Joint Rate and Power Allocation for NOMA With Statistical CSI

Jinho Choi, *Senior Member, IEEE*

***Abstract*— Nonorthogonal multiple access (NOMA) can effectively improve the spectral efficiency by exploiting the power difference and employing successive interference cancellation (CSI) at receivers. In NOMA, the power allocation is crucial and can be performed with known CSI. It is also possible to carry out the power allocation with statistical CSI to meet target outage probabilities for given transmission rates. In this paper, we consider joint rate and power allocation to minimize the total transmission power with individual throughput constraints in NOMA when statistical CSI is available. An approach to find the optimal solution to the joint rate and power allocation problem is proposed. Based on simulation results, it is shown that NOMA can have a much lower total transmission power than OMA for given target throughput of each user.**

***Index Terms*— Non-orthogonal multiple access, rate and power allocation, throughput.**

I. Introduction

DOWNLINK nonorthogonal multiple access (NOMA) has been extensively studied in [1]–[4] as it can improve the spectral efficiency compared to orthogonal multiple access (OMA). NOMA is to exploit the power difference between multiple users when they share a radio resource block, and the successive interference cancellation (SIC) plays a crucial role in mitigating multiple access interference (MAI) together with power allocation. In [5], practical NOMA schemes, called multiuser superposition transmission (MUST) schemes, are considered for downlink transmissions (with two users). Due to SIC, the receiver's complexity can increase at users. Thus, NOMA without SIC in [6] would be helpful to decrease the receiver's complexity at users.

In [3], NOMA is employed for coordinated multipoint (CoMP) downlink in order to support a cell-edge user without degrading the spectral efficiency. In addition, in [7], an opportunistic base station (BS) or access point (AP) selection is studied for CoMP with NOMA to improve the spectral efficiency. The performance comparison with multiuser diversity schemes (e.g., the opportunistic user selection scheme in [8]) is also an important issue in terms of the tradeoff between spectral efficiency and fairness. To address this issue, in [9], NOMA and multiuser diversity schemes are compared when a proportional fairness scheduler is employed. In [10], power allocation for NOMA is studied from a fairness point of view.

NOMA is extended to multiple-input multiple-output (MIMO) systems in [4] and [11] and capacity analysis of MIMO-NOMA can be found in [12]. For MIMO-NOMA, power allocation as well as precoding are studied for small packet transmissions in [13].

Since NOMA is to exploit the power difference, the power allocation based on known channel state information (CSI) plays a key role in providing certain guaranteed performances. For example, in [12] and [14], instantaneous CSI is assumed to be known at a transmitter. In [15], a suboptimal power control policy to maximize the effective capacity with delay constraints in NOMA is proposed with known CSI. In [13] and [16], however, limited feedback of CSI is considered for a more realistic environment in NOMA. In particular, in [16], the power allocation for NOMA subject to short and long-term power constraints is studied with one-bit CSI feedback.

It is possible to perform the power allocation for NOMA with statistical CSI. In this case, the outage probability is usually employed as a performance measure as in [10] and [17]–[19]. Under fading environments, the distance between a user and a transmitter (i.e., BS for downlink NOMA) affects the statistical properties of CSI. In particular, for Rayleigh fading channels, the second order statistics that are decided by the distance are sufficient to characterize the statistical properties of CSI (or the distribution of fading coefficient). Thus, the BS is able to have statistical CSI of users without any explicitly feedback of CSI as the received signals from users can be used to estimate the distances [20].

In this paper, we study NOMA with statistical CSI as in [10] and [17]–[19]. However, instead of outage probability for a performance measure, we consider the throughput that is the product of transmission rate and the probability of successful decoding. The throughput has been used as a performance measure in various wireless systems [21]–[23]. When short coded blocks or packets are transmitted over fading channels in conjunction with automatic repeat request (ARQ) protocols [24], [25], the throughput becomes the average successful transmission rate. In general, to maximize the throughput, we need to jointly decide (code) rate and (transmission) power. This implies that joint rate and power allocation is to be performed when the throughput is considered in NOMA. This results in different problem formulations from those in [10] and [17]–[19], where only power allocation is considered (for given rates). In particular, in [18], for fixed rates, the

Manuscript received March 26, 2017; revised May 16, 2017 and June 20, 2017; accepted June 21, 2017. Date of publication June 27, 2017; date of current version October 16, 2017. This work was supported by the "Climate Technology Development and Application" research project (K07732) through a grant provided by GIST in 2017. The associate editor coordinating the review of this paper and approving it for publication was S. Durrani.

The author is with the School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju 61005, South Korea (e-mail: jchoi0114@gist.ac.kr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2017.2720176

TABLE I

TABLE OF SYMBOLS

K:	the number of users
P_k:	the allocated power to user k
R_k:	the transmission (or code) rate of the signal to user k
$C_{l;k}$:	the achievable rate for the signal to user l at user k with SIC, where $l \geq k$
h_k:	the channel coefficient from the BS to user k
α_k:	the channel (power) gain from the BS to user k
$\bar{\alpha}_k$:	the average channel (power) gain from the BS to user k
η_k:	the throughput to user k
$\bar{\eta}_k$:	the target throughput to user k
d_k:	the distance between the BS and the kth nearest user

power allocation problems are formulated to minimize the total transmission power with outage probability constraints. In this case, as the rates are fixed, the throughput performances may not be fully optimized. To the best of our knowledge, joint rate and power allocation is not studied yet in NOMA with throughput.

The main contribution of the paper is to provide the optimal solution to the joint rate and power allocation problem that minimizes the total transmission power with throughput constraints for downlink NOMA. From this, NOMA can guarantee certain throughputs of users with the minimum total transmission power when only statistical CSI is available at the BS.

The rest of the paper is organized as follow. In Section II, we present the system model for downlink NOMA and discuss power allocation problems with known CSI. In Section III, a joint rate and power allocation problem is formulated to minimize the total transmission power subject to throughput constraints for NOMA when only statistical CSI is given at the BS. We propose an approach for joint rate and power allocation to find the optimal solution in Section IV. We present numerical results in Section V. The paper is concluded in Section VI with some remarks.

Notation: Matrices and vectors are denoted by upper- and lower-case boldface letters, respectively. $\mathbb{E}[\cdot]$ and $\text{Var}(\cdot)$ denote the statistical expectation and variance, respectively. $\mathcal{CN}(\mathbf{a}, \mathbf{R})$ represents the distribution of circularly symmetric complex Gaussian (CSCG) random vector with mean vector $\mathbf{a}$ and covariance matrix $\mathbf{R}$.

For convenience, we present the list of symbols used in the paper in Table I.

II. SYSTEM MODEL

In this section, we present a system model consisting of a BS and multiple users for downlink NOMA. Throughout the paper, we assume that the BS and users are equipped with single antennas.

Suppose that there are K users in the same resource block for downlink. Let $s_{k,t}$ and h_k denote the data symbol at time t and channel coefficient from the BS to user k, respectively. The block of data symbols, $[s_{k,0} \ \ldots \ s_{k,T-1}]^{\mathrm{T}}$, where T is the length of data block, is assumed to be a codeword of a capacity-achieving code. Furthermore, T is shorter than the coherence time so that h_k remains unchanged for the duration of a block transmission. Suppose that superposition coding [26] is employed for NOMA and the signal to be transmitted by the BS is $\sum_{k=1}^{K} s_{k,t}$. Then, at user k, the received signal is given by

$$y_{k,t} = h_k \sum_{m=1}^{K} s_{m,t} + n_{k,t}, \quad t = 0, \ldots, T-1, \tag{1}$$

where $n_{k,t} \sim \mathcal{CN}(0, 1)$ is the independent background noise (here, the variance of $n_{k,t}$ is normalized for convenience). Let $\alpha_k = |h_k|^2$ and $P_k = \mathbb{E}[|s_{k,t}|^2]$ (with $\mathbb{E}[s_{k,t}] = 0$). Then, α_k becomes the channel (power) gain from the BS to user k and P_k becomes the transmission power allocated to $s_{k,t}$.

Throughout the paper, we assume capacity-achieving codes and no instantaneous CSI at the BS. However, we assume that statistical CSI is available. That is, the probability distribution functions (pdfs) of α_k's are known at the BS, while their values or realizations (i.e., instantaneous CSI) are not available. Under this assumption, in this paper, we consider the throughput that is the product of the transmission rate and the probability of successful transmission, which is widely used as a performance metric over fading channels [21]–[23]. In particular, if the length of codeword is short and less than the coherence time, the throughput might be a reasonable performance measure in conjunction with ARQ protocols [22].

Let R_k denote the transmission rate of $s_{k,t}$ and $C_{l;k}$ denote the achievable rate for the signal to user l at user k with SIC, where $l \geq k$. Then, it can be shown that

$$C_{l;k} = \log_2 \left(1 + \frac{\alpha_k P_l}{\alpha_k \sum_{m=1}^{l-1} P_m + 1} \right). \tag{2}$$

For illustrative purposes, we consider the case of $K = 2$. Suppose that the transmission rates, R_k, and powers, P_k, are given. At user 2, the probability of successful decoding is given by

$$\mathbb{P}_2 = \Pr(C_{2;2} > R_2) = \Pr\left(\alpha_2 > \frac{\tau_2}{P_2 - \tau_2 P_1} \right), \tag{3}$$

where $\tau_k = 2^{R_k} - 1$, if $P_2 > \tau_2 P_1$. Note that if $P_2 \leq \tau_2 P_1$, $\mathbb{P}_2 = 0$. At user 1, the probability of successful decoding is slightly different from (3) as the signal to user 2 has to be decoded for SIC as well, which is given by

$$\begin{aligned} \mathbb{P}_1 &= \Pr(C_{2;1} > R_2, C_{1;1} > R_1) \\ &= \Pr\left(\alpha_1 > \max\left\{ \frac{\tau_2}{P_2 - \tau_2 P_1}, \frac{\tau_1}{P_1} \right\} \right). \end{aligned} \tag{4}$$

Then, the average transmission rate or throughput of each user becomes

$$\eta_k = R_k \mathbb{P}_k. \tag{5}$$

That is, the throughput is the average (successful) transmission rate. In general, if R_k is too high, the probability of successful decoding can be low, which results in a low throughput. On the other hand, if R_k is too low, although a high probability of successful decoding can be achieved, the throughput cannot be high due to a low transmission rate, R_k. Thus, the determination of R_k (together with the transmit power, P_k) plays a crucial role in the throughput maximization, which will be studied in the next section.

III. Problem Formulation

In this section, we consider a transmit power maximization problem to decide the transmission rates and power with target throughputs.

A. Minimum Total Power Minimization for NOMA

Various throughput maximization problems with η_k can be considered without instantaneous CSI, $\{\alpha_k\}$, as long as the pdf of α_k (i.e., statistical CSI) is available. Throughout the paper, we consider the following assumption for fading channels (which is widely used, e.g., [17], [18]).

A) The channel gain, α_k, is independent and factorized into the large and small scale fading terms as $\alpha_k = \bar{\alpha}_k \upsilon_k$, where $\bar{\alpha}_k = \mathbb{E}[\alpha_k] = \mathrm{Var}(h_k)$ and υ_k is an exponential random variable with $\mathbb{E}[\upsilon_k] = 1$. That is, independent Rayleigh fading is assumed.

For example, let us consider the case of $K = 2$. Under **A**, from (3), since α_2 is an exponential random variable, the throughput of user 2 becomes

$$\begin{aligned}\eta_2 &= R_2 \Pr\left(\alpha_2 > \frac{\tau_2}{P_2 - \tau_2 P_1}\right)\\ &= R_2 \exp\left(-\frac{\tau_2}{\bar{\alpha}_2(P_2 - \tau_2 P_1)}\right)\\ &= \log_2(1+\tau_2)\exp\left(-\frac{\tau_2}{\bar{\alpha}_2(P_2 - \tau_2 P_1)}\right). \end{aligned} \tag{6}$$

For given P_1 and P_2, τ_2 or R_2 that maximizes η_2 can be found, which is denoted by τ_2^* or R_2^*, respectively. From (4), the throughput of user 1 is given by

$$\eta_1 = \log_2(1+\tau_1) e^{-\frac{1}{\bar{\alpha}_1}\max\left\{\frac{\tau_2}{P_2 - \tau_2 P_1}, \frac{\tau_1}{P_1}\right\}}. \tag{7}$$

Unfortunately, the throughput maximization in this case is different from that for user 2 as η_1 depends on both τ_1 and τ_2 for given for given $\{P_1, P_2\}$. This demonstrates that the rate allocation with statistical CSI even if $\mathbf{p} = [P_1 \ \ldots \ P_K]^{\mathrm{T}}$ is given is not straightforward, and may need a multi-dimensional optimization.

Since we assume that statistical CSI is available, with individual throughput constraints, the following problem can be formulated:

$$\begin{aligned}&\min ||\mathbf{p}||_1\\ &\text{subject to } \eta_k \geq \bar{\eta}_k, \quad k = 1, \ldots, K,\end{aligned} \tag{8}$$

where $||\mathbf{x}||_1 = \sum_i |x_i|$ denotes the 1-norm of vector $\mathbf{x}$, $\bar{\eta}_k$ is the target throughout of user k, and η_k is the throughput under **A** that is given by

$$\begin{aligned}\eta_k &= R_k \Pr(C_{l;k} > R_l, \ l = k, \ldots, K)\\ &= R_k \Pr(\alpha_k > \max\{\omega_k, \ldots, \omega_K\})\\ &= \log_2(1+\tau_k) e^{-\frac{1}{\bar{\alpha}_k}\max\{\omega_k, \ldots, \omega_K\}}, \end{aligned} \tag{9}$$

where

$$\omega_k = \frac{\tau_k}{P_k - \tau_k \sum_{l=1}^{k-1} P_l}. \tag{10}$$

The problem in (8) is involved as the rates should also be optimized. In other words, joint rate and power allocation has to be considered for (8). In the next section, we present a computationally efficient approach that finds the solution to (8) using multiple one-dimensional search.

There are few remarks.

- The throughput in (9) is not a convex function. Therefore, the problem in (8) is not a convex problem, which means that standard optimization solvers cannot be used.
- In [10] and [17]–[19], statistical CSI is considered for NOMA. In particular, in [18], the power allocation is studied to minimize the total transmission power with outage probability constraints for given rates when only statistical CSI is available at the BS (as in this paper). On the other hand, in this paper, we consider throughput constraints as in (8). The resulting optimization requires joint rate and power allocation, which is more involved than the optimization in [18] where only power allocation is considered.

B. Joint Rate and Power Allocation for Total Power Minimization of OMA

In this subsection, for comparison purposes, we briefly consider joint rate and power allocation to solve the problem in (8) for OMA.

For OMA, the problem in (8) can be readily solved as there is no MAI. We assume that the resource block is divided into K subblocks. Thus, under **A**, the throughput of user k is given by

$$\begin{aligned}\eta_k &= \frac{1}{K} R_k \Pr(\log_2(1+\alpha_k P_k) \geq R_k)\\ &= \frac{1}{K}\log_2(1+\tau_k) e^{-\frac{\tau_k}{\bar{\alpha}_k P_k}}. \end{aligned} \tag{11}$$

Let

$$\eta_k^*(P_k) = \max_{\tau_k \geq 0} \eta_k. \tag{12}$$

It can be readily shown that η_k is unimodal in τ_k, which means that the optimal value of τ_k that maximizes η_k is unique, although η_k is not concave in τ_k. Furthermore, we can show that

$$\frac{d\eta_k}{d\tau_k} = \frac{1}{K \ln 2} e^{-\frac{\tau_k}{\bar{\alpha}_k P_k}} \left(\frac{1}{1+\tau_k} - \frac{\ln(1+\tau_k)}{\bar{\alpha}_k P_k}\right). \tag{13}$$

Since $\ln(1+x) \geq \frac{x}{1+x}$, $\frac{d\eta_k}{d\tau_k} < 0$ if $\tau_k > \bar{\alpha}_k P_k$. From this, the feasible set in (12) can be bounded and (12) is modified as

$$\eta_k^*(P_k) = \max_{0 \leq \tau_k \leq \bar{\alpha}_k P_k} \eta_k. \tag{14}$$

Then, the minimum power by joint rate and power allocation is given by

$$P_k^* = \min\{P_k \mid \eta_k^*(P_k) \geq \bar{\eta}_k\}. \tag{15}$$

Noting that η_k is an increasing function of P_k, we can see that $\eta_k^*(P_k)$ is also an increasing function of P_k. Thus, using simple one-dimensional search techniques (e.g., the bi-section method [27]) can be used to find P_k^*.

IV. An Approach to Find Optimal Solution

In this section, we present the main results of the paper. In particular, we provide an approach to solve (8) under a certain condition with the rate allocation in descending order.

Throughout the paper, we assume that the user ordering is given as follows:

$$\bar{\alpha}_1 \geq \ldots \geq \bar{\alpha}_K. \tag{16}$$

A. Rate Allocation

Throughout this subsection, we assume that the transmission powers are given and consider the rate allocation.

Lemma 1: Suppose that $\mathbf{p}$ is given with $0 < P_k < \infty$. The optimal rate that maximizes the throughput of each user is to be decided in descending order (i.e., from $k = K$ to $k = 1$). That is, the optimal rate, R_k, of user k is decided for given $\mathbf{p}$ as well as $R_{k+1}, \ldots, R_K$, because η_k in (9) is a function of R_l, $l = k+1, \ldots, K$. For convenience, those optimal rates are denoted by R_k^*. Then, τ_k that maximizes the throughput, denoted by τ_k^*, can be obtained as follows:

$$\tau_k^* = \underset{\tau_k}{\operatorname{argmax}}\, \eta_k(\tau_k, \mathbf{p}_k) \text{ subject to } \tau_k \in I_k, \tag{17}$$

where $\mathbf{p}_k = [P_1 \ \ldots \ P_k]^{\mathrm{T}}$,

$$\eta_k(\tau_k, \mathbf{p}_k) = \log_2(1+\tau_k) e^{-\frac{\tau_k}{\bar{\alpha}_k (P_k - \tau_k \sum_{l=1}^{k-1} P_l)}}, \tag{18}$$

and

$$I_k = \left\{ \tau \mid \frac{\omega_{k+1}^* P_k}{1 + \omega_{k+1}^* \sum_{l=1}^{k-1} P_l} \leq \tau < \frac{P_k}{\sum_{l=1}^{k-1} P_l} \right\}. \tag{19}$$

Here, ω_{k+1}^* is ω_{k+1} in (10) with τ_{k+1}^*, which is decided from the previous stage (i.e., $\omega_{k+1}^* = \frac{\tau_{k+1}^*}{P_k - \tau_{k+1}^* \sum_{l=1}^{k} P_l}$), and $\omega_{K+1} = 0$. In addition, we assume that the powers are sufficiently large such that $\omega_k > 0$, $k = 1, \ldots, K$.

Proof: See Appendix A. ■

Lemma 1 shows that the optimal solution to the rate allocation problem can be found as in (17) for given $\mathbf{p}$. To carry out the optimization in (17), we need to understand the properties of the function $\eta_k(\tau_k, \mathbf{p}_k)$. To this end, we have the following results.

Lemma 2: For $A, B > 0$, the following function has a unique maximum:

$$\psi(x) = \ln(1+x) \exp\left(-\frac{x}{A - xB}\right), \quad 0 \leq x < \frac{A}{B}. \tag{20}$$

In addition, $\psi(x)$ is increasing on the interval $[0, x^*)$ and decreasing on the interval $(x^*, A/B)$, where x^* denotes the maximum point of $\psi(x)$. That is, $\psi(x)$ is unimodal on the interval $[0, A/B)$.

Proof: See Appendix B. ■

According to Lemma 2, we can see that $\eta_k(\tau_k, \mathbf{p})$ is unimodal and the optimal solution to (17) can be readily found using simple one-dimensional search methods.

B. Power Allocation

According to Lemma 1, we are able to perform the rate allocation to maximize the throughput of each user in descending order for given $\mathbf{p}$. Although this allows us to maximize each user's throughput for given power allocation, $\mathbf{p}$, it does not solve the problem in (8) (as $||\mathbf{p}||_1$ is to be minimized). To consider the power allocation in relation to the rate allocation, we study a joint rate and power allocation approach under a total transmission power constraint in this subsection. Then, by minimizing the total transmission power, we show that the solution to (8) can be found.

Suppose that the *strict* total transmission power constraint is imposed as

$$\sum_{k=1}^{K} P_k = P_{\mathrm{T}}. \tag{21}$$

In addition, suppose that certain transmit powers, $P_l = \bar{P}_l$, $l = k+1, \ldots, K$, are given such that $\sum_{l=k+1}^{K} \bar{P}_l < P_{\mathrm{T}}$. Let

$$P_{\mathrm{T},k} = P_{\mathrm{T}} - \sum_{l=k+1}^{K} \bar{P}_l. \tag{22}$$

Then, for given $\bar{P}_l$, $l = k+1, \ldots, K$, the throughput in (18) is rewritten as

$$\begin{aligned} \eta_k(\tau_k, \mathbf{p}_k) &= \eta_k(\tau_k, P_k; P_{\mathrm{T},k}) \\ &= \log_2(1+\tau_k) e^{-\frac{\tau_k}{\bar{\alpha}_k (P_k - \tau_k (P_{\mathrm{T},k} - P_k))}}, \end{aligned} \tag{23}$$

which shows that $\eta_k(\tau_k, \mathbf{p}_k)$ is a function of τ_k and P_k under the strict total transmission power constraint in (21), i.e., $\eta_k(\tau_k, P_k; P_{\mathrm{T},k})$, where $P_{\mathrm{T},k}$ is fixed as in (22) (not a variable). From this, the maximum throughput of user k in (17) becomes a function of only one variable, P_k, which is denoted by $\eta_k^*(P_k, \bar{P}_{k+1}, \ldots, \bar{P}_K)$, i.e.,

$$\begin{aligned} \eta_k^*(P_k, \bar{P}_{k+1}, \ldots, \bar{P}_K) &= \max_{\tau_k \in I_k} \eta_k(\tau_k, \mathbf{p}_k) \\ &= \max_{\tau_k \in I_k} \eta_k(\tau_k, P_k; P_{\mathrm{T},k}). \end{aligned} \tag{24}$$

Here, according to Lemma 1, I_k in (19) can be expressed in terms of P_k and $P_{\mathrm{T},k}$ as follows:

$$I_k = \left\{ \tau \mid \frac{\omega_{k+1} P_k}{1 + \omega_{k+1}(P_{\mathrm{T},k} - P_k)} \leq \tau < \frac{P_k}{P_{\mathrm{T},k} - P_k} \right\}, \tag{25}$$

because from (22), we have

$$\sum_{l=1}^{k-1} P_l = P_{\mathrm{T}} - \sum_{l=k}^{K} P_l = P_{\mathrm{T},k} - P_k.$$

The significance of (23) is that the rate allocation can be carried out without a pre-determined power allocation, $\mathbf{p}$. To see this clearly, we note that the rate allocation is carried out when $\mathbf{p}$ is given or fixed as the rate allocation for user k depends on the powers of users $1, \ldots, k-1$ in Lemma 1. On the other hand, (23) shows that the rate allocation for user k depends on the powers that have decided in the previous stages (if the power and rate allocation is performed in descending order), not $\{P_1, \ldots, P_{k-1}\}$. This allows us to perform joint rate and power allocation. For example, we can decide the minimum power P_k that satisfies the throughput constraint, $\eta_k \geq \bar{\eta}_k$ in (8). As shown in the following result, this approach can minimize the partial sum of powers in joint rate and power allocation under individual throughput constraints.

Lemma 3: Suppose the total power constraint is imposed as (21), where P_{T} is sufficiently[1] large. The minimum transmission power of user $k \in \{2, \dots, K\}$ that is recursively found in descending order as follows:

$$P_k^* = \min_{P_k \geq 0} \{\eta_k^*(P_k, P_{k+1}^*, \dots, P_K^*) \geq \bar{\eta}_k\}, \quad k = K, \dots, 2, \tag{26}$$

where

$$\eta_k^*(P_k, P_{k+1}^*, \dots, P_K^*) = \max_{\tau_k \in I_k} \eta_k(\tau_k, \mathbf{p}_k), \tag{27}$$

leads to the minimum partial sum of powers, $\sum_{k=2}^K P_k$, satisfying individual throughput constraints for users $k = 2, \dots, K$. In addition, $\eta_k^*(P_k, P_{k+1}^*, \dots, P_K^*)$ is a nondecreasing function of P_k.

Proof: See Appendix C. ∎

The optimal transmission power P_k^* in (26) can be found by a simple numerical technique that performs one-dimensional search (e.g., the bi-section method) since $\eta_k^*(P_k, P_{k+1}^*, \dots, P_K^*)$ is a nondecreasing function of P_k. The significance of Lemma 3 is that P_k^* as well as τ_k^* can be found recursively so that the complexity grows *linearly* with K.

Note that due to the total transmission power constraint, the power of user 1 is decided as $P_1 = P_{\mathrm{T}} - \sum_{k=2}^K P_k^*$. From (17), we have

$$\begin{aligned}\eta_1^*(P_1, P_2^*, \dots, P_K^*) &= \max_{\tau_1 \in I_1} \eta_1(\tau_1, P_1) \\ &= \max_{\tau_1 \in I_1} \log_2(1 + \tau_1) e^{-\frac{\tau_1}{\bar{\alpha}_1 P_1}}.\end{aligned}$$

Thus, if P_{T} is sufficiently large, there might be P_1^* such that

$$P_1^* = \min_{P_1 \geq 0} \{\eta_1^*(P_1, P_2^*, \dots, P_K^*) \geq \bar{\eta}_1\}. \tag{28}$$

Consequently, if

$$\sum_{k=1}^K P_k^* \leq P_{\mathrm{T}} \tag{29}$$

holds, P_{T} is feasible (with respect to the throughput constraints). Thus, in order to find the solution to (8), we need to find the minimum P_{T} in a feasible set. In the following result, we show that (8) can be solved using the joint rate and power allocation in Lemma 3. Note that since P_k^* depends on P_{T}, P_k^* can be replaced with $P_k^*(P_{\mathrm{T}})$.

Theorem 1: Suppose that $P_k^*(P_{\mathrm{T}})$ is obtained by Lemma 3, $k = 2, \dots, K$ and $P_1^*(P_{\mathrm{T}})$ is obtained as (28) for given P_{T}. If P_{T} satisfies the following condition:

$$\sum_{k=1}^K P_k^*(P_{\mathrm{T}}) = P_{\mathrm{T}}, \tag{30}$$

it is the minimum total transmission power in (8).

Proof: See Appendix D. ∎

[1]To be a sufficiently large P_{T}, it should satisfy (29). This will be discussed later.

According to Theorem 1, we are able to find the solution to (8). For convenience, let

$$\mathcal{P} = \left\{ P_{\mathrm{T}} \mid \sum_{k=1}^K P_k^*(P_{\mathrm{T}}) \leq P_{\mathrm{T}} \right\}.$$

If a feasible $P_{\mathrm{T}} \in \mathcal{P}$ satisfies $P_{\mathrm{T}} - \sum_{k=1}^K P_k^*(P_{\mathrm{T}}) > 0$, we may decrease P_{T} to find a better solution. On the other hand, if P_{T} is not feasible, we need to increase P_{T}. Based on this, we can derive an iterative algorithm. For convenience, we show the following pseudo-code based on the bi-section search to find the minimum total transmission power in (8) as follows.

Algorithm

*) Let $P_{\mathrm{T,max}} = \bar{P}$ ($\bar{P}$ is a sufficiently large power[2] so that P_k^* exists for all k) and $P_{\mathrm{T,min}} = 0$.
0) Let $P_{\mathrm{T}} = (P_{\mathrm{T,max}} + P_{\mathrm{T,min}})/2$.
1) Find $P_k^*(P_{\mathrm{T}})$ according to (26) and (28).
2) If there is no feasible solution for any k, set $P_{\mathrm{T,min}} = P_{\mathrm{T}}$ and move to Step 0.
3) If $|P_{\mathrm{T}} - \sum_{k=1}^K P_k^*(P_{\mathrm{T}})| < \epsilon$, where $\epsilon > 0$ (degree of tolerance) is sufficiently small, stop.
4) $P_{\mathrm{T,max}} = P_{\mathrm{T}}$ and move to Step 0.

Note that **Algorithm** has a low computational complexity. For each k, in (26) (and (28)), the optimal rate for given power can be obtained by the golden-section search or other one-dimensional search algorithm [27]. Then, the bi-section is used to find P_k^*. The outer optimization to find the minimum P_{T} can be based on the bi-section as in **Algorithm**. Consequently, the complexity is proportional to $O(\log^3(1/\epsilon))$.

C. A Suboptimal Approach With Fixed Rates

A suboptimal approach can be derived using outage probability constraints which is similar to the approaches in [10] and [18]. If the transmission rates are fixed with outage probability constraints, certain throughputs can be guaranteed. Let $\tilde{R}_k$ denote the pre-determined transmission rate of user k, $k = 1, \dots, K$. For given target outage probability of user k, denoted by $\mathbb{P}_{\mathrm{out},k}$, the throughput of user k becomes

$$\eta_k = \tilde{R}_k (1 - \mathbb{P}_{\mathrm{out},k}).$$

Consequently, if the target throughput is given with fixed transmission rate $\tilde{R}_k$, we have the following outage probability constraints:

$$\mathbb{P}_{\mathrm{out},k} \leq 1 - \frac{\bar{\eta}_k}{\tilde{R}_k}, \quad k = 1, \dots, K. \tag{31}$$

The power allocation can be considered to minimize the total transmission power with the above outage probability constraints. For convenience, the resulting power allocation is referred to as NOMA suboptimal allocation (NOMA-SOA), while the joint rate and power allocation in Subsection IV-B is referred to as NOMA optimal allocation (NOMA-OA).

[2]For finite $\bar{\eta}_k$, the minimum total transmission power can be obtained by only power allocation as in Subsection IV-C with a non-zero outage probability. This minimum total transmission power can be used as $\bar{P}$ since it is higher than the minimum total transmission power obtained by joint rate and power allocation.

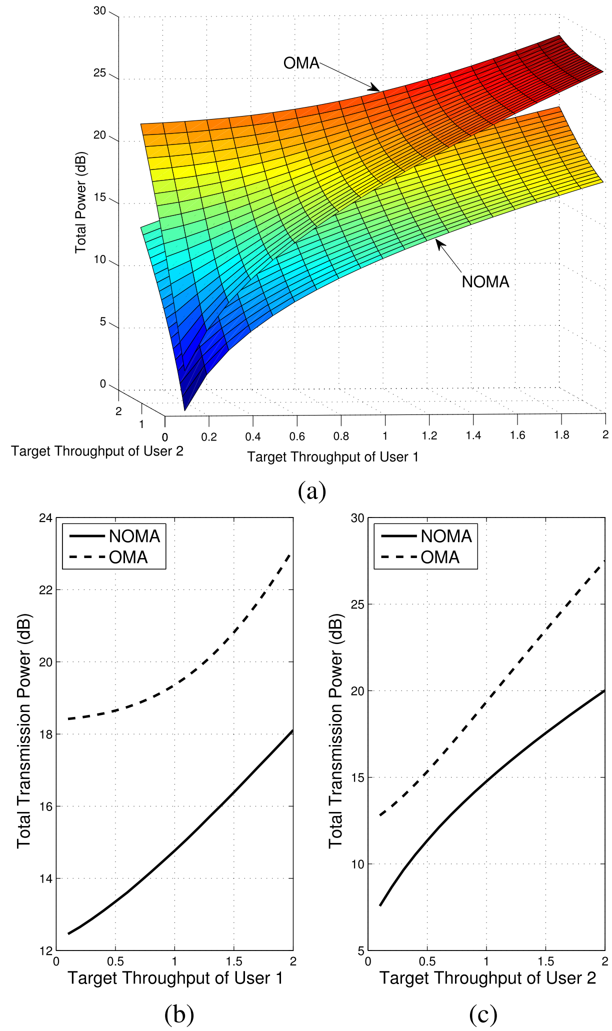


Fig. 1. Minimum total transmission powers, P_{T}, of OMA and NOMA for various values of target throughput when $\{\bar{\alpha}_1, \bar{\alpha}_2\} = \{1, 1/4\}$: (a) Surfaces of P_{T} of OMA (the upper one) and NOMA (the lower one) for $(\bar{\eta}_1, \bar{\eta}_2) \in \{(x, y)|0 < x, y < 2\}$; (b) P_{T} for $\bar{\eta}_2 = 1$ and $\bar{\eta}_1 \in \{0, 2\}$; (c) P_{T} for $\bar{\eta}_1 = 1$ and $\bar{\eta}_2 \in \{0, 2\}$.

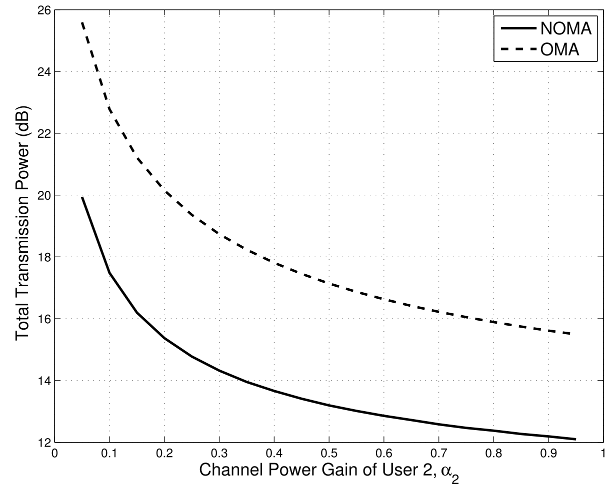


Fig. 2. Minimum total transmission powers, P_{T}, of OMA and NOMA for various values of α_2 when $\alpha_1 = 1$ with $\bar{\eta}_k = 1$, $k = 1, 2$.

It is noteworthy that if the same outage probability constraint is assumed (i.e., $\mathbb{P}_{\mathrm{out},k} = \mathbb{P}_{\mathrm{out}}$ for all k), according to [18, Th. 1], the user ordering in (16) is optimal in NOMA-SOA. However, in NOMA-OA, we cannot claim that (16) is optimal in NOMA-OA, and the optimal user ordering is an open problem.

V. Numerical Results

In this section, we present numerical results to compare OMA and NOMA when the joint rate and power allocation problem to minimize total transmission power with throughput constraints is considered. We also compare the proposed approach with the approach in [18], i.e., NOMA-SOA.

A. Case of $K = 2$

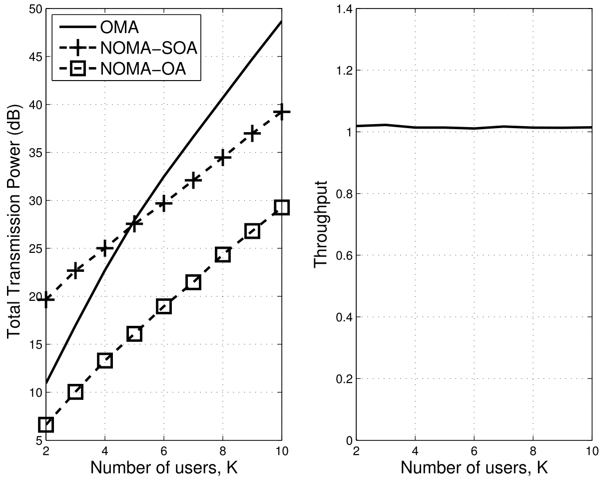
We consider the minimum total transmission powers of OMA and NOMA via joint rate and power allocation when $K = 2$ in this subsection. The minimum total transmission power of OMA is obtained according to (15) for given target throughput for user k, $\bar{\eta}_k$, $k = 1, 2$. To find the minimum total transmission power of NOMA, we use **Algorithm** in Subsection IV-B (i.e., NOMA-OA is considered). Fig. 1 (a) shows the minimum total transmission powers of OMA and NOMA for given set of target throughputs with $\{\bar{\alpha}_1, \bar{\alpha}_2\} = \{1, 1/4\}$. For any pair of target throughputs, $\{\bar{\eta}_1, \bar{\eta}_2\}$, we can confirm that the minimum total transmission power of NOMA is lower than that of OMA. In Figs. 1 (b) and (c), we fix one target throughput and set various values for the other target throughput to see how the minimum total transmission power varies. As any target throughput increases, the minimum total transmission power increases. By comparing Figs. 1 (b) and (c), we can observe that the increase of $\bar{\eta}_2$ (with fixed $\bar{\eta}_1 = 1$) requires a more increase of total transmission power than the increase of $\bar{\eta}_1$ (with fixed $\bar{\eta}_2 = 1$), which results from the fact that $\bar{\alpha}_2$ is lower than $\bar{\alpha}_1$.

Fig. 2 shows the minimum total transmission powers of OMA and NOMA with target throughputs $\bar{\eta}_k = 1$, $k = 1, 2$, for different values of $\bar{\alpha}_2$ when $\bar{\alpha}_1 = 1$ is fixed. As the channel power gain of user 2, $\bar{\alpha}_2$, increases, the total transmission power decreases in both OMA and NOMA. We can also see that the total transmission power of NOMA is always lower than that of OMA.

B. Case of $K \geq 2$

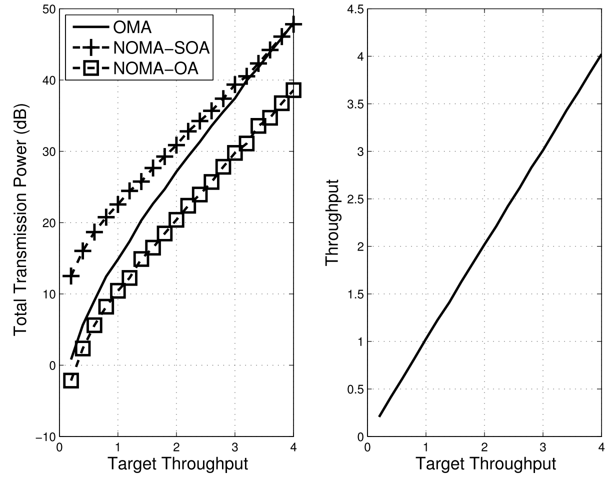
In this subsection, we assume that K users are uniformly and randomly located in a cell of radius 1 under the assumption of **A**. The large scale fading term, i.e., $\bar{\alpha}_k$, becomes a random variable and given by $\bar{\alpha}_k = d_k^{-\kappa}$, where κ is the path loss exponent and d_k is the kth smallest distance from the origin (where the BS is located), i.e.,

$$d_1 \leq \ldots \leq d_K.$$



(a) (b)

Fig. 3. Performance of OMA, NOMA-SOA, and NOMA-OA for different numbers of users, K, when $\bar{\eta}_k = 1$ for all k: (a) minimum total transmission power; (b) throughputs of NOMA-OA from simulations.



(a) (b)

Fig. 4. Performance of OMA, NOMA-SOA, and NOMA-OA for different values of target throughput, $\bar{\eta}$, when $K = 3$: (a) minimum total transmission power; (b) throughputs of NOMA-OA from simulations.

Here, the d_k's are order statistics of the random variables that are independently generated from the distribution[3], $f(d) = 2d$. In addition, for the simulation, we assume that the path loss exponent is 3.5, i.e., $\kappa = 3.5$. Consequently, in conjunction with the assumption of **A**, we consider the following channel gains:

$$\alpha_k = \upsilon_k \bar{\alpha}_k = \upsilon_k d_k^{-3.5}, \quad (32)$$

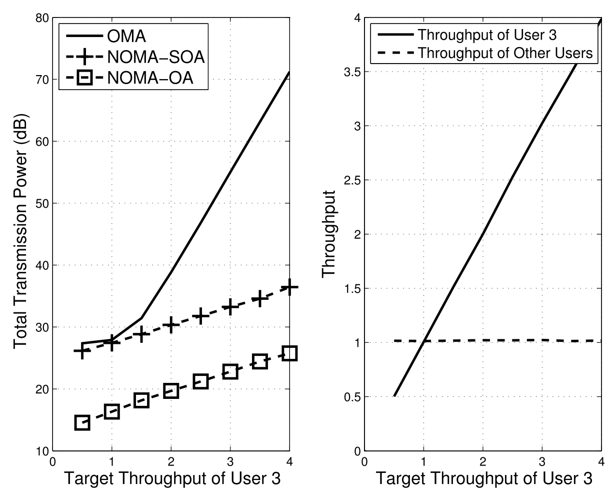
where υ_k is an independent exponential random variable with $\mathbb{E}[\upsilon_k] = 1$ for small-scaling (Rayleigh) fading.

In this subsection, we also present simulation results of NOMA-SOA with the target outage probability $\mathbb{P}_{\text{out},k} = 10^{-2}$ for comparison purposes. In this case, the transmission rate in NOMA-SOA becomes fixed and given by $\tilde{R}_k = \frac{\bar{\eta}_k}{1-\mathbb{P}_{\text{out},k}} = \frac{\bar{\eta}_k}{0.99}$.

Fig. 3 (a) shows the minimum total transmission powers of OMA, NOMA-SOA [18], and NOMA-OA for different numbers of users, K, when $\bar{\eta}_k = 1$ for all k. In general, as K increases, the minimum total transmission power increases in all cases. We can confirm again that NOMA-OA performs better than OMA in terms of the total transmission power. Furthermore, NOMA-OA outperforms NOMA-SOA due to joint rate and power allocation. Interestingly, we find that OMA can perform better than NOMA-SOA. This results from the fact that joint rate and power allocation is used for OMA, while only power allocation is used for NOMA-SOA (with fixed rates).

In addition, we can observe that the gap between the minimum total transmission powers of OMA and NOMA-OA increases with K. In other words, NOMA-OA becomes more energy (or power) efficient than OMA as more users share the same resource block for given target throughput. In Fig. 3 (b), we show the throughput of NOMA-OA from simulations with the allocated rates and powers by **Algorithm** in Subsection IV-B. We can see that the target throughput, $\bar{\eta}_k = 1$, can be achieved in NOMA-OA using the proposed joint rate and power allocation. Note that the achieved throughput in Fig. 3 (b) is slightly higher than the target throughput, $\bar{\eta}_k = 1$. In **Algorithm**, since we perform (26) (and (28)) with a small tolerant parameter, i.e., $\eta_k \geq \bar{\eta}_k + \delta$, where $\delta > 0$ is a small tolerant parameter, it results in a slightly higher throughput than the target throughput.

[3]This is the pdf for distance d from the center of a point picked at random in a unit disk.



(a) (b)

Fig. 5. Performance of OMA, NOMA-SOA, and NOMA-OA for different values of target throughput of user 3, $\bar{\eta}_3$, while the other users' target throughputs are set to 1 (i.e., $\bar{\eta}_k = 1$, $k \in \{1, 2, 4, 5\}$) when $K = 5$: (a) minimum total transmission power; (b) throughputs of NOMA-OA from simulations.

Fig. 4 (a) shows the minimum total transmission powers of OMA, NOMA-SOA, and NOMA-OA for different values of the target throughput, $\bar{\eta} = \bar{\eta}_k$, $k = 1, \ldots, K$ when $K = 3$. For a higher target throughput $\bar{\eta}$, the minimum total transmission power is higher, while the minimum total transmission power of NOMA-OA is lower than that of OMA and that of

NOMA-SOA. In Fig. 4 (b), we can also confirm that the target throughput can be achieved in NOMA-OA using the proposed joint rate and power allocation approach.

In Fig. 5, we assume that the 3rd user's target throughput varies while the other users' target throughputs are set to 1 (i.e., $\bar{\eta}_k = 1,\ k \in \{1, 2, 4, 5\}$) when $K = 5$. As the target throughput of user 3 increases, the minimum total transmission power increases in all OMA, NOMA-SOA, and NOMA-OA. Interestingly, the increase of the minimum total transmission power of OMA is significant, while those of NOMA-SOA and NOMA-OA are not. This shows that the allocation of power (and rate) across multiple users in NOMA-SOA (and NOMA-OA, respectively) can be effective to support a user's higher target throughput without a significant increase of the total transmission power. On the other hand, in OMA, the rate and power allocation for a user alone cannot be effective to accommodate a higher target throughput.

VI. Concluding Remarks

We studied joint rate and power allocation to minimize the total transmission power with throughput constraints in NOMA and proposed an approach to find the optimal solution in this paper. Since the throughput is not a concave function of power, we could not rely on existing approaches for convex optimization. Thus, we first considered the optimal rate allocation for given powers. Then, we showed that the power allocation can be carried out in conjunction of the rate allocation for a given total transmission power. Finally, an iterative algorithm was derive to minimize the total transmission power. Based on simulation results with optimal joint rate and power allocation, we confirmed that NOMA can have a much lower total transmission power than OMA for given target throughput of each user. In addition, due to joint rate and power allocation, it was shown that the total transmission power can be lower than that obtained by only power allocation with fixed rates in NOMA.

Appendix A
Proof of Lemma 1

First of all, we verify that I_k in (17) is not empty, which can be proved by showing the following inequality:

$$\frac{\omega_{k+1} P_k}{1 + \omega_{k+1} \sum_{l=1}^{k-1} P_l} < \frac{P_k}{\sum_{l=1}^{k-1} P_l},$$

which becomes after some manipulations,

$$\omega_{k+1} < \frac{1}{\sum_{l=1}^{k-1} P_l} + \omega_{k+1}.$$

Since $P_l < \infty$, the inequality is valid and this shows that I_k is not empty.

Suppose that τ_l^*, $l = K, \ldots, k+1$, are found. Then, for k, successful decoding is guaranteed if

$$C_{l;k} > R_l^*,\ l = K, \ldots, k+1, \text{ and } C_{k;k} > R_k,$$

which is equivalent to

$$\alpha_k > \max \left\{ \omega_K^*, \ldots, \omega_{k+1}^*, \frac{\tau_k}{P_k - \tau_k \sum_{l=1}^{k-1} P_l} \right\}. \tag{33}$$

We now claim that

$$\omega_{k+1}^* \le \omega_k^*, \tag{34}$$

which will be proved later. Then, (33) is reduced to

$$\alpha_k > \max \left\{ \omega_{k+1}^*, \frac{\tau_k}{P_k - \tau_k \sum_{l=1}^{k-1} P_l} \right\}. \tag{35}$$

From (9), in order to maximize the throughput of user k with respect to τ_k, we now consider two different cases.

Case A: Let $\omega_{k+1}^* \le \frac{\tau_k}{P_k - \tau_k \sum_{l=1}^{k-1} P_l}$. Then, from (9), the maximum throughput is given by

$$\eta_k^{(a)} = \max_{\tau_k} \log_2(1+\tau_k) e^{-\frac{\tau_k}{\bar{\alpha}_k (P_k - \tau_k \sum_{l=1}^{k-1} P_l)}}$$
$$\text{subject to } \begin{cases} \omega_{k+1}^* \le \dfrac{\tau_k}{P_k - \tau_k \sum_{l=1}^{k-1} P_l} \\ P_k - \tau_k \sum_{l=1}^{k-1} P_l > 0. \end{cases} \tag{36}$$

The constraint in (36) becomes

$$\frac{\omega_{k+1}^* P_k}{1 + \omega_{k+1}^* \sum_{l=1}^{k-1} P_l} \le \tau_k < \frac{P_k}{\sum_{l=1}^{k-1} P_l}. \tag{37}$$

Case B: Let $\omega_{k+1}^* > \frac{\tau_k}{P_k - \tau_k \sum_{l=1}^{k-1} P_l}$. In this case, the throughput becomes $\eta_k = R_k e^{-\frac{\omega_{k+1}^*}{\bar{\alpha}_k}}$. Since ω_{k+1}^* is given, the maximum throughput in this case becomes

$$\eta_k^{(b)} = \max_{\tau_k < \frac{\omega_{k+1}^* P_k}{1+\omega_{k+1}^* \sum_{l=1}^{k-1} P_l}} \log_2(1+\tau_k) e^{-\frac{\omega_{k+1}^*}{\bar{\alpha}_k}}. \tag{38}$$

From (36) and (38), we can show that

$$\eta_k^{(b)} \le \log_2 \left(1 + \frac{\omega_{k+1}^* P_k}{1 + \omega_{k+1}^* \sum_{l=1}^{k-1} P_l} \right) e^{-\frac{\omega_{k+1}^*}{\bar{\alpha}_k}} \le \eta_k^{(a)}. \tag{39}$$

In (39), the first inequality is due to the fact that $\log_2(1+x)$ is an increasing function. The second inequality is due to the following:

$$\max_{\frac{\omega_{k+1}^* P_k}{1+\omega_{k+1}^* \sum_{l=1}^{k-1} P_l} \le \tau_k < \frac{P_k}{\sum_{l=1}^{k-1} P_l}} \log_2(1+\tau_k) e^{-\frac{\tau_k}{\bar{\alpha}_k (P_k - \tau_k \sum_{l=1}^{k-1} P_l)}}$$
$$(= \eta_k^{(a)})$$
$$\ge \log_2(1+\tau_k) e^{-\frac{\tau_k}{\bar{\alpha}_k (P_k - \tau_k \sum_{l=1}^{k-1} P_l)}} \Bigg|_{\tau_k = \frac{\omega_{k+1}^* P_k}{1+\omega_{k+1}^* \sum_{l=1}^{k-1} P_l}}$$
$$= \log_2 \left(1 + \frac{\omega_{k+1}^* P_k}{1 + \omega_{k+1}^* \sum_{l=1}^{k-1} P_l} \right) e^{-\frac{\omega_{k+1}^*}{\bar{\alpha}_k}}.$$

From (39), we can see that the maximum throughput of user k can be decided as in (36). That is, τ_k^* can be found as (17). Since this result is based on the claim in (34), we need to verify it. For $k = K$, (34) is obvious as ω_{K+1}^* is assumed to be 0. According to (39), we see that τ_k^* is decided according to Case A. Thus, we have

$$\omega_{k+1}^* \le \frac{\tau_k^*}{P_k - \tau_k^* \sum_{l=1}^{k-1} P_l} = \omega_k^* \tag{40}$$

by the definition of ω_k in (10), which proves the claim in (34).

APPENDIX B
PROOF OF LEMMA 2

We first note that $\psi(x)$ is not a concave function of x. Thus, we cannot simply claim that $\psi(x)$ has a unique maximum. The first derivative is given by

$$\frac{d\psi(x)}{dx} = \frac{e^{-\frac{x}{A-xB}}}{(1+x)(A-Bx)^2} G(x), \tag{41}$$

where $G(x) = (A-Bx)^2 - A(1+x)\ln(1+x)$. Since $\frac{e^{-\frac{x}{A-xB}}}{(1+x)(A-Bx)^2} > 0$, the roots of $\psi(x)$ can be found by finding x that satisfies $G(x)$. We can observe that $G(0) = A^2 > 0$ and $\lim_{x \to A/B} G(x) = -A(1 + A/B)\ln(1 + A/B) < 0$. Furthermore, from

$$\frac{dG(x)}{dx} = -2B(A-Bx) - A\left(\ln(1+x)+1\right) < 0, \quad 0 \le x < \frac{A}{B},$$

we can find that $G(x)$ is a decreasing function of x, $0 \le x < \frac{A}{B}$. Consequently, we observe that $\frac{d\psi(x)}{dx} > 0$ on the interval $[0, x^*)$ and $\frac{d\psi(x)}{dx} < 0$ on the interval $(x^*, A/B)$, where x^* is one extreme point of $\psi(x)$ or the unique solution of $\frac{d\psi(x)}{dx} = 0$. This implies that $\psi(x)$ is increasing and then decreasing with the maximum at x^*.

APPENDIX C
PROOF OF LEMMA 3

The throughput in (23), $\eta_k(\tau_k, \mathbf{p}_k)$, is an increasing function of P_k. Let $P'_k < P''_k$. Then, it follows

$$\eta_k(\tau_k, P'_k, P_{\mathrm{T},k}) \le \eta_k(\tau_k, P''_k, P_{\mathrm{T},k}),$$

which implies that $\eta^*_k(P_k, P^*_{k+1}, \dots, P^*_K)$ is a nondecreasing function of P_k. From this, we can show that the minimum transmission power of user k that satisfies the throughput constraint in (8) is given by (26).

We now show that the descending order to decide the power is optimal. Since the throughput of user K is dependent only on τ_K, the minimum power of P_K is P^*_K from (26). Consider $P^\sharp_K$ which is greater than P^*_K, i.e., $P^\sharp_K > P^*_K$, which results in a higher throughput of user K than $\bar{\eta}_K$ (thus, the throughput constraint for user K is satisfied). Let

$$\begin{aligned} P_{\mathrm{T},K-1} &= P_{\mathrm{T}} - P^*_K \\ P^\sharp_{\mathrm{T},K-1} &= P_{\mathrm{T}} - P^\sharp_K. \end{aligned} \tag{42}$$

Clearly, since $P_{\mathrm{T},K-1} > P^\sharp_{\mathrm{T},K-1}$, for a subproblem of the power allocation for users $1, \dots, K-1$, with $P_K = P^*_K$, the total power, P_{T}, that satisfies the throughput constraints for users $1, \dots, K-1$ can be lower than that with $P_K = P^\sharp_K$. Thus, to minimize the transmission power, we can decide P^*_K and set $P_K = P^*_K$ first. Then, P^*_{K-1} can be found for given $P_K = P^*_K$, and so on. Thus, P^*_k can be found recursively in descending order, and it results in the minimum partial sum, $\sum_{k=2}^K P_k$.

Finally, note that P^*_k in (26) may not exist unless P_{T} is sufficiently large. That is, if

$$\sum_{l=k+1}^K P^*_l < P_{\mathrm{T}} < \sum_{l=k}^K P^*_l,$$

the power allocation violates the total transmission power. To avoid this, P_{T} has to be sufficiently large.

APPENDIX D
PROOF OF THEOREM 1

For the equality in (30), consider (23), where we can see that $\eta_k(\tau_k, P_k, P_{\mathrm{T},k})$ is a decreasing function of P_{T} for $k \in \{2, \dots, K\}$. According to Lemma 3, $\eta^*_k(P_k, P^*_{k+1}, \dots, P^*_K)$ is a nondecreasing function of P_k. Thus, $P^*_k(P_{\mathrm{T}})$, $k = 2, \dots, K$, is a nonincreasing function of P_{T}. This implies that $\sum_{k=2}^K P^*_k(P_{\mathrm{T}})$ is also a nonincreasing function of P_{T}.

For two feasible total transmission powers, $P'_{\mathrm{T}}, P''_{\mathrm{T}} \in \mathcal{P}$, if $P'_{\mathrm{T}} < P''_{\mathrm{T}}$, we have

$$\sum_{k=2}^K P^*_k(P'_{\mathrm{T}}) \ge \sum_{k=2}^K P^*_k(P''_{\mathrm{T}}), \tag{43}$$

because $\sum_{k=2}^K P^*_k(P_{\mathrm{T}})$ is a nonincreasing function of P_{T} as shown above. In addition, it can be shown that

$$P'_{\mathrm{T}} - \sum_{k=2}^K P^*_k(P'_{\mathrm{T}}) = P'_1 < P''_{\mathrm{T}} - \sum_{k=2}^K P^*_k(P''_{\mathrm{T}}) = P''_1. \tag{44}$$

From this, if $P'_1 = P^*_1$ (i.e., $\sum_{k=1}^K P^*_k(P'_{\mathrm{T}}) = P'_{\mathrm{T}}$ or (30) holds with P'_{T}), P'_{T} becomes the smallest P_{T} that satisfies all the throughput constraints.

On the other hand, if we assume that $P''_1 = P^*_1(P''_{\mathrm{T}})$ (i.e., P''_{T}, which is greater than another feasible P'_{T}, satisfies (30) or becomes a solution to (30)), we have $\sum_{k=1}^K P^*_k(P''_{\mathrm{T}}) = P''_{\mathrm{T}}$. In this case, from (43), we have

$$\sum_{k=2}^K P^*_k(P'_{\mathrm{T}}) + P^*_1(P''_{\mathrm{T}}) \ge \sum_{k=2}^K P^*_k(P''_{\mathrm{T}}) + P^*_1(P''_{\mathrm{T}}) = P''_{\mathrm{T}}. \tag{45}$$

Since $P'_1 + \sum_{k=2}^K P^*_k(P'_{\mathrm{T}}) = P'_{\mathrm{T}}$ from (44), (45) becomes

$$P'_{\mathrm{T}} + P^*_1(P''_{\mathrm{T}}) - P'_1 = P'_{\mathrm{T}} + P''_1 - P'_1 \ge P''_{\mathrm{T}}. \tag{46}$$

However, since $P'_1 < P''_1$, (46) implies that $P'_{\mathrm{T}} \ge P''_{\mathrm{T}}$, which contradicts the assumption that $P'_{\mathrm{T}} < P''_{\mathrm{T}}$. This shows that the P_{T} that satisfies (30) is the minimum of P_{T}, which completes the proof.

REFERENCES

[1] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. 77th IEEE VTC-Spring*, Dresden, Germany, Jun. 2013, pp. 1–5.

[2] B. Kim *et al.*, "Non-orthogonal multiple access in a downlink multiuser beamforming system," in *Proc. IEEE Military Commun. Conf. (MILCOM)*, Nov. 2013, pp. 1278–1283.

[3] J. Choi, "Non-orthogonal multiple access in downlink coordinated two-point systems," *IEEE Commun. Lett.*, vol. 18, no. 2, pp. 313–316, Feb. 2014.

[4] Z. Ding, F. Adachi, and H. V. Poor, "The application of MIMO to non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 537–552, Jan. 2016.

[5] *TP for Classification of MUST Schemes*, document 3GPP R1-154999, 2015.

[6] S.-L. Shieh and Y.-C. Huang, "A simple scheme for realizing the promised gains of downlink nonorthogonal multiple access," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1624–1635, Apr. 2016.

[7] Y. Tian, A. R. Nix, and M. Beach, "On the performance of opportunistic NOMA in downlink CoMP networks," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 998–1001, May 2016.

[8] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. IEEE ICC*, Jun. 1995, pp. 331–335.

[9] J. Choi, "Power allocation for max-sum rate and max-min rate proportional fairness in NOMA," *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 2055–2058, Oct. 2016.

[10] S. Timotheou and I. Krikidis, "Fairness for non-orthogonal multiple access in 5G systems," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1647–1651, Oct. 2015.

[11] J. Choi, "On the power allocation for MIMO-NOMA systems with layered transmissions," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3226–3237, May 2016.

[12] Q. Sun, S. Han, C. L. I., and Z. Pan, "On the ergodic capacity of MIMO NOMA systems," *IEEE Wireless Commun. Lett.*, vol. 4, no. 4, pp. 405–408, Aug. 2015.

[13] Z. Ding and H. V. Poor, "Design of massive-MIMO-NOMA with limited feedback," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 629–633, May 2016.

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

[15] J. Choi, "Effective capacity of NOMA and a suboptimal power control policy with delay QoS," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1849–1858, Apr. 2017.

[16] P. Xu, Y. Yuan, Z. Ding, X. Dai, and R. Schober, "On the outage performance of non-orthogonal multiple access with 1-bit feedback," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 6716–6730, Oct. 2016.

[17] Z. Yang, Z. Ding, P. Fan, and G. K. Karagiannidis, "On the performance of non-orthogonal multiple access systems with partial channel information," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 654–667, Feb. 2016.

[18] J. Cui, Z. Ding, and P. Fan, "A novel power allocation scheme under outage constraints in NOMA systems," *IEEE Signal Process. Lett.*, vol. 23, no. 9, pp. 1226–1230, Sep. 2016.

[19] S. Shi, L. Yang, and H. Zhu, "Outage balancing in downlink nonorthogonal multiple access with statistical channel state information," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4718–4731, Jul. 2016.

[20] R. Zekavat and R. M. Buehrer, *Handbook of Position Location: Theory, Practice and Advances*. New York, NY, USA: Wiley, 2011.

[21] J. Choi, "Capacity/throughput optimization for H-BLAST with SC receiver over MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 3, pp. 1016–1024, Mar. 2008.

[22] J. Zhang and J. S. Lehnert, "Throughput-optimal precoding and rate allocation for MISO systems with noisy feedback channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2139–2155, May 2008.

[23] M. Haenggi, "Outage, local throughput, and capacity of random wireless networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4350–4359, Aug. 2009.

[24] S. Lin and D. J. Costello, *Error Control Coding: United States Edition*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1983.

[25] S. B. Wicker, *Error Control Systems for Digital Communication and Storage*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1995.

[26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.

[27] D. G. Luenberger, *Linear and Nonlinear Programming*, 2nd ed. New York, NY, USA: Springer, 2003.

Jinho Choi (SM'02) was born in Seoul, South Korea. He received the B.E. (*magna cum laude*) degree in electronics engineering from Sogang University, Seoul, in 1989, and the M.S.E. and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology, Daejeon, in 1991 and 1994, respectively. In 2013, he was with the College of Engineering, Swansea University, U.K., as a Professor/Chair in wireless. He is currently with the Gwangju Institute of Science and Technology as a Professor. He authored two books published by Cambridge University Press in 2006 and 2010. His research interests include wireless communications and array/statistical signal processing. He received the 1999 Best Paper Award for Signal Processing from EURASIP and the 2009 Best Paper Award from WPMC (Conference). Currently, he is an Editor of the IEEE TRANS. COMMUNICATIONS and had served as an Associate Editor or an Editor of other journals, including the IEEE COMMUNICATIONS LETTERS, the *Journal of Communications and Networks*, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and the ETRI Journal.