mathematics

Article

User Grouping, Precoding Design, and Power Allocation for MIMO-NOMA Systems

Byungjo Kim [1] and Jae-Mo Kang [2,*]

[1] Electronics and Telecommunications Research Institute, Daejeon 34129, Republic of Korea
[2] Department of Artificial Intelligence, Kyungpook National University, Daegu 41566, Republic of Korea
* Correspondence: jmkang@knu.ac.kr

Abstract: In this paper, we study user grouping, precoding design, and power allocation for multiple-input multiple-output (MIMO) nonorthogonal multiple access (NOMA) systems. An optimization problem is formulated to the maximize the sum rate under a transmit power constraint at a base station and rate constraints on users, which are nonconvex and combinatorial and thus very challenging to solve . To tackle this problem, we carry out the optimization in two steps. In the first step, exploiting the machine learning technique, we develop an efficient iterative algorithm for user grouping and precoding design. In the second step, we develop a power-allocation scheme in closed form by recasting the original problem into a useful and tractable convex form. The numerical results demonstrate that the proposed joint scheme, including user grouping, precoding design, and power allocation, considerably outperforms the existing schemes in terms of sum rate maximization, which increases the sum-rate up to 8–18%. In addition, the results show the larger the number of antennas or users, the bigger the performance gap, at the cost of less computational complexity.

Keywords: MIMO; machine learning; NOMA; power allocation; precoding design; user clustering; sum rate maximization; power resources

MSC: 94A05

Citation: Kim, B.; Kang, J.-M. User Grouping, Precoding Design, and Power Allocation for MIMO-NOMA Systems. *Mathematics* **2023**, *11*, 995. https://doi.org/10.3390/math11040995

Academic Editor: Lingfei Mo

Received: 13 December 2022
Revised: 10 February 2023
Accepted: 14 February 2023
Published: 15 February 2023

1. Introduction

Nonorthogonal multiple access (NOMA) is a promising technique for 5G and B5G systems due to higher spectral efficiency over the traditional orthogonal multiple access [1–4]. NOMA can achieve superior spectral efficiency based on superposition coding (SC) at the transmitter and successive interference cancellation (SIC) at the receiver [2–4]. Multiple-input multiple-output (MIMO) is another technique to enhance the spectrum efficiency. MIMO has been adopted in many wireless standards and is also a promising technique for 5G systems since it offers additional diversity/multiplexing gains without increasing power and bandwidth. Therefore, combining the MIMO and the NOMA, namely, the MIMO-NOMA, is a natural and synergistic research direction towards achieving high performance in 5G systems [1,2,4,5] and much higher spectral efficiency can be achieved.

For the MIMO-NOMA system, there are several critical issues to be addressed. Specially, studies on MIMO NOMA systems can be mainly classified into three types: user grouping, precoding and power allocation. First, in the MIMO-NOMA system, a base station (BS) can support more users than the number of beams (or antennas) because multiple users can share a single beam [6–8]. Thus, the system performance critically depends on which beam the users are assigned to (i.e., user grouping) [1,2,4,6–12]. Second, to achieve the high performance gain offered by the MIMO, the design of precoding (i.e., the set of beams) is essential [1,2,13]. Third, since the NOMA exploits the power domain to serve multiple users nonorthogonally, a proper power allocation is crucial for the system performance [1,2,4,12,14]. Overall, the user grouping, precoding design, and power allocation are very important issues for the MIMO-NOMA system. To the best of our knowledge, in the literature, the problem of joint-user grouping, precoding design, and power allocation has not been studied for the MIMO-NOMA system. This motivated our work.

2. Related Work

In recent works [6–8], the user grouping for the MIMO-NOMA system was studied. It has been shown that user grouping can enormously improve NOMA performance compared to OMA. In [6], a dynamic user-clustering problem was investigated from a fairness perspective, and three sub-optimal algorithms were proposed for the solution. In [7], the sort-based user paring and power allocation are deployed to reduce the interference and maximize the system capacity in the downlink of MIMO-NOMA system. However, in [6–8], the grouping algorithms were developed in a heuristic manner, and the precoding design was not considered. Thus, the performance of the schemes of [6–8] is not satisfactory. In [6,7,14], several power-allocation schemes were studied for the MIMO-NOMA system. However, the schemes of [6,7,14] are limited only to the case when there are two users in a single group. The extension of [6,7,14] to the general scenario with more than two users and multiple groups is never straightforward. In [11,15–18], the joint design schemes of beamforming and power allocation were studied for the MIMO-NOMA system. In [11], the joint design of beamforming and power allocation is proposed for the MIMO NOMA network in which the users are divided into two groups according to their QoS requirements. In [15,16], the joint design of beamforming and power allocation was studied on in two-user scenarios. In [15], using decomposed optimization, the joint design of power allocation, and beamforming were proposed for MIMO NOMA system in two-user scenarios. A joint design for artificial-noise-aided beamforming and power allocation was studied in [16]. However, joint-user grouping, precoding design, and power allocation in multi-user scenarios are very challenging topics in the MIMO-NOMA system.

In this paper, we study the joint optimization of user grouping, precoding design, and power allocation for the MIMO-NOMA system, which is formulated to maximize the sum rate under the transmit power constraint at the BS and the rate constraints at the users. This problem is combinatorial and nonconvex. Thus, it is generally very difficult to solve. To overcome the challenge, we divide the optimization into two parts: the user grouping and precoding design are carried out in the first part, and the power allocation is performed in the second part. Compared to [4–7,11,14], the main contributions of our work are summarized as follows. First, we develop an effective and high-performing algorithm for the user grouping and precoding design based on the machine learning technique. Second, given the user groups and precoding matrix, we recast the original problem into a useful and tractable convex form; then, we develop a power-allocation scheme, which works for any number of users and groups.

For the sake of clarity, we summarize the main symbols and their descriptions used in this paper in Table 1.

Table 1. Symbols and description.

Symbols	Descriptions
N	Number of users
K	Number of antennas at Base station (BS)
M	Number of antennas at each user
W	Precoding matrix
$\mathcal{N}$	Set of user groups
x	Transmitted signal vector from BS
$\boldsymbol{w}_k$	Beamforming vector for the users in the kth group
$s_{k,n}$	Message intended to the nth user in the kth group

Table 1. *Cont.*

Symbols	Descriptions
$p_{k,n}$	Power assigned to the nth user in the kth group
$\boldsymbol{H}_{k,n}$	MIMO channel matrix between the BS and the nth user in the kth group
$\boldsymbol{v}_{k,n}$	Linear receive filter at the nth user in the kth group
$\boldsymbol{z}_{k,n}$	Additive Gaussian noise vector with zero mean and covariance matrix $\sigma^2 \boldsymbol{I}_M$
$\boldsymbol{R}_{k,n}$	Achievable rate of the nth user in the kth group
$r_{k,n}$	Rate threshold for the nth user in the kth group
P	Maximum transmit power at the BS

Notation: The conjugate transpose of a matrix $\boldsymbol{A}$, the cardinality of a set $\mathcal{A}$ and the Euclidean norm of a vector $\boldsymbol{a}$ are denoted by $\boldsymbol{A}^H$, $|\mathcal{A}|$ and $\|\boldsymbol{a}\|$ respectively.

3. System Model and Problem Formulation

3.1. System Model

We consider the downlink of the MIMO-NOMA system with a BS and N users. The BS is equipped with K antennas, and each user is eqipped with M antennas. There are K user groups, $\mathcal{N}_k \subseteq \mathcal{N}$, $k = 1, \cdots, K$, where $\mathcal{N} \triangleq \{1, \cdots, N\}$ denotes the set of all of the users. It is assumed that $\bigcup_{k=1}^{K} \mathcal{N}_k = \mathcal{N}$ and $\mathcal{N}_k \cap \mathcal{N}_j = \emptyset$, $1 \le k \ne j \le K$.

The transmitted signal vector $\boldsymbol{x} \in \mathbb{C}^{K\times 1}$ from the BS is given by [6,7]

$$\boldsymbol{x} = \boldsymbol{W}\boldsymbol{s} = \sum_{k=1}^{K} \boldsymbol{w}_k \sum_{n\in\mathcal{N}_k} \sqrt{p_{k,n}} s_{k,n} \tag{1}$$

where $\boldsymbol{w}_k \in \mathbb{C}^{K\times 1}$ is the beamforming vector for the users in the kth group and $\boldsymbol{W} \triangleq [\boldsymbol{w}_1, \cdots, \boldsymbol{w}_K] \in \mathbb{C}^{K\times K}$ is the precoding matrix. Additionally, $\boldsymbol{s} \triangleq \left[\sum_{n\in\mathcal{N}_1} \sqrt{p_{1,n}} s_{1,n}, \cdots, \sum_{n\in\mathcal{N}_K} \sqrt{p_{K,n}} s_{K,n}\right]^T \in \mathbb{C}^{K\times 1}$, where $s_{k,n} \in \mathbb{C}$ and $p_{k,n} \ge 0$ denote the message intended to and the power assigned to the nth user in the kth group, respectively, $\forall n \in \mathcal{N}_k$, $\forall k$. It is assumed that $\mathbb{E}[|s_{k,n}|^2] = 1$, $\mathbb{E}[s_{k,n}s_{k,i}^*] = 0$, $\forall n \ne i$, and $\|\boldsymbol{w}_k\| = 1$, $\forall k$.

The received signal $\boldsymbol{y}_{k,n} \in \mathbb{C}^{M\times 1}$ at the nth user in the kth group is given by

$$\boldsymbol{y}_{k,n} = \boldsymbol{H}_{k,n}\boldsymbol{x} + \boldsymbol{z}_{k,n} \tag{2}$$

where $\boldsymbol{H}_{k,n} \in \mathbb{C}^{M\times K}$ is the MIMO channel matrix between the BS and the nth user in the kth group, and $\boldsymbol{z}_{k,n} \in \mathbb{C}^{M\times 1}$ is the additive Gaussian noise vector with zero mean and covariance matrix $\sigma^2 \boldsymbol{I}_M$. The nth user in the kth group uses a linear receive filter $\boldsymbol{v}_{k,n} \in \mathbb{C}^{M\times 1}$ to process its received signal as follows:

$$\begin{aligned} \tilde{y}_{k,n} = {} & \boldsymbol{v}_{k,n}^H \boldsymbol{H}_{k,n} \boldsymbol{w}_k \sum_{i\in\mathcal{N}_k} \sqrt{p_{k,i}} s_{k,i} \\ & + \boldsymbol{v}_{k,n}^H \boldsymbol{H}_{k,n} \sum_{j=1, j\ne k}^{K} \boldsymbol{w}_j \sum_{n\in\mathcal{N}_j} \sqrt{p_{j,n}} s_{j,n} + \boldsymbol{v}_{k,n}^H \boldsymbol{z}_{k,n}. \end{aligned} \tag{3}$$

To eliminate the inter-group interference (i.e., the second summation term in (3)), $\boldsymbol{v}_{k,n}$ is chosen to satisfy $\boldsymbol{v}_{k,n}^H \boldsymbol{H}_{k,n} \boldsymbol{w}_j = \boldsymbol{0}$, $\forall k \ne j$ [6,7]. To ensure the existence of such $\boldsymbol{v}_{k,n}$, it is assumed that $M \le K$ [6,7]. Let $\boldsymbol{U}_{k,n} \in \mathbb{C}^{M\times(M-K+1)}$ denote the matrix formed by the basis vectors of the left null space of $\boldsymbol{H}_{k,n}\boldsymbol{W}_{-k}$, where $\boldsymbol{W}_{-k} \triangleq [\boldsymbol{w}_1, \cdots, \boldsymbol{w}_{k-1}, \boldsymbol{w}_{k+1}, \cdots, \boldsymbol{w}_K] \in \mathbb{C}^{K\times(K-1)}$. Then, we have $\boldsymbol{v}_{k,n} = \boldsymbol{U}_{k,n}\boldsymbol{c}_{k,n}$, $\forall n, k$, where $\boldsymbol{c}_{k,n} \in \mathbb{C}^{(M-K+1)\times 1}$ is a normalized vector. Without a loss of generality, $\boldsymbol{c}_{k,n}$ is selected to maximize the received power as follows [6,7]: $\boldsymbol{c}_{k,n} = \frac{\boldsymbol{U}_{k,n}^H \boldsymbol{H}_{k,n} \boldsymbol{w}_k}{\|\boldsymbol{U}_{k,n}^H \boldsymbol{H}_{k,n} \boldsymbol{w}_k\|}$, $\forall n, k$. Consequently, (3) becomes

$$\tilde{y}_{k,n} = \|\boldsymbol{U}_{k,n}^H \boldsymbol{H}_{k,n} \boldsymbol{w}_k\| \sum_{i \in \mathcal{N}_k} \sqrt{p_{k,i}} s_{k,i} + \tilde{z}_{k,n} \tag{4}$$

where $\tilde{z}_{k,n}$ is the complex Gaussian noise with zero mean and variance σ^2. In the sequel, we will refer to the term $\|\boldsymbol{U}_{k,n}^H \boldsymbol{H}_{k,n} \boldsymbol{w}_k\|$ in (4) as the effective channel gain.

At the users in each group, the successive interference cancellation (SIC) is carried out for decoding. Suppose that in each group the effective channel gains (i.e., $\|\boldsymbol{U}_{k,n}^H \boldsymbol{H}_{k,n} \boldsymbol{w}_k\|$, $\forall n \in \mathcal{N}_k$, for each k) are sorted in ascending order. Then, the nth user in the kth group first decodes $s_{k,i}$, $i < n$, by treating $s_{k,i}$, $i > n$, as interference, and then the decoded $s_{k,i}$, $i < n$ are cancelled from the received signal $\tilde{y}_{k,n}$ to decode its desired message $s_{k,n}$. The achievable rate of the nth user in the kth group is thus given by

$$\begin{aligned} &R_{k,n}(\mathcal{N}_k, \boldsymbol{W}, \boldsymbol{p}_k) \\ &= \log_2\left(1 + \frac{\|\boldsymbol{U}_{k,n}^H \boldsymbol{H}_{k,n} \boldsymbol{w}_k\|^2 p_{k,n}}{\|\boldsymbol{U}_{k,n}^H \boldsymbol{H}_{k,n} \boldsymbol{w}_k\|^2 \sum_{i \in \mathcal{N}_k \setminus \{1,\cdots,n\}} p_{k,i} + \sigma^2}\right) \end{aligned} \tag{5}$$

where $\boldsymbol{p}_k$ denotes the vector composed of $p_{k,n}$, $\forall n \in \mathcal{N}_k$.

3.2. Problem Formulation

In this paper, our goal is to maximize the sum rate under the transmit power constraint at the BS and the rate constraints at the users by jointly optimizing the user groups $\{\mathcal{N}_k\}$, the precoding matrix $\boldsymbol{W}$, and the transmit power $\{\boldsymbol{p}_k\}$ as follows:

$$(\mathrm{P1}): \quad \max_{\{\mathcal{N}_k\}, \boldsymbol{W}, \{\boldsymbol{p}_k\}} \quad \sum_{k=1}^{K} \sum_{n \in \mathcal{N}_k} R_{k,n}(\mathcal{N}_k, \boldsymbol{W}, \boldsymbol{p}_k) \tag{6}$$

$$\text{s.t.} \quad R_{k,n}(\mathcal{N}_k, \boldsymbol{W}, \boldsymbol{p}_k) \geq r_{k,n}, \ \forall n \in \mathcal{N}_k, \ \forall k, \tag{7}$$

$$p_{k,n} \geq 0, \ \forall n \in \mathcal{N}_k, \ \forall k, \ \sum_{k=1}^{K} \sum_{n \in \mathcal{N}_k} p_{k,n} \leq P, \tag{8}$$

$$\|\boldsymbol{w}_k\| = 1, \ \forall k, \ \mathcal{N}_k \subseteq \mathcal{N}, \ \forall k, \ \sum_{k=1}^{K} |\mathcal{N}_k| = N \tag{9}$$

where $r_{k,n}$ is the rate threshold for the nth user in the kth group and P is the maximum transmit power at the BS.

Note that the problem (P1) is nonconvex and combinatorial, and thus, it is very difficult to solve (P1) directly or even numerically. Specifically, in (P1), the optimal user groups $\{\mathcal{N}_k\}$ must be searched over $\frac{(N+K-1)!}{N!(K-1)!} K^N$ possibilities, of which complexity is prohibitively high in practice. For example, there are 8.7617×10^{11} possibilities when $K = 4$ and $N = 15$. Moreover, even when $\{\mathcal{N}_k\}$ are fixed, the problem (P1) is still difficult to solve due to the nonconvexity in both $\{\boldsymbol{w}_k\}$ and $\{\boldsymbol{p}_k\}$. To address the complexity and nonconvexity issues, in the next section, we propose an efficient and effective approach to carry out the joint optimization of (P1).

4. User Grouping, Precoding Design, and Power Allocation

In this section, we carry out the optimization of (P1) in two stages.

4.1. User Grouping and Precoding Design

In the first stage, we carry out the user grouping and precoding design very efficiently by exploiting the machine learning technique [19]. The key idea of the proposed approach is to separate the user grouping and precoding design (i.e., determinations of $\{\mathcal{N}_k\}$ and $\boldsymbol{W}$) from the power allocation. To this end, we develop a clustering algorithm based on a new design criterion, namely, the weighted sum of the effective channel gains:

$$J(\{\mathcal{N}_k\}, \boldsymbol{W}) \triangleq \sum_{k=1}^{K} \sum_{n \in \mathcal{N}_k} \mu_{k,n} \|\boldsymbol{U}_{k,n}^H \boldsymbol{H}_{k,n} \boldsymbol{w}_k\|^2 \tag{10}$$

where $\mu_{k,n} > 0$ is the weight for the nth user in the kth group. Note that this design criterion (i.e., the weighted sum of the effective channel gains) is dependent only on the variables $\{\mathcal{N}_k\}$ and $\boldsymbol{W}$, but independent of the variables $\{\boldsymbol{p}_k\}$; thus, it is very useful to conduct the user grouping and precoding design separately from the power allocation.

The proposed algorithm is summarized in Algorithm 1, in which the user grouping and precoding design are carried out as follows. First, given the precoding matrix $\boldsymbol{W}$, each user is assigned to a group yielding the largest effective channel gain as follows:

$$\begin{aligned} \mathcal{N}_k = \{n \in \mathcal{N} : \ & \mu_{k,n} \|\boldsymbol{U}_{k,n}^H \boldsymbol{H}_{k,n} \boldsymbol{w}_k\|^2 \\ & \geq \mu_{j,n} \|\boldsymbol{U}_{j,n}^H \boldsymbol{H}_{j,n} \boldsymbol{w}_j\|^2, \forall k \neq j\}, \ \forall k. \end{aligned} \tag{11}$$

Second, given the user groups $\{\mathcal{N}_k\}$, the precoding matrix $\boldsymbol{W}$ is designed to maximize the weighted sum $J(\{\mathcal{N}_k\}, \boldsymbol{W})$ of the effective channel gains as follows:

$$\begin{aligned} \boldsymbol{w}_k &= \arg \max_{\|\boldsymbol{w}\|=1} \sum_{n \in \mathcal{N}_k} \mu_{k,n} \|\boldsymbol{U}_{k,n}^H \boldsymbol{H}_{k,n} \boldsymbol{w}\|^2 \\ &= \boldsymbol{v}_k, \ \forall k, \end{aligned} \tag{12}$$

where $\boldsymbol{v}_k$ is the principle eigenvector of the matrix $\left(\sum_{n \in \mathcal{N}_k} \mu_{k,n} \boldsymbol{H}_{k,n}^H \boldsymbol{U}_{k,n} \boldsymbol{U}_{k,n}^H \boldsymbol{H}_{k,n}\right)$ corresponding to the largest eigenvalue. The above procedures are repeated until convergence.

Since the value of $J(\{\mathcal{N}_k\}, \boldsymbol{W})$ increases or remains the same after each step, the proposed algorithm is guaranteed to converge, where the convergent point might be a local optimum. To achieve the (near) global optimal performance, one can select the best result after running the algorithm multiple times with different initializations. The computational complexities in (Step 1) and (Step 2) of Algorithm 1 are given by $\mathcal{O}(NM^3)$ and $\mathcal{O}(K^4)$, respectively. Thus, the total complexity of Algorithm 1 is given by $\mathcal{O}\big(N_{\text{iter}}(K^4 + NM^3)\big)$, where N_{iter} denotes the number of iterations. It is important to note that the complexity of the proposed algorithm is polynomial in both K and N; thus, the complexity reduction by the proposed algorithm is indeed significant in practice, particularly when K or N large.

Algorithm 1 Proposed algorithm for user grouping and precoding design

Initialize $\boldsymbol{W}$.
Repeat the following steps until convergence:
(Step 1): Update $\{\mathcal{N}_k\}$ according to (11).
(Step 2): Update $\boldsymbol{W}$ according to (12).

4.2. Power Allocation

Once $\{\mathcal{N}_k\}$ and $\boldsymbol{W}$ are determined by Algorithm 1, the remaining task is the power allocation (i.e., optimizing $\{\boldsymbol{p}_k\}$), which is performed in the second stage. However, even when $\{\mathcal{N}_k\}$ and $\boldsymbol{W}$ are given, the problem (P1) is still nonconvex in $\{\boldsymbol{p}_k\}$. Fortunately and interestingly, in the following, we can show that given $\{\mathcal{N}_k\}$ and $\boldsymbol{W}$, the problem (P1), can be recast into a useful and tractable convex form by changing the optimization variables.

Theorem 1. *Given the variables $\{\mathcal{N}_k\}$ and $\boldsymbol{W}$, by changing the variables $\{\boldsymbol{p}_k\}$ to $\{\boldsymbol{q}_k\}$, where $\boldsymbol{q}_k \triangleq [q_{k,1}, \cdots, q_{k,N_k}]^T$, $q_{k,n} \triangleq \sum_{i=n}^{N_k} p_{k,i}$, $\forall n, k$, and $N_k \triangleq |\mathcal{N}_k|$, the problem (P1) can be recast into the following convex problem:*

$$
\begin{aligned}
(\text{P2}): \quad & \max_{\{\boldsymbol{q}_k\}} \quad \sum_{k=1}^{K} \sum_{n=1}^{N_k} f_{k,n}(q_{k,n}) && (13)\\
\text{s.t.} \quad & g_{k,n}(q_{k,n}) \geq \gamma_{k,n} g_{k,n}(q_{k,n+1}),\ 1 \leq n < N_k,\ \forall k, && (14)\\
& g_{k,N_k}(q_{k,N_k}) \geq \gamma_{k,N_k} g_{k,N_k}(0),\ \forall k, && (15)\\
& q_{k,n} \geq q_{k,n+1},\ 1 \leq n < N_k,\ q_{k,N_k} \geq 0,\ \forall k, && (16)\\
& \sum_{k=1}^{K} q_{k,1} \leq P && (17)
\end{aligned}
$$

where $f_{k,1}(x) \triangleq \log_2\left(1 + \frac{h_{k,1}x}{\sigma^2}\right)$ and $f_{k,n}(x) \triangleq \log_2\left(1 + \frac{(h_{k,n} - h_{k,n-1})x}{h_{k,n-1}x + \sigma^2}\right)$, $n = 2, \cdots, N_k$. Additionally, $g_{k,n}(x) \triangleq h_{k,n}x + \sigma^2$, $h_{k,n} \triangleq \|\boldsymbol{U}_{k,n}^H \boldsymbol{H}_{k,n} \boldsymbol{w}_k\|^2$, and $\gamma_{k,n} \triangleq 2^{r_{k,n}}$, $\forall k, n$.

Proof. See Appendix A. □

The result of Theorem 1 is very useful in practice because the power allocation can be performed very efficiently. Specifically, the solution to (P2) can be obtained in closed form, which is presented in the following.

Theorem 2. *The optimal solution $\{\boldsymbol{q}_k^\star\}$ to (P2) is given by*

$$
q_{k,1}^\star = \max\left(\frac{1}{\lambda} - \frac{\sigma^2}{h_{k,1}}, \psi_{k,N_k}\right),\ \forall k, \tag{18}
$$

$$
q_{k,n}^\star = \frac{1}{\Gamma_{k,n-1}}(q_{k,1} - \psi_{k,n-1}),\ n = 2, \cdots, N_k,\ \forall k, \tag{19}
$$

where $\psi_{k,n} \triangleq \sum_{i=1}^{n} \frac{\sigma^2}{h_{k,i}}(\gamma_{k,i} - 1)\phi_{k,i-1}$ and $\phi_{k,n} \triangleq \prod_{i=1}^{n} \gamma_{k,i}$, $\forall k, n$. Additionally, the constant $\lambda > 0$ is determined such that $\sum_{k=1}^{K} q_{k,1}^* = P$.

Proof. See Appendix B. □

From the result of Theorem 2, we can obtain the power allocation solution (i.e., the solution of $\boldsymbol{p}$) as follows:

$$
p_{k,n}^\star = q_{k,n}^\star - q_{k,n+1}^\star,\ 1 \leq n < N_k, \tag{20}
$$

$$
p_{k,N_k}^\star = q_{k,N_k}^\star. \tag{21}
$$

From the results of (18)–(21), we can obtain the following useful insights: First, the term $q_{k,1}^*$ in (18) denotes the total power assigned to all of the users in the kth group. Thus, it turns out that the amount of power assigned to all of the users in a group depends on the effective channel gain of the weakest user (i.e., user 1) in that group. Additionally, in each group, all of the users (except for user N_k) are assigned the power such that their rate constraints are satisfied with the equality (i.e., $R_{k,n} = r_{k,n}$, $1 \leq n < N_k$, $\forall k$). The remaining power is assigned to user N_k (i.e., the strongest user with the largest effective channel gain).

5. Experimental Results

In this section, the performance of the proposed scheme is evaluated and is compared to that of the existing schemes of [4,7] through the numerical results. In the simulations, we randomly generate the elements of $\{\boldsymbol{H}_{k,n}\}$ according to the Rayleigh fading model with average power gain of 0 dBW. Additionally, we set $\sigma^2 = 1$ and $R_{k,n} = 0.5$, $\forall n, k$.

In Figure 1, when $K = M = 2$ and $N = 100$, we plot the beamforming vectors $\{\boldsymbol{w}_k : k = 1, 2\}$ determined by Algorithm 1 and the filtered channel vectors $\{\boldsymbol{h}_{k,n} \triangleq \boldsymbol{H}_{k,n}^H \boldsymbol{v}_{k,n} : n = 1, \cdots, 100, k = 1, 2\}$ of the users in order to illustrate the result of the user grouping and precoding design, where the dotted line denotes the unit circle. In Figure 1, for the illustration purpose, the elements of $\{\boldsymbol{w}_k\}$ and $\{\boldsymbol{h}_{k,n}\}$ are restricted to be real-valued (i.e., $\boldsymbol{w}_k, \boldsymbol{h}_{k,n} \in \mathbb{R}^{2\times 1}$, $\forall k, n$). From Figure 1, it can be seen that the users, of which filtered channel vectors are aligned in the similar directions, are grouped into the same group, which accords with the intuition. Specifically, a user, of which $\boldsymbol{h}_{k,n}$ is aligned with $\boldsymbol{w}_1$ (or $\boldsymbol{w}_2$), is grouped into group 1 (or group 2).

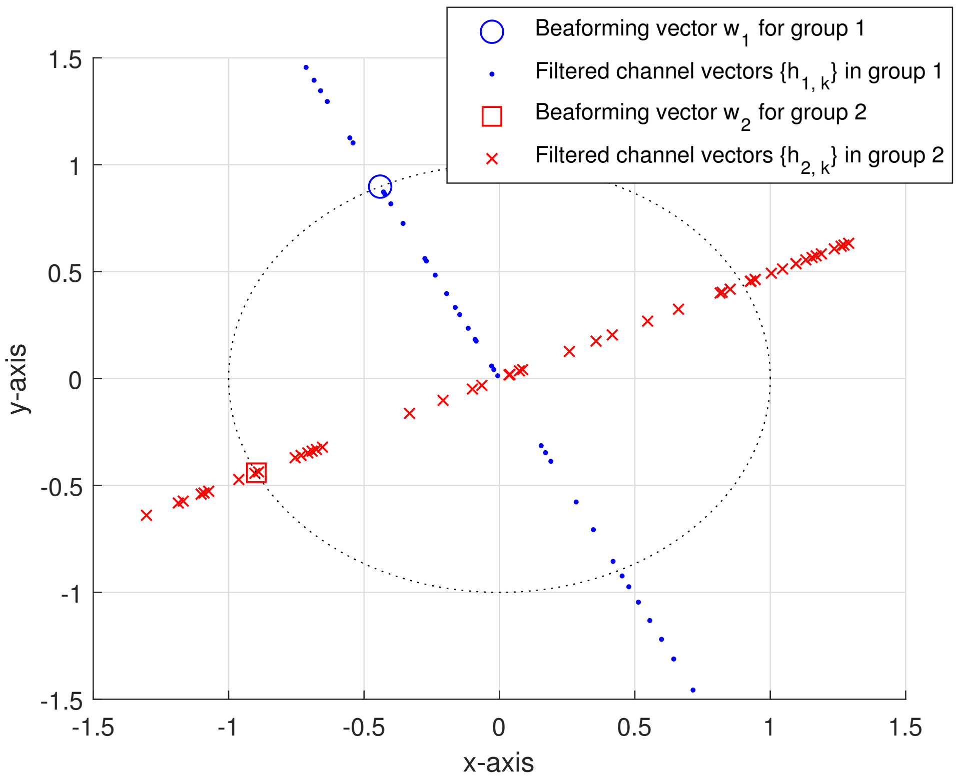

Figure 1. Beamforming vectors $\{\boldsymbol{w}_k\}$ determined by Algorithm 1 and filtered channel vectors $\{\boldsymbol{h}_{k,n}\}$ of the users when $K = M = 2$ and $N = 100$.

In Figure 2, the sum rates of the proposed and existing schemes are shown versus the transmit power P when $K = M \in \{4, 6, 8\}$ and $N \in \{15, 20, 30\}$, where the results are averaged over 10^4 channel realizations. From Figure 2, one can see that the proposed scheme considerably outperforms the existing schemes because the proposed scheme carries out the joint optimization of the user grouping, the precoding matrix, and the power allocation. The performance gain becomes more pronounced when the number of antennas or the number of users increases.

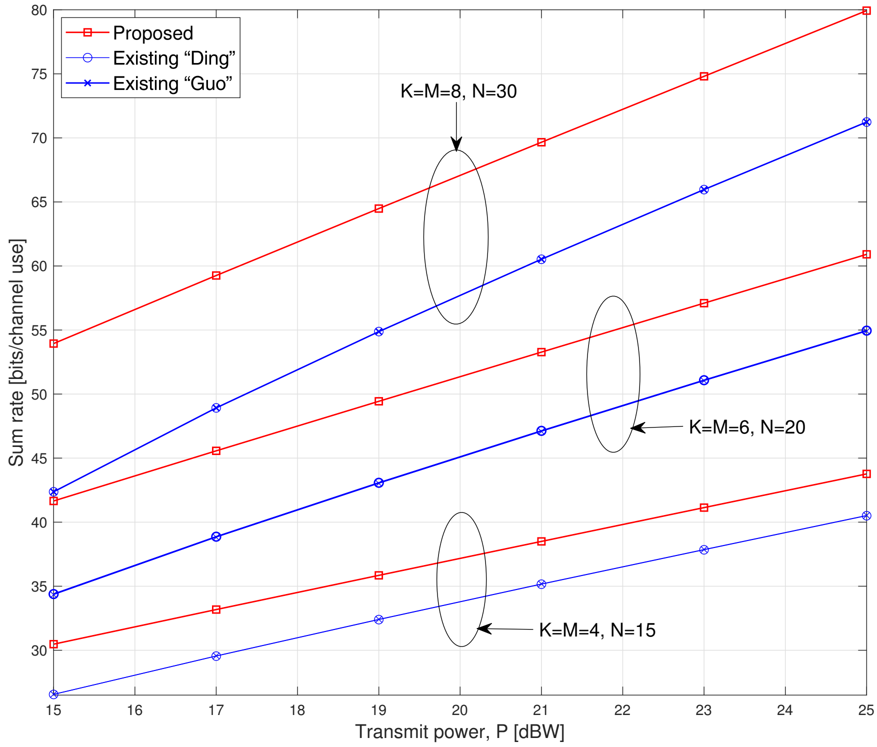

Figure 2. Sum rate versus the transmit power P when $K = M \in \{4, 6, 8\}$ and $N \in \{15, 20, 30\}$ [4,7].

6. Conclusions

In this paper, the problem of joint-user grouping, precoding design, and power allocation was investigated for the MIMO NOMA system to maximize the sum rate with the transmit power constraint at the BS and the rate constraints at the users. To effectively and efficiently carry out this challenging non-convex optimization problem, we first developed the user grouping and precoding design algorithm based on the clustering technique. Then, we developed the power-allocation scheme in closed form, which can derive the optimal solution efficiently. In practice, the proposed scheme effectively reduced the complexity, which is indeed significant, particulary when K or N is large. The numerical results demonstrate that the proposed joint scheme, including user grouping, precoding design, and power allocation, considerably outperforms the existing schemes in terms of sum rate maximization, which increases the sum-rate up to 8–18%. The superior performance of the proposed schemes was demonstrated by the numerical results and outperforms the existing scheme, but it is much more computationally efficient.

Author Contributions: Conceptualization, B.K. and J.-M.K.; Methodology, B.K. and J.-M.K.; Data curation, B.K. and J.-M.K.; Software, B.K. and J.-M.K.; Visualization, B.K.; Validation, B.K. and J.-M.K.; Formal Analysis, B.K. and J.-M.K.; Writing-Original Draft Preparation, B.K.; Writing-Review and Editing, B.K. and J.-M.K.; Supervision, J.-M.K.; Funding Acquisition, J.-M.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2022R1A4A1033830), in part by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2023-2020-0-01808) supervised by the IITP (Institute of Information & Communications Technology Planning & Evaluation) and in part by the National Research Foundation (NRF), Korea, under project BK21 FOUR.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of Theorem 1

Let us define $q_{k,n} \triangleq \sum_{i=n}^{N_k} p_{k,i}$, $\forall n,k$. Then, the rate of (5) can be written as $R_{k,n} = \log_2\left(1 + \frac{h_{k,n} p_{k,n}}{h_{k,n}\sum_{i=n+1}^{N_k} p_{k,i} + \sigma^2}\right) = \log_2\left(\frac{h_{k,n} q_{k,n} + \sigma^2}{h_{k,n} q_{k,n+1} + \sigma^2}\right)$, $1 \le n < N_k$ and $R_{k,N_k} = \log_2\left(1 + \frac{h_{k,N_k} p_{k,N_k}}{\sigma^2}\right) = \log_2\left(\frac{h_{k,N_k} q_{k,N_k} + \sigma^2}{\sigma^2}\right)$, $\forall k$. From this, the objective function of (P1) can be written as in (13), and the rate constraints in (7) can be written as in (14) and (15). Additionally, the power constraints in (8) can be written as in (16) and (17). Thus, given $\{\mathcal{N}_k\}$ and $\boldsymbol{W}$, (P1) can be recast into (P2). Since the objective function in (13) is concave and the constraints in (14) and (15) are linear, (P2) is a convex optimization problem.

Appendix B. Proof of Theorem 2

The objective function of (P2) is increasing in $q_{k,n}$, $1 < n \le N_k$, $\forall k$. Thus, from the constraints in (14), we have $q_{k,n+1} = \frac{q_{k,n}}{\gamma_{k,n}} - \frac{(\gamma_{k,n}-1)\sigma^2}{\gamma_{k,n} h_{k,n}}$, $1 \le n < N_k$, $\forall k$. By recursion, we have the result of (19). Substituting (19) into (P2), we have the following optimization problem: $\max_{\{q_{k,1}\}} \quad \sum_{k=1}^{K} \log_2\left(1 + \frac{h_{k,1} q_{k,1}}{\sigma^2}\right)$ s.t. $q_{k,1} \ge \psi_{k,N_k}$, $\forall k$, $\sum_{k=1}^{K} q_{k,1} \le P$. The solution to this problem can be obtained from the Lagrange duality method as in (18).

References

1. Islam, S.M.R.; Avazov, N.; Dobre, O.A.; Kwak, K.S. Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges. *IEEE Commun. Surv. Tuts.* **2015**, *19*, 721–742. [CrossRef]
2. Ding, Z.; Lei, X.; Karagiannidis, G.K.; Schober, R.; Yuan, J.; Bhargava, V.K. A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends. *IEEE J. Select. Areas Commun.* **2017**, *35*, 2181–2195. [CrossRef]
3. Ding, Z.; Schober, R.; Poor, H.V. A General MIMO Framework for NOMA Downlink and Uplink Transmission Based on Signal Alignment. *IEEE Trans. Wirel. Commun.* **2016**, *15*, 4438–4454. [CrossRef]
4. Ding, Z.; Adachi, F.; Poor, H. The application of MIMO to non-orthogonal multiple access. *IEEE Trans.Wirel. Commun.* **2016**, *15*, 537–552. [CrossRef]
5. Sun, Q.; Han, S.; I, C.L.; Pan, Z. On the ergodic capacity of MIMO NOMA systems. *IEEE Wirel. Commun. Lett.* **2015**, *4*, 405–408. [CrossRef]
6. Liu, Y.; Elkashlan, M.; Ding, Z.; Karagiannidis, G.K. Fairness of user clustering in MIMO non-orthogonal multiple access systems. *IEEE Wirel. Commun. Lett.* **2016**, *20*, 1465–1468. [CrossRef]
7. Guo, J.; Wang, X.; Yang, J.; Zheng, J.; Zhao, B. User pairing and power allocation for downlink non-orthogonal multiple access. In Proceedings of the IEEE Globecom Workshops, Washington, DC, USA, 4–8 December 2016; pp. 1–6.
8. Ghafouri, N.; Movahhedinia, N.; Khayyambashi, M.R. A Novel User Grouping Algorithm for Downlink NOMA. *Wirel. Pers Commun.* **2022**, *126*, 1895–1914. [CrossRef]
9. Kang, J.M.; Kim, I.M. Optimal user grouping for downlink NOMA. *IEEE Wirel. Commun. Lett.* **2018**, *7*, 724–727. [CrossRef]
10. Chen, X.; Gong, F.K.; Li, G.; Zhang, H.; Song, P. User Pairing and Pair Scheduling in Massive MIMO-NOMA Systems. *IEEE Commun. Lett.* **2018**, *22*, 788–791. [CrossRef]
11. Ding, X.; Chen, Y.; Lu, Y. Joint User Ordering, Beamforming and Power Allocation for Downlink MIMO-NOMA Systems. In Proceedings of the 2020 12th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP), Porto, Portugal, 20–22 July 2020; pp. 1–6.
12. Jia, B.; Hu, H.Z.Y.; Xu, T.; Chen, H. Joint User Pairing and Power Allocation in Virtual MIMO Systems. *IEEE Trans. Wirel. Commun.* **2018**, *17*, 3697–3708. [CrossRef]
13. Li, G.; Hong, Z.; Xu, Y.; Huang, Z. Resource allocation for sum-rate maximization in NOMA-based generalized spatial modulation. *Digit. Commun. Netw.* **2022**, *8*, 1077–1084. [CrossRef]
14. Wang, C.-L.; Chen, J.-Y.; Chen, Y.-J. Power allocation for a downlink non-orthogonal multiple access system. *IEEE Wirel. Commun. Lett.* **2016**, *5*, 532–535. [CrossRef]
15. Xiao, Z.; Zhu, L.; Choi, J.; Xia, P.; Xia, X.G. Joint Power Allocation and Beamforming for Non-Orthogonal Multiple Access (NOMA) in 5G Millimeter Wave Communications. *IEEE Trans. Wirel. Commun.* **2018**, *17*, 2961–2974. [CrossRef]
16. Cao, K.; Wang, B.; Ding, H.; Li, T.; Tian, J.; Gong, F. Secure Transmission Designs for NOMA Systems Against Internal and External Eavesdropping. *IEEE Trans. Inf. Forensics Secur.* **2020**, *15*, 2930–2943. [CrossRef]
17. Lin, Z.; Lin, M.; Wang, J.B.; de Cola, T.; Wang, J. Joint Beamforming and Power Allocation for Satellite-Terrestrial Integrated Networks With Non-Orthogonal Multiple Access. *IEEE J. Sel. Top. Signal Process.* **2019**, *13*, 657–670.[CrossRef]

18. Xiu, Y.; Zhao, J.; Sun, W.; Renzo, M.D.; Gui, G.; Zhang, Z.; Wei, N. Reconfigurable Intelligent Surfaces Aided mmWave NOMA: Joint Power Allocation, Phase Shifts, and Hybrid Beamforming Optimization. *IEEE Trans. Wirel. Commun.* **2021**, *20*, 8393–8409. [CrossRef]
19. MacKay, D.J.C. *Information Theory, Inference, and Learning Algorithms*; Cambridge Univ. Press: Cambridge, UK, 2003.