Channel Estimation for Reconfigurable Intelligent Surface Aided Multi-User MIMO Systems

Jie Chen, *Student Member, IEEE*, Ying-Chang Liang, *Fellow, IEEE*, Hei Victor Cheng, *Member, IEEE*, and Wei Yu, *Fellow, IEEE*

Abstract—Channel acquisition is one of the main challenges for the deployment of reconfigurable intelligent surface (RIS) aided communication system. This is because RIS has a large number of reflective elements, which are passive devices without active transmitting/receiving and signal processing abilities. In this paper, we study the uplink channel estimation for the RIS aided multi-user multi-input multi-output (MIMO) system. Specifically, we propose a novel channel estimation protocol for the above system to estimate the cascade channel, which consists of the channels from the base station (BS) to the RIS and from the RIS to the user. Further, we recognize the cascaded channels are typically sparse, this allows us to formulate the channel estimation problem into a sparse channel matrix recovery problem using the compressive sensing (CS) technique, with which we can achieve robust channel estimation with limited training overhead. In particular, the sparse channel matrixes of the cascaded channels of all users have a common row-column-block sparsity structure due to the common channel between BS and RIS. By considering such a common sparsity, we further propose a two-step procedure based multi-user joint channel estimator. In the first step, by considering common column-block sparsity, we project the signal into the common column subspace for reducing complexity, quantization error, and noise level. In the second step, by considering common row-block sparsity, we apply all the projected signals to formulate a multi-user joint sparse matrix recovery problem, and we propose an iterative approach to solve this non-convex problem efficiently. Moreover, the optimization of the training reflection sequences at the RIS is studied to improve the estimation performance.

Index Terms—Reconfigurable intelligent surface, common row-column sparsity, multi-user joint channel estimation, compressive sensing.

I. INTRODUCTION

Reconfigurable intelligent surface (RIS) is a promising technique to achieve high spectrum- and energy- efficiency [1]–[4]. Specifically, RIS is a uniform planar array with a large number of reflective elements, each of which can induce a phase shift of the incident signal and reflect it passively. Hence, by adaptively adjusting the phase shift matrix of RIS, it can enhance the transmission quality of the intended incident-reflection signal [5], which is also called as passive beamforming [6]. Compared with traditional *amplify-and-forward* (AF) relay beamforming techniques [7], [8], RIS can reconfigure the reflective coefficients in real-time and reflect the incident signal passively without additional energy consumption [9]. Besides, RIS can be equipped with a large number of reflective elements for achieving high array/passive beamforming gain without requiring much hardware cost [10].

Due to the above promising advantages, RIS has been introduced into various wireless communication systems. In particular, the key design issue of the RIS aided wireless communication system is to jointly optimize the beamformer at the transceiver and the phase shift matrix induced by RIS to achieve various objectives [11]–[19]. Specifically, for a downlink multi-user *multiple-input multiple-output* MIMO systems [11]–[16], the energy-efficiency maximization was studied in [11] subject to the individual *Quality-of-Service* (QoS) constraint. In [12], the minimum *signal-to-interference-plus-noise ratio* (SINR) subject to a maximum power constraint was studied by considering both rank-one and full-rank channel matrix between the *base station* (BS) and the RIS. Then, the weighted sum-rate maximization problems were studied in [13] for a single-cell scenario and in [14] for a multi-cell scenario, respectively. Moreover, the downlink achievable rate maximization problem was studied in [19] for wideband *orthogonal frequency division multiplexing* (OFDM) system. Besides, the channel capacity optimization problems were studied in [16] with single RIS and in [15] with multiple RISs, respectively, and then extended into *millimeter-wave* (mmWave) environment in [17], [18].

However, the above studies [11]–[19] focus on the joint design of the beamformer at the BS and the phase shift matrix induced by RIS under the assumption that *channel state information*s (CSIs) are perfectly known at the BS, which is not practical for the RIS aided wireless communication system. Compared with the traditional active devices (i.e., AF relay) aided communication systems, the channel estimation in the RIS aided system is quite a challenging problem. This is because in the active devices aided communication system, the CSI can be estimated by enabling the active devices to send training sequences. However, RIS is a passive device with a large number of passive reflective elements, which cannot perform active transmitting/receiving and signal processing. Thus, the CSIs with such a large number of unknown parameters can only be estimated at the active BS or users, which makes the channel estimation quite difficult [4]. This motivates us to find innovative channel estimation methods to tackle these new challenges.

Recently, there are some literature investigating the channel estimation for the RIS aided single user communication sys-

J. Chen and Y.-C. Liang are with the Center for Intelligent Networking and Communications (CINC), University of Electronic Science and Technology of China (UESTC), Chengdu 611731, China (e-mails: chenjie.ay@gmail.com, liangyc@ieee.org).

H. Cheng and W. Yu are with the Electrical and Computer Engineering Department, University of Toronto, Toronto, ON M5S 3G4, Canada (e-mails: hei.cheng@utoronto.ca, weiyu@ece.utoronto.ca).

tems [20]–[25]. Specifically, the binary reflection method was proposed in [20], [21], where the RIS turns on each reflective element successively, while keeping the rest reflective elements closed. Then, the BS successively estimates the cascaded channel, which consists of the channels from the BS to one typical reflective element and from this element to the users. In [22], a minimum variance unbiased channel estimator was proposed by turning on all reflective elements in the entire training period, where the optimal phase shift matrix induced by RIS was shown to be a discrete Fourier transform matrix. This method was further extended in [23], [24], where the authors assume that the surface can be divided into multiple sub-surface, where each sub-surface consists of some adjacent reflective elements sharing the common reflection coefficient. However, the training overhead of the above methods in [20]–[24] scales up with the number of reflective elements (or sub-surfaces), which causes intractable training overhead and degrades the spectrum-efficiency. In [25], some active elements are randomly deployed at the RIS to perform channel estimation. Then, the full CSIs are recovered by using the estimated CSIs from the active elements due to channel sparsity. This method, in fact, can reduce the training overhead, but it also increases the hardware cost and complexity due to the deployment of active elements.

Motivated by the above reasons, in this paper, we study the channel estimation for the RIS aided multi-user MIMO communication system operated in *time division duplex* (TDD) mode. To highlight the main contributions, we summarize the paper as follows:

- We propose a novel uplink channel estimation protocol and apply compressive sensing (CS) technique to estimate the cascaded channels of the RIS aided multi-user MIMO system with limited training overhead. Specifically, we first investigate the sparsity representation of the cascaded channels. Since the BS and the RIS are usually mounted at a height, there are only limited scatters around the BS and the RIS. This indicates that the cascaded channel has only a few *angle of arrival* (AoA) and *angle of departure* (AoD) array steering vectors, and thus it can be represented by a row-column-block sparse channel matrix. This specific sparsity structure is quite different from the conventional (mmWave) MIMO communication systems, whose sparse channel matrix is usually just row-block sparsity, because there are only limited scatters at the BS but rich scatters at the receivers [26]–[28].

- We further find that the cascaded channels of all users have the common row-column-block sparsity structure due to the common channel from the BS to the RIS. However, the conventional CS-based channel estimators, i.e., *single measurement vector* (SMV) [29] and *multiple measurement vectors* (MMV) [26]–[28], recover the sparse channel matrix of each user individually without considering the specific common row-column-block sparsity. Hence, these estimators usually require more training overhead to guarantee recovery performance when the sparsity level (the number of spatial paths) increases.

- To avoid the drawbacks of the conventional estimators,

we apply the common row-column-block sparsity to jointly estimate the cascaded channels. Specifically, if we recover the sparse matrix by considering the common row-column-block sparsity simultaneously, we need to quantize the AoA and AoD with high resolutions to reduce the quantization errors caused by the discrete grid of AoAs/AoDs. This leads to intractable computational complexity. To further deal with this issue, we propose the following two-step procedure based multi-user joint channel estimator. In particular, in the first step, we exploit the common column-block sparsity to estimate the common subspace spanned by AoD array steering vectors. Then, we project our received signals into this common AoD subspace, which can reduce the number of zero columns of the original sparse matrix and transform it as a row-block sparsity matrix. Hence, this procedure can achieve lower complexity due to less unknown columns in the sparse matrix, lower quantization error due to without quantizing AoD, and higher SNR due to reducing the influence of noise on the null space of common AoD subspace. In the second step, we exploit the common row-block sparsity to formulate a MMV-based multi-user joint sparse matrix recovery problem. Since we use the received signals of all users to recover the sparse matrix jointly, a better recovery performance can be achieved.

- Since the optimization variables are coupled in the formulated multi-user joint sparse matrix recovery problem, which is non-convex and hard to solve, we propose an approach based on the principles of alternative optimization and iterative reweighted algorithm to solve it efficiently. Besides, we analyze the convergence of the proposed algorithm. Moreover, we design a training reflection coefficient sequence optimization method based on minimizing the mutual coherence of the equivalent dictionary. Finally, the simulation results validate the effectiveness of the proposed estimation scheme.

The rest of this paper is organized as follows. Section II presents the system model and channel estimation protocol for the RIS aided multi-user MIMO system. Section III investigates the common sparsity of the cascaded channels, and shows the drawbacks of the conventional CS-based techniques. Section IV studies the two-step procedure based multi-user joint channel estimator and Section V provides its detail solution. Section VI studies the training reflection coefficient optimization method. Section VII provides simulation results to validate the effectiveness of the proposed scheme. Finally, Section VIII concludes the paper.

Notations: The scalar, vector, and matrix are lowercase, bold lowercase, and bold uppercase, i.e., a, $\boldsymbol{a}$, and $\boldsymbol{A}$, respectively. $(\cdot)^T$, $(\cdot)^H$, $\mathrm{Tr}\,(\cdot)$, and $\mathrm{rank}\,(\cdot)$ denote transpose, conjugate transpose, trace, and rank, respectively. $\boldsymbol{A}^{\dagger}$ is the Moore-Penrose pseudoinverse matrix. $[\boldsymbol{a}]_i$ and $\boldsymbol{A}^{i,j}$ denote the i-th component of vector $\boldsymbol{a}$ and i-th row j-th column component of matrix $\boldsymbol{A}$, respectively. $\boldsymbol{A}^{:,i}$ and $\boldsymbol{A}^{:,\Omega}$ denote the i-th column vector of matrix $\boldsymbol{A}$ and the sub-matrix consisting of the columns of matrix $\boldsymbol{A}$ with indices in set Ω. $\boldsymbol{I}_M$ denotes the M-by-M identity matrix and $\boldsymbol{1}_{M \times N}$ denotes M-
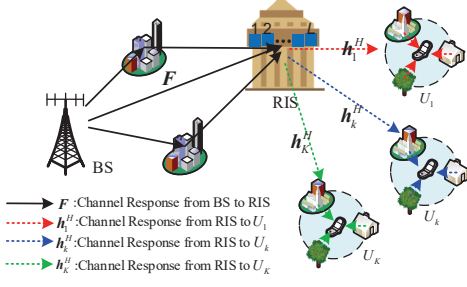
Fig. 1: A RIS aided multi-user MIMO system consisting of one BS with M antennas, one RIS with L reflective elements, and K single-antenna users.
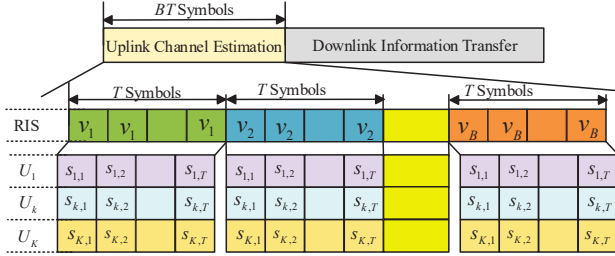

Fig. 2: Channel estimation protocol and frame structure.

by-N matrix whose elements are equal to 1. $\mathrm{span}(\boldsymbol{A})$ is the space spanned by the column vectors of matrix $\boldsymbol{A}$. Besides, $\mathcal{CN}\left(\mu, \sigma^2\right)$ denotes the distribution of *circularly symmetric complex Gaussian* (CSCG) with mean μ and variance σ^2.

II. SYSTEM MODEL

A. Cascaded Channel Model

As shown in Fig. 1, we consider a block-fading multi-user MIMO system operated in TDD mode, where a BS with the assistance of RIS serves K users. The BS and the RIS are equipped with M antennas and L reflective elements, respectively, while the users are all equipped with a single antenna each. The users are denoted by $U_1, \cdots, U_K$. The channel responses from the BS to the RIS and from the RIS to U_k are denoted by $\boldsymbol{F} \in \mathbb{C}^{L \times M}$ and $\boldsymbol{h}_k^H \in \mathbb{C}^{1 \times L}$, respectively. The reflective channel at the RIS is usually referred to as the dyadic backscatter channel, where each reflective element combines all the received signals and then transmits them to U_k acting as a point source by reflection. Thus, the reflection coefficient channel matrix [13] is given by $\boldsymbol{V}_o = \mathrm{diag}(\boldsymbol{v}) \in \mathbb{C}^{L \times L}$ with $\boldsymbol{v} = [v_1, v_2, \cdots, v_L]^T \in \mathbb{C}^{L \times 1}$ where $v_l = e^{j\vartheta_l}$ is the reconfigurable reflection coefficient on the l-th reflective element.

Next, the whole channel response from BS to U_k through RIS is denoted by $\boldsymbol{h}_k^H \boldsymbol{V}_o \boldsymbol{F} \in \mathbb{C}^{1 \times M}$. Hence, the transmission quality of the intended incident-reflection signal can be enhanced by adaptively adjusting $\boldsymbol{V}_o$ at the RIS. However, the joint design of beamformer at the BS and reflection coefficient matrix $\boldsymbol{V}_o$ at the RIS for downlink data transmission requires the CSIs of $\boldsymbol{h}_k^H$ and $\boldsymbol{F}$, simultaneously. Moreover, RIS has a large number of reflective elements, which are passive devices without active transmitting/receiving and signal processing

abilities. Hence, it is quite a challenging problem to achieve channel estimation in the studied system.

Therefore, we propose the following innovative method to tackle these new challenges. Specifically, due to $\boldsymbol{V}_o = \mathrm{diag}(\boldsymbol{v})$ and $\boldsymbol{h}_k^H \mathrm{diag}(\boldsymbol{v}) = \boldsymbol{v}^T \mathrm{diag}(\boldsymbol{h}_k^H)$, it is straightforward to know that the joint beamformer and reflection coefficient design is only dependent on the CSI of the following cascaded channel [4], [10]:

$$\boldsymbol{G}_k = \mathrm{diag}(\boldsymbol{h}_k^H)\boldsymbol{F} \in \mathbb{C}^{L \times M}. \tag{1}$$

In the subsequent parts, we propose a novel uplink channel estimation protocol for the studied system and use the *least square* (LS) method [22] to estimate the cascaded channel $\boldsymbol{G}_k$.

B. Channel Estimation Protocol

For TDD systems, the CSI of the downlink channel can be obtained by estimating the CSI of the uplink channel due to channel reciprocity. Since $L \geq M \geq K$ exists in most wireless communications, in this paper, we propose the following uplink channel estimation protocol.

To begin with, note that the whole channel response from BS to U_k can be expressed by $\boldsymbol{v}^T \mathrm{diag}(\boldsymbol{h}_k^H)\boldsymbol{F} = \boldsymbol{v}^T \boldsymbol{G}_k \in \mathbb{C}^{1 \times M}$. Hence, in order to separate $\boldsymbol{v}$ and $\boldsymbol{G}_k$ from the received training signals, we need to obtain enough individual observations with different reflection coefficient $\boldsymbol{v}$. Moreover, we need to enable each user to transmit orthogonal pilot sequence to separate $\boldsymbol{G}_k$ from each user without interference.

Inspired from the above characteristics, we propose the estimation protocol as shown in Fig. 2. Specifically, the frame structure is divided into two phases, i.e., Phase-I for uplink channel estimation and Phase-II for downlink data transmission. In this paper, we only focus on the uplink channel estimation of $\boldsymbol{G}_k$ in Phase-I. In particular, Phase-I consists of B sub-frames and each sub-frame consists of T symbol durations ($T \geq K$). Specifically, the RIS keeps T copies of the training reflection coefficient vector $\boldsymbol{v}_b = [v_{b,1}, v_{b,2}, \cdots, v_{b,L}]^H \in \mathbb{C}^{L \times 1}$ in the b-th sub-frame, and then adjusts the value of reflection coefficient to make it different in each sub-frame. U_k transmits B copies of the k-th orthogonal pilot sequence in B sub-frames, where each pilot sequence is with length T, i.e., $\boldsymbol{s}_k^H = [s_{k,1}, s_{k,2}, \cdots, s_{k,T}] \in \mathbb{C}^{1 \times T}$ with $\boldsymbol{s}_{k_1}^H \boldsymbol{s}_{k_2} = 0$ for $1 \leq k_1, k_2 \leq K$ and $k_1 \neq k_2$.

Specifically, in the b-th sub-frame, the received T pilot signals at the BS, i.e., $\boldsymbol{Y}_b \in \mathbb{C}^{M \times T}$, can be written as

$$\boldsymbol{Y}_b = \sum_{k=1}^{K} \boldsymbol{F}^H \mathrm{diag}(\boldsymbol{v}_b)\boldsymbol{h}_k \boldsymbol{s}_k^H + \boldsymbol{U}_b$$
$$\overset{(a)}{=} \sum_{k=1}^{K} \boldsymbol{G}_k^H \boldsymbol{v}_b \boldsymbol{s}_k^H + \boldsymbol{U}_b, \tag{2}$$

where (a) is because of $\boldsymbol{G}_k = \mathrm{diag}(\boldsymbol{h}_k^H)\boldsymbol{F}$. Note that $\boldsymbol{s}_k^H \boldsymbol{s}_k = PT$ is the transmit energy constraints for training sequence of U_k and P is the transmit power of each user. $\boldsymbol{U}_b \in \mathbb{C}^{M \times T}$ is received Gaussian noise with assuming $\boldsymbol{U}_b \sim \mathcal{CN}\left(0, \delta^2 \boldsymbol{I}_M\right)$.

C. Conventional LS Estimator

With the above estimation protocol, we can apply the conventional LS estimator to estimate the cascaded channel. Specifically, since all s_k^H in (2) are orthogonal pilot sequences, we have

$$\tilde{y}_{b,k} \triangleq \frac{1}{PT} Y_b s_k = G_k^H v_b + u_{b,k}, \tag{3}$$

where $u_{b,k} = \frac{1}{PT} U_b s_k \in \mathbb{C}^{M \times 1}$. Let $\widetilde{Y}_k = [\tilde{y}_{1,k}, \tilde{y}_{2,k}, \cdots, \tilde{y}_{B,k}] \in \mathbb{C}^{M \times B}$, $V = [v_1, v_2, \cdots, v_B] \in \mathbb{C}^{L \times B}$, and $\widetilde{U}_k = [u_{1,k}, u_{2,k}, \cdots, u_{B,k}] \in \mathbb{C}^{M \times B}$. Then, we can rewrite (3) into the following matrix form:

$$\widetilde{Y}_k = G_k^H V + \widetilde{U}_k. \tag{4}$$

Using the conventional LS channel estimator [30], the cascaded channel is estimated by

$$\hat{G}_k = (\widetilde{Y}_k V^{\dagger})^H, \tag{5}$$

where $V^{\dagger} = V^H (V V^H)^{-1}$.

It is worth noting that the above LS estimator in (5) requires $B \geq L$. Then, it causes intractable training overhead when the RIS is equipped with a large number of reflective elements. Therefore, it motivates us to investigate the efficient channel estimation schemes to reduce the training overhead.

III. CASCADED CHANNEL SPARSITY MODEL

In this section, we recognize the cascaded channels are typically sparse, which motivates us to apply CS-based techniques to achieve robust channel estimation with limited training overhead. Specifically, we first investigate the cascaded channel sparsity representation for the individual user. Then, we analyze the specific sparsity structure of the cascaded channel and the drawbacks of the conventional CS-based techniques in the studied scenario. Finally, we study the common sparsity structure for the further multi-user joint channel estimator design.

A. Individual User Channel Sparsity Representation

In this part, we investigate the sparsity representation of the cascaded channel in the RIS aided communication system.

Assume BS and RIS are each equipped with a *uniform linear array* (ULA). By applying the physical propagation structure of wireless channel [26], the channels h_k and F are given by

$$F = \sqrt{\frac{LM}{N_f}} \sum_{p=1}^{N_f} \alpha_p a_L(\frac{2\varpi}{\rho} \sin(\phi_p^{\text{AoA}})) a_M^H(\frac{2\varpi}{\rho} \sin(\phi_p^{\text{AoD}})), \tag{6}$$

$$h_k = \sqrt{\frac{L}{N_{h,k}}} \sum_{q=1}^{N_{h,k}} \beta_{k,q} a_L\left(\frac{2\varpi}{\rho} \sin(\varphi_{k,q})\right), \tag{7}$$

respectively, where α_p and $\beta_{k,q}$ denote the complex gains of p-th spatial path between the BS and the RIS and q-th spatial path between the RIS and U_k, respectively. ϕ_p^{AoD} and ϕ_p^{AoA} are the p-th AoD from the BS and AoA to the RIS, respectively, and φ_q is the q-th AoD from the RIS to U_k. N_f is the number of spatial paths between the BS and the RIS, and $N_{h,k}$ is

the number of spatial paths between the RIS and U_k. ϖ is the antenna spacing and ρ is the carrier wavelength, and we set $\varpi/\rho = 1/2$ for simplicity. $a_X(\varphi) \in \mathbb{C}^{X \times 1}$ is the array steering vector, i.e.,

$$a_X(\varphi) = \frac{1}{\sqrt{X}}[1, e^{j\pi\varphi}, \cdots, e^{j\pi\varphi(X-1)}]^H. \tag{8}$$

From (6) and (7), $G_k = \text{diag}(h_k^H) F$ can be rewritten as

$$G_k = \sum_{p=1}^{N_f} \sum_{q=1}^{N_{h_k}} \sqrt{\frac{L^2 M}{N_f N_{h,k}}} \alpha_p \beta_{k,q} a_L(\sin(\varphi_{k,q}) - \sin(\phi_p^{\text{AoD}}))$$
$$\times a_M^H(\sin(\theta_p^{\text{AoD}})). \tag{9}$$

Note that $a_L(\sin(\varphi_{k,q}) - \sin(\phi_p^{\text{AoD}}))$ is the (p, q)-th cascaded AoA/AoD at RIS/user side, which is called as the cascaded AoA in the following contents for simplicity.

To design the CS-based channel estimator, we approximate the cascaded channel in (9) by using the *virtual angular domain* (VAD) representation, i.e.,

$$G_k = A_R X_k A_T^H, \tag{10}$$

where $A_R \in \mathbb{C}^{L \times G_r}$ and $A_T \in \mathbb{C}^{M \times G_t}$ are the dictionary matrices for the angular domain with angular resolutions G_r and G_t, respectively, i.e., each column of A_R and A_T represent the array steering vectors (array steering vectors) corresponding to one specific cascaded AoA at RIS/user side and one specific AoD at the BS side, respectively[1], i.e.,

$$A_R = \left[a_L(-1), a_L(-1 + \tfrac{2}{G_r}), \cdots a_L(1 - \tfrac{2}{G_r}) \right], \tag{11}$$

$$A_T = \left[a_M(-1), a_M(-1 + \tfrac{2}{G_t}), \cdots a_M(1 - \tfrac{2}{G_t}) \right]. \tag{12}$$

Besides, $X_k \in \mathbb{C}^{G_r \times G_t}$ is the angular domain sparse channel matrix, where the non-zero (i, j)-th component corresponds to the complex gain on the channel consisting of the i-th cascaded AoA array steering vector at RIS/user side and the j-th AoD array steering vector at the BS side.

B. Sparsity Structure Analysis and Related CS-based Techniques

In this part, we analyze the sparsity structure of the cascaded channel and show the drawbacks of applying the conventional CS-based techniques to recover the sparse channel matrix X_k in the RIS aided communications.

1) Sparsity Structure Analysis: Since both BS and RIS are usually mounted at a height to assist the communications between the BS and users, there are only limited scatters around the BS and the RIS. This indicates that there are only a few AoDs and cascaded AoAs, i.e., both N_f and N_{h_k} should be small. Hence, X_k has the row-column-block sparsity structure, i.e., there are only a few column/row vectors in X_k being non-zero, as shown in Fig. 3. This specific structure of RIS aided communication system is quite different from the conventional (mmWave) MIMO communication system,

---

[1] Although $\sin(\varphi_{k,q}) - \sin(\phi_p^{\text{AoD}}) \in [-2, 2]$, it is straightforward to know that we only need to quantize $\sin(\varphi_{k,q}) - \sin(\phi_p^{\text{AoD}})$ in the domain $[-1, 1]$ due to the term of $e^{-j\pi(\sin(\varphi_{k,q}) - \sin(\phi_p^{\text{AoD}}))}$.
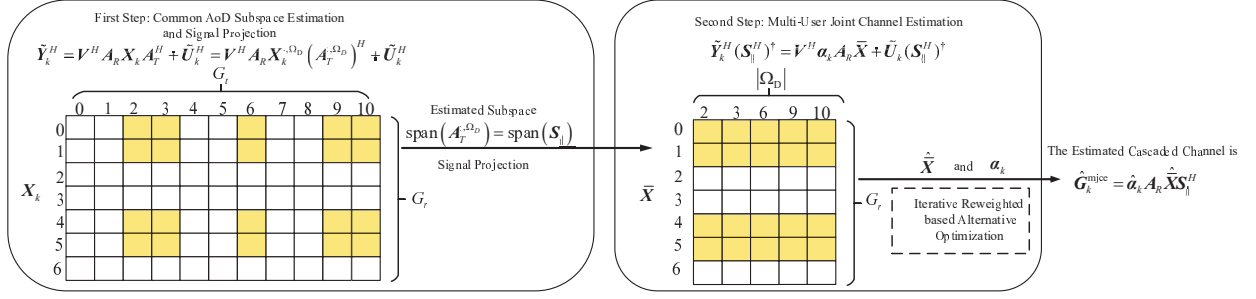
Fig. 3: Two-step procedure based multi user joint channel estimation.

whose sparse channel matrix is usually just row-block sparsity. This is because there are only limited scatters at the BS but rich scatters at the users in the conventional (mmWave) communications [26]–[28]. Then, the sparse channel matrix consists of only a few AoD array steering vectors but each corresponds to all AoA array steering vectors, thus it has a row-block sparsity structure.

2) Related CS-based Techniques: From (4) with the obtained sparsity channel representation of the cascaded channels (10), the received signal can be rewritten as

$$\widetilde{\boldsymbol{Y}}_k^H = \boldsymbol{V}^H \boldsymbol{G}_k + \widetilde{\boldsymbol{U}}_k^H = \boldsymbol{V}^H \boldsymbol{A}_R \boldsymbol{X}_k \boldsymbol{A}_T^H + \widetilde{\boldsymbol{U}}_k^H. \quad (13)$$

With (13), we can formulate the channel estimation problem into a sparse channel matrix $\boldsymbol{X}_k$ recovery problem using the CS-based techniques. One straightforward approach is to ignore the block sparsity structure of the sparse channel matrix, and vectorize the signal $\widetilde{\boldsymbol{Y}}_k^H$ and sparse channel matrix $\boldsymbol{X}_k$ into vectors directly. So, it can be formulated into a conventional SMV-based recovery problem, which can be solved efficiently by applying *orthogonal matching pursuit* (OMP) algorithm [27].

However, the above SMV-based channel estimator has the following disadvantages: 1) This estimator usually requires more training overhead to guarantee the estimation performance when the sparsity level increases due to ignoring the block sparsity structure . 2) This estimator needs to discrete both AoAs and AoDs on the grid due to the VAD representation in (10). Hence, in order to reduce the quantization errors of the AoA and AoD, both G_r and G_t should be much larger than the number of antennas at the BS or reflective elements at the RIS, which introduces intractable computational complexity.

Instead of ignoring the block sparsity, another better method is to ignore the only column-block sparsity, and use $\widetilde{\boldsymbol{X}}_k = \boldsymbol{X}_k \boldsymbol{A}_T^H \in \mathbb{C}^{G_r \times M}$ to represent a new sparse channel matrix. Thus, $\widetilde{\boldsymbol{X}}_k$ has a row-block sparsity structure and can be formulated into a conventional MMV-based recovery problem [26]–[28], which can be solved efficiently by applying simultaneous OMP (SOMP) algorithm [26], [27], as shown in Algorithm 1. Particularly, this estimator only needs to recover AoA array steering vectors with $\widetilde{\boldsymbol{X}}_k$, where the dimension of optimization variables and the quantization error for AoD representation are both reduced. As aforementioned that row-block sparsity usually exists in the spare channel matrix of conventional

---

Algorithm 1 MMV-based Channel Estimator by using SOMP

Input: $\widetilde{\boldsymbol{Y}}_k^H = \boldsymbol{V}^H \boldsymbol{A}_R \widetilde{\boldsymbol{X}}_k + \widetilde{\boldsymbol{U}}_k^H, 1 \le k \le K$, $\boldsymbol{D} = \boldsymbol{V}^H \boldsymbol{A}_R$, and $\epsilon = \frac{MB\delta^2}{PT}$.
Output: $\hat{\boldsymbol{G}}_k^{\text{mmv}}, 1 \le k \le K$.
1: **for** $k = 1$ to K **do**
2:     Initialize the residual $\boldsymbol{R}_k^{(0)} = \widetilde{\boldsymbol{Y}}_k^H$, the corresponding index set $\Omega_k^{(0)} = \emptyset$, and the iteration counter $t = 0$.
3:     **repeat**
4:       Set $t = t + 1$ and estimate the support $i_k^{(t)} = \arg\max_{i \notin \Omega_k^{(t-1)}} \left\| (\boldsymbol{D}^{:,i})^H \boldsymbol{R}_k^{(t-1)} \right\|_2^2$.
5:       Update index set $\Omega_k^{(t)} = \Omega_k^{(t-1)} \cup \{i_k^{(t)}\}$.
6:       Update residual $\boldsymbol{R}_k^{(t)} = (\boldsymbol{I} - \boldsymbol{D}^{:,\Omega_k^{(t)}} (\boldsymbol{D}^{:,\Omega_k^{(t)}})^\dagger) \widetilde{\boldsymbol{Y}}_k^H$.
7:     **until** $\left\| \boldsymbol{R}_k^{(t)} \right\|_2^2 \le \epsilon$.
8:     Denote the estimated index set $\hat{\Omega}_k = \Omega_k^{(t)}$. The channel is estimated by $\hat{\boldsymbol{G}}_k^{\text{mmv}} = \boldsymbol{A}_R \widetilde{\hat{\boldsymbol{X}}}_k$, where $\widetilde{\hat{\boldsymbol{X}}}_k = (\boldsymbol{D}^{:,\hat{\Omega}})^\dagger \widetilde{\boldsymbol{Y}}_k^H$.
9: **end for**

(mmWave) MIMO communications, the MMV-based channel estimator is thus usually adopted in these scenarios.

However, applying the above conventional SMV- and MMV-based estimators into the RIS aided multi-user communication system usually leads to performance loss due to ignoring its specific row-column-block sparsity structure. Hence, it motivates us to redesign a channel estimator for the studied system by taking advantages of both row- and column- block sparsity.

C. Common Channel Sparsity Representation of Multiple-User

It is worth noting that the cascaded channels have the common block sparsity structure brought by the common channel between BS and RIS. Therefore, another main issue of the above conventional estimators is that they recover the sparse channel matrix of each user individually without considering this common sparsity.

Hence, in this subsection, we investigate the common block sparsity, which together with row-column-block sparsity will be applied to design a multi-user joint channel estimator in Section IV.

1) **Common Column-Block Sparsity due to Common Scatters at the BS***:* The cascaded channels exist a common channel, whose spatial paths departing from the BS to the RIS go through the common scatters. Hence, the AoD array

steering vectors of each cascaded channel should be the same as each other. Specifically, let the AoD index set corresponding to the non-zero columns of $\boldsymbol{X}_k$ be $\Omega_{\mathrm{D},k}$, where each element is associated with a typical AoD array steering vector at the BS side. Then, from (1) and (9), we have

$$\Omega_{\mathrm{D},1}=\Omega_{\mathrm{D},2}\cdots=\Omega_{\mathrm{D},K}\overset{\Delta}{=}\Omega_{\mathrm{D}}, \quad (14)$$

where Ω_D is defined as the common AoD support index set.

*2) **Common Row-Block Sparsity due to Scaling Property***: Since the cascaded AoAs for each user are different, there is no straightforward common row-block sparsity. Hence, in this part, we rewrite the VAD presentation of $\boldsymbol{G}_k$ to establish the common row-block sparsity.

First, we have the following challenge to establish the common row-block sparsity.

Challenge 1: Denoting the VAD representation of $\boldsymbol{F}$ by $\boldsymbol{F}=\boldsymbol{A}_R\boldsymbol{X}_F\boldsymbol{A}_T^H$, we have $\boldsymbol{G}_k=\mathrm{diag}(\boldsymbol{h}_k^H)\boldsymbol{A}_R\boldsymbol{X}_F\boldsymbol{A}_T^H$. Then, if $\boldsymbol{h}_k^H$ is known, we can jointly estimate the common sparsity channel matrix $\boldsymbol{X}_F$ existed in all cascaded channels. However, it is challenging to obtain $\boldsymbol{h}_k^H$ since there exists the ambiguity that we cannot separate $\mathrm{diag}(\boldsymbol{h}_k^H)$ and $\boldsymbol{F}$ from $\boldsymbol{G}_k$.

To deal with Challenge 1 and obtain the common row-column-block sparsity representation, we propose the following joint scaling property. Specifically, we observe that the cascaded channels through one arbitrary reflective element of all users are the common channel scaled by different scalars, i.e., $\boldsymbol{G}_k^{l,:}=[\boldsymbol{h}_k^H]_l\boldsymbol{F}^{l,:}$. Hence, we know each element in the l-th row vector of $\boldsymbol{G}_{k_1}$ divided by each element in the l-th row vector of $\boldsymbol{G}_{k_2}$ has the same scaling factor $\frac{[\boldsymbol{h}_{k_1}^H]_l}{[\boldsymbol{h}_{k_2}^H]_l}$, i.e.,

$$\frac{[\boldsymbol{h}_{k_1}^H]_l}{[\boldsymbol{h}_{k_2}^H]_l}=\frac{\boldsymbol{G}_{k_1}^{l,1}}{\boldsymbol{G}_{k_2}^{l,1}}=\frac{\boldsymbol{G}_{k_1}^{l,2}}{\boldsymbol{G}_{k_2}^{l,2}}\cdots=\frac{\boldsymbol{G}_{k_1}^{l,M}}{\boldsymbol{G}_{k_2}^{l,M}}. \quad (15)$$

This is called as the joint scaling property, which implies that all cascaded channels can be represented by one arbitrary cascaded channel with different scalars. For example, we can use the cascaded channel of U_1, i.e., $\boldsymbol{G}_1$, to represent the cascaded channels of the rest users, i.e.,

$$\boldsymbol{G}_k=\boldsymbol{\alpha}_k\boldsymbol{G}, \quad (16)$$

where the lower script "1" of $\boldsymbol{G}_1$ is ignored for notation simplicity, i.e., $\boldsymbol{G}=\boldsymbol{G}_1$, and

$$\boldsymbol{\alpha}_k=\mathrm{diag}\left(\left[\frac{[\boldsymbol{h}_k^H]_1}{[\boldsymbol{h}_1^H]_1},\frac{[\boldsymbol{h}_k^H]_2}{[\boldsymbol{h}_1^H]_2},\cdots,\frac{[\boldsymbol{h}_k^H]_L}{[\boldsymbol{h}_1^H]_L}\right]\right)\in\mathbb{C}^{L\times L}, \quad (17)$$

is referred to as scaling matrix. Observed from (16), we know all the cascaded channels have the common channel part $\boldsymbol{G}$, which means that all cascaded channels have the following common row-block sparse VAD representation, i.e.,

$$\boldsymbol{G}_k=\boldsymbol{\alpha}_k\boldsymbol{A}_R\boldsymbol{X}\boldsymbol{A}_T^H, \quad (18)$$

where the lower script "1" of $\boldsymbol{X}_1$ is ignored for notation simplicity, i.e., $\boldsymbol{X}=\boldsymbol{X}_1$.

From (18), we know that all $\boldsymbol{G}_k$ have the common sparse matrix $\boldsymbol{X}$ and we have established the common row-block sparsity. Note that the main reason, we use $\boldsymbol{\alpha}_k\boldsymbol{G}$ to establish common row-block sparsity instead of $\mathrm{diag}(\boldsymbol{h}_k^H)\boldsymbol{F}$, is that we

can use (15) to separate $\boldsymbol{\alpha}_k$ and $\boldsymbol{G}$ from $\boldsymbol{G}_k$ to solve Challenge 1, which cannot be solved using the term $\mathrm{diag}(\boldsymbol{h}_k^H)\boldsymbol{F}$. This fact is important for the initialization in the following alternative optimization algorithm for solving the multi-user joint channel estimator designed in the next section.

IV. TWO-STEP PROCEDURE BASED MULTI-USER JOINT CHANNEL ESTIMATOR

A. Overview of Two-Step Procedure

To begin with, it is worth noting that if we apply the common row-block sparsity and common column-block sparsity, simultaneously, to formulate a sparse channel matrix recovery problem, we need to quantize the AoA and AoD simultaneously with high resolutions to reduce the quantization errors caused by the VAD representation. This leads to intractable complexity.

To deal with the above issue, we propose the following two-step procedure based multi-user joint channel estimator by applying common row-/column- block sparsity, respectively.

- In the first step, we exploit the common column-block sparsity to jointly estimate the common subspace spanned by AoD array steering vectors. This is because we will show that it is sufficient to represent $\boldsymbol{G}_k$ by using arbitrary N_f basis in the subspace spanned by AoD array steering vectors in Section IV-B. Then, we project our received signal into the common AoD subspace for reducing the number of zero columns of $\boldsymbol{X}_k$ from G_t to N_f. This projection procedure reduces the number of columns of the sparse matrix for complexity reduction, reduce the quantization error due to without quantizing the AoD, and achieves higher SNR due to reducing the influence of noise on the null space of the common AoD subspace.

- In the second step, we exploit the common row-block sparsity to design a MMV-based multi-user joint sparse matrix recovery problem, in which we recover the cascaded AoA array steering vectors with common sparse channel matrix and the scaling matrix $\boldsymbol{\alpha}_k$ jointly. In this step, since we use the received signals of all users to recover the sparse matrix jointly, a better recovery performance can be achieved due to the increased number of measurements.

B. First Step: Subspace Estimation and Signal Projection

In this subsection, we propose the common AoD subspace estimation method and project the signals into this subspace.

To begin with, denote the common AoD subspace by $\mathrm{span}(\boldsymbol{A}_T^{:,\Omega_{\mathrm{D}}})$, which is the linear span of N_f AoD array steering vectors, i.e., $\boldsymbol{A}_T^{:,\Omega_{\mathrm{D}}}\in\mathbb{C}^{M\times N_f}$, where Ω_{D} is the corresponding AoD index set, as aforementioned in (14).

Note that if we apply CS-based methods to estimate the exact N_f AoD array steering vectors $\boldsymbol{A}_T^{:,\Omega_{\mathrm{D}}}$ over a discrete grid of AoDs, and then project the signals into the subspace spanned by the estimated AoD array steering vectors, there exist both quantization error and estimation error, both of which degrade the estimation performance in the second step.

In fact, we do not need to estimate the exact N_f AoD array steering vectors $\boldsymbol{A}_T^{:,\Omega_{\mathrm{D}}}$ to represent common AoD subspace, and we only need to estimate N_f vectors, $\boldsymbol{S}_\| \in \mathbb{C}^{M \times N_f}$, whose linear span includes the common AoD subspace, i.e., $\mathrm{span}(\boldsymbol{A}_T^{:,\Omega_{\mathrm{D}}}) \subseteq \mathrm{span}(\boldsymbol{S}_\|)$. Then, it is clear that $\boldsymbol{S}_\|$ is sufficient to represent $\boldsymbol{A}_T^{:,\Omega_{\mathrm{D}}}$ for the sparse representation of $\boldsymbol{G}_k$, i.e.,

$$\begin{aligned}
\boldsymbol{G}_k &= \boldsymbol{A}_R \boldsymbol{X}_k \boldsymbol{A}_T^H = \boldsymbol{A}_R \boldsymbol{X}_k^{:,\Omega_{\mathrm{D}}}(\boldsymbol{A}_T^{:,\Omega_{\mathrm{D}}})^H \\
&\overset{(a)}{=} \boldsymbol{A}_R \boldsymbol{X}_k^{:,\Omega_{\mathrm{D}}} \boldsymbol{M}^H \boldsymbol{S}_\|^H \overset{(b)}{=} \boldsymbol{A}_R \bar{\boldsymbol{X}}_k \boldsymbol{S}_\|^H,
\end{aligned} \tag{19}$$

where in (a) we use the fact that there exists a matrix $\boldsymbol{M} \in \mathbb{C}^{N_f \times N_f}$ such that $\boldsymbol{A}_T^{:,\Omega_{\mathrm{D}}} = \boldsymbol{S}_\| \boldsymbol{M}$ if $\mathrm{span}(\boldsymbol{A}_T^{:,\Omega_{\mathrm{D}}}) \subseteq \mathrm{span}(\boldsymbol{S}_\|)$, and in (b) we use $\bar{\boldsymbol{X}}_k = \boldsymbol{X}_k^{:,\Omega_{\mathrm{D}}} \boldsymbol{M}^H \in \mathbb{C}^{G_r \times N_f}$, which is a row-block sparsity matrix consisting of N_f columns. Specifically, $\boldsymbol{S}_\|$ can be estimated by eigenvalue decomposition of the signal's covariance matrix without requiring a discrete grid of AoDs, and we do not need to estimate the linear transform matrix $\boldsymbol{M}$ for $\boldsymbol{A}_T^{:,\Omega_{\mathrm{D}}} = \boldsymbol{S}_\| \boldsymbol{M}$ because we only need to estimate the combined row-block sparsity matrix $\bar{\boldsymbol{X}}_k$ in the second step.

Therefore, we can reduce the number of zero columns by projecting the signal into the estimated subspace $\mathrm{span}(\boldsymbol{S}_\|)$ without estimating the exact AoD array steering vectors $\boldsymbol{A}_T^{:,\Omega_{\mathrm{D}}}$.

1) Subspace Estimation: In this part, we propose the following lemmas to estimate the common AoD subspace.

Lemma 4.1: Denote $\boldsymbol{Y} = [\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots, \boldsymbol{Y}_B] \in \mathbb{C}^{M \times BT}$. Then, $\hat{\boldsymbol{C}} = \frac{1}{BT} \boldsymbol{Y} \boldsymbol{Y}^H$ is a sufficient statistics for estimating the AoD subspace $\mathrm{span}(\boldsymbol{A}_T^{:,\Omega_{\mathrm{D}}})$ if $K \to \infty$.

Proof: Please refer to Appendix A-A. ∎

Lemma 4.2: For $K \to \infty$, by maximizing the likelihood function of $\hat{\boldsymbol{C}}$ for a given N_f, the optimal solution of subspace estimation is the linear span of the eigenvalue vectors corresponding to the N_f largest eigenvalue of $\hat{\boldsymbol{C}}$, i.e.,

$$\boldsymbol{S}_\| = \left[\boldsymbol{S}^{:,1}, \boldsymbol{S}^{:,2}, \cdots, \boldsymbol{S}^{:,N_f} \right], \tag{20}$$

where $\boldsymbol{S}\boldsymbol{\Theta}\boldsymbol{S}^{\mathrm{H}}$ is the eigenvalue decomposition of $\hat{\boldsymbol{C}}$, and $\boldsymbol{\Theta} = \mathrm{diag}([\theta_1, \theta_2, \cdots, \theta_M]) \in \mathbb{C}^{M \times M}$ is the eigenvalue matrix where the eigenvalue θ_m is ordered in a decreasing order of magnitude.

Proof: Please refer to Appendix A-B. ∎

Note that the decision on the number of AoD array steering vectors, i.e., N_f, is a model selection problem, which is addressed by information theoretic criteria [31]. In addition, there are several well-known decision rules, such as likelihood ratio [32], *Akaike's information criterion* (AIC), and *minimum description length* (MDL) [33]. In this paper, we just adopt MDL scheme [33] to estimate N_f, i.e.,

$$\hat{N}_f = \arg\min_n \left\{ -\log \left(\frac{\prod\limits_{i=n+1}^{M} \theta_i^{\frac{1}{M-n}}}{\frac{\sum\limits_{i=n+1}^{M} \theta_i}{M-n}} \right)^{(M-n)BT} + 2n(2M-n) \right\}. \tag{21}$$

Algorithm 2 Two-Step Procedure (Subspace) based Multi-User Joint Channel Estimation (S-MJCE)

Input: $\tilde{\boldsymbol{Y}}_k^H, 1 \le k \le K$.
Output: $\hat{\boldsymbol{G}}_k^{\mathrm{mjce}}, 1 \le k \le K$.
1: **First Step: Subspace Estimation and Signal Projection**
   • **Subspace Estimation**
   Compute covariance matrix $\hat{\boldsymbol{C}} = \frac{1}{BT} \boldsymbol{Y} \boldsymbol{Y}^H$.
   Apply the eigenvalue decomposition as $\hat{\boldsymbol{C}} = \boldsymbol{S}\boldsymbol{\Theta}\boldsymbol{S}^H$.
   According to the MDL, the number of AoD vectors is estimated by (21) and denoted by $\hat{N}_f$.
   According to Lemma 4.2, the AoD subspace is estimated by the first $\hat{N}_f$ columns of $\boldsymbol{S}$ in (20).
   • **Signal Projection**
   Project signal as $\tilde{\boldsymbol{Y}}_k^H(\boldsymbol{S}_\|^H)^\dagger = \boldsymbol{V}^H \boldsymbol{A}_R \bar{\boldsymbol{X}}_k + \tilde{\boldsymbol{U}}_k(\boldsymbol{S}_\|^H)^\dagger$ in (22).
2: **Second Step: multi-user joint channel estimation (MJCE)**
   Solving problem (24) by Algorithm 3, the cascaded channels can be estimated by $\hat{\boldsymbol{G}}_k^{\mathrm{mjce}} = \hat{\boldsymbol{\alpha}}_k \boldsymbol{A}_R \hat{\bar{\boldsymbol{X}}} \hat{\boldsymbol{S}}_\|^H$.

---

2) Signal Projection: By applying (19), we can project the received signals into common AoD subspace $\mathrm{span}(\boldsymbol{S}_\|)$, i.e.,

$$\begin{aligned}
\bar{\boldsymbol{Y}}_k^H \triangleq \tilde{\boldsymbol{Y}}_k^H (\boldsymbol{S}_\|^H)^\dagger &= \boldsymbol{V}^H \boldsymbol{A}_R \bar{\boldsymbol{X}}_k \boldsymbol{S}_\|^H(\boldsymbol{S}_\|^H)^\dagger + \tilde{\boldsymbol{U}}_k(\boldsymbol{S}_\|^H)^\dagger \\
&\overset{(a)}{=} \boldsymbol{V}^H \boldsymbol{A}_R \bar{\boldsymbol{X}}_k + \tilde{\boldsymbol{U}}_k(\boldsymbol{S}_\|^H)^\dagger,
\end{aligned} \tag{22}$$

where (a) is due to $\boldsymbol{S}_\|^H(\boldsymbol{S}_\|^H)^\dagger = \boldsymbol{I}_{N_f}$. Compared with (13) and (22), the noise on the null space of $\mathrm{span}(\boldsymbol{S}_\|)$ has been reduced, and the noise power is reduced from $\mathbb{E}(\|\tilde{\boldsymbol{U}}_k\|_F^2) = MB\delta^2/PT$ to $\mathbb{E}(\|\tilde{\boldsymbol{U}}_k(\boldsymbol{S}_\|^H)^\dagger\|_F^2) = N_f B\delta^2/PT$.

C. Second Step: Multi-User Joint Sparse Matrix Recovery

Specifically, applying the common row-block sparsity (16), we can rewrite (22) as

$$\begin{aligned}
\bar{\boldsymbol{Y}}_k^H &= \boldsymbol{V}^H \boldsymbol{\alpha}_k \boldsymbol{G}(\boldsymbol{S}_\|^H)^\dagger + \tilde{\boldsymbol{U}}_k(\boldsymbol{S}_\|^H)^\dagger \\
&= \boldsymbol{V}^H \boldsymbol{\alpha}_k \boldsymbol{A}_R \bar{\boldsymbol{X}} + \tilde{\boldsymbol{U}}_k(\boldsymbol{S}_\|^H)^\dagger.
\end{aligned} \tag{23}$$

In order to estimate $\boldsymbol{\alpha}_k$ and the real cascaded AoAs associated with the sparse channel matrix $\bar{\boldsymbol{X}}$ from (23), we can formulate the following MMV-based multi-user joint sparse matrix recovery problem, i.e.,

$$\min_{\bar{\boldsymbol{X}}, \boldsymbol{\alpha}_k} \left\| \mathrm{diag}\left(\bar{\boldsymbol{X}}\bar{\boldsymbol{X}}^H \right) \right\|_0 = \sum\nolimits_{i=1}^{G_r} \left\| \bar{\mathbf{x}}_i^H \bar{\mathbf{x}}_i \right\|_0 \tag{24}$$

$$\text{s.t.} \left\| \tilde{\boldsymbol{Y}}_k^H - \boldsymbol{V}^H \boldsymbol{\alpha}_k \boldsymbol{A}_R \bar{\boldsymbol{X}} \right\|_2^2 \le \bar{\epsilon}, 1 \le k \le K,$$

where $\bar{\boldsymbol{x}}_i^H$ is the i-th row vector of $\bar{\boldsymbol{X}}$, and $\bar{\epsilon} \ge \frac{N_f B\delta^2}{PT}$ is the tolerance upper bound related to the combined noise.

Denote the solutions of problem (24) as $\hat{\bar{\boldsymbol{X}}}$ and $\hat{\boldsymbol{\alpha}}_k$, respectively. Then, the cascaded channels can be estimated by

$$\hat{\boldsymbol{G}}_k^{\mathrm{mjce}} = \hat{\boldsymbol{\alpha}}_k \boldsymbol{A}_R \hat{\bar{\boldsymbol{X}}} \boldsymbol{S}_\|^H. \tag{25}$$

Finally, the detail steps of the above two-step based multi-user joint channel estimation are summarized in Algorithm 2.

However, Algorithm 2 requires to solve the non-convex problem (24), where the optimization variables $\bar{\boldsymbol{X}}$ and $\boldsymbol{\alpha}_k$ are further coupled. Hence, the conventional OMP/SOMP method

can not be applied to solve the 1-norm relaxation of problem (24). In the next section, we develop an algorithm based on the principles of alternative optimization and iterative reweighted algorithm to solve this problem efficiently.

V. SOLUTION TO MULTI-USER JOINT SPARSE MATRIX RECOVERY PROBLEM

A. Algorithm Development

To deal with the coupled optimization variables $\bar{X}$ and α_k, we develop an alternative optimization method to solve the multi-user joint sparse matrix recovery problem (24). From [34], [35], theoretical results and experimental results have shown that Log-sum penalty function has superiority over the 1-norm penalty function for sparse signal recovery. Hence, in this section, we relax the 0-norm function as the log-sum function for the alternative sparsity-promoting function design, i.e.,

$$\min_{\bar{X},\alpha_k} \quad \mathcal{Q}(\bar{X}) = \sum_{i=1}^{G_r} \log\left(\bar{x}_i^H \bar{x}_i + \varsigma\right) \tag{26}$$
$$\text{s.t.} \quad \left\|\tilde{Y}_k^H - V^H \alpha_k A_R \bar{X}\right\|_2^2 \leq \bar{\epsilon}, 1 \leq k \leq K,$$

where $\varsigma > 0$ is a small positive parameter to guarantee that the function is meaningful. The choice of ς can be related to [36] for details.

Then, by introducing non-negative penalty factor λ_k, we reformulate problem (26) as the following unconstrained optimization problem,

$$\min_{\bar{X},\alpha_k} \quad \mathcal{L}(\bar{X},\alpha_k) \triangleq \mathcal{Q}(\bar{X}) + \sum_{k=1}^{K} \lambda_k \left\|\tilde{Y}_k^H - V^H \alpha_k A_R \bar{X}\right\|_2^2. \tag{27}$$

Note that λ_k is the penalty factor balancing the tradeoff between data fitting and the sparsity of the solutions. The value of λ_k should be related to the *signal-to-noise ratio* (SNR) and can be set as $\lambda_k = \frac{PTd}{\delta^2 \log G_r}$ [36]–[38], where d is a constant scaling factor. Since the received signal for each user has the same SNR, we set $\lambda = \lambda_1 = \lambda_2 = \cdots = \lambda_K$ in the following contents.

Then, to further deal with the coupled optimization variables, we apply alternative optimization to decouple problem (27) into the following two unconstrained subproblems as

$$\min_{\alpha_k} \mathcal{L}(\bar{X},\alpha_k), \tag{28}$$

which is an optimization problem of α_k for given $\bar{X}$, and

$$\min_{\bar{X}} \mathcal{L}(\bar{X},\alpha_k), \tag{29}$$

which is an optimization problem of $\bar{X}$ for given α_k. Then, we alternatively solve (28) and (29) until the objective function converges.

In the following subsequent parts, we first develop algorithms to obtain the solutions of problem (28) and problem (29), respectively. Then, we provide the convergence analysis and initialization method of the proposed algorithm.

B. Optimization of α_k

In this subsection, we provide the optimal solution of (28) with fixed $\bar{X}$. To begin with, problem (28) can be equivalently reformulated as the following K individual optimization problems, i.e.,

$$\min_{\alpha_k} \left\|\tilde{Y}_k^H - V^H \alpha_k A_R \bar{X}\right\|_2^2. \tag{30}$$

Then, let $\bar{z}_k = \text{vec}(\bar{Y}_k^H) \in \mathbb{C}^{BM \times 1}$, and $H = (A_R \bar{X})^T \otimes V^H \in \mathbb{C}^{BN_f \times L^2}$. Note that there are only L non-zero elements in the vector $\text{vec}(\alpha_k)$ with the following index set Ω_α, i.e.,

$$\Omega_\alpha = \{i + (i-1)L \,|\, i \in \{1, 2, \cdots, L\}\}, \tag{31}$$

where $|\Omega_\alpha| = L$. Hence, problem (30) can be rewritten as, i.e.,

$$\min_{\alpha_k} \left\|[\bar{z}_k]_{\Omega_\alpha} - H^{:,\Omega_\alpha}[\text{vec}(\alpha_k)]_{\Omega_\alpha}\right\|_2^2. \tag{32}$$

Since problem (32) is a convex optimization problem, the optimal solution can be obtained by letting the first derivative of the objection equal to zero, which is given by

$$\alpha_k = \text{diag}\left((H^{:,\Omega_\alpha})^\dagger [\bar{z}_k]_{\Omega_\alpha}\right). \tag{33}$$

$(H^{:,\Omega_\alpha})^\dagger = ((H^{:,\Omega_\alpha})^H H^{:,\Omega_\alpha})^{-1}(H^{:,\Omega_\alpha})^H$. Note that the rank of matrix $H^{:,\Omega_\alpha}$ should be large than L in (33), which means that $\text{rank}(\bar{X}) \geq \frac{L}{B}$.

C. Iterative Reweighted Algorithm for Optimizing $\bar{X}$

In this section, we find the solution of (29) for fixed α_k. In fact, $\mathcal{L}(\bar{X},\alpha_k)$ is a non-convex function, which is still hard to solve and may require intractable complexity to find the optimal solutions. To find an efficient solution of non-convex problem (29), we apply the iterative reweighted technique to solve it in an iterative manner. The basic principle of this method is to iteratively approximate the non-convex objective function as a convex function with reweighted coefficients and solve the approximated convex problem.

Specifically, in the t-th iteration of iterative reweighted algorithm, the upper bound approximated function of $\mathcal{Q}(\bar{X})$ at $\bar{X}(t)$ is given by

$$\mathcal{Q}(\bar{X}) = \sum_{i=1}^{G_r} \log\left(\bar{x}_i^H \bar{x}_i + \varsigma\right)$$
$$\leq \sum_{i=1}^{G_r} \left(\frac{\bar{x}_i^H \bar{x}_i + \varsigma}{\bar{x}_{i,t}^H \bar{x}_{i,t} + \varsigma} + \log\left(\bar{x}_{i,t}^H \bar{x}_{i,t} + \varsigma\right) - 1\right)$$
$$\triangleq \mathcal{Q}^{\text{ub}}(\bar{X}, \bar{X}(t)), \tag{34}$$

where $\bar{x}_{i,t}^H$ is the i-th row vector of constant matrix $\bar{X}(t)$, and $\bar{x}_{i,t}^H \bar{x}_{i,t} + \varsigma$ can be regarded as the reweighted coefficient in the t-th iteration on the objective function. This inequality can be proved by calculating its first derivation, which is omitted here for brevity. Then, the upper bound approximated function

of $\mathcal{L}(\bar{\boldsymbol{X}})$ in the t-th iteration of iterative reweighted algorithm is given by

$$
\begin{aligned}
\mathcal{L}(\bar{\boldsymbol{X}}, \boldsymbol{\alpha}_k) \\
\leq \mathcal{Q}^{\mathrm{ub}}(\bar{\boldsymbol{X}}, \bar{\boldsymbol{X}}(t)) + \lambda \sum_{k=1}^{K} \left\| \tilde{\boldsymbol{Y}}_k^H - \boldsymbol{V}^H \boldsymbol{\alpha}_k \boldsymbol{A}_R \bar{\boldsymbol{X}} \right\|_2^2 \\
\overset{\triangle}{=} \mathcal{L}^{\mathrm{ub}}(\bar{\boldsymbol{X}}, \boldsymbol{\alpha}_k; \bar{\boldsymbol{X}}(t)).
\end{aligned} \tag{35}
$$

Note that the equality holds in (35) if $\bar{\boldsymbol{X}} = \bar{\boldsymbol{X}}(t)$.

Next, minimizing $\mathcal{L}^{\mathrm{ub}}(\bar{\boldsymbol{X}}, \boldsymbol{\alpha}_k; \bar{\boldsymbol{X}}(t))$ is equivalent to solving the following problem, i.e.,

$$
\min_{\bar{\boldsymbol{X}}} \mathrm{Tr}\left(\bar{\boldsymbol{X}}^H \boldsymbol{\Lambda} \bar{\boldsymbol{X}} \right) + \lambda \sum_{k=1}^{K} \left\| \tilde{\boldsymbol{Y}}_k^H - \boldsymbol{V}^H \boldsymbol{\alpha}_k \boldsymbol{A}_R \bar{\boldsymbol{X}} \right\|_2^2, \tag{36}
$$

where $\boldsymbol{\Lambda} \in \mathbb{C}^{G_r \times G_r}$ is a diagonal matrix, i.e.,

$$
\boldsymbol{\Lambda} = \mathrm{diag}\left(\frac{1}{\bar{\boldsymbol{x}}_{1,t}^H \bar{\boldsymbol{x}}_{1,t} + \varsigma}, \cdots, \frac{1}{\bar{\boldsymbol{x}}_{G_r,t}^H \bar{\boldsymbol{x}}_{G_r,t} + \varsigma} \right). \tag{37}
$$

It is straightforward to know that problem (36) is a convex optimization problem, and the optimal solution can be obtained by letting the first derivative of the objection equal to zero. Then, the optimal solution can be given by

$$
\bar{\boldsymbol{X}} = \left(\frac{\boldsymbol{\Lambda}}{\lambda} + \sum_{k=1}^{K} (\boldsymbol{V}\boldsymbol{\alpha}_k \boldsymbol{A}_R)^H \boldsymbol{V}\boldsymbol{\alpha}_k \boldsymbol{A}_R \right)^{-1} \sum_{k=1}^{K} (\boldsymbol{V}\boldsymbol{\alpha}_k \boldsymbol{A}_R)^H \tilde{\boldsymbol{Y}}_k^H. \tag{38}
$$

D. Convergence and Initialization Analysis

1) Convergence Analysis: In this part, we analyze the convergence of the proposed iterative reweighted based alternative optimization algorithm, which is summarized in Algorithm 3 with a double loop structure. Specifically, the inner loop is from step 4 to step 6 for optimizing $\bar{\boldsymbol{X}}$ and the outer loop is from step 1 to step 9 for alternatively optimizing $\bar{\boldsymbol{X}}$ and $\boldsymbol{\alpha}_k$. In the following, we show $\mathcal{L}(\bar{\boldsymbol{X}}^{(r+1)}, \boldsymbol{\alpha}_k^{(r+1)}) \leq \mathcal{L}(\bar{\boldsymbol{X}}^{(r)}, \boldsymbol{\alpha}_k^{(r)})$, where r is the outer loop iteration index, which guarantees the convergence of Algorithm 3 [39] .

Specifically, the proof is given as follows.

$$
\begin{aligned}
\mathcal{L}(\bar{\boldsymbol{X}}^{(r)}, \boldsymbol{\alpha}_k^{(r)}) &\overset{(a)}{=} \mathcal{L}^{\mathrm{ub}}(\bar{\boldsymbol{X}}^{(r)}, \boldsymbol{\alpha}_k^{(r)}; \bar{\boldsymbol{X}}(0)) \\
&\geq \min_{\bar{\boldsymbol{X}}} \mathcal{L}^{\mathrm{ub}}(\bar{\boldsymbol{X}}, \boldsymbol{\alpha}_k^{(r)}; \bar{\boldsymbol{X}}(0)) \overset{(b)}{=} \mathcal{L}^{\mathrm{ub}}(\bar{\boldsymbol{X}}(1), \boldsymbol{\alpha}_k^{(r)}; \bar{\boldsymbol{X}}(0)) \cdots \\
&\geq \min_{\bar{\boldsymbol{X}}} \mathcal{L}^{\mathrm{ub}}(\bar{\boldsymbol{X}}, \boldsymbol{\alpha}_k^{(r)}; \bar{\boldsymbol{X}}(t-1)) \overset{(c)}{=} \mathcal{L}^{\mathrm{ub}}(\bar{\boldsymbol{X}}(t), \boldsymbol{\alpha}_k^{(r)}; \bar{\boldsymbol{X}}(t-1)) \\
&\overset{(d)}{\geq} \mathcal{L}(\bar{\boldsymbol{X}}(t), \boldsymbol{\alpha}_k^{(r)}) \overset{(e)}{=} \mathcal{L}(\bar{\boldsymbol{X}}^{(r+1)}, \boldsymbol{\alpha}_k^{(r)}) \\
&\geq \min_{\boldsymbol{\alpha}_k} \mathcal{L}(\bar{\boldsymbol{X}}^{(r+1)}, \boldsymbol{\alpha}_k^{(r)}) \overset{(f)}{=} \mathcal{L}(\bar{\boldsymbol{X}}^{(r+1)}, \boldsymbol{\alpha}_k^{(r+1)}),
\end{aligned} \tag{39}
$$

where in (a) we initialize $\bar{\boldsymbol{X}}^{(r)} = \bar{\boldsymbol{X}}(0)$ and the equality holds in (35) if $\bar{\boldsymbol{X}}^{(r)} = \bar{\boldsymbol{X}}(0)$, in (b) and (c) we use the fact that $\bar{\boldsymbol{X}}(t)$ is the optimal solution of (36), in (d) we use (35), in (e) we update $\bar{\boldsymbol{X}}^{(r+1)}$ as $\bar{\boldsymbol{X}}(t)$, and in (f) we use the fact that $\boldsymbol{\alpha}_k$ is the optimal solution of (30).

Therefore, we have $\mathcal{L}(\bar{\boldsymbol{X}}^{(r+1)}, \boldsymbol{\alpha}_k^{(r+1)}) \leq \mathcal{L}(\bar{\boldsymbol{X}}^{(r)}, \boldsymbol{\alpha}_k^{(r)})$, which guarantees the convergence.

Algorithm 3 Iterative Reweighted based Alternative Optimization

1: Initialize $\bar{\boldsymbol{X}}^{(0)}$ as $\boldsymbol{1}_{G_r \times \hat{N}_f}$, $\boldsymbol{\alpha}_k^{(0)}$ as (41), and the outer loop iteration counter $r = 0$.
2: **repeat**
3:     Set $\boldsymbol{\alpha}_k = \boldsymbol{\alpha}_k^{(r)}$ and $\bar{\boldsymbol{X}}(0) = \bar{\boldsymbol{X}}^{(r)}$, initialize the inner loop iteration counter $t = 0$.
4:     **repeat**
5:        Given $\bar{\boldsymbol{X}}(t)$, calculate $\bar{\boldsymbol{X}}(t+1)$ based on (38), and then set $t = t + 1$.
6:     **until** $\mathcal{L}^{\mathrm{ub}}(\bar{\boldsymbol{X}}, \boldsymbol{\alpha}_k; \bar{\boldsymbol{X}}(t))$ converges.
7:     Set $r = r + 1$ and update $\bar{\boldsymbol{X}}^{(r)} = \bar{\boldsymbol{X}}(t)$, then given $\bar{\boldsymbol{X}} = \bar{\boldsymbol{X}}^{(r)}$, calculate $\boldsymbol{\alpha}_k^{(r)}$ based on (33).
8: **until** $\mathcal{L}(\bar{\boldsymbol{X}}^{(r)}, \boldsymbol{\alpha}_k^{(r)})$ converges.

---

2) Initialization Analysis: It is worth noting that the initialization in Algorithm 3 is important for the successful recovery of the sparse channel matrix, especially for the initialization of $\boldsymbol{\alpha}_k$. This is because only when the initialized $\boldsymbol{\alpha}_k$ is as close as possible to the real value, all the cascaded channels have the similar sparse channel matrix, and thus it can be jointly recovered due to (16). Otherwise, there is no solution of $\bar{\boldsymbol{X}}$ to make the equality hold in (16).

In fact, we can initialize $\boldsymbol{\alpha}_k$ by first estimating the cascaded channels $\boldsymbol{G}_k$ individually, i.e., using the SMV method, using the MMV method, or solving the single user version of (24). Then, we use the estimated $\hat{\boldsymbol{G}}_k$ to initialize $\boldsymbol{\alpha}_k$ according to the scaling property (15), i.e.,

$$
\boldsymbol{\alpha}_k^{l,l} = \frac{[\boldsymbol{h}_k^H]_l}{[\boldsymbol{h}_1^H]_l} \approx \frac{\hat{\boldsymbol{G}}_k^{l,1}}{\hat{\boldsymbol{G}}_1^{l,1}} \approx \frac{\hat{\boldsymbol{G}}_k^{l,2}}{\hat{\boldsymbol{G}}_1^{l,2}} \cdots \approx \frac{\hat{\boldsymbol{G}}_k^{l,M}}{\hat{\boldsymbol{G}}_1^{l,M}}, \forall k, \forall l. \tag{40}
$$

Thus, we can initialize $\boldsymbol{\alpha}_k^{(0)}$ as

$$
\left(\boldsymbol{\alpha}_k^{(0)} \right)^{i,l} = \begin{cases} \frac{1}{M} \sum_{m=1}^{M} \frac{\hat{\boldsymbol{G}}_k^{l,m}}{\hat{\boldsymbol{G}}_1^{l,m}}, & \text{if } 1 \leq i = l \leq L, \\ 0, & \text{Others.} \end{cases} \tag{41}
$$

VI. TRAINING REFLECTION COEFFICIENT OPTIMIZATION

In this section, we optimize the training reflection coefficient sequence at the RIS to improve the estimation performance. Motivated by the fact that a better recovery performance can be achieved if the mutual coherence of the equivalent dictionary is smaller [40]. Specifically, denoting the equivalent dictionary by $\boldsymbol{D} = \boldsymbol{V}^H \boldsymbol{A}_R$ from (13), the corresponding mutual coherence is denoted by $\mu(\boldsymbol{D})$, which can be written as

$$
\mu(\boldsymbol{D}) = \max_{i \neq j, 1 \leq i,j \leq G_r} \left\{ \frac{(\boldsymbol{D}^{:,i})^H \boldsymbol{D}^{:,j}}{\|\boldsymbol{D}^{:,i}\|_2 \|\boldsymbol{d}^{:,j}\|_2} \right\}. \tag{42}
$$

This fact implies that the columns of $\boldsymbol{D}$ should be as orthogonal as possible. Equivalently, it requires that we need to design $\boldsymbol{D}$ to make $\boldsymbol{D}^H \boldsymbol{D}$ as similar as possible to identity matrix, i.e.,

$$
\boldsymbol{D}^H \boldsymbol{D} = \boldsymbol{A}_R^H \boldsymbol{V} \boldsymbol{V}^H \boldsymbol{A}_R \approx B \boldsymbol{I}_{G_t G_r}, \tag{43}
$$

where B in the right term is applied for normalization.

The solution for problem (43) has been investigated in [41]. However, since RIS just induces a phase shift on the

incident signal without changing its amplitude, the reflection coefficients should satisfy the following constraint [42], i.e.,

$$\left|\boldsymbol{V}^{b,l}\right| = 1, \forall b, \forall l. \tag{44}$$

Thus, the method in [41] cannot be applied to find the solution of $\boldsymbol{V}$ in (43) directly. In the following, we modify the method in [41] to solve problem (43) subject to constraint (44). Firstly, we can transfer (43) into the following equation, i.e.,

$$\boldsymbol{A}_R \boldsymbol{A}_R^H \boldsymbol{V} \boldsymbol{V}^H \boldsymbol{A}_R \boldsymbol{A}_R^H \approx B \boldsymbol{A}_R \boldsymbol{A}_R^H. \tag{45}$$

Let $\boldsymbol{U}_R \boldsymbol{\Xi} \boldsymbol{U}_R^H$ be the eigenvalue decomposition of $\boldsymbol{A}_R \boldsymbol{A}_R^H$, and $\boldsymbol{\Xi} = \text{diag}(\gamma_1, \gamma_2, \cdots, \gamma_L) \in \mathbb{C}^{L \times L}$ be the eigenvalue matrix where the eigenvalue γ_l is ordered in a decreasing order of magnitude. Then, we have the following optimization problem,

$$\min_{\boldsymbol{V}} \left\| B\boldsymbol{\Xi} - \boldsymbol{\Xi} \boldsymbol{U}_R^H \boldsymbol{V} \boldsymbol{V}^H \boldsymbol{U}_R \boldsymbol{\Xi} \right\|_2^2 \text{ s.t. } \left|\boldsymbol{V}^{b,l}\right| = 1, \forall b, \forall l. \tag{46}$$

Then, define $\boldsymbol{Q} = [\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots, \boldsymbol{q}_B] = \boldsymbol{\Xi} \boldsymbol{U}_R^H \boldsymbol{V}$ and $\boldsymbol{E}_b = B\boldsymbol{\Xi} - \sum_{i \neq b}^{B} \boldsymbol{q}_i \boldsymbol{q}_i^H$. Thus, (46) can be rewritten as

$$\min_{\boldsymbol{V}} \left\| \boldsymbol{E}_b - \boldsymbol{q}_b \boldsymbol{q}_b^H \right\|_2^2 \text{ s.t. } \left|\boldsymbol{V}^{b,l}\right| = 1, \forall b, \forall l. \tag{47}$$

Let $\boldsymbol{U}_{E,b} \boldsymbol{\Psi}_b \boldsymbol{U}_{E,b}^H$ be the eigenvalue decomposition of $\boldsymbol{E}_b$, and $\boldsymbol{\Psi}_b = \text{diag}(\xi_{b,1}, \xi_{b,2}, \cdots, \xi_{b,L}) \in \mathbb{C}^{L \times L}$ be the eigenvalue matrix where the eigenvalue $\xi_{b,l}$ is ordered in a decreasing order of magnitude. Then, we can make $\boldsymbol{q}_b = \sqrt{\xi_{b,1}} \boldsymbol{u}_{E,b}$, where $\boldsymbol{u}_{E,b}$ is the first column vector of $\boldsymbol{U}_{E,b}$ corresponding to the largest eigenvalue. Thus, the largest error in (46) is eliminated.

Next, denote $\boldsymbol{q}_b = [\gamma_1 \tilde{q}_{b,1}, \gamma_2 \tilde{q}_{b,2}, \cdots, \gamma_L \tilde{q}_{b,L}]^T \in \mathbb{C}^{L \times 1}$, where $\tilde{q}_{b,l}$ is the l-th component of $\tilde{\boldsymbol{q}}_b$. Then, we can recover $\tilde{\boldsymbol{q}}_b$. Note that since matrix $\boldsymbol{E}_b$ is not full-rank in general, we only need to update the components in $\tilde{\boldsymbol{q}}_b$ corresponding to the positive eigenvalue.

By considering (44), we need to further solve the following projection problem:

$$\min_{\boldsymbol{v}_b} \left\| \boldsymbol{U}_R^H \boldsymbol{v}_b - \tilde{\boldsymbol{q}}_b \right\|_2^2 \text{ s.t. } |\boldsymbol{V}(b,l)| = 1, \forall l. \tag{48}$$

From [10], the optimal solution of (48) is given by

$$\boldsymbol{v}_b = e^{j\angle(\boldsymbol{U}_R \tilde{\boldsymbol{q}}_b)}. \tag{49}$$

Then, we substitute $\boldsymbol{v}_b$ into (46) and repeat the above steps B times to compute $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_B$, successively. Finally, we can obtain the optimized training reflection coefficient $\boldsymbol{V}$.

VII. SIMULATION RESULTS

In this section, we show the simulation results to validate the effectiveness of the proposed algorithm. In these simulations, we assume α_p and $\beta_{k,q}$ follow complex Gaussian distribution with unit power. ϕ_p^{AoA}, ϕ_p^{AoD}, and $\varphi_{k,q}$ are continuous and uniformly distributed over $[0, 2\pi)$. Parameters ς, d, and G_r are set as 10^{-9}, 0.1 and 512, respectively. In addition, the noise power δ^2 is normalized to one, and the transmit power P (dB)

is described as a relative value of the noise power. Specifically, we define the normalized mean square error (NMSE) as the performance metric [26], which is given by

$$\text{NMSE} = \mathbb{E}\left[\left\|\hat{\boldsymbol{G}}_k - \boldsymbol{G}_k\right\|_2^2 \bigg/ \left\|\boldsymbol{G}_k\right\|_2^2\right]. \tag{50}$$

In the following simulations, we compare the NMSE performances of the following channel estimation schemes. In addition, each result is obtained over 500 Monte Carlo trials.

- *LS*: The channels are estimated by using the estimation protocol in Fig. 2 and the LS estimator in (5) with the optimal training sequences studied in [22].
- *Binary Reflection* [20]: The channels are estimated one by one through turning on only one reflective element and keeping the rest reflective elements closed.
- *MMV*: The channels are estimated by formulating a MMV problem with the solution of SOMP algorithm as shown in Algorithm 1.
- *S-MMV*: The channels are estimated by projecting the received signals into the common subspace firstly and using the MMV estimator.
- *S-SMV*: The channels are estimated by projecting the received signals into the common subspace firstly and formulating a SMV problem with the solution of OMP algorithm [26].
- *S-MJCE*: The channels are estimated by two-step procedure (subspace) based *multi-user joint channel estimation* (MJCE), which is given in Algorithm 2.
- *S-Genie-aided LS*: The cascaded channels are estimated by assuming that the BS knows the exact angles of the cascaded AoAs at the RIS/user side and the AoDs at the BS side, and projecting the received signals into the known common AoD subspace with the solution of LS estimator, which is the performance upper bound that can not be achieved.

Fig. 4 shows the impacts of training overhead B on the NMSE. Firstly, we can observe that the estimation performances of all estimation schemes increase as the training overhead B. This is because a better successful recovery and estimation accuracy can be achieved with a large number of measurements. Next, the performance of the binary reflection method is much worse than that of other baselines. This is because in each transmit symbol duration, the power of training reflection coefficient at the RIS of this method is 1 while that of other schemes is L. Next, the increased performance gap between the MMV and the S-MMV implies that subspace projection for estimation performance is significant with a large B for the studied system. Moreover, the decreased performance gap between the S-SMV and the S-MMV shows that the performance loss by ignoring block sparsity is significant when B is small. Finally, the proposed method outperforms other baseline schemes significantly, and achieves similar estimation performance of the Genie-aided LS method. This validates the effectiveness of the proposed estimator.

Fig. 5 shows the impacts of the number of scatters between the BS and the RIS on the NMSE. Firstly, we can observe that the performances of all estimation schemes decrease as
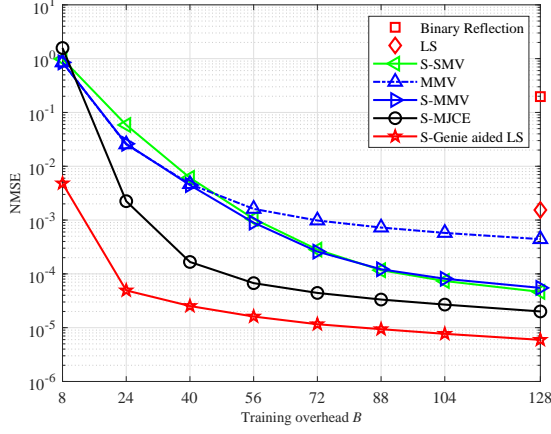
Fig. 4: Effect of training overhead B on the NMSE: $P = 10\text{dB}$, $K = T = 4$, $M = L = 128$, $N_f = 8$, and $N_{h_k} = 1$.
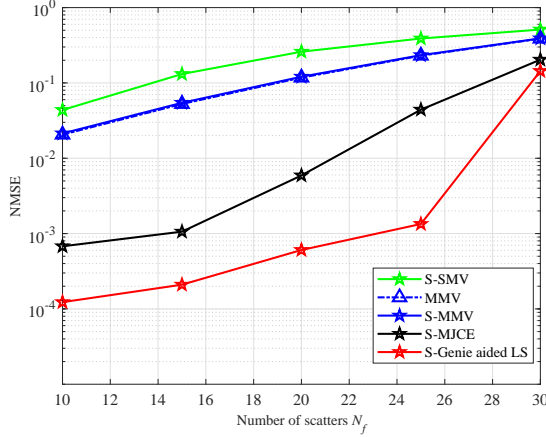

Fig. 5: Effect of the number of scatters between the BS and the RIS on the NMSE: $P = 10\text{dB}$, $B = 30$, $K = T = 4$, $M = L = 128$, $N_f = 8$, and $N_{h_k} = 1$.
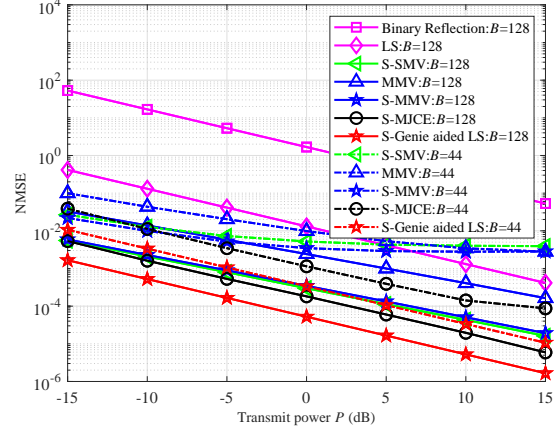

Fig. 6: Effect of the transmit power on the NMSE for different system parameters: $K = T = 4$, $M = L = 128$, $N_f = 8$, and $N_{h_k} = 1$.
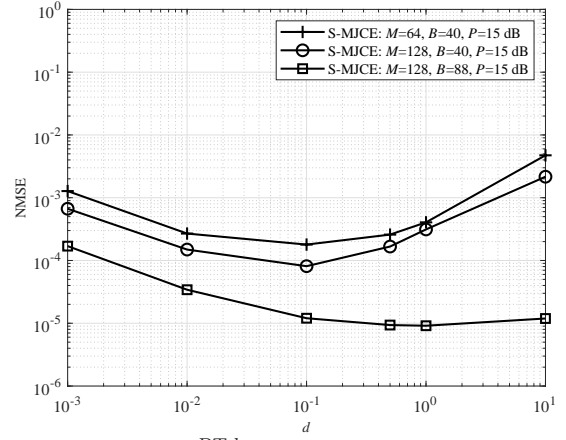

Fig. 7: Effect of $\lambda = \frac{PTd}{\log G_r}$ on the NMSE with different system parameters: $K = T = 4$, $L = 128$, $N_f = 8$, and $N_{h_k} = 1$.

the number of scatters grows. This is because the number of unknown parameters required to be estimated increases with the number of scatters. Also, the performance gaps among these methods decrease as the number of scatters grows. This is because the estimation performance is limited by the number of measurements and SNR. Besides, the proposed method outperforms other baseline schemes especially when the number of scatters is small, which also validates the effectiveness of the proposed estimator.

Fig. 6 shows the impacts of transmit power on the NMSE. We can observe that the performances of all estimation schemes increase as the transmit power grows, which is because the better recovery/estimation performance can be achieved with a higher SNR. Also, the performance gaps between the proposed estimator with other baseline schemes increase as transmit power grows, especially when the training overhead is small. This implies that our method is superior for reducing training overhead, which further demonstrates the effectiveness of the proposed estimator.

Fig. 7 shows the impacts of penalty factor λ on the NMSE. We can observe that the performance of the proposed estimator

decreases first and then increases as λ grows. This is because penalty factor is applied for balancing the tradeoff between data fitting and the sparsity of the solutions. Hence, a smaller λ causes a more sparse matrix but does not fit the data, while a larger λ fits the data well but loses the sparsity. Both of these two scenarios degrade the recovery performance of the proposed estimator.

VIII. Conclusions

This paper studies the channel estimation problem for the RIS aided multi-user MIMO system and proposes a novel uplink channel estimation protocol to estimate the cascaded channels directly. Specifically, we recognize the cascaded channel are typically sparse, and formulate the channel estimation problem into a sparse channel matrix recovery problem using CS techniques, in order to achieve achieve robust channel estimation with limited training overhead. In particular, the sparse channel matrixes of the cascaded channels of all users have a common row-column-block sparsity structure due to the common channel between BS and RIS. Hence, by taking

advantages of such a sparsity structure, we propose a two-step procedure based multi-user joint channel estimator. Moreover, the optimization of the training reflection coefficient sequences at the RIS is studied to improve the estimation performance. Finally, the simulation results validate the effectiveness of the proposed estimator.

APPENDIX A

A. Derivation of Sufficient Statistics

To begin with, we define $\boldsymbol{A} = \boldsymbol{A}_T^{:,\Omega_{\mathrm{D}}}$, $\boldsymbol{C}_b = \boldsymbol{A}(\sum_{k=1}^{K} \boldsymbol{X}_k^H \boldsymbol{A}_R^H \boldsymbol{v}_b)(\sum_{k=1}^{K} \boldsymbol{X}_k^H \boldsymbol{A}_R^H \boldsymbol{v}_b)^H \boldsymbol{A}^H + \delta^2 \boldsymbol{I}_M$, and $\boldsymbol{P}_{b,t} = \sum_{k=1}^{K} \boldsymbol{X}_k^H \boldsymbol{A}_R^H \boldsymbol{v}_b [\boldsymbol{s}_k]_t$ for notation simplicity. Then, if $K \to \infty$, we have

$$\boldsymbol{P}_{b,t} = \boldsymbol{0}, \text{ and } \boldsymbol{C}_b = \xi \boldsymbol{A} \boldsymbol{A}^H + \delta^2 \boldsymbol{I}_M, \tag{51}$$

where $\xi = PKL$. Then, from (2), the likelihood function is given by

$$\mathrm{P}(\boldsymbol{Y}|\boldsymbol{A})$$
$$= \frac{\exp\left(-\sum_{b=1}^{B}\sum_{t=1}^{T}(\boldsymbol{Y}_b^{:,t} - \boldsymbol{A}\boldsymbol{P}_{b,t})^H \boldsymbol{C}_b^{-1}(\boldsymbol{Y}_b^{:,t} - \boldsymbol{A}\boldsymbol{P}_{b,t})\right)}{\pi^{MBT} \prod_{b=1}^{B} \det\left|\boldsymbol{C}_b^{-1}\right|^T}$$
$$\stackrel{(a)}{=} \frac{\exp\left(-BT\mathrm{Tr}\left((\xi \boldsymbol{A}\boldsymbol{A}^H + \delta^2 \boldsymbol{I}_M)^{-1}\hat{\boldsymbol{C}}\right)\right)}{\pi^{MBT} \det\left|\xi \boldsymbol{A}\boldsymbol{A}^H + \delta^2 \boldsymbol{I}_M\right|^{BT}}, \tag{52}$$

where in (a) we use (51) for $K \to \infty$. Therefore, it is clear that the likelihood function depends on $\boldsymbol{D}$ only via $\hat{\boldsymbol{C}}$. According to the Fischer-Neyman factorization theorem [43], [44], it follows that $\hat{\boldsymbol{C}}$ is a sufficient statistics for estimating the subspace spanned by $\boldsymbol{A}$.

B. Derivation of Basis Optimization for Subspace Estimation

Since $\hat{\boldsymbol{C}}$ is a sufficient statistics, we can design an optimal *maximum-likelihood* (ML) estimator by minimizing its likelihood function as

$$\min_{\boldsymbol{A}} \mathrm{Tr}((\xi \boldsymbol{A}\boldsymbol{A}^H + \delta^2 \boldsymbol{I}_M)^{-1}\hat{\boldsymbol{C}}) + \det|\xi \boldsymbol{A}\boldsymbol{A}^H + \delta^2 \boldsymbol{I}_M|. \tag{53}$$

Let $\boldsymbol{W}_{\|}\boldsymbol{\Delta}\boldsymbol{W}_{\|}^H$ be the eigenvalue decomposition of $\xi \boldsymbol{A}\boldsymbol{A}^H$, and $\boldsymbol{\Delta} = \mathrm{diag}(\Delta_1, \Delta_2, \cdots, \Delta_{N_f})$ be the eigenvalue matrix where the eigenvalue Δ_n is ordered in a decreasing order of magnitude. It is straightforward to know that $\mathrm{span}(\boldsymbol{A})$ is equal to the linear span of $\boldsymbol{W}_{\|} \in \mathbb{C}^{M \times N_f}$, i.e.,

$$\mathrm{span}(\boldsymbol{A}) = \mathrm{span}(\boldsymbol{W}_{\|}). \tag{54}$$

Further denoting $\boldsymbol{W}_{\perp}$ be the rest $(M - N_f)$ normalized orthogonal basis, we have

$$\left(\xi \boldsymbol{A}\boldsymbol{A}^H + \delta^2 \boldsymbol{I}_M\right)^{-1} = \boldsymbol{W}_{\|}\tilde{\boldsymbol{\Delta}}\boldsymbol{W}_{\|}^H + \frac{1}{\delta^2}\boldsymbol{W}_{\perp}\boldsymbol{W}_{\perp}^H, \tag{55}$$

where $\tilde{\boldsymbol{\Delta}} = \mathrm{diag}(\frac{1}{\Delta_1 + \delta^2}, \frac{1}{\Delta_2 + \delta^2}, \cdots, \frac{1}{\Delta_{N_f} + \delta^2}) \in \mathbb{C}^{N_f \times N_f}$. As aforementioned, we only need to estimate subspace $\boldsymbol{W}_{\|}$

without requiring exact AoD vectors $\boldsymbol{A}$. Hence, substituting (55) into (53), we have the following equivalent problem of (53), i.e.,

$$\min_{\boldsymbol{W}_{\|}} \mathrm{Tr}\left(\tilde{\boldsymbol{\Delta}}\boldsymbol{W}_{\|}^H \boldsymbol{S}\boldsymbol{\Theta}\boldsymbol{S}^H \boldsymbol{W}_{\|}\right). \tag{56}$$

Since $\begin{bmatrix} \boldsymbol{W}_{\|}^H \\ \boldsymbol{W}_{\perp}^H \end{bmatrix} \boldsymbol{S}\boldsymbol{S}^H \begin{bmatrix} \boldsymbol{W}_{\|}\boldsymbol{W}_{\perp} \end{bmatrix} = \boldsymbol{I}_M$, and $\boldsymbol{S}^H \begin{bmatrix} \boldsymbol{W}_{\|}\boldsymbol{W}_{\perp} \end{bmatrix} \begin{bmatrix} \boldsymbol{W}_{\|}^H \\ \boldsymbol{W}_{\perp}^H \end{bmatrix} \boldsymbol{S} = \boldsymbol{I}_M$, problem (56) is equivalent to the following problem

$$\min_{0 \le a_{n,m}} \sum_{n=1}^{N_f}\sum_{m=1}^{M} a_{n,m}\frac{\theta_m}{\Delta_n + \delta^2} \tag{57}$$
$$\text{s.t. } \sum_{m=1}^{m} a_{n,m} = 1, 1 \le n \le N_f,$$
$$\sum_{n=1}^{N_f} a_{n,m} \le 1, 1 \le m \le M,$$

where $a_{n,m} = |(\boldsymbol{W}_{\|}^{:,n})^H \boldsymbol{S}^{:,m}|^2$. This is a convex problem, which can be solve by applying Lagrangian method. Then, the optimal solutions of $a_{n,m}$ is given by

$$a_{n,m} = \begin{cases} 1, 1 \le n = m \le N_f, \\ 0, \text{Others}. \end{cases} \tag{58}$$

Note that $a_{n,m} = 1$ means $\boldsymbol{W}_{\|}^{:,n} = \boldsymbol{S}^{:,m}$, which indicates the optimal solution is $\boldsymbol{W}_{\|} = \boldsymbol{S}_{\|} \stackrel{\Delta}{=} \left[\boldsymbol{S}^{:,1}, \boldsymbol{S}^{:,2}, \cdots, \boldsymbol{S}^{:,N_f}\right]$.

REFERENCES

[1] H. Yang, X. Cao, F. Yang, J. Gao, S. Xu, M. Li, X. Chen, Y. Zhao, Y. Zheng, and S. Li, "A programmable metasurface with dynamic polarization, scattering and focusing control," *Scientific reports*, vol. 6, p. 35692, 2016.

[2] E. Basar, M. Di Renzo, J. de Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *arXiv preprint arXiv:1906.09490*, 2019.

[3] M. Di Renzo, M. Debbah, D.-T. Phan-Huy, A. Zappone, M.-S. Alouini, C. Yuen, V. Sciancalepore, G. C. Alexandropoulos, J. Hoydis, H. Gacanin *et al.*, "Smart radio environments empowered by reconfigurable AI meta-surfaces: an idea whose time has come," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, pp. 1–20, 2019.

[4] Y.-C. Liang, R. Long, Q. Zhang, J. Chen, H. V. Cheng, and H. Guo, "Large intelligent surface/antennas (LISA): Making reflective radios smart," *J. Commun. Inf. Netw.*, vol. 4, no. 2, Jun. 2009.

[5] J. Zhao, "A survey of reconfigurable intelligent surfaces: Towards 6G wireless communication networks with massive mimo 2.0," *arXiv preprint arXiv:1907.04789*, 2019.

[6] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun., to be published, doi: 10.1109/TWC.2019.2936025*, 2019.

[7] R. Zhang, Y.-C. Liang, C. C. Chai, and S. Cui, "Optimal beamforming for two-way multi-antenna relay channel with analogue network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 699–712, 2009.

[8] K. Ntontin, M. Di Renzo, J. Song, F. Lazarakis, J. de Rosny, D.-T. Phan-Huy, O. Simeone, R. Zhang, M. Debbah, G. Lerosey *et al.*, "Reconfigurable intelligent surfaces vs. relaying: Differences, similarities, and performance comparison," *arXiv preprint arXiv:1908.08747*, 2019.

[9] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *arXiv preprint arXiv:1905.00152*, 2019.

[10] J. Chen, Y.-C. Liang, Y. Pei, and H. Guo, "Intelligent reflecting surface: A programmable wireless environment for physical layer security," *IEEE ACCESS*, vol. 7, pp. 82 599 – 82 612, Jun. 2019.

[11] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, 2019.

[12] Q.-U.-A. Nadeem, A. Kammoun, A. Chaaban, M. Debbah, and M.-S. Alouini, "Large intelligent surface assisted MIMO communications," *arXiv preprint arXiv:1903.08127*, 2019.

[13] H. Guo, Y.-C. Liang, J. Chen, and E. G. Larsson, "Weighted sum-rate optimization for intelligent reflecting surface enhanced wireless networks," *arXiv preprint arXiv:1905.07920*, 2019.

[14] C. Pan, H. Ren, K. Wang, W. Xu, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Multicell MIMO communications relying on intelligent reflecting surface," *arXiv preprint arXiv:1907.10864*, 2019.

[15] M. Jung, W. Saad, M. Debbah, and C. S. Hong, "On the optimality of reconfigurable intelligent surfaces (RISs): Passive beamforming, modulation, and resource allocation," *arXiv preprint arXiv:1910.00968*, 2019.

[16] G. Zhou, C. Pan, H. Ren, K. Wang, W. Xu, and A. Nallanathan, "Intelligent reflecting surface aided multigroup multicast miso communication systems," *arXiv preprint arXiv:1909.04606*, 2019.

[17] P. Wang, J. Fang, X. Yuan, Z. Chen, H. Duan, and H. Li, "Intelligent reflecting surface-assisted millimeter wave communications: Joint active and passive precoding design," *arXiv preprint arXiv:1908.10734*, 2019.

[18] N. S. Perović, M. Di Renzo, and M. F. Flanagan, "Channel capacity optimization using reconfigurable intelligent surfaces in indoor mmwave environments," *arXiv preprint arXiv:1910.14310*, 2019.

[19] Y. Yang, S. Zhang, and R. Zhang, "IRS-enhanced OFDM: Power allocation and passive array optimization," *arXiv preprint arXiv:1905.00604*, 2019.

[20] D. Mishra and H. Johansson, "Channel estimation and low-complexity beamforming design for passive intelligent surface assisted MISO wireless energy transfer," in *IEEE ICASSP*, 2019, pp. 4659–4663.

[21] Z.-Q. He and X. Yuan, "Cascaded channel estimation for large intelligent metasurface assisted massive MIMO," *arXiv preprint arXiv:1905.07948*, 2019.

[22] T. L. Jensen and E. De Carvalho, "On optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator," *arXiv preprint arXiv:1909.09440*, 2019.

[23] B. Zheng and R. Zhang, "Intelligent reflecting surface-enhanced OFDM: Channel estimation and reflection optimization," *arXiv preprint arXiv:1909.03272*, 2019.

[24] C. You, B. Zheng, and R. Zhang, "Intelligent reflecting surface with discrete phase shifts: Channel estimation and passive beamforming," *arXiv preprint arXiv:1911.03916*, 2019.

[25] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," *arXiv preprint arXiv:1904.10136*, 2019.

[26] C.-R. Tsai, Y.-H. Liu, and A.-Y. Wu, "Efficient compressive channel estimation for millimeter-wave large-scale antenna systems," *IEEE Trans. Signal Process.*, vol. 66, no. 9, pp. 2414–2428, 2018.

[27] X. Rao and V. K. Lau, "Distributed compressive CSIT estimation and feedback for fdd multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261–3271, 2014.

[28] Y. Ding and B. D. Rao, "Dictionary learning-based sparse channel representation and estimation for fdd massive mimo systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5437–5451, 2018.

[29] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[30] J. Chen, L. Zhang, and Y.-C. Liang, "Exploiting Gaussian mixture model clustering for full-duplex transceiver design," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5802–5816, May 2019.

[31] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Selected papers of hirotugu akaike*. Springer, 1998, pp. 199–213.

[32] A. Paulraj, B. Ottersten, R. Roy, A. Swindlehurst, G. Xu, and T. Kailath, "subspace methods for directions-of-arrival estimation," *Handbook of Statistics*, vol. 10, pp. 693–739, 1993.

[33] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 387–392, 1985.

[34] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl*, vol. 14, no. 5-6, pp. 877–905, Dec.

[35] D. Wipf and S. Nagarajan, "Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 317–329, Apr. 2010.

[36] J. Fang, F. Wang, Y. Shen, H. Li, and R. S. Blum, "Super-resolution compressed sensing for line spectral estimation: An iterative reweighted approach," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4649–4662, Sep. 2016.

[37] R. E. Carrillo and K. E. Barner, "Iteratively re-weighted least squares for sparse signal reconstruction from noisy measurements," in *IEEE Annual Conference on Information Sciences and Systems*, 2009, pp. 448–453.

[38] M. E. Tipping, "Sparse bayesian learning and the relevance vector machine," *Journal of machine learning research*, vol. 1, no. Jun, pp. 211–244, 2001.

[39] J. Chen, L. Zhang, Y.-C. Liang, X. Kang, and R. Zhang, "Resource allocation for wireless-powered IoT networks with short packet communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1447–1461, Feb. 2019.

[40] M. Elad, "Optimized projections for compressed sensing," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5695–5702, Dec. 2007.

[41] J. M. Duarte-Carvajalino and G. Sapiro, "Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1395–1408, Jul. 2009.

[42] S. Abeywickrama, R. Zhang, and C. Yuen, "Intelligent reflecting surface: Practical phase shift model and beamforming optimization," *arXiv preprint arXiv:1907.06002*, 2019.

[43] S. M. Kay, *Fundamentals of statistical signal processing*. Prentice Hall PTR, 1993.

[44] S. Haghighatshoar and G. Caire, "Massive MIMO channel subspace estimation from low-dimensional projections," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 303–318, 2016.