

Channel Estimation for Intelligent Reflecting Surface-Assisted Millimeter Wave MIMO Systems

Tian Lin^{*}, Xianghao Yu[†], Yu Zhu^{*}, and Robert Schober[†],

^{*}Fudan University, China, [†]Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

Email: ^{*}{lint17, zhuyu}@fudan.edu.cn, [†]{xianghao.yu, robert.schober}@fau.de

Abstract—Intelligent reflecting surfaces (IRSs) are regarded as promising enablers for future millimeter wave (mmWave) wireless communication, due to their ability to create favorable line-of-sight (LoS) propagation environments. In this paper, we investigate channel estimation in downlink IRS-assisted mmWave multiple-input multiple-output (MIMO) systems. By leveraging the sparsity of mmWave channels, we formulate the channel estimation problem as a fixed-rank constrained non-convex optimization problem. To tackle the non-convexity, an efficient algorithm is proposed by capitalizing on alternating minimization and manifold optimization (MO), which yields a locally optimal solution. Simulation results show that the proposed MO-based estimation (MO-EST) algorithm significantly outperforms two benchmark schemes and demonstrate the robustness of the MO-EST algorithm with respect to imperfect knowledge of the sparsity level of the channels in practical implementations.

I. INTRODUCTION

Due to its enormous potential for overcoming the spectrum crunch, millimeter wave (mmWave) communications has become a promising technology for future wireless cellular systems [1]. However, mmWave communication is vulnerable to blockages due to the limited scattering effects at mmWave frequencies. Furthermore, in conventional mmWave communication systems, the propagation environment is uncontrollable, and therefore, the quality of service (QoS) is significantly degraded when the line-of-sight (LoS) links are blocked.

Recently, intelligent reflecting surfaces (IRSs) have been incorporated into wireless communication systems, mainly due to their capability of customizing favorable wireless propagation environments [2]. Equipped with a large number of low-cost *passive* reflective elements, e.g., dipoles and phase shifters, IRSs enable the adaptation of wireless propagation environments with limited power consumption [3]. This property of IRSs can be exploited in mmWave systems [4]. Specifically, when the direct LoS links between the transceivers are blocked, the IRSs can reflect the incident signals to provide an effective virtual LoS link for mmWave communications. With well-designed reflecting IRS elements, the communication performance can be further enhanced via programmable and reconfigurable signal reflections [3], [5].

Nevertheless, the introduction of IRSs brings new challenges, among which the acquisition of channel state information (CSI) may be the most demanding task. In particular, in addition to the conventional direct channel between the base station (BS) and the user equipment (UE), two IRS-assisted channels need to be estimated, i.e., the BS-IRS channel and IRS-UE channel. Furthermore, since radio frequency (RF)

chains are not available at the passive IRSs, it is not possible to estimate the two IRS-assisted channels directly by regarding the IRS as a conventional RF chain-driven transceiver. Therefore, the classical channel estimation techniques are not applicable in the newly-emerged IRS-assisted communications systems [5].

Recently, several works have investigated channel estimation in IRS-assisted wireless systems [5]–[10]. The authors of [5] characterized the minimum pilot sequence length for channel estimation in IRS-assisted multi-user multiple-input single-output (MISO) systems based on the least square (LS) criterion. A two-timescale estimation scheme was proposed in [6], where the high-dimensional BS-IRS channel and the low-dimensional IRS-UE channel are estimated in a large timescale and a small timescale, respectively. To further reduce the pilot overhead, by exploiting the sparsity of the channels, compressive sensing techniques were utilized in [7]–[9] to solve the estimation problem. However, the algorithms proposed in these existing works are only applicable in wireless systems with single-antenna users. Multiple-input multiple-output (MIMO) systems were studied in [10], where a channel estimation algorithm for IRS-assisted systems was developed based on parallel factor decomposition (PARAFAC). While this approach, designed for sub-6 GHz bands, is also applicable in mmWave MIMO systems, a significant performance loss is expected as the unique channel characteristics of mmWave MIMO systems are not considered, e.g., the sparsity of mmWave channels.

In this paper, we propose a novel channel estimation algorithm for IRS-assisted point-to-point mmWave MIMO systems. By exploiting the sparsity of mmWave channels, we formulate the channel estimation problem as a non-convex optimization problem with fixed-rank constraints. Then, we apply the alternating minimization principle to divide the original problem into two subproblems, which target the estimation of the BS-IRS channel and the IRS-UE channel, respectively. Finally, manifold optimization (MO) is employed to address the non-convex rank constraint and the subproblems are solved iteratively. The developed algorithm guarantees monotonic convergence to a locally optimal solution. Simulations results clearly illustrate the performance improvement of the proposed MO-based estimation (MO-EST) algorithm over two benchmark schemes including the state-of-the-art PARAFAC approach in [10]. We also demonstrate the robustness of the proposed MO-EST algorithm with respect to different channel sparsity levels.

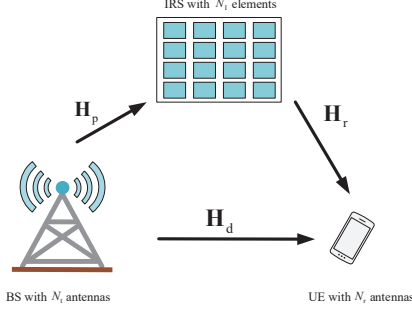


Fig. 1. A downlink IRS-assisted mmWave MIMO communication system.

Notations: In this paper, the imaginary unit of a complex number is denoted by $j = \sqrt{-1}$. The set of nonnegative integers is denoted by $\mathbb{N} = \{0, 1, \dots\}$. $\mathbb{C}^{m \times n}$ denotes the set of all $m \times n$ complex-valued matrices. Matrices and vectors are denoted by boldface capital and lower-case letters, respectively. The i -th element of vector \mathbf{x} is denoted by x_i . \mathbf{I}_N denotes the $N \times N$ identity matrix. $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$, $\text{rank}(\cdot)$, $\text{tr}(\cdot)$, $\text{vec}(\cdot)$, and $\|\cdot\|_F$ denote the conjugate, transpose, conjugate transpose, rank, trace, vectorization, and Frobenius norm of a matrix, respectively. The Khatri-Rao matrix product is represented by \odot . $\Re(\cdot)$ and $\mathbb{E}(\cdot)$ denote the real part of a complex number and expectation, respectively. $\text{diag}(\mathbf{x})$ is a diagonal matrix with the entries of \mathbf{x} on its main diagonal. $\mathcal{CN}(\mathbf{0}, \Sigma)$ denotes the circularly symmetric complex Gaussian distribution with zero mean and covariance matrix Σ .

II. SYSTEM MODEL

As shown in Fig. 1, we consider the downlink of an IRS-assisted point-to-point mmWave MIMO communication system. BS and UE are equipped with uniform planar arrays (UPAs) consisting of N_t and N_r antennas, respectively. In addition, a passive IRS that employs N_I phase shifters is deployed in the network to facilitate mmWave communications.

A. Channel Estimation Protocol

The channel estimation protocol adopted in this paper is shown in Fig. 2. Specifically, the time available for estimation is divided into B blocks, and each block consists of T pilot symbol durations. The reflection coefficient vectors of the IRS may be different in different blocks but are constant within one block [7], [10]. The pilots received at the UE in T consecutive time slots of the b -th block, denoted by $\mathbf{R}_b \in \mathbb{C}^{N_r \times T}$, are compactly written as

$$\mathbf{R}_b = (\mathbf{H}_r \text{diag}(\mathbf{v}_b) \mathbf{H}_p + \mathbf{H}_d) \mathbf{X}_b + \mathbf{Z}_b, \quad (1)$$

where the BS-UE, BS-IRS, and IRS-UE channel matrices are denoted by $\mathbf{H}_d \in \mathbb{C}^{N_r \times N_t}$, $\mathbf{H}_p \in \mathbb{C}^{N_I \times N_t}$, and $\mathbf{H}_r \in \mathbb{C}^{N_r \times N_I}$, respectively. $\mathbf{Z}_b = [\mathbf{z}_1, \dots, \mathbf{z}_T] \in \mathbb{C}^{N_r \times T}$ denotes the received Gaussian noise with $\mathbf{z}_t \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_{N_r})$, $\forall t \in \{1, \dots, T\}$. $\mathbf{v}_b = [v_{b,1}, \dots, v_{b,N_I}]^T \in \mathbb{C}^{N_I}$ is the training reflection coefficient vector in the b -th block. Since the IRS is implemented by phase shifters [3], the reflecting elements can

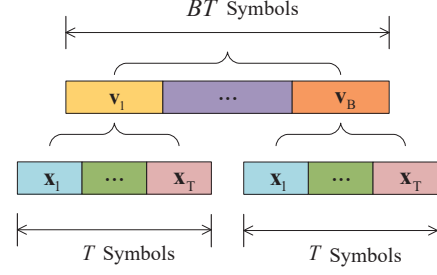


Fig. 2. Frame structure of the channel estimation protocol.

only change the phases of the received signals, i.e., $|v_{b,n}| = 1$. Furthermore, we assume that the pilot sequences $\mathbf{X}_b \in \mathbb{C}^{N_t \times T}$ transmitted by the BS are orthogonal to each other, namely, $\mathbf{X}_b \mathbf{X}_b^H = T \mathbf{I}_{N_t}$ [10]. Thus, after removing the pilot symbols at the UE, we have

$$\hat{\mathbf{Y}}_b \triangleq \mathbf{R}_b \mathbf{X}_b^H = (\mathbf{H}_r \text{diag}(\mathbf{v}_b) \mathbf{H}_p + \mathbf{H}_d) + \hat{\mathbf{Z}}_b, \quad (2)$$

where $\hat{\mathbf{Z}}_b = \mathbf{Z}_b \mathbf{X}_b^H \in \mathbb{C}^{N_r \times N_t}$.

Remark 1: By switching off all IRS elements, the direct BS-UE channel can be estimated via traditional algorithms, e.g., [5]. Therefore, in this paper, we assume that \mathbf{H}_d is known and focus on the estimation of \mathbf{H}_p and \mathbf{H}_r , which is the main challenge in channel estimation for IRS-assisted systems. Thus, the relevant part of $\hat{\mathbf{Y}}_b$ is given by

$$\mathbf{Y}_b = \mathbf{H}_r \text{diag}(\mathbf{v}_b) \mathbf{H}_p + \hat{\mathbf{Z}}_b. \quad (3)$$

We further concatenate the signals received in all B sub-frames as $\tilde{\mathbf{Y}}_1 = [\mathbf{Y}_1^T, \dots, \mathbf{Y}_B^T]^T \in \mathbb{C}^{BN_r \times N_t}$, which leads to

$$\tilde{\mathbf{Y}}_1 = (\mathbf{V} \odot \mathbf{H}_r) \mathbf{H}_p + \tilde{\mathbf{Z}}_1, \quad (4)$$

where $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_B]^T \in \mathbb{C}^{B \times N_I}$ and $\tilde{\mathbf{Z}}_1 = [\hat{\mathbf{Z}}_1^T, \dots, \hat{\mathbf{Z}}_B^T]^T \in \mathbb{C}^{BN_r \times N_t}$.

B. MmWave Channel Model

Before formulating the estimation problem, we introduce the channel model for mmWave propagation. The mmWave propagation environment is well characterized by the Saleh-Valenzuela model [9], which is given by

$$\begin{aligned} \mathbf{H}_r &= \sqrt{\frac{N_r N_I}{P}} \sum_{p=1}^P \alpha_p \mathbf{a}_r(\theta_r^p, \phi_r^p) \mathbf{a}_t^H(\theta_t^p, \phi_t^p), \\ \mathbf{H}_p &= \sqrt{\frac{N_t N_I}{Q}} \sum_{q=1}^Q \beta_q \mathbf{a}_r(\psi_r^q, \varphi_r^q) \mathbf{a}_t^H(\psi_t^q, \varphi_t^q), \end{aligned} \quad (5)$$

where α_p , θ_r^p (ϕ_r^p), and θ_t^p (ϕ_t^p) denote the complex gain, azimuth (elevation) angle of arrival (AoA), and azimuth (elevation) angle of departure (AoD) of the p -th path of the IRS-UE channel. Similarly, β_q , ψ_r^q (φ_r^q), and ψ_t^q (φ_t^q) denote the complex gain, azimuth (elevation) AoA, and azimuth (elevation) AoD of the q -th path of the BS-IRS channel. In

addition, \mathbf{a}_r and \mathbf{a}_t denote the receive and transmit array response vectors, respectively. The array response vector of a half-wavelength spaced UPA with $M \times N$ elements is given as follows

$$\mathbf{a}(\theta, \phi) = \frac{1}{\sqrt{MN}} \left[1, \dots, e^{j\pi(n \sin \theta \sin \phi + m \cos \phi)}, \dots, e^{j\pi((\sqrt{N}-1) \sin \theta \sin \phi + (\sqrt{M}-1) \cos \phi)} \right]^T, \quad (6)$$

where m and n are the antenna element indices in the 2-dimensional plane. An important property of mmWave channels is presented in the following lemma.

Lemma 1. Suppose $\min(N_t, N_r, N_I) \geq \max(P, Q)$, then we have

$$\text{rank}(\mathbf{H}_r) = P, \quad \text{rank}(\mathbf{H}_p) = Q. \quad (7)$$

Proof: Please refer to Appendix A. ■

Remark 2: In Section III, the numbers of paths P and Q are assumed to be known at the BS, and hence, the achieved performance is an upper bound for the scenario where P and Q are not available or cannot be accurately estimated. In practice, the numbers of paths can be estimated via low-complexity compressive sensing methods, e.g., the orthogonal matching pursuit (OMP) method in [9]. In Section IV, we consider the case where P and Q are not exactly known to evaluate the robustness of the proposed algorithm with respect to a mismatched number of paths.

C. Problem Formulation

According to [11], the minimum variance unbiased estimators of \mathbf{H}_p and \mathbf{H}_r can be obtained based on the LS criterion. By taking the sparsity of the channels into account and leveraging Lemma 1, we formulate the channel estimation problem in IRS-assisted mmWave MIMO systems as follows

$$\begin{aligned} & \underset{\hat{\mathbf{H}}_r, \hat{\mathbf{H}}_p}{\text{minimize}} \quad f = \left\| \tilde{\mathbf{Y}}_1 - (\mathbf{V} \odot \hat{\mathbf{H}}_r) \hat{\mathbf{H}}_p \right\|_F^2 \\ & \text{subject to} \quad \text{rank}(\hat{\mathbf{H}}_r) = P, \text{rank}(\hat{\mathbf{H}}_p) = Q, \end{aligned} \quad (8)$$

where $\hat{\mathbf{H}}_r$ and $\hat{\mathbf{H}}_p$ denote the estimates of \mathbf{H}_r and \mathbf{H}_p , respectively. Due to the fixed-rank constraints, problem (8) is a highly non-convex problem and a globally optimal solution would entail a very high computational complexity. Besides, the coupling of the two optimization variables in the objective function further complicates the problem. Thus, in the following, we propose an efficient algorithm that achieves a locally optimal solution of problem (8).

III. PROPOSED MO-EST ALGORITHM

To tackle the coupling of the optimization variables in (8), we first decouple the two variables by applying the alternating minimization principle [3], [12]. Specifically, we first fix $\hat{\mathbf{H}}_r$ and minimize f with respect to the single variable $\hat{\mathbf{H}}_p$. The corresponding subproblem is given by

$$\begin{aligned} & \underset{\hat{\mathbf{H}}_p}{\text{minimize}} \quad f = \left\| \tilde{\mathbf{Y}}_1 - (\mathbf{V} \odot \hat{\mathbf{H}}_r) \hat{\mathbf{H}}_p \right\|_F^2 \\ & \text{subject to} \quad \text{rank}(\hat{\mathbf{H}}_p) = Q. \end{aligned} \quad (9)$$

To address the non-convex fixed-rank constraint, we apply the MO technique to solve problem (9). Different from traditional compressive sensing methods, e.g., the OMP and basis pursuit (BP) methods, the proposed MO-based algorithm guarantees convergence to a locally optimal solution of problem (9).

A. Preliminaries of MO

First, we note that the feasible set of problem (9) can be represented as a fixed-rank manifold

$$\mathcal{M}_Q \triangleq \{ \mathbf{X} \in \mathbb{C}^{N_t \times N_t} : \text{rank}(\mathbf{X}) = Q \}, \quad (10)$$

which is a smooth complex Riemannian manifold. The Riemannian optimization method for the real-valued fixed-rank manifold has been studied in [13]. By extending the definitions of the fixed-rank manifold to the complex domain, we introduce the key operations that are necessary for the Riemannian optimization method for \mathcal{M}_Q .

1) *Inner product:* By endowing the complex space $\mathbb{C}^{N_t \times N_t}$ with the Euclidean metric, the inner product between two points $\mathbf{X}_1, \mathbf{X}_2 \in \mathcal{M}_Q$ is defined as

$$\langle \mathbf{X}_1, \mathbf{X}_2 \rangle = \Re \{ \text{tr}(\mathbf{X}_1^H \mathbf{X}_2) \}. \quad (11)$$

2) *Tangent space:* For a point $\mathbf{X} \in \mathcal{M}_Q$ on the manifold, its tangent space $T_{\mathbf{X}}\mathcal{M}_Q$, which is composed of all the vectors that tangentially pass through \mathbf{X} , is given by [13]

$$T_{\mathbf{X}}\mathcal{M}_Q \triangleq \{ \mathbf{X}_U \mathbf{M} \mathbf{X}_V^H + \mathbf{U}_p \mathbf{X}_V^H + \mathbf{X}_U \mathbf{V}_p^H : \mathbf{M} \in \mathbb{C}^{Q \times Q} \}, \quad (12)$$

where $\mathbf{X}_U \in \mathbb{C}^{N_t \times Q}$ and $\mathbf{X}_V \in \mathbb{C}^{N_t \times Q}$ denote the semi-unitary matrices containing the first Q left and right singular vectors of \mathbf{X} , respectively. In addition, $\mathbf{U}_p \in \mathbb{C}^{N_t \times Q}$ and $\mathbf{V}_p \in \mathbb{C}^{N_t \times Q}$ lie in the null spaces of \mathbf{X}_U and \mathbf{X}_V , respectively, i.e., $\mathbf{U}_p^H \mathbf{X}_U = \mathbf{0}$, $\mathbf{V}_p^H \mathbf{X}_V = \mathbf{0}$.

3) *Orthogonal projection:* The orthogonal projection of a point $\mathbf{J} \in \mathbb{C}^{N_t \times N_t}$ onto the tangent space of \mathbf{X} , $T_{\mathbf{X}}\mathcal{M}_Q$, is given by

$$P_{T_{\mathbf{X}}\mathcal{M}_Q}(\mathbf{J}) = \mathbf{P}_U \mathbf{J} \mathbf{P}_V + \mathbf{P}_U^\perp \mathbf{J} \mathbf{P}_V + \mathbf{P}_U \mathbf{J} \mathbf{P}_V^\perp, \quad (13)$$

where $\mathbf{P}_U = \mathbf{X}_U \mathbf{X}_U^H$, $\mathbf{P}_V = \mathbf{X}_V \mathbf{X}_V^H$, $\mathbf{P}_U^\perp = \mathbf{I}_{N_t} - \mathbf{P}_U$, and $\mathbf{P}_V^\perp = \mathbf{I}_{N_t} - \mathbf{P}_V$.

4) *Retraction:* Retraction is a mapping from the tangent space to the manifold itself. Particularly, for a point $\tilde{\mathbf{X}} \in T_{\mathbf{X}}\mathcal{M}_Q$, the retraction operation can be formulated via a truncated singular value decomposition (SVD)

$$\mathcal{R}(\tilde{\mathbf{X}}) \triangleq T_{\mathbf{X}}\mathcal{M}_Q \mapsto \mathcal{M}_Q : \tilde{\mathbf{X}} \mapsto \sum_{i=1}^Q \sigma_i \mathbf{u}_i \mathbf{v}_i^H, \quad (14)$$

where σ_i , \mathbf{u}_i , and \mathbf{v}_i are the ordered singular values, left singular vectors, and right singular vectors of $\tilde{\mathbf{X}}$, respectively.

B. Conjugate Gradient Method on \mathcal{M}_Q

With the basic definitions of the key operations on \mathcal{M}_Q at hand, we can formulate the counterpart of the classic conjugate gradient (CG) algorithm in the Euclidean space on the manifold \mathcal{M}_Q [12], [14]. The main idea is illustrated in Fig. 3. In each iteration, we first find a local minimizer in the tangent space, and then project the obtained point back to the

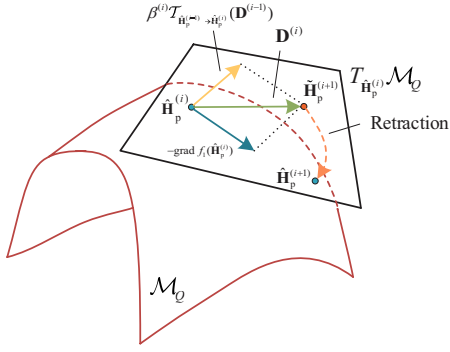


Fig. 3. Illustration of the generalized CG method for the fixed-rank manifold.

manifold. For problem (9), the update rule of the CG method is given by [13]

$$\tilde{\mathbf{H}}_p^{(i+1)} = \hat{\mathbf{H}}_p^{(i)} + \alpha^{(i)} \mathbf{D}^{(i)}, \quad (15)$$

where $\alpha^{(i)}$ denotes the Armijo backtracking step size in the i -th iteration [15, Eq. (59)], $\hat{\mathbf{H}}_p^{(i)}$ is the estimate in the i -th iteration, and $\tilde{\mathbf{H}}_p^{(i+1)}$ is the updated local minimizer in the $(i+1)$ -th iteration. In addition, $\mathbf{D}^{(i)}$ is the conjugate direction in the i -th iteration, given by

$$\mathbf{D}^{(i)} = -\text{grad} f(\hat{\mathbf{H}}_p^{(i)}) + \beta^{(i)} T_{\hat{\mathbf{H}}_p^{(i-1)} \rightarrow \hat{\mathbf{H}}_p^{(i)}}(\mathbf{D}^{(i-1)}), \quad (16)$$

where the first term is the negative Riemannian gradient representing the steepest descent direction of the objective function in the tangent space $T_{\hat{\mathbf{H}}_p^{(i)}} \mathcal{M}_Q$, and $\beta^{(i)}$ represents the chosen Polak-Ribiere parameter [15, p. 42]. Since the conjugate direction in the previous iteration $\mathbf{D}^{(i-1)}$ does not lie in $T_{\hat{\mathbf{H}}_p^{(i)}} \mathcal{M}_Q$, the sum operation in (16) can not be performed directly. To this end, we introduce the *vector transport* operation to project $\mathbf{D}^{(i-1)}$ to the current tangent space $T_{\hat{\mathbf{H}}_p^{(i)}} \mathcal{M}_Q$. According to (13), the vector transport for \mathcal{M}_Q is given by

$$T_{\hat{\mathbf{H}}_p^{(i-1)} \rightarrow \hat{\mathbf{H}}_p^{(i)}} = P_{T_{\hat{\mathbf{H}}_p^{(i)}} \mathcal{M}_Q}(\mathbf{D}^{(i-1)}). \quad (17)$$

Therefore, the remaining task to determine the conjugate direction in (16) is to derive the Riemannian gradient. Since \mathcal{M}_Q is embedded in $\mathbb{C}^{N_t \times N_t}$, the Riemannian gradient is obtained by projecting the Euclidean gradient onto the tangent space [14], i.e.,

$$\text{grad} f(\hat{\mathbf{H}}_p) = P_{T_{\hat{\mathbf{H}}_p} \mathcal{M}_Q}(\mathbf{G}_1). \quad (18)$$

The Euclidean gradient \mathbf{G}_1 of f with respect to $\hat{\mathbf{H}}_p$ is given by

$$\mathbf{G}_1 = (\mathbf{V} \odot \hat{\mathbf{H}}_r)^H \left((\mathbf{V} \odot \hat{\mathbf{H}}_r) \hat{\mathbf{H}}_p - \tilde{\mathbf{Y}}_1 \right). \quad (19)$$

After updating the local minimizer in the $(i+1)$ -th iteration according to (15), we need to map this minimizer $\tilde{\mathbf{H}}_p^{(i+1)}$ back to \mathcal{M}_Q to obtain the estimate in the $(i+1)$ -th iteration, which is achieved by the retraction operation shown in (14), i.e.,

$$\hat{\mathbf{H}}_p^{(i+1)} = \mathcal{R}(\tilde{\mathbf{H}}_p^{(i+1)}). \quad (20)$$

Algorithm 1 CG-MO Algorithm

Input: $\hat{\mathbf{H}}_p^{(0)} \in \mathcal{M}_Q$, $\hat{\mathbf{H}}_r$, \mathbf{V} , $\tilde{\mathbf{Y}}_1$

- 1: Set $i = 0$ and $f^{(0)} = f(\hat{\mathbf{H}}_p^{(0)})$;
 - 2: **repeat**
 - 3: Compute the Riemannian gradient $\text{grad} f(\hat{\mathbf{H}}_p^{(i)})$ according to (18) and (19);
 - 4: Compute the conjugate direction $\mathbf{D}^{(i)}$ according to (16);
 - 5: Update $\tilde{\mathbf{H}}_p^{(i+1)}$ according to (15);
 - 6: Retract $\tilde{\mathbf{H}}_p^{(i+1)}$ to obtain $\hat{\mathbf{H}}_p^{(i+1)}$ according to (20);
 - 7: $i \leftarrow i + 1$;
 - 8: $f^{(i)} = f(\hat{\mathbf{H}}_p^{(i)})$;
 - 9: **until** $f^{(i-1)} - f^{(i)} \leq \epsilon$;
 - 10: Update $\hat{\mathbf{H}}_p^{(i)}$ as the estimate of \mathbf{H}_p .
-

The proposed generalized CG method for the fixed-rank manifold, referred to as the **CG-MO algorithm**, is summarized in **Algorithm 1**, where ϵ is the convergence threshold.

C. Estimation of \mathbf{H}_r

In this subsection, we present the optimization of $\hat{\mathbf{H}}_r$ for given $\hat{\mathbf{H}}_p$. First, we establish the following equality

$$\|\tilde{\mathbf{Y}}_1 - (\mathbf{V} \odot \hat{\mathbf{H}}_r) \hat{\mathbf{H}}_p\|_F^2 = \|\tilde{\mathbf{Y}}_2 - (\mathbf{V} \odot \hat{\mathbf{H}}_p^T) \hat{\mathbf{H}}_r^T\|_F^2, \quad (21)$$

where $\tilde{\mathbf{Y}}_2 = [\mathbf{Y}_1, \dots, \mathbf{Y}_B]^T \in \mathbb{C}^{BN_t \times N_r}$. The subproblem that optimizes $\hat{\mathbf{H}}_r$ for given $\hat{\mathbf{H}}_p$ is then formulated as follows

$$\begin{aligned} &\underset{\hat{\mathbf{H}}_r}{\text{minimize}} && f = \|\tilde{\mathbf{Y}}_2 - (\mathbf{V} \odot \hat{\mathbf{H}}_p^T) \hat{\mathbf{H}}_r^T\|_F^2 \\ &\text{subject to} && \text{rank}(\hat{\mathbf{H}}_r) = P. \end{aligned} \quad (22)$$

Thus, the **CG-MO algorithm** is also applicable to solving problem (22). The main modification compared to the optimization of $\hat{\mathbf{H}}_p$ is the replacement of the Euclidean gradient in (19) by the Euclidean gradient of f with respect to $\hat{\mathbf{H}}_r$, which is given by

$$\mathbf{G}_2 = \left(\hat{\mathbf{H}}_r (\mathbf{V} \odot \hat{\mathbf{H}}_p^T)^T - \tilde{\mathbf{Y}}_2^T \right) (\mathbf{V} \odot \hat{\mathbf{H}}_p^T)^*. \quad (23)$$

Finally, the overall estimation scheme is referred to as the **MO-EST algorithm** and summarized in **Algorithm 2**. With the proposed algorithm, the objective values f achieved by the sequence $\{\hat{\mathbf{H}}_p^{(k)}, \hat{\mathbf{H}}_r^{(k)}\}_{k \in \mathbb{N}}$ form a non-increasing sequence that converges to a stationary value, and any limit point of the sequence $\{\hat{\mathbf{H}}_p^{(k)}, \hat{\mathbf{H}}_r^{(k)}\}_{k \in \mathbb{N}}$ is a stationary point of problem (9) [13].

IV. SIMULATION RESULTS

In this section, we provide simulation results for performance evaluation of the proposed MO-EST algorithm. The signal-to-noise-ratio (SNR) is defined as $\frac{1}{\sigma^2}$. Square UPAs are equipped at both the BS and UE. For both \mathbf{H}_p and \mathbf{H}_r , the same number of paths are assumed, i.e., $P = Q \triangleq C$. According to the channel model in (5), without loss of

Algorithm 2 MO-EST Algorithm

Input: \mathbf{V} , $\tilde{\mathbf{Y}}_1$, $\tilde{\mathbf{Y}}_2$

- 1: Randomly initialize $\hat{\mathbf{H}}_r^{(0)} \in \mathcal{M}_P$ and $\hat{\mathbf{H}}_p^{(0)} \in \mathcal{M}_Q$, set $k = 0$ and $f^{(0)} = f(\hat{\mathbf{H}}_r^{(0)}, \hat{\mathbf{H}}_p^{(0)})$;
 - 2: **repeat**
 - 3: $k \leftarrow k + 1$;
 - 4: Optimize $\hat{\mathbf{H}}_p^{(k)}$ for given $\hat{\mathbf{H}}_r^{(k-1)}$ by solving problem (9) with the CG-MO algorithm;
 - 5: Optimize $\hat{\mathbf{H}}_r^{(k)}$ for given $\hat{\mathbf{H}}_p^{(k)}$ by solving problem (22) with the CG-MO algorithm;
 - 6: $f^{(k)} = f(\hat{\mathbf{H}}_r^{(k)}, \hat{\mathbf{H}}_p^{(k)})$;
 - 7: **until** $f^{(k-1)} - f^{(k)} \leq \epsilon$;
 - 8: Update $\hat{\mathbf{H}}_p^{(k)}$ and $\hat{\mathbf{H}}_r^{(k)}$ as the estimates of \mathbf{H}_p and \mathbf{H}_r .
-

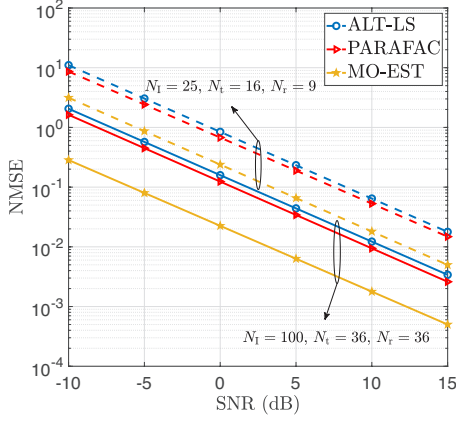


Fig. 4. NMSE versus SNR for different estimation algorithms when $C = 3$.

generality, we let $p = 1$ and $q = 1$ represent the indices of the LoS components in \mathbf{H}_r and \mathbf{H}_p . The complex channel gains are distributed as $\alpha_1 (\beta_1) \sim \mathcal{CN}(0, 1)$ and $\alpha_i (\beta_i) \sim \mathcal{CN}(0, 10^{-0.5})$ for $i = 2, \dots, C$ [9]. The azimuth and elevation AoAs/AoDs, i.e., θ_r^p (θ_t^p) and ϕ_r^p (ϕ_t^p), are generated uniformly distributed in $[0, \pi]$ and $[-\pi/2, \pi/2]$, respectively. For the reflecting elements, we set $N_l = B$ and use the discrete Fourier transform (DFT) matrix as \mathbf{V} . The normalized mean square error (NMSE) is adopted as the performance metric. The NMSE is defined as $\mathbb{E} \left\{ \|\mathbf{H}_c - \hat{\mathbf{H}}_c\|_F^2 / \|\mathbf{H}_c\|^2 \right\}$, where $\mathbf{H}_c = \mathbf{H}_r \mathbf{H}_p$ and $\hat{\mathbf{H}}_c = \hat{\mathbf{H}}_r \hat{\mathbf{H}}_p$ denote the cascaded channel and its estimate, respectively¹. The convergence threshold in both **Algorithm 1** and **2** is set as $\epsilon = 10^{-3}$. To show the effectiveness of the proposed MO-EST algorithm, the PARAFAC algorithm [10] is adopted as a benchmark. In addition, by dropping the rank constraints, the LS problems in (9) and (22) can be alternately solved in closed form. This approach is also adopted as a benchmark and is referred to as the **ALT-LS algorithm**.

In Fig. 4, the NMSE is plotted as a function of SNR

¹As the \mathbf{H}_r and \mathbf{H}_p are coupled in the received signal \mathbf{Y}_b , there inevitably exist scaling ambiguities between $\hat{\mathbf{H}}_r$ and $\hat{\mathbf{H}}_p$. Therefore, the NMSE of $\hat{\mathbf{H}}_c$ is adopted as performance metric to avoid the scaling issues [10].

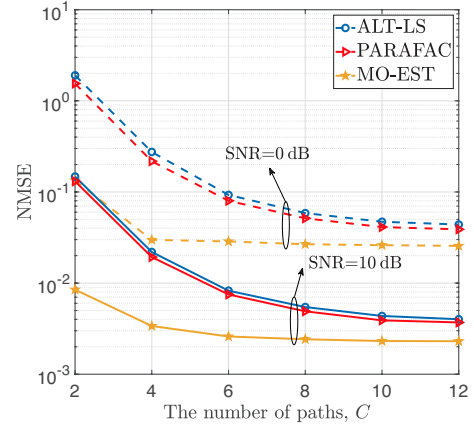


Fig. 5. Effect of the number of paths for different estimation algorithms when $N_r = 16$, $N_t = 36$, and $N_l = 64$.

when $C = 3$. It can be observed that our proposed MO-EST algorithm achieves a significant performance gain of more than 4 dB compared to the two benchmark schemes. This is mainly because the proposed MO-EST algorithm exploits the sparsity of the involved mmWave channels. In contrast, the ALT-LS algorithm yields the highest NMSE. This phenomenon highlights the importance of incorporating the rank constraints into the alternating optimization algorithm for channel estimation in IRS-assisted mmWave MIMO systems. Furthermore, as more antennas and reflection elements provide more spatial degrees of freedom for channel estimation, the performance of all three algorithms is improved for larger values of N_t , N_r , and N_l .

In Fig. 5, we investigate the impact of the number of paths of the estimated channels, i.e., C , when $N_r = 16$, $N_t = 36$, and $N_l = 64$. As can be observed, MO-EST outperforms the two benchmark algorithms for all considered values of C . The performance gain is especially significant in the high sparsity regime. This is because the performance gain mainly comes from the exploitation of channel sparsity. As the number of paths of the estimated channels increases, the channel sparsity level decreases. Therefore, the performance gap is larger when C is small, which is typically the case for mmWave channels where scattering is very limited.

In Fig. 6, we consider the case where the number of paths, C , is not perfectly known for channel estimation and test the robustness of the MO-EST algorithm with respect to the resulting uncertainty. The parameters are set as $C = 3$, $N_r = 16$, $N_t = 36$, $N_l = 64$, and the estimated number of paths is denoted as \hat{C} . As can be observed, with the proposed MO-EST algorithm, the lowest NMSE is achieved when $\hat{C} = C$, i.e., the number of paths is perfectly known. In contrast, the performance of the two benchmark algorithms does not depend on the number of paths, and therefore the achieved NMSEs are independent of \hat{C} . In addition, for the MO-EST algorithm, the mismatch between the estimated \hat{C} and the true value of C leads to a performance loss, which, nevertheless, is limited especially when $\hat{C} \geq C$. In particular, the channel matrix \mathbf{H}_c and its estimate $\hat{\mathbf{H}}_c$ can

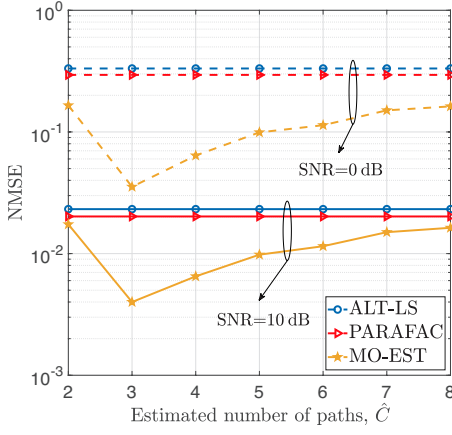


Fig. 6. Effect of mismatches of \hat{C} on the NMSE of the MO-EST algorithm when $C = 3$, $N_r = 16$, $N_t = 36$, and $N_I = 64$.

be decomposed by their SVDs, i.e., $\mathbf{H}_c = \Sigma_{c=1}^C \sigma_c \mathbf{u}_c \mathbf{v}_c^H$ and $\hat{\mathbf{H}}_c = \Sigma_{c=1}^{\hat{C}} \hat{\sigma}_c \hat{\mathbf{u}}_c \hat{\mathbf{v}}_c^H$, where σ_c ($\hat{\sigma}_c$), \mathbf{u}_c ($\hat{\mathbf{u}}_c$), and \mathbf{v}_c ($\hat{\mathbf{v}}_c$) denote the ordered singular values, left singular vectors, and right singular vectors, respectively. In order to minimize the objective function in (8) based on the LS criterion, the MO-EST algorithm chooses the C largest singular values of $\hat{\mathbf{H}}_c$ and the corresponding singular vectors to be close to the true values while making the remaining $\hat{C} - C$ singular values small. In other words, the solution obtained by the MO-EST algorithm satisfies $\hat{\sigma}_c \approx \sigma_c$, $\hat{\mathbf{u}}_c \approx \mathbf{u}_c$, $\hat{\mathbf{v}}_c \approx \mathbf{v}_c$ for $c = 1, \dots, C$, and $\hat{\sigma}_c \approx 0$ for $c = C + 1, \dots, \hat{C}$, which still maintains a satisfactory estimation performance when $\hat{C} \geq C$. Hence, the proposed MO-EST algorithm is robust with respect to imperfect knowledge of the exact number of paths of the estimated channels.

V. CONCLUSION

In this paper, we investigated the channel estimation problem for IRS-assisted mmWave MIMO systems. By exploiting the sparsity of the mmWave channel, a manifold optimization-based alternating optimization algorithm, i.e., the MO-EST algorithm, was developed to effectively estimate the BS-IRS and IRS-UE channels. Simulation results showed the achieved performance improvements compared to two existing benchmark schemes, even when the sparsity level of the channels was not accurately unknown. As a next step, it is of great interest to extend this work to multi-user and broadband scenarios.

APPENDIX A

According to (6), the receive array response vector $\mathbf{a}_r(\theta_r^p, \phi_r^p)$ can be written as

$$\mathbf{a}_r(\theta_r^p, \phi_r^p) = \text{vec} \left(\frac{1}{\sqrt{MN}} \Psi \left[\bar{\mathbf{a}}_{r,0}^p, \dots, \bar{\mathbf{a}}_{r,m}^p, \dots, \bar{\mathbf{a}}_{r,M-1}^p \right]^T \right), \quad (24)$$

where $\Psi = \text{diag}([e^{j\pi^0 \cos \phi_r^p}, \dots, e^{j\pi(M-1) \cos \phi_r^p}])$ and $\bar{\mathbf{a}}_{r,m}^p = [1, \dots, e^{j\pi(N-1) \sin \theta_r^p \sin \phi_r^p}]^T$. When $\min(N_t, N_r, N_I) \geq \max(P, Q)$, it can be shown that matrix $\bar{\mathbf{A}}_{r,n} = [\bar{\mathbf{a}}_{r,m}^1, \dots, \bar{\mathbf{a}}_{r,m}^P]$ is an $N \times P$ Vandermonde matrix,

whose column vectors are linearly independent. Therefore, the vectors $\mathbf{a}_r(\theta_r^1, \phi_r^1), \dots, \mathbf{a}_r(\theta_r^P, \phi_r^P)$ are also linearly independent and matrix $\mathbf{A}_r = [\mathbf{a}_r(\theta_r^1, \phi_r^1), \dots, \mathbf{a}_r(\theta_r^P, \phi_r^P)]$ satisfies $\text{rank}(\mathbf{A}_r) = P$. Similarly, $\mathbf{A}_t = [\mathbf{a}_t(\theta_t^1, \phi_t^1), \dots, \mathbf{a}_t(\theta_t^P, \phi_t^P)]$ also satisfies $\text{rank}(\mathbf{A}_t) = P$. According to (5), \mathbf{H}_r can be expressed as

$$\mathbf{H}_r = \mathbf{A}_r \Sigma \mathbf{A}_t^H, \quad (25)$$

where $\Sigma = \text{diag}(\alpha_1, \dots, \alpha_P)$ is also a rank- P matrix. We have the following inequalities

$$\begin{aligned} \text{rank}(\mathbf{AB}) &\geq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) - k, \\ \text{rank}(\mathbf{AB}) &\leq \min\{\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B})\} \end{aligned} \quad (26)$$

for arbitrary matrices $\mathbf{A} \in \mathbb{C}^{m \times k}$ and $\mathbf{B} \in \mathbb{C}^{k \times n}$. Combining the results in (25) and (26), it can be shown that

$$\text{rank}(\mathbf{H}_r) = P, \quad (27)$$

and similarly we can prove $\text{rank}(\mathbf{H}_p) = Q$.

REFERENCES

- [1] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499-1513, Mar. 2014.
- [2] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: intelligent reflecting surface aided wireless network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106-112, Jan. 2020.
- [3] X. Yu, D. Xu, and R. Schober, "Enabling secure wireless communications via intelligent reflecting surfaces," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Waikoloa, HI, USA, 2019, pp. 1-6.
- [4] Y. Xiu, Y. Zhao, Y. Liu, J. Zhao, O. Yagan, and N. Wei, "IRS-assisted millimeter wave communications: Joint power allocation and beamforming design", Jan. 2020, [Online]. Available: <https://arxiv.org/abs/2001.07467>.
- [5] Z. Wang, L. Liu, and S. Cui, "Channel estimation for intelligent reflecting surface assisted multiuser communications: Framework, algorithms, and analysis", Dec. 2019, [Online]. Available: <https://arxiv.org/abs/1912.11783>.
- [6] C. Hu and L. Dai, "Two-timescale channel estimation for reconfigurable intelligent surface aided wireless communications", Dec. 2019, [Online]. Available: <https://arxiv.org/abs/1912.07990>.
- [7] J. Chen, Y. Liang, H. Cheng, and W. Yu, "Channel estimation for reconfigurable intelligent surface aided multi-user MIMO systems", Dec. 2019, [Online]. Available: <https://arxiv.org/abs/1912.03619>.
- [8] Z. Wan, Z. Gao, and M. Alouini, "Broadband channel estimation for intelligent reflecting surface aided mmWave massive MIMO systems", Feb. 2020, [Online]. Available: <https://arxiv.org/abs/2002.01629>.
- [9] P. Wang, J. Fang, H. Duan, and H. Li, "Compressed channel estimation and joint beamforming for intelligent reflecting surface-assisted millimeter wave systems", Nov. 2019, [Online]. Available: <https://arxiv.org/abs/1911.07202>.
- [10] G. Arajo and A. Almeida, "PARAFAC-based channel estimation for intelligent reflective surface assisted MIMO system", Jan. 2020, [Online]. Available: <https://arxiv.org/abs/2001.06554>.
- [11] T. L. Jensen and E. De Carvalho, "An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator", Nov. 2019, [Online]. Available: <http://arxiv.org/abs/1909.09440>.
- [12] T. Lin, J. Cong, Y. Zhu, J. Zhang, and K. B. Letaief, "Hybrid beamforming for millimeter wave systems using the MMSE criterion," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3693-3708, May 2019.
- [13] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2009.
- [14] B. Vandereycken, "Low-rank matrix completion by Riemannian optimization—extended version", Sept. 2012, [Online]. Available: <https://arxiv.org/abs/1209.3834>.
- [15] J. R. Shewchuk, "An introduction to the conjugate gradient method without the agonizing pain, <http://www.cs.cmu.edu/quake-papers/painless-conjugate-gradient.pdf>, 1994.