

Power Efficient IRS-Assisted NOMA

Jianyue Zhu, Yongming Huang, Jiaheng Wang, Keivan Navaie, Zhiguo Ding

Abstract

In this paper, we propose a downlink multiple-input single-output (MISO) transmission scheme, which is assisted by an intelligent reflecting surface (IRS) consisting of a large number of passive reflecting elements. In the literature, it has been proved that nonorthogonal multiple access (NOMA) can achieve the capacity region when the quasi-degradation condition is satisfied, which is characterized by users' channels. However, in a conventional communication scenario, it is difficult to guarantee the quasi-degradation, because the channels are determined by the propagation environments and cannot be reconfigured. To overcome this difficulty, we focus on an IRS-assisted MISO NOMA system, where the wireless channels can be effectively tuned. We optimize the beamforming vectors and the IRS phase shift matrix for minimizing transmission power. Furthermore, we propose an improved quasi-degradation condition by using IRS, which can ensure that NOMA achieves the capacity region with high possibility. For a comparison, we study zero-forcing beamforming (ZFBF) as well, where the beamforming vectors and the IRS phase shift matrix are also jointly optimized. Comparing NOMA with ZFBF, it is shown that, with the same IRS phase shift matrix and the improved quasi-degradation condition, NOMA always outperforms ZFBF. At the same time, we identify the condition under which ZFBF outperforms NOMA, which motivates the proposed hybrid NOMA transmission. Simulation results show that the proposed IRS-assisted MISO system outperforms the MISO case without IRS, and the hybrid NOMA transmission scheme always achieves better performance than orthogonal multiple access.

Index Terms

Multiple-input single-output, nonorthogonal multiple access, intelligent reflecting surface, zero-forcing beamforming, quasi-degradation

J. Zhu, Y. Huang, J. Wang are with the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China. (email: {zhuji, huangym, jhwang}@seu.edu.cn).

K. Navaie is with the School of Computing and Communications, Lancaster University, Lancaster, United Kingdom (email: k.navaie@lancaster.ac.uk).

Z. Ding is with the School of Electrical and Electronic Engineering, Manchester University, Manchester, UK (email: zhiguo.ding@manchester.ac.uk).

I. INTRODUCTION

In the beyond fifth generation (B5G) communication systems, there are high requirements in spectrum efficiency, energy consumption, and massive connectivity [1]. In order to meet these high demands, various technologies, such as massive multiple-input multiple-output (MIMO), millimeter wave, and small cell, are being investigated for the B5G communication systems. In addition, nonorthogonal multiple access (NOMA) has also been introduced as a promising multiple access candidate for future mobile networks [2], [3]. Different from the conventional multiple access scheme, i.e., orthogonal multiple access (OMA), NOMA allows multiple users sharing the same resources, such as time, frequency, space, and code, and hence significantly improves the spectrum efficiency [4]–[6].

In addition, being able to provide the flexibility and spatial degrees of freedom needed for efficient resource allocation, multi-antenna techniques have been widely considered in the B5G communication systems [7]. In the literature, there are many works focusing on the coexistence of NOMA and multi-antenna techniques. For example, in [8], the authors applied NOMA in MIMO systems and showed that MIMO-NOMA outperforms conventional MIMO-OMA. In [9], the authors proposed the quasi-degradation condition, under which NOMA can achieve the same performance as dirty paper coding (DPC). This work also demonstrated that NOMA is not always preferable. If users' channels are orthogonal, zero-forcing beamforming (ZFBE) is a more preferred option than NOMA. On the other hand, the users' channels having the same directions is the ideal case for the implementation of NOMA.

Conventionally, in NOMA systems, the directions of users' channels cannot be tuned. This is because the users' channels are determined by propagation environments and hence are highly stochastic. Motivated by this, in this paper, we apply the intelligent reflecting surface (IRS) to NOMA systems to reconfigure the channels.

The IRS, consisting of a large number of passive elements, is introduced as a promising technology for the B5G communication systems [10], [11]. The reflecting elements of the IRS can affect the electromagnetic behavior of the wireless propagation channel. Usually, the IRS operates as a multi-antenna relay but it is fundamentally different from a conventional relay. Specifically, the IRS functions as a reconfigurable scatterer, which requires no dedicated energy resource for decoding, channel estimation, and transmission.

[In the literature, there were a large number of works studying the IRS-assisted communication](#)

system. For example, [12] provided an overview of the IRS technology, including its main applications in wireless communication, competitive advantages over existing technologies, and hardware architecture. In [13], the authors studied the issue of security in the IRS-assisted communication system, where the beamforming vectors, artificial noise covariance matrix, and the phase shifters at the IRSs are jointly optimized. One of the key challenges in adopting IRS-assisted communication systems is to design the IRS phase shift matrix [14]–[16]. In this paper, we focus on the design of beamforming vectors and the IRS phase shift matrix for an IRS-assisted multiple-input single-output (MISO) NOMA communication system.

Among the existing works, there is a few similar works, e.g., [17]–[21], investigating the application of IRS in NOMA systems. In [17], the authors optimized the beamforming vectors and the IRS phase shift matrix for minimizing downlink transmission power in the IRS-empowered NOMA system. [18] proposed a simple design of IRS-assisted NOMA transmission, where the authors characterized the performance of practical IRS NOMA transmission. [19] maximized the sum rate of all users by jointly optimizing the active beamforming at the BS and the passive beamforming at the IRS, where the ideal and non-ideal IRS assumptions are both studied. In a multi-cluster IRS-aided MISO NOMA communication network, [21] jointly optimized the beamforming vectors and the IRS element matrix for minimizing the transmission power. In [20], the authors discussed the comparison between IRS OMA and IRS NOMA. However, all the above works have not found that the IRS-assisted MISO NOMA can achieve the same performance as DPC.

Different from all the above mentioned previous works, in this paper, the advantage of the employment of IRS in NOMA systems is demonstrated and the best achievable performance achieved by using IRS-assisted NOMA is obtained. We optimize the beamforming vectors and the IRS phase shift matrix to minimize the transmission power. Particularly, note that the quasi-degradation condition is used to characterize the gap between the optimal performance of DPC and that achieved by NOMA for the special case of two users [9], [22]. Therefore, for the formulated optimization problem, we consider a quasi-degradation constraint to guarantee that NOMA achieves the same performance as DPC, which can realize the capacity region [23].

Conventionally, whether the quasi-degradation condition can be satisfied depends on the predetermined channels, which means NOMA might not always achieve the capacity region. In addition, DPC is difficult to be implemented in practical communication systems, due to its prohibitively high complexity [23]. By employing the IRS-assisted NOMA, we design a

more practical transmission scheme, which achieves the same performance as DPC. We further investigate optimizing beamforming vectors and the IRS phase shift matrix for IRS-assisted zero-forcing beamforming (ZFBF) scheme and show that it is preferred to IRS-assisted NOMA in certain condition.

In this paper, we study the joint optimization of beamforming vectors and the IRS element matrix respectively for IRS-assisted NOMA and IRS-assisted ZFBF and focus on the fundamental two users' case. The contributions are summarized as follows.

- In IRS-assisted NOMA systems, we study the optimization of beamforming vectors and the IRS phase shift matrix for minimizing the BS transmission power. Particularly, the quasi-degradation is considered as a constraint in the optimization problem, which guarantees that NOMA can achieve the same performance as DPC.
- For IRS-assisted NOMA, we provide an improved quasi-degradation condition under which the feasibility of the optimization problem is guaranteed and the same performance as DPC is obtained.
- We show that the improved quasi-degradation condition for IRS-assisted NOMA can be satisfied with a higher possibility than conventional NOMA without IRS.
- The optimal beamforming solutions with given IRS phase shift matrix are provided according to our previous papers [22] and [9]. Then, with the optimal beamforming, we further optimize the IRS phase shift matrix using semidefinite relaxation (SDR) and quadratic transform.
- For IRS-assisted ZFBF, we also characterize the optimal beamforming in a closed form with a given IRS phase shift matrix. The IRS phase shift matrix is further optimized using fractional programming and successive convex approximation (SCA).
- Comparing IRS-assisted NOMA with IRS-assisted ZFBF, we show that, given the same IRS phase shift matrix and the improved quasi-degradation condition, NOMA achieves better performance than that of ZFBF. However, ZFBF outperforms NOMA under certain condition and hence we provide the hybrid NOMA transmission scheme.

The rest of the paper is organized as follows. In Section II, we describe the system model and formulate the minimizing transmission power problem. In Section III, we analyze the feasibility of the formulated problem in IRS-assisted NOMA systems, and then the beamforming vectors and the IRS phase shift matrix are optimized. In Section IV, we design the beamforming vectors

and the IRS phase shift matrix for IRS-assisted ZFBF. In Section V, we compare the schemes of IRS-assisted NOMA and IRS-assisted ZFBF and propose the hybrid NOMA transmission scheme. Finally, in Section VI, we present simulation results followed by conclusion in Section VII.

Notations: The matrices and vectors are respectively denoted as boldface capital and lower case letters. $\text{Tr}(\mathbf{A})$ and $\text{Rank}(\mathbf{A})$ represent the trace and rank of matrix \mathbf{A} , respectively; \mathbf{a}^H is the Hermitian transpose vector \mathbf{a} ; $\mathbf{A} \succeq \mathbf{0}$ indicates that \mathbf{A} is a positive semidefinite matrix; \mathbf{I}_N is the $N \times N$ identity matrix; \mathbb{C}^T denotes the set of complex numbers; $\|\cdot\|$ denotes the absolute value of the Euclidean vector norm.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model

As shown in Fig. 1, we consider the downlink MISO system wherein a base station (BS) equipped with M antennas serves two single-antenna users. In this MISO system, the communication is assisted by an IRS with N phase shift elements. Let s_i be the message intended to be received by user i with $E[|s_i|^2] = 1$ and $\mathbf{w}_i \in \mathbb{C}^M$ be the complex beamforming vector for user i . In this IRS-assisted MISO system, we consider that the BS exploits superposition coding and hence the received signal at each user is

$$y_k = \mathbf{h}_k^H (\mathbf{w}_1 s_1 + \mathbf{w}_2 s_2) + n_k, k = 1, 2, \quad (1)$$

where $n_k \sim \mathcal{CN}(0, \sigma^2)$ is the additive zero average white Gaussian noise (AWGN) with variance of σ^2 , and

$$\mathbf{h}_k^H = \mathbf{h}_{rk}^H \mathbf{\Theta} \mathbf{G} + \mathbf{h}_{dk}^H \in \mathbb{C}^{1 \times M}, \quad (2)$$

is the channel of user k . In (2), $\mathbf{h}_{rk}^H \in \mathbb{C}^{1 \times N}$ is the channel between the IRS and user k , $\mathbf{G} \in \mathbb{C}^{N \times M}$ denotes the channel between the BS and the IRS, and $\mathbf{h}_{dk}^H \in \mathbb{C}^{1 \times M}$ represents the direct channel between the BS and user k ; $\mathbf{\Theta} = \text{diag}(e^{j\theta_1}, \dots, e^{j\theta_N})$ is the IRS phase shift matrix, where j denotes the imaginary unit and $\theta_n \in [0, 2\pi]$ is the phase shift.

B. Problem Formulation

1) *NOMA Transmission Scheme:* It has been shown that with quasi-degraded channels, the same performance as DPC, i.e., the capacity region of MISO broadcast channels [23], can be

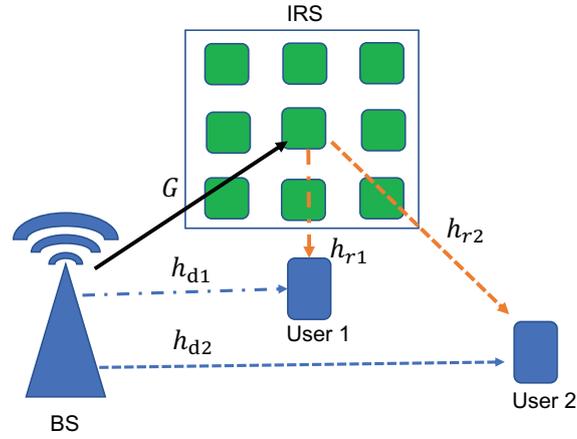


Figure 1. A downlink IRS-assisted MISO system.

achieved by using NOMA [9], [22]. Considering the prohibitively high complexity, DPC is difficult to be implemented in practical communication systems. Thus, we focus on designing a more practical downlink transmission scheme, i.e., NOMA, which can outperform OMA and simultaneously yield a performance close to the capacity region. In addition, it's worthy pointing out, considering the complexity of successive interference cancellation (SIC), always two users are paired to perform NOMA [4], [6]. In this work, we consider the basic two users' scenario by incorporating the matching theory [4].

It has been shown in [9], [22] that whether the quasi-degradation condition can be satisfied depends on the channels between the BS and users, which cannot be controlled. However, in this paper, by using the IRS, the channels of users can be efficiently tuned and we might find the IRS phase shift matrix, i.e., Θ , to guarantee the quasi-degradation. Thus, considering the quasi-degradation constraint with variable Θ , the same performance as DPC is achieved by using NOMA.

In NOMA systems, the successive interference cancellation (SIC) is employed for the users to decode their signals. In single antenna NOMA systems, the optimal signal decoding order is same as the ascending order of the channel gains. However, in the IRS-assisted MISO NOMA system, it is difficult to know the optimal decoding order. It's worthy pointing out that, in this paper, our focus is not the optimal decoding order. To reduce the complexity of beamforming

design, the decoding order is assumed to be the ascending order of the channel gains, which was also adopted in other similar works [9], [24], [25].

In this paper, user 1 is supposed to be the cell center user and user 2 is the cell edge user. Thus, we assume a fixed decoding order, (2, 1). Specifically, user 1 first decodes the signal of user 2 and subtracts this from its received signal, which means user 1 can decode its signal without interference. For user 2, it does not perform SIC and simply decodes its signal by treating user 1's signal as interference. Therefore, the achievable rate of these two users are respectively given by

$$\begin{cases} R_1 = \ln \left(1 + \frac{\mathbf{h}_1^H \mathbf{w}_1 \mathbf{w}_1^H \mathbf{h}_1}{\sigma^2} \right), \\ R_2 = \min \{ \ln (1 + \text{SINR}_{2,1}), \ln (1 + \text{SINR}_{2,2}) \}, \end{cases} \quad (3)$$

where

$$\text{SINR}_{2,1} = \frac{\mathbf{h}_1^H \mathbf{w}_2 \mathbf{w}_2^H \mathbf{h}_1}{\mathbf{h}_1^H \mathbf{w}_1 \mathbf{w}_1^H \mathbf{h}_1 + \sigma^2}, \quad (4)$$

$$\text{SINR}_{2,2} = \frac{\mathbf{h}_2^H \mathbf{w}_2 \mathbf{w}_2^H \mathbf{h}_2}{\mathbf{h}_2^H \mathbf{w}_1 \mathbf{w}_1^H \mathbf{h}_2 + \sigma^2}, \quad (5)$$

respectively denote the signal interference noise ratio (SINR) of user 1 to decode user 2 and user 2 to decode itself.

The performance of the IRS-assisted MISO NOMA system relies on the design of beamforming vectors and the IRS phase shift matrix. In this paper, we focus on the joint optimization of beamforming vectors and the IRS phase matrix in IRS-assisted MISO NOMA systems.

Note that, in this paper, we assume the perfect channel state information (CSI) is available at all nodes. This assumption was also adopted in a large number of works [13], [17], [25]. In the literature, several methods have been proposed to solve the problem of channel estimation in the IRS-assisted wireless system [12], [26], [27]. For example, in [27], the authors proposed that the IRS elements can be divided into multiple groups and each group consists of adjacent elements, which share a common reflection coefficient. Then, using pilot training, the combined channel associated with each group is estimated sequentially.

In the formulated optimization problem, we consider minimizing the transmission power (MTP) subject to the quality of service (QoS) constraints of the two users. In addition, we incorporate an additional important constraint, i.e., the quasi-degradation constraint. This constraint

ensures that the NOMA scheme can achieve the same performance as DPC. The optimization problem is thus formulated as:

$$P^{NOMA} = \min_{\{\mathbf{w}_1, \mathbf{w}_2, \Theta\}} \|\mathbf{w}_1\|^2 + \|\mathbf{w}_2\|^2, \quad (6a)$$

$$\text{s.t. } \frac{\mathbf{h}_1^H \mathbf{w}_1 \mathbf{w}_1^H \mathbf{h}_1}{\sigma^2} \geq r_1^{\min}, \quad (6b)$$

$$\min \left\{ \frac{\mathbf{h}_1^H \mathbf{w}_2 \mathbf{w}_2^H \mathbf{h}_1}{\mathbf{h}_1^H \mathbf{w}_1 \mathbf{w}_1^H \mathbf{h}_1 + \sigma^2}, \frac{\mathbf{h}_2^H \mathbf{w}_2 \mathbf{w}_2^H \mathbf{h}_2}{\mathbf{h}_2^H \mathbf{w}_1 \mathbf{w}_1^H \mathbf{h}_2 + \sigma^2} \right\} \geq r_2^{\min}, \quad (6c)$$

$$\frac{1 + r_1^{\min}}{\cos^2 \alpha} - \frac{r_1^{\min} \cos^2 \alpha}{(1 + r_2^{\min} (1 - \cos^2 \alpha))^2} \leq \frac{\|\mathbf{h}_1\|^2}{\|\mathbf{h}_2\|^2}, \quad (6d)$$

$$0 \leq \theta_i \leq 2\pi, i = 1, \dots, N, \quad (6e)$$

where r_i^{\min} is the corresponding target signal-to-noise ratio (SNR) level, i.e., $\ln(1 + r_i^{\min}) = R_i^{\min}$, $i = 1, 2$, and α is the angel between the two users' channels with

$$\cos^2 \alpha = \frac{\mathbf{h}_1^H \mathbf{h}_2 \mathbf{h}_2^H \mathbf{h}_1}{\|\mathbf{h}_1\|^2 \|\mathbf{h}_2\|^2}. \quad (7)$$

In problem (6a), (6b) and (6c) are respectively the QoS constraints of user 1 and user 2 and (6e) is due to the IRS phase shift matrix, i.e., $\Theta = \text{diag}(e^{j\theta_1}, \dots, e^{j\theta_N})$. In addition, (6d) is the quasi-degradation constraint, which was proposed in [9], [22]. Specifically, from (6d), given the ratio $\frac{\|\mathbf{h}_1\|^2}{\|\mathbf{h}_2\|^2}$, the channels becomes quasi-degraded by simply choosing the angle that satisfies (6d). Actually, aligning one user's channel with the others' is the idea case for using NOMA [18].

To the best of our knowledge, there are very limited works investigating the IRS-assisted NOMA system, e.g., [17], [18], [20]. The similar works also studied the optimization of beamforming vectors and the IRS phase shift matrix for MTP. However, they considered the optimization problem without the quasi-degradation constraint, which means the proposed scheme might not achieve the best performance using NOMA. In contrast here, we fully exploit the advantage of IRS, i.e., the directions of users' channel can be effectively tuned to obtain the same performance as DPC.

2) *ZFBF Transmission Scheme*: Here, we benchmark our proposed IRS-assisted NOMA system against an IRS-assisted ZFBF system. ZFBF is a widely used transmission scheme in the literature. In the ZFBF strategy, the users transmit data in the null space of other users' channels, which mitigates the multi-user interference. If the channels of the two users are orthogonal, the user interference can be avoided by using ZFBF. Thus, we can achieve the best performance using

ZFBF if the channels of users are orthogonal. However, the channels are not always orthogonal. In this paper, we investigate the IRS-assisted ZFBF transmission scheme, where the IRS phase shift matrix can be obtained to make the channels orthogonal under certain condition.

In the IRS-assisted MISO ZFBF system, the MTP problem with QoS constraints is formulated as:

$$P^{ZF} = \min_{\{\mathbf{w}_1, \mathbf{w}_2, \Theta\}} \|\mathbf{w}_1\|^2 + \|\mathbf{w}_2\|^2, \quad (8a)$$

$$\text{s.t. } \frac{\mathbf{h}_1^H \mathbf{w}_1 \mathbf{w}_1^H \mathbf{h}_1}{\sigma^2} \geq r_1^{\min}, \quad (8b)$$

$$\frac{\mathbf{h}_2^H \mathbf{w}_2 \mathbf{w}_2^H \mathbf{h}_2}{\sigma^2} \geq r_2^{\min}, \quad (8c)$$

$$\mathbf{h}_1^H \mathbf{w}_2 = 0, \quad \mathbf{h}_2^H \mathbf{w}_1 = 0, \quad (8d)$$

$$0 \leq \theta_i \leq 2\pi, i = 1, \dots, N, \quad (8e)$$

where (8b) and (8c) are respectively the QoS constraints of user 1 and user 2. In addition, the constraint (8d) is due to the ZFBF principle, i.e., the users transmit data in the null space of other users' channels.

In previous works on IRS, e.g., [18], also used the ZFBF beamforming. [18] employed the IRS to make ZFBF an ideal beamforming for maximizing weak user's rate. In this paper, we focus on the beamforming and the IRS phase shift matrix design to minimize the transmission power. Moreover, in the following, we will show that, under certain condition, the IRS can be used to make the channels orthogonal, which is the ideal case for ZFBF.

III. BEAMFORMING AND IRS PHASE SHIFT DESIGN: IRS-ASSISTED NOMA

In this section, we first investigate the feasibility of the formulated problem (6a). It is worthy pointing out that only if problem (6a) is feasible, the NOMA transmission can lead to the same performance as DPC. Then, given the proposed feasibility condition, we focus on the joint optimization of beamforming vectors and the IRS phase shift matrix. Specifically, we first optimize the beamforming vectors with given IRS phase shift matrix, Θ , and then Θ is optimized.

A. The Feasibility of the Quasi-degradation Constraint

Different from the MISO NOMA system without IRS, the IRS-assisted MISO NOMA system is able to adjust the angle between the channels of user 1 and user 2 to satisfy the quasi-degradation condition. In the following, we show that, compared to the literature MISO NOMA

system without IRS, the IRS-assisted MISO NOMA system has a greater possibility to satisfy the quasi-degradation condition. Here, we derive a sufficient condition to guarantee the feasibility of the quasi-degradation constraint.

Proposition 1. *The quasi-degradation constraint (6d) is feasible if*

$$\lambda_{\max}(\Upsilon_1 - \Upsilon_2) \geq 0, \quad (9)$$

where

$$\Upsilon_k = \begin{bmatrix} \Phi_k \Phi_k^H & \Phi_k \mathbf{h}_{dk} \\ \mathbf{h}_{dk}^H \Phi_k & \mathbf{h}_{dk}^H \mathbf{h}_{dk} \end{bmatrix}, \quad (10)$$

$\Phi_k = \text{diag}(\mathbf{h}_{rk}) \mathbf{G}$ for $k = 1, 2$, and $\lambda_{\max}(\Upsilon_1 - \Upsilon_2)$ denotes the maximum eigenvalue of matrix $\Upsilon_1 - \Upsilon_2$.

Proof. See Appendix A. □

According to Proposition 1, given the improved quasi-degradation condition (9), the IRS phase shift matrix can always be found to satisfy the quasi-degradation constraint (6d). Thus, under condition (9), the proposed MTP problem in (6a) using NOMA is always feasible and the IRS-assisted MISO NOMA scheme can obtain the same performance as DPC.

Corollary 1. *Suppose $\mathbf{h}_1^H \mathbf{h}_2 \neq 0$, the quasi-degradation condition, (6d), is always satisfied if the IRS is located at the position of user 1.*

Proof. See Appendix B. □

In practice, the location of the IRS can be easily controlled [28]. Hence, to realize the capacity region using NOMA, the IRS would better be located close to user 1.

Remark 1. In the following, we further show that, adopting IRS, the quasi-degradation condition would be satisfied with a greater possibility.

To better illustrate the advantage of the employment of IRS in MISO NOMA systems, we provide the following example. In Fig. 2 and Fig. 3, the BS is located at point $(0, 0)$, the IRS central element is located at point $(5, 5)$ and the single-antenna users are randomly placed within a rectangular around the BS. The channel between the IRS and user k is given by $\mathbf{h}_{rk} = d_{rk}^{-\alpha} \mathbf{g}_{rk}$, where d_{rk} is the distance between user k and the BS, $\alpha = 3$ is the path loss exponent, \mathbf{g}_{rk} follows a Rayleigh distribution. Similarly, the channel between the BS and the IRS is $\mathbf{G} = d_r^{-\alpha} \mathbf{G}_r$, and

the direct channel between the BS and user k is $\mathbf{h}_{dk} = d_k^{-\alpha} \mathbf{g}_k$. Assume that the location of user 1 is fixed, we focus on describing different locations of user 2 to satisfy the quasi-degradation condition without IRS [9], i.e.,

$$\frac{1 + r_1^{\min}}{\cos^2 \alpha} - \frac{r_1^{\min} \cos^2 \alpha}{(1 + r_2^{\min} (1 - \cos^2 \alpha))^2} \leq \frac{\|\mathbf{h}_{d1}\|^2}{\|\mathbf{h}_{d2}\|^2}, \quad (11)$$

where

$$\cos^2 \alpha = \frac{\mathbf{h}_{d1}^H \mathbf{h}_{d2} \mathbf{h}_{d2}^H \mathbf{h}_{d1}}{\|\mathbf{h}_{d1}\|^2 \|\mathbf{h}_{d2}\|^2}, \quad (12)$$

and the improved quasi-degradation condition provided in Proposition 1 as (9). In addition, in Fig. 2 and Fig. 3, the black area denotes the locations of user 2 satisfying the quasi-degradation condition without IRS or the improved quasi-degradation condition.

Fig. 2 and Fig. 3 (a) respectively show the region of user 2 to satisfy quasi-degradation condition without IRS, and the improved quasi-degradation condition. In these two figures, the location of user 1 is fixed at (5, 5.5). From Fig. 2 and Fig.3 (a), one can easily find that, with the help of IRS, the black area becomes larger. Therefore, compared with the conventional MISO NOMA systems without IRS, the quasi-degradation condition is more likely to be satisfied in IRS-assisted MISO NOMA systems. Thus, the use of IRS in MISO systems facilitates to achieve the best performance by using NOMA transmission scheme.

Moreover, Fig. 3 (b) depicts the region of user 2 to satisfy the improved quasi-degradation condition. Different from Fig. 3 (a), the location of user 1 is fixed at (5, 5.2). Hence, compared with Fig. 3 (a), the location of user 1 in Fig. 3 (b) is closer to the IRS. From Fig. 3 (a) and Fig. 3 (b), it is easy to see that if user 1 is near the IRS, the quasi-degradation condition is more likely to be satisfied. Actually, according to Corollary 1, if user 1 and the IRS are at the same location, the quasi-degradation condition can always be satisfied.

B. Beamforming Vectors and the IRS Phase Shift Matrix Design for MTP

In this subsection, we focus on optimizing the beamforming vectors and the IRS phase shift matrix for MTP. The infeasibility of problem (6a) lies on the quasi-degradation constraint [17]. Therefore, given (9), the proposed optimization problem (6a) is always feasible. However, problem (6a) is a non-convex problem due to the non-convex constraints with respect to \mathbf{w}_1 , \mathbf{w}_2 , and Θ . In the following, we first optimize the beamforming vectors, i.e., \mathbf{w}_1 , \mathbf{w}_2 , for any given Θ and the results are given in the following Lemma.

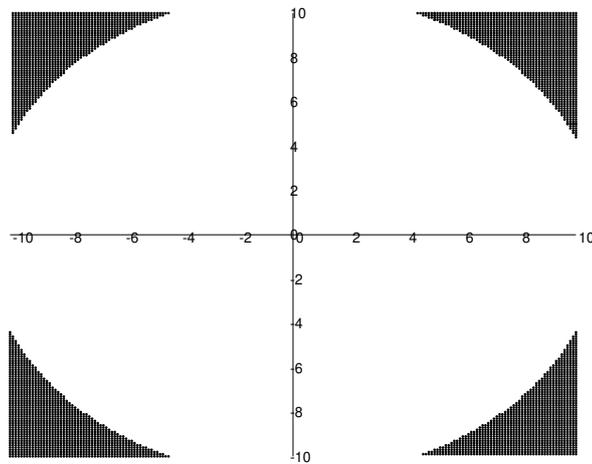
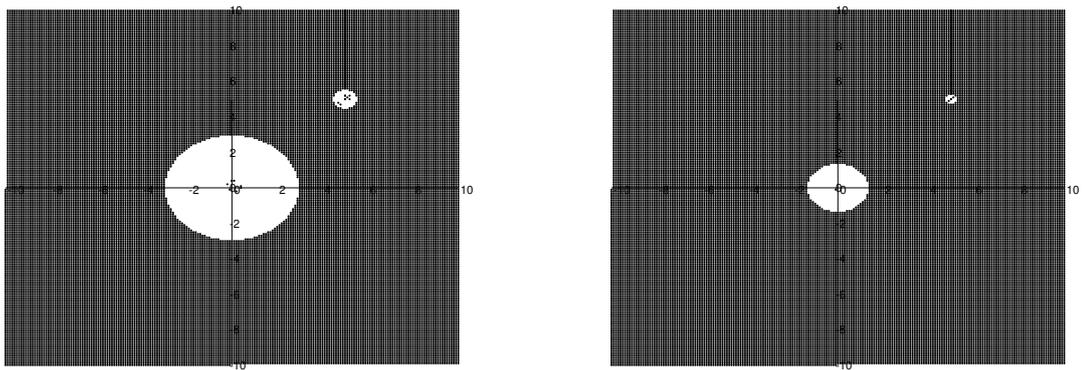


Figure 2. Quasi-degradation region of user 2 without using IRS with fixed location of user 1 at (5, 5.5).



(a) Fixed locations of user 1 at (5, 5.5) and IRS at (5, 5). (b) Fixed locations of user 1 at (5, 5.2) and IRS at (5, 5).

Figure 3. Improved quasi-degradation region of user 2.

Lemma 1. *Given the improved quasi-degradation condition (9), for any given Θ , the optimal beamforming solutions to problem (6a) is*

$$\begin{cases} \mathbf{w}_1^* = \phi_1 (1 + A_2) \mathbf{e}_1 - r_2^{\min} \mathbf{e}_2^H \mathbf{e}_1 \mathbf{e}_2, \\ \mathbf{w}_2^* = \phi_2 \mathbf{e}_2, \end{cases} \quad (13)$$

where

$$\begin{cases} \mathbf{e}_1 = \frac{\mathbf{h}_1}{\|\mathbf{h}_1\|}, \mathbf{e}_2 = \frac{\mathbf{h}_2}{\|\mathbf{h}_2\|}, \\ \phi_1^2 = \frac{r_1^{\min} \sigma^2}{\|\mathbf{h}_1\|^2} \frac{1}{(1+r_2^{\min} \sin^2 \alpha)^2}, \\ \phi_2^2 = \frac{r_2^{\min} \sigma^2}{\|\mathbf{h}_2\|^2} + \frac{r_1^{\min} \sigma^2}{\|\mathbf{h}_1\|^2} \frac{r_2^{\min} \cos^2 \alpha}{(1+r_2^{\min} \sin^2 \alpha)^2}. \end{cases} \quad (14)$$

Proof. See the proof of Proposition 1 in [9]. \square

From Lemma 1, for any given Θ , the optimal beamforming is characterized in a closed form expression. Then, using the closed-form beamforming solutions, we optimize the IRS phase shift matrix, i.e., $\Theta = \text{diag}(e^{j\theta_1}, \dots, e^{j\theta_N})$. The corresponding optimization problem is

$$\min_{\{\Theta\}} F, \quad (15a)$$

$$\text{s.t.} \quad \frac{1+r_1^{\min}}{\cos^2 \alpha} - \frac{r_1^{\min} \cos^2 \alpha}{(1+r_2^{\min} (1-\cos^2 \alpha))^2} \leq \frac{\|\mathbf{h}_1\|^2}{\|\mathbf{h}_2\|^2}, \quad (15b)$$

$$0 \leq \theta_i \leq 2\pi, i = 1, \dots, N, \quad (15c)$$

where

$$\begin{aligned} F &= \|\mathbf{w}_1^*\|^2 + \|\mathbf{w}_2^*\|^2, \\ &= \phi_1^2 \left((1+r_2^{\min})^2 - (2+r_2^{\min}) r_2^{\min} \cos^2 \alpha \right) + \phi_2^2, \end{aligned} \quad (16)$$

and ϕ_1^2, ϕ_2^2 have been written in (14).

To solve problem (15a), we introduce variable $\mathbf{Q} = \tilde{\mathbf{v}}\tilde{\mathbf{v}}^H$, with $\tilde{\mathbf{v}} = [\mathbf{v}; 1]$, and $\mathbf{v} = [e^{j\theta_1}, \dots, e^{j\theta_N}]^H$. Therefore, for $k = 1, 2$, $\mathbf{h}_k^H = \mathbf{h}_{r_k}^H \Theta \mathbf{G} + \mathbf{h}_{d_k}^H = \mathbf{v}^H \Phi_k + \mathbf{h}_{d_k}^H$, with $\Phi_k = \text{diag}(\mathbf{h}_{r_k}) \mathbf{G}$ and

$$\|\mathbf{h}_k\|^2 = \text{Tr}(\mathbf{Q}\Upsilon_k), \quad (17)$$

$$\mathbf{h}_1^H \mathbf{h}_2 = \text{Tr}(\mathbf{Q}\mathbf{R}), \quad (18)$$

where Υ_k has been given in (10) and

$$\mathbf{R} = \begin{bmatrix} \Phi_1 \Phi_2^H & \Phi_1 \mathbf{h}_{d2} \\ \mathbf{h}_{d1}^H \Phi_2^H & \mathbf{h}_{d1}^H \mathbf{h}_{d2} \end{bmatrix}. \quad (19)$$

Therefore, problem (15a) is reduced to the following

$$\min_{\{\mathbf{Q} \succeq \mathbf{0}\}} \phi_1^2 \left((1 + r_2^{\min})^2 - (2 + r_2^{\min}) r_2^{\min} \cos^2 \alpha \right) + \phi_2^2, \quad (20a)$$

$$\text{s.t. } \frac{1 + r_1^{\min}}{\cos^2 \alpha} - \frac{r_1^{\min} \cos^2 \alpha}{(1 + r_2^{\min} (1 - \cos^2 \alpha))^2} \leq \frac{\text{Tr}(\mathbf{Q}\Upsilon_1)}{\text{Tr}(\mathbf{Q}\Upsilon_2)}, \quad (20b)$$

$$\mathbf{Q}_{i,i} = 1, i = 1, \dots, N + 1, \quad (20c)$$

$$\text{Rank}(\mathbf{Q}) = 1, \quad (20d)$$

where

$$\cos^2 \alpha = \frac{\text{Tr}(\mathbf{Q}\mathbf{R}) \text{Tr}(\mathbf{R}^H \mathbf{Q}^H)}{\text{Tr}(\mathbf{Q}\Upsilon_1) \text{Tr}(\mathbf{Q}\Upsilon_2)}, \quad (21)$$

$$\phi_1^2 = \frac{r_1^{\min} \sigma^2}{\text{Tr}(\mathbf{Q}\Upsilon_1)} \frac{1}{(1 + r_2^{\min} \sin^2 \alpha)^2}, \quad (22)$$

and

$$\phi_2^2 = \frac{r_2^{\min} \sigma^2}{\text{Tr}(\mathbf{Q}\Upsilon_2)} + \frac{r_1^{\min} \sigma^2}{\text{Tr}(\mathbf{Q}\Upsilon_1)} \frac{r_2^{\min} \cos^2 \alpha}{(1 + r_2^{\min} \sin^2 \alpha)^2}. \quad (23)$$

Note that the constraints in (20c) and (20d) come from the introduced variable substitution: $\mathbf{Q} = \tilde{\mathbf{v}}\tilde{\mathbf{v}}^H$. Problem (20a) is very difficult to solve due to the tricky objective function and the quasi-degradation constraint. In practice, the objective function and the quasi-degradation constraint are both nonconvex due to the complex expression of $\cos^2 \alpha$. In the following, we first deal with the quasi-degradation constraint, which will be transformed into a convex constraint.

Proposition 2. *The quasi-degradation constraint can be satisfied if*

$$\chi_1 \mathbf{Q} \chi_1^H - (1 + r_1^{\min}) \chi_2 \mathbf{Q} \chi_2^H \succeq \mathbf{0}, \quad (24)$$

where $\chi_k = [\Phi_k^H, \mathbf{h}_{dk}]$, $k = 1, 2$.

Proof. See Appendix C. □

The quasi-degradation constraint (24) in Proposition 2 is convex. We then focus on the complex nonconvex objective function. An upper bound of the objective function is given in Proposition 3.

Proposition 3. *The upper bound of the objective function in problem (15a) is given by*

$$F \leq \frac{\sigma^2 (r_1^{\min} (1 + r_2^{\min}))}{\text{Tr}(\mathbf{Q}\Upsilon_1)} + \frac{\sigma^2 r_2^{\min}}{\text{Tr}(\mathbf{Q}\Upsilon_2)}. \quad (25)$$

Proof. See Appendix D. □

Using Proposition 2 and Proposition 3, an approximation solution to problem (20a) is obtained by solving the following problem

$$\min_{\{\mathbf{Q} \succeq \mathbf{0}\}} \frac{\sigma^2 (r_1^{\min} (1 + r_2^{\min}))}{\text{Tr}(\mathbf{Q}\mathbf{Y}_1)} + \frac{\sigma^2 r_2^{\min}}{\text{Tr}(\mathbf{Q}\mathbf{Y}_2)}, \quad (26a)$$

$$\text{s.t. } \boldsymbol{\chi}_1 \mathbf{Q} \boldsymbol{\chi}_1^H - (1 + r_1^{\min}) \boldsymbol{\chi}_2 \mathbf{Q} \boldsymbol{\chi}_2^H \succeq \mathbf{0}, \quad (26b)$$

$$\mathbf{Q}_{i,i} = 1, i = 1, \dots, N + 1, \quad (26c)$$

$$\text{Rank}(\mathbf{Q}) = 1. \quad (26d)$$

Note that (26a) is actually a sum of ratios problem. In order to solve this multiple-ratio fractional programming, we exploit the quadratic transform, which is described in the following Lemma [29].

Lemma 2. [29] *The sum-of-ratio problem*

$$\max_{\{\mathbf{x}\}} \sum_{m=1}^M \frac{A_m(\mathbf{x})}{B_m(\mathbf{x})}, \quad (27)$$

$$\text{s.t. } \mathbf{x} \in \mathcal{X},$$

is equivalent to

$$\max_{\{\mathbf{x}, \mathbf{y}\}} \sum_{m=1}^M \left(2y_m \sqrt{A_m(\mathbf{x})} - y_m^2 B_m(\mathbf{x}) \right), \quad (28)$$

$$\text{s.t. } \mathbf{x} \in \mathcal{X}, y_m \in \mathbb{R}, m = 1, \dots, M,$$

where \mathbf{y} refers to a collection of variables $\{y_1, \dots, y_M\}$.

Using Lemma 2, problem (26a) is then reduced to

$$\max_{\{\mathbf{Q} \succeq \mathbf{0}, y_1, y_2\}} f, \quad (29a)$$

$$\text{s.t. } \boldsymbol{\chi}_1 \mathbf{Q} \boldsymbol{\chi}_1^H - (1 + r_1^{\min}) \boldsymbol{\chi}_2 \mathbf{Q} \boldsymbol{\chi}_2^H \succeq \mathbf{0}, \quad (29b)$$

$$\mathbf{Q}_{i,i} = 1, i = 1, \dots, N + 1, \quad (29c)$$

$$\text{Rank}(\mathbf{Q}) = 1, \quad (29d)$$

where

$$f = -2y_1 \sqrt{\sigma^2 (r_1^{\min} (1 + r_2^{\min}))} + y_1^2 \text{Tr}(\mathbf{Q}\mathbf{Y}_1) - 2y_2 \sqrt{\sigma^2 r_2^{\min}} + y_2^2 \text{Tr}(\mathbf{Q}\mathbf{Y}_2).$$

Here, we propose to optimize the primal variable \mathbf{Q} and the auxiliary variables y_1, y_2 iteratively. Specifically, we first optimize \mathbf{Q} by fixing the auxiliary variables y_1, y_2 and then further optimize y_1, y_2 . However, by fixing y_1, y_2 , (29a) is also nonconvex due to the rank constraint (29d). Here, we exploit the semidefinite relaxation (SDR) [30] method to solve the nonconvex problem (29a) and the relaxed problem is therefore:

$$\max_{\{\mathbf{Q} \succeq \mathbf{0}\}} y_1^2 \text{Tr}(\mathbf{Q}\Upsilon_1) + y_2^2 \text{Tr}(\mathbf{Q}\Upsilon_2), \quad (30a)$$

$$\text{s.t. } \chi_1 \mathbf{Q} \chi_1^H - (1 + r_1^{\min}) \chi_2 \mathbf{Q} \chi_2^H \succeq \mathbf{0}, \quad (30b)$$

$$\mathbf{Q}_{i,i} = 1, i = 1, \dots, N + 1. \quad (30c)$$

One can easily show that problem (30a) is a standard semidefinite programming (SDP), which can be efficiently solved through convex optimization tools, e.g., CVX. However, the solution to problem (30a) cannot always satisfy the rank constraint, i.e., $\text{Rank}(\mathbf{Q}) = 1$. In general, if the solution to the relaxed problem in (30a) is a set of rank-one matrices, then it will be also the optimal solution to the problem (29a). Otherwise, the randomization technique, see, e.g., [31], can be used to generate a set of rank-one solutions.

In addition, with fixed \mathbf{Q} , the optimal y_1, y_2 are obtained in closed form as

$$y_1^* = \frac{\sqrt{\sigma^2 (r_1^{\min} (1 + r_2^{\min}))}}{\text{Tr}(\mathbf{Q}\Upsilon_1)}, \quad (31)$$

$$y_2^* = \frac{\sqrt{\sigma^2 r_2^{\min}}}{\text{Tr}(\mathbf{Q}\Upsilon_2)}. \quad (32)$$

Therefore, the optimal \mathbf{Q} can thus be efficiently obtained through numerical convex optimization. The steps of the proposed approach are summarized in Algorithm 1.

Note that, using Lemma 2, in problem (26a), each $A_m(\mathbf{x})$ is a constant and every $B_m(\mathbf{x})$ is convex. Moreover, after using SDR, the constraints in problem (26a) are all convex. Thus, the proposed Algorithm 1 is guaranteed to converge to a local optimum of (29a) [29].

In addition, the complexity of Algorithm 1 is given as follows. Since step 3 is a SDR procedure, the complexity for solving (26a) is $\mathcal{O}(\sqrt{2(N+1)})$ [32]. Thus, the overall computational complexity is $\mathcal{O}(L_1(\sqrt{2(N+1)}))$, where L_1 is the number of iterations of Algorithm 1.

Additional steps are needed to obtain the beamforming vectors and IRS elements. Specifically, we first obtain the eigenvalue decomposition of \mathbf{Q} as $\mathbf{Q} = \mathbf{U}\Lambda\mathbf{U}$, where $\mathbf{U} = [e_1, \dots, e_{N+1}]$ is

a unitary matrix and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_N)$ is a diagonal matrix. Therefore, the IRS elements are given as

$$e^{j\theta_k} = e^{j\arg\left(\frac{\tilde{v}_k}{\tilde{v}_{N+1}}\right)}, k = 1, \dots, N. \quad (33)$$

In (33), $\tilde{\mathbf{v}} = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{r}$ and $\mathbf{r} \in \mathbb{C}^{(N+1) \times 1}$ is chosen from a lot of randomly generalized vectors satisfying $\mathbf{r} \in \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N+1})$. Then, the beamforming vectors can be easily obtained using Lemma 1.

Note that the proposed solution to the original problem (20a) is suboptimal. Due to the tricky objective function and the quasi-degradation constraint, we use the approximation method and SDR, which results in the suboptimal solution of the IRS element matrix. However, the proposed beamforming with given IRS element matrix is optimal and can even achieve the best performance as DPC, which is the main contribution in this work.

Algorithm 1 Iterative Approach for solving (29a)

1: Initialize y_1, y_2 to a feasible value.

2: **Repeat**

3: Update \mathbf{Q} by solving the SDR problem in (30a).

4: Update y_1, y_2 by using (31) and (32).

5: **Until** The value of objective function in (30a) converges or the maximum number of iterations is reached.

IV. BEAMFORMING AND IRS PHASE SHIFT DESIGN: IRS-ASSISTED ZFBF

The transmission scheme of ZFBF is in fact a special case of OMA as the spatial degrees of freedom are used for interference avoidance. Here, we investigate the IRS-assisted ZFBF transmission scheme, where the beamforming vectors and the IRS phase shift matrix are optimized to achieve the minimum transmission power.

In order to solve problem (8a), we first optimize the beamforming vectors, \mathbf{w}_1 and \mathbf{w}_2 , by fixing Θ . Lemma 3 gives the optimal beamforming solutions.

Lemma 3. For any given Θ , the optimal beamforming solutions to problem in (8a) are

$$\begin{cases} \mathbf{w}_1^* = \frac{\sqrt{r_1^{\min}} \mathbf{a}_1}{\|\mathbf{h}_1\|^2 \|\mathbf{h}_2\|^2 \sin^2 \alpha}, \\ \mathbf{w}_2^* = \frac{\sqrt{r_2^{\min}} \mathbf{a}_2}{\|\mathbf{h}_1\|^2 \|\mathbf{h}_2\|^2 \sin^2 \alpha}, \end{cases} \quad (34)$$

where

$$\sin^2 \alpha = 1 - \frac{\mathbf{h}_1^H \mathbf{h}_2 \mathbf{h}_2^H \mathbf{h}_1}{\|\mathbf{h}_1\|^2 \|\mathbf{h}_2\|^2}, \quad (35)$$

and

$$\begin{cases} \mathbf{a}_1 = \|\mathbf{h}_2\|^2 \mathbf{h}_1 - (\mathbf{h}_2^H \mathbf{h}_1) \mathbf{h}_2, \\ \mathbf{a}_2 = -(\mathbf{h}_1^H \mathbf{h}_2) \mathbf{h}_1 + \|\mathbf{h}_1\|^2 \mathbf{h}_2. \end{cases} \quad (36)$$

Proof. See Appendix E. \square

From Lemma 3, with a given Θ , the optimal beamforming solutions to problem (8a) is characterized in closed form expressions. Thereby, using the closed-form beamforming solutions, we further optimize the IRS phase shift matrix and the corresponding optimization problem is

$$\min_{\{\Theta\}} W = \frac{\sigma^2}{\sin^2 \alpha} \left(\frac{r_1^{\min}}{\|\mathbf{h}_1\|^2} + \frac{r_2^{\min}}{\|\mathbf{h}_2\|^2} \right), \quad (37a)$$

$$\text{s.t. } 0 \leq \theta_i \leq 2\pi, i = 1, \dots, N. \quad (37b)$$

Similarly, to solve problem (37a), we also use the introduced variables, i.e., $\mathbf{Q} = \tilde{\mathbf{v}} \tilde{\mathbf{v}}^H$ with $\tilde{\mathbf{v}} = [\mathbf{v}; 1]$ and $\mathbf{v} = [e^{j\theta_1}, \dots, e^{j\theta_N}]^H$. Then, problem (37a) is rewritten as

$$\min_{\{\mathbf{Q} \succeq \mathbf{0}\}} W = \frac{\sigma^2}{\sin^2 \alpha} \left(\frac{r_1^{\min}}{\text{Tr}(\mathbf{Q}\Upsilon_1)} + \frac{r_2^{\min}}{\text{Tr}(\mathbf{Q}\Upsilon_2)} \right), \quad (38a)$$

$$\text{s.t. } \mathbf{Q}_{i,i} = 1, i = 1, \dots, N + 1, \quad (38b)$$

$$\text{Rank}(\mathbf{Q}) = 1, \quad (38c)$$

where

$$\sin^2 \alpha = 1 - \frac{\text{Tr}(\mathbf{Q}\mathbf{R}) \text{Tr}(\mathbf{R}^H \mathbf{Q}^H)}{\text{Tr}(\mathbf{Q}\Upsilon_1) \text{Tr}(\mathbf{Q}\Upsilon_2)}. \quad (39)$$

The objective function in (38a) is a nonconvex and fractional objective, which is rewritten as

$$\begin{aligned} W &= \frac{\sigma^2}{\sin^2 \alpha} \left(\frac{r_1^{\min}}{\text{Tr}(\mathbf{Q}\Upsilon_1)} + \frac{r_2^{\min}}{\text{Tr}(\mathbf{Q}\Upsilon_2)} \right), \\ &= \frac{\sigma^2 r_1^{\min} \text{Tr}(\mathbf{Q}\Upsilon_2) + \sigma^2 r_2^{\min} \text{Tr}(\mathbf{Q}\Upsilon_1)}{\text{Tr}(\mathbf{Q}\Upsilon_1) \text{Tr}(\mathbf{Q}\Upsilon_2) - \text{Tr}(\mathbf{Q}\mathbf{R}) \text{Tr}(\mathbf{R}^H \mathbf{Q}^H)}. \end{aligned} \quad (40)$$

In order to solve the problem in (38a), we introduce the following objective function:

$$G(\mathbf{Q}, \eta) = G_1(\mathbf{Q}) - \eta G_2(\mathbf{Q}), \quad (41)$$

where η is a positive parameter,

$$G_1(\mathbf{Q}) = \sigma^2 r_1^{\min} \text{Tr}(\mathbf{Q}\Upsilon_2) + \sigma^2 r_2^{\min} \text{Tr}(\mathbf{Q}\Upsilon_1), \quad (42)$$

and

$$G_2(\mathbf{Q}) = \text{Tr}(\mathbf{Q}\Upsilon_1) \text{Tr}(\mathbf{Q}\Upsilon_2) - \text{Tr}(\mathbf{Q}\mathbf{R}) \text{Tr}(\mathbf{R}^H \mathbf{Q}^H). \quad (43)$$

We then study the following optimization problem with given η :

$$\min_{\{\mathbf{Q} \succeq \mathbf{0}\}} G(\mathbf{Q}, \eta) = G_1(\mathbf{Q}) - \eta G_2(\mathbf{Q}), \quad (44a)$$

$$\text{s.t. } \mathbf{Q}_{i,i} = 1, i = 1, \dots, N+1, \quad (44b)$$

$$\text{Rank}(\mathbf{Q}) = 1, \quad (44c)$$

The relation between problem (38a) and (44a) is given by the following lemma.

Lemma 4. [33] *Let $G^*(\eta)$ be the optimal objective value of problem (44a) and $\mathbf{Q}^*(\eta)$ be the optimal solution of problem (44a). Then, $\mathbf{Q}^*(\eta)$ is the optimal solution to (38a) if and only if $G^*(\eta) = 0$.*

According to Lemma 4, the optimal solution to (38a) can be found by solving problem (44a) parameterized by η and then updating η until $G^*(\eta) = 0$. Therefore, we first focus on solving (44a) for a given η .

For a given η , the non-convexity of problem (44a) lies on the rank constraint and part of the objective function $G_2(\mathbf{Q}, \eta)$. In the following, for a given η , we will solve the nonconvex problem (44a) by using successive approximation (SCA) [34] and SDR.

Firstly, we construct global under-estimators of $G_2(\mathbf{Q}, \eta)$. Specifically, for any feasible point \mathbf{Q}^i , the function $G_2(\mathbf{Q}, \eta)$ satisfies the following inequality:

$$\begin{aligned} G_2(\mathbf{Q}, \eta) &\geq G_2(\mathbf{Q}^i) + \text{Tr}\left(\left(\nabla_{\mathbf{Q}} G_2(\mathbf{Q}^i)\right)^H (\mathbf{Q} - \mathbf{Q}^i)\right), \\ &= \tilde{G}_2(\mathbf{Q}, \mathbf{Q}^i), \end{aligned} \quad (45)$$

where

$$\nabla_{\mathbf{Q}} G_2(\mathbf{Q}^i) = \Upsilon_1^H \text{Tr}(\mathbf{Q}^i \Upsilon_2) + \Upsilon_2^H \text{Tr}(\mathbf{Q}^i \Upsilon_1) - 2\mathbf{R}^H \text{Tr}(\mathbf{Q}^i \mathbf{R}). \quad (46)$$

Therefore, for any given \mathbf{Q}^i , an upper bound of problem (44a) can be obtained by solving the following optimization problem

$$\min_{\{\mathbf{Q} \succeq \mathbf{0}\}} G(\mathbf{Q}, \eta) = G_1(\mathbf{Q}) - \eta \tilde{G}_2(\mathbf{Q}, \mathbf{Q}^i), \quad (47a)$$

$$\text{s.t. } \mathbf{Q}_{i,i} = 1, i = 1, \dots, N+1, \quad (47b)$$

$$\text{Rank}(\mathbf{Q}) = 1. \quad (47c)$$

Optimization problem (47a) 's non-convexity only lies on the rank constraint (47c). To tackle this issue, we remove the rank constraint (47c) by applying SDR. Then, the relaxed version of (47a) can be efficiently solved via standard optimization tools, such as CVX.

Note that (47a) serves as an upper bound of (44a). The process of SCA is summarized in Algorithm 2, where we iteratively tighten the upper bound and obtain a sequence of solutions, i.e., \mathbf{Q} .

In Algorithm 2, the solution is a KKT point of problem (47a) without the rank constraint. This is because the following conditions are satisfied:

- (1) $G_1(\mathbf{Q}) - \eta G_2(\mathbf{Q}) \leq G_1(\mathbf{Q}) - \eta \tilde{G}_2(\mathbf{Q}, \mathbf{Q}^i)$;
- (2) $G_1(\mathbf{Q}^i) - \eta G_2(\mathbf{Q}^i) = G_1(\mathbf{Q}^i) - \eta \tilde{G}_2(\mathbf{Q}^i, \mathbf{Q}^i)$;
- (3) $\frac{\partial(G_1(\mathbf{Q}^i) - \eta G_2(\mathbf{Q}^i))}{\partial \mathbf{Q}^j} = \frac{\partial(G_1(\mathbf{Q}^i) - \eta \tilde{G}_2(\mathbf{Q}^i, \mathbf{Q}^i))}{\partial \mathbf{Q}^j}$;

Thus, according to [35], SCA converges to a KKT point of the original problem (44a) without the rank constraint. And the complexity of Algorithm 2 is to solve (47a), whose complexity is $\mathcal{O}\left(\sqrt{3(N+1)}\right)$ [32].

Algorithm 2 Successive Convex Approximation Algorithm

- 1: **Initialize** iteration index $i = 1$ and a feasible \mathbf{Q}^1 .
 - 2: **Repeat**
 - 3: Solve problem (47a) for a given \mathbf{Q}^i and store the solution \mathbf{Q} ;
 - 4: Set $i = i + 1$ and $\mathbf{Q}^i = \mathbf{Q}$;
 - 5: **Until** The value of objective function in (47a) converges or the maximum number of iterations is reached.
 - 6: $\mathbf{Q}^* = \mathbf{Q}^i$.
-

After the solution to problem (44a) is obtained, we shall find an η such that $G^*(\eta) = 0$, which can be achieved by Algorithm 3. In Algorithm 3, δ denotes a small positive threshold.

Algorithm 3 is guaranteed to converge to the desirable η [33]. In addition, note that Algorithm 3 iteratively uses Algorithm 2, hence the complexity of Algorithm 3 is $\mathcal{O}\left(L_2 \sqrt{3(N+1)}\right)$, where L_2 is the number of iterations of Algorithm 3.

Finally, we can achieve the solution of IRS elements by using the SVD of \mathbf{Q}^* and then the beamforming vectors are obtained using Lemma 3. Similar to the IRS-assisted NOMA, the proposed solution of the IRS element matrix may not be optimal, which is due to the use of SDR

and SCA. However, with given IRS phase shift matrix, the proposed solution of beamforming vector is optimal.

Algorithm 3 The solution to problem (44a)

- 1: **Initialize** $\eta_{ini} = 0$, $G^*(\eta) = \infty$, and precision $\delta > 0$.
 - 2: **While** $|G^*(\eta)| > \delta$ **do**
 - 3: Find the solution \mathbf{Q}^* using Algorithm 2;
 - 4: Calculate $G^*(\eta)$;
 - 5: Update $\eta = \frac{G_1(\mathbf{Q}^*)}{G_2(\mathbf{Q}^*)}$;
 - 6: **Return** η and \mathbf{Q}^* .
-

V. COMPARISON OF IRS-ASSISTED NOMA AND IRS-ASSISTED ZFBF

From the previous sections, we have respectively investigated the joint optimization of beamforming vectors and the IRS phase shift matrix for IRS-assisted NOMA and IRS-assisted ZFBF. The application of IRS in multi-antenna systems facilitates the implementation of NOMA and ZFBF, which is because the directions of users' channel vectors can be effectively aligned. In this section, we further study the comparison between schemes of IRS-assisted NOMA and IRS-assisted ZFBF.

According to Lemma 1, with improved quasi-degraded condition, the minimum transmission power achieved by NOMA is

$$\begin{aligned} P^{NOMA} &= \phi_1^2 \left((1 + r_2^{\min})^2 - (2 + r_2^{\min}) r_2^{\min} \cos^2 \alpha \right) + \phi_2^2, \\ &= \frac{r_1^{\min} (1 + r_2^{\min}) \sigma^2}{\|\mathbf{h}_1\|^2 (1 + r_2^{\min} \sin^2 \alpha)} + \frac{r_2^{\min} \sigma^2}{\|\mathbf{h}_2\|^2}. \end{aligned} \quad (48)$$

In addition, from Lemma 3, the minimum transmission power achieved by ZFBF is given by

$$P^{ZFBF} = \frac{\sigma^2}{\sin^2 \alpha} \left(\frac{r_1^{\min}}{\|\mathbf{h}_1\|^2} + \frac{r_2^{\min}}{\|\mathbf{h}_2\|^2} \right). \quad (49)$$

In the following, we compare P^{NOMA} and P^{ZFBF} .

Theorem 1. *Given the same IRS phase shift matrix Θ and improved quasi-degraded condition, we always have*

$$P^{NOMA} \leq P^{ZFBF}. \quad (50)$$

Proof. See Appendix F.

According to Proposition 1, given the same IRS phase shift matrix and the improved quasi-degraded condition, the scheme of NOMA always achieves better performance than ZFBF. In addition, in the following, we consider a specific situation where the channels of user 1 and user 2 are orthogonal, i.e., $\mathbf{h}_1^H \mathbf{h}_2 = 0$. \square

Proposition 4. *The improved quasi-degradation condition is invalid when $\mathbf{h}_1^H \mathbf{h}_2 = 0$, which can be satisfied if and only if*

$$\mathbf{R} = \begin{bmatrix} \Phi_1 \Phi_2^H & \mathbf{h}_{d1}^H \Phi_2^H \\ \Phi_1 \mathbf{h}_{d2} & \mathbf{h}_{d1}^H \mathbf{h}_{d2} \end{bmatrix} = \mathbf{0}, \quad (51)$$

and ZFBF achieves the optimal value.

Proof. See Appendix G. \square

From Proposition 4, the IRS phase shift matrix can be found to make the channels of user 1 and user 2 orthogonal only with condition (51). In practice, the condition (51) is very difficult to satisfy due to the random \mathbf{h}_{rk} , \mathbf{h}_{dk} , for $k = 1, 2$, and \mathbf{G} . However, as shown in Subsection III-A, the improved quasi-degradation condition for MISO NOMA is more easily satisfied, especially for cases where the IRS is located close to user 1.

Remark 2. Based on the provided results, the hybrid NOMA precoding scheme can be given as follows:

- Given the improved quasi-degradation condition, which can be satisfied with high possibility, NOMA achieves the optimal performance and hence we prefer to use NOMA transmission scheme.
- If $\mathbf{R} = \mathbf{0}$, under which the IRS phase shift matrix can be found to satisfy $\mathbf{h}_1^H \mathbf{h}_2 = 0$, the ZFBF is preferred.
- Assume a case where the improved quasi-degradation condition is violated and $\mathbf{R} \neq \mathbf{0}$, the performance loss is inevitable. Considering the computational complexity, one may prefer to use ZFBF as the closed form beamforming cannot be obtained by using NOMA.

VI. SIMULATION RESULTS

In this section, the performance of the proposed solution to the MTP problem is evaluated. In our simulation, the BS is located at the point (0m, 0m), and single-antenna users are randomly placed in the half right-hand side square (200m \times 200m) [36] around the BS. The channel

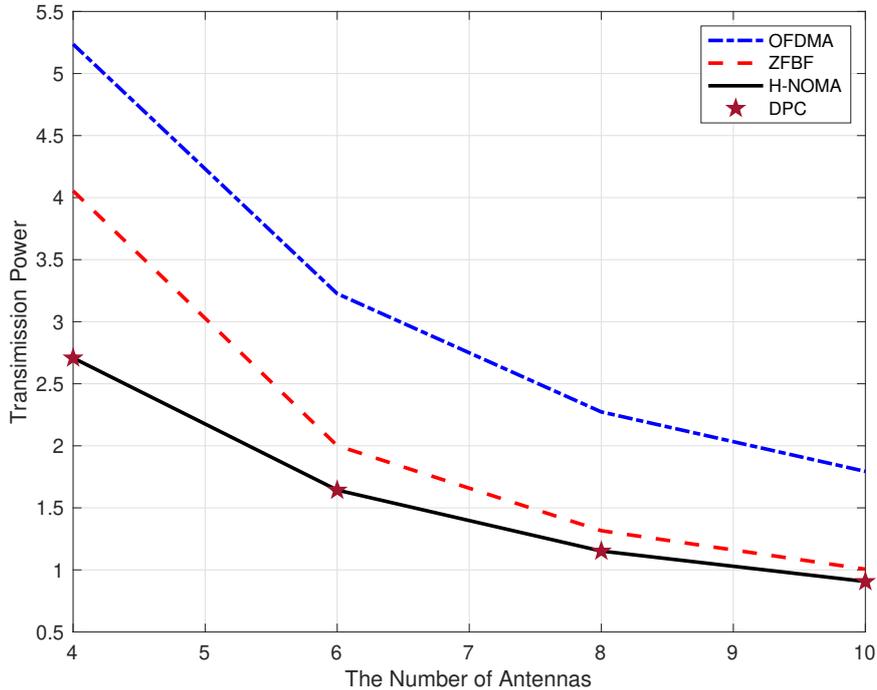


Figure 4. Transmission power versus the number of antennas.

between the IRS and user k is given by $\mathbf{h}_{rk} = d_{rk}^{-\alpha} \mathbf{g}_{rk}$, where d_{rk} is the distance between user k and the BS, $\alpha = 3$ is the path loss exponent, \mathbf{g}_{rk} follows a Rayleigh distribution. Similarly, the channel between the BS and the IRS is $\mathbf{G} = d_r^{-\alpha} \mathbf{G}_r$ and the direct channel between the BS and user k is $\mathbf{h}_{dk} = d_k^{-\alpha} \mathbf{g}_k$ [4]. The noise power is given by $\sigma^2 = BN_0$ with bandwidth $B = 10\text{MHz}$ and the noise power spectral density $N_0 = -174\text{dBm}$. The target rates for both users are the same ($R_k^{\min} = 1\text{bps/Hz}$ for $k = 1, 2$).

Fig. 4 depicts the BS transmission power versus the number of antennas respectively using the schemes of hybrid NOMA (H-NOMA), ZFBF, and OFDMA. In Fig. 4, the IRS is located at the point (20m, 20m) and the number of IRS elements is $N = 100$. Here, note that all the schemes are assisted by IRS. The minimum transmission powers achieved by NOMA and ZFBF use our provided solution. In the IRS-assisted OFDMA scheme, we first optimize the beamforming with given Θ and then further optimize Θ using SDP. As expected, for the three schemes, the total transmission power decreases with the increase of the number of antennas. In comparison with ZFBF and OFDMA schemes, H-NOMA yields a significant performance gain, which is because

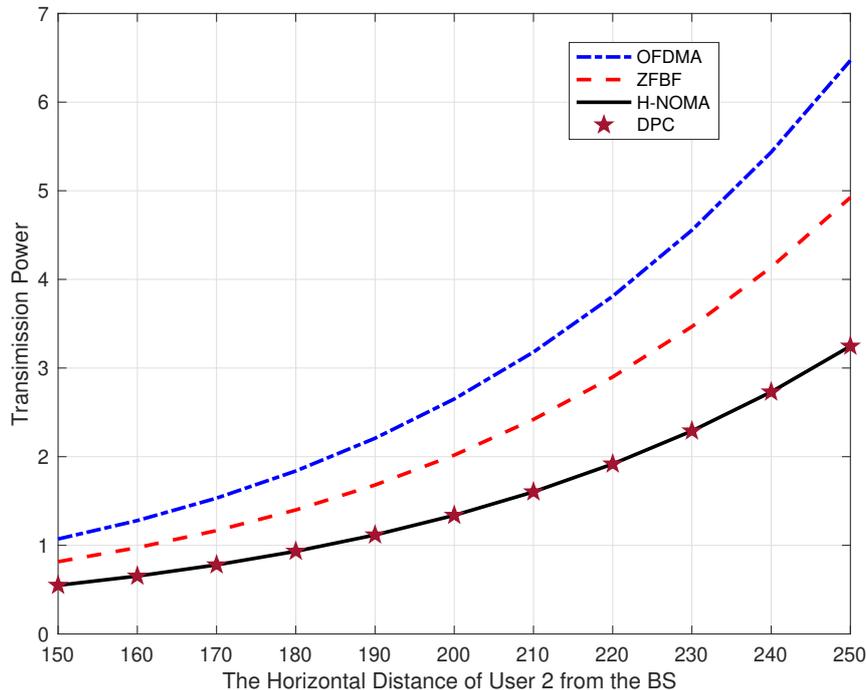


Figure 5. Transmission power versus the horizontal distance of user 2 from the BS.

NOMA allows the users sharing the same spectrum and spatial resources and hence improves the performance. Furthermore, with limited number of antennas, the advantage of H-NOMA scheme is more obvious. In addition, one can easily find that the performance achieved by the proposed H-NOMA scheme is nearly same as DPC, which is owing to the employment of IRS.

In Fig. 5, we display the BS transmission power versus the horizontal distance of user 2 from the BS. Here, the location of user 2 is $(Dm, 150m)$ and D is the horizontal distance of user 2 from the BS, which ranges from 150m to 250m in the simulation. In addition, in Fig. 5, the locations of user 1 and IRS is respectively $(100m, 100m)$ and $(20m, 20m)$, the number of antennas is $M = 4$, and the number of IRS elements is $N = 100$. As can be seen in this figure, when user 2 is farther from the BS, the BS needs to consume more power, which is in line with our expectation. Meanwhile, some similar phenomenon as Fig. 4 that H-NOMA performs better than ZFBF and OFDMA and the performance achieved by H-NOMA is nearly same as DPC. Moreover, one can easily find out that, as the distance between the two users becomes larger, the H-NOMA scheme yields a significant performance gain.

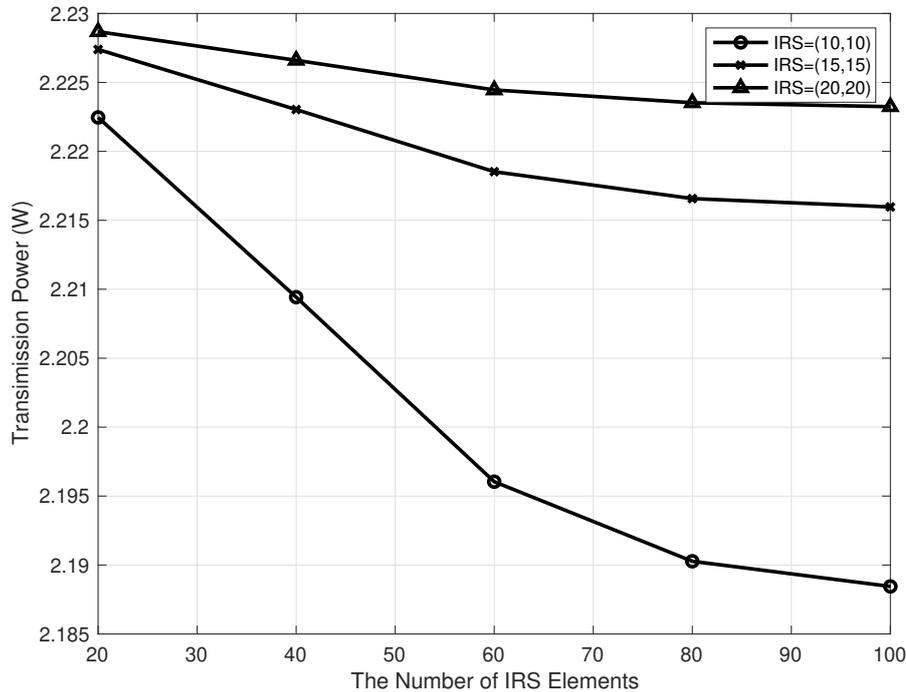


Figure 6. Transmission power versus the number of IRS elements.

Fig. 6 shows the transmission power versus different number of IRS elements with different IRS locations using the provided H-NOMA transmission scheme. Here, the number of antennas is $M = 4$. Obviously, one can achieve better performance by increasing N , which is because a larger N enables more reflecting elements to receive the signal energy from the BS. Furthermore, if the IRS is close to the BS, a significant performance gain can be achieved by increasing N . Conversely, there has no obvious performance gain by increasing N when the user is far from the BS. Therefore, the number of IRS elements can be properly selected according to the location of IRS.

In Fig. 7, the performance both in the IRS-assisted MISO NOMA scheme and the literature MISO NOMA scheme without IRS is evaluated. The number of IRS elements is $N = 100$. In comparison with the literature MISO NOMA, the use of IRS in MISO NOMA yields a significant performance gain, which is because the quasi-degradation condition might be guaranteed and hence the IRS-assisted MISO NOMA achieves the same performance as DPC. Moreover, the performance gain is especially high when the number of antennas is small, which in line with

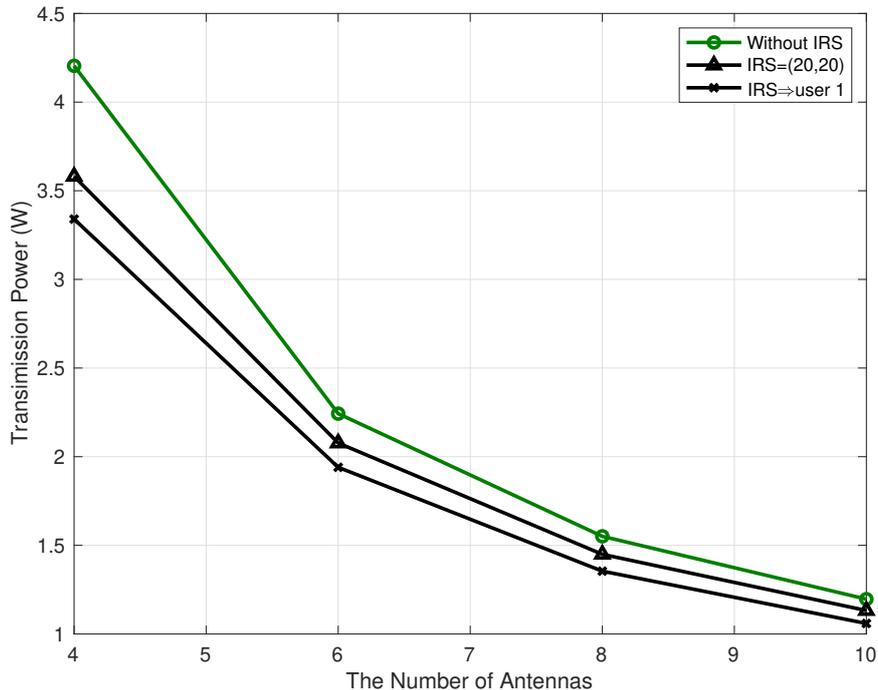


Figure 7. Transmission power versus the number of antennas.

the advantage of NOMA, i.e., supporting overloaded transmission with limited resources. In addition, in Fig. 7, we also present the performance with different locations of IRS. It is easy to see that if the location of IRS is set very close to user 1, the performance is improved. This also re-proves the Corollary 1 via simulation. Actually, when the location of IRS is set very close to user 1, the quasi-degradation condition can always be satisfied and the performance region can be obtained in the IRS-assisted MISO NOMA system.

VII. CONCLUSION

In this paper, we optimized the beamforming vectors and the IRS phase shift matrix respectively for IRS-assisted NOMA and IRS-assisted ZFBF. For IRS-assisted NOMA, we provided the improved quasi-degradation condition under which NOMA can always achieve the same performance as DPC. Furthermore, we show that the improved quasi-degradation condition can be satisfied with greater possibility than the original quasi-degradation condition without IRS. Both for IRS-assisted NOMA and IRS-assisted ZFBF, we characterized the optimal beamforming

with given IRS phase shift matrix and proposed algorithms to optimize the phase shift matrix. The IRS-assisted NOMA and the IRS-assisted ZFBF were further compared. It is shown that, with the same IRS phase shift matrix and improved quasi-degradation condition, NOMA always outperforms ZFBF. At the same time, we provided certain condition under which the IRS phase shift matrix can be found to generate orthogonal channels and hence the hybrid NOMA transmission scheme was proposed. Simulation results show that the IRS-assisted NOMA system not only outperforms the NOMA system without IRS but also the conventional OMA. In addition, in the future work, we would like to extend the two users case to multiple users case using matching theory [4].

APPENDIX

A. Proof of Proposition 1

We consider the feasibility of the quasi-degradation constraint in a converse way. Assume that for each possible \mathbf{h}_{rk} , \mathbf{h}_{dk} , r_k^{\min} , $k = 1, 2$, and Θ , we always have

$$\frac{1 + r_1^{\min}}{\cos^2 \alpha} - \frac{r_1^{\min} \cos^2 \alpha}{(1 + r_2^{\min} (1 - \cos^2 \alpha))^2} > \frac{\|\mathbf{h}_1\|^2}{\|\mathbf{h}_2\|^2}. \quad (52)$$

Since $0 \leq \cos^2 \alpha \leq 1$, we should have

$$\begin{aligned} & \frac{1 + r_1^{\min}}{\cos^2 \alpha} - \frac{r_1^{\min} \cos^2 \alpha}{(1 + r_2^{\min} (1 - \cos^2 \alpha))^2} \\ & \geq \frac{1 + r_1^{\min}}{\cos^2 \alpha} - r_1^{\min} \cos^2 \alpha > \frac{\|\mathbf{h}_1\|^2}{\|\mathbf{h}_2\|^2}, \end{aligned} \quad (53)$$

which is equivalently rewritten as

$$\frac{(1 + r_1^{\min}) \|\mathbf{h}_2\|^4}{\mathbf{h}_2^H \mathbf{h}_1 \mathbf{h}_1^H \mathbf{h}_2} - \frac{r_1^{\min} \mathbf{h}_2^H \mathbf{h}_1 \mathbf{h}_1^H \mathbf{h}_2}{\|\mathbf{h}_1\|^4} > 1. \quad (54)$$

Using (54) and notify $\mathbf{h}_2^H \mathbf{h}_1 \mathbf{h}_1^H \mathbf{h}_2 \leq \mathbf{h}_1^H \mathbf{h}_1 \mathbf{h}_2^H \mathbf{h}_2$, we then obtain

$$\begin{aligned} & \frac{(1 + r_1^{\min}) \|\mathbf{h}_2\|^4}{\mathbf{h}_2^H \mathbf{h}_1 \mathbf{h}_1^H \mathbf{h}_2} - \frac{r_1^{\min} \mathbf{h}_2^H \mathbf{h}_1 \mathbf{h}_1^H \mathbf{h}_2}{\|\mathbf{h}_1\|^4}, \\ & \geq \frac{(1 + r_1^{\min}) \mathbf{h}_2^H \mathbf{h}_2}{\mathbf{h}_1^H \mathbf{h}_1} - \frac{r_1^{\min} \mathbf{h}_2^H \mathbf{h}_2}{\mathbf{h}_1^H \mathbf{h}_1} = \frac{\mathbf{h}_2^H \mathbf{h}_2}{\mathbf{h}_1^H \mathbf{h}_1} > 1, \end{aligned} \quad (55)$$

implying

$$\mathbf{h}_2^H \mathbf{h}_2 - \mathbf{h}_1^H \mathbf{h}_1 > 0. \quad (56)$$

Let $\tilde{\mathbf{v}} = [\mathbf{v}; 1]$ and $\mathbf{v} = [e^{j\theta_1}, \dots, e^{j\theta_N}]^H$, we can write

$$\begin{aligned}
\mathbf{h}_k^H \mathbf{h}_k &= (\mathbf{h}_{rk}^H \Theta \mathbf{G} + \mathbf{h}_{dk}^H) (\mathbf{h}_{rk}^H \Theta \mathbf{G} + \mathbf{h}_{dk}^H)^H, \\
&= (\mathbf{v}^H \Phi_k + \mathbf{h}_{dk}^H) (\mathbf{v}^H \Phi_k + \mathbf{h}_{dk}^H)^H \\
&= \mathbf{v}^H \Phi_k \Phi_k^H \mathbf{v} + \mathbf{v}^H \Phi_k \mathbf{h}_{dk} + \mathbf{h}_{dk}^H \Phi_k^H \mathbf{v} + \mathbf{h}_{dk}^H \mathbf{h}_{dk}, \\
&= \tilde{\mathbf{v}}^H \Upsilon_k \tilde{\mathbf{v}}.
\end{aligned} \tag{57}$$

Then, for each $\tilde{\mathbf{v}}$, we have

$$\mathbf{h}_2^H \mathbf{h}_2 - \mathbf{h}_1^H \mathbf{h}_1 = \tilde{\mathbf{v}}^H (\Upsilon_2 - \Upsilon_1) \tilde{\mathbf{v}} > 0. \tag{58}$$

Therefore, $\Upsilon_2 - \Upsilon_1 \succ \mathbf{0}$. Hence, one can easily obtain $\lambda_{\min}(\Upsilon_2 - \Upsilon_1) > 0$, where $\lambda_{\min}(\Upsilon_2 - \Upsilon_1)$ denotes the minimum eigenvalue of matrix $\Upsilon_2 - \Upsilon_1$.

Therefore, if $\lambda_{\min}(\Upsilon_2 - \Upsilon_1) > 0$, the IRS phase shift matrix cannot be found to satisfy the quasi-degradation constraint. Conversely, if $\lambda_{\min}(\Upsilon_2 - \Upsilon_1) \leq 0$, i.e., $\lambda_{\max}(\Upsilon_1 - \Upsilon_2) \geq 0$, the quasi-degradation constraint is always feasible. This completes the proof.

B. Proof of Corollary 1

Assume the IRS is set at the location of user 1, we note that $\|\mathbf{h}_{r1}\|^2 \rightarrow +\infty$, and hence $\frac{\|\mathbf{h}_1\|^2}{\|\mathbf{h}_2\|^2} = \frac{\|\mathbf{h}_{r1}\Theta\mathbf{G} + \mathbf{h}_{d1}\|^2}{\|\mathbf{h}_{r2}\Theta\mathbf{G} + \mathbf{h}_{d2}\|^2} \rightarrow +\infty$. In addition, the function

$$L(\cos^2 \alpha) = \frac{1 + r_1^{\min}}{\cos^2 \alpha} - \frac{r_1^{\min} \cos^2 \alpha}{(1 + r_2^{\min} (1 - \cos^2 \alpha))^2} \tag{59}$$

is monotonically decreasing function of $\cos^2 \alpha$. The supposition $\mathbf{h}_1^H \mathbf{h}_2 \neq 0$ implies that $\cos^2 \alpha \geq \kappa > 0$. Therefore, $L(\cos^2 \alpha) \leq L(\kappa) < +\infty$, this in turn implies that the quasi-degradation condition, (6d), can always be satisfied.

C. Proof of Proposition 2

Since $\mathbf{h}_k \mathbf{h}_k^H = \chi_k \mathbf{Q} \chi_k^H$ for $k = 1, 2$, using condition (24), we always have

$$\mathbf{h}_1 \mathbf{h}_1^H - (1 + r_1^{\min}) \mathbf{h}_2 \mathbf{h}_2^H \succeq \mathbf{0}, \tag{60}$$

and hence

$$\mathbf{h}_2^H (\mathbf{h}_1 \mathbf{h}_1^H - (1 + r_1^{\min}) \mathbf{h}_2 \mathbf{h}_2^H) \mathbf{h}_2 \geq 0, \tag{61}$$

which is equivalently written as

$$\frac{1 + r_1^{\min}}{\cos^2 \alpha} \leq \frac{\|\mathbf{h}_1\|^2}{\|\mathbf{h}_2\|^2}. \tag{62}$$

In addition, according to the quasi-degradation condition (6d), with $0 \leq \cos^2 \alpha \leq 1$, we have

$$\frac{1 + r_1^{\min}}{\cos^2 \alpha} - \frac{r_1^{\min} \cos^2 \alpha}{(1 + r_2^{\min} (1 - \cos^2 \alpha))^2} \leq \frac{1 + r_1^{\min}}{\cos^2 \alpha}. \quad (63)$$

Combining (62) and (63), we have

$$\frac{1 + r_1^{\min}}{\cos^2 \alpha} - \frac{r_1^{\min} \cos^2 \alpha}{(1 + r_2^{\min} (1 - \cos^2 \alpha))^2} \leq \frac{\|\mathbf{h}_1\|^2}{\|\mathbf{h}_2\|^2}, \quad (64)$$

which is in fact the quasi-degradation constraint shown in (20b).

D. Proof of Proposition 3

Since $0 \leq \cos^2 \alpha \leq 1$, we have

$$\begin{aligned} F &= \phi_1^2 \left((1 + r_2^{\min})^2 - (2 + r_2^{\min}) r_2^{\min} \cos^2 \alpha \right) + \phi_2^2, \\ &= \frac{\sigma^2 (r_1^{\min} (1 + r_2^{\min}))}{\text{Tr}(\mathbf{Q}\mathbf{Y}_1)} \frac{1}{1 + r_2^{\min} (1 - \cos^2 \alpha)} + \frac{\sigma^2 r_2^{\min}}{\text{Tr}(\mathbf{Q}\mathbf{Y}_2)}, \\ &\leq \frac{\sigma^2 (r_1^{\min} (1 + r_2^{\min}))}{\text{Tr}(\mathbf{Q}\mathbf{Y}_1)} + \frac{\sigma^2 r_2^{\min}}{\text{Tr}(\mathbf{Q}\mathbf{Y}_2)}, \end{aligned} \quad (65)$$

which completes the proof.

E. Proof of Lemma 3

By using the least square property of Moore-Penrose inverse [37], the optimal solution to problem (8a) with given Θ is trivially obtained. Specifically, let $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2]$, the optimal solution is given as

$$\begin{aligned} [\mathbf{w}_1^*, \mathbf{w}_2^*] &= \mathbf{H}^{H\dagger} \begin{bmatrix} \sqrt{r_1^{\min}} & 0 \\ 0 & \sqrt{r_2^{\min}} \end{bmatrix}, \\ &= \frac{[\sqrt{r_1^{\min}} \mathbf{a}_1, \sqrt{r_2^{\min}} \mathbf{a}_2]}{\|\mathbf{h}_1\|^2 \|\mathbf{h}_2\|^2 \sin^2 \alpha}, \end{aligned} \quad (66)$$

where \dagger represents the Moore-Penrose inverse.

F. Proof of Theorem 1

Note that

$$\frac{r_2^{\min} \sigma^2}{\|\mathbf{h}_2\|^2} \leq \frac{\sigma^2}{\sin^2 \alpha} \frac{r_2^{\min} \sigma^2}{\|\mathbf{h}_2\|^2}, \quad (67)$$

and

$$\begin{aligned} \frac{r_1^{\min} (1 + r_2^{\min}) \sigma^2}{\|\mathbf{h}_1\|^2 (1 + r_2^{\min} \sin^2 \alpha)} &\leq \frac{r_1^{\min} (1 + r_2^{\min}) \sigma^2}{\|\mathbf{h}_1\|^2 (\sin^2 \alpha + r_2^{\min} \sin^2 \alpha)}, \\ &\leq \frac{r_1^{\min} \sigma^2}{\|\mathbf{h}_1\|^2 \sin^2 \alpha}. \end{aligned} \quad (68)$$

Therefore,

$$\begin{aligned} P^{NOMA} &= \frac{r_1^{\min} (1 + r_2^{\min}) \sigma^2}{\|\mathbf{h}_1\|^2 (1 + r_2^{\min} \sin^2 \alpha)} + \frac{r_2^{\min} \sigma^2}{\|\mathbf{h}_2\|^2}, \\ &\leq \frac{r_1^{\min} \sigma^2}{\|\mathbf{h}_1\|^2 \sin^2 \alpha} + \frac{\sigma^2}{\sin^2 \alpha} \frac{r_2^{\min} \sigma^2}{\|\mathbf{h}_2\|^2}, \\ &= \frac{\sigma^2}{\sin^2 \alpha} \left(\frac{r_1^{\min}}{\|\mathbf{h}_1\|^2} + \frac{r_2^{\min}}{\|\mathbf{h}_2\|^2} \right) = P^{ZFBB}. \end{aligned} \quad (69)$$

G. Proof of Proposition 4

Given $\mathbf{h}_1^H \mathbf{h}_2 = 0$, we have $\cos^2 \alpha = 0$ and hence

$$\frac{1 + r_1^{\min}}{\cos^2 \alpha} - \frac{r_1^{\min} \cos^2 \alpha}{(1 + r_2^{\min} (1 - \cos^2 \alpha))^2} \rightarrow +\infty,$$

which implying the quasi-degradation condition in (6d) is invalid.

Here, we prove that $\mathbf{h}_1^H \mathbf{h}_2 = 0$ can be satisfied if and only if $\mathbf{R} = \mathbf{0}$. First, we prove the necessary condition. If $\mathbf{h}_1^H \mathbf{h}_2 = 0$, i.e., the channel vectors of the two users are orthogonal, we have

$$\mathbf{h}_1^H \mathbf{h}_2 = \tilde{\mathbf{v}}^H \mathbf{R} \tilde{\mathbf{v}} = 0. \quad (70)$$

Since $\tilde{\mathbf{v}} \neq \mathbf{0}$, (51) implies that \mathbf{R} is a zero matrix, i.e., condition (51).

We then prove the sufficiency. Given $\mathbf{R} = \mathbf{0}$, we have $\mathbf{v}^H \mathbf{R} \mathbf{v} = 0$, for any possible \mathbf{v} . In addition, from (70), one can easily find out that

$$\mathbf{v}^H \mathbf{R} \mathbf{v} = \mathbf{h}_1^H \mathbf{h}_2 = 0. \quad (71)$$

From Lemma 3, if $\mathbf{h}_1^H \mathbf{h}_2 = 0$, the minimum transmission power of the BS achieved by using ZFBF is

$$p = \sigma^2 \left(\frac{r_1^{\min}}{\|\mathbf{h}_1\|^2} + \frac{r_2^{\min}}{\|\mathbf{h}_2\|^2} \right), \quad (72)$$

which also satisfies

$$p \leq P^{ZFBB} = \frac{\sigma^2}{\sin^2 \alpha} \left(\frac{r_1^{\min}}{\|\mathbf{h}_1\|^2} + \frac{r_2^{\min}}{\|\mathbf{h}_2\|^2} \right), \quad (73)$$

and

$$p \leq P^{NOMA} = \frac{r_1^{\min} (1 + r_2^{\min}) \sigma^2}{\|\mathbf{h}_1\|^2 (1 + r_2^{\min} \sin^2 \alpha)} + \frac{r_2^{\min} \sigma^2}{\|\mathbf{h}_2\|^2}. \quad (74)$$

From (73) and (74), one can easily find that if the channels of user 1 and user 2 are orthogonal, the ZFBF scheme obtains the optimal performance.

REFERENCES

- [1] K. David and H. Berndt, "6G vision and requirements: Is there any need for beyond 5G?" *IEEE Veh. Technol. Mag.*, vol. 13, no. 3, pp. 72–80, Sep. 2018.
- [2] 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, Sept. 2015.
- [3] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2294–2323, May 2018.
- [4] J. Zhu, J. Wang, Y. Huang, S. He, X. You, and L. Yang, "On optimal power allocation for downlink non-orthogonal multiple access systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2744–2757, Dec. 2017.
- [5] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, C. I. and H. V. Poor, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, Feb. 2017.
- [6] Z. Ding, M. Peng, and H. V. Poor, "Cooperative non-orthogonal multiple access in 5G systems," *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1462–1465, Aug. 2015.
- [7] 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.
- [8] Z. Ding, F. Adachi, and H. V. Poor, "The application of MIMO to non-orthogonal multiple access," *IEEE Trans. on Wireless Commun.*, vol. 15, no. 1, pp. 537–552, Jan. 2016.
- [9] Z. Chen, Z. Ding, X. Dai, and G. K. Karagiannidis, "On the application of quasi-degradation to MISO-NOMA downlink," *IEEE Trans. on Signal Process.*, vol. 64, no. 23, pp. 6174–6189, Dec. 2016.
- [10] W. Tang, M. Z. Chen, J. Y. Dai, Y. Zeng, X. Zhao, S. Jin, Q. Cheng, and T. J. Cui, "Wireless communications with programmable metasurface: New paradigms, opportunities, and challenges on transceiver design," *IEEE Wireless Commun.*, vol. 27, no. 2, pp. 180–187, Mar. 2020.
- [11] W. Tang, M. Z. Chen, X. Chen, J. Y. Dai, Y. Han, M. Di Renzo, Y. Zeng, S. Jin, Q. Cheng, and T. J. Cui, "Wireless communications with reconfigurable intelligent surface: Path loss modeling and experimental measurement," *arXiv preprint arXiv:1911.05326*, 2019.
- [12] 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, Nov. 2020.
- [13] 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.*, pp. 1–1, Jul. 2020.
- [14] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design," in *Proc. IEEE Global Communications Conf.*, Abu Dhabi, Dec. 2018, pp. 1–6.
- [15] D. Xu, X. Yu, Y. Sun, D. W. K. Ng, and R. Schober, "Resource allocation for secure IRS-assisted multiuser MISO systems," in *Proc. IEEE Globecom Workshops*, Waikoloa, HI, Dec. 2019, pp. 1–6.
- [16] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Trans. on Commun.*, vol. 68, no. 3, pp. 1838–1851, Dec. 2020.

- [17] M. Fu, Y. Zhou, and Y. Shi, "Intelligent reflecting surface for downlink non-orthogonal multiple access networks," in *Proc. IEEE Globecom Workshops*, Waikoloa, HI, Dec. 2019, pp. 1–6.
- [18] Z. Ding and H. V. Poor, "A simple design of IRS-NOMA transmission," *IEEE Commun. Lett.*, pp. 1–1, Feb. 2020.
- [19] X. Mu, Y. Liu, L. Guo, J. Lin, and N. Al-Dhahir, "Exploiting intelligent reflecting surfaces in multi-antenna aided NOMA systems," *arXiv preprint arXiv:1910.13636*, 2019.
- [20] B. Zheng, Q. Wu, and R. Zhang, "Intelligent reflecting surface-assisted multiple access with user pairing: NOMA or OMA?" *IEEE Commun. Lett.*, vol. 24, no. 4, pp. 753–757, Jan. 2020.
- [21] Y. Li, M. Jiang, Q. Zhang, and J. Qin, "Joint beamforming design in multi-cluster MISO NOMA intelligent reflecting surface-aided downlink communication networks," *arXiv preprint arXiv:1909.06972*, 2019.
- [22] J. Zhu, J. Wang, Y. Huang, K. Navaie, Z. Ding, and L. Yang, "On optimal beamforming design for downlink MISO NOMA systems," *IEEE Trans. on Veh. Technol.*, vol. 69, no. 3, pp. 3008–3020, Jan. 2020.
- [23] H. Weingarten, Y. Steinberg, and S. S. Shamai, "The capacity region of the gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.
- [24] Z. Chen, Z. Ding, P. Xu, and X. Dai, "Optimal precoding for a QoS optimization problem in two-user MISO-NOMA downlink," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1263–1266, Jun. 2016.
- [25] T. Hou, Y. Liu, Z. Song, X. Sun, Y. Chen, and L. Hanzo, "Reconfigurable intelligent surface aided NOMA networks," *IEEE J. Sel. Areas Commun.*, pp. 1–1, Jul. 2020.
- [26] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," *arXiv preprint arXiv:1904.10136*, 2019.
- [27] Y. Yang, B. Zheng, S. Zhang, and R. Zhang, "Intelligent reflecting surface meets OFDM: Protocol design and rate maximization," *IEEE Trans. on Commun.*, vol. 68, no. 7, pp. 4522–4535, Mar. 2020.
- [28] M. Di Renzo, M. Debbah, D.-T. Phan-Huy, A. Zappone, M.-S. Alouini, C. Yuen, V. Sciancalepore, G. C. Alexandropoulos, J. Hoydis, H. Gacanin *et al.*, "Smart radio environments empowered by AI reconfigurable meta-surfaces: An idea whose time has come," *arXiv preprint arXiv:1903.08925*, 2019.
- [29] K. Shen and W. Yu, "Fractional programming for communication systems-part I: Power control and beamforming," *IEEE Trans. on Signal Process.*, vol. 66, no. 10, pp. 2616–2630, May 2018.
- [30] W.-K. K. Ma, "Semidefinite relaxation of quadratic optimization problems and applications," *IEEE Signal Process. Mag.*, vol. 1053, no. 5888/10, 2010.
- [31] Y. Huang and D. P. Palomar, "Rank-constrained separable semidefinite programming with applications to optimal beamforming," *IEEE Trans. on Signal Process.*, vol. 58, no. 2, pp. 664–678, Feb. 2010.
- [32] J. Gondzio and T. Terlaky, "A computational view of interior point methods," *JE Beasley. Advances in linear and integer programming. Oxford Lecture Series in Mathematics and its Applications*, vol. 4, pp. 103–144, 1996.
- [33] W. Dinkelbach, "On nonlinear fractional programming," *Management science*, vol. 13, no. 7, pp. 492–498, 1967.
- [34] H. Kurtaran, A. Eskandarian, D. Marzougui, and N. Bedewi, "Crashworthiness design optimization using successive response surface approximations," *Computational mechanics*, vol. 29, no. 4-5, pp. 409–421, 2002.
- [35] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Journal of the Operations Research Society of America*, vol. 26, no. 4, pp. 681–683, Aug. 1978.
- [36] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design," in *Proc. IEEE Global Communications Conference*, Dec. 2018, pp. 1–6.
- [37] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.