Reconfigurable Intelligent Surface Empowered Downlink Non-Orthogonal Multiple Access

Min Fu, *Student Member, IEEE*, Yong Zhou, *Member, IEEE*, Yuanming Shi, *Senior Member, IEEE*, and Khaled B. Letaief, *Fellow, IEEE*

***Abstract*— Power-domain non-orthogonal multiple access (NOMA) has become a promising technology to exploit the new dimension of the power domain to enhance the spectral efficiency of wireless networks. However, most existing NOMA schemes rely on the strong assumption that users' channel gains are quite different, which may be invalid in practice. To unleash the potential of power-domain NOMA, we propose a reconfigurable intelligent surface (RIS)-empowered NOMA scheme to introduce desirable channel gain differences among the users by adjusting the phase shifts at the RIS. Our goal is to minimize the total transmit power by jointly optimizing the beamforming vectors at the base station, the phase-shift matrix at the RIS, and user ordering. To address challenge due to the highly coupled optimization variables, we present an alternating optimization framework to decompose the non-convex bi-quadratically constrained quadratic problem under a specific user ordering into two rank-one constrained matrices optimization problems via matrix lifting. To accurately detect the feasibility of the non-convex rank-one constraints and improve performance by avoiding early stopping in the alternating optimization procedure, we equivalently represent the rank-one constraint as the difference between nuclear norm and spectral norm. A difference-of-convex (DC) algorithm is further developed to solve the resulting DC programs via successive convex relaxation, followed by establishing the convergence of the proposed DC-based alternating optimization method. We further propose an efficient user ordering scheme with closed-form expressions, considering both the channel conditions and users' target data rates. Simulation results validate the ability of an RIS in enlarging the channel-gain difference when the users' original channel conditions are similar and the superiority of the proposed DC-based alternating optimization method in reducing the total transmit power.**

***Index Terms*— Reconfigurable intelligent surface, non-orthogonal multiple access, joint beamformer and phase-shift matrix design, difference-of-convex programming.**

Manuscript received August 6, 2020; revised January 14, 2021; accepted March 9, 2021. Date of publication March 17, 2021; date of current version June 16, 2021. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 62001294. This article has been presented in part at the IEEE Globecom Workshops, Waikoloa, Hawaii, December 2019. The associate editor coordinating the review of this article and approving it for publication was L. Song. *(Corresponding author: Yuanming Shi.)*

Min Fu is with the School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China, also with the Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 200050, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: fumin@shanghaitech.edu.cn).

Yong Zhou is with the School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China (e-mail: zhouyong@shanghaitech.edu.cn).

Yuanming Shi is with the School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China, and also with Yoke Intelligence, Shanghai 201210, China (e-mail: shiym@shanghaitech.edu.cn).

Khaled B. Letaief is with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong, and also with the Peng Cheng Laboratory, Shenzhen 518066, China (e-mail: eekhaled@ust.hk).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCOMM.2021.3066587.

Digital Object Identifier 10.1109/TCOMM.2021.3066587

I. Introduction

With the upsurge of diversified wireless services and applications such as Internet of Things (IoT) and mobile Internet, various innovative technologies are expected to keep pace with the exponential growth of the mobile data traffic generated by billions of connected devices in the fifth generation (5G) networks and beyond [2]. Therein and to meet the demand of enormous data traffic, the design of appropriate multiple access techniques has been under intense consideration in both academia and industry [3]. Power-domain non-orthogonal multiple access (NOMA) is recognized as a key enabling technology that enables the base station (BS) to simultaneously serve multiple users in the same physical resource block (e.g., time and frequency), thereby significantly improving the spectral efficiency and connection density [3]–[5]. The main idea of downlink power-domain NOMA is that the BS applies superposition coding at the transmitter with the transmit powers as weight factors while each user performing successive interference cancellation (SIC) at the receiver to remove co-channel interference from the received signal before decoding its own signal [6], [7].

However, most of the existing studies on NOMA assumed that users' channel gain are quite different [8]. For example, the authors in [9] assumed the NOMA user pair includes a user located close to the BS and the other far from the BS. The BS allocates a higher transmit power to the user with worse channel condition. As demonstrated in [10], an appropriate users' channel-gain difference is important to unleash the potential of NOMA. In particular, the performance gain of NOMA over orthogonal multiple access (OMA) is limited if there are small channel gain differences among the users. Unfortunately, the simultaneously served users in NOMA networks may not always have diverse channel conditions in practical scenarios because this depends on the propagation environment (e.g., path loss), which are uncontrollable. For instance, in the Internet of vehicles scenarios [11], multiple users with diversified quality of service (QoS) requirements need to be simultaneously served by the BS even if their channel conditions are similar.

Fortunately, with the theoretical and experimental breakthroughs in micro electromechanical systems and metamaterials (e.g., metasurface) [12], reconfigurable intelligent surface (RIS), as an emerging cost-effective technology, has recently been proposed as a powerful solution to enhance the spectrum-efficiency and energy-efficiency of wireless networks [13], [14]. In particular, the RIS is implemented as an array of low-cost scattering elements, each of which being able to induce an adjustable phase shift to the incident signal to be reflected, thereby reconfiguring the reflected signal propagations [15]. Different from amplify-and-forward (AF) relay, backscatter communication, and active intelligent surface based massive multiple-input multiple-output (MIMO), the RIS operates in full-duplex mode and only reflects the received signals as a passive array without the need of any transmit radio-frequency (RF) chains to provide spectrum-efficient and cost-effective communications [14], [15]. This, thus, motivates the study of RIS-empowered downlink NOMA transmission, where the RIS is capable of inducing desirable channel differences among the users to enhance the performance of NOMA.

Energy-efficient communication is a critical design aspect for future wireless networks [16]. Besides, user ordering has a significant impact on power consumption in NOMA networks. In this paper, we will focus on the power minimization problem for an RIS-empowered downlink multi-user multiple-input single-output (MISO) NOMA network by jointly optimizing both the transmit beamforming vectors at the BS, the phase-shift matrix at the RIS, and user ordering by taking into account the users' target data rates. The unique challenges of the power minimization problem arise from the non-convex bi-quadratic QoS constraints (due to the highly coupled transmit beamforming vectors and the phase-shift matrix), the non-convex modulus constraints (due to the RIS hardware setup), and the challenging user ordering design in RIS-empowered multi-user MISO NOMA networks. Generally, in NOMA networks without RIS, the users are usually ordered based on their channel conditions with respect to the BS. However, in RIS-empowered NOMA networks, the RIS-related channels and the RIS configuration based phase-shift matrix are coupled with each other. Due to the diverse data rate requirements as well as the combined channel conditions among users, the design of user ordering becomes much more complicated.

A. Contributions

To address the aforementioned unique challenges, we shall develop an effective optimization framework for the total power minimization problem, followed by proposing a low-complexity user ordering method in the RIS-empowered multi-user MISO NOMA networks. The main contributions of this paper are summarized as follows.

- This paper is one of the early attempts to study the transmit power minimization problem in RIS-empowered multi-user MISO NOMA networks, where an RIS is deployed to introduce appropriate channel gain differences among users, thereby unleashing the potential of NOMA without relying on the assumption of the users' diverse channel conditions. Moreover, we propose an efficient user ordering scheme with a closed-form ordering criterion by taking into account both the combined channel conditions and the target data rates.
- To support efficient algorithm design, we adopt the alternating optimization method to decompose the original bi-quadratically constrained quadratic problem into two subproblems with non-convex quadratic constraints, i.e., a non-convex quadratically constrained quadratic programming (QCQP) subproblem for optimizing transmit beamforming vectors and a non-convex QCQP feasibility subproblem for optimizing the phase-shift matrix.
- We develop a unified difference-of-convex (DC) method to solve the aforementioned subproblems with the capability of accurately detecting the feasibility of non-convex quadratic constraints for the transmit beamforming vectors and phase-shift matrix design, which can avoid early stopping in the alternating optimization procedure, thereby considerably improving the performance compared with the state-of-the-art methods. The main idea is to reformulate the resultant non-convex QCQP subproblems as multiple rank-one constrained matrices optimization problems via matrix lifting, followed by equivalently representing the rank-one constraint as the difference between the nuclear norm and the spectral norm.
- We further present an efficient DC algorithm for the resulting non-convex DC programs via successive convex relaxation. By representing the objective functions of the resultant DC programs as the difference of two strongly convex functions, we prove that the DC algorithm converges to the stationary solution for DC programs and that the DC-based alternating optimization method (namely, alternating DC method) always converges.

Simulation results demonstrate the ability of an RIS in enlarging the channel-gain difference when the users' original channel conditions are similar, leading to transmit power reduction. Simulation results still validate the superiority of the proposed alternating DC method in reducing the total transmit power. Besides, the proposed user ordering criterion is a good option for large-size networks, which provides comparable performance to the exhaustive search.

B. Related Works

1) RIS-Empowered OMA Networks: The research on RIS-empowered wireless networks has recently received considerable attention in vast applications, e.g., coverage extension [17], energy-efficient beamforming [18]–[20], physical layer security [21], [22], and massive connectivity [23]. For coverage extension, the objective of RIS is to create indirect links with the BS and user, in the scenario where there is no direct link between the BS and users, or the direct link is severely blocked by obstacles. Therein, [17] demonstrated that RIS-based networks achieved higher energy efficiency than conventional AF relays. The RIS is deployed to assist the efficient beamforming design that compensates for the signal attenuation from the BS or co-channel interference from neighboring BS by jointly optimizing the active beamformer

at BS and the phase-shift matrix at RIS. In particular, [18] showed that an RIS-empowered MISO system significantly reduced the transmit power consumption compared with the system without using an RIS, and achieved the same rate performance as a conventional massive MIMO system, but with considerably decreased active antennas/RF chains. [19] also showed that the RIS-empowered network outperformed the system without using RIS in terms of the weighted sum rate. Besides, the author in [20] applied RIS to reduce the distortion between the decoding signal and the ground-truth signal for over-the-air computation. The use of RIS for improving physical layer security aims to cancel out the signal from the BS at the eavesdropper. Specifically, [21] validated that deploying large-scale RISs achieved higher significant performance gains than increasing the number of the antenna at the transmitter in terms of secrecy rate and energy efficiency for a simple scenario with one legitimate receiver and one eavesdropper. [22] further confirmed that the security enhancement provisioning for RIS-empowered networks to keep the signals secret from multiple eavesdroppers in a broadcast system. The author in [23] considered an RIS-empowered IoT network to support massive connectivity under the grant-free random access protocol.

2) RIS-Empowered NOMA Networks: Upon the completion of this work, the application of RIS in NOMA networks was investigated in some parallel works [24]–[30]. Specifically, the authors in [24] analyzed the transmission reliability of RIS-aided NOMA transmission. For a multi-user NOMA network, the authors in [25] optimized the transmit beamforming vectors at the BS and the phase shifts at RIS to enhance the user fairness. In addition, the authors in [26] considered an RIS-empowered downlink MISO NOMA network to minimize the power consumption, where zero-forcing precoding was employed at the BS to cancel the inter-pair interference. In [27], the authors studied the power efficiency of RIS-empowered MISO NOMA system under additional quasi-degraded channels constraints with two users' case. In [28], the authors investigated the sum rate maximization problem in RIS-empowered NOMA networks. The authors in [29] focused on the maximization of the system throughput over the channel assignment, power allocation, and ideal reflection coefficient in a single-input single-output (SISO) NOMA network. In [30], the author proposed a low complexity machine learning approach to maximize the energy efficiency of RIS-empowered systems by jointly designing the deployment policy and phase shift policy of RISs while considering the time-varying data demand of users.

C. Organization and Notations

The remainder of this paper is organized as follows. Section II describes the system model and problem formulation. We present an alternating DC method to solve the power minimization problem under a specific user ordering in Section III. Section IV proposes a low-complexity user ordering scheme with closed-form expressions. Section V presents the numerical results and Section VI concludes this paper.

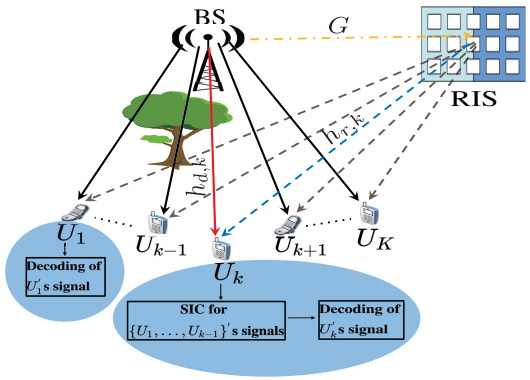


Fig. 1. An RIS-empowered downlink MISO NOMA network with K users. User U_k is allocated the k-th highest transmit power, and has to decode and remove the signals intended for users $U_1, \ldots, U_{k-1}$ before decoding its own signal.

Notations: $\mathbb{E}(\cdot)$ denotes the statistical expectation. $(\cdot)^{\mathsf{H}}$ and $(\cdot)^{\mathsf{T}}$ denote the conjugate transpose and transpose, respectively. For a complex-valued vector $\boldsymbol{x}$, $\|\boldsymbol{x}\|$ denotes its Euclidean norm and $\text{diag}(\boldsymbol{x})$ denotes a diagonal matrix with each diagonal entry being the corresponding element in $\boldsymbol{x}$. $[\boldsymbol{x}]_{(1:N)}$ denotes the first N elements of vector $\boldsymbol{x}$. For a matrix $\boldsymbol{X}$, $\|\boldsymbol{X}\|_F$, $\|\boldsymbol{X}\|_*$, and $\|\boldsymbol{X}\|_2$ denote its the Frobenius norm, the nuclear norm and the spectral norm, respectively. j denotes the imaginary unit. Finally, $\Re(\cdot)$ denotes the real part of a complex number.

II. System Model and Problem Formulation

A. System Model

As shown in Fig. 1, we consider an RIS-empowered multi-user MISO power-domain NOMA system, where an RIS with N passive reflecting elements is deployed to assist the data transmission from an M-antenna BS to K single-antenna users by providing additional channel paths to introduce diverse channel conditions between the users and the BS. To account for the increasing number of users and the limited spectrum resource, we consider fully- and over-loaded scenarios, where the number of users is not smaller than the number of antennas at the BS, i.e., $K \geq M$. We denote $s_k \in \mathbb{C}$ and $\boldsymbol{w}_k \in \mathbb{C}^M$ as the signal and linear beamforming vector at the BS for user U_k, respectively, where $k \in \mathcal{K}$ with $\mathcal{K} = \{1, 2, \ldots, K\}$. Without loss of generality, signal s_k is assumed to have zero mean and unit variance, i.e., $\mathbb{E}[s_k s_k^{\mathsf{H}}] = 1, \forall\, k \in \mathcal{K}$. Under universal frequency reuse, the BS superimposes and transmits the signals intended for K users. After the reflection of the RIS, the signal received at user U_k is given by

$$y_k = \left(\boldsymbol{h}_{r,k}^{\mathsf{H}} \boldsymbol{\Theta} \boldsymbol{G} + \boldsymbol{h}_{d,k}^{\mathsf{H}}\right) \sum_{j=1}^{K} \boldsymbol{w}_j s_j + e_k, \quad \forall\, k \in \mathcal{K}, \tag{1}$$

where $\boldsymbol{h}_{d,k} \in \mathbb{C}^M, \boldsymbol{G} \in \mathbb{C}^{N \times M}$, and $\boldsymbol{h}_{r,k} \in \mathbb{C}^N$ denote the channel responses from the BS to user U_k, from the BS to the RIS, and from the RIS to user U_k, respectively. Likewise, $e_k \sim \mathcal{CN}(0, \sigma^2)$ is the additive white Gaussian noise (AWGN) with σ^2 being the noise power. Note that quasi-static flat fading model is considered for all channels. In addition, $\boldsymbol{\Theta} = \text{diag}(\beta e^{j\theta_1}, \ldots, \beta e^{j\theta_N}) \in \mathbb{C}^{N \times N}$ denotes the

diagonal phase-shift matrix of the RIS, where $\theta_n \in [0, 2\pi), \forall n$ and $\beta \in [0, 1]$ denote the phase shift of element n and the amplitude reflection coefficient on the incident signal, respectively. As each element on the RIS is designed to boost the received signals, we assume $\beta = 1$ without loss of generality, similar to [17]–[22], [24]–[28]. Due to the severe path loss, the power of the signals that are reflected by the RIS two or more times is assumed to be negligible [17]–[29]. In this paper, to characterize the performance gain achieved by RIS, we study the transmit power minimization problem and under the assumption that the CSI of all channels is perfectly known, as in most of the exiting studies on resource allocation in RIS-assisted systems [17]–[22], [24]–[29]. Although it is generally difficult to obtain perfect CSI, various efficient channel estimation methods proposed for RIS-assisted wireless networks can be adopted to provide accurate CSI the channels, e.g., brute-force method [31], compressive-sensing based method [32], and deep learning based method [33]. Note that the proposed algorithm developed in this paper with perfect CSI not only serves as the performance upper bound for the practical scenarios with imperfect CSI,[1] but also provides useful insights on the design of RIS-empowered NOMA schemes.

The SIC decoding order is an essential issue in NOMA systems, where each user sequentially performs SIC to remove the interference in a specific order before decoding its own signal. It has been shown in [10] that the decoding order plays a critical role in determining the overall system performance. In SISO NOMA networks without RIS, the users are usually ordered based on their channel conditions with respect to the BS. In MISO NOMA networks without RIS, the user ordering becomes more complicated and is determined by the product of the channel gain and the beamforming gain of each user [39]. Furthermore, in RIS-empowered MISO NOMA networks, obtaining the optimal user ordering is further complicated, as the concatenated channel response $(\boldsymbol{h}_{r,k}^{\mathsf{H}}\boldsymbol{\Theta}\boldsymbol{G} + \boldsymbol{h}_{d,k}^{\mathsf{H}})$ depends not only on $\boldsymbol{h}_{r,k}$ but also on the coupled matrix (i.e., $\boldsymbol{G}$, $\boldsymbol{h}_{d,k}$, and $\boldsymbol{\Theta}$) while the phase-shift matrix simultaneously affects the concatenated channel strengths of all users. Compared to the traditional analog beamforming in mmWave NOMA system, the reflection-dominated propagation and the RIS configuration based phase-shift matrix are coupled with each other in RIS-aided NOMA networks. Furthermore, we study the transmit power minimization problem in NOMA networks, taking into account both the constant-modulus constraint of phase shifters at the RIS and the QoS constraints of all users, leading to a more complex phase-shifter design at the RIS. Therefore, the traditional analog beamforming schemes with a separate channel matrix and analog beamformer cannot be applied. Therefore, the optimal user ordering may be any one of the $K!$ different decoding orders for K users. We denote π as the decoding ordering operator. In particular, $\pi(k)$ denotes the user index whose signal is the k-th signal (i.e., $s_{\pi(k)}$) to be decoded by users $U_{\pi(l)}$ for all $1 \le k \le l \le K$. For instance, if $\pi(k) = i$, then the signal intended for user U_i is the k-th signal to be decoded by users $U_{\pi(l)}$ for all $l \ge k$. According to the NOMA decoding strategy, user $U_{\pi(l)}$, $\forall l > 1$ needs to sequentially decode and remove the co-channel interference introduced by the signals intended for $U_{\pi(j)}$ for all $j < l$, i.e., $\{s_{\pi(1)}, \ldots, s_{\pi(l-1)}\}$, before successfully decoding its own signal, whereas the signals intended for other users are treated as noise. Specifically, for $1 \le k \le l \le K$, after removing signals $\{s_{\pi(1)}, \ldots, s_{\pi(k-1)}\}$, the remaining signal at user $U_{\pi(l)}$ can be expressed as

$$y_{\pi(l)}^{\pi(k)} = \left(\boldsymbol{h}_{r,\pi(l)}^{\mathsf{H}}\boldsymbol{\Theta}\boldsymbol{G} + \boldsymbol{h}_{d,\pi(l)}^{\mathsf{H}}\right)\sum_{j=k}^{K}\boldsymbol{w}_{\pi(j)}s_{\pi(j)} + e_{\pi(l)}. \quad (2)$$

According to (2), the achievable SINR for user $U_{\pi(l)}$ to decode the signal intended for $U_{\pi(k)}$ can be expressed as

$$\mathrm{SINR}_{\pi(l)}^{\pi(k)} = \frac{|(\boldsymbol{h}_{r,\pi(l)}^{\mathsf{H}}\boldsymbol{\Theta}\boldsymbol{G} + \boldsymbol{h}_{d,\pi(l)}^{\mathsf{H}})\boldsymbol{w}_{\pi(k)}|^2}{\sum_{j=k+1}^{K}|(\boldsymbol{h}_{r,\pi(l)}^{\mathsf{H}}\boldsymbol{\Theta}\boldsymbol{G} + \boldsymbol{h}_{d,\pi(l)}^{\mathsf{H}})\boldsymbol{w}_{\pi(j)}|^2 + \sigma^2}, \quad \forall l \ge k. \quad (3)$$

In other words, to ensure successful SIC for a specific user ordering π, the signal intended for user $U_{\pi(k)}$ should be successfully decoded at user $U_{\pi(l)}$ for all $l \ge k$. To successfully decode signal $s_{\pi(k)}$ at user $U_{\pi(l)}$ for all $l \ge k$, the achievable SINR of decoding signal $s_{\pi(k)}$ at user $U_{\pi(l)}$ for all $k \le l$ should be no smaller than the predefined target SINR $\gamma_{\pi(k)}$ of signal $s_{\pi(k)}$, which can be mathematically expressed as

$$\min_{l\in[k,K]} \mathrm{SINR}_{\pi(l)}^{\pi(k)} \ge \gamma_{\pi(k)}, \forall k. \quad (4)$$

This criterion is widely adopted in existing works investigating the power minimization on NOMA [40]–[42]. Therefore, the achievable rate $R_{\pi(k)}$ for user $U_{\pi(k)}$ can be expressed as

$$R_{\pi(k)} = \log_2\left(1 + \min_{l\in[k,K]} \mathrm{SINR}_{\pi(l)}^{\pi(k)}\right), \quad \forall k, \quad (5)$$

where the channel bandwidth is normalized to 1.

B. Problem Formulation

In this subsection, we formulate a total transmit power minimization problem by jointly optimizing the beamforming vectors (i.e., $\{\boldsymbol{w}_k, k \in \mathcal{K}\}$) at the BS, the phase-shift matrix (i.e., $\boldsymbol{\Theta}$) at the RIS, and user ordering (i.e., π) by taking into account the data rate requirements of all users and the unit modulus constraints of all reflecting elements. The total transmit power minimization problem is formulated as

$$\mathscr{P}: \ \underset{\{\boldsymbol{w}_{\pi(k)}\},\boldsymbol{\Theta},\pi}{\text{minimize}} \sum_{k=1}^{K}\|\boldsymbol{w}_{\pi(k)}\|^2$$
$$\text{subject to } \log_2\left(1+\min_{l\in[k,K]} \mathrm{SINR}_{\pi(l)}^{\pi(k)}\right) \ge R_{\pi(k)}^{\min}, \quad \forall k, \quad (6)$$
$$|\boldsymbol{\Theta}_{n,n}| = 1, \quad \forall n, \quad (7)$$
$$\pi \in \mathcal{S}, \quad (8)$$

where $\pi = \{\pi(1), \ldots, \pi(K)\}$ is the user ordering index vector, $\|\boldsymbol{w}_{\pi(k)}\|^2$ is the transmit power allocated to user $U_{\pi(k)}$,

[1]To account for the imperfect CSI in RIS-assisted systems, the worst-case robust optimization [34]–[36] and the stochastic optimization [19], [37], [38] are two main methods that can be adopted. Such an extension will be left as our future work.

$R^{\min}_{\pi(k)}$ denotes the target data rate of user $U_{\pi(k)}$, and $\mathcal{S}$ denotes the combination set of all possible user orderings. Constraints (6) represent the target data rate requirements of all users and guarantees that SIC can be performed successfully. Constraints (7) denote unit modulus' phase shifters at the RIS.

However, the joint optimization of beamforming vectors, the phase-shift matrix, and user ordering in $\mathscr{P}$ is much more complicated than the power minimization problem in [18], [41] due to the non-convex constraints (6) with highly coupled optimization variables and minimization operator, non-convex unit-modulus constraints (7), and highly challenging user ordering design controlled by the phase-shift matrix. In the following, we will analyze and reformulate $\mathscr{P}$.

C. Problem Analysis

Since the total number of user ordering combinations is a finite value and known via permutation, the required transmit power consumption can be obtained by first solving Problem $\mathscr{P}$ for all possible decoding orders and then selecting the user ordering that minimizes the objective function's value. Therefore, to make Problem $\mathscr{P}$ tractable, we solve this problem in two steps, i.e., obtaining the decoding order and solving a power minimization problem given the obtained decoding order. In particular, we first consider the transmit power minimization problem with a given user ordering π. To assist an efficient algorithm design, we rewrite constraints (6) as

$$\min_{l\in[k,K]} \text{SINR}^{\pi(k)}_{\pi(l)} \geq \gamma^{\min}_{\pi(k)}, \quad \forall\, k, \tag{9}$$

where $\gamma^{\min}_{\pi(k)} = 2^{R^{\min}_{\pi(k)}} - 1$ is the minimum SINR required to successfully decode signal $s_{\pi(k)}$. To eliminate the minimization operator, constraints (9) can be further rewritten as

$$\gamma^{\min}_{\pi(k)} \left(\sum_{j=k+1}^{K} \left| (\boldsymbol{h}^{\mathsf{H}}_{r,\pi(l)} \boldsymbol{\Theta} \boldsymbol{G} + \boldsymbol{h}^{\mathsf{H}}_{d,\pi(l)}) \boldsymbol{w}_{\pi(j)} \right|^2 + \sigma^2 \right) \leq \left| (\boldsymbol{h}^{\mathsf{H}}_{r,\pi(l)} \boldsymbol{\Theta} \boldsymbol{G} + \boldsymbol{h}^{\mathsf{H}}_{d,\pi(l)}) \boldsymbol{w}_{\pi(k)} \right|^2, \quad \forall k, l \geq k. \tag{10}$$

Therefore, given the user ordering π, the transmit power minimization problem is given by

$$\mathscr{P}_1: \ \underset{\{\boldsymbol{w}_{\pi(k)}\}, \boldsymbol{\Theta}}{\text{minimize}} \sum_{k=1}^{K} \|\boldsymbol{w}_{\pi(k)}\|^2 \quad \text{subject to constraints (7), (10),} \tag{11}$$

with the optimal solution denoted by $P^\star(\pi)$.

As a result, the transmit power minimization problem $\mathscr{P}$ can be solved by searching over all possible user orderings in set $\mathcal{S}$, i.e., $P^\star = \min_{\pi\in\mathcal{S}} P^\star(\pi)$.

However, Problem $\mathscr{P}_1$ is still highly intractable due to the non-convex bi-quadratic constraints (10), in which the beamforming vectors and the phase-shift matrix are highly coupled, and the non-convex unit modulus constraints (7). To address these unique challenges, we present an efficient method with algorithmic advantages to solve Problem $\mathscr{P}_1$ in Section III.

Unfortunately, the exhaustive search method needs to solve $K!$ power minimization problems, each of which is for a particular user ordering. This method is exponential in the number of users and is computationally prohibitive for practical applications. To further reduce the computational complexity of the exhaustive search method, we develop an efficient user ordering method in a closed-form by taking into account both the combined channel conditions and the target data rates in Section IV. The required transmit power consumption can be obtained by solving the power minimization Problem $\mathscr{P}_1$ under the proposed user ordering strategy.

We will demonstrate the superiority of the proposed algorithm for the power minimization problem and effectiveness of the proposed user ordering along with the deployment of RIS in NOMA networks in Section V.

III. Alternating DC Method for Beamformers Design

In this section, we develop efficient an algorithm to solve Problem $\mathscr{P}_1$, namely alternating DC method. The main motivation is to employ alternating optimization to decompose the joint optimization Problem $\mathscr{P}_1$ into two non-convex QCQP subproblems to address the highly coupled optimization variables, followed by reformulating the subproblems into DC programmings to avoid early stopping in alternating procedure, thereby achieving performance improvement.

A. Alternating Optimization Framework

Recently, most the existing works, e.g., [17]–[19] applied a powerful alternating optimization to design the active beamforming at the BS and the phase-shift matrix at the RIS, which always yields convex constraints given the phase-shift matrix (i.e., affine constraints [17], second-order cone (SOC) constraints [18], and quadratic constraints [19]). Inspired by this, in this subsection, to decouple the optimization variables, we shall employ alternating optimization to decompose the joint optimization Problem $\mathscr{P}_1$ into two subproblems but with non-convex constraints, i.e., a non-convex QCQP subproblem for transmit beamforming vectors and a non-convex QCQP feasibility subproblem for the phase-shift matrix. For notational ease, we omit the user ordering index π in the sequel. Let k denote the user index whose signal is the k-th signal to be decoded by users U_l for all $l \geq k$.

1) Transmit Beamforming Vectors Optimization: For a given phase-shift matrix $\boldsymbol{\Theta}$, the concatenated channel response $\boldsymbol{h}^{\mathsf{H}}_l = \boldsymbol{h}^{\mathsf{H}}_{r,l} \boldsymbol{\Theta} \boldsymbol{G} + \boldsymbol{h}^{\mathsf{H}}_{d,l} \in \mathbb{C}^{1\times M}$ is fixed, and hence $\mathscr{P}_1$ is simplified as the following non-convex QCQP problem

$$\mathscr{P}_{1.1}: \ \underset{\{\boldsymbol{w}_k\}}{\text{minimize}} \sum_{k=1}^{K} \|\boldsymbol{w}_k\|^2 \quad \text{subject to } \gamma^{\min}_k \left(\sum_{j=k+1}^{K} |\boldsymbol{h}^{\mathsf{H}}_l \boldsymbol{w}_j|^2 + \sigma^2 \right) \leq |\boldsymbol{h}^{\mathsf{H}}_l \boldsymbol{w}_k|^2, \quad \forall\, k, l = k, \ldots, K. \tag{12}$$

While problem (12) appears similar to the beamforming design in [18], it cannot be equivalently transformed into a

second-order cone program (SOCP) since it is impossible to have a phase rotation to simultaneously satisfy $\Re(\boldsymbol{h}_k^{\mathsf{H}}\boldsymbol{w}_k) = \Re(\boldsymbol{h}_l^{\mathsf{H}}\boldsymbol{w}_k) = 0, k < l \leq K$.

2) Phase-Shift Matrix Optimization: Given the beamforming vectors $\{\boldsymbol{w}_k\}$, we denote $b_{l,k} = \boldsymbol{h}_{d,l}^{\mathsf{H}}\boldsymbol{w}_k$ and $\boldsymbol{a}_{l,k} = \mathrm{diag}(\boldsymbol{h}_{r,l}^{\mathsf{H}})\boldsymbol{G}\boldsymbol{w}_k,\ \forall k, l \geq K$. Hence, we have $(\boldsymbol{h}_{r,l}^{\mathsf{H}}\boldsymbol{\Theta}\boldsymbol{G} + \boldsymbol{h}_{d,l}^{\mathsf{H}})\boldsymbol{w}_k = \boldsymbol{v}^{\mathsf{H}}\boldsymbol{a}_{l,k} + b_{l,k}$, where $\boldsymbol{v} = [e^{j\theta_1}, \ldots, e^{j\theta_N}]^{\mathsf{H}}$. Thus, Problem $\mathscr{P}_1$ is simplified as the following homogeneous feasibility detection problem by introducing an auxiliary variable t, given by

$$\begin{aligned}\mathscr{P}_{1.2}:\ \text{Find}\ &\tilde{\boldsymbol{v}}\\ \text{subject to}\ &\gamma_k^{\min}\left(\sum_{j=k+1}^{K}\left(\tilde{\boldsymbol{v}}^{\mathsf{H}}\boldsymbol{R}_{l,j}\tilde{\boldsymbol{v}} + |b_{l,j}|^2\right) + \sigma^2\right)\\ &\leq \tilde{\boldsymbol{v}}^{\mathsf{H}}\boldsymbol{R}_{l,k}\tilde{\boldsymbol{v}} + |b_{l,k}|^2, \quad \forall k, l = k, \ldots, K, \end{aligned} \tag{13}$$

$$|\tilde{\boldsymbol{v}}_n| = 1, \quad \forall n = 1, \ldots, N+1, \tag{14}$$

where $\boldsymbol{R}_{l,k} = \begin{bmatrix}\boldsymbol{a}_{l,k}\boldsymbol{a}_{l,k}^{\mathsf{H}} & \boldsymbol{a}_{l,k}b_{l,k}^{\mathsf{H}}\\ b_{l,k}\boldsymbol{a}_{l,k}^{\mathsf{H}} & 0\end{bmatrix} \in \mathbb{C}^{(N+1)\times(N+1)}$ and $\tilde{\boldsymbol{v}} = \left[\boldsymbol{v}^{\mathsf{H}}, t^{\mathsf{H}}\right]^{\mathsf{H}}$. If we obtain a feasible solution, denoted as $\tilde{\boldsymbol{v}}^\star$, by solving Problem $\mathscr{P}_{1.2}$, then a feasible solution for phase shifter vector can immediately be obtained by setting $\boldsymbol{v}^\star = [\tilde{\boldsymbol{v}}^\star/\tilde{\boldsymbol{v}}_{N+1}^\star]_{(1:N)}$.

3) Discussion: In the subsection, we provide some discussions on the impact of the solution about the above two subproblems.

For Problem $\mathscr{P}_{1.1}$, there exists two state-of-the-arts methods. One exploits the successive convex approximation (SCA) technique to relax the optimization problem as an SOCP problem, which, however, results in a suboptimal solution with performance degradation [41]. To further improve the performance, the semidefinite relaxation (SDR) technique [43] can be exploited by formulating the optimization problem as a semidefinite programming (SDP) form by lifting $\boldsymbol{w}_k$ into a positive semidefinite (PSD) matrix $\boldsymbol{W}_k \in \mathbb{C}^{M\times M}$, where $\boldsymbol{W}_k = \boldsymbol{w}_k\boldsymbol{w}_k^{\mathsf{H}}$ and $\mathrm{rank}(\boldsymbol{W}_k) = 1, \forall k$, followed by dropping rank-one constraints [41], [42]. The optimal beamforming vectors of the Problem $\mathscr{P}_{1.1}$ can be obtained from the SDR method if it yields rank-one solutions.

For Problem $\mathscr{P}_{1.2}$, the state-of-the-art method is to formulate $\mathscr{P}_{1.2}$ as an SDP optimization problem by lifting $\tilde{\boldsymbol{v}}$ into a PSD matrix $\boldsymbol{V} \in \mathbb{C}^{(N+1)\times(N+1)}$, where $\boldsymbol{V} = \tilde{\boldsymbol{v}}\tilde{\boldsymbol{v}}^{\mathsf{H}}$ and $\mathrm{rank}(\boldsymbol{V}) = 1$, followed by dropping the rank-one constraint via the SDR technique, similar to [18].

To solve problems $\mathscr{P}_{1.1}$ and $\mathscr{P}_{1.2}$ with good performance, both subproblems shall be solved by the SDR method in alternating procedure (namely, alternating SDR method). However, such an SDR technique often yields a solution which fails to satisfy the rank-one constraint, especially in the cases where either the dimension of the optimization variable or the number of users is large, as demonstrated in [44], [45]. If so, an optimal beamforming vector cannot be obtained which leads to performance degradation. Furthermore, for the phase-shifter vector, if the SDR method turn out not to be rank-one, the suboptimal solution obtained by the Gaussian randomization technique [43] may not guarantee meeting the original quadratic constraints (14) in $\mathscr{P}_{1.2}$, and thus results in early stopping in the alternating procedure. It is also equally important to note that the existing methods fail to accurately detect the feasibility of rank-one constraints, which may yield performance degradation as suboptimal beamforming vectors and the early stopping in the procedure of alternating optimization.

Based on the above discussions and different from previous works [17]–[19] here we endeavour to address the following coupled challenges to solve the total transmit power minimization problem for RIS-empowered multi-user MISO NOMA networks:

- For the transmit beamforming vectors optimization, we need to efficiently solve a non-convex QCQP optimization problem with the capability of accurately detecting the feasibility of multiple rank-one constraints in the lifted matrix space to obtain the high-quality beamforming vectors, thereby achieving the performance gains;
- For the phase-shift matrix optimization, we need to accurately detect the feasibility of a non-convex quadratically constrained problem, which can avoid the early stopping of the alternating optimization procedure, thereby enhancing the performance compared with the state-of-the-art methods.

In summary, we shall develop an efficient method to accurately detect the feasibility of the non-convex quadratic constraints to push the alternating optimization procedure for the transmit beamforming design and phase shifters design. In the next subsection, to address the limitations of existing methods, we present a unified DC method to solve the aforementioned subproblems.

B. A Unified Difference-of-Convex Programming Method

In this subsection, we propose a unified DC method to the non-convex QCQP subproblems to accurately detect the feasibility of non-convex quadratic constraints for the transmit beamforming vectors optimization and the phase-shift matrix optimization while avoiding the alternative procedure to stop early, thereby achieving the performance improvement.

1) DC Representation for Rank-One Constraint: Firstly, we present an exact DC representation for the rank-one constraint. For a matrix $\boldsymbol{X} \in \mathbb{C}^{N\times N}$, the rank-one constraint can be rewritten as $\left\|[\sigma_1(\boldsymbol{X}), \ldots, \sigma_i(\boldsymbol{X}) \ldots, \sigma_N(\boldsymbol{X})]\right\|_0 = 1$, where $\sigma_i(\boldsymbol{X})$ is the i-th largest singular value of matrix $\boldsymbol{X}$, and $\|\cdot\|_0$ is the l_0-norm of a vector. It is noteworthy that the rank function is a discontinuous function. To reformulate a continuous function, we introduce an exact DC representation for the rank-one constraint in Proposition 1.

Proposition 1: For a PSD matrix $\boldsymbol{X} \in \mathbb{C}^{N\times N}$ with $\mathrm{Tr}(\boldsymbol{X}) > 0$, we have $\mathrm{rank}(\boldsymbol{X}) = 1 \Leftrightarrow \|\boldsymbol{X}\|_* - \|\boldsymbol{X}\|_2 = 0$.

Proof. Please refer to Appendix VI-A. □

It is noteworthy that the DC representation is a continuous function.

2) Proposed Unified DC Programming Method: The main idea of our proposed DC method is to first reformulate Problems $\mathscr{P}_{1.1}$ and $\mathscr{P}_{1.2}$ into matrices optimization problems

via the matrix lifting technique, then apply the DC presentation to accurately detect the rank-one constraints.

Given the phase-shift matrix $\boldsymbol{\Theta}$, by lifting $\boldsymbol{w}_k$ into a PSD matrix $\boldsymbol{W}_k \in \mathbb{C}^{M\times M}$, where $\boldsymbol{W}_k = \boldsymbol{w}_k\boldsymbol{w}_k^{\mathsf{H}}$ and $\mathrm{rank}(\boldsymbol{W}_k) = 1, \forall k$, we reformulate Problem $\mathscr{P}_{1.1}$ into the following DC program to obtain K rank-one matrices

$$\begin{aligned}\mathscr{P}_{1.3}: \ \underset{\{\boldsymbol{W}_k\}}{\text{minimize}} & \sum_{k=1}^{K}\mathrm{Tr}(\boldsymbol{W}_k) + \rho\sum_{k=1}^{K}\Big(\|\boldsymbol{W}_k\|_* - \|\boldsymbol{W}_k\|_2\Big)\\ \text{subject to } & \gamma_k^{\min}\left(\sum_{j=k+1}^{K}\mathrm{Tr}(\boldsymbol{H}_l^{\mathsf{H}}\boldsymbol{W}_j) + \sigma^2\right)\\ & \le \mathrm{Tr}(\boldsymbol{H}_l^{\mathsf{H}}\boldsymbol{W}_k), \quad \forall\, k, l = k,\ldots,K, \quad (15)\\ & \boldsymbol{W}_k \succcurlyeq 0, \quad \forall\, k, \quad (16)\end{aligned}$$

where $\rho > 0$ is a penalty parameter and $\boldsymbol{H}_l = \boldsymbol{h}_l\boldsymbol{h}_l^{\mathsf{H}} \in \mathbb{C}^{M\times M}$. By enforcing the penalty term to be zero, Problem $\mathscr{P}_{1.3}$ induces K rank-one matrices. After solving $\mathscr{P}_{1.3}$, we can recover the beamforming vectors $\boldsymbol{w}_k$ for Problem $\mathscr{P}_{1.1}$ through the SVD, i.e., $\boldsymbol{W}_k^\star = \boldsymbol{w}_k\boldsymbol{w}_k^{\mathsf{H}}$, $\forall\, k \in \mathcal{K}$, where $\{\boldsymbol{W}_k^\star, k \in \mathcal{K}\}$ denotes the solution of Problem $\mathscr{P}_{1.3}$.

Similarly, given the beamforming vectors $\{\boldsymbol{w}_k, k \in \mathcal{K}\}$, we minimize the difference between the nuclear norm and the spectral norm by lifting $\tilde{\boldsymbol{v}}$ into a PSD matrix $\boldsymbol{V} \in \mathbb{C}^{N+1\times N+1}$, where $\boldsymbol{V} = \tilde{\boldsymbol{v}}\tilde{\boldsymbol{v}}^{\mathsf{H}}$, $\mathrm{rank}(\boldsymbol{V}) = 1$. That is,

$$\begin{aligned}\mathscr{P}_{1.4}: \ \underset{\boldsymbol{V}}{\text{minimize}} & \ \|\boldsymbol{V}\|_* - \|\boldsymbol{V}\|_2\\ \text{subject to} & \ \gamma_k^{\min}\left(\sum_{j=k+1}^{K}\big(\mathrm{Tr}(\boldsymbol{R}_{l,j}\boldsymbol{V}) + |b_{l,j}|^2\big) + \sigma^2\right)\\ & \le \mathrm{Tr}(\boldsymbol{R}_{l,k}\boldsymbol{V}) + |b_{l,k}|^2, \quad \forall k, l = k,\ldots,K, \quad (17)\\ & \boldsymbol{V}_{n,n} = 1, \quad \forall\, n = 1,\ldots,N+1, \quad (18)\\ & \boldsymbol{V} \succcurlyeq 0. \quad (19)\end{aligned}$$

Specifically, when the objective value of Problem $\mathscr{P}_{1.4}$ becomes zero, we obtain an exact rank-one feasible solution, denoted as $\boldsymbol{V}^\star$. Using the SVD operator $\boldsymbol{V}^\star = \tilde{\boldsymbol{v}}\tilde{\boldsymbol{v}}^{\mathsf{H}}$, we then obtain a feasible solution $\tilde{\boldsymbol{v}}$ to Problem $\mathscr{P}_{1.2}$.

Although the above DC programs are still non-convex, they have the algorithmic advantage. In the next subsection, we will develop the DC algorithm for problems $\mathscr{P}_{1.3}$ and $\mathscr{P}_{1.4}$ via successive convex relaxation and prove that the sequence solutions of the algorithm converges to the stationary point. The superior performance has been shown in vast applications, e.g., degrees-of-freedom maximization for data shuffling in wireless distributed computing [46] and model aggregation via over-the-air computation for federated learning [47].

C. Alternating DC Algorithm

In this subsection, we shall propose an efficient alternating DC algorithm to obtain high-quality solutions for the beamforming vectors and the phase-shift matrix.

1) Difference of Strongly Convex Functions Representation: Although the DC programs $\mathscr{P}_{1.3}$ and $\mathscr{P}_{1.4}$ are non-convex, they have a good structure that can be exploited to develop an efficient algorithm by using successive convex approximation [48]. In order to establish some important properties of the algorithm, we represent the objective function as the difference of two strongly convex functions. Specifically, we rewrite $\mathscr{P}_{1.3}$ as

$$\begin{aligned}\underset{\{\boldsymbol{W}_k\}}{\text{minimize}}\ f_1 = & \sum_{k=1}^{K}\mathrm{Tr}(\boldsymbol{W}_k) + \rho\sum_{k=1}^{K}\Big(\|\boldsymbol{W}_k\|_* - \|\boldsymbol{W}_k\|_2\Big)\\ & + I_{\mathcal{C}_1}(\{\boldsymbol{W}_k\}), \quad (20)\end{aligned}$$

and Problem $\mathscr{P}_{1.4}$ as

$$\underset{\boldsymbol{V}}{\text{minimize}}\ f_2 = \|\boldsymbol{V}\|_* - \|\boldsymbol{V}\|_2 + I_{\mathcal{C}_2}(\boldsymbol{V}), \quad (21)$$

where $\mathcal{C}_1$ and $\mathcal{C}_2$ denote the PSD cones that satisfy the constraints in problems $\mathscr{P}_{1.3}$ and $\mathscr{P}_{1.4}$, respectively, and both $I_{\mathcal{C}_1}(\{\boldsymbol{W}_k\})$ and $I_{\mathcal{C}_2}(\boldsymbol{V})$ is the indicator functions. We rewrite the DC functions f_1 and f_2 as the difference of two strongly convex functions, i.e., $f_1 = g_1 - h_1$ and $f_2 = g_2 - h_2$, where $g_1 = \sum_{k=1}^{K}\mathrm{Tr}(\boldsymbol{W}_k) + \rho\sum_{k=1}^{K}\|\boldsymbol{W}_k\|_* + I_{\mathcal{C}_1}(\{\boldsymbol{W}_k\}) + \frac{\eta}{2}\sum_{k=1}^{K}\|\boldsymbol{W}_k\|_F^2$, $h_1 = \rho\sum_{k=1}^{K}\|\boldsymbol{W}_k\|_2 + \frac{\eta}{2}\sum_{k=1}^{K}\|\boldsymbol{W}_k\|_F^2$, $g_2 = \|\boldsymbol{V}\|_* + I_{\mathcal{C}_2}(\boldsymbol{V}) + \frac{\eta}{2}\|\boldsymbol{V}\|_F^2$, $h_2 = \|\boldsymbol{V}\|_2 + \frac{\eta}{2}\|\boldsymbol{V}\|_F^2$.

Because of the additional quadratic terms (i.e., $\frac{\eta}{2}\sum_{k=1}^{K}\|\boldsymbol{W}_k\|_F^2$ and $\frac{\eta}{2}\|\boldsymbol{V}\|_F^2$), g_1, h_1, g_2, and h_2 are all η-strongly convex functions. It turns out that problems (20) and (21) have the unified structure of minimizing the difference of two strongly convex functions, i.e.,

$$\underset{\boldsymbol{Z}\in\mathbb{C}^{m\times m}}{\text{minimize}}\ f_i = g_i(\boldsymbol{Z}) - h_i(\boldsymbol{Z}), \quad i = 1, 2. \quad (22)$$

To solve the non-convex DC program, we present a DC algorithm to construct a sequence of candidates to the primal and dual solutions via successive convex relaxations in the sequel.

2) DC Algorithm for Problem (22)*:* According to the Fenchel's duality [49], the dual problem of Problem (22) is equivalent to

$$\underset{\boldsymbol{Y}\in\mathbb{C}^{m\times m}}{\text{minimize}}\ h_i^*(\boldsymbol{Y}) - g_i^*(\boldsymbol{Y}), \quad i = 1, 2, \quad (23)$$

where g_i^* and h_i^* are the conjugate functions of g_i and h_i, respectively. The conjugate function $h_i^*(\boldsymbol{Y})$ is defined as $h_i^*(\boldsymbol{Y}) = \sup_{\boldsymbol{Y}}\{\langle\boldsymbol{Z},\boldsymbol{Y}\rangle - h_i(\boldsymbol{Z}) : \boldsymbol{Z} \in \mathcal{Z}\}, i = 1, 2$, where the inner product is given by $\langle\boldsymbol{X},\boldsymbol{Y}\rangle = \Re(\mathrm{Tr}(\boldsymbol{X}^{\mathsf{H}}\boldsymbol{Y}))$ according to Wirtinger calculus [49] in the complex domain and $\mathcal{Z}$ denotes $\boldsymbol{Z}$'s feasible solution region. Since the primal problem (22) and its dual problem (23) are non-convex, the DC algorithm iteratively updates both the primal and dual variables via successive convex approximations. Specifically, in the r-th iteration, we have

$$\boldsymbol{Y}^r = \underset{\boldsymbol{Y}}{\arg\inf}\ h_i^*(\boldsymbol{Y}) - \big[g_i^*(\boldsymbol{Y}^{r-1}) + \langle\boldsymbol{Y} - \boldsymbol{Y}^{r-1}, \boldsymbol{Z}^r\rangle\big], \quad (24)$$

$$\boldsymbol{Z}^{r+1} = \underset{\boldsymbol{Z}}{\arg\inf}\ g_i(\boldsymbol{Z}) - \big[h_i(\boldsymbol{Z}^r) + \langle\boldsymbol{Z} - \boldsymbol{Z}^r, \boldsymbol{Y}^r\rangle\big]. \quad (25)$$

Based on the Fenchel bi-conjugation theorem [49], the solution to Problem (25) can be written as $\boldsymbol{Y}^r \in \partial_{\boldsymbol{Z}^r}h_i$, where $\partial_{\boldsymbol{Z}^r}h_i$ is the sub-gradient of h_i with respect to $\boldsymbol{Z}$ at $\boldsymbol{Z}^r$. Thus,

Algorithm 1: Proposed Alternating DC Algorithm for Solving Problem $\mathscr{P}_1$

Input: $\{\boldsymbol{h}_{r,k}\}$, $\{\boldsymbol{h}_{d,k}\}$, $\boldsymbol{G}$, $\{R_k^{\min}\}$, σ, ϵ, and π.
Initialize: $\boldsymbol{\Theta}^0 = \boldsymbol{\Theta}^{\text{initial}}$.
repeat
  Given $\boldsymbol{\Theta}^{t-1}$, solve Problem $\mathscr{P}_{1.3}$, set $r = 1$.
  while penalty component of $\mathscr{P}_{1.3}$ is not zero **do**
    Compute subgradient $\partial_{\boldsymbol{W_k}^{r-1}} \|\boldsymbol{W_k}\|_2, k \in \mathcal{K}$, obtain solution $\{\boldsymbol{W}_k^r\}$ by solving (26).
  end while
  Obtain $\{\boldsymbol{w}_k^t, k \in \mathcal{K}\}$ via SVD where $\boldsymbol{W}_k^r = \boldsymbol{w}_k^t (\boldsymbol{w}_k^t)^{\mathsf{H}}$.
  Given $\{\boldsymbol{w}_k^t, k \in \mathcal{K}\}$, solve $\mathscr{P}_{1.4}$, set $r = 1$.
  while objective value of $\mathscr{P}_{1.4}$ is non-zero **do**
    Compute subgradient $\partial_{\boldsymbol{V}^{r-1}} \|\boldsymbol{V}\|_2$, obtain solution $\{\boldsymbol{V}^r\}$ by solving (27).
  end while
  Obtain $\tilde{\boldsymbol{v}}^t$ via SVD where $\boldsymbol{V}^r = \tilde{\boldsymbol{v}}^t (\tilde{\boldsymbol{v}}^t)^{\mathsf{H}}$, $\boldsymbol{v}^t = [\tilde{\boldsymbol{v}}^t / \tilde{\boldsymbol{v}}_{N+1}^t]_{(1:N)}$, $\boldsymbol{\Theta}^t = \text{diag}((\boldsymbol{v}^t)^{\mathsf{H}})$
until Decrease of the transmit power is below ϵ or Problem $\mathscr{P}_{1.2}$ becomes infeasible.
Output: $\{\boldsymbol{w}_k\}$ and $\boldsymbol{\Theta}$

$\{\boldsymbol{W}_k^r, k \in \mathcal{K}\}$ at the r-th iteration for $\mathscr{P}_{1.3}$ can be obtained by solving the following convex problem

$$\begin{aligned} &\underset{\{\boldsymbol{W}_k\}}{\text{minimize}}\, g_1 - \sum_{k=1}^{K} \langle \boldsymbol{W_k}, \partial_{\boldsymbol{W_k}^{r-1}} h_1 \rangle \\ &\text{subject to constraints (15), (16).} \end{aligned} \tag{26}$$

Similarly, $\boldsymbol{V}^r$ at the r-th iteration for $\mathscr{P}_{1.4}$ can be obtained by solving the following convex optimization problem

$$\begin{aligned} &\underset{\boldsymbol{V}}{\text{minimize}}\, g_2 - \langle \boldsymbol{V}, \partial_{\boldsymbol{V}^{r-1}} h_2 \rangle \\ &\text{subject to constraints (17), (18), (19).} \end{aligned} \tag{27}$$

Problems (26) and (27) are convex and can be efficiently solved by using CVX [50]. Note that $\partial_{\boldsymbol{W_k}^{r-1}} h_1$ and $\partial_{\boldsymbol{V}^{r-1}} h_2$ are $\partial_{\boldsymbol{W_k}^{r-1}} h_1 = \rho \partial_{\boldsymbol{W_k}^{r-1}} \|\boldsymbol{W}_k\|_2 + \eta \boldsymbol{W}_k^{r-1}$, $\partial_{\boldsymbol{V}^{r-1}} h_2 = \partial_{\boldsymbol{V}^{r-1}} \|\boldsymbol{V}\|_2 + \eta \boldsymbol{V}^{r-1}$, respectively. It is worth noting that the sub-gradient of $\|\boldsymbol{X}\|_2$ at $\boldsymbol{X}^r \in \mathbb{C}^{N \times N}$ (i.e., $\boldsymbol{\partial}_{\boldsymbol{X}^r} \|\boldsymbol{X}\|_2$) can be efficiently computed according to the following proposition.

Proposition 2: For a PSD matrix $\boldsymbol{X}$, the sub-gradient of $\|\boldsymbol{X}\|_2$ at $\boldsymbol{X}^r$ can be efficiently computed as $\boldsymbol{u}_1 \boldsymbol{u}_1^{\mathsf{H}}$, where $\boldsymbol{u}_1 \in \mathbb{C}^N$ is the eigenvector corresponding to the largest eigenvalue $\sigma_1(\boldsymbol{X}^r)$.

The efficient DC algorithm is developed by successively solving the convex relaxation of the primal and dual problems of DC programming. The overall algorithm, solving problems $\mathscr{P}_{1.3}$ and $\mathscr{P}_{1.4}$ in an alternative approach, which is referred to as the alternating DC algorithm as presented in Algorithm 2. Specifically, Algorithm 1 optimizes $\{\boldsymbol{w}_k, k \in \mathcal{K}\}$ and $\boldsymbol{\Theta}$ alternatively, where the presented DC algorithm is adopted to obtain the beamforming vectors and the phase shifts in the lifted matrix space that satisfy the rank-one constraints. For a fair comparison, the alternating DC algorithm terminates when the decrease of the objective value of Problem $\mathscr{P}_1$ is smaller than ϵ, which is a predetermined convergence threshold, or Problem $\mathscr{P}_{1.2}$ becomes infeasible. We shall prove the convergence of Algorithm 2 in the sequel.

Algorithm 2: User Ordering Optimization Algorithm

Input: $\{\boldsymbol{h}_{r,k}\}$, $\{\boldsymbol{h}_{d,k}\}$, $\boldsymbol{G}$, $\{R_k^{\min}\}$, and σ.
Calculate the eigenvectors $\{\boldsymbol{u}_k\}$ corresponding to the largest eigenvalue of matrix $\{\boldsymbol{Q}_k\}$.
Calculate all users' required minimal transmit power $\{\hat{p}_k\}$ according to (34) and sort them in the descending order: $\hat{p}_{\pi(1)} \geq \ldots \geq \hat{p}_{\pi(K)}$.
Output: user ordering vector $\pi = \{\pi(1), \ldots, \pi(K)\}$.

D. Alternating DC Algorithm Convergence Analysis

Before proving the convergence of the proposed alternating DC algorithm, we present some important properties of the solutions obtained by solving the convex relaxation of the primal and dual problems of DC programming in the following proposition.

Proposition 3: For any $r = 0, 1, \ldots,$ the sequence $\{\boldsymbol{W}_k^r, k \in \mathcal{K}\}$ generated by iteratively solving Problem (26) has the following properties:

(i) The sequence $\{\boldsymbol{W}_k^r, k \in \mathcal{K}\}$ converges to a stationary point of f_1 in Problem (20) from an arbitrary initial point, and the sequence $\{f_1^r\}$ is strictly decreasing and convergent.

(ii) For any $r = 0, 1, \ldots,$ we have

$$\text{Avg}\Big(\big\|\boldsymbol{W}_k^r - \boldsymbol{W}_k^{r+1}\big\|\Big) \leq \frac{f_1^0 - f_1^\star}{\eta(r+1)}, \quad \forall\, k = 1, \ldots, K, \tag{28}$$

where $f_1^\star$ is the global minimum of f_1 and $\text{Avg}\big(\big\|\boldsymbol{W}_k^r - \boldsymbol{W}_k^{r+1}\big\|\big)$ denotes the average of the sequence $\{\|\boldsymbol{W}_k^i - \boldsymbol{W}_k^{i+1}\|_F^2\}_{i=0}^r$.

Likewise, for any $r = 0, 1, \ldots,$ the sequence $\{\boldsymbol{V}^r\}$ generated by iteratively solving Problem (27) has the following properties:

(iii) The sequence $\{\boldsymbol{V}^r\}$ converges to a stationary point of f_2 in Problem (21) from an arbitrary initial point, and the sequence of $\{f_2^r\}$ is strictly decreasing and convergent.

(iv) For any $r = 0, 1, \ldots,$ we have

$$\text{Avg}\Big(\big\|\boldsymbol{V}^r - \boldsymbol{V}^{r+1}\big\|\Big) \leq \frac{f_2^0 - f_2^\star}{\eta(r+1)}, \tag{29}$$

where $f_2^\star$ is the global minimum of f_2.

Proof: Please refer to Appendix A. □

Based on Proposition 3, the convergence analysis of Algorithm 2 is given in proposition 4.

Proposition 4: The objective value of Problem $\mathscr{P}_1$ in (11) decreases as the number of iteration increases until convergence by applying the proposed alternating DC algorithm.

Proof: Please refer to Appendix B. □

IV. Low-Complexity User Ordering Scheme

The optimal decoding order will be any one of the $K!$ different decoding orders and $\mathscr{P}_1$ must be solved $K!$ times. Therefore, an exhaustive search is needed over all the decoding

orders which is computationally prohibitive when K is large. To address this issue, we shall develop a low-complexity user ordering scheme to determine the decoding order of the users for RIS-empowered MISO NOMA networks. The existing studies mainly focused on either the channel condition [10] or QoS [51] based user ordering scheme for NOMA transmission. In contrast to the existing studies, to decouple the users' influence on each other, we order the users in the descending order of all users' minimum required transmit power, each of which is obtained by optimizing the phase-shift matrix at RIS and corresponding transmit beamforming without intra-cell interference. Note that the proposed user ordering strategy takes both the combined channel conditions and the target data rates into account. With the proposed strategy, the required transmit power can be obtained by only solving the power minimization problem once.

A. Proposed User Ordering Optimization

Specifically, the minimum transmit power required at the BS to serve user U_k can be obtained by solving the following problem

$$\begin{aligned} &\underset{\boldsymbol{w}_k,\boldsymbol{\Theta}}{\text{minimize}} \ \|\boldsymbol{w}_k\|^2 \\ &\text{subject to } \|(\boldsymbol{h}_{r,k}^{\mathsf{H}}\boldsymbol{\Theta}\boldsymbol{G}+\boldsymbol{h}_{d,k}^{\mathsf{H}})\boldsymbol{w}_k\|^2 \geq \gamma_k^{\min}\sigma^2, \\ &\qquad |\boldsymbol{\Theta}_{n,n}|=1, \quad \forall\, n. \end{aligned} \tag{30}$$

Although [18] has provided a SDR-based alternating optimization approach to solve (30), this user ordering scheme needs to solve K SDP problems and suffers from a very high computation complexity. To reduce this computation complexity, we propose a method to derive solutions for $\boldsymbol{w}_k$ and $\boldsymbol{\Theta}$ with closed-form expressions in this following.

For a given $\boldsymbol{\Theta}$, it is well-known that the maximum-ratio transmission (MRT) is the optimal transmit beamforming solution to Problem (30) [18], i.e., $\boldsymbol{w}_k^\star = \sqrt{p_k}\frac{(\boldsymbol{h}_{r,k}^{\mathsf{H}}\boldsymbol{\Theta}\boldsymbol{G}+\boldsymbol{h}_{d,k}^{\mathsf{H}})^{\mathsf{H}}}{\|\boldsymbol{h}_{r,k}^{\mathsf{H}}\boldsymbol{\Theta}\boldsymbol{G}+\boldsymbol{h}_{d,k}^{\mathsf{H}}\|}$, where p_k is the transmit power of the AP for user U_k. Furthermore, the optimal transmit power $p_k^\star$ satisfies $p_k^\star = \frac{\gamma_k^{\min}\sigma^2}{\|\boldsymbol{h}_{r,k}^{\mathsf{H}}\boldsymbol{\Theta}\boldsymbol{G}+\boldsymbol{h}_{d,k}^{\mathsf{H}}\|^2}$, which is determined by the target data rate and combined channel gains. Therefore, minimizing the transmit power is equivalent to maximizing the combined channel power gain, which is given by

$$\begin{aligned} &\underset{\boldsymbol{\Theta}}{\text{maximize}} \|\boldsymbol{h}_{r,k}^{\mathsf{H}}\boldsymbol{\Theta}\boldsymbol{G}+\boldsymbol{h}_{d,k}^{\mathsf{H}}\|^2 \\ &\text{subject to } |\boldsymbol{\Theta}_{n,n}|=1, \quad \forall\, n. \end{aligned} \tag{31}$$

Similar to Problem $\mathscr{P}_{1.2}$, by introducing $\tilde{\boldsymbol{v}}$, Problem (31) can be rewritten as

$$\begin{aligned} &\underset{\tilde{\boldsymbol{v}}\in\mathbb{C}^{N+1}}{\text{maximize}} \tilde{\boldsymbol{v}}^{\mathsf{H}}\boldsymbol{Q}_k\tilde{\boldsymbol{v}} \\ &\text{subject to } |\tilde{\boldsymbol{v}}|_n=1, \quad \forall n=1,\ldots,N+1, \end{aligned} \tag{32}$$

where $\boldsymbol{Q}_k = \begin{bmatrix} \text{diag}(\boldsymbol{h}_{r,k}^{\mathsf{H}})\boldsymbol{G}\boldsymbol{G}^{\mathsf{H}}\text{diag}(\boldsymbol{h}_{r,k}) & \text{diag}(\boldsymbol{h}_{r,k}^{\mathsf{H}})\boldsymbol{G}\boldsymbol{h}_{d,k} \\ \boldsymbol{h}_{d,k}^{\mathsf{H}}\boldsymbol{G}^{\mathsf{H}}\text{diag}(\boldsymbol{h}_{r,k}) & \boldsymbol{h}_{d,k}^{\mathsf{H}}\boldsymbol{h}_{d,k} \end{bmatrix}$. Problem (32) has a concave objective function with non-convex unit modulus constraints. The authors [25] applied the SDR technique to reformulate Problem (32) into a convex SDP problem. To further reduce the computational complexity, we relax the unit modulus constraints in Problem (32) as a norm constraint, i.e., $\|\tilde{\boldsymbol{v}}\|^2 = N+1$. Such a relaxation yields a closed-form expression of $\boldsymbol{\Theta}$ for Problem (32). The relaxed optimization problem is given by

$$\begin{aligned} &\underset{\tilde{\boldsymbol{v}}\in\mathbb{C}^{N+1}}{\text{maximize}} \tilde{\boldsymbol{v}}^{\mathsf{H}}\boldsymbol{Q}_k\tilde{\boldsymbol{v}} \\ &\text{subject to } \|\tilde{\boldsymbol{v}}\|^2 = N+1. \end{aligned} \tag{33}$$

Problem (33) is an eigenvalue problem and its optimal solution is $\sqrt{N+1}\boldsymbol{u}_k$, where $\boldsymbol{u}_k \in \mathbb{C}^{N+1}$ is the eigenvector corresponding to the largest eigenvalue of matrix $\boldsymbol{Q}_k$. Then, we perform a phase extraction of this solution to form a unit modulus vector $\boldsymbol{v}_k^\star = [\text{unit}(\boldsymbol{u}_k)/\text{unit}(\boldsymbol{u}_k(N+1))]_{(1:N)}$. The objective value of Problem (30) is approximated by

$$\hat{p}_k = \frac{\gamma_k^{\min}\sigma^2}{\|\boldsymbol{h}_{r,k}^{\mathsf{H}}\text{diag}((\boldsymbol{v}_k^\star)^{\mathsf{H}})\boldsymbol{G}+\boldsymbol{h}_{d,k}^{\mathsf{H}}\|^2}, \tag{34}$$

which depends on the target data rate and the channel condition. We order K users in the descending order of power $\hat{p}_k$. Specifically, the user with the largest value of $\hat{p}_k$ decodes its own signal first, while the user with the smallest value of $\hat{p}_k$ needs to decode all other users' signals before decoding its own signal, and the corresponding algorithm is presented in Algorithm 2.

B. Complexity Analysis

The computational complexity of the proposed closed-form user ordering algorithm is $\mathcal{O}(KN^3)$ due to the SVD operator. In contrast, to solve the SDP optimization problem resulted from the state-of-the-art SDR approach [43], the computational complexity of the interior-point method is $\mathcal{O}(KN^{6.5}\log(1/\epsilon))$ with accuracy $\epsilon > 0$, which is much higher than the proposed algorithm. The effectiveness of the proposed user ordering scheme will also be shown in the next section.

V. Numerical Results

In this section, we present sample numerical results to demonstrate the ability of an RIS in enlarging the channel-gain difference and the effectiveness of the proposed alternating DC method. We consider a three-dimensional (3D) Cartesian coordinate system, where the BS is assumed to be equipped with a uniform linear array (ULA) located on the y-axis while the RIS is assumed to be equipped with a uniform planar array (UPA) located parallel to the $x-z$ plane. The number of RIS elements is set to be $N = N_xN_z$, where N_x and N_z are the number of RIS elements along the x-axis and z-axis, respectively. We set $N_x = 5$ and increase N_z linearly with N. The antenna spacing is a half wavelength. We consider an RIS-empowered network, whose horizontal projections are illustrated in Fig. 2, where K single-antenna users are uniformly and randomly distributed in a circle centered at (60, 0, 0) meter with a radius of 15 meters. At the same time, the RIS and BS are fixed at (50, 15,10) meter and (0, 0, 10) meter, respectively. We denote d_{BU}^k, d_{IU}^k, and d_{IB} as the distances between user U_k and the BS, between user U_k and the RIS, and between the BS and the RIS, respectively. The distance-dependent path loss for all channel is modeled

TABLE I
SUMMARY OF SIMULATION PARAMETERS

	BS-RIS link, $\boldsymbol{G}$	RIS-user link, $\boldsymbol{h}_{r,i}$
AoA	$\theta_{\mathrm{TR}}^{\mathrm{A}} = \arctan\left(\frac{50}{15}\right)$, $\psi_{\mathrm{TI}}^{\mathrm{A}} = 0$	$\theta_{k,\mathrm{IU}}^{\mathrm{A}} = \arctan\left(\frac{\lvert x_k - 50\rvert}{15 - y_k}\right)$
AoD	$\theta_{\mathrm{TR}}^{\mathrm{D}} = \frac{\pi}{2} - \theta_{\mathrm{TR}}^{\mathrm{A}}$	$\theta_{k,\mathrm{IR}}^{\mathrm{D}} = \frac{\pi}{2} - \theta_{\mathrm{k,IU}}^{\mathrm{A}}$, $\psi_{k,\mathrm{IR}}^{\mathrm{D}} = \arctan\left(\frac{-10}{d_{IU}^{k}}\right)$
LoS component	$\boldsymbol{G}^{\mathrm{LoS}} = \mathrm{a_I}(\theta_{\mathrm{TR}}^{\mathrm{A}}, \psi_{\mathrm{TR}}^{\mathrm{A}})\mathrm{a_T}(\theta_{\mathrm{TR}}^{\mathrm{D}})^H$	$\boldsymbol{h}_{r,i}^{\mathrm{LoS}} = \mathrm{a_R}(\theta_{k,\mathrm{IU}}^{\mathrm{A}})\mathrm{a_I}(\theta_{k,\mathrm{IR}}^{\mathrm{D}}, \psi_{k,\mathrm{IR}}^{\mathrm{D}})^H$

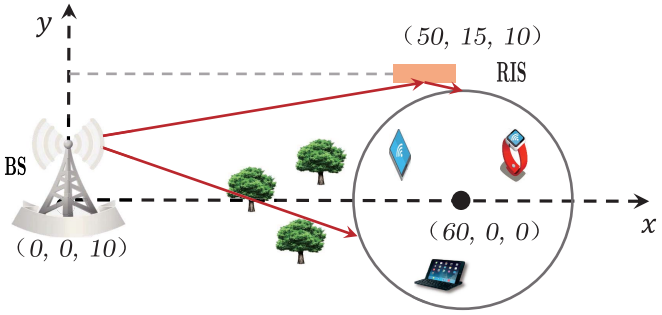


Fig. 2. Simulation setup.

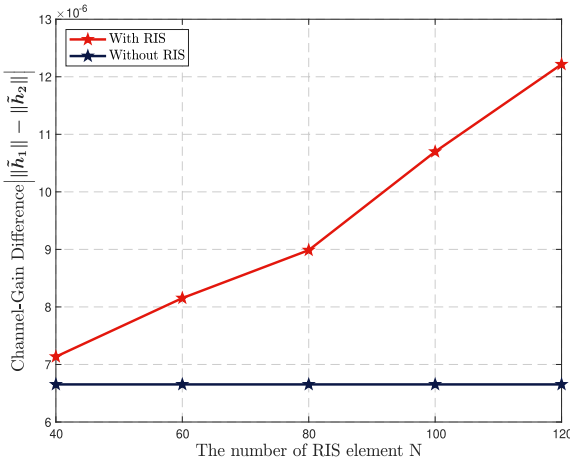


Fig. 3. Channel-gain difference versus N.

as $L(d) = T_0 d^{-\alpha}$, where $T_0 = -30$ dB is the path loss at reference distance one meter, d is the link distance, and α is the path loss exponent. To account for the small-scale fading, we assume Rician fading for all RIS-related channels since the RIS can be predeployed on tall buildings and in the vicinity of users. Due to the relatively large distance resulting in severe blockages and rich random scattering between BS and users, the BS-user channel is assumed to follow Rayleigh fading, which are almost the same as those in [18], [19]. In addition, the path loss exponents for the BS-user link, the BS-RIS link, and the RIS-user link are set to 3.67, 2.2, and 2.2, respectively, according to the 3GPP propagation environment [52]. Channels including pathloss and small-scale fading are denoted as $\boldsymbol{G} = \sqrt{L(d_{\mathrm{IB}})}\left(\sqrt{\frac{\beta_{BR}}{1+\beta_{BR}}}\boldsymbol{G}^{\mathrm{LOS}} + \sqrt{\frac{1}{1+\beta_{BR}}}\tilde{\boldsymbol{G}}\right)$, $\boldsymbol{h}_{r,k} = \sqrt{L(d_{\mathrm{IU}}^{k})}\left(\sqrt{\frac{\beta_{RU}}{1+\beta_{RU}}}\boldsymbol{h}_{r,k}^{\mathrm{LOS}} + \sqrt{\frac{1}{1+\beta_{RU}}}\tilde{\boldsymbol{h}}_{r,k}\right)$, $\boldsymbol{h}_{d,k} = \sqrt{L(d_{\mathrm{BU}}^{k})}\tilde{\boldsymbol{h}}_{d,k}$, where β_{BR} and β_{RU} are the corresponding Rician factors, $\tilde{\boldsymbol{G}}$, $\tilde{\boldsymbol{h}}_{r,k}$, and $\tilde{\boldsymbol{h}}_{d,k}$ denote Rayleigh fading components whose elements are generated according to $\mathcal{CN}(0,1)$, and $\boldsymbol{G}^{\mathrm{LOS}}$ and $\{\boldsymbol{h}_{r,k}^{\mathrm{LOS}}\}$ denote the deterministic line-of-sight (LoS) components of the BS-RIS and RIS-user, respectively. In particular, the LoS component is modeled as the product of the array responses at two sides. For ULA at the BS, the array response is modeled as $\mathrm{a_T}(\theta) \in \mathbb{C}^M$, with $[\mathrm{a_T}(\theta)]_m = e^{j\pi(m-1)\sin\theta}$, $\forall m$, where $\theta \in [0, 2\pi]$ denotes the angle-of-arrival (AoA) or angle-of-departure (AoD). For UPA at RIS, the array response is modeled as $\mathrm{a_I}(\theta,\psi) \in \mathbb{C}^{N\times 1}$ with $[\mathrm{a_I}(\theta,\psi)]_n = e^{j\pi(\lfloor\frac{n}{N_x}\rfloor \sin\psi\sin\theta + (n-\lfloor\frac{n}{N_x}\rfloor N_x)\sin\psi\cos\theta)}, \forall n$, where $\theta \in [0, 2\pi)$ and $\psi \in [-\pi/2, \pi/2)$ denote the azimuth AoA or AoD and elevation AoA or AoD [53], respectively and $\lfloor x \rfloor$ denotes the maximum integer no larger than a real number x. The AoA or AoD for RIS-related channels and their corresponding LoS components are summarized in Table II.

Unless specified otherwise, we set $R_k^{\min} = 1.5$ bits per channel use, $\forall\, k \in \mathcal{K}$, $\rho = 10$, $\epsilon = 10^{-4}$, channel bandwidth of 10 MHz, the noise power density of -174 dBm/Hz, Rician factor of $\beta_{BR} = \beta_{RU} = 3$. In spite of the Fig. 4 and Fig. 9, all curves in others are obtained by adopting the proposed user ordering design according to Algorithm 2 in Section IV, wherein the phase-shift matrix $\boldsymbol{\Theta}$ is a zero matrix in the scenario without RIS. All results in Figs. 3-4 and Figs. 6-10 are obtained by averaging over 100 random small-scale fading realizations.

A. Effect of Introducing RIS in NOMA Networks

In the following, we conduct numerical experiments to justify our motivation of the proposed RIS-empowered NOMA. We consider two users in the NOMA networks. In the simulations, we fix user U_1's location (i.e., (60, 10, 0) meter) and $M = 2$. The coordinate of user U_2 is (x_2, 0, 0) meter. We study the ability of an RIS in enlarging the channel-gain difference in MISO-NOMA networks when the original users' channel conditions are similar, i.e., user U_2's x-coordinate is equal to 70. The channel-gain difference is defined as the absolute value of the difference between the norm of the effective channel coefficient vectors (i.e., $\left|\|\tilde{\boldsymbol{h}}_i\|_2 - \|\tilde{\boldsymbol{h}}_j\|_2\right|$). Note that both the effective channel gains $\|\tilde{\boldsymbol{h}}_i\|_2$ and $\|\tilde{\boldsymbol{h}}_j\|_2$ incorporate the impact of the phase-shift matrix at the RIS. Fig. 3 illustrates the channel-gain difference between two users introduced by deploying an RIS for different number of RIS elements. Note that the order of the channel magnitude of each user is about 10^{-6}, which can be viewed as a reference for the average channel-gain difference. As shown in Fig. 3, the average channel-gain difference goes up as N increases in the scenario with RIS. In particular, when

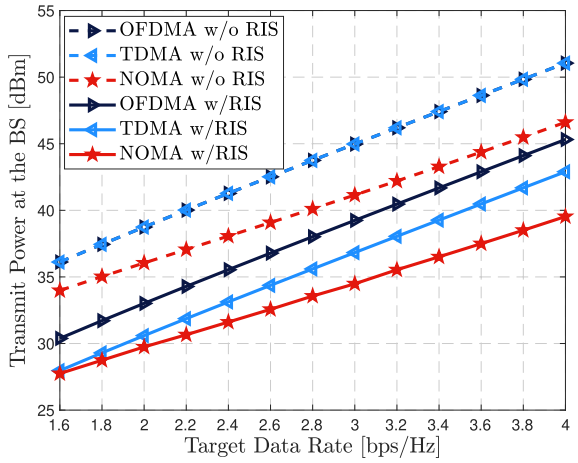


Fig. 4. Performance comparison of NOMA and OMA.

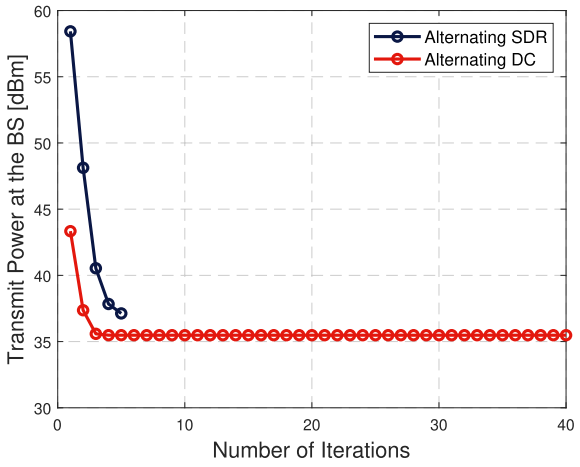


Fig. 5. Convergence behaviors of difference algorithms.

deploying an RIS, although N is relatively small (e.g., 60), the average channel-gain difference increases 23% compared to that without RIS.

In Fig. 4, we compare the performance of NOMA and OMA (i.e., TDMA and OFDMA) in an RIS-assisted system when $K = 2$, $M = 2$, and $N = 50$. We set the target data rate of all the users to be the same (i.e., $R_1^{\min} = R_2^{\min}$). In addition, the user ordering in NOMA adopts the optimal user ordering. As expected, the transmit power of the BS is significantly reduced with the assistance of RIS. Furthermore, the transmit power of the BS with RIS-assisted NOMA is always lower than that with RIS-assisted TDMA and RIS-assisted OFDMA since NOMA achieves higher spectrum efficiency than OMA.

B. Performance Comparison of Different Methods

We compare the following algorithms: 1) **Alternating SDR**: This method leverages the SDR technique to solve problems $\mathscr{P}_{1.1}$ and $\mathscr{P}_{1.2}$ alternatively. 2) **Random phase shift**: With this method, the phase for each reflection element is uniformly and independently generated from $[0, 2\pi]$ and kept fixed when solving the transmit power minimization problem $\mathscr{P}_{1.3}$; 3) **Without RIS**: With this method, the phase-shift matrix $\mathbf{\Theta}$ is set to zero matrix when solving the transmit power minimization problem $\mathscr{P}_{1.3}$.

Fig. 5 illustrates the convergence behaviors of the proposed alternating DC method and the alternating SDR method when $K = 5$, $M = 4$, and $N = 60$. The transmit power obtained by the alternating SDR method is higher than that obtained by the alternating DC method at the first iteration when the same initial phase-shift matrix is given. This is because removing the rank-one constraints incurs performance degradation, while the proposed alternating DC method ensures that the rank-one constraints hold. Furthermore, it can be observed that the alternating SDR method with Gaussian randomization fails to return a feasible solution to Problem $\mathscr{P}_{1.2}$ after the fifth iteration and thus early terminates the alternating optimization procedure. In contrast, the proposed alternating DC method is able to induce exact rank-one solutions, and hence accurately detects the feasibility of Problem $\mathscr{P}_{1.3}$, which avoids the early stopping in the alternating optimization procedure, thereby considerably reducing the transmit power consumption compared with alternating SDR method.

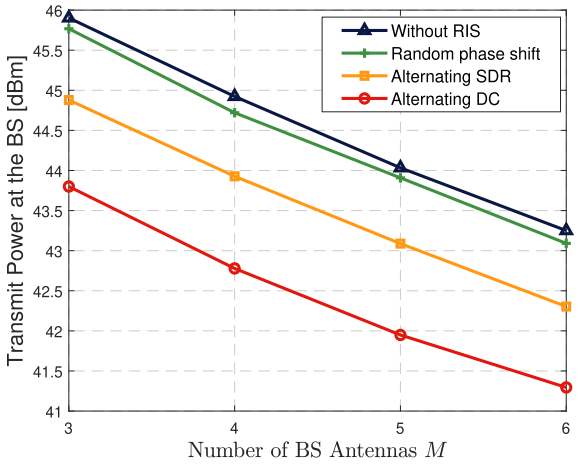


Fig. 6. Transmit power versus M ($N = 60$, $K = 7$).

Fig. 6 shows the impact of the number of BS antennas (i.e., M) on the total transmit power when $N = 60$ and $K = 7$. The total transmit power of the BS decreases as the value of M increases, which indicates that more antennas at the BS lead to a better performance by achieving a higher diversity gain. As shown in the figure, the RIS-empowered network outperforms the network without RIS, which demonstrates the effectiveness of deploying RIS in cellular networks. In addition, both the proposed alternating DC method and the alternating SDR method significantly outperform the random phase shift method. Hence demonstrating the necessity of jointly optimizing the beamforming vectors at the BS and the phase-shift matrix at the RIS. Furthermore, due to the superiority of the proposed DC representation, the proposed alternating DC method consumes much less transmit power than the alternating SDR method.

Fig. 7 illustrates the impact of the number of passive reflecting elements at the RIS (i.e., N) on the total transmit power when $M = 6$ and $K = 6$. For alternating DC and alternating SDR, the total transmit power decreases quickly

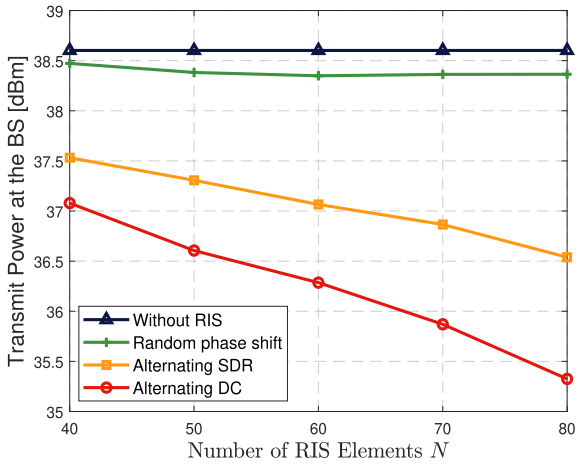


Fig. 7. Transmit power versus N ($M = 6$ and $K = 6$).

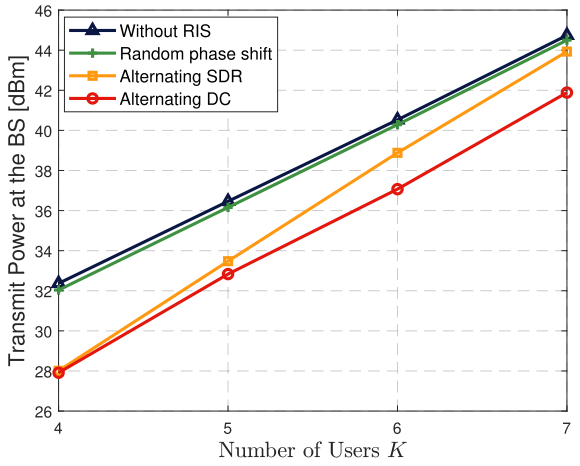


Fig. 8. Transmit power versus K ($M = 4$ and $N = 80$).

as the value of N increases by jointly optimizing transmit beamforming vector at the BS and phase-shift matrix at RIS. This is because an RIS with more reflecting elements significantly can enhance the receiving power at the users by optimizing the phase-shift matrix and introduces more channel differences among the users. Therefore, a larger number of passive reflecting elements leads to a higher energy-efficiency. Furthermore, as shown in the figure, the performance gap of alternating DC over alternating SDR goes up as N increases. This is because SDR method hardly return a feasible solution for Problem $\mathscr{P}_{1.2}$ as the dimension of phase shifter (i.e., N) increase and thus early terminates the alternating optimization procedure.

Fig. 8 shows the performance of downlink NOMA networks with and without RIS when $M = 4$ and $N = 80$. The total transmit power of the BS quickly increases as the value of K increases. Furthermore, we observe that the performance gap of alternating DC over alternating SDR goes up as K increases since the rank-one probability of SDP solution by using SDR method for the transmit beamforming optimization gradually becomes low and thus incurs performance degradation as the number of users increases.

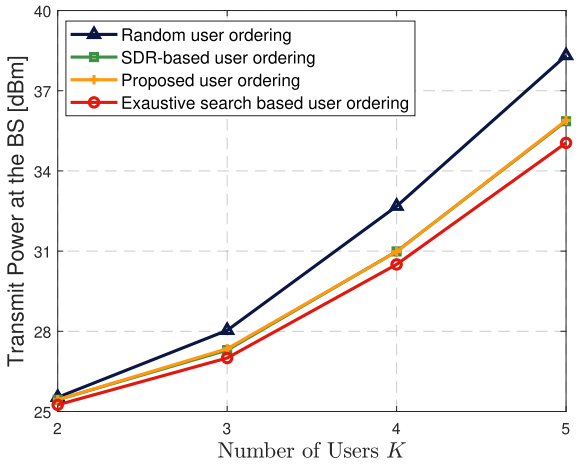


Fig. 9. Transmit power versus K for different user ordering schemes.

C. Performance Comparison of Different User Ordering Schemes

We compare the performance of the proposed user ordering scheme with following schemes: 1) **Random user ordering**: The users ordering is randomly selected from user ordering set $\mathcal{S}$; 2) **Exhaustive search based user ordering**: This user ordering scheme finds the optimal decoding order that achieves the best performance by exhaustively searching over all $K!$ possible decoding orders; 3) **Proposed user ordering**: The proposed user ordering scheme with closed-form solutions is presented in Algorithm 2; 4) **SDR-based user ordering**: Different from the proposed scheme, this user ordering scheme obtains the ordering criterion by solving Problem (31) using the SDR technique [43].

Fig. 9 compares the performance of our proposed user ordering scheme with three benchmarks when $N = 60$ and $M = 2$. It is observed that the user ordering has a significantly impact on the transmit power consumption in RIS-empowered NOMA networks. In particular, the performance gaps between the optimal user ordering scheme and other three user ordering schemes increase as the number of users increases. However, the optimal user ordering scheme needs to exhaustively search all $K!$ possible decoding orders, and its computational complexity is extremely high. Although the random user ordering scheme has the lowest complexity, it suffers from a larger performance degradation than the proposed and the SDR-based user ordering schemes. This is because the last two schemes capture the effects of users' target data rates and combined channels which have an impact on the transmit power at BS for the each user. When $K = 5$, the transmit powers of the random, the proposed, and the SDR-based user ordering schemes are 3.27 dBm, 0.84 dBm, and 0.81 dBm higher than the optimal user ordering scheme, respectively. It shows that the proposed user ordering provides comparable performance to the optimal user ordering. Moreover, as shown in Fig. 9, the proposed user ordering scheme achieves almost the same performance as the SDR-based user ordering scheme. However, the SDR-based user ordering scheme needs to solve K SDP problems for K users, the complexity of which is much higher than the proposed user ordering scheme with closed-form solutions.

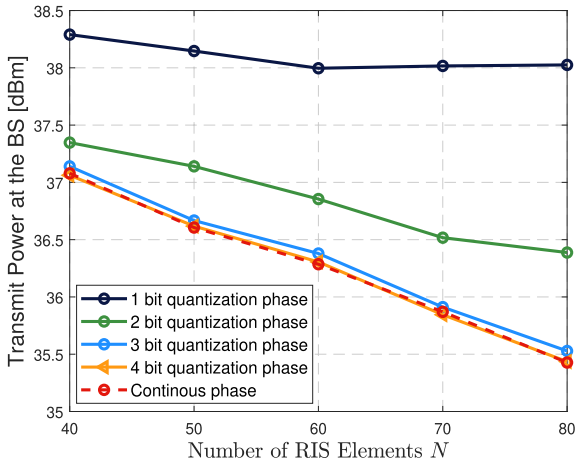


Fig. 10. Transmit power versus N with discrete and continuous phase shifts.

D. Performance Comparison of Discrete and Continuous Phase Shifts

In practical systems, the RIS with a large number of passive elements has a finite phase resolution, which depends on the number of quantization bits, denoted as B [54], [55]. We numerically investigate the effect of RIS's finite phase resolution on the total transmit power consumption at the BS by the alternating DC algorithm. Specifically, at the t-th iteration, the optimized continuous phase shifts are quantized to their nearest discrete values denoted by $\boldsymbol{v}_\Phi$ in the set $\left\{0, \frac{2\pi}{2^B}, \ldots, \frac{2\pi\times(2^B-1)}{2^B}\right\}$. Although the quantized phase shifts $\boldsymbol{v}_\Phi$ may not satisfy the QoS constraints, we can optimize the active beamforming $\{\tilde{\boldsymbol{w}}_k\}$ to satisfy the QoS constraints by solving $\mathscr{P}_{1.3}$ given $\boldsymbol{v}_\Phi$. To make the objective value to be non-increasing after each iteration for discrete phase shifts, we update $\boldsymbol{v}^t$ and $\{\boldsymbol{w}_k^t\}$ only when $\sum_{k=1}^K \|\tilde{\boldsymbol{w}}_k\|^2 \leq \sum_{k=1}^K \|\boldsymbol{w}_k^{t-1}\|^2$.

Fig. 10 plots the total transmit power consumption of the proposed alternating DC method versus the number of RIS elements for different phase quantization bits when $K = 6$, $M = 6$. We observe that the total transmit power consumption of the network with a discrete phase-shift RIS is greater than that of the network with a continuous phase-shift RIS. As the value of B increases, the total transmit power consumption decreases. With 1-3-bit phase shifters, the performance gap between the continuous and discrete phase shifts increases as the number of passive elements increases. Moreover, 4-bit phase shifters is practically sufficient to achieve almost the same performance as the continuous phase shifters.

VI. Conclusion

In this paper, we studied an RIS-empowered NOMA network to minimize the total transmit power by jointly optimizing the beamforming vectors at the BS and phase-shift matrix at RIS, where the RIS is capable of inducing desirable channel differences among the users to enhance the performance of NOMA. To address the unique challenges of highly coupled optimization variables and non-convex quadratic constraints, we proposed an alternating DC method to solve the non-convex bi-quadratically constrained quadratic problem. This is achieved by introducing an exact DC representation for the rank-one constraints in the lifted non-convex QCQP problems to accurately detect the feasibility of non-convex quadratic constraints for the transmit beamforming vectors and phase-shift matrix design. Therefore, early stopping can be avoided in the procedure of alternating optimization, thereby considerably improving the performance. We also developed an efficient DC algorithm with convergence guarantee to solve the resulting DC programming problems via successive convex relaxation. We further proposed a low-complexity user ordering scheme, which achieves a comparable performance to the exhaustive search method. Simulation results demonstrated that the proposed alternating DC method outperforms the state-of-the-art methods in terms of total transmit power minimization.

This initial investigation demonstrated the effectiveness of deploying an RIS in NOMA networks for energy efficiency enhancement. Numerical results also showed that RIS with low phase resolution can achieve almost the same performance as RIS with continuous phase shifts. More works are needed to fully exploit the benefits of RIS-empowered NOMA systems, including theoretical analysis, channel estimation, large-scale optimization. For future studies, the resource allocation framework developed in this paper will be extended to the scenario with multiple BSs, while taking into account imperfect CSI for practical implementations.

Appendix

A. Proof of Proposition 1

If matrix $\boldsymbol{X}$ is a rank-one PSD matrix, the nuclear norm is equal to the spectral norm since $\sigma_i(\boldsymbol{X}) = 0$ for all $i \geq 2$. Hence, $\|\boldsymbol{X}\|_* - \|\boldsymbol{X}\|_2 = \sum_{i=2}^N \sigma_i(\boldsymbol{X}) = 0$ implies that $\left\|[\sigma_1(\boldsymbol{X}), \ldots, \sigma_N(\boldsymbol{X})]\right\|_0 \leq 1$. Because of $\mathrm{Tr}(\boldsymbol{X}) > 0$, we have $\sigma_1(\boldsymbol{X}) > 0$. Therefore, $\mathrm{rank}(\boldsymbol{X}) = 1$ is equivalent to $\|\boldsymbol{X}\|_* - \|\boldsymbol{X}\|_2 = 0$. □

B. Proof of Proposition 3

Without loss of generality, we shall only present the proofs of properties (i) and (ii), while properties (iii) and (iv) can be proved similarly.

We first present the proof of property (i). For the sequence $\{\boldsymbol{W}_k^r, k \in \mathcal{K}\}$ generated by iteratively solving problem (26), we denote the dual variables as $\boldsymbol{Y}_k^r \in \partial_{\boldsymbol{W_k}^r} h_1$. Due to the strong convexity of h_1, we have

$$h_1^{r+1} \geq h_1^r + \sum_{k=1}^K \langle \Delta_r \boldsymbol{W}_k, \boldsymbol{Y}_k^r \rangle + \frac{\eta}{2} \sum_{k=1}^K \|\Delta_r \boldsymbol{W}_k\|_F^2, \tag{35}$$

$$\sum_{k=1}^K \langle \boldsymbol{W}_k^r, \boldsymbol{Y}_k^r \rangle = h_1^r + (h_1^*)^r, \tag{36}$$

where $\Delta_r \boldsymbol{W}_k = \boldsymbol{W}_k^{r+1} - \boldsymbol{W}_k^r$. By adding g_1^{r+1} to both sides of (35), we obtain

$$f_1^{r+1} \leq g_1^{r+1} - h_1^r - \sum_{k=1}^{K} \langle \Delta_r \boldsymbol{W}_k, \boldsymbol{Y}_k^r \rangle - \frac{\eta}{2} \sum_{k=1}^{K} \|\Delta_r \boldsymbol{W}_k\|_F^2, \tag{37}$$

For the update of primal variable $\{\boldsymbol{W}_k, k \in \mathcal{K}\}$ according to (25), we have $\boldsymbol{Y}_k^r \in \partial_{\boldsymbol{W_k}^{r+1}} g_1$, which implies that

$$g_1^r \geq g_1^{r+1} + \sum_{k=1}^{K} \langle -\Delta_r \boldsymbol{W}_k, \boldsymbol{Y}_k^r \rangle + \frac{\eta}{2} \sum_{k=1}^{K} \|\Delta_r \boldsymbol{W}_k\|_F^2, \tag{38}$$

$$\sum_{k=1}^{K} \langle \boldsymbol{W}_k^{r+1}, \boldsymbol{Y}_k^r \rangle = g_1^{r+1} + (g_1^*)^r. \tag{39}$$

Similarly, by subtracting h_1^r from both sides of (38), we have

$$f_1^r \geq g_1^{r+1} - h_1^r + \sum_{k=1}^{K} \langle -\Delta_r \boldsymbol{W}_k, \boldsymbol{Y}_k^r \rangle + \frac{\eta}{2} \sum_{k=1}^{K} \|\Delta_r \boldsymbol{W}_k\|_F^2. \tag{40}$$

By subtracting (39) from (36), we have

$$(f_1^*)^r = (h_1^*)^r - (g_1^*)^r = g_1^{r+1} - h_1^r + \sum_{k=1}^{K} \langle -\Delta_r \boldsymbol{W}_k, \boldsymbol{Y}_k^r \rangle, \tag{41}$$

After combining (37) and (41), we have

$$(f_1^*)^r \geq f_1^{r+1} + \frac{\eta}{2} \sum_{k=1}^{K} \|\Delta_r \boldsymbol{W}_k\|_F^2. \tag{42}$$

Similarly, after combining (40) and (41), we have

$$f_1^r \geq (f_1^*)^r + \frac{\eta}{2} \sum_{k=1}^{K} \|\Delta_r \boldsymbol{W}_k\|_F^2. \tag{43}$$

Based on (42) and (43), we conclude that

$$f_1^r \geq f_1^{r+1} + \eta \sum_{k=1}^{K} \|\Delta_r \boldsymbol{W}_k\|_F^2. \tag{44}$$

Therefore, the sequence $\{f_1^r\}$ is non-increasing. Since $f_1 \geq 0$ always holds, we conclude that the sequence $\{f_1^r\}$ is strictly decreasing until convergence, i.e., $\lim_{r\to\infty} \sum_{k=1}^{K} \|\Delta_r \boldsymbol{W}_k\|_F^2 = 0$.

When the sequence $\{f_1^r\}$ converges at the limit point $\big(\{\boldsymbol{W}_k, k \in \mathcal{K}\}\big)$, for every limit point, the distances between $\boldsymbol{W}_k^{r+1}$ and $\boldsymbol{W}_k^r$ satisfy $\|\boldsymbol{W}_k^{r+1} - \boldsymbol{W}_k^r\|_F^2 = 0, \forall\, k = 1, \ldots, K$. Based on inequalities (43) and (44), the equalities $f_1^{r+1} = f_1^r = (f_1^*)^r$ hold.

Based on the definitions of f_1 and f_1^*, we have $f_1^{r+1} = g_1^{r+1} - h_1^{r+1}$ and $(f_1^*)^r = (h_1^*)^r - (g_1^*)^r$. Therefore, it follows that

$$(h_1^*)^r + h_1^{r+1} = (g_1^*)^r + g_1^{r+1}. \tag{45}$$

By combining (39) and (45), we obtain $(h_1^*)^r + h_1^{r+1} = \sum_{k=1}^{K} \langle \boldsymbol{W}_k^{r+1}, \boldsymbol{Y}_k^r \rangle$. Since h_1 is a convex function, we have $\boldsymbol{Y}_k^r \in \partial_{\boldsymbol{W_k}^{r+1}} h_1, \forall k = 1, \ldots, K$. Therefore, we have $\boldsymbol{Y}_k^r \in \partial_{\boldsymbol{W_k}^{r+1}} g_1 \cap \partial_{\boldsymbol{W_k}^{r+1}} h_1, \forall k = 1, \ldots, K$. It is concluded that $\big(\{\boldsymbol{W}_k^{r+1}\}\big)$ is a critical point of $f_1 = g_1 - h_1$.

We now present the proof of property (ii). Based on the above analysis, we have

$$\begin{aligned} \mathrm{Avg}\Big(\sum_{k=1}^{K} \big\|\boldsymbol{W}_k^r - \boldsymbol{W}_k^{r+1}\big\|_F^2\Big) &\leq \sum_{i=0}^{r} \frac{1}{\eta(r+1)} (f_1^i - f_1^{i-1}) \\ &\leq \frac{1}{\eta(r+1)} (f_1^0 - f_1^{r+1}). \end{aligned} \tag{46}$$

We denote the optimal value of f_1 as $f_1^\star$. Since inequality $f_1^\star \leq f_1^{r+1}$ holds, we have

$$\frac{1}{\eta(r+1)} (f_1^0 - f_1^{r+1}) \leq \frac{1}{\eta(r+1)} (f_1^0 - f_1^\star). \tag{47}$$

According to (46) and (47), we conclude that property (ii) holds, i.e.,

$$\mathrm{Avg}\Big(\big\|\boldsymbol{W}_k^r - \boldsymbol{W}_k^{r+1}\big\|_F^2\Big) \leq \frac{f_1^0 - f_1^\star}{\eta(r+1)}, \quad \forall\, k = 1, \ldots, K. \tag{48}$$

This completes the proof.

C. Proof of Proposition 4

We denote $f\big(\{\boldsymbol{w}_k\}, \boldsymbol{\Theta}\big)$ as the objective value of $\mathscr{P}_1$ for a feasible solution $\big(\{\boldsymbol{w}_k\}, \boldsymbol{\Theta}\big)$. We denote $\big(\{\boldsymbol{w}_k^t\}, \boldsymbol{\Theta}^t\big)$ as a feasible solution of $\mathscr{P}_1$ at the t-th iteration. For a given $\boldsymbol{\Theta}^t$, we apply the presented DC algorithm to obtain a solution $\{\boldsymbol{W}_k^r\}$ for problem $\mathscr{P}_{1.3}$, based on which we obtain $\big(\{\boldsymbol{w}_k^t(\boldsymbol{w}_k^t)^{\mathsf{H}}\}\big)$ as the initial point for the $(t+1)$ iteration. Because the DC algorithm can accurately detect the feasibility of rank-one constraints, the solution $\{\boldsymbol{w}_k^{t+1}\}$ can be obtained via cholesky decomposition, where $\boldsymbol{W}_k^r = \boldsymbol{w}_k^{t+1}(\boldsymbol{w}_k^{t+1})^{\mathsf{H}}$. Hence, we have $f_1\big(\{\boldsymbol{w}_k^{t+1}(\boldsymbol{w}_k^{t+1})^{\mathsf{H}}\}, \boldsymbol{\Theta}^t\big) = f\big(\{\boldsymbol{w}_k^{t+1}\}, \boldsymbol{\Theta}^t\big)$ and $f_1\big(\{\boldsymbol{w}_k^t(\boldsymbol{w}_k^t)^{\mathsf{H}}\}, \boldsymbol{\Theta}^t\big) = f\big(\{\boldsymbol{w}_k^t\}, \boldsymbol{\Theta}^t\big)$. According to Proposition 3, the object value of $\mathscr{P}_{1.3}$ is strictly decreasing over the iterations. Hence, we have $f_1\big(\{\boldsymbol{w}_k^{t+1}(\boldsymbol{w}_k^{t+1})^{\mathsf{H}}\}, \boldsymbol{\Theta}^t\big) < f_1\big(\{\boldsymbol{w}_k^t(\boldsymbol{w}_k^t)^{\mathsf{H}}\}, \boldsymbol{\Theta}^t\big)$. Based on Algorithm 2, we have

$$f\big(\{\boldsymbol{w}_k^{t+1}\}, \boldsymbol{\Theta}^t\big) < f\big(\{\boldsymbol{w}_k^t\}, \boldsymbol{\Theta}^t\big). \tag{49}$$

For a given $\{\boldsymbol{w}_k^{t+1}, k \in \mathcal{K}\}$, we also apply the duality-based DC algorithm to solve problem $\mathscr{P}_{1.4}$. Based on Algorithm 2, if there exists a feasible solution $\boldsymbol{V}^{t+1}$ to problem $\mathscr{P}_{1.4}$, it is also feasible to problem $\mathscr{P}_{1.2}$, i.e., $\big(\{\boldsymbol{w}_k^{t+1}\}, \boldsymbol{\Theta}^{t+1}\big)$ exists. It follows that

$$f\big(\{\boldsymbol{w}_k^{t+1}\}, \boldsymbol{\Theta}^t\big) = f\big(\{\boldsymbol{w}_k^{t+1}\}, \boldsymbol{\Theta}^{t+1}\big), \tag{50}$$

where the equality holds as the value of f is independent of $\boldsymbol{\Theta}$ but only depends on $\{\boldsymbol{w}_k, k \in \mathcal{K}\}$. Based on (49) and (50), we further have $f\big(\{\boldsymbol{w}_k^{t+1}\}, \boldsymbol{\Theta}^{t+1}\big) < f\big(\{\boldsymbol{w}_k^t\}, \boldsymbol{\Theta}^t\big)$, which demonstrates that the objective value of problem $\mathscr{P}_1$ is always decreasing over iterations. Therefore, the proposed alternating DC algorithm converges. This completes the proof.

REFERENCES

[1] M. Fu, Y. Zhou, and Y. Shi, "Intelligent reflecting surface for downlink non-orthogonal multiple access networks," in *Proc. IEEE Global Commun. Conf. (Globecom) Workshops*, Waikoloa, HI, USA, Dec. 2019, pp. 1–6. [Online]. Available: https://arxiv.org/abs/1906.09434.

[2] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J.-A. Zhang, "The roadmap to 6G: AI empowered wireless networks," *IEEE Commun. Mag.*, vol. 57, no. 8, pp. 84–90, Aug. 2019.

[3] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal multiple access for 5G and beyond," *Proc. IEEE*, vol. 105, no. 12, pp. 2347–2381, Dec. 2017.

[4] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-Lin, and Z. Wang, "Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.

[5] Y. Zhou, V. W. S. Wong, and R. Schober, "Coverage and rate analysis of millimeter wave NOMA networks with beam misalignment," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8211–8227, Dec. 2018.

[6] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, 2nd Quart., 2017.

[7] Z. Ding *et al.*, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, Feb. 2017.

[8] Z. Ding, L. Dai, and H. V. Poor, "MIMO-NOMA design for small packet transmission in the Internet of Things," *IEEE Access*, vol. 4, pp. 1393–1405, Apr. 2016.

[9] Y. Zhou, V. W. S. Wong, and R. Schober, "Dynamic decode-and-forward based cooperative NOMA with spatially random users," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 3340–3356, May 2018.

[10] Z. Ding, P. Fan, and V. Poor, "Impact of user pairing on 5G non-orthogonal multiple access downlink transmissions," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, Sep. 2016.

[11] N. Lu, N. Cheng, N. Zhang, X. Shen, and J. W. Mark, "Connected vehicles: Solutions and challenges," *IEEE Internet Things J.*, vol. 1, no. 4, pp. 289–299, Aug. 2014.

[12] T. J. Cui, M. Q. Qi, X. Wan, J. Zhao, and Q. Cheng, "Coding metamaterials, digital metamaterials and programmable metamaterials," *Light Sci. Appl.*, vol. 3, no. 10, p. e218, Oct. 2014.

[13] M. D. Renzo *et al.*, "Smart radio environments empowered by reconfigurable AI meta-surfaces: An idea whose time has come," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 1, May 2019.

[14] C. Huang *et al.*, "Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends," *IEEE Wireless Commun.*, vol. 27, no. 5, pp. 118–125, Oct. 2020.

[15] 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.

[16] Y. Shi, J. Zhang, and K. B. Letaief, "Group sparse beamforming for green cloud-RAN," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2809–2823, May 2014.

[17] 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.

[18] 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.

[19] 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.

[20] T. Jiang and Y. Shi, "Over-the-air computation via intelligent reflecting surfaces," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Waikoloa, HI, USA, Dec. 2019, pp. 1–6.

[21] X. Yu, D. Xu, and R. Schober, "Enabling secure wireless communications via intelligent reflecting surfaces," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Waikoloa, HI, USA, Dec. 2019, pp. 1–6.

[22] J. Chen, Y.-C. Liang, Y. Pei, and H. Guo, "Intelligent reflecting surface: A programmable wireless environment for physical layer security," *IEEE Access*, vol. 7, pp. 82599–82612, Jun. 2019.

[23] S. Xia and Y. Shi, "Intelligent reflecting surface for massive device connectivity: Joint activity detection and channel estimation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Barcelona, Spain, May 2020, pp. 5175–5179.

[24] Z. Ding and H. V. Poor, "A simple design of IRS-NOMA transmission," *IEEE Commun. Lett.*, vol. 24, no. 5, pp. 1119–1123, May 2020.

[25] G. Yang, X. Xu, and Y.-C. Liang, "Intelligent reflecting surface assisted non-orthogonal multiple access," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Seoul, South Korea, May 2020, pp. 1–6.

[26] Y. Li, M. Jiang, Q. Zhang, and J. Qin, "Joint beamforming design in multi-cluster MISO NOMA reconfigurable intelligent surface-aided downlink communication networks," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 664–674, Jan. 2021, doi: 10.1109/TCOMM.2020.3032695.

[27] J. Zhu, Y. Huang, J. Wang, K. Navaie, and Z. Ding, "Power efficient IRS-assisted NOMA," *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 900–913, Feb. 2021, doi: 10.1109/TCOMM.2020.3029617.

[28] X. Mu, Y. Liu, L. Guo, J. Lin, and N. Al-Dhahir, "Exploiting intelligent reflecting surfaces in NOMA networks: Joint beamforming optimization," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6884–6898, Oct. 2020.

[29] J. Zuo, Y. Liu, Z. Qin, and N. Al-Dhahir, "Resource allocation in intelligent reflecting surface assisted NOMA systems," *IEEE Trans. Commun.*, vol. 68, no. 11, pp. 7170–7183, Nov. 2020.

[30] X. Liu, Y. Liu, Y. Chen, and H. V. Poor, "RIS enhanced massive non-orthogonal multiple access networks: Deployment and passive beamforming design," *IEEE J. Sel. Areas Commun.*, early access, Aug. 24, 2020, doi: 10.1109/JSAC.2020.3018823.

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

[32] H. Liu, X. Yuan, and Y.-J.-A. Zhang, "Matrix-calibration-based cascaded channel estimation for reconfigurable intelligent surface assisted multiuser MIMO," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2621–2636, Nov. 2020.

[33] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," in *Proc. IEEE Global Commun. Conf. (Globecom)*, Waikoloa, HI, USA, Dec. 2019, pp. 1–6.

[34] G. Zhou, C. Pan, H. Ren, K. Wang, M. Di Renzo, and A. Nallanathan, "Robust beamforming design for intelligent reflecting surface aided MISO communication systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1658–1662, Oct. 2020.

[35] X. Lu, W. Yang, X. Guan, Q. Wu, and Y. Cai, "Robust and secure beamforming for intelligent reflecting surface aided mmWave MISO systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 12, pp. 2068–2072, Dec. 2020.

[36] 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.

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

[38] W. Fang, M. Fu, K. Wang, Y. Shi, and Y. Zhou, "Stochastic beamforming for reconfigurable intelligent surface aided over-the-air computation," in *Proc. GLOBECOM IEEE Global Commun. Conf.*, Dec. 2020, pp. 1–6.

[39] Y. Liu, H. Xing, C. Pan, A. Nallanathan, M. Elkashlan, and L. Hanzo, "Multiple-antenna-assisted non-orthogonal multiple access," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 17–23, Apr. 2018.

[40] J. Choi, "Minimum power multicast beamforming with superposition coding for multiresolution broadcast and application to NOMA systems," *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 791–800, Mar. 2015.

[41] F. Alavi, K. Cumanan, Z. Ding, and A. G. Burr, "Beamforming techniques for nonorthogonal multiple access in 5G cellular networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9474–9487, Oct. 2018.

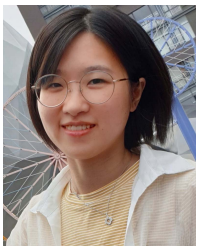
[42] J. Zhu, J. Wang, Y. Huang, K. Navaie, Z. Ding, and L. Yang, "On optimal beamforming design for downlink MISO NOMA systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 3008–3020, Mar. 2020.

[43] Z.-Q. Luo, W.-K. Ma, A. 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.

[44] E. Chen and M. Tao, "ADMM-based fast algorithm for multi-group multicast beamforming in large-scale wireless systems," *IEEE Trans. Commun.*, vol. 65, no. 6, pp. 2685–2698, Jun. 2017.

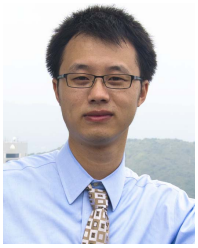
[45] L. Chen, X. Qin, and G. Wei, "A uniform-forcing transceiver design for over-the-air function computation," *IEEE Wireless Commun. Lett.*, vol. 7, no. 6, pp. 942–945, Dec. 2018.

[46] K. Yang, Y. Shi, and Z. Ding, "Data shuffling in wireless distributed computing via low-rank optimization," *IEEE Trans. Signal Process.*, vol. 67, no. 12, pp. 3087–3099, Jun. 2019.
[47] K. Yang, T. Jiang, Y. Shi, and Z. Ding, "Federated learning via over-the-air computation," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 2022–2035, Mar. 2020.
[48] P. D. Tao and L. T. H. An, "Convex analysis approach to DC programming: Theory, algorithms and applications," *Acta Math. Vietnam.*, vol. 22, no. 1, pp. 289–355, 1997.
[49] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ, USA: Princeton Univ. Press, 2015.
[50] M. Grant and S. Boyd. (Mar. 2014). *CVX: MATLAB Software for Disciplined Convex Programming, Version 2.1*. [Online]. Available: http://cvxr.com/cvx
[51] Y. Zhou, V. W. S. Wong, and R. Schober, "Stable throughput regions of opportunistic NOMA and cooperative NOMA with full-duplex relaying," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5059–5075, Aug. 2018.
[52] *Further Advancements for E-Utra Physical Layer Aspects (Release 9)*, document 3GPP TS 36.814, Mar. 2010.
[53] S. Zhang and R. Zhang, "Capacity characterization for intelligent reflecting surface aided MIMO communication," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1823–1838, Aug. 2020.
[54] B. Di, H. Zhang, L. Song, Y. Li, Z. Han, and H. V. Poor, "Hybrid beamforming for reconfigurable intelligent surface based multi-user communications: Achievable rates with limited discrete phase shifts," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1809–1822, Aug. 2020.
[55] 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.



Min Fu (Student Member, IEEE) received the B.S. degree in smart grid from the Nanjing University of Science and Technology, Nanjing, China, 2017. She is currently pursuing the Ph.D. degree with the School of Information Science and Technology, ShanghaiTech University, Shanghai, China. Her research interests include optimization and reconfigurable intelligent surfaces and their applications to 6G.

Yong Zhou (Member, IEEE) received the B.Sc. and M.Eng. degrees from Shandong University, Jinan, China, in 2008 and 2011, respectively, and the Ph.D. degree from the University of Waterloo, Waterloo, ON, Canada, in 2015. From November 2015 to January 2018, he worked as a Postdoctoral Research Fellow with the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, Canada. He is currently an Assistant Professor with the School of Information Science and Technology, ShanghaiTech University, Shanghai, China. His research interests include the Internet of Things, edge computing, and reconfigurable intelligent surfaces.



Yuanming Shi (Senior Member, IEEE) received the B.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 2011, and the Ph.D. degree in electronic and computer engineering from The Hong Kong University of Science and Technology (HKUST) in 2015. Since September 2015, he has been with the School of Information Science and Technology, ShanghaiTech University, where he is currently a tenured Associate Professor. He visited the University of California, Berkeley, CA, USA, from October 2016 to February 2017. His research areas include optimization, statistics, machine learning, signal processing, and their applications to 6G, the IoT, and AI. He was a recipient of the 2016 IEEE Marconi Prize Paper Award in Wireless Communications, and the 2016 Young Author Best Paper Award by the IEEE Signal Processing Society. He is also an Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS.

Khaled B. Letaief (Fellow, IEEE) received the B.S. (Hons.), M.S., and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, USA, in December 1984, August 1986, and May 1990, respectively.

From 1990 to 1993, he was a Faculty Member with the University of Melbourne, Australia. Since 1993, he has been with The Hong Kong University of Science and Technology (HKUST). While at HKUST, he has held many administrative positions, including the Dean of Engineering, the Head of the Electronic and Computer Engineering Department, the Director of the Wireless IC Design Center, the Founding Director of the Huawei Innovation Laboratory, and the Director of the Hong Kong Telecom Institute of Information Technology. He also served as consultants for different organizations, including Huawei, ASTRI, ZTE, Nortel, PricewaterhouseCoopers, and Motorola. He has also been involved in organizing many flagship international conferences. He is also an Internationally Recognized Leader in wireless communications and networks with research interest in artificial intelligence, big data analytics systems, mobile cloud and edge computing, tactile Internet, 5G systems and beyond. In these areas, he has over 630 papers with over 38 600 citations and an h-index of 87 along with 15 patents, including 11 US inventions.

Dr. Letaief is currently a member of the United States National Academy of Engineering, fellow of The Hong Kong Institution of Engineers, and a member of the Hong Kong Academy of Engineering Sciences. He is also recognized by Thomson Reuters as an ISI Highly Cited Researcher and was listed among the 2020 top 30 of AI 2000 Internet of Things Most Influential Scholars. He was a recipient of many distinguished awards and honors, including the 2019 Distinguished Research Excellence Award by the HKUST School of Engineering (Highest research award and only one recipient/3 years is honored for his/her contributions), the 2019 IEEE Communications Society and Information Theory Society Joint Paper Award, the 2018 IEEE Signal Processing Society Young Author Best Paper Award, the 2017 IEEE Cognitive Networks Technical Committee Publication Award, the 2016 IEEE Signal Processing Society Young Author Best Paper Award, the 2016 IEEE Marconi Prize Paper Award in Wireless Communications, the 2011 IEEE Wireless Communications Technical Committee Recognition Award, the 2011 IEEE Communications Society Harold Sobol Award, the 2010 Purdue University Outstanding Electrical and Computer Engineer Award, the 2009 IEEE Marconi Prize Award in Wireless Communications, the 2007 IEEE Communications Society Joseph LoCicero Publications Exemplary Award, and over 16 IEEE best paper awards. He is also well recognized for his dedicated service to professional societies and IEEE, where he has served in many leadership positions, including a Treasurer for the IEEE Communications Society, the IEEE Communications Society Vice-President for Conferences, the Chair of IEEE Committee on Wireless Communications, the elected member of IEEE Product Services and Publications Board, and the IEEE Communications Society Vice-President for Technical Activities. He also served as a President of the IEEE Communications Society from 2018 to 2019, the world's leading organization for communications professionals with headquarter in New York and members in 162 countries. He is also the Founding Editor-in-Chief of the prestigious IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and has served on the editorial board of other premier journals, including the Editor-in Chief for IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS—WIRELESS SERIES.