IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 22, 2017, accepted January 7, 2018, date of publication January 23, 2018, date of current version March 16, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2797209

Beamforming Design for Downlink Non-Orthogonal Multiple Access Systems

FUSHENG ZHU[1], ZHAOHUA LU[2], JIANYUE ZHU[1], JIAHENG WANG[1], (Senior Member, IEEE), AND YONGMING HUANG[1], (Senior Member, IEEE)
[1]National Mobile Communications Research Laboratory, Southeast University, Nanjing 210000, China
[2]Zhongxing Telecom Equipment, Shenzhen 518000, China

Corresponding author: Yongming Huang (huangym@seu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61720106003, Grant 61531011, Grant 61571107, and Grant 61711540305, and in part by the Natural Science Foundation of Jiangsu Province under Grant BK20160069.

ABSTRACT This paper investigates the beamforming design for multiple-input single-output nonorthogonal multiple access (NOMA) downlink systems. The NOMA beamforming design is formulated as a weighted sum rate maximization (WSRM) problem with decoding order constraints on the beamforming vectors and quality-of-service constraints for each user. We first investigate the feasibility of the NOMA WSRM problem and then propose an efficient method to achieve the optimized beamforming vectors. Furthermore, we consider the situation where user channels are homogeneous, which is particularly suitable for using NOMA. In this case, we show that the NOMA WSRM beamforming problem admits the favorable convexity and thus the optimal beamforming solution can be efficiently found. We further show that, with homogeneous channels, the optimal power allocation of the users can be analytically characterized.

INDEX TERMS Beamforming design, multiple-input single-output, nonorthogonal multiple access, quality-of-service, weighted sum rate.

I. INTRODUCTION

As a result of the constantly growing demand of new service and data traffic for wireless communications, the fifth generation (5G) communication systems propose higher requirements in data rates, lower latency, and massive connectivity [1]. In order to meet these high demands, some potential technologies, such as massive multiple-input multiple-output (MIMO) [2], small cell [3], millimeter wave [4] and device to device communication [5], [6] will be introduced to 5G communication systems. Specifically, multiantenna techniques will be widely adopted in 5G communication systems, since it provides the flexibility and degrees of freedom needed for efficient resource allocation. Most existing works, such as [7] and [8], consider orthogonal multiple access (OMA) techniques, which impose orthogonal policies for radio resource allocation. In this way, subcarriers and spatial resources are allocated exclusively to each user to avoid multiuser interference. However, such traditional resource allocation strategies cannot fully exploit limited spectral resources. To overcome this shortcoming, non-orthogonal multiple access (NOMA), which can support overloaded transmission over limited spectral resources and further improve the spectral efficiency [9], [10], will be introduced into 5G communication systems. Particularly, by using superposition coding at the transmitter with successive interference cancellation (SIC) at the receiver, NOMA allows multiple users to share the same (frequency, time, code, or spatial) resources at the same time, which can bring better performance in terms of spectral efficiency. Therefore, NOMA has received much attention recently.

Because of the non-orthogonality of users, resource allocation is a much more challenging task in NOMA systems compared to OMA systems [11]. Nevertheless, Zhu *et al.* [9] showed that single-input single-output NOMA (SISO NOMA) systems can achieve higher throughput compared to conventional SISO OMA systems. However, most existing works studying resource allocation in NOMA systems focused on SISO NOMA and the results cannot be applied to multiple-input single-output (MISO) NOMA systems directly. In fact, the transmit beamforming design in MISO NOMA systems leads to a rank constrained optimization problem [12], which is more challenging to solve than the power allocation for SISO NOMA systems.

In the literature, there are some works that investigated the beamforming design in MISO NOMA systems. Specifically, Cui *et al.* [13] studied the NOMA beamforming design for maximizing the sum rate of the strong users, and the NOMA beamforming for minimizing the transmit power was studied in [12] and [14]. In this paper, we investigate the NOMA beamforming design for maximizing the weighted sum rate of all users under decoding order constraints and quality-of-service (QoS) constraints.

In [15]–[17], there are some related works. Sun *et al.* [15] considered weighted sum rate maximization with QoS constraints in MISO NOMA systems and exploited a monotonic optimization method to solve the formulated problem. However, in [15] only two users (on one channel) were considered and the proposed beamforming method has an exponential complexity, which can only serves as a system benchmark. The problem of maximizing the sum rate in a downlink MISO NOMA system was investigated, where the nonconvex optimization problem was solved via minorization-maximization algorithm (MMA), but Hanif *et al.* [17] did not consider user weights and QoS requirements. Cai and Jin [16] investigated the user selection, beamforming and power allocation problem for maximizing the sum rate with QoS constraints in downlink MISO NOMA systems. However, this work was based on the assumption of strong channel gains and the solution is not applicable to the general case. Therefore, so far, there are no efficient method that can achieve the beamforming vectors for maximizing the weighted sum rate of MISO NOMA systems with QoS constraints.

In this paper, we investigate the beamforming design for maximizing weighted sum rate in downlink MISO NOMA systems. The technical contributions in this paper are summarized in the following.

- We consider maximizing weighted sum rate of MISO NOMA systems and take into account decoding order constraints on the beamforming vectors and QoS constraints for each user, which are often absent in existing works.
- The sufficient and necessary condition for the feasibility of the formulated NOMA beamforming problem is provided, and then an efficient method is proposed to solve the nonconvex NOMA WSRM beamforming problem.
- The situation of homogeneous channels is considered. In this case, we show that the NOMA WSRM beamforming problem admits the favorable convexity and thus the optimal beamforming solution can be efficiently found.
- Furthermore, with homogeneous channels, we are able to analytically characterize the optimal power allocation of the users.

The rest of the paper is organized as follows. Section II introduces the MISO NOMA system model and the problem formulation of the MISO NOMA beamforming design. Section III provides the feasibility condition and optimization method for the NOMA WSRM beamforming problem. Section IV investigates the MISO NOMA beamforming design with homogeneous channels. The performance of the proposed beamforming design is evaluated in Section V by simulation and the conclusion is drawn in Section VI.

Notations: We use boldface capital and lower case letters to denote matrices and vectors, respectively. $\mathrm{Tr}(\mathbf{A})$ and $\mathrm{Rank}(\mathbf{A})$ denote the trace and rank of matrix $\mathbf{A}$, respectively; $\boldsymbol{a}^H$ and $\boldsymbol{a}^T$ denote the Hermitian transpose and the transpose of vector $\boldsymbol{a}$, respectively; $\mathbf{A} \succeq 0$ indicates that $\mathbf{A}$ is a positive semidefinite matrix; $\mathbf{I}_N$ is the $\mathrm{N} \times \mathrm{N}$ identity matrix; $\mathbb{C}$ denotes the set of complex numbers; $|.|$ and $\|.\|$ denote the absolute value of a complex scalar and the Euclidean vector norm, respectively; $\nabla_x f(x)$ denotes the gradient vector of function $f(x)$ whose components are the partial derivatives of $f(x)$.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. SYSTEM MODEL

Consider a downlink NOMA network wherein a base station (BS) is equipped with T antennas and serves N single-antenna users. Let s_i be the message intended by user i with $E[|s_i|^2] = 1$ and let $\mathbf{w}_i \in \mathbb{C}^T$ be the complex beamforming vector for user i. In NOMA systems, the BS exploits the superposition coding and hence the received signal at each user k is

$$y_k = \sum_{i=1}^{N} \mathbf{h}_k^H \mathbf{w}_i s_i + n_k, \quad k \in \mathcal{N}$$

where $\mathcal{N} = \{k = 1, \cdots, N\}$, $\mathbf{h}_k = d_k^{-\alpha} \mathbf{g}_k \in \mathbb{C}^T$ (column vector) contains the channel coefficients from the BS to user k, d_k is the distance between user k and the BS, α is the path loss exponent, $\mathbf{g}_k$ follows a Rayleigh distribution, and $n_k \sim \mathcal{CN}(0, \sigma^2)$ is the additive white Gaussian noise (AWGN).

According to the NOMA principle, the individual users employ SIC to decode their signals. For a SISO system, if the channels are arranged in an increasing (decreasing) order, user i is able to decode signals of user k for $k < i$ ($k > i$) and remove them from its own signal, but treats the signals from user k for $k > i$ ($k < i$) as interference [9], [10]. However, the SISO ordering cannot be directly applied to MISO NOMA systems, wherein additional constraints have to be imposed to guarantee a similar decoding order. In order to study the design of the complex beamforming vectors, i.e., $\{\mathbf{w}_k\}_{k=1}^N$, we assume that the channel state information (CSI) is perfectly known at all notes and the user ordering is given.

B. PROBLEM FORMULATION

Assume that user 1 is the weakest user and not able to decode any interfering signals, while user N is the strongest user and able to nullify all other users' signals by performing SIC. The other users are placed in an increasing order with respect to their indices. Then, according to the NOMA principle, the achievable rate after performing SIC at user k is

$$R_k = \log\left(1 + \frac{\left|\mathbf{h}_k^H \mathbf{w}_k\right|^2}{\sum_{j=k+1}^{N} \left|\mathbf{h}_k^H \mathbf{w}_j\right|^2 + \sigma^2}\right), \quad k \in \mathcal{V}$$

$$R_N = \log\left(1 + \frac{\left|\mathbf{h}_N^H \mathbf{w}_N\right|^2}{\sigma^2}\right),$$

where $\mathcal{V} = \{k = 1, \cdots, N-1\}$. NOMA systems exploit the power domain for multiple access where different users are served at different power levels. Hence, user k can successfully decode and remove user j's signals (for $j = k+1, \cdots, N$) only if the following inequalities hold:

$$\begin{aligned} g_{k,j} &= \log\left(1+\frac{\left|\mathbf{h}_j^H\mathbf{w}_k\right|^2}{\sum_{m=k+1}^N\left|\mathbf{h}_j^H\mathbf{w}_m\right|^2+\sigma^2}\right) \\ &\quad - \log\left(1+\frac{\left|\mathbf{h}_k^H\mathbf{w}_k\right|^2}{\sum_{m=k+1}^N\left|\mathbf{h}_k^H\mathbf{w}_m\right|^2+\sigma^2}\right) \geq 0, \quad k,j\in\mathcal{M}, \end{aligned}$$

where $\mathcal{M} = \{k = 1, \cdots, N-1,\ j = k+1, \cdots, N\}$. The design of the beamforming vectors, i.e., $\{\mathbf{w}_k\}_{k=1}^N$, is the key to exploit the potential of the MISO NOMA system. In this paper, we investigate optimization of the beamforming vectors aiming to maximize the weighted sum rate of all users. The weighted sum rate maximization (WSRM) problem is formulated as

$$\begin{aligned} &\max_{\{\mathbf{w}_k\}} R_{\text{sum}} = \sum_{k=1}^N w_k R_k \\ &\text{s.t. C1}: \ g_{k,j}\geq 0,\ k,j\in\mathcal{M} \\ &\quad \text{C2}: \ \sum_{k=1}^N \|\mathbf{w}_k\|^2 \leq P \\ &\quad \text{C3}: \ \left|\mathbf{h}_k^H\mathbf{w}_1\right|^2 \geq \cdots \geq \left|\mathbf{h}_k^H\mathbf{w}_N\right|^2,\ k\in\mathcal{N} \\ &\quad \text{C4}: \ R_k \geq r_k,\ k\in\mathcal{N}. \end{aligned} \tag{1}$$

Here, w_k is the weight of user k. Constraint C1 is to guarantee successful SIC at user k and C2 is the total power constraint with power budget P. Particularly, to allocate non-trivial data rates to the weak users, which presents a lower decoding capability in a given order, constraint C3 must also be satisfied. In addition, in C4, $R_k \geq r_k$ denotes the QoS constraint with threshold $r_k \geq 0$ for each user k.

In problem (1), the weights can represent the priorities of different users and the QoS constraints can guarantee a minimum rate for each user. Although some works investigated beamforming designs for sum rate maximization, e.g., [16] and [17], the user weights and the QoS constraints were not considered. Note that the formulated NOMA WSRM beamforming problem in (1) is a difficult nonconvex problem, whose globally optimal solution is hard to find. In this paper, we devise an efficient method to solve this difficult problem. Furthermore, we show that, under some conditions of channels and user weights, problem (1) exhibits favorable convexity. In this case, we are able to find the optimal beamforming solution and analytically characterize the optimal power allocation.

III. NOMA WSRM BEAMFORMING DESIGN

In this section, we first investigate the feasibility of the NOMA WSRM problem in (1). Then, we transform problem (1) into a more tractable form by introducing new matrix variables. Based on the equivalent problem formulation, we propose an efficient method to find a locally optimal beamforming solution.

A. FEASIBILITY OF NOMA WSRM BEAMFORMING

A basic question about the NOMA WSRM Beamforming problem in (1) is whether it is feasible or not. The following result provides a necessary and sufficient condition on the feasibility of (1).

Proposition 1: Problem (1) *is feasible if and only if* $P \geq \sum_{k=1}^N \varphi_k$, *where* $\varsigma_k = 2^{r_k} - 1$ *and*

$$\varphi_k = \begin{cases} \max\left\{\varphi_{k+1}, \varsigma_k\left(\sum_{j=k+1}^N \varphi_j + \sigma^2/\mathbf{h}_k^H\mathbf{h}_k\right)\right\}, & k\in\mathcal{V} \\ \varsigma_N\sigma^2/\mathbf{h}_N^H\mathbf{h}_N, & k=N. \end{cases}$$

In particular, if $r_k \geq 1$ *for* $k = 1, \cdots, N$, *then*

$$\varphi_k = \begin{cases} \varsigma_k\left(\sum_{j=k+1}^N \varphi_j + \sigma^2/\mathbf{h}_k^H\mathbf{h}_k\right), & k\in\mathcal{V} \\ \varsigma_N\sigma^2/\mathbf{h}_N^H\mathbf{h}_N, & k=N. \end{cases}$$

Proof: First let $\varphi_k = \mathrm{Tr}\left(\mathbf{w}_k\mathbf{w}_k^H\right)$, from constraint $R_k \geq r_k$, we have $\varphi_k \geq \varsigma_k\left(\sum_{j=k+1}^N \varphi_j + \sigma^2/\mathbf{h}_k^H\mathbf{h}_k\right)$ for $k\in\mathcal{V}$ and $\varphi_N \geq \varsigma_N\sigma^2/\mathbf{h}_N^H\mathbf{h}_N$. Thus, the minimum transmit power of user N is $\varphi_N = \varsigma_N\sigma^2/\mathbf{h}_N^H\mathbf{h}_N$. With constraint C3, i.e., $|\mathbf{h}_k\mathbf{w}_1|^2 \geq \cdots \geq |\mathbf{h}_k\mathbf{w}_N|^2$, $k\in\mathcal{N}$, the transmit power of user k is $\varphi_k = \max\left\{\varphi_{k+1}, \varsigma_k\left(\sum_{j=k+1}^N \varphi_j + \sigma^2/\mathbf{h}_k^H\mathbf{h}_k\right)\right\}$ for $k\in\mathcal{V}$. If $r_k \geq 1$, so $2^{r_k}-1 \geq 1$ and $\varsigma_k\left(\sum_{j=k+1}^N \varphi_j + \sigma^2/\mathbf{h}_k^H\mathbf{h}_k\right) \geq \varphi_{k+1}$.

According to Proposition 1, to guarantee problem (1) feasible, the power budget of the BS can not be too small. This is to avoid that the order constraint C3 and QoS constraint C4 contradict each other. Therefore, the QoS threshold and total power budget of the BS should take rational values as indicated in Proposition 1. □

B. EQUIVALENT PROBLEM TRANSFORMATION

Assuming that the feasibility condition in Proposition 1 holds, now we consider how to solve problem (1). As mentioned above, problem (1) is a difficult nonconvex problem, and its original form is intractable. To address it, we introduce the following matrix variables:

$$\boldsymbol{Q}_k = \frac{1}{\sigma^2}\sum_{j=k}^N \mathbf{w}_j\mathbf{w}_j^H, \quad k\in\mathcal{N}.$$

In this way, we have

$$\begin{aligned} R_{\text{sum}} &= \sum_{k=1}^{N-1} w_k \log\left(\frac{\sum_{j=k}^N\left|\mathbf{h}_k^H\mathbf{w}_j\right|^2+\sigma^2}{\sum_{j=k+1}^N\left|\mathbf{h}_k^H\mathbf{w}_j\right|^2+\sigma^2}\right) \\ &\quad + w_N\log\left(1+\mathbf{h}_N^H\mathbf{Q}_N\mathbf{h}_N\right) \end{aligned}$$

$$= \sum_{k=1}^{N-1} w_k \left(\log\left(1+\mathbf{h}_k^H \boldsymbol{Q}_k \mathbf{h}_k\right) - \log\left(1+\mathbf{h}_k^H Q_{k+1}\mathbf{h}_k\right)\right) + w_N \log\left(1+\mathbf{h}_N^H \boldsymbol{Q}_N \mathbf{h}_N\right) = \sum_{k=1}^{N} f_k(\boldsymbol{Q}_k),$$

where $f_1(\boldsymbol{Q}_1) = w_1 \log\left(1+\mathbf{h}_1^H \boldsymbol{Q}_1 \mathbf{h}_1\right)$ and for $k = 2, \cdots, N$

$$f_k(\boldsymbol{Q}_k) = w_k \log\left(1+\mathbf{h}_k^H \boldsymbol{Q}_k \mathbf{h}_k\right) - w_{k-1} \log\left(1+\mathbf{h}_{k-1}^H \boldsymbol{Q}_k \mathbf{h}_{k-1}\right). \quad (2)$$

Meanwhile, the constraints can also be transferred into the equivalent forms. In particular, constraint C1 for successful SIC at user k is equal to

$$\text{C1}': \; g_{k,j}(\boldsymbol{Q}_k, \boldsymbol{Q}_{k+1}) = \log\left(\frac{\mathbf{h}_j^H \boldsymbol{Q}_k \mathbf{h}_j + 1}{\mathbf{h}_j^H \boldsymbol{Q}_{k+1} \mathbf{h}_j + 1}\right) - \log\left(\frac{\mathbf{h}_k^H \boldsymbol{Q}_k \mathbf{h}_k + 1}{\mathbf{h}_k^H \boldsymbol{Q}_{k+1} \mathbf{h}_k + 1}\right) \geq 0, \quad k, j \in \mathcal{M}.$$

The power constraint C2 is equal to C2$'$: $\mathrm{Tr}\left(\boldsymbol{Q}_1\right) \leq P$. Constraints C3 and C4 can both be linearized into

$$\begin{aligned} \text{C3}': \; & \mathbf{h}_k^H \left(\boldsymbol{Q}_1 - \boldsymbol{Q}_2\right) \mathbf{h}_k \geq \mathbf{h}_k^H \left(\boldsymbol{Q}_2 - \boldsymbol{Q}_3\right) \mathbf{h}_k \\ & \geq \cdots \geq \mathbf{h}_k^H \boldsymbol{Q}_N \mathbf{h}_k \geq 0, \quad k \in \mathcal{N} \\ \text{C4}': \; & \mathbf{h}_k^H \left(\boldsymbol{Q}_k - a_k \boldsymbol{Q}_{k+1}\right) \mathbf{h}_k + 1 - a_k \geq 0, \quad k \in \mathcal{V} \\ & \mathbf{h}_N^H \boldsymbol{Q}_N \mathbf{h}_N \geq a_N - 1, \end{aligned}$$

respectively, where $a_k = 2^{r_k}$.

Consequently, the NOMA WSRM problem in (1) can be equivalently transformed into the following problem:

$$\begin{aligned} \max_{\{\boldsymbol{Q}_k\}} \; & \sum_{k=1}^{N} f_k(\boldsymbol{Q}_k) \\ \text{s.t. } & \text{C1}', \; \text{C2}', \; \text{C3}', \; \text{C4}', \\ & \text{C5}: \mathrm{Rank}(\boldsymbol{Q}_k - \boldsymbol{Q}_{k+1}) \leq 1, \quad k \in \mathcal{N} \\ & \text{C6}: \boldsymbol{Q}_k \succeq 0, \; k \in \mathcal{N}. \end{aligned} \quad (3)$$

Constraint C5 follows from the fact that $\mathrm{Rank}(\boldsymbol{Q}_k - \boldsymbol{Q}_{k+1}) = \mathrm{Rank}\left(\mathbf{w}_k \mathbf{w}_k^H\right) \leq 1$. Note that problem (3) is still a nonconvex problem, as the objective function is not concave and constraints C1$'$ and C5 are not convex.

C. ALGORITHM DESIGN

In this subsection, we try to solve the equivalent problem in (3). As the NOMA WSRM Beamforming problem as well as its equivalent form in (1) is a nonconvex problem, finding its globally optimal solution generally requires exhaustive search, which leads to prohibitive complexity and cannot be used in practice. Therefore, our focus is on devising an efficient algorithm that can reach a locally optimal solution. For this purpose, we first introduce the following intermediate results.

Lemma 1: A lower bound of $f_k(\boldsymbol{Q}_k)$ at any feasible $\boldsymbol{Q}_k^0$ for $k = 2, \cdots, N$ is given by

$$F_k(\boldsymbol{Q}_k) = w_k \log\left(1+\mathbf{h}_k^H \boldsymbol{Q}_k \mathbf{h}_k\right) - \mathrm{Tr}\left(\boldsymbol{A}\left(\boldsymbol{Q}_k^0\right)\boldsymbol{Q}_k\right) - B\left(\boldsymbol{Q}_k^0\right),$$

where

$$\boldsymbol{A}\left(\boldsymbol{Q}_k^0\right) = \frac{1}{\ln 2} w_{k-1} \mathbf{h}_{k-1}\left(1+\mathbf{h}_{k-1}^H \boldsymbol{Q}_k^0 \mathbf{h}_{k-1}\right)^{-1} \mathbf{h}_{k-1}^H,$$

$$B\left(\boldsymbol{Q}_k^0\right) = w_{k-1} \log\left(1+\mathbf{h}_{k-1}^H \boldsymbol{Q}_k^0 \mathbf{h}_{k-1}\right) - \mathrm{Tr}\left(\boldsymbol{A}\left(\boldsymbol{Q}_k^0\right)\boldsymbol{Q}_k^0\right).$$

Proof: We have

$$w_{k-1} \log\left(1+\mathbf{h}_{k-1}^H \boldsymbol{Q}_k \mathbf{h}_{k-1}\right) \leq w_{k-1} \log\left(1+\mathbf{h}_{k-1}^H \boldsymbol{Q}_k^0 \mathbf{h}_{k-1}\right) + \mathrm{Tr}\left(\boldsymbol{A}\left(\boldsymbol{Q}_k^0\right)\left(\boldsymbol{Q}_k - \boldsymbol{Q}_k^0\right)\right)$$

and hence $f_k(\boldsymbol{Q}_k) \geq F_k(\boldsymbol{Q}_k)$. □

Lemma 2: A lower bound of $g_{k,j}(\boldsymbol{Q}_k, \boldsymbol{Q}_{k+1})$ at any feasible $(\boldsymbol{Q}_k^0, \boldsymbol{Q}_{k+1}^0)$ is given by

$$\begin{aligned} G_{k,j}(\boldsymbol{Q}_k, \boldsymbol{Q}_{k+1}) = \; & \log\left(\mathbf{h}_j^H \boldsymbol{Q}_k \mathbf{h}_j + 1\right) + \log\left(\mathbf{h}_k^H \boldsymbol{Q}_{k+1} \mathbf{h}_k + 1\right) \\ & - \mathrm{Tr}\left(\boldsymbol{E}_{k,j}\left(\boldsymbol{Q}_{k+1}^0\right)\boldsymbol{Q}_{k+1} + \boldsymbol{F}_{k,j}\left(\boldsymbol{Q}_k^0\right)\boldsymbol{Q}_k\right) \\ & - D_{k,j}(\boldsymbol{Q}_k^0, \boldsymbol{Q}_{k+1}^0), \quad k, j \in \mathcal{M} \end{aligned}$$

where

$$\boldsymbol{E}_{k,j}\left(\boldsymbol{Q}_{k+1}^0\right) = \frac{\mathbf{h}_j \mathbf{h}_j^H}{\ln 2\left(1+\mathbf{h}_j^H \boldsymbol{Q}_{k+1}^0 \mathbf{h}_j\right)},$$

$$\boldsymbol{F}_{k,j}\left(\boldsymbol{Q}_k^0\right) = \frac{\mathbf{h}_k \mathbf{h}_k^H}{\ln 2\left(1+\mathbf{h}_k^H \boldsymbol{Q}_k^0 \mathbf{h}_k\right)},$$

$$\begin{aligned} D_{k,j}(\boldsymbol{Q}_k^0, \boldsymbol{Q}_{k+1}^0) = \; & \log\left(\mathbf{h}_j^H \boldsymbol{Q}_k^0 \mathbf{h}_j + 1\right) + \log\left(\mathbf{h}_k^H \boldsymbol{Q}_{k+1}^0 \mathbf{h}_k + 1\right) \\ & - \mathrm{Tr}\left(\boldsymbol{E}_{k,j}\left(\boldsymbol{Q}_{k+1}^0\right)\boldsymbol{Q}_{k+1}^0 + \boldsymbol{F}_{k,j}\left(\boldsymbol{Q}_k^0\right)\boldsymbol{Q}_k^0\right). \end{aligned}$$

Proof: It follows from C1$'$ that

$$\begin{aligned} g_{k,j}(\boldsymbol{Q}_k, \boldsymbol{Q}_{k+1}) = \; & \log\left(\mathbf{h}_j^H \boldsymbol{Q}_k \mathbf{h}_j + 1\right) + \log\left(\mathbf{h}_k^H \boldsymbol{Q}_{k+1} \mathbf{h}_k + 1\right) \\ & - \log\left(\mathbf{h}_j^H \boldsymbol{Q}_{k+1} \mathbf{h}_j + 1\right) \\ & - \log\left(\mathbf{h}_k^H \boldsymbol{Q}_k \mathbf{h}_k + 1\right) \geq 0, \quad k, j \in \mathcal{M}. \end{aligned}$$

Similarly, we have

$$\begin{aligned} & \log\left(\mathbf{h}_j^H \boldsymbol{Q}_{k+1} \mathbf{h}_j + 1\right) + \log\left(\mathbf{h}_k^H \boldsymbol{Q}_k \mathbf{h}_k + 1\right) \\ & \leq \log\left(\mathbf{h}_j^H \boldsymbol{Q}_{k+1}^0 \mathbf{h}_j + 1\right) + \log\left(\mathbf{h}_k^H \boldsymbol{Q}_k^0 \mathbf{h}_k + 1\right) \\ & + \mathrm{Tr}\left(\boldsymbol{E}_{k,j}\left(\boldsymbol{Q}_{k+1}^0\right)\left(\boldsymbol{Q}_{k+1} - \boldsymbol{Q}_{k+1}^0\right)\right) \\ & + \mathrm{Tr}\left(\boldsymbol{F}_{k,j}\left(\boldsymbol{Q}_k^0\right)\left(\boldsymbol{Q}_k - \boldsymbol{Q}_k^0\right)\right) \end{aligned}$$

and hence $g_{k,j}(\boldsymbol{Q}_k, \boldsymbol{Q}_{k+1}) \geq G_{k,j}(\boldsymbol{Q}_k, \boldsymbol{Q}_{k+1})$. □

Using the results in Lemmas 1 and 2, we are able to bound the nonconcave objective function and constraints by concave ones. Specifically, the objective function can be lower bounded by $f_1(\boldsymbol{Q}_1) + \sum_{k=2}^{N} F_k(\boldsymbol{Q}_k)$, and constraint C1′ can be inner bounded by

$$\overline{\mathrm{C1}'}: \ G_{k,j}\left(\boldsymbol{Q}_k, \boldsymbol{Q}_{k+1}\right) \geq 0.$$

Therefore, given a feasible point $\{\boldsymbol{Q}_k^0\}$, we seek to solve the following approximated problem:

$$\begin{aligned} &\max_{\{\boldsymbol{Q}_k\}} f_1(\boldsymbol{Q}_1) + \sum_{k=2}^{N} F_k(\boldsymbol{Q}_k) \\ &\text{s.t. } \overline{\mathrm{C1}'}, \mathrm{C2}', \mathrm{C3}', \mathrm{C4}', \mathrm{C5}, \mathrm{C6}. \end{aligned} \tag{4}$$

Note that the approximated problem in (4) is still a nonconvex problem due to the rank constraint C5. Nevertheless, we show that the rank constraint can be removed without affecting the optimality of the obtained solution.

Theorem 1: In the absence of constraint C5, the solution to problem (4) *always satisfies Rank*$(\boldsymbol{Q}_k - \boldsymbol{Q}_{k+1}) \leq 1, k \in \mathcal{N}$.

Proof: First, with $\sum_{j=k}^{N} \boldsymbol{T}_j = \boldsymbol{Q}_k$, $\boldsymbol{T}_j = \mathbf{w}_j \mathbf{w}_j^H$ and

$$K_{k,j} = G_{k,j}\left(\sum_{j=k}^{N} \boldsymbol{T}_j, \sum_{j=k+1}^{N} \boldsymbol{T}_j\right),$$

problem (4) in the absence of constraint C5 can be written equivalently as

$$\max_{\{\boldsymbol{T}_k\}} f_1\left(\sum_{j=1}^{N} \boldsymbol{T}_j\right) + \sum_{k=2}^{N} F_k\left(\sum_{j=k}^{N} \boldsymbol{T}_j\right) \tag{5}$$

$$\text{s.t.} \quad K_{k,j} \geq 0, \quad k, j \in \mathcal{M} \tag{6}$$

$$\mathrm{Tr}\left(\sum_{j=1}^{N} \boldsymbol{T}_j\right) \leq P \tag{7}$$

$$\mathbf{h}_k^H \boldsymbol{T}_1 \mathbf{h}_k \geq \cdots \geq \mathbf{h}_k^H \boldsymbol{T}_N \mathbf{h}_k \geq a_N - 1, \quad k \in \mathcal{N} \tag{8}$$

$$\mathbf{h}_k^H \left(\sum_{j=k}^{N} \boldsymbol{T}_j - a_k \sum_{j=k+1}^{N} \boldsymbol{T}_j\right) \mathbf{h}_k + 1 - a_k \geq 0, \quad k \in \mathcal{V} \tag{9}$$

$$\boldsymbol{T}_k \succeq \mathbf{0}, \quad k \in \mathcal{N} \tag{10}$$

where (6), (7), (8), (9), and (10) correspond to $\overline{\mathrm{C1}}'$, C2′, C3′, C4, and C6. Note that problem (5) is a convex problem, whose optimal solution is characterized by the KKT conditions [18]. In particular, the Lagrangian of problem (5) is given by

$$\begin{aligned} L = &f_1\left(\sum_{j=1}^{N} \boldsymbol{T}_j\right) + \sum_{k=2}^{N} F_k\left(\sum_{j=k}^{N} \boldsymbol{T}_j\right) \\ &+ \sum_{k=1}^{N-1} \sum_{j=k+1}^{N} \mu_{k,j} K_{k,j} - \theta \left(\mathrm{Tr}\left(\sum_{j=1}^{N} \boldsymbol{T}_j\right) - P\right) \\ &+ \sum_{k=1}^{N-2} \sum_{j=1}^{N} \sigma_{k,j} \left(\mathbf{h}_j^H \boldsymbol{T}_k \mathbf{h}_j - \mathbf{h}_j^H \boldsymbol{T}_{k+1} \mathbf{h}_j\right) \\ &+ \sum_{k=1}^{N-1} \rho_k \left(\mathbf{h}_k^H \left(\sum_{j=k}^{N} \boldsymbol{T}_j - a_k \sum_{j=k+1}^{N} \boldsymbol{T}_j\right) \mathbf{h}_k + 1 - a_k\right) \\ &+ \rho_N \left(\mathbf{h}_N^H \boldsymbol{T}_N \mathbf{h}_N + 1 - a_N\right) + \sum_{k=1}^{N} \mathrm{Tr}\left(\boldsymbol{T}_k \boldsymbol{Y}_k\right), \end{aligned}$$

where $\mu_{k,j}$, θ, $\sigma_{k,j}$, ρ_k and $\boldsymbol{Y}_k$ are Lagrange multipliers associated with corresponding constraints. The KKT conditions for the optimal $\boldsymbol{T}_k^*$, $k \in \mathcal{N}$ are given by

$$\begin{aligned} \mu_{k,j}^*, \theta^*, \sigma_{k,j}^*, \rho_k^* \geq 0, \quad \boldsymbol{Y}_k^* \succeq \mathbf{0}, \\ \boldsymbol{Y}_k^* \boldsymbol{T}_k^* = \mathbf{0}, \quad \nabla_{\boldsymbol{T}_k^*} L = \mathbf{0}, \end{aligned}$$

where $\mu_{k,j}^*$, θ^*, $\sigma_{k,j}^*$, ρ_k^* and $\boldsymbol{Y}_k^*$ are the optimal Lagrange multipliers and $\nabla_{\boldsymbol{T}_k^*} L$ denotes the gradients of L with respect to $\boldsymbol{T}_k^*$. The condition $\nabla_{\boldsymbol{T}_k^*} L = \mathbf{0}$ can be expressed as

$$\theta^* \boldsymbol{I} = \varrho \mathbf{h}_k \mathbf{h}_k^H - \boldsymbol{Y}_k^*, \tag{11}$$

where ϱ is a scalar function the optimal Lagrange multipliers and optimal $\boldsymbol{T}_k^*$. Multiplying both sides of (11) by $\boldsymbol{T}_k^*$ and utilizing $\boldsymbol{Y}_k^* \boldsymbol{T}_k^* = \mathbf{0}$, we have $\theta^* \boldsymbol{T}_k^* = \varrho \mathbf{h}_k \mathbf{h}_k^H \boldsymbol{T}_k^*, k \in \mathcal{N}$ and θ^* is always positive. Applying basic rank inequalities for matrices, for $k \in \mathcal{N}$, the following relation hold:

$$\begin{aligned} \mathrm{Rank}\left(\boldsymbol{T}_k^*\right) &= \mathrm{Rank}\left(\theta^* \boldsymbol{T}_k^*\right) \\ &= \mathrm{Rank}\left(\varrho \mathbf{h}_k \mathbf{h}_k^H \boldsymbol{T}_k^*\right) \leq \mathrm{Rank}\left(\mathbf{h}_k \mathbf{h}_k^H\right) \leq 1. \end{aligned}$$

Thus, $\mathrm{Rank}(\boldsymbol{T}_k^*) = \mathrm{Rank}(\boldsymbol{Q}_k^* - \boldsymbol{Q}_{k+1}^*) \leq 1$, which completes the proof. □

Theorem 1 indicates that the relaxed version of (4) without constraint C5 is tight. Therefore, we are able to find the optimal solution to (4) by solving a convex problem, which can be addressed by a number of convex optimization methods and software such as CVX.

Algorithm 1 The Solution of Beamforming Vectors

1: **Initialization**: $\boldsymbol{Q}^0 = \mathbf{I}_N$.
2: **Repeat**
3: Obtain the lower bound of $f_k(\boldsymbol{Q}_k)$ and $g_{k,j}\left(\boldsymbol{Q}_k, \boldsymbol{Q}_{k+1}\right)$ according to Lemma 1 and Lemma 2.
4: Solve the WSRM problem (4) using CVX and achieve the solution $\boldsymbol{Q}^*$.
5: Set $\boldsymbol{Q}^0 = \boldsymbol{Q}^*$.
6: **Until** convergence
7: Obtain the beamforming vectors $\{\mathbf{w}_k\}_{k=1}^{N}$ via

$$\mathbf{w}_k \mathbf{w}_k^H = \boldsymbol{Q}_k^* - \boldsymbol{Q}_{k+1}^*, k \in \mathcal{V} \text{ and } \mathbf{w}_N \mathbf{w}_N^H = \boldsymbol{Q}_{N-1}^*.$$

Using the above results, we propose Algorithm 1 to solve the original problem (1) or (3). Specifically, Algorithm 1 iteratively updates the approximated point by solving problem (4). As it belongs to the class of successive convex approximation methods, Algorithm 1 is guaranteed to converge to a locally optimal solution to (3). After obtaining the solution $\left\{\boldsymbol{Q}_k^*\right\}_{k=1}^{N-1}$, we can obtain the beamforming vectors via the singular value decomposition (SVD) of $\mathbf{w}_k\mathbf{w}_k^H = \boldsymbol{Q}_k^* - \boldsymbol{Q}_{k+1}^*, k \in \mathcal{V}$ and $\mathbf{w}_N\mathbf{w}_N^H = \boldsymbol{Q}_{N-1}^*$.

IV. NOMA WSRM BEAMFORMING WITH HOMOGENEOUS CHANNELS

In problem (1), constraint C1 is to guarantee successful SIC at each user k. It ensures that users with high SINRs are able to decode the messages of the weak ones in the superposition coded signal. This constraint, however, makes (1) a difficult nonconvex problem. In this subsection, we show that under some channel conditions, this constraint may be automatically satisfied. In this case, we find the hidden convexity of the NOMA WSRM problem and consequently the optimal beamforming solution can be efficiently found.

We say the user channel are homogeneous if $\boldsymbol{h}_{k+1} = c_{k+1}\boldsymbol{h}_k,\ k \in \mathcal{V}$, where c_{k+1} for $k \in \mathcal{V}$ are complex constants and $|c_{k+1}|^2 \geq 1$. The homogeneous channel condition was first introduced in [17], which actually states the most suitable situation where NOMA beamforming can be applied. Indeed, NOMA beamforming is expected to achieve the largest performance gain when users are transmitting along the similar directions. Therefore, in this section, we focus on investigating beamforming design in this important situation. Without loss of generality (w.l.o.g.), the homogeneous channels $\boldsymbol{h}_{k+1} = c_{k+1}\boldsymbol{h}_k,\ k \in \mathcal{V}$ are ordered as $H_1 \leq \cdots \leq H_N$.

We also study the problem (1) and consider its equivalent problem with variable transformation $\boldsymbol{Q}_k = 1/\sigma^2 \sum_{j=k}^N \mathbf{w}_j\mathbf{w}_j^H$, which is described in detail in part III-B. The considered NOMA WSRM problem is

$$\max_{\{\boldsymbol{Q}_k\}} \sum_{k=1}^{N} f_k(\boldsymbol{Q}_k) \quad \text{s.t. C1}',\ \text{C2}',\ \text{C3}',\ \text{C4}',\ \text{C6} \tag{12}$$

where the rank constraint C5 is omitted and later we will show such a relaxation does not lose optimality. The following result states that, with homogeneous channels, condition C1′ is always satisfied.

Proposition 2: Given $\mathbf{h}_{k+1} = c_{k+1}\mathbf{h}_k,\ k \in \mathcal{V}$ and $|c_{k+1}|^2 \geq 1$, $g_{k,j} \geq 0,\ k,j \in \mathcal{M}$ always holds.

Proof: For $j = k+1$, with $\mathbf{h}_{k+1} = c_{k+1}\mathbf{h}_k$, $g_{k,j}$ can be given by

$$g_{k,j} = \log\left(1 + \frac{\left|\mathbf{h}_k^H\mathbf{w}_k\right|^2}{\sum_{m=k+1}^N \left|\mathbf{h}_k^H\mathbf{w}_m\right|^2 + 1/\left|c_{k+1}\right|^2}\right) - \log\left(1 + \frac{\left|\mathbf{h}_k^H\mathbf{w}_k\right|^2}{\sum_{m=k+1}^N \left|\mathbf{h}_k^H\mathbf{w}_m\right|^2 + 1}\right) \geq 0.$$

Due to $|c_{k+1}|^2 \geq 1$, we have $g_{k,j} \geq 0$. In addition, for other $j = k+2, \cdots, N$, we have $\boldsymbol{h}_j = c_jc_{j-1}\cdots c_{k+1}\boldsymbol{h}_k$ with $\left|c_jc_{j-1}\cdots c_{k+1}\right|^2 \geq 1$ and hence $g_{k,j} \geq 0$. Finally, $g_{k,j} \geq 0$ can be generalized to all $k = 1, \cdots, N-1$, which completes the proof. □

Proposition 2 indicates that with homogeneous channels, each user k is able to decode and remove user j's signals (for $j = k+1, \cdots, N$). Therefore, constraint C1′ in (12) can be removed and problem (12) can be written equivalently as

$$\max_{\{\boldsymbol{Q}_k\}} \sum_{k=1}^{N} f_k(\boldsymbol{Q}_k) \quad \text{s.t. C2}',\ \text{C3}',\ \text{C4}',\ \text{C6}. \tag{13}$$

In the following, we show that the optimal solution to problem (12) can be efficiently found and the optimal power allocation can even be analytically characterized.

A. CONVEXITY OF NOMA WSRM BEAMFORMING

With homogeneous channels, the NOMA WSRM problem in (12) possesses the following favorable property.

Theorem 2: Given $\mathbf{h}_k = c_k\mathbf{h}_{k-1}$, $|c_k|^2 \geq 1$ for $k \in \mathcal{O}$, problem (12) is a convex problem if one of the following condition holds :

T1: $w_{k-1} \leq w_k$

T2: $1 < \frac{w_{k-1}}{w_k} \leq \frac{\left(1+P\mathbf{h}_k^H\mathbf{h}_k\right)^2}{\left(1/|c_k|^2+P\mathbf{h}_k^H\mathbf{h}_k\right)^2}$.

Proof: First, according to Proposition 2, constraint C1′ can always be satisfied and constraints C2′, C3′, C4′ and C6 in problem (13) are linear. Therefore, it suffices to investigate the concavity of the objective function. According to (12), the objective function is $R_{\text{sum}} = \sum_{k=1}^N f_k(\boldsymbol{Q}_k)$ and hence we need to investigate the concavity of $f_k(\boldsymbol{Q}_k)$ for $k \in \mathcal{O}$. The second-order derivative of $f_k(\boldsymbol{Q}_k)$ is given by

$$\frac{d^2 f_k(\boldsymbol{Q}_k)}{d\boldsymbol{Q}_k^2} = \frac{w_{k-1}}{\left(1 + \mathbf{h}_{k-1}^H\boldsymbol{Q}_k\mathbf{h}_{k-1}\right)^2}\boldsymbol{H}_{k-1} - \frac{w_k}{\left(1 + \mathbf{h}_k^H\boldsymbol{Q}_k\mathbf{h}_k\right)^2}\boldsymbol{H}_k,$$

where $\boldsymbol{H}_k = vec(\mathbf{h}_k\mathbf{h}_k^H)\left(vec(\mathbf{h}_k\mathbf{h}_k^H)\right)^T$. Given $\mathbf{h}_k = c_k\mathbf{h}_{k-1}$, $|c_k|^2 \geq 1$ for $k \in \mathcal{O}$, we have

$$\frac{d^2 f_k(\boldsymbol{Q}_k)}{d\boldsymbol{Q}_k^2} = \alpha\left(\boldsymbol{Q}_k\right)\boldsymbol{H}_{k-1},$$

where

$$\alpha\left(\boldsymbol{Q}_k\right) = \frac{ST}{\left(1 + \mathbf{h}_{k-1}^H\boldsymbol{Q}_k\mathbf{h}_{k-1}\right)^2\left(1/\left|c_k\right|^2 + \mathbf{h}_{k-1}^H\boldsymbol{Q}_k\mathbf{h}_{k-1}\right)^2}$$

$$S = \sqrt{w_{k-1}}/\left|c_k\right|^2 - \sqrt{w_k} + \left(\sqrt{w_{k-1}} - \sqrt{w_k}\right)\mathbf{h}_{k-1}^H\boldsymbol{Q}_k\mathbf{h}_{k-1}$$

$$T = \sqrt{w_{k-1}}/\left|c_k\right|^2 + \sqrt{w_k} + \left(\sqrt{w_{k-1}} + \sqrt{w_k}\right)\mathbf{h}_{k-1}^H\boldsymbol{Q}_k\mathbf{h}_{k-1}.$$

If T1 holds, i.e., $w_{k-1} < w_k$, then $d^2 f_k(\boldsymbol{Q}_k)/d\boldsymbol{Q}_k^2$ is negative semi-definite and hence $f_k(\boldsymbol{Q}_k)$ is concave. On the other hand, from the variable transformation, we have $\text{Tr}\left(\boldsymbol{Q}_N\right) \leq \text{Tr}\left(\boldsymbol{Q}_{N-1}\right) \leq \cdots \leq \text{Tr}\left(\boldsymbol{Q}_1\right) \leq P$. Thus if T2 holds, we have

$$S \leq \sqrt{w_{k-1}}/\left|c_k\right|^2 - \sqrt{w_k} + \left(\sqrt{w_{k-1}} - \sqrt{w_k}\right)P\mathbf{h}_{k-1}^H\mathbf{h}_{k-1} \leq 0,$$

which also implies the concavity of $f_k(\boldsymbol{Q}_k)$. □

Remark 1: Theorem 2 indicates that the NOMA WSRM problem in (12) is a convex problem with homogeneous channels. In particular, the convexity of (12) depends on the user weights. From T1, if the user weights are in the same order as the channel gains, i.e., $w_1 \leq w_2 \leq \cdots \leq w_N$, then the objective function is concave and the problem is thus convex. Note that this situation includes the most common sum rate as a special case. On the other hand, the user weights can also be in the inverse order of the channel gains, i.e., $w_1 \geq w_2 \geq \cdots \geq w_N$, but in this case the ratio between w_k and w_{k-1} cannot be too large according to T2.

Consequently, if the feasibility condition in Proposition 1 and T1 or T2 in Theorem 2 hold, we are able to find the optimal solution to problem (12) via a number of convex optimization methods and software such as CVX. On the other hand, we will show that the NOMA WSRM in problem (12) in the absence of the rank constraint is tight.

Theorem 3: In the absence of constraint C5, the solution to problem (12) *always satisfies Rank*$(\boldsymbol{Q}_k - \boldsymbol{Q}_{k+1}) \leq 1$, $k \in \mathcal{N}$.

Proof: The proof is similar to that of Theorem 1, which is based on investigating the Lagrangian of problem (12) with $\sum_{j=k}^{N} \boldsymbol{T}_j = \boldsymbol{Q}_k$, $\boldsymbol{T}_j = \mathbf{w}_j \mathbf{w}_j^H$. Thus, the proof is omitted. □

Theorem 3 indicates that the solution of the convex problem in (12) always satisfies the rank constraint in C5. Therefore, we can provide the optimal beamforming solution to original problem (1) or (3) with homogeneous channels by solving a convex problem. Specifically, one can first obtain the optimal $\{\boldsymbol{Q}_k^*\}_{k=1}^{N-1}$ to problem (12) by using, e.g., the interior point method or the software CVX. Then, using the SVD $\mathbf{w}_k \mathbf{w}_k^H = \boldsymbol{Q}_k^* - \boldsymbol{Q}_{k+1}^*, k \in \mathcal{V}$ and $\mathbf{w}_N \mathbf{w}_N^H = \boldsymbol{Q}_{N-1}^*$, one can obtain the optimal beamforming vectors for the NOMA WSRM problem.

B. OPTIMAL POWER ALLOCATION

Given the convexity of the NOMA WSRM problem, one may wonder if its optimal solution can be analytically found. Unfortunately, a closed-formed expression for the optimal beamforming vectors does not exist. Nevertheless, we are able to analytically characterize the optimal power allocation of the users in this subsection.

For this purpose, let $\mathbf{w}_k = \sqrt{p_k}\bar{\mathbf{w}}_k$, where $p_k = \mathrm{Tr}\left(\mathbf{w}_k \mathbf{w}_k^H\right)$ is the transmit power of user k, and $\bar{\mathbf{w}}_k$ is the normalized beamforming vector with $\|\bar{\mathbf{w}}_k\|^2 = 1$. Suppose that $\{\bar{\mathbf{w}}_k\}$ are given and the NOMA WSRM problem with homogeneous channels reduces to the following power allocation problem:

$$\begin{aligned}
\max_{\{p_k\}} \; & R_{\mathrm{sum}} = \sum_{k=1}^{N} w_k R_k \\
\text{s.t.} \; & \sum_{k=1}^{N} p_k \leq P \\
& p_1 \geq p_2 \geq \cdots \geq p_N \\
& R_k \geq r_k, \quad k \in \mathcal{N}
\end{aligned} \tag{14}$$

where

$$\begin{aligned}
R_k &= \log\left(1 + \frac{p_k H_k}{\sum_{j=k+1}^{N} p_j H_k + 1}\right), \quad k \in \mathcal{V} \\
R_N &= \log\left(1 + p_N H_N\right),
\end{aligned}$$

and $H_k = \mathbf{h}_k^H \mathbf{h}_k / \sigma^2$. Assume w.l.o.g., the homogeneous channels $\boldsymbol{h}_{k+1} = c_{k+1}\boldsymbol{h}_k$, $k \in \mathcal{V}$ are ordered as $H_1 \leq \cdots \leq H_N$. Then, similarly, let $q_k = \sum_{j=k+1}^{N} p_j$ and hence the power allocation problem in (14) can be written as

$$\begin{aligned}
\max_{\{q_k\}} \; & \sum_{k=1}^{N} f_k(q_k) \\
\text{s.t.} \; & q_1 \leq P \\
& q_1 - q_2 \geq q_2 - q_3 \geq \cdots \geq q_N \geq (a_N - 1)/H_N \\
& q_{k+1} \leq q_k / a_k - \varepsilon_k, \quad k \in \mathcal{V}
\end{aligned} \tag{15}$$

where $a_k = 2^{r_k}$, $\varepsilon_k = \left(1 - 2^{-r_k}\right)/H_k$, $f_1(q_1) = w_1 \log\left(1 + q_1 H_1\right)$ and for $k \in \mathcal{O}$

$$f_k(q_k) = w_k \log\left(1 + q_k H_k\right) - w_{k-1} \log\left(1 + q_k H_{k-1}\right).$$

In the following parts, we will discuss the optimal power allocation in two cases, i.e., without or with QoS constraints.

1) OPTIMAL POWER ALLOCATION WITHOUT QoS

First, we consider the WSRM problem without QoS constraints, which is a simplified version of (15). According to (15), the corresponding problem is given by

$$\begin{aligned}
\max_{\{q_k\}} \; & \sum_{k=1}^{N} f_k(q_k) \\
\text{s.t.} \; & q_1 \leq P \\
& q_1 - q_2 \geq q_2 - q_3 \geq \cdots \geq q_N \geq 0
\end{aligned} \tag{16}$$

According to Theorem 2, if T1 or T2 holds, we can find the optimal beamforming vectors, however, which can not be shown in closed form. For problem (16), we show that it admits a closed form solution in some situations.

Proposition 3: Given $\mathbf{h}_{k+1} = c_{k+1}\mathbf{h}_k$, $k \in \mathcal{V}$ *and* $|c_{k+1}|^2 \geq 1$ *and* $w_1 < \cdots < w_N$, *the optimal solution to problem* (16) *is* $q_k^* = \frac{N-k+1}{N} P$ *for* $k \in \mathcal{N}$.

Proof: With homogeneous channels condition, i.e. $\mathbf{h}_{k+1} = c_{k+1}\mathbf{h}_k$, $k \in \mathcal{V}$ and $|c_{k+1}|^2 \geq 1$, we obtain $H_k = |c_k|^2 H_{k-1}$. Then the first order derivative of $f_k(q_k)$ is given by

$$\begin{aligned}
f_k'(q_k) &= \frac{1}{\ln 2}\left(\frac{w_k}{1 + q_k H_k} H_k - \frac{w_{k-1}}{1 + q_k H_{k-1}} H_{k-1}\right) \\
&= \frac{1}{\ln 2}\left(\frac{w_k}{1/|c_k|^2 + q_k H_{k-1}} - \frac{w_{k-1}}{1 + q_k H_{k-1}}\right) H_{k-1}.
\end{aligned} \tag{17}$$

With $|c_k|^2 \geq 1$, for $k \in \mathcal{O}$ and $w_1 < w_2 < \cdots < w_N$, $f_k'(q_k) \geq 0$ is always satisfied and then $f_k(q_k)$ is nondecreasing with respective to q_k. From $q_{N-1} - q_N \geq q_N$, we obtain $2q_N = q_{N-1}$. Then from $q_{N-2} - q_{N-1} \geq q_{N-1} - q_N$, we obtain $3q_N = q_{N-2}$ and so on until $Nq_N = q_1 = P$. Finally, we have $q_k = \frac{N-k+1}{N}P$. □

Remark 2: From Proposition 3 and $q_k = \sum_{j=k+1}^{N} p_j$, we have $p_k = P/N$ for $k = 1, \cdots, N$. Therefore, the optimal power allocation for NOMA WSRM is equal among all users under the weights condition $w_1 < \cdots < w_N$.

Proposition 4: Given $\mathbf{h}_{k+1} = c_{k+1}\mathbf{h}_k$, $k \in \mathcal{V}$ *and* $|c_{k+1}|^2 \geq 1$ *and*

$$T3: \frac{w_{k-1}}{w_k} \geq |c_k|, \quad k = 2, \cdots, N,$$

the optimal solution q_k *to problem* (16) *satisfies* $q_1 = P$ *and* $q_k = 0$ *for* $k \in \mathcal{O}$.

Proof: According to the first order derivative of $f_k(q_k)$, i.e., (17), it can be verified that if $\frac{w_{k-1}}{w_k} \geq |c_k|$, $f_k'(q_k) \leq 0$, $k \in \mathcal{O}$. Thus, each $f_k(q_k)$ is nonincreasing with respective to q_k in this case. Therefore, we have $q_1 = P$ and $q_k = 0$ for $k \in \mathcal{O}$. □

From proposition 4, under the weights condition T3 , the optimal solution is to allocate all power to the weakest user. However, from the fairness perspective, this situation is not desirable in NOMA systems.

2) OPTIMAL POWER ALLOCATION WITH QoS

In this part, we consider the optimal power allocation in term of WSRM with QoS constrains, i.e., problem (15). Now, we assume the feasibility condition in Proposition 1 and T1 or T2 in Theorem 2 hold, and seek the optimal power allocation to problem (15). In this case, although WSRM problem (15) is convex, a closed-form optimal solution is difficult to find. In the following, we can provide a closed-form solution when the number of users is limited to two.

Lemma 3: Given $\mathbf{h}_{k+1} = c_{k+1}\mathbf{h}_k$, $k \in \mathcal{V}$ *and* $|c_{k+1}|^2 \geq 1$, $N = 2$ *and* $w_1 \leq w_2$, *the optimal solution to problem* (15) *is* $q_1^* = P$, $q_2^* = \min\{P/2, P/a_1 - \rho_1\}$.

Proof: From problem (15), we have $q_2 \leq P/a_1 - \varepsilon_1$, $q_1 - q_2 \geq q_2$, implying $q_2 \leq q_1/2 = P/2$, and hence $q_2 \leq \min\{P/2, P/a_1 - \varepsilon_1\}$, which takes the equality as $f_2(q_2)$ is nondecreasing. □

Actually, due to the power order constraint $p_1 \geq p_2 \geq \cdots \geq p_N$, it is difficulty to analytically characterize the optimal powers. Indeed, in the absence of the power order constraint, there exists a closed-form power allocation solution to problem (14), which is given below.

Lemma 4: Given $\mathbf{h}_{k+1} = c_{k+1}\mathbf{h}_k$, $k \in \mathcal{V}$ *and* $|c_{k+1}|^2 \geq 1$ *and the order constraint is absent in problem* (14) *and* (15), *the solution to problem* (15) *is*

$$\tilde{q_k} = \begin{cases} P, & k = 1 \\ \tilde{q_{k-1}}/a_{k-1} - \varepsilon_{k-1}, & k \in \mathcal{O}, \end{cases} \tag{18}$$

and the solution to problem (14) *is*

$$\tilde{p_k} = \begin{cases} (1 - 1/a_k)\tilde{q_k} + \varepsilon_k, & k \in \mathcal{M} \\ \tilde{q_N}, & k = N. \end{cases} \tag{19}$$

Proof: Given $H_1 \leq \cdots \leq H_N$ and $w_1 < \cdots < w_N$, $f_k(q_k)$ is nondecreasing. In the absence of the power order constraint, the constraint $q_{k+1} \leq q_k/a_k - \varepsilon_k, k \in \mathcal{V}$ will take the equality, implying $q_{k+1} = q_k/a_k - \varepsilon_k, k \in \mathcal{V}$. Consequently, we obtain $p_k = q_k - q_{k-1} = (1 - 1/a_k)\tilde{q_k} + \varepsilon_k$ for $k \in \mathcal{M}$ and $\tilde{p_N} = \tilde{q_N}$. □

However, the above solution may not be optimal with the power order constraint. For example, assuming the number of users is $N = 2$ and the power order constraint is omitted, the solution will be $p_1 = (1 - 1/a_1)P + \varepsilon_1$ and $p_2 = P/a_1 - \varepsilon_1$ according to Lemma 4. Then, if $a_1 < 2$ and $P > 2\varepsilon_1 (2/a_1 - 1)$, we will have $p_1 < p_2$, which violates the power order constraint $p_1 \geq p_2$. Therefore, it is necessary to investigate the condition when the solution in Lemma 4 is really optimal.

Theorem 4: The solution in (19) *is optimal for problem* (14) *if*

$$T4: r_k \geq \log\left(2 - 2^{-r_{k+1}}\right), \quad i = 1, \cdots, N-2$$

and $r_{N-1} \geq 1$.

Proof: From (18) and (19), we obtain $\tilde{p_k} - \tilde{p_{k+1}} = (1 - 1/a_k)\tilde{q_k} + \varepsilon_k - (1 - 1/a_{k+1})\tilde{q_{k+1}} - \varepsilon_{k+1} = (1 - 1/a_k)\tilde{q_k} + \varepsilon_k - (1 - 1/a_{k+1})(\tilde{q_k}/a_k - \varepsilon_k) - \varepsilon_{k+1} = (1 - 2/a_k + 1/(a_k a_{k+1}))\tilde{q_k} + (2 - 1/a_{k+1})\varepsilon_k - \varepsilon_{k+1}$. It can be verified that if $1/a_k \leq 1/(2 - 1/a_{k+1})$, which implies $(2 - 1/a_{k+1})\varepsilon_k \geq \varepsilon_{k+1}$, we have $\tilde{p_k} \geq \tilde{p_{k+1}}$. Similarly, $\tilde{p_{N-1}} - \tilde{p_N} = (1 - 1/a_{N-1})\tilde{q_{N-1}} + \varepsilon_{N-1} - \tilde{q_N} = (1 - 1/a_{N-1})\tilde{q_{N-1}} + \varepsilon_{N-1} - \tilde{q_{N-1}}/a_{N-1}\tilde{q_{N-1}} = (1 - 2/a_{N-1})\tilde{q_{N-1}} + 2\varepsilon_{N-1}$, which is nonnegative if $a_{N-1} \geq 2$. □

Corollary 1: Condition T4 holds if $r_k \geq 1$ *for* $k \in \mathcal{M}$.

From Theorem 4, the optimal solution to problem (14) can be achieved only if the QoS thresholds of the first $N-1$ users are not small. Specifically, according to Corollary 1, the QoS threshold of the first $N-1$ users are required to be no less than 1bps/Hz, which is usually satisfied in practice. Therefore, the optimal power allocation is analytically characterized by Lemma 4.

V. NUMERICAL RESULTS

In this section, we evaluate the performance of the proposed solution to the weighted sum rate maximization problem in MISO NOMA systems. In simulations, for a given set of antennas $T = 2$ and users $N = 4$, the users are equally spaced in a cell with a radius of 100m and the BS is located in the center. For each channel, i.e., $\mathbf{h}_k = d_k^{-\alpha}\mathbf{g}_k$, the channel coefficient follows an i.i.d Gaussian distribution as $\mathbf{g}_k \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I})$ for $k \in \mathcal{N}$ and the path loss exponent is $\alpha = 3$. The noise power is $\sigma^2 = BN_0$, where the bandwidth is $B = $ 10MHz and the noise power spectral density is $N_0 = -174$dBm. The power budget of BS is 41dBm.

In Fig.1, we display the weighted sum rate achieved by Algorithm 1 and compare MISO NOMA scheme with the conventional MISO OMA scheme. We set the user weights to be $\boldsymbol{w}1 = [0.4, 0.3, 0.2, 0.1]$ or $\boldsymbol{w}2 = [0.25, 0.25, 0.25, 0.25]$ and the QoS thresholds to be $\boldsymbol{r}1 = [1, 1, 1, 1]$ bps/Hz or $\boldsymbol{r}2 = [2, 2, 2, 2]$ bps/Hz. As expected, our proposed MISO NOMA scheme outperforms MISO OMA. Furthermore, one can

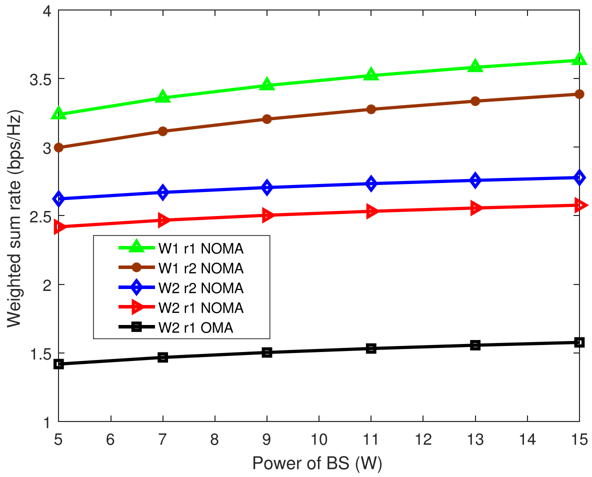

FIGURE 1. Weighted sum rate with different weights and QoS thresholds.

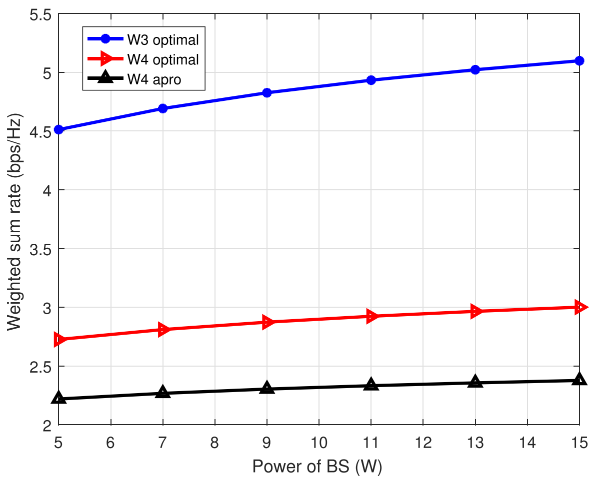

FIGURE 2. Weighted sum rate using two proposed solutions.

observe that the weighted sum rate can perform better if the weak users are equipped with more weights. In addition, if the users are equipped with equal weights, higher QoS thresholds will bring higher weighted sum rate. These results are owing to the fairness in NOMA systems, which is guaranteed by the decoding order constraints on beamforming vectors.

Fig.2 shows the weighted sum rate versus BS power using two proposed solutions. One solution uses Algorithm 1, another is the optimal solution to NOMA WSRM problem with homogeneous channels , i.e., $\mathbf{h}_k = c_k\mathbf{h}_{k-1},\ k \in \mathcal{O}$ and $|c_k|^2 \geq 1$ ($|c_k|^2 = 4$). According to Theorem 2, we let $\boldsymbol{w}3 = [0.4, 0.3, 0.2, 0.1]$ (T1 holds) and $\frac{w4_{k-1}}{w4_k} = \frac{(1+P\mathbf{h}_k^H\mathbf{h}_k)^2}{(1/|c_k|^2+P\mathbf{h}_k^H\mathbf{h}_k)^2}, k \in \mathcal{V}$ with $1^T\boldsymbol{w}4 = 1$ (T2 holds). As expected, the optimal solution performs well than the approximation method. Furthermore, one can see that if the weak user, such as user 1, is equipped with more weight, the system weighted sum rate will increase a lot. This is because more power is allocated to the weak user in NOMA systems.

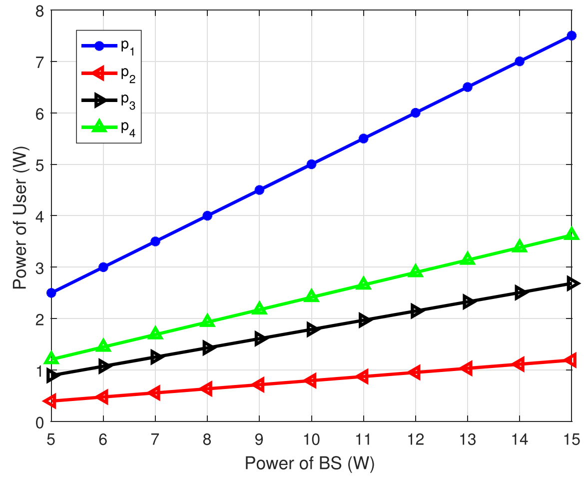

FIGURE 3. Power allocation of each other with QoS thresholds $\boldsymbol{r} = [1, 0.5, 0.5, 3]$ bps/Hz.

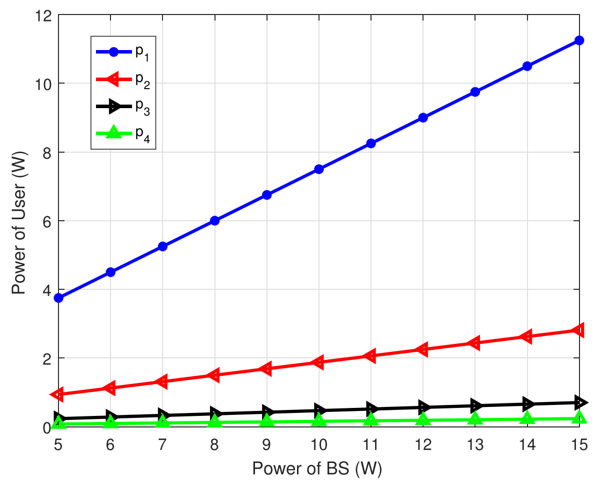

FIGURE 4. Power allocation of each other with QoS thresholds $\boldsymbol{r} = [1, 1, 1, 1]$ bps/Hz.

Fig.3 displays the power allocation of each user with equal weight (T1 holds) and QoS thresholds $\boldsymbol{r}3 = [1, 0.5, 0.5, 3]$ bps/Hz (T4 does not holds). On the other hand, Fig.4 shows the power allocation of each user with equal weight (T1 holds) and QoS thresholds $\boldsymbol{r}1 = [1, 1, 1, 1]$ bps/Hz (T4 holds). Fig.4 is the optimal power allocation, while the Fig.3 is the solution in Lemma 4. As can be seen in Fig.3, with small QoS thresholds, the solution causes $p_2 < p_3 < p_4$, violating the power order constraint $p_1 \geq \cdots \geq p_4$. Hence, the power order constraint cannot be omitted if T4 does not holds.

VI. CONCLUSION

In this paper, we have studied the beamforming design for MISO NOMA downlink systems aiming to maximize the weighted sum rate. The decoding order constraints and QoS constraints have been explicitly taken into account. We provided the necessary and sufficient condition characterizing the feasibility of the NOMA WSRM beamforming problem and proposed an efficient method optimize the beamforming vectors. Furthermore, we have studied the situation of homogeneous channels, where the user channels have similar directions. In this case, we have shown that the NOMA WSRM beamforming problem has a hidden convexity and its optimal solution can be efficiently found. In addition, we have also analytically characterized the optimal powers allocated to users with homogeneous channels.

REFERENCES

[1] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[2] V. Jungnickel *et al.*, "The role of small cells, coordinated multipoint, and massive MIMO in 5G," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 44–51, May 2014.

[3] N. Bhushan *et al.*, "Network densification: The dominant theme for wireless evolution into 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 82–89, Feb. 2014.

[4] S. He, Y. Huang, J. Wang, L. Yang, and W. Hong, "Joint antenna selection and energy-efficient beamforming design," *IEEE Signal Process. Lett.*, vol. 23, no. 9, pp. 1165–1169, Sep. 2016.

[5] K. Doppler, M. Rinne, C. Wijting, C. B. Ribeiro, and K. Hugl, "Device-to-device communication as an underlay to LTE-advanced networks," *IEEE Commun. Mag.*, vol. 47, no. 12, pp. 42–49, Dec. 2009.

[6] J. Guo, S. Durrani, X. Zhou, and H. Yanikomeroglu, "Device-to-device communication underlaying a finite cellular network region," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 332–347, Jan. 2017.

[7] G. Zheng, K.-K. Wong, and T.-S. Ng, "Throughput maximization in linear multiuser MIMO–OFDM downlink systems," *IEEE Trans. Veh. Technol.*, vol. 57, no. 3, pp. 1993–1998, May 2008.

[8] D. W. K. Ng, E. S. Lo, and R. Schober, "Dynamic resource allocation in MIMO-OFDMA systems with full-duplex and hybrid relaying," *IEEE Trans. Wireless Commun.*, vol. 60, no. 5, pp. 1291–1304, May 2012.

[9] J. Zhu, J. Wang, Y. Huang, S. He, X. You, and L. Yang, "On optimal power allocation for downlink non-orthogonal multiple access systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2744–2757, Dec. 2017.

[10] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.

[11] L. Dai, B. Wang, Y. Yuan, S. Han, C. L. I, and Z. Wang, "Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.

[12] Z. Chen, Z. Ding, P. Xu, and X. Dai, "Optimal precoding for a QoS optimization problem in two-user MISO-NOMA downlink," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1263–1266, Jun. 2016.

[13] J. Cui, Z. Ding, and P. Fan, "Beamforming design for MISO non-orthogonal multiple access systems," *IET Commun.*, vol. 11, no. 5, pp. 720–725, Apr. 2017.

[14] Z. Chen, Z. Ding, X. Dai, and G. K. Karagiannidis, "On the application of quasi-degradation to MISO-NOMA downlink," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6174–6189, Dec. 2016.

[15] Y. Sun, D. W. K. Ng, and R. Schober, "Optimal resource allocation for multicarrier MISO-NOMA systems," in *Proc. IEEE Int. Commun. Conf.*, Paris, France, May 2017, pp. 1–7.

[16] W. Cai and Y. Jin, "Low complexity BF scheme for downlink MISO-NOMA systems," *IET Commun.*, vol. 11, no. 10, pp. 1566–1573, 2017.

[17] M. F. Hanif, Z. Ding, T. Ratnarajah, and G. K. Karagiannidis, "A minorization-maximization method for optimizing sum rate in the downlink of non-orthogonal multiple access systems," *IEEE Trans. Signal Process.*, vol. 64, no. 1, pp. 76–88, Jan. 2016.

[18] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

FUSHENG ZHU is currently pursuing the Ph.D. degree in information and communication engineering with the School of Information Science and Engineering, Southeast University, Nanjing. His current research interests include massive multiple-input multiple-output systems, millimeter wave wireless communications, and nonorthogonal multiple access.

ZHAOHUA LU received the Ph.D. degree from Tianjin University in 2006. He is a Senior Wireless Communication System Research Expert with ZTE Corporation, where he has been involved in the field of wireless communication system design and the key technologies of the physical layer. He has many technical contributions, papers, and patents in interference mitigation in the multiple-input multiple-output field.

JIANYUE ZHU received the B.S. degree from Nanjing Agricultural University, Nanjing, China, in 2015. She is currently pursuing the Ph.D. degree in information and communication engineering with the School of Information Science and Engineering, Southeast University, Nanjing. Her current research interests include nonorthogonal multiple access and optimization theory.

JIAHENG WANG received the B.E. and M.S. degrees from Southeast University, Nanjing, China, in 2001 and 2006, respectively, and the Ph.D. degree in electronic and computer engineering from The Hong Kong University of Science and Technology, Hong Kong, in 2010.

He is currently a Full Professor with the National Mobile Communications Research Laboratory, Southeast University, Nanjing. From 2010 to 2011, he was with the Signal Processing Laboratory, KTH Royal Institute of Technology, Stockholm, Sweden. He also held visiting positions with the Friedrich Alexander University Erlangen–Nürnberg, Nürnberg, Germany; and the University of Macau, Macau. He has authored over 90 articles on international journals and conferences. His research interests are mainly on optimization in signal processing and wireless communications.

Dr. Wang was a recipient of the Humboldt Fellowship for Experienced Researchers, and the Best Paper Award in WCSP 2014. He serves as an Associate Editor for the IEEE Signal Processing Letters.

YONGMING HUANG received the B.S. and M.S. degrees from Nanjing University, Nanjing, China, in 2000 and 2003, respectively, and the Ph.D. degree in electrical engineering from Southeast University, Nanjing, in 2007. From 2008 to 2009, he visited the Signal Processing Laboratory, School of Electrical Engineering, Royal Institute of Technology, Stockholm, Sweden. Since 2007, he has been a Faculty Member with the School of Information Science and Engineering, Southeast University, where he is currently a Full Professor. He has authored over 200 peer-reviewed papers, holds over 50 invention patents, and submitted over ten technical contributions to the IEEE standards. His current research interests include multiple-input multiple-output wireless communications, cooperative wireless communications, and millimeter wave wireless communications. He has served as an Associate Editor for the IEEE Transactions on Signal Processing, the IEEE Wireless Communications Letters, the *EURASIP Journal on Advances in Signal Processing*, and the *EURASIP Journal on Wireless Communications and Networking*.