Article:

Xu, Peng and Cumanan, Kanapathippillai orcid.org/0000-0002-9735-7019 (2017) Optimal Power Allocation Scheme for Non-Orthogonal Multiple Access with α-Fairness. IEEE Journal on Selected Areas in Communication. 7986959. pp. 2357-2369. ISSN 1558-0008

https://doi.org/10.1109/JSAC.2017.2729780

Optimal Power Allocation Scheme for Non-Orthogonal Multiple Access with α-Fairness

Peng Xu and Kanapathippillai Cumanan, *Member, IEEE*

***Abstract*—This paper investigates the optimal power allocation scheme for sum throughput maximization of non-orthogonal multiple access (NOMA) system with α-fairness. In contrast to the existing fairness NOMA models, α-fairness can only utilize a single scalar to achieve different user fairness levels. Two different channel state information at the transmitter (CSIT) assumptions are considered, namely, statistical and perfect CSIT. For statistical CSIT, fixed target data rates are predefined, and the power allocation problem is solved for sum throughput maximization with α-fairness, through characterizing several properties of the optimal power allocation solution. For perfect CSIT, the optimal power allocation is determined to maximize the instantaneous sum rate with α-fairness, where user rates are adapted according to the instantaneous channel state information (CSI). In particular, a simple alternate optimization (AO) algorithm is proposed, which is demonstrated to yield the optimal solution. Numerical results reveal that, at the same fairness level, NOMA significantly outperforms the conventional orthogonal multiple access (MA) for both the scenarios with statistical and perfect CSIT.**

***Index Terms*—Non-orthogonal multiple access, α-fairness, outage probability, ergodic rate, power allocation.**

I. Introduction

Non-orthogonal multiple access (NOMA) enables to realize a balanced tradeoff between spectral efficiency and user fairness, which has been recognized as a promising multiple access (MA) technique for future fifth generation (5G) networks [1]–[18]. In contrast to the conventional MA (e.g., time-division multiple access (TDMA), etc.), NOMA exploits power domain to simultaneously serves multiple users at different power levels, where power allocation at the base station plays a key role in determining the overall performance of the system. Downlink NOMA combines superposition coding at the transmitter and successive interference cancellation (SIC) decoding at each receiver, which can be considered as a special case of the conventional broadcast channel (BC) [19]. To maintain user fairness, NOMA always allocates more power to the users with weaker channel gains.

Based on the superposition coding, the works in [20] and [21] explored the capacity region of the degraded discrete memoryless BC and the Gaussian BC with single-antenna terminals, respectively. On the other hand, the ergodic capacity and the outage capacity/probability of the fading BC with perfect channel state information at the transmitter (CSIT) were established in [22] and [23], respectively. For the concept of ergodic capacity, user rates can be adapted according to the instantaneous channel state information (CSI); while the concept of outage is more appropriate for applications with stringent delay constraints as a predefined rate is assumed for each transmission. In [24], the performance of outage capacity was analyzed without CSIT. However, these works for conventional BCs have not taken into account the issue of user fairness, which is different from NOMA with fairness constraints.

Recently, the issue of user fairness has received considerable attention in a series of NOMA systems [11]–[18]. The works in [11]–[14], [25]–[28] adopted fixed power allocation approaches to guarantee user fairness, which can only ensure that the users with weaker channel gains are allocated with more power and might suffer from poor user fairness when some users have very poor channel conditions. In order to enhance user fairness, an appropriate power allocation should be adopted at the base station for each user message in the superposition coding, similar to the works in [15]–[18]. In [15], the max-min and min-max power allocation schemes are proposed to maximize the ergodic rate and minimize the outage probability, respectively, whereas the common outage probability of NOMA with one-bit feedback is minimized in [16]. A throughput maximization scheme for a multiple-input multiple-output (MIMO) NOMA system is presented in [17] by solving the max-min fairness problem. However, the schemes proposed in [15]–[17] can only achieve absolute fairness[1], where the system throughput is limited by the user with the worst channel gain. In [18], the power allocation approach has been proposed to maximize the minimum weighted success probability, where a weighting vector is exploited to adjust fairness levels. However, the design of the optimal weighting vector is a challenging issue, which has not been addressed in [18]. Most recently, a proportional fairness over a time-domain window size has been presented for NOMA in [29].

The main objective of this paper is to investigate the optimal power allocation scheme for sum throughput maximization of the NOMA system with α-fairness constraints. In existing fairness models in [15]–[17], only absolute fairness can be

Peng Xu is with the Chongqing Key Laboratory of Mobile Communications Technology, Chongqing University of Posts and Telecommunications (CQUPT), Chongqing, 400065, P. R. China. Kanapathippillai Cumanan is with Department of Electronic Engineering, University of York, York, UK, YO10 5DD.

The work of Peng Xu was supported in part by the Scientific and Technological Research Program of Chongqing Municipal Education Commission under Grant KJ1704088, and in part by Doctoral Initial Funding of Chongqing University of Posts and Telecommunications under Grant A2016-84. The work of K. Cumanan was supported by H2020- MSCARISE-2015 under grant number 690750.

[1]The term "absolute fairness" means that all users have the same performance (e.g., the same outage probability or the same ergodic rate).

achieved; while in [18], a weighting vector are exploited to adjust the fairness level. However, α-fairness only utilizes a single scalar, denoted as α, to achieve different user fairness levels and well-known efficiency-fairness tradeoffs [30]. The concept of α-fairness was first introduced in [31] for a fair end-to-end congestion control, which generalizes proportional and max-min fairness approaches. Since then, α-fairness has been widely incorporated in a series of fairness optimization models for resource allocation and congestion control (e.g., [32]–[34]). More details on fairness in wireless networks can be found in [30] and the references therein. In general, increasing α results in higher user fairness [32]. For instance, maximum efficiency can be achieved by setting $\alpha = 0$, whereas proportional and max-min fairness can be achieved by setting $\alpha = 1$ [35], [36], and $\alpha \to \infty$ [31], respectively.

In this paper, a downlink NOMA system with two different CSIT assumptions are considered: statistical and perfect CSIT. For statistical CSIT, fixed target data rates should be predefined for all users, for which, we first analyze the outage probability of each user, and then formulate the power allocation optimization framework for sum throughput maximization with α-fairness. However, this optimization problem is not convex in nature due to the non-convex objective function. To circumvent this non-convex issue, we reformulate the original problem into an equivalent problem with a simple expression. Analysis reveals that the equivalent transformed problem is convex for the case of $\alpha \geq 1$ and still non-convex for the case of $\alpha < 1$. However, for the case $\alpha < 1$, the structure of the optimal solution is characterized based on some properties of the optimal power allocation solution, which demonstrates that the problem turns out to be convex if we fix the first power parameter and the number of power parameters that are below $(1-\alpha)/2$.

For perfect CSIT, the power allocation problem is formulated to maximize the instantaneous sum rate with α-fairness, where user rates are adapted according to instantaneous CSI. We first transform this optimization problem into an equivalent problem by setting a series of parameters to denote the sum power allocated to a group of users. Then, we demonstrate that there exists only one solution to satisfy the Karush-Kuhn Tucker (KKT) conditions. Furthermore, a simple alternate optimization (AO) algorithm is proposed to yield the optimal solution through solving KKT conditions. The algorithm is developed based on the idea of AO approach, where each KKT condition is solved by fixing the other corresponding parameters. In addition, it is shown that each variable is monotonically increasing in each iteration of the algorithm and therefore it converges.

Numerical results reveal that parameter α can adjust the fairness level in terms of fairness index [37] for both NOMA and TDMA. In addition, for the same required fairness index , NOMA outperforms TDMA in terms of both the sum throughput with statistical CSIT and ergodic sum rate with perfect CSIT. Moreover, the proposed algorithm for ergodic rate maximization converges with less number of iterations than the conventional interior point algorithm in most scenarios.

Throughout this paper, $\mathbb{P}(\cdot)$ and $\mathbb{E}(\cdot)$ are used to denote the probability of an event and the expectation of a random variable. Moreover, $[1:K]$ represents the set $\{1, \cdots, K\}$, and $\{x_i\}$ indicates the sequence formed by all the possible x_i's. Furthermore, $\log(\cdot)$ and $\ln(\cdot)$ stands for the logarithm with base 2 and the natural logarithm, whereas $\exp(\cdot)$ denotes the exponential function.

II. System Model and Problem Formulations

A downlink NOMA system is considered with one single-antenna base station and K single-antenna users. For this network setup, quasi-static block fading is assumed, where the channel gains from the base station to all users are constant during one fading block, but change independently from one fading block to the next fading block. The base station transmits K messages to the users using the NOMA scheme, i.e., it sends a superposition codeword $x = \sum_{k=1}^{K} \sqrt{\tilde{P}_k} s_k$ during each fading block, where s_k is the signal intended for user k with $E[|s_k|^2] = 1$ and $\tilde{P}_k$ is the power allocated to user k, which satisfies $\sum_{k=1}^{K} \tilde{P}_k \leq P$. The received signal at user k can be expressed as

$$y_k = h_k \sum_{i=1}^{K} \sqrt{\tilde{P}_i} s_i + n_k, \ k \in [1:K], \tag{1}$$

where the noise n_k at user k is assumed to be an additive white Gaussian noise with zero mean and unit variance, and h_k denotes the channel gain from the base station to user k. Specifically, $h_k = d_k^{-\beta/2} g_k$, where g_k is a normalized Rayleigh fading channel gain with unit variance, d_k is the distance between the base station and user k, and β is the path loss exponent. Without loss of generality, it is assumed that $d_1 > d_2 > \cdots > d_K$. In addition, it is also assumed that noises and channel gains associated with all users are mutually independent from each other. In this paper, we consider the case where each superposition codeword spans only a single fading block.

The users employ SIC to decode their messages, where the user order (or equivalently, decoding order) is determined by the base station according to the CSIT assumption discussed later in this section. It can be assumed without loss of generality that user k is allocated with index k. In the SIC process, user k will sequentially decode the messages of users l, $l \in [1:k]$ and then successively remove these messages from its received signal. When user k decodes the message of user l, the signal-to-interference-plus-noise ratio (SINR) can be written as

$$\gamma_l^{(k)} = \frac{\tilde{P}_l H_k}{H_k \sum_{m=l+1}^{K} \tilde{P}_m + 1}, \ l \in [1:k], \tag{2}$$

where we define $H_k = |h_k|^2$, $\forall k \in [1:K]$, for simplicity; obviously, H_k follows an exponential distribution with a mean $d_k^{-\beta}$.

Next, we will investigate optimal power allocation from a fairness perspective, under two main CSI assumptions of statistical and perfect CSIT. To model fairness, we adopt α-fair utility function [37]

$$u_\alpha(x) \triangleq \begin{cases} \ln(x), & \text{if } \alpha = 1, \\ \frac{x^{1-\alpha}}{1-\alpha}, & \text{if } \alpha \neq 1, \alpha \geq 0. \end{cases}, \ x > 0, \tag{3}$$

where x could be throughput or instantaneous rate shown later in this section, and different values of α represents different fairness levels. Note that the choices of $\alpha = 0$ and $\alpha \to \infty$ represent no fairness and absolute fairness requirements, respectively.

A. NOMA with Statistical CSIT

For the statistical CSIT scenario, only statistics of fading channels (including channel distributions, means and variances) are available at the transmitter, and hence fixed target data rates should be predefined for all users. The overhead cost in this scenario would be low as the variation of channel statistics is much more slower than that of instantaneous CSI. Moreover, the user order is determined based on the distance from the base station to each user, where a user with a larger distance is assigned with a smaller order index. Since it is assumed that $d_1 > d_2 \cdots > d_K$ previously in this section, user k is always allocated with order index k. Assume that the base station transmits one message to each user in each block with the same fixed target rate r_0 bits per channel use (BPCU). [2] For this transmission scenario, the outage probability needs to be evaluated, and the outage probability for user k can be expressed as

$$\begin{aligned}\mathbb{P}_k &= \mathbb{P}\left\{\gamma_l^{(k)} < \hat{r}_0, \text{ for some } l \in [1:k]\right\}\\ &= \mathbb{P}\left\{H_k < \max\left\{\frac{\hat{r}_0}{\hat{P}_1}, \cdots, \frac{\hat{r}_0}{\hat{P}_k}\right\}\right\}\\ &= 1 - \exp\left(-\max\left\{\frac{\hat{r}_0 d_k^{\beta}}{\hat{P}_1}, \cdots, \frac{\hat{r}_0 d_k^{\beta}}{\hat{P}_k}\right\}\right), \end{aligned} \tag{4}$$

where $\hat{r}_0 \triangleq 2^{r_0} - 1$, $\hat{P}_k \triangleq \tilde{P}_k - \hat{r}_0 \sum_{m=k+1}^{K} \tilde{P}_m$ which can be considered as an *equivalent* power for user k. Note that in (4), it is implicitly assumed that

$$\tilde{P}_k \geq \hat{r}_0 \sum_{m=k+1}^{K} \tilde{P}_m, \ \forall k \in [1:K-1]. \tag{5}$$

This power constraint is widely incorporated in general for NOMA systems as in [11], [13], [15], [18], where more power is allocated to a user with weak channel gains to guarantee user fairness.

The power constraint can be rewritten as [15], [16]

$$\sum_{k=1}^{K} (\hat{r}_0 + 1)^{k-1} \hat{r}_0 \hat{P}_k \leq P. \tag{6}$$

Furthermore, the throughput of user k is denoted as

$$\begin{aligned}F_k(\{\hat{P}_k\}) &\triangleq r_0(1 - \mathbb{P}_k)\\ &= r_0 \exp\left(-\max\left\{\frac{\hat{r}_0 d_k^{\beta}}{\hat{P}_1}, \cdots, \frac{\hat{r}_0 d_k^{\beta}}{\hat{P}_k}\right\}\right). \end{aligned} \tag{7}$$

[2] Note that setting different fixed target rates for different users can improve the sum throughput. However, user fairness will be affected by such a different-rate scheme. For instance, absolute fairness is difficult to be achieved if the data rates of the users are not the same. Motivated by this, the same fixed rate r_0 for each user is assumed in this paper.

To investigate the sum throughput maximization with α-fairness, we formulate the following optimization problem:

$$\text{(F.P1)} \ \max_{\{\hat{P}_k\}} \sum_{k=1}^{K} u_\alpha\left(F_k(\{\hat{P}_k\})\right) \tag{8a}$$

$$\text{s.t. (6)}, \ \hat{P}_k \geq 0, \ k \in [1:K]. \tag{8b}$$

B. NOMA with Perfect CSIT

In the scenario of perfect CSIT in each block, user's data rates can be adapted according to the channel conditions without any outage. However, the base station needs to estimate each channel gain based on pilot symbols transmitted by the users, which is different from the scenario of statistical CSIT assumption in the previous subsection. The user order is determined based on instantaneous CSI at the beginning of each fading block. It is assumed without loss of generality that $H_1 \leq H_2 \leq \cdots \leq H_K$. The instantaneous rate for user k can be expressed as [21]

$$R_k(\{\tilde{P}_i\}) = \ln\left(\frac{1 + H_k \sum_{i=k}^{K} \tilde{P}_i}{1 + H_k \sum_{i=k+1}^{K} \tilde{P}_i}\right), \ k \in [1:K], \tag{9}$$

where the rate is measured in nats per channel user (NPCU). Note that NPCU has been adopted here for mathematical brevity, however, it can be easily converted into BPCU. The ergodic sum rate can be expressed as $\mathbb{E}\left[\sum_{k=1}^{K} R_k\right]$.

To determine the optimal power allocation to maximize the instantaneous sum rate with α-fairness, we formulate the following optimization problem:

$$\text{(R.P1)} \ \max_{\{\tilde{P}_i\}} \sum_{k=1}^{K} u_\alpha(R_k(\{\tilde{P}_i\})) \tag{10a}$$

$$\text{s.t.} \ \sum_{i=1} \tilde{P}_i \leq P, \tag{10b}$$

$$\tilde{P}_i \geq 0, \ i \in [1:K]. \tag{10c}$$

Remark 1: Although the Rayleigh fading channel model is considered in this paper, the formed optimization problems can be easily extended to more practical channel models, such as the widely used Saleh-Valenzuela multi-path model [38]. In particular, the extensions of problem (R.P1) to the other channel models are straightforward as the instantaneous rates in (9) are also valid for any other channel distributions; whereas the outage probabilities in (4) as well as problem (F.P1) should be modified according to the channel distribution. The study of the other multi-path channel models is out of the scope of this paper.

Remark 2: Note that the power allocation problems with α-fairness will be more complicated for the scenario of multiple antennas at the base station, where the optimal user ordering scheme in the SIC process is still an open problem for MIMO-NOMA [7]–[9]. Thus, for MIMO-NOMA with α-fairness, a possible solution approach is to utilize the sub-optimal user ordering schemes in [7]–[9], and then form the precoding optimization problems at the base station. More details of MIMO-NOMA with α-fairness are out of the scope of this paper, which would be an interesting future direction.

III. Optimal Power Allocation with Statistical CSIT

In this section, we solve problem (F.P1) to obtain the optimal power allocation scheme for sum throughput maximization with α-fairness.

A. Problem Transformation

In this subsection, we first convert the problem (F.P1) into a more simple tractable optimization framework. As the first step in this transformation, we can prove the following inequality condition on the optimal power allocation of problem (F.P1) [15], [16]:

$$\hat{P}_1 \geq \hat{P}_2 \geq \cdots \geq \hat{P}_K. \tag{11}$$

The details of the proof are omitted here for simplicity. In addition, F_k in (7) can be simplified as

$$F_k(\hat{P}_k) = r_0 \exp\left(-\frac{\hat{r}_0 d_k^{\beta}}{\hat{P}_k}\right). \tag{12}$$

By denoting $P_k \triangleq \hat{P}_k/(\hat{r}_0 d_k^{\beta})$, F_k in (12) can be represented as

$$F_k(P_k) = r_0 \exp\left(-\frac{1}{P_k}\right). \tag{13}$$

On the other hand, the constraints in (6) and (11) can be rewritten as

$$\hat{r}_0^2 \sum_{k=1}^{K} \Gamma_k P_k \leq P, \text{ where } \Gamma_k \triangleq (\hat{r}_0+1)^{k-1} d_k^{\beta}, \tag{14}$$

$$d_1^{\beta} P_1 \geq d_2^{\beta} P_2 \geq \cdots \geq d_K^{\beta} P_K, \tag{15}$$

respectively. Now, problem (F.P1) can be reformulated as

$$\text{(F.P2)} \ \max_{\{P_k\}} \sum_{k=1}^{K} u_{\alpha}(F_k(P_k)) \tag{16a}$$

$$\text{s.t. (14) and (15)}, \ P_k \geq 0, \ k \in [1:K]. \tag{16b}$$

B. Optimal Power Allocation

In this subsection, we solve the power allocation problem (F.P2) for different cases with the corresponding values of α. Note that it is assumed that the distances of the users are significantly different from each other, such that[3]

$$\frac{d_i^{\beta}}{d_j^{\beta}} > \frac{(\hat{r}_0+1)^{j-1}}{(\hat{r}_0+1)^{i-1}}, \text{ i.e., } \Gamma_i > \Gamma_j, \ \forall i < j. \tag{17}$$

1) Case $0 \leq \alpha < 1$: In this case, based on (3) and (13), problem (F.P2) can be expressed as

$$\text{(F.P3)} \ f_{(F.P3)} \triangleq \max_{\{P_k\}} \sum_{k=1}^{K} \exp\left(-\frac{1-\alpha}{P_k}\right) \tag{18a}$$

$$\text{s.t. (14) and (15)}, \ P_k \geq 0, \ k \in [1:K]. \tag{18b}$$

Problem (F.P3) is challenging to solve due to the non-convex objective function. To tackle this issue, we first present the following propositions on the objective function and the optimal solution.

Proposition 1: When $0 \leq \alpha \leq 1$, the function $G(x) \triangleq \exp\left(-\frac{1-\alpha}{x}\right)$ is convex for $x \in [0, \frac{1-\alpha}{2})$, and concave for $x \in [\frac{1-\alpha}{2}, \infty)$.

Proof: The second derivative of $G(x)$ can be derived as

$$G''(x) = -\frac{2(1-\alpha)}{x^3}\exp\left(-\frac{1-\alpha}{x}\right) + \frac{(1-\alpha)^2}{x^4}\exp\left(-\frac{1-\alpha}{x}\right). \tag{19}$$

Thus, one can observe that $G''(x) > 0$ if $x \in [0, \frac{1-\alpha}{2})$, $G''(x) = 0$ if $x = \frac{1-\alpha}{2}$, and $G''(x) < 0$ if $x \in [\frac{1-\alpha}{2}, \infty)$. ■

Proposition 2: At the optimal solution of problem (F.P3), $P_k^* \geq P_{k-1}^*$, $\forall k \in [2:K]$.

Proof: This proposition can be proven by *reduction to absurdity*. Suppose that for the optimal power $\{P_k^*\}$ of problem (F.P3), there exist i and j, $i,j \in [1:K]$, such that $i < j$ and $P_i^* > P_j^*$. Now, consider another power pair $(P_i, P_j) \triangleq (P_j^*, P_i^* + \epsilon)$, where we define

$$\epsilon \triangleq (P_i^* - P_j^*)\left(\frac{\Gamma_i}{\Gamma_j} - 1\right), \tag{20}$$

such that

$$P_i\Gamma_i + P_j\Gamma_j = P_i^*\Gamma_i + P_j^*\Gamma_j. \tag{21}$$

From (17), one can observe that $\epsilon > 0$. Furthermore, it can be obtained that

$$G(P_i^*) + G(P_j^*) < G(P_i) + G(P_j), \tag{22}$$

since $\epsilon > 0$ and $G(x)$ is a monotonically increasing function, which contradicts with the optimality of (P_i^*, P_j^*). This completes the proof of this proposition. ■

Proposition 3: For the optimal solution of problem (F.P3), if there are k_0 power values, P_k^*'s, that are below $\frac{\alpha-1}{2}$, then the constraint in (15) is binding for these power values, i.e., $d_i^{\beta} P_i^* = d_j^{\beta} P_j^*$, $\forall i,j \in [1:k_0]$.

Proof: Please refer to Appendix A. ■

Remark 3: Based on Propositions 3, it follows that the optimal solution of problem (F.P3) should have the following structure: there are k_0 power values, $(P_1, \cdots, P_{k_0})$, satisfying $P_k < \frac{\alpha-1}{2}$ and $P_k = \frac{d_1^{\beta}}{d_k^{\beta}} P_1$, $\forall k \in [1:k_0]$, and the rest of $(K-k_0)$ power values satisfying $P_k \geq \frac{\alpha-1}{2}$, $\forall k \in [k_0+1, K]$. Therefore, the maximum value of the objective function can be expressed as $f_{(F.P3)}^* = \max_{k_0, P_1, \{P_{k_0+1}, \cdots, P_K\}} \sum_{k=1}^{k_0} G\left(\frac{d_1^{\beta}}{d_k^{\beta}} P_1\right) + \sum_{k=k_0+1}^{K} G(P_k)$.

[3] The assumption on the distances in (17) is reasonable in practical NOMA systems. For example, multiuser superposition transmission (MUST), a downlink two-user version of NOMA, has been included in 3rd generation partnership project long-term evolution advanced (3GPP-LTE-A) networks [39]. For MUST, the base station selects two users, which are far from and near to the base station, respectively. Obviously, the distance difference between these two selected users is significantly large.

From problem (F.P3) and Remark 3, one can observe that if (k_0, P_1) is fixed, the optimal values of $(P_{k_0+1}, \cdots, P_K)$ can be obtained by solving the following optimization problem:

$$\text{(F.P4)}\ f_{(F.P4)}(k_0, P_1) \triangleq \max_{\{P_{k_0+1},\cdots,P_K\}} \sum_{k=k_0+1}^{K} G(P_k) \tag{23a}$$

$$\text{s.t. } \hat{r}_0^2 \sum_{k=k_0+1}^{K} \Gamma_k P_k \leq P - \hat{r}_0 P_1 d_1^{\beta}((\hat{r}_0+1)^{k_0}-1), \tag{23b}$$

$$d_{k_0+1}^{\beta} P_{k_0+1} \geq \cdots \geq d_K^{\beta} P_K, \tag{23c}$$

$$P_k \geq \frac{\alpha-1}{2},\ k \in [k_0+1:K]. \tag{23d}$$

where $(k_0, P_1) \in \mathcal{S}$, and $\mathcal{S}$ is defined as

$$\mathcal{S} \triangleq \Big\{(k_0, P_1):\ k_0 \in [0:K],\ 0 \leq P_1 \leq \frac{(\alpha-1)d_{k_0}^{\beta}}{2d_1^{\beta}},$$
$$\hat{r}_0 P_1 d_1^{\beta}((\hat{r}_0+1)^{k_0}-1) + \frac{\hat{r}_0^2(\alpha-1)}{2} \sum_{k=k_0+1}^{K} \Gamma_k \leq P\Big\}, \tag{24}$$

such that constraints (23b) and (23d) can be satisfied.

Closed-form solution to problem (F.P4) is in general not possible. However, it can be easily shown that problem (F.P4) is convex since $G(x)$ is concave when $x \in \left[\frac{1-\alpha}{2}, \infty\right)$ as presented in Proposition 1. Thus, for a fixed pair (k_0, P_1), problem (F.P4) will be solved later in Section VI with the help of corresponding numerical solvers.

The following work is to find optimal values of k_0 and P_1, denoted as as (k_0^*, P_1^*), which can be expressed as

$$(k_0^*, P_1^*) = \arg \max_{(k_0,P_1)\in\mathcal{S}} \sum_{k=1}^{k_0} G\left(\frac{d_1^{\beta}}{d_k^{\beta}} P_1\right) + f_{(F.P4)}^*(k_0, P_1), \tag{25}$$

where $f_{(F.P4)}^*(k_0, P_1)$ is the maximum value of the objective function in problem (F.P4) for a fixed pair (k_0, P_1). Specifically, in order to find (k_0^*, P_1^*), a two-dimensional exhaustive search over k_0 and P_1 should be carried out. Since k_0 is an integer in $[0:K]$ as shown in (24), the computational complexity of this two-dimensional exhaustive search is $O((K+1)\delta)$, where δ is the step size when searching P_1^* (i.e., δ denotes the searching accuracy of P_1^*).

2) Case $\alpha = 1$: In this case, based on (3) and (13), problem (F.P2) can be expressed as

$$\text{(F.P5)}\ \min_{\{P_k\}} \sum_{k=1}^{K} \frac{1}{P_k} \tag{26a}$$

$$\text{s.t. (14) and (15)},\ P_k \geq 0,\ k \in [1:K]. \tag{26b}$$

The following lemma provides the closed-form expression for the optimal solution of the problem.

Lemma 1: The optimal solution for problem (F.P5) is given by

$$P_k = \frac{1}{\hat{r}_0\sqrt{\omega\Gamma_k}},\ \text{where } \omega = \left(\frac{\hat{r}_0}{P}\sum_{k=1}^{K}\sqrt{\Gamma_k}\right)^2. \tag{27}$$

Proof: Please refer to Appendix B. ∎

3) Case $\alpha > 1$: In this case, based on (3) and (13), problem (F.P2) can be expressed as

$$\text{(F.P7)}\ \min_{\{P_k\}} \sum_{k=1}^{K} \exp\left(\frac{\alpha-1}{P_k}\right) \tag{28a}$$

$$\text{s.t. (14) and (15)},\ P_k \geq 0,\ k \in [1:K]. \tag{28b}$$

The convexity of this problem can be verified through deriving the Hessian matrix of the objective function. Obviously a closed-form expression for the optimal solution of problem (F.P7) is difficult to obtain, however, this problem will be solved later in Section VI using corresponding numerical solvers. On the other hand, we can verify that, when $\alpha \to \infty$, absolute user fairness in terms of throughput can be obtained in the following Lemma.

Lemma 2: When $\alpha \to \infty$, $F_i(P_i^*) = F_j(P_j^*)$, $\forall i, j \in [1:K]$, where $(P_1^*, \cdots, P_K^*)$ is the optimal solution of problem (F.P7).

Proof: Please refer to Appendix C. ∎

IV. Optimal Power Allocation with Perfect CSIT

In this section, we determine the optimal power allocation to maximize the instantaneous sum rate with α-fairness by solving problem (R.P1).

A. Problem Transformation

By denoting K variables as: $b_k \triangleq \sum_{i=k}^{K} \tilde{P}_i$, $k \in [1:K]$, from (9), the instantaneous rate of user k can be expressed as

$$R_k(b_k, b_{k+1}) = \ln\left(\frac{1+H_k b_k}{1+H_k b_{k+1}}\right),\ k \in [1:K], \tag{29}$$

where it is defined $b_{K+1} \triangleq 0$ for the sake of brevity.

In addition, the power constraint in (10b) is obviously binding at the optimal solution of problem (R.P1), i.e., $\sum_{i=1}^{K} \tilde{P}_i = P$ and $b_1 = P$. Thus, problem (R.P1) can be reformulated into the following optimization framework:

$$\text{(R.P2)}\ \max_{\{b_2,\cdots,b_K\}} \sum_{k=1}^{K} u_{\alpha}(R_k(b_k, b_{k+1})) \tag{30a}$$

$$\text{s.t. } b_k \geq b_{k+1},\ \forall k \in [1:K], \tag{30b}$$

$$b_1 = P,\ b_{K+1} = 0. \tag{30c}$$

The following lemma is required to represent the KKT conditions of problem (R.P2).

Lemma 3: The KKT conditions of problem (R.P2) can be transformed into the following K equations:

$$f_{1,k}(b_k, b_{k+1}, b_{k+2}) \triangleq \frac{R_{k+1}\left(b_{k+1}, b_{k+2}\right)}{R_k\left(b_k, b_{k+1}\right)} - \left(\frac{b_{k+1}+\frac{1}{H_k}}{b_{k+1}+\frac{1}{H_{k+1}}}\right)^{1/\alpha} = 0,$$
$$b_{k+2} < b_{k+1} < b_k,\ \forall k \in [1:K-1], \tag{31}$$

Proof: Please refer to Appendix D. ■

Remark 4: From Lemma 3, it can be observed that absolute user fairness in terms of instantaneous rate can be obtained when $\alpha \to \infty$. Specifically, $R_{k+1} = R_k$ holds in (31), $\forall k \in [1:K-1]$, as long as $\alpha \to \infty$.

To obtain the solution through the KKT conditions of problem (R.P2), the following theorem is presented.

Theorem *1:* There is only a unique solution for the $K-1$ equations in (31), denoted as $(\hat{b}_2, \cdots, \hat{b}_K)$.

Proof: Please refer to Appendix E. ■

Remark 5: Theorem 1 shows that the KKT conditions of problem (R.P2) are sufficient to determine the optimal solution, i.e., $(\hat{b}_2, \cdots, \hat{b}_K)$ is the optimal solution of problem (R.P2). Thus, the conventional interior point algorithm can be utilized to solve problem (R.P2). Alternatively, a simple algorithm can be developed to solve the $K-1$ equations in (31), as provided in the next subsection.

B. Proposed Algorithm

In this subsection, a simple algorithm is developed to solve $K-1$ equations in (31), which yields the optimal solution of the original problem in (30).

Lemma 4: For a fixed pair (b_k, b_{k+2}), $k \in [1:K-1]$, only a unique b_{k+1} satisfies the k-th equation in (31), which is the unique root of the following function:

$$\tilde{f}_{1,k}(x) \triangleq \frac{\ln\left(\frac{1+H_{k+1}x}{1+H_{k+1}b_{k+2}}\right)}{\ln\left(\frac{1+H_k b_k}{1+H_k x}\right)} - \left(\frac{x+\frac{1}{H_k}}{x+\frac{1}{H_{k+1}}}\right)^{1/\alpha},\ b_{k+2} < x < b_k, \tag{32}$$

where function $\tilde{f}_{1,k}$ is defined as the same as $f_{1,k}$ in (31), except that $\tilde{f}_{1,k}$ is a single-variable function whereas $f_{1,k}$ is a multi-variable function.

Proof: We will show that function $\tilde{f}_{1,k}(x)$ is monotonically increasing when $b_{k+2} < x < b_k$, and $\tilde{f}_{1,k}(x) = 0$ has only a unique root over (b_{k+1}, b_k). Specifically, $\frac{x+\frac{1}{H_k}}{x+\frac{1}{H_{k+1}}} = 1 + \frac{\frac{1}{H_k}-\frac{1}{H_{k+1}}}{x+\frac{1}{H_{k+1}}}$, which decreases with x for $x > 0$. Recalling (32), $\tilde{f}_{1,k}(x)$ is obviously a monotonically increasing function when $b_{k+2} \le x \le b_k$. Furthermore, $\tilde{f}_{1,k}(x) < 0$ as $x \to b_{k+2}$; $\tilde{f}_{1,k}(x) \to +\infty$ as $x \to b_k$. Therefore, equation $\tilde{f}_{1,k}(x) = 0$ has only a unique root, which is denoted as b^*_{k+1}. Based on the definitions of $\tilde{f}_{1,k}$ and $f_{1,k}$, b^*_{k+1} is the unique value that satisfies the k-th equation in (31) for a fixed pair (b_k, b_{k+2}). ■

Remark 6: As discussed in the proof of Lemma 4, $\tilde{f}_{1,k}(x)$ is a monotonically increasing function, hence a simple bisection method can be utilized to determine the root of equation (32), which is summarized in Algorithm I.

Algorithm I: Root Search for Fixed (b_k, b_{k+2}) in (32)

1: Initialize $b_{lb} = b_{k+2}$, $b_{ub} = b_k$;
2: **while** ($|\tilde{f}_{1,k}(b_{k+1})| > \epsilon_1$) **do**
3: Set $b_{k+1} = (b_{lb} + b_{ub})/2$, and calculate $\tilde{f}_{1,k}(b_{k+1})$;
4: **if** $\tilde{f}_{1,k}(b_{k+1}) > \epsilon_1$ **then** $b_{ub} = b_{k+1}$;
5: **else** $b_{lb} = b_{k+1}$;
6: **until** $|\tilde{f}_{1,k}(b_{k+1})| < \epsilon_1$;

Algorithm II: Proposed Alternate Algorithm for Problem (R.P2)

1) Initialize $t = 1$, $b_k^{(0)} = 0$, $\forall k \in [2:K]$;
2) The t-th iteration:
Set $k = 1$, $b_1^{(t)} = P$, and $b_{K+1}^{(t-1)} = 0$;
Repeat
a) Fix $(b_k^{(t)}, b_{k+2}^{(t-1)})$, then find the root of the k-th equation in (31), i.e., b^*_{k+1}, using Algorithm I.
b) Set $b_{k+1}^{(t)} = b^*_{k+1}$;
c) Update $k = k+1$.
Until $k = K-1$;
3) Update $t = t+1$ and repeat Step 2) until Norm $\left[\mathbf{f}_1^{(t)}\right] \le \epsilon_2$;

Motivated by Lemma 4, a simple AO algorithm is summarized in Algorithm II, where $b_k^{(t)}$ denotes the value of b_k in the t-th iteration. The basic idea is to alternately solve the k-th equation in (31) by fixing the other corresponding variables. Specifically, in each iteration t, the root of the k-th equation in (31) is determined using Algorithm I for a fixed pair $(b_k^{(t)}, b_{k+2}^{(t-1)})$, $\forall k \in [1:K-1]$. By denoting such a root as b^*_{k+1}, the value of b_{k+1} in iteration t is updated as $b_{k+1}^{(t)} = b^*_{k+1}$, until the required accuracy is achieved. Note that Norm $\left[\mathbf{f}_1^{(t)}\right] \le \epsilon_2$ is utilized as the stopping criterion, where

$$\mathbf{f}_1^{(t)} \triangleq \left(f_{1,1}\left(b_1^{(t)}, b_2^{(t)}, b_3^{(t)}\right), \cdots, f_{1,K-1}\left(b_{K-1}^{(t)}, b_K^{(t)}, b_{K+1}^{(t)}\right)\right), \tag{33}$$

and Norm $[\cdot]$ is the Euclidean distance of a vector. In addition, the KKT conditions can be obviously satisfied as Norm $\left[\mathbf{f}_1^{(t)}\right] \to 0$, as provided in Lemma 3.

Next, we analyze the convergence and optimality of the proposed algorithm. To verify the convergence of the algorithm, the following theorem is required.

Theorem *2:* For Algorithm II, $b_k^{(t)}$ is monotonically increasing with t, $\forall k \in [2:K]$.

Proof: Please refer to Appendix F. ■

Lemma 5: The proposed AO algorithm in Algorithm II converges.

Proof: From Theorem 2, it can be seen that $b_k^{(t)}$ increases with t and its upper bound can be defined by $b_1^{(t)} = P$. Therefore, $\lim_{t\to\infty} b_k^{(t)}$ exists, $\forall k \in [2:K]$, and the proposed algorithm in Algorithm II converges. ■

To validate the optimality of Algorithm II, the following lemma is provided.

Lemma 6: The proposed algorithm achieves the optimal solution for problem (R.P2).

Proof: Since Algorithm II converges as shown in Theorem 2 and Lemma 5, limit $\bar{b}_k = \lim_{t\to\infty} b_k^{(t)}$ exists, $k \in [2 : K]$, and $\tilde{f}_{1,k}(\bar{b}_{k+1}) = 0$ for the given pair $(\bar{b}_k, \bar{b}_{k+2})$ in (32), $\forall k \in [1 : K-1]$. Thus, from Lemma 3, it can be observed that solution $(\bar{b}_2, \cdots, \bar{b}_K)$ satisfies the KKT conditions of problem (R.P2). Furthermore, we know from Theorem 1 that solution $(\bar{b}_2, \cdots, \bar{b}_K)$ is the unique solution of the KKT functions in (31), i.e., Algorithm II yields the optimal solution for problem (R.P2). ■

C. Complexity of Algorithm II

The complexity of Algorithm II is mainly determined by two crucial parameters: the number of arithmetic operations in each iteration and the speed of convergence.

For each iteration, the number of arithmetic operations involved in the proposed algorithm is $O((K-1)\log(1/\epsilon_1))$ since $K-1$ bisection searches are required with ϵ_1 solution accuracy in Algorithm I. In contrary, the conventional interior point algorithm requires $O((K-1)^3)$ arithmetic operations for each iteration [40], which does not have any impact by ϵ_1, however significantly increases with K.

The convergence speed of Algorithm II is difficult to estimate due to the very complicated expression of the functions in (31). However, we demonstrate the speed of the convergence with the help of numerical results later in Section VI, which reveals that the proposed algorithm converges faster than the interior point algorithm in most scenarios.

V. Discussion

In this section, we discuss an appropriate evaluation criterion of the proposed α-fairness scheme. The α-fairness is a qualitative fairness measure of user throughput or instantaneous rate [30]. To evaluate quantitative fairness, there is a widely used measurement, known as "Jain's Index" or "Fairness Index" (FI), which is defined as [37]

$$\mathrm{FI}(\{x_k\}) \triangleq \left(\sum_{k=1}^{K} x_k\right)^2 \Big/ \left(K \sum_{k=1}^{K} x_k^2\right), \tag{34}$$

where x_k could be either F_k or R_k, and FI could take any value over $[1/K, 1]$. A larger FI generally represents a higher fairness level; the case FI=1 corresponds to absolute fairness. Moreover, for statistical CSIT, FI turns out to be long-term fairness within a large number of blocks; while for perfect CSIT, FI represents short-term fairness within each block.

In general, different values of FI can be achieved by adjusting α [32]. For instance, as shown in Lemma 2 and Remark 4, $\mathrm{FI}(\{F_k\}) = \mathrm{FI}(\{R_k\}) = 1$ as $\alpha \to \infty$. Therefore, we can appropriately choose α to achieve the fairness index requirement (FIr), where FIr$\in [1/K, 1]$. The corresponding optimization problem can be defined as follows:

$$\max_{0\le\alpha\le1} \sum_{k=1}^{K} x_k^* \tag{35a}$$

$$\text{s.t. } \mathrm{FI}(\{x_k^*\}) \ge \mathrm{FIr},\ \alpha \ge 0, \tag{35b}$$

where for a given α, $x_k^* = F_k(P_k^*)$ in the case of statistical CSIT and $x_k^* = R_k(b_k^*, b_{k+1}^*)$ in the case of perfect CSI. Note that the optimal solutions of problem (F.P2) and (R.P2) are denoted as $\{P_k^*\}$ and $\{b_k^*\}$, respectively.

Note that increasing α does not necessarily increase FI as shown in [41]. Thus, in general, a one-dimensional search is required to find the optimal value of α, denoted as α^*, for the problem defined in (35). However, in most scenarios, $\sum_{k=1}^{K} x_k^*$ and FI decreases and increases with α, respectively, as shown in many existing works (e.g., [32], [34], [42]). Thus, a simple bisection method will be utilized to find α^* in most scenarios later in Section VI.

VI. Numerical Results

In this section, computer simulation results are provided to evaluate the sum throughput and the ergodic sum rate of NOMA with α-fairness. In these simulations, some parameters for the considered NOMA system are set as follows. The small scale fading gain is Rayleigh distributed, i.e., $g_i \sim \mathcal{CN}(0, 1)$. Furthermore, the noise at each user is assumed to be an additive white Gaussian variable with zero mean and unit variance. In addition, the distance between the base station and user k is defined as $d_k = 1.5^{K-k}$, and the path loss exponent is chosen as 2 to reflect a favorable propagation condition. [4] Since the variance of noise power is unity, the transmit signal-to-noise-ratio (SNR) is equivalent to the transmit power P.

A. Benchmark Schemes

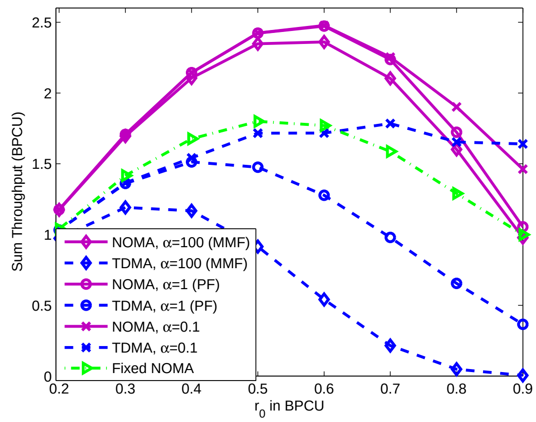
Two benchmark transmission schemes of TDMA and NOMA with fixed power allocation (i.e., fixed NOMA) are considered as explained in the following.

1) TDMA Scheme: The TDMA transmission method is chosen as one of the benchmark schemes in this evaluation, as it is equivalent to any orthogonal MA scheme [43, Sec. 6.1.3]. For TDMA transmission, each fading block is assumed to be equally divided into K time slots, where user k occupies the k-th time slot. By defining the power allocated to user k as P_k^T, the power constraint for the TDMA scheme can be expressed as

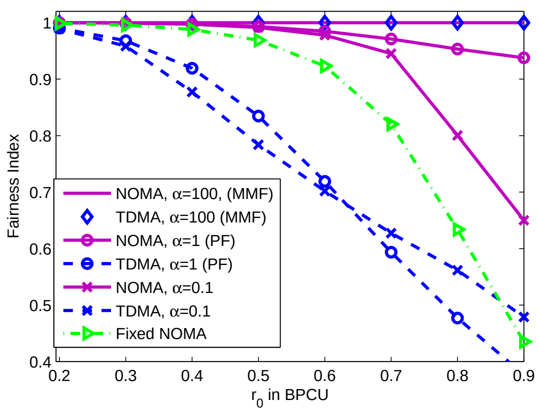
$$\frac{1}{K}\sum_{k=1}^{K} P_k^T \le P.$$

Now, similar to the problems (F.P1) an (R.P1) in Section II, one can formulate two power allocation problems for the TDMA scheme with statistical and perfect CSIT, respectively. Furthermore, these two new TDMA power allocation problems can be solved using similar approaches as in Section III and IV. The details of these approaches are omitted here due to space limitations.

[4] The parameter settings $d_k = 1.5^{(K-k)}$ and $\beta = 2$ show that the assumption in (17) is valid even when the distance-ratio d_{k+1}/d_k and the path loss exponent β are small or moderate. Due to space limitation, the other choices of parameters have not been considered in this paper.



(a) Sum throughput vs r_0



(b) Fairness Index vs r_0

Fig. 1. Sum throughput and fairness index (FI) vs the transmission rate r_0 in BPCU, where SNR = 20 dB, $K = 6$, α = 100, 1, 0.1; MMF and PF denotes max-min fairness and proportional fairness, respectively.

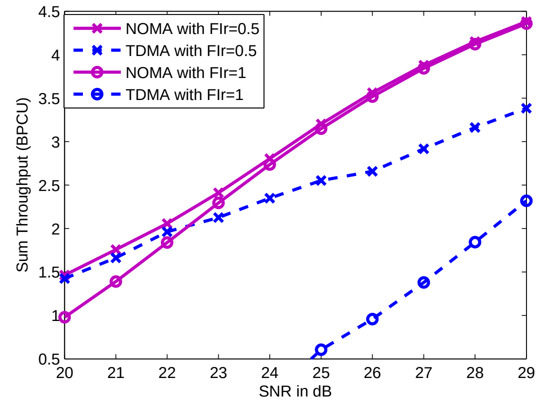


Fig. 2. Sum throughput vs SNR, where $r_0 = 0.9$ BPCU, $K = 6$, FIr = 0.5, 1.

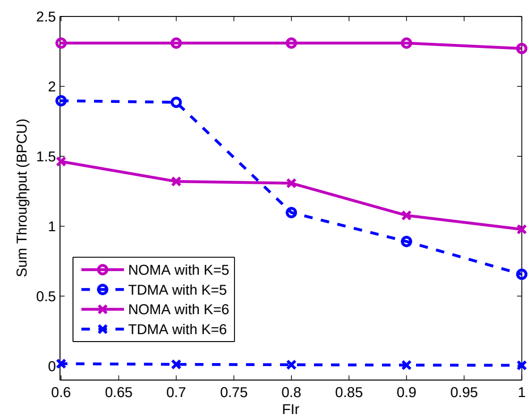


Fig. 3. Sum throughput vs the fairness index requirement (FIr), where $r_0 =$ 0.9 BPCU, SNR = 20 dB. K = 5, 6.

2) Fixed NOMA: In order to demonstrate the benefits of power allocation, NOMA with fixed power allocation is used as another benchmark scheme. In particular, the NOMA transmission scheme in Section II is also utilized, but the power allocation scheme is fixed as

$$\tilde{P}_k = \frac{2^{K-k}P}{2^K - 1}, \ \ k \in [1:K],$$

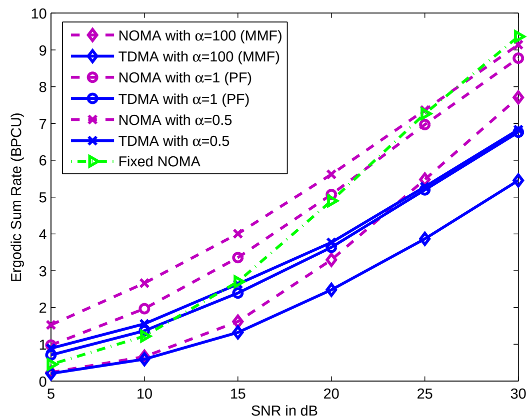
for both statistical and perfect CSIT. Note that this fixed power allocation scheme is similar to the one in [14] with a slight modification.

B. Statistical CSIT

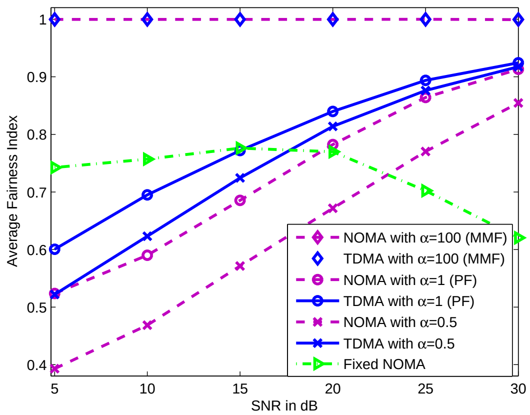
This subsection focuses on the sum throughput performance of NOMA with α-fairness and statistical CSIT. Figs. 1(a) and 1(b) compare the sum throughput and FI of NOMA employing optimal power allocation proposed in Section III with the benchmark schemes as a function of the transmission rate r_0, where we set $K = 6$, SNR = 20 dB, and α = 100, 1, 0.1. [5] As seen in these two sub-figures, NOMA with optimal power allocation enjoys both larger sum throughput and FI than the fixed NOMA scheme and the TDMA scheme with optimal power allocation, for $\alpha = 100$ or 1. Moreover, increasing α decreases sum throughput and increases FI for NOMA, and absolute fairness can be achieved with $\alpha = 100$, which supports the discussions in Lemma 2. For $\alpha = 0.1$, although TDMA with optimal power allocation has a larger sum throughout when $r_0 = 0.9$ BPCU as shown in Fig. 1(a), its FI (0.48) is lower than the one achieved by NOMA (0.65). This is due to the fact that an additional power constraint is imposed on NOMA in (5), which might reduce the sum throughput with small values of α, however, it can guarantee the fairness level of NOMA. From Fig. 1(b), one can observe that decreasing α from 1 to 0.1 results in the improvement of FI for TDMA when $r_0 \geq 0.6$, which is consistent with the conclusion made in [41] that increasing α does not necessarily increase FI.

For a fair comparison between NOMA and TDMA schemes, the same FI is required in Figs. 2 and 3. Specifically, we utilize α to adjust the value of FI as shown in problem (35), where a bisection search is adopted by NOMA, whereas an exhaustive search needs to be adopted by TDMA since its FI does not necessarily increase with α as shown in Fig. 1(b). In Fig. 2, the sum throughput is depicted as a function of

[5] Note that max-min fairness (MMF) and proportional fairness (PF) can be achieved when $\alpha = 100$ (i.e., α is sufficiently large) and $\alpha = 1$, respectively [30].



(a) Ergodic sum rate vs SNR



(b) Average fairness Index vs SNR

Fig. 4. Ergodic sum rate and average FI vs SNR in dB, where $K = 5$, α = 100, 1, 0.5.

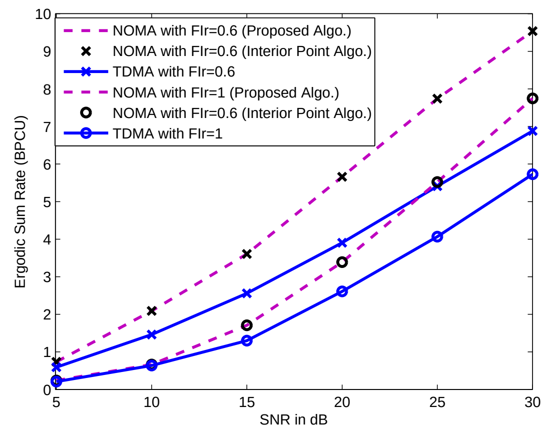


Fig. 5. Ergodic sum rate vs SNR in dB, where $K = 5$, SNR = 20 dB, FIr = 0.6, 1.

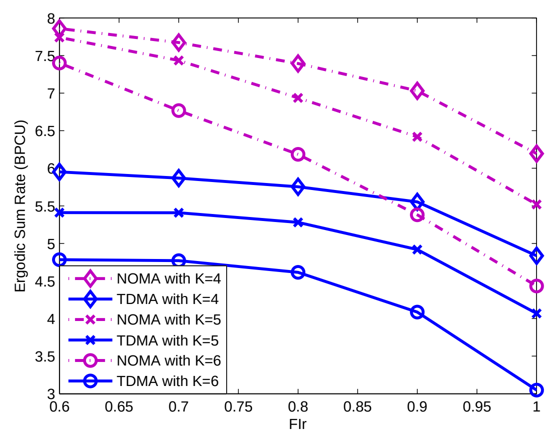


Fig. 6. Ergodic sum rate vs FIr, where SNR = 20 dB, and K = 4, 5, 6.

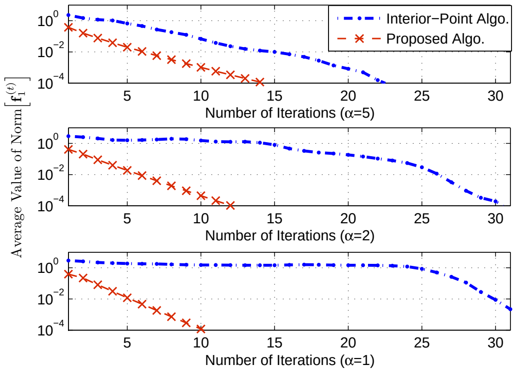
SNR, where $r_0 = 0.9$ BPCU, $K = 6$, and the required FI is set as FIr = 0.5 or 1. In Fig. 3, the sum throughput is presented as a function of FIr, where $r_0 = 0.9$ BPCU, SNR = 20 dB, $K = 5$ or 6. From these two figures, one can observe that moderate or high FIr significantly decreases the sum throughput of TDMA, however, it has less impact on NOMA, i.e., NOMA provides a significant performance gain compared to TDMA with moderate or high FIr.

C. Perfect CSIT

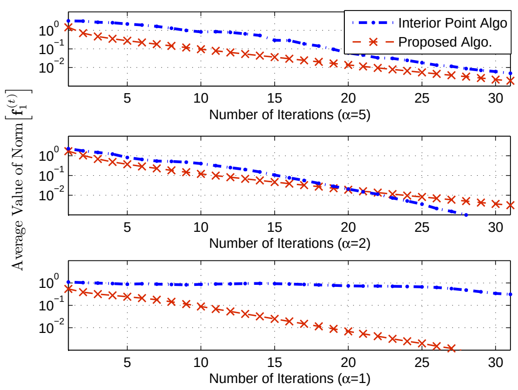
This subsection focuses on the ergodic rate performance of NOMA with α-fairness and perfect CSIT. Figs. 4(a) and 4(b) compare the sum throughput and FI of NOMA employing optimal AO power allocation algorithm proposed in Section IV with the benchmark schemes as a function of SNR, where the parameters are set as $K = 5$ and α = 100, 1, 0.5. As seen in these two sub-figures, the fixed NOMA has a large ergodic sum rate but a very poor average FI when SNR = 30 dB. On the other hand, NOMA with optimal power allocation has a larger ergodic sum rate with a low average FI compared to the TDMA scheme. For both NOMA and TDMA, increasing α decreases ergodic sum rate, however increases average FI. The absolute fairness can be achieved with $\alpha = 100$, which validates the discussions in Remark 4.

In order to make a fair comparison between NOMA and TDMA schemes, the same FI is required in Figs. 5 and 6. Specifically, we utilize α to adjust the value of FI as shown in problem (35), where a bisection search is adopted by both NOMA and TDMA schemes. In Fig. 2, the ergodic sum rate is depicted as a function of SNR, where $K = 5$, and FIr = 0.6 or 1. In Fig. 3, the ergodic sum rate is depicted as a function of FIr, with SNR = 20 dB, $K = 4$, 5 or 6. As seen in these two figures, one can observe that NOMA provides a significant performance gain than the TDMA scheme in terms of ergodic sum rate at the same required fairness level. Moreover, the proposed power allocation algorithm achieves the same ergodic sum rate as the conventional interior point algorithm, as shown in Fig. 5.

Figs. 7(a) and 7(b) compare convergence speeds of the proposed algorithm in Section IV (i.e., Algorithm II) and the conventional interior point algorithm with $K = 4$ and 8, respectively. Since Norm $\left[\mathbf{f}_1^{(t)}\right]$ is utilized as the stopping criterion for each fading block, we depict its average value as a function of the number of iterations, where the required accuracy of Algorithm I (involved in Algorithm II) is set as $\epsilon_1 = 10^{-5}$, and α = 5, 2, 1. As evidenced by these two



(a) $K = 4$



(b) $K = 8$

Fig. 7. The average value of Norm$\left[\mathbf{f}_1^{(t)}\right]$ vs the number of iterations, where SNR = 20 dB; K = 4, 8; α = 5, 2, 1.

sub-figures, one can observe that the proposed algorithm converges more faster than the interior point algorithm in most scenarios, except with $K = 8$, $\alpha = 2$ and the number of iterations is larger than 20.

VII. Conclusions

This paper investigated α-fairness based power allocation schemes for sum throughput and ergodic rate maximization problems in a downlink NOMA system with statistical and perfect CSIT. For statistical CSIT, the outage probability of each user was analyzed, and the power allocation strategy was developed for sum throughput maximization with α-fairness. Specifically, the original non-convex sum throughput maximization problem was converted into an equivalent problem and demonstrated that the transformed equivalent problem is convex for the case of $\alpha \geq 1$. In addition, it was shown that the problem turns out to be convex for $\alpha < 1$ by fixing the first power parameter and the number of power parameters that are below $\frac{1-\alpha}{2}$. Next, the instantaneous sum rate maximization with α-fairness was solved for perfect CSIT, for which it was proven that there exists only one solution to satisfy the corresponding KKT conditions. Then, a simple AO algorithm was developed to solve these KKT equations. As this work only considered single antenna NOMA, an interesting future work is to extend to MIMO NOMA with fairness constrains. Moreover, considering user fairness for the other more practical channel model (e.g., Saleh-Valenzuela multipath model [38]) or considering user fairness over a time-domain window would be also one of the possible future directions.

Appendix A
Proof of Proposition 3

This proposition can be proven by *reduction to absurdity*. Denote the optimal power of problem (F.P3) as $\{P_k^*\}$. Without loss of generality, it can be assumed that there exist $i < j$, $i, j \in [1 : k_0]$, such that the constraint in (15) is not binding, i.e., $d_i^\beta P_i^* > d_j^\beta P_j^*$. From Proposition 2 and the definition of k_0, $0 < P_i^* < P_j^* < \frac{1-\alpha}{2}$ can be obtained. Now, consider another power pair $\left(P_i^* - \epsilon_1, P_j^* + \epsilon_2\right)$, where $(\epsilon_1, \epsilon_2) \triangleq \left(\frac{\epsilon}{\Gamma_i}, \frac{\epsilon}{\Gamma_j}\right)$, and ϵ satisfies

$$0 < \epsilon < \min\left\{\Gamma_i P_i^*, \Gamma_j\left(\frac{1-\alpha}{2} - P_j^*\right), \frac{d_i^\beta P_i^* - d_j^\beta P_j^*}{d_i^\beta/\Gamma_i + d_j^\beta/\Gamma_j}\right\}. \tag{36}$$

Obviously we have

$$0 \leq P_i^* - \epsilon_1 < P_j^* + \epsilon_2 < \frac{1-\alpha}{2}, \tag{37}$$

$$d_i^\beta(P_i^* - \epsilon_1) \geq d_j^\beta(P_j^* + \epsilon_2), \tag{38}$$

$$\Gamma_i(P_i^* - \epsilon_1) + \Gamma_j(P_j^* + \epsilon_2) = \Gamma_i P_i^* + \Gamma_j P_j^*, \tag{39}$$

where (37) implies that the value of k_0 will remain the same by replacing the power pair (P_i^*, P_j^*) by $(P_i^* - \epsilon_1, P_j^* + \epsilon_2)$; (38) and (39) ensure that $(P_i^* - \epsilon_1, P_j^* + \epsilon_2)$ satisfies the power constraints in (14) and (15), respectively. Next, we need to verify that $G(P_i^* - \epsilon_1) + G(P_j^* + \epsilon_2) > G(P_i^*) + G(P_j^*)$, where $G(x)$ is defined in Proposition 1.

Based on *Lagrange mean value theorem*, there exists some $\varepsilon_1 \in (P_i^* - \epsilon_1, P_i^*)$ and $\varepsilon_2 \in (P_j^*, P_j^* + \epsilon_2)$ such that

$$G(P_i^*) - G(P_i^* - \epsilon_1) = \epsilon_1 G'(\varepsilon_1), \tag{40}$$

$$G(P_j^* + \epsilon_2) - G(P_j^*) = \epsilon_2 G'(\varepsilon_2). \tag{41}$$

Note that since $\varepsilon_1 < \varepsilon_2$ and $G''(x) > 0$ if $x \in [0, \frac{1-\alpha}{2})$, as shown in Proposition 1, $G'(\varepsilon_1) < G'(\varepsilon_2)$ holds; furthermore, since $G'(x) = \frac{1-\alpha}{x^2}\exp\left(-\frac{1-\alpha}{x}\right)$, $x \geq 0$, $0 < G'(\varepsilon_1) < G'(\varepsilon_2)$ can be obtained. In addition, $\epsilon_1 < \epsilon_2$ holds since $\Gamma_i > \Gamma_j$ shown in (17). Thus, from (40) and (41), one can observe that $G(P_i^*) - G(P_i^* - \epsilon_1) < G(P_j^* + \epsilon_2) - G(P_j^*)$.

In summary, power pair $\left(P_i^* - \epsilon_1, P_j^* + \epsilon_2\right)$ yields a larger value of the objective function in problem (F.P3), which contradicts with the optimality of $\left(P_i^*, P_j^*\right)$. Therefore, the case $d_i^\beta P_i^* > d_j^\beta P_j^*$ is not optimal, and $d_i^\beta P_i^* = d_j^\beta P_j^*$ holds at the optimal solution of problem (F.P3).

Appendix B
Proof of Lemma 1

To solve problem (F.P5), we first consider the following problem by relaxing the constraint in (15) of problem (F.P5):

$$\text{(F.P6)} \min_{\{P_k\}} \quad \sum_{k=1}^{K} \frac{1}{P_k} \tag{42a}$$

$$\text{s.t. (14)}, \ \ P_k \geq 0, \ \ k \in [1 : K]. \tag{42b}$$

The Lagrangian function for this problem is defined as:

$$\mathcal{L}(\{P_k\},\omega,\{\lambda_k\}) \triangleq \sum_{k=1}^{K}\frac{1}{P_k}+\omega\left[\hat{r}_0^2\sum_{k=1}^{K}\Gamma_k P_k - P\right] - \sum_{k=1}^{K}\lambda_k P_k, \tag{43}$$

where $\omega, \lambda_k \geq 0$ are Lagrange multipliers. The KKT conditions are given by

$$\frac{\partial \mathcal{L}}{\partial P_k} = -\frac{1}{P_k^2}+\hat{r}_0^2\omega\Gamma_k-\lambda_k = 0. \tag{44}$$

In addition, from the complementary slackness conditions (omitted here for simplicity), obviously we have $\lambda_k = 0$ and $\omega > 0$, and the power constraint in (14) is binding. Therefore, from (14) and (44), the optimal solution of problem (F.P6) can be obtained as shown in (27).

From (27), one can observe that $d_k^{\beta}P_k$ decreases with k, which means that the constraint in (15) is satisfied. Thus, problems (F.P5) and (F.P6) have the same optimal solution.

Appendix C
Proof of Lemma 2

Lemma 2 can also be proven by *reduction to absurdity*. Denote the optimal power of problem (F.P7) as $\{P_k^*\}$; based on Proposition 2, it holds that $P_i^* \leq P_j^*, \forall i<j,\ i,j \in [1:K]$. Assume without loss of generality that there exist i and j satisfying $i<j,\ i,j \in [1:K]$, such that $0 < P_i^* < P_j^*$. Consider another power pair $\left(P_i^*+\epsilon_1, P_j^*-\epsilon_2\right)$, where $\epsilon_1+\epsilon_2 < P_j^*-P_i^*$ and $(\epsilon_1,\epsilon_2) \triangleq \left(\frac{\epsilon}{\Gamma_i},\frac{\epsilon}{\Gamma_j}\right)$ for $\epsilon > 0$. Obviously we have

$$P_i^*+\epsilon_1 < P_j^*-\epsilon_2, \tag{45}$$

$$d_i^{\beta}(P_i^*+\epsilon_1) \geq d_j^{\beta}(P_j^*-\epsilon_2), \tag{46}$$

$$\Gamma_i(P_i^*+\epsilon_1)+\Gamma_j(P_j^*-\epsilon_2) = \Gamma_i P_i^*+\Gamma_j P_j^*, \tag{47}$$

where (46) and (47) ensure that $\left(P_i^*+\epsilon_1, P_j^*-\epsilon_2\right)$ satisfies the power constraints in (14) and (15), respectively. Denote the function $G_1(x) \triangleq \exp\left(\frac{\alpha-1}{x}\right)$, where $x>0$, so the objective function in problem (F.P7) can be expressed as $\sum_{k=1}^{K} G_1(P_k)$. Next, we will verify that $G_1(P_i^*+\epsilon_1)+G_1(P_i^*-\epsilon_2) > G_1(P_i^*)+G_1(P_j^*)$.

Based on *Lagrange mean value theorem*, there exists some $\varepsilon_1 \in (P_i^*, P_i^*+\epsilon_1)$ and $\varepsilon_2 \in (P_j^*-\epsilon_2, P_j^*)$ such that

$$G_1(P_i^*+\epsilon_1)-G_1(P_i^*) = \epsilon_1 G_1'(\varepsilon_1), \tag{48}$$

$$G_1(P_j^*)-G_1(P_i^*-\epsilon_2) = \epsilon_2 G_1'(\varepsilon_2). \tag{49}$$

Since the derivative of $G_1(x)$ is

$$G_1'(x) = -\frac{\alpha-1}{x^2}\exp\left(\frac{\alpha-1}{x}\right) = -(\alpha-1)\exp\left(-2\ln(x)+\frac{\alpha-1}{x}\right),$$

we have

$$\frac{\epsilon_1 G_1'(\varepsilon_1)}{\alpha-1} = -\exp\left(\ln(\epsilon_1)-2\ln(\varepsilon_1)+\frac{\alpha-1}{\varepsilon_1}\right), \tag{50}$$

$$\frac{\epsilon_2 G_1'(\varepsilon_2)}{\alpha-1} = -\exp\left(\ln(\epsilon_2)-2\ln(\varepsilon_2)+\frac{\alpha-1}{\varepsilon_2}\right). \tag{51}$$

Furthermore, from (45), one can easily obtain that $\varepsilon_1 < \varepsilon_2$. Thus, from (50) and (51), we have

$$\epsilon_1 G_1'(\varepsilon_1) < \epsilon_2 G_2'(\varepsilon_2) < 0 \text{ as } \alpha \to \infty. \tag{52}$$

Now, combing (48), (49) with (52), $G_1(P_i^*+\epsilon_1)+G_1(P_i^*-\epsilon_2) > G_1(P_i^*)+G_1(P_j^*)$ holds when $\alpha \to \infty$. In summary, power pair $\left(P_i^*+\epsilon_1, P_j^*-\epsilon_2\right)$ yields a smaller value of the objective function for problem (F.P7), which contradicts with the optimality of $\left(P_i^*, P_j^*\right)$ for problem (F.P7). Therefore, when $\alpha \to \infty$, the inequality $P_i^* < P_j^*$ does not hold, i.e., $P_i^* \geq P_j^*$. Based on Proposition 2, $P_i^* = P_j^*$ can be obtained.

Appendix D
Proof of Lemma 3

The Lagrangian function of problem (R.P2) is first expressed as

$$\mathcal{L}(\{b_i\},\{\lambda_i\}) \triangleq \sum_{i=1}^{K} u_{\alpha}(R_i(b_i,b_{i+1})) - \sum_{i=1}^{K}\lambda_i(b_{i+1}-b_i) \tag{53}$$

where we define $\{b_i\} \triangleq \{b_2,\cdots,b_K\}$ and $\{\lambda_i\} \triangleq \{\lambda_1,\cdots,\lambda_K\}$, $\lambda_i \geq 0$, are Lagrange multipliers. Based on the definition of $\mu_{\alpha}(x)$ in (3), the KKT conditions are given by

$$\frac{\partial \mathcal{L}}{\partial b_{k+1}} = -\frac{(R_k(b_k,b_{k+1}))^{-\alpha}}{b_{k+1}+\frac{1}{H_k}}+\frac{(R_{k+1}(b_{k+1},b_{k+2}))^{-\alpha}}{b_{k+1}+\frac{1}{H_{k+1}}} - \lambda_k+\lambda_{k+1} = 0,\ \forall k \in [1:K-1]. \tag{54}$$

The complementary slackness conditions can be written as

$$\lambda_{k+1}(b_{k+1}-b_{k+2}) = 0, \tag{55}$$

$$\lambda_k(b_{k+1}-b_k) = 0. \tag{56}$$

Note that $R_{k+1}\left(b_{k+1},b_{k+2}\right) = 0$ if $b_{k+1} = b_{k+2}$, and $R_k\left(b_k,b_{k+1}\right) = 0$ if $b_k = b_{k+1}$. However, from (54), $R_{k+1}\left(b_{k+1},b_{k+2}\right), R_k\left(b_k,b_{k+1}\right) > 0$ needs to be satisfied, so we have $b_{k+2} < b_{k+1} < b_k$ at the optimal solution, and hence $\lambda_k = \lambda_{k+1} = 0$. Thus, from (54),

$$-\frac{(R_k(b_k,b_{k+1}))^{-\alpha}}{b_{k+1}+\frac{1}{H_k}}+\frac{(R_{k+1}(b_{k+1},b_{k+2}))^{-\alpha}}{b_{k+1}+\frac{1}{H_{k+1}}} = 0,\quad \forall k \in [1:K-1]. \tag{57}$$

The above equation can be equivalently transformed to $f_{1,k}(b_k,b_{k+1},b_{k+2}) = 0$ as defined in (31), which completes the proof of this lemma.

Appendix E
Proof of Theorem 1

Denote $(\hat{b}_2,\cdots,\hat{b}_K)$ as a solution of the KKT functions in (31). Now, we verify that $(\hat{b}_2,\cdots,\hat{b}_K)$ is the unique solution of these functions. To prove this theorem, *reduction to absurdity* is adopted. In particular, we assume that, beyond $(\hat{b}_2, \cdots, \hat{b}_K)$, there also exists another solution $(\hat{\hat{b}}_2, \cdots, \hat{\hat{b}}_K)$ satisfying the KKT conditions in (31). Assume without loss of generality that $\hat{\hat{b}}_K > \hat{b}_K$. Let $k = K-1$ in (31), then we have

$$\ln\left(1+\hat{b}_{K-1}H_{K-1}\right) = \ln\left(1+\hat{b}_K H_{K-1}\right) + f_{2,K-1}\left(\hat{b}_K\right)\ln\left(1+\hat{b}_K H_K\right), \tag{58}$$

$$\ln\left(1+\hat{\hat{b}}_{K-1}H_{K-1}\right) = \ln\left(1+\hat{\hat{b}}_K H_{K-1}\right) + f_{2,K-1}\left(\hat{\hat{b}}_K\right)\ln\left(1+\hat{\hat{b}}_K H_K\right), \tag{59}$$

where function $f_{2,K-1}(x)$ is defined as

$$f_{2,k}(x) \triangleq \left(\frac{x+\frac{1}{H_{k+1}}}{x+\frac{1}{H_k}}\right)^{1/\alpha} = \left(1-\frac{\frac{1}{H_k}-\frac{1}{H_{k+1}}}{x+\frac{1}{H_k}}\right)^{1/\alpha}, \quad k\in[1:K-1]. \tag{60}$$

Since $f_{2,K-1}(x)$ increases with x when $x>0$, we can obtain

$$\hat{\hat{b}}_{K-1} > \hat{b}_{K-1}, \quad \text{and } \ln\left(\frac{1+\hat{\hat{b}}_{K-1}H_{K-1}}{1+\hat{\hat{b}}_K H_{K-1}}\right) > \ln\left(\frac{1+\hat{b}_{K-1}H_{K-1}}{1+\hat{b}_K H_{K-1}}\right). \tag{61}$$

Now, let $k=K-2$ in (31), we have

$$\ln\left(1+\hat{b}_{K-2}H_{K-2}\right) = \ln\left(1+\hat{b}_{K-1} H_{K-2}\right) + f_{2,K-2}\left(\hat{b}_{K-1}\right)\ln\left(\frac{1+\hat{b}_{K-1}H_{K-1}}{1+\hat{b}_K H_{K-1}}\right), \tag{62}$$

$$\ln\left(1+\hat{\hat{b}}_{K-2}H_{K-2}\right) = \ln\left(1+\hat{\hat{b}}_{K-1} H_{K-2}\right) + f_{2,K-2}\left(\hat{\hat{b}}_{K-1}\right)\ln\left(\frac{1+\hat{\hat{b}}_{K-1}H_{K-1}}{1+\hat{\hat{b}}_K H_{K-1}}\right). \tag{63}$$

Based on (61), (62) and (63), we have

$$\hat{\hat{b}}_{K-2} > \hat{b}_{K-2}, \quad \text{and } \ln\left(\frac{1+\hat{\hat{b}}_{K-2}H_{K-2}}{1+\hat{\hat{b}}_{K-1} H_{K-2}}\right) > \ln\left(\frac{1+\hat{b}_{K-2}H_{K-2}}{1+\hat{b}_{K-1} H_{K-2}}\right). \tag{64}$$

By analogy, $\hat{\hat{b}}_k > \hat{b}_k$ can be verified from $k=K-3$ to $k=1$, i.e., $\forall k\in[1:K-3]$. However, $\hat{\hat{b}}_{K-1} = \hat{b}_{K-1} = P$ holds for problem (R.P2), which contradicts with the result that $\hat{\hat{b}}_k > \hat{b}_k$, $\forall k\in[1:K-3]$. Therefore, only a unique solution $(\hat{b}_2,\cdots,\hat{b}_K)$ of problem (R.P2) exists to satisfy the KKT conditions in (31).

Appendix F
Proof of Theorem 2

This theorem is proven based on the *inductive method*. Specifically, for a given $t_0 \geq 1$, we assume that $b_k^{(t_0)} > b_k^{(t_0-1)}$, $\forall k\in[2:K]$, and then we prove that $b_k^{(t_0+1)} > b_k^{(t_0)}$, $\forall k\in[2:K]$. First, recall that function $f_{2,K-1}(x)$ is defined (60), which increases with x. Next, three different cases are considered.

A. Case $k=1$

In the t-th iteration, from Algorithm II and Lemma 4, we have

$$\ln\left(\tilde{c}_1^{(t)}\right) = f_{2,1}\left(b_2^{(t)}\right)\ln\left(\tilde{\tilde{c}}_2^{(t)}\right), \tag{65}$$

when $k=1$, where we define

$$\tilde{c}_k^{(t)} \triangleq \frac{1+H_k b_k^{(t)}}{1+H_k b_{k+1}^{(t)}}, \tilde{\tilde{c}}_{k+1}^{(t)} \triangleq \frac{1+H_{k+1} b_{k+1}^{(t)}}{1+H_{k+1} b_{k+2}^{(t-1)}}, \quad \forall k\in[1:K-1]. \tag{66}$$

Next, we consider two cases: $\tilde{\tilde{c}}_2^{(t_0+1)} \geq \tilde{\tilde{c}}_2^{(t_0)}$ and $\tilde{\tilde{c}}_2^{(t_0+1)} < \tilde{\tilde{c}}_2^{(t_0)}$.

If $\tilde{\tilde{c}}_2^{(t_0+1)} \geq \tilde{\tilde{c}}_2^{(t_0)}$, since we have assumed that $b_3^{(t_0)} > b_3^{(t_0-1)}$, obviously $b_2^{(t_0+1)} > b_2^{(t_0)}$ holds based on (66).

If $\tilde{\tilde{c}}_2^{(t_0+1)} < \tilde{\tilde{c}}_2^{(t_0)}$, we adopt *reduction to absurdity* to prove that $b_2^{(t_0+1)} > b_2^{(t_0)}$. Specifically, we assume that $b_2^{(t_0+1)} \leq b_2^{(t_0)}$, so we have $f_{2,1}\left(b_2^{(t_0+1)}\right) \leq f_{2,1}\left(b_2^{(t_0)}\right)$. Thus, $f_{2,1}\left(b_2^{(t_0+1)}\right)\ln(\tilde{\tilde{c}}_2^{(t_0+1)}) < f_{2,1}\left(b_2^{(t_0)}\right)\ln\left(\tilde{\tilde{c}}_2^{(t_0)}\right)$ can be obtained. From (65), $\ln\left(\tilde{c}_1^{(t_0+1)}\right) < \ln\left(\tilde{c}_1^{(t_0)}\right)$ holds. However, from (66), $\ln\left(\tilde{c}_1^{(t_0+1)}\right) \geq \ln\left(\tilde{c}_1^{(t_0)}\right)$ under the assumption that $b_2^{(t_0+1)} \leq b_2^{(t_0)}$, since $b_1^{(t_0+1)} = b_1^{(t_0)} = P$. This implies that the assumption $b_2^{(t_0+1)} \leq b_2^{(t_0)}$ does not hold, and thus $b_2^{(t_0+1)} > b_2^{(t_0)}$.

B. Case $k\in[2:K-2]$

Similarly, in the t-th iteration, from Algorithm II and Lemma 4, we have

$$\ln\left(\tilde{c}_2^{(t)}\right) = f_{2,2}\left(b_3^{(t)}\right)\ln\left(\tilde{\tilde{c}}_3^{(t)}\right), \tag{67}$$

when $k=2$. As in the previous case, $b_2^{(t_0+1)} > b_2^{(t_0)}$ if $\tilde{\tilde{c}}_3^{(t_0+1)} \geq \tilde{\tilde{c}}_3^{(t_0)}$.

Now, *reduction to absurdity* is also adopted if $\tilde{\tilde{c}}_3^{(t_0+1)} < \tilde{\tilde{c}}_3^{(t_0)}$. Specifically, similar to the previous case $k-1$, $\ln\left(\tilde{c}_2^{(t_0+1)}\right) < \ln\left(\tilde{c}_2^{(t_0)}\right)$ can be obtained, if we assume that $b_2^{(t_0+1)} \leq b_2^{(t_0)}$ in (67). However, from (66), $\ln\left(\tilde{c}_2^{(t_0+1)}\right) \geq \ln\left(\tilde{c}_2^{(t_0)}\right)$ under the assumption that $b_3^{(t_0+1)} \leq b_3^{(t_0)}$, since $b_2^{(t_0+1)} > b_2^{(t_0)}$ as verified in the previous case. This implies that the assumption $b_3^{(t_0+1)} \leq b_3^{(t_0)}$ does not hold, and thus $b_3^{(t_0+1)} > b_3^{(t_0)}$.

Similarly, $b_k^{(t_0+1)} > b_k^{(t_0)}$ can be proven iteratively, for $k\in[3:K-2]$.

C. Case $k=K-1$

From Algorithm II and Lemma 4, we have

$$\ln\left(\tilde{c}_{K-1}^{(t)}\right) = f_{2,K-1}\left(b_K^{(t)}\right)\ln\left(\tilde{c}_K^{(t)}\right), \tag{68}$$

when $k = K - 1$. Note that $b_K^{(t_0+1)} > b_K^{(t_0)}$ can be proven using almost the same steps to the previous two cases. There is only a slight difference that is $b_{K+1}^{t_0} = b_{K+1}^{(t_0-1)} = 0$. In order to show that $b_K^{(t_0+1)} > b_K^{(t_0)}$ if $\tilde{\tilde{c}}_2^{(t_0+1)} \geq \tilde{\tilde{c}}_2^{(t_0)}$, we only need to verify that $\tilde{\tilde{c}}_2^{(t_0+1)} \neq \tilde{\tilde{c}}_2^{(t_0)}$. Specifically, from (66), $b_K^{(t_0+1)} = b_K^{(t_0)}$ if $\tilde{\tilde{c}}_2^{(t_0+1)} = \tilde{\tilde{c}}_2^{(t_0)}$, so $\tilde{c}_2^{(t_0+1)} = \tilde{c}_2^{(t_0)}$ from (68) and $b_{K-1}^{(t_0)} = b_{K-1}^{(t_0-1)}$ can be obtained from (66). However, $b_{K-1}^{(t_0+1)} > b_{K-1}^{(t_0)}$ as verified in the previous case, which means that $\tilde{\tilde{c}}_2^{(t_0+1)} = \tilde{\tilde{c}}_2^{(t_0)}$ does not hold, i.e., $\tilde{\tilde{c}}_2^{(t_0+1)} \neq \tilde{\tilde{c}}_2^{(t_0)}$.

References

[1] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System-level performance evaluation of downlink non-orthogonal multiple access (NOMA)," in *Proc. IEEE International Symposium Personal, Indoor and Mobile Radio Communications (PIMRC)*, (London, U.K.), pp. 611–615, Sep. 2013.

[2] Q. Li, H. Niu, A. Papathanassiou, and G. Wu, "5G network capacity: Key elements and technologies," *IEEE Vehicular Technology Magzine*, vol. 9, pp. 71–78, Mar. 2014.

[3] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, C. L. I, and H. V. Poor, "Application of Non-Orthogonal Multiple Access in LTE and 5G Networks," *IEEE Communications Magazine*, vol. 55, pp. 185–191, Feb. 2017.

[4] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-Lin, and Z. Wang, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends," *IEEE Communications Magazine*, vol. 53, pp. 74–81, Sep. 2015.

[5] B. Wang, L. Dai, Y. Zhang, T. Mir, and J. Li, "Dynamic compressive sensing-based multi-user detection for uplink grant-free NOMA," *IEEE Communications Letters*, vol. 20, pp. 2320–2323, Nov. 2016.

[6] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Transactions on Vehicular Technology*, vol. 65, pp. 6010–6023, Aug. 2016.

[7] Z. Ding, F. Adachi, and H. Poor, "The application of MIMO to non-orthogonal multiple access," *IEEE Transactions on Wireless Communications*, vol. 15, pp. 537–552, Jan. 2016.

[8] Z. Ding, R. Schober, and H. V. Poor, "A general MIMO framework for NOMA downlink and uplink transmission based on signal alignment," *IEEE Transactions on Wireless Communications*, vol. 15, pp. 4438–4454, June 2016.

[9] Z. Ding and H. V. Poor, "Design of massive-mimo-noma with limited feedback," *IEEE Signal Processing Letters*, vol. 23, pp. 629–633, May 2016.

[10] F. Alavi, K. Cumanan, Z. Ding and A.G. Burr, "Robust beamforming techniques for non-orthogonal multiple access systems with bounded channel uncertainties", IEEE Communications Letters, accepted for publication, 2017.

[11] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Processing Letters*, vol. 21, pp. 1501–1505, Sep. 2014.

[12] Y. Liu, Z. Ding, M. Elkashlan, and H. V. Poor, "Cooperative non-orthogonal multiple access with simultaneous wireless information and power transfer," *IEEE Journal on Selected Areas in Communications*, vol. 34, pp. 938–953, Mar. 2016.

[13] Z. Ding, M. Peng, and H. Poor, "Cooperative non-orthogonal multiple access in 5G systems," *IEEE Communications Letters*, vol. 19, no. 8, pp. 1462–1465, 2015.

[14] Z. Yang, Z. Ding, P. Fan, and G. Karagiannidis, "On the performance of 5G non-orthogonal multiple access systems with partial channel information," *IEEE Transactions on Communications*, vol. 64, pp. 654–667, Feb. 2016.

[15] S. Timotheou and I. Krikidis, "Fairness for non-orthogonal multiple access in 5G systems," *IEEE Signal Processing Letters*, vol. 22, pp. 1647–1651, Oct. 2015.

[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 Transactions on Wireless Communications*, vol. 15, pp. 6716–6730, Oct. 2016.

[17] Y. Liu, M. Elkashlan, Z. Ding, and G. K. Karagiannidis, "Fairness of user clustering in MIMO non-orthogonal multiple access systems," *IEEE Communications Letters*, vol. 20, pp. 1465 – 1468, Jul. 2016.

[18] S. Shi, L. Yang, and H. Zhu, "Outage balancing in downlink non-orthogonal multiple access with statistical channel state information," *IEEE Transactions on Wireless Communications*, vol. 15, pp. 4718–4731, Jul. 2016.

[19] P. Xu, Z. Ding, X. Dai, and H. V. Poor, "A new evaluation criterion for non-orthogonal multiple access in 5G software defined networks," *IEEE Access*, vol. 3, pp. 1633–1639, Oct. 2015.

[20] T. Cover, "Broadcast channels," *IEEE Transactions on Information Theory*, vol. 18, pp. 2–14, Jan. 1972.

[21] P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise (corresp.)," *IEEE Transactions on Information Theory*, vol. 20, pp. 279–280, Mar. 1974.

[22] L. Li and A. J. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels - Part I: Ergodic capacity," *IEEE Transactions on Information Theory*, vol. 47, pp. 1083–1102, Mar. 2001.

[23] L. Li and A. J. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels - Part II: Outage capacity," *IEEE Transactions on Information Theory,*, vol. 47, pp. 1103–1127, Mar. 2001.

[24] W. Zhang, S. P. Kotagiri, and J. N. Laneman, "On downlink transmission without transmit channel state information and with outage constraints," *IEEE Transactions on Information Theory*, vol. 55, pp. 4240–4248, Sept. 2009.

[25] K. Cumanan, L. Musavian, S. Lambotharan, and A. B. Gershman, "SINR balancing technique for downlink beamforming in cognitive radio networks," *IEEE Signal Process. Lett.*, vol. 17, no. 2, pp. 133–136, Feb. 2010.

[26] Y. Rahulamathavan, K. Cumanan, and S. Lambotharan, "A mixed SINR-balancing and SINR-target-constraints-based beamformer design technique for spectrum-sharing networks," *IEEE Trans. Vehicular Technol.*, vol. 60, no. 9, pp. 4403–4414, Nov. 2011.

[27] K. Cumanan, J. Tang, and S. Lambotharan, "Rate balancing based linear transceiver design for multiuser MIMO system with multiple linear transmit covariance constraints," in *Proc. IEEE ICC 2011, Kyoto, Japan*, pp. 1–5, Jun. 2011.

[28] K. Cumanan, R. Krishna, Z. Xiong, and S. Lambotharan, "SINR balancing technique and its comparison to semidefnite programming based QoS provision for cognitive radios," in *Proc. IEEE VTC, Barcelona, Spain*, pp. 1–5, Apr. 2009.

[29] F. Liu, P. Manonen, and M. Petrova, "Proportional fairness-based power allocation and user set selection for downlink NOMA systems," in *Proc. IEEE International Conference on Communications (ICC)*, May 2016.

[30] H. Shi, R. V. Prasad, E. Onur, and I. Niemegeers, "Fairness in wireless networks: Issues, measures and challenges," *IEEE Communications Surveys and Tutorials*, vol. 16, pp. 5–24, First Quarter 2014.

[31] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Transactions on Networking*, vol. 8, pp. 556–567, Oct. 2000.

[32] T. Lan, D. Kao, M. Chiang, and A. Sabharwal, "An axiomatic theory of fairness in network resource allocation," in *Proc. IEEE INFOCOM*, Mar. 2010.

[33] M. Uchida and J. Kurose, "An information-theoretic characterization of weighted alpha-proportional fairness," in *Proc. IEEE INFOCOM*, pp. 1053–1061, Apr. 2009.

[34] Z. Song, Q. Ni, K. Navaie, S. Hou, and S. Wu, "Energy- and spectral-efficiency tradeoff with alpha-fairness in downlink OFDMA systems," *IEEE Communications Letters*, vol. 19, pp. 1265–1268, Jul. 2015.

[35] F. Kelly, A. K. Maulloo, and D. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, pp. 237–252, Mar. 1998.

[36] T. Nguyen and Y. Han, "A proportional fairness algorithm with QoS provision in downlink OFDMA systems," *IEEE Communications Letters*, vol. 10, pp. 760–762, Dec. 2006.

[37] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared systems", Digital Equipment Corporation, Technical Report DEC-TR-301, Tech. Rep., 1984.

[38] X. Gao, L. Dai, S. Han, I. Chih-Lin, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for MmWave MIMO systems with large antenna arrays," *IEEE Journal on Selected Areas in Communications*, vol. 34, pp. 998–1009, Apr. 2016.

[39] 3GPP TR 36.859: "Study on Downlink Multiuser Superposition Transmission (MUST) for LTE", Nov. 2015.

[40] Y. Ye, Interior Point Algorithms. Theory and Analysis. John Wiley & Sons, 1997.

[41] A. B. Sediq, R. H. Gohary, R. Schoenen, and H. Yanikomeroglu, "Optimal tradeoff between sum-rate efficiency and Jain's fairness index

in resource allocation," *IEEE Transactions on Wireless Communications*, vol. 12, pp. 3496–3509, Jul. 2013.

[42] H. T. Cheng and W. Zhuang, "An optimization framework for balancing throughput and fairness in wireless networks with QoS support," *IEEE Transactions on Wireless Communications*, vol. 7, pp. 584–593, Feb. 2008.

[43] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. New York, NY, USA: Cambridge University press, 2005.

Peng Xu received the B.Eng. and the Ph.D. degrees in electronic and information engineering from the University of Science and Technology of China, Anhui, China, in 2009 and 2014, respectively. Since July 2014, he was working as a postdoctoral researchers with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, China. From July 2016, he has been working at the School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications (CQUPT), Chongqing, China. His current research interests include cooperative communications, information theory, information-theoretic secrecy, and 5G networks. Dr. Peng Xu received IEEE Wireless Communications Letters Exemplary Reviewer 2015.

Kanapathippillai Cumanan (M'10) received the BSc degree with first class honors in electrical and electronic engineering from the University of Peradeniya, Sri Lanka in 2006 and the PhD degree in signal processing for wireless communications from Loughborough University, Loughborough, UK, in 2009.

He is currently a lecturer at the Department of Electronic Engineering, University of York, UK. From March 2012 to November 2014, he was working as a research associate at School of Electrical and Electronic Engineering, Newcastle University, UK. Prior to this, he was with the School of Electronic, Electrical and System Engineering, Loughborough University, UK. In 2011, he was an academic visitor at Department of Electrical and Computer Engineering, National University of Singapore, Singapore. From January 2006 to August 2006, he was a teaching assistant with Department of Electrical and Electronic Engineering, University of Peradeniya, Sri Lanka. His research interests include physical layer security, cognitive radio networks, relay networks, convex optimization techniques and resource allocation techniques.

Dr. Cumanan was the recipient of an overseas research student award scheme (ORSAS) from Cardiff University, Wales, UK, where he was a research student between September 2006 and July 2007.