Research Article

Stochastic Optimal Control for Buffer-Aided Cooperative Relaying Systems Using Nonorthogonal Multiple Access

Zhishan Deng,[1] Qianyun Gong,[1] Quanzhong Li [iD],[2] and Jiayin Qin[3]

[1]School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou 510006, Guangdong, China
[2]School of Data and Computer Science, and also with the Guangdong Key Laboratory of Information Security,
 Sun Yat-sen University, Guangzhou 510006, Guangdong, China
[3]School of Electronics and Information Technology, and also with the Xinhua College, Sun Yat-sen University,
 Guangzhou 510006, Guangdong, China

Correspondence should be addressed to Quanzhong Li; liquanzh@mail.sysu.edu.cn

In this paper, a multiple-cluster downlink multiple-input single-output (MISO) nonorthogonal multiple access (NOMA) system is considered. In each cluster, there are one central user and one cell-edge user. The central user has a data buffer with finite storage units, which will decode the cell-edge user's message and store it at the data buffer. To enhance the performance of the cell-edge user, the central user operates as a relay and helps forward the message to the cell-edge user. Our objective is to maximize the long-term average sum rates for the cell-edge users by designing the beamforming vectors and online power control, under the constraints of the data buffer causality, required information rates for central users, and transmit power at the base station and central users. Based on the current buffer state and the channel state information, we propose a low-complexity online Lyapunov optimization algorithm combined with a constrained concave-convex procedure (CCCP) to solve the causal and nonconvex problem. Furthermore, we verify the asymptotic optimality of the proposed online Lyapunov optimization algorithm. Simulation results demonstrate that our proposed scheme performs better than the greedy algorithm and the orthogonal multiple access (OMA) scheme.

1. Introduction

Recently, nonorthogonal multiple access (NOMA) scheme is considered as a breakthrough key technology for the fifth generation networks and has attracted great attention [1, 2]. Combined with successive interference cancellation (SIC) and beamforming technique, the NOMA scheme allows a single base station to serve more users with higher spectral efficiency [3, 4]. NOMA also has an extension to a cooperative relaying system, where users with poor capabilities can be improved by requesting the ones with strong capabilities acting as relays, which can decode and forward (DF) the messages to the poor ones [5, 6]. By using cooperative relaying and maximum ratio combining, the spatial diversity of NOMA systems can be enhanced [7, 8]. Furthermore, Zhang et al. [9] had shown

that the data buffer, which can store messages sent by the base station, applied in the cooperative relays can improve the performances of NOMA cooperative relaying systems significantly.

In the practical and causal NOMA cooperative relaying system, the current information is reachable only when it arrives, which is involved with the online resource allocation problem. In [10–12], the authors manage to solve the resource allocation problem by merely utilizing the current information and applying dynamic programming algorithms [13]. Because of the high computational complexity of dynamic programming algorithms, some studies propose the Lyapunov optimization approach [14–16]. The authors of [14] propose a low-complexity optimization algorithm using Lyapunov functions to achieve close-to-optimal utility performance in energy-harvesting networks. Mao et al. [15]

aimed to minimize the long-term average network service cost on hybrid energy supply networks by optimizing the base station selection and power allocation. Choi and Kim [16] considered a wireless powered system, and its objective was to minimize the expected energy transmission power while stabilizing data queue for all nodes. To the best of our knowledge, the stochastic optimal control for multiple-input single-output (MISO) NOMA cooperative relaying systems has not been studied.

A cellular downlink MISO NOMA system is being considered in our study, which consists of multiple clusters. In each cluster, there are one central user and one cell-edge user. The central user has a data buffer with finite storage units and operates as a relay to help forward the message to the cell-edge user. For cell-edge users with high quality of service (QoS) requirement, by designing optimal beamforming vectors and power allocation, our objective is to maximize the long-term average achievable information bits for the cell-edge users under the constraints of required number of achievable information bits for the central users and transmit power constraints. The optimization problem is causal and nonconvex and thus hard to solve. Based on the current buffer state and the channel state information, we propose a low-complexity online Lyapunov optimization algorithm combined with a constrained concave-convex procedure (CCCP). The asymptotic optimality of the proposed online Lyapunov optimization algorithm is also verified.

The rest of this paper is organized as follows: Section 2 presents the system model in detail. Section 3 presents the proposed low-complexity online Lyapunov optimization algorithm. Section 4 verifies the asymptotic optimality of the proposed online Lyapunov optimization algorithm. Section 5 provides the simulation results. Finally, Section 6 concludes the paper.

Notations. For a vector, $\mathbf{a}$, $\mathbf{a}^{\dagger}$, and $||\mathbf{a}||$ denotes its conjugate, transpose, and Frobenius norm, respectively. $\mathbb{E}[x]$ and $\text{Re}\{x\}$ denote the expectation and the real part of x, respectively. $\mathscr{CN}(a, b)$ denotes the Gaussian variable with mean a and variance b.

2. System Model and Problem Formulation

We consider a buffer-aided cooperative NOMA downlink system consisting of a base station and $2K$ users as shown in Figure 1, where the base station is equipped with N antennas and each user is equipped with a single antenna. The clustering algorithm is a significant factor in the multiuser NOMA system since it will influence the system performance [17–19]. In [17], a clustering algorithm was proposed based on selecting two highly correlated users with a large channel-gain difference. In [18], the authors selected users for satellite by a channel quality-based scheme and proposed a user pairing method by maximizing the minimum channel correlation between users in the same group. The authors in [19] investigated three user clustering schemes from a fairness perspective by
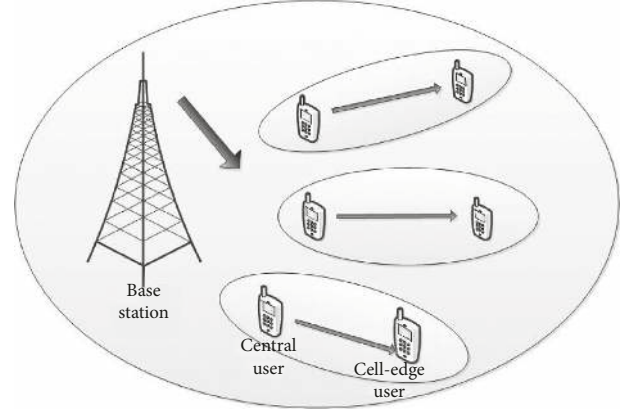

FIGURE 1: The model for a buffer-aided cooperative NOMA downlink transmission system.

maximizing the throughput of the worst user. However, the aforementioned studies focused on the one-time-slot problem without considering the long-term performance of the system. By employing the clustering scheme proposed in [19], we group the $2K$ users into K clusters in advance and formulate a multitime-slot online stochastic optimization problem in the following. In the kth cluster, $k \in \mathscr{K} = \{1, 2, \ldots, K\}$, two users are included, i.e., a central user and a cell-edge user.

The base station transmits signals to central users directly while transmits signals to cell-edge users with the cooperation of central users. We assume that the transmission time duration is partitioned into M slots with equal length of T. Each time slot is further partitioned into $(K + 1)$ phases. In each time slot, the base station transmits signals to central and cell-edge users using the NOMA protocol during the first phase, and all the central users decode and forward the signals to cell-edge users using the timed-division multiple access scheme during the remaining K phases. Thus, during the first phase of the mth time slot, $m \in \mathscr{M} = \{1, 2, \ldots, M\}$, the transmitted signal from the base station to the central and cell-edge users is expressed as

$$\mathbf{x}_m = \sum_{k=1}^{K} \left(\mathbf{v}_{k,m} s_{k,m,1} + \mathbf{w}_{k,m} s_{k,m,2} \right), \quad (1)$$

where $s_{k,m,1}$ and $s_{k,m,2}$ denote the signals intended to the central and cell-edge users of the kth cluster with $\mathbb{E}[|s_{k,m,1}|^2] = 1$ and $\mathbb{E}[|s_{k,m,2}|^2] = 1$, respectively, and $\mathbf{v}_{k,m} \in \mathbb{C}^{N \times 1}$ and $\mathbf{w}_{k,m} \in \mathbb{C}^{N \times 1}$ are the corresponding beamforming vectors. Accordingly, the received signals at the central and cell-edge users of the kth cluster are

$$\begin{aligned} y_{k,m,1} &= \mathbf{h}_{k,m}^{\dagger} \mathbf{x}_m + n_{k,m,1}, \\ y_{k,m,2} &= \mathbf{g}_{k,m}^{\dagger} \mathbf{x}_m + n_{k,m,2}, \end{aligned} \quad (2)$$

where $\mathbf{h}_{k,m} \in \mathbb{C}^{N \times 1}$ and $\mathbf{g}_{k,m} \in \mathbb{C}^{N \times 1}$ denote the channel response from the base station to the central and cell-edge users of the kth cluster, respectively, and $n_{k,m,1} \sim \mathscr{CN}(0, \sigma^2)$ and $n_{k,m,2} \sim \mathscr{CN}(0, \sigma^2)$ denote the additive Gaussian noises at the central and cell-edge users of the kth cluster.

In the kth cluster, the central user first decodes $s_{k,m,2}$ by treating $s_{k,m,1}$ as interference and then removes $s_{k,m,2}$ using successive interference cancellation to decode $s_{k,m,1}$. The signal-to-interference-and-noise ratios (SINRs) for decoding $s_{k,m,2}$ and $s_{k,m,1}$ at the kth central user, $k \in \mathcal{K}$, are

$$\gamma_{k,m,1,2} = \frac{\mathbf{h}_{k,m}^{\dagger} \mathbf{w}_{k,m} \mathbf{w}_{k,m}^{\dagger} \mathbf{h}_{k,m}}{\xi_{k,m} + \mathbf{h}_{k,m}^{\dagger} \mathbf{v}_{k,m} \mathbf{v}_{k,m}^{\dagger} \mathbf{h}_{k,m} + \sigma^2}, \tag{3}$$

$$\gamma_{k,m,1,1} = \frac{\mathbf{h}_{k,m}^{\dagger} \mathbf{v}_{k,m} \mathbf{v}_{k,m}^{\dagger} \mathbf{h}_{k,m}}{\xi_{k,m} + \sigma^2}, \tag{4}$$

where $\xi_{k,m}$ is the intercluster interference and

$$\xi_{k,m} = \sum_{i \neq k, i \in \mathcal{K}} \mathbf{h}_{k,m}^{\dagger} \left(\mathbf{w}_{i,m} \mathbf{w}_{i,m}^{\dagger} + \mathbf{v}_{i,m} \mathbf{v}_{i,m}^{\dagger} \right) \mathbf{h}_{k,m}. \tag{5}$$

Accordingly, at the kth central user, the numbers of achievable information bits intended for the kth central and cell-edge users in the first phase are

$$R_{k,m,1} = \alpha \log_2 \left(1 + \gamma_{k,m,1,1} \right),$$
$$C_{k,m,1} = \alpha \log_2 \left(1 + \gamma_{k,m,1,2} \right), \tag{6}$$

where $\alpha = BT/(K+1)$ and B denotes the bandwidth. At the kth cell-edge user, $k \in \mathcal{K}$, the SINR for decoding $s_{k,m,2}$ is expressed as

$$\gamma_{k,m,2,2} = \frac{\mathbf{g}_{k,m}^{\dagger} \mathbf{w}_{k,m} \mathbf{w}_{k,m}^{\dagger} \mathbf{g}_{k,m}}{\zeta_{k,m} + \mathbf{g}_{k,m}^{\dagger} \mathbf{v}_{k,m} \mathbf{v}_{k,m}^{\dagger} \mathbf{g}_{k,m} + \sigma^2}, \tag{7}$$

where

$$\zeta_{k,m} = \sum_{i \neq k, i \in \mathcal{K}} \mathbf{g}_{k,m}^{\dagger} \left(\mathbf{w}_{i,m} \mathbf{w}_{i,m}^{\dagger} + \mathbf{v}_{i,m} \mathbf{v}_{i,m}^{\dagger} \right) \mathbf{g}_{k,m}. \tag{8}$$

During the remaining K phases of the mth time slot, the kth central user, $k \in \mathcal{K}$, decodes and forwards the signals to the kth cell-edge user in any phase among the remaining K phases. The received signal at the kth cell-edge user is

$$y_{k,m,3} = f_{k,m} \sqrt{P_{k,m}} s_{k,m,2} + n_{k,m,3}, \tag{9}$$

where $f_{k,m}$ denotes the channel response from the kth central user to the kth cell-edge user, $P_{k,m}$ denotes the transmit power, and $n_{k,m,3} \sim \mathcal{CN}(0, \sigma^2)$ denotes the additive Gaussian noise at the kth cell-edge user. Accordingly, the number of achievable information bits transmitted from the kth central user to the kth cell-edge user is

$$C_{k,m,2} = \frac{BT}{K+1} \log_2 \left(1 + \gamma_{k,m,2,3} \right), \tag{10}$$

where the denominator K is included because the orthogonal multiple access (OMA) transmission scheme among K clusters is employed and $\gamma_{k,m,2,3}$ denotes the SINR for decoding $s_{k,m,2}$ as

$$\gamma_{k,m,2,3} = \frac{\left| f_{k,m} \right|^2 P_{k,m}}{\sigma^2}. \tag{11}$$

The central users can store the information bits at the data buffers for later transmission. We assume the data buffers at the central users have finite storage units, denoted as $D^{\max}$. Denote $D_{k,m}$ as the number of storage units and stored information bits in the kth central user, $k \in \mathcal{K}$, at the end of the first phase of the mth time slot, $m \in \mathcal{M}$. Thus, considering the data buffer's states and channel states, the central users will transmit the information bits to the cell-edge users with the following data causal constraint:

$$C_{k,m,2} \leq D_{k,m} \leq D^{\max}. \tag{12}$$

The dynamics of data buffer is

$$D_{k,m+1} = D_{k,m} + C_{k,m+1,1} - C_{k,m,2}. \tag{13}$$

At the kth cell-edge user, maximum ratio combining is employed to combine the signals transmitted from the base station and the kth central user. Thus, in the mth time slot, the number of achievable information bits intended for the kth cell-edge user is

$$R_{k,m,2} = \alpha \log_2 \left(1 + \min \left(\gamma_{k,m,1,2}, \gamma_{k,m,2,2} + \gamma_{k,m,2,3} \right) \right). \tag{14}$$

For the cell-edge users with high QoS requirement, our objective is to maximize the long-term average achievable information bits for the cell-edge users under the constraints of required number of achievable information bits for the central users and transmit power constraints, which is formulated as

$$P_0: \quad \max_{\substack{\{\mathbf{v}_{k,m}, \\ \mathbf{w}_{k,m}, \\ P_{k,m}\}}} \lim_{M \longrightarrow \infty} \frac{1}{M} \sum_{m=1}^{M} \mathbb{E}\left[\widehat{R}_m \right], \tag{15a}$$

$$\text{s.t.} \quad (12), \mathbb{E}\left[R_{k,m,1} \right] \geq r_k, \tag{15b}$$

$$\sum_{k=1}^{K} \left(\left\| \mathbf{v}_{k,m} \right\|^2 + \left\| \mathbf{w}_{k,m} \right\|^2 \right) \leq P_b^{\max}, \tag{15c}$$

$$P_{k,m} \leq P_c^{\max}, \quad \forall k \in \mathcal{K}, m \in \mathcal{M}, \tag{15d}$$

where r_k denotes the required number of achievable information bits for the kth central user in any time slot; $P_b^{\max}$ and $P_c^{\max}$ mean the transmit power constraints at the base station and central users, respectively; and

$$\widehat{R}_m = \sum_{k=1}^{K} R_{k,m,2}. \tag{16}$$

Problem P_0 is a stochastic optimization problem which needs all the channel state and buffer state of all time duration ahead, it is impractical and challenging. Therefore, we will solve this problem with an online algorithm.

3. Online Lyapunov Algorithm

In this section, an online Lyapunov algorithm to solve problem P_0 is proposed. The main idea is to convert the causal constraints into virtual buffers and maintain their stability.

According to the Lyapunov optimization [20], we first introduce a virtual buffer queue $V_{k,m}$ for the actual buffer state $D_{k,m}$ at the kth user as $V_{k,m} = D_{k,m} - \theta_k$, $\quad \forall m \in \mathcal{M}$, where θ_k denotes the perturbation parameter. Specifically, Neely [20] had proved that maintaining the stability of all the virtual buffers $V_{k,m}$ is equivalent to meeting their causality requirements (13). Then, we define the quadratic Lyapunov function with respect to the data buffer at the kth central user as

$$L_{k,m} = \frac{\left(V_{k,m}\right)^2}{2} = \frac{1}{2}\left(D_{k,m} - \theta_k\right)^2, \quad \forall m \in \mathcal{M}. \tag{17}$$

Accordingly, the per-time-slot Lyapunov drift with respect to the data buffer at the kth central user is

$$\Delta_{k,m} = \mathbb{E}\left[L_{k,m+1} - L_{k,m} \mid D_{k,m}\right], \quad \forall m \in \mathcal{M}, \tag{18}$$

which describes the expected change in the Lyapunov function over one time slot. Based on the Lyapunov optimization framework [20], we consider the minimization of the drift-plus-penalty function rather than using the objective of (15a)–(15d) directly, which is a technique to maintain the stability of the queue and optimize the long-term average objective function in the meantime. Therefore, the Lyapunov drift-plus-penalty function is expressed as

$$U_m = \sum_{k=1}^{K} \Delta_{k,m} - \rho\mathbb{E}\left[\widehat{R}_m \mid \mathcal{D}_m\right], \tag{19}$$

where ρ denotes a control parameter and

$$\mathcal{D}_m = \left\{D_{1,m}, \ldots, D_{K,m}\right\}. \tag{20}$$

Because of the dynamics involved in $\Delta_{k,m}$, it is still difficult to minimize the Lyapunov drift-plus-penalty function (19) directly. Instead, we turn to minimize the upper bound of (19). Thus, we present the following lemma to provide an upper bound on the Lyapunov drift-plus-penalty function.

Lemma 1. *For any feasible $\left\{\mathbf{v}_{k,m}, \mathbf{w}_{k,m}, P_{k,m}\right\}$, the drift-plus-penalty function has an upper bound as*

$$U_m \leq \mathbb{E}\left[\sum_{k=1}^{K}\left(D_{k,m} - \theta_k\right)\left(C_{k,m+1,1} - C_{k,m,2}\right) - \rho\widehat{R}_m \mid \mathcal{D}_m\right]$$
$$+ \frac{1}{2}\mathbb{E}\left[\sum_{k=1}^{K}\left(C_{k,m+1,1}^{\max}\right)^2 + \left(C_{k,m,2}^{\max}\right)^2\right], \tag{21}$$

where

$$C_{k,m+1,1}^{\max} = \max_{\left\{\mathbf{v}_{k,m+1}, \mathbf{w}_{k,m+1}\right\}} C_{k,m+1,1}, \tag{22}$$

$$C_{k,m,2}^{\max} = \max_{P_{k,m}} C_{k,m,2}. \tag{23}$$

Proof. From the dynamics of data buffer (13), we have

$$\Delta_{k,m} = \frac{1}{2}\mathbb{E}\left[\left(D_{k,m+1} - \theta_k\right)^2 - \left(D_{k,m} - \theta_k\right)^2\right]$$

$$= \frac{1}{2}\mathbb{E}\left[\left(D_{k,m} + C_{k,m+1,1} - C_{k,m,2} - \theta_k\right)^2 - \left(D_{k,m} - \theta_k\right)^2\right]$$

$$= \frac{1}{2}\mathbb{E}\left[\left(C_{k,m+1,1} - C_{k,m,2}\right)^2\right.$$
$$\left. + 2\left(D_{k,m} - \theta_k\right)\left(C_{k,m+1,1} - C_{k,m,2}\right)\right]$$

$$\leq \frac{1}{2}\mathbb{E}\left[\left(C_{k,m+1,1}\right)^2 + \left(C_{k,m,2}\right)^2\right.$$
$$\left. + 2\left(D_{k,m} - \theta_k\right)\left(C_{k,m+1,1} - C_{k,m,2}\right)\right], \tag{24}$$

where the identity $(a-b)^2 \leq a^2 + b^2$ is used.

From (22) and (23), we have

$$\Delta_{k,m} \leq \frac{1}{2}\mathbb{E}\left[\left(C_{k,m+1,1}^{\max}\right)^2 + \left(C_{k,m,2}^{\max}\right)^2\right.$$
$$\left. + 2\left(D_{k,m} - \theta_k\right) \cdot \left(C_{k,m+1,1} - C_{k,m,2}\right)\right]. \tag{25}$$

Substituting (25) into (19), we obtain (21).

The Lyapunov algorithm only needs the current system state, and thus, we turn it into a per-slot problem and minimize the drift-plus-penalty function's upper bound. Since $C_{k,m+1,1}^{\max}$ and $C_{k,m,2}^{\max}$ are constants, given $\mathcal{D}_m$, then $C_{k,m+1,1}^{\max}$, $C_{k,m,2}^{\max}$, and the expectation can be removed. Thus, the optimization problem P_0 is reduced to

$$\min_{\substack{\mathbf{v}_{k,m}, \mathbf{w}_{k,m}, \\ P_{k,m}}} \sum_{k=1}^{K}\left(D_{k,m} - \theta_k\right)\left(C_{k,m+1,1} - C_{k,m,2}\right) - \rho\widehat{R}_m, \tag{26}$$

s.t. $R_{k,m,1} \geq r_k$, (15c), (15d), $\forall k \in \mathcal{K}$.

In problem (26), the constraint $R_{k,m,1} \geq r_k$ is equivalent to

$$\xi_{k,m} + \sigma^2 \leq \frac{\mathbf{h}_{k,m}^\dagger\mathbf{v}_{k,m}\mathbf{v}_{k,m}^\dagger\mathbf{h}_{k,m}}{2^{r_k/\alpha} - 1}. \tag{27}$$

Using (3), (7), and (11), problem (26) is recast as

$$P_1: \quad \min_{\substack{\mathbf{v}_{k,m}, \mathbf{w}_{k,m}, \\ P_{k,m}, d_{k,m} \geq 0}} \alpha \sum_{k=1}^{K}\left(\left(D_{k,m} - \theta_k\right) \cdot \Theta_k - \rho d_{k,m}\right), \tag{28a}$$

s.t.   (29), (17c), (17d), \tag{28b}

$$\xi_{k,m} + \mathbf{h}_{k,m}^\dagger\mathbf{v}_{k,m}\mathbf{v}_{k,m}^\dagger\mathbf{h}_{k,m} + \sigma^2 \leq \frac{\mathbf{h}_{k,m}^\dagger\mathbf{w}_{k,m}\mathbf{w}_{k,m}^\dagger\mathbf{h}_{k,m}}{d_{k,m}}, \tag{28c}$$

$$\zeta_{k,m} + \mathbf{g}_{k,m}^\dagger \mathbf{v}_{k,m} \mathbf{v}_{k,m}^\dagger \mathbf{g}_{k,m} + \sigma^2 \le \frac{\mathbf{g}_{k,m}^\dagger \mathbf{w}_{k,m} \mathbf{w}_{k,m}^\dagger \mathbf{g}_{k,m}}{d_{k,m} - \left(\left|f_{k,m}\right|^2 P_{k,m}/\sigma^2\right)},$$
$$\forall k \in \mathcal{K}, \tag{28d}$$

where $d_{k,m}$ is an introduced slack variable and

$$\Theta_k = \log_2\left(1 + d_{k,m}\right) - \log_2\left(1 + \frac{\left|f_{k,m}\right|^2 P_{k,m}}{\sigma^2}\right). \tag{29}$$

Problem P_1 is still nonconvex on account of the objective function, constraints (27), (28c), and (28d). To work out the problems (28a)–(28d), we employ an iterative algorithm based on CCCP [21].

Note that when $D_{k,m} - \theta_k \le 0$, $(D_{k,m} - \theta_k) \cdot \Theta_k$ in the objective function is convex and when $D_{k,m} - \theta_k > 0$, $(D_{k,m} - \theta_k) \cdot \Theta_k$ is concave. To proceed, we define it by its first-order Taylor expansion around the point $\widetilde{Y}_{k,m} = (\widetilde{P}_{k,m}, \widetilde{d}_{k,m})$ as [22]

$$\widehat{\Theta}_k = \begin{cases} \log_2\left(1 + d_{k,m}\right) - \left(P_{k,m} - \widetilde{P}_{k,m}\right)\dfrac{\left|f_{k,m}\right|^2}{\left(\left|f_{k,m}\right|^2 + \sigma^2\right)\ln 2} - \log_2\left(1 + \dfrac{\left|f_{k,m}\right|^2 \widetilde{P}_{k,m}}{\sigma^2}\right), & \text{if } D_{k,m} - \theta_k \le 0, \\[3ex] -\log_2\left(1 + \dfrac{\left|f_{k,m}\right|^2 P_{k,m}}{\sigma^2}\right) + \log_2\left(1 + \widetilde{d}_{k,m}\right) + \left(d_{k,m} - \widetilde{d}_{k,m}\right)\dfrac{1}{\left(1 + \widetilde{d}_{k,m}\right)\ln 2}, & \text{if } D_{k,m} - \theta_k > 0. \end{cases} \tag{30}$$

The right-hand side of (27) can be approximated by its first-order Taylor expansion around the point $\widetilde{\mathbf{v}}_{k,m}$ as

$$\psi_{k,m}\left(\mathbf{v}_{k,m}; \widetilde{\mathbf{v}}_{k,m}\right) = \frac{2\mathrm{Re}\left\{\mathbf{h}_{k,m}^\dagger\left(\mathbf{v}_{k,m} - \widetilde{\mathbf{v}}_{k,m}\right)\widetilde{\mathbf{v}}_{k,m}^\dagger \mathbf{h}_{k,m}\right\}}{2^{r_k/\alpha} - 1}. \tag{31}$$

Similarly, the right-hand sides of (28c) and (28d) can also be approximated by their first-order Taylor expansions around the points $\widetilde{\Omega}_{k,m} = (\widetilde{\mathbf{w}}_{k,m}, \widetilde{d}_{k,m}, \widetilde{P}_{k,m})$ as

$$\phi_{k,m}\left(\Omega_{k,m}; \widetilde{\Omega}_{k,m}\right) = \frac{2\mathrm{Re}\left\{\mathbf{h}_{k,m}^\dagger \mathbf{w}_{k,m} \widetilde{\mathbf{w}}_{k,m}^\dagger \mathbf{h}_{k,m}\right\}}{\widetilde{d}_{k,m}} + \frac{d_{k,m} \mathbf{h}_{k,m}^\dagger \widetilde{\mathbf{w}}_{k,m} \widetilde{\mathbf{w}}_{k,m}^\dagger \mathbf{h}_{k,m}}{\widetilde{d}_{k,m}^2},$$

$$\varphi_{k,m}\left(\Omega_{k,m}; \widetilde{\Omega}_{k,m}\right) = \frac{2\mathrm{Re}\left\{\mathbf{g}_{k,m}^\dagger\left(\mathbf{w}_{k,m} - \widetilde{\mathbf{w}}_{k,m}\right)\widetilde{\mathbf{w}}_{k,m}^\dagger \mathbf{g}_{k,m}\right\}}{\widetilde{d}_{k,m} - \left(\left|f_{k,m}\right|^2 \widetilde{P}_{k,m}/\sigma^2\right)}$$
$$+ \frac{\left(\left|f_{k,m}\right|^2\left(P_{k,m} - \widetilde{P}_{k,m}\right)/\sigma^2\right) - \left(d_{k,m} - \widetilde{d}_{k,m}\right)}{\left(\widetilde{d}_{k,m} - \left(\left|f_{k,m}\right|^2 \widetilde{P}_{k,m}/\sigma^2\right)\right)^2}$$
$$\cdot \mathbf{g}_{k,m}^\dagger \widetilde{\mathbf{w}}_{k,m} \widetilde{\mathbf{w}}_{k,m}^\dagger \mathbf{g}_{k,m}, \tag{32}$$

where $\Omega_{k,m} = (\mathbf{w}_{k,m}, d_{k,m}, P_{k,m})$.

Thus, in our iterative algorithm, given $\{\mathbf{v}_{k,m}^{(l)}, \mathbf{w}_{k,m}^{(l)}, P_{k,m}^{(l)}, d_{k,m}^{(l)}\}$ which is optimal in the lth iteration, we solve the following problem in the $(l+1)$th iteration:

$$\min_{\substack{\mathbf{v}_{k,m}, \mathbf{w}_{k,m}, \\ P_{k,m}, d_{k,m} \ge 0}} \quad \alpha \sum_{k=1}^{K}\left(\left(D_{k,m} - \theta_k\right)\cdot\widehat{\Theta}_k - \rho d_{k,m}\right),$$

$$\text{s.t.} \quad \xi_{k,m} + \sigma^2 \le \psi_{k,m}\left(\mathbf{v}_{k,m}; \mathbf{v}_{k,m}^{(l)}\right), \quad (17c), \ (17d),$$

$$\xi_{k,m} + \mathbf{h}_{k,m}^\dagger \mathbf{v}_{k,m} \mathbf{v}_{k,m}^\dagger \mathbf{h}_{k,m} + \sigma^2$$
$$\le \phi_{k,m}\left(\Omega_{k,m}; \Omega_{k,m}^{(l)}\right),$$

$$\zeta_{k,m} + \mathbf{g}_{k,m}^\dagger \mathbf{v}_{k,m} \mathbf{v}_{k,m}^\dagger \mathbf{g}_{k,m}$$
$$+ \sigma^2 \le \varphi_{k,m}\left(\Omega_{k,m}; \Omega_{k,m}^{(l)}\right), \quad \forall k \in \mathcal{K}, \tag{33}$$

where $\Omega_{k,m} = (\mathbf{w}_{k,m}^{(l)}, d_{k,m}^{(l)}, P_{k,m}^{(l)})$. Problem (33) is convex and can be solved effectively using the interior point method [23].

Now, we summarize the proposed online Lyapunov algorithm in Algorithm 1. $\qquad\square$

Remark 1. (*Complexity Analysis for Algorithm* 1): the computational complexity for the online Lyapunov algorithm combined with CCCP is mainly from solving problem (33), which is a semidefinite programming (SDP). From [24], the complexity of solving an SDP is $\mathcal{O}\left(m_{\mathrm{sdp}} n_{\mathrm{sdp}}^{7/2} + m_{\mathrm{sdp}}^2 n_{\mathrm{sdp}}^{5/2} + m_{\mathrm{sdp}}^3 n_{\mathrm{sdp}}^{1/2}\right)$, where m_{sdp} denotes the number of semidefinite cone constraints and n_{sdp} denotes the dimension of the semidefinite cone. For problem (33), $m_{\mathrm{sdp}} = 3K$ and $n_{\mathrm{sdp}} = N$. Thus, the complexity of solving problem (33) is $\mathcal{O}\left(3KN^{7/2} + (3K)^2 N^{5/2} + (3K)^3 N^{1/2}\right)$. Denote the iteration number of CCCP iterations as $\mathscr{L}$, the complexity of the proposed Algorithm 1 is $\mathcal{O}\left(\mathscr{L}\left(3KN^{7/2} + (3K)^2 N^{5/2} + (3K)^3 N^{1/2}\right)\right)$, which indicates the complexity of Algorithm 1 grows as the number of clusters increases.

Remark 2. (*Application Issue of Algorithm* 1): actually, the expected number of users that the system can support depends on the implementation hardware. For example, under our experimental circumstance (the central processing unit is Intel Core i7-4790K with 4.0 GHz and the random access memory is 8 GB), the average time required to solve one CCCP procedure is about 0.7 ms for the scenario that $K = 4$, i.e., 8 users. Thus, for each time slot, $T = 1$ ms is a pessimistic setting in our experiment, which aims to adapt our proposed Lyapunov algorithm to suit other conditions. Furthermore, as the unit for data transmission and scheduling, the subframe lasts for 1 ms, which is called transmission time interval (TTI) in the LTE standard [25, 26]. Therefore, in most cases, TTI requires our proposed algorithm should be executed within 1 ms. In reality, it is feasible to implement our proposed algorithm on real base stations since they have much more powerful processing ability to execute the

```
(1) Initialize: m = 0, D_{k,m} = 0;
(2) While m ≤ M
        Initialize: l = 0, v_{k,m}^{(0)}, w_{k,m}^{(0)}, P_{k,m}^{(0)};
        Repeat:
          Solve (33) to obtain {v_{k,m}^{(l+1)}, w_{k,m}^{(l+1)}, P_{k,m}^{(l+1)}, d_{k,m}^{(l+1)}};
        l := l + 1;
        Until: Convergence;
        m := m + 1; Update D_{k,m}.
(3) End While
```

ALGORITHM 1: The proposed online Lyapunov algorithm.

algorithm within such quite short time slot. Specifically, our proposed algorithm aims at the downlink NOMA system considering the user clustering scenario [19]. In most existing cellular systems [1], with powerful computing hardware equipped in the base station (such as macro base station), our proposed algorithm can still be applicable even when more users are involved in the system.

From (11), if the channel response between the kth central user and the kth cell-edge user, $f_{k,m}$, is a circularly symmetric complex Gaussian random variable, $\gamma_{k,m,2,3}$ is unbounded given a positive $P_{k,m}$ because $|f_{k,m}|^2$ is an exponentially distributed random variable. To work out this problem, we propose a modified algorithm from Algorithm 1 to provide a feasible solution for the case when $\gamma_{k,m,2,3} > \gamma_{\max}$, where $\gamma_{\max}$ is defined as the maximum gain [27].

Firstly, we give the feasible set and its complement on the possible range of $\gamma_{k,m,2,3}$ as $\mathcal{G} \triangleq [0, \gamma_{\max}]$ and $\mathcal{G}^c \triangleq (\gamma_{\max}, \infty)$, respectively. The case $\gamma_{k,m,2,3} \in \mathcal{G}^c$ is defined as an outage event. Let ε denote the outage probability, i.e., $\Pr(\gamma_{k,m,2,3} \in \mathcal{G}^c) = \varepsilon$. When $\gamma_{k,m,2,3} \in \mathcal{G}$, Algorithm 1 still provides the feasible solution to problem P_0. When $\gamma_{k,m,2,3} \in \mathcal{G}^c$, the causal constraint (12) may be violated and the solution obtained by Algorithm 1 may be infeasible. Thus, the transmit power of the kth central user is given by

$$P_{k,m}^* = \min\left\{ \frac{\sigma^2 \left(2^{D_{k,m}/\alpha} - 1 \right)}{|f_{k,m}|^2}, P_c^{\max} \right\}. \tag{34}$$

Remark 3. The main idea of this modified algorithm above is that we use the buffer state $D_{k,m}$ to determine the transmit power of the central user in the current time slot. To maintain the causal constraint, i.e., $C_{k,m,2} \leq D_{k,m}$, and the transmit power constraint (15d), we can choose $P_{k,m}$ according to (34) adaptively. Specifically, when the central user transmits all the messages of the data buffer, the critical value of the transmit power can be obtained by setting $C_{k,m,2} = D_{k,m}$. To avoid the SINR $\gamma_{k,m,2,3}$ reaching the infeasible region, the transmit power should be less than $P_c^{\max}$, which is the power limit we have predefined.

4. Performance Analysis

In this section, the online Lyapunov algorithm's performance is analyzed. From [27], the perturbation parameters, $\{\theta_k\}$, are defined as

$$\theta_k = \alpha \log_2 \left(1 + \gamma_{\max} \right) + \rho, \quad k \in \mathcal{K}. \tag{35}$$

The asymptotic optimality of our proposed online Lyapunov algorithm is verified as follows.

We first define the following optimization problem related to P_0 as

$$P_2 : \max_{\substack{\{v_{k,m}, \\ w_{k,m}, \\ P_{k,m}\}}} \lim_{M \to \infty} \frac{1}{M} \sum_{m=1}^{M} \sum_{k=1}^{K} \mathbb{E}\left(R_{k,m,1} + R_{k,m,2} \right), \tag{36a}$$

$$\text{s.t.} \lim_{M \to \infty} \frac{1}{M} \sum_{m=1}^{M} \mathbb{E}\left(C_{k,m,1} - C_{k,m,2} \right) = 0, \tag{36b}$$

$$R_{k,m,2} \geq b_k, \tag{36c}$$

$$\sum_{k=1}^{K} \left(\|v_{k,m}\|^2 + \|w_{k,m}\|^2 \right) \leq P_b^{\max}, \tag{36d}$$

$$P_{k,m} \leq P_c^{\max}, \quad \forall k \in \mathcal{K}, \; m \in \mathcal{M}, \tag{36e}$$

where the data causal constraint (12) are replaced by constraint (36b), i.e., in a long time, the average number of data units stored at the kth central user equals to that transmitted to the kth cell-edge user.

Lemma 2. *Problem P_2 is the relaxed problem of P_0.*

Proof. For any feasible solution of problem P_0, because of the dynamics of the data buffer (13), we have

$$\lim_{M \to \infty} \frac{1}{M} \mathbb{E}\left(D_{k,M+1} \right) - \lim_{M \to \infty} \frac{1}{M} \mathbb{E}\left(D_{k,1} \right)$$

$$= \lim_{M \to \infty} \frac{1}{M} \sum_{m=1}^{M} \mathbb{E}\left(C_{k,m,1} - C_{k,m,2} \right). \tag{37}$$

Since $D_{k,M+1} < \infty$, $D_{k,1} = 0$, we have $\lim_{M \to \infty} 1/M\mathbb{E}(D_{k,M+1}) = \lim_{M \to \infty} 1/M\mathbb{E}(D_{k,1}) = 0$, i.e., (36b) is satisfied. Therefore, P_2 is the relaxed problem of P_0, and any feasible solution for (15a)–(15b) is also feasible for P_2.

To continue, we have the following result for problem P_2 whose proof can be found in ([20], (Appendix 4.A)). □

Lemma 3. *Let $\Psi = \{v_{k,m}, w_{k,m}, P_{k,m}\}$. There exists a solution to problem P_2 without the constraint (36b), denoted as Ψ_1, which satisfies the following equalities:*

$$\mathbb{E}\left[\widehat{R}_m \right]\Big|_{\Psi=\Psi_1} = \tau_1 + \varepsilon, \quad \forall m \in \mathcal{M},$$

$$\mathbb{E}\left(C_{k,m,1} - C_{k,m,2} \right) = \eta_k, \quad \forall k \in \mathcal{K}, \; m \in \mathcal{M}, \tag{38}$$

where τ_1 denotes the optimal objective value of problem P_2, η_k is an arbitrary small positive number, and $\varepsilon \geq 0$.

Then the asymptotic optimality of our proposed online Lyapunov algorithm is verified in the following lemma.

Lemma 4. *Denote* τ_2 *and* τ_3 *as the optimal objective values of problems* P_0 *and* P_1, *respectively. We have*

$$\tau_2 \leq \tau_3 + \frac{\varsigma}{\rho}, \tag{39}$$

where

$$\varsigma = \kappa + \sum_{k=1}^{K} \mathbb{E}\left(C_{k,m+1,1}^{\max}\theta_k + C_{k,m,2}^{\max}D^{\max}\right),$$

$$\kappa = \frac{1}{2}\mathbb{E}\left[\sum_{k=1}^{K}\left(C_{k,m+1,1}^{\max}\right)^2 + \left(C_{k,m,2}^{\max}\right)^2\right]. \tag{40}$$

Proof. Since the objective of problem P_1 is minimization of the drift-plus-penalty function, we have

$$\sum_{k=1}^{K}\mathbb{E}\left[L_{k,m+1} - L_{k,m} \mid D_{k,m}\right] - \rho\tau_3 \leq U_m. \tag{41}$$

From Lemma 1, we have

$$U_m \leq \mathbb{E}\left[\sum_{k=1}^{K}\left(D_{k,m} - \theta_k\right)\left(C_{k,m+1,1} - C_{k,m,2}\right)\right] - \mathbb{E}\left[\rho\widehat{R}_m \mid \mathcal{D}_m\right]$$

$$+ \kappa \leq KD^{\max}\eta - \rho\left(\tau_1 + \varepsilon\right) + \kappa, \tag{42}$$

where the last inequality comes from Lemma 3 and $\eta = \max_{k \in \mathcal{K}} \eta_k$. Combining (41) and (42), we have

$$\tau_1 - \tau_3 \leq \frac{1}{\rho}\left(KD^{\max}\eta - \sum_{k=1}^{K}\mathbb{E}\left[L_{k,m+1} - L_{k,m} | D_{k,m}|\right] - \rho\varepsilon + \kappa\right)$$

$$\leq \frac{1}{\rho}\left(KD^{\max}\eta + \kappa + \sum_{k=1}^{K}\mathbb{E}\left(C_{k,m+1,1}^{\max}\theta_k + C_{k,m,2}^{\max}D^{\max}\right)\right). \tag{43}$$

Since problem P_2 is a relaxation of problem P_0, we have $\tau_2 \geq \tau_1$. Let $\eta \longrightarrow 0$, we can obtain

$$\tau_2 - \tau_3 \leq \tau_1 - \tau_3 \leq \frac{\varsigma}{\rho}, \tag{44}$$

where $\varsigma = \kappa + \sum_{k=1}^{K}\mathbb{E}\left(C_{k,m+1,1}^{\max}\theta_k + C_{k,m,2}^{\max}D^{\max}\right)$. Summing up the equations for $m = 0, \ldots, M - 1$, then dividing by M and letting $M \longrightarrow \infty$, and when $\varsigma/\rho \longrightarrow 0$, the asymptotic optimality is proved. $\square$

5. Simulation Results

In this section, we provide the simulation results of our proposed algorithm. We consider the NOMA system with complex Gaussian random channels, where the channel responses $\mathbf{h}_{k,m}, \mathbf{g}_{k,m}$, and $f_{k,m}$ are modeled as Gaussian random variables with zero-mean and variance $\sigma_h^2 = 1, \sigma_g^2 = 0.4$, and $\sigma_f^2 = 2$, respectively. We assume that there are 3 clusters in the system and the base station has 4 antennas, and we set $T = 1$ ms, $B = 1$ MHz, $D^{\max} = 50$ bits, and

$P_b^{\max}/\sigma^2 = 10$ dB. The central users transmit signal with $P_c^{\max} = 0.1$ W, and the central users' QoS constraint is $r_k = 0.3$ bps/Hz. Moreover, the greedy algorithm and the conventional OMA are used for comparison. The greedy algorithm maximizes the sum rate during each time slot as local optimal solution. For the conventional OMA scheme, time-division multiple access is utilized, which means the base station serves all the users independently for different time slots. All the simulation works are conducted in the MATLAB-based framework. To solve the convex problem, we use the CVX optimization software [28].

Figures 2 and 3 show the long-term average sum rate of cell-edge users and all users versus time slot respectively, with the control parameter $\rho = 5$. As we can see, the overall performance of the proposed online Lyapunov algorithm is superior to the greedy algorithm and conventional OMA scheme because our proposed online Lyapunov algorithm can decode and forward messages according to the channel states and buffer states, i.e., when the channel states are good and buffers have enough storage, more information bits will be transmitted; when the channel states are poor and buffers are insufficient, the information will be stored for later transmission. However, the greedy algorithm can only optimize the current sum rate without considering the channel states and buffer states.

From the perspective of time slots, Figures 2 and 3 also reveal that our proposed Lyapunov algorithm is worse than the greedy method and the OMA scheme in the first several tens of time slots while outperforms them afterwards. This is because the information of the channel states and buffer states is inadequate in the beginning, which is adverse to the online optimization in the Lyapunov algorithm. Besides, the average sum rate obtained by our proposed Lyapunov algorithm tends to be stable after approximately 400 time slots, which is the time needed to reach the saturation, i.e., rate stability [20]. In order to get more insights into how the number of users impact on that time, we further evaluate the system performance based on the proposed Lyapunov algorithm with respect to different numbers of users. As a result, for the scenarios $K = 1, 2, 3, 4$, i.e., 2, 4, 6, 8 users, the required time slots to achieve stability are 130, 250, 400, 620, respectively. Concretely, it takes more time slots to reach the sum rate saturation if we increase the number of users in the system because more users included will add up the number of the causal constraints (12), which requires the additional Lyapunov optimization process to achieve the stability of the system.

Figure 4 describes the data buffer dynamic versus time slot with the control parameter $\rho = 5$. As Figure 4 shows, the data buffers' levels of the three clusters are confined within the range of θ_k under $D_{\max}$ and verify that the proposed online Lyapunov algorithm can stabilize the buffer length. The weight parameter ρ aims to remain the stability of the buffer queue in the Lyapunov drift-plus-penalty function. Actually, the choice of the value of ρ depends on the specific conditions. In the problem we formulated, we figure out that setting $\rho = 5$ can better strike the balance of good system performance and fast convergence rate. Thus, we suggest choosing the value of ρ based on the specific problem.
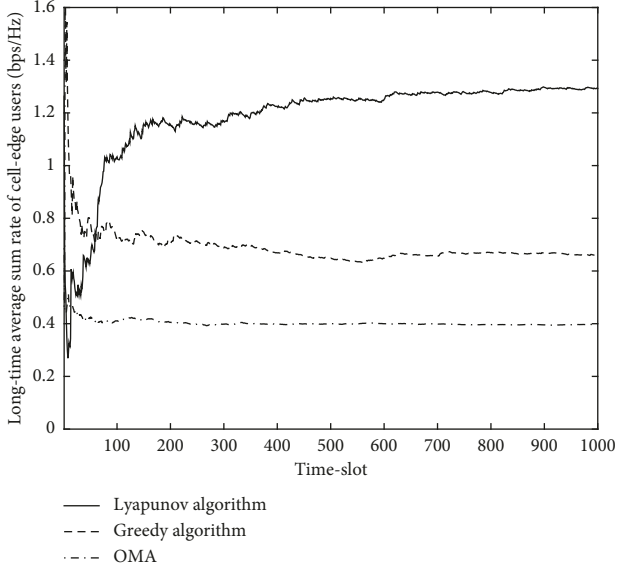
FIGURE 2: Long-term average sum rate of cell-edge users versus time slot.
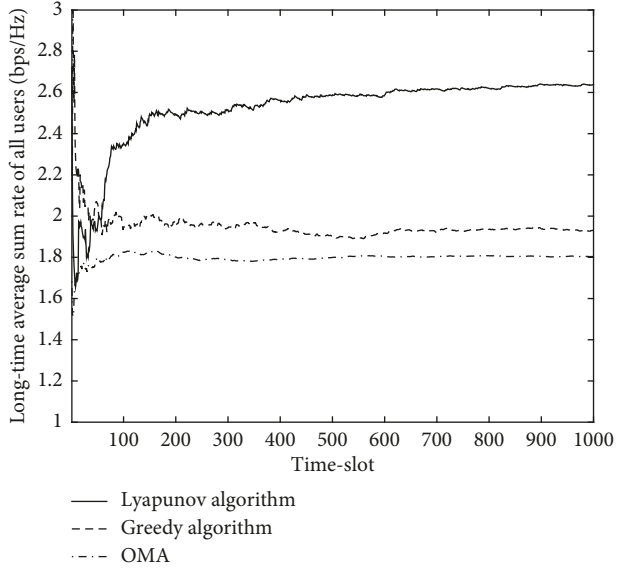

FIGURE 3: Long-term average sum rate of all users versus time slot.

FIGURE 4: Data buffer dynamics versus time slot.
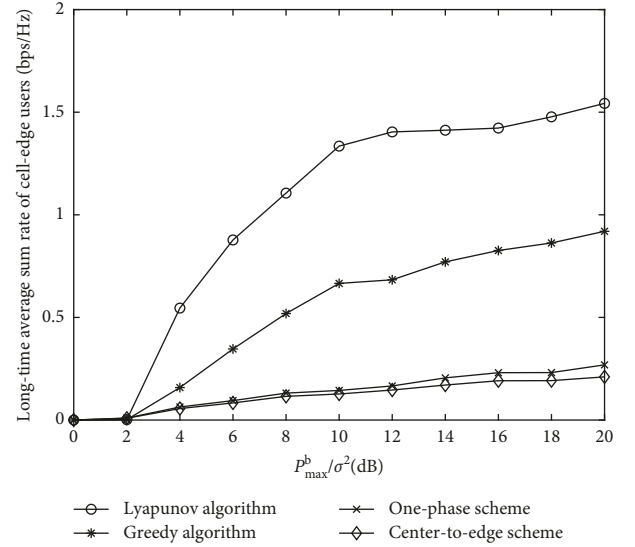
FIGURE 5: Long-term average sum rate of cell-edge users versus transmit power constraints.

Figure 5 shows the long-term average sum rate of cell-edge users with 400 time slots versus transmit power constraints at the base station. Here, we include another two schemes, where all the central users decode and forward the signals to cell-edge users in one phase after the base station transmission (denoted as "one-phase scheme" in the legend) and the cell-edge users only receive signals from central users while the base station can transmit signals in each phase (denoted as "center-to-edge scheme"). It is seen from Figure 5 that the long-term average sum rate increases with the growth of P_b^{max}. Moreover, Figure 5 also shows that our proposed online Lyapunov algorithm has a significant performance improvement over the greedy algorithm, the one-phase scheme 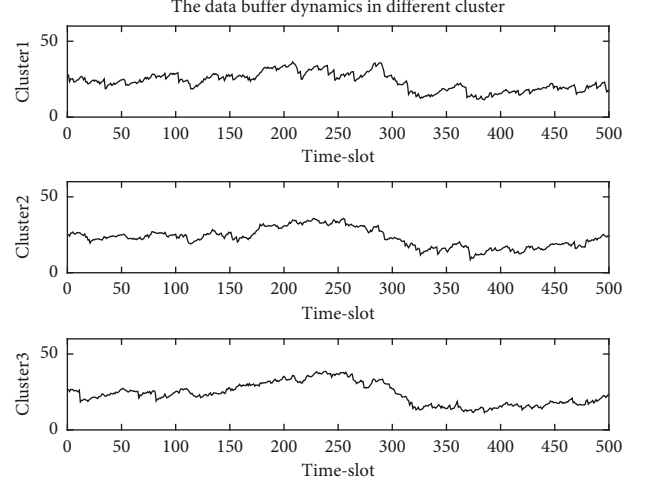and the center-to-edge scheme, especially when the transmit power is large. This can be explained as follows. For the one-phase scheme, when the central users transmit signals simultaneously in one phase, extra-interferences are introduced in the SINRs of the cell-edge users, which will cause degradation on their channels. For the center-to-edge scheme, except for the extrainterferences, the cell-edge users only receive signals from the central users, which act as relays. Thus, the amount of the transmitted information bits is reduced since the BS does not support the direct transmission to the cell-edge users. Instead, our proposed method can avoid the interference channels and obtain more effective performance by completing the transmission with $K + 1$ phases.

6. Conclusions

Considering the NOMA system with buffer-aided cooperative relaying, we have proposed the online Lyapunov

algorithm combined with the constrained concave-convex procedure to solve the causal long-term average transmission sum rate maximization problem and verified the asymptotic optimality of the proposed online Lyapunov. Simulation results have shown that the proposed online Lyapunov algorithm outperforms the greedy algorithm and the conventional OMA scheme.

Data Availability

No data were used to support this study.

Disclosure

Conflicts of Interest

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

Acknowledgments

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 *Proceedings of the 2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*, pp. 1–5, Dresden, Germany, June 2013.

[2] J. G. Andrews, S. Buzzi, W. Choi et al., "What will 5G be?," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[3] Q. Zhang, Q. Li, and J. Qin, "Robust beamforming for nonorthogonal multiple-access systems in MISO channels," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 12, pp. 10231–10236, 2016.

[4] Y. Li, M. Jiang, Q. Zhang, Q. Li, and J. Qin, "Secure beamforming in downlink MISO nonorthogonal multiple access systems," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 8, pp. 7563–7567, 2017.

[5] J. Chen, L. Yang, and M.-S. Alouini, "Physical layer security for cooperative NOMA systems," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 4645–4649, 2018.

[6] J. Chen, L. Yang, and M.-S. Alouini, "Performance analysis of cooperative NOMA schemes in spatially random relaying networks," *IEEE Access*, vol. 6, pp. 33159–33168, 2018.

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

[8] J.-B. Kim and I.-H. Lee, "Capacity analysis of cooperative relaying systems using non-orthogonal multiple access," *IEEE Communications Letters*, vol. 19, no. 11, pp. 1949–1952, 2015.

[9] Q. Zhang, Z. Liang, Q. Li, and J. Qin, "Buffer-aided non-orthogonal multiple access relaying systems in Rayleigh fading channels," *IEEE Transactions on Communications*, vol. 65, no. 1, pp. 95–106, 2017.

[10] C. K. Ho and R. Zhang, "Optimal energy allocation for wireless communications with energy harvesting constraints," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4808–4818, 2012.

[11] I. Ahmed, A. Ikhlef, R. Schober, and R. K. Mallik, "Power allocation for conventional and buffer-aided link adaptive relaying systems with energy harvesting nodes," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1182–1195, 2014.

[12] L. P. Qian, G. Feng, and V. C. M. Leung, "Optimal transmission policies for relay communication networks with ambient energy harvesting relays," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3754–3768, 2016.

[13] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, Athena Scientific, Belmont, MA, USA, 3rd edition, 2006.

[14] L. Huang and M. J. Neely, "Utility optimal scheduling in energy-harvesting networks," *IEEE/ACM Transactions on Networking*, vol. 21, no. 4, pp. 1117–1130, 2013.

[15] Y. Mao, J. Zhang, and K. B. Letaief, "A Lyapunov optimization approach for green cellular networks with hybrid energy supplies," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 12, pp. 2463–2477, 2015.

[16] K. W. Choi and D. I. Kim, "Stochastic optimal control for wireless powered communication networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 686–698, 2016.

[17] B. Kim, S. Lim, H. Kim et al., "Non-orthogonal multiple access in a downlink multiuser beamforming system," in *Proceedings of the IEEE MILCOM*, pp. 1278–1283, San Diego, CA, USA, November 2013.

[18] X. Zhu, C. Jiang, L. Kuang, N. Ge, and J. Lu, "Non-orthogonal multiple access based integrated terrestrial-satellite networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2253–2267, 2017.

[19] 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, no. 7, pp. 1465–1468, 2016.

[20] M. J. Neely, *Stochastic Network Optimization with Application to Communication and Queueing Systems*, Morgan & Claypool, San Rafael, CA, USA, 2010.

[21] Q. Li and L. Yang, "Robust optimization for energy efficiency in MIMO two-way relay networks with SWIPT," *IEEE Systems Journal*, pp. 1–12, 2019.

[22] Q. Li, Q. Zhang, and J. Qin, "Secure relay beamforming for SWIPT in amplify-and-forward two-way relay networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 11, pp. 9006–9019, 2016.

[23] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.

[24] I. Polik and T. Terlaky, "Interior point methods for nonlinear optimization," in *Nonlinear Optimization*, G. Di Pillo and F. Schoen, Eds., Springer-Verlag, New York, NY, USA, 1st edition, 2010.

[25] A. Damnjanovic, J. Montojo, Y. Wei et al., "A survey on 3GPP heterogeneous networks," *IEEE Wireless Communications*, vol. 18, no. 3, pp. 10–21, 2011.

[26] A. Ghosh, R. Ratasuk, B. Mondal, N. Mangalvedhe, and T. Thomas, "LTE-advanced: next-generation wireless broadband technology [Invited Paper]," *IEEE Wireless Communications*, vol. 17, no. 3, pp. 10–22, 2010.

[27] F. Amirnavaei and M. Dong, "Online power control optimization for wireless transmission with energy harvesting and storage," *IEEE Transactions on Wireless Communications*, vol. 15, no. 7, pp. 4888–4901, 2016.

[28] M. Grant and S. Boyd, "CVX: matlab software for disciplined convex programming," 2014, http://cvxr.com/cvx.