Joint Array Diagnosis and Channel Estimation for RIS-Aided mmWave MIMO System

BINRUI LI[1], **ZHONGPEI ZHANG**[1], **(Member, IEEE), ZHENZHEN HU**[2], **AND YANG CHEN**[1]

[1]National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China
[2]College of Communication Engineering, Chengdu University of Information Technology, Chengdu 610225, China

Corresponding author: Zhongpei Zhang (zhangzp@uestc.edu.cn)

ABSTRACT In this paper, we consider a reconfigurable intelligent surface (RIS) aided millimeter wave (mmWave) multiple-input multiple-output (MIMO) system. The system can obtain the huge gain via joint active beamforming at the base station (BS) and passive beamforming at the RIS. However, due to weather and atmospheric effects, outdoor RIS antenna elements are subject to full or partial blockages from a plethora of particles like dirt, salt, ice, and water droplets. These blockages can cause an approximate squared power/SNR loss for the system. Different from the conventional array diagnosis, the RIS has no signal processing capability. Thus, we propose the joint array diagnosis and channel estimation techniques containing two stages to solve the problem. At the first stage the channel parameters at user equipment (UE) and BS are estimated using an iterative reweighted (IR) method. At the second stage, the array blockage coefficient vector and the effective sparse channel parameters at RIS are jointly estimated via solving a two-timescale non-convex optimization problem. We propose two algorithms, i.e., a batch algorithm (BA) and a two-timescale online joint array diagnosis and channel estimation (TOJADCE) algorithm to solve the problem and compare the performance of these two algorithms. Finally, to speed up the convergence of long-term variable and improve estimation performance, we propose a noise reduction (NR) algorithm. The simulations verify the effectiveness of our proposed algorithms.

INDEX TERMS mmWave MIMO system, reconfigurable intelligent surface, joint array diagnosis and channel estimation, two-timescale non-convex optimization.

I. INTRODUCTION

Millimeter wave (mmWave) multiple-input multiple-output (MIMO) system is a promising candidate technology for 5G communication system [1]. Reconfigurable intelligent surface (RIS), as an innovative technology, is considered to have great potential for cost-effectively improving the performance of mmWave MIMO system recently. The RIS is composed of reconfigurable and nearly passive reflecting antennas which are made of newly developed metamaterials. Different from traditional reflecting surfaces with only fixed phase shifters, each element of RIS can independently reflect incident signals with reconfigurable amplitudes and phase shifts by control of a smart micro controller. In many previous literatures, the passive beamforming has been recommended

The associate editor coordinating the review of this manuscript and approving it for publication was Mohammad Zia Ur Rahman.

to enhance the performance of system. In [2], it is verified that the proposed joint active and passive beamforming method can offer significant performance improvement comparing to the traditional massive MIMO. Considering the power consumption of the RIS in practical systems, a self-sustainable RIS system is proposed where the self-sustainability is realized by harvesting wireless power with each RIS elements in [3]. To fully harvest the benefit of RIS-assisted massive MIMO system with passive beamforming, knowledge of channel state information (CSI) is an essential requirement [2], [3]. Though the CSI acquisition in the traditional mmWave MIMO has been well studied by many articles, such as [4], [5]. In [4], a common support is proposed to improve performance of compressive sensing (CS) channel estimation. In [5], a channel estimation algorithm is proposed based on effective channel in hybrid mmWave system and a performance analysis with hardware impairments for channel

IEEE *Access*

estimation is discussed. Unfortunately, channel estimation is still challenging for the RIS-aided MIMO system since there is no signal processing capability at the RIS. Channel estimation for RIS-aided MIMO system has been recently studied in many previous works. In [6], [7], they proposed a least squares (LS) estimator based on/off method through switching groups of RIS elements on and off. To reduce the training overhead, CS techniques and deep learning methods were proposed in [8]. The parallel factor decomposition method was presented for RIS-aided multi-user MISO system in [9]. Besides, The reference [10] has presented a two-stage method, including sparse matrix factorization and matrix completion stage, to complete channel estimation. Recently, the channel estimation problem was divided into two subproblems and the authors applied an iterative reweighted (IR) method to find the estimates of the channel parameters sequentially in [11].

In the above works, the ideal hardware is assumed. However, in practice, the hardware impairments such as phase noise, quantization errors, In-phase/Quadrature imbalance, amplifier non-linearities and antenna array blockages are inevitable and lead to a serious degradation in signal processing performance [5], [12]–[14]. Specially, due to weather and atmospheric effects, outdoor mmWave antenna elements are subject to blockages from flying debris or particles found in the air [15]–[17]. It is noteworthy that the term blockage here represents a physical object partially or completely blocking a subset of antenna elements and is different from the mmWave channel blockage obstructed by the obstacles which will be appeared later. Partial or complete blockage of some of the antenna elements reduces the amount of energy incident on the antenna [18]. The water droplets, thin water film, snowflakes, dry and damp sand particles, and ice stones on the antenna array are reported to cause attenuation and a phase shift on mmWave signals. Therefore it is important to monitor the antenna array of mmWave MIMO system via the array diagnosis techniques, which are capable of detecting the blocked antennas and the corresponding signal power loss and phase shifts caused by the blocking particles. Several array diagnosis techniques, which are based on genetic algorithms [19], exhaustive search [20], MUSIC [21] and convex optimization [22], have been proposed in the literatures to identify the locations of faulty antenna elements. In the case of large antenna arrays, these techniques involve a large number of samples to achieve reliable results. Therefore, compressed sensing (CS) has recently been proposed in [23]–[25] for diagnosing large antenna arrays using reduced number of samples. The group CS approach in [26], [27] was proposed to estimate the locations of the blocked antennas and the induced the complex blockage coefficients including attenuation and phase shifts jointly.

In the conventional mmWave MIMO system, all the array diagnosis techniques are designed for array at the BS. However, in the RIS-aided mmWave MIMO system, we have to pay attention to the RIS, since the RIS plays a more important role and it is more likely to be blocked. Particularly, according

to the investigation of [28], an asymptotic receive signal power or signal-to-noise ratio (SNR) gain in the order of the squared of the number of antennas can be obtained via joint active beamforming at the BS and passive beamforming at the RIS due to the fact that the RIS combines the functionalities of both receive and transmit arrays for energy harvesting and reflect beamforming, respectively. Therefore, in the presence of complete blockages, the system can suffer a squared power/SNR loss with the number of complete blockages antennas. Moreover, considering that the outdoor RIS is usually deployed on the external wall of high buildings without additional protection, the RIS would suffer the precipitation, snowflakes, dry and damp sand particles, and ice stones, like the external wall, and the blockage caused by them. Unfortunately, to the best of our knowledge, no literature has been proposed to study the array diagnosis for RIS. Meanwhile, the conventional array diagnosis techniques, such as [26], [27], which have to work in the array under test (AUT) model, can not be applied to the RIS. The AUT model requires that the BS transmits a sequence of pilot to a reference array for array diagnosis, where the perfect CSI between them is assumed. However the RIS only reflects signals by a certain phase shift and has no signal processing capability. Hence the pilot has to transmit by the BS or the reference UE. Moreover, to apply such array diagnosis techniques to RIS, the BS-RIS channel and RIS-UE channel information need to be known. However, in practice, the CSI acquisition for the BS-RIS-UE cascaded channel is still a open and difficult issue, due to no signal processing capability for RIS. Therefore, motivated by these reasons, in this paper, comparing to existing works [26], [27], we propose a new algorithm to avoid the AUT model and complete the array diagnosis and channel estimation jointly.

In this paper, we focus on the joint array diagnosis and channel estimation problem for RIS-aided mmWave MIMO system. we divide the problem into two stages, where a standard line spectrum estimation formulation and a two-timescale line spectral estimation formulation are proposed to the first stage and second stage, respectively. In the first stage, an IR algorithm is proposed to estimation of effective channel parameters for the BS-RIS-UE cascaded channel. The two-timescale optimization approaches in the second stage are proposed to jointly complete array diagnosis and remaining channel parameters estimation jointly. Further, we propose a noise reduction algorithm to speed up the convergence of the two-timescale optimization approaches and improve the estimation performance. The main contributions of the paper can be summarized as follows:

- comparing with research about the hardware impairments for mmWave system in the existing works, such as [5], [12]–[14], We specially investigate the effects of random blockages on the signal power/SNR gain at the RIS and provide a metric to efficiently evaluate the impact of the blockage probability and the maximum amplitude absorption value on the RIS power/SNR gain.
- Different from the ideal hardware assumption for the RIS-aided mmWave MIMO system in [2], [3], [6]–[11],

we consider the RIS-aided mmWave MIMO system in the presence of the blockages at the RIS. For channel estimation problem in the system, we formulate it as joint antenna diagnosis and channel estimation. The new formulation, which is a two-timescale non-convex optimization formulation, can eliminate the dependence of traditional antenna diagnosis algorithms proposed in [26], [27] on AUT model.

- We propose the batch algorithm (BA) to solve the challenging two-timescale non-convex optimization problem, which can identify the locations and the induced attenuation and phase shifts caused by blockages and simultaneously implement channel estimation. Next, a two-timescale online joint array diagnosis and channel estimation (TOJADCE) algorithm is also proposed to reduce the computational complexity and memory cost over the batch algorithm.

- We exploit the characteristic of blockage vector to design a Noise Reduction (NR) algorithm, which speeds up the convergence of long-term variable estimation and improves estimation performance with a little additional complexity.

The rest of the paper is organized as follows. In Section II, we introduce the system model and channel model. In Section III, we formulate the joint array diagnosis and channel estimation as a two-stage estimation problem. In Section IV, we present IR algorithm for first stage. Then, the BA and TOJADCE algorithm are proposed for second stage, respectively. Finally, the NR algorithm is proposed. The analysis and convergence proof of the all proposed algorithms involved in this paper are discussed in Section V. In Section VI, we provide simulation results. And conclusions are drawn in Section VII.

Notations: We use the following notations throughout this paper: A is a matrix and a is a vector. $|a|, \|a\|$, $a(k)$ are the cardinality, 2-norm, k-th element of a, respectively. Whereas A^T, A^H, A^*, $[A]_{a,b}$ are its transpose, Hermitian (conjugate transpose), conjugate, entry at the a-th row and b-th column respectively. $|a|$ represents the absolute value of a. I is the identity matrix. $\mathcal{CN}(m,n)$ is a complex Gaussian random vector with mean m and covariance n. diag(a) is the diagonal matrix created from vector a. $\lceil \cdot \rceil$ represents the ceiling function. $a \odot b$ represents the element-wise product of a and b.

II. SYSTEM MODEL
A. TRANSMISSION WITH RIS BLOCKAGE
We consider a narrowband mmWave MIMO system as illustrated in Fig. 1 where the BS is equipped with M antennas to serve the UE equipped with N antennas and one RF chain. Due to the high cost and power consumption suffered by the fully-digital beamformer/combiner, the fully-connected hybrid digital is assumed to be equipped in the BS similar to [29]–[31]. Meanwhile since the high pathloss during mmWave propagation, the mmWave MIMO system is considered to be sensitive to LoS-blocking obstacles [11]. Thus in order to overcome this, the RIS consisting of L reflecting
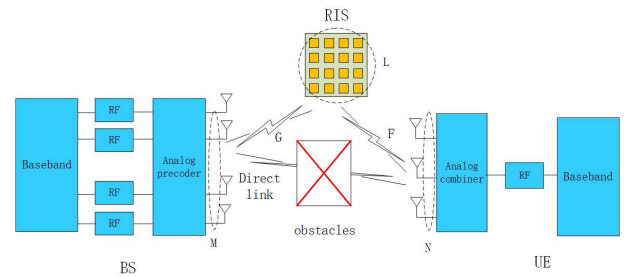

FIGURE 1. RIS-aided mmWave MIMO system.

elements is deployed to assist communication between the BS and UE in mmWave MIMO system. Specifically, it is noted that the channel consists of two parts, direct link and reflecting link, since the RIS is additionally installed at a height to enhance the signal coverage and communication quality by providing a tunable strong reflecting antenna array. Noticeably, the direct link and reflecting link are independent of each other. Therefore many previous works, such as [30], [31], have handled issues related to the direct link, which can be estimated via traditional channel estimation methods when the RIS is turned off. The direct link can be cancelled from the channel model during reflecting link channel training stage. Besides, the direct link often suffers from LoS-blocking obstacles and the remaining NLoS components are too weak to provide an effective communication path between BS and UE. Hence due to the above two reasons, the reflecting link will be the only focus in this paper.

For the reflecting link, each reflecting element of the RIS can dynamically adjust the reflecting signal with a reconfigurable phase shift via a smart controller in a passive way. Therefore the reflecting link channel is modelled as a cascaded channel and can be expressed as

$$H = FSG, \qquad (1)$$

where $F \in \mathbb{C}^{N \times L}$ and $G \in \mathbb{C}^{L \times M}$ denote the channel from RIS to UE (RIS-UE) and the channel from BS to RIS (BS-RIS), respectively. And S is defined as diag(s), where $s \triangleq [\beta_1 e^{j\theta_1}, \ldots, \beta_L e^{j\theta_L}]^T \in \mathbb{C}^L$ represents the phase-shift vector of the RIS. In s, θ_l and $\beta_l \in \{0, 1\}$ denote the phase shift and on/off state associated with the l-th passive element of the RIS, respectively.

For an ideal RIS-aided wireless system with a single user, an asymptotic receive signal power or signal-to-noise ratio (SNR) gain in the order of $\mathcal{O}(L^2)$ can be obtained via joint active beamforming at the BS and passive beamforming at the RIS [28]. Such a squared power/SNR gain is larger than that of massive MIMO, i.e., $\mathcal{O}(L)$, due to the fact that the RIS combines the functionalities of both receive and transmit arrays for energy harvesting and reflect beamforming, respectively. In addition, with conventional MIMO relays (even assuming their full-duplex operation with perfect self-interference cancellation), the SNR at the user receiver increases with the number of active antennas, L, with $\mathcal{O}(L)$
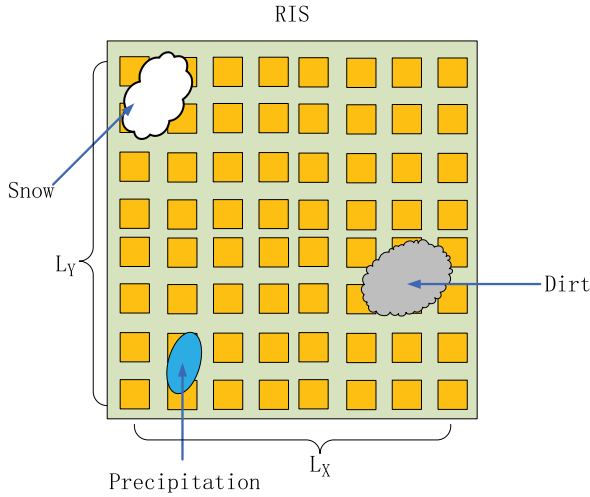
FIGURE 2. An example of an outdoor millimeter wave RIS with different suspended particles partially blocking the array.

FIGURE 3. The RIS power/SNR gain versus blockage probability P_b with different κ.

due to the noise effect at the relay, which is also lower than $\mathcal{O}(L^2)$ of the RIS due to its full-duplex and noise-free signal reflection [28].

However, due to weather and atmospheric effects, outdoor elements of RIS are subject to blockages from flying debris or particles found in the air as shown in Fig. 2. Meanwhile a physical object partially or completely blocking a subset of antenna elements can be caused by a thin water film, snowflakes, ice stones, and dry and damp sand particles on the antenna array [26]. In the presence blockages of RIS, the reflecting link channel can be expressed as

$$H = FESG, \qquad (2)$$

where $E = \text{diag}(e)$ represents the blockage coefficient matrix and the l-th entry of the blockage coefficient vector e is defined as

$$e_l = \begin{cases} \alpha_l, & \text{if the } l\text{-th element is blocked} \\ 1, & \text{otherwise,} \end{cases} \qquad (3)$$

where $\alpha_l = \kappa_l e^{j\Phi_l}$, in which $0 \leq \kappa_l \leq \kappa$ and $0 \leq \Phi_l \leq 2\pi$), are the resultant complex blockage coefficients at the l-th element. A value of κ_l represents amplitude absorption at the l-th element, where κ is the possible maximum value of κ_l, and the scattering coefficient Φ_l measures the phase-shift caused by the particle suspended on the l-th element. Specially, $\kappa_l = 0$ represents the complete blockage at the l-th element. This makes e_l a random variable, i.e. $e_l = \alpha_l$ with probability P_b if the l-th antenna is blocked and $e_l = 1$ with probability $1 - P_b$ otherwise.

For the RIS-aided mmWave MIMO system in the azimuth/elevation direction with AOA of incident signal, ϕ_A/ψ_A, and the azimuth/elevation AOD of reflected signal, ϕ_D/ψ_D at the RIS, the practice signal power/SNR gain can be defined as

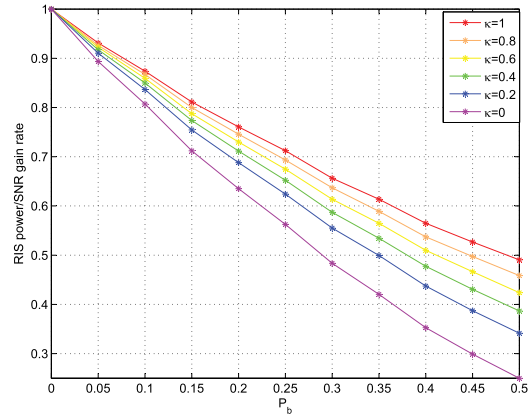$$\gamma = |a_i^H(\phi_D, \psi_D)ESa_i(\phi_A, \psi_A)|^2 \qquad (4)$$

where a_i denotes the antenna array response vector of the RIS. It is assumed that the UPA is equipped at the RIS as shown in Fig. 2. There are L_x equally spaced elements along the x-axis and L_y equally spaced elements along the y-axis at the RIS, BS and UE, respectively. Each antenna element is described by its position along the x and y axis. Thus in the case of the UPA, the array response vector can be written as

$$\begin{aligned} &\mathbf{a}_{\text{UPA}}(\phi, \psi) \\ &= \frac{1}{\sqrt{L_x L_y}} \Big[1, \cdots, e^{j\frac{2\pi}{\lambda}d(l_x \sin(\phi)\cos(\psi) + l_y \sin(\psi))}, \\ &\quad \cdots, e^{j\frac{2\pi}{\lambda}d((L_x-1)\sin(\phi)\cos(\psi) + (L_y-1)\sin(\psi))} \Big]^T. \end{aligned} \qquad (5)$$

From Eq. (4), we can find that power/SNR gain is the function of maximum amplitude absorption and the probability of blockages. From Fig. 3, we observe that power/SNR gain decreases with the probability P_b under different maximum amplitude absorption value. In Fig. 3, the RIS power/SNR gain rate is defined as γ^*/L^2, where $\gamma^* = \sum_{l=1}^{L} |e_l|^2$ represents the maximum of γ under the optimal RIS configuration proposed in [2]. Specially, in the presence of complete blockages, the system can suffer a squared power/SNR loss with the number of complete blockages antennas. Therefore, it is important to continuously monitor the RIS, reveal any abnormalities, and take corrective measures to maintain efficient operation of the system. However, since the RIS has no signal processing capability, the AUT model can not work in the conventional diagnosis methods as mentioned in the previous section. Hence, we propose a method for joint array diagnosis and channel estimation in the RIS-aided mmWave MIMO system.

Before we proceed with the proposed techniques, we list the challenges and lay down some assumptions for the system firstly. The main challenges come from two aspects as follow:

- (i) compared with the conventional array diagnosis techniques, such as [26], the AUT model can not be implemented in the RIS-aided mmWave MIMO system since

the RIS only reflects signals by a certain phase shift and has no baseband modules to transmit and receive signals.

- (ii) Since a lot data are needed to deal with during continuously monitoring the RIS, it can induce a huge computational complexity and memory cost.

According to reference [26], [27], we make some assumptions as follow:

- (i) The number of blockages is assumed to be small compared to the array size.
- (ii) the channel suffers from block fading, i.e., the channel changes independently per block.
- (iii) the blockage coefficient vector e remains constant for B blocks, which is as large as dozens or even hundreds, since the weather and atmospheric circumstance at the RIS always vary in a much slower process compared with the channel.

B. SPARSE CHANNEL MODEL

The narrowband geometric channel model is adopted to characterize both the BS-RIS channel G and the RIS-UE channel F similar to [11], [29], [32].

$$G = \sum_{c=0}^{C^g} \alpha_c^g a_i(\phi_c^{g,i}, \psi_c^{g,i}) a_b^H(\phi_c^{g,b}, \psi_c^{g,b}), \quad (6)$$

where C^g is the total number of paths including the LOS path. The α_0^g, $\phi_0^{g,i}/\psi_0^{g,i}$ and $\phi_0^{g,b}/\psi_0^{g,b}$ are the complex path gain, azimuth/elevation AoD and AoA of the LOS path, respectively. The α_c^g, $\phi_c^{g,i}/\psi_c^{g,i}$ and $\phi_c^{g,b}/\psi_c^{g,b}$ are the complex path gain, azimuth/elevation AoD and AoA of the c-th NLOS path, respectively. Additionally, $\alpha_0^g \sim \mathcal{CN}(0, 1)$ and $\alpha_c^g \sim \mathcal{CN}(0, 10^{-\mu})$, where μ is the power distribution ratio of LoS to NLoS multipath components. The $a_i(\phi_c^{g,i}, \psi_c^{g,i})$ and $a_b(\phi_c^{g,b}, \psi_c^{g,b})$ denote the antenna array response vector of the RIS and BS, respectively.

For the RIS-UE channel F, it can be similarly modeled as

$$F = \sum_{d=0}^{D^f} \alpha_d^f a_u(\phi_d^{f,u}, \psi_d^{f,u}) a_i^H(\phi_d^{f,i}, \psi_d^{f,i}), \quad (7)$$

where D^f is the number of paths. The α_0^f, $\phi_0^{f,i}/\psi_0^{f,i}$ and $\phi_0^{f,u}/\psi_0^{f,u}$ are the complex path gain, azimuth/elevation AoD and AoA of the LOS path, respectively. The α_d^f, $\phi_d^{f,i}/\psi_d^{f,i}$ and $\phi_d^{f,u}/\psi_d^{f,u}$ denote the complex path gain, azimuth/elevation AoD and AoA of the d-th path, respectively. Besides, $\alpha_0^f \sim \mathcal{CN}(0, 1)$ and $\alpha_d^f \sim \mathcal{CN}(0, 10^{-\mu})$, where μ is the power distribution ratio of LoS to NLoS multipath components. The $a_u(\phi_d^{f,u}, \psi_d^{f,u})$ and $a_i^H(\phi_d^{f,i}, \psi_d^{f,i})$ denote the antenna array response vector of the UE and RIS, respectively.

Similar to the RIS, the BS and UE are assumed to equip the UPAs. There are M_x, N_x and M_y, N_y equally spaced elements along the x-axis and y-axis at BS and UE, respectively. Therefore, their antenna array response vector can refer to the UPA response vector as Eq. (5).

Meanwhile, since the LoS path is typically much stronger than the NLoS paths according to the field measurements

in [11], we ignore the NLoS paths in the BS-RIS and RIS-UE links. By substituting Eq. (6) and (7) into Eq.(2), we can approximate the cascaded channel H as

$$H \approx g_0 a_u(\phi_0^{f,u}, \psi_0^{f,u}) a_b^H(\phi_0^{g,b}, \psi_0^{g,b}), \quad (8)$$

where g_0 denotes the effective gain, which is defined as

$$g_0 = \alpha_0^g \alpha_0^f a_i^H(\phi_0^{f,i}, \psi_0^{f,i}) ES a_i(\phi_0^{g,i}, \psi_0^{g,i})$$

which is able to be configured via adjusting the phase-shift vector of RIS s.

C. EFFECTS OF BLOCKAGES ON SYSTEM PERFORMANCE

Based on the channel model, during the downlink data transmission, signal $x \in \mathbb{C}^{M \times K}$ is send by a beamforming vector $z \in \mathbb{C}^{M \times 1}$ at the BS, where K denotes the data length. The transmission signal is reflected by the RIS with configuring the phase-shift vector S and received at the UE through a combining vector $W \in \mathbb{C}^{N \times 1}$. Therefore the received signal at the UE is given by

$$\begin{aligned} y &= w^H H z x + w^H n_0 \\ &= w^H g_0 a_u(\phi_0^{f,u}, \psi_0^{f,u}) a_b^H(\phi_0^{g,b}, \psi_0^{g,b}) z x + w^H n_0 \quad (9) \end{aligned}$$

where $n_0 \in \mathbb{C}^{N \times 1}$ denotes the noise at the receiving antennas.

Based on the transmission expression, the ergodic achievable downlink rate of system can be expressed as [2]

$$\bar{R} = \mathbb{E}\left\{ \log_2 \left\{ 1 + \frac{|g_0|^2 \beta_{BS} \beta_{UE} P_s}{\sigma_{UE}^2} \right\} \right\} \quad (10)$$

where $\beta_{BS} = |a_b(\phi_0^{g,b}, \psi_0^{g,b})^H z|^2$ and $\beta_{UE} = |a_u(\phi_0^{f,u}, \psi_0^{f,u})^H w|^2$ denote the array gain at the BS and MS, respectively. Besides, the $P_s = \frac{E_s}{M}$ and σ_{UE}^2 represents the average transmitted power at the BS and the effective noise variance at the UE, where E_s denotes the total transmitted power. In the presence blockages of RIS and the large numbers of RIS antennas regime, i.e., $L \to \infty$, a perfect CSI including blockages matrix of RIS is assumed at the BS and UE. Thus, $\beta_{BS} = M^2$ and $\beta_{UE} = N^2$ can be obtained with beamforming and combining vector $z = a_b(\phi_0^{g,b}, \psi_0^{g,b})$ and $w = a_u(\phi_0^{f,u}, \psi_0^{f,u})$, respectively [2]. Therefore, for the case of optimal RIS configuration, the achievable downlink rate can be obtained as

$$\begin{aligned} \bar{R}_1 &= \mathbb{E}\left\{ \log_2 \left\{ 1 + \frac{|\alpha_0^g|^2 |\alpha_0^f|^2 \gamma \beta_{BS} \beta_{UE} P_s}{\sigma_{UE}^2} \right\} \right\} \\ &= \log_2 \left\{ 1 + \frac{\frac{\pi^2}{16}|(1 - P_b)L + P_b L \mathbb{E}\{|\alpha_l|\}|^2 MN^2 E_s}{\sigma_{UE}^2} \right\}, \end{aligned} \quad (11)$$

where $\mathbb{E}\{|\alpha_l|\}$ denotes the expectation of $|\alpha_l|$, since $|\alpha_0^g|$ and $|\alpha_0^f|$ are statistically independent and both follow Rayleigh distribution with mean values $\frac{\sqrt{\pi}}{2}$.

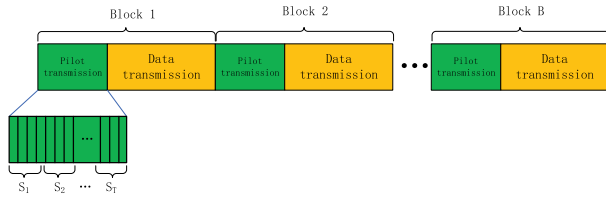Without the information of blockages matrix of RIS at the BS, the phase of blockages element can not be corrected,

FIGURE 4. A typical structure of the virtual angular domain representation of the effective channel and its vectorization.

which can be viewed to suffer a random phase shifts configuration as [2]. Therefore, the achievable downlink rate can be obtained as

$$\bar{R}_2 = \log_2 \left\{ 1 + \frac{((1 - P_b)L + P_b L \mathbb{V}\{\alpha_l\})MN^2 E_s}{\sigma_{UE}^2} \right\}, \quad (12)$$

where $\mathbb{V}\{\alpha_l\}$ denotes the variance of α_l. By utilizing the Lindeberg-Levy central limit theorem, similar derivation of results can be found in [2] and are omitted for space limitation. It is worthy to noticed that the achievable downlink rate of system reduce significantly suffering to the blockage of RIS.

Comparing between Eq. 11 and Eq. 12, it is worthy to noticed that the achievable downlink rate of system reduce significantly when the system suffer the blockage of RIS. Specially, the system can suffer a squared power/SNR loss with the number of complete blockages antennas. In addition, it is worthy to noticed that power allocation is ignored during the derivation the rate expression, since the elements of RIS are passive devices which the energy of reflected signal is restricted to be no more than the one of incident signal. Therefore new calibration method design is required for the RIS the RIS-aided mmWave MIMO system, which is not discussed in this paper, but will be our the future work.

III. PROBLEM FORMULATION

Based on the system model and channel model above, we consider a pilot-aid approach to simplify joint array diagnosis and channel estimation. In this process, one channel coherence block is divided into two phases, the first one for transmitting pilot to joint array diagnosis and channel estimation and the second for data transmission as depicted in Fig. 4. Besides, the first phase of block b is further divided into T subblocks. In each subblock $t = 1, \ldots T$, the BS sends a training matrix $\mathbf{Z}_{b,t} \in \mathbb{C}^{M \times P_z}$, which is reflected by the RIS with configuring the phase-shift vector $\mathbf{S}_{b,t}$ and received at the UE through a combining matrix $\mathbf{W}_{b,t} \in \mathbb{C}^{N \times P_w}$. The P_z and P_w represent the sequence length for training matrix and combining matrix, respectively. Thus, the received signal at the UE is given as

$$\mathbf{Y}_{b,t} = \mathbf{W}_{b,t}^H \mathbf{H}(\mathbf{S}_{b,t})\mathbf{Z}_{b,t} + \mathbf{W}_{b,t}^H \mathbf{N}_{b,t} \quad (13)$$

where $\mathbf{N}_{b,t}$ denotes the noise in the receiving antenna and $\mathbf{H}(\mathbf{S}_{b,t})$ is considered as a function of $\mathbf{S}_{b,t}$, which is defined

---

Algorithm 1 Two-Stage Algorithm for Joint Array Diagnosis and Channel Estimation

1: **First Stage:**
2: **Input:** Received signals $\mathbf{Y}_{b,t}$, combining matrices $\mathbf{W}_{b,t}$ and training matrices $\mathbf{Z}_{b,t}$, $\forall t = 1, \ldots T$, $b = 1, \ldots B$.
3: Run the Algorithm 2
4: **Output:** $\hat{\mathbf{g}}_b, \phi_b^{f,u}, \psi_b^{f,u}, \phi_b^{g,b}$ and $\psi_b^{g,b}, \forall b = 1, \ldots B$.
5: **Second Stage:**
6: **Input:** the output of first stage $\hat{\mathbf{g}}_b$ and phase control matrices $\mathbf{S}_{b,t}, \forall t = 1, \ldots T$, $b = 1, \ldots B$.
7: Run the Algorithm 3 or Algorithm 4
8: **Output:** $\mathbf{h}_b, \forall b = 1, \ldots B$ and $\mathbf{e}$.

---

by Eq. (8) as

$$\mathbf{H}(\mathbf{S}_{b,t}) = g_{b,t}\mathbf{a}_u(\phi_b^{f,u}, \psi_b^{f,u})\mathbf{a}_b^H(\phi_b^{g,b}, \psi_b^{g,b}), \quad (14)$$

where

$$g_{b,t} = \alpha_b^g \alpha_b^f \mathbf{a}_i^H(\phi_b^{f,i}, \psi_b^{f,i})\mathbf{E}\mathbf{S}_{b,t}\mathbf{a}_i(\phi_b^{g,i}, \psi_b^{g,i}), \quad (15)$$

in which the subscript 0 is ignored for convenience.

Next, inspired by literature [11], we propose a two-stage approach to complete both array diagnosis and channel estimation, which is described in Algorithm 1. The details of the algorithm will be provided in the following subsection.

A. FIRST STAGE

In the first stage, we estimate the azimuth/elevation AoD of the BS-RIS link $\phi_b^{g,b}/\psi_b^{g,b}$, the azimuth/elevation AoA of the RIS-UE link $\phi_b^{f,u}/\psi_b^{f,u}$ and the effective propagation path gain of the reflection link $g_{b,t}$ based on the received signals $\mathbf{Y}_{b,t}$. Thus the problem in the first stage is formulated as

$$\min_{\phi_b^{g,b}, \psi_b^{g,b}, \phi_b^{f,u}, \psi_b^{f,u}, \mathbf{g}_b} \mathcal{G}(\phi_b^{g,b}, \psi_b^{g,b}, \phi_b^{f,u}, \psi_b^{f,u}, \mathbf{g}_b) \quad (16)$$

where

$$\mathcal{G} = ||\mathbf{g}_b||_0 + \sum_{t=1}^{T} \lambda ||\mathbf{Y}_{b,t} - \mathbf{W}_{b,t}^H \mathbf{H}(\mathbf{S}_{b,t})\mathbf{Z}_{b,t}||_F^2$$

and $\mathbf{g}_b = [g_{b,1}, \ldots, g_{b,T}]^T$, $||\mathbf{g}_b||_0$ denotes the 0-norm of $\mathbf{g}_b$ which is equal to the number of the nonzero components of $\mathbf{g}_b$, λ controls the trade-off between the sparsity and data fitting. For conventional CS techniques, the continuous angel space has to be discretized into a finite set of grid points. The unknown angle components are assumed to lie on some of the discretized grid points. However, in practice, the true parameters do not necessarily lie on the discretized grid. Thus, the recovery signal suffers from the power leakage originating from basis mismatching and interference from neighboring paths. The accuracy recovery signals is very useful for estimating the remaining parameters, which will be explained in subsequent section. Thus, here we adopt the same IR method as in [33], [34] to estimate all the azimuth/elevation AoD/AoA and path gains.

Using the IR method, Eq. (16) can be further formulated as

$$\min_{\phi_b^{g,b},\psi_b^{g,b},\phi_b^{f,u},\psi_b^{f,u},g_b} \mathcal{G}(\phi_b^{g,b},\psi_b^{g,b},\phi_b^{f,u},\psi_b^{f,u},g_b) \quad (17)$$

where

$$\mathcal{G} = \sum_{t=1}^{T} \ln(|g_{b,t}|^2 + \varepsilon) + \lambda||Y_{b,t} - W_{b,t}^H H(S_{b,t})Z_{b,t}||_F^2$$

and parameter ε is a positive parameter to ensure that the argument of $\ln(\cdot)$ is positive. In this step, the log-sum function is used to relax the 0-norm function, which accounts for the fact that due to the random generation of $S_{b,t}$, some elements in g_b may be much smaller than others. Besides, for sparse signal recovery, the sparsity-inducing log-sum function term has superiority over the 1-norm penalty function.

B. SECOND STAGE
In the second stage, we complete the array diagnosis and remaining channel parameters estimation based on the final estimate of g_b obtained in the first stage, denoted as $\hat{g}_b$. Obtaining a separate estimate of the azimuth/elevation AoA for the BS-RIS link $\phi_b^{g,i}/\psi_b^{g,i}$ and the azimuth/elevation AoD for the RIS-MS link $\phi_b^{f,i}/\psi_b^{f,i}$ seems infeasible. The same holds for the propagation path gains in the BS-RIS and RIS-MS links α_b^g and α_b^f. Thus, we will estimate the product of the propagation path gains and the difference of azimuth/elevation directional trigonometric sines and cosines.

According to Eq. (15), we rewrite $g_{b,t}$ as

$$g_{b,t} = s_{b,t}^T E\alpha_b^g \alpha_b^f(a_i^*(\phi_b^{f,i},\psi_b^{f,i}) \odot a_i(\phi_b^{g,i},\psi_b^{g,i})) \quad (18)$$

where $s_{b,t} \in \mathbb{C}^{L \times 1}$ is the phase-shift vector configuration of the RIS at the subblock t of block b. By stacking the T different $s_{b,t}^T$ row-wise as $\Xi_b = [s_{b,1}, \ldots, s_{b,T}]^T$ and introducing $\alpha_b = \alpha_b^g \alpha_b^f$, $\phi_b^\Delta = \sin\left(\phi_b^{g,i}\right)\cos\left(\psi_b^{g,i}\right) - \sin\left(\phi_b^{f,i}\right)\cos\left(\psi_b^{f,i}\right)$ and $\psi_b^\Delta = \sin\left(\psi_b^{g,i}\right) - \sin\left(\psi_b^{f,i}\right)$, the Eq. (18) can be further expressed as

$$g_b = \Xi_b E\alpha_b(a_i^*(\phi_b^{f,i},\psi_b^{f,i}) \odot a_i(\phi_b^{g,i},\psi_b^{g,i}))$$
$$= \Xi_b E\alpha_b a(\phi_b^\Delta,\psi_b^\Delta) \quad (19)$$

where $a(\phi_b^\Delta,\psi_b^\Delta)$ follows the same form as Eq. (5).
Further, we assume that

$$\hat{g}_b = g_b + n_b$$
$$= \Xi_b E\alpha_b a(\phi_b^\Delta,\psi_b^\Delta) + n_b$$
$$= \Xi_b E h_b + n_b, \quad (20)$$

where the estimation error n_b from the first stage is modelled as additive Gaussian noise, independent of h_b. Besides $h_b = \alpha_b a(\phi_b^\Delta,\psi_b^\Delta)$, which can be considered as a channel with one angular sparsity. For fixed E and treating $\Xi_b E$ as the effective observation matrix, Eq. (20) is similar to the problem formulation with the line spectrum estimation. Besides, recalling that the channel changes rapidly with time while the blockage

coefficient vector e varies in a slower process. Thus, Given the observations of all B blocks $\hat{g}_b, \forall b = 1, \ldots, B$, the estimation problem can be formulated as a two-timescale line spectral estimation formulation and expressed as

$$\min_{e,\{h_b\}_{b=1}^B} \mathcal{F}(e,\{h_b\}_{b=1}^B)$$
$$s.t. \, ||e||_\infty \leq 1 \quad (21)$$

where

$$\mathcal{F} = \min_{e,\{h_b\}_{b=1}^B} \sum_{b=1}^{B} \ln(|\alpha_b|^2 + \epsilon) + \bar{\lambda}||\hat{g}_b - \Xi_b E h_b||_2^2$$

and $\bar{\lambda}$ is a parameter that controls the trade-off between the sparsity and data fitting. And ϵ is a positive parameter to ensure that the argument of $\ln(\cdot)$ is positive. Besides, $E = \text{diag}(e)$, and the value of e is constrained to Eq. (3).

Therefore our goal in the second stage is to recover the short-term signal h_b, and also learn the long-term unknown parameter e, from the measurement vectors $\hat{g}_b, \forall b = 1, \ldots, B$, at all the previous blocks. The two-timescale optimization approach is proposed for the joint array diagnosis and channel estimation.

IV. PROPOSED ALGORITHMS
A. IR ALGORITHM FOR FIRST STAGE
To solve (13), we resort to majorization-minimization (MM) method as [33], [34]. Due to that $\ln(|g_{b,t}|^2 + \varepsilon)$ is concave in $|g_{b,t}|^2$, we take its first order Taylor expansion as the surrogate function. Thus, the Eq. (17) can be further formulated as

$$\min_{\phi_b^{g,b},\psi_b^{g,b},\phi_b^{f,u},\psi_b^{f,u},g_b} \bar{\mathcal{G}}(\phi_b^{g,b},\psi_b^{g,b},\phi_b^{f,u},\psi_b^{f,u},g_b) \quad (22)$$

where

$$\bar{\mathcal{G}} = g_b^H G^{(i)} g_b + \sum_{t=1}^{T} \lambda||Y_{b,t} - W_{b,t}^H H(S_{b,t})Z_{b,t}||_F^2$$

and

$$G^i = \text{diag}([\frac{1}{|g_{b,1}^{(i)}|^2 + \varepsilon}, \ldots, \frac{1}{|g_{b,T}^{(i)}|^2 + \varepsilon}]^T),$$

$g_{b,t}^{(i)}$ is an estimate of $g_{b,t}$ at the i-th iteration. For $\forall t = 1, \ldots, T$, setting the first-order partial derivative of $\mathcal{G}(\phi_b^{g,b},\psi_b^{g,b},\phi_b^{f,u},\psi_b^{f,u},g_b)$ in Eq. (22) with respect to $g_{b,t}$ to zero yields

$$\hat{g}_{b,t}^{(i+1)} = (\frac{1}{\lambda(|g_{b,t}^{(i)}|^2 + \varepsilon)} + ||A_{b,t}Z_{b,t}||_F^2)^{-1} \sum_{p_z=1}^{P_z} z_{b,t,p_z}^H A_{b,t}^H y_{b,t,p_z} \quad (23)$$

where z_{b,t,p_z} and y_{b,t,p_z} is the p_z-th column of $Z_{b,t}$ and $Y_{b,t}$, respectively, and $A_{b,t} = W_{b,t}^H a(\phi_b^{g,b},\psi_b^{g,b})a(\phi_b^{f,u},\psi_b^{f,u})^H$.

Algorithm 2 Iterative Reweighted Algorithm

Input: Received signals $\boldsymbol{Y}_{b,t}$, combining matrices $\boldsymbol{W}_{b,t}$ and training matrices $\boldsymbol{Z}_{b,t}$, $\forall t = 1, \ldots T, b = 1, \ldots B$.

Output: $\hat{\boldsymbol{g}}_b, \phi_b^{f,u}, \psi_b^{f,u}, \phi_b^{g,b}$ and $\psi_b^{g,b}, \forall b = 1, \ldots B$.

1: **Initialize:** $\phi_b^{f,u}, \psi_b^{f,u}, \phi_b^{g,b}$ and $\psi_b^{g,b}$ for $i = 0$ by the SVD.

2: **repeat**

3:   Compute $\hat{g}_{b,t}^{(i+1)}$ by Eq. (23).

4:   Construct the function according to Eq. (24).

5:   Search for new angle estimates using the gradient descent method.

6: **until** The maximum number of iterations **or** stopping criterion reached.

Thus, for given $\hat{\boldsymbol{g}}_b^{(i+1)} = [\hat{g}_{b,1}^{(i+1)}, \ldots, \hat{g}_{b,T}^{(i+1)}]^T$, according to Eq. (23), the Eq. (22) can be written as

$$\min_{\phi_b^{g,b}, \psi_b^{g,b}, \phi_b^{f,u}, \psi_b^{f,u}, \boldsymbol{g}_b} \bar{\mathcal{G}}(\phi_b^{g,b}, \psi_b^{g,b}, \phi_b^{f,u}, \psi_b^{f,u}, \boldsymbol{g}_b) \quad (24)$$

where

$$\bar{\mathcal{G}} = \sum_{t=1}^{T} (-\beta_{b,t} x_{b,t}^H x_{b,t} + \sum_{p_z=1}^{P_z} \boldsymbol{y}_{b,t,p_z}^H \boldsymbol{y}_{b,t,p_z}),$$

$\beta_{b,t} = (\frac{1}{\lambda(|g_{b,t}^{(i)}|^2 + \varepsilon)} + ||\boldsymbol{A}_{b,t}\boldsymbol{Z}_{b,t}||_F^2)^{-1}$ and $x_{b,t} = \sum_{p_z=1}^{P_z} \boldsymbol{z}_{b,t,p_z}^H \boldsymbol{A}_{b,t}^H \boldsymbol{y}_{b,t,p_z}$.

An analytical solution of the the above optimization problem (24) is difficult to obtain. Since $\bar{\mathcal{G}}(\phi_b^{g,b}, \psi_b^{g,b}, \phi_b^{f,u}, \psi_b^{f,u}, \boldsymbol{g}_b)$ is differentiable for our case, we use a gradient descent algorithm to find estimates for the angles. The calculation of the required first-order partial derivatives associated with the angles can be calculated according to Appendix A. The initial values can be determined by the singular value decomposition (SVD) based on approach in [33] and parameters λ, ε can be adaptively updated according to [34]. To this end, the proposed IR algorithm is summarized in Algorithm 2.

B. JOINT ARRAY DIAGNOSIS AND CHANNEL ESTIMATION ALGORITHM FOR SECOND STAGE

In this section, to jointly estimate the long-term array blockage coefficient vector and the short-term channel vectors in the second stage, we propose a batch algorithm which optimizes the current parameter based on all received signals firstly. However, its computational complexity and memory cost increase over time. Next, we propose a TOJADCE algorithm, which can achieve a compromise between complexity and estimation accuracy.

1) BATCH ALGORITHM

Recalling the two-timescale optimization Eq. (21), the batch algorithm solve such a problem by optimizing $\boldsymbol{h}_b, \forall b = 1, \ldots, B$ and $\boldsymbol{e}$ alternately with all available observations. Therefore, in the $(i + 1)$-th iteration, given the long-term variable $\boldsymbol{e}^{(i)}$ and the short-term variables $\boldsymbol{h}_b^{(i)}, \forall b = 1, \ldots, B$,

the new estimate of short-term channel vectors can be obtained by solving a sequence of subproblems, respectively. Since each of the subproblems is similar to the line spectral estimation, we can follow the IR method to find the new estimate of $\boldsymbol{h}_b$. Thus, the objective function of subproblems can be expressed as

$$\min_{\boldsymbol{h}_b} \bar{\mathcal{F}}(\boldsymbol{h}_b|\boldsymbol{e}^{(i)}, \hat{\boldsymbol{g}}_b)$$
$$= \min_{\boldsymbol{h}_b} \frac{\alpha_b \alpha_b^H}{(\hat{\alpha}_b^{(i)})^2 + \epsilon} + \bar{\lambda}||\hat{\boldsymbol{g}}_b - \boldsymbol{\Xi}_b \boldsymbol{E}^{(i)} \boldsymbol{h}_b||_2^2, \quad \forall b = 1, \ldots, B$$
(25)

where $\hat{\alpha}_b^{(i)}$ is an estimate of α_b at the i-th iteration. Setting the derivative of the objective function $\bar{\mathcal{F}}(\boldsymbol{h}_b|\boldsymbol{e}^{(i)}, \hat{\boldsymbol{g}}_b)$ in Eq. (25) with respect to α_b to zero yields

$$\hat{\alpha}_b^{(i+1)} = \zeta_b \boldsymbol{a}(\phi_b^\Delta, \psi_b^\Delta)^H (\boldsymbol{E}^{(i)})^H \boldsymbol{\Xi}_b^H \hat{\boldsymbol{g}}_b \quad (26)$$

where

$$\zeta_b = (\frac{1}{\bar{\lambda}(|\alpha_b^{(i)}|^2 + \epsilon)} + ||\boldsymbol{a}(\phi_b^\Delta, \psi_b^\Delta)^H (\boldsymbol{E}^{(i)})^H \boldsymbol{\Xi}_b^H \boldsymbol{\Xi}_b \boldsymbol{E}^{(i)} \boldsymbol{a}(\phi_b^\Delta, \psi_b^\Delta)||_F^2)^{-1}.$$

For given $\hat{\alpha}_b^{(i+1)}$ and $\boldsymbol{e}^{(i)}$, the objective function can be further written as

$$\bar{\mathcal{F}}(\{\boldsymbol{h}_b\}|\boldsymbol{e}^{(i)}) = -\zeta_b \tau_b + \hat{\boldsymbol{g}}_b^H \hat{\boldsymbol{g}}_b \quad (27)$$

where $\tau_b = \boldsymbol{a}(\phi_b^\Delta, \psi_b^\Delta)^H (\boldsymbol{E}^{(i)})^H \boldsymbol{\Xi}_b^H \hat{\boldsymbol{g}}_b \hat{\boldsymbol{g}}_b^H \boldsymbol{\Xi}_b \boldsymbol{E}^{(i)} \boldsymbol{a}(\phi_b^\Delta, \psi_b^\Delta)$. Based on the first-order derivative in Appendix B, the gradient descent algorithm is applied to obtain a suboptimal $\phi_b^\Delta, \psi_b^\Delta$ via minimizing the objective function in Eq. (27).

Then the optimization of objective function associated with long-term variable $\boldsymbol{e}^{(i+1)}$ is given by

$$\min_{\boldsymbol{e}} \bar{\mathcal{F}}(\boldsymbol{e}|\boldsymbol{h}_b^{(i+1)}, \hat{\boldsymbol{g}}_b) = \min_{\boldsymbol{e}} \sum_{b=1}^{B} ||\hat{\boldsymbol{g}}_b - \boldsymbol{\Xi}_b \boldsymbol{E} \boldsymbol{h}_b^{(i+1)}||_2^2$$
$$s.t. ||\boldsymbol{e}||_\infty \leq 1 \quad (28)$$

The optimization problem (28) can be reformulated as

$$\min \sum_{b=1}^{B} ||\hat{\boldsymbol{g}}_b - \boldsymbol{\Upsilon}_b^{(i+1)} \boldsymbol{e}||_2^2$$
$$s.t. ||\boldsymbol{e}||_\infty \leq 1 \quad (29)$$

where $\boldsymbol{\Upsilon}_b^{(i+1)} = \boldsymbol{\Xi}_b \mathrm{diag}(\boldsymbol{h}_b^{(i+1)})$. It is obvious that the optimization problem (28) is converted to a quadratic minimization problem, which can be solve with standard convex optimization methods. Hence, the overall batch algorithm is summarized in Algorithm 3.

2) TOJADCE ALGORITHM

The batch algorithm requires to estimate all the previous short term variables in each iteration, which leads that the memory requirements and complexity become too high to accept over the number of blocks. Therefore, a more efficient online algorithm called TOJADCE is proposed in this section.

Algorithm 3 Batch Algorithm

Input: the output of first stage $\hat{g}_b$ and phase control matrices $S_{b,t}, \forall t = 1, \ldots T, b = 1, \ldots B$.

Output: $h_b, \forall b = 1, \ldots B$ and e.

1: **Initialize:** ϕ_b^Δ, ψ_b^Δ and $E = \text{diag}(e) = I_L$.
2: **repeat**
3:    **repeat**
4:       Compute $\hat{\alpha}_b^{(i+1)}$ by Eq. (26).
5:       Construct the function according to Eq. (27).
6:       Search for new angle estimates using the gradient descent method.
7:    **until** The maximum number of iterations **or** stopping criterion reached.
8:    Update e by solving the quadratic minimization problem Eq. (29) with standard convex optimization method.
9:    **optional noise reduction algorithm:** Update e according to Eq. (41).
10: **until** The maximum number of iterations **or** stopping criterion reached.

The key idea of TOJADCE algorithm is to find a stationary point of the original objective function Eq. (25) as $B \to \infty$ with much lower memory requirements and complexity. The details about TOJADCE algorithm is provided as follows.

According to [35], For large b, the original objective function Eq. (25) can converges to the following two-timescale stochastic non-convex optimization problem

$$\mathcal{P}: \min_{e,\Omega} \tilde{\mathcal{F}}(e, \Omega)$$
$$= \min_{e,\Omega} \mathbb{E}\left\{\ln(|\alpha|^2 + \epsilon) + \bar{\lambda}\|\hat{g} - \Xi E h\|_2^2\right\} \quad (30)$$

where $\mathbb{E}\{ \cdot \}$ term represents that the expectation is taken w.r.t. $h(\hat{g})$. Besides $\Omega = (h(\hat{g}), \forall \hat{g})$ is the collection of the short-term variables corresponding to all possible observations $\hat{g}$ with known pilot matrix Ξ. Problem $\mathcal{P}$ is a non-convex stochastic optimization problem in which both short-term and long-term variables require to estimate.

At first, using primal decomposition [35], the Problem $\mathcal{P}$ can be divided into a sequence of short-term subproblems and a long-term master problem. In the short-term subproblem, given the $\hat{g}$ and the estimation of long-term variable e, the effective line spectral estimation can be expressed as

$$\min_h \tilde{\mathcal{F}}(h|e, \hat{g}) = \min_h \ln(|\alpha|^2 + \epsilon) + \bar{\lambda}\|\hat{g} - \Xi E h\|_2^2 \quad (31)$$

Besides, long-term master problem focuses on the array diagnosis problem based on the stochastic optimal of h, and can be expressed as

$$\min_e \tilde{\mathcal{F}}(e|\Omega^*(e))$$
$$s.t. \|e\|_\infty \leq 1 \quad (32)$$

where $\Omega^*(e) = (h^*(e, \hat{g}), \forall \hat{g})$ and $h^*(e, \hat{g})$ is the solution of Eq. (31) for the different blocks.

It can be observed that the short-term subproblem is similar to Eq. (25). Thus using the iterative reweighted algorithm, we can achieve an online and estimation of h_b for each block b based on the current observations $\hat{g}_b$ and the estimation of the long-term variable $e_{(b-1)}$ at block $b - 1$. Therefore Eq. (31) is further formulated as

$$\min_{h_b} \tilde{\mathcal{F}}(h_b|e_{(b-1)}, \hat{g}_b)$$
$$= \min_{h_b} \frac{\alpha_b \alpha_b^H}{(\hat{\alpha}_b^{(i)})^2 + \epsilon} + \bar{\lambda}\|\hat{g}_b - \Xi_b E_{(b-1)} h_b\|_2^2 \quad (33)$$

where $\hat{\alpha}_b^{(i)}$ is an estimate of the i-th inner iteration at b-th outer iteration. The inner iteration and outer iteration represent iterative process in the IR algorithm and TOJADCE algorithm. Thus, we can follow the IR algorithm to find the estimate of h_b. Setting the derivative of the objective function $\tilde{\mathcal{F}}(h_b|E_{(b-1)}, \hat{g}_b)$ in Eq. (33) with respect to α_b to zero yields

$$\hat{\alpha}_b^{(i+1)} = \zeta_b a(\phi_b^\Delta, \psi_b^\Delta)^H E_{(b-1)}^H \Xi_b^H \hat{g}_b \quad (34)$$

where ζ_b is defined as

$$\left(\frac{1}{\bar{\lambda}(|\alpha_b^{(i)}|^2 + \epsilon)} + \|a(\phi_b^\Delta, \psi_b^\Delta)^H\right.$$
$$\left.(E_{(b-1)})^H \Xi_b^H \Xi_b E_{(b-1)} a(\phi_b^\Delta, \psi_b^\Delta)\|_F^2\right)^{-1}.$$

For fixed $\hat{\alpha}_b^{(i+1)}$, the objective function can be further written as

$$\tilde{\mathcal{F}}(h_b|e_{(b-1)}, \hat{g}_b) = -\zeta_b \tau_b + \hat{g}_b^H \hat{g}_b \quad (35)$$

where $\tau_b = a(\phi_b^\Delta, \psi_b^\Delta)^H E_{(b-1)}^H \Xi_b^H \hat{g}_b \hat{g}_b^H \Xi_b E_{(b-1)} a(\phi_b^\Delta, \psi_b^\Delta)$. Based on the first-order derivative in Appendix B, gradient descent method is applied to find a suboptimal ϕ_b^Δ, ψ_b^Δ via minimizing the objective function in Eq. (35).

For a long-term subproblem, at the b-th block, we design a recursive convex approximation $\mathcal{H}_b$ of the objective function in Eq. (32), which is updated by

$$\mathcal{H}_b = (1 - \rho_b)\mathcal{H}_{(b-1)} + \rho_b \tilde{\mathcal{F}}(e|h_b, \hat{g}_b), \quad (36)$$

where $\mathcal{H}_{-1} = 0$, $\rho_b \to 0$, $\sum_{b \to \infty} \rho_b = \infty$, $\sum_{b \to \infty}(\rho_b)^2 < \infty$. Besides, $\tilde{\mathcal{F}}(e|h_b, \hat{g}_b)$ is defined as $\|\hat{g}_b - \Upsilon_b e\|_2^2$, where $\Upsilon_b = \Xi_b \text{diag}(h_b)$. Then we obtain an intermediate variable $\tilde{e}_b$ by solving the problem:

$$\min_e \mathcal{H}_b(e)$$
$$s.t. \|e\|_\infty \leq 1 \quad (37)$$

Specifically, define

$$Q_b = (1 - \rho_b)Q_{(b-1)} + \rho_b(\Upsilon_b^H \Upsilon_b),$$
$$p_b = (1 - \rho_b)p_{(b-1)} + \rho_b(\Upsilon_b^H \hat{g}_b),$$
$$r_b = (1 - \rho_b)r_{(b-1)} + \rho_b(\hat{g}_b^H \hat{g}_b), \quad (38)$$

then, the objective function in Eq. (37) can be further formulated as

$$\mathcal{H}_b(e) = e^H Q_b e - p_b^H e - e^H p_b + r_b$$
$$= \bar{e}^H \bar{Q}_b \bar{e}$$

where $\bar{e} = [e; 1]$ and

$$\bar{Q}_b = \begin{bmatrix} Q_b & -p_b \\ -p_b^H & r_b. \end{bmatrix}$$

Thus Problem (37) is equivalent to the following real-valued optimization problem:

$$\min_{c} c^T \tilde{Q}_b c$$
$$s.t. \, c^T R_l c \leq 1, \forall l = 1, \ldots, L \qquad (39)$$

where

$$c = [\Re(\bar{e}); \Im(\bar{e})],$$
$$\tilde{Q}_b = \begin{bmatrix} \Re(\bar{Q}_b) & -\Im(\bar{Q}_b) \\ \Im(\bar{Q}_b) & \Re(\bar{Q}_b) \end{bmatrix}$$

and R_l is a $2L \times 2L$ matrix given by

$$R_l(i,j) = \begin{cases} 1, & \text{if } i = j = l \\ 1, & \text{if } i = j = l + L \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, problem (39) is a convex quadratically constrained quadratic program (QCQP), which can be solved efficiently using standard convex optimization methods. After that the intermediate variable $\tilde{e}_b = c^*(1 : L) + jc^*(L + 1 : 2L + 1)$ is obtained, where c^* is the optimum of problem (39). And then the long-term variable e_b is updated according to

$$e_b = (1 - \tau_b)e_{(b-1)} + \tau_b \tilde{e}_b, \qquad (40)$$

where τ_b satisfies $\tau_b \rightarrow 0$, $\sum_{b\rightarrow\infty} \tau_b = \infty$, $\sum_{b\rightarrow\infty}(\tau_b)^2 < \infty$.

The overall TOJADCE algorithm is summarized in Algorithm 4. In the next section, we will propose an improvement algorithm for long-term variable estimation problem.

C. NOISE REDUCTION ALGORITHM

For the two-timescale optimization approaches, how to speed up the convergence of long-term variable and estimation performance is one of the core problems. Due to the impact of noise, the estimated value of e deviates from the true one. Therefore the noise reduction algorithm is proposed. Base on the fact that the number of blockages is usually small compared to the array size, which implys that there are parts of the elements in e being equal to 1 and others being complex numbers with amplitude less than 1. However, the elements in the solutions of Eq. (29) and (40) are corrupted by the noise. Specilly, many elements in e whose true values are equal to 1 is reduced by a perturbation caused by the noise. Hence we propose a noise reduction algorithm based on the hard threshold to corrects the estimated value as

$$e_b(l) = \begin{cases} 1, & \text{if } e_b(l) > \iota \\ e_b(l), & \text{otherwise.} \end{cases} \qquad (41)$$

where ι is the hard threshold value and can be equal to the $\lceil \eta L \rceil$-th largest amplitude of elements in e. The proposed noise reduction algorithm can improve the performance of

---

Algorithm 4 TOJADCE Algorithm

Input: the output of first stage $\hat{g}_b$ and phase control matrices $S_{b,t}, \forall t = 1, \ldots T$ at block b.

Output: h_b and e_b at block b.

1: **Initialize:** ϕ_b^Δ, ψ_b^Δ and $E = \text{diag}(e) = I_L$.
2: **repeat**
3:     Compute $\hat{\alpha}_b^{(i+1)}$ by Eq. (34).
4:     Construct the function according to Eq. (35).
5:     Search for new angle estimates using the gradient descent method.
6: **until** The maximum number of iterations **or** stopping criterion reached.
7: Construct the recursive convex approximation function according to Eq. (36).
8: Update the intermediate variable $\tilde{e}_b$ by solving the problem Eq. (39) with standard convex optimization methods
9: Compute the long-term variable e_b according to Eq. (40)
10: **optional noise reduction algorithm:** Update e according to Eq. (41).

---

the two-timescale optimization approaches for joint array diagnosis and channel estimation with a little complexity. Therefore it is embedded in the step 9 of algorithm 3 and the step 10 of algorithm 4.

V. CONVERGENCE ANALYSIS AND DISCUSSION
A. CONVERGENCE ANALYSIS

The IR algorithm is adopted to complete the line spectral estimation in the first stage and second stage. As mentioned in the previous section, since the original estimation problem is a non-convex problem, the IR algorithm resorts to MM approach to iteratively approximate the non-convex objective function as a convex function with reweighted coefficients and solve the approximated convex problem. Specifically, in the i-th iteration of IR algorithm, the approximated function can be proved to be an upper bound to the original objective function according to [34]. Besides, for the process of estimating angles, a gradient descent algorithm can guarantee a monotonically decreasing objective function value. However, the convergence property of the IR algorithm still remains an open issue.

In the second stage, the batch algorithm is proposed to solve the the two-timescale optimization problem Eq. (25) firstly. According to [12], the batch algorithm can be considered as a special case of the block successive upper-bound minimization (BSUM) algorithm proposed in [36] to solve the original non-convex and non-differentiable objective function in Eq. (25). Besides, it can be proved that the local approximation satisfies all the conditions of Assumption 2 in [36]. Then using the conclusion of Theorem 2 in [36], every limit point of the iterates generated by the batch algorithm is a stationary point of Eq. (25).

We propose the TOJADCE algorithm to reduce the high computational complexity and memory cost of batch algorithm. The convergence of the TOJADCE algorithm has been established in the works in [35]. This subsection is to ensure the completeness, state the properties of the step sizes, and check whether the surrogate functions in this paper satisfy the convergence condition or not. For the short-term problem, the iterative reweighted algorithm is adopted to solve the line spectral estimation with a fixed e. It is required to satisfy the Assumption 2 in [35], i.e.

1) The initial point is $h_b^0(e)$ Lipschitz continuous on e.
2) the iterative reweighted algorithm is Lipschitz continuous on h for any inner iteration number.
3) For any fixed feasible e, the solution $h_b^*(e)\forall b = 1, \ldots, B$ converges to a stationary point of short-term problem.

Fortunately, the iterative reweighted algorithm can be consider as a extension to the MM Algorithm, which is verified as a typical example to satisfy Assumption 2. Therefore, for clarification, the proof procedure is omitted here. And readers who are interested in it can refer to [35]. Besides, for the long-term problem, we set the recursive step parameters according to the Assumption 5 in [35] to satisfy the following conditions:

1) $\mathcal{H}_{-1} = 0, \rho_b \to 0, \sum_{b\to\infty} \rho_b = \infty, \sum_{b\to\infty}(\rho_b)^2 < \infty$,
2) $\tau_b \to 0, \sum_{b\to\infty} \tau_b = \infty, \sum_{b\to\infty}(\tau_b)^2 < \infty$,
3) $\lim_{b\to\infty} \tau_b/\rho_b = 0$.

With the recursive step parameters setting like this, the convergence and properties of long-term surrogate functions can be guaranteed according to [35], i.e.,

$$\lim_{t\to\infty} |\mathcal{H}_b(e_b) - \tilde{\mathcal{F}}(e_b|\Omega^*(e_b))| = 0,$$
$$\lim_{t\to\infty} ||\nabla_e \mathcal{H}_b(e_b) - \nabla_e \tilde{\mathcal{F}}(e_b|\Omega^*(e_b))|| = 0. \quad (42)$$

Meanwhile, for all $b = 1, \ldots, B$, $\mathcal{H}_b(e)$ satisfies the following properties.

1) $\mathcal{H}_b(e)$ is uniformly strongly convex in e.
2) $\mathcal{H}_b(e)$ is Lipschitz continuous on e, and $|\mathcal{H}_{b_1}(e) - \mathcal{H}_{b_2}(e)| \leq B_f||e_{b_1} - e_{b_2}|| + \delta(b_1, b_2), \forall ||e||_\infty \leq 1$ for some constant $B_f > 0$, where $\lim_{b_1,b_2\to\infty} \delta(b_1, b_2) = 0$.
3) The derivative and second-order derivative of $\mathcal{H}_b(e)$ are uniformly bounded.

Proof: See Appendix C.

Based on the above conditions, for the large b, the solutions of short-term algorithm and long-term algorithm in Algorithm 4 converge to the stationary point of subproblems almost surely, respectively.

B. COMPUTATIONAL COMPLEXITY

Without loss of generality, we discuss the computational complexity of the proposed IR algorithm in first stage, batch algorithm and TOJADCE algorithm in second stage.

For the IR algorithm in first stage, the computational complexity in each iteration lies in calculating the gradient, which is equal to $\mathcal{O}(P_z P_w(M + N))$.

For the short-term problem in second stage, the batch algorithm solves $\{h_1, \ldots, h_b\}$ according to the iterative reweighted algorithm with b observations $\{\hat{g}_1, \ldots, \hat{g}_b\}$, while the TOJADCE algorithm only needs to solve one short-term problem with the current observation at b block. Since both the batch and TOJADCE algorithms utilize the same the IR algorithm to solve the short-term variables, it is reasonably assumed that they converge with the same iteration number. The computational complexity of the batch algorithm for short-term variables is $\mathcal{O}(bTL)$ per iteration, which is b times that of TOJADCE. For the long-term problem, the complexity of the batch algorithm is $\mathcal{O}(bT^2L)$, where the standard QCQP problem is solved by the CVX MATLAB tools. With the same MATLAB tools, the complexity of the TOJADCE algorithms is $\mathcal{O}(T^2L)$, since it only need to utilize the current observation at b block. It is worthy to noticed that there is b times relationship for the memory cost between the batch algorithm and TOJADCE algorithm.

VI. SIMULATION RESULTS

In this section, we conduct numerical simulations to evaluate the performance of the proposed techniques. The NMSE performance of IR algorithm in the first stage stage is similar to the first stage stage in [11]. Besides, it only provides the estimation of g_b for second stage and can be consider as an intermediate process independent of the second stage. Hence we will not show simulation and evaluation about IR algorithm in first stage and focus on the the NMSE performance in the second stage. And readers interested in IR algorithm can refer to the literature [11]. We let $L = 128, L_x = 16$ and $L_x = 8$ at the RIS. All of the AoA, AoD and the phase of elements of $S_{b,t}$ are randomly and uniformly distributed between 0 and 2π. Besides, the path gains α^l and ρ^d are drawn from distribution $\mathcal{CN}(0, 1)$. The product of the propagation path gains α_b follows $\mathcal{CN}(0, 1)$ and each element of n_b follows $\mathcal{CN}(0, \delta^2)$. The SNR is defined as $\frac{1}{\delta^2}$. For the random blockages, we set its probability and maximum absorption as $P_b = 0.4$ and $\kappa = 1$, respectively. $\rho_b = \frac{1}{b^{0.4}}$ and $\tau_b = \frac{1}{b^{0.65}}$ is adopted as an example in our simulation. However, an arbitrary ρ_b and τ_b meeting the conditions mentioned in the previous section can make the algorithm work well. The metric used for short-term subproblem performance comparison is the normalized mean square error (NMSE) of channel estimation which is defined as

$$\text{NMSE} = 10\log_{10} ||h_b - \hat{h}_b||^2 / ||h_b||^2, \quad (43)$$

and the metric used for long-term subproblem is NMSE of blockage coefficient vector estimation which is defined as

$$\text{NMSE} = 10\log_{10} ||e - \hat{e}||^2 / ||e||^2. \quad (44)$$

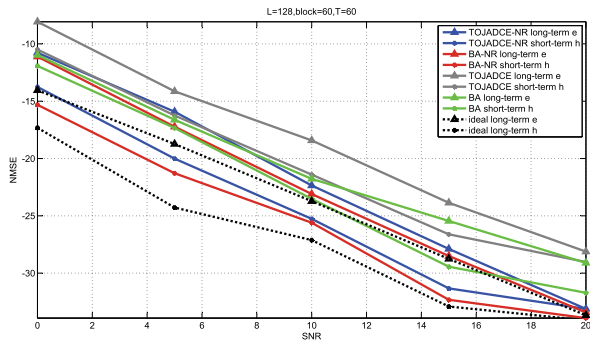In Fig. 5, we plot the performance for BA and TOJADCE algorithm with or without NR algorithm against SNR, where

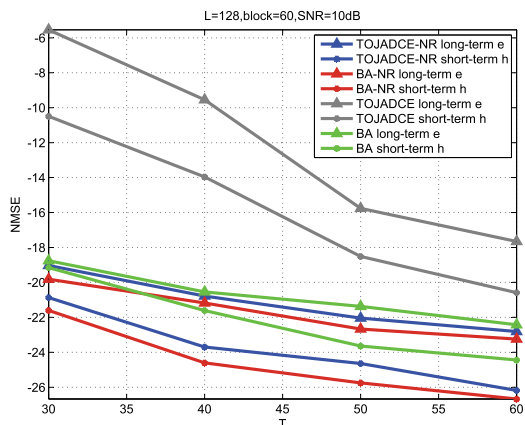**FIGURE 5.** NMSE versus SNR with $T = 60$ and $B = 60$.

FIGURE 6. NMSE versus pilot with $SNR = 10dB$ and $B = 60$.

FIGURE 7. NMSE versus block with $SNR = 10dB$ and $T = 60$.

Fig. 7 investigates the NMSE of short-term and long-term variable estimation for BA, BA-NR, TOJADCE and TOJADCE-NR algorithms against the number of blocks. The number of pilot is fixed to $T = 60$ and the SNR is chosen as $SNR = 10dB$. It can be observed that compared with TOJADCE and BA, TOJADCE-NR and BA-NR algorithms can converge to a stationary solution with a lower NMSE, respectively. The performances of all proposed algorithms are similar before the $B = 15$. when the number of block is larger than 15, the convergence rate of TOJADCE is lower than that of other algorithms. And the performance of TOJADCE-NR algorithm can approach or exceed that of batch algorithm with increase of blocks. Besides, it can be observed that performance of all the algorithms for the short-term variable suffers a subtle ups and downs. We infer that this is because in the inner iteration algorithm (IR algorithm), a gradient descent algorithm only guarantee a suboptimal angle estimation and its maximum number of iterations is fixed to 300.

VII. CONCLUSION

In this paper, we considered the effects of blockages on RIS in the RIS-aided mmWave MIMO system. We showed that the blockages of RIS can distort the power/SNR gain more heavily than the ones of BS array. Specially, in the presence of blockages, the system can suffer a squared power/SNR loss with the number of complete blockages antennas. Meanwhile, since the RIS only reflect signals by a certain phase shift and has no signal processing capability, the conventional array diagnosis methods can not work. Therefore we proposed a two-stage joint array diagnosis and channel estimation algorithm to simplify this problem. In the first stage, the iterative reweighted algorithm is adopted to estimate effective channel parameters for the cascaded channel. And then, we convert the remaining estimation problem to a two-timescale problem. The batch algorithm and TOJADCE algorithm is proposed to solve this problem based on the final estimate value of first stage. Specially, the batch algorithm solves both short-term and long-term problems

the number of pilot is $T = 60$ and the number of blocks is $B = 60$. Meanwhile the lower bound (ideal estimation) for short-term subproblem is obtained with the assumption that perfect information of blockage coefficient vector is known. Correspondingly, the lower bound (ideal estimation) for long-term subproblem can obtained under the condition that perfect information of channel all blocks is known. It is observed that the NMSE of short-term h and long-term e in the proposed BA and TOJADCE algorithm with or without NR algorithm decrease with the increase of SNR. Meanwhile we observe that the performance gap between BA and TOJADCE algorithm decreases as SNR increasing. Besides, the NR algorithm can improve the performance of BA and TOJADCE algorithm obviously. Specially, the BA-NR and TOJADCE-NR algorithm for long-term variable estimation can approach the lower bound at the high SNR.

In Fig. 6, we plot the performance against the number of pilot under the condition $SNR = 10dB$ and $B = 60$ with different algorithms. As the number of pilot increase, a better performance can be obtained by all proposed algorithms. And a obvious performance gap exits between BA and BA-NR and between TOJADCE and TOJADCE-NR for different number of pilot. However, the performance gap decrease as the number of pilot increase.
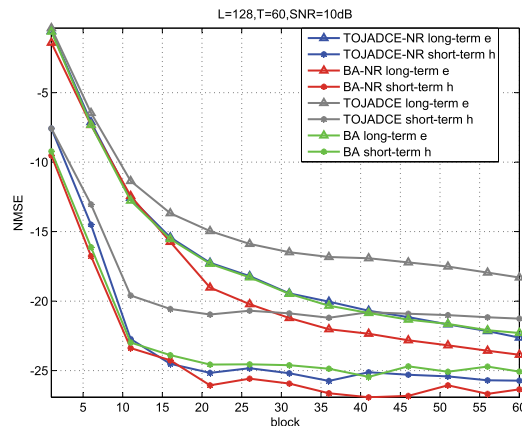
involving all the available samples in each iteration higher computational complexity and memory cost. For handling the shortcoming of the batch algorithm, TOJADCE algorithm is proposed to run a short-term algorithm only associated with the current observation, and updates the long-term variable according to the constructed surrogate function and the solution of the short-term problem. Simulations verify that our proposed algorithms can achieve preferable NMSE over various baselines. Besides, in order to speed up the convergence of long-term and estimation performance, we propose a NR algorithm, which is verified to improve the estimate performance with a little additional complexity by simulation results. Finally, base on the estimated CSI and array diagnosis results of RIS, the existing compensation methods for RIS-aided mmWave MIMO system can be investigated in future.

APPENDIX A
DERIVATIVES IN THE FIRST STAGE

The first-order partial derivative of $\bar{\mathcal{G}}(\phi_b^{g,b}, \psi_b^{g,b}, \phi_b^{f,u}, \psi_b^{f,u}, g_b)$ in Eq. (24) with respect to $\phi_b^{g,b}$ is written as

$$\frac{\partial \bar{\mathcal{G}}}{\partial \phi_b^{g,b}} = \sum_{t=1}^{T} -\frac{\partial \beta_{b,t}}{\partial \phi_b^{g,b}} x_{b,t}^H x_{b,t} - \beta_{b,t} \frac{\partial x_{b,t}^H}{\partial \phi_b^{g,b}} x_{b,t} - \beta_{b,t} x_{b,t}^H \frac{\partial x_{b,t}}{\partial \phi_b^{g,b}}$$

where

$$\frac{\partial \beta_{b,t}}{\partial \phi_b^{g,b}} = -\beta_{b,t}^2 \sum_{p_x=1}^{P_x} z_{b,t,p_x}^H \frac{\partial A_{b,t}^H}{\partial \phi_b^{g,b}} A_{b,t} z_{b,t,p_x}$$

$$+ z_{b,t,p_x}^H A_{b,t}^H \frac{\partial A_{b,t}}{\partial \phi_b^{g,b}} z_{b,t,p_x},$$

$$\frac{\partial x_{b,t}^H}{\partial \phi_b^{g,b}} = \sum_{p_x=1}^{P_x} y_{b,t,p_x}^H \frac{\partial A_{b,t}}{\partial \phi_b^{g,b}} z_{b,t,p_x},$$

$$\frac{\partial x_{b,t}}{\partial \phi_b^{g,b}} = \sum_{p_x=1}^{P_x} z_{b,t,p_x} \frac{\partial A_{b,t}^H}{\partial \phi_b^{g,b}} y_{b,t,p_x}^H.$$

and

$$\frac{\partial A_{b,t}}{\partial \phi_b^{g,b}} = A_{b,t} D_{\phi_b^{g,b}},$$

where

$$D_{\phi_b^{g,b}} = \mathrm{diag}(0, \ldots, e^{j\frac{2\pi}{\lambda} d\left(m_x \cos\left(\phi_b^{g,b}\right) \cos\left(\psi_b^{g,b}\right) + m_y \sin\left(\psi_b^{g,b}\right)\right)},$$

$$\ldots, e^{j\frac{2\pi}{\lambda} d\left(M_x \cos\left(\phi_b^{g,b}\right) \cos\left(\psi_b^{g,b}\right) + M_y \sin\left(\psi_b^{g,b}\right)\right)}).$$

Similarly, we can write the first-order partial derivative of $\bar{\mathcal{G}}(\phi_b^{g,b}, \psi_b^{g,b}, \phi_b^{f,u}, \psi_b^{f,u}, g_b)$ in Eq. (24) with respect to $\psi_b^{g,b}$ is written as

$$\frac{\partial \bar{\mathcal{G}}}{\partial \psi_b^{g,b}} = \sum_{t=1}^{T} -\frac{\partial \beta_{b,t}}{\partial \psi_b^{g,b}} x_{b,t}^H x_{b,t} - \beta_{b,t} \frac{\partial x_{b,t}^H}{\partial \psi_b^{g,b}} x_{b,t} - \beta_{b,t} x_{b,t}^H \frac{\partial x_{b,t}}{\partial \psi_b^{g,b}}$$

where

$$\frac{\partial A_{b,t}}{\partial \psi_b^{g,b}} = A_{b,t} D_{\psi_b^{g,b}},$$

where

$$D_{\psi_b^{g,b}} = \mathrm{diag}(0, \ldots, e^{j\frac{2\pi}{\lambda} d\left(m_x \sin\left(\phi_b^{g,b}\right) \sin\left(\psi_b^{g,b}\right) + m_y \cos\left(\psi_b^{g,b}\right)\right)},$$

$$\ldots, e^{j\frac{2\pi}{\lambda} d\left(M_x \sin\left(\phi_b^{g,b}\right) \sin\left(\psi_b^{g,b}\right) + M_y \cos\left(\psi_b^{g,b}\right)\right)}).$$

Besides, for $\bar{\mathcal{G}}(\phi_b^{g,b}, \psi_b^{g,b}, \phi_b^{f,u}, \psi_b^{f,u}, g_b)$ in Eq. (24), its first-order partial derivative with respect to $\phi_b^{f,u}$ and $\psi_b^{f,u}$ can be written similar to the above process replacing the corresponding variables.

APPENDIX B
DERIVATIVES IN THE SECOND STAGE

The first-order partial derivative of $\bar{\mathcal{F}}(\{h_b\} | e^{(i)})$ in Eq. (27) with respect to ϕ_b^Δ is written as

$$\frac{\partial \bar{\mathcal{F}}(\{h_b\} | e^{(i)})}{\partial \phi_b^\Delta} = -\frac{\partial \zeta_b}{\partial \phi_b^\Delta} \tau_b - \zeta_b \frac{\partial \tau_b}{\partial \phi_b^\Delta}$$

where

$$\frac{\partial \zeta_b}{\partial \phi_b^\Delta} = -\zeta_b^2(a(\phi_b^\Delta, \psi_b^\Delta)^H D_{\phi_b^\Delta}^H (E^{(i)})^H \Xi_b^H \Xi_b E^{(i)} a(\phi_b^\Delta, \psi_b^\Delta)$$

$$+ a(\phi_b^\Delta, \psi_b^\Delta)^H (E^{(i)})^H \Xi_b^H \Xi_b E^{(i)} D_{\phi_b^\Delta} a(\phi_b^\Delta, \psi_b^\Delta)),$$

$$\frac{\partial \tau_b}{\partial \phi_b^\Delta} = a(\phi_b^\Delta, \psi_b^\Delta)^H D_{\phi_b^\Delta}^H (E^{(i)})^H \Xi_b^H \hat{g}_b \hat{g}_b^H \Xi_b E^{(i)} a(\phi_b^\Delta, \psi_b^\Delta)$$

$$+ a(\phi_b^\Delta, \psi_b^\Delta)^H (E^{(i)})^H \Xi_b^H \hat{g}_b \hat{g}_b^H \Xi_b E^{(i)} D_{\phi_b^\Delta} a(\phi_b^\Delta, \psi_b^\Delta),$$

and $D_{\phi_b^\Delta}$ is defined as

$$D_{\phi_b^\Delta} = \mathrm{diag}(0, \ldots, e^{j\frac{2\pi}{\lambda} d\left(l_x \cos\left(\phi_b^\Delta\right) \cos\left(\psi_b^\Delta\right) + l_y \sin\left(\psi_b^\Delta\right)\right)},$$

$$\ldots, e^{j\frac{2\pi}{\lambda} d\left(L_x \cos\left(\phi_b^\Delta\right) \cos\left(\psi_b^\Delta\right) + L_y \sin\left(\psi_b^\Delta\right)\right)}).$$

Similarly, we can write the first-order partial derivative of $\bar{\mathcal{F}}(\{h_b\} | e^{(i)})$ in Eq. (27) with respect to ψ_b^Δ is written as

$$\frac{\partial \bar{\mathcal{F}}(\{h_b\} | e^{(i)})}{\partial \psi_b^\Delta} = -\frac{\partial \zeta_b}{\partial \psi_b^\Delta} \tau_b - \zeta_b \frac{\partial \tau_b}{\partial \psi_b^\Delta}$$

where

$$\frac{\partial \zeta_b}{\partial \psi_b^\Delta} = -\zeta_b^2(a(\phi_b^\Delta, \psi_b^\Delta)^H D_{\psi_b^\Delta}^H (E^{(i)})^H \Xi_b^H \Xi_b E^{(i)} a(\phi_b^\Delta, \psi_b^\Delta)$$

$$+ a(\phi_b^\Delta, \psi_b^\Delta)^H (E^{(i)})^H \Xi_b^H \Xi_b E^{(i)} D_{\psi_b^\Delta} a(\phi_b^\Delta, \psi_b^\Delta)),$$

$$\frac{\partial \tau_b}{\partial \phi_b^\Delta} = a(\phi_b^\Delta, \psi_b^\Delta)^H D_{\psi_b^\Delta}^H (E^{(i)})^H \Xi_b^H \hat{g}_b \hat{g}_b^H \Xi_b E^{(i)} a(\phi_b^\Delta, \psi_b^\Delta)$$

$$+ a(\phi_b^\Delta, \psi_b^\Delta)^H (E^{(i)})^H \Xi_b^H \hat{g}_b \hat{g}_b^H \Xi_b E^{(i)} D_{\psi_b^\Delta} a(\phi_b^\Delta, \psi_b^\Delta),$$

and $\boldsymbol{D}_{\psi_b^\Delta}$ is defined as

$$\boldsymbol{D}_{\psi_b^\Delta} = \mathrm{diag}\big(0, \ldots, e^{j\frac{2\pi}{\lambda}d\left(l_x \cos\left(\phi_b^\Delta\right)\cos\left(\psi_b^\Delta\right)+l_y \sin\left(\psi_b^\Delta\right)\right)},$$
$$\ldots, e^{j\frac{2\pi}{\lambda}d\left(L_x \cos\left(\phi_b^\Delta\right)\cos\left(\psi_b^\Delta\right)+L_y \sin\left(\psi_b^\Delta\right)\right)}\big).$$

Here, all the first-order partial derivatives of objective function are completed.

APPENDIX C
PROOF OF PROPERTY 1

Since the update term of the recursive long-term function $\tilde{\mathcal{F}}(\boldsymbol{e}|\boldsymbol{h}_b, \hat{\boldsymbol{g}}_b)$ is a quadratic function on $\boldsymbol{e}$, it is convex in $\boldsymbol{e}$ and the first property can be satisfied.

Then, for the second property, we define a function

$$\dot{\mathcal{F}}(\boldsymbol{e}|\boldsymbol{h}_b, \hat{\boldsymbol{g}}_b) = \mathbb{E}\big\{\tilde{\mathcal{F}}(\boldsymbol{e}|\boldsymbol{h}_b, \hat{\boldsymbol{g}}_b)\big\}.$$

following a similar analysis to Appendix A in [35],

$$\lim_{b\to\infty} |\mathcal{H}_b(\boldsymbol{e}) - \dot{\mathcal{F}}(\boldsymbol{e}|\boldsymbol{h}_b, \hat{\boldsymbol{g}}_b)| = 0.$$

Beside, $\dot{\mathcal{F}}(\boldsymbol{e}|\boldsymbol{h}_b, \hat{\boldsymbol{g}}_b)$ Lipschitz continuous in $\boldsymbol{e}$ according to [35], i.e.

$$|\dot{\mathcal{F}}(\boldsymbol{e}_{b_1}|\boldsymbol{h}_b, \hat{\boldsymbol{g}}_b) - \dot{\mathcal{F}}(\boldsymbol{e}_{b_2}|\boldsymbol{h}_b, \hat{\boldsymbol{g}}_b)| \le B_f ||\boldsymbol{e}_{b_1} - \boldsymbol{e}_{b_2}||$$
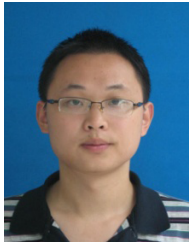
for some constant B_f. Combining the above two equations, it is obvious that $|\mathcal{H}_{b_1}(\boldsymbol{e}) - \mathcal{H}_{b_2}(\boldsymbol{e})| \le B_f ||\boldsymbol{e}_{b_1} - \boldsymbol{e}_{b_2}|| + \delta(b_1, b_2), \forall ||\boldsymbol{e}||_\infty \le 1$ for some constant $B_f > 0$, where $\lim_{b_1, b_2 \to \infty} \delta(b_1, b_2) = 0$.

For the third property, according to the lemma 2 in [10], the derivative and second-order derivative of $\min_{\boldsymbol{e}} \tilde{\mathcal{F}}(\boldsymbol{e}|\Omega^*(\boldsymbol{e}))$ w.r.t. $\boldsymbol{e}$ are bounded. Considering Eq. (42), The derivative and second-order derivative of $\mathcal{H}_b(\boldsymbol{e})$ are also uniformly bounded.

REFERENCES

[1] T. L. Marzetta, E. G. Larsson, Y. Hong, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge, U.K.: Cambridge Univ. Press, 2016.

[2] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1838–1851, Mar. 2020.

[3] S. Hu, Z. Wei, Y. Cai, D. Wing Kwan Ng, and J. Yuan, "Sum-rate maximization for multiuser MISO downlink systems with self-sustainable IRS," 2020, *arXiv:2005.11663*. [Online]. Available: http://arxiv.org/abs/2005.11663

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

[5] L. Zhao, D. W. K. Ng, and J. Yuan, "Multi-user precoding and channel estimation for hybrid millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1576–1590, Jul. 2017.

[6] Q.-U.-U. Nadeem, A. Kammoun, A. Chaaban, M. Debbah, and M.-S. Alouini, "Intelligent reflecting surface assisted wireless communication: Modeling and channel estimation," 2019, *arXiv:1906.02360*. [Online]. Available: http://arxiv.org/abs/1906.02360

[7] T. Lindstrøm Jensen and E. De Carvalho, "An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator," 2019, *arXiv:1909.09440*. [Online]. Available: http://arxiv.org/abs/1909.09440

[8] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," 2019, *arXiv:1904.10136*. [Online]. Available: http://arxiv.org/abs/1904.10136

[9] L. Wei, C. Huang, G. C. Alexandropoulos, and C. Yuen, "Parallel factor decomposition channel estimation in RIS-assisted multi-user MISO communication," 2020, *arXiv:2001.09413*. [Online]. Available: http://arxiv.org/abs/2001.09413

[10] 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.

[11] J. He, M. Leinonen, H. Wymeersch, and M. Juntti, "Channel estimation for RIS-aided mmWave MIMO channels," 2020, *arXiv:2002.06453*. [Online]. Available: http://arxiv.org/abs/2002.06453

[12] Y. Teng, L. Jia, A. Liu, and V. K. N. Lau, "Joint estimation of channel and I/Q imbalance in massive MIMO: A two-timescale optimization approach," *IEEE Trans. Wireless Commun.*, vol. 18, no. 10, pp. 4723–4737, Oct. 2019.

[13] E. Bjornson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, Nov. 2014.

[14] S. Zarei, W. H. Gerstacker, J. Aulin, and R. Schober, "I/Q imbalance aware widely-linear receiver for uplink multi-cell massive MIMO systems: Design and sum rate analysis," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3393–3408, May 2016.

[15] Y. Kuga, F. T. Ulaby, T. F. Haddock, and R. D. DeRoo, "Millimeter-wave radar scattering from snow 1. Radiative transfer model," *Radio Sci.*, vol. 26, no. 2, pp. 329–341, Mar. 1991.

[16] A. Arage, W. M. Steffens, G. Kuehnle, and R. Jakoby, "Effects of water and ice layer on automotive radar," in *Proc. German Microw. Conf.*, 2006, pp. 1–5.

[17] F. T. Ulaby, T. F. Haddock, R. T. Austin, and Y. Kuga, "Millimeter-wave radar scattering from snow: 2. Comparison of theory with experimental observations," *Radio Sci.*, vol. 26, no. 2, pp. 343–351, Mar. 1991.

[18] M. Li and Y. Lu, "Source bearing and steering-vector estimation using partially calibrated arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 45, no. 4, pp. 1361–1372, Oct. 2009.

[19] R. Iglesias, F. Ares, M. Fernandez-Delgado, J. Rodriguez, J. Bregains, and S. Barro, "Element failure detection in linear antenna arrays using case-based reasoning," *IEEE Antennas Propag. Mag.*, vol. 50, no. 4, pp. 198–204, Aug. 2008.

[20] J. A. Rodriguez-Gonzalez, F. Ares-Pena, M. Fernandez-Delgado, R. Iglesias, and S. Barro, "Rapid method for finding faulty elements in antenna arrays using far field pattern samples," *IEEE Trans. Antennas Propag.*, vol. 57, no. 6, pp. 1679–1683, Jun. 2009.

[21] A. Buonanno and M. D'urso, "On the diagnosis of arbitrary geometry fully active arrays," in *Proc. 4th Eur. Conf. Antennas Propag.*, Jun. 2010, pp. 1–4.

[22] O. Aldayel, "Constrained optimization for millimeter array diagnosis," in *Proc. IEEE Radar Conf. (RadarConf)*, Apr. 2019, pp. 1–6.

[23] M. D. Migliore, "A compressed sensing approach for array diagnosis from a small set of near-field measurements," *IEEE Trans. Antennas Propag.*, vol. 59, no. 6, pp. 2127–2133, Jun. 2011.

[24] G. Oliveri, P. Rocca, and A. Massa, "Reliable diagnosis of large linear arrays—A Bayesian compressive sensing approach," *IEEE Trans. Antennas Propag.*, vol. 60, no. 10, pp. 4627–4636, Oct. 2012.

[25] B. Friedlander and T. Strohmer, "Bilinear compressed sensing for array self-calibration," in *Proc. 48th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2014, pp. 363–367.

[26] M. E. Eltayeb, T. Y. Al-Naffouri, and R. W. Heath, "Compressive sensing for millimeter wave antenna array diagnosis," *IEEE Trans. Commun.*, vol. 66, no. 6, pp. 2708–2721, Jun. 2018.

[27] S. Ma, W. Shen, J. An, and L. Hanzo, "Antenna array diagnosis for millimeter-wave MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4585–4589, Apr. 2020.

[28] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, Nov. 2019.

[29] Y. Xiu, W. Wang, and Z. Zhang, "A message passing approach to acquire mm-wave channel state information based on Out-of-Band data," *IEEE Access*, vol. 6, pp. 45665–45680, 2018.

[30] J. Lee, G.-T. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2370–2386, Jun. 2016.

[31] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1259–1262, Jun. 2016.

[32] H. Liu, X. Yuan, and Y.-J. Angela Zhang, "Matrix-calibration-based cascaded channel estimation for reconfigurable intelligent surface assisted multiuser MIMO," 2019, *arXiv:1912.09025*. [Online]. Available: http://arxiv.org/abs/1912.09025

[33] C. Hu, L. Dai, T. Mir, Z. Gao, and J. Fang, "Super-resolution channel estimation for mmWave massive MIMO with hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8954–8958, Sep. 2018.

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

[35] A. Liu, V. K. N. Lau, and M.-J. Zhao, "Online successive convex approximation for two-stage stochastic nonconvex optimization," *IEEE Trans. Signal Process.*, vol. 66, no. 22, pp. 5941–5955, Nov. 2018.

[36] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, Jan. 2013.

ZHENZHEN HU received the Ph.D. degree in communication and information systems from the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu, China, in 2017. From 2014 to 2017, she was a Visiting Student with The University of British Columbia. She is currently a Faculty Member with the Chengdu University of Information Technology. Her research interests include cognitive radio networks, cooperative relay networks, the Internet-of-Things, artificial intelligence, and machine learning techniques.

BINRUI LI received the B.Eng. degree from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in July 2011, where he is currently pursuing the Ph.D. degree with the National Key Laboratory of Science and Technology on Communication. His research interests include wireless communications and signal processing, including massive multiple-input multiple-output channel estimation, beamforming/precoding, compressed sensing, and pilot contamination elimination/reduction techniques.

ZHONGPEI ZHANG (Member, IEEE) received the Ph.D. degree in traffic information engineering from Southwest Jiaotong University, Chengdu, China, in 2000. From 2001 to 2003, he held a postdoctoral position with Tsinghua University. From March 2004 to February 2005, he was a Research Scientist with the University of Oulu, Oulu, Finland. He is currently a Professor with the University of Electronic Science and Technology of China. His research interests include wireless communication networks, equalization and iterative receiver, space-time coding, and non-coherent detection algorithm.

YANG CHEN received the B.S. degree in electrical engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2011, and the M.S. degree in communication and information systems from Southwest Jiaotong University, Chengdu, in 2015. He is currently pursuing the Ph.D. degree in communication and information system with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China. His research interests include massive multiple-input multiple-output and physical layer security.

• • •