Multiuser Uplink MIMO Communications Assisted by Multiple Reconfigurable Intelligent Surfaces

Yang Lv, *Member, IEEE,* Zhiqiang He, *Member, IEEE,* and Yue Rong, *Senior Member, IEEE*

***Abstract*—This letter focuses on the design of multiuser uplink multiple-input multiple-output (MIMO) communications assisted by multiple reconfigurable intelligent surfaces (RISs), where we consider both linear minimal mean-squared error (MMSE) and nonlinear MMSE-decision feedback equalization (DFE) receivers. The RISs, consisting of massive low-cost passive elements, can reflect the incident signals with either continuous or discrete phase shifts. Adopting the sum mean-squared error (MSE) minimization criterion, we jointly optimize the RIS phase shifts, the linear or nonlinear receiving matrices, and the source precoding matrices in iterative manner. Simulation results show that, when the signals in direct user-access point (AP) links suffer severe attenuation, the introduction of indirect user-RIS-AP links can significantly improve the system performance.**

***Index Terms*—Reconfigurable intelligent surfaces, MIMO communications, multiuser, MMSE, DFE, semidefinite relaxation.**

I. Introduction

In recent years, for the ability to considerably enhance the quality of wireless communications along with high spectrum and energy efficiency, reconfigurable intelligent surfaces (RISs) [1], also called intelligent reflecting surfaces [2], etc., have drawn wide attention in academia and industry. An RIS is generally a planar array of low-cost passive elements, each of which can independently induce an adjustable phase shift and/or amplitude change to its incident electromagnetic wave, thereby reconfiguring the characteristics of wireless channels and making the concept of "smart radio environments" emerge.

Considering a downlink single-user multiple-input single-output (MISO) system, [3] investigated the joint design of the active beamforming at the access point (AP) and the passive beamforming at the RIS. References [4] and [5] then extended the single-user scenario in [3] to the multiuser ones with continuous and discrete RIS phase shifts, respectively, and both of them focused on minimizing the transmission power of the AP. Targeting at maximizing all users' weighted sum-rate, [6] studied both perfect and imperfect channel state information (CSI) circumstances. In [7], the direct links between the base station and users were neglected due to the poor conditions for signal propagation, and two energy efficiency maximization algorithms were presented. Towards a single-user uplink single-input multiple-output (SIMO) system, [8] developed a passive beamforming and information transfer (PBIT) technique where the RIS utilizes the on/off states of passive elements to carry its private data. Such technique was further expanded in [9] to apply to multiuser multiple-input multiple-output (MIMO) systems. To guard against the eavesdroppers, [10] addressed the issue of secure communications with the aid of RISs.

This letter investigates the design of multiple RISs-assisted multiuser uplink MIMO communications with both continuous and discrete RIS phase shifts take into consideration, where either the linear minimal mean-squared error (MMSE) receiver or the nonlinear MMSE-decision feedback equalization (DFE) receiver [11] can be adopted at the AP. Our design criterion is to minimize the sum mean-squared error (MSE) for the signal waveform estimation of all data streams and the transmission power of each user is limited. To solve the formulated nonconvex problem, we employ the block coordinate decent (BCD) method [12, Sec. 2.7] to iteratively optimize the signal phase shifts at all RISs, the linear or nonlinear receiving matrices at the AP, and the source precoding matrices at all users, where the semidefinite relaxation (SDR) technique [13], the QR factorization [14, Sec. 2.1], and the Karush-Kuhn-Tucker (KKT) conditions [15] are respectively used. Simulation results show that, when direct user-AP communication links are unfavorable, in comparison to the systems without RISs, significant performance improvement can be observed for RISs-assisted systems, and such advantage becomes more prominent as the number of passive elements in each RIS increases.

Notations: $\mathbb{C}^n$ is the space of complex column vectors with dimension n and $\mathbb{C}^{m\times n}$ is that of m-by-n complex matrices. $\triangleq$ denotes "defined as". $\sim$ signifies "distributed as". $j \triangleq \sqrt{-1}$ is the imaginary unit. $(a)_{m\times n}$ denotes an m-by-n matrix with all its elements being scalar a. $(\cdot)^*$, $|\cdot|$, and $\angle(\cdot)$ stand for the complex conjugate, modulus, and phase angle of a scalar, respectively. $\boldsymbol{I}_n$ is the nth-order identity matrix. $(\cdot)^T$ and $(\cdot)^H$ represent the transpose and Hermitian transpose of a vector or matrix, respectively. $\mathrm{vec}(\boldsymbol{X})$ is a column vector made up of stacked columns of matrix $\boldsymbol{X}$ [16]. $[\boldsymbol{a}]_{1:n}$ denotes a subvector of vector $\boldsymbol{a}$, containing its first n elements. $[\boldsymbol{X}]_{m,n}$ indicates the mth row and nth column element of $\boldsymbol{X}$. $\mathrm{diag}(\cdot)$ and $\mathrm{bd}(\cdot)$ are a diagonal and a block diagonal matrix, respectively, with the diagonal elements given in parentheses. $\otimes$ is the Kronecker product operator. $\mathrm{tr}(\cdot)$, $(\cdot)^{-1}$, $(\cdot)^{\dagger}$, and $(\cdot)^{\frac{1}{2}}$ stand for the trace, inverse, Moore-Penrose inverse, and square root of a matrix [14], respectively. $\boldsymbol{X} \succeq \boldsymbol{0}$ means that $\boldsymbol{X}$ is Hermitian positive semidefinite (PSD). The distribution of a circularly symmetric complex Gaussian random variable (vector) having zero mean and variance a (covariance matrix $\boldsymbol{X}$) is denoted by $\mathcal{CN}(0,a)$ ($\mathcal{CN}(\boldsymbol{0},\boldsymbol{X})$). $\mathrm{E}[\cdot]$ stands for the statistical expectation.

This work was supported by the National Key Research and Development Program of China under Grant 2018YFB1801101.

Yang Lv and Zhiqiang He are with the Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: lvyangcn@foxmail.com; hezq@bupt.edu.cn).

Yue Rong is with the School of Electrical Engineering, Computing and Mathematical Sciences, Curtin University, Bentley, WA 6102, Australia (e-mail: y.rong@curtin.edu.au).

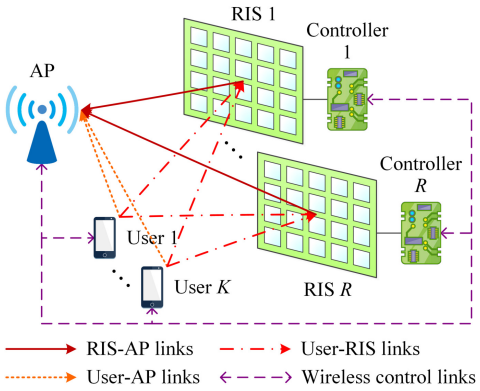

Fig. 1. Multiuser uplink MIMO communications assisted by multiple RISs.

II. System Model and Problem Formulation

As shown in Fig. 1, we consider an R RISs-assisted uplink MIMO communication system that serves K users. For $k = 1, \cdots, K$, the kth user transmits N_k data streams via $N_{s,k}$ antennas ($N_{s,k} \geqslant N_k$) with the source precoding matrix $\boldsymbol{B}_k \in \mathbb{C}^{N_{s,k} \times N_k}$. By assuming that the modulated signal vector $\boldsymbol{s}_k \in \mathbb{C}^{N_k}$ satisfies $\mathrm{E}[\boldsymbol{s}_k] = (0)_{N_k \times 1}$ and $\mathrm{E}\big[\boldsymbol{s}_k \boldsymbol{s}_k^H\big] = \boldsymbol{I}_{N_k}$, the mean power of the transmitted signal vector $\boldsymbol{x}_k = \boldsymbol{B}_k \boldsymbol{s}_k$ is derived as $\mathrm{tr}\big(\mathrm{E}\big[\boldsymbol{x}_k \boldsymbol{x}_k^H\big]\big) = \mathrm{tr}\big(\boldsymbol{B}_k \boldsymbol{B}_k^H\big)$, which should not be greater than the power budget p_k. Additionally, the AP is equipped with N_d antennas. For $r = 1, \cdots, R$, the rth RIS consists of M_r passive elements, each behaving like a keyhole [3], which receives superposed multi-path signals at a physical point and then scatters the combined signal as if from a point source [2]. This kind of RIS is called the reflectarray-based RIS in [1]. Let the passive reflection matrix of the rth RIS be written as

$$\boldsymbol{\Theta}_r \triangleq \mathrm{diag}\big(\beta_{r,1} e^{j\theta_{r,1}}, \cdots, \beta_{r,M_r} e^{j\theta_{r,M_r}}\big) \tag{1}$$

where for $m_r = 1, \cdots, M_r$, $\beta_{r,m_r} \in [0,1]$ and $\theta_{r,m_r} \in [0, 2\pi)$ are, respectively, the amplitude coefficient and the phase shift generated by the m_rth passive element for its combined signal. As in [5], [9], [10], for simplicity and maximizing the signal reflection power of RISs, we set $\beta_{r,m_r} = 1, \forall\, r, m_r$. Besides, $\theta_{r,m_r}, \forall\, r, m_r$ can take any values in $[0, 2\pi)$ for continuous RIS phase shifts, and if there are only b bits of discrete RIS phase shifts due to the limitation of hardware costs, we have $\theta_{r,m_r} \in \mathcal{L} \triangleq \Big\{0, \frac{2\pi}{L}, \cdots, \frac{2\pi(L-1)}{L}\Big\}, \forall\, r, m_r$ with $L \triangleq 2^b$. Note that, compared to a single RIS, multiple RISs can provide more passive elements, along with additional spatial diversity, and thus can further improve the system performance.

Each RIS is connected to a controller that has multiple responsibilities. For example, the rth controller can (a) exchange information with other devices via dedicated wireless control links to coordinate communications, (b) regulate the switching between two working modes of the rth RIS, i.e., the reflecting mode for data transmission and the receiving mode for CSI estimation, and (c) adjust the phase shifts of the rth RIS once $\boldsymbol{\Theta}_r$ is updated [3]–[5]. The baseband equivalent channels for the kth user-AP, the kth user-rth RIS, and the rth RIS-AP data links are denoted by $\boldsymbol{H}_{sd,k} \in \mathbb{C}^{N_d \times N_{s,k}}$, $\boldsymbol{H}_{s,rk} \in \mathbb{C}^{M_r \times N_{s,k}}$, and $\boldsymbol{H}_{d,r} \in \mathbb{C}^{N_d \times M_r}$, respectively, which are assumed to be quasi-static flat-fading and perfectly known through various CSI acquisition means as discussed in [1]–[4]. Here, we can equip the RISs with receiving radio frequency (RF) chains. After all users send orthogonal pilot signals, the AP and the RIS controllers can estimate $\boldsymbol{H}_{sd,k}$ and $\boldsymbol{H}_{s,rk}$, respectively. After the AP sends a pilot signal, the RIS controllers can estimate $\boldsymbol{H}_{d,r}$ by exploiting the channel reciprocity. Via control links, all CSI can be collected at one of the RIS controllers, which is selected to execute our proposed algorithms and deliver the optimized system parameters to their corresponding devices.

According to [6], [8], [9], the received signal at the AP is the superposition of different components coming from two types of communication links, i.e., the direct links and the RIS reflection links, among which the time delay difference is assumed to be negligible. Specifically, for $k = 1, \cdots, K$ and $r = 1, \cdots, R$, $\boldsymbol{H}_{sd,k} \boldsymbol{x}_k$ is the signal component received directly from the kth user, and $\boldsymbol{H}_{d,r} \boldsymbol{\Theta}_r \boldsymbol{H}_{s,rk} \boldsymbol{x}_k$ is that received indirectly from the kth user via the reflection of the rth RIS. Hence, the total signal received at the AP is given by $\boldsymbol{y} = \sum_{k=1}^{K} \big(\boldsymbol{H}_{sd,k} + \sum_{r=1}^{R} \boldsymbol{H}_{d,r} \boldsymbol{\Theta}_r \boldsymbol{H}_{s,rk}\big) \boldsymbol{B}_k \boldsymbol{s}_k + \boldsymbol{n}$, where $\boldsymbol{n} \sim \mathcal{CN}\big(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_{N_d}\big)$ denotes the additive white Gaussian noise (AWGN) vector at the AP. Moreover, via setting $N \triangleq \sum_{k=1}^{K} N_k$, $N_s \triangleq \sum_{k=1}^{K} N_{s,k}$, and $M \triangleq \sum_{r=1}^{R} M_r$ as well as $\boldsymbol{s} \triangleq \big[\boldsymbol{s}_1^T, \cdots, \boldsymbol{s}_K^T\big]^T \in \mathbb{C}^N$, $\boldsymbol{B} \triangleq \mathrm{bd}(\boldsymbol{B}_1, \cdots, \boldsymbol{B}_K) \in \mathbb{C}^{N_s \times N}$, $\boldsymbol{H}_{sd} \triangleq [\boldsymbol{H}_{sd,1}, \cdots, \boldsymbol{H}_{sd,K}] \in \mathbb{C}^{N_d \times N_s}$, $\boldsymbol{H}_d \triangleq [\boldsymbol{H}_{d,1}, \cdots, \boldsymbol{H}_{d,R}] \in \mathbb{C}^{N_d \times M}$, $\boldsymbol{\Theta} \triangleq \mathrm{bd}(\boldsymbol{\Theta}_1, \cdots, \boldsymbol{\Theta}_R) \in \mathbb{C}^{M \times M}$, $\boldsymbol{H}_{s,k} \triangleq \big[\boldsymbol{H}_{s,1k}^T, \cdots, \boldsymbol{H}_{s,Rk}^T\big]^T \in \mathbb{C}^{M \times N_{s,k}}$ for $k = 1, \cdots, K$, and $\boldsymbol{H}_s \triangleq [\boldsymbol{H}_{s,1}, \cdots, \boldsymbol{H}_{s,K}] \in \mathbb{C}^{M \times N_s}$, we have

$$\boldsymbol{y} = (\boldsymbol{H}_{sd} + \boldsymbol{H}_d \boldsymbol{\Theta} \boldsymbol{H}_s) \boldsymbol{B} \boldsymbol{s} + \boldsymbol{n} = \boldsymbol{A} \boldsymbol{s} + \boldsymbol{n} \tag{2}$$

with $\boldsymbol{A} \triangleq (\boldsymbol{H}_{sd} + \boldsymbol{H}_d \boldsymbol{\Theta} \boldsymbol{H}_s) \boldsymbol{B} \in \mathbb{C}^{N_d \times N}$. For convenience, we further make $\boldsymbol{\Theta} = \mathrm{diag}\big(e^{j\theta_1}, \cdots, e^{j\theta_M}\big)$ and $\boldsymbol{\theta} \triangleq [\theta_1, \cdots, \theta_M]^T$, where for $m = 1, \cdots, M$, $\theta_m \in [0, 2\pi)$ or $\theta_m \in \mathcal{L}$, and for $r = 1, \cdots, R$ and $m_r = 1, \cdots, M_r$, θ_{r,m_r} is just the $\big(\sum_{\zeta=0}^{r-1} M_\zeta + m_r\big)$th element of $\boldsymbol{\theta}$ with $M_0 = 0$.

At the AP, when we use the linear MMSE receiver, the source signal vector $\boldsymbol{s}$ is estimated as $\bar{\boldsymbol{s}} = \bar{\boldsymbol{W}}^H \boldsymbol{y}$, where $\bar{\boldsymbol{W}} \in \mathbb{C}^{N_d \times N}$ is the linear receiving matrix and $\bar{\boldsymbol{s}}$ is the linearly estimated signal vector. Thus the sum MSE for estimating all symbols in $\boldsymbol{s}$ can be expressed as $\bar{E}_s \triangleq \mathrm{tr}\big(\mathrm{E}\big[(\bar{\boldsymbol{s}} - \boldsymbol{s})(\bar{\boldsymbol{s}} - \boldsymbol{s})^H\big]\big) = \mathrm{tr}\Big(\big(\bar{\boldsymbol{W}}^H \boldsymbol{A} - \boldsymbol{I}_N\big)\big(\bar{\boldsymbol{W}}^H \boldsymbol{A} - \boldsymbol{I}_N\big)^H + \sigma^2 \bar{\boldsymbol{W}}^H \bar{\boldsymbol{W}}\Big)$. When we use the nonlinear MMSE-DFE receiver instead, $\boldsymbol{s}$ is estimated as $\hat{\boldsymbol{s}} = \boldsymbol{W}^H \boldsymbol{y} - \boldsymbol{D} \tilde{\boldsymbol{s}}$, where $\boldsymbol{W} \in \mathbb{C}^{N_d \times N}$ is the decision feed-forward matrix with its ith column vector denoted by $\boldsymbol{w}_i$ for $i = 1, \cdots, N$, $\boldsymbol{D} \in \mathbb{C}^{N \times N}$ is the decision feedback matrix which is strictly upper-triangular, $\tilde{\boldsymbol{s}} = [\tilde{s}_1, \cdots, \tilde{s}_N]^T$ is the detected signal vector, and $\hat{\boldsymbol{s}} = [\hat{s}_1, \cdots, \hat{s}_N]^T$ is the nonlinearly estimated signal vector. Specifically, we first estimate the Nth symbol of $\boldsymbol{s}$ as $\hat{s}_N = \boldsymbol{w}_N^H \boldsymbol{y}$, which is then detected as $\tilde{s}_N$. Following that, we estimate the ith symbol of $\boldsymbol{s}$ for $i = N-1, \cdots, 1$, one by one, as $\hat{s}_i = \boldsymbol{w}_i^H \boldsymbol{y} - \sum_{l=i+1}^{N} [\boldsymbol{D}]_{i,l} \tilde{s}_l$, which is then detected as $\tilde{s}_i$. Besides, like in [11], during our mathematical derivations, we assume that there is no error propagation within the MMSE-DFE receiver, i.e., $\tilde{\boldsymbol{s}} = \boldsymbol{s}$, thus we obtain $\hat{\boldsymbol{s}} = \boldsymbol{W}^H \boldsymbol{y} - \boldsymbol{D} \boldsymbol{s}$. Consequently, the sum MSE is now given by $E_s \triangleq \mathrm{tr}\big(\mathrm{E}\big[(\hat{\boldsymbol{s}} - \boldsymbol{s})(\hat{\boldsymbol{s}} - \boldsymbol{s})^H\big]\big) = \mathrm{tr}\Big(\big(\boldsymbol{W}^H \boldsymbol{A} - \boldsymbol{U}\big)\big(\boldsymbol{W}^H \boldsymbol{A} - \boldsymbol{U}\big)^H + \sigma^2 \boldsymbol{W}^H \boldsymbol{W}\Big)$, where $\boldsymbol{U} \triangleq \boldsymbol{D} + \boldsymbol{I}_N$ is also called the decision feedback matrix hereafter.

At this point, with the above nonlinear receiving structure, we can formulate the system optimization problem as follows.

$$\min_{\{\theta_m\},\{\boldsymbol{W},\boldsymbol{U}\},\{\boldsymbol{B}_k\}} \ \mathrm{tr}\Big(\big(\boldsymbol{W}^H\boldsymbol{A}-\boldsymbol{U}\big) \times \big(\boldsymbol{W}^H\boldsymbol{A}-\boldsymbol{U}\big)^H+\sigma^2\boldsymbol{W}^H\boldsymbol{W}\Big) \quad (3)$$

$$\text{s.t. } \theta_m\in[0,2\pi) \text{ or } \theta_m\in\mathcal{L}, \quad m=1,\cdots,M, \quad (4)$$

$$[\boldsymbol{U}]_{i,l}=\begin{cases}0, & i>l,\\ 1, & i=l,\end{cases} \quad i,l=1,\cdots,N, \quad (5)$$

$$\mathrm{tr}\big(\boldsymbol{B}_k\boldsymbol{B}_k^H\big)\leqslant p_k, \quad k=1,\cdots,K \quad (6)$$

where $\{\theta_m\}\triangleq\{\theta_1,\cdots,\theta_M\}$ and $\{\boldsymbol{B}_k\}\triangleq\{\boldsymbol{B}_1,\cdots,\boldsymbol{B}_K\}$. Noteworthily, when $\boldsymbol{W}$ is substituted by $\bar{\boldsymbol{W}}$ and $\boldsymbol{U}$ is set to be $\boldsymbol{I}_N$, the problem (3)–(6) turns into the one with linear receiving structure. Besides, since only the phase angles of the diagonal elements of $\boldsymbol{\Theta}_r$ are adjustable and the RISs can not increase the power of their reflected signals, it is more challenging to design the systems with RISs than with amplify-and-forward relays [11]. In the next section, following the BCD method, we shall develop iterative algorithms to solve the problem (3)–(6).

III. Algorithm Design

Here, the three blocks of variables for the problem (3)–(6), i.e., $\{\theta_m\}$, $\{\boldsymbol{W},\boldsymbol{U}\}$, and $\{\boldsymbol{B}_k\}$, are iteratively optimized one by one with the other blocks fixed. At the beginning of such iterative procedures, we initialize $\boldsymbol{W}=(1)_{N_d\times N}$, $\boldsymbol{U}=\boldsymbol{I}_N$, and $\boldsymbol{B}_k=\big[\sqrt{p_k/N_k}\boldsymbol{I}_{N_k},(0)_{N_k\times(N_{s,k}-N_k)}\big]^T$ for $k=1,\cdots,K$.

Firstly, with fixed $\{\boldsymbol{W},\boldsymbol{U}\}$ and $\{\boldsymbol{B}_k\}$, we optimize $\{\theta_m\}$ by solving the problem as below.

$$\min_{\{\theta_m\}} \ \mathrm{tr}\Big(\big(\boldsymbol{W}^H\boldsymbol{A}-\boldsymbol{U}\big)\big(\boldsymbol{W}^H\boldsymbol{A}-\boldsymbol{U}\big)^H\Big) \quad (7)$$

$$\text{s.t. } \theta_m\in[0,2\pi) \text{ or } \theta_m\in\mathcal{L}, \quad m=1,\cdots,M \quad (8)$$

which, via $\mathrm{tr}\big(\boldsymbol{X}\boldsymbol{X}^H\big)=\mathrm{vec}(\boldsymbol{X})^H\mathrm{vec}(\boldsymbol{X})$ and $\mathrm{vec}(\boldsymbol{X}\boldsymbol{Y}\boldsymbol{Z})=\big(\boldsymbol{Z}^T\otimes\boldsymbol{X}\big)\mathrm{vec}(\boldsymbol{Y})$ [16, Lemma 4.3.1], can be rewritten as

$$\min_{\{\theta_m\}} \ \mathrm{vec}(\boldsymbol{\Theta})^H\boldsymbol{C}\,\mathrm{vec}(\boldsymbol{\Theta})+\mathrm{vec}(\boldsymbol{\Theta})^H\boldsymbol{c}+\boldsymbol{c}^H\mathrm{vec}(\boldsymbol{\Theta}) \quad (9)$$

$$\text{s.t. } \theta_m\in[0,2\pi) \text{ or } \theta_m\in\mathcal{L}, \quad m=1,\cdots,M \quad (10)$$

with $\boldsymbol{C}\triangleq\big[(\boldsymbol{H}_s\boldsymbol{B})^T\otimes\big(\boldsymbol{W}^H\boldsymbol{H}_d\big)\big]^H\big[(\boldsymbol{H}_s\boldsymbol{B})^T\otimes\big(\boldsymbol{W}^H\boldsymbol{H}_d\big)\big]$ and $\boldsymbol{c}\triangleq\big[(\boldsymbol{H}_s\boldsymbol{B})^T\otimes\big(\boldsymbol{W}^H\boldsymbol{H}_d\big)\big]^H\mathrm{vec}\big(\boldsymbol{W}^H\boldsymbol{H}_{sd}\boldsymbol{B}-\boldsymbol{U}\big)$. Now, by setting $\boldsymbol{v}\triangleq[v_1,\cdots,v_M]^T$ along with $v_m\triangleq e^{j\theta_m}$ for $m=1,\cdots,M$ and $\mathcal{V}\triangleq\Big\{1,e^{j\frac{2\pi}{L}},\cdots,e^{j\frac{2\pi(L-1)}{L}}\Big\}$, we are able to reformulate the problem (9)–(10) as

$$\min_{\boldsymbol{v}} \ \boldsymbol{v}^H\boldsymbol{C}_0\boldsymbol{v}+\boldsymbol{v}^H\boldsymbol{c}_0+\boldsymbol{c}_0^H\boldsymbol{v} \quad (11)$$

$$\text{s.t. } |v_m|=1 \text{ or } v_m\in\mathcal{V}, \quad m=1,\cdots,M \quad (12)$$

where $\boldsymbol{C}_0\in\mathbb{C}^{M\times M}$ is made up of all the elements at the intersections of the $[(i-1)M+i]$th rows and the $[(l-1)M+l]$th columns of matrix $\boldsymbol{C}$ for $i,l=1,\cdots,M$ and $\boldsymbol{c}_0\in\mathbb{C}^M$ is made up of all the $[(m-1)M+m]$th elements of vector $\boldsymbol{c}$ for $m=1,\cdots,M$. To homogenize the above nonconvex quadratic programming problem, we bring in an auxiliary variable α and gives the following problem with $\dot{\boldsymbol{v}}\triangleq[\dot{v}_1,\cdots,\dot{v}_M]^T$.

$$\min_{\dot{\boldsymbol{v}},\alpha} \ \dot{\boldsymbol{v}}^H\boldsymbol{C}_0\dot{\boldsymbol{v}}+\dot{\boldsymbol{v}}^H\boldsymbol{c}_0\alpha+\alpha^*\boldsymbol{c}_0^H\dot{\boldsymbol{v}} \quad (13)$$

$$\text{s.t. } |\dot{v}_m|=1,\ |\alpha|=1 \text{ or } \dot{v}_m,\alpha\in\mathcal{V}, \quad m=1,\cdots,M. \quad (14)$$

It can be readily observed that, if $\hat{\boldsymbol{v}}\triangleq\big[\dot{\boldsymbol{v}}^T,\alpha\big]^T$ solves the problem (13)–(14), then $\boldsymbol{v}=\alpha^*\dot{\boldsymbol{v}}$ solves the problem (11)–(12). Via further setting $\hat{\boldsymbol{C}}_0\triangleq\begin{bmatrix}\boldsymbol{C}_0 & \boldsymbol{c}_0\\ \boldsymbol{c}_0^H & 0\end{bmatrix}$ and $\hat{\boldsymbol{v}}=[\hat{v}_1,\cdots,\hat{v}_M,\hat{v}_{M+1}]^T$, we can rewrite the problem (11)–(12) as

$$\min_{\hat{\boldsymbol{v}}} \ \hat{\boldsymbol{v}}^H\hat{\boldsymbol{C}}_0\hat{\boldsymbol{v}} \quad (15)$$

$$\text{s.t. } |\hat{v}_m|=1 \text{ or } \hat{v}_m\in\mathcal{V}, \quad m=1,\cdots,M,M+1 \quad (16)$$

of which the objective $\hat{\boldsymbol{v}}^H\hat{\boldsymbol{C}}_0\hat{\boldsymbol{v}}$ is equivalent to $\mathrm{tr}\big(\hat{\boldsymbol{C}}_0\hat{\boldsymbol{v}}\hat{\boldsymbol{v}}^H\big)=\mathrm{tr}\big(\hat{\boldsymbol{C}}_0\hat{\boldsymbol{V}}\big)$ with $\hat{\boldsymbol{V}}\triangleq\hat{\boldsymbol{v}}\hat{\boldsymbol{v}}^H\in\mathbb{C}^{(M+1)\times(M+1)}$ required to satisfy $\hat{\boldsymbol{V}}\succeq\boldsymbol{0}$ and $\mathrm{rank}\big(\hat{\boldsymbol{V}}\big)=1$. Since the rank-one constraint of $\hat{\boldsymbol{V}}$ is nonconvex, by applying the SDR technique, we leave it out to acquire the following relaxed version of the problem (15)–(16) for both continuous and discrete RIS phase shifts.

$$\min_{\hat{\boldsymbol{V}}} \ \mathrm{tr}\big(\hat{\boldsymbol{C}}_0\hat{\boldsymbol{V}}\big) \quad (17)$$

$$\text{s.t. } \hat{\boldsymbol{V}}\succeq\boldsymbol{0},\ \big[\hat{\boldsymbol{V}}\big]_{m,m}=1, \quad m=1,\cdots,M,M+1. \quad (18)$$

The above is a convex semidefinite programming (SDP) problem that can be optimally solved via, e.g., the convex optimization software: CVX [17]. Because the optimal solution of the problem (17)–(18) may not meet $\mathrm{rank}\big(\hat{\boldsymbol{V}}\big)=1$, the optimal objective value of the problem (17)–(18) only provides a lower bound of that of the problem (15)–(16). Here, to get $\hat{\boldsymbol{v}}$ from $\hat{\boldsymbol{V}}$, we adopt the Gaussian randomization method [13] as below.

On the basis of the eigenvalue decomposition (EVD) of $\hat{\boldsymbol{V}}$, i.e., $\hat{\boldsymbol{V}}=\boldsymbol{\Upsilon}\boldsymbol{\Lambda}\boldsymbol{\Upsilon}^H$ with $\boldsymbol{\Upsilon},\boldsymbol{\Lambda}\in\mathbb{C}^{(M+1)\times(M+1)}$ being a unitary and a diagonal matrix, respectively, we generate T random vectors as $\boldsymbol{z}^{(t)}=\boldsymbol{\Upsilon}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\beta}^{(t)}=\big[z_1^{(t)},\cdots,z_M^{(t)},z_{M+1}^{(t)}\big]^T$, where $\boldsymbol{\beta}^{(t)}\sim\mathcal{CN}(\boldsymbol{0},\boldsymbol{I}_{M+1})$ for $t=1,\cdots,T$. Then, for constructing the feasible points $\hat{\boldsymbol{v}}^{(t)}=\big[\hat{v}_1^{(t)},\cdots,\hat{v}_M^{(t)},\hat{v}_{M+1}^{(t)}\big]^T$, in the case of continuous RIS phase shifts, we set $\hat{v}_m^{(t)}=z_m^{(t)}\big/\big|z_m^{(t)}\big|$ for $m=1,\cdots,M,M+1$, meanwhile, in the case of discrete RIS phase shifts, we set $\hat{v}_m^{(t)}=e^{j2\pi\xi/L}$ if $\mathrm{angle}\big(z_m^{(t)}\big)\in\big[(2\xi-1)\pi/L,(2\xi+1)\pi/L\big)$ with $\xi\in\{0,1,\cdots,L-1\}$. Among all $\hat{\boldsymbol{v}}^{(t)}$, we determine $\hat{t}=\mathrm{argmin}_{t=1,\cdots,T}\big(\hat{\boldsymbol{v}}^{(t)}\big)^H\hat{\boldsymbol{C}}_0\big(\hat{\boldsymbol{v}}^{(t)}\big)$ and set $\hat{\boldsymbol{v}}=\hat{\boldsymbol{v}}^{(\hat{t})}$ as the approximate solution of the problem (15)–(16). Following that, the original problem (11)–(12) can thus be approximately solved by $\boldsymbol{v}=\hat{v}_{M+1}^*[\hat{\boldsymbol{v}}]_{1:M}$. As pointed out in [13], the accuracy of the obtained solution depends on the number of randomizations T. Via simulation tests, setting $T=100(M+1)$ is deemed appropriate here.

Secondly, with fixed $\{\boldsymbol{B}_k\}$ and $\{\theta_m\}$, we optimize $\{\boldsymbol{W},\boldsymbol{U}\}$ by using the QR factorization $\begin{bmatrix}\sqrt{1/\sigma^2}\boldsymbol{A}\\ \boldsymbol{I}_N\end{bmatrix}=\boldsymbol{Q}\boldsymbol{R}=\begin{bmatrix}\dot{\boldsymbol{Q}}\\ \ddot{\boldsymbol{Q}}\end{bmatrix}\boldsymbol{R}$, where $\boldsymbol{Q}\in\mathbb{C}^{(N_d+N)\times N}$ consists of orthonormal columns and its first N_d rows and last N rows are denoted by $\dot{\boldsymbol{Q}}$ and $\ddot{\boldsymbol{Q}}$, respectively, and $\boldsymbol{R}\in\mathbb{C}^{N\times N}$ is upper-triangular with positive diagonal elements. From Theorem 1 in [11], along with $\boldsymbol{D}_{\boldsymbol{R}}\triangleq\mathrm{diag}([\boldsymbol{R}]_{1,1},\cdots,[\boldsymbol{R}]_{N,N})$, the optimal $\boldsymbol{W}$ and $\boldsymbol{U}$ are given by $\boldsymbol{W}=\sqrt{1/\sigma^2}\dot{\boldsymbol{Q}}\boldsymbol{D}_{\boldsymbol{R}}^{-H}$ and $\boldsymbol{U}=\boldsymbol{D}_{\boldsymbol{R}}^{-1}\boldsymbol{R}$.

Thirdly, with fixed $\{\theta_m\}$ and $\{\boldsymbol{W},\boldsymbol{U}\}$, we optimize $\{\boldsymbol{B}_k\}$ via solving the problem as below.

$$\min_{\{\boldsymbol{B}_k\}} \ \mathrm{tr}\big((\boldsymbol{F}\boldsymbol{B}-\boldsymbol{U})(\boldsymbol{F}\boldsymbol{B}-\boldsymbol{U})^H\big) \quad (19)$$

$$\text{s.t. } \mathrm{tr}\big(\boldsymbol{B}_k\boldsymbol{B}_k^H\big)\leqslant p_k, \quad k=1,\cdots,K \quad (20)$$

TABLE I
PROCEDURES OF THE RIS(-b)-DFE ALGORITHM

1) Set $n=0$. Initialize $\{\boldsymbol{W},\boldsymbol{U}\}^{\langle n\rangle}$ and $\{\boldsymbol{B}_k\}^{\langle n\rangle}$ as $\boldsymbol{W}=(1)_{N_d\times N}$, $\boldsymbol{U}=\boldsymbol{I}_N$, and $\boldsymbol{B}_k=\left[\sqrt{p_k/N_k}\boldsymbol{I}_{N_k},(0)_{N_k\times(N_{s,k}-N_k)}\right]^T$, $k=1,\cdots,K$.
2) Obtain $\{\theta_m\}^{\langle n+1\rangle}$ with fixed $\{\boldsymbol{W},\boldsymbol{U}\}^{\langle n\rangle}$ and $\{\boldsymbol{B}_k\}^{\langle n\rangle}$ by solving the problem (7)–(8) for either continuous or discrete RIS phase shifts.
3) Obtain $\{\boldsymbol{W},\boldsymbol{U}\}^{\langle n+1\rangle}$ with fixed $\{\boldsymbol{B}_k\}^{\langle n\rangle}$ and $\{\theta_m\}^{\langle n+1\rangle}$ by computing $\boldsymbol{W}=\sqrt{1/\sigma^2}\dot{\boldsymbol{Q}}\boldsymbol{D}_{\boldsymbol{R}}^{-H}$ and $\boldsymbol{U}=\boldsymbol{D}_{\boldsymbol{R}}^{-1}\boldsymbol{R}$.
4) Obtain $\{\boldsymbol{B}_k\}^{\langle n+1\rangle}$ with fixed $\{\theta_m\}^{\langle n+1\rangle}$ and $\{\boldsymbol{W},\boldsymbol{U}\}^{\langle n+1\rangle}$ by solving the problem (19)–(20).
5) Let $n=n+1$. Continue to Step 2) if $n<20$. Otherwise, end the algorithm.

where $\boldsymbol{F}\triangleq\boldsymbol{W}^H(\boldsymbol{H}_{sd}+\boldsymbol{H}_d\boldsymbol{\Theta}\boldsymbol{H}_s)\in\mathbb{C}^{N\times N_s}$. Now, we further set $\boldsymbol{F}=[\boldsymbol{F}_1,\cdots,\boldsymbol{F}_K]$ and $\boldsymbol{U}=[\boldsymbol{U}_1,\cdots,\boldsymbol{U}_K]$. Here, for $k=1,\cdots,K$, $N_{s,k}$ columns of $\boldsymbol{F}$, from the $\left(\sum_{i=0}^{k-1}N_{s,i}+1\right)$th to the $\left(\sum_{i=0}^{k}N_{s,i}\right)$th one with $N_{s,0}=0$, together make up $\boldsymbol{F}_k\in\mathbb{C}^{N\times N_{s,k}}$, and N_k columns of $\boldsymbol{U}$, from the $\left(\sum_{i=0}^{k-1}N_i+1\right)$th to the $\left(\sum_{i=0}^{k}N_i\right)$th one with $N_0=0$, make up $\boldsymbol{U}_k\in\mathbb{C}^{N\times N_k}$. Thus, we can decompose the problem (19)–(20) into

$$\min_{\boldsymbol{B}_k}\ \mathrm{tr}\left((\boldsymbol{F}_k\boldsymbol{B}_k-\boldsymbol{U}_k)(\boldsymbol{F}_k\boldsymbol{B}_k-\boldsymbol{U}_k)^H\right) \tag{21}$$

$$\text{s.t.}\ \ \mathrm{tr}\left(\boldsymbol{B}_k\boldsymbol{B}_k^H\right)\leqslant p_k \tag{22}$$

for $k=1,\cdots,K$. With the Lagrange multiplier μ_k, the KKT conditions of the above problem are $\mu_k\geqslant 0$, $\mathrm{tr}\left(\boldsymbol{B}_k\boldsymbol{B}_k^H\right)\leqslant p_k$, $\mu_k\left[\mathrm{tr}\left(\boldsymbol{B}_k\boldsymbol{B}_k^H\right)-p_k\right]=0$, and $\left(\boldsymbol{F}_k^H\boldsymbol{F}_k+\mu_k\boldsymbol{I}_{N_{s,k}}\right)\boldsymbol{B}_k=\boldsymbol{F}_k^H\boldsymbol{U}_k$, which can be solved under two possible cases. One case is $\mu_k=0$, where we have $\boldsymbol{B}_k=\left(\boldsymbol{F}_k^H\boldsymbol{F}_k\right)^\dagger\boldsymbol{F}_k^H\boldsymbol{U}_k$ and it is just the optimal solution so long as $\mathrm{tr}\left(\boldsymbol{B}_k\boldsymbol{B}_k^H\right)\leqslant p_k$ is satisfied. Otherwise, we turn to the other case $\mu_k>0$, where we have $\boldsymbol{B}_k=\left(\boldsymbol{F}_k^H\boldsymbol{F}_k+\mu_k\boldsymbol{I}_{N_{s,k}}\right)^{-1}\boldsymbol{F}_k^H\boldsymbol{U}_k$ and it should be substituted into $\mathrm{tr}\left(\boldsymbol{B}_k\boldsymbol{B}_k^H\right)=p_k$ for determining μ_k. Given that $\mathrm{tr}\left(\boldsymbol{B}_k\boldsymbol{B}_k^H\right)$ is a decreasing function with respect to μ_k, we can obtain the optimal μ_k via the bisection method [18].

So far, one iteration of our nonlinear receiver-based algorithm has been completed. Hereafter, such iterative algorithm is called "the RIS-DFE algorithm" for continuous RIS phase shifts or "the RIS-b-DFE algorithm" for b bits of discrete RIS phase shifts, whose procedures are listed in Table I with the output of the nth iteration marked by superscript $\langle n\rangle$. When the linear MMSE receiver is used, so long as we replace $\boldsymbol{W}$ with $\bar{\boldsymbol{W}}$, which is also initialized by $(1)_{N_d\times N}$, and set $\boldsymbol{U}=\boldsymbol{I}_N$, those optimization processes of $\{\theta_m\}$ and $\{\boldsymbol{B}_k\}$ remain unaffected. For optimizing $\bar{\boldsymbol{W}}$ with fixed $\{\theta_m\}$ and $\{\boldsymbol{B}_k\}$, by referring to Theorem 4 in [11], the optimal $\bar{\boldsymbol{W}}$ is derived as $\bar{\boldsymbol{W}}=\left(\boldsymbol{A}\boldsymbol{A}^H+\sigma^2\boldsymbol{I}_{N_d}\right)^{-1}\boldsymbol{A}$. Such linear receiver-based iterative algorithm is hereafter called "the RIS(-b)-L algorithm".

When the primal-dual path-following interior-point method is employed to solve the SDP problem (17)–(18) [13], via utilizing the analytical approach in [11], we are able to obtain the computational complexity order (CCO) for one iteration of the RIS-L/DFE algorithm as $\mathcal{O}\left(M^{4.5}+M^4N_m^2+N_m^3+TM^3\right)$, where $N_m\triangleq\max\{N,N_s,N_d\}$, and the CCO for one iteration of the RIS-b-L/DFE algorithm as $\mathcal{O}\left(M^{4.5}+M^4N_m^2+N_m^3+TM^3+2^bTM\right)$. From simulation tests, after each iteration of the algorithms proposed above, the value of the objective, E_s or $\bar{E}_s$, monotonically decreases, which is also bounded below by 0, so the convergence of our algorithms is guaranteed due to Proposition A.3 in [12]. Besides, it is recommended to execute the iterations of our algorithms for 20 times, since after that amount, the performance gains are hardly visible.

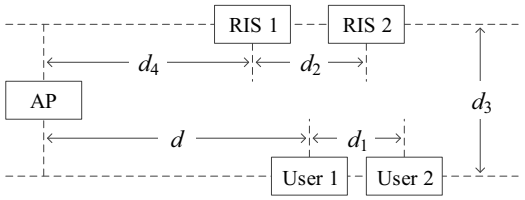

Fig. 2. System architecture configuration for numerical simulations.

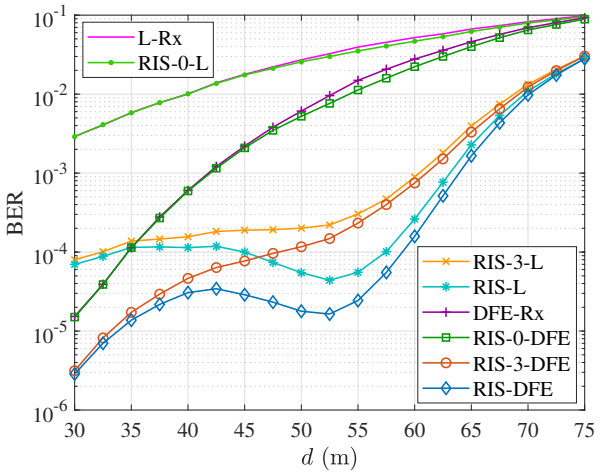

Fig. 3. BER versus d performance comparisons with $\tilde{M}=80$.

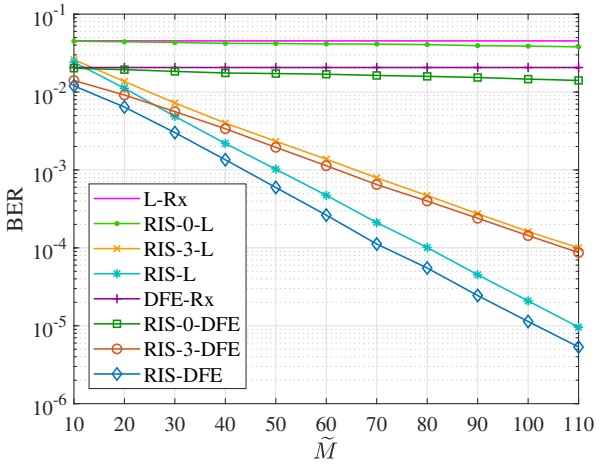

Fig. 4. BER versus $\tilde{M}$ performance comparisons with $d=57.5$ m.

IV. NUMERICAL SIMULATIONS

This section verifies the bit-error-rate (BER) performance of the proposed algorithms via Monte Carlo simulations. Here, we make comparisons among (a) the RIS(-b)-DFE and RIS(-b)-L algorithms, (b) the nonlinear and linear receiver-based algorithms that do not optimize the RIS phase shifts and fix $\theta_m=0$ for $m=1,\cdots,M$, called "the RIS-0-DFE algorithm" and "the RIS-0-L algorithm", and (c) the nonlinear and linear receiver-based algorithms without the assistance of RISs, i.e., fixing $\boldsymbol{\Theta}=(0)_{M\times M}$, called "the DFE-Rx algorithm" and "the L-Rx algorithm", where the algorithms (b) and (c) follow the

same way to optimize $\{\boldsymbol{B}_k\}$ and $\{\boldsymbol{W}, \boldsymbol{U}\}$ or $\bar{\boldsymbol{W}}$ as in Section III and also execute their iterations for 20 times to ensure the fairness of comparisons. All the numerical simulation results are acquired through the average of 1000 independent channel realizations, and for each of them, half a million information bits per data stream are transmitted from every user with the QPSK modulation mode. Besides, in the MMSE-DFE receiver, the symbols fed back are the practical ones affected by the phenomenon of error propagation, which are regenerated from the information bits detected before with detection errors.

Similar to [3]–[5], the system architecture configuration for our simulations is illustrated by Fig. 2 with $K = R = 2$, where there are two users located $d_1 = 2.5\,\mathrm{m}$ apart on a horizontal line, which is parallel to the line that has two RISs located $d_2 = 5\,\mathrm{m}$ apart, and the AP is equidistant from these two lines which are vertically $d_3 = 8\,\mathrm{m}$ apart. In addition, the horizontal distance between the AP and the first RIS is $d_4 = 55\,\mathrm{m}$, and that between the AP and the first user is denoted by d. Like in [1], [7], [8], we set the small-scale fading of the channels $\boldsymbol{H}_{sd,k}$, $\boldsymbol{H}_{s,rk}$, and $\boldsymbol{H}_{d,r}$ for $k = 1, \cdots, K$ and $r = 1, \cdots, R$ to follow the Rayleigh fading with their elements independently generated from $\mathcal{CN}(0, 1)$. The large-scale fading is modelled by $\Gamma(\tilde{d}) = Q_0(\tilde{d}/d_0)^{-\tilde{\gamma}}$ where $Q_0 = -30\,\mathrm{dB}$ stands for the path loss at the reference distance $d_0 = 1\,\mathrm{m}$, $\tilde{d}$ is the distance of signal propagation, and $\tilde{\gamma}$ represents the path loss exponent (PLE) [4], [5], [9]. Here, due to numerous obstacles, the direct user-AP links are assumed to have severe signal attenuation, so we set the PLEs of $\boldsymbol{H}_{sd,k}$ as $\gamma_{sd,k} = \gamma_{sd} = 3.7$. Meanwhile, the user-RIS links are assumed to be in good condition with the PLEs of $\boldsymbol{H}_{s,rk}$ set as $\gamma_{s,rk} = \gamma_s = 2.6$. Since the RISs are usually placed to ensure high-quality RIS-AP links, we set the PLEs of $\boldsymbol{H}_{d,r}$ as $\gamma_{d,r} = \gamma_d = 2.4$. Furthermore, we also set $N_k = \tilde{N} = 4$, $N_{s,k} = \tilde{N}_s = 4$, $p_k = P = 13\,\mathrm{dBm}$, and $M_r = \tilde{M}$, together with $N_d = 8$, $\sigma^2 = -90\,\mathrm{dBm}$, and $b = 3$.

For all the considered algorithms, Fig. 3 exhibits their BER performance with $\tilde{M} = 80$ and $d \in [30\,\mathrm{m}, 75\,\mathrm{m}]$, additionally, Fig. 4 exhibits that with $d = 57.5\,\mathrm{m}$ and $\tilde{M} \in [10, 110]$. It can be seen that the nonlinear receiver-based algorithms perform better than their linear receiver-based counterparts. Compared to the algorithms without using the RISs, the algorithms that fix $\{\theta_m\}$ to be zeros can merely provide little gain, while those algorithms that have $\{\theta_m\}$ optimized can significantly improve the performance, among which, the algorithms with continuous $\{\theta_m\}$ outperform their counterparts with discrete $\{\theta_m\}$. In Fig. 3, as d increases, the BER performance of all algorithms gets worse in general due to the growth of path loss in direct links. However, when d ranges from 42.5 m to 52.5 m, the performance of the RIS-L/DFE algorithm becomes better, which is because the users are approaching the RISs that enhance the quality of communications via the optimization of continuous $\{\theta_m\}$. In Fig. 4, since the L/DFE-Rx algorithm does not utilize the RISs, its performance is unrelated to $\tilde{M}$, while the performance of the other algorithms becomes better as $\tilde{M}$ increases. When $\tilde{M} \geqslant 100$, compared to the L/DFE-Rx and RIS-0-L/DFE algorithms, the RIS-3-L/DFE algorithm can improve the performance by nearly two orders of magnitude, and such superiority expands to three orders of magnitude for the RIS-L/DFE algorithm.

V. Conclusion

Aiming at enhancing the performance of multiuser uplink MIMO communications via multiple RISs, this letter developed iterative algorithms to jointly optimize the continuous or discrete RIS phase shifts, the MMSE or MMSE-DFE receiving matrices, and the source precoding matrices. Simulations confirmed the effectiveness of our proposed algorithms.

References

[1] M. A. ElMossallamy, H. Zhang, L. Song, K. G. Seddik, Z. Han, and G. Y. Li, "Reconfigurable intelligent surfaces for wireless communications: Principles, challenges, and opportunities," *IEEE Trans. Cognitive Commun. Netw.*, vol. 6, no. 3, pp. 990–1002, Sep. 2020.

[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] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design," in *Proc. 2018 IEEE Global Commun. Conf. (GLOBECOM)*, Abu Dhabi, United Arab Emirates, Dec. 9–13, 2018, pp. 1–6.

[4] 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.

[5] 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.

[6] H. Guo, Y.-C. Liang, J. Chen, and E. G. Larsson, "Weighted sum-rate maximization for reconfigurable intelligent surface aided wireless networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3064–3076, May 2020.

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

[8] W. Yan, X. Yuan, and X. Kuai, "Passive beamforming and information transfer via large intelligent surface," *IEEE Wireless Commun. Lett.*, vol. 9, no. 4, pp. 533–537, Apr. 2020.

[9] W. Yan, X. Yuan, Z.-Q. He, and X. Kuai, "Passive beamforming and information transfer design for reconfigurable intelligent surfaces aided multiuser MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1793–1808, Aug. 2020.

[10] X. Yu, D. Xu, Y. Sun, D. W. K. Ng, and R. Schober, "Robust and secure wireless communications via intelligent reflecting surfaces," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2637–2652, Nov. 2020.

[11] Y. Lv, Z. He, and Y. Rong, "Two-way AF MIMO multi-relay system design using MMSE-DFE techniques," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 389–405, Jan. 2021.

[12] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, Massachusetts: Athena Scientific, 1999.

[13] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.

[14] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. New York, USA: Cambridge Univ. Press, 2013.

[15] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge Univ. Press, 2004, sec. 5.5.3, pp. 243–246.

[16] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, UK: Cambridge Univ. Press, 1991.

[17] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.2," CVX Research, Inc., Jan. 2020. [Online]. Available: http://cvxr.com/cvx

[18] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. Cambridge, UK: Cambridge Univ. Press, 2007.