

Research Article

Joint User Pairing and Power Allocation Scheme Based on Transmission Mode Switching between NOMA-Based Maximum Ratio Transmission and MMSE Beamforming in Downlink MISO Systems

Zhixin Zhao ¹, Dong Wang ², Hongwei Zhang ¹, and Haitao Sang ¹

¹School of Information Engineering, Lingnan Normal University, Zhanjiang 524048, China

²School of Life Science & Technology, Lingnan Normal University, Zhanjiang 524048, China

Correspondence should be addressed to Zhixin Zhao; zhaozhixin0830@gmail.com

Received 2 January 2021; Revised 6 February 2021; Accepted 14 February 2021; Published 9 March 2021

Academic Editor: Ding Xu

Copyright © 2021 Zhixin Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper investigates the design of the joint user pairing and power allocation scheme with transmission mode switching (TMS) in downlink multiple-input-single-output (MISO) systems. Firstly, the closed-form expressions of the boundary of achievable rate region of two candidate transmission modes, i.e., non-orthogonal multiple access based maximum ratio transmission (NOMA-MRT) and minimum mean square error beamforming (MMSE-BF), are obtained. By obtaining the outer boundary of the union of the achievable rate regions of the two transmission modes, an adaptive switching method is developed to achieve a larger rate region. Secondly, based on the idea that the solution to the problem of weighted sum rate (WSR) optimization must be on the boundary of the achievable rate region, the optimal solutions to the problem of WSR optimization for NOMA-MRT and MMSE-BF are obtained for the two-user case, respectively. Subsequently, by exploiting the aforementioned optimal solutions for two transmission modes and the high efficiency of TMS, a suboptimal user pairing and power allocation algorithm (JUPA) is proposed to further improve the sum rate performance for the multiuser case. Compared with the exhaustive search-based user pairing and power allocation algorithm (ES-PPA), the proposed JUPA can enjoy a much lower computational complexity and only suffers a slight sum rate performance loss, and it outperforms other traditional schemes. Finally, numerical results are provided to validate the analyses and the proposed algorithms.

1. Introduction

With the continuous emergence of new application scenarios, one of the challenges faced by the future wireless communication systems is how to provide higher-speed downlink transmission, restricted to the scarce spectrum resources. The traditional downlink transmission schemes used in mobile communication systems are based on orthogonal multiple access (OMA) technology, e.g., frequency division multiple access (FDMA) for the first generation (1G), time division multiple access (TDMA) for 2G, code division multiple access (CDMA) for 3G, and orthogonal frequency division multiple access (OFDMA) [1, 2] for 4G. These conventional multiple access schemes can mitigate or

avoid the interuser interference by allocating orthogonal resources (frequency/time/code) to different users, which result in insufficient use of spectrum resources.

Recently, non-orthogonal multiple access (NOMA) has been considered as a promising multiple access scheme for 5th Generation (5G) wireless communication systems, owing to its higher spectrum efficiency compared with the conventional orthogonal multiple access schemes [3–5]. Note that Third Generation Partnership Project (3GPP) has considered NOMA as a study item for 5G new radio (NR) in Release 15 and decided to leave it for possible use in Beyond 5G (B5G) [6]. NOMA enables multiple users to share the same time-frequency resource block on the same spatial layer. NOMA is achieved by the combination of

superposition encoding and successive interference cancellation (SIC), which is a method to reach the boundary of the capacity region of degraded broadcast channel [7].

In order to further enhance system performance, beamforming (BF) was combined with NOMA in multiple-input-single-output (MISO) downlink [8–14]. In [8], the sum rate maximization problem with BF vector being the optimization variable was studied and an one-dimensional iterative algorithm was proposed. However, for each iteration, a second-order cone program convex problem needs to be solved, which results in very high computational complexity. Moreover, as the signals of all users are superposed on one resource block, the algorithm may suffer large process delay and error propagation of SIC for the system with a large number of users. In [9], the sum rate optimization problem with minimum user rate constraint was investigated, and therefore a low-complexity BF scheme and a user clustering scheme were proposed. Since the presented BF scheme and user clustering scheme were designed separately, some sum rate performance loss is suffered. In [10], the problem for maximization of the number of users with an ergodic user rate constraint was considered. A power allocation scheme to satisfy ergodic user rate constraints was proposed and then a user admission algorithm that achieves the maximum number of users was developed to guarantee the minimum user rate requirements. However, as the MRT beamforming is adopted for downlink transmission, the optimization of BF vector is not considered. By the duality between the multiple access channel (MAC) and broadcast channel (BC), the duality scheme for sum rate optimization was developed in [11], which also suffers rather high computational complexity because of needing to solve a quadratic constrained quadratic programs convex problem. Furthermore, since the duality scheme is a quasi-degraded solution to the problem of sum rate optimization actually, it is only feasible in the case, where the channel state information (CSI) of the scheduled users meets the quasi-degraded property. As a result, the application of the duality scheme in practical systems may be heavily restricted. In [12], the robust BF design problem to optimize the worst-case achievable sum rate constrained by the total transmit power was studied. In [13], the optimal BF design problem which minimizes the total transmission power subject to a pair of target interference-level constraints was investigated for two-user MISO-NOMA downlink. Based on the results obtained by Chen et al. [13], Chen et al. [14] further proved that the minimum transmit power of the NOMA transmission scheme was equal to that of dirty-paper coding in two-user case, under the condition of the broadcast channel being quasi-degraded. Furthermore, a hybrid NOMA (H-NOMA) precoding algorithm with low complexity is proposed by combining NOMA with zero-forcing beamforming (ZFBB). The summary of comparison of this paper with the existing works [8–14] is shown in Table 1.

Dynamic user clustering and user pairing are effective methods to achieve the viable benefits of NOMA in providing high spectral efficiency [15–17]. In [15], the impact of

user pairing on the performance of the two-user NOMA systems is characterized. It was shown that the performance gain of NOMA over conventional multiple access can be further enlarged by selecting users whose channel conditions are more distinctive. In [16], the problem of jointly optimizing user association and power control to maximize the overall spectral efficiency was investigated for NOMA-based full-duplex MISO systems. To be spectrally efficient, the tensor model was introduced to optimize uplink users' decoding order and downlink users' clustering. In [17], a novel hybrid user pairing beamforming (HUP) scheme was proposed to maximize the achievable spectral and energy efficiencies for multiuser MISO-NOMA downlink systems with simultaneous wireless information and power transfer (SWIPT). In the proposed HUP scheme, two information users with distinct channel conditions are paired while energy users located near the BS perform energy harvesting.

Due to the equivalence between the maximization of the weighted sum rate (WSR) and the acquisition of the maximum WSR point on achievable rate region of the MISO downlink, the problems of characterization of the achievable rate region for MISO broadcast channel [18, 19] and for MISO interference channel [20–22] were investigated, respectively. Specifically, in [18], the set of BF vectors which achieve points on the boundary of achievable rate region of two-user MISO broadcast channel is characterized by a single real valued parameter per user. In [19], the design of adaptive transmission mode switching to derive the larger rate region for the two-user MISO broadcast channel is investigated. An explicit characterization of the boundary of achievable rate region for multiuser MISO interference channel was obtained in [20]. A general framework for finding the maximum sum rate operating points on the boundary of the achievable rate region for the two-user MISO interference channel was proposed in [21]. In [22], the achievable rate region of two-user MISO interference channel for single user detection was obtained.

Although the aforementioned schemes can provide efficient solutions with several advantages, they bring few insights about their optimality, compared to the achievable rate regions for downlink MISO systems with transmission mode switching. In this paper, we compare the achievable rate regions obtained by NOMA-MRT and MMSE-BF. Then, we find that when the channel vectors of the two users are greatly correlated, the achievable rate region of NOMA-MRT includes that of MMSE-BF completely. However, when the two channels are almost orthogonal, the opposite is true. Consequently, we focus on the design of adaptive transmission mode switching (TMS) between NOMA-MRT and MMSE-BF to maximize the sum rate for multiuser MISO downlink by utilizing the idea of user pairing.

The main contributions of this paper are listed as follows:

- (1) By combining MRT beamforming and NOMA, a novel transmission scheme referred to as NOMA-MRT for MISO downlink is proposed and the corresponding closed-form expression of the boundary of the achievable rate region is achieved. Moreover,

TABLE 1: Comparison of this paper with the existing works [8–14].

Research works	[8]	[9]	[10]	[11]	[12]	[13]	[14]	This paper
Scenario	MU MISO	MU MISO	MU MISO	TU MIMO	MU MISO	TU MISO	MU MISO	MU MISO
Main objectives	Sum rate	Sum rate	Number of users	Sum rate	WC sum rate	Transmit power	Transmit power	Sum rate
Optimization variable	BV	BV + CM	Power	Precoding matrix	BV + power	BV	BV + CM	BV + UPM + CP
Joint optimization	–	No	–	–	No	–	Yes	Yes
User number in a cluster	MU	MU	MU	TU	MU	TU	TU	TU
MAC-BC duality	Not used	Not used	Not used	Used	Not used	Not used	Not used	Used
Each BF vector serves single user	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Complexity	High	Medium	Medium	High	Medium	Medium	Low	Low
Need for iteration	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Main challenges	PD and EP	PD and EP	PD and EP	Only feasible in QDC	PD and EP	Spectral efficiency	Spectral efficiency	Spectral efficiency

“TU” and “MU” represent two users and multiple users, respectively. “BV” and “WC sum rate” represent BF vector and worst-case sum rate, respectively. “UPM” and “CM” represent user pairing relationship matrix and clustering relationship matrix, respectively. “CP” and “QDC” represent control parameter for TMS and quasi-degraded channel, respectively. “PD” and “EP” represent processing delay and error propagation of SIC, respectively.

building on the duality of MAC and BC, the closed-form expression for the boundary of the achievable rate region of MISO broadcast channel with MMSE-BF adopted at BS is given. Subsequently, by obtaining the outer boundary of the union of the achievable rate regions of the two transmission schemes, an adaptive transmission mode switching method is developed to achieve a larger rate region for the two-user case.

- (2) Based on the idea that the solution to the problem of WSR optimization must be on the corresponding boundary of the achievable rate region and the closed-form expression of the rate region boundary in any channel condition, the global maximum value of the transformed WSR (the unary function) can be found by selecting the maximum among the function values at the end of the interval and the local maximum values of WSR. Consequently, two optimal power allocation algorithms, i.e., power allocation algorithm for NOMA-MRT mode (NOMA-MRT-PA) and power allocation algorithm for MMSE-BF mode (MMSE-BF-PA), are proposed by focusing on the case with two users.
- (3) Building on the NOMA-MRT-PA and MMSE-BF-PA, a novel joint user pairing and power allocation algorithm (JUPA) is proposed for the multiuser case. Subsequently, by the combination of JUPA and the idea of TMS between NOMA-MRT and MMSE-BF, a practical transmission method is developed. By making use of the closed-form expression of the optimal solution by NOMA-MRT-PA (MMSE-BF-PA), JUPA can be performed with a low computational complexity, while the exhaustive search-based user pairing and power allocation algorithm (ES-PPA) requires extremely high complexity, compared with which JUPA only suffers a slight performance loss.

The remainder of this paper is organized as follows. Section 2 briefly describes the system model and introduces the NOMA-based beamforming scheme. In Section 3, the closed-form expressions of the achievable rate region boundary of MISO broadcast channel by using NOMA-MRT and MMSE-BF are obtained. Consequently, an adaptive switching method is proposed in Section 4. In Section 5, two optimal power allocation algorithms for NOMA-MRT mode and MMSE-BF mode are presented, by which a joint user pairing and power allocation algorithm is also developed, respectively. The numerical results are illustrated in Section 6, and finally conclusions are drawn in Section 7.

Before proceeding, we introduce the following notation. Throughout the paper, we denote column vectors \mathbf{x} and matrices \mathbf{X} by bold lower-case and upper-case letters, respectively. $(\cdot)^H$ represents the complex conjugate transpose of a vector or matrix. The absolute value of a scalar is denoted by $|\cdot|$ and the norm of a vector is denoted by $\|\cdot\|$. $\mathbf{x} \in \mathbb{C}^{M \times 1}$ means that \mathbf{x} is an $M \times 1$ complex vector. $\mathcal{E}\mathcal{N}(\mu, \sigma^2)$ denotes a complex Gaussian random variable with mean μ and variance σ^2 . $\langle \mathbf{x}, \mathbf{y} \rangle$ and $\angle(\mathbf{x}, \mathbf{y})$ denote the inner product and the angle of two complex vectors \mathbf{x} and \mathbf{y} , respectively.

2. Problem Description

2.1. System Model. We consider a downlink communication system with one M -antenna base station (BS) and $2K$ (assumed to be an even number) single-antenna users, from which K user pairs are selected and the two users in a user pair share the same spectrum. The BS serves one user pair in each time-frequency resource block, and the data transmission for the user pair can adaptively switch between two transmission modes, i.e., NOMA-MRT and MMSE-BF, according to the channel state, as shown in Figure 1. When the channel vectors of the two

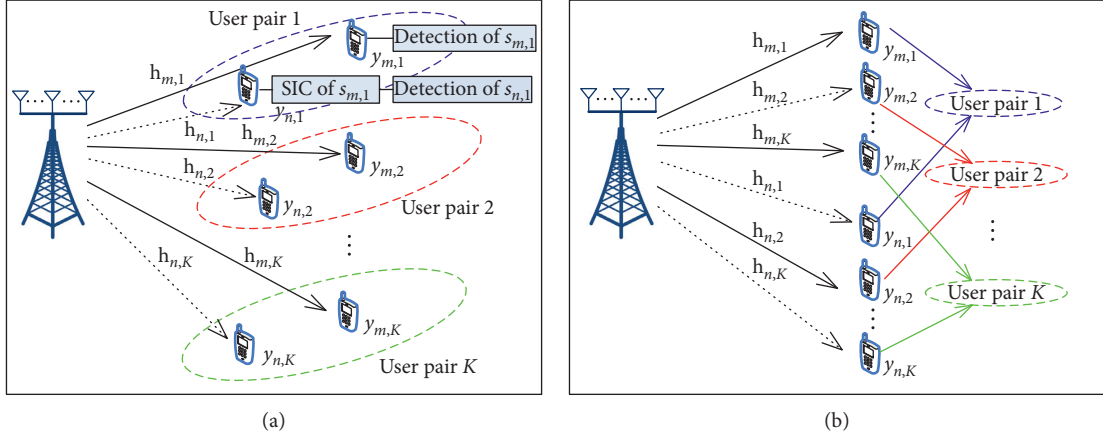


FIGURE 1: System model with transmission mode switching between NOMA-MRT and MMSE-BF. (a) NOMA-MRT mode. (b) MMSE-BF mode.

users in a user pair are greatly correlated, the NOMA-MRT mode is usually selected to achieve larger sum rate in the system. However, when the two channels are almost orthogonal, the MMSE-BF mode is selected normally (see Section 6.1 for details). In the NOMA-MRT mode, MRT beamforming is performed at the BS and SIC is executed at the user with strong channel conditions in the user pair. Specifically, for example, user n with strong channel conditions in user pair 1 first decodes the signal $s_{m,1}$ of user m in the user pair 1 and subtracts this from its received signal $y_{n,1}$. As a result, user n can decode its own signal $s_{n,1}$ without the interference caused by user m . However, user m in user pair 1 simply treats $s_{n,1}$ as noise and decodes its own signal $s_{m,1}$. In the MMSE-BF mode, MMSE beamforming is adopted for the user pair at the BS. $2K$ users are assumed to be uniformly located within a cell with a radius of R , and the BS is deployed at the center of the cell. Here, we consider user pairing, i.e., selecting two users to form a group, in the system model for the following reasons. In the NOMA downlink, users cancel the co-channel interference by performing SIC. However, with the number of users in a group growing, the processing complexity and delay at users dramatically increase [23]. As a result, we consider only grouping two users and adopting user pairing scheme to decrease processing complexity and delay in the NOMA downlink. The observation at user i in the user pair j is given by

$$y_{i,j} = \mathbf{h}_{i,j}^H \mathbf{x} + z_{i,j}, \quad i = m, n; j = 1, 2, \dots, K, \quad (1)$$

where $\mathbf{h}_{i,j} \sim \mathcal{CN}(0, \sigma_{i,j}^2 \mathbf{I}_M)$ is the channel vector from BS to user i in the user pair j , $z_{i,j} \sim \mathcal{CN}(0, N_0)$ is the additive white Gaussian noise (AWGN) at user i in the user pair j , and N_0 is the corresponding noise power. For the channel vector $\mathbf{h}_{i,j}$, the variance of the channel from BS to user i in the user pair j is modeled as [14, 24]

$$\sigma_{i,j}^2 = \begin{cases} d_{i,j}^{-\alpha'}, & \text{if } d_{i,j} > d_0, \\ d_0^{-\alpha'}, & \text{otherwise,} \end{cases} \quad (2)$$

where $d_{i,j}$ denotes the distance between BS and user i in the user pair j , α' denotes the path loss exponent, and the

parameter d_0 avoids the singularity when $d_{i,j}$ is small. Furthermore, $\mathbf{x}_j = \sqrt{p_{m,j}} \mathbf{w}_{m,j} s_{m,j} + \sqrt{p_{n,j}} \mathbf{w}_{n,j} s_{n,j}$ is the signal transmitted by the BS for the user pair j , where $s_{i,j}$ and $\mathbf{w}_{i,j}$ are the scalar signal and normalized BF vector for user i in the user pair j , respectively. $p_{i,j}$ is the transmit power allocated to user i ($i = n, m$) in the user pair j . Assume that the BS can obtain the perfect CSI.

For convenience, we omit the subscript j of the variables mentioned above in the following. In this paper, we attempt to maximize the sum rate of users by jointly optimizing the pairing relationship matrix \mathbf{U} as well as power allocation vector $\mathbf{p}_{m,n}$ and control parameter $T_{m,n}$ for TMS of each user pair. Mathematically, the optimization problem can be formulated as

$$(P1): \max_{\mathbf{p}_{m,n}, \mathbf{U}, T_{m,n}} \sum_{m=1}^{2K} \sum_{n=m+1}^{2K} u_{m,n} (R_m + R_n), \quad (3a)$$

$$\text{s.t. } T_{m,n} \in \{0, 1\}, \quad 1 \leq m, n \leq 2K, \quad (3b)$$

$$u_{m,n} \in \{0, 1\}, \quad 1 \leq m, n \leq 2K, \quad (3c)$$

$$u_{m,n} = u_{n,m}, \quad 1 \leq m, n \leq 2K, \quad (3d)$$

$$u_{m,m} = 0, \quad 1 \leq m, n \leq 2K, \quad (3e)$$

$$\sum_{m=1}^{2K} u_{m,n} = 1, \quad 1 \leq n \leq 2K, \quad (3f)$$

$$\sum_{n=1}^{2K} u_{m,n} = 1, \quad 1 \leq m \leq 2K, \quad (3g)$$

$$0 \leq u_{m,n} (p_m + p_n) \leq P, \quad 1 \leq m, n \leq 2K, \quad (3h)$$

$$\|\mathbf{w}_m\| = \|\mathbf{w}_n\| = 1, \quad 1 \leq m, n \leq 2K, \quad (3i)$$

where R_i is the data rate of user i in the user pair $\{m, n\}$, $i = m, n$. \mathbf{U} is the user pairing relationship matrix with $2K$ dimensions, in which the m -th row and n -th column element

is denoted by $u_{m,n}$. $\mathbf{p}_{m,n} = [p_m, p_n]$ and $T_{m,n}$ are power allocation vector and control parameter for transmission mode switching of the user pair $\{m, n\}$. As shown in (3b), $T_{m,n} = 1$ denotes that the transmission mode NOMA-MRT is selected for the user pair $\{m, n\}$ at BS. Otherwise, MMSE-BF is performed for the user pair $\{m, n\}$. In (3c), $u_{m,n} = 1$ represents that users m and n are paired together and constitute a user pair. Otherwise, users m and n are not paired together. (3d) and (3e) imply that U is a symmetric matrix, and the diagonal elements are zero as a user cannot be paired with itself. (3f) and (3g) indicate that a user can be paired with only one user. Furthermore, (3h) represents the total power allocated to a user pair $\{m, n\}$, which is upper bounded to P . (3i) denotes that the BF vectors are normalized in this paper.

It is hard to achieve the optimal solution to the problem (P1) due to binary constraints of $u_{m,n}$ and the nonconvex property of the achievable rate of user i , which is the data rate of user i with the equality holding in (4) or (7). Therefore, for the treatability of the problem (P1), we select the BF scheme (i.e., control parameter for TMS) with the larger sum rate for a user pair, among MMSE-BF and NOMA-based BF, instead of jointly optimizing BF vector for all users, in the process of jointly optimizing the pairing relationship matrix \mathbf{U} , power allocation vector $\mathbf{p}_{m,n}$, and control parameter $T_{m,n}$ of each user pair.

2.2. Achievable Rate Region of the Existing BF Schemes. For conventional linear BF scheme, such as ZFBF, MRT, and MMSE-BF, given a fixed normalized BF vector \mathbf{w}_i , the achievable rate region of the MISO broadcast channel (BC) is given by the set of rate tuples (R_m, R_n) satisfying

$$R_i \leq \log(1 + \gamma_i^B), \quad (4)$$

where

$$\gamma_i^B = \frac{p_i |\mathbf{h}_i^H \mathbf{w}_i|^2}{N_0 + p_{\varphi(i)} |\mathbf{h}_i^H \mathbf{w}_{\varphi(i)}|^2}, \quad (5)$$

is the signal-to-interference-plus-noise ratio (SINR) of user i with $p_m + p_n \leq P$, $i = m, n$. The mapping function $\varphi(i)$ is defined by

$$\varphi(i) = \begin{cases} n, & i = m, \\ m, & i = n, \end{cases} \quad (6)$$

and P is the total transmit power for each user pair. When the equality holds in (4), the rate tuple (R_m, R_n) achieves on the boundary of the achievable rate region with $p_m + p_n = P$.

2.3. Achievable Rate Region of NOMA-Based BF Scheme. For two-user MISO downlink with the NOMA-based linear BF scheme adopted at the BS, SIC is implemented at the user with strong channel conditions. Therefore, for a fixed normalized BF vector \mathbf{w}_i , when SIC is carried out at user n with strong channel conditions, the achievable rate region of

the MISO broadcast channel by using NOMA-based linear BF at the BS can be formulated by the set of rate tuples satisfying

$$\begin{cases} R_n \leq \log\left(1 + \frac{p_n |\mathbf{h}_n^H \mathbf{w}_n|^2}{N_0}\right), \\ R_m \leq \min\{\log(1 + \gamma_{m,n}^B), \log(1 + \gamma_m^B)\}, \end{cases} \quad (7)$$

where $\log(1 + \gamma_{m,n}^B)$ denotes the achievable rate for user n to detect user m 's message and $\gamma_{m,n}^B$ is the corresponding SINR, i.e.,

$$\gamma_{m,n}^B = \frac{p_m |\mathbf{h}_n^H \mathbf{w}_m|^2}{N_0 + p_n |\mathbf{h}_n^H \mathbf{w}_m|^2}. \quad (8)$$

In this paper, the angle between \mathbf{h}_m and \mathbf{h}_n is denoted by α , where

$$\cos^2(\alpha) = \frac{|\mathbf{h}_n^H \mathbf{h}_m|^2}{\|\mathbf{h}_m\|^2 \|\mathbf{h}_n\|^2}. \quad (9)$$

Without loss of generality, we set $\alpha \in [0, \pi)$.

2.4. The Equivalence between Maximizing WSR and Enlarging Achievable Rate Region. For a given weight vector, the solution of optimization problem of maximizing WSR must be on the boundary of achievable rate region of the two-user MISO broadcast channel [19]. Thus, maximizing the WSR in MISO downlink systems is equivalent to enlarging the corresponding achievable rate region as much as possible.

In this paper, from the perspective of achievable rate region, we investigate the design of the suboptimal algorithm for the solution to the problem (P1) based on transmission mode switching between NOMA-based BF and MMSE-BF.

3. Analysis on Rate Regions of NOMA-MRT and MMSE-BF

3.1. Achievable Rate Region of NOMA-MRT. Geometrically, the BF vector \mathbf{w}_i of user i with MRT is aligned with the spatial direction of \mathbf{h}_i to maximize the length of the projection of \mathbf{w}_i onto \mathbf{h}_i . Therefore, the maximum of the signal-to-noise ratio for user i can be achieved by using MRT. The normalized BF vector of user i by using MRT is given by

$$\mathbf{w}_i^{\text{MRT}} = \frac{\mathbf{h}_i}{\|\mathbf{h}_i\|}, \quad i = m, n. \quad (10)$$

NOMA-MRT is the combination of NOMA and MRT, i.e., performing MRT beamforming at the transmitter and executing SIC at the receiver with strong channel conditions for two-receiver case.

Lemma 1. When $\|\mathbf{h}_m\| < \|\mathbf{h}_n\|$, the achievable rate region boundary of NOMA-MRT for MISO broadcast channel can be expressed as follows:

$$r_n = \log\left(1 + \tilde{\rho}_n \|\mathbf{h}_n\|^2\right),$$

$$r_m = \begin{cases} \log\left(1 + \frac{\tilde{\rho}_m \|\mathbf{h}_n\|^2 (1 - \theta)}{1 + \tilde{\rho}_n \|\mathbf{h}_n\|^2}\right), & \zeta_{n,m} \leq 0 \text{ or } \zeta_{n,m} < \tilde{\rho}_n \leq \rho, \\ \log\left(1 + \frac{\tilde{\rho}_m \|\mathbf{h}_m\|^2}{1 + \tilde{\rho}_n \|\mathbf{h}_m\|^2 (1 - \theta)}\right), & \zeta_{n,m} \geq \rho \text{ or } 0 \leq \tilde{\rho}_n \leq \zeta_{n,m}, \end{cases} \quad (11)$$

where

$$\zeta_{n,m} = \frac{\|\mathbf{h}_n\|^2 (1 - \theta) - \|\mathbf{h}_m\|^2}{\|\mathbf{h}_m\|^2 \|\mathbf{h}_n\|^2 (2\theta - \theta^2)}, \quad (12)$$

where $\theta = \sin^2 \alpha$, $\rho = P/N_0$, $\tilde{\rho}_m = \rho - \tilde{\rho}_n$, and $\tilde{\rho}_i = p_i/N_0$, $i = m, n$.

Case (2). $\theta = 0$.

$$\begin{cases} r_n = \log\left(1 + \tilde{\rho}_n \|\mathbf{h}_n\|^2\right), \\ r_m = \log\left(1 + \frac{\tilde{\rho}_m \|\mathbf{h}_m\|^2}{1 + \tilde{\rho}_n \|\mathbf{h}_m\|^2}\right). \end{cases} \quad (13)$$

Proof. See the Appendix. \square

Lemma 2. When $\|\mathbf{h}_m\| = \|\mathbf{h}_n\| = l$, the achievable rate region boundary of NOMA-MRT for MISO broadcast channel can be expressed as

$$\begin{cases} r_n = \log\left(1 + \frac{P_n l^2}{N_0}\right), \\ r_m = \log\left(1 + \frac{P_m l^2}{N_0 + P_n l^2}\right). \end{cases} \quad (14)$$

Proof. See the Appendix. \square

3.2. Achievable Rate Region of MMSE-BF. The MMSE-BF can optimally trade off fighting interference to other users and the background Gaussian noise, i.e., the MMSE-BF can maximize the output SINR for any value of signal-to-noise ratio (SNR) [25]. Such a beamformer looks like the zero-forcing beamformer when the interuser interference is large and like the MRT beamformer when the interference is small. The geometric description of normalized BF vectors for ZFBF, MRT, and MMSE-BF is shown in Figure 2.

Case (1). $\theta > 0$.

By MAC-BC duality theory [25], the MMSE beamformer in BC is exactly the MMSE receiver filters in dual MAC. Therefore, the sets of achievable SINRs are the same in both cases with the same total transmit power constraint. Consequently, we have the following lemma.

Lemma 3. By the duality of MAC and BC, the achievable rate region boundary of BC by using MMSE-BF can be written in terms of that of the dual MAC with MMSE receiver filter. The set of rate tuple $(\tilde{r}_m, \tilde{r}_n)$ on the achievable rate region boundary of BC satisfies

$$\tilde{r}_i = \log\left(1 + \gamma_i^M\right), \quad i = m, n, \quad (15)$$

where

$$\gamma_i^M = \frac{\rho_i \|\mathbf{h}_i\|^2 \left(1 + \rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2 \theta\right)^2}{\varsigma_i + \rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2 (1 - \theta)}, \quad (16)$$

where $\varsigma_i = \rho_{\varphi(i)}^2 \|\mathbf{h}_{\varphi(i)}\|^4 \theta + 2\rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2 \theta + 1$, $\rho_i = q_i/N_0$, and q_i is the transmit power of user i in the dual MAC, satisfying $q_i + q_{\varphi(i)} = P$.

Proof. According to the duality between MAC and BC [25], the normalized BF vectors of user i with MMSE-BF in MISO broadcast channel are given by

$$\mathbf{w}_i^{\text{MMSE}} = \frac{\left(N_0 \mathbf{I} + q_{\varphi(i)} \mathbf{h}_{\varphi(i)} \mathbf{h}_{\varphi(i)}^H\right)^{-1} \mathbf{h}_i}{\left\| \left(N_0 \mathbf{I} + q_{\varphi(i)} \mathbf{h}_{\varphi(i)} \mathbf{h}_{\varphi(i)}^H\right)^{-1} \mathbf{h}_i \right\|}. \quad (17)$$

By the matrix inversion lemma [26], equation (17) can be written as

$$\mathbf{w}_i^{\text{MMSE}} = \frac{\left(\mathbf{I} - \left(\rho_{\varphi(i)} \mathbf{h}_{\varphi(i)} \mathbf{h}_{\varphi(i)}^H\right) / \left(1 + \rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2\right)\right) \mathbf{h}_i}{\left\| \left(\mathbf{I} - \left(\rho_{\varphi(i)} \mathbf{h}_{\varphi(i)} \mathbf{h}_{\varphi(i)}^H\right) / \left(1 + \rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2\right)\right) \mathbf{h}_i \right\|}. \quad (18)$$

The SINR γ_i^M of user i in dual MAC with MMSE receiver filter has the following form:

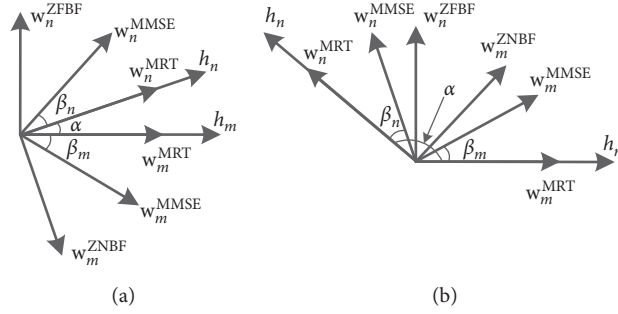


FIGURE 2: The geometric description of BF vectors for ZFBF, MRT, and MMSE-BF. For ZFBF scheme, $\mathbf{w}_i^{\text{ZFBF}}$ is orthogonal to $\mathbf{h}_{\varphi(i)}$, $i = m, n$. (a) $0 \leq \alpha \leq \pi/2$. (b) $\pi/2 \leq \alpha \leq \pi$.

$$\gamma_i^M = \frac{q_i \|\mathbf{h}_i^H \mathbf{w}_i^{\text{MMSE}}\|^2}{N_0 + q_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}^H \mathbf{w}_i^{\text{MMSE}}\|^2}. \quad (19)$$

According to the definition of inner product, equation (19) can be rewritten as

$$\gamma_i^M = \frac{q_i \|\mathbf{h}_i\|^2 \cos^2 \beta_i}{N_0 + q_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2 \cos^2 (\alpha + \beta_i)}, \quad \alpha \in [0, (\pi/2)), \quad (20)$$

$$\gamma_i^M = \frac{q_i \|\mathbf{h}_i\|^2 \cos^2 \beta_i}{N_0 + q_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2 \cos^2 (\alpha - \beta_i)}, \quad \alpha \in [(\pi/2), \pi), \quad (21)$$

where β_i represents the angle between $\mathbf{w}_i^{\text{MMSE}}$ and \mathbf{h}_i , $\alpha + \beta_i$ is the angle between $\mathbf{w}_i^{\text{MMSE}}$ and $\mathbf{h}_{\varphi(i)}$ when $\alpha \in [0, (\pi/2))$, and $\alpha - \beta_i$ is that when $\alpha \in [(\pi/2), \pi)$, $i = n, m$, as illustrated in Figure 2.

The square of the cosine value of β_i can be expressed as

$$\cos^2 \beta_i = \frac{|\langle \mathbf{w}_i^{\text{MMSE}}, \mathbf{h}_i \rangle|^2}{\|\mathbf{h}_i\|^2}. \quad (22)$$

By substituting (18) into (22), we can derive

$$\cos^2 \beta_i = \frac{\|\mathbf{h}_i\|^2 (1 + \rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2 \theta)^2}{\|\mathbf{s}_i\|^2 (1 + \rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2)^2}, \quad (23)$$

where

$$\mathbf{s}_i = \left(\mathbf{I} - \frac{\rho_{\varphi(i)} \mathbf{h}_{\varphi(i)} \mathbf{h}_{\varphi(i)}^H}{1 + \rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2} \right) \mathbf{h}_i. \quad (24)$$

According to the definition of inner product, \mathbf{s}_i can be written as

$$\mathbf{s}_i = \mathbf{h}_i - \frac{\rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2}{1 + \rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2} \frac{\langle \mathbf{h}_i, \mathbf{h}_{\varphi(i)} \rangle \mathbf{h}_{\varphi(i)}}{\|\mathbf{h}_{\varphi(i)}\|^2} \quad (25)$$

$$= \mathbf{h}_i - \frac{\rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2}{1 + \rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2} \prod_{\mathbf{h}_{\varphi(i)}} \mathbf{h}_i,$$

where $\prod_{\mathbf{h}_{\varphi(i)}} \mathbf{h}_i$ represents the orthogonal projection of \mathbf{h}_i onto $\mathbf{h}_{\varphi(i)}$. The norm of \mathbf{s}_i is given by

$$\|\mathbf{s}_i\| = \|\mathbf{h}_i\| \sqrt{1 + (1 - \theta)(g_i^2 - 2g_i)}, \quad (26)$$

where $g_i = \rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2 / (1 + \rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2)$. By substituting (26) into (23), we can obtain

$$\cos^2 \beta_i = \frac{(1 + \rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2 \theta)^2}{(1 + \rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2)^2 (1 + (1 - \theta)(g_i^2 - 2g_i))}. \quad (27)$$

When α belongs to the interval $[0, (\pi/2))$, the angle between $\mathbf{w}_i^{\text{MMSE}}$ and $\mathbf{h}_{\varphi(i)}$ is $\alpha + \beta_i$ and the square of the cosine value of $\alpha + \beta_i$ can be expressed as

$$\cos^2 (\alpha + \beta_i) = (1 - \theta) \cos^2 \beta_i + \theta \sin^2 \beta_i - 2\sqrt{\theta(1 - \theta)} \sin \beta_i \cos \beta_i. \quad (28)$$

Then, we consider $\beta_i \in [0, (\pi/2))$. In this case, by substituting (27) into (28), we can derive

$$\begin{aligned} \cos^2 (\alpha + \beta_i) &= \frac{1 - \theta}{\rho_{\varphi(i)}^2 \|\mathbf{h}_{\varphi(i)}\|^4 \theta + 2\rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2 \theta + 1} \\ &= \frac{1 - \theta}{\varsigma_i}. \end{aligned} \quad (29)$$

When α belongs to the interval $[(\pi/2), \pi)$, the angle between $\mathbf{w}_i^{\text{MMSE}}$ and $\mathbf{h}_{\varphi(i)}$ is $\alpha - \beta_i$. Similarly, we have

$$\cos^2 (\alpha - \beta_i) = \frac{1 - \theta}{\varsigma_i}. \quad (30)$$

From (29) and (30), we can find that the expression of $\cos^2(\alpha + \beta_i)$ with $\alpha \in [0, (\pi/2))$ and that of $\cos^2(\alpha - \beta_i)$ with $\alpha \in [(\pi/2), \pi)$ are the same. Therefore, equations (20) and (21) can be combined into one equation. We substitute (27) into (20) and (21), and then (29) and (30) into (20) and (21), respectively. Subsequently, by combining (20) and (21), we can obtain (15). \square

4. Achievable Rate Region of the Adaptive Switching Method

In this section, we first derive the intersection points of achievable rate region boundaries of NOMA-MRT and MMSE-BF. Based on the intersection points, we proposed the adaptive switching method for the case with two users, which obtains the larger achievable rate region than that derived by employing NOMA-MRT or MMSE-BF only.

4.1. The Intersection Points of NOMA-MRT's and MMSE-BF's Rate Region Boundaries. According to Lemmas 1–3, we can compare the achievable rate regions obtained by NOMA-MRT and MMSE-BF. Then, we can find that neither MMSE-BF nor NOMA-MRT is optimal in all channel states. Consequently, an adaptive switching method is preferred, which can achieve a larger rate region than that derived by MMSE-BF or NOMA-MRT only. Combined with the concept of time-sharing [25], the adaptive switching method can achieve a convex hull of the union of MMSE-BF and NOMA-MRT's achievable rate region.

Definition 1. If $r_m(\tilde{\rho}_{n,k}) = \tilde{r}_m(\rho_{n,k})$ and $r_n(\tilde{\rho}_{n,k}) = \tilde{r}_n(\rho_{n,k})$, $\tilde{\rho}_{n,k} \in [0, \rho]$, $\rho_{n,k} \in [0, \rho]$, we call that the rate region boundaries of NOMA-MRT and MMSE-BF have intersection points at $\rho_{n,k}$, $k = 1, 2, \dots, \kappa$, where κ is the number of the intersection points.

The intersection points of NOMA-MRT and MMSE-BF's rate region boundaries are given as follows.

Case (1). When $\|\mathbf{h}_m\| < \|\mathbf{h}_n\|$.

By combining (11) with (15), we obtain two equation sets

$$\tilde{\rho}_n \|\mathbf{h}_n\|^2 = \frac{\rho_n \|\mathbf{h}_n\|^2 \left(1 + \rho_m \|\mathbf{h}_m\|^2 \theta\right)^2}{\varsigma_n + \rho_m \|\mathbf{h}_m\|^2 (1 - \theta)}, \quad (31)$$

$$\frac{\tilde{\rho}_m \|\mathbf{h}_m\|^2 (1 - \theta)}{1 + \tilde{\rho}_n \|\mathbf{h}_n\|^2} = \frac{\rho_m \|\mathbf{h}_m\|^2 \left(1 + \rho_n \|\mathbf{h}_n\|^2 \theta\right)^2}{\varsigma_m + \rho_n \|\mathbf{h}_n\|^2 (1 - \theta)}; \quad (32)$$

$$\tilde{\rho}_n \|\mathbf{h}_n\|^2 = \frac{\rho_n \|\mathbf{h}_n\|^2 \left(1 + \rho_m \|\mathbf{h}_m\|^2 \theta\right)^2}{\varsigma_n + \rho_m \|\mathbf{h}_m\|^2 (1 - \theta)}, \quad (33)$$

$$\frac{\tilde{\rho}_m \|\mathbf{h}_m\|^2}{1 + \tilde{\rho}_n \|\mathbf{h}_n\|^2 (1 - \theta)} = \frac{\rho_m \|\mathbf{h}_m\|^2 \left(1 + \rho_n \|\mathbf{h}_n\|^2 \theta\right)^2}{\varsigma_m + \rho_n \|\mathbf{h}_n\|^2 (1 - \theta)}. \quad (34)$$

By substituting (31) and (33) into (32) and (34), respectively, the intersection points can be derived by solving the following two equations:

$$c_{1,6} \rho_n^6 + c_{1,5} \rho_n^5 + c_{1,4} \rho_n^4 + c_{1,3} \rho_n^3 + c_{1,2} \rho_n^2 + c_{1,1} \rho_n + c_{1,0} = 0, \quad (35)$$

where $c_{1,6} = -\theta^4 l_{n,m}^3 l_m^3$, $c_{1,5} = \theta^3 l_{n,m}^2 l_m^2 (3l_n - 3l_m - \theta l_n + 3\rho \theta l_n l_m)$, $c_{1,4} = -\theta^2 l_{n,m} l_m (3\rho^2 \theta^2 l_{n,m}^2 l_m^2 - 2\rho \theta^2 l_{n,m}^2 l_m + 5\rho \theta l_{n,m}^2 l_m - 9\rho \theta l_{n,m}^2 l_m + \rho l_{n,m}^2 l_m + \theta^2 l_{n,m} l_m - 2\theta l_{n,m}^2 - \theta l_{n,m} l_m + 3l_n^2 - 6l_n l_m + 3l_m^2)$, $c_{1,3} = -\theta(-\rho^3 \theta^3 l_{n,m}^3 l_m^3 + \rho^2 \theta^3 l_{n,m}^3 l_m^2 - \rho^2 \theta^2 l_{n,m}^3 l_m^2 + 9\rho^2 \theta^2 l_{n,m}^3 l_m^2 - 2\rho^2 \theta l_{n,m}^3 l_m^2 - 2\rho \theta^3 l_{n,m}^2 l_m^2 + 3\rho \theta^2 l_{n,m}^2 l_m^2 + \rho \theta^2 l_{n,m}^2 l_m^2 - 3\rho \theta l_{n,m}^2 l_m^2 + 12\rho \theta l_{n,m}^2 l_m^2 - 9\rho \theta l_{n,m} l_m^3 - \rho l_{n,m}^2 l_m + \rho l_{n,m}^2 l_m^2 - 2\theta^2 l_{n,m}^2 l_m + \theta^2 l_{n,m}^2 l_m + \theta l_n^3 + \theta l_n^2 l_m - 3\theta l_n l_m^2 - l_n^3 + 4l_n^2 l_m - 4l_n l_m^2 + l_m^3)$, $c_{1,2} = -\rho^2 \theta^3 l_{n,m}^2 l_m^2 - \rho^2 \theta^2 l_{n,m}^2 l_m^2 - 3\rho^2 \theta^2 l_{n,m}^2 l_m^2 + 3\rho^2 \theta^2 l_{n,m}^2 l_m^2 + \rho^2 \theta l_{n,m}^2 l_m^2 - \rho^2 \theta l_{n,m}^2 l_m^2 + \rho \theta^3 l_{n,m}^2 l_m^2 - \rho \theta^2 l_{n,m}^2 l_m^2 + \rho \theta^2 l_{n,m}^2 l_m^2 + \rho \theta l_{n,m}^2 l_m^2 - 2\rho \theta l_{n,m}^2 l_m^2 - \rho \theta l_{n,m}^2 l_m^2 + 6\rho \theta l_{n,m}^2 l_m^2 + 2\rho \theta l_{n,m}^2 l_m^2 - 2\rho \theta l_{n,m}^2 l_m^2 - \rho l_{n,m}^2 l_m^2 + \theta^2 l_{n,m}^2 l_m^2 - 2\theta^2 l_{n,m}^2 l_m^2 + 4\theta l_{n,m}^2 l_m^2 - \theta l_{n,m}^2 l_m^2 - l_{n,m}^2 l_m^2 + l_n l_m^2 - l_m^2 l_m^2$, $c_{1,1} = 2\rho \theta l_{n,m} l_m^2 l_m^2 - \rho \theta^2 l_{n,m} l_m^2 l_m^2 + \rho \theta l_{n,m} l_m^2 l_m^2 - \rho \theta l_{n,m} l_m^2 l_m^2 + \rho \theta^2 l_{n,m} l_m^2 l_m^2 + \rho^2 \theta l_{n,m} l_m^2 l_m^2 + 2\rho l_{n,m} l_m^2 l_m^2 - \rho l_{n,m}^2 l_m^2 l_m^2 - \rho \theta l_{n,m}^2 l_m^2 l_m^2 - \theta l_{n,m}^2 l_m^2 l_m^2 - \rho l_{n,m}^2 l_m^2 l_m^2 - l_m^2 l_m^2 l_m^2$, and $c_{1,0} = \rho v_m (\rho l_m + 1) (l_m - l_n + \theta l_n)$, $v_m = \rho \theta l_m + 1$, $v_n = \rho l_m + 1$.

$$c_{2,6} \rho_n^6 + c_{2,5} \rho_n^5 + c_{2,4} \rho_n^4 + c_{2,3} \rho_n^3 + c_{2,2} \rho_n^2 + c_{2,1} \rho_n + c_{2,0} = 0. \quad (36)$$

where $c_{2,6} = -\theta^4 (1 - \theta) l_{n,m}^2 l_m^2$; $c_{2,5} = \theta^3 l_{n,m} l_m^2 (1 - \theta) (2l_n - 2l_m + 3\rho \theta l_n l_m)$; $c_{2,4} = -\theta^2 (1 - \theta) l_m (3\rho^2 \theta^2 l_{n,m}^2 l_m^2 + 4\rho \theta l_{n,m}^2 l_m^2 - 6\rho \theta l_{n,m}^2 l_m^2 + \rho l_{n,m}^2 l_m^2 + 2l_n^2 - 3l_n l_m + l_m^2)$; $c_{2,3} = \theta (1 - \theta) (\rho^3 \theta^3 l_{n,m}^3 l_m^3 + 2\rho^2 \theta^2 l_{n,m}^2 l_m^2 - 6\rho^2 \theta^2 l_{n,m}^2 l_m^2 + 2\rho^2 \theta^2 l_{n,m}^2 l_m^2 + 3\rho \theta l_{n,m}^2 l_m^2 - 6\rho \theta l_{n,m}^2 l_m^2 + 3\rho \theta l_{n,m}^2 l_m^2 + \rho l_{n,m}^2 l_m^2 - \rho l_{n,m}^2 l_m^2 + l_n^2 - 2l_n l_m + l_m^2)$, $c_{2,2} = (1 - \theta) (2\rho^3 \theta^3 l_{n,m}^3 l_m^3 - \rho^3 \theta^2 l_{n,m}^2 l_m^2 - \rho^2 \theta^2 l_{n,m}^2 l_m^2 + 3\rho^2 \theta^2 l_{n,m}^2 l_m^2 - 3\rho^2 \theta^2 l_{n,m}^2 l_m^2 - \rho^2 \theta l_{n,m}^2 l_m^2 + 2\rho \theta l_{n,m}^2 l_m^2 - \rho \theta l_{n,m}^2 l_m^2 + 3\rho \theta l_{n,m}^2 l_m^2 - 2\rho \theta l_{n,m}^2 l_m^2 + \rho l_{n,m}^2 l_m^2 + l_n)$, and $c_{2,1} = \rho (1 - \theta) (\rho \theta l_m + 1) (\rho \theta l_m^2 - \rho l_n l_m - l_n)$, $c_{2,0} = 0$.

According to equation (31) (equation (33)) with $\rho_m = \rho - \rho_n$, we can obtain the result

$$\tilde{\rho}_n = \frac{\rho_n \left(1 + (\rho - \rho_n) \|\mathbf{h}_m\|^2 \theta\right)^2}{\varsigma_n + (\rho - \rho_n) \|\mathbf{h}_m\|^2 (1 - \theta)}. \quad (37)$$

Here $\varsigma_n = \|\mathbf{h}_m\|^4 \theta (\rho - \rho_n)^2 + 2\|\mathbf{h}_m\|^2 \theta (\rho - \rho_n) + 1$. If the rate region boundaries of NOMA-MRT and MMSE-BF have intersection point at $\rho_{n,k}$, $k = 1, 2, \dots, \kappa$, then $\rho_{n,k}$ must satisfy equation (37)

For equation (35), we consider $\rho_n \in [0, \rho]$, $\tilde{\rho}_n \in [0, \rho]$ when $\zeta_{n,m} \leq 0$, or consider $\rho_n \in [0, \rho]$, $\tilde{\rho}_n \in (\zeta_{n,m}, \rho]$ when $0 < \zeta_{n,m} < \rho$, as shown in (11). In this case, the intersection point $\rho_{n,k}$ ($k \in \{1, 2, \dots, \kappa\}$) must satisfy $\rho_{n,k} \in [0, \rho]$ and $\tilde{\rho}_{n,k} \in [0, \rho]$ when $\zeta_{n,m} \leq 0$ or $\rho_{n,k} \in [0, \rho]$ and $\tilde{\rho}_{n,k} \in (\zeta_{n,m}, \rho]$ when $0 < \zeta_{n,m} < \rho$, where

$$\tilde{\rho}_{n,k} = \frac{\rho_{n,k} \left(1 + (\rho - \rho_{n,k}) \|\mathbf{h}_m\|^2 \theta \right)^2}{\varsigma_{n,k} + (\rho - \rho_{n,k}) \|\mathbf{h}_m\|^2 (1 - \theta)}, \quad (38)$$

$$\varsigma_{n,k} = \|\mathbf{h}_m\|^4 \theta (\rho - \rho_{n,k})^2 + 2 \|\mathbf{h}_m\|^2 \theta (\rho - \rho_{n,k}) + 1. \quad (39)$$

Similarly, for equation (36), the intersection point $\rho_{n,k}$ ($k \in \{1, 2, \dots, \kappa\}$) must satisfy $\rho_{n,k} \in [0, \rho]$, $\tilde{\rho}_{n,k} \in [0, \rho]$ when $\zeta_{n,m} \geq \rho$ or $\rho_{n,k} \in [0, \rho]$, $\tilde{\rho}_{n,k} \in (0, \zeta_{n,m}]$ when $0 < \zeta_{n,m} < \rho$.

Since deriving the analytic solution of (35) and (36) is not a trivial thing, we use numerical method to derive the solutions.

Case (2). When $\|\mathbf{h}_m\| = \|\mathbf{h}_n\| = l$.

By combining (14) with (15), we derive the equation set

$$\tilde{\rho}_n l^2 = \frac{\rho_n l^2 (1 + \rho_m l^2 \theta)^2}{\varsigma_n + \rho_m l^2 (1 - \theta)}, \quad (40)$$

$$\frac{\tilde{\rho}_n l^2}{1 + \tilde{\rho}_m l^2} = \frac{\rho_m l^2 (1 + \rho_n l^2 \theta)^2}{\varsigma_m + \rho_n l^2 (1 - \theta)}. \quad (41)$$

By substituting (40) into (41), the intersection points can be obtained by solving the following equation:

$$\left(\rho_n^2 - \rho \rho_n - \frac{2\rho\theta l^2 + 2\theta - \rho l^2 + 1}{\theta^2 l^4} \right) (\rho_n - \rho) \cdot \left(\rho_n + \frac{1}{\theta l^2} \right) \left(\rho_n - \frac{\rho\theta l^2 + 1}{\theta l^2} \right) \rho_n = 0. \quad (42)$$

By solving equation (42), we can derive the results

$$\begin{aligned} \rho_{n,1} &= 0, \\ \rho_{n,2} &= \rho, \\ \rho_{n,3} &= \frac{1}{\theta l^2}, \\ \rho_{n,4} &= \frac{(\rho\theta l^2 + 1)}{\theta l^2}, \\ \rho_{n,(5,6)} &= \frac{\rho\theta l^2 \pm \sqrt{\rho^2 \theta^2 l^4 + 8\rho\theta l^2 - (4\rho l^2 - 8\theta + 4)}}{2\theta l^2}, \end{aligned} \quad (43)$$

where $\rho_{n,k} \in [0, \rho]$, $k = 1, 2, \dots, 6$.

In the following, we denote by \tilde{p}_i (p_i) the power allocated to user i in the BC with MMSE-BF (NOMA-MRT) scheme, $i = m, n$. Since the transmit power q_i of user i in dual MAC is not equal to the allocated power \tilde{p}_i for user i in the BC, we should obtain the allocated power $\tilde{p}_{i,k}$ of user i to design the switching method in the BC, which corresponds to the intersection point $\rho_{i,k} = q_{i,k}/N_0$. According to the duality between MAC

and BC, the SINR at each point on the achievable rate region boundary of BC by performing MMSE-BF is equal to that on the rate region boundary of MAC with MMSE receiver filter at the BS, i.e.,

$$\gamma_i^B = \gamma_i^M. \quad (44)$$

We substitute (18) into (5), then substitute (5) and (16) into (44), and replace p_i with \tilde{p}_i by abuse of notation. As a result, \tilde{p}_i can be expressed as

$$\tilde{p}_i = \frac{q_i e_{\varphi(i)} \left(e_i + \rho \|\mathbf{h}_i\|^2 (1 - \theta) \right)}{e_i e_{\varphi(i)} + (1 - \theta) \left(\rho_i \|\mathbf{h}_i\|^2 e_{\varphi(i)} + \rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2 e_i \right)}, \quad (45)$$

where $e_i = \|\mathbf{h}_i\|^4 \rho_i^2 \theta + 2 \|\mathbf{h}_i\|^2 \rho_i \theta + 1$, $e_{\varphi(i)} = \rho_{\varphi(i)}^2 \|\mathbf{h}_{\varphi(i)}\|^4 \theta + 2 \rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2 \theta + 1$, $i = m, n$.

4.2. Adaptive Switching Method to Achieving a Larger Rate Region. Based on the aforementioned intersection points, we develop an adaptive switching method to achieve a larger rate region, i.e., Algorithm 1, which outputs some parameters such as the user's transmit power and BF vector in any channel condition. Combined with time-sharing, the switching method can achieve a convex hull of the union of NOMA-MRT and MMSE-BF's achievable rate region.

5. Joint User Pairing and Power Allocation with Transmission Mode Switching

In this section, for the problem of WSR maximization, we first proposed two optimal power allocation algorithms based on the concept of the achievable rate region, i.e., power allocation algorithm for NOMA-MRT mode (NOMA-MRT-PA) and power allocation algorithm for MMSE-BF mode (MMSE-BF-PA), by focusing on two-user case. Then, based on the two aforementioned proposed algorithms, a joint user pairing and power allocation algorithm (JUPA) is then developed for multiuser case. Finally, a practical transmission method is proposed by combining JUPA with transmission mode switching between NOMA-MRT and MMSE-BF.

5.1. Achieving Maximum WSR on the Rate Region Boundary for NOMA-MRT Mode. According to [19], the solution to the problem of WSR maximization must be on the boundary of achievable rate region of two-user MISO downlink systems. As a result, when the NOMA-MRT mode is employed at BS, the problem of WSR maximization is formulated as

$$\begin{aligned} (P2): \quad & \max_{\mathbf{p}_{m,n}} U(\mathbf{p}_{m,n}) := \mu_m r_m + \mu_n r_n, \\ & \text{s.t. } \mathbf{p}_{m,n} \in \mathcal{P}_{m,n}, \end{aligned} \quad (46)$$

where r_i ($i = m, n$) is the achievable rate of user i , defined in Lemma 1 or Lemma 2, and $\mathcal{P}_{m,n} = \{\mathbf{p}_{m,n} | 0 \leq p_i \leq P, p_i + p_{\varphi(i)} = P, i = m, n\}$ is the feasible set of power allocation vector for user m and n , in which the power allocation vector corresponding to rate point on the rate region boundary satisfies full power allocation, i.e., $p_i + p_{\varphi(i)} = P$.

Input: $\mathbf{h}_i \in \mathbb{C}^{M \times 1}$ (assuming $\|\mathbf{h}_m\| < \|\mathbf{h}_n\|$), $i = m, n$.
Output: $\mathbf{p}_{m,n}$, \mathbf{w}_i , $i = m, n$.

- (1) **Step 1:** obtain the intersection point $\rho_{n,k}$, $k = 1, 2, \dots, \kappa$.
- (2) **if** $\|\mathbf{h}_m\| < \|\mathbf{h}_n\|$
- (3) Calculate $\rho_{n,k}$ using (35) and (36)
- (4) **else if** $\mathbf{h}_m = \|\mathbf{h}_n\|$
- (5) Obtain $\rho_{n,k}$ using (43)
- (6) **end if**
- (7) **Step 2:** define as the mode switching point the intersection point excluding the ends of the interval $[0, \rho]$ and sort the mode switching points in ascending order, i.e., $\rho_{n,1} \leq \dots \leq \rho_{n,j} \leq \dots \leq \rho_{n,\Gamma}$, where Γ is the number of mode switching points. Divide rate region boundary into $\Gamma + 1$ sections by $\rho_{n,j}$.
- (8) **if** $\Gamma = 0$
- (9) **if** (15) is larger compared to (11) or (14), in $[0, \rho]$
- (10) Go to **Step 3**
- (11) **Else**
- (12) Go to **Step 4**
- (13) **end if**
- (14) **else** divide region boundary (15) into $\Gamma + 1$ sections by $\rho_{n,j}$, $j = 1, 2, \dots, \Gamma$.
- (15) **for** each section between $\rho_{n,j}$ and $\rho_{n,j+1}$ ($\rho_{n,0} = 0, \rho_{n,\Gamma+1} = \rho$)
- (16) **if** (15) is larger compared to (11) or (14), in this section
- (17) Go to **Step 3**
- (18) **else**
- (19) Go to **Step 4**
- (20) **end if**
- (21) **end for**
- (22) **end if**
- (23) **Step 3: Applying MMSE-BF**
- (24) Beamform with $\mathbf{w}_m, \mathbf{w}_n$ described in (18) and transmit $\mathbf{x} = \sqrt{\tilde{p}_m} \mathbf{w}_m s_m + \sqrt{\tilde{p}_n} \mathbf{w}_n s_n$, with $\mathbf{p}_{m,n} = [P - \tilde{p}_n, \tilde{p}_n]$, where \tilde{p}_n is calculated by (45), with $q_n = \rho_n N_0$, $\rho_m = \rho - \rho_n$ and $\rho_n \in (\rho_{n,j}, \rho_{n,j+1})$.
- (25) **Step 4: performing NOMA-MRT**
- (26) Beamform with $\mathbf{w}_m, \mathbf{w}_n$ expressed in (10) and transmit \mathbf{x} with $\mathbf{p}_{m,n} = [P - p_n, p_n]$, where $p_n = \tilde{p}_n N_0$, $\tilde{p}_n \in (\tilde{\rho}_{n,j}, \tilde{\rho}_{n,j+1})$ and $\tilde{\rho}_{n,j}$ is calculated by (38) (replacing subscript k with subscript j).

ALGORITHM 1: Adaptive switching method.

For the simplicity of notation, we let $U(\tilde{\rho}_n)$ stand for the WSR $U(\mathbf{p}_{m,n})$ in (46) with $p_n = \tilde{\rho}_n N_0$ and $p_m = P - p_n$. The motivations behind proposing NOMA-MRT-PA are as following:

- (1) According to Lemmas 1 and 2, the expression of the achievable rate region boundary of NOMA-MRT can be given in any channel conditions. Furthermore, the achievable rate r_i is a differentiable or piecewise differentiable function of $\tilde{\rho}_n$ in the interval $[0, \rho]$.
- (2) The global maximum value of $U(\tilde{\rho}_n)$ in the interval $[0, \rho]$ can be found by selecting the maximum among $U(0), U(\rho)$ and those corresponding to which the first-order derivative of $U(\tilde{\rho}_n)$ is zeros in the interval $(0, \rho)$. As we know, the local maximum values of $U(\tilde{\rho}_n)$ in the interval $(0, \rho)$ satisfy that the first-order derivative of $U(\tilde{\rho}_n)$ is equal to zero.

Theorem 1. When $\|\mathbf{h}_m\| < \|\mathbf{h}_n\|$, the solution of Problem (P2) is as follows.

Case (1). $\zeta_{n,m} \leq 0$.

$$\mathbf{p}_{m,n}^* = [p_m^*, p_n^*], \quad (47)$$

where $p_n^* = \tilde{\rho}_n^* N_0$, $p_m^* = P - p_n^*$, $\tilde{\rho}_n^* = \operatorname{argmax}_{\tilde{\rho}_n} \{U(0), U(\rho), U(\tilde{\rho}_{n,1})\}$, and

$$\tilde{\rho}_{n,1} = \frac{\mu_n \|\mathbf{h}_n\|^2 \rho \theta - (1 + \rho \|\mathbf{h}_n\|^2) (\mu_n - \mu_m (1 - \theta))}{\mu_n \|\mathbf{h}_n\|^2 \theta},$$

$$\tilde{\rho}_{n,1} \in [0, \rho]. \quad (48)$$

Case (2). $\zeta_{n,m} \geq \rho$.

$$\mathbf{p}_{m,n}^* = [p_m^*, p_n^*], \quad (49)$$

where $p_n^* = \tilde{\rho}_n^* N_0$, $p_m^* = P - p_n^*$, $\tilde{\rho}_n^* = \operatorname{argmax}_{\tilde{\rho}_n} \{U(0), U(\rho), U(\tilde{\rho}_{n,2}), U(\tilde{\rho}_{n,3})\}$, and

$$\tilde{\rho}_{n,(2,3)} = \frac{-\epsilon_1 \pm \sqrt{\epsilon_1^2 - 4\epsilon_2\epsilon_0}}{2\epsilon_2}, \quad (50)$$

where $\epsilon_2 = \mu_n \|\mathbf{h}_n\|^2 \|\mathbf{h}_m\|^4 \theta (1 - \theta)$, $\epsilon_1 = \|\mathbf{h}_n\|^2 (\mu_m \epsilon_3 - \mu_n \epsilon_4)$, $\epsilon_0 = \mu_m \epsilon_3 - \mu_n \|\mathbf{h}_n\|^2 (1 + \|\mathbf{h}_m\|^2 \rho)$, $\epsilon_3 = \|\mathbf{h}_m\|^2 (\|\mathbf{h}_m\|^2 \rho (1 - \theta) + 1)$, $\epsilon_4 = \|\mathbf{h}_m\|^2 (\|\mathbf{h}_m\|^2 \rho (1 - \theta) - 2\theta + 1)$, $\tilde{\rho}_{n,2} \in [0, \rho]$, and $\tilde{\rho}_{n,3} \in [0, \rho]$.

Case (3). $0 < \zeta_{n,m} < \rho$.

$$\mathbf{p}_{m,n}^* = [p_m^*, p_n^*], \quad (51)$$

where $p_n^* = \tilde{p}_n^* N_0$, $p_m^* = P - p_n^*$, $\tilde{p}_n^* = \operatorname{argmax}_{\tilde{p}_n} \{U(0), U(\rho), U(\zeta_{n,m}), U(\tilde{p}_{n,1}), U(\tilde{p}_{n,2}), U(\tilde{p}_{n,3})\}$, $\tilde{p}_{n,1} \in (\zeta_{n,m}, \rho]$, $\tilde{p}_{n,2} \in [0, \zeta_{n,m}]$, and $\tilde{p}_{n,3} \in [0, \zeta_{n,m}]$.

Proof. See the Appendix. \square

Theorem 2. When $\|\mathbf{h}_m\| = \|\mathbf{h}_n\|$, the solution of Problem (P2) is $\mathbf{p}_{m,n}^* = [P, 0]$ if $\mu_m \geq \mu_n$. Otherwise, the solution of Problem (P2) is $\mathbf{p}_{m,n}^* = [0, P]$.

Proof. See the Appendix.

For the simplicity of description of the proposed NOMA-MRT-PA, we assume $\|\mathbf{h}_m\| \leq \|\mathbf{h}_n\|$. By Theorems 1 and 2, NOMA-MRT-PA is proposed, i.e., Algorithm 2. \square

5.2. Achieving Maximum WSR on Rate Region Boundary for MMSE-BF Mode. When the MMSE-BF mode is performed at BS, the problem of WSR maximization can be formulated as

$$(P3): \max_{\mathbf{q}_{m,n}} U(\mathbf{q}_{m,n}) := \mu_m \tilde{r}_m + \mu_n \tilde{r}_n, \quad (52)$$

$$\text{s.t. } \mathbf{q}_{m,n} \in \mathcal{Q}_{m,n},$$

where $\mathbf{q}_{m,n} = (q_m, q_n)$ is transmit power vector for users m and n in the dual MAC, \tilde{r}_i and q_i are defined in Lemma 3, and $\mathcal{Q}_{m,n} = \{\mathbf{q}_{m,n} | 0 \leq q_i \leq P, q_i + q_{\varphi(i)} = P, i = m, n\}$ is the feasible set of transmit power vector for users m and n .

For the simplicity of notation, we let $U(\rho_m)$ stand for the WSR $U(\mathbf{q}_{m,n})$ in (52) with $q_m = \rho_m N_0$ and $q_n = P - q_m$. With the similar idea to that, by which Theorem 1 was proposed, we have the following theorem.

Theorem 3. The solution of Problem (P3) is $\mathbf{q}_{m,n}^* = [q_m^*, q_n^*]$ in dual two-user MAC, where $q_m^* = \rho_m^* N_0$, $q_n^* = P - q_m^*$, $\rho_m^* = \operatorname{argmax}_{\rho_m} \{U(0), U(\rho), U(\rho_{m,1}), U(\rho_{m,2}), \dots, U(\rho_{m,9})\}$, and $\rho_{m,i} \in [0, \rho]$, which are the roots of the nine-degree equation $U'(\rho_m) = 0$, which is given by (54).

Proof. According to Lemma 3, the WSR $U(\rho_m)$ can be expressed as

$$U(\rho_m) = \mu_m \log \left(1 + \frac{\rho_m \|\mathbf{h}_m\|^2 (1 + \rho_n \|\mathbf{h}_n\|^2 \theta)^2}{\varsigma_m + \rho_n \|\mathbf{h}_n\|^2 (1 - \theta)} \right) + \mu_n \log \left(1 + \frac{\rho_n \|\mathbf{h}_n\|^2 (1 + \rho_m \|\mathbf{h}_m\|^2 \theta)^2}{\varsigma_n + \rho_m \|\mathbf{h}_m\|^2 (1 - \theta)} \right), \quad (53)$$

where $\rho_n = \rho - \rho_m$. In this case, $U(\rho_m)$ is differentiable function of ρ_m in the interval $[0, \rho]$. By letting $U'(\rho_m) = 0$, we can obtain a nine-degree equation:

$$a_9 \rho_m^9 + a_8 \rho_m^8 + a_7 \rho_m^7 + a_6 \rho_m^6 + a_5 \rho_m^5 + a_4 \rho_m^4 + a_3 \rho_m^3 + a_2 \rho_m^2 + a_1 \rho_m + a_0 = 0. \quad (54)$$

Since the coefficients of equation (54) are very complicated, we omit them here. In this case, the global maximum value of $U(\rho_m)$ in the interval $[0, \rho]$ can be found by selecting the maximum among $U(0), U(\rho), U(\rho_{m,1}), U(\rho_{m,2}), \dots, U(\rho_{m,9})$. Consequently, the solution of Problem (P3) is $\mathbf{q}_{m,n}^* = [q_m^*, q_n^*]$ in dual two-user MAC, where $q_m^* = \rho_m^* N_0$, $q_n^* = P - q_m^*$, $\rho_m^* = \operatorname{argmax}_{\rho_m} \{U(0), U(\rho), U(\rho_{m,1}), U(\rho_{m,2}), \dots, U(\rho_{m,9})\}$, and $\rho_{m,i} \in [0, \rho]$, $i = 1, 2, \dots, 9$. If $\rho_{m,i} \notin [0, \rho]$, $i = 1, 2, \dots, 9$, we let $U(\rho_{m,i}) = 0$.

By Theorem 3, MMSE-BF-PA is proposed, i.e., Algorithm 3. \square

5.3. Joint User Pairing and Power Allocation (JUPA) Algorithm. With the aim of maximizing the total sum rate, in the step of forming a user pair in JUPA, we compare the two sum rates obtained by NOMA-MRT-PA and MMSE-BF-PA, select the transmission mode with the larger sum rate, and then choose the corresponding user pair to send data. By focusing on the expression of rate region boundary of NOMA-MRT in (11), it is worthwhile noticing the following properties:

- (1) Considering the original user pairing problem which is how to pair user i with a user from the set of other unpaired users to obtain the maximum sum rate, we can divide all of unpaired users besides user i into 3 subsets, i.e., Φ_1^i , Φ_2^i , and Φ_3^i , according to $\zeta_{i,i'}$. Particularly, user i' in Φ_1^i , Φ_2^i , and Φ_3^i , satisfies $\zeta_{i,i'} \leq 0$, $\zeta_{i,i'} \geq \rho$ and $0 < \zeta_{i,i'} < \rho$, respectively. The original user pairing problem is equivalent to 3 problems, which are how to pair user i with a user to obtain the local-maximum sum rate from Φ_1^i , Φ_2^i , and Φ_3^i , respectively. Therefore, the solution of the original user pairing problem is the user pair, which corresponds to the maximum among the local maximum sum rates for Φ_1^i , Φ_2^i , and Φ_3^i .
- (2) Considering the problem which is how to pair user i with a user from Φ_1^i (Φ_2^i) to achieve local maximum sum rate, user i should be paired with the user, whose channel vector can form the minimum (maximum) angle with that of user i , according to (11).
- (3) For the problem which is how to pair user i with a user from Φ_3^i to achieve local maximum sum rate, we divide Φ_3^i into 2 subsets, i.e., $\Phi_{3,1}^i$ and $\Phi_{3,2}^i$. The user i' in $\Phi_{3,1}^i$ and $\Phi_{3,2}^i$ satisfies $0 < \zeta_{i,i'} \leq (\rho/2)$ and $(\rho/2) < \zeta_{i,i'} < \rho$, respectively. According to (11), the formula $r_{i'} = \log(1 + (\tilde{\rho}_{i'} \|\mathbf{h}_i\|^2 (1 - \theta) / (1 + \tilde{\rho}_i \|\mathbf{h}_i\|^2)))$ provides dominant contribution to rate region boundary formed by user i and i' , when $0 < \zeta_{i,i'} \leq (\rho/2)$. In this case, for the treatability of the problem, we use the formula $r_{i'} = \log(1 + (\tilde{\rho}_{i'} \|\mathbf{h}_i\|^2 (1 - \theta) / (1 + \tilde{\rho}_i \|\mathbf{h}_i\|^2)))$ as the expression of $r_{i'}$ in the whole interval $[0, \rho]$, instead of the piecewise formula as shown in (11). Thus, for the problem which is how to pair user i with

Input: $\mu_i, \mathbf{h}_i \in \mathbb{C}^{M \times 1}$ (assuming $\|\mathbf{h}_m\| < \|\mathbf{h}_n\|$), $i = m, n$.

Output: $\mathbf{p}_{m,n}^*$.

- (1) **if** $\|\mathbf{h}_m\| < \|\mathbf{h}_n\|$
- (2) Calculate $\zeta_{n,m}$ using (12)
- (3) **if** $\zeta_{n,m} \leq 0$
- (4) $\mathbf{p}_{m,n}^* = [P - \tilde{\rho}_n^* N_0, \tilde{\rho}_n^* N_0]$, where $\tilde{\rho}_n^* = \operatorname{argmax}_{\tilde{\rho}_n} \{U(0), U(\rho), U(\tilde{\rho}_{n,1})\}$, $U(\tilde{\rho}_{n,1})$ is calculated by (48), $\tilde{\rho}_{n,1} \in [0, \rho]$.
- (5) **else if** $\zeta_{n,m} \geq \rho$
- (6) $\mathbf{p}_{m,n}^* = [P - \tilde{\rho}_n^* N_0, \tilde{\rho}_n^* N_0]$, where $\tilde{\rho}_n^* = \operatorname{argmax}_{\tilde{\rho}_n} \{U(0), U(\rho), U(\tilde{\rho}_{n,2}), U(\tilde{\rho}_{n,3})\}$, $U(\tilde{\rho}_{n,2})$ and $U(\tilde{\rho}_{n,3})$ are calculated by (50), $\tilde{\rho}_{n,2} \in [0, \rho]$, $\tilde{\rho}_{n,3} \in [0, \rho]$.
- (7) **else**
- (8) $\mathbf{p}_{m,n}^* = [P - \tilde{\rho}_n^* N_0, \tilde{\rho}_n^* N_0]$, where $\tilde{\rho}_n^* = \operatorname{argmax}_{\tilde{\rho}_n} \{U(0), U(\rho), U(\zeta_{n,m}), U(\tilde{\rho}_{n,1}), U(\tilde{\rho}_{n,2}), U(\tilde{\rho}_{n,3})\}$, $\tilde{\rho}_{n,1} \in (\zeta_{n,m}, \rho]$, $\tilde{\rho}_{n,2} \in [0, \zeta_{n,m}]$ and $\tilde{\rho}_{n,3} \in [0, \zeta_{n,m}]$
- (9) **end if**
- (10) **else**
- (11) **If** $\mu_m \geq \mu_n$, $\mathbf{p}_{m,n}^* = [P, 0]$. Otherwise, $\mathbf{p}_{m,n}^* = [0, P]$.
- (12) **end if**

ALGORITHM 2: NOMA-MRT-PA.

Input: $\mu_i, \mathbf{h}_i \in \mathbb{C}^{M \times 1}$, $i = m, n$.

Output: $\mathbf{p}_{m,n}^*$.

- (1) Calculate $\rho_{m,i}$ using (54), $i = 1, 2, \dots, 9$.
- (2) $\rho_m^* = \operatorname{argmax}_{\rho_m} \{U(0), U(\rho), U(\rho_{m,1}), U(\rho_{m,2}), \dots, U(\rho_{m,9})\}$, where $\rho_{m,i} \in [0, \rho]$.
- (3) $\mathbf{p}_{m,n}^* = [\tilde{p}_m^*, P - \tilde{p}_m^*]$, where \tilde{p}_m^* is calculated by (45), with $q_m^* = \rho_m^* N_0$.

ALGORITHM 3: MMSE-BF-PA.

Input: $\mathbf{h}_i \in \mathbb{C}^{M \times 1}$ (assuming $\|\mathbf{h}_1\| \leq \dots \leq \|\mathbf{h}_{2K}\|$), $\mu_i = 1$, $i = 1, 2, \dots, 2K$.

Output: Ω and \mathbf{U}

Initial: $\Omega = \emptyset$ and $\mathbf{U} = \mathbf{0}_{2K \times 2K}$.

- (1) **for** $i = 1: 2K - 1$ **do**
- (2) **if** user i has been paired
- (3) Go to line 1
- (4) **end if**
- (5) **for** $j = i + 1: 2K$ **do**
- (6) **if** user j has been paired
- (7) Go to line 5
- (8) **end if**
- (9) **if** \mathbf{h}_i and \mathbf{h}_j is orthogonal
- (10) Obtain $\mathbf{p}_{i,j}$ by MMSE-BF-PA, $\Omega = \Omega \cup \{(i, j, \mathbf{p}_{i,j}, 0)\}$ and go to line 1
- (11) **end if**
- (12) **end for**
- (13) Find $k_1 = \operatorname{argmax}_k \cos^2 \angle(\mathbf{h}_i, \mathbf{h}_{\Phi_1^i(k)})$ and obtain the power allocation solution $\mathbf{p}_{i, \Phi_1^i(k_1)}$ for the user pair $\{i, \Phi_1^i(k_1)\}$ by NOMA-MRT-PA
- (14) Find $k_2 = \operatorname{argmin}_k \cos^2 \angle(\mathbf{h}_i, \mathbf{h}_{\Phi_2^i(k)})$ and obtain the power allocation solution $\mathbf{p}_{i, \Phi_2^i(k_2)}$ for the user pair $\{i, \Phi_2^i(k_2)\}$ by NOMA-MRT-PA
- (15) Find $k_{3,1} = \operatorname{argmax}_k \cos^2 \angle(\mathbf{h}_i, \mathbf{h}_{\Phi_{3,1}^i(k)})$ and obtain the power allocation solution $\mathbf{p}_{i, \Phi_{3,1}^i(k_{3,1})}$ for the user pair $\{i, \Phi_{3,1}^i(k_{3,1})\}$ by NOMA-MRT-PA
- (16) Find $k_{3,2} = \operatorname{argmin}_k \cos^2 \angle(\mathbf{h}_i, \mathbf{h}_{\Phi_{3,2}^i(k)})$ and obtain the power allocation solution $\mathbf{p}_{i, \Phi_{3,2}^i(k_{3,2})}$ for the user pair $\{i, \Phi_{3,2}^i(k_{3,2})\}$ by NOMA-MRT-PA
- (17) Obtain the power allocation vector with the maximum WSR, from the solutions obtained by lines 13–16:
 $\mathbf{p}^N = \operatorname{argmax}_{\mathbf{p}} \{U(\mathbf{p}_{i, \Phi_1^i(k_1)}), U(\mathbf{p}_{i, \Phi_2^i(k_2)}), U(\mathbf{p}_{i, \Phi_{3,1}^i(k_{3,1})}), U(\mathbf{p}_{i, \Phi_{3,2}^i(k_{3,2})})\}$.
Let $\{i, j^N\}$ denote the user pair which corresponds to \mathbf{p}^N .
- (18) Find $j^M = \operatorname{argmin}_j \cos \angle(\mathbf{h}_i, \mathbf{h}_j)$ and obtain the power allocation solution \mathbf{p}_{i, j^M} for the user pair $\{i, j^M\}$ by MMSE-BF-PA.
- (19) **if** $U(\mathbf{p}^N) \geq U(\mathbf{p}_{i, j^M})$

ALGORITHM 4: Continued.

```

(20)    $\Omega = \Omega \cup \{(i, j^N, \mathbf{p}^N, 1)\}$  and  $u_{i,j^N} = u_{j^N,i} = 1$ .
(21)   else
(22)    $\Omega = \Omega \cup \{(i, j^M, \mathbf{p}_{i,j^M}, 0)\}$  and  $u_{i,j^M} = u_{j^M,i} = 1$ .
(23)   end if
(24) end for

```

ALGORITHM 4: Joint user pairing and power allocation algorithm (JUPA).

a user from $\Phi_{3,1}^i$, user i should be paired with the user whose channel vector can form the minimum angle with that of user i . With the similar analysis, user i should be paired with the user in $\Phi_{3,2}^i$ whose channel vector can form the maximum angle with that of user i .

For the simplicity of description of the proposed JUPA (Algorithm 4), assume that users are ordered as $\|\mathbf{h}_1\| \leq \dots \leq \|\mathbf{h}_{2K}\|$. Let $\Phi_1^i(k)$ ($\Phi_2^i(k)$) denote the index of k -th entry in Φ_1^i (Φ_2^i). Besides, let $\Phi_{3,1}^i(k)$ ($\Phi_{3,2}^i(k)$) denote the index of k -th entry in $\Phi_{3,1}^i$ ($\Phi_{3,2}^i$).

In Algorithm 4, line 13 (14, 15, and 16) is the step for selecting a user to be paired with user i and deriving the power allocation solution for the obtained user pair by NOMA-MRT-PA, in the set Φ_1^i (Φ_2^i , $\Phi_{3,1}^i$, and $\Phi_{3,2}^i$). Line 18 is the step for similar operation by MMSE-BF-PA. The output of Algorithm 4, Ω , is termed the pairing and power allocation configuration, in which K entries are included and 4 elements are contained in each entry. The first two elements of i -th entry in Ω denote the indexes of the two users in i -th user pair, and the third and fourth element stand for power allocation vector and transmission mode adopted by the i -th user pair, respectively.

When the number of users is odd, the last remained user can be trivially served by OMA. Specifically, if the channel vector of the last remained user is \mathbf{h}_l , then MRT BF is performed at the BS and the BF vector is given by (10). As a result, the maximum signal-to-noise ratio can be achieved.

5.3.1. Complexity Analysis. The computational complexity of the proposed JUPA mainly comes from the computation of the angle between channels of two users and from solving the nine-degree equation (54), by which the optimal power allocation scheme is obtained for MMSE-BF mode.

Note that the computational complexity for computing the angle between channels of two users is $\mathcal{O}(M)$. In addition, we use upper-case letter ‘‘C’’ to indicate the computational complexity for solving the nine-degree equation (54). As the number of computing the angle is $2K - 2(i - 1) - 1$ in i -th outer ‘‘for’’ loop in Algorithm 4, for the MMSE-BF mode or NOMA-MRT mode, the computational complexity for the user pairing process is $\mathcal{O}((2K/2)^2 M)$. On the other hand, the computational complexity for the power allocation process is $\mathcal{O}((2K/2)C)$. Consequently, the overall computational complexity of Algorithm 4 is $\mathcal{O}((2K/2)^2 M) + \mathcal{O}((2K/2)C)$.

For the exhaustive search-based scheme, i.e., exhaustive search-based user pairing and power allocation (ES-PPA) described in Section 6.2, the number of all possible pairing

schemes is $(2K - 1)!!$. In each pairing scheme, there are K user pairs. Then, for each user pair, the angle between channels of two users and the allocated power of two users for MMSE-BF mode should be calculated. Therefore, the total computational complexity of exhaustive search-based scheme is $\mathcal{O}((2K - 1)!!(2K/2)M) + \mathcal{O}((2K - 1)!!(2K/2)C)$.

5.4. JUPA/TMS: A Practical Transmission Method. By combining JUPA with transmission mode switching (TMS) between NOMA-MRT and MMSE-BF, a practical transmission scheme termed JUPA/TMS is proposed. JUPA/TMS is described in Algorithm 5, where S is the control parameter for transmission mode switching and $\Omega(i)_j$ denotes the j -th element in the i -th entry in the set Ω . Noting that TDMA(FDMA) is adopted in Algorithm 5 (line10), it is efficient when M is small. When the scenario with sufficient number of antennas at base station is considered, e.g., $M \geq 2K$, the performance can be further improved by combining the proposed JUPA with SDMA, which is beyond the scope of this paper.

6. Results and Discussion

In this section, we first provide numerical results on the rate regions of different schemes (ZFBB, MRT, MMSE-BF, and NOMA-MRT) under various channel conditions and validate the performance of the proposed adaptive switching method by comparing with other schemes. Then, we demonstrate the optimality of NOMA-MRT-PA and MMSE-BF-PA. Furthermore, we also verify the performance of the proposed JUPA by comparing with traditional transmission schemes in simulation.

6.1. The Adaptive Switching Method. In Figure 3, numerical results of performance comparison among NOMA-MRT, MMSE-BF, MRT, and ZFBB are derived in various channel states. When $\|\mathbf{h}_n\| = 2$ and $\alpha = \pi/8$ implying that the angle between two users’ channel vectors is smaller in space, NOMA-MRT gets better performance than MMSE-BF absolutely and the rate region of MMSE-BF is completely within that of NOMA-MRT as illustrated in Figure 3(a). In Figure 3(b), when $\|\mathbf{h}_n\| = 2$ and $\alpha = \pi/4$, larger R_n can be obtained by implementing NOMA-MRT in high R_m area, while in low R_m area, larger R_n can be derived by using MMSE-BF. When $\|\mathbf{h}_n\| = 2$ and $\alpha = 3\pi/8$ which implies that the angle between channel vectors of two users becomes larger, MMSE-BF performs better than NOMA-MRT completely as shown in Figure 3(c). In symmetric case where

```

Input:  $\mathbf{h}_i \in \mathbb{C}^{M \times 1}$ ,  $s_i$ ,  $i = 1, 2, \dots, 2K$ .
Output:  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K$ .
(1) Perform JUPA to obtain pairing and power allocation configuration  $\Omega$ 
(2) for  $i = 1: K$  do
(3)    $(m, n, \mathbf{p}_{m,n}, T_{m,n}) \leftarrow (\Omega(i)_1, \Omega(i)_2, \Omega(i)_3, \Omega(i)_4)$ 
(4)   if  $T_{m,n} = 1$ 
(5)     Calculate  $\mathbf{w}_m$  and  $\mathbf{w}_n$  with (10)
(6)   else if  $T_{m,n} = 0$ 
(7)     Calculate  $\mathbf{w}_m$  and  $\mathbf{w}_n$  with (18)
(8)   end if
(9)    $\mathbf{x}_i \leftarrow \sqrt{P_m} \mathbf{w}_m s_m + \sqrt{P_n} \mathbf{w}_n s_n$ 
(10)  Transmit  $x_i$  in  $i$ -th frequency/time slot
(11) end for

```

ALGORITHM 5: JUPA/TMS.

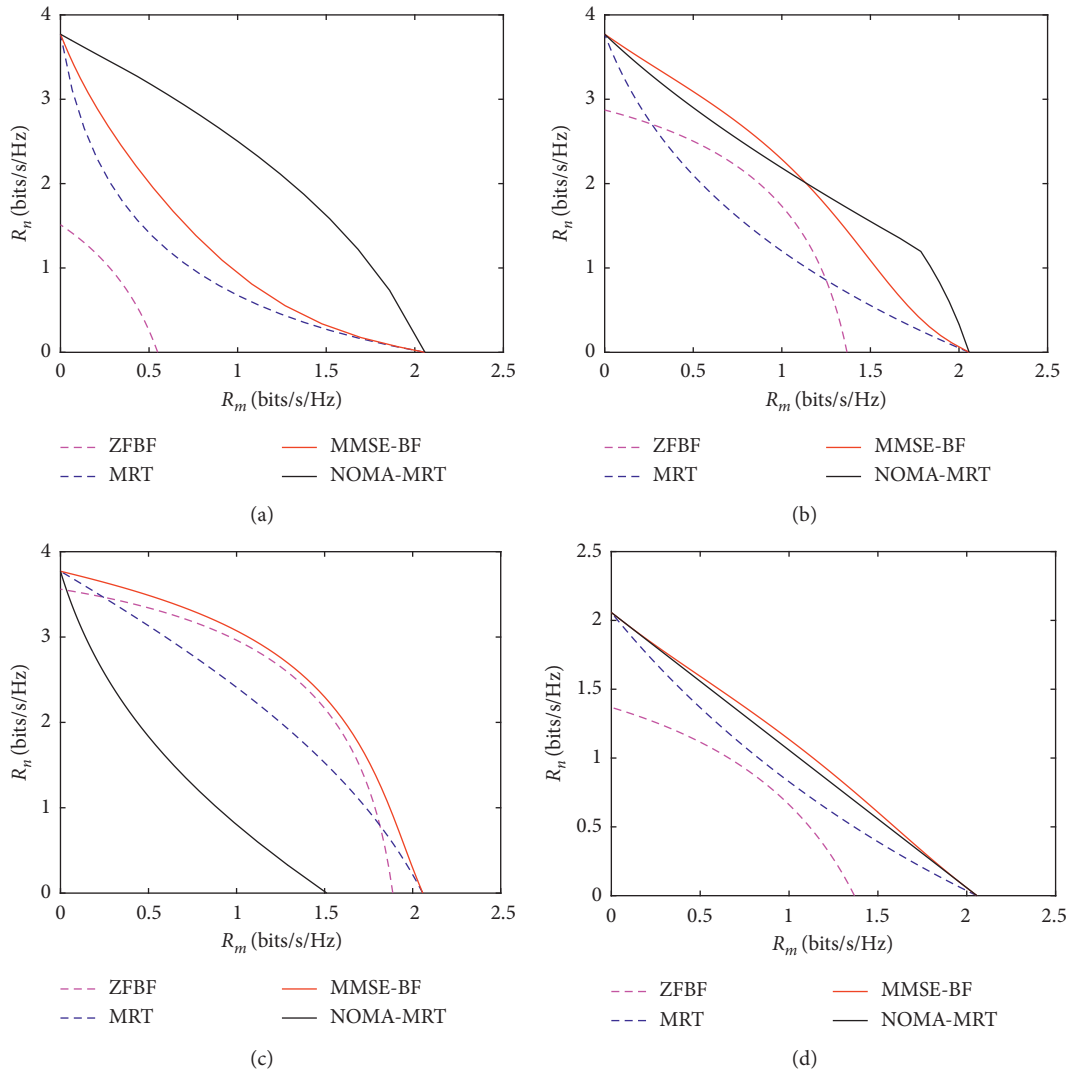


FIGURE 3: Rate region comparison of ZFBF, MRT, MMSE-BF, and NOMA-MRT in various channel states. $P/N_0 = 5$ dB, $\|\mathbf{h}_m\| = 1$. (a) $\|\mathbf{h}_n\| = 2$, $\alpha = \pi/8$. (b) $\|\mathbf{h}_n\| = 2$, $\alpha = \pi/4$. (c) $\|\mathbf{h}_n\| = 2$, $\alpha = 3\pi/8$. (d) $\|\mathbf{h}_n\| = 1$, $\alpha = \pi/4$. Rate region boundaries of the four schemes is without time-sharing.

$\|\mathbf{h}_m\| = \|\mathbf{h}_n\| = 1$ and $\alpha = \pi/4$, the rate region of MMSE-BF covers that of NOMA-MRT. From Figures 3(a)–3(d), as MMSE-BF can optimally trade off fighting interference to other users and the background Gaussian noise, MMSE-BF gets better performance than ZFBF and MRT absolutely, in any channel state.

As demonstrated in Figure 4, the adaptive switching method can obtain a larger rate region than the other schemes, which can be illustrated as the convex hull of the union of NOMA-MRT and MMSE-BF’s rate regions by combining with time-sharing.

6.2. Joint User Pairing and Power Allocation with Transmission Mode Switching. In Figure 5, with different weight vectors, the rate points on rate region boundary and the corresponding maximum WSR points are obtained by NOMA-MRT-PA, MMSE-BF-PA, and exhaustive search, respectively. When weight vector \mathbf{u} is $(0.5, 0.5)$, $(0.33, 0.67)$, or $(0.25, 0.75)$, rate points obtained by NOMA-MRT-PA or MMSE-BF-PA are superposed together with those derived by exhaustive search, which means that NOMA-MRT-PA and MMSE-BF can derive optimal rate point for different weight vector as shown in Figure 5(a). The rate points obtained by NOMA-MRT-PA for the case $\mathbf{u} = (0.33, 0.67)$ and $\mathbf{u} = (0.25, 0.75)$ are superposed together, since the rate points to achieve the maximum WSR for $\mathbf{u} = (0.33, 0.67)$ and $\mathbf{u} = (0.25, 0.75)$ both occur at the inflection point of the achievable rate region boundary of NOMA-MRT. In Figure 5(b), with $\mathbf{u} = (0.5, 0.5)$, the WSRs obtained by NOMA-MRT (red curve) and MMSE-BF (blue curve) are given, respectively, when the allocated power of user m varies, with P being normalized to 1. The maximum WSR points corresponding to the rate point with $\mathbf{u} = (0.5, 0.5)$ in Figure 5(a) can be obtained by NOMA-MRT-PA and MMSE-BF-PA, respectively, each of which is also superposed with that derived by exhaustive search. Figures 5(c) and 5(d) are the cases for $\mathbf{u} = (0.33, 0.67)$ and $\mathbf{u} = (0.25, 0.75)$, respectively.

To validate the effectiveness of the proposed JUPA scheme, other schemes which combine the existing user pairing algorithms (exhaustive search, greedy algorithm, correlation pairing [27, 28], and random pairing) with NOMA-MRT-PA and MMSE-BF-PA are also considered for comparison in simulation, including exhaustive search-based user pairing and power allocation algorithm (ES-PPA), greedy algorithm-based user pairing and power allocation algorithm (GA-PPA), correlation pairing-based user pairing and power allocation algorithm (COR-PPA), and random pairing-based user pairing and power allocation algorithm (RAN-PPA). In the following, we assume noise power $N_0 = -10$ dB, the cell radius $R = 30$ m, and the path loss exponent $\alpha' = 3$, and TDMA technology is adopted. The bandwidth is normalized to one. Assume that the weights of all users are set to one. The key idea of GA-PPA is to pair the two users with the maximum sum rate obtained by NOMA-MRT-PA or MMSE-BF-PA. Mathematically, for a fixed user i , we pair it with user j^* , if

$$j^* = \arg \max_j \left\{ \max \left\{ U_{i,j}^{\text{NOMA-MRT}}, U_{i,j}^{\text{MMSE-BF}} \right\} \right\}, \quad j \neq i, \quad (55)$$

where $U_{i,j}^{\text{NOMA-MRT}}$ and $U_{i,j}^{\text{MMSE-BF}}$ stand for the sum rate of users i and j , obtained by NOMA-MRT-PA and MMSE-BF-PA, respectively. With the similar architecture to the outer “for” loop in the Algorithm 4, the numbers of computing the angle between channels of two users used by NOMA-MRT-PA and the number of solving the equation (54) used by MMSE-BF-PA are both $2K - 2(i - 1) - 1$ in i -th outer “for” loop for GA-PPA. Therefore, its computational complexity is $\mathcal{O}((2K/2)^2 M) + \mathcal{O}((2K/2)^2 C)$ in total. In RAN-PPA and COR-PPA, we first obtain user pairing solution by random pairing algorithm and correlation pairing algorithm, respectively, and then derive the sum rate for each user pair by selecting the larger one among the sum rates achieved by NOMA-MRT-PA and MMSE-BF-PA.

The key idea of correlation pairing algorithm is only to pair two users with the maximum channel correlation. Specifically, for a fixed user i , we pair it with user j^* satisfying

$$j^* = \arg \max_j \frac{|\mathbf{h}_j^H \mathbf{h}_i|^2}{\|\mathbf{h}_i\|^2 \|\mathbf{h}_j\|^2}, \quad j \neq i. \quad (56)$$

Since the ideas of forming a user pair in COR-PPA and the proposed JUPA are both based on the angle between channels of the two users, COR-PPA and JUPA have the same computational complexity (considering the complexity of computing the angle between two users’ channels and solving equation (54)), i.e., $\mathcal{O}((2K/2)^2 M) + \mathcal{O}((2K/2)C)$. In ES-PPA, we first obtain $(2K - 1)!!$ possible pairing schemes by exhaustive search and then select the pairing scheme with maximum total sum rate, which is the sum of sum rates of K user pairs.

For summary, JUPA and COR-PPA have roughly the same computational complexity. However, GA-PPA and ES-PPA require much higher computational complexity than JUPA, especially for ES-PPA. Due to the random pairing method adopted in RAN-PPA, its computational complexity is the lowest, which only results from computing the angle between two users’ channels used by NOMA-MRT-PA and from solving equation (54) utilized by MMSE-BF-PA, for each user pair. The computational complexity of these schemes is summarized in Table 2.

In Figures 6 and 7, the sum rates of different schemes, including ES-PPA, RAN-PPA, GA-PPA, COR-PPA, and the proposed JUPA, are compared, when the transmit power P and the number of antennas at the BS vary, respectively. In Figure 6, the sum rates of all schemes grow with the transmit power P increasing. Compared with ES-PPA using exhaustive search to obtain the user pairing solution, the proposed JUPA only suffers very slight performance loss. However, the computational complexity of JUPA is only $\mathcal{O}((2K/2)^2 M) + \mathcal{O}((2K/2)C)$, while that of ES-PPA is $\mathcal{O}((2K - 1)!!(2K/2)M) + \mathcal{O}((2K - 1)!!(2K/2)C)$. In comparison with other schemes, the JUPA derives better performance and the performance gain boosts with P increasing. Further, JUPA outperforms GA-PPA, since GA does not consider the performance loss caused by the users with poor channels. COR-PPA performs worse than RAN-

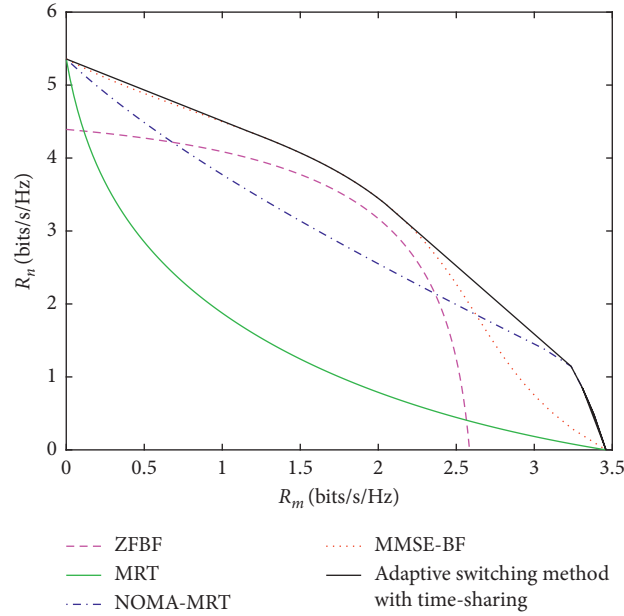


FIGURE 4: Rate region of the adaptive switching method with time-sharing. $P/N_0 = 10$ dB, $\|\mathbf{h}_m\| = 1$, $\|\mathbf{h}_n\| = 2$, $\alpha = \pi/4$. Also plotted for comparison are rate regions of ZFBF, MRT, MMSE-BF, and NOMA-MRT, without time-sharing.

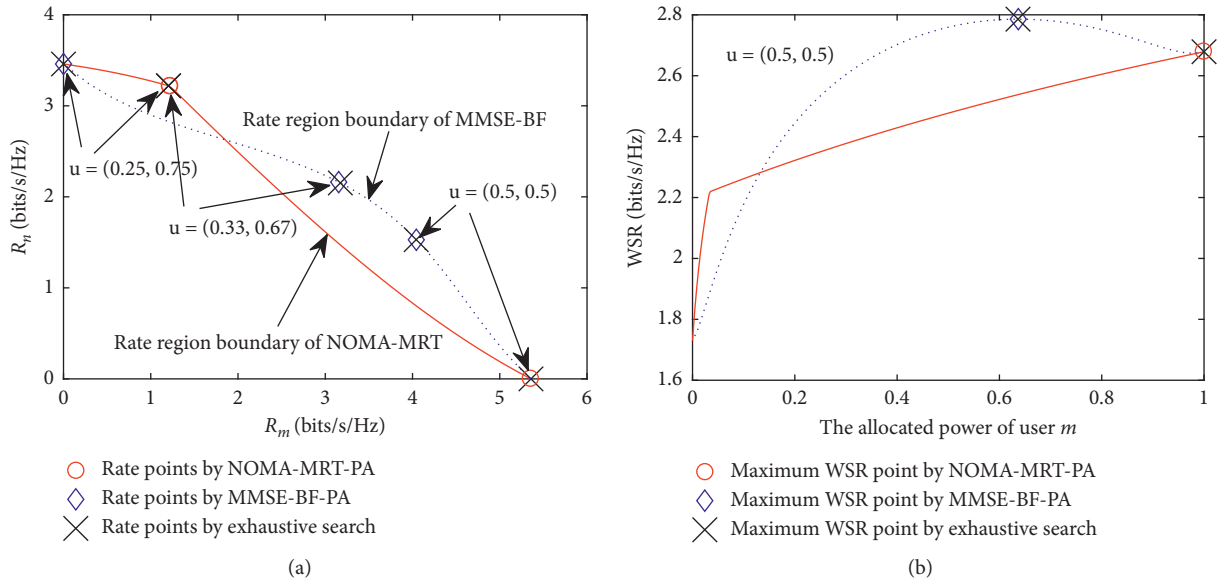


FIGURE 5: Continued.

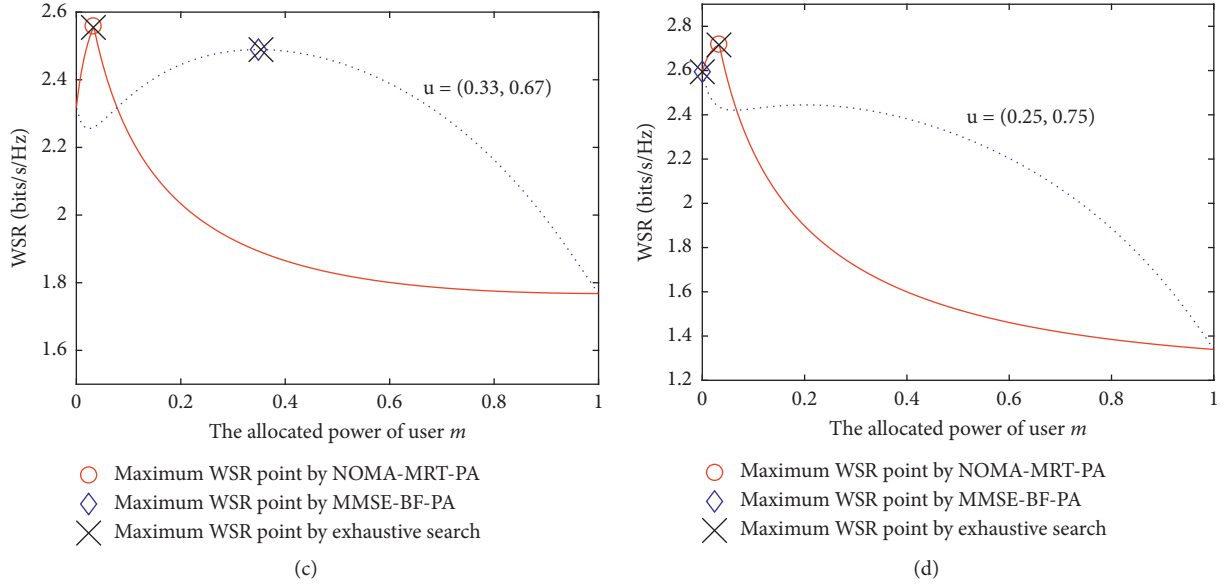


FIGURE 5: Rate points and the corresponding maximum WSR points obtained by NOMA-MRT-PA, MMSE-BF-PA, and exhaustive search. $P/N_0 = 10$ dB, $\|\mathbf{h}_m\| = 2$, $\|\mathbf{h}_n\| = 1$, $\alpha = \pi/4$. (a) Rate points under different weight vectors. Maximum WSR point with (b) $\mathbf{u} = (0.5, 0.5)$, (c) $\mathbf{u} = (0.33, 0.67)$, and (d) $\mathbf{u} = (0.25, 0.75)$.

TABLE 2: Computational complexity summary for various schemes.

User pairing and power allocation schemes	Computational complexity
GA-PPA	$\mathcal{O}((2K/2)^2 M) + \mathcal{O}((2K/2)^2 C)$
COR-PPA	$\mathcal{O}((2K/2)^2 M) + \mathcal{O}((2K/2)C)$
RAN-PPA	$\mathcal{O}((2K/2)M) + \mathcal{O}((2K/2)C)$
ES-PPA	$\mathcal{O}((2K-1)!!(2K/2)M) + \mathcal{O}((2K-1)!!(2K/2)C)$
JUPA	$\mathcal{O}((2K/2)^2 M) + \mathcal{O}((2K/2)C)$

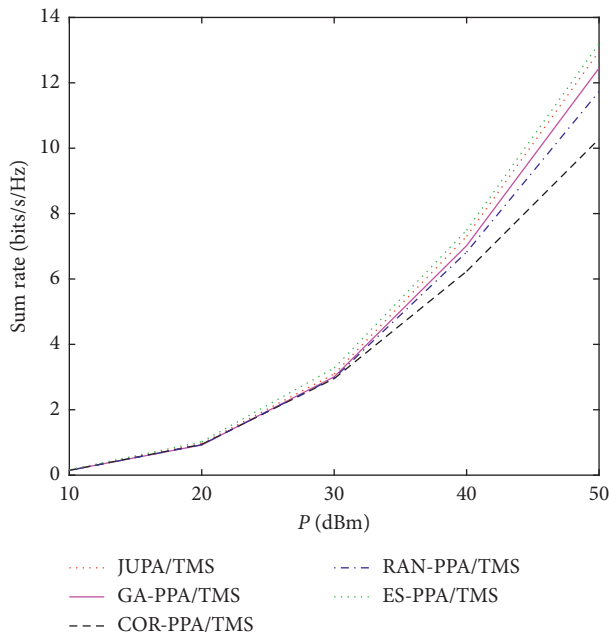


FIGURE 6: Sum rate versus P for different transmission schemes ($M = 2$, $2K = 8$).

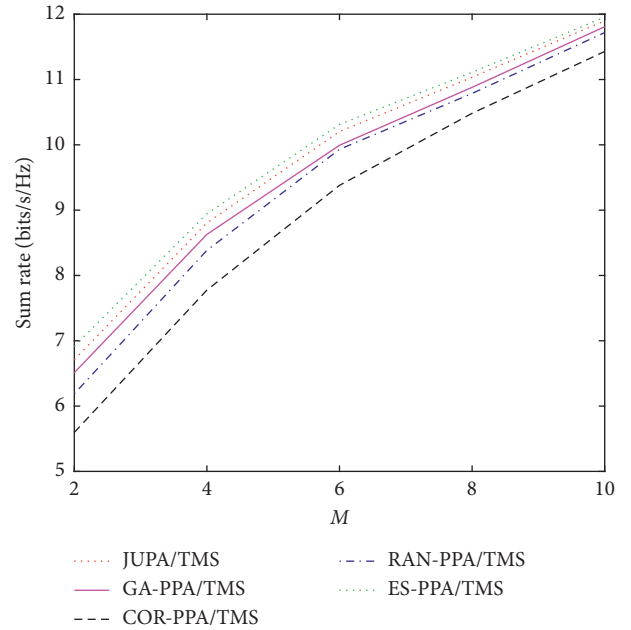


FIGURE 7: Sum rate versus the number of the antennas for several transmission schemes ($P = 40$ dBm, $2K = 8$).

PPA, since correlation-based user pairing algorithm breaks the channels orthogonality, i.e., the users with orthogonal channel are never paired together.

As demonstrated in Figure 7, it can be observed that the proposed JUPA results in a slight performance loss compared with ES-PPA. However, the JUPA outperforms other schemes and the performance gain over other schemes decreases as the number of antennas at the BS increases.

7. Conclusion

In this paper, the design of adaptive TMS between NOMA-MRT and MMSE-BF to maximize the sum rate for downlink MISO systems was investigated. Firstly, the closed-form expressions of the boundary of achievable rate region for NOMA-MRT and MMSE-BF were obtained. It has been shown that when the channel vectors of the two users are greatly correlated, the achievable rate region of NOMA-MRT includes that of MMSE-BF completely. However, when the two channels are almost orthogonal, the opposite conclusion can be drawn. As a result, an adaptive switching method is developed to achieve a larger rate region for the two-user case. Subsequently, the optimal power allocation algorithms to maximize weighted sum rate for both NOMA-MRT mode and MMSE-BF mode were presented. Finally, the low-complexity JUPA is consequently developed to further improve the sum rate performance for the multiuser case. Compared with the exhaustive search-based scheme with the computational complexity of $\mathcal{O}((2K-1)!(2K/2)M) + \mathcal{O}((2K-1)!(2K/2)C)$, the proposed JUPA can obtain a much lower complexity of $\mathcal{O}((2K/2)^2M) + \mathcal{O}((2K/2)C)$ and only suffers a slight sum rate performance loss, and it outperforms other conventional schemes.

Appendix

Proof. of Lemma 1.

Case (1). $\theta > 0$.

When the equality holds for (7), the achievable rate region boundary of NOMA-based BF for MISO broadcast channel can be achieved. As a result, the achievable rate region boundary of NOMA-MRT can be given by

$$r_n = \log(1 + \tilde{\rho}_n \|\mathbf{h}_n\|^2), \quad (\text{A.1})$$

$$r_m = \min\{\log(1 + \gamma_{m,n}^B), \log(1 + \gamma_m^B)\}, \quad (\text{A.2})$$

where

$$\begin{cases} \gamma_{m,n}^B = \frac{\tilde{\rho}_m \|\mathbf{h}_n\|^2 (1 - \theta)}{1 + \tilde{\rho}_n \|\mathbf{h}_n\|^2}, \\ \gamma_m^B = \frac{\tilde{\rho}_m \|\mathbf{h}_m\|^2}{1 + \tilde{\rho}_n \|\mathbf{h}_m\|^2 (1 - \theta)} \end{cases} \quad (\text{A.3})$$

and $\tilde{\rho}_m = \rho - \tilde{\rho}_n$.

Define

$$f(\tilde{\rho}_n) = \gamma_{m,n}^B - \gamma_m^B. \quad (\text{A.4})$$

We let $f(\tilde{\rho}_n) = 0$. Subsequently, we can derive a quadratic equation with respect to $\tilde{\rho}_n$.

$$m_1 \tilde{\rho}_n^2 + m_2 \tilde{\rho}_n + m_3 = 0, \quad (\text{A.5})$$

where $m_1 = \|\mathbf{h}_n\|^2 \|\mathbf{h}_m\|^2 (2\theta - \theta^2)$, $m_2 = -(\|\mathbf{h}_n\|^2 (1 - \theta) - \|\mathbf{h}_m\|^2) - \rho \|\mathbf{h}_n\|^2 \|\mathbf{h}_m\|^2 (2\theta - \theta^2)$, and $m_3 = \rho (\|\mathbf{h}_n\|^2 (1 - \theta) - \|\mathbf{h}_m\|^2)$.

As $\theta > 0$, we solve equation (A.5) with quadratic formula. The discriminant of equation (A.5) can be expressed as

$$\begin{aligned} \Delta &= m_2^2 - 4m_1m_3 \\ &= \left(\rho \|\mathbf{h}_n\|^2 \|\mathbf{h}_m\|^2 (2\theta - \theta^2) - \left(\|\mathbf{h}_n\|^2 (1 - \theta) - \|\mathbf{h}_m\|^2 \right) \right)^2. \end{aligned} \quad (\text{A.6})$$

As $\Delta \geq 0$, equation (A.5) has two roots both of which are real numbers and the roots are given by

$$\begin{aligned} \tilde{\rho}_{n,1}^{(\Delta)} &= \rho, \\ \tilde{\rho}_{n,2}^{(\Delta)} &= \zeta_{n,m}, \end{aligned} \quad (\text{A.7})$$

where $\zeta_{n,m} = (\|\mathbf{h}_n\|^2 (1 - \theta) - \|\mathbf{h}_m\|^2) / \|\mathbf{h}_n\|^2 \|\mathbf{h}_m\|^2 (2\theta - \theta^2)$. When $\Delta = 0$, $\tilde{\rho}_{n,1}^{(\Delta)} = \tilde{\rho}_{n,2}^{(\Delta)}$.

Assuming $m_1 > 0$, we have the following cases.

Case (a). When $\rho \leq \zeta_{n,m}$, i.e., $\tilde{\rho}_{n,1}^{(\Delta)} \leq \tilde{\rho}_{n,2}^{(\Delta)}$, the function $f(\tilde{\rho}_n) \geq 0$ with $\tilde{\rho}_n \in [0, \rho]$, which means that $\log(1 + \gamma_{m,n}^B) \geq \log(1 + \gamma_m^B)$ with $\tilde{\rho}_n \in [0, \rho]$. In this case, $r_m = \log(1 + \gamma_m^B)$ according to (A.2). When $\tilde{\rho}_n = \rho$, the equality in $f(\tilde{\rho}_n) \geq 0$ holds.

Case (b). When $\zeta_{n,m} \leq 0$, i.e., $\tilde{\rho}_{n,2}^{(\Delta)} \leq 0$, the function $f(\tilde{\rho}_n) \leq 0$ with $\tilde{\rho}_n \in [0, \rho]$, which means that $\log(1 + \gamma_{m,n}^B) \leq \log(1 + \gamma_m^B)$ with $\tilde{\rho}_n \in [0, \rho]$. In this case, $r_m = \log(1 + \gamma_{m,n}^B)$ according to (A.2). When $\zeta_{n,m} = 0$ and $\tilde{\rho}_n = 0$ or $\tilde{\rho}_n = \rho$, the equality in $f(\tilde{\rho}_n) \geq 0$ holds.

Case (c). When $0 < \tilde{\rho}_{n,2}^{(\Delta)} = \zeta_{n,m} < \rho$, the function $f(\tilde{\rho}_n) \geq 0$ with $\tilde{\rho}_n \in [0, \zeta_{n,m}]$ and $f(\tilde{\rho}_n) \leq 0$ with $\tilde{\rho}_n \in (\zeta_{n,m}, \rho]$, which means that $\log(1 + \gamma_{m,n}^B) \geq \log(1 + \gamma_m^B)$ with $\tilde{\rho}_n \in [0, \zeta_{n,m}]$ and $\log(1 + \gamma_{m,n}^B) \leq \log(1 + \gamma_m^B)$ with $\tilde{\rho}_n \in (\zeta_{n,m}, \rho]$.

By combining Cases (a), (b), and (c), we can derive (11).

Case (2). $\theta = 0$.

As $\theta = 0$, we can obtain the result

$$\begin{cases} \gamma_{m,n}^B = \frac{\tilde{\rho}_m \|\mathbf{h}_n\|^2}{1 + \tilde{\rho}_n \|\mathbf{h}_n\|^2}, \\ \gamma_m^B = \frac{\tilde{\rho}_m \|\mathbf{h}_m\|^2}{1 + \tilde{\rho}_n \|\mathbf{h}_m\|^2}. \end{cases} \quad (\text{A.8})$$

In this case, because $\|\mathbf{h}_m\| < \|\mathbf{h}_n\|$, $f(\tilde{\rho}_n) > 0$. As a result, we can derive (13). \square

Proof. of Lemma 2.

The upper bound $r_i^{\text{upper}} = \log(1 + (P l^2 / N_0))$ of achievable rate of user i can be derived by allocating all the power P to user i with the other user's rate being zero. Then, we obtain two extreme rate tuples $(r_m^{\text{upper}}, 0)$, $(0, r_n^{\text{upper}})$. As a

result, when $\|\mathbf{h}_m\| = \|\mathbf{h}_n\| = l$, by time-sharing between the two extreme rate tuples, the achievable rate region boundary of MISO broadcast channel can be expressed as [25]

$$\begin{cases} r_n^{\text{TS}} = \lambda \log\left(1 + \frac{Pl^2}{N_0}\right), \\ r_m^{\text{TS}} = (1 - \lambda) \log\left(1 + \frac{Pl^2}{N_0}\right), \end{cases} \quad (\text{A.9})$$

where $\lambda \in [0, 1]$ represents the fraction of the time allocated to user n . We consider the case in which SIC is executed at user n . According to (A.1) and (A.2), when $\|\mathbf{h}_m\| = \|\mathbf{h}_n\| = l$, the achievable rate region boundary can be characterized as

$$\begin{cases} r_n = \log\left(1 + \frac{p_n l^2}{N_0}\right), \\ r_m = \log\left(1 + \frac{p_m l^2 \cos^2 \alpha}{N_0 + p_n l^2}\right), \end{cases} \quad (\text{A.10})$$

where $\forall l \in (0, \infty)$, $\alpha \in [0, \pi)$, $\lambda \in [0, 1]$, $p_m \in [0, P]$, and $p_n \in [0, P]$; we let

$$r_n^{\text{TS}} = r_n. \quad (\text{A.11})$$

We define

$$h(p_m) = r_m^{\text{TS}} - r_m. \quad (\text{A.12})$$

By substituting (A.11) into (A.12), we can get the result

$$h(p_m) = \log\left(\frac{N_0 + Pl^2}{N_0 + Pl^2 - l^2 p_m (1 - \cos^2 \alpha)}\right) \geq 0. \quad (\text{A.13})$$

From (A.13), when $\cos^2 \alpha = 1$, (A.10) is equivalent to (A.9). Consequently, we obtain (14). When SIC is executed at user m , the same result can be achieved. \square

Proof. of Theorem 1.

Case (1). $\zeta_{n,m} \leq 0$.

According to Lemma 1, when $\zeta_{n,m} \leq 0$, the WSR $U(\bar{\rho}_n)$ can be expressed as

$$\begin{aligned} U(\bar{\rho}_n) &= \mu_n \log\left(1 + \bar{\rho}_n \|\mathbf{h}_n\|^2\right) \\ &+ \mu_m \log\left(1 + \frac{\bar{\rho}_m \|\mathbf{h}_m\|^2 (1 - \theta)}{1 + \bar{\rho}_n \|\mathbf{h}_n\|^2}\right), \end{aligned} \quad (\text{A.14})$$

where $\bar{\rho}_m = \rho - \bar{\rho}_n$. In this case, $U(\bar{\rho}_n)$ is a differentiable function of $\bar{\rho}_n$ in the interval $[0, \rho]$. By letting $U'(\bar{\rho}_n) = 0$, we can obtain a linear equation

$$\varepsilon_1 \bar{\rho}_n + \varepsilon_0 = 0, \quad (\text{A.15})$$

where $\varepsilon_1 = \mu_n \|\mathbf{h}_n\|^2 \theta$, $\varepsilon_0 = (1 + \rho \|\mathbf{h}_n\|^2)(\mu_n - \mu_m (1 - \theta)) - \mu_n \|\mathbf{h}_n\|^2 \rho \theta$. As a result, we can obtain (48). In this

case, the global maximum value of $U(\bar{\rho}_n)$ in the interval $[0, \rho]$ can be found by selecting the maximum among $U(0)$, $U(\rho)$, and $U(\bar{\rho}_{n,1})$. Therefore, the solution of Problem (P2) is $\mathbf{p}_{m,n}^* = [p_m^*, p_n^*]$, where $p_n^* = \bar{\rho}_{n,1}^* N_0$, $p_m^* = P - p_n^*$, $\bar{\rho}_{n,1}^* = \operatorname{argmax}_{\bar{\rho}_n} \{U(0), U(\rho), U(\bar{\rho}_{n,1})\}$.

Case (2). $\zeta_{n,m} \geq \rho$.

According to Lemma 1, when $\zeta_{n,m} \geq \rho$, the WSR $U(\bar{\rho}_n)$ can be expressed as

$$\begin{aligned} U(\bar{\rho}_n) &= \mu_n \log\left(1 + \bar{\rho}_n \|\mathbf{h}_n\|^2\right) \\ &+ \mu_m \log\left(1 + \frac{\bar{\rho}_m \|\mathbf{h}_m\|^2}{1 + \bar{\rho}_n \|\mathbf{h}_m\|^2 (1 - \theta)}\right), \end{aligned} \quad (\text{A.16})$$

where $\bar{\rho}_m = \rho - \bar{\rho}_n$. By letting $U'(\bar{\rho}_n) = 0$, we can obtain a quadratic equation

$$\varepsilon_2 \bar{\rho}_n^2 + \varepsilon_1 \bar{\rho}_n + \varepsilon_0 = 0, \quad (\text{A.17})$$

where $\varepsilon_2 = \mu_n \|\mathbf{h}_n\|^2 \|\mathbf{h}_m\|^4 \theta (1 - \theta)$, $\varepsilon_1 = \|\mathbf{h}_n\|^2 (\mu_m \varepsilon_3 - \mu_n \varepsilon_4)$, $\varepsilon_0 = \mu_m \varepsilon_3 - \mu_n \|\mathbf{h}_n\|^2 (1 + \|\mathbf{h}_m\|^2 \rho)$, $\varepsilon_3 = \|\mathbf{h}_m\|^2 (\|\mathbf{h}_m\|^2 \rho (1 - \theta) + 1)$, and $\varepsilon_4 = \|\mathbf{h}_m\|^2 (\|\mathbf{h}_m\|^2 \rho (1 - \theta) - 2\theta + 1)$. By using the quadratic formula to solve equation (A.17), we can obtain (50). Consequently, the solution of Problem (P2) is $\mathbf{p}_{m,n}^* = [p_m^*, p_n^*]$, where $p_n^* = \bar{\rho}_{n,2}^* N_0$, $p_m^* = P - p_n^*$, $\bar{\rho}_{n,2}^* = \operatorname{argmax}_{\bar{\rho}_n} \{U(0), U(\rho), U(\bar{\rho}_{n,2}), U(\bar{\rho}_{n,3})\}$.

Case (3). $0 < \zeta_{n,m} < \rho$.

When $0 \leq \bar{\rho}_n \leq \zeta_{n,m}$, $U(\bar{\rho}_n)$ can be expressed as (A.16) according to Lemma 1. Therefore, the power allocation vector for the global maximum value of $U(\bar{\rho}_n)$ in the interval $[0, \zeta_{n,m}]$ is the one corresponding to the maximum among $U(0)$, $U(\zeta_{n,m})$, $U(\bar{\rho}_{n,2})$, and $U(\bar{\rho}_{n,3})$, where $\bar{\rho}_{n,2} \in [0, \zeta_{n,m}]$, $\bar{\rho}_{n,3} \in [0, \zeta_{n,m}]$.

When $\zeta_{n,m} < \bar{\rho}_n \leq \rho$, $U(\bar{\rho}_n)$ can be expressed as (A.14). Therefore, the power allocation vector for the global maximum value of $U(\bar{\rho}_n)$ in the interval $(\zeta_{n,m}, \rho]$ is the one corresponding to the maximum among $U(\zeta_{n,m})$, $U(\rho)$, and $U(\bar{\rho}_{n,1})$, where $\bar{\rho}_{n,1} \in (\zeta_{n,m}, \rho]$.

We combine the two cases above, in which $0 \leq \bar{\rho}_n \leq \zeta_{n,m}$ and $\zeta_{n,m} < \bar{\rho}_n \leq \rho$. Consequently, in the whole interval $[0, \rho]$, the solution of Problem (P2) is $\mathbf{p}_{m,n}^* = [p_m^*, p_n^*]$, where $p_n^* = \bar{\rho}_{n,1}^* N_0$, $p_m^* = P - p_n^*$, $\bar{\rho}_{n,1}^* = \operatorname{argmax}_{\bar{\rho}_n} \{U(0), U(\rho), U(\zeta_{n,m}), U(\bar{\rho}_{n,1}), U(\bar{\rho}_{n,2}), U(\bar{\rho}_{n,3})\}$, $\bar{\rho}_{n,1} \in (\zeta_{n,m}, \rho]$, $\bar{\rho}_{n,2} \in [0, \zeta_{n,m}]$, and $\bar{\rho}_{n,3} \in [0, \zeta_{n,m}]$. \square

Proof. of Theorem 2.

If $\mu_m \geq \mu_n$, the WSR $\mu_m r_m + \mu_n r_n \leq \mu_m (r_m + r_n)$. According to Lemma 2, $r_m + r_n = \log(1 + \rho l^2)$. Consequently, if $\mu_m \geq \mu_n$, $\mu_m r_m + \mu_n r_n \leq \mu_m \log(1 + \rho l^2)$, i.e., $U(0) = \mu_m \log(1 + \rho l^2)$ is the global maximum value of $U(\bar{\rho}_n)$ in the interval $[0, \rho]$, implying that the solution of

Problem (P2) is $\mathbf{p}_{m,n}^* = [P, 0]$. If $\mu_m < \mu_n$, the proof is similar to that of the above case, in which $\mu_m \geq \mu_n$. \square

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study was supported by the Nonfunded Science and Technology Research Project of Zhanjiang City (grant nos. 2020B01067, 2020B01037, and 2019B01076), the Special Foundation of Talents Project at Lingnan Normal University (grant no. ZL2048), and the Fundamental Research Funds Project for Provincial Universities in Heilongjiang Province (grant no. Hkdxp201804).

References

- [1] N.-T. Le, L.-N. Tran, Q.-D. Vu, and D. Jayalath, "Energy-efficient resource allocation for OFDMA heterogeneous networks," *IEEE Transactions on Communications*, vol. 67, no. 10, pp. 7043–7057, 2019.
- [2] Z. Zhao, M. Jia, and Q. Guo, "Resource allocation algorithm based on bandwidth demands for OFDMA systems with imperfect channel quality information," *Sensors & Transducers*, vol. 167, no. 3, p. 151, 2014.
- [3] J. Zeng, T. Lv, R. P. Liu et al., "Investigation on evolving single-carrier NOMA into multi-carrier NOMA in 5G," *IEEE Access*, vol. 6, pp. 48268–48288, 2018.
- [4] S. 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 Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 721–742, 2016.
- [5] G. Liu, Z. Wang, J. Hu, Z. Ding, and P. Fan, "Cooperative NOMA broadcasting/multicasting for low-latency and high-reliability 5G cellular V2X communications," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 7828–7838, 2019.
- [6] B. Makki, K. Chitti, A. Behravan, and M.-S. Alouini, "A survey of NOMA: current status and open research challenges," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 179–189, 2020.
- [7] T. M. Cover and J. A. Thomas, *Elements of Information Theory [Internet]*, John Wiley & Sons, New York, NY, USA, 2012.
- [8] M. F. Hanif, Z. Ding, T. Ratnarajah, and G. K. Karagiannidis, "A minorization-maximization method for optimizing sum rate in the downlink of non-orthogonal multiple access systems," *IEEE Transactions on Signal Processing*, vol. 64, no. 1, pp. 76–88, 2015.
- [9] C. Chen, W. Cai, X. Cheng, L. Yang, and Y. Jin, "Low complexity beamforming and user selection schemes for 5G MIMO-NOMA systems," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 12, pp. 2708–2722, 2017.
- [10] S. H. Amin, A. H. Mehana, S. S. Soliman, and Y. A. Fahmy, "User capacity in downlink MISO-NOMA systems," in *Proceedings of the 2018 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–7, IEEE, Abu Dhabi, UAE, 2018.
- [11] Q. Sun, S. Han, Z. Xu, S. Wang, I. Chih-Lin, and Z. Pan, "Sum rate optimization for MIMO non-orthogonal multiple access systems," in *Proceedings of the 2015 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 747–752, IEEE, Las Vegas, NV, USA, 2015.
- [12] Q. Zhang, Q. Li, and J. Qin, "Robust beamforming for nonorthogonal multiple-access systems in MISO channels," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 12, pp. 10231–10236, 2016.
- [13] Z. Chen, Z. Ding, P. Xu, and X. Dai, "Optimal precoding for a QoS optimization problem in two-user MISO-NOMA downlink," *IEEE Communications Letters*, vol. 20, no. 6, pp. 1263–1266, 2016.
- [14] Z. Chen, Z. Ding, X. Dai, and G. K. Karagiannidis, "On the application of quasi-degradation to MISO-NOMA downlink," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6174–6189, 2016.
- [15] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 6010–6023, 2015.
- [16] H. V. Nguyen, V. D. Nguyen, O. A. Dobre, D. N. Nguyen, E. Dutkiewicz, and O. S. Shin, "Joint power control and user association for NOMA-based full-duplex systems," *IEEE Transactions on Communications*, vol. 67, no. 11, pp. 8037–8055, 2019.
- [17] T. V. Nguyen, V. D. Nguyen, D. B. da Costa, and B. An, "Hybrid user pairing for spectral and energy efficiencies in multiuser MISO-NOMA networks with SWIPT," *IEEE Transactions on Communications*, vol. 68, no. 8, pp. 4874–4890, 2020.
- [18] E. A. Jorswieck and E. G. Larsson, "Linear precoding in multiple antenna broadcast channels: efficient computation of the achievable rate region," in *Proceedings of the 2008 International ITG Workshop on Smart Antennas*, pp. 21–28, IEEE, Darmstadt, Germany, 2008.
- [19] Z. Zhao and W. Chen, "An adaptive switching method for sum rate maximization in downlink MISO-NOMA systems," in *Proceedings of the GLOBECOM 2017-2017 IEEE Global Communications Conference*, pp. 1–6, IEEE, Singapore, 2017.
- [20] E. A. Jorswieck, E. G. Larsson, and D. Danev, "Complete characterization of the pareto boundary for the MISO interference channel," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 5292–5296, 2008.
- [21] E. A. Jorswieck and E. G. Larsson, "Monotonic optimization framework for the two-user MISO interference channel," *IEEE Transactions on Communications*, vol. 58, no. 7, pp. 2159–2168, 2010.
- [22] X. Shang, B. Chen, and H. V. Poor, "Multiuser MISO interference channels with single-user detection: optimality of beamforming and the achievable rate region," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4255–4273, 2011.
- [23] Z. Ding, Y. Liu, J. Choi et al., "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 185–191, 2017.
- [24] B. Sklar, "Rayleigh fading channels in mobile digital communication systems. I. characterization," *IEEE Communications Magazine*, vol. 35, no. 7, pp. 90–100, 1997.
- [25] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, Cambridge, UK, 2005.
- [26] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 2012.

- [27] S. M. Tseng, Y. F. Chen, and H. H. Fang, "Cross PHY/APP layer user grouping and power allocation for uplink multi-antenna NOMA video communication systems," *IEEE Systems Journal*, vol. 14, no. 3, pp. 3351–3359, 2019.
- [28] H. V. Nguyen, H. M. Kim, G. M. Kang, K. H. Nguyen, V. P. Bui, and O. S. Shin, "A survey on non-orthogonal multiple access: from the perspective of spectral efficiency and energy efficiency," *Energies*, vol. 13, no. 16, p. 4106, 2020.