Utility Maximization for Large-Scale Cell-Free Massive MIMO Downlink

Muhammad Farooq, *Graduate Student Member, IEEE*, Hien Quoc Ngo, *Senior Member, IEEE*, Een-Kee Hong, *Senior Member, IEEE*, and Le-Nam Tran, *Senior Member, IEEE*

***Abstract*— We consider utility maximization problems in the downlink cell-free massive multiple-input multiple-output (MIMO) whereby a large number of access points (APs) simultaneously serve a group of users. Four fundamental maximization objectives are of interest: (i) average spectral efficiency (SE), (ii) proportional fairness, (iii) harmonic-rate, and (iv) minimum SE of all users, subject to a sum power constraint at each AP. As considered problems are non-convex, existing solutions normally rely on successive convex approximation (SCA) and use off-the-shelf convex solvers, which implement an interior-point algorithm, to solve derived convex problems. The complexity of such methods scales quickly with the problem size. Therefore, we propose an accelerated projected gradient method to solve the considered problems. Particularly, each iteration of the proposed solution is given in a closed form and only requires the first order oracle of the objective, rather than the Hessian matrix as in known solutions, and thus is much more memory efficient. Numerical results demonstrate that our proposed solution achieves the same utility performance but with far less runtime, compared to the SCA method. Simulation results show that large-scale cell-free massive MIMO has the intrinsic user fairness, i.e. the four utility functions can deliver nearly uniformed services to all users.**

***Index Terms*— Cell-free massive MIMO, sum-rate, power control, gradient.**

I. Introduction

MULTIPLE-INPUT multiple-output (MIMO) is the underlying technology in the physical layer of many modern wireless communications standards. The use of multiple antennas at transceivers can offer high data rates and high reliability by exploiting spatial and diversity gains [2]–[4]. To meet a set of requirements for 5G networks, MIMO has evolved into so-called massive MIMO where a very large number of antennas are deployed at each base station (BS) to serve many users at the same time [5], [6]. In particular, massive MIMO has been implemented in the first version of 5G NR [7]. Since 5G still follows the conventional design of a *cellular* network like its predecessors, inter-cell interference remains a fundamental problem, and thus massive MIMO cannot be unlocked to its full potential [8].

There are two types of massive MIMO in terms of the service area: colocated massive MIMO and distributed massive MIMO. For the former, all the antennas are placed in a small area and therefore, the processing complexity requirement is very low. For the latter, on the other hand, the antennas are distributed to serve a relatively much larger area. These systems are more diverse against the shadow fading and they have a large coverage area [9]. There is no doubt that the distributed massive MIMO is better than colocated massive MIMO but due to the more processing complexity and high cost requirements [10], the scalability remains an active area of research in distributed systems.

Cell-free massive multiple-input multiple-output (MIMO) was introduced in [11] as a major leap of massive MIMO technology to overcome the inter-cell interference which is the main inherent limitation of cellular-based networks. In cell-free massive MIMO, many access points (APs) distributed over the whole network serve many users in the same time-frequency resource. In practice, the APs are more irregularly distributed rather than a uniform distribution [12]. There are no cells, and hence, no boundary effects. Unlike colocated massive MIMO, each AP in cell-free massive MIMO is equipped with just a few antennas. But an important point is that when the number of APs is very large, cell-free massive MIMO is still able to exploit the favorable propagation and channel hardening properties, like colocated massive MIMO. In particular, with favorable propagation, APs can use simple linear processing techniques to combine the signals in the uplink, and precode the symbols in the downlink. With channel hardening, decoding the signals using the channel statistics (large-scale fading coefficients) can provide good performance. Note that in some propagation environments, the level of channel hardening in cell-free massive MIMO is lesser than that in colocated massive MIMO [13].

The research on cell-free massive MIMO is still in its infancy and thus deserves more extensive and thorough studies. We discuss here some of the important and related studies in the literature. In [11], Ngo *et al.* considered the problem

Manuscript received January 7, 2021; revised May 15, 2021; accepted July 8, 2021. Date of publication July 16, 2021; date of current version October 18, 2021. This publication has emanated from research supported in part by a Grant from Science Foundation Ireland under Grant number 17/CDA/4786. The work of Hien Quoc Ngo was supported by the U.K. Research and Innovation Future Leaders Fellowships under Grant MR/S017666/1. This article was presented at 2020 IEEE International Symposium on Personal, Indoor and Mobile Radio Communications [1]. The associate editor coordinating the review of this article and approving it for publication was X. Chen. *(Corresponding author: Muhammad Farooq.)*

Muhammad Farooq and Le-Nam Tran are with the School of Electrical and Electronic Engineering, University College Dublin, Dublin 4, D04 V1W8 Ireland (e-mail: muhammad.farooq@ucdconnect.ie; nam.tran@ucd.ie).

Hien Quoc Ngo is with the Institute of Electronics, Communications and Information Technology, Queen's University Belfast, Belfast BT3 9DT, U.K. (e-mail: hien.ngo@qub.ac.uk).

Een-Kee Hong is with the Department of Electronic Engineering, Kyung Hee University, Yongin 130701, Republic of Korea (e-mail: ekhong@khu.ac.kr).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCOMM.2021.3097718.

Digital Object Identifier 10.1109/TCOMM.2021.3097718

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

of minimum rate maximization to provide uniformly good services to all users. The problem was then solved using a bisection search and a sequence of second-order cone feasibility problems. In [14], Nguyen *et al.* adopted zero-forcing precoding and studied the energy efficiency maximization (EEmax) problem. In this work, an iterative method based on successive convex approximation (SCA) was derived. In [15], both the max-min fairness and sum-rate maximization problems for user-centric cell-free massive MIMO were considered and solved by SCA. The SCA-based method was also used in [16] and [17] to solve the EEmax and max-min fairness power controls with different cell-free massive MIMO setups, respectively. In [18], Alonzo *et al.* proposed an energy-efficient power control algorithm for cell-free and user-centric massive MIMO at millimeter wave frequencies using a sequential optimization method. In [19], Bashar *et al.* investigated the problem of max-min signal to interference plus noise ratio (SINR) for the uplink of cell-free massive MIMO, which is solved by an alternating optimization method. In [20], spectral efficiency (SE) and energy efficiency were optimized for full-duplex cell-free massive MIMO by a combination of the inner approximation framework and Dinkelbach's method. In [21], dynamic resource allocation for the uplink of both colocated and cell-free massive MIMO using Lyapunov optimization techniques. Furthermore, the integration of cell-free massive MIMO with other radio access techniques can bring further advantages. For example, the integration of cell-free massive MIMO with non-orthogonal multiple access can support significantly more users compared to the orthogonal multiple access counterpart [22].

A common feature of all the above mentioned pioneer studies on cell-free massive MIMO is the use of a second-order interior-point method for the optimization of several performance measures in cell-free massive MIMO. This requires the computation of the Hessian matrix of the objective, and thus their computational complexity and memory requirement makes them impossible to implement and investigate the performance of large-scale cell-free massive MIMO. To motivate our proposed method, let us consider an example where 2000 APs are deployed to serve 200 users over an area of 1 km^2, which is typical for an urban area in our vision. The power control problem arising from this scenario has 4×10^5 optimization variables. Consequently, we would basically need 160 GB of memory to store the resulting Hessian matrix, assuming a single-precision floating-point format. It is this immense memory requirement of the existing power control methods that only allows us to implement as well as to characterize the performance of cell-free massive MIMO for a relatively small-scale system. For example, the work of [11] was able to consider an area of 1 km^2, consisting of 100 APs serving 40 users. Numbers with the same order of magnitude were also observed in the above mentioned papers. The performance of these scenarios fractionally represents the full potential of cell-free massive MIMO.

In this paper, we consider the downlink of cell-free massive MIMO. Similar to many previous studies (e.g. [16]), the conjugate beamforming (based on small-scale fading) is adopted at each AP due to its simplicity. Under this setup, we consider power control problems (based on large-scale fading) to maximize four system-wide utility functions: the total SE, proportional fairness, harmonic-rate, and the minimum rate.[1] The considered power control problems are known to be nonconvex and thus difficult to solve. To this end, we propose a novel low-complexity method, called the accelerated projected gradient (APG) method, which is derived from a gradient based method for nonconvex programming in [23]. A similar method has been used in [24] to solve the EEmax problem. In particular, the proposed method can deal with the four considered utility functions in a unified framework. Being a first order method, the proposed method is memory efficient. Specifically, referring to the motivating example in the preceding paragraph, our proposed method only requires a memory of 8 MB, which is affordable by most, if not all, modern desktops. Consequently, the proposed method is more practically feasible and also allows us to study the performance limits of cell-free massive MIMO for large-scale settings that have not been reported previously. In particular, our main contributions are as follows

- We present the APG method, a special variant of an accelerated proximal gradient method introduced in [23] for nonconvex programming, to efficiently solve the considered utility maximization problems in a unified framework. Particularly, each iteration of the proposed iterative power control algorithms is given in closed form and can be done in parallel. To achieve this, we reformulate the considered problems so that the gradient of the objective is Lipschitz continuous and the projection is still efficient to compute.
- We provide a complexity and convergence analysis of the proposed method. Specifically, the per-iteration complexity of our proposed method is only $\mathcal{O}(K^2M)$ as compared to the per-iteration complexity of $\mathcal{O}\left(\sqrt{K+M}M^3K^4\right)$ for the SCA-method in [16], where M and K are the numbers of APs and users, respectively. Accordingly, the proposed method takes much reduced run time to return a solution as numerically shown in Section V. As a result, the propose method can lay the foundation to numerically analyze the performance of large-scale cell-free massive MIMO.
- We carry out extensive numerical experiments to draw useful insights into the performance of large-scale cell-free massive MIMO regarding the four utility metrics above. In particular we find that, in the domain of large-scale cell-free massive MIMO, per-user rates are quite comparable for the four above utility functions, which means that large-scale cell-free massive MIMO

[1]Though our considered cell-free massive MIMO system follows a similar transmission protocol as the one in colocated massive MIMO, there are many differences between these two systems which have been well explained in for example [11] such as the network topology and channel state information (CSI) exchange. Most importantly, the large-scale fading coefficients from a given user to different service antennas (APs for cell-free systems or BS antennas for colocated systems) are different in cell-free massive MIMO, but are reasonably assumed to be the same in colocated massive MIMO. As a result, the downlink achievable rates of these systems have totally different forms, and hence, the power control methods extensively studied in colocated massive MIMO cannot be applied to our considered cell-free massive MIMO systems.

can deliver universally good services to all users. Also, in terms of per-user rate, it is more beneficial to use a higher number of APs with a fewer antenna per AP than to use a smaller number of APs with more antennas per APs.

Notations: Bold lower and upper case letters represent vectors and matrices. $\mathcal{CN}(0,a)$ denotes a complex Gaussian random variable with zero mean and variance a. $\mathbf{X}^*$, $\mathbf{X}^{\mathrm{T}}$ and $\mathbf{X}^\dagger$ stand for the conjugate, transpose and Hermitian of $\mathbf{X}$, respectively. x_i is the i-th entry of vector $\mathbf{x}$; $[\mathbf{X}]_{i,j}$ is the (i,j)-th entry of $\mathbf{X}$. $\nabla f(\mathbf{x})$ represents the gradient of $f(\mathbf{x})$ and $\frac{\partial}{\partial \mathbf{x}_i} f(\mathbf{x})$ is the partial gradient with respect to $\mathbf{x}_i$. $\langle \mathbf{x}, \mathbf{y}\rangle \triangleq \mathbf{x}^{\mathrm{T}}\mathbf{y}$ is the inner product of vectors $\mathbf{x}$ and $\mathbf{y}$. $[\mathbf{x}]_+$ denotes the projector onto the positive orthant, i.e., $[\mathbf{x}]_+ = [\max(x_1,0);\max(x_2,0);\ldots;\max(x_n,0)]$. In other words, projecting $\mathbf{x}$ onto the positive orthant means setting the negative entries of $\mathbf{x}$ to zero and keep the remaining entries the same. $||\cdot||$ represents the Euclidean norm; $|\cdot|$ is the absolute value of the argument.

II. System Model and Problem Formulation

A. System Model

We consider the downlink of a cell-free massive MIMO system model as in [16]. In particular, there are M APs serving K single-antenna users in time division duplex (TDD) mode. Each AP is equipped with N antennas. All the APs and the users are assumed to be distributed in a large area. The APs cooperate through one or several central processing units (CPUs) to deal with resource allocations (i.e. power control, user scheduling or AP selections) on the large-scale fading time scale [16], [25]. However, each AP works independently to cope with small-scale fading. More specifically, each AP estimates the channels and applies a beamforming scheme locally. There is no need to exchange the small-scale fading coefficients between the APs, and between the APs and the CPUs. As TDD operation is adopted, APs first estimate the channels using pilot sequences from the uplink (commonly known as uplink training) and then apply a beamforming technique to transmit signals to all users in the downlink, or use a matched filter technique to combine signals in the uplink. Since this work focuses on the downlink transmission, we neglect the uplink payload transmission phase. Let us denote by T_c and T_p the length of the coherence interval and the uplink training phase in data symbols, respectively. The uplink training and downlink payload transmission phases are summarized as follows. The interested reader is referred to [16] for further details.

1) Uplink Training: We assume the channel is reciprocal, i.e., the channel gains on the uplink and on the downlink are the same. Consequently, APs can estimate the downlink channel based on the pilot sequences sent by all users on the uplink. Let $\sqrt{T_p}\boldsymbol{\psi}_k \in \mathbb{C}^{T_p\times 1}$, where $||\boldsymbol{\psi}_k||^2 = 1$, be the pilot sequence transmitted from the k-th user, $k = 1,\ldots,K$. Note that T_p is the length of the pilot sequences, which is the same for all users. The received signal at the m-th AP is given by

$$\mathbf{R}_{\text{up},m} = \sqrt{\zeta_p T_p}\sum_{k=1}^{K}\mathbf{g}_{mk}\boldsymbol{\psi}_k^\dagger + \mathbf{W}_{\text{up},m}, \tag{1}$$

where ζ_p is the normalized transmit signal-to-noise ratio (SNR) of each pilot symbol, and $\mathbf{W}_{\text{up},m} \in \mathbb{C}^{N\times T_p}$ is the noise matrix whose entries are independent and identically distributed (i.i.d.) drawn from $\mathcal{CN}(0,1)$, and $\mathbf{g}_{mk} \in \mathbb{C}^{N\times 1}$ is the channel between the m-th AP and the k-th user. As in [16], we model $\mathbf{g}_{mk}$ as

$$\mathbf{g}_{mk} = \beta_{mk}^{1/2}\mathbf{h}_{mk}, \tag{2}$$

where β_{mk} represents the large-scale fading (i.e. including path loss and shadowing effects) and $\mathbf{h}_{mk} \in \mathbb{C}^{N\times 1}$ comprises of small-scale fading coefficients between the N antennas of the m-th AP and the k-th user. We further assume that the entries of $\mathbf{h}_{mk}$ follows i.i.d. $\mathcal{CN}(0,1)$.

Next the m-th AP needs to estimate the channel $\mathbf{g}_{mk}$, $k = 1,2,\ldots,N$, based on the received pilot signal $\mathbf{R}_{\text{up},m}$. To do so, the m-th AP projects $\mathbf{R}_{\text{up},m}$ onto $\boldsymbol{\psi}_k$, producing

$$\mathbf{r}_{mk} = \mathbf{R}_{\text{up},m}\boldsymbol{\psi}_k = \sqrt{\zeta_p T_p}\sum_{i=1}^{K}\mathbf{g}_{mi}\boldsymbol{\psi}_i^\dagger\boldsymbol{\psi}_k + \tilde{\mathbf{w}}_{mk}, \tag{3}$$

where $\tilde{\mathbf{w}}_{mk} \triangleq \mathbf{W}_{\text{up},m}\boldsymbol{\psi}_k \in \mathbb{C}^{N\times 1}$ has entries following i.i.d. $\mathcal{CN}(0,1)$. We remark that the pilot sequences are not necessarily to be mutually orthogonal. Thus, from an algorithmic point of view, our proposed solutions to be presented in following section are applicable to both orthogonal and non-orthogonal pilot sequences. However, when possible, it is desirable to achieve orthogonal pilot sequences to mitigate the effect of pilot contamination, thereby increasing the achievable data rates.

Given $\mathbf{r}_{mk}$, the minimum mean-square error (MMSE) of the channel estimate of $\mathbf{g}_{mk}$ is [16]

$$\begin{aligned}\hat{\mathbf{g}}_{mk} &= \mathbb{E}\{\mathbf{g}_{mk}\mathbf{r}_{mk}^\dagger\}\Big(\mathbb{E}\{\mathbf{r}_{mk}\mathbf{r}_{mk}^\dagger\}\Big)^{-1}\mathbf{r}_{mk}\\ &= \frac{\sqrt{\zeta_p T_p}\beta_{mk}}{1+\zeta_p T_p\sum_{i=1}^{K}\beta_{mi}\left|\boldsymbol{\psi}_i^\dagger\boldsymbol{\psi}_k\right|^2}\mathbf{r}_{mk}.\end{aligned} \tag{4}$$

Note that the expectations in the above equation are carried out with respect to small-scale fading and also that elements of $\hat{\mathbf{g}}_{mk}$ are independent and identical Gaussian distribution. The mean square of any element of $\hat{\mathbf{g}}_{mk}$ is given by

$$\nu_{mk} = \frac{\zeta_p T_p\beta_{mk}^2}{1+\zeta_p T_p\sum_{i=1}^{K}\beta_{mi}\left|\boldsymbol{\psi}_i^\dagger\boldsymbol{\psi}_k\right|^2}. \tag{5}$$

2) Downlink Payload Data Transmission: For downlink payload data transmission, the APs use the channel estimates obtained in (4) to form separate radio beams to the K users. As mentioned earlier we adopt conjugate beamforming in this paper, which is due to two main reasons. First, conjugate beamforming is computationally simple and can be done locally at each AP. Second, conjugate beamforming offers excellent performance for a large number of APs (relatively compared to the number of users). Denote the symbol to be sent to the k-th user by c_k and the power control coefficient between the m-th AP and the k-th user by η_{mk}. For conjugate beamforming, the transmitted symbols from the m-th AP are

contained in the vector $\mathbf{x}_m$ given by

$$\mathbf{x}_m = \sqrt{\zeta_d} \sum_{k=1}^{K} \sqrt{\eta_{mk}} \hat{\mathbf{g}}_{mk}^* c_k, \tag{6}$$

where ζ_d is the maximum downlink transmit power at each AP normalized to noise power. Note that the total power at each AP is

$$\mathbb{E}\{||\mathbf{x}_m||^2\} = \zeta_d N \sum_{k=1}^{K} \eta_{mk} \nu_{mk}. \tag{7}$$

The received signal at the k-th user is written as

$$r_k = \sum_{m=1}^{M} \mathbf{g}_{mk}^{\mathrm{T}} \mathbf{x}_m + w_k = \sqrt{\zeta_d} \sum_{m=1}^{M} \sqrt{\eta_{mk}} \mathbf{g}_{mk}^{\mathrm{T}} \hat{\mathbf{g}}_{mk}^* c_k + \sqrt{\zeta_d} \sum_{i \neq k}^{K} \sum_{m=1}^{M} \sqrt{\eta_{mi}} \mathbf{g}_{mk}^{\mathrm{T}} \hat{\mathbf{g}}_{mi}^* c_i + w_k, \tag{8}$$

where w_k is the white Gaussian noise with zero mean and unit variance.

3) Signal Detection Based on Channel Statistics and Spectral Efficiency: Ideally, to detect c_k, the k-th user needs to know the effective channel gain $\sqrt{\zeta_d} \sum_{m=1}^{M} \sqrt{\eta_{mk}} \mathbf{g}_{mk}^{\mathrm{T}} \hat{\mathbf{g}}_{mk}^*$. However this is impossible since there are no downlink pilots. Instead, the k-th user will rely on the mean of the effective channel gain to detect c_k. To see this we rewrite (8) as

$$\begin{aligned} r_k = & \sqrt{\zeta_d} \mathbb{E}\left\{ \sum_{m=1}^{M} \sqrt{\eta_{mk}} \mathbf{g}_{mk}^{\mathrm{T}} \hat{\mathbf{g}}_{mk}^* \right\} c_k \\ & + \sqrt{\zeta_d} \left(\sum_{m=1}^{M} \sqrt{\eta_{mk}} \mathbf{g}_{mk}^{\mathrm{T}} \hat{\mathbf{g}}_{mk}^* - \mathbb{E}\left\{ \sum_{m=1}^{M} \sqrt{\eta_{mk}} \mathbf{g}_{mk}^{\mathrm{T}} \hat{\mathbf{g}}_{mk}^* \right\} \right) c_k \\ & + \sum_{i \neq k}^{K} \sqrt{\zeta_d} \sum_{m=1}^{M} \sqrt{\eta_{mi}} \mathbf{g}_{mk}^{\mathrm{T}} \hat{\mathbf{g}}_{mi}^* c_i + w_k. \end{aligned} \tag{9}$$

In the above equation the second term in the right side can be seen as the beamforming uncertainty, which is due to treating the mean of the effective channel gain as the true channel. We remark that by the law of large numbers which holds for our system model, with high probability, this term is much smaller compared to the mean of the effective channel gain. By further treating this and the inter-user interference as the Gaussian noise, we can express the SINR at the k-th user as [16, Appendix A]

$$\gamma_k(\bar{\boldsymbol{\eta}}) = \frac{\zeta_d N^2 \left| \boldsymbol{\nu}_{kk}^{\mathrm{T}} \bar{\boldsymbol{\eta}}_k \right|^2}{\zeta_d N^2 \sum_{i \neq k}^{K} \left| \boldsymbol{\nu}_{ik}^{\mathrm{T}} \bar{\boldsymbol{\eta}}_i \right|^2 + \zeta_d N \sum_{i=1}^{K} ||\mathbf{D}_{ik} \bar{\boldsymbol{\eta}}_i||_2^2 + 1}, \tag{10}$$

where $\bar{\boldsymbol{\eta}}_k = [\sqrt{\eta_{1k}}; \ldots; \sqrt{\eta_{Mk}}] \in \mathbb{R}_+^M$ consists of all power control coefficients associated with user k, $\bar{\boldsymbol{\eta}} = [\bar{\boldsymbol{\eta}}_1; \bar{\boldsymbol{\eta}}_2; \ldots; \bar{\boldsymbol{\eta}}_K] \in \mathbb{R}_+^{MK}$, $\mathbf{D}_{ik} \in \mathbb{R}_+^{M \times M}$ is a diagonal matrix with $[\mathbf{D}_{ik}]_{m,m} = \sqrt{\nu_{mi} \beta_{mk}}$, and

$$\boldsymbol{\nu}_{ik} \triangleq \left| \boldsymbol{\psi}_i^{\dagger} \boldsymbol{\psi}_k \right| \left[\nu_{1i} \frac{\beta_{1k}}{\beta_{1i}}; \nu_{2i} \frac{\beta_{2k}}{\beta_{2i}}; \ldots; \nu_{Mi} \frac{\beta_{Mk}}{\beta_{Mi}} \right]. \tag{11}$$

Accordingly, the spectral efficiency of the k-th user is given by

$$\mathrm{SE}_k(\bar{\boldsymbol{\eta}}) = \left(1 - \frac{T_p}{T_c}\right) \log\left(1 + \gamma_k(\bar{\boldsymbol{\eta}})\right) \quad \text{(nat/s/Hz)}. \tag{12}$$

Note that for mathematical convenience we use the natural logarithm in (12), and thus the resulting unit of the SE is nat/s/Hz. However, for the numerical results presented in Section V, we instead use the logarithm to base 2 to compute the SE and the corresponding unit is bit/s/Hz.

B. Problem Formulation

To formulate the considered problem and to facilitate the development of the proposed algorithm, we define $\boldsymbol{\mu}_m \in \mathbb{R}_+^K$ to be the vector of all power control coefficients associated with the m-th AP as

$$\boldsymbol{\mu}_m \triangleq [\mu_{m1}; \mu_{m2}; \ldots; \mu_{mK}], \tag{13}$$

where $\mu_{mk} = \sqrt{\eta_{mk} \nu_{mk}}, m = 1, \ldots, M, k = 1, \ldots, K$. We also define $\boldsymbol{\mu} \triangleq [\boldsymbol{\mu}_1; \boldsymbol{\mu}_2; \ldots; \boldsymbol{\mu}_M] \in \mathbb{R}_+^{MK \times 1}$ to include the power control coefficients of all APs. To express the spectral efficiency in (12) as a function of $\boldsymbol{\mu}$, we denote by $\bar{\boldsymbol{\mu}}_k = [\mu_{1k}; \mu_{2k}; \ldots; \mu_{Mk}]$ the vector of power control coefficients associated with user k. Thus we can write $\boldsymbol{\nu}_{ik}^{\mathrm{T}} \boldsymbol{\eta}_i$ as $\bar{\boldsymbol{\nu}}_{ik} \bar{\boldsymbol{\mu}}_k$, where

$$\bar{\boldsymbol{\nu}}_{ik} \triangleq \left| \boldsymbol{\psi}_i^{\dagger} \boldsymbol{\psi}_k \right| \left[\sqrt{\nu_{1i}} \frac{\beta_{1k}}{\beta_{1i}}; \sqrt{\nu_{2i}} \frac{\beta_{2k}}{\beta_{2i}}; \ldots; \sqrt{\nu_{Mi}} \frac{\beta_{Mk}}{\beta_{Mi}} \right]. \tag{14}$$

Similarly, we can write $\mathbf{D}_{ik} \bar{\boldsymbol{\eta}}_i$ as $\bar{\mathbf{D}}_i \bar{\boldsymbol{\mu}}_i$, where $\bar{\mathbf{D}}_i$ is the diagonal matrix with the m-th diagonal entry equal to $\sqrt{\beta_{mi}}$. Now the spectral efficiency of the k-th user (in nat/s/Hz) can be expressed as

$$\mathrm{SE}_k(\boldsymbol{\mu}) = \left(1 - \frac{T_p}{T_c}\right) \log\left(1 + \gamma_k(\boldsymbol{\mu})\right), \tag{15}$$

where $\gamma_k(\boldsymbol{\mu})$ is the SINR of the k-th user given by

$$\gamma_k(\boldsymbol{\mu}) = \frac{\zeta_d (\bar{\boldsymbol{\nu}}_{kk}^{\mathrm{T}} \bar{\boldsymbol{\mu}}_k)^2}{\zeta_d \left(\sum_{i \neq k}^{K} (\bar{\boldsymbol{\nu}}_{ik}^{\mathrm{T}} \bar{\boldsymbol{\mu}}_i)^2 + \frac{1}{N} \sum_{i=1}^{K} \left\| \bar{\mathbf{D}}_i \bar{\boldsymbol{\mu}}_i \right\|_2^2 \right) + \frac{1}{N^2}}. \tag{16}$$

The total spectral efficiency of the system is defined as

$$\mathrm{SE}(\boldsymbol{\mu}) \triangleq \sum_{k=1}^{K} \mathrm{SE}_k(\boldsymbol{\mu}). \tag{17}$$

In this paper, we consider a power constraint at each AP which is given by $||\boldsymbol{\mu}_m||^2 \leq \frac{1}{N}, m = 1, 2, \ldots, M$, which follows from (7). For the problem formulation purpose, we define the following set

$$\mathcal{S} = \left\{ \boldsymbol{\mu} | \boldsymbol{\mu} \geq 0; \left\| \boldsymbol{\mu}_m \right\|^2 \leq \frac{1}{N}, m = 1, 2, \ldots, M \right\}, \tag{18}$$

which is nothing but the feasible set of the utility maximization problems to be presented. In this paper, we consider the following four common power control utility optimization problems [26], namely

- The problem of average spectral efficiency maximization (SEmax)

$$\boxed{(\mathcal{P}_1) : \underset{\boldsymbol{\mu}}{\text{maximize}} \left\{ (1/K) \sum_{k=1}^{K} \mathrm{SE}_k(\boldsymbol{\mu}) \mid \boldsymbol{\mu} \in \mathcal{S} \right\}.}$$

- The problem of proportional fairness maximization (PFmax)

$$\boxed{(\mathcal{P}_2): \underset{\boldsymbol{\mu}}{\text{maximize}} \left\{\sum\nolimits_{k=1}^{K} \log \mathrm{SE}_k(\boldsymbol{\mu}) \mid \boldsymbol{\mu} \in \mathcal{S}\right\}.}$$

Note that the above problem is equivalent to maximizing $\left(\prod_{k=1}^{K} \mathrm{SE}_k(\boldsymbol{\mu})\right)^{1/K}$ over the set $\mathcal{S}$. Thus it is also known as the problem of geometric-rate maximization.

- The problem of the harmonic rate maximization (HRmax)

$$\boxed{(\mathcal{P}_3): \underset{\boldsymbol{\mu}}{\text{maximize}} \left\{K\Big(\sum_{k=1}^{K} \mathrm{SE}_k(\boldsymbol{\mu})^{-1}\Big)^{-1} \;\Big|\; \boldsymbol{\mu} \in \mathcal{S}\right\}.}$$

- The problem of maximizing the minimum rate (MRmax) among all users (also known as max-min fairness maximization)

$$\boxed{(\mathcal{P}_4): \underset{\boldsymbol{\mu}}{\text{maximize}} \left\{\min_{1\le k\le K} \mathrm{SE}_k(\boldsymbol{\mu}) \mid \boldsymbol{\mu} \in \mathcal{S}\right\}.}$$

We note that the above four problems are noncovex and thus difficult to solve. For such problems, a pragmatic goal is to derive a low complexity high-performance solution, rather than a globally optimal solution. To this end, SCA has proved to be very effective and gradually become a standard mathematical tool [11], [16]. The idea of SCA is to approximate a non-convex program by a series of convex sub-problems. In all known solutions for the considered problems or related ones, interior point methods (through the use of off-the-shelf convex solvers) are invoked to solve these convex problems [15]–[17], which do not scale favorably with the problem size. Thus the existing solutions are unable to characterize the performance limits of cell-free massive MIMO systems where the number of APs can be in the order of thousands, even from an off-line design perspective. In this paper, we propose methods that can tackle this scalability problem. In particular, our proposed methods are based on first order optimization methods which are presented in the following section.

III. Proposed Solutions

In this section, we present the proposed APG method to solve $(\mathcal{P}_1)$ to $(\mathcal{P}_4)$, which is a variant of the accelerated proximal gradient method (AProxG) introduced in [23]. In general the four considered problems can be written in a compact form as

$$\underset{\boldsymbol{\mu}}{\text{maximize}}\ f(\boldsymbol{\mu}) \tag{19a}$$

$$\text{subject to}\ \ \boldsymbol{\mu} \in \mathcal{S}, \tag{19b}$$

where $f(\boldsymbol{\mu}) = (1/K)\sum_{k=1}^{K} \mathrm{SE}_k(\boldsymbol{\mu})$ for problem $(\mathcal{P}_1)$, $f(\boldsymbol{\mu}) = \sum_{k=1}^{K} \log \mathrm{SE}_k(\boldsymbol{\mu})$ for problem $(\mathcal{P}_2)$, $f(\boldsymbol{\mu}) = K\big(\sum_{k=1}^{K} \mathrm{SE}_k(\boldsymbol{\mu})^{-1}\big)^{-1}$ for $(\mathcal{P}_3)$, and $f(\boldsymbol{\mu}) = \min_{1\le k\le K} \mathrm{SE}_k(\boldsymbol{\mu})$ for $(\mathcal{P}_4)$. We remark that the method described in [23] concerns the following problem

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}}\ \{F(\mathbf{x}) \equiv f(\mathbf{x}) + g(\mathbf{x})\}, \tag{20}$$

where $f(\mathbf{x})$ is differentiable (but possibly *nonconvex*) and $g(\mathbf{x})$ can be both nonconvex and *nonsmooth*. Further assumptions on $f(\mathbf{x})$ are listed below:

- A1: $f(\mathbf{x})$ is a proper function with Lipschitz continuous gradients. A function f is said to have an L-Lipschitz continuous gradient if there exists some $L > 0$ such that

$$||\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})|| \le L||\mathbf{x} - \mathbf{y}||, \quad \forall \mathbf{x}, \mathbf{y}. \tag{21}$$

- A2: $f(\mathbf{x})$ is coercive i.e. $f(\mathbf{x})$ is bounded from below and $f(\mathbf{x}) \to \infty$ when $\mathbf{x} \to \infty$.

To solve the general nonconvex problem in (20), Li *et al.* proposed the monotone AProxG [23, Algorithm 1]. In essence, the AProxG method moves the current point, say $\mathbf{y}^n$, along the gradient of the objective $\nabla f(\mathbf{y}^n)$ with a proper step size (denoted by α). At the resulting point, the AProxG method finds the proximal operator of g(cf. [23, Eqn. (10)]). This step is repeated until some stopping criterion is met. More specifically, the proximal operator of g at a given $\mathbf{x}$ is defined as

$$\mathrm{prox}_{\alpha g}(\mathbf{x}) = \mathrm{argmin}_{\mathbf{u}}\ \ g(\mathbf{u}) + \frac{1}{2\alpha}\,\|\mathbf{x}\|^2\,.$$

To customize the AProxG to solve the considered problems, we let $g(\mathbf{x})$ be the indicator function of $\mathcal{S}$, defined as

$$\delta_{\mathcal{S}}(\mathbf{x}) = \begin{cases} 0 & \mathbf{x} \in \mathcal{S} \\ +\infty & \mathbf{x} \notin \mathcal{S}, \end{cases} \tag{22}$$

then (20) is actually equivalent to (19). In this way, the proximal operator of $\delta_{\mathcal{S}}(\mathbf{x})$ becomes the Euclidean projection onto $\mathcal{S}$. Specifically, the Euclidean projection of $\mathbf{u}$ onto $\mathcal{S}$, denoted by $P_{\mathcal{S}}(\mathbf{u})$, is defined as $P_{\mathcal{S}}(\mathbf{u}) = \arg\min\Big\{||\mathbf{x}-\mathbf{u}|| \mid \mathbf{x} \in \mathcal{S}\Big\}$. We are now in a position to apply the AProxG method to solve the considered problems. The details are given in the subsequent subsections.

Remark 1: One may ask why we have made a change of variables from $\bar{\boldsymbol{\eta}}$ to $\boldsymbol{\mu}$ in Section II-B. The question is relevant since the projection onto the feasible set expressed in terms of $\bar{\boldsymbol{\eta}}$ can also be done efficiently. However, if the objective function of the four considered problems is written as a function of $\bar{\boldsymbol{\eta}}$, then there are two difficulties arising. Firstly, the expression of the gradient of the objective becomes very complicated. Secondly and more importantly, the gradient of the objective is not Lipschitz continuous since the term $\sqrt{\eta_{mk}}$ would appear in the denominator of the gradient. Note that η_{mk} can be zero, which can make the gradient unbounded.

A. Proposed Solution to $(\mathcal{P}_1)$

Since $f(\boldsymbol{\mu})$ for $(\mathcal{P}_1)$ is differentiable, the proposed algorithm for solving $(\mathcal{P}_1)$ follows closely the monotone AProxG method in [23], which is outlined in Algorithm 1. In Algorithm 1, $\alpha > 0$ is called the step size which should be sufficiently small to guarantee its convergence, L_f is the Lipschitz constant of $\nabla f(\boldsymbol{\mu})$, and t_n is an extrapolation parameter recursively defined in Line 7, which determines how the current and the past iterates are linearly combined to obtain the next iterate to improve the convergence. The superscripts in Algorithm 1 denote the iteration count. Note that we adapt the monotone AProxG method in [23] for minimization to the context of maximization for our problems.

An intuitive description of the proposed APG method is as follows. From the current iterate denoted by $\boldsymbol{\mu}^n$, we compute the extrapolated point $\mathbf{y}^n$ as in Line 3 of Algorithm 1, which is used for convergence acceleration. From $\mathbf{y}^n$, we move along its the gradient with the step size α, and then project the resulting point (i.e., $\mathbf{y}^n + \alpha\nabla f(\mathbf{y}^n)$) onto the feasible set to obtain $\mathbf{z}^{n+1}$. However, unlike conventional APG methods for convex programming, $\mathbf{y}^n$ *can be a bad extrapolation* in the sense that $f(\mathbf{z}^{n+1})$ does not improve the objective sequence. To prevent this, a monitor and correction process is proposed in [23]. Specifically, we also calculate an extra point $\mathbf{v}^{n+1}$ as in Line 5 of Algorithm 1. The next iterate $\boldsymbol{\mu}^{n+1}$ is obtained by comparing the objective values of $\mathbf{v}^{n+1}$ and $\mathbf{z}^{n+1}$ as done in Line 6.

Algorithm 1 Accelerated Projected Gradient Algorithm for Solving $(\mathcal{P}_1)$-$(\mathcal{P}_4)$

1: Input: $\boldsymbol{\mu}^0 >= 0, t_0 = 0, t_1 = 1, \frac{1}{L_f} > \alpha > 0, \boldsymbol{\mu}^1 = \mathbf{z}^1 = \boldsymbol{\mu}^0$
2: **for** $n = 1, 2, \ldots$ **do**
3: $\quad \mathbf{y}^n = \boldsymbol{\mu}^n + \frac{t_{n-1}}{t_n}(\mathbf{z}^n - \boldsymbol{\mu}^n) + \frac{t_{n-1}-1}{t_n}(\boldsymbol{\mu}^n - \boldsymbol{\mu}^{n-1})$
4: $\quad \mathbf{z}^{n+1} = P_{\mathcal{S}}(\mathbf{y}^n + \alpha\nabla f(\mathbf{y}^n))$
5: $\quad \mathbf{v}^{n+1} = P_{\mathcal{S}}(\boldsymbol{\mu}^n + \alpha\nabla f(\boldsymbol{\mu}^n))$
6: $\quad \boldsymbol{\mu}^{n+1} = \begin{cases} \mathbf{z}^{n+1} & f(\mathbf{z}^{n+1}) \geq f(\mathbf{v}^{n+1}) \\ \mathbf{v}^{n+1} & \text{otherwise} \end{cases}$
7: $\quad t_{n+1} = 0.5\left(\sqrt{4t_n^2 + 1} + 1\right)$
8: **end for**
9: Output: $\boldsymbol{\mu}^n$

We now give the details for the two main operations of Algorithm 1, namely: the projection onto the feasible set $\mathcal{S}$ and the gradient of $f(\boldsymbol{\mu})$.

1) Projection Onto $\mathcal{S}$: We show that the projection in Steps 4 and 5 in Algorithm 1 can be done *in parallel* and by *closed-form expressions*. Recall that for given a $\mathbf{x} \in \mathbb{R}^{MK\times 1}$, $P_{\mathcal{S}}(\mathbf{x})$ is the solution to the following problem

$$\underset{\boldsymbol{\mu}\in\mathbb{R}^{MK\times 1}}{\text{minimize}} \left\{ ||\boldsymbol{\mu} - \mathbf{x}||^2 \;\middle|\; \boldsymbol{\mu} \geq 0; ||\boldsymbol{\mu}_m||^2 \leq \frac{1}{N}, \quad m = 1, 2, \ldots, M \right\}. \tag{23}$$

It is easy to see that the above problem can be decomposed into sub-problems at each AP m as

$$\underset{\boldsymbol{\mu}_m\in\mathbb{R}^{K\times 1}}{\text{minimize}} \left\{ ||\boldsymbol{\mu}_m - \mathbf{x}_m||^2 \;\middle|\; \boldsymbol{\mu}_m \geq 0; ||\boldsymbol{\mu}_m||^2 \leq \frac{1}{N} \right\}. \tag{24}$$

The above problem is in fact the projection onto the intersection of a ball and the positive orthant. Interestingly, the analytical solution to this problem can be found by applying [27, Theorem 7.1], which produces

$$\boldsymbol{\mu}_m = \frac{\sqrt{1/N}}{\max\left(\sqrt{1/N}, ||[\mathbf{x}_m]_+||\right)}[\mathbf{x}_m]_+. \tag{25}$$

The above expression means that we first project $\mathbf{x}_m$ onto the positive orthant and then onto the Euclidean ball of radius $\sqrt{1/N}$. A simpler way to prove (25) is detailed in Appendix A.

2) Gradient of $f(\boldsymbol{\mu})$ for $(\mathcal{P}_1)$: To implement Algorithm 1, we also need to compute $\nabla_{\boldsymbol{\mu}} f(\boldsymbol{\mu})$, which is derived in what follows. We know that the gradient of a multi-variable function is the vector of all its partial derivatives, i.e.

$$\nabla f(\boldsymbol{\mu}) = \left[\frac{\partial}{\partial\bar{\boldsymbol{\mu}}_1} f(\boldsymbol{\mu}); \frac{\partial}{\partial\bar{\boldsymbol{\mu}}_2} f(\boldsymbol{\mu}), \ldots, \frac{\partial}{\partial\bar{\boldsymbol{\mu}}_K} f(\boldsymbol{\mu})\right], \tag{26}$$

where $\frac{\partial}{\partial\bar{\boldsymbol{\mu}}_i} f(\boldsymbol{\mu}) = (1/K)\sum_{k=1}^K \frac{\partial}{\partial\bar{\boldsymbol{\mu}}_i}\mathrm{SE}_k(\boldsymbol{\mu})$. Thus, it basically boils down to finding $\frac{\partial}{\partial\bar{\boldsymbol{\mu}}_i}\mathrm{SE}_k(\boldsymbol{\mu})$. Let us define $b_k(\boldsymbol{\mu}) = \zeta_d(\bar{\boldsymbol{\nu}}_{kk}^{\mathrm{T}}\bar{\boldsymbol{\mu}}_k)^2$ and $c_k(\boldsymbol{\mu}) = \zeta_d\left(\sum_{i\neq k}^K(\bar{\boldsymbol{\nu}}_{ik}^{\mathrm{T}}\bar{\boldsymbol{\mu}}_i)^2 + \frac{1}{N}\sum_{i=1}^K ||\bar{\mathbf{D}}_i\bar{\boldsymbol{\mu}}_i||_2^2\right) + \frac{1}{N^2}$. Then we can rewrite $\mathrm{SE}_k(\boldsymbol{\mu})$ as

$$\mathrm{SE}_k(\boldsymbol{\mu}) = \log\big(b_k(\boldsymbol{\mu}) + c_k(\boldsymbol{\mu})\big) - \log c_k(\boldsymbol{\mu}). \tag{27}$$

The gradient of $\mathrm{SE}_k(\boldsymbol{\mu})$ with respect to $\bar{\boldsymbol{\mu}}_i$, $i = 1, 2, \ldots, K$, is found as

$$\frac{\partial}{\partial\bar{\boldsymbol{\mu}}_i}\mathrm{SE}_k(\boldsymbol{\mu}) = \frac{\frac{\partial}{\partial\bar{\boldsymbol{\mu}}_i}\left(b_k(\boldsymbol{\mu}) + c_k(\boldsymbol{\mu})\right)}{b_k(\boldsymbol{\mu}) + c_k(\boldsymbol{\mu})} - \frac{\frac{\partial}{\partial\bar{\boldsymbol{\mu}}_i}c_k(\boldsymbol{\mu})}{c_k(\boldsymbol{\mu})} \tag{28}$$

Now we recall the following identity $\nabla||\mathbf{A}\mathbf{x}||^2 = 2\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{x}$ for any symmetric matrix $\mathbf{A}$, and thus $\nabla_{\bar{\boldsymbol{\mu}}_i} b_k(\boldsymbol{\mu})$ and $\nabla_{\bar{\boldsymbol{\mu}}_i} c_k(\boldsymbol{\mu})$ are respectively given by

$$\frac{\partial}{\partial\bar{\boldsymbol{\mu}}_i} b_k(\boldsymbol{\mu}) = \begin{cases} 2\zeta_d\bar{\boldsymbol{\nu}}_{kk}\bar{\boldsymbol{\nu}}_{kk}^{\mathrm{T}}\bar{\boldsymbol{\mu}}_k, & i = k \\ 0, & i \neq k \end{cases} \tag{29}$$

$$\frac{\partial}{\partial\bar{\boldsymbol{\mu}}_i} c_k(\boldsymbol{\mu}) = \begin{cases} 2(\zeta_d/N)\bar{\mathbf{D}}_k^2\bar{\boldsymbol{\mu}}_k, & i = k \\ 2\zeta_d\bar{\boldsymbol{\nu}}_{ik}\bar{\boldsymbol{\nu}}_{ik}^{\mathrm{T}}\bar{\boldsymbol{\mu}}_i + \dfrac{2\zeta_d}{N}\bar{\mathbf{D}}_i^2\bar{\boldsymbol{\mu}}_i, & i \neq k. \end{cases} \tag{30}$$

B. Improved Convergence With Line Search

For $(\mathcal{P}_1)$, from (28), (29), and (30), it easy to check that $\nabla f(\boldsymbol{\mu})$ is Lipschitz continuous, or equivalently $f(\boldsymbol{\mu})$ has Lipschitz continuous gradient. That is, there exists a constant $L_f > 0$ such that

$$||\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})|| \leq L_f||\mathbf{x} - \mathbf{y}|| \quad \forall \mathbf{x}, \mathbf{y} \in \mathcal{S}. \tag{31}$$

Further details are given in Appendix B.

In practice, we in fact do not need to compute a Lipschitz constant of $\nabla f(\boldsymbol{\mu})$ for two reasons. First, the best Lipschitz constant of $\nabla f(\boldsymbol{\mu})$ (i.e. the smallest L such that (31) holds) is hard to find. Second, the conditions $\alpha < \frac{1}{L_f}$ is sufficient but not necessary for Algorithm 1 to converge. Thus, we can allow α to take on larger values to speed up the convergence of Algorithm 1 by means of a linear search procedure. In this paper we use line search with the Barzilai-Borwein (BB) rule to compute step sizes for Algorithm 1. The APG method with line search backtracking line search is summarized in Algorithm 2. The step sizes α_y and α_μ computed in Steps 5 and 7 can be viewed as local estimate of the optimal Lipschitz constant of the gradient at $\mathbf{y}^{n-1}$ and $\boldsymbol{\mu}^{n-1}$, respectively.

Algorithm 2 APG Method With Line Search for Solving $(\mathcal{P}_1)$-$(\mathcal{P}_4)$

Input: $\boldsymbol{\mu}^0 >= 0, t_0 = 0, t_1 = 1, \alpha_\mu > 0, \alpha_y > 0, \delta > 0, \rho < 1, \boldsymbol{\mu}^1 = \mathbf{z}^1 = \boldsymbol{\mu}^0$
for $n = 1, 2, \ldots$ **do**
  $\mathbf{y}^n = \boldsymbol{\mu}^n + \frac{t_{n-1}}{t_n}(\mathbf{z}^n - \boldsymbol{\mu}^n) + \frac{t_{n-1}-1}{t_n}(\boldsymbol{\mu}^n - \boldsymbol{\mu}^{n-1})$
  $\mathbf{s}^n = \mathbf{z}^n - \mathbf{y}^{n-1}, \mathbf{r}^n = \nabla f(\mathbf{z}^n) - \nabla f(\mathbf{y}^{n-1})$
  $\alpha_y = \langle \mathbf{s}^n, \mathbf{s}^n \rangle / \langle \mathbf{s}^n, \mathbf{r}^n \rangle$ or $\alpha_y = \langle \mathbf{s}^n, \mathbf{r}^n \rangle / \langle \mathbf{r}^n, \mathbf{r}^n \rangle$
  $\mathbf{s}^n = \mathbf{v}^n - \boldsymbol{\mu}^{n-1}, \mathbf{r}^n = \nabla f(\mathbf{v}^n) - \nabla f(\boldsymbol{\mu}^{n-1})$
  $\alpha_\mu = \langle \mathbf{s}^n, \mathbf{s}^n \rangle / \langle \mathbf{s}^n, \mathbf{r}^n \rangle$ or $\alpha_\mu = \langle \mathbf{s}^n, \mathbf{r}^n \rangle / \langle \mathbf{r}^n, \mathbf{r}^n \rangle$
  repeat
    $\mathbf{z}^{n+1} = P_{\mathcal{S}}(\mathbf{y}^n + \alpha_y \nabla f(\mathbf{y}^n))$
    $\alpha_y = \alpha_y \times \rho$
  until $F(\mathbf{z}^{n+1}) \geq F(\mathbf{y}^n) + \delta ||\mathbf{z}^{n+1} - \mathbf{y}^n||^2$
  repeat
    $\mathbf{v}^{n+1} = P_{\mathcal{S}}(\boldsymbol{\mu}^n + \alpha_\mu \nabla f(\boldsymbol{\mu}^n))$
    $\alpha_\mu = \alpha_\mu \times \rho$
  until $F(\mathbf{v}^{n+1}) \geq F(\boldsymbol{\mu}^n) + \delta ||\mathbf{v}^{n+1} - \boldsymbol{\mu}^n||^2$
  $\boldsymbol{\mu}^{n+1} = \begin{cases} \mathbf{z}^{n+1} & F(\mathbf{z}^{n+1}) \geq F(\mathbf{v}^{n+1}) \\ \mathbf{v}^{n+1} & \text{otherwise} \end{cases}$
  $t_{n+1} = 0.5\left(\sqrt{4t_n^2 + 1} + 1\right)$
end for
Output: $\boldsymbol{\mu}^n$

C. Customization for Solving $(\mathcal{P}_2)$ and $(\mathcal{P}_3)$

We remark that Algorithms 1 and 2 are unified in the sense that they can be easily modified to solve the remaining considered problems. In this subsection we explain how to apply Algorithms 1 and 2 to solve problems $(\mathcal{P}_2)$ and $(\mathcal{P}_3)$. First note that the objective functions in $(\mathcal{P}_2)$ and $(\mathcal{P}_3)$ are differentiable and the application of Algorithms 1 and 2 is straightforward. Specifically, for the PFmax problem (i.e. $(\mathcal{P}_2)$), the objective is

$$f(\boldsymbol{\mu}) = \sum\nolimits_{k=1}^{K} \log \mathrm{SE}_k(\boldsymbol{\mu}), \tag{32}$$

and thus $\nabla f(\boldsymbol{\mu})$ is given by (26), where $\nabla_{\bar{\boldsymbol{\mu}}_i} f(\boldsymbol{\mu})$ is found as

$$\frac{\partial}{\partial \bar{\boldsymbol{\mu}}_i} f(\boldsymbol{\mu}) = \sum_{k=1}^{K} \frac{1}{\mathrm{SE}_k(\boldsymbol{\mu})} \frac{\partial}{\partial \bar{\boldsymbol{\mu}}_i} \mathrm{SE}_k(\boldsymbol{\mu}), \tag{33}$$

where $\frac{\partial}{\partial \bar{\boldsymbol{\mu}}_i} \mathrm{SE}_k(\boldsymbol{\mu})$ is provided in (28). For the HRmax problem (i.e. $(\mathcal{P}_3)$), the objective is

$$f(\boldsymbol{\mu}) = K\Big(\sum_{k=1}^{K} \mathrm{SE}_k(\boldsymbol{\mu})^{-1}\Big)^{-1}. \tag{34}$$

The gradient of the function can be found similarly where $\nabla_{\bar{\boldsymbol{\mu}}_i} f(\boldsymbol{\mu})$ is written as

$$\frac{\partial}{\partial \bar{\boldsymbol{\mu}}_i} f(\boldsymbol{\mu}) = K\Big(\sum_{k=1}^{K} \mathrm{SE}_k(\boldsymbol{\mu})^{-1}\Big)^{-2} \times \sum_{k=1}^{K} \frac{1}{\big(\mathrm{SE}_k(\boldsymbol{\mu})\big)^2} \frac{\partial}{\partial \bar{\boldsymbol{\mu}}_i} \mathrm{SE}_k(\boldsymbol{\mu}). \tag{35}$$

where $\frac{\partial}{\partial \bar{\boldsymbol{\mu}}_i} \mathrm{SE}_k(\boldsymbol{\mu})$ is again provided in (28).

Remark 2: For $(\mathcal{P}_2)$ and $(\mathcal{P}_3)$, the utility functions are not Lipschitz continuous gradient in principle since the data rate $\mathrm{SE}_k(\boldsymbol{\mu})$ can be zero for some user k. Consequently, the gradient of the objective becomes unbounded due to the term $\frac{1}{\mathrm{SE}_k(\boldsymbol{\mu})}$ in (33) and (35). In practice, to fix this problem we simply add a fixed regularization parameter ϵ (say, $\epsilon = 10^{-6}$) and consider $f(\boldsymbol{\mu}) = \sum_{k=1}^{K} \log\big(\epsilon + \mathrm{SE}_k(\boldsymbol{\mu})\big)$ for $(\mathcal{P}_2)$ and $f(\boldsymbol{\mu}) = K\Big(\sum_{k=1}^{K}\big(\epsilon + \mathrm{SE}_k(\boldsymbol{\mu})\big)^{-1}\Big)^{-1}$ for $(\mathcal{P}_3)$. In this way $f(\boldsymbol{\mu})$ is Lipschitz continuous gradient.

D. Proposed Solution to $(\mathcal{P}_4)$

Problem $(\mathcal{P}_4)$ deserves further discussions since the objective is nonsmooth. We recall that for $(\mathcal{P}_4)$ the objective function is

$$f(\boldsymbol{\mu}) = \min_{1 \leq k \leq K} \mathrm{SE}_k(\boldsymbol{\mu}), \tag{36}$$

which is *non-differentiable*. Thus a straightforward application of the APG method is impossible. To overcome this issue, we adopt a smoothing technique. In particular, $f(\boldsymbol{\mu})$ is approximated by the following log-sum-exp function given by [28]

$$f_\tau(\boldsymbol{\mu}) = -\frac{1}{\tau} \log\big(\frac{1}{K} \sum\nolimits_{k=1}^{K} \exp\big(-\tau \mathrm{SE}_k(\boldsymbol{\mu})\big), \tag{37}$$

where $\tau > 0$ is the positive *smoothness* parameter. To obtain (37), we have used the fact that $\min_{1 \leq k \leq K} \mathrm{SE}_k(\boldsymbol{\mu}) = -\max_{1 \leq k \leq K}(-\mathrm{SE}_k(\boldsymbol{\mu}))$. It was proved in [28] that $f(\boldsymbol{\mu}) + \frac{\log K}{\tau} \geq f_\tau(\boldsymbol{\mu}) \geq f(\boldsymbol{\mu})$. In other words, $f_\tau(\boldsymbol{\mu})$ is a differentiable approximation of $f(\boldsymbol{\mu})$ with a numerical accuracy of $\frac{\log K}{\tau}$. Thus, with a sufficiently high τ, we can find an approximate solution to $(\mathcal{P}_4)$ by running Algorithm 1 with $f(\boldsymbol{\mu})$ being replaced by $f_\tau(\boldsymbol{\mu})$ in (37). In this regard, the gradient of $f_\tau(\boldsymbol{\mu})$ is easily found as $\nabla_{\bar{\boldsymbol{\mu}}} f_\tau(\boldsymbol{\mu}) = [\frac{\partial}{\partial \bar{\boldsymbol{\mu}}_1} f_\tau(\boldsymbol{\mu}), \frac{\partial}{\partial \bar{\boldsymbol{\mu}}_2} f_\tau(\boldsymbol{\mu}), \ldots, \frac{\partial}{\partial \bar{\boldsymbol{\mu}}_K} f_\tau(\boldsymbol{\mu})]$ where $\frac{\partial}{\partial \bar{\boldsymbol{\mu}}_i} f_\tau(\boldsymbol{\mu})$ is given by

$$\frac{\partial}{\partial \bar{\boldsymbol{\mu}}_i} f_\tau(\boldsymbol{\mu}) = \frac{\sum_{k=1}^{K}\Big(\exp\big(-\tau \mathrm{SE}_k(\boldsymbol{\mu})\big) \frac{\partial}{\partial \bar{\boldsymbol{\mu}}_k} \mathrm{SE}_k(\boldsymbol{\mu})\Big)}{\sum_{k=1}^{K} \exp\big(-\tau \mathrm{SE}_k(\boldsymbol{\mu})\big)}. \tag{38}$$

IV. Complexity and Convergence Analysis of Proposed Methods

A. Complexity Analysis

We now provide the complexity analysis of the proposed algorithms for one iteration using the big-O notation. It is clear that the complexity of Algorithm 1 is dominated by the computation of three quantities: the objective, the gradient, and the projection. It is easy to see that KM multiplications are required to compute $\mathrm{SE}_k(\boldsymbol{\mu})$. Therefore, the complexity of finding $f(\boldsymbol{\mu}) = \sum_{k=1}^{K} \mathrm{SE}_k(\boldsymbol{\mu})$ is $\mathcal{O}(K^2M)$. In a similar way, we can find that the complexity of $\nabla f(\boldsymbol{\mu})$ is also $\mathcal{O}(K^2M)$. The projection of $\boldsymbol{\mu}$ onto $\mathcal{S}$ is given in (25), which requires the computation of the l_2-norm of $K \times 1$ vector $\mathbf{x}_m$ at each AP, and thus the complexity of the projection is $\mathcal{O}(KM)$. In summary, the per-iteration complexity of the proposed algorithm for solving $(\mathcal{P}_1)$ is $\mathcal{O}(K^2M)$. Similarly, the per-iteration complexity for solving $(\mathcal{P}_2)$, $(\mathcal{P}_3)$ and $(\mathcal{P}_4)$ is also $\mathcal{O}(K^2M)$.

To appreciate the low-complexity of the proposed methods, we now provide the per-iteration complexity of the SCA-based

method for solving $(\mathcal{P}_1)$ derived from [16]. Note that the iterative method presented in [16] is dedicated to the problem of total energy efficiency maximization but it can be easily customized to solve $(\mathcal{P}_1)$. Specifically, the convex sub-problem at the iteration $n+1$ of the SCA-based method reads

$$\underset{\boldsymbol{\mu}\geq 0, \mathbf{t}\geq 0}{\text{minimize}} \quad \left(t_1 t_2 \ldots t_K\right)^{1/K} \tag{39a}$$

$$\text{subject to} \quad F(\boldsymbol{\mu}, t_k; \boldsymbol{\mu}^n, t_k^n) \geq \zeta_d N^2 \sum_{i\neq k}^{K} (\bar{\boldsymbol{\nu}}_{ik}^{\mathrm{T}} \bar{\boldsymbol{\mu}}_i)^2 + \zeta_d N \sum_{k=1}^{K} \left\|\bar{\mathbf{D}}_k \bar{\boldsymbol{\mu}}_k\right\|_2^2 + 1, \quad k = 1, \ldots, K \tag{39b}$$

$$\|\boldsymbol{\mu}_m\|^2 \leq \frac{1}{N}, \quad m = 1, \ldots, M, \tag{39c}$$

where

$$F(\boldsymbol{\mu}, t_k; \boldsymbol{\mu}^n, t_k^n) = f(\boldsymbol{\mu}^n, t_k^n) + \nabla_{\boldsymbol{\mu}} f(\boldsymbol{\mu}^n, t_k^n)^{\mathrm{T}} (\boldsymbol{\mu} - \boldsymbol{\mu}^n) + \partial_{t_k} f(\boldsymbol{\mu}^n, t_k^n)(t_k - t_k^n), \tag{40}$$

and

$$f(\boldsymbol{\mu}, t_k) \triangleq \frac{\zeta_d N^2 \sum_{i=1}^{K} (\bar{\boldsymbol{\nu}}_{ik}^{\mathrm{T}} \bar{\boldsymbol{\mu}}_i)^2 + \zeta_d N \sum_{k=1}^{K} \left\|\bar{\mathbf{D}}_k \bar{\boldsymbol{\mu}}_k\right\|_2^2 + 1}{t_k}.$$

We remark that the objective admits a second order cone reformulation and thus (39) is a second order cone program. According to [29, Sect. 6.6.2], the complexity to solve (39) is $\mathcal{O}\big(\sqrt{K+M} M^3 K^4\big)$, which is much larger than $\mathcal{O}(K^2 M)$ for the proposed method, especially when M and K are large.

B. Convergence Analysis

We now discuss the convergence result of Algorithms 1 and 2 for solving $(\mathcal{P}_1)$, which is stated in the following lemma.

Lemma 3: Let $\{\boldsymbol{\mu}^n\}$ be the sequence produced by Algorithms 1 or 2. Then the following properties hold.

- *The sequence of the objective values $\{f(\boldsymbol{\mu}^n)\}$ is nondecreasing and convergent.*
- *The sequence $\{\boldsymbol{\mu}^n\}$ is bounded and any limit point of $\{\boldsymbol{\mu}^n\}$ is a critical point of $(\mathcal{P}_1)$.*

Proof: Please see Appendix C. ■

The same convergence result applies to Algorithms 1 and 2 for solving $(\mathcal{P}_2)$, $(\mathcal{P}_3)$ and $(\mathcal{P}_4)$.

V. Numerical Results

In this section, we evaluate the performance of the proposed methods in terms of computational complexity and achieved spectral efficiency. All simulations results are obtained using Algorithm 2 since it has a faster convergence rate. The users and the APs are uniformly dropped over a $D \times D$ km^2. The large-scale fading coefficient between the m-th AP and the k-th user is generated as

$$\beta_{mk} = \mathrm{PL}_{mk}.z_{mk},$$

where PL_{mk} and z_{mk} represent the path loss and log-normal shadowing with mean zero and standard deviation σ_{sh}, respectively. In this paper, we adopt the three-slope path loss model as in [16], in which PL_{mk} (in dB) equals $-L - 15\log_{10}(d_1) - 20\log_{10}(d_0)$ if $d_{mk} < d_0$, equals $-L - 15\log_{10}(d_1) - 20\log_{10}(d_{mk})$ if $d_0 < d_{mk} < d_1$, and equals $-L - 35\log_{10}(d_{mk})$ otherwise, where L is a constant dependent on carrier frequency, d_{mk} (in km) is the distance between the m-th AP and the k-th user, and d_0 and d_1 (both in km) are reference distances. The pilot sequences are generated as follows. First, note that we can create T_p mutually orthogonal vectors of length T_p (in symbols), e.g using the singular vectors resulting from the singular value decomposition of a square complex random matrix of size T_p. Thus, when $K \leq T_p$, we simply select K out of these T_p orthogonal vectors as pilot sequences for K users. In this way, pilot sequences are mutually orthogonal and thus pilot contamination does not occur. However, when $K > T_p$, it is impossible to obtain orthogonal pilot sequences among all users. In such a case, we randomly assign these T_p mutually orthogonal vectors to K users. Similar to [16], we choose the parameters given in Table I, if not stated otherwise.

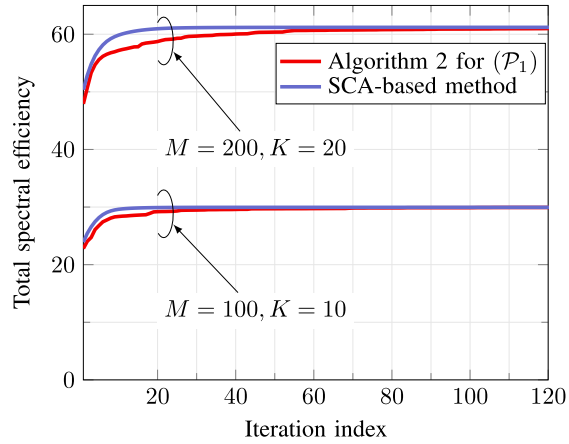
TABLE I

Table of Simulation Parameters

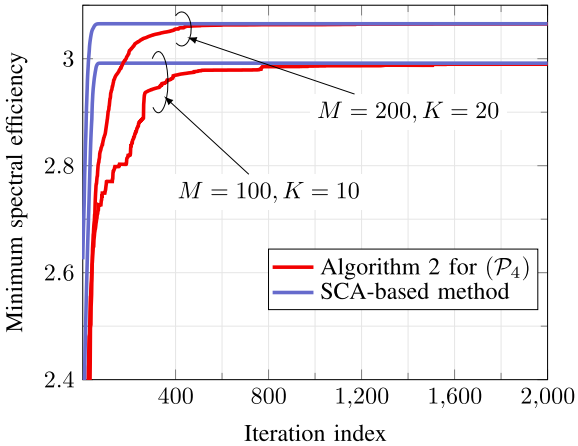
Parameter	Value	Parameter	Value
L	140.7 dB	Noise figure	9 dB
d_0	0.01 km	T_p	20
d_1	0.05 km	T_c	200
Bandwidth	20 MHz	ζ_d	1 W
N_0	174 (dBm/Hz)	ζ_p	0.2 W

In the first numerical experiment, we compare the convergence rate of the proposed method with the SCA-based method presented in [16] as explained in Section IV-A. To solve (39), we use convex conic solver MOSEK [30] through the modeling tool YALMIP [31]. In particular, Figs. 1(a) and 1(b) show the convergence of Algorithm 2 and the SCA-based method for the total spectral efficiency and the min-rate maximization problem, respectively. We can see that Algorithm 2 and SCA-based methods achieve the same performance but the SCA-based method requires fewer iterations to return a solution. However, the main advantage of Algorithm 2 over the SCA-based method is that each iteration of the proposed method is very memory efficient and can be done by closed-form expressions, and hence, is executed very fast. As a result, the total run-time of the proposed method is far less than that of the SCA-based method as shown in Table II. In Table II, we report the actual run-time of both methods to solve the SEmax problem. Here, we execute our codes on a 64-bit Windows operating system with 16 GB RAM and Intel CORE i7, 3.7 GHz. Both iterative methods are terminated when the difference of the objective for the last 5 iterations is less than 10^{-3}.

We next take advantage of the proposed methods to explore the spectral efficiency performance of large-scale cell-free massive MIMO that can cover, e.g. a large metropolitan area in our vision. In particular, we investigate the performance of cell-free massive MIMO for two cases $D = 1$ and $D = 10$. To obtain a fair comparison, we keep the AP density (defined as the number of APs per square kilometer) same for both cases. Note that the AP density of 1000 means there are 10 000 APs for the case of $D = 10$, which has not been studied in the



(a) SEmax (i.e. $(\mathcal{P}_1)$)



(b) Max-min fairness (i.e. $(\mathcal{P}_4)$)

Fig. 1. Total SE and minimum SE versus the number of iterations. The values of M and K are given explicitly in the figure. Each AP is equipped with a single antenna.

TABLE II

COMPARISON OF RUN-TIME (IN SECONDS) BETWEEN ALGORITHM 2 AND THE SCA-BASED METHOD FOR SOLVING THE SEMAX PROBLEM. THE VALUES OF OTHER PARAMETERS ARE TAKEN AS $K = 40, N = 1$ AND $D = 1$

number of APs	SCA-based method	Algorithm 2
200	330.84	**2.88**
400	408.94	**9.42**
800	1115.18	**18.07**
1600	1648.09	**49.45**

literature previously. To appreciate the proposed method for this large-scale scenario, we compare it with the SCA method and the equal power allocation (EPA) method where the power control coefficient η_{mk} is given by, $\eta_{mk} = (\sum_{i=1}^{k} \nu_{mi})^{-1}$. The results in Fig. 2 are interesting. First, increasing the AP density expectedly improves the total spectral efficiency of the system. Second, for the same AP density, a larger area provides a better sum spectral efficiency. The reason is that for a larger area, the users that are served by the APs become far apart each other. As a result, the inter-user inference becomes weaker, leading to an improved sum spectral efficiency. On the other hand, the EPA method yields smaller spectral efficiency as the in terms of the achieved utility performance.

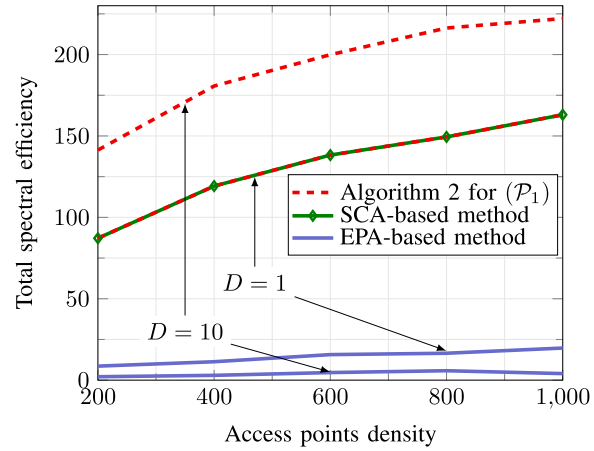


Fig. 2. Total spectral efficiency versus AP density. The number of users is $K = 40$.

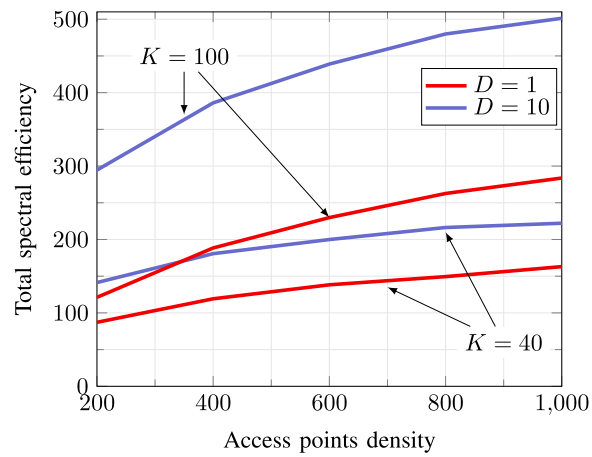
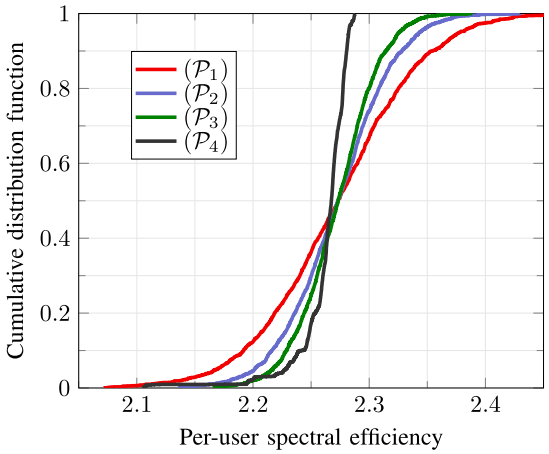


Fig. 3. Total spectral efficiency versus AP density and number of users.

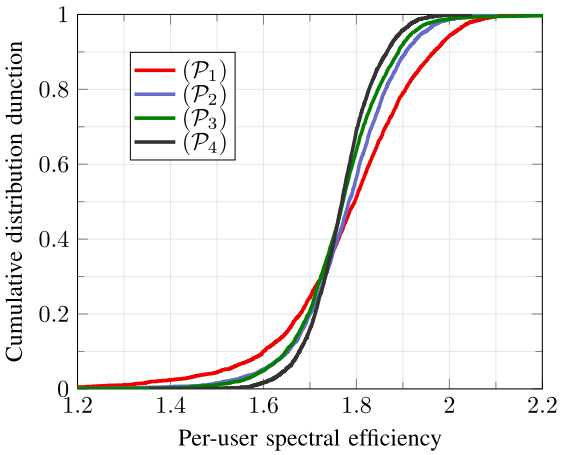
coverage area is larger because more power should be to spent to the users having small path loss. Although, the SCA-based method produces the same spectral efficiency as the proposed APG method for $D = 1$, it is unable to run for $D = 10$ on the system specifications mentioned above, for which the proposed method is still viable. The above facts show that the proposed scheme outperforms the SCA-based method in terms of the required run time, and outperforms the EPA methods in terms of the achieved utility performance.

In Fig. 3, we again investigate the spectral efficiency performance of the AP density but for different number of users (i.e. $K = 100$ and $K = 40$). It can be seen that for a given AP density, the SE is increased if the number of users becomes larger. The gain is more profound for larger AP density due to the fact that more APs allow for more efficient exploitation of multiuser diversity gains.

Next we plot the cumulative distribution function (CDF) of the per-user SE obtained by the four considered utility functions. Two settings are examined: $M = 100$, $K = 20$, $T_c = 200$ symbols and $T_p = 20$ symbols (cf. Fig. 4(a)), and $M = 2000$, $K = 500$, $T_c = 1000$ and $T_p = 200$



(a) Small-scale problem: $K = 20$ and $M = 100$



(b) Large-scale problem: $K = 500$ and $M = 2000$

Fig. 4. CDF of per-user spectral efficiency for $(\mathcal{P}_1)$-$(\mathcal{P}_4)$ for small scale and large-scale problems.

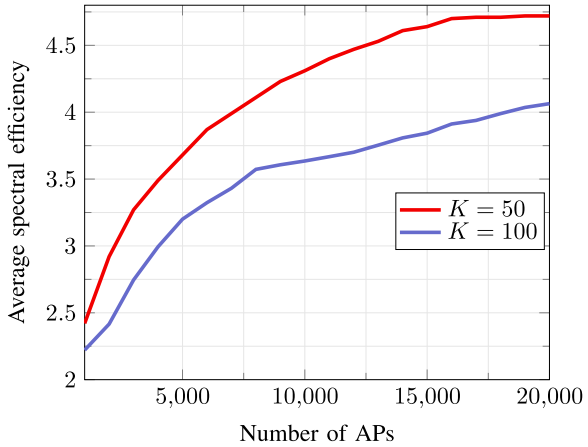


Fig. 5. Average spectral efficiency versus number of APs for $K = 100$ and $K = 50$ users for $D = 1$.

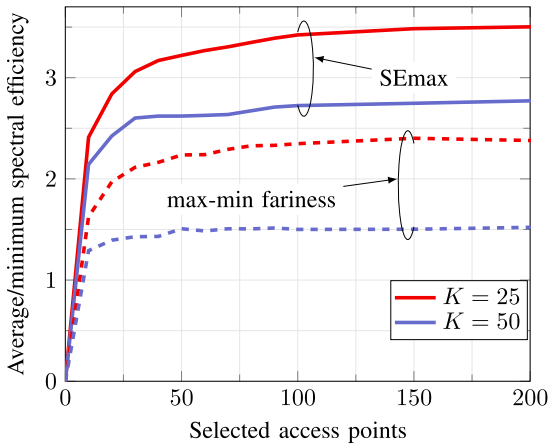


Fig. 6. Average spectral efficiency versus number of selected APs for $M = 500$ and $D = 1$.

(cf. Fig. 4(b)). We can observe in Fig. 4 that the median values of the total achieved SEs are more or less the same for all the four utility functions. The 95%-likely achievable downlink SE is decreasing from $(\mathcal{P}_1)$ to $(\mathcal{P}_4)$ which can be explained by the fact that the following inequality holds: $\mathrm{SE}^{(\mathcal{P}_1)} > \mathrm{SE}^{(\mathcal{P}_2)} > \mathrm{SE}^{(\mathcal{P}_3)} > \mathrm{SE}^{(\mathcal{P}_4)}$, where $\mathrm{SE}^{(\mathcal{P}_i)}$ denotes the per-user spectral efficiency obtained by solving $(\mathcal{P}_i)$ [26]. It is also known that the order is reversed in terms of fairness. Consequently, the CDF of the per-user SE of $(\mathcal{P}_4)$ (i.e. the MRmax problem) has the steepest slope and that of the SEmax problem is more spread. It is particularly interesting to see that the difference on the CDF of the per-user SE of all four utility metrics is marginal for large-scale cell-free massive MIMO. This simulation result again confirms that cell-free massive MIMO can deliver universally good services to all users in the system. In the next experiment, we investigate how the average SE varies as a function of the number of APs. In particular, Fig. 5 shows the average spectral efficiency as a function of the number of APs for $K = 100$ and $K = 50$ users in an area of $D = 1$. We can see that the average SE increases quickly when the number of APs is less 1000 for both $K = 50$ and $K = 100$ users, and it starts to saturate when the number of APs is larger. The reason is that for a given user, there should be a certain number of APs (i.e closest APs) that truly provides macro diversity gains to that user in terms of SE. Thus, a user may be served by a subset of APs to achieve similar SE performance. This can help reduce the overhead in cell-free massive MIMO. Further insights into this are discussed in the next numerical experiment.

In Fig. 6, we consider a scenario with $M = 500$ APs and study how many APs are effectively required for each user. In particular, we plot the average spectral efficiency as a function of the number of selected APs per user for two utility functions: SEmax and MRmax. For a given user, a number of APs is simply selected based on their large-scale fading coefficients. From Fig. 6, we can observe that not all 500 APs are needed to serve 25 or 50 users in an area of 1 km^2. Instead a smaller number of APs per user can can yield nearly the same performance. For example, we only need to assign less than 100 APs to a user to achieve 95% of the SE of the full system.

Finally, we investigate the effect of increasing the number of antennas per AP on the sum SE and minimum SE. Specifically, we plot the average SE and minimum SE with respect to the number of antennas for both $M = 500$ and $M = 1000$ APs. The number of users is fixed to $K = 50$. Expectedly, the SE increases with the increase in the number of antennas per APs

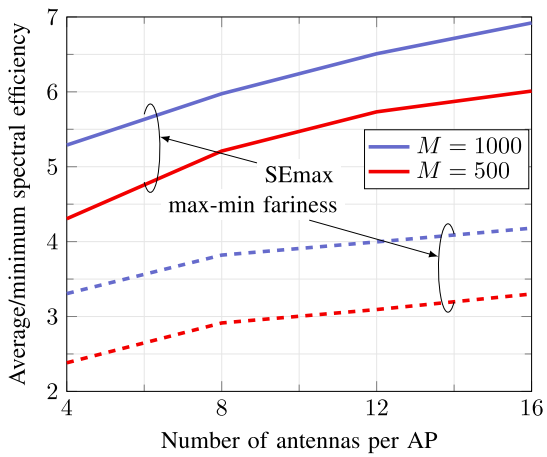


Fig. 7. Average spectral efficiency with respect to the number of antenna at each AP for $K = 50, D = 1$.

but the increase tends to be small when the number of antennas is sufficiently large. The reason is that for a large number of APs channel hardening and favorable propagation can be achieved by a few antennas per AP. Specifically, we can see that the SE for the case of 1000 APs and 4 antennas per AP is larger than the SE for the case of 500 APs and 8 antennas per AP. Therefore, for large-scale cell-free massive MIMO, having more APs with a few antennas each seems to be more beneficial than having fewer APs with more antennas each.

VI. Conclusion

We have considered the downlink of cell-free massive MIMO and aimed to maximize four utility functions, subject to a power constraint at each AP. Conjugate beamforming has been adopted, resulting in a power control problem for which an accelerated project gradient method has been proposed. Particularly, the proposed solutions only requires the first order information of the objective and, in particular, each iteration of the proposed solutions can be computed by closed-form expressions. We have numerically shown that the proposed methods can achieve (nearly) the same SE as a known SCA-based method but with much lesser run time. For the first time, we have evaluated the SE performance of cell-free massive MIMO for an area of 10 km^2, consisting of up to 10 000 APs, whereby the achieved SE can be up to 200 (bit/s/Hz). We have also found that the SE performance of the four utility functions is quite similar for large-scale cell-free massive MIMO, confirming again that cell-free massive MIMO can provide uniformed services to all users.

Appendix A
Solution to the Projection Onto $\mathcal{S}$

The solution of (24) can be found using the KKT conditions given by

$$\nabla_{\boldsymbol{\mu}_m}\mathcal{L} = 2(\boldsymbol{\mu}_m - \mathbf{x}_m) + 2\lambda\boldsymbol{\mu}_m = 0, \tag{41a}$$

$$\lambda(||\boldsymbol{\mu}_m||^2 - \frac{1}{N}) = 0, \tag{41b}$$

$$||\boldsymbol{\mu}_m||^2 \leq \frac{1}{N}, \tag{41c}$$

$$\lambda \geq 0. \tag{41d}$$

Applying the constraint $\boldsymbol{\mu}_m \geq 0$ to (41a), we get

$$\mathbf{x}_m = (1+\lambda)\boldsymbol{\mu}_m \geq 0. \tag{42}$$

If $\lambda = 0$, the stationary condition in (41a) results in

$$\boldsymbol{\mu}_m = \mathbf{x}_m, \tag{43}$$

and (41c) gives $\boldsymbol{\mu}_m = \mathbf{x}_m \leq \frac{1}{N}$, which corresponds to the case where $\mathbf{x}_m$ lies in $\mathcal{S}$. If $\lambda > 0$, the complementary slackness in (41b) implies that $||\boldsymbol{\mu}_m||^2 = \frac{1}{N}$. The equality in (42) can further be written as

$$\boldsymbol{\mu}_m^{\mathrm{T}}\mathbf{x}_m = (1+\lambda)\boldsymbol{\mu}_m^{\mathrm{T}}\boldsymbol{\mu}_m = (1+\lambda)\frac{1}{N}. \tag{44}$$

Since $\lambda > 0$, we get $N\boldsymbol{\mu}_m^{\mathrm{T}}\mathbf{x}_m - 1 > 0$ or $\boldsymbol{\mu}_m^{\mathrm{T}}\mathbf{x}_m > \frac{1}{N}$. By using the Cauchy–Schwartz inequality, we can write $||\boldsymbol{\mu}_m||\,||\mathbf{x}_m|| > \frac{1}{N}$ or $||\mathbf{x}_m||^2 > \frac{1}{N}$, which refers to the case where $\mathbf{x}_m$ lies outside $\mathcal{S}$. Furthermore, substituting $\lambda = N\boldsymbol{\mu}_m^{\mathrm{T}}\mathbf{x}_m - 1$ into (42), we get

$$\mathbf{x}_m = N(\boldsymbol{\mu}_m^{\mathrm{T}}\mathbf{x}_m)\boldsymbol{\mu}_m, \tag{45}$$

which means that $\boldsymbol{\mu}_m$ is parallel to $\mathbf{x}_m$ or $\boldsymbol{\mu}_m = a\mathbf{x}_m$ where a is a constant. Next using (45) gives $a = \frac{1/\sqrt{N}}{||\mathbf{x}_m||^2}$ and therefore

$$\boldsymbol{\mu}_m = \frac{1/\sqrt{N}}{||\mathbf{x}_m||^2}\mathbf{x}_m. \tag{46}$$

Combining (42), (43) and (46) results in

$$\mu_m = \frac{\sqrt{1/N}}{\max\left(\sqrt{1/N}, ||[\mathbf{x}_m]_+||\right)}[\mathbf{x}_m]_+, \tag{47}$$

which completes the proof.

Appendix B
Lipschitz Constant of $\nabla f(\boldsymbol{\mu})$

Assessing the Lipschitz constant of $\nabla f(\boldsymbol{\mu})$ for the four problems discussed in the paper boils down to finding the Lipschitz constant of $\nabla \mathrm{SE}_k(\boldsymbol{\mu})$. A convenient way to do this is to rewrite the function $\mathrm{SE}_k(\boldsymbol{\mu})$ in the form of single variable $\boldsymbol{\mu}$ which can be done by denoting $\mathbf{A}_i \triangleq \mathbf{I}_M \otimes \mathbf{e}_i^{\mathrm{T}}$. Then, we can write $\bar{\boldsymbol{\mu}}_i$ as $\mathbf{A}_i\boldsymbol{\mu}$, and thus $\gamma_k(\boldsymbol{\mu})$, the SINR of the k-th user, can be rewritten as

$$\gamma_k(\boldsymbol{\mu}) = \frac{\overbrace{\zeta_d(\bar{\boldsymbol{\nu}}_{kk}^{\mathrm{T}}\mathbf{A}_k\boldsymbol{\mu})^2}^{b_k(\boldsymbol{\mu})}}{\underbrace{\zeta_d\left(\sum_{i\neq k}^{K}(\bar{\boldsymbol{\nu}}_{ik}^{\mathrm{T}}\mathbf{A}_i\boldsymbol{\mu})^2 + \frac{1}{N}\sum_{i=1}^{K}||\bar{\mathbf{D}}_k\mathbf{A}_i\boldsymbol{\mu}||_2^2\right)}_{c_k(\boldsymbol{\mu})} + \frac{1}{N^2}}. \tag{48}$$

The gradient of $\mathrm{SE}_k(\boldsymbol{\mu})$ with respect to $\boldsymbol{\mu}$ is found as

$$\begin{aligned}\nabla \mathrm{SE}_k(\boldsymbol{\mu}) &= \nabla_{\boldsymbol{\mu}}\log(1+\gamma_k(\boldsymbol{\mu})) \\ &= \frac{\nabla_{\boldsymbol{\mu}}\left(b_k(\boldsymbol{\mu}) + c_k(\boldsymbol{\mu})\right)}{b_k(\boldsymbol{\mu}) + c_k(\boldsymbol{\mu}) + \frac{1}{N^2}} - \frac{\nabla_{\boldsymbol{\mu}}c_k(\boldsymbol{\mu})}{c_k(\boldsymbol{\mu}) + \frac{1}{N^2}},\end{aligned} \tag{49}$$

where $\nabla_{\boldsymbol{\mu}} b(\boldsymbol{\mu})$ and $\nabla_{\boldsymbol{\mu}} c(\boldsymbol{\mu})$ are given by

$$\nabla_{\boldsymbol{\mu}} b_k(\boldsymbol{\mu}) = 2\zeta_d \mathbf{A}_k^{\mathrm{T}} \bar{\boldsymbol{\nu}}_{kk} \bar{\boldsymbol{\nu}}_{kk}^{\mathrm{T}} \mathbf{A}_k \boldsymbol{\mu}, \tag{50}$$

$$\nabla_{\boldsymbol{\mu}} c_k(\boldsymbol{\mu}) = 2\zeta_d \left(\sum_{i \neq k}^{K} \mathbf{A}_i^{\mathrm{T}} \bar{\boldsymbol{\nu}}_{ik} \bar{\boldsymbol{\nu}}_{ik}^{\mathrm{T}} \mathbf{A}_i + \frac{1}{N} \sum_{i=1}^{K} \mathbf{A}_i^{\mathrm{T}} \bar{\mathbf{D}}_k^2 \mathbf{A}_i \right) \boldsymbol{\mu}. \tag{51}$$

Now the Lipschitz continuity of $\nabla_{\boldsymbol{\mu}} b_k(\boldsymbol{\mu})$ and $\nabla_{\boldsymbol{\mu}} c_k(\boldsymbol{\mu})$ is obvious, and so is that of $\nabla \mathrm{SE}_k(\boldsymbol{\mu})$. Although we can compute the Lipschitz constant of $\nabla \mathrm{SE}_k(\boldsymbol{\mu})$, this is quite involved and not necessary since a line search is used to find a proper step size.

Appendix C
Convergence Proof of Algorithm 1

The proof is due to [23]. First, we note that since $\nabla f(\boldsymbol{\mu})$ is Lipschitz continuous and a line search is used to find a proper step size in Algorithm 2, it is sufficient to prove the convergence Algorithm 1. We begin with by recalling an important inequality of a L-smooth function. Specifically, for a function $f(\mathbf{x})$ has the Lipschitz continuous gradient with a constant L_f, the following inequality holds

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \left\langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \right\rangle - \frac{L_f}{2} ||\mathbf{y} - \mathbf{x}||^2. \tag{52}$$

The projection in Step 5 of Algorithm 1 can be written as

$$\begin{aligned} \mathbf{v}^{n+1} &= \arg\min_{\boldsymbol{\mu} \in \mathcal{S}} \left\| \boldsymbol{\mu} - \boldsymbol{\mu}^n - \alpha \nabla f(\boldsymbol{\mu}^n) \right\|^2 \\ &= \arg\max_{\boldsymbol{\mu} \in \mathcal{S}} \left\langle \nabla f(\boldsymbol{\mu}^n), \boldsymbol{\mu} - \boldsymbol{\mu}^n \right\rangle - \frac{1}{2\alpha} ||\boldsymbol{\mu} - \boldsymbol{\mu}^n||^2, \end{aligned} \tag{53}$$

where we have used the fact that $||\mathbf{a} - \mathbf{b}||^2 = ||\mathbf{a}||^2 + ||\mathbf{b}||^2 + 2\left\langle \mathbf{a}, \mathbf{b} \right\rangle$. Note that when $\boldsymbol{\mu} = \boldsymbol{\mu}^n$, the objective in the above problem is 0, and thus we have

$$\left\langle \nabla f(\boldsymbol{\mu}^n), \mathbf{v}^{n+1} - \boldsymbol{\mu}^n \right\rangle - \frac{1}{2\alpha} ||\mathbf{v}^{n+1} - \boldsymbol{\mu}^n||^2 \geq 0. \tag{54}$$

Applying (52) yields

$$\begin{aligned} f(\mathbf{v}^{n+1}) &\geq f(\boldsymbol{\mu}^n) + \left\langle \nabla f(\boldsymbol{\mu}^n), \mathbf{v}^{n+1} - \boldsymbol{\mu}^n \right\rangle \\ &\quad - \frac{L_f}{2} \left\| \mathbf{v}^{n+1} - \boldsymbol{\mu}^n \right\|^2 \\ &\geq f(\boldsymbol{\mu}^n) + \left(\frac{1}{2\alpha} - \frac{L_f}{2} \right) \left(\left\| \mathbf{v}^{n+1} - \boldsymbol{\mu}^n \right\|^2 \right. . \end{aligned} \tag{55}$$

It is easy to see that $f(\mathbf{v}^{n+1}) \geq f(\boldsymbol{\mu}^n)$ if $\alpha < \frac{1}{L_f}$. From Step 6, if $f(\mathbf{z}^{n+1}) \geq f(\mathbf{v}^{n+1})$, then

$$\boldsymbol{\mu}^{n+1} = \mathbf{z}^{n+1}, f(\boldsymbol{\mu}^{n+1}) = f(\mathbf{z}^{n+1}) \geq f(\mathbf{v}^{n+1}). \tag{56}$$

Similar if $f(\mathbf{z}^{n+1}) < f(\mathbf{v}^{n+1})$, then

$$\boldsymbol{\mu}^{n+1} = \mathbf{v}^{n+1}, f(\boldsymbol{\mu}^{n+1}) = f(\mathbf{v}^{n+1}). \tag{57}$$

From (55), (56), and (57) we have

$$f(\boldsymbol{\mu}^{n+1}) \geq f(\mathbf{v}^{n+1}) \geq f(\boldsymbol{\mu}^n). \tag{58}$$

Since the feasible set of the considered problems is compact convex, the iterates $\{\mathbf{v}^n\}$ and $\{\boldsymbol{\mu}^n\}$ are both bounded and thus $\{\boldsymbol{\mu}^n\}$ has accumulation points. As shown above, $f(\boldsymbol{\mu}^n)$ is nondecreasing, f has the same value, denoted by f^*, at all the accumulation points. From (55), we have

$$\begin{aligned} f(\boldsymbol{\mu}^{n+1}) - f(\boldsymbol{\mu}^n) &\geq f(\mathbf{v}^{n+1}) - f(\boldsymbol{\mu}^n) \\ &\geq \left(\frac{1}{2\alpha} - \frac{L_f}{2} \right) \left(\left\| \mathbf{v}^{n+1} - \boldsymbol{\mu}^n \right\|^2 \right. , \end{aligned} \tag{59}$$

which results in

$$\infty > f^* - f(\boldsymbol{\mu}^1) \geq \sum_{n=1}^{\infty} \left(\frac{1}{2\alpha} - \frac{L_f}{2} \right) \left(\left\| \mathbf{v}^{n+1} - \boldsymbol{\mu}^n \right\|^2 \right. . \tag{60}$$

Since $\alpha < \frac{1}{L_f}$, we can conclude that

$$\left\| \mathbf{v}^{n+1} - \boldsymbol{\mu}^n \right\| \to 0 \quad \text{as } n \to \infty. \tag{61}$$

The convergence proof of Algorithm 1 to a critical point of $(\mathcal{P}_1)$ follows the same arguments as those in [23] and thus, is omitted for the sake of brevity.

References

[1] M. Farooq, H. Q. Ngo, and L. N. Tran, "Accelerated projected gradient method for the optimization of cell-free massive MIMO downlink," in *IEEE PIMRC*, Oct. 2020, pp. 1–6.

[2] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, pp. 585–598, Nov. 1999.

[3] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[4] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[5] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[6] F. Boccardi, R. W. Heath, Jr., A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[7] E. Dahlman, S. Parkvall, and J. Skold, *5G NR: The Next Generation Wireless Access Technology*, 1st ed. New York, NY, USA: Academic, 2018.

[8] A. Lozano, R. W. Heath, Jr., and J. G. Andrews, "Fundamental limits of cooperation," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5213–5226, Sep. 2013.

[9] H. Yang and T. L. Marzetta, "A macro cellular wireless network with uniformly high user throughputs," in *Proc. IEEE VTC-Fall*, Sep. 2014, pp. 1–5.

[10] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, "Massive MIMO is a reality—What is next?: Five promising research directions for antenna arrays," *Digit. Signal Process.*, vol. 94, pp. 3–20, Nov. 2019.

[11] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.

[12] A. Papazafeiropoulos, P. Kourtessis, M. D. Renzo, S. Chatzinotas, and J. M. Senior, "Performance analysis of cell-free massive MIMO systems: A stochastic geometry approach," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 3523–3537, Apr. 2020.

[13] Z. Chen and E. Björnson, "Channel hardening and favorable propagation in cell-free massive MIMO with stochastic geometry," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5205–5219, Nov. 2018.

[14] L. D. Nguyen, T. Q. Duong, H. Q. Ngo, and K. Tourki, "Energy efficiency in cell-free massive MIMO with zero-forcing precoding design," *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1871–1874, Aug. 2017.

[15] S. Buzzi and A. Zappone, "Downlink power control in user-centric and cell-free massive MIMO wireless networks," in *Proc. IEEE PIMRC*, Oct. 2017, pp. 1–6.

[16] H. Q. Ngo, L.-N. Tran, T. Q. Duong, M. Matthaiou, and E. G. Larsson, "On the total energy efficiency of cell-free massive MIMO," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 1, pp. 25–39, Mar. 2018.

[17] G. Interdonato, H. Q. Ngo, P. Frenger, and E. G. Larsson, "Downlink training in cell-free massive MIMO: A blessing in disguise," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5153–5169, Nov. 2019.

[18] M. Alonzo, S. Buzzi, A. Zappone, and C. D'Elia, "Energy-efficient power control in cell-free and user-centric massive MIMO at millimeter wave," *IEEE Trans. Green Commun. Netw.*, vol. 3, no. 3, pp. 651–663, Sep. 2019.

[19] M. Bashar, K. Cumanan, A. G. Burr, M. Debbah, and H. Q. Ngo, "On the uplink max–min SINR of cell-free massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2021–2036, Apr. 2019.

[20] H. V. Nguyen *et al.*, "On the spectral and energy efficiencies of full-duplex cell-free massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1698–1718, Aug. 2020.

[21] Z. Chen, E. Björnson, and E. G. Larsson, "Dynamic resource allocation in co-located and cell-free massive MIMO," *IEEE Trans. Green Commun. Netw.*, vol. 4, no. 1, pp. 209–220, Mar. 2020.

[22] F. Rezaei, C. Tellambura, A. A. Tadaion, and A. R. Heidarpour, "Rate analysis of cell-free massive MIMO-NOMA with three linear precoders," *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3480–3494, Jun. 2020.

[23] H. Li and Z. Lin, "Accelerated proximal gradient methods for non-convex programming," in *Advances in Neural Information Processing Systems*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2015, pp. 379–387.

[24] L.-N. Tran and H. Q. Ngo, "First-order methods for energy-efficient power control in cell-free massive MIMO: Invited paper," in *Proc. IEEE ACSSC*, Nov. 2019, pp. 848–852.

[25] J. Zhang, S. Chen, Y. Lin, J. Zheng, B. Ai, and L. Hanzo, "Cell-free massive MIMO: A new next-generation paradigm," *IEEE Access*, vol. 7, pp. 99878–99888, 2019.

[26] Z.-Q. Luo and S. Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 57–73, Feb. 2008.

[27] H. H. Bauschke, M. N. Bui, and X. Wang, "Projecting onto the intersection of a cone and a sphere," *SIAM J. Optim.*, vol. 28, no. 3, pp. 2158–2188, Jan. 2018.

[28] Y. Nesterov, "Smooth minimization of non-smooth functions," *Math. Program., A*, vol. 103, no. 1, pp. 127–152, 2005.

[29] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2001, doi: 10.1137/1.9780898718829.

[30] *The MOSEK Optimization Toolbox for MATLAB Manual Version 7.1 (Revision 28)*, MOSEK ApS, Copenhagen, Denmark, 2015.

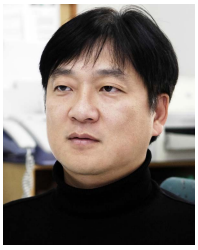
[31] J. Löfberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," in *Proc. CACSD Conf.*, 2004, pp. 284–289.

Muhammad Farooq (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from the Department of Electrical Engineering, University of Engineering and Technology Lahore, Pakistan, in 2014 and 2018, respectively. He is currently pursuing the Ph.D. degree with the School of Electrical and Electronic Engineering, University College Dublin, Ireland.

He has worked as a Lecturer with the University of Engineering and Technology Lahore in 2018 and the University of Sargodha, Pakistan, from 2015 to 2018. He has authored and coauthored conference papers in IEEE PIMRC 2020, IWCMC 2020, IEEE VTC 2021-Spring, and EUSIPCO 2021. His main research interests include optimization techniques, massive MIMO, signal processing, machine learning, physical layer security, and the Internet of Things. He also works as a Reviewer for IEEE COMMUNICATION LETTERS, IEEE WIRELESS COMMUNICATION LETTERS, and IEEE conferences.

Hien Quoc Ngo (Senior Member, IEEE) is currently a Lecturer with Queen's University Belfast, U.K. He has coauthored many research articles in wireless communications and coauthored the textbook *Fundamentals of Massive MIMO* (Cambridge University Press, 2016). His main research interests include massive (large-scale) MIMO systems, cell-free massive MIMO, physical layer security, and cooperative communications.

Dr. Ngo received the IEEE ComSoc Stephen O. Rice Prize in Communications Theory in 2015, the IEEE ComSoc Leonard G. Abraham Prize in 2017, and the Best Ph.D. Award from EURASIP in 2018. He also received the IEEE Sweden VT-COM-IT Joint Chapter Best Student Journal Paper Award in 2015. He was awarded the UKRI Future Leaders Fellowship in 2019. He serves as an Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE WIRELESS COMMUNICATIONS LETTERS, *Digital Signal Processing*, and *Physical Communication* (Elsevier). He was a Guest Editor of *IET Communications* and a Guest Editor of IEEE ACCESS in 2017.



Een-Kee Hong (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Yonsei University in 1989, 1991, and 1995, respectively. He was a Senior Research Engineer with SK Telecom from September 1995 to February 1999 and a Visiting Senior Engineer with NTT DoCoMo from October 1997 to December 1998. From 2006 to 2007, he was a Visiting Professor with Oregon State University. Since 1999, he has been a Professor and served as the Vice Dean with the College of Electronics and Information Engineering, Kyung Hee University, South Korea. His research interests are in physical layer in wireless communication, radio resource management, and spectrum engineering.

Prof. Hong was awarded the Best Paper Award, the Institute of Information Technology Assessment, and the Haedong Best Paper Award, KICS, and the Order of Merit of the Republic of Korea for his contribution to information and communication.

Le-Nam Tran (Senior Member, IEEE) received the B.S. degree in electrical engineering from the Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam, in 2003, and the M.S. and Ph.D. degrees in radio engineering from Kyung Hee University, Seoul, South Korea, in 2006 and 2009, respectively.

He is currently an Assistant Professor with the School of Electrical and Electronic Engineering, University College Dublin, Ireland. Prior to this, he was a Lecturer with the Department of Electronic Engineering, Maynooth University, Ireland. From 2010 to 2014, he held a postdoctoral positions with the Signal Processing Laboratory, ACCESS Linnaeus Centre, KTH Royal Institute of Technology, Stockholm, Sweden, and the Centre for Wireless Communications, University of Oulu, Finland. His research interests are mainly on applications of optimization techniques for wireless communications design. Some recent topics include energy efficient communications, physical layer security, cloud radio access networks, cell-free massive MIMO, and reconfigurable intelligent surfaces. He has published more than 110 papers in international journals and conference proceedings. He was a recipient of the Career Development Award from Science Foundation Ireland in 2018. He has served on the technical program committees of several IEEE major conferences. He is an Associate Editor of *EURASIP Journal on Wireless Communications and Networking*. He was the Symposium Co-Chair of Cognitive Computing and Networking Symposium of International Conference on Computing, Networking and Communication (ICNC 2016) and the Co-Chair of the Workshop on Scalable Massive MIMO Technologies for Beyond 5G at IEEE ICC 2020.