Manifold Optimization Empowered Two-Timescale Channel Estimation for RIS-Assisted Systems

Zheng Huang, *Graduate Student Member, IEEE*, Chen Liu, *Member, IEEE*, Yunchao Song, Hong Wang, Haibo Zhou, *Senior Member, IEEE*, and Sherman Shen, *Fellow, IEEE*

***Abstract*—In this paper, we propose a manifold optimization empowered two-timescale channel estimation (MO-TTCE) scheme for reconfigurable intelligent surface (RIS)-assisted multi-user millimeter wave massive MIMO systems. Since the base station (BS) and RIS remain relatively stationary while the user equipments (UEs) are mobile, the coherence time of the RIS-BS channel is significantly longer than that of the UE-related channels, namely the UE-RIS and UE-BS channels. Therefore, it is sufficient to estimate the RIS-BS channel only once in the large timescale, while frequently estimating the UE-related channels in the small timescale. We leverage the sparse and low-rank properties of the RIS-BS channel and transform the channel estimation problem into a series of sparse low-rank matrix recovery (SLRMR) problems, specifically ℓ_1-norm regularized constrained optimization problems with the feasible region being a complex bounded-rank (CBR) matrix set. To ensure the differentiability of the objective function, we employ a differentiable Huber-γ function as a substitute for the ℓ_1-norm. To handle the non-convex nature of the CBR matrix set, we treat the complex fixed-rank (CFR) matrix set as a CFR manifold and consider the CBR matrix set as a collection of CFR manifolds, thereby employing manifold optimization techniques to solve the problem. Furthermore, in the small timescale, we utilize the downlink pilot transmission and uplink feedback scheme to simultaneously estimate all UE-RIS channels. Simulation results show that the proposed MO-TTCE scheme can enhance the accuracy of channel estimation and reduce pilot overhead.**

***Index Terms*—Channel estimation, reconfigurable intelligent surface, manifold optimization, low-rank constraint.**

I. Introduction

RECONFIGURABLE intelligent surfaces (RISs) have recently emerged as a promising technology for enhancing communication system performance by dynamically reconfiguring the wireless propagation environment [1], [2], [3]. A RIS consists of multiple reconfigurable passive reflecting elements and an intelligent controller connected to a base station (BS). Each reflecting element can be independently controlled to adjust the phase shift of the incident signal, enabling the enhancement of desired signals or suppression of interfering signals [4]. In this manner, RISs have the potential to improve system performance in wireless communication systems.

Obtaining precise channel state information (CSI) is essential to fully realize the benefits of RISs [5]. However, channel estimation in RIS-assisted systems presents new challenges due to the lack of active components in the RIS architecture. On the one hand, conventional approaches cannot independently estimate RIS-related channels, namely the channel between user equipment (UE) and RIS and the channel between RIS and the BS, due to the inability to sense signals. On the other hand, the increasing number of RIS elements increases the channel dimensionality, leading to significant pilot overhead. To address these challenges, the concept of cascaded channels has been introduced, representing the combined channels of UE-RIS and RIS-BS [6]. Mishra et al. initially proposed estimating cascaded channels using a least square (LS)-based approach, sequentially turning on and off each RIS element [7]. However, the channel estimation accuracy is compromised by the low signal-to-noise ratio (SNR) at the receiver, as only one RIS element is turned on at a time. To address this limitation, Jensen et al. proposed a method in which all RIS elements are turned on during training, and the reflection vectors are selected from the columns of the discrete Fourier transform (DFT) matrix [8]. Additionally, Araújo et al. extended the MISO scenario to the MIMO scenario and introduced parallel factor (PARAFAC)-based channel estimation methods to estimate the cascaded channel by decomposing it into BS-RIS and RIS-UE channels using the PARAFAC tensor model [9].

Several works have focused on further reducing the pilot overhead in RIS-assisted systems. Yang et al. proposed the RIS element grouping method, which groups neighboring RIS elements to reduce the effective number of RIS elements and consequently reduces the pilot overhead [10]. Wang et al. extended this method to the multi-user scenario, exploiting the correlation between user channels (i.e., all users share the same RIS-BS channel) to further reduce the pilot overhead [11]. Hu et al. considered the relative stationarity of the BS and RIS compared to mobile UEs and proposed a two-timescale channel

Manuscript received 28 November 2023; revised 9 March 2024; accepted 8 May 2024. Date of publication 28 May 2024; date of current version 17 October 2024. This work was supported in part by the National Science Foundation of China under Grant 62371249 and in part by the Postgraduate Research and Practice Innovation Program of Jiangsu Province under Grant KYCX23_0990. The review of this article was coordinated by Prof. Bin Lin. *(Corresponding author: Chen Liu.)*

Zheng Huang, Chen Liu, and Yunchao Song are with the College of Electronic and Optical Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China (e-mail: 2022020403@njupt.edu.cn; liuch@njupt.edu.cn; songyc@njupt.edu.cn).

Hong Wang is with the School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China (e-mail: wanghong@njupt.edu.cn).

Haibo Zhou is with the College of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China (e-mail: haibozhou@nju.edu.cn).

Sherman Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L3G1, Canada (e-mail: sshen@uwaterloo.ca).

Digital Object Identifier 10.1109/TVT.2024.3405955

0018-9545 © 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: University of Waterloo. Downloaded on December 06,2024 at 18:52:20 UTC from IEEE Xplore. Restrictions apply.

estimation scheme that reduces the long-term pilot overhead by estimating the RIS-BS channel on a large timescale and the UE-related channels (i.e., the UE-RIS and UE-BS channels) on a small timescale [12]. Subsequent studies extended these works to millimeter wave systems, which are characterized by a limited number of available paths and therefore exhibit sparse and low-rank properties. Wang and Lin et al. leveraged the sparsity of millimeter wave channels to propose compressed sensing (CS)-based channel estimation approaches, effectively reducing pilot overhead [13], [14]. Additionally, He et al. leveraged the fixed-rank property of millimeter wave channels to formulate the RIS-related channel estimation problem as a combination of matrix factorization and fixed-rank matrix completion problem, reducing pilot overhead [6]. Lin et al. utilized both the sparse and fixed-rank properties of millimeter wave channels to propose a complex fixed-rank (CFR) manifold optimization-based channel estimation algorithm, enhancing the accuracy of channel estimation [15]. However, this work directly extends conclusions from the real fixed-rank manifold to the CFR manifold without a robust theoretical foundation. Furthermore, all the aforementioned works relied on unrealistic prior knowledge of the channel rank, necessitating a shift from the fixed-rank assumption to a more flexible low-rank assumption. Consequently, effectively combining the sparse and low-rank properties to enhance the accuracy of channel estimation while reducing pilot overhead remains an unresolved challenge. Further research is needed to address this issue.

This paper proposes a manifold optimization empowered two-timescale channel estimation (MO-TTCE) scheme for multi-user millimeter wave massive MIMO systems assisted by RIS. Over a large timescale, a dual-link pilot transmission is implemented to estimate the RIS-BS channels. During this transmission, the BS operates in full-duplex mode, enabling the transmission of pilots from a single BS antenna to the RIS through the downlink channel. Simultaneously, the RIS reflects the received signal back to the remaining BS antennas via the uplink channel. This forms a cascaded BS-RIS-BS channel, known as the dual-link channel, which encompasses crucial information about the RIS-BS channel. The dual-link channel estimation problem is formulated as a sparse low-rank matrix recovery (SLRMR) problem, and a manifold optimization-based sparse low-rank matrix recovery (MO-SLRMR) algorithm is proposed to solve this problem. Based on the estimated dual-link channels, we can then estimate the RIS-BS channel. At a small timescale, a downlink pilot transmission and uplink feedback scheme are adopted to estimate the frequently changing UE-related channels, reducing the pilot overhead. In summary, the contributions of this paper can be outlined as follows:

- The paper leverages the sparse and low-rank properties of millimeter wave channels to formulate the dual-link channel estimation problem as an SLRMR problem. By incorporating these properties, the paper achieves a more precise characterization of millimeter wave channels, resulting in enhanced channel estimation accuracy. Additionally, imposing sparsity and low-rank constraints into the target solution reduces the pilot overhead required to ensure the uniqueness of the solution compared to schemes that rely on either the sparse or low-rank property alone.
- Conventional ℓ_1-norm regularization terms, commonly used to promote sparsity, impose limitations on the applicability of widely used smooth optimization algorithms due to their non-smooth nature. To overcome this limitation, we propose a substitution using the differentiable Huber-γ function. Additionally, considering the non-convexity of the complex bounded-rank (CBR) matrix set, we treat it as a collection of CFR manifolds and utilize manifold optimization techniques to optimize within the CBR matrix set. To ensure a seamless transition between different CFR manifolds, we further derive the geometric properties of the CBR matrix set. Finally, we develop the MO-SLRMR algorithm to solve the SLRMR problem.
- In scenarios where direct UE-BS channels exist, a downlink pilot transmission and uplink feedback scheme is designed to estimate the UE-RIS channels, with the pilot overhead being proportional to the number of RIS elements. This approach offers a reduction in pilot overhead compared to traditional uplink pilot transmission schemes, which require pilot overhead proportional to the product of the number of RIS elements and the number of UEs.

Organization: Section II introduces the system model of RIS-assisted wireless communication and discusses the dual-link pilot transmission under full-duplex modes. Section III presents the proposed MO-TTCE scheme for the multi-user millimeter wave system using the two-timescale channel estimation framework. The details of the core MO-SLRMR algorithm are omitted for brevity but will be discussed later. Section IV derives the geometric properties of the CBR matrix set based on those of the CFR manifold, providing the necessary geometric foundation for algorithm implementation. Section V details the proposed MO-SLRMR algorithm. Section VI presents the simulation results obtained from our experiments. Section VII concludes the paper.

Notation: Matrices are represented by bold uppercase letters, vectors are denoted by bold lowercase letters, and scalars are represented by the normal font. The i-th element of vector $\mathbf{a}$ is denoted by $[\mathbf{a}]_i$, the (i,j)-th element of matrix $\mathbf{A}$ is denoted by $[\mathbf{A}]_{i,j}$, and the i-th row of matrix $\mathbf{A}$ is denoted as $[\mathbf{A}]_{i,:}$. $\mathrm{rank}(\cdot)$, $\mathrm{tr}(\cdot)$, $\cdot^T$, $\cdot^H$, $\cdot^*$, and $\cdot^\dagger$ are the rank, trace, transpose, Hermitian transpose, complex conjugate, and pseudo-inverse of matrices, respectively. $\mathbf{1}_N$ and $\mathbf{0}_N$ denote $N \times 1$ all-ones and all-zeros vector, respectively. $\mathbf{0}_{N\times N}$ and $\mathbf{I}_{N\times N}$ denote $N \times N$ zero and identity matrices, respectively. $\otimes$, $\odot$, and $\bullet$ denote the Kronecker product, Hadamard product, and transpose Khatri-Rao product, respectively. $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}_N, \mathbf{\Sigma})$ represents a circularly symmetric complex Gaussian vector $\mathbf{x}$ with zero mean and covariance matrix $\mathbf{\Sigma}$.

II. System Model

In this paper, we investigate a RIS-assisted multi-user millimeter wave massive MIMO system, as depicted in Fig. 1. The BS employs a uniform planar array (UPA) consisting of

Authorized licensed use limited to: University of Waterloo. Downloaded on December 06,2024 at 18:52:20 UTC from IEEE Xplore. Restrictions apply.

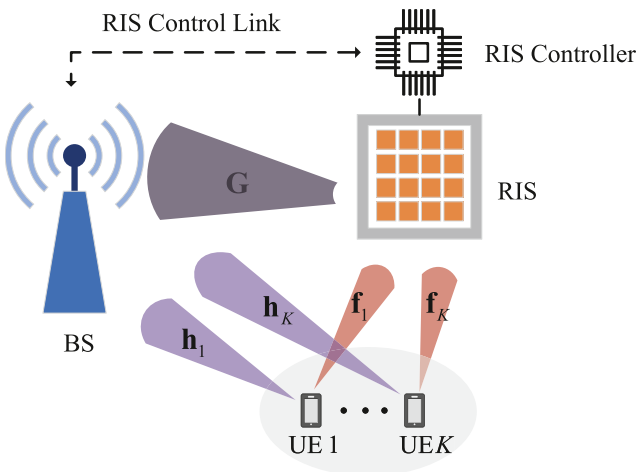

Fig. 1. System model of the considered RIS-assisted multi-user millimeter wave massive MIMO system.

$M = M^{\mathrm{x}}M^{\mathrm{y}}$ antennas, where M^{x} and M^{y} represent the number of antennas in the elevation and azimuth domains, respectively. Similarly, the RIS is equipped with a UPA consisting of $N = N^{\mathrm{x}}N^{\mathrm{y}}$ passive reflecting elements. Moreover, the RIS controller, situated within the RIS itself, communicates wirelessly with the BS (represented by the dashed line) to enable information exchange and control operations. The system serves K UEs, with each UE being equipped with a single antenna. The system operates under the time division duplex (TDD) protocol. Here, we assume that the channels experience block fading, which implies that the channels remain constant over specific time blocks and undergo changes across different time blocks. Specifically, considering that the BS and RIS are generally stationary while the UEs are mobile, the channels related to the UEs change more frequently compared to the RIS-BS channel. As a result, the coherence time of the UE-related channels, represented by T^{u}, is significantly smaller than that of the RIS-BS channel, denoted as T^{b}, i.e., $T^{\mathrm{b}} \gg T^{\mathrm{u}}$ [12]. Within their respective coherence times, the corresponding channels remain constant.

A. Millimeter Wave Channel Model

Let $\mathbf{G} \in \mathbb{C}^{M\times N}$ represent the public channel from the RIS to the BS. The channels from the k-th UE to the RIS and from the k-th UE to the BS are denoted as $\mathbf{f}_k \in \mathbb{C}^N$ and $\mathbf{h}_k \in \mathbb{C}^M$, respectively, where $k = 1, \ldots, K$. These channels $\mathbf{G}$, $\{\mathbf{f}_k\}_{k=1}^K$ and $\{\mathbf{h}_k\}_{k=1}^K$ remain unchanged within their respective coherence times, which can be modeled using the narrowband geometric channel model described in [13]:

$$\mathbf{G} = \sqrt{\frac{MN}{Q}} \sum_{q=1}^{Q} \alpha_q \mathbf{a}_{\mathrm{BS}}\left(\theta_q^{\mathrm{az}}, \theta_q^{\mathrm{el}}\right) \mathbf{a}_{\mathrm{RIS}}^H\left(\phi_q^{\mathrm{az}}, \phi_q^{\mathrm{el}}\right), \tag{1}$$

$$\mathbf{f}_k = \sqrt{\frac{N}{P_k}} \sum_{p_k=1}^{P_k} \beta_{p_k} \mathbf{a}_{\mathrm{RIS}}\left(\psi_{p_k}^{\mathrm{az}}, \psi_{p_k}^{\mathrm{el}}\right), \tag{2}$$

$$\mathbf{h}_k = \sqrt{\frac{M}{L_k}} \sum_{l_k=1}^{L_k} \gamma_{l_k} \mathbf{a}_{\mathrm{BS}}\left(\omega_{l_k}^{\mathrm{az}}, \omega_{l_k}^{\mathrm{el}}\right), \tag{3}$$

where Q, P_k, and L_k denote the numbers of paths in the respective channels, $\alpha_q, \theta_q^{\mathrm{az}}(\theta_q^{\mathrm{el}})$, and ϕ_q^{az} (ϕ_q^{el}) denote the complex path gain, the azimuth (elevation) angle of arrival (AoA) at the BS, and the azimuth (elevation) angle of departure (AoD) at the RIS of the q-th RIS-BS path, respectively. Similarly, β_{p_k} and $\psi_{p_k}^{\mathrm{az}}(\psi_{p_k}^{\mathrm{el}})$ denote the complex path gain and the azimuth (elevation) AoA at the RIS of the p_k-th UE k-RIS path, respectively. γ_{l_k} and $\omega_{l_k}^{\mathrm{az}}(\omega_{l_k}^{\mathrm{el}})$ denote the complex path gain and the azimuth (elevation) AoA at the BS of the l_k-th UE k-BS path, respectively. Moreover, $\mathbf{a}_{\mathrm{BS}}(\cdot,\cdot)$ and $\mathbf{a}_{\mathrm{RIS}}(\cdot,\cdot)$ denote the array steering vectors of the UPA at BS and RIS, respectively. The steering vector of a half-wavelength spaced UPA with size $M^{\mathrm{x}} \times M^{\mathrm{y}}$ is expressed as [16]:

$$\mathbf{a}\left(\theta^{\mathrm{az}}, \theta^{\mathrm{el}}\right) = \frac{1}{\sqrt{M}} \left[1, e^{ju}, \ldots, e^{j(M^{\mathrm{x}}-1)u}\right]^T \otimes \left[1, e^{jv}, \ldots, e^{j(M^{\mathrm{y}}-1)v}\right]^T, \tag{4}$$

where $u \triangleq \pi\cos(\theta^{\mathrm{el}})$ and $v \triangleq \pi\sin(\theta^{\mathrm{el}})\cos(\theta^{\mathrm{az}})$.

Given the serious path loss of millimeter wave channels, the number of paths Q in the channel $\mathbf{G}$ is significantly smaller than the dimension of $\mathbf{G}$, i.e., $Q \ll \min(M, N)$. Moreover, the AoAs and AoDs are typically well-separated, and the corresponding array steering vectors in (1) are linearly independent. Therefore, we can conclude that $\mathrm{rank}(\mathbf{G}) = Q$.

Next, we present the angular domain channel model of $\mathbf{G}$, which exhibits sparse and low-rank characteristics. To achieve this, we initiate a fine sampling procedure for the value ranges of θ_q^{az}, θ_q^{el}, ϕ_q^{az} and ϕ_q^{az}. Using the obtained discrete sampling points, we can form the over-complete codebooks $\mathbf{F}_{\mathrm{BS}} \in \mathbb{C}^{M\times M_{\mathbf{G}}}$ $(M_{\mathbf{G}} \geq M)$ and $\mathbf{F}_{\mathrm{RIS}} \in \mathbb{C}^{N\times N_{\mathbf{G}}}$ $(N_{\mathbf{G}} \geq N)$. Specifically, taking $\mathbf{F}_{\mathrm{BS}}$ as an example, each column of $\mathbf{F}_{\mathrm{BS}}$ take the form of $\mathbf{a}_{\mathrm{BS}}(\theta_q^{\mathrm{az}}, \theta_q^{\mathrm{el}})$ with θ_q^{az} and θ_q^{el} are sampled from the ranges $[0, \pi)$ and $[-\pi/2, \pi/2)$, respectively. This design ensures that $\mathrm{rank}(\mathbf{F}_{\mathrm{BS}}) = M$ and $\mathrm{rank}(\mathbf{F}_{\mathrm{RIS}}) = N$. Then, the angular domain representation of $\mathbf{G}$ is as follows:

$$\mathbf{G} = \mathbf{F}_{\mathrm{BS}} \mathbf{\Sigma}_{\mathbf{G}} \mathbf{F}_{\mathrm{RIS}}^H, \tag{5}$$

where $\mathbf{\Sigma}_{\mathbf{G}} \in \mathbb{C}^{M_{\mathbf{G}}\times N_{\mathbf{G}}}$ represents the sparse angular domain channel representation of $\mathbf{G}$. Considering the rank property of matrix products [17], we can deduce that

$$\mathrm{rank}\left(\mathbf{\Sigma}_{\mathbf{G}}\right) = \mathrm{rank}\left(\mathbf{G}\right) = Q. \tag{6}$$

This implies that the sparse angular domain channel representation $\mathbf{\Sigma}_{\mathbf{G}}$ preserves the low-rank property of the original channel matrix $\mathbf{G}$. It is worth noting that when the true angles are located on the discrete grid points obtained before, $\mathbf{\Sigma}_{\mathbf{G}}$ contains Q non-zero elements corresponding to Q path components. However, in practical scenarios, the true angles will lie between the sampled grid points, leading to an increase in the number of non-zero elements in the sparsest representation $\mathbf{\Sigma}_{\mathbf{G}}$. These additional non-zero elements primarily concentrate on the codewords corresponding to the location of sampled grid points

Authorized licensed use limited to: University of Waterloo. Downloaded on December 06,2024 at 18:52:20 UTC from IEEE Xplore. Restrictions apply.

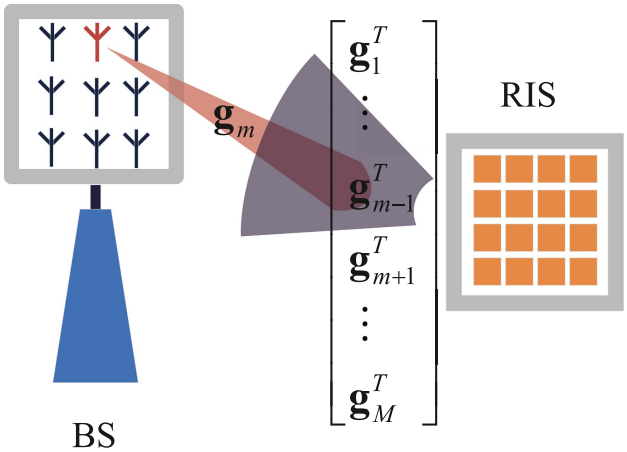

Fig. 2. Illustration of the dual-link pilot transmission scheme.

closest to the true angles [18]. This behavior can be partially captured by the low-rank property.

B. The Dual-Link Pilot Transmission

The independent estimation of the channels related to the RIS presents a significant challenge due to the absence of active components within the RIS. However, by introducing a full-duplex BS, the BS side can autonomously sense the RIS-BS channel $\mathbf{G}$. Here, let BS_m refer to the m-th ($m = 1, \ldots, M$) antenna of BS, and $\mathbf{g}_m \in \mathbb{C}^N$ represents the channel from BS_m to the RIS. According to the channel reciprocity under TDD mode [19], we have $\mathbf{g}_m = [\mathbf{G}]_{m,:}^T$. The dual-link pilot transmission process is illustrated in Fig. 2. In the t-th slot, BS_m operates in transmit mode, sending the pilot signal. Simultaneously, the RIS reflects the received signal back to the remaining $M-1$ antennas of the BS, which operate in receive mode. This forms a cascaded BS-RIS-BS channel, known as the dual-link channel. It is important to note that apart from the RIS reflection, other objects in the environment can also reflect signals back to the remaining $M-1$ antennas of the BS. This forms the environment reflection channel. Moreover, since the BS works in full-duplex mode, there is self-interference among the BS antennas. The received signal at the remaining $M-1$ antennas of the BS after the self-interference mitigation is given by [20]:

$$\begin{aligned}\mathbf{y}_m[t] &= [\mathbf{G}]_{\mathcal{M}\backslash m,:}\,\mathrm{diag}\left(\mathbf{v}[t]\right)\mathbf{g}_m x_m[t] + \mathbf{s}_m x_m[t] \\ &\quad + \mathbf{i}_m[t] + \mathbf{n}_m[t] \\ &= [\mathbf{G}]_{\mathcal{M}\backslash m,:}\,\mathrm{diag}\left(\mathbf{g}_m\right)\mathbf{v}[t] x_m[t] + \mathbf{s}_m x_m[t] \\ &\quad + \mathbf{i}_m[t] + \mathbf{n}_m[t] \\ &= \mathbf{G}_m^T \mathbf{v}[t] x_m[t] + \mathbf{s}_m x_m[t] + \mathbf{i}_m[t] + \mathbf{n}_m[t], \end{aligned} \tag{7}$$

where the set $\mathcal{M}\backslash m$ represents the collection $\{1, \ldots, m-1, m+1, \ldots, M\}$, $\mathrm{diag}(\cdot)$ denotes the diagonalization operation, and $\mathbf{G}_m \triangleq \mathrm{diag}(\mathbf{g}_m)[\mathbf{G}]_{\mathcal{M}\backslash m,:}^T$ represents the dual-link channel from BS_m to the other $M-1$ antennas of BS through the RIS. Moreover, $\mathbf{s}_m$ represents the environment reflection channel from BS_m to the other $M-1$ antennas of BS. The remaining self-interference term is denoted by $\mathbf{i}_m[t] \sim \mathcal{CN}(\mathbf{0}_{M-1}, \sigma_{\mathrm{i}}^2 \mathbf{I}_{(M-1)\times(M-1)})$, while $\mathbf{n}_m[t] \sim \mathcal{CN}(\mathbf{0}_{M-1}, \sigma_{\mathrm{n}}^2 \mathbf{I}_{(M-1)\times(M-1)})$ represents the additive white Gaussian noise (AWGN) [21]. By employing the mitigation technique proposed in [22], the power of $\mathbf{i}_m[t]$ can be reduced to twice that of $\mathbf{n}_m[t]$, i.e., $\sigma_{\mathrm{i}} = \sqrt{2}\sigma_{\mathrm{n}}$. Furthermore, the RIS reflection vector is denoted by $\mathbf{v}[t] \in \mathbf{C}^{N\times 1}$, where each element $[\mathbf{v}[t]]_n$ represents the reflecting coefficient of the n-th RIS reflecting element. To maximize the signal reflection, we assume that $|[\mathbf{v}[t]]_n| \equiv 1, n = 1, \ldots, N$. The pilot signal $x_m[t]$ has an average transmit power of P_{b}, i.e., $\mathbb{E}\{|x_m[t]|^2\} = P_{\mathrm{b}}$. Here, the notation $\mathbb{E}\{\cdot\}$ represents the statistical expectation operation.

Next, we demonstrate that the rank of $\mathbf{G}_m$ is equal to Q. Considering the expression $\mathbf{G}_m = \mathrm{diag}(\mathbf{g}_m)[\mathbf{G}]_{\mathcal{M}\backslash m,:}^T$, it can be observed that $\mathrm{diag}(\mathbf{g}_m)$ is non-singular since $\mathbf{g}_m$ contains no zeros. Therefore, we have $\mathrm{rank}(\mathbf{G}_m) = \mathrm{rank}([\mathbf{G}]_{\mathcal{M}\backslash m,:})$. Furthermore, expressing $[\mathbf{G}]_{\mathcal{M}\backslash m,:}$ using the definition of $\mathbf{G}$, we have:

$$\begin{aligned}[\mathbf{G}]_{\mathcal{M}\backslash m,:} = \sqrt{\frac{MN}{Q}} \sum_{q=1}^{Q} \alpha_q \quad & \left[\mathbf{a}_{\mathrm{BS}}\left(\theta_q^{\mathrm{az}}, \theta_q^{\mathrm{el}}\right)\right]_{\mathcal{M}\backslash m} \\ & \cdot \mathbf{a}_{\mathrm{RIS}}^H\left(\phi_q^{\mathrm{az}}, \phi_q^{\mathrm{el}}\right). \end{aligned} \tag{8}$$

As the basis vectors $\{\mathbf{a}_{\mathrm{BS}}(\theta_1^{\mathrm{az}}, \theta_1^{\mathrm{el}}), \ldots, \mathbf{a}_{\mathrm{BS}}(\theta_Q^{\mathrm{az}}, \theta_Q^{\mathrm{el}})\}$ are linearly independent and $Q \ll M$, it can be easily deduced that the basis vectors $\{[\mathbf{a}_{\mathrm{BS}}(\theta_1^{\mathrm{az}}, \theta_1^{\mathrm{el}})]_{\mathcal{M}\backslash m}, \ldots, [\mathbf{a}_{\mathrm{BS}}(\theta_Q^{\mathrm{az}}, \theta_Q^{\mathrm{el}})]_{\mathcal{M}\backslash m}\}$ are also linearly independent. Then, the submatrix $[\mathbf{G}]_{\mathcal{M}\backslash m,:}$ maintains the same rank as the full matrix $\mathbf{G}$, known to be Q. Consequently, we have $\mathrm{rank}(\mathbf{G}_m) = \mathrm{rank}([\mathbf{G}]_{\mathcal{M}\backslash m,:}) = Q$.

Furthermore, we can derive the angular domain channel model of $\mathbf{G}_m$, which also exhibits sparse and low-rank characteristics:

$$\mathbf{G}_m = \mathbf{D}_{\mathrm{RIS}} \boldsymbol{\Sigma}_{\mathbf{G}_m} \mathbf{F}_{\mathrm{BS}_m}^T, \tag{9}$$

where $\mathbf{D}_{\mathrm{RIS}} \triangleq [\mathbf{F}_{\mathrm{RIS}}^* \bullet \mathbf{F}_{\mathrm{RIS}}^*]_{:,1:N_{\mathbf{G}}}$, $\mathbf{F}_{\mathrm{BS}_m} \triangleq [\mathbf{F}_{\mathrm{BS}}]_{\mathcal{M}\backslash m,:}$, and $\boldsymbol{\Sigma}_{\mathbf{G}_m} \in \mathbb{C}^{N_{\mathbf{G}}\times M_{\mathbf{G}}}$ represents the sparse angular domain channel representation of $\mathbf{G}_m$ with rank equal to Q. A detailed proof of this statement can be found in Appendix A.

III. The Proposed MO-TTCE Scheme

In this section, we propose the MO-TTCE scheme to reduce the long-term pilot overhead by exploiting the variation in coherence times among different channels. This scheme leverages the sparse and low-rank properties of channels within the process. The transmission frame associated with this scheme is divided into two distinct Phases. During Phase 1, the public RIS-BS channel $\mathbf{G}$ is estimated. Subsequently, in Phase 2, the UE-related channels $\{\mathbf{f}_k\}_{k=1}^K$ and $\{\mathbf{h}_k\}_{k=1}^K$ are periodically estimated using the previously estimated $\hat{\mathbf{G}}$, after which the transmission stage is initiated. Estimating the high-dimensional and high pilot overhead channel matrix $\mathbf{G}$ only once can effectively reduce the pilot overhead in the long run.

Authorized licensed use limited to: University of Waterloo. Downloaded on December 06,2024 at 18:52:20 UTC from IEEE Xplore. Restrictions apply.

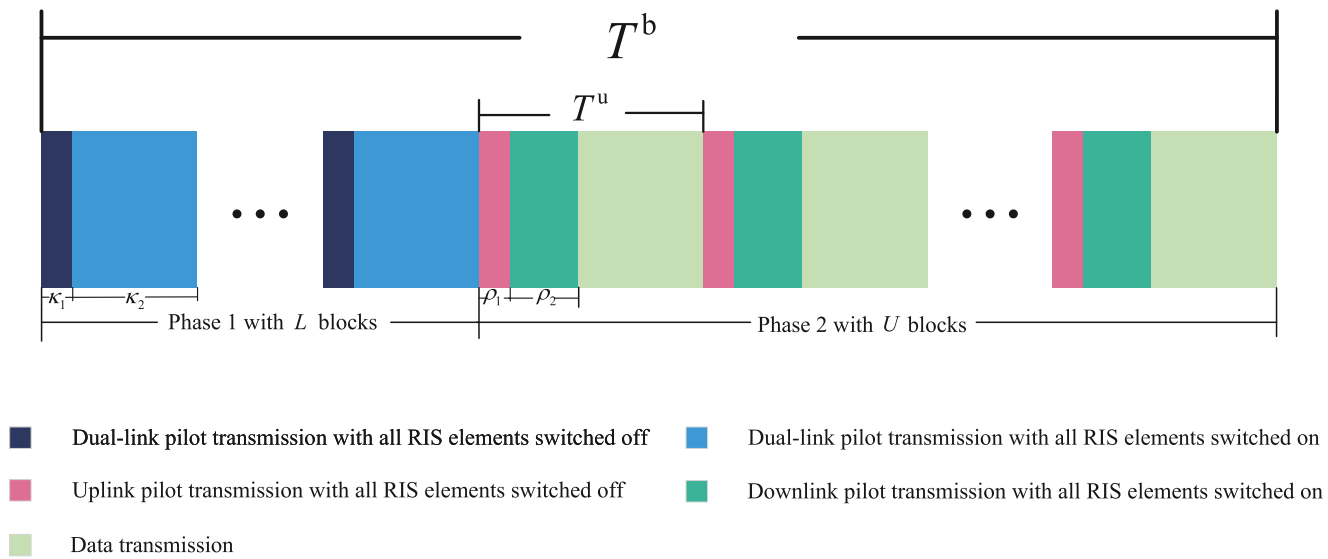

Fig. 3. The proposed two-timescale channel estimation and data transmission frame.

A. Two-Timescale Transmission Frame

As illustrated in Fig. 3, the two-timescale transmission frame comprises two distinct Phases, spanning a total of T^{b} slots. Phase 1, dedicated to estimating $\mathbf{G}$, is divided into L time blocks. Each time block corresponds to a selected BS antenna BS_{m_l} working in transmit mode, where $l = 1, \ldots, L$, and $m_l \in \{1, \ldots, M\}$. Notably, each time block employs a different BS antenna for transmission. Each time block consists of two parts: 1) Dual-link pilot transmission with all RIS elements switched off. This part is aimed at estimating the environment reflections $\mathbf{s}_{m_l}$ and occupies κ_1 slots within the time block. 2) Dual-link pilot transmission with all RIS elements switched on. This part is used to estimate the dual-link channel $\mathbf{G}_{m_l}$ and occupies κ_2 slots within the time block. Finally, based on the estimated $\{\hat{\mathbf{G}}_{m_l}\}_{l=1}^{L}$ in the L time blocks, the $\mathbf{G}$ can be estimated without requiring additional pilot overhead.

Building upon the estimated $\hat{\mathbf{G}}$ obtained in Phase 1, Phase 2 focuses on estimating $\{\mathbf{f}_k\}_{k=1}^{K}$ and $\{\mathbf{h}_k\}_{k=1}^{K}$. Phase 2 is further divided into $U = (T^{\mathrm{b}} - L(\kappa_1 + \kappa_2))/T^{\mathrm{u}}$ time blocks, each occupying T^{u} slots in total. For the sake of simplicity, we assume that $(T^{\mathrm{b}} - L(\kappa_1 + \kappa_2))/T^{\mathrm{u}}$ is an integer. Each time block in Phase 2 consists of three parts: 1) Uplink pilot transmission with all RIS elements switched off. This part is used to estimate UE-BS channels $\{\mathbf{h}_k\}_{k=1}^{K}$ and occupies ρ_1 slots. 2) Downlink pilot transmission with all RIS elements switched on. This part is combined with an uplink feedback to estimate UE-RIS channels $\{\mathbf{f}_k\}_{k=1}^{K}$ and occupies ρ_2 slots. 3) Data transmission. Since all the channels are known, the remaining $T^{\mathrm{u}} - \rho_1 - \rho_2$ slots are used for data transmission.

B. The Estimation of the RIS-BS Channel

To simplify slot counting, we denote the beginning of each time block as 1 in the subsequent discussion. During the l-th time block, a specific BS_{m_l} is activated in transmit mode. In the first part, BS_{m_l} continuously transmits pilot signals $x_{m_l}[t] \equiv \sqrt{P_{\mathrm{b}}}$ for $t = 1, \ldots, \kappa_1$ with all RIS elements turned off. According to (7), during the t-th slot, the remaining antennas of the BS receive signals after the self-interference mitigation, given by:

$$\mathbf{y}_{m_l}[t] = \mathbf{s}_{m_l}\sqrt{P_{\mathrm{b}}} + \mathbf{i}_{m_l}[t] + \mathbf{n}_{m_l}[t]. \tag{10}$$

It is evident that the estimation of $\mathbf{s}_{m_l}$ using the LS criterion can be expressed as:

$$\hat{\mathbf{s}}_{m_l} = \frac{\sum_{t=1}^{\kappa_1} \mathbf{y}_{m_l}[t]}{\kappa_1\sqrt{P_{\mathrm{b}}}}, \tag{11}$$

where at least one pilot is required to estimate $\mathbf{s}_{m_l}$ (i.e., $\kappa_1 \geq 1$).

In the subsequent part, BS_{m_l} continues to transmit the pilot signals $x_{m_l}[t] \equiv \sqrt{P_{\mathrm{b}}}$ for $t = \kappa_1 + 1, \ldots, \kappa_1 + \kappa_2$ with all RIS elements turned on. Using the obtained $\hat{\mathbf{s}}_{m_l}$, we define the effective receive signal $\tilde{\mathbf{y}}_{m_l}[t]$ as:

$$\begin{aligned}\tilde{\mathbf{y}}_{m_l}[t] &\triangleq \mathbf{y}_{m_l}[t] - \hat{\mathbf{s}}_{m_l}\sqrt{P_{\mathrm{b}}} \\ &= \mathbf{G}_{m_l}^{T}\mathbf{v}[t]\sqrt{P_{\mathrm{b}}} + \mathbf{u}_{m_l}[t],\end{aligned} \tag{12}$$

where $\mathbf{u}_{m_l}[t] \triangleq (\mathbf{s}_{m_l} - \hat{\mathbf{s}}_{m_l})\sqrt{P_{\mathrm{b}}} + \mathbf{i}_{m_l}[t] + \mathbf{n}_{m_l}[t]$ can be treated as noise. Furthermore, by stacking κ_2 effective received signals $\{\tilde{\mathbf{y}}_{m_l}[t]\}_{t=\kappa_1+1}^{\kappa_1+\kappa_2}$, we have:

$$\begin{aligned}\tilde{\mathbf{Y}}_{m_l} &\triangleq [\tilde{\mathbf{y}}_{m_l}[\kappa_1 + 1], \ldots, \tilde{\mathbf{y}}_{m_l}[\kappa_1 + \kappa_2]] \\ &= \mathbf{G}_{m_l}^{T}\mathbf{V}_{m_l} + \mathbf{U}_{m_l},\end{aligned} \tag{13}$$

where $\mathbf{V}_{m_l} \triangleq \sqrt{P_{\mathrm{b}}}[\mathbf{v}[\kappa_1 + 1], \ldots, \mathbf{v}[\kappa_1 + \kappa_2]]$, and $\mathbf{U}_{m_l} \triangleq [\mathbf{u}_{m_l}[\kappa_1 + 1], \ldots, \mathbf{u}_{m_l}[\kappa_1 + \kappa_2]]$.

Typically, the conventional LS-based channel estimation scheme requires a minimum of $\kappa_2 = N$ pilot lengths to guarantee a full row-rank $\mathbf{V}_{m_l}$, ensuring the uniqueness of $\mathbf{G}_{m_l}$. However, this LS-based scheme presents two drawbacks: 1) High pilot overhead: It necessitates a large number of pilot transmissions corresponding to the number of RIS elements, resulting in a substantial pilot overhead. 2) Noise sensitivity: The resulting estimation is susceptible to significant noise impact, as it ignores the inherent low-rank property of channels. To address

Authorized licensed use limited to: University of Waterloo. Downloaded on December 06,2024 at 18:52:20 UTC from IEEE Xplore. Restrictions apply.

these limitations, we expand (13) to:

$$\tilde{\mathbf{Y}}_{m_l} = \mathbf{F}_{\mathrm{BS}_{m_l}} \mathbf{\Sigma}_{\mathbf{G}_{m_l}}^T \mathbf{D}_{\mathrm{RIS}}^T \mathbf{V}_{m_l} + \mathbf{U}_{m_l}. \tag{14}$$

By leveraging both the sparse and low-rank properties of the angle-domain channel $\mathbf{\Sigma}_{\mathbf{G}_{m_l}}$, we transform the dual-link channel estimation problem into an SLRMR problem:

$$\begin{aligned} \mathcal{P}1: \hat{\mathbf{\Sigma}}_{\mathbf{G}_{m_l}} = \arg\min_{\mathbf{\Sigma}_{\mathbf{G}_{m_l}}} \quad & f_{m_l}\left(\mathbf{\Sigma}_{\mathbf{G}_{m_l}}\right) \\ \text{s.t.} \quad & \operatorname{rank}\left(\mathbf{\Sigma}_{\mathbf{G}_{m_l}}\right) \leq Q_{\max}, \end{aligned} \tag{15}$$

where $f_{m_l}(\mathbf{\Sigma}_{\mathbf{G}_{m_l}}) \triangleq \frac{1}{2}\|\tilde{\mathbf{Y}}_{m_l} - \mathbf{F}_{\mathrm{BS}_{m_l}} \mathbf{\Sigma}_{\mathbf{G}_{m_l}}^T \mathbf{D}_{\mathrm{RIS}}^T \mathbf{V}_{m_l}\|_F^2 + \mu\|\mathrm{vec}(\mathbf{\Sigma}_{\mathbf{G}_{m_l}})\|_1$. Moreover, $Q_{\max}$ denotes the maximum number of paths between RIS and BS, which can be achieved from long-term statistical channel information [23], $\|\cdot\|_F$ denotes the Frobenius norm, $\|\cdot\|_1$ denotes the ℓ_1-norm, $\mathrm{vec}(\cdot)$ indicates column-wise stacked operation, and μ denotes the regularization parameter. For the effectiveness of the SLRMR problem, the regularization parameter μ should satisfy the following inequality [24]:

$$\mu < \left\| \mathrm{vec}\left(\mathbf{D}_{\mathrm{RIS}}^H \mathbf{V}_{m_l}^* \left(\tilde{\mathbf{Y}}_{m_l}\right)^T \mathbf{F}_{\mathrm{BS}_{m_l}}^* \right) \right\|_\infty, \tag{16}$$

where $\|\cdot\|_\infty$ denotes the ℓ_∞-norm. Additionally, the design of $\mathbf{V}_{m_l}$ must adhere to the restricted isometry property (RIP) condition [25]. To achieve this, the columns of $\mathbf{V}_{m_l}$ can be selected from the Hadamard matrix.

Solving $\mathcal{P}1$ presents two main difficulties: The non-differentiability of the ℓ_1-norm term in the objective function and the non-convex nature of the feasible region. To address these problems, we propose the MO-SLRMR algorithm, described in detail in Section V. Once $\mathcal{P}1$ is solved, we obtain an estimation of the dual-link channel:

$$\hat{\mathbf{G}}_{m_l} = \mathbf{D}_{\mathrm{RIS}} \hat{\mathbf{\Sigma}}_{\mathbf{G}_{m_l}} \mathbf{F}_{\mathrm{BS}_{m_l}}^T. \tag{17}$$

Since $\mathbf{G}_{m_l} = [\mathbf{g}_{m_l} \odot \mathbf{g}_1, \ldots, \mathbf{g}_{m_l} \odot \mathbf{g}_{m_l-1}, \mathbf{g}_{m_l} \odot \mathbf{g}_{m_l+1}, \ldots, \mathbf{g}_{m_l} \odot \mathbf{g}_M]$, we can estimate $\{\mathbf{g}_m\}_{m=1}^M$ using the estimated $\{\hat{\mathbf{G}}_{m_l}\}_{l=1}^L$. However, when $L = 1$, only $M - 1$ equations are available for M unknowns $\{\mathbf{g}_m\}_{m=1}^M$:

$$\mathbf{g}_{m_1} \odot \mathbf{g}_1 = \left[\hat{\mathbf{G}}_{m_1}\right]_{:,1}, \ldots, \mathbf{g}_{m_1} \odot \mathbf{g}_{m_1-1} = \left[\hat{\mathbf{G}}_{m_1}\right]_{:,m_1-1},$$

$$\mathbf{g}_{m_1} \odot \mathbf{g}_{m_1+1} = \left[\hat{\mathbf{G}}_{m_1}\right]_{:,m_1+1}, \ldots, \mathbf{g}_{m_1} \odot \mathbf{g}_M = \left[\hat{\mathbf{G}}_{m_1}\right]_{:,M},$$

leading to multiple possible solutions. To address this, the number L of dual-link channels to be estimated must be at least 2. Subsequently, we can invoke the coordinate descent-based iterative refinement (CD-IR) algorithm proposed in [12] to estimate $\mathbf{G}$ from the estimated $\{\hat{\mathbf{G}}_{m_l}\}_{l=1}^L$. However, due to the property $\mathbf{g}_{m_l} \odot \mathbf{g}_{m_{l'}} = -\mathbf{g}_{m_l} \odot -\mathbf{g}_{m_{l'}}$ for any $l, l' \in \{1, \ldots, L\}$, the overall ± 1 error exists. Therefore, disregarding the impact of noise on channel estimation accuracy, we have:

$$\hat{\mathbf{G}} = \mathbf{G}\mathbf{P}, \tag{18}$$

where $\mathbf{P} \triangleq \mathrm{diag}(\pm 1, \ldots, \pm 1) \in \mathbb{R}^{N \times N}$, $\mathbf{P}^T = \mathbf{P}$, and $\mathbf{P}^T\mathbf{P} = \mathbf{I}_{N \times N}$.

C. The Estimation of the UE-Related Channels

Existing channel estimation techniques have thoroughly investigated the estimation of UE-BS channels $\{\mathbf{h}_k\}_{k=1}^K$. Since these channels are not the primary focus of our work, we employ a simple LS-based channel estimation algorithm to estimate them. Specifically, we implement an uplink pilot transmission scheme with all RIS elements turned off. In this scheme, all users transmit pilot signals $x_{\mathrm{u}_k}[t]$ $(k = 1, \ldots, K)$ to the BS with an average power of P_{u}. Then, at the t-th $(t = 1, \ldots, \rho_1)$ slot, the received signal at the BS is given by:

$$\mathbf{y}[t] = [\mathbf{h}_1, \ldots, \mathbf{h}_K]\, \mathbf{x}_{\mathrm{u}}[t] + \mathbf{n}[t], \tag{19}$$

where $\mathbf{x}_{\mathrm{u}}[t] \triangleq [x_{\mathrm{u}_1}[t], \ldots, x_{\mathrm{u}_K}[t]]^T$, and $\mathbf{n}[t] \sim \mathcal{CN}(\mathbf{0}_M, \sigma_{\mathrm{n}}^2 \mathbf{I}_{M \times M})$ denotes the AWGN. By collecting ρ_1 effective received signals $\{\mathbf{y}[t]\}_{t=1}^{\rho_1}$, we have

$$\begin{aligned} \mathbf{Y} &\triangleq [\mathbf{y}[1], \ldots, \mathbf{y}[\rho_1]] \\ &= [\mathbf{h}_1, \ldots, \mathbf{h}_K]\, \mathbf{X}_{\mathrm{u}} + [\mathbf{n}[1], \ldots, \mathbf{n}[\rho_1]], \end{aligned} \tag{20}$$

where $\mathbf{X}_{\mathrm{u}} \triangleq [\mathbf{x}_{\mathrm{u}}[1], \ldots, \mathbf{x}_{\mathrm{u}}[\rho_1]]$. To ensure the uniqueness of the LS estimator, the number of pilots ρ_1 must be at least K. The LS estimates of the UE-BS channels are then given by:

$$\left[\hat{\mathbf{h}}_1, \ldots, \hat{\mathbf{h}}_K\right] = \mathbf{Y}\mathbf{X}_{\mathrm{u}}^\dagger, \tag{21}$$

where $\mathbf{X}_{\mathrm{u}}$ can be any full column-rank matrix with each entry having an amplitude of $\sqrt{P_{\mathrm{u}}}$.

The next step involves estimating the UE-RIS channels $\{\mathbf{f}_k\}_{k=1}^K$. Since the RIS cannot sense signals, the estimations of $\mathbf{G}$ and $\{\mathbf{f}_k\}_{k=1}^K$ are inherently coupled. Therefore, the estimation objective during Phase 1, $\mathbf{GP}$, leads to $\{\mathbf{P}\mathbf{f}_k\}_{k=1}^K$ being the desired estimation result during Phase 2. However, $\mathbf{P}$ is unknown, and the angular representations of $\{\mathbf{P}\mathbf{f}_k\}_{k=1}^K$ are not sparse, rendering sparse-inspired algorithms ineffective for estimation. Consequently, we employ conventional LS-based algorithms to estimate them. Two cases require further discussion. In the first case, when the direct links between UEs and BS are blocked, uplink pilot transmission becomes the only viable option. Consider the simplified scenario where the k-th UE transmits constant signals $x_{\mathrm{u}_k}[t] \equiv 1$ while other UEs remain silent. In this case, the received signal at the BS can be expressed as:

$$\mathbf{y}[t] = \mathbf{G}\mathrm{diag}\left(\mathbf{v}[t]\right) \mathbf{f}_k + \mathbf{n}[t]. \tag{22}$$

Since the rank of $\mathbf{G}\mathrm{diag}(\mathbf{v}[t])$ is equal to the rank of $\mathbf{G}$ (which is Q), the LS estimator of $\mathbf{f}_k$ cannot guarantee a unique solution. Hence, the low-rank property of channel $\mathbf{G}$ necessitates a higher pilot overhead to ensure the uniqueness of the LS estimator. To be more specific, the minimum pilot overhead required for LS estimation of the UE-RIS channels scales with $K\lceil N/\mathrm{rank}(\mathbf{G})\rceil$, where $\lceil\cdot\rceil$ denotes the rounding-up operation. In the second case, when the direct links between UEs and BS exist, we can leverage the communication established by the direct channel between the BS and UEs to propose a downlink pilot transmission and uplink feedback scheme that enables the estimation of all UE-RIS channels using a pilot length of only N. In the following, we introduce this scheme in detail.

Authorized licensed use limited to: University of Waterloo. Downloaded on December 06,2024 at 18:52:20 UTC from IEEE Xplore. Restrictions apply.

At the t-th ($t = \rho_1 + 1, \ldots, \rho_1 + \rho_2$) slot, the BS sends pilot signals $\mathbf{x}[t]$ to all UEs with an average power of P_{b}. Then, the received signal $y_{\mathrm{u}_k}[t]$ of k-th UE can be expressed as:

$$\begin{aligned} y_{\mathrm{u}_k}[t] &= \mathbf{f}_k^T \mathrm{diag}\left(\mathbf{v}[t]\right) \mathbf{G}^T \mathbf{x}[t] + \mathbf{h}_k^T \mathbf{x}[t] + n_k[t] \\ &= \mathbf{x}[t]^T \mathbf{G} \mathrm{diag}\left(\mathbf{v}[t]\right) \mathbf{f}_k + \mathbf{h}_k^T \mathbf{x}[t] + n_{\mathrm{u}_k}[t], \end{aligned} \tag{23}$$

where $n_{\mathrm{u}_k}[t] \sim \mathcal{CN}(\mathbf{0}, \sigma_{\mathrm{n}}^2)$ denotes the AWGN. By removing the direct link components estimated by (21) from (23), we define the effective received signal $\tilde{y}_{\mathrm{u}_k}[t]$ as:

$$\begin{aligned} \tilde{y}_{\mathrm{u}_k}[t] &\triangleq y_{\mathrm{u}_k}[t] - \hat{\mathbf{h}}_k^T \mathbf{x}[t] \\ &= \mathbf{x}[t]^T \mathbf{G} \mathrm{diag}\left(\mathbf{v}[t]\right) \mathbf{f}_k + u_{\mathrm{u}_k}[t], \end{aligned} \tag{24}$$

where $u_{\mathrm{u}_k}[t] \triangleq (\mathbf{h}_k - \hat{\mathbf{h}}_k)^T \mathbf{x}[t] + n_{\mathrm{u}_k}[t]$.

Then, stack the receive signals $\{\tilde{y}_{\mathrm{u}_k}[t]\}_{t=\rho_1+1}^{\rho_1+\rho_2}$ as:

$$\begin{aligned} \tilde{\mathbf{y}}_{\mathrm{u}_k} &\triangleq [\tilde{y}_{\mathrm{u}_k}[\rho_1 + 1], \ldots, \tilde{y}_{\mathrm{u}_k}[\rho_1 + \rho_2]]^T \\ &= \mathbf{A}\mathbf{f}_k + \mathbf{u}_{\mathrm{u}_k}, \end{aligned} \tag{25}$$

where

$$\mathbf{A} \triangleq \begin{bmatrix} \mathbf{x}[\rho_1 + 1]^T \mathbf{G} \mathrm{diag}\left(\mathbf{v}[\rho_1 + 1]\right) \\ \vdots \\ \mathbf{x}[\rho_1 + \rho_2]^T \mathbf{G} \mathrm{diag}\left(\mathbf{v}[\rho_1 + \rho_2]\right) \end{bmatrix},$$

and $\mathbf{u}_{\mathrm{u}_k} \triangleq [u_{\mathrm{u}_k}[\rho_1 + 1], \ldots, u_{\mathrm{u}_k}[\rho_1 + \rho_2]]^T$.

To ensure the uniqueness of the LS estimator, we guarantee that $\mathbf{A}$ is a full column-rank matrix by randomly selecting reflection coefficient vectors $\{\mathbf{v}[t]\}_{t=\rho_1+1}^{\rho_1+\rho_2}$ from the columns of DFT matrix, and pilots $\{\mathbf{x}[t]\}_{t=\rho_1+1}^{\rho_1+\rho_2}$ from the columns of DFT matrix normalized by the transmitted power P_{b}, with $\rho_2 \geq N$. Then, the LS estimation of $\mathbf{f}_k$ is given by:

$$\hat{\mathbf{f}}_k = \hat{\mathbf{A}}^{\dagger} \tilde{\mathbf{y}}_{\mathrm{u}_k}, \tag{26}$$

where $\hat{\mathbf{A}}$ is obtained by inserting estimated $\hat{\mathbf{G}}$ in Phase 1 into the definition of $\mathbf{A}$.

Since the BS does not know $\tilde{\mathbf{y}}_{\mathrm{u}_k}$, each UE should feed back the received signals $y_{\mathrm{u}_k}[t]$ to the BS through the UE-BS channel. In conclusion, the pilot overhead for estimating the UE-related channels is $(\rho_1 + \rho_2) \geq (N + K)$ in the second case, and the total feedback overhead is $K\rho_2$.

IV. The Geometric Properties Derivation

We first define the CBR matrix set, denoted by $\mathbb{C}_{\leq r}^{a_1 \times a_2}$, as the set comprising all complex matrices of dimensions $a_1 \times a_2$ with rank at most r, i.e., $\mathbb{C}_{\leq r}^{a_1 \times a_2} \triangleq \{\mathbf{X} \in \mathbb{C}^{a_1 \times a_2} : \mathrm{rank}(\mathbf{X}) \leq r\}$. Similarly, the CFR matrix set, denoted by $\mathbb{C}_k^{a_1 \times a_2}$, consists of complex matrices of dimensions $a_1 \times a_2$ with rank k, i.e., $\mathbb{C}_k^{a_1 \times a_2} \triangleq \{\mathbf{X} \in \mathbb{C}^{a_1 \times a_2} : \mathrm{rank}(\mathbf{X}) = k\}$. $\mathbb{C}_{\leq r}^{a_1 \times a_2}$ can be effectively represented by the collection $\{\mathbb{C}_k^{a_1 \times a_2}\}_{k=0}^{r}$.

Notably, $\mathbb{C}_k^{a_1 \times a_2}$ exhibits both non-linear and non-convex characteristics, rendering traditional convex optimization algorithms ineffective for optimization. However, the inherent manifold properties of $\mathbb{C}_k^{a_1 \times a_2}$ render it well-suited for manifold optimization techniques. While traditional convex optimization algorithms rely on linear vector space structures, manifold optimization leverages the non-linear geometry of the feasible region. Despite this distinction, manifold optimization frameworks share similarities with traditional convex optimization frameworks. The process operates iteratively, starting from an initial point on the manifold, and repeats the following steps until a local minimum is found: i) Determine a search direction within the tangent space; ii) Perform a search on the manifold in the direction of the determined search direction; iii) Cease searching upon encountering the next iteration point that causes the objective function to descend appropriately. The search performed on the manifold is realized by the retraction operation, a geometric technique that utilizes matrix algebra to approximate geodesic motion. In conclusion, the effective implementation of manifold optimization on $\mathbb{C}_k^{a_1 \times a_2}$ hinges on deriving the necessary geometric properties, including the tangent space, tangent space projection, retraction, and other relevant geometric concepts.

A. The CFR Manifold $\mathbb{C}_k^{a_1 \times a_2}$

First, we demonstrate that the CFR matrix set $\mathbb{C}_k^{a_1 \times a_2}$ forms a manifold.

Theorem 1: The CFR matrix set $\mathbb{C}_k^{a_1 \times a_2}$ is a manifold of dimension $2k(a_1 + a_2 - k)$.

Proof: See Appendix B.

Then, we derive the tangent space and normal space expression of $\mathbb{C}_k^{a_1 \times a_2}$.

Proposition 1: Endow the complex space $\mathbb{C}^{a_1 \times a_2}$ with the standard inner product $\langle \mathbf{X}_1, \mathbf{X}_2 \rangle = \Re\{\mathrm{tr}(\mathbf{X}_1^H \mathbf{X}_2)\}$ for all $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{C}^{a_1 \times a_2}$ [26]. Then, the tangent space $T_{\mathbf{X}}\mathbb{C}_k^{a_1 \times a_2}$ and normal space $(T_{\mathbf{X}}\mathbb{C}_k^{a_1 \times a_2})^{\perp}$ of the manifold $\mathbb{C}_k^{a_1 \times a_2}$ at any $\mathbf{X} \in \mathbb{C}_k^{a_1 \times a_2}$ can be expressed as follows:

$$T_{\mathbf{X}}\mathbb{C}_k^{a_1 \times a_2} = \left\{ \begin{array}{l} \mathbf{U}_{\mathbf{X}} \mathbf{M} \mathbf{V}_{\mathbf{X}}^H + \mathbf{U}_{\mathbf{X},\mathrm{p}} \mathbf{V}_{\mathbf{X}}^H \\ + \mathbf{U}_{\mathbf{X}} \mathbf{V}_{\mathbf{X},\mathrm{p}}^H : \forall \mathbf{M} \in \mathbb{C}^{k \times k}; \\ \forall \mathbf{U}_{\mathbf{X},\mathrm{p}} \in \mathbb{C}^{a_1 \times k}, \\ \text{and } \mathbf{U}_{\mathbf{X},\mathrm{p}}^H \mathbf{U}_{\mathbf{X}} = \mathbf{0}_{k \times k}; \\ \forall \mathbf{V}_{\mathbf{X},\mathrm{p}} \in \mathbb{C}^{a_2 \times k}, \\ \text{and } \mathbf{V}_{\mathbf{X},\mathrm{p}}^H \mathbf{V}_{\mathbf{X}} = \mathbf{0}_{k \times k}. \end{array} \right\}, \tag{27}$$

$$\left(T_{\mathbf{X}}\mathbb{C}_k^{a_1 \times a_2}\right)^{\perp} = \left\{ \begin{array}{l} \mathbf{U}_{\mathbf{X}}^{\perp} \mathbf{W} (\mathbf{V}_{\mathbf{X}}^{\perp})^H : \\ \forall \mathbf{W} \in \mathbb{C}^{(a_1 - k) \times (a_2 - k)}. \end{array} \right\}, \tag{28}$$

where $\mathbf{U}_{\mathbf{X}} \in \mathbb{C}^{a_1 \times k}$ and $\mathbf{V}_{\mathbf{X}} \in \mathbb{C}^{a_2 \times k}$ represent the unitary matrices that contain the first k left and right singular vectors of $\mathbf{X}$, respectively. Moreover, $\mathbf{U}_{\mathbf{X}}^{\perp} \in \mathbb{C}^{a_1 \times (a_1 - k)}$ and $\mathbf{V}_{\mathbf{X}}^{\perp} \in \mathbb{C}^{a_2 \times (a_2 - k)}$ denote unitary matrices satisfying $\mathbf{U}_{\mathbf{X}}^H \mathbf{U}_{\mathbf{X}}^{\perp} = \mathbf{0}_{k \times (a_1 - k)}$ and $\mathbf{V}_{\mathbf{X}}^H \mathbf{V}_{\mathbf{X}}^{\perp} = \mathbf{0}_{k \times (a_2 - k)}$, respectively.

Proof: See Appendix C.

Building upon the expressions for the tangent space and normal space, we derive the corresponding subspace projection operation to determine the search direction within the tangent space.

Proposition 2: For any vector $\zeta \in T_{\mathbf{X}}\mathbb{C}^{a_1 \times a_2} = \mathbb{C}^{a_1 \times a_2}$, where $\mathbf{X} \in \mathbb{C}_k^{a_1 \times a_2}$, the tangent space projection of ζ onto

Authorized licensed use limited to: University of Waterloo. Downloaded on December 06,2024 at 18:52:20 UTC from IEEE Xplore. Restrictions apply.

$T_{\mathbf{X}}\mathbb{C}_k^{a_1\times a_2}$ is given by:

$$P_{T_{\mathbf{X}}\mathbb{C}_k^{a_1\times a_2}}\zeta = \mathbf{U}_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}^H\zeta\mathbf{V}_{\mathbf{X}}\mathbf{V}_{\mathbf{X}}^H + \mathbf{U}_{\mathbf{X}}^{\perp}\left(\mathbf{U}_{\mathbf{X}}^{\perp}\right)^H\zeta\mathbf{V}_{\mathbf{X}}\mathbf{V}_{\mathbf{X}}^H + \mathbf{U}_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}^H\zeta\mathbf{V}_{\mathbf{X}}^{\perp}\left(\mathbf{V}_{\mathbf{X}}^{\perp}\right)^H, \tag{29}$$

and the normal space projection of ζ onto $(T_{\mathbf{X}}\mathbb{C}_k^{a_1\times a_2})^{\perp}$ is given by:

$$P_{\left(T_{\mathbf{X}}\mathbb{C}_k^{a_1\times a_2}\right)^{\perp}}\zeta = \mathbf{U}_{\mathbf{X}}^{\perp}\left(\mathbf{U}_{\mathbf{X}}^{\perp}\right)^H\zeta\mathbf{V}_{\mathbf{X}}^{\perp}(\mathbf{V}_{\mathbf{X}}^{\perp})^H. \tag{30}$$

Proof: See Appendix D.

Finally, the retraction concept, denoted by $R_{\mathbf{X}}(\cdot)$, which facilitates the search on $\mathbb{C}_k^{a_1\times a_2}$, is derived employing the metric projection approach:

$$\begin{aligned} R_{\mathbf{X}}(\xi) &= \underset{\hat{\mathbf{X}}\in\mathbb{C}_k^{a_1\times a_2}}{\arg\min}\left\|(\mathbf{X}+\xi)-\hat{\mathbf{X}}\right\|_F \\ &= \sum_{i=1}^{k} s_i\mathbf{u}_i\mathbf{v}_i^H, \end{aligned} \tag{31}$$

where s_i, $\mathbf{u}_i$ and $\mathbf{v}_i$ are the ordered singular values, left singular vectors, and right singular vectors of $(\mathbf{X}+\xi)$, respectively.

B. The CBR Matrix Set $\mathbb{C}_{\leq R}^{a_1\times a_2}$

The feasible region of $\mathcal{P}1$ is a CBR matrix set. As mentioned earlier, $\mathbb{C}_{\leq r}^{a_1\times a_2}$ is a collection of multiple CFR manifolds $\mathbb{C}_k^{a_1\times a_2}$ $(k=0,\ldots,r)$. Searching over $\mathbb{C}_{\leq r}^{a_1\times a_2}$ necessitates searching through multiple CFR manifolds, making the transition between different CFR manifolds critically important. This transition can be accomplished using feasible direction methods, necessitating the determination of feasible directions at any given point $\mathbf{X}\in\mathbb{C}_k^{a_1\times a_2}$. These feasible directions are defined by the concept of the tangent cone. The following Proposition describes the tangent cone expression for $\mathbb{C}_{\leq r}^{a_1\times a_2}$.

Proposition 3: Endow the complex space $\mathbb{C}^{a_1\times a_2}$ with the standard inner product $\langle\mathbf{X}_1,\mathbf{X}_2\rangle = \Re\{\mathrm{tr}(\mathbf{X}_1^H\mathbf{X}_2)\}$ for all $\mathbf{X}_1,\mathbf{X}_2\in\mathbb{C}^{a_1\times a_2}$. The tangent cone of set $\mathbb{C}_{\leq r}^{a_1\times a_2}$ at any $\mathbf{X}\in\mathbb{C}_k^{a_1\times a_2}$ $(k\leq r)$, denoted as $T_{\mathbf{X}}\mathbb{C}_{\leq r}^{a_1\times a_2}$, can be expressed as the following orthogonal decomposition:

$$T_{\mathbf{X}}\mathbb{C}_{\leq r}^{a_1\times a_2} = T_{\mathbf{X}}\mathbb{C}_k^{a_1\times a_2}\oplus\left(T_{\mathbf{X}}\mathbb{C}_k^{a_1\times a_2}\right)_{\leq(r-k)}^{\perp}, \tag{32}$$

where $\oplus$ represents the direct sum operation, and $(T_{\mathbf{X}}\mathbb{C}_k^{a_1\times a_2})_{\leq(r-k)}^{\perp} \triangleq \{\xi\in(T_{\mathbf{X}}\mathbb{C}_k^{a_1\times a_2})^{\perp} : \mathrm{rank}(\xi)\leq (r-k)\}$.

Proof: See Appendix E.

Expanding on the expressions for the tangent cone, we derive the corresponding tangent cone projection operation to calculate the feasible descent direction. For any descend direction $\zeta\in T_{\mathbf{X}}\mathbb{C}^{a_1\times a_2}=\mathbb{C}^{a_1\times a_2}$, where $\mathbf{X}\in\mathbb{C}_k^{a_1\times a_2}$, the tangent cone projection of ζ onto $T_{\mathbf{X}}\mathbb{C}_{\leq r}^{a_1\times a_2}$ is given by:

$$\begin{aligned} P_{T_{\mathbf{X}}\mathbb{C}_{\leq r}^{a_1\times a_2}}\zeta &\in \underset{\xi\in T_{\mathbf{X}}\mathbb{C}_{\leq r}^{a_1\times a_2}}{\arg\min}\|\zeta-\xi\|_F \\ &= P_{T_{\mathbf{X}}\mathbb{C}_k^{a_1\times a_2}}\zeta + P_{\left(T_{\mathbf{X}}\mathbb{C}_k^{a_1\times a_2}\right)_{\leq(r-k)}^{\perp}}\zeta, \end{aligned} \tag{33}$$

where $P_{(T_{\mathbf{X}}\mathbb{C}_k^{a_1\times a_2})_{\leq(r-k)}^{\perp}}\zeta$ is a best rank-$(r-k)$ approximation of $P_{(T_{\mathbf{X}}\mathbb{C}_k^{a_1\times a_2})^{\perp}}\zeta$.

V. The Proposed MO-SLRMR Algorithm

In this section, we replace the non-differentiable ℓ_1-norm with the Huber-γ function $\Psi(\cdot)$, a differentiable convex lower bound of the ℓ_1-norm [27]. The Huber-γ function is defined as follows:

$$\begin{aligned} \Psi(\mathbf{x}) &= \sum_i \psi([\mathbf{x}]_i), \\ \psi([\mathbf{x}]_i) &= \begin{cases} [\mathbf{x}]_i^*[\mathbf{x}]_i/2\gamma, & |[\mathbf{x}]_i|<\gamma, \\ |[\mathbf{x}]_i|-\gamma/2, & |[\mathbf{x}]_i|\geq\gamma, \end{cases} \end{aligned} \tag{34}$$

where $\gamma>0$ is a real-valued parameter that trades off the fitting error with the ℓ_1-norm and its smoothness. Note that $\Psi(\cdot)$ is convex and has a first-order continuously differentiable property. The derivative of the Huber-γ function is given by:

$$[\nabla\Psi(\mathbf{x})]_i = \begin{cases} [\mathbf{x}]_i/\gamma, & |[\mathbf{x}]_i|<\gamma, \\ [\mathbf{x}]_i/|[\mathbf{x}]_i|, & |[\mathbf{x}]_i|\geq\gamma. \end{cases} \tag{35}$$

By replacing the ℓ_1-norm with the Huber-γ function and the variable $\mathbf{\Sigma}_{\mathbf{G}_{m_l}}$ with $\mathbf{X}$ to maintain consistent variable notation, we transform the optimization problem $\mathcal{P}1$ into the following problem $\mathcal{P}2$:

$$\begin{aligned} \mathcal{P}2: \hat{\mathbf{\Sigma}}_{\mathbf{G}_{m_l}} &= \underset{\mathbf{X}}{\arg\min}\ \tilde{f}_{m_l}(\mathbf{X}) \\ &\quad \text{s.t.} \quad \mathbf{X}\in\mathbb{C}_{\leq Q_{\max}}^{N_{\mathbf{G}}\times M_{\mathbf{G}}}, \end{aligned} \tag{36}$$

where $\tilde{f}_{m_l}(\mathbf{X}) \triangleq \frac{1}{2}\|\tilde{\mathbf{Y}}_{m_l}-\mathbf{F}_{\mathrm{BS}_{m_l}}\mathbf{X}^T\mathbf{D}_{\mathrm{RIS}}^T\mathbf{V}_{m_l}\|_F^2+\mu\Psi(\mathrm{vec}(\mathbf{X}))$.

Based on the geometric properties of the manifold $\mathbb{C}_k^{N_{\mathbf{G}}\times M_{\mathbf{G}}}$ $(k\leq Q_{\max})$ and the set $\mathbb{C}_{\leq Q_{\max}}^{N_{\mathbf{G}}\times M_{\mathbf{G}}}$ discussed in Section IV, we propose the MO-SLRMR algorithm to solve $\mathcal{P}2$. The MO-SLRMR algorithm consists of three main parts: the **CFR MO** algorithm, the **Rank Increase** algorithm, and the **Rank Reduction** algorithm. The CFR MO algorithm is used to move on the CFR manifold. The Rank Increase algorithm and the Rank Reduction algorithm are used to switch between the different CFR manifolds seamlessly. Specifically, since the number of paths is unknown, we initialize the estimated value of Q to be 0 and alternate between the Rank Increase algorithm and the CFR MO algorithm to estimate new path components. If the Rank Increase algorithm does not significantly affect the objective function, we can infer that the path component currently being estimated is small or comparable to the noise component. In this case, we apply the Rank Reduction algorithm to ignore this path component and obtain a final estimation of Q. Finally, we apply the CFR MO algorithm again to improve accuracy.

Next, we will detail these three main parts involved.

1) The CFR MO Algorithm: We utilize the Riemannian conjugate gradient (RCG) method [28] to optimize on the manifold $\mathbb{C}_k^{N_{\mathbf{G}}\times M_{\mathbf{G}}}$. The RCG method is an iterative algorithm, with each iteration employing the line-search method using the following

Authorized licensed use limited to: University of Waterloo. Downloaded on December 06,2024 at 18:52:20 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1: CFR MO Algorithm.

Input: Optimization objective $\tilde{f}_{m_l}$, initial iteration value $\mathbf{X}_0 \in \mathbb{C}_k^{N_\mathbf{G}\times M_\mathbf{G}}$, and stopping criterion parameter ϵ;
1: Set $j=0$, $\mathbf{X}^{(j)}=\mathbf{X}_0$;
2: **repeat**
3: Compute the Euclidean Gradient $\nabla \tilde{f}_{m_l}(\mathbf{X}^{(j)})$ according to (40);
4: Compute the conjugate search direction $\xi^{(j)}$ according to (38);
5: Compute the Armijo backtracking step-size $\alpha^{(j)}$;
6: Update $\mathbf{X}^{(j+1)}$ according to (37);
7: $j \leftarrow j+1$;
8: **until** $\|\mathbf{G}_k(\mathbf{X}^{(j)})\|_F \le \epsilon$;
Output: Optimization result $\mathbf{X}^{(j)} \in \mathbb{C}_k^{N_\mathbf{G}\times M_\mathbf{G}}$.

update formula:

$$\mathbf{X}^{(j+1)} = R_{\mathbf{X}^{(j)}}\left(\alpha^{(j)}\xi^{(j)}\right). \tag{37}$$

Here, $\mathbf{X}^{(j)}$ represents the current point, $\xi^{(j)} \in T_{\mathbf{X}^{(j)}}\mathbb{C}_k^{N_\mathbf{G}\times M_\mathbf{G}}$ denotes the current search direction, $\alpha^{(j)}$ denotes the Armijo backtracking step-size, and $\mathbf{X}^{(j+1)}$ represents the next iteration point. The search direction $\xi^{(j)}$ is determined as a linear combination of the Riemannian gradient $\mathbf{G}_k(\mathbf{X}^{(j)})$ at $\mathbf{X}^{(j)}$ and the previous search direction $\xi^{(j-1)}$, as expressed in the following equation:

$$\xi^{(j)} = \begin{cases} -\mathbf{G}_k\left(\mathbf{X}^{(j)}\right) + \beta^{(j)}\mathcal{T}_{\mathbf{X}^{(j-1)}\to\mathbf{X}^{(j)}}\left(\xi^{(j-1)}\right), j \ge 1, \\ -\mathbf{G}_k\left(\mathbf{X}^{(j)}\right), \quad j=0. \end{cases} \tag{38}$$

In the above equation, the coefficient $\beta^{(j)}$ is selected as the Hestenes-Stiefel parameter, $\mathcal{T}_{\mathbf{X}^{(j-1)}\to\mathbf{X}^{(j)}}(\xi^{(j-1)})$ represents the vector transport operator, defined as:

$$\mathcal{T}_{\mathbf{X}^{(j-1)}\to\mathbf{X}^{(j)}}\left(\xi^{(j-1)}\right) = P_{T_{\mathbf{X}^{(j)}}\mathbb{C}_k^{N_\mathbf{G}\times M_\mathbf{G}}}\left(\xi^{(j-1)}\right), \tag{39}$$

and $\mathbf{G}_k(\mathbf{X}^{(j)}) \triangleq P_{T_{\mathbf{X}^{(j)}}\mathbb{C}_k^{N_\mathbf{G}\times M_\mathbf{G}}}(\nabla \tilde{f}_{m_l}(\mathbf{X}^{(j)}))$, where $\nabla \tilde{f}_{m_l}(\mathbf{X}^{(j)})$ denotes the Euclidean gradient of function $\tilde{f}_{m_l}$ at $\mathbf{X}^{(j)}$, given by:

$$\nabla \tilde{f}_{m_l}(\mathbf{X}^{(j)}) = -\mathbf{D}_{\mathrm{RIS}}^H \mathbf{V}_{m_l}^*(\tilde{\mathbf{Y}}_{m_l} - \mathbf{F}_{\mathrm{BS}_{m_l}}(\mathbf{X}^{(j)})^T \mathbf{D}_{\mathrm{RIS}}^T \mathbf{V}_{m_l})^T \mathbf{F}_{\mathrm{BS}_{m_l}}^* + \mu \nabla \Psi(\mathbf{X}^{(j)}). \tag{40}$$

Overall, the CFR MO algorithm is summarized in Algorithm 1, where ϵ denotes the convergence threshold.

2) The Rank Increase Algorithm: In order to increase the rank of $\mathbf{X}^{(k)} \in \mathbb{C}_k^{N_\mathbf{G}\times M_\mathbf{G}}$, we apply a rank-1 normal correction step, which belongs to the category of feasible direction methods. This step involves utilizing a feasible search direction $\zeta^{(k)}$ that lies within the subspace of normal space $(T_{\mathbf{X}^{(k)}}\mathbb{C}_k^{N_\mathbf{G}\times M_\mathbf{G}})^\perp$, denoted by $\mathbf{N}_{Q_{\max}-k}(\mathbf{X}^{(k)}) \triangleq P_{(T_{\mathbf{X}^{(k)}}\mathbb{C}_k^{N_\mathbf{G}\times M_\mathbf{G}})^\perp_{\le(Q_{\max}-k)}}(\nabla \tilde{f}_{m_l}(\mathbf{X}^{(k)}))$, which is confined to the tangent cone $T_{\mathbf{X}^{(k)}}\mathbb{C}_{\le r}^{N_\mathbf{G}\times M_\mathbf{G}}$. The determination of the feasible search direction $\zeta^{(k)}$ involves finding the best rank-1 approximation of $\mathbf{N}_{Q_{\max}-k}(\mathbf{X}^{(k)})$, which is formulated as the following optimization problem:

$$\zeta^{(k)} \in \underset{\mathrm{rank}(\zeta)=1}{\arg\min} \left\|\mathbf{N}_{Q_{\max}-k}(\mathbf{X}^{(k)}) - \zeta\right\|_F, \tag{41}$$

whose closed-form solution is rank-1 truncated singular value decomposition (SVD) of $\mathbf{N}_{Q_{\max}-k}(\mathbf{X}^{(k)})$. Then, along the search direction $\zeta^{(k)}$, we update the next iteration point as:

$$\mathbf{X}^{(k+1)} = \mathbf{X}^{(k)} + \alpha^{(k)}\zeta^{(k)}, \tag{42}$$

where $\alpha^{(k)}$ denotes the Armijo backtracking step-size.

3) The Rank Reduction Algorithm: In order to reduce the rank of $\mathbf{X}^{(k)} \in \mathbb{C}_k^{N_\mathbf{G}\times M_\mathbf{G}}$, we adopt the rank-1 reduction step by solving the following problem:

$$\mathbf{X}^{(k-1)} \in \underset{\mathrm{rank}(\mathbf{X})=k-1}{\arg\min} \left\|\mathbf{X}^{(k)} - \mathbf{X}\right\|_F, \tag{43}$$

whose closed-form solution is rank-$(k-1)$ truncated SVD of $\mathbf{X}^{(k)}$.

With the details of the three main parts provided, the proposed MO-SLRMR algorithm is summarized in Algorithm 2.

Algorithm 2: MO-SLRMR Algorithm.

Input: Optimization objective $\tilde{f}_{m_l}$, stopping criterion parameter Δ of the Rank Increase algorithm, and the different stopping criterion parameter ϵ_1 and ϵ_2 of the MO-RCG algorithm.
1: Set $k=0$, $\mathbf{X}^{(0)}=\mathbf{0}$, and $f^{(0)}=\tilde{f}_{m_l}(\mathbf{X}^{(0)})$;
2: **repeat**
3: Compute the Euclidean gradient $\nabla \tilde{f}_{m_l}(\mathbf{X}^{(k)})$ according to (40);
4: Perform the **Rank Increase** algorithm: Compute the feasible search direction $\zeta^{(k)}$ according to (41), update $\mathbf{X}^{(k+1)}$ according to (42), and $k \leftarrow k+1$;
5: Update $\mathbf{X}^{(k)}$ via the **CFR MO** algorithm on $\mathbb{C}_k^{N_\mathbf{G}\times M_\mathbf{G}}$ with initial iteration value set to $\mathbf{X}^{(k)}$ and stopping criterion parameter set to $\epsilon=\epsilon_1$;
6: $f^{(k)}=\tilde{f}_{m_l}(\mathbf{X}^{(k)})$;
7: **until** $f^{(k-1)} - f^{(k)} \le \Delta$;
8: Perform the **Rank Reduction** algorithm: Update $\mathbf{X}^{(k-1)}$ according to (43), and $k \leftarrow k-1$;
9: Update $\mathbf{X}^{(k)}$ via the **CFR MO** algorithm on $\mathbb{C}_k^{N_\mathbf{G}\times M_\mathbf{G}}$ with initial iteration value set to $\mathbf{X}^{(k)}$ and stopping criterion parameter set to $\epsilon=\epsilon_2$;
Output: Optimization result $\mathbf{X}^{(k)}$.

VI. Simulations

In this section, we describe the parameter setup and performance metrics used to evaluate the proposed MO-TTCE scheme. The default system parameters are set as $M=16$, $N=64$, $K=5$, $P_\mathrm{b}=10$, and $P_\mathrm{u}=1$. The noise variance σ_n is set according to the setting of SNR, where $\mathrm{SNR} \triangleq 10\log_{10}(P_\mathrm{u}/\sigma_\mathrm{n}^2)(\text{in dB})$. The default channel parameters are set as follows: $Q=P_k=L_k=3$, $Q_{\max}=6$, the propagation gains of all line-of-sight (LoS) paths (α_1, β_1 and γ_1) are distributed as $\mathcal{CN}(0,1)$, the propagation gains of other non-LoS paths are distributed as

Authorized licensed use limited to: University of Waterloo. Downloaded on December 06,2024 at 18:52:20 UTC from IEEE Xplore. Restrictions apply.

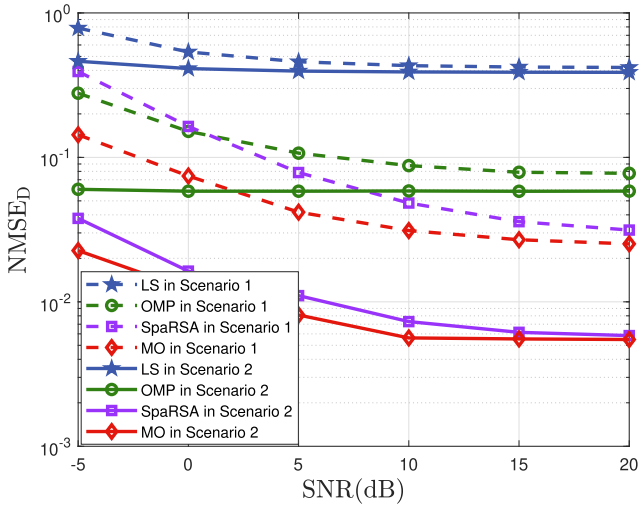

Fig. 4. The $\mathrm{NMSE_D}$ v.s. the SNR (dB) in two different system setups. Scenario 1: $M = 16, N = 64$, and $K = 5$, with transmission frame parameters being $L = 2$, $\kappa_1 = 2$, $\kappa_2 = 39$, $\rho_1 = 5$, and $\rho_2 = 90$. Scenario 2: $M = 64, N = 144$, and $K = 10$, with transmission frame parameters being $L = 2$, $\kappa_1 = 2$, $\kappa_2 = 87$, $\rho_1 = 10$, and $\rho_2 = 202$.

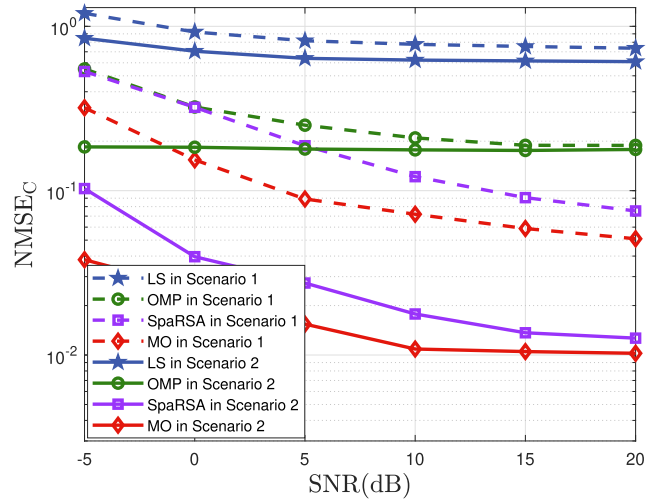

Fig. 5. The $\mathrm{NMSE_C}$ v.s. the SNR (dB) in two different system setups. Scenario 1: $M = 16, N = 64$, and $K = 5$, with transmission frame parameters being $L = 2$, $\kappa_1 = 2$, $\kappa_2 = 39$, $\rho_1 = 5$, and $\rho_2 = 90$. Scenario 2: $M = 64, N = 144$, and $K = 10$, with transmission frame parameters being $L = 2$, $\kappa_1 = 2$, $\kappa_2 = 87$, $\rho_1 = 10$, and $\rho_2 = 202$.

$\mathcal{CN}(0, 10^{-0.5})$, and the azimuth and elevation angles are randomly generated from $[0, \pi)$ and $[-\pi/2, \pi/2)$, respectively. The codebook sizes $M_{\mathbf{G}}$ and $N_{\mathbf{G}}$ are set to $\lceil 1.44\, M \rceil$ and $\lceil 1.44\, N \rceil$, respectively. Moreover, the default parameters of the proposed scheme are listed as follows: $L = 2$, $\kappa_1 = 2$, $\kappa_2 = 39$, $\rho_1 = 5$, $\rho_2 = 90$, the real-valued parameter of Huber-γ function is $\gamma = 1e^{-4}$, and the regularization parameter μ is set according to SNR and system scale. Moreover, the antenna of the full-duplex BS activated in transmit mode is designed as $\mathrm{BS}_{m_l} = \mathrm{BS}_l$. The default stopping criterion parameters in the proposed MO-SLRMR algorithm are set as $\epsilon_1 = 1\mathrm{e}^{-1}$, $\epsilon_2 = 1\mathrm{e}^{-3}$ for distinct accuracy requirement, and Δ is chosen according to SNR and system scale.

The normalized mean square error (NMSE) is chosen as a metric to evaluate the performance of channel estimation algorithms. Specifically, the NMSE of dual-link channels $\{\mathbf{G}_{m_l}\}_{l=1}^{L}$ is defined by:

$$\mathrm{NMSE_D} = \mathbb{E}\left\{ \frac{\sum_{l=1}^{L} \|\mathbf{G}_{m_l} - \hat{\mathbf{G}}_{m_l}\|_F^2}{\sum_{l=1}^{L} \|\mathbf{G}_{m_l}\|_F^2} \right\}, \tag{44}$$

and that of RIS-BS channel $\mathbf{G}$ and UE-RIS channels $\{\mathbf{f}_k\}_{k=1}^{K}$ is defined by:

$$\mathrm{NMSE_C} = \mathbb{E}\left\{ \frac{\sum_{k=1}^{K} \|\mathbf{G}\mathbf{f}_k - \hat{\mathbf{G}}\hat{\mathbf{f}}_k\|_2^2}{\sum_{k=1}^{K} \|\mathbf{G}\mathbf{f}_k\|_2^2} \right\}. \tag{45}$$

To demonstrate the advantages of the proposed MO-based scheme, we compare it with the following three classical benchmark schemes: LS-based scheme [29], orthogonal matching pursuit (OMP)-based scheme [30], and SpaRSA-based scheme [31].

A. Estimation Accuracy Analysis

Figs. 4 and 5 show the performance of $\mathrm{NMSE_D}$ and $\mathrm{NMSE_C}$ versus SNR under two scenarios: Scenario 1: $M = 16, N = 64, K = 5$; Scenario 2: $M = 64, N = 144, K = 10$. In both scenarios, the NMSE of all schemes decreases with the increasing SNR. The LS-based scheme performs the worst because the pilot length κ_2 is insufficient to guarantee the uniqueness of the LS estimator. The OMP- and SpaRSA-based schemes perform better than the LS-based scheme by exploiting sparsity information. The SpaRSA-based scheme, which balances the relationship between fitting error and sparsity, outperforms the OMP-based scheme. The proposed MO-based scheme achieves the best NMSE performance among all schemes by additionally considering the low-rank property of the channel. Moreover, the proposed MO-based scheme requires lower pilot overhead to achieve the same channel estimation accuracy, thereby reducing pilot overhead. All schemes demonstrate superior NMSE performance in Scenario 2 compared to Scenario 1. This improvement can be attributed to the larger number of antennas, which provide a higher power gain, resulting in an enhanced actual SNR. Additionally, the larger antenna array also increases the angular resolution, thereby reducing energy leakage in the angular domain. Consequently, the overall NMSE performance of SpaRSA- and MO-based schemes in Scenario 2 gets greatly improved. However, the gap in NMSE performance between the SpaRSA- and MO-based schemes becomes narrower. This is because the sparser channel representation in the angular domain weakens the significance of considering the low-rank property of the channel, which is a fundamental advantage of MO-based schemes.

Figs. 7 and 6 show the performance of $\mathrm{NMSE_D}$ and $\mathrm{NMSE_C}$ versus the number κ_2 of pilots, respectively. As the length of the pilot increases, the NMSE of all schemes decreases due to the availability of more signals for angular domain channel

Authorized licensed use limited to: University of Waterloo. Downloaded on December 06,2024 at 18:52:20 UTC from IEEE Xplore. Restrictions apply.

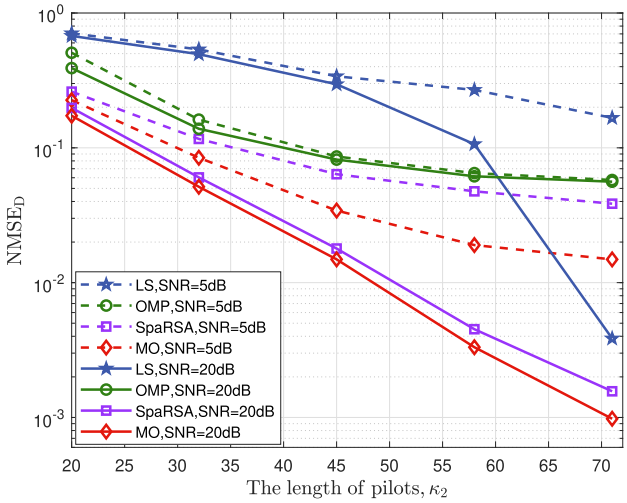

Fig. 6. The $\mathrm{NMSE_C}$ v.s. the length of pilots κ_2 under different SNRs. $M = 16$, $N = 64$ and $K = 5$, with transmission frame parameters being $L = 2$, $\kappa_1 = 2$, $\rho_1 = 5$, and $\rho_2 = 90$.

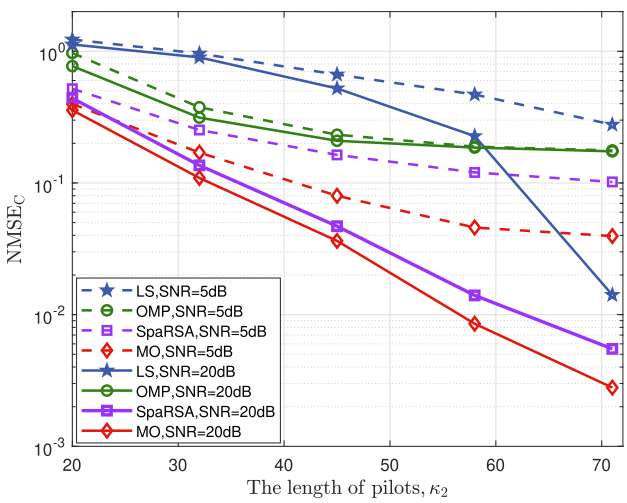

Fig. 7. The $\mathrm{NMSE_D}$ v.s. the length of pilots κ_2 under different SNRs. $M = 16$, $N = 64$ and $K = 5$, with transmission frame parameters being $L = 2$, $\kappa_1 = 2$, $\rho_1 = 5$, and $\rho_2 = 90$.

analysis. More importantly, the proposed MO-based scheme outperforms other schemes at different SNRs. It should be noted that when $\kappa_2 = 72$ and $\mathrm{SNR} = 20\mathrm{dB}$, the LS-based scheme exhibits a considerable performance gain over the OMP-based scheme. This is because, in this case, the length of the pilot is large enough to guarantee the uniqueness of the LS estimator, while the OMP-based scheme focuses too heavily on sparsity. However, at a lower SNR of 5 dB, the LS-based scheme did not exhibit significant improvement, primarily due to the substantial influence of noise.

Fig. 8 compares the $\mathrm{NMSE_C}$ performance of different schemes as a function of the number of dual-link channels L. As L increases, the $\mathrm{NMSE_C}$ decreases, approaching a lower bound for sufficiently large values of L. This indicates that acquiring an

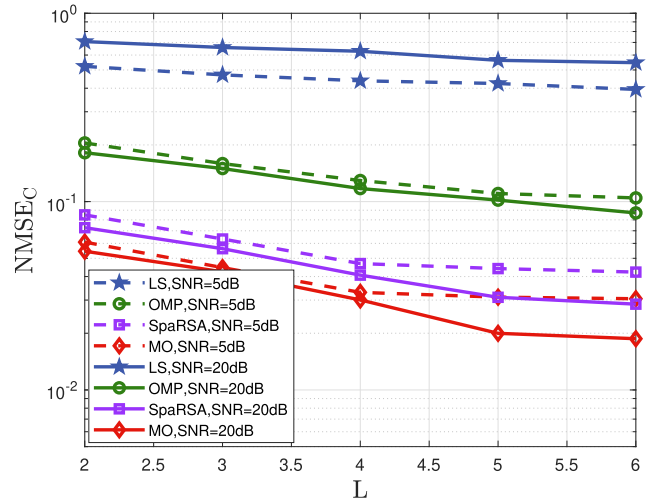

Fig. 8. The $\mathrm{NMSE_C}$ v.s. L under different SNRs. $M = 16$, $N = 64$ and $K = 5$, with transmission frame parameters being $\kappa_1 = 2$, $\kappa_2 = 39$, $\rho_1 = 5$, and $\rho_2 = 90$.

excessive number of dual-link channels does not further enhance the channel estimation performance.

B. Complexity Analysis

We now analyze the computational complexity of the proposed schemes in terms of floating-point operations [32]. The estimation of the quasi-static RIS-BS channel involves the implementation of two key algorithms: the MO-SLRMR algorithm and the CD-IR algorithm. Specifically, at the k-th iteration of the MO-SLRMR algorithm, the computational complexity of the Rank Increase algorithm is $\mathcal{O}(N_{\mathbf{G}}(M+N)\kappa_2 + M_{\mathbf{G}}N_{\mathbf{G}}^2 + M_{\mathbf{G}}^2N_{\mathbf{G}} + M_{\mathbf{G}}^3 + N_{\mathbf{G}}^3)$, the complexity of the Rank Decrease algorithm is $\mathcal{O}(N_{\mathbf{G}}M_{\mathbf{G}}k)$, the complexity of the CFR-MO algorithm is $I_{\mathrm{MO},k}(N_{\mathbf{G}}(M+N)\kappa_2 + M_{\mathbf{G}}N_{\mathbf{G}}^2 + M_{\mathbf{G}}^2N_{\mathbf{G}} + M_{\mathbf{G}}^3 + N_{\mathbf{G}}^3)$, where $I_{\mathrm{MO},k}$ denotes the number of iterations of the CFR-MO algorithms, and the complexity of CD-IR algorithm is $\mathcal{O}(MNLI_{\mathrm{IR}})$, where I_{IR} denotes the number of iterations of iterative refinement. The complexity of channel estimation for UE-BS channels is $\rho_1^2k + \rho_1^3 + M\rho_1 K$, while the complexity of channel estimation for UE-RIS channels is $K(N\rho_2 + N^2\rho_2 + N^3)$. The overall computational complexity is within an acceptable bound.

VII. Conclusion

In this paper, we have proposed the MO-TTCE scheme by exploiting both the sparse and low-rank properties of the channels in RIS-assisted multi-user millimeter wave massive MIMO systems. By incorporating constraints on both low-rankness and sparsity, our MO-TTCE scheme provides a more precise characterization of the channels, resulting in improved channel estimation via a manifold optimization-based algorithm. Simulation results showed that the proposed MO-TTCE scheme provided a higher channel estimation accuracy than the existing

Authorized licensed use limited to: University of Waterloo. Downloaded on December 06,2024 at 18:52:20 UTC from IEEE Xplore. Restrictions apply.

TTCE schemes. In our future work, we will investigate the channel estimation for a wideband RIS-assisted massive MIMO system.

Appendix A
Proof of the Angular Domain Channel Representation of $\mathbf{G}_m$

$$\begin{aligned}\mathbf{G}_m &= \operatorname{diag}(\mathbf{g}_m)\left[\mathbf{F}_{\mathrm{BS}}\boldsymbol{\Sigma}_{\mathbf{G}}\mathbf{F}_{\mathrm{RIS}}^H\right]_{\mathcal{M}\backslash m,:}^T \\ &= \mathbf{g}_m \bullet \left(\left[\mathbf{F}_{\mathrm{BS}}\right]_{\mathcal{M}\backslash m,:}\boldsymbol{\Sigma}_{\mathbf{G}}\mathbf{F}_{\mathrm{RIS}}^H\right)^T \\ &= \left(\left[\mathbf{F}_{\mathrm{BS}}\right]_{m,:}\boldsymbol{\Sigma}_{\mathbf{G}}\mathbf{F}_{\mathrm{RIS}}^H\right)^T \bullet \left(\left[\mathbf{F}_{\mathrm{BS}}\right]_{\mathcal{M}\backslash m,:}\boldsymbol{\Sigma}_{\mathbf{G}}\mathbf{F}_{\mathrm{RIS}}^H\right)^T \\ &= \left(\mathbf{F}_{\mathrm{RIS}}^*\boldsymbol{\Sigma}_{\mathbf{G}}^T\left[\mathbf{F}_{\mathrm{BS}}\right]_{m,:}^T\right) \bullet \left(\mathbf{F}_{\mathrm{RIS}}^*\boldsymbol{\Sigma}_{\mathbf{G}}^T\left[\mathbf{F}_{\mathrm{BS}}\right]_{\mathcal{M}\backslash m,:}^T\right) \\ &= \left(\mathbf{F}_{\mathrm{RIS}}^* \bullet \mathbf{F}_{\mathrm{RIS}}^*\right)\left(\left(\boldsymbol{\Sigma}_{\mathbf{G}}^T\left[\mathbf{F}_{\mathrm{BS}}\right]_{m,:}^T\right) \otimes \boldsymbol{\Sigma}_{\mathbf{G}}^T\right)\left[\mathbf{F}_{\mathrm{BS}}\right]_{\mathcal{M}\backslash m,:}^T.\end{aligned} \tag{46}$$

Since the matrix $(\mathbf{F}_{\mathrm{RIS}}^* \bullet \mathbf{F}_{\mathrm{RIS}}^*) \in \mathbb{C}^{N\times N_{\mathbf{G}}^2}$ contains only $N_{\mathbf{G}}$ distinct columns correspond to its first $N_{\mathbf{G}}$ columns, (46) can be further simplified as:

$$\mathbf{G}_m = \mathbf{D}_{\mathrm{RIS}}\boldsymbol{\Sigma}_{\mathbf{G}_m}\mathbf{F}_{\mathrm{BS}_m}^T, \tag{47}$$

where $\mathbf{D}_{\mathrm{RIS}} \triangleq [\mathbf{F}_{\mathrm{RIS}}^* \bullet \mathbf{F}_{\mathrm{RIS}}^*]_{:,1:N_{\mathbf{G}}}$, $\mathbf{F}_{\mathrm{BS}_m} \triangleq [\mathbf{F}_{\mathrm{BS}}]_{\mathcal{M}\backslash m,:}$, $\boldsymbol{\Sigma}_{\mathbf{G}_m} \in \mathbb{C}^{N_{\mathbf{G}}\times M_{\mathbf{G}}}$ is a merged version of $(\boldsymbol{\Sigma}_{\mathbf{G}}^T[\mathbf{F}_{\mathrm{BS}}]_{m,:}^T) \otimes \boldsymbol{\Sigma}_{\mathbf{G}}^T$ with rank equals Q.

Appendix B
Proof of Theorem 1

Using the complex number field version of the implicit function theorem, we can construct a series of local defining functions to cover the entire set $\mathbb{C}_k^{a_1\times a_2}$ $(k \le a_1, a_2)$, such that each local area of $\mathbb{C}_k^{a_1\times a_2}$ is diffeomorphic to an open subset of $\mathbb{C}^{k(a_1+a_2-k)}$, which implies that $\mathbb{C}_k^{a_1\times a_2}$ is a submanifold of $\mathbb{C}^{a_1\times a_2}$ with dimension $2k(a_1+a_2-k)$.

Given any $\mathbf{X} \in \mathbb{C}_k^{a_1\times a_2}$, $\mathbf{X}$ must contain an invertible submatrix of size $k\times k$. For simplicity, we consider the case where $\mathbf{X}$ can be written in the following block form:

$$\mathbf{X} = \begin{bmatrix}\mathbf{X}_{11} & \mathbf{X}_{12}\\ \mathbf{X}_{21} & \mathbf{X}_{22}\end{bmatrix}, \tag{48}$$

where $\mathbf{X}_{11} \in \mathbb{C}^{k\times k}$ is invertible, $\mathbf{X}_{12} \in \mathbb{C}^{k\times(a_2-k)}$, $\mathbf{X}_{21} \in \mathbb{C}^{(a_1-k)\times k}$, and $\mathbf{X}_{22} \in \mathbb{C}^{(a_1-k)\times(a_2-k)}$.

Since $\operatorname{rank}(\mathbf{X}) = k$ and $\operatorname{rank}(\mathbf{X}_{11}) = k$, the last a_2-k column vectors of $\mathbf{X}$ lie in the subspace spanned by the first k column vectors of $\mathbf{X}$, namely exist a matrix $\mathbf{W} \in \mathbb{C}^{k\times(a_2-k)}$ such that

$$\begin{bmatrix}\mathbf{X}_{12}\\ \mathbf{X}_{22}\end{bmatrix} = \begin{bmatrix}\mathbf{X}_{11}\\ \mathbf{X}_{21}\end{bmatrix}\mathbf{W}. \tag{49}$$

We can derive that $\mathbf{W} = {\mathbf{X}_{11}}^{-1}\mathbf{X}_{12}$, and $\mathbf{X}_{22} = \mathbf{X}_{21}\mathbf{W} = \mathbf{X}_{21}{\mathbf{X}_{11}}^{-1}\mathbf{X}_{12}$. Using this formula, we define a differential mapping h as:

$$\begin{aligned} h:\quad \mathcal{U} \subset \mathbb{C}^{a_1\times a_2} &\to \mathbb{C}^{(a_1-k)\times(a_2-k)} \\ \mathbf{X} &\mapsto \mathbf{X}_{21}{\mathbf{X}_{11}}^{-1}\mathbf{X}_{12} - \mathbf{X}_{22},\end{aligned} \tag{50}$$

where $\mathcal{U}$ denotes the subset of $\mathbb{C}^{a_1\times a_2}$ consisting all matrices whose upper-left submatrix of size $k\times k$ is invertible. By discussion above, we have $h^{-1}(\mathbf{0}) = \mathbb{C}_k^{a_1\times a_2} \cap \mathcal{U}$. Furthermore, the differential of h at $\mathbf{X} \in \mathbb{C}_k^{a_1\times a_2} \cap \mathcal{U}$ is

$$\begin{aligned}\mathrm{D}h(\mathbf{X})[\xi] = \xi_{22} - \xi_{21}\mathbf{X}_{11}^{-1}\mathbf{X}_{12} + \mathbf{X}_{21}\mathbf{X}_{11}^{-1}\xi_{11}\mathbf{X}_{11}^{-1}\mathbf{X}_{12} \\ - \mathbf{X}_{21}\mathbf{X}_{11}^{-1}\xi_{12},\end{aligned} \tag{51}$$

where ξ has the same block structure as $\mathbf{X}$. It is easy to see that $\mathrm{D}h$ is surjective for $\forall \mathbf{X} \in \mathbb{C}_k^{a_1\times a_2} \cap \mathcal{U}$, i.e., h has constant rank $(a_1-k)(a_2-k)$ in $\mathbb{C}_k^{a_1\times a_2} \cap \mathcal{U}$. Referring to the Implicit Function Theorem in [33], it follows that $\forall \mathbf{X} \in \mathbb{C}_k^{a_1\times a_2} \cap \mathcal{U}$, and there exist a neighborhood $\mathcal{X} \in \mathbb{C}^{a_1\times a_2}$ of $\mathbf{X}$, a neighborhood $\mathcal{I} \subset \mathbb{C}^{k\times k} \times \mathbb{C}^{k\times(a_2-k)} \times \mathbb{C}^{(a_1-k)\times k}$ of $(\mathbf{X}_{11}, \mathbf{X}_{12}, \mathbf{X}_{21})$, a neighborhood $\mathcal{D} \subset \mathbb{C}^{(a_1-k)\times(a_2-k)}$ of $\mathbf{X}_{22}$, and a holomorphic function $g: \mathcal{I} \to \mathcal{D}$ such that

$$\begin{aligned}&\begin{bmatrix}\tilde{\mathbf{X}}_{11}, \tilde{\mathbf{X}}_{12}\\ \tilde{\mathbf{X}}_{21}, \tilde{\mathbf{X}}_{22}\end{bmatrix} \in \mathcal{X} \cap \left(\mathbb{C}_k^{a_1\times a_2} \cap \mathcal{U}\right) \\ &\Leftrightarrow \tilde{\mathbf{X}}_{22} = g\left(\tilde{\mathbf{X}}_{11}, \tilde{\mathbf{X}}_{12}, \tilde{\mathbf{X}}_{21}\right),\end{aligned} \tag{52}$$

where $(\tilde{\mathbf{X}}_{11}, \tilde{\mathbf{X}}_{12}, \tilde{\mathbf{X}}_{21}) \in \mathcal{I}$, $\tilde{\mathbf{X}}_{22} \in \mathcal{D}$. This indicates that there contains only $k(a_1+a_2-k)$ free complex variables in the neighborhood $\mathcal{X} \cap (\mathbb{C}_k^{a_1\times a_2} \cap \mathcal{U})$ of $\mathbf{X}$. Thus, we can define a local defining function around $\mathbf{X}$ as:

$$\begin{aligned}\phi:\ \mathcal{X} \cap \left(\mathbb{C}_k^{a_1\times a_2} \cap \mathcal{U}\right) &\to \mathcal{I} \\ \tilde{\mathbf{X}} &\mapsto \left(\tilde{\mathbf{X}}_{11}, \tilde{\mathbf{X}}_{12}, \tilde{\mathbf{X}}_{21}\right).\end{aligned} \tag{53}$$

Similarly, we can construct local defining functions for any other choice of submatrix through the same procedure. These local defining functions cover the entire set $\mathbb{C}_k^{a_1\times a_2}$, and the change of local defining functions is smooth and holomorphic. Hence, $\mathbb{C}_k^{a_1\times a_2}$ is indeed a manifold of dimension $2k(a_1+a_2-k)$.

Appendix C
Proof of Proposition 1

To study the geometric properties of $\mathbb{C}_{\le r}^{a_1\times a_2}$, we first introduce the isomorphism mapping $\mathcal{R}(\cdot)$ [34], which allows us to investigate $\mathbb{C}_k^{a_1\times a_2}$ by studying its real counterpart $\mathcal{R}(\mathbb{C}_k^{a_1\times a_2})$. The isomorphism mapping $\mathcal{R}$ is defined as follows:

$$\mathcal{R}: \mathbf{X} \mapsto \begin{bmatrix}\Re(\mathbf{X}) & \Im(\mathbf{X})\\ -\Im(\mathbf{X}) & \Re(\mathbf{X})\end{bmatrix}, \tag{54}$$

where $\Re(\cdot)$ and $\Im(\cdot)$ denote the real and imaginary parts of the argument, respectively.

The key properties of $\mathcal{R}(\cdot)$ that will be used later are listed below:

(1) $\mathcal{R}(\mathbf{X}_1+\mathbf{X}_2) = \mathcal{R}(\mathbf{X}_1) + \mathcal{R}(\mathbf{X}_2)$,

(2) $\mathcal{R}(\mathbf{X}_1\mathbf{X}_3) = \mathcal{R}(\mathbf{X}_1)\mathcal{R}(\mathbf{X}_3)$,

(3) $\mathcal{R}(\mathbf{X}_1^H) = \mathcal{R}(\mathbf{X}_1)^T$,

(4) $\operatorname{tr}(\mathcal{R}(\mathbf{X}_1)) = 2\Re(\operatorname{tr}(\mathbf{X}_1))$,

Authorized licensed use limited to: University of Waterloo. Downloaded on December 06,2024 at 18:52:20 UTC from IEEE Xplore. Restrictions apply.

where $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{C}^{a_1\times a_2}$, and $\mathbf{X}_3 \in \mathbb{C}^{a_2\times a_3}$. To simplify subsequent discussions, we will refer to $\mathcal{R}(\mathbf{X})$ as $\bar{\mathbf{X}}$.

By utilizing the SVD decomposition of $\mathbf{X} \in \mathbb{C}_k^{a_1\times a_2}$, represented as $\mathbf{X} = \mathbf{U_X}\mathbf{\Sigma_X}\mathbf{V_X}^H$, we can express $\mathcal{R}(\mathbb{C}_k^{a_1\times a_2})$ as follows:

$$\mathcal{R}\left(\mathbb{C}_k^{a_1\times a_2}\right) = \left\{\begin{array}{l} \bar{\mathbf{X}} = \bar{\mathbf{U}}_\mathbf{X}\bar{\mathbf{\Sigma}}_\mathbf{X}\bar{\mathbf{V}}_\mathbf{X}^T : \\ \bar{\mathbf{U}}_\mathbf{X} \in \mathcal{R}\left(\mathrm{St}(a_1,k,\mathbb{C})\right), \\ \bar{\mathbf{V}}_\mathbf{X} \in \mathcal{R}\left(\mathrm{St}(a_2,k,\mathbb{C})\right), \\ \bar{\mathbf{\Sigma}}_\mathbf{X} = \mathrm{diag}(v_1,\ldots,v_k, \\ \qquad v_1,\ldots,v_k), \\ v_1 \geq \ldots \geq v_k. \end{array}\right\}, \tag{55}$$

where $\mathbf{U_X}$ is a complex unitary matrix of size $a_1 \times k$, belonging to the complex Stiefel manifold $\mathrm{St}(a_1,k,\mathbb{C})$ [34]. And $\mathrm{St}(a_1,k,\mathbb{C}) \triangleq \{\mathbf{U_X} \in \mathbb{C}^{a_1\times k} : \mathbf{U_X}^H\mathbf{U_X} = \mathbf{I}_k\}$. Similarly, $\mathbf{V_X} \in \mathrm{St}(a_2,k,\mathbb{C})$. Moreover, $\mathbf{\Sigma_X}$ is a $k \times k$ diagonal matrix with $\{v_i\}_{i=1}^k$ being its diagonal elements.

Next, we define a curve on $\mathcal{R}(\mathbb{C}_k^{a_1\times a_2})$ and find its velocity vector, which represents the tangent vector [33]. For any $\bar{\mathbf{X}} = \bar{\mathbf{U}}_\mathbf{X}\bar{\mathbf{\Sigma}}_\mathbf{X}\bar{\mathbf{V}}_\mathbf{X}^T \in \mathcal{R}(\mathbb{C}_k^{a_1\times a_2})$, we define $\bar{\mathbf{U}}_\mathbf{X}(t)$ as a smooth curve on $\mathcal{R}(\mathrm{St}(a_1,k,\mathbb{C}))$, satisfying $\bar{\mathbf{U}}_\mathbf{X}(0) = \bar{\mathbf{U}}_\mathbf{X}$. Similarly, we define $\bar{\mathbf{V}}_\mathbf{X}(t)$ as a smooth curve on $\mathcal{R}(\mathrm{St}(a_2,k,\mathbb{C}))$, with $\bar{\mathbf{V}}_\mathbf{X}(0) = \bar{\mathbf{V}}_\mathbf{X}$. Additionally, we introduce $\bar{\mathbf{\Sigma}}_\mathbf{X}(t)$ as a smooth curve on $\mathrm{GL}(2k) \cap \mathcal{SP}(k,k)$, satisfying $\bar{\mathbf{\Sigma}}_\mathbf{X}(0) = \bar{\mathbf{\Sigma}}$. Here, $\mathrm{GL}(2k) \triangleq \{\mathbf{X} \in \mathbb{R}^{2k\times 2k} : \mathrm{rank}(\mathbf{X}) = 2k\}$ denotes the general linear group, and $\mathcal{SP}(k,k) \triangleq \mathcal{R}(\mathbb{C}^{k\times k})$ denotes the quasi-symplectic matrix set. Then $\bar{\mathbf{X}}(t) \triangleq \bar{\mathbf{U}}_\mathbf{X}(t)\bar{\mathbf{\Sigma}}_\mathbf{X}(t)\bar{\mathbf{V}}_\mathbf{X}(t)^T$ is a smooth curve on $\mathcal{R}(\mathbb{C}_k^{a_1\times a_2})$ such that $\bar{\mathbf{X}}(0) = \bar{\mathbf{X}}$. The velocity vector $\bar{\mathbf{X}}'(0)$ of $\bar{\mathbf{X}}(t)$ at $t = 0$ is

$$\begin{aligned}\bar{\mathbf{X}}'(0) = {} & \bar{\mathbf{U}}'_\mathbf{X}(0)\bar{\mathbf{\Sigma}}_\mathbf{X}\bar{\mathbf{V}}_\mathbf{X}^T + \bar{\mathbf{U}}_\mathbf{X}\bar{\mathbf{\Sigma}}'_\mathbf{X}(0)\bar{\mathbf{V}}_\mathbf{X}^T \\ & + \bar{\mathbf{U}}_\mathbf{X}\bar{\mathbf{\Sigma}}_\mathbf{X}\bar{\mathbf{V}}'_\mathbf{X}(0)^T,\end{aligned} \tag{56}$$

where $\bar{\mathbf{U}}'_\mathbf{X}(0) \in T_{\bar{\mathbf{U}}_\mathbf{X}}\mathcal{R}(\mathrm{St}(a_1,k,\mathbb{C})) \triangleq \{\bar{\mathbf{U}}_\mathbf{X}\bar{\mathbf{\Omega}}_1 + \bar{\mathbf{U}}_\mathbf{X}^\perp\bar{\mathbf{K}}_1 \in \mathcal{SP}(a_1,k) : \bar{\mathbf{\Omega}}_1 = -\bar{\mathbf{\Omega}}_1^T \in \mathcal{SP}(k,k), \bar{\mathbf{K}}_1 \in \mathcal{SP}(a_1-k,k)\}$, with $\bar{\mathbf{U}}_\mathbf{X}^\perp \in \mathcal{SP}(a_1,a_1-k)$, $(\bar{\mathbf{U}}_\mathbf{X}^\perp)^T\bar{\mathbf{U}}_\mathbf{X}^\perp = \mathbf{I}_{2(a_1-k)\times 2(a_1-k)}$, and $\bar{\mathbf{U}}_\mathbf{X}^T\bar{\mathbf{U}}_\mathbf{X}^\perp = \mathbf{0}_{2k\times 2(a_1-k)}$ [34]. Similarly, $\bar{\mathbf{V}}'_\mathbf{X}(0) \in T_{\bar{\mathbf{V}}_\mathbf{X}}\mathcal{R}(\mathrm{St}(a_2,k,\mathbb{C})) \triangleq \{\bar{\mathbf{V}}_\mathbf{X}\bar{\mathbf{\Omega}}_2 + \bar{\mathbf{V}}_\mathbf{X}^\perp\bar{\mathbf{K}}_2 \in \mathcal{SP}(a_2,k) : \bar{\mathbf{\Omega}}_2 = -\bar{\mathbf{\Omega}}_2^T \in \mathcal{SP}(k,k), \bar{\mathbf{K}}_2 \in \mathcal{SP}(a_2-k,k)\}$, with $\bar{\mathbf{V}}_\mathbf{X}^\perp \in \mathcal{SP}(a_2,a_2-k)$, $(\bar{\mathbf{V}}_\mathbf{X}^\perp)^T\bar{\mathbf{V}}_\mathbf{X}^\perp = \mathbf{I}_{2(a_2-k)\times 2(a_2-k)}$, and $\bar{\mathbf{V}}_\mathbf{X}^T\bar{\mathbf{V}}_\mathbf{X}^\perp = \mathbf{0}_{2k\times 2(a_2-k)}$. Moreover, $\bar{\mathbf{\Sigma}}'_\mathbf{X}(0) \in T_{\bar{\mathbf{\Sigma}}_\mathbf{X}}(\mathrm{GL}(2k) \cap \mathcal{SP}(k,k)) \triangleq \{\bar{\mathbf{A}} : \bar{\mathbf{A}} \in \mathcal{SP}(k,k)\}$. To further expand (56), we have

$$\bar{\mathbf{X}}'(0) = \bar{\mathbf{U}}_\mathbf{X}\bar{\mathbf{M}}\bar{\mathbf{V}}_\mathbf{X}^T + \bar{\mathbf{U}}_{\mathbf{X},\mathrm{p}}\bar{\mathbf{V}}_\mathbf{X}^T + \bar{\mathbf{U}}_\mathbf{X}\bar{\mathbf{V}}_{\mathbf{X},\mathrm{p}}^T, \tag{57}$$

where $\bar{\mathbf{M}} \triangleq \bar{\mathbf{\Omega}}_1\bar{\mathbf{\Sigma}}_\mathbf{X} + \bar{\mathbf{A}} - \bar{\mathbf{\Sigma}}_\mathbf{X}\bar{\mathbf{\Omega}}_2 \in \mathcal{SP}(k,k)$, $\bar{\mathbf{U}}_{\mathbf{X},\mathrm{p}} \triangleq \bar{\mathbf{U}}_\mathbf{X}^\perp\bar{\mathbf{K}}_1\bar{\mathbf{\Sigma}}_\mathbf{X} \in \mathcal{SP}(a_1,k)$ satisfying $\bar{\mathbf{U}}_{\mathbf{X},\mathrm{p}}^T\bar{\mathbf{U}}_\mathbf{X} = \mathbf{0}_{2k\times 2k}$, and $\bar{\mathbf{V}}_{\mathbf{X},\mathrm{p}} \triangleq \bar{\mathbf{V}}_\mathbf{X}^\perp\bar{\mathbf{K}}_2\bar{\mathbf{\Sigma}}_\mathbf{X}^T \in \mathcal{SP}(a_2,k)$ satisfying $\bar{\mathbf{V}}_{\mathbf{X},\mathrm{p}}^T\bar{\mathbf{V}}_\mathbf{X} = \mathbf{0}_{2k\times 2k}$.

Note that $\bar{\mathbf{M}}$, $\bar{\mathbf{U}}_{\mathbf{X},\mathrm{p}}$ and $\bar{\mathbf{V}}_{\mathbf{X},\mathrm{p}}$ contain $2k^2 + 2ka_1 + 2ka_2$ variables and $4k^2$ linear constrains. As a result, there are $2k(a_1 + a_2 - k)$ free variables, i.e., we find a linear subspace of dimension $2k(a_1 + a_2 - k)$ included in the tangent space $T_{\bar{\mathbf{X}}}\mathcal{R}(\mathbb{C}_k^{a_1\times a_2})$, which equals the dimension of $T_{\bar{\mathbf{X}}}\mathcal{R}(\mathbb{C}_k^{a_1\times a_2})$. Consequently, the linear subspace is the entire tangent space, and we have

$$T_{\bar{\mathbf{X}}}\mathcal{R}\left(\mathbb{C}_k^{a_1\times a_2}\right) = \left\{\begin{array}{l} \bar{\mathbf{U}}_\mathbf{X}\bar{\mathbf{M}}\bar{\mathbf{V}}_\mathbf{X}^T + \bar{\mathbf{U}}_{\mathbf{X},\mathrm{p}}\bar{\mathbf{V}}_\mathbf{X}^T \\ +\bar{\mathbf{U}}_\mathbf{X}\bar{\mathbf{V}}_{\mathbf{X},\mathrm{p}}^T : \\ \forall\bar{\mathbf{M}} \in \mathcal{SP}(k,k); \\ \forall\bar{\mathbf{U}}_{\mathbf{X},\mathrm{p}} \in \mathcal{SP}\left(a_1,k\right), \\ \text{and } \bar{\mathbf{U}}_{\mathbf{X},\mathrm{p}}^T\bar{\mathbf{U}}_\mathbf{X} = \mathbf{0}_{k\times k}; \\ \forall\bar{\mathbf{V}}_{\mathbf{X},\mathrm{p}} \in \mathcal{SP}\left(a_2,k\right), \\ \text{and } \bar{\mathbf{V}}_{\mathbf{X},\mathrm{p}}^T\bar{\mathbf{V}}_\mathbf{X} = \mathbf{0}_{k\times k}. \end{array}\right\}. \tag{58}$$

Define the inner product of $\mathcal{SP}(a_1,a_2)$ as [34]:

$$\left\langle\bar{\mathbf{X}}_1,\bar{\mathbf{X}}_2\right\rangle \triangleq \frac{1}{2}\mathrm{tr}\left(\bar{\mathbf{X}}_1^T\bar{\mathbf{X}}_2\right), \forall\bar{\mathbf{X}}_1,\bar{\mathbf{X}}_2 \in \mathcal{SP}\left(a_1,a_2\right). \tag{59}$$

Then, the orthogonal complement of $T_{\bar{\mathbf{X}}}\mathcal{R}(\mathbb{C}_k^{a_1\times a_2})$ in $T_{\bar{\mathbf{X}}}\mathcal{SP}(a_1,a_2)$ is defined as $(T_{\bar{\mathbf{X}}}\mathcal{R}(\mathbb{C}_k^{a_1\times a_2}))^\perp \triangleq \{\bar{\chi} \in T_{\bar{\mathbf{X}}}\mathcal{SP}(a_1,a_2) : \langle\bar{\chi},\bar{\xi}\rangle = 0, \forall\bar{\xi} \in T_{\bar{\mathbf{X}}}\mathcal{R}(\mathbb{C}_k^{a_1\times a_2})\}$. From (57), it's clear that

$$\left(T_{\bar{\mathbf{X}}}\mathcal{R}\left(\mathbb{C}_k^{a_1\times a_2}\right)\right)^\perp = \left\{\begin{array}{l} \bar{\mathbf{U}}_\mathbf{X}^\perp\bar{\mathbf{W}}(\bar{\mathbf{V}}_\mathbf{X}^\perp)^T : \\ \forall\bar{\mathbf{W}} \in \mathcal{SP}\left(a_1-k,a_2-k\right). \end{array}\right\}. \tag{60}$$

Finally, by utilizing the inverse mapping $\mathcal{R}^{-1}$, Proposition 1 can be demonstrated.

Appendix D
Proof of Proposition 2

For any vector $\zeta \in T_\mathbf{X}\mathbb{C}^{a_1\times a_2} = \mathbb{C}^{a_1\times a_2}$, where $\mathbf{X} \in \mathbb{C}_k^{a_1\times a_2}$, ζ can be decomposed into the sum of its components belonging to $T_\mathbf{X}\mathbb{C}_k^{a_1\times a_2}$ and $(T_\mathbf{X}\mathbb{C}_k^{a_1\times a_2})^\perp$. This decomposition is expressed as follows:

$$\zeta = P_{T_\mathbf{X}\mathbb{C}_k^{a_1\times a_2}}\zeta + P_{\left(T_\mathbf{X}\mathbb{C}_k^{a_1\times a_2}\right)^\perp}\zeta, \tag{61}$$

where $P_{T_\mathbf{X}\mathbb{C}_k^{a_1\times a_2}}$ denotes the tangent space projection onto $T_\mathbf{X}\mathbb{C}_k^{a_1\times a_2}$ and $P_{(T_\mathbf{X}\mathbb{C}_k^{a_1\times a_2})^\perp}$ denotes the normal space projection onto $(T_\mathbf{X}\mathbb{C}_k^{a_1\times a_2})^\perp$.

Based on (27), we can derive the expression for normal space projection $P_{(T_\mathbf{X}\mathbb{C}_k^{a_1\times a_2})^\perp}\zeta$. This projection has the following form: $P_{(T_\mathbf{X}\mathbb{C}_k^{a_1\times a_2})^\perp}\zeta = \mathbf{U}_\mathbf{X}^\perp\mathbf{W}(\mathbf{V}_\mathbf{X}^\perp)^H$, where $\mathbf{W} = (\mathbf{U}_\mathbf{X}^\perp)^H\zeta\mathbf{V}_\mathbf{X}^\perp$. Substituting this expression into the original equation, we have:

$$P_{\left(T_\mathbf{X}\mathbb{C}_k^{a_1\times a_2}\right)^\perp}\zeta = \mathbf{U}_\mathbf{X}^\perp\left(\mathbf{U}_\mathbf{X}^\perp\right)^H\zeta\mathbf{V}_\mathbf{X}^\perp(\mathbf{V}_\mathbf{X}^\perp)^H. \tag{62}$$

Based on (61), we can further derive the tangent space projection as:

$$\begin{aligned}P_{T_\mathbf{X}\mathbb{C}_k^{a_1\times a_2}}\zeta = {} & \mathbf{U_X}\mathbf{U_X}^H\zeta\mathbf{V_X}\mathbf{V_X}^H + \mathbf{U}_\mathbf{X}^\perp\left(\mathbf{U}_\mathbf{X}^\perp\right)^H\zeta\mathbf{V_X}\mathbf{V_X}^H \\ & + \mathbf{U_X}\mathbf{U_X}^H\zeta\mathbf{V}_\mathbf{X}^\perp\left(\mathbf{V}_\mathbf{X}^\perp\right)^H.\end{aligned} \tag{63}$$

Authorized licensed use limited to: University of Waterloo. Downloaded on December 06,2024 at 18:52:20 UTC from IEEE Xplore. Restrictions apply.

Appendix E
Proof of Proposition 3

The tangent cone $T_{\bar{\mathbf{X}}}\mathcal{R}(\mathbb{C}_{\leq r}^{a_1\times a_2})$ of $\mathcal{R}(\mathbb{C}_{\leq r}^{a_1\times a_2})$ at a given point $\bar{\mathbf{X}}$ is defined as:

$$T_{\bar{\mathbf{X}}}\mathcal{R}\left(\mathbb{C}_{\leq r}^{a_1\times a_2}\right)=\left\{\begin{array}{l}\bar{\xi}\in\mathcal{SP}\left(a_1,a_2\right):\\ \exists\left\{\bar{\mathbf{X}}_n\right\}\subseteq\mathcal{R}\left(\mathbb{C}_{\leq r}^{a_1\times a_2}\right),\\ \left\{a_n\right\}\subseteq\mathbb{R}^+,\\ \text{s.t. }\bar{\mathbf{X}}_n\to\bar{\mathbf{X}},\\ a_n\left(\bar{\mathbf{X}}_n-\bar{\mathbf{X}}\right)\to\bar{\xi}.\end{array}\right\}. \tag{64}$$

Utilizing the connection between the tangent cone of $\mathcal{R}(\mathbb{C}_{\leq r}^{a_1\times a_2})$ and the tangent space of $\mathcal{R}(\mathbb{C}_{k}^{a_1\times a_2})$, the tangent cone $T_{\bar{\mathbf{X}}}\mathcal{R}(\mathbb{C}_{\leq r}^{a_1\times a_2})$ can be simplified as:

$$\begin{aligned}T_{\bar{\mathbf{X}}}\mathcal{R}\left(\mathbb{C}_{\leq r}^{a_1\times a_2}\right)&=T_{\bar{\mathbf{X}}}\mathcal{R}\left(\mathbb{C}_{k}^{a_1\times a_2}\right)\\&\oplus\left\{\begin{array}{l}\bar{\xi}_{r-k}\in\left(T_{\bar{\mathbf{X}}}\mathcal{R}\left(\mathbb{C}_{k}^{a_1\times a_2}\right)\right)^{\perp}:\\ \text{rank}\left(\bar{\xi}_{r-k}\right)\leq 2(r-k).\end{array}\right\}.\end{aligned} \tag{65}$$

Here, $\bar{\mathbf{X}}\in\mathcal{R}(\mathbb{C}_{\leq r}^{a_1\times a_2})$ and $\text{rank}(\bar{\mathbf{X}})=2k$. The following process is used to establish (65). To demonstrate the "$\supseteq$" inclusion, let $\bar{\xi}$ be an element of the set on the right side of the equality, i.e., $\bar{\xi}=\bar{\xi}_k+\bar{\xi}_{r-k}$ with $\bar{\xi}_k\in T_{\bar{\mathbf{X}}}\mathcal{R}(\mathbb{C}_{k}^{a_1\times a_2})$, $\bar{\xi}_{r-k}\in(T_{\bar{\mathbf{X}}}\mathcal{R}(\mathbb{C}_{k}^{a_1\times a_2}))^{\perp}$, and $\text{rank}(\bar{\xi}_{r-k})\leq 2(r-k)$. Then, there exist a sequence $\{\bar{\mathbf{Y}}_n\}\subseteq\mathcal{R}(\mathbb{C}_{k}^{a_1\times a_2})$ and a sequence $\{a_n\}\subseteq\mathbb{R}^+$ such that $\bar{\mathbf{Y}}_n\to\bar{\mathbf{X}}$ and $a_n(\bar{\mathbf{Y}}_n-\bar{\mathbf{X}})=\bar{\xi}_k$. We can assume $a_n\to\infty$. Further, define a sequence $\bar{\mathbf{X}}_n\triangleq\bar{\mathbf{Y}}_n+a_n^{-1}\bar{\xi}_{r-k}$ in $\mathcal{R}(\mathbb{C}_{\leq r}^{a_1\times a_2})$. It's easy to show that $\{\bar{\mathbf{X}}_n\}$ convergences to $\bar{\mathbf{X}}$ and $a_n(\bar{\mathbf{X}}_n-\bar{\mathbf{X}})$ converges to $\bar{\xi}$. Hence, $\bar{\xi}\in T_{\bar{\mathbf{X}}}\mathcal{R}(\mathbb{C}_{\leq r}^{a_1\times a_2})$.

To show the "$\subseteq$" inclusion, let $\bar{\xi}$ be an element of the set on the left side of the equality. Based on (64), we know $\exists\bar{\mathbf{X}}_n\to\bar{\mathbf{X}}$ in $\mathcal{R}(\mathbb{C}_{\leq r}^{a_1\times a_2})$ and $\{a_n\}\subseteq\mathbb{R}^+$, s.t. $\bar{\xi}=\lim_{n\to\infty}a_n(\bar{\mathbf{X}}_n-\bar{\mathbf{X}})$. Furthermore, utilizing the orthogonal decomposition of $\mathcal{R}(\mathbb{C}_{k}^{a_1\times a_2})$ yields:

$$\begin{aligned}a_n\left(\bar{\mathbf{X}}_n-\bar{\mathbf{X}}\right)&=P_{T_{\bar{\mathbf{X}}}\mathcal{R}\left(\mathbb{C}_{k}^{a_1\times a_2}\right)}\left(a_n\left(\bar{\mathbf{X}}_n-\bar{\mathbf{X}}\right)\right)\\&\quad+P_{\left(T_{\bar{\mathbf{X}}}\mathcal{R}\left(\mathbb{C}_{k}^{a_1\times a_2}\right)\right)^{\perp}}\left(a_n\left(\bar{\mathbf{X}}_n-\bar{\mathbf{X}}\right)\right).\end{aligned} \tag{66}$$

Using this decomposition, we can demonstrate that both terms converge separately. Denote their limits by $\bar{\xi}_k$ and $\bar{\xi}_{r-k}$, it's obvious that $\bar{\xi}=\bar{\xi}_k+\bar{\xi}_{r-k}$ with $\bar{\xi}_k\in T_{\bar{\mathbf{X}}}\mathcal{R}(\mathbb{C}_{k}^{a_1\times a_2})$, $\bar{\xi}_{r-k}\in(T_{\bar{\mathbf{X}}}\mathcal{R}(\mathbb{C}_{k}^{a_1\times a_2}))^{\perp}$, and $\text{rank}(\bar{\xi})\leq 2r$, $\text{rank}(\bar{\xi}_k)=2k$. Based on the property of rank, i.e., $\text{rank}(\bar{\xi})\leq\text{rank}(\bar{\xi}_k)+\text{rank}(\bar{\xi}_{r-k})$, it can be derived that $\text{rank}(\bar{\xi}_{r-k})\leq 2(r-k)$.

Till now, we have proved (65). Finally, by utilizing the inverse mapping $\mathcal{R}^{-1}$, Proposition 3 can be demonstrated.

References

[1] Y. Cui, T. Lv, W. Ni, and A. Jamalipour, "Digital twin-aided learning for managing reconfigurable intelligent surface-assisted, uplink, user-centric cell-free systems," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 10, pp. 3175–3190, Oct. 2023.

[2] P. Wang, J. Fang, X. Yuan, Z. Chen, and H. Li, "Intelligent reflecting surface-assisted millimeter wave communications: Joint active and passive precoding design," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 14960–14973, Dec. 2020.

[3] Y. Han, W. Tang, S. Jin, C.-K. Wen, and X. Ma, "Large intelligent surface-assisted wireless communication exploiting statistical CSI," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 8238–8242, Aug. 2019.

[4] W. Mei, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless networks: From single-reflection to multi-reflection design and optimization," *Proc. IEEE*, vol. 110, no. 9, pp. 1380–1400, Sep. 2022.

[5] H. Zhou, X. Kang, Y.-C. Liang, S. Sun, and X. Shen, "Cooperative beamforming for reconfigurable intelligent surface-assisted symbiotic radios," *IEEE Trans. Veh. Technol.*, vol. 71, no. 11, pp. 11677–11692, Nov. 2022.

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

[7] D. Mishra and H. Johansson, "Low-complexity beamforming designs and channel estimation for passive-intelligent-surface-assisted MISO energy transfer," *IEEE Internet Things J.*, vol. 10, no. 9, pp. 8286–8304, May 2023.

[8] T. L. Jensen and E. De Carvalho, "An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2020, pp. 5000–5004.

[9] G. T. de Araújo and A. L. F. de Almeida, "PARAFAC-based channel estimation for intelligent reflective surface assisted MIMO system," in *Proc. IEEE 11th Sensor Array Multichannel Signal Process. Workshop*, 2020, pp. 1–5.

[10] Y. Yang, B. Zheng, S. Zhang, and R. Zhang, "Intelligent reflecting surface meets OFDM: Protocol design and rate maximization," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4522–4535, Jul. 2020.

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

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

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

[14] T. Lin, X. Yu, Y. Zhu, and R. Schober, "Channel estimation for IRS-assisted millimeter-wave MIMO systems: Sparsity-inspired approaches," *IEEE Trans. Commun.*, vol. 70, no. 6, pp. 4078–4092, Jun. 2022.

[15] T. Lin, X. Yu, Y. Zhu, and R. Schober, "Channel estimation for intelligent reflecting surface-assisted millimeter wave MIMO systems," in *Proc. IEEE Glob. Commun. Conf.*, 2020, pp. 1–6.

[16] J. Hu, H. Yin, and E. Björnson, "Mmwave MIMO communication with semi-passive RIS: A low-complexity channel estimation scheme," in *Proc. IEEE Glob. Commun. Conf.*, 2020, pp. 1–6.

[17] G. H. Golub and C. F. Van Loan *Matrix Computations*. Baltimore, MA, USA: JHU Press, 2013.

[18] J. Wang, Y.-C. Liang, Y. Pei, and X. Shen, "Reconfigurable intelligent surface as a micro base station: A novel paradigm for small cell networks," *IEEE Trans. Wireless Commun.*, vol. 22, no. 4, pp. 2338–2351, Apr. 2023.

[19] J. Tan and L. Dai, "Channel feedback in TDD massive MIMO systems with partial reciprocity," *IEEE Trans. Veh. Technol.*, vol. 70, no. 12, pp. 12960–12974, Dec. 2021.

[20] M. Vaezi, H. Inaltekin, W. Shin, H. V. Poor, and J. Zhang, "Social-aware user cooperation in full-duplex and half-duplex multi-antenna systems," *IEEE Trans. Commun.*, vol. 66, no. 8, pp. 3309–3321, Aug. 2018.

[21] Y. Song, C. Liu, W. Zhang, Y. Liu, H. Zhou, and X. Shen, "Two stage beamforming in massive MIMO: A combinatorial multi-armed bandit based approach," *IEEE Trans. Veh. Technol.*, vol. 72, no. 5, pp. 6794–6799, May 2023.

[22] E. Ahmed and A. M. Eltawil, "All-digital self-interference cancellation technique for full-duplex systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3519–3532, Jul. 2015.

[23] T. S. Rappaport et al., "Millimeter wave mobile communications for 5G cellular: It will work!," *IEEE Access*, vol. 1, pp. 335–349, 2013.

[24] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale ℓ_1-regularized least squares," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 606–617, Dec. 2007.

[25] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[26] T. M. Rassias, *Inner Product Spaces and Applications*, vol. 376. Boca Raton, FL, USA: CRC Press, 1997.

[27] M. C. Akkaya and D. Pinar, "Minimizers of sparsity regularized huber loss function," *J. Optim. Theory Appl.*, vol. 187, no. 1, pp. 205–233, 2020.

Authorized licensed use limited to: University of Waterloo. Downloaded on December 06,2024 at 18:52:20 UTC from IEEE Xplore. Restrictions apply.

[28] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, PA, USA: Princeton Univ. Press, 2008.
[29] H. Zamiri-Jafarian and S. Pasupathy, "Robust and improved channel estimation algorithm for MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 6, no. 6, pp. 2106–2113, Jun. 2007.
[30] 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, Dec. 2007.
[31] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, "Sparse reconstruction by separable approximation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2008, pp. 3373–3376.
[32] Y. Song et al., "Domain selective precoding in 3-D massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 5, pp. 1103–1118, Sep. 2019.
[33] L. Hormander, *An Introduction to Complex Analysis in Several Variables*. Amsterdam, The Netherlands: Elsevier, 1973.
[34] H. Sato and T. Iwai, "A complex singular value decomposition algorithm based on the Riemannian Newton method," in *Proc. IEEE 52nd Conf. Decis. Control*, 2013, pp. 2972–2978.

Zheng Huang (Graduate Student Member, IEEE) received the B.S. degree from the College of Networks and Telecommunications Engineering, Jinling Institute of Technology, Nanjing, China, in 2019. He is currently working toward the Ph.D. degree with the Nanjing University of Posts and Telecommunications, Nanjing. His main research interests include reconfigurable intelligent surface, channel estimation, and manifold optimization techniques.

Chen Liu (Member, IEEE) received the B.E. degree in electrical and information engineering from the Nanjing Institute of Technology, Southeast University, Nanjing, China, in 1985, the M.S. degree in circuits and systems from Anhui University, Hefei, China, in 1988, and the Ph.D. degree in signal and information processing from Southeast University, in 2005. In 1988, he joined Nanjing University of Posts and Telecommunications, Nanjing, China, where he has been a Professor since 2002. His research focuses on massive MIMO systems.

Yunchao Song received the B.E. degree in electronic science and technology and the Ph.D. degree in circuits and systems from the Nanjing University of Posts and Telecommunications (NUPT), Nanjing, China, in 2010 and 2016, respectively. Since 2017, he has been a faculty with the College of Electronic and Optical Engineering, NUPT, where he is currently an Associate Professor. From 2018 to 2019, he was a Visiting Scholar with the BBCR Lab, Department of Electronics and Communication Engineering, University of Waterloo, ON, Canada. His research focuses on massive MIMO systems.

Hong Wang received the B.S. degree in communication engineering from Jiangsu University, Zhenjiang, China, in 2011, and the Ph.D. degree in information and communication engineering from Nanjing University of Posts and Telecommunications (NUPT), Nanjing, China, in 2016. From 2014 to 2015, he was a Research Assistant with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong. From 2016 to 2018, he was a Senior Research Associate with the State Key Laboratory of Millimeter Waves and Department of Electronic Engineering, City University of Hong Kong. Since 2019, he has been an Associate Professor with the Department of Communication Engineering, NUPT. His research interests include broadband wireless communications, particularly in interference management in small cell networks, non-orthogonal multiple access, and intelligent reflecting surface.

Haibo Zhou (Senior Member, IEEE) received the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2014. He is currently a Full Professor with the School of Electronic Science and Engineering, Nanjing University, Nanjing, China. His research interests include resource management and protocol design in B5G/6G networks, vehicular ad-hoc networks, and space-air-ground integrated networks. He was the recipient of the 2019 IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award, (2023–2024) IEEE ComSoc Distinguished Lecturer, and (2023–2025) IEEE VTS Distinguished Lecturer.

Sherman Shen (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 1990. He is currently a Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His research interests include network resource management, wireless network security, Internet of Things, 5G and beyond, and vehicular networks. Dr. Shen was the recipient of the West Lake Friendship Award from Zhejiang Province in 2023, President's Excellence in Research from the University of Waterloo in 2022, Canadian Award for Telecommunications Research from the Canadian Society of Information Theory in 2021, R.A. Fessenden Award in 2019 from IEEE, Canada, Award of Merit from the Federation of Chinese Canadian Professionals (Ontario) in 2019, James Evans Avant Garde Award in 2018 from the IEEE Vehicular Technology Society, Joseph LoCicero Award in 2015 and Education Award in 2017 from the IEEE Communications Society (ComSoc), and Technical Recognition Award from Wireless Communications Technical Committee in 2019 and AHSN Technical Committee in 2013. He was also the recipient of the Excellent Graduate Supervision Award in 2006 from the University of Waterloo and Premier's Research Excellence Award in 2003 from the Province of Ontario, Canada. He is/was the General Chair of the 6G Global Conference'23 and ACM Mobihoc'15, Technical Program Committee Chair/Co-Chair of IEEE Globecom'24, 16, and 07, IEEE Infocom'14, and IEEE VTC'10 Fall, and Chair of the IEEE ComSoc Technical Committee on Wireless Communications. He is the President of the IEEE ComSoc. He was the Vice President for Technical and Educational Activities, Vice President for Publications, a Member-at-Large on the Board of Governors, the Chair of the Distinguished Lecturer Selection Committee, and a member of the IEEE Fellow Selection Committee of the ComSoc. He was the Editor-in-Chief of IEEE INTERNET OF THINGS JOURNAL, IEEE NETWORK, and *Peer-to-Peer Networking and Applications*. He is a registered Professional Engineer of Ontario, Canada, an Engineering Institute of Canada Fellow, a Canadian Academy of Engineering Fellow, Royal Society of Canada Fellow, Chinese Academy of Engineering Foreign Member, and Distinguished Lecturer of the IEEE Vehicular Technology Society and Communications Society.

Authorized licensed use limited to: University of Waterloo. Downloaded on December 06,2024 at 18:52:20 UTC from IEEE Xplore. Restrictions apply.