On the Optimality of Power Allocation for NOMA Downlinks with Individual QoS Constraints

Zhaohui Yang, Wei Xu, Cunhua Pan, Yijin Pan and Ming Chen

Abstract—This letter investigates a power allocation problem in a downlink single-input single-output (SISO) non-orthogonal multiple access (NOMA) system. Our goal is to maximize the sum rate of users subject to minimum user rate requirements. We rigorously prove the optimal user decoding order, and show that the sum rate maximization problem is convex which guarantees the globally optimal solution. Further by analyzing the Karush-Kuhn-Tucker (KKT) conditions, we reveal that the optimal power allocation strategy is to allocate additional power to the user with the best channel gain, while other users are allocated with minimum power to maintain their minimum rate requirements. Numerical results validate the performance gain by the proposed NOMA compared to conventional schemes.

Index Terms—Non-orthogonal multiple access (NOMA), power allocation, sum rate maximization.

I. INTRODUCTION

Non-orthogonal multiple access (NOMA) has been recognized as a promising multiple access scheme for future communication systems [1]–[3]. By using superposition coding at transmitter and successive interference cancelation (SIC) at users, NOMA can achieve higher spectral efficiency than conventional orthogonal multiple access (OMA), such as time-division multiple access (TDMA).

NOMA can simultaneously serve multiple users by splitting them in the power domain. In [4], the ergodic sum rate and the outage performance of single-input single-output (SISO) NOMA were analyzed in a cellular downlink scenario with randomly deployed users. The outage probability and diversity order achieved by the cooperative NOMA were studied in [5] where users with better channel conditions have prior information about the messages of other users. Besides, the authors in [6] investigated the impact of power allocation on the fairness of a downlink SISO NOMA system and a new power allocation scheme was proposed in [7] for a downlink SISO NOMA system with two users.

Recently, energy efficiency (EE) optimization becomes attractive especially for 5G networks, e.g., using NOMA serving multiple users. To maximize the EE, the authors in [8] proposed an energy-efficient power allocation strategy which involved the sum rate maximization problem. Although the power allocation was optimized in [8], the authors treated the sum rate maximization problem as a nonconvex multivariate optimization problem via nonlinear programming approaches.

In this paper, we revisit the sum rate maximization problem in the downlink SISO NOMA system. We mathematically prove that it is optimal for each user to decode the messages of users with lower channel gains before decoding its own message. Even with individual user quality of service (QoS) constraints, we succeed in showing that the problem is convex.

II. SYSTEM MODEL AND PROBLEM STATEMENT

Consider a downlink NOMA system with one single-antenna base station (BS) and $M$ single-antenna users. The channel gain between the BS and user $m$ is denoted by $h_m$. Without loss of generality, the channels are sorted as $|h_1|^2 \leq \cdots \leq |h_M|^2$. According to the NOMA principle, the BS simultaneously transmits signal $s$ to all the users. The transmitted signal $s$ can be expressed as

$$s = \sum_{m=1}^{M} \sqrt{p_m}s_m,$$

where $s_m$ and $p_m$ are the message and allocated power for user $m$, respectively.

The observation at user $m$ is given by

$$y_m = \sum_{l=1}^{M} \sqrt{p_m|h_m|^2} y_{ml} + n_m,$$ (2)

where $n_m$ represents the additive zero-mean Gaussian noise with variance $\sigma^2$. With NOMA, successive detection is carried out at users. According to [4], the achievable rate of user $m$ can be given by

$$r_m = \log_2 \left(1 + \frac{|h_m|^2 p_m}{|h_m|^2 \sum_{l=m+1}^{M} p_l + \sigma^2}\right).$$ (3)

Denote $R_m$ as the minimum rate requirement of user $m$. Applying (3), $r_m \geq R_m$ is equivalent to:

$$|h_m|^2 \sum_{l=m+1}^{M} p_l + \sigma^2 \geq 2^{R_m} \left(|h_m|^2 \sum_{l=m+1}^{M} p_l + \sigma^2\right).$$ (4)

Our objective is to optimize the power allocation in order to maximize the sum rate under the total power constraint and individual minimum rate requirements. Mathematically, the sum rate maximization problem can be formulated as

$$\max_{\mathbf{p}} \sum_{m=1}^{M} \log_2 \left(\frac{|h_m|^2 a_m + \sigma^2}{|h_m|^2 a_{m+1} + \sigma^2}\right)$$ (5a)

subject to

$$\sum_{m=1}^{M} p_m \leq P_{\text{max}}$$ (5b)

$$|h_m|^2 a_m + \sigma^2 \geq 2^{R_m} \left(|h_m|^2 a_{m+1} + \sigma^2\right), \quad \forall m$$ (5c)

$$\mathbf{p} \geq 0,$$ (5d)

where $\mathbf{p} = (p_1, \cdots, p_M)^T$, $a_m = \sum_{l=m}^{M} p_l$, $m = 1, \cdots, M$, $a_{M+1} = 0$, and $P_{\text{max}}$ is the maximum transmit power of BS.
III. OPTIMAL POWER ALLOCATION

We first analyze the optimal decoding order for users. Then, we prove that the rate maximization problem in (5) is convex, and obtain the optimal power allocation in closed form.

It is widely assumed that each user should first decode the messages of users with lower channel gains [2]-[8]. This kind of choice is mainly assumed to simplify the successful decoding in SIC, whose optimum in terms of performance is, however, rarely discussed. In the following theorem, this choice is proved to be fortunately also the optimal solution for achieving maximal sum rate with NOMA.

Theorem 1: To maximize the sum rate, the optimal decoding sequence is determined by the channel gain order. Specifically, when the users are sorted in the ascending order of channel gains, i.e., \(|h_1|^2 \leq \cdots \leq |h_M|^2\), the optimal decoding order of user \(m\) is \(\{1, 2, \cdots, m\}\).

Proof: Please refer to Appendix A.

Theorem 2: Problem (5) is a convex problem. There exists at least one feasible solution to (5) if and only if the user rate constraints satisfy the following condition:

\[
\sum_{i=1}^{M} \left( \prod_{j=1}^{i-1} \frac{1}{2R_j} \right) (2R_i - 1) \alpha^2 \leq P_{\text{max}}.
\]

(6)

Proof: Please refer to Appendix B.

Since problem (5) is convex, we can obtain the globally optimal solution by solving the KKT conditions [9], [10]. The Lagrangian penalty of problem (5) can be written by

\[
\mathcal{L}(\mathbf{p}, \alpha, \beta) = \sum_{m=1}^{M} \log_2 \left( \frac{|h_m|^2 a_m + \sigma^2}{|h_m|^2 a_{m+1} + \sigma^2} \right) + \alpha \left( P_{\text{max}} - \sum_{m=1}^{M} p_m \right) + \sum_{m=1}^{M} \beta_m \left( |h_m|^2 a_m + \sigma^2 - 2R_m |h_m|^2 a_{m+1} - 2R_m \sigma^2 \right),
\]

where \(\alpha\) and \(\beta = (\beta_1, \cdots, \beta_M)^T\) are, respectively, the non-negative Lagrange multipliers associated with constraints (5b) and (5c). The KKT conditions of problem (5) are listed as:

\[
\frac{\partial \mathcal{L}}{\partial p_m} = \ln 2 \sum_{k=2}^{m} \left( \frac{|h_k|^2}{|h_k|^2 a_k + \sigma^2} - \frac{|h_{k-1}|^2}{|h_{k-1}|^2 a_k + \sigma^2} \right) + \left( \frac{|h_1|^2}{|h_1|^2 a_1 + \sigma^2} \right) - \alpha + \beta_1 |h_1|^2 + \sum_{k=2}^{m} (\beta_k |h_k|^2 - \beta_{k-1} 2R_{k-1} |h_{k-1}|^2) = 0, \quad \forall m
\]

(7a)

\[
\alpha \left( P_{\text{max}} - \sum_{m=1}^{M} p_m \right) = 0
\]

(7b)

\[
\beta_m (|h_m|^2 a_m + \sigma^2 - 2R_m |h_m|^2 a_{m+1} - 2R_m \sigma^2) = 0, \quad \forall m
\]

(7c)

First, from (7a) with \(m = 1\), we have

\[
\alpha = \frac{|h_1|^2}{(\ln 2)(|h_1|^2 a_1 + \sigma^2)^2} + \beta_1 |h_1|^2 > 0.
\]

(8)

Hence, further considering (7b), we can obtain \(\sum_{m=1}^{M} p_m = P_{\text{max}}\), equivalently

\[
a_1 = P_{\text{max}}.
\]

Here, we assume that \(|h_1|^2 < \cdots < |h_M|^2\). The special case with \(|h_{m-1}|^2 = |h_m|^2\) is considered later. Using (7a), we obtain

\[
\frac{\partial \mathcal{L}}{\partial p_m} - \frac{\partial \mathcal{L}}{\partial p_{m-1}} = \frac{1}{\ln 2} \left( \frac{|h_m|^2}{|h_m|^2 a_m + \sigma^2} - \frac{|h_{m-1}|^2}{|h_{m-1}|^2 a_m + \sigma^2} \right) + \beta_m |h_m|^2 - \beta_{m-1} 2R_{m-1} |h_{m-1}|^2 = 0, \quad m = 2, \cdots, M.
\]

(10)

Considering \(|h_{m-1}|^2 < |h_m|^2\), we have

\[
\beta_{m-1} 2R_{m-1} |h_{m-1}|^2 - \beta_m |h_m|^2 = 0.
\]

Proof: Please refer to Appendix B.

Since \(\beta_m \geq 0\) for all \(m = 1, \cdots, M\), we have \(\beta_{m-1} 2R_{m-1} |h_{m-1}|^2 > \beta_m |h_m|^2 \geq 0\). Thus, we can obtain \(\beta_{M-1} > 0, \ldots, \beta_1 > 0\), owing to the fact that the probability of \(h_m = 0\) is zero for all \(m = 1, \cdots, M\). If \(\beta_M > 0\), we assure that the constraints in (7c) are satisfied via \(|h_m|^2 a_m + \sigma^2 - 2R_m |h_m|^2 a_{m+1} - 2R_m \sigma^2 = 0\) for all \(m = 1, \cdots, M\), which implies that the minimum rate constraints (5c) hold with equality for all users. Thus, the optimal value of problem (5) is \(\sum_{m=1}^{M} R_m\), and the rate maximization problem (5b) becomes finding the linear feasible set, which can be effectively solved. Since problem (5) can be easily solved when \(\beta_M > 0\), we consider the case \(\beta_M = 0\) in the following.

Owing to that minimum rate constraints (5c) hold with equality for all \(m < M\), we find that the additional power is allocated to the user with the highest channel gain and other users are allocated with the minimum transmission power to meet the minimum rate requirements. Now, it remains optimal to solve constraints (5c) with equality for \(m < M\). We have

\[
a_{m+1} = \frac{a_m}{2R_m} - \frac{(2R_m - 1) \sigma^2}{2R_m |h_m|^2}, \quad \forall 1 \leq m < M.
\]

(12)

Using (9), it gives

\[
ap_m = \frac{\sum_{i=1}^{m} (\frac{R_i - 1) \sigma^2}{|h_i|^2 2R_i}, \quad \forall 1 \leq m < M.\]

Recalling that \(a_m = \sum_{i=m}^{M} p_i\), we can obtain

\[
p_m = \begin{cases} a_m - a_{m+1} & \text{if } m < M \\ a_m & \text{if } m = M. \end{cases}
\]

(14)

By inserting (13) into (14), we have

\[
p_{\text{max}} = \begin{cases} \frac{\sum_{i=1}^{m} (\frac{R_i - 1) \sigma^2}{|h_i|^2 2R_i}}{\prod_{i=1}^{m} |h_i|^2 2R_i} & \text{if } m < M \\ \frac{\sum_{i=1}^{M} (\frac{R_i - 1) \sigma^2}{|h_i|^2 2R_i}}{\prod_{i=1}^{M} |h_i|^2 2R_i} & \text{if } m = M. \end{cases}
\]

(15)

Substituting (15) into objective function (5a), we can obtain the optimal sum rate of problem (5) in closed form as

\[
\log_2 \left( 1 + \frac{P_{\text{max}} |h_M|^2}{\prod_{i=1}^{M-1} 2R_i} \left( \sum_{i=1}^{M-1} \frac{|h_i|^2 2R_i}{|h_i|^2 2R_i} \right) + \sum_{m=1}^{M-1} R_m \right),
\]

where the first term is the rate of user \(M\) with largest channel gain and the second term is the sum rate of all other users.

Now, we consider the special case with equal channel gains remained to be discussed. Assume that there are two users with the same channel gain, i.e., \(|h_m|^2 = |h_{m+1}|^2\). Under this case, we can define a new user, \(m'\), with \(R_{m'} = R_{m+1} + R_m\), and then calculate the optimal power allocation strategy, i.e., \(p_1, \cdots, p_{m-2}, p_{m'}, p_{m+1}, \cdots, p_M\) for users \(1, \cdots, m-2, m', m+1, \cdots, M\) according to (15). Based on (3), we have
Further from (17), we have
\[ R \sim \text{standard deviation of shadow fading is} \]
the propagation environment, the large-scale path loss is
user
optimal power for user
arbitrarily allocate power
that it is optimal for each user to decode the messages of users
maximizing the sum rate of a SISO NOMA system. We prove
rate demand transmission.

In summary, we conclude that the globally optimal power
allocation for problem (5) is determined in closed form in (15),
and the corresponding optimal sum rate is given in (16).

IV. NUMERICAL RESULTS

We consider that there are 10 users uniformly distributed in a square area of size 300 m \( \times \) 300 m. In modeling
the propagation environment, the large-scale path loss is
\[ L(d) = 37 + 30 \log(d), \]
where the unit of \( d \) is meter, and the standard deviation of shadow fading is 4 dB. The noise power
is assumed to be -104 dBm. We consider equal minimum rate
requirement, i.e., \( R_1 = \cdots = R_M = R_{\text{min}} \).

The proposed optimal power allocation scheme for rate
maximization of NOMA is labeled as ‘OPA-NOMA’. We compare the proposed algorithm with the fairness power allocation scheme of NOMA [6], which is labeled as ‘FPA-NOMA’. For the comparison between NOMA and conventional OMA, we use a TDMA system [11, Sec. 5.3] as a baseline. The optimal power allocation of TDMA system is labeled as ‘OPA-TDMA’.

In Fig. 1, we investigate the sum rate versus minimum rate requirement. For both OPA-NOMA and OPDA-TDMA, we find that the sum rate decreases with the minimum rate requirement due to the fact that the increase of minimum rate requirement requires the BS to allocate more power to the users with worse channel gains. For the FPA-NOMA, the sum rate remains the same with the minimum rate requirement because that the rates for all users are set the same to ensure user fairness [6]. The proposed OPA-NOMA outperforms the FPA-NOMA in terms of sum rate especially when the minimum rate requirement is small. Moreover, the sum rate approaches zero faster for the OPA-TDMA. This is because the transmission power of the BS is not large enough for satisfying high rate demand by TDMA, which implies that NOMA is more suitable for high rate demand transmission.

V. CONCLUSION

In this paper, we have investigated the power allocation of maximizing the sum rate of a SISO NOMA system. We prove
that it is optimal for each user to decode the messages of users
with poorer channel gains first. We also prove that this sum
rate maximization problem is a convex problem, and we have obtained the optimal power allocation in closed form.

APPENDIX A

Before proving Theorem 1, we present the following lemma.

Lemma 1: The optimal decoding order of user 1 is \{1\}.

Proof: There is an arbitrary strategy \{D, p\}, where \( D = \{D_1, \cdots, D_M\}, D_m \) is the decoding order of user \( m \). For the sake of clear exposition, we temporarily first consider the special case that there are only two users, i.e., \( M = 2 \). Assume that \( D_1 \neq \{1\}, \) i.e., \( D_1 = \{2, 1\} \). Denote a new decoding order set \( D' = \{D'_1, D'_2\}, \) where \( D'_1 = \{1\}, D'_2 = \{1, 2\} \).

In the following, we show that the sum rate with strategy \( \{D', p\} \) is no less than the sum rate with strategy \( \{D, p\} \), which indicates that arranging decoding order strategy \( D \) would always increase the sum rate with given power \( p \).

With strategy \( \{D, p\} \), the achievable rate of user 1 to detect the message of user 2 and user 1 can be expressed as
\[ r_{12} = \log_2 \left( 1 + \frac{|h_1|^2 p_2}{|h_1|^2 p_1 + \sigma^2} \right), \quad r_{11} = \log_2 \left( 1 + \frac{|h_1|^2 p_1}{\sigma^2} \right), \]
respectively. Then, we have
\[ \sum_{k \in D_1} r_{1k} = \log_2 \left( \left| h_1 \right|^2 \left( p_1 + p_2 \right) + \sigma^2 \right) \cdot \left| h_2 \right|^2. \] (18)

With strategy \( \{D', p\} \), the achievable rate of user \( k \in D_1 \) to detect the message of user \( j \leq k \) is
\[ r'_{kj} = \log_2 \left( 1 + \frac{|h_k|^2 p_j}{\left| h_k \right|^2 \sum_{l=j+1}^M p_l + \sigma^2} \right). \] (19)
Since \( |h_1|^2 \leq |h_2|^2 \), we can obtain
\[ r'_{11} = \log_2 \left( 1 + \frac{p_1}{p_2 + \frac{\sigma^2}{|h_1|^2}} \right) \leq \log_2 \left( 1 + \frac{p_1}{p_2 + \frac{\sigma^2}{|h_2|^2}} \right) = r_{21}. \] (20)

Denote the achievable rate of user \( m \) with strategy \( \{D, p\} \) and \( \{D', p\} \) as \( r_m \) and \( r'_m \), respectively. Then, to guarantee the successful decoding order, we have
\[ r_m = \min_{l \in \{s|m \in D_s\}} r_{lm}, \quad r'_m = \min_{l \in \{s|m \in D'_s\}} r'_{lm} \] (21)
for \( m = 1, 2 \). Based on (18)-(21), we have
\[ r'_1 + r'_2 = r'_{11} + r'_{22} = \log_2 \left( 1 + \frac{p_1}{p_2 + \frac{\sigma^2}{|h_1|^2}} \right) + \log_2 \left( 1 + \frac{p_2}{\frac{\sigma^2}{|h_2|^2}} \right) \geq \log_2 \left( \frac{|h_1|^2 \left( p_1 + p_2 \right) + \sigma^2}{\sigma^2} \right) = r_{12} + r_{11} \geq r_1 + r_2. \] (22)
where (a) follows from the fact that \( |h_1|^2 \leq |h_2|^2 \). From (22), Lemma 1 can be proved for the special case with two users.
The philosophy of proving the Lemma 1 for generally multiple users, i.e., $M > 2$, is analogous to that of $M = 2$ exemplified above. Specially, for $M > 2$, we assume that user 1 needs to decode $n \geq 1$ users before decoding its own message, and the decoding order of user 1 is $D_1 = \{i_1, i_2, \ldots, i_{n+1}\}$, where $i_{n+1} = 1$. Let set $\{i_1, i_2, \ldots, i_{n+1}\}$ rearrange in an increasing order $\{j_1, j_2, \ldots, j_{n+1}\}$ where $j_1 = 1$ and $j_1 < j_2 < \cdots < j_{n+1}$. Denote $D' = \{D'_1, D'_2, \ldots, D'_M\}$ where $D'_k = \{j_1, j_2, \ldots, j_k\}$ for $k = 1, \ldots, n+1$, and $D'_l = D_l$ for all $l \in \{1, \ldots, M\} \setminus D_1$. By using the same method in (18)-(22), we can show that the sum rate with strategy $\{D', p\}$ is no less than the sum rate with strategy $\{D, p\}$. 

According to Lemma 1, we have proven that the optimal decoding order of user 1 is $\{1\}$. Based on (20) and (21), we can observe that the sum rate of the system is non-decreasing if user $k > 1$ can decode the message of user 1. Hence, we can prove that the optimal decoding order of user 2 is $\{1, 2\}$ by using the same method in (18)-(22). With the recursion method, we can prove that the optimal decoding order of user $m$ is $\{1, \ldots, m\}$.

**APPENDIX B**

We first prove that problem (5) is convex. Since the constraints of problem (5) are linear, we only need to check that the objective function (5a) is concave. In the following, we assume that the objective function (5a) is concave. As a result, problem (5) is convex.

Then, we prove the feasibility condition for (5). To prove this, we denote

$$Q = \min_{P \geq 0} \sum_{m=1}^{M} p_m \quad \text{s.t.} \quad \sum_{m=1}^{M} |h_m|^2 a_m + \sigma^2 \geq 2R_m (|h_m|^2 a_{m+1} + \sigma^2), \forall m.$$  

From (5b), (5c), and (5d), we can find that problem (5) is feasible if and only if $Q \leq P_{\max}$. If there exists $m$ such that $\sum_{m=1}^{M} |h_m|^2 a_m + \sigma^2 > 2R_m (|h_m|^2 a_{m+1} + \sigma^2)$, we can further strictly decrease $a_m$ while keeping the inequality still valid. Then, we can also sequentially decrease $a_{m-1}, \ldots, a_1$. Since $Q = a_1$, we observe that constraints (5c) hold with equality for all $m$, as otherwise $Q$ can be further improved. Setting constraints (5c) with equality yields $a_{m+1} = a_m - \frac{2R_m - (\frac{2}{|h_m|^2} - 1)\sigma^2}{2R_m |h_m|^2}$, \quad \forall 1 \leq m \leq M. \quad (27)$

Owing to that $a_{M+1} = 0$, we can obtain $a_M = \frac{2R_1 - (\frac{2}{|h_1|^2} - 1)\sigma^2}{2R_1 |h_1|^2}$. By using the recursion method, we have $Q = \sum_{m=1}^{M} p_m = a_1 = \sum_{m=1}^{M} \frac{|h_m|^2}{|h_1|^2 (\frac{2}{|h_m|^2} - 1)\sigma^2}$. Hence, the feasibility condition for (5) is achieved as (6).

**REFERENCES**


