Cost-Efficient RIS-Aided Channel Estimation via Rank-One Matrix Factorization

Wei Zhang, *Member, IEEE* and Wee Peng Tay, *Senior Member, IEEE*

Abstract

A reconfigurable intelligent surface (RIS) consists of massive meta elements, which can improve the performance of future wireless communication systems. Existing RIS-aided channel estimation methods try to estimate the cascaded channel directly, incurring high computational and training overhead especially when the number of elements of RIS is extremely large. In this paper, we propose a cost-efficient channel estimation method via rank-one matrix factorization (MF). Specifically, if the RIS is employed near base station (BS), it is found that the RIS-aided channel can be factorized into a product of low-dimensional matrices. To estimate these factorized matrices, we propose alternating minimization and gradient descent approaches to obtain the near optimal solutions. Compared to directly estimating the cascaded channel, the proposed MF method reduces training overhead substantially. Finally, the numerical simulations show the effectiveness of the proposed MF method.

Index Terms

Channel estimation, matrix factorization, reconfigurable intelligent surface.

I. Introduction

Due to the increasing bandwidth requirement of modern wireless communication systems, many new technologies such as millimeter wave, hybrid precoding, and massive multi-input and multi-output (MIMO) have been introduced. Though these technologies can provide high spectral efficiency, the hardware cost of these technologies is also large [1]. Recently, the reconfigurable intelligent surface (RIS) is proposed as an aid to the wireless communications. An RIS directly reflects the received signal with many low-cost meta elements [2], through which the performance of existing systems is improved without high hardware cost. Another advantage of RIS is that it establishes a reflected transmission path for the communication if the direct transmission is blocked [2]. By carefully deploying the RIS near the boundary of a cell, RIS also helps to improve the performance of cell-edge users [3]. Overall, to achieve these benefits introduced by RIS, channel state information is key.

The authors are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore (e-mail: weizhang@ntu.edu.sg, wptay@ntu.edu.sg).

In RIS-aided communication systems, the signals are transmitted by the BS to RIS through the BS-RIS link, and reflected passively by RIS, then received by UE through the RIS-UE link. The channel estimation task is to obtain the estimate of the cascaded channel which consists of the BS-RIS link and RIS-UE link. The recent works focusing on estimation of the RIS-aided channel can be categorized into two types. The first type focuses on estimating the cascaded channel as a whole [4]–[6]. The work in [4] proposed a minimum mean square error channel estimation approach. To handle the high training overhead, the work in [5] divided the high-dimensional channel matrix into low-dimensional sub-groups. In [6], a compressed sensing based method was proposed to estimate the angle-of-arrival (AoA) and angle-of-departure (AoD) associated with the RIS-aided channel. Although the cascaded channel is sufficient to design transmit and passive beamforming [7], [8], the structure of the cascaded channel is not fully considered in [4]–[6]. The second type of existing works focus on estimating the channels of BS-RIS link and RIS-UE link [9]–[11]. A tensor decomposition method was proposed in [9], which leverages the parallel factor tensor modeling of the received signals. The work [10] proposed a two-timescale framework based on the characteristic that the BS-RIS link is quasi-static, while the RIS-UE link is mobile. The authors of [11] presented a two-stage method that utilizes the techniques of sparse matrix factorization and matrix completion. However, these methods [9]–[11] require high overhead [9] or additional system constraints, such as dual-link pilot transmission in [10] or sparsity of RIS phase shifts in [11].

In this work, we develop a cost-efficient method for RIS-aided channel estimation by factorizing the channel matrix as a product of low-dimensional matrices, which helps to reduce the training overhead without any additional system constraints. Specifically, we perform rank-one matrix factorization (MF) to formulate the RIS-aided channel estimation as a phase retrieval problem, which is solved by gradient descent or alternating minimization methods. We then extend the proposed MF method into multi-user scenarios. The simulations illustrate that the proposed MF method achieves more accurate estimation than existing works and has low training overhead.

Notations: A bold lower case letter $\mathbf{a}$ is a vector and a bold capital letter $\mathbf{A}$ is a matrix. $\mathbf{A}^T$, $\mathbf{A}^*$, $\mathbf{A}^H$, $\mathbf{A}^{-1}$, $\mathrm{tr}(\mathbf{A})$, $\|\mathbf{A}\|_F$ and $\|\mathbf{a}\|_2$ are, respectively, the transpose, conjugate, Hermitian, inverse, trace, Frobenius norm of $\mathbf{A}$, and 2-norm of $\mathbf{a}$. $[\mathbf{A}]_{:,i}$ and $[\mathbf{A}]_{i,:}$ are, respectively, the ith column and ith row of $\mathbf{A}$. $\mathrm{vec}(\mathbf{A})$ stacks the columns of $\mathbf{A}$ and forms a long column vector. $\mathrm{diag}(\mathbf{a})$ returns a square diagonal matrix with the vector $\mathbf{a}$ on the main diagonal. $\mathrm{Re}\{z\}$ and $\mathrm{Im}\{z\}$ are the real and imaginary parts of the complex number z, respectively. $\mathbf{A} \otimes \mathbf{B}$ and $\mathbf{A} \circ \mathbf{B}$ denote the Kronecker and Hadamard product of $\mathbf{A}$ and $\mathbf{B}$, respectively.

II. CHANNEL MODEL

In this section, we introduce the signal and channel model of an RIS-aided communication system.

Suppose the BS has N antennas, the RIS has M elements, and the UE has one antenna. The BS-RIS channel is $\mathbf{G} \in \mathbb{C}^{M \times N}$, and RIS-UE channel is $\mathbf{h}_r \in \mathbb{C}^{M \times 1}$, and the BS-UE channel is $\mathbf{h}_d \in \mathbb{C}^{N \times 1}$. For the RIS, $\boldsymbol{\theta} \in \mathbb{C}^{M \times 1}$ denotes the phase shifts, i.e., $\boldsymbol{\theta} = [\beta_1 e^{j\theta_1}, \ldots, \beta_m e^{j\theta_m}, \ldots, \beta_M e^{j\theta_M}]^T$, where $\theta_m \in [0, 2\pi)$ and $\beta_m \in [0, 1]$ are the phase-shift value and amplitude reflection coefficient of the mth element. Here, we let $\beta_m = 1, \forall m$ to maximize the signal reflection [4], [8].

In this work, we assume there is line-of-sight (LOS) path[1] between BS and RIS. Then, the channel model $\mathbf{G}$ is given by

$$\mathbf{G} = \beta_{BR} \mathbf{a}_R(\phi) \mathbf{a}_B^H(\psi), \tag{1}$$

where β_{BR} is the complex path gain, $\phi \in [0, 1)$ and $\psi \in [0, 1)$ are the effective AoA and AoD of RIS and BS, respectively. In this work, it is assumed BS and RIS have uniform linear antenna array (ULA) [6], [11]. Thus,

$$\mathbf{a}_B(\psi) = \frac{1}{\sqrt{N}} [1, e^{-j2\pi\psi}, \ldots, e^{-j2\pi\psi(N-1)}]^T \in \mathbb{C}^{N \times 1},$$

$$\mathbf{a}_R(\phi) = \frac{1}{\sqrt{M}} [1, e^{-j2\pi\phi}, \ldots, e^{-j2\pi\phi(M-1)}]^T \in \mathbb{C}^{M \times 1},$$

are the array response vectors of BS and RIS. The proposed method in this paper can be extended to the planar antenna array at RIS by expressing $\mathbf{a}_R(\cdot)$ above in a 2D form [1].

For the channel $\mathbf{h}_r$ between RIS and UE, we assume that there is no LOS path and it is the Rayleigh fading channel [12], i.e., $\mathbf{h}_r = \beta_{RU} \mathbf{z}$, where β_{RU} is the complex path gain, and $\mathbf{z} \in \mathbb{C}^{M \times 1}$ is a complex Gaussian random variable $\mathcal{CN}(0, 1)$.

III. RIS-AIDED CHANNEL ESTIMATION VIA MF

In this section, we discuss the down-link and single-user RIS-aided channel estimation by using the MF method.

A. Problem Formulation

In the down-link transmission, the received signal at UE is

$$r = (\mathbf{h}_d^H + \mathbf{h}_r^H \boldsymbol{\Theta} \mathbf{G})\mathbf{x} + n, \tag{2}$$

where $\boldsymbol{\Theta} = \mathrm{diag}(\boldsymbol{\theta}) \in \mathbb{C}^{M \times M}$, $\mathbf{x} \in \mathbb{C}^{N \times 1}$ is the transmitted signal, and $n \in \mathbb{C}$ is noise distributed as $\mathcal{CN}(0, \sigma^2)$.

We assume that the channel of direct path $\mathbf{h}_d$ is accurately estimated, which can be done by using traditional channel estimation methods before enabling the RIS. Here, we suppose there are $K \geq M$ transmit pilots for channel estimation. Then, the received signal at the kth training transmission is

$$r_k = \mathbf{h}_r^H \boldsymbol{\Theta}_k \mathbf{G} \mathbf{x}_k + n_k = \boldsymbol{\theta}_k^T \mathrm{diag}(\mathbf{h}_r^H) \mathbf{G} \mathbf{x}_k + n_k, \tag{3}$$

[1]The LOS path between RIS and BS can be achieved by deploying RIS near BS. We extend it to other channel models in Section III-E.

Algorithm 1 Down-link and single-user RIS-aided channel estimation via matrix factorization

1: Input: BS antennas N, RIS elements M, transmitted signals $\{\mathbf{x}_k\}_{k=1}^K$, received signals $\{r_k\}_{k=1}^K$, step size η.

2: Initialization: $\bar{\mathbf{a}}^{(0)}$ and $\psi^{(0)}$ are from (8) and (6).

3: **for** $p = 0, 1, 2, \ldots$ **do**

4:     Update $\bar{\mathbf{a}}^{(p)}$, $\psi^{(p)}$ to $\bar{\mathbf{a}}^{(p+1)}$, $\psi^{(p+1)}$ according to gradient method in (11), (12) and (13) or the alternating minimization method in (10) and (9).

5:     If the iteration statures or the maximum iteration is attained, it terminates the iteration.

6: **end for**

7: Calculate the estimation of cascaded channel: $\widehat{\mathbf{H}}_e = \bar{\mathbf{a}}^{(p+1)} \mathbf{a}_B^H(\psi^{(p+1)})$.

8: Output: $\widehat{\mathbf{H}}_e$.

---

where $\boldsymbol{\theta}_k$ is the kth phase shifts of RIS, and $\mathbf{x}_k$ is the kth transmitted signal with $\|\mathbf{x}_k\|_2 = 1$, thus SNR $= 1/\sigma^2$. The effective cascaded channel is expressed as $\mathbf{H}_e = \mathrm{diag}(\mathbf{h}_r^H)\mathbf{G} \in \mathbb{C}^{M \times N}$. Plugging the model of $\mathbf{G}$ in (1), we express (3) as

$$r_k = \boldsymbol{\theta}_k^T \bar{\mathbf{a}} \mathbf{a}_B^H(\psi)\mathbf{x}_k + n_k, \tag{4}$$

where $\bar{\mathbf{a}} = \mathrm{diag}(\mathbf{h}_r^H)\beta_{BR}\mathbf{a}_R(\phi) \in \mathbb{C}^{M \times 1}$, and $\mathbf{H}_e = \bar{\mathbf{a}}\mathbf{a}_B^H(\psi)$. Then, the task in this work is to estimate ψ and $\bar{\mathbf{a}}$ from $\{r_k\}_{k=1}^K$. From the model expressed in (4), the problem formulation of maximal likelihood estimation is,

$$\min_{\bar{\mathbf{a}}, \psi} \sum_{k=1}^K \left| \boldsymbol{\theta}_k^T \bar{\mathbf{a}} \mathbf{a}_B^H(\psi)\mathbf{x}_k - r_k \right|^2. \tag{5}$$

For convenience, the objective function in (5) is defined as $J(\bar{\mathbf{a}}, \psi)$. If we ignore the structure of $\mathbf{a}_B(\psi)$, the problem in (5) is a conventional phase retrieval problem, which can be solved by low-rank matrix recovery techniques [13]–[15]. However, these approaches do not take the structure of $\mathbf{a}_B(\psi)$ into account. In this work, we aim to estimate $\bar{\mathbf{a}}$ and ψ instead. As (5) is a non-convex problem, we consider the use of alternating minimization and gradient descent approaches to find its solution. The details of the proposed RIS-aided channel estimation are shown in Algorithm 1.

B. Initialization

Solving the non-convex optimization problem (5) using either alternating minimization or gradient descent approach requires good initialization. In the following, we discuss how to initialize these variables.

According to (4), we define $\mathbf{S} = \sqrt{N}/K \sum_{k=1}^K r_k \boldsymbol{\theta}_k^* \mathbf{x}_k^H \in \mathbb{C}^{M \times N}$. Then, the initialization of ψ is given by

$$\psi^{(0)} = \underset{\psi}{\arg\max} \|\mathbf{S}\mathbf{a}_B(\psi)\|_2^2, \tag{6}$$

which means that $\mathbf{a}_B(\psi^{(0)})$ is most correlated with $\mathbf{S}$. The problem in (6) is one-dimensional, which can be solved by using coarse grid search and then refining the result gradually.

To initialize $\bar{\mathbf{a}}$, we solve (5) with the initialized $\psi^{(0)}$,

$$\bar{\mathbf{a}}^{(0)} = \arg\min_{\bar{\mathbf{a}}} \sum_{k=1}^{K} \left| \boldsymbol{\theta}_k^T \bar{\mathbf{a}} \mathbf{a}_B^H(\psi^{(0)}) \mathbf{x}_k - r_k \right|^2, \tag{7}$$

which is a common least squares (LS) problem. Define $\mathbf{r} = [r_1, \ldots, r_K]^T \in \mathbb{C}^{K \times 1}$ and $\mathbf{B}^{(0)} \in \mathbb{C}^{K \times M}$ with $[\mathbf{B}^{(0)}]_{k,:} = \mathbf{a}_B^H(\psi^{(0)}) \mathbf{x}_k \boldsymbol{\theta}_k^T$, the problem in (7) is rewritten as

$$\bar{\mathbf{a}}^{(0)} = \arg\min_{\bar{\mathbf{a}}} \|\mathbf{B}^{(0)} \bar{\mathbf{a}} - \mathbf{r}\|_2^2, \tag{8}$$

where the solution is $\bar{\mathbf{a}}^{(0)} = ((\mathbf{B}^{(0)})^H \mathbf{B}^{(0)})^{-1} (\mathbf{B}^{(0)})^H \mathbf{r}$. Because we assume that $K \geq M$, $(\mathbf{B}^{(0)})^H \mathbf{B}^{(0)}$ is invertible.

C. Alternating Minimization

After initializing the ψ and $\bar{\mathbf{a}}$ with $\psi^{(0)}$ and $\bar{\mathbf{a}}^{(0)}$, respectively, we iteratively refine these variables as follows:

- Fix $\bar{\mathbf{a}}$, solve the optimization problem (5) over ψ.
- Fix ψ, solve the optimization problem (5) over $\bar{\mathbf{a}}$.

We denote p as the index of iteration. For the optimization of ψ with fixed $\bar{\mathbf{a}}^{(p)}$, the sub-problem is given by

$$\psi^{(p+1)} = \arg\min_{\psi} \sum_{k=1}^{K} \left| \boldsymbol{\theta}_k^T \bar{\mathbf{a}}^{(p)} \mathbf{a}_B^H(\psi) \mathbf{x}_k - r_k \right|^2.$$

By denoting $\mathbf{A}^{(p)} \in \mathbb{C}^{N \times K}$ with $[\mathbf{A}^{(p)}]_{:,k} = \boldsymbol{\theta}_k^T \bar{\mathbf{a}}^{(p)} \mathbf{x}_k$, the problem above can be simplified as

$$\psi^{(p+1)} = \arg\min_{\psi} \|\mathbf{a}_B^H(\psi) \mathbf{A}^{(p)} - \mathbf{r}^T\|_2^2, \tag{9}$$

which can be solved by using similar technique as (6). For optimizing $\bar{\mathbf{a}}$ with fixed $\psi^{(p+1)}$, the sub-problem is $\bar{\mathbf{a}}^{(p+1)} = \arg\min_{\bar{\mathbf{a}}} \sum_{k=1}^{K} |\boldsymbol{\theta}_k^T \bar{\mathbf{a}} \mathbf{a}_B^H(\psi^{(p+1)}) \mathbf{x}_k - r_k|^2$, whose solution is

$$\bar{\mathbf{a}}^{(p+1)} = ((\mathbf{B}^{(p+1)})^H \mathbf{B}^{(p+1)})^{-1} (\mathbf{B}^{(p+1)})^H \mathbf{r}, \tag{10}$$

with $[\mathbf{B}^{(p+1)}]_{k,:} = \mathbf{a}_B^H(\psi^{(p+1)}) \mathbf{x}_k \boldsymbol{\theta}_k^T$.

D. Gradient Descent

Apart from the alternating minimization, we can also employ a gradient descent approach to obtain the local optimal solution of (5). The gradient $J(\bar{\mathbf{a}}, \psi)$ in (5) with respect to the real and imaginary parts of $\bar{\mathbf{a}}$ are given by

$$\frac{dJ}{d\operatorname{Re}\{\bar{\mathbf{a}}\}} = \operatorname{Re}\left\{ \sum_{k=1}^{K} 2\mathbf{x}_k^H \mathbf{a}_B(\psi) \boldsymbol{\theta}_k^* (\mathbf{a}_B^H(\psi) \mathbf{x}_k \boldsymbol{\theta}_k^T \bar{\mathbf{a}} - r_k) \right\},$$

$$\frac{dJ}{d\operatorname{Im}\{\bar{\mathbf{a}}\}} = \operatorname{Im}\left\{ \sum_{k=1}^{K} 2\mathbf{x}_k^H \mathbf{a}_B(\psi) \boldsymbol{\theta}_k^* (\mathbf{a}_B^H(\psi) \mathbf{x}_k \boldsymbol{\theta}_k^T \bar{\mathbf{a}} - r_k) \right\},$$

respectively. From the chain rule, the gradient of $J(\bar{\mathbf{a}}, \psi)$ with respect to ψ is given by

$$\frac{dJ}{d\psi} = \left(\frac{dJ}{d\operatorname{Re}\{\mathbf{a}_B\}}\right)^T \frac{d\operatorname{Re}\{\mathbf{a}_B\}}{d\psi} + \left(\frac{dJ}{d\operatorname{Im}\{\mathbf{a}_B\}}\right)^T \frac{d\operatorname{Im}\{\mathbf{a}_B\}}{d\psi},$$

where

$$\frac{dJ}{d\operatorname{Re}\{\mathbf{a}_B\}} = \operatorname{Re}\left\{\sum_{k=1}^{K} 2\boldsymbol{\theta}_k^T \bar{\mathbf{a}} \mathbf{x}_k (\bar{\mathbf{a}}^H \boldsymbol{\theta}_k^* \mathbf{x}_k^H \mathbf{a}_B - r_k^H)\right\},$$

$$\frac{dJ}{d\operatorname{Im}\{\mathbf{a}_B\}} = \operatorname{Im}\left\{\sum_{k=1}^{K} 2\boldsymbol{\theta}_k^T \bar{\mathbf{a}} \mathbf{x}_k (\bar{\mathbf{a}}^H \boldsymbol{\theta}_k^* \mathbf{x}_k^H \mathbf{a}_B - r_k^H)\right\},$$

$$\frac{d\operatorname{Re}\{\mathbf{a}_B\}}{d\psi} = \frac{-1}{\sqrt{N}}\sin(\psi\mathbf{z}) \circ \mathbf{z}, \quad \frac{d\operatorname{Im}\{\mathbf{a}_B\}}{d\psi} = \frac{-1}{\sqrt{N}}\cos(\psi\mathbf{z}) \circ \mathbf{z},$$

with $\mathbf{z} = 2\pi[0, 1, \ldots, N-1]^T$. Therefore, the updating rules of $\bar{\mathbf{a}}$ and ψ are explicitly,

$$\operatorname{Re}\{\bar{\mathbf{a}}^{(p+1)}\} = \operatorname{Re}\{\bar{\mathbf{a}}^{(p)}\} - \eta \frac{dJ}{d\operatorname{Re}\{\bar{\mathbf{a}}\}}\bigg|_{\bar{\mathbf{a}}=\bar{\mathbf{a}}^{(p)}}, \tag{11}$$

$$\operatorname{Im}\{\bar{\mathbf{a}}^{(p+1)}\} = \operatorname{Im}\{\bar{\mathbf{a}}^{(p)}\} - \eta \frac{dJ}{d\operatorname{Im}\{\bar{\mathbf{a}}\}}\bigg|_{\bar{\mathbf{a}}=\bar{\mathbf{a}}^{(p)}}, \tag{12}$$

$$\psi^{(p+1)} = \psi^{(p)} - \eta \frac{dJ}{d\psi}\bigg|_{\psi=\psi^{(p)}}, \tag{13}$$

where η is the step size of gradient descent.

E. Discussions of the Proposed MF Method

In this subsection, we talk about the extension of proposed MF method when the channel model has the different structures as provided Section II.

Remark 1: When there are more than one paths in BS-RIS link, i.e., $\mathbf{G} = \sum_{l=1}^{L} \beta_{BR}^{(l)} \mathbf{a}_R(\phi^{(l)}) \mathbf{a}_B^H(\psi^{(l)})$, where L is the number of paths, the cascaded channel can be expressed as

$$\mathbf{H}_e = \sum_{l=1}^{L} \operatorname{diag}(\mathbf{h}_r^H) \beta_{BR}^{(l)} \mathbf{a}_R(\phi^{(l)}) \mathbf{a}_B^H(\psi^{(l)}) = \sum_{l=1}^{L} \mathbf{H}_e^{(l)}.$$

Each $\mathbf{H}_e^{(l)}$ can be estimated by the proposed MF method after subtracting the contributions of the previously estimated paths.

Remark 2: When the RIS is located near the UE resulting in a LOS path between RIS and UE and no LOS path between BS and RIS. The channel of RIS-UE link is $\mathbf{h}_r = \beta_{RU} \mathbf{a}_R(\phi)$, where β_{RU} is the complex path gain and ϕ is the AoD of RIS. The channel $\mathbf{G} \in \mathbb{C}^{M \times N}$ of BS-RIS link is Rayleigh fading. Thus, the received signal at kth transmission in (3) is

$$r_k = (\beta_{RU} \mathbf{a}_R(\phi))^H \boldsymbol{\Theta}_k \mathbf{G} \mathbf{x}_k + n_k$$

$$= \mathbf{a}_R(\phi)^H \boldsymbol{\Theta}_k \bar{\mathbf{G}} \mathbf{x}_k + n_k,$$

where $\bar{\mathbf{G}} = \beta_{RU} \mathbf{G}$. In this scenario, the channel problem is to estimate ϕ and $\bar{\mathbf{G}}$. The proposed MF methods can also be applicable. It is worth noting that the number of minimal required training pilots is MN because $\bar{\mathbf{G}}$ is full rank.

Algorithm 2 Up-link and multi-user RIS-aided channel estimation via matrix factorization

---

1: Input: BS antennas N, RIS elements M, number of users Q, transmitted signals $x_{q,k,t}, \forall q \leq Q, k \leq K, t \leq T$, received signals $\{\mathbf{R}_k\}_{k=1}^K$, phase shift $\{\boldsymbol{\theta}_k\}_{k=1}^K$.

2: Obtain the signal of qth UE $\mathbf{S}_q$ through (16).

3: Estimate $\widehat{\psi}$ of the BS-RIS link through (19).

4: **for** $q = 1, 2, \ldots, Q$ **do**

5:     Estimate $\{\hat{\bar{\mathbf{a}}}_q\}_{q=1}^Q$ of the RIS-UE link through (21).

6:     Calculate the estimation of cascaded channel of qth UE: $\widehat{\mathbf{H}}_q = \mathbf{a}_B(\widehat{\psi})\hat{\bar{\mathbf{a}}}_q^H$.

7: **end for**

8: Output: $\{\widehat{\mathbf{H}}_q\}_{q=1}^Q$.

IV. EXTENSION TO MULTI-USER SCENARIOS

In the down-link and multi-user RIS-aided systems, each user can estimate the channel separately, which is a trivial extension of the method discussed in Section III. Therefore, in this section, we mainly focus on the up-link and multi-user RIS-aided channel estimation by using the MF method.

A. Problem Formulation

Suppose there are Q users, the channel between BS and RIS is $\mathbf{G} \in \mathbb{C}^{N \times M}$, the channel between the RIS and UE q is $\mathbf{h}_q \in \mathbb{C}^{M \times 1}$, which are all defined in a similar way as down-link scenario in Section III. The received signal at the BS is the summation of Q UEs, i.e., $\mathbf{r} = \sum_{q=1}^Q \mathbf{G} \operatorname{diag}(\boldsymbol{\theta})\mathbf{h}_q x_q + \mathbf{n} \in \mathbb{C}^{N \times 1}$, where $|x_q| = 1$ denotes the transmitted signal of qth UE and $\mathbf{n} \in \mathbb{C}^{N \times 1}$ is noise distributed as $\mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_N)$. For the channel estimation framework, it is assumed that there are K transmitted blocks. In each block, the RIS fixes the phase shifts, and each UE sends T symbols. Precisely, in the kth block, the received signal at BS is $\mathbf{R}_k \in \mathbb{C}^{N \times T}$, given by

$$\mathbf{R}_k = \sum_{q=1}^Q \mathbf{G} \operatorname{diag}(\boldsymbol{\theta}_k)\mathbf{h}_q \mathbf{x}_{q,k}^T + \mathbf{N}_k, \tag{14}$$

where $\mathbf{x}_{q,k} = [x_{q,k,1}, \ldots, x_{q,k,T}]^T \in \mathbb{C}^{T \times 1}$ is the transmitted signal of qth UE in kth block, $\boldsymbol{\theta}_k \in \mathbb{C}^{M \times 1}$ denotes the kth RIS phase shifts, and $\mathbf{N}_k = [\mathbf{n}_{k,1}, \ldots, \mathbf{n}_{k,T}] \in \mathbb{C}^{N \times T}$ denotes the noise in kth block.

We assume the signals transmitted by different UEs are orthogonal, i.e., $\mathbf{x}_{q,k}^H \mathbf{x}_{p,k} = 0, \forall q \neq p$ and $\mathbf{x}_{q,k}^H \mathbf{x}_{q,k} = T, \forall q$, which can be achieved if $T \geq Q$. Thus, we right multiply $\mathbf{x}_{q,k}^*$ for (14) to obtain the signal associated with qth UE as

$$\mathbf{R}_k \mathbf{x}_{q,k}^* = T\mathbf{G} \operatorname{diag}(\boldsymbol{\theta}_k)\mathbf{h}_q + \mathbf{N}_k \mathbf{x}_{q,k}^*. \tag{15}$$

By defining $\mathbf{s}_{q,k} = (1/T)\mathbf{R}_k \mathbf{x}_{q,k}^* \in \mathbb{C}^{N \times 1}$ and $\mathbf{n}_{q,k} = (1/T)\mathbf{N}_k \mathbf{x}_{q,k}^* \in \mathbb{C}^{N \times 1}$, we rewrite the expression in (15) as

$$\mathbf{s}_{q,k} = \mathbf{G} \operatorname{diag}(\boldsymbol{\theta}_k)\mathbf{h}_q + \mathbf{n}_{q,k} = \mathbf{G} \operatorname{diag}(\mathbf{h}_q)\boldsymbol{\theta}_k + \mathbf{n}_{q,k},$$

where the entries in $\mathbf{n}_{q,k}$ are i.i.d. with distribution of $\mathcal{CN}(0, \sigma^2/T)$. We then collect K blocks $\{\mathbf{s}_{q,k}\}_{k=1}^K$

of the qth UE as $\mathbf{S}_q = [\mathbf{s}_{q,1}, \ldots, \mathbf{s}_{q,K}] \in \mathbb{C}^{N \times K}$ in the following,

$$
\begin{aligned}
\mathbf{S}_q &= \mathbf{G}\,\mathrm{diag}(\mathbf{h}_q)[\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K] + [\mathbf{n}_{q,1}, \ldots, \mathbf{n}_{q,K}] \\
&= \mathbf{G}\,\mathrm{diag}(\mathbf{h}_q)\bar{\boldsymbol{\Theta}} + \mathbf{N}_q,
\end{aligned}
\tag{16}
$$

where $\bar{\boldsymbol{\Theta}} = [\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K] \in \mathbb{C}^{M \times K}$, and $\mathbf{N}_q = [\mathbf{n}_{q,1}, \ldots, \mathbf{n}_{q,K}] \in \mathbb{C}^{M \times K}$. Therefore, from (16), the cascaded channel for qth UE is $\mathbf{H}_q = \mathbf{G}\,\mathrm{diag}(\mathbf{h}_q) \in \mathbb{C}^{N \times M}$. Since $\mathbf{G} = \beta_{BR}\mathbf{a}_B(\psi)\mathbf{a}_R^H(\phi)$ in (1), we express $\mathbf{S}_q$ in (16) as

$$
\mathbf{S}_q = \mathbf{a}_B(\psi)\bar{\mathbf{a}}_q^H \bar{\boldsymbol{\Theta}} + \mathbf{N}_q,
\tag{17}
$$

where $\bar{\mathbf{a}}_q = (\beta_{BR}\mathbf{a}_R^H(\phi)\,\mathrm{diag}(\mathbf{h}_q))^H \in \mathbb{C}^{M \times 1}$. Then, the cascaded channel for qth UE is given by $\mathbf{H}_q = \mathbf{a}_B(\psi)\bar{\mathbf{a}}_q^H$. Therefore, the channel estimation task is to estimate ψ and $\{\bar{\mathbf{a}}_q\}_{q=1}^Q$. Because of the Guassian distribution of $\mathbf{N}_q$, we consider the following optimization problem:

$$
\min_{\psi, \{\bar{\mathbf{a}}_q\}_{q=1}^Q} \sum_{q=1}^Q \|\mathbf{S}_q - \mathbf{a}_B(\psi)\bar{\mathbf{a}}_q^H \bar{\boldsymbol{\Theta}}\|_F^2.
\tag{18}
$$

In the following, we complete the estimation of ψ and $\{\bar{\mathbf{a}}_q\}_{q=1}^Q$ from (18) in two steps. In the first step, we estimate the channel associated with BS-RIS link, i.e., ψ, which is shared among multiple users. In the second step, we estimate channel associated with the RIS-UE link, i.e., $\{\bar{\mathbf{a}}_q\}_{q=1}^Q$.

B. Multi-User Channel Estimation

1) Estimate ψ: We collect the observations of Q UEs as $\mathbf{S} = [\mathbf{S}_1, \ldots, \mathbf{S}_Q] \in \mathbb{C}^{N \times KQ}$. It is from (17) that $\mathbf{S}$ is a rank-one matrix, whose column subspace is spanned by $\mathbf{a}_B(\psi)$. Thus, the estimation of ψ from (18) is formulated as follows,

$$
\widehat{\psi} = \arg\max_{\psi} \|\mathbf{a}_B(\psi)^H \mathbf{S}\|_2^2,
\tag{19}
$$

which can be solved by using similar technique as (6).

2) Estimate $\{\bar{\mathbf{a}}_q\}_{q=1}^Q$: With the estimation of $\mathbf{a}_B(\widehat{\psi})$, solving the problem in (18) with respect to $\{\bar{\mathbf{a}}_q\}_{q=1}^Q$ can then be separated. In particular, for the qth UE, we now need to solve the following problem,

$$
\hat{\bar{\mathbf{a}}}_q = \arg\min_{\bar{\mathbf{a}}_q} \|\mathbf{S}_q - \mathbf{a}_B(\widehat{\psi})\bar{\mathbf{a}}_q^H \bar{\boldsymbol{\Theta}}\|_F^2.
\tag{20}
$$

By vectorizing the notation as $\mathrm{vec}(\mathbf{a}_B(\widehat{\psi})\bar{\mathbf{a}}_q^H \bar{\boldsymbol{\Theta}}) = \bar{\boldsymbol{\Theta}}^T \otimes \mathbf{a}_B(\widehat{\psi})\,\mathrm{vec}(\bar{\mathbf{a}}_q^H) = \bar{\boldsymbol{\Theta}}^T \otimes \mathbf{a}_B(\widehat{\psi})\bar{\mathbf{a}}_q^*$, the problem in (20) can be rewritten as $\hat{\bar{\mathbf{a}}}_q = \arg\min_{\bar{\mathbf{a}}_q} \|\mathrm{vec}(\mathbf{S}_q) - \bar{\boldsymbol{\Theta}}^T \otimes \mathbf{a}_B(\widehat{\psi})\bar{\mathbf{a}}_q^*\|_2^2$. Since this is a LS problem, and the solution is given by

$$
\hat{\bar{\mathbf{a}}}_q = \left((\mathbf{V}(\widehat{\psi})^H \mathbf{V}(\widehat{\psi}))^{-1}\mathbf{V}(\widehat{\psi})^H \,\mathrm{vec}(\mathbf{S}_q)\right)^*,
\tag{21}
$$

where $\mathbf{V}(\widehat{\psi}) = \bar{\boldsymbol{\Theta}}^T \otimes \mathbf{a}_B(\widehat{\psi}) \in \mathbb{C}^{NK \times M}$. We can check that $\|\mathbf{V}(\widehat{\psi})\|_F^2 = MK$. Given $\widehat{\psi}$ in (19) and $\hat{\bar{\mathbf{a}}}_q$ in (21), the estimate of cascaded channel of qth UE is $\widehat{\mathbf{H}}_q = \mathbf{a}_B(\widehat{\psi})\hat{\bar{\mathbf{a}}}_q^H$. The details of the multi-user RIS channel estimation are in Algorithm 2.

TABLE I: Training Overhead of Channel Estimation Methods

Channel Estimation Methods	Minimal Training Pilots
Proposed MF-GD	M
Proposed MF-AM	M
LS [4]	MN
LR [14]	$M + N$
KBF [9]	MN

3) Design of Phase Shifts of RIS: By denoting $\mathbf{V}(\widehat{\psi})^{\dagger} = (\mathbf{V}(\widehat{\psi})^{H}\mathbf{V}(\widehat{\psi}))^{-1}\mathbf{V}(\widehat{\psi})^{H} \in \mathbb{C}^{M \times NK}$, we can calculate the mean-square error (MSE) of $\hat{\bar{\mathbf{a}}}_q$ as follows,

$$\mathbb{E}\|\mathbf{V}(\widehat{\psi})^{\dagger}\text{vec}(\mathbf{S}_q) - \bar{\mathbf{a}}_q^*\|_2^2 = \mathbb{E}\|\mathbf{V}(\widehat{\psi})^{\dagger}(\mathbf{V}(\psi)\bar{\mathbf{a}}_q^* + \text{vec}(\mathbf{N}_q)) - \bar{\mathbf{a}}_q^*\|_2^2,$$

where the equality holds from $\mathbf{S}_q = \mathbf{V}(\psi)\bar{\mathbf{a}}_q^* + \text{vec}(\mathbf{N}_q)$. Here, we assume that the estimation of ψ is accurate, so we have $\mathbf{V}(\widehat{\psi})^{\dagger}\mathbf{V}(\psi) \approx \mathbf{I}_M$. Then, the MSE can be approximated as

$$\text{MSE}(\bar{\mathbf{a}}_q) \approx \mathbb{E}\|\mathbf{V}(\widehat{\psi})^{\dagger}\text{vec}(\mathbf{N}_q)\|_2^2 = \frac{\sigma^2}{T}\text{tr}((\mathbf{V}(\widehat{\psi})^{H}\mathbf{V}(\widehat{\psi}))^{-1}). \tag{22}$$

Therefore, to minimize the MSE with respect to $\bar{\mathbf{a}}_q$, we need to solve the following problem,

$$\min_{\bar{\boldsymbol{\Theta}}} \text{tr}((\mathbf{V}(\widehat{\psi})^{H}\mathbf{V}(\widehat{\psi}))^{-1}), \text{ subject to } \|\mathbf{V}(\widehat{\psi})\|_F^2 = MK. \tag{23}$$

The following lemma states the condition for $\bar{\boldsymbol{\Theta}}$, which achieves the minimal MSE of $\bar{\mathbf{a}}_q$ in (22).

Lemma 1: When the phase shift matrix $\bar{\boldsymbol{\Theta}}$ in (16) satisfies $\bar{\boldsymbol{\Theta}}\bar{\boldsymbol{\Theta}}^{H} = K\mathbf{I}_M$, the minimal MSE of $\bar{\mathbf{a}}_q, q = 1, \ldots, Q$ in (22) can be achieved, whose value is $\text{MSE}(\bar{\mathbf{a}}_q) = \sigma^2 M/(KT)$.

Proof: Suppose the eigenvalues of $\mathbf{V}(\widehat{\psi})^{H}\mathbf{V}(\widehat{\psi})$ are given by $\lambda_1, \ldots, \lambda_M$. Because we have $\|\mathbf{V}(\widehat{\psi})\|_F^2 = MK$, then $\lambda_1 +, \ldots, + \lambda_M = MK$. Thus, the problem in (23) becomes

$$\min \sum_{i=1}^{M} \frac{1}{\lambda_i}, \text{ subject to } \sum_{i=1}^{M} \lambda_i = MK, \ \lambda_i \geq 0.$$

The above problem is minimized when $\lambda_i = K, \forall i$, and the minimum is M/K. The optimality condition means that $\mathbf{V}(\widehat{\psi})^{H}\mathbf{V}(\widehat{\psi}) = K\mathbf{I}_M$, precisely,

$$(\bar{\boldsymbol{\Theta}}^{T} \otimes \mathbf{a}_B(\widehat{\psi}))^{H}\bar{\boldsymbol{\Theta}}^{T} \otimes \mathbf{a}_B(\widehat{\psi}) = K\mathbf{I}_M. \tag{24}$$

Using the fact $\mathbf{a}_B(\widehat{\psi})^{H}\mathbf{a}_B(\widehat{\psi}) = 1$, we can simplify the optimality condition in (24) as $\bar{\boldsymbol{\Theta}}\bar{\boldsymbol{\Theta}}^{H} = K\mathbf{I}_M$. Substituting the design of $\bar{\boldsymbol{\Theta}}$ into (22), the minimized MSE is given by $\sigma^2 M/(KT)$. This concludes the proof. ∎
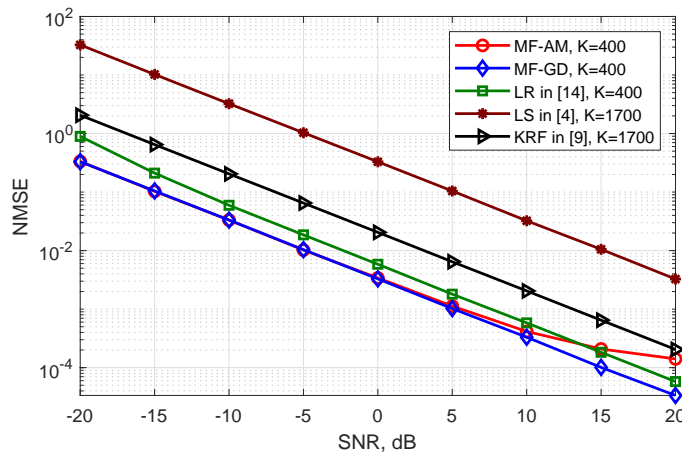
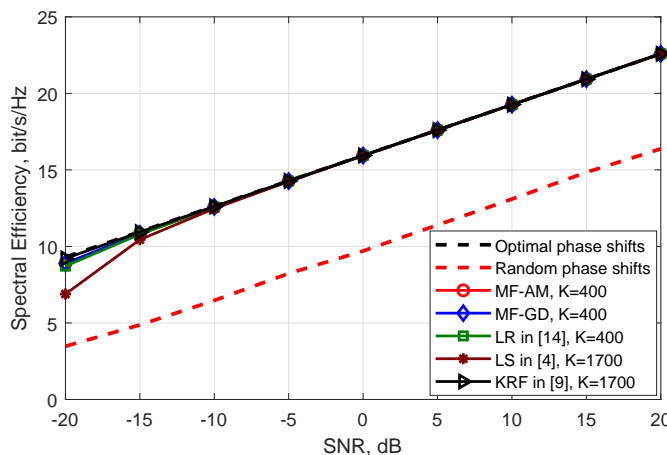Fig. 1: NMSE vs. SNR (dB) in down-link and single-user scenario ($N = 32, M = 50$).

Fig. 2: Spectral efficiency vs. SNR (dB) in down-link and single-user scenario ($N = 32, M = 50$).

V. NUMERICAL RESULTS

A. Training Pilot Overhead

In Table I, we compare the minimal training pilots of our proposed MF with gradient descent (MF-GD) and MF with alternating minimization (MF-AM) with the existing LS method [4], low rank matrix recovery (LR) method [14], and KBF method [9] in the down-link and single-user scenario. As shown in Table I, the proposed MF-GD and MF-AM only require minimal M training pilots, which are much less than that of the existing works in [4], [9], [14].

B. Single-User Scenario

In Fig. 1, we compare the estimation accuracy of our proposed MF-GD and MF-AM with the LS method [4], LR method [14], and KBF method [9] in the down-link and single-user scenario. The normalized MSE (NMSE) is defined as $\mathbb{E}[\|\mathbf{H}_e - \widehat{\mathbf{H}}_e\|_F^2 / \|\mathbf{H}_e\|_F^2]$. Since LS and KBF require $K \geq MN$ to obtain a
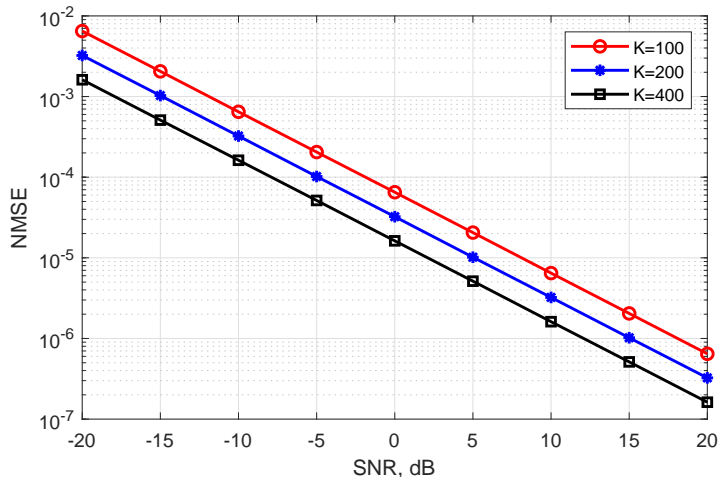
Fig. 3: NMSE vs. SNR (dB) in up-link and multi-user scenario ($N = 32, M = 50, Q = 5, T = 5$).

valid channel estimation, in this work, we let $K = 1700$ for LS and KBF methods, but $K = 400 \ll MN$ for our proposed MF methods. Observed from Fig. 1, though utilizing much less training pilots, the proposed MF methods still outperform LS and KBF. Meanwhile, the proposed MF methods also outperform LR method because they leverage on both the low-rank property and channel structure of the RIS-aided channel. Moreover, compared to MF-AM, the MF-GD can achieve even more accurate estimation as SNR increases.

In Fig. 2, we compare the spectral efficiency achieved by our proposed MF methods with benchmarks [4], [9], [14]. The phase shifts of RIS are based on the estimated $\widehat{\mathbf{H}}_e$ through the method in [8]. The performances of random and optimal phase shifts are also plotted as benchmarks, where phase shifts of the former are uniformly distributed in $[0, 2\pi)$ with zero training pilots, and the latter is designed from the true $\mathbf{H}_e$. As illustrated in Fig. 2, the proposed MF methods can achieve near optimal spectral efficiency with much less training overhead.

C. Multi-User Scenario

In Fig. 3, we illustrate the estimation accuracy of proposed MF method in Section IV for up-link and multi-user RIS-aided channel. The NMSE is defined as $1/Q \sum_{q=1}^{Q} \mathbb{E}[\|\mathbf{H}_q - \widehat{\mathbf{H}}_q\|_F^2 / \|\mathbf{H}_q\|_F^2]$. The simulation parameters are $N = 32, M = 50, Q = 5, T = 5$. As can be seen from Fig. 3, with increasing transmitted blocks K, the NMSE of cascaded channel decreases, which is consistent with our analysis.

VI. CONCLUSION

In this paper, we have investigated the RIS-aided channel estimation via MF. By using the proposed MF method, it only estimates low-dimensional matrices to obtain the estimation of RIS-aided channel.

Compared to the existing works which directly estimate the cascaded channel, the proposed method achieves more accurate estimation with low training overhead.

References

[1] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 3, pp. 436–453, April 2016.

[2] E. Bjornson, O. Ozdogan, and E. G. Larsson, "Intelligent reflecting surface versus decode-and-forward: How large surfaces are needed to beat relaying?" *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 244–248, 2020.

[3] C. Pan *et al.*, "Multicell MIMO communications relying on intelligent reflecting surfaces," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5218–5233, 2020.

[4] 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. ICASSP*, May, 2020, pp. 5000–5004.

[5] C. You, B. Zheng, and R. Zhang, "Channel estimation and passive beamforming for intelligent reflecting surface: Discrete phase shift and progressive refinement," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2604–2620, 2020.

[6] J. Chen, Y.-C. Liang, H. V. Cheng, and W. Yu, "Channel estimation for reconfigurable intelligent surface aided multi-user MIMO systems," *arXiv preprint arXiv:1912.03619*, 2019.

[7] W. Chen, X. Ma, Z. Li, and N. Kuang, "Sum-rate maximization for intelligent reflecting surface based terahertz communication systems," in *Proc. ICCC Workshops*, Aug. 2019, pp. 153–157.

[8] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design," in *Proc. GLOBECOM*, Dec. 2018, pp. 1–6.

[9] G. T. de Araujo, A. L. F. de Almeida, and R. Boyer, "Channel estimation for intelligent reflecting surface assisted MIMO systems: A tensor modeling approach," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 3, pp. 789–802, 2021.

[10] C. Hu, L. Dai, S. Han, and X. Wang, "Two-timescale channel estimation for reconfigurable intelligent surface aided wireless communications," *IEEE Trans. Commun.*, pp. 1–1, 2021.

[11] 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, 2020.

[12] Q.-U.-A. Nadeem, H. Alwazani, A. Kammoun, A. Chaaban, M. Debbah, and M.-S. Alouini, "Intelligent reflecting surface-assisted multi-user MISO communication: Channel estimation and beamforming design," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 661–680, 2020.

[13] W. Zhang, T. Kim, D. J. Love, and E. Perrins, "Leveraging the restricted isometry property: Improved low-rank subspace decomposition for hybrid millimeter-wave systems," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5814–5827, 2018.

[14] P. Jain, P. Netrapalli, and S. Sanghavi, "Low-rank matrix completion using alternating minimization," in *Proc. STOC*, Jun. 2013, pp. 665–674.

[15] Y. Chi, Y. M. Lu, and Y. Chen, "Nonconvex optimization meets low-rank matrix factorization: An overview," *IEEE Trans. Signal Process.*, vol. 67, no. 20, pp. 5239–5269, 2019.