right\|_F^2 + \eta\mathrm{tr}\left(\mathbf{P}_\mu\left(\hat{\boldsymbol{Y}}\right)\right) \\ &= \tfrac{1}{2}\left\|\boldsymbol{\Sigma}_{\boldsymbol{Y}}-\boldsymbol{\Phi}\right\|_F^2 + \eta\mathrm{tr}\left(\mathbf{P}_\mu\left(\boldsymbol{\Phi}\right)\right). \end{aligned} \tag{15}$$

Substituting (15), the problem (14) may be recast as

$$\min_{\boldsymbol{\Phi}}\left\{\frac{1}{2}\left\|\boldsymbol{\Sigma}_{\boldsymbol{Y}}-\boldsymbol{\Phi}\right\|_F^2 + \eta\mathrm{tr}\left(\mathbf{P}_\mu\left(\boldsymbol{\Phi}\right)\right)\right\}. \tag{16}$$

For the first portion of the problem (16), the following exists

$$\begin{aligned} \left\|\boldsymbol{\Sigma}_{\boldsymbol{Y}}-\boldsymbol{\Phi}\right\|_F^2 &= \left\|\boldsymbol{\Sigma}_{\boldsymbol{Y}}\right\|_F^2 + \left\|\boldsymbol{\Phi}\right\|_F^2 - 2\mathrm{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{Y}}\boldsymbol{\Phi}^H\right) \\ &\overset{(a)}{\geq} \left\|\boldsymbol{\Sigma}_{\boldsymbol{Y}}\right\|_F^2 + \left\|\boldsymbol{\Sigma}_{\boldsymbol{\Phi}}\right\|_F^2 - 2\mathrm{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{Y}}\boldsymbol{\Sigma}_{\boldsymbol{\Phi}}{}^H\right) \\ &= \left\|\boldsymbol{\Sigma}_{\boldsymbol{Y}}-\boldsymbol{\Sigma}_{\boldsymbol{\Phi}}\right\|_F^2 , \end{aligned} \tag{17}$$

where operation (a) is based on the von Neumann trace inequality [10], and $\boldsymbol{\Sigma}_{\boldsymbol{\Phi}}$ denotes the diagonal matrix of singular values derived by SVD of $\boldsymbol{\Phi}$.

We can deduce this further and get the following

$$\left\|\boldsymbol{\Sigma}_{\boldsymbol{Y}}-\boldsymbol{\Phi}\right\|_F^2 + \eta\mathrm{tr}\left(\mathbf{P}_\mu\left(\boldsymbol{\Phi}\right)\right) \geq \left\|\boldsymbol{\Sigma}_{\boldsymbol{Y}}-\boldsymbol{\Sigma}_{\boldsymbol{\Phi}}\right\|_{\mathrm{F}}^2 + \eta\mathrm{tr}\left(\mathbf{P}_\mu\left(\boldsymbol{\Phi}\right)\right), \tag{18}$$

where equality holds if and only if $\boldsymbol{\Phi} = \boldsymbol{\Sigma}_{\boldsymbol{\Phi}}$. Based on (18), problem (16) is equal to the following

$$\min_{\boldsymbol{\Sigma}_{\boldsymbol{\Phi}}}\left\{\frac{1}{2}\left\|\boldsymbol{\Sigma}_{\boldsymbol{Y}}-\boldsymbol{\Sigma}_{\boldsymbol{\Phi}}\right\|_F^2 + \eta\mathrm{tr}\left(\mathbf{P}_\mu\left(\boldsymbol{\Phi}\right)\right)\right\}. \tag{19}$$

Further, the Frobenius norm and Appendix A allow us to write the problem (19) as follows

$$\begin{aligned} &\min_{\boldsymbol{\Sigma}_{\boldsymbol{\Phi}}} \tfrac{1}{2}\left\|\boldsymbol{\Sigma}_{\boldsymbol{Y}}-\boldsymbol{\Sigma}_{\boldsymbol{\Phi}}\right\|_F^2 + \eta\mathrm{tr}\left(\mathbf{P}_\mu\left(\boldsymbol{\Phi}\right)\right) \\ &= \min_{\boldsymbol{\Sigma}_{\boldsymbol{\Phi}}} \tfrac{1}{2}\left\|\boldsymbol{\Sigma}_{\boldsymbol{Y}}-\boldsymbol{\Sigma}_{\boldsymbol{\Phi}}\right\|_F^2 + \eta\sum_{i=1}^{L}\tfrac{\sigma_i^2(\boldsymbol{\Phi})}{\sigma_i^2(\boldsymbol{\Phi})+\mu} \\ &= \min_{\boldsymbol{\Sigma}_{\boldsymbol{\Phi}}} \tfrac{1}{2}\left\|\boldsymbol{\Sigma}_{\boldsymbol{Y}}\right\|_F^2 + \tfrac{1}{2}\left\|\boldsymbol{\Sigma}_{\boldsymbol{\Phi}}\right\|_F^2 - \left\langle\boldsymbol{\Sigma}_{\boldsymbol{Y}}, \boldsymbol{\Sigma}_{\boldsymbol{\Phi}}\right\rangle_F + \eta\sum_{i=1}^{L}\tfrac{\sigma_i^2(\boldsymbol{\Phi})}{\sigma_i^2(\boldsymbol{\Phi})+\mu} \\ &= \min_{\sigma_1(\boldsymbol{\Phi}),\cdots,\sigma_L(\boldsymbol{\Phi})} \tfrac{1}{2}\sum_{i=1}^{L}\left(\sigma_i\left(\boldsymbol{\Sigma}_{\boldsymbol{Y}}\right)-\sigma_i\left(\boldsymbol{\Phi}\right)\right)^2 + \eta\sum_{i=1}^{L}\tfrac{\sigma_i^2(\boldsymbol{\Phi})}{\sigma_i^2(\boldsymbol{\Phi})+\mu}. \end{aligned} \tag{20}$$

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

$$\min_{\sigma_i(\boldsymbol{\Phi})}\left\{f\left(\sigma_i\left(\boldsymbol{\Phi}\right)\right) = \frac{1}{2}\left(\sigma_i\left(\boldsymbol{\Sigma}_{\boldsymbol{Y}}\right)-\sigma_i\left(\boldsymbol{\Phi}\right)\right)^2 + \eta\frac{\sigma_i^2\left(\boldsymbol{\Phi}\right)}{\sigma_i^2\left(\boldsymbol{\Phi}\right)+\mu}\right\}. \tag{21}$$

This is a scalar minimization problem, which is easy to solve. As a result, we can write the answer to (13) as

$$\hat{\boldsymbol{Y}} = \boldsymbol{U}\mathrm{diag}(\sigma_1\left(\boldsymbol{\Phi}\right), \cdots, \sigma_L\left(\boldsymbol{\Phi}\right))\boldsymbol{V}^H. \tag{22}$$

Inspired by [11], the initial value of the first loop is set to $\mu{=}4\min\limits_i|\sigma_i(\boldsymbol{Y})|$. 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 is summarized in Algorithm 1.

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

```
Input: The observed noisy signal $\boldsymbol{Y}_\Upsilon$.
Initialization: $\hat{\boldsymbol{Y}}^{(0)} = \boldsymbol{Y}_\Upsilon$, $c = 0.5$, $\rho$, $k = 1$ and $K$;
While $k < K$ do
    Iterative regularization:$\boldsymbol{Y}_\Upsilon^{(k)} = \hat{\boldsymbol{Y}}^{(k-1)} + \rho(\boldsymbol{Y}_\Upsilon - \hat{\boldsymbol{Y}}^{(k-1)})$;
    if $k = 1$
        $\mu = 4\min_i|\sigma_i(\boldsymbol{Y}_\Upsilon)|$;
    else
        Update the $\mu^{(k)} = c\mu^{(k-1)}$;
    end if
    Update the $\hat{\boldsymbol{Y}}^{(k)}$ via (13);
    $k = k + 1$;
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 (8), the beamspace model of user-BS, user-RIS, and RIS-BS channels can be written as

$$\begin{aligned} \boldsymbol{H}_d &= \boldsymbol{A}_{dA}\boldsymbol{H}_{dv}\boldsymbol{A}_{dD}^H, \\ \boldsymbol{H}_r &= \boldsymbol{A}_{rA}\boldsymbol{H}_{rv}\boldsymbol{A}_{rD}^H, \\ \boldsymbol{G} &= \boldsymbol{A}_{GA}\boldsymbol{H}_{Gv}\boldsymbol{A}_{GD}^H, \end{aligned} \tag{23}$$

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 (23) into the sampling model (9), we have

$$\begin{aligned} \hat{\boldsymbol{Y}} &= \boldsymbol{Z}^H(\boldsymbol{A}_{dR}\boldsymbol{H}_{dv}\boldsymbol{A}_{dD}^H)\boldsymbol{F} \\ &\quad + \boldsymbol{Z}^H(\boldsymbol{A}_{GR}\boldsymbol{H}_{Gv}\boldsymbol{A}_{GD}^H)\boldsymbol{\Theta}(\boldsymbol{A}_{rR}\boldsymbol{H}_{rv}\boldsymbol{A}_{rD}^H)\boldsymbol{F} + \boldsymbol{\Omega} \\ &= (\boldsymbol{Z}^H\boldsymbol{A}_{dR})\boldsymbol{H}_{dv}(\boldsymbol{A}_{dD}^H\boldsymbol{F}) \\ &\quad + (\boldsymbol{Z}^H\boldsymbol{A}_{GR})(\boldsymbol{H}_{Gv}\boldsymbol{A}_{GD}^H\boldsymbol{\Theta}\boldsymbol{A}_{rR}\boldsymbol{H}_{rv})(\boldsymbol{A}_{rD}^H\boldsymbol{F}) + \boldsymbol{\Omega} \\ &= \boldsymbol{C}_d\boldsymbol{H}_{dv}\boldsymbol{D}_d \\ &\quad + \boldsymbol{C}_c(\boldsymbol{H}_{Gv}\boldsymbol{A}_{GD}^H\boldsymbol{\Theta}\boldsymbol{A}_{rR}\boldsymbol{H}_{rv})\boldsymbol{D}_c + \boldsymbol{\Omega}, \end{aligned} \tag{24}$$

$$\begin{aligned} \mathrm{vec}(\hat{\boldsymbol{Y}}) &= \mathrm{vec}(\boldsymbol{C}_d\boldsymbol{H}_{dv}\boldsymbol{D}_d) + \mathrm{vec}(\boldsymbol{C}_c(\boldsymbol{H}_{Gv}\boldsymbol{A}_{GD}^H\boldsymbol{\Theta}\boldsymbol{A}_{rA}\boldsymbol{H}_{rv})\boldsymbol{D}_c) + \mathrm{vec}(\boldsymbol{\Omega}) \\ &= (\boldsymbol{D}_d^T\otimes\boldsymbol{C}_d)\mathrm{vec}(\boldsymbol{H}_{dv}) + (\boldsymbol{D}_c^T\otimes\boldsymbol{C}_c)\mathrm{vec}(\boldsymbol{H}_{Gv}\boldsymbol{A}_{GD}^H\boldsymbol{\Theta}\boldsymbol{A}_{rA}\boldsymbol{H}_{rv}) + \mathrm{vec}(\boldsymbol{\Omega}) \\ &= (\boldsymbol{D}_d^T\otimes\boldsymbol{C}_d)\mathrm{vec}(\boldsymbol{H}_{dv}) + (\boldsymbol{D}_c^T\otimes\boldsymbol{C}_c)(\boldsymbol{H}_{rv}^T\otimes\boldsymbol{H}_{Gv})\mathrm{vec}(\boldsymbol{A}_{GD}^H\boldsymbol{\Theta}\boldsymbol{A}_{rA}) + \mathrm{vec}(\boldsymbol{\Omega}) \\ &= (\boldsymbol{D}_d^T\otimes\boldsymbol{C}_d)\mathrm{vec}(\boldsymbol{H}_{dv}) + ((\mathrm{vec}(\boldsymbol{A}_{GD}^H\boldsymbol{\Theta}\boldsymbol{A}_{rA}))^T\otimes(\boldsymbol{D}_c^T\otimes\boldsymbol{C}_c))\mathrm{vec}(\boldsymbol{H}_{rv}^T\otimes\boldsymbol{H}_{Gv}) + \mathrm{vec}(\boldsymbol{\Omega}) \\ &= \left[\begin{array}{cc} (\boldsymbol{D}_d^T\otimes\boldsymbol{C}_d) & (\mathrm{vec}(\boldsymbol{A}_{GD}^H\boldsymbol{\Theta}\boldsymbol{A}_{rA}))^T\otimes(\boldsymbol{D}_c^T\otimes\boldsymbol{C}_c) \end{array}\right]\left[\begin{array}{c} \mathrm{vec}(\boldsymbol{H}_{dv}) \\ \mathrm{vec}(\boldsymbol{H}_{rv}^T\otimes\boldsymbol{H}_{Gv}) \end{array}\right] + \mathrm{vec}(\boldsymbol{\Omega}) \\ &= \boldsymbol{\psi}\boldsymbol{h} + \mathrm{vec}(\boldsymbol{\Omega}), \end{aligned} \tag{26}$$

with

$$\begin{aligned}\boldsymbol{C}_d = \boldsymbol{Z}^H\boldsymbol{A}_{dR}, \boldsymbol{D}_d = \boldsymbol{A}_{dD}^H\boldsymbol{F},\\ \boldsymbol{C}_c = \boldsymbol{Z}^H\boldsymbol{A}_{GR}, \boldsymbol{D}_c = \boldsymbol{A}_{rD}^H\boldsymbol{F}.\end{aligned} \tag{25}$$

Then, employing the Kronecker product operation and matrix vectorization operator to further reduce the complexity, we obtain (26) shown at the bottom of the last page, with

$$\boldsymbol{\psi}{=}\left[\left(\boldsymbol{D}_d^T\otimes\boldsymbol{C}_d\right)\ \left(\mathrm{vec}(\boldsymbol{A}_{GD}^H\boldsymbol{\Theta}\boldsymbol{A}_{rA})\right)^T{\otimes}(\boldsymbol{D}_c^T\otimes\boldsymbol{C}_c)\right], \tag{27}$$

$$\boldsymbol{h} = \begin{bmatrix}\mathrm{vec}(\boldsymbol{H}_{dv})\\ \mathrm{vec}(\boldsymbol{H}_{rv}^T\otimes\boldsymbol{H}_{Gv})\end{bmatrix}, \tag{28}$$

where $\otimes$ is the Kronecker product operation, and $\boldsymbol{I}$ is the identity matrix. Here, the channel estimation is converted into a sparse signal recovery problem,

$$\begin{aligned}&\min_{\boldsymbol{h}} \|\boldsymbol{h}\|_1\\ &s.t.\left\|\mathrm{vec}(\hat{\boldsymbol{Y}}) - \boldsymbol{\psi}\boldsymbol{h}\right\|_2 \leq \varepsilon_2,\end{aligned} \tag{29}$$

where ε_2 is the precise threshold.

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

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

```
Require: The observed noisy signal $\boldsymbol{Y}_{\boldsymbol{\Upsilon}}$ and the coefficient
    matrices $\boldsymbol{\psi}$.
1: Recover $\hat{\boldsymbol{Y}}$ based on the Algorithm 1 by solving
       $\min_{\hat{\boldsymbol{Y}}} \mathrm{rank}(\hat{\boldsymbol{Y}})$
       $s.t.\left\|\boldsymbol{Y}_{\boldsymbol{\Upsilon}} - \hat{\boldsymbol{Y}}\right\|_F^2 \leq \varepsilon_1,$
2: Estimate $\hat{\boldsymbol{h}}$ via
       $\min_{\boldsymbol{h}} \|\boldsymbol{h}\|_1$
       $s.t.\left\|\mathrm{vec}(\hat{\boldsymbol{Y}}) - \boldsymbol{\psi}\boldsymbol{h}\right\|_2 \leq \varepsilon_2.$
```

IV. Simulation Result

In this section, we perform simulations to evaluate the proposed two-stage channel estimation scheme. Considering a RIS-assisted uplink mmWave MIMO system (see Fig. 1), both the user and BS employ uniform linear array (ULA) antennas where the distance of adjacent units is half the signal wavelength. RIS is a square panel consisting of $M \times M$ uniform rectangular arrays. We set $M = 16$, $N_{BS} = 32$ and $N_{US} = 16$ in the general case. The mmWave channels are generated by the geometric channel in (7). Taking the sparse scattering of mmWave into account, the number of clusters in each transmission link is set to $L = 2$ [8]. 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.

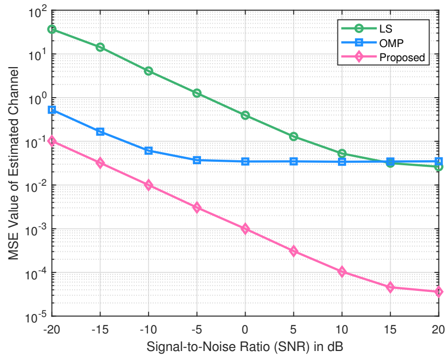

Fig. 2. The MSE performance versus SNR.

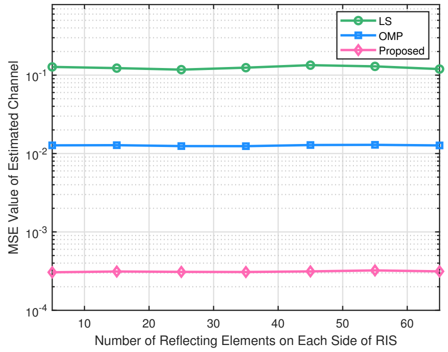

Fig. 3. The MSE performance versus the number of RIS reflecting elements.

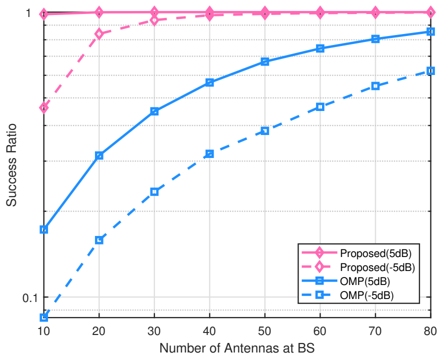

Fig. 4. The success ratio versus the number of antennas at BS.

The beamforming/combining matrices $\boldsymbol{F}$ and $\boldsymbol{Z}$ are randomly selected in the experiment. In addition, the classical LS and orthogonal matching pursuit (OMP) methods are chosen as comparisons, and the sparsity K is set to 10. 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. The formula is shown as

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{h}_i - \hat{\boldsymbol{h}}_i\right)^2, \tag{30}$$

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 performance of the estimation method.

We first examine the estimation accuracy varying with signal-to-noise ratio (SNR), as illustrated in Fig. 2. We can observe that the MSEs of all schemes improve 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. The proposed scheme reconstructs the noisy received signal by low-rank matrix approximation, fitting in with the precondition of the compressed sensing theory, thus bringing more obvious performance improvement.

The increase of reflecting elements brings a growing coefficient matrix, making the channel estimation more complicated. With M representing the number of elements on either side, the experimental results with the SNR of 5dB are plotted against M. As the number of elements grows, Fig. 3 indicates that the MSEs of the three schemes stay almost constant. However, the proposed two-stage scheme consistently maintains the best estimation accuracy. The results demonstrate that the proposed scheme is robust and can perform accurate channel estimation under large-scale channel matrices, even if the reflecting element is significantly increased.

Given the severe path loss in mmWave channels, the value of SNR before beam alignment is typically low, sometimes below 0dB. To track the success ratio under the low SNR region, we set SNR to 5dB and -5dB. Specifically, when MSE is less than 10^{-2}, the trial is marked as a success. The success ratio is defined as the ratio of the number of successful trials N_{succ} to the total number of all trials N_{total}, that is, N_{succ}/N_{total}. The numerical results in Fig. 4 demonstrate that the accuracy is improved as the number of antennas increases, and the success ratios exhibit upward trends. The proposed scheme consistently maintains a high accuracy of almost 100% when SNR is set to 5dB. Even with SNR=-5dB, the success ratio is greater than 90% with 30 antennas and 99.49% with 80 antennas. The consistently high success ratio 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.

V. 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 consisted of sparse and low-rank characteristics. In the first stage, we utilized low-rank characteristics to reconstruct the noisy observed 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 two-stage scheme could effectively perform accurate channel estimation and was robust for different channel environments.

Appendix A
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

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

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

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

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] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] L. Mirsky, “A trace inequality of john von neumann,” *Monatshefte für Mathematik*, vol. 79, no. 4, pp. 303–306, 1975.

[11] 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.